EVALUATION OF THE SPACE-TIME VARIABILITY OF SOIL SALINITY BY STATISTICAL, GEOSTATISTICAL AND BAYESIAN MAXIMUM ENTROPY METHODS

EVALUATIE VAN DE RUIMTE-TIJD VARIABILITEIT VAN BODEMVERZILTING VIA STATISTISCHE, GEOSTATISTISCHE EN BAYESIAANSE MAXIMUM ENTROPIE METHODEN

by / door

ir. Ahmed DOUAIK

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (Ph.D.) in Applied Biological Sciences
Section: Land- and Forest Management

Proefschrift voorgedragen tot het bekomen van de graad van
Doctor in de Toegepaste Biologische Wetenschappen
Optie: Land- en Bosbeheer

op gezag van
Rector: **Prof. dr. A. DE LEENHEER**

Decaan: Promotoren:

Prof. dr. ir. H. VAN LANGENHOVE   **Prof. dr. ir. M. VAN MEIRVENNE**
                                                                            Dr. T. TÓTH

COPYRIGHT

Gent, March 2005

De auteur / The author   De promotoren / The promoters

ir. A. Douaik   Prof. dr. ir. M. Van Meirvenne   Dr. T. Toth

وَفَوْقَ كُلِّ ذِى عِلْمٍ عَلِيمٌ ..

يوسف - 76

... but over all those endowed with knowledge is the All-Knowing.

YUSUF - 76

CONTENTS

LIST OF FIGURES

Chapter 8

LIST OF TABLES

LIST OF ABBREVIATIONS

TDS   : Total Dissolved Salts

TSS   : Total Soluble Salts

ppm   : part per million (mg/l or μg/ml)

CEC   : Cation Exchange Capacity

ESP   : Exchange Sodium Percentage

SAR   : Sodium Adsorption Ratio

EC   : Electrical Conductivity

EC_a : Bulk soil or Apparent Electrical Conductivity

EC_e : Electrical Conductivity from a saturated paste Extract

EC_x : Electrical Conductivity from a Extract from a given soil:water ratio where x represents the proportion of water, for example 1, 2.5, etc.

EC_θ : Electrical Conductivity determined at the temperature θ

EC25   : Electrical Conductivity determined at the reference temperature of 25°C

EMI   : ElectroMagnetic Induction

TDR   : Time Domain Reflectometry

ATV   : All Terrain Vehicle

GPS   : Global Positioning System

GIS   : Geographic Information System

RS   : Remote Sensing

ANOVA   : ANalysis Of Variance

STRF   : Space Time Random Field

SRF   : Spatial Random Field

TRF   : Temporal Random Field

LMC   : Linear Model of Coregionalization

BME   : Bayesian Maximum Entropy

GHP   : Great Hungarian Plain

HNP   : Hortobagy National Park

CASMM   : Catchment Average Soil Moisture Monitoring

CV   : Coefficient of Variation

OLS   : Ordinary Least Squares

SAM   : Spatial Autoregressive Model

SEM   : Spatial Error Model

SGM   : Spatial General Model

GWR   : Geographically Weighted Regression

Min   : Minimum

Max   : Maximum

HK   : Kriging with Hard data only

HSK   : Kriging with Hard and Soft data

ME   : Mean Error

MSE   : Mean Squared Error

cdf   : cumulative distribution function

pdf   : probability distribution function

SB   : Square of the Bias

MSV   : Mean Squared Variation

LCS   : Lack of positive correlation

SDSD   : Standard Deviation Standard Deviation

HMIK   : Kriging with Hard and Mid Interval data

MMSE   : Minimum Mean Squared Error

LIST OF SYMBOLS

d	: inter electrode distance, in centimetre
R	: resistance, in ohm
E	: potential, in volt
I	: current intensity, in ampere
z	: a given soil attribute, the outcome of a random function
x	: space coordinates
t	: time coordinate
μ	: grand mean, or population mean
$\varepsilon(x, t)$	: random, spatially and temporally uncorrelated error term
n	: sample size
σ^2	: population variance
δ_{ij}	: relative difference at space location i and time instant j
$\overline{\delta}_i$	: temporal average of the relative difference at the space location i
$\sigma\left(\overline{\delta}_i\right)$	: temporal standard deviation of the relative difference at the space location i
r_s	: Spearman rank correlation coefficient
$X(\boldsymbol{p})$	: Space-time random field at the space-time point $\boldsymbol{p}$
χ	: one realization of $X(\boldsymbol{p})$
$F_x(\chi)$	: the cumulative distribution function of $X(\boldsymbol{p})$
$f_x(\chi)$	: the probability distribution function of $X(\boldsymbol{p})$
$m_x(\boldsymbol{p})$	: the mean function of $X(\boldsymbol{p})$
$c_x(\boldsymbol{p}, \boldsymbol{p}')$	: the covariance function of $X(\boldsymbol{p})$ at two points $\boldsymbol{p}$ and $\boldsymbol{p}'$
$\gamma_x(\boldsymbol{p}, \boldsymbol{p}')$	: the variogram of $X(\boldsymbol{p})$ at two points $\boldsymbol{p}$ and $\boldsymbol{p}'$
$c_x(\boldsymbol{h}, \tau)$	: covariance function with spatial lag $\boldsymbol{h}$ and temporal lag τ
$\gamma_x(\boldsymbol{h}, \tau)$	: variogram with spatial lag $\boldsymbol{h}$ and temporal lag τ
$\hat{c}_x(\boldsymbol{h}, \tau)$	: moments-based estimator of the covariance function
$\hat{\gamma}_x(\boldsymbol{h}, \tau)$	: moments-based estimator of the variogram
φ	: space-time anisotropy ratio
ψ	: space anisotropy ratio
$c_j(\boldsymbol{h})$	: spatial covariance function for the time instant j
$\hat{c}_j(\boldsymbol{h})$	: moments-based estimator of the spatial covariance function

$\gamma_j(\boldsymbol{h})$: spatial variogram for the time instant j

$\hat{\gamma}_j(\boldsymbol{h})$: moments-based estimator of the spatial variogram

$c_i(\tau)$: temporal covariance function for the space location i

$\hat{c}_i(\tau)$: moments-based estimator of the temporal covariance function

$\gamma_i(\tau)$: temporal variogram for the space location i

$\hat{\gamma}_i(\tau)$: moments-based estimator of the temporal variogram

$c_{jj'}(h)$: spatial cross covariance function between time instants j and j'

$\hat{c}_{jj'}(h)$: moments-based estimator of the spatial cross covariance function

$\gamma_{jj'}(\boldsymbol{h})$: spatial cross variogram between time instants j and j'

$\hat{\gamma}_{jj'}(\boldsymbol{h})$: moments-based estimator of the spatial cross variogram

$c_{ii'}(\tau)$: temporal cross covariance function between space location i and i'

$\hat{c}_{ii'}(\tau)$: moments-based estimator of the temporal cross covariance function

$\gamma_{ii'}(\tau)$: temporal cross variogram between space location i and i'

$\hat{\gamma}_{ii'}(\tau)$: moments-based estimator of the temporal cross variogram

$\gamma_{21}{}^{p}(\boldsymbol{h})$: spatial pseudo cross variogram between time instants t_2 and t_1

$\hat{\gamma}_{21}{}^{p}(\boldsymbol{h})$: moments-based estimator of the spatial pseudo cross variogram

χ_{soft} : soft data

χ_{hard} : hard data

χ_{data} : total (hard and soft) data

$f_G(\boldsymbol{\chi})$: G-based or prior pdf

$f_S(\chi_{\text{soft}})$: S-based pdf

$f_K(\chi_{\text{k}})$: posterior pdf at the space-time point χ_k

PART I

GENERAL INTRODUCTION, LITERATURE REVIEW, DATA DESCRIPTION AND METHODOLOGY

CHAPTER 1

GENERAL INTRODUCTION

1.1. Overview

To satisfy the food needs in the future of an increasing population, two options are possible: improving the management of soils already under cultivation or using the potentially arable soils which are not yet cultivated. The latter option becomes more and more difficult to realize as the potentially arable areas have reduced drastically during the last decades. So every effort should be directed towards the first option. Soil salinity is a major factor which impedes the increase of crop yields. To circumvent its limiting effects, the United Nations Conference on Desertification, held in Nairobi, Kenya, in 1977, adopted the following recommendations: "It is recommended that management measures be taken to combat desertification in the irrigated lands by preventing and controlling waterlogging, salinization, and sodification by modifying techniques to increase productivity in a regular and sustained way; by developing new irrigation and drainage schemas where appropriate always using an integrated approach; and through improvement of the social and economic conditions of people dependent on irrigated agriculture".

It is estimated that one million hectares of irrigated land is abandoned yearly as a consequence of the adverse effects of irrigation due to secondary salinization and sodification. The areas lost due to the natural or primary salinization and sodification need to be added to the above amount.

The effective control of soil salinity and waterlogging requires, among others, the knowledge of the magnitude, the extent, and the distribution of root zone salinity (inventory), the knowledge of the changes and trends of soil salinity over time (monitoring), and the ability to determine the impact of management changes upon saline conditions (Rhoades et al., 1999a).

The sustainability of agriculture is an important concept of soil quality (Warkentin, 1995), which implies that soil quality must be maintained over time (Aon et al., 2001). The initial assessment of spatial and future spatio-temporal changes in soil quality is the basis for evaluating this sustainability (Corwin and Lesch, 2003).

Precision agriculture is one of the potential means of attaining sustainable agriculture. Assessing the impact of soil salinity at regional and local scales is a key component to achieving this sustainable agriculture. This assessment involves the detection of change in soil salinity over time, which shows the usefulness of easily measured soil salinity parameter like the apparent electrical conductivity.

In addition to their use for crop production, most saline and sodic soils have native vegetation providing a wide variety of wild plants, which along with their nature animals, contribute greatly to the biological diversity (Brady and Weil, 1999).

The delineation of field salinity and the knowledge of its variability over an area can assist in designing experiments such as selecting plot size and layout (Van Es et al., 1989; Gotway and Cressie, 1990; Bhatti et al., 1991; Zimmerman and Harville, 1991; Fagroud and Van Meirvenne, 2002; Johnson et al., 2003), sampling strategy (Burgess et al., 1981; Faltman et al., 1987; Fagroud, 2001), or developing practices to manage and reclaim salt-affected soils (Chang et al., 1988).

The spatial variability and the dynamic nature of soil salinity is the outcome of the effects and interactions of different edaphic factors (soil permeability, water table depth, salinity of the groundwater, topography, and parent material), management-related factors (irrigation, drainage, and tillage), and climatic factors (rainfall, wind, and relative humidity). Due to the spatio-temporal variability of soil salinity, numerous samples need to be taken and the measurements need to be repeated as conditions change or to determine if they are changing. We will see that there are ways to optimize the sampling effort if there is temporal persistence or stability.

1.2. Problem definition and research objectives

In the precedent section, we stressed the importance of inventorying and monitoring soil salinity. This soil salinity is customarily evaluated in the laboratory from soil samples by determining the electrical conductivity of an extract from a saturated soil paste (EC_e) or, cheaper and less time consuming, by determining the electrical conductivity from a soil:water ratio (EC_x, where x can be, for example, 1, 2, 2.5, 5, etc., representing the water proportion). We chose to use $EC_{2.5}$ as we were mostly interested in the relative change of soil salinity over time instead of the absolute soil salinity values. Alternatively soil salinity can be evaluated by measuring, in the field, the apparent electrical conductivity, using either the four-electrode or electromagnetic induction devices. The former sensor was selected because at the time of the first sampling campaign (November 1994), the electromagnetic induction device was not yet available and our study focussed on the surface layers of the soil.

A first main objective of our research project was to monitor the change of soil salinity over time. In this sense we used various statistical approaches to reach this objective. First of all,

we used the concept of temporal stability (Vachaud et al., 1985) to check if the spatial pattern of soil salinity can persist from one time instant to another. If it is the case, the sampling effort could be reduced to a limited number of locations representative of low, average, and high saline conditions. The locations representative of these saline conditions were linked to the salt accumulation processes (leaching, accumulation, and waterlogging) and to easily measured covariables like elevation and vegetation pattern in order to identify these locations. The mean temporal change was checked using the Student paired-*t* test and the first test of Lesch et al. (1998) while the dynamic spatial variation, i.e. different change in soil salinity between two time instants for the space locations, was checked using the second test of the last authors.

In the precedent approach we were mainly interested in identifying locations representative of, for example, the average field soil salinity. This average value is required for predictive salinity models (Oosterbaan, 1997). However, land users and managers are also interested in mapping soil salinity in space and/or time by predicting it at unobserved space locations and/or time instants. In order to satisfy this second main objective, we used classical and modern geostatistical methods. In a first tentative, we rescaled the apparent electrical conductivity (EC_a) into soil salinity ($EC_{2.5}$) based on regression equations determined for each time instant. The estimated $EC_{2.5}$ values, obtained by converting the intensively sensed EC_a using the regression equations, were then analyzed for their space time variability and interpolated using the space time kriging. As the rescaled $EC_{2.5}$ values are entailed with uncertainty, we assessed other alternative methods to interpolate soil salinity. We compared the interpolated values obtained from two prediction methods, each with two different algorithms. Space time kriging was used based on only observed $EC_{2.5}$ and on both observed and estimated $EC_{2.5}$. To take into account the difference in accuracy between observed and estimated $EC_{2.5}$, we used the Bayesian maximum entropy method with either interval or probabilistic soft data.

1.3. Outline of the thesis

This thesis is based on a collection of papers that have been published or were submitted for publication in international peer-reviewed journals. Some changes were made to standardize the layout, using the same notations, renumbering figures sand tables, moving the references to the end of the dissertation and to avoid the overlapping between chapters. As the

consequence of the latter point, the different methods applied to the data sets were put together in a separate chapter.

This thesis is organized in four main parts. In the first part, after a general introduction (chapter 1), the literature on soil salinity is reviewed (chapter 2). The focus of the literature review is particularly on how soil salinity is measured and on the selection of appropriate parameters to characterize it. In chapter 3, the study site and the data used in the subsequent analyses are presented. Also, we reviewed in chapter 4 some of the available statistical/geostatistical methods to analyze space time data and focused on the ones that we used in the subsequent parts of this dissertation. Applications of the classical statistical methods (part two) and classical and modern geostatistical methods (part three) followed the literature review. In chapter 5, the temporal stability, based on the concepts of Spearman rank correlation and relative differences, is discussed. A refinement of the concept of relative differences along with the concept of dynamic spatial variation is presented in chapter 6. The third part involves two chapters. The space time kriging of the rescaled soil salinity was compared to the space kriging in chapter 7. The difference in accuracy between observed and estimated soil salinity was considered in chapter 8 for the case of Bayesian maximum entropy (BME) with interval and probabilistic soft data. The two algorithms of BME were compared to space time kriging using only observed soil salinity or using both observed and estimated soil salinity.

Finally chapter 9 summarizes the main results and major conclusions of the thesis, and gives some recommendations for further research emphasizing research problems that still need to be tackled.

CHAPTER 2

SOIL SALINITY: A LITERATURE REVIEW

2.1. Introduction

This chapter reviews briefly the literature on soil salinity whereby section 2.2 deals with total soluble or dissolved salts; soil sodicity is discussed in section 2.3 while soil salinity sensu stricto is presented in section 2.4.

2.2. Total soluble or dissolved salts

There are three forms of inorganic salts in the soil: solid or precipitated, solved in soil solution, and adsorbed on colloid surfaces. Soil salinity refers to the total amount of soluble or dissolved salts in soil. Total soluble (*TSS*) or dissolved (*TDS*) salts are measures of soluble components in soil solution. The most common soluble salts in the natural soils are compounds of the cations Ca^{2+}, Mg^{2+}, K^+, and Na^+, and the anions Cl^-, SO_4^{2-}, CO_3^{2-} and HCO_3^-. The TSS is obtained by weighing the residue after evaporating a given volume of water or soil extract. It is expressed in *ppm* (mg/kg or µg/g of soil). This procedure requires long time, a sensible balance and high temperatures may volatilise some salts. It is for these reasons and others that soil salinity is quantified in terms of electrical conductivity. Before defining soil electrical conductivity and how it is measured in the laboratory and in the field, we make first a distinction between soil sodicity and soil salinity.

2.3. Soil sodicity

2.3.1. Exchangeable sodium percentage

The total amount of exchangeable cations that a soil can retain is termed cation exchange capacity (*CEC*). It is often more convenient to express the relative amount of a given exchange cation adsorbed as a percentage of the *CEC*. Consequently, exchangeable sodium percentage (*ESP*) is calculated as the ratio between exchangeable sodium content and the *CEC* and is expressed as a percentage:

$$ESP = 100 \frac{Na_{exch}}{CEC} \qquad (1)$$

where Na_{exch} represents the exchangeable sodium.

ESP is a measure of soil sodicity and is calculated from direct measurements of each cation (Carter, 1993). It identifies the degree to which the exchangeable complex is saturated with sodium (Brady and Weil, 1999). Its determination is time consuming and the routine analysis has numerous pitfalls. Sodic soils are defined as soils that have an *ESP* > 15 (Richards, 1954).

2.3.2. Sodium adsorption ratio

The Sodium adsorption ration (*SAR*) is another, yet easily measured soil property that is becoming more widely used than *ESP* to identify sodic soils. It gives information on the comparative concentrations of Ca^{2+}, Mg^{2+}, and Na^+ in soil solutions. It is calculated as follows:

$$SAR = \frac{\left[Na^+\right]}{\sqrt{\frac{1}{2}\left(\left[Ca^{2+}\right]+\left[Mg^{2+}\right]\right)}} \tag{2}$$

with $[Na^+]$, $[Ca^{2+}]$, and $[Mg^{2+}]$ representing the concentration of the corresponding cations in milliequivalents per liter.

The *SAR* takes into consideration that the adverse effects of sodium are moderated by the presence of calcium and magnesium (Brady and Weil, 1999). It is considered as a useful index of the sodicity or relative sodium status in soil solutions, aqueous extracts, or water in equilibrium with soil (Carter, 1993). Soils with a *SAR* > 13 are considered to be sodic (SSSA, 1984).

The *ESP* can be estimated from the *SAR*, determined from saturated paste extracts, using the following equation (Kamphorst and Bolt, 1978):

$$0.015 * SAR = \frac{ESP}{100 - ESP} \tag{3}$$

or $ESP = \frac{1.5 * SAR}{1 + 0.015 * SAR} \tag{4}$

For more diluted extracts like 1:5 soil: water ratio, the relation is as follows (Rengasamy et al., 1984):

$$ESP = 1.95 * SAR + 1.8 \tag{5}$$

Sodic soils have a pH > 8.5, the stability of the soil aggregates is deteriorated, the soil colloids disperse and plug the soil's drainage pores implying that the downward percolation of water is prevented. When soil dries out, sodic soils form hard massive structures, which prevent roots to penetrate and the plants are allowed only a small percentage of the total possible soil water

(physiological drought). Also in sodic soils, due to the swelling pressure and dispersion of organic matter at high pH, the sodium-clay complex becomes unstable and deflocculates. It is illuviated into the lower profile when it is flocculated by the higher salt concentration. The B horizon that develops is very poorly drained and has a massive columnar structure. This process of clay defloculation and illuviation is called solonization (sodification) and the soil formed is a solonetz (sodic soil).

2.4. Soil salinity

Traditionally soil salinity is expressed as the electrical conductivity (*EC*) of a solution extracted from soil at water saturation. Its value is given in unit of dS/m. The usual method to quantify the soluble salts concentration in soils is to measure the *EC* of a soil-water extract or in the soil solution. The *EC* refers to the ability of a material or solution to conduct an electric current. Liquids, which carry an electric current, are referred to as electrolytic conductors. Water is a poor conductor. However, when salts are dissolved in it, its aptitude to conduct current increases dramatically.

The electrical conductivity of a given solution will change with a change in temperature. The temperature used as reference is 25°C, and to convert *EC* measured at a temperature θ to *EC* at the reference temperature, the following formula is used:

$$EC_\theta = EC_{25} * \left[1 + 0.021 * (\theta - 25)\right] \tag{6}$$

with θ in degree Celsius,

EC_θ, *EC* at the temperature θ, and

EC_{25}, *EC* at 25°C.

The *EC* can be measured in the laboratory as well as in the field.

2.4.1. Laboratory measurement of the electrical conductivity

Laboratory methods used to quantify soluble salts in soils include:
- measuring *EC* in a saturated paste of soil and water (Richards, 1954),
- measuring *EC* in a soil-water extract based on a fixed soil:water ratio, e.g. 1:1, 1:2, 1:5, etc. The corresponding *EC* will be noted EC_x, *x* is the second part of the ratio.

Ideally, the salt present in the soil should be determined by analysis of the soil water under field moisture conditions. This cannot be done conveniently and the salinity status of the soil

is evaluated from a sample under laboratory conditions. A soil sample is saturated with distilled water, then vacuum filtration is used to obtain the extract, the latter is used for the determination of the electrical conductivity, whereas ordinary filtration or centrifugation can be used when higher moisture levels are used.

The saturated paste provides a more representative measurement of *TSS* in the soil solution because it approximates the water content of the soil under field conditions, thus it is more related to plant response. However, this approach, although more precise, is more time consuming, expensive, and more susceptible to errors due to variability between analysts in the preparation of the saturated paste.

In cases where large numbers of soil samples must be processed, *EC* can be measured at a fixed and more dilute soil: water ratio because it is rapid, low cost, easily done, and reproducible across a wide range of soils. Extracts obtained from higher soil: water ratios need to be calibrated with extracts of saturated soil paste. For example, Richards (1954) gave a relationship between EC_e and EC_1. In the early sixties, Agarwal et al. (1961) established a relationship between EC_e and EC_5. Other ratios and media were used. For example, a good relationship was found between, from one side, EC_e and, from the other sides, EC_1 and $EC_{2.5}$ extracts and suspensions (Winsor and Davis, 1956), EC_1 soil extracts and suspensions (Kamaliddin et al., 1961), EC_1 soil extracts (Wagenet and Jurinak, 1978), EC_1 soil suspensions (Fowler and Hamm, 1980), EC_2 soil extracts (McKenzie et al., 1983), EC_1 and EC_2 from soil extracts and suspensions (Hogg and Henry, 1984), $EC_{2.5}$ (Carter and Pearen, 1985), EC_5 (Triantafilis et al., 2000). In addition, Pittman et al (2001) established relationships between the electrical conductivity and ions from 1:1 and saturated paste extracts using regression equations. They found that these soil properties are highly correlated but the correlation depends on ionic charge and soil texture (fine texture with less than 60% of sand as opposed to coarse texture).

When relative salinity is of interest, the soil: water ratios can be used to advantage (Jurinak and Suarez, 1990).

A strong link was found between *EC* and *TSS*. For example, White (1997) found that:

$$TSS(mg\,/\,l) = 640 * EC_e(dS\,/\,m) \qquad (7)$$

while Shaw et al. (1986) and White (1997) showed that:

$$TSS(\%) = 0.34 * EC_{1.5}(dS\,/\,m) \qquad (8)$$

Soil is considered saline when *TSS* is > 2500 ppm or equivalently EC_e > 4 dS/m.

Natural saline soils are called white alkali soils or solonchaks. They have a uniform dark brown Ah horizon containing superficial salt efflorescences. Salts are brought up by the capillary rise of the saline groundwater. The clay remains flocculated and soil structure is stable, provided that a high salt concentration is maintained.

2.4.2. Field measurement of electrical conductivity

As an alternative to the determination of *EC* in the laboratory, soil salinity can be evaluated in the field by reporting the apparent or soil bulk *EC* (EC_a). The latter can be measured in three different ways using electrode probes (Rhoades and Ingvalson, 1971; El Oumri and Vieillefon, 1983), electromagnetic induction (De Jong et al., 1979; Rhoades and Corwin, 1981), and time domain reflectometry (Dalton and Van Genuchten, 1986; Topp et al., 1988). We present them in the following sections.

2.4.2.1. Electrode sensors

This kind of sensors has two main configurations: horizontal or surface array configuration (Rhoades and Ingvalson, 1971, Read and Cameron, 1979), and insertion probes (Rhoades and Van Schilfgaarde, 1976).

2.4.2.1.1. Horizontal or surface array configuration.

In the horizontal configuration, four electrodes are equally spaced, placed on a straight line, and inserted a few centimetres into the soil. The outer electrodes provide a constant electric current while the inner electrodes are used to measure the potential. The depth of current penetration is about equal to the interelectrode distance, *d*. The EC_a is determined using the equation (Rhoades and Ingvalson, 1971):

$$EC_a = \frac{1}{2\pi dR} \qquad (9)$$

with *R*, the resistance in ohms, and *d*, the inter electrode distance in cm.

The soil volume penetrated by the current is given by (Rhoades and Van Schilfgraade, 1976):

$$V = \frac{5\pi d^3}{6} \qquad (10)$$

So the expansion of the inter electrode distance, *d*, increases the depth and volume of soil salinity measurements.

As these sensors require a contact with soil, it is recommended to measure EC_a just following irrigation or rain (so at field capacity). They have the advantages to be simple, rapid, inexpensive, and eliminate the recourse to the tedious soil sampling and laboratory analysis (once an EC_e - EC_a calibration equation is determined).

This method was used to determine soil salinity in artificially salinized plots (Rhoades and Ingvalson, 1971), to identify potential saline seep areas (Halvorson and Rhoades, 1974), and to delineate the surface and subsurface boundaries of encroaching and developed saline seeps (Halvorson and Rhoades, 1976).

2.4.2.1.2. Insertion or salinity probes

The insertion probes are useful when more accurate measurement of soil salinity distribution within the rootzone is desired, particularly when the salinity is not uniform laterally. In this configuration the four electrodes are mounted as annular rings. The probe is inserted into the soil to a desired depth. An electric current with intensity *I* is induced between the outer electrodes and the potential, *E*, is measured between the inner electrodes. The soil resistance, $R = \dfrac{E}{I}$, is converted to EC_a using a cell constant determined experimentally. The volume of soil salinity measurement has the shape of an ellipse encircling the probes.

Soil-depth interval salinities have been estimated using measurements at increasing interelectrode distances in the Horizontal configuration while the insertion probe was used to evaluate the accuracy of such estimates. The agreement between measurements from both configurations is excellent. However the Horizontal array is less suitable when the soil is highly stratified or nonhomogeneous laterally while the insertion probe is not so limited and is more generally reliable in such cases.

Comparing the Horizontal array and the insertion probe, Rhoades and Halvorson (1976) concluded that the latter can be used to more accurately determine soil salinity of a discrete depth interval and very useful for establishing EC_e - EC_a calibration equations while the former is better suited to provide an index of bulk soil salinity.

The measured EC_a is function of many soil properties like moisture content, texture, and salt content (Halvorson et al., 1977).

2.4.2.2. Electromagnetic induction sensors

With electromagnetic induction the current is supplied to the soil through induction. The sensor employs two kinds of coils: a transmitter and a receiver. As the transmitter coil is energized, circular currents are induced in the soil. The magnitude of loops of the induced current is proportional to EC_a. The current loops produce a magnetic field, the strength of which is proportional to the current flowing in the loops. A fraction of the magnetic field is intercepted by the receiver and creates an output related to EC_a. The soil depth characterised is function of the frequency of the transmitter, the intercoil spacing, the coil orientation relative to the soil surface, and of the soil conductivity.

There are two main commercially available instruments: EM31 and EM38. For the EM31, the intercoil spacing is 3.70 m. In vertical position (coils perpendicular to the soil surface), the measurement depth (6 m) is twice as deep as in horizontal position (coils parallel to the soil surface). The variation in EC_a with depth can be assessed by varying the height of the instrument and the coil orientation relative to the soil surface (horizontal or vertical position). The technique has considerable potential for providing a rapid and easy method for detecting and mapping saline areas. However it has less depth-resolution than the horizontal array instrument.

For EM38, the intercoil spacing is one meter. When placed at ground level, it permits EC_a measurement to effective depths of approximately one and two meters when the instrument is in horizontal and vertical positions, respectively. When the instrument is placed in the horizontal position, 75% of its response comes from less than 90 cm while in the vertical position; the corresponding depth is less than 190 cm. The instrument can be used to establish EC_a -depth distribution by holding the device at various heights above the soil surface.

Job et al. (1995) presented an EMI device with an intercoil distance of 0.60 m, smaller and easily manipulable than the existing instruments. It is more sensitive to salinity of the top one meter of soil and less sensible to the salinity of the groundwater.

2.4.2.3. Time domain reflectometry

The EC_a measurement is based on the velocity of propagation and reflection of an electromagnetic pulse along transmission lines. The dielectric constant and the impedance of the medium (the soil) can be measured once the transmission lines are placed inside the soil. The first parameter (the dielectric constant) is function of the soil water content while the

impedance of the transmission lines in the soil produces an attenuation of the signal, the latter can be related to EC_a. The TDR can be advantageously used to measure simultaneously both water content and electrical conductivity of the soil (Dalton et al., 1984; Dasberg and Dalton, 1985; Dalton and Van Genuchten, 1986). In this way the dependence of the electrical conductivity on water content can be taken into account. However, Topp et al. (1988) noted that the precedent works ignored the effect of multiple reflections of signal within the soil transmission line and the attenuation of the signal by the impedance-matching transformer was not accounted for. They found that the frequency dependence of the dielectric constant and attenuation needs to be measured independently in order to determine the relative contribution of the dielectric constant to the EC measurements made by TDR.

2.4.3. Calibration of laboratory and field electrical conductivity

Once the EC_a is measured whatever the instrument used (electrode probes, EMI, or TDR), we still need to calibrate these measurements with EC_e /EC_x, where x represents the ratio of water (1, 2, 2.5, 5, etc.). In this way different methods were proposed.

2.4.3.1. Electrode measured EC_a

As early as the beginning of the seventies, Rhoades and Ingvalson (1971) found that, based on measurements made using a horizontal configuration, a very close relationship exists between EC_e and EC_a. The correlation coefficients varied between 0.981 and 0.997 for different depths and different methods of calculation of EC_a. In addition, Halvorson and Rhoades (1974) found a high significant correlation between EC_e and EC_a with a correlation coefficient of 0.98 for a first date, when the soil profile was at or near field capacity, and a coefficient of 0.96 for a second date, when the soil profile was drier. For the same configuration, Halvorson and Rhoades (1976) found correlation coefficients higher than 0.94 for different depths (0 to 120 cm by 30 cm increments).

Halvorson et al. (1977) assessed whether the EC_e - EC_a relationship is different when using different methods of calibrations for soils of different parent material and with different textures. In this way, the EC_a was determined by three different methods: soil sampling method (Halvorson and Rhoades, 1974), four electrode cell technique method (Rhoades et al., 1977), and the EC-probe method (Rhoades and Van Schilfgaarde, 1976). The authors found that the slope of the linear regression line linking EC_e to EC_a decreases as clay content

increases and, to a lesser degree, as the soil water content increases. The parent material had a limited influence on the correlation between EC_e and EC_a. They concluded also that the cell and EC probe methods required less work and limited samples and were easier to use compared to the soil sampling technique. Both former methods were more accurate than the latter as they used the same soil volume for the determination of EC_e and EC_a.

To take into account the effect of soil properties like clay, sand, and water content on the EC_a values, Read and Cameron (1979) used the stepwise multiple regression to relate EC_e to both EC_a, determined using a horizontal configuration, and soil properties. They evaluated this technique in dryland soils in the Canadian Prairies. They found that, for uniform texture and moist soil, EC_a was highly correlated with EC_e (r=0.96 to 0.98). However, for more heterogeneous conditions (differences in salt and texture between depths and soil moisture less than field capacity), lower correlations were found (r=0.61). The inclusion of the clay, sand, and water contents in the EC_e - EC_a calibration equations, for the heterogeneous areas, improved the correlation coefficient to 0.84.

Rhoades (1981) went a step further in the consideration of the soil properties that affect EC_a, and thus the EC_e - EC_a calibration. He estimated the slope and intercept of the calibration equations from several soil properties. He found that the slope was better predicted from saturation percentage (or field capacity) and from clay+silt percentage while the intercept was best predicted from clay percentage. He concluded that as saturation percentage (or field capacity) is related to soil texture, the EC_e - EC_a calibration equations could be approximated from the knowledge of the soil texture.

All the above research works linked EC_e to EC_a. We noted in section 2.2.3.1 that EC_e is more reliable measure of soil salinity than electrical conductivity determined from other soil: water ratios. However EC_e is time and labour consuming, and is costly. In this regard, McKenzie et al. (1983) determined the correlation between four soil EC determinations: EC_e, EC_2, and EC_a measured using horizontal array and insertion probe. They found that the 1:2 soil: water extract method gave the best agreement with EC_e, with coefficients of determination ranging between 0.85 and 0.93 while these coefficients ranged between 0.64 and 0.78 from EC_e and EC_a obtained with insertion probe. The agreement was better for the shallower samples than for the deeper ones. For the horizontal configuration, the coefficients were lower ranging between 0.59 and 0.67.

2.4.3.2. EMI measured EC_a

The EMI technique for soil salinity measurement, with an EM31 instrument, was first used at the end of the seventies. De Jong et al. (1979) found that the coefficient of correlation between EC_e and EC_a was 0.83 when the instrument was carried at 80 cm above soil surface and 0.86 when it was on soil surface. For both cases, the readings were on vertical position. The authors determined also the coefficient of correlation between EC_a determined using EMI instrument (with different positions) and horizontal configuration (with different interelectrode distances). The correlation coefficients ranged between 0.63 and 0.90. They found that the correlation decreased, for a given depth, with interelectrode distance due to differences in measured soil volume.

In addition, Rhoades and Corwin (1981) correlated EC_a measured with EMI placed on the ground surface to EC_a measured with either a horizontal configuration or insertion probe. The relationships were strong between EC_a determined by EMI and insertion probe, except for the first 30 cm soil depth. The coefficients of determination ranged between 0.85 and 0.95 and increased with depth. For the horizontal configuration, the coefficients of determination were lower than for the insertion probe and decreased with the interelectrode spacing, ranging between 0.62 and 0.82. Using a horizontal configuration with an interelectrode distance of 1.52 m resulted in an almost 1:1 relationship between its EC_a and that determined using EMI under uniform soil conditions (mainly salt and water contents).

Wollenhaupt et al. (1986) used the same device, positions, and depths as McNeill (1980) except that, in determining the partial contributions, they ignored the depths lower than 1.20 m and 1.80 m for the horizontal and vertical position readings, respectively. In addition, they rescaled the partial contributions of the remaining depths to sum to one.

McKenzie et al. (1989) studied the influence of temperature, soil texture, and moisture on the conversion of EC_a to EC_e. They used the method of Wollenhaupt et al. (1986) to determine first the depth-weighted EC_e. The authors found that temperature corrections are essential for accurate conversions to EC_e. They noted also that the variability of EC_a to EC_e conversions were greater on coarse than on fine textured soils. Thus it is wise to determine separately calibration equations for soils with dissimilar parent materials (different textures). Also better calibration equations can be obtained if soil moisture is higher than 30%.

In a somewhat complex method, Rhoades et al. (1990) determined EC_e from the EC_a measurements determined with horizontal array, insertion probe, and EMI as well as from the estimates of some soil properties like soil water content, bulk density, and surface

conductance. The three instruments give depth-weighted EC_a measurements. The weights are function of the configuration of the electrodes or the EMI coils, the frequency of the electrical current, the distribution of EC_a within the depths of the soil profile, etc. (Lesch et al., 1992).

There are mainly two approaches to determine soil salinity (EC_e) by depth from EMI measurements (EC_a). A salinity prediction model was developed by Rhoades et al. (1989) to estimate EC_e using EMI or insertion probe EC_a measurements supposing that soil water and clay content are known. An alternative method, proposed by Lesch et al. (1992), establishes a direct relationship between EC_e and EC_a for homogeneous areas versus water content, bulk density, and other soil properties except soil salinity. From the horizontal and vertical EM38 readings (EC_a), the authors selected a limited number of locations for soil sampling to determine EC_e. The EC_e - EC_a data were used to determine the parameters of the prediction model and the computed model is used to rescale the intensive EM38 data into EC_e values. The authors recommend using geostatistical methods in cases where fields have substantial soil property heterogeneity other than salinity.

Diaz and Herrero (1992) related the EC_e and EC_5 to EC_a determined from EMI for both horizontal and vertical positions. They used simpler equations than McNeill (1980) by considering only two depths: the depth at which the samples were collected and a depth larger than it. They found that, in general, EC_e correlated better to EMI than to EC_5. Also EC_e and EC_5 correlated better with EMI horizontal readings than with vertical measurements. Moreover, the correlation between the observed and the predicted (from EMI horizontal and vertical positions) values were higher for EC_e than for $EC_{1:5}$.

Similarly, Cannon et al. (1994) found a good agreement between observed and determined EC_e from EMI readings calculated according to Wollenhaupt et al. (1986). The coefficient of correlation was 0.95.

All the precedent techniques used discrete (discontinuous) depths for the measurement of EC_a. In contrast, Acworth (1999) used the technique of EC image, which reflects the evolution of the EC_a continuously with depth.

The multiple coefficients approach of Rhoades and Corwin (1981) has some drawbacks. It requires a large number of EMI measurements made at different heights above the ground level in the horizontal and vertical modes, which means that this method is time consuming. Also the result, the prediction curve, is discontinuous with depth because the fitting is done for each depth separately. To circumvent these disadvantages, Corwin and Rhoades (1982, 1984) developed the established-coefficients approach, which requires only EMI measurements at the ground surface in horizontal and vertical modes instead of measurements

at different heights above the soil surface. However the method is empirical, the prediction is linear, and the coefficients are different for normal (salinity increases with depth) and inverted (salinity decreases with depth) profiles. A more robust calibration approach was developed by Lesch et al. (1992) that does not depend on the profile shape. More recently, Lesch et al. (1995b, c) developed a new approach called multiple linear regression calibration, which is able to produce different types of soil salinity estimates such as point estimates, conditional probability estimates, and field average estimates.

Triantafilis et al. (2000) proposed to fit a logistic model instead of linear models because most of the shapes of salinity profiles are sigmoidal. They compared the predictive capabilities of their model to those of the linear regression and established coefficients models. They found that the logistic model is less time consuming during the fitting step and provides better predictions.

Corwin and Lesch (2003) stated that the EC_e is inferred from EC_a by two approaches (Rhoades et al., 1999a, b): deterministic and stochastic approaches. In the deterministic approach, the models are theoretically or empirically determined, they are static meaning that all model parameters are considered known and no EC_e data need to be determined. Rhoades et al. (1989) gives an example of this deterministic approach. This approach is preferred when significant and localized variations in soil type exist in the field. It requires the knowledge of additional soil properties (water content, saturation percentage, bulk density, soil temperature, etc.). In the stochastic approach, the statistical modelling techniques (spatial regression or cokriging) are used to directly predict soil salinity (EC_e) from EC_a. Models are dynamic in the sense that model parameters are estimated using soil sample data collected during the soil survey. The calibration equations are developed by acquiring soil salinity data (or other soil properties like saturation percentage, texture, bulk density, etc.) from a small percentage of the sensed locations and estimating a stochastic prediction model for the established model. Stochastic and deterministic calibration approaches are described in more detail in Lesch et al. (1995b, c) and Rhoades et al. (1999a) and incorporated into the ESAP computer software application (Lesch et al., 1995a, 2000).

2.4.3.3. TDR measured EC_a

Van Loon et al. (1990) found a slope of about one for regression equations between EC_e and EC_a determined using TDR instrument for water solution and sandy and loamy saturated soils. The intercept was significant for the water solution but not for the two soils. The

correlation coefficients were very high and amounted to 0.993 for both the water solution and the sandy soil and to 0.968 for the loamy soil. The same strength of the EC_e - EC_a relationships was noted by Topp et al. (1988). They found coefficients of determination ranging between 0.972 for silty loam soil and 0.999 for clayey loam soil, using a coaxial configuration. Using a thin sample approach, the coefficients of determination exceeded 0.99 for both types of soils.

2.4.4. Mobile apparent electrical conductivity measurement systems

The EC_a sensors (electrode probes or EMI) can be mounted on a small tractor or on all terrain vehicle (ATV) with a global positioning system (GPS) and a datalogger to obtain a mobile system able to measure EC_a and register the coordinates of the locations where EC_a were measured. In this sense, Rhoades (1992, 1993) incorporated a GPS with a four-electrode probe on a small tractor to measure EC_a. Also, Carter et al. (1993) incorporated EM38 and horizontal array sensors to GPS while Jaynes et al. (1993), Cannon et al. (1994), and Sudduth et al. (2001) used only the EM38 sensor. In contrast, Rhoades et al. (1997) presented a system based on electrode probes providing EC_a measurements to a soil depth of 1.30 m by 0.10 m increments while Lund et al. (1999) used a mobile system equipped with six coulters (electrodes) which is able to provide shallow (0 to 30cm) and deep (0 to 100 cm soil depth) measurements of EC_a. Recently, Triantafilis et al. (2002) conceived a mobile system, which integrates EM31 and EM38 sensors along with a computer datalogger and two GPS units. The system was used to provide information for managing soil salinization in a cotton field of about 26 ha in the Lower Namoi Valley of New South Wales, Australia.

2.4.5. Other uses of the apparent electrical conductivity

Apparent electrical conductivity, mainly that determined using EMI devices, was used to study the spatial distribution of different soil properties like clay content (Williams and Hoey, 1987; Triantafilis et al., 2001), soil moisture (Kachanoski et al., 1988), soil salinity (Lesch et al., 1992; Vaughan et al., 1995), nutrient status of soil (Sudduth et al., 1995), depth to a clay layer (Brus et al., 1992; Doolittle et al., 1994; Sudduth et al., 2001), determination of organic matter fraction (Jaynes et al., 1995), forest soil quality (McBride et al., 1990), and lateral changes in texture (Zalasiewicz et al., 1985).

2.4.6. Use of GIS and remote sensing to evaluate salinization risk

Soil salinization is a major problem worldwide and researchers are interested in evaluating its risk in order to adapt management techniques to the situation to reduce the occurrence of salinization. The integration of the new information acquisition and manipulation tools like GPS, GIS, and remote sensing (RS) can enormously help in characterizing and evaluating the risk. Bui et al. (1996) used soil survey information and water balance simulations in a GIS framework to identify areas at risk of salinization. They determined three permeability and drainage classes (low, medium, and high) and recharge and discharge areas from the soil survey information (parent material, soil thickness, depth to bedrock, etc.). The study of the association between electrical conductivity point measurements (or %*TSS*) and the recharge and discharge areas allowed to assess the salinity hazard.

In addition to GIS, RS can be very useful in the delineation of saline areas. In this way, Peng (1998) used Landsat Thematic Mapper images, which were transformed and classified. The classification was based on the generalized Bayes analysis of multisource data (Swain et al., 1985). Using JERS-1 SAR images and fuzzy k-means classification algorithm (Burrough, 1989), Metternicht (1998) mapped salt-affected soils. The class determination of saline, sodic, and saline-sodic areas was based on the EC_e threshold values of 4 dS/m (Richards, 1954). In a recent work, Meternicht (2003) refined his precedent approach, following the World Reference Base for Soil Resources (Spaargaren, 1994) in which anion ratios, instead of EC_e, are used. He used a supervised maximum likelihood classification technique where the membership grades of the saline fuzzy classes are incorporated as prior probabilities to classify Landsat Thematic Mapper data sets.

2.5. Conclusions

In this chapter, we described how salt-affected soils are characterized. We focused specifically on soil salinity which is our main concern.

We conclude, from this chapter, that soil salinity can be accurately measured in the laboratory by determining either electrical conductivity from an extract from a saturated paste of soil (EC_e) or from a soil:water ratio, in our case 1:2.5 ($EC_{2.5}$). We choose the latter because it is easily measured and less time-consuming than the former and also because we are interested in the relative change of soil salinity over time instead in its absolute values. In addition, the

laboratory measurements were complemented by the survey of the apparent electrical conductivity (EC_a) using a 4-electrode probe device. This soil propriety, although not a direct measure of soil salinity, is strongly correlated to it and needs just a simple calibration based on, for example, linear regression models to convert it into soil salinity ($EC_{2.5}$).

CHAPTER 3

STUDY SITE AND DATA DESCRIPTION

3.1. Study site description

Salt-affected soils in Hungary cover more than 10 % of the territory of the country (exceed one million hectares) and more than 95 % of these soils are located in the Great Hungarian Plain (Szabolcs, 1989). Hortobagy National Park, where lies the study area (Fig. 3.1), forms a subregion of this plain, called also Alföld. The area of salt-affected soils in the HNP exceeds 60000 ha (Tóth and Rajkai, 1994). Therefore the main environmental features of the Great Hungarian Plain (GHP) that are pertinent to the occurrence and the formation of salt-affected soils will be described.
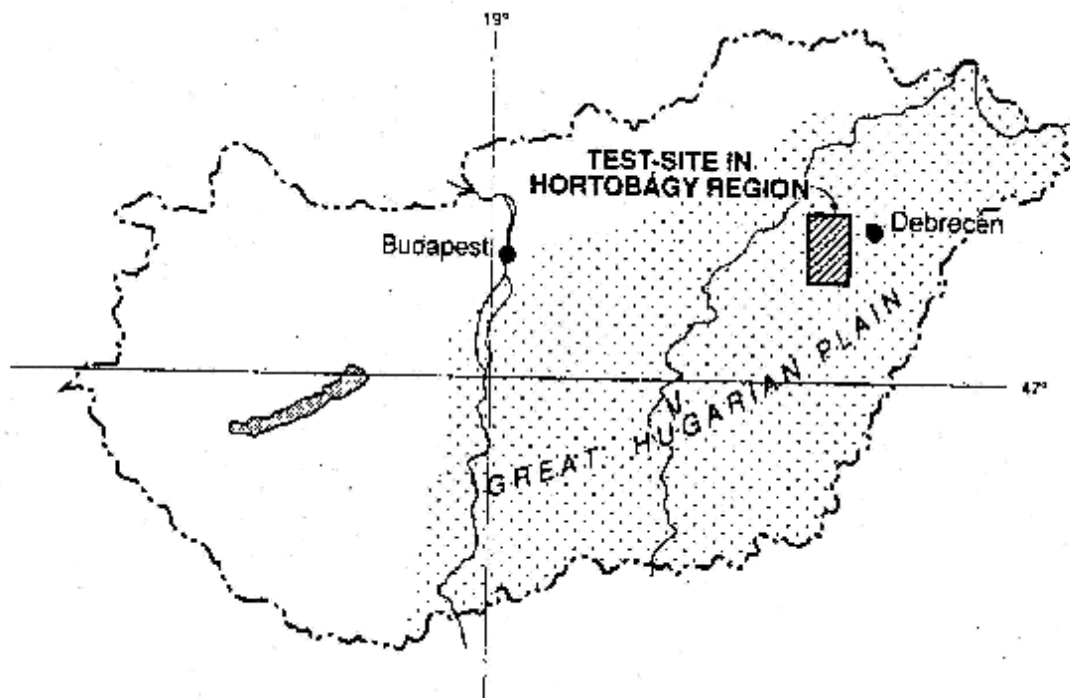
Fig. 3.1. Map of Hungary with the Great Hungarian Plain and the study site.

3.1.1. Geology and geomorphology

The Alpine orogenesis is behind the origin of the GHP. This orogenesis contributed, around the end of the Miocene, to the formation of the Carpathian basin. This basin, also known as Sarmathian or Pannonian sea and which communicated initially with the Mediterranean Sea, was reduced, during the Pliocene, to a lake without outlet. The sea and the lakes derived from it were filled up, during the end of the Tertiary, by river sediments (Abraham and Bocskai, 1971; Tóth et al., 1991; Bui et al., 1998).

The characteristic soil-forming loess material of the GHP is the product of the Quaternary during which the glaciers did not extend in the plain but this latter was strongly influenced by the glacial era. Water is another factor that played a decisive role in the formation of the parent material of the GHP. The loess formed under its influence is called lowland or influence loess, to be distinguished from the loess deposited by wind on dry surfaces (Abraham and Bocskai, 1971; Tóth et al., 1991; Bui et al., 1998).

The height above sea level of the GHP ranges between 85 and 120 m. The plain is formed of many smaller basins, most of them occupied by sodic (formerly alkali) soils. The height of these basins is 95 to 97 m in the Danube valley, 90 m in Hortobagy, and 80 to 82 m along the lower regions of the Tisza river. Tóth and Rajkai (1994), and Tóth and Jozefaciuk (2002) noted that a small difference in elevation (some decimeters) results in large difference of salt accumulation. The Salic Solonetz (World Reference Base soil nomenclature) is the most sodic and saline part of the toposequence.

3.1.2. Meteorology

The Carpathian basin is characterized by a temperate climate, but the GHP is the hottest and the driest region of this basin as it is the most continental part of Hungary. The Hortobagy region is characterized by a cold winter, 170 to 200 days per year of frost, a temperature range of 50-55 °C, and an average sunshine duration of 1900 to 2000 hours per year (Abraham and Bocskai, 1971; Tóth et al., 1991). Additional main climate features are reported in table 3.1 (Tóth et al., 1991).

Table 3.1. Some climatologic parameters of the GHP Central Region. P: precipitation (mm), PE: potential evaporation (mm), AE: Actual evaporation (mm), T: Air temperature (°C), DI: drought index.

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Year
P	30	30	28	41	51	71	53	50	34	33	46	46	527
PE	12	19	40	78	112	136	156	144	106	58	25	14	900
DI	0.4	0.63	1.43	1.90	2.20	1.92	2.94	2.88	3.12	1.64	0.54	0.30	1.71
AE	11	15	27	63	102	91	76	58	35	21	16	12	527
T	-1.8	0.5	5.2	10.9	16	19.7	21.3	20.5	16.4	10.7	5.3	0.6	10

The yearly values are sums for precipitation and potential and actual evaporation whereas they are means for drought index and air temperature.

The rainiest period is from May till August, with a peak of 71 mm in June while the period from January to March is the least rainy. The hottest months are June to August and the coldest are December to February, with an average monthly temperature below zero in January. More interesting is the drought index (DI), the ratio of potential evaporation to precipitation. This index exceeds unity for most of the months meaning that for these months (from March to October) the amount of rainfall is less than the potential evaporation. This will have an important impact on the salinisation of the plain.

3.1.3. Hydrology and Groundwater

Sediments deposited by rivers (mainly Danube and Tisza) and wind filled up the basin of the GHP. There were many floods that created marshlands. So, surface water was an important factor of soil formation (Tóth et al., 1991).

Hortobagy is a recharge area of a saline groundwater originating from northern mountains. This groundwater is the main source of salt accumulation in the area. Waterlogging induces a rise in the groundwater level during the wet season. This results in a flow towards the Salic Solonetz elevated zone (Tóth and Jozefaciuk, 2002). However the groundwater flows downward during the dry season.

Groundwater, and more precisely the water table depth, played a decisive role in the formation of salt-affected soils in the GHP. As there are regional differences in the composition and the concentration of the groundwater, this resulted in different types of salt-affected soils. This will be discussed later.

The groundwater table depth varies between 0.5 and 4 m and it fluctuates, on average, between 0.5 and 2 m. Waterlogging, which is the consequence of the shallow water table, fine texture, and uneven precipitation distribution, is frequent in lower parts of the plain. This surface water pounding appears also in low-lying and low permeable areas at the end of the winter as the snow melts and/or during high precipitation periods.

The groundwater has high salt content and a high sodium adsorption ratio. These two features contribute to the salinisation and alkalinisation of the soils of the GHP.

3.1.4. Soils

There are mainly three soil types. In the hilly marginal regions of the plain, there are soils of the Atlantic region (brown forest soils). In the inner plateaus of the plain, soils of the steppe region (chernozems) are found. The third type includes the salt-affected soils, meadow soils, and the alluvial soils.

Three main categories of salt-affected soils can be distinguished (Szabolcs, 1989): potentially salt-affected soils, saline soils, and sodic soils. The latter, which is found in the study area, is characterized by high clay content and unfavorable physical properties, low content of salt, high *ESP*, and high pH in the B horizon. There is a clear difference between the A and B horizons. The A or E (eluvial) horizon is most of the time without salt, with a pH of about 7 or slightly less. It is homogeneous, with dense roots, a texture ranging between sandy and silty loam, and it has a laminar structure. In contrast, the B (illuvial) horizon is characterized by alkalinity and a finer texture (clay-loam) and it has a columnar structure. The C horizon of the sodic (Solonetz) soils is not strongly affected by the different processes of salt accumulation.

3.1.5. Ecology and botanics

The GHP is a flat area, with almost no trees. It has an area of 45000 km^2 and is covered by marshes, rangelands, fishponds, abandoned watercourses, meadows, reeds, and some woody patches (Tóth et al., 1991).

Soil salinity/sodicity and its correlation with the vegetation have been studied in the area by many authors, among them Bodrogközy (1965), Tóth et al (1991, 2002b), Van Meirvenne et al. (1995).

As it was noted before, the basin was filled up with sediments from the Danube and Tisza rivers during the Pleistocene. The thickness of these sediments varies between 100 and 200 m. Upon this layer, 4 to 5 m of loess or meadow clay has been deposited.

The GHP, in the past, was inundated, 2 to 3 times per year, by the Tisza river. It had open forest and marshy vegetation separated by grassland.

The large-scale formation of the szik soil (solonetz) began in the Holocene. It was due to the salts transported by the groundwater to soil surface layers. The area was deforested and dams and canals have been built. These anthropogenic factors induced the decreasing of waterlogging, and the desiccation of the area and resulted in considerable extension of salt-

affected soils. Waterlogging may still occur as there is a fast snowmelt, heavy showers can occur, and the hydraulic conductivity of the B horizon is very low.

Chernozems and meadow chernozems (Hungarian soil nomenclature) are found in high sites of the GHP while solonetz soils occur in lower lying locations.

The main catenas of salt-affected soils are reported in table 3.2 and Fig. 3.2 (Tóth et al., 1991).

Table 3.2. Main types of surface elements, erosion form, plant associations, and soil types.

Surface element	Erosion form	Plant association	Soil type
Loess plateau	No	*Cynodonti-Poetum augustifoliae*	Meadow chernozem
Grassy lower place	Slight	*Achilleo-Festucetum pseudovinae*	(Steppising) meadow solonetz
Wormwoody rangeland	Torn grasscover	*Artemisio-Festucetum pseudovinae*	Meadow solonetz
Bare spot on lower part	No A horizon	*Camphoresmetum annuae*	Decapitated meadow solonetz
Bare spot on lower temporary water course	1-2 cm of A horizon	*Puccinellietum limosae*	Solonchaky decapited solonetz
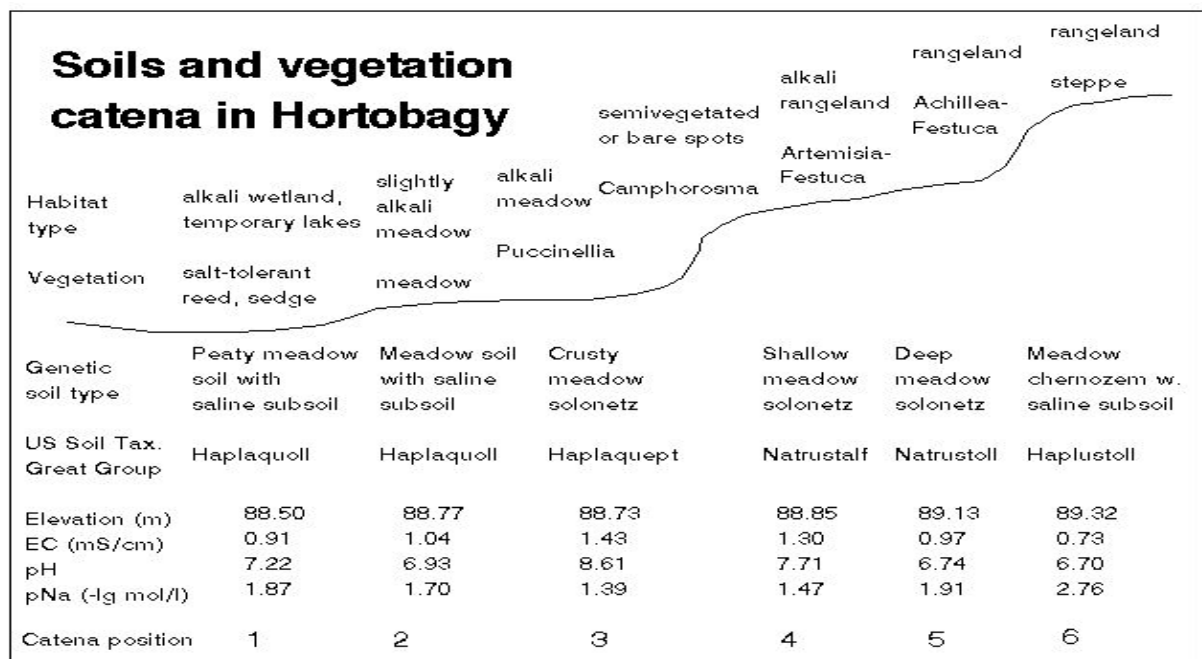
Fig. 3.2. Soil and vegetation catena in Hortobagy National Park (source: Toth and Kertesz, 1996).

