M.A. Oliver
Editor

Geostatistical Applications for Precision Agriculture

Springer

Geostatistical Applications for Precision Agriculture

M.A. Oliver
Editor

Geostatistical Applications for Precision Agriculture

Springer

Editor
M.A. Oliver
Visiting Professor in Soil Science
Department of Soil Science
The University of Reading
Whiteknights, Reading RG6 6DW
United Kingdom
m.a.oliver@reading.ac.uk

ISBN 978-90-481-9132-1 e-ISBN 978-90-481-9133-8
DOI 10.1007/978-90-481-9133-8
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010931684

Cover legend: Harvesting a variable-rate nitrogen experiment on Cashmore Field, Silsoe, England. Photograph provided by Silsoe Research institute.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book brings together two dynamic subjects, precision agriculture and geostatistics, that have spatial variation at their core. Geostatistics is applied to many aspects of precision agriculture (PA) including sampling, prediction, mapping, decision-making, variable-rate applications, economics and so on. Contributions from experts in several fields of study illustrate how geostatistics can and has been used to advantage with data such as yield, soil, crops, pests, aerial photographs, remote and proximal imagery. Geostatistical techniques applied include variography, ordinary-, disjunctive-, factorial-, indicator-, regression-, simple-, space-time- and co-kriging, and geostatistical simulation. This book was requested by participants at the Sixth European Conference on Precision Agriculture in Skiathos, 2007 because the link between geostatistics and PA will increase as more intensive information on the soil and crops becomes available from sensors and on-the-go technology. This is not a recipe book, but is intended to guide readers in the use of appropriate techniques for the types of data and needs of the farmer in managing the land. All chapters include one or more case studies to illustrate the techniques.

Chapter 1 sets the scene for the two main topics of the book. The two core techniques of geostatistics, variography and kriging, are described, together with examples of how they can be applied. Sampling for geostatistics is an important issue because it underpins sound results. Chapter 2 considers the importance of spatial scale in sampling, the use of ancillary data, a nested survey and existing variograms of soil or crop properties to guide sampling. Chapter 3 demonstrates the potential to optimize the design of soil sampling schemes if the variation of the target property is represented by a linear mixed model. Chapter 4 describes how calibrated yield data from monitors can be used to target crop and soil investigations and nutrient applications, and for on-farm experiments. This chapter uses spatial statistics rather than only geostatistics because it lends itself better to econometrics. Many environmental variables that are relevant to precision agriculture, such as crop and soil properties and climate, vary in both time and space; Chapter 5 explains the basic elements of space-time geostatistics. Chapter 6 provides an overview of mobile proximal sensors, such as those used to measure apparent soil electrical conductivity (EC_a), and how geostatistics can be used to direct soil sampling to create site-specific management units. Three geostatistical methods to incorporate secondary information into the mapping of soil and crop attributes to improve

the accuracy of their predictions are the topic of Chapter 7. For soil and crop properties that require costly sampling and analysis, there are often insufficient data for geostatistical analyses and Chapter 8 shows how management zones can provide an interim solution to more comprehensive site-specific management. Weeds and plant-parasitic nematodes occur in patches in agricultural fields; Chapter 9 describes how standard geostatistical methods have been used successfully to analyse counts of both weed seedlings and nematodes in the soil and to map their distributions from kriged predictions. Chapter 10 shows how geostatistics can play an important role in analysing experiments for site-specific crop management. Two broad classes of experimental design for precision agriculture (management-class experiments and local-response experiments) are considered and how each may be analysed geostatistically. Geostatistical simulation provides a means to mimic the spatial and or temporal variation of processes that are relevant to precision agriculture, and Chapter 11 shows how it can incorporate uncertainty into modelling to obtain a more realistic impression of the variation. The book has raised several issues, ideas and questions, which are summarized in Chapter 12. Geostatistics needs to be tailored better to the needs of the various groups involved; farmers, advisors and researchers which have their own particular requirements. The potential for geostatistics and precision agriculture for the rest of the twenty-first century appears great.

The Appendix gives examples of software that can be used for geostatistical analyses, and there are brief descriptions of GenStat, VESPER and SGeMS.

Reading, United Kingdom Margaret A. Oliver

Contents

Contributors

Thomas F.A. Bishop Australian Centre for Precision Agriculture, Faculty of Agriculture, The University of Sydney, 1 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015, Australia, t.bishop@usyd.edu.au

Dennis L. Corwin USDA-ARS U.S. Salinity Laboratory, 450 West Big Springs Road, Riverside, CA 92507-4617, USA, Dennis.Corwin@ars.usda.gov

Sytze de Bruin Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands, sytze.deBruin@wur.nl

Fenny M. van Egmond The Soil Company, Leonard Springerlaan 9, 9727 KB Groningen, The Netherlands, fenny@medusa-online.com

Zoë L. Frogbrook Environment Agency Wales, Ty Cambria, 29 Newport Road, Cardiff CF24 0TP, United Kingdom, zoe.frogbrook@environment-agency.wales.gov.uk

William J. Gangloff Department of Soil & Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA, billgangloff@gmail.com

Robin Gebbers Department of Engineering for Crop Production, Leibniz-Institute for Agricultural Engineering, Max-Eyth-Allee 100, D-14469 Potsdam, Germany, rgebbers@atb-potsdam.de

Pierre Goovaerts BioMedware Inc3526 W Liberty, Suite 100, Ann Arbor, MI 48104, USA, goovaerts@terraseer.com

Terry Griffin Division of Agriculture, University of Arkansas, 2301 S University Avenue, Little Rock, AR 72204, USA, tgriffin@uaex.edu

Gerard B.M. Heuvelink Environmental Sciences Group, Wageningen University and Research Centre, PO Box 47, 6700 AA Wageningen, The Netherlands, Gerard.Heuvelink@wur.nl

Ruth Kerry Department of Geography, Brigham Young University, 690 SWKT, Provo, UT 84602, USA, ruth_kerry@byu.edu

Raj Khosla Department of Soil & Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA, Raj.Khosla@colostate.edu

R. Murray Lark Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom, murray.lark@bbsrc.ac.uk

Scott M. Lesch Riverside Public Utilities, Resource Division, 3435 14th Street, Riverside, CA 9250, USA, SLesch@riversideca.gov

Alex B. McBratney Australian Centre for Precision Agriculture, The University of Sydney, John Woolley Building, Sydney, NSW 2006, Australia, Alex.McBratney@sydney.edu.au

Jaskaran S. Mahal Department of Farm Power and Machinery, Punjab Agricultural University, Ludhiana 141004, India, jsmahal@gmail.com

Ben P. Marchant Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom, ben.marchant@bbsrc.ac.uk

Budiman Minasny Australian Centre for Precision Agriculture, John Woolley Building, The University of Sydney, NSW 2006, Australia, b.minasny@sydney.edu.au

Margaret A. Oliver Department of Soil Science, The University of Reading, Whiteknights, Reading RG6 6DW, United Kingdom, m.a.oliver@reading.ac.uk

Matthew J. Pringle Department of Environment and Resource Management, QCCCA Building, 80 Meiers Road, Indooroopilly, QLD 4068, Australia, matthew.pringle@derm.qld.gov.au

Robin M. Reich Department of Forest, Range and Watershed Stewardship, Colorado State University, Fort Collins, CO 80523, USA, robin@warnercnr.colostate.edu

John K. Schueller Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6300, schuejk@ufl.edu

Richard Webster Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom, richard.webster@bbsrc.ac.uk

Dwayne G. Westfall Department of Soil & Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA, Dwayne.Westfall@colostate.edu

Brett M. Whelan Australian Centre for Precision Agriculture, The University of Sydney, John Woolley Building, Sydney, NSW 2006, Australia, brett.whelan@sydney.edu.au

Chapter 1
An Overview of Geostatistics and Precision Agriculture

M.A. Oliver

Abstract This chapter sets the scene for the two main topics of this book, namely geostatistics and precision agriculture. The aim is to provide readers with a foundation for what is to come in the other chapters and an understanding of why the subjects make suitable companions. The history and basic theory of geostatistics are covered, together with the history of precision agriculture and of geostatistics in precision agriculture. The two core techniques of geostatistics, variography and kriging, are described, together with examples of how they can be applied. Methods of estimating the variogram and fitting an authorized model to the experimental values are explained and illustrated. There are many types of kriging; ordinary and disjunctive kriging are described briefly in this chapter, and others types are portrayed in subsequent chapters. The application of the variogram and kriging are illustrated with a case study of an arable field in England.

Keywords History of geostatistics · History of precision agriculture · Theory of geostatistics · Stationarity · Variograghy · Variogram modelling · Ordinary kriging · Disjunctive kriging · Factorial kriging

1.1 Introduction

It is no coincidence that the two essential topics of this book, namely geostatistics and precision agriculture, have come together. When I became interested in precision agriculture (PA) in the mid-1990s, I was surprised to discover that geostatistics was already quite well established for analysing various types of data. It came as a relief that I could propose geostatistical analyses of the data involved in PA without having to convince people of its value. Having come into PA from a background in soil science, my previous experiences were quite different. Soil scientists were reluctant to appreciate the benefits of geostatistics even though Richard Webster

M.A. Oliver (✉)
Department of Soil Science, The University of Reading, Whiteknights,
Reading RG6 6DW, United Kingdom
e-mail: m.a.oliver@reading.ac.uk

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_1,

(another soil scientist) was one of the first to transfer its methods from the mining industry to the environmental sciences. This distinction remains to some extent even now; the PA community has embraced geostatistics to explore the many kinds of data that farmers work with.

This chapter gives an overview of the development and theory of geostatistics, a brief history of precision agriculture and geostatistics in PA. The variogram is the central tool of geostatistics; Section 1.2.2 describes how to compute, model and interpret it. Kriging is a generic term for geostatistical prediction; there are many types of kriging, but ordinary kriging remains the most widely used form. In addition to Section 1.2.3.1 on ordinary kriging, Section 1.2.3.4 describes disjunctive kriging which has the potential to enable farmers to manage their land objectively and to give certain areas priority for treatment. Disjunctive kriging is not described in any other chapters. The methods described are put into practice in a case study of a field on the Shuttleworth Estate in Berkshire, England. The data analysed include yield and soil data.

1.1.1 A Brief History of Geostatistics

What we now regard as geostatistics applies to a specific set of models and techniques developed largely by Matheron (1963) in the 1960s to evaluate recoverable reserves for the mining industry. Many of the ideas had arisen previously in other fields; in fact they have a long history stretching back to the work of Mercer and Hall (1911) at Rothamsted Experimental Station (now Rothamsted Research). Mercer and Hall (1911) examined variation in the yields of crops in numerous small plots in the historical fields at Rothamsted, Harpenden, England. Mercer and Hall were interested in the optimal plot size for experiments on crop yields. Student in an appendix to Mercer and Hall's (1911) paper showed even greater foresight in his observation that yields in adjacent plots were more similar to one another than between other plots further away. He suggested that there were two sources of variation; one that was autocorrelated, i.e. spatially correlated or dependent as we would now refer to it, and the other completely random or spatially uncorrelated, i.e. the nugget effect. In spite of Student's early understanding of the components of spatial variation, these ideas had little impact until the middle of the last century.

As a consequence of the burgeoning amount of data from the agricultural field trials at Rothamsted there was a need for more statistical expertise, and R. A. Fisher was employed in 1919 to develop appropriate methods of analysis. He was interested in the design of experiments and his aim was to be able to estimate the responses of crop yields to different agronomic treatments and varieties. Although he recognized the existence of spatial variation, he regarded it as a nuisance. Therefore, he designed his experiments in such a way as to remove the effects of short-range variation by using large plots and of long-range variation by blocking. The result was that spatial variation was regarded as of little consequence. However, two agronomists Youden and Mehlich (1937) adapted Fisher's (Fisher 1925) analysis of variance to

estimate the variance associated with different sample spacings (see Section 2.2.1). Their aim was to plan further sampling based on this knowledge to avoid wasted effort. As with Student, their ideas were not followed up for several decades. The main revival of Youden and Mehlich's (1937) ideas came after Miesch (1975) showed the equivalence of the components of variance from a hierarchical analysis of variance based on distances with the semivariances of geostatistics. Webster and Butler (1976) also applied this approach in soil science.

Russian research in meteorology by Kolmogorov (1941) led to the recognition of spatial autocorrelation for which he developed the 'structure function' (now the variogram). He also worked out how to use the structure function for optimal interpolation, i.e. without bias and with minimum variance (now kriging). In the early 1950s Krige observed that variation in the block grade in the South African gold mines was considerably less than that of the averaged core samples, and that the block and core-sample grades were correlated. He saw that this relation could improve prediction using regression (Krige 1951). This technique was effectively the first use of kriging and which he called later simple elementary kriging. Matheron (1963) expanded Krige's empirical ideas, in particular the concept that neighbouring samples could be used to improve prediction, and put them into the theoretical framework of regionalized variable theory that underpins geostatistics. This theory provides the basis for solving the pressing problem in the environmental sciences of the need to predict from sparse data. Matheron (1963) first used the term 'kriging', although it derives from Pierre Carlier's use of 'krigeage' in the late 1950s, for geostatistical prediction in recognition of D. G. Krige's contribution to improving the precision of estimating concentrations of gold and other metals in ore bodies. Geostatistics has since become a principal branch of the wider body of spatial statistics (Cressie 1993). It has been applied in many different fields, such as agriculture, fisheries, hydrology, geology, meteorology, petroleum, remote sensing, soil science and of importance here precision agriculture.

1.1.2 A Brief History of Precision Agriculture

As with geostatistics, the background to precision agriculture is multifarious and apparently novel. Precision agriculture, however, has been carried out by farmers since the early days of agriculture. Subsistence farmers worked (and still do in parts of the world) on small patches of ground, the characteristics of which they knew well. They divided their landholdings into smaller areas, fields, to grow crops where the conditions were most suitable. For them precision was about ensuring enough food to sustain the family; a life or death matter. The work of Gilbert and Lawes and their successors at Rothamsted was also about precision farming; they wanted to assess the benefits of different combinations and amounts of crop nutrients and of crop varieties. The aim was explicitly to increase yields; cheap fertilizers could achieve this and concerns about their impact on the environment were not an issue at the time. Until the 1980s precise or site-specific management was at the farm level

and the management unit was the field. The soil of a field was sampled to determine the mean value of crop nutrients and pH, and it was amended uniformly over the field. The crop yield was based on the total weight taken from the field.

The term precision agriculture appears to have been used first in 1990 as the title of a workshop held in Great Falls, Montana, sponsored by Montana State University. Before this, the terms 'site-specific crop management' or 'site-specific agriculture' were used. In fact, the first two international conferences on 'precision agriculture' referred to site-specific management in the title, but by the third conference in 1996 the term precision agriculture was being used. By the mid-1990s, what we now regard as the new paradigm in agriculture was being referred to as precision agriculture. The concept of modern precision agriculture has been driven forward and is underpinned by technological changes based on information technology (Schueller 1997). This enabled more precise local management, and consequently the unit of management has now become the field and variation within that unit has become the focus. This reflects a change in the scale of operation from the farm to the field, but there is more to it than this. With the increase in size of machinery being used in agriculture in the developed countries, farmers removed field boundaries and merged fields into increasingly larger units. The original fields, which had probably been created because of a particular set of soil or landscape conditions, were now parts of larger fields and their inherent variation was added together. The increase in field size was accompanied, therefore, by an increase in within-field variability.

Robert (1999) mentioned that in the mid-1970s to early 1980s there was a greater awareness among farmers of the potential benefits of better farm record keeping and understanding of soil and crop input requirements. He described how in the late 1970s CENEX (Farmers Union Central Exchange, Inc.) and the Control Data Corporation started a joint venture called "CENTROL – Farm Management Services". The outcome of their study was a better awareness of within-field variation in properties of the soil and crop, and of the potential benefits of management within fields by zones. This then led to a project by SoilTeq (Luellan 1985) to create a spreader that could change the blend and rate of fertilizer on-the-go, i.e. what we now know as variable-rate application (VRT). The first VRT machines were used in 1985 by CENEX. David Mulla (personal communication) worked with SoilTeq to write specialized GIS software for mapping spatial patterns in crop nutrients in 1986, and the first maps were produced from this software in 1987. Robert (1999) said that by the mid-1980s microprocessors made possible the development of computers for farm equipment and controllers, positioning of machines with global positioning systems (GPS), development of the first sensors, electronic acquisition and processing of spatial data for farm geographic record keeping systems and the use of GIS to produce site-specific management maps. Robert (1999) said that since the introduction of the concept of PA in the mid-1980s, it has spread rapidly and widely, albeit the adoption has been variable. There has been a mushrooming of technology and services in response to the needs of this approach to agriculture. Schueller (1997) described how one major manufacturer of grain harvesters claimed that a third of its new combines were equipped with yield monitors. At a similar time I recollect

being told at the International Geostatistics Congress held in Avignon in 1988 that tractors had onboard computers that would be able to use kriged maps to guide fertilizer and other applications. At the time this seemed far-fetched and fanciful as few people were even familiar with microprocessors which were essential for the types of portable operations associated with agriculture.

It was during the 1980s that huge changes in the way that we perceive agriculture began (Cochrane 1993). There was a quiet revolution underway based on information technology. One of the most significant steps was the introduction of a yield meter by Massey Ferguson in 1982. This device was mounted on a clean grain elevator of one of their combine harvesters. This meant that yield could be recorded continuously for the first time. In 1984 Massey Ferguson carried out a field trial in the United Kingdom to investigate whether the yield meter could be used to measure yield variation and whether yield varied within fields (http://www.fieldstar.dk/agco/FieldStar/FieldStarUK/System/HistoryFeatures/YieldMapping.htm). At this time GPS was not available, and the company did its experiment as follows. It set out a 10-m grid in a field of wheat, and the start of each grid square was marked by a pole in the ground. The field was harvested with two men on the combine; one to drive and the other recorded the yield manually from the meter as they passed over each grid square. The yield varied over the field by 10 t ha^{-1}. In the 1990s with the advent of GPS, yield mapping became fairly routine. The first GPS were available on tractors in 1991, but they had an accuracy of only 100 m which was not good enough for mapping (http://www.fieldstar.dk/agco/FieldStar/FieldStarUK/System/HistoryFeatures/GPSBrings.htm). By the mid-1990s with differential GPS (DGPS) accuracy improved to 5–10 m, and this has improved further since 2000 when the US Department of Defence turned off selective availability.

Before the 1990s maps, other than of the soil and possibly landscape, played little part in agricultural management. Schafer et al. (1984) said at this time that maps of soil type and topography could be used to control fertilizer and pesticide applications and tillage operations. The first yield map of Searcy et al. (1989) showed the effect of compaction from farm machinery on yield (see Section 1.3.3.2).

The National Research Council (1997, p. 17) gave a clear definition of PA as follows: "Precision agriculture is a management strategy that uses information technologies to bring data from multiple sources to bear on decisions associated with crop production". They suggested that PA has three components: obtaining data at an appropriate scale, interpretation and analyses of the data, and implementation of a management response at an appropriate scale and time. The intensity and resolution of the spatial information involved in PA means that the revolution to modern PA is essentially about a change in the scale of operation and management. The ability to determine within-field variation and to manage it are central. The data used in PA are often at a large spatial resolution, for example yield, proximal sensor data, remotely sensed data, digital elevation models and so on. A major stumbling block to the wider spread and adoption of PA is the sparsity of soil and crop information, although there have been examples of on-the-go measurement of pH (Viscarra Rossel and McBratney 1997). The National Research Council (1997, p. 4) also makes the point that "current mapping techniques are limited by a lack of understanding of

the geostatistics necessary for displaying spatial variability of crops and soils", and on p. 59 "An increased knowledge base in geostatistical methods should improve interpretation of precision agriculture data".

It is clear from the quotations given above that the value of geostatistics had already been established in PA. In fact, the early applications came from scientists already conversant with geostatistical methods. The marriage of geostatistics and PA was an easy one because geostatistics requires enough data at an interval that resolves the variation adequately to compute reliable variograms. These demands can be satisfied by the kinds of data widely available in PA, apart from some soil and crop data.

1.1.3 A Brief History of Geostatistics in Precision Agriculture

David Mulla was the first person to apply geostatistics explicitly in precision agriculture (see Mulla and Hammond 1988). This paper describes a study that aimed to: (i) introduce geostatistics for mapping patterns in soil P and K, (ii) determine the nature and extent of spatial variation in these crop nutrients in a large irrigation circle and (iii) determine what sampling intensity is necessary to identify the major patterns in the soil. Mulla and Hammond stated that the variable-rate programme needed appropriate sampling and an accurate map of the crop nutrients. Their recommendations were that if the soil is variable farmers should avoid uniform applications. They used geostatistics to interpolate between measured values based on the work of Warrick et al. (1986) in soil science. Richard Webster (Chapter 9) and his team, namely Burgess and McBratney, were applying geostatistics to soil data in the early 1980s (Burgess and Webster 1980a, b; Burgess et al. 1981). Their aim was to quantify the spatial structure in the variation with the variogram and to use its parameters with the data for prediction by kriging to produce a map of the variation in soil properties. Although their work was earlier than that of Mulla and Hammond (1988) and was explicitly related to agriculture, it was not directed towards the modern concept of PA.

During the same period Miller et al. (1988) and Webster and Oliver (1989) were applying geostatistics in an agricultural context; they were on the track of PA, but not in an explicit way. Miller et al. (1988) attempted to understand crop growth and yield by examining the spatial relations of soil physical and chemical properties that were altered by soil erosion. In addition to soil information, they also had data on above ground biomass and the grain yield at each sampling site. Their aim was to provide information that would enable better management. They computed variograms and cross variograms of several properties and also used kriging to produce maps. They showed the importance of landscape position in relation to soil properties and crop growth. Webster and Oliver (1989) were the first to apply disjunctive kriging (see Section 1.2.3.4) to an agricultural problem. At this time, however, we knew nothing of the revolution in agriculture that was occurring in the USA.

Mulla's work during the late 1980s and early 1990s laid the foundations for the adoption of geostatistics in PA (Mulla 1989, 1991, 1993; Bhatti et al. 1991). The last paper in the list was presented at the first workshop on site-specific crop management (Robert et al. 1993) held at the University of Minnesota. At this workshop there were only 21 papers, and Mulla's was the only one that referred to geostatistics. At the second international conference on site-specific management for agricultural systems (Robert et al. 1995) there were 67 papers. Two years later this conference was renamed the international conference on precision agriculture (ICPA) and this has since taken the form of a biennial meeting. At the first conference on precision agriculture in Europe in 1997 (Stafford 1997), there were some 15 papers in which geostatistics was used. By the fourth ICPA in 1998 there were 19 papers with geostatistical applications and in 2008 there were 23. There has been a similar increase in the application of geostatistics in PA at the European conferences, culminating in over 20 papers in the 2005 proceedings, but then it declined to about 12 in 2007and 8 in 2009. The decline might reflect that geostatistics has become more commonplace in PA and that authors no longer mention explicitly how their maps have been obtained. It also indicates that many more types of analysis are now being used in PA as researchers adapt to the needs of farmers.

The journal *Precision Agriculture* has had, and continues to have, a steady stream of papers in which geostatistics is applied. The subjects in which it is applied have broadened from soil and crop properties to remotely and proximally sensed data, weeds, yields of tree crops and so on. A wide range of geostatistical techniques is applied to these data, for example variogram analysis, ordinary kriging, cokriging, simulation, intrinsic random function (IRF-k) kriging and the indicator approach.

In the rest of this chapter I shall set the scene for what is to come in subsequent chapters by summarizing the core geostatistical methods of the variogram and kriging. I illustrate these with a case study from the Shuttleworth Estate, Bedfordshire, England. In addition, I give a brief introduction to disjunctive kriging as a management technique and illustrate its potential with the same case study.

1.2 The Theory of Geostatistics

Geostatistics as we now know it has developed from Matheron's (1963) coherent theoretical underpinning of Krige's empirical observations (for more detail see Journel and Huijbregts 1978; Goovaerts 1997; Webster and Oliver 2007). The spatial variation of most properties on, above or beneath the Earth's surface is so complex that it led Matheron to find an alternative approach to the traditional deterministic one for their analysis. The approach he adopted was one that could deal with the inherent uncertainty of spatial data in a stochastic way. The basis of modern geostatistics is to treat the variable of interest as a random variable. This implies that at each point, $\mathbf{x}$, in space there is a series of values for a property, $Z(\mathbf{x})$, and the one observed, $z(\mathbf{x})$, is drawn at random according to some law, from some probability distribution. At $\mathbf{x}$, a property $Z(\mathbf{x})$ is a random variable with a mean, μ, and

variance, σ^2. The set of random variables, $Z(\mathbf{x}_1)$, $Z(\mathbf{x}_2), \ldots$, is a random process, and the actual value of Z observed is just one of potentially any number of realizations of that process.

To describe the variation of the underlying random process, we can use the fact that the values of regionalized variables at places near to one another tend to be autocorrelated. Therefore, we can estimate the spatial covariance to describe this relation between pairs of points; for a random variable this is given by

$$C(\mathbf{x}_1, \mathbf{x}_2) = \mathrm{E}[\{\mathrm{Z}(\mathbf{x}_1) - \mu(\mathbf{x}_1)\}\{\mathrm{Z}(\mathbf{x}_2) - \mu(\mathbf{x}_2)\}], \tag{1.1}$$

where $\mu(\mathbf{x}_1)$ and $\mu(\mathbf{x}_2)$ are the means of Z at $\mathbf{x}_1$ and $\mathbf{x}_2$, and E denotes the expected value. As there is only ever one realization of Z at each point, this solution is unavailable because the means are unknown. To proceed we have to invoke assumptions of stationarity.

1.2.1 Stationarity

Under the assumptions of stationarity certain attributes of the random process are the same everywhere. We assume that the mean, $\mu = \mathrm{E}[Z(\mathbf{x})]$, is constant for all $\mathbf{x}$, and so $\mu(\mathbf{x}_1)$ and $\mu(\mathbf{x}_2)$ can be replaced by μ, which can be estimated by repetitive sampling. When $\mathbf{x}_1$ and $\mathbf{x}_2$ coincide, Eq. 1.1 defines the variance (or the a priori variance of the process), $\sigma^2 = \mathrm{E}[\{\mathrm{Z}(\mathbf{x}) - \mu\}^2]$,which is assumed to be finite and, as for the mean, the same everywhere. When $\mathbf{x}_1$ and $\mathbf{x}_2$ do not coincide, their covariance depends on their separation and not on their absolute positions, and this applies to any pair of points $\mathbf{x}_i$, $\mathbf{x}_j$ separated by the lag $\mathbf{h} = \mathbf{x}_i - \mathbf{x}_j$ (a vector in both distance and direction), so that

$$\begin{aligned} C(\mathbf{x}_i, \mathbf{x}_j) &= \mathrm{E}\left[\{\mathrm{Z}(\mathbf{x}_i) - \mu\}\{\mathrm{Z}(\mathbf{x}_j) - \mu\}\right] \\ &= \mathrm{E}\left[\{\mathrm{Z}(\mathbf{x})\}\{\mathrm{Z}(\mathbf{x} + \mathbf{h})\} - \mu^2\right] \\ &= C(\mathbf{h}), \end{aligned} \tag{1.2}$$

which is also constant for a given $\mathbf{h}$. This constancy of the first and second moments of the process constitutes second-order or weak stationarity. Equation 1.2 indicates that the covariance is a function of the lag, and it describes quantitatively the dependence between values of Z with changing separation or lag. The autocovariance depends on the scale on which $Z$is measured; therefore, it is often converted to the dimensionless autocorrelation, $\rho(\mathbf{h})$, by

$$\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0}), \tag{1.3}$$

where $C(0) = \sigma^2$, the covariance at lag 0.

1.2.1.1 Intrinsic Variation and the Variogram

The mean often appears to change across a region and the variance will appear to increase indefinitely as the extent of the area increases. Consequently, there is no value for μ to insert into Eq. 1.2 and the covariance cannot be defined. This is a departure from weak stationarity. Matheron's (1965) solution to this was the weaker *intrinsic hypothesis* of geostatistics. Although the general mean might not be constant, it would be for small lag distances, and so the expected differences would be zero as follows:

$$\mathrm{E}[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] = 0, \tag{1.4}$$

and he replaced the covariances by the expected squared differences

$$\mathrm{E}\left[\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\}^2\right] = \mathrm{var}\left[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\right] = 2\gamma(\mathbf{h}). \tag{1.5}$$

The quantity $\gamma(\mathbf{h})$ is known as the semivariance at lag $\mathbf{h}$, or the variance per point when points are considered in pairs. As for the covariance, the semivariance depends only on the lag and not on the absolute positions of the data. As a function of $\mathbf{h}$, $\gamma(\mathbf{h})$ is the semivariogram or nowadays more usually termed the variogram.

If the process $Z(\mathbf{x})$ is second-order stationary, the variogram and covariance are equivalent:

$$\begin{aligned}\gamma(\mathbf{h}) &= C(\mathbf{0}) - C(\mathbf{h}) \\ &= \sigma^2\{1 - \rho(\mathbf{h})\}.\end{aligned} \tag{1.6}$$

However, if the process is intrinsic only there is no equivalence because the covariance function does not exist. The variogram is valid, however, and it can be applied more widely than the covariance function. This makes the variogram a valuable tool and as a consequence it is at the core of geostatistics.

1.2.2 The Variogram

1.2.2.1 Estimating the Variogram

Matheron's (1965) method of moments (MoM) estimator is the usual method of computing the empirical semivariances from data, $z(\mathbf{x}_1)$, $z(\mathbf{x}_2)$, ... Its equation is

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})}\sum_{i=1}^{m(\mathbf{h})}\{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2, \tag{1.7}$$

where $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i + \mathbf{h})$ are the actual values of Z at places $\mathbf{x}_i$ and $\mathbf{x}_i + \mathbf{h}$, and $m(\mathbf{h})$ is the number of paired comparisons at lag $\mathbf{h}$. The experimental or sample variogram

is obtained by changing **h**. If the data are on a regularly sampled transect or grid, the semivariances can be computed for integral multiples of the sampling interval. For a transect the lag becomes a scalar, $h = |\mathbf{h}|$, and the maximum lag should be no more than one third of the transect length. For irregularly sampled data in one or more dimensions, or to compute the omnidirectional variogram of data on a regular grid, the separations between pairs of points are placed into bins with limits in both separating distance and direction (see p. 72 of Webster and Oliver 2007).

Webster and Oliver (1992) showed that at least 100 sampling points are required to estimate the method of moments variogram reliably. Their results show clearly the penalty of too few data. Pardo-Igúzquiza (1997) suggested the maximum likelihood (ML) approach as an alternative to Matheron's estimator. He also suggested that where the number of data is small (a few dozen), the ML variogram estimator offers an alternative that gives an estimate of the variogram parameters and of their uncertainty (Pardo-Igúzquiza 1998). Kerry and Oliver (2007a) have shown that with about 50 to <100 data the residual maximum likelihood (REML) variogram estimator (Pardo-Igúzquiza 1997) provides a more accurate variogram (see Section 2.5.1). However, this estimator still performs better with about 100 data.

Although linear geostatistics does not require a normal distribution, the variogram is based on variances and any asymmetry in the distribution signalled by a skewness coefficient >1 or < −1 should be examined. Departures from normality can arise from a long tail of larger or smaller values in the underlying process or from one or more extreme values from a secondary process that contaminates the primary one. Box-plots and histograms should be examined to see which of these is the cause because, as Kerry and Oliver (2007b, c) showed, they require a different approach to computing the variogram.

1.2.2.2 Features of the Variogram

Most environmental variables vary in a spatially *continuous* way; therefore we should expect $\gamma(\mathbf{h})$ to pass through the origin at $\mathbf{h} = \mathbf{0}$ (Fig. 1.1a). However, the variogram often appears to approach the ordinate at some positive value as **h** approaches $\mathbf{0}(|h| \to 0)$, Fig. 1.1b, which suggests that the process is discontinuous. This discrepancy is known as the *nugget variance*. (The features described in this section are illustrated in one dimension in Fig. 1.1 where $h = |\mathbf{h}|$.) For properties that vary continuously the nugget variance usually includes some measurement error, but it comprises mainly variation that occurs over distances less than the shortest sampling interval. Figure 1.1c is a pure nugget variogram which usually indicates that the sampling interval is too large to resolve the variation present. The semivariances increase with increasing lag distance (*monotonic increasing*) as shown in Fig. 1.1a, b. The small values of $\gamma(\mathbf{h})$ at short lag distances show that the values of $Z(\mathbf{x})$ are similar, but as the lag distance increases they become increasingly dissimilar on average. A variogram with a monotonic increasing slope indicates that the process is *spatially dependent or autocorrelated*. Variograms that reach an upper bound after the initial slope as in Fig. 1.1b describe a second-order stationary process. This maximum is known as the *sill variance*. It is the a priori variance, σ^2,

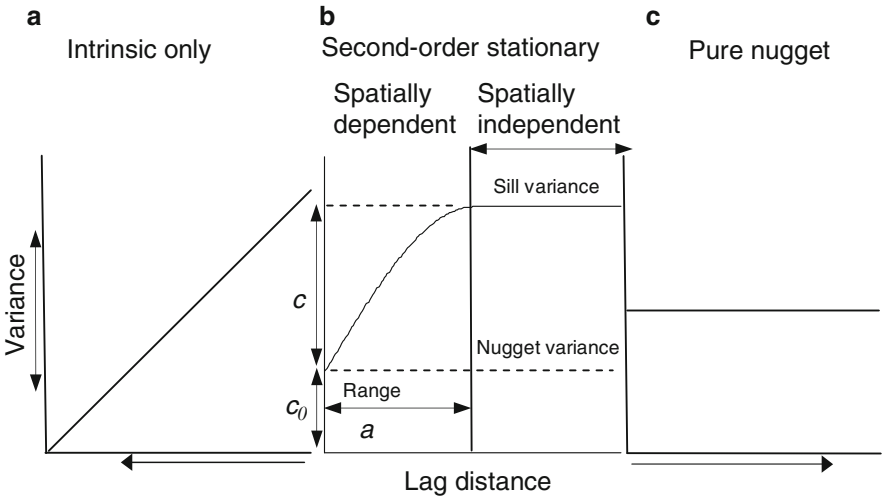

Fig. 1.1 Three variogram forms: (**a**) unbounded, (**b**) bounded and (**c**) pure nugget

of the process. The sill variance ($c + c_0$) comprises any nugget variance (c_0) and the spatially correlated variance (c). The finite distance at which some variograms reach their sill is the *range* (a), i.e. the range of spatial dependence. Places further apart than the range are spatially independent, Fig. 1.1b. Some variograms do not have a finite range, and the variogram approaches its sill asymptotically. If the variogram increases indefinitely (*unbounded variogram*) with increasing lag distance as in Fig. 1.1a, the process is intrinsic only.

A variogram that fluctuates in a periodic way with increasing lag distance indicates regular repetition in the variation. If the variation differs according to direction, it is *anisotropic* and the variogram will be also. The anisotropy is *geometric* if the initial gradient or variogram range change with direction and a simple transformation of the coordinates will remove the effect. Z*onal anisotropy* cannot be dealt with as readily; it is present if the sill variance fluctuates with changes in direction, which might indicate the presence of preferentially orientated zones with different means. Variation in the environment may occur at several spatial scales simultaneously, and patterns in the variation can be nested within one another. The experimental variogram will appear more complex if more than one spatial scale is present (see Fig. 1.7). Nested variation is often observed when there are many data, for example from remote or proximal sensing or from yield data. A combination of two or more simple models that are authorized (a nested model) can be used to model such variation, Fig. 1.7.

1.2.2.3 Modelling the Variogram

The experimental method of moments variogram estimates the underlying variogram, which is a continuous function, as a set of discrete points at particular lag

intervals. These estimates often fluctuate from point to point because they are subject to error that arises largely from the sampling. To describe the spatial variation, we fit a model to the experimental values. The model must be conditional negative semi-definite (CNSD) so that it will not give rise to negative variances when random variables are combined (see Webster and Oliver 2007 for more detail on this). The function must also be able to represent the variogram features described above.

There are a few simple functions that satisfy the above conditions. They include bounded functions, which represent processes that are second-order stationary, and unbounded ones that are intrinsic only. Webster and Oliver (2007) give several functions and there are many examples of the different models in the following chapters. The most commonly fitted models are the exponential and spherical ones (Figs. 1.5a, b and 1.6, respectively, and Eqs. 1.26 and 1.27, respectively).

Fitting models to the experimental values is controversial, and we recommend practitioners to avoid fitting by eye. This is because the accuracy of the semivariances varies and the experimental variogram might fluctuate considerably from point to point. Fitting a suitable model is fundamental in geostatistics because it affects subsequent analyses (see Webster and Oliver 2007). A weighted least squares approach is advisable for fitting because it takes account of the accuracy of the individual semivariances and the residual sum of squares provides a means of selecting the best fitting function.

1.2.3 Geostatistical Prediction: Kriging

Kriging is often known as a best linear unbiased predictor (BLUP); it is a method of optimal prediction or estimation in geographical space. It is optimal in the sense of unbiasedness and minimum variance. It is the geostatistical method of interpolation of sparse data for random spatial processes. Most features of the environment (atmosphere, ores, soil, vegetation, water and oceans) can be measured at any of an infinite number of places, but for reasons of economy they are usually measured at few. However, kriging has also been shown to be of value for reducing the point-to-point variation or noise in intensive data, such as satellite imagery and yield, to gain insight into the structure of the variation (Oliver et al. 2000; Oliver and Carroll 2004).

Several mathematical methods of interpolation are available, for example, Thiessen polygons, triangulation, natural-neighbour interpolation, inverse functions of distance, least-squares polynomials (trend surfaces) and splines. Laslett et al. (1987) compared several of these methods with kriging, and showed that kriging performed the best. Kriging overcomes many of the shortcomings of the mathematical methods of interpolation by taking into account the way a property varies in space through the variogram or covariance function. In addition, kriging provides not only predictions but also the kriging variances or errors. Kriging can be regarded simply as a method of local weighted moving averaging of the observed values of a random variable, Z, within a neighbourhood, V. It can be done for point (punctual kriging) or block supports of various size (block kriging), depending upon the aims of prediction, even though the sample information is often for points.

Kriging is now used in many disciplines that use spatial prediction and mapping, such as mining, petroleum engineering, meteorology, soil science, precision agriculture, pollution control, public health, monitoring fish stocks and other animal densities, remote sensing, ecology, geology, hydrology, and so on. As a consequence, kriging has become a generic term for a range of BLUP least-squares methods of spatial prediction in geostatistics. The original formulation of kriging, now known as ordinary kriging (Journel and Huijbregts 1978), is the most robust method and the one most often used.

1.2.3.1 Ordinary Kriging

Ordinary kriging is based on the assumption that the mean is unknown. Consider that a random variable, Z, has been measured at sampling points, $\mathbf{x}_i$, $i = 1, \ldots, N$. We use this information to estimate its value at a point $\mathbf{x}_0$ by punctual kriging with the same support as the data by

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i), \tag{1.8}$$

where n usually represents the data points within the local neighbourhood, V, and is much smaller than the total number in the sample, N, and λ_i are the weights. To ensure that the estimate is unbiased the weights are made to sum to one,

$$\sum_{i=1}^{n} \lambda_i = 1, \tag{1.9}$$

and the expected error is $\mathrm{E}[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$. The prediction variance is

$$\begin{aligned} \mathrm{var}\left[\hat{Z}(\mathbf{x}_0)\right] &= \mathrm{E}\left[\left\{\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)\right\}^2\right] \\ &= 2\sum_{i=1}^{n} \lambda_i \gamma\left(\mathbf{x}_i, \mathbf{x}_0\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma\left(\mathbf{x}_i, \mathbf{x}_j\right), \end{aligned} \tag{1.10}$$

where $\gamma(\mathbf{x}_i, \mathbf{x}_j)$ is the semivariance of Z between points $\mathbf{x}_i$ and $\mathbf{x}_j$, $\gamma(\mathbf{x}_i, \mathbf{x}_0)$ is the semivariance between the ith sampling point and the target point $\mathbf{x}_0$. The semivariances are derived from the variogram model because the experimental semivariances are discrete and at limited distances.

Kriged predictions are often required over areas that are larger than the sample support of the data for which block kriging is used. The estimate is still a weighted average of the data, $z(\mathbf{x}_1)$, $z(\mathbf{x}_2)$, …, $z(\mathbf{x}_n)$, at the unknown block, B,

$$\hat{Z}(B) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i). \tag{1.11}$$

The estimation variance of $\hat{Z}(B)$ is:

$$\begin{aligned}\text{var}\left[\hat{Z}(B)\right] &= \text{E}\left[\left\{\hat{Z}(B) - Z(B)\right\}^2\right] \\ &= 2\sum_{i=1}^{n}\lambda_i\bar{\gamma}(\mathbf{x}_i, \mathbf{B}) - \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma\left(\mathbf{x}_i, \mathbf{x}_j\right) - \bar{\gamma}(B, B). \quad (1.12)\end{aligned}$$

where $\bar{\gamma}(\mathbf{x}_i, B)$ is the average semivariance between data point $\mathbf{x}_i$ and the target block B, and $\bar{\gamma}(B, B)$ is the average semivariance within B, the within-block variance.

Equation 1.10 for a point leads to a set of $n+1$ equations in the $n+1$ unknowns:

$$\begin{aligned}\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{x}_i, \mathbf{x}_j) + \psi(\mathbf{x}_0) &= \gamma(\mathbf{x}_j, \mathbf{x}_0) \quad \text{for all } j, \\ \sum_{i=1}^{n}\lambda_i &= 1, \quad (1.13)\end{aligned}$$

where the Lagrange multiplier, $\psi(\mathbf{x}_0)$, is introduced to achieve minimization. The weights, λ_i, are inserted into Eq. 1.8 to give the prediction of Z at $\mathbf{x}_0$. The kriging (prediction or estimation) variance is then obtained as

$$\sigma^2(\mathbf{x}_0) = \sum_{i=1}^{n}\lambda_i\gamma(\mathbf{x}_i, \mathbf{x}_0) + \psi(\mathbf{x}_0). \quad (1.14)$$

Punctual kriging is an exact interpolator – the kriged value at a sampling site is the observed value there and the prediction variance is then zero.

The equivalent kriging system for blocks is

$$\begin{aligned}\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{x}_i, \mathbf{x}_j) + \psi(B) &= \bar{\gamma}(\mathbf{x}_j, B) \qquad \text{for all } j \quad (1.15) \\ \sum_{i=1}^{n}\lambda_i &= 1,\end{aligned}$$

and the block kriging variance is obtained as

$$\sigma^2(B) = \sum_{i=1}^{n}\lambda_i\bar{\gamma}(\mathbf{x}_i, B) + \psi(B) - \bar{\gamma}(B, B). \quad (1.16)$$

Block kriging results in smoother estimates and smaller estimation variances overall because the nugget variance is contained entirely in the within-block variance, $\bar{\gamma}(B, B)$, and does not contribute to the block-kriging variance.

For many environmental applications, including PA, kriging is most likely to be used for interpolation and mapping. The values of the property are usually estimated at the nodes of a fine grid, and the variation can then be displayed by isarithms or by layer shading. The kriging variances or standard errors can be mapped similarly: they are a guide to the reliability of the estimates. Where sampling is irregular, such a map may indicate if there are parts of a region where sampling should be increased to improve the estimates.

1.2.3.2 Kriging Weights

The kriging weights depend on the variogram and the configuration of the sampling. The distribution of the weights within the search radius is one feature that makes kriging different from classical interpolators where the weights are applied arbitrarily. Webster and Oliver (2007) illustrate how the weights vary according to changes in the nugget:sill ratio, the variogram range, type of model, sampling configuration and the effect of anisotropy. The weights are particularly sensitive to the nugget variance and anisotropy. Weights close to the point or block to be estimated carry more weight than those further away, which shows that kriging is a local predictor. As the nugget:sill ratio increases the weights near to the target decrease and those further away increase. For a pure nugget variogram, the kriging weights are all the same and the estimate is simply the mean of the values in the neighbourhood. The effect of the range is more complex than for the nugget:sill ratio because it is also affected by the type of variogram model. In general, however, as the range increases the weights increase close to the target. For data that are irregularly distributed, points that are clustered carry less weight individually than those that are isolated. The fact that the points nearest to the target generally carry the most weight has practical implications. It means that the search neighbourhood need contain no more than 16–20 data points.

1.2.3.3 Other Types of Kriging

As mentioned above, the term kriging is now used generically because the method has been adapted to tackle increasingly varied problems that have arisen. Ordinary kriging assumes that the mean is unknown and that the process is locally stationary, whereas simple kriging assumes that the mean is known. As a consequence it is used little because the mean is generally unknown. Simple kriging is used in indicator and disjunctive kriging in which the data are transformed to have known means. Lognormal kriging is ordinary kriging of the logarithms of strongly positively skewed data that approximate a lognormal distribution. Kriging with trend (5.3) enables data with a strong deterministic component (non-stationary process) to be analysed; Matheron (1969) originally introduced universal kriging for this purpose, followed by intrinsic random function kriging with drift of order k (IRF-k kriging), Matheron (1973). The state of the art is empirical-BLUP (Stein 1999), which uses

a variogram estimated by REML (Lark et al. 2006). Matheron (1982) developed factorial kriging or kriging analysis for variation that is nested. It estimates the long- and short-range components of the variation separately, but in a single analysis. Ordinary cokriging (Matheron 1965) (Section 7.2.2) is the extension of ordinary kriging to two or more variables that are spatially correlated. If some property that can be measured cheaply at many sites is spatially correlated or coregionalized with others that are expensive to measure and recorded at many fewer sites, the latter can be estimated more precisely by cokriging with the spatial information from the former. Disjunctive kriging (Matheron 1976) (Section 1.2.3.4) is a non-linear parametric method of kriging. It is valuable for decision-making because the probabilities of exceeding (or not) a predefined threshold are determined in addition to the kriged estimates. Indicator kriging (Journel 1983) (Section 9.3.3) is a non-linear, non-parametric form of kriging in which continuous variables are converted to binary ones (indicators). It can handle distributions of almost any kind and can also accommodate 'soft' qualitative information to improve prediction. Probability kriging was proposed by Sullivan (1984) because indicator kriging does not take into account the proximity of a value to the threshold, but only its geographic position. Bayesian kriging was introduced by Omre (1987) for situations in which there is some prior knowledge about the drift or trend.

1.2.3.4 Disjunctive Kriging

Disjunctive kriging (DK), proposed Matheron (1976), is described here because it is not applied in any other chapter, and also Webster and Oliver (1989) have shown its suitability for agricultural management. Precision agriculture aims to apply sufficient plant nutrients, lime and other agrochemicals, but no more, both to limit damage to the environment and for economic reasons. Recommendations on the amount of nutrients to apply are based on their concentrations in the soil, and for lime on the soil's pH. If a nutrient or the pH is less than a particular value or critical threshold, z_c, the farmer is advised to apply fertilizer or lime. The amounts recommended may vary according to the nutrient concentrations and pH of the soil and the type of crop. To apply fertilizer or lime at variable rates requires accurate local information on the nutrient status of the soil and its pH, which is usually based on estimates from sample information that are more or less in error.

The variogram and ordinary kriging can provide accurate local information if the sampling is sound (Chapters 2 and 3). Kriged predictions, although optimal, are smoothed, especially where there is a large nugget variance, and this can have an adverse effect on their use for decision making. In general, decisions are easy where the estimated values of the properties are much less than or much greater than the specified threshold, or where the kriging variance is small, or both. Difficulty arises where the estimate is close to the threshold. The consequence is that the farmer might attempt to remedy a nutrient deficiency that does not exist or fail to remedy a real one. The manager needs to know the risks of taking the estimates at face value. In other words what is the probability that the true values exceed or fall short of the

critical values. Disjunctive kriging provides a solution. For each estimate, it enables the probability that the true value exceeds (or is less than) a threshold to be estimated through non-linear rescaling of the original data.

Theory

Disjunctive kriging is a non-linear method of prediction based on transforming the data to indicator functions in relation to a predefined threshold, z_c. This dissects the scale of a property of interest, Z, into two parts: one for which $Z(\mathbf{x}) < z_c$ and the other for which $Z(\mathbf{x}) \geq z_c$, and we can assign the values 1 and 0 to these, respectively. This is known as indicator coding. The indicator function, $\Omega[Z(\mathbf{x}) < z_c]$, is a random variable, $Y(\mathbf{x})$, which has a variogram $\gamma_{z_c}^{\Omega}(\mathbf{h})$. The most common type of DK is Gaussian disjunctive kriging, which I use in the case study.

The assumptions underlying Gaussian DK are more stringent than those of OK. The first is that $z(\mathbf{x})$ is a realization of a second-order stationary random process $Z(\mathbf{x})$ with a mean, μ, and variance, σ^2. Therefore, the underlying variogram must be bounded. Secondly, the bivariate distribution for $n+1$ variates, i.e. for the target site and the sample locations in its neighbourhood, is known and it is stable throughout the region. If the distribution of $Z(\mathbf{x})$ is normal and the process is second-order stationary then we can assume that the bivariate distribution for each pair of points is also normal. Webster and Oliver (2007) describe the theory of Gaussian DK in more detail, and Rivoirard (1994) provides a full account of non-linear geostatistics.

Hermite Polynomials Transformation to a standard normal distribution, $Y(\mathrm{x})$, is given by

$$Z(\mathbf{x}) = F\left[Y(\mathbf{x})\right], \tag{1.17}$$

where Φ is a linear combination of Hermite polynomials. This transform is invertible, which means that the results are expressed in the same units as the original measurements. The transform can be expressed as

$$\begin{aligned} Z(\mathbf{x}) = \Phi[Y(\mathbf{x})] &= \varphi_0 H_0\{Y(\mathbf{x})\} + \varphi_1 H_1\{Y(\mathbf{x})\} + \varphi_2 H_2\{Y(\mathbf{x})\} + \cdots \\ &= \sum_{k=0}^{\infty} \varphi_k H_k\{Y(\mathbf{x})\}. \end{aligned} \tag{1.18}$$

The H_k are an infinite series of Hermite polynomials and the ϕ_k are coefficients that can be evaluated by Hermite integration. The Hermite polynomials are kriged separately and have only to be summed to give the disjunctively kriged estimates:

$$\hat{Z}^{\mathrm{DK}}(\mathbf{x}) = \varphi_0 + \varphi_1 \hat{H}_1^{\mathrm{K}}\{Y(\mathbf{x})\} + \varphi_2 \hat{H}_2^{\mathrm{K}}\{Y(\mathbf{x})\} + \cdots . \tag{1.19}$$

If there are n points in the neighbourhood of $\mathbf{x_0}$, the target point, the Hermite polynomials are estimated by

$$\hat{H}_k^{\mathrm{K}}\{Y(\mathbf{x_0})\} = \sum_{i=1}^{n} \lambda_{ik} H_k\{Y(\mathbf{x}_i)\}, \tag{1.20}$$

which are then inserted into Eq. 1.19. The kriging weights, λ_{ik}, are found by solving the equations for simple kriging because we assume that the mean is known:

$$\sum_{i=1}^{n} \lambda_{ik} \text{cov}[H_k\{Y(\mathbf{x}_j)\}, H_k\{Y(\mathbf{x}_i)\}] = \text{cov}[H_k\{Y(\mathbf{x}_j)\}, H_k\{Y(\mathbf{x}_0)\}] \quad \text{for all } j. \tag{1.21}$$

The procedure enables us to estimate $Z(\mathrm{x}_0)$ by

$$\hat{Z}(\mathbf{x}_0) = \Phi\{\hat{Y}(\mathbf{x}_0)\} = \varphi_0 + \varphi_1[\hat{H}_1^{\mathrm{K}}\{y(\mathbf{x}_0)\}] + \varphi_2[\hat{H}_2^{\mathrm{K}}\{y(\mathbf{x}_0)\}] + \cdots . \tag{1.22}$$

The disjunctive kriging variance of $\hat{f}[Y(\mathbf{x}_0)]$ is

$$\sigma^2_{\mathrm{DK}}(\mathbf{x}_0) = \sum_{k=1}^{\infty} f_k^2 \sigma_k^2(\mathbf{x}_0). \tag{1.23}$$

Conditional Probability Once the Hermite polynomials have been estimated at a target point, the conditional probability that the true value there exceeds or is less than the critical value, z_{c}, is calculated. The transformation $Z(\mathbf{x}) = \Phi[Y(\mathbf{x})]$ means that z_{c} has an equivalent y_{c} on the standard normal scale. The probability of exceeding the threshold is:

$$\begin{aligned} \hat{\Omega}^{\mathrm{DK}}[z(\mathbf{x}_0) \geq z_c] &= \hat{\Omega}^{\mathrm{DK}}[y(\mathbf{x}_0) \geq y_c] \\ &= 1 - G(y_{\mathrm{c}}) - \sum_{k=1}^{L} \frac{1}{\sqrt{k}} H_{k-1}(y_{\mathrm{c}}) g(y_{\mathrm{c}}) \hat{H}_k^{\mathrm{K}}\{y(\mathbf{x}_0)\}, \end{aligned} \tag{1.24}$$

and of being less than the threshold it is

$$\begin{aligned} \hat{\Omega}^{\mathrm{DK}}[z(\mathbf{x}_0) < z_c] &= \hat{\Omega}^{\mathrm{DK}}[y(\mathbf{x}_0) < y_c] \\ &= G(y_{\mathrm{c}}) + \sum_{k=1}^{L} \frac{1}{\sqrt{k}} H_{k-1}(y_{\mathrm{c}}) g(y_{\mathrm{c}}) \hat{H}_k^{\mathrm{K}}\{y(\mathbf{x}_0)\}, \end{aligned} \tag{1.25}$$

The probabilities can be mapped in the same way as the predictions and disjunctive kriging variances.

1.3 Case Study: Football Field

Football field is an 11 ha field on the Shuttleworth Estate in Bedfordshire, England (UK Ordnance Survey grid reference TL 142447). Physiographically the field comprises a gently sloping plateau in the south and a central portion with a steeper slope that leads to a level area in the north. The Lower Greensand (Cretaceous)

Fig. 1.2 Sampling schemes for soil properties at Football Field, Bedfordshire, England: (**a**) topsoil properties (0–15 cm) and (**b**) sub-soil properties (30–60 cm)

underlies about two thirds of the field to the south and the Oxford Clay (Upper Jurassic) underlies the rest to the north. The soil in the northern area is a clay loam, whereas the texture for the rest of the field is a sandy loam. The sampling frame was placed within the field to avoid the headlands. The topsoil (0–15 cm) was sampled on a square grid of 20 × 20 m with additional samples at an interval on 10 m from randomly selected grid nodes (Fig. 1.2). The aim of the additional sampling was to identify any local variation and to reduce the nugget variance of the variogram. The additional samples were taken along both the eastings and northings to avoid bias. Subsoil sampling was less intensive on a basic 40-m grid, but with additional samples at 20 and 10 m. For the topsoil, 10 cores of soil were taken from an area of 5×2 m around each grid node and were mixed to form a bulked sample. Three cores of soil were taken from a 1 m^2 area around the sampling sites of the subsoil sampling scheme and these were also bulked. Several soil properties were determined in the laboratory, but top- and sub-soil extractable potassium (K) and topsoil pH only are examined in this Chapter (see Oliver and Carroll 2004 for a full description of the analyses). The analyses followed standard laboratory methods (MAFF 1986). The yield for winter wheat 1999 is described for this Chapter, although other years of yield were also available. The yield was measured with a Massey Ferguson system.

1.3.1 Summary Statistics

Table 1.1 lists the summary statistics. The range in values for topsoil K is larger than that for the subsoil; however, the variance for subsoil K is substantially

Table 1.1 Summary statistics

Statistic	Topsoil (K mg l^{-1})	Subsoil (K mg l^{-1})	Subsoil (K mg l^{-1}) outliers removed	Subsoil (log_{10}K mg l^{-1})	Topsoil pH	Yield 1999 (t ha^{-1})
Number of samples	244	112	108	244	244	7921
Mean	166.2	145.0	138.8	2.208	6.562	4.785
Median	165.0	139.5	136.0	2.218	6.600	4.679
Minimum	49.90	69.49	69.49	1.698	4.990	1.753
Maximum	356.7	335.7	239.1	2.552	7.590	7.742
Range	306.8	266.2	169.6	0.854	2.600	5.989
Variance	1636.1	2137.4	1140.2	0.011	0.138	1.425
Standard deviation	40.45	46.23	33.8	0.107	0.372	1.194
Skewness	0.970	1.678	0.500	−0.536	0.682	0.249

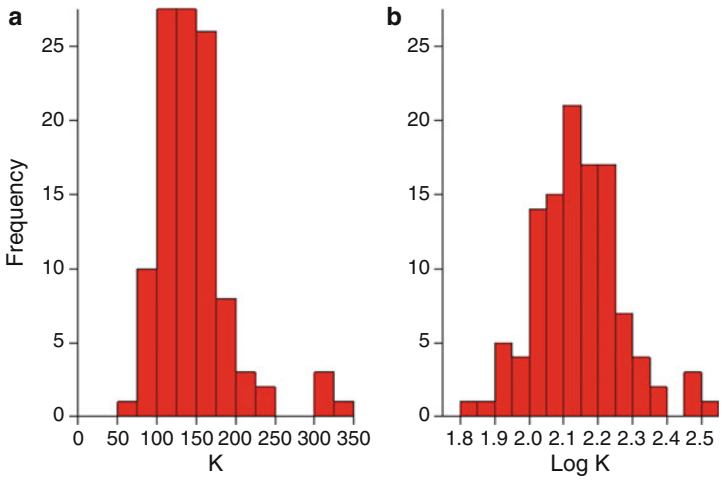

Fig. 1.3 Histograms of subsoil K mg $K\,l^{-1}$: (**a**) raw data and (**b**) transformed to common logarithms, log^{10} K

larger. The range for pH is large, 2.6 units, and also for yield for which it is almost 6t ha^{-1}. Table 1.1 shows that subsoil K has a coefficient of skewness >1 and is the only variable here that requires further investigation of its asymmetry. The histogram, Fig. 1.3a, gives a weak indication that there are four outliers with values >300 mg $K\,l^{-1}$ rather than a long tail of larger values.

1.3.2 Variography

Experimental variograms were computed for all the variables by Matheron's estimator, Eq. 1.7. To explore the data for any anisotropy, i.e. directional variation,

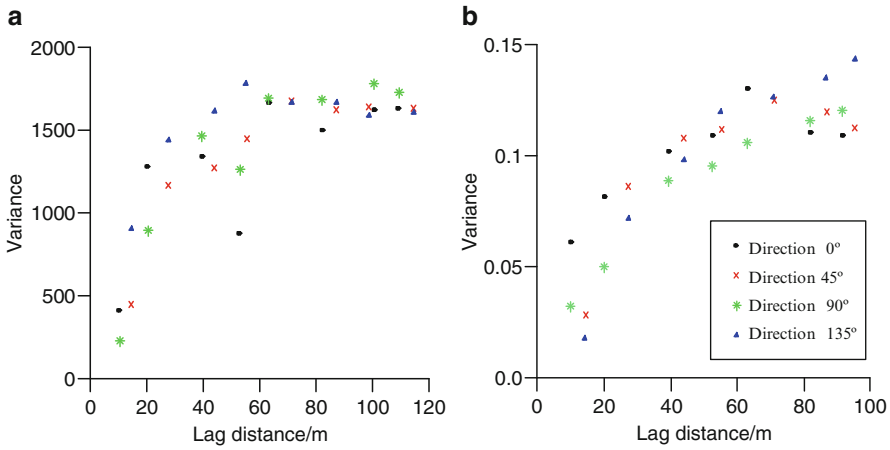

Fig. 1.4 Directional variograms computed over four directions with an angular discretization of 22.5° for: (**a**) topsoil K and (**b**) pH

experimental variograms were computed in four directions on the raw data for topsoil K and pH. With four directions I set the angular discretization to 22.5°; this angle can be decreased if there appears to be anisotropy. Figure 1.4 shows the directional variogram for topsoil K and pH; there is no sign of any anisotropy and the variation may be treated as isotropic.

Omnidirectional variograms were computed and **h** is replaced by h. Several models were fitted to the experimental values by least squares approximation in GenStat (Payne 2008). The best fitting model was selected as the one with the smallest residual sum of squares (RSS). For topsoil K the variogram was computed first with a lag interval of 10 m, which is the smallest sampling interval on the short transects. This resulted in a somewhat erratic experimental variogram that can be difficult to model (Fig. 1.5a). Therefore, the variogram was recomputed with a lag interval of 15 m which is midway between the grid interval and the transects. This has resulted in a much smoother variogram (Fig. 1.5b). The best fitting model to both experimental variograms is an exponential function with no nugget variance; its equation is

$$\gamma(h) = c\left\{1 - \exp\left(-\frac{h}{r}\right)\right\}, \tag{1.26}$$

where $\gamma(h)$ is the semivariance at lag h, c is the a priori variance of the autocorrelated process and r is a distance parameter for this function. The exponential model approaches its sill gradually and also asymptotically so that it does not have a finite range. In practice, an effective range is assigned as the distance at which the function has reached 95% of c. The effective range, a', is $3r$. Table 1.2 gives the parameters of these models; there is little difference between them. The sill variance

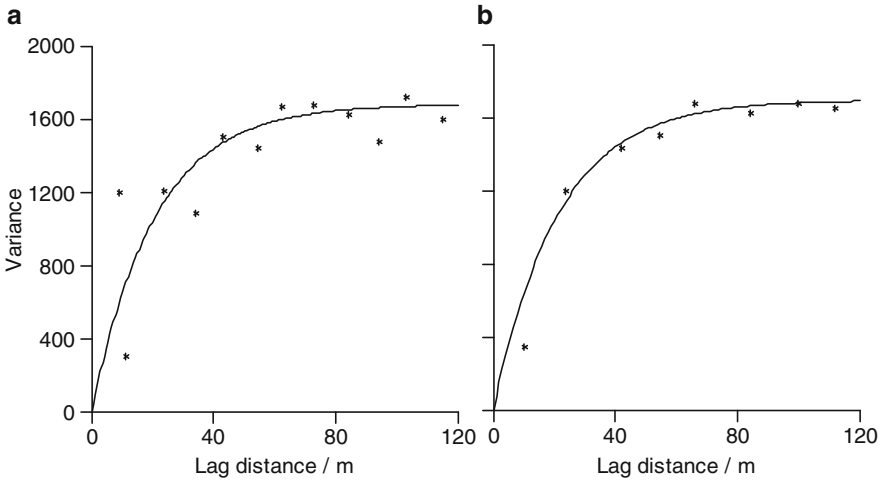

Fig. 1.5 Experimental variograms (*symbols*) and fitted exponential models (*solid lines*) for topsoil K computed with a lag interval of: (**a**) 10 m and (**b**) 15 m

and distance parameter of the model fitted to the variogram computed with a lag of 15 m are a little larger than are those for variogram computed with a lag of 10 m.

For topsoil pH, a spherical model with a nugget variance provided the best fit to the experimental variogram computed with a lag of 15 m, Fig. 1.6. The equation for this function is

$$\gamma(h) = \begin{cases} c_0 + c\left\{\dfrac{3h}{2a} + \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3\right\} & \text{for } h \leq a, \\ c_0 + c & \text{for } h > a, \\ 0 & \text{for } h = 0, \end{cases} \tag{1.27}$$

where c_0 is the nugget variance, which represents the spatially uncorrelated variation at distances less than the sampling interval and measurement error, and a is the range of spatial dependence or spatial autocorrelation. Table 1.2 gives the model parameters.

The four outliers with values $>300\,\text{mg K l}^{-1}$, Fig. 1.3a, were removed and summary statistics were computed on the remaining data. The removal of the four large values (<4% of the data) has had a marked effect on the summary statistics. The skewness coefficient has decreased to 0.5 (Table 1.1) and the variance has also decreased markedly to almost half of that with all the data. The data were also transformed to common logarithms ($\log_{10}$) and the summary statistics of the transformed data are also given in Table 1.1. The skewness coefficient is again substantially smaller than that for the raw data, 0.536, nevertheless, the histogram of

Table 1.2 Parameters of the models fitted to the experimental variograms

		Model parameters			
Variable	Model type	Nugget variance c_0	Correlated component $c_1, c_2^{\blacklozenge}$	Range a_1(m) a_2(m)$^+$ or r(m)*	Sill variance $c_0 + c$
Topsoil K (0–15 cm) Lag 10	Exponential	0	1683.0	20.70 (62.10)	1683.0
Topsoil K (0–15 cm) Lag 15	Exponential	0	1698.0	21.18 (63.54)	1698.0
Topsoil K Hermite polynomials	Exponential	0	1.00	24.09 (72.27)	1.00
Subsoil K (30–60 cm)	Exponential	0	2447.0	25.49 (76.47)	2447.0
Subsoil K No outliers	Spherical	201.3	1159	124.1	1360.3
Subsoil $\log_{10}$ K	Exponential	0	0.0178	33.13 (99.39)	0.0178
Topsoil pH	Spherical	0.0292	0.0938	74.93	0.1230
Topsoil pH Hermite polynomials	Penta.	0.1911	0.8089	83.20	1.00
Yield 1999	Double sph.	0.1975	0.4318 0.8415$^{\blacklozenge}$	33.88 137.8$^+$	1.471

Penta. is pentaspherical and sph. is spherical

$^{\blacklozenge}$ is the spatially correlated variance of the long-range spatial component.

$^+$is the range of the long-range spatial component.

*is the distance parameter of the exponential function; to obtain a working range $a' = 3r$ (values in parentheses).

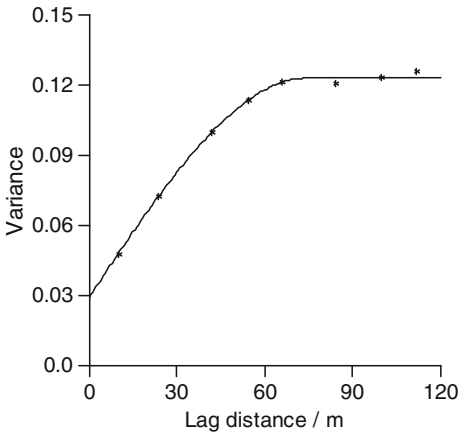

Fig. 1.6 Experimental variogram (*symbols*) and fitted spherical model (*solid line*) for topsoil pH computed with a lag interval of 15

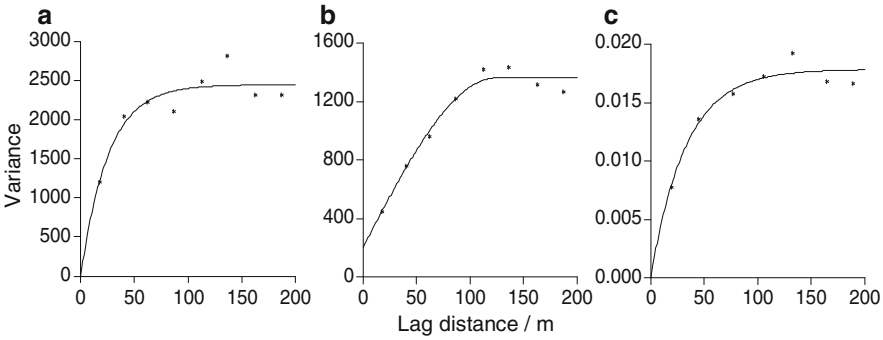

Fig. 1.7 Experimental (*symbols*) and model variograms (*line*) of subsoil K mg l^{-1} computed from: (**a**) the raw data, (**b**) data with the four largest values >300 mg K l^{-1} removed and (**c**) data transformed to $\log_{10}$

$\log_{10}$ K shows that the outliers are still present (Fig. 1.3b). Experimental variograms were computed on the raw data, the data with outliers removed and on data transformed to $\log_{10}$ for subsoil K, Fig. 1.7a–c.

The variogram of the raw data is the most erratic, but only marginally so, and the one computed on the transformed data is the most smooth. The exponential function provides the best fit to the raw and $\log_{10}$ data; Table 1.2 gives the model parameters. Both exponential models have a zero nugget variance, but the distance parameter of the variogram of the transformed data is about 25% longer. The spherical function provides the best fit to the experimental variogram with the outliers removed. The range of spatial dependence for this function is considerably larger than for the raw data, and the sill variance is also much less (Table 1.2).

I used cross-validation to determine the best fitting model for subsoil K. The approach I used involved leaving a sampling point out and kriging the value at that point with the model and the data within the search neighbourhood. The mean error (ME), mean squared error (MSE) and the mean squared deviation ratio (MSDR) were determined by the following equations, respectively:

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}\{z(\mathbf{x}_i) - \hat{z}(\mathbf{x}_i)\}, \tag{1.28}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\{z(\mathbf{x}_i) - \hat{z}(\mathbf{x}_i)\}^2, \tag{1.29}$$

and

$$\mathrm{MSDR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\{z(\mathbf{x}_i) - \hat{z}(\mathbf{x}_i)\}^2}{\hat{\sigma}^2(\mathbf{x}_i)}. \tag{1.30}$$

where N is the number of data values, $z(\mathbf{x}_i)$ is the true value at $\mathbf{x}_i$ and $\hat{z}(\mathbf{x}_i)$ is the estimated value there, and $\hat{\sigma}^2(\mathbf{x}_i)$ is the kriging variance. The closer the MSDR is to 1 the better is the model for kriging. Table 1.3 gives the results of cross-validation for the three models fitted to subsoil K. The mean errors are small because kriging is unbiased, but there is a considerable difference in the MSEs between the raw data and the data with the four largest values removed. The MSDRs, however, indicate that the model fitted to the raw data is the best for kriging and the one fitted to the variogram with the four largest values removed is the poorest, but by only a small margin. Overall, the results show that all three models could be used for kriging without incurring large errors. The results also illustrate the importance of examining the model by cross-validation before proceeding to transform the data or to remove values considered as outliers. Kerry and Oliver (2007b, c) showed that transformation is not always necessary for skewness levels of the kind we have here for subsoil K.

The experimental variogram computed on the yield data for 1999 suggests that there might be more than one scale of variation present (Fig. 1.8). There is a change in the slope of the experimental values at a lag of about 45 m that suggests more than

Table 1.3 Mean errors (ME), mean squared errors (MSE) and the mean squared deviation ratio (MSDR) for subsoil K

Model	ME	MSE	MSDR
Subsoil K raw data	0.4386	1365	1.014
Subsoil K raw data – outliers removed	0.2735	664.4	1.178
Log_{10} subsoil K	0.0008766	0.009068	1.120

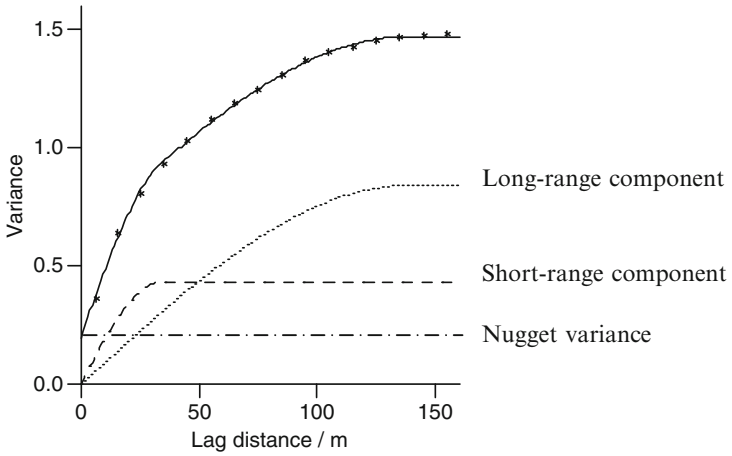

Fig. 1.8 Experimental variogram (*symbols*) and the fitted nested spherical model (*solid line*) of yield 1999. The model was decomposed to illustrate the individual components as shown by the ornamented lines

one scale of variation might be present. A nested spherical model was fitted to the experimental values and this function provided the best fit in terms of the residual mean square (RMS). Its equation is given by

$$\gamma(h) = \begin{cases} c_0 + c_1 \left\{ \dfrac{3h}{2a_1} - \dfrac{1}{2}\left(\dfrac{h}{a_1}\right)^3 \right\} + c_2 \left\{ \dfrac{3h}{2a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3 \right\} & \text{for } 0 < h \le a_1, \\ c_0 + c_1 + c_2 \left\{ \dfrac{3h}{2a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3 \right\} & \text{for } a_1 < h \le a_2, \\ c_0 + c_1 + c_2 & \text{for } h \le a_2, \\ 0 & \text{for } h = 0. \end{cases} \tag{1.31}$$

where c_1 and a_1 are the sill and range of the short-range component of the variation, respectively, and c_1 and a_1 are the sill and range of the long-range component respectively. A nugget component can also be added as above. Figure 1.8 shows the components of the nested model fitted to yield; the nugget, short-range and long-range components.

1.3.3 Kriging

1.3.3.1 Ordinary Kriging

The models for topsoil K and pH were used for ordinary block kriging with the data; this is the method that farmers would be most likely to use for precision agriculture because they are dealing with areas. Although the kriging variances are much smaller than for punctual kriging, the estimates are usually little different. Predictions were made over blocks of 5×5 m for topsoil K. An arbitrary block size was chosen in this case, but a farmer would usually choose a size that relates to the operating dimensions of the relevant machine. The map of block kriged predictions (Fig. 1.9a) shows that there are large areas of the field with values of K that would trigger the need for K applications. There are also other areas where there is no need for any additional K. This map shows the potential in this field for variable-rate (VRT) fertilizer applications. The predictions on which this map is based are accurate because of the intensive sampling, and such maps have an important role to play in VRT applications. The map of kriging variances (Fig. 1.9b) shows that the smallest variances are along the short transects where sampling was at an interval of 10 m and the largest ones are near the field boundary where there are fewer samples in the kriging neighbourhood.

The map of ordinary block kriged estimates of pH (Fig. 1.10) shows that there is considerable variation in pH in this field and that there is scope for variable-rate applications of lime. The map of kriging variances shows a similar pattern to that in Fig. 1.9b and so is not included.

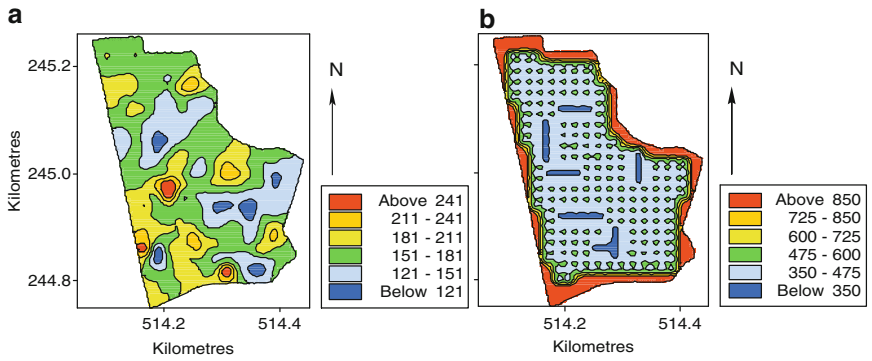

Fig. 1.9 Maps of topsoil K (mg $K\,l^{-1}$) from ordinary block kriging: (**a**) estimates and (**b**) kriging variances

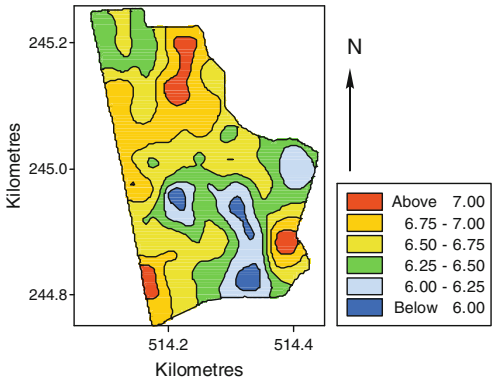

Fig. 1.10 Map of ordinary block kriged estimates of topsoil pH

The model parameters for the three sets of data for subsoil K were used for ordinary block kriging, and Fig. 1.11 shows the resulting maps of the kriged predictions. There is little difference in the three maps of predictions; the one based on the raw data with the variogram computed on data with the four largest values removed is the most different (Fig. 1.11c). Maps of the kriging variances, Fig. 1.11b, d, are more different from one another. The map based on the raw data, Fig. 1.11b, has larger kriging variances than the one in Fig. 1.11d where the variogram was computed on data with the four largest values (>300 mg $K\,l^{-1}$) removed. The difference reflects the larger variance of the data with the outliers present. The small kriging variances associated with the sampling are very clear in Fig. 1.11b. These results suggest that analysts should not adopt a mechanical approach to removal of either the largest or smallest values, or to transforming their data. The variograms should be examined for the various scenarios first. If that of the raw data is more erratic and difficult to model than one computed from transformed data, it would be prudent to use the

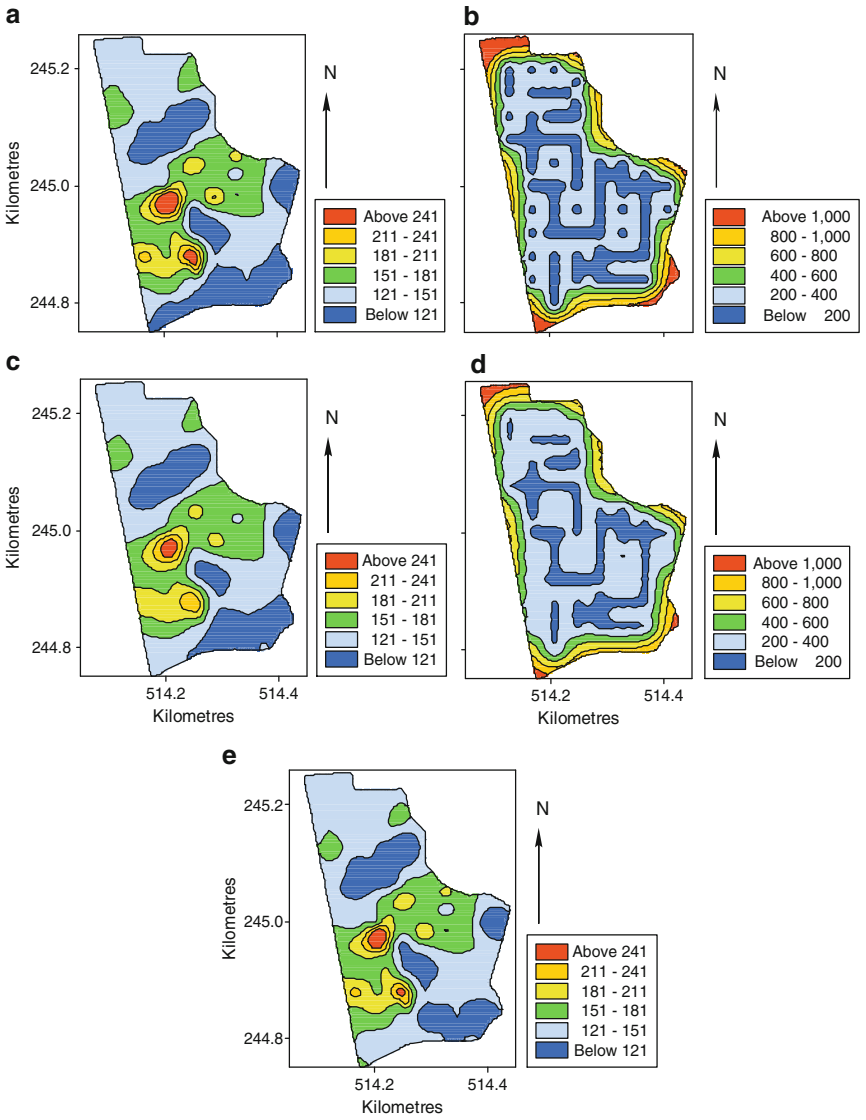

Fig. 1.11 Maps of ordinary block kriged estimates of subsoil K (mg $K\,l^{-1}$) from: (**a**) raw data (112 sites), (**c**) raw data (112 sites) and variogram computed on data with the four largest values (>300 mg $K\,l^{-1}$) removed and (**e**) back-transformed values from $\log_{10}$. Maps of kriging variances for: (**b**) raw data and (**d**) data with four largest values removed

transformed data. Oliver et al. (2002) showed that skewness values outside the envelope of the common limits of ± 1 do not necessarily signify a need to transform data for further analysis in linear geostatistics, especially where the number of data is large.

1.3.3.2 Disjunctive Kriging

As the maps of ordinary kriged estimates showed the potential for VRT, I used disjunctive kriging to examine the probabilities that the true values would be less than a threshold value for topsoil K and pH. The data were transformed first to Hermite polynomials to ensure a Gaussian distribution (see Webster and Oliver (2007). The experimental variogram was computed on the Hermite transformed data for K, and the best fitting model is an exponential function (see Fig. 1.12a). Table 1.2 gives the parameters of the model. The block disjunctively kriged estimates (the analysis was done in ISATIS (www.geovariances.fr)) of K are not included because they are so similar to those from ordinary kriging. Figure 1.12b shows the probabilities of the true values' being $<181\,\mathrm{mg\,K\,l^{-1}}$, which is equivalent to 2+ in the MAFF RB209 (MAFF 2000) guidelines for cereal production. Areas of the field with probabilities >0.3 would need additional K. They are in fairly well defined parts of the field which would make variable-rate application feasible. This approach would reduce fertilizer requirements compared with uniform application. The results show that farmers could use this approach to decide between uniform and VRT application (Fig. 1.12b).

The topsoil data for pH were also transformed to Hermite polynomials and the variogram computed on the transformed data. Figure 1.13a shows the experimental variogram and the fitted pentaspherical model; Table 1.2 gives the model parameters. The equation for this function is given by

$$\gamma(h) = \begin{cases} c_0 + c\left\{\dfrac{15h}{8a} - \dfrac{5}{4}\left(\dfrac{h}{a}\right)^3 + \dfrac{3}{8}\left(\dfrac{h}{a}\right)^5\right\} & \text{for } h \le a, \\ c_0 + c & \text{for } h > a, \\ 0 & \text{for } h = 0. \end{cases} \tag{1.32}$$

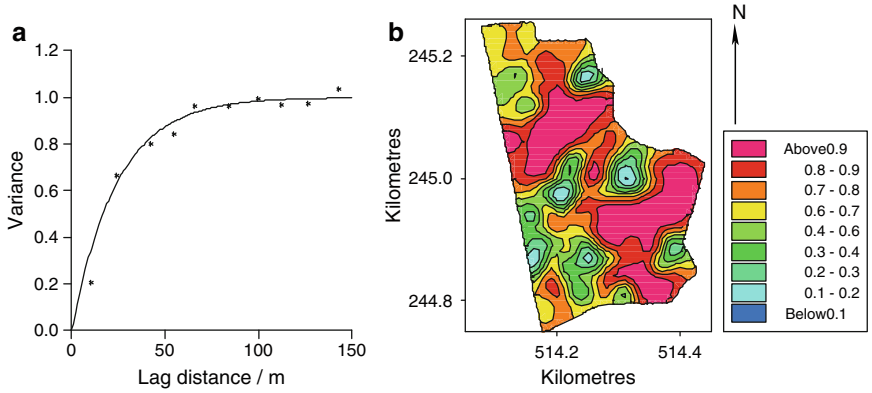

Fig. 1.12 (**a**) Experimental variogram (*symbols*) and the fitted exponential function (*solid line*) for the Hermite polynomials of topsoil K, (**b**) map of probabilities that the true values of topsoil K are less than a threshold of $181\,\mathrm{mg\,K\,l^{-1}}$

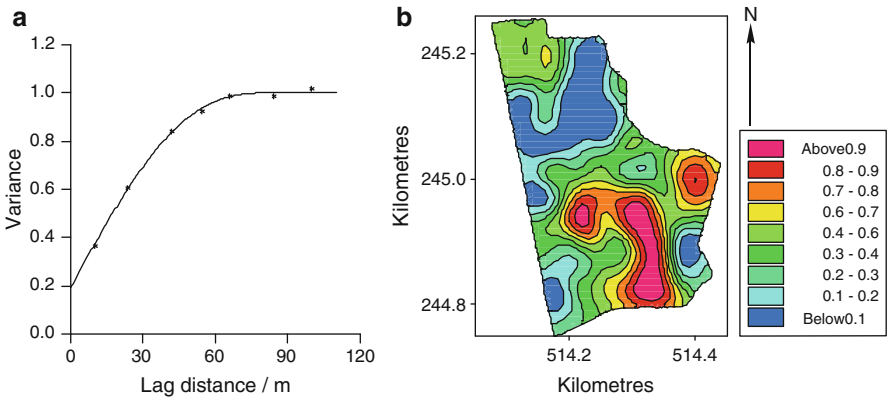

Fig. 1.13 (**a**) Experimental variogram (*symbols*) computed on the Hermite polynomials of topsoil pH and the fitted exponential function (*solid line*) and (**b**) probabilities that the true values of topsoil pH are less than a threshold of 6.5

This model curves more gently as it approaches its sill than does the spherical one, Fig. 1.6. The optimal pH for continuous arable cropping is 6.5. The probabilities of the true values of pH being less than this threshold are shown in Fig. 1.13b. This map shows very specifically where the farmer needs to apply lime and where applications would be unnecessary. Areas with probabilities <0.3 would not require lime based on the measured pH and this would represent a considerable saving in resources and the farmer's time. This map could also be used to prioritize management. The south central and eastern parts of the field clearly have a pH that is too low for optimal cereal growth; the farmer could focus attention here in the short term.

1.3.3.3 Factorial Kriging

As the variogram of the yield data indicates variation at two spatial scales, I used factorial kriging to filter out and estimate the long- and short-range components separately (see Webster and Oliver 2007 for more information on this method). Figure 1.14a shows the ordinary kriged predictions of yield (1999); there is no strong evidence of the two spatial scales of variation in this map. The map of long-range estimates, Fig. 1.14b, is similar to the ordinary kriged map; it is slightly less detailed however. The map of the short-range estimates, Fig. 1.14b, shows clearly that this component of variation is the result of management. There are lines in the variation that relate to tramlines that are parallel to the long axis of the field and to the effect of traffic on the headlands in the southern part of the field. The green lineations are where the yield is less because of compaction from the farm machinery.

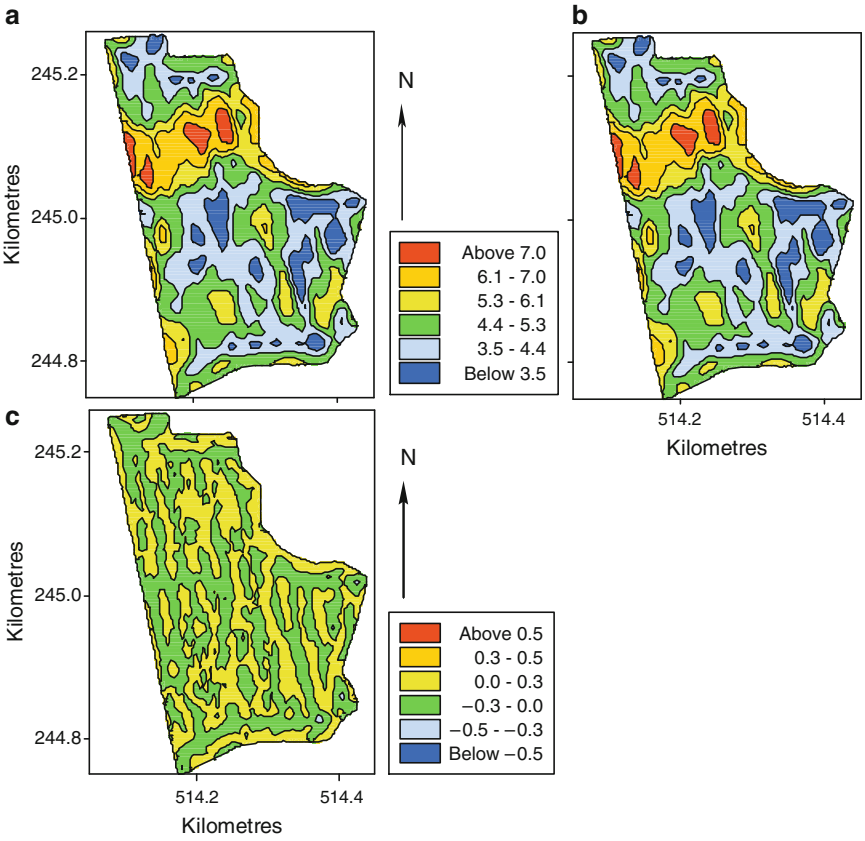

Fig. 1.14 Maps of predictions of yield 1999 for the Yattendon Estate from factorial kriging: (**a**) map of predictions based on the nested spherical model, (**b**) map of long-range component of the variation and (**c**) map of short-range component

1.3.4 Conclusions

This case study describes the stages that a geostatistical analysis should go through. First, an exploratory data analysis to determine the distribution of the data and whether there is a need to remove outliers or to transform the data to obtain a near-normal distribution. The choice of a suitable lag interval and the effects on the model parameters are illustrated. The cross-validation analysis showed that transformation or the removal of outliers when skewness levels are outside the usual bounds is not always necessary. It is always better to work with the raw data if possible. When data are transformed, they usually need to be back-transformed for mapping so that the predictions are on the original measurement scale for the farmer to work with. Disjunctive kriging provides an objective way of prioritizing management where resources and or time are scarce. It also shows the farmer which areas really require

treatment and which do not. In this way excess applications can be avoided to limit environmental impact and improve the farm's economy. Factorial kriging in the example here could be used to identify localized areas for subsoiling where traffic pressure has reduced yields in the vicinity of the tramlines.

References

Bhatti, A. U., Mulla, D. J., & Frazier, B. E. (1991). Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images. *Remote Sensing of the Environment*, *37*, 181–191.

Burgess, T. M., & Webster, R. (1980a). Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *Journal of Soil Science*, *31*, 315–331.

Burgess, T. M., & Webster, R. (1980b). Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *Journal of Soil Science*, *31*, 333–341.

Burgess, T. M., Webster, R., & McBratney, A. B. (1981). Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science*, *32*, 643–654.

Cochrane, W. W. (1993). *The development of American agriculture. A historical analysis.* Minneapolis, MN, USA: University of Minnesota Press.

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York: Wiley.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Journel, A. G. (1983). Non-parametric estimation of spatial distributions. *Journal of the International Association of Mathematical Geology*, *15*, 445–468.

Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*. London: Academic.

Kerry, R., & Oliver, M. A. (2007a). Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma, Pedometrics 2005*, *140*, 383–396.

Kerry, R., & Oliver, M. A. (2007b). Determining the effect of asymmetric data on the variogram. I. Underlying asymmetry. *Computers and Geosciences*, *33*, 1212–1232.

Kerry, R., & Oliver, M. A. (2007c). Determining the effect of asymmetric data on the variogram. II. Outliers. *Computers and Geosciences*, *33*, 1233–1260.

Kolmogorov, A. N. (1939). Sur l'interpolation et l'extrapolation des suites stationnaires. *Comptes Rendus de l'Académie des Sciences de Paris*, *208*, 2043–2045.

Kolmogorov, A. N. (1941). The local structure of turbulence in an incompressible fluid at very large Reynolds numbers. *Doklady Academii Nauk SSSR*, *30*, 301–305.

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

Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On optimal prediction of soil properties in the presence of spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, *57*, 787–799.

Laslett, G. M., McBratney, A. B., Pahl, P. J., & Hutchinson, M. F. (1987). Comparison of several spatial prediction methods for soil pH. *Journal of Soil Science*, *38*, 325–341.

Luellan, W. R. (1985). Fine-tuned fertility. Tomorrow's technology here today. *Crops and Soils Magazine*, *38*, 18–22.

MAFF. (1986). *MAFF reference book 427: The analysis of agricultural materials* (3rd ed.). London: Her Majesty's Stationery Office.

MAFF. (2000). *Fertiliser recommendations for agricultural and horticultural crops: RB209* (7th ed.). London: Her Majesty's Stationery Office.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*, 1246–1266.
Matheron, G. (1965). *Les variables régionalisées et leur estimation*, *une application de la théorie de fonctions aléatoires aux sciences de la nature*. Paris: Masson et Cie.
Matheron, G. (1969). *Le krigeage universel*. Cahiers du Centre de Morphologie Mathématique, No. 1. Fontainebleau: Ecole des Mines de Paris.
Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, *5*, 439–468.
Matheron, G. (1976). A simple substitute for conditional expectation: The disjunctive kriging. In M. Guarascio, M. David, & C. Huijbregts (Eds.), *Advanced geostatistics in the mining industry* (pp. 221–236). Dordrecht: D. Reidel.
Matheron, G. (1982). *Pour une analyse krigeante de données régionalisées*. Note N-732 du Centre de Géeostatistique. Fontainebleau: Ecole des Mines de Paris.
Mercer, W. B., & Hall, A. D. (1911). The experimental error of field trials. *Journal of Agricultural Science*, *4*, 107–132.
Miesch, A. T. (1975). Variograms and variance components in geochemistry and ore evaluation. *Geological Society of America Memoir*, 142, 333–340.
Miller, M.P., Singer, M.J., & Nielsen, D.R. (1988). Spatial variability of wheat yield and soil properties on complex hills. *Soil Science Society of America Journal*, *52*, 1133–1141.
Mulla, D. J. (1989). Using geostatistics to manage spatial variability in soil fertility. In C. M. Renard, R. J. Van den Beldt, & J. F. Parr (Eds.), *Soil, crop and water management in the Sudano-Sahelian zone* (pp. 241–254). Pantcheru, India: ICRISAT.
Mulla, D. J. (1991). Using geostatistics and GIS to manage spatial patterns in soil fertility. In G. Kranzler (Ed.), *Proceedings of symposium on automated agriculture for the 21st century* (pp. 336–345). St Joseph, MI: American Society of Agricultural Engineers.
Mulla, D. J. (1993). Mapping and managing spatial patterns in soil fertility and crop yield. In P. C. Robert, W. E. Larson and R. H. Rust (Eds.), *Proceedings of Site Specific Crop Management: A Workshop on Research and Development Issues* (pp. 15–26). Madison, WI: American Society of Agronomy; Crop Science Society of America; Soil Science Society of America.
Mulla, D. J., & Hammond, M. W. (1988). Mapping soil test results from large irrigation circles. In J. S. Jacobsen (Ed.), *Proceedings of the 39th Annual Far West Regional Fertilizer Conference* (pp. 169–171). Pasco, WA: Agricultural Experimental Station Technical Paper No. 8597.
National Research Council. (1997). *Precision agriculture in the 21st century*. Washington DC, USA: National Academy Press.
Oliver, M. A., & Carroll, Z. L. (2004). *Description of spatial variation in soil to optimize cereal management*. Project Report No. 330. HGCA, London.
Oliver, M. A., Webster, R., & Slocum, K. (2000). Filtering SPOT imagery by kriging analysis. *International Journal of Remote Sensing*, *21*, 735–752.
Oliver, M. A., Loveland, P. J., Frogbrook, Z. L., Webster, R., & McGrath, S. P.(2002). *Statistical and geostatistical analysis of the national soil inventory of England and Wales* (MAFF project SPO124, available on CD ROM from National Soil Resources Institute, Cranfield University, UK).
Omre, H. (1987). Bayesian kriging – merging observations and qualified guesses in kriging. *Mathematical Geology*, *19*, 25–39.
Pardo-Igúzquiza, E. (1997). MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. *Computers and Geosciences*, *23*, 153–162.
Pardo-Igúzquiza, E. (1998). Inference of spatial indicator covariance parameters by maximum likelihood using MLREML. *Computers and Geosciences*, *24*, 453–464.
Payne, R. W. (2008). *The guide to GenStat for GenStat release 10: Part 2, statistics*. Hemel Hempstead, UK: VSN International.
Rivoirard, J. (1994). *An introduction to disjunctive kriging and non-linear geostatsitics.* Oxford: Clarendon.
Robert, P. C. (1999). Precision agriculture: Research needs and status in the USA. In J. V. Stafford (Ed.), *Precision Agriculture'99* (pp. 19–33). Sheffield, UK: Sheffield Academic Press.

Robert, P. C., Rust, R. H., & Larson, W. E. (Eds.) (1993). *Proceedings of Site Specific Crop Management: A Workshop on Research and Development Issues.* Madison, WI: American Society of Agronomy; Crop Science Society of America; Soil Science Society of America.

Robert, P. C., Rust, R. H., & Larson, W. E. (Eds.) (1995). *Proceedings of the Second International Conference on Site-Specific Management for Agricultural Systems.* Madison, WI: American Society of Agronomy; Crop Science Society of America; Soil Science Society of America.

Schafer, R. L., Young, S. C., Hendrick, J. G., & Johnson, C. E. (1984). Control concepts for tillage systems. *Soil & Tillage Research*, *4*, 313–320.

Schueller, J. K. (1997). Technology for precision agriculture. In J. V. Stafford (Ed.), *Precision Agriculture'97* (pp. 19–33). Oxford, UK: BIOS Scientific Publishers.

Searcy, S. W., Schueler, J. K., Bae, Y. H., Borgelt, S. C., & Stout, B. A. (1989). Mapping of spatially-variable yield during grain combining. *Transactions of the ASAE*, *32*, 826–829.

Stafford, J. V. (Ed.) (1997). *Precision Agriculture'97*. Oxford, UK: BIOS Scientific Publishers.

Stein, M. L. 1999. *Interpolation of spatial data: Some theory for kriging*. New York: Springer.

Sullivan, J., (1984). Conditional recovery estimation through probability kriging: Theory and practice. In G. Verly, M. David, A. G. Journel, & A. Marechal (Eds.), *Geostatistics for natural resources characterization* (pp. 365–384). Dordrecht: D. Reidel.

Viscarra Rossel, R. A., & McBratney, A. B. (1997). Preliminary experiments towards the evaluation of a suitable soil sensor for continuous 'on-the-go' field pH measurements. In J. V. Stafford (Ed.), *Precision agriculture'97* (pp. 493–501). Oxford, UK: BIOS Scientific Publishers.

Warrick, A. W., Myers, D. E., & Nielsen, D. R. (1986). Geostatistical methods applied to soil science. In A. Klute (Ed.), *Methods of soil analysis: Part 1, physical and mineralogical methods, agronomy monograph no. 9* (2nd ed.). (pp. 53–82). Madison, WI: American Society of Agronomy and Soil Science Society of America.

Webster, R., & Butler, B. E. (1976). Soil survey and classification studies at Ginninderra. *Australian Journal of Soil Research*, *14*, 1–26.

Webster, R., & Oliver, M. A. (1989). Disjunctive kriging in agriculture. In M. Armstrong (Ed.), *Geostatistics* (Vol. 1, pp. 421–432). Dordrecht: Kluwer.

Webster, R., & Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, *43*, 177–192.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. Chichester, UK: Wiley.

Youden, W. J., & Mehlich, A. (1937). Selection of efficient methods for soil sampling. *Contributions of the Boyce Thompson Institute for Plant Research*, *9*, 59–70.

Chapter 2
Sampling in Precision Agriculture

R. Kerry, M.A. Oliver and Z.L. Frogbrook

Abstract This chapter considers the importance of spatial scale in sampling and investigates various methods by which the variogram can be used to determine an appropriate sampling scheme or interval for grid sampling. When no prior information is available on the scale of variation, and the variable of interest is unlikely to be strongly correlated to available ancillary data, a nested survey and analysis provides a first approximation to the variogram and the approximate spatial scale. If the variable of interest appears related to ancillary data such as aerial photographs or elevation, variograms of these data can provide an indication of the likely scale of variation in the soil or crop. Existing variograms of soil or crop properties can be used to determine how many cores of soil or samples from plants should be taken to form a composite (bulked) sample to reduce the local noise. Such variograms can also be used with the kriging equations to determine a grid sampling interval with a specific tolerable error, or an interval of less than half the variogram range can be used to ensure a spatially dependent sample. Finally, if the scale of variation is large in relation to the field size, a variogram estimated by residual maximum likelihood (REML) or standardized variograms from ancillary data can be used to krige data from a small, but spatially dependent sample. Each of the methods investigated is illustrated with a case study.

Keywords Sampling · Method of moments variogram (MoM) · Residual maximum likelihood (REML) · Nested sample design · Bulking strategy · Soil data · Ancillary data

R. Kerry (✉)
Department of Geography, Brigham Young University, 690 SWKT, Provo, UT 84602, USA
e-mail: ruth_kerry@byu.edu

M.A. Oliver
Department of Soil Science, The University of Reading, Whiteknights,
Reading RG6 6DW, United Kingdom
e-mail: m.a.oliver@reading.ac.uk

Z.L. Frogbrook
Environment Agency Wales, Ty Cambria, 29 Newport Road, Cardiff CF24 0TP,
United Kingdom
e-mail: zoe.frogbrook@environment-agency.wales.gov.uk

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_2,

2.1 Introduction

Accurate information about the variation in soil and crop attributes within fields is crucial for precise management in agriculture. This is the essence of precision agriculture (PA) which began in the early 1990s, (see papers in Robert et al. 1995, 1996; Schueller 1997; Stafford 1997, 1999). The value of geostatistics to predict accurately for digital mapping of soil and crop attributes was recognized at a similar time (Blackmore 1994; Whelan et al. 1996; Oliver and Frogbrook 1998). The accuracy of such predictions, however, depends on the quality of information on the soil and crops. Many soil and crop attributes have to be determined physically from samples in the field, which can be time-consuming and expensive. The geostatistical method of local prediction, kriging, depends on having accurate variograms and spatially dependent data from which to predict. All methods of interpolation assume implicitly that data are spatially dependent, which means that sampling should be at an interval that is well within the correlation range of spatial variation. Despite the popularity of geostatistics in PA, we have found that the methods have been, and continue to be, applied to unsuitable data. Therefore, the aim of this chapter is to guide potential users of geostatistics in precision agriculture on the sampling requirements and cost effective approaches to sampling. Chapter 3 continues the theme of sampling with examples of more elaborate methods.

Planning the sampling for surveys appears to be the 'Cinderella' of many environmental studies; the temptation is to rush into the field without adequate preparation. However, sampling underpins the quality and accuracy of subsequent analyses and decision making because these depend on the data being suitable for the purposes intended. It is notable that there are few papers on sampling in the proceedings of either the International or European Conferences on Precision Agriculture and in the *Journal of Precision Agriculture*. We need a more objective approach to sampling in PA to provide the quantitative information that is required. McKenzie et al. (2008) reinforce the need for a clear and consistent conceptual framework for sampling. Although much of the information used in agriculture in the past was quantitative, it was based on field averages which were only adequate for uniform applications and management at the field level. Management of the variation within fields, however, requires more detailed knowledge, which can be obtained only by intensive sampling.

The variation in soil and crop properties within fields comprises variation over short distances of a few metres and that over longer distances of tens or hundreds of metres. It is this last component of variation that the precision farmer wants to resolve for management, and we can regard the short-range variation as 'noise' or a sampling effect. Sampling for traditional farm surveys in the United Kingdom and many other countries was at about one sample per hectare and sometimes it was even more sparse. This approach has also become widely used in precision farming surveys (Godwin and Miller 2003) because it is considered a sampling density that the farmer can afford. However, sampling at this density takes no account of either the scale of variation or of how many sampling points might be needed for further analyses. Geostatistics, and in particular the variogram, can be used to guide

sampling. The importance of spatial scale in sampling is considered in Section 2.1.1. For some attributes, especially of the soil where variation is complex and cannot be observed at the surface, for example nematodes (see Chapter 9), and there is no prior information, a nested survey and analysis can provide a first approximation to the variogram. This approximation can then be used to determine the approximate scale(s) of variation and a suitable sampling interval for a more detailed survey (see Section 2.2). For other attributes, such as weeds, plant diseases, some soil properties, etc., the variation might be evident visually in the field or in remote and proximally sensed imagery. The latter types of data are often referred to as ancillary data and variation in these is often linked to properties of the soil and of the crop. Variograms computed from ancillary data can be used to determine the approximate scale of spatial variation, and so can be used to guide sampling based on the 'rule of thumb' of sampling at less than half the variogram range (see Section 2.2.2).

Although we more or less dismiss the local sample to sample variation above as 'noise', we cannot disregard it because it can affect the accuracy of any future predictions. Local variation can be smoothed by mixing together several small cores of soil or, for example, by taking several cotton bolls from one or more plants, etc. over a given sample support to create a bulked or composite sample. The number of samples to take for bulking can be optimized using the variogram (see Section 2.3).

Sampling should match the objectives of the survey; for precision farming this is often to produce accurate digital maps. For this, a grid survey has been the usual choice (Viscarra-Rossel and McBratney 1998) because of its efficiency for sample collection in the field, prediction and mapping. If variograms of soil and crop properties from previous surveys exist for an area with a similar soil parent material and similar crop, they can be used with the kriging equations to determine an optimal grid interval.

If the sampling intervals recommended by the above methods are large, there may be too few data from which to compute a reliable variogram by the usual method of moments estimator. Kerry and Oliver (2007) showed that a variogram estimated by residual maximum likelihood (REML) could provide more accurate predictions with fewer data than one estimated by the conventional method of moments (see Section 2.5.1). A standardized variogram based on ancillary data or existing variograms of soil properties can also be used to krige spatially dependent data that are too sparse from which to compute a variogram (see Section 2.5.2).

