ELSEVIER

Geoderma 97 (2000) 293–327

GEODERMA

www.elsevier.nl/locate/geoderma

An overview of pedometric techniques for use in soil survey ☆

Alex B. McBratney [a,b,*], Inakwu O.A. Odeh [b], Thomas F.A. Bishop [a], Marian S. Dunbar [a,b], Tamara M. Shatar [a]

[a] *Department of Agricultural Chemistry and Soil Science, The University of Sydney, Ross St. Building A03, Sydney, NSW, 2006, Australia*

[b] *Australian Cotton CRC, Department of Agricultural Chemistry and Soil Science, The University of Sydney, Ross St. Building A03, Sydney, NSW, 2006, Australia*

Received 17 November 1998; received in revised form 16 November 1999; received in revised form 5 April 2000; accepted 5 April 2000

Abstract

Quantitative techniques for spatial prediction in soil survey are developing apace. They generally derive from geostatistics and modern statistics. The recent developments in geostatistics are reviewed particularly with respect to non-linear methods and the use of all types of ancillary information. Additionally analysis based on non-stationarity of a variable and the use of ancillary information are demonstrated as encompassing modern regression techniques, including generalised linear models (GLM), generalised additive models (GAM), classification and regression trees (RT) and neural networks (NN). Three resolutions of interest are discussed. Case studies are used to illustrate different pedometric techniques, and a variety of ancillary data. The case studies focus on predicting different soil properties and classifying soil in an area into soil classes defined a priori. Different techniques produced different error of interpolation. Hybrid methods such as CLORPT with geostatistics offer powerful spatial prediction methods, especially up to the catchment and regional extent. It is shown that the use of each pedometric technique depends on

☆ Keynote paper Symposium 17, 16th World Congress of Soil Science, Montpellier delivered on August 24th 1998.

* Corresponding author. Department of Agricultural Chemistry and Soil Science, The University of Sydney, Ross St. Building A03, NSW 2006 Australia. Tel.: +61-2-9351-3214; fax: +61-2-9351-3706.

E-mail address: alex.mcbratney@acss.usyd.edu.au (A.B. McBratney).

PII: S0016-7061(00)00043-4

the purpose of the survey and the accuracy required of the final product.

Keywords: soil science; soil survey; pedometrics; soil geostatistics; spatial prediction; generalised additive models; regression trees

1. Introduction

> . . . In that Empire, the Art of Cartography reached such Perfection that the map of one Province alone took up the whole of a City, and the map of the Empire the whole of a Province. In time, those Unconscionable Maps did not satisfy and the Colleges of Cartographers set up a Map of the Empire which had the size of the Empire itself. and coincided with it point by point. Less Addicted to the Study of Cartography, Succeeding Generations understood that this Widespread Map was Useless and not without Impiety they abandoned it to the Inclemencies of the Sun and of the Winters. In the deserts of the West some mangled Ruins of the Map lasted on, inhabited by Animals and Beggars, in the whole country there are no other relics of the Discipline of Geography. Jorge Luis Borges, El hacedor

Effective soil management requires an understanding of soil distribution patterns within the landscape. This is particularly so as the knowledge of soil allows land use to be kept within its constraints, and thus enable wise decisions regarding land use by planners and policy makers. Conventionally, soil survey can be considered as inventories of soil, including field description and laboratory analysis and subsequent classification and mapping. The ultimate products of soil survey are soil maps, incorporating reports on and interpretations of the soil mapping units or series. Traditionally, soil management and land-use planning have been the main broad aims of soil survey at all scales. However, with increasing concern on environmental issues related to our planet, soil survey has moved from its traditional subjective conjecture to more quantitative modelling with accompanying accuracy and uncertainty issues.

Traditional soil survey may be thought of as a modelling exercise involving scientific methods and an element of art (Wilding, 1985). Generally, a field survey involves the development of a mental model, which relates the soil with landform conditions, followed by formulation of hypotheses, which are then tested by ground-truth survey. The model could be revised, reformulated and a new soil–landscape portrait could emerge. This is analogous to a painter transferring a mental image of a subject to the canvas (Wilding, 1985).

Conventional soil survey methods have in the past been criticised, perhaps justifiably, for being too qualitative in character. In response to these criticisms, quantitative models have been developed, especially within the last 30 years or so, which are being used to describe, classify and study the spatial distribution patterns of soil in a more objective way. These quantitative methods enable precise statements about the soil to be made. The methods are collectively categorised in the emerging field of soil science known as *pedometrics*. What is pedometrics sensu stricto? Webster (1994) gave a formal definition: ''Pedometrics is a neologism derived from the Greek roots, pedos [soil] and metron [measurement], and is formed and used analogously to other words such as biometrics, psychometrics, econometrics, chemometrics and the oldest of all geometrics''. The definition covers two main ideas. First, the metric part has been restricted to quantitative mathematical and statistical methods, and the soil or pedo part corresponds roughly to that branch of soil science we call pedology.

Webster (1994) also suggested an alternative problem-oriented definition, which he paraphrased as ''soil science under uncertainty''. In this sense, pedometrics deals with uncertainty in soil models that are due to deterministic or stochastic variation, vagueness and lack of knowledge of soil properties and processes. Thus, mathematical, statistical and numerical methods could be applied to resolve the uncertainty and complexity inherent in the soil system, including numerical approaches to classification, which deals with supposedly deterministic variation.

Pedometrics is not new, as mathematical and statistical methods have been applied to soil studies since at least the 1960s. However, it was first formally recognised as a different branch of soil science to traditional pedology just over a decade ago. Over time, the use of computers has increased in both fields, and the difference between the two has decreased, and in some cases overlapped (as shown in Fig. 1). Traditional pedology has, of necessity, become more quantita-

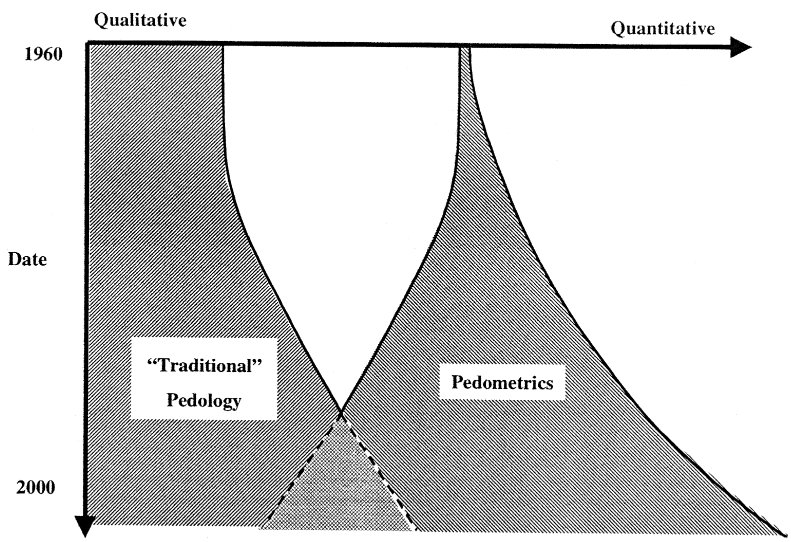

Fig. 1. A time line of the growth of pedology and pedometrics.

Fig. 2. Postage stamp issued for the 9th Congress of Soil Science, Adelaide, Australia, 1968.

tive through the increased use of computerised soil information systems. Pedometrics has developed quantitative methods, which attempt to account for conceptual pedological models of soil variation. Now there is a strong and growing overlap and synthesis between traditional pedology and pedometrics. Although technology has changed dramatically since the late 1960s, the promise of computerised soil information and associated pedometric techniques was alluded to in a postage stamp issued 30 years ago for the 9th Congress of Soil Science held in Adelaide in, Australia (Fig. 2).

At the same time as the increase in the use of pedometric techniques there has been a corresponding increase in the demand for quantitative information. This information is required for different purposes, and at a variety of extents and resolutions (Table 1). This paper will review a range of pedometric techniques that are being used for soil survey data analysis at present, with case studies. First, some thoughts on how increasing availability of ancillary data have provided new vistas for soil inventories.