The limits of the former river beds are continuously eroded by water derived from heavy spring rains and/or snow melting. Trampling animals and vehicles induce erosion of the A horizon in the higher locations. This horizon is not saline, however when it is partially or fully eroded, the plant roots live in thinner surface horizon, or even on compact, saline B

horizon of columnar or prismatic structure. These are the causes of the heterogeneous spatial vegetal cover. There are mainly 4 groups of plant associations in the sodic grasslands of Hortobagy (Bodrogközy, 1965; Tóth et al., 1991):

- *Beckmannion* : salt-affected meadow with *Agrostis alba*
- *Puccinellion limosae* : association of bare salt-affected spots
- *Festucion pseudovinae* : salt-affected rangeland
- *Festucion sulcatae* : slightly salt-affected rangeland

Tóth and Jozefaciuk (2002) studied 3 soils in a sodic toposequence (Mollic Solonetz, Salic Solonetz and Haplic Solonetz according to the World Reference Base soil nomenclature) from the Hortobagy National Park (HNP) and found that there exists a close relation between the position on the toposequence and the vegetation. Grassland occurs at high locations (Mollic Solonetz), meadow in low position (Haplic Solonetz), and short grass in intermediate position (Salic Solonetz).

3.1.6. Salt-affected soils of Hungary

Szabolcs (1989) defined salt-affected soils as formations under the dominating influence of different electrolytes in solid or liquid phases that alter the physical, chemical, and biological properties, and eventually the fertility of the soil. He distinguished 5 groups depending on the kind of electrolytes that cause salinity and/or alkalinity:

- Saline soils: the main electrolytes are sodium chloride and sulphate
- Alkali soils: sodium ions capable of alkaline hydrolysis
- Magnesium soil: magnesium ions
- Gypsiferous soils: calcium ions (mainly calcium sulphate)
- Acid sulphate soils: ferric and aluminium ions (mainly sulphates)

From the 5 above groups, Hungary is mostly concerned with the second group, as shown in table 3.3 (Szabolcs, 1989).

A further subdivision, based on the classification accepted by the previous subcomission on salt-affected soils of the International Union of Soil Science, is to distinguish alkali soils which have a structural B horizon from those missing this horizon. The former is again

subdivided in the two types depending on the nature of the parent material: it can be calcareous or not.

Table 3.3. Extent of salt-affected soils in Hungary. Calc: calcareous; With/Without B: with/without structural B horizon; SAS: salt-affected soils. Areas in 1000 ha.

Type	Saline soils	Alkali soils			Potential SAS	Total
		With B		Without B		
		Non-calc	Calc			
Area	1.6	294	31	58.6	885.5	1271
Percentage	0.1	23.1	2.5	4.7	69.6	100

The geographical distribution of salt-affected soils in Hungary is given in Fig. 3.3.
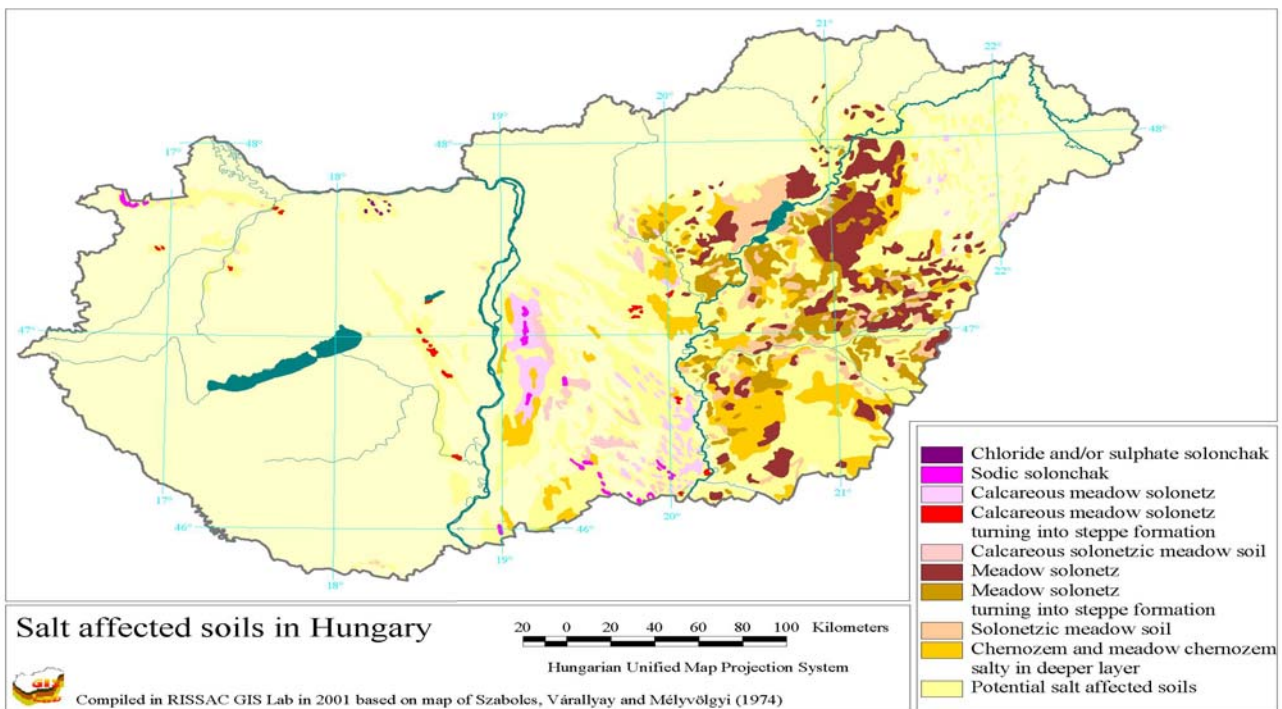
Fig. 3.3. Salt-affected soils in Hungary (based on Szabolcs, 1974 and digitally compiled by GIS lab of RISSAC)

In alkali soils without a structural B horizon, the harmfulness of the sodium salts capable of alkaline hydrolysis because of the inhibition of plant growth while in alkali soils with a structural B horizon, alkalinity causes physical soil properties which interfere with the water supply (Szabolcs, 1989). Alkali soils without a structural B horizon are characterized by high concentration of water-soluble salts capable of alkaline hydrolysis found even in the top

layers. They are similar to saline soils, for example both soils have not readily distinguishable horizons separated by abrupt boundaries, they have salic horizon and their electrical conductivity exceeds 4 dS/m. However they differ in their pH: alkali soils have bigger pH (more than 8.5) than the saline soils. The alkalinity of alkali soils results in compactness, lack of structure, low humus and plant nutrients content, and impermeability of the top layers.

Alkali soils with structural B horizon have very low concentration of water-soluble salts capable of alkaline hydrolysis, particularly in the eluvial A horizon, which is less compact and less developed than the B horizon. The illuvial or accumulation horizon B is of prismatic or columnar structure, is well distinguished from the A horizon (this one can be fully eroded), has an *ESP* bigger than 15, and is rich in humus with presence of clay particles (Szabolcs, 1971). The high level of *ESP* implies poor physical and water regime properties, and compact structure of the B horizon.

The thickness of the A horizon is an important parameter as it determines the amount of retained water and made available to plants. Based on this factor, alkali soils with a structural B horizon can be divided in 3 types (Szabolcs, 1971):

- Shallow solonetz soil: the depth of the B horizon is less than 7 cm
- Medium solonetz soil: the depth ranges between 7 and 16 cm
- Deep solonetz soil: the depth exceeds 16 cm

The sodic solonchak-solonetz soils occur in the Danube-Tisza interfluve, in lower zones where the water table is high (about 1 m). Calcium carbonate is present in the whole profile. The parent material is loessial or calcareous sand, loam, or clay. *Lepidium cartilagineum* and *draba, Camphorosma ovata*, and *Puccinellia limosa* are the frequent plant species in these soils (Abraham and Bocskai, 1971).

The meadow solonetz soils are the most frequent type of the salt-affected soils in Hungary. They are found east from the Tisza river, with however in small patches in the west of the Danube. If the A horizon is eroded, they are called crusty meadow solonetz. The parent material is loess or calcareous clay or loam. The water table is at a depth of 1.50 to 3.50 m. The characteristic plant species are *Artemisia salina, Santonicum*, and *Festuca pseudovinae* (Abraham and Bocskai, 1971).

The water table of meadow solonetz turning into steppe formation is deeper than the two precedent soil types (3 to 4 m). It has as parent material loess or carbonate-bearing clay or loam. *Achillea setacea*, and *Festuca pseudovinea* are examples of frequent plant species.

3.2. Data description

3.2.1. Soil salinity

The study area covers about 25 ha in the HNP, in the east of Hungary with the central coordinates 47°30" N and 21°30" E (Fig. 3.4).
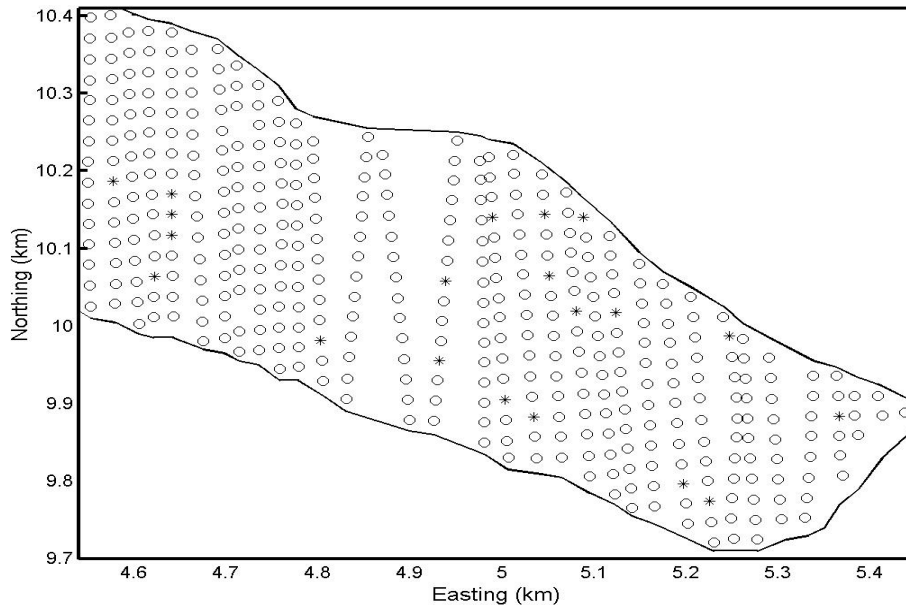
Figure 3.4. Extent of the study area. Circles represent locations where EC_a was measured and asterisks represent locations where both EC_a and $EC_{2.5}$ were measured.

Initially the sampling design was chosen in a way to check if there is a temporal change in soil salinity between two sampling campaigns. In this sense two data sets have been collected. The first data set, termed 'data set to be calibrated', involves the measurement of the bulk soil or apparent electrical conductivity (EC_a) in the field at 413 locations in a pseudo-regular grid of 25 by 25 m; with mainly in the middle of the study area some larger spacings due to the presence of a well where much disturbance occurred.

The EC_a has been measured using an electrical conductivity probe (Fig. 3.5), equipped with 4 electrodes (Rhoades and Van Schilfgaarde, 1976). The spacing between the inner sensing electrodes is 90 cm and between the sensing/receiver (outer pair) and sensing (inner pair) electrodes there is 10 cm distance both sides. The electrodes are inserted in the soil at two different depths: 8 and 13 cm. The corresponding EC_a measures (dS/m) are characteristic for

the 0-20 and 0-40 cm soil depth respectively. At each calibration point, there were always 3 parallel measures of EC_a but only one from normal points. It is the first reported extensive use of such probes in native grasslands, which are characterized by heterogeneous vegetation, elevation and which can undergo large fluctuations in soil moisture.

Fig. 3.5. The 4 electrode probe used during work in Hortobagy.

A spatial site selection algorithm based on response surface design (Lesch et al., 1995c) was followed to identify a minimal number of calibration sites. The selected sites are based on the spatial configuration of locations for which EC_a was measured and on the values of these measurements. These are spatially representative of the study area and allow accurate estimation of the calibration parameters. These selected sites constitute the second data set, called 'calibration data set'. Soil samples from all sites have been collected, put in plastic bags, and analysed in laboratory. Samples were air dried and crushed to pass through a 2 mm sieve. The 1:2.5 soil:water suspensions were prepared, and shacked. After 16 hours, pH and *EC* (electrical conductivity) were measured. $EC_{2.5}$ is reported after conversion to a standard 25 °C temperature. In addition to the EC_a field values, measurements of electrical conductivity from the 1:2.5 pH suspension ($EC_{2.5}$, in dS/m), gravimetric moisture (%), and pH were available. Gravimetric moisture was determined by drying soils collected in air-tight containers at 105 °C until mass stabilized.

The $EC_{2.5}$ is a simple proxy of the water-saturated soil-paste extract (EC_e), which is the conventional measure of soil salinity (Soil and Plant Analysis Council, 1992).

At the calibration points, soil samples were collected down to 40 cm, by 10 cm increments. Therefore, bulked samples were taken from two augerings located between the pairs of electrodes of the conductivity probe at a distance of 50 cm.

The sampling of the 'data set to be calibrated' and the 'calibration data set', has been repeated at 21 time instants: November 1994; March, June, September, and December 1995; March, June, September, and October 1996; March, June, September, and December 1997; September 1998; April, July, and September 1999; April, and December 2000; and March, and June 2001. However, at some dates, September and October 1996, we did not have EC_a data, because of instrumental malfunction and vandalism, and consequently these dates were not used. Since the area is an open pasture it is very easily accessed. During the observation period the identifying sticks were removed by vandals twice. After the first case new sticks have been installed but there was some spatial shift. The second resulted in the ending of the observation period. Therefore, in total the useful sampling was repeated 19 times. The average temporal lag is 3 months but ranged from 2 to 9 months.

In general 20 calibration sites have been selected and it was tried to keep the same sites for the future sampling campaigns, but at a few time instants some of calibration sites have not been measured due to standing water. So the number of calibration sites varied between 13 and 20 for the different sampling campaigns.

To have an idea about how soil salinity varies in space and time, before any formal (geo) statistical analysis and interpolation, EC_a measurements at the 413 space locations and at the 19 time instants are shown in Fig. 3.6. From this figure some points need to be highlighted. First of all, the study site was not equally surveyed during all the time instants; for example it was markedly undersampled during March and June 1997, and to a lesser extent in April 2000. Regarding the temporal variability, EC_a measurements were low for some time instants like November 1994, June 1996, or September 1997 whereas these measurements were very high compared to the general trend for time instants like September 1998 and April and July 1999. Concerning the spatial variability of soil salinity, it can be noted that some space locations are characterized by high/low values for most or all of the time instants. This is an indication of temporal stability or persistence which will be demonstrated in chapters 5 and 6.
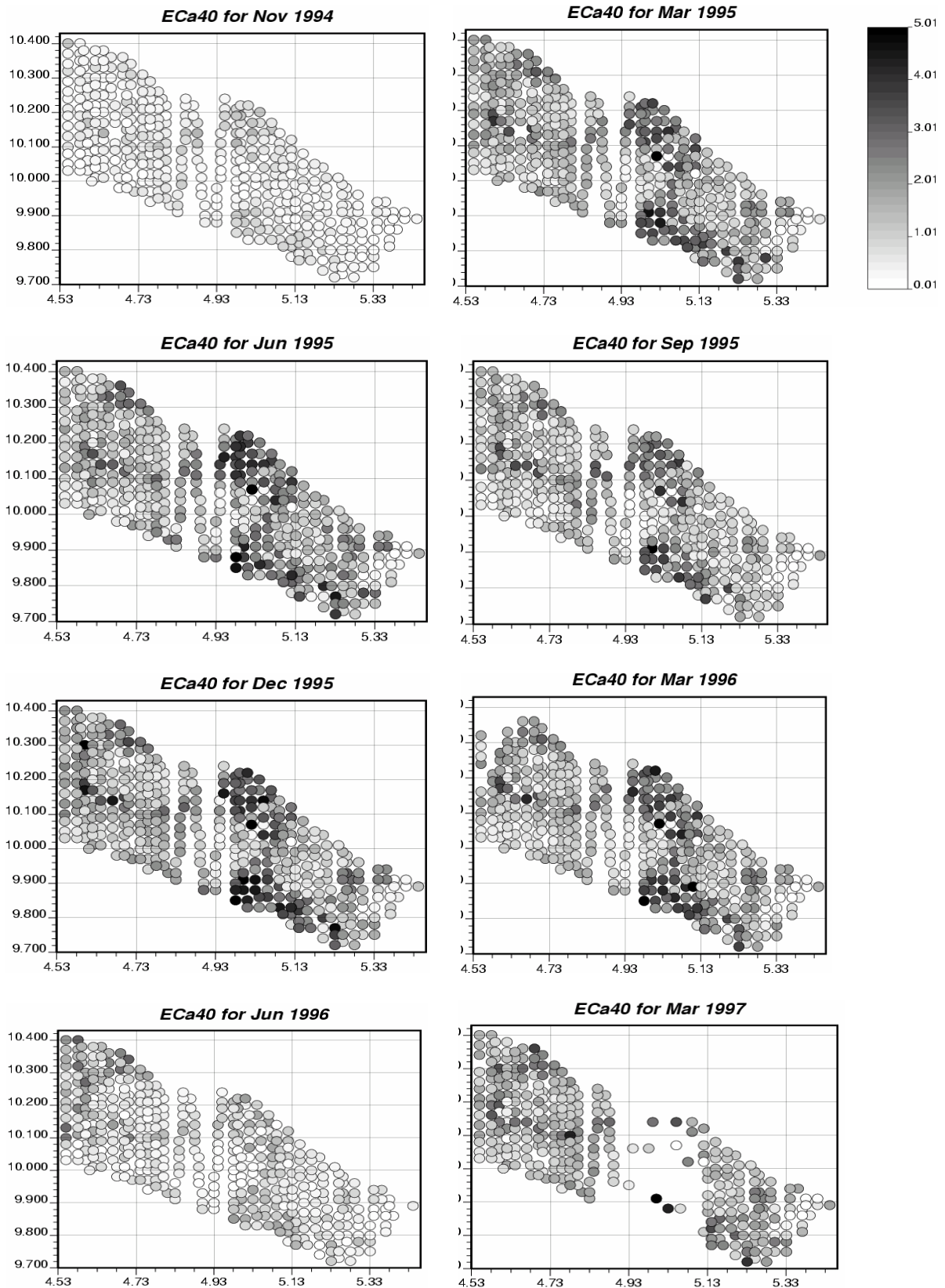
Fig. 3.6. Location of the samples where EC_a was surveyed. The level of tone reflects the measured values (see legend above).

Fig. 3.6. (continued).

ECa40 for Dec 2000

ECa40 for Mar 2001

ECa40 for Jun 2001
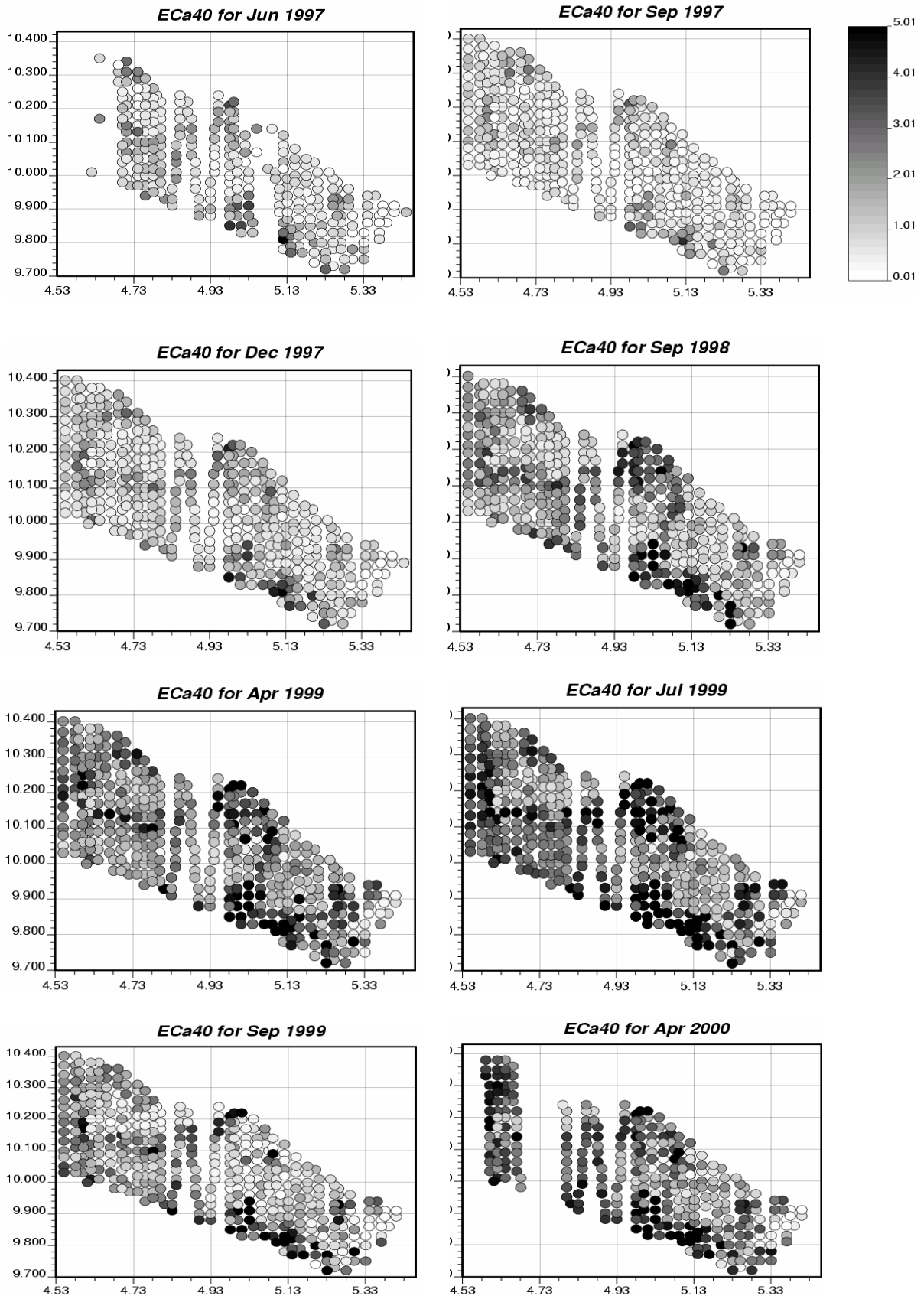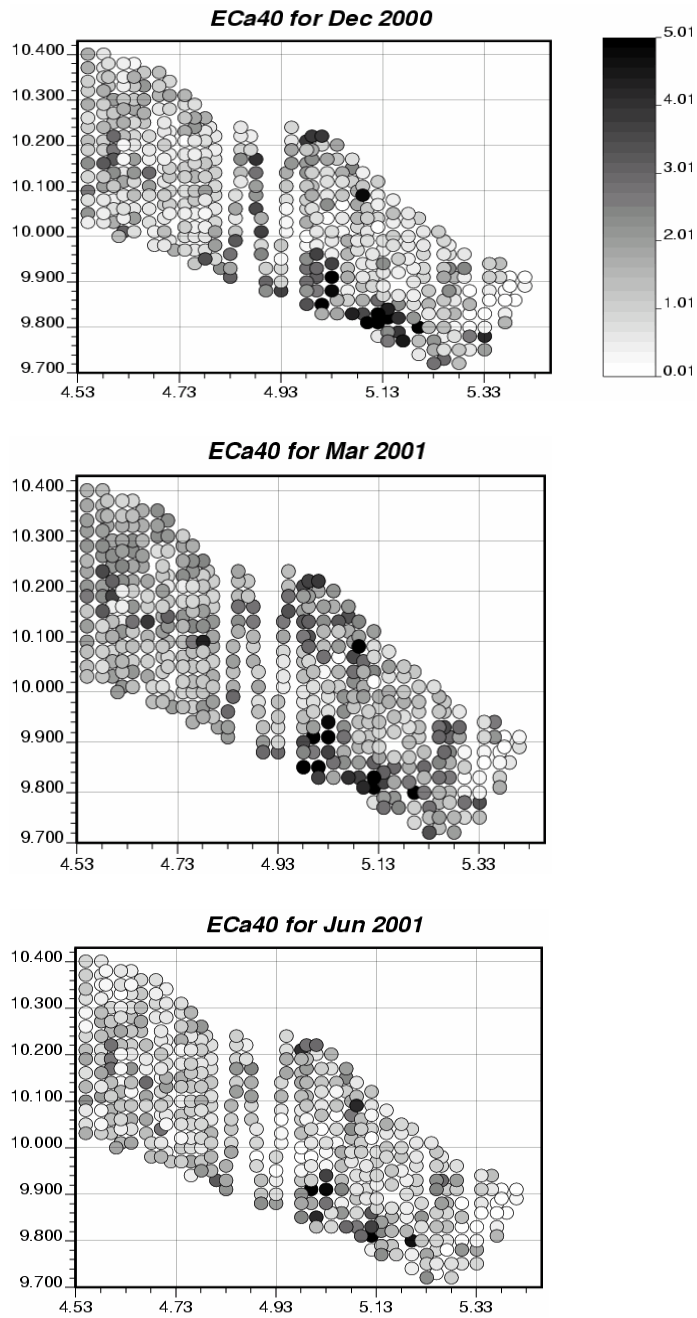
Fig. 3.6. (continued and end).

In addition to the basic soil salinity data sets, other related data were available including profile description, vegetal cover, meteorological data, data on groundwater positions, etc.

3.2.2. Vegetation

The vegetal cover was identified for the 413 locations by botanical categorization based on plant composition. It was reported as detailed description of 14 categories or summarized inventory of 6 plant associations.

The categories are:

 1: *Phragmitetum*

 2: *Typha sp.*

 3: *Schoenoplectus tabernaemonatani*

 4: *Bolboschoenetum maritimi*

 5: *Agrosti-Glycerietum p.*

 6: *Agrosti-Beckmannietum e.*

 7: *Eleochari-Alopecuretum geniculati*

 8: *Agrosti-Alopecuretum pratensis*

 9: *Trifolio-Poetum puccinelliet.*

10: *Gypsophilo-Artemisietum*

11: *Trifolio-Poetum Artemisio-Festucetum*

12: *Achilleo-Festucetum*

13: *Cynodonti-Poetum angustifoliae*

14: *Salvio-Festucetum rup.* or *Astragalo-Poetum angustifoli*

The numbering order follows ascending elevation.

The plant associations are the regrouping of the categories:

I: categories from 1 to 7

II: category 8

III: categories 9 and 10

IV: category 11

V: category 12

VI: categories 13 and 14.

The different plant categories along with the space coordinates were used as explanatory variables in the regression models relating $EC_{2.5}$ to EC_a.

3.2.3. Groundwater

The depth of the groundwater under the surface (cm) and its electrical conductivity (*EC*, in mS/cm) were measured in 4 wells, which were at the elevations of 90 m, 89.3 m, 88.8 m, and 88.7 m above the sea level. The groundwater was monitored at different time instants during the sampling campaigns; these time instants are reported in table 3.4.

Table 3.4. Frequency of the monitoring of the groundwater depth and its electrical conductivity (EC).

Year	1995	1996	1997	Total
Depth	33	46	46	125
EC	8	12	4	24

3.2.4. Profile description

At the stick numbers 6, 249, and 419, the description of the soil profile was made. It includes 4 genetic horizons and their depth (cm); the cations exchange capacity (*CEC*, in meq/100g); the exchangeable and soluble calcium, magnesium, sodium, and potassium (meq/100g); exchangeable sodium percentage (*ESP*, in %); $CaCO_3$ (%); saturation (%); pH; EC_e; anions (CO_2, HCO_3, Cl, SO_4, and total; in meq/l); cations (Mg, Ca, Na, K, and total; in meq/l); sodium adsorption ratio (*SAR*); particle size fractions (%); saturated hydraulic conductivity (cm/day); pF0; and bulk density (g/cm^3).

3.2.5. Meteorology

Some climatic data were available like daily measured precipitation (mm), daily calculated evapotranspiration (mm), from a nearby meteorological station. Also monthly precipitation and temperature for November 1994 and the years 1995 to 1997 were available. Additionally, for the Hortobagy National Park, meteorological data were available on a decade basis between 1988 and 2000. The following parameters were measured: air temperature (°C), precipitation (mm), evapotranspiration (mm), soil moisture deficit, and mean water loss from the soil (0-100 cm) through calculated evaporation (mm).

CHAPTER 4

SPACE-TIME DATA ANALYSIS: METHODOLOGY

4.1. Introduction

This chapter will review the literature on statistical methods for analysis of space-time data and the focus is made on the approaches applied to the available data sets. It is divided into four main parts: classical statistical methods (section 4.2), structural analysis (section 4.3), kriging (section 4.4), and Bayesian maximum entropy (section 4.5).

4.2. Classical statistical methods

Classical statistical methods are basically aspatial and atemporal methods where the spatial and temporal coordinates are ignored. Four assumptions must be checked before any use of these methods: linearity of the relationship between dependent and explanatory variables, normality of the distribution, equality of variances (homoscedasticity), and independence of the residuals. We present in the next section (4.2.1.) the tools used to check these assumptions, and then the different approaches: coefficient of variation (4.2.2), variance component analysis (4.2.3), temporal stability (4.2.4), and dynamic spatial variation (4.2.5).

4.2.1. Assumptions

4.2.1.1. Linearity of relationship

Linearity means that the amount of change, or rate of change, between values on two variables is constant for the entire range of values for the variables. There are both graphical and statistical methods for evaluating linearity. Graphical methods include the examination of scatterplots, often overlaid with a trend line. This is the first and informal way to check the linearity. Plotting the standardized residuals from the regression against the standardized predicted values is another useful graphical tool. The absence of nonlinearity should show a random pattern in this plot. Statistical methods include diagnostic hypothesis tests for linearity, a rule of thumb that says a relationship is linear if the difference between the linear correlation coefficient (r) and the nonlinear correlation coefficient is small and examining patterns of correlation coefficients. The correlation between, on one hand, the dependent variable and its transformations (root square, inverse, or logarithm) and, on the other hand, the original and transformed variations of the explanatory variable provides us with a pattern that

is easier to interpret. A brief check of these correlations and their significance can help in identifying which transformation is required in order to verify the linearity and to increase the relationship between the dependent and the explanatory variables. Additionally, when the standard deviation of the residuals exceeds the standard deviation of the dependent variable, one may suspect a possible nonlinearity. Also, another way to check for nonlinearity is to add to a model nonlinear terms like squares or cubes of the explanatory variable and see if the coefficient of correlation (determination) improves significantly compared to the restricted linear model.

4.2.1.2. Normality of distribution

Again, this assumption can be checked either graphically or statistically. Among the graphic tools, there is the histogram of frequencies, for a normal distribution it looks like a well bell-shaped curve, or equivalently the cumulative distribution function which has a sigmoidal shape for normal distribution, the normal probability plot (P-P plot), the Q-Q plot, the steam-and-leaf plot, etc.

A first statistical test for normality is to check if kurtosis and skewness coefficients range between -1 and +1, or equivalently if these coefficients, when divided by their respective standard deviations, belong to the intervals [-3, 3] and [-2, 2] respectively. For a more formal test, one can use either the Kolmogorov-Smirnov test or Lillefors test (which is an adjustment of the former when the mean and variance are unknown) for large samples or the Shapiro-Wilk test for small samples, say ≤ 50.

4.2.1.3. Homoscedasticity

The equality of variances or homoscedasticity can be, as for the two above assumptions, checked visually using graphs or formally using statistical tests. The most useful graphs in this case are the plot of observed or standardized predicted values as function of the standardized residuals or the residuals as function of the observed values. A random pattern in these graphs is an indicator of homoscedasticity. Again for a formal test, one can use, for example, the traditional Bartlett test which requires normality of the distribution or the Levene test which is robust to the departure from normality.

4.2.1.4. Spatial/temporal independence

When the plot of residuals as function of the identification number, reflecting the order (spatial or temporal) of data collection, shows no particular pattern, one may consider that the observations are independent. The Durbin-Watson statistic can be used to formally check independence, mainly temporal while the Moran test is recommended to check spatial independence. This test is presented in section 4.2.5 when dealing with the dynamic spatial variation.

Now we have identified the main assumptions required to be checked before any use of the classical statistical methods, we present some of these methods and focus on the ones that will be used in our application.

4.2.2. Coefficient of variation

It is assumed that the observed value for a soil attribute (x) at any location (s) at a given time instant (t) is:

$$x(\boldsymbol{s},t) = \mu + \varepsilon(\boldsymbol{s},t) \tag{1}$$

with $\mathrm{E}\big[x(\boldsymbol{s},t)\big] = \mu$: the expected value or the population mean and

$\varepsilon(\boldsymbol{s},t)$: random, spatially and temporally uncorrelated error assumed normally distributed with zero mean and variance σ^2.

The population mean, μ, is estimated from a sample:

$$\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

with n: the sample size

and the variance, σ^2, is estimated by:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \tag{3}$$

Its root square, $\hat{\sigma}$, is the standard deviation.

The coefficient of variation is defined as:

$$CV = 100 * \frac{\hat{\sigma}}{\overline{x}} \tag{4}$$

The coefficient of variation can be calculated for the temporal probability distribution function (pdf) at each space location as well as for the spatial pdf at each time instant. As an

example, Van Wesenbeeck and Kachanoski (1988) studied the spatial distribution (from 100 locations) of the surface soil water content throughout the growing season (65 measurement dates). They concluded that there exists a temporal dependence of the spatial variability and a spatial dependence of the temporal variability. They also suggested that random sampling of the locations should be avoided. James et al. (2003) defined the temporal variability of soil moisture content as the coefficient of variation at each site across the whole year (52 time instants) and used an analysis of variance (ANOVA) to test for differences between habitats (three sites) in the temporal variability of soil moisture content. They found that soil moisture varied significantly with time but not between habitats while the interaction between time and habitat was significant which reflects differences between habitats in temporal pattern of soil moisture. The coefficients of variation were significantly different between habitats. Other examples of the use of the coefficient of variation for the assessment of the spatial and temporal variabilities can be found in Ehrenfeld et al. (1997) and Guo et al. (2002).

The standard deviation and, hence, the coefficient of variation do not incorporate spatial and/or temporal information and therefore don't provide a representation of the nature of the spatial and/or temporal behaviour of a given soil property. It is considered as an indicator of global variability as opposed to local variability (Carter and Pearen, 1985), the latter is identified using correlogram, covariance function, or variogram which will be presented later in this chapter (section 4.3).

4.2.3. Variance component analysis

In the variance component analysis approach, the spatial and temporal contributions are considered as factors with random effects in the ANOVA model. In this way the different components of the total variance can be determined. Campbell et al. (1989) adopted the following model, for the log concentrations of a given solute ($\log c$) in subsample l (3) from sampling area k (2) at site j (2) and at time i (6 dates over 20 days):

$$\log c_{ijkl} = \mu + T_i + S_j + A_k + (A.U)_{kl} + \varepsilon_{ijkl} \tag{5}$$

with μ: the grand mean; T: the time effect; S: the site effect; A: the sampling area effect; $(A.U)$: the subsampling within sampling area effect, and ε: an error term.

The temporal and spatial variances were compared to the sampling area variance while the latter was compared to the subsampling within sampling area variance. They found that the temporal variation was far less significant than the variation between the two sites. Also they

found that the variation between areas a few meters apart was greater than that between the two sites even though the texture and the management of the sites were quite different. Other examples of the evaluation of the temporal and spatial variabilities based on the concept of variance component analysis can be found in Van Es (1993) and Van Es et al. (1999).

The variance component analysis is directly related to the coefficient of variation presented in the precedent section. The variance component estimates for each variability source are computed, and then the square root of these estimates is divided by the mean value. The output is exactly the definition of the coefficient of variation. Thus, the variance component analysis is suffering from the same drawbacks as the coefficient of variation approach.

4.2.4. Temporal stability

The concept of temporal stability was first introduced by Vachaud et al. (1985). It is defined as the time invariant association between spatial location and statistical parameters of soil properties. They distinguished mainly two approaches.

4.2.4.1. Relative differences

The first one is based on the concept of relative differences.

Let x_{ij} be the observed value of a soil property at location i (i=1, …, n) and time j (j=1, …, m). The relative differences δ_{ij} is defined as:

$$\delta_{ij} = \frac{x_{ij} - \overline{x}_j}{\overline{x}_j} \tag{6}$$

with $\overline{x}_j$: the spatial average for the time instant j.

Using δ_{ij} we can estimate, for each location i, the temporal average:

$$\overline{\delta}_i = \frac{1}{m} \sum_{j=1}^{m} \delta_{ij} \tag{7}$$

and its corresponding temporal standard deviation is defined as:

$$\sigma\left(\overline{\delta}_i\right) = \sqrt{\frac{1}{m-1} \sum_{j=1}^{m} \left(\delta_{ij} - \overline{\delta}_i\right)^2} \tag{8}$$

A zero value for $\overline{\delta}_i$ indicates that the temporal average $\overline{x}_j$ represents the average value over the whole study area at any time. The field average value is overestimated if $\overline{\delta}_i > 0$ while it is

underestimated when $\bar{\delta}_i < 0$. A more time stable location will be indicated by a small value of $\sigma\left(\bar{\delta}_i\right)$ whereas a high value of the latter is an indication of a less time stable location. The $\bar{\delta}_i$ values can be plotted against their rank with the corresponding temporal standard deviations.

Based on the relative differences, Kachanoski and De Jong (1988) showed that a useful test for temporal stability is the Pearson correlation coefficient between soil property values measured at consecutive time instants. They also showed that temporal stability exists if the relative differences remain constant between two time instants:

$$\delta_{i2} = \delta_{i1}. \tag{9}$$

This equation implies that

$$x_{i2} = \frac{\bar{x}_2}{\bar{x}_1} x_{i1}. \tag{10}$$

$\bar{x}_1$ and $\bar{x}_2$ being spatial average for the time instants $j = 1$ and 2, respectively.

Equation (10) establishes a linear relationship between the soil property at two different times with an intercept equal to zero and a slope equal to the ratio of the mean value observed at the second time frame to the mean value observed at the first time frame. Consequently a good test for the temporal stability is the correlation between soil property values measured at consecutive time frames (Kachanoski and De Jong, 1988).

Furthermore, Eq. (10) implies that the regression between relative differences from two consecutive time frames should have a zero intercept and a unity slope. This is another way to check the existence of temporal stability of the spatial pattern of a variable. However in general the regression between x_{i2} and x_{i1} is of the following form:

$$x_{i2} = a * x_{i1} + b \tag{11}$$

Using Eq. (10) and (11) we expect that if temporal stability exists, the following conditions are verified:

$$a = \frac{\bar{x}_2}{\bar{x}_1} \qquad \text{and} \qquad b = 0 \tag{12}$$

So another way to check the temporal stability is to test for a zero intercept and a unity slope of the regression between two consecutive time instants.

If there is a constant increase or decrease in the soil property, *w*, in all the locations between two time frames, we expect from Eq. (11) that (Kachanoski and De Jong, 1988):

$$a = 1 \qquad \text{and} \quad b = w \tag{13}$$

Eq. (13) indicates that if the slope of the regression between values at two consecutive time instants is not significantly different from one, then the regression intercept represents the constant change that has occurred between the two time instants.

These refinements of the concept of temporal stability were used by, among others, Van Wesenbeeck et al. (1988), Goovaerts and Chiang (1993), Ehrenfeldt et al. (1997), Da Silva et al. (2001), and Petrone et al. (2004).

The concept of relative differences is based on the same statistical parameters as the two precedent approaches (mean and standard deviation). However, it uses the standardised data (relative differences) for each time instant instead of the original data and the mean and standard deviation are computed for each space location. This concept is based on the assumptions of Gaussian distribution of data and the absence of autocorrelation. Both assumptions must be checked (see section 4.2.1) before the use of relative differences in the study of temporal stability. The normality can be checked using the standard statistical tests and the independence can be verified using either a test for autocorrelation like the Moran's I or by computing the variogram and determining its range. Normality can be gained via data transformation and independence by the selection of sample locations separated by distances larger than the autocorrelation range. For non normally distributed data, but still non autocorrelated, the Spearman rank correlation can be used. The latter is presented in the next section.

4.2.4.2. Spearman rank correlation

The second concept used to assess the temporal stability is the Spearman rank correlation coefficient. It refers to the tendency of a soil property, measured at different locations in space, to maintain their relative ranking over time. It is defined as:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\left(R_{ij} - R_{ik}\right)^2}{n\left(n^2 - 1\right)} \tag{14}$$

with R_{ij} and R_{ik} the ranks of z_{ij} observed at location i on time instants j and k, respectively. A value of this coefficient equal one indicates a perfect temporal stability between time instants j and k, and thus identity of ranks for any location whereas a lack of temporal stability implies that $r_s = 0$.

Many researchers used these two concepts to evaluate the temporal stability of different soil properties, among them, Reichardt et al. (1993), Farley and Fitter (1999), Wendroth et al. (1999), Campbell et al. (2001), Guo et al. (2002), and Si (2002).

The Spearman rank correlation coefficient is a non parametric statistics therefore it can be used even for non normally distributed data. However in the existence of autocorrelation, its use is not recommended and it will be cautious to use statistical methods which consider the spatial and temporal nature of the given soil property. In addition, even for independent data, the Spearman rank correlation inform us only about the relative ranking of space locations between two time instants and do not provide a quantification of the temporal change. This quantification can be obtained using either paired-*t* test for comparing mean values or the approach of Lesch et al. (1998). The former is presented in what follows while the latter is discussed in section 4.2.5.

4.2.4.3. Paired-t test

Equation [10] just indicates if there is a constant change or not between two time instants, but we still need to check if the mean values of *EC* between these two times are significantly different or not. To answer this question, the paired-*t* test (McClave and Sincich, 2000) is used. The test statistic *t* is computed as:

$$t = \frac{\bar{d}}{s_d / n^{1/2}}$$

(15)

with $\bar{d}$ the mean of the differences, d_i, between the observations for the second and the first time frames:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i = \frac{1}{n} \sum_{i=1}^{n} \left(x_{i2} - x_{i1} \right),$$

(16)

s_d is the standard deviation of the differences:

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(d_i - \bar{d} \right)^2}$$

(17)

and *n* the number of the differences, i.e. the number of locations.

The test statistic *t* has (*n*-1) degrees of freedom. The calculated statistic is compared to a tabulated *t* value with the same degrees of freedom and a probability of error of type I (taken generally equal to 5%). Equivalently the corresponding probability to the test statistic *t* is compared to 5% to check if the difference of mean values is significant or not.

4.2.5. Dynamic spatial variation

The concept of temporal stability processes each variable (EC_a and $EC_{2.5}$) separately and, as described before (section 4.2.4), it can quantify the temporal change even without testing if this change is significant or not. Lesch et al. (1998) discussed another approach that can test for change in the mean value as well as for a dynamic spatial variation between two time instants. Dynamic spatial variation refers to the spatially variable change of a soil property between two time instants, i.e., the change is not the same between two time instants for each location.

To test for the existence of dynamic spatial variation, Lesch et al. (1998) identifies different steps.

First of all, the field (EC_a) and the laboratory ($EC_{2.5}$) electrical conductivity data obtained at a first time instant, from n space locations, are used to establish a calibration equation based on regression model:

$$EC_{2.5} = b_1 EC_a + b_0 \tag{18}$$

The parameters of this regression equation are estimated using the ordinary least squares method.

In a second step, the residuals of the model are tested to check their spatial independence, normality, homoscedasticity and the linearity of the relationship. The three last assumptions can be verified using classical regression tools (Myers, 1986), see sections 4.2.1.1 to 4.2.1.3.

The assumption of spatial independence can be checked using the Moran test (Moran, 1950; Lesch et al., 1995b) which is summarized in what follows.

The Moran test is defined by:

$$I_M = \frac{e'We}{e'e} \tag{19}$$

where e represents the vector of the residuals from the regression model and W is a proximity matrix reflecting the neighborhood of observations. It incorporates the prior structure of dependence between space locations. Frequently, for n space locations, the elements of this matrix, w_{ij}, are function of the distance, d_{ij}, separating two locations. Thus:

$w_{ij} = 0$ when $i = j$, a location with itself and

$$w_{ij} = \frac{\frac{1}{d_{ij}^2}}{\sum_{i=1}^{n} \frac{1}{d_{ij}^2}} \text{ for } i \neq j \tag{20}$$

The expected value, $E[I_M]$, and the variance, $\sigma^2[I_M]$, of the Moran test are (Brandsma and Ketellapper, 1979):

$$E[I_M] = \frac{tr(MW)}{n-p} \qquad (21)$$

$$\sigma^2[I_M] = \frac{tr(MWMW') + tr(MWMW) + \{tr(MW)\}^2}{(n-p)(n-p+2)} - \{E[I_M]\}^2 \qquad (22)$$

where p equals the number of regression parameters, $tr(\ldots)$ denotes the trace of a matrix, and

$$M = I - X(X'X)^{-1}X' \qquad (23)$$

where I is an identity matrix and X is the design matrix (intercept and explanatory variables, in our case EC_a).

Cliff and Ord (1981) showed that, in the case of residuals from ordinary least squares regression, the Moran test follows an asymptotic normal distribution. Thus, to test its significance, it is first standardized:

$$Std[I_M] = \frac{I_M - E[I_M]}{\sigma[I_M]} \qquad (24)$$

and then this standardized statistic is compared to the standard normal distribution.

In the next step, a mixed ANOVA model is set:

$$x_{ij} = \mu + t_i + b_j + \varepsilon_{ij} \qquad (25)$$

$i=1, 2$ are the two time instants and $j=1, \ldots, n$ are the locations.

This model is equivalent to:

$$x_{2j} - x_{1j} = d_j = d_0 + \eta_j + (\varepsilon_{2j} - \varepsilon_{1j}) \qquad (26)$$

with $d_0 = \bar{t}_2 - \bar{t}_1$ is the difference between mean values for the second and first time instants and η_j independently and identically normally distributed with zero mean and variance equal to σ_{tb}^2: $\eta_j \approx N(0, \sigma_{tb}^2)$.

If x_{1j} represents the observed laboratory electrical conductivity at the first time instant and at locations j ($j=1, \ldots, n$) and x_{2k} represents the same property observed at the second time instant at the locations k ($k=1, \ldots, m$), a regression model for the first time is:

$$x_1 | X_1 = X_1\beta + \varepsilon_1 \text{ with } \varepsilon_1 \approx N(0, \sigma^2 I). \qquad (27)$$

For the second time instant, and using the explanatory variables of the first time instant, the model is:

$$x_2 \big| X_1 = X_1\beta + d_0 + \eta + \varepsilon_2 \tag{28}$$

with $\eta_j \approx N\left(0, \theta^2 I\right)$ and $\varepsilon_2 \approx N\left(0, \sigma^2 I\right)$

Thus

$$x_2 \big| X_1 - x_1 \big| X_1 = d_0 + \eta + \left(\varepsilon_2 - \varepsilon_1\right) \tag{29}$$

X_1 being the design matrix with two columns: ones for the intercept and the observed values of EC_a surveyed during the first time frame at N space locations with $N > n, m$.

Equation (29) is equivalent to

$$x_{2j} - x_{1j} = d_j = d_0 + \eta_j + \left(\varepsilon_{2j} - \varepsilon_{1j}\right) \tag{30}$$

with d_0 representing the shift in the average value between the two time instants and η representing the dynamic spatial variation.

Based on the calibration data from the first time instant, equation (27) is estimated:

$$\hat{x}_1 \big| X_1 = X_n b \quad \text{where} \tag{31}$$

$\hat{x}_1$ represents the predicted $EC_{2.5}$,

X_n represents the design matrix associated with the n calibration locations, and

b represents the estimated regression parameters.

Using the EC_a data from the first time frame and the regression parameters determined above, $EC_{2.5}$ for the second time frame is predicted. Let $\hat{x}_1$ denotes this predicted $EC_{2.5}$ and x_2 denotes the $EC_{2.5}$ observed at the second time frame, both, at m space locations, m may be equal or different than n.

The prediction error associated with m new locations would be distributed as multivariate normal with a mean zero and a variance-covariance matrix, cov(m):

$$\text{cov}(m) = \sigma^2(I_m + H_m) \text{ with } I_m \text{ an identity matrix of dimension } m \text{ x } m \text{ and}$$

$$H_m = X'_m(X'_n X_n)X_m \text{ with } X_m \text{ the design matrix associated with the } m \text{ new locations.}$$

From equation (28), it can be deduced that the prediction error associated with the m sites would be normally distributed with a mean value d_0 and a variance-covariance matrix $\theta^2 I_m + \sigma^2(I_m + H_m)$. This means that the differences between the observed values at the second time instant and the predicted values at the first time instant could be different than zero and they will contain two sources of error.

The differences between the observed and the predicted $EC_{2.5}$ at the second time frame are:

$$d_i = x_{2i} - \hat{x}_{1i} \tag{32}$$

with $i=1,\ldots, m$. These differences have the mean:

$$\bar{u} = \frac{1}{m}\sum_{i=1}^{m} d_i \qquad (33)$$

which represents an unbiased estimate of d_0. The corresponding variance is:

$$w^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(d_i - \bar{u}\right)^2 . \qquad (34)$$

Additionally, let s^2 represents the mean square error related to equation (31) with n-p-1 degrees of freedom with p being the number of regression parameters. To test if $\theta^2 > 0$, the statistic φ is computed as follows:

$$\varphi = \frac{\left(d - \bar{u}\right)' \Sigma^{-1}\left(d - \bar{u}\right)}{\left(m-1\right)s^2} \qquad (35)$$

where $\Sigma = I_m + H_m$.

The statistic φ is compared to an F distribution with (m-1) and (n-p-1) degrees of freedom. Furthermore, the estimated value of θ^2 is:

$$v^2 = w^2 - s^2\left(1 + \lambda_1 - \lambda_2\right) \qquad (36)$$

with

$$\lambda_1 = \frac{1}{m}\sum_{i=1}^{m} h_{ii} \qquad (37)$$

and $\lambda_2 = \dfrac{1}{m(m-1)}\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} h_{ij} \,\forall i \neq j \qquad (38)$

where h_{ij} represents the ith, j^{th} element of H_m.

The shift in the average value $d_0 = 0$ can be tested using an approximate t-test:

$$c = \frac{\bar{u}}{g} \qquad (39)$$

where $g = \sqrt{\dfrac{1}{m}v^2 + 2s^2\left(\dfrac{1}{m} + h_\mu\right)} \qquad (40)$

with $h_\mu = x'_\mu\left(X'_n X_n\right)^{-1} x_\mu \qquad (41)$

where $x_\mu = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m} x_i . \qquad (42)$

The statistic c is compared with a t distribution with n-p-1 degrees of freedom.

An application of the approach of Lesch et al. (1998) can be found in Tóth et al. (2002a).

The approach of Lesch et al. (1998) is based on the determination of a linear regression between $EC_{2.5}$ and EC_a, hence it assumes that the residuals from the regression model are normally distributed with equality of variances and spatially non autocorrelated. All these assumptions must be checked (see section 4.2.1). If spatial autocorrelation is present in the residuals the estimates of the regression parameters and their variances are biased and the statistical tests are non longer valid. In this special case, spatial regression (see section 7.3) and/or geostatistical approaches (sections 4.3.3 to 4.3.5) are recommended.

4.3. Structural analysis

4.3.1. Definitions

The analysis of space-time data requires the definition of a space-time continuum with a coordinate system and a measure of space-time distance and models and techniques which make possible the link between spatio-temporally distributed data (Christakos et al., 2002).

A spatio-temporal continuum E is a set of points associated with a continuous arrangement of events combined with their temporal order in which space represents the order of coexistence of events while time represents the order of their successive existence. For a continuum to be useful, it requires to be equipped with a coordinate system. The latter will allow to identify precisely any point in space and time along with a metric which allows to measure distances in the space-time domain. The spatial coordinates are defined, in general, in two dimensions $s = (s_1, s_2) \in S \subset R^2$, S the spatial domain, and the temporal coordinate t along the temporal axis $T \subset R^1$ such that the space-time coordinates are:

$$p = (s, t) \in E = SxT .$$

The most used coordinate system is the Euclidian with its particular case the rectangular or Cartesian coordinates. The geographic coordinates (latitude and longitude) are an example of currently used coordinates. The term georeferencing is used to design the process of registering data to a coordinate system using, for example, a GPS device.

4.3.2. Space-time random field model

Christakos et al. (2002) defined space-time random field (STRF) model as 'a mathematical construction that rigorously represents the distribution of natural phenomena across space and time. The STRF model provides scientifically meaningful representation, explanation, and prediction of these phenomena in uncertain environments. These uncertainties may be due to measurement errors, heterogeneous databases, erratic fluctuations in the space-time variation of the underlying process, and insufficient knowledge'.

The STRF model was presented in more detail in Christakos (1992) and Christakos and Hristopoulos (1998), and we provide here a brief description.

A STRF, $X(\boldsymbol{p})$, is a collection of realizations χ of the space-time distribution of a natural variable. It can be viewed as a collection of correlated random variables $\boldsymbol{x}=[x_1, \ldots, x_n]'$ at the space-time points $\boldsymbol{p}=[p_1, \ldots, p_n]'$ and a realization of the STRF at these points is $\chi=[\chi_1, \ldots, \chi_n]'$.

It assigns a probability that a realization χ in n dimensions will occur following the multivariate pdf of $X(\boldsymbol{p})$:

$$\text{Prob}(\boldsymbol{\chi}) = P_x\left[\chi_1 \leq x_1 \leq \chi_1 + d\chi_1, \ldots, \chi_n \leq x_n \leq \chi_n + d\chi_n\right] = f_x(\boldsymbol{\chi})d\boldsymbol{\chi} \tag{43}$$

The pdf is the derivative of the probability that a STRF realization assumes values $\leq \chi$:

$$F_x(\boldsymbol{\chi}) = P_x\left[x_1 \leq \chi_1, \ldots, x_n \leq \chi_n\right] = \int_{-\infty}^{\chi_1} \ldots \int_{-\infty}^{\chi_n} d\chi_1 \ldots d\chi_n f_x(\boldsymbol{\chi}) \tag{44}$$

The equation above represents the cumulative distribution function (cdf).

The above two definitions represent a complete characterization of the STRF. A partial but sufficient characterization of the STRF is provided by its space-time statistical moments.

The first order moment, which is the mean function, expresses trends or systematic structures in space-time:

$$m_x(\boldsymbol{p}) = E\left[X(\boldsymbol{p})\right] = \int \chi f_x(\boldsymbol{\chi})d\boldsymbol{\chi} \tag{45}$$

The moment of second order, the covariance function, expresses correlations and dependencies between two different points $\boldsymbol{p}$ and $\boldsymbol{p}'$:

$$\begin{aligned} c_x(\boldsymbol{p}, \boldsymbol{p}') &= E\left\{\left[X(\boldsymbol{p}) - m_x(\boldsymbol{p})\right]\left[X(\boldsymbol{p}') - m_x(\boldsymbol{p}')\right]\right\} \\ &= \iint \left[\chi - m_x(\boldsymbol{p})\right]\left[\chi' - m_x(\boldsymbol{p}')\right] f_x(\chi, \chi')d\chi d\chi' \end{aligned} \tag{46}$$

with $f_x(\chi, \chi')$ the bivariate pdf of the random function taken between points $\boldsymbol{p}$ and $\boldsymbol{p}'$.

In addition to these two moments, an STRF can be also characterized by its variogram:

$$\gamma_x(\boldsymbol{p}, \boldsymbol{p}') = \tfrac{1}{2} E\left\{\left[X(\boldsymbol{p}) - X(\boldsymbol{p}')\right]^2\right\}$$

$$= \tfrac{1}{2}\iint (\chi - \chi')^2 f_x(\chi, \chi')d\chi d\chi' \tag{47}$$

4.3.3. Conceptual approaches

Christakos and Raghu (1996) reported that some forms of averaging over time or space are done before the analysis of the spatial or temporal random fields. In doing so, the joint space-time variability is not accounted for and there is some loss of information. Initially to consider both space and time variability, some authors resorted to some simplifications. For example, Bilonick (1985) and Egbert and Lettenmaeir (1986) decomposed their STRF in a purely spatial and a purely temporal components whereas Rodriguéz-Iturbe and Eagelson (1987) considered separable models where spatial variations where assumed to be independent of temporal variations. Some of these simplifications with some generalizations are discussed in section 4.3.6 in which covariance function and variogram models are presented.