2.1.1 The Importance of Spatial Scale for Sampling

Soil and crop properties can vary at markedly different spatial scales both within and between fields. Therefore, it is essential when designing a sampling scheme that the spatial scales of variation in the properties of most importance for PA management are used to guide sampling. Figure 2.1a,b illustrates the effect of spatial scale in the pixel maps of two random processes simulated with a spherical variogram. Consider that each pixel represents information on a 5-m sampling grid and the area is a 25-ha field. The variation in Fig. 2.1a occurs over short distances, whereas that in Fig. 2.1b

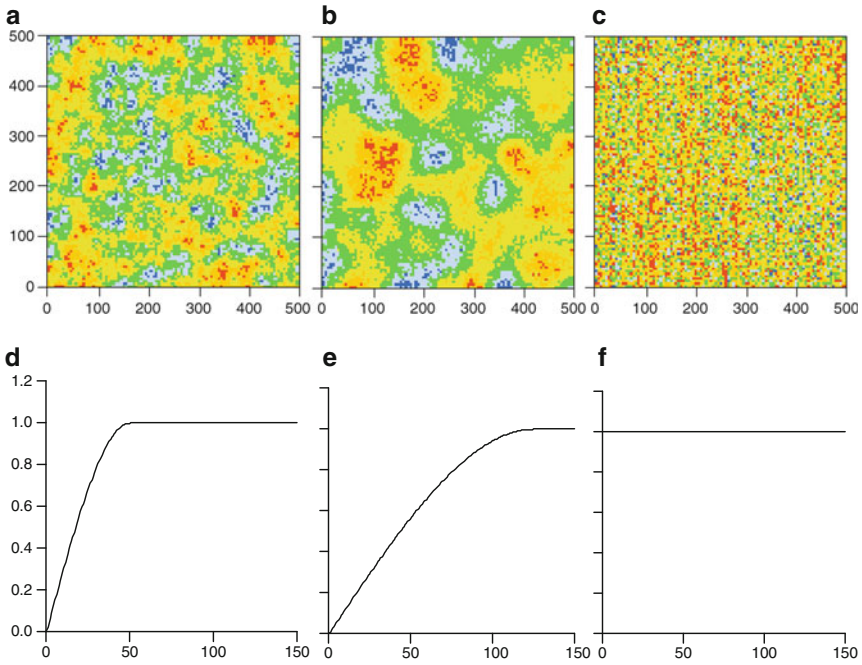

Fig. 2.1 Fields simulated with a spherical (sph) function with zero nugget variance, a sill variance of 1.0 and ranges of: (**a**) 50 m and (**b**) 125 m; (**c**) field simulated with pure nugget variogram; variogram functions used to generate the simulated fields: (**d**) sph 50 m, (**e**) sph 125 m and (**f**) pure nugget

occurs over much longer distances. The variation represented in Fig. 2.1c is 'white noise'. To sample to provide spatially dependent data would require quite different sampling intensities for the fields in Fig. 2.1a,b, whereas for Fig. 2.1c the regional mean only could be estimated. If these random processes were superimposed on one another they would result in nested variation as described for yield in Chapter 1. Therefore, we should need to choose which of the processes illustrated in Fig. 2.1a or b was the more important to resolve, and how many samples to bulk from to remove the effect of the 'noise' in Fig. 2.1c. Figure 2.1d–f shows the variograms used to generate the simulated fields; for the short-range variation the range was 50 m and for the long-range variation it was 125 m. Figure 2.1c was generated by a pure nugget variogram (Fig. 2.1f).

2.1.2 How Can Geostatistics Help?

Geostatistics embodies techniques to describe spatial autocorrelation of a regionalized variable, Z, and that use this information for local prediction by kriging. Kriging requires a model of the spatial correlation structure derived from either the

covariance function or more usually the variogram. These functions are not known a priori, and they must be estimated from sample data. Sampling for subsequent geostatistical analysis must serve two purposes: first estimation and modelling of the variogram and second local prediction by kriging. To satisfy the first purpose, sampling must be sufficient to estimate the semivariances precisely.

2.1.3 How can the Variogram be Used to Guide Sampling?

Webster and Oliver (1992) showed that at least 100 samples are required to estimate the variogram reliably, but for some agricultural fields this number might be too few to resolve the variation present if the scale is short, whereas it might be too many where the scale is long. In some situations there are no visible signs of variation because of the nature of the property of interest, and there are no clues to the approximate scale of variation. So how can we deal with these situations?

1. In the absence of any prior information about the spatial scale of variation and no visible signs of the variation in the property of interest, a nested survey and analysis can provide a first approximation to the variogram. This approximation can then be used to determine a suitable sampling interval for a more detailed survey. This is described in Section 2.2.1 together with an example.
2. In the absence of existing variograms of soil or crop properties and the property of interest appears related to ancillary data, such as those from remote and proximal sensing, digital elevation models, etc., they could be used to compute variograms from which to judge the approximate spatial scale of variation and a suitable sampling interval.
3. If variograms of the soil or crop properties are available, they can be used to determine how many cores of soil or samples from plants should be taken over a given support to form composite samples for laboratory analyses. The aim is to reduce local 'noise'.

   > The sample support is the area or volume of material on which measurements are made. It has size and shape, and it may have orientation. In crop surveys it might be a specified area or a given number of plants (see Willers et al. 2009), and in remote sensing the pixel is the support. In soil survey it is the volume or core of soil taken from the ground, or it might be a specified area from which several cores may be taken.

4. If there are variograms of the soil or crop properties from previous surveys, they can be used with the kriging equations to determine an optimal sampling interval, for a future grid survey with a specified tolerable error to avoid over- or undersampling. Existing variograms can also be used with the 'half variogram range rule of thumb' to ensure that survey data will be spatially dependent.
5. If the scale of variation is large and there are 50–100 samples, a variogram can be estimated by residual maximum likelihood (REML) and could be used to krige data from a small, but spatially dependent sample. Alternatively, existing variograms of ancillary data, soil or crop properties could be used to compute standardized average variograms to krige sparse data that have been standardized.

2.2 Variograms to Guide Sampling

2.2.1 Nested Survey and Analysis: Reconnaissance Variogram

A nested survey and analysis is advantageous in precision agriculture when there are few or no clues as to the spatial scale of variation in the property of interest and the costs of management are large. The nested survey enables several orders of magnitude of spatial scale to be examined in a single analysis (Fig. 2.2a) to determine the approximate scale of variation with no more than about 108 samples. Webster and Oliver (2007) describe nested analysis in detail, and Section 9.3.2 provides some of the theoretical background to the method. The idea underlying the model of nested variation is that a population can be divided into classes at two or more categoric levels or stages in a hierarchy. The population can then be sampled with a nested scheme to estimate the variance at each stage, i.e. the components of variance. The individual component for a given stage measures the variation attributable to that stage, and the components sum to the total variance. Miesch (1975) showed that there is a link between the results of a hierarchical analysis of variance and the semivariances of geostatistics. If the components of variance are accumulated, starting with the smallest spacing, they are equivalent over the same range of distances to the semivariances of geostatistics.

The nested analysis provides a first approximation to the variogram (Oliver and Webster 1986; Webster and Oliver 2007). It can indicate the range of spatial scales over which most of the variation occurs making it a valuable reconnaissance tool. It indicates the spatial scale at which most of the variation occurs and this information can be used to guide sampling for an overall survey or to obtain a more accurate variogram. For properties for which we have no clues about their scale of variation, such as nematode infestations, a reconnaissance variogram could avoid wasted sampling effort and costs of analysis by indicating a suitable sampling interval for a more detailed survey. The samples from the nested survey could be integrated with the later samples provided the time interval was not too great for temporally variable properties such as nematodes. For more permanent properties of the soil, the temporal aspect would not be an issue.

2.2.1.1 Unequal Sampling

Youden and Mehlich's (1937) sampling design was fully balanced with replication at each stage; the sample size doubled for each additional stage. Oliver and Webster (1986) showed that full replication at each stage is not necessary because the mean squares for the lower stages are estimated more precisely than those of the higher ones. Economy can be achieved by replication of only a proportion of the sampling centres in the lower stages. Such a scheme is then unbalanced, which makes estimating the components more complex (see Gower 1962). Webster et al. (2006) have shown that computing the hierarchical analysis of variance by residual maxi-

mum likelihood (REML), which is a general method of model fitting, is preferable for unbalanced nested surveys because it gives a unique result, whereas there are several methods for finding the components by the analysis of variance (ANOVA). In general, the unbalanced approach would be preferable in PA because more stages can be used with no more sampling effort.

Components of Variance by Residual Maximum Likelihood

For balanced designs ANOVA and REML give the same results, but for unbalanced ones they do not in general (Pettitt and McBratney (1993). If the random effects are assumed to be normally distributed, maximum likelihood estimates of the variance components can be calculated from Eq. 9.6. Patterson and Thompson (1971) developed the method of residual maximum likelihood (REML) to adjust for the fixed degrees of freedom before estimating the variance components. In the context here there is only one fixed effect, the grand mean, μ; therefore the differences between the estimates from REML and ANOVA are small. Readers are referred to Webster et al. (2006) for a full description of the method.

Case Study

Oliver and Webster (1987) used an unbalanced nested sampling strategy to determine the spatial variation in clay content of the soil in part of the Wyre Forest (1440 ha) of central England. Their scheme had five stages covering sampling intervals from 6 to 600 m increasing in a geometrical progression of approximately threefold increments. The survey had nine main centres on a 3×3 square grid with a spacing of 600 m; the grid was orientated randomly over the region. All other sampling positions were located from these grid nodes on random orientations as follows (see Fig. 2.2a for the plan at one main centre). From each main centre, a second site was chosen 190 m away to provide the second stage. From each of the now 18 points another site was chosen 60 m away (stage 3) and the procedure was repeated at stage 4 to locate points 19 m away. At stage 5 only half of the stage 4 sites were replicated by sampling 6 m away. The result was a sample size of 108 rather than 144 for a balanced design. Since this survey Oliver and Badr (1995) applied a nested design to a survey of soil radon concentration and Webster and Boag (1992) to a survey of nematodes; both had seven stages with only 108 sampling sites.

We estimated the variance components by REML (Webster et al. 2006). Table 2.1 gives the accumulated components of variance and also the percentage variance explained by each stage. Figure 2.2b shows the accumulated components of variance for clay content at each of four sampling depths plotted against distance on a logarithmic scale to provide a first approximation to the variogram.

Figure 2.2 shows that about 80% of the variance occurs within 60 m, and that stages 1 and 2, i.e. distances of 190–600 and 60–190 m, respectively, account for less than 20%. For these variables and many others at this site, the components

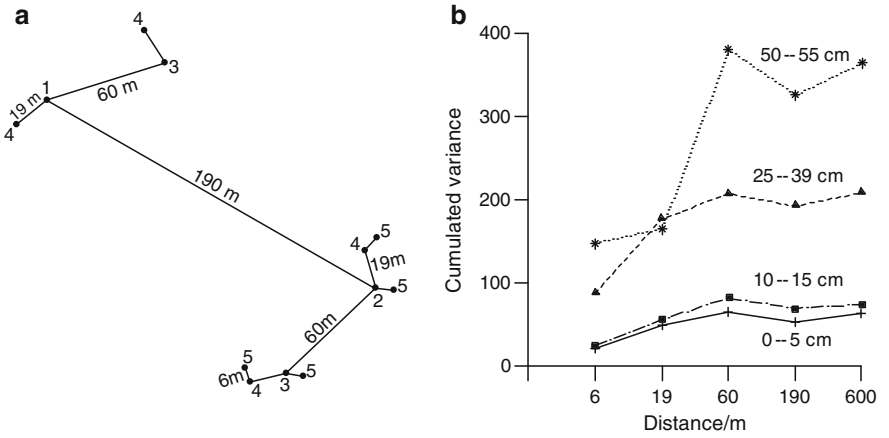

Fig. 2.2 Nested sampling scheme in the Wyre Forest, England: (**a**) sampling plan for one of the main centres and (**b**) accumulated components of variance estimated by REML for clay content in the soil at four depths in the profile, with the lag distance on a logarithmic scale

Table 2.1 Accumulated components of variance estimated by REML for clay content in the soil of the Wyre Forest at four depths in the profile, and the percentage variance accounted for by each stage

	Accumulated components of variance estimated by REML (percentage variance explained at each stage)			
Stage	Depth 0–5 cm	Depth 10–15 cm	Depth 25–30 cm	Depth 50–55 cm
1 (600 m)	63.82 (17.02)	74.73 (6.25)	209.3 (7.72)	365.1 (10.57)
2 (190 m)	52.96 (−20.01)	70.06 (−16.82)	193.1 (−7.18)	326.5 (−14.93)
3 (60 m)	65.73 (23.38)	82.63 (34.59)	208.1 (14.56)	381.0 (58.97)
4 (19 m)	50.81 (46.44)	56.78 (41.69)	177.6 (42.63)	165.7 (4.60)
5 (6 m)	21.17 (33.17)	25.92 (34.69)	88.41 (42.24)	148.9 (40.78)

for stage 2 are negative. These suggest that either there is some repetition in the variation of the property at that distance or there is no contribution to the variance at this stage. At the lowest stage, there is a considerable contribution to the variance, especially for clay at depths of 25–30 and 50–55 cm. This represents the unresolved variation within 6 m plus any errors of measurement. This is similar to the nugget variance in geostatistics (see Chapter 1). From this information we could design a survey to estimate the variogram more precisely by sampling along transects or we could design an overall survey with a maximum sampling interval of less than half the correlation range of 60 m identified (see Oliver and Webster (1987) for a full account of these results). Although this example is not agricultural, the principles are the same and the study area here would be equivalent to doing a nested survey of a whole farm. A case study described in Chapter 9 of this book shows an application of nested sampling and analysis to determine a suitable grid sampling interval to estimate the pattern of variation in the cereal cyst nematode, *Heterodera avenae*.

2.2.2 Variograms from Ancillary Data

Many of the more permanent soil properties, such as soil texture, appear to co-vary spatially with ancillary data, which are usually intensive and cheaper to obtain. Ancillary data include digitized aerial photographs, electrical conductivity (EC_a), yield, remotely and proximally sensed data, and digital elevation models. Each type of ancillary data is likely to be more strongly correlated with some soil properties than with others. In other words they are coregionalized which suggests that we can use these data to indicate the approximate spatial scale of variation in more expensive variables. Variograms computed from intensive ancillary data can then be used to guide sampling of the variables of interest. Chapter 7 describes how inexpensive coregionalized secondary information can be used to improve the accuracy of predictions of the primary variable by cokriging and other multivariate geostatistical methods. These approaches are likely to be of value to precision farmers because they often have suitable ancillary data or they can obtain it with little additional expense.

Experimental variograms computed from ancillary data and modelled can be used to guide sampling with an often used 'rule of thumb' of sampling at less than half of the variogram range (see Section 2.4.2). This use of ancillary data should avoid over- or under-sampling, both of which will result in wasted effort. The case study described below shows that the variogram ranges of the more permanent soil properties and ancillary data are reasonably consistent.

2.2.2.1 Case Study

Data from a field at Wallingford with soil developed on the plateau gravels of the Thames valley near to Oxford, England illustrate the above approach. The site was sampled and observations were made in the winter of 2000 on a 30-m grid. Six samples of the topsoil (0–15 cm) were taken from a 1 m^2 support and bulked at each grid node. Several soil properties of the air-dried <2 mm soil fraction were measured using standard methods. The property of interest here is loss on ignition (LOI). Ancillary data were obtained as follows. Aerial photographs of the bare soil from aerial surveys were digitized to give a ground pixel size of 3.5×3.5 m. Apparent soil electrical conductivity (EC_a) was recorded with a Geonics EM38 instrument in the vertical position about every 1 m along transects about 20-m apart. Elevation data were obtained from the z coordinate of differential global positioning system (DGPS) surveys of the fields, and yield data from several years were obtained with the Massey Ferguson Fieldstar system (www.masseyferguson.com).

Omni-directional variograms were computed for each of the soil and ancillary variables at the site. Multivariate variograms (Bourgault and Marcotte 1991) were computed from the aerial photograph data by Eq. 2.2. The multivariate variogram is defined by:

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}[\{\mathbf{Z}(\mathbf{x}) - \mathbf{Z}(\mathbf{x}+\mathbf{h})\}^{\mathrm{T}}\mathbf{M}\{\mathbf{Z}(\mathbf{x}) - \mathbf{Z}(\mathbf{x}+\mathbf{h}\}], \tag{2.1}$$

where E is the expectation, $\mathbf{Z}(\mathbf{x})$ and $\mathbf{Z}(\mathbf{x}+\mathbf{h})$ are vectors of random variables at positions $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ separated by the lag, $\mathbf{h}$, T is the transpose, and $\mathbf{M}$ is a $p \times p$ positive-definite symmetric matrix defining the relations between the variables. The experimental variogram was calculated by the standard formula adapted for the multivariate case:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \{\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_i + \mathbf{h})\}^{\mathrm{T}} \mathbf{M} \{\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_i + \mathbf{h})\}, \qquad (2.2)$$

where $\mathbf{z}(\mathbf{x}_i)$ and $\mathbf{z}(\mathbf{x}_i + \mathbf{h})$ are the vectors of observations at $\mathbf{x}_i$ and $\mathbf{x}_i + \mathbf{h}$.

Table 2.2 gives the parameters of the models fitted to each experimental variogram. All soil properties and most ancillary variables have variogram ranges of similar magnitudes, except for yield. Yield shows more complex variation with two scales of variation for most years. The average variogram range for the soil properties is 238 m and for the ancillary data excluding yield it is 221 m. The difference in the variogram ranges for yield can be explained by the fact that several factors other than the soil, such as disease, weeds, pests, management practices, weather, etc., affect the yield in a given year. At Wallingford the physiography is fairly complex, and the soil properties appear to vary in harmony with it. Our analyses at other sites showed that in general aerial photograph and EC_a data had variogram ranges that were more similar to those of the soil properties than elevation and yield data. Therefore, we do not recommend yield and elevation data to guide sampling unless there is evidence that these data are consistently related to patterns of soil variation in the field of interest.

Figure 2.3a, c shows the aerial photograph and EC_a data for Wallingford, respectively. The patterns of variation for both sets of data are similar. The paler areas have smaller conductivities and vice versa; they indicate places where the soil is particularly gravelly, or where the gravel is largely calcareous. The experimental variograms and the fitted models for the aerial photograph and EC_a data (Fig. 2.3) have similar ranges. There are also some similarities in the patterns observed in the aerial photograph, EC_a and the soil data (compare Figs. 2.3a, c and 2.4a).

Based on the variogram results in Table 2.2 and sampling at an interval of less than half the range of a variogram from appropriate ancillary data, grid intervals of 100–120-m should suffice at the Wallingford site. Data on the 30-m grid were sub-sampled to give a range of coarser sampling intervals: 60-m (70 sites), 90-m (36 sites), 120-m (23 sites), 150-m (14 sites) and 120 + 60-m (50 sites). The sub-sampled data were then used with the model parameters of the variogram computed from the 30-m data for kriging. The variogram for the 30-m data was used because it was computed from >100 data, whereas the sub-samples would provide too few data from which to compute accurate variograms. Figure 2.4 shows the kriged maps of loss on ignition (LOI) for the full and sub-sampled data. The main features of the variation are preserved for sampling intervals of 120-m and less, although there is some loss of detail for all sub-samples. The main features of the variation are no longer evident when the sampling interval is 150-m (Fig. 2.4e). The results suggest

Table 2.2 Variogram model parameters for soil and ancillary variables at Wallingford, Oxfordshire

Variable	Model	c_0	c_1	c_2	α	a_1 (m)	r_1 (m) ($3r_1$ (m))	a_2 (m)	r_2 (m) ($3r_2$ (m))
Clay	Spherical	0.04556	28.87			134.8			
Depth	Exponential	310.00	271.30				77.46 (232.4)		
LOI	Pentaspherical	0.0535	0.4610			226.1			
MCF	Circular	0.03091	0.3151			185.7			
Munsell value	Pentaspherical	0.1273	0.3967			218.9			
Sand	Pentaspherical	5.894	81.39			231.2			
Stoniness	Exponential	11.73	205.8				89.05 (267.2)		
VWC	Pentaspherical	1.891	4.770			202.2			
Aerial[66]	Stable exponential	0	538.3		1.459		68.20 (204.6)		
Aerial[97]	Exponential	1.861	1.180				82.57 (247.7)		
EC_a	Pentaspherical	3.525	43.19			201.4			
Elevation	Circular	0.2309	10.03			228.5			
Yield[96]	Double spherical	0.6027	0.4363	2.281		62.62	56.60 (169.8)	650.0	
Yield[97]	Exponential	0.2082	0.3993						
Yield[98]	Double spherical	0	0.216	0.074		22.63			
Yield[99]	Double spherical	0	0.7842	0.361		14.31		316.0	
Yield[00]	Double exponential	0	0.9201	0.776		11.37		89.40	31.13 (93.39)

LOI is loss on ignition; MCF is moisture correction factor and VWC is volumetric water content.

The parameters are: c_0, the nugget variance; c, the sill of the autocorrelated variance; a, the range of the spatial dependence; for the exponential and stable exponential functions, r is the distance parameter of the model and because the sill is asymptotic an approximate range is determined as $3r$ (value in parentheses).

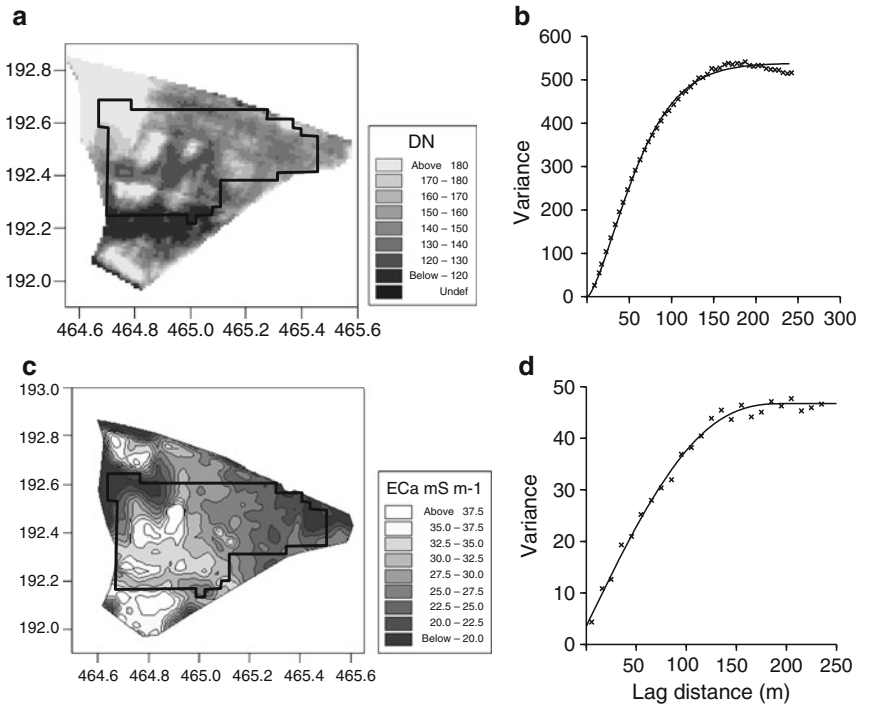

Fig. 2.3 Wallingford field site: (**a**) aerial photograph, (**b**) multivariate variogram of the Red, Green and Blue wavebands from the aerial photograph, (**c**) map of EC_a, the bold line within the field is the limit of the sampled area and (**d**) variogram of EC_a

that a sampling interval of 100–120-m would be adequate for precise management in PA, which corresponds with less than half the range of all ancillary variables apart from yield (Table 2.2). Figure 2.4f suggests that additional sampling at half the grid interval for randomly selected nodes has increased the observed detail. It would also improve the accuracy of the variogram near to the origin.

As the pattern of variation is generally unknown at the sampling stage the importance of any given sample location cannot be known. Taking additional samples at shorter intervals reduces the potential loss of information and reduces the nugget component. Additional samples can be targeted according to the variation in a key ancillary variable to avoid missing important features in the soil variation. The further samples are from one another, the larger the nugget is likely to be due to unresolved variation at distances less than the sampling interval. Accurate kriged predictions depend on accurate estimation of the nugget variance. The nugget variance can also be reduced by bulking soil or plant samples to reduce the effects of variation at distances much less than the sampling interval.

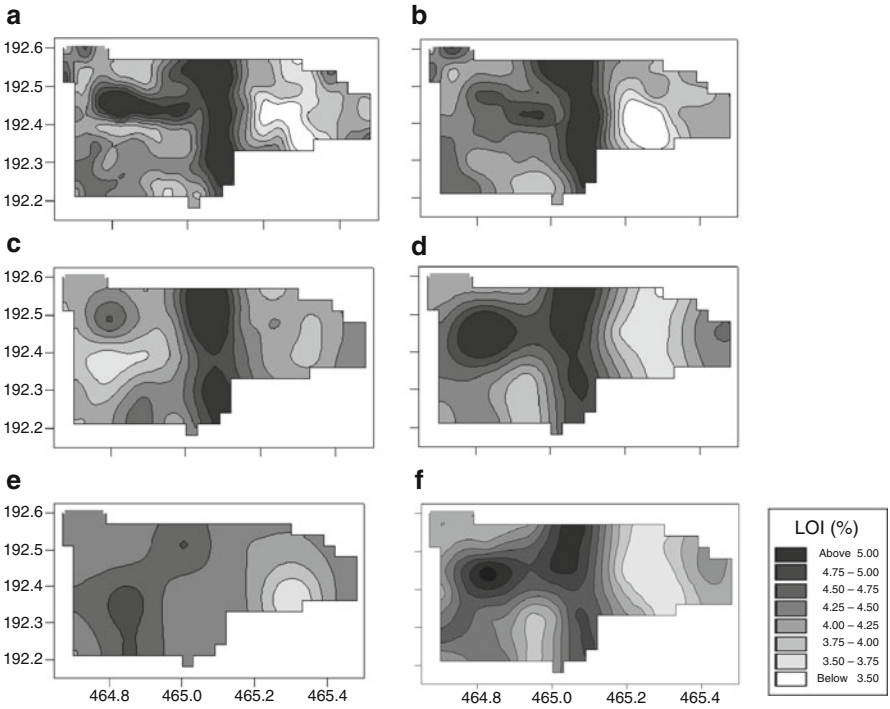

Fig. 2.4 Maps of kriged predictions of loss on ignition (LOI) at Wallingford based on the (**a**) 30-m data and 30-m model, (**b**) 60-m data and 30-m model, (**c**) 90-m data and 30-m model, (**d**) 120-m data and 30-m model, (**e**) 150-m data and 30-m model and (**f**) 120-m +60-m data and 30-m model

2.3 Use of the Variogram to Guide Sampling for Bulking

Soil and crop attributes can vary considerably over short distances. This local fluctuation in values between points might mask the variation over the tens or hundreds of metres that is of greatest interest to the precision farmer. The material taken or the observations made at a sampling point are intended to represent the property of interest reliably over the surrounding area, the size of which will depend on the degree of spatial variation. If there is sufficient information from several cores of soil or several plants then this short-range fluctuation can be smoothed by local averages. However, obtaining so many individual measurements would be too costly in practice. The alternative is to bulk several cores of soil from a small area (Oliver et al. 1997) and mix them thoroughly for analysis or bulk material from several plants (Willers et al. 2009). The concentration measured in a bulked sample should equal the arithmetic mean of the individual cores or plants contributing to it, unless some chemical reaction takes place within the sample.

The number of soil cores or plants to be bulked depends on the local variation and the error that can be tolerated. Burgess and Webster (1984) described how to

determine the number of cores or plants for bulking if the variogram of the property is known. The variogram model is used to determine the variance for various combinations of support and configuration of sampling points within the blocks. The estimation variances (or errors) are calculated for a range of sample sizes and configurations, and the combination that meets the tolerance can be determined.

For a small area B, we can estimate the average of property Z, $\hat{\mu}(B)$, in B from n known values, $z(\mathbf{x}_i)$, at positions $\mathbf{x}_i$, $i = 1, 2, \ldots, n$ within B by

$$\hat{\mu}(B) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i), \tag{2.3}$$

where λ_i are the weights associated with positions $\mathbf{x}_i$ (Burgess and Webster 1984).

The estimation variance of $\hat{\mu}(B)$ is

$$\sigma^2(B) = 2 \sum_{i=1}^{n} \lambda_i \overline{\gamma}(\mathbf{x}_i, B) - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j) - \overline{\gamma}(B, B), \tag{2.4}$$

where $\gamma(\mathbf{x}_i, \mathbf{x}_j)$ is the semivariance between points $\mathbf{x}_i$ and $\mathbf{x}_j$, $\bar{\gamma}(\mathbf{x}_i, B)$ is the average semivariance between data point $\mathbf{x}_i$ and the block B, and $\bar{\gamma}(B, B)$ is the within-block variance.

In principle, bulking is equivalent to averaging the values at the positions from which the individual samples of soil or plants are taken in B. Every sample is the same size and shape, so the weights, λ_i, are equal. The number and positions of the cores can be varied to achieve a particular precision expressed in terms of the estimation variance. For practical purposes the estimation variance is minimized for a given n when the sampling points are on a centred regular grid (Burgess and Webster 1984). With the parameters of an existing variogram, Eq. 2.4 can be solved for a range of n on a square grid. To determine the number of cores required for a bulked sample, the calculated estimation variance or error is plotted against n. The smallest value of n can be determined from the graph to satisfy a predefined tolerable error. Tolerable errors can be determined with Eq. 2.5 as in Section 2.3.1 if none is known beforehand.

2.3.1 Case Study

An arable field on a chalk plateau on the Yattendon Estate, Berkshire, England, was sampled (Oliver et al. 1997). The sampling scheme comprised six nodes 100-m apart. At each node there were two orthogonal transects, the mid-points of which coincided with the node. The transects were 14 m long, and the soil was sampled at 1-m intervals to a depth of 15 cm. Twenty nine samples were taken at each grid node, giving 174 samples in total. The samples were air-dried and sieved, and the

Table 2.3 Variogram model parameters for the soil survey on the Yattendon Estate, Berkshire, England

		Variogram model parameters				
Property	Model	c_0	c	a (m)	w	α
K	Spherical	159.10	298.00	3.09		
P	Power	9.82			1.228	1.19
Mg	Circular	5.24	6.43	4.38		

The parameters are: c_0, the nugget variance; c, the sill of the autocorrelated variance; a, the range of the spatial dependence; w and α are the intensity and exponent, respectively, of the power function. The equation for the spherical function is given in Section 1.3.2, Eq. 1.27. The equations for the power and circular functions are: $\gamma(h) = c_0 + wh^{\alpha}$ and $\gamma(h) = c_0 + c\left\{1 - 2/\pi \cos^{-1}(h/a) + 2h/\pi a\sqrt{1 - h^2/a^2}\right\}$, respectively.

available potassium (K), phosphorus (P) and magnesium (Mg) were determined by the standard methods of MAFF (1986). Experimental variograms were computed for the variables by the usual method of moments estimator, and models were fitted in GenStat (Payne 2008).

The variogram model parameters for each property in Table 2.3 were used with Eq. 2.4 to determine the estimation errors for three block sizes (of side 2-, 5- and 10-m) and sample sizes ($n = 4, 9, 16, 25, 36, 49$). The errors plotted against n for each block size are shown in Fig. 2.5. All the blocks were square, and the results are for square sampling configurations. There are no intermediate values of n with these configurations; the lines are drawn to guide the eye and aid interpretation. In all cases the estimation errors decrease roughly in inverse proportion to the sample size, and as there is more variation in the larger blocks their errors are always somewhat larger than those of the smaller blocks.

To determine the optimal number of cores for bulking, an acceptable margin of error is needed. If this value is known, or has been determined by Eq. 2.5, it can be represented as a horizontal line drawn across the graph at the specified tolerance (Fig. 2.5). The intersections of the line with the graphs indicate the number of cores from which to bulk. Since the block is square the next largest square number from the intersection is used.

Recommended tolerance values were used for this study. For K, the value was $7\,\mathrm{mg\,l^{-1}}$ and the line intersects the graph at $n \approx 4$ for the 2×2 m block. Bulking from this number of cores in a centrally placed square configuration will ensure that the tolerable error is not exceeded. For the 5×5 m block the line intercepts the graph at $n \approx 6$ and the next largest square number, nine, would be used. The 10×10 m block would also require a bulked sample of nine cores. For P the tolerance was $1\,\mathrm{mg\,l^{-1}}$, and Fig. 2.5 shows that a sample of sixteen cores is needed for all block sizes. A tolerance of $3\,\mathrm{mg\,l^{-1}}$ was used for Mg (the line for this is not shown because it is much larger than the largest error), and Fig. 2.5 shows that a bulked sample of four cores would be sufficient.

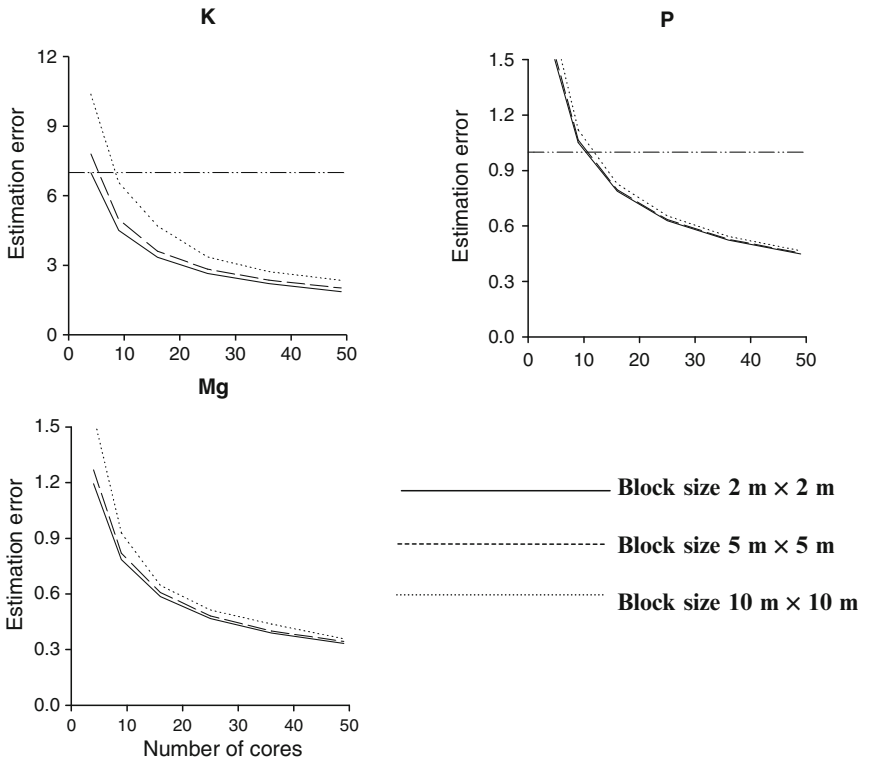

Fig. 2.5 Graphs of estimation error against size of sample to bulk from for P, K and Mg for a field on the Yattendon Estate, Berkshire, England

If there are no recommended values for the tolerance then we suggest that the departure from the mean for a 95% level of confidence is calculated by

$$L = \frac{1.96\sigma}{\sqrt{n}}, \tag{2.5}$$

where L is the tolerance, σ is the standard deviation of the variate and n is the sample number (Webster and Oliver 1990).

Different nutrients require different sampling schemes; therefore, the number of samples should be based on the nutrient that has most effect on the crop to be grown. In many instances, phosphorus may be limiting and, therefore, a bulked sample of 16 should be used. Although this number results in over sampling for potassium and magnesium, it is clear from Fig. 2.5 that a sample of 16 reduces the variance considerably and will improve the estimates of all these nutrients.

The results from other fields with different soil types showed that the sample size needed for bulking also varies with soil type. On more variable soil, such as the alluvium that Oliver et al. (1997) examined, the optimal number of cores was

much larger. In the absence of prior knowledge, and no variogram from which to compute the estimation variances, a reasonable 'rule of thumb' would be to bulk from 16 cores.

2.4 The Variogram to Guide Grid-Based Sampling

Sampling on a grid is favoured in geostatistics and PA because it provides an even cover of values and minimizes the maximum estimation variance (or error) for a given grid interval. It is also easy to implement in the field. However, practitioners must still decide how many sample sites are needed and what the interval between them should be. The variogram can provide the answer to the latter. Two approaches for identifying a suitable interval for grid-based soil sampling are considered here and are illustrated with case studies.

2.4.1 The Variogram and Kriging Equations

McBratney et al. (1981) and McBratney and Webster (1981) showed how the variogram and kriging equations could be used to determine an optimal sampling interval for prediction by kriging before obtaining new data from a survey. The basis of their approach is that the kriging weights, and also the kriging variances (see Chapter 1) depend on the configuration of the sampling points in relation to the target point or block and on the variogram. They do not depend on the observed values at these points. If we have a variogram function from a previous survey of a field then we can determine the kriging errors for any grid size before sampling. It is possible to optimize the sampling by designing a scheme to meet a specified tolerance or precision. Although a triangular grid is the most efficient, square or rectangular grids are preferred because they are easier to implement, and there is little difference in precision between them in practice.

Sampling can be optimized for punctual or block kriging (see Eqs. 1.8–1.16). Webster and Oliver (2007) provide more detail on this; here we consider only block kriging. Precision farmers want to manage areas (blocks) of their fields that relate to their machinery. For block kriging practitioners must decide where to determine the kriging variances, i.e. whether for blocks centred on grid cells or ones centred on grid nodes. This is because the position at which the kriging variance is greatest varies according to block size, and it is the largest variances that should be used.

2.4.1.1 Case Study

The topsoil (0–15 cm) of the same field on the Yattendon Estate in Section 2.3.1 (Frogbrook 2000) was sampled in December 1997 and 1998, and available K,

Table 2.4 Variogram model parameters for K, Mg and P for a field on the Yattendon Estate

Variable	Model	c_0	c	r	a(m) (3r (m))
K^{97}	Spherical	411.1	450.1		171.4
K^{98}	Spherical	648.5	299.1		150.0
Mg^{97}	Exponential	0	55.49	22.02	(66.06)
Mg^{98}	Spherical	12.01	21.13		50.68
P^{97}	Spherical	49.26	132.0		206.6
P^{98}	Pentaspherical	28.41	72.84		182.8

The model parameters are as given above. For the exponential function, the sill is asymptotic and an approximate range is determined as $3r$, where r is the distance parameter of the model.

Mg and P were determined. There was no anisotropy in the variation, and omnidirectional experimental variograms were computed from the measurements on a bulked sample of 10 cores of soil from a support of 2×5 m taken at the nodes of a 20-m grid. The experimental values were modelled, and the model parameters are given in Table 2.4. The parameter values are similar for both years for each nutrient, which suggests that a tolerance can be fixed in the absence of any marked temporal variation.

The model parameters for K, Mg and P (Table 2.4) were used with the kriging equations to provide estimates of the kriging standard errors associated with sampling intervals on a square grid between 2- and 200-m and blocks of side 20-, 40- and 60-m. These block sizes are ones that farmers might use to manage inputs and are associated with the size of areas over which farm machinery operates. Figure 2.6 shows that for each nutrient and block size, as the grid spacing increases, the kriging error increases until it reaches a maximum. For all examples in Fig. 2.6, when the sampling interval is very short the maximum estimation variances decrease from some small value and then increase again. This is an artefact that arises because only observations nearest the centres of the blocks have been used for estimation.

Tolerable kriging errors of 2, 5 and 10 mg kg^{-1} for P, Mg and K, respectively, were used to determine sampling intervals for these nutrients. These concentrations were considered to be limiting to cereal yield for each nutrient. Table 2.5 summarizes the results for the various block sizes and tolerance values used in Fig. 2.6. The sampling intervals indicated for K and P are consistent between years, they are greater for K than for P and they increase as block size increases. Thus a variogram from a survey done in a previous year could be used to improve sampling at a later stage. For Mg, the largest interval is for the 40-m block in 1997; the tolerance exceeds the estimation error for the 60-m block in 1997 and for all blocks for 1998. The variation of Mg in this field is small compared to the tolerable error and a large sampling interval would achieve the specified error.

It is possible to take anisotropy into account by adjusting the grid spacing so that sampling is more intense in the direction of maximum variation (see Webster and Oliver 2007, for more detail). The intervals can be determined as above with the model parameters of an anisotropic model.

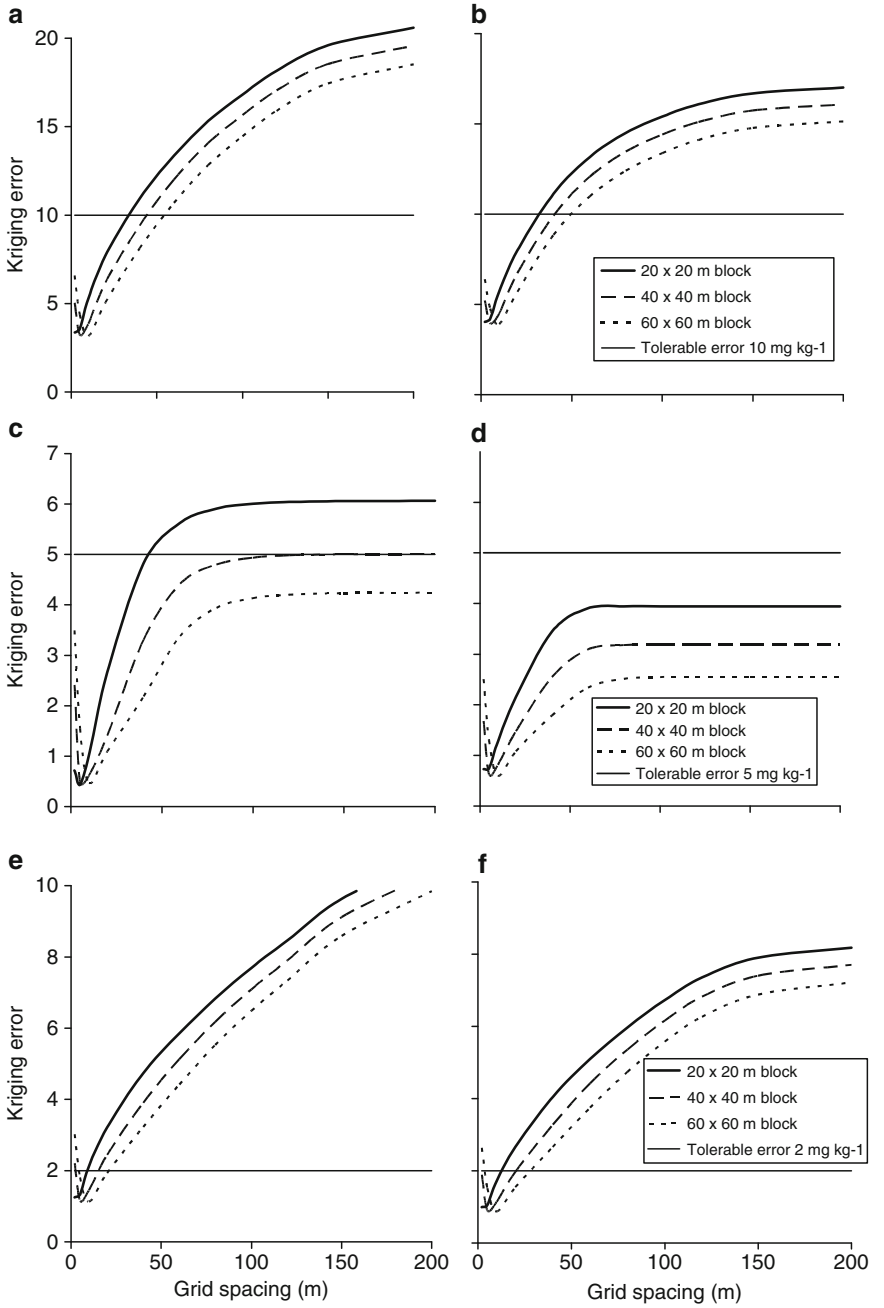

Fig. 2.6 Graphs of kriging error against grid spacing for: (**a**) K^{97}, (**b**) K^{98}, (**c**) Mg^{97}, (**d**) Mg^{98}, (**e**) P^{97} and (**f**) P^{98} on the Yattendon Estate, Berkshire

Table 2.5 Optimal grid spacing (m) for three different block sizes at the Yattendon Estate

Property and tolerance ($mg\,kg^{-1}$)	20-m block		40-m block		60-m block	
	1997	1998	1997	1998	1997	1998
K ± 10	35	32	45	42	55	52
Mg ± 5	43	*	130	*	*	*
P ± 2	10	12	15	20	22	28

*Tolerance exceeds the estimation error or the variation in the field is smaller than the tolerable error so a large sampling interval would suffice.

2.4.2 *Half the Variogram Range 'Rule of Thumb' as a Guide to Sampling Interval*

The case study on ancillary data in Section 2.2.2.1 suggests that a sampling interval of less than half the range of a variogram from such data can be used to guide sampling. We tested this 'rule of thumb' on large fields simulated by the turning bands method (Journel and Huijbregts 1978). Following the approach of Webster and Oliver (1992) the fields were sub-sampled randomly 100 times for a given sampling interval based on the variogram range. The sub-samples were designed to use a datum no more than once and to cover the same area. The minimum sample size was 144 so that we could discount the effects on the accuracy of too small a sample to compute a reliable MoM variogram (Webster and Oliver 1992). Variograms were computed for each random sub-sample, and confidence limits were computed based on the 100 experimental variograms. These limits narrowed as the sampling interval decreased (Fig. 2.7). The variograms computed on data with an interval of 0.66 of the range appeared to be almost all pure nugget (Fig. 2.7a), and the confidence intervals were almost twice as wide as those for the other sampling intervals.

Standard deviations of the observed semivariances (not shown) were similar for sampling intervals of 0.5 and 0.4 of the variogram range when there was no nugget component. However, they were larger for 0.5 of the range when there was a nugget component. Nevertheless, they were still markedly less than those for an interval of 0.66 of the range. These results show that a sampling interval of 0.66 of the range is too large and will result in large errors in the predictions and probably a pure nugget variogram. An interval of 0.5 of the range would give acceptably accurate predictions for site-specific management. The results for intervals of 0.4 and 0.33 of the range, which are similar, would give more accurate predictions, especially when there is a nugget component. Therefore, sampling at an interval of 0.4 of the variogram range or less should be adopted if possible.

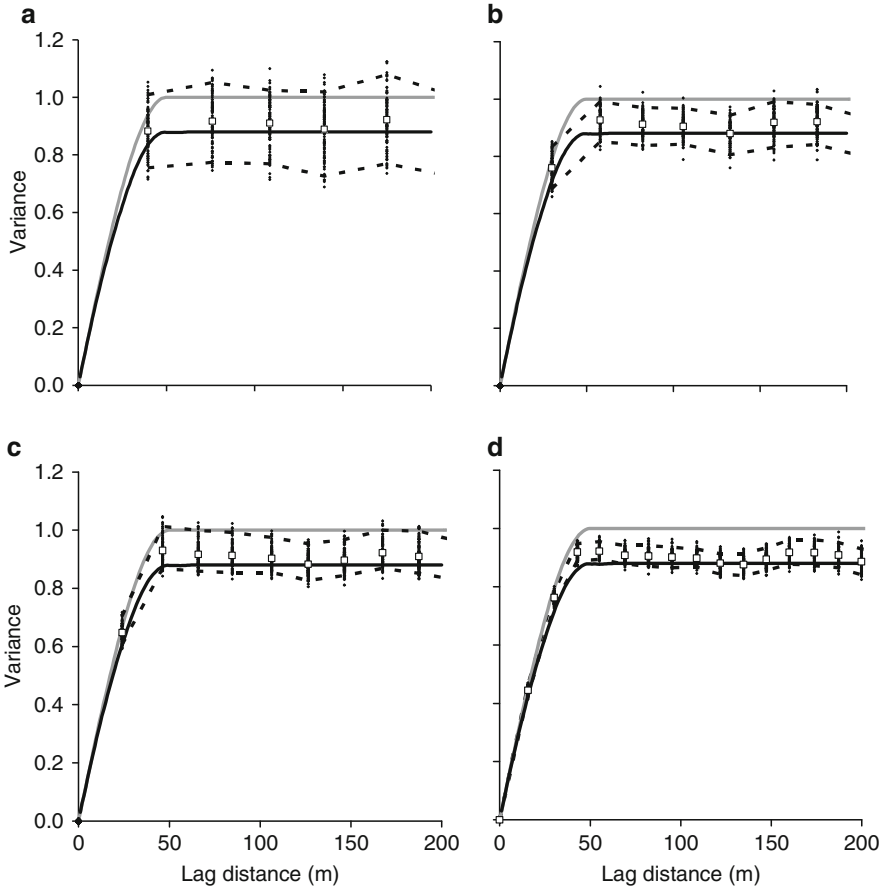

Fig. 2.7 Observed semivariances and the 90% confidence limits (*dashed lines*) for various sampling intervals for a spherical field with a range of 50 m and zero nugget: (**a**) 0.66 of range (33 m), (**b**) 0.5 of range (25 m), (**c**) 0.4 of range (20 m), and (**d**) 0.33 of range (17 m) (the *grey line* is the function used for simulation and the *black line* is the model fitted to the exhaustive variogram of the sub-sample)

2.5 Variograms to Improve Predictions from Sparse Sampling

2.5.1 Residual Maximum Likelihood (REML) Variogram Estimator

We describe above how a suitable sampling interval can be determined from variograms of ancillary data or existing variograms of relevant soil properties. However, if this interval is large compared with the extent of the field, then there will be too

few data from which to compute Matheron's (1965) method of moments (MoM) variogram reliably (Webster and Oliver 1992). Some potential solutions to this situation are illustrated below with a case study.

2.5.1.1 Case Study

For the Wallingford site (see Section 2.2.2.1), variograms from ancillary data suggested that an interval of 120-m would resolve the spatial variation in this 43.5 ha field. This sampling interval results in only 23 sampling points, which is too few data from which to compute a reliable MoM variogram. Pardo-Igúzquiza (1998a) suggested that a reliable variogram could be computed from a 'few dozen' data by maximum likelihood or residual maximum likelihood (REML). Lark (2000) and Kerry and Oliver (2007) examined this idea further and suggested that 50–60 data might suffice. Based on a simulation study, Kerry et al. (2008) showed that the accuracy of a variogram estimated from 100 data by MoM was similar to one estimated from 50 data by REML. Nevertheless, the variogram estimated from 100 data by REML was more reliable.

Figure 2.8 shows the variograms computed by MoM and REML from data with sampling intervals between 30- and 120-m at Wallingford, and Table 2.6 gives the model parameters. All variograms have been standardized to a sill of 1 so that the parameters can be compared with the approach in Section 2.5.2. The MoM variogram for the 90-m grid is unbounded, and that for the 120-m grid appears as pure nugget, whereas the variograms estimated by REML for these intervals are similar to that for the 30-m grid. Table 2.6 shows that when data on the 120-m grid are supplemented with additional points at 60-m to give 50 data, the model parameters for the MoM and REML estimators are closer to those for the 30-m data. Although a plausible model was fitted to the experimental MoM variogram computed from these data, Fig. 2.8e shows that the semivariances are erratic and the model is a poor fit.

Table 2.7 gives the results of cross-validation for the standardized LOI data. Data on the 30-m grid were used with the model parameters of variograms estimated from the subsets to determine how appropriate these models were for representing the variation in LOI. For the MoM variogram there is a marked increase in mean squared error (MSE), Eq. 1.29, as the sampling interval increases and the mean squared deviation ratio (MSDR), Eq. 1.30, is markedly less than 1. For the variogram estimated by REML the increase in MSE with increasing sampling interval is less marked. The MSDRs for some sub-samples and the REML estimator are around 2, which show that the kriging variance is underestimated by the relevant model. The MSEs, however, indicate that sample size affects variograms estimated by REML less and as a consequence there is less loss of accuracy in the predictions. This result has important implications for PA because farmers often cannot afford to sample intensively. Pardo-Igúzquiza (1998b) provides a published program to compute the variogram by REML.

Loss on ignition was kriged using data from the various sub-samples and their associated MoM and REML variograms, however, for pure nugget variograms

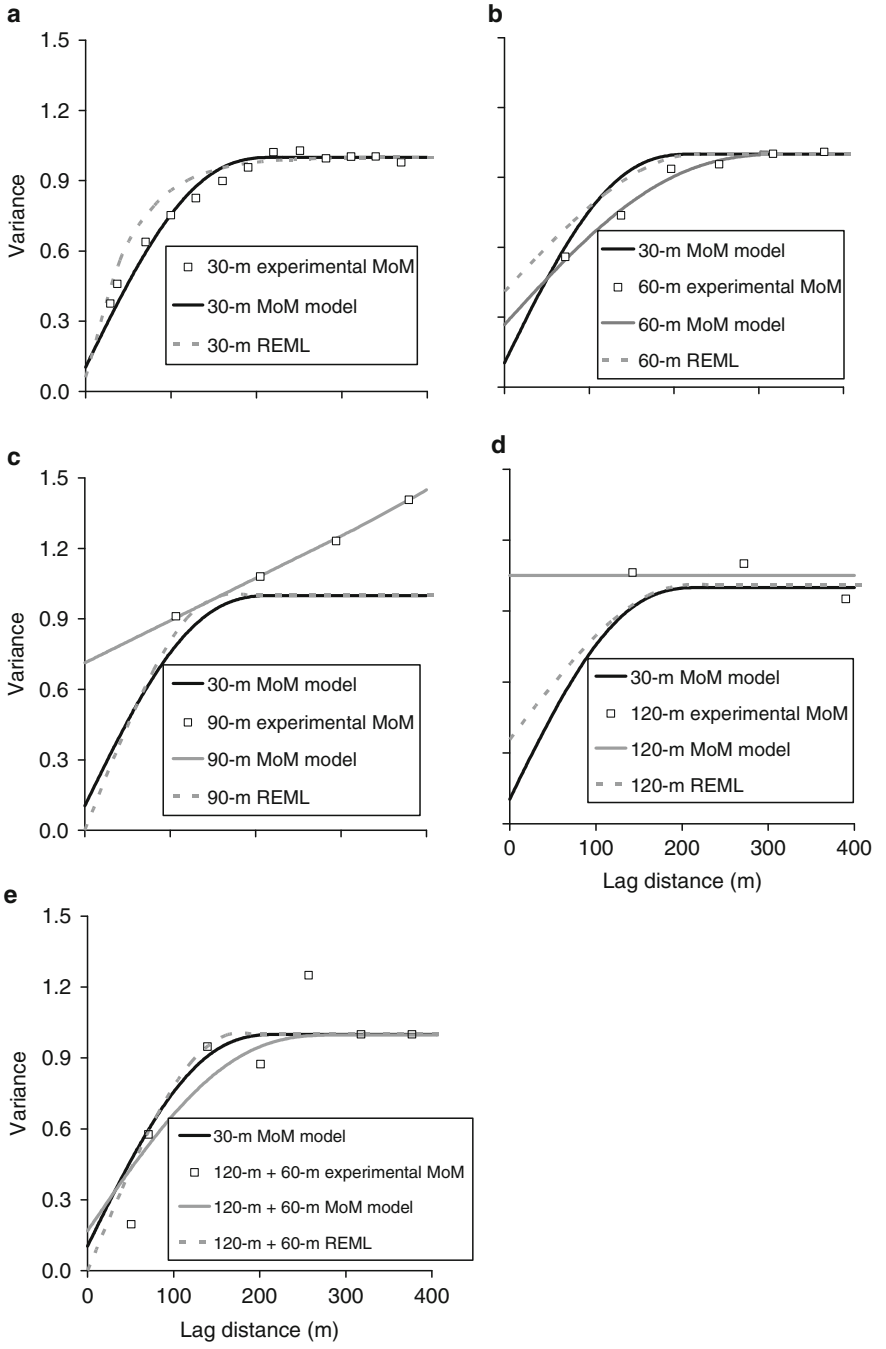

Fig. 2.8 Variograms for percentage loss on ignition (LOI) at Wallingford estimated by the method of moments (MoM) and residual maximum likelihood (REML) for: (**a**) 30-m grid (296 data), (**b**) 60-m grid (70 data), (**c**) 90-m grid (36 data), (**d**) 120-m grid (23 data) and (**e**) 120-m + 60-m grid (50 data)

Table 2.6 Variogram model parameters for residual maximum likelihood (REML), method of moments (MoM) and standardized (Std) variograms of percentage loss on ignition (LOI) at Wallingford

Grid interval (number of data)	Method	Model	c_0	c_1	w	a (m)	r (m), (3r (m))
30-m (296)	REML	Exp.	0.04	0.96			50.07, (159.2)
60-m (70)	REML	Spher.	0.41	0.59		227.8	
90-m (36)	REML	Spher.	0	1		163.1	
120-m (23)	REML	Spher.	0.36	0.64		205.9	
120-m + 60-m (50)	REML	Spher.	0	1		170.6	
120-m + HML Std (38)	REML	Spher.	0	1		206.8	
30-m (296)	MoM	Penta.	0.10	0.89		226.1	
60-m (70)	MoM	Penta.	0.27	0.73		344.4	
90-m (36)	MoM	Linear	0.71		0.002		
120-m (23)	MoM	Nugg.					
120-m + 60-m (50)	MoM	Penta.	0.17	0.83		291.3	
120-m + LMS (38)	Std.	Exp.	0	1			68.2, (204.6)

The models are: Exp. exponential, Spher. spherical, Penta. pentaspherical and Nugg. pure nugget. The model parameters are: c_0 the nugget variance; c the sill of the autocorrelated variance; a the range of the spatial dependence and w the intensity of the linear function. For the exponential function, the sill is asymptotic and an approximate range is determined as 3r (values in parentheses), where r is the distance parameter of the model, and LMS is large, medium or small digital numbers (DNs).

Table 2.7 Cross-validation results for Wallingford for loss on ignition (LOI) data with model parameters estimated by residual maximum likelihood (REML), the method of moments (MoM) and standardized (Std) variograms

Variogram estimator	Number of data	MSE	MSDR
30-m REML	296	0.471	0.970
60-m REML	70	0.505	0.887
90-m REML	36	0.478	2.091
120-m REML	23	0.497	0.929
120-m + 60-m REML	50	0.477	2.188
120-m + LMS REML	38	0.477	2.657
30-m MoM	296	0.477	1.504
60-m MoM	70	0.495	1.191
90-m MoM	36	0.543	0.661
120-m MoM	23	0.574	0.521
120-m + 60-m MoM	50	0.482	1.444
120 m + LMS Std	38	0.471	1.367

bilinear interpolation was used; Figs. 2.9 and 2.10, respectively, show the mapped predictions. These maps contrast with those in Fig. 2.4 where the model for the 30-m data was used to krige the sub-sampled data. Figure 2.9 shows that there is considerable loss of detail in the maps based on the sub-sampled data and that this is much greater than in Fig. 2.4. The variation in Fig. 2.9e is the least degraded, which

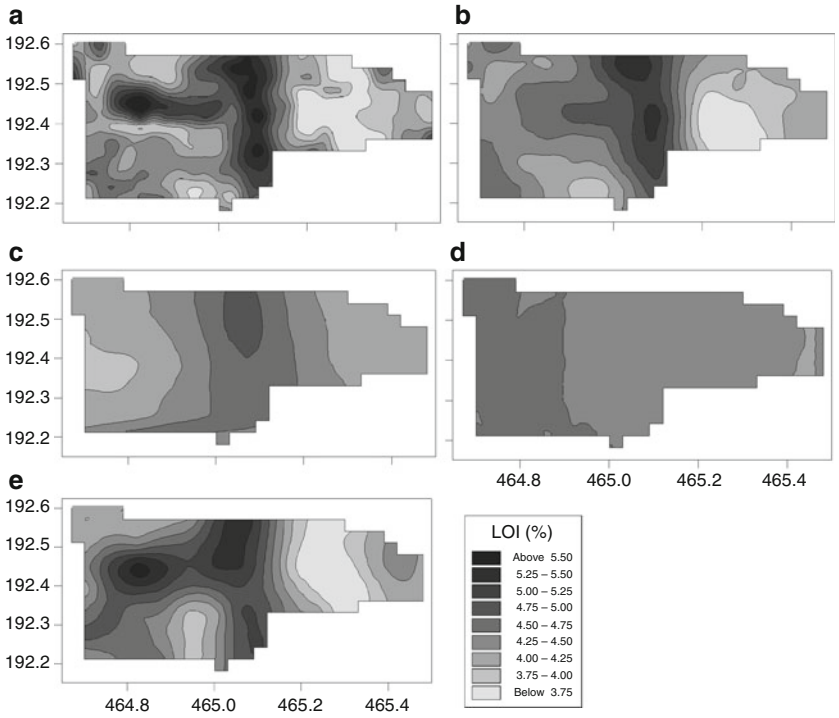

Fig. 2.9 Kriged maps of LOI for Wallingford for sub-samples with grid spacings of: (**a**) 30-m, (**b**) 60-m, (**c**) 90-m, (**d**) 120-m and (**e**) 120-m + 60-m data and associated MoM variograms

indicates the advantage of the additional sampling at half the grid interval (Fig. 2.4f) when sampling is sparse. Figure 2.10 shows that the patterns of variation are far less degraded for the small sample sizes when the variogram estimated by REML is used.

2.5.2 Standardized Variograms

As an alternative to estimating the variogram by REML, Kerry and Oliver (2008) suggested using standardized variograms from ancillary data with the sill scaled to 1 and nugget:sill ratios that were appropriate for different soil properties. These authors found that the most reliable method of determining the nugget:sill ratio was to compute a variogram by REML from the soil data, together with data from 15 additional sampling sites targeted to areas in an aerial photograph with large, medium and small (LMS) digital numbers.

Figure 2.11a shows samples selected from the 30-m grid based on large, medium or small (LMS) digital numbers (DNs) in Fig. 2.3a. Five samples were selected for

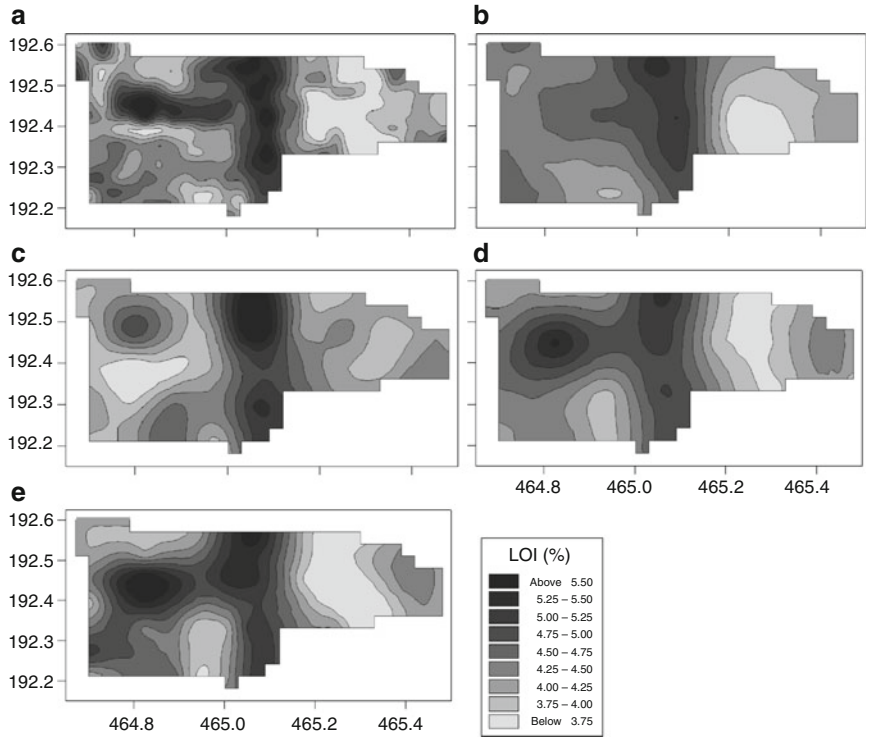

Fig. 2.10 Kriged maps of LOI for Wallingford for sub-samples with grid spacings of: (**a**) 30-m, (**b**) 60-m, (**c**) 90-m, (**d**) 120-m and (**e**) 120-m + 60-m grid data and associated variograms estimated by REML

each of the three levels of reflectance. The data from these samples were used with those from the 120-m grid (38 samples in total) to estimate the variogram by REML to determine the nugget:sill ratio for the standardized variogram. Table 2.6 gives the model parameters of this variogram and those of the standardized variogram (last line). Table 2.7 gives the associated cross-validation results for comparison with the other sub-samples. The MSE is small for the variogram estimated by REML from data on the 120-m grid and targeted LMS data, but the MSDR is large. By contrast, the MSE is slightly smaller and MSDR is markedly smaller (Table 2.7) using the range and model type of the standardized variogram determined from ancillary data and the nugget:sill ratio from the variogram estimated by REML from just 38 data. The results for the standardized variogram are comparable to those for the 30-m data and better than those for each of the other sub-samples. There was also a strong correlation between the predictions based on the standardized variogram model and those from the 30-m model. Figure 2.11b shows the variation in LOI using the standardized variogram model and the 38 data for kriging. The map from the 30-m data (Fig. 2.9a) and that for the 38 data based on the standardized variogram (Fig. 2.11b)

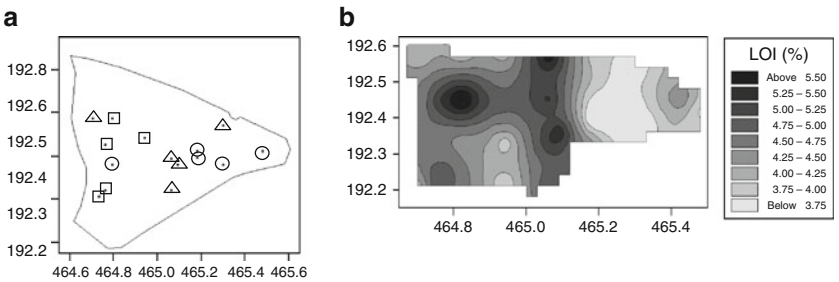

Fig. 2.11 Maps of: (**a**) targeted sample locations based on large, medium and small (L O, M □, S Δ) digital numbers and (**b**) loss on ignition (LOI) for Wallingford using 120-m grid plus targeted data and a standardized variogram model from ancillary data

are more similar than for any of the other sub-sample results (Figs. 2.9 and 2.10). These results support the need for some additional samples at shorter intervals, and suggest that a targeted sample based on ancillary data might reduce the number required.

2.6 Conclusions

This chapter has shown the importance of planning the sampling and analyses to be done before going into the field and the problems of applying a prescriptive approach to sampling based on economic considerations (i.e. one sample per hectare). This book focuses on geostatistical applications in PA and this chapter on sampling for them. Readers should not forget, however, that the message is the same for other types of data analysis although the requirements will be different. Nested sampling and analysis provide a first approximation to the variogram when little or nothing is known about the spatial scale of variation. If there is a relation between soil or crop properties and ancillary data, variograms of the latter can indicate the approximate scale of variation in the former. Existing variograms of soil or crop properties can be used to determine how many samples to bulk from to reduce sampling effects on predictions and also to design optimal sampling schemes on a grid for kriging. The half variogram range 'rule of thumb' can guide sampling to provide predictions that are accurate enough for management in PA. If the sampling intervals indicated are large, there might be too few samples to estimate the variogram accurately by MoM. Variograms estimated by REML and standardized variograms from ancillary data, together with a few judiciously located additional samples, can improve the accuracy of predictions. These approaches provide an interim solution until more intensive information about soil and crop properties can be obtained more cheaply. A general message that emerges from the case studies is that when sampling for eventual geostatistical analysis there is a need for balance between overall cover of

the field, for which a grid survey at an appropriate interval is suitable, and resolution of localized variation which requires some degree of nesting in the sample configuration. This is discussed further in Chapter 3. When there are fewer than 100 soil or crop data, the variogram should be estimated by REML rather than MoM.

References

Blackmore, S. (1994). Precision farming: An introduction. *Outlook on Agriculture*, *4*, 275–280.

Bourgault, G., & Marcotte, D. (1991). Multivariable variogram and its application to the linear model of coregionalization. *Mathematical Geology*, *23*, 899–928.

Burgess, T. M., & Webster, R. (1984). Sampling and bulking strategies for estimating soil properties in small regions. *Journal of Soil Science*, *35*, 127–140.

Frogbrook, Z. L. (2000). *Geostatistics as an aid to soil management for precision agriculture*. Unpublished PhD thesis, University of Reading, Reading, England.

Godwin, R. J., & Miller, P. C. H. (2003). A review of the technologies for mapping within-field variability. *Biosystems Engineering*, *84*, 393–407.

Gower, J. C. (1962). Variance component estimation for unbalanced hierarchical classification. *Biometrics*, *18*, 168–182.

Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*. London: Academic.

Kerry, R., & Oliver, M. A. (2007). Sampling requirements for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, *140*, 383–396.

Kerry, R., & Oliver, M. A. (2008). Determining nugget:sill ratios of standardized variograms from aerial photographs to krige sparse soil data. *Precision Agriculture*, *9*, 33–56.

Kerry, R., Ingram, B. R., Goovaerts, P., & Oliver, M. A. (2008). How many samples are required to estimate a reliable REML variogram? In J. M. Ortiz, & X. Emery (Eds.), *Geostats 2008. Proceedings of the Eighth International Geostatistics Congress* (pp. 1155–1160). Santiago, Chile: Gecamin Ltd.

Lark, R. M. (2000). Estimating variograms of soil properties by the method-of- moments and maximum likelihood. *European Journal of Soil Science*, *51*, 717–728.

McBratney, A. B., & Webster, R. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables II. Program and examples. *Computers & Geosciences*, *7*, 335–365.

McBratney, A. B., Webster, R., & Burgess, T. M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and method. *Computers & Geosciences*, *7*, 331–334.

MAFF. (1986). *The analysis of agricultural materials* (3rd ed.). Reference Book 427. London: Her Majesty's Stationery Office.

McKenzie, N. J., Grundy, M. J., Webster, R., & Ringrose-Voase, A. J. (Eds.) (2008). *Guidelines for surveying soil and land resources* (2nd ed.). Collingwood, Australia: CSIRO Publishing.

Matheron, G. (1965). *Les variables régionalisées et leur estimation*. Paris: Masson et Cie.

Miesch, A. T. (1975). Variograms and variance components in geochemistry and ore evaluation. *Geological Society of America Memoir*, *142*, 333–340.

Oliver, M. A., & Badr, I. (1995). Determining the spatial scale of variation in soil radon concentration. *Mathematical Geology*, *27*, 893–922.

Oliver, M. A., & Frogbrook, Z. L. (1998). *Sampling to estimate soil nutrients for precision agriculture*. York, UK: The International Fertiliser Society.

Oliver, M. A., Frogbrook, Z. L., Webster, R., & Dawson, C. J. (1997). A rational strategy for determining the number of cores for bulked sampling of soil. In J. V. Stafford (Ed.), *Precision agriculture '97. Volume I, spatial variability in soil and crop* (pp. 155–162). Oxford: BIOS Scientific Publishers.

Oliver, M. A., & Webster, R. (1986). Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables. *Geographical Analysis*, *18*, 227–242.

Oliver, M. A., & Webster, R. (1987). The elucidation of soil pattern in the Wyre Forest of the West Midlands, England. II. Spatial distribution. *Journal of Soil Science*, *38*, 293–307.

Pardo-Igúzquiza, E. (1998a). Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, *30*, 95–107.

Pardo-Igúzquiza, E. (1998b). MLREML4: A program for the inference of the power variogram model by maximum likelihood and restricted maximum likelihood. *Computers & Geosciences*, *24*, 537–543.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545–209.

Payne, R. W. (2008). *The guide to GenStat for GenStat release 10: Part 2, statistics*. Hemel Hempstead, UK: VSN International.

Pettitt, A. N., & McBratney, A. B. (1993). Sampling designs for estimating variance components. *Applied Statistics*, *42*, 185–209.

Robert, P. C., Rust, R. H., & Larson, W. E. (Eds.) (1995). *Site-specific management for agricultural systems. Proceedings of the 2nd International Conference*. Madsion, WI: Agronomy Society of America, Crop Science Society of America, Soil Science Society of America.

Robert, P. C., Rust, R. H., & Larson, W. E. (Eds.) (1996). *Precision agriculture, Proceedings of the Third International Conference on Precision Agriculture*. Madsion, WI: Agronomy Society of America, Crop Science Society of America, Soil Science Society of America.

Schueller, J. K. (1997). Technology for precision agriculture. In J. V. Stafford (Ed.), *Precision agriculture'97. Volume I, spatial variability in soil and crop* (pp. 19–33). Oxford, UK: BIOS Scientific Publishers.

Stafford, J. V. (Ed.) (1997). *Precision agriculture '97. Proceedings of the 1st European Conference on Precision Agriculture, Volumes I and II*. Oxford: BIOS Scientific Publications.

Stafford, J. V. (Ed.) (1999). *Precision agriculture '99. Proceedings of the 2nd European Conference on Precision Agriculture, Volumes I and II*. Sheffield, UK: Sheffield Academic Press.

Viscarra-Rossel, R. A., & McBratney, A. B. (1998). Soil chemical analytical accuracy and costs: Implications from precision agriculture. *Australian Journal of Experimental Agriculture*, *38*, 765–775.

Webster, R., & Boag, B. (1992). A geostatistical analysis of cyst nematodes in soil. *Journal of Soil Science*, *43*, 583–595.

Webster, R., & Oliver, M. A. (1990). *Statistical methods in soil and land resource survey*. Oxford: Oxford University Press.

Webster, R., & Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, *43*, 177–192.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. Chichester: Wiley.

Webster, R., Welham, S. J., Potts, J. M., & Oliver, M. A. (2006). Estimating the spatial scales of regionalized variables by nested sampling, hierarchical analysis of variance and residual maximum likelihood. *Computers & Geosciences*, *32*, 1320–1333.

Whelan, B. M., McBratney, A. B., & Viscarra-Rossel, R. A. (1996). Spatial prediction for precision agriculture. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision agriculture* (pp. 331–342). *Proceedings of the 3rd International Conference*. Madison, WI: Agronomy Society of America, Crop Science Society of America, Soil Science Society of America.

Willers, J., Jenkins, J. N., McKinion, J. M., Gerard, P., Hood, K. B., Bassie, J. R., & Cauthen, M. D. (2009). Methods of analysis for georeferenced sample counts of tarnished plant bugs in cotton. *Precision Agriculture*, *10*, 189–212.

Youden, W. J., & Mehlich, A. (1937). Selection of efficient methods for soil sampling. *Contributions of the Boyce Thompson Institute for Plant Research*, *9*, 59–70.

Chapter 3
Sampling in Precision Agriculture, Optimal Designs from Uncertain Models

B.P. Marchant and R.M. Lark

Abstract If farmers are to manage the soil–crop system efficiently through variable application of fertilizers within fields they require information of the within-field variation of soil properties. To ensure that precision agriculture is cost-effective, soil sampling must be as efficient as possible. This chapter demonstrates the potential to optimize the design of soil sampling schemes if the variation of the target property is represented by a linear mixed model. If the parameters of the model are known prior to sampling we see that it is possible to optimize the sampling design with a numerical algorithm known as spatial simulated annealing. In general the parameters are unknown when the sample scheme is designed and the model is fitted to the observations. However it can be sufficient to assume a model which was fitted to a previous survey of the target variable over a similar landscape. When we do not have existing information about the variation of the target variable multi-phase adaptive sampling schemes may be used. We describe such a scheme for a survey of top-soil water content. The data are analysed as they are collected, and the sample design is modified to ensure that it is suitable for the particular target variable. The technologies described in this chapter represent the state of the art for sampling design in the geostatistical context. We discuss the developments required for them to be implemented as standard tools for precision agriculture.

Keywords Sampling · Linear mixed models (LMM) · Spatial simulated annealing · Adaptive sampling · Residual maximum likelihood (REML)

3.1 Introduction

The objective of precision agriculture is to improve the management of crop production by responding to variations of the soil–crop system at within-field scales. An implication of this is that we shall require information on this soil–crop system

B.P. Marchant (✉) and R.M. Lark
Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK
e-mail: ben.marchant@bbsrc.ac.uk; murray.lark@bbsrc.ac.uk

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_3,

at fine scales of resolution. Geostatistical prediction by kriging is an obvious way to obtain such information from a set of observations. The observations must be obtained by an appropriate sampling scheme.

There are considerable costs involved in observing soil properties at within-field scales. These include travel expenses of visiting the field, time taken to move around the field and extract soil cores, and the costs of laboratory analyses of the soil cores. If within-field precision agriculture is to be cost-effective then these costs must be minimized. In this chapter we describe 'state of the art' techniques for optimizing sampling schemes to provide information of the required precision at the minimum cost. Kerry et al. (Chapter 2) show that the sampling requirements are different for different target variables. In general the optimal sampling scheme for a particular target variable is not known prior to sampling. We discuss strategies for overcoming this problem either by making the best use of limited information about the target variable available prior to sampling or by dividing the sampling process into several phases so that the sampling scheme may be modified as more information is obtained.

One advantage of geostatistics over other prediction techniques is that, given a variogram model, we can determine a sampling scheme that allows us to predict the variable of interest with specified precision (McBratney et al. 1981). In geostatistics we make the explicit assumption that the data can be treated as random because they are a realization of a random model, rather than justifying this by the random selection of sample locations (Brus and de Gruijter 1997). Hence the sample sites need not be selected at random. Sampling can be optimized in various ways, depending on the objective function that we choose to minimize or maximize and the constraints that we impose. For example, we could minimize the average kriging variance across our study region, given a fixed sample size. In practice we might be more interested to find the sample size n_{opt} such that a target kriging variance across the study region is just achieved when a sample of this size is distributed optimally. In this case we minimize our sampling costs to achieve predictions of desired precision.

The problem with the simple approach of McBratney et al. (1981) is that it assumes the variogram is known without error. In practice we must estimate the variogram from sample data. We might undertake an initial survey to obtain a variogram, but this estimate will have errors that will make our sampling scheme suboptimal. We need a framework for geostatistical estimation, inference and prediction that allows us to handle uncertainty in the statistical model of spatial variability. Classical geostatistics does not do this, but geostatistics based upon linear mixed models (LMMs) does (Stein 1999). A LMM is a stochastic model to describe the spatial variation of our target variable, and the parameters of this model are estimated by likelihood methods. In this chapter we outline this approach, and illustrate it with examples in which sampling schemes are optimized for the prediction of soil properties at within-field scales.

3.2 The Linear Mixed Model: Estimation, Predictions and Uncertainty

3.2.1 The Model

The approach that we propose in this chapter is based on the LMM, rather than on classical geostatistics. The LMM is so-called because it consists of both fixed and random effects. The latter are equivalent to the random variables of standard geostatistics, and the former consist of a linear combination of q covariates which model the expected value of our target variable at any location. In the simplest LMM $q = 1$, and the fixed effect is just a mean value constant at any location.

We may write the LMM as

$$\mathbf{z} = \mathbf{M}\boldsymbol{\beta} + \mathbf{Yu} + \boldsymbol{\varepsilon}. \tag{3.1}$$

The vector $\mathbf{z}$ contains N observations of our target variable from M locations. Commonly $N = M$ but this need not be the case. The matrix $\mathbf{M}$ $(N \times q)$ is called the design matrix for the fixed effects in the model, and the vector $\boldsymbol{\beta}$, of length q, contains the fixed effects coefficients. As an example, consider a simple case where our target variable is crop yield and we also have measurements of a covariate such as a remotely sensed vegetation index. If it is reasonable to predict yield as a linear function of the vegetation index then we may treat the index as a fixed effect. Each row of the design matrix, $\mathbf{M}$, corresponds to one of our observations. In a simple linear model the first column in row n of the design matrix will take the value 1 (for any n), and the second column will contain the value of the vegetation index that corresponds to our nth observation. In this case the first value in $\boldsymbol{\beta}$ will correspond to the intercept of a linear regression of yield on the vegetation index, and the second value is the regression coefficient. In more complex models we might propose different intercepts for K different management zones of a field. In this case $q = K + 1$, and the entries in the design matrix for an observation in the second management zone will be all zero except for a 1 in the second column and the value of the vegetation index in the qth column.

The random effect, $\mathbf{u}$, is a spatially correlated random variable with variance c_1, and the matrix $\mathbf{Y}$ is the design matrix for the random effects in the model. For simplicity we will assume here that the target variable can be measured only once at any location (i.e. it is a variable such as crop yield or soil pH that is measured destructively). In this case $N = M$ and $\mathbf{Y}$ is an $N \times N$ identity matrix (all values are zero except on the main diagonal where they are all 1). In general this assumption is not necessary and the model can handle situations where duplicate measurements are made at some or all locations (e.g. with a remote sensing device). In general $\mathbf{Y}$ is an $N \times M$ matrix where M is the number of distinct sites at which the target variable is measured and $M \leq N$. Each row of $\mathbf{Y}$ corresponds to one measurement and contains $M-1$ zero entries and a 1 in the column corresponding the site at which the measurement was made. The $M \times 1$ vector $\mathbf{u}$ contains random effects at each observation site. These have zero mean and a covariance matrix $\mathbf{G}$. The $N \times 1$ vector

$\boldsymbol{\varepsilon}$ contains the residuals, which have a covariance matrix $\mathbf{R}$ and are independent of the random effects so that the overall covariance matrix of our variable is

$$\mathbf{V} = \mathbf{Y}\mathbf{G}\mathbf{Y}^{\mathrm{T}} + \mathbf{R}.$$

When $\mathbf{Y}$ is an $N \times N$ identity matrix our usual assumption is that the term $\boldsymbol{\varepsilon}$ comprises both measurement error and the variance in the target variable that is not correlated over distances equal to or larger than the smallest lag distances in our data set. This is the nugget variance in classical geostatistics and so the matrix $\mathbf{R}$ has no off-diagonal terms (the values in $\boldsymbol{\varepsilon}$ are not correlated), and under assumptions of second-order stationarity (Webster and Oliver 2007) all the elements on the main diagonal are equal to the nugget variance, c_0.

The $M \times M$ spatial covariance matrix of the random effects in our model, $\mathbf{G}$, may be determined from an appropriate covariance function, $C(h)$, which gives the covariance for two elements of $\mathbf{u}$ separated in space by the lag distance h (we assume here that the covariance depends on distance only, and not direction, but anisotropic variation can be modelled). This depends on the assumption that the random effect is a second-order stationary random variable so that the variance $C(0)$ exists (see Section 1.2.1 for more explanation). We assume that most readers are more familiar with the variogram than the covariance function, so note that the variogram of a second-order stationary random variable is a function, $\gamma(h)$, bounded at the sill variance, $C(0)$. There are various parametric functions that can be used to represent both the variogram and the covariance function, but in general, when the spatially dependent variance of $\mathbf{u}$ is c_1

$$C(h) = c_1 - \gamma(h). \tag{3.2}$$

We can use this expression to compute any entry in the covariance matrix, $\mathbf{G}$, given the lag distance, h, between the corresponding observations. In a simple case, where $\gamma(h)$ is the familiar exponential variogram with distance parameter a,

$$\begin{aligned} \gamma(h) &= c_1 \left\{ 1 - \exp\left(-\frac{h}{a} \right) \right\} \quad \text{for } h > 0 \\ &= 0 \quad \text{for } h = 0. \end{aligned} \tag{3.3}$$

We denote by $\boldsymbol{\theta}$ the vector of variance parameters in the LMM so in this case $\boldsymbol{\theta} = (c_0, c_1, a)$ but more complex models can be used, and $\boldsymbol{\theta}$ might include parameters to describe anisotropy.

3.2.2 Estimation

To apply the LMM we need to estimate the unknown fixed effects coefficients, given our observations and known design matrices. If we know the variance parameters, and so can compute $\mathbf{V}$, then this can be done by generalized least squares

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{M}^{\mathrm{T}} \mathbf{V}^{-1} \mathbf{M} \right)^{-1} \mathbf{M}^{\mathrm{T}} \mathbf{V}^{-1} \mathbf{z}. \tag{3.4}$$

The problem is to estimate the variance parameters. In the model-based approach we do this by providing a likelihood function for $\mathbf{z}$. The likelihood function is equivalent to a probability density function (pdf) for $\mathbf{z}$. Where the pdf is a function of $\mathbf{z}$, with parameters including the fixed effect coefficients and variance parameters, the likelihood function is a function of these unknown parameters, with the observations $\mathbf{z}$ considered fixed. Maximum likelihood (ML) estimates of the parameters are those for which this likelihood function is largest, given the data. We actually use the log-likelihood, which is at a maximum when the likelihood is at a maximum. Under an assumption of normality of the sum of the random effects and residuals, the likelihood is

$$l(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{z}) = \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{M}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{V}^{-1} (\mathbf{z} - \mathbf{M}\boldsymbol{\beta})\right\}, \tag{3.5}$$

where $|\mathbf{V}|$ is the determinant of $\mathbf{V}$. For convenience we generally work with the negative log likelihood function, $L(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{z}) = -\log\{\ell(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{z})\}$. To calculate L for a possible value of $\boldsymbol{\theta}$, we must compute the generalized least-squares estimate of the fixed effects coefficients with Eq. 3.4, then substitute the resulting estimates, $\widehat{\boldsymbol{\beta}}$, into Eq. 3.5. The ML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are obtained by iteration of this process in a numerical optimization method to find the values that minimize the negative log-likelihood. Thus the estimates of the variance parameters depend on the estimates of the fixed effects coefficients. The latter are said, therefore to be 'nuisance parameters' for the estimation of the former because the variance parameter estimates depend in a non-linear way on the estimated fixed effects coefficients their estimates are subject to bias.

Consider a simple situation in which a variable, z, sampled at regular intervals along a transect, comprises a linear trend along the transect together with superimposed random variation with a mean of zero. The mean of the variable, and the slope of the trend (the fixed effects) are unknown. If we replace each observation $z(x_i)$ with the difference $d'(x_i) = z(x_i) - z(x_{i+1})$ it is easy to show that the expected value (i.e. the mean) of the resulting variable depends on the slope of the trend, but not the mean of z. We say that the mean has been filtered out. If we repeat the differencing operation to compute $d''(x_i) = d'(x_i) - d'(x_{i+1})$ it is again easy to see that both fixed effects have now been removed, the expectation of $d''(x_i)$ is zero. We do not need to estimate the fixed effects to filter them out in this way, we simply need to know the design matrix. The principle of residual maximum likelihood (REML), due to Patterson and Thompson (1971), is to generalize this filtering procedure for any design matrix of a linear model to obtain a new variable of known expectation (zero) which is a linear function of the observed variable. A likelihood function can then be written down for this new variable in terms of the covariance matrix of the original variable, $\mathbf{V}$. This is called the residual likelihood, and the parameters for $\mathbf{V}$, which maximize it, are the REML estimates. The REML estimates do not depend on the unknown fixed effects, since the non-random component of the new variable has known expectation (zero), so the estimates are substantially less biased than ML estimates.

For a given design matrix, $\mathbf{M}$, the residual negative log likelihood is

$$L_R(\boldsymbol{\theta}|\mathbf{z}) = \text{Constant} + \frac{1}{2}\left(\log|\mathbf{V}| + \log\left|\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{M}\right| + \mathbf{z}^{\mathrm{T}}\mathbf{P}\mathbf{z}\right), \tag{3.6}$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{M}\left(\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}.$$

The REML estimator is fully described by Diggle and Ribeiro (2007).

The likelihood functions in Eqs. 3.5 and 3.6 make the explicit assumption that the random terms in the LMM are N-variate normal. We have only a single N-variate realization, so cannot test this assumption directly; but the histogram of our observations, or of the residuals if the fixed effects are more complex than an overall mean, indicates whether the assumption is plausible. When it is not a transformation should be found. However, Pardo-Igúzquiza (1998) argues that the maximum likelihood criterion for estimation is generally a good one in a maximum entropy sense even when we cannot verify assumptions of normality. Furthermore, Kitanidis (1985) reports simulation studies in which maximum likelihood gave the best estimates for spatial covariance parameters even for variables generated from processes known to be non-normal.

3.2.3 Prediction

Once we have estimated $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ we are in a position to obtain predictions of the target variable at unsampled sites. These predictions are empirical best linear unbiased predictions (E-BLUPs), empirical because they are based on estimated variance parameters. These are equivalent to universal kriging predictions (sometimes called kriging with external drift when the fixed effects are not terms in a model of spatial trend). The E-BLUPs at a set of N_{p} prediction sites, $\mathbf{x}_{\mathrm{p}}$, are given by

$$\begin{aligned}\tilde{\mathbf{z}}(\mathbf{x}_{\mathrm{p}}) &= \left(\mathbf{M}_{\mathrm{p}} - \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\mathbf{M}\right)\widehat{\boldsymbol{\beta}} + \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\mathbf{z}, \\ &= \left\{\left(\mathbf{M}_{\mathrm{p}} - \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\mathbf{M}\right)\left(\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1} + \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\right\}\mathbf{z}, \\ &= \boldsymbol{\lambda}\mathbf{z}, \end{aligned} \tag{3.7}$$

where the $N_{\mathrm{p}} \times q$ matrix $\mathbf{M}_{\mathrm{p}}$ is the design matrix for the prediction sites and $\mathbf{V}_{\mathrm{po}}$ is an $N_{\mathrm{p}} \times N$ covariance matrix in which element $\{i, j\}$ contains the covariance of the ith prediction site with the jth observation site, computed from the covariance function and estimated variance parameters by Eq. 3.2. In the same way we can define the $N_{\mathrm{p}} \times N_{\mathrm{p}}$ covariance matrix of the prediction sites, $\mathbf{V}_{\mathrm{pp}}$. The covariance matrix of the prediction errors is

$$\mathbf{C} = \left(\mathbf{M}_{\mathrm{p}} - \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\mathbf{M}\right)\mathbf{H}^{-1}\left(\mathbf{M}_{\mathrm{p}} - \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\mathbf{M}\right)^{\mathrm{T}} + \mathbf{V}_{\mathrm{pp}} - \mathbf{V}_{\mathrm{po}}\mathbf{V}^{-1}\mathbf{V}_{\mathrm{po}}^{\mathrm{T}}, \tag{3.8}$$

where

$$\mathbf{H} = \mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{M},$$

so the diagonal elements of $\mathbf{C}$ are the prediction variances which are equivalent to σ^2_{K}, the kriging variances of classical geostatistics calculated at each of the N_{p} prediction sites.

3.2.4 Uncertainty

The LMM framework for estimating spatial models and predicting with them from data has several advantages over classical geostatistics. One of these is the fact that classical geostatistics has no general and theoretically satisfactory way of estimating variance parameters for the random effects and residuals when the fixed effect is anything more complex than a simple mean (Lark et al. 2006). Of particular relevance to our interests in spatial sampling is that an account of the uncertainty of the variance parameters, and the implications of this uncertainty for sampling decisions and estimation, is tractable in the model-based approach.

Consider a negative log-likelihood function $L(\boldsymbol{\theta}|\mathbf{z})$, minimized to find an estimate, $\widehat{\boldsymbol{\theta}}$ of the parameters. Element $\{i, j\}$ of the Fisher information matrix is defined as follows

$$\left[\mathbf{F}\left(\widehat{\boldsymbol{\theta}}\right)\right]_{ij} = -\mathrm{E}\left[\frac{\partial^2 L\left(\widehat{\boldsymbol{\theta}}|\mathbf{z}\right)}{\partial\theta_i \partial\theta_j^{\mathrm{T}}}\right]. \tag{3.9}$$

The inverse of this matrix is the covariance matrix of $\widehat{\boldsymbol{\theta}}$, as discussed in standard texts such as Dobson (1990) and also the appendix to Marchant and Lark (2006). Under an assumption of normality in the estimation errors of the parameters we can compute confidence intervals from this matrix. However, Marchant and Lark (2004) showed that the assumptions behind this approach may break down for parameters in which the covariance function is non-linear (e.g. the range parameter), and that the uncertainty of distance parameters may be underestimated as a result.

In practice we may compute the Fisher information matrix for a given $\boldsymbol{\theta}$ and set of observation sites with an expression due to Kitanidis (1987)

$$[\mathbf{F}(\boldsymbol{\theta})]_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j\right], \tag{3.10}$$

where $\mathrm{Tr}[\cdot]$ denotes the trace – the sum of the entries on the main diagonal – of the matrix in the brackets and $\mathbf{V}_i$ denotes $\partial\mathbf{V}/\partial\theta_i$.

3.3 Optimizing Sampling Schemes by Spatial Simulated Annealing

3.3.1 Spatial Simulated Annealing

We observed above that to optimize a geostatistical survey we need to find a way to distribute sample points across the area of interest that maximizes some measure of the utility of the outcome (the objective function) subject to constraints. We might find, for example, a sampling scheme that minimizes the average prediction variance at sites across a region subject to the constraint that only $N_{\max}$ sample points are used. Alternatively, we might find the most efficient way to ensure that the average prediction variance at sites across a region is no larger than some threshold, which requires the minimization of the same objective function (average prediction variance) for increasing values of $N_{\max}$ until we find the smallest sample size that meets our overall objective to be met.

Given an objective function, and constraints, how can the optimal sampling scheme be determined? A practical solution to the problem is spatial simulated annealing (SSA), proposed by Van Groenigen et al. (1999) and used subsequently in various case studies. A detailed discussion of simulated annealing is beyond the scope of this chapter and the reader is referred to Kirkpatrick et al. (1983), Aarts and Korst (1989) and Press et al. (1992), for detailed information.

Simulated annealing is a method of numerical optimization to find the values of a set of parameters of an objective function, ϕ, that optimize (minimize or maximize) that function. In SSA the parameters are the co-ordinates of a set of sample points, and the objective function could be the average prediction error variance evaluated over a fine network of points in our sample area. Simulated annealing proceeds by random perturbation of an initial set of parameter values, in SSA this means moving each sample point randomly in turn. Perturbations that improve the objective function are all accepted, whereas those that make it worse are accepted or rejected at random where the probability of acceptance, p_{a}, depends on a function (the Metropolis criterion) that simulates the statistical mechanics of atoms in a molten metal. Thus, if the proposed change in the system results in a change in the objective function from ϕ_i to ϕ_j, where $\phi_j > \phi_i$, then the probability of acceptance of the change is

$$p_{\mathrm{a}} = \exp\left(\frac{\phi_i - \phi_j}{\kappa}\right), \tag{3.11}$$

where κ is a parameter analogous to the temperature of the metal.

If the parameters are perturbed at random many times in sequence at a fixed temperature, with changes accepted or rejected according to this criterion, then the system approaches a thermal equilibrium such that the value of the objective function is a sample from Boltzmann's distribution (Aarts and Korst 1989). In simulated annealing the system is taken through many such sequences of perturbations, and

at the end of each the temperature is reduced. Reduction of the temperature reduces the probability of acceptance of a change in the parameters that makes the objective function worse by some given amount. The aim is to emulate the slow cooling of a molten metal that will cause it to 'anneal', i.e. to reach an energy state that is a global minimum – a regular crystalline solid. The particular advantage of simulated annealing as a method of optimization is that the Metropolis criterion allows the system in effect to jump over a barrier that could trap it at a solution that is only locally optimal. This could not be achieved by a procedure in which any change that makes the objective function worse is rejected.

A successful simulated annealing requires a good 'cooling schedule'. If the temperature is reduced too quickly then the parameter estimates are likely to become stuck at a solution that is only locally optimal, whereas if the temperature is reduced too slowly then considerable computing time may be needed before the parameter estimates converge to a solution. Most applications of SSA have used a cooling schedule proposed by Kirkpatrick et al. (1983). The initial temperature of the system, κ_1 is chosen so that the proportion of proposed changes accepted before the first reduction in temperature is in the range 0.90–0.99 and the new temperature of the system κ_{m+1} after the mth cooling step is $\alpha_c \kappa_m$ where $\alpha_c = 0.95$. The cooling step takes place after a fixed number of perturbations of each parameter of the objective function. Similarly the maximum distance that a sample point can be perturbed may also be reduced at each cooling step. The algorithm stops when a given number of successive cooling steps have taken place with no further change in the objective function larger than some threshold. The lengths of the chains of perturbations and the stopping criteria are found by trial and error; Lark and Papritz (2003) give some diagnostics to help with this.

A particular advantage of simulated annealing as a method to optimize spatial sampling is that an irregularly-shaped sampling region, and constraints on sampling (e.g. ponds or buildings in the middle of the area to be sampled) are easily accommodated. If a random perturbation of a sample point takes it outside the region or into an area that cannot be sampled then it is simply returned to its original position and moved randomly again.

3.3.2 Objective Functions from the LMM

Lark (2002) illustrated the use of SSA to optimize spatial sampling for variogram estimation in terms of an objective function based on the covariance matrix of the variance parameters. He envisaged a prediction location at the centre of a cell in a sample grid. Errors in the variance parameters of the LMM (i.e. variogram parameters) will lead to error in our computed value of the prediction variance at that location. Lark (2002) gave an objective function that was the error variance of the prediction error variance (i.e. the mean squared error of the kriging variance). This can be computed from the covariance matrix of the variance parameters, computed in turn by inverting the Fisher information matrix. Given the location of the sample

points and a set of true variance parameters the Fisher information matrix is obtained using Eq. 3.10. The objective function is

$$\phi_{\mathrm{E}}(\mathbf{S};\boldsymbol{\theta}) = \sum_{i=1}^{q}\sum_{j=1}^{q} \mathrm{Cov}\left(\theta_i, \theta_j\right) \frac{\partial \sigma_{\mathrm{K}}^2}{\partial \theta_i} \frac{\partial \sigma_{\mathrm{K}}^2}{\partial \theta_j}, \tag{3.12}$$

where $\mathbf{S}$ is the vector of sampling locations.

Figure 3.1 shows sampling schemes that have been optimized for 49 points in a square region with different underlying variance parameters. This example serves two purposes. First, it illustrates SSA for a case where statistics obtained from the LMM enable us to manage uncertainty in the variance parameters. Second, it motivates much of what follows. Figure 3.1 shows how optimal sampling schemes for the variogram depend strongly on the underlying spatial covariance. If the distance parameter is small relative to the sampled area, or the proportion of the overall variance that is spatially correlated is small, or both, then sampling points tend to be clustered in the optimal arrays. If the distance parameter is large and the nugget

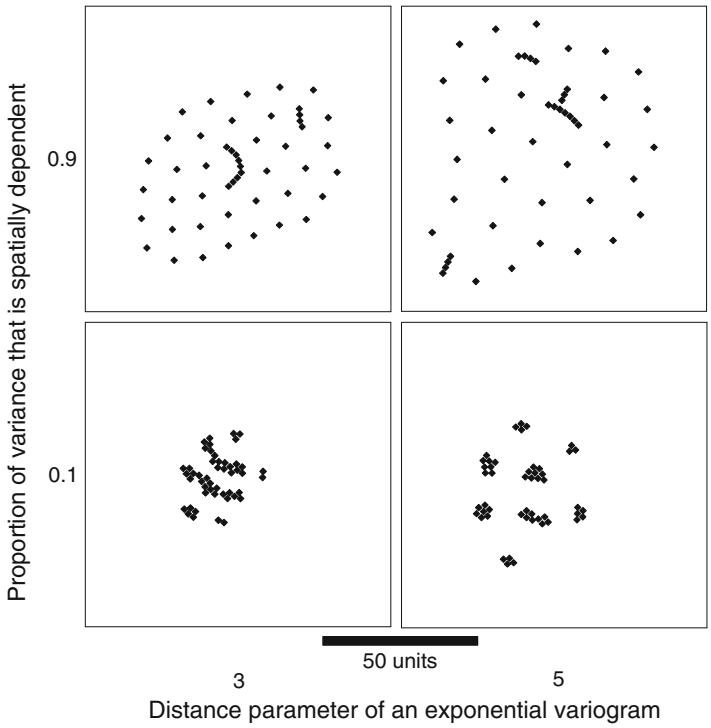

Fig. 3.1 Surveys optimized by the criterion of Lark (2002) for exponential variogram estimation. They are arranged according to the specified spatial structure (*columns*) and the ratio of spatial dependence $c_1/(c_0 + c_1)$ (*rows*)

variance relatively small, then the sampling points tend to be dispersed, with a few chains of close points. In intermediate conditions clusters of two points are dispersed across the region.

Van Groenigen et al. (1999) optimized sample schemes for prediction. Their objective function was

$$\phi_{\mathrm{P}}(\mathbf{S};\boldsymbol{\theta}) = \text{mean}\left(\boldsymbol{\sigma}_{\mathrm{K}}^2\right), \tag{3.13}$$

where $\boldsymbol{\sigma}_{\mathrm{K}}^2$ is the prediction variance calculated over a fine grid of prediction sites. An example of the sampling scheme that results from this objective function is shown in Fig. 3.2 (KV) for a spherical variogram with parameters $c_0 = 0.127$, $c_1 = 1.0$ and $a = 240.0$ m. Here and in all optimizations presented in this chapter the SSA procedure was repeated five times to ensure that convergence to a global minimum had occurred. By contrast to the schemes for variogram estimation, the sampling locations are dispersed evenly over the region and this pattern is generally insensitive to changes in the variogram parameters. Brus and Heuvelink (2007) extended this approach to prediction when there was a linear relationship between the property of interest and an exhaustively sampled covariate. This may be applicable in precision agriculture if exhaustive electrical conductivity or remotely sensed data are available. The resulting schemes were a trade-off between ensuring good spatial cover and that the extremes of the covariate were well sampled so that the linear relationship could be fitted. In practice a geostatistical survey of a region must be suitable for estimating both the variogram and returning reliable local predictions. It has not been clear how much of the sampling effort should be devoted to variogram estimation and how much to prediction.

In the early days of the application of geostatistics to soil mapping it was suggested that an exploratory survey might return information on the variogram to plan the rest of the survey. For example, McBratney et al. (1981) suggested that the variogram should be approximated from previous surveys of the target soil property on cognate soil types or estimated from an exploratory survey, and that the main survey should consist of measurements made on a regular grid with spacing, $I(\boldsymbol{\theta})$, chosen to ensure that the kriging variance at the centre of each grid cell is less than some pre-specified threshold. However, it is unlikely that we will often have a prior survey that we can confidently assume represents the variation in our region of interest. The results of Lark (2002) show that when we start in ignorance of the variogram we do not know how best to select the number and location of observations in an exploratory survey. Optimal sampling schemes to estimate the variogram may differ markedly between different fields. Spatially nested sampling may be used efficiently to identify the important spatial scales of variation in a particular variable (see, for example Chapters 2 and 9), but considerable uncertainty remains about the form of the variogram. We consider practical solutions to these dilemmas in the following sections.

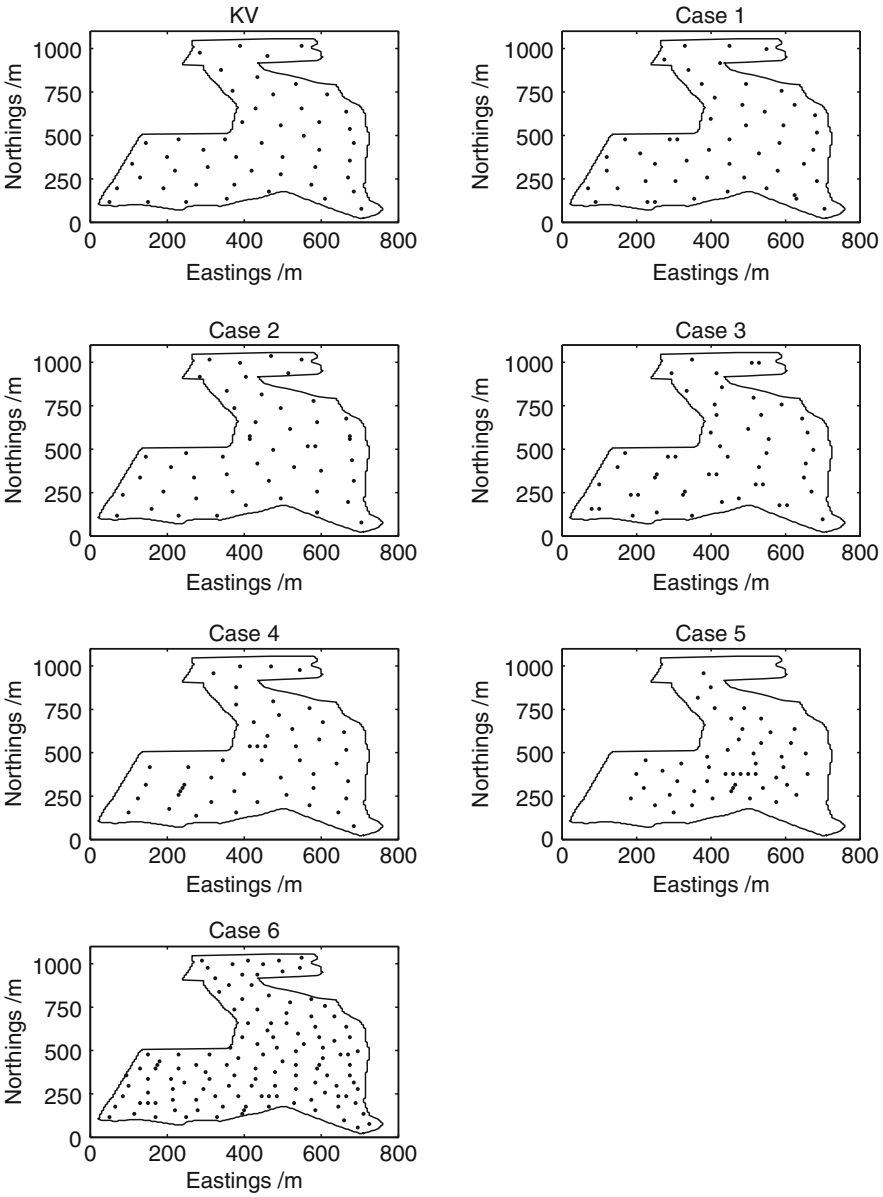

Fig. 3.2 Sampling scheme optimized to minimize the mean kriging variance across the trial region for a spherical variogram with parameters $c_0 = 0.127$, $c_1 = 1.0$ and $a = 240.0$ m (KV) and schemes optimized to minimize the expected MSE across the trial region for spherical variograms with the parameters listed in Table 3.1 (Cases 2–6)

3.3.3 Optimized Sample Scheme for Single Phase Geostatistical Surveys

We have seen above that optimal sampling schemes for variogram estimation (Fig. 3.1) are quite different from those for prediction (Fig. 3.2, KV). When we design a single sampling scheme to perform both of these tasks we must decide how much of our sampling resources should be allocated to each.