Table 1
New demands for quantitative soil information at various resolutions and extents

	Demand	Typical resolution	Typical extent
National/Continental/Global	Global climate/food security models	>2 km	>200 km
Catchment/Landscape	Environmental input/watershed	20 m–2 km	2–200 km
Local	Precision agriculture	<20 m	2 km

Table 2
Examples of sources of exogenous data for soil inventory

Carrier	Sensor / Scanner	Sensed data	Land / Soil information
Air-borne: (aeroplanes and balloons)	Photogrammetric / videographic cameras	spectral imageries (albedo)	DEM, crop growth, vegetation
	SLAR	radiance energy	moisture, landscape
	NIR	radiance energy	moisture, clay, etc.
	Gamma-radiometer	gamma radiation	K, U and Th isotopes
Space-borne	Multi-spectral thematic mapper, etc.	radiance energy and albedo	vegetation, moisture
Landsat: SPOT Satellites	High resolution visible (panchromatic and multispectral)	radiance energy and visible spectral imageries (albedo)	DEM, landscape, vegetation and moisture
Landsat: NOAA Satellites	Advanced very high resolution radiometers(AVHRR)	surface reflectance, surface temperature	soil, vegetation, moisture (drought)
Proximal Sensors and Scanners:	Time domain reflectrometer (TDR)	apparent dielectric constant	moisture
(humans, vehicular, etc.)	Electromagnetic induction (EM)	apparent conductivity	salinity, clay and moisture
	Gamma-radiometer	gamma radiation	K, U and Th isotopes
	Soil chemical sensors	soil chemical composition	soil fertility, organic carbon, etc
	Ground-penetrating radar (GPR)	radiance energy	soil conductivity, soil layers, etc
	Yield monitors	crop yield (ha^{-1}) and quality	yield data associated with soil variability
Humans	Various, human	existing ground-truth data and maps, etc.	e.g., topographic, geologic, vegetation

2. The increasing richness of data sources

The task of a soil survey is to provide soil information for either general purposes or for a specific use, with the latter task being dominant. In the past, soil scientists based their approach on the qualitative analysis of the landscape either by physiographic analysis or by aerial photographic interpretation or both. These were all attempts to enrich the soil information through the use of exogenous data. Due to increasing awareness of environmental pollution and associated problems by the general community, quantitative soil information is now required to enable more precise statements on the status of the environment to be made. Pedometric techniques have been developed to meet these requirements. Basically, these techniques stem from the classical approach when the soil scientist generally would study the climate, geology, geomorphology, vegetation, and land use and land use history, prior to any soil survey. Lately, technical advances have provided us with a wealth of new environmental data sources, which are summarised in Table 2.

Air-borne and space sensors are now generating terabytes of data on a daily basis. The air-borne sensors vary from the traditional aerial photography to air-borne videography (Table 2). The space sensors generally consist of series of satellites, many of which were originally designed and launched for general studies of the various earth resources. Increasingly, proximal sensors are now being used for the purpose of research and development for soil-specific crop management (McBratney and Pringle, 1997).

There is a variety of in situ soil-measuring systems, and they may determine variability internally or externally from the soil. For example, internal detectors include the neutron probe and time domain reflectometry (TDR), both providing an estimate of soil moisture. The soil can be analysed from an external, or remote position, and gammaradiometrics (e.g., Cook et al., 1996), ground-penetrating radar, or electromagnetic induction (EM) (Lesch et al., 1995) provide another quantitative measurement for use in pedometrics. There has been an attempt in recent years to obtain the quantitative data in a continuous fashion across the landscape, and at the catchment and continental extent. This may be provided by satellite digital images (Curran, 1998). At the local extent, newly developed sensors attached to mobile machinery are providing on the go continuous measurements of a field. Examples of this include the conductivity cart (Triantafilis and McBratney, 1998), which measures the EC of soil. We now examine pedometric techniques that are used for the analysis of the sensed data for more efficient and inexpensive, quantitative soil inventories.

3. Pedometric techniques, including newer methods

There are a variety of techniques available for analysing the spatial distribution of soil, with the most common methods in use at present being geostatistics, classical statistics and the combination of the two. Each of these will be discussed, but first it is worth mentioning what should be considered as some of the pioneering efforts in pedometrics, and the two more recently emergent techniques-fuzzy sets or fuzzy logic, and pedodiversity.

3.1. Numerical classification

The main purpose of any classification is data reduction whereby a complex system represented by some set of data is made more explicit. Almost all soil surveys are accompanied by some forms of grouping, be it the so-called ‘‘natural’’ system of classification or the technically interpretative form. However, these classifications are composed of mutually exclusive classes in order to conform to the discontinuous soil variation embedded in the traditional soil surveys. But, soil variation is more continuous than discrete. The pioneer work in pedometrics, the computer-based numerical classification (Hole and Hironaka, 1960; Moore and Russell, 1967; de Gruijter 1977), was designed to address this limitation among others.

Since the early studies in numerical soil classification, its application has broadened considerably. Spatial and geostatistical analysis (e.g., McBratney and Webster, 1981), soil database management (de Gruijter, 1977), discriminant analysis for improved identification (Webster and Burrough, 1974) are some of the application of numerical classification in soil science. While the applications of numerical soil classification to soil studies are, to some extent, based on continuous representation of soil in space, their results are still interpreted in terms of discontinuous classes (Odeh et al., 1992). There is also lack of any spatial coherence for the classes to be mapped (Burrough, 1986). Moreover, numerical classification methods are based on linear interrelationships. Recent advances are based on fuzzy sets for optimised prediction quality of the resulting classification, and which take cognisance of the continuous nature of soil variation (McBratney and Odeh, 1997).

3.1.1. Fuzzy sets and fuzzy logic

Many quantitative models in soil science are characterized by multiple, usually conflicting attributes; subjective uncertain conception of preferences (of the modeller), and uncertain, imprecise information on data used in the models. While classical statistical methods, based on determinism and involving imposition of some specific field designs and treatments to minimize the effect of uncertainty (Fisher, 1954), have been quite successful, uncertainty is considered a removable artifact which should disappear with increasing knowledge (Bardossy and Duckstein, 1995). However, uncertainty, imprecision and ambiguity are inevitable or inherent parts of natural systems such as soil. In many cases, the complexity of the models stems from overemphasized precision, which does not always mean greater truth. Applications of fuzzy-set theory in soil science, especially in soil-pattern recognition, deal with uncertainty that is mainly due to imprecise boundaries between categories. Even in cases where the model is considered to be precise, fuzziness may be a concomitant of the complexity (Kandel, 1986).

The first application of fuzzy-set theory in soil survey was principally for classification. In general, the purpose of classification is to reduce a complex system, represented by some sets of data, into explicitly defined classes. Two different but complementary approaches to grouping individuals into fuzzy sets or classes are now used in soil science. The first is based on *fuzzy c-means* (FCM, also known as *fuzzy k-means*) (Bezdek, 1981) partitioning of observations in multivariate space into relatively stable naturally occurring groups. De Gruijter and McBratney (1988) modified the FCM algorithm for improved predictive classification by providing for membership to an extragrade class, which McBratney et al. (1992) termed, *continuous classification with extragrades*. The modified objective function, defining the within-class sum-of-square errors J_{E}, is expressed as:

$$J_{\mathrm{E}}(M,c) = \alpha \sum_{i=1}^{n} \sum_{j=1}^{c} m_{ij}^{\varphi} d_{ij}^{2} + (1-\alpha) \sum_{j=1}^{c} m_{i}^{*} \sum_{j=1}^{c} d_{ij}^{-2} \tag{1}$$

where c is the number of classes, n is the number of individuals or pedons; m_{ij} is the membership of an individual i in class j; φ is the fuzziness exponent $(1 < \varphi < \infty)$; d_{ij} is the character space between the feature value of an individual, i, and the feature centroidal value for class j; α is the parameter that determines the mean value of m_i^*, which is the membership value of an individual, i, in the extragrade class. The application of the FCM algorithm for soil classification is gaining wide acceptance in the soil science community (see examples in de Gruijter et al., 1997).

Another approach in using the fuzzy sets for soil classification is based on what is termed the Semantic Import model (SI) (Burrough et al., 1992), whereby a membership function is defined without reference to the data, but with the class limits specified. The class limit specifications are defined a priori based on expert knowledge or conventionally imposed definitions before multi-attribute individuals are allocated on the basis of how close they match the requirements of the classes. An example of classes that are defined a priori is the FAO Framework for Land Evaluation (FAO, 1976).

Fuzzy-set theory is required for land evaluation, as defined in the FAO framework, because basic soil information used for land evaluation is mostly described by seemingly vague terms such as ‘‘poorly drained’’, ‘‘slightly susceptible to soil erosion,’’ ‘‘moderate nutrient availability’’ etc. (Burrough, 1989). Even when these terms are defined precisely, the qualitative ambiguity remains. Usually, the land evaluator’s aim is to produce a set of clearly defined classes of land qualities based on specified land use requirements (FAO, 1976).