Three main approaches for the analysis of the geostatistical space-time data can be distinguished (Stein et al., 1998; Stein and Sterk, 1999; Kyriakidis and Journel, 1999):
- methods using an STRF, thus considering the joint space-time variability;
- methods based on vectors of spatial random fields (SRF) for the case of many space locations and few time instants (Goovaerts and Sonnet, 1993; Papritz and Fluhler, 1994; Bogaert and Christakos, 1997). It does not include the temporal dependence existing between observations and can predict only at the observed time instants. The spatial variability is modelled either through a separable spatial variogram for each time instant, or by a single spatial variogram considering time instants as replicates;
- methods based on vectors of time series (temporal random fields, TRF) for the case of long time series with few space locations (Solow and Gorelick, 1986; Rouhani and Wackernagel, 1990). This approach doesn't take into account the spatial dependence and it predicts only at the observed locations. In a similar way as above, independent TRF or TRF as replicates in space can be considered.

For the second approach, the focus of the analysis is on smooth interpolated soil attribute maps over specific time instants. The intention is to capture single instantaneous snapshots (a static picture) of the natural process. The objective can be also the comparison of the various

maps or the detection of the temporal persistence or change in the spatial patterns (Goovaerts and Chiang, 1993; Van Meirvenne et al., 1996).

The intention of the third group is to capture a sequence of successive snapshots at single space locations, which give temporal profiles.

Only the first group of methods includes both the spatial and temporal dependencies so the interpolation is more precise and can be done for unsampled time instants at unsampled space locations. The focus is here on video sequence of successive spatial pictures like a movie. This approach was used to analyze our salinity data set.

As a fully satisfactory stochastic model should involve explicitly both spatial and temporal aspects, we present in the following sections the STRF concept and, subsequently, some special cases. We present in section 4.3.4 the joint STRF while the spatial/temporal random fields are presented in sections 4.3.5 and simplifications and other classes of models in 4.3.6. Finally the space-time cross covariance function and cross variogram are discussed in section 4.3.7.

4.3.4. Single space-time random field

4.3.4.1. Space-time covariance function

For spatially homogeneous and temporally stationary random fields, the mean function is constant:

$$m_x(\boldsymbol{p}) = m_x \tag{48}$$

and the covariance function depends only on the spatial lag $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}'$ and the temporal lag $\tau = t - t'$ between any two points $\boldsymbol{p} = (\boldsymbol{s}, t)$ and $\boldsymbol{p}' = (\boldsymbol{s}', t') = (\boldsymbol{s} + \boldsymbol{h}, t + \tau)$:

$$c_x(\boldsymbol{p}, \boldsymbol{p}') = c_x(\boldsymbol{s} - \boldsymbol{s}', t - t') = c_x(\boldsymbol{h}, \tau) \tag{49}$$

The space-time covariance function can be written:

$$c_x(\boldsymbol{h}, \tau) = E\left\{ [X(\boldsymbol{s}, t) - m_x][X(\boldsymbol{s} + \boldsymbol{h}, t + \tau) - m_x] \right\} = E[X(\boldsymbol{s}, t)X(\boldsymbol{s} + \boldsymbol{h}, t + \tau)] - m_x^2 \tag{50}$$

If the mean function is known (in fact, it is estimated from data), m, the moments estimator of the space-time covariance function is:

$$\hat{c}_x(\boldsymbol{h}, \tau) = \frac{1}{N(\boldsymbol{h}, \tau)} \sum_{i,j=1}^{N(\boldsymbol{h}, \tau)} \left\{ [x(\boldsymbol{s}_i, t_j) - m][x(\boldsymbol{s}_i + \boldsymbol{h}, t_j + \tau) - m] \right\} \tag{51}$$

with $N(\boldsymbol{h}, \tau)$ the number of pairs of data separated by the spatial lag $\boldsymbol{h}$ and the temporal lag τ.

This estimation does not require that the space locations are collocated for each time instant. However, in the case of collocation, the same n_s space locations observed at each of the n_t time instant, the estimate of the above function for zero spatial (temporal) lag is equivalent to the average of the estimated temporal (spatial) covariance functions calculated separately for the n_s (n_t) different space locations (time instants).

This definition of space-time covariance function was applied to our salinity data set in chapters 7 and 8.

Examples of application of the space-time covariance function include Bennett (1975) for the analysis of population diffusion and Bell (1987) for the study of rainfall.

For spatially non homogeneous and temporally non stationary random fields, more general forms of STRFs were considered. For example, the STRF v/μ was developed by Christakos (1991; 1992) and applied by Christakos and Bogaert (1996), Christakos and Raghu (1996), Vyas and Christakos (1997), and Christakos and Vyas (1998).

The STRF v/μ is based on a mathematical operator which filters out space-time trend functions involving polynomials of degree v in space and μ in time. The resulting residual STRF is then spatially homogeneous and temporally stationary, and the models reviewed here can be applied to them. The space-time continuity orders, v and μ, allow a quantitative assessment of the average continuity. The distribution of the difference $v - \mu$ over the study area informs about the relative trends in space and time. A positive difference implies a more complex spatial structure and the process can be handled as a space random function while a negative difference indicates a more complex temporal trends and the process can be considered as a temporal random function.

4.3.4.2. Space-time variogram

If the increments $X(s+h, t+\tau) - X(s,t)$ are second order stationary, the space-time variogram is defined as:

$$\gamma_x(\boldsymbol{h}, \tau) = \frac{1}{2} \text{var}\left[X(\boldsymbol{s}+\boldsymbol{h}, t+\tau) - X(\boldsymbol{s},t)\right] = \frac{1}{2} E\left\{\left[X(\boldsymbol{s}+\boldsymbol{h}, t+\tau) - X(\boldsymbol{s},t)\right]^2\right\} \quad (52)$$

Its moment-based estimator is given by (Stein et al., 1998) :

$$\hat{\gamma}_x(\boldsymbol{h}, \tau) = \frac{1}{2N(\boldsymbol{h}, \tau)} \sum_{i,j=1}^{N(\boldsymbol{h}, \tau)} \left[x(\boldsymbol{s}_i, t_j) - x(\boldsymbol{s}_i + \boldsymbol{h}, t_j + \tau)\right]^2 \quad (53)$$

with $x(s_i, t_j)$ and $x(s_i + h, t_j + \tau)$ are pairs of observations with a spatial distance equal **h** at the time instant t_j, the total number of such pairs is $N_j(h)$, and a temporal distance equal τ, the number of such pairs is $N(\tau)$.

Bilonick (1985) and Stein et al. (1998) are examples of application of the space-time variogram.

To introduce a space-time anisotropy, the spatial and temporal distances are combined to give a space-time distance, *d*, as follows:

$$d = \sqrt{h^2 + \varphi \tau^2} \tag{54}$$

with φ the space-time anisotropy ratio. In addition, the spatial anisotropy can be also incorporated by modifying the above distance by including the spatial anisotropy ratio ψ:

$$d = \sqrt{(h_x^2 + \psi h_y^2) + \varphi \tau^2} \tag{55}$$

with h_x and h_y spatial distances along the *x* and *y* axes.

For a second-order stationary STRF, the space-time variogram and space-time covariance function are related to each other:

$$\gamma_x(h, \tau) = \sigma_x^2 - c_x(h, \tau) \tag{56}$$

with $\sigma_x^2 = c_x(0, 0)$, the variance of the STRF.

4.3.5. Spatial/temporal random fields

4.3.5.1. Spatial/temporal covariance function

For each of the n_t time instants corresponds a spatial random field $X(s, t_j)$, j=1, …, n_t. The spatial variability is modelled by the covariance function (Stein and Sterk, 1999):

$$c_j(h) = E\left[X(s, t_j) X(s + h, t_j) \right] - \mu_j^2 \tag{57}$$

where μ_j is the mean value at the time instant t_j.

Its moment-based estimator is given by:

$$\hat{c}_j(h) = \frac{1}{N_j(h)} \sum_{i=1}^{N_j(h)} \left[x(s_i, t_j) - \mu_j \right]\left[x(s_i + h, t_j) - \mu_j \right] \tag{58}$$

The space-time covariance function reduces to the spatial covariance function when the temporal lag is zero:

$$c_j(h) = c_x(h, 0) \tag{59}$$

In a similar way to the SRF and by symmetry, the temporal covariance function, $c_i(\tau)$, is defined and estimated for each space location s_i, i=1, …, n_s, to which corresponds the temporal random field $X(s_i, t)$.

The temporal covariance function is a special case of the space-time covariance function when the spatial lag is zero:

$$c_i(\tau) = c_x(\mathbf{0}, \tau) \tag{60}$$

Both the spatial and the temporal covariance functions, on which was built the space-time covariance function, were used in chapters 7 and 8.

4.3.5.2. Spatial/temporal cross covariance function

The finite set of time instants, t_j, j=1, …, n_t, corresponds to simultaneous responses observed on all space locations, s_i, i=1, …, n_s. The relationship between two SRFs $X(s, t_j)$ and $X(s, t_{j'})$, observed at time instants t_j and $t_{j'}$ is characterized by the spatial cross covariance function:

$$c_{jj'}(\mathbf{h}) = E\left[X(s, t_j) X(s + \mathbf{h}, t_{j'})\right] - E\left[X(s, t_j)\right] E\left[X(s + \mathbf{h}, t_{j'})\right] \tag{61}$$

It is estimated as follows (Ettema et al., 2000):

$$\hat{c}_{jj'}(\mathbf{h}) = \frac{1}{N_\tau(\mathbf{h})} \sum_{i=1}^{N_\tau(\mathbf{h})} \left[x(s_i, t_j) - m_j\right]\left[X(s_i + \mathbf{h}, t_{j'}) - m_{j'}\right] \tag{62}$$

with $m_j = \dfrac{1}{n_s} \sum_{i=1}^{n_s} x(s_i, t_j)$ is the estimate of the mean of the SRF $X(s, t_j)$ at the time instant t_j.

$m_{j'}$ is defined in the same way for the time instant $t_{j'}$.

For a second-order stationary STRF, the spatial cross covariance function can be derived from the space-time covariance function by fixing the latter at a given time lag, $\tau = t_{j'} - t_j$:

$$c_{jj'}(\mathbf{h}) = c_\tau(\mathbf{h}) = c_x(\mathbf{h}, \tau) \qquad \text{for a fixed } \tau. \tag{63}$$

The spatial cross covariance function is then easily obtained.

By analogy to the definition of spatial cross covariance function, the finite set of space locations corresponds to simultaneous responses observed on all time instants and the relationship between two TRFs observed at locations, s_i and $s_{i'}$, is characterized by the temporal cross covariance function $c_{ii'}(\tau)$.

For a second-order stationary STRF, the temporal cross covariance function can be derived from the space-time covariance function by fixing the latter at a given space lag, $\mathbf{h} = s_{i'} - s_i$:

$$c_{ii'}(\tau) = c_{\boldsymbol{h}}(\tau) = c_x(\boldsymbol{h},\tau) \quad \text{for a fixed } \boldsymbol{h}. \tag{64}$$

Consequently the temporal cross covariance function can be easily obtained.

The assumption of the space-time independence of data can be tested by calculating the spatial (temporal) cross covariance functions between all pairs of time instants (space locations) and compare them to the spatial (temporal) covariance functions computed for each time instant (space location).

4.3.5.3. Spatial/temporal variogram

For a fixed time instant, t_j, the spatial variogram is defined as:

$$\gamma_j(\boldsymbol{h}) = \frac{1}{2}\text{var}\left[X(\boldsymbol{s},t_j) - X(\boldsymbol{s}+\boldsymbol{h},t_j)\right] = \frac{1}{2}E\left\{\left[X(\boldsymbol{s},t_j) - X(\boldsymbol{s}+\boldsymbol{h},t_j)\right]^2\right\} \tag{65}$$

Its corresponding moment estimator is obtained by:

$$\hat{\gamma}_j(\boldsymbol{h}) = \frac{1}{2N_j(\boldsymbol{h})}\sum_{i=1}^{N_j(\boldsymbol{h})}\left[x(\boldsymbol{s}_i,t_j) - x(\boldsymbol{s}_i+\boldsymbol{h},t_j)\right]^2 \tag{66}$$

with $N_j(\boldsymbol{h})$ is the number of pairs [$x(\boldsymbol{s}_i,t_j)$, $x(\boldsymbol{s}_i+\boldsymbol{h},t_j)$] of data separated by a spatial distance $\boldsymbol{h}$, at the time instant t_j.

As above, the spatial variogram is a special case of the space-time variogram, i.e. when the temporal lag is zero:

$$\gamma_j(\boldsymbol{h}) = \gamma_x(\boldsymbol{h},0) \tag{67}$$

The spatial variograms include the spatial dependence but the dependence in time is not accounted for. It is useful in the case of rich data in space and scarce data in time.

The above procedure was used, for example, by Sterk and Stein (1997) and Ettema et al. (1998), among others.

Special cases of the approach above can be found in the literature. For example, Petitgas (1997) inferred, upon the condition that the spatial variograms were similar for the different time instants, a time-averaged spatial variogram by calculating the mean semi-variances (from all the time instants) for each spatial lag.

Many applications of separate spatial variograms, one for each time instant, can be found in the soil science literature. Among others, soil carbon (Van Meirvenne et al., 1996; Chevallier

et al., 2000), soil salinity and sodicity (Agrawal et al., 1995 ; Mostafa and Yomota, 1998 ; Utset and Castellanos, 1999 ; Moameni and Stein, 2002 ; Cetin and Kirda, 2003), soil water (Wendroth et al., 1999; Schume et al., 2003), nitrate (Bruckler, 1997), soil biological activity (Gorres et al., 1997), soil strength (Castrignanò et al., 2002), and soil fertility (Delcourt et al, 1996; Shi et al., 2002) were analyzed for their spatial and temporal change.

By analogy with the spatial variogram, the temporal variogram, $\gamma_i(\tau)$, at location s_i, i=1, …, n_s, can be defined and estimated.

Similarly to the spatial variogram, the temporal variogram is a special case of the space-time variogram when the spatial lag is zero:

$$\gamma_i(\tau) = \gamma_x(\mathbf{0}, \tau) \tag{68}$$

The temporal dependence is included whereas the spatial dependence is not considered. Thus it is useful if many data are collected in time and few in space. Petitgas (1997) used the same procedure as for spatial variogram to compute a space-averaged temporal variogram.

4.3.5.4. Spatial/temporal cross variogram

Following the same approach as for section 4.3.5.2, the spatial cross variogram is defined, for second order stationary increments, as follows:

$$\gamma_{jj'}(\mathbf{h}) = \frac{1}{2} E\left\{\left[X(\mathbf{s},t_j) - X(\mathbf{s}+\mathbf{h},t_j)\right]\left[X(\mathbf{s},t_{j'}) - X(\mathbf{s}+\mathbf{h},t_{j'})\right]\right\} \tag{69}$$

Its moment-based estimator is given by:

$$\hat{\gamma}_{jj'}(\mathbf{h}) = \frac{1}{2N_\tau(\mathbf{h})} \sum_{i=1}^{N_\tau(\mathbf{h})} \left[x(\mathbf{s}_i,t_j) - x(\mathbf{s}_i+\mathbf{h},t_j)\right]\left[x(\mathbf{s}_i,t_{j'}) - x(\mathbf{s}_i+\mathbf{h},t_{j'})\right] \tag{70}$$

with $N_\tau(\mathbf{h})$ is defined as $N(\mathbf{h},\tau)$ for a particular fixed temporal lag τ.

When $\tau = 0$, the space-time variogram is equal to the spatial variogram or the spatial cross variogram between time t_j and itself:

$$\gamma_{jj}(\mathbf{h}) = \gamma_x(\mathbf{h},0) \qquad \forall j = 1,...,n_t \tag{71}$$

The same is valid for the covariance functions:

$$c_{jj}(\mathbf{h}) = c_x(\mathbf{h},0) \qquad \forall j = 1,...,n_t \tag{72}$$

In a similar way, the temporal cross variogram, $\gamma_{ii'}(\tau)$, is defined and estimated.

When $h = 0$, the space-time variogram is equal to the temporal variogram or the temporal cross variogram between space location s_i and itself. The same is valid for the covariance functions.

The definition of the spatial (temporal) cross variogram requires that data are observed at the same space locations for two different time instants (the same time instants for two different space locations). If it is not the case, the spatial (temporal) pseudo cross variogram (Papritz et al., 1993) needs to be used. The spatial pseudo cross variogram is presented in the next section.

4.3.5.5. Spatial/temporal pseudo cross variogram

The spatial pseudo cross variogram between two SRFs $X(s_i, t_1)$ and $X(s_{i'}, t_2)$, observed at two time instants t_1 and t_2 but not necessarily at the same space locations (s_i and $s_{i'}$) is defined by (Clark et al, 1989; Myers, 1991; Cressie, 1993):

$$\gamma_{21}{}^p(\boldsymbol{h}) = \frac{1}{2}\operatorname{var}\left[X(\boldsymbol{s}, t_1) - X(\boldsymbol{s} + \boldsymbol{h}, t_2)\right] = \frac{1}{2}E\left\{\left[X(\boldsymbol{s}, t_1) - X(\boldsymbol{s} + \boldsymbol{h}, t_2)\right]^2\right\} \tag{73}$$

Its moment-based estimator is given by (Papritz et al., 1993; Papritz and Fluhler, 1994):

$$\hat{\gamma}_{21}{}^p(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{i=1}^{N(\boldsymbol{h})}\left[x(\boldsymbol{s}_i, t_1) - x(\boldsymbol{s}_i + \boldsymbol{h}, t_2)\right]^2 \tag{74}$$

Zhang et al. (1992, 1999) used cokriging with pseudo cross variograms to estimate spatial distributions of soil chemicals.

The temporal pseudo cross variogram can be defined in a similar way between two TRFs $X(s_1, t_j)$ and $X(s_2, t_{j'})$, observed at two space locations s_1 and s_2 but not necessarily at the same tim instants (t_j and $t_{j'}$).

4.3.6. Other classes of space-time covariance models

De Cesare et al. (2001b) and De Iaco et al. (2002) distinguished five types of space-time covariance models:

- the metric model (Dimitrakopouolos and Luo, 1994):

$$c_{st}(\boldsymbol{h}, \tau) = C(a^2|\boldsymbol{h}|^2 + b^2\tau^2) \quad \text{with } a, b \in R^2 \tag{75}$$

In this equation, the same type of model is assumed for the spatial and temporal covariances, with possibly different ranges.

- the product or separable model (Rodriguez-Iturbe and Mejia, 1974; De Cesare et al., 1997; Bourgine et al., 2001; Heuvelink et al. 1997):

$$c_{st}(\boldsymbol{h}, \tau) = c_s(\boldsymbol{h}) c_t(\tau) \qquad (76)$$

In this case the spatial dependence is separated from the temporal one. However this model is severely limited, since for any pair of space locations the cross covariance function of the two time series always has the same shape. The same is valid for any pair of time instants and the cross covariance function of the two SRFs.

- Another form of the separability is the linear model (Rouhani and Hall, 1989; Heuvelink et al., 1997) where the spatial and temporal covariances are additive:

$$c_{st}(\boldsymbol{h}, \tau) = c_s(\boldsymbol{h}) + c_t(\tau) \qquad (77)$$

For some data configuration, the covariance matrices are singular (Myers and Journel, 1990; Rouhani and Myers, 1990)

- The integrated product or the nonseparable model (Cressie and Huang, 1999)
- The product sum model (De Cesare et al, 2001b):

$$c_{st}(\boldsymbol{h}, \tau) = k_1 c_s(\boldsymbol{h}) c_t(\tau) + k_2 c_s(\boldsymbol{h}) + k_3 c_t(\tau) \qquad (78)$$

with k_1, k_2, and k_3 are constants function of the different sill variances (spatial, temporal, and spatio-temporal).

De Cesare et al. (2001b) showed that the product model is obtained from the product sum model by setting $k_2 = k_3 = 0$, and the linear model is obtained by setting $k_1 = 0$.

Fortran programs (De Cesare et al., 2002) are available for the computation of the product and the product sum covariance models.

The product sum covariance model was generalized to a general product sum model (De Iaco et al., 2001a) which provides a large new class of models and is easily modelled using techniques similar to those used for modelling spatial variograms.

The space-time generalized product sum variogram model is given by (De Iaco et al., 2001a):

$$\gamma_{st}(\boldsymbol{h}, \tau) = \gamma_{st}(\boldsymbol{h}, 0) + \gamma_{st}(\boldsymbol{0}, \tau) - k \gamma_{st}(\boldsymbol{h}, 0) \gamma_{st}(\boldsymbol{0}, \tau) \qquad (79)$$

with k a constant, function of different sills, and $\gamma_{st}(\boldsymbol{h}, 0)$ and $\gamma_{st}(\boldsymbol{0}, \tau)$ are spatial and temporal bounded variograms, respectively.

De Iaco et al. (2001b, 2002) applied the generalized product sum model for the mapping of total air pollution in Milan, Italy.

The product, product sum, and integrated product models were further generalized to provide more non separable models (De Iaco et al., 2002).

4.3.7. Space-time cross covariance function and cross variogram

De Iaco et al. (2004) presented the multivariate STRF with the special case of the bivariate STRF. Let

$$X(s,t) = \left[X_1(s,t), X_2(s,t) \right] = Y(s,t) + M(s,t) \tag{80}$$

be a bivariate STRF with

$M(s,t)$ representing the trend component such that $M(s,t) = \left[M_1(s,t), M_2(s,t) \right] = M$, and

$Y(s,t)$ is a second order stationary STRF with its expected vector $E\left[Y(s,t) \right] = 0$, its covariance function $c_y(h,\tau)$, and its variogram $\gamma_y(h,\tau)$.

The covariance function is given by:

$$c_y(h,\tau) = \mathrm{cov}\left[X(s+h,t+\tau), X(s,\tau) \right] = E\left\{ \left[X(s,t) - M \right] \left[X(s+h,t+\tau) - M \right]' \right\} = c(h,\tau) = \left[c_{\alpha\beta}(h,\tau) \right] \tag{81}$$

where $c_{\alpha\beta}(h,\tau)$ are the cross covariance between the space-time random variables $X_\alpha(s,t)$ and $X_\beta(s,+h, t+\tau)$ when $\alpha \neq \beta$, and the autocovariance when $\alpha = \beta$.

The variogram is defined as:

$$\gamma_y(h,\tau) = \frac{1}{2} E\left\{ \left[Y(s+h,t+\tau) - Y(s,\tau) \right] \left[Y(s+h,t+\tau) - Y(s,\tau) \right]' \right\} = \left[\gamma_{\alpha\beta}(h,\tau) \right] \tag{82}$$

where $\gamma_{\alpha\beta}(h,\tau)$ are the cross variogram between the space-time random variables $X_\alpha(s,t)$ and $X_\beta(s,+h, t+\tau)$ when $\alpha \neq \beta$, and the autovariogram when $\alpha = \beta$.

4.3.8. Purely and joint spatial/temporal components

More complex natural phenomena require the integration of different components, mostly if some form of nonergodicity is suspected. In this way, Bogaert (1996a) developed a regressive space-time model which includes, in addition to a space-time mean function ($\mu(s,t)$), purely spatial ($M_s(s)$), purely temporal ($M_t(t)$), and a joint space-time ($Y(s,t)$) components:

$$X(s,t) = \mu(s,t) + M_s(s) + M_t(t) + Y(s,t) \tag{83}$$

The three stochastic components are considered to be independent and hence the covariance function and variogram of STRF are given, respectively, by:

$$c_x(\boldsymbol{h}, \tau) = c_s(\boldsymbol{h}) + c_t(\tau) + c_y(\boldsymbol{h}, \tau) \tag{84}$$

$$\gamma_x(\boldsymbol{h}, \tau) = \gamma_s(\boldsymbol{h}) + \gamma_t(\tau) + \gamma_y(\boldsymbol{h}, \tau) \tag{85}$$

Bogaert and Christakos (1997) applied the above model to the analysis of solute content from the Dyle river, Belgium while Beckers (1997) used this model in the analysis of space-time field experiments.

4.3.9. Comparison of single STRF with multiple SRF/TRF

Using the single STRF, the spatio-temporal variability is modelled by a joint space-time covariance function whereas the approach of multiple SRF/TRF models this variability via the linear model of coregionalisation, LMC (Journel and Huijbregts, 1978). The single approach allows prediction (space-time kriging, see section 4.4.4) at any location in space and any instant in time whereas the prediction through cokriging (section 4.4.3) is limited, under LMC, to the observed time instants in the case of n_t correlated SRF and to the observed space locations in the case of n_s correlated TRF. The joint space-time covariance function cannot be estimated reliably if the number of space locations and/or time instants is limited. Conversely, if this number is large, the LMC becomes cumbersome as it requires the computation and modelling of many direct and cross covariance functions.

Now we have presented the different tools for the description and fitting of the space-time variability of soil salinity, we can examine some geostatistical techniques which use these tools in order to interpolate data in space and time.

4.4. Geostatistical interpolation: Space-time kriging

4.4.1. Introduction

When a theoretical model is fitted to the experimental space-time covariance function or variogram, it becomes possible to tackle the problem of predicting attributes at unsampled space locations and/or time instants.

Consider the problem of predicting the value of a continuous attribute x at any unsampled space location s_0 and time instant t_0, $x(s_0, t_0)$, using only the x-data available over the space region S and time domain T, say, $n = \sum_{j=1}^{n_t} n_j$ data $\left\{ x(s_{ij}, t_j), j = 1,...,n_t; i = 1,...,n_j \right\}$.

Kriging is a family of generalized least squares regression algorithms that allows to predict $x(s_0, t_0)$. The latter is considered, in geostatistics, as a realization of the STRF $X(s_0, t_0)$. The problem of predicting $x(s_0, t_0)$ is defined as (Rouhani and Myers, 1990):

$$\hat{X}(s_0, t_0) = \sum_j \sum_i \lambda_{ij} X(s_{ij}, t_j) \tag{86}$$

with λ_{ij}, the kriging weights for the time instant t_j and the space location s_{ij}.

This is the general form and there are some simplifications which are related to the kind of STRF considered. We present first the simple cases where we consider only SRF or TRF, and then we present the more general case for the single STRF.

4.4.2. Spatial/temporal kriging

In case of spatial (temporal) random fields (section 4.3.5) we saw that the dependence in data is modelled via a spatial (temporal) covariance function or variogram. For example, in the case of an SRF, we are dealing in the space domain and the different kriging algorithms defined for SRF can be applied.

Let $X(s, t_0)$ denotes the SRF at the time instant t_0. The predictor $X(s_0, t_0)$, at the space location s_0, is a linear combination of the n_s observations $x_1, \ldots, x_{ns}$ available at the time instant t_0:

$$\hat{X}(s_0, t_0) = \sum_{i=1}^{n_s} \lambda_i X(s_i, t_0) \tag{87}$$

The n_s weights are calculated such that $X(s_0, t_0)$ is unbiased and that the variance of the prediction error is minimal. A detailed description of the computation procedure can be found in geostatistical handbooks (Journel and Huibregts, 1978; Cressie, 1993; Goovaerts, 1997).

Three kriging algorithms can be distinguished (Goovaerts, 1997):

- Simple kriging: the mean is known and constant through the space region;
- Ordinary kriging: the mean is unknown but constant in the subdomains of the space region;
- Kriging with trend model or universal kriging: the mean is unknown, is local and varies smoothly in the subdomains of the space region.

In a similar way, the predictor $X(s_0, t_0)$, at the time instant t_0, is a linear combination of the n_t observations $x_1, \ldots, x_{nt}$ available at the space location s_0.

The prediction is done for each time instant (space location) separately and independently from the other time instants (space locations).

Many applications can be found in the literature: ordinary kriging using variograms converted from a basic pooled spatial variogram on the sampling time instants (Sterk and Stein, 1997; Stein et al., 1997; Stein and Sterk, 1999); ordinary kriging and universal kriging (Ettema et al., 1998; Figueira et al., 1999, 2001); and ordinary log normal kriging (Ettema et al., 2000). The kriging in the space domain will be used in chapter 7 and compared to space-time kriging (see section 4.4.4.2).

4.4.3. Spatial/temporal cokriging

For the same cases as before (spatial/temporal random fields) more accurate estimates, based on cokriging, can be obtained by using the information available at a precedent time instant as well as the contemporaneous data for spatial prediction (D'Agostino et al., 1998) or information from two different TRFs observed at two different space locations. Cokriging can also be used to estimate the temporal change for a spatially correlated soil property between two time instants using the pseudo cross variogram (Papritz and Fluhler, 1994; Lark, 2002).

Under the separability hypothesis and the LMC, Bogaert (1996b) showed that the simple cokriging system between n_t coregionalized space variables (one for each time instant) is equivalent to the simple space-time kriging. This equivalence was shown for a more general form of the covariance functions (Kyriakidis and Journel, 1999). Bogaert (1996b) demonstrated also that space ordinary cokriging system with one unbiasedness condition is equivalent to the space-time ordinary kriging system when it is expressed in terms of covariance functions. Pseudo cross variograms should replace the cross variograms if the cokriging system has to be written in terms of variograms. However, the use of the unique unbiasedness constraint is restricted to the second order stationary situations. The author noted also that space-time ordinary kriging is preferable to space ordinary cokriging whenever it is possible because, although the differences in the prediction variances are negligible when data are abundant, it is not the case when some time instants involve limited data.

We recall that for SRF, the prediction can be done at unsampled locations only for one or more of the observed time instants. As examples of applications, we cite Bogaert and Christakos (1997) and D'Agostino et al. (1998). In the case of TRF, cokriging allows

forecasting and hindcasting, the latter is useful in the estimation of missing values. Examples of application can be found in Rouhani et al. (1992a, b). Cokriging was used also to estimate a missing observation at a given time instant by considering the observations at this time instant as the variable of interest and the observations at other time instants as covariables (Stein, 1998).

Let a given variable been sampled at n_1 space locations during the first time instant, $X_1(s_i)$, $i = 1, \ldots, n_1$, and at n_2 (less than n_1) space locations later during the second time instant, $X_2(s_j)$, $j = 1, \ldots, n_2$. The kriging estimator of X_2 at an unsampled location s_0 is (Goovaerts, 1997):

$$\hat{X}_2(s_0) = \sum_{i=1}^{n_1} \lambda_{1i} X_1(s_i) + \sum_{j=1}^{n_2} \lambda_{2j} X_2(s_j) \tag{88}$$

with λ_{1i} and λ_{2j} are the kriging weights relative to the variables X_1 and X_2, respectively.

Under the hypothesis of stationarity, the weights must satisfy the following conditions:

$$\sum_{i=1}^{n_1} \lambda_{1i} = 0 \text{ and } \sum_{j=1}^{n_2} \lambda_{2j} = 1 \tag{89}$$

Fore cases where the primary and secondary variables relate to the same attribute, the unknown mean can be filtered out and the single unbiasedness constraint

$$\sum_{i=1}^{n_1} \lambda_{1i} + \sum_{j=1}^{n_2} \lambda_{2j} = 1 \tag{90}$$

replaces the above two unbiasedness conditions.

4.4.4. Space-time kriging

For the more general single STRF, two kriging algorithms can be distinguished, i.e. two step space-time kriging and anisotropic space-time kriging. They are presented in the next sections.

4.4.4.1. Two step space-time kriging

The prediction of the value of $X_0 = X(s_0, t_0)$ is a linear combination of n_t predictors in space. In the first step, at each of the n_t time instants $t_j, j = 1, \ldots, n_t$, the predictor of the value of $X(s_0, t_j)$ is determined as (Stein et al., 1994):

$$\hat{X}(\boldsymbol{s}_0, t_j) = \sum_{i=1}^{n_j} \lambda_{ij} X(\boldsymbol{s}_{ij}, t_j) \tag{91}$$

The kriging weights are determined in the standard manner for kriging in the space domain (Goovaerts, 1997).

In the second step $X(\boldsymbol{s}_0, t_0)$ is predicted, using a linear combination of the predictors determined at the first step. This predictor is (Stein et al., 1994):

$$\hat{X}(\boldsymbol{s}_0, t_0) = \sum_{j=1}^{n_t} \theta_j \hat{X}(\boldsymbol{s}_0, t_j) = \sum_{j=1}^{n_t} \sum_{i=1}^{n_j} \theta_j \lambda_{ij} X(\boldsymbol{s}_{ij}, t_j) \tag{92}$$

with θ_j are the kriging weights from the second step.

The predictor $\hat{X}(\boldsymbol{s}_0, t_0)$ explicitly uses spatial variograms that change over time. This algorithm was applied in different situations (Stein et al., 1994, 1997, 1998; Stein, 1998).

4.4.4.2. Anisotropic space-time kriging

For this algorithm, the prediction is done in one and unique step capitalizing on a space-time variogram model which incorporates the spatial and temporal dependencies with a space-time anisotropy ratio (Rouhani and Myers, 1990; Stein et al., 1994, 1997, 1998; Bechini et al., 2000). The advantage of the anisotropic over the two step space-time kriging is that the former allows predictions to be made at every point in space and time. Also it avoids the uncertainties related to the predictors determined at the first step for the latter. This algorithm was used in our application (chapter 7 and 8) and was compared to kriging limited to the spatial domain (chapter 7) and to BME (chapter 8).

4.4.5. Space-time cokriging

De Iaco et al. (2004) extended the LMC to the space-time domain using generalized product sum models for the basic variograms. In addition they proposed an extension of ordinary cokriging to the space-time domain. Although De Iaco et al. (2004) discussed the more general case of several secondary variables, we present here only the case of one variable of interest and one covariable.

A linear space-time predictor of $X(\boldsymbol{s}, t)$ at an unsampled point $(\boldsymbol{s}_0, t_0)$ in the space-time domain is:

$$\hat{X}(s_0, t_0) = \sum_{i=1}^{n} \Lambda_i(s_0, t_0) X(s_i, t_i) \tag{93}$$

where $\Lambda_i(s_0, t_0)$ are the (2x2) matrices of kriging weights whose elements $\lambda_i^{\alpha\beta}(s, t)$ are the weights assigned to the value of the βth variable, $\beta = 1, 2$, at the ith sampled point to predict the αth variable, $\alpha = 1, 2$, at the unsampled point (s_0, t_0).

The predicted value includes two components:

$$\hat{X}(s_0, t_0) = \left[\hat{X}_1(s_0, t_0), \hat{X}_2(s_0, t_0) \right] \tag{94}$$

such that

$$\hat{X}_\alpha(s_0, t_0) = \sum_{i=1}^{n} \sum_{\beta=1}^{2} \lambda_i^{\alpha\beta}(s_0, t_0) X_\beta(s_i, t_i), \, \alpha = 1, 2 \tag{95}$$

This corresponds to an extension of ordinary cokriging to the space-time domain. More detail on how to determine the kriging weights $\lambda_i^{\alpha\beta}(s, t)$ can be found in De Iaco et al. (2004). In their work, they compared space-time ordinary kriging to space-time ordinary cokriging. They found that the correlation coefficients between observed and predicted values were higher when the latter algorithm was used. Also the cokriging error variances were lower than those from kriging.

4.4.6. Other forms of space-time kriging

Additional forms of space-time kriging exist in the literature but will not be used in the application to our data sets. Among these variants, indicator kriging (Bilonick, 1988), kriging with external drift (Snepvangers et al., 2003), and factorial kriging (Rouhani et al., 1992a, b; Van Meirvenne and Goovaerts, 2002; De Iaco et al., 2003) were used in the literature.

Now the classical geostatistical techniques of space-time interpolation have been laid out, we present a versatile and more general method of interpolation, i.e., BME.

4.5. Bayesian Maximum Entropy

4.5.1. Introduction

For most of the classical geostatistical methods of interpolation, i.e. kriging, the prediction is based solely on the hard data. However, other data sources can be useful, mostly, for space

locations and time instants where hard data are missing and they can improve the accuracy of the predictions. Such data can be different types of soft data (interval, probabilistic, etc.), physical laws, moments of higher order, etc.

Bayesian Maximum Entropy (BME) (Christakos, 1990, 2000; Christakos and Li, 1998) is a recent approach developed for the spatio-temporal mapping of natural processes using uncertain information. It offers the flexibility to incorporate various sources of physical knowledge. This incorporation of physical knowledge bases enables global prediction features and the adoption of probability distributions without a need for any assumptions like for example to be gaussian. The BME was also extended to handle vector and functional variables and allows multipoint mapping (Choi et al., 1998; Christakos, 1998; Christakos and Hristopoulos, 1998; Christakos et al., 1999).

4.5.2. Knowledge bases

Christakos (2000) defined a knowledge base as 'a collection of knowledge sources relevant to the problem at hand to be involved by a reasoning process aiming at the solution of the problem'. In the BME framework, the total knowledge (K) available regarding a natural process is considered to be formed from two main bases: the general knowledge (G) and the specificatory knowledge (S) such that $K = G \cup S$.

4.5.2.1. General knowledge base

The general knowledge represents the knowledge that one has about the distribution of the natural variable to be mapped before any specific data. It encompasses physical laws, statistical moments of any order (including the mean and variogram or covariance function), multipoint statistics, etc. It is said to be general because it does not depend on the specific random field realisation at hand.

The objective in the space-time mapping is to predict the values of an STRF, $X(\boldsymbol{p})$, at a point in space and time, $\boldsymbol{p}_k$, given data at space-time points, $\boldsymbol{p}_i$, $i=1$, …, m. Let $\boldsymbol{\chi}_{\mathrm{map}} = [\chi_1,...,\chi_m,\chi_k]'$ be a realisation of the STRF at the points $\boldsymbol{p}_{\mathrm{map}} = [\boldsymbol{p}_1,...,\boldsymbol{p}_m,\boldsymbol{p}_k]'$. The joint cdf of the m observed and the prediction points is defined by:

$$F_x(\chi_1,...,\chi_m,\chi_k) = \mathrm{Pr}\,ob[x_1 \le \chi_1,...,x_m \le \chi_m, x_k \le \chi_k) \tag{96}$$

and its corresponding pdf by:

$$f_x(\chi_1,...,\chi_m,\chi_k) = \partial^{m+1} F_x(\chi_1,...,\chi_m,\chi_k)/\partial\chi_1...\partial\chi_m\partial\chi_k \tag{97}$$

This pdf forms the prior pdf, which can be derived by an estimation process that considers physical constraints provided by prior information or knowledge (*G*). These physical constraints are given by:

$$E[g_\alpha] = \int g_\alpha(\boldsymbol{\chi}_{\text{map}}) f_G(\boldsymbol{\chi}_{\text{map}}) d\boldsymbol{\chi}_{\text{map}}, \qquad \alpha = 0,...,N_c \tag{98}$$

where $f_G(\boldsymbol{\chi}_{\text{map}})$ has the same definition as Eq.(97), except that *x* is replaced by *G*. It is, in fact, the unknown multivariate pdf associated with the general knowledge. The functions g_α are chosen such that the general knowledge base, *G*, is taken account of in full in the prediction process, and their expectations, $E[g_\alpha]$, provide the space-time statistical moments of interest (means, variances, covariances, etc.).

For example, $\alpha = 0$ defines a normalization constraint such that

$$\int f_G(\boldsymbol{\chi}_{\text{map}}) d\boldsymbol{\chi}_{\text{map}} = 1 \tag{99}$$

This leads to $g_0(\boldsymbol{\chi}_{\text{map}}) = 1$ and $E[g_0] = 1$

To take into account the means $m_i = E[X(\boldsymbol{p}_i)] = E[x_i]$, at points $\boldsymbol{p}_i$ (*i=1, ..., m, k*), we let

$$g_\alpha(\chi_i) = \chi_i \text{ and } E[g_\alpha] = m_i, \text{ for } \alpha = 1, ..., m+1. \tag{100}$$

The covariance function

$$c_{ij} = E\left\{[X(\boldsymbol{p}_i) - m_i][X(\boldsymbol{p}_j) - m_j]\right\} \tag{101}$$

between two points $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$, is considered by letting

$$g_\alpha(\chi_i, \chi_j) = [\chi_i - m_i][\chi_j - m_j] \tag{102}$$

and

$$E[g_\alpha] = E\left\{[x_i - m_i][x_j - m_j]\right\} \tag{103}$$

with *i, j* = 1, ..., *m, k* and α = *m+2*, ..., *(m+1)(m+4)/2*.

It is important to note that BME allows also the use of higher moments (Christakos et al., 2002, page 37) by letting $g_\alpha(\chi_i) = \chi_i^q$, *q* is the order of the statistical moment of interest. In addition the prior pdf can be derived from physical laws, for example from stochastic differential equations (Christakos et al., 1999; Serre et al., 1998; Kolovos et al, 2000).

4.5.2.2. Specificatory knowledge base

The specificatory knowledge includes the data specific to a given experiment or situation. It refers to a particular occurrence of the natural variable at a particular space location and a particular time instant. It is divided into two main categories depending on the accuracy of the data.

The hard data $\chi_{hard} = [\chi_1, ..., \chi_h]'$ (104)

are exact measurements of the natural process and considered to be error-free. It encompasses accurate measurements obtained from real-time observation devices, computational algorithms, simulation processes, etc. Their accuracy means that the probability that the vector x_{hard} takes the values χ_{hard} is one:

$$\Pr ob[x_{hard} = \chi_{hard}] = 1.$$ (105)

The soft data $\chi_{soft} = [\chi_{h+1}, ..., \chi_m]'$ (106)

involve uncertain observations, empirical charts, assessments by experts, etc. They are incomplete or qualitative data linked to opinions, intuition, etc. Different kinds of soft data are available, often in the form of interval domain data or of a probabilistic nature. The interval soft data is given by:

$$\chi_{soft} = \left\{ [\chi_{h+1}, ..., \chi_m]' : \chi_i \in I_i = [l_i, u_i], i = h+1, ..., m \right\}$$ (107)

with l_i and u_i representing the lower and upper limits of the interval I_i, respectively. This means that the unknown exact values χ_i, $i = h+1, ..., m$, lie within the known intervals I_i is equal one:

$$\Pr ob[\chi_i \in I_i = (l_i, u_i)] = 1$$ (108)

The probabilistic soft data is given by:

$$\chi_{soft} : P_S(x_{soft} \leq \xi) = \int_{-\infty}^{\xi} f_S(\chi_{soft}) d\chi_{soft}$$ (109)

with $f_S(\chi_{soft})$, the specificatory pdf.

If the specificatory pdf follows a uniform distribution, the probabilistic soft data are, in fact, equivalent to interval soft data.

The total data available for mapping is then the union of both hard and soft data:

$$\chi_{data} = [\chi_{hard}, \chi_{soft}]'$$ (110)

To illustrate these notions of hard and soft data, we give here an example that is used in the application part of this dissertation (see chapter 8). As reported in chapter 2, soil salinity can be assessed by measuring the electrical conductivity of the soil. It is measured either in the laboratory ($EC_{2.5}$) or in the field (EC_a). The measurement in the laboratory is done with high precision in standard conditions and is directly related to soil salinity. Thus it is considered as an accurate measurement of the soil salinity and represents our hard data. The apparent electrical conductivity measured in the field (EC_a) is only an indirect measurement of soil salinity and still needs to be calibrated with $EC_{2.5}$ values. Thus, based on a calibration model (frequently a regression model), EC_a values are converted into 'predicted' $EC_{2.5}$ values. As these values are only predictions from a stochastic model, they are entailed with some uncertainty, and are consequently considered as soft data. The predicted $EC_{2.5}$ values and their corresponding standard deviations allow us to define the two kinds of soft data: the confidence intervals are built which give the interval soft data while the two statistic parameters are used to describe Gaussian pdfs which represent our probabilistic soft data.

4.5.3. The three steps of BME analysis

The BME has a double goal: informativeness (prior information maximization given the general knowledge) and cogency (posterior probability maximization given specificatory knowledge).

The BME analysis is done in three main steps of knowledge acquisition, integration and processing: the structural or prior step, the meta-prior step, and the integration or posterior step. At the end of these three steps, we get the posterior pdf:

$$f_K(\chi_k) = A^{-1} \Upsilon_S \left[(\chi_{soft}), \Upsilon_G (\chi_{map}) \right] \tag{111}$$

where Υ_G represents the operator processing the general knowledge, G, Υ_S the operator processing the specificatory kwonledge, S, and A is a normalization constant. The two operators, Υ_G and Υ_S are considered to be the two legs on which the BME equations stand (Christakos, 2000). This is the general form of the posterior pdf, and it will takes different forms depending on the nature of available general and spacificatory knowledge bases. We present some specific cases in which follows.

4.5.3.1. The structural or prior step

The goal is the maximization of the information content considering only the general knowledge before any use of the data, which corresponds to the first goal of the BME. This step assumes an inverse relation between information and probability: the more informative an evaluation of a mapping situation is, the less probable it is to occur. This means that if a theory is general and vague, it includes more alternatives so it is more probable however it is less informative. This inversion relation is expressed in the equation below.

A random field is completely defined by its multivariate probability distribution function (pdf), which forms the prior pdf. The latter should be derived by means of an estimation process that takes into consideration physical constraints under the form of prior information or knowledge. This information is measured, in the context of BME, using the Shannon's entropy function (Shannon, 1948), thus the E in BME, which expresses the given information in the random vector x_{map} as:

$$\text{Info}(x_{map}) = -\ln\left[f_G(\chi_{map})\right] \tag{112}$$

This equation represents the uncertainty (in the form of the pdf $f_G(\chi_{map})$) regarding the random vector x_{map}: the higher the probability, the lower the uncertainty about x_{map} and the lesser the amount of information provided by the pdf about x_{map}.

The expected information is given by the following entropy function:

$$\text{E}\left[\text{Info}(x_{map})\right] = -\int \ln\left[f_G(\chi_{map})\right]f_G(\chi_{map})d\chi_{map} \tag{113}$$

This function needs to be maximized subject to the physical constraints provided by prior information or knowledge G (equation 98), which justifies the M in the BME acronym. This maximization requires the use of the Lagrange multipliers. For the physical constraints, g_α, provided by equation 98, the operator processing the general knowledge, G, is given by:

$$\Upsilon_G(\chi_{map}) = \sum_{\alpha=1}^{N_c} \mu_\alpha g_\alpha(\chi_{map}) \tag{114}$$

where μ_α are Lagrange multipliers calculated by substituting equation (98) into equation (114).

At this stage the prior or G-based multivariate pdf is:

$$f_G(\chi_{map}) = Z^{-1}\exp\left[\Upsilon_G(\chi_{map})\right] \tag{115}$$

The entropy function, equation (113) modified to include the Lagrange multipliers, is as follows:

$$E_{modif}\left[Info(\boldsymbol{x}_{map})\right] = -\int \ln\left[f_G(\boldsymbol{\chi}_{map})\right]f_G(\boldsymbol{\chi}_{map})d\boldsymbol{\chi}_{map} + \sum_{\alpha=0}^{N_c}\mu_\alpha E\left[g_\alpha(\boldsymbol{\chi}_{map})\right] \quad (116)$$

This function, when maximized, gives the prior pdf:

$$f_G(\boldsymbol{\chi}_{map}) = Z^{-1}\exp\left[\sum_{\alpha=1}^{N_c}\mu_\alpha g_\alpha(\boldsymbol{\chi}_{map})\right] \quad (117)$$

where $Z = \int\exp\left[\sum_{\alpha=1}^{N_c}\mu_\alpha g_\alpha(\boldsymbol{\chi}_{map})\right]d\boldsymbol{\chi}_{map} = \exp(-\mu_0)$ \quad (118)

is a normalization constant and μ_0 is the first Lagrange multiplier.

Substituting the expressions (100) and (102) for the functions $g_\alpha(\boldsymbol{\chi}_{map})$ into equation (117) gives the following prior pdf:

$$f_G(\boldsymbol{\chi}_{map}) = Z^{-1}\exp\left[\sum_{i=1}^{m,k}\mu_i\chi_i + \sum_{i,j=1}^{m,k}\mu_{ij}(\chi_i - m_i)(\chi_j - m_j)\right] \quad (119)$$

The constraints, means and covariances, are written:

$$m_i = \int \chi_i f_G(\boldsymbol{\chi}_{map})d\boldsymbol{\chi}_{map} \qquad \text{for } i = 1, 2, \ldots, m, k \quad (120)$$

$$c_{ij} = \int(\chi_i - m_i)(\chi_j - m_j)f_G(\boldsymbol{\chi}_{map})d\boldsymbol{\chi}_{map} \qquad \text{for } i,j = 1, 2, \ldots, m, k \quad (121)$$

As these constraints are known (mostly estimated from data), the above constraints, equations (120) and (121), are solved for the Lagrange multipliers μ_i and μ_{ij}. Solving for these multipliers results in the following prior pdf, which is multivariate Gaussian (Serre, 1999):

$$f_G(\boldsymbol{\chi}_{map}) = \frac{1}{(2\pi)^{\frac{m+1}{2}}\sqrt{|\boldsymbol{C}_{map}|}}\exp\left[-\frac{1}{2}(\boldsymbol{\chi}_{map} - \boldsymbol{m}_{map})'\boldsymbol{C}_{map}^{-1}(\boldsymbol{\chi}_{map} - \boldsymbol{m}_{map})\right] \quad (122)$$

where $\boldsymbol{m}_{map}$ represents the vector of means and $\boldsymbol{C}_{map}$ is the variance-covariance matrix.

Expression (A7) can be rewritten:

$$f_G(\boldsymbol{\chi}_{map}) = \phi(\boldsymbol{\chi}_{map}; \boldsymbol{m}_{map}, \boldsymbol{C}_{map}) \quad (123)$$

where $\phi(\boldsymbol{\chi}_{map}; \boldsymbol{m}_{map}, \boldsymbol{C}_{map})$ is the multivariate Gaussian pdf fully characterized by its vector of means $\boldsymbol{m}_{map}$ and its variance-covariance matrix $\boldsymbol{C}_{map}$.

4.5.3.2. The meta-prior step

During this step, the specificatory knowledge is collected and organized in appropriate quantitative forms that can be easily incorporated in the BME framework. The available data can be divided into two main types: the hard data and the soft data. The hard data are considered exact and accurate measurements of the natural process while the soft data are indirect and inaccurate measures of the variable of interest. The soft data may be in the form of intervals, probability distribution functions, etc (see section 4.5.2.2).

4.5.3.3. The integration or posterior step

In this step the two knowledge bases (G and S) are integrated. The goal is the maximization of the posterior pdf given the total knowledge K, which is the second goal of BME. The G-based pdf is updated, by considering the available site-specific knowledge (the data). This updating is performed using a Bayesian conditionalization, thus the B in BME:

$$f_K(\chi_k|\chi_{\text{data}}) = f_G(\chi_{\text{map}})/f(\chi_{\text{data}})$$ (124)

where $f_K(\chi_k|\chi_{\text{data}})$ and $f_G(\chi_{\text{map}})$ are the posterior and the prior pdfs, respectively. The posterior pdf should be maximized with respect to χ_k. This stage yields the K-based pdf, $f_K(\chi_k)$.

In the case of interval soft data, equation (107), the operator processing the specificatory knowledge is:

$$\Upsilon_S(\chi_{\text{map}}) = Z^{-1}\int_l^u \exp\left[\Upsilon_G(\chi_{\text{map}})\right]d\chi_{\text{soft}}$$ (125)

with a multiple integration over a hypercube bounded by the lower (l) and the upper (u) limits.

Writing equation (125) in terms of the prior pdf, equation (115), and combining with equation (111) leads to the following posterior pdf:

$$f_K(\chi_k) = \frac{\int_l^u f_G(\chi_{\text{map}})d\chi_{\text{soft}}}{\int_l^u f_G(\chi_{\text{data}})d\chi_{\text{soft}}}$$ (126)

where $f_G(\boldsymbol{\chi}_{\mathrm{data}}) = \int f_G(\boldsymbol{\chi}_{\mathrm{map}}) d\chi_k$ (127)

Equation (126) gives a general form for specificatory knowledge involving hard and interval soft data. This needs to be combined with the prior pdf, $f_G(\boldsymbol{\chi}_{\mathrm{map}})$, leading to the posterior BME pdf $f_K(\chi_k)$.

For probabilistic soft data, the operator processing specificatory knowledge is given by:

$$\Upsilon_S(\boldsymbol{\chi}_{\mathrm{map}}) = Z^{-1} \int f_S(\boldsymbol{\chi}_{\mathrm{soft}}) \exp\left[\Upsilon_G(\boldsymbol{\chi}_{\mathrm{map}})\right] d\chi_{\mathrm{soft}}$$ (128)

Writing equation (128) in terms of the prior pdf, equation (115) and combining with equation (111) leads to the following posterior pdf:

$$f_K(\chi_k) = \frac{\int f_S(\boldsymbol{\chi}_{\mathrm{soft}}) f_G(\boldsymbol{\chi}_{\mathrm{map}}) d\chi_{\mathrm{soft}}}{\int f_S(\boldsymbol{\chi}_{\mathrm{soft}}) f_G(\boldsymbol{\chi}_{\mathrm{data}}) d\chi_{\mathrm{soft}}}$$ (129)

Again, equation (129) gives a general form for specificatory knowledge involving hard and probabilistic soft data. This needs to be combined with the prior pdf, $f_G(\boldsymbol{\chi}_{\mathrm{map}})$, leading to the posterior BME pdf $f_K(\chi_k)$.

As hard and both interval and probabilistic soft data are available in our case study, equations (126) and (129) will be used in chapter 8 in conjunction with equation (122). The latter defines the prior pdf when the general knowledge is limited to the statistical moments of order one and two. This is, in fact, our situation. The results of BME will be compared to those from space-time kriging.

4.5.4. Some BME estimators

The posterior pdf, which is not limited to the Gaussian type, describes fully the random field at the estimation point. It provides a complete picture of the mapping situation as well as different estimators and their associated estimation uncertainty. The prediction points lie, in general, on a regular grid and the predictions are used to create space-time maps. The different estimates presented in the following sections are specific to a general knowledge involving the first two statistical moments and a specificatory knowledge encompassing hard data as well as interval and probabilistic soft data. These two kinds of knowledge bases are, in fact, available for our application in chapter 8.