Marchant and Lark (2006) and Zhu and Stein (2006) addressed this problem in a similar manner. They optimized sampling schemes by SSA with an objective function that approximated $\boldsymbol{\sigma}_{\mathrm{T}}^2$, the total prediction variance in a geostatistical survey. This objective function was the sum of the kriging variance and the prediction variance due to variogram estimation. Zimmerman and Cressie (1992) approximated the component of variance in predictions of the target variable due to variogram estimation by

$$\sigma_{\mathrm{V}}^2 = \mathrm{Tr}\left\{\mathbf{F}^{-1}(\boldsymbol{\theta})\left(\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\theta}}\right)\mathbf{V}\left(\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\theta}}\right)\right\}. \tag{3.14}$$

The objective function of Marchant and Lark (2006) and Zhu and Stein (2006) was

$$\phi_{\mathrm{E,P}}(\mathbf{S}) = \mathrm{mean}\left(\boldsymbol{\sigma}_{\mathrm{T}}^2\right) = \mathrm{mean}\left(\boldsymbol{\sigma}_{\mathrm{V}}^2 + \boldsymbol{\sigma}_{\mathrm{K}}^2\right), \tag{3.15}$$

where $\boldsymbol{\sigma}_{\mathrm{V}}^2$ and $\boldsymbol{\sigma}_{\mathrm{K}}^2$ are the vectors of prediction variances due to variogram estimation and prediction, respectively, evaluated at the prediction sites. In general, the $\partial\boldsymbol{\lambda}/\partial\theta_i$ derivatives in Eq. 3.14 must be approximated by numerical techniques which can be computationally intensive. Marchant and Lark (2007) noted that an exact expression for these derivatives exists for the model with no fixed effects from classical geostatistics as

$$\frac{\partial\boldsymbol{\lambda}}{\partial\theta_i} = \mathbf{V}^{-1}\left(\frac{\partial\mathbf{V}_{\mathrm{po}}}{\partial\theta_i} - \frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{V}^{-1}\mathbf{V}_{\mathrm{po}}\right). \tag{3.16}$$

It can be determined without numerical approximation using analytical expressions for the derivatives of the covariance functions with respect to their parameters.

The objective function used by Marchant and Lark (2007) was Eq. 3.15 with the $\partial\boldsymbol{\lambda}/\partial\theta_i$ derivatives determined from Eq. 3.16. This approach might be considered as a compromise between the classical approach and LMMs as the variogram uncertainty is determined assuming REML estimation, but the prediction variance is determined assuming ordinary kriging. The approach of Zhu and Stein (2006) assumed that LMMs were used for both variogram estimation and prediction.

Marchant and Lark (2007) optimized sampling schemes with 50 observations for properties with different spherical variograms. The data were from a field of 42.3 ha at Silsoe, Bedfordshire, UK. The optimized schemes are shown in Fig. 3.2 (Cases 1–6) and the variogram parameters are given in Table 3.1. The pattern of sampling locations changes with the variogram parameters. In Case 1 where $c_0 = 0.127$, $c_1 = 1.0$ and $a = 240.0$ m the majority of locations are evenly dispersed across the study region as required for local prediction. However there are also three close pairs

Table 3.1 Specified spherical variogram parameters c_0, c_1 and a and number of observations, N, for each optimized sampling scheme featured in Fig. 3.2

Case	c_0	c_1	a /m	N
1	0.127	1.000	240.0	50
2	0.127	1.500	240.0	50
3	0.127	0.500	240.0	50
4	0.127	1.000	120.0	50
5	0.127	1.000	90.0	50
6	0.127	1.000	90.0	125

of points separated by 20.0 m, the minimum separation distance allowed in these schemes. Kerry et al. in Section 2.5.1 demonstrate the benefit of a small number of close pairs in their case study. The close pairs reduce the estimation variance of the variogram parameters. The pattern of sampling locations is similar when the sill parameter is increased to 1.5 in Case 2. However, when the nugget to sill ratio is increased in Case 3 the number of close pairs increases.

The number of close pairs also increases when the range is decreased to 120 m in Case 4 and 90 m in Case 5. Short transects of observations are evident in these schemes. In Case 5 some portions of the study region close to the boundary are unsampled. There are many close pairs to estimate the variogram parameters adequately thus insufficient observations remain to ensure good local prediction everywhere. The objective function is a trade-off between both considerations. For this variogram 50 observations are insufficient to manage the demands of both effectively. Therefore, in Case 6 a sampling scheme of 125 observations is optimized for the same variogram parameters. Here there is both good spatial cover and sufficient close pairs to estimate the variogram.

Thus the approach suggested by Marchant and Lark (2007) and Zhu and Stein (2006) is suitable for optimizing sampling schemes for both estimating the variogram parameters and for prediction. The resulting schemes have good spatial cover and also generally contain several close pairs of sampling points to aid variogram estimation. This approach requires the variogram parameters as an input and the pattern of the optimized schemes is different for different variograms. For example, the number of close pairs in the optimized schemes increases if the range of spatial correlation of the variogram decreases. Therefore, if a scheme is optimized based upon an assumed set of variogram parameters then it is likely to be suboptimal. In the next section we explore strategies for ensuring that sampling schemes are optimal for the particular soil property being measured.

3.3.4 Adaptive Exploratory Surveys to Estimate the Variogram

We have identified two problems with exploratory surveys to choose the grid spacing in a geostatistical survey. First the optimal sampling scheme for estimation of

the variogram depends upon that same underlying variogram. Thus if we were to design an optimal exploratory survey based on an assumed variogram then that survey is likely to be suboptimal. Second, we do not know how to decide how many measurements should be included in the exploratory survey. Marchant and Lark (2006) suggested that both of these problems could be accounted for by a Bayesian adaptive sampling approach. The Bayesian approach considers a statistical distribution of variogram parameters over all plausible parameter values. This distribution is modified as successive stages of sampling provide new evidence about the spatial variation of the variable of interest, and the integrated effect of this uncertainty on our uncertainty about any quantity that depends on the variogram parameters can be computed.

Bayesian adaptive sampling divides the exploratory survey into several phases. Before the first sampling phase there are no data from which the variogram can be estimated. The Bayesian approach requires a prior distribution for the parameters of the variogram, which reflects our initial beliefs about what values these might take. The prior distribution could be based on previous surveys of the same target variable over similar study areas. Marchant and Lark (2006) initially assumed that, for each variogram parameter, all values were equally likely over some plausible range and that the values of the different parameters were independent of each other. The bounded region of parameter space and the prior distribution are denoted $\boldsymbol{\Theta}$ and $p(\boldsymbol{\theta})$, respectively. The objective function of Marchant and Lark (2006), to be minimized in the design of initial phases of the sampling scheme, was based on ϕ_{E} (Eq. 3.12). This objective function is a measure of our uncertainty about the kriging variance at the centre of a regular grid cell which arises from uncertainty about the variogram parameters (it is an approximation to the estimation variance of the kriging variance). The design of the initial phases of the sampling scheme therefore aims to minimize our uncertainty about how good the kriging predictions based on the resulting variogram will be. However, rather than minimizing ϕ_{E} for a single assumed variogram, Marchant and Lark (2006) minimized it over the whole prior distribution of variograms $p(\boldsymbol{\theta})$. Also, for each set of variogram parameters in $\boldsymbol{\Theta}$ they calculated ϕ_{E} for observations made on a grid with spacing $I(\boldsymbol{\theta})$ – the spacing that leads to the threshold on the kriging variance being satisfied for parameter vector $\boldsymbol{\theta}$ – rather than for an arbitrary spacing. Thus the objective function was written

$$\phi_{\mathrm{B}}(\mathbf{S}) = \int_{\boldsymbol{\Theta}} \phi_{\mathrm{E}}(\mathbf{S};\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}. \tag{3.17}$$

The integral in this objective function is approximated by discretizing $\boldsymbol{\Theta}$, the plausible region of the parameter space, on a regular mesh and calculating $\phi_{\mathrm{E}}(\mathbf{S};\boldsymbol{\theta})\, p(\boldsymbol{\theta})$ at each node of this mesh.

If it is not practical to have more than one phase in the exploratory survey then it may be optimized with Eq. 3.17. Such a survey is referred to as a Bayesian survey and is designed to ensure that the sample scheme is suitable for a variable with any plausible variogram. However Marchant and Lark (2006) consider the situation where several exploratory sampling phases are possible and do what we refer to as a Bayesian adaptive survey. They use Eq. 3.17 to optimize the initial phase. Then

having measured the target variable at each of the optimized sites they use the observations to update their prior distribution by Bayes's rule (Pardo-Igúzquiza and Dowd 2003)

$$\tilde{p}\left(\boldsymbol{\theta}|\mathbf{z}\right) \propto p\left(\boldsymbol{\theta}\right) l\left(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{z}\right), \tag{3.18}$$

where $\tilde{p}\left(\boldsymbol{\theta}|\mathbf{z}\right)$ is the conditional pdf of $\boldsymbol{\theta}$ given $\mathbf{z}$. The conditional pdf is calculated at each node on the regular grid which approximates $\boldsymbol{\Theta}$. The constant of proportionality may be determined by integrating $p\left(\boldsymbol{\theta}|\mathbf{z}\right)$ over $\boldsymbol{\Theta}$ since $\int_{\boldsymbol{\Theta}} \tilde{p}\left(\boldsymbol{\theta}\right) = 1$.

Thus following the first phase of the exploratory survey one may estimate a distribution for the variogram parameters $\tilde{p}\left(\boldsymbol{\theta}|\mathbf{z}\right)$ rather than a single estimate. From this distribution it is possible to calculate the distribution of the sampling interval for the main survey since each $\boldsymbol{\theta}$ corresponds to a sampling interval $I\left(\boldsymbol{\theta}\right)$. If, having examined the distribution of $I\left(\boldsymbol{\theta}\right)$, the practitioner is satisfied that he or she knows the optimal sampling interval with adequate certainty then he may stop the exploratory survey and select a sampling interval for the main survey. We suggest that a conservative interval is selected such as I^{95-} the lower 95% confidence limit of $I\left(\boldsymbol{\theta}\right)$. If the practitioner is not satisfied with the certainty of their estimate of $I\left(\boldsymbol{\theta}\right)$ another exploratory phase may be optimized. The objective function is Eq. 3.17 with $p\left(\boldsymbol{\theta}\right)$ replaced by $\tilde{p}\left(\boldsymbol{\theta}|\mathbf{z}\right)$. Thus as the exploratory survey proceeds the sampling locations become more suited to the actual underlying variogram.

Marchant and Lark (2006) compared the performance of the Bayesian adaptive design, the Bayesian design and the exploratory survey design of Lark (2002) in a series of tests on simulated data. They found that the Bayesian adaptive design was more efficient than the Bayesian one which in turn was more efficient than that of Lark (2002). Figure 3.3 shows two examples of exploratory surveys that result from the Bayesian adaptive design. Both are based on spherical variograms. The design in (a) results when the underlying variogram has a range of spatial correlation of 10 units, whereas the underlying variogram leading to (b) has a range of 30 units. Both optimized surveys gave comparisons between observations separated by both

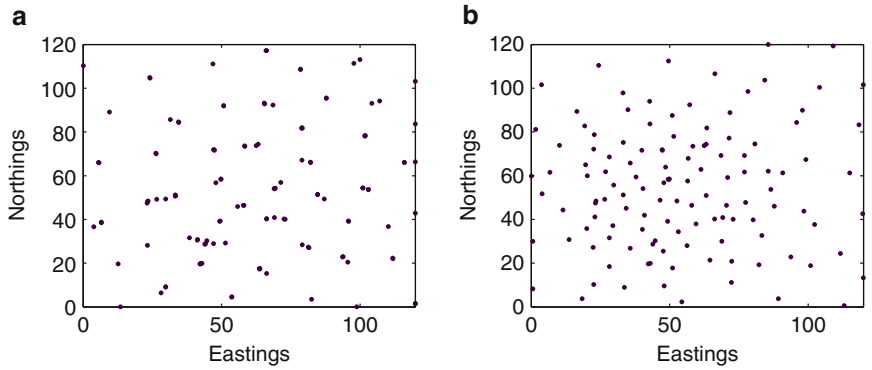

Fig. 3.3 Examples of the Bayesian Adaptive surveys for 124 observations resulting from the 'short-range' (**a**) and 'long-range' (**b**) spherical variogram functions

short and long lags. Thus they are suitable for estimating the degree of spatial correlation over all relevant scales. This is because the initial phase of sampling is optimized to provide information about the form of the variogram, given considerable uncertainty (expressed in the prior distribution) about what this might be. In this way the early phase in Bayesian adaptive sampling is comparable to the nested sampling approach used in Chapters 2 and 9. There are differences between the pattern of locations that result from the information gained during early phases and from the adaptive nature of the survey designs. The survey for the 'short-range' variable contains some close pairs of observations and some short transects with a small separation distance. Observations of the 'long-range' variable are more evenly dispersed with fewer close pairs.

3.4 A Case Study in Soil Sampling

Marchant and Lark (2006) demonstrated their Bayesian adaptive sampling algorithm in a survey of top soil water content (%) over a 90 × 60 m field in Silsoe, Bedfordshire, UK (lat.= 52°0′N, long.= 0°24′W). The aim of the entire survey was to predict top soil water content across the field with a prediction variance of less than $21\%^2$ at each prediction site on a square grid with a spacing of 1 m. The Bayesian adaptive sampling algorithm was run on a portable computer connected to a GPS (Fig. 3.4). At the start of each sampling phase the locations were selected by the algorithm, then the software directed users to each of these locations where top soil water content was measured by a Theta probe (Delta-T Devices 1999) and recorded in the computer. At the end of each sampling phase the pdfs of $\boldsymbol{\theta}$ and $I(\boldsymbol{\theta})$ were calculated. If the optimal value of $I(\boldsymbol{\theta})$ was known with adequate certainty, the exploratory survey was halted and the main survey was designed. Otherwise another phase of the exploratory survey was designed.

The main survey was made on a regular triangular grid. Topsoil water content was a convenient target variable for this demonstration because it could be measured rapidly in the field. The whole survey was completed in less than 3 h. Adaptive sampling algorithms can also be practical for variables that are determined by laboratory analysis (e.g. Demougeot-Renard et al. 2004) provided that the variable remains reasonably constant over the time required for field measurement and laboratory analysis. However, we envisage that the adaptive algorithm will be most effective with soil sensors that return data in real time. Such sensors have been developed in the context of precision agriculture. For example Adamchuk et al. (2005) evaluated sensors based on ion-sensitive electrodes for soil pH, macronutrients (potassium and nitrate nitrogen) and a micronutrient (sodium). Adsett et al. (1999) developed an on-the-go sensor for nitrate based on ion-sensitive electrodes. Viscarra Rossel et al. (2005) developed a system that uses ion-sensitive electrodes and on-the-go processing and analysis to measure soil pH and lime-requirement.

The above survey resulted in 79 observations which ensured with 95% confidence that the threshold on the prediction variance was satisfied. The exploratory

Fig. 3.4 A field system for adaptive sampling

portion of the survey consisted of 75 observations divided into four phases (Fig. 3.5). The remaining four observations were required to complete the main survey. The cumulative distribution function (cdf) of $I(\boldsymbol{\theta})$ after each phase is shown in Fig. 3.6. Table 3.2 gives overall descriptors of confidence in the estimate of $I(\boldsymbol{\theta})$ after each phase. These are I^{95-}, the lower 95% confidence limit on the optimal interval, and N_1^{95-}, the number of observations required to complete a regular survey with this interval, $I(\hat{\boldsymbol{\theta}})$, the interval corresponding to the REML estimate of the variogram parameters, and N_1^{R}, the number of observations required to complete a regular survey with this interval.

Phase 1 consisted of 30 observations. Following this phase $I^{95-} = 6.4$ m and $I(\hat{\boldsymbol{\theta}}) = 15.6$ m. In terms of sampling requirements for the main survey these intervals corresponded to $N_1^{95-} = 142$ for I^{95-} compared with $N_1^{\mathrm{R}} = 25$ for $I(\hat{\boldsymbol{\theta}})$. Thus if the exploratory survey had been halted after one phase and the main survey designed with interval I^{95-} then there is potential for wasted sampling effort. Therefore a second exploratory phase consisting of a further 30 observations was designed.

After Phase 2 the value of N_1^{95-} decreased to 37 compared with $N_1^{\mathrm{R}} = 0$ (i.e. the exploratory observations are sufficient to complete the survey). Thus the sampling intensity required in the main survey was still uncertain so a third phase

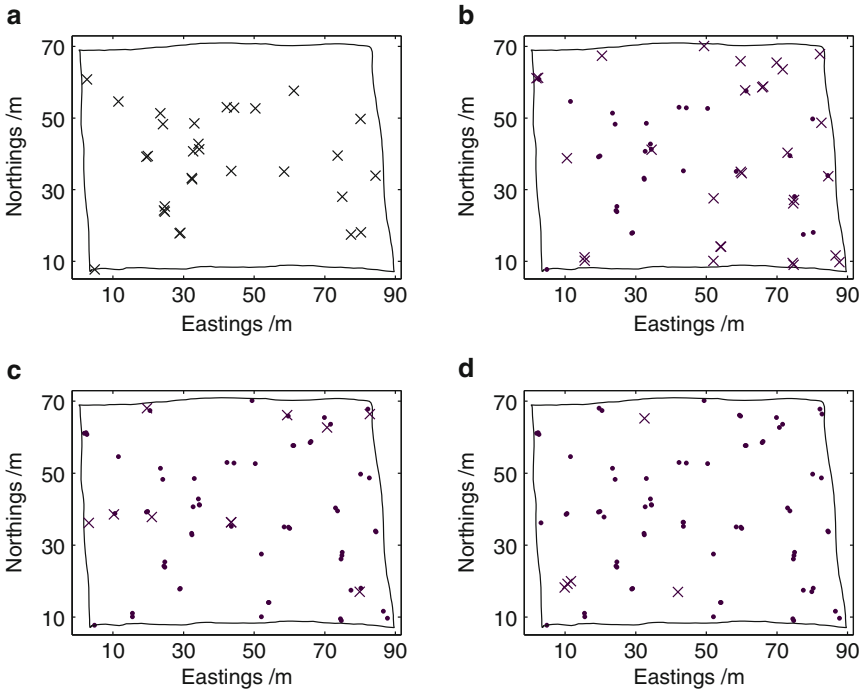

Fig. 3.5 The locations of observations for: (**a**) Phase 1, (**b**) Phase 2, (**c**) Phase 3 and (**d**) Phase 4 of the exploratory survey for the Silsoe case study. The *irregular solid line* is the field boundary. The *dots* denote previous observations and the *crosses* observations in that phase. Coordinates in metres relative to an origin at 508 980, 235 670 on the UK Ordnance Survey National Grid

of 10 points was designed. Following the third phase N_1^{95-} decreased further to 22, but as $N_1^{\mathrm{R}} = 1$ a fourth phase of five points was designed. This phase reduced N_1^{95-} by four to 18, whereas $N_1^{\mathrm{R}} = 1$. Since N_1^{95-} had decreased by less than the sampling effort that had been added to the exploratory survey in the fourth phase, the exploratory survey was halted.

The Bayesian adaptive algorithm suggested a main survey with 18 observations on a regular triangular grid. However Marchant and Lark (2006) considered how further efficiencies could be made. First, they noted that some portions of the field had already been sampled intensively during the exploratory survey and therefore some of the main survey observations within these portions of the field may be superfluous. They tested this by reducing the size of the main survey iteratively by one and then using SSA to minimize the maximum prediction variance on the 1-m square prediction grid whilst constraining the observations to positions on the triangular grid with spacing $I(\boldsymbol{\theta})$. This algorithm used the variogram parameters c_0^{95-}, c_1^{95-} and a^{95-}, which were those with the largest likelihood and corresponded to a sampling interval equal to I^{95-}. In this way the number of main survey locations was reduced to 15. The resulting sampling scheme is shown in Fig. 3.7a. Furthermore,

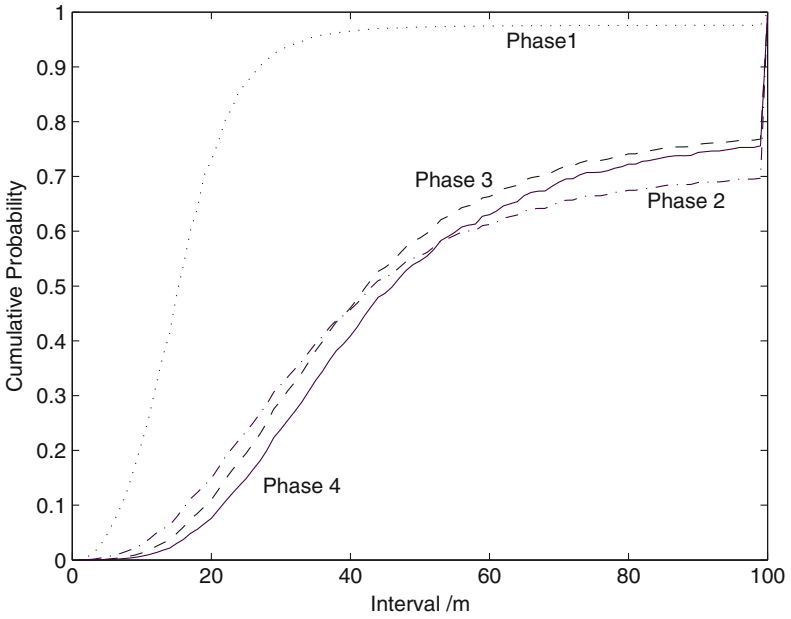

Fig. 3.6 The cdf of the optimal sampling interval after each phase of the Silsoe case study

Table 3.2 Results from the Silsoe case study after each phase of the exploratory survey

N_{E} [a]	I^{95-} [b]	N_{I}^{95-} [c]	$I\left(\hat{\boldsymbol{\theta}}\right)$ [d]	$N_{\mathrm{I}}^{\mathrm{R}}$ [e]
30	6.4	142	15.6	25
60	12.8	37	100.0	0
70	16.0	22	66.5	1
75	17.8	18	75.6	1

[a] N_{E} the number of exploratory survey observations.
[b] I^{95-} the lower 95% confidence limit on $I(\boldsymbol{\theta})$ /m.
[c] N_{I}^{95-} the number of observations on a triangular grid with interval I^{95-} required to cover the field.
[d] $I(\hat{\boldsymbol{\theta}})$ the REML estimate of $I(\boldsymbol{\theta})$ /m.
[e] $N_{\mathrm{I}}^{\mathrm{R}}$ the number of observations on a triangular grid with separation $I(\hat{\boldsymbol{\theta}})$ required to cover the field.

Marchant and Lark (2006) found that if the iterative SSA was repeated without constraining points to the triangular grid then a survey of four observations could complete the survey in a way that would ensure the prediction variance threshold is satisfied with 95% confidence. Thus the Bayesian adaptive approach was able to determine efficiently that relatively little sampling effort was required to predict soil moisture content with the required precision and prevented costly over-sampling.

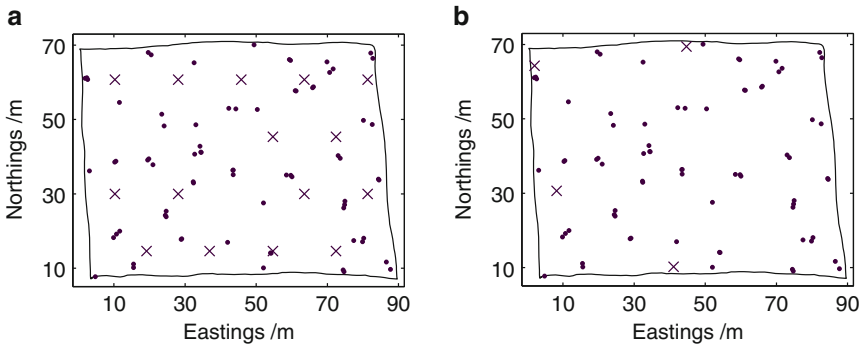

Fig. 3.7 The suggested sampling positions for the complete survey of moisture content at Silsoe. The *dots* denote the exploratory survey. The *crosses* in (**a**) denote the regular survey with superfluous locations removed and the *crosses* in (**b**) denote the irregular survey which ensures that the kriging variance threshold is not exceeded anywhere in the field. Coordinates in metres are relative to an origin at 508 980, 235 670 on the UK Ordnance Survey National Grid

3.5 Conclusions

In this chapter we have seen that we can optimize sampling schemes for geostatistical surveys when we present the problem in terms of a LMM with parameters estimated by REML. The major advantage of this approach over classical geostatistics is that such methods can approximate uncertainty in the survey due to both variogram parameter estimation and prediction. Thus we are able to account for this uncertainty when we optimize sampling schemes.

A major difficulty with optimizing geostatistical surveys is that the optimal survey depends on the variogram of the property being measured which is not known prior to sampling. If the target property is amenable to multi-phase sampling we have shown that a Bayesian adaptive sampling design can select the number and location of observations required efficiently to ensure with a pre-specified confidence (here 95%) that a threshold on the kriging variance is met. Often however multi-phase surveys of soil properties are not cost-effective. In such cases it is necessary to optimize the survey based upon an assumed variogram or an assumed distribution of the variogram parameters. Such surveys may be suboptimal but by using the techniques described in this chapter we can ensure that they are adequate for estimating the expected variogram and predicting the target property across the region.

This chapter has described the ‘state of the art’ for sampling design in the geostatistical context. Various questions remain for further development in the context of precision agriculture.

First, Chapter 2 suggests that ancillary variables commonly available in precision agriculture, such as apparent electrical conductivity of the soil, yield maps and remote sensor imagery of the crop canopy, can be used to guide grid-sampling of the soil by rules of thumb. It might be possible to use such ancillary variables more formally, perhaps to define narrower prior distributions for the parameters of soil

variograms than the uniform distributions used by Marchant and Lark (2006). More informative prior distributions should reduce the sampling effort in the exploratory phases, and so increase the overall efficiency of Bayesian adaptive sampling.

Second, all discussion of geostatistical sampling assumes that a target kriging variance for predictions may be easily specified. We believe that this requires closer attention in the context of precision agriculture. Increasing sample effort will reduce the kriging variance averaged over a field. At what point does the cost of extra effort match the benefit of improved predictions? This will depend on the form of a loss function that describes how the cost to the farmer of uncertainty about the value of a property at some location depends on the prediction variance of the property at that location. This will depend, in turn, on the costs associated with under- or over-application of an input at any location, and so on the costs of the input, crop responses, prices of the crop and the costs of externalities (such as pollution from excessive application of a fertilizer or pesticide). Such a sophisticated decision model is needed as a basis for the rational management of sampling in precision agriculture, and for the full exploitation of the power of geostatistical approaches to sampling and prediction.

Acknowledgements The work reported in this chapter was largely undertaken in a project funded by the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC), Grant 204/D15335, as part of an Industry Partnership Award with the Home-Grown Cereals Authority (Grant 2453). BPM and RML are funded by the BBSRC in the Mathematical and Computational Biology programme at Rothamsted Research. We are grateful to Miss Helen Wheeler and Mr Peter Richards for their contribution to the development and operation of the field system. Figures 3.1 and 3.3–3.7 are reproduced from Marchant and Lark (2006) by kind permission of Wiley-Blackwell. Figure 3.2 is reproduced from Marchant and Lark (2007) by kind permission of Springer Science and Business Media.

References

Aarts, E., & Korst, J. (1989). *Simulated annealing and Boltzmann machines – A stochastic approach to combinatorial optimization and neural computing.* New York: Wiley.

Adamchuk, V. I., Lund, E. D., Sethuramasamyraja, B., Morgan, M. T., Doberman, A., & Marx, D. B. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. *Computers and Electronics in Agriculture, 48*, 272–294.

Adsett, J. F., Thottan, J. A., & Sibley, K. J. (1999). Development of an automated on-the-go soil nitrate monitoring system. *Applied Engineering in Agriculture, 15,* 351–356.

Brus, D. J., & de Gruijter, J. J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma, 80*, 1–44.

Brus, D. J., & Heuvelink, G. B. M. (2007). Optimization of sample patterns for universal kriging of environmental variables. *Geoderma, 138*, 86–95.

Delta-T Devices. (1999). *Theta probe moisture sensor type ML2x user manual.* Cambridge: Delta-T Devices.

Demougeot-Renard, H., de Fouquet, C., & Fritsch, M. (2004). Optimizing sampling for acceptable accuracy levels on remediation and cost estimations: an iterative approach, based on geostatistical methods, illustrated on a former smelting works polluted with lead. In X. Sanchez Vila,

J. Carrera & J. J. Gomez Hernandez (Eds.), *GEOENV IV – Geostatistics for environmental applications: Proceedings* (pp. 283–294). *Book series: Quantitative geology and geostatistics 13.* Dordrecht: Kluwer.

Diggle, P. J., & Ribeiro Jr., P. J. (2007). *Model-based geostatistics.* New York: Springer.

Dobson, A. J. (1990). *An introduction to generalized linear models* (2nd ed.). London: Chapman and Hall.

Kirkpatrick, S., Gellat, C. D., & Vecchi, M. P. (1983). Optimisation by simulated annealing. *Science, 220*, 671–680.

Kitanidis, P. K. (1985). Minimum-variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association of Mathematical Geology, 17*, 195–208.

Kitanidis, P. K. (1987). Parametric estimation of covariances of regionalised variables. *Water Resources Bulletin, 23*, 557–567.

Lark, R. M. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma, 105*, 49–80.

Lark, R. M., & Papritz, A. (2003). Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma, 115*, 245–260.

Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science, 57*, 787–799.

Marchant, B. P., & Lark, R. M. (2004). Estimating variogram uncertainty. *Journal of the International Association of Mathematical Geology, 36*, 867–898.

Marchant, B. P., & Lark, R. M. (2006). Adaptive sampling for reconnaissance surveys for geostatistical mapping of the soil. *European Journal of Soil Science, 57*, 831–845.

Marchant, B. P., & Lark, R. M. (2007). Optimized sample schemes for geostatistical surveys. *Journal of the International Association of Mathematical Geology, 39*, 113–134.

McBratney, A. B., Webster, R., & Burgess, T. M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalised variables. *Computers and Geosciences, 7*, 331–334.

Pardo-Igúzquiza, E. (1998). Maximum likelihood estimation of spatial covariance parameters. *Journal of the International Association of Mathematical Geology, 30*, 95–108.

Pardo-Igúzquiza, E., & Dowd, P. A. (2003). Assessment of the uncertainty of spatial covariance parameters of soil properties and its use in applications. *Soil Science, 168*, 769–782.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58*, 545–554.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes (Fortran)* (2nd ed.). Cambridge: Cambridge University Press.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for Kriging.* New York: Springer.

Van Groenigen, J. W., Siderius, W., & Stein, A. (1999). Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma, 87*, 239–259.

Viscarra Rossel, R. A., Gilbertson, M., Thylen, L., Hansen, O., McVey, S., & McBratney, A. B. (2005). Field measurements of soil pH and lime requirement using an on-the-go soil pH and lime requirement measurement system. In J. V. Stafford (Ed.), *Precision agriculture 05, Proceedings of the 5th European conference on precision agriculture* (pp. 511–520). Wageningen: Wageningen Academic Publishers.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester: Wiley.

Zhu, Z., & Stein, M. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological and Environmental Statistics, 11*, 24–44.

Zimmerman, D. L., & Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics, 44*, 27–43.

Chapter 4
The Spatial Analysis of Yield Data

T.W. Griffin

Abstract Yield data are now recorded automatically for a wide variety of crops including cereal grains, oilseeds, fiber, forage, biomass, fruits and vegetables. Yield monitors for grain crops were developed in the 1980s and commercialized by 1992 before global positioning systems (GPS) had become fully operational for civilian use. Since the initial agricultural use of GPS, farmers have made intensive use of yield monitors. This chapter describes how yield data from monitors must be calibrated and how measurement errors can be addressed. Yield data can be used to target crop and soil investigations, nutrient applications and for on-farm experiments. Challenges concerning the use of yield data for decision making are considered; important amongst these are the difficulties associated with spatially correlated data for traditional statistical analyses and the alignment of data in different spatial layers. This Chapter uses spatial statistics rather than only geostatistics; the latter comes within this broader category. Some links between the different approaches in spatial analysis are illustrated. Spatial statistics lends itself better to econometrics than does geostatistics, and the importance of the economics of precision agriculture is discussed. The methods described are applied to a case study of soya bean data.

Keywords Economics · Farm management · On-farm experiments · Profitability · Site-specific · Spatial analysis · Spatial econometrics · Spatial statistics · Technology adoption · Yield data · Yield monitor

4.1 Introduction

Site-specific yield data have been collected from crops including cereal grains, oilseeds, fiber, forage, biomass, fruits and vegetables. Spatial variation in site-specific yield data often reflects the variation in factors affecting yield, which may

T.W. Griffin (✉)
Division of Agriculture, University of Arkansas, 2301 S University Avenue,
Little Rock, AR 72204, USA
e-mail: tgriffin@uaex.edu

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_4,

or may not be measured feasibly or managed effectively by cultural practices. Spatial variation in yield provides an indication of whether there are potential benefits in exploring the causes of variation with site-specific technology. In some cases the marginal benefit of addressing the spatial variation may not necessitate the investment of resources; in such cases the rational decision maker would avoid the costs of site-specific sampling and manage the field uniformly. Since yield data tend to be spatially variable and geographically dense, traditional methods of analysis based on assumptions such as independence are precluded from the spatial analyst's toolkit. In general, spatial statistical methods are appropriate for analysing site-specific yield data especially from on-farm experiments. Spatial statistical analyses provide a sound basis for exploring and understanding the causes of variation in yield, leading to objective analyses. Spatial technologies have also been used to examine yield quality, in particular the protein content of wheat and fiber quality of cotton. The focus of this chapter is on the spatial analysis of yield monitor data of field-scale on-farm experiments. The spatial statistical analyses discussed in this chapter are analogous to the geostatistical methods applied in the remainder of this book. The differences between these two techniques of objective analysis are discussed relative to nomenclature, development and estimation, and how each provides a sound basis for exploring and understanding the causes of variation in yield. The chapter culminates with a case study adapted from Griffin (2006).

4.2 Background of Site-Specific Yield Monitors

Precision agriculture technologies can be categorized into one of two broad categories; information-intensive and embodied-knowledge. Information-intensive technologies include those that provide more information, such as yield monitors, but at the cost of requiring additional ability in management to make practical use of the technology. Information-intensive technology has been readily adopted on farms, but not as quickly as embodied-knowledge technologies. The latter include, for example, global positioning system (GPS) automated guidance and automated spray boom controls; they require less management ability to make effective use of the technology.

Yield monitors have been used by farmers and researchers for the gamut of crops; however most of data on adoption has focused on grains, oilseeds and cotton. Precision agriculture technologies have spread rapidly around the world. In the United States 28% of maize and 22% of soya bean areas were harvested in 2005 and 2002, respectively, with a yield monitor (Griffin 2009a). Griffin and Lowenberg-DeBoer (2005) provided world-wide estimates of yield monitor adoption, comparing the USA, EU and Latin America by estimating the number of yield monitors per million arable hectares (Table 4.1). Germany is projected to have the highest density of yield monitors in the world (523 M ha^{-1}) followed by the United States (335 M ha^{-1}), Denmark (247 M ha^{-1}), Sweden (119 M ha^{-1}) and the United Kingdom (107 M ha^{-1}).

Table 4.1 Number of yield monitors by country

Country	Estimated number	Year of estimate	Yield monitors per million hectares
Americas			
United States	30 000	2000	335
Argentina	4500	2008	180
Brazil	250	2008	5
Chile	12	2000	19
Uruguay	15	2000	12
Europe			
U.K.	400	2000	107
Denmark	400	2000	247
France	50	2000	5
Germany	4250	2003	523
Netherlands	6	2000	27
Sweden	150	2000	119
Belgium	6	2000	17
Spain	5	2003	1
Portugal	4	2003	6
Other			
Australia	800	2000	42
South Africa	15	2000	3

Source: Adapted from Griffin and Lowenberg-DeBoer (2005).

It is estimated that more than 90% of yield monitors in Argentina are associated with a GPS (Instituto Nacional de Tecnología Agropecuaria 2008), whereas in the USA most are not (Table 4.2). It is likely that most yield monitors in the rest of the world have GPS, although the gap may not be this wide. Although the value of yield monitor data is greatly diminished without the georeferenced data, there are still some practical uses for it.

The National Research Council defines yield mapping as "the process of collecting georeferenced data on crop yield and characteristics, such as moisture content, while the crop is being harvested" (Sonka et al. 1997, p. 137). The commercialization of yield monitors has occurred at different times for each type of crop or harvester. The first widely commercialized yield monitors for the grain combine became available in 1992 (Griffin et al. 2004), over 2 years before GPS equipment was fully operational for civilian uses (United States Naval Observatory). The cotton yield monitor became commercially available in 1998, at a time when over 20% of USA maize and soya beans were harvested with yield monitors. It has been expected that the adoption of yield monitors would occur more quickly for higher valued crops that provide an opportunity to achieve greater net returns. Yield monitors are most often associated with grain harvesters because of the relatively higher adoption rates, and they were commercialized several years before yield monitors for cotton (Vellidis et al. 2003), grapes (Bramley and Williams 2001), sugar beet (Konstantinovic et al. 2007), tomatoes (Pelletier and Upadhyaya 1999), fruits (Alchanatis et al. 2007; Ampatzidis et al. 2009), forages (Kumhala et al. 2005;

Table 4.2 Percentage of planted hectares of crops on which yield monitor technologies were used from 1996 to 2005 in the USA

	Soya beans	Cotton	Barley	Durum wheat	Spring wheat	Winter wheat	Maize	Potatoes	Sunflower	Rice
Yield monitor without GPS										
1996	14			9	3	2				
1997	10			6	11	6	12			
1998	15	a		4	6	6	12			
1999	17	4				17	16	3	8	
2000	21	1		a	9	10	18			18
2001							19			
2002	22									
2003		2	13							
2004				16	14	10				
2005							28			
Yield monitor with GPS										
1996	3			a	a	a				
1997	4			a	a	1	5			
1998	6	a		a	1	a	3			
1999	6	a				7	6	3	a	
2000		a			a	3	6			6
2001							7			
2002	8									
2003		2	4							
2004				7	4	2				
2005										

[a]Less than 1%.
Adapted from Griffin (2009a) using USDA-ARMS data.

Maguire et al. 2003; Wild et al. 2003; Wild and Auernhammer 1999), citrus (Schueller et al. 1999), peanuts (Durrence et al. 1999; Vellidis et al. 2001), baling hay and other crops. Although yield monitors for grains (maize and wheat) have been discussed in the literature more frequently than those for other crops (Griffin et al. 2004), producers of higher value crops are relatively faster at adopting them for production decisions.

With the commercialization of yield monitors and global positioning system (GPS) equipment, many georeferenced yield observations can be recorded inexpensively. Data from yield monitors include yield measurements from mass flow or volumetric methods, and may also include data on moisture. The data from GPS provide location in terms of coordinates (latitude and longitude), time, elevation, and their derivatives such as speed and heading.

Site-specific yield data provide farmers and their advisors with additional metrics to examine the factors that affect yield such as drainage, soil management and planned on-farm experiments. These methods of analysis have included both subjective visual observations as well as objective quantitative analysis with spatial statistical methods. As with most other information technologies, the adoption of yield monitors is influenced by how many other farmers use the technology and this

is indirectly related to complementary services. Many potential users may not adopt the technology until there are sufficient complementary services, such as consultancy firms offering support for the hardware and or data analysis. Until a critical mass of complementary support and data analysis services exist, many farmers perceive no benefit in adopting yield monitor technology.

4.2.1 Concept of a Yield Monitor

In some parts of the World, yield is reported in volumetric units but crops are marketed on standard sized volumetric units typically expressed as weight. For instance in the USA, maize is sold by a volumetric unit (bushels), but is standardized to a '56 pound bushel' (25.4 kg). In reality maize is marketed in 25.4 kg increments rather than by the volumetric 'bushel' because grain density or test weight varies from field to field as well as within fields. Test weight is the density of grain and is reflected in the weight of the volumetric unit. Therefore, it is worth understanding the concept of a yield monitor and how calculations are made. For grain yield monitors, yield is expressed as in Eq. 4.1:

$$Y = K\frac{M}{W * S}, \tag{4.1}$$

where Y is instantaneous yield reported as volume per unit area, M is grain flow measured as mass per unit time, W is the harvester width (cutting width), S is travel velocity in distance per unit time and K is a conversion coefficient. Grain harvesters typically use physical sensors to measure grain flow, whereas remotely sensed data have been somewhat successful for monitoring yield in biomass crops such as forages, vegetables and fruits. Cotton monitors use microwave or near-infrared sensors to measure volume.

Variation in test weight has an impact on volumetric units because the yield monitor measures mass per unit time and farmers market crops based on weight. However, test weight conversion issues are of no consequence to subsequent farm management decisions.

Cutting width can be set by the operator or measured with a sensor. Software has been developed to detect the effective swath width based upon whether the harvester has already passed over a particular area. Ultrasonic sensors are available for measuring from the outside of the harvester width.

Travel velocity or ground speed can be measured by radar, ultrasonic measures, a wheel or driveshaft rotation counter or GPS. In many cases, rotation counters suffer from inaccuracies because of slippage and tyre distortion under varying loads. Speed derived from GPS location also has limitations including momentary lack of differential correction and the relatively lower refresh rate than is needed by the velocity sensor. The preferred method of measuring travel velocity is with ultrasonic speed sensors, which do not have the limitations of other methods.

It should be noted that yield monitor data are recorded as points and are then converted to aerial units rather than measuring the yield per areal unit as is the common perception. The distinction between the two ideas is important when considering methods of spatial analysis.

4.2.2 Calibration and Errors

Although calibration has been covered *ad nauseum*, it deserves to be reemphasized because of its importance in relation to recording yield data that are suitable to make the best farm management decisions possible. Even with a properly calibrated yield monitor, several sources of potential error remain. The exact crop width entering the harvester is likely to be unknown, except under the best case scenario. Another source of error is the time lag of grain flowing from the header through the threshing mechanism to the measurement area, especially for grain combines. The accuracy of sensors depends on calibration and regular inspection. Excess dirt, plant residue or moisture can result in inaccurate measurements from yield and moisture sensors. Another source of error in yield monitor data originates from operator behaviour or error. Changes in travel velocity is a leading cause of error in grain combines. Although yield observations recorded while the harvester is changing velocity can be removed using the 'smooth velocity' parameter setting of Yield Editor (Sudduth and Drummond 2007), data are lost. Yield Editor can be downloaded free from the USDA-ARS and is useful for researchers, industry and farmers to make the most of yield monitor data. It is recommended that the harvester is operated as consistently as possible to maintain a constant mass flow rate of grain, such as was entering the harvester during calibration (Grisso et al. 2002; Shearer et al. 1999; Blackmore and Marshall 1996).

Calibration of the harvester has been recommended at the beginning of the harvest season for each type of grain harvested (Grisso et al. 2002, p. 4). Grisso et al. go on to say that recalibration may be needed as crop conditions, such as moisture levels, change. Four elements are typically evaluated when calibrating a yield monitor: distance, header height, mass flow rate and moisture content (Grisso et al. 2002; Shearer et al. 1999; Blackmore and Marshall 1996). It is useful to understand what the data are from the yield monitor.

> Shearer et al. (p. 8–9) stated it appropriately as follows:
>
> "The following 'Advanced' export data format is nearly universal:
>
> ddd.dddddd,dd.dddddd,mm.mm,tttttttt,n,lll,www,cc.c,kk,ppppp,ssssss,Fnn:bb
>
> bbbbbb,Lnn:bbbbbbbb,gggggggggg,sss, ppp,aaaa",

where **ddd.dddddd** = longitude (degrees, + East and – West), **dd.dddddd** = latitude (degrees, + North and – South), **mm.mm** = grain mass flow (mg per second), **tttttttt** = GPS time (seconds), **n** = cycle period (seconds), **lll** = distance travelled in cycle period (cm), **www** = effective swath width (cm), **cc.c** = moisture

content (percent wet basis), **kk** = status (bits 0 to 4 – header down), **ppppp** = pass number, **ssssss** = yield monitor serial number, **Fnn:bbbbbbbb** = field ID (number and name), **Lnn.bbbbbbbb** = load ID (number and name), **ggggggggggg** = grain type, **sss** = GPS status, **ppp** = point dilution of precision (PDOP) and **aaaa** = altitude (m).

This 'advanced' export data format is the same as that required by Yield Editor, although other data formats can be used. Agricultural economists use yield data as the basis for economic analyses and also use quality data for some crops such as wheat, cotton and fruits where there is a difference in value between categories of quality. Assessment of the profitability of precision agricultural technologies usually includes data from instantaneous yield monitors.

4.2.3 Common Uses of Yield Monitor Data

Yield monitors were initially commercialized to document yields although farmers have extended the frontier for how the data can be used to its potential. The most common uses of grain yield data from monitors with a GPS are to monitor crop moisture, conduct on-farm experiments and tile drainage management (Table 4.3) (Griffin 2009b). In some areas, crop nutrient applications are based on crop removal using yield data. Monitoring crop moisture at harvest is important for deciding which storage facility to use. Another example of harvest logistics is the filling of cotton modules by tracking overall volume of cotton in the harvester. Seed cotton is stored temporarily in the field in a semi-compressed 'module' that holds a certain volume of cotton. Cotton yield monitor data give information to the cotton

Table 4.3 Use of yield monitor data for selected crops with and without GPS, 2002–2005

Crop year sampled	Soya beans (2002)		Cotton (2003)		Barley (2003)		Durum wheat (2004)		Spring wheat (2004)		Winter wheat (2004)		Maize (2005)	
With GPS?	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Monitor crop moisture	68	86	a	a	68	67	100	52	60	63	60	85	91	83
Document yields	50	40	25	41	76	38	69	65	54	37	41	29	51	30
Conduct field experiments	42	23	37	a	32	5	a	13	53	9	14	9	46	28
Tile drainage	32	8	5	3	6	6	a	a	7	a	32	2	31	7
Negotiate new crop lease	9	1	1	3	5	a	53	a	21	a	a	1	5	2
Divide crop production	6	7	7	54	12	11	a	48	a	3	7	8	12	11
Irrigation	4	a	4	8	24	3	a	a	a	a	a	a	4	3
Other uses	7	13	1	19	15	8	53	a	6	20	a	7	7	5

[a]Less than 1%.

Adapted from Griffin (2009b) using USDA-ARMS data.

picker operator to help decide whether to continue with the current module, or to complete the module and begin another. Contrary to early anticipation from farm managers, negotiating leases, apportioning crop shares and managing irrigation were not leading uses of yield monitor data.

One common use of georeferenced yield data, especially for initial use after adopting the technology, is to guide farmers and their advisors to specific locations within fields based on relative differences in yield, i.e. directed scouting. At these sites, they evaluate crop nutrient levels that correspond to relatively small and large yields. In practice, many areas with small nutrient concentrations are associated with larger yields than areas with large concentrations where the factor(s) limiting yield is other than nutrient availability. Thus there is a probable inverse relationship between crop yield and fertility level. Under conventional uniform rate applications, fertilizer applied to areas where yield limiting factors are severe accumulate nutrients because of limited nutrient uptake.

4.2.4 Profitability of Yield Monitors

The profitability of yield monitors depends on if, when and how the technology is used. Studies on profitability should focus on how farm management decisions can be improved with yield data, whether by quantifying the effects of poor drainage or by conducting on-farm experiments to decide which inputs or production system should be used in subsequent years. Assigning a value directly to yield monitor technology is difficult. The value of farm management decisions attributed to the yield monitor can be partly assessed, whereas ownership costs can be readily estimated by summing straight-line depreciation (Eq. 4.2) and interest expense costs (Eq. 4.3). Straight-line depreciation is a frequently used method to estimate annual depreciation of a fixed asset (Kay et al. 2008).

$$\text{Depreciation} = \frac{\text{cost} - \text{salvage}}{\text{ownership life}}, \tag{4.2}$$

$$\text{Interest expense} = \frac{\text{cost} + \text{salvage}}{2} * \text{interest rate}. \tag{4.3}$$

Two variables that deserve further mention are salvage value and ownership life. The salvage value of spatial technology is often assumed to be zero. The debate on ownership life has focused on whether it should be similar to the useful life of the machinery or closer to that of computer technology. In addition to ownership costs, taxes, insurance and operating costs such as repairs and maintenance should be considered.

In addition to ownership and operating costs, costs of yield monitor use need to be understood. The latter are primarily a function of calibration costs that can best be described as a classic downtime model during harvest (Griffin et al. 2009). Proper calibration requires time that could be used to harvest crops; if this causes

Table 4.4 Costs of delayed harvest resulting from yield monitor calibration

Time (in days) to calibrate yield monitor during October 11–31 time period		
0	0.5	1.0
$0	$859	$1818

Adapted from Griffin et al. (2009).

the harvest to be delayed, yield penalties may occur. It was estimated that diverting harvest equipment away from field operations for half a day reduced whole-farm returns by nearly $900 USD (Table 4.4). It is plausible that two calibration events could require half a day of good harvest days.

4.2.5 Quantity and Quality of Product

In addition to recording site-specific yield quantity, it is also possible to obtain data on quality. Grain protein sensors based on near-infrared reflectance have been used to estimate protein content of wheat (Stewart et al. 2002; Taylor et al. 2005). Although instantaneous on-the-go quality sensors have not yet been commercialized for cotton, it is possible to measure cotton lint quality at within-field locations using identity tracking and radio-frequency identification (RFID) (Griffin and Slinsky 2008). It is also expected that RFID technologies will enable georeferenceing of quantity and quality for crops such as tomatoes and fruits.

4.3 Managing Yield Monitor Data

Applied researchers and consultants have to deal with yield monitor data from a spectrum of yield monitor types, manufacturers and file formats. This is problematic when the analyst requires access to the original unprocessed data, e.g. in 'Advanced' export format, in order to have more control over which observations are excluded. To prevent maintaining multiple software packages to import, view and export the raw data, the use of Field Operation Data Models (FODM) using FOViewer (Field Operation Viewer, MapShots, Cummins, GA, USA) allows users to handle yield data from different file formats.

4.3.1 Quality of Yield Monitor Data

Yield monitors have provided farmers and researchers with many georeferenced yield observations at relatively lower costs compared to previous techniques. Successful use of such data, however, typically requires more management ability

to convert the data into information that is suitable for making farm management decisions. Erroneous data from yield monitors result when some measurements are made where the monitor cannot measure yield correctly (Noack et al. 2005; Simbahan et al. 2004). For grain and cotton yield monitors, post-processing of the data is possible with mapping or yield monitor software. Data analysts, however, should be cautious when accepting default post-processing parameters imposed by farm-level mapping software by understanding the parameters and how the default settings affect the quality of the data. The post-processing of the 'Advanced' export format data can be subjected to an objective correction procedure using Yield Editor. Data preparation procedures, referred to as 'data cleaning' or 'data filtering', do little more than correct the location of observations and remove measurements that are known to be erroneous because of harvester machine dynamics or operator behavior and are not an unethical modification or manipulation of the data.

Standard parameter settings can be imposed to filter yield data, but this is not recommended. The analyst's intuition, experience, prior knowledge and skill should guide the setting of parameters. No single parameter setting is likely to be universally appropriate, even with the same harvester and operator. The Yield Editor user interface includes a map of the data points so that the analyst can assess the data visually, albeit subjectively. The final Yield Editor parameter settings may or may not be farmer or field specific and can vary between fields for the same harvester and operator. Adjusting flow delay, start pass delay and end pass delay are the most difficult and may be the most important to the quality of the data for spatial statistical analyses. Since the grain yield monitor requires up to 30 m of harvester travel and crop intake to make accurate measurements, start and end pass delays arise from 'ramping up and down' of the harvester at the beginning and end of rows (Arslan and Colvin 2002).

Parameters do not have to be positive, negative values are sometimes correct when the data have been previously subjected to processing by the farm-level mapping software. Setting the flow delay is easiest when the operator harvests three to eight passes in one direction and alternates the pattern across the field (see Fig. 4.1 for an example of five adjacent passes before alternating the direction). This allows a visual reference wide enough to be seen on the Yield Editor map. Although the most dramatic differences are the start and end pass delays, within-field variability is useful to assess visually the proper flow delay parameter values. Alternating direction between individual passes may not provide the required visual reference unless the field has some distinct variation such as rice levees, tramlines or centre pivot irrigation (see Fig. 4.2 for an example of how irrigation system tracks can assist in adjusting flow delays).

It is assumed that filtering improves the quality of the data. Yield monitor data filtered consciously with Yield Editor can lead to different production recommendations from those based on data processed with the default correction parameters. Griffin et al. (2008) analysed seven field-scale on-farm experiments conducted by farmers and concluded that five experiments would have led to different farm management recommendations depending upon whether the yield data were consciously corrected.

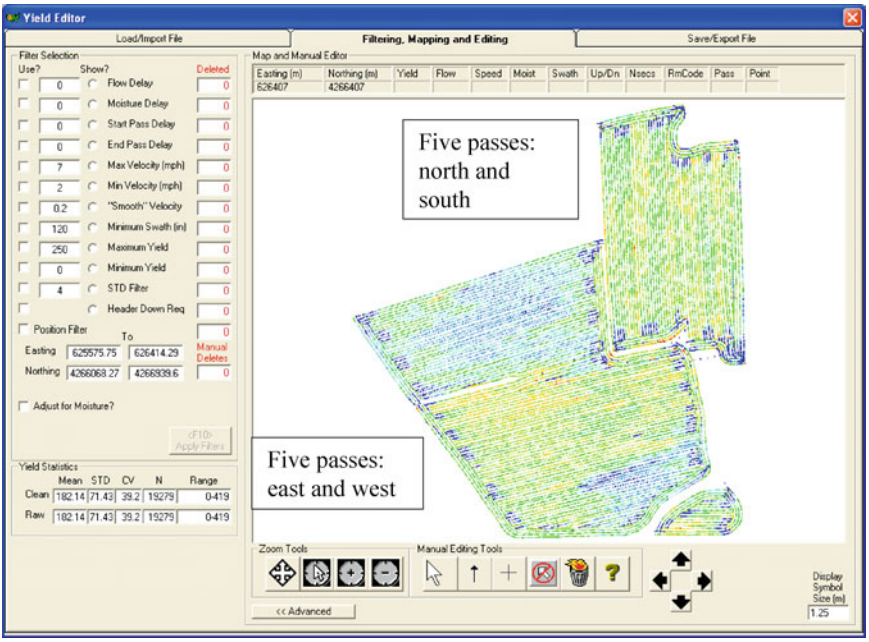

Fig. 4.1 Yield Editor screen capture and example of a suggested harvest pattern

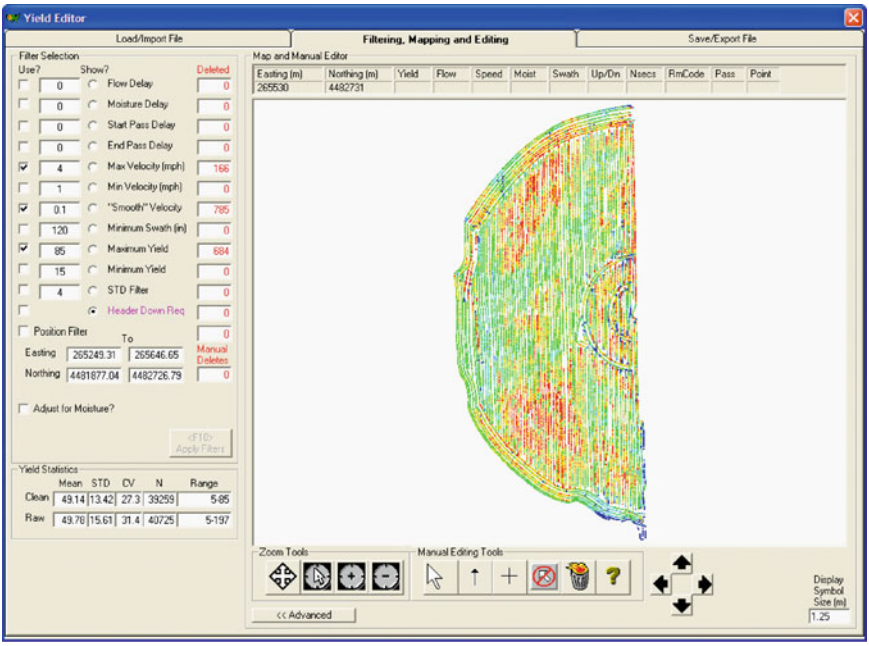

Fig. 4.2 Yield Editor screen capture and centre pivot irrigation system tracks

4.3.2 Challenges in the Use of Yield Data for Decision Making

Precision agriculture data tend to be dense for yield and sparse for soil or other factors affecting yield. Statistically significant differences can be estimated for yield data because of the large number of data even though the true difference is arbitrarily small or may not exist in reality (Hicks et al. 1997). The analyst must be aware of the distinction between statistical significance and practical significance (Ziliak and McCloskey 2008).

Spatially intensive data tend to result in spatially correlated residuals for ordinary least squares (OLS) regression (Bell and Bockstael 2000, p. 72; Bockstael 1996). Additional aspatial problems arise from measurement errors in attributes and location. Inherent spatial autocorrelation and spatial heterogeneity of intensive data can be disadvantageous for traditional statistical analyses. Data with spatial effects contain less information than independent observations of the same sample size (Anselin 1989; Florax and Nijkamp 2004). Anselin states that "the loss of information that results from the dependence in the observation should be accounted for" (1989, p. 68). Therefore, spatial statistical analyses are needed to model the spatial structure of the data explicitly to improve inference based on the site-specific data.

4.3.3 Aligning Spatially Disparate Spatial Data Layers

Before statistical analysis of precision agriculture data, individual data points from different attribute layers must align, especially when analysing data from several years. A common grid (or other system of location) must be imposed on the data for spatio-temporal analysis. Several methods have been used in research and in commercial practice to remedy disparate spatial data layers and the modifiable areal unit problem.

To integrate disparate data layers, observations from the denser data can be averaged within the neighbourhood of locations of the most sparse data. The averaged values then have the coordinates of the sparse data locations. Half the average distance between adjacent sampling points can be used as the radius of the neighbourhood. Values may be estimated with Thiessen polygons or alternatively interpolation by kriging or inverse squared distance is often used to ensure that all data layers align even though yield monitor data are more dense than other site-specific data. If interpolation or Thiessen polygons are used to predict all values in different data layers to a common grid a systematic error with spatial structure, i.e. error increases in proportion to increases in distance to the measured location, is introduced into the predictions (Anselin 2001). When the interpolations are fewer, or the original number of observations of data are based on multiple original measurements, then the systematic errors (Anselin 2001) may be at a reasonable level. Although there are more cautions than solutions, the problem of disparate spatial data layers must be addressed in the analysis of field-scale on-farm experiments.

When yield data, which tend to be spatially dense and variable, are used for farm management decisions, they are suitable for rigorous geo-spatial statistical analysis. Such analyses provide a sound basis for exploring and understanding the causes of variation in yield and how to address them.

4.4 Spatial Statistical Analysis of Yield Monitor Data

A major use of yield monitor data is to evaluate planned on-farm experiments (Griffin 2009b). The most current and rigorous analyses of data from on-farm experiments use advanced spatial statistics to account for spatial effects such as autocorrelation and heteroscedasticity. These spatial econometric techniques have been used with site-specific data from yield monitors, small-plot harvest equipment (Lambert et al. 2006), as well as hand harvested crops such as citrus and subsistence millet production (Florax et al. 2002).

Spatial statistical analysis of yield monitor data has become more common than for hand-harvested crops. Four studies provide examples of the appropriate use of spatial statistical methods. The most influential field experiment for spatial analysis is at Los Rosas in Argentina (see Anselin et al. 2004; Bongiovanni 2002; Lambert et al. 2004), which evaluated the effects of nitrogen rates on maize. Parameters estimated from the Los Rosas study were used in Monte Carlo simulations by Griffin (2006) to evaluate experimental designs and spatial statistical methods for fields with different levels of spatial autocorrelation. Ntsikane Maine (see Maine et al. 2007) has led the way in the use of spatial statistics in South Africa for her evaluation of variable rates of phosphorus for maize production. In Europe, Andreas Meyer-Aurich (see Meyer-Aurich et al. 2008) has been prominent in field-scale evaluation of site-specific fertility applications using spatial econometric methods. In the USA, Griffin et al. (2008) collaborated with farmers to analyse yield monitor data from farmers' field-scale experiments for crops such as corn, soya bean, popcorn, rice and cotton.

4.4.1 Explicit Modelling of Spatial Effects

Two general approaches of addressing spatial autocorrelation are the simultaneous approach, which models spatial autocorrelation explicitly with spatial process models, and the conditional approach in which observations are removed until there is no spatial autocorrelation in the data. This distance at which observations are no longer spatially autocorrelated is known as the spatial range and can be determined from the correlogram or variogram. At distances greater than the spatial range, data can be analysed with aspatial models although information is lost by omitting observations.

Spatial statistics were developed to deal explicitly with spatial effects. Several spatial regression methods are available (Anselin 1988; Cressie 1993), and some

have been available for many years (Papadakis 1937). There are two broad groups of spatial statistical methods: the continuous geostatistical one (Cressie 1993) and spatial process approaches (Anselin 1988). The geostatistical approach relies on data to identify spatial characteristics by examining pair-wise observations (Cressie 1993). The experimental variogram is based on the squared differences between all pairs of data within a given distance or lag apart (see Section 1.2.2). The variogram function describes how the differences between values change with increasing separation. Spatial process methods rely on theory and an exogenously chosen specification of a neighbourhood structure (Anselin 1988).

Dubin (2003) compared spatial statistical methods and showed that geostatistical methods outperformed spatial process models when the true form of spatial variation was unknown. Lambert et al. (2004) compared ordinary least squares regression and four spatial regression methods on the Los Rosas data (see Bongiovanni 2002 or Anselin et al. 2004 for details of the field study). Their results indicated that all four spatial regression methods provided similar estimates, but the spatial process and geostatistical techniques modelled the treatment effects better than aspatial regression methods. The spatial process model has the advantages of being computed in one step, it requires fewer observations than the variogram and can model spatial autocorrelation in the dependent variable, error term or explanatory variables. However, a disadvantage is that the spatial interaction structure may be based on prior knowledge or a priori theoretical assumptions. Geostatistics is familiar to many agronomists and can be implemented with SAS software (Lambert et al. 2004), whereas the spatial process models are less well known.

Site-specific yield monitor data, as with most other agricultural data obtained at high resolutions, are expected to be spatially structured, i.e. autocorrelated and heteroscedastic, which violates the assumptions of classical statistics such as independence of observations and homoscedastic error terms. To correct for spatial effects in the residuals from a linear model estimated by OLS, methods that adjust for spatial dependence and will give more accurate estimates should be chosen. Anselin (1988) suggests two spatial process models: the spatial error model (Eqs. 4.4 and 4.5) and the spatial lag model (Eqs. 4.6 and 4.7). The spatial error model is given as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} = \lambda\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu}, \tag{4.4}$$

or in reduced form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\mu}, \tag{4.5}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of dependent variables, $\mathbf{X}$ is an $n \times k$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of residuals, λ is a spatial autoregressive parameter, $\mathbf{W}$ is an $n \times n$ spatial weights matrix and μ is a non-heteroscedastic uncorrelated (well behaved) error term (Anselin 1988). When the spatial autoregressive term, λ, is 0, the spatial error model reverts to the aspatial model. The spatial error process can be characterized by the autoregressive (AR) or the moving average (MA) error process resulting in global and local spatial externalities or 'spillovers', respectively. When the spatial

error model is appropriate, OLS estimates are unbiased but inefficient. A positive value for λ implies that neighbouring values exert an influence on a particular location's outcome through some unobserved or otherwise unmeasured factor; and in this case the structure of the unmeasured factor is 'absorbed' in the error term. In the familiar linear model, the variation in the dependent variable that is not explained by the explanatory variables is included in the error term, $\boldsymbol{\varepsilon}$. When the unexplained portion of the dependent variable has a spatial structure, then the residuals absorbed in the error term will have spatial structure.

The spatial lag model is given as

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}, \tag{4.6}$$

or in reduced form

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}], \tag{4.7}$$

where ρ is the spatial autoregressive parameter and the others are as above (Anselin 1988). As above, the spatial lag model reverts to the aspatial model when the spatial autoregressive term, ρ, is 0. A positive value for ρ implies that a location's observed value relates to those of its neighbours, i.e. spatially dependent. Spatial lags result in global spatial externalities and have a substantive economic interpretation. These models are sensitive to localized interventions that affect the whole system through the spatial multiplier, $(I - \rho W)^{-1}$. The OLS estimator is inconsistent for purely spatial autoregressive processes (Lee 2002), i.e. the unbiasedness of the estimate does not improve as sample sizes increase.

Both of the above models have been used with precision agriculture data. Anselin et al. (2004), Lambert et al. (2004) and Griffin et al. (2008) used the spatial error process model in their analyses, whereas Florax et al. (2002) used the spatial lag process model. Theory and a priori information suggest that when yield is the dependent variable, spatially autocorrelated error terms are expected rather than the contagion existing in the dependent variables. This suggests that the spatial analyst would opt to use spatial error process models to address the spatial effects explicitly. To determine empirically which model is more appropriate, spatial diagnostics such as Moran's I and Lagrange Multiplier tests of the OLS residuals provide insight into the underlying contagion; however, the spatial interaction structure must be specified first.

4.4.2 Spatial Interaction Structure

To model spatial autocorrelation explicitly in a spatial process model, the spatial interaction structure is defined separately in a spatial weights matrix, **W**, such that $w_{ij} > 0$ for neighbours and $w_{ij} = 0$ for non-neighbours where ij denotes location. In general, the connectedness, or 'relations' among observations specified by **W** influence the estimation and inference of the model. Under-specification

of **W**, i.e. limited connectedness, causes more errors in estimation than over-specification, i.e. more connectedness than is appropriate (Florax and Rey 1995). Bell and Bockstael (2000) add that the specification of **W** affects parameter estimates more than the choice of estimator. In practice, when the spatial weights matrix is under-specified (i.e. spatial connectedness is too limited), spatial diagnostics may inappropriately suggest the spatial process model, but when it is properly or over-specified the spatial error process model tends to be identified as more appropriate.

The connectedness of **W** is influenced by the matrix form (e.g. binary contiguity, binary distance, k-nearest neighbours, inverse distance) and the relevant distance measure (e.g. first- or higher-order contiguity, distance, number of neighbours). Inverse distance weights matrices are often chosen a priori for precision agriculture data because they imply a smooth distance decay of spatial correlation, and the relevant distance band can be determined empirically. Distance bands are based on specified distances from a central location, similar to the layers of an onion.

4.4.3 Empirical Determination of Spatial Neighbourhood Structure

Spatial autocorrelation is tested for in exploratory spatial data analysis and can be useful for interpretation in intermediate analyses. Three distinct yet analogous techniques based on spatial econometrics and geostatistics for determining empirically the spatial range of data are compared. The techniques include the spatial correlogram based on Moran's I test statistic, Lagrange Multiplier test and variogram. The de facto method of determining the spatial range has been the variogram; however, Griffin (2006) used correlograms for his seven studies to determine the limit of the spatial interaction structure.

The Moran's I test statistic (Eq. 4.8) tests for global spatial autocorrelation in a random variable (Anselin 1988; Cliff and Ord 1981) and is given by:

$$I = \frac{n}{S_o} \frac{\mathbf{x}'\mathbf{W}\mathbf{x}}{\mathbf{x}'\mathbf{x}}, \tag{4.8}$$

where **x** is an $n \times 1$ vector of deviations from the mean, **W** is an $n \times n$ spatial weights matrix as before and S_o is the sum of the elements of **W** (Anselin 1988; Cliff and Ord 1981). If Moran's I is positive, neighbouring values are interpreted as being large (small), if it is negative neighbouring values are both large and small, and if zero the distribution is spatially random. A local indicator of spatial association (LISA) (Anselin 1995), or the so-called local Moran's I, tests for local spatial autocorrelation. Moran's I test statistic of OLS residuals indicates whether estimation could be improved by correcting for spatial structure in the data.

The Lagrange Multiplier (LM) test can be used to determine whether the spatial error process or spatial lag process model should be used. Four LM tests for spatial

autocorrelation diagnostics are commonly used on OLS residuals; they include LM error (LM_λ), robust LM error (LM_λ^*), LM lag (LM_ρ) and robust LM lag (LM_ρ^*). The LM_λ and LM_ρ tests are unidirectional tests with the spatial error and lag models as alternative hypotheses, respectively. The LM_λ^* and LM_ρ^* tests take into account the potential presence of spatial lag and spatially correlated residuals, respectively, i.e. they take into account spatial autocorrelation of the other spatial process form. The LM_λ, LM_λ^*, LM_ρ and LM_ρ^* all follow an asymptotic χ_1^2 distribution.

The LM test with the largest χ^2 value, or alternatively the smallest p-value, between LM_λ and LM_ρ indicates whether the spatial error or lag model is appropriate. The model parameters of the correlogram or variogram computed on OLS residuals from the 'full' model define the spatial range (i.e neighbourhood) for each dataset. Spatial correlograms of Moran's I (Fig. 4.3) or LM test χ^2 values (Fig. 4.4) for autocorrelation in OLS residuals indicate an appropriate distance band over which to determine **W**.

Figure 4.4 shows the LM χ^2 coefficients for each iteration of contiguity orders of a 'queen' weights matrix. Contiguity weights are defined based on the direction

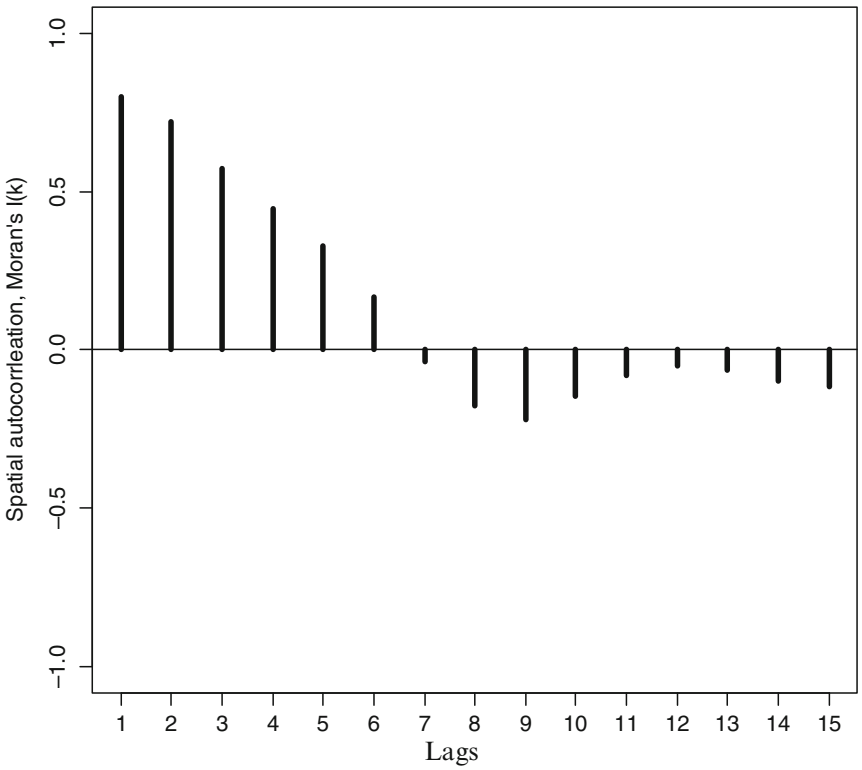

Fig. 4.3 Spatial correlogram of OLS residuals, Moran's $I(k)$ over k contiguity orders, estimated using the spcorrelogram function of spdep (Bivand 2009) contributed package to R (R Development Core Team 2009)

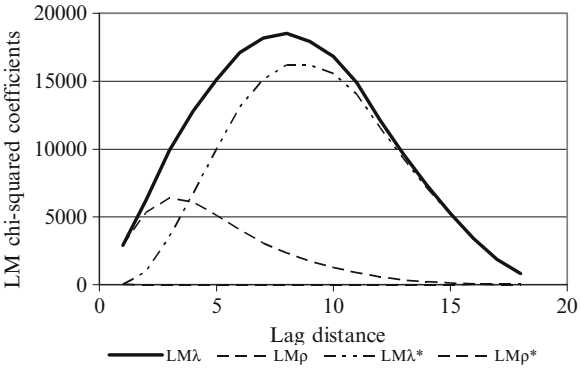

Fig. 4.4 Conceptual relationship between Lagrange Multiplier χ^2 values over contiguity orders for a random variable, values estimated using lm.LMtests function of spdep (Bivand 2009) contributed package to R (R Development Core Team 2009)

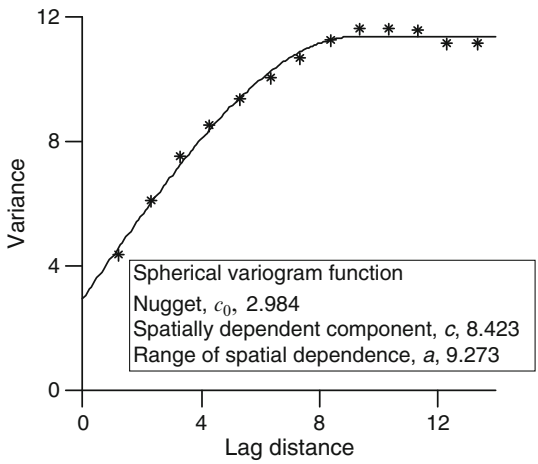

Fig. 4.5 Experimental variogram (*symbols*) of OLS residuals and spherical model (*solid line*) fitted by weighted least squares

of their connectedness in relation to chess pieces; queen contiguity has neighbours in all directions, rook has north–south or east–west neighbours only and bishop has neighbours only in northeast–southwest and northwest–southeast directions. Correlograms can be computed for lag intervals as well as contiguity orders. The general shape and relationship of the function in Fig. 4.4 is that expected for site-specific yield monitor data. For the LM tests, χ^2, the appropriate spatial range is determined by the distance at which the test with the largest coefficient reaches its maximum.

The spatial econometric analyses discussed in this chapter are analogous to the geostatistical methods applied in the remainder of this book; however there are some differences in nomenclature and development. An experimental variogram computed on the same residuals as those used for the correlogram (Fig. 4.3) and Lagrange Multiplier χ^2 values (Fig. 4.4) is shown by the symbols in Fig. 4.5. The variogram model was fitted by weighted least squares approximation (WLS).

The spatial ranges were estimated by spatial correlograms, LM tests and variograms. The spatial range is determined by identifying the distance at which the Moran's I test statistic goes to zero, i.e. lag 7, the peak of the LM test with the largest value, i.e. lag 9, and where the variogram reaches its sill, i.e. lag 9.3. The ranges determined by the three methods are similar (Figs. 4.3–4.5). The three metrics (LM test, spatial correlogram and variogram) evaluate the residuals of the full regression model estimated by OLS. The spatial structure of the correlogram and LM tests rely on the exogenously chosen **W**, whereas the variogram uses a pairwise comparison of all data pairs at a series of separating distances. The spatial econometric technique is used in the following example.

4.5 Case Study: Spatial Analysis of Yield Monitor Data from a Field-Scale Experiment

The data in this study include spatially correlated yield monitor data, treatments and elevation; they have been adapted from Griffin (2006). Treatments of 198, 247, 297, 346 and 395 thousand seeds per hectare were planted in a strip-trial design; the strip was 530 m long about 18.3 m wide with 24 rows 0.762 m apart (Fig. 4.6). This study is pertinent because seed costs are becoming a larger proportion of variable costs, there are changes in soya bean price to seed cost ratio and in cultural practices, specifically weed control.

4.5.1 Case Study Data

The yield monitor data were filtered using Yield Editor; Table 4.5 gives the parameters for filtering and the number of observations deleted. These parameters were based on prior experience and trial and error with these data. The resulting 3897 yield data are summarized in Table 4.6 by soil mapping unit and seeding rate treatment, and in Fig. 4.6.

Before analysing the yield, it is important to have data that might explain the spatial variation in the field. Soya bean seeding rates were assigned to each yield observation to indicate which rate was represented by the yield measurement. For each of the three soil mapping units in the field (USDA-NRCS soil map) (Fig. 4.6), a binary variable was assigned to the yield observation. The number 1 was assigned if the yield observation was from the particular soil and 0 otherwise. The soil series were: Soil A, a silt loam (Fine-silty, mixed, superactive, mesic Aeric Endoaqualfs; Fine, mixed, active, mesic Aeric Epiaqualfs), Soil B, a silt loam (Fine-silty, mixed, superactive, mesic Aeric Epiaqualfs; Fine-loamy, mixed, active, mesic Oxyaquic Hapludalfs) and Soil C, a silt loam (Fine-loamy, mixed, active, mesic Typic Hapludalfs; Fine-loamy, mixed, active, mesic Typic Hapludalfs). Soil A was the predominant soil series followed by Soil B, representing 68.3% and 28.1%

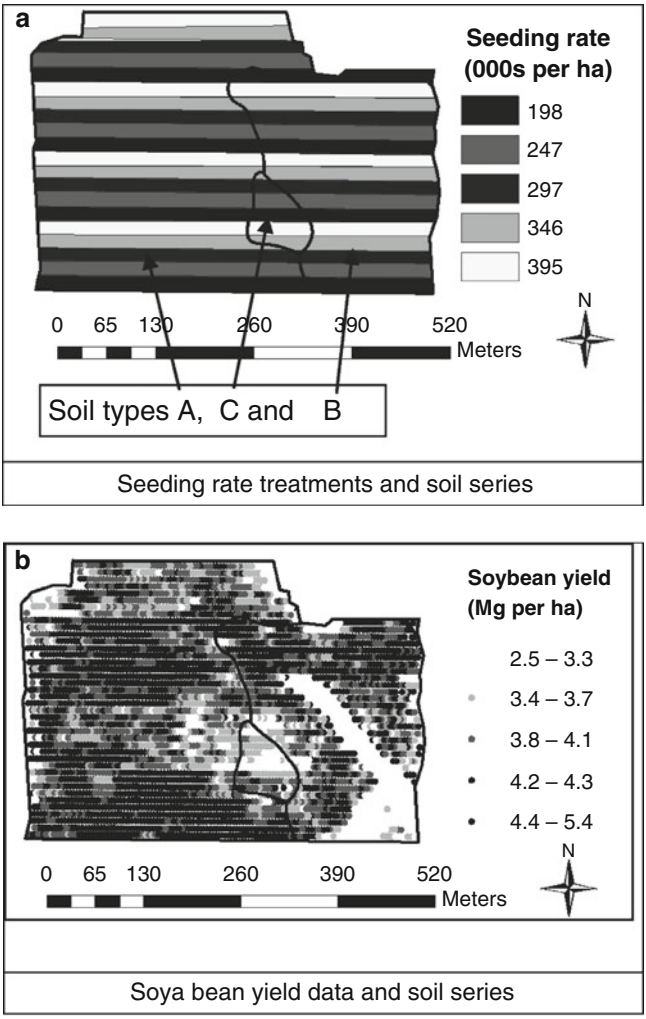

Fig. 4.6 Maps from the 2004 soya bean seeding rate study

of observations, respectively. Soil C was a minor component with only 3.6% of the observations and was expected to have a lower yield potential than the rest of the field. Since previous research indicated that small portions of a field may influence whole field profitability, Soil C was analysed thoroughly.

Elevation data were obtained from the real-time kinetic (RTK) GPS measurements of the combine harvester (Fig. 4.6 and Table 4.6) and were used as an explanatory variable. Elevation data provide information that relates to depth of topsoil, such as eroded hilltops, and relative position in the terrain. In addition to elevation, the square of elevation was used to determine if yield responses differed over the range of observed elevation. Relative elevation was calculated by identifying the

Table 4.5 Parameter values and number of yield data points deleted by filtering

Filtering parameter	Parameter value	Number of points deleted[a]
Maximum velocity (kph)	8.5	8
Minimum velocity (kph)	6.4	1802
Smooth velocity	0.2	408
Minimum swath (m)	0	0
Maximum yield (Mg ha^{-1})	5	49
Minimum yield (Mg ha^{-1})	0	0
Standard deviation filter	4	1524
[b]Flow delay (s)	3	561
[b]Start pass delay (s)	4	745
[b]End pass delay (s)	0	0

[a]Points deleted are not cumulative, i.e. the 'same' point can be deleted by several criteria.
[b]Flow delay, start and end pass delays of 12, 4, 4 were selected when importing data into farm-level mapping software by the farmer with additional filtering set as above.

Table 4.6 Descriptive statistics of selected continuous variables

	Yield mean (Mg ha^{-1})	Yield std dev (Mg ha^{-1})	Elevation mean (m)	Elevation std dev (m)	Elevation min (m)	Elevation max (m)
Whole-field (all treatments)	4.00	0.38	256.7	1.4	251.6	260.1
By soil mapping unit						
Soil A	4.04	0.35	257.0	1.1	253.9	260.1
Soil B	3.99	0.40	255.8	1.5	251.6	259.1
Soil C	3.49	0.40	258.6	1.0	256.4	260.1
By seeding rate (000s ha^{-1})						
198	4.01	0.44	256.5	1.5	252.1	259.7
247	4.00	0.36	256.6	1.6	251.6	260.1
297	4.02	0.36	256.6	1.4	252.6	260.1
346	4.02	0.38	256.8	1.3	252.6	259.4
395	3.97	0.33	257.2	1.1	252.7	259.7

localized weighted elevation (average elevation of immediate surrounding observations) and subtracting the elevation measurement for the given location. Relative elevation provides the relative micro-scale elevation (measures whether the observation was higher or lower than neighbouring observations) (Lowenberg-DeBoer et al. 2006).

Interaction terms between elevation, seeding rate and soil type were included to determine if treatment response varied with elevation and or soil. All soya bean seeding rates were represented across the whole range of topography and all soil map units (Table 4.6 and Fig. 4.6); however, it is questionable if adequate yield observations were recorded for Soil C.

Table 4.7 Description of variables

Variables	Description
POP	Seeding population in thousands
POP_SQ	Seeding population squared
Soil A	Soil A binary variable
Soil B	Soil B binary variable
Soil C	Soil C binary variable
POP_Soil C	Population by Soil C binary variable interaction
POP_Soil B	Population by Soil B binary variable interaction
Elevation by Soil C	Elevation by Soil C binary variable interaction
Elevation by Soil B	Elevation by Soil B binary variable interaction
Elevation	Standardized elevation (minimum elevation = 0) (m)
Elevation squared	Elevation squared
POP by elevation	Population by elevation interaction
Relative elevation	Relative elevation

Average yield is consistent across seeding rate treatments, differing by only 0.044 Mg ha^{-1} across all five rates and only 0.005 Mg ha^{-1} for the four lowest seeding rates (Table 4.6). Although the 395K seeding rate has the smallest mean yield, it also has the smallest standard deviation (Table 4.6). The lowest seeding rate tested, 198K seeds, has the greatest variability with a standard deviation of 0.439 Mg ha^{-1} (Table 4.6).

4.5.2 Data Analysis

Precision agriculture data in general are spatially correlated and are not independent. Traditional methods of statistical analysis cannot estimate the variability of the estimate properly (predicted yield in this case) when observations are spatially correlated, whereas spatial methods can model the spatial variation explicitly and accurately to estimate reliable treatment responses. Analyses were done with both traditional and spatial techniques to provide a comparison of results and to illustrate the decision that would have been made without adequate spatial analysis. The spatial error process model was calculated using the 45-m inverse distance weights matrix determined empirically by LM tests of OLS residuals. The regression output from the spatial error process model is given in Table 4.8.

The economic analysis was based on the results of the spatial error process model (Table 4.9). The estimated coefficients from the model output were put into a quadratic yield response equation, and then converted into a profitability function by transforming values from physical yield to monetary units. The price ratio was based on the assumed soya bean price of $216.5 per Mg and soya bean seed costs of $0.20 per thousand seeds (Table 4.9). Profit maximization occurs when the marginal costs equate to the marginal revenue.

Table 4.8 Regression output from the spatial error process model[a]

Variable	Coefficient	Standard error	Z-value	Probability (p)
Intercept	43.17	5.302	8.141	0
POP	0.292	0.067	4.335	0
POP_SQ	−0.001	0.000	−4.687	0
MeB	−23.20	4.802	−4.830	0
FgB2	−3.545	2.229	−1.591	0.112
POP_MeB	0.116	0.018	6.321	0
POP_FgB2	0.020	0.009	2.196	0.028
Elevation by MeB	0.342	0.188	1.818	0.069
Elevation by FgB2	−0.051	0.109	−0.473	0.637
Elevation	1.089	0.296	3.673	0
Elevation squared	−0.053	0.008	−6.828	0
POP by elevation	0.002	0.001	1.641	0.101
Relative elevation	0.957	0.156	6.134	0

[a]Evaluated at minimum relative elevation, i.e. zero (N = 3897). See Table 4.7 for a list of abbreviations.

Table 4.9 Optimal agronomic and economic seeding rates

	Portion of field			
	Field	Soil A	Soil B	Soil C
Optimal seeding rates (1000 seeds ha^{-1})				
Agronomic	319	309	329	420
Economic[a]	284	274	294	385

[a]Soya bean price of $216.5 per Mg and seed cost of $0.20 per 1000 seeds. Soils as defined by USDA-NRCS.

Yield response to soya bean seeding rates was estimated to have a quadratic functional form which can be expressed as $y_s = pop + pop^2$ where y_s is soya bean yield and pop and pop^2 are seeding population and population squared. Regression coefficients were used to calculate yield maximizing soya bean population levels, or what is commonly known as the agronomic maximum. Maximization of yield, however, does not equate with profit maximization unless the input, soya bean seed, is free. To calculate profit maximization levels, or the economic optimal levels, the profit function, $\pi = R - C$, is used where π is profit, R is revenue and C is cost. The profit function can be expanded to $\pi = p_y y - p_x x$, where p_x is the price of the input, x. So the equation for profit from a soya bean population rate study may be $\pi_s = p_y \left(pop + pop^2\right) - p_s \left(pop\right)$ where π_s is profit from soya bean and p_s is the price of soya bean seed. Yield maximization and profit maximization levels are found by setting the first derivative of the respective equation to zero and solving for the optimum level of input use. The profit maximization level of soya bean seeds for the generic equation is given by $pop = \frac{p_s - p_y}{2p_y}$.

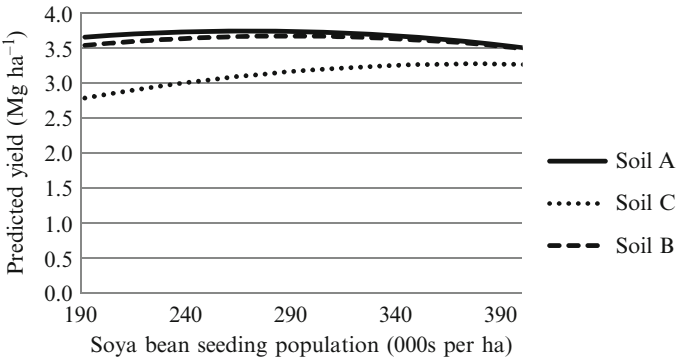

Fig. 4.7 Predicted yield response to soya bean seeding rate using spatial analysis

4.5.3 Case Study Results

The results of both traditional and spatial analyses were compared to show the differences in predicted yield and decisions that would have been made. Figure 4.7 shows the quadratic yield response functions estimated by spatial statistical methods. Optimal seeding rates varied in relation to both elevation and soil for both traditional and spatial statistical methods. The agronomic optimal seeding rate (the rate that maximizes yield) for the field is 319K seeds ha^{-1} and ranged from 309K on Soil A to 420K on Soil C (Table 4.6). It should be noted that the 420K seeding rate was higher than the largest seeding rate included in the experiment and might not have been properly estimated. Soil B has an agronomic optimal rate of 329K seeds ha^{-1} similar to Soil A. The economic analyses indicate a field profit maximization rate (economic) of 284K seeds ha^{-1} (Table 4.9). The rates for Soil A and B are similar at 274K and 294K seeds ha^{-1}, respectively; however that for Soil C is 385K seeds ha^{-1} (Table 4.9).

4.5.4 Case Study Summary

In general, soya bean yields were largest on Soil A and smallest on Soil C. Soil B responded similarly to Soil A, although there were minor differences including a slightly higher, but significantly different, optimal seeding rate. The results of this study suggest that soya bean seeding rates may be as low as 272 to 297K seeds ha^{-1} and still maximize profits.

Caution must be used in making decisions on Soil C for two reasons. First, it should be noted that Soil C may not have been a large enough portion of the field to have an adequate number of yield observations. Second, higher seeding rates are necessary to estimate the yield response properly on Soil C and to make appropriate recommendations. Future research on a wider range of soil types with sufficient

observations per soil type might provide information that is suitable for making variable-rate application prescriptions for seeding. In addition, a precise delineation of soil units or 'management zones' based upon soil characteristics affecting yield response might enable greater precision in estimating yield response and making decisions. Without spatial analysis, which explicitly modelled the spatial structure of the error term, evaluation of this site-specific dataset would not have produced practical results.

4.6 Conclusion

Yield monitors and GPS have provided farmers with a management tool with which to make farm management decisions; albeit tools and data that require specialized skills and more management ability. To make the most of yield data, appropriate statistical techniques are necessary. Spatial statistical methods provide the foundation to develop techniques to analyse site-specific yield monitor data.

Yield monitor data have been used by many farmers to accomplish several farmer-related tasks. Farmers have made use of the data in ways that the developers and manufacturers may not have anticipated. Researchers are beginning to understand how yield monitor data can be useful and how best to make use of them with spatial statistical techniques.

References

Alchanatis, V., Safren, O., Levi, O., & Ostrovsky, V. (2007). Apple yield mapping using hyperspectral machine vision. In J. Stafford (Ed.), *Precision agriculture'07* (pp. 555–562). Wageningen, The Netherlands: Wageningen Academic Publishers.

Ampatzidis, Y. G., Vougioukas, S. G., Bochtis, D. D., & Tsatsarelis, C. A. (2009). A yield mapping system for hand-harvested fruits based on RFID and GPS location technologies: field testing. *Precision Agriculture, 10*, 63–72.

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht, The Netherlands: Kluwer.

Anselin, L. (1989). *What is special about spatial data*. Technical Report 89–4. Santa Barbara: University of California, Santa Barbara, National Center for Geographic Information and Analysis.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.

Anselin, L. (2001). Spatial effects in econometric practice in environmental and resource economics. *American Journal of Agricultural Economics, 83*, 705–710.

Anselin, L., Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. *American Journal of Agricultural Economics, 86*, 675–687.

Arslan, S., & Colvin, T. S. (2002). An evaluation of the response of yield monitors and combines to varying yields. *Precision Agriculture, 3*, 107–122.

Bell, K. P., & Bockstael, N. E. (2000). Applying the generalized-moment estimation approach to spatial problems involving micro level data. *The Review of Economics and Statistics, 82*, 72–82.

Bivand, R. (with contributions by Luc Anselin, Renato Assunçaõ, Olaf Berke, Andrew Bernat, Marilia Carvalho, Yongwan Chun, Carsten Dormann, Stéphane Dray, Rein Halbersma, Elias Krainski, Nicholas Lewin-Koh, Hongfei Li, Jielai Ma, Giovanni Millo, Werner Mueller, Hisaji Ono, Pedro Peres-Neto, Markus Reder, Michael Tiefelsdorf and Danlin Yu). (2009). spdep: Spatial dependence: Weighting schemes, statistics and models. R package version 0.4–34.

Blackmore, S., & Marshall, C. J. (1996). Yield mapping: errors and algorithms. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Proceedings of the Third International Conference on Precision Agriculture* (pp. 403). Madison, WI: ASA,CSSA,SSSA.

Bockstael, N. E. (1996). Modeling economics and ecology: the importance of a spatial perspective. *American Journal of Agricultural Economics, 78*, 1168–1180.

Bongiovanni, R. (2002). *A spatial econometric approach to the economics of site-specific nitrogen management in corn production*. PhD dissertation. West Lafayette, IN: Purdue University,

Bramley, R. G. V., & Williams, S. K. (2001). A protocol for winegrape yield maps. In G. Grenier, & S. Blackmore (Eds.), *Proceedings of the Third European Conference on Precision Agriculture* (pp. 773–767). Montpellier: Agro Montpellier.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: models and applications*. London: Pion.

Cressie, N. A. C. (1993). *Statistics for spatial data* (Revised ed.). New York: Wiley.

Dubin, R. A. (2003). Robustness of spatial autocorrelation specifications: some Monte Carlo evidence. *Journal of Regional Science, 43*, 221–248.

Durrence, J. S., Hamrita, T. K., & Vellidis, G. (1999). A load cell based yield monitor for peanut feasibility study. *Precision Agriculture, 1*, 301–317.

Florax, R. J. G. M., & Nijkamp, P. (2004). Misspecification in linear spatial regression models. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 695–707). San Diego: Academic Press.

Florax, R. J. G. M., & Rey, S. (1995). The impacts of misspecified spatial interaction in linear regression models. In L. Anselin, & R J. G. M. Florax (Eds.), *New directions in spatial econometrics* (pp. 111–135). Berlin: Springer.

Florax, R. J. G. M., Voortman, R. L., & Brouwer, J. (2002). Spatial dimensions of precision agriculture: a spatial econometric analysis of millet yield on Sahelian coversands. *Agricultural Economics, 27*, 425–443.

Griffin, T. W. (2006). *Decision-making from on-farm experiments: spatial analysis of precision agriculture data*. PhD dissertation. West Lafayette, IN: Purdue University,

Griffin, T. W. (2009a). *Adoption of yield monitor technology for crop production*. University of Arkansas Division of Agriculture Factsheet FSA37.

Griffin, T. W. (2009b). *Farmers' use of yield monitors*. University of Arkansas Division of Agriculture Factsheet FSA36.

Griffin, T. W., Dobbins, C. L., Vyn, T., Florax, R. J. G. M., & Lowenberg-DeBoer, J. (2008). Spatial analysis of yield monitor data: case studies of on-farm trials and farm management decision-making. *Precision Agriculture, 9*, 269–283.

Griffin, T. W., & Lowenberg-DeBoer, J. (2005). Worldwide adoption and profitability of precision agriculture: Implications for Brazil. *Revista de Politica Agricola, 14*, 20–37.

Griffin, T. W., Lowenberg-DeBoer, J., Lambert, D. M., Peone, J., Payne, T., & Daberkow, S. G. (2004). *Adoption, profitability, and making better use of precision farming data*. Staff Paper No. 04–06. West Lafayette, IN: Department of Agricultural Economics, Purdue University.

Griffin, T. W., Mark, T. B., Dobbins, C. L., & Lowenberg-DeBoer, J. (2009). Whole farm profitability impact from implementing and harvesting on-farm trials with precision agriculture technologies. In *Proceedings of the 17th International Farm Management Congress*, Bloomington, IL, July 19–24, 2009.

Griffin, T. W., & Slinsky, S. P. (2008). On-board module builders, on-farm research, and precision agriculture: information technology from cotton fields to cotton gins to cotton warehouses and back again. In *Proceedings of the 9th International Conference on Precision Agriculture*, Denver, CO July 21–23, 2008.

Grisso, R., Alley, M., & McClellan, P. (2002). *Precision farming tools: yield monitor*. Publication 442–502. Virginia Cooperative Extension.

Hicks, D. R., van den Heuvel, R. M., & Forek, Z. Q. (1997). Analysis and practical use of information from on-farm strip trials. *Better Crops, 81*, 18–21.

Instituto Nacional de Tecnología Agropecuaria. (2008). Proyecto Agricultura de Precisión, INTA. 2008. 8vo. Curso Internacional de Agricultura de Precisión. Octubre de 2008. INTA Manfredi, Córdoba, Argentina. Available online http://www.agriculturadeprecision.org/ (Last accessed January 2010).

Kay, R. D., Edwards, W. M., & Duffy, P. A. (2008). *Farm management*. Singapore: McGraw-Hill.

Konstantinovic, M., Woeckel, S., Schulze Lammers, P., & Sachs, J. (2007). Influence of the sugar beet spatial arrangement on yield mapping of sugar beet using UWB radar. In J. V. Stafford (Ed.), *Precision agriculture'07* (pp. 341–348). Wageningen, The Netherlands: Wageningen Academic Publishers.

Kumhala, F., Kroulik, M., Masek, J., Prochazka, P., & Kviz, Z. (2005). Evaluation of forage yield map techniques on a mowing-conditioning machine. In Stafford, J. V. (Ed.), *Precision agriculture'05* (pp. 401–408). Wageningen, The Netherlands: Wageningen Academic Publishers.

Lambert, D. M., Lowenberg-DeBoer, J., & Bongiovanni, R. (2004). A comparison of four spatial regression models for yield monitor data: a case study from Argentina. *Precision Agriculture, 5*, 579–600.

Lambert, D. M., Lowenberg-DeBoer, J., & Malzer, G. L. (2006). Economic analysis of spatial-temporal patterns in corn and soybean response to nitrogen and phosphorus. *Agronomy Journal, 98*, 43–54.

Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory, 18*, 252–277.

Lowenberg-DeBoer, J., Griffin, T. W., & Florax, R. G. J. M. (2006). Local spatial autocorrelation in precision agriculture settings accounting for micro-scale topography differences, In *Proceedings of the 8th International Conference on Precision Agriculture and Other Resource Management*, Minneapolis Minnesota, July, 2006.

Maguire, S., Godwin, R. J., Smith, D. F., & O'Dogherty, M. J. (2003). Hay and forage measurement for mapping. In J. V. Stafford, & A. Werner (Eds.), *Precision agriculture* (pp. 379–384). Wageningen, The Netherlands: Wageningen Academic Publishers.

Maine, N, Nell, W. T., Lowenberg-DeBoer, J., & Alemu, Z. G. (2007). Economic analysis of phosphorus applications under variable and single-rate applications in the Bothaville District. *Agrekon, 46*, 532–547.

Meyer-Aurich, A., Muhammad, N., & Herbst, R. (2008). On-farm experimentation for identification of site-specific fertilization potentials. *In Proceedings of the 9th International Conference on Precision Agriculture*, Denver, CO, July 20–23, 2008.

Noack, P. O., Muhr, T., & Demmel, M. (2005). Effect of interpolation methods and filtering on the quality of yieldmaps. In J. V. Stafford, J. (Ed.), *Precision agriculture'05* (pp. 701–706). Wageningen, The Netherlands: Wageningen Academic Publishers.

Papadakis, J. S. (1937). Méthode statistique pour des expériences sur champ. Bulletin Scientifique, 23. Thessalonique, Greece: Thessalonique Institut d'amélioration des plantes.

Pelletier, G., & Upadhyaya, S. K. (1999). Development of a tomato load/yield monitor. *Computers and Electronics in Agriculture, 23*, 103–117.

R Development Core Team. (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. (www.r-project.org).

Shearer, S. A., Fulton, J. P., McNeill, S. G., Higgins, S. F., & Mueller, T. G. (1999). *Elements of precision agriculture: basics of yield monitor installation and operation*. PA-1. University of Kentucky Cooperative Extension Service.

Simbahan, G. C., Dobermann, A., & Ping, J. L. (2004). Screening yield monitor data improves grain yield maps. *Agronomy Journal, 96*, 1091–1102.

Sonka, S. T., Bauer, M. E., Cherry, E. T., Colburn, J. W. Jr., Heimlich, R. E., et al. (1997). *Precision agriculture in the 21st century. Geospatial and information technologies in crop management*. Washington, DC: National Academy Press.

Stewart, C. M., McBratney, A. B., & Skerritt, J. H. (2002). Site-specific durum wheat quality and its relationship to soil properties in a single field in northern New South Wales. *Precision Agriculture, 3*, 155–168.

Schueller, J. K., Whitney, J. D., Wheaton, T. A., Miller, W. M., & Turner, A. E. (1999). Low-cost automatic yield mapping in hand-harvested citrus. *Computers and Electronics in Agriculture, 23*, 145–153.

Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: software for removing errors from crop yield maps. *Agronomy Journal, 99*, 1471–1482.

Taylor, J., Whelan, B., Thylen, L., Gilbertsson, M., & Hassall, J. 2005. Monitoring wheat protein content on-harvester: Australian experiences. In J. V. Stafford (Ed.), *Precision agriculture'05* (pp. 369–375). Wageningen, The Netherlands: Wageningen Academic Publishers.

United States Naval Observatory. USNO NAVSTAR Global Positioning System Available online at: http://tycho.usno.navy.mil/gpsinfo.html (Last accessed January 2010).

Vellidis, G., Perry, C. D., Durrence, J. S., Thomas, D. L., Hill, R. W., Kvien, C. K., Hamrita, T. K., & Rains, G. C. (2001). The peanut yield monitoring system. *Transactions of the ASAE, 44*, 775–785.

Vellidis, G., Perry, C. D., Rains, G., Thomas, D. L., Wells, N., & Kvien, C. K. (2003). Simultaneous assessment of cotton yield monitors. *Applied Engineering in Agriculture, 19*, 259–272.

Wild, K., & Auernhammer, H. (1999). A weighing system for local yield monitoring of forage crops in round balers. *Computers and Electronics in Agriculture, 23*, 119–132.

Wild, K., Ruhland, S., & Haedicke, S. (2003). Pulse radar systems for yield measurements in forage harvesters. In J. V. Stafford, & A. Werner (Eds.), *Precision agriculture* (pp. 739–744). Wageningen, The Netherlands: Wageningen Academic Publishers.

Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: how the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: University of Michigan Press.

Chapter 5
Space–Time Geostatistics for Precision Agriculture: A Case Study of NDVI Mapping for a Dutch Potato Field

G.B.M. Heuvelink and F.M. van Egmond

Abstract Many environmental variables that are relevant to precision agriculture, such as crop and soil properties and climate, vary both in time and space. Farmers can often benefit greatly from accurate information about the status of these variables at any particular point in time and space to aid their management decisions on irrigation, fertilizer and pesticide applications, and so on. Practically, however, it is not feasible to measure a variable exhaustively in space and time. Space–time geostatistics can be useful to fill in the gaps. This chapter explains the basic elements of space–time geostatistics and uses a case study on space–time interpolation of the normalized difference vegetation index (NDVI) as an indicator of biomass in a Dutch potato field. Space–time geostatistics proves to be a useful extension to spatial geostatistics for precision agriculture, although theoretical as well as practical advances are required to mature this subject area and make it ready to be used for within-season, within-field decision making by farmers.

Keywords Space–time geostatistics · Variogram · Kriging · NDVI · Interpolation · Spatial variability · Temporal variability · Prediction · Uncertainty · Mapping · GPS · Crop growth · Sampling

5.1 Introduction

Many of the soil properties that are relevant to precision agriculture vary both in space and time. Examples are soil moisture, soil nutrient concentrations and pH. Clearly, variation in space–time is not restricted to the soil, but extends to other domains such as crops and climate. For instance, the biomass and protein concentration

G.B.M. Heuvelink (✉)
Environmental Sciences Group, Wageningen University and Research Centre,
PO Box 47, 6700 AA Wageningen, The Netherlands
e-mail: Gerard.Heuvelink@wur.nl

F.M. van Egmond
The Soil Company, Leonard Springerlaan 9, 9727 KB Groningen, The Netherlands
e-mail: fenny@medusa-online.com

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_5,

of the root, stem and leaves of a crop depend on location and time, and so do precipitation and temperature when the study area is large. Farmers can benefit greatly from accurate information about the status of these variables at any particular point in time and space. It can aid their management decisions on irrigation, fertilizer and pesticide applications and so on. Practically, however, it is not feasible to measure a variable exhaustively in space and time. For example, Snepvangers et al. (2003) used time domain reflectometry to record about 100,000 observations of the topsoil water content in a 0.36 ha grassland plot during a 30 day period, but the resulting data were still sparse and interpolation was needed to describe the space–time variation adequately. In many practical cases, even in precision agriculture where measurements may be abundant, the size of the dataset will be much smaller than in the example above and the space–time domain will be much larger. Thus, a common problem is to construct a high-resolution space–time representation of a variable that varies in space and time from a limited number of observations. Space–time geostatistics can be useful for this.

Space–time geostatistics is a natural extension of 'spatial' geostatistics as described in Chapter 1. It begins by characterizing the variation in space and time with variograms. Next, these variograms are used to predict the target variable at unmeasured points by kriging. The prediction error can be quantified and trends incorporated to reduce this error. Trends in space and or time are incorporated when part of the variation in the target variable can be explained by explanatory variables, such as when soil type is used to explain spatial variation in soil texture or when precipitation is used to explain temporal variation in soil moisture. The difference from 'spatial' geostatistics is that variation occurs in space and time, and both these sources of variation must be modelled and their effects on prediction taken into account. Variation in space might be much less than that in time. For instance, rainfall events are likely to pass over the entire field, but they vary considerably over time. The opposite is also common; soil bulk density can differ markedly between locations with different types of soil or landuse, but it changes little over a timespan of a few years. In fact, the bulk density example hints at a situation where temporal variation may be negligibly small compared to spatial variation. In such a case one might decide to disregard the temporal variation completely and return to conventional 'spatial' geostatistics. When spatial variation is negligible a time-series analysis will suffice. Thus, space–time geostatistics is required when neither spatial nor temporal variation can be ignored, but this does not mean that both must be equally large. Also, the lengths of spatial and temporal dependence will be different, if only because space and time have different measurement units.