These subsequently provide the means of transferring information about the soil and its use. As land qualities are complex attributes that are derived from land characteristics such as topography, soil, water, or biological and human activity, subsequent Boolean logical operations in the process of land evaluation tend to throw away much of useful information (Burrough et al., 1992).

Chang and Burrough (1987) were the first to apply fuzzy sets and logic to land evaluation. Burrough (1989) and Burrough et al. (1992) used fuzzy classification to determine land suitability for various purposes. All these studies involved the complex combination operator as defined for fuzzy systems (McBratney and Odeh, 1997). Burrough et al. (1992) reported that fuzzy methods were much better in producing suitability classifications than Boolean methods. Fuzzy applications in the analysis for specific soil qualities is becoming common, for example, soil fertility (Dobermann and Oberthur, 1997) and soil pollutants (Hendricks Franssen et al., 1997).

The use of fuzzy sets and fuzzy logic is advantageous in soil survey as it allows for the spatial variability of soil to be determined in a continuous way as it is in the real world. Recent advances in the use of fuzzy sets in soil science involve three-dimensional rule-based modelling in a continuous manner (Ameskamp, 1997). De Gruijter et al. (1997) provide a full account of examples of fuzzy sets in soil science.

3.2. Pedodiversity

Pedodiversity analysis is a relatively new technique (McBratney, 1992), and is a new way of looking at variability of soil. Ibanez et al. (1998) took a bold step in determining the pedodiversity of the continents. This was based on the available soil maps and Shannon's index, which measures the richness and evenness of the major soil groups within each continent. Using the data from Ibanez et al. (1998) we have plotted the diversity against the area for each continent, to produce the plot in Fig. 3. For example it can be seen in Fig. 3 that, on a unit areal basis it seems Australasia, one of the oldest continents, is the least diverse, while the younger continents, such as South and Central America, are the most diverse. Pedodiversity may be considered from a taxonomic or functional point of view. It is expected that pedodiversity will become an important index of soil quality and its estimation will become an integral part of soil-resource assessment.

3.3. Pedometric methods used for land evaluation and soil quality assessment

Although soil-quality assessment is often used as a misnomer for land evaluation, both could be regarded as an interpretative phase of soil survey. Eventhough the focus of this paper is primarily on analysing soil data obtained

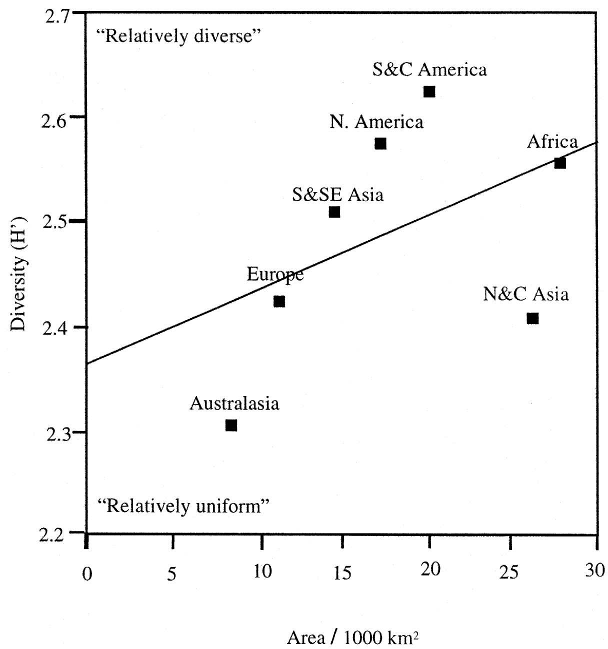

Fig. 3. Diversity versus area of continents, (from Ibanez et al., 1998) (with best fit line of ''average'' diversity).

from soil surveys and transforming them into either digital or map form, a number of pedometric techniques used for land evaluation and soil quality assessment are worth mentioning. A widely acceptable definition of soil quality is given by Doran and Parkin (1994) which states that soil quality is ''the capacity of the soil to function within ecosystem boundaries to sustain biological productivity, maintain environmental quality, and promote plant and animal health''. This definition emphasises soil processes and ecological functions of soil, including agricultural productivity and human health. Mostly pedometric techniques are used for soil-quality assessment, which could be single issue-, single process- or function-based. Pedo-transfer functions (Renger, 1971; Vereecken et al., 1989) have largely been used for such process- and function-related analyses. Many examples of process-orientated soil-quality assessment using pedometric techniques are provided in Finke et al. (1989). Bouma et al. (1998) provide a global perspective of land-quality assessment and the world food supply.

3.4. Generic pedometric techniques

Focusing on the basic types of pedometric techniques used for spatial prediction of soil, and hence in a general sense soil survey, two generic techniques come to mind. The first, the classical approach, collectively referred to here as the CLORPT methods, and the second, geostatistical methods. To these we add a third group, hybrid methods, which are some combination of techniques from the two generic techniques above. The combination is carried out optimising the prediction of soil properties. These are treated in Section 3.5. The older and the emerging techniques and the style of each are summarised in Fig. 4.

3.4.1. CLORPT techniques

The CLORPT methods are based on the empirical-deterministic models that originated from Jenny's (1941) ''Factors of Soil Formation'' (Fig. 3). Jenny's (1941) state-factor equation is expressed as:

$$S = f(\mathrm{CL}, O, R, P, T) \tag{2}$$

where S is some soil properties as a function of the state factors: CL as climate, O as organisms, R as relief, P as parent material, and T as time. Soil spatial variability is therefore considered as being causative realisations of the complex combinations of soil-forming processes as influenced by the soil forming factors. The CLORPT function (Eq. (2)), earlier in the 19th century, stimulated numerous studies. Much of the earlier studies, and indeed some recent ones, were based on general and bivariate-simple linear regression (e.g., Furley, 1971;

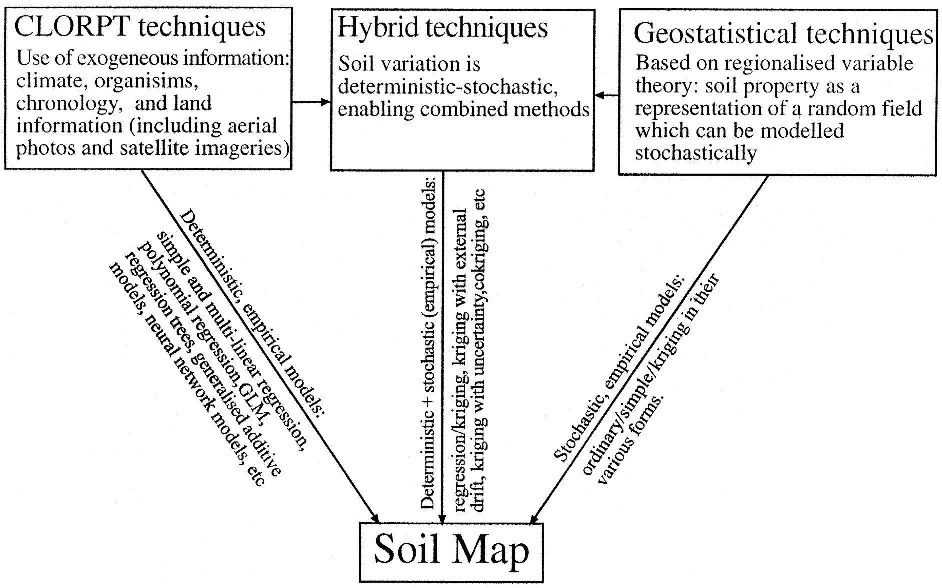

Fig. 4. The generic pedometric techniques, their hybrid and their styles with respect to mode of spatial prediction.

Moore et al., 1993), although multiple polynomial regression models were applied by Ruhe and Walker (1968). The realisation that many of these studies do not accommodate the non-linearity in the relations has led to recent application of the more robust methods such as generalised linear models (GLM), generalised additive models (GAM) and regression trees (RT) (Odeh et al., 1994, Gessler et al., 1995). Another developing method is the artificial neural network (NN) model, which is a non-parametric modelling technique, which mimics the neurons of the brain (Venables and Ripley, 1994). The networks are composed of processing units, or neurons, which are organised into layers, i.e., input, hidden and output layers. Neural networks are ''trained'' by the data used to create them. Inputs are distributed from the input units into the hidden layer where the inputs are weighted and summed, a bias added and a fixed function of the result taken. The neural network fits iteratively, minimising differences between predictions and actual values during training (Venables and Ripley, 1994). But, the question to ask is while the classical models or the more robust methods may take care of the deterministic relations, do they account for spatial autocorrelations of the soil properties, especially at the local level? To answer this question, the pioneer pedometricians initiated the application of *geostatistics* (which was primarily developed for the mining industry), the compendium of which is given below.