4.5.4.1. The mode estimate

The mode, $\tilde{\chi}_k$, represents the most probable situation. It corresponds to the maximization of the posterior pdf, $f_K(\chi_k)$, with respect to χ_k:

$$\left[\frac{\partial f_K(\chi_k)}{\partial \chi_k}\right]_{\chi_k=\tilde{\chi}_k} = 0 \tag{130}$$

4.5.4.2. The conditional mean estimate

The conditional mean estimate, which is in general a nonlinear function of the data, is suitable for mapping situations where one is interested in minimizing the mean square estimation error:

$$\bar{\chi}_k = \int f_K(\chi_k)\chi_k d\chi_k \tag{131}$$

In the case of interval data, it has the following expression (Serre, 1999):

$$\bar{\chi}_k = \frac{\int \chi_k d\chi_k \int_l^u \phi(\boldsymbol{\chi}_{\mathrm{map}}; \boldsymbol{m}_{\mathrm{map}}, \boldsymbol{C}_{\mathrm{map}}) d\boldsymbol{\chi}_{\mathrm{soft}}}{\int_l^u \phi(\boldsymbol{\chi}_{\mathrm{data}}; \boldsymbol{m}_{\mathrm{data}}, \boldsymbol{C}_{\mathrm{hs,hs}}) d\boldsymbol{\chi}_{\mathrm{soft}}} \tag{132}$$

where $\boldsymbol{m}_{\mathrm{data}}$ is defined as $\boldsymbol{m}_{\mathrm{map}}$, except that the estimation point is excluded and $\boldsymbol{C}_{\mathrm{hs,hs}}$ is a partition of $\boldsymbol{C}_{\mathrm{map}}$ relative to hard and interval data such that :

$$\boldsymbol{C}_{\mathrm{map}} = \begin{bmatrix} \boldsymbol{C}_{\mathrm{hs,hs}} & \boldsymbol{C}_{\mathrm{hs,k}} \\ \boldsymbol{C}_{\mathrm{k,hs}} & \boldsymbol{C}_{\mathrm{k,k}} \end{bmatrix} \tag{133}$$

For probabilistic soft data, it is given by (Serre, 1999):

$$\bar{\chi}_k = \frac{\int \chi_k d\chi_k \phi \int (\boldsymbol{\chi}_{\mathrm{map}}; \boldsymbol{m}_{\mathrm{map}}, \boldsymbol{C}_{\mathrm{map}}) f_S(\boldsymbol{\chi}_{\mathrm{soft}}) d\boldsymbol{\chi}_{\mathrm{soft}}}{\int \phi(\boldsymbol{\chi}_{\mathrm{data}}; \boldsymbol{m}_{\mathrm{data}}, \boldsymbol{C}_{\mathrm{hs,hs}}) f_S(\boldsymbol{\chi}_{\mathrm{soft}}) d\boldsymbol{\chi}_{\mathrm{soft}}} \tag{134}$$

4.5.4.3. The uncertainty estimates

A measure of the uncertainty associated with the estimated values is provided by the variance of the estimation error:

$$\hat{\sigma}_k^2 = \int (\chi_k - \bar{\chi}_k)^2 f_K(\chi_k) d\chi_k \tag{135}$$

For interval soft data, it takes the following expression (Serre, 1999):

$$\hat{\sigma}_k^2 = \frac{\int (\chi_k - \bar{\chi}_k)^2 d\chi_k \int_l^u \phi(\chi_{\text{map}}; \boldsymbol{m}_{\text{map}}, \boldsymbol{C}_{\text{map}}) d\chi_{\text{soft}}}{\int_l^u \phi(\chi_{\text{data}}; \boldsymbol{m}_{\text{data}}, \boldsymbol{C}_{\text{hs,hs}}) d\chi_{\text{soft}}} \tag{136}$$

In the case of probabilistic soft data, it is given by (Serre, 1999):

$$\hat{\sigma}_k^2 = \frac{\int (\chi_k - \bar{\chi}_k)^2 d\chi_k \phi \int (\chi_{\text{map}}; \boldsymbol{m}_{\text{map}}, \boldsymbol{C}_{\text{map}}) f_S(\chi_{\text{soft}}) d\chi_{\text{soft}}}{\int \phi(\chi_{\text{data}}; \boldsymbol{m}_{\text{data}}, \boldsymbol{C}_{\text{hs,hs}}) f_S(\chi_{\text{soft}}) d\chi_{\text{soft}}}. \tag{137}$$

This variance is data-dependent, whereas in kriging it is data-free. Using this uncertainty estimate and the conditional mean, one can compute the confidence intervals assuming a Gaussian distribution. For example, for an error of type I of 5%, the confidence limits are:

$$[\bar{\chi}_k - 1.96\hat{\sigma}_k, \bar{\chi}_k - 1.96\hat{\sigma}_k] \tag{138}$$

In addition, from the posterior pdf, one can compute directly the confidence intervals (Serre and Christakos, 1999a), which provide a more realistic assessment of the estimation error than the error variance or the confidence limits derived from equation (138).

4.5.5. Kriging as a special case of BME

When the general knowledge is limited to the mean and covariance functions and the specificatory knowledge is restricted to the only hard data, the BME posterior pdf is Gaussian (the mean and the mode are equal) and the BME mean estimate is equivalent to the kriging (best minimum mean squared error, MMSE) estimate, which is the conditional mean:

$$\hat{\chi}_{k,MMSE} = E[X(\boldsymbol{p}_k)|\chi_{\text{hard}}] \tag{139}$$

When the space-time random field is Gaussian, eq.(139) becomes linear and optimal among all MMSE estimators, and expressed as:

$$\hat{\chi}_{k,MMSE} = \lambda' \chi_{hard} \tag{140}$$

where λ is a vector of weights associated with the data points and involving the space-time mean and covariance functions. Thus BME is a more general interpolation approach and kriging is a special case in limiting situations.

4.5.6. Examples of BME application

The BME approach has been applied in different research fields. Examples of such applications can be found in agricultural sciences (D'Or et al., 2001; Bogaert and D'Or, 2002; D'Or and Bogaert, 2003), environmental sciences (Christakos, 1998; Serre and Christakos, 1999b; Christakos and Serre, 2000; Serre et al., 2001; Choi et al., 2003), etc.

4.6. Conclusions

We presented, in this chapter, different statistical methods suitable for the analysis of space-time data. We focused on three groups of methods. First, we described classical statistical methods which ignore the space-time nature of the data. Then, we reviewed some tools useful for characterizing the space-time variability of data along followed by geostatistical interpolation techniques. Finally, we presented the modern geostatistical method of BME.

From the classical statistical methods, we retained the concepts of temporal stability and dynamic spatial variation which will be applied to our data set in the following chapters. The spatial, temporal and space-time covariance functions were considered in our application for the description and modelling of soil salinity variability. Finally, we choose spatial as well as space-time kriging and BME for the prediction of soil salinity at unsampled space locations and time instants.

PART II

CLASSICAL STATISTICAL APPROACHES

CHAPTER 5

TEMPORAL STABILITY OF SPATIAL PATTERNS OF SOIL SALINITY DETERMINED FROM LABORATORY AND FIELD ELECTRICAL CONDUCTIVITY

Abstract

We elaborated a procedure for the assessment of the temporal stability of soil salinity and the optimization of the sampling effort. Soil electrical conductivity data obtained from field electrode probes and laboratory analysis were compared and analyzed to check the temporal stability of salinity patterns. Therefore sampling of 20 locations at different depths was repeated 19 times over a period from November 1994 to June 2001. Although the use of electrical conductivity probes for the assessment of soil salinity is suggested for irrigated fields we found that it is applicable in semi-humid natural grassland. Both determination methods showed a strong temporal stability. The Spearman rank correlation confirmed the persistence of the ranking of the different locations. Additionally, using the technique of relative differences, we were able to identify three classes: (1) the low saline locations, (2) locations which are representative of the average field soil salinity, and (3) the high saline locations. The latter included the least time stable locations while the first class contained the most time stable ones. The low saline locations were identified as belonging to the zones of waterlogging and/or salt leaching, the high saline locations belonged exclusively to the zone of salt accumulation, while locations representative of the average soil salinity belonged to all three possible zones. We investigated also how precise the selected locations representing the average soil salinity can estimate this average. We found that using only two locations from the 20 available, the average was adequately estimated with a difference smaller than 0.3 dS/m. This representativity was also checked by splitting the measurements into two temporal subsamples. We found for both subsamples that the same locations were representative of the average soil salinity as when all measurement dates were considered.

5.1. Introduction

In Hungary salt-affected soils exceed one million hectares, covering more than 10% of the territory of the country. More than 95% of these soils are located in the Great Hungarian Plain (GHP). Hortobágy National Park, where our study area was located, forms a subregion of this plain. Hortobágy is a recharge area of saline groundwater originating from northern mountains. This groundwater is the main source of salt accumulation in the area, so its dynamics strongly influences soil salinization (Tóth et al., 2002b).

Conventionally soil salinity is determined in the laboratory by measuring the electrical conductivity of a solution extracted from a water-saturated soil paste (EC_e). Alternatively it can be measured in the field, and then it is called apparent electrical conductivity (EC_a), using electrode probes. This approach is easier, less time consuming and cheaper than the laboratory approach.

For an efficient management of salt-affected soils, we need to measure soil salinity; the latter is spatially variable and dynamic. This variability is the outcome of different pedological factors like water table depth, topography, parent material, etc. As a consequence of the spatial and temporal variability, we need measurements from numerous samples from different locations and during different time instants. However as soil salinity does not change noticeably during short time in natural conditions, the observed spatial pattern could be time stable and can persist from one time instant to another. If this is so, the sampling effort could be reduced to a limited number of locations representative of the mean, low and high saline conditions.

The concept of temporal stability, or persistence, has been used almost exclusively in the context of soil water content. Vachaud et al. (1985) were the first to introduce the concept. They analyzed the temporal stability of soil water for three crops (grass, olive trees, and wheat) to check if time-invariant characteristic statistical properties of the probability distribution functions can be assigned to individual locations. Comegna and Basile (1994) analyzed the temporal stability of soil water storage in a cultivated sandy soil in Italy. Grayson and Western (1998) defined the concept of Catchment Average Soil Moisture Monitoring (CASMM) sites using the approach of Vachaud et al. (1985). More recently Gómez-Plaza et al. (2000) studied the temporal stability of the spatial pattern of soil moisture encompassing different scales from a semi-arid region of Spain. Mohanty and Skaggs (2001) analyzed the influence of factors like soil, slope, and vegetation on the temporal stability of soil moisture within three remote sensing footprints. The temporal stability of soil water

content and matric potential were investigated by Van Pelt and Wierenga (2001). The most recent work on time stability of soil moisture was done by Martínez-Fernández and Ceballos (2003).

All the previous studies focused on soil water. There is one exception (Castrignanò et al. 1994). The authors studied the temporal persistence of three indicators of soil salinity: electrical conductivity (EC), sodium content and sodium adsorption ratio (SAR) of the soil saturation extract. They collected 28 soil samples from an agricultural field of 80 x 350 m at three different depths during 8 time instants for a period over two years.

The first objective of our research was to apply the concepts of temporal stability to soil salinity determined by two techniques (a field and laboratory determinations of electrical conductivity), at different depths sampled within a large field and repeated over a long temporal domain. The second objective was to study the relationship between the salt accumulation processes and their temporal stability. We also tried to find simple covariables (like vegetation pattern or elevation) with soil salinity that would allow the identification of optimal sampling locations without an extensive sampling effort.

5.2. Material and methods

5.2.1. Data description

In this chapter, only data from the 'calibration data set' were used (see chapter 3, section 3.2.1). The 20 locations where EC_a was surveyed and $EC_{2.5}$ measured are given in Fig. 5.1. Also, we consider only the EC_a at the depth 0-40 cm and the mean $EC_{2.5}$ calculated over the four depths.
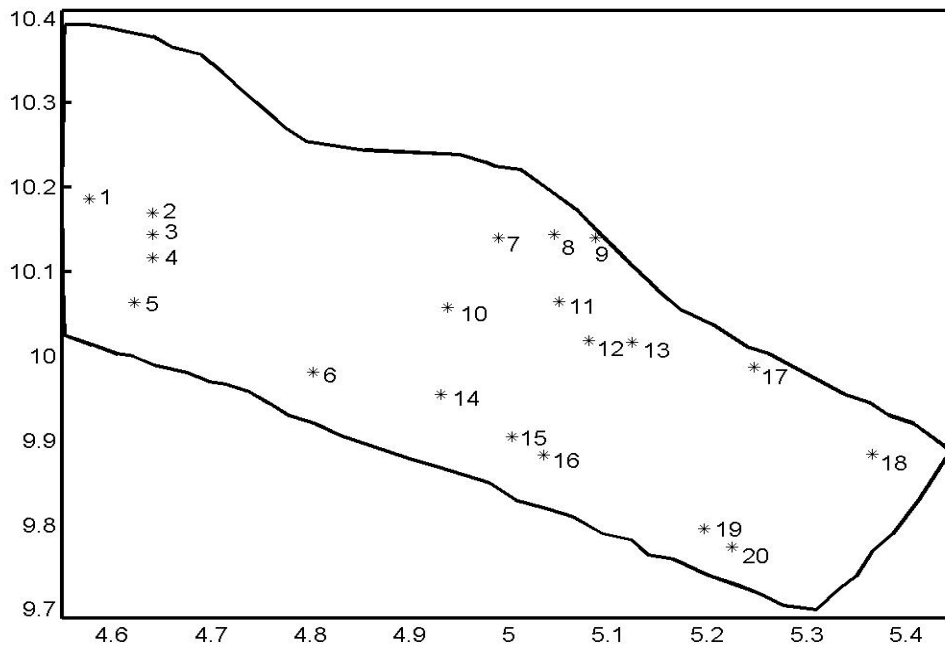
Fig.5.1. Spatial locations (with their number) where EC_a was sensed and soil was sampled for the determination of $EC_{2.5}$

5.2.2. Temporal stability

The two concepts of temporal stability were presented in detail in chapter 4 (sections 4.2.3.1 and 4.2.3.2). This chapter used the first definition of temporal stability based on relative differences (Vachaud et al, 1985) as well as the Spearman rank correlation.

5.3. Results and discussion

5.3.1. Exploratory data analysis

Basic statistics about the apparent (EC_a) and laboratory ($EC_{2.5}$) electrical conductivity, calculated using all locations (20) and all time instants (19) are reported in table 5.1.

Table 5.1. Descriptive statistics for EC_a and $EC_{2.5}$ (in dS/m) measured at different depths. Min: minimum; Max: maximum; Stdev: standard deviation; CV: coefficient of variation.

Property	Depth (cm)	Mean	Median	Min	Max	Stdev	CV %
EC_a	0-20	1.40	0.90	0.02	9.19	1.64	117
	0-40	1.94	1.31	0.03	10.05	1.99	103
$EC_{2.5}$	0-10	1.29	1.02	0.10	13.50	1.29	100
	10-20	1.72	1.37	0.08	8.30	1.41	82
	20-30	2.33	1.92	0.07	13.70	1.77	76
	30-40	2.69	2.28	0.07	11.50	1.84	68
	0-40 (mean)	2.01	1.67	0.10	7.50	1.38	69

Since bulk soil electrical conductivity is affected by spatial variability of texture and soil moisture, EC_a has higher CV's compared to $EC_{2.5}$, and for both soil properties the CV decreased, and the mean value increased, with depth. Regarding the extreme values, there are some non saline locations (minimum of 0.02 to 0.10 dS/m) as well as some highly saline sites (maximum values ranging between 8.3 and 13.7 dS/m).

5.3.2. Temporal stability using the Spearman's rank order correlation

As the Spearman rank order correlation is a non parametric (free distribution) test it is less restrictive than the Pearson linear correlation. It indicates the strength and the direction of a rising or falling relationship between two different variables, or the same variable observed at two different time instants.

Table 5.2 gives the rank correlation coefficients with regard to $EC_{2.5}$ for the 19 time instants. Rank correlation coefficients ranged between 0.46 and 0.96 and were generally greater than 0.85. The table shows that $EC_{2.5}$ presented time stable spatial patterns across the whole study period. This is indicated by the values of order correlation which were highly to very highly significant in most of the cases. For example only one of the 171 coefficients was not significant at 5%, while 153 coefficients were significant at the 0.1%. Also, the loss of information between two measurement times was small. The rank correlations of the other variables (results not shown) were mostly significant, although the number of non significant coefficients was larger for some of the depths.

5.3.3. Temporal stability using the mean relative differences

Using the Pearson linear and Spearman rank correlation coefficients, we were able to show that there was a strong temporal stability of the spatial pattern of the soil salinity across the whole spatial domain and the whole study period. However we still need to quantify this temporal stability by identifying locations which are time stable and simultaneously which are representative of the mean and/or extreme saline conditions. Therefore the concept of relative differences (Vachaud et al., 1985) was used.

The plots of the mean relative differences ranked in ascending order are reported in Fig. 5.2 for EC_a and in Fig. 5.3 for $EC_{2.5}$. The corresponding temporal standard deviations (vertical bars represent +/- one standard deviation) were also drawn to indicate the dispersion around the mean relative differences.

Fig. 5.2 shows that at locations 1, 2, and 12 the mean field EC_a was observed consistently within +/- 0.1 dS/m at any time instant. At the locations 10, 11, 13, 14, and 18 this mean was systematically underestimated by more than 0.5 dS/m. From these locations 11, 13, and 18 were the least saline. On the other hand at locations 8, 15, 16, 19, and 20 the spatial mean EC_a was overestimated by more than 0.5 dS/m, with locations 15, 16, and 19 being the most saline. The temporal stability, indicated by the vertical bars corresponding to ± one temporal standard deviation, is very strong for the locations with low salinity, intermediate for locations representative of the mean field EC_a, and low for locations with high salinity.

The locations representative of the least saline conditions in terms of $EC_{2.5}$ were identified from Fig. 5.3. These are locations 11, 13, and 18. They maintain their temporal stability, although these locations were less time stable as the temporal standard deviations were 0.06, 0.02, and 0.25 dS/m for EC_a, and 0.21, 0.19, and 0.30 dS/m for $EC_{2.5}$. When the locations representative of the most saline conditions, based on $EC_{2.5}$, were selected from Fig. 4.3, the same locations were obtained as for EC_a except one (location 8). Also, as for EC_a, they displayed a weak temporal stability although the standard deviations were smaller (e.g. the temporal standard deviation for location 15 is 0.73 for EC_a, and 0.50 for $EC_{2.5}$). The locations most representative of the average field salinity were 2, 3, 6, and 12.
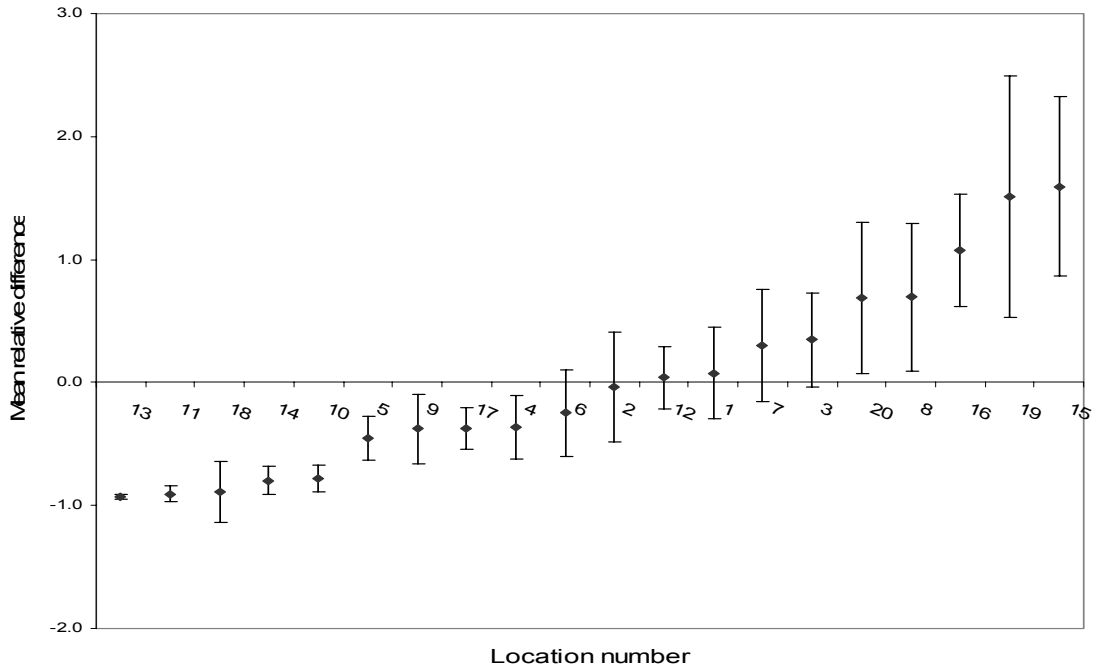
Fig. 5.2. Mean relative differences for EC$_a$. Vertical bars represent +/- one standard deviation.
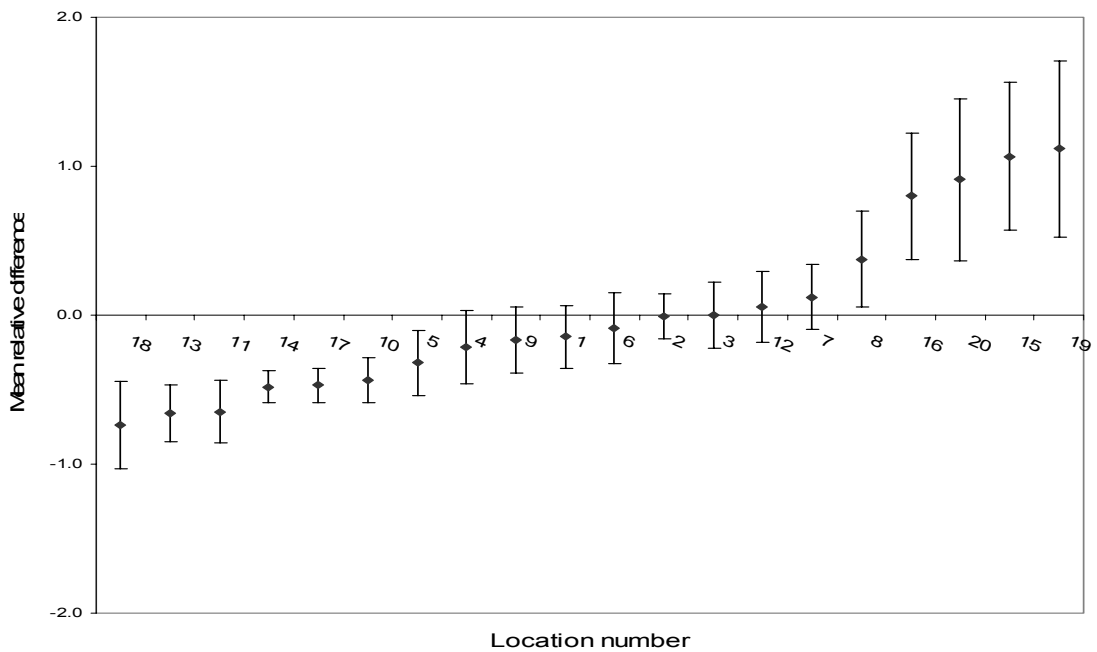
Fig. 5.3. Mean relative differences for EC$_{2.5}$. Vertical bars represent +/- one standard deviation.

Table 5.2. Spearman rank order correlation coefficients for EC$_{2.5}$ at the 19 time instants.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	0.65																	
3	0.77	0.72																
4	0.89	0.71	0.94															
5	0.87	0.72	0.96	0.96														
6	0.69	0.66	0.94	0.92	0.94													
7	0.87	0.73	0.92	0.96	0.94	0.88												
8	0.59	0.71	0.85	0.88	0.85	0.89	0.83											
9	0.82	0.79	0.94	0.96	0.95	0.92	0.92	0.94										
10	0.79	0.71	0.89	0.94	0.92	0.91	0.92	0.89	0.95									
11	0.69	0.72	0.95	0.93	0.94	0.96	0.91	0.93	0.95	0.93								
12	0.76	0.67	0.86	0.84	0.87	0.85	0.80	0.85	0.88	0.89	0.88							
13	0.70	0.54	0.85	0.90	0.87	0.90	0.84	0.86	0.90	0.84	0.86	0.80						
14	0.46	0.74	0.88	0.86	0.85	0.78	0.87	0.86	0.93	0.84	0.84	0.71	0.85					
15	0.85	0.74	0.95	0.94	0.94	0.93	0.92	0.87	0.95	0.95	0.95	0.90	0.87	0.85				
16	0.77	0.73	0.92	0.93	0.93	0.91	0.91	0.89	0.96	0.95	0.94	0.90	0.90	0.89	0.95			
17	0.70	0.78	0.89	0.85	0.88	0.84	0.87	0.86	0.95	0.92	0.90	0.86	0.76	0.88	0.94	0.93		
18	0.91	0.64	0.90	0.94	0.95	0.89	0.92	0.80	0.90	0.90	0.89	0.85	0.86	0.82	0.91	0.91	0.79	
19	0.76	0.66	0.82	0.90	0.88	0.90	0.88	0.84	0.95	0.96	0.85	0.87	0.81	0.83	0.90	0.94	0.88	0.87

1 to 19 refer to: Nov 1994, March, June, Sept, and Dec 1995, March and June 1996, March, June, Sept, and Dec 1997, Sept 1998, Apr, Jul and Sept 1999, Apr and Dec 2000, March and June 2001 respectively.

5.3.4. Classification of the locations as low, average or highly saline

Analyzing corn yield data, Taylor et al. (2000) considered locations with a temporal standard deviation smaller than the mean relative difference to be consistently different from the mean. They further subdivided this group by separating locations for which the mean relative difference is negative from those for which this mean is positive. Also they distinguished a third group, locations similar to the mean, for which the temporal standard deviation is larger than the mean relative difference and overlapped with the mean value (this is illustrated in Fig. 5.2 and 5.3 by the x axis intersecting with the vertical bars). Based on these definitions we classified the locations as low, average or highly saline (table 5.3).

Table 5.3. Location membership to low, average, and high salinity groups.

Property	Low salinity	Average salinity	High salinity
EC_a	4, 5, 9, 10, 11, 13, 14, 17, 18	1, 2, 3, 6, 7, 12	8, 15, 16, 19, 20
$EC_{2.5}$	5, 10, 11, 13, 14, 17, 18	1, 2, 3, 4, 6, 7, 9, 12	8, 15, 16, 19, 20

These groups contain more locations than when we used a restrictive criterion, like +/- 0.1 dS/m for the mean and +/- 0.5 dS/m for the extreme values (section 5.3.3). We included almost the same locations for the three classes. For example the low saline class contained locations 5, 10, 11, 13, 14, 17, and 18; the average salinity class grouped locations 1, 2, 3, 6, 7, and 12 while locations 8, 15, 16, 19, and 20 belonged to the high salinity class.

The average soil salinity for the three classes and the total proportion of locations in each class are shown in table 5.4.

Table 5.4. Mean salinity and proportion of locations for the three salinity classes.

Property	Salinity class	Mean salinity (dS/m)	Proportion of locations (%)			
			Total	Waterlogging	Accumulation	Leaching
EC_a	Low	0.64	45	15	0	30
	Average	2.00	30	5	20	5
	High	4.25	25	0	25	0
$EC_{2.5}$	Low	0.89	35	15	0	20
	Average	1.89	40	5	20	15
	High	3.77	25	0	25	0

The proportion of locations classified as highly saline fluctuates between 15% and 25% while this proportion ranges between 25 and 45% for the low saline class, and 30 and 60% for the average saline class. Consequently most of the 20 locations were classified as average saline, followed by low saline and lastly highly saline.

5.3.5. Salt accumulation processes and temporal stability

The geology, the natural vegetation, and a conceptual model of salt accumulation of the study site were described in Tóth et al. (2002a). In the study area the maximal difference in elevation is 1.76 m. Although this difference is small, it was found that elevation is a major factor in the soil salinization. The other factors are the groundwater depth and its chemical composition, which are related to the elevation. All these three factors contribute to the mosaic distribution of the natural vegetation.

Tóth and Kuti (2002) used a *k*-means clustering procedure (Burrough, 1989) to classify the 413 locations where EC_a was measured into one of the three strata of salt accumulation. These strata were identified as waterlogging, salt accumulation, and salt leaching (Tóth et al, 2001a). For 3 locations (numbers 8, 10, and 11) a morphologic description and horizon-wise sampling for laboratory analysis was carried out as a way for validation. The waterlogging zone corresponds to the wet area with the lowest elevation, the natural vegetation is a meadow, and location 10 represents a typical profile.

The zone of salt accumulation is intermediate in elevation, is the most sodic and saline, covered with short grass, and location 8 can be considered as a typical profile. The salt leaching zone has the highest elevation, is the least sodic, the natural vegetation is a tall grass, and location 11 is a typical profile.

All the 413 locations were subject to a *k*-means clustering, but only the membership of the 20 locations considered in this study was reported (table 5.5).

Table 5.5. k-means membership of the locations to the three salt accumulation strata.

Salt accumulation stratum	Waterlogging	Salt accumulation	Salt leaching
Location number	2, 10, 14, 17	1, 3, 7, 8, 12, 15, 16, 19, 20	4, 5, 6, 9, 11, 13, 18

From tables 5.4 and 5.5 we can conclude that the locations classified as low saline belong to the zones of waterlogging or leaching, the latter being more frequent than the former. Locations classified as highly saline originate exclusively from the accumulation zone while locations representative of the average field salinity encompass the three possible zones with the predominance of the accumulation zone, leaching and waterlogging zones are equally represented but less than the accumulation zone.

The A horizon of the salt accumulation zone (location 8) has a very limited hydraulic conductivity compared to the waterlogging zone (location 10) and the salt leaching zone (location 11). Also, the range of soil moisture change is the largest for the accumulation zone. The region is characterized by sudden showers during summer with large amount of rainfall. The soil shows swelling/shrinking properties which results in cracking. The cracks, open on the dry surface, allow a sudden leaching of salts. Consequently it can be expected that the large temporal deviations (less time stable) for locations in the accumulation zone are related to the cracking and the subsequent leaching processes.

Waterlogging and leaching zones showed soil patterns which are strongly time stable. During the wet season, in the leaching zone, the changes in soil salinity occur at greater depth because the groundwater table is deep, and during the warm season, there is no change in salinity as the groundwater is too deep. However in the waterlogging zone the changes occur at depth during the wet season because waterlogging keeps salinity low while during warm season there is no change because groundwater rise is controlled by strong rain infiltration.

Based on the description above, it is possible to optimize the selection of locations used to monitor salinity in the future. The selection of the locations should consider the 3 salt

accumulation zones, which can be identified considering the elevation and the vegetation pattern. The characterization of low saline areas will be based on samples taken from the lowest locations covered with meadow (waterlogging zone) or from the highest zones covered with tall grass (leaching zone) while the investigation of highly saline areas considers the locations with intermediate elevations and covered with short grass (accumulation zone). However if we are interested in characterizing the average field salinity, we will need to obtain samples from the three zones.

5.3.6. How good the selected sites are representative of the average salinity?

To investigate how good the selected sites are representative of the average salinity we focused on $EC_{2.5}$. The average salinity is an important input parameter for salinity predictive models. For example Saltmod (Oosterbaan, 1997) uses the cumulative Gumbel distribution which is assumed to fit the cumulative probability distribution of the root zone salinity. This distribution requires the average salinity and the standard deviation which is function of the average salinity.

Locations 2 and 3 had a mean relative difference approaching zero and the smallest temporal standard deviation. The mean relative difference and its corresponding standard deviation were -0.01 dS/m and 0.15 dS/m for location 2, and 0 dS/m and 0.22 dS/m for location 3.

Comparing the two series of means for using, either all 20 locations or only locations 2 and 3, we found that they agreed in most of the 19 cases. For example allowing for a difference of +/- 0.3 dS/m, they agreed for 14 out of the 19 cases, and in all the cases the difference is not more than 0.5 dS/m. The differences would be smaller if locations 2 and 3 had reduced temporal standard deviations, implying if they were more time stable.

To explore the possibility to use only a limited number of locations to estimate the average $EC_{2.5}$ instead of using all the 20 locations, we split our data set into two subsamples (Grayson and Western, 1998): the first subsample involving only the 9 first measurement times (excluding November 1994), and the second subsample covering the last 9 time instants. We computed the relative differences, their means and their standard deviations for each subsample separately. Then we plotted the average relative differences corresponding to the first subsample ranked from the smallest to the largest difference (Fig. 5.4) and overlayed the corresponding differences computed from the second subsample.
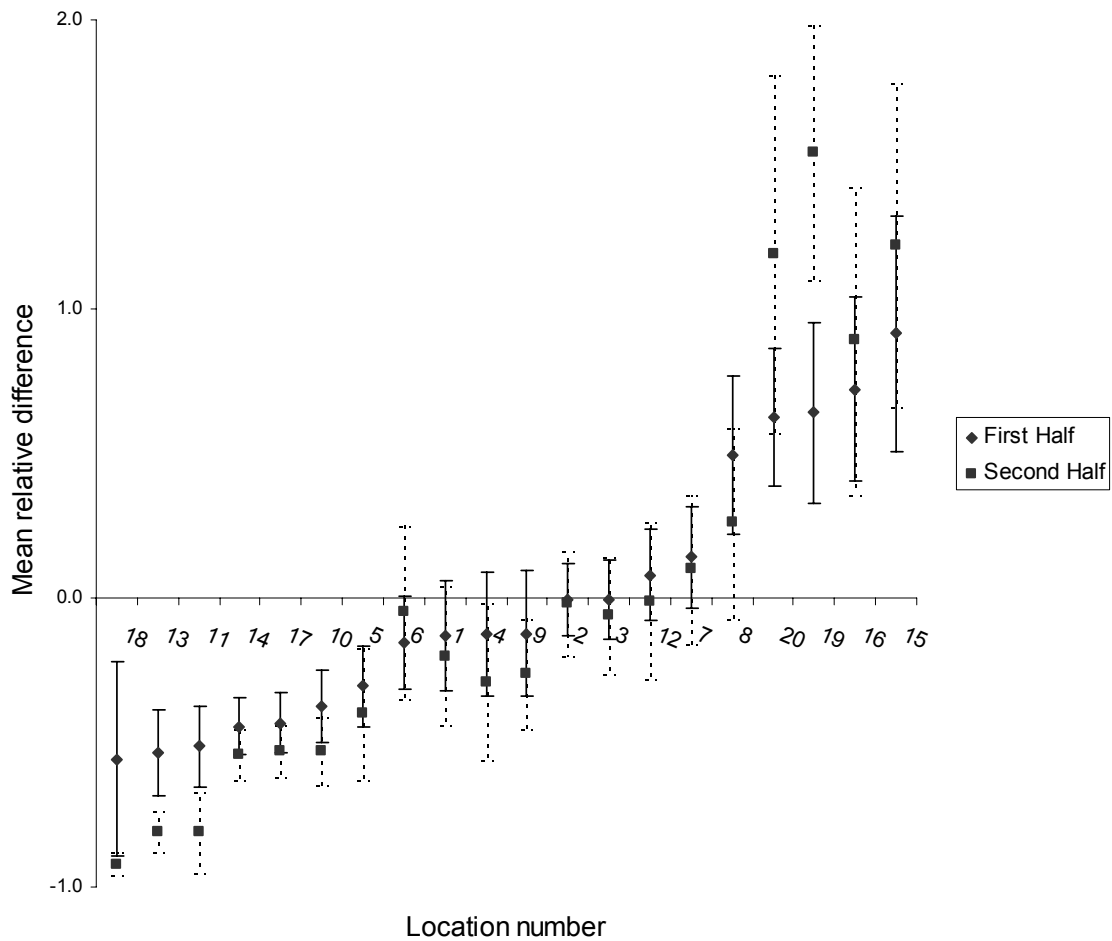
Fig. 5.4. Mean relative differences for $EC_{2.5}$ for the first half of the data, the equivalent values for the second half are overlayed. Vertical bars represent +/- one standard deviation.

We note from this figure that, using only half time coverage data, we reached the same locations representative of the average $EC_{2.5}$ for both halves; these locations were also selected using the whole data (Fig. 5.3). This result provides more confidence in the ability of the method of relative differences to consistently identify and select locations which seems representative of the average field $EC_{2.5}$.

5.4. Conclusions

We applied the concept of temporal stability to soil salinity measurements provided by laboratory analysis ($EC_{2.5}$) and field probes (EC_a). The samples were collected from 20 locations at four depths while EC_a was sensed at two depths. Additionally we analyzed the mean $EC_{2.5}$ over the four depths. The sampling was repeated at 19 time instants over 7 years. The Spearman rank order correlation showed a temporal persistence of the spatial pattern of both properties at all depths. It indicates the strength and the direction of a rising or falling relationship between measurements made at two different time instants. As the Spearman correlation measures only the degree of concordance between two rankings and to find out the locations which were time stable we applied the technique of relative differences. We found no temporal stability of the complete soil salinity pattern. However the low saline conditions were the most time stable while the locations representative of the average soil salinity had intermediate time stability and the high saline locations were the least time stable. Also, the low saline locations were related to zones of waterlogging and/or salt leaching while the high saline locations were related to the zone of salt accumulation. The locations representative of average soil salinity were present in the three zones: salt accumulation, salt leaching, and waterlogging. The concept of temporal stability allowed us to select a limited number of locations (as small as two), which were used to estimate the average soil salinity instead of using the 20 available locations. So this technique can be used to select the locations representative of average and extreme saline conditions, which will be used as 'ground truth' for the calibration and the validation of remote sensing data for the determination, for example, of a soil salinity index. According to the results on the temporal stability of soil salinity in the studied native solonetzic landscape, the zonation of the toposequence (vegetation pattern and elevation) was a very good indicator of the differences in soil salinity and its temporal stability and supports the previous/ongoing soil mappings based on vegetation type.

The elaborated procedure is a general one and can be used in different ecosystems (agricultural field, native vegetation like grassland, bushland, forest), for different soil proprieties (moisture, salinity, soluble nutrients), and under different climatic zones (semi-humid, arid, semi-arid).

CHAPTER 6

DETECTING AND MONITORING TEMPORAL CHANGES OF SPATIAL PATTERNS OF SOIL SALINITY USING ELECTRODE PROBES

Abstract

Soil salinity was evaluated in the east of Hungary by measuring electrical conductivity from 1:2.5 soil:water extract ($EC_{2.5}$) and using 4-electrode probes (EC_a) at 13 to 20 locations during 19 time instants between November 1994 and June 2001. Spearman rank correlation and a stronger criterion, the test of significance of the slope and the intercept of the regression, were used to test the temporal stability of soil salinity. To check the significance of the mean temporal change we used the paired-t test and the first test of Lesch et al. (1998) while their second test was used to check if there is a dynamic spatial variation. Spearman correlation revealed an overall temporal stability in EC_a and $EC_{2.5}$. The stronger criterion showed that the relative differences did not remain constant for some pairs of time instants while it did for others. In the latter case, the change in the spatial pattern was constant for EC_a between March and June 2001 (a decrease of 0.66 dS m^{-1}) while there was a constant decrease for $EC_{2.5}$ between March and June 1995 (0.28 dS m^{-1}) and between March and June 1996 (0.18 dS m^{-1}). The paired-*t* test indicated that the mean difference in EC_a was significant in 14 out of the 18 pairs while for $EC_{2.5}$ it was significant only for 4 pairs. The first test of Lesch et al. (1998) confirmed, for 3 of the 4 pairs, the results of the paired-*t* test. Their second test revealed that there was a dynamic spatial variation for 3 pairs while for the remaining pairs, there was no change or the change was proportional. The results indicate that the temporal variability of soil salinity is complex and unstable.

6.1. Introduction

Soil salinity is one of the most important factors of worldwide land degradation. To control its harmful effect, we need to monitor it both in space and time. Soil salinity can be assessed in the laboratory by determining the electrical conductivity of a solution extracted from a water-saturated soil paste (EC_e). However, this procedure is time consuming and expensive. As an alternative, it can be evaluated by measuring the apparent electrical conductivity (EC_a) in the field using electrode probes or electromagnetic induction instruments. This approach is fast and cheaper, and allows a more intensive surveying of soil salinity. Nevertheless it still needs the collection of soil samples for their analysis in the laboratory to establish calibration equations linking EC_a to EC_e.

As soil salinity is variable in space and time, its monitoring requires numerous measurements. However the spatial pattern of soil salinity can manifest some persistence over time. In this case we can identify certain measuring locations where deviations of soil salinity measurements from the field mean value are nearly constant at all times. This implies that a reduced number of locations could be identified which are able to characterize the soil salinity status of the field and consequently the sampling and survey effort can be reduced.

To check the presence of temporal persistence, also known as temporal stability, Vachaud et al. (1985) introduced two concepts. The first one is based on the Spearman rank correlation. It consists in the determination of the rank correlation coefficient between two time instants. The spatial pattern is considered to be persistent or stable when this correlation coefficient is close to one. The lack of temporal stability is indicated by a coefficient near zero.

The second concept is based on the relative difference. Kachanoski and De Jong (1988) used the relative differences and showed that a good test for the temporal persistence is the Pearson correlation and the simple linear regression between values observed at two consecutive time instants. This finding was applied by many researchers. Van Wesenbeek et al. (1988) examined the temporal persistence of the spatial pattern of soil water storage by computing the coefficient of determination between successive pairs of measurement dates. In addition, Winkel et al. (1995) studied the temporal stability of the shoot biomass production of grapevine response to a soil texture gradient. Da Silva et al. (2001) used the regression between successive dates to determine factors that contribute to the temporal stability in spatial pattern of the water content while the temporal stability of soil moisture was studied by Martínez-Fernández and Ceballos (2003). Similarly, the temporal persistence of the spatial variability of soil moisture was studied by Petrone et al. (2004).

Using the procedure based on the relative differences and the regression between consecutive times, it is possible to check if there is persistence in the spatial pattern of a soil property between two time instants. This temporal persistence implies that the values, measured at various locations, tend to retain their relative ranking over time even though the mean value might have increased or decreased between two measurement instants. We still need to know and test if the mean value has significantly increased or decreased. The latter can be achieved by comparing the mean difference based on the paired-t test (McClave and Sincich, 2000). This approach was used by Kenny et al. (2002) to check the temporal trend of the mean thickness of the Ap horizon.

Lesch et al. (1998) introduced a procedure to check the temporal change between two time instants by combining EC_a and $EC_{2.5}$. Their approach was based on the determination of a calibration equation relating EC_a to $EC_{2.5}$ from the first time frame. Using this equation, they predicted $EC_{2.5}$ for the second time frame and tested for two kinds of change. The first test is used to check if the observed mean value at the second time frame and the mean of the estimated values, from the calibration equation of the first time frame, are significantly different. The second test allows detecting a dynamic spatial variation, i.e. the change in soil salinity was different from one location to another. If this was not the case then the spatial variation was not dynamic meaning either no change or that the change occurred in the same proportion between two time frames for all locations.

The objectives of our study were to check if:

- the mean soil salinity changed (increased or decreased) between two time frames using the paired-t test and the first test of Lesch et al. (1998);
- the change was the same or not from location to location, i.e. if this change was dynamic using the concept of temporal stability and the second test of Lesch et al. (1998).

The paired-t test and the temporal stability were applied to EC_a and $EC_{2.5}$ while the tests of Lesch et al. (1998) were applied only to $EC_{2.5}$ as the latter was predicted from EC_a.

6.2. Material and methods

6.2.1. Data Description

The same data sets and the same variables considered in the last chapter are used here (see chapter 5, section 5.2.1 and chapter 3, section 3.2.1).

6.2.2. Methods

For more detail about the different statistical approaches used in this chapter, see sections 4.2.3 and 4.2.4.

6.3. Results

6.3.1. Correlation Coefficients

The Pearson linear and Spearman rank correlation coefficients between successive time instants for EC_a and $EC_{2.5}$ are reported in table 6.1.

All Pearson correlation coefficients for both variables are very highly significant indicating the presence of temporal stability in the spatial pattern of EC_a and $EC_{2.5}$. The coefficients ranged between 0.81 and 0.98 for EC_a and between 0.58 and 0.94 for $EC_{2.5}$. The Spearman correlation coefficients varied between 0.69 and 0.98 for EC_a and between 0.65 and 0.95 for $EC_{2.5}$. The probability that ranks were not preserved between any two consecutive times was less than 0.001, except for $EC_{2.5}$ between the two first times (0.02), which indicates a very high degree of temporal persistence for the ranking of both variables.

The square of the Pearson correlation coefficients, the coefficients of determination, indicate how much of the spatial variance observed at one time can be explained by the variance present at the precedent time. The values ranged between 66 and 96% for EC_a and between 34 and 88% for $EC_{2.5}$. All the coefficients of determination, except the one for $EC_{2.5}$ between the first two times, were more than 60%, which means that more than 60% of the spatial pattern persisted over time.

Table 6.1. Pearson linear and Spearman rank correlation coefficients for EC_a and $EC_{2.5}$.

Times	Pearson correlation		Spearman correlation	
	EC_a	$EC_{2.5}$	EC_a	$EC_{2.5}$
	dS m^{-1}			
1 - 2	0.85	0.58	0.69	0.65
2 - 3	0.87	0.78	0.82	0.72
3 - 4	0.91	0.93	0.91	0.94
4 - 5	0.93	0.92	0.92	0.96
5 - 6	0.98	0.93	0.98	0.94
6 - 7	0.96	0.84	0.96	0.88
7 - 8	0.88	0.78	0.90	0.83
8 - 9	0.88	0.92	0.92	0.94
9 - 10	0.92	0.93	0.95	0.95
10 -11	0.88	0.90	0.91	0.93
11 - 12	0.86	0.90	0.92	0.88
12 - 13	0.94	0.82	0.94	0.80
13 - 14	0.94	0.94	0.92	0.85
14 - 15	0.81	0.91	0.80	0.85
15 - 16	0.87	0.94	0.89	0.95
16 - 17	0.94	0.94	0.93	0.93
17 - 18	0.83	0.79	0.89	0.79
18 - 19	0.86	0.85	0.91	0.87

1 to 19 refer to: November 1994; March, June, September, and December 1995; March and June 1996; March, June, September, and December 1997; September 1998; April, July and September 1999; April and December 2000; March and June 2001, respectively.

6.3.2. Linear regression models

The intercept of the fitted linear regression models and the slope are reported in table 6.2 for EC_a and in table 6.3 for $EC_{2.5}$ together with some related statistics.

Table 6.2. Regression intercepts and slopes (computed and expected) between EC_a values for consecutive times. I: intercept; p(I=0): probability that the intercept is equal 0; S: slope, p(S=1): probability that the slope is equal 1; ES: expected slope; p(S=ES): probability that the slope is equal to the expected value.

Times	I	$p(I=0)$	S	ES	$p(S=1)$	$p(S=ES)$
	dS m^{-1}		dS m^{-1}	dS m^{-1}		
1 - 2	0.07	0.74	3.94	4.23	0.002	0.69
2 - 3	0.12	0.75	1.17	1.23	0.30	0.71
3 - 4	0.07	0.79	0.82	0.85	0.07	0.77
4 - 5	0.31	0.20	0.94	1.10	0.53	0.10
5 - 6	0.11	0.42	0.80	0.85	0.0004	0.26
6 - 7	-0.03	0.70	0.38	0.37	0.0001	0.71
7 - 8	0.44	0.09	2.13	2.75	0.0007	0.04
8 - 9	0.02	0.92	0.58	0.60	0.0004	0.85
9 - 10	-0.03	0.76	0.62	0.59	0.0001	0.71
10 -11	0.45	0.007	1.15	1.78	0.33	0.0004
11 - 12	0.30	0.47	1.85	2.09	0.005	0.38
12 - 13	-0.01	0.98	1.20	1.19	0.07	0.94
13 - 14	0.78	0.17	1.06	1.28	0.65	0.09
14 - 15	-0.95	0.31	0.75	0.54	0.15	0.22
15 - 16	1.41	0.002	0.75	1.36	0.03	0.0001
16 - 17	-0.68	0.02	0.76	0.54	0.002	0.005
17 - 18	1.11	0.0005	0.64	1.31	0.003	0.0001
18 - 19	-0.66	0.11	1.01	0.70	0.96	0.052

From tables 6.2 and 6.3 we note that the regression intercept was significantly different from zero for 4 and 6 consecutive times for EC_a and $EC_{2.5}$, respectively.

The temporal stability implies that there is a small variation and temporal independence in the relative differences between two time instants. This is indicated by a regression intercept equal to zero and a slope equal to the ratio of the mean value of the observed data at the second time frame to the mean for the first time frame. We note that, for EC_a, this is not checked for the consecutive times 7-8, 10-11, 15-16, 16-17, and 17-18.

Table 6.3. Regression intercepts and slopes (computed and expected) between $EC_{2.5}$ values for consecutive times. The abbreviations are the same as in table 6.2.

Times	I	$p(I=0)$	S	ES	$p(S=1)$	$p(S=ES)$
	dS m^{-1}		dS m^{-1}	dS m^{-1}		
1 - 2	1.05	0.008	0.52	1.62	0.052	0.006
2 - 3	-0.28	0.56	1.00	0.88	1.00	0.54
3 - 4	0.33	0.09	0.84	1.00	0.058	0.06
4 - 5	0.15	0.43	0.84	0.91	0.07	0.40
5 - 6	0.54	0.008	0.78	1.06	0.02	0.004
6 - 7	-0.18	0.64	1.00	0.92	0.98	0.64
7 - 8	0.59	0.01	0.55	0.88	0.0004	0.005
8 - 9	-0.24	0.38	1.34	1.18	0.052	0.33
9 - 10	-0.57	0.06	1.23	1.60	0.13	0.02
10 -11	0.37	0.047	0.76	0.98	0.01	0.02
11 - 12	0.32	0.45	1.86	2.06	0.001	0.38
12 - 13	-0.33	0.45	0.66	0.56	0.006	0.37
13 - 14	0.69	0.01	0.72	1.04	0.005	0.002
14 - 15	-0.20	0.68	1.23	1.14	0.22	0.63
15 - 16	0.24	0.27	0.78	0.88	0.006	0.16
16 - 17	0.02	0.93	1.15	1.16	0.17	0.90
17 - 18	0.63	0.04	0.49	0.75	0.0001	0.01
18 - 19	-0.25	0.54	1.27	1.13	0.17	0.47

Regarding $EC_{2.5}$, the intercept and the slope were significantly different from the expected values for the consecutive times 1-2, 5-6, 7-8, 9-10, 10-11, 13-14, and 17-18. This means that for these pairs of times there was no temporal stability and consequently the relative ranking of the locations did not remain the same between two consecutive times. We deduce also that, for the remaining pairs of times, the relative differences remained constant and the relative ranking of the locations remained approximately the same. However we still don't know if the change between the above two successive measurements was the same for all the locations or if it was different from one location to another. If there was a change which affected uniformly all the spatial positions between consecutive times, the regression slope should be equal to one and the intercept represents the constant change that has occurred (Kachanoski

and De Jong, 1988; Winkel et al., 1995). Based on the results from tables 2 and 3, we can note that, for EC_a, a constant change occurred between consecutive times 2-3, 3-4, 4-5, 12-13, 13-14, 14-15, and 18-19. This variable increased between 2-3 (0.12 dS m^{-1}), 3-4 (0.07 dS m^{-1}), 4-5 (0.31 dS m^{-1}), and 13-14 (0.78 dS m^{-1}) while it decreased between 12-13 (0.01 dS m^{-1}), 14-15 (0.95 dS m^{-1}), and 18-19 (0.66 dS m^{-1}). For $EC_{2.5}$, the constant change has occurred between consecutive times 2-3 to 4-5, 6-7, 8-9, 14-15, 16-17, and 18-19 with a uniform increase between 3-4 (0.33 dS m^{-1}), 4-5 (0.15 dS m^{-1}), and 16-17 (0.02 dS m^{-1}) and a constant decrease between 2-3 (0.28 dS m^{-1}), 6-7 (0.18 dS m^{-1}), 8-9 (0.24 dS m^{-1}), 14-15 (0.20 dS m^{-1}), and 18-19 (0.25 dS m^{-1}).

The precedent results were based on a probability of the error of type I of 5%. However the regression slope was equal to one, senso stricto, only between times 18 and 19 for EC_a, and between times 2 and 3, and 6 and 7 for $EC_{2.5}$. The mean difference values between two consecutive times are reported in table 5.4. We note from tables 2 to 4 that the mean difference value was equal to the intercept only for the consecutive times for which the regression slope was effectively equal to one but not in the statistical significance sense. The discrepancy between these two values became larger as the significance level of the unity slope became smaller.

6.3.3. Temporal mean change

The results of the paired-t test applied to EC_a and $EC_{2.5}$ are reported in table 6.4. We note that the mean EC_a difference between two consecutive times was significantly different from zero for 14 out of the 18 pairs of times. For these cases the mean value decreased or increased for the same number of occasions (7). However the results were different for $EC_{2.5}$. For this variable the hypothesis that the mean difference was not equal to zero was rejected only for 4 consecutive times, from which the mean value decreased between one pair of times while it increased for the three others.

Table 6.4. Paired-t test for consecutive EC_a and $EC_{2.5}$ measurements. MD: mean difference; p(MD=0): probability that the mean difference is not significantly different from zero.

Times	EC_a		$EC_{2.5}$	
	MD	*p(MD=0)*	*MD*	*p(MD=0)*
	dS m^{-1}		dS m^{-1}	
1 - 2	0.77	0.0005	0.37	0.006
2 - 3	0.43	0.06	-0.27	0.09
3 - 4	-0.33	0.08	-0.01	0.97
4 - 5	0.20	0.21	-0.17	0.06
5 - 6	-0.36	0.01	0.11	0.24
6 - 7	-1.25	0.0001	-0.17	0.20
7 - 8	1.24	0	-0.22	0.09
8 - 9	-0.73	0.003	0.27	0.03
9 - 10	-0.44	0.002	-0.17	0.25
10 -11	0.55	0	-0.03	0.83
11 - 12	1.37	0.0002	1.70	0
12 - 13	0.51	0.03	-1.46	0
13 - 14	0.98	0.01	0.08	0.72
14 - 15	-2.06	0.003	0.32	0.26
15 - 16	0.84	0.02	-0.29	0.10
16 - 17	-1.44	0	0.33	0.06
17 - 18	0.52	0.06	-0.58	0.06
18 - 19	-0.65	0.007	0.22	0.33

The first approach of Lesch et al. (1998) compares the mean value for the observed $EC_{2.5}$ at the second time frame to the mean of the predicted values based on the regression modeling using EC_a and $EC_{2.5}$ from the first time frame. Table 6.5 shows the parameters (intercept and slope), the coefficient of determination and the mean square error of the regression calibration equations linking $EC_{2.5}$ to EC_a for each time frame.

Table 6.5. Parameters of the calibration regression models linking $EC_{2.5}$ to EC_a with the coefficient of determination (r^2) and the mean square error (MSE).

Time	Intercept	Slope	r^2	MSE
	dS m^{-1}	dS m^{-1}		
1	0.81	2.56	0.72	0.07
2	1.29	0.54	0.83	0.13
3	0.87	0.50	0.77	0.29
4	0.97	0.54	0.89	0.11
5	0.83	0.47	0.85	0.13
6	1.20	0.44	0.75	0.17
7	0.87	1.35	0.76	0.20
8	0.82	0.41	0.80	0.08
9	0.91	0.80	0.69	0.32
10	0.51	1.59	0.88	0.19
11	0.36	0.99	0.82	0.21
12	0.95	0.89	0.72	1.37
13	-0.10	0.62	0.86	0.44
14	0.38	0.43	0.83	0.47
15	0.85	0.62	0.83	0.65
16	0.20	0.61	0.89	0.32
17	0.83	0.92	0.87	0.49
18	0.12	0.77	0.95	0.08
19	0.70	0.86	0.74	0.84

Based on the parameters from table 6.5, $EC_{2.5}$ was predicted for successive times and the differences between the observed and the predicted $EC_{2.5}$ values with their mean were computed. The latter with the test statistic c and its probability of significance are reported in table 6.6.

Table 6.6. Mean difference values and their significance level for consecutive times. MD: mean difference; p(MD=0): probability that the mean difference is not significantly different from zero.

Times	MD	Statistic c	$p(MD=0)$
	$dS\ m^{-1}$		
1 - 2	0.37	2.32	0.04
2 - 3	-0.28	-1.48	0.16
3 - 4	0	-0.02	0.99
4 - 5	-0.18	-1.02	0.32
5 - 6	0.15	0.83	0.42
6 - 7	-0.17	-0.86	0.40
7 - 8	-0.22	-1.05	0.31
8 - 9	0.22	1.28	0.22
9 - 10	-0.17	-0.55	0.59
10 -11	-0.03	-0.12	0.90
11 - 12	1.70	4.41	0
12 - 13	-1.45	-2.84	0.01
13 - 14	0.21	0.67	0.51
14 - 15	0.32	0.73	0.48
15 - 16	-0.18	-0.49	0.63
16 - 17	0.33	1.15	0.27
17 - 18	-0.56	-1.48	0.16
18 - 19	0.23	0.90	0.38

The shift in the mean field salinity was significantly different for three pairs of time frames with an increase between times 1-2 (0.37 dS m^{-1}) and 11-12 (1.70 dS m^{-1}) while the mean $EC_{2.5}$ decreased between 12-13 (1.45 dS m^{-1}). For all the remaining pairs of times, the mean change in $EC_{2.5}$ may be considered as not significantly different from zero.

Comparing the mean differences from tables 6.4 and 6.6 we can note a good agreement between the two series of means even if the methods of computation are basically different. Also we note that the mean difference was significantly different from zero for three pairs of measurements instead of four pairs when the paired-t test was used (table 6.4).

6.3.4. Dynamic spatial variation

The approach of Lesch et al. (1998) allows to test, in addition to the mean change, if the salinity pattern has changed in a spatially variable manner, i.e. if the change was different from location to location, between two time frames (space-time interaction). The results from this test are shown in table 6.7.

Table 6.7. Test of significance of the dynamic spatial variation: statistic φ and its significance p(φ=0).

Times	Statistic φ	$p(\varphi=0)$
1 - 2	1.95	0.13
2 - 3	2.19	0.051
3 - 4	0.58	0.88
4 - 5	2.18	0.053
5 - 6	1.40	0.24
6 - 7	0.65	0.81
7 - 8	1.09	0.43
8 - 9	2.52	0.03
9 - 10	1.12	0.43
10 -11	1.42	0.23
11 - 12	9.62	0
12 - 13	0.80	0.68
13 - 14	0.66	0.76
14 - 15	2.06	0.11
15 - 16	1.02	0.48
16 - 17	1.57	0.18
17 - 18	1.88	0.10
18 - 19	10.84	0

From the results of table 6.7, we can consider that the spatial variation in $EC_{2.5}$ was different from location to location for three pairs of times: between times 8-9, 11-12, and 18-19. For all the other consecutive times, there was either no change between the pairs of times or the

spatial variation was constant or proportional from location to location between each pair of times.