This chapter explains the basic elements of space–time geostatistics and uses a case study on space–time interpolation of the normalized difference vegetation index (NDVI) as an indicator of biomass in a Dutch potato field. We begin with a description of the study site and discuss the positional correction of GPS data from the field, followed by an exploratory data analysis of the NDVI data. Next we present the theory of space–time geostatistics. We apply it to the case study, discuss results and draw conclusions.

5.2 Description of the Lauwersmeer Study Site and Positional Correction of NDVI Data

The study site is near the village of Vierhuizen in the Lauwersmeer area to the North of the Netherlands at 6.27° longitude and 53.35° latitude. The climate is temperate maritime; the annual rainfall in 2006 was 661 mm with an annual average temperature of 11.7°C. In 2006, the growing season for potatoes was relatively warm and dry except for August, which was colder and wetter than usual. The soil is characterized as a fluvisol formed on young calcareous marine sediments. Soil texture ranges from loamy sand to sandy clay loam with 1–4% organic matter. Elevation ranges from 0.5 to 1.5 m above sea level with lower areas on the W and SE parts of the field (Fig. 5.1). The lower parts of the field are more sandy. Typical crop rotations consist of seed and consumption potatoes, sugar beet, wheat and onions.

A 10 ha field with two potato varieties, Innovator and Sofista, was studied during the growing season of 2006. Innovator is a mid-early variety with fast and good biomass development, whereas Sofista is classified as mid-late with good biomass development.

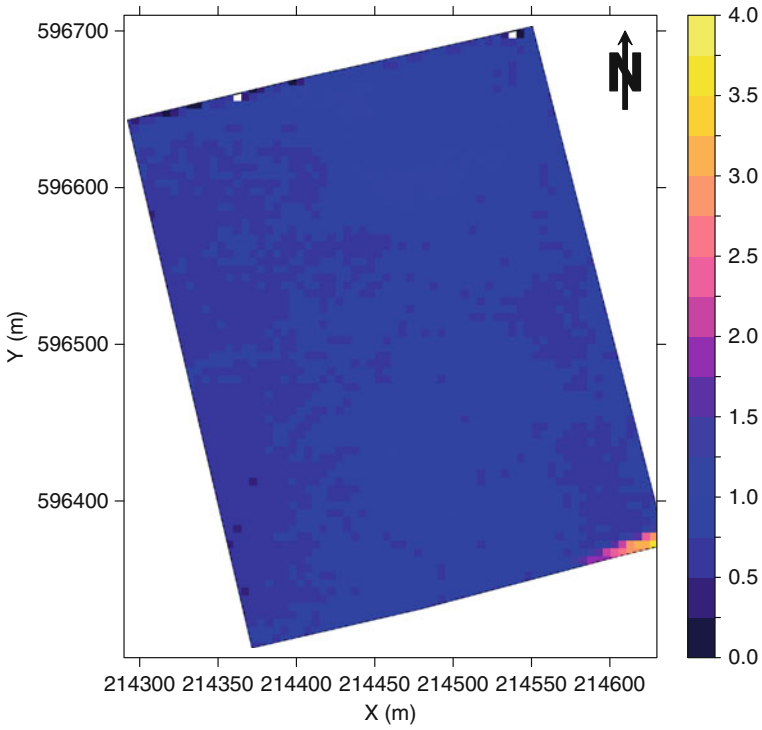

Fig. 5.1 Digital elevation map (metres above sea level) of the study site, obtained from the Dutch AHN (Actueel Hoogtebestand Nederland 2009)

A Crop Circle sensor (Holland Scientific 2008) measured spectral reflectance each time the crop was sprayed against phytophtora, a potato disease. The active spectral reflectance sensor has a spatial support of about 1 m^2. It emits at 650 and 880 nm and measures reflectance in the visible (between 400 and 680 nm) and near-infrared part of the spectrum (800–1100 nm). The sensor calculates the NDVI based on those reflectances. The sensor- and GPS data are logged every second. Data from headlands are not taken into account. The number of tramlines recorded and the driving speed differs with measurement date (Table 5.1).

The NDVI data were obtained from reflectances recorded on 16 different days during the growing season of 2006. The Crop Circle sensor was mounted 8.5 m behind the tractor on a spraying boom and 6.5 m left of the centre. The GPS was located on the tractor roof. The difference in position between the sensor and GPS caused an error in the positions of recorded values that depends on the driving direction. The GPS coordinates logged were corrected for this difference (Fig. 5.2a). The accuracy of a GPS is fairly constant during any one day, but locations may shift slightly when measuring the same coordinates on different days. This effect is clearly visible from the wide transects in Fig. 5.2a. The farmer measures NDVI when he sprays the field and uses exactly the same tramlines each time to minimize crop damage. Hence the corrected coordinates measured on different days should be positioned along the same lines. The corrected coordinates were therefore corrected a second time for the 'temporal' GPS error. When the coordinates of a tramline differed by more than 0.5 m from the centre of the tramline, they were corrected in a perpendicular direction to the driving direction to match the centre of the tramline and to minimize unwanted displacement of the coordinates in the driving direction. The results are shown in Fig. 5.2b.

5.3 Exploratory Data Analysis of Lauwersmeer Data

The main interest of the farmer is in the distribution of the potato growth over time and space. Accurate information on growth and variation in growth enables the farmer to fine-tune fertilizer application and the timing and amount of chemicals to remove above ground biomass at the end of the growing season. The space–time distribution of below ground biomass, and hence tuber yield, is correlated to the above ground green biomass, of which NDVI is an indicator (Baret and Guyot 1991; Carlson and Ripley 1997). The values of NDVI range between 0 and 1. An NDVI of 0.2 indicates bare soil, a value of 0.65 indicates that the crop canopy is closed and a further increase in NDVI indicates an increase in the number of green leaf layers. The index 'saturates' at 0.9 which represents a leaf area index (LAI) of 3–4. Throughout the growing season NDVI increases until tuber formation starts and then it decreases as the above ground biomass dies and becomes yellow (Wu et al. 2007).

Figure 5.3 shows a scatter plot of NDVI against day of year (DOY). The NDVI increases rapidly in spring and gradually decreases towards the end of the summer. The spatial variation is large for all dates, as indicated by the wide spread in values. There appear to be outliers with small NDVI values when overall NDVI values

Table 5.1 Descriptive statistics of the Lauwersmeer farm data

	DOY	Crop type	Number of tramlines	Number of observations	Mean	Median	Minimum	Maximum	Standard deviation	Skewness	Correlation with NDVI
NDVI	149	Sofista	1	353	0.08	0.08	0.06	0.12	0.01	0.21	
	149	Innovator	5	1689	0.16	0.16	0.05	0.32	0.04	−0.07	
	154	Sofista	3	1237	0.19	0.19	0.07	0.29	0.03	0.10	
	154	Innovator	5	1408	0.26	0.27	0.06	0.37	0.05	−0.75	
	160	Sofista	1	370	0.28	0.28	0.08	0.40	0.05	−0.27	
	160	Innovator	5	1834	0.44	0.44	0.13	0.59	0.05	−0.67	
	167	Innovator	5	1605	0.69	0.70	0.08	0.75	0.05	−7.44	
	168	Sofista	4	1304	0.60	0.61	0.28	0.73	0.06	−1.29	
	174	Sofista	4	1231	0.72	0.72	0.45	0.77	0.03	−3.20	
	174	Innovator	5	1532	0.75	0.76	0.09	0.79	0.04	−11.00	
	181	Sofista	4	1312	0.76	0.76	0.35	0.79	0.02	−12.71	
	181	Innovator	5	1471	0.78	0.78	0.28	0.80	0.03	−13.62	
	187	Sofista	4	971	0.77	0.78	0.50	0.81	0.02	−6.10	
	187	Innovator	5	1274	0.79	0.79	0.12	0.81	0.03	−17.21	
	193	Innovator	5	1430	0.78	0.78	0.74	0.80	0.01	−0.49	
	196	Sofista	4	1263	0.79	0.79	0.58	0.82	0.02	−3.10	
	202	Sofista	4	850	0.78	0.78	0.64	0.82	0.02	−1.49	
	202	Innovator	5	1210	0.80	0.80	0.70	0.82	0.01	−2.42	
	208	Sofista	4	966	0.75	0.76	0.58	0.83	0.04	−0.95	
	208	Innovator	5	1299	0.77	0.77	0.15	0.80	0.04	−13.14	
	217	Sofista	3	747	0.73	0.73	0.50	0.81	0.04	−0.76	
	217	Innovator	5	1248	0.78	0.78	0.67	0.81	0.01	−1.07	
	224	Sofista	4	755	0.70	0.71	0.53	0.79	0.05	−0.83	
	224	Innovator	5	1074	0.77	0.77	0.41	0.80	0.02	−5.02	

(continued)

Table 5.1 (continued)

	DOY	Crop type	Number of tramlines	Number of observations	Mean	Median	Minimum	Maximum	Standard deviation	Skewness	Correlation with NDVI
NDVI	231	Sofista	4	940	0.65	0.68	0.34	0.79	0.09	−1.05	
	231	Innovator	6	1275	0.77	0.77	0.57	0.81	0.02	−1.77	
	243	Sofista	4	1295	0.33	0.30	0.17	0.78	0.12	1.24	
	243	Innovator	5	1806	0.74	0.75	0.11	0.80	0.03	−10.37	
Clay (%)				33749	15.80	16.26	7.50	22.50	3.13	−0.44	0.035
Sand (%)				33749	62.27	60.70	55.00	84.41	6.59	0.87	−0.029
M0 (%)				33749	68.93	68.60	60.00	84.96	5.27	0.37	−0.030
Organic matter (%)				33749	2.61	2.56	1.20	4.20	0.73	0.16	0.009
Bulk density				33749	1.43	1.43	1.33	1.55	0.04	0.34	−0.026
Elevation (m ASL)				92152	0.812	0.810	−0.1	3.83	0.165	6.88	0.005

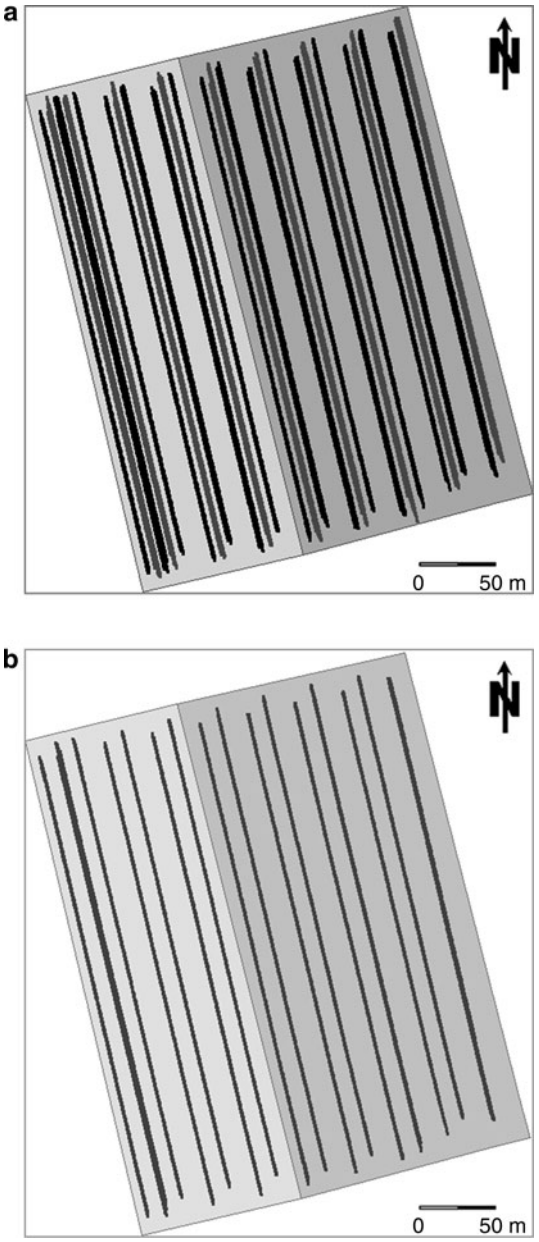

Fig. 5.2 (**a**) The original GPS coordinates (dark grey) and the coordinates after the first correction (*black*) and (**b**) GPS coordinates after second correction. Background colour represents the two potato species (Sofista *light grey*; Innovator dark grey)

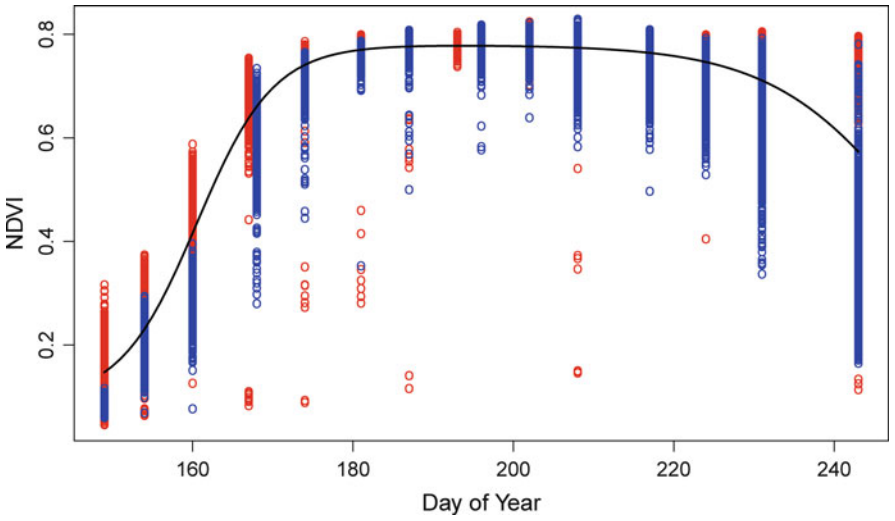

Fig. 5.3 The NDVI observations (Innovator in *red*, Sofista in *blue*) against day of year in 2006. The *solid line* is the fitted temporal trend

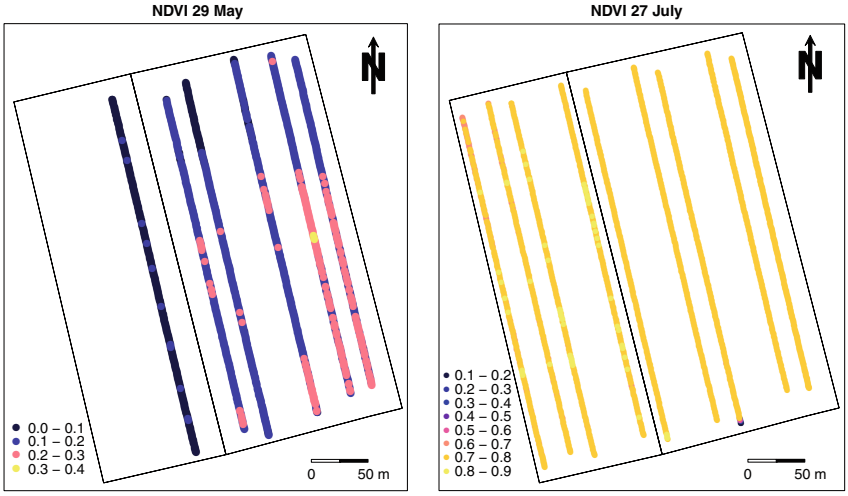

Fig. 5.4 Spatial distribution of NDVI observations for May 29 (DOY 149) and July 27 (DOY 208)

are large, showing that the crop was not fully developed in part of the field. This is confirmed in Fig. 5.4, where the spatial distribution of NDVI is shown for two dates. There are also a few small values in one area near the field boundary on July 27. Note also that the Sofista crop has developed little biomass on May 29. Figure 5.3 shows that the Sofista crop develops later than Innovator and starts to show a decline in above ground biomass earlier. Innovator has a larger average NDVI in late summer when Sofista has already started to decline.

Factors that might influence crop growth, and hence NDVI, can be divided into temporally constant but spatially varying factors and spatially constant but temporally varying factors. The first category includes soil properties, elevation and the spatial allocation of the two different crop types. Soil properties that might influence potato growth are the amount of clay, sand and organic matter, median grain size (M0), potential bulk density and potential water retention. Calculation of the latter two is based on pedotransfer functions (Wösten et al. 2001). Summary statistics of the NDVI, soil and elevation data are given in Table 5.1. Digital elevation data are available at 1-m resolution from the AHN (Actueel Hoogtebestand Nederland 2009) the national altitude dataset of the Netherlands, Fig. 5.1. The second category of spatially constant explanatory variables includes precipitation, temperature and fertilizer application. The crop received a spatially homogeneous nitrogen application on 14 June 2006 (DOY 155).

5.4 Space–Time Geostatistics

Consider a variable $z = \{z(\mathbf{s}, t)|\mathbf{s} \in S, t \in T\}$ that varies within a spatial domain S and a time interval T. Let z be observed at n space–time points $(\mathbf{s}_i, t_i), i = 1, \ldots, n$. In the case study described in Sections 5.2 and 5.3, these space–time observations are transects of observations along the tramlines recorded at several instants in time. Although the total number of observations, n, may be very large, to observe z at each and every combination of time and space is not feasible. To obtain a complete space–time surface of z requires some form of prediction. Therefore, the objective is to obtain a prediction of $z(\mathbf{s}_0, t_0)$ at a point $(\mathbf{s}_0, t_0)$ at which z was not observed, where $(\mathbf{s}_0, t_0)$ typically is associated with the nodes of a fine space–time grid. To do this, z is assumed to be a realization of a random function Z (Webster 2000). The random function Z is characterized by a statistical model that must describe the structure of dependence in space–time. Once the model is fully defined, $Z(\mathbf{s}_0, t_0)$ may be predicted from the observations by kriging as described for the spatial variables in Chapter 1.

The space–time variation of Z can be characterized by first decomposing it into a deterministic trend m and a zero-mean stochastic residual ε as follows:

$$Z(\mathbf{s}, t) = m(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t). \tag{5.1}$$

The trend m is a deterministic, structural component that represents large scale variation. The residual is a stochastic component representing small scale, 'noisy' variation. Alternatively, the trend may be thought of as that part of Z that can be explained physically or empirically by auxiliary information. The residual component still holds important information on the variation when it is correlated in space and or time. Note that the decomposition of Z into a trend and a residual is a scale-dependent, subjective choice made by a modeller (Diggle and Ribeiro 2007).

5.4.1 Characterization of the Trend

As noted in Section 5.1, often the behaviour of a variable over time is entirely different from its behaviour in space. This difference can be represented by the trend in the specification of a space–time model. For instance, crop growth in an agricultural field is causally dependent on factors such as soil, climate and management. This information should be incorporated into the model if possible when predicting crop growth or biomass accumulation in space and time. Ideally, the trend would be a process-oriented, physical-deterministic model. However, when deterministic modelling is not feasible due to a lack of understanding of the underlying governing processes or because external forces and boundary conditions are unknown or unreliably known, one may rely on a regression-type model relating the dependent to the explanatory variables in an empirical way. The simplest approach is to assume that the trend is a linear function of the (possibly transformed) explanatory variables, as in linear multiple regression:

$$m(\mathbf{s}, t) = \sum_{i=1}^{p} \beta_i \cdot f_i(\mathbf{s}, t), \tag{5.2}$$

where the β_i are regression coefficients, the f_i are explanatory variables that must be known exhaustively over the space–time domain and p is the number of explanatory variables.

Estimation of the regression coefficients can be done using common least squares algorithms that minimize the sum of squared differences between the observations and predictions (Montgomery et al. 2006), such as implemented in statistical software packages. Alternatively, maximum likelihood methods may be used (see Chapter 3, also).

After a trend has been specified and estimated, it may be subtracted from Z so that attention can be directed to the space–time stochastic residual ε. With this approach, uncertainties in the detrending procedure are not taken into account in the subsequent analysis. As a consequence, in kriging this causes the uncertainty in predictions to appear smaller than it is. This problem can be avoided by accounting for uncertainties in the trend coefficients, such as is done in universal kriging (Diggle and Ribeiro 2007); this method is also referred to as regression kriging (Hengl et al. 2004).

5.4.2 Characterization of the Stochastic Residual

Throughout this chapter, we assume that the zero-mean stochastic residual ε is multivariate normally distributed. Although a distributional assumption (i.e. normality) is not strictly needed for kriging, the assumption ascertains a completely-specified

statistical model and allows us to compute prediction intervals from the kriging results. Given this assumption, the only information lacking is the autocovariance function, C,

$$C(\mathbf{s}_i, t_i, \mathbf{s}_j, t_j) = \mathrm{E}[\varepsilon(\mathbf{s}_i, t_i) \cdot \varepsilon(\mathbf{s}_j, t_j)], \tag{5.3}$$

where E is the mathematical expectation. Alternatively, we may characterize the second-order properties of ε with the variogram, γ, as follows (see also Chapter 1):

$$\gamma(\mathbf{s}_i, t_i, \mathbf{s}_j, t_j) = \frac{1}{2}\mathrm{E}[(\varepsilon(\mathbf{s}_i, t_i) - \varepsilon(\mathbf{s}_j, t_j))^2]. \tag{5.4}$$

In practice, to estimate γ from observations, some additional simplifying assumptions are necessary. A common approach is to introduce the assumption of second-order stationarity, which posits that the semivariance of $\varepsilon(\mathbf{s}, t)$ and $\varepsilon(\mathbf{s} + \mathbf{h}, t + u)$ depends only on the separating distance in space, $\mathbf{h}$, and that in time, u, between the points: $\gamma(\mathbf{s}, t, \mathbf{s} + \mathbf{h}, t + u) = \gamma(\mathbf{h}, u)$. Although $\mathbf{h}$ is usually a vector in two or more dimensions, it can be replaced by Euclidean distance if isotropy is assumed in space, allowing both h and u to be treated as scalars. In practice a further simplification is needed to be able to estimate a space–time variogram from a set of observations. Here we impose a sum-metric variogram model on the space–time residual (Bilonick 1988; Dimitrakopoulos and Luo 1994; Snepvangers et al. 2003):

$$\gamma(h, u) = \gamma_S(h) + \gamma_T(u) + \gamma_{ST}\left(\sqrt{h^2 + (\alpha \cdot u)^2}\right). \tag{5.5}$$

The first two terms on the right-hand side of Eq. 5.5 allow for zonal space–time anisotropies (i.e. variogram sills that are not the same in all directions). Zonal anisotropy means that the amount of variation in time is smaller or larger than that in space, and or that in joint space–time. If variation in space dominates variation in time, then γ_S will have larger values than γ_T, which may be the case in the bulk density example given in Section 5.1. The opposite would hold for the rainfall example, also mentioned above. The third term on the right-hand side of Eq. 5.5 represents a joint space–time structure. It contains a geometric anisotropy ratio α to match distance in time with distance in space. Geometric anisotropy is needed here because a unit of distance in space is not the same as a unit of distance in time. For instance, if $\alpha = 20$ m per day, then two points that are separated by 100 m in space and zero days in time have the same correlation as two points that are 5 days apart in time and zero metres apart in space, or as two points that are separated by 60 m in space and 4 days in time (Fig. 5.5).

The sum-metric model simplifies the space–time variogram to a form such that its parameters can be estimated from a space–time dataset, but there are other approaches too. Among alternative models the product-sum model is often used (De Cesare et al. 2001). Research into which models are most appropriate for which situations is ongoing (see the discussion in Section 5.6).

Once the trend and variogram of the residuals have been specified, space–time prediction can be done in the usual way by ordinary or simple kriging (see Chapter 1). Kriging not only provides the best linear unbiased predictor of $Z(\mathbf{s}_0, \mathrm{t}_0)$

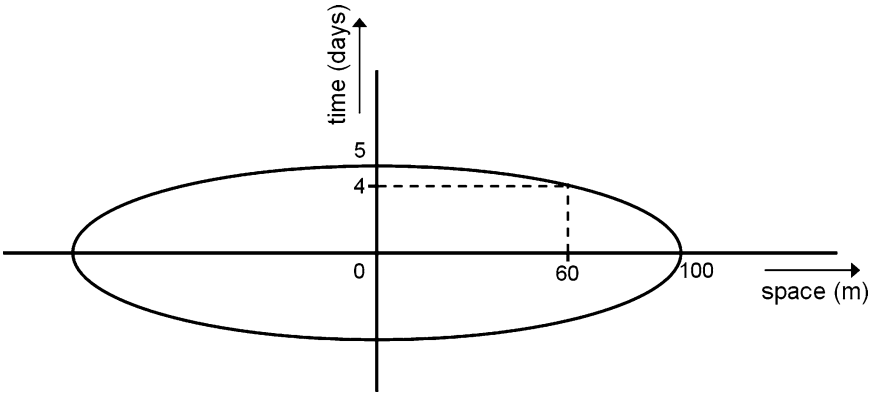

Fig. 5.5 Graphical illustration of geometric space–time anisotropy with an anisotropy ratio of 20 m per day. All points on the ellipse have the same semivariance

at any space–time point $(\mathbf{s}_0, t_0)$, but it also quantifies the kriging prediction error with the kriging variance from which maps of the kriging standard deviation or lower and upper limits of prediction intervals can be derived. In space–time kriging, this gives results in three dimensions. These can be presented as a series of two-dimensional maps by taking time slices (which may be presented as a moving image when the distance between subsequent times is sufficiently small), or as time-series of predictions and prediction error standard deviations at different locations. The advantage of space–time kriging over spatial kriging for separate time points is that all observations are used rather than just the observations at the particular point in time, and that predictions in between measurement times can be made.

5.5 Application of Space–Time Geostatistics to the Lauwersmeer Farm Data

5.5.1 Characterization of the Trend

The absolute value of the correlation coefficients between soil properties and NDVI, and between elevation and NDVI are small (Table 5.1). Although these values were larger when the correlation was computed for separate days, they remained small (data not shown). However, the correlation between NDVI and crop type is greater (0.085) and there is also a clear relationship between NDVI and time (Fig. 5.3). We decided, therefore, to define the trend as a linear combination of crop type and a time effect:

$$m(\mathbf{s}, t) = \beta_S \cdot \delta(crop\ type = Sofista) + \beta_I \cdot \delta(crop\ type = Innovator) + f(DOY). \tag{5.6}$$

Here, the function δ is an indicator transform that is one if the condition is satisfied and zero otherwise; and f is a double logistic function that has been applied successfully to model temporal variation of NDVI (Fischer 1994):

$$f(DOY) = v_{\min} + v_{\max}\left(\frac{1}{1+e^{m_1-n_1\cdot DOY}} - \frac{1}{1+e^{m_2-n_2\cdot DOY}}\right). \qquad (5.7)$$

The six parameters of the double logistic function were fitted on the NDVI data using a least squares minimization approach; the parameters are $v_{\min} = 0.091$, $v_{\max} = 0.689$, $m_1 = 33.7$, $n_1 = 0.2099$, $m_2 = 28.1$ and $n_2 = 0.1123$. The graph of the fitted temporal trend is shown in Fig. 5.3. Next, the remaining two regression coefficients were estimated on the residuals after removal of the temporal trend by unweighted least squares; the coefficients are $\beta_S = -0.0497$ and $\beta_I = 0.0335$. Estimation of the trend could be improved by the simultaneous fitting of all parameters or by using residual maximum likelihood (REML) to estimate the parameters (Chapter 3), however, this may be computationally difficult because Eq. 5.7 is nonlinear in its parameters. A more complex trend including interactions might also have been used to take into account that the effect of crop type varies with time. Absolute values of the correlation coefficients between the NDVI residuals and soil properties and elevation were all smaller than 0.10, indicating that there was no substantial predictive power left in the auxiliary information.

The histogram of the NDVI residuals is shown in Fig. 5.6. There are outliers to the left of the distribution in Fig. 5.6b. These correspond to locations at the boundary of the parcels where the crop did not develop fully (see Figs. 5.2 and 5.3, also). We decided to remove all values < -0.3 (214 observations, 0.6% of the total) because the outliers would impair the geostatistical analysis too much. The choice of the -0.3 threshold was subjective, however, if a value had been chosen closer to zero many more observations would have to be removed (Fig. 5.6b). As a consequence of discarding these observations, the resulting maps will be biased to larger

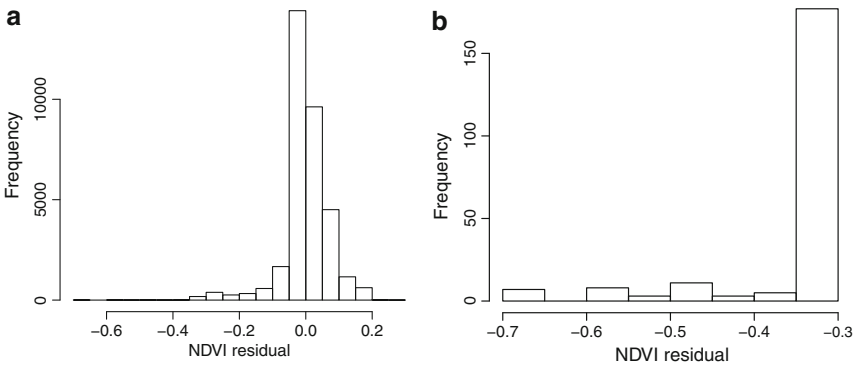

Fig. 5.6 (**a**) Histogram of NDVI residuals after removal of a double logistic temporal trend and crop type means and (**b**) histogram of a subset of the same data with only values smaller than −0.3

NDVI values at the boundary of the field. Alternatively, all the observations could be retained and transformed prior to variogram estimation and kriging, for instance using a disjunctive kriging approach with Hermite polynomial transformation as described in Chapter 1.

5.5.2 Characterization of the Stochastic Residual

The experimental variogram of the NDVI residuals was calculated for various temporal and spatial lags. The temporal lag size was chosen to be fairly large because there were only 16 measurement dates. Six time lags were used, each having a width of 10 days. The spatial lag size used was 15 m. Marginal spatial and temporal variograms were also calculated; these variograms are specific to only the space or time dimension. For instance, the marginal spatial variogram is computed on NDVI residuals for which the distance in time is zero (i.e. observations collected on the same day). The marginal and space–time experimental variograms are shown in Fig. 5.7.

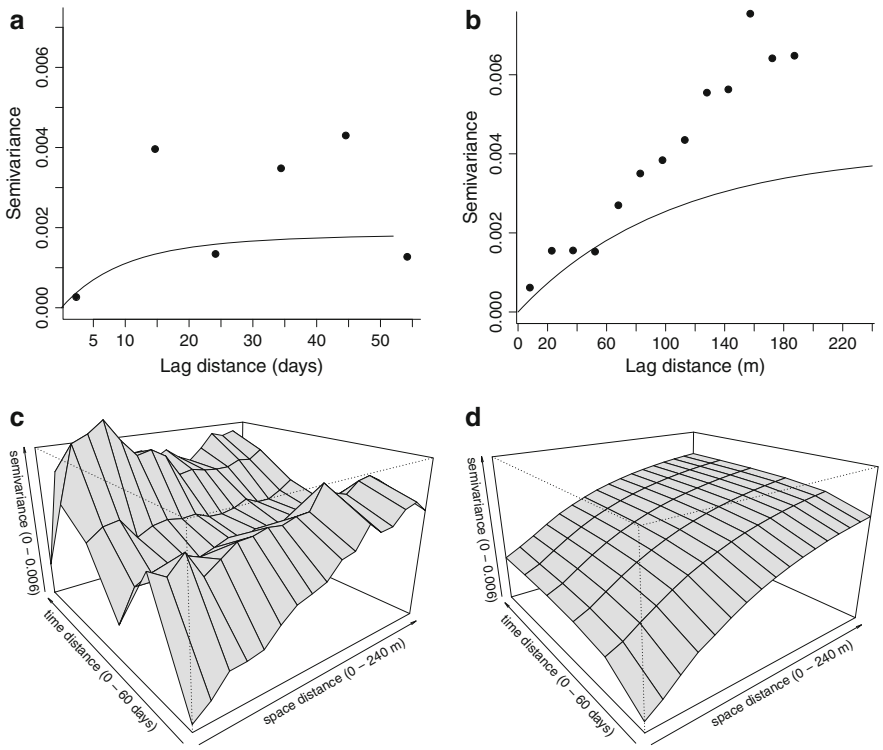

Fig. 5.7 Marginal experimental variogram (*dots*) and fitted model (*solid lines*): (**a**) in time direction and (**b**) space direction. Perspective plots of: (**c**) 3D experimental variogram and (**d**) fitted space–time variogram model

Table 5.2 Parameters of the fitted space–time variogram model

Variogram component	Nugget	Sill	Range	Anisotropy ratio
Spatial	0	0.00283	100 m	NA
Temporal	0	0.00114	8 days	NA
Spatio-temporal	0	0.00047	120 m	6 m/day

The experimental variogram in the time direction, in particular, is noisy. Nevertheless, the variograms show convincingly that the NDVI residuals are correlated in space and time. Near the origin, where distances in time and space are small, the semivariances are smaller than at larger distances. Spatial and temporal variation are of the same order of magnitude; there is somewhat more variation in space than in time. Thus, after removal of the trend, the NDVI residuals tend to vary more between locations in space than between instants in time.

The metric space–time variogram model was fitted using an exponential function for all three variogram components. Fitting was done using a quasi-Newton method with box constraints (Byrd et al. 1995). Quasi-Newton methods seek the minimum of a function by setting its gradient to zero and box constraints impose minima and maxima for the parameters in the search space (for instance, the nugget, sill and range parameter must all be equal to or greater than zero). The results are shown in Fig. 5.7b and also as the solid lines in Fig. 5.7a, b. The fitting was based on all space–time lags, which explains why the fitted curves do not reproduce the marginal variograms as well as might have been achieved if the only consideration had been to reproduce the marginal experimental variograms. The parameters of the variogram models are given in Table 5.2. The nugget variance is zero, which implies that the NDVI residual is perfectly correlated at very short distances in time and space. The spatial sill is the largest, which confirms that spatial variation in the NDVI residual dominates temporal variation. The space–time component has a moderate sill, although its contribution is not negligible.

5.5.3 Space–Time Kriging

Space–time kriging was done using the Gstat library (www.gstat.org) within the R statistical software (Bivand et al. 2008). A small subset only of the entire space–time cube of predictions and prediction error standard deviations are shown here. Maps of the predicted NDVI are shown for three time instants in Fig. 5.8. These show that NDVI is small on DOY 165, especially for Sofista, large on DOY 200 for both crops and small for Sofista and still large for Innovator on DOY 235. This agrees with the results presented in Fig. 5.3. The boundary between the two parcels is clearly visible in all the maps of predictions shown here, and so are the patches with small NDVI values near the parcel boundaries. A stripe effect is also apparent, particularly for DOY 165. This is probably an artefact caused by systematic measurement errors between the NDVI observations of neighbouring tramlines, although true systematic differences between neighbouring crop rows cannot be ruled out. One kriging

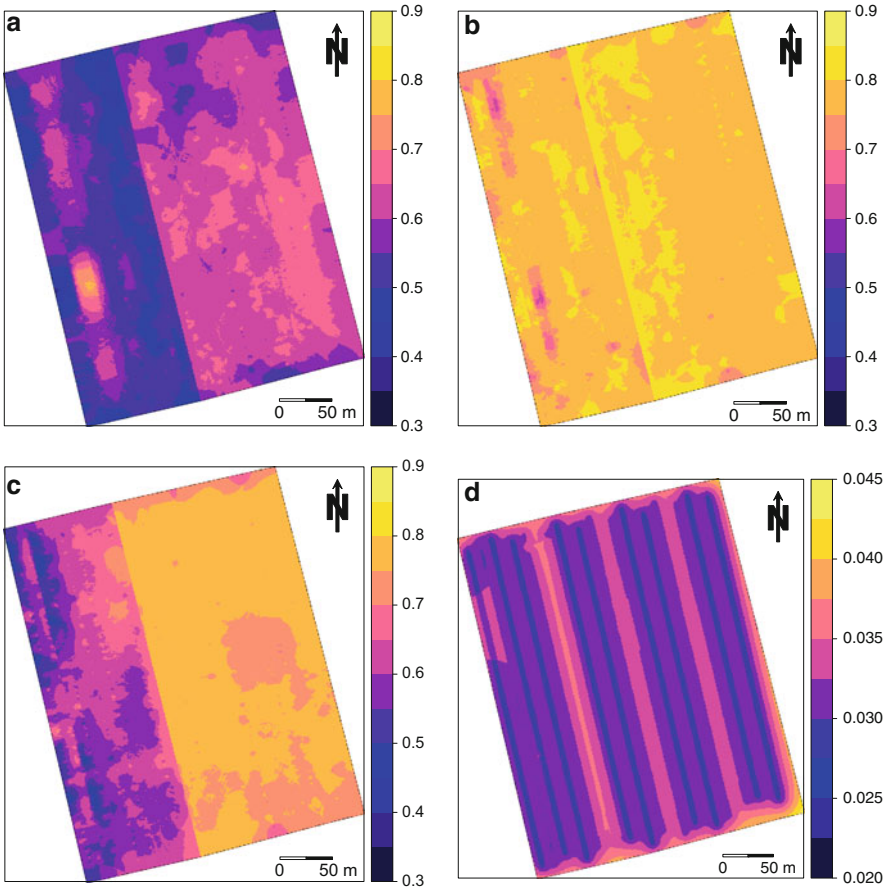

Fig. 5.8 Space–time kriged predictions for three arbitrary days: (**a**) DOY 165, (**b**) DOY 200 and (**c**) DOY 235; (**d**) kriging standard deviations for DOY 235

prediction error standard deviation map is shown for DOY 235 in Fig. 5.8d. The standard deviations are small compared to the predicted values (note the different legend entries) and are smaller near observations than further away from them. Note also that the standard deviation map indicates that only 9 of the 16 tramlines (see Fig. 5.8) were used near DOY 235 (see also Table 5.1). The standard errors are small because of the large number of observations and strong space–time correlations.

Space–time kriging results are also shown as time-series for three arbitrary locations in Fig. 5.9. Two of these are from the Sofista parcel where NDVI predictions become smaller towards the end of the observation period. Note also that the intervals between predictions are smaller near points in time where observations were made (i.e. one of the 16 days where NDVI was measured). For the corner location in the Sofista parcel, some observations fall outside the 95% prediction interval, and

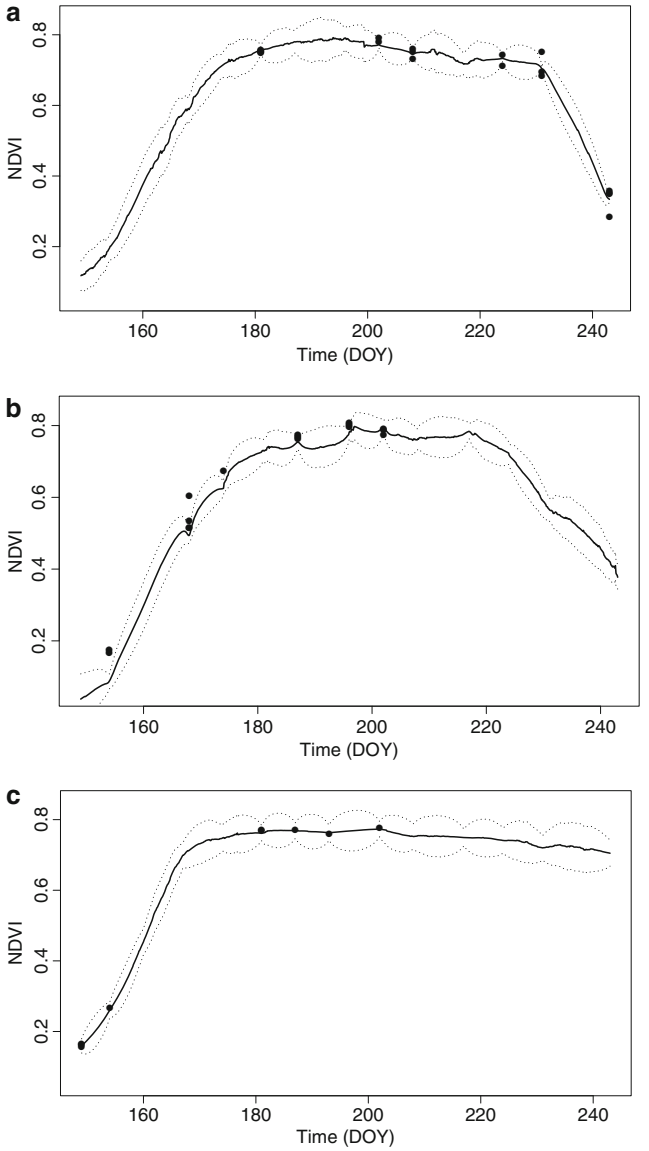

Fig. 5.9 Space–time kriging results for three arbitrary locations: (**a**) centre and (**b**) north-west corner of the Sofista parcel, and (**c**) centre of the Innovator parcel. The *solid line* is the regression kriging prediction, *dotted lines* represent the 95% prediction intervals derived from the kriging standard deviation. The dots are NDVI observations within a circular neighbourhood with radius 2 m

this happens more often than the expected 1 out of 20 cases. This might be caused by the fact that there may be a separation distance of up to 2 m between the observation and prediction location (a tolerance of 2 m was used to ensure that each graph

has observations), but may also be caused by the fact that outliers were removed before computing and modelling the space–time variogram (see Section 5.2), which led to a systematic underestimation of the variation.

5.6 Discussion and Conclusions

Space–time geostatistics is a useful extension to spatial geostatistics for precision agriculture because many of the variables addressed in precision agriculture vary in time as well as in space. Space–time kriging can provide predictions at a high space–time resolution and can be used, for example, to produce a time-series of spatial maps. When such maps are shown in animation mode these images may be useful to farmers to gain insight into crop growth. The data and results of the case study were only available after the growing season, but if available during the growing season the results can aid the farmer's understanding of how crop variety and management practices affect crop growth. It will show where and when within a parcel anomalies caused by drought, wetness or diseases occur. For instance, this will enable the farmer to adjust the chemical spraying against the potato disease Phytophtera in space and time according to the actual occurrence of the disease. Compared to the current practice of weather dependent precautionary spraying, this can reduce substantially the amount of chemicals needed for optimal crop growth. The trend analysis, which is part of the development of a space–time model, might also provide important insight into which factors influence the variation in space and time in target variables. This can contribute to a better timing of the final fertilizer N application, or to determine the best DOY to remove the aboveground green biomass at harvest.

Space–time kriging is not intended to make forecasts in time. The primary purpose of kriging is interpolation; it is not intended for extrapolation, be that in space or time. Extrapolation would result in large uncertainties, as exemplified by the kriging variance. However, when implemented in real-time mode, space–time kriging can provide farmers with almost instant imagery of crop- and soil-related properties for the past and present. Process-based modelling, possibly augmented with data assimilation functionality (e.g. Hoeben and Troch 2000; Heuvelink et al. 2006), is needed to make reliable forecasts. Data assimilation techniques merge information from observations with information from models, taking the relative uncertainty associated with each of the sources of information into account.

The Lauwersmeer farm case study addressed in this chapter was limited to NDVI, but the methodology is generic and applies to many more properties that are relevant to precision agriculture, provided sufficient observations are available. In addition, the spatial extent and time period are not restricted. The case study addressed spatial variation within a 10 ha field during the growing season, but much larger space–time domains can be handled too. In the case study we assumed a trend that was a combination of a constant-in-time crop variety effect and a constant-in-space seasonal effect. A more elaborate model would let these effects vary in space and time, would

consider interactions between effects and would make a more careful assessment of the effects of other factors, such as soil type, previous crop growth variation and terrain form. Also, effects from parcel boundaries and anisotropy resulting from tillage, planting and fertilizer application may be included.

Practical application of space–time kriging is not as straightforward as it is for spatial kriging. This is partly because modelling variation in space–time is more difficult than modelling that in space, and because 'off-the-shelf' software is not yet available. The software Gstat was used in the case study, but it is not tailored to space–time geostatistics and cannot handle non-metric variograms or 3D space–time prediction. For instance, some additional programming was needed to compute the experimental variogram. More flexible alternatives are GSLIB (Deutsch and Journel 1998) extensions (De Cesare et al. 2002) and SEKS-GUI (Yu et al. 2007). The modelling of space–time variation also needs further development, both in terms of the choice of structure of the space–time variogram and in terms of the improved and user-friendly estimation of its parameters. Recently, many advanced methods have been published in the statistical literature to define classes of valid space–time covariance structures (e.g. Cressie and Huang 1999; Gneiting 2002; Huang et al. 2007; Fuentes et al. 2008; Ma 2008). A comprehensive review of these methods is beyond the scope of this chapter, but it is important to note that the variogram modelling approach used in this chapter is only one of many possibilities. One particular problem that confronts the modelling is that observations are generally spread unevenly over space and time. Data sets will be generally sparse in space and abundant in time (e.g. Snepvangers et al. 2003). In precision agriculture applications, however, data may be sparse in time and abundant in space: in the case study there were only 16 measurements in time and hundreds of spatial measurements for each instant in time. The uneven spread of observations over the space–time domain complicates the variogram modelling process and subsequent kriging. In the future the lack of temporal data could well change in agriculture and other applications. For instance, high resolution satellite or aerial imagery such as Ikonos and radar maps of rainfall may become readily available and instantly downloadable at affordable prices. Also the application and further development of 'smart dust' soil temperature and moisture sensors could provide a more even distribution of abundant data in space and time in precision agriculture. Space–time kriging may become a valuable tool in precision agriculture and other applications, combined with the current developing practice of using continuous soil moisture sensors and high resolution soil maps to fine-tune irrigation or the use of precise soil temperature data to determine sowing and planting dates.

In parallel with the theoretical advances, practical application of space–time geostatistics also needs further development. More real-world applications of space–time kriging will not only develop the maturity of this subject area and encourage the development of user-friendly software, but it might also reveal stable patterns in the type of model to be used for specific applications. If this were the case, then a model structure developed in one year for a given area, soil type or crop variety may be used in future years for similar areas, soil types or crop varieties, thus saving time on the cumbersome modelling stage. Thus, the farmer may use the model

developed in the case study for future use on his farm to create space–time surfaces of NDVI in near real-time. This will allow him to delineate those parts of the parcel where additional management is required (e.g. parts where the NDVI has not increased sufficiently during the past fortnight) and take action. The incentive for the further development of space–time techniques must come primarily from the field of applications. Data availability in precision agriculture is ever increasing and reliable methods are needed to distil useful information from the data and present the information in a tangible and efficient way so that farmers and their advisers can comprehend and use it. In this respect, precision agriculture may well be an important application field that can stimulate and steer the further development of space–time geostatistics.

Acknowledgements We thank Mr. Claassen for use of the Lauwersmeer farm data. We thank Allard de Wit (Alterra) for his suggestion to use the double logistic function and assistance with fitting it to the NDVI data.

References

Actueel Hoogtebestand Nederland. (9 July 2009). http://www.ahn.nl/index.php.

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment, 35*, 161–173.

Bilonick, R. A. (1988). Monthly hydrogen in deposition maps for the northeastern U.S. from July 1982 to September 1984. *Atmospheric Environment*, *22*, 1909–1924.

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York, USA: Springer.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing, 16*, 1190–1208.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment, 62*, 241–252.

Cressie, N., & Huang, H. C. (1999). Classes of nonseperable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association, 94*, 1–53.

De Cesare, L., Myers, D. E., & Posa, D. (2001). Product-sum covariance for space–time modeling: An environmental application. *Environmetrics, 12*, 11–23.

De Cesare, L., Myers, D. E., & Posa, D. (2002). FORTRAN programs for space–time modeling. *Computers and Geosciences, 28*, 205–212.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.

Diggle, P. J., & Ribeiro Jr., P. J. (2007). *Model-based geostatistics*. New York: Springer.

Dimitrakopoulos, R. & Luo, X. (1994) Spatiotemporal modeling: Covariances and ordinary kriging systems In R. Dimitrakopoulos (Ed.), *Geostatistics for the next century* (pp. 88–93). Dordrecht: Kluwer.

Fischer, A. (1994). A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. *Remote Sensing of Environment, 48*, 220–230.

Fuentes, M., Chen, L. & Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics, 19*, 487–507.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association, 97*, 590–600.

Hengl, T., Heuvelink, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil properties based on regression-kriging. *Geoderma, 120*, 75–93.

Heuvelink, G. B. M., Schoorl, J. M., Veldkamp, A., & Pennock, D. J. (2006). Space–time Kalman filtering of soil redistribution. *Geoderma, 133*, 124–137.

Hoeben, R. & Troch, P. A. (2000). Assimilation of active microwave observation data for soil moisture profile estimation. *Water Resources Research, 36*, 2805–2819.

Holland Scientific (27 November 2008). www.hollandscientific.com.

Huang, H.-C., Martinez, F., Mateu, J, & Nontes, F. (2007). Model comparison and selection for stationary space–time models. *Computational Statistics & Data Analysis, 51*, 4577–4596.

Ma, C. (2008). Recent developments on the construction of spatio-temporal covariance models. *Stochastic Environment Research and Risk Assessment, 22*, S39–S47.

Montgomery, D. C., Peck, E. A, & Vining, G. G. (2006). *Introduction to linear regression analysis* (4th ed). New York: Wiley.

Snepvangers, J. J. J. C., Heuvelink, G. B. M., & Huisman, J. A. (2003). Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma, 112*, 253–271.

Webster, R. (2000). Is soil variation random? *Geoderma, 97*, 149–163.

Wösten, J. H. M., Veerman, G. J., De Groot, W. J. M., & Stolte, J. (2001). *Waterretentie- en doorlatendheidskarakteristieken van boven- en ondergronden in Nederland: de Staringreeks*. Vernieuwde uitgave 2001. Wageningen, The Netherlands: Alterra, Report 153.

Wu, J., Wang, D., & Bauer, M. E. (2007). Assessing broadband vegetation indices and QuickBird data in estimating leaf area index of corn and potato canopies. *Field Crops Research, 102*, 33–42.

Yu, H.-W., Kolovos, A., Christakos, G., Chen, J.-C., Warmerdam, S., & Dev, B. (2007). Interactive spatiotemporal modelling of health systems: the SEKS–GUI framework. *Stochastic Environment Research and Risk Assessment, 21*, 555–572.

Chapter 6
Delineating Site-Specific Management Units with Proximal Sensors

D.L. Corwin and S.M. Lesch

Abstract Conventional farming manages fields uniformly with no consideration for spatial variation. This causes reduced productivity, misuse of finite resources (e.g. water and fertilizers) and detrimental impacts on the environment. Site-specific management units (SSMUs) have been proposed as a way of resolving the spatial variation of various factors (i.e. soil, climate, management, pests, etc.) that affect variation in crop yield. Mobile proximal sensors, such as those used to measure apparent soil electrical conductivity (EC_a), can be used to characterize the spatial variation of soil properties that affect crop yield. This Chapter provides an overview of the work by the authors that has led to the delineation of SSMUs based on edaphic and anthropogenic properties, with particular emphasis given to the geostatistical techniques needed to direct soil sampling to characterize the spatial variation. The approach uses geospatial proximal sensor measurements to locate the positions of soil samples to characterize the variation in soil properties that affect crop yield within a field. A crop yield response model is developed and maps of SSMUs based on soil and crop yield information are produced. The methodology for delineating SSMUs can be used whenever the proximal sensor measurements correlate with yield. Maps of SSMUs provide the vital information for variable-rate technology (e.g. site-specific fertilizer and irrigation water application).

Keywords Soil salinity · Apparent electrical conductivity (ECa) · ECa-directed sampling · Response surface sampling design · Electromagnetic induction · Electrical resistivity

D.L. Corwin (✉)
USDA-ARS U.S. Salinity Laboratory, 450 West Big Springs Road, Riverside, CA 92507-4617, USA
e-mail: Dennis.Corwin@ars.usda.gov

S.M. Lesch
Riverside Public Utilities, Resource Division, 3435 14th Street, Riverside, CA 92501, USA
e-mail: SLesch@riversideca.gov

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_6,

6.1 Introduction

6.1.1 The Need for Site-Specific Management

Tremendous strides have been made to expand the world's supply of food. Even though the world population has doubled over this time period, food production has risen even faster with per capita food supplies increasing from less than 2000 calories per day in 1962 to more than 2500 calories in 1995 (World Resources Institute 1998). The rise in global food production has been credited to better seeds, expanded irrigation, and greater fertilizer and pesticide use, commonly referred to as the Green Revolution. However, the prospect of feeding a projected additional 3 billion people over the next 30 years poses more challenges than have been encountered in the past 30 years. In the short term, global resource experts predict that there will be adequate global food supplies, but the distribution of those supplies to malnourished people will be the primary problem. Longer term, however, the obstacles become more formidable, though not insurmountable. Although total yields continue to rise on a global basis, there is a disturbing decline in the growth of yield with some major crops such as wheat and maize reaching a 'yield plateau' (World Resources Institute 1998).

Sustainable agriculture is viewed as the most viable means of meeting the food demands of the projected world's population, barring unexpected technological breakthroughs. The concept of sustainable agriculture is predicated on a delicate balance of maximizing crop productivity to keep pace with population growth and maintaining economic stability while minimizing the use of finite natural resources (e.g. water, fertilizers and pesticides) and the detrimental environmental impacts of associated agrichemical pollutants. Arguably, the most promising approach for attaining sustainable agriculture is precision agriculture or site-specific crop management.

Site-specific crop management, or more specifically site-specific management (SSM) attempts to manage the soil, pests and crops based upon spatial variation within a field (Larson and Robert 1991), whereas conventional farming treats a field uniformly, ignoring the naturally inherent variability of soil and crop conditions between and within fields. There is well-documented evidence that spatial variation within a field is highly significant and amounts to a factor of 3–4 or more for crops (Birrel et al. 1995; Verhagen et al. 1995) and up to an order of magnitude or more for soil (Corwin et al. 2003a). Specifically, SSM is the management of agricultural crops at a spatial scale smaller than the whole field that takes account of local variation to cost effectively balance crop productivity and quality, detrimental environmental impacts and the use of resources (e.g. water, fertilizer, pesticides, etc.) by applying them when, where and in the amount needed. Spatial variation in crops is the result of a complex interaction of biological (e.g. pests, earthworms, microbes), edaphic (e.g. salinity, organic matter, nutrients, texture), anthropogenic (e.g. leaching efficiency, soil compaction due to farm equipment), topographic (e.g. slope, elevation) and climatic (e.g. relative humidity, temperature, rainfall) factors.

6.1.2 Definition of Site-Specific Management Unit (SSMU)

Site-specific management units (SSMUs) have been proposed as a means of dealing with the spatial variation of edaphic (i.e. soil related) properties that affect crop productivity (or quality) to achieve the goals of SSM. A SSMU is simply a mapped unit within a field that could be based on soil properties, landscape units, past yield, etc. that is managed to achieve the goals of SSM. To manage within-field variation site-specifically, geo-referenced areas (or units) that are similar with respect to a specified characteristic must be identified (van Uffelen et al. 1997). Ideally, a site-specific management unit (SSMU) will account for the spatial variation of all factors that affect variation in crop yield, including edaphic, meteorological, biological, anthropogenic and topographic factors. To achieve this, the delineation of SSMUs would be extremely complicated because all these must be considered. One means of simplifying the complexity is to delineate SSMUs based on a single factor, such as edaphic properties, and determine the extent of variation in yield related to this factor.

The extent and conditions under which these spatial patterns are stable should also be established. Yield maps provide information on the integrated effects of the physical, chemical, and biological processes under certain weather conditions (van Uffelen et al. 1997), and the spatial patterns of crop productivity provide a basis for implementing SSM by indicating where varying crop inputs are needed (Long 1998). However, the inputs required to optimize crop productivity and minimize impacts on the environment can be determined only if the factors that gave rise to the observed spatial crop patterns are known (Long 1998). Yield maps alone cannot provide information to distinguish between the various sources of variation and cannot give clear guidelines for management without information on the effects of variation in weather, pests and diseases, and soil physical and chemical properties on the variability of a crop for a particular year (van Uffelen et al. 1997). Each factor that affects within-field variation in yield needs to be characterized spatially to be able to manage a crop on a site-specific basis. The spatial characterization of these factors can be achieved with spatial measurements from a spectrum of proximal sensors.

6.1.3 Proximal Sensors

Ground-based proximal sensors generally include sensors that take measurements from within a distance of 2 m from the soil surface. They may take measurements of the soil, such as electrical, electromagnetic or radiometric sensors, or of plants, such as crop yield or spectral sensors. Adamchuk et al. (2004) reviewed on-the-go proximal soil sensors for precision agriculture and Barnes et al. (2003) provided a concise review of ground-based sensor techniques as well as remote imagery sensors for mapping soil properties.

According to Adamchuk et al. (2004), proximal sensors fall into six main categories: electrical and electromagnetic, optical and radiometric, mechanical, acoustic,

Table 6.1 Selected recent references using proximal soil sensors to map soil properties for applications in precision agriculture. Modification of tables from Adamchuk et al. (2004)

Category of proximal sensor	Review article	Sensor	Technical reference
Electrical and EMI	Corwin and Lesch (2005a)	ER	Corwin and Lesch (2003)
		EMI	Corwin and Lesch (2005b,c)
		Capacitance	Andrade et al. (2001)
Optical	Ben-Dor et al. (2009)[a]	Single wavelength	Shonk et al. (1991)
		Multi- or Hyperspectral	Maleki et al. (2008), Mouazen et al. (2007)
Radiometric	Huisman et al. (2003)	GPR	Lunt et al. (2005)
		Microwave	Whalley and Bull (1991)
Mechanical	Hemmat and Adamchuk (2008)	Draft	Ehrhardt et al. (2001), Mouazen and Roman (2006)
		Load cells and penetrometers	Chung et al. (2003), Verschoore et al. (2003)
Acoustic and pneumatic		Microphone	Liu et al. (1993)
		Air pressure transducer	Clement and Stombaugh (2000)
Electrochemical		ISFET	Birrell and Hummel (2001), Viscarra Rossel and Walter (2004)
		ISE	Adamchuk et al. (2005), Sethuramasamyraja et al. (2008)

EMI, electromagnetic induction; ER, electrical resistivity; GPR, ground penetrating radar; ISFET, ion-selective field effect transistor; ISE, ion-selective electrode.
[a]Review includes remote and proximal sensors.

pneumatic and electrochemical. Several studies have been conducted using proximal sensors with just a few of the more current ones listed in Table 6.1. The output from each sensor is typically affected by more than one agronomic soil property. Table 6.2 outlines the soil properties influencing each category of proximal sensor.

Electrical and electromagnetic sensors include electrical resistivity (ER), electromagnetic induction (EMI), time domain reflectometry (TDR) and capacitance sensors. The most commonly used for field-scale on-the-go measurements are ER and EMI (Corwin and Lesch 2005a). Electrical resistivity and EMI measure the electrical conductivity of the bulk soil, which is referred to as the apparent soil electrical conductivity (EC_a). Corwin and Lesch (2005a) have provided a review of EC_a measurements in agriculture. Apparent soil electrical conductivity is affected by a variety of soil properties including salinity, texture, water content, organic matter, cation exchange capacity (CEC) and bulk density (Corwin and Lesch

Table 6.2 Soil properties that influence proximal sensors. Modified from Adamchuk et al. (2004)

	Agronomic soil property									
Category of proximal sensor	Texture (sand, silt, clay content)	OM	θ	EC or Na	Cp or ρ_b	Depth of topsoil or hard pan	pH	Residual NO_3 or total N	Other macro-nutrients	CEC
Electrical and EMI	X	X	X	X	X	X		X		X
Optical and radiometric	X	X	X				X	X		X
Mechanical					X	X				
Acoustic and pneumatic	X				X	X				
Electrochemical				X			X	X	X	

EMI, electromagnetic induction; OM, soil organic matter; θ, water content; EC, electrical conductivity (salinity); Na, sodium content; Cp, compaction; ρ_b, bulk density; CEC, cation exchange capacity.

2005a). Capacitance sensors and TDR use the dielectric constant or relative permittivity to infer the volumetric water content. There are commercially available on-the-go ER (e.g. Veris 3100) and EMI units (e.g. Geonics EM38-MK2).

Optical sensors comprise single wavelength and hyperspectral reflectance sensors, whereas radiometric sensors include microwave sensors and ground penetrating radar (GPR). Like electrical and electromagnetic sensors, optical and radiometric sensors are frequently influenced by a variety of soil properties (see Table 6.2). However, there is a potential advantage of optical and radiometric measurements in that the response in different parts of the spectral range may be affected to varying degrees by different soil properties, enabling the separation of effects (Adamchuk et al. 2004). As indicated by Baumgardner et al. (1985), soil reflectance is influenced by a variety of properties including parent material, salts, iron oxides, organic matter, particle size, moisture and mineral composition. Radiometric sensors have been widely used to establish the spatial distribution of soil water content.

Mechanical sensors such as a strain gauge, load cell, or horizontal cone and wedge penetrometer are used to measure soil mechanical resistance or soil compaction, which in turn provides information on soil moisture, texture and bulk density. Similarly, acoustic and pneumatic sensors have been correlated to soil texture (Liu et al. 1993) and compaction (Clement and Stombaugh 2000).

Electrochemical sensors use either an ion-selective electrode (ISE) or ion-selective field effect transistor (ISFET) to provide a direct means of measuring pH or nutrient content (e.g. K^+ or NO_3^-) to evaluate soil fertility. Electrochemical sensors have the distinct disadvantage of requiring a significant amount of time for equilibrium between the sensor and the soil or soil solution.

To a varying extent from one field to the next, crop patterns are affected by edaphic properties. Bullock and Bullock (2000) indicated that efficient methods for measuring within-field variation accurately in soil physical and chemical properties are important for precision agriculture. No single sensor will measure all the soil properties that affect crop yield variation; therefore, combinations of sensors are recommended, resulting in a mobile multi-sensor platform. Of all of the proximal sensors, EMI and ER sensors are arguably the most thoroughly researched and commonly used for measuring the edaphic properties that affect crop yield (Corwin and Lesch 2003, 2005a).

6.1.4 Objective

This chapter aims to provide the general knowledge and understanding to delineate SSMUs based on edaphic and anthropogenic factors influencing crop yield that have been identified and spatially defined using geo-referenced proximal sensor data. Because the measurement of EC_a is one of the most widely used and well-understood soil measurements (Corwin and Lesch 2003, 2005a), it has been singled out in this chapter to represent ground-based proximal sensors. However, the

methodology that is described in this chapter for delineating SSMUs can be applied to any of the sensors. In addition, this chapter illustrates the use of spatial and geostatistical analysis to calibrate and interpret geo-referenced proximal sensor data.

6.2 Directed Sampling with a Proximal Sensor

6.2.1 Complexity of Proximal Sensor Measurements and the Role of Geostatistics

Numerous studies have related proximal sensors to crop yield (or quality) in a precision agriculture context. A short list of some recent proximal sensor studies associated directly with SSM includes Adamchuk et al. (2007), Yan et al. (2007a, b), Corwin et al. (2008), Vitharana et al. (2008), Morari et al. (2009), as well as those listed in Table 6.1.

Corwin and Lesch (2003) warned of the complexity of proximal sensor measurements, specifically spatial measurements of EC_a, and provided guidance for the application of EC_a to precision agriculture. However, even now some of the most recent proximal sensor studies demonstrate a lack of understanding of the complexity of proximal sensor measurements. For example, the work by Yan et al. (2007a, b) relates yield to EC_a rather than to the edaphic properties affecting the EC_a measurement that concomitantly influence crop yield (or crop quality). By basing SSMUs directly on EC_a, rather than on the properties affecting its measurement at a field site, SSMUs can be defined erroneously, in particular where more than one soil property dominates the EC_a measurement and affects crop yield or quality. In addition, basing SSMUs on EC_a rather than on the properties that affect it does not enable associated management recommendations because increases or decreases in EC_a involve changes in all the properties affecting it at a particular site.

Because proximal sensors are typically affected by more than one agronomic property (i.e. soil- or plant-related properties), spatial measurements with proximal sensors are best used to develop a sampling plan to characterize the spatial distribution of those properties that affect the sensor and that, in turn, influence crop yield (or quality). The proximal sensor directed sampling approach aims to identify sample locations that reflect the range and variability of agronomic properties that affect the sensor measurement. Apparent soil electrical conductivity is not the property that affects crop yield (or quality); rather it is the edaphic properties influencing EC_a (i.e. salinity, water content, texture, organic matter, bulk density) that directly affect crop yield (or quality). Nevertheless, information from the proximal sensor can be used to direct soil (or plant) sampling. Spatial statistics plays a crucial role in establishing the sampling locations from geo-referenced proximal sensor data from which soil (or plant) properties that directly affect yield are determined. It is these latter data that enable the delineation of SSMUs with their associated management recommendations to maximize yield (or quality).

6.2.2 Practical Consideration of Differences in Support

Differences in support are important when using proximal sensors to direct soil (or plant) sampling for site-specific management. First there is a difference in support between the proximal sensor (few m^2 or less) and yield (generally tens of m^2) measurements, and between the soil (or plant) sample volume (0.075 m^3) and the proximal sensor's volume of measurement (e.g. Geonics EM38 measures roughly 1–1.5 m^3). In many respects differences in support are strongly influenced by practical considerations of resources (i.e. time, labor and cost). As a rule-of-thumb, a minimum number of samples needs to be taken at each scale to enable a comparison of local (a few metres) and field-scale variation (tens to hundreds of metres). For example, where local-scale variation is significantly less than field-scale variation sampling directed by a proximal sensor will be viable, but as the scale of local variation approaches the observed field-scale variation, the approach becomes less tenable. In other words, the proximal sensor can resolve local variation because of its support and intensity of measurement, whereas the yield monitor can resolve only the larger scale variation that occurs within fields. For the soil and plant samples, regardless of support, the variation that they resolve will depend on the intensity of sampling, which cannot be as intensive as the sensor because of practical considerations.

6.3 Delineation of SSMUs with a Proximal Sensor

6.3.1 Geostatistical Mixed Linear Model

In a typical field survey where proximal sensor readings such as EC_a are recorded, the sensor data are often used to help predict a specific, unobserved soil property. For instance, assume a dense grid of proximal sensor data has been acquired across a field and soil samples have been taken at some locations so that the data from both sources can be used to estimate a model that can predict the detailed spatial pattern of the soil property measured by or correlated with the proximal sensor measurement. Assume that the relationship between the soil property measurement and sensor data can be approximated adequately using the following geostatistical mixed linear model (Haskard et al. 2007):

$$\mathbf{y} = \mathbf{X}\beta + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{6.1}$$

where $\mathbf{y}$ represents an $(n \times 1)$ vector of observed soil property data, $\mathbf{s}$ is the corresponding vector of paired (s_x, s_y) survey location coordinates, $\mathbf{X}$ represents an $(n \times p)$ fixed data matrix that includes observed functions of sensor readings and possibly also the coordinates, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown parameter estimates, $\eta(\mathbf{s})$ represents a zero mean, second-order stationary spatial Gaussian error process and $\boldsymbol{\varepsilon}(\mathbf{s})$ is a vector of jointly independent normal $(0, \sigma_n^2)$ random variables. Typical

stationary spatial structures for $\eta(\mathbf{s})$ are well documented in the spatial statistical and geostatistical literature; examples in two dimensions include the isotropic and anisotropic exponential and spherical covariance structures, as well as the Matérn class of covariance functions (Cressie 1993; Wackernagel 1998; Schabenberger and Gotway 2005; Webster and Oliver 2007). Note also that the second $\boldsymbol{\varepsilon}(\mathbf{s})$ error component is usually referred to as the 'nugget' effect in geostatistics (Webster and Oliver 2007).

Equation 6.1 represents a versatile spatial linear prediction model that can incorporate various types of modelling assumptions. The deterministic component of the model ($\mathbf{X}\boldsymbol{\beta}$) can be defined to include trend surface parameters and or additional collocated soil-property measurements, in addition to various hypothesized target property and sensor relationships. As noted above, the stochastic error terms ($\eta(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s})$) can be parameterized to match the geostatistical covariance functions commonly used in kriging. Indeed, Eq. 6.1 is identical to universal kriging when ($\mathbf{X}\boldsymbol{\beta}$) contains only trend surface parameters, and kriging with external drift when ($\mathbf{X}\boldsymbol{\beta}$) contains only sensor readings. In addition, both ordinary kriging and regression kriging models can also be derived as special cases of Eq. 6.1 (Schabenberger and Gotway 2005; Haskard et al. 2007).

In the most general case, ($\mathbf{X}\boldsymbol{\beta}$) may contain multiple fixed effects and the residual errors are assumed to be spatially autocorrelated. Assume that the corresponding residual errors follow a Gaussian (e.g. multivariate normal) distribution defined as

$$\begin{aligned}
&\eta(\mathbf{s}) \approx G(\mathbf{0}, \sigma_s^2 \mathbf{C}(\theta)), \\
&\varepsilon(\mathbf{s}) \approx G(\mathbf{0}, \sigma_n^2 \mathbf{I}), \\
&\operatorname{cov}\{\eta(\mathbf{s}), \boldsymbol{\varepsilon}(\mathbf{s})\} = \mathbf{0} \\
&\Rightarrow \\
&\operatorname{var}\{\eta(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s})\} = \sigma_s^2 \mathbf{C}(\theta) + \sigma^2 n \mathbf{I} = \Sigma,
\end{aligned} \tag{6.2}$$

where Σ is assumed to be positive definite and $\mathbf{C}(\theta)$ represents the correlation function of a second-order stationary error process (for example, $\mathbf{C}(\theta)$ could represent an isotropic exponential correlation function with range parameter θ). When the covariance structure is known up to a proportionality constant in the geostatistical mixed linear model (i.e. $\Sigma = \tau^2 \mathbf{V}$, where $\mathbf{V}$ is assumed to be known a priori), $\boldsymbol{\beta}$ of Eq. 6.1 can be estimated by generalized least squares (Rao and Toutenburg 1995). However, the specific Σ hyper-parameter values are rarely known a priori. In practice, $\boldsymbol{\beta}$ and the variance structure $\boldsymbol{\Sigma}$ are jointly estimated from the sample data, typically by maximum likelihood (ML) or residual maximum likelihood (REML) estimation (Littell et al. 1996; Lark et al. 2006). The ML or REML Σ hyper-parameter estimates are then returned to the model to compute the fixed effect parameter estimates, $\boldsymbol{\beta}$, and model predictions.

Conditional on a known covariance structure, standard mixed linear modelling theory (Cressie 1993) can be used to show that the best linear unbiased estimator for β is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}} \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \Sigma^{-1} \mathbf{y}, \tag{6.3}$$

with a corresponding variance of

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{\mathrm{T}}\Sigma^{-1}\mathbf{X})^{-1}. \tag{6.4}$$

Likewise, one can show that the best linear unbiased prediction for $\mathbf{y}_z$ (where $\mathbf{y}_z$ represents the remaining (non-sampled) survey locations can be expressed as

$$\hat{\mathbf{y}}_z = (\mathbf{X}_z\boldsymbol{\beta} + \Sigma_{yz}\Sigma^{-1}(\mathbf{y} - \mathbf{X}), \tag{6.5}$$

where $\mathbf{X}_z$ represents the design matrix associated with $\mathbf{y}_z$ and Σ_{yz} represents the model-based covariance matrix between $\mathbf{y}_z$ and the observed sample data $\mathbf{y}$. In addition, the corresponding variance estimate associated with this prediction vector is

$$\begin{aligned}\text{var}(\mathbf{y}_z - \hat{\mathbf{y}}_z) = {} & \boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{yz}^{\mathrm{T}} \\ & + \left[\mathbf{X}_z - \boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]\left(\mathbf{X}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\left[\mathbf{X}_z - \boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]^{\mathrm{T}},\end{aligned} \tag{6.6}$$

where Σ_z represents the model-based variance matrix of $\mathbf{y}_z$ (Cressie 1993). Once again, these predictions and variance estimates are identical to those obtained from universal kriging and or kriging with external drift models (when the design matrix is specified appropriately to give such models).

6.3.2 *Soil Sampling Strategies Based on Geo-Referenced Proximal Sensor Data*

A minimum number of sites for soil (or plants) must be sampled to calibrate the geostatistical mixed linear model following the proximal sensor survey. In general, the most common strategies currently used can be classified as either probability-based (design-based) or prediction-based (model-based) sampling approaches. A brief description of each of these approaches is given below.

Probability sampling includes techniques such as simple random, stratified random and cluster sampling. Thompson (1992) provides a review of these. Probability sampling has a well developed underlying theory (Thompson 1992; Brus and de Gruijter 1993), but it was not designed specifically for estimating models. Indeed, most probability sampling strategies explicitly avoid incorporating any parametric modelling assumptions; they rely instead on the principles of randomization that are built into the design for drawing statistical inference.

Prediction-based sampling strategies, which are adopted in geostatistics and time-series analysis, are focused explicitly towards model estimation. The underlying theory behind this approach for finite population sampling and inference is discussed in detail in Valliant et al. (2000). More generally, response surface and optimal experimental design theory are closely related areas of statistical research in which sampling designs are studied specifically from the viewpoint of model estima-

tion (Myers and Montgomery 2002). Techniques from these two subject areas have been applied to the optimal collection of spatial data by Müller (2001), the specification of optimal designs for variogram estimation by Müller and Zimmerman (1999), the estimation of spatially referenced regression models by Lesch et al. (1995) and Lesch (2005), and the estimation of geostatistical linear models by Zhu and Stein (2006) and Brus and Heuvelink (2007). Conceptually similar types of non-random sampling designs for variogram estimation have been introduced by Russo (1984) and Warrick and Myers (1987).

Sampling on a grid has been used for many years in soil science; however, it is not strictly randomized even when a random starting point is used. As a consequence there is no direct way of estimating the standard errors of the mean from a design-based viewpoint. Grid sampling has generally been favored in model-based sampling designs and has also been commonly used in precision agriculture because it is easy to implement and results in an even distribution of sample sites. Grid sampling is often used when kriging is to be used for analysis and mapping because it is an effective way to minimize the average interpolation error (Burgess et al. 1981; Burgess and Webster 1984).

Theoretically, any of the above sampling approaches can be used to estimate a spatial or geostatistical model, although each approach has various strengths and weaknesses. Lesch (2005) compares and contrasts probability- and prediction-based sampling strategies in more detail, and highlights some of the strengths of the prediction-based sampling approach.

The prediction-based sampling approach discussed by Lesch (2005) was designed specifically for use with ground-based EC_a sensor readings. A minimum number of samples for calibration is selected based on the observed magnitudes and spatial locations of the EC_a data. These sites are chosen in an iterative, non-random way to (i) optimize the estimation of a regression model (i.e. minimize the mean square prediction errors produced by the calibration function) and (ii) maximize simultaneously the average separation between adjacent sampling locations to reduce the possibility of spatially correlated residual errors. Intuitively, this sampling approach represents a hybrid of a response surface sampling technique (Myers and Montgomery 2002) with a space-filling algorithm (Müller 2001). Lesch (2005) demonstrated that such a sampling approach can substantially outperform probability-based sampling with respect to several important model-based prediction criteria, particularly optimal estimation of the fixed-effect part of a spatial (or geostatistical) linear model. Response surface sampling design software, known as ESAP, has been developed specifically for use with EC_a measurements and other proximal sensors (Lesch et al. 2000). See http://www.ars.usda.gov/services/software/software.htm for this open access software.

There are two main advantages of the response surface approach. First, the number of samples required for estimating a calibration function can be reduced substantially in comparison to more traditional design-based sampling. Response surface designs are commonly used to minimize the estimation variance of linear statistical models in the non-spatial setting. Second, this approach lends itself naturally to the analysis of proximal sensor data. Indeed, many types of ground-,

airborne- and satellite-based remotely sensed data are often collected specifically because one expects them to correlate strongly with some property of interest (e.g. crop stress, soil type, soil salinity, etc.). Nevertheless, the exact parameter estimates associated with the calibration model may still need to be determined by some type of site-specific sampling design. The response surface approach explicitly optimizes this site selection process.

6.3.3 Applications of Geostatistical Mixed Linear Models to Proximal Sensor Directed Surveys

Geostatistical mixed linear models can be used effectively to delineate SSMUs using one of two approaches. In the first (and more common) approach, the model is used directly to map one or more specific soil (or plant) properties. Such an approach is useful when the SSMU can be defined effectively by only a few properties, and each of these properties correlates reasonably well with the sensor readings. Some well-known examples of application include the mapping of field-scale soil salinity and or soil texture patterns, typically for leaching or reclamation of the soil using EC_a measurements. Corwin and Lesch (2005b, c) and Lesch (2005) discuss the survey protocols associated with this approach in detail, together with various case studies.

When a geostatistical mixed linear model is used to produce detailed maps of just one or two primary soil (or plant) properties by direct prediction using proximal sensor data, the delineation of SSMUs is straightforward. For a single property, the resulting map defines the SSMU boundaries. Likewise, if two or three properties are considered, a GIS overlay (or similar operation) of the predicted values can usually be used to define and determine the SSMUs. Note that the 'optimal' boundaries and or size of the units are nearly always application specific and subject to the operational constraints of the associated farming management practices.

In the second approach, proximal sensor data are again used to direct soil (or plant) sampling. Soil (or plant tissue) from the selected sampling locations is then analysed for several secondary soil chemical and physical properties (or plant properties), and it is these measurements that are used for prediction in the geostatistical model. This approach was originally suggested by Corwin and Lesch (2003); it is well suited for determining the primary SSMUs influencing a crop response function. Note that in this case the proximal sensor data are not used directly in the geostatistical model as explicit predictor variables. Rather, the model relates the collocated soil chemical and physical properties (or plant properties) to the crop response levels, which enables us to relate the SSMUs better to these individual properties. It is the secondary soil properties that affect EC_a (i.e. salinity, water content, etc.) that are used as the predictor variables, rather than the sensor data themselves.

If the geostatistical model is used to estimate a crop response equation, which in turn is a function of measured soil chemical and physical properties, the delineation of the SSMUs can become more complex. Crop response equations can often include many different soil chemical and physical property effects, and these

individual effects may not all be spatially well defined or easily predicted from the sensor data. In addition, the overlaying of many soil properties tends to produce overly complex mosaic maps that are not easily interpreted or delineated into contiguous SSMUs (see Chapter 8). In such a situation, considerable subjective intuition may be needed to define a useful set of SSMUs.

6.4 Case Study Using Apparent Soil Electrical Conductivity (EC_a) – San Joaquin Valley, CA

The objective of this case study is (i) to use an intensive EC_a survey to direct soil sampling and to identify edaphic properties that affect cotton yield and (ii) to use this spatial information to make recommendations for SSM of cotton by delineating SSMUs based solely on the edaphic and anthropogenic properties that affect cotton yield. This paper draws from previous more detailed work conducted and published by Corwin and colleagues (Corwin and Lesch 2003, 2005b; Corwin and Lesch 2003).

6.4.1 Materials and Methods

6.4.1.1 Study Site

The study site is a 32.4 ha field in the Broadview Water District on the west side of the San Joaquin Valley in central California. The soil at the site is a Panoche silty clay (thermic Xerorthents), which is slightly alkaline with good surface and subsurface drainage. The subsoil is thick, friable, calcareous, and easily penetrated by roots and water. In the arid southwestern USA the primary soil properties influencing crop yield are salinity, soil texture and structure, plant-available water, trace elements (particularly B), and ion toxicity from Na^+ and Cl^- (Tanji 1996).

6.4.1.2 EC_a-Directed Soil Sampling Protocols for Site-Specific Management

General survey protocols for EC_a-directed soil sampling developed by Corwin and Lesch (2005b, c) were followed to characterize soil spatial variation. The basic elements of a field-scale EC_a survey applied specifically to precision agriculture include: (i) site description and EC_a survey design, (ii) geo-referenced EC_a data collection, (iii) soil sampling strategies based on geo-referenced EC_a data, (iv) soil sample collection, (v) physical and chemical analysis of pertinent soil properties, (vi) statistical and spatial analysis, (vii) geographic information system (GIS) database development and (viii) approaches for delineating SSMUs. The basic steps within each component are outlined in Table 6.3 and discussed in detail in Corwin and Lesch (2005b).

Table 6.3 Outline of steps for an EC_a field survey for precision agriculture applications. (Modified from Corwin and Lesch 2005b)

1. Site description and EC_a survey design
   (a) Record site metadata
   (b) Define project's or survey's objective
   (c) Establish site boundaries
   (d) Select GPS coordinate system
   (e) Establish EC_a measurement intensity
2. EC_a data collection with mobile GPS-based equipment
   (a) Geo-reference site boundaries and significant physical geographic features with GPS
   (b) Measure geo-referenced EC_a data at the pre-determined spatial intensity and record associated metadata
3. Soil sampling strategies based on geo-referenced EC_a data
   (a) Statistically analyse EC_a data using an appropriate statistical sampling design to establish the soil sample site locations
   (b) Establish sampling depth, sample depth increments and number of cores per site
4. Soil core sampling at specified sites designated by the sample design
   (a) Obtain measurements of soil temperature through the profile at selected sites
   (b) At randomly selected locations obtain duplicate soil cores within a 1-m distance of one another to establish local-scale variation of soil properties
   (c) Record soil core observations (e.g. mottling, horizonation, textural discontinuities, etc.)
5. Laboratory analyses of appropriate soil physical and chemical properties defined by project objectives
6. Statistical and spatial analyses to determine the soil properties that affect EC_a and crop yield (provided EC_a correlates with crop yield):
   (a) Perform a basic statistical analysis of physical and chemical data by depth increment and by composite depths
   (b) Determine the correlation between EC_a and physico-chemical soil properties by depth increment and by composite depths
   (c) Determine the correlation between crop yield and physical/chemical soil properties by depth and by composite depths to determine depth of concern (i.e. depth with consistently highest correlation, whether positive or negative, of soil properties to yield) and the soil properties that have a significant effect on crop yield (or crop quality)
   (d) Conduct an exploratory graphical analysis to determine the relationship between the significant physical and chemical properties and crop yield (or crop quality)
   (e) Formulate a spatial linear regression (SLR) model that relates soil properties (independent variables) to crop yield or crop quality (dependent variable)
   (f) Adjust this model for spatial autocorrelation, if necessary, using residual maximum likelihood (REML) or some other technique
   (g) Conduct a sensitivity analysis to establish dominant soil property affecting yield or quality
7. GIS database development and graphic display of spatial distribution of soil properties
8. Approaches for delineating site-specific management units

For the protocols to be applicable to SSM, EC_a must be correlated to crop yield (or quality), which would indicate that EC_a is measuring some edaphic property (or properties) that affect crop yield (or quality). The correlation coefficient (r) for yield and EC_a was $r = 0.51$ ($p < 0.01$).

6.4.1.3 Yield Monitoring and EC_a Survey

Spatial variation of cotton yield was measured at the study site in August 1999 using a four-row cotton picker equipped with a yield sensor and global positioning system (GPS). The yield sensors measured average seed cotton yield. All subsequent references to cotton yield are with respect to seed cotton yield. A total of 7706 cotton yield readings were recorded (Fig. 6.1a). Each yield observation represented an area of approximately 42 m^2. From August 1999 to March 2000 the field was fallow.

On March 2000 an intensive EC_a survey was conducted using mobile fixed-array electrical resistivity equipment developed by Rhoades and colleagues (Rhoades 1992; Carter et al. 1993) that measured EC_a at 9-m intervals (4000 EC_a readings). The fixed-array electrodes were spaced to measure EC_a to a depth of 1.5 m using a Wenner array electrode configuration with an inter-electrode spacing of 1.5 m. A map of the EC_a measurements is shown in Fig. 6.1b.

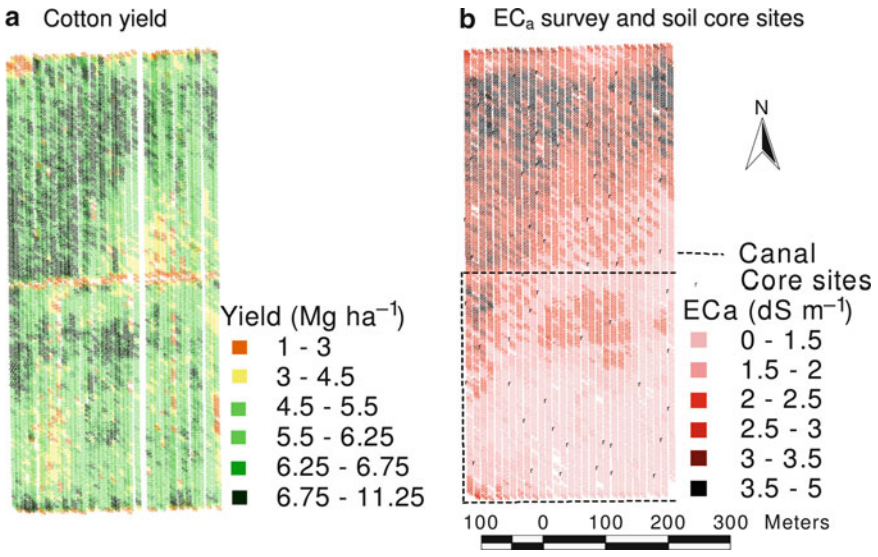

Fig. 6.1 Maps of: (**a**) cotton yield and (**b**) EC_a measurements including 60 soil sampling sites (Modified from Corwin and Lesch (2003) with permission)

6.4.1.4 Sample Site Selection, Soil Sampling and Soil Analyses

Data from the EC_a survey were used to direct the selection of 60 sample sites. The statistics software ESAP-95 version 2.01 (Lesch et al. 2000) was used to determine the sample sites from the EC_a survey data. The software uses a model-based response-surface sampling strategy. The selected sites reflect the observed signal variation in EC_a while simultaneously maximizing the spatial uniformity of the sampling design across the study area. Figure 6.1b shows the spatial EC_a survey data and the locations of the 60 soil sampling sites. Soil samples were taken at 0.3-m increments to a depth of 1.8 m and were analysed for the physical and chemical properties thought to influence cotton yield. They included gravimetric water content (θ_g), bulk density (ρ_b), pH, B, NO_3-N, Cl^-, electrical conductivity of the saturation extract (EC_e), leaching fraction (LF), % clay and saturation percentage (SP). The laboratory analyses followed the methods outlined in Agronomy Monograph No. 9 (Page et al. 1982).

The cotton yield data were not collocated with the EC_a or soil data; therefore, cotton yield was predicted at the 60 soil sampling sites by ordinary kriging. The experimental variogram computed by the usual method of moments on the yield data was fitted by an isotropic exponential function with a large nugget effect (Fig. 6.2). The considerable variation in yield over distances less than the sample spacing was most likely due to large measurement errors caused by the yield-monitoring dynamics (see Chapter 4). Nonlinear least-squares estimation was used to derive the three variogram model parameter estimates (and standard errors): nugget (c_0) = 0.76 (0.02), partial sill (c) = 1.08 (0.02) and distance parameter of the exponential function (r) = 109.3 (6.0) (approximate range, $3r$ = 327.9 m). The mean estimated

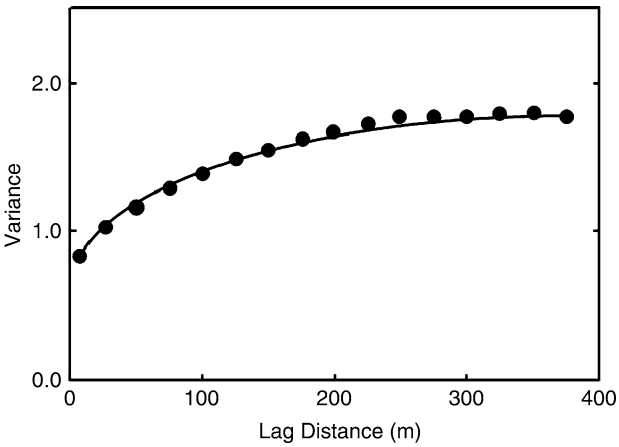

Fig. 6.2 Variogram of cotton yield. The points are the experimental variogram computed on all 7706 yield data and the solid line is the fitted exponential variogram model (see Section 1.3.2) (Taken from Corwin and Lesch (2003) with permission)

yield for the 60 sample sites was 5.95 Mg ha^{-1}, and individual estimates ranged from 3.40 to 7.41 Mg ha^{-1}. The associated kriging standard errors were from 0.93 to 0.96 Mg ha^{-1}.

6.4.1.5 Statistical and Spatial Analyses

The statistical analyses done using SAS software (SAS Institute 1999) were: (i) correlation analysis between EC_a and interpolated cotton yield using data from the 60 sites, (ii) exploratory statistical analysis to identify the significant soil properties that affect cotton yield and (iii) development of a crop yield response model using REML estimation techniques. Exploratory statistical analysis was done to determine the soil properties that have a significant effect on cotton yield and to establish the general form of the cotton yield response model. This required two stages of analysis: (i) a correlation analysis in conjunction with scatter plots of yield versus potentially significant soil properties and (ii) a preliminary multiple linear regression (MLR) analysis.

The commercial GIS software ArcView 3.3 (ESRI 2002) was used to compile, manipulate, organize and display all spatial data. The final delineation of SSMUs was done using the GIS, after exploratory statistical analyses and estimating a crop yield response model adjusted for spatial autocorrelation. A sensitivity analysis of the adjusted crop yield response model was used to identify the most significant property influencing crop yield. This analysis calculated how much the predicted yield decreased when the value for each soil property was shifted up (or down) by 1 standard deviation from its mean (Corwin and Lesch 2003).

6.4.2 Results and Discussion

6.4.2.1 Correlation Between Crop Yield and EC_a

The correlation between EC_a and yield at the 60 soil sampling sites was 0.51 (r coefficient of correlation). The moderate correlation between yield and EC_a suggests that some soil property(ies) affect both EC_a and cotton yield making an EC_a-directed soil sampling strategy potentially viable at this site. The visual similarity in the spatial distributions of EC_a and cotton yield in Fig. 6.1 confirms their close relationship.

6.4.2.2 Exploratory Statistical Analysis

Both preliminary MLR and correlation analysis showed that the 0–1.5 m soil depth resulted in the strongest correlations between yield and soil properties and best fit of the MLR to the data for the various depths considered (i.e. 0–0.3, 0–0.6, 0–0.9,

Table 6.4 Simple correlation coefficients between EC_a and soil properties and between cotton yield and soil properties. Modified from Corwin and Lesch (2003)

Soil property[a]	Fixed-array EC_a[b]	Cotton yield[c]
θ_g	0.79	0.42
EC_e	0.87	0.53
B	0.88	0.50
pH	0.33	−0.01
% clay	0.76	0.36
ρ_b	−0.38	−0.29
NO_3-N	0.22	−0.03
Cl^-	0.61	0.25
LF	−0.50	−0.49
SP	0.77	0.38

[a]Properties averaged over 0–1.5 m.
[b]Pearson correlation coefficients based on 60 observations.
[c]Pearson correlation coefficients based on 59 observations.
θ_g, gravimetric water content; EC_e, electrical conductivity of the saturation extract (dS m^{-1}); LF, leaching fraction; SP, saturation percentage.

0–1.2 and 0–1.5 m); 0–1.5 m was considered to correspond to the active root zone. The correlation analysis indicated that the following soil properties are those most significantly related to cotton yield: EC_e, LF, pH, % clay, θ_g and ρ_b. Table 6.4 shows that the correlation coefficients between EC_a and θ_g, EC_e, B, % clay, ρ_b, Cl^-, LF and SP are significant at the 0.01 level. The strongest correlations are between EC_a and θ_g, EC_e, B, % clay and SP. Note that B is not measured directly by EC_a. The strong correlation between B and EC_a is an artifact due to its close correspondence to salinity (i.e. EC_e) as a consequence of leaching. The strong correlation between EC_a and both % clay and SP is expected because it reflects the effect of texture on the EC_a. In this particular field, EC_a is strongly correlated with salinity, θ_g and texture. Table 6.4 also gives the correlation between cotton yield and the soil properties; the strongest correlation is with salinity (EC_e).

A scatter plot of EC_e and yield indicates a quadratic relationship where yield increases and then decreases (Fig. 6.3a). The scatter plot of LF and yield shows a negative, curvilinear relationship (Fig. 6.3b). Yield shows a minimal response to LF below 0.4 and it declines rapidly for LF > 0.4. Clay percentage, θ_g and ρ_b appear to be linearly related to yield to various degrees (Figs. 6.3c, f, respectively). Although there is clearly no correlation between yield and pH ($r = -0.01$, Table 6.4; Fig. 6.3d); pH became significant in the presence of the other variables, which became apparent in both the preliminary MLR analysis and final yield response model.

Based on the exploratory statistical analysis, an empirical cotton yield response model was specified as:

$$Y = \beta_0 + \beta_1 (EC_e) + \beta_2 (EC_e)^2 + \beta_3 (LF)^2 + \beta_4 (pH) + \beta_5 (\%\ clay) + \beta_6 (\theta_g) + \beta_7 (\rho_b) + \varepsilon. \quad (6.7)$$

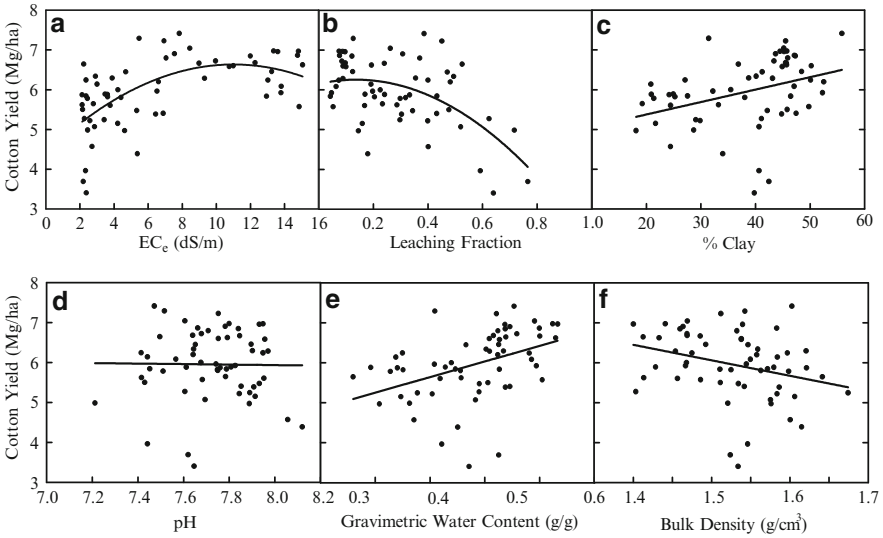

Fig. 6.3 Scatter plots of soil properties and cotton yield: (**a**) electrical conductivity of the saturation extract (ECe, dS m^{-1}), (**b**) leaching fraction, (**c**) percentage clay, (**d**) pH, (**e**) gravimetric water content (g g^{-1}) and (**f**) bulk density (g cm^{-3}) (Taken from Corwin and Lesch (2003) with permission)

In this model, the relationships between cotton yield (Y) and pH, % clay, θ_g and ρ_b are assumed to be linear; the relationship between yield and EC_e is assumed to be quadratic; the relationship between yield and LF is assumed to be curvilinear; β_0, β_1, $\beta_2, \ldots, \beta_7$ are the regression model parameters and ε represents the random error component.

6.4.2.3 Crop Yield Response Model Development

The initial estimation of Eq. 6.7 by ordinary least squares resulted in the following simplified crop yield response model:

$$Y = 20.90 + 0.38\,(EC_e) - 0.02\,(EC_e)^2 - 3.15\,(LF)^2 - 2.22\,(pH) + 9.27\left(\theta_g\right) + \varepsilon. \tag{6.8}$$

In this initial analysis, the parameter estimates for % clay and ρ_b were not significant in the t-tests and were dropped from the regression model (all other parameters were significant near or below the 0.05 level). The R^2 value for Eq. 6.8 was 0.61 indicating that 61% of the estimated spatial variation in yield could be described successfully by this model. However, a variogram of the residuals from the fitted function (Fig. 6.4) indicates that the errors are clearly spatially correlated, implying that Eq. 6.8 should be refitted using REML to adjust for spatial autocorrelation.

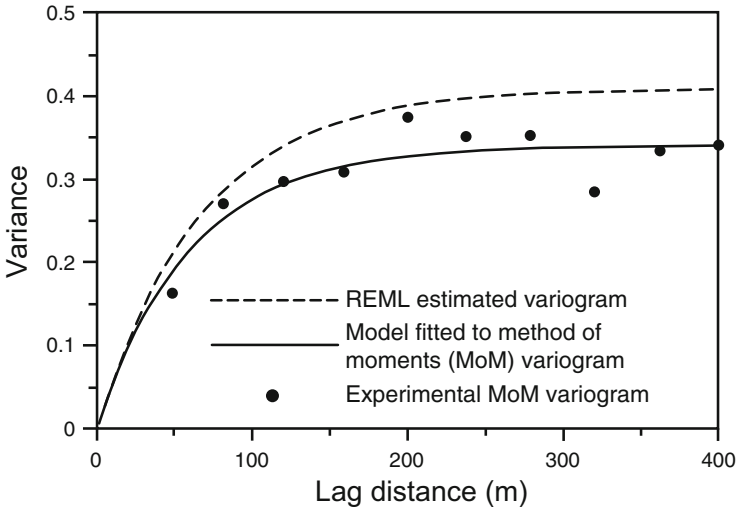

Fig. 6.4 Variograms estimated on the residuals from the ordinary least-squares yield regression model (Eq. 6.8) by residual maximum likelihood (REML) (*dashed line*) and method of moments (MoM). The symbols are the experimental variogram estimated by MoM for the 59 calibration locations and the solid line is the fitted model (Modified from Corwin and Lesch (2003) with permission)

After re-fitting Eq. 6.8 using an isotropic exponential covariance structure without a nugget effect, the following crop yield response model estimated by REML was obtained:

$$Y = 19.28 + 0.22\,(\mathrm{EC_e}) - 0.02\,(\mathrm{EC_e})^2 - 4.42\,(\mathrm{LF})^2 - 1.99\,(\mathrm{pH}) + 6.93\,(\theta_g) + \varepsilon. \tag{6.9}$$

The dashed line in Fig. 6.4 represents the variogram model estimated by REML (sill = 0.39, distance parameter = 66.2 m (working range = 198.6 m)). Note that the sill variance is larger than for the method-of-moments variogram of the residuals because the residuals from the trend are biased and the variogram is underestimated (Rao and Toutenburg 1995). The bias increases with increasing lag distance (Cressie 1993); this occurs in Fig. 6.4 to the distance at which the asymptotic sill of the exponential function is reached.

Figure 6.5 shows the observed versus predicted cotton yield estimates for Eq. 6.9. Figure 6.5 suggests that the estimated regression relationship is reasonably successful at reproducing the predicted yield estimates. A sensitivity analysis showed that LF was the single most significant factor affecting cotton yield; the degree of predicted yield sensitivity to a one standard deviation change in the $\mathrm{EC_e}$, LF, pH and θ_g resulted in % yield reductions of 4.6%, 9.6%, 5.8% and 5.1%, respectively. The point of maximum yield with respect to salinity was calculated by setting the first partial derivative of Eq. 6.9 to zero with respect to $\mathrm{EC_e}$. We note in passing that the value of 7.17 dS m^{-1} obtained is quite similar to the salinity threshold for cotton (7.7 dS m^{-1}) reported by Maas and Hoffman (1977).

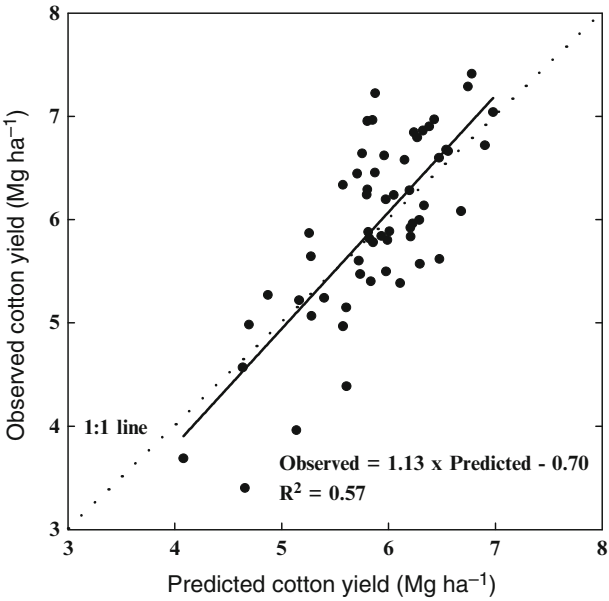

Fig. 6.5 Observed versus predicted estimates of cotton yield using Eq. 6.9. *Dotted line* is a 1:1 relationship (Taken from Corwin and Lesch (2003) with permission)

6.4.2.4 Site-Specific Management Units

Figure 6.6a–d shows the ordinary kriged maps of the four significant soil properties (0–1.5m) that affect cotton yield: (a) soil salinity (EC_e, dS m^{-1}), (b) leaching fraction (LF), (c) gravimetric water content (θ_g, kg kg^{-1}) and (d) soil pH. Ideally, if each of these four soil properties can be suitably adjusted, then in theory an optimal cotton yield could be achieved across the entire field. Based on Eq. 6.9, scatter plots of cotton yield against soil properties (Fig. 6.2) and the corresponding soil property maps (Fig 6.6), management recommendations were made that prescribed spatially what could be done to increase cotton yield in those areas with less than the optimal yield. Four recommendations can be made to improve cotton productivity at the study site: (i) reduce the LF in highly leached areas (i.e. areas where LF $>$ 0.5), (ii) reduce salinity by increased leaching in areas where the average root zone (0–1.5 m) salinity is $>$ 7.17 dS m^{-1}, (iii) increase the plant-available water in coarse-textured areas by more frequent irrigation and (iv) reduce the pH where it is $>$7.9. The rationale for each recommendation is discussed in Corwin and Lesch (2003).

Corwin and Lesch (2005a) subsequently delineated the SSMUs shown in Fig. 6.7 that indicate those areas that are pertinent to the above recommendations. All four recommendations can be accomplished by improving water application timing and distribution with variable-rate irrigation technology and by the precise application of soil amendments. Strongly leached zones were delineated where the LF needed to be reduced to $\leq$ 0.5; markedly saline areas were defined where the salinity needed

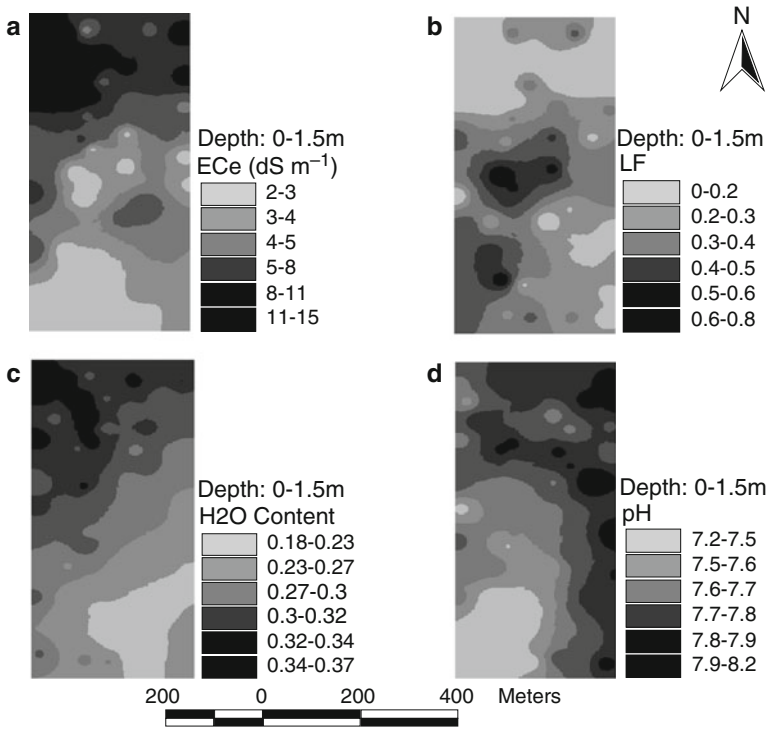

Fig. 6.6 Kriged maps of the four most significant soil properties (0–1.5 m) that affect cotton yield: (**a**) electrical conductivity of the saturation extract (EC_e, dS m^{-1}), (**b**) leaching fraction (LF), (**c**) gravimetric water content (θ_g, kg kg^{-1}) and (**d**) pH (Taken from Corwin and Lesch (2003) with permission)

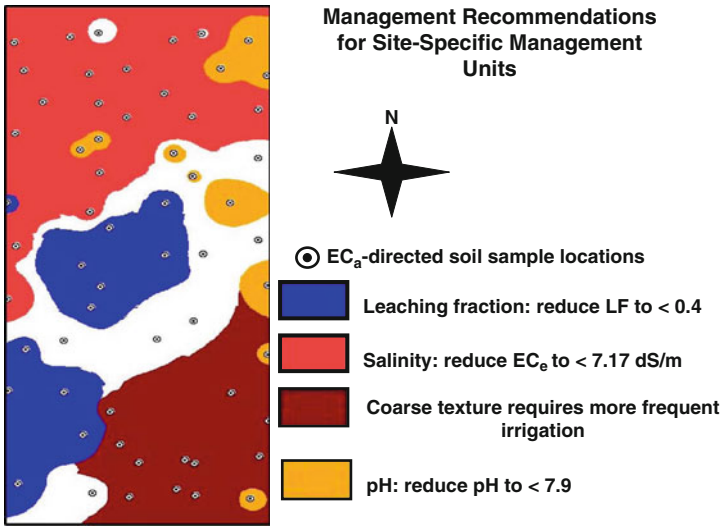

Fig. 6.7 Site-specific management units (SSMUs) for a 32.4-ha cotton field in the Broadview Water District of central California's San Joaquin Valley. Recommendations associated with the SSMUs are for: (**a**) leaching fraction, (**b**) salinity, (**c**) texture and (**d**) pH (Taken from Corwin and Lesch 2005a)

to be reduced below the salinity threshold for cotton, which was established from Eq. 6.9 to be $EC_e = 7.17\,dS\,m^{-1}$ for this field; areas of coarse texture were defined that needed more frequent irrigation and areas were identified where the pH needed to be reduced below a pH 8 with a soil amendment such as OM. Although this work has delineated within-field units where associated site-specific management recommendations would optimize the yield, it still falls short of integrating meteorological, economic and environmental impacts on within-field crop yield variation.