3.4.2. Geostatistical techniques

Geostatistical methods are based on the *theory of regionalised variables* (Matheron, 1965), which allows us to consider spatial variability of a soil property as a realisation of a random function represented by a stochastic model. The geostatistical method of spatial interpolation is termed *kriging.* The first major applications of ordinary kriging (OK) in soil studies emerged in the early 1980s (e.g., Burgess and Webster, 1980). Since then, ordinary kriging has been widely used in various sub-fields of soil science: soil reclamation (Samara and Singh, 1990), in soil classification (e.g., Odeh et al., 1992; Burrough et al., 1992), soil salinity study (Bourgault et al., 1997) and soil pollution studies (e.g, Hendricks Frassen et al., 1997), etc. Major limitations of the univariate geostatistical technique of kriging are due to the assumptions of stationarity which are not often met by the field-sampled data sets and, of course, the often cited requirement of large amount of data to define the spatial autocorrelation. However, with increasing availability of ancillary information, the lack of adequate samples has been partially solved. The univariate usage of kriging is also limiting in situations of complex terrain where the soil-forming processes are themselves complex. In such situations, there is the need to model both the structured and the spatially dependent components of the soil variable. Also there are economic and logistic reasons for including the ancillary influencing soil variability, especially if the latter are more readily and cheaply available. As both the soil and the exogenous factors are multivariate, the most obvious choices are appropriate combinations of multivariate/univariate analysis using the CLORPT factors and the geostatistical methods. These combinations constitute the hybrid techniques (Fig. 4).

3.5. Hybrid techniques

The hybrid techniques for soil survey are based on various combinations of the geostatistical and multivariate or univariate CLORPT methods. Let us suppose that a data vector describing a soil property is a random variable Z, determined at locations in a region, $X = x_1, \ldots, x_N$, and consisting of three components as

$$Z(x) = m + Z_1(x) + \varepsilon(x) \tag{3}$$

where m is the local mean for the region, $Z_1(x)$ is the spatially dependent component and ε the residual error term, spatially independent. Now there may be situations where m is varying and dependent on some exogenous factors such as the CLORPT factors. In other words it is deterministically related to some causative factors (in geostatistics parlance, the variable is said to exhibit a trend). Wherever trend exists, the ordinary univariate kriging is inappropriate. Several methods have been designed to accommodate the trend.

3.5.1. Universal kriging

Universal kriging (Matheron, 1969) has been the commonly used method to accommodate the trend or the ''changing drift'', as it is sometimes known, in a soil variable. The universal kriging is a combination of the standard model of multiple-linear regression and the geostatistical method of ordinary kriging (Webster, 1994), which is also analogous to combining CLORPT methods with the univariate kriging using the geographical coordinates as determining the drift. More recently, a more advanced approach, the Intrinsic Random Function of Order k (IRF-k), has been used to accommodate the varying nature of the trend in a regionalised soil variable (McBratney et al., 1991). The term k represents the order of polynomial trends — k = 0 means constant drift, and the IRF-k is equivalent to ordinary kriging system of equations; if $k = 1$, we have linear drift; $k = 2$ yields quadratic drift. But, where there is not trend but deterministic relationships with some known or readily available and inexpensive covariates (CLORPT factors) or other easy-to-measure soil variables, cokriging has played a major role in efficiently predicting the target soil variable (Stein et al., 1989; Odeh et al., 1995). Universal cokriging is also possible when considering the trend and covariation with one or more secondary variables (Stein et al., 1988).

3.5.2. Cokriging

Cokriging is the multivariate extension of kriging that allows the inclusion of more readily available and inexpensive attributes in the prediction process. There are many instances in soil survey where the CLORPT factors such as topography, time, variable parent material, etc, are easily discernible or are either readily available and/or cheap to obtain. The most efficient way to predict the expensive-to-measure target soil variable, the variation of which is affected by the CLORPT factors, is to use the factors in cokriging the target soil variable-sampled at fewer locations, into dense grid nodes. This is termed as *heterotopic cokriging* (HCOK) (Wackernagel, 1994) in comparison with *isotopic cokriging* which requires that data on both the target variable and co-variables be available at all sample locations. A variant of both is the *generalised cokriging* (Myers, 1982) that involves simultaneous prediction of all the correlated variables into more dense locations. Heterotopy can either be complete or partial (Wackernagel, 1995). The complete case is the case where the covariates and the target variable do not share any common locations. A third is termed *collocated cokriging*, whereby covariates are available at all interpolation nodes, eventhough the target variable is available at only a few locations. This is often the case with using exogenous variables especially landform attributes derived from DEM (Odeh et al., 1995) and satellite imageries for predicting the target soil variable. The partial heterotopy involves cases where there is some coincidence of the locations of the target variable and the ancillary variables. The latter is often the case when other soil covariates are used (McBratney and Webster, 1983).

3.5.3. Regression kriging

Regression-kriging (RK) is another hybrid method that combines either a simple or multiple-linear regression model (or a variant of GLM, GAM and regression trees) with ordinary, or simple, kriging of the regression residuals (Odeh et al., 1995; Goovaerts, 1997). The assumption here is that the deterministic component (m in Eq. (3)) of the target (soil) variable is accounted for by the regression model, while the model residuals represent the spatially varying but dependent component (Z_1 in Eq. (3)). If the exogenous variables used in the regression equation are available at denser locations than the target variable, the equation can then be used to predict the m onto those locations. The Z_1 can also be predicted to the same locations by ordinary kriging system of equations, and then added to the m to obtain Z^*. Odeh et al. (1995) and Odeh and McBratney (2000) have demonstrated the superiority of RK to other prediction methods such as ordinary kriging, universal kriging, multiple-linear regression and cokriging. A variant of RK is *kriging with uncertainty* (Ahmed and DeMarsily, 1987). Kriging with uncertainty introduces the regression residuals (as representing model uncertainty) into the kriging system, which is then used to predict the target soil variable (Knotters et al., 1995). This reduces the extrema of the target variable and therefore produces a smoother function of the predicted values. While Odeh et al. (1995) found kriging with uncertainty not as good as regression-kriging, they nevertheless reported it to be better than ordinary kriging or cokriging alone.

3.5.4. Kriging with external drift

Kriging with external drift (KED) is a somewhat different hybrid technique which integrates the universality conditions into the kriging system using one or more of the ancillary drift variables (Wackernagel, 1995). It is similar to universal kriging, but it uses an ancillary variable to represent the trend (Goovaerts, 1997). These variables could be digitised covariates derived from digital elevation model (DEM), or rainfall data or scanned images. The universality conditions need to be known not only at the sampled locations but also at the prediction locations. KED has not been widely used in soil science but, as remotely sensed data become more readily available, it may well be the method to be used along with regression-kriging (Odeh and McBratney, 2000), especially for soil inventory at the regional/catchment (20 m–2 km) resolution.

3.5.5. Factorial kriging

Wackernagel (1988) and Goovaerts (1992) first introduced factorial kriging (FK) to soil science. The method involves a combination of classical multivariate analysis and geostatistics in which multivariate variogram modelling, *principal component analysis* and cokriging are carried out on a multitude of soil variables. The assumption behind FK is that many of the soil variables have the same communality (as defined by soil forming processes influenced by the CLORPT factors) that enables principal component (or its variants) analysis of the variance–covariance matrices of the variables which are themselves associated with spatial scales (Goovaerts, 1992). Prior to this, McBratney and Webster (1981) transformed their soil data by principal component analysis before embarking on spatial analysis — a form of factorial kriging. Also, closely related to factorial kriging is the application of *fuzzy set theory* for classification of the soil into continuous classes. Various combinations of fuzzy logic and classification with kriging have been adopted for soil mapping (see examples in de Gruijter et al., 1997). New methods involving *fuzzy-kriging integral* and *fuzzy inference* (e.g., Pham, 1997) are emerging that could prove useful for optimal spatial prediction of soil variables. A major problem with FK is the linearity assumption which is often not met by many soil variables, but the problem can largely be solved by some transformations or by using *correspondence analysis* or even by using fuzzy integral as mentioned above. In general, FK, and indeed many of the multivariate statistical techniques, are mainly exploratory tools for revealing the correlation structure of the soil variables but are not spatial prediction techniques per se.

4. Three case studies at spatial extents of interest

In Australia, one of the pressing needs in environmental modelling and land use planning is soil information. Good quality soil data are now required for accurate planning of the land resource use, minimising environmental impact and monitoring of the environment. We provide examples or case studies in Australia, which illustrate how the necessary soil data are being acquired at different spatial extents for their respective purposes.