Based on the results from tables 6.6 and 6.7, we note that we have four different scenarios in the spatial pattern of $EC_{2.5}$:

- For the pairs of times 1-2, and 12-13: there was a significant change in the mean value but this change was not different from location to location,

- For the pairs of times 2-3 to 7-8, and 13-14 to 17-18: the mean change was not significantly different from zero and this change was static (not different from location to location),

- For the pairs of times 8-9, and 18-19: the mean change was not significantly different from zero however this change was dynamic (different from location to location), but when averaged over all the locations it was not significant,

- For the pair of times 11-12: the mean change was significantly different from zero and this change was dynamic.

6.4. Discussion and conclusions

The utility of the concept of temporal stability lies in the possibility to identify particular locations that represent important statistics of the variable under study as the mean value or the mean value ± one standard deviation. This implies that the sampling effort for monitoring soil proprieties could be noticeably reduced. The Spearman rank correlation and the relative differences, two techniques used to check the temporal stability, allow to test if the relative ranking of locations remained the same between two time instants. However we still need to know if the change that occurred is the same for all locations between the two time instants or if it was different. The linear transformation between a variable measured at two consecutive times, as a consequence of the existence of the temporal stability of the spatial pattern, can be used to check this hypothesis. When the slope of the regression between values from two consecutive times is equal to the ratio of the means and the intercept is zero, we are able to conclude that the temporal stability exists, the relative differences remain constant, and the relative ranking of locations remains the same. Nevertheless the relative ranking does not inform us about the nature of the change that occurred because the relative ranking can remain the same for the cases of no change, a constant change (increase or decrease) or even a different change from location to location. Using the concept of temporal stability, the only

situation for which we are able to know the nature of the change, is when there is no change or constant change between two times. In this case the constant change is also the mean change that occurred between two measurements. We still need an approach to check if the change in the mean value is significant. The paired-t test is used to answer this question. It can be also checked using the first test of Lesch et al. (1998). Moreover their second test allows checking the existence of a dynamic spatial variation, meaning if the soil salinity changed differently from location to location. The usefulness of their approach lies in the fact that only the EC_a - $EC_{2.5}$ pairs of values from the first time frame are needed along with the $EC_{2.5}$ values from the second time frame. If it is found that there was no change, or the change was constant, between the two time frames, the regression equation computed using the first time frame data can be used to determine the $EC_{2.5}$ values for the second time frame using a larger EC_a survey data set. However if a dynamic spatial variation was detected, then a new regression equation should be estimated using the EC_a - $EC_{2.5}$ pairs of values from the second time frame.

Based on the Spearman rank correlation coefficient, we found that the spatial pattern of EC_a and $EC_{2.5}$ appeared to be stable over time. However using the stronger criterion of the equality of the relative differences between two time instants, we found that for some pairs of time frames, the relative differences did not remain constant even though the ranking remained almost the same. This was the case for EC_a between the time instants 7-8, 10-11, 15-16, 16-17, and 17-18 while for $EC_{2.5}$ between time frames 1-2, 5-6, 7-8, 9-10, 10-11, 13-14, and 17-18. The spatial pattern of EC_a and $EC_{2.5}$ was found to be time stable for the other pairs of times. Among the latter, the change in the spatial pattern was constant between time instants 18-19 (a decrease of 0.66 dS m^{-1}) for EC_a and between 2-3 (a decrease of 0.28 dS m^{-1}) and 6-7 (a decrease of 0.18 dS m^{-1}) for $EC_{2.5}$. These constant changes reflected also the mean change. To further test if the mean change was significant or not, we used the paired-t test and the first test of Lesch et al. (1998). Both techniques identified a significant mean difference for three common pairs of times: mean change between time instants 1-2, 11-12, and 12-13, while the paired-t test identified an additional pair of times (8-9) for which the mean change was also significant. So both techniques agreed in their results even if they are based on different data sources. The dynamic spatial variation was checked using the second test of Lesch et al. (1998). We found that the spatial variation was different from location to location for three pairs of measurements: 8-9, 11-12, and 18-19. For the remaining couples of times, the spatial pattern was considered to have not been changed or changed in a constant or proportional manner.

The temporal stability concept requires a large number of measurements, but once it is checked, future measurements are reduced to locations that represent important statistics. Therefore the sampling effort can be reduced. Likewise, based on the Lesch et al. (1998) approach, if it is found that there is no dynamic spatial variation, the sampling effort can be further reduced to the EC_a survey for the first time frame while the laboratory analysis can be done for two time frames but at a reduced number of locations. So the joint use of the concept of temporal stability and the two tests of Lesch et al (1998) could drastically reduce the sampling effort.

PART III

GEOSTATISTICAL AND BAYESIAN MAXIMUM ENTROPY APPROACHES

CHAPTER 7

SPATIO-TEMPORAL KRIGING OF SOIL SALINITY RESCALED FROM BULK SOIL ELECTRICAL CONDUCTIVITY

Abstract.

Our spatial data consist of 413 measurements of the apparent electrical conductivity (EC_a) obtained with electrical probes in the east of Hungary. Additionally, a limited subset of the locations (15 to 20) was sampled for laboratory analysis of soil electrical conductivity of 1:2.5 soil:water suspension ($EC_{2.5}$), a simple proxy for the electrical conductivity of soil saturation extract (EC_e). The latter formed our calibration data set. This procedure was repeated 17 times between November 1994 and December 2000 yielding a large spatio-temporal database.

The first step was to rescale $EC_{2.5}$ from EC_a, based on the calibration data sets, using classical and spatial regression models. The residuals of the ordinary least squares model were tested for the absence of spatial dependence using the Moran's *I* test. This hypothesis was accepted, the $EC_{2.5}$ was rescaled using the classical regression model. The next step was to identify the structure of the variability of the rescaled $EC_{2.5}$ by computing and modeling the spatial, the temporal, and the spatio-temporal covariance functions. Finally, soil salinity maps were produced for the study area and for any time instant using spatio-temporal kriging. The estimates were more precise compared to the ones obtained using only the spatial covariance function computed and modeled separately for each time instant.

7.1. Introduction

The effective control of soil salinity requires the knowledge of its magnitude and extent, and also its changes over time. Detecting trends, which occur in salinity conditions over time, is an important step to identify emerging problems, and to determine the progress of reclamation efforts.

Soil salinity assessment requires inventory and monitoring of soil salinity, since it is spatially variable and temporally dynamic in nature.

Soil salinity is conventionally determined by measurements of the electrical conductivity of the extract of a water-saturated soil-paste (EC_e). This property can also be observed indirectly from measurements of the apparent electrical conductivity (EC_a) of the bulk soil. The latter is measured in the field using electrical probes.

The conventional soil sampling and laboratory analysis procedure is very expensive. A cost-effective way is to use mobile techniques for rapidly measuring EC_a as a function of the spatial position, to infer EC_e from EC_a, and to map EC_e at any location in space and any instant in time.

Lesch et al. (1998) developed a statistical monitoring strategy. It requires the estimation of a conditional regression model to predict EC_e from EC_a, and the use of 2 statistical tests: one for detecting dynamic spatial variation in the salinity pattern and the other for detecting a change in the field median salinity level with time. The drawback of this approach is that we get salinity maps only for the observed time instants, and at the observed locations.

We propose in this work to use an alternative approach, based on geostatistical tools, which is capable of using the spatial and temporal dependencies as well as producing maps for any location in space and any time instant.

7.2. Data sets

Both 'calibration data set' and 'data set to be calibrated' were used in this chapter (see chapter 3, section 3.2.1). A lot of research on salinity/sodicity and its correlation to the vegetation has been done in this natural ecosystem (Toth et al., 1991; Van Meirvenne et al., 1995; Toth et al., 1998; Toth et al., 2001b; and Toth et al., 2002b).

7.3. Analysis

The histograms of the EC_a and also of $EC_{2.5}$ showed skewed distributions and after a logarithmic transformation, the distributions became less asymmetric. The linearity of their relationship was also improved (see section 7.4.1). All the analysis was based on the transformed data. The calibration data set was used to compute the calibration equations (one equation for each time instant) as a first step. We tried to relate $EC_{2.5}$ to EC_a using 2 approaches. The first method is the classical ordinary least squares regression (OLS):

$$\ln(EC_{2.5}) = a + b\ln(EC_a) + \xi$$

where a and b are the regression coefficients and ξ represents the independently Gaussian errors.

The other approach is the spatial regression (Anselin, 1988). For this method we checked 4 different models:

- the spatial autoregressive model (SAM):

$$\ln(EC_{2.5}) = \rho W\ln(EC_{2.5}) + b\ln(EC_a) + \xi$$

It is also known under the name of spatial lag model. The spatial dependence is incorporated, in this case, as an additional regressor in the form of a spatially lagged dependent variable $[W\ln(EC_{2.5})]$. The coefficient related to this additional regressor reflects the spatial dependence inherent in the sample data and measures the average influence of neighbouring or contiguous observations on the observed dependent values. This model is appropriate when the concern is the assessment of the existence and strength of spatial interaction

- the spatial error model (SEM):

$$\ln(EC_{2.5}) = b\ln(EC_a) + \omega \qquad\text{with } \omega = \lambda W\omega + \xi \text{ or } (I - \lambda W)\omega = \xi, I \text{ being the identity}$$
matrix.

For this model, the spatial dependence is incorporated in the error structure, which means that $E\left[\xi_i\xi_j\right] \neq 0$. It is appropriate when the focus of interest is with correcting for the potentially biasing influence of the spatial autocorrelation. It captures the influence of unmeasured independent variables.

- the spatial general model (SGM) which is the combination of the 2 above models:

$$\ln(EC_{2.5}) = \rho W \ln(EC_{2.5}) + b \ln(EC_a) + \omega \qquad \text{with} \quad \omega = \lambda W \omega + \xi \text{ or } (I - \lambda W)\omega = \xi.$$

In these equations ρ and λ are the spatial autocorrelation parameters, ω represents the errors with spatial dependence, and W is the matrix of the spatial weights build from the distance separating 2 observations using the Delaunay triangulation algorithm (see section 4.2.5).

- The geographically weighted regression, GWR (Brundson et al., 1996):

$$W_i^{1/2} \ln(EC_{2.5}) = W_i^{1/2} b \ln(EC_a) + \xi$$

The approach is, in essence, non-parametric. It is a method of analysing spatially varying relationships, assuming that data are spatially non-stationary. It uses distance-weighted sub-samples of the data to produce locally linear regression estimates for every point in space. Each estimated set of parameters is based on a distance-weighted sub-sample of neighbouring observations based on distances separating the observations. The distance-based weights can be determined based on a vector of distance between a given observation (space location) and all other observations in the sample in a similar way to the definition of the proximity matrix (see section 4.2.5).

The residuals of the OLS regression were tested for the presence of spatial autocorrelation using the Moran's *I* test (Cliff and Ord, 1981). Further, we used maximum likelihood-based tests, on the results of the spatial regression, to check the significance of the spatial autocorrelation parameters (ρ and λ) and to choose the most adequate model.

At the end of this step, we got a data matrix of 17 columns (time instants) and 413 rows (locations) of $EC_{2.5}$ values with some missing values corresponding to the locations for which EC_a was not available.

This data matrix can be considered as a space-time random field, STRF (Christakos, 1992):

$$X(s, t), (s, t) \in D \times T$$

with $D \subset \Re^2$ (real numbers set) and $T \subset \Re_+$ (positive real numbers set)

with s : 2-D spatial coordinates and t : temporal coordinate.

The second step in our analysis was to model the spatial, the temporal and the spatio-temporal dependence of the salinity data matrix.

The procedure is as follows (Christakos et al., 2002):

- First, the space-time mean trend is estimated. The smoothed spatial components (one for each location) were computed using an exponential spatial filter applied to the averaged measurements (for each location, over all the time instants). We computed also the smoothed temporal components (one for each time instant) using an exponential temporal filter applied to the averaged measurements (for each time instant, over all the locations);

- Then, the above components of the space-time mean trend were interpolated to the data grid giving $m(s, t)$;

- The residuals were computed as the space-time mean trend subtracted from the original data matrix: $R(s, t) = X(s, t) - m(s, t)$;

- The residual data matrix was used to compute the spatial $c(h, \tau=0)$, temporal $c(h=0, \tau)$ and spatio-temporal $c(h, \tau)$ covariance functions :

$$c(\boldsymbol{h},0) = \mathrm{cov}\big[R(\boldsymbol{s}+\boldsymbol{h}), R(\boldsymbol{s})\big],$$

$$c(\boldsymbol{0},\tau) = \mathrm{cov}\big[R(t+\tau), R(t)\big],$$

$$c(\boldsymbol{h},\tau) = \mathrm{cov}\big[R(\boldsymbol{s}+\boldsymbol{h}, t+\tau), R(\boldsymbol{s},t)\big]$$

where $\boldsymbol{h}$ and τ are the spatial and temporal lags, respectively, and cov is the covariance function.

- Finally we fitted theoretical models to the computed experimental covariance functions. The available data were sparse and so we were restricted to an isotropic model, $c(\boldsymbol{h},\tau) = c(|\boldsymbol{h}|,\tau)$.

The last step was the estimation at unobserved locations and time instants using space-time kriging (Chiles and Delfiner, 1999; Christakos, 1992) and the fitted covariance functions.

As we used the residual covariance functions, the resulting estimated data corresponded to the residual values. To get the values in the original scale, we interpolated the spatial and temporal components of the space-time mean trend to the kriging grid. These estimated values were added to the interpolated space-time mean trend values to obtain the kriged values in the original scale.

The classical regression was done using the SAS software (SAS Institute, 1993), the spatial regression was fitted using the Econometrics Toolbox (Lesage, 1999) running under Matlab software. The geostatistical computations were handled using the BMElib library (Christakos and al., 2002). The toolbox and the library are built on the Matlab software (Mathworks, 1999).

7.4. Results

7.4.1. Exploratory data analysis

As there was only up to 20 data available for $EC_{2.5}$ and EC_a from the calibration data set at each of the 19 time instants, their histograms were not so well behaved. Even tough, we can still note the tail heaviness of the distributions of the data on the original scale whereas these distributions looked less asymmetric and even like a normal distribution when a logarithmic transformation was applied to data. Some examples are shown in Fig. 7.1 for EC_a and in Fig. 7.2 for $EC_{2.5}$ on both original and logarithmic scales for September and December 1997. Additionally, the histograms of the EC_a (Fig. 7.3) from all the surveyed locations (413) from which the 20 space locations above were selected for soil sampling, showed that for all time instants, the distribution is markedly skewed. When their cumulative histograms where plotted on a logarithmic scale (Fig. 7.4), the distributions have a sigmoidal shape, which is a characteristic of normal distributions. These graphical assertions were confirmed using the statistical test of Shapiro-Wilk. The results of this test are given, for $EC_{2.5}$, in table 7.1 on both original and logarithmic scales. The Shapiro-Wilk statistic W may be thought of as the correlation between $EC_{2.5}$ and their corresponding normal scores, with $W = 1$ when $EC_{2.5}$ data are perfectly normal in distribution. When W is significantly smaller than 1, the assumption of normality is not met. Equivalently, a significant W statistic, for example $p(W)$ less than 5%, causes the rejection of the assumption that the distribution of $EC_{2.5}$ is normal. Consequently, all analyses were done based on the logarithmic transformed data. From table 7.1, we note that the logarithmic transformation contributed to the increase of W for 14 out of the 19 time instants. Also, this transformation reduced the rejection of the hypothesis of normality from 6 to only 2 time instants.

Table 7.1. Shapiro-Wilk test of normality for $EC_{2.5}$. W: test statistic, p(W): probability of significance of W. Bold values for time instant for which the assumption of normality was rejected.

Time	$EC_{2.5}$		$\ln(EC_{2.5})$	
	W	p(W)	W	p(W)
Nov-94	0.87	0.05	0.83	**0.02**
Mar-95	0.90	0.05	0.94	0.30
Jun-95	0.90	0.05	0.97	0.67
Sep-95	0.93	0.14	0.96	0.51
Dec-95	0.89	**0.03**	0.92	0.12
Mar-96	0.95	0.49	0.94	0.35
Jun-96	0.92	0.13	0.93	0.15
Mar-97	0.97	0.78	0.92	0.12
Jun-97	0.91	0.16	0.96	0.67
Sep-97	0.88	**0.02**	0.97	0.68
Dec-97	0.96	0.47	0.93	0.19
Sep-98	0.93	0.14	0.88	**0.02**
Apr-99	0.79	**0.00**	0.97	0.81
Jul-99	0.92	0.21	0.93	0.36
Sep-99	0.87	**0.01**	0.91	0.07
Apr-00	0.91	0.08	0.93	0.19
Dec-00	0.89	**0.02**	0.90	0.05
Mar-01	0.90	0.06	0.91	0.06
Jun-01	0.83	**0.00**	0.93	0.16
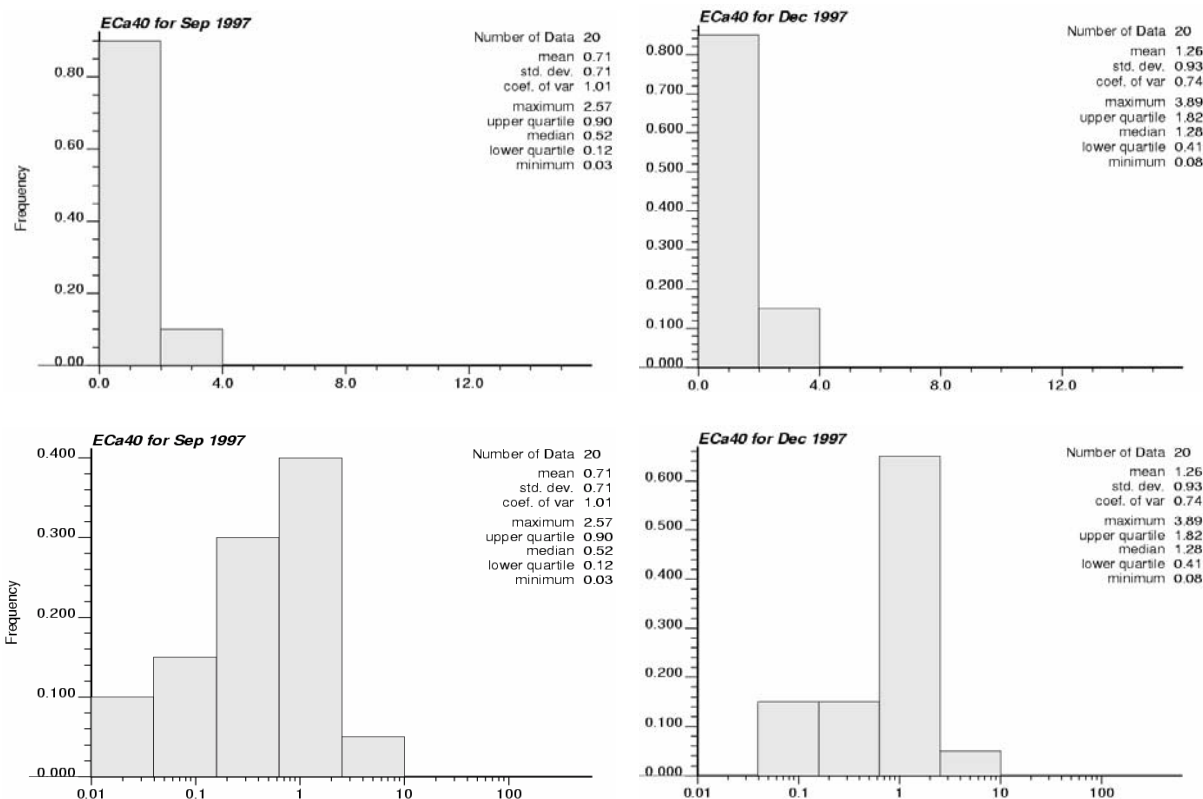
Fig. 7.1. Histograms of EC_a from calibration locations. Above: original scale, below: logarithmic scale.
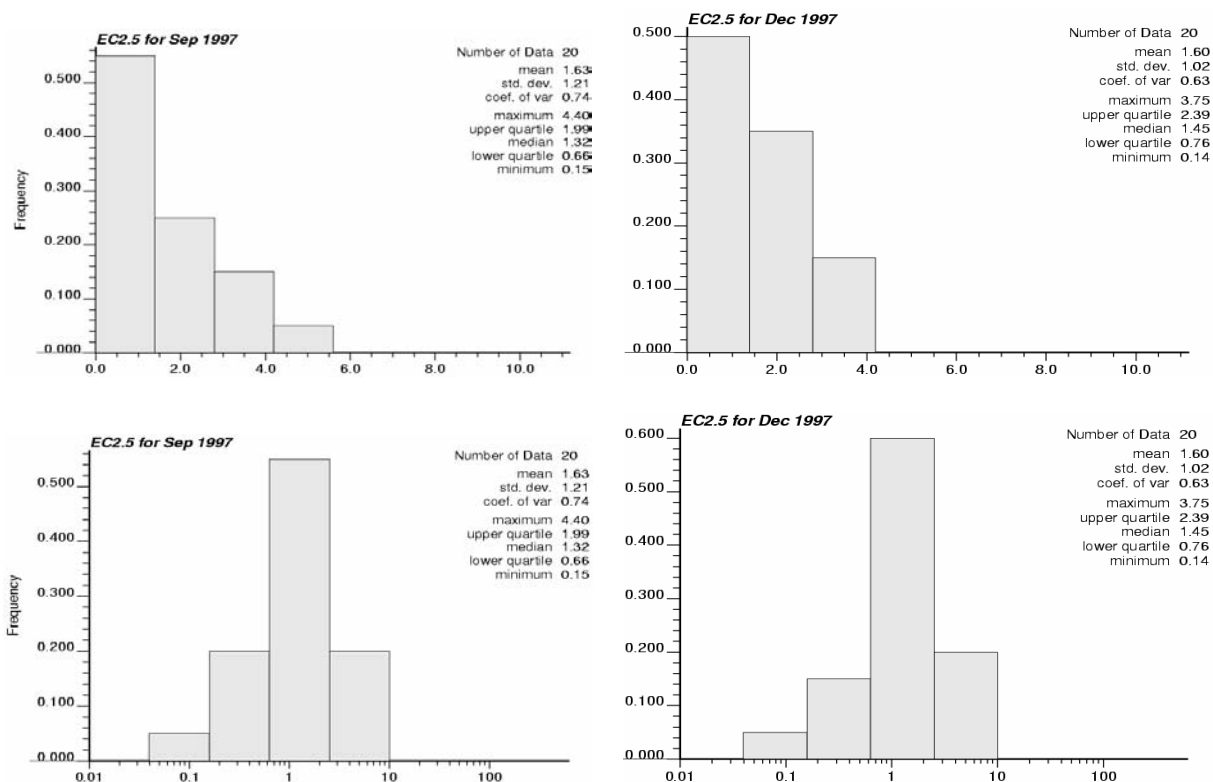
Fig. 7.2. Histograms of $EC_{2.5}$ from calibration locations. Above: original scale, below: logarithmic scale.
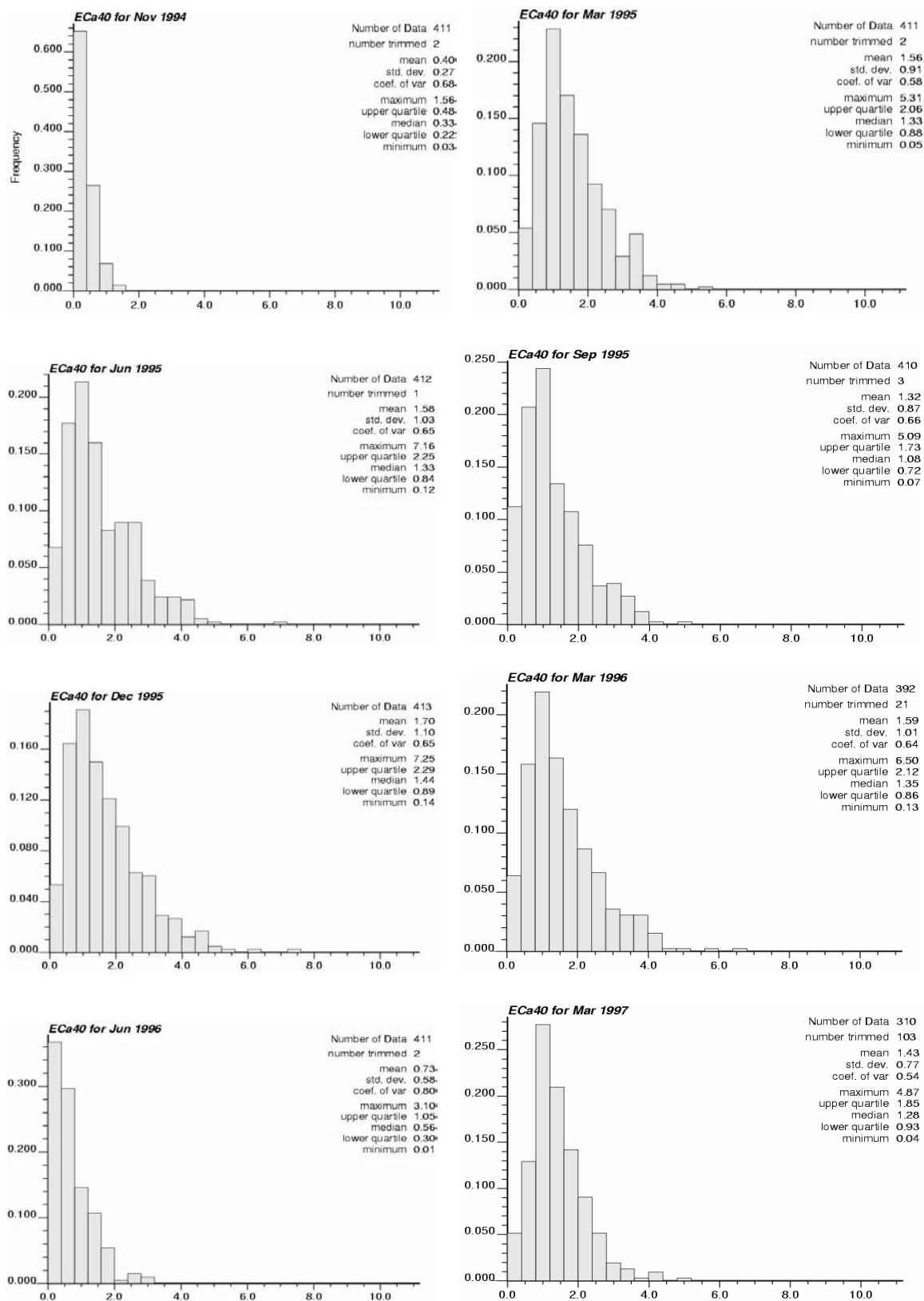
Fig. 7.3. Histograms of EC_a from all the space locations (413).
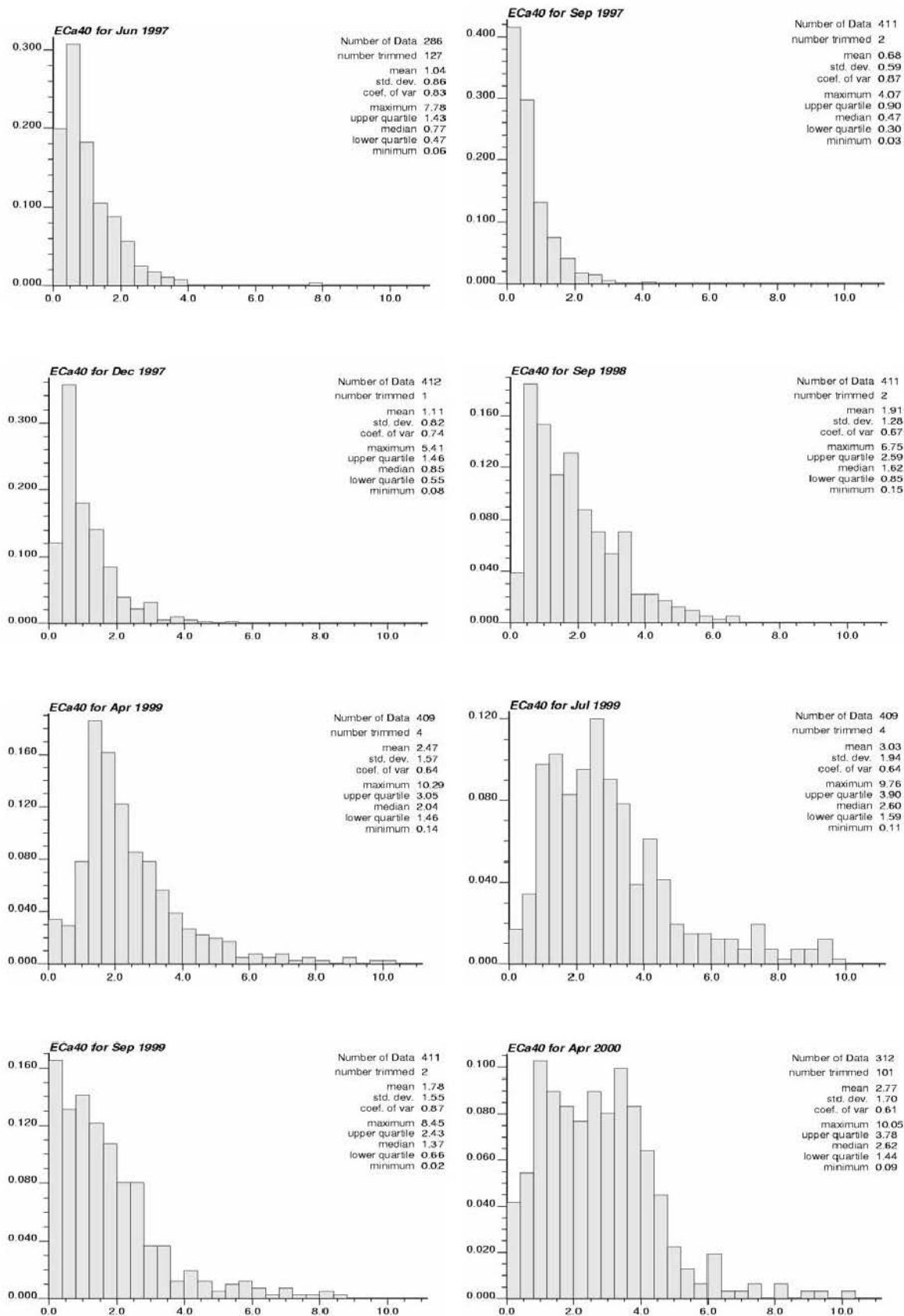
Fig. 7.3. Histograms of EC_a. (continued).
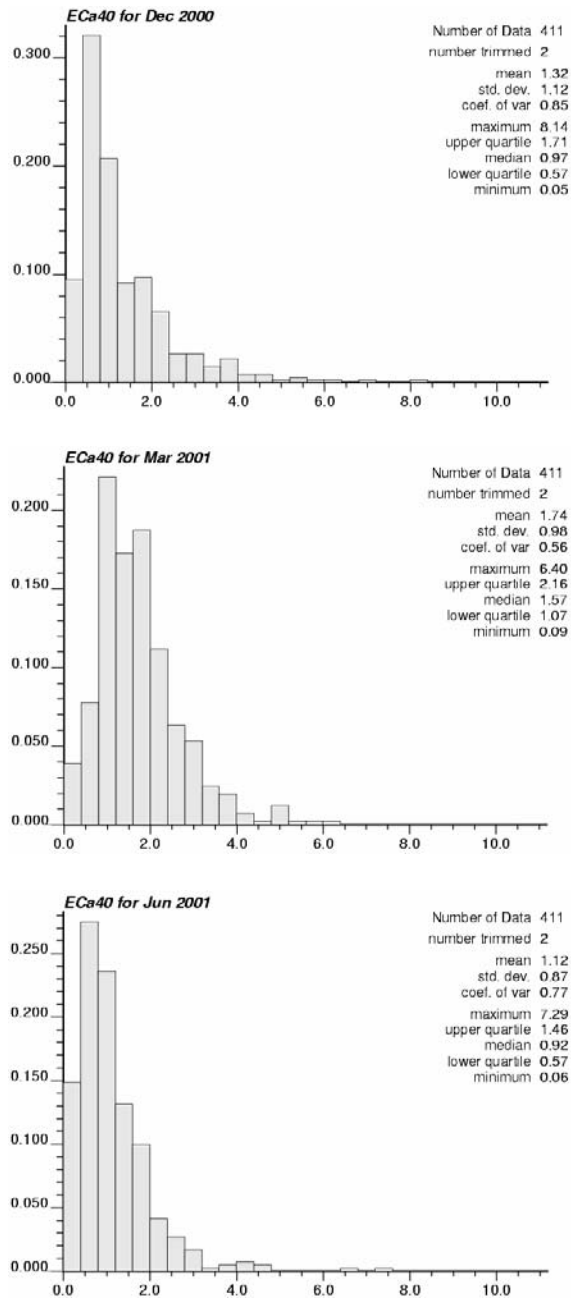
Fig. 7.3. Histograms of EC$_a$. (continued and end).
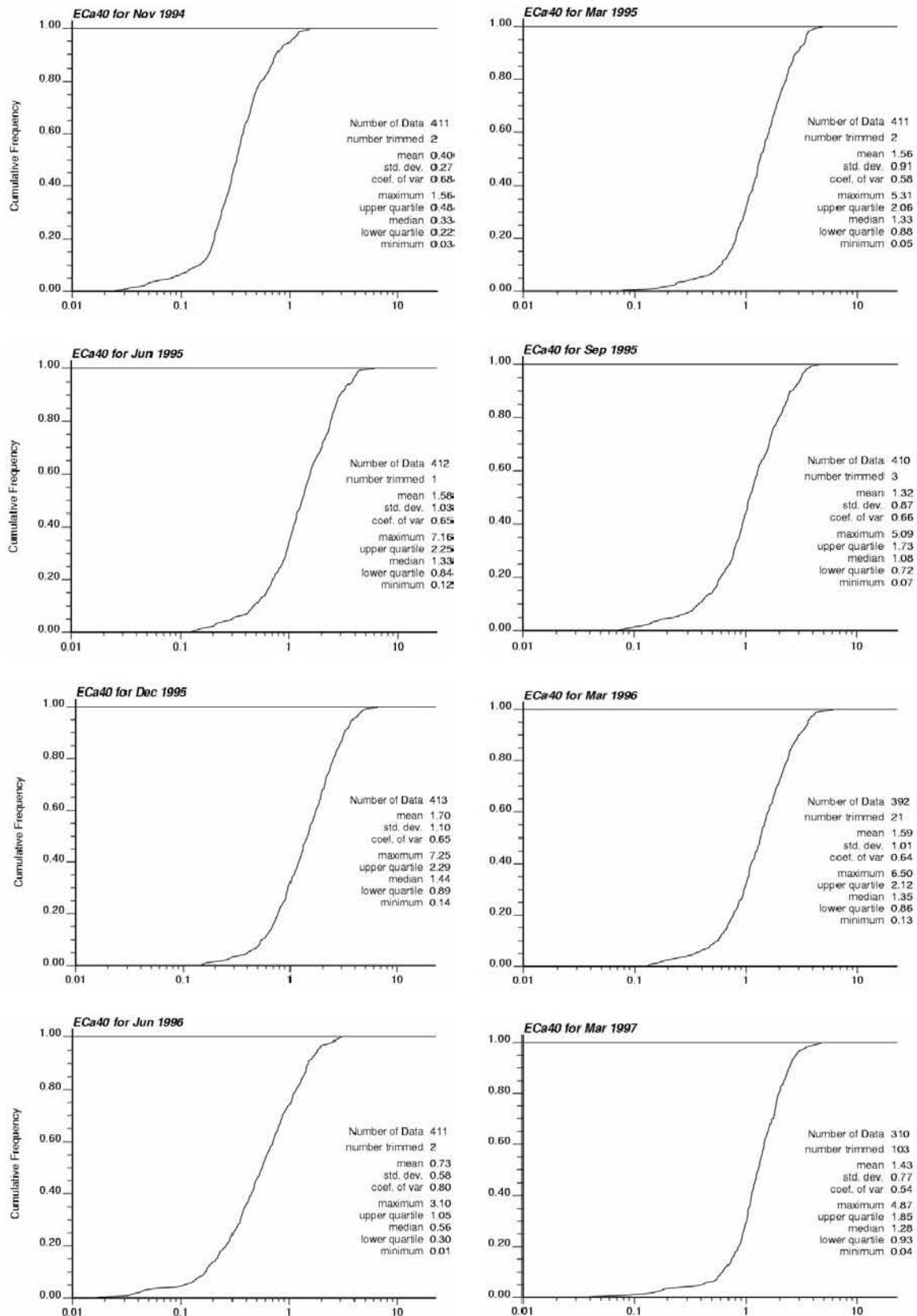
Fig. 7.4. Cumulative distribution functions of EC$_a$ from all the space locations, plotted on a logarithmic scale.
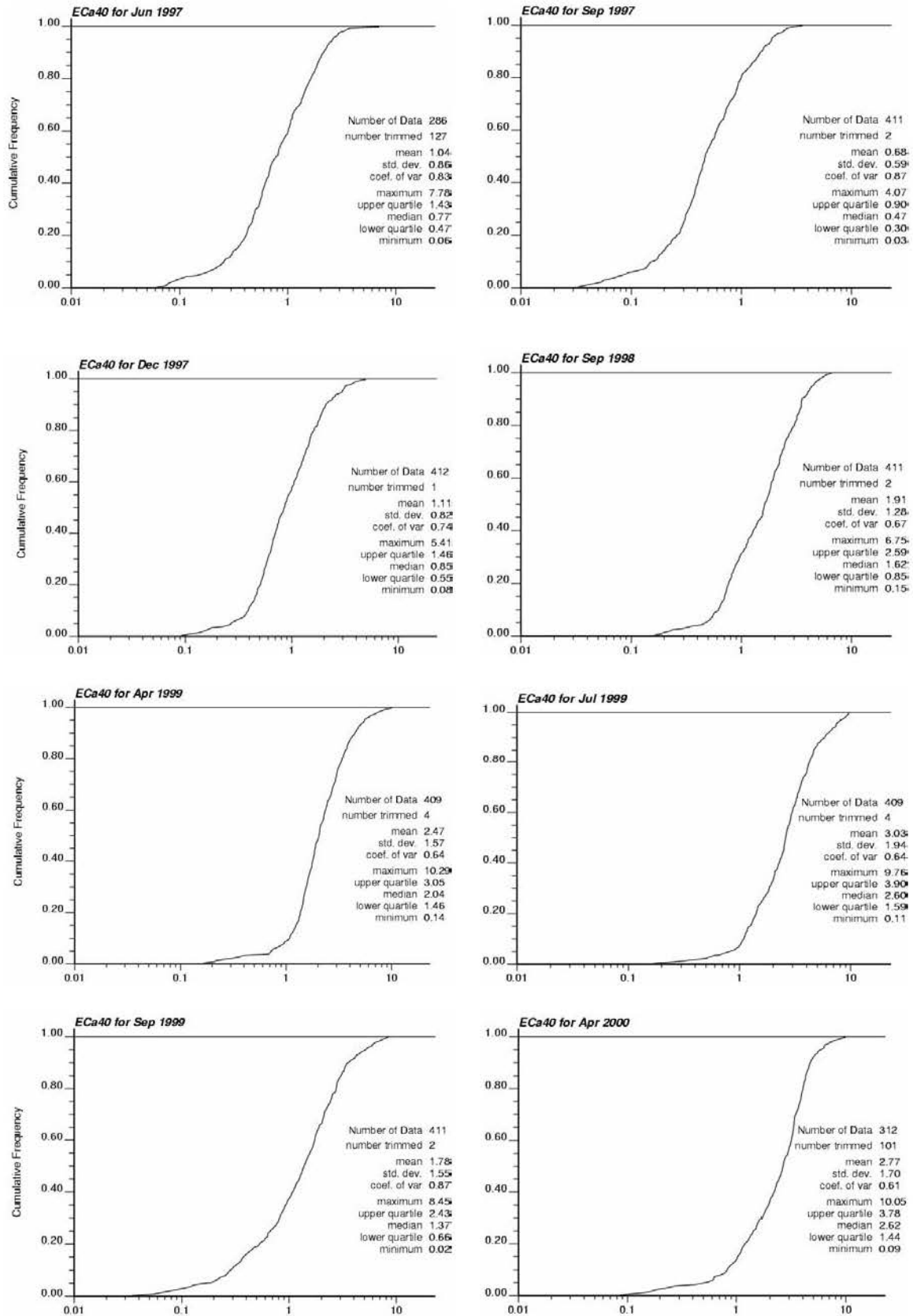
Fig. 7.4. Cumulative distribution functions of EC_a (continued).
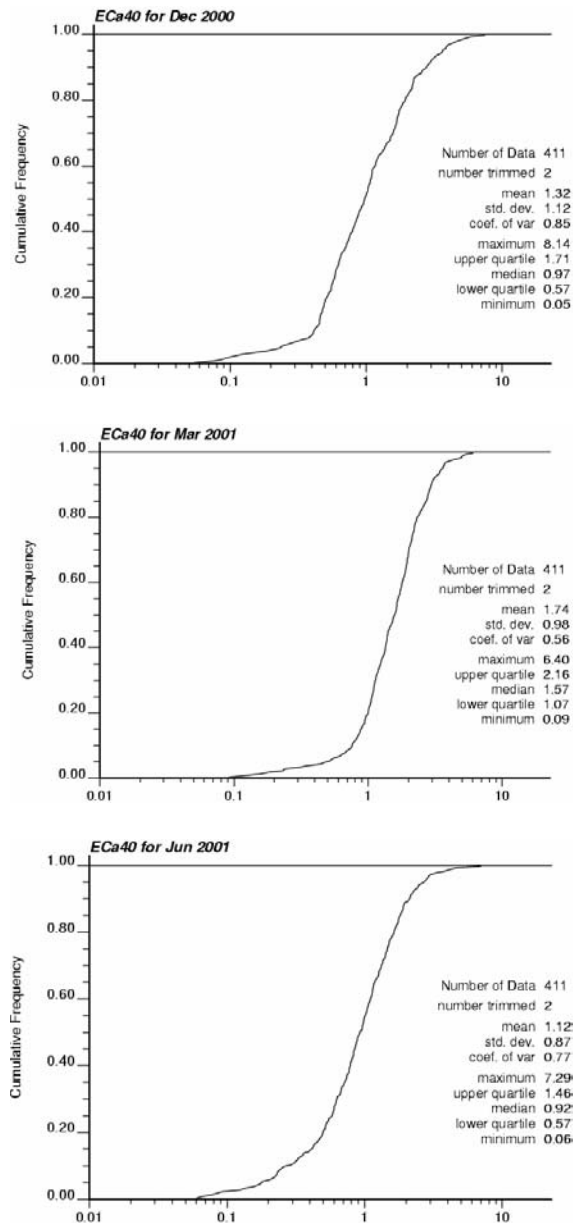
Fig. 7.4. Cumulative distribution functions of EC_a (continued and end).

7.4.2. Calibration Equations

The OLS residuals showed no significant spatial dependence. This result was confirmed by the maximum likelihood-based tests of the non-appropriateness of an additional spatial parameter in the spatial regression models as well as the Moran's *I* test. The results of the latter are given in table 7.2. When the standardized *I* statistic is larger than 1.96 or, equivalently, when its corresponding probability is lesser than 5 %, the hypothesis of no spatial correlation is rejected. In our case, this hypothesis was rejected only 3 out of the 19 time instants. However, when we fitted a first autoregressive model to the full EC_a survey

from the 'data set to be calibrated' (EC_a regressed on its neighbours), we found a significant spatial dependence. The absence of spatial autocorrelation in the $EC_{2.5}$- EC_a relationship may be due to the fact that we have very few locations (13 to 20) which are far apart comparatively to the EC_a data (286 to 413). Consequently, we adopted the classical OLS regression model to relate $EC_{2.5}$ to EC_a. The scatterplots of these two variables, one for each time instant, are shown in Fig. 7.5 for data on original scale and in Fig. 7.6 for log transformed data.

Table 7.2. Moran's I test for spatial independence of residuals from OLS regression linking $EC_{2.5}$ to EC_a. Moran's I : statistic of Moran, E(I) : its expected value, Var(I) : its variance, Std(I) : standardized Moran's I, Proba : probability of significance of Std(I). Bold values for time instant for which the assumption of no spatial correlation was rejected.

Time	Moran's I	E(I)	Var(I)	Std(I)	Proba
Nov-94	0.34	-0.06	0.03	2.50	**0.02**
Mar-95	0.23	-0.04	0.02	1.82	0.08
Jun-95	0.16	-0.04	0.02	1.36	0.16
Sep-95	0.21	-0.04	0.02	1.64	0.10
Dec-95	-0.01	-0.04	0.02	0.21	0.39
Mar-96	0.15	-0.05	0.04	1.04	0.23
Jun-96	0.22	-0.03	0.02	1.72	0.09
Mar-97	0.31	-0.04	0.02	2.36	**0.02**
Jun-97	0.11	-0.05	0.02	1.13	0.21
Sep-97	0.11	-0.03	0.02	0.97	0.25
Dec-97	-0.10	-0.04	0.02	-0.40	0.37
Sep-98	-0.18	-0.04	0.02	-0.89	0.27
Apr-99	-0.19	-0.04	0.02	-1.02	0.24
Jul-99	-0.20	-0.05	0.03	-0.93	0.26
Sep-99	0.54	-0.04	0.02	3.78	**0.00**
Apr-00	0.00	-0.04	0.03	0.29	0.38
Dec-00	-0.05	-0.03	0.02	-0.11	0.40
Mar-01	-0.11	-0.04	0.02	-0.43	0.36
Jun-01	-0.23	-0.03	0.02	-1.34	0.16
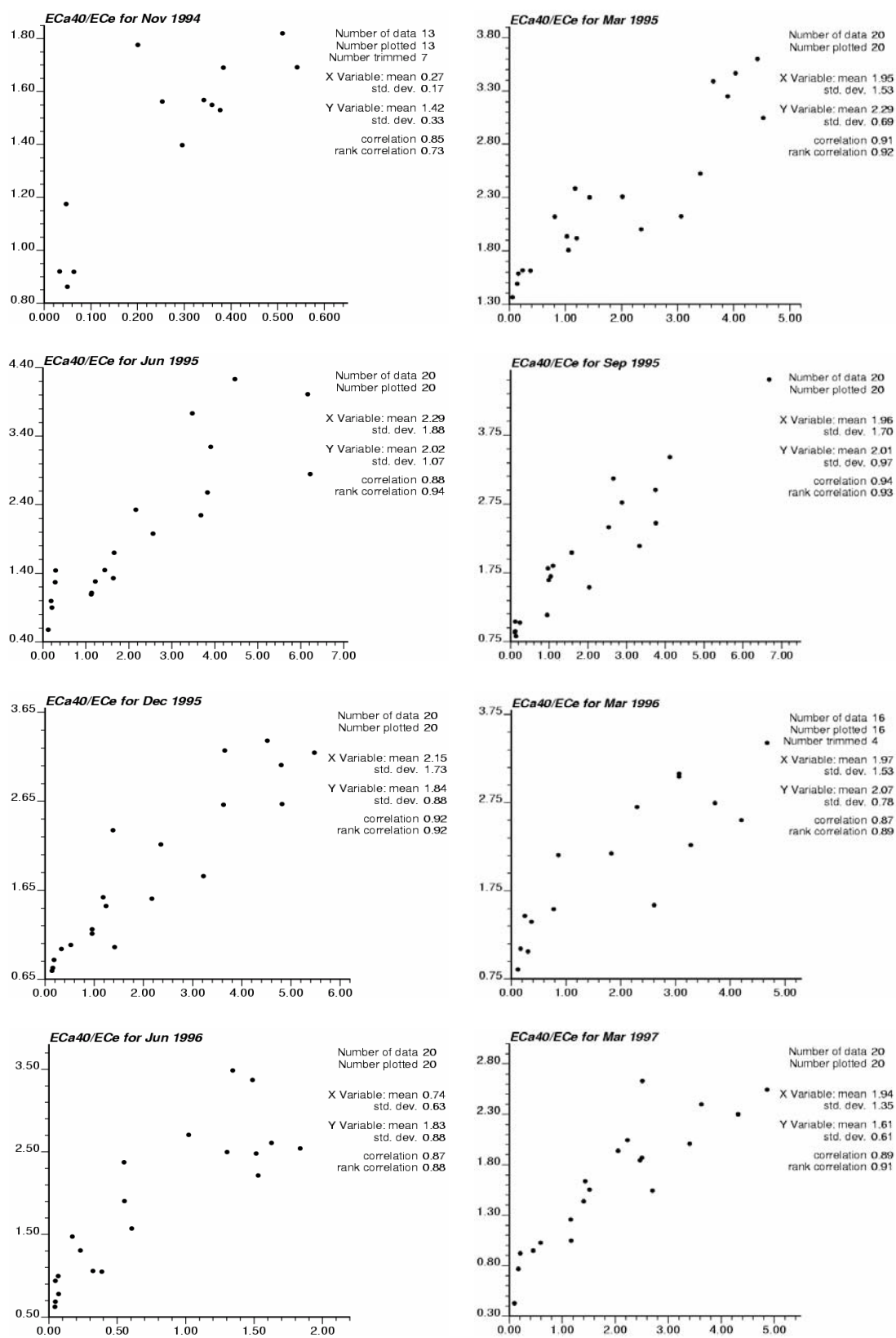
Fig. 7.5. Scatterplot of EC$_{2.5}$ as function of EC$_a$ on the original scale.
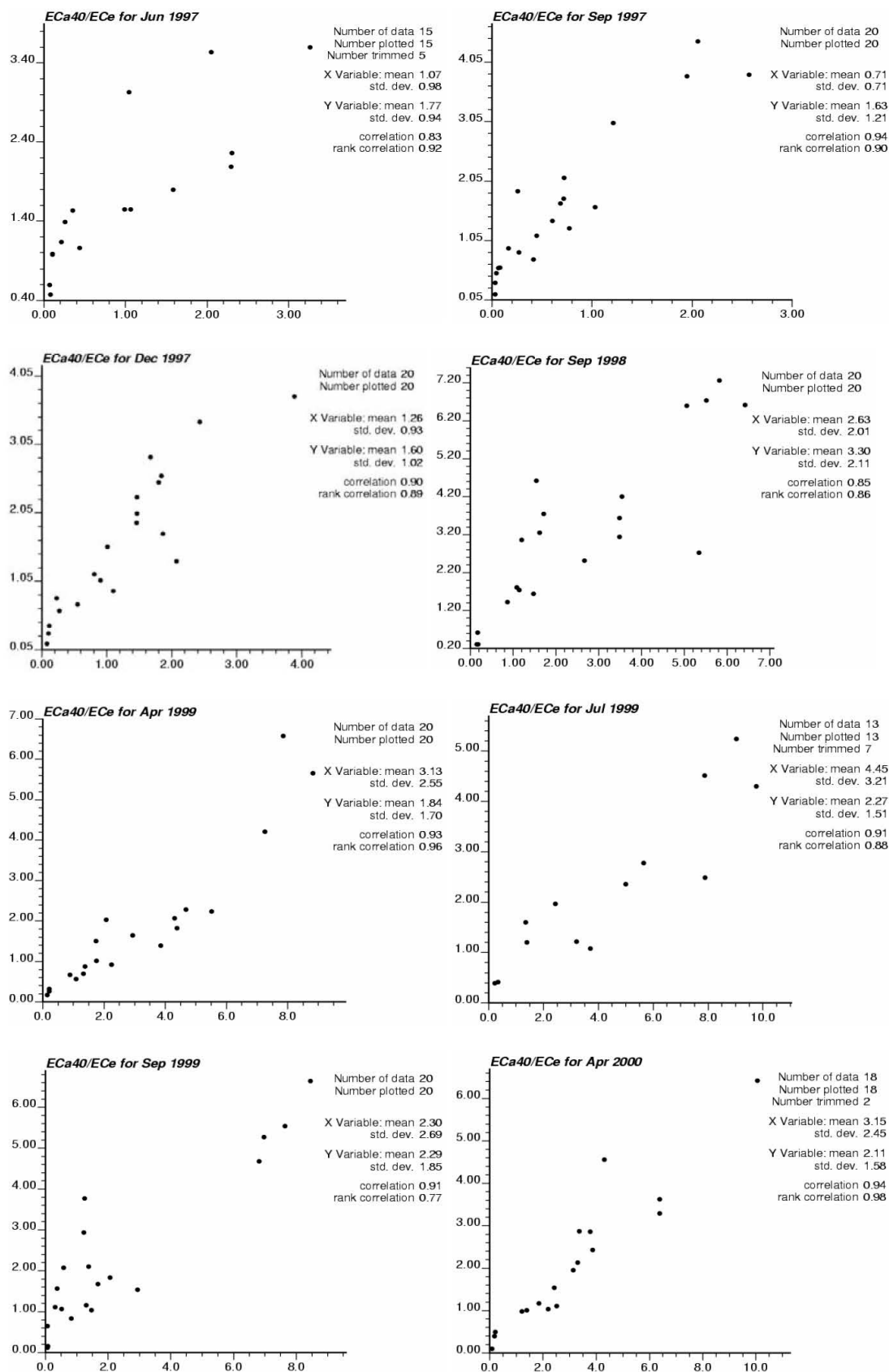
Fig. 7.5. Scatterplot of $EC_{2.5}$ as function of EC_a on the original scale (continued).
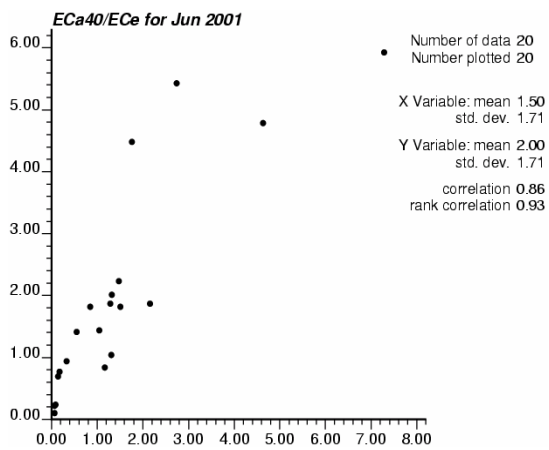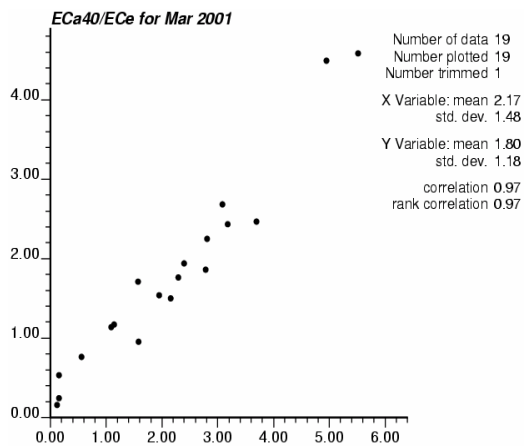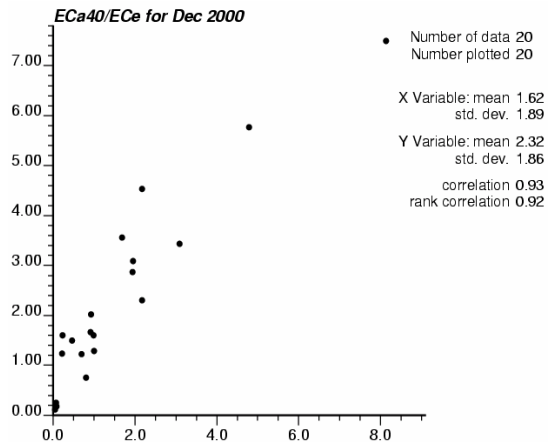
Fig. 7.5. Scatterplot of EC$_{2.5}$ as function of EC$_a$ on the original scale (continued).
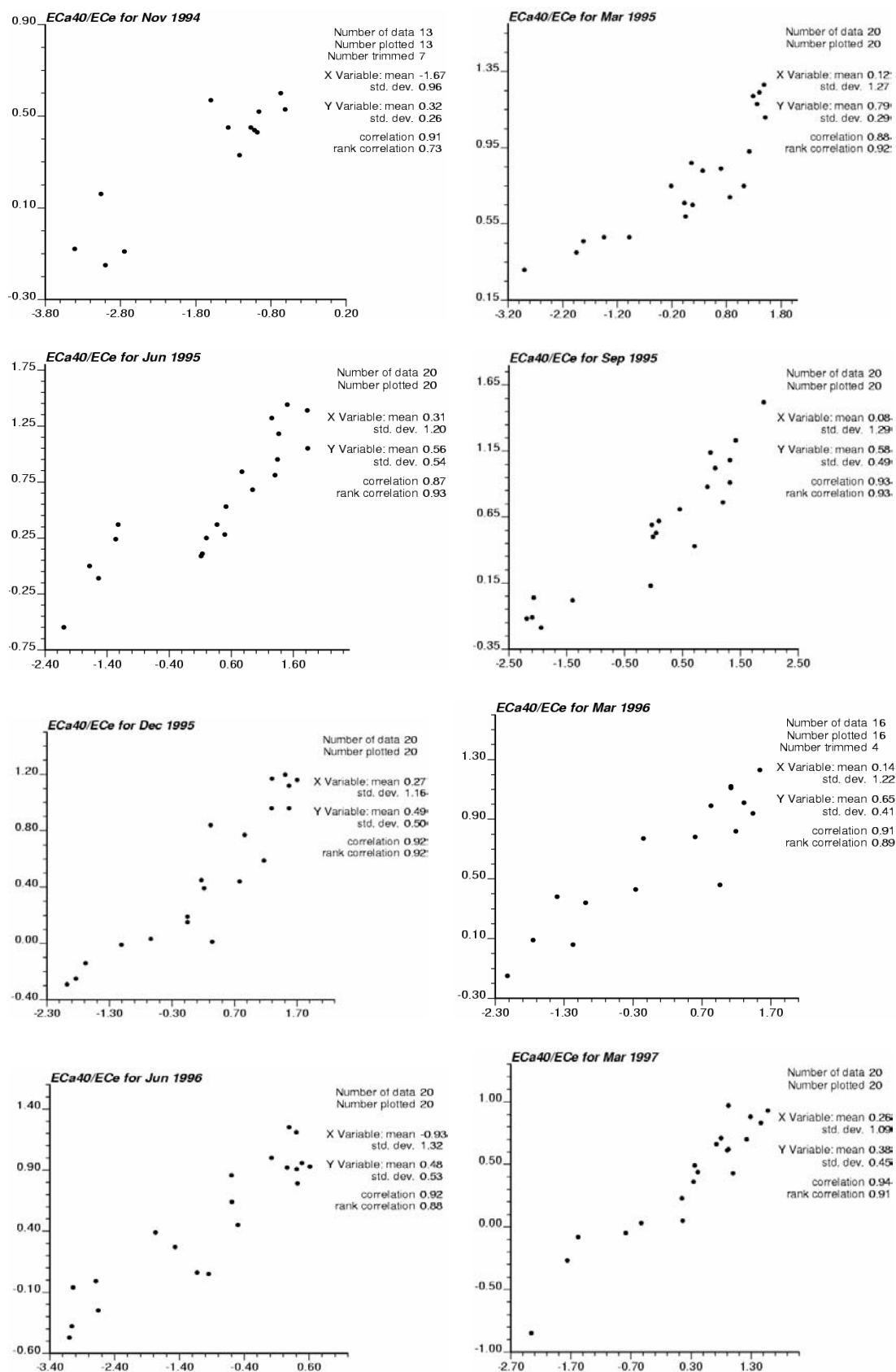
Fig. 7.6. Scatterplot of EC$_{2.5}$ as function of EC$_a$, both variables log transformed.
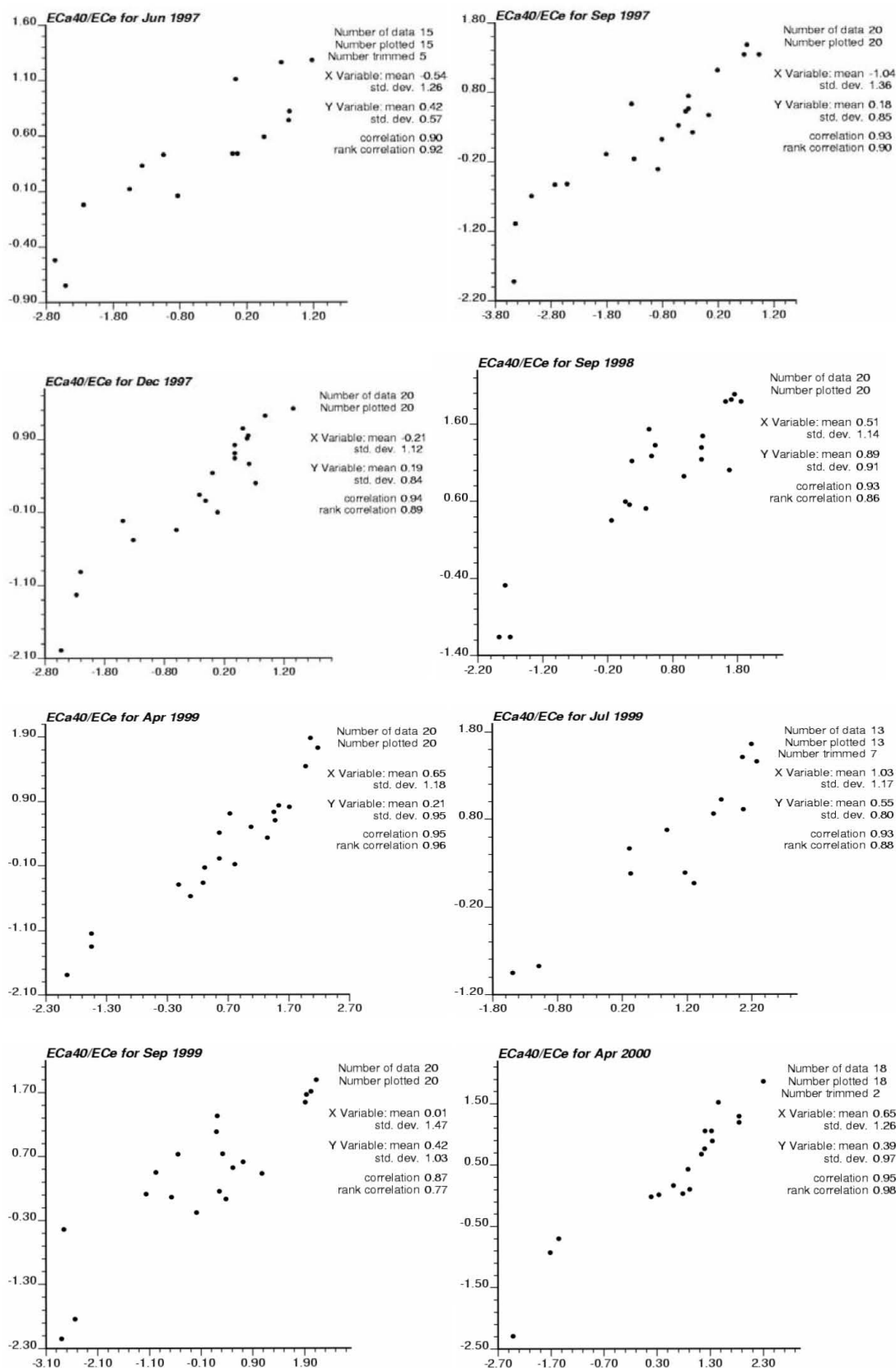
Fig. 7.6. Scatterplot of $EC_{2.5}$ as function of EC_a (continued).
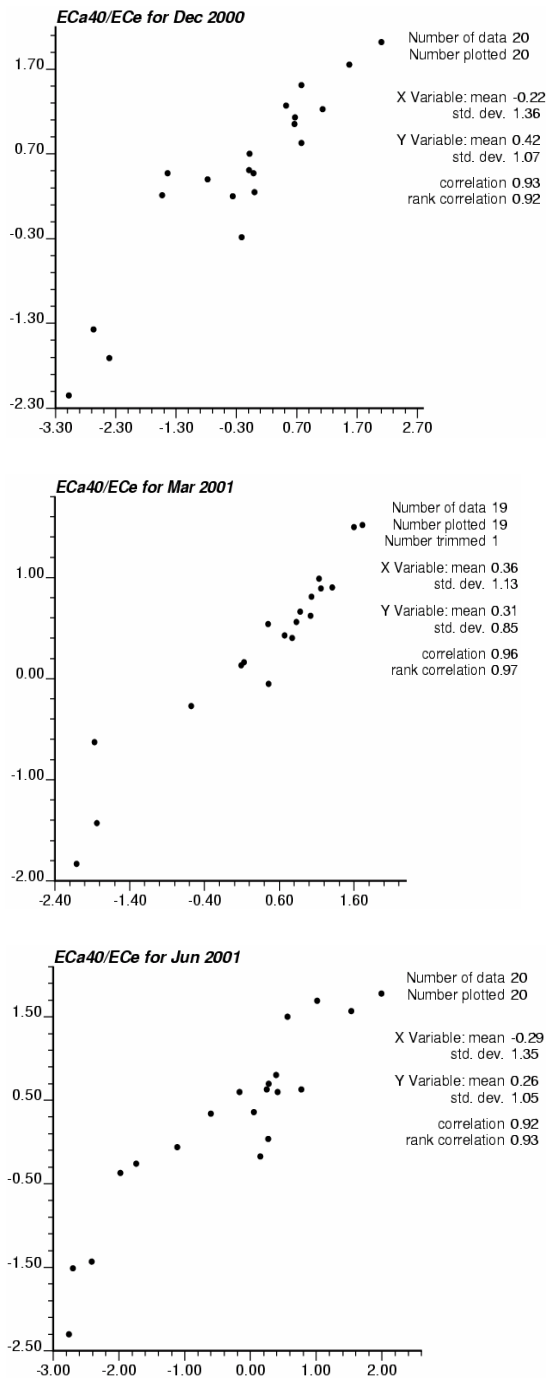
Fig. 7.6. Scatterplot of $EC_{2.5}$ as function of EC_a (continued and end).