6.5 Conclusion

Since all proximal sensors can be, and generally are, influenced by more than one property that can affect plant yield (or quality), the most appropriate use of georeferenced proximal sensor data is to direct soil (or plant) sampling to determine the spatial distribution of properties affecting crop yield (or quality). Directed soil (or plant) sampling with proximal sensor data provides a means of establishing the properties that have most effect in crop yield (or quality) and of mapping the distribution of these properties. In addition, it provides sufficient information to develop a crop yield (or quality) response model that relates yield to edaphic or other properties affecting yield. The spatial distribution of the properties that have most effect on yield (or quality) together with a crop yield (or quality) response model provide sufficient information to delineate SSMUs with associated recommendations to increase yield (or improve quality).

Even though EC_a-directed soil sampling provides a viable means of identifying some soil properties that affect within-field variation of yield, it is only one piece of a complicated puzzle of interacting factors that result in the observed within-field variation in crops. Crop yield is affected by complex interactions of meteorological, biological, anthropogenic, topographic and edaphic factors. Furthermore, SSM requires more than just a myopic look at crop productivity. It must balance sustainability, profitability, crop productivity and quality, optimization of inputs and minimization of environmental impacts.

Mobile platforms containing multiple proximal sensors are currently being developed and tested to provide the full complement of spatial data needed to identify and spatially characterize not only edaphic but anthropogenic, topographic, meteorological and biological properties that influence plant growth. These platforms will provide multiple layers of spatial information enabling the delineation of SSMUs well beyond the capability of single proximal sensor platforms.

References

Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture, 44*, 71–91.

Adamchuk, V. I., Lund, E. D., Sethuramasamyraja, B., Morgan, M. T., Dobermann, A., & Marx, D. B. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. *Computers and Electronics in Agriculture, 48*, 272–294.

Adamchuk, V. I., Marx, D. B., Kerby, A. T., Samal, A. K., Soh, L. K., Ferguson, R. B., & Wortmann, C. S. (2007). Guided soil sampling enhanced analysis of georeferenced sensor-based data. Geocomputation Conference 2007, 3–5 Sept. 2007, National Centre for Geocomputation, Maynooth, Ireland. Available at http://ncg.nuim.ie/geocomputation/sessions/2A/2A2.pdf (verified 23 July 2009).

Andrade, P., Aguera, J., Upadhyaya, S. K., Jenkins, B. M., Rosa, U.A., & Josiah, M. (2001). Evaluation of a dielectric based moisture and salinity sensor for in-situ applications. *ASAE Paper No. 01-1010*. St. Joseph, MI, USA: ASAE.

Barnes, E. M., Sudduth, K. A., Hummel, J. W., Lesch, S. M., Corwin, D. L., Yang, C., Daughtry, C. S. T., & Bausch, W. C. (2003). Remote- and ground-based sensor techniques to map soil properties. *Photogrammetric Engineering & Remote Sensing, 69*, 619–630.

Baumgardner, M. F., Silva, L. F., Beihl, L. L., & Stoner, E. R. (1985). Reflectance properties of soils. *Advances in Agronomy, 38*, 1–44.

Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment, 113*, S38–S55.

Birrel, S. J., Borgelt, S. C., & Sudduth, K. A. (1995). Crop yield mapping: Comparison of yield monitors and mapping techniques. In P. C. Robert, R. H. Rust, & W.E. Larson (Eds.), *Proceedings of 2nd International Conference on Site-Specific Management for Agricultural Systems* (pp. 15–32). Madison WI, USA: ASA-CSSA-SSSA.

Birrel, S. J., & Hummel, J. W. (2001). Real-time multi-ISFET/FIA soil analysis system with automatic sample extraction. *Computers and Electronics in Agriculture*, 32(1), 45–67.

Brus, D. J., & de Gruijter, J. J. (1993). Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science. *Environmetrics, 4*, 123–152.

Brus, D. J., & Heuvelink, G. B. M. (2007). Optimization of sample patterns for universal kriging of environmental variables. *Geoderma, 138*, 86–95.

Bullock, D. S., & Bullock, D. G. (2000). Economic optimality of input application rates in precision farming. *Precision Agriculture, 2*, 71–101.

Burgess, T. M., & Webster, R. (1984). Optimal sampling strategy for mapping soil types. I. Distribution of boundary spacings. *Journal of Soil Science, 35*, 641–654.

Burgess, T. M., Webster, R., & McBratney, A. B. (1981). Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science, 32*, 643–654.

Carter, L. M., Rhoades, J. D., & Chesson, J.H. (1993). Mechanization of soil salinity assessment for mapping. *ASAE Paper No. 93-1557*. St. Joseph, MI, USA: ASAE.

Chung, S., Sudduth, K. A., & Hummel, J. W. (2003). On-the-go soil strength profile sensor using a load cell array. *ASAE Paper No. 03-1071*. St. Joseph, MI, USA: ASAE.

Clement, B. R., & Stombaugh, T. S. (2000). Continuously-measuring soil compaction sensor development. *ASAE Paper No. 00-1041*. St. Joseph, MI, USA: ASAE.

Corwin, D. L., Lesch, S. M. (2003). Application of soil electrical conductivity to precision agriculture: Theory, principles, and guidelines. *Agronomy Journal, 95*, 455–471.

Corwin, D. L., & Lesch, S. M. (2005a). Apparent soil electrical conductivity measurements in agriculture, *Computers and Electronics in Agriculture, 46*, 11–43.

Corwin, D. L., & Lesch, S. M. (2005b). Characterizing soil spatial variability with apparent soil electrical conductivity: I. Survey protocols. *Computers and Electronics in Agriculture, 46*, 103–133.

Corwin, D. L., & Lesch, S. M. (2005c). Characterizing soil spatial variability with apparent soil electrical conductivity: II. Case study. *Computers and Electronics in Agriculture, 46*, 135–152.

Corwin, D. L., Lesch, S. M., Shouse, P. J., Soppe, R., & Ayars, J.E. (2008). Delineating site-specific management units using geospatial EC_a measurements. In B. J. Allred, J. J. Daniels, & M. R. Ehsani (Eds.), *Handbook of agricultural geophysics* (pp. 247–254). Boca Raton, FL, USA: CRC.

Corwin, D. L., Kaffka, S. R., Hopmans, J. W., Mori, Y., Lesch, S. M., & Oster, J. D. (2003a). Assessment and field-scale mapping of soil quality properties of a saline-sodic soil. *Geoderma, 114*, 231–259.

Corwin, D. L., Lesch, S. M., Shouse, P. J., Soppe, R., & Ayars, J. E. (2003b). Identifying soil properties that influence cotton yield using soil sampling directed by apparent soil electrical conductivity. *Agronomy Journal, 95*, 352–364.

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York, NY: Wiley.

Ehrhardt, J. P., Grisso, R. D., Kocher, M. F., Jasa, P. J., & Schinstock, J. L. (2001). Using the Veris electrical conductivity art as a draft predictor. *ASAE Paper No. 01-1012*. St. Joseph, MI: ASAE.

ESRI. (2002). *ArcView 3.3*. Redlands, CA: ESRI.

Haskard, K. A., Cullis, B. R., & Verbyla, A. P. (2007). Anisotropic Matèrn correlation and spatial prediction using REML. *Journal of Agricultural, Biological, and Environmental Statistics, 12*, 147–160.

Hemmat, R., & Adamchuk, V. I. (2008). Sensor systems for measuring soil *compaction: Review and analysis. Computers and Electronics in Agriculture, 63*, 89–103.

Huisman, J. A., Hubbard, S. S., Redman, J. D., & Annan, A. P. (2003). Measuring soil water content with ground penetrating radar: A review. *Vadose Zone Journal, 2*, 476–491.

Lark, R. M., Cullis, B. R., & Welham, S. J. (2006). On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science, 57*, 787–799.

Larson, W. E., & Robert, P. C. (1991). Farming by soil. In R. Lal, & F. J. Pierce (Eds.), *Soil management for sustainability* (pp. 103–112). Ankeny, IA: Soil and Water Conservation Society.

Lesch, S. M. (2005). Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. *Computers and Electronics in Agriculture, 46*, 153–179.

Lesch, S. M., Corwin, D. L., & Robinson, D. A. (2005). Apparent soil electrical conductivity mapping as an agricultural management tool in arid zone soils. *Computers and Electronics in Agriculture*, 46, 351–378.

Lesch, S. M., Rhoades, J. D., & Corwin, D. L. (2000). *ESAP-95 version 2.01R: User manual and tutorial guide*. Research Rpt. 146. Riverside, CA: USDA-ARS, U.S. Salinity Laboratory.

Lesch, S. M., Strauss, D. J., & Rhoades, J. D. (1995). 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 Resources Research, 31*, 387–398.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute Inc.

Liu, W., Gaultney, L. D., & Morgan, M. T. (1993). Soil texture detection using acoustic methods. *ASAE Paper No. 93-1015*. St. Joseph, MI: ASAE.

Long, D. S. (1998). Spatial autoregression modeling of site-specific wheat yield. *Geoderma, 85*, 181–197.

Lunt, I. A., Hubbard, S. S., & Rubin, Y. (2005). Soil moisture content estimation using ground-penetrating radar reflection data. *Journal of Hydrology, 307*, 254–269.

Maas, E. V., & Hoffman, G. J. (1977). Crop salt tolerance – current assessment. *Journal of the Irrigation and. Drainage Division, American Society of Civil Engineers, 103*, 115–134.

Maleki, M. R., Mouazen, A. M., De Ketelaere, B., Ramon, H., & De Baerdemaeker, J. (2008). On-the-go variable-rate phosphorus fertilization based on a visible and near-infrared soil sensor. *Biosystems Engineering, 99*, 35–46.

Morari, F., Castrignano, A., & Paglirin, C. (2009). Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. *Computers and Electronics in Agriculture, 68*, 97–107.

Mouazen, A. M., & Roman, H. (2006). Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture. *Soil & Tillage Research, 86*, 218–229.

Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., & Ramon, H. (2007). On-line measurement of some selected soil properties using a VIS-NIR sensor. *Soil & Tillage Research, 93*, 13–17.

Müller, W. G. (2001). Collecting spatial data: Optimum design of experiments for random fields (2nd ed.). Heidelberg, Germany: Physica-Verlag.

Müller, W. G., & Zimmerman, D. L. (1999). Optimal designs for variogram estimation. *Environmetrics, 10*, 23–37.

Myers, R. H., & Montgomery, D. C. (2002). *Response surface methodology: process and product optimization using designed experiments* (2nd ed.). New York, NY: Wiley.

Page, A. L., Miller, R. H., & Kenney, D. R. (Eds.). (1982). *Methods of soil analysis, Part 2 – Chemical and microbiological properties* (2nd ed.). Agron. Monogr. No. 9. Madison, WI: ASA-CSSA-SSSA.

Rao, C. R., & Toutenburg, H. (1995). *Linear models: least squares and alternatives*. New York, NY: Springer.

Rhoades, J. D. (1992). Instrumental field methods of salinity appraisal. In G. C. Topp, W. D. Reynolds, & R. E. Green (Eds.), *Advances in measurement of soil physical properties: bring theory into practice* (pp. 231–248). SSSA Special Publication No. 30. Madison, WI: SSSA.

Russo, D. (1984). Design of an optimal sampling network for estimating the variogram. *Soil Science Society of America Journal, 48*, 708–716.

SAS Institute. (1999). *SAS software*, version 8.2. Cary, NC: SAS Institute.

Schabenberger, O., & Gotway, C. A. (2005). *Statistical methods for spatial data analysis*. Boca Raton, FL: CRC.

Sethuramasamyraja, B., Adamchuk, V. I., Dobermann, A., Marx, D. B., Jones, D. D., & Meyer, G. E. (2008). Agitated soil measurement method for integrated on-the-go mapping of soil pH, potassium and nitrate contents. *Computers and Electronics in Agriculture, 60*, 212–225.

Shonk, J. L., Gaultney, L. D., Schulze, D. G., & Van Scoyoc, G. E. (1991). Spectroscopic sensing of soil organic matter content. *Transactions of the American Society of Agricultural Engineers, 34*, 1978–1984.

Tanji, K. K. (Ed.). (1996). *Agricultural salinity assessment and management*. New York, NY: ASCE.

Thompson, S. K. (1992). *Sampling*. New York, NY: Wiley.

Valliant, R., Dorfman, A. H., & Royall, R. M. (2000). *Finite population sampling and inference: a prediction approach*. New York, NY: Wiley.

Van Uffelen, C. G. R., Verhagen, J., & Bouma, J. (1997). Comparison of simulated crop yield patterns for site-specific management. *Agricultural Systems, 54*, 207–222.

Verhagen, A., Booltink, H. W. G., & Bouma, J. (1995). Site-specific management: balancing production and environmental requirements at farm level. *Agricultural Systems, 49*, 369–384.

Verschoore, R., Pieters, J.G., Seps, T., Spriet, Y., & Vangeyte, J. (2003). Development of a sensor for continuous soil resistance measurement. In J. Stafford, & A. Werner (Eds.), *Precision agriculture* (pp. 689–695). Wageningen, The Netherlands: Wageningen Academic Publishers.

Viscarra Rossel, R. A., & Walter, C. (2004). Rapid, quantitative and spatial field measurements of soil pH using an ion sensitive field effect transistor. *Geoderma, 119*, 9–20.

Vitharana, U. W. A., Van Meirvenne, M., Simpson, D., Cockx, L., & De Baerdemaeker, J. (2008). Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. *Geoderma, 143*, 206–215.

Wackernagel, H. (1998). *Multivariate geostatistics* (2nd ed.). Berlin, Germany: Springer.

Warrick, A. W., & Myers, D. E. (1987). Optimization of sampling locations for variogram calculations. *Water Resources Research, 23*, 496–500.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). New York, NY: Wiley.

Whalley, W. R., & Bull, C. R. (1991). An assessment of microwave reflectance as a technique for estimating the volumetric water content of soil. *Journal of Agricultural Engineering Research, 50*, 315–326.

World Resources Institute. (1998). *1998–99 world resources – a guide to the global environment*. New York, NY: Oxford University Press.

Yan, L., Zhou, S., & Feng, L. (2007a). Delineation of site-specific management zones based on temporal and spatial variability of soil electrical conductivity. *Pedosphere, 17*, 156–164.

Yan, L., Zhou, S., Feng, L., & Hong-Yi, L. (2007b). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture, 56*, 174–186.

Zhu, Z., & Stein, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics, 11*, 24–44.

Chapter 7
Using Ancillary Data to Improve Prediction of Soil and Crop Attributes in Precision Agriculture

P. Goovaerts and R. Kerry

Abstract This chapter describes three geostatistical methods to incorporate secondary information into the mapping of soil and crop attributes to improve the accuracy of their predictions. The application of the methods is illustrated in two case studies. Cokriging is the multivariate extension of the well known ordinary kriging. It does not require ancillary data to be available at all nodes of the interpolation grid, whereas kriging with external drift and simple kriging with local means do. Cokriging, however, is more demanding in terms of variogram inference and modelling. The other two methods use ancillary data to model the spatial trend of the primary variable. Kriging with an external drift can account for local changes in the linear correlation between primary and secondary variables. Simple kriging with local means, which applies kriging to regression residuals and adds the kriged residual to the regression estimate, is the most straightforward of these methods to implement. The prediction performance of each technique was evaluated by cross-validation. As the results are site-specific, the choice of technique for a given site should be guided by the results of cross-validation.

Keywords Cokriging · Cross-variogram · Kriging with external drift · Simple kriging with local means · Soil · Ancillary data

7.1 Introduction

Contour maps that characterize the variation in soil and crop attributes accurately within fields are a fundamental requirement for precise crop management. Chapter 2 in this book indicates that the accuracy of predictions used for such maps depends

P. Goovaerts
BioMedware Inc3526 W Liberty, Suite 100, Ann Arbor, MI 48104, USA
e-mail: goovaerts@terraseer.com

R. Kerry (✉)
Department of Geography, Brigham Young University, 690 SWKT, Provo, UT 84602, USA
e-mail: ruth_kerry@byu.edu

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_7,

on the quality of sample information. Many soil and crop attributes, however, are recorded by the time-consuming and expensive methods of sampling in the field followed by laboratory analyses. Consequently, many precision farmers take only a small number of samples. One sample per hectare has become almost an industry standard (Godwin and Miller 2003), but this sampling intensity is governed by economic concerns and bears no relationship to the sampling requirements to resolve the variation present in a particular field. Often ancillary information is correlated with the soil and crop information and can indicate the approximate scale of variation present in these. Chapter 2 describes how the scale of variation can be characterized, with little additional expense, by ancillary data to determine an appropriate sampling interval. It also examines approaches to minimize the number of samples that need to be obtained, but it does not investigate the merits of ancillary data to improve prediction. Ancillary data provide a dense cover of the field yet are relatively inexpensive to obtain. Several sources of ancillary data such as yield data, electrical conductivity (EC_a), remotely sensed images, digital elevation models and information on soil series from polygon maps are available to the precision farmer.

This chapter is concerned primarily with the merits of various methods of incorporating ancillary data into the kriging procedure, namely by: cokriging (CK), kriging with an external drift (KED) and simple kriging with local means (SKlm) using soil type or other ancillary data. Each of these methods is explained in detail by Goovaerts (1997). The existing literature shows that these methods of incorporating ancillary data into the kriging procedure have been widely adopted in the geosciences, hydrology, soil science and climatology. They have not been widely applied in precision agriculture, but there have been studies that illustrate the value of these techniques. For example, Dobermann and Ping (2004) compared the merits of each of the above techniques for improving yield maps with information from remotely sensed images, and Ge et al. (2007) investigated regression kriging (RK) and visible and near infrared spectroscopy (VNIR) for estimating soil properties. Regression kriging is a name often used for SKlm. It has been used by Kravchenko and Robertson (2007) with topographic and yield data to improve estimates of soil carbon, and by Triantafilis et al. (2001), together with CK to estimate soil salinity from electrical conductivity (EC_a) data. Kriging with an external drift and RK have been used by Lesch and Corwin (2008) with EC_a and remotely sensed imagery to improve predictions of various soil properties, whereas Baxter and Oliver (2005) used KED with elevation data to improve the prediction of soil nitrogen. Kozar et al. (2002), Tarr et al. (2005) and Vitharana et al. (2006) used CK to improve the mapping of selected soil properties with terrain information and EC_a as the secondary data. Some of the studies mentioned above use only one covariate in the interpolation process, but the use of more than one in combination may improve the accuracy of prediction.

This chapter illustrates how several geostatistical methods that incorporate ancillary data can improve the accuracy of predicting soil properties. We provide a short background to the geostatistical theory of CK, KED and SKlm. The results of applying these methods for prediction are compared with those from ordinary kriging (OK) and are illustrated with two case studies. The effects of decreasing sample

size on the prediction performances of the various methods of kriging will also be compared. The practicalities of using each method in terms of data and software availability are also discussed.

7.2 Theory

The techniques described in this chapter use the spatial autocorrelation between soil samples and their cross-correlation or coregionalization with ancillary data to predict at unsampled locations. The (co)kriging systems introduced below are written in terms of covariances, but common practice tends to infer and model the variogram (which measures the dissimilarity between observations) rather than the covariance function because it requires a weaker assumption (intrinsic stationarity instead of second-order stationarity). The covariance is easily retrieved from the variogram model by subtracting the variogram values from the sill (bounded model) or pseudo-sill (unbounded model). Regardless of the choice of interpolation algorithm, the first step is to select the subset of covariates that is the most informative. Secondary variables should correlate well with the primary variable, but weakly among themselves to avoid redundancy and associated collinearity issues.

7.2.1 Variogram and Cross-Variogram

The experimental variogram, $\hat{\gamma}_Z(\mathbf{h})$, of the soil attribute Z (primary variable) for a given lag vector $\mathbf{h}$ is estimated as:

$$\hat{\gamma}_Z(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2, \tag{7.1}$$

where $N(\mathbf{h})$ is the number of data pairs within the class of distance and direction used for the lag vector $\mathbf{h}$. A continuous function must be fitted to $\hat{\gamma}_Z(\mathbf{h})$ to obtain semivariances for any possible lag $\mathbf{h}$ required by the prediction algorithms (Goovaerts 1999).

Joint variation between primary and secondary variables Z and Y can be characterized by the experimental cross-variogram defined as:

$$\hat{\gamma}_{ZY}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})] [y(\mathbf{u}_\alpha) - y(\mathbf{u}_\alpha + \mathbf{h})] \ . \tag{7.2}$$

Note that to compute the cross variogram there must be locations in common where both variables have been measured. If both attributes are positively correlated spatially, an increase (decrease) in the values of Z from $\mathbf{u}_\alpha$ to $\mathbf{u}_\alpha + \mathbf{h}$ tends to be

associated with an increase (decrease) in the values of Y. Conversely, a negative correlation between attributes would arise from an increase (decrease) in Z and an associated decrease (increase) in Y. Modelling the coregionalization between two variables Z and Y involves choosing and fitting functions to the two auto-variograms $\hat{\gamma}_Z(\mathbf{h})$ and $\hat{\gamma}_Y(\mathbf{h})$ and to the cross-variogram $\hat{\gamma}_{ZY}(\mathbf{h})$. The three models cannot be fitted independently from one another (Goovaerts 1997). The easiest approach is to model the three variograms as linear combinations of the same set of basic variogram functions under the constraint that the matrix of sills of these models are all positive semi-definite. This is generally known as the linear model of coregionalization (LMC); see Goovaerts (1997) and Wackernagel (1995) for a description of this. As for the direct variogram, the cross-covariance value is computed by subtracting the cross-variogram value from its modelled sill.

7.2.2 Cokriging

One way to incorporate secondary information once the LMC has been fitted is to use a multivariate extension of kriging known as cokriging. In the simplest case of a single secondary attribute Y, the ordinary cokriging estimate is written as the following linear combination of both neighbouring primary and secondary data:

$$z^*_{CK}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha) + \sum_{\alpha'=1}^{m} \lambda_{\alpha'}(\mathbf{u}) y(\mathbf{u}_{\alpha'}), \tag{7.3}$$

where some of the secondary data might have been measured at possibly different locations $\mathbf{u}_{\alpha'}$. As for ordinary kriging, the objective is to minimize the error variance under an unbiasedness constraint, which gives the following system of linear equations:

$$\begin{aligned}
&\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) C_Z(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{\beta'=1}^{m} \lambda_{\beta'}(\mathbf{u}) C_{ZY}(\mathbf{u}_\alpha - \mathbf{u}_{\beta'}) + \mu_Z(\mathbf{u}) \\
&\qquad = C_Z(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1, 2, \ldots, n \\
&\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) C_{YZ}(\mathbf{u}_{\alpha'} - \mathbf{u}_\beta) + \sum_{\beta'=1}^{m} \lambda_{\beta'}(\mathbf{u}) C_Y(\mathbf{u}_{\alpha'} - \mathbf{u}_{\beta'}) + \mu_Y(\mathbf{u}) \\
&\qquad = C_{YZ}(\mathbf{u}_{\alpha'} - \mathbf{u}) \quad \alpha' = 1, 2, \ldots, m \\
&\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) = 1 \\
&\sum_{\beta'=1}^{m} \lambda_{\beta'}(\mathbf{u}) = 0.
\end{aligned} \tag{7.4}$$

There are two Lagrange parameters, $\mu_Z(\mathbf{u})$ and $\mu_Y(\mathbf{u})$, to account for the constraints on weights of the primary and secondary data, and information from the LMC provides values of the direct and cross-covariances for different lags.

The application of cokriging with N_v secondary variables requires the estimation and joint modelling of $(N_v + 1)(N_v + 2)/2$ direct and cross-variograms; a task that rapidly becomes daunting as the number of secondary variables increases. A less demanding approach in terms of variogram inference and modelling is to use the secondary information to estimate the local mean of the primary attribute Z (e.g. through linear regression), and then krige the corresponding residuals. Unlike the cokriging approach, only one residual variogram needs to be estimated and modelled regardless of the number of secondary variables. However, the auxiliary variables must be known at all locations where the primary variable is known, as well as at all places on the prediction grid. If the secondary variables were not sampled exhaustively, this requirement could be met by interpolating the secondary variables to these additional locations at the outset. The two main geostatistical algorithms for incorporating secondary information to model the local mean are simple kriging with local means (SKlm) and kriging with an external drift (KED).

7.2.3 Simple Kriging with Local Means

This procedure starts by modelling the local mean of the primary variable Z from the secondary information. Two variants of the linear regression model were applied in this chapter: (1) a multiple linear regression model was fitted to ancillary data (SKlm-h), i.e. $\mathrm{E}[Z(\mathbf{u})] = f(Y_1(\mathbf{u}), \ldots, Y_{Nv}(\mathbf{u})) = m^*(\mathbf{u})$ and (2) the local mean within each soil series was assumed constant and equal to the arithmetic average of the values of Z within a given soil series (SKlm-s). The kriging estimate is then expressed as a linear combination of the neighbouring primary z-data and the local mean estimated at these n locations and the location $\mathbf{u}$ being predicted:

$$\begin{aligned} z^*_{SKlm}(\mathbf{u}) &= \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) \left[z(\mathbf{u}_\alpha) - m^*(\mathbf{u}_\alpha) \right] + m^*(\mathbf{u}) \\ &= \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) r(\mathbf{u}_\alpha) + m^*(\mathbf{u}) \,, \end{aligned} \tag{7.5}$$

where $r(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha) - m^*(\mathbf{u}_\alpha)$ are referred to as residuals. The kriging weights are obtained by solving the following simple kriging system:

$$\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1, 2, \ldots, n \,, \tag{7.6}$$

where $C_R(\mathbf{h})$ is the covariance function of the residual random function $R(\mathbf{u})$, not that of the variable Z itself.

7.2.4 Kriging with an External Drift

There are two main differences between KED and SKlm. These are that in KED: (1) the relationship between primary and secondary variables needs to be linear (if it is not an appropriate transformation of the secondary variables could make the relation linear) and (2) the coefficients of the linear model are determined implicitly through the kriging system within each search neighbourhood (these coefficients are assumed to be constant across the study area for SKlm). The KED estimate is computed as follows:

$$z^*_{KED}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha)\,. \tag{7.7}$$

The weights $\lambda_\alpha(\mathbf{u})$, assigned to each observation, are computed by solving the following kriging system for the simple case with one secondary variable Y:

$$\begin{aligned}
&\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_0(\mathbf{u}) + \mu_1(\mathbf{u}) y(\mathbf{u}_\alpha) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1, 2, \ldots, n\,, \\
&\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) = 1, \\
&\sum_{\beta=1}^{n} \lambda_\beta(\mathbf{u}) y(\mathbf{u}_\beta) = y(\mathbf{u}),
\end{aligned} \tag{7.8}$$

where $\mu_0(\mathbf{u})$ and $\mu_1(\mathbf{u})$ are two Lagrange parameters accounting for the constraints on the weights. Inference of the residual covariance is not as straightforward as for SKlm because to compute the residuals requires the kriging system, Eq. 7.8, to be solved and prior knowledge of the residual covariance. Inference methods of varying complexity are available (Wackernagel 2003), which range from maximum likelihood estimation of the residual variogram to a straightforward method of computing the variogram of the variable using only pairs of z-values that are unaffected or only slightly affected by the trend. In this chapter, as in previous studies (i.e. Goovaerts 2000), KED was performed using the same covariance model as for SKlm.

7.3 Case Study 1: The Yattendon Site

7.3.1 Site Description and Available Data

The Yattendon site is a 15.3 ha field on the Chalk ridgeway in Berkshire, southern England. The field comprises a plateau area to the north and a south-facing slope with gradients of 8–15% in the south. A previous survey of the area identified six

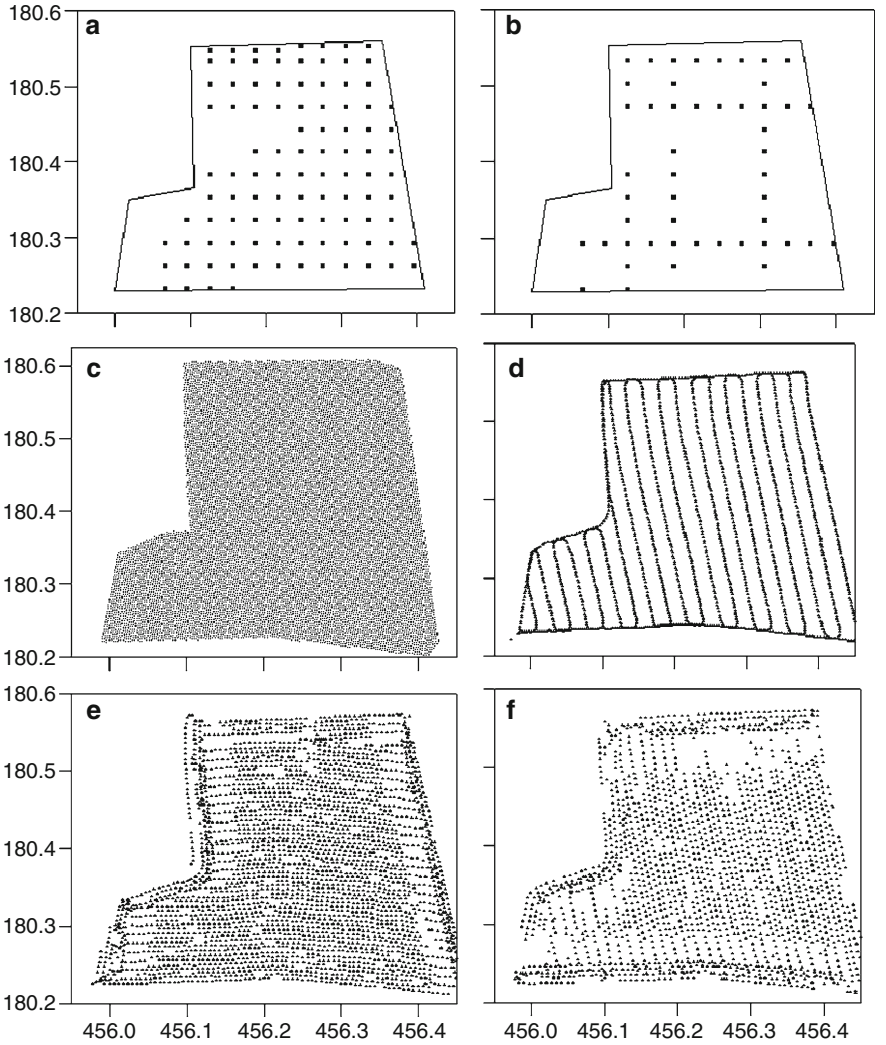

Fig. 7.1 Soil and ancillary data sampling schemes at the Yattendon site for: (**a**) the full soil data (102 sites), (**b**) sub-sampled soil data (50 sites), (**c**) aerial[91] data (14 683 points), (**d**) EC_a and elevation data (1927 points), (**e**) yield[96] data (3106 points), (**f**) yield[98] data (2136 points). Coordinates are Eastings and Northings in kilometres

soil series within this field (Heming 1997). Topsoil samples (0–15 cm) were taken on a 30-m square grid (Fig. 7.1a). At each grid node six samples of soil were taken within 1 m^2 and bulked for further analyses. Several soil properties were measured for each of the 102 samples. Percentage sand of the air-dry <2 mm soil fraction determined by laser granulometry (Coulter LS 230 laser granulometer) is used here as the primary variable.

Several types of ancillary data were available at the Yattendon site and these are typical of the information available to precision farmers. Aerial photographs of the bare soil from standard surveys (Aerofilms Ltd., www.aerofilms.com) from 1986 and 1991 (aerial^{86} and aerial^{91}) were scanned at a resolution of 75 dpi to give a ground pixel size of 3.4 m (Fig. 7.1c) and geo-corrected to British Ordnance Survey coordinates using Erdas Imagine (www.Erdas.com). Digital numbers (DNs, 0–255) were then extracted for the red, green and blue wavebands; the green DNs only are used here because analysis of the individual wavebands by Kerry (2004) suggested that this was the best of the visible wavebands to use. Electrical conductivity (EC_a) data were recorded every 1–2 m along transects spaced approximately 20-m apart (Fig. 7.1d) using the Geonics EM38 instrument (www.geonics.com) in the vertical position. Location and elevation were recorded with a differential global positioning system (DGPS) which was attached to the EM38 (Fig. 7.1d). Yield data from 1996 and 1998 (yield^{96} and yield^{98}) (Fig. 7.1e, f) were available and were recorded using the RDS Technology (www.rdstechnology.ltd.uk) 'Ceres 2' yield meter and DGPS. This volume based sensor was calibrated with grain densities to determine the mass of the yield. The yield data were also pre-processed to remove probable erroneous data, for example, locations with very small yield values which had been passed over twice by the combine (see Kerry 2004 for more detail). Figure 7.1c–f shows the number and location of data points for each type of ancillary information. There are about 20 times more observations for the ancillary data than for the soil (Fig. 7.1a); the aerial image provides the most intensive data and the most complete cover of the field.

7.3.2 Data Preparation

Experimental variograms were computed from the ancillary data by Eq. 7.1 and modelled. Ordinary kriging was then used to predict the ancillary data to the nodes of a 5-m grid which had points in common with the 30-m soil sampling grid. The kriged maps are shown in Fig. 7.2 c–h. Pixel maps rather than contour maps are used to show the complex spatial patterns and to avoid smoothing of the micro-scale variation. The kriged ancillary values at the nodes of the 30-m soil sampling grid were extracted and used for correlation analysis. Table 7.1 shows the correlations between percentage sand and the various types of ancillary data at the Yattendon site. Correlations were moderate for aerial^{91} and EC_a, and weak for the other variables. Nevertheless, there are distinct similarities in the spatial patterns of all data, in particular the contrast between the south and east of the field (the south-facing slope) and the north and west (the plateau area) (Fig. 7.2). The main difference between the patterns in sand and the ancillary data is the size of patches of different values in the northwest of the field. These are markedly smaller in Fig. 7.2e and f than Fig. 7.2a for percentage sand; this suggests that more intensive soil sampling might have resulted in the identification of smaller patches of small and large percentage sand, Fig. 7.2a. The patterns for aerial^{86} are similar to those for aerial^{91},

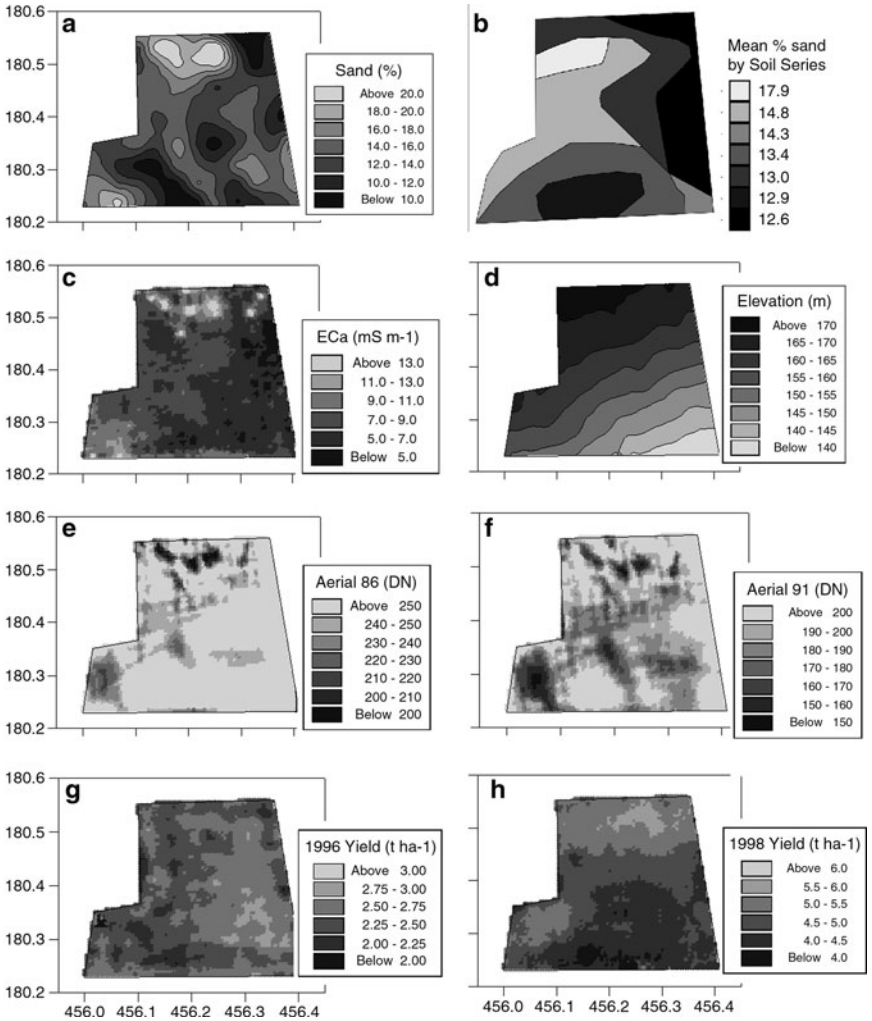

Fig. 7.2 Maps of ordinary kriged sand (**a**) and available ancillary data at the Yattendon site: (**b**) mean sand by soil series, (**c**) EC_a, (**d**) elevation, (**e**) DNs for aerial^{86}, (**f**) DNs for aerial^{91}, (**g**) 1996 yield data, (**h**) 1998 yield data

Table 7.1 Correlations between selected soil and ancillary variables at the Yattendon site

	Elevation	Aerial^{86}	Aerial^{91}	EC_a	Yield^{96}	Yield^{98}	Sand
Elevation	1.00						
Aerial^{86}	−0.30	1.00					
Aerial^{91}	0.01	0.28	1.00				
EC_a	0.48	−0.53	−0.32	1.00			
Yield^{96}	0.02	−0.08	−0.21	0.07	1.00		
Yield^{98}	0.80	−0.29	0.05	0.41	0.04	1.00	
Sand	0.07	−0.16	−0.49	0.38	0.08	0.11	1.00

but the former shows less detail in the east of the field, which probably explains the weaker correlation with the soil data (compare Fig. 7.2e, f). Some of the literature cited above suggested that some authors had found elevation and yield data useful. However, EC_a and aerial imagery are the most strongly correlated with sand in this field, but in addition elevation was used for SKlm-h and KED, and yield[96] for CK because the spatial patterns of all ancillary data are generally similar.

The mean sand percentage for each soil series (Fig. 7.2b) was determined from sand data on the 30-m grid using the aggregation procedure of Terraseer STIS (http://www.terraseer.com/products_stis.php) and the soil series boundaries. This procedure was also used to assign the mean percentage sand by soil series to each node of the 5-m interpolation grid.

7.3.3 Variograms

The usual method of moments variogram (Eq. 7.1) was computed on the raw sand data. All variogram models were fitted by weighted least squares approximation with more weight being assigned to the first few lags. The experimental variogram of the raw sand data is shown as symbols in Fig. 7.3a and the fitted model as a solid line. The parameters of the model are given in Table 7.2. Simple linear regression and multiple linear regression were performed with sand as the dependent variable and aerial[91], EC_a and elevation as independent variables, both separately and in combinations of two or three variables. Both linear and squared terms (i.e. no interaction) were used for each of the ancillary variables in the regressions. The residuals from these regressions (rr) of the 30-m soil and ancillary data were used to compute variograms. These are shown in Fig. 7.3c–f and their model parameters are given in Table 7.2. A variogram was also computed and modelled after subtracting the soil series mean sand percentage from the observed sand data (soil series residuals, Fig. 7.3b and Table 7.2).

Each of the variograms computed from the 30-m grid (Fig. 7.3) has a range between 50 and 70 m (Table 7.2), except the one for aerial[91] and EC_arr which shows little spatial structure as >80% of the variance is nugget (Table 7.2). The nugget:sill ratio for aerial[91], EC_a and elevation rr is also greater than 50% (Fig. 7.3c). The large nugget:sill ratio suggests that using these combinations of ancillary data for regression explains much of the spatially structured variation in the sand data. The nugget:sill ratios for the raw soil data, elevation rr and aerial[91], and elevation rr variograms are 0% (Table 7.2 and Fig. 7.3f) and those for soil series residuals, EC_a rr, elevation rr and EC_a, and elevation rr (Table 7.2 and Fig. 7.3e) are 20–40%. The full 30-m soil dataset was sub-sampled to a 60-m interval and then supplemented with additional samples at 30-m (usually along small transects) to give a sample size of 50. Variograms computed on this sub-sample of 50 data (Table 7.2) have a slightly longer range and generally smaller nugget:sill ratios (Fig. 7.3g–i and Table 7.2). This

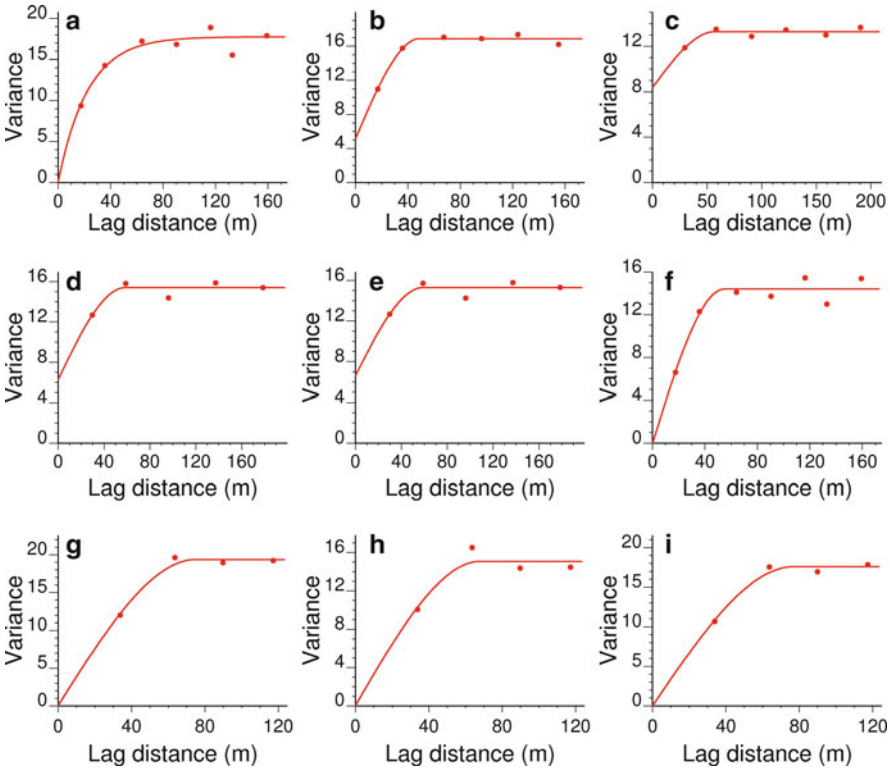

Fig. 7.3 Variograms for 30-m soil data: (**a**) sand (raw data), (**b**) soil series residuals, (**c**) aerial^{91}, EC_a and elevation rr, (**d**) EC_a rr, (**e**) EC_a and elevation rr, (**f**) aerial^{91} and elevation rr. Variograms for the subset of 30-m soil data with 50 samples: (**g**) soil series residuals (50), (**h**) aerial^{91}, EC_a and elevation rr (50) and (**i**) aerial^{91} and EC_a rr (50)

illustrates the effect of the number of observations on variogram estimation, and the smaller nugget:sill ratio shows that regression with this smaller dataset explains less of the spatially structured variation present.

Auto- and cross-variograms for sand and aerial^{91}, EC_a and yield^{96} were computed and fitted with a linear model of coregionalization (LMC) in ISATIS (www.geovariances.fr). As the variograms of elevation and yield^{98} are unbounded and that for sand is bounded, the former were unsuitable for fitting the LMC with sand. The auto- and cross-variograms for sand with aerial^{91} and EC_a are shown in Fig. 7.4 and their model parameters are given in Table 7.3. A spherical function with a nugget component and range of 85 m was used for the LMC. This fits best for the aerial^{91} data (Fig. 7.4f) and similarly well for the sand and EC_a data (Fig. 7.4a,c). Following Wackernagel (2003), the degree of coregionalization for each cross-variogram can be assessed visually by plotting the hull of perfect positive and negative correlation. The hull is obtained by replacing the cross-variogram sills by the square roots of the product of the sills of the two corresponding auto-variograms.

Table 7.2 Parameters of models fitted to variograms of raw data and regression residuals (rr)

Data	Model	c_0	c_1	$a, 3r$ (m)
30-m data				
Sand (raw data)	Exponential	0	17.76	67.97
Soil series residuals	Spherical	5.081	11.78	50.40
Aerial^{91} rr	Spherical	0	14.42	54.71
EC_a rr	Spherical	6.253	9.154	58.22
Elevation rr	Spherical	2.293	14.97	54.81
Aerial^{91}, EC_a rr	Exponential	11.08	2.334	115.1
Aerial^{91}, EC_a, elevation rr	Spherical	8.380	4.914	57.08
EC_a, elevation rr	Spherical	6.604	8.695	58.55
Aerial^{91}, elevation rr	Spherical	0	14.42	54.71
Sub-sample of 50 data				
Sand (raw data)	Spherical	1.295	16.80	80.95
Soil series residuals	Spherical	0	19.38	75.26
Aerial^{91}, EC_a, elevation rr	Spherical	0	15.09	68.51
Aerial^{91}, EC_a rr	Spherical	0	17.59	77.18

The model parameters are: c_0, the nugget variance; c, the sill of the autocorrelated variance; a, the range of the spatial dependence. For the exponential model, because the sill is asymptotic an approximate range is determined as $3r$,where r is the distance parameter of the model.

The hulls indicate that there is a weak-moderate positive spatial relationship between EC_a and sand, and a moderate negative one between sand and aerial^{91}. These results confirm the patterns observed in the maps in Fig. 7.2 and the correlation coefficients listed in Table 7.1.

7.3.4 Leave-One-Out Cross-Validation

The parameters of the various models given in Table 7.2 and the associated data were used for leave-one-out (LOO) cross-validation by OK, KED, SKlm-h and SKlm-s in Terraseer STIS. Each soil datum is removed in turn and its value is predicted using all collocated ancillary data and the remaining soil data in the neighbourhood. Accuracy of prediction can be evaluated using the mean error (ME, Eq. 1.28) which should be close to zero, the mean absolute error (MAE, Eq. 1.29) which should be as small as possible and the mean squared deviation ratio (MSDR, Eq. 1.30) which should be close to one if an appropriate model has been used. The MSDR is essentially the mean of the ratio of the squared prediction errors to the kriging variance. If an inappropriate model has been used, the kriging variance will either markedly under- or over-estimate the true squared errors. Finally, cross-validation for CK was performed following the same procedure and using the same evaluation statistics in ISATIS.

Cross-validation results for the various methods and combinations of ancillary data are given in Table 7.4. In this table and hereafter, the ME is excluded as an evaluation criterion as values for all methods were close to zero indicating little

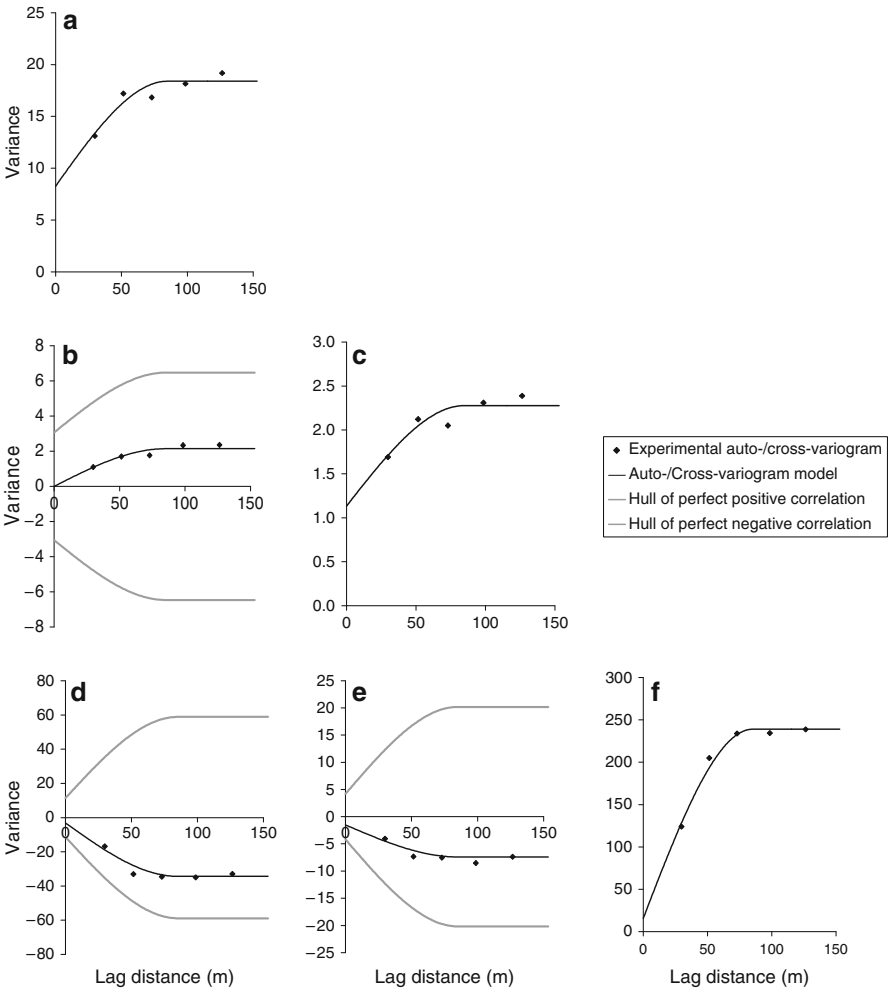

Fig. 7.4 Cross- (**b**, **d** and **e**) and auto-variograms (**a**, **c** and **f**) for soil and ancillary data at the Yattendon site fitted using a linear model of coregionalization with spherical structure, nugget component and range of 85 m. (**a**) sand, (**b**) sand and EC_a, (**c**) EC_a, (**d**) sand and aerial^{91}, (**e**) EC_a and aerial^{91} and (**f**) aerial^{91}

bias in the predictions. For reference, Table 7.4 shows the MAE (2.93) and MSDR (0.996) for OK. For CK and all combinations of ancillary data the MSDR is close to 1 and the MAE is smaller than that for OK, except when EC_a and yield^{96} data are used on their own. The MAE is smallest when aerial^{91} and EC_a data are used together in the LMC. Overall, the best MAEs for CK are generally smaller than for the other methods and the smallest MAE was obtained for CK. For KED, the MSDR values are generally further from one than for OK and CK, but they are still reasonable. The MAEs for KED are less than for OK with only the EC_a and aerial^{91},

Table 7.3 Auto- and cross-variograms fitted using the linear model of coregionalization with a spherical function and a range of 85 m

Variable 1	Variable 2	c_0	c
Sand	Sand	8.274	10.12
Sand	EC_a	−0.004	2.142
Sand	Aerial[91]	−3.089	−31.30
EC_a	Sand	−0.004	2.142
EC_a	EC_a	1.133	1.145
EC_a	Aerial[91]	−1.517	−5.902
Aerial[91]	Sand	−3.087	−31.30
Aerial[91]	EC_a	−1.517	5.902
Aerial[91]	Aerial[91]	15.52	223.5

Table 7.4 Leave-one-out cross-validation results for ordinary kriging (OK), cokriging (CK), simple kriging with local means (SKlm) and kriging with an external drift (KED) using the full soil data at the Yattendon site

Method	Ancillary variables	R^2 for regression model	Mean absolute error (MAE)	Mean squared deviation ratio (MSDR)
CK	Aerial[91]		2.728	1.030
CK	EC_a		2.891	1.009
CK	Yield[96]		2.945	1.007
CK	Aerial[91], EC_a		2.674	1.046
CK	Aerial[91], yield[96]		2.722	1.080
CK	Aerial[91], EC_a, yield[96]		2.684	1.062
CK	EC_a, yield[96]		2.855	1.009
KED	Aerial[91]	0.182	2.955	1.248
KED	EC_a	0.172	2.865	0.954
KED	Elevation	0.017	3.097	1.077
KED	Aerial[91], EC_a	0.281	2.777	0.875
KED	Aerial[91], elevation	0.221	3.061	1.241
KED	Aerial[91], EC[a], elevation	0.312	3.141	0.811
KED	EC[a], elevation	0.177	3.088	1.014
OK	–		2.930	0.996
SKlm-h	Aerial[91]	0.182	2.982	1.253
SKlm-h	EC_a	0.172	2.828	0.974
SKlm-h	Elevation	0.017	2.971	1.003
SKlm-h	Aerial[91], EC_a	0.281	2.813	0.942
SKlm-h	Aerial[91], elevation	0.221	2.982	1.254
SKlm-h	Aerial[91], EC_a, elevation	0.312	2.805	1.078
SKlm-h	EC_a, elevation	0.177	2.834	0.964
SKlm-s	Soil series	0.112	2.824	0.865

and EC_a models (Table 7.4). The MSDR values for SKlm-h are similar to those for KED, and the MAEs are slightly less than for OK for all combinations of ancillary data apart from aerial[91] and elevation on their own. Nevertheless, the MAEs are generally not as small as those for CK. For SKlm-h, the best results are obtained for

a combination of aerial^{91}, EC_a and elevation as ancillary data. As one might expect for SKlm-h, because it assumes stationarity of the regression coefficients across the entire site, there is a moderate negative relationship (−0.562) between the R^2 of the regression model and the MAEs from cross-validation, whereas for KED this relationship is much weaker (−0.210). This suggests that the R^2 of the regression model can be used as a guide for choosing an appropriate combination of variables for use with SKlm-h, whereas this is not advisable for KED where the regression coefficients are locally re-evaluated.

The MAE for SKlm-s (2.824) is promising and less than that for OK. The MSDR is less than one which indicates the prediction errors are over-estimated by the kriging variance, nevertheless, the value is reasonable and of a similar magnitude to that for KED (Table 7.4). These results show that SKlm-s using soil series data is also a viable alternative to OK.

7.3.5 Patterns of Variation

Figure 7.5 shows the maps of percentage sand predicted at the nodes of a 5-m grid by: (1) ordinary kriging of the 30-m sand data, (2) simple kriging of soil series residuals followed by the addition of mean sand content for the corresponding soil series at each node, (3) after determining the best combination of ancillary data by cross-validation for SKlm-h (Table 7.4), and adding the kriged regression residuals to sand values obtained using the coefficients from the appropriate regression model, (4) for KED, the residual variogram (Table 7.2) for the best combination of ancillary data (Table 7.4) was used with the appropriate ancillary data (external drifts) for kriging sand content to the 5-m grid and (5) ordinary cokriging of the 30-m sand data using the best combination of ancillary data (Table 7.4). All maps show similar patterns of variation to those for sand predicted by OK (Fig. 7.5a), particularly those based on SKlm-s and CK (Fig. 7.5c and e). The maps of KED and SKlm-h predictions show the more detailed patterns present in aerial^{91} in the north and west of the field (Fig. 7.1f). This level of detail might be unnecessary for precision farming because it will depend on the smallest area the machinery can manage reliably.

The widespread application of CK has been hindered by the strict requirements of the LMC. In the above analysis elevation and yield^{98} data could not be used for CK because the variograms were unbounded, whereas those for sand and the other ancillary data are bounded. If a significant proportion of the variation in a property is accounted for by trend, even though a LMC can be fitted to the residuals from the trend, there is often little or no improvement in prediction because the trend might account for most of the coregionalization (Kerry 2004). Kerry (2004) analysed data from seven fields, and fitted LMCs to trend residuals where >20% of variation was accounted for by trend. She found that improved accuracy of prediction by CK over OK was small for data with marked trend and greatest at the sites with least trend. Cokriging is usually implemented with a local search window, and requires an assumption of stationarity only within that neighbourhood (quasi-stationarity). Some

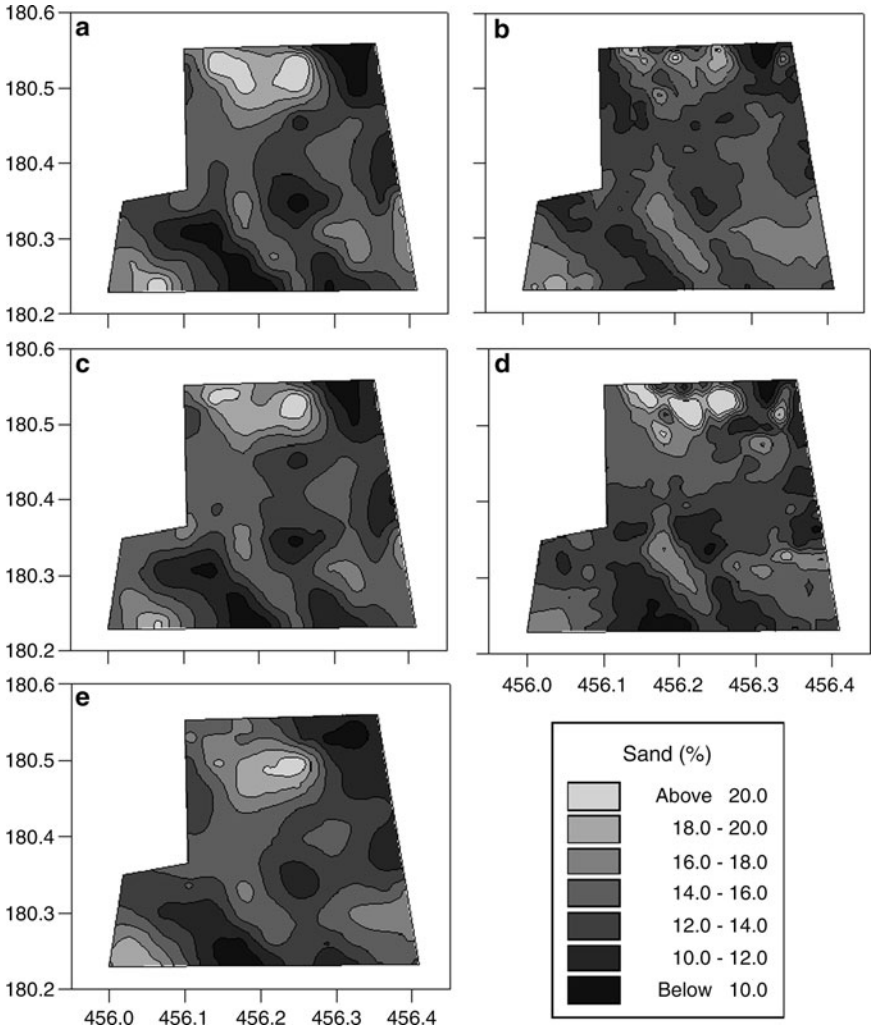

Fig. 7.5 Maps of sand content at Yattendon produced using: (**a**) ordinary kriging (OK), (**b**) simple kriging with local means (SKlm-h) with aerial[91], EC_a and elevation (**c**) simple kriging with local means (SKlm-s), (**d**) kriging with an external drift (KED) with aerial[91] and (**e**) cokriging (CK with aerial[91] and EC_a).

authors such as Bishop and McBratney (2001) recommend the use of SKlm or KED when there is marked trend in the ancillary data. Alternatively fit an unbounded LMC and use CK with a local search window and then trend need not be removed from the data prior to modelling the cross-variogram and cokriging.

The real test of methods that use ancillary data is the accuracy of prediction when there are few data for the primary variable, for example 50 data, and when parts of the field, the centre in Fig.7.1b, are sparsely sampled. Figure 7.6 enables

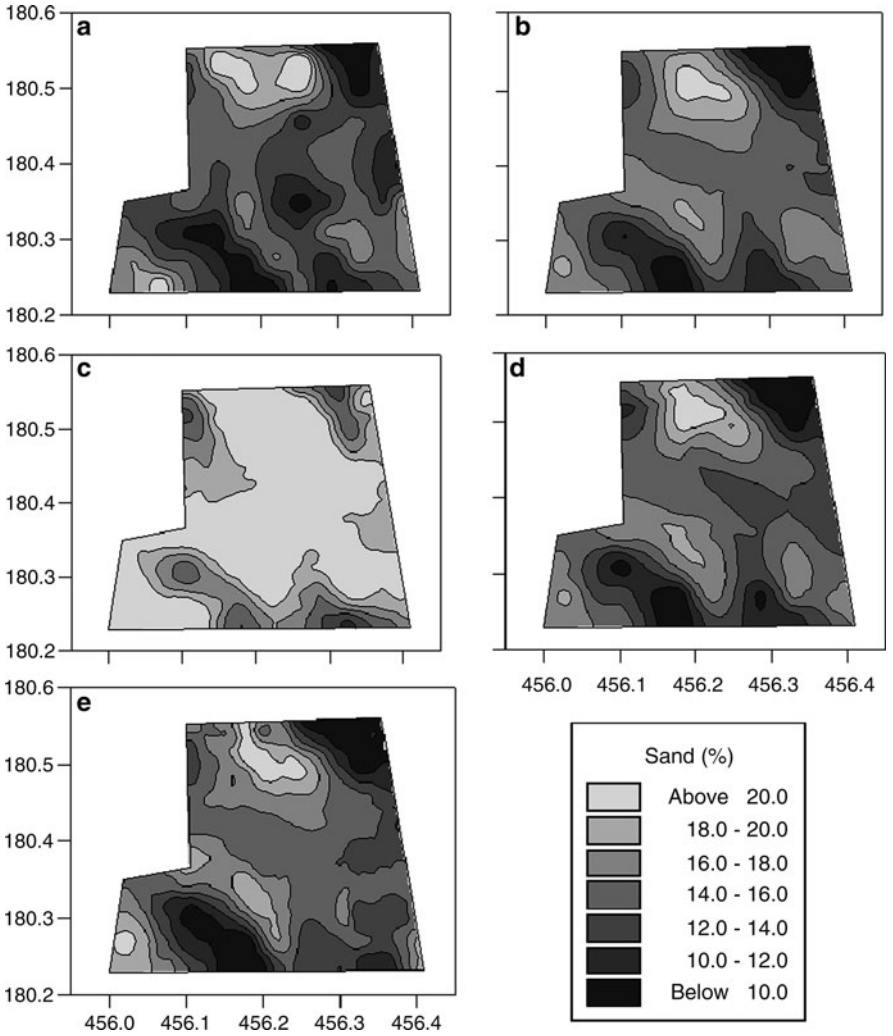

Fig. 7.6 Maps of sand content at Yattendon produced using: (**a**) ordinary kriging (OK) and the full data, (**b**) OK and 50 data, (**c**) simple kriging with local means (SKlm-hard data) using aerial[91], EC_a and elevation with 50 data, (**d**) simple kriging with local means (SKlm-soft data) and 50 data, and (**e**) kriging with an external drift (KED) using aerial[91] and EC_a with 50 data

a comparison of the maps of predictions produced by the above methods with only 50 data to that from OK with the 30-m data. Except for SKlm-h (Fig. 7.6c), the maps of sand content based on 50 data and ancillary data are more similar to the reference map created with the 30-m data, than is the ordinary kriged map of sand based on 50 data. This is especially so in the center of the field.

7.3.6 How Small Can the Sample Size of Primary Data be when Secondary Data are Available?

The best combination of ancillary data identified by cross-validation (Table 7.4), namely aerial^{91} and EC_a for KED and aerial^{91}, EC_a and elevation in combination for SKlm-h was used in a jack-knife procedure. This involved making 100 random selections of a given sample size from the original 102 soil data (Fig. 7.1a). The selected points (the prediction points) and the associated ancillary data were used to compute the appropriate regression model and residual variogram, and then SKlm-h or KED was carried out at the remaining unselected points (the validation points). The jack-knife procedure is valuable for determining the best methods of prediction because 100 random samples of different size are investigated rather than just one sample of a given size as in LOO cross-validation. This approach was taken because if only one random sub-sample of a given size is selected, it does not necessarily give a good indication of how other possible sub-samples with that number of observations might perform. For OK, KED and SKlm-h the jack-knife procedure was repeated for sample sizes of 10, 20, 30, 40, 50, 60, 70, 80 and 90 from which to predict. The sample size of 10 was included to show that once the sample falls below a certain size, the increase in errors is almost exponential. We suggest that such a small sample size should not be considered for geostatistical analysis either with or without secondary information. The statistics used to evaluate the results from this are the same as those for LOO cross-validation, except that they were calculated 100 times for each random sub-sample. The mean and standard deviation of these are used in Figs. 7.7–7.10 to summarize the results.

For all methods the ME is close to zero for all sample sizes apart from 10, therefore it is not considered further here because it was of little value in determining which methods performed best.

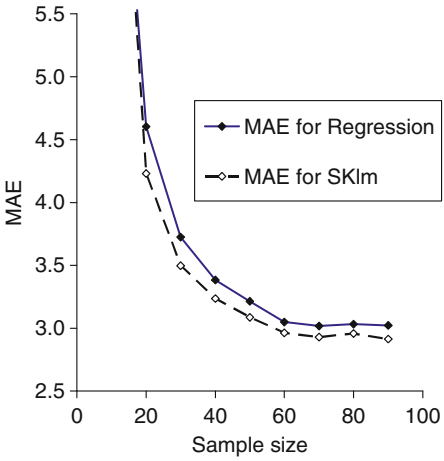

Fig. 7.7 Comparison of the mean absolute error from regression (MAE-reg) and simple kriging with local means (hard data) (MAE-SKlm) for different sizes of sub-sample from a jack-knife procedure. Both procedures use aerial^{91}, EC_a and elevation as the ancillary data

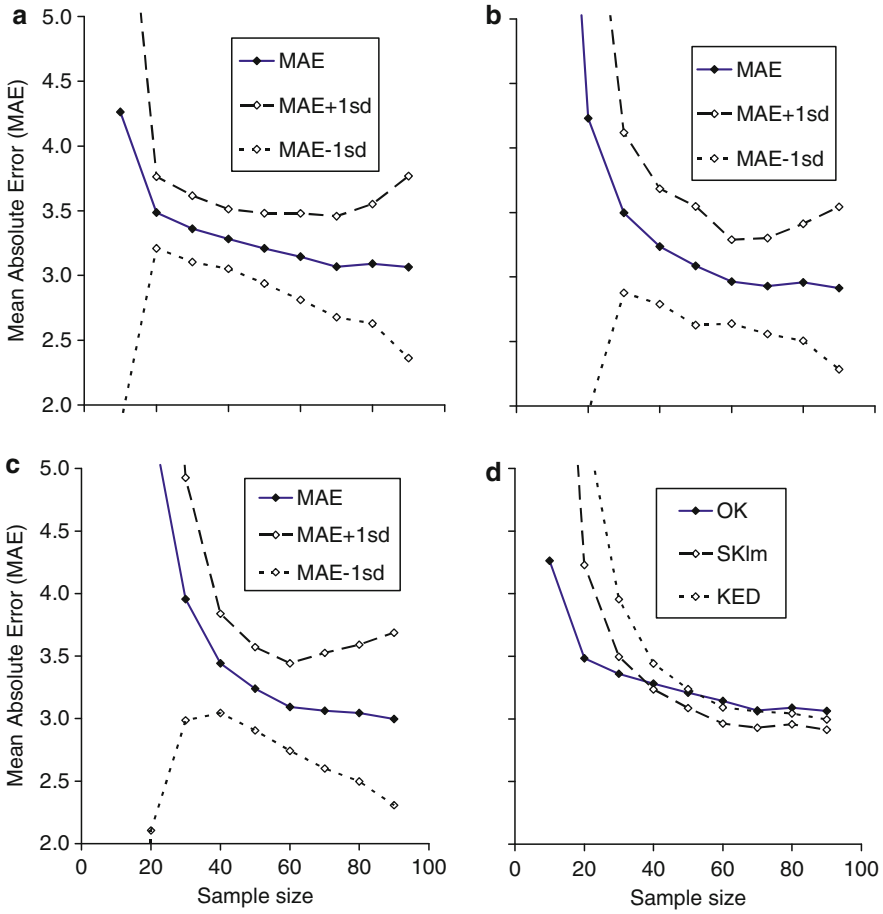

Fig. 7.8 Graphs of mean absolute error (MAE) for different sizes of sub-sample in a jack-knife procedure using (**a**) ordinary kriging (OK), (**b**) simple kriging with local means (SKlm-h) using aerial[91], EC_a and elevation, (**c**) kriging with an external drift (KED) using aerial[91] and EC_a and (**d**) a comparison of methods (**a**–**c**)

The MAE is consistently larger for regression than for SKlm-h (Fig. 7.7). Thus, kriging the residuals and adding these predictions to the output of the regression model is more accurate than simple regression for all sample sizes. The MAEs for all methods follow the same pattern: as expected the statistic decreases sharply as more data are added until there are 60 data at which point the benefit of adding more levels off.

Figure 7.8 shows the MAEs for OK, SKlm-h and KED (Fig. 7.8a–c). For OK the MAEs are relatively constant for sample sizes between 70 and 90, whereas for SKlm-h and KED they are similar for sample sizes between 60 and 90. This result suggests that similar prediction accuracy can be achieved with slightly smaller sample sizes when ancillary data are incorporated into the interpolation process. For all

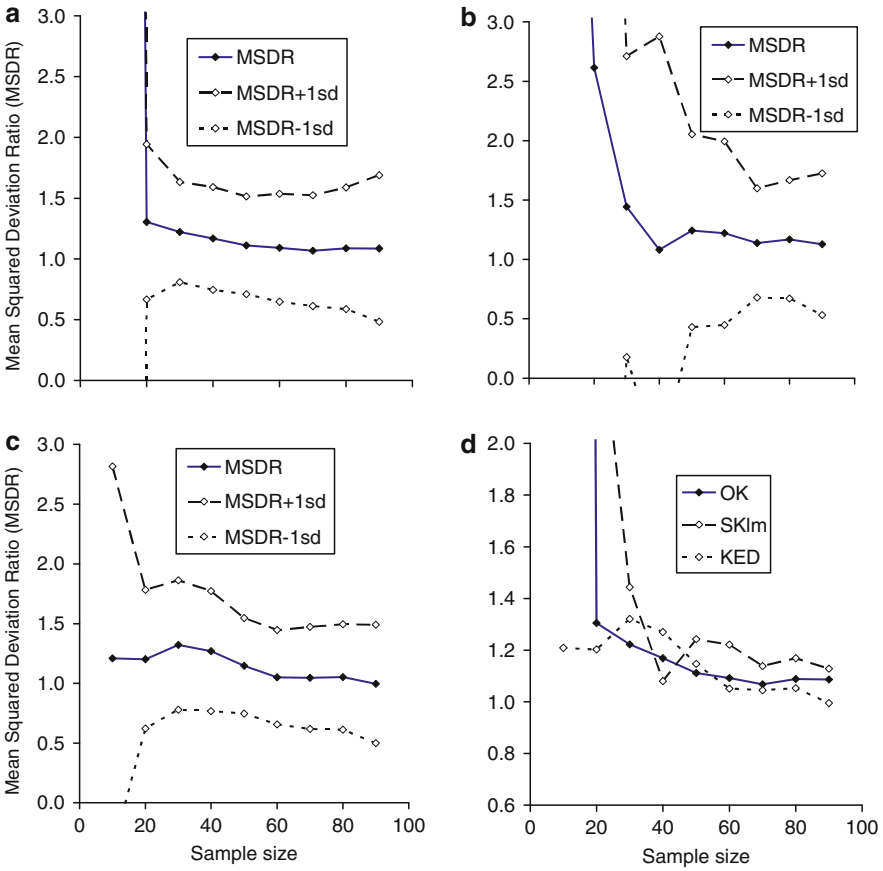

Fig. 7.9 Graphs of mean squared deviation ratio (MSDR) for different sizes of sub-sample in a jack-knife procedure for: (**a**) ordinary kriging (OK), (**b**) simple kriging with local means (SKlm-h) using aerial[91], EC_a and elevation, (**c**) kriging with an external drift (KED) using aerial[91] and EC_a and (**d**) a comparison of methods (**a**–**c**)

methods the MAE increases greatly for a sample size of 10, but this also occurs for a sample size of 20 for SKlm-h and KED. These results show, as one might expect, that these sample sizes are far too small even when ancillary data are available. There is a marked increase in MAE with SKlm-h and KED for sample sizes <40. Figure 7.8d shows that for sample sizes of 40 or more SKlm-h and KED result in smaller MAEs than OK, but for sample sizes <40 OK performs better than SKlm-h and KED. This result suggests the need for enough observations to calibrate the relationship between ancillary data and the primary variable. The MAEs for KED are consistently larger than for SKlm-h (Fig. 7.8d) for all sample sizes, which indicates that the regression coefficients are reasonably stable across the field. Theoretically, with the same ancillary variables and a global search window, KED will always perform better than SKlm-h (Papritz 2008). However, using a global search window is

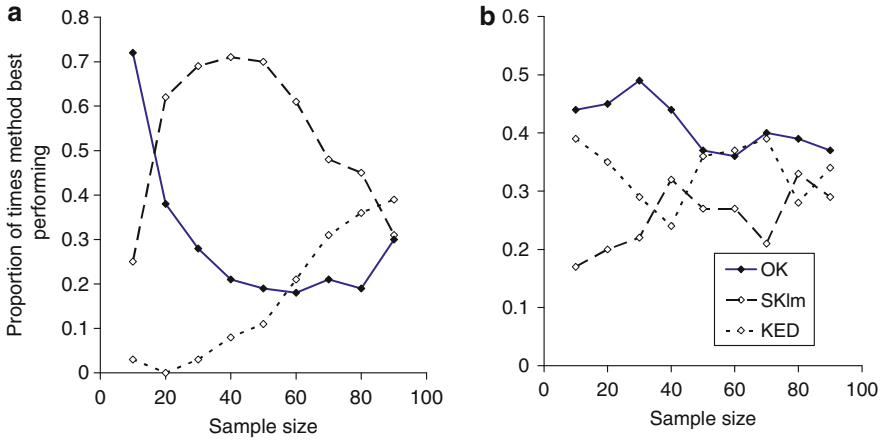

Fig. 7.10 Number of times three methods (OK, SKlm-h and KED) are the best performing in a jack-knife procedure using 100 randomly selected samples of different size at the Yattendon site. Jack-knife evaluation criteria: (**a**) mean absolute error (MAE) and (**b**) mean squared deviation ratio (MSDR)

computationally inefficient, especially for large ancillary datasets. In addition, the relationships between soil properties and other covariates typically change across the site, which violates the assumption of stationarity of the regression coefficients within the search window. A moving correlation analysis would establish whether the correlation is stationary for the whole site before using SKlm-h or KED, and if so KED could be used with a global search window rather than SKlm-h.

Figure 7.9 shows the MSDR values for various sample sizes and methods of prediction. For OK, this statistic is close to one for all sample sizes other than 10 and increases only slightly for sample sizes <50 (Fig.7.9a). For SKlm-h, the MSDRs are close to one for sample sizes >30, but they are quite variable (Fig. 7.9b). For KED, the MSDRs depart more from 1 as the sample size decreases (Fig. 7.9c) and they are similar to those for OK for sample sizes of 60–90 (Fig. 7.9d). Although OK appears to perform best overall based on its MSDRs, none of the values is vastly different from one unless the sample size is <20 and so all models could be considered reasonable given the small sample size. The MSDR is slightly >1 for each method.

The same random sub-samples were used by each method of prediction, therefore, it is possible to determine, for a given sample size, the proportion of times (i.e. percentage of 100 subsets) one method outperforms the others. The criteria were the smallest MAE and MSDR closest to 1. For MAE, SKlm-h is consistently the best method (Fig. 7.10a) 45–70% of the time for all sample sizes except 10 and 90. The smallest MAE is given more often for KED than OK when the sample size is 60 or greater. Ordinary kriging only performs best about 20% of the time for all sample sizes greater than 40. Small sample sizes seem to affect the results of KED and SKlm-h more than OK (Fig. 7.10a). For the MSDRs, all three methods perform

similarly (Fig. 7.10b). However, OK is generally the best performing by a small margin, and this margin increases for smaller sample sizes (Fig.7.10b). Again, this result suggests the need to have enough observations to calibrate the relationship between the ancillary data and primary variable; otherwise ordinary kriging might be the most conservative choice.

The ME and MSDR values were generally best for OK, but the values of these statistics for all methods were similar. The greatest contrast among methods was observed for the MAE: SKlm-h was the best performing method for all sample sizes and KED was preferable to OK for sample sizes greater than 60. The jack-knife cross-validation results indicate that on average for a sample size of 60 and ancillary data, the MAEs are similar to those obtained by OK with 90 data, but it is not advisable to use fewer than 60 data.

Precision farming practitioners could use information from this analysis to determine which methods to use for a given sample size. However as mentioned in Chapter 2, simple random sampling (as used in the jack-knife procedure) is generally not a good sampling strategy for geostatistical analysis. Most precision farmers sample on a square grid because it is practical to work with and can provide good cover of the field. Chapter 2 suggests that some degree of nested sampling within a grid will help to resolve the small scale variation.

7.4 Case Study 2: The Wallingford Site

The final step in evaluating the merits of the different interpolation methods is a case study where LOO cross-validation results for KED, OK, SKlm-h and SKlm-s are computed for sampling grids with different intervals and with some nesting.

7.4.1 Site Description and Available Data

The study site is a 43.5 ha field with soil developed on the plateau gravels of the Thames valley near Wallingford in Oxfordshire England. The topsoil (0–15 cm) was sampled using the same 30-m grid spacing and bulking strategy as at Yattendon. Anomalous areas of the field, such as a disused gravel pit in the far north-west and the headlands, were not sampled. The soil property of interest was loss on ignition (LOI), which is often used as an indication of soil organic matter content. The LOI (500°C) was determined for the air-dry <2 mm fraction of soil (Avery and Bascomb 1974). Available ancillary data included aerial photographs from 1966 (see Fig. 2.3a in Chapter 2) and 1997 (aerial^{66} and aerial^{97}, respectively), EC_a (see Fig. 2.3c in Chapter 2), elevation and yield for 1996–2000 (yield^{96} etc.). All ancillary data were recorded and processed as for Yattendon, except that the Massey Ferguson Fieldstar system (www.masseyferguson.com) was used to record yield. A soil series map from a previous survey (Fordham 1985) was also available; six soil series were identified in this field.

Table 7.5 Correlations between loss on ignition (LOI) and various ancillary data at the Wallingford site

	Aerial^{66}	Aerial^{97}	EC_a	Elevation	LOI	Yield^{96}	Yield^{97}	Yield^{98}	Yield^{99}	Yield^{00}
Aerial^{66}	1.00									
Aerial^{97}	0.02	1.00								
EC_a	−0.34	0.37	1.00							
Elevation	0.16	−0.55	−0.65	1.00						
LOI	−0.14	0.37	0.12	−0.26	1.00					
Yield^{96}	−0.54	0.02	0.67	−0.35	0.03	1.00				
Yield^{97}	−0.41	0.09	0.63	−0.51	0.11	0.71	1.00			
Yield^{98}	0.10	0.13	0.32	−0.19	0.18	0.35	0.30	1.00		
Yield^{99}	−0.11	−0.03	0.10	0.11	0.09	0.29	0.20	0.21	1.00	
Yield^{00}	−0.30	0.32	0.65	−0.62	0.29	0.55	0.59	0.41	0.27	1.00

The 296 soil sampling points on a 30-m grid were sub-sampled to create 60-m, 90-m, 120-m and 150-m grids with 70, 36, 23 and 15 points, respectively. The variogram ranges for soil and ancillary data, except yield, were consistently around 200–250 m (Table 2.2 and Fig. 2.3b, d in Chapter 2). Using half the range of the variogram of ancillary data suggests that 120-m would be a suitable soil sampling interval (see Chapter 2). As this is a large spacing, additional samples at shorter intervals should be included as recommended in Chapter 2. Therefore, two other sampling schemes were considered: (1) the 120-m soil sampling grid was supplemented with samples at a 60-m spacing to give 50 samples in total and (2) five sample locations (Fig. 2.11a in Chapter 2) were selected in areas of large, medium and small (LMS) DNs of an aerial photograph (Fig. 2.3a in Chapter 2) and added to those on the 120-m grid, resulting in a sample size of 38.

Variograms were computed and modelled for each ancillary variable. Ordinary kriging was then used to predict the ancillary variables at the nodes of a 5-m grid which had points in common with the 30-m soil sampling grid. Ancillary data were extracted at the grid nodes that coincided with the 30-m soil sampling grid and correlation coefficients were calculated (Table 7.5). The results indicate that all correlations with LOI are weak, but they are strongest for aerial^{97}, yield^{00} and elevation.

7.4.2 Leave-One-Out Cross-Validation Using Grid Sampled Data

Variograms were computed and modelled for LOI for each of the sub-samples. The parameters of the fitted models and the associated sub-sampled soil data were used for LOO cross-validation with OK. Table 2.6 shows how the form of the variogram changes as the sampling interval increases and sample size decreases. The method of moments variogram computed from spatially correlated data on the 120-m grid is essentially pure nugget because 23 data are too few to compute a reliable variogram (Webster and Oliver 1992). In cases where the variogram was essentially

Table 7.6 Leave-one-out cross-validation results for ordinary kriging (OK), simple kriging with local means (SKlm) and kriging with an external drift (KED) using the full and sub-sample data at the Wallingford site

Sample	Number of data	Mean absolute error (MAE)				Mean squared deviation ratio			
		OK	SKlm-h	SKlm-s	KED	OK	SKlm-h	SKlm-s	KED
30-m	296	0.41	0.37	0.43	0.41	1.05	0.87	0.95	0.88
60-m	70	0.45	0.42	0.46	0.52	0.99	0.82	0.88	0.81
120-m+60-m	50	0.23	0.22	0.39	0.26	0.60	0.40	0.80	0.31
120-m+LMS	38	0.42	0.37	0.40	0.58	0.92	0.60	0.69	0.56
90-m	36	0.46	0.43	0.41	0.60	0.91	0.75	0.72	0.75
120-m	23	0.71	0.50	0.52	1.02	0.93	0.63	0.72	0.81
150-m	15	0.50	0.14	0.40	0.01	1.74	0.46	0.62	0.05

pure nugget (120-m and 150-m data), an automatic weighted least squares algorithm was used to fit unbounded models with a very low gradient and these were used for cross-validation with OK. This practice enabled comparison of these results with the larger sub-samples; however, we do not recommend the use of such small data sets for geostatistics.

A regression model with linear and squared terms for aerial^{97}, yield^{00} and elevation was computed for each sub-sample of the soil data and variograms of the regression residuals were also computed. The parameters of the models fitted to these variograms were then used together with the appropriate soil and sub-sampled ancillary data for LOO cross-validation with KED and SKlm-h. The mean LOI for each soil series was also determined as described above for Yattendon for each sub-sample, and the residual variograms were computed and modelled. These were then used for cross-validation with SKlm-s for each sub-sample.

The results for the various methods of prediction and sub-samples are given in Table 7.6. In general, the MSDR values are closest to one for OK, but depart increasingly from one for all methods as the sample size decreases and grid spacing increases. The MSDRs for the larger sample sizes using SKlm-h, SKlm-s and KED are acceptable. The MAEs (Table 7.6) show that SKlm-h outperforms OK for all sub-samples, and SKlm-s outperforms OK for small sample sizes such as for the 90-m, 120-m+LMS DNs, 120-m and 150-m grids. Finally, KED has larger MAEs than OK for all sub-samples. These results suggest that ancillary data improve prediction with SKlm-h and SKlm-s when they are available, especially when the number of soil samples is small.

7.4.3 Patterns of Variation

The LOI data from various sub-samples were interpolated to a 5-m grid by OK. For SKlm-h the regression was applied to the 5-m grid of covariates and the resulting mean values were combined with the kriged residuals according to Eq. 7.5. Figure 7.11a shows the map of ordinary kriged predictions from the 296 data on

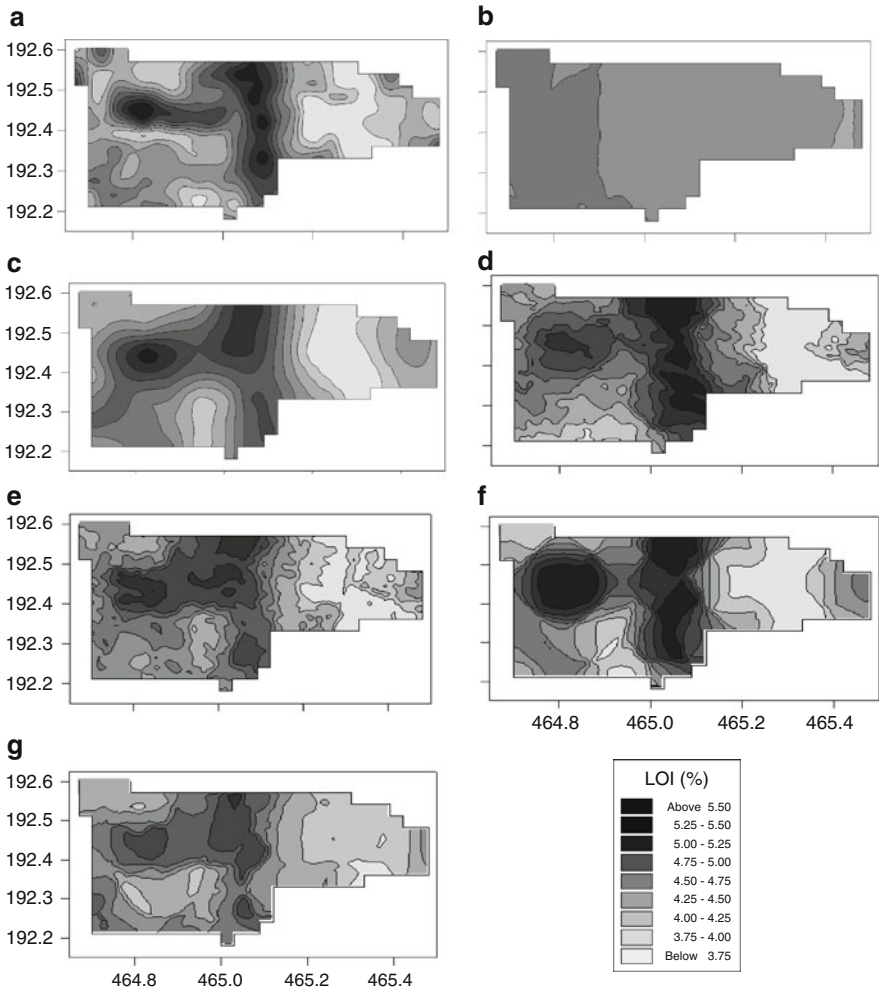

Fig. 7.11 Maps of LOI at the Wallingford site from predictions by: (**a**) ordinary kriging (OK) and the full data (296 sites), (**b**) OK and the 120-m data (23 sites), (**c**) ordinary kriging (OK) with the 120-m plus 60-m data (50 sites), (**d**) simple kriging with local means (SKlm-h) using aerial^{97}, yield^{00}, elevation and the 120-m data (23 sites), (**e**) SKlm-h using aerial^{97}, yield^{00}, elevation and the 120-m plus 60-m data (50 sites), (**f**) SKlm-s with the 120-m data (23 sites), (g) SKlm-s with the 120-m plus 60-m data (50 sites)

the 30-m grid. This can be used as a reference to compare the variation resolved by various sub-samples and interpolation methods. We focus on the 120-m sampling interval and 120-m plus 60-m data as these would be recommended to the practitioner based on the ancillary variograms (see Chapter 2 of this book). When data on the 120-m grid (23 points) are used alone for OK (Fig. 7.11b), the main patterns of variation are no longer evident. However, the main patterns of variation

are preserved when the 120+60-m data (50 points) are used for OK (Fig. 7.11c). Figure 7.11d shows that when SKlm-h is used to interpolate the 120-m data with ancillary data the patterns of variation are even more similar to those for the 30-m soil data (Fig. 7.11a). There is little improvement over these results when the 120-m plus 60-m data are used for SKlm-h (Fig. 7.11e). The improvement with ancillary data is most evident for the 120-m data and this is also reflected in the LOO cross-validation results where the difference between the MAEs for OK and SKlm-h is greatest for 120-m grid and smallest for the 120-m plus 60-m data (Table 7.6). Figure 7.11f and g show the predictions from the 120-m grid and 120-m plus 60-m data using SKlm-s. More of the variation has been resolved than with OK for the 120-m data, but not for the 120-m plus 60-m data. These results again confirm those from cross-validation (Table 7.6). These results illustrate to practitioners that if the only ancillary data available to them are soil series maps, they can be used to improve estimates from sparse, spatially structured soil data.

The LOO cross-validation and jack-knife statistics give a good indication of which method performs best. However, these cross-validation statistics do not convey the full benefit of ancillary data for mapping soil properties, especially when the sample size is small because only the ancillary data that are collocated with the soil data are used and not the full dataset. When the sample size is very small, the jack-knife results show that it can be difficult to calibrate the regression model on which KED and SKlm-h are based using only the nearest ancillary data to the soil data. The maps in Fig. 7.6 and particularly Fig. 7.11 show that when all the data from ancillary variables are used to interpolate to a dense grid, it is generally beneficial to use a multivariate technique rather than OK even when the sample size is small.

7.5 Conclusions

This Chapter has introduced the three most common geostatistical approaches for incorporating ancillary data into the prediction of soil properties: cokriging (CK), kriging with an external drift (KED) and simple kriging with local means (SKlm). Each method has its advantages and disadvantages, and their prediction performances are site-specific as illustrated by the two cross-validation studies. Cokriging, unlike KED and SKlm, does not require the ancillary data to be at all locations on the prediction grid, however, its application is more demanding in terms of computing and modelling the variogram. It also becomes more complex when there is significant trend in the primary or ancillary data; some authors recommend the use of KED and SKlm if trend is present. Simple kriging with local means allows the straightforward incorporation of both categorical and continuous attributes and can be implemented using common geostatistical software once regression residuals have been computed. Kriging with an external drift allows one to account for spatial changes in the relationship between variables, but this relationship must be linear and most software accommodates only a single covariate. Correlation or regression analyses with a moving window could be used to determine whether KED or SKlm

is the more appropriate. If the correlations between variables are stationary over the field, then KED with a global search window outperforms SKlm, but this can be computationally intensive. If the secondary variables are not sampled exhaustively, practitioners should use either cokriging or predict the variables to the nodes of the prediction grid before applying KED or SKlm.

Cross-validation should be used to select the method that provides the most accurate predictions for a given site. The results described in this chapter suggest that any method of prediction that incorporates ancillary data with only weak-moderate (0.25–0.5) correlation with the soil property is preferable to ordinary kriging, especially when there are few data. Such methods can improve the quality of the digital maps that farmers need based on sparse, yet spatially structured soil data with little additional expense. Most farmers have some kind of ancillary data and incorporating them into the prediction process would improve information on the variation at little or no extra expense. We recommend that practitioners use a practical approach to incorporating secondary information into the interpolation process that considers the available data, the strength of correlation between primary and secondary variables, and their access to and expertise in using appropriate software.

References

Avery, B. W., & Bascomb, C. L. (1974). *Soil survey laboratory methods*. Soil survey technical monograph No. 6. Harpenden, UK: Soil Survey of England and Wales.

Baxter, S. J., & Oliver, M. A. (2005). The spatial prediction of soil mineral N and potentially available N using elevation. *Geoderma*, *128*, 325–339.

Bishop, T. F. A., & McBratney, A. B. (2001). A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, *103*, 149–160.

Dobermann, A., & Ping, J. L. (2004). Geostatistical integration of yield monitor data and remote sensing improves yield maps. *Agronomy Journal, 96*, 285–297.

Fordham, S. J. (1985). *Soils of Crowmarsh Battle Farms*. Vol. private, Soil Survey of England and Wales.

Ge, Y., Thomasson, J. A., Morgan, C. L., & Searcy, S. W. (2007). VNIR diffuse reflectance spectroscopy for agricultural soil property determination based on regression-kriging. *Transactions of the ASABE, 50*, 1081–1092.

Godwin, R. J., & Miller, P. C. H. (2003). A review of the technologies for mapping within-field variability. *Biosystems Engineering, 84*, 393–340.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Goovaerts, P. (1999). Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89, 1–45.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228, 113–129.

Heming, S. (1997). *Soil survey and soil data base of part of Westridge Farm and Mapletons Farm for Yattendon Estates*. Winchester, UK: Soil Services Limited (private report).

Kerry, R. (2004). *Determining the effect of parent material and topography on the spatial structure of variation in soil properties for precision agriculture*. Unpublished PhD. Thesis, University of Reading, Reading, England.

Kravchenko, A. N., & Robertson, G. P. (2007). Can Topographical and yield data substantially improve total soil carbon mapping by regression kriging. *Agronomy Journal*, *99*, 12–17.

Kozar, B., Lawrence, R., & Long, D. S. (2002). Soil phosphorus and potassium mapping using a spatial correlation model incorporating terrain slope gradient. *Precision Agriculture, 3*, 407–417.

Lesch, S. M., & Corwin, D. L. (2008). Prediction of spatial soil property information from ancillary sensor data using ordinary linear regression: model derivations, residual assumptions and model validation tests. *Geoderma, 148*, 130–140.

Papritz, A. (2008). Standardized vs. customary ordinary cokriging: some comments on the article "The geostatistical analysis of experiments at the landscape-scale" by T. F. A. Bishop and R. M. Lark. *Geoderma, 146*, 391–396.

Tarr, A. B., Moore, K. J., Burras, C. L., Bullock, D. G., & Dixon, P. M. (2005). Improving map accuracy of soil variables using soil electrical conductivity as a covariate. *Precision Agriculture, 6*, 255–270.

Triantafilis, J., Odeh, I. O. A., & McBratney, A. B. (2001). Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. *Soil Science Society of America Journal, 65*, 869–878.

Vitharana, U. W. A., Van Meirvenne, M., Cockx, L., & Bourgeois, J. (2006). Identifying potential management zones in a layered soil using several sources of ancillary information. *Soil Use and Management, 22*, 405–413.

Wackernagel, H. (1995). *Multivariate geostatistics*. Berlin: Springer.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications, 3rd edition*. Berlin: Springer.

Webster, R., & Oliver, M.A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science, 43*, 177–192.

Chapter 8
Spatial Variation and Site-Specific Management Zones

R. Khosla, D.G. Westfall, R.M. Reich, J.S. Mahal and W.J. Gangloff

Abstract Many approaches have been proposed over the last two decades for managing the spatial variation of soil and crops. In this chapter we discuss the importance of quantifying and managing spatial variation in crop production fields to implement site-specific crop management. We outline the challenges that soil and crop scientists have addressed since the inception of precision agriculture (PA) in terms of managing soil spatial variation, and the development of simple, stable and inexpensive techniques for quantifying and managing it with tools such as site-specific management zones. This chapter summarizes and cites the work of several scientists who have worked in the area of development and evaluation of site-specific management zones from around the world. Geostatistics is being applied increasingly in PA because of the need for accurate maps on which to base site-specific management. For soil and crop properties that require costly sampling and analysis, there are often insufficient data for geostatistical analyses and this chapter shows how management zones can provide an interim solution to more comprehensive site-specific management. Physical and chemical soil properties have been the most widely used properties for delineating management zones, however, intensive data from remote and proximal sensors are being used increasingly. The case study describes methods of delineating and evaluating management zones.

Keywords Management zones · Delineation · Site-specific crop management · Physical and chemical soil properties · GIS data layers · Bare soil imagery · Yield

R. Khosla (✉), D.G. Westfall and W.J. Gangloff
Department of Soil & Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA
e-mail: Raj.Khosla@colostate.edu; dwayne.westfall@colostate.edu; billgangloff@gmail.com

R.M. Reich
Department of Forest, Range and Watershed Stewardship, Colorado State University, Fort Collins, CO 80523, USA
e-mail: robin@warnercnr.colostate.edu

J.S. Mahal
Department of Farm Power and Machinery, Punjab Agricultural University, Ludhiana 141004, India
e-mail: jsmahal@gmail.com

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_8,

8.1 Introduction

Spatial variation in soil properties exists within fields, farms and across landscapes. Although spatial variation in agricultural fields has received considerable attention recently, its importance and impact on crop management has been discussed for over a century. Mercer and Hall (1911) examined the variation in crop yields in small plots in fields at Rothamsted Research, Harpenden, England. Later Waynick and Sharp (1919) reported variation in soil nitrogen and carbon in field trials in a series of studies and its impact on the accuracy of field trials. They found that the differences between samples taken at shorter intervals were significantly less than those for the population as a whole (Waynick and Sharp 1919). Several studies since then have shown that soil properties, as well as many natural resource properties, vary continuously in space (Haradine 1949; Hammond et al. 1958; Peterson and Calvin 1986; Cipra et al. 1972; Vieira et al. 1982). Likewise, crop producers and farm managers have observed spatial and temporal variability in their fields for a long time. Until recently, they did not have access to appropriate tools and technology to manage the variation. Since variation in soil properties greatly influences crop productivity across a field, it was evident early on that successful implementation of precision agriculture technologies for managing agricultural fields would require the spatial variation to be quantified accurately.

Geostatistical techniques have been adopted with some enthusiasm in PA because of their suitability for quantifying and predicting the spatial variation of soil, crop and landscape properties (see Chapter 1 for some of the background). Natural systems in the environment usually show structured or periodic variation in time or space (i.e. spatial or temporal dependence, see Chapter 5). This is particularly true for soil systems where patterns develop as a result of variation in topography, parent material, climate and biology. The consequence of spatial dependence is that samples separated by small distances tend to be more similar than those further apart. Classical statistical procedures on the other hand assume that data are spatially independent. Geostatistics is a collection of statistical methods that have been used for some time in the geosciences. The basis of the methods is to describe and model spatial dependence or autocorrelation among sample data, and to use this information for various types of spatial prediction. There is some overlap with GIS (geographic information systems) and spatial statistics more generally.

There are two major components of a geostatistical analysis: modelling spatial dependence in the form of a correlogram or variogram, and predicting variable values at unsampled locations with techniques such as kriging or cokriging (see Chapters 1 and 7, in particular for more theoretical background to these techniques). Geostatistics can provide accurate maps for the successful implementation of variable-rate prescription for site-specific nutrient management and other applications in precision agriculture, such as irrigation.

If the objective is to quantify the spatial variation in a given field, the sample design used to obtain data is important. Geostatistics places a different emphasis on the approach to sampling from that used in conventional statistics (see Chapters 2 and 3). Classical methods of sampling based on randomization of the

sampling positions aim to avoid spatial correlation because of the assumptions that underpin many conventional statistical techniques. In geostatistics, the aim of sampling is to ensure that the data will be spatially correlated and randomization is no longer a requirement. Randomization in geostatistics is a feature of the model rather than a property of the phenomenon of interest. Furthermore, geostatistics changes the emphasis from the estimation of regional averages in classical statistics to the local estimation of spatially distributed variables using techniques such as kriging and cokriging.

The costs of obtaining enough data for geostatistical analysis often preclude its application and there is a need for an alternative approach to site-specific management that does not necessarily depend on spatially dependent information about the soil. In the last decade there has been considerable research in PA on the delineation of zones for crop management. They enable site-specific crop management without the high costs associated with sampling and analysis for geostatistical analyses. The basis for delineating management zones is to identify soil and crop properties that: (i) are easy and or inexpensive to measure; (ii) are temporally stable (do not change in the short term, i.e. year to year) and (iii) can characterize variation in crop yield accurately. The resulting zones should be simple, stable, accurate and inexpensive to identify, and enable within-field spatial variation to be managed. Zones within a field can be classified into areas that need to be managed differently (Fig. 8.7), and this is the main focus of this chapter.

8.2 Quantifying Spatial Variation in Soil and Crop Properties

To interpolate or predict spatially a continuous or discrete variable, it is important to know how it changes throughout the area of interest. For example, if the variable is distributed randomly, then the best local estimate would be the global mean. If, as is usually the case, the variable shows some spatial structure, for example patches of soil with large or small organic matter (OM) contents, it should be possible to describe this change quantitatively. To describe the variation requires a suitable sampling design, otherwise the results of statistical analysis and or interpolation could be disappointing and sometimes misleading.

Several sampling designs have been proposed and tested to quantify the spatial variation of soil properties in PA. For example, simple random sampling, systematic sampling on a grid and stratified random sampling where a field is stratified into m squares. The latter has been popular and widely reported in PA, and is sometimes referred to as grid sampling as is systematic sampling (Thom et al. 2003). The scale at which sampling is done commercially for soil surveys by agricultural service providers is at about one sample per hectare or more due to economic constraints (Landwise Inc. 2006; Thom et al. 2003; Rehm et al. 2001; Pachepsky 1998). However, an important consideration when designing a survey for spatial interpolation is the scale of variation of the property at the level of investigation, i.e. the field for PA. The main limitation of using such large separating distances between sample locations is that the scales of variation associated with soil and crop variables are

much smaller (Gangloff 2004). For example, if one soil sample per hectare is taken in a field where variation is over distances of say 30–40 m, the sampling design will not resolve the variation adequately for interpolation. It should be noted that any method of interpolation, geostatistical or otherwise requires data that are spatially correlated or dependent.

Mueller et al. (2001) concluded that a commercial sampling approach on a grid of 100 × 100 m (1 ha) was inadequate for developing sound soil nutrient maps, whereas Franzen and Peck (1995) found that distance of 65-m (0.42 ha) between samples was necessary. Hammond (1993) also found that a 60 × 60 m (0.36 ha) grid was adequate for developing nutrient concentration maps. Gangloff (2004) investigated the spatial variation of selected soil properties using a stratified random sample with a grid of 76 × 76 m (0.58 ha) and a finer grid of 15 × 15 m (0.023 ha). Gangloff (2004) used a G-statistic (Agterberg 1984) to evaluate the 'goodness-of-prediction' from kriging selected soil properties. For the coarser grid the G-statistic was close to zero, and Gangloff (2004) concluded that the poor results were because some sample locations appeared to be anomalous making it difficult to predict accurately near to them and the interval of the coarse grid was too large. Analysis of the data on the fine grid indicated that the scale of spatial dependence for the soil properties was 20–30 m, which is much smaller than the interval of the coarse grid. In practical terms, these results suggest that the grid size used by commercial practitioners makes accurate prediction of soil properties impossible because of a lack of spatial dependence among the sample data. McBratney and Pringle (1999) in a separate study previously suggested that strata of 20- to 30-m are required for site-specific applications of agricultural inputs. While it is difficult to suggest an ideal grid size to characterize soil variation accurately, variation in soil properties appears to occur at a much finer scale than the 1 hectare strata commonly used in PA.

A coarse grid also limits the number of samples available to estimate the variogram; sample sizes generally range from 35 to 76 observations. Small sample sizes, such as these, affect the accuracy of the variogram estimated by the usual method of moments (Webster and Oliver 1992). Residual maximum likelihood (REML) (Cressie 1993) is a possible alternative when there are >50 and <100 data (see Kerry and Oliver 2007). Irrespective of the method used to estimate the variogram, if the resulting model does not describe the spatial structure of the variable adequately, any kind of interpolation of soil or crop properties to produce maps is unlikely to represent variation in the field accurately.

The difficulties associated with using geostatistical techniques to interpolate soil properties in PA mean that most commercial software used by crop consultants and other practitioners disregards the importance of spatial dependence. They rely on deterministic techniques such as inverse distance weighting for interpolation without considering that this method makes little sense without spatially dependent data. The problem is that such software can interpolate from any set of data regardless of whether they are spatially dependent, and so the resulting surfaces may or may not reflect the true variation in the population. Discussions with consultants and practitioners (personal communication) indicate that "*sampling at scales needed to attain spatial dependency is time consuming, laborious, expensive and not necessarily*

advantageous given the size of the large commercial farm equipment they work with to apply variable rate nutrients across a field". As sampling for geostatistics (or interpolation more generally) requires skilled labour, is time consuming, labour intensive and expensive (Khosla and Alley 1999; Khosla et al. 2002), there has been a need for alternative techniques that are efficient, simple, accurate and economic (Bullock and Bullock 2000) to characterize the spatial variation in soil and crop properties accurately. This has been achieved with management zones.

8.3 Site-Specific Management Zones

The most widely cited definition of a management zone was provided by Thomas Doerge in the late 1990s. According to Doerge (1999) "Management Zones" are "*sub-regions of a field that express a homogeneous combination of yield limiting factors for which a single crop input is appropriate to attain maximum efficiency of farm inputs*". This definition became a 'mantra' for those who started developing and evaluating techniques for delineating management zones for precision nutrient management and later for precision management of other inputs also (herbicides, water, seeds, manure, etc.). Figure 8.7 shows a field classified into separate regions of low, medium and high productivity. It illustrates some of the unique features of management zones: (i) the areas of different productivity potential may or may not be contiguous and (ii) the number of zones delineated is subjective and is a function of the technique used to delineate them and the scale of variation observed in that field. The overarching idea is to characterize within-field variation, identify yield limiting factors and classify homogenous areas into zones to manage them separately to enhance production, and to improve the efficient use of inputs and economic returns (Khosla et al. 2002; Koch et al. 2004).

Based on the number of publications (>200) found in the Agricola® and other scientific databases that reported work on some aspect of management zones, coupled with the diversity of management zone techniques reported in the literature, it may appear that they have been around for decades. However, the concept of managing crop production inputs within zones is relatively new. A review of the literature indicates that Yost et al. (1982) delineated '*zones of influence*', i.e. areas of soil with similar properties based on the 'range' of the variogram. However, the focus of that investigation was on mapping soil properties using geostatistics and not necessarily to aid crop management decisions.