Three spatial extents and the associated resolutions are considered here. Each of these extents has different implications for the soil information required. At the local extent information is needed for agricultural management, conservation and precision agriculture. As the area is increased to the catchment or regional extent soil knowledge is required for environmental monitoring, and for studying changes brought about by a disruption of ecosystems. At the national and continental extents, which are not treated here, soil information is required to allow study of the global climate and to model food production and supply.

The case studies provide examples of different pedometric techniques at different extents and their associated resolutions. The locations of the case studies are shown in Fig. 5. The areas are all located in the northwest of New South Wales, Australia, an area that is primarily of arable agriculture.

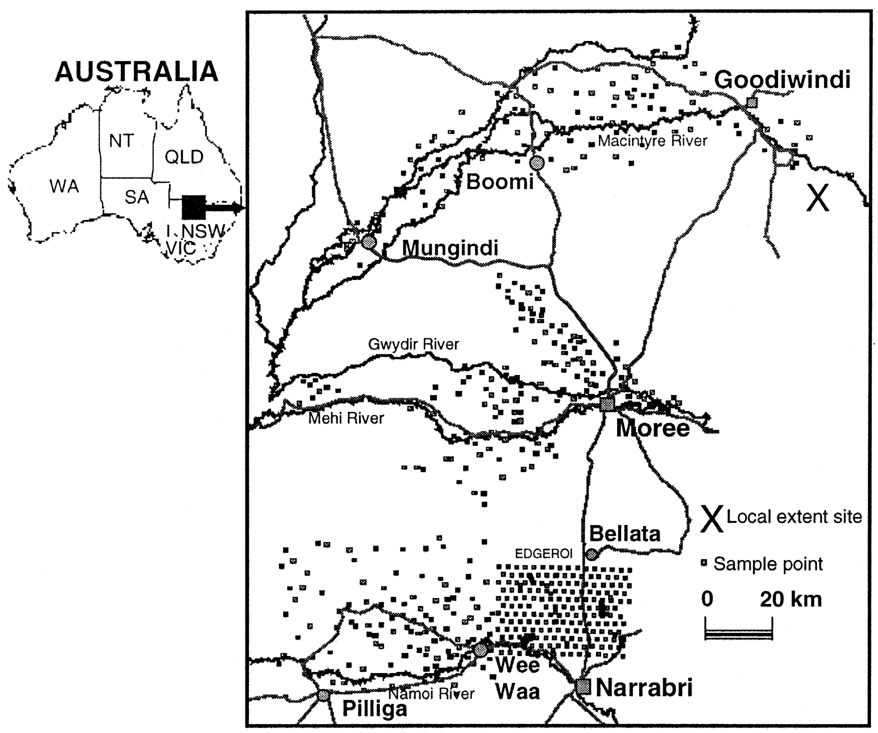

Fig. 5. The locations of the three study areas within Australia and New South Wales, and the sampling pattern.

4.1. Local extent

The emerging field of precision agriculture has brought about increased interest and the need for the development of methods for analysing the within-field variability of crop and soil attributes and their management. The ultimate requirement of these methods is real-time ‘‘on-the-go’’ soil and crop sensing which is at varying stages of development — the most developed being the grain-yield sensors. Furthermore, advances in remote sensing techniques and the increasing availability of high-resolution, multi-spectral satellite images (as high as 2 m pixels) and new data sources, such as radar imagery, have provided new vistas for within-field monitoring of soil and crop attributes. This section deals with how these new exogenous data could be used for producing more precise soil maps at high resolution, required for precision agriculture. This is especially pertinent to the Australian wheat belt where soil information is scarce to non-existent. In the examples here, we compare the quality of information produced using different ancillary data and different pedometric techniques.

4.1.1. The study site and the available ancillary information

This study was performed on a 42-ha field near Moree in Northern NSW in Australia. Ninety-five soil samples were taken at a depth of 15–30 cm and the clay content measured using the hydrometer method. The sampling layout is shown in Fig. 6. In carrying out the comparative study, 10 samples were randomly selected as the validation set while the remaining 85 samples were used in different prediction methods with varying combinations of ancillary variables.

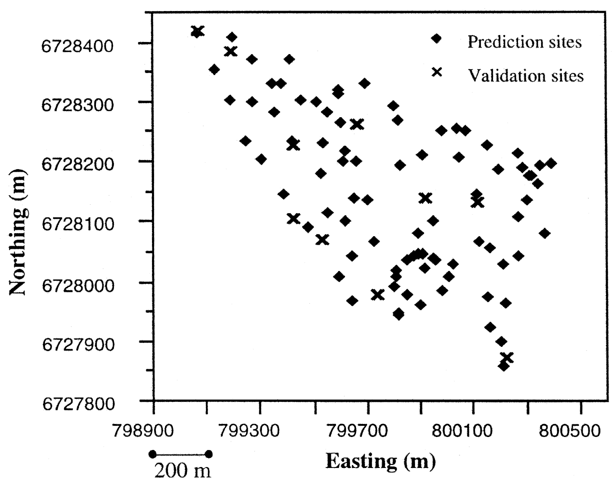

Fig. 6. Prediction and validation sites within field W80.

Ancillary data acquired for the field are:

- electromagnetic (EM) data from a Mobile Electromagnetic Soil-sensing System (MESS) with GPS
- sorghum yield data for 1997, obtained using an AgLeader yield monitor with GPS
- point elevation, derived from a real-time (Ashtech) kinematic GPS

The EM data were obtained using both the EM31 and EM38 in both the horizontal and vertical mode. Thus, there are four different data values at each point. Whereas the EM38 measures bulk soil conductivity in the top 1 m in the horizontal and top 2 m in the vertical mode, the EM31 measures bulk soil conductivity of the top 4 and 7 m in the horizontal and vertical modes, respectively (Rhoades and Corwin, 1981; Corwin and Rhoades, 1990). Apart from soil solutes, EM measurements are also affected by soil texture, mineralogy, structure and moisture (Rhoades, 1992). Hence, the MESS could generally be considered a crude real-time ‘‘on-the-go’’ sensor for several soil attributes.

A crop yield map reflects the interaction of the soil and atmospheric environment with the plant. For this reason, one or more soil attributes have been used to predict crop yield (Sudduth et al., 1997; Shatar and McBratney, 1999). It was therefore considered reasonable to examine the reverse and use the yield data for predicting soil properties. This is especially relevant in the case of clay content as it is relatively highly correlated with other soil properties, such as soil drainability, moisture content, nutrient status, etc., which affect plant growth and therefore yield.

The use of elevation data and derived terrain attributes for the prediction of soil properties and the rationale behind it has been reported in the literature (Bell et al., 1995; Moore et al., 1993). Several studies have used only simple or multiple-linear regression techniques (MLR) to predict soil properties from terrain attributes (Tomer et al., 1995; Moore et al., 1993). Recently emerging and more advanced pedometric techniques, such as heterotopic cokriging and RK, have demonstrated increased benefits in terms of resolution and precision when digital terrain models are used (Odeh et al., 1995; 1996). Our aim here was to compare several of these methods to determine the most precise and least biased of the methods.

4.1.2. Quantitative spatial prediction and results

The raw ancillary data was interpolated onto a common 2-m grid prior to their use for soil prediction. This may be thought of as a downscaling process (Bouma and Hoosbeek, 1998). For the raw elevation data the TOPOGRID tool in Arc Info (ESRI, 1997) was used to generate a 2-m raster DEM. The GRID module of Arc Info was then used to derive the terrain attributes of slope,

Table 3
Summary statistics of prediction and validation sets of clay content (dag kg^{-1})

	Prediction set	Validation set
Maximum	60.18	62.07
Minimum	32.39	34.34
Mean	45.34	48.28
Standard deviation	6.93	8.43
Number of samples	85	10

upslope area, profile curvature, plan curvature (Zeverbergen and Thorne, 1987) and compound topographic index.

In the case of the EM and yield data, local block kriging based on an exponential variogram (Minasny et al., 1999) was used to predict onto the 2-m grid. For block kriging of the EM data, a 5-m block size was used to smooth out both positional errors from the GPS and instrumental errors. A 20-m block size was chosen for the yield data because at this block size a balance was reached between minimising the uncertainty in the yield measurements and over-smoothing of the yield data. If the size of the block is too large, over-smoothing occurs in the yield data to the point where site specificity in yield measurements is lost (Whelan and McBratney, 1999).