It can be inferred from figures 7.5 and 7.6 that the relationship between $EC_{2.5}$ and EC_a can be considered as linear. Also, this relationship is very strong. Most of the correlation coefficients were higher than 0.85 (for all except one of the 19 time frames) with a maximum value of 0.97. The logarithmic transformation required to obtain normal or, at least, less skewed distributions improved the correlation between both variables. This coefficient was increased for 11 out of the 19 time instants by as much as 0.08.

Different models were tried by adding other covariates than EC_a, for example the coordinates and the vegetal coverage. The best model, chosen among all possible combinations between the covariables (EC_a, x and y coordinates and the vegetal cover), which had the highest adjusted coefficient of determination, the lowest mean square error and all its coefficients being significant was the following:

$$\ln(EC_{2.5}) = b_0 + b_1 \ln(EC_a) + b_2 u + b_3 u^2 + b_4 v + b_{5j} cover_j$$

$$u = \frac{(x - m_x)}{s_x}$$ with m_x and s_x: mean and standard deviation of the x coordinate

$$v = \frac{(y - m_y)}{s_y}$$ with m_y and s_y: mean and standard deviation of the y coordinate

$cover_j$, j=1,…,4, represents the 4 categories of vegetal coverage.

We fitted this model at each time instant separately, so finally we obtained 17 equations corresponding to the 17 time instants.

7.4.3. Descriptive Statistics

The main statistic parameters of the predicted $EC_{2.5}$ data are summarized in table 7.3. The mean $EC_{2.5}$ varied between 1.39 (November 1994) and 2.74 dS.m^{-1} (September 1998). The minimum is enclosed between 0.06 (December 2000) and 0.68 (March 1996) and maximum varying between 2.59 (November 1994) and 9.41 (December 2000). The data are moderately to highly variable with coefficients of variation ranging between 0.28 (March 1997) and 0.64 (December 1997 and 2000).

For all the time instants, the range in salinity values was large in comparison to the mean indicating that soil salinity is highly variable in space. Moreover, the differences in the statistic parameters (mean, median and range) between time instants are an indication of a temporal variation.

Table 7.3. Statistic parameters of predicted salinity data ($EC_{2.5}$ in dS m^{-1}). N: number of observations, cv: coefficient of variation, min: minimum, med: median, max: maximum.

$EC_{2.5}$	N	mean	cv	min	med	max
Nov-94	411	1.39	0.37	0.45	1.38	2.59
Mar-95	411	2.03	0.32	0.60	1.94	4.78
Jun-95	412	1.74	0.39	0.48	1.62	4.66
Sep-95	410	1.77	0.33	0.54	1.70	3.66
Dec-95	413	1.65	0.38	0.53	1.53	3.93
Mar-96	392	1.96	0.29	0.68	1.86	3.30
Jun-96	411	1.54	0.42	0.31	1.41	4.52
Mar-97	310	1.48	0.28	0.42	1.43	2.91
Jun-97	286	1.69	0.59	0.24	1.48	8.33
Sep-97	411	1.50	0.58	0.13	1.32	5.83
Dec-97	412	1.44	0.64	0.13	1.18	6.99
Sep-98	411	2.74	0.55	0.31	2.47	8.35
Apr-99	409	1.43	0.63	0.17	1.19	6.90
Jul-99	409	1.96	0.50	0.16	1.81	5.85
Sep-99	411	1.93	0.63	0.11	1.58	6.39
Apr-00	312	1.78	0.58	0.09	1.57	7.07
Dec-00	411	2.10	0.64	0.06	1.76	9.41
Overall	6640	1.78	0.54	0.06	1.59	9.41

7.4.4. Covariography

The purely spatial and purely temporal experimental covariance functions as well as the theoretical models fitted to the observed $EC_{2.5}$ data are shown in Fig. 7.7. Although, the spatial covariance tended towards zero as the spatial lag tended towards infinity, the temporal covariance function did not when the temporal lag tended towards infinity. This represents the existence of a trend in the data, which need to be detrended before their use for structural analysis. For illustration, the mean temporal trend is given in Fig. 7.8 for four different space locations. The observed $EC_{2.5}$ (in fact, the predicted $EC_{2.5}$ using the adopted multiple

regression models) as well as the mean spatial trend and the resulting residual $EC_{2.5}$ are shown in Fig. 7.9 for two time instants.
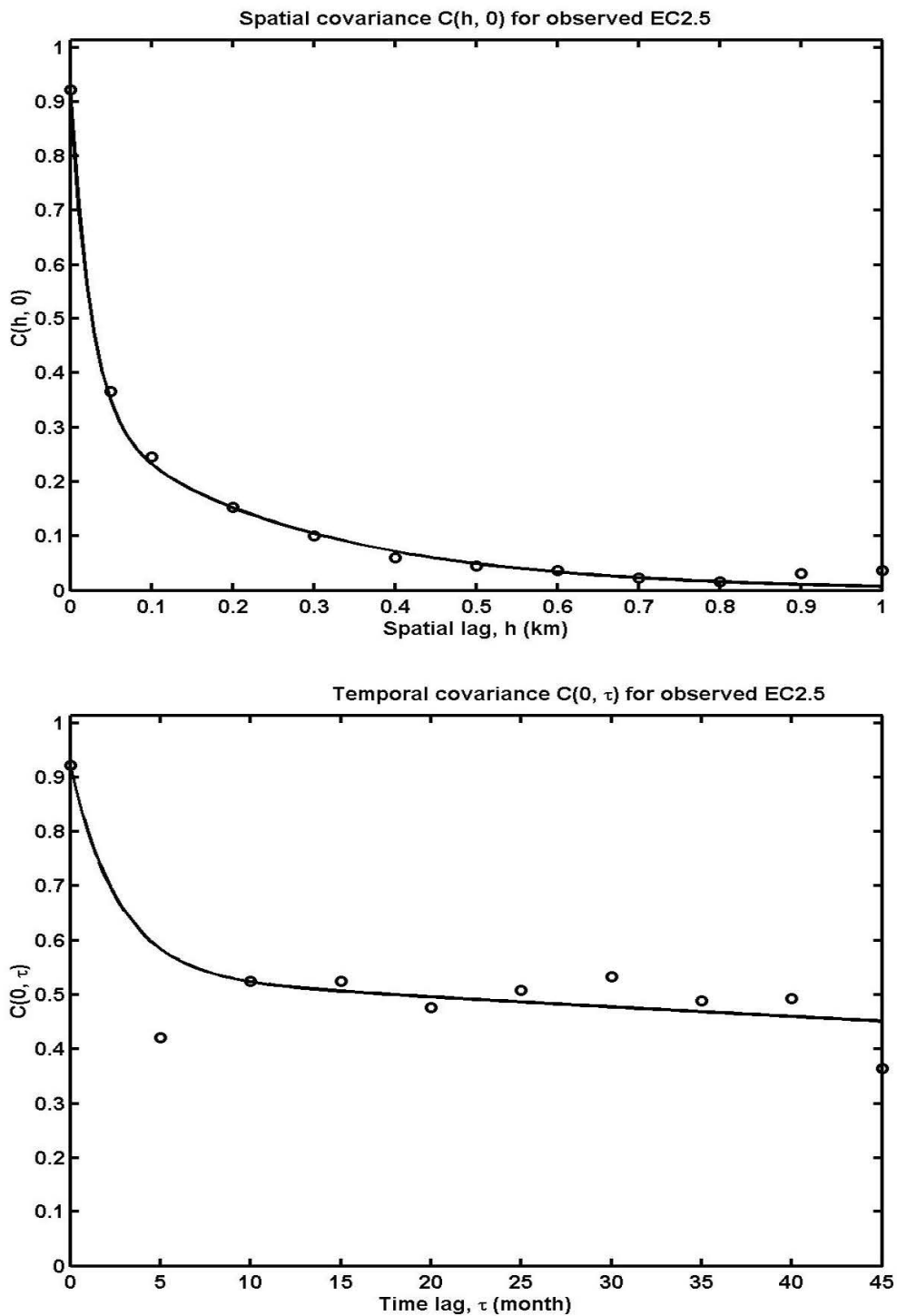
Fig. 7.7. Purely spatial and purely temporal covariance functions for the observed $EC_{2.5}$ data. Circles: experimental values and solid lines: fitted theoretical models.
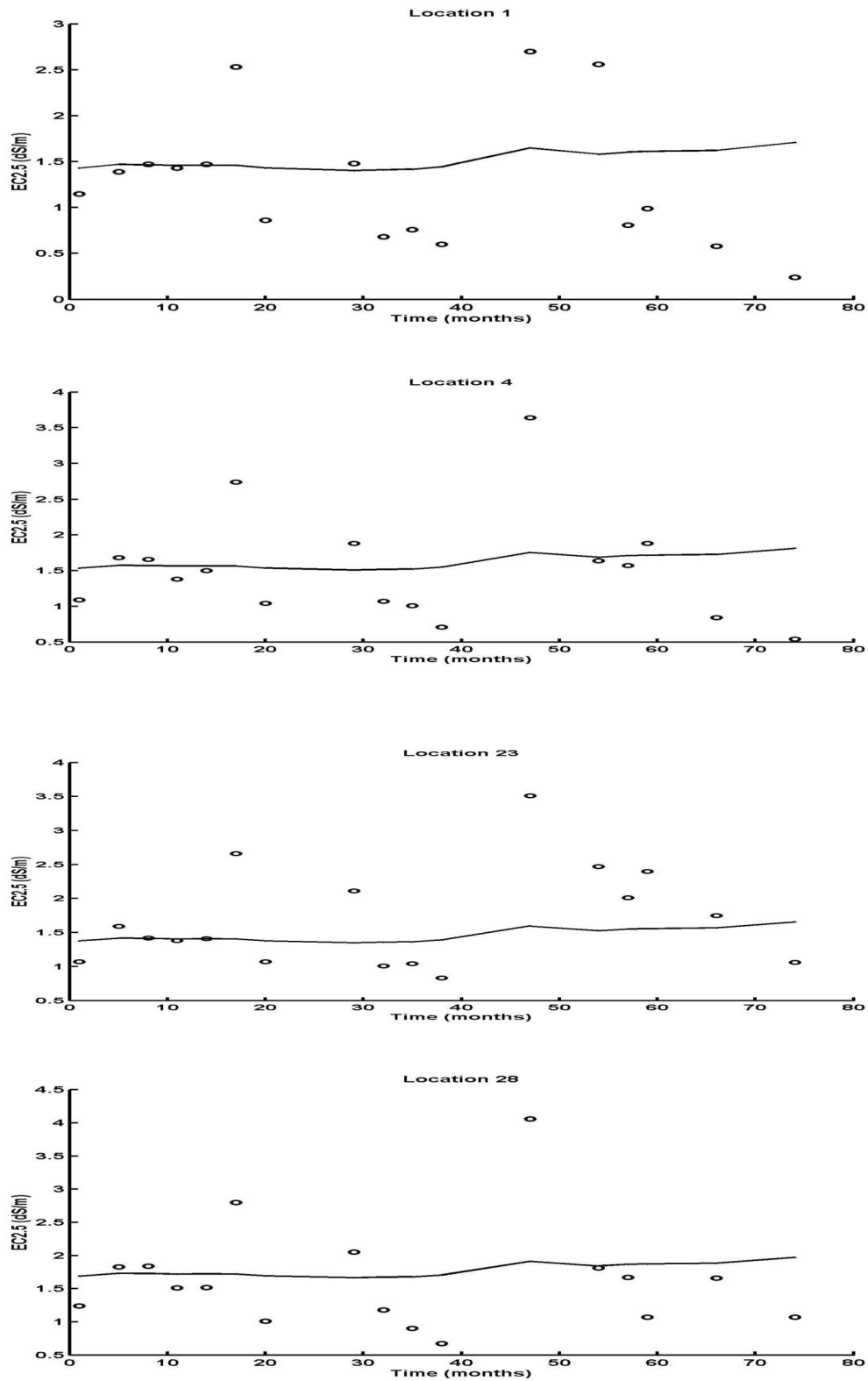
Fig. 7.8. Temporal mean trend for some space locations. Circles: observed EC$_{2.5}$, solid line: mean trend.
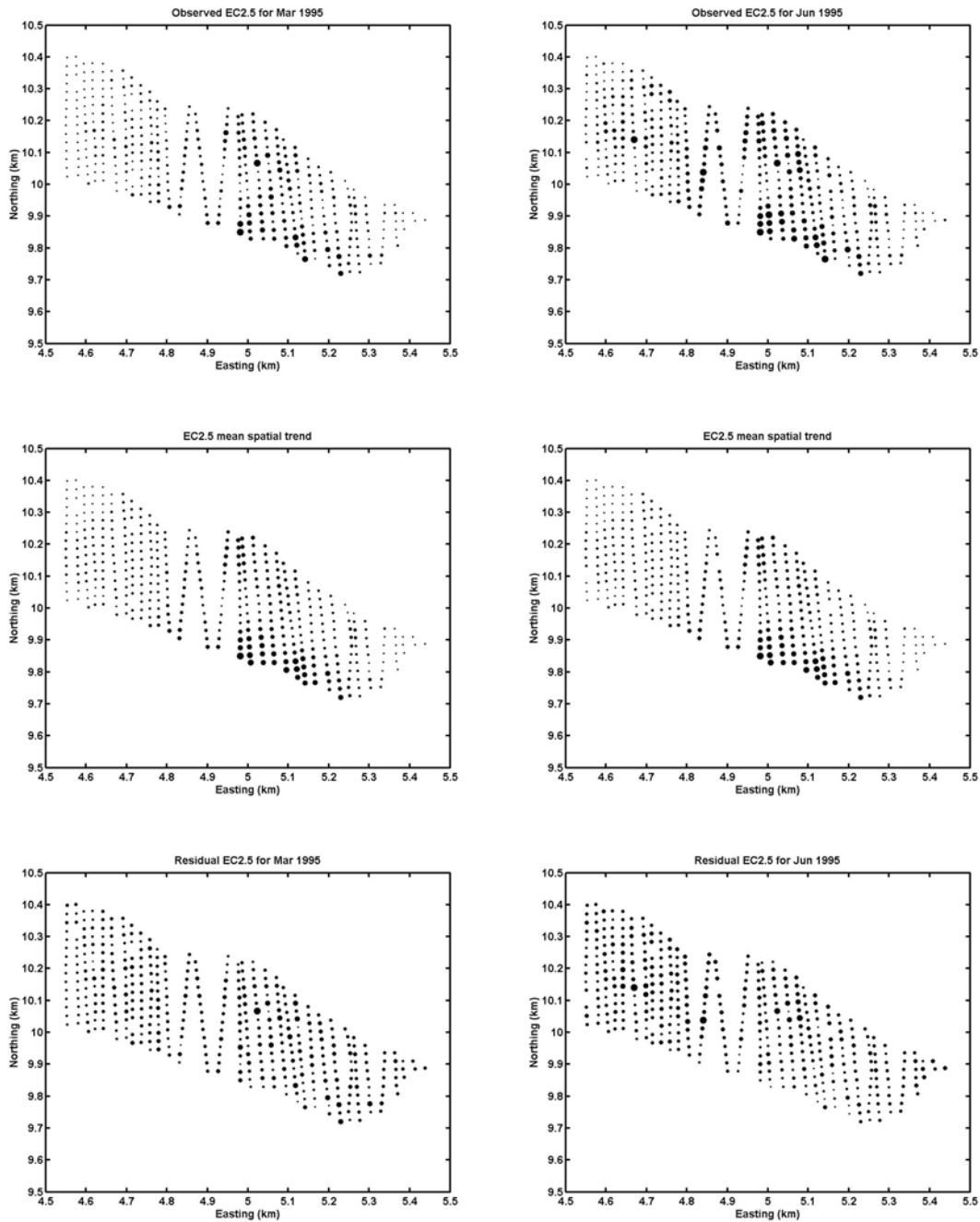
Fig. 7.9. Observed EC$_{2.5}$, mean spatial trend and residual EC$_{2.5}$ for time instants 3 and 4 (March and June 1995). The size of the points is function of the EC$_{2.5}$ values.

The spatial, temporal and spatio-temporal dependencies in the salinity data were described and modelled using covariance functions after the removal of the mean space-time trend. The spatial covariance function was fitted with a nested exponential model as is illustrated in Fig. 7.10 (a). The small-scale range is about 250 m with a sill equal to 0.27 (dS m^{-1})^2, which

represents 79.4% of the total variance and the large-scale range is beyond the dimensions of the study area (1500 m):

$$c(\boldsymbol{h},0) = c_{01} \exp(-3\boldsymbol{h}/as_1) + c_{02} \exp(-3\boldsymbol{h}/as_2)$$

with c_{01} and c_{02} the sills of the nested models and as_1 and as_2 their corresponding ranges.
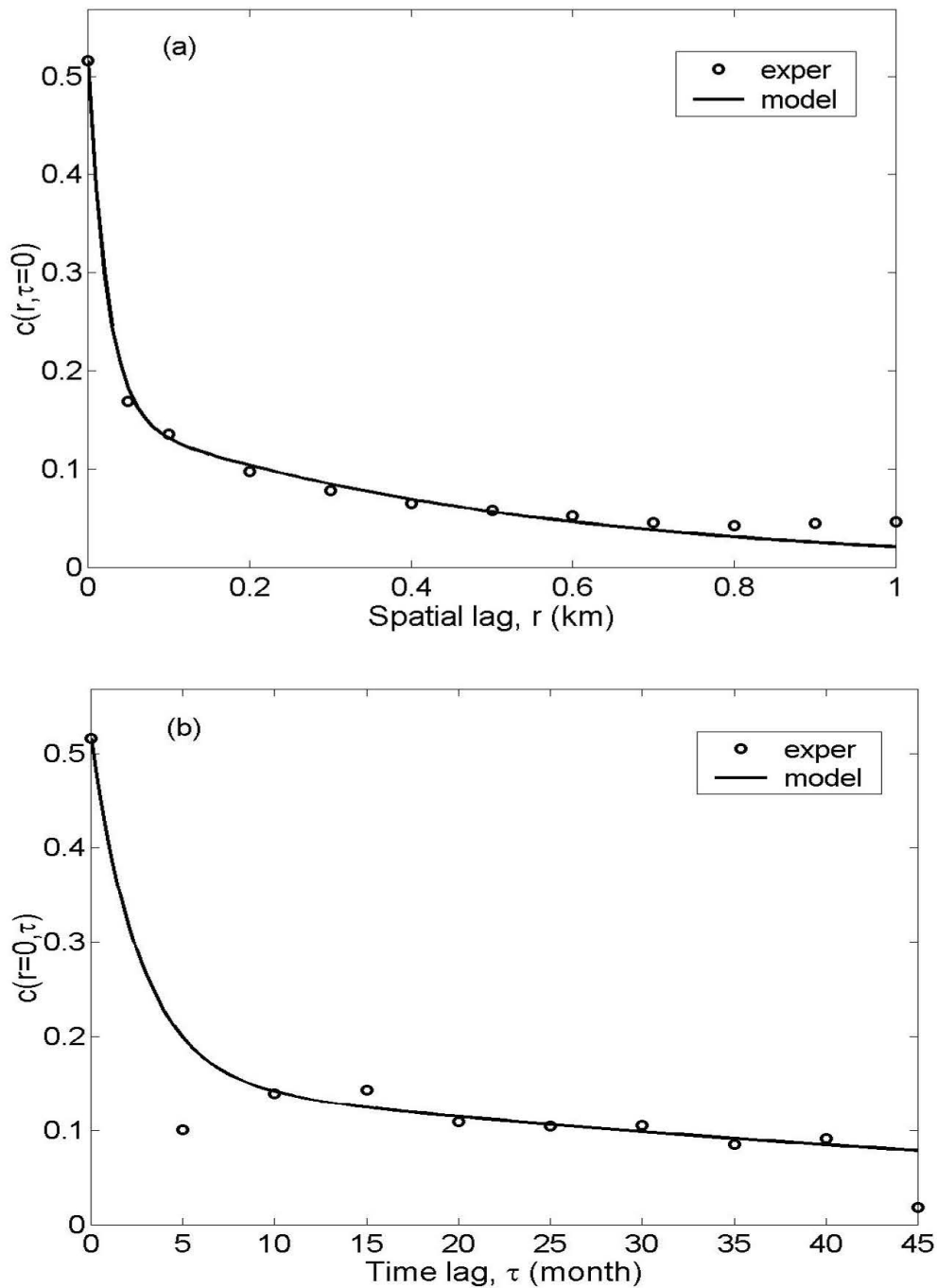
Figure 7.10. (a): Spatial covariance function; (b): temporal covariance function. Circles: experimental covariance function; curve: fitted model.

The same nested model was used to fit the temporal covariance function, Fig. 7.10 (b), with a small-scale range of 8 months and a large-scale range far beyond the time period covered (200 months):

$$c(\mathbf{0}, \tau) = c_{01} \exp(-3\tau / at_1) + c_{02} \exp(-3\tau / at_2)$$

with at_1 and at_2 the small-scale and large-scale ranges

The space-time covariance function (Fig. 7.11) is a nested structure of two space-time separable covariance models:

$$c(\mathbf{h}, \tau) = c_{01} \exp(-3\mathbf{h} / as_1) \exp(-3\tau / at_1) + c_{02} \exp(-3\mathbf{h} / as_2) \exp(-3\tau / at_2)$$

The non-separable space-time covariance function of Fig. 7.11 provides a more accurate representation of the correlation structure of salinity in both space and time than that described by a purely spatial covariance model, or a covariance model where time is taken as an additional spatial coordinate. Its parameters are summarized in table 7.4.
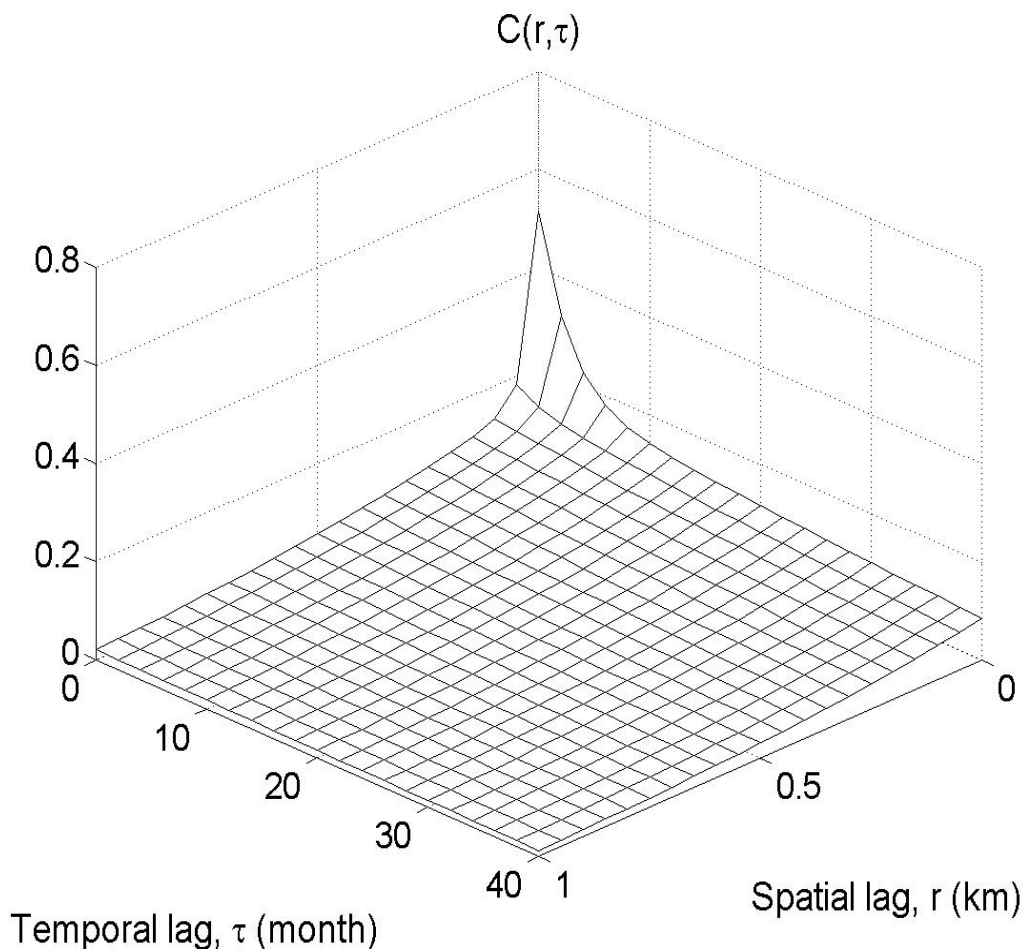
Figure 7.11. Space-time covariance function of the residual data R(s, t).

Table 7.4. Parameters of the fitted space-time covariance model.

Component	Spatial range (m)	Temporal range (month)	Sill (dS m^{-1})^2
First nested model	250	8	0.27
Second nested model	1500	200	0.07

7.4.5. Space-time Kriging

As the spatio-temporal dependence of $EC_{2.5}$ was modelled using a nested separable space-time covariance function, it was used to estimate soil salinity at any location in space and any instant in time by defining a search neighbourhood.

For illustration purposes, we show only results for the most frequently observed month (September of the years 1995, 1997, 1998, and 1999). We estimated on a dense spatial grid including the 413 locations for which we have the observed $EC_{2.5}$ values for September from 1995 to 2000 (Fig. 7.12).

First, we note that for the non-observed time instants (September 1996 and 2000), the smoothing effect is stronger than for the observed time instants (September 1995, 1997, 1998, and 1999). This is due mainly to the fact that for the latter ones, the neighbours come mostly from the simultaneous time instant but for the former ones the neighbours are from different time instants. Also, there is a net general increase in soil salinity from September 1995 to 1999.

To check the contribution of the additional temporal dependence, we compared the space-time kriging to a simple spatial kriging by modeling independently and separately the spatial dependence for each time instant. Sterk and Stein (1997) computed a single spatial variogram pooling the data of the 4 time instants that they had. This was required to circumvent the lack of sufficient observations (100 or more as reported by Webster and Oliver, 1992) to compute a reliable variogram. Ettema et al. (2000) adopted the same procedure. As we had sufficient observations at each time instant (at least 286), we computed the spatial variograms separately for each time instant. The results of the comparison are reported in Fig. 7.13 for the histograms of the estimated $EC_{2.5}$ values and in Fig. 7.14 for those of their corresponding estimation errors. The estimated values are more or less the same but it is clear that the spatio-temporal estimates are more precise comparatively to the spatial estimates. Ettema et al.

(2000) reached the same conclusions in their study of the spatio-temporal patchiness of nematode species.

7.5. Conclusions

The space-time kriging estimates were more precise than the estimates obtained using only the spatial component of the soil salinity dependence (the most frequent estimation error is bigger for the latter than for the former). Also, the smoothing effect seems to be more pronounced in the case of the spatial kriging than in the space-time kriging (the extreme values are lesser for the former than for the latter). These conclusions were deduced from the graphic representation of the estimates and their estimation errors for the 2 approaches. For a more formal comparison, it would be better to use some quantitative criteria. So in this sense, it may be suitable to leave some locations for a validation data set that will be used in the computation of, for example, the mean error or the mean square error. Another possible improvement is to fit the product-sum model of De Cesare et al. (2001a) to the experimental covariance function.
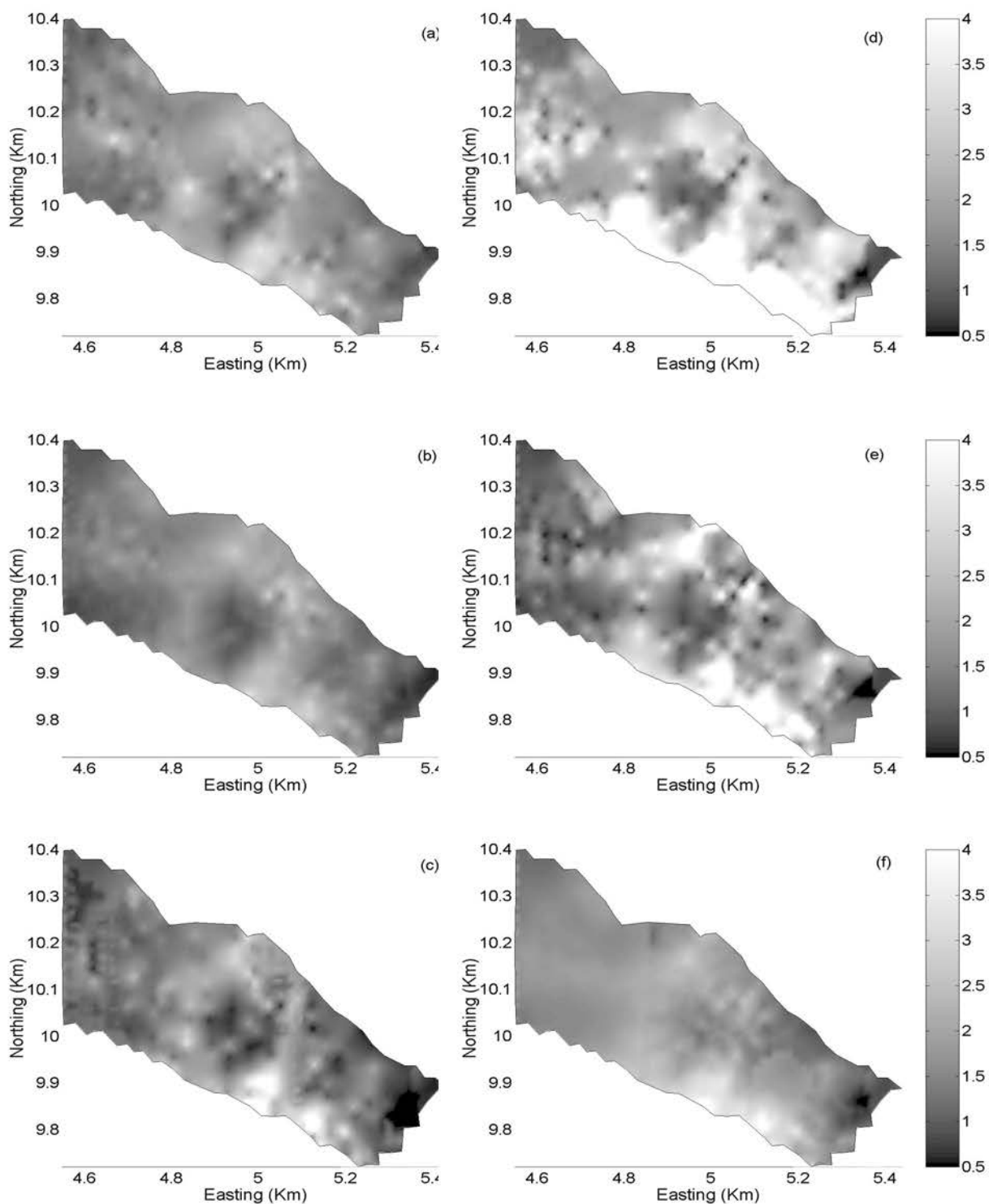
Figure 7.12. EC$_{2.5}$ estimates (dS m^{-1}) using the space-time covariance models for each September between 1995 (a) and 2000 (f).
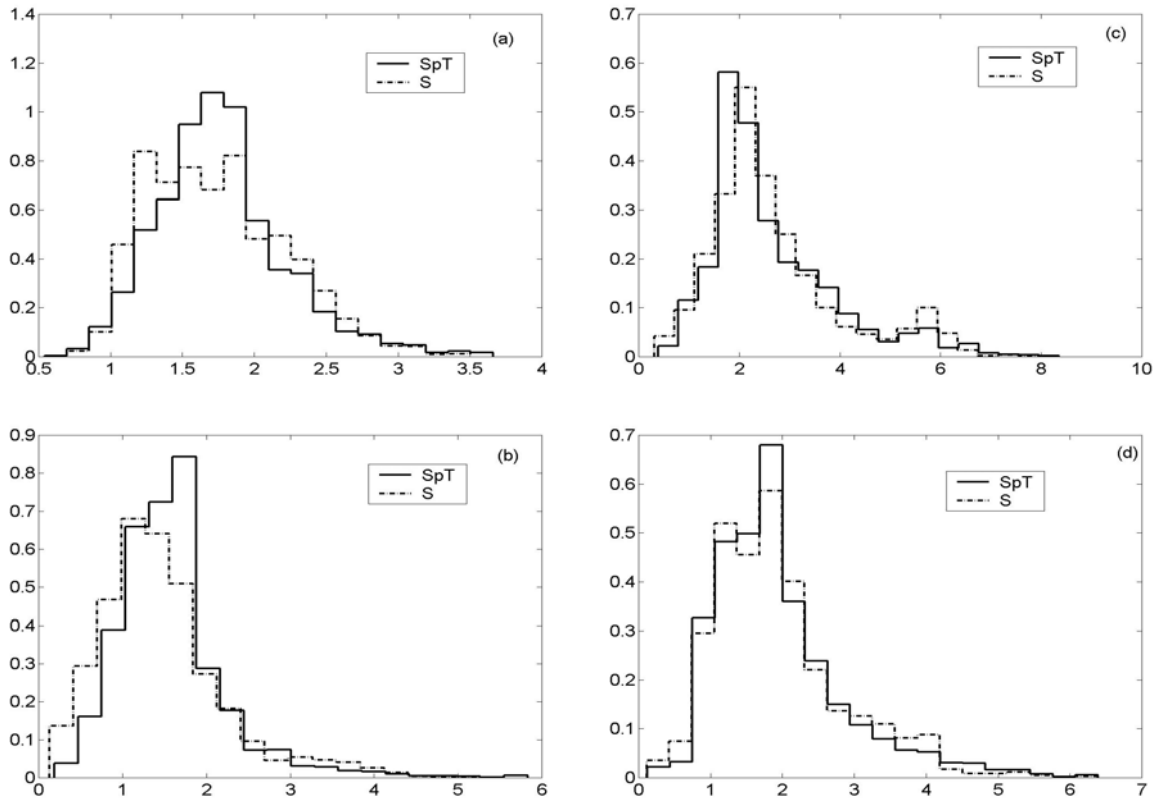
Figure 7.13. Spatial (S) and spatio-temporal (SpT) estimates of EC$_{2.5}$ (dS m^{-1}) for September 1995 (a), 1997 (b), 1998 (c), and 1999 (d).

Figure 7.14. Spatial (S) and spatio-temporal (SpT) estimation errors of EC$_{2.5}$ (dS m^{-1}) for September 1995 (a), 1997 (b), 1998 (c), and 1999 (d).

148

CHAPTER 8

SOIL SALINITY MAPPING USING SPATIO-TEMPORAL KRIGING AND BAYESIAN MAXIMUM ENTROPY

8.1. Introduction

Soil salinity limits food production in many countries of the world. There are mainly two kinds of soil salinity: naturally occurring dryland salinity and human-induced salinity caused by the low quality of water. In both cases the development of plants and soil organisms are limited leading to low yields. In Hungary, where more than 10% of the land is affected by salt, groundwater is the major cause of salinization.

Saline and sodic soils have particular physical and chemical properties that require specific management. As a first step for the better management of salt-affected soils, soil salinity needs to be monitored in space as well as in time to determine where it is, where it is spreading to, and the rate at which it is spreading. Therefore, we need to sample the soil for laboratory analysis to determine the electrical conductivity of the saturated soil paste extract (EC_e). The latter is a measure of soil salinity. This conventional procedure (Soil and Plant Analysis Council, 1992) is expensive, time consuming, and provides an incomplete view of the extent of soil salinity.

An alternative to laboratory analysis is to assess soil salinity in the field by determining the apparent electrical conductivity (EC_a). This can be done using sensors such as the four-electrode probes (Rhoades and van Schilfgaarde, 1976) or by electromagnetic induction instruments (McNeil, 1980). This procedure is cheaper and less time-consuming than the conventional one, and the sensors can be mounted on a small vehicle enabling a more intensive survey of the study area.

The appraisal of space-time variability of soil salinity has been approached in different ways. Lesch et al. (1998) used a classical statistical method to monitor the temporal change of soil salinity between two time periods. The approach can be applied easily to a few measurement times. However it is of limited practical use for many time periods as the procedure must be repeated for each pair of times. In addition, the technique takes no account of any possible temporal correlation between two or more successive measurements.

Douaik et al. (2004) proposed an alternative approach. They rescaled the EC_a measurements into $EC_{2.5}$ (the electrical conductivity determined by laboratory analysis from 1:2.5 soil-water suspensions, which is a simple representation of the electrical conductivity of the water-saturated soil-paste extract, EC_e) using calibration equations based on regression models. This was followed by spatio-temporal kriging to predict soil salinity at unknown places and times. The approach takes into account the spatial and the temporal correlations between the soil

salinity measurements. However the resulting $EC_{2.5}$ values from the calibration equations are estimates of the actual soil salinity. This means that they have some degree of uncertainty which needs to be considered in the analysis.

The method of Bayesian maximum entropy (BME) (Christakos, 1990; 2000) enabled a rigorous analysis of our data by distinguishing formally between the accuracy of the laboratory and the field electrical conductivity measurements. The former are direct and accurate measurements of the soil salinity; they are considered as hard data. The latter are indirect measurements that represent uncertain estimates of soil salinity. They provide less accurate values of soil salinity and can be considered as soft data since they are determined, not only by the soil salinity ($EC_{2.5}$), but also by the soil moisture content, temperature, particle size distribution, etc.

Bayesian maximum entropy provides a general framework for space-time interpolation. It can incorporate different physical knowledge bases such as statistical moments (not limited to the second-order), multipoint statistics, physical laws, hard and soft data, etc. Kriging, the classical geostatistical method of interpolation, is a special case of BME. When physical knowledge is restricted to the second-order statistical moments (mean and covariance or variogram functions) and to the hard data, kriged and BME predictions are equivalent (Christakos and Li, 1998; Lee and Ellis, 1997).

BME has been successfully applied in different areas. D'Or and Bogaert (2001), D'Or et al. (2001), Bogaert and D'Or (2002), and D'Or and Bogaert (2003) mapped soil texture using BME in the space domain. Serre and Christakos (1999a) studied the water-table elevations of an aquifer in Kansas while Christakos and Serre (2000) analyzed the distribution of particulate matter in North Carolina. These two studies have been done in the space-time context. Bogaert (2002) extended the approach to include categorical variables.

This study has three main objectives: (i) to apply the BME method to soil salinity data set using interval and probabilistic soft data, (ii) to compare the prediction performance of BME with two types of kriging: ordinary kriging with hard data only (HK), and ordinary kriging with hard data and the mid-interval of the soft data (HSK), and (iii) to determine the probability, based on BME and kriging, that soil can be considered saline in the study site.

8.2 Data description

As in the precedent chapter, both data sets were used (see chapter 3, section 3.2.1 and chapter 7, section 7.2). The calibration data sets were used to determine the relationships between $EC_{2.5}$ and EC_a. These relationships were used to convert EC_a data from data to be calibrated sets' into $EC_{2.5}$. The measured $EC_{2.5}$ represents the hard data whereas the predicted $EC_{2.5}$ represents the soft data.

8.3. Methods

8.3.1. Data analysis

The adopted methodology to determine the mean trend and the covariance function were presented in the precedent chapter (section 7.3). They were used as general knowledge for both space-time interpolation methods presented in this study.

In our case study, the observed $EC_{2.5}$ are the hard data available at up to $h = 20$ locations in space. Two kinds of soft data were used in the BME framework: interval and probabilistic data whereas the mid-point of interval data was considered in the second approach of kriging. The calibration data set was used to calculate the interval data. The pairs of data values of EC_a and $EC_{2.5}$ were used to determine the calibration equations, one for each time instant, by calculating simple ordinary least squares regression models:

$$\ln(EC_{2.5}) = a + b\ln(EC_a),\qquad\qquad(1)$$

where a is the intercept and b is the slope of the regression model.

These calibration equations were applied to the 'data set to be calibrated' to give the expected values and their standard deviations for all locations (413) and 17 time periods; the two remaining periods (sampling campaigns 18 and 19, which correspond to March and June 2001, respectively) were kept for validation. These parameters were used to determine the 95% confidence intervals; their lower and upper limits form our interval soft data:

$$\chi_{\text{soft}} = \left\{ \left[\chi_{h+1},...,\chi_m \right]' : \chi_i \in I_i = \left[l_i, u_i \right], i = h+1,...,m \right\}\qquad(2)$$

This means that the non observed exact values χ_i have probabilities of one within known intervals I_i with l_i and u_i the lower and upper limits, respectively. All the intervals constitute the definition domain $\boldsymbol{I} = I_{h+1} \cup I_{h+2} \cup ... \cup I_m$.

The interval midpoint data used in kriging were calculated simply as the average of the lower and upper limits, which is in fact the expected value.

The same expected values and their corresponding standard errors were used to determine the soft pdfs, $f_S(\xi)$, for each of the 413*17 points assuming a Gaussian distribution. These soft

pdfs were used as probabilistic soft data: $\chi_{\text{soft}} : P_S(x_{\text{soft}} \leq \xi) = \int_{-\infty}^{\xi} f_S(\chi_{\text{soft}}) d\chi_{\text{soft}}$ \hfill (3)

For illustration, the probabilistic soft data for some points are reported in Fig. 8.1. The soft data from location 1 at November 1994 is smaller in magnitude than the 3 others, however it is less uncertain as its pdf is less dispersed around the mean value.
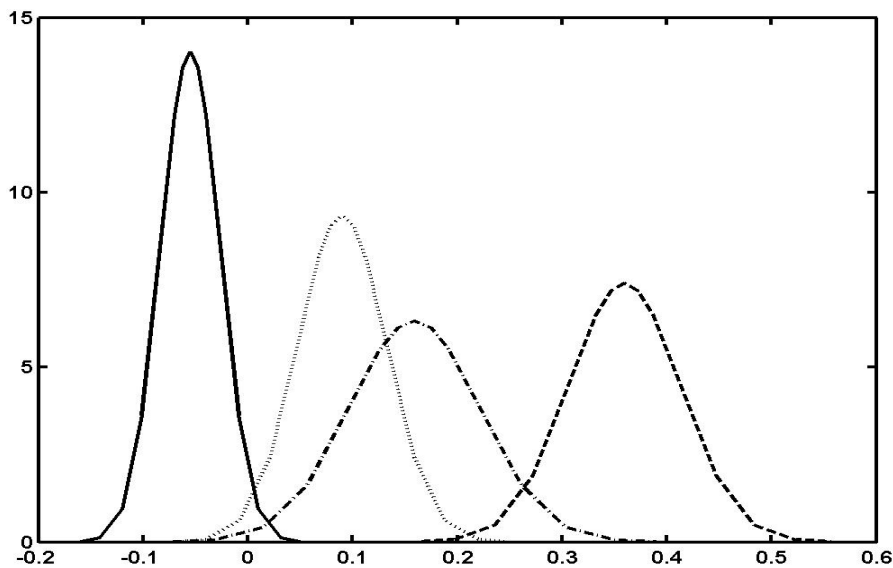
Fig. 8.1. Examples of probabilistic soft data (based on the residuals). Full curve: location 1 in November 1994; dashed curve: location 2 in March 1995; dash-dotted curve: location 3 in June 1995; dotted curve: location 4 in September 1995.

We compared two methods of space-time prediction, each with two variants. They differ in the way the soft data are processed:

(1) Ordinary kriging using only the hard data (HK), which provides no direct way of integrating soft data and ignores them;

(2) Ordinary kriging using hard data and the midpoint of the interval soft data, regarding the latter as if it was a hard datum (HSK), and disregards their uncertainty. The difference between the two types of kriging is in the number of data considered during the analysis.

Ordinary kriging (HK) is limited to the hard (observed $EC_{2.5}$) data only, whereas HSK treats both sets of data (observed and predicted $EC_{2.5}$) as hard essentially;

(3) Bayesian maximum entropy using the hard and interval soft data (BMEI), which integrates the interval soft data in the prediction as they are and maintaining the difference in the degree of uncertainty between hard and soft data;

(4) Bayesian maximum entropy using the hard and interval soft data (BMEP), which integrates the full distribution of the soft data in the prediction. Because BME processes the full pdf of the soft data rather than just the mid point of its confidence interval, it uses more information and it is expected to lead to more accurate predictions.

The methods were compared by cross-validation on observed $EC_{2.5}$ data of sampling campaigns 18 and 19.

8.3.2. Validation and comparison criteria

Soil salinity was predicted at each of the sites (19 for March 2001 and 20 for June 2001), which have not been used in all the previous computations (structural analysis and neighborhood) and for which we had measurements, by deleting in turn the value of each location where the prediction was being made. This gave pairs of estimated-observed soil salinity values for the two time periods. This was done for the three first methods (HK, HSK, and BMEI) and not for BMEP. Three quantitative criteria were computed from these pairs of values: the Pearson correlation coefficient (r), the mean error or bias (ME), and the mean squared error (MSE). The first, r, measures the strength of the linear relation between the estimated and the observed soil salinity values, and should be close to one for an accurate prediction. The ME should be close to zero, and the MSE should be as small as possible. We also represented graphically the distribution of the estimation errors for a visual comparison of the three methods.

The MSE can be divided further into components that identify and quantify the deviation of estimated values from the observations. They represent different aspects of the discrepancy between the estimates and the measurements (Kobayachi and Us Salam, 2000). Let x_i and y_i ($i = 1, ..., n$) represent the estimated and the observed soil salinity values, respectively, and $d_i = x_i - y_i$ the deviation of the estimated values from the observations. The mean error (ME) or bias is defined by:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) = \bar{x} - \bar{y}, \tag{4}$$

where $\bar{x}$ and $\bar{y}$ represent the means of the estimated and the observed values, respectively, and n is the number of locations for which observations are available. The mean squared error (*MSE*) is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 = (\bar{x} - \bar{y})^2 + \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \bar{x}) - (y_i - \bar{y})\right]^2 . \tag{5}$$

The first term on the right is the square of the bias (*SB*):

$$SB = (\bar{x} - \bar{y})^2 = ME^2 . \tag{6}$$

The second term is the mean squared difference between the estimates and the measured values with respect to the deviation from means. It is known as the mean squared variation (*MSV*) and represents the proportion of the *MSE* that is not due to the bias. A larger *MSV* indicates that the model did not estimate the variability of the observed values around their mean adequately, i.e. the precision of the predicted values is poor.

Equation (5) can be rewritten as:

$$MSE = SB + MSV . \tag{7}$$

The *MSV* can be divided, in turn, into two components:

$$MSV = (SD_e - SD_o)^2 + 2SD_e SD_o (1 - r) = SDSD + LCS , \tag{8}$$

where $SD_e = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $\tag{9}$

which is the standard deviation of the estimated values, and

$$SD_o = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} , \tag{10}$$

is the standard deviation of the observed values. The *SDSD* is the difference in the magnitude of fluctuation between the estimated and measured values. A larger value implies that the model failed to estimate the magnitude of fluctuation among the measurements. The *LCS* is the lack of positive correlation (1-*r*) weighted by the standard deviations. A large value means that the model did not estimate the degree of fluctuation in the observations.