A review of the literature does not provide a clear indication of when the first zone-based variable-rate nutrient management was initiated. Mulla et al. (1992) reported the results of their study in the late 1980s on nutrient management based on management zones in the Pacific Northwest of the USA. In studies prior to that by Mulla et al. (1992), recommendations were made by soil scientists for variable rates of fertilizer application based on patterns in soil fertility (Dow and James 1973). However, such recommendations could not be implemented because of the limitations associated with the technology available to map patterns in soil fertility and

to apply fertilizer at a variable rate (Mulla et al. 1992). With developments in agricultural equipment, new fertilizer applicators were introduced in the late 1980s that were capable of map-based fertilizer application (Fairchild and Hammond 1988). Around the same time advances were made in within-field mapping of soil properties with the aid of GPS and GIS, enabling map-based variable-rate fertilizer application to become a reality (Mulla and Hammond 1988; Hammond et al. 1988; Mulla et al. 1992). Grid sampling of the soil continued to play a major role in quantifying within-field variation and producing variable-rate fertilizer prescription maps, whereas management zones were in the early stages of development and evaluation. It took several years and many studies to indicate the success of management zones for quantifying within-field variation of soil and crop properties accurately before they were recognized as a viable tool for precision nutrient management.

8.3.1 Soil Properties, Crops and Geographic Distribution of Management Zones

Table 8.1 gives the frequency of various soil, crop and other properties that have been used, individually or in combination, to delineate site-specific management zones for crop management. The list of properties and frequency of their occurrence (a total of 162) were collated from over 100 refereed publications from around the world published between 1992 and 2008. Table 8.1 clearly reflects the diversity of properties (46 in total) that have been used by researchers and practitioners to delineate management zones. These properties can be divided into eight broad categories (Table 8.1). It is evident that, both physical and chemical soil properties are the most widely reported properties used for delineating management zones, followed closely by those from sensing technologies, crop properties and landscape attributes. Although sensing is a separate category in Table 8.1, the methods were used to measure aspects of both soil and crop properties. In addition, sensing technologies, such as the Veris soil electrical conductivity unit (Veris Tech., Salina, KS), remote-sensing for bare soil imagery, normalized difference vegetation index (NDVI), etc., are reported in more recent literature (Kitchen et al. 2005; Khosla et al. 2002; Inman et al. 2008) and are a reflection of the progress that has been made (i.e. accessibility to innovative sensors and tools) in the last decade in site-specific crop management. Table 8.1 also gives a few of the rare properties that have been used for delineating management zones, such as tillage depth, tillage force, weed populations and soya bean cyst nematode densities, for site-specific crop management.

Tables 8.2 and 8.3 give the frequency of the various crops and geographical locations, respectively, where site-specific management zones have been developed and evaluated or applied for site-specific crop management. Table 8.2 shows that the majority of management zone research has focused on ‘row-crops’ primarily maize (*Zea mays*), followed by soya beans (*Glycine max*), cotton (*Gossypium hirsutum*), wheat (*Triticum aestivum*), potatoes (*Solanum tuberosum*) and rice (*Oryza sativa*). Inclusion of other crops, fruits, vegetables and tree species illustrate unique

Table 8.1 A list of properties used in various techniques of delineating site-specific management zones

Category	Properties used to delineate management zones	Number of occurrences
I. Soil properties		[61][a]
Chemical	Soil organic matter and soil carbon	8
	Nitrogen	7
	Phosphorous	7
	Potassium(4), Magnesium(1) and Calcium(1)[b]	6
	Cation exchange capacity	2
	Grid sampling	2
	Soil pH	1
	Targeted sampling	1
	Gypsum requirement	1
Physical	Soil texture	13
	Soil type	7
	Soil colour	6
	Soil moisture content	4
	Aggregate stability in water test	1
	Hard pan	1
	Penetration resistance	1
	Water content/holding capacity	2
II. Landscape attributes		[18]
	Topography	14
	Aspect	2
	Curvature	1
	Other	1
III. Crop properties		[28]
	Yield map (Spatial)	14
	Yield map (Temporal)	9
	Shoot density	3
	Ground based leaf area index (LAI)	1
	Protein content (wheat)	1
IV. Sensing		[40]
	Soil electrical conductivity sensor	18
	Satellite imaginary and other platforms	12
	Normalized differential vegetative index (NDVI)	5
	Digital photography (crop canopy)	4
	Electromagnetic induction sensors	1
V. Management practice		[4]
	No-tillage, chisel-plow, with and without traffic	1
	Tillage depth	1
	Tillage force	1
	Weeding, mulching and traffic-aisles	1

(continued)

Table 8.1 (continued)

Category	Properties used to delineate management zones	Number of occurrences
VI. Weed and pest management		[4]
	Weed population	2
	Soya bean cyst nematode densities	1
	Silverleaf whitefly population	1
VII. Subjective approach		[4]
	Self made zones for soil management	4
VIII. Modeling		[3]
	Crop/simulation/GIS	3
Total		162

[a]Total of each category is bracketed.
[b]Number of occurrences for each nutrient is in parenthesis.

applications of management zones (Table 8.2). Proliferation in the use of management zones around the world has been fairly rapid. Table 8.3 shows that research on management zones has been reported from several countries across six continents. While it comes as no surprise that the majority of research has been reported from the USA, followed by Europe, it is still encouraging that use of management zones is proliferating in other parts of the world.

Although the original primary purpose of management zones was to replace grid sampling for site-specific nutrient management, a review of the literature (Tables 8.1–8.3) indicates that management zones have been developed for managing a wide variety of crop inputs. A partial list includes management of crop irrigation, manure, weeds and pests, tillage, in addition to crop nutrients such as N, P, K, Ca, Mg and Fe, ameliorants such as gypsum, and characteristics of crop quality such as the protein content of wheat or wine quality from grapes, etc. Tables 8.1–8.3 provide an overview and insight into the diversity as well as the similarities among the parameters used globally in developing and delineating management zones for various crops that have been reported in the literature.

8.3.2 Techniques of Delineating Management Zones

There are numerous techniques for delineating management zones across a field. Over the years, these techniques have evolved and have been transformed from being primarily soil property based to minimally intrusive (i.e. do not rely strictly on soil sampling) and, therefore, have the potential to be more economically feasible than grid sampling for variable-rate management (Hornung et al. 2006). Management zones may be delineated based on a single soil or crop property (such as soil texture or yield) or a combination of several that are known to affect crop productivity and yield (Table 8.1). Likewise, some techniques are based on a simple process

Table 8.2 Frequency distribution of crops, fruits, vegetables and trees, reported to be managed with site-specific management zones in the literature

Category	Name of crop	Botanical name	Number of occurrences
Crops			[106][a]
	Maize	*Zea mays*	34
	Wheat	*Triticum aestivum*	17
	Soya beans	*Glycine max*	15
	Cotton	*Gossypium hirsutum*	12
	Potato	*Solamum tuberosum*	6
	Barley	*Hordeum vulgare*	4
	Rice	*Oryza sativa*	4
	Sugar beet	*Beta vulgaris*	4
	Rape	*Brassica compestris*	2
	Sunflower	*Helianthus annuus*	2
	Bean	*Phaseolus vulgaris*	1
	Flax	*Linum usitatissimum*	1
	Millet	*Pennisetum americanum*	1
	Sweet potato	*Ipomea batatus*	1
	Sorghum	*Sorghum bicolor*	1
	Sugarcane	*Saccharum officinarum*	1
Vegetables			[5]
	Cassava	*Manihot esculenta*	2
	Arrow roots	*Maranta arundinacea*	1
	Cowpeas	*Vigna unguiculata*	1
	Tomato	*Lycopersicon esculentum*	1
Fruits			[3]
	Banana	*Musa paradisiacal*	1
	Grapes	*Vitis vinifera*	1
	Pineapples	*Ananas comosus*	1
Trees			[2]
	Oil palm	*Elaeis guineensis*	2
Others			[2]
	Pasture		2
Total			[118]

[a]Total of each category is shown in brackets.

that involves only a clustering algorithm with one property, such as soil electrical conductivity, to classify the field into zones (Fleming et al. 2004). Conversely, there are other complex techniques that may involve a variety of GIS data layers (i.e. remotely sensed red, green and near infra-red bands; soil organic matter; soil cation exchange capacity; soil sand, silt and clay content; and the previous year's digital yield data) to create the final management zone surface (Hornung et al. 2006).

Digital soil survey maps are becoming increasingly available in the USA and have been investigated as a means of generating site-specific management zones (Franzen et al. 2002; Kitchen et al. 1998; Anderson-Cook et al. (2002). Franzen et al. (2002) found that although order 2 digital soil survey maps (i.e. map scales

Table 8.3 Frequency distribution of countries where the site-specific management zone approach has been reported in the literature

Continent	Country name	Number of occurrences
North America		[59][a]
	USA	56
	Canada	3
Europe		[16]
	England	5
	Belgium	2
	Italy	2
	Austria	1
	Czech Republic	1
	Finland	1
	France	1
	Hungary	1
	Germany	1
	Spain	1
South America		[2]
	Argentina	1
	Chile	1
Africa		[2]
	Kenya	1
	South Africa	1
Australia		[7]
	Australia	6
	New Zealand	1
Asia		[11]
	Bangladesh	2
	China	2
	Pakistan	2
	Iran	1
	Japan	1
	Malaysia	1
	Thailand	1
	Papua New Guinea	1
Total		[97]

[a]Total of each category is shown in brackets.

of 1:15 840 to 1:30 000) are readily available to farmers, they were inadequate for developing N management zones. However, Franzen et al. (2002) further reported that order 1 soil survey maps (i.e. map scales of 1:5000 to 1:10 000) were useful for developing N management zones. In a similar study, Kitchen et al. (1998) compared soil survey maps of order 2, order 1 and enhanced order 1 with a scale of 1:5000. They found that while order 1 soil survey maps (both, enhanced and standard) were

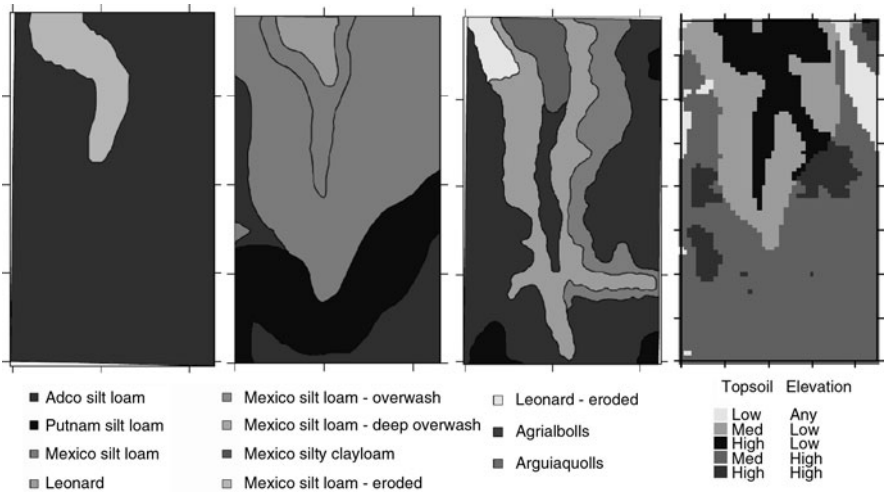

Fig. 8.1 Soil surveys conducted on the research field. From *left to right*: order 2 soil survey (1989–91), order 1 soil survey (1993), 'enhanced' order 1 soil survey (1997) and management zones created by topsoil depth and elevation (modified and adapted from Kitchen et al. 1998)

unquestionably better (Fig. 8.1) than order 2 maps, they concluded that order 2 maps were better than having no subfield delineation at all (Kitchen et al. 1998). It is important to note that order 1 soil surveys are generally not available free of charge to the public, whereas order 2 soil surveys are. Therefore, a soil consultant would have to generate a custom soil survey order 1 map of the area, which could be expensive depending on the size of the area (Hornung et al. 2006). Although order 2 soil survey maps may be a starting point for sub-field management, order 1 or enhanced order 1 soil survey maps would be needed for site-specific crop management.

As order 1 soil survey maps are expensive to generate, Anderson-Cook et al. (2002) investigated an alternative technique for soil mapping. They compared order 1 soil type maps with apparent soil electrical conductivity (EC_a) measurements and found that it was possible to classify the soil type correctly 62–81% of the time from EC_a values alone. When EC_a was used with crop yield data, the accuracy increased to between 80% and 91% (Table 8.4). This is a major contribution because soil EC_a measurements are relatively inexpensive to record compared to obtaining order 1 soil survey maps.

Previous studies with soil EC_a have shown that in addition to identifying variation in soil texture, it relates closely to other properties that often determine a field's productivity (Lund et al. 1999). Heermann et al. (1999) found that soil EC_a was the best predictor of crop yield when compared with many other common soil and crop properties. Fleming et al. (2004) used only soil EC_a to delineate management zones and found that it consistently identified areas of different productivity potential across a field. Johnson et al. (2003), however, indicated that the soil properties that control soil EC_a do not necessarily correspond to yield limiting factors. They

Table 8.4 Percentage of correct classification of order 1 soil type maps versus apparent electromagnetic conductivity (EC_a) and combined EC_a and crop yield for four soil types

Number of observations	EC_a alone	EC_a and crop yield
	% Correctly classified	
129	85.3	91.5
197	91.4	96.4
67	95.5	95.5
211	87.2	93.4
259	86.6	90.3

Modified and adapted from Anderson-Cook et al. (2002).

found that patterns on soil EC_a maps were correlated weakly with variation in corn yield and that there was no consistent relationship between EC_a-based management zones and corn grain yield (Johnson et al. 2003). It is evident that EC_a alone may not be appropriate under all crop production systems. However, there is a potential for using soil EC_a when it is combined with other soil and crop properties.

Variation in landscape attributes (topography, aspect, slope, curvature, etc.) have also been a focus of investigation for the delineation of management zones. Even before the advent of management zones, field topography was used to generate variable-rate nutrient application maps. Previous studies have described the link between field topography and soil nitrogen content (Bruulsema et al. 1996; Cassel et al. 1996) as well as topography and yield variation (Ciha 1984; Verity and Anderson 1990). Kravenchenko and Bullock (2000) found that topography together with other data such as organic matter, cation exchange capacity, phosphorus and potassium accounted for 40% of the variation in grain yield. Nolan et al. (2000) found that management zones based on elevation, curvature and slope could account for as much as 51% of the variation in crop yield. These are important findings because topography is a stable property and relatively easy and inexpensive to measure with current high resolution GPS technology.

Crop yield can be mapped easily by yield monitoring devices and the result is a reflection of within-field variation at a fine resolution. Hence, many have tried to use this valuable information to classify fields into areas of different productivity to aid management decisions. Most researchers who have used yield maps as a data layer for management zone delineation have concluded that yield maps alone are not a suitable basis for delineating zones, primarily because yield patterns are inconsistent between growing seasons (Welsh et al. 2003a,b; Godwin et al. 2003). Nevertheless, they provide valuable ancillary data (Stafford et al. 1998). For example, combining grain yield data with other soil variables could potentially explain variation associated with both crop and soil properties. This hypothesis was tested by Hornung et al. (2006) with the previous year's yield data and eight other data layers (red, green and near infra-red bands; soil organic matter; cation exchange capacity; sand, silt and clay content). They reported only marginal success with the management zone technique, primarily because of the temporally unstable nature

of the yield data. They concluded that perhaps a weighting system is needed in the delineation process of management zones that would allocate different weights to different data layers on the basis of their importance to the variation in crop production. Nanna and Franzen (2003), in a separate study, considered a weighted classification method for nitrogen zone delineation. Recently, studies have focused on several years of yield data to generate management zones and have reported significant success (Lauzon et al. 2005; Jaynes et al. 2003; Bakhsh et al. 2005). In a previous study, Moore and Wolcott (2000) correctly suggested that management zones based on several years of yield maps should be generated only after the stability of yield zones within a specific field has been tested.

Remote sensing platforms (aerial and or satellite-based passive remote sensing or ground based active remote sensing) are promising alternatives to intensive grid sampling and analysis for characterizing the spatial variation of soil and crop properties for management zone delineation and variable-rate nutrient application. Bare soil imagery, together with topography and the farmer's experience of the farm were used to delineate management zones by Khosla et al. (2002). This technique characterized grain yields accurately (Fig. 8.2) into high, medium and low productivity potential management zones in 9 out of 10 site years (Inman et al. 2005; Khosla et al. 2008). The management zones also characterized the economic returns accurately that followed the productivity potential of each zone closely (Fig. 8.3). Mzuku et al. (2005) evaluated several soil physical and chemical properties across management zones that were delineated using bare soil imagery. They found that soil properties had a significant ($p < 0.05$) correlation with low, medium and high management zones where bulk density decreased, organic carbon increased, sand content decreased and silt content increased significantly from low to high productivity potential zones (Fig. 8.4). More recent studies have focused on using remote sensing for in-season crop management. Inman et al. (2005) compared NDVI and relative maize grain yield with management zones and found similar spatial patterns among the, low, medium and high potential yield zones. They concluded that

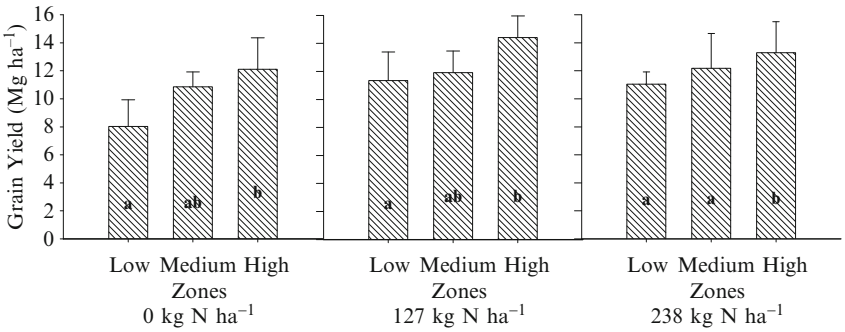

Fig. 8.2 Mean grain yield for each N application rate across site-specific management zones. *Bars with a different letter* are significantly different at $p \leq 0.05$, *bars with similar letters* are not significantly different at $p \leq 0.05$ (adapted from Inman et al., 2005)

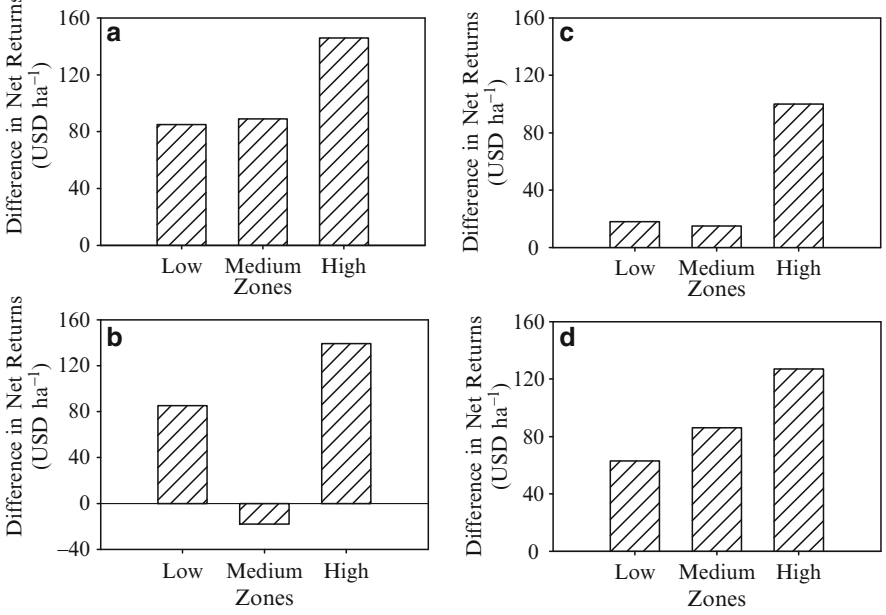

Fig. 8.3 Average differences in net returns (USD ha^{-1}) between the uniform and variable-rate nitrogen management that closely followed the productivity potential of management zones across site years: (**a**) I, (**b**) II, (**c**) III and (**d**) all years. A positive difference in net returns indicates that the variable-rate strategy performed better than the uniform rate. (Adapted from Inman et al. 2008)

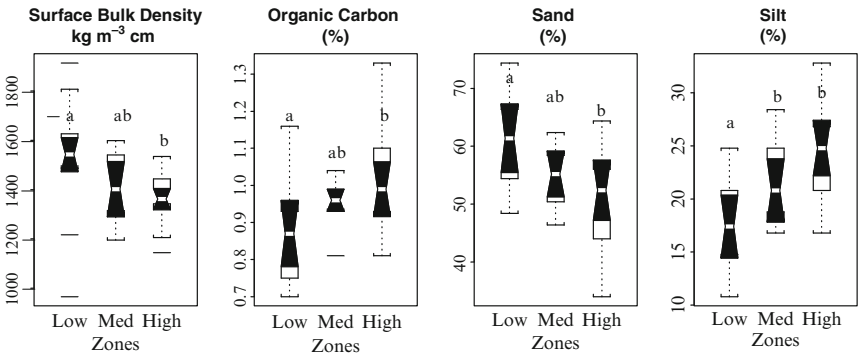

Fig. 8.4 *Box plots* of soil physical properties across site-specific management zones. Within a plot, *boxes with different letters* are statistically different at $p \leq 0.05$ (modified and adapted from Mzuku et al. 2005)

NDVI can potentially be used to model grain yield as early as the six- to eight-leaf crop growth stage in irrigated maize. Remote sensing applications in precision nutrient management with or without management zones continue to be important. Ground based active remote sensing devices hold even further potential to optimize crop nutrient management.

Techniques of management zone delineation proposed in the literature are numerous. However, to date there are few studies that have compared management zone delineation techniques on the basis of their relative performance to characterize areas of different crop productivity. A case study comparing four different management zone delineation techniques that uses a diversity of properties and complex geostatistical approaches is presented below.

8.4 Statistical Evaluation of Management Zone Delineation Techniques: A Case Study

Irrespective of the technique used to create management zones, it must be possible to characterize within-field spatial variation and classify crop yields correctly into separate productivity classes, such as low, medium and high potential production management zones. The best way to select the most appropriate management zone delineation technique would be to compare several techniques on the same field, replicated over time and space, yet few studies have done this (Fleming et al. 2004; Hornung et al. 2006; Derby et al. 2007). Gangloff (2004) did a comprehensive study on a maize field of 58 ha with centre pivot irrigation that compared four techniques for delineating management zones.

Four techniques for delineating management zones were applied to create three zones of low, medium and high productivity potential for each field. Management zone techniques ranged from simple (technique 1) to complex (technique 4). Technique 1, referred to as the soil colour management zone technique (SCMZ), uses bare soil imagery, field topography and the farmer's past management experience as three GIS data layers to delineate zones (see Khosla et al. 2002 for the detail). Technique 2, uses apparent soil electrical conductivity generated by Veris® (model 3100 EC) as a single GIS data layer to delineate management zones (ECMZ) (see Fleming et al. 2004 for the detail). Technique 3, referred to as yield based management zones (YBMZ), uses several GIS data layers, which included: multi-spectral bare soil imagery, OM, CEC, soil texture (sand, silt and clay content) and previous years' yield monitor data (see Hornung et al. 2006). Technique 4 is new and is referred to as the remotely sensed data and cluster sampling management zone technique (RCMZ). Geostatistics is used to analyse bare soil imagery and soil sample data, which are combined to delineate management zones (see Gangloff 2004 for more detail).

The RCMZ was the most complex technique among the four techniques that were compared because it uses remote sensing, soil sampling and geostatistical procedures to create the final zone map. Bare soil imagery was used to develop a directed soil sampling procedure to characterize small scale variation associated with selected soil properties (i.e. organic matter, nitrate-N (NO_3-N), zinc, electric conductivity and ammonium –N (NO_4-N)). Soil reflectance for each bare soil image was quantified into three distinct bands: blue (0.48–0.50 μm), green (0.55–0.60 μm) and red (0.62–0.68 μm) with Imagine software (Leica Geosystems® 2003). Bare

soil imagery was used to identify regions or strata with similar spectral properties. The number of strata delineated in a field was determined through a subjective iteration procedure.

1. Ten strata were initially delineated over each field using an unsupervised classification algorithm.
2. If the image appeared too pixilated and lacked visually obvious and distinctly contiguous regions, the clustering algorithm was run again after reducing the number of strata by one.
3. This procedure was repeated until visually obvious and distinctly contiguous regions were apparent.

After stratification, three sets of soil samples were obtained by sampling randomly within each stratum. At each sample location, three soil samples were taken at the vertices of an equilateral triangle and were about 5 m apart. The variation identified by these small clusters of sampling points is the small-scale spatial variation in soil properties, whereas the variation between clusters is the large-scale spatial variation.

A stepwise Akaike information criterion (AIC) was used to identify the combination of satellite bands (red, green and blue) and x,y-coordinates to include in the regression model to describe soil properties in the field. Regression and variance model parameters were estimated simultaneously by maximum likelihood (O'Connell and Wolfinger 1997). The variogram model that minimized the AIC was selected to krige the residuals from the regression model of a given soil property (fit.geospatial; S-PLUS©, Reich and Davis 1998). A spherical function provided the best fit to the experimental variograms of the regression residuals (Table 8.5 and Fig. 8.5). The variogram ranges of the soil properties have a similar magnitude, with an average range of 131 m. Figure 8.6 shows the grey scale maps from regression kriging of the five soil properties for one field. The spatial variation in these properties illustrates the difficulty of delineating management zones from several properties at the same time.

Table 8.5 Model parameters of spherical variograms for regression residuals of soil properties used as input to the RCMZ management zone delineation technique

		Model parameters			
Soil property	Regression variables[a]	Model	Nugget	Spatially correlated component	Range (m)
Organic matter	Red, green X^2	Spherical	0.0	0.034	141.0
Nitrate-N (NO_3-N)	X,Y	Spherical	0.0	7.615	137.7
Zinc	Green, blue, Y,Y^2	Spherical	0.005	0.239	101.1
Electrical conductivity	X,Y^2	Spherical	0.0	0.003	150.0
Ammonium-N (NH_4-N)	Green, X,Y,X^2	Spherical	0.243	1.694	126.9

[a]Red, green, blue – spectral bands of bare soil imagery; X,Y – geographical coordinates of soil sample.

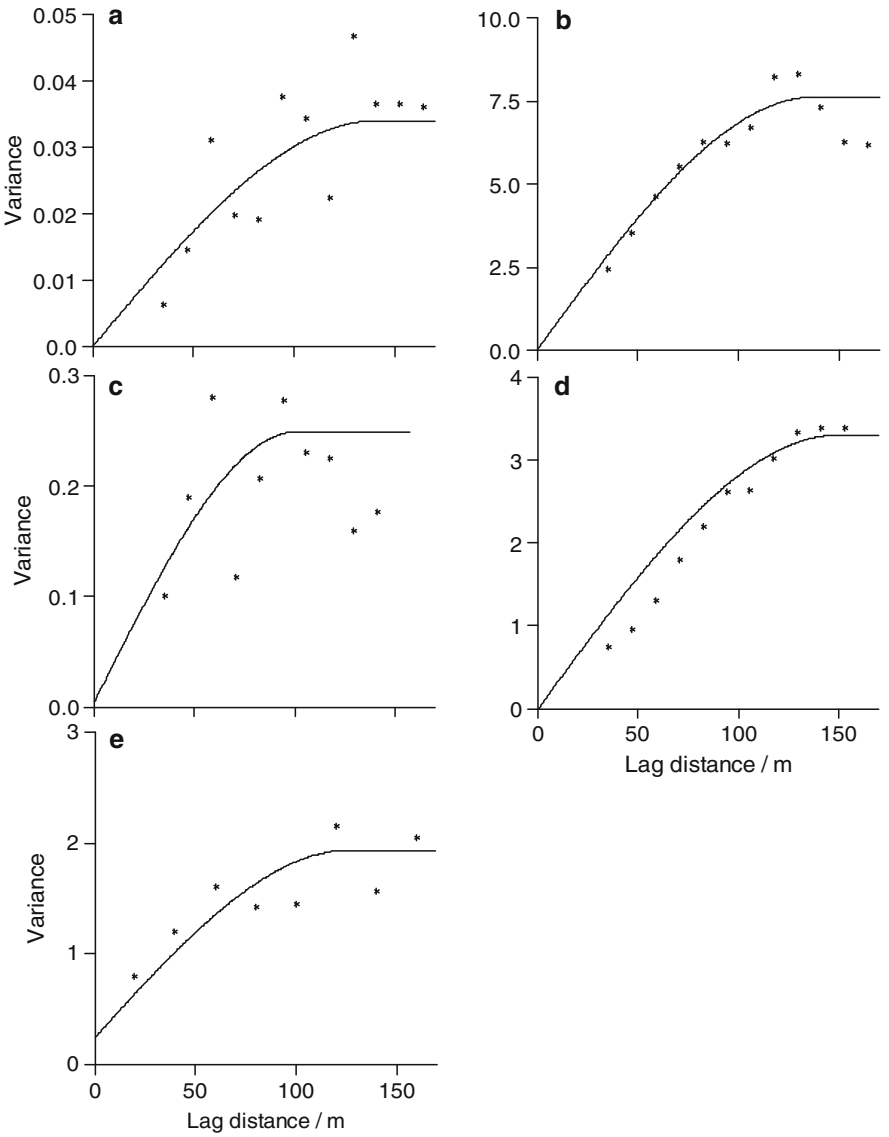

Fig. 8.5 Experimental variograms (*symbols*) and fitted spherical model (*solid line*) for regression residuals of: (**a**) organic matter, (**b**) nitrate-N (NO_3-N), (**c**) zinc, (**d**) electric conductivity and (**e**) ammonium –N (NH_4-N)

The predictions for the five soil properties were then analysed with a nonhierarchical k-means clustering algorithm for spatial data to create three management regions for each field (MSCA; S-PLUS©, Reich and Davis 1998). The algorithm groups the sites in the field into management zones using spatial attributes. In k-means clustering the grouping aims to minimize or maximize some criterion, in this

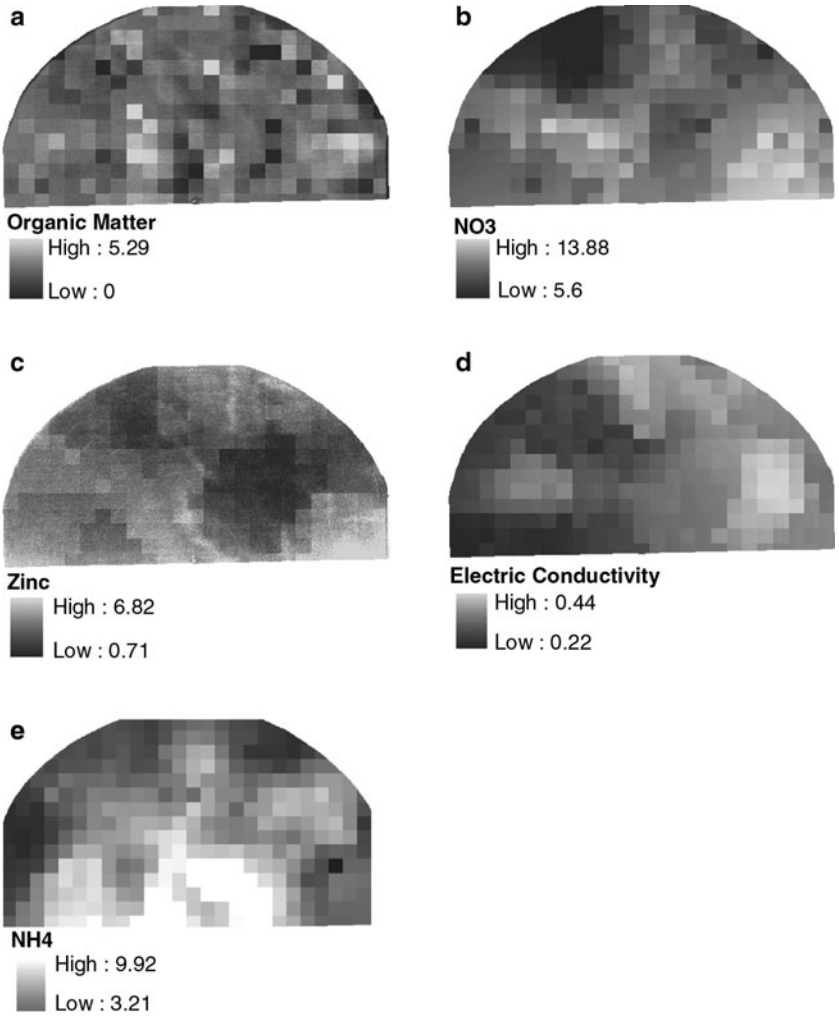

Fig. 8.6 Pixel maps of predictions from regression kriging for: (**a**) organic matter, (**b**) nitrate-N (NO_3-N), (**c**) zinc, (**d**) electric conductivity and (**e**) ammonium –N (NH_4-N)

case minimizing the sum of squares of distances between data and corresponding cluster centroid. Regions of high productivity potential were identified as regions with higher soil OM and NO_3-N levels, and the converse was the case for low productivity management zones. The resulting management zone surface map is a noisy representation of the zones, which would be difficult for a farmer to work with practically. The minimum size and shape of a zone for management is limited by the ability of the farmer and available implements to apply nutrients variably across a field. Therefore, some form of spatial filtering or contiguity constraint is necessary to reduce the fragmentation of the original classes to create smoother management

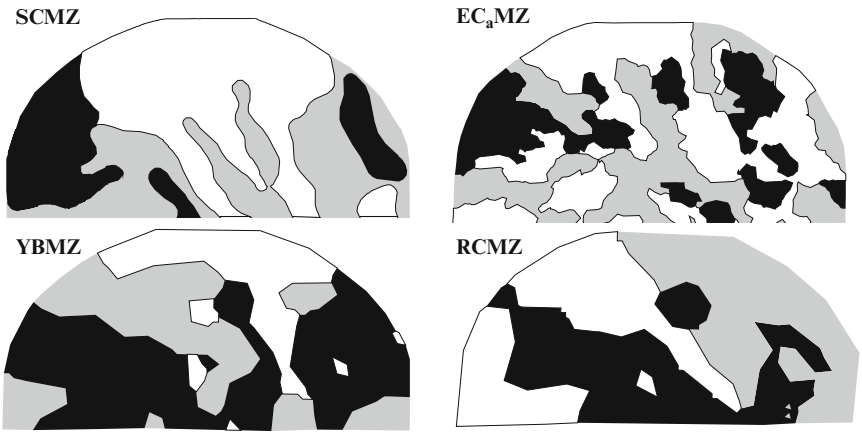

Fig. 8.7 Four techniques of management zone delineated on one centre pivot irrigated field. SCMZ, soil colour management zone technique; EC_aMZ, apparent soil electrical conductivity technique; YBMZ, yield based management zone technique; RCMZ, remotely sensed data and cluster sampling technique. Low productivity zone is *white*, medium productivity zone is *grey*, and high productivity zone is *black* (adapted from Gangloff 2004)

zones (Zhang et al. 2002; Kvien and Pocknee 2000). In this example, smoothing was accomplished by applying a focal majority function to the three initial k-means classes. The function finds the majority class value (the value that appears most often) for each location within a specified neighborhood and this become the class value at the corresponding location for the smoothed management zone map. An alternative method is to apply a contiguity constraint to the classification and this can be done using the variogram as described by Frogbrook and Oliver (2007).

Figure 8.7 shows the management zones delineated by the four techniques for one field. A visual comparison shows that there are similarities and differences in identifying areas of different productivity potential in the same field (Fig. 8.7). All four techniques delineated the north-west section of the field as a low productivity zone; similarly, high productivity zones were delineated in the south-central and western section of the field. These correspond to some extent with the areas of large and small values on the plots for the five soil variables (Fig. 8.6). Overall, the medium productivity zones show the least correspondence from a visual comparison. Quantitative assessments are needed, however, to identify the best technique for delineating management zones.

To evaluate how accurately the four management zone techniques delineate zones, a variety of statistical procedures were used (Gangloff 2004). First, an S-statistic was used to test the null hypothesis that management zones consist of a random collection of yields. The S-statistic is a median-based non-parametric, absolute deviation statistic designed to evaluate whether the within-management zone variability in yields is minimized over the field. If the zones differentiate yields into low, medium and high productivity regions, one would expect the variation in yield within management zones to be at a minimum and that the S-statistic would

be small. By contrast, if yields were assigned randomly to management zones, one would expect the variation in yield within management zones to be large, resulting in a large S-statistic. The results of the analysis showed a significant organization ($p < 0.05$) in grain yields in 30 out of 36 comparisons for the full analysis (three fields, three zones, four techniques) indicating that the yield patterns under analysis did not occur by chance and are worth classifying into low, medium and high classes.

The delineation techniques were analysed further using areal association statistics (Rees 2008), which determine how well the management zone delineation techniques worked. Each delineated management zone (potentially low, medium and high yield) was evaluated against spatially referenced yield data, which were classified into three classes (i.e. low, medium and high) by the following three approaches. The first used an objective, k-means clustering algorithm to classify each yield value as high, medium or low. The second approach used an objective, non-parametric classification procedure. Yield values were sorted and classified as 'high' if they were greater than the 3rd quartile yield value; 'medium' if they were within the 1st and 3rd quartiles; and 'low' if they were below the 1st quartile value. The third yield classification approach was subjective and based on knowledge of maize yields in the region; this was accomplished with the help of cooperating farmers. Yield values were classified as high ($>11.9\,\mathrm{Mg\ ha^{-1}}$), medium (8.8–11.9 $\mathrm{Mg\ ha^{-1}}$) or low ($<8.8\,\mathrm{Mg\ ha^{-1}}$) based on farmers' knowledge.

The classified yield value at each location was compared with the associated management zone classification, and summarized in square error matrices with classified yield in the columns and management zone in the rows, where the columns and rows are the categories in the classification. The diagonal elements of the matrix are the number of times the two data sets agree. Non-diagonal elements give the number of misclassified times by category. The sum of the diagonal cells on the matrix represents the total number of correctly classified yield observations The proportion of the total number of correctly classified yield observations in the matrix gives the overall aerial agreement for a given delineation technique. A chi-square goodness-of-fit test and Kappa statistic were used to compare the areal agreement against that which might be expected by chance. The Kappa statistic can be thought of as a chance-corrected proportional agreement that ranges from $+1$ for perfect agreement, to 0 for no agreement above that expected by chance and to -1 for complete disagreement. The areal agreement analysis was done for all approaches to yield classification, but only the results for the clustering approach that proved superior to the traditional grid-sampling approach are discussed. Cluster sampling accounted for, on the average, 34% (minimum = 1% and maximum = 74%) more of the variability in soil properties compared to grid sampling. Estimates of bias were small and estimates of the root mean squared error suggested that cluster sampling captured more of the small-scale variation in soil properties than the traditional grid sampling approach. Overall percentage agreement for the cluster sampling approach ranged from 12% to 49% (Table 8.6). All comparisons were significant based on the chi-squared goodness-of-fit test. Overall, the RCMZ technique has the greatest

Table 8.6 Areal agreement (in percent) between yield classes (high, medium or low) using a clustering approach to classify yield and four management zone delineation techniques as evaluated with the chi-square goodness-of-fit test

Management zone techniques	Low yield vs low MZ	Medium yield vs medium MZ	High yield vs high MZ	Weighted mean
	Percentage			
[a]SCMZ	36	46	32	42
[b]ECMZ	33	40	39	36
[c]YBMZ	23	40	12	22
[d]SSMZ	49	49	45	49

[a]SCMZ refers to soil colour based management zone technique.
[b]ECMZ refers to soil electrical conductivity based management zone technique.
[c]YBMZ refers to yield based management zone technique.
[d]RCMZ refers to remote sensing and soil sampling based management zone technique.

overall percentage agreement when compared with the other delineation techniques. The SCMZ technique has the second highest percentage agreements. The RCMZ was the only technique that used soil samples in the delineation process. The bare-soil imagery was used to develop a directed soil sampling procedure to capture the small-scale variation associated with soil properties. This information was used to interpolate soil properties. The ability to interpolate soil properties was the main reason the RCMZ technique was superior to other delineations techniques.

It is interesting to observe in the Gangloff (2004) study, that although the areal associations between the grain yield classes are significant (Table 8.6), the agreements in general may appear quite small (12–49%). This may be attributed to the intentional smoothing or spatial filtering that is part of the process of management zone delineation (Zhang et al. 2002). Smoothing removes the isolated islands of low, medium and high zones distributed throughout the field, thereby decreasing the areal association between management zones and yield classes.

When multi-sectional fertilizer sprayer booms with individual nozzle control become economically feasible, perhaps there will be little to no need for spatial filtering of management zones to create a smooth surface. Also, with the availability of a suite of active remote sensors that can be mounted on tractors to assess the health and vigour of the plant while 'in-motion' there is potential for further improvements in the optimization of input applications for site-specific crop management. "*Farming by the foot*" as perceived from the early precision agriculture concept may perhaps then become a reality.

8.5 Conclusions

Spatial variation in agricultural fields has been recognized worldwide. However, there have been limitations associated with quantifying within-field spatial variation in an easy, inexpensive and accurate way. Site-specific management zones have

been developed and evaluated as a successful tool that characterizes within-field spatial variation, and groups homogeneous areas of a field into small regions (called zones) so that each zone can be managed differently based upon the limiting factor of that zone. Some may argue that the concept of management zones may be perceived as a setback from the original concept of precision agriculture, i.e. "*farming by the foot*" or micro-management (Zhang et al. 2002). However, site-specific crop management across management zones has been shown to maintain or enhance crop yields, nutrient use efficiency, to be environmentally suitable and economically feasible. With the advent of new technologies and a suite of active sensors, when coupled with management zones, there is a tremendous potential to improve further the efficiency, economics and overall crop production systems. Nevertheless, management zones are probably an interim measure to fully variable-rate management. The latter will be become more economically feasible when spatially intensive soil and crop information become available cheaply from on-the-go measurement devices for geostatistical analysis and mapping.

References

Anderson-Cook, C. M., Alley, M. M., Roygard, J. F. K., Khosla, R., Noble, R., & Doolittle, J. A. (2002). Differentiating soil types using electromagnetic induction and crop yield maps. *Soil Science Society of American Journal, 66*, 1562–1570.

Agterberg, F. P. (1984). Trend surface analysis. In G. L. Gaile, & C. J. Willmott (Eds.), *Spatial statistics and models* (pp. 147–171). Dordrecht, The Netherlands: Reidel.

Bakhsh, A., Kanwar, R. S., Jaynes, D. B., Colvin, T. S., & Ahuja, L. R. (2005). Modeling precision agriculture for better crop productivity and environmental quality. *International Agricultural Engineering Journal, 14*, 235–243.

Bruulsema, T. W., Malzer, G. L., Robert, P. C., Davis, J. G., & Copeland, P. J. (1996). Spatial relationships of soil nitrogen with corn yield response to applied nitrogen. In P. C. Roberts, R. H. Rust, & W. E. Larson (Eds.), *Proceedings of Third International Conference on Precision Agriculture* (pp. 505–512). Minneapolis, MN, 23–26 June 1996. Madison, WI: ASA, CSSA, SSA.

Bullock, D. S., & Bullock, D. G. (2000). Economic optimality of input application rates in precision farming. *Precision Agriculture, 2*, 71–101.

Cassel, D. K., Kamprath, E. J., & Simmons, F.W. (1996). Nitrogen–sulfur relationships in corn as affected by landscape attributes and tillage. *Agronomy Journal, 88*, 133–140.

Ciha, A. J. (1984). Slope position and grain yield of soft-white winter wheat. *Agronomy Journal, 76*, 193–196.

Cipra, J. E., Bidwell, O. W., Whitney, D. A., & Feyerhem, A. M. (1972). Variations with distance in selected fertility measurements of pedons of Western Kansas Ustoll. *Soil Science Society America Proceedings, 36*, 111–115.

Cressie, N. (1993). *Statistics for spatial data, revised edition*. New York: Wiley.

Derby, N. E., Casey, F. X. M., & Franzen, D. W. (2007). Comparison of nitrogen management zone delineation methods for corn grain yield. *Agronomy Journal, 99*, 405–414.

Doerge, T. (1999). Defining management zones for precision farming. *Crop Insight, 8*, 21. Pioneer Hi-Bred International Inc.

Dow, A. I., & James, D. W. (1973). *Intensive soil sampling. A principle of soil fertility management in intensive irrigation agriculture*. Washington Agricultural Experiment Station Bulletin, 781. Pullman, WA: Washington State University.

Fairchild, D., & Hammond, M. (1988). Using computerized fertilizer application equipment for efficient soil fertility management. In J. S. Jacobsen (Ed.), *Proceedings of 39th Annual Regional Fertilizer Conference* (pp. 77–82). Bozeman, MT, 11–13 July 1988. Pasco, WA: Far West Fertilizer and Agrichemical Association.

Fleming, K. L., Heermann, D. F., & Westfall, D. G. (2004). Evaluating soil color with farmer input and apparent soil electrical conductivity for management zone delineation. *Agronomy Journal, 96*, 1581–1587.

Franzen, D. W., & Peck, T. R. (1995). Sampling for site-specific application. In P. C. Robert, R. H. Rust, & W.E. Larson (Eds.), *Proceedings of Second International Conference on Soil Specific Management for Agricultural Systems* (pp. 535–551). Minneapolis, MN, 27–30 March 1994. Madison, WI: ASA, CSSA, and SSSA.

Franzen D. W., Hopkins, D. H., Sweeney, M. D., Ulmer, M. K., & Halvorson, A. D. (2002). Evaluation of soil survey scale for zone development of site-specific nitrogen management. *Agronomy Journal, 94*, 381–389.

Frogbrook, Z. L., & Oliver, M. A. (2007). Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. *Journal of Soil Use and Management, 23*, 40–51.

Gangloff, W. J. (2004). *Spatial statistical analysis of soil properties and crop yield for precision agriculture applications*. PhD dissertation, Fort Collins, CO: Colorado State University.

Godwin, R. J., Wood, G. A., Taylor, J. C., Knight, S. M., & Welsh, J. P. (2003). Precision farming of cereal crops: a review of a six year experiment to develop management guidelines. *Biosystems Engineering, 84*, 375–391.

Jaynes, D. B., Kaspar, T. C., Colvin, T. S., & James, D. E. (2003). Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. *Agronomy Journal, 95*, 574–586.

Johnson, C. K., Mortensen, D. A., Wienhold, B. J., Shanahan, J. F., & Doran, J. W. (2003). Site-specific management zones based on soil electrical conductivity in a semiarid cropping system. *Agronomy Journal, 95*, 303–315.

Hammond, L. C., Pritchett, W. L., & Chew, V. (1958). Soil sampling in relation to soil heterogeneity. *Soil Science Society America Proceedings, 22*, 548–552.

Hammond, M. W. (1993). Cost analysis of variable fertility management of phosphorus and potassium for potato production in central Washington. In P. C. Robert, R. H. Rust, & W. E. Larson. (Eds.), *Proceedings of First Workshop on Soil Specific Crop Management* (pp. 213–228). Minneapolis, MN, 14–16 April 1992. Madison, WI: ASA, CSSA, and SSSA.

Hammond, M. W., Mulla, D. J., & Fairchild, D. S. (1988). Development of management maps for spatially variable soil fertility. In J. S. Jacobsen (Ed.), *Proceedings of 39th Annual Regional Fertilizer Conference* (pp. 67–76). Bozeman, MT, 11–13 July, 1988. Pasco, WA: Far West Fertilizer and Agrichemical Association.

Haradine, F. F. (1949). The variation of soil properties in relation to stage of profile development. *Soil Science Society America Proceedings, 14*, 302–311.

Heermann, D. F., Hoeting, J., Duke, H. R., Westfall, D. G., Buchleiter, G. W., Westra, P., Peairs, F. B., & Fleming, K. L. (1999). Interdisciplinary irrigation precision farming team research. In J. V. Stafford (Ed.), *Precision Agriculture'99, Vol. 1* (pp. 212–130. Sheffield, UK: Sheffield Academic Press.

Hornung, A., Khosla, R., Reich, R., Inman, D., & Westfall, D. G. (2006). Comparison of site-specific management zones: soil color based and yield based. *Agronomy Journal, 98*, 405–417.

Inman, D., Khosla, R., Lefsky, M., & Westfall, D. G. (2005). Early season grain yield prediction using site-specific management zones and remote sensing. In J. V. Stafford (Ed.), *Precision agriculture'05* (pp. 835–841). Wageningen, The Netherlands: Wageningen Academic Publishers.

Inman, D., Khosla, R., Reich, R., & Westfall, D. G. (2008). Normalized difference vegetation index and soil color-based management zones in irrigated maize. *Agronomy Journal, 100*, 60–66.

Kerry, R., & Oliver, M. A. (2007). Determining the effect of asymmetric data on the variogram. I. Underlying asymmetry. *Computers and Geosciences, 33*, 1212–1232.

Khosla, R., & Alley, M. M. (1999). Soil-specific nitrogen management on mid-atlantic coastal plain soils. *Better Crops with Plant Food, 83*, 6–7.

Khosla, R., Fleming, K., Delgado, J., Shaver, T., & Westfall, D. G. (2002). Use of site specific management zones to improve nitrogen management for precision agriculture. *Journal of Soil Water Conservation, 57*, 513–518.

Khosla, R., Inman, D., Westfall, D. G., Riech, R., Frasier, W. M., Mzuku, M., Koch, B., & Hornung, A. (2008). A synthesis of multi-disciplinary research in precision agriculture: site-specific management zones in the semi-arid western Great Plains of the USA. *Precision Agriculture, 9*, 85–100.

Kitchen, N. R., Sudduth, K. A., & Drummond, S. T. (1998). An evaluation of methods for determining site-specific management zones. *Proceedings of North Central Extension Industry Soil Fertility Conference* (pp. 133–139). Brookings, SD: Potash and Phosphate Institute.

Kitchen, N. R., Sudduth, K. A., Myers, D. B., Drummond, S. T., & Hong, S. Y. (2005). Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Computers and Electronics in Agriculture, 46*, 285–308.

Koch, B., Khosla, R., Frasier, W. M., Westfall, D. G., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal, 96*, 1572–1580.

Kravenchenko, A. N., & Bullock, D. G. (2000). Correlation of corn and soybean grain yield with topography and soil properties. Agronomy Journal, 92, 75–83.

Kvien, C., & Pocknee, S. (2000). *Introduction to management zones*. University of Georgia, USA: National Environmental Sound Production Agriculture Laboratory (NESPAL), College of Agricultural and Environmental Science.

Landwise, Inc. (2006). http://landwise.ca/Nutrient/Step2SoilSampling.htm. Accessed 17 April 2009.

Lauzon, J. D., Fallow, D. J., O'Halloran, I. P., Gregory, S. D. L., & von Bertoldi, A. P. (2005). Assessing the temporal stability of spatial patterns in crop yields using combine yield monitor data. *Canadian Journal of Soil Science, 85*, 439–451.

Leica Geosystems. (2003). *Image analysis for ArcGIS. Image analysis software for personal computers*. Atlanta, GA: Leica Geosystems.

Lund, E. D., Christy, C. D., & Drummond, P. E. (1999). Practical applications of soil electrical conductivity mapping. In J. V. Stafford (Ed.), *Precision Agriculture '09* (pp. 771–779). Sheffield, UK: Sheffield Academic Press.

McBratney, A. B., & Pringle, M. J. (1999). Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture, 1*, 212–236.

Mercer, W. B., & Hall, A. D. (1911). The experimental error of field trials. *Journal of Agricultural Science, 4*, 107–132.

Moore, S. H., & Wolcott, M. C. (2000). *Using yield maps to create management zones in field crops*, Vol. 43, 12–13. Baton Rouge: Louisiana Agricultural Experiment Station.

Mueller, T. G., Pierce, F. J., Schabengerger, O., & Warncke, D. D. (2001). Map quality for site-specific fertility management. *Soil Science Society America Journal, 65*, 1547–1558.

Mulla, D. J., Bhatti, A. U., Hammond, M. W., & Benson, J. A. (1992). A comparison of winter wheat yield and quality under uniform versus spatially variable fertilizer management. *Agriculture Ecosystems & Environment, 38*, 301–311.

Mulla, D. J., & Hammond, M. W. (1988). Mapping of soil test results from large irrigation circles. In J. S. Jacobsen (Ed.), *Proceedings of 39th Annual Regional Fertilizer Conference* (pp. 169–176). Bozeman, MT, 11–13 July 1988. Pasco, WA: Far West Fertilizer and Agrichemical Association.

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., & MacDonald, L. (2005). Spatial variation of measured soil properties across site specific management zones. *Soil Science Society America Journal, 69*, 1572–1579.

Nanna, T., & Franzen, D. W. (2003). A weighted classified method for nitrogen zone delineation. *Proceedings of North Central Extension-Industry Soil Fertility Conference* (pp. 177–184). Des Moines, IA, 19–20 November 2004. Brookings, SD: Potash & Phosphate Institute.

Nolan, S. C., Goddard, T. W., & Lohstraeter, G. (2000). Assessing management units on rolling topography. In P.C. Roberts et al. (Ed.), *CD-ROM Proceedings of Fifth International Conference on Precision Agriculture*. Bloomington, MN, 16–19 July 2000. Madison, WI: ASA, CSSA, and SSA.

O'Connell, M., & Wolfinger, R. (1997). Spatial regression models, response surfaces and process optimization. *Journal of Computational and Graphical Statistics, 6*, 224–241.

Pachepsky, Y., & Acock, B. (1998). Stochastic imaging of soil parameters to assess variation and uncertainty of crop yield estimates. *Geoderma, 85*, 213–229.

Peterson, R. G., & Calvin, L. D. (1986). Sampling. In A. Klute (Ed.), *Methods of soil analysis. Part I. Physical and mineralogical properties, including statistics of measurement and sampling* (pp. 33–51). Madison, WI: American Society of Agronomy.

Rees, W. G. (2008). Comparing the spatial content of thematic maps. *International Journal of Remote Sensing, 29*, 3833–3844.

Rehm, G. W., Mallarino, A., Reid, K., Franzen, D., & Lamb, J. (2001). *Soil sampling for variable rate fertilizer and lime application*. St. Paul, MN: University of Minnesota Extension Publication, SB-7647

Reich, R. M., & Davis, R. A. (1998). *On-line spatial library for the S-PLUS statistical software package*. Fort Collins, CO: Colorado State University.

Stafford, J. V., Lark, R. M., & Bolam, H. C. (1998). Using yield maps to regionalize fields into potential management units. In P. C. Roberts, R. H. Rust, & W. E. Larson. (Eds.), *Proceedings of Fourth International Conference on Precision Agriculture* (pp. 225). St. Paul, MN, 19–22 July 1998. Madison, WI: ASA, CSSA, SSA.

Thom, W. O., Schwab, G. J., Murdock, L. W., & Sikora, F. J. (2003). Taking soil test samples. Lexington, KY: *University of Kentucky Cooperative Extension Service Publication*, AGR-16.

Verity, G. E., & Anderson, D. W. (1990). Soil erosion effects on soil quality and yield. *Canadian Journal Soil Science, 70*, 471–484.

Vieira, S. R., Neilsen, D. R., & Biggar, J. W. (1982). Spatial variation of field-measured infiltration rate. *Soil Science Society America Journal, 45*, 1040–1048.

Waynick, D. D., & Sharp, L. T. (1919). Variation in soils and its significance to past and future soil investigations. II. Variations in nitrogen and carbon in field soils and their relation to the accuracy of field trials. *University of California Publications in Agricultural Sciences* (pp. 121–139). Vol 4, No. 5.

Welsh, J. P., Wood, G. A., Godwin, R. J., Taylor, J. C., Earl, R., Blackmore, B. S., & Knight, S. M. (2003a). Developing strategies for spatially variable nitrogen application in cereals, Part I: Winter barley. *Biosystems Engineering, 84*, 481–494.

Welsh, J. P., Wood, G. A., Godwin, R. J., Taylor, J. C., Earl, R., Blackmore, B. S., & Knight, S. M. (2003b). Developing strategies for spatially variable nitrogen application in cereals, Part II: Wheat. *Biosystems Engineering, 84*, 495–511.

Webster, R., & Oliver, M. A. (1992). Sample adequately to estimate variograms of soil properties. *Journal of Soil Science, 43*, 177–192.

Yost, R. S., Uehara, G., & Fox, R. L. (1982). Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. *Soil Science Society America Journal, 46*, 1028–1032.

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture – a worldwide overview. *Computers and Electronics in Agriculture, 36*, 113–132.

Chapter 9
Weeds, Worms and Geostatistics

R. Webster

Abstract Weeds and plant-parasitic nematodes occur in patches in agricultural fields. Farmers can control them with chemicals. They can do so precisely and prevent competition (from weeds) and predation (by nematodes) provided they know where the pests are early in the lives of their crops or before sowing or planting them. Standard geostatistical methods have been used successfully to analyse counts of both weed seedlings and nematodes in the soil and to map their distributions from kriged estimates. The application is technologically sound. The most serious obstacle to its application in farming is that sampling must be intense, with spacings between sampling points of 20–40 m. This means that the cost of sampling and counting the pests is greater than the savings from not applying herbicides or nematicides. Only for potatoes is the effort and cost of estimating the burdens of the parasitic cyst nematodes of the genus *Globodera* justified economically. For precise control of weeds proximal sensing at the seedling stage seems more promising.

Keywords Pest control · Weeds · Nematodes · Potatoes · Soya beans · Cereal crops · Geostatistics · Nested sampling · Analysis of variance

9.1 Introduction

Infestations of weeds, agricultural pests and diseases, which collectively I shall call 'pests', vary in intensity within fields and paddocks. In many fields their occurrences are patchy, varying from none in some patches to numbers that seriously diminish crop yields in others. In these circumstances farmers want to kill the offenders with sprays etc. where they occur and not to treat land where they are absent. For weeds they may also wish to vary the amount or type of herbicide according to the degree or kind of infestation. Their desires clearly lie within the realm of precision agriculture.

The 1990s brought new opportunities for precise application of agricultural chemicals. They saw the advent of global positioning systems (GPS) with local

R. Webster (✉)
Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, Great Britain
e-mail: richard.webster@bbsrc.ac.uk

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_9,

accuracies of 1–2 m on the farm and the development of sprayers capable of varying the amounts pesticide applied 'on the run'. These together appeared to enable farmers to control pests selectively where they occurred. The 1990s were also a decade during which society became increasingly concerned about the impact of excesses of pesticide, fungicide and herbicide in the environment. Society wanted farmers to use no more than necessary of these chemicals and not to use them where there were no pests, diseases or weeds or too few to be of consequence.

To take advantage of the opportunities and to satisfy society's concerns farmers had to know where the pests were and how many; they needed maps of infestations. So quantitative mapping became a third strand for success in this branch of precision agriculture. As it happened, agronomists 'discovered' geostatistics also in the 1990s and saw in it the means by which they could interpolate their sample counts and scores of infestation so as to create maps.

Weed scientists had been making maps visually before that. Many weeds, once grown, are readily distinguished from the crop. By that stage, however, they can already have retarded the growth of the crop, and the need is to map them early in their growth so that they can be treated before they have any serious effect. Automated detection of weeds is being investigated, and Gerhards and co-workers in Germany in particular (see Gerhards and Oebel 2006) report some success, though, in the view of Lutman and Miller (2007), their techniques are not yet practicable on the farm. The only effective way of mapping for pre-emptive local control is to count or otherwise estimate density at the seedling stage, or to count seeds in the soil, in sample quadrats and to interpolate from those counts. Geostatistics would play its part, and weed agronomists explored its suitability for the purpose.

Plant-parasitic nematodes are equally significant agricultural pests. Like weeds, their distributions tend to be patchy. Unlike weeds, their presence becomes obvious only after their damage has been done, either in sickly crop plants or poor yields. Control patch by patch therefore depends on farmers' knowing where the nematodes are before crops are sown. The chemicals for killing nematodes are expensive and unpleasant, and this adds to farmers' desires to treat the land only where the pests are prevalent. Commercial companies have responded by offering packages whereby they estimate nematode densities by sampling, typically, in hectare blocks within fields and then treating (or not) the fields with nematicide block by block. Several nematologists in collaboration with geostatisticians have investigated the distributions of nematodes more rigorously, and for potato crops we can now see the economic opportunities for more precise control.

Other agricultural pests, notably fungi, occur in patches. There are few accounts of local control following any form of geostatistical analysis, however.

9.2 Weeds

By the mid-1990s several papers on the use of geostatistics for characterizing and mapping the spatial distribution of weeds had appeared. Donald (1994), for example, mapped the density of shoots and root growth of the thistle (*Cirsium arvense*),

a perennial weed, in Dakota. The area was small, 12 × 12 m, and was sampled at 1.8-m intervals. Johnson et al. (1996) applied the standard technology for mapping the weeds *Helianthus annuus* and *Abutilon theophrasti* in maize and soya beans in Nebraska. The counts were strongly skewed, but the authors did not transform them.

Cardina et al. (1995) explored the technology more thoroughly. They took two approximately 25 × 90 m areas in two adjacent soya-bean fields in Ohio and sampled them on grids of 3 × 7 m (in 1990) and 3 × 6 m (in 1993) and counted seedlings of *Chenopodium album*. They transformed their counts to logarithms to remove skewness and stabilize the variances. Then they recognized trends, which they removed before doing geostatistical analyses on the residuals. Finally, they added the trends to their kriged estimates of the residuals and back-transformed the logarithms to the original scales for their final maps.

Trend is often overlooked, and it will be well to consider it now. The basic model underlying the usual geostatistical analysis is

$$Z(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}). \tag{9.1}$$

Here $Z(\mathbf{x})$ is a random variable at a place $\mathbf{x}$, μ is the mean and $\varepsilon(\mathbf{x})$ is a random component drawn from a distribution with mean zero and variogram

$$\begin{aligned}\gamma(\mathbf{h}) &= \frac{1}{2}\mathrm{E}\left[\{\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}+\mathbf{h})\}^2\right] \\ &= \frac{1}{2}\mathrm{E}\left[\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\}^2\right] \quad \text{for all } \mathbf{h},\end{aligned} \tag{9.2}$$

where $\mathbf{h}$ is the lag separating pairs of points. The model holds only if the mean is constant, at least locally, i.e. if the process is intrinsically stationary. If there is trend, i.e. if the mean changes systematically from place to place, then we have a somewhat more complex model such as

$$Z(\mathbf{x}) = m(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{9.3}$$

in which a trend term $m(\mathbf{x})$ has replaced the mean μ. The equivalence in Eq. 9.2 no longer holds, but we still need the variogram as defined there for analysis. There are several ways of dealing with the problem, the most recent and satisfactory of which is to estimate the variogram and the trend simultaneously by residual maximum likelihood (REML) – see Webster and Oliver (2007). A feasible alternative in the 1990s was to remove the trend first and to compute and model the variogram of the residuals for later use in kriging. The most popular method was to fit a global trend surface, usually a low-degree polynomial, compute and model the variogram of the residuals, krige the residuals with the fitted model and then add the trend surface back to the kriged residuals. The resulting estimates were unbiased but usually with underestimated variances.

Cardina et al. removed the trends from the logarithms of their counts in a different way, namely by median polish. This technique, developed by Cressie (1993) for gridded data, proceeds as follows. The medians of the values in each and every row

and column of the grid are found. Denote these by $\tilde{z}_{i.}$ for the ith column and $\tilde{z}_{.j}$ for the jth row, and denote the general median of all the data by $\tilde{z}_{..}$. Then compute

$$y_{ij} = z_{ij} - \tilde{z}_{.j} - \tilde{z}_{i.} + \tilde{z}_{..} \quad \text{for all } i \text{ and } j. \tag{9.4}$$

The resulting y_{ij} are effectively residuals, and one can compute and model their variograms and krige from them by standard technique.

Cardina et al. computed experimental variograms of their residuals by the method of moments and fitted isotropic spherical models to them. The spherical model is

$$\begin{aligned} \gamma(h) &= c_0 + c\left\{\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right\} \quad \text{for } 0 < h \le a \\ &= c_0 + c \quad \text{for } h > a \\ &= 0 \quad \text{for } h = 0. \end{aligned} \tag{9.5}$$

Here c_0 is the nugget variance, c is the sill of the correlated variance, and a is a distance parameter, the range. Figure 9.1, redrawn from Fig. 3 of Cardina et al. (1995), shows variograms for 1990 and 1993. The black discs are the experimental values for the residuals of the natural logarithms of the counts, and the solid lines are the fitted models. Table 9.1 lists the parameter values. The authors computed and modelled the variograms of the logarithms of the counts with the trends still present, and I have retained their results in Fig. 9.1 for completeness. As we should expect, the variances are larger and so too are the fitted ranges. As above, the authors used the models to estimate statistical surfaces of the residuals by ordinary kriging, added the trends and then back-transformed the results. Figure 9.2 shows their final estimated weed surfaces for the 2 years. The weed densities are exceptionally large at the back left-hand corners as they appear in the figure, and it is easy to understand that the original counts would be both strongly skewed and contain trend.

There were similar studies in Europe. One of the more thorough ones was that of Heisel et al. (1996) on the weeds *Lamium* spp. and *Veronica* spp. in an arable field in Denmark. They sampled at the intersections of a square grid at intervals of 10 m and counted seedlings in 0.25-m^2 quadrats. The frequency distributions were strongly skewed, and the authors took logarithms to stabilize the variances. They then did a standard geostatistical analysis on the logarithms of the counts, and after kriging on to a fine grid they back-transformed the kriged predictions to counts per 0.25 m^2 and mapped the results.

The authors mention a feature that is common with counts: their records for some species contained many zeros, and there was no way that they could transform those distributions to approximate normality. They proposed an indicator approach for modelling such data with a threshold at, for example, 4 weed plants per 0.25 m^2, and mapping the resulting indicator to identify patches that should be sprayed with herbicide. They did not pursue the idea; but it is worth pursuing, and I return to the idea below for nematode populations.

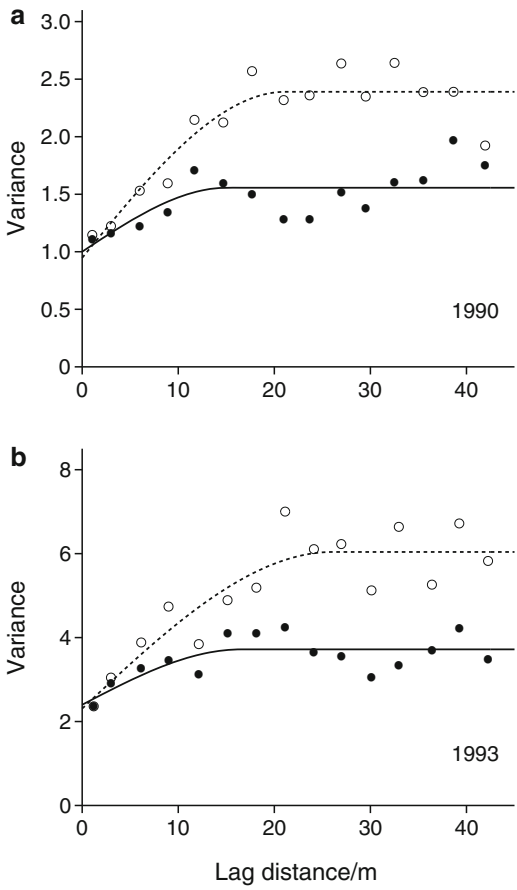

Fig. 9.1 Variograms of the natural logarithms of counts of *Chenopodium album* in a soya-bean field in Ohio. The *circles* and *dashed lines* are the experimental values and fitted spherical models with trend present; the *black discs* and *solid lines* are the corresponding values and models after the removal of trend by median polish. The parameter values are listed in Table 9.1. The figure is redrawn from Figure 3 of Cardina et al. (1995) with the permission of the authors and the Weed Science Society of America, the publishers of *Weed Science*

Table 9.1 Parameter values of the spherical models fitted to the variograms shown in Fig. 9.1. The symbols are c_0 the nugget variance, c the sill of the correlated variance and a the range, see Eq. 9.5

	ln(counts)			Residuals		
Year	c_0	c	a/m	c_0	c	a/m
1990	0.947	1.442	21.2	1.001	0.559	15.1
1993	2.309	3.731	26.1	2.400	1.320	16.6

Heisel et al. (1996) realized that sampling at 10-m intervals was time-consuming and expensive. So to see whether they could economize they removed 5/6 of the sampling points and analysed the counts for the remaining 38 points on what then became a 30 × 20 m grid. As we might expect (see Chapter 2), the resulting variograms were erratic. The authors mistrusted the models that they fitted, and they judged their kriged maps to be poor representations of reality. Sampling would

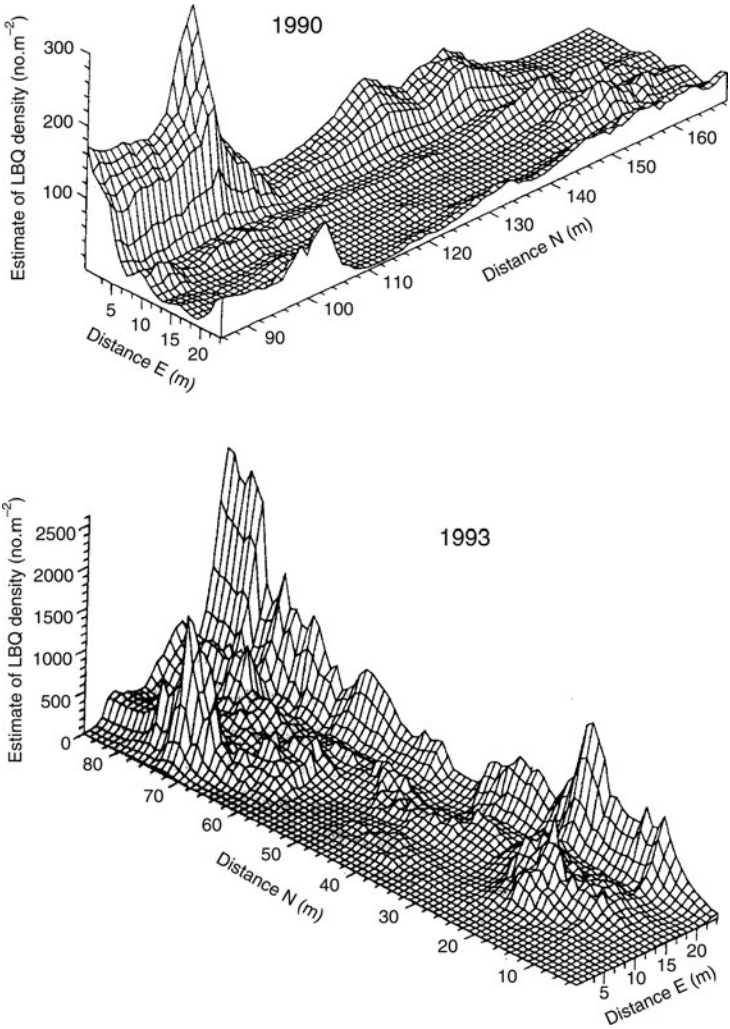

Fig. 9.2 Statistical surfaces of *Chenopodium album* in two soya-bean fields in Ohio made by kriging. The *vertical scales* are in counts m^{-2}, with means of 28 and 400 plants m^{-2} in 1990 and 1993 respectively. Reproduced from Cardina et al. (1995) in *Weed Science*, volume 43, with permission of the authors and the Weed Science Society of America.

have to be at approximately their original intensity to produce reliable maps – unless they could take advantage of covariates. Somewhat later, Cousens et al. (2002) found that even a 10-m grid was too coarse for reliably mapping weed seedlings in a wheat field in Victoria, Australia.

In a second paper Heisel et al. (1999) explored the potential of cokriging with soil variables as subsidiary predictors. They found that the abundance of *Lamium* spp. was correlated with the soil's silt content ($r = 0.30$–0.38) and that this was strong

enough for them to use silt as a covariate and to cokrige *Lamium* with a reduction of 11–15% in the prediction variance. I find that result somewhat surprising with such weak correlation, but perhaps the structural correlation was stronger; the authors do not say. Despite their mention of modelling the coregionalization, the authors appear not to have fitted such a model, and their back-transformation from logarithms to the original scale seems wrong. We should treat their results with caution, therefore. The same idea of using the correlation between soil properties and weed densities to help predict the latter was pursued by Walter et al. (2002), also in Denmark. These authors sampled two arable fields each of approximately 10 ha on 20-m grids in three consecutive years, 1993, 1994 and 1995. They found some cross correlations between some species and the soil's pH and phosphorus content in some years and also between different years for some species. They appear not to have attempted to fit a linear model of coregionalization, however. They concluded that the correlations were neither strong enough nor sufficiently consistent for them to cokrige reliably to advantage.

By the year 2000 it was clear that weed infestations could be sampled and analysed by standard geostatistical methods to produce maps at scales suitable for patch spraying. The question then became no longer whether it was possible, but whether it was economic: was the cost of sampling, counting or otherwise measuring the degree of infestation offset by the saving in herbicide?

Jurado-Expósito et al. (2003) explored one aspect of cost, namely that of the herbicide. They sampled sunflower fields in Spain on grids of 7×7 m and counted the weeds within quadrats at the grid nodes. From the data they kriged and mapped infestations that exceeded the economic threshold for spraying. In one of their fields this would have led to a reduction of 61% in herbicide. They make two points in their final section. One is that if weeds are persistent then mapping in 1 year might be used in subsequent years to guide control, so that the cost of mapping would be discounted over several years. The second is that a cheap method of weed detection is crucial for site-specific weed control 'because conventional weed sampling in a grid ... is very time consuming and expensive'. The authors give no figures for the cost of the sampling, but clearly they imply that it would be more than the savings on herbicide.

The theme of economy was pursued by Rew et al. (2001) with grid surveys of weeds in arable land on four farms in New South Wales. They sampled areas of various sizes with spacings ranging from 1×1 m to 10×10 m, giving sample sizes ranging from 395 to 841 quadrats according the areas covered. They counted the weeds in each quadrat, and they recorded the time spent counting. The authors split each set of their data into a 'test set' and a 'real set'. They analysed their counts in the test sets, transformed to logarithms where desirable, by standard geostatistical methods to produce maps of weed infestations. They compared their kriged predictions with the true values in their real sets of data.

As we might expect, the smaller were the proportions of sites retained in the test sets, and hence the sparser sampling, the poorer were the variograms estimated and the worse were the predictions. The authors could not say in general what sampling intensity might be satisfactory, however, because the weed densities and the

correlation ranges were unique to each field. Also, kriging underestimated the greatest densities and overestimated the smallest. Such effects are well-known attributes of the technique, but they could have important consequences for weed control. Patches free of weeds were mapped as infested, albeit sparsely, so that farmers might spray them unnecessarily; in contrast, farmers might apply too little herbicide to the most infested patches.

The most serious criticism of the technology levelled by the authors concerns the cost of sampling. The actual costs of sampling and counting varied from field to field, depending on the sampling density and the numbers of weeds. The authors give a figure of Aus$1200 for sampling an area somewhat larger than 1 ha on a 25-m grid, assuming a cost in labour of Aus$25 per hour; that was in 2000. If you multiply this figure by the area of a field, say in the range 10–50 ha, you see that the cost could be anything from Aus$12 000 to Aus$60 000 per field. This far exceeds any possible saving in the cost of herbicide.

9.3 Nematodes

The plant-parasitic nematodes, of which there are many species, are serious agricultural pests. They are responsible for losses to agriculture of the order US$$100 \times 10^9$$ per year worldwide. The distributions of nematodes in the soil have for long been studied by statisticians. Emphasis was originally on marginal frequencies and indices of aggregation and their implications for sampling and estimation – see, for example, Anscombe (1950) and Seinhorst (1982). Taylor's power law (Taylor 1984) has been much invoked to describe aggregated distributions. In the 1990s nematologists and statisticians turned their attention to geostatistics both to describe spatial distributions and for mapping, and more recently with a view to local control of nematodes by patch treatment within fields.

9.3.1 Lives of Nematodes

It might help readers to understand the precise control of plant-parasitic nematodes in agriculture if I first summarize the life forms and cycles of the most damaging genera.

The juvenile forms of these pests are small and wormlike; they are commonly known as 'eel worms'. They are typically 0.2–1.0 mm long, and they live freely in the rhizosphere of the soil. They burrow into the soft tissues of plant roots where they then live as parasites, thereby damaging and debilitating their hosts. They also create lesions through which fungal pathogens can invade and further debilitate the host plants. The females of several genera lay their eggs in the roots, and these eggs then find themselves back in the rhizosphere when the roots die. The females of the root-knot nematodes of the genus *Meloidogyne* lay their eggs into a gelatinous matrix which forms canals through the outer root tissue, and so the eggs pass into

the soil. Juveniles hatch from the eggs and they reinvade the roots of the host plants. Another major group of plant-parasitic nematodes comprise the cyst nematodes, e.g. *Heterodera* spp. and *Globodera* spp. They behave somewhat differently in that in due course the mature females, after living and feeding within the roots expand and break through to the roots' outer surfaces. There they swell into spherical cysts, typically 0.3–0.5 mm in diameter but sometimes larger, containing eggs, which are destined to become the next generation. When the roots die the cysts, still containing the eggs, remain in the soil. If a new host crop is sown the eggs within the cysts respond to substances exuded by the hosts and hatch to start the cycle afresh. Otherwise, in the absence of a new host crop the cysts and their eggs eventually die.

Some nematodes that parasitize crop plants have alternative hosts. One such species is the cereal cyst nematode *Heterodera avenae*. It lives not only on cereals but so too on many native grasses, and these help to maintain its population even in the absence of a cereal crop. As it happens, there are also fungi in the soil that prey on the nematodes and can keep their numbers acceptably small. In contrast, the cyst nematode *Heterodera glycines* which parasitizes soya beans in the USA and South America has no natural predators there. It can reduce yields by as much as 75% and causes huge financial losses to farmers. Chemical control is possible, but the costs in relation to the benefits are prohibitive. However, because the nematode has no alternative host in the Americas farmers can control it by rotating soya beans with other crops.

Potatoes are parasitized by two other species of cyst nematode that were introduced with them into Europe from South America. These nematodes are *Globodera rostochiensis* and *G. pallida*. They are both persistent and cause substantial losses of yield. In Europe they have no alternative host crops, though they can infect tomatoes and aubergines. So, as with the soya-bean cyst nematode, farmers can eradicate them simply by not growing potatoes for a long time or by growing resistant varieties that trigger the eggs to hatch but on which the nematodes cannot subsequently feed. For potatoes, however, the value of the crop is such as to justify chemical control with nematicide, especially if farmers know where the nematodes are.

9.3.2 Geostatistical Applications

As with weeds, the first investigators wanted to see whether geostatistics could be applied to nematode infestation at all. They already knew that nematodes were aggregated at some scales, and, as above, they could see some of the effects as patchy debilitation of crops. They sampled the soil, counted the nematodes, computed and modelled variograms, and kriged to map the infestations. Wallace et al. (1993) and Wallace and Hawkins (1994) did this for several plant-parasitic species in a field under reed canary-grass in the USA. They sampled the topsoil at 10-m intervals along transects by taking cores and extracting the nematodes from 116 cm^3 of soil. Six of the seven species they counted had bounded variograms with correlation ranges between 40 and 160 m. In like vein, though at a much finer scale, Rossi et al. (1996) analysed four species of parasitic nematodes in a sugar cane crop. They

sampled the soil at 20-cm intervals between pairs of rows of the crop. Their mean counts were in the range 53–316 per 100 cm^3 of soil. From their counts they computed variograms, but all were unbounded within the dimensions of their small plot, and their kriged maps revealed strong trends with maxima under the rows.

Boag and I had similar ideas (Webster and Boag 1992). Unlike Wallace et al. and Rossi et al., however, we did not presume to know how big or how far apart any patches might be; i.e. we did not wish to guess the spatial scale of variation of any infestation. We wanted, therefore, to determine that scale first, and we adapted the nested design and analysis of variance (ANOVA) devised by Youden and Mehlich (1937) for the purpose (see Section 2.2.1 also).

The design is basically a hierarchy at the top of which is the whole region of interest; in precision agriculture it is likely to be a particular field or part of a field. The region is divided into segments; these form the first stage or level in the hierarchy and in each is placed initially one sampling point. In a second stage one new sampling point is chosen in each segment at a fixed distance and random direction from the first point. In a third stage the process may be repeated for each point chosen in the second stage, and so on for as many stages as desired. The resulting design is balanced in the sense that all branches of the hierarchy are the same below any given level. As the number of stages is increased so the number of sampling points doubles at each stage. To avoid this and the concomitant expense one can economize by replicating only a proportion of the sampling nodes in the lower branches of the hierarchy. The result is an unbalanced design. Whether the design is balanced or not observations are then made at every point in the scheme, and they can be analysed by ANOVA.

The model for analysis with s stages is

$$Z_{ijk\ldots m} = \mu + A_i + B_{ij} + C_{ijk} + \cdots + \varepsilon_{ijk\ldots m}. \tag{9.6}$$

Here $Z_{ijk\ldots m}$ is the value of the mth unit in $\ldots$, in the kth class at stage 3, in the jth class at stage 2, and in the ith class at stage 1. The quantity μ is the general mean; A_i is the difference between μ and the mean of class i in the first stage, B_{ij} is the difference between the mean of the jth subclass within class i and the mean of class i, and so on. The residual $\varepsilon_{ijk\ldots m}$ represents the difference between the observed value from its class mean at the last stage of subdivision. The effects $A_i,\ B_{ij},\ C_{ijk}, \ldots, \varepsilon_{ijk\ldots m}$ are assumed to be independent random variables with means of zero and variances $\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2, \ldots, \sigma_s^2$. The latter are components of variance, and when estimated by ANOVA they describe the variances attributable to the distances, $d_s,\ d_{s-1},\ d_{s-2}, \ldots, d_1$, in the design. Further, when accumulated from the lowest stage upwards they form a variogram, albeit somewhat crude, thus:

$$\begin{aligned} \widehat{\sigma}_s^2 &= \widehat{\gamma}(d_s) \\ \widehat{\sigma}_{s-1}^2 + \widehat{\sigma}_s^2 &= \widehat{\gamma}(d_{s-1}) \\ \widehat{\sigma}_{s-2}^2 + \widehat{\sigma}_{s-1}^2 + \widehat{\sigma}_s^2 &= \widehat{\gamma}(d_{s-2}), \end{aligned} \tag{9.7}$$

and so on. By plotting the $\widehat{\gamma}(d_s)$, $\widehat{\gamma}(d_{s-1}), \ldots, \widehat{\gamma}(d_1)$ against the corresponding distances one can see within which range of distance most of the variance lies. Webster et al. (2006) and Webster and Oliver (2007) describe the analysis and its interpretation fully.

9.3.3 Case Study

Boag and I (Webster and Boag 1992) designed a sampling scheme along the above lines to discover at what spatial scales the nematode *Heterodera avenae* varied in cereal fields in eastern Scotland. Our design was an unbalanced one with seven stages spanning distances from 70 mm to 50 m with intermediate spacings of 0.2, 0.6, 1.8, 5.6 and 16.7 m. We sampled five fields with this design, and at each sampling point we took a core of topsoil (about 200 g) and counted the cysts of *H. avenae* in it. By a hierarchical ANOVA of our counts we estimated the components of variance, which we then accumulated to form variograms. Table 9.2 summarizes the statistics, and Fig. 9.3 shows their accumulated components of variance. Note the geometric progression of the sampling intervals and that the scale for distance on the abscissa is logarithmic.

One finding from the study was that there is no common range of distance within which most of the variance lies. Each field is unique, and the spatial scale of the nematode burden in it must be estimated roughly before one can sample with confidence to estimate the variogram accurately and to krige.

The field at Invergowrie, field E, was the one that most obviously lent itself to such analysis at a practicable scale, and we sampled a portion of it to explore the feasibility of mapping. The portion was a square of side 100 m, which we sampled at 7.14-m intervals to give 225 cyst counts. Table 9.3 summarizes the data, and Fig. 9.4a shows a conventional variogram for the counts with the sample estimates shown by the black discs and an isotropic spherical model, Eq. 9.5, fitted to them by the solid curve. The parameter values are $c_0 = 32.6$ counts^2, $c = 36.5$ counts^2 and $a = 29.7$ m, and they are listed in Table 9.4. The distribution of the counts is strongly skewed with many zeros. Transformation of the counts to common logarithms (with the addition of 0.1) gives a more symmetrical distribution, summarized

Table 9.2 Summary statistics of counts of cysts of *Heterodera avenae* per 200 g of topsoil of five Scottish fields each sampled with 108 cores with a nested design

	Field				
	A	B	C	D	E
Mean	43.4	23.7	15.9	28.9	17.4
Median	42	22	14	26	12
Standard deviation	20.3	12.7	8.6	18.9	20.5
Variance	412.02	161.12	74.35	358.11	420.47
Skewness	0.6	0.5	0.8	1.5	2.3

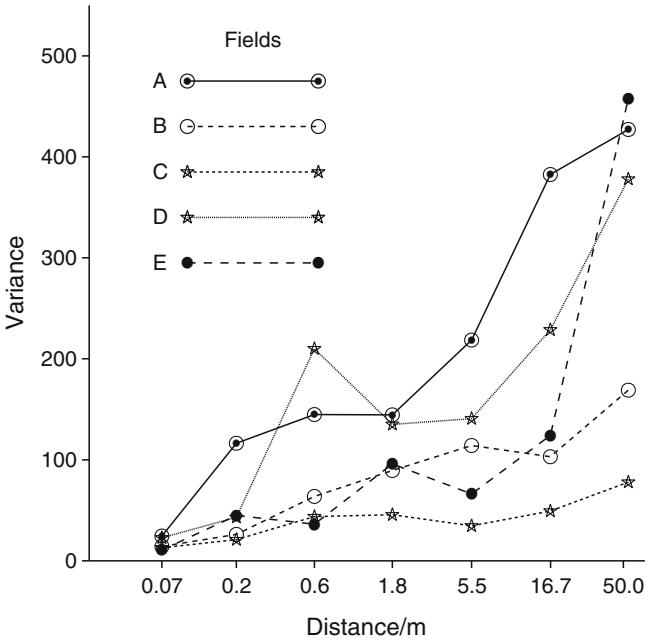

Fig. 9.3 Variograms of counts of the cysts of the nematode *Heterodera avenae* per 200 g soil in five Scottish arable fields

Table 9.3 Summary statistics of counts of cysts of *Heterodera avenae* per 200 g soil and their logarithms for the 1-ha plot at Invergowrie (in field E in Table 9.1) sampled with 225 cores on a square grid

	Count	Log_{10}count
Mean	4.15	−0.213
Median	0	−1
Standard deviation	8.01	0.910
Variance	64.89	0.8280
Skewness	2.6	0.5

in Table 9.3, and the variogram in Fig. 9.4(b). The model fitted is again spherical with parameter values $c_0 = 0.309$, $c = 0.601$ and $a = 37.5$ m (Table 9.4). Note that the proportion of nugget variance appears diminished by the transformation (0.31 instead of 0.47) and that the estimated range is greater. Figure 9.5 is a map of infestation made from the data and the model by log-normal punctual kriging. It shows two largish patches of infested soil and one small one. The more extensive area is almost free of the nematodes.

The above part of the investigation is a standard application of geostatistics. Maps such as that in Fig. 9.5 are subject to error, and farmers might reasonably want to know the probabilities that there are nematodes in the soil and whether they

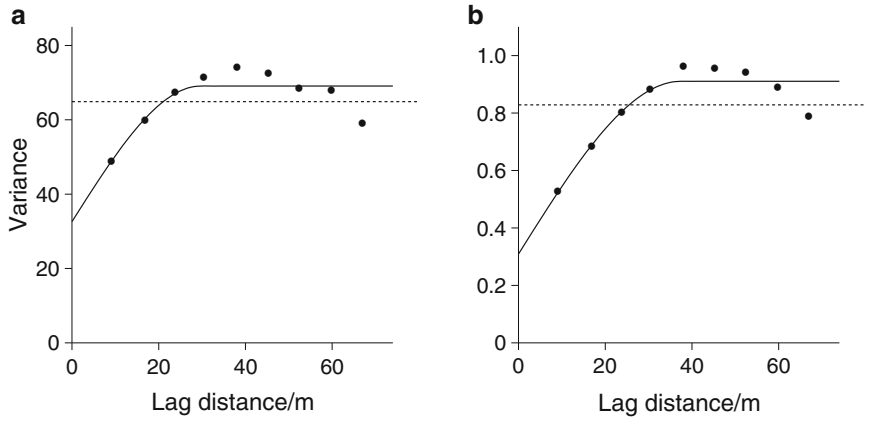

Fig. 9.4 Variograms of the cysts of *Heterodera avenae* in 1 ha of an arable field at Invergowrie, Scotland: (**a**) on the original counts per 200 g soil, and (**b**) on the their common logarithms. The points are the sample estimates and the curves are of the best fitting spherical models. The *horizontal lines* are the sample variances

Table 9.4 Parameter values, c_0, c and a, of spherical models, Eq. 9.5, fitted to variograms of counts of cysts of *Heterodera avenae*, their logarithms and indicators for the 1-ha plot at Invergowrie (in field E in Table 9.1)

			Indicator	
	Count	Log_{10}count	$I(\mathbf{h}) \geq 1$	$I(\mathbf{h}) \geq 4$
c_0	32.6	0.309	0.142	0.063
c	36.5	0.601	0.127	0.140
a/m	29.7	37.5	43.3	29.8

are present in sufficient numbers as to damage the crops. For continuous variables such as the soil's pH the probabilities are best estimated by disjunctive kriging (see Section 1.2.3.4). The method involves transforming the data to approximate a normal (Gaussian) distribution or some other known distribution. This is not feasible with count data with many zeros. One way that is feasible is by indicator kriging.

An indicator is a variable that takes values of 1 and 0 only. The values may indicate presence and absence respectively. In geostatistics indicators are often created by the division of a continuous scale or of a scale of counts at one or more specified thresholds. If the original variable is assumed to be random then so is the indicator variable. If therefore we define a threshold, z_c, for the random variable $Z(\mathbf{x})$ then we create the indicator variable $I(\mathbf{x})$ from it as

$$\begin{aligned} I(\mathbf{x}) &= 0 \quad \text{if } Z(\mathbf{x}) < z_c \\ &= 1 \quad \text{otherwise.} \end{aligned} \tag{9.8}$$

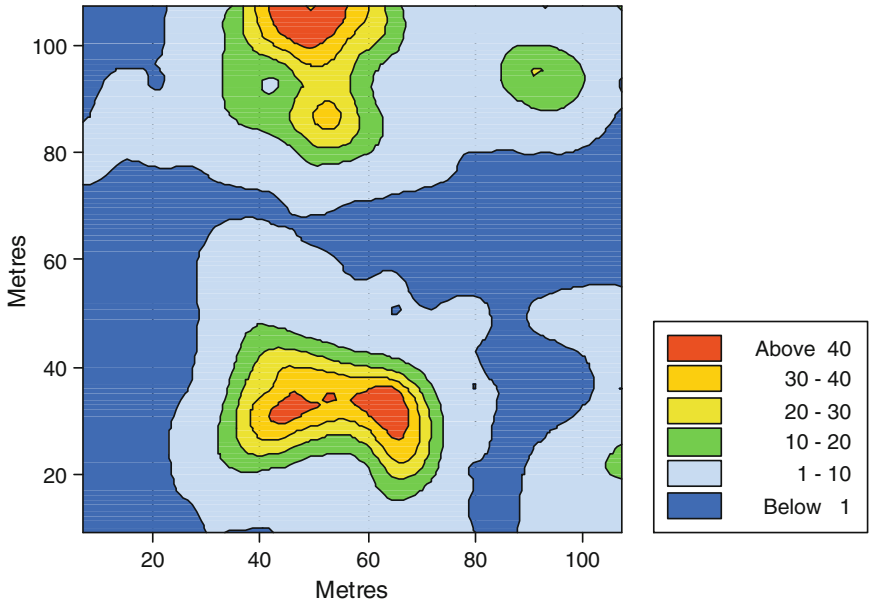

Fig. 9.5 Map of estimated numbers of the cysts of *Heterodera avenae* per 200 g soil in 1 ha of an arable field at Invergowrie, Scotland

Variograms of indicators may be computed and modelled in the same way as are those of continuous variables, and they may be used with the indicator data to krige. The kriged results are effectively probabilities that the values of z at the target sites exceed or not the defined thresholds.

The subject is a large one, and is covered in detail by Goovaerts (1997). Here I take the study in field E at Invergowrie just one step further to show some of what is possible.

Figure 9.6 displays the indicator variograms for *H. avenae* with (a) $I(\mathbf{x}) \geq 1$ and (b) $I(\mathbf{x}) \geq 4$. As before, the black discs are the sample values, $\widehat{\gamma}(\mathbf{h})$, and the curves are the fitted models, which are again spherical. Table 9.4 lists the parameter values. Figure 9.7a and b are the corresponding isarithmic ('contour') maps of the punctually kriged estimates. Figure 9.7a shows rather little of the area where the probability of the soil's being free of the nematode is small (less than 0.2). If we increase the threshold to four cysts per core then approximately half the area has a probability less than 0.2, Fig. 9.7b.

9.3.4 Economics

By the end of the 1990s it had become clear that nematode infestations could be analysed geostatistically and mapped by kriging. However, it is expensive to sample the soil, to separate the nematodes and to count them, and this cost must

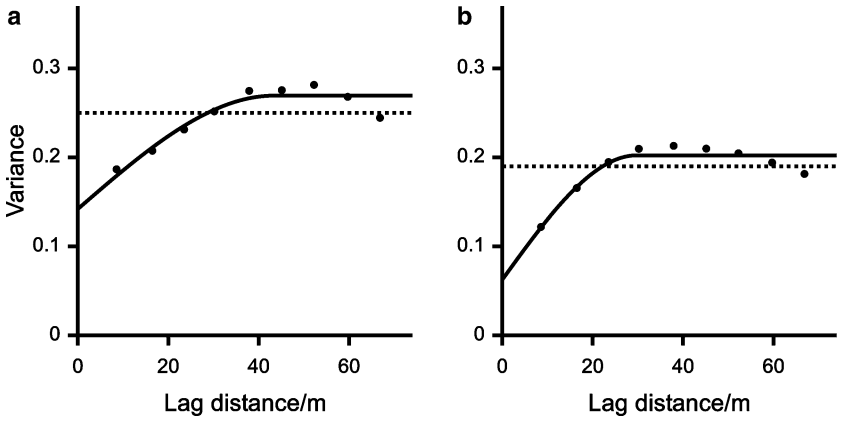

Fig. 9.6 Indicator variograms of the cysts of *Heterodera avenae* in 1 ha of an arable field at Invergowrie, Scotland: (**a**) with the threshold set at one cyst per 200 g soil core and (**b**) at four cysts per 200 g. The points are the sample estimates and the curves are of the best fitting spherical models. The horizontal lines are the sample variances

be set against the savings that can accrue from patch treatment with the knowledge gained. So the question then became, as in weed agronomy: was the procedure cost-effective?

Farias et al. (2002) attempted an answer for cotton growing in Brazil where damage by the nematode *Rotylenchulus reniformis* in the 1990s was a principal cause of the 50% reduction in cotton production. They took a field 48 × 32 m and sampled its topsoil, both before germination of the cotton and at harvest, on a grid at intervals of 6 m in one direction and 4 m in the orthogonal direction to give 64 sampling points. Their variograms of the nematode counts seemed bounded, and they fitted spherical models. In addition to mapping the infestations by ordinary punctual kriging they did 100 conditional simulations to estimate the probabilities that the damage threshold was exceeded and so identify patches that should be treated with nematicide. They drew attention to the large nugget variances, which led to imprecise estimates and simulations and the need for some more closely spaced sampling to estimate the variogram better at short lag distances. They could not complete a risk–benefit analysis because they lacked the information on prices. Nevertheless, their study of risk by simulation was clearly a step in the right direction.

Wyse-Pester et al. (2002) and Avendaño et al. (2004) recognized that the cost of sampling and counting could militate against mapping nematode infestations for site-specific management. They sought therefore to economize by taking advantage of correlations between nematode counts and more stable properties of the soil. Wyse-Pester et al. considered the correlations with particle-size fractions and organic matter. They sampled the topsoil of two large fields under maize, one of 71 ha and the other 53 ha, at intervals of 76 m and counted the numbers of *Helicotylenchus* spp., *Tylenchorhynchus capitatus* and *Pratylenchus neglectus*. They recorded strong spatial dependence in the nematode counts and could fit spherical models with small nugget variances and with ranges between 140 and 630 m.

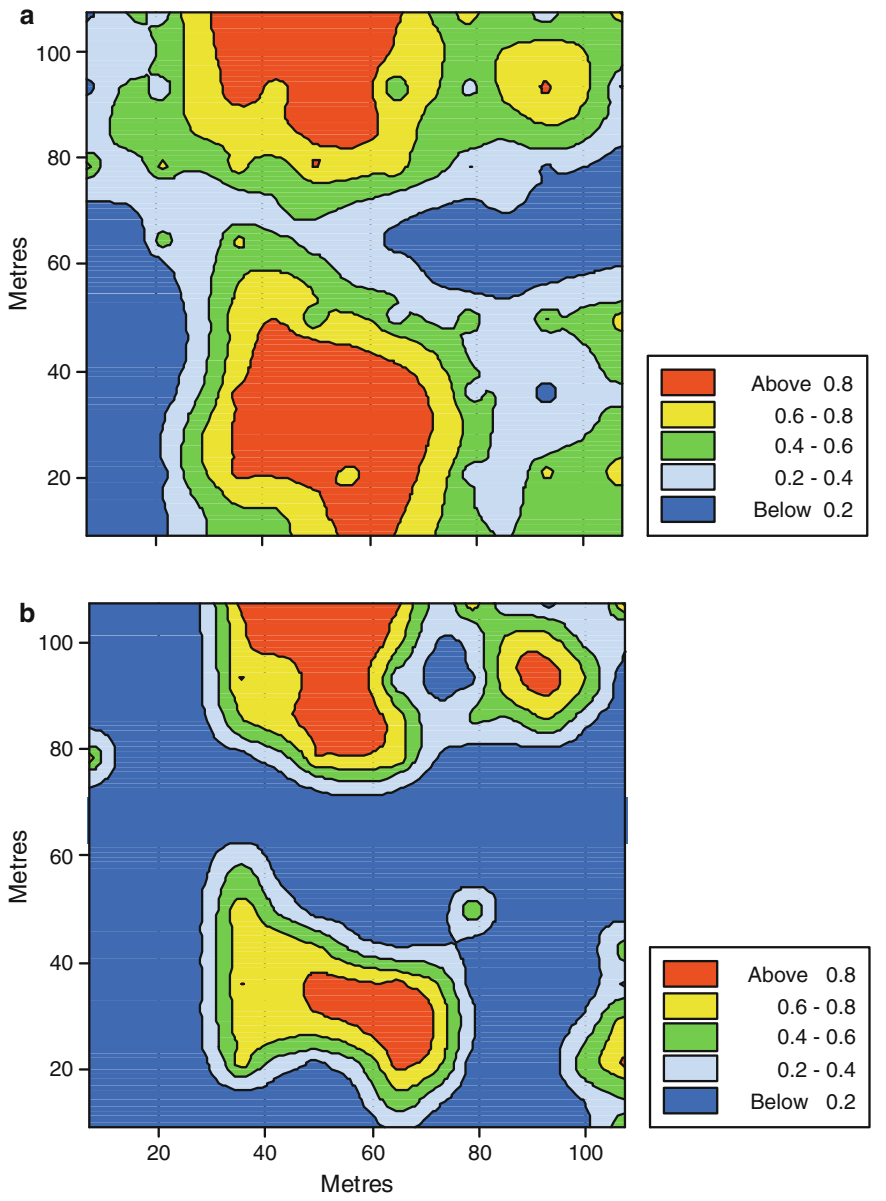

Fig. 9.7 Maps of estimated probabilities that the soil contains (**a**) at least one cyst of *Heterodera avenae* per 200 g soil and (**b**) at least four cysts per 200 g in 1 ha of an arable field at Invergowrie, Scotland

However, the authors judged the correlations between the counts and the proportions of sand, silt, clay and organic matter too weak for there to be any economic advantage in substituting them for nematode counts. Avendaño et al. (2004) investigated

infestations of *Heterodera glycines* in several fields growing soya beans. They found significant linear correlations between nematode counts and the soil's pH, with coefficients, r, mainly between 0.2 and 0.5, and thought that farmers might divide their fields according to pH and general fertility and manage each division separately. Neither group of investigators seems to have considered using the subsidiary soil variables for cokriging or kriging with external drift – perhaps fortunately in view of the only weak to moderate correlations.

As mentioned above, potatoes are parasitized by two species of cyst nematode, *Globodera pallida* and *G. rostochiensis*. These nematodes have no alternative hosts in Europe, and farmers can control infestations simply by not growing the crop a second time on the same land until the nematode population has declined to some non-damaging size. That can take 10 or more years, and farmers are unwilling to wait that long. Potato growing is highly specialized and requires expensive equipment for planting, cultivation and harvesting, and it can be very profitable. Farmers therefore want to grow the crop as frequently as they can on the land they have available to get the most out of their capital investment, and they are prepared to treat the soil with nematicide to do so. The chemicals are expensive, and by applying them only where necessary, i.e. only where the nematode burden exceeds some threshold, potato farmers can save themselves a lot of money. The only question then is whether that saving is more than the cost of sampling and counting the nematodes to make maps.

Evans et al. (2002, 2003) studied this aspect of precise control of nematodes. They monitored infestations of potato cyst nematodes (PCN) on several commercial farms in eastern England by sampling the topsoil and counting the nematode eggs both before and after the fields had grown potatoes. The fields, varying in size from 6 to 10 ha, were sampled on square grids at 20-m intervals. The resulting variograms were similar in form to those in Fig. 9.4, and the fitted spherical models had correlation ranges of 40–60 m. The authors could thus expect infestations to occur in patches roughly 40–60 m across, and this was confirmed in the maps made by ordinary punctual kriging.

In all instances except one the numbers of PCN eggs increased enormously as a result of potato growing without nematicide, and they appeared to have done so from isolated patches identified by the pre-cropping sampling. (The exception was in a field where there seemed to be also the nematophagous fungus *Verticillium chlamydosporium* = *Pochonia chlamydosporia* acting as a natural control.) One example from Evans (2003), summarized in Table 9.5, illustrates what can happen.

Table 9.5 Summary of counts (numbers of eggs g^{-1} soil) of potato cyst nematode (*Globodera* spp.) at Ram Farm before potatoes were planted and immediately after harvest, from Evans et al. (2003)

	Before planting	After harvest
Minimum	0	0
Maximum	160.2	495.6
Mean	8.4	65.6
Median	2.5	45.3
Variance	522.3	5042.4
Standard deviation	22.8	71.0
Skewness	4.9	2.1

Table 9.6 Proportions of a 10-ha field that a farmer would have to leave untreated with nematicide to justify sampling at various intensities. Costs, at 1999 prices, are £360 ha^{-1} for granular nematicide, £560 ha^{-1} for fumigant and £6 to take and process each sample of soil (from Evans et al. 2002)

			Proportion of land left untreated		
Sample spacing/m	Number of samples	Cost of sampling/£	Granular nematicide	Fumigant	Granular + fumigant
20	250	1500	0.42	0.27	0.16
40	63	380	0.11	0.068	0.041
60	28	170	0.047	0.030	0.018
80	15	90	0.025	0.016	0.010
100	10	60	0.017	0.011	0.007

The median egg count in the field increased from 2.5 per gram of soil before planting to 45.3 at harvest, and the maximum increased from 160.2 g^{-1} to 495.6. The rate at which the nematode eggs die is about 30% per year, and so the farmer could expect to wait 14 years before the count has declined to less than five eggs per gram, the critical threshold in the tactical management for a single crop. Had the farmers treated the patches with nematicide they would almost certainly have prevented the increases in nematodes and loss of yield from those patches.

All the maps of pre-cropping infestations showed large areas where there were fewer than five eggs per gram of soil. If the intensity is less than this threshold then the farmer should not need to apply nematicide; if it is more then he should. Using this information and knowing the cost of sampling and nematicide, Evans et al. (2002) calculated the savings that were likely to accrue from the information gained. They are summarized in Table 9.6 in terms of the proportion of a 10-ha field sampled on grids of various intervals. The table shows that if a farmer applied both granular and fumigant nematicide he could justify sampling at 20-m intervals if only 16% of a field has fewer nematodes than the threshold, and even if he were to apply only the granular nematicide he would save money if 42% did not require treatment. The costs listed are those pertaining in 1999, and although prices have changed since then, as have also the chemicals permitted, the costs of sampling and chemical treatment relative to the value of the crop are likely to be much the same.

The table appears to show that much larger savings are possible with sparser sampling. Evans et al. (2002) warn against such an inference. They successively removed samples from the full set of data for one 8-ha field to leave counts at the nodes of 40-, 60-, 80- and 100-m grids and mapped from them. As one might expect for a correlation range of 40–60 m, important detail was lost. At an interval of 40 m the essentials of the pattern were still evident; but the authors pointed out that small patches of infestation would be likely to go undetected, and the nematodes in those if untreated would significantly diminish crop yield. They concluded that, given the potential economies, sampling should be at a 20-m interval.

9.4 The Future for Geostatistics in Precise Pest Control

As numerous investigators have shown, the infestations of both weeds and nematodes are spatially correlated at scales varying from that between the rows of crops, of the order of 1 m, to 10s of metres. Sample counts can be analysed geostatistically and converted to maps of infestations by kriging, and the maps can be used for the treatment with pesticides patch by patch. The basic geostatistical technology works for counts of these pests.