Table 3 shows the summary statistics of the prediction and validation set. Both sets have similar range, mean and standard deviation.

Correlation between the clay content in the prediction set and the ancillary variables are displayed in Table 4. Note that the upslope area was log-transformed to obtain a more normal distribution than exhibited by the raw data.

Table 4
Correlation coefficients of ancillary information with clay content at 15–30 cm depth

Ancillary data	Correlation coefficient
Elevation	−0.358
EM31-vertical mode	−0.188
Compound topographic index	0.162
Sorghum yield	−0.153
EM31-horizontal mode	−0.107
EM38-vertical mode	0.087
EM38-horizontal mode	−0.078
Upslope area[a]	0.071
Plan curvature	0.041
Slope	0.024
Profile curvature	0.001

[a] Upslope area was log transformed to give it a normal distribution.

To determine the benefits of the different ancillary data for soil map creation they were divided into three groups for the prediction of clay content; (1) digital terrain attributes, (2) EM data, (3) yield data. In addition they were considered all together as one group.

The following pedometric techniques were tested in combination with each different group of ancillary variables:

- Regression techniques: MLR, GLM, GAM, RT and neural networks (NN).
- Geostatistical techniques: ordinary kriging, heterotopic cokriging.
- Hybrid techniques: RK (regression techniques as mentioned above combined with ordinary kriging of their residuals).

No trend was evident in the clay content data so geostatistical techniques such as IRF-*k* kriging or kriging with external drift were not performed.

Regression methods were performed using S-PLUS Statistical Software (Statistical Sciences, 1995). A feed-forward neural network with one hidden layer was used. Kriging and co-kriging were performed using ISATIS (Geovariances, 1997). All the techniques are described above.

To test for prediction performance, the various data sets were used to predict clay content at the validation sites using the different pedometric techniques. The prediction quality of the methods was determined based on the root mean square error (RMSE) and mean error (ME) of prediction, respectively expressed as:

$$\mathrm{RMSE} = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}\left(z(x_i) - z^*(x_i)\right)^2\right]} \tag{4}$$

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(z(x_i) - z^*(x_i)\right) \tag{5}$$

where $z(x_i)$ = actual clay content, and $z(x_i^*)$ = predicted clay content. The results are summarised in Table 5.

4.1.3. Discussion

Based on the RMSE values as the criterion for prediction performance, the results are similar to those reported by Odeh et al. (1996). RK generally outperformed the generic geostatistical techniques of kriging and the hybrid technique of co-kriging. As expected ordinary kriging performed poorly due to its reliance solely on the sparsely distributed clay data. Partial heterotopic co-kriging also performed poorly, probably due to the low correlation of clay with the covariates and the problems with fitting a common variogram model to

Table 5
Prediction performance at the local extent: root mean square error (RMSE) and mean error (ME) of of prediction by different pedometric methods and different ancillary data combinations (values in parentheses are the regression kriging result)

Pedometric technique	RMSE (dag kg^{-1})	ME (dag kg^{-1})
Regression technique / ancillary variable		
Regression tree: terrain data	8.88 (8.94)	2.29 (1.67)
Regression tree: yield data	8.65 (9.04)	2.79 (2.11)
Regression tree: EM data	7.93 (7.06)	0.96 (−0.06)
Regression tree: all data	8.19 (7.57)	1.53 (0.15)
Generalised additive models: terrain data	6.16 (6.00)	1.46 (−0.33)
Generalised additive models: yield data	8.34 (8.01)	2.29 (0.80)
Generalised additive models: EM data	7.51 (7.19)	1.81 (0.64)
Generalised additive models: all data	6.90 (7.03)	1.55 (0.33)
Multiple linear regression: terrain data	7.94 (7.27)	3.41 (2.29)
Multiple linear regression: EM data	7.52 (7.11)	1.59 (0.56)
Multiple linear regression: yield data	8.58 (7.58)	2.92 (1.50)
Multiple linear regression: all data	7.82 (7.41)	2.92 (2.29)
Generalised linear models: terrain data	7.35 (6.88)	2.85 (2.05)
Generalised linear models: yield data	8.58 (7.94)	2.92 (1.57)
Generalised linear models: EM data	7.52 (6.75)	1.59 (0.84)
Generalised linear models: all data	8.42 (7.90)	2.31 (1.23)
Neural networks: terrain data	7.85 (7.20)	3.05 (1.58)
Neural networks: yield data	8.40 (7.91)	2.47 (0.93)
Neural networks: EM data	7.64 (7.56)	1.76 (0.61)
Neural networks: all data	7.44 (7.28)	2.38 (1.36)
Geostatistical technique		
Ordinary kriging	8.15	−1.59
Partial-heterotopic co-kriging: elevation	7.74	−1.30
Partial-heterotopic co-kriging: sorghum yield	8.12	−1.31
Partial-heterotopic co-kriging: EM31 vertical mode	7.60	−0.80

the clay data and the covariate data. Clearly from the results (Table 5), the use of GAMs alone or as the regression component of the RK with terrain data is the best technique/ancillary data combination. Artifacts in the clay prediction map correspond to contour banks in the field where the soil was overturned during construction. The terrain data obviously reflect the change in landform structure (Figs. 7 and 8).

Although the simple neural networks used performed quite well, there is potential for improvement through use of more sophisticated neural nets. For example, the number of hidden layers could be increased to incorporate the complexity in the soil/environment interrelations. Regression trees were the poorest performing of the regression techniques in terms of RMSE. This may be due to their predictions being composed of a small number of discrete values equal to the number of terminating nodes. This gives unrealistic prediction

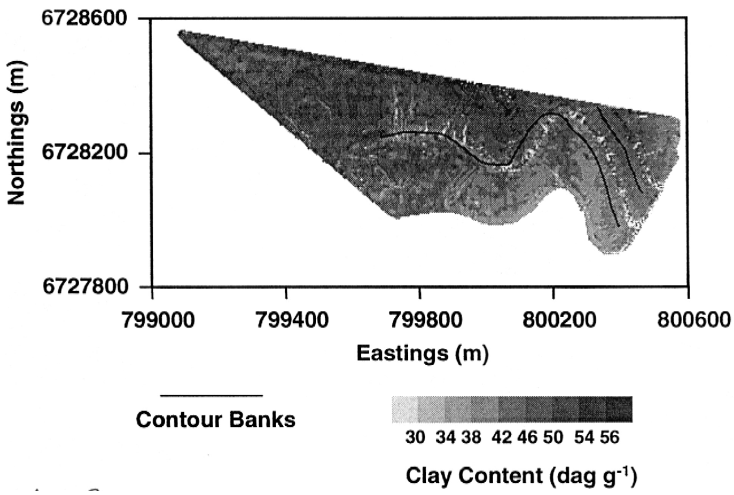

Fig. 7. Prediction of clay content (dag kg^{-1}) using a Generalised Additive Model (GAM).

values, as it would be expected that in reality variation in clay content would be continuous rather than discrete. As a result, there was some loss of information.

When considering the worth of the different types of ancillary data, the elevation data is clearly the most valuable. This is simply because it quantifies topography which is one of the main soil-forming factors. On a local scale (~ 40-ha area) and in a stable landscape, such as the case here, other soil-forming factors such as parent material and climate are quite uniform. Soil

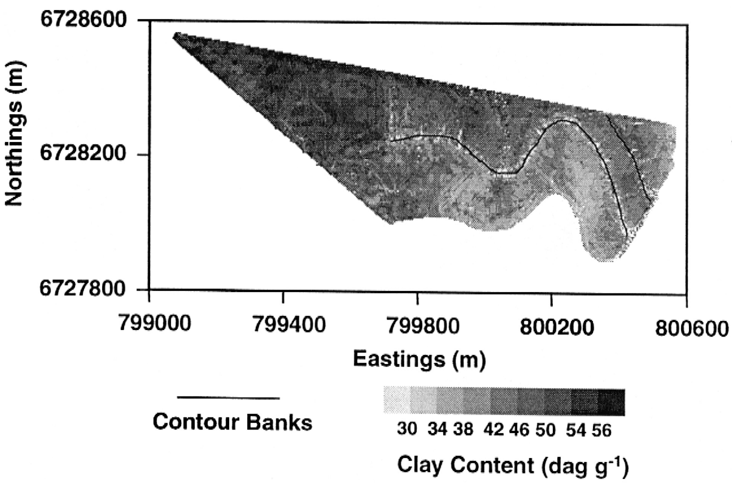

Fig. 8. Prediction of clay content (dag kg^{-1}) using Generalised Additive Model (GAM) and kriged residuals.

variation, therefore, could mainly be attributed to topographic effects. The EM data is a new source of potentially valuable information for predicting soil attributes, as exemplified by its enhancement of prediction precision of clay content. Other than being a surrogate predictor of soil texture (and of course electrical conductivity), it can potentially be used as a predictor of water holding capacity and nutrient status of soil. Using all available ancillary information for prediction produced reasonable prediction values, although it is not advantageous because using terrain data alone produced better prediction values.