Equation (7) can be rewritten as:

$$MSE = SB + SDSD + LCS \tag{11}$$

The validation study was followed by predicting soil salinity for time instant n° 12 (September 1998) along with the corresponding estimation variances. The data for this time instant were included in the structural analysis. We used a fine estimation grid (10x10 m). The

probability that the soil salinity could exceed 4 dS m^{-1} was mapped; this value is the threshold used to distinguish between saline and non-saline soils (Spaargaren, 1994; USDA, 1996).

All the analyses were done using the BMElib toolbox (Christakos et al., 2002) written for Matlab (MathWorks, 1999).

8.4. Results and discussion

8.4.1. Descriptive statistics

Table 8.1 gives the summary statistics of soil salinity ($EC_{2.5}$) for the different time periods. The mean values, which vary between 1.42 dS m^{-1} for November 1994 and 3.30 dS m^{-1} for September 1998, suggest the presence of strong temporal variability. There is also considerable spatial variation shown by the large ranges between the minimum and maximum values for the different time periods. For example, December 2000 has a range of 7.38 dS m^{-1}, which is the largest for the data examined.

The Pearson correlation coefficients between EC_a and $EC_{2.5}$ are strong. They vary from 0.83 to 0.97.

8.4.2. Covariography

The structural analysis was fully described in the precedent chapter (see section 7.4.3).

Table 8.1. Summary statistics for the hard data EC$_{2.5}$ (dS m^{-1}).

$EC_{2.5}$	N	Mean	std	Range	r	Skewness	Kurtosis
				Calibration data			
Nov-94	13	1.42	0.34	0.96	0.85	-0.71	-1.00
Mar-95	20	2.29	0.70	2.23	0.91	0.66	-0.78
Jun-95	20	2.02	1.10	3.66	0.88	0.80	-0.50
Sep-95	20	2.01	0.99	3.73	0.94	0.89	0.67
Dec-95	20	1.84	0.90	2.59	0.92	0.45	-1.31
Mar-96	16	2.07	0.81	2.57	0.87	0.08	-1.26
Jun-96	20	1.83	0.90	2.86	0.87	0.28	-1.12
Mar-97	20	1.61	0.63	2.20	0.89	-0.08	-0.87
Jun-97	15	1.77	0.97	3.12	0.83	0.78	-0.23
Sep-97	20	1.63	1.24	4.25	0.94	1.06	0.20
Dec-97	20	1.60	1.04	3.61	0.90	0.51	-0.62
Sep-98	20	3.30	2.17	6.95	0.85	0.50	-0.69
Apr-99	20	1.84	1.74	6.41	0.93	1.70	2.48
Jul-99	13	2.27	1.57	4.86	0.91	0.73	-0.52
Sep-99	20	2.29	1.89	6.52	0.91	1.07	0.10
Apr-00	18	2.11	1.63	6.32	0.94	1.19	1.51
Dec-00	20	2.32	1.91	7.38	0.93	1.31	1.71
				Validation data			
Mar-01	19	1.80	1.21	4.42	0.97	1.05	1.18
Jun-01	20	1.99	1.75	5.83	0.86	1.22	0.42

N is the number of observations, std is the standard deviation, and r is the Pearson correlation coefficient.

8.4.3. Comparison of results

Soil salinity was predicted for two time periods, March and June 2001, using the four approaches discussed above. The cross validation criteria *ME, MSE,* and *r* are given in table 8.2 for both times. Figure 8.2 gives the distributions of the errors for March 2001.

The HSK results are the poorest; they have the largest bias (*ME*) compared to the three other approaches, although it is still not significantly different from zero, and the largest *MSE* for both time periods. Figure 8.2 shows that HSK has the broadest error distribution, mostly on the negative side of the curve. This implies that this method is likely to produce larger errors than the other two. The errors for BME have a higher mode and a narrower distribution compared with both kriging techniques (Fig. 8.2): this is confirmed by its having the smallest *MSE* values (table 8.2). The *MSE* for HK is between those of BME and HSK, therefore this method provides more accurate estimates than HSK but less accurate ones than both BME techniques. However, HK gives estimates that are less biased than BME (mainly for March 2001). The estimates are strongly correlated with the observations for the three techniques and both time periods.

Table 8.2. Quantitative criteria for the comparison of the three approaches.

Criterion	Time	HK	HSK	BMEI	BMEP
r	March 2001	0.87	0.92	0.93	0.94
	June 2001	0.93	0.93	0.95	0.95
ME $(dS\ m^{-1})$	March 2001	-0.176	-0.323	-0.226	-0.173
	June 2001	0.062	-0.057	-0.017	-0.019
MSE $(dSm^{-1})^2$	March 2001	0.489	0.650	0.387	0.254
	June 2001	0.378	0.513	0.337	0.223

r is the Pearson correlation coefficient, ME is the mean error, and MSE is the mean squared error.
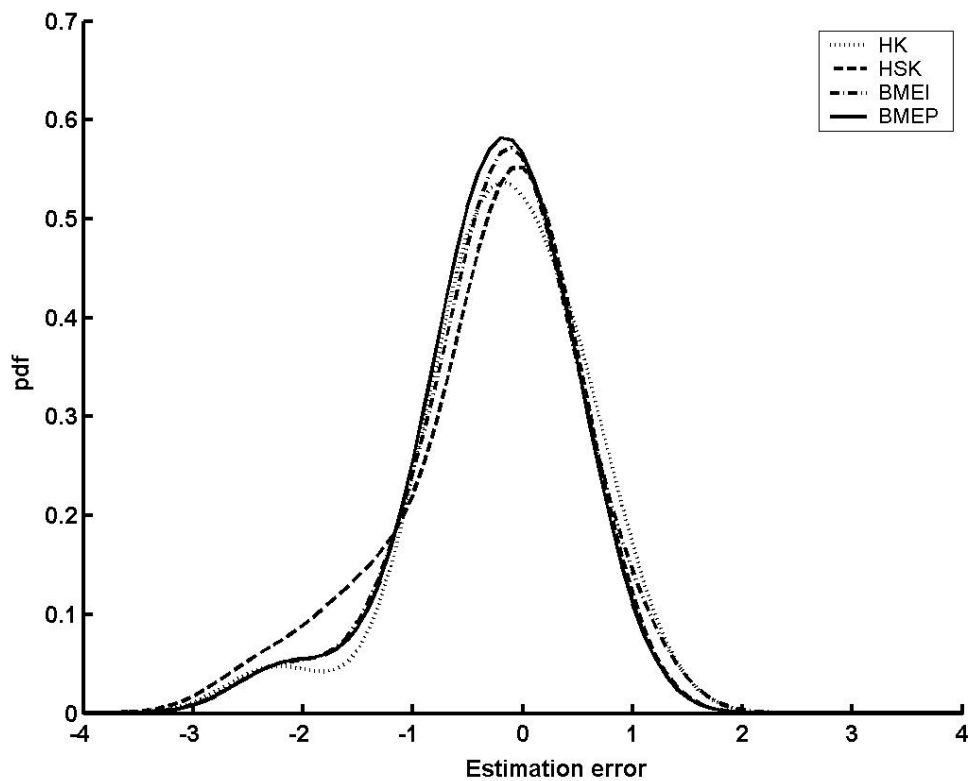
Fig. 8.2. Distributions of estimation error for March 2001. The solid line is for BMEP, the dash-dotted line is for BMEI, the dotted line is for HK, and the dashed line is for HSK.

When BME is used without any soft data, the results are strictly equivalent to HK. This is in accordance with the theory (Christakos and Li, 1998). Bayesian maximum entropy can also give estimates of soil salinity in the absence of hard data; the results of this cross validation are given in table 8.3. The estimates show more bias (but still negligible) and are slightly less accurate (particularly for March 2001), but the differences in the results when hard (the neighborhood research was limited to a maximum of the 10 nearest data) and soft data were used are not significantly greater (BMEI column of table 8.2, and table 8.3). This is a useful feature of BME, and D'Or and Bogaert (2003) used this property to map soil texture using only the intervals defined from a textural triangle.

Table 8.3. Quantitative criteria for BMEI when hard data were excluded.

Criterion	March 2001	June 2001
r	0.91	0.94
ME	0.294	-0.075
MSE	0.701	0.348

To investigate further the incorporation of soft data by HSK and BME, we analyzed, in addition to the hard data, only data with the largest intervals (from 30 spatial locations for each time period) rather than the full set of 393 spatial locations for each time period (see table 8.2 and Fig. 8.2). Table 8.4 gives the results for March 2001 and the distributions of the estimation errors are shown in Fig. 8.3.

Table 8.4. Quantitative criteria to compare HSK and BMEI methods of prediction using only the largest interval data.

Criterion	HSK	BMEI
r	0.23	0.86
ME	-88.4	-0.192
MSE	106079.4	0.612

HK is not reported in table 8.4 as this technique takes no account of the soft data. The *ME* for BME, -0.192, is not markedly different from that in table 8.2 when all the interval data were used (-0.226), but the *MSE* is larger, 0.612 instead of 0.387 for the full interval data. The results for HSK show that the estimates are biased and inaccurate, giving erroneous estimates for some spatial locations. The quantitative criteria confirm the graphical representation in Fig. 8.3.
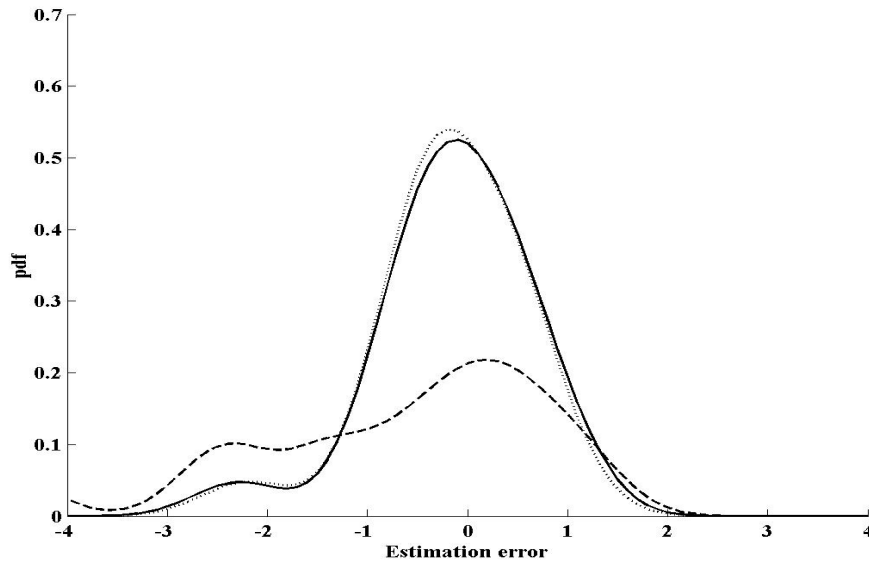
Fig. 8.3. Distributions of the estimation errors for March 2001, using only the largest interval data. The solid line is for BMEI, the dotted line is for HK, and the dashed line is for HSK.

The components of *MSE* can provide more information about the difference between the estimated and the observed values. They are given in absolute values in table 8.5 and shown as proportions in Fig. 8.4. The HSK has the largest *MSE* for both time periods, whereas BME has the smallest values. This suggests that BME provides more accurate predictions than the two methods of kriging used. In addition, the contribution of the bias to the *MSE* is almost zero for the three approaches for June 2001, but for March 2001 it is largest for HSK (16 %), followed by BME (13.2 %), and finally HK (6.3 %). This confirms that HK estimates are the least biased and those of HSK are the most biased.

The lack of positive correlation (*LCS*) is the component that contributed the most to the *MSE* of the three interpolation methods for both time periods (table 8.5 and Fig. 8.4); the larger contribution is for June 2001. In particular, it contributed the most to the *MSE* of the HK estimates with 87 % and 95 %, whereas it contributed least to that of HSK (47 % and 62 %), and its contribution was intermediate for BME (63 % and 76 %); the first value refers to March and the second to June 2001. This suggests that HK failed to estimate the degree of fluctuation in the observed soil salinity even if its *MSE* is smaller than that of HSK. However, the *SDSD* contributed more to the *MSE* of HSK (36.8 % and 36.5 %) than to that of HK (6.7 % and 4 %) and BME (23.8 % and 24.3 %), which indicates that HSK failed to estimate the magnitude of fluctuation in the measured electrical conductivity. The three components of *MSE* (*SB*, *SDSD*, and *LCS*) for BME are intermediate to those of HK and HSK. The *MSE* of

BME is the smallest, indicating that it performs better than the two kriging techniques, and the *SDSD* shows that it represents the degree of fluctuation in the observations reasonably.

The small *MSE* for BME in March 2001 can be explained as follows. The standard deviation SD_o for this time period is 1.18 dS m^{-1}. Since *SB* (0.051) and *SDSD* (0.092) are negligible, *LCS* (0.244) is the component that contributes most to the *MSE* (0.387). As the Pearson correlation coefficients (0.92 for HSK and 0.93 for BME) are the same and SD_o is the same, the larger SD_e for HSK (1.67) led to a larger *LCS*, and hence larger *MSE*. The smaller SD_e for BME (1.48) resulted in a small *LCS*, and hence small *MSE*. The overall deviation (*MSE*) between the measurements and the BME predictions for March 2001 is small because the predicted soil salinity shows limited variation (compared to HSK estimates) for the 19 sites and BME predicted the observations with a reduced bias (─0.226 against ─0.323 for HSK).

Table 8.5. The components of MSE as absolute values.

Criterion	Time	HK	HSK	BMEI
SB	March 2001	0.031	0.104	0.051
	June 2001	0.003	0.004	0.000
SDSD	March 2001	0.033	0.239	0.092
	June 2001	0.015	0.187	0.082
LCS	March 2001	0.425	0.307	0.244
	June 2001	0.360	0.322	0.255
MSE	March 2001	0.489	0.650	0.387
	June 2001	0.378	0.513	0.337

SB is the squared bias, SDSD is the squared difference between standard deviations, and LCS is the weighted lack of positive correlation.

(a)
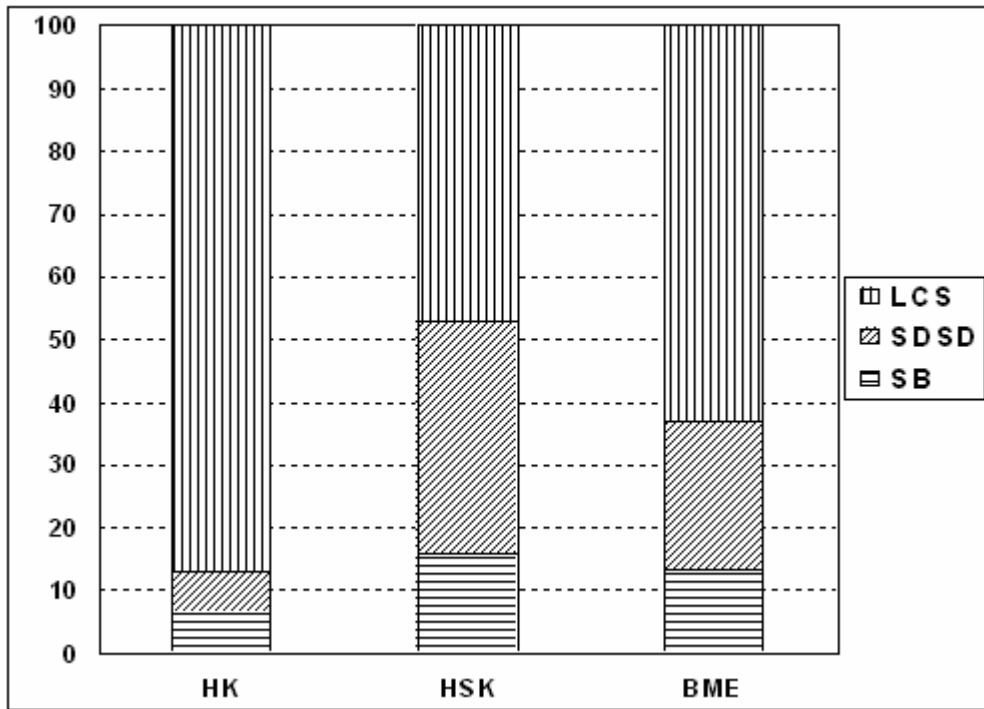
(b)
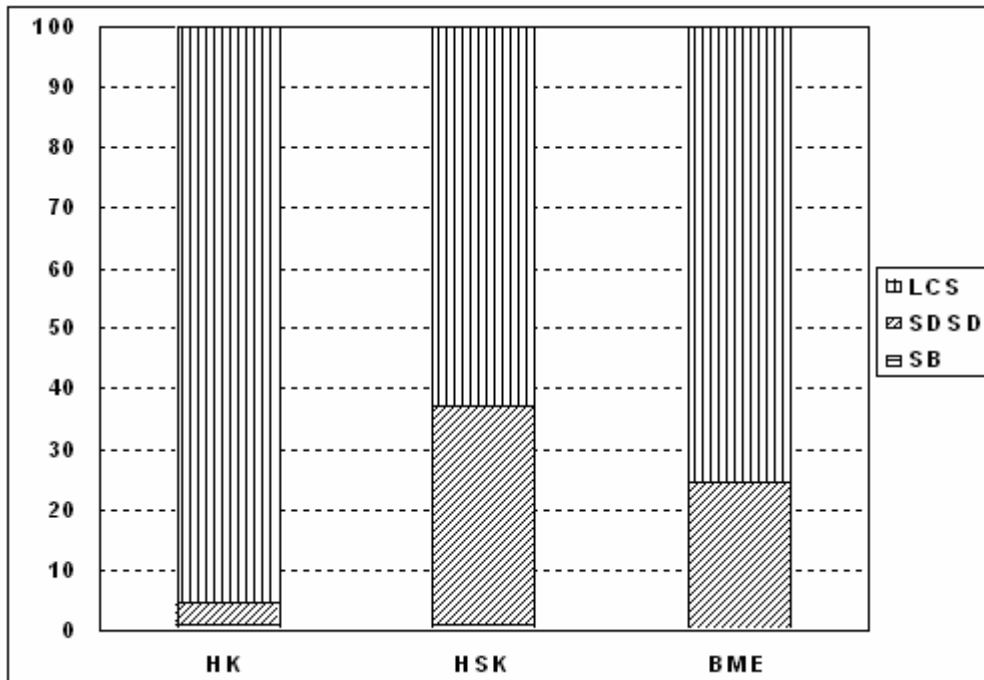
Fig. 8.4. Contribution of the components of MSE to its total. (a): March 2001, (b): June 2001.
The SB is the squared bias, SDSD is the squared difference between standard deviations,
and LCS is the weighted lack of positive correlation.

Using only the largest interval data instead of all of them, table 8.6 and Fig. 8.5, the *SDSD* becomes the major component of *MSE* for HSK (92.1 %). This means that HSK failed to estimate the magnitude of fluctuation accurately among the observed electrical conductivity values. The minimum and maximum values of the latter for March 2001 are 0.16 and 4.58 dSm^{-1}, respectively, whereas for the HSK estimates they are 0 and 1408 dSm^{-1}. This large range for the latter resulted in the large SD_e of 313.74 dS m^{-1} compared to the small SD_o of 1.18 dSm^{-1}; these values explain the large *SDSD*. In comparison BME performed well. Its minimum and maximum predictions are 0.10 and 5.24 dSm^{-1}, respectively. This is a consequence of how the soft data are integrated into the prediction process. The HSK used only the midpoint of intervals, disregarding their range and the uncertainty associated with them, whereas BME considers the full information provided by the soft data. It takes into account the upper and lower limits of the interval data, and the uncertainty associated with them. More importantly, BME distinguishes clearly between accurate (hard) and uncertain (soft) data and processes them differently.

Table 8.6. The components of MSE as absolute values, using only the largest interval data for March 2001.

Component	HK	HSK	BME
SB	0.031	7813.7	0.037
SDSD	0.033	97695.5	0.086
LCS	0.425	570.2	0.489
MSE	0.489	106079.4	0.612

SB is the squared bias, SDSD is the squared difference between standard deviations, and LCS is the weighted lack of positive correlation.
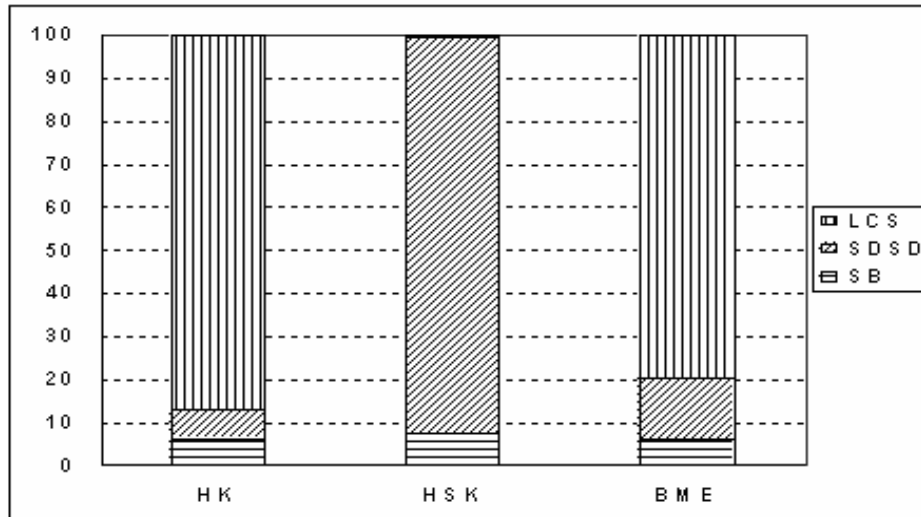
Fig. 8.5. Contribution of the components of MSE to its total, using only the largest interval data for March 2001. The SB is the squared bias, SDSD is the squared difference between standard deviations, and LCS is the weighted lack of positive correlation.

8.4.4. Space-time mapping of soil salinity

We predicted the soil salinity for September 1998 to compare the HK, HSK, and BMEP graphically. The soil salinity maps are given in Fig. 8.6 and their corresponding estimation variances in Fig. 8.7.

The soil salinity is strongly smoothed in the map obtained by HK. This is the consequence of the limited number of hard data (19, in fact as we interpolated in the space-time domain, we have the 19 locations of some previous time instants as well; however, geographically, we used the same 19 locations). In contrast, the maps of HSK and BME show much more detail due to the additional data ('hardened' for HSK or used as soft for BME). Furthermore, for the BME, we mapped the conditional mean (minimizing a least square criterion), Fig. 8.6 (c), as well as the mode (the most probable value) of the posterior pdf, Fig. 8.6 (d). The map of the BME mode estimate is less smooth than that of the BME mean estimate.

The estimation variance maps (Fig. 8.7) reflect the difference in data availability. The HK map shows a zero estimation variance at the locations where hard data were present and a gradual increase in the estimation variances as one goes further away from these hard data locations. In the case of HSK, the estimation variance is zero for the hard as well as soft data points, resulting in a small estimation variance for the entire study area (except in the centre where the locations are more spaced), which results in an under-prediction of the true uncertainty. On the other hand, BME map has a zero estimation variance at the hard data

points only, and a small but non-zero estimation variance at the soft data points, which is a better representation of the true uncertainty in the estimated map.

We also mapped the probability that the estimated soil salinity exceeds 4 dS m^{-1}, which is considered as a critical threshold separating non-saline from saline conditions (Fig. 8.8). As the estimation standard errors for HK are higher than those for HSK and BME, and as the estimated soil salinity is smoothed, a larger area of saline soil was delineated. In contrast we obtained, for HSK and BME, clearly delimited and smaller areas of saline soil.
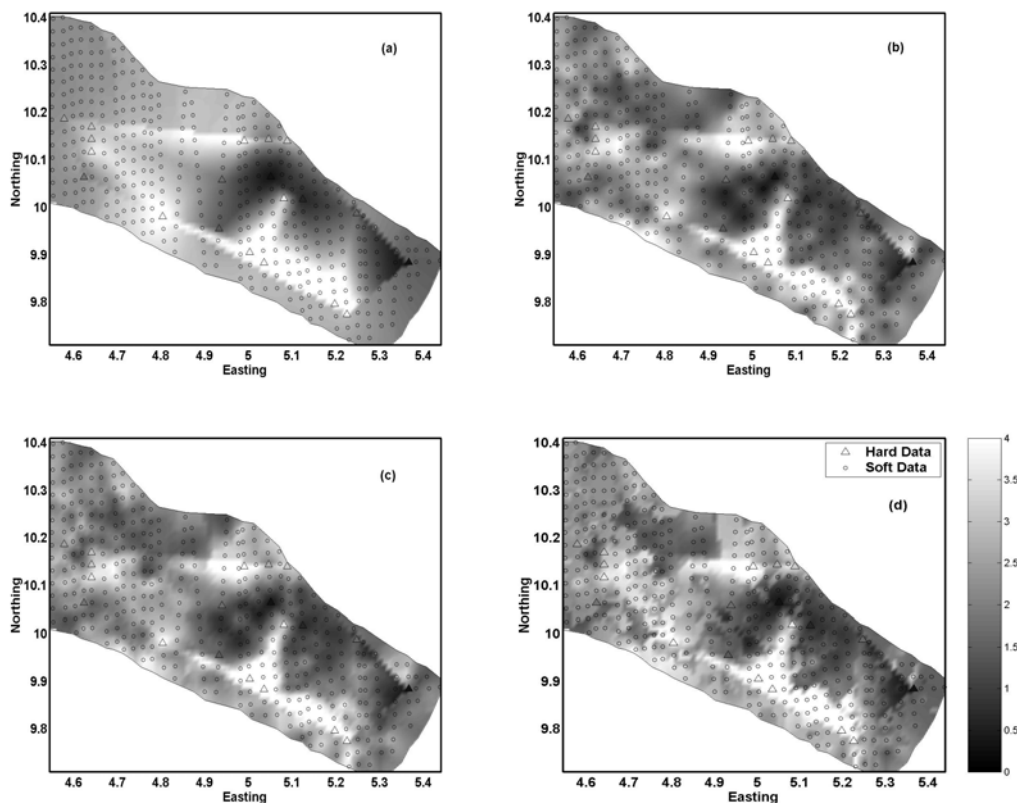
Fig. 8.6. Soil salinity (EC$_{2.5}$ in dS m^{-1}) estimates for September 1998; (a): HK, (b): HSK, (c): BME mean, (d): BME mode. On the maps, triangles and circles indicate points where hard and soft data were available, respectively.
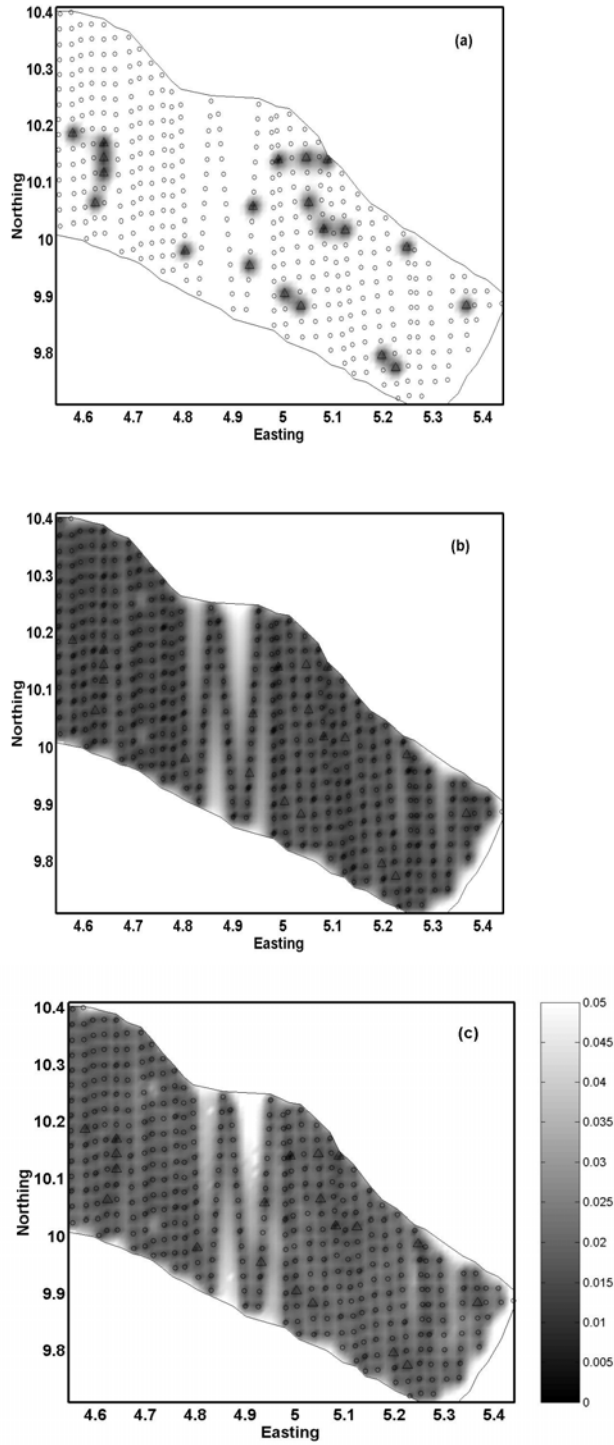
Fig. 8.7. Soil salinity estimation variances for September 1998; (a): HK, (b): HSK, (c): BMEP.
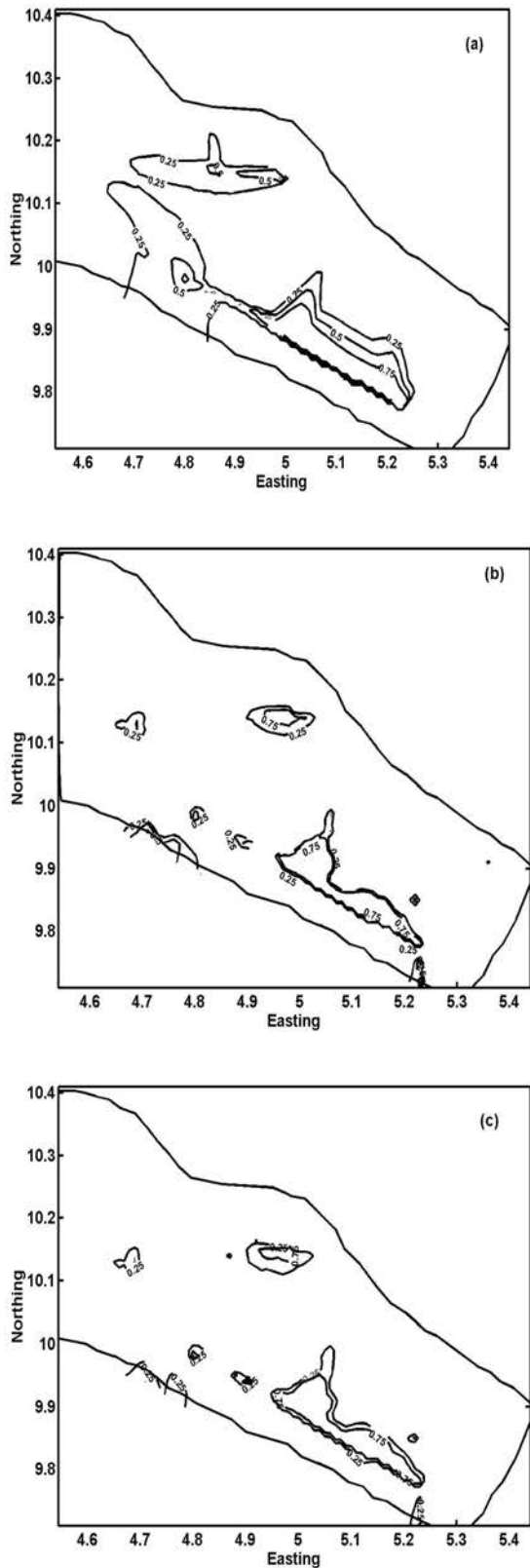
Fig. 8.8. Probability that estimated soil salinity exceeds 4 dS m⁻¹ for September 1998; (a): HK, (b): HSK, (c): BMEP mean.

8.5. Conclusions

The main aim of this work was to compare the performance of two prediction techniques, each with two variants: BME which incorporates either interval or probabilistic soft data, and two variants of kriging (one using only hard data and the other using hard data as well as the midpoint value of interval soft data, treating them as if they were hard data). The four approaches were evaluated by cross-validation for two different time periods which had not been used in the previous analyses.

The BME provided reliable estimates even in the absence of any hard data. When no soft data were used, the BME estimates were strictly equivalent to those from kriging (HK). Based on the *ME* and *MSE,* we can conclude that the predictions from BME are less biased and more accurate than those from the two kriging techniques. Of these two techniques, the one using the soft data (HSK) resulted in more bias and less accuracy in the predictions. The results showed that BME improved substantially the accuracy of the predictions compared to kriging, by taking into account soft data (interval or probabilistic).

The Pearson correlation coefficients were of the same magnitude for HSK and BME. However, by dividing the *MSE* into three components, we found that HSK gave more biased estimates (large *SB*) and failed to reproduce the true magnitude of fluctuation among the observations.

Also the BME estimation error distribution showed a higher peak around zero than the two other techniques, indicating that the probability of obtaining an estimate equal to the observed soil salinity is higher for BME than for the two versions of kriging. In addition, BME allows one to delineate more rigorously saline areas from non-saline zones.

The failure of HSK to incorporate the soft information was more pronounced when we used only the largest interval data, in addition to the hard data, instead of the full interval data. In this case HSK produced some unrealistic predictions of electrical conductivity (very large and unreliable predictions). In contrast, the way that BME integrates soft data into the prediction process resulted in more accurate predictions, whether we used the full interval data or only the largest ones.

Ancillary data are cheap and readily available, sometimes for the whole study area (exhaustive secondary information). This secondary information can be used in an efficient way to complement the scarcity of the direct measurements of a soil property. This work showed that BME can incorporate and process soft data rigorously, leading to more accurate

predictions. It can use cheap, dense and easily obtained data (like EC_a), to estimate with less bias and more accuracy, a scarce, time consuming and expensive soil property of interest (such as $EC_{2.5}$).

PART IV

GENERAL CONCLUSIONS AND
FURTHER RESEARCH

CHAPTER 9

GENERAL CONCLUSIONS AND FURTHER RESEARCH

9.1. Introduction

Our research work had two main objectives: (i) evaluating temporal stability and monitoring temporal change of the spatial pattern of soil salinity and (ii) predicting this soil salinity at unsampled space locations and time instants.

In this chapter, we briefly summarize our major findings and the contributions of this research work to the accomplishment of these objectives. Then, some recommendations for further research are given.

9.2. Major findings

In a first application (chapters 5 and 6), we examined the temporal stability of soil salinity in terms of electrical conductivity determined, in the laboratory, in an extract from a 1:2.5 soil:water ratio, $EC_{2.5}$ and apparent electrical conductivity, EC_a, surveyed in the field using a 4-electrode probe sensor.

Based on the first concept of temporal stability, the Spearman rank correlation coefficient, we found that there exists temporal persistence in the spatial pattern of soil salinity. However, as this statistical tool is just measuring the degree of concordance between two series of rankings, it cannot be used to foresee the positions of the measurement locations. This goal can be achieved using the second concept of temporal stability, i.e., the mean relative difference. The analysis of deviations between individual and field average soil salinity showed that some space locations are time stable. If, in addition, these locations have a mean relative difference about zero, they are representative of the field average soil salinity and thus the sampling effort can be drastically reduced in the future to these field average soil salinity monitoring sites. The soil salinity average value is an important parameter for mathematical modelling of solute transport if quasi steady state assumptions can be made. For example, Oosterbann (1997) developed the SALTMOD model based on the cumulative Gumbel distribution. This distribution is assumed to fit the cumulative distribution function of the root zone salinity and requires the mean value and the standard deviation (it is function of the mean value) as parameters for its full characterization. In addition, time stable locations with positive (negative) mean relative difference can be useful to identify more (less) than average saline locations. Both these groups of locations with those representative of the field average

value can be used to determine the probability distribution function of soil salinity and then determine its main parameters like the mean or the standard deviation from a very limited number of measurement locations. Moreover, temporal stability of spatial patterns of EC_a may be relevant to the management of agricultural fields leading to the challenging site-specific management or precision agriculture. However, precision agriculture depends on identifying parts of a field that will be managed differently (management units or zones) and heavily on the size and the stability of the units through time. The next requirement in precision agriculture would be to map the management units. This can be readily accomplished via EC_a which can be easily and intensively surveyed and which was proven to be strongly correlated to soil salinity ($EC_{2.5}$).

The time stable locations may also be relevant to the calibration and validation of remote sensing products like aerial photographs or satellite images. These locations could be used to gather 'ground truth' data required for calibration and validation of remote sensed materials in order to determine, for example, a soil salinity index.

A key problem to resolve, in connection with time stable locations, is how to locate easily in the field these sites. We found that the different salt processes (accumulation, leaching, and waterlogging) are strongly linked to easily measurable factors like vegetation pattern or elevation and the level of soil salinity is, in turn, related to these processes. Therefore, based on vegetation pattern and elevation, one can easily identify high, average or low saline locations with a minimum sampling.

The temporal stability of soil salinity was assessed using both variables (EC_a and $EC_{2.5}$) separately and using measurements of these variables for a given number of time instants (in our case, 19). However, as $EC_{2.5}$ is strongly correlated with EC_a the sampling effort for monitoring the temporal change of soil salinity can be further reduced provided that there is no dynamic spatial variation. The latter refers to no or a proportional change for all space locations between two time instants. In this case, one requires surveying EC_a only at the first time instant and, based on a response surface spatial design algorithm, selecting a limited number of space locations (in general 12 to 20) from which soil samples will be analyzed in the laboratory for the determination of $EC_{2.5}$. The latter will be determined for both time instants. Thus the EC_a surveying at the second time instant is avoided and consequently the sampling effort reduced.

However, if dynamic spatial variation is detected between two time instants, a new calibration equation should be estimated. This means that EC_a should be surveyed at both time instants in

contrast to one EC_a surveying in the absence of dynamic spatial variation. Thus, in the presence of dynamic spatial variation between any pair of consecutive time instants, the regression linking $EC_{2.5}$ to EC_a should be estimated in order to convert the intensively measured EC_a into 'predicted' $EC_{2.5}$.

The second objective of our research work was achieved when the space-time variability of the 'predicted' $EC_{2.5}$ values was determined and modelled and then used in geostatistical algorithms to predict soil salinity at unobserved space locations and time instants.

In this way, we determined and modelled the joint space-time variability by using the space-time covariance function. Afterwards, we compared first space-time kriging to spatial kriging and then space-time kriging to BME. For both space-time prediction methods, we used two approaches: kriging based only on hard data (observed $EC_{2.5}$) and kriging using observed as well as mid interval of soft ('predicted' $EC_{2.5}$) data while for BME we used hard data and two kinds of soft data, i.e., interval and probabilistic. These soft data were derived from the 'predicted' $EC_{2.5}$ values.

The results of this work showed that the space-time covariance function provides a more accurate and realistic representation of the correlation structure of the soil salinity both in space and time than that described by a purely spatial or a purely temporal covariance function. Also, and as a consequence of the precedent result, the space-time kriging provided more accurate predictions than kriging limited to the spatial domain. In addition, space-time kriging using hard and mid interval soft data was the worst with the highest bias and the least accuracy compared to the other approaches. Moreover, BME produced less biased and more accurate predictions than kriging based only on hard data. Finally, BME based on hard and probabilistic soft data gave the most accurate predictions among all compared approaches.

These results confirm the conclusions from precedent works which compared BME to kriging.

The BME is a versatile space-time prediction method and more general than the classical geostatistical prediction method, i.e., kriging which has many limitations. It has many advantages over kriging which are reported in what follows:

- Kriging uses only statistical moments of first and second order (mean and covariance function) and hard data whereas BME uses these moments as well as moments of higher order, hard and soft data of different kinds, and physical laws can be easily integrated in the knowledge processing rule. Moreover, soft data at prediction points can be taken into account;

- Kriging is limited to single point estimation. In contrast, BME offers single as well as multipoint mapping;

- Kriging is the Best Linear Unbiased Estimator (BLUE) so it is restricted to linear estimators whereas BME provides estimators that are, in general, non linear and thus BME is simply Best Unbiased Estimator (BUE);

- Additional constraints like unbiasedness are imposed by kriging. Conversly BME does not require these constraints;

- Kriging is limited to interpolation whereas BME allows interpolation and extrapolation;

- Most of the kriging techniques produce as output the conditional mean and the estimation error variance; only indicator kriging provides a step cumulative distribution function with some severe drawbacks. In contrast, BME provides the full pdf from which different estimators like the conditional mean, the variance, the mode, any percentile, etc. can be easily derived and mapped;

All these points emphasize the flexibility of BME and its more general character which make it a very powerful method for space-time mapping.

9.3. Recommendations for further research

Other data sources can be readily used to improve the space-time analysis of soil salinity. Also, there are a number of areas arising from the work presented in this thesis that demand further investigation, specifically in data processing.

9.3.1. Data sources

In this thesis, soil salinity data were analyzed for their space-time variability. We used electrical conductivity determined in the laboratory ($EC_{2.5}$) and measured in the field (EC_a) using 4-electrode probe. An additional consideration might be to complement EC_a values from 4-electrode probe with those measured using an electromagnetic induction device since the latter needs no contact with soil, hence extensive measurements can be done within a similar time. This is expected to be done during 2005 using a dipole EM38 device in our study site.

Also, additional and easily collected ancillary data can be made available via remote sensing like aerial photographs or even better, satellite images. Remote sensing is a valuable tool for obtaining relevant data on soil salinity. Multi-temporal images are a suitable way to detect the changing state of soil salinity by considering them as secondary covariables in the cokriging paradigm or as explanatory variables in the regression calibration models.

Finally, as was noted before, BME is able to incorporate different sources of data. General knowledge in the form of a physical law can be accounted for. This allows obtaining space-time predictions by considering the law that governs the natural process (soil salinity) while the predictions also embody the case-specific data (hard and soft data).

9.3.2. Data processing

As data become available for more time instants and space locations, using for example electromagnetic induction or remote sensing, other geostatistical methods can be used. As example, the space-time cokriging (sections 4.3.7 and 4.4.5) can apply to our data by considering $EC_{2.5}$ as the property of interest and EC_a and/or remote sensing data as covariables. This will increase the accuracy and the reliability of the predictions. Also, the different covariance functions would be fitted using more elaborate models like the ones reported in section 4.3.6. Moreover, in our application we considered only the space-time component in the model, but it may be more cautious to add a purely spatial and/or a purely temporal component to the model (Bogaert, 1996a; see also section 4.3.8), especially when some non ergodicity is present in the data. Also, the covariables can be used in the BME framework which is then called vectorial BME. Additionally, as we noted in the precedent section, BME can be used to incorporate the physical law governing the processes related to soil salinity. Space-time simulation algorithms can also be used to evaluate the space-time uncertainty in soil salinity and to forecast different scenarios and thus to choose between different soil management alternatives. Finally, it would be challenging to find connexions between the temporal stability and the covariance function even though these two concepts are based on completely different assumptions (spatial/temporal independence for the former method whereas spatial/temporal correlation for the latter) and have somewhat contradictory objectives, i.e., the former tries to reduce the sampling effort as the interest is in the average or extreme values whereas the computation of the covariance function requires numerous space-time data and the objective is to map the entire study site at different time instants (in the past as well as in the future).

SUMMARY

SAMENVATTING

RESUME

SUMMARY

Evaluation of the space-time variability of soil salinity by statistical, geostatistical and Bayesian maximum entropy methods.

by **ir Ahmed DOUAIK**

Aims

This thesis aims to the development of statistical and geostatistical methods for the analysis of space-time data and their application to soil salinity. Special emphasis was put on how to characterize soil salinity. Also, the focus was on three groups of methods, i.e. statistical, geostatistical and Bayesian maximum entropy (BME) approaches. Although, the methods were applied to soil salinity, they can be applied to other fields of research as well.

In summary, the two main objectives of this study were:

-   monitoring the temporal change of the spatial pattern of soil salinity using classical statistical methods;

-   mapping of soil salinity at unobserved space locations and time instants using classical and modern geostatistical methods like the BME.

Outline

After a general introduction (chapter 1), a literature review on soil salinity was presented (chapter 2). It was deduced from this chapter that soil salinity can be measured in the laboratory or in the field. For laboratory analysis, soil salinity is determined either from an extract from a saturated soil paste (EC_e) or from an extract from a given soil:water ratio, for example 1:1, 1:2.5, 1:5. In the field, soil salinity can be evaluated by measuring the apparent or bulk soil electrical conductivity (EC_a) using 4-electrode probe or electromagnetic induction devices. In our case study we used $EC_{2.5}$ and EC_a from 4-electrode probe. These soil properties were measured at different space locations and repeated in time from November 1994 to March 2001. The study site and data were described in chapter 3.

In the review of methods for analysis of space-time data (chapter 4), we distinguished three groups of methods. First, classical statistical methods which ignore the space and time

coordinates were considered. In this group, we presented the coefficient of variation, the analysis of variance and the variance component analysis, the temporal stability, and the dynamic spatial variation. The last two methods were used in our soil salinity application (chapters 5 and 6). The second group of methods involves classical geostatistical approaches including structural analysis (variogram and covariance function) and interpolation (kriging) in the space-time domain while the third group of methods includes modern geostatistical approaches, i.e. BME. Both the second and third approaches were applied to our soil salinity data (chapters 7 and 8).

Classical statistical methods

Soil electrical conductivity values obtained, at different locations and times, from field electrode probes and laboratory analysis were compared and analysed to check the temporal stability of salinity patterns. The statistical methods used were the Spearman rank correlation, the technique of relative differences, the significance of the parameters of the regression linking data from one period to those from its precedent, the paired-*t* test, and the test of significance of the dynamic spatial variation. Based on the Spearman correlation, we found that the ranking of the different locations persisted over time. We identified, based on the relative differences, three salinity classes with different temporal stability: the most time stable low saline area belonging to the zones of waterlogging and/or salt leaching, the least time stable high saline area belonging to the salt accumulation zone, and finally an area with locations which are representative of the average field soil salinity with an intermediate temporal stability belonging to the zones of salt accumulation, salt leaching and waterlogging. A dynamic spatial variation was found only for three pairs of times while for the remaining pairs there was no change or the change was proportional.

Geostatistical and Bayesian maximum entropy methods

A step further was reached when the field electrical conductivity was converted to laboratory electrical conductivity based on calibration equations linking laboratory to field electrical conductivity values. The converted laboratory electrical conductivity values were analysed for their space-time variability using variograms and interpolated using the kriging algorithm in a space-time domain. The space-time kriging was compared to kriging limited to the space domain. In addition, two variants of space-time kriging (using, on one hand, only the

observed laboratory electrical conductivity data and, on the other hand, the latter as well as the converted laboratory electrical conductivity data) were compared to the method of BME. BME is an interpolation technique which distinguishes between the difference in the accuracy of the observed and the converted laboratory electrical conductivity data as the latter are less accurate (as they are estimates) than the former. We found that the prediction in space-time domain was not different from those from the space domain but the former predictions were more accurate than the latter ones. Also we found that the BME is less biased, more accurate, and giving estimates which were better correlated with the observed laboratory electrical conductivity values than the two kriging techniques. Finally, using probabilistic soft data instead of intervals in the BME framework reduced the bias and increased the accuracy of the predictions.

The dissertation finishes with a concluding section (chapter 9) stressing the major results and giving recommendations for further research.

SAMENVATTING

Evaluatie van de ruimte-tijd variabiliteit van bodemverzilting via statistische, geostatistische en Bayesiaanse maximum entropie methoden.

door **ir Ahmed DOUAIK**

Doel

Deze thesis poogt de ontwikkeling van statistische en geostatistische methodes voor de analyse van ruimte-tijd gegevens en hun toepassing voor bodemverzilting. Speciale nadruk werd gelegd op het karakteriseren van bodemverzilting. De aandacht werd gericht op drie groepen van methodes: statische, geostatische en Bayesiaanse maximum entropie (BME) benaderingen. Hoewel de methodes werden toegepast op gegevens over het zoutgehalte van de bodem, zijn ze ook van nut voor andere onderzoeksdomeinen.

Samengevat, de twee hoofddoelen van deze studie waren:
- het opvolgen van de temporele verandering van het ruimtelijk patroon van bodemzouten op basis van klassieke statistische methoden;
- het in kaart brengen van bodemzouten op niet geobserveerde locaties en ogenblikken gebruik makend van klassieke en moderne geostatistische methodes zoals BME.

Overzicht

Na een algemene inleiding (hoofdstuk 1) werd een literatuurstudie over bodemzouten gegeven (hoofdstuk 2). Uit dit hoofdstuk werd afgeleid dat het zoutgehalte van de bodem zowel in het laboratorium als op het veld kan opgemeten worden. Voor laboratorium analyses wordt het zoutgehalte bepaald ofwel op basis van een extract van een verzadigde bodempasta (EC_e) ofwel op basis van een extract met een bepaalde bodem:water ratio, zoals bijvoorbeeld 1:1, 1:2.5, 1:5. Op het veld word het zoutgehalte bepaald met een 4-electrode sonde of met electromagnetische inductie instrumenten die de schijnbare of bulk bodem elektrische geleidbaarheid (EC_a) meten. In onze studie gebruikten we $EC_{2.5}$ en EC_a gemeten met een 4-elektrode sonde. Deze bodemeigenschappen werden gemeten op verschillende locaties en

verschillende tijdstippen gaande van november 1994 tot maart 2001. Het studiegebied en de gegevens werden beschreven in hoofdstuk 3.

De methodes voor analyseren van ruimte-tijd gegevens (hoofdstuk 4) hebben we opgesplitst in drie groepen. De eerste groep, de klassieke statistische methodes, negeren ruimtelijke en temporele coordinaten. Hierin kwamen aan bod de variatiecoefficient, de variantie analyse en de variantie component analyse, de temporele stabiliteit en de dynamische ruimtelijke variabiliteit. Deze laatste twee methodes werden toegepast op onze bodemzoutgegevens (hoofdstuk 5 en 6). De tweede groep methodes bestaat uit de klassieke geostatistische methodes met inbegrip van de structuuranalyse (variogram en covariantie functie) en interpolatie (kriging) in het ruimte-tijd domein. De derde groep methodes omvat de moderne geostatistische benaderingen, i.e. BME. Zowel de tweede als derde benadering werden toegepast op onze bodemzoutgegevens (hoofdstuk 7 en 8).

Klassieke statistische methodes

Metingen van de bodem elektrische geleidbaarheid, bekomen op verschillende locaties en tijdstippen met zowel de elektrode sondes als via laboratiumanalyses, werden vergeleken en geanalyseerd om de temporele stabiliteit van zoutpatronen te onderzoeken. De gebruikte statistische methodes waren 'Spearman rank' correlatie, de techniek van relatieve verschillen, de significantie van de regressie parameters die gegevens van één periode koppelt aan die van de voorgaande periode, de gepaarde t-test en de significantie test van de dynamische ruimtelijke variabiliteit. Gebaseerd op de Spearman correlatie konden we concluderen dat de rangorde van de verschillende locaties stabiel bleef doorheen de tijd. Op basis van de relatieve verschillen werden drie verziltingsklassen met elk een verschillende temporele stabiliteit geïndentificeerd. Het meest temporeel stabiele verziltingsgebied had lage zoutgehaltes en omvatte de zones van waterverzadiging en/of zoutuitloging, de minst temporeel stabiele gebieden waren de zoutaccumulatiezones met een hoog zoutgehalte en tot slot een gebied met locaties die representatief waren voor het gemiddelde bodemzoutgehalte met een intermediaire temporele stabiliteit waaronder de zones van zoutaccumulatie, zoutuitloging en waterverzadiging vallen.

Een dynamisch ruimtelijk patroon werd gevonden voor drie paren in de tijd terwijl er bij de overblijvende paren geen verandering was of de verandering was proportioneel

Geostatistische and Bayesiaanse maximum entropie methoden

In een volgende stap werd de veld elektrische geleidbaarheid omgezet naar laboratorium elektrische geleidbaarheid op basis van kalibratievergelijkingen die laboratorium en veld elektrische geleidbaarheidwaardes relateren. De ruimte-tijd variabiliteit van de geconverteerde laboratorium elektrische geleidbaarheidwaardes werd onderzocht met variogrammen en geinterpoleerd met het kriging algoritme in het ruimte-tijd domein. Deze ruimte-tijd kriging werd vergelijken met een kriging beperkt tot enkel het ruimte domein. Bovendien werden twee varianten van ruimte-tijd kriging (gebruik makend van enerzijds geobserveerde laboratorium elektrische geleidbaarheid en anderzijds de laatstgenoemde uitgebreid met de geconverteerde laboratorium elektrische geleidbaarheid) vergeleken met de methode van BME. BME is een interpolatietechniek die onderscheid maakt in de nauwkeurigheid van de geobserveerde en geconverteerde laboratorium elektrische geleidbaarheid gegevens daar de laatstgenoemde minder nauwkeurig is dan de eerstgenoemde (het zijn immers schattingen). We ontdekten dat de voorspelling in het ruimte-tijd domein niet verschilde van de voorspelling in het ruimte domein maar de voorspellingen in het ruimte-tijd domein waren nauwkeuriger. BME was ook minder vertekend, nauwkeuriger en de schattingen waren beter gecorreleerd met de geobserveerde laboratorium elektrische geleidbaarheid waardes dan de twee kriging technieken. Tenslotte, gebruik maken van probabilistische zachte gegevens in plaats van intervallen in de BME procedure verminderde de vertekening en vergrootte de nauwkeurigheid van de voorspellingen.

De verhandeling sluit af met een concluderende sectie (hoofdstuk 9) die de voornaamste resultaten benadrukt en aanbevelingen geeft voor verder onderzoek.

RESUME

Evaluation de la variabilité spatio-temporelle de la salinité du sol par des méthodes statistiques, géostatistiques, et d'entropie maximale bayesienne.

par **ir Ahmed DOUAIK**

Objectifs

Cette thèse a pour objectif le développement de méthodes statistiques et géostatistiques pour l'analyse de données spatio-temporelles et leur application à la salinité du sol. Un intérêt spécial a été porté sur comment caractériser la salinité du sol. Aussi, le focus était sur trois groupes de méthodes d'analyse, à savoir les méthodes statistiques classiques, les méthodes géostatistiques classiques et les méthodes géostatistiques modernes, c'est à dire l'entropie maximale bayesienne. Quoique les méthodes en question aient été appliquées à la salinité du sol, elles pourront être appliquées également à d'autres domaines de recherche.

En résumé, les deux principaux objectifs de notre étude étaient :
- suivi de l'évolution dans le temps du pattern spatial de la salinité du sol par l'utilisation de méthodes statistiques classiques ;
- cartographie de la salinité du sol aux endroits et aux instants non échantillonnés par l'utilisation des méthodes géostatistiques classiques et modernes comme l'entropie maximale bayesienne.

Aperçu

Après une introduction générale (chapitre 1), une étude bibliographique sur la salinité du sol a été présentée (chapitre 2). Il ressort de ce chapitre que la salinité du sol peut être mesurée soit au laboratoire soit au champ. Au laboratoire, on peut déterminer la conductivité électrique d'un extrait de pâte saturée (EC_e) ou d'un extrait d'un rapport sol:eau donné, par exemple 1:1, 1:2.5, 1:5, etc. Quant au champ, la salinité du sol peut être évaluée par la mesure de la conductivité électrique apparente (EC_a) via un quadripôle ou un instrument d'induction électromagnétique. Dans notre étude nous avons opté pour l'utilisation de $EC_{2.5}$ et EC_a mesurée par un quadripôle. Ces propriétés du sol ont été mesurées en divers sites et les

mesures ont été répétées dans le temps entre novembre 1994 et mars 2001. Le site d'étude et les données ont été décrits au chapitre 3.

Concernant l'analyse des données spatio-temporelles (chapitre 4), on a distingué trois groupes de méthodes. D'abord les méthodes statistiques classiques, qui ignorent les coordonnées dans l'espace et dans le temps, ont été présentées. Elles regroupent le coefficient de variation, l'analyse de la variance et l'analyse des composantes de variance, la stabilité temporelle, et la variation spatiale dynamique. Les deux dernières méthodes ont été appliquées à nos données (chapitres 5 et 6). Le deuxième groupe de méthodes est l'approche géostatistique classique incluant l'analyse structurale (variogramme et fonction de covariance) et l'interpolation (krigeage) dans le domaine spatio-temporel. Le troisième groupe concerne les méthodes géostatistiques modernes, c'est à dire l'entropie maximale bayesienne. Les méthodes géostatistiques classiques et modernes ont été appliquées à nos données (chapitres 7 et 8).