We need to be cautious, however. We need to beware of looking into a mirror and seeing ourselves, seeing what we want to see, as A.R. Ammons warns in his short poem, *Reflective*. You can read this in full at http://poemfortoday.wordpress.com/20/01/27/reflective-by-a-r-ammons/. We must recognize that the biggest stumbling block to the application is not statistical but economic, largely because of the cost of sampling. For weed control the sampling costs far exceed possible savings in herbicides and gains from increased yields in treated patches. Now agronomists, having realized this, have turned their attention to other methods such as comprehensive proximal sensing of weeds. Most recently Guillot et al. (2009) investigated the combination of direct counting of weeds and estimates from digital images obtained with a camera held at breast height above the ground. They fitted a linear model of coregionalization to the two sets of data and cokriged to estimate weed density. Their errors were smaller than estimates made from counts alone, but the method is still expensive in that one needs counts and images from at least 100 quadrats in a field to estimate the variograms plus additional images to improve on autokriging.

With nematodes the situation is less clear cut. For cereals and soya beans, for example, the costs of sampling the soil, separating the nematodes or their cysts and counting them outweigh any savings from patch treatment with nematicides. For potatoes, however, a crop of much greater value, the costs can be more than matched by savings on expensive nematicides.

There are in addition technical difficulties that most investigators have tended to ignore. One is drift or trend. In many instances ordinary kriging is sufficiently robust that trend can be ignored for a first mapping. But where trend dominates the spatial distribution, as in the example of the nematodes between the rows of sugar cane, it should be taken into account, nowadays by residual maximum likelihood (REML) techniques (Webster and Oliver 2007).

Geostatistics works best for normally distributed variables. As investigators have discovered, counts of weeds and nematodes are strongly positively skewed in many instances. Some authors have transformed their counts to logarithms to stabilize variances and analysed the logarithms geostatistically before finally back-transforming their estimates to the original scales. Others have not.

One problem with counts has not been adequately explored and solved; it is that caused by large numbers of zeros. A first line of attack on the problem is to use indicator kriging with a threshold set at some count, say z_c, at less than which damage is unlikely; i.e. set the indicator, $i(\mathbf{x})$, to 0 if the count, $z(\mathbf{x})$, is less than z_c, and

to 1 otherwise, as above. Heisel et al. (1996) suggested this for weeds, and I show above how it might apply to nematode infestation. On the practical side, farmers do not want to treat uninfested patches with expensive and noxious chemicals. So any method of mapping should honour the absence of pests, and it is here that indicator and disjunctive kriging could find application.

References

Anscombe, F. J. (1950). Soil sampling for potato root eel worm cysts. *Annals of Applied Biology*, *37*, 286–295.

Avendaño, F., Pierce, F. J., & Melakeberhan, H. (2004). Spatial analysis of soybean yield in relation to soil texture, soil fertility and soybean cyst nematode. *Nematology*, *6*, 527–545.

Cardina, J., Sparrow, D. H., & McCoy, E. L. (1995). Analysis of spatial distribution of common lambsquarters (*Chenopodium album*) in no-till soybean (*Glycine max*). *Weed Science*, *43*, 258–268.

Cousens, R. D., Brown, R. W., McBratney, A. B., Whelan, B., & Moerkerk, M. (2002). Sampling strategy is important for producing weed maps: A case study using kriging. *Weed Science*, *50*, 542–546.

Cressie, N. A. C. (1993). *Statistics for spatial data*, revised edition. New York: Wiley.

Donald, W. W. (1994). Geostatistics for mapping weeds, with a Canada thistle (*Cirsium arvense*) patch as a case study. *Weed Science*, *42*, 648–657.

Evans, K., Webster, R. M., Halford, P. D., Barker, A. D., & Russell, M. D. (2002). Site-specific management of nematodes – pitfalls and practicalities. *Journal of Nematology*, *34*, 194–199.

Evans, K., Webster, R., Barker, A., Halford, P, Russell, M., Stafford, J., & Griffin, S. (2003). Mapping infestations of potato cyst nematodes and the potential for spatially varying application of nematicides. *Precision Agriculture*, *4*, 149–162.

Farias, P. R. S., Sánchez-Vila, X., Barbosa, J. C., Vieira, S. R., Ferraz, L. C. C. B., & Solís-Delfin, J. (2002). Using geostatistical analysis to evaluate the presence of *Rotylenchulus reniformis* in cotton crops in Brazil: economic implications. *Journal of Nematology*, *34*, 232–238.

Gerhards, R., & Oebel, H. (2006). Practical experiences with a system for site specific weed control in arable crops using real-time image analysis and GPS-controlled patch spraying. *Weed Research*, *43*, 385–392.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Guillot, G., Lorén, N., & Rudemo, M. (2009). Spatial prediction of weed intensities from exact count data and image-based estimates. *Applied Statistics*, *58*, 525–542.

Heisel, T., Andreasen, C., & Ersbøll, A. E. (1996). Annual weed distributions can be mapped with kriging. *Weed Research*, *36*, 325–337.

Heisel, T., Ersbøll, A. E., & Andreasen, C. (1999). Weed mapping with co-kriging using soil properties. *Precision Agriculture*, *1*, 39–52.

Johnson, G. A., Mortensen, D. A., & Gotway, C. A. (1996). Spatial and temporal analysis of weed seedling populations using geostatistics. *Weed Science*, *44*, 704–710.

Jurado-Expósito, M., López-Granados, F., García-Torres, L., García-Ferrer, A., Sánchez de la Orden, M., & Atenciano, S. (2003). Multi-species weed spatial variability and site-specific management maps in cultivated sunflower. *Weed Science*, *51*, 319–328.

Lutman, P. J. W., & Miller, P. C. H. (2007). *Spatially variable herbicide application technology: Opportunities for herbicide minimisation and protection of beneficial weeds*. Research Review No 62. London: Home-Grown Cereals Authority.

Rew, L. J., Whelan, B., & McBratney, A. B. (2001). Does kriging predict weed distributions accurately enough for site-specific weed control? *Weed Research*, *41*, 245–263.

Rossi, J.-P., Delaville, L., & Quénéhervé, P. (1996). Microspatial structure of a plant-parasitic nematode community in a sugarcane field in Martinique. *Applied Soil Ecology*, *3*, 17–26.

Seinhorst, J. W. (1982). The distribution of cysts of *Globodera rostochiensis* in small plots and the resulting sampling errors. *Nematologica*, *23*, 285–297.

Taylor, L. R. (1984). Assessing and interpreting the spatial distribution of insect populations. *Annual Review of Entomology*, *29*, 321–357.

Wallace, M. K., & Hawkins, D. M. (1994). Applications of geostatistics in plant nematology. *Journal of Nematology*, *26*, 626–634.

Wallace, M. K., Rust, R. H., Hawkins, D. M., & MacDonald, D. H. (1993). Correlation of edaphic factors with plant-parasitic nematode population densities in a forage field. *Journal of Nematology*, *25*, 642–653.

Walter, A. M., Christensen, S., & Simmelsgaard, S. E. (2002). Spatial correlation between weed species densities and soil properties. *Weed Research*, *42*, 26–38.

Webster, R., & Boag, B. (1992). A geostatistical analysis of cyst nematodes in soil. *Journal of Soil Science*, *43*, 583–595.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester: Wiley.

Webster, R., Welham, S. J., Potts, J. M., & Oliver, M. A. (2006). Estimating the spatial scales of regionalized variable by nested sampling, hierarchical analysis of variance and residual maximum likelihood. *Computers & Geosciences*, *32*, 1320–1333.

Wyse-Pester, D. Y., Wiles, L. J., & Westra, P. (2002). The potential for mapping nematode distributions for site-specific management. *Journal of Nematology*, *34*, 80–87.

Youden, W. J., & Mehlich, A. (1937). Selection of efficient methods for soil sampling. *Contributions of the Boyce Thompson Institute for Plant Research*, *9*, 59–70.

Chapter 10
The Analysis of Spatial Experiments

M.J. Pringle, T.F.A. Bishop, R.M. Lark, B.M. Whelan and A.B. McBratney

Abstract Anyone with an interest in precision agriculture has already formed a hypothesis that the field is a sub-optimum management unit for cropping. The role of experimentation is to test this hypothesis. Geostatistics can play an important role in analysing experiments for site-specific crop management: put simply, spatial autocorrelation must be accounted for if one is to draw valid inferences. We provide here some background to the basic concepts of agronomic experimentation. We then consider two broad classes of experimental design for precision agriculture (management-class experiments and local-response experiments), and show, with the aid of case studies, how each may be analysed geostatistically. Ultimately though, if farmers are compelled to use relatively simple designs and less formal analyses, then researchers must follow and adapt their geostatistical analyses accordingly.

Keywords Management-class experiments · Local-response experiments · Polynomial response function · Residual maximum likelihood (REML) · Cokriging · Trend comparison

M.J. Pringle (✉)
Remote Sensing Centre, Department of Environment and Resource Management, QCCCA Building, 80 Meiers Road, Indooroopilly, QLD 4068, Australia
e-mail: matthew.pringle@derm.qld.gov.au

T.F.A. Bishop
Australian Centre for Precision Agriculture, Faculty of Agriculture, The University of Sydney, 1 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015, Australia
e-mail: thomas.bishop@sydney.edu.au

R.M. Lark
Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom
e-mail: murray.lark@bbsrc.ac.uk

B.M. Whelan and A.B. McBratney
Australian Centre for Precision Agriculture, Faculty of Agriculture, The University of Sydney, John Woolley Building, Sydney, NSW 2006, Australia
e-mail: brett.whelan@sydney.edu.au; Alex.McBratney@sydney.edu.au

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_10,

10.1 Introduction

Geostatistics can play an important role in the analysis of precision agriculture-based agronomic experiments. It is commonly accepted that the principal role of geostatistics in precision agriculture (PA) is to create maps of the different kinds of data observed at different locations and supports (e.g. the results of soil testing, the output of a yield monitor and so on). In addition to its mapping applications, geostatistics enables the analysis of agronomic experiments in a spatially explicit context. A spatial analysis should, as a matter of course, accompany a spatial experiment.

But why should practitioners of precision agriculture be concerned with experiments anyway? The influence of soil conditions on plant growth is well-known (Russell 1976). It is also well-known that crop and soil attributes vary spatially (e.g. Mercer and Hall 1911; Beckett and Webster 1971). Other environmental attributes that affect the growth of a crop, such as topography and pest infestations, also vary spatially. The impact of spatial variation in agriculture has been exacerbated by a long-term trend towards larger fields due to increased mechanization. Consequently, it seems arbitrary for conventional agronomy to treat *the field* as an indivisible unit for uniform management. An apparently logical improvement on uniform management might be to sample soil and plants at various locations within a field, and then manage crop nutrition spatially on the basis of laboratory recommendations. Unfortunately this approach has its own inherent problem: the response functions used by a laboratory will be spatially generalized, and possibly irrelevant to the locations where the samples were collected (Cook and Bramley 2000). Even if the response function is correct, how does one assess the benefit of site-specific nutrient management compared with uniform management? The only way to address these issues is to conduct some form of experiment in the field of interest. Anyone with an interest in precision agriculture has already formed a hypothesis that *the field* is a sub-optimum management unit. The role of experimentation is to test this hypothesis. Lark and Wheeler (2003) noted that it is a plausible hypothesis, which, if correct, may lead to economic and environmental benefits.

It would be unrealistic to hope that, by reading this chapter, farmers and their consultants will grasp the intricacies of experimental inference and geostatistics; however, we can use this opportunity to demonstrate some practical applications of geostatistics to experimentation, and the consequent insights it provided. In this chapter we give some background to the basic concepts of agronomic experimentation. We then consider two broad classes of experimental design for PA (management-class experiments and local-response experiments), and demonstrate with the aid of case studies how each could be analysed geostatistically. Our focus is primarily on the geostatistical analysis of experiments that have been conducted with the aid of variable-rate technology and a yield monitor; some alternative approaches are discussed at the end of the chapter.

10.2 Background

An experimenter will, inevitably, view his or her observations and ask, "How much of this overall variation can be attributed to the treatments I applied?" In an agronomic context, an obvious treatment, due to its cost to the farmer, is the amount of fertilizer. Let us say we have applied different amounts of fertilizer to various plots scattered about a field. The yield of a crop is affected by a multitude of different factors, just one of which is the applied fertilizer. The simplest way to separate the effects of the treatment from the remaining variation is the decomposition:

$$z_{i,j} = \bar{z} + \delta_j + \varepsilon_{i,j}, \tag{10.1}$$

where $z_{i,j}$ is the yield of the ith plot that has been subjected to the jth level (of t levels) of fertilizer, $\bar{z}$ is the mean yield for all plots, δ_j is the change in yield attributed to the jth level (the *treatment effect*) and $\varepsilon_{i,j}$ is an error term that represents the combined effect on yield of all other causal factors. The experimental design associated with the above model is referred to as a *completely randomized design*. For much of the twentieth century experimenters used the techniques pioneered by R.A. Fisher (1890–1962), which we now know as design-based statistics, to estimate δ_j. Central to design-based estimation of δ_j is the analysis of variance (ANOVA), performed on the results of an experiment whose design has appropriate replication, randomization and in some cases blocking. Replication is required in any experiment to ensure confidence that we can estimate the effect of a particular treatment, not confounded with other sources of variation. For example, in an experiment where we simply divided a field in half and applied a treatment to one half and a control to the other, we could not be confident that the difference between them reflected real treatment effects and not just inherent differences between the two halves of the field. Randomization justifies our assumption that our observations for each treatment constitute a set of random variables, the mean of which is the treatment effect (plus any other fixed effects such as $\bar{z}$); thus the treatment effect can be estimated from the data without bias. Furthermore, by appropriate randomization we ensure that errors can be treated as independent random variables, which is a necessary assumption for design-based estimation and inference about the treatment effects. In a completely randomized design all inherent variation in the field contributes to the error term. It is against this background variation that treatment effects must be detected. If the error is large the power of the experiment to detect a treatment effect is diminished. Blocking is used to address this issue. When a field is divided into blocks, treatment effects are estimated from within-block comparisons only. As a result, between-block variation is removed from the error term. In effect, blocking alters Eq. 10.1 with a further decomposition:

$$z_{i,j,k} = \bar{z} + \delta_j + b_k + \varepsilon_{i,j,k}, \tag{10.2}$$

where b_k is the component of variation associated with the kth block. The error term now corresponds to within-block variation only. The experimental design

associated with Eq. 10.2 is referred to as a *randomized block design*. The between-block component comprises the spatial variation of the experimental domain. The ANOVA estimate of δ_j will be the same as in Eq. 10.1 but, because $\varepsilon_{i,j,k}$ is smaller than $\varepsilon_{i,j}$, we have a better chance of detecting a statistically significant response.

For the best part of a century government departments and private companies have analysed agronomic experiments with design-based statistics, and have extended the results to local farmers. Design-based statistics arose in part because of contemporary constraints on the gathering and processing of information. In Fisher's time, agronomic experiments were necessarily small, partly because mechanization was a luxury and partly because what we now consider a 'one-click' analysis might have then meant hours of laborious calculations. In the field, specialists in white coats attempted to regulate the experimental conditions as much as possible. In the office, design-based statistics, underpinned by a strong theoretical foundation, offered a tool to interpret and extrapolate from small samples. In general, these methods of experimental analysis have been tremendously successful and have become the convention; however, they have two disadvantages with regard to the extrapolation of results. First, the methods require that we assume additivity of treatment effects and block differences. Under this assumption the optimum application of an input does not vary spatially; in other words the PA hypothesis is excluded a priori. Second, the design-based approach requires complete control over the experimental design – in the form of independent, random treatment allocation – which might not always be possible.

With time, our ability to gather and process agronomic information has grown enormously. This has brought two important developments: (i) on-farm experimentation and (ii) model-based statistics.

On-farm experimentation actively involves the farmer and the farmer's own fields. The general goal of on-farm experimentation is to compare the results of a new practice (i.e. something that offers the potential for improvement) with the farmer's current method, and to consider the economic and social consequences of a change (Petersen 1994). The benefits of on-farm experimentation have been recognized for over 30 years, but for a long time the concept was limited by data collection and computing issues. In the mid-1990s, soon after the first yield maps appeared, researchers realized that yield monitors and variable-rate input applicators could be combined with experimental techniques. Reetz (1996) saw the potential: "Treatments can be applied with field-scale equipment in large plots that can be harvested with field-scale harvesters equipped with yield monitors. Such capabilities open new opportunities for farmers to do more reliable research on their own farms and for researchers to move their studies to the farm".

The difference between design-based and model-based statistics concerns how randomness is incorporated into an experiment. In design-based statistics treatments are allocated to plots at random, conditional on the blocking structure, but we consider the value observed at each location to be fixed. In model-based statistics the sampling locations are fixed, but we consider the values at the locations as a realization of a modelled random function (Brus and de Gruijter 1997). The implication of the latter point is that, in contrast to design-based analysis, sampling for model-based analysis need not rely on randomization (Lark and Cullis 2004). In the context

of agronomic experiments, model-based statistics are effectively a geostatistical survey of the treatment response; therefore we should prefer the relatively even spatial cover that is offered by a systematic sampling design. We realize, however, that relative to design-based analysis, a geostatistical analysis of an experiment is more data-demanding, requires greater computational effort and does not have the advantage of being design-unbiased. De Gruijter et al. (2006) note that to gain advantage over design-based analysis, the data used for model-based statistical analysis must satisfy three conditions: (i) the variable of interest must be spatially autocorrelated; (ii) many samples must be taken at spatial intervals much smaller than the range of the variogram and (iii) there must be many data. On-farm experimentation, when used in conjunction with precision agriculture technology (e.g. variable-rate input-applicators and yield monitors), satisfies these conditions.

Cook and Bramley (2000) drew a disturbing analogy between the decline of the British steel industry in the second half of the twentieth century and the situation experienced by many farmers today. The root of the problem for the former was its inability to control a highly variable production process compared with its competitors. Cook and Bramley described the lesson: "... improved control based on sound information is an essential feature of industrial process. Few, if any, modern industries operate without knowing, in detail, about the processes which go on within them. This knowledge enables managers to maximise efficiency and to respond to external requirements." Cook et al. (1999) likened the principles of PA-based agronomic experimentation to those applied by modern industries, e.g. the Taguchi philosophy (Peace 1993): a deliberate perturbation of a controllable input to a system (e.g. a fertilizer experiment with a suitable design), followed by monitoring of the effect of the perturbation on the system's output (i.e. a real-time yield monitor) will provide vast magnitudes of response data that can aid decision-making and help to improve control of the production process. With this in mind, a farmer might approach experimentation with one of three aims in order of increasing complexity (Lark and Wheeler 2003): (i) to find the optimum uniform treatment for a field of interest (a *whole-field* experiment); (ii) to find the optimum treatment at the spatial resolution of within-field management classes (a *management-class* experiment) or (iii) to find the optimum treatment as it varies spatially across a field (a *whole-of-block* or *local-response* experiment). With regard to management-class and local-response experiments, different designs and model-based statistical techniques are required for each, which we discuss in more detail below. We deal no further with model-based analysis of a whole-field experiment; the analysis will proceed as for a management-class experiment.

10.3 Management-Class Experiments

The principal goals of a management-class experiment are to compare how a crop responds to the applied treatments, between and within classes. The agronomic inputs used as treatments in a management-class experiment may be categorical (such as *types* of fertilizer, *dates* of sowing) or continuous (such as *rates* of fertilizer,

depths of sowing). As long as we collect enough data to make an inference about the results, the experimental design does not need to cover the whole area of interest. Technically, we should refer to them as *potential management-class experiments* until we are presented with evidence that site-specific crop management is justified.

When a farmer does a management-class experiment, an important issue to be aware of is that each yield observation does not represent an independent source of information. This lack of independence arises from two sources: (i) spatial autocorrelation between plots (the stronger is the autocorrelation, the less is the information provided by each additional observation, and vice versa) and (ii) the fact that harvesters generally record many more than one yield observation per plot. With regard to the latter, each yield observation associated with an individual plot is *not* an additional degree-of-freedom for the experiment; to consider the observations in this way constitutes pseudo-replication (Hurlbert 1984), and increases the chance of finding statistically significant treatment effects simply by virtue of the spatial resolution of a yield monitor.

In the context of design-based analysis, we might analyse a management-class experiment with a linear model:

$$\mathbf{z} = \mathbf{M}\boldsymbol{\beta} + \mathbf{e}, \tag{10.3}$$

where $\mathbf{z}$ is an $n \times 1$ vector of observed crop yields resulting from an experiment and $\mathbf{M}$ is an n-row design matrix that associates each observation with a value of one of q fixed effects in the linear model. The fixed effects may include the class means and terms of a polynomial function of the applied inputs. In a management-class experiment that addresses the hypothesis of precision agriculture, it is necessary that there are potentially different response functions in the different classes. In a simple case of two classes with independent quadratic response functions, $q = 6$. The coefficients of the fixed effects in $\boldsymbol{\beta}$ will be the class means and coefficients of the respective polynomial terms (by class). The $n \times 1$ vector $\mathbf{e}$ contains independent random errors. The parameters of $\boldsymbol{\beta}$ are estimated conventionally by ordinary least-squares (OLS), which is justified by the assumption that the errors are independent identically distributed random variables.

In the context of model-based analysis, we re-define Eq. 10.3 as a linear mixed model. Unlike Eq. 10.3, the linear mixed model deals explicitly with spatial autocorrelation and pseudo-replication. According to the theory of geostatistics, crop yield is now a random function of its spatial coordinates, with observations denoted $\mathbf{z}(\mathbf{x})$. The linear mixed model that we use is:

$$\mathbf{z}(\mathbf{x}) = \mathbf{M}\boldsymbol{\beta} + \mathbf{u}(\mathbf{x}), \tag{10.4}$$

where $\mathbf{M}$ and $\boldsymbol{\beta}$ retain the same meaning as above; however, in contrast to Eq. 10.3 the error term has been replaced by the $n \times 1$ vector $\mathbf{u}(\mathbf{x})$, which contains a realization of a second-order stationary, normally distributed, spatially correlated random function, $U(\mathbf{x})$. Thus the variation in yield is divided into a deterministic component (the *fixed effects*) and a random component (the *random effects*). The variation

in $\mathbf{u}(\mathbf{x})$ is described by an $n \times n$ covariance matrix, $\mathbf{V}$, with individual elements $V_{i,j} = \sigma^2 - \gamma(\mathbf{h})$, where σ^2 is the a priori variance of $U(\mathbf{x})$ and $\gamma(\mathbf{h})$ is the variogram of $\mathbf{u}(\mathbf{x})$ described with some bounded authorized function (Webster and Oliver 2007).

We can estimate the coefficients in $\boldsymbol{\beta}$ and the parameters of the variogram of $U(\mathbf{x})$ (denoted $\boldsymbol{\alpha}$) by residual maximum likelihood (REML) (Patterson and Thompson 1971). Pardo-Igúzquiza and Dowd (1998) suggested that a second-order stationary variogram could be fitted by finding $\boldsymbol{\alpha}$ that maximizes a likelihood function, $l(\boldsymbol{\beta}, \boldsymbol{\alpha}|\mathbf{z})$ (we herein drop the spatial coordinates from $\mathbf{z}$ for clarity). Parameters of the linear mixed model could be estimated by this maximum likelihood approach, although the estimates will be biased in the presence of a deterministic trend (Lark and Cullis 2004). The REML approach reduces bias in the estimated variogram parameters by projecting $\mathbf{z}$ into a residual space where all the fixed effects have zero expectation. The residual log-likelihood of the projected data is (after Lark and Cullis 2004):

$$l_R\left(\boldsymbol{\alpha}|\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}\right) = \text{constant} - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{Q}| - \frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{I} - \mathbf{M}\mathbf{Q}^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1})\mathbf{z}, \tag{10.5}$$

where 'constant' represents terms in the likelihood that do not depend on $\boldsymbol{\alpha}$ (and can therefore be ignored); $\mathbf{V}$ is calculated from the estimates of the variogram parameters, $\hat{\boldsymbol{\alpha}}$; $\mathbf{Q}$ is given by $\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{M}$ and estimates of the fixed effects, $\hat{\boldsymbol{\beta}}$, are derived by generalized least squares:

$$\hat{\boldsymbol{\beta}} = \mathbf{Q}^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{z}. \tag{10.6}$$

The residual log-likelihood is generally multiplied by -1.0 and minimized by a numerical algorithm to determine $\hat{\boldsymbol{\alpha}}$ (Lark and Cullis 2004).

Two studies that have used REML to analyse PA-based experiments are Lambert et al. (2004) and Minasny and McBratney (2007). The latter described REML-estimated yield response functions for an Australian cotton field subjected to a nitrogen fertilizer management-class experiment. Unfortunately the design of the experiment was flawed in that the smallest rate of nitrogen applied was not substantially different from the field's average application. Consequently, the value of the response functions was questionable. Lambert et al. (2004) found for a field in Argentina statistically significant interactions between the yield of maize, the rate of applied nitrogen (defined according to a quadratic response function) and the management class (defined according to the field's topography). They showed that, by using REML to account for spatially correlated model error, the profitability of PA was greater than when the same model was implemented with OLS. We now present a case study to demonstrate why, from a statistical viewpoint, we prefer to apply a REML-based analysis rather than an OLS-based one for a management-class experiment.

10.3.1 Case Study I: REML-Based Analysis of a Management-Class Experiment

Rosewood is a 75 ha field in northern NSW, Australia. Two management classes were delineated for the field – based on data from yield mapping in 2003, a profile-depth soil EC_a survey and a digital elevation model – according to the protocol described by Taylor et al. (2007) (Fig. 10.1). A nitrogen fertilizer experiment was established in Rosewood field in 2004. The crop was winter wheat. In each management-class, six plots 100 m long and four harvester-swaths wide were established and aligned in the direction of field operations. In each class the plots were placed randomly, subject to the constraint that there was no overlap between plots and no overlap with the class boundaries. Two replicates of the three treatments (0, 27 and 83 kg N ha^{-1}) were allocated randomly to the six plots in each class.

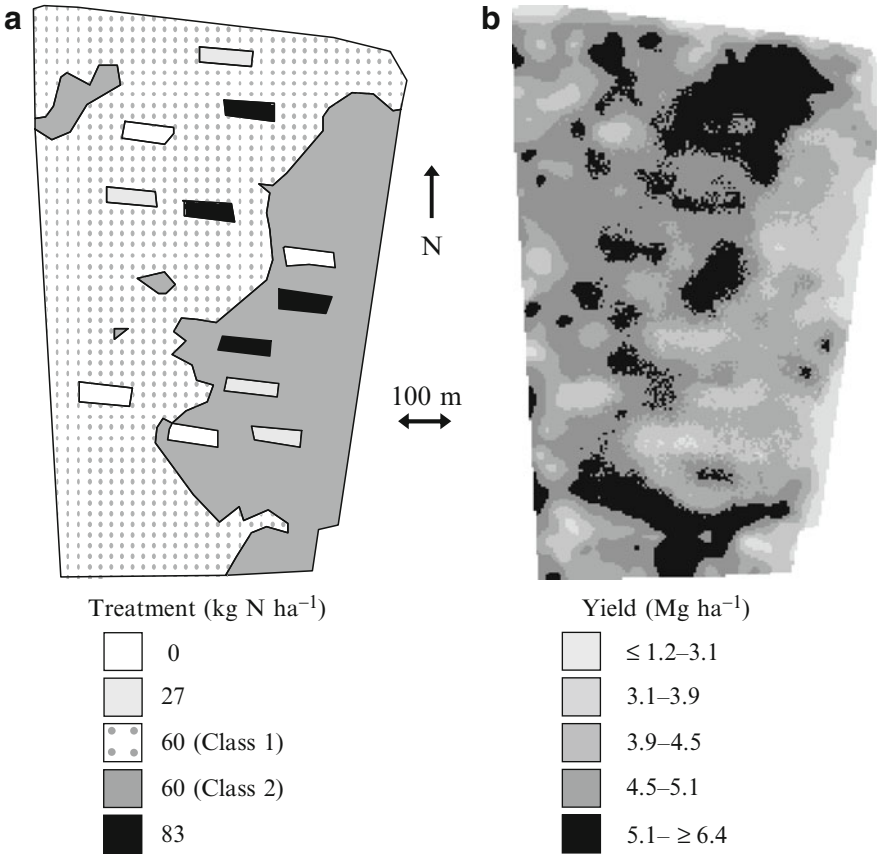

Fig. 10.1 Rosewood field: (**a**) the spatial arrangement of the nitrogen-fertilizer experiment and locations of the management-classes and (**b**) the yield of wheat at the end of the growing season

The rest of the area in both classes received a fertilizer application of 60 kg N ha^{-1}, which was the uniform rate determined by the farm's agronomist to achieve the field's targeted grain yield of 4.5 Mg ha^{-1}. The 60 kg N ha^{-1} rate has an important implication for the statistical analysis: it has no chance of being found in the small plots (and conversely the rates applied to the plots have no chance of being found in the rest of the field). The design as a whole is not truly randomized. Design-based statistics are not appropriate, therefore, to analyse this experiment we must use model-based statistics instead.

The specific null hypothesis that we wish to investigate is that each coefficient of a polynomial yield response function (i.e. $\boldsymbol{\beta}$ from the description above) for each management-class is zero. We estimate the coefficients in $\boldsymbol{\beta}$ with REML. Despite the fact that OLS is inappropriate for the experiment, for comparison we use it here also to estimate $\boldsymbol{\beta}$. We implemented REML with the aid of the GeoR library (Ribeiro and Diggle 2001) developed for the R software package (R Development Core Team 2008). We considered only a spherical covariance function (Webster and Oliver 2007) for use with REML.

There were 11 523 observations in the original yield map of the field. Unfortunately, REML cannot handle this large number of data. Prior to our implementation of REML and OLS we sub-sampled the data associated with the 60 kg N ha^{-1} treatment that covered the majority of the field. We divided the histogram of yield associated with the 60 kg N ha^{-1} treatment into 1000 strata, and took one random observation from each. It was not necessary to sub-sample the yield data associated with the 0, 27 and 83 kg N ha^{-1} treatments. Following sub-sampling, Class 1 comprised 1 343 yield observations and Class 2 comprised 1 209.

Table 10.1 gives the best-fitting polynomial response function for each management-class, for each method of estimation. At a threshold of $p = 0.05$, according to OLS a quadratic response function is justified for Class 2, but only a linear function is justified for Class 1. The results for REML agree with the OLS result for Class 1; however, for Class 2 REML shows that the quadratic term is not justified. The OLS-estimate of β_2 in Class 2 leads us to reject the null hypothesis for the coefficient, a false positive result. The corresponding coefficient estimates were similar for each method, but the standard errors returned by REML are larger than for OLS. This is because REML accounts for any autocorrelation, so consequently one observation equates to something less than one piece of information about yield response. Therefore, relative to the result from OLS, we remain more uncertain about the true value of each REML-estimated parameter. The response functions are presented graphically in Fig. 10.2. Assuming that the aim is to maximise yield, in Class 1 the method that we use to estimate the response function has no effect on the optimum fertilizer rate: both OLS and REML suggest that the optimum is somewhat greater than the maximum used in the experiment (83 kg N ha^{-1}). In Class 2 our choice of the optimum fertilizer rate varies according to the method we used to estimate the yield response function: OLS suggests that 60 kg N ha^{-1} maximizes yield (i.e. the slope of the function at this input is zero), but REML suggests 83 kg N ha^{-1}. Thus OLS leads the farmer to think (incorrectly) that there is a benefit in decreasing the amount of nitrogen applied to Class 2. As for Class 1,

Table 10.1 Rosewood field. Polynomial yield-response functions estimated with ordinary least-squares (OLS) and residual maximum likelihood (REML)

Class	Method	Coefficient	Estimate	Standard error	*t*-statistic	*p*
1	OLS	β_0	3.70×10^{0}	6.98×10^{-2}	52.98	<0.001
		β_1	1.87×10^{-2}	1.21×10^{-3}	15.40	<0.001
	REML	β_0	3.83×10^{0}	1.80×10^{-1}	21.28	<0.001
		β_1	1.52×10^{-2}	2.73×10^{-3}	5.57	<0.001
2	OLS	β_0	3.41×10^{0}	8.72×10^{-2}	39.10	<0.001
		β_1	2.20×10^{-2}	4.04×10^{-3}	5.44	<0.001
		β_2	-1.80×10^{-4}	5.00×10^{-5}	−3.72	<0.001
	REML	β_0	3.54×10^{0}	1.87×10^{-1}	18.95	<0.001
		β_1	7.88×10^{-3}	2.54×10^{-3}	3.10	0.002

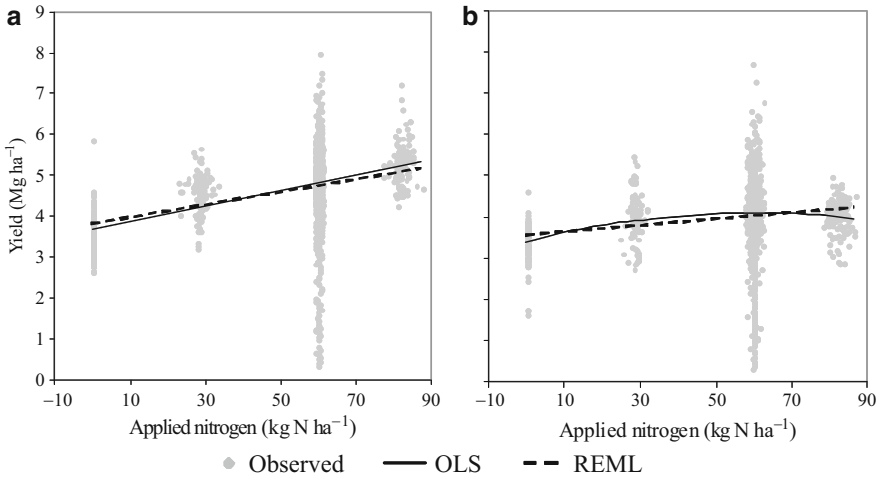

Fig. 10.2 Wheat yield on Rosewood field as a function of applied nitrogen: (**a**) Class 1 and (**b**) Class 2. The models for each class correspond to the coefficients given in Table 10.1

the REML-based estimates suggest that if the aim is to maximize yield, then further experiments with applied nitrogen >83 kg N ha^{-1} would be justified.

Residual maximum likelihood has two disadvantages. The first is a disadvantage also shared by OLS: the response function must be linear in its parameters (such as a polynomial). This requirement disregards the common observation that yield response, to nitrogen at least, is generally asymptotic, which would necessitate that a model should have a non-linear parameter (such as an exponential term). The implication is that polynomial response functions are convenient, rather than sensible (Boyd et al. 1976). (Indeed, they are a convenience that recurs through this chapter!) But a biologically sensible model need not preclude the use of REML, provided that the user is careful: Lark and Wheeler (2003) showed how to combine successfully a likelihood-based analysis and an asymptotic response function (discussed below). The second disadvantage of REML is that it is limited by the number of data it

can analyse in an acceptable time. This number will vary according to the computer and the software available, and the time constraints on a given project. The reason for the limitation is that every iterated value in $\hat{\boldsymbol{\alpha}}$ requires a computationally expensive inversion of the covariance matrix, $\mathbf{V}$ (Eq. 10.5). A less-demanding approach is approximate maximum likelihood (Pardo-Igúquiza and Dowd 1997; Stein et al. 2004), but it is unclear whether the principle can be translated successfully to REML.

10.4 Local-Response Experiments

Unlike management-class experiments, local-response experiments must cover the extent of the area of interest. The principal goal of a local-response experiment is to examine the fine-scale spatial variation of crop response to one (or more) agronomic input(s) and to derive a fine-scale, spatially-variable optimum application (e.g. Davis et al. 1996; Bramley et al. 1999). In an arable field the inputs used as treatments in a local-response experiment will almost certainly be continuous (e.g. *rates* of fertilizer, *depths* of sowing); machinery constraints should preclude inputs that are categorical (e.g. *types* of fertilizer, *dates* of sowing). (However, Panten et al. 2010, have demonstrated that it *is* possible to devise a local-response experiment with treatments of qualitative inputs for intensively managed horticultural crops – in their case various ground-cover management strategies.) Statistical analysis of a local-response experiment is complex because we must determine if there are significant differences in the treatment effects at a particular location, subject to the constraint that we can apply only one treatment per location. As well as dealing with inference issues for a field-scale experiment, this is in addition an interpolation issue.

McBratney (1985) proposed that kriging be used to interpolate the yield observations of a particular treatment to the locations associated with other treatments. This would create maps – one for each treatment – of what the yield would have been had each treatment been applied uniformly to the extent of the experimental domain. Bruulsema et al. (1996) used this idea on a local-response experiment. They fitted a quadratic response function through each of the interpolated values, weighted by the inverse of the kriging variance. Doerge and Gardner (1999) also interpolated across treatments, but the interpolator was not specified.

Pringle et al. (2004) studied the response of wheat grain yield to various local-response experiments. Their preferred interpolator was cokriging, with the yield map of the previous season's crop used as a covariate. Goovaerts (1997) and Webster and Oliver (2007) show that cokriging can return more accurate estimates than ordinary kriging, and does so with a smaller estimation variance. With this in mind, the rationale of Pringle et al. was that, because they saw similar spatial patterns to yield between seasons, the yield information of the first season could be used to guide the interpolation of yield in the second season. (It is possible that a pair of consecutive yield maps may be related non-linearly due to strong inter-season climatic variability, for example. In this case a prior yield map would not be a good choice

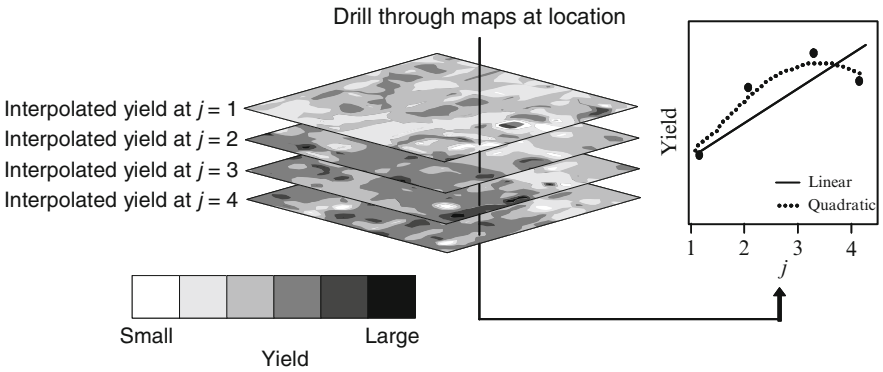

Fig. 10.3 The yield of a hypothetical crop at four levels of a treatment j, obtained by kriging across each treatment. The local yield response is estimated by the best-fitting polynomial response function (adapted from Pringle et al. 2004)

of covariate – Cook and Bramley 2000.) The basic method of Pringle et al. (2004) is shown in Fig. 10.3. Consider that we wish to estimate the local yield response to an experiment with $j = 1, \ldots, 4$ treatment levels. The yield data associated with the experiment are interpolated at the nodes of a grid. The yield data associated with the yield map of the previous season are interpolated at the same nodes. Collocated cokriging (Goovaerts 1997), a computationally fast variant of ordinary cokriging, is then used to interpolate the yield associated with the locations of $j = 1$ at locations $j \neq 1$, and so on for all j. The grid nodes of the resulting maps are then 'drilled' so that the yield estimates can be modelled as a first- or second-order polynomial function of the treatments, weighted by the inverse of the cokriging estimation variance for each treatment. The most parsimonious polynomial is chosen to describe the treatment effect, which is taken as the difference between the observed yield at the location of interest and the maximum yield according to the response function.

The method of Pringle et al. (2004) has four important shortcomings. First, they interpolated their data twice to derive estimates of local yield response (once to a set of common grid nodes, then again across treatments with collocated cokriging). Consequently, much of the observed fine-scale yield variation would be smoothed and only the strongest responses would remain. (The first interpolation is unnecessary if one manages to associate each yield monitor observation with a treatment; this would be a relatively easy operation in a GIS.) The second shortcoming follows from the first: the estimation variance associated with a collocated cokriging prediction would be greater than that of an ordinary cokriging prediction (Papritz 2008), so what Pringle et al. gained in computing efficiency they lost in statistical precision. Third, while their approach ensured optimum predictions of the response to an individual treatment, optimum predictions for the difference between treatments (i.e. the *contrasts*) were not considered. Fourth, Lark and Wheeler (2003) noted that the hypothesis testing used by Pringle et al. was ad hoc because the latter's method

did not obtain correct estimates for the standard errors of the parameters of the yield response functions.

An alternative approach was presented by Lark and Wheeler (2003) in an explicitly model-based context. Their central idea was to use raw yield-monitor observations, rather than interpolated estimates. For a local-response experiment with various nitrogen application rates, they considered that crop yield at any location was a function of two models: (i) a model that described how the yield recorded by the monitor related to yield on the ground (Whelan and McBratney 2002) and (ii) an asymptotic yield-response function:

$$z = a + bR^N, \tag{10.7}$$

where z is yield, N is the rate of applied nitrogen, and a, b and R are parameters of the function. Lark and Wheeler used the first model to correct the yield monitor data and to estimate a field-average value for the non-linear parameter (R) of the yield-response function. They then estimated by maximum likelihood the (linear) parameters a and b using only the nearest 90 yield observations to each node on a fine grid across the field's extent. The residuals of the localized yield response functions were assumed to be a realization of a random function described by an exponential variogram, whose parameters were estimated simultaneously during the maximum likelihood process. A bootstrap estimate of the standard deviation of b enabled Lark and Wheeler to reject a null hypothesis that there was no spatial variation in the yield response functions across the field. A map of the economically optimum N rate for the field, derived from the analysis, showed substantial spatial variation.

Bishop and Lark (2006) improved on the interpolation procedure of Pringle et al. (2004) to analyse local-response experiments. Their principle was to consider the yield responses of each individual treatment as coregionalized variables that can be mapped simultaneously by cokriging. This ensures that the estimates of individual treatment responses are optimal, as are the estimates of their contrasts. A further advantage of the method of Bishop and Lark (2006) is that the information provided by the responses at all treatment levels is used to improve the prediction quality of any individual treatment.

From the cokriging prediction variances one can obtain the kriging variance associated with a contrast; this could be a simple pair-wise contrast (Bishop and Lark 2006) or a more complex contrast such as the comparison of a control with the mean of two other treatments (Bishop and Lark 2007). A test statistic is derived from the contrast kriging variance:

$$\hat{s}_k(\mathbf{x_0}) = \frac{\hat{d}_k(\mathbf{x_0})}{\sqrt{\hat{\sigma}_k^2(\mathbf{x_0})}}, \tag{10.8}$$

where $\hat{s}_k(\mathbf{x_0})$ is the test statistic for location $\mathbf{x_0}$, $\hat{d}_k(\mathbf{x_0})$ is the cokriged estimate of the kth contrast and $\hat{\sigma}_k^2(\mathbf{x_0})$ is the estimate of the kriging variance of the kth contrast. The null hypothesis, i.e. that the contrast $= 0$, will be rejected if

$|\hat{s}_k(\mathbf{x_0})| > 1.96$ (assuming normally distributed data). The relation between $\hat{d}_k(\mathbf{x_0})$ and the cokriged estimates of the response to the jth treatment, $z_j(\mathbf{x_0})$, for a simple pair-wise contrast is:

$$\hat{d}_k(\mathbf{x_0}) = c_2 z_2(\mathbf{x_0}) + c_1 z_1(\mathbf{x_0}), \tag{10.9}$$

where c_1 and c_2 are coefficients equal to 1 and -1, respectively. Equation 10.9 can be generalized to a form capable of accommodating contrasts that involve more than two treatments:

$$\hat{d}_k(\mathbf{x_0}) = \sum_{j=1}^{t} c_j \hat{z}_j(\mathbf{x_0}), \tag{10.10}$$

where t is the number of treatments. For example, if we want to compare a control treatment with the mean of two other treatments the contrast coefficients would be -2 for the control and 1 for the other treatments, i.e. $1z_2 + 1z_3 - 2z_1$.

But which contrasts should be considered? For example, for an experiment with four treatments there will be six simple pair-wise contrasts (Eq. 10.9) that can be examined, in addition to other more complex linear combinations of the treatment responses. Some contrasts are not independent so, without care, we will have trouble in interpreting the results. Furthermore, the simultaneous testing of many (or all) possible comparisons increases the chance of rejecting a null hypothesis incorrectly (a false positive result). One way to avoid these issues is to split the treatment sum of squares (SS) into meaningful components in which each component is associated with a particular independent (orthogonal) treatment contrast. The contrasts should be pre-planned based on the specific hypotheses being tested, and should be reflected in the choice of treatment structure for the experiment. The number of orthogonal treatment contrasts that are permitted is $t - 1$ (i.e. the treatment degrees of freedom).

To have a set of orthogonal treatment contrasts, that is, to have the sum of the SS for each treatment contrast equal to the treatment SS, we must satisfy two constraints on the response parameters. First, the coefficients for a particular contrast must sum to zero:

$$\sum_{j=1}^{t-1} c_j = 0. \tag{10.11}$$

Second, the sum of the cross-products for the set of orthogonal treatment contrasts must equal zero:

$$\sum_{j=1}^{t-1} c_{p,j} c_{q,j} = 0, \tag{10.12}$$

where $c_{p,j}$ and $c_{q,j}$ are coefficients for the pth and qth contrasts.

Bishop and Lark (2007) used orthogonal contrasts to compare means. In experiments with quantitative treatments and continuity between treatment levels it is better to examine the functional relation between the response and treatment levels.

Table 10.2 Contrast coefficients (c) for a trend comparison with four treatments

Treatment (kg N ha^{-1})	Linear	Quadratic	Cubic
0: 0	−3 [−3]	+1 [+1]	−1 [−1]
1: 100	−1 [−0.58]	−1 [−1.47]	+3 [+4.37]
2: 170	+1 [+1.12]	−1 [−0.75]	−3 [−5.84]
3: 225	+3 [+2.46]	+1 [+1.22]	+1 [+2.47]

[] = coefficients used in Case study II.

This type of analysis is called a *trend comparison*. For t treatment levels, $t - 2$ is the highest order polynomial whose pure contribution to the treatment effect can be estimated. One additional comparison of the sum of squares can be estimated that represents all higher-order components, e.g. for $t = 4$ this would be referred to as a cubic trend comparison. A significant cubic trend indicates that the yield variation is more complex than a quadratic trend. Table 10.2 presents coefficient values for a trend comparison with $t = 4$. They are standard values that can be found in most statistical textbooks (e.g. Gomez and Gomez 1984), and are appropriate when the treatments have an even separation interval. Importantly, these coefficients satisfy the constraints in Eqs. (10.11) and (10.12). The coefficients in Table 10.2 are just one subset of all possible orthogonal contrasts that can be tested, except in this case where the coefficients correspond to particular trends.

10.4.1 Case Study II: Analysis of a Local-Response Experiment

In 2000–2001 a nitrogen-response experiment was performed in a field of 8.9 ha, known as 'Bypass', in Bedfordshire, United Kingdom. It was a randomized block design with four treatments (0, 100, 170 and 225 kg N ha^{-1}). Plots were approximately 12 × 15 m. Wheat yield was measured with a yield monitor at the season's end. Figure 10.4 shows the locations of yield observations associated with each treatment. Since the interval between the treatment levels is not even in the Bypass experiment, the standard contrast coefficients (Table 10.2) cannot be used. In this case a new set of coefficients must be solved for each order of polynomial, based on Eqs. 10.11–10.12 and the basic form of linear, quadratic and cubic polynomial equations (see p. 229–233 of Gomez and Gomez 1984). The coefficients calculated for the experiments are shown in square brackets in Table 10.2. Summary statistics of yield for each treatment (Fig. 10.5) show that any application of nitrogen fertilizer increases the mean yield by 1–1.2 Mg ha^{-1} compared with the 0 kg N ha^{-1} treatment. Also, the addition of nitrogen results in a slight negative skewness compared with an apparently normal distribution when no nitrogen is applied.

The treatment responses for both experiments were interpolated at the nodes of a 2-m grid, using 2-m blocks and a 50 m search radius. The interpolator was standardized cokriging, which returned local estimates of the treatment contrast

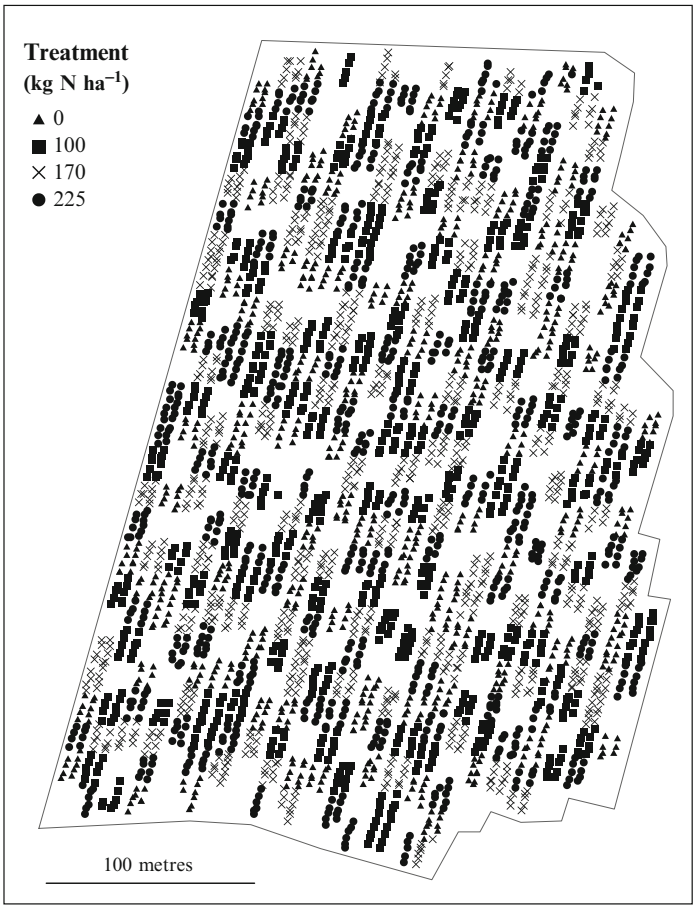

Fig. 10.4 The locations of yield observations associated with the nitrogen fertilizer treatments for Bypass field

(Fig. 10.6). Standardized cokriging is a variant of ordinary cokriging, but where the latter accounts explicitly for the different means of the different treatments. Standardized cokriging essentially scales each treatment to the mean of the treatment of interest (Goovaerts 1997). The method has been criticized strongly by Papritz (2008), but this does not preclude us from illustrating our point. The choice of dimension for a cokriging block is subjective. We used 2-m because it decreased the estimation variance relative to point-based cokriging, but still preserved much of the inherent variability of the data. The 2-m grid and block scheme effectively creates raster-based maps for each treatment. The 0 kg N ha^{-1} treatment clearly has the smallest yields, but there is less difference between the other treatment response maps. Figure 10.6 shows that there are two distinct regions of yield variation in the field that are common to all treatments: the north-east corner has the smallest yields,

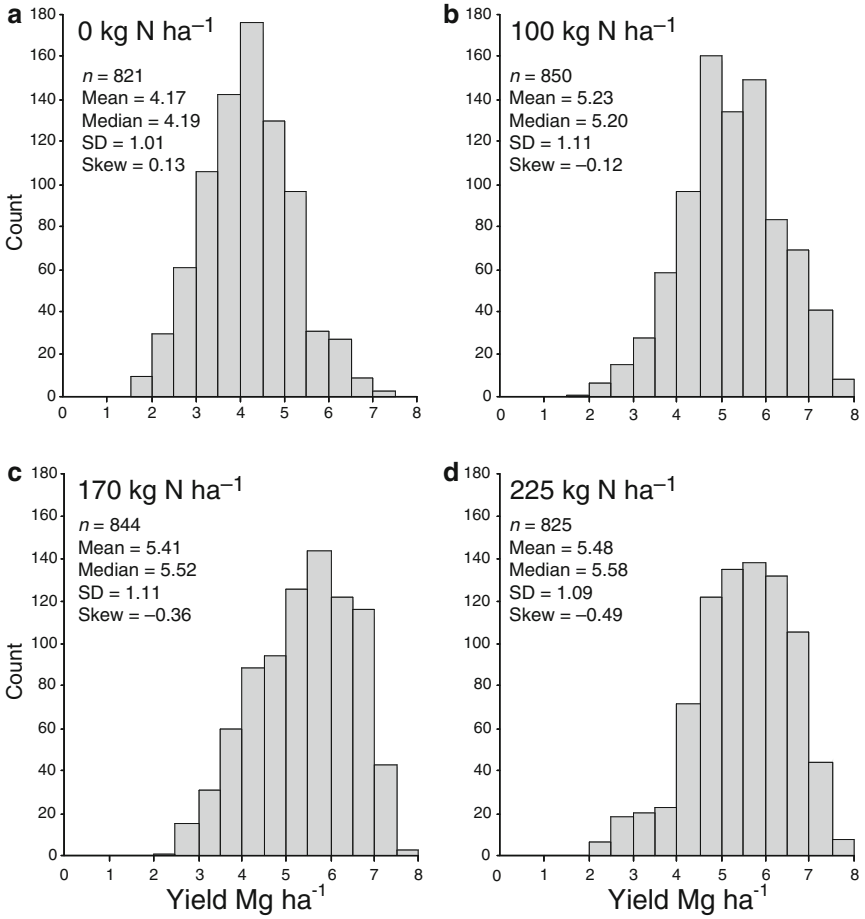

Fig. 10.5 Summary statistics for the yield observed for each nitrogen fertilizer treatment for Bypass field: (**a**) 0 kg N ha^{-1}, (**b**) 100 kg N ha^{-1}, 170 kg N ha^{-1} and (**d**) 225 kg N ha^{-1}

but in the rest of the field the yields are larger. Based on limited soil sampling, these regions correspond to different soil series: the north-east region has a coarse texture, possibly from glacial drifts overlying the Gault Clay, whereas the rest of the field is a heavy clay.

Equation 10.8 was used to test the significance of the trend comparisons for each 2-m block, based on the coefficients in Table 10.2. The maps of $\hat{s}_k(\mathbf{x_0})$ are shown in Fig. 10.7. The individual hypotheses being tested are whether the yield response has a linear, quadratic or more complex component at each location of interest. The maps of $\hat{s}_k(\mathbf{x_0})$ show that most of the field showed at least a linear response to the nitrogen fertilizer (Fig. 10.7a). A significant quadratic trend is identified in many parts of the field that exhibit a linear trend (Fig. 10.7b). Parts of the field also show a more complex yield response than linear and quadratic trends (Fig. 10.7c). As noted

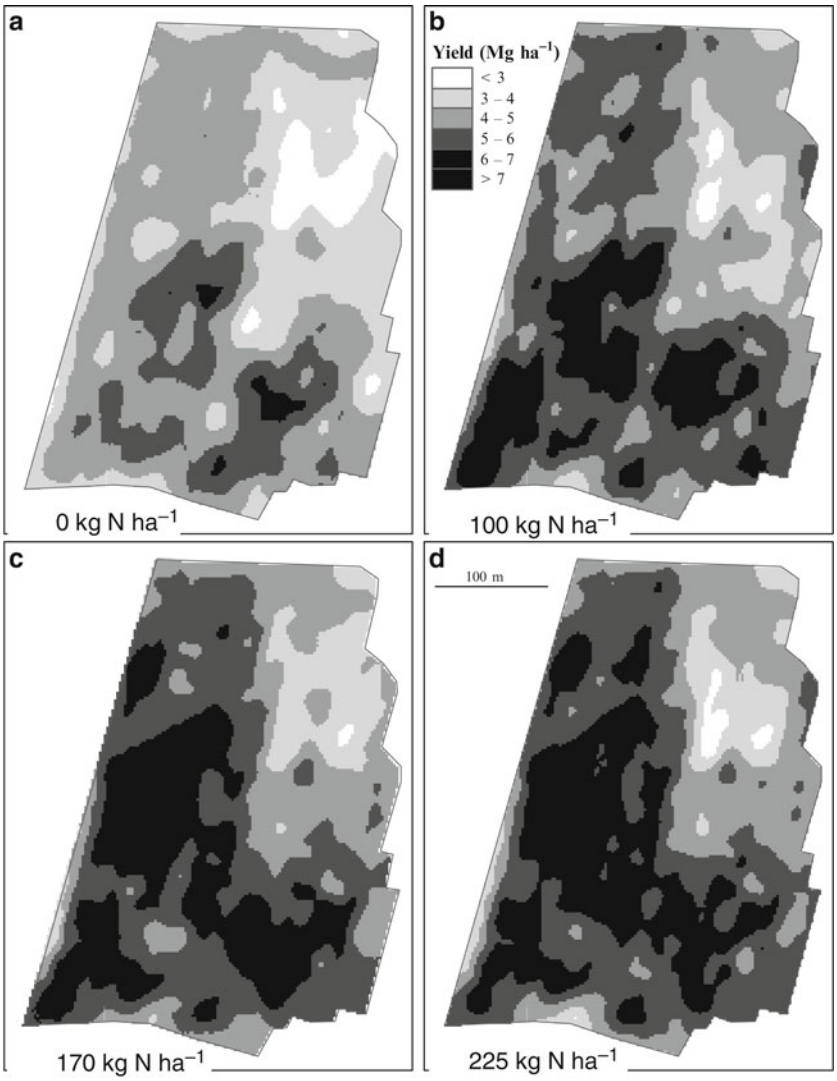

Fig. 10.6 Yield associated with each treatment level, interpolated by standardized cokriging for Bypass field: (**a**) 0 kg N ha^{-1}, (**b**) 100 kg N ha^{-1}, 170 kg N ha^{-1} and (**d**) 225 kg N ha^{-1}

above, the cubic trend is a residual term and we can only state that in these regions the trend contains higher-order terms than a quadratic.

The comparison of trend could be extended in two ways. First, we may use it to derive polynomial coefficients. From these it would be possible to derive yield response functions, perhaps following the method of Lark and Wheeler (2003). Second, secondary information could be used as a covariate for cokriging or included as a fixed effect.

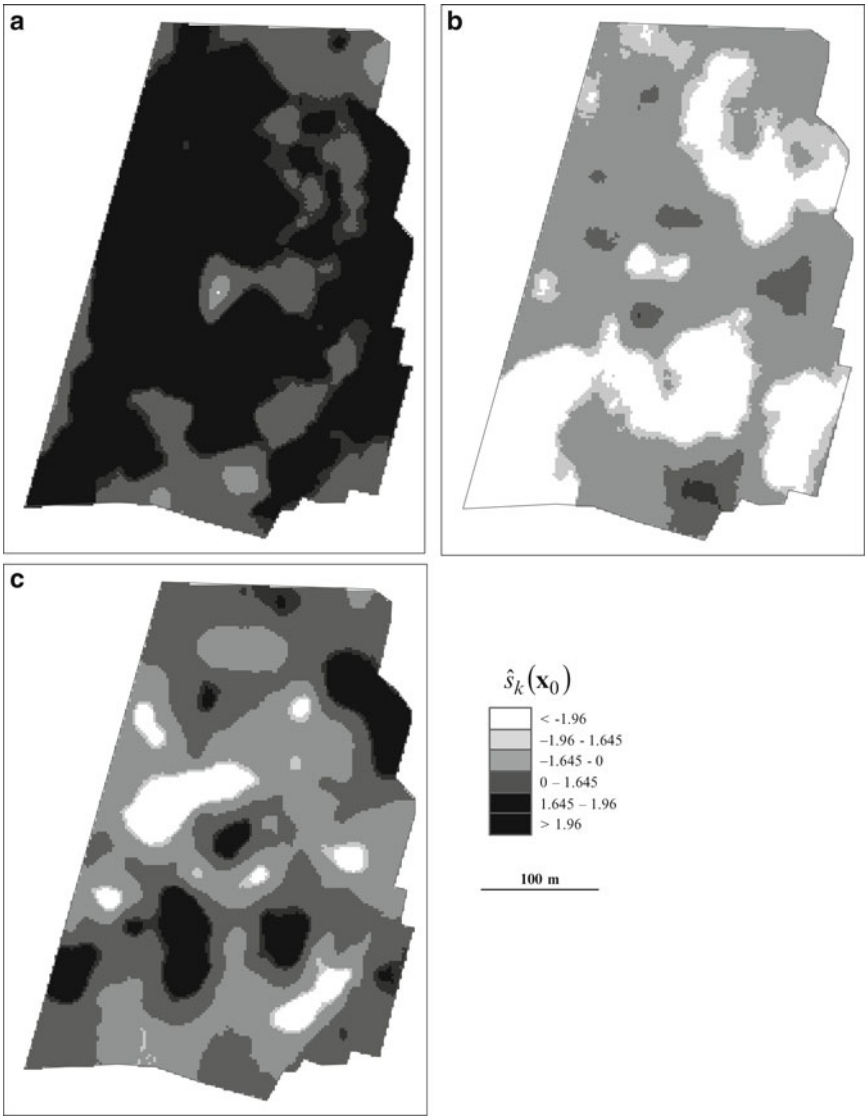

Fig. 10.7 Comparison of trend for Bypass field: (**a**) linear, (**b**) quadratic and (**c**) more complex. The trend is significant when $|\hat{s}_k(\mathbf{x}_0)| > 1.96$

10.5 Alternative Approaches to Experimentation

We have discussed the geostatistical analysis of PA-based agronomic experiments conducted with the aid of variable-rate technology and yield monitors. In our experience this is the norm under which this kind of experimentation operates; however, we acknowledge two important alternatives.

First, what happens when a farmer has an interest in PA but does not have variable-rate technology or a yield monitor? It is not necessary to possess either. For example, Panten et al. (2010) analysed a viticultural experiment geostatistically that related under-vine ground-cover treatments to hand-gathered mid-season crop attributes. This laborious and intense information-gathering was feasible for a vineyard, but it is doubtful whether a similar approach could be applied to an arable field. Undoubtedly the scope for experimentation in large-scale grain or fibre cropping will be limited if a farmer does not invest in variable-rate technology or a yield monitor. Nevertheless, there are still ways in which information can be gathered that do not require a substantial financial outlay. For example, a farmer might obtain a map of his or her farm and delineate those sub-regions believed to be consistently large or small yielding for individual fields. A replicated strip trial – which does not require variable-rate technology to apply – could then be applied to the fields. In the absence of a yield monitor, the farmer may obtain a mid-season aerial photograph or a satellite image of the farm (particularly now that imagery such as Landsat can now be downloaded freely), and judge the results of the trials. If these results agree with intuition, the farmer may decide to change his or her operation accordingly in the following growing season; the benefit of any changes made could be evaluated with another experiment. Some – particularly statisticians – may be uncomfortable with this kind of informal approach to PA, but others may find it empowering because it develops at a level of commitment that the farmer is comfortable with, and requires little more than common sense to implement.

Second, an alternative to geostatistical analysis, not easily categorized as either a design-based or a model-based method, is discrete spatial autoregression (Anselin 1988). This approach has been used to illustrate the economic benefits of PA (e.g. Anselin et al. 2004; Lambert et al. 2004; Hurley et al. 2005). The principal difference between the two methods concerns the definition of spatial covariance: in geostatistics the spatial covariance is considered to be a continuous function of separation distance; in discrete spatial autoregression the covariance arises through the inter-relationships between discrete units (such as management classes), which have a pre-defined neighbourhood structure. In discrete spatial autoregression, in order to match the observations to the neighbourhood structure, irregularly spaced data must be interpolated at the nodes of a regular grid, which entails some loss of information. The studies of Lambert et al. (2004) and Hurley et al. (2005) showed that, regardless of whether the PA-experiments were analysed with discrete spatial autoregression or with geostatistics (in the form of REML), the profitability of PA was similar. Discrete spatial autoregression is computationally efficient compared with geostatistics; however, the continuous covariance model of geostatistics is arguably more suitable for variables such as crop yield whose variation in space is clearly continuous. Furthermore, geostatistical models can be used in a fully model-based way, which makes them less arbitrary and more suitable for inference based on the uncertainty of the covariance parameters (Diggle and Ribeiro 2007).

10.6 Issues for the Future

A challenging area for future research concerns the optimum experimental design for a particular situation. By *optimum* we mean a balance between the information gained from the experiment and the inconvenience caused to the farmer. This kind of optimization will inevitably require a forecast of the value of the information generated by the experiment – a notoriously difficult concept to quantify (Pannell 1998).

In a provocative paper, Whelan et al. (2003) dismissed local-response experiments on the basis that they are overly invasive to farm operations; instead, they preferred to concentrate on management-class experiments. They listed some machinery-related issues that constrain designs for PA-based agronomic experimentation. It is worth re-iterating the list here. First, plots must be aligned in the direction of sowing and harvesting. Second, the width of a plot must be at least three times the cutting-width of the harvester; the exact width will be a multiple of the narrowest machinery used on the field (variable-rate input-applicator or harvester). Third, the plots must be long enough to counter spatial mixing of the crop as the harvester moves from plot to plot; 80–100 m is recommended as a general rule, with the first and last 20 m of the plot discarded from all statistical analysis. Whelan et al. (2003) used spatial simulated annealing (Van Groenigen and Stein 1998), in conjunction with prior knowledge about site variability and some simple economic assumptions, to address simultaneously the physical constraints on an experiment and the aforementioned information and inconvenience trade-off. They considered the resulting 'fleck' design (as it might be called) to be the economically, agronomically and biometrically optimum experimental design for a field.

Whelan et al. (2003) were arguing in the context of arable cropping. Their notions may not apply to horticulture. For example, Dr R. Bramley (personal communication) has indicated that vineyard managers are surprisingly positive about their experiences with local-response experiments: his view is that any cost to the manager is outweighed by the experience of "being able to wander through the experiment and see for himself how treatment responses varied spatially". With regard to the optimization of a local-response experiment, Bishop and Lark (2006) have shown how the optimum spatial arrangement of the treatments can be very different depending on whether the goal is to map the treatment response or the contrasts. If we are interested in estimating the contrast then the plots of different treatments should be placed as close together as possible; if, however, we wish to map the individual treatment responses then the plots should be spread across the experimental area. Spatial simulated annealing can be used to ensure an optimum spread of the treatments. Bishop and Lark (2006) used the kriging variance as the criterion to derive the optima; the value of the information generated by the experiment was not considered.

An additional issue is Bayesian statistics. Consider REML; we have shown the statistical benefit of REML above, but by definition it returns information about only one point in the joint distribution of the model parameters: the point of maximum likelihood, which is then assumed to be known and 'plugged in' to further

analyses. Bayesian statistics offer an alternative – arguably more easily interpretable – framework by which one may examine yield responses from an entirely probabilistic viewpoint. Besag and Higdon (1999) described the principles of the Bayesian analysis of agricultural plot trials; the same principles are applicable to large-scale, on-farm experiments, although some modifications may be needed to cope with the greater volume of data. Bayesian statistics are especially adept at dealing with prior knowledge, which farmers have in abundance. The Bayesian Maximum Entropy (BME) approach (Christakos and Li 1998; Christakos 2000) may ultimately prove to be more flexible than classical geostatistics or the linear mixed model. It is a probability-based method that formally incorporates observed data, physical theories and prior knowledge into a map. In the context of experimental analysis for PA, it seems like a perfect marriage. Unfortunately, at this time a lack of wide understanding about the underlying mathematics of BME constrains its application. (It is probable that this was once said of kriging too!) To see PA experiments analysed in a Bayesian context accounting for prior information and with yield responses quantified for a range of probabilities would be an exciting development for precision agriculture.

Finally, we have seen that, in Australia at least, farming groups often express a desire to use experimental techniques to complement their PA activities, but they are reluctant to use anything more complex than a strip trial. This risks creating an absurd disconnection where academics continue research on efficient designs for spatial experiments, but nobody uses the results. Ultimately, the onus is on researchers to adapt their geostatistical analyses to the demands of the farming community.

10.7 Conclusions

Model-based statistical analysis is crucial for inference from experiments conducted within the PA paradigm. Residual maximum likelihood is important for analysis of management-class experiments, whereas trend comparisons are important for analysis of local-response experiments. Geostatistics has an important role to play in each of these. Put simply, spatial autocorrelation must be accounted for if one is to draw a valid inference from a spatial experiment. The optimum experimental design to meet a particular purpose, given the constraints on production and the spatial variability of an area of interest, is an area that requires more research. Ultimately though, if farmers are compelled to use relatively simple designs and less-formal analyses, then researchers must follow, and adapt their geostatistical analyses accordingly.

Acknowledgements Thanks to Mr. Michael Ledingham, 'Merinda', NSW, for the data for Rosewood field. The contribution of RML is part of the programme in Mathematical and Computational Biology at Rothamsted Research, funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC). The analyses for Bypass field were undertaken as part of a project funded by BBSRC Grant D20191, using data collected in another BBSRC-funded project (Grant 204/11563). Thanks also to Dr Rob Bramley for providing the comments that helped shape the final version.

References

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht, The Netherlands: Kluwer.

Anselin, L., Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. *American Journal of Agricultural Economics*, *86*, 675–687.

Beckett, P. H. T., & Webster, R. (1971). Soil variability: a review. *Soils and Fertilizers*, *34*, 1–15.

Besag, J., & Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *61*, 691–746.

Bishop, T. F. A., & Lark, R. M. (2006). The geostatistical analysis of experiments at the landscape-scale. *Geoderma*, *133*, 87–106.

Bishop, T. F. A., & Lark, R. M. (2007). A landscape-scale experiment on the changes in available potassium over a winter wheat cropping season. *Geoderma*, *141*, 384–396.

Boyd, D. A., Yuen, Lowsing, T. K., & Needham, P. (1976). Nitrogen requirement of cereals. 1. response curves. *Journal of Agricultural Science, Cambridge*, *87*, 149–162.

Bramley, R. G. V., Cook, S. E., Adams, M. L., & Corner, R. J. (1999). *Designing your own experiments: how precision agriculture can help*. Canberra, Australia: Grains Research and Development Corporation.

Brus, D. J., & de Gruijter, J. J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). *Geoderma*, *80*, 1–44.

Bruulsema, T. W., Malzer, G. L., Robert, P. C., Davis, J. G., & Copeland, P. J. (1996). Spatial relationships of soil nitrogen with corn yield response to applied nitrogen. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision Agriculture: Proceedings of the 3rd International Conference* (pp. 505–512). Madison, WI: ASA, CSSA, SSSA.

Christakos, G. (2000). *Modern spatiotemporal geostatistics*. New York: Oxford University Press.

Christakos, G., & Li, X. (1998). Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators. *Mathematical Geology, 130*, 435–462.

Cook, S. E., & Bramley, R. G. V. (2000). Coping with variability in agricultural production-implications for soil testing and fertiliser management. *Communications in Soil Science and Plant Analysis*, *31*, 1531–1551.

Cook, S. E., Adams, M. L., & Corner, R. J. (1999). On-farm experimentation to determine site-specific responses to variable inputs. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision Agriculture: Proceedings of the 4th International Conference* (pp. 611–621). Madison, WI: ASA, CSSA, SSSA.

Davis, J. G., Malzer, G. L., Copeland, P. J., Lamb, J. A., & Robert, P. C. (1996). Using yield variability to characterize spatial crop response to applied N. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision Agriculture: Proceedings of the 3rd International Conference* (pp. 513–519). Madison, WI: ASA-CSSA-SSSA.

de Gruijter, J. J., Brus, D. J., Bierkens, M. F. P., & Knotters, M. (2006). *Sampling for natural resource monitoring*. Berlin: Springer.

Diggle, P. J., & Ribeiro, Jr, P. J. (2007). *Model-based geostatistics*. New York: Springer.

Doerge, T. A., & Gardner, D. L. (1999). On-farm testing using the adjacent strip comparison method. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision Agriculture: Proceedings of the 4th International Conference* (pp. 603–609). Madison, WI: ASA, CSSA, SSSA.

Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural research*. New York: Wiley.

Goovaerts, P. (1997). *Geostatistics for natural resource evaluation*. New York, NY: Oxford University Press.

Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, *54*, 187–211.

Hurley, T. M., Oishi, K., & Malzer, G. L. (2005). Estimating the potential value of variable rate nitrogen applications: a comparison of spatial econometric and geostatistical models. *Journal of Agricultural and Resource Economics*, *30*, 231–249.

Lambert, D. M., Lowenberg-DeBoer, J., & Bongiovanni, R. (2004). Comparison of four spatial regression models for yield monitor data: a case study from Argentina. *Precision Agriculture*, *5*, 579–600.

Lark, R. M., & Cullis, B. R. (2004). Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*, *55*, 799–813.

Lark, R. M., & Wheeler, H. C. (2003). A method to investigate within-field variation of the response of combinable crops to an input. *Agronomy Journal*, *95*, 1093–1104.

McBratney, A. B. (1985). The role of geostatistics in the design and analysis of field experiments with reference to the effect of soil properties on crop yield. In D. R. Nielsen, & J. Bouma (Eds.), *Soil Spatial Variability: Proceedings of a Workshop of the ISSS and the SSSA. Las Vegas, USA. 30th November–1st December 1984* (pp. 3–8). Wageningen, The Netherlands: Pudoc.

Mercer, W. B., & Hall, A. D. (1911). The experimental error of field trials. *Journal of Agricultural Science, Cambridge*, *4*, 107–127.

Minasny, B., & McBratney, A. B. (2007). Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma*, *140*, 324–336.

Pannell, D. J. (1998). On the estimation of on-farm benefits of agricultural research. *Agricultural Systems*, *61*, 123–134.

Panten, K., Bramley, R. G. V., Lark, R. M., & Bishop, T. F. A. (2010). Enhancing the value of field experimentation through whole-of-block designs. *Precision Agriculture*, *11*, 18–213.

Papritz, A. (2008). Standardised vs. customary ordinary cokriging: some comments on the article "The geostatistical analysis of experiments at the landscape-scale" by T.F.A. Bishop and R.M. Lark. *Geoderma*, *146*, 391–396.

Pardo-Igúquiza, E., & Dowd, P.A. (1997). AMLE3D: A computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. *Computers & Geosciences*, *23*, 793–805.

Pardo-Igúzquiza, E., & Dowd, P. A. (1998). Maximum likelihood inference of spatial covariance parameters of soil properties. *Soil Science*, *163*, 212–219.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*, 545–554.

Peace, G. S. (1993). *Taguchi methods*. Boston, MA: Addison-Wesley.

Petersen, R. G. (1994). *Agricultural field experiments: design and analysis*. New York, NY: Marcel Dekker.

Pringle, M. J., McBratney, A. B., & Cook, S. E. (2004). Field-scale experiments for site-specific crop management. Part II: a geostatistical analysis. *Precision Agriculture*, *5*, 625–645.

R Development Core Team. (2008). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (URL: http://www.R-project.org; last accessed 19.8.2009).

Reetz, H. F. (1996). On-farm research opportunities through site-specific management. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Precision Agriculture: Proceedings of the 3rd International Conference* (pp. 1173–1176). Madison, WI: ASA-CSSA-SSSA.

Ribeiro, Jr, P. J., & Diggle, P. J. (2001). geoR: A package for geostatistical analysis. *R-NEWS,* 1, 15–18 (URL: http://cran.r-project.org/doc/Rnews/Rnews_2001–2.pdf; last accessed 19.8.2009).

Russell, E. W. (1976). *Soil conditions and plant growth* (10th ed.). London, UK: Longman.

Stein, M. L., Chi, Z., & Welty, L. J. (2004). Approximating likelihood methods for large spatial data sets. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *66*, 275–296.

Taylor, J. A., McBratney, A. B., & Whelan, B. M. (2007). Establishing management classes for broadacre agricultural production. *Agronomy Journal*, *99*, 1366–1376.

Van Groenigen, J. W., & Stein, A. (1998). Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality*, *27*, 1078–1086.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester, UK: Wiley.

Whelan, B. M., & McBratney, A. B. (2002). A parametric transfer function for grain-flow within a conventional combine harvester. *Precision Agriculture*, *3*, 123–134.

Whelan, B. M., McBratney, A. B., & Stein, A. (2003). On-farm field experiments for precision agriculture. In J. V. Stafford, & A. Werner (Eds.), *Precision Agriculture: Papers from the 4th European Conference on Precision Agriculture, Berlin, Germany, 15–19 June 2003* (pp. 731–737). Wageningen, the Netherlands: Wageningen Academic Publishers.

Chapter 11
Application of Geostatistical Simulation in Precision Agriculture

R. Gebbers and S. de Bruin

Abstract Geostatistical simulation provides a means to mimic spatial and or temporal variation of processes that are relevant to precision agriculture. Simulation by computer models aids decision making when it is too difficult, time consuming, costly or dangerous to perform real-world experiments. Spatio-temporal processes are often considered as uncertain because it is impossible to make accurate and comprehensive observations. Geostatistical simulation incorporates uncertainty into modelling to obtain a more realistic impression of the variation. This chapter provides a short introduction to the background of geostatistical simulation and explains sequential Gaussian simulation in more detail because it is the method most commonly applied. Three case studies demonstrate the application of geostatistical simulation in precision agriculture. They deal with the risk of under- and over-liming because of uncertainty about the accuracy of a pH map, the economic costs of GPS errors and the identification of factors that are most relevant to the accuracy of mapping.

Keywords Sequential Gaussian simulation · Stochastic modelling · Conditional simulation · Unconditional simulation · Scenarios · Spatio-temporal variability · Smoothing · Positional error · Error propagation · Error model · Prediction error · Uncertainty · Risk assessment

R. Gebbers (✉)
Department of Engineering for Crop Production, Leibniz-Institute for Agricultural Engineering, Max-Eyth-Allee 100, D-14469 Potsdam, Germany
e-mail: rgebbers@atb-potsdam.de

S. de Bruin
Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands
e-mail: sytze.deBruin@wur.nl

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_11,

11.1 Introduction

Geostatistical simulation is a specific form of stochastic simulation that addresses spatial and or temporal uncertainty. Simulation can be considered a part of modelling. According to Shiflet and Shiflet (2006, p. 6), "modeling is the application of methods to analyse complex, real-world problems in order to make predictions about what might happen with various actions." Simulation aims to imitate reality with hypothetical situations; in scientific situations this is usually done by the application of a given model with certain inputs. Simulation by computer models is of particular value when it is too difficult, time consuming, costly or dangerous to perform real-world experiments. They enable certain situations and alternatives to be examined to aid decision making (Shiflet and Shiflet 2006). By simulating a process, one can consider various scenarios and test the effect of each. Deterministic models produce the same results with a given input, whereas stochastic models generate different results from the same input. Both of these basic types of model can be combined. However, the output of a combined model will remain random. By incorporating the element of chance into modelling it is possible to obtain more realistic simulation results.

Geostatistical simulation is used to reproduce spatial and sometimes temporal variation and uncertainty. In this respect it differs from kriging which is intended to produce optimum, albeit smoothed, predictions. Variation is present everywhere in the environment – agricultural fields are no exception. Uncertainty results from the fact that we cannot measure entire populations or domains (like fields) and that data can suffer from measurement errors. The simulation of spatial variation and uncertainty can be useful for understanding the effects of certain precision agriculture (PA) techniques and to identify components of PA that require improvement. For example, a farmer may want to assess the risk of applying the wrong fertilizer rate because of uncertainty about the accuracy of a soil nutrient map (Section 11.2), to quantify the benefits of using more accurate GPS for field boundary survey (Section 11.3) or to determine whether sampling density or laboratory error has the greater impact on the accuracy of predictions for mapping (Section 11.4).

The first part of this chapter provides a short introduction to geostatistical simulation and focuses on the most commonly applied method. The rest of the chapter includes three case studies to illustrate the application of geostatistical simulation in the context of precision agriculture. For those who want to learn more there are two textbooks that provide excellent practical introductions to simulation by Deutsch and Journel (1998) and Remy et al. (2009). Both books have associated software: GSLIB by Deutsch and Journel (1998) is a collection of Fortran routines and SGeMS by Remy et al. (2009) is user-friendly software with a graphical interface. There are more mathematical books by Chilès and Delfiner (1999) and Lantuéjoul et al. (2002); the first contains examples mainly from geology and the mining industry, whereas the latter is entirely theoretical. Goovaerts (1997) includes a large chapter on geostatistical simulation, which is recommended as an

introduction. Other introductory books with shorter chapters on simulation are Webster and Oliver (2007) and Leuangthong et al. (2008). Even though geostatistical simulation has a value of its own, it should be seen in the broader context of error propagation. Heuvelink's comprehensive book on error propagation in spatial modelling (Heuvelink 1998) integrates these objectives.

11.1.1 Basics of Geostatistical Simulation

Geostatistical simulation was established more than 30 years ago (Remy et al. 2009). Algorithms for geostatistical simulation were developed for a probabilistic assessment of 2-D and 3-D representations of spatial processes, but they can be extended easily to temporal or spatio-temporal processes. In geostatistical simulation random numbers are drawn from a predefined probability distribution (the stochastic model). The purpose is to generate equally probable realizations of a process described by functions of spatial dependence, i.e. the variogram, and the cumulative frequency distributions of the variables under consideration.

The most common methods of geostatistical simulation are based on the variogram, however, some methods do not use it to describe spatial dependence (see later in this section). Simulation can be unconditional or conditional; the latter integrates existing knowledge into the simulation. In this case, the simulation should honour observations at the sampling points. Geostatistical simulation can be applied in all four dimensions provided that the generating function (variogram) is conditional negative semidefinite for the relevant dimensions. By generating one-dimensional realizations (Figs. 11.1f and 11.2f), for example, autocorrelated time-series can be simulated. We return to this in Section 11.3. The typical outcome of a geostatistical simulation is a map, sometimes called a stochastic image. Figure 11.1a–c shows three realizations resulting from unconditional geostatistical simulation, whereas those in Fig. 11.2a–c result from conditional geostatistical simulation in which data values at given locations are honoured. The unconditional simulation results in Fig. 11.1 are based on the frequency distribution function of pH, the variogram and the field dimensions of the case study in Section11.2. The maps in Fig. 11.1 show different patterns of variation as do the transects (Fig. 11.1f), which are the values along the row at a y-position of 316 m. The frequency distributions (kernel density plot, Fig. 11.1d) and variograms (Fig. 11.1e) of each realization appear fairly similar. This is how it should be because each realization honours the predefined statistical model given by the frequency distribution and the variogram, but the variation in the maps and outcomes at single locations are random. The frequency distribution at a given location can be assessed by evaluating the pixel values there from a large number of simulations.

The maps of conditional simulation of the pH data in Fig. 11.2, based on the same frequency distribution and variogram as above, appear much more similar to one another than do those in Fig. 11.1f. In particular, readers should take a closer

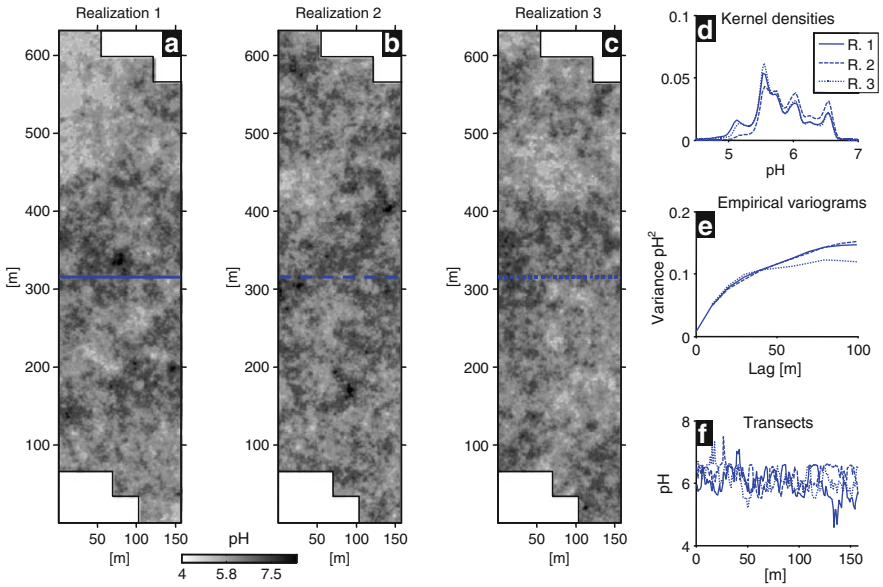

Fig. 11.1 (**a**) to (**c**) Three examples of unconditional geostatistical simulation based on a non-normal distribution and an omni-directional exponential variogram function with a nugget, c_0, of 0.018, a spatially dependent component, c, of 0.106 and an approximate range, a', of 65.5 m (a' is three times the distance parameter of the model), (**d**) the respective frequency distributions, (**e**) empirical variograms and (**f**) values from transects along the 316 m y-coordinate

look at the transect plots in Figs. 11.1f and 11.2f. The transects were placed to cross five conditioning data on the 316 m y-coordinate (see the line on maps in Figs. 11.1 and 11.2a–c). The pH values at these points are marked by dots on the transect plot. Even though the values vary considerably, the conditioning data are honoured by the line passing through them. The frequency distributions of the conditional simulations are similar (Fig. 11.2d), whereas the variograms (Fig. 11.2e) are more different from each other than are those in Fig. 11.1e.

As geostatistical simulation has not been widely used in PA, the reader might ask what are the differences between kriging and simulation, and what are the benefits? Kriging is intended to predict values of the target variable at unobserved locations, whereas geostatistical simulation addresses the variance. Kriging is a best linear unbiased predictor (BLUP), which tends to smooth values at unobserved locations, i.e. it overestimates values that are smaller than average and underestimates those that are larger. As a consequence, the global variance of the kriged estimates is much less (about half) than the original variance; in other words kriging loses variance (see Webster and Oliver 2007 for the mathematical proof). Stochastic simulation, however, retains the global variance and reflects the variation present, but the value simulated at a given location is, unlike kriging, not intended to give an accurate prediction. Instead, repeated simulations at a given location can reveal

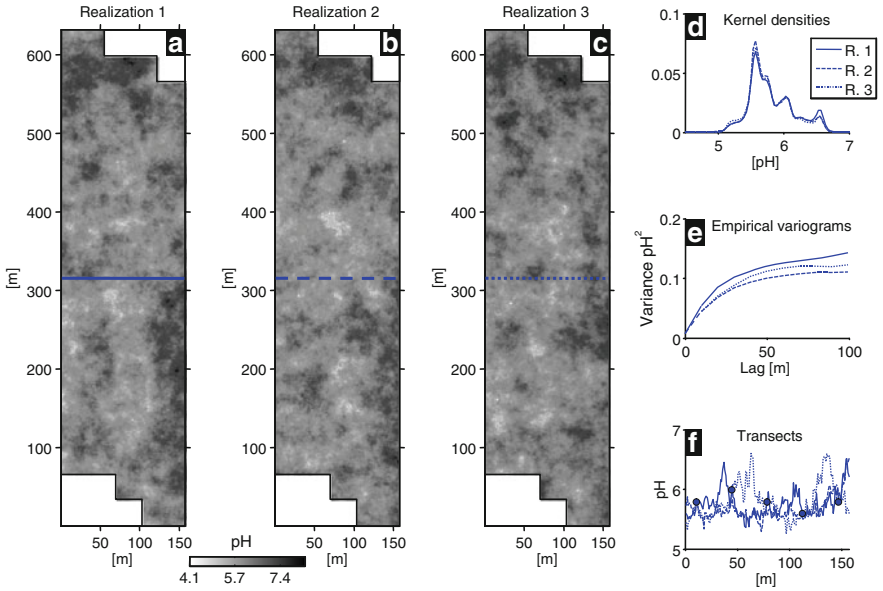

Fig. 11.2 (**a**) to (**c**) Three examples of conditional geostatistical simulation defined by a non-normal distribution and an omni-directional exponential variogram function with a nugget, c_0, of 0.018, a spatially dependent component, c, of 0.106 and an approximate range, a', of 65.5 m (a' is three times the distance parameter of the model), (**d**) the frequency distributions, (**e**) empirical variograms and (**f**) values from transects along the 316 m y-coordinate

the local probability distribution and the moments that can be derived from that. Reproduction of the local and global variance is essential for the two main uses of geostatistical simulation: uncertainty analysis of interpolated maps and error propagation in complex models that use spatial or temporal data or a combination of both. Other applications include scaling issues such as change of support and inverse modelling in which input data can be estimated from a given model and the output data (Chilès and Delfiner 1999; Goovaerts 1999).

Section 11.2 gives an example of conditional simulation for uncertainty analysis of interpolated values. Complex modelling with spatio-temporal input data can use conditional and or unconditional geostatistical simulation; examples are given in Sections 11.3 and 11.4. Other applications in agriculture are uncertainty in crop yield response (Faechner et al. 1999), the effect of soil compaction on corn yield (Lapen et al. 2001), design of field trials (Fagroud and Van Meirvenne 2002), soil erosion (Favis-Mortlock et al. 2000), decision support (Pokrajac et al. 2002) and the irrigation of maize (Zanolin et al. 2007).

The lack of user-friendly software and its poor integration into GIS might explain why simulation has been little used in PA. However, SGeMS (Remy et al. 2009) makes the application of geostatistical simulation more attractive to practitioners. Examples of software for simulation are given in the Appendix to this book.

11.1.2 Theory

11.1.2.1 Spatial Random Variable and Spatial Random Function

Geostatistical simulation is based on the same assumptions about random processes as kriging (see Chapter 1). A spatial random function $Z(\mathbf{x})$ could be characterized by its multivariate distribution function, Prob$\{Z(\mathbf{x}), \mathbf{x} \in S\}$. However, an analytical description of such a multivariate distribution is usually impractical because there might be parameters that vary from one location to another (Remy et al. 2009, p. 34). Thus, several assumptions are necessary to simplify the modelling of the joint distribution by the spatial random function. For example, instead of modelling the probability function of all unknowns simultaneously this can be done individually for each $z_i(\mathbf{x})$, step by step. This is exactly how the sequential simulation algorithm works. For in depth discussion of the random function and the joint distribution the reader is referred to Goovaerts (1997) and Remy et al. (2009).

11.1.2.2 Stochastic Simulation

Stochastic simulation is the process of drawing random numbers from a predefined probability distribution. This kind of simulation is part of the Monte Carlo approach, which aims to solve mathematical problems by modelling random variables. Stochastic simulation requires the generation of a large set of random numbers, which has only become feasible with the advent of modern computers. However, random numbers generated by computer are so called 'pseudo-random numbers' because they are not truly random. They are usually generated by sequential congruential algorithms, which produce sequences that depend on an initial value known as the 'seed'. Using the same seed will reproduce the same sequence of numbers; therefore, experiments based on pseudo-random number generators are repeatable. Pseudo-random numbers can reproduce target distributions quite well, and cannot be distinguished from real random samples by common statistical tests. Readers are referred to Gentle (1989) for more information about pseudo random numbers and Monte Carlo methods.

11.1.2.3 Overview of Methods for Geostatistical Simulation

There are several methods of geostatistical simulation and sometimes different names are used for the same method. Thus, some categorization may help to distinguish the different algorithms and to identify the appropriate method for the problem at hand. We follow the classification given by Vann et al. (2002) with some additions by Deutsch and Journel (1998), Chilès and Delfiner (1999) and Remy et al. (2009). Detailed explanations of the simulation methods are given in the last three references and in the literature cited herein.

- Pixel-based methods (variogram-based or two-point simulation methods)

In pixel-based methods simulations are generated point by point (or pixel by pixel for simulating raster data). They are conditioned by two-point statistics as modelled by the variogram or covariance functions. Pixel-based methods can be used for continuous and categorical variables. Among these methods Gaussian-based methods are the most popular (multiGaussian approach). They rely on Gaussian random functions and the gradual transition between high and low values. Sequential Gaussian, covariance matrix decomposition, turning bands, truncated Gaussian and pluriGaussian methods belong to this class. Non-parametric pixel-based methods do not require a Gaussian or some other distribution, e.g. sequential indicator simulation is a non-parametric method.

- Object-based methods

Object-based methods simulate the occurrence of objects and their properties at random points. Point processes or point patterns deal with the presence or absence of objects, sometimes called 'events'. An example would be the occurrence of weeds. Boolean simulation is an extension of point process modelling. Here the object's geometry is attached to the points obtained by point pattern simulation. An example might be the distribution and orientation of leaves in a crop's canopy.

- Multi-point methods

Multi-point methods combine pixel- and object-based methods. They operate pixel-wise and conditional probabilities for each pixel value are obtained as conditional proportions from a training image that depicts the geometry and distribution of objects expected in reality (Remy et al. 2009). Multi-point methods can include soft data. An application for PA would be the simulation of soil types with soil horizons as categorical variables.

Pixel-based simulation algorithms should be the preferred option when it is important to reproduce local data. However, such two-point methods can only reproduce the variogram and not shapes and patterns. Object-based simulation algorithms are ideal for creating maps (or images) with strong spatial structures and patterns, but they are notoriously difficult to condition to local data.

11.1.3 Sequential Gaussian Simulation

Sequential Gaussian simulation is the most straightforward algorithm for the simulation of continuous variables. In the case of conditional simulation we have k conditioning data $Z(\mathbf{y}_k)$ and m nodes at which we want to simulate a regionalized variable $Z_{Sn}(\mathbf{x}_m)$. The $m = 1, \ldots, l$ nodes will be visited in a sequence $n = 1, \ldots, l$. In the case of unconditional simulation the algorithm will start with a random value at a random location. The procedure for two-dimensional simulation is presented as a flowchart in Fig. 11.3.

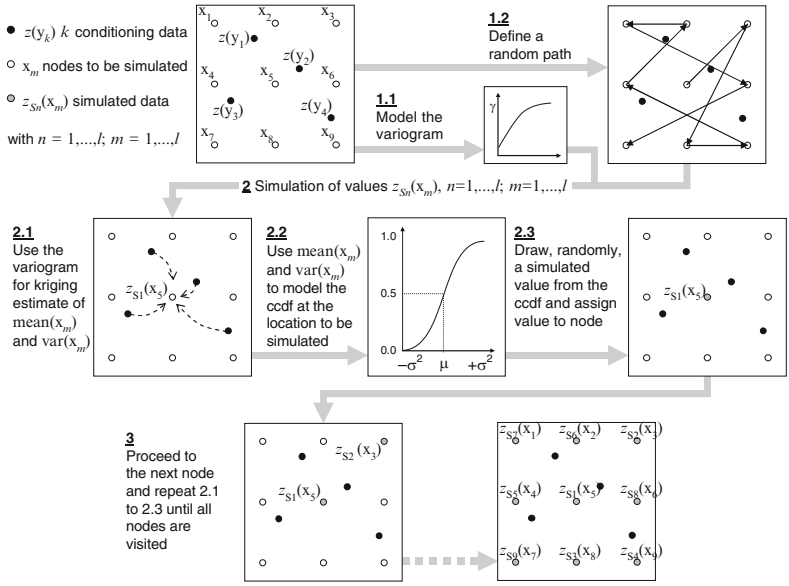

Fig. 11.3 Flowchart of sequential Gaussian simulation in a two-dimensional domain

The procedure is described as follows:

0. Pre-processing (can be omitted in the case of unconditional simulation): examine conditioning data $Z(\mathbf{y}_k)$ for normality. Transform $Z(\mathbf{y}_k)$ to a standard normal distribution if necessary.
1. Initial definitions
   1.1. Definition of a variogram: compute and model the experimental variogram from the normally distributed conditioning data $Z(\mathbf{y}_k)$ or select a model and parameters if there are no conditioning data.
   1.2. Definition of a random path: create a path $n = 1, \ldots, l$ which passes through all nodes $\mathbf{x}_m$, where n is ordered randomly.
2. Simulation at station n
   2.1. Kriging: use simple or ordinary kriging with the variogram model at a node $\mathbf{x}_m$ to estimate $\hat{Z}(\mathbf{x}_i)$ and the variance, $\sigma^2_{\mathrm{K}}(\mathbf{x}_i)$. The estimate will be based on conditioning data $Z(\mathbf{y}_k)$ and previously simulated data $Z_{\mathrm{S}(1\ldots n-1)}(\mathbf{y}_k)$ within the search radius.
   2.2. Model a conditional cumulative distribution function (ccdf): use the kriged estimate and variance to form a Gaussian cumulative distribution function.
   2.3. Assign a random number: generate a pseudo-random number from the ccdf and assign this number to $Z_{Sn}(\mathbf{x}_m)$. Proceed to station n + 1 on the path.

3. Repeat steps 2.1 to 2.3 until all nodes have been visited.
4. Post-processing (optional): back-transform values if they are required on the original scale of measurement.

Sequential Gaussian simulation can produce large fields of values because it is not constrained to keeping the covariance matrix of the entire domain in memory. It reduces the demand on memory by limiting the neighbourhood using a specified search radius and or a limited number of points to be included in the kriging system. In this case, the covariance is only approximated. Thus, the neighbourhood should be large enough otherwise reproduction of the variogram will be poor. It is recommended that the field being simulated is larger than the range of the variogram (Vann et al. 2002). The approach can be extended easily to indicator kriging.

11.1.4 Transformation of Probability Distributions

Regionalized variables are sometimes non-normal. A normal distribution is required for Gaussian simulation described above; this is usually achieved by a transformation to normal scores so that the target cdf is a standard normal distribution with zero mean and unit variance (Fig. 11.4). Normal scores transformation and the reverse, back-transformation, are critical steps in Gaussian simulation methods.

Common methods for transformation and back-transformation are described in Deutsch and Journel (1998) and Remy et al. (2009). These methods include interpolation to estimate probabilities where no raw values are available and extrapolation for cases where the sample data may not cover the full range of possible values. We demonstrate this in Section 11.2.2. In particular, the lower and upper tail of the distribution can be modelled by extrapolation with various functions (typically power, exponential or hyperbolic, see Remy et al. 2009). This is a critical step because we must make assumptions about the minimum and maximum values. Alternatively, Bourgalt (1997) describes an extension of sequential simulation for non-Gaussian distributions that avoids transformation and back-transformation.

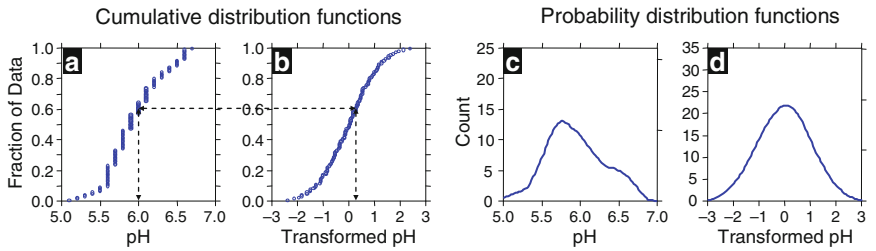

Fig. 11.4 Example of non-parametric transformation to normality and back-transformation using the pH test data from Section 11.2. Distribution functions of the raw data (**a, c**) and of the transformed data (**b, d**)

11.2 Case Study I: Uncertainty of a pH Map

Assessment of interpolation error in maps is one of the classical applications of geostatistical simulation. Nevertheless, it has rarely been used for this purpose in PA. This section describes an application of simulation and its benefits with a practical example.

11.2.1 Introduction

In PA, maps of pH are required for making site-specific decisions on the application of lime to correct soil acidity. Soil pH affects the availability of most plant nutrients, which can be reduced where pH is either too low or too high (MAFF 2000). The optimum pH of a soil depends on soil texture, organic matter content and on the crop grown. In many countries ranges of optimum soil pH are published by official advisory boards (e.g. in the U.K. by the Department for Environment, Food, & Rural Affairs (DEFRA) or in Germany by the 'Verband Deutscher Landwirtschaftlichen Untersuchungs- und Forschungsanstalten' (VDLUFA)). For economic reasons, a farmer's field cannot be sampled exhaustively to map the pH and so values are predicted from a few samples by kriging or some other method of interpolation. The predicted values are in error because they are based on a sample and so there is a risk of under- or over-liming. This kind of uncertainty in the predictions can be addressed by geostatistical simulation. The objectives of this case study are to:

- Compare the results of kriging with those of simulation.
- Determine the accuracy of an interpolated map of pH.
- Derive probabilities that a pH value is below or above the optimum.

11.2.2 Materials and Methods

The sampling site is on a commercial farm at Kassow in northeast Germany, about 30 km south of the Baltic Sea (lat. 53.8671°, long. 12.0697°). The landscape was formed during the last ice age about 10 000 years ago and by post-Quaternary processes. As a typical ground moraine area, the sampling site is undulating, has a predominantly sandy soil and spatially variable soil properties.

Soil samples were collected within a sampling frame of 137 × 612 m on a 73 ha field (KSG 111). A basic sampling grid with a spacing of 30-m was established related to the average distance between the tramline-system of the field (Fig. 11.5a). Samples were taken to one side of the tramlines, and were georeferenced by a high precision RTK-dGPS. There were 85 sampling points and these will provide the test

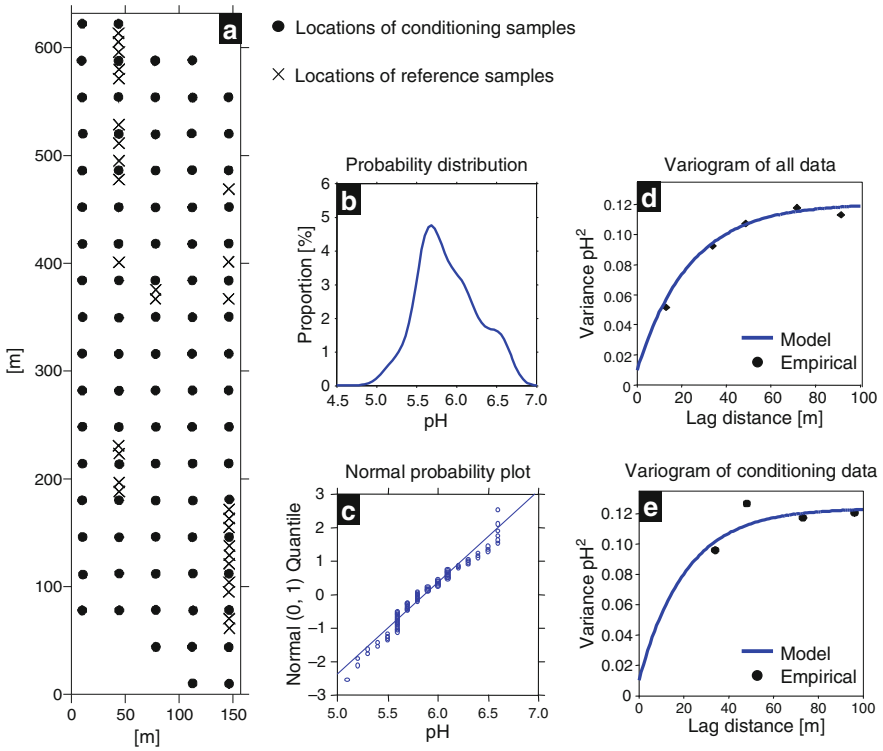

Fig. 11.5 (**a**) Sampling design, (**b, c**) plots of the statistical distribution of all observed pH values and (**d, e**) experimental variogram and model of all data and of the conditioning data, respectively

data set in this study. In addition, 27 samples were taken at irregular intervals based on a map of apparent soil electrical conductivity (EC_a). Locations were chosen to cover a representative range of EC_a values and to obtain information at small sampling intervals, which is important for variogram modelling. For convenience, global coordinates of the sampling points were transformed into a local Cartesian system with the sampling grid arranged in parallel to the x- and y-axis. Sampling depth was 30 cm, which is approximately equivalent to the depth of the plough horizon A_p. At every location, six sub-samples were taken within a radius of about 1 m and bulked to obtain 0.5 to 1 kg of soil. The samples were analysed for pH by a glass electrode in a suspension of 10 g soil and 25 ml of 0.01 M CaCl solution (VDLUFA 2000a).

Data analysis included summary statistics, variography, sequential geostatistical simulation and post-processing of the simulation results. In the preceding analysis, summary statistics and Kolgomorov–Smirnov tests for normality were done with SYSTAT (Systat Software Inc.). Transformation of the conditioning data to normality and back-transformation of simulated values (Fig. 11.4) were done by the SGeMS function 'trans'. We have used the power function, which produced the best

results according to the statistical tests for normality (data not shown). The lower tail of the cdf was approximated between the minimum $z_{\min}$ and the first threshold z_1 by:

$$\left(\frac{z_1 - z}{z_1 - z_{\min}}\right)^{\omega} \quad \forall\, z \in (z_{\min}, z_1), \tag{11.1}$$

where ω controls the decrease of the function, with the constraint $\omega \geq 1$. The greater is ω, the less likely are small values close to $z_{\min}$ (Remy et al. 2009, p. 106).

The upper tail was modelled by:

$$\left(\frac{z - z_L}{z_{\max} - z_L}\right)^{\omega} \quad \forall\, z \in (z_L, z_{\max}), \tag{11.2}$$

where z_L is the upper threshold and $z_{\max}$ is the maximum. Parameter ω controls the decrease of the function, with the constraint $\omega \in [0, 1]$. The smaller is the value, the less likely are extreme values close to $z_{\max}$ (Remy et al. 2009, p. 106).

Parameters for the approximation of the normal scores for the lower tail were $z_{\min} = -3$, $\omega = 3$ and were $z_{\max} = +3$, $\omega = 0.333$ for the upper tail. For approximation to the distribution of the raw data, the parameters were set to: 0 pH, $\omega = 3$ for the lower tail and to +14 pH, $\omega = 0.333$ for the upper tail.

The experimental omnidirectional variogram was modelled by ordinary least squares approximation. Kriging and simulation were done on a 1-m grid using SGeMS. The kriging standard errors were obtained for comparison with other results. Conditional sequential Gaussian simulation (Section 11.1.3) was used to create 100 realizations of pH from the normal score transform (SGeMS). The search radius was set to 100 m and a maximum of 100 conditioning points. After back-transformation the statistical moments of the simulated data were calculated and mapped in MATLAB (The MathWorks Inc.). Kernel density estimation was used instead of a histogram to describe the frequency distribution (Härdle et al. 2004).