Using the yield data was the poorest performer of the ancillary data, even worse than using ordinary kriging alone. Still, with the increasing popularity of precision agriculture it is becoming the most readily available source of ancillary information. Therefore, yield data, as a source of soil information, should not be ignored especially in countries where soil information is extremely limited. The poor performance is probably due to our using only one season of yield data in the prediction model. Further studies using multi-year yield data would better characterise the clay–yield relationship.

In conclusion, at the field scale the RK outperformed the purely geostatistical techniques of kriging and co-kriging. In particular, the best regression components of the RK were the GAMs and neural networks. This is very promising for precision-agriculture practitioners who, in general, have insufficient soil data for accurate variogram estimation and therefore are very much reliant on regression techniques, e.g., GAMs. The most valuable ancillary information was the elevation data which has been extensively used for prediction of soil attributes (e.g., Moore et al., 1993). The EM data performed quite well demonstrating its potential use as a crude real time 'on-the-go' soil sensor. Unfortunately, no remotely sensed data was used for this study. As other sections of this paper demonstrate, remote sensing is a valuable source of ancillary information for soil prediction at the catchment and regional extents. As technology advances result in increase in spatial resolution of remotely sensed digital images, there is no reason the same cannot be said for the within-field soil variation modelling.

4.2. Sub-catchment extent

At the medium or catchment extent the soil survey information of Edgeroi 1:50,000 topographic map sheet was used. The Edgeroi area, located in the lower Namoi valley in north-western NSW, was surveyed using an equilateral triangular grid with a spacing of 2.8 km (Fig. 5) so that site placement was random with respect to landscape or landform (McGarry et al., 1989). The main aim of the survey was to provide more accurate soil information for management and extension purposes, but has now found a more useful purpose in environmental modelling. At each site, the soil was characterised fully to a depth of 2.6 m (where possible). In using the data for our examples here, two separate analyses were carried out, viz.: (1) Classification of soil into existing classes based on the Australian Soil Classification System (Isbell, 1996), and by using exogenous attributes of Advanced Very High-Resolution Radiometric (AVHRR) data and physiographic information. We termed this as allocation. (2) Prediction of topsoil CEC using the AVHRR data. This variable will be an important input into a nutrient flux model for a total catchment management project for the area.

4.2.1. Allocation into pre-existing classes using classification tree

The soil characteristic information was first used to group the soil individuals at the sample locations into various classes at the suborder level of the Australian Soil Classification (Isbell, 1996). The AVHRR coverage of the area, recorded on May 22, 1996, during a fallow period, was obtained to maximise the registration of soil exposure to the satellite radiometer and thus provide optimal data for soil classification and prediction. The AVHRR data is at an approximate spatial resolution of 1 km. Before using the satellite images the data were kriged onto a target grid of 200-m spacing (Odeh and McBratney, 2000). In addition, a physiographical map of the area was digitised from the geologic and geomorphic map produced by Ward (personal communication) onto the same grid matrix. The features used to predict the soil classes were NDVI, mid infrared (MIR), second thermal infrared band (TIR2), elevation (also interpolated onto a 200-m grid), profile curvature, plan curvature and the physiographic data. These features were used in a classification tree model to rapidly predict the classes at the 200-m grid nodes, based on soil classes obtained at 209 sites.

The resulting soil class map is shown in Fig. 9. As shown on the map, two soil types, the Black Vertosol and the Grey Vertosol, are predominant in the area, occurring mainly in the central to western parts of the sub-catchment. To the east are mainly the Rudosols and the Kandosols, interspersed with Vertosols, Red Dermosols and Red Chromosols. Thus, soil variability is much higher in this part than to the west of it. Physiologically there are two main landforms in the study area that are shaping the soil distribution patterns in the sub-catchment: the plains to the west, and the hilly region in the east. Looking at the map in Fig. 9 almost the whole of the plains are classified as Vertosols with few patches of Dermosols and Kandosols and Kurosols occurring in slightly elevated levees and prior-stream formations (McGarry et al., 1989). A variable mantle, related to the baseline geology, underlies the hilly and undulating land to the east. The soil types in this part are mainly a function of geology and topography, hence, the high variability of soil types. Generally, the pattern of soil distribution is almost well aligned to the physiographical setting of the hills and valleys in the area. Eventhough misclassification by our model was shown to be low (15%), the extent to which the classification tree model rapidly classified the soil into classes based on coincidence of soil classes with physiographic unit, topography and AVHRR attributes, will need to be corroborated further.

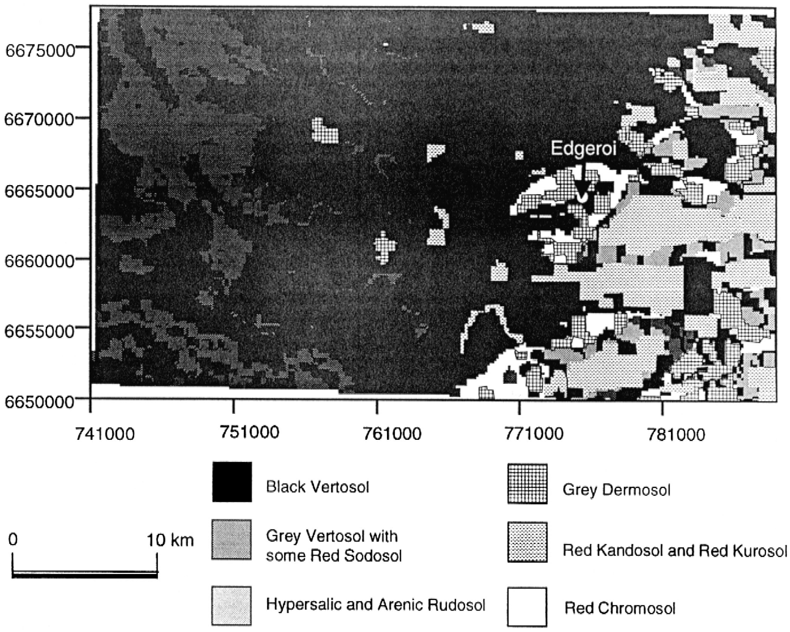

Fig. 9. Predicted soil classes in accordance with the Australia Soil Classification in the Edgeroi area.

4.2.2. Spatial prediction of CEC from AVHRR data

The CEC for the Edgeroi area was also predicted from the AVHRR data onto the same grid spacing as described above. The CEC for the topsoil (0–0.1 m) was determined by McGarry et al. (1989). As was the case with the local extent several spatial prediction models were tested in predicting the topsoil CEC in order to determine the most accurate, based on RMSE (Eq. (4)) and the least biased (ME — Eq. (5)). To test the models, 50 out of a total of 229 CEC records were set aside for validation. The remaining 189 were used in the following prediction models:

- MLR using a combination of AVHRR data and land attributes
- Kriging based on IRF-1
- KED using each of the elevation or AVHRR bands: MIR, TIR and NDVI, as the external drift
- RK, combining MLR with ordinary kriging of the MLR residuals.

As previously explained, RK combines the results of prediction by MLR and those of ordinary kriging of the MLR residuals in order to minimize the uncertainty of the regression model (Odeh et al., 1995).

Initially, all bands of AVHRR data and landform attributes were used to determine the best MLR model. Based on a stepwise regression procedure the following formula was selected as the best MLR model, using the prediction set of 189:

$$\begin{aligned}\text{CEC} = & -1315.71 - 1.56 \times \text{red} - 0.65 \times \text{NIR} - 0.66 \times \text{MIR} + 1.44 \\ & \times \text{TIR2} + 0.30 \times \text{Elavation} + 2837.10 \times \text{Plan Curvature} \\ & + 4364.57 \times \text{Profile Curvature} \end{aligned} \tag{6}$$

As for the RK prediction, the MLR residuals were interpolated onto 200-m grid nodes, following the MLR model prediction onto the same grid nodes. The two values, i.e., the MLR model value and the residual value were summed at each node to give the new predicted CEC value. The aim was to minimise the uncertainty due to MLR model.

Drift identification was carried with each of the predictor variables used in Eq. (6) to test for a common trend with our target variable, the topsoil CEC. It was found that only the MIR, TIR2, NDVI and elevation showed some trend with CEC.