Méthodes statistiques classiques

Les valeurs de la conductivité électrique obtenues, en différents sites et instants, à partir de l'analyse au laboratoire ($EC_{2.5}$) et d'un quadripôle (EC_a) ont été comparées et analysées pour vérifier la stabilité temporelle des patterns spatiaux de la salinité du sol. Les méthodes statistiques utilisées ont été la corrélation de rang de Spearman, la technique des différences relatives, le test de signification des paramètres de la régression liant les données d'une période à celle de la période précédente, le test t de Student par paires, et le test de signification de la variation spatiale dynamique. Basé sur la corrélation de Spearman, on a trouvé que le classement des différents sites persiste avec le temps. On a identifié, basé sur les différences relatives, trois classes de salinité ayant différentes stabilités temporelles. Les sites de faible salinité sont les plus stables avec le temps et appartiennent aux zones de stagnation de l'eau et de lessivage des sels, les sites de forte salinité sont les moins stables et appartiennent aux zones d'accumulation des sels, alors que les sites représentatifs de la salinité moyenne du champ ont une stabilité temporelle intermédiaire et appartiennent aux trois zones de processus liés à la salinisation.

Une variation spatiale dynamique a été identifiée pour trois paires de temps alors que pour les autres paires, il n'y a pas eu de changement dans la salinité ou bien ce changement a été constant pour tous les sites entre les deux instants considérés.

Méthodes géostatistiques et d'entropie maximale bayésienne

La conductivité électrique apparente (EC_a) a été convertie en salinité du sol ($EC_{2.5}$) en utilisant les équations de calibration basées sur des modèles de régression. Ces valeurs converties ont été analysées pour leur variabilité spatio-temporelle en utilisant les variogrammes / fonctions de covariance et ont été interpolées via le krigeage dans le domaine spatio-temporel. Ce dernier a été comparé au krigeage limité au domaine spatial. En outre, deux variantes du krigeage spatio-temporel, utilisant d'un côté uniquement les valeurs observées et de l'autre aussi bien les valeurs observées que les valeurs converties, ont été comparées à la méthode de l'entropie maximale bayesienne. Cette dernière est une technique d'interpolation qui distingue entre la précision des données observées de celle des données converties. On a trouvé que la prédiction via le krigeage spatio-temporel ne différait pas de celle du krigeage limité au domaine spatial, cependant la précision du premier était plus grande. On a trouvé, aussi, que l'entropie maximale bayesienne fournit des prédictions moins biaisées, plus précises, et sont mieux corrélées aux valeurs observées que les prédictions fournies par les deux variantes du krigeage.

La dissertation se termine par une section de conclusion (chapitre 9) en résumant les principales conclusions et en présentant quelques recommandations de recherche pour le future.

REFERENCES

Abraham L and Bocskai J. (1971). *Utilization and reclamation of Hungarian salt-affected soils* (in Hungarian). OMMI: Budapest, Hungary.

Acworth RI. (1999). Investigation of dryland salinity using the electrical image method. *Aust J. Soil Res.* 37: 623-636.

Agarwal RR, Das SK, and Mehrotra CL. (1961). Interrelationship between electrical conductivity 1:5 and saturation extracts and total soluble salts in saline-alkali soils of the Gangetic alluvium in Uttar Pradesh. *Indian J. Agric. Sci.* 31: 284-294.

Agarwal OP, Rao KGGK, Chauhan HS, and Khandelwal MK. (1995). Geostatistical analysis of soil salinity improvement with subsurface drainage system. *Trans. ASAE*, 38: 1427-1433.

Anselin L. (1988). *Spatial econometrics: methods and models*. Boston: Kluwer Academic.

Aon MA, Sarena DE, Burgos JL, and Cortassa S. (2001). (Micro)biological, chemical and physical properties of soils subjected to conventional or no-till management : an assessment of their quality status. *Soil Tillage Res.* 60: 173-186.

Bechini L, Ducco G, Donatelli M, and Stein A. (2000). Modelling, interpolation, and stochastic simulation in space and time of global solar radiation. *Agric. Ecosys. Environ.* 81: 29-42.

Beckers F. (1997). *On the spatial and space-time analysis of field experiments*. PhD thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. 178 pp.

Bell TL. (1987). A space-time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res*. 92 (D8): 9631-9643.

Bennett RJ. (1975). The representation and identification of spatio-temporal systems: an example of population diffusion in North-West England. *Trans. Inst. Brit. Geogr.*, 66: 73-94.

Bhatti AU, Mulla DJ, Koehler FE, and Gurmani AH. (1991). Identifying and removing spatial correlation from yield experiments. *Soil Sci. Soc. Am. J.* 55: 1523-1528.

Bilonick RA. (1985). The space-time distribution of sulphate deposition in the Northern US. *Atmos. Environ.* 19: 1829-1845.

Bilonick RA. (1988). Monthly hydrogen ion deposition maps for the North Eastern US from July 1982 to September 1984. *Atmos. Environ.* 22: 1909-1924.

Bodrogközy G. (1965). Ecology of the halophilic vegetation of the Pannonicum. II. Correlation between alkali ("szik") plant communities and genetic soil classification in the northern Hortobagy. *Acta Botan. Hung.* 11: 1-51.

Bogaert P. (1996a). *Geostatistics applied to the analysis of space-time data*. PhD thesis. Université Catholique de Louvain, Louvain-la-Neuve, Belgium. 176p.

Bogaert P. (1996b). Comparison of kriging in a space-time context. *Math. Geol.* 28: 73-86.

Bogaert P. (2002). Spatial prediction of categorical variables: the Bayesian Maximum Entropy. *Stoch. Environ. Res. Risk Assess.* 16: 425-448.

Bogaert P and Christakos G. (1997). Stochastic analysis of spatio-temporal solute content measurements using a regression model. *Stoch. Hydrol. Hydraul.* 11: 267-295.

Bogaert P and D'Or D. (2002). Estimating soil properties from thematic soil maps: the Bayesian Maximum Entropy. *Soil Sci. Soc. Am. J.* 66: 1492-1500.

Bourgine B, Chiles JP, and Watremez P. (2001). Space-time modelling of sand beach data: a geostatistical approach. In: Monestiez P, Allard D, and Froidevaux R (eds). *GeoEnv III: Geostatistics for Environmental applications*. Kluwer Academic Publishers: Dordrecht, The Netherlands, pp: 101-101.

Brady NC and Weil RR. (1999). *The nature and properties of soils*. 12th ed, Prentice Hall: Upper Saddle River, New Jersey.

Brandsma AS and Ketellapper RH. (1979). Further evidence on alternative procedures for testing of spatial autocorrelation amongst regression disturbances. In: Bartels CPA and Ketellaper RH (eds). *Exploratory and explanatory statistical analysis of spatial data*. Martinus Nijhoff: Hingham, MA, pp: 113-136.

Bruckler L, De Cockborne AM, Renault P, and Claudot P. (1997). Spatial and temporal variability of nitrate in irrigated salad crops. *Irrig. Sci.* 17: 53-61.

Brundson C, Fotheringham AS and Charlton ME. (1996). Geographically weighted regression: a method for exploring spatial non-stationarity. *Geog. Analysis*, 28: 281-298.

Brus DJ, Knotters M, Van Dooremolen WA, Van Kernebeek P, and Van Seeters RJM. (1992). The use of electromagnetic measurements of apparent soil electrical conductivity to predict the boulder clay depth. *Geoderma*, 55: 79-93.

Burgess TM, Webster R, and McBratney AB. (1981). Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *J. Soil Sci.* 32: 643-659.

Bui EN, Smettem KRJ, Moran CJ, and Williams J. (1996). Use of soil survey information to assess regional salinization risk using geographic information systems. *J. Environ. Qual.* 25: 433-439.

Bui EN, Krogh L, Lavado RS, Nachtergaele FO, Toth T, and Fitzpatrick RW. (1998). Distribution of sodic soils: the world scene. In: Sumner ME and Naidu R. (eds). *Sodic soils: distribution, properties, management and environmental consequences*. Oxford University Press: New York, pp: 19-33.

Burrough PA. (1989). Fuzzy mathematical methods for soil survey and land evaluation. *J. Soil Sci.* 40: 477-492.

Campbell CG, Ghodrati M, and Garrido F. (2001). Temporal consistency of solute transport in a heterogeneous field plot. *Soil Sci.* 166: 491-506.

Campbell DJ, Kinnigurgh DG, and Beckett PHT. (1989). The soil solution chemistry of some Oxfordshire soils: temporal and spatial variability. *J. Soil Sci.* 40: 321-339.

Cannon ME, McKenzie RC, and Lachapelle G. (1994). Soil salinity mapping with electromagnetic induction and satellite-based navigation methods. *Can J. Soil Sci.* 74: 335-343.

Carter MR (ed). (1993). *Soil sampling and methods of analysis*. Lewis Publishers: Boca Raton, Florida.

Carter MR and Pearen JR. (1985). General and spatial variability of Solonetzic soils in north central Alberta. *Can J. Soil Sci.* 65: 157-167.

Carter LM, Rhoades JD, and Chesson JH. (1993). Mechanization of soil salinity assessment for mapping. *Proc. ASAE Winter Meetings*, Chicago, IL, Dec 12-17, ASAE paper n° 931557, ASAE, St Joseph, MI.

Castrignanò A, Lopez G, Stelluti M. (1994). Temporal and spatial variability of electrical conductivity, Na content and sodium adsorption ratio of saturation extract measurements. *Europ. J. Agron.* 3: 221-226.

Castrignanò A, Maiorana M, Fornaro F, and Lopez N. (2002). 3D spatial variability of soil strength and its change over time in a durum wheat field in southern Italy. *Soil Tillage Res.* 65: 95-108.

Cetin M and Kirda C. (2003). Spatial and temporal changes of soil salinity in a cotton field irrigated with low-quality water. *J. Hydrol.* 272: 238-249.

Chevallier T, Voltz M, Blanchart E, Chotte JL, Eschenbrenner V, Mahieu M, and Albrecht A. (2000). Spatial and temporal changes of soil carbon after establishment of a pasture on a long-term cultivated vertisol (Martinique). *Geoderma*, 94: 43-58.

Chang C, Sommerfeldt TG, and Entz T. (1988). Soil salinity and sand content variability determined by two statistical methods in an irrigated saline soil. *Can. J. Soil Sci.* 68: 209-221.

Chiles JP and Delfiner P. (1999). *Geostatistics: modeling spatial uncertainty*. Wiley: New York.

Choi K-M, Christakos G, and Serre ML. (1998). Recent developments in vectorial and multipoint BME analysis. In: Buccianti A, Nardi G, and Potenza R (eds). *Proc. 4th Conf. of IAMG*, Naples, Italy, pp: 91-96.

Choi K-M, Serre ML, and Christakos G. (2003). Efficient mapping of California mortality fields at different spatial scales. *J. Expos. Anal. Environ. Epidem.* 13: 120-133.

Christakos G. (1990). A Bayesian/maximum entropy view to the spatial estimation problem. *Math. Geol.* 22: 763-776.

Christakos G. (1991). Certain classes of spatiotemporal random fields with applications to space-time data processing. *IEEE Trans. Syst. Man Cyber.* 21: 261-275.

Christakos G. (1992). *Random field models in earth sciences*. Academic Press: San Diego, California.

Christakos G. (1998). Spatiotemporal information systems in soil and environmental sciences. *Geoderma*, 85: 141-179.

Christakos G. (2000). *Modern spatiotemporal geostatistics*. Oxford University Press, New York.

Christakos G and Bogaert P. (1996). Spatiotemporal analysis of springwater ion processes derived from measurements at the Dyle basin in Belgium. *IEEE Trans. Geosci. Remote Sens.* 34: 1-17.

Christakos G, Bogaert P, and Serre M.L. (2002). *Temporal GIS: advanced functions for field-based applications*. Springer-Verlag: New York.

Christakos G and Hristopoulos DT. (1998). *Spatiotemporal environmental health modelling: a tractatus stochasticus*. Kluwer Academic Publisher, Boston, MA.

Christakos G, Hristopoulos DT, and Serre ML. (1999). BME studies of stochastic differential equations representing physical laws - Part I. In: Lippard SJ, Naess A, and Sinding-Larsen R (eds). *Proc. 5th Conf. of IAMG*, Trondheim, Norway, volume1, pp: 63-68.

Christakos G and Li X. (1998). Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators. *Math. Geol.* 30: 435-462.

Christakos G and Raghu VR. (1996). Dynamic stochastic estimation of physical variables. *Math. Geol.* 28: 341-365.

Christakos G and Serre M.L. (2000). BME analysis of spatiotemporal particulate matter distributions in North Carolina. *Atmos. Environ.* 34: 3393-3406.

Christakos G and Vyas V. (1998). A novel method for studying population health impacts of spatiotemporal ozone distribution. *Soc. Sci. Med*. 47: 1051-1066.

Clark I, Basinger KL, and Harper WV. (1989). MUCK- a novel approach to cokriging. In: Buxton BE (ed). *Proc. Conf. on Geostatistics, Sensitivity, and Uncertainty Methods for growndwater flow and radionuclide transport modelling*. Battelle Press, Colombus, OH, pp: 473-493.

Cliff AD and Ord JK. (1981). *Spatial processes: models and applications*. Pion: London.

Comegna V and Basile A. (1994). Temporal stability of spatial patterns of soil water storage in a cultivated Vesuvian soil. *Geoderma,* 62: 299-310.

Corwin DL and Lesch SM. (2003). Application of soil electrical conductivity to precision agriculture: theory, principles, and guidelines. *Agron. J.* 95: 455-471.

Corwin DL and Rhoades JD. (1982). An improved technique for determining soil electrical conductivity–depth relations from above-ground electromagnetic measurements. *Soil Sci. Soc. Am. J.* 46: 517-520.

Corwin DL and Rhoades JD. (1984). Measurement of inverted electrical conductivity profiles using electromagnetic induction. *Soil Sci. Soc. Am. J.* 48: 288-291.

Cressie N. (1993). *Statistics for spatial data*. Wiley: New York.

Cressie N and Huang HC. (1999). Classes of nonseparable spatiotemporal stationary covariance functions. *J. Am. Stat. Assoc.* 94: 1330-1340.

D'Agostino V, Greene EA, Passarella G, and Vurro M. (1998). Spatial and temporal study of nitrate concentration in groundwater by means of coregionalisation. *Environ. Geol.* 36: 285-295.

D'Or D and Bogaert P. (2001). Fine scale soil texture estimation using soil maps and profile descriptions. In: Monestiez P, Allard D, and Froidevaux R (eds). *GeoEnv III: Geostatistics for Environmental applications*. Kluwer Academic Publishers: Dordrecht, The Netherlands, pp: 453-462.

D'Or D and Bogaert P. (2003). Continuous-valued map reconstruction with the Bayesian Maximum Entropy. *Geoderma*, 112: 169-178.

D'Or D Bogaert P, and Christakos G. (2001). Application of the BME approach to soil texture mapping. *Stoch. Environ. Res. Risk Assess.* 15: 87-100.

Da Silva AP, Nadler A, and Kay BD. (2001). Factors contributing to temporal stability in spatial patterns of water content in the tillage zone. *Soil Tillage Res.* 58: 207-218.

Dalton FN, Herkelrath WN, Rawlins DS, and Rhoades JD. (1984). TDR: simultaneous measurement of soil water and electrical conductivity with a single probe. *Science*, 224: 988-990.

Dalton FN and Van Genuchten MT. (1986). The TDR for measuring soil water content and salinity. *Geoderma*, 38: 237-250.

Dasberg S and Dalton FN. (1985). TDR field measurements of soil water content and electrical conductivity. *Soil Sci. Soc. Am. J.* 49: 293-297.

De Cesare L, Myers DE, and Posa D. (1997). Spatiotemporal modelling of SO_2 in Milan district. In: Baafi EY and Schofield N (eds). *Geostatistics Wollongong'96*, Kluwer Academic Publisher, Dordrecht, The Netherlands, volume 2, pp: 1310-1342.

De Cesare L, Myers D, and Posa D. (2001a). Estimating and modeling space-time correlation structures. *Stat. Proba. Letters*, 51: 9-14.

De Cesare L, Myers DE, and Posa D. (2001b). Product sum covariance for space-time modelling: an environmental application. *Environmetrics*, 12: 11-23.

De Cesare L, Posa D, and Myers DE. (2002). Fortran programs for space-time modelling. *Computers Geosci.* 28: 205-212.

De Iaco S, Myers DE, and Posa D. (2001a). Space-time analysis using a general product sum model. *Stat. Proba. Letters*, 52: 21-28.

De Iaco S, Myers DE, and Posa D. (2001b). Total air pollution and space-time modelling. In: Monestiez P, Allard D, and Froidevaux R (eds). *GeoEnv III: Geostatistics for Environmental applications*. Kluwer Academic Publisher, Dordrecht, The Netherlands, pp: 45-56.

De Iaco S, Myers DE, and Posa D. (2002). Space-time variograms and a functional form for total air pollution measurements. *Comput. Stat. Data Anal.* 41: 311-328.

De Iaco S, Myers DE, and Posa D. (2003). The linear coregionalization model and the product sum space-time variogram. *Math. Geol.* 35: 25-38.

De Iaco S, Palma M, and Posa D. (2004). Modelling and prediction of multivariate space-time random fields. *Comput. Stat. Data Anal.* 48: 525-547.

De Jong E, Ballantyne AK, Cameron DR, and Read DWL. (1979). Measurement of apparent electrical conductivity of soils by an electromagnetic induction probe to aid salinity surveys. *Soil Sci. Soc. Am. J.* 43: 810-812.

Delcourt H, Darius PL, and De Baerdemaker J. (1996). The spatial variability of some aspects of topsoil fertility in two Belgian fields. *Computers Electronics Agric.* 14: 179-196.

Diaz L and Herrero J. (1992). Salinity estimates in irrigated soils using electromagnetic induction. *Soil Sci.* 154: 151-157.

Dimitrakopoulos R and Luo X. (1994). Spatiotemporal modelling: covariances and ordinary kriging systems. In: Dimitrakopoulos R (ed). *Geostatistics for the next century*. Kluwer Academic Publisher, Dordrecht, The Netherlands, pp: 88-93.

Doolittle JA, Sudduth KA, Kitchen NR, and Indorante SJ. (1994). Estimating depths to claypans using electromagnetic induction methods. *J. Soil Water Conserv.* 49: 572-575.

Douaik A, Van Meirvenne M, and Tóth T. (2004). Spatio-temporal kriging of soil salinity rescaled from bulk soil electrical conductivity. In: Sanchez-Vila X, Carrera J, and Gomez-Hernandez J (eds). *GeoEnv IV: Geostatistics for Enironmental Applications.* Kluwer Academic Publishers: Dordrecht, The Netherlands, pp: 413-424.

Egbert GD and Lettenmaeir DP. (1986). Stochastic modelling of the space-time structure of atmospheric chemical deposition. *Water Resour. Res.* 22: 165-179.

Ehrenfeld JG, Han X, Parsons WFJ, and Zhu W. (1997). On the nature of environmental gradients: temporal and spatial variability of soils and vegetation in the New Jersey pinelands. *J. Ecol.* 85: 785-798.

El Oumri M and Vieillefon J. (1983). Etude expérimentale de la conductivité électrique globale des sols : application à l'estimation de leur salinité. *Cahiers ORSTOM*, *Pédologie*, 20: 91-108.

Ettema CH, Coleman DC, Vellidis G, Lowrance R, and Rathbun SL. (1998). Spatiotemporal distributions of bacterivorous nematodes and soil resources in a restored riparian wetland. *Ecology*, 79: 2721-2734.

Ettema CH, Rathbun SL, and Coleman DC. (2000). On spatiotemporal patchiness and the existence of 5 species of Chronogaster (Nematoda Chronogasteridae) in a riparian wetland. *Oecologia*, 125: 444-452.

Fagroud M. (2001). *L'application des modèles de simulation stochastique pour l'optimisation de l'expérimentation au champ et de l'échantillonnage des sols*. Thèse de doctorat, Université de Gand, Gand, Belgique.

Fagroud M and Van Meirvenne M. (2002). Accounting for soil spatial autocorrelation in the design of experimental trials. *Soil Sci. Soc. Am. J.* 66: 1134-1142.

Faltman GT, Englund EJ and Yfantis AA. (1987). Geostatistical approaches to the design of sampling regimes. In: Keith LH (ed.). *Principles of environmental sampling.* ACS Professional Reference Book. American Chemistry Society, pp: 73-92.

Farley RA and Fitter AH. (1999). Temporal and spatial variation in soil resources in a deciduous woodland. *J. Ecol.* 87: 688-696.

Figueira R, Sousa AJ, Pacheco AMG, and Catarino F. (1999). Space-time geostatistical modelling: a case study of sea-salt measured on lichens. In: Gomez-Hernandez J, Soares A, and Froidevaux R (eds).*GeoEnv II: Geostatistics for Enironmental Applications.* Kluwer Academic Publisher, Dordrecht, The Netherlands, pp: 53-64.

Figueira R, Sousa AJ, Pacheco AMG, and Catarino F. (2001). Use of secondary information in space-time statistics for biomonitoring studies of saline deposition. *Environmetrics*, 12: 203-217.

Fowler DB and Hamm JW. (1980). Crop response to saline soil conditions in the parkland area of Saskatchewan. *Can. J. Soil Sci.* 60: 439-449.

Gómez-Plaza A, Alvarez-Rogel J, Albaladejo J, and Castillo VM. (2000). Spatial patterns and temporal stability of soil moisture across a range of scales in a semi-arid environment. *Hydrol. Process.* 14: 1261-1277.

Goovaerts P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press, New York.

Goovaerts P and Chiang CN. (1993). Temporal persistence of spatial patterns for mineralizable nitrogen and selected soil properties. *Soil Sci. Soc. Am. J.* 57: 372-381.

Goovaerts P and Sonnet P. (1993). Study of spatial and temporal variations of hydrochemical variables using factorial kriging analysis. In Soares (ed). *Geostatistics Troia'92*, vol. 2, Kluwer: Dordrecht, The Netherlands, pp: 745-756.

Gorres JH, Dichiaro MJ, Lyons JB, and Amador JA. (1997). Spatial and temporal patterns of soil biological activity in a forest and an old field. *Soil Biol. Biochem.* 30: 219-230.

Gotway CA and Cressie N. (1990). A spatial analysis of variance applied to soil-water infiltration. *Water Resour. Res*. 26: 2695-2703.

Grayson RB and Western AW. (1998). Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *J. Hydrol*. 207: 68-82.

Guo D, Mou P, Jones RH, and Mitchell RJ. (2002). Temporal changes in spatial patterns of soil moisture following disturbance: an experimental approach. *J. Ecol.* 90: 338-347.

Halvorson AD and Rhoades JD. (1974). Assessing soil salinity and identifying potential saline-seep areas with field soil resistance measurements. *Soil Sci. Soc. Am. Proc.* 38: 576-581.

Halvorson AD and Rhoades JD. (1976). Field mapping soil conductivity to delineate dryland saline seeps with four-electrode technique. *Soil Sci. Soc. Am. J.* 40: 571-575.

Halvorson AD, Rhoades JD, and Reule CA. (1977). Soil salinity–four-electrode resistivity relationships for soils of the Northern Great Plains. *Soil Sci. Soc. Am. J.* 41: 966-971.

Heuvelink GBM, Musters P, and Pebesma EJ. (1997). Spatiotemporal kriging of soil water content. In: Baafi EY and Schofield N (eds). *Geostatistics Wollongong'96*, Kluwer Academic Publisher, Dordrecht, The Netherlands, volume 2, pp: 1020-1030.

Hogg TJ and Henry JL. (1984). Comparison of 1:1 and 1:2 suspensions and extracts with the saturation extract in estimating salinity in Saskatchewan soils. *Can. J. Soil Sci.* 64: 699-704.

James SE, Partel M, Wilson SD, and Peltzer DA. (2003). Temporal heterogeneity of soil moisture in grassland and forest. *J. Ecol.* 91: 234-239.

Jaynes DB, Colvin TS, and Ambuel J. (1993). Soil type and crop yield determinations from ground conductivity surveys. *Proc. ASAE Winter Meetings*, Chicago, IL, Dec 12-17, ASAE paper n° 933552, ASAE: St Joseph, MI.

Jaynes DB, Novak JM, Moorman TB, and Cambardella CA. (1995). Estimating herbicide partition coefficients from electromagnetic induction measurements. *J. Environ. Qual.* 24: 36-41.

Job JO, Tabbagh A, and Hachicha M. (1995). Détermination par méthode électromagnétique de la concentration en sel d'un sol irrigué. *Can. J. Soil Sci.* 75: 463-469.

Johnson CK, Eskridge KM, Wienhold BJ, Doran JW, Peterson GA, and Buchleiter GW. (2003). Using electrical conductivity classification and within-field variability to design field-scale research. *Agron. J.* 95: 602-613.

Journel AG and Huibregts CJ. (1978). *Mining geostatistics*. Academic Press: New York.

Jurinak JJ and Suarez DL. (1990). The chemistry of salt-affected soils and waters. In: Tanji KK (ed). *Agricultural salinity assessment and management*. ASCE Manual and Reports on Engineering Practice n° 71, ASCE, New York, pp: 42-63.

Kachanoski RG and De Jong E. (1988). Scale dependence and the temporal persistence of spatial patterns of soil water storage. *Water Resour. Res.* 24: 85-91.

Kachanoski RG, Gregorich EG, and Van Wesenbeeck IJ. (1988). Estimating spatial variations of soil water content using no contacting electromagnetic inductive methods. *Can. J. Soil Sci.* 68: 715-722.

Kamaliddin AR, Sharif B, and Hardan A. (1961). Electrical conductivity relationships for soil pastes and 1:1 soil:water suspensions and their extracts. *Proc. Symp. Salinity Problems in Arid Zones*, Tehran, pp: 299-303.

Kamphorst A and Bolt GH. (1978). Saline and sodic soils. In: Bolt GH and Bruggenwert MGM (eds). *Soil chemistry: a basic elements*. Second edition. Elsevier Scientific, Amsterdam: The Netherlands, pp: 171-191.

Kenny EA, Hall JW, and Wang C. (2002). Temporal trends in soil physical properties at a soil quality benchmark site in British Columbia, Canada. *Proc. 17th World Cong. Soil Sci.* paper 291.

Kobayachi K and Us Salam M. (2000). Comparing simulated and measured values using mean squared deviation and its components. *Agron. J.* 92: 345-352.

Kolovos A, Christakos G, and Serre ML. (2000). Incorporation of physical laws and other forms of knowledge in spatiotemporal prediction of hydrologic processes. *EOS Trans. AGU*, San Francisco, CA, Dec 15-19.

Kyriakidis PC and Journel AG. (1999). Geostatistical space-time models: a review. *Math. Geol.* 31: 651-684.

Lark RM. (2002). Robust estimation of the pseudo-cross variogram for cokriging soil properties. *Europ. J. Soil Sci.* 53: 253-270.

Lee Y-M and Ellis JH. (1997). On the equivalence of kriging and maximum entropy estimators. *Math. Geol.* 29: 131-151.

Lesage JP. (1999). Applied econometrics using Matlab. The toolbox and the manual are available on the *internet*.

Lesch SM, Herrero J and Rhoades JD. (1998). Monitoring for temporal changes in soil salinity using electromagnetic induction techniques. *Soil Sci. Soc. Am. J.* 62: 232-242.

Lesch SM, Rhoades JD, and Corwin DL. (2000). ESAP-95 version 2.01 R: user manual and tutorial guide. *Research Report 146*, USDA-ARS George E. Brown Jr. Salinity Laboratory, Riverside, CA.

Lesch SM, Rhoades JD, Lund LJ, and Corwin DL. (1992). Mapping soil salinity using calibrated electromagnetic measurements. *Soil Sci. Soc. Am. J.* 56: 540-548.

Lesch SM, Rhoades JD, Strauss DJ, Lin K, and Co MAA. (1995a). The ESAP user manual and tutorial guide. Version 1.0. *USSL Research Report 138*.

Lesch SM, Strauss DJ, and Rhoades JD. (1995b). Spatial prediction of soil salinity using electromagnetic induction techniques.1 Statistical prediction models: a comparison of multiple linear regression and cokriging. *Water Resour. Res.* 31: 373-386.

Lesch SM, Strauss DJ, and Rhoades JD. (1995c). Spatial prediction of soil salinity using electromagnetic induction techniques. 2 An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation. *Water Resour. Res.* 31: 387-398.

Lund ED, Christy D, and Drummond PE. (1999). Applying soil electrical conductivity technology to precision agriculture. In: Robert PC et al. (eds). *Proc. 4th Internat. Conf. Precision Agriculture.* ASA-CSSA-SSSA, Madison, WI, pp: 1089-1100.

Martínez-Fernández J and Ceballos A. (2003). Temporal stability of soil moisture in a large-field experiment in Spain. *Soil Sci. Soc. Am. J.* 67: 1647-1656.

MathWorks. (1999). *Using Matlab*, version 5. The MathWorks Inc., Natick : MA.

McBride RA, Gordon AM, and Shrive SC. (1990). Estimating forest soil quality from terrain measurements of apparent electrical conductivity. *Soil Sci. Soc. Am. J.* 54: 290-293.

McClave JT and Sincich T. (2000). *Statistics*. 8th ed. Prentice Hall, Upper Saddle River: NJ.

McKenzie RC, Chomistek W, and Clark NF. (1989). Conversion of electromagnetic inductance readings to saturated paste extract values in soils for different temperature, texture, and moisture conditions. *Can. J. Soil Sci.* 69: 25-32.

McKenzie RC, Sprout CH, and Clark NF. (1983). The relationship of the yield of irrigated barley to soil salinity as measured by several methods. *Can. J. Soil Sci.* 63: 519-528.

McNeill JD. (1980). Electromagnetic terrain conductivity measurement at low induction numbers. *Technical Note TN-6*, Geonics limited, Mississauga, Ontario, Canada.

Metternicht GI. (1998). Fuzzy classification of JERS-1 SAR data: an evaluation of its performance for soil salinity mapping. *Ecol. Model.* 111: 61-74.

Metternicht GI. (2003). Categorical fuzziness: a comparison between crisp and fuzzy class boundary modelling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios. *Ecol. Model.* 168: 371-389.

Moameni A and Stein A. (2002). Modelling spatiotemporal changes in soil salinity and waterlogging in the Marvdasht plain, Iran. *Proc. 17th World Cong. Soil Sci.* Bangkok, Thailand, August 2002, Symposium 48, paper 674.

Mohanty BP and Skaggs T.H. (2001). Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Adv. Water Resour*. 24: 1051-1067.

Moran PAP. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37: 17-23.

Mostafa MM and Yomota A. (1998). Use of a covariance variogram to investigate influence of subsurface drainage on spatial variability of soil-water properties. *Agric. Water Manag.* 37: 1-19.

Myers DE. (1991). Pseudo-cross variograms, positive-definiteness and cokriging. *Math. Geol.* 23: 805-816.

Myers DE and Journel AG. (1990). Variograms with zonal anisotropies and non invertible kriging systems. *Math. Geol.* 22: 779-785.

Myers RH. (1986). *Classical and modern regression with applications*. Duxbury Press: Boston, MA.

Oosterbaan RJ. (1997). Saltmod: a tool for interweaving of irrigation and drainage for salinity control. In: Snellen WB. (ed.). *Towards integration of irrigation and drainage management*. Proceedings of the Jubilee Symposium at the occasion of the 40th anniversary of ILRI, Wageningen, The Netherlands, pp: 43-49.

Papritz A and Fluhler H. (1994). Temporal change of spatially autocorrelated soil properties: optimal estimation by kriging. *Geoderma*, 62: 29-43.

Papritz A, Kunsch HR, and Webster R. (1993). On the pseudo-cross variogram. *Math. Geol.* 25: 1015-1026.

Peng W. (1998). Synthetic analysis for extracting information on soil salinity using remote sensing and GIS: a case study of Yanggao basin in China. *Environ. Manag.* 22: 153-159.

Petitgas P. (1997). Sol egg distributions in space and time characterized by a geostatistical model and its estimation variance. *ICES J. Marine Sci.* 54: 213-225.

Petrone RM, Price JS, Carey SK, and Waddington JM. (2004). Statistical characterization of the spatial variability of soil moisture in a cutover peatland. *Hydrol. Process.* 18: 41-52.

Pittman JJ, Kress MW, and Zhang H. (2001). Comparison of two soil salinity extraction methods. In: Pittman JJ (ed). *Proc. 8th Internat. Petroleum Environ. Conf.* Houston, Texas, November 6-9, 2001.

Read DWL and Cameron DR. (1979).Relationship between salinity and Wenner resistivity for some dryland soils. *Can. J. Soil Sci.* 59: 381-385.

Reichardt K, Bacchi OOS, Villagra MDLM, Turrati AL, and Pedrosa ZO. (1993). Hydraulic variability in space and time in a dark red latosol of the tropics. *Geoderma*, 60: 159-168.

Rengasamy P, Greene RSB, Ford GW, and Mehammi AH. (1984). Identification of dispersive behaviour and the management of red-brown earths. *Aust. J. Soil Res.* 22: 413-431.

Rhoades JD. (1981). Predicting bulk soil electrical conductivity versus saturation paste extract electrical conductivity calibrations from soil properties. *Soil Sci. Soc. Am. J.* 45: 42-44.

Rhoades JD. (1992). Instrumental field methods of salinity appraisal. In: Topp GC, Reynolds WD, and Green RE. (eds). *Advances in measurement of soil physical properties: bringing theory into practice*. SSSA special publication n° 30, pp: 231-248.

Rhoades JD. (1993). Electrical conductivity methods for measuring and mapping soil salinity. *Adv. Agron.* 49: 201-251.

Rhoades JD, Chanduvi F, and Lesch SM. (1999a). Soil salinity assessment: methods and interpretation of electrical conductivity measurements. *FAO Irrigation and Drainage Paper n° 57*, Rome, Italy.

Rhoades JD and Corwin DL. (1981). Determining soil electrical conductivity-depth relations using an inductive electromagnetic soil conductivity meter. *Soil Sci. Soc. Am. J.* 45: 255-260.

Rhoades JD and Corwin DL. (1999b). Geospatial measurements of soil electrical conductivity to assess soil salinity and diffuse salt loading from irrigation. In: Corwin DL, Loague K, and Ellsworth TR (eds). *Assessment of non-point source pollution in the vadose zone*. Geophysical Monograph 108, AGU, Washington, DC, pp: 197-215.

Rhoades JD and Halvorson AD. (1976). Detecting and delineating saline seeps with soil resistance measurements. *Proc. Saline Seep Control Symp.* Montana State University, Bozeman, Bulletin n° 1132, pp: 19-34.

Rhoades JD and Ingvalson RD. (1971). Determining salinity in field soils with soil resistance measurements. *Soil Sci. Soc. Am. Proc.* 35: 54-60.

Rhoades JD, Kaddah MT, Halvorson AD, and Prather RJ. (1977). Establishing soil electrical conductivity-salinity calibrations using four-electrode cells containing undisturbed soil cores. *Soil Sci.* 123: 137-141.

Rhoades JD, Lesch SM, LeMert RD, and Alves WJ. (1997). Assessing irrigation/drainage/salinity management using spatially referenced salinity measurements. *Agric. Water Manag.* 35: 147-165.

Rhoades JD, Lesch SM, Shouse PJ, and Alves WJ. (1989). New calibrations for determining soil electrical conductivity–depth relations from electromagnetic measurements. *Soil Sci. Soc. Am. J.* 53: 74-79.

Rhoades JD, Shouse PJ, Alves WJ, Manteghi NA, and Lesch SM. (1990). Determining soil salinity from soil electrical conductivity using different models and estimates. *Soil Sci. Soc. Am. J.* 54: 46-54.

Rhoades JD and Van Schiffgaarde J. (1976). An electrical conductivity probe for determining soil salinity. *Soil Sci. Soc. Am. J.* 40: 647-651.

Richards LA. (ed.) (1954). Diagnosis and improvement of saline and alkali soils. USSL staff *Agriculture handbook n° 60*, USDA, Washington DC, 160p.

Rodriguez-Iturbe I and Eagelson PS. (1987). Mathematical models of rainstorm events in space and time. *Water Resour. Res.* 23: 181-190.

Rodriguez-Iturbe I and Mejia JM. (1974). The design of rainfall networks in time and space. *Water Resour. Res.* 10: 713-728.

Rouhani S, Ebrahimpour RM, Yaqub I, and Gianella E. (1992a). Multivariate geostatistical trend detection and network evaluation of space-time acid deposition data. – I. Methodology. *Atmos. Environ.* 26: 2603-2614.

Rouhani S, Ebrahimpour RM, Yaqub I, and Gianella E. (1992b). Multivariate geostatistical trend detection and network evaluation of space-time acid deposition data. –II. Application to NADP/NTN data. *Atmos. Environ.* 26A: 2615-2626.

Rouhani S and Hall TJ. (1989). Space-time kriging of groundwater data. In: Armstrong M (ed). *Geostatistics*, Kluwer Academic Publisher, Dordrecht, The Netherlands, volume 2, pp: 639-651.

Rouhani S and Myers DE. (1990). Problems in space-time kriging of geohydrological data. *Math. Geol.* 22: 611-623.

Rouhani S and Wackernagel H. (1990). Multivariate geostatistical approach to space-time data analysis. *Water Resour. Res.* 26: 585-591.

SAS Institute. (1993). *SAS/STAT user's guide, version 6*, 2nd ed. SAS Institute, Cary : NC.

Schume H, Jost G, and Katzensteiner K. (2003). Spatiotemporal analysis of the soil water content in a mixed Norway spruce (Picea abies (L) Karst.)-European beech (Fagus sylvatica L.) stand. *Geoderma*, 112: 273-287.

Serre ML. (1999). *Environmental spatiotemporal mapping and ground water flow modeling using the Bayesian maximum entropy and space transformation methods*. PhD thesis, University of North Carolina, Chapel Hill, NC. 241p.

Serre ML, Bogaert P, and Christakos G. (1998). Latest Computational Results in Spatiotemporal Prediction Using the Bayesian Maximum Entropy Method. In Buccianti A, Nardi G, and Potenza R. (eds), *Proc. 5th Conf. of IAMG*, vol. 1, pp: 117-122, De Frede Editore: Napoli.

Serre ML and Christakos G. (1999a). Modern geostatistics: computational BME in the light of uncertain physical knowledge – The Equus Beds study. *Stoch. Environ. Res. Risk Assess.* 13: 1-26.

Serre ML and Christakos G. (1999b). BME studies of stochastic differential equations representing physical laws - Part II. In: Lippard SJ, Naess A, and Sinding-Larsen R (eds). *Proc. 5th Conf. of IAMG*, Trondheim, Norway, vol. 1, pp: 93-98.

Serre ML, Christakos G, Howes J, and Abdel-Rehiem AG. (2001). Powering an Egyptian air quality information system with the BME space-time analysis toolbox: results from the Cairo baseline year study. In: Monestiez P, Allard D, and Froidevaux R (eds). *GeoEnvIII: Geostatistics for Enironmental Applications.* Kluwer Academic Publisher, Dordrecht, The Netherlands, pp: 91-101.

Shannon CE. (1948). A mathematical theory of communication. *Bell System Tech. J.* 27: 379-423.

Shaw RJ, Hughes KK, Dowling AJ, and Thorburn PJ. (1986). Principles of landscape, soil, and water salinity-processes and management options. Part A. In: *Landscape, soil and water salinity. Proc Burdekin Regional Salinity Workshop*, Ayr, Queensland, QDPI Publication n° QC86003, Brisbane, Australia.

Shi Z, Wang K, Bailey JS, Jordan C, and Higgins AH. (2002). Temporal changes in the spatial distributions of some soil properties on a temperate grassland site. *Soil Use Manag.* 18: 353-362.

Si BC. (2002). Spatial and statistical similarities of local soil water fluxes. *Soil Sci. Soc. Am. J.* 66: 753-759.

Snepvangers JJJC, Heuvelink GBM, and Huisman JA. (2003). Soil water content interpolation using spatiotemporal kriging with external drift. *Geoderma*, 112: 253-271.

Soil and Plant Analysis Council. (1992). *Handbook on reference methods for soil analysis*. Georgia University Station, Athens GA., 202 p.

Solow AR and Gorelick SM. (1986). Estimating monthly streamflow values by cokriging. *Math. Geol.* 18: 785-810.

Spaargaren OC. (1994). *World reference base for soil resources*. ISSS-ISRIC-FAO, Rome, Italy.

SSSA. (1984). *Glossary of soil science terms*. SSSA Inc., Madison: Wisconsin.

Stein A. (1998). Analysis of space-time variability in agriculture and the environment with geostatistics. *Stat. Neerlandica*, 52: 18-41.

Stein A, Hoosbeek MR, and Sterk G. (1997). Space-time statistics for decision support to smart farming. In: Lake JV, Bock JR and Goode JA. (eds). *Precision agriculture: spatial and temporal variability of environmental quality*. Ciba foundation symposium n° 210, Wiley: Chichester, pp: 120-133.

Stein A, Kocks CG, Zadocks JC, Frinking HD, Ruissen MA, and Myers DE. (1994). A geostatistical analysis of the spatiotemporal development of downy mildew epidemics in cabbage. *Phytopathology*, 84: 1227-1239.

Stein A, Van Groenigen JW, Jeger MJ, and Hoosbeek MR. (1998). Space-time statistics for environmental and agricultural related phenomena. *Environ. Ecol. Stat.* 5: 155-172.

Stein A and Sterk G. (1999). Modelling space and time dependence in environmental studies. *JAG*, 1: 109-121.

Sterk G and Stein A. (1997). Mapping wind-blown mass transport by modeling variability in space and time. *Soil Sci. Soc. Am. J.* 61: 232-239.

Sudduth KA, Drummond ST, and Kitchen NR. (2001). Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. *Computers Electronics Agric.* 31: 239-264.

Sudduth KA, Kitchen NR, Hughes DF, and Drummond ST. (1995). Electromagnetic induction sensing as an indicator of productivity on claypan soils. In: Robert PG, Rust RIH, and Larson WE (eds). *Proc. 2nd Internat. Conf. on Site-Specific Management for Agric. Systems*, pp: 671-681.

Swain PH, Richards JA, and Lee T. (1985). Multisource data analysis in remote sensing and geographic information processing. *Proc. 11th Internat. Symp. on Machine Processing of Remote Sensed Data*, Purdue University, LARS technical report n° 062685, pp: 211-218.

Szabolcs I. (1971). *European solonetzic soils and their reclamation*. Budapest, Hungary.

Szabolcs I. (1974). *Salt-affected soils in Europe*. Martinus Nijhoff: The Hague, The Netherlands and RISSAC: Budapest, Hungary.

Szabolcs I. (1989). *Salt-affected soils*. CRC Press: Boca Raton, Florida.

Taylor RK, Zhang N, Schrock M, Schmidt JP, and Kluitenberg GJ. (2000). Classification of yield monitor data to determine yield potential. Presented at July 9-12, 2000, *Annual Internat. Meeting sponsored by ASAE*, paper n° 001087. ASAE, 2950 Niles Rd, St. Joseph, MI 49085-9659 USA.

Topp GC, Yanuka M, Zebchuk WD, and Zegelin S. (1988). Determination of electrical conductivity using Time Domain Reflectometry: soil and water experiments in coaxial lines. *Water Resour. Res.* 24: 945-952.

Tóth T, Csillag F, Biehl LL, and Micheli E. (1991). Characterization of semi-vegetated salt-affected soils by means of field remote sensing. *Remote Sens. Environ*. 37: 167-180.

Tóth T and Jozefaciuk G. (2002). Physicochemical properties of a solonetzic toposequence. *Geoderma*, 106: 137-159.

Tóth T and Kertsz M. (1996). Application of soil-vegetation correlation to optimal resolution mapping of solonetzic rangeland. *Arid Soil Res. Rehabil.* 10: 1-12.

Tóth T Kertsz M, and Pasztor L. (1998). New approaches in salinity/sodicity mapping in Hungary. *Agrokem. Talajtan*, 47: 76-86.

Tóth T and Kuti L. (2002). Testing alternative techniques of numerical simulation versus repeated field instrumental measurements for assessing soil salinity status in a sodic grassland. *Agrokem. Talajtan,* 51: 243-252.

Tóth T, Kuti L, Fórizs I, and Kabos S. (2001a). Changes in the factors of salt accumulation in the study site 'Nyírölapos' of Hortobágy region. *Agrokem. Talajtan,* 50: 409-426, (in Hungarian).

Tóth T, Kuti L, Kabos L, and Pasztor L. (2001b). Use of digitalized hydrogeological maps for evaluation of salt-affected soils of large areas. *Arid Land Res. Manag.* 15: 329-346.

Tóth T, Kuti L, Forizs I, Kabos S, and Douaik A. (2002a). Spatial and temporal aspects of soil salinization in a sodic grassland. In: Fez Cano A, Ortiz Silla R, and Mermut AR. (eds.). *Sustainable use and management of soils in arid and semiarid regions*, vol. 1, pp :276-288.

Tóth T, Pasztor L, Kabos S, and Kuti L. (2002b). Statistical prediction of the presence of salt-affected soils by using digitalized hydrogeological maps. *Arid Land Res. Manag.* 16: 55-68.

Tóth T and Rajkai K. (1994). Soil and plant correlations in a solonetzic grassland. *Soil Sci.* 157: 253-262.

Triantafilis J, Ahmed MF, and Odeh IOA. (2002). Application of a mobile electromagnetic sensing system (MESS) to assess cause and management of soil salinization in an irrigated cotton-growing field. *Soil Use Manag.* 18: 330-339.

Triantafilis J, Huckel IA, and Odeh IOA. (2001). Comparison of statistical prediction methods for estimating field-scale clay content using different combinations of ancillary variables. *Soil Sci.* 166: 415-427.

Triantafilis J, Laslett GM, and McBratney AB. (2000). Calibrating an electromagnetic induction instrument to measure salinity in soil under irrigated cotton. *Soil Sci. Soc. Am. J.* 64: 1009-1017.

USDA. (1996). *Keys to soil taxonomy*. 7th ed., USDA: NRCS.

Utset A and Castellanos A. (1999). Drainage effects on spatial variability of soil electrical conductivity in a vertisol. *Agric. Water Manag.* 38: 213-222.

Vachaud G, Passerat De Silans A, Balabanis P, and Vauclin M. (1985). Temporal stability of spatially measured soil water probability density function. *Soil Sci. Soc. Am. J.* 49: 822-828.

Van Es HM. (1993). Evaluation of temporal, spatial, and tillage-induced variability for parametrization of soil infiltration. *Geoderma*, 60: 187-199.

Van Es HM, Ogden CB, Hill RL, Schindelbeck RR, and Tsegaye T. (1999). Integrated assessment of space, time, and management-related variability of soil hydraulic properties. *Soil Sci. Soc. Am. J.* 63: 1599-1608.

Van Es HM, Van Es CL, and Cassel DK. (1989). Application of regionalized variable theory to large-plot field experiments. *Soil Sci. Soc. Am. J.* 53: 1178-1183.

Van Loon WKP, Perfect E, Groenevelt PH, and Kay BD. (1990).A new method to measure bulk electrical conductivity in soils with time domain reflectometry. *Can. J. Soil Sci.* 70: 403-410.

Van Meirvenne M, De Groote P, Kertesz M, Tóth T, and Hofman G. (1995). Multivariate geostatistical inventory of sodicity hazard in the Hungarian Puszta. In: Escadafal R,

Mulders M, and Thiombiano L. (eds). *Monitoring soils in the environment with remote sensing and GIS*, ORSTOM éditions: Paris, pp: 293-305.

Van Meirvenne M and Goovaerts P. (2002). Accounting for spatial dependence in the processing of multi-temporal SAR images using factorial kriging. *Internat. J. Remote Sens.* 23: 371-387.

Van Meirvenne M, Pannier J, Hofman G, and Louwagie G. (1996). Regional characterization of the long-term change in soil organic carbon under intensive agriculture. *Soil Use Manag.* 12: 86-94.

Van Pelt R and Wierenga PJ. (2001). Temporal stability of spatially measured soil matric potential probability density function. *Soil Sci. Soc. Am. J.* 65: 668-677.

Van Wesenbeeck IJ and Kachanoski RG. (1988). Spatial and temporal distribution of soil water in the tilled layer under a corn crop. *Soil Sci. Soc. Am. J.* 52: 363-368.

Van Wesenbeek I.J., Kachanoski R.G., and Rolston D.E. (1988). Temporal persistence of spatial patterns of soil water content in the tilled layer under a corn crop. *Soil Sci. Soc. Am. J.* 52: 934-941.

Vaughan PJ, Lesch SM, Corwin DL, and Cone DG. (1995). Water content effect on soil salinity prediction: a geostatistical study using cokriging. *Soil Sci. Soc. Am. J.* 59: 1146-1156.

Vyas V and Christakos G. (1997). Spatiotemporal analysis and mapping of sulfate deposition data over eastern USA. *Atmos. Environ.* 31: 3623-3633.

Wagenet RJ and Jurinak JJ. (1978). Spatial variability of soluble salt content in a Mancos shale watershed. *Soil Sci.* 126: 342-349.

Warkentin BP. (1995). The changing concept of soil quality. *J. Soil Water Conserv.* 50: 226-228.

Webster R and Oliver M. (1992). Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* 43: 177-192.

Wendroth O, Pohl W, Koszinski S, Rogasik H, Ritsema CJ, and Nielsen DR. (1999). Spatio-temporal patterns and covariance structures of soil water status in two north east German field sites. *J. Hydrol.* 215: 38-58.

White RE. (1997). *Principles and practice of soil science: the soil as a natural resource*. 3rd ed. Blackwell Science: Carlton, Victoria, Australia.

Williams BG and Hoey D. (1987). The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soil. *Aust. J. Soil Res.* 25: 21-27.

Winkel T, Rambal S, and Bariac T. (1995). Spatial variation and temporal persistence of grapevine response to a soil texture gradient. *Geoderma,* 68: 67-78.

Winsor GW and Davis JN. (1956). A critical re-examination of the conventional conductivity (PC) test for soluble salts in soils. *Glasshouse Crops Res. Institute Annual Report*, pp: 84-92.

Wollenhaupt NC, Richardson JL, Foss JE, and Doll EC. (1986). A rapid method for estimating weighted soil salinity from apparent soil electrical conductivity measured with an aboveground electromagnetic induction meter. *Can. J. Soil Sci.* 66: 315-321.

Zalasiewics JA, Mathers SJ, and Cornwell JD. (1985). The application of ground conductivity measurements to geological mapping. *Q. J. Engin. Geol.* 18: 139-148.

Zhang, R, Meyers, DE and Warrick AW. 1992. Estimation of the spatial distribution of soil chemicals using pseudo cross variograms. *Soil Sci. Soc. Am. J.* 56: 1444-1452

Zhang R, Shouse P and Yates S. (1999). Estimation of soil nitrate distribution using cokriging with pseudo-cross-variograms. *J. Environ. Qual.* 28: 424-428.

Zimmerman DL and Harville DA. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics,* 47: 223-239.

BIBLIOGRAPHY

SYMPOSIA AND CONFERENCES

- Pedometrics'01: 4th international conference on Pedometrics, International Union of Soil Science,

    * Organized by my promoter (Prof. Marc Van Meirvenne), Ghent, 19 to 21 September 2001

- International Symposium on Sustainable Use and Management of Soils in Arid and Semiarid Regions

    * Organized by University of Murcia, Cartagena, Spain, 22 to 26 September 2002

    * oral contribution under title: *Spatial and temporal aspects of soil salinization in a sodic grassland*, presented by the coauthor T. Toth.

- GeoEnv02: 4th European conference on Geostatistics with environmental applications

    * Organized by Polytechnic University of Catalonia, Barcelona, Spain, 27 to 29 November 2002

    * oral presentation entitled: *Spatio-temporal kriging of soil salinity rescaled from bulk soil electrical conductivity*

- Pedometrics'03: 5th international conference on Pedometrics, International Union of Soil Science

    * Organized by Department of Soil Sciences, University of Reading, UK, 10 to 12 September 2003

    * Oral presentation under title: *Soil salinity mapping using spatio-temporal kriging and bayesian maximum entropy*

- 8th international meeting on soils with Mediterranean type of climate

    * Organized by Association Marocaine des Sciences des Sols (Moroccan Soil Science Society), Ecole Nationale d'Agriculture (National School of Agriculture), and Institut Agronomique et Vétérinaire Hassan II, Marrakech, 9 to 11 February 2004

    * Oral presentation entitled: *Detecting and monitoring temporal changes of spatially variable soil salinity*

Conference proceedings

- Toth T., Kuti L., Forizs I., Kebos S. and **Douaik A**. (2002). Spatial and temporal aspects of soil salinization in a sodic grassland. In: Fez Cano A., Ortiz Sella R. et Mermut AR (eds). *Sustainable use and management of soils in arid and semi-arid regions*, volume 1: 276-288.

- **Douaik A**., Van Meirvenne M. and Toth T. (2003). Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy. *Book of abstracts of the 5th conference on Pedometrics*, Reading University, UK.

- **Douaik A**., Van Meirvenne M. and Toth T. (2004). Spatio-temporal kriging of soil salinity rescaled from bulk soil electrical conductivity. In : Sanchez-Vila X., Carrera J. et Gómez-Hernández J. (eds.). *GeoEnvIV: Geostatistics for Enironmental Applications*. Kluwer Academic Publishers: Dordrecht, The Netherlands, volume 13: 413-424.

- **Douaik A**., Van Meirvenne M. and Toth T. (2004). Detecting and monitoring temporal changes of spatially variable soil salinity. In: Bouabid R, Badraoui M, and Bamouh A. (eds.). *Proceeding of 8th International Meeting on Soils with Mediterranean Type of Climate*, Marrakech, pages: 114-116.

Per reviewed publications

- **Douaik A**., Van Meirvenne M., Toth T. and Serre ML. (2004). Space-Time mapping of soil salinity using probabilistic Bayesian maximum entropy. *Stochastic Environmental Research and Risk assessment*, 18: 219-227.

- **Douaik A**., Van Meirvenne M. and Toth T. (2004). Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. Accepted for publication in *Geoderma.*

- **Douaik A**., Van Meirvenne M. and Toth T. (2004). Temporal stability of spatial patterns of soil salinity determined from laboratory and field electrical conductivity. Submitted for publication in *Arid Land Research and Management*.

- **Douaik A**., Van Meirvenne M. and Toth T. (2004). Detecting and monitoring temporal changes of spatial patterns of soil salinity. Submitted for publication in *Soil Science Society of America Journal*.

<u>Book chapter</u>

- Van Meirvenne M., Vernaillen L., **<u>Douaik A</u>**., Verhoest N.E.C. and Callens M. (2005). Geostatistical procedures for characterizing soil processes. In: Alvarez Benedi J. and Munoz-Carpena R. (eds.), ***Soil-Water-Solute Process Characterization: An Integrated Approach***. CRC Press: Boca Raton, USA, pp: 585-615. ISBN: 1-5667-0657-2.

CURRICULUM VITAE

Ahmed DOUAIK was born on December 7, 1965 in Temara, 15 km south of Rabat. He completed his secondary education at the same city at the Agricultural School and obtained the French "**Baccalauréat D'**, **Agricultural Sciences**". He entered the Hassan II Institute of Agricultural Sciences and Veterinary Medicine, Rabat, in September 1985. After 4 years at this institute, he obtained "**Diplôme d'Agronomie Générale**", and then he moved in September 1989 to the University of Agricultural Sciences of Gembloux, Belgium to specialise in Biometrics. He got the diploma of '**Ingénieur Agronome**' in September 1991. After a short period of work at a public agricultural company (SO.DE.A), he joined, in August 1992, the National Institute of Agricultural Research (INRA) at the Department of Computer Science and Biometrics. He acted as a consulting statistician and helped the researchers of INRA in experimental design, sampling strategy, data analysis, and training in statistics. He started his PhD study in November 2000 under the supervision of Prof. dr. ir. Marc Van Meirvenne, at the Department of Soil Management and Soil Care, Faculty of Bioscience Engineering, Ghent University, Belgium with the cosupervision of dr. Tibor Toth from the Research Institute for Soil Science and Agricultural Chemistry, Hungarian Academy of Sciences, Budapest. During the period of the PhD research, he followed different academic activities in the frame of the doctoral training. He followed courses (in Ghent University and Université Catholique de Louvain), made oral presentations in international conferences (Pedometrics, Geostatistics with Applications to Environment, Soils with Mediterranean Type of Climate), published in international peer-reviewed journals, and visited twice the Hungarian Academy of Sciences. At the completion of these activities he was awarded the '**Certificate of Doctoral Training in Applied Biological Sciences**'. He is currently employed as a scientific researcher at the Research Unit on Environment and Conservation of Natural Resources, INRA, Rabat. He is member of the International Biometric Society, the Belgian Soil Science Society, and the International Union of Soil Science.