11.2.3 Results and Discussion

There is little difference between the descriptive statistics for all of the data and the conditioning data (Table 11.1). The median pH is 5.9 and 5.8, respectively. This value is at the lower limit of the optimum pH range for loamy sand, which is between pH 5.8–6.3 for all crops according to VDLUFA (2000a, b). The minimum pH in the data is 5.1, which indicates a need for liming, whereas the maximum is 6.7, which is higher than it should be on loamy sand. The kernel density and the normal probability plots (Fig. 11.5b, c) show that the distribution is slightly skewed and that it may also be bimodal. The skewness coefficients are given in Table 11.1. The statistical test indicates significant deviation from normality (Table 11.1). Thus, we decided to transform the raw data before proceeding with conditional Gaussian simulation.

Table 11.1 Descriptive statistics of all data and conditioning data for pH

	All data	Conditioning data
Number of cases	116	87
Minimum	5.1	5.1
Maximum	6.7	6.6
Range	1.6	1.5
Median	5.9	5.8
Arithmetic mean	5.93	5.89
Standard deviation	0.36	0.37
Variance	0.13	0.13
Skewness	0.245	0.322
Kurtosis	−0.466	−0.492
Kolmorogov–Smirnow p-value	0.001	0.001

Table 11.2 Parameters of omnidirectional variogram models fitted to all data and to the conditioning data for pH before and after normal scores transformation

Data	Variogram model	Nugget c_0	Spatially dependent component c	Effective range a' [m]
All data, no transformation	Exponential	0.0100	0.1107	70.6
Conditioning data, normal scores	Exponential	0.0414	0.8641	59.6

a' is three times the distance parameter of the exponential model.

Experimental variograms were computed and modelled for the complete set of raw and conditioning data before and after transformation to normal scores. Figure 11.5d, e shows the variogram for these, and Table 11.2 gives the parameters of the variogram models. The fit of the model for the conditioning data was constrained to the nugget variance derived from the complete data because the separating intervals of the conditioning data are too wide to obtain reliable estimates of the variogram near the origin.

11.2.3.1 Simulation

Kernel density, quantile–quantile and experimental variograms of three realizations are plotted to assess how well the simulations reproduce the conditioning data in Fig. 11.6. Kernel density curves show some differences at the peak near the mean and at the shoulder around pH 6.5 (Fig. 11.6a). They are, however, all close enough for them to be assumed to come from the same distribution. The similarity between the conditioning data and the realizations is confirmed by the quantile–quantile plot overlaid by the 1:1 line (Fig. 11.6b). The variogram model of the conditioning

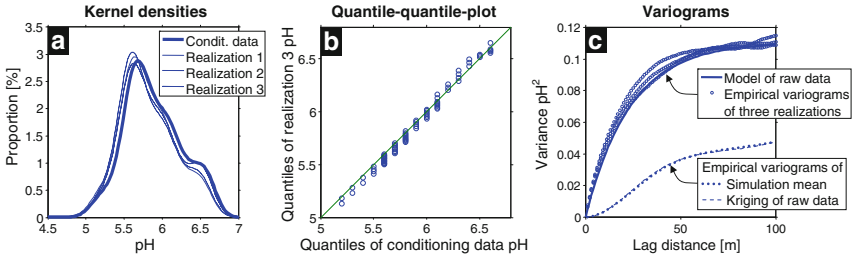

Fig. 11.6 Evaluation of simulation results: (**a**) comparison of kernel densities from three simulations and the conditioning data, (**b**) quantile–quantile-plot of a simulation versus the conditioning data and (**c**) comparison of the variogram model from the conditioning data with empirical variograms from three simulations, the mean of all runs of simulation and kriging of raw conditioning data (note that the empirical variograms of the simulation mean and kriging are almost identical)

data and the empirical variograms from these realizations are also very similar (Fig. 11.6c), which indicates that variation in the conditioning data was reproduced sufficiently by simulation.

11.2.3.2 Comparison of Kriging and Simulation

Ordinary kriging was done with the raw conditioning data and the parameters of the variogram model given in Table 11.2. A comparison of the empirical variograms from the three realizations with that from the kriged predictions in Fig. 11.6c demonstrates how kriging smoothes. The empirical variogram of the kriged predictions approaches its sill at about half that of the empirical variogram of the simulated realizations. The variogram of kriged predictions also has a longer range. However, if simulations are averaged over a large number of realizations variance is also lost and the shape of the empirical variogram becomes identical to that of the kriged predictions. Hence, only single realizations from geostatistical simulation mimic the variation in the observations.

Maps of kriged predictions and standard deviations are shown in Fig. 11.7a, c. For further comparison of kriging with simulation the realizations have to be summarized. This is done by so called E-type estimates (conditional expectation), which are point-wise statistics computed at each simulation point (e.g. the arithmetic mean or the variance) (Remy et al. 2009, p. 37). These E-type estimates were based on 100 realizations (Fig. 11.7b, d); they can be compared visually with their ordinary kriging equivalents (Fig. 11.7a, c).

The contour lines of the kriged map are smooth, whereas those for the E-type mean are irregular. Nevertheless, both maps show the same spatial structures, which is confirmed by the scatter plot in Fig. 11.7e. Table 11.3 gives the summary statistics of the kriged and E-type estimates, which show that overall the E-type mean is slightly smaller.

The maps of standard deviation (SD) from kriging and E-type estimation differ much more than the kriged predictions and E-type mean. The map of kriging

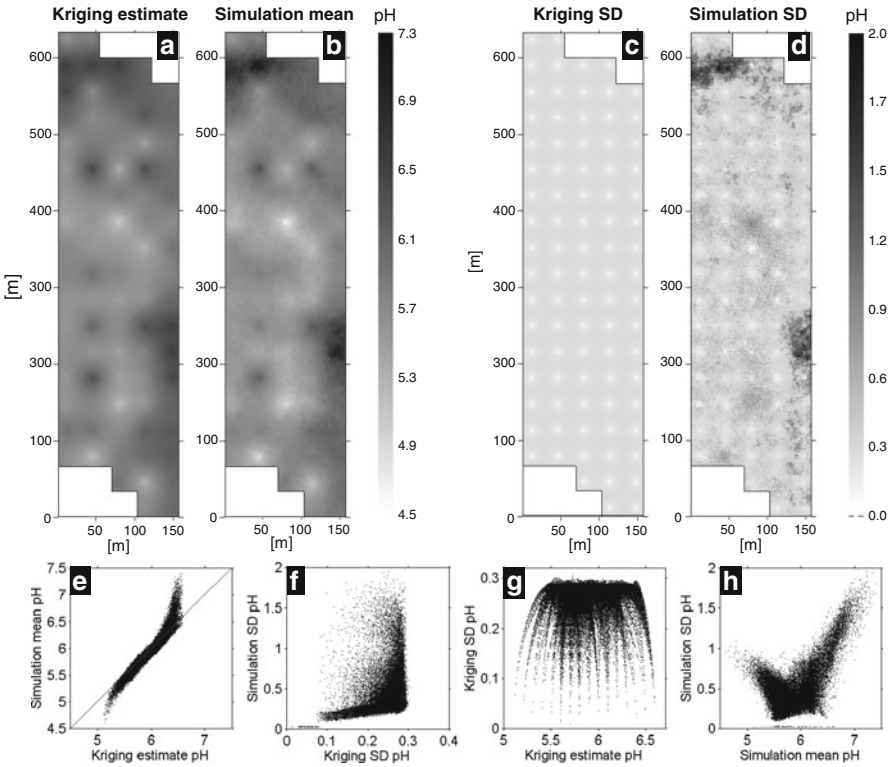

Fig. 11.7 Comparison of the results of kriging and simulation (the individual components of this figure (**a** – **h**) are explained in the text)

Table 11.3 Summary statistics of kriged predictions and E-type estimates from simulations of pH

	Kriged predictions	E-type mean	Kriging SD	E-type SD
Mean	5.87	5.84	0.26	0.40
Median	5.82	5.80	0.27	0.33
Variance	0.06	0.08	0.0009	0.04
Standard deviation (SD)	0.24	0.29	0.03	0.21
Minimum	5.11	4.61	0.01	0.02
Maximum	6.59	7.40	0.31	2.03
Range	1.48	2.79	0.30	2.01

SD (Fig. 11.7c) indicates that kriging standard errors (derived from the kriging variance) depend only on the distance from the sampling location; they are independent of the kriged predictions as confirmed by the scatterplot (Fig. 11.7g). By contrast, the plot of E-type SD versus E-type mean (Fig. 11.7h) shows a non-linear trend with a decrease from pH 5 to 5.5 and an increase between pH 6.3 and 6.7.

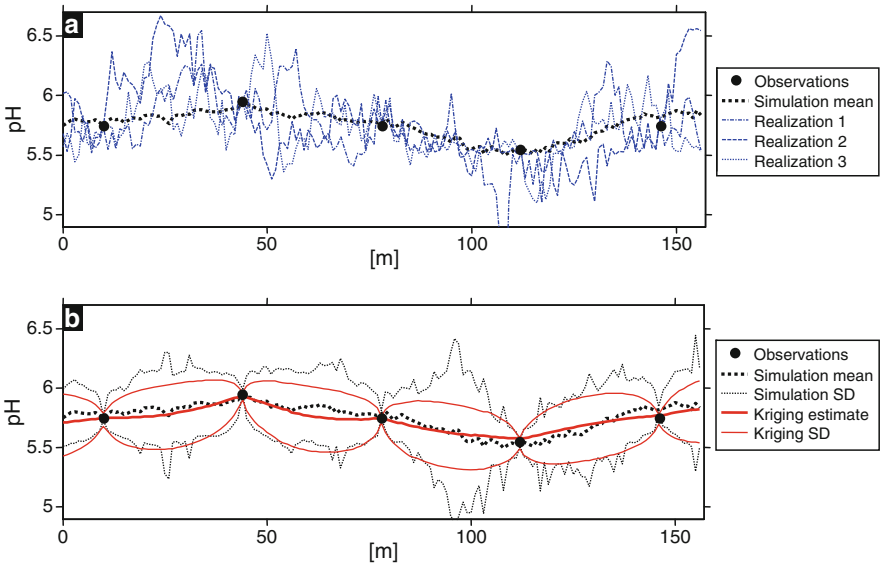

Fig. 11.8 Detailed results of kriging and geostatistical simulation from a transect along y = 316 m. (**a**) Average of 100 simulations and three examples of single simulations and (**b**) comparison of means and standard deviations (SD) from 100 simulations with the estimates and standard deviations from kriging

Values between pH 5.5 and 6.3 are associated with a smaller SD. Large SDs (dark areas) are associated with high pH values near the upper and right-hand borders (around y = 150–250 m), Fig. 11.7a and b, and where pH is low around x = 75 m, y = 150 m and x = 75 m, y = 375 m. The E-Type SD also (but not only) depends on distance from the conditioning points; there is a regular grid of small SDs (light values) near these points in Fig. 11.7d. The E-type SD is generally larger than the kriging SD (Table 11.3), the latter is limited to 0.31 pH units as can be seen in Fig. 11.7f, whereas the simulation SDs show more fluctuation.

Figure 11.8 provides a more detailed comparison of the kriging and simulation results along a transect that crosses five conditioning points. Figure 11.8a shows the large fluctuations in individual realizations of geostatistical simulation beyond the conditioning points, whereas Fig. 11.8b shows the smoothing by kriging. Although both methods honour the conditioning values they behave differently between the conditioning points. That is, the E-type means show more roughness than the kriged estimates.

11.2.3.3 Prediction Error of the Interpolated Map

To compare the estimation errors predicted by geostatistical stimulation and kriging we used the 27 samples of the validation data $z_{\text{val}}(\mathbf{x}_i)$. We obtained the true absolute estimation error of ordinary kriging by:

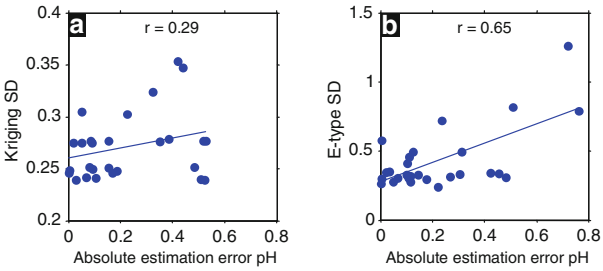

Fig. 11.9 Scatter plots of absolute estimation error of the validation data versus standard deviation from ordinary kriging (**a**) and geostatistical simulation (**b**) with the linear regression line and Pearson's coefficient of correlation, r

$$\left| z^*_{\mathrm{OK}}(\mathbf{x}_i) - z_{\mathrm{val}}(\mathbf{x}_i) \right|, \tag{11.3}$$

where $z^*_{\mathrm{OK}}(\mathbf{x}_i)$ are the kriged estimates at the locations of the validation data. The true absolute estimation error by geostatistical simulation is given by:

$$\left| z^*_{\mathrm{mean}}(\mathbf{x}_i) - z_{\mathrm{val}}(\mathbf{x}_i) \right|, \tag{11.4}$$

where $z^*_{\mathrm{mean}}(\mathbf{x}_i)$ is the E-type estimate of the mean at the locations of the validation data. These true estimation errors were compared with the kriging standard errors and the E-type SD, respectively. The results are summarized in the scatter plot of Fig. 11.9a, b. The linear regression line indicates a positive relationship between predicted and true errors for both kriging and simulation. The correlation coefficients suggest that error prediction by simulation ($r = 0.65$) is better than that by kriging ($r = 0.29$). However, the results are probably not as clear as expected because of the small number and an unfavourable dispersion of the validation points in this case study.

11.2.3.4 Probability that a pH Value is Outside the Optimal Range

Based on cumulative probability distributions from several simulated realizations, the probabilities of pH values being below or above a threshold can be calculated. The optimum range of pH for loamy sand is 5.8–6.3 (VDLUFA 2000a, b). The probabilities of pH being below 5.8 are shown in Fig. 11.10a. Areas with large probabilities require lime and are likely to show a positive response to its addition. The probability of exceeding pH 6.3 (Fig. 11.10b) highlights a few areas where no lime is required and could be counter productive. The farmer might consider decreasing pH in these areas by the use of acid fertilizers.

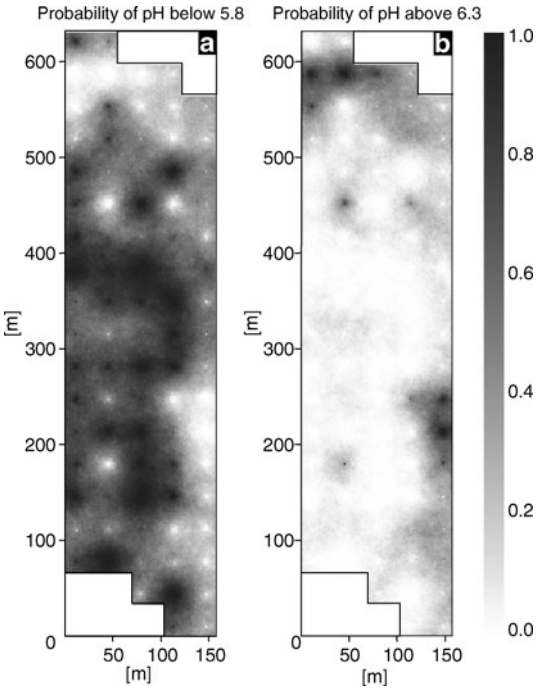

Fig. 11.10 Probabilities based on geostatistical simulation of pH being: (**a**) below 5.8 and (**b**) above 6.3

11.2.4 Summary and Conclusions

The kriging variances or standard errors are unsuitable indications of the errors in interpolated maps. Geostatistical simulation allows one to generate probability distribution functions for every pixel in a map by producing many equally probable realizations that honour the distribution and the variogram of the raw data. The E-type estimates derived from these realizations can be used to calculate parametric and non-parametric measures of dispersion, such as the mean, standard deviation, median, interquartile range and probabilities of having values below or above a certain threshold. This can be used in decision making for soil remediation.

There are other methods that can determine probabilities in relation to thresholds as described for sequential Gaussian simulation, such as disjunctive and indicator kriging or indicator simulation. Sequential Gaussian simulation is often preferred because it is easy to understand and implement, and because it can simulate large fields of values. Although the application of this method has become easier with the availability of user-friendly software such as SGeMS, there are problems that the user must be aware of. First, conditional simulation results cannot be better than the input data and the variogram derived from them. As with kriging, simulation methods based on two-point statistics (Section 11.1) rely on the quality of the variogram model. In Gaussian simulation the requirements for normality are stricter than for linear geostatistics (Vann et al. 2002). Thus, transformation to normality and back-transformation are critical steps. Selection of the parameters for

transformation can be difficult and may have a large effect on the realizations and the uncertainty distribution derived from them (Gotway and Rutherford 1996). In addition, even with modern computers, geostatistical simulation is computationally demanding and depends on the choice of appropriate parameters. Reducing the search radius and maximum number of conditioning data can accelerate the simulation, but may lead to poorer results.

11.3 Case Study II: Uncertainty in the Position of Geographic Objects

11.3.1 Introduction

Global positioning system (GPS)-based tracking, area measurement and machine guidance are among the most widely adopted technologies in precision agriculture (Bramley 2009; Reichardt et al. 2009). However, GPS-positions are not error free and positional errors may propagate through precision agriculture operations and result in wasted inputs, harm to the environment, unharvested crops, inefficient use of the land and even accidents. Although GPS errors are often considered to be independent and random, they are time-dependent which requires that temporal correlation of errors is taken into account in the assessment of positional uncertainty.

The GPS measurement errors are dominated by systematic errors caused by clock error, satellite orbit, atmospheric and multipath effects, which are different for each satellite. Even after correcting for these errors, the residuals tend to change slowly over time, which adds to the effect of temporal filtering that is applied within GPS receivers (Bona 2000; Olynik et al. 2002; Wang et al. 2002; Amiri-Simkooei and Tiberius 2007). Therefore, GPS measurement errors tend to be temporally correlated. This is acknowledged by manufacturers of agricultural navigation systems, who typically list different values for *track-to-track* accuracy and *absolute* accuracy of GPS receivers. As an example of temporal correlation, Fig. 11.11 shows a scatterplot of GPS positions acquired at a stationary position over a period of 6 h at 2 s interval by the Joint Research Centre (JRC) of the European Union in Ispra (Italy). The plotted positions correspond to single frequency (L1) code phase GPS data which were augmented using European Geostationary Navigation Overlay Service (EGNOS) (European Space Agency et al. 2005). Observations close in time appear more or less spatially clustered, which indicates temporal correlation.

This section considers propagation of positional errors in measured field boundaries that are used to guide field operations, for example on the headlands. Errors in the field boundaries cause errors of commission and omission, where:

1. Commission errors correspond to areas outside the true field geometry which are erroneously mapped as belonging to the field.
2. Omission errors refer to areas belonging to the field that are erroneously excluded from the mapped field.

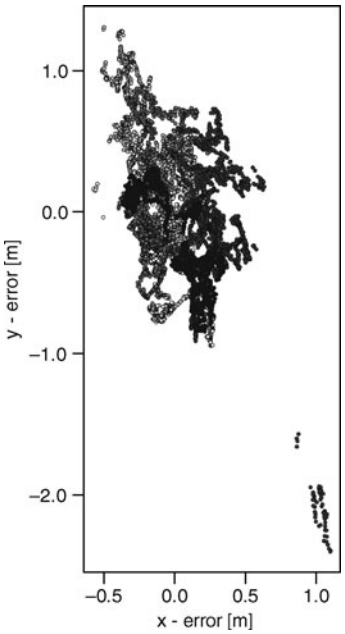

Fig. 11.11 Scatterplot of EGNOS-corrected GPS positions acquired at a fixed position over a period of 6 h. Measurements were acquired at 2 s interval and are time-coded using different shades of grey

Both types of error involve costs owing to ineffectual inputs and loss of net income, as well as other effects. The propagation of positional error into these costs is assessed following the three steps of error propagation, i.e. (1) definition of the error model, (2) identification of the error model and (3) application of the model (Heuvelink 1999). The paragraphs on methods of this section are largely taken from (De Bruin et al. 2008), but the material differs in the data and the implementation software used.

11.3.2 Methods

11.3.2.1 Definition of Positional Error Model[1]

Methods for defining positional uncertainties in geographic objects include partial and full applications of probability theory to vector data. They range in complexity from the simple 'epsilon band' approach, where a buffer of radius ε is imposed

[1] Reprinted from Computers and Electronics in Agriculture, 63/2, S. de Bruin, G.B.M. Heuvelink and J.D. Brown, Propagation of positional measurement errors to agricultural field boundaries and associated costs, pp 247–248, Copyright (2008), with permission from Elsevier.

around each line segment, to map perturbation functions (Kiiveri 1997), and the estimation of joint probability distribution functions for the elementary vertices of line segments, which include the spatial correlation between the positional errors of the vertices. Following the latter approach, the positional uncertainty of area objects in a GIS is expressed as a function of uncertainty in the coordinates of their elementary vertices (Shi and Liu 2000; Zhang and Kirby 2000; Bogaert et al. 2005; Heuvelink et al. 2007). The coordinates of vertices are subject to observational error which in two-dimensional Cartesian space can be represented by the random variables X and Y, with marginal cumulative probability distribution functions (MPDFs) F_X and F_Y

$$F_X(x) = \text{Prob}(X \leq x) \quad \text{and} \quad F_Y(y) = \text{Prob}(Y \leq y), \tag{11.5}$$

where x and y are real numbers. When the errors in the x- and y-direction are correlated, the joint (cumulative) pdf is required

$$F_{XY}(x, y) = \text{Prob}(X \leq x \text{ and } Y \leq y). \tag{11.6}$$

The expectations or means μ_X and μ_Y can be estimated from repeated measurements at a location with known coordinates to provide information on positional bias. The standard deviations σ_X and σ_Y of X and Y are measures of the random positional uncertainty in the x and y errors, respectively. A spatial object is treated as 'deformable' if its component vertices can move with a degree of independence (Heuvelink et al. 2007). Description of the positional uncertainty of a deformable object composed of n vertices requires a $2n$-dimensional joint probability distribution function (JPDF). This JPDF contains the MPDFs for the coordinates of the individual vertices, together with all the auto- and cross-correlations between them

$$\begin{aligned} &F_{X_1 Y_1 \ldots X_n Y_n}(x_1, y_1, \ldots, x_n, y_n) \\ &\quad = \text{Prob}\,(X_1 \leq x_1, Y_1 \leq y_1, \ldots, X_n \leq x_n, Y_n \leq y_n). \end{aligned} \tag{11.7}$$

As explained above, GPS surveys may introduce temporal correlations of measurement errors in the vector data sets, and these will affect the correlations represented in Eq. 11.7.

11.3.2.2 Identification of the Model

Estimation of Eq.11.7 typically relies on the assumption of second-order stationarity, whereby the second-order properties of the positional errors (means and covariances) do not vary under translation. This usually involves modelling with a joint normal distribution, which is completely defined by its means and covariances,

and a covariance function that depends only on the relative distances between points (Goovaerts 1997; Heuvelink et al. 2007).

Geostatistical error models typically consider spatial correlations of random variables. However, as referred to above, in the case of GPS positional errors it is more straightforward and realistic to model temporal correlations than the resulting spatial correlations, because the temporal lag between the recordings is the main force behind the correlation. Therefore, our error model considers the temporal correlation of positional errors. These were described by temporal variograms

$$\gamma_X(\tau) = \tfrac{1}{2}\text{var}[X(t+\tau) - X(t)], \tag{11.8}$$

where t is time, τ is a time lag and $\gamma X(\tau)$ is the variogram of positional errors in the x-direction. The same function can be defined in the y-direction, whereas a cross-variogram $\gamma XY(\tau)$ (see Chapter 7) is needed to characterize the cross-correlation between positional errors in the x- and y-directions.

Variogram analysis of the experimental data described below was performed using R (http://www.r-project.org) and the gstat package for R (Pebesma 2004; Bivand et al. 2008). In the case of cross-correlations between the x and y errors, a linear model of coregionalization was used (Goovaerts 1997). We assumed the errors to be normally distributed as did Bogaert et al. (2005), however, unlike them we allowed for different variances for the GPS errors in the x and y directions. This was considered relevant because GPS satellite orbits cross the equator at an angle of 55°, which reduces the signal availability from the northern (y) direction in the Netherlands (52°N latitude), so that typically $\sigma_Y > \sigma_X$. This effect can also be appreciated in Fig. 11.11.

11.3.2.3 Application

The areas (A) corresponding to errors of commission and omission errors were computed using

$$A_{Commission} = A_{Measured} - A_{Measured} \cap A_{Reference}, \tag{11.9}$$

$$A_{Omission} = A_{Reference} - A_{Measured} \cap A_{Reference}, \tag{11.10}$$

where $\cap$ denotes a spatial intersection, *Reference* refers to a true geometry and *Measured* is a measured geometry. The areas obtained in this way were multiplied by the associated costs.

Heuvelink (1999) refers to two major error propagation methods: the Taylor series and Monte Carlo. The Taylor series method requires that the spatial operation is continuously differentiable, which is not the case when computing the intersection of polygon geometries. Therefore we applied a Monte Carlo method, as described below.

11.3.2.4 Data and Scenarios

Two GPS datasets were used to parameterize measurement scenarios for our error model. The RTK-GPS comprises survey-quality data acquired with dual frequency (L1 and L2) carrier phase Real Time Kinematic GPS equipment. The data were acquired in Wageningen, The Netherlands at a stationary position on May 24, 2007 between 9:44:13 and 12:08:31 UTC time. The EGNOS GPS are augmented GPS data from a stationary position in Ispra (Italy). These are single frequency (L1) code phase GPS data, with differential correction based on networks of widely spaced reference stations and transmitted by geostationary satellites. The data were acquired by JRC between August 25, 2007 at 00:00:01 and September 3, 2007 at 23:59:59 UTC time. Figure 11.11 shows a subset of the latter data set.

We also considered a scenario in which field geometry is acquired by manual digitizing. Following Van Buren et al. (2003) and De Bruin et al. (2008) we assumed zero mean positional errors in the x- and y-directions with a standard deviation of 2 m in each direction ($\sigma_X = \sigma_Y = 2\,\text{m}$) and no temporal correlation or cross correlation.

The financial consequences of errors of omission and commission (Eqs. 11.9 and 11.10) were assessed by multiplying $A_{Commission}$ with €0.2023 m^{-2} and $A_{Omission}$ with €0.1923 m^{-2}. The first value corresponds to the net return rate of a potato crop (one season) in the Hoeksche Waard (Praktijkonderzoek Plant en Omgeving 2006). The second value represents loss resulting from inefficiently used inputs (seed-potatoes, fertilizers, fungicides, herbicides, insecticides, fuel, soil analyses, interest and crop insurance). This loss would be higher if the costs of labour, wear of machinery, risks of damage to equipment, humans and the environment were also included, but these were not considered here.

11.3.3 Study Site

Figure 11.12 shows our study site; an irregularly shaped field of about 16 ha in the Hoeksche Waard, The Netherlands. The (x_i, y_i) coordinates in the Dutch grid system of $n = 59$ vertices of this field were measured using RTK-GPS equipment. The resulting coordinates and mapped field boundaries were used as the reference geometry in this study. Note that any observation error in these locations is of no consequence for the results because the reference geometry constitutes our 'true' geometry in all subsequent calculations.

Under the measurement scenarios mentioned above, however, the coordinates of vertices are subject to observational error, which were considered at the original 59 vertices in all cases. The survey of reference geometry took 15 min; individual vertices were assumed to be equally separated in time, at 15 s (unfortunately, the time of the measurements was not recorded). Since the geostatistical software used assumes coordinates are defined in two or three dimensions, a constant dummy time coordinate was added.

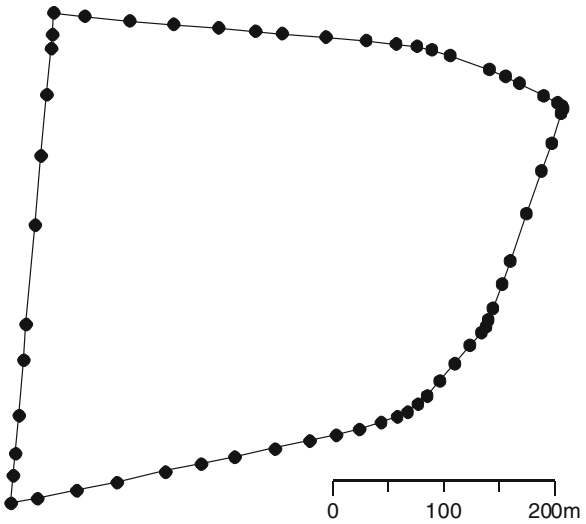

Fig. 11.12 Reference geometry including 59 vertices of a 16 ha field in the Hoeksche Waard, The Netherlands

11.3.3.1 Software

The geostatistical analyses were done in R (R Development Core Team 2008) using the gstat and rgdal packages (Pebesma 2004; Bivand et al. 2008), and Python Scripting (http://www.python.org) using the GDAL/OGR library (http://www.osgeo.org/gdal_ogr) was used to compute the intersection required in Eqs. 11.9 and 11.10. Note that the above software can be downloaded free from the internet. Self intersections of the field boundary occurring under the manual digitizing scenario were corrected using ArcGIS® by adding vertices at intersections (Repair Geometry tool), converting multiple geometries to single geometries and deleting sliver polygons. We ran 1000 Monte Carlo simulations using sequential Gaussian simulation (GPS scenarios) or by drawing from a bivariate Gaussian distribution (manual digitizing scenario).

11.3.3.2 Results and Discussion

Model Identification

Figures 11.13 and 11.14 show the experimental variograms of the x and y errors and their cross-correlations for the RTK-GPS and EGNOS GPS data. Parameters of nested variogram models fitted using the linear model of coregionalization (Goovaerts 1997) are listed in Table 11.4. Temporal cross-correlation between x and y errors in the RTK-GPS data was ignored because of its small magnitude and

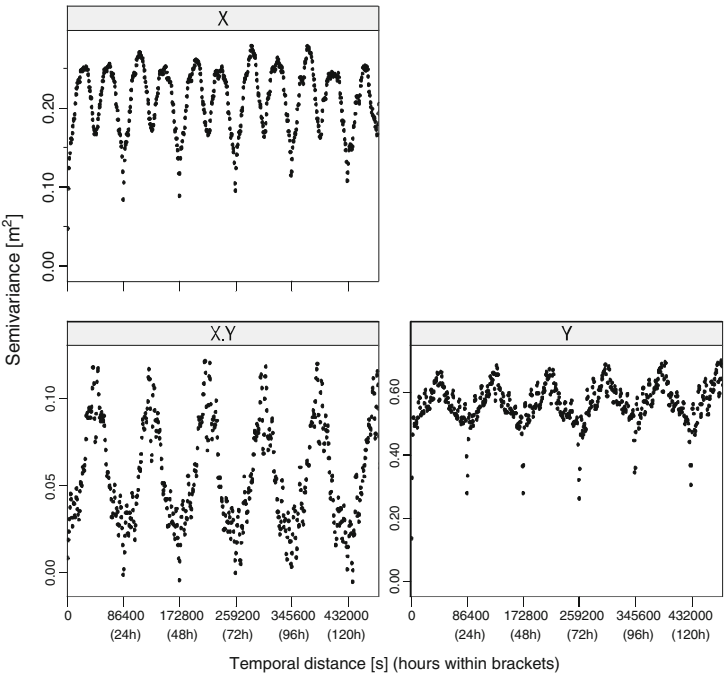

Fig. 11.13 Experimental variograms and cross-variogram of the RTK-GPS time series

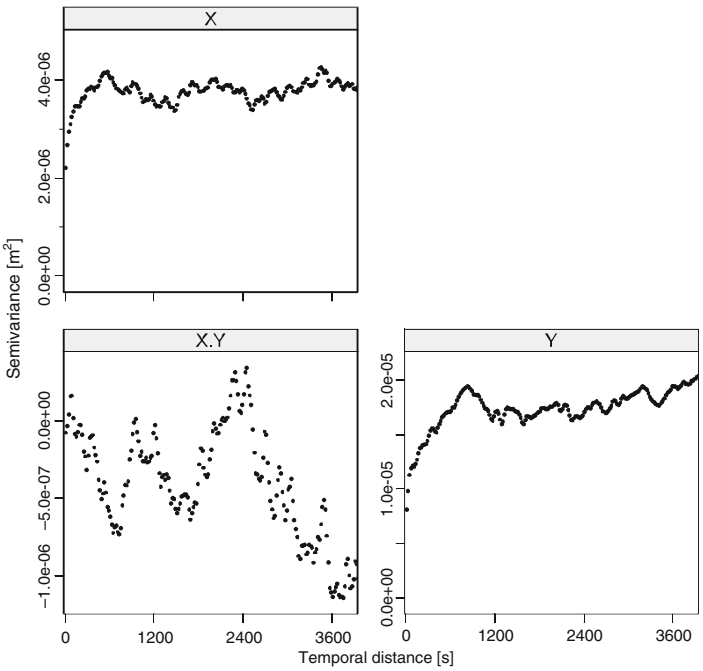

Fig. 11.14 Experimental variograms and cross-variogram of the EGNOS GPS time series

Table 11.4 Fitted variogram models

	EGNOS			RTK		
Element	Model type	Partial sill or amplitude [m^2][a]	Range or wavelength [s][a]	Model parameters	Partial sill or amplitude [m^2][a]	Range or wavelength [s][a]
X[1]	Periodic	0.0225	86 164	Nugget	2.08e-06	–
X[2]	Periodic	0.0398	43 082	Exponential	1.74e-06	85.2
X[3]	Exponential	0.1470	1100	–	–	–
Y[1]	Periodic	0.0925	86 164	Nugget	9.46e-06	–
Y[2]	Periodic	0.0000	43 082	Spherical	8.38e-06	739.0
Y[3]	Exponential	0.4700	1100	–	–	–
X.Y[1]	Periodic	0.0353	86 164	–	–	–
X.Y[2]	Periodic	0.0000	43 082	–		
X.Y[3]	Exponential	0.0263	1100	–	–	–

[a]Second option for periodic structure.

erratic structure. The EGNOS GPS data show periodicity with a main period of 86 164s (almost 1 day) corresponding with the period over which the GPS satellites complete two orbits and the Earth completes one revolution, so that the constellation returns to the same geometry (Agnew and Larson 2007).

Model Application

Figure 11.15 shows histograms of simulated loss (1000 realizations) resulting from positional uncertainty (three measurement scenarios) and summary statistics of the distributions are given in Table 11.5. Recall that we assumed positional errors to be normally distributed. The non-Gaussian distributions of the histograms of the output of the GPS scenarios (Fig. 11.15a, b) are symptomatic of the non-linear operations performed on the data. However, the histogram of the manual digitizing scenario (Fig. 11.15c) appears more or less Gaussian.

For the 16-ha field considered here, a manually digitized map would involve an expected loss of almost € 408 with respect to reference geometry. The expected loss associated with RTK-GPS surveyed geometry is <1 €. This implies that if an RTK-GPS survey cost <€ 407 the farmer would be better off using this than a manually digitized map for machine guidance on the headlands. Table 11.5 also indicates that there is more than a 99% chance that the farmer would benefit from a map acquired under the EGNOS scenario rather than a manually digitized map of the field as $P_{90(EGNOS)} < P_{10(Manual)}$. The GPS survey would not need to be done every year yet the benefits may last for several years. Note that the benefits of accurate mapping will increase if other risks such as damage to equipment and infrastructure and externalities (e.g. environmental effects) are also taken into account.

Finally, Fig. 11.16 illustrates a shortcoming of the error model in which uncertainty in area features is attributed solely to uncertainty in the coordinates of their elementary vertices which are assumed to be connected by straight lines. As a

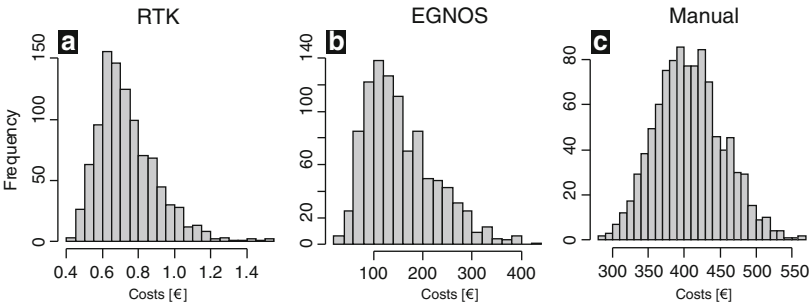

Fig. 11.15 Histograms of the costs resulting from positional uncertainty for: (a) RTK-GPS, (b) EGNOS and (c) Manual measurements

Table 11.5 Summary statistics of financial loss [€] with respect to error-free geometry under three scenarios

			Percentile		
Scenario	Mean	SD	P_{10}	P_{50}	P_{90}
RTK	0.74	0.16	0.56	0.70	0.96
EGNOS	154.4	70.34	76.70	139.3	255.4
Manual	408.0	47.58	347.4	405.8	471.8

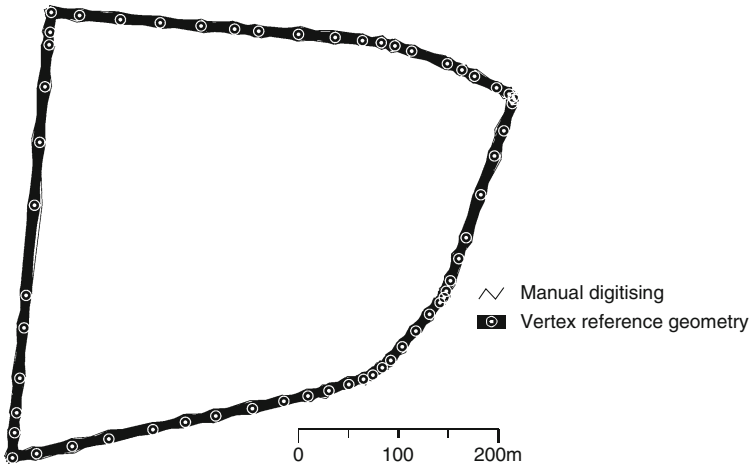

Fig. 11.16 One thousand simulated field geometries of the manual digitizing scenario. Notice that most uncertainty (largest spread) occurs at the vertices

consequence of this assumption, positional uncertainty is largest at the measured positions of the vertices and diminishes with distance from the vertex points along the line segments; the smallest value is midway between the measured locations. Such an approach disregards uncertainty caused by sampling and approximation of a curvilinear feature by a sequence of straight line segments. De Bruin et al. (2008)

suggest a method that allows for the latter type of uncertainty by modelling random rectangular deviations from the conventional straight line segments, but application of that method was considered beyond the scope of the current chapter.

11.3.4 Conclusions

This case study has shown that geostatistical techniques such as variogram modelling and stochastic simulation, which are typically used to model spatial correlation of observed data, can also be used for modelling temporal correlation. Our variogram analysis of stationary GPS time series revealed substantial temporal correlation in the positional errors for RTK-GPS and EGNOS corrected GPS data. Our data demonstrate considerable differences between the losses in PA associated with specific accuracies of the position of field boundaries, even when only direct losses for a farmer are considered. The expected benefits of accurate maps would probably be even larger if secondary losses (e.g. damage to equipment and environmental pollution) were also taken into account.

11.4 Case Study III: Uncertainty Propagation in Soil Mapping[2]

11.4.1 Introduction

Uncertainty in the process of soil mapping can be caused by several factors. Some sources of error can be modelled by geostatistical simulation, and this section illustrates how it can be integrated with other methods to form a complex model for assessing accuracy in soil mapping. Maps of the crop nutrients N, P, K and Mg, and pH from which to derive site-specific fertilizer and lime recommendations are the outcomes of several consecutive steps involving sampling design, sample location, laboratory analysis and interpolation to create a digital map. We call this the soil mapping process. The accuracy or uncertainty of the resulting maps is affected by factors such as the choice of methods and instruments, and the inherent variability of soil. Uncertainty is not only due to randomness (stochastic uncertainty), but also to the lack of knowledge about the outcomes of alternative methods (structural uncertainty). The effect of individual factors on map accuracy, such as sampling density, the sampling scheme or method of interpolation, has been evaluated by several researchers (Gotway et al. 1996; Demougeot-Renard et al. 2004; Hoskinson et al. 2004; Mallarino and Wittry 2004; Farahani and Flynn 2007). However, there

[2] Gebbers, R., Herbst, R., & Wenkel, K.-O. (2009). Sensitivity analysis of soil nutrient mapping. In E. J. van Henten, D. Goense, & C. Lokhorst (Eds.), *Precision Agriculture '09. Proceedings of the 7th European Conference on Precision Agriculture* (pp. 513–519). Wageningen, The Netherlands: Wageningen Academic Publishers.

have been few attempts to compare the effects of these factors on the accuracy of maps of soil nutrient concentration (Mueller et al. 2004). To optimize soil mapping it is essential to understand how the variation in these factors is propagated simultaneously through the mapping process. This will then help to identify the most important factors with respect to errors in soil nutrient maps.

11.4.2 Materials and Methods

The effect of the factors above on the errors of mapping nutrient concentrations in soil was investigated by modelling the mapping process by computer. The stochastic uncertainty is simulated by Monte Carlo methods and inserted into the model. The structural uncertainty, e.g. the use of different sampling designs, is implemented by alternative algorithms (branches) inside the model.

Our model of the soil mapping process includes sample design and location, chemical analysis and the creation of digital maps. Soil nutrient concentrations of a 36 ha virtual field of 1 by 1 m pixels were created by geostatistical simulation using Gstat (Pebesma 2001). These digital fields represented realistic nutrient concentrations and were regarded as the true dispersion that should be predicted by the mapping process. Spherical variogram models with ranges of 30, 60, 90, 120, 150 and 180 m were used to model differences in spatial variation. These ranges were representative of soil spatial variation in different regions of Germany for plant available P, K, Mg and for pH (Herbst et al. 2001). Here we consider only K, but the results of this uncertainty analysis can be extended to the mapping process of other crop nutrients and pH. Field average nutrient concentration was set to the optimum soil index for K of 11 mg K 100 g^{-1} (double lactate extract, according to VDLUFA 1996). The standard deviation of K was set to 5 mg K 100 g^{-1} to obtain a variation of plus or minus one soil index class. A Gaussian distribution was assumed. Soil sampling was simulated by point queries on the simulated fields. Two main sampling schemes were modelled: (a) subdivision of the field into quadratic management zones and bulking 20 samples within each zone and (b) sensor based online data collection by point sampling and instant chemical analysis. Bulk sampling (scheme a) was either done on diagonal lines crossing each zone or on circular lines of 10 m radius around the centre of each zone. The size of the sampling zones was varied from 5 ha, which is the maximum size allowed in Germany (Dungeverordnung et al., 2007), to 1 ha, which is a typical zone size in precision farming practice, to 0.05 ha, which is a sampling density that can be achieved by an on-line soil sensor such as the Veris MSP (Veris Technologies, Salina, USA), see Adamchuk et al. (2007). To account for accuracy of the GPS, we simulated positioning errors of 2 m by pseudo random number generation using a zero-centred normal distribution. A standard error of 2 m is typical for L1 band differential GPS with a beacon or EGNOS correction. We did not account for temporal autocorrelation in the GPS errors because this is less relevant in the context of soil mapping. Errors in chemical analysis were obtained by simulating cases with normally distributed errors of 0%, 4% or 20%. An accuracy of 2–4% is what our laboratory achieved.

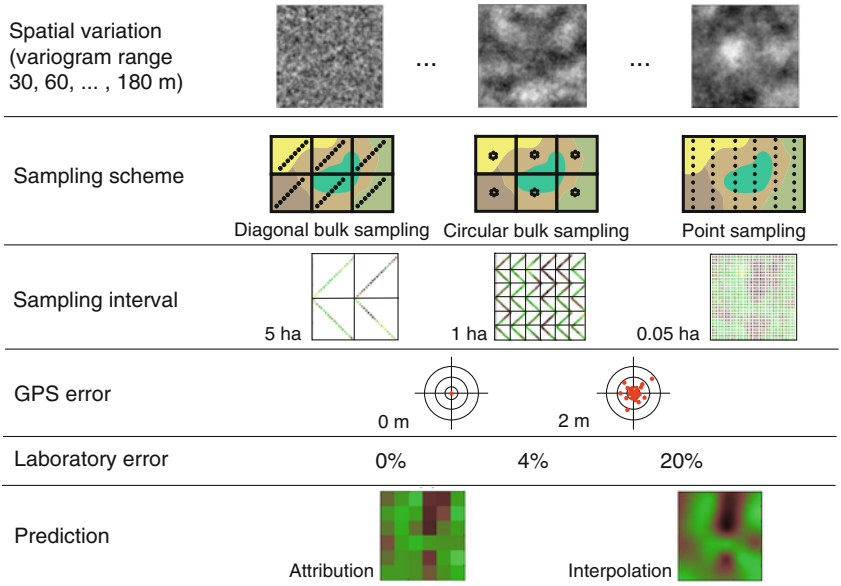

Fig. 11.17 Model inputs and the respective range of possible values (input parameter space) of the model of the soil mapping process

An error of 20% is assumed to be a worst-case scenario (e.g. an on-line sensor with poor calibration). Two methods were applied to create maps of predictions from the samples: (a) attribution of the data to the previously defined zones and (b) interpolation. For zone sampling, the centres of zones were used as reference points for interpolation. Interpolation was done by biharmonic splines. Model inputs and the respective range of possible values (input parameter space) are shown in Fig. 11.17.

The combination of six ranges of autocorrelation, three sampling schemes, three sampling intervals, two levels of positioning error, three levels of laboratory error, two methods of regionalization and two replications resulted in 1296 runs of the soil mapping model. For each map simulation, the results were compared with the values of the simulated 'true' nutrient distribution (Fig. 11.18). The standard deviation of the differences between these values was calculated to summarize the mapping error.

11.4.3 Results and Discussion

Based on the standard deviation of the differences between the 'true' nutrient distribution and the simulation of the mapping process, a sensitivity analysis as described in Gebbers et al. (2009) was conducted to identify the most important factors affecting the accuracy of maps based on sample information. The three most important factors for accurate soil mapping were sampling interval, inherent spatial varia-

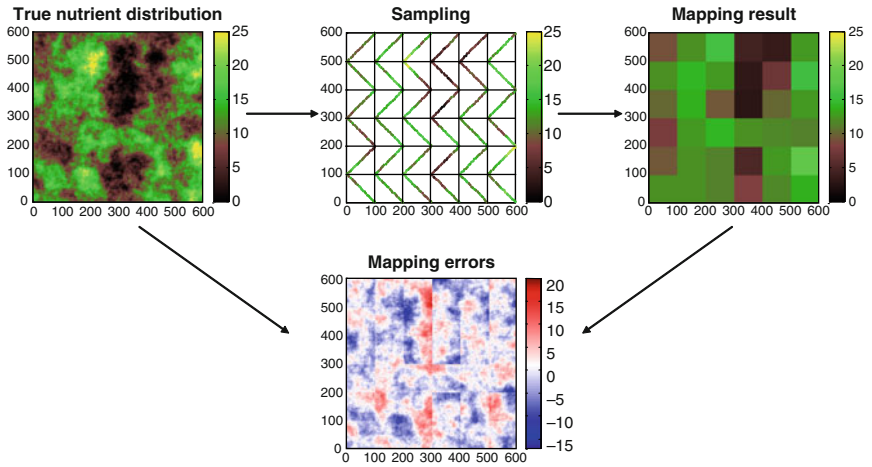

Fig. 11.18 Example of the calculation of mapping errors of K [mg K 100 g^{-1}] by simulation of the soil mapping process. It shows zone sampling from 1 ha zones by bulking 20 samples per zone taken along diagonal lines. To create a map, the K concentrations derived from the samples were attributed to the zones

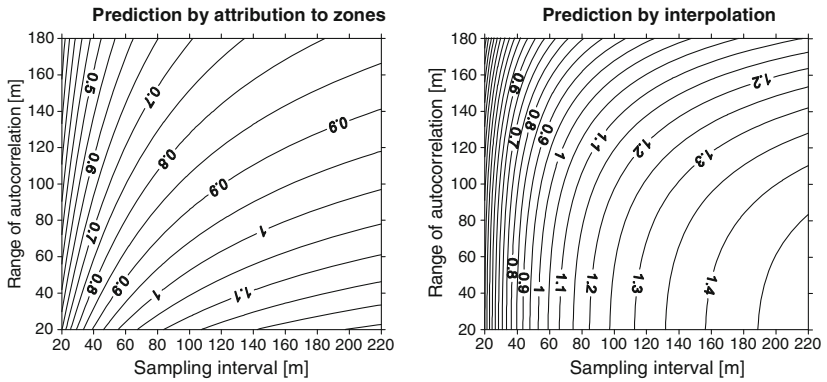

Fig. 11.19 Relative mapping error as a function of sampling interval, range of spatial autocorrelation and regionalization method

tion of the soil property and method of prediction (in this order). These factors were used to construct nomographs of the relative mapping error (Fig. 11.19). The latter is the standard deviation of the mapping error divided by the standard deviation of the simulated images, which represent the true K concentration. A relative mapping error of <1 means that a map showing variation is more accurate than a uniform value, which represents the average of the true K concentration of the field. The reader can derive the potential mapping error from the nomographs in Fig. 11.19 when the range of spatial autocorrelation of the variable, the sampling interval and the prediction method are known. Conversely, if one decides to accept

a certain mapping error, the respective contour line identifies the combination of range parameter and sampling interval associated with it. In this way, the nomographs in Fig. 11.19 can help to determine the optimum combination of sampling interval and prediction method given the range of spatial autocorrelation of the soil property under consideration.

In addition, Fig. 11.19 provides insight into the interactions between the three most important factors for mapping accuracy. The sampling interval has a larger effect on the mapping error when interpolation rather than attribution of values to zones is used, especially at intervals of <100 m. Therefore, interpolation should only be used when samples are taken at small intervals.

11.4.4 Conclusions

The key to improving nutrient maps is an appropriate sampling density (see Chapters 2 and 3). In the soil mapping scenario investigated here, errors of chemical analysis were unimportant compared to inappropriate sampling density. Thus, more reliable maps can be obtained by increasing sampling density even at the expense of accuracy of chemical analysis. Our results advocate strongly the collection of soil nutrient data by online sensors such as the Veris MSP pH sensor, which might result in larger errors in the chemical analysis but enables small sampling intervals. However, one problem remains: The range of spatial autocorrelation has to be determined before one can decide on an appropriate sampling density. Kerry and Oliver (2003) suggested the use of intensive ancillary data such as remotely sensed images or soil electrical conductivity to determine an approximate spatial scale of variation in the soil, but this approach requires that target parameters and ancillary data are correlated.

11.5 Application of Geostatistical Simulation in Precision Agriculture: Summary

Geostatistical simulation helps to optimise site-specific management by including the inherent variation of natural processes in our models. The outcome can have immediate practical implications. Farmers can benefit directly from the case studies presented here. Section 11.2 shows how uncertainty of a pH map is translated into maps of the risk of under- and over-liming. Section 11.3 quantifies the economic costs of GPS errors. The modelling of soil mapping in Section 11.4 identifies the factors that are most relevant to improving mapping accuracy.

Geostatistics offers a large variety of methods to simulate spatio-temporal processes. It might be difficult to identify the appropriate methods for the problem at hand and to find suitable software. However, PA can profit from geostatistical simulation and an increasing demand by the PA community may promote the publication of user-friendly texts and software.

References

Adamchuk, V. I., Lund, E. D., Reed, T. M., & Ferguson, R. B. (2007). Evaluation of an on-the-go technology for soil pH mapping. *Precision Agriculture, 8*, 139–149.

Agnew, D., & Larson, K. (2007). Finding the repeat times of the GPS constellation. *GPS Solutions, 11*, 71–76.

Amiri-Simkooei, A. R., & Tiberius, C. C. J. M. (2007). Assessing receiver noise using GPS short baseline time series. *GPS Solutions, 11*, 21–35.

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer.

Bogaert, P., Delincé, J., & Kay, S. (2005). Assessing the error of polygonal area measurements: a general formulation with applications to agriculture. *Measurement Science & Technology, 16*, 1170–1178.

Bona, P. (2000). Precision, cross correlation, and time correlation of GPS phase and code observations. *GPS Solutions, 4*, 3–13.

Bourgault, G. (1997). Using non-Gaussian distributions in geostatistical simulation. *Mathematical Geology, 29*, 315–334.

Bramley, R. G. V. (2009). Lessons from nearly 20 years of precision agriculture research, development, and adoption as a guide to its appropriate application. *Crop and Pasture Science, 60*, 197–217.

Chilès, J-P., & Delfiner, P. (1999). *Geostatistics. Modeling spatial uncertainty*. New York: Wiley.

de Bruin, S. (2008). Modelling positional uncertainty of line features by accounting for stochastic deviations from straight line segments. *Transactions in GIS, 12*, 165–177.

de Bruin, S., Heuvelink, G. B. M., & Brown, J. D. (2008). Propagation of positional measurement errors to agricultural field boundaries and associated costs. *Computers and Electronics in Agriculture, 63*, 245–256.

Demougeot-Renard, H., de Fouquet, C., & Renard, P. (2004). Forecasting the number of soil samples required to reduce remediation costs uncertainty. *Journal of Environmental Quality, 33*, 1694–1702.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.

Düngeverordnung (2007). Verordnung über die Anwendung von Düngemitteln, Bodenhilfsstoffen, Kultursubstraten und Pflanzenhilfsmitteln nach den Grundsätzen der guten Fachlichen Praxis beim Düngen (Düngeverordnung – DüV). Neufassung der Düngeverordnung (27.02.2007). *Bundesgesetzblatt I, 2007, 221 (Federal act on the use of fertilizers in Germany)*.

European Space Agency (2005). The EGNOS signal explained, EGNOS fact sheet 12. EGNOS fact sheet http://www.egnos-pro.esa.int/Publications/2005%20Updated%20Fact%20Sheets/fact_sheet_12.pdf (accessed August 18 2009).

Faechner, T., Pyrcz, M., & Deutsch, C. V. (1999). Soil remediation decision making in presence of uncertainty in crop yield response. *Geoderma, 97*, 21–38.

Fagroud, M., & Van Meirvenne, M. (2002). Accounting for soil spatial autocorrelation in the design of experimental trials. *Soil Science Society of America Journal, 66*, 1134–1142.

Farahani, H. J., & Flynn, R. L. (2007). Map quality and zone delineation as affected by width of parallel swaths of mobile agricultural sensors. *Biosystems Engineering, 96*, 151–159.

Favis-Mortlock, D. T., Boardman, J., Parsons, A. J., & Lascelles, B. (2000). Emergence and erosion: a model for rill initiation and development. *Hydrological Processes, 14*, 2173–2205.

Gebbers, R., Herbst, R., & Wenkel, K.-O. (2009). Sensitivity analysis of soil nutrient mapping. In E. J. van Henten, D. Goense, & C. Lokhorst (Eds.), *Precision Agriculture '09. Proceedings of the 7th European Conference on Precision Agriculture* (pp. 513–519). Wageningen, The Netherlands: Wageningen Academic Publishers.

Gentle, J. E. (1989). *Random number generation and Monte Carlo methods*. New York: Springer.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Goovaerts, P. (1999). Geostatistical tools for deriving block-averaged values of environmental attributes. *Journal of Geographical Information Sciences, 5*, 88–96.

Gotway, C.A., & Rutherford, B. M. (1996). *The components of geostatistical simulation*. In Mower, H.T. (Ed.), *Proceedings of the Second International Symposium on Spatial Accuracy in Natural Resources and Environmental Sciences*. Fort Collins: USDA Forest Service. (www.osti.gov/bridge/servlets/purl/228463-cOfkGn/webviewable/228463.pdf, accessed September 29 2009).

Gotway, C. A., Ferguson, R. B., Hergert, G. W., & Peterson, T. A. (1996). Comparisons of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America Journal, 60*, 1237–1247.

Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and semiparametric models*. Berlin: Springer.

Herbst, R., Lamp, J., & Reimer, G. (2001). Inventory and spatial modelling of soils on PA pilot fields. In: G. Grenier, & S. Blackmore (Eds.), *Third European Conference on Precision Agriculture* (pp. 395–400). Montpellier, France: Agro Montpellier.

Heuvelink, G. B. M. (1998). *Error propagation in environmental modelling with GIS*. London, UK: Taylor & Francis.

Heuvelink, G. B. M. (1999). Propagation of error in spatial modelling with GIS. In Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (Eds.), *Geographical information systems* (2nd ed., pp. 207–217). New York: Wiley.

Heuvelink, G. B. M., Brown, J. D., & van Loon, E. E. (2007). A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science, 21*, 497–513.

Hoskinson, R. L., Rope, R. C., Blackwood, L. G., Lee, R. D., & Fink, R. K. (2004). The impact of soil sampling errors on variable rate fertilization. In D. J. Mulla (Ed.), *Proceedings of 7th International Conference of Precision Agriculture* (pp. 1645–1654). Minneapolis, USA: Precision Agriculture Center, University of Minnesota, Department of Soil, Water and Climate.

Kerry, R., & Oliver, M. A. (2003). Variograms of ancillary data to aid sampling for soil surveys. *Precision Agriculture, 4*, 261–278.

Kiiveri, H. T. (1997). Assessing, representing and transmitting positional uncertainty in maps. *International Journal of Geographical Information Science, 11*, 33–52.

Lantuéjoul, C. (2002). *Geostatistical simulation. Models and algorithms*. Berlin: Springer.

Lapen, D. R., Topp, G. C., Hayhoe, H. N., Gregorich, E. G., & Curnoe, W. E. (2001). Stochastic simulation of soil strength/compaction and assessment of corn yield risk using threshold probability patterns. *Geoderma, 104*, 325–343.

Leuangthong, O., Khan, K. D., & Deutsch, C. V. (2008). *Solved problems in geostatistics*. Hoboken: Wiley.

MAFF (2000). *Fertilizer recommendations for agricultural and horticultural crops*. London: The Stationery Office.

Mallarino, A. P., & Wittry, D. J. (2004). Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. *Precision Agriculture, 5*, 131–144.

MATLAB: *The MathWorks*, Natick, MA. (www.mathworks.com).

Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., & Barnhisel, R. I. (2004). Site-specific soil fertility management: a model for map quality. *Soil Science Society of America Journal, 68*, 2031–2041.

Olynik, M., Petovello, M., Cannon, M., & Lachapelle, G. (2002). Temporal impact of selected GPS errors on point positioning. *GPS Solutions, 6*, 47–57.

Pebesma, E. J. (2001). *Gstat users' manual*. Utrecht, The Netherlands: Department of Physical Geography, Utrecht University. http://www.gstat.org/gstat.pdf. Accessed 28. Jan. 2009.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences, 30*, 683–691.

Pokrajac, D., Fiez, T., & Obradovic, Z. (2002). A data generator for evaluating spatial issues in precision agriculture. *Precision Agriculture, 3*, 259–281.

Praktijkonderzoek Plant en Omgeving (2006). *Kwantitatieve informatie akkerbouw en vollegrondsgroenteteelt 2006 (KWIN 2006)*. Lelystad: Praktijkonderzoek Plant en Omgeving.

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. (www.r-project.org).

Reichardt, M., Jürgens, C., Klöble, U., Hüter, J., & Moser, K. (2009). Dissemination of precision farming in Germany: acceptance, adoption, obstacles, knowledge transfer and training activities. *Precision Agriculture* (DOI 10.1007/s11119-009-9112-6, accessed October 1 2009).

Remy, N., Boucher, A., & Wu, J. (2009). *Applied geostatistics with SGeMS*. New York: Cambridge University Press.

Shi, W., & Liu, W. (2000). A stochastic process-based model for the positional error of line segments in GIS. *International Journal of Geographical Information Science, 14*, 51–66.

Shiflet, A. B., & Shiflet, G. W. (2006). *Introduction to computational science*. Princeton: Princeton University Press.

SYSTAT: Systat Software Inc., Chicago, IL. (www.systat.com).

Van Buren, J., Westerik, A., & Olink, E. J. H. (2003). *Kwaliteit TOP10vector – De geometrische kwaliteit van het bestand TOP10vector van de Topografische Dienst*. Kadaster – Concernstaf Vastgoedinformatie en Geodesie: 12.

Vann, J., Bertoli, O., & Jackson, S. (2002). An overview of geostatistical simulation for quantifying risk. In S. M. Searston, & R. J. Warner (Eds.), *Quantifiying risk and error*. Geostatistical Association of Australasia Symposium. (http://www.qgeoscience.com/images/downloads/vann_bertoli_jackson_simulation_for_risk_distribution.pdf, accessed September 29 2009).

VDLUFA (1996). *Kalium-Düngung nach Bodenuntersuchung und Pflanzenbedarf. Richtwerte für die Gehaltsklasse C*. Darmstadt, Germany: VDLUFA-Verlag (Potassium fertilization based on soil analysis and crop requirements. Guidelines to maintain optimum soil index).

VDLUFA (2000a). *Bestimmung des Kalkbedarfs von Acker- und Grünlandböden*. Darmstadt: Verband Deutscher Landwirtschaftlicher Untersuchungs- und Forschungsanstalten (VDLUFA). (Determination of lime needs for arable and grassland soils) (http://www.vdlufa.de/joomla/Dokumente/Standpunkte/0-9-kalk.pdf, accessed August 18 2009).

VDLUFA (2000b). *Bestimmung des Kalkbedarfs von Acker- und Grünlandböden. Appendix. Richtwerte für das Rahmenschema zur Kalkbedarfsermittlung in Deutschland*. Darmstadt: Verband Deutscher Landwirtschaftlicher Untersuchungs- und Forschungsanstalten (VDLUFA) (Recommendation tables for lime) (http://www.vdlufa.de/joomla/Dokumente/Standpunkte/0-9-kalkanl.pdf, accessed 18.08.2009).

Wang, J., Satirapod, C., & Rizos, C. (2002). Stochastic assessment of GPS carrier phase measurements for precise static relative positioning. *Journal of Geodesy, 76*, 95–104.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester: Wiley.

Zanolin, A., de Fouquet, C., Granier, J., Ruelle, P., & Nicoullaude, B. (2007). Geostatistical simulation of the spatial variability of an irrigated maize farm plot. *Comptes Rendus Geosciences, 339*, 430–440.

Zhang, J., & Kirby, R. P. (2000). A geostatistical approach to modelling positional errors in vector data. *Transactions in GIS, 4*, 145–159.

Chapter 12
Geostatistics and Precision Agriculture: A Way Forward

J.K. Schueller

Abstract Geostatistics became an integral part of precision agriculture (PA) early in its history. The essence of this chapter is to identify some future trends in the application of geostatistics to PA, but predicting the future of these subjects is difficult because of their dynamic nature. There is a need to accommodate the variation in weather and temporal changes in the theory and practice of geostatistics. An important issue is to make space-time geostatistics more accessible to practitioners. Geostatistics needs to be tailored better to the needs of the various groups involved; farmers, advisors and researchers who have their own particular requirements. The book has raised several issues, ideas and questions, which are summarized in this chapter. For example, data quality, more transparency in the packages for analysis, more advanced methods of geostatistics and the use of Bayesian techniques, more user-friendly software that is inexpensive, education in geostatistics, more automatic soil and crop data recording, greater use of ancillary data and better understanding of the relations with the soil and yield, and so on. The potential for geostatistics and precision agriculture for the rest of the twenty-first century appears great.

Keywords Agricultural community · Anisotropy · Spatial and temporal variation · Education · Future · User-friendly

12.1 Introduction

Research in precision agriculture started to be documented in the mid-1980s, by authors such as Lullen (1985) and Schueller and Bae (1987). As was shown by the inclusion of Mulla (1989) in one of the first precision agriculture review papers (Schueller 1992), geostatistics was soon an integral part of precision agriculture.

J.K. Schueller (✉)
Department of Mechanical and Aerospace Engineering,
University of Florida, Gainesville, FL 32611-6300
e-mail: schuejk@ufl.edu

M.A. Oliver (ed.), *Geostatistical Applications for Precision Agriculture*,
DOI 10.1007/978-90-481-9133-8_12,

This book takes on the daunting task of examining the intersection of a new and changing field of operational agriculture (precision agriculture) with a fairly new and changing set of techniques of analysis and prediction (geostatistics). The dynamic nature of the new fields of precision agriculture and geostatistics make accurate projections of their future interactions difficult. Nevertheless, some future trends and needs can be identified.

The preceding chapters show how geostatistical methods can be applied in precision agriculture and how geostatistics has contributed to its advancement. Geostatistics is a relatively recent technology dating mainly from the 1960s. Therefore, while technologies for applying geostatistics in precision agriculture are being developed, geostatistics itself also continues to develop. Because there have not been many years of experience with geostatistics in agriculture and geostatistical education and training are not widespread, the application of geostatistics to precision agriculture is more challenging than the application of conventional statistics.

The authors of the previous chapters do an excellent job of reviewing the past, presenting the present and making some suggestions for the future. They do this in addition to formalizing how precision agriculture can and should use geostatistics. The purpose of this chapter is to provide further suggestions.

12.2 Weather, Time and Space

Crop production agriculture was treated historically as fixed and deterministic. For example, the yield of a crop in a field or region was often expressed as 'x tons per hectare'. Of course, the yield at any particular position in the field or region was almost never 'x'; there was always variability. It is possible to characterize the variability as having three dominant categories of cause: weather-related, temporal and spatial.

The old canard that 'climate is what you expect, weather is what you get' applies to crop production agriculture. Differences in weather are the main cause of variation from one growing season to the next. In many production situations where only one crop is produced per year, this results in substantial annual differences in crop production, even for the same commodity grown at the same place. Within each crop growing season there is continuous variation in temperature, precipitation, solar radiation, humidity and other features of the weather. This variation is confounded with, and interacts with, temporal and spatial variation.

Agronomic and soil characteristics, such as yield, crop quality, crop moisture content, soil moisture content and soil nutrient status change in time over both short and long time periods. This gives rise to temporal variation. Understanding of the different types of temporal variations is often absent and usually is not taken into account in agronomic decision-making.

Spatial variation is essentially the reason for the development of precision agriculture. In fact, 'spatially-variable crop production' is one of the many names that is better than 'precision agriculture' for these technologies. Precision agriculture

attempts to maximize agronomic productivity and efficiency by consciously considering spatial variation and acting accordingly. Spatial variation and the statistical tools to deal with that variation are the focus of this book and of this chapter. Nevertheless, it appears desirable, and perhaps necessary, that precision agriculture should accommodate the other two dominant causes of variation. Scientists and engineers in precision agriculture need to consider variation in both weather and over time, and how they interact with spatial variation. It seems that this is a current 'grand challenge' for geostatistics and precision agriculture.

Studies, such as that by Colvin et al. (1997), show how the variation in yield itself varies greatly from year-to-year. Bakhsh et al. (2000) point out that the lack of temporal stability in variation requires the analysis of long series of yield data for precision agriculture to gain insight into the weather conditions that have a major effect from year to year. More studies are needed to determine the temporal nature of spatial variation in agriculture.

Studies of yield usually assume that it is stationary with respect to time, but the climate and technologies continuously change. The potentially rapid changes in climate, i.e. commonly referred to as climate change, may have effects on yield that will lead to greater non-stationarity in yield data. In addition to climate, agricultural management such as new seeds, techniques for applying inputs and other technologies, as well as effects such as salinization and erosion, changes the yield over time. Spatio-temporal geostatistical techniques are needed in precision agriculture that can detrend the temporal effects, and possibly prioritize recent (and therefore probably more relevant) data. Perhaps there should be an exponential decay weighting of data over time.

It is known that certain weather events can have a localized spatial variability. For example, summer thunderstorms in North America can produce very variable patterns of precipitation. Another example is damage from freezing in Florida, which can vary considerably depending upon the nature of the event and the local topography. Localized variation in weather is not taken into account in current precision agriculture or in geostatistical models. Given the advent of wireless micro-climate sensors, however, the inclusion of localized weather information in geostatistical models would be feasible and useful.

As mentioned above, the quantities of interest in agronomy and horticulture change with time. Consider nitrogen, for example; elemental, nitrate, nitrite or ammonium nitrogen forms can vary greatly both spatially and temporally. The nitrogen available to a plant will depend upon where the nitrogen is in relation to the plant's roots and in what form it is present in at that time. Crop yield and quality can also vary with time. For example, green pea quality and yield change drastically within a day. The peas can be very tender in the early morning, but be hard 'bullets' by the evening on a hot day.

Time alone is not the issue. The issue is to put time and space together. Scientists and engineers need to provide both spatial and temporal data. Experiments, tests, and other data gathering should be planned so that geostatistical models can have temporal components.

Time also has an effect through the dynamics of the equipment. Dynamic effects are well-known for yield mapping and variable-rate application. These dynamics can be countered with proper adjustments. For example, there can be postprocessing compensation for delays in yield and other mapping. Control theory concepts, such as 'feedforward control' and 'precommand', can be used to counteract temporal delays in application machines by issuing application rate commands the correct small amount of time before the responses to these commands should be achieved. Overall, the integration of time and space is a much needed and potentially very rewarding area of understanding that is still in its infancy. Chapter 5 in this book is an example of the developing discussion. Many models for precision agriculture should include a temporal component.

12.3 Farmers, Advisors and Researchers

The potential of precision agriculture and geostatistics appeals to diverse groups in the agricultural community. The expectations of precision agriculture and geostatistics also vary greatly because of the diversity of experiences and needs in this community. There is a wide variation in crops, technological capabilities, geographic locations, educational backgrounds and job functions. It is difficult and perhaps foolhardy to make broad generalizations, but such generalizations will be made here to facilitate discussion. Accordingly, the users of precision agriculture and geostatistics can be divided into three categories which may be roughly termed 'farmer', 'advisor' and 'researcher'. When projecting into the future, it might be instructive to examine what each can expect.

The 'farmer' category represents the agricultural producer who may be a family farmer, a farm manager of a large farm or a worker on a farm. These people are responsible for producing and harvesting the crop with physical or management labour. They have to perform many different operational and management tasks. They do not have the time, or often the inclination, to become experts in geostatistics. Geostatistics must be essentially invisible to them. The use of geostatistics should be embedded in the generation of information. For example, maps of yield, soil properties and crop conditions should be generated automatically with the geostatistical and other technical details hidden from the farmer. Nevertheless, the farmer must consider properly what kinds of data are required and the constraints involved in the underlying analysis. It means that the equipment designers and software providers must make their interfaces 'user-friendly' so that the embedded geostatistical capabilities and procedures are clear to understand and implement.

The 'advisor' category represents those practitioners in the public and private sectors who provide advice to agricultural producers. This is a varied group, including extension workers, private consultants and commercial salesmen. They must understand more than the 'farmer' category in order to be able to help the farmer. The advisors will often perform analyses and make recommendations. Because they must be paid and their time is valuable, they need tools that can improve

their efficiency and productivity. However, advisors will commonly have had more training and experience in geostatistics, and have a clearer understanding of when and how it will be of value. They will often want to do more sophisticated analyses. Therefore, their tools must be more flexible. For example, an advisor doing yield mapping may want to have more choice of geostatistical model than the average farmer would want to consider.

The 'researcher' category represents those doing research to understand precision agriculture and to develop new techniques. This category is probably easier to understand if it is divided into two sub-categories: users of geostatistics and experts in geostatistics.

Many soil scientists, agronomists, economists and engineers who work in precision agriculture and wish to apply geostatistics do not have a strong geostatistical background. This must be remedied by a twofold approach. First, new researchers must have more training in geostatistics as part of their university courses. Researchers in crop production agriculture must be expected to know more statistics than just regression, ANOVA and simple experimental design. The agricultural universities have been very slow to react to the need for training in geostatistics. This must, and probably will, change in the near term. Instruction in geostatistics will become the norm for the more respected courses in the agricultural sciences. Secondly, the lack of knowledge about geostatistics among existing researchers in crop production agriculture must be dealt with. Courses are available based upon the instructors' notes or books such as Goovaerts (1997), but such courses are often costly. Research administrators must realize that their staff needs this training and the administrators should ensure that it is provided at little or no cost to the researchers. Existing researchers need to understand geostatistical concepts and have as much proficiency with the appropriate geostatistical tools as they have with the conventional statistical tools in SAS, SPSS, R, etc. They need to have tools available that are appropriate for agriculture and that deal with weather, and temporal and spatial variability.

For the farmers to have the necessary tools and techniques, they need to be taught by advisors who have the necessary agronomic and statistical knowledge. The farmers must have at least some conceptual knowledge of variation to fine-turn their processes of production and to learn how to apply the tools and techniques. The advisors, however, must know what variation there is in local crop production and what hardware, software and management tools are available. These tools, in turn, must come from the researchers who are users of geostatistics. Those researchers depend upon the geostatistical and precision agriculture experts. So there is an order of dependence from farmers to advisors to geostatistics users to experts. If the experts in geostatistics and precision agriculture develop the necessary tools and techniques, they will eventually percolate down to the farmers with appropriate assistance.

Hence, it is crucial that there are sufficient researchers with expertise in both geostatistics and precision agriculture who are working at the intersection of those two topics. Given the recognition of the importance of this intersection, it is likely that such a community of researchers will eventually be developed, particularly if university instruction in geostatistics becomes more widely available. Techniques

and computerized tools will be likely to result from the work of these individuals. Those researchers who are geostatistics experts will provide the foundation upon which the integration of geostatistics and precision agriculture is based.

12.4 Issues, Ideas and Questions

The previous chapters of this book present many issues and ideas that suggest further ideas, and also pose many questions.

Some of these are:

- One definition of precision agriculture is that it is a management strategy that uses data from multiple sources to bear on decisions associated with crop production. Is spatial variation a necessary condition for precision agriculture? Some researchers now include all achievements of crop production excellence, such as maintaining a consistent planting depth, as part of 'precision agriculture'. Therefore, the breadth of the term 'precision agriculture' should be established by widespread consensus to facilitate more accurate communication and understanding.
- How will anisotropy be handled in widespread practice? Although there have been some examples of effective handling of anisotropy, for example in Chapter 11 of this book, eventually the preponderance of anisotropy in precision agriculture will have to be reflected in the geostatistical tools and techniques that are used in widespread practice by farmers and advisors. To achieve practical implementation of anisotropy in commercial tools and techniques, researchers must investigate how anisotropy can be more effectively and efficiently handled.
- Similarly, the importance and quality of sources of data vary widely. There needs to be effective and objective weighting of data. This might be achieved first according to general guidelines developed by researchers from local empirical data, but later through automatic adaptive weighting based upon experiences with local data. Different data are likely to have different impacts on the models and these effects need to be understood.
- Mixed models, cokriging, indicator kriging, Bayesian and other relatively advanced techniques must be developed further for application in precision agriculture. As conventional statistics is becoming more complex and more Bayesian, so must geostatistics. Nevertheless, concern about the reliability and complexity of these approaches should be taken into account. Geostatistics in precision agriculture must avoid the common plague of overfitting and overtransforming data that is prevalent in conventional statistics. This plague affects many of those who currently use conventional computer statistics package. Researchers must develop appropriate procedures and guidelines that are clear, straightforward and maintain a high standard of analysis. Many geostatistical applications in mapping packages are treated as a 'black boxes' and so the user has no idea whether the analysis is acceptable because the important intervening steps are hidden. This approach should be avoided.

- It was pointed out correctly in an earlier chapter that the analyst should not adopt a mechanical approach to data analysis. This is especially true for the researchers. However, to achieve widespread use in commercial production agriculture, geostatistics must be easy-to-use and user-friendly. Tools have to be developed that have defaults that are robust and do not require many inputs. The defaults, however, must be able to be changed permanently for permanent local conditions and temporarily changeable for transient conditions.
- Besides ease-of-use, software cost is important. Increasingly, there is inexpensive or free software available. But does that software have support and an improvement path? There will probably be a range of software costs and related performances and reliabilities. The various users need to make the proper choices and this requires expert guidance.
- Soil and pest sampling should be automated to reduce cost. With advances in robotics, automated robots should be able to scout fields and take soil and plant samples and automatically process them. Initially, humans will be 'in-the-loop' planning sampling, setting up the analyses and studying the results. Eventually, however, those aspects should be automated, including the geostatistical parts. Geostatistics should drive sampling to facilitate analysis of both dependent and independent data.
- Geostatistics requires enough data to establish a reliable variogram. For crop data there has been much progress, but low-cost automated soil and pest sampling and analysis is needed to facilitate the gathering of ample data for reliable geostatistical analyses. The likely development of such automated systems will increase the use of geostatistics.
- The variograms used in precision agriculture should use auxiliary data to supplement sampled data as described in Chapters 2 and 7. The process of obtaining and using such data should be automated. For example, dense remote sensing and electrical conductivity data should aid the interpolation of sparse soil sampling data.
- Accordingly, there needs to be significantly more research into the correlations between measured values that are difficult or costly to obtain and cheaper covariates. Applied researchers need to determine what the reliabilities of covariates are for local conditions.
- A major problem in precision agriculture is how to merge space and time. Time needs to be dealt with adequately, whereas at present it is largely disregarded. Space and time vary in different ways and a further complication is that both vary at several scales, i.e. multiscale variation. A profitable area for fundamental research is to develop geostatistics so that it can handle the dynamics of time more readily than at present.
- In addition, geostatistics and other precision agriculture analyses need to be able to handle crop-to-crop, within-crop and equipment dynamics. The effects of these dynamics should be removed. Variation in the weather must be dealt with similarly.
- Great progress continues to be made in such fields as signal processing, image processing and control theory with the development of new understandings,

techniques and tools. These techniques and tools should be applied more widely to precision agriculture. For example, pattern recognition could be used with kriged data.
- There are times, such as in computer vision techniques and remote sensing, when there are possibly too many data. Data reduction might then be guided by geostatistics (e.g. Oliver et al. 2005). This area has potential for further development.
- The techniques for making estimates of errors and uncertainties need to be developed further in precision agriculture and geostatistics, and applied to representative situations. Chapter 11 described how geostatistical simulation can be used in this context. It is not widely known in precision agriculture at present.

Dealing with these various issues above should propel geostatistics to become an even more integral part of precision agriculture.

12.5 Past, Present and Future

Pioneering innovators of the past developed the exciting and useful techniques of geostatistics and precision agriculture in the late twentieth century. Diligent practitioners, scientists and engineers have refined, expanded and further applied the techniques in the first decade of the twenty-first century. The potential for geostatistics and precision agriculture in the remainder of the twenty-first century is great. Pressing forward enthusiastically to answer questions with further development and widespread applications will realize that potential.

References

Bakhsh, A., Jaynes, D. B., Colvin, T. S., & Kanwar, R. S. (2000) Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. *Transactions of the ASAE, 43*, 31–38.

Colvin, T. S., Jaynes, D. B., Karlen, D. L., Laird, D. A., & Ambuel, J.R. (1997). Yield variability within a central Iowa field. *Transactions of the ASAE, 40*, 883–889.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

Lullen, W. R. (1985). Fine-tuned fertility: tomorrow's technology here today. *Crops and Soils, 38*, 18–22.

Mulla, D. J. (1989). Using geostatistics to manage spatial variability in soil fertility. In C. M. Renard, R. J. Van den Beldt, & J. F. Parr (Eds.) *Soil, crop and water management in the Sudano-Sahelian Zone* (pp. 241–254). Patancheru, India: ICRISAT.

Oliver, M. A., Shine, J. A. and Slocum, K. R. (2005). Using the variogram to explore imagery at two different spatial resolutions. *International Journal of Remote Sensing, 26*, 225–3240.

Schueller, J. K. (1992). A review and integrating analysis of spatially-variable control of crop production. *Fertilizer Research, 33*, 1–34.

Schueller, J. K. and Y. H. Bae (1987). Spatially attributed automatic combine data acquisition. *Computers and Electronics in Agriculture, 2*, 119–127.

Appendix: Software

A.1 Geostatistics in GenStat

R. Webster

Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, United Kingdom

GenStat is a powerful, flexible and numerically sound system for statistical analysis programmed to the most up-to-date professional standards. The main geostatistical tasks are built into it and can be accomplished from menus or with single commands. Other tasks can be programmed in the GenStat language, and all can be combined to run sequentially in single submissions. GenStat is a general comprehensive system for analysing data from experiments and surveys. It was originally devised in the late 1960s to analyse data from designed experiments. The system is marketed by VSN International (www.vsn-intl.com). In due course geostatistical subroutines were added to the system, and these now enable users to tackle the standard analyses with confidence and to explore spatial data by ever more advanced techniques. The current version of GenStat, released in July 2009, is the 12th.

The following basic tasks in geostatistical investigation and analysis can all be readily accomplished in GenStat both by command and by menu.

1. Posting. These are maps of the positions of sampling points, and bounding outlines of regions can be added.
2. Statistical summaries and marginal distributions, and display as histograms and box-and-whisker plots and of cumulative distributions.
3. Auto-variograms. Experimental, or sample, variograms can be computed from univariate data on regular transects and grids and from irregularly scattered data. Users can choose binwidths in both distance and direction (for two-dimensional data). They can also choose whether to use the usual method-of-moments estimator or one of the robust estimators of Cressie and Hawkins (1980), Dowd (1984) or Genton (1998).

The output consists of the estimated semivariances for the nominated lags (in distance and direction) and the numbers of paired comparisons (counts) contributing to the estimates. The general graphics commands can be used to display the variograms with semivariance plotted against lag distance.

4. Model fitting. Having computed experimental variograms users can fit models to them. With a single command a user can fit all of the popular variogram functions and several that are not so popular, including Whittle's elementary correlation function, the pentaspherical function and the cubic function, and take into account anisotropy. Estimates are weighted in proportion to the counts by default, but users can weight the estimates equally or inversely as the expected values. The experimental values and the fitted functions can be displayed graphically.

The command calls into play the more general routine for fitting non-linear functions, and users may write their own code in the GenStat language. This enables them to specify models that are not available with the single command.

The output from both approaches lists the fitted values of the parameters, including standard errors of the non-linear ones, and an analysis-of-variance table. There is the option to monitor the iteration, valuable if the procedure does not converge properly so that the cause might be identified.

5. Kriging. Ordinary auto-kriging is the 'workhorse' of geostatistics, and it is the basic form of kriging in GenStat. The user has many options to control the computations. There are those such as the area within which kriged predictions are required and whether estimates are required for points or blocks. In addition the user can choose the search radius for data or the minimum and maximum number of data points for each prediction. Universal kriging is a straightforward extension of it. Variograms may be any of the standard ones, though only the power function is included at present for anisotropic cases. The output comprises principally the predictions and their associated variances. Both can be mapped by calls to graphics commands, but users may prefer to transfer the results to special mapping programs for final display and printing. The Lagrange multipliers can be saved; this enables users to back-transform predictions to their original scales, in particular after log-normal kriging. There are other output options to enable users to monitor the kriging process.
6. Cross-validation. Cross-validation proceeds by the leave-one-out principle. Each data point is omitted in turn, the value there is predicted by kriging with the chosen model and the other data in the neighbourhood, and the results are summarized in terms of mean error, mean squared error and the mean ratio of the squared errors to the kriging variances.
7. Coregionalization. For two or more variates, GenStat will compute all experimental auto- and cross-variograms and fit to them a linear model of coregionalization. The user chooses the basic components of the model and starting values for the non-linear parameters. The program fits the chosen model by iteration, ensuring that the result is conditionally negative semi-definite at every stage. The commands for forming all the experimental variograms and for fitting the models to them are simple extensions of those for the univariate case.

8. Cokriging. GenStat will predict the values of any one variate from values of it and others coregionalized with it at points or in blocks by ordinary cokriging using the models fitted by the coregionalization directive.
9. Nested analysis. Geostatisticians are recognizing the merits of nested sampling and analysis for revealing the gross spatial structure of regionalized variables (see Chapters 2 and 9). The data can be analysed by ANOVA, but the solution is not unique; they are better analysed by residual maximum likelihood (REML) as devised by Patterson and Thompson (1971). GenStat has a comprehensive suite of facilities for analysing data from such designs.

Webster and Oliver (2007) provide examples of GenStat code for these tasks. The standard operations can be done either by written commands in the GenStat language or by menu. In the latter case GenStat generates the code and saves it in an output log. This means that you have a full record, step by step, of what you have done and that you can modify the code to create more comprehensive programs.

A.2 VESPER

Budiman Minasny, Alex B. McBratney and Brett M. Whelan

Australian Centre for Precision Agriculture, The University of Sydney, NSW 2006, Australia

A.2.1 Background

VESPER (Variogram Estimation and Spatial Prediction plus Error) is a user-friendly PC-windows software program that can calculate and model global local variograms and do global and local kriging in either punctual or block form.

VESPER was developed to deal with the large volume of intensive data collected by on-the-go proximal soil and crop sensors (approximately 5000–65 000 data points per km^2). The purpose is to represent the data (yield data or soil electrical conductivity, etc.) as a digital map at a regular grid interval. In most geostatistical software, spatial interpolation usually involves two separate steps: calculating and modelling of the variogram for the whole area followed by prediction at unsampled points by kriging on a regular grid over the area. There are several shortcomings to this approach: first the time taken to calculate an empirical variogram of the whole area can be excessive (e.g. a variogram for 100 000 data points can take hours to calculate), and secondly information is lost by assuming a single variogram model for the whole area, and which results in a smooth map. VESPER can accommodate the large number of data and take into account the local spatial structure.

Kriging with local variograms, also known as kriging and automated variogram modelling within a moving window, involves searching for the closest neighbourhood for each prediction site, estimating the empirical variogram from the neighbourhood, fitting a variogram model to the data automatically by a non-linear least squares approach, kriging with local neighbourhood and variogram parameters and calculating the uncertainty of kriging prediction. The program adapts itself spatially in the presence of distinct differences in local structure over the whole field. Local variogram estimation and kriging can preserve the true local spatial variation in the predictions. In most cases, local variograms could circumvent the problems of anisotropy and the need for trend analysis.

A.2.2 The Software

VESPER comes as executable files consisting of an interface (written in Visual basic) and a main computation program (written in Fortran). The execution of the program is through the interface. Figure A.1a shows the main interface panel, where input and output files are controlled. Input data containing Cartesian coordinates and the variable of interest are required as an ASCII text file. The output files record the specific session details, variogram model parameters, the prediction locations, values and associated prediction error.

The variogram panel (Fig. A.1b) provides the choice of global (whole-area) or local variogram estimation. The variogram is estimated by Matheron's method-of-moments. A comprehensive range of models (Fig. A.1c) can be fitted to the empirical variogram using four possible weighting procedures (Fig. A.1c). Nonlinear least-squares estimation is used in the model fitting process, minimizing (Jian et al. 1996):

$$R = \sum_{i=1}^{n} w_i \left[\hat{\gamma}\left(h_i\right) - \hat{\gamma}^*\left(h_i\right)\right]^2, \tag{A.1}$$

where w_i is the weighting option, $\hat{\gamma}(h_i)$ is the estimated semivariance at distance h_i and $\hat{\gamma}(h_i^*)$ is the semivariance predicted by the model. The 'goodness of fit' of the models can be compared with the Akaike Information Criterion, AIC, and sum of squared errors.

If a global variogram is required, the 'Fit Variogram' button provides access to an interactive calculation and modelling panel (Fig. A.1d); model parameters can be extracted for subsequent kriging. The global modelling panel also allows subjective model fitting through interactive parameter control bars. This is useful with small data sets and applications where emphasis is needed at particular sample separations.

Local variograms are calculated automatically for each neighbourhood during the local kriging process, but the maximum distance and number of lags required for estimating them may be set through the Variogram panel. Experience at the Australian Centre for Precision Agriculture has shown that an exponential model is

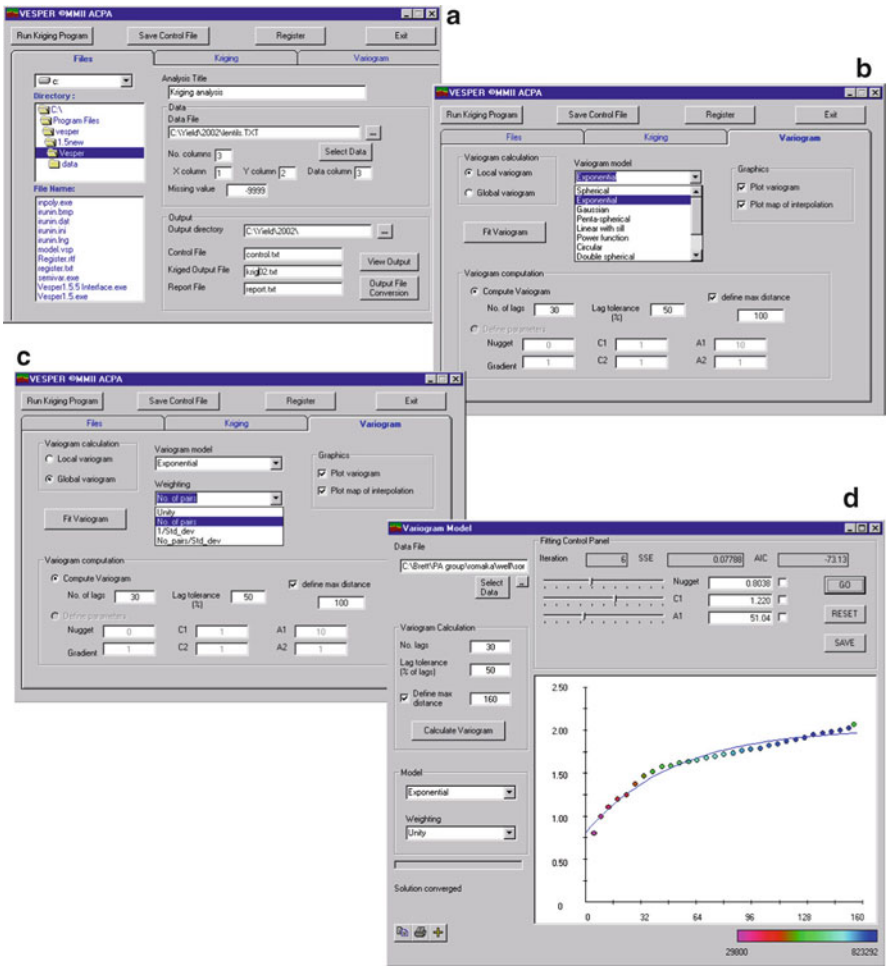

Fig. A.1 Operational panels of VESPER: (**a**) file input/output control panel, (**b**) variogram panel showing available models, (**c**) weighting options for model fitting and (**d**) global variogram operation window

usually the best for local variogram estimation of yield data. More complex models, e.g. Matérn models, often become unstable with automatic local fitting (the covariance matrix can become non positive-definite in some local areas). We recommend limiting model selection to either exponential or spherical models.

The kriging panel (Fig. A.2) provides punctual or block options. It is possible to define the block size (if relevant), set neighbourhood limits based on radial distance or number of data points and manipulate the kriging region. For yield data, a block size close to the swath width is recommended (a 10 m^2 block is sufficient for most combine harvester fronts at present).

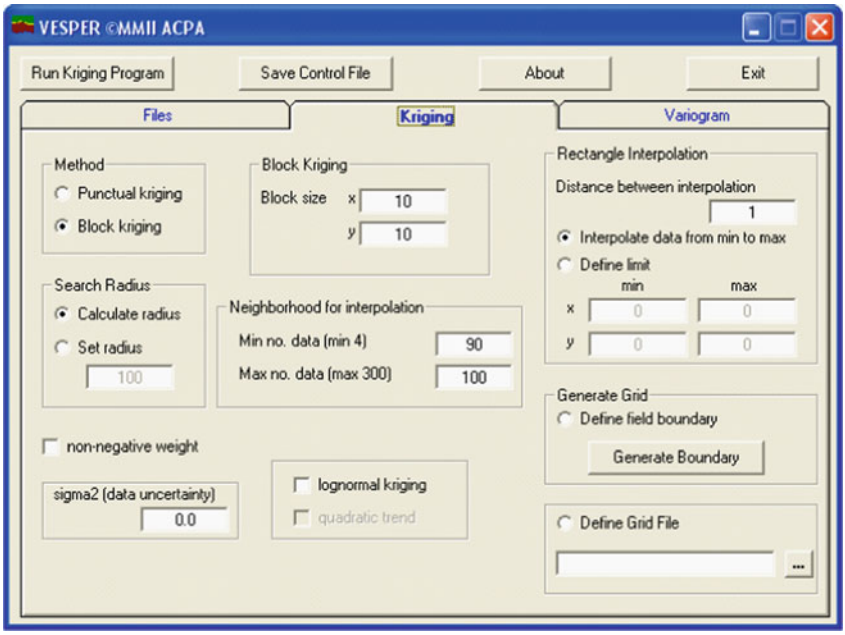

Fig. A.2 Kriging panel of VESPER – showing controls for punctual or block kriging, neighbourhood definition, boundary, grid and advanced options

An interpolation grid is used to define the location of the points to which the data will be predicted. The interpolation grid allows data that are collected at different intervals to be collocated and analysed further. The interpolation grid can be specified in one of the following options:

- When the field has a rectangular shape, a grid with regular distance can be defined.
- When the field has an irregular shape, the boundary can be manually defined and a grid with regular distance can be generated that is confined to the boundary area.
- A file containing a pre-defined grid can be specified. Users can define the spatial coordinates of the prediction.

For most precision agriculture applications, the field boundary will provide the limits of the kriging region. VESPER provides the option of importing an existing boundary file or describing the field boundary using an interactive drawing tool. The prediction grid (at user-defined distances) may then be produced with the software or an existing grid file can be imported. These features are important for the continuity of prediction sites through time within a field.

In addition, for specific applications, VESPER can also perform:

- Lognormal Kriging – transforms lognormal data before performing the interpolation process.

- Non-negative weight – used to ensure 'extreme' values do not produce irrational results e.g. negative yield values. VESPER uses the method of Deutsch (1996) for correcting negative weights.
- Sigma2 (data uncertainty) – is a user-defined estimate of the variance or uncertainty in the data.

In operation, VESPER provides a window displaying the operational progress (Fig. A.3). For all forms of kriging a prediction progress map is produced together

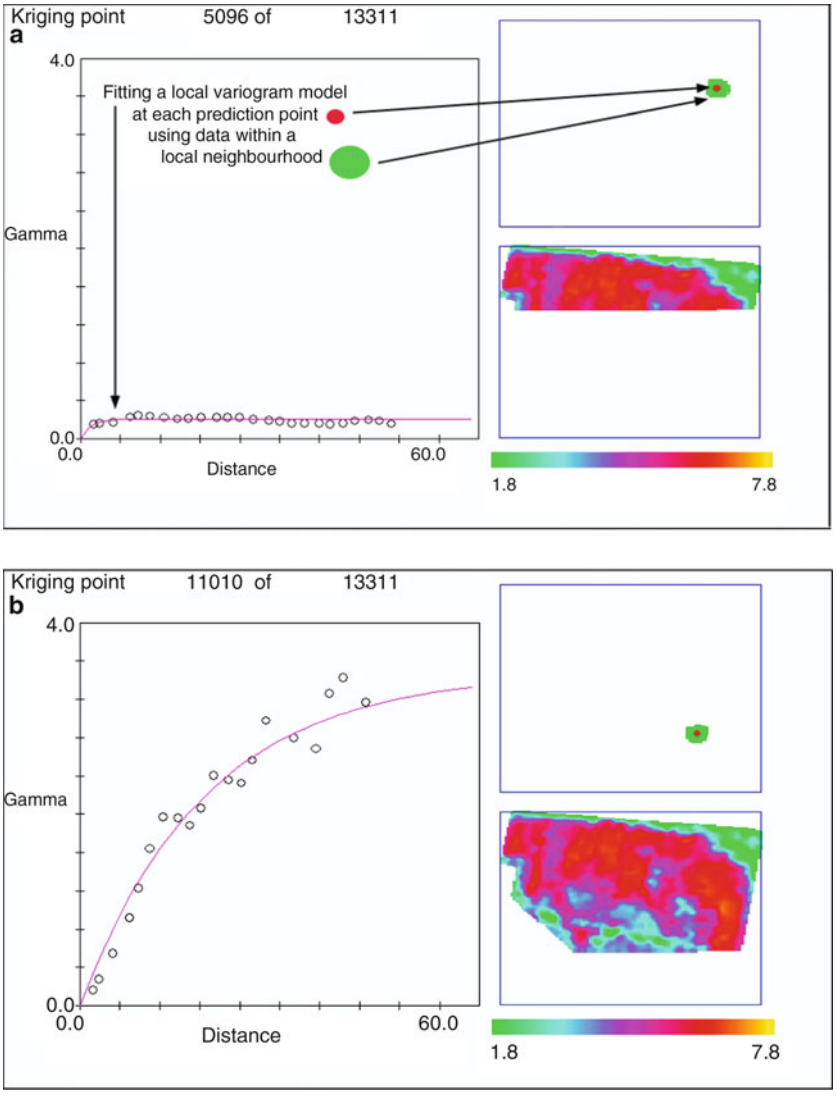

Fig. A.3 Local variogram, data neighbourhood and prediction point progress map for an area with: (**a**) low variability and (**b**) greater variability

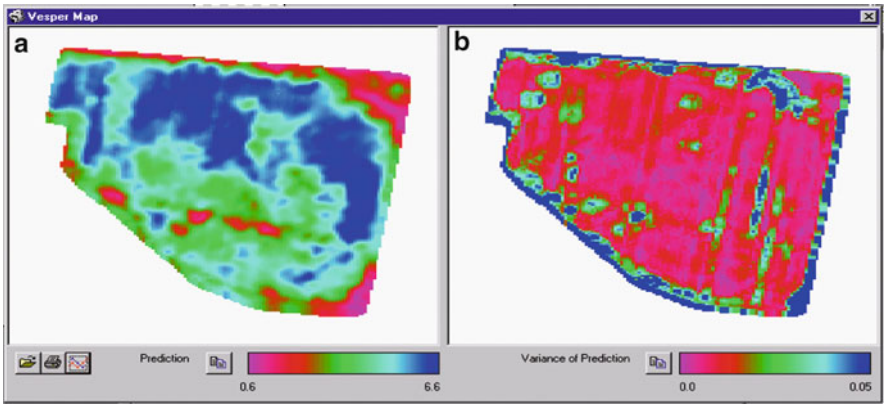

Fig. A.4 Output maps for: (**a**) kriged predictions and (**b**) kriging variances

with a count of visited versus total prediction sites. For local kriging, individual variograms and the fitted models are displayed for the search neighbourhood around each prediction point. Note in Fig. A.3a, b that this local method allows changes in local variability to be reflected in the variogram parameters for each prediction. The graphical progress facilities can be disengaged to increase the speed of the prediction process.

The output for all kriging operations is an ASCII text file containing the prediction point location coordinates, the predicted value and the kriging variance. An input file detailing the exact settings for each prediction session is also saved together with a report file logging global variogram parameters or the parameters of each local variogram depending on the operation. Other details of the data and kriging session are also recorded in this file for future reference. Maps of estimates and prediction variances (Fig. A.4) can be obtained at the end of kriging.

VESPER is available as freeware from the ACPA at www.usyd.edu.au/su/agric/acpa The CSIRO Precision Viticulture group produces a PostVesper tool which automates the process of converting file output from the Vesper kriging program into raster format in ArcGIS.

A.2.3 Applications

VESPER has been used both for research and practical applications. Google Scholar identified about 135 papers that mentioned the use of VESPER for kriging. In Australia, it is being used routinely for making yield maps with data from commercially available yield monitors for cereals and grapes, for example. VESPER is used routinely to format multi-year and multi-sensor data onto a single grid for multivariate analysis and the creation of potential management classes.

A.3 SGeMS and Other Software

R. Gebbers

Department of Engineering for Crop Production, Leibniz-Institute for Agricultural Engineering, Max-Eyth-Allee 100, D-14469 Potsdam, Germany

A.3.1 SGeMS

The Stanford Geostatistical Modeling Software (SGeMS) is an open-source computer package developed by Nicolas Remy with contributions from Alexandre Boucher, Jianbing Wu and Ting Li. SGeMS offers a wide range of functions and is a successor to the well-known Geostatistical Software Library (GSLIB; (Deutsch and Journel 1998, www.gslib.com). A Windows version of SGeMS can be downloaded from http://sgems.sourceforge.net/. The SGeMS homepage provides additional resources such as source code for developers and people who need to build SGeMS on Linux or Apple OS. Although there is a manual at http://sgems.sourceforge.net/old/index.html, the supplemental textbook by Remy et al. (2009) is recommended. The strengths of SGeMS are its user-friendliness (a graphical user interface allows access to every function), the 3-D data visualisation and range of kriging and simulation methods (Fig. A.5).

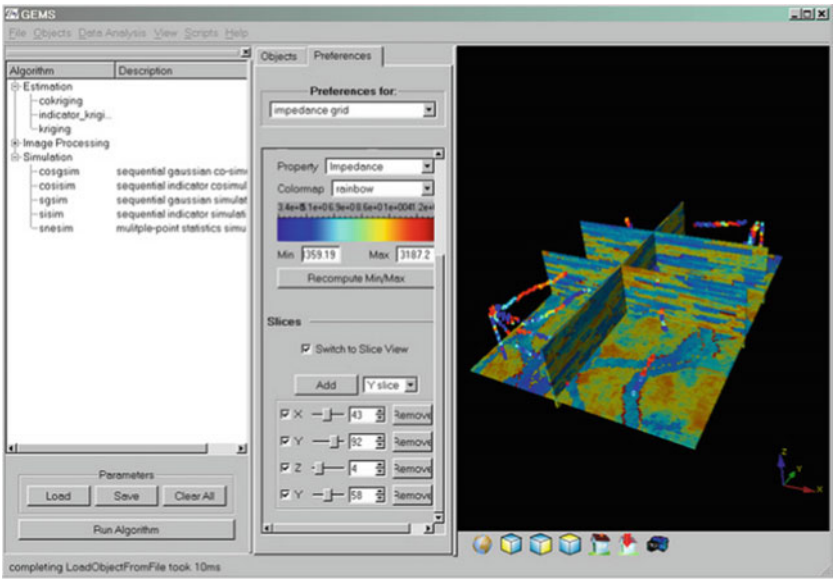

Fig. A.5 Kriging panel of SGeMS – showing controls for kriging, image processing and simulation

A weakness of this package is the variography; it provides only visual aids for variogram fitting and it does not calculate any goodness-of-fit indicators. Four variogram models only are available: nugget, spherical, exponential and Gaussian. SGeMS loads and stores point and grid data in a binary format and the ASCII GSLIB format. Tools for exploratory data analysis include histograms, QQ-plots, scatter plots and descriptive statistics. Directional univariate and cross-variograms can be calculated from point and grid data in two and three dimensions. SGeMS provides univariate, indicator, co- and block-kriging. The point kriging options include simple, ordinary, trend and local means kriging. For cokriging, the user may choose from three coregionalization models (linear, Markov I and II). SGeMS simulates smooth surfaces or volumes by variogram based methods such as LU decomposition, sequential Gaussian, co-simulation, indicator, block and block error simulation. Discrete spatial structures such as horizontal soil layers can be simulated by multipoint methods. SGeMS allows for unconditional or conditional simulation. Variography, kriging and simulation may require pre- and post-processing. SGeMS can be automated and extended by its internal scripting language, plug-ins written in Phyton and a MATLAB interface.

A.3.2 Other Software

Other software packages that include geostatistical functions or may be extended by plug-ins are the Gstat library (www.gstat.org), SYSTAT (www.systat.com), SAS (www.sas.com/software) S-PLUS (www.insightful.com), R (R Development core team (www.r-project.org)) and Terraseer STIS (http://www.terraseer.com/products_stis.php). Among these, R is outstanding: it is freeware and there is a large number of extensions, in particular, there are specific classes and methods for spatial data analysis (see Diggle and Ribeiro 2007; Bivand et al. 2008). ISATIS (www.geovariances.fr) is dedicated geostatistical software. MATLAB (www.mathworks.com) is frequently used for geostatistics and can be linked to SGeMS and Gstat by the mGstat toolbox. The BMElib toolbox extends classical geostatistical simulation with Bayesian approaches for space-time applications (Christakos et al. 2002). Response surface sampling design software, known as ESAP, has been developed specifically for use with EC_a measurements and other proximal sensors (Lesch et al. 2000; see http://www.ars.usda.gov/services/software/software.htm for this open access software.

References

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R.* New York: Springer.

Christakos, G., Bogaert, P., & Serre, M.L. (2002). *Temporal GIS. advanced functions for field-based applications.* New York: Springer.

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram. *Journal of the International Association of Mathematical Geology, 12*, 115–125.

Deutsch, C. V. (1996). Correcting for negative weights in ordinary kriging. *Computers & Geosciences,* 22, 765–773.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.

Diggle, P. J., & Ribeiro Jr., P. J. (2007). *Model-based geostatistics.* New York: Springer.

Dowd, P. A. (1984). The variogram and kriging: robust and resistant estimators. In G. Verly, M. David, A. G. Journel, & A. Marechal (Eds.), *Geostatistics for natural resources characterization* (pp. 91–106). Dordrecht: D. Reidel.

Genton, M. G. (1998). Highly robust variogram estimation. *Mathematical Geology, 30*, 213–221.

Jian, X., Olea, R. A., & Yu, Y.-S. (1996). Semivariogram modeling by weighted least squares. *Computers & Geosciences, 22*, 387–397.

Lesch, S. M., Rhoades, J. D., & Corwin, D. L. (2000). *ESAP-95 version 2.01R: User manual and tutorial guide. Research Rpt. 146.* Riverside, CA: USDA-ARS, U. S. Salinity Laboratory.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58*, 545–554.

R Development Core Team (2008). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (www.r-project.org).

Remy, N., Boucher, A., & Wu, J. (2009). *Applied geostatistics with SGeMS*. New York: Cambridge University Press.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester: Wiley.

Index