The results of prediction performance by the methods are shown in Table 6. Of all the methods used, KED with elevation as the external drift and IRF-1 produced the least RMSE (66 and 86 mmol$^+$/kg, respectively). The better performance by the two methods is probably due to the linear trend of CEC from east to west, which is well modelled by the intrinsic random function and the elevation. It is obvious that KED, with NDVI as the external drift, was the poorest performer (RMSE = 147.6 mmol$^+$/kg), followed by KED with MIR as the external drift (RMSE = 133.50 mmol$^+$/kg). Comparing the averages of actual and predicted CEC values in Table 6, it is clear that IRF-1 and KED with elevation as the external drift, are further confirmed to be the best predictors, with the average produced by each of these methods being closest to the actual average. One interesting point is that all seven methods used, on the average, over-predicted topsoil CEC. This is evident by their average predicted CEC values at the validation sites being lower than the average of the actual values. From the ME values shown in Table 6, it also clear that IRF-1 and all the KED models also over-smoothed the prediction surface, more so than MLR and RK.

Two examples of topsoil CEC maps predicted by two methods, KED with elevation as the external drift and RK, are shown in Fig. 10(a) and (b), respectively. The two maps show remarkably similar patterns of variation of topsoil CEC across the sub-catchment. The CEC trend, increasing from east to west, is also evident. This confirms that the model that best fit this trend provides the most precise predicted topsoil CEC and, in this case, the method is KED using elevation as the external drift. Relief of the area, best described by elevation, increases from the northwest to southeast, which is well reflected in the distribution patterns of topsoil CEC shown in Fig. 10. Unreliable CEC

Table 6
Summary statistics of actual and the predicted values of CEC by different methods at the validation set extracted from the sub-catchment data

	Actual	IRF-1	KED Elevation	KED MIR	KED NDVI	KED TIR2	MLR	RK
Minimum (mmol$^+$/kg)	101.15	45.69	26.01	53.77	62.11	56.67	142.20	69.62
Maximum (mmol$^+$/kg)	488.05	493.43	486.40	654.79	516.32	532.20	443.09	460.86
Mean (mmol$^+$/kg)	344.76	325.55	329.96	318.96	311.76	318.38	320.59	323.33
Mean Error (ME) of prediction (mmol$^+$/kg)	–	19.2	14.8	25.8	33.0	26.4	24.17	21.42
Root Mean Square Error (RMSE) of prediction (mmol$^+$/kg)	–	85.9	66.2	115.3	147.6	118.0	108.07	95.81

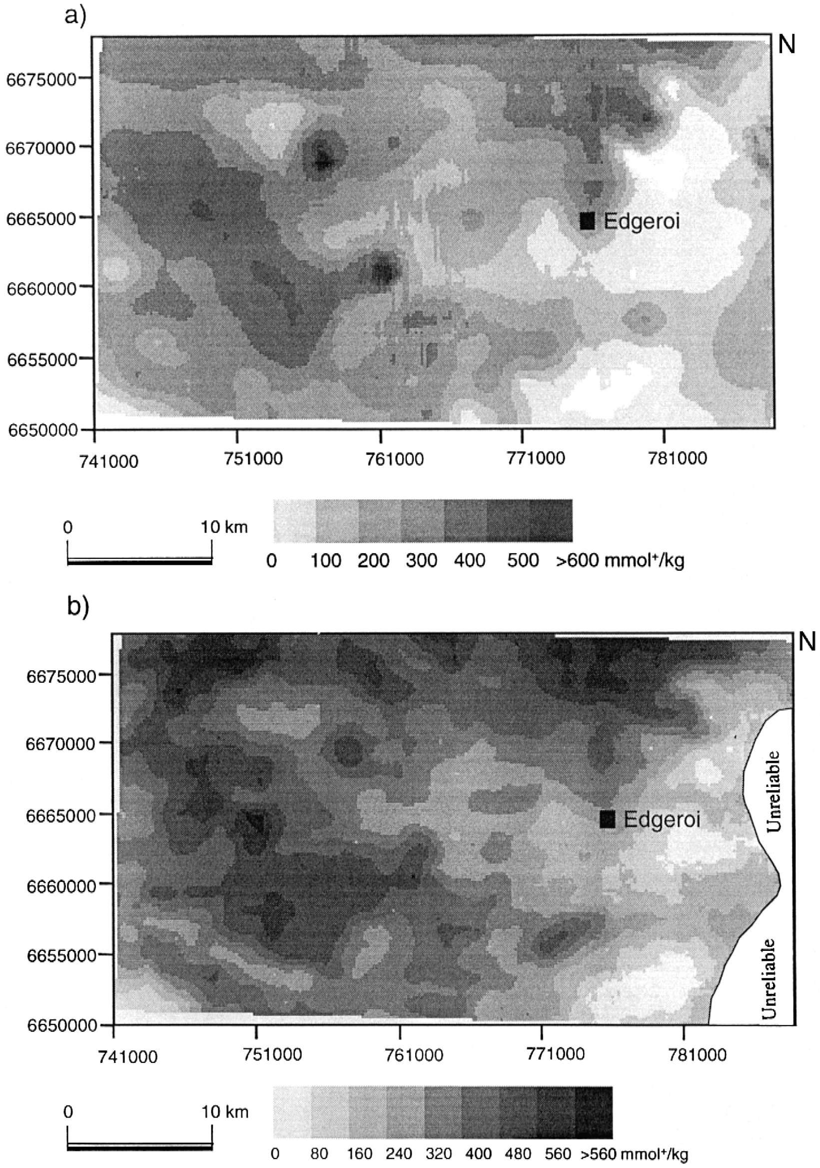

Fig. 10. Predicted CEC (mmol/kg) for the Edgeroi area using (a) Kriging with elevation as the external drift and (b) Regression kriging (RK).

values are predicted in some parts of the eastern section of the sub-catchment. This is true for all the methods, but less so for KED with elevation as the external drift (Fig. 10a). The high prediction error in this section of Edgeroi is probably due to different topology in the study area, which results in higher model uncertainty in the section than in the western part. However, the models, especially KED with elevation as the external drift (Fig. 10a), have produced reasonable predictions of CEC values for much of the Edgeroi areas suitable for further analysis for sub-catchment management.

4.3. Regional extent

The example for regional survey is based on the combined area of the Lower Macintyre, Gwydir and Namoi valleys, including data from the detailed survey of the Edgeroi area. The total area covered is approximately 45,600 km^2. This region has been sampled extensively for a major study of irrigated cotton regions of eastern Australia (Fig. 5). The objective was to obtain a soil database so that a quantitative statement on the status of soil could be made. For this purpose, a total of 734 locations were visited and sampled.

The main purpose of the project was to predict several soil attributes from samples taken from 734 sites (Fig. 5). This is required for the basic regional planning and total catchment management strategy being envisaged for the region. We provide a simple example, again with spatial prediction of the topsoil CEC. First, the AVHRR and elevation data were interpolated onto a grid matrix of 500-m spacing. This resolution is considered more than adequate for a regional survey. Again, we used the AVHRR data combined with the elevation to predict the CEC over the entire region using only the MLR and RK.

Fig. 11 shows the map of CEC predicted by the RK model. The distribution patterns highlight the influence of physiography on the soil types — and hence the CEC. Clearly the light-textured soil on the leveés are identified by relatively low CEC (CEC < 150 mmol$^+$/kg). These areas are depicted by dark grey to dark patches on the map in Fig. 11, which represent areas of relatively large CEC values, are mainly low depressions which are components of the clay plains, mostly consisting of cracking clays (Stannard and Kelly, 1977).

The predicted CEC values in some areas of the map (these are specified as ‘‘unreliable’’ in Fig. 11) are spurious to say the least. This is especially true for the eastern part of the map. The eastern section is the fringe of the Great Dividing Range, which is undulating and hilly. The landforms are quite dissimilar to the clay plains. This accounts for low precision of prediction for the section as the models are not suitable for topsoil CEC/ancillary variables interrelationships. The section was also scantly sampled, or not sampled at all. It thus shows how poor the models, such as MLR and RK, are in extrapolating to areas outside the sampled region. An area just east of Pilliga is characterised by very low topsoil CEC (Fig. 11). This area is underlain by Pilliga sandstone, which produced some light-textured soil types characterised by low CEC. Otherwise, the topsoil CEC is generally large and adequate as a soil quality indicator for soil nutrient-holding capacity for plant growth. The map generally depicts a reasonably acceptable distribution of CEC for the region, adequate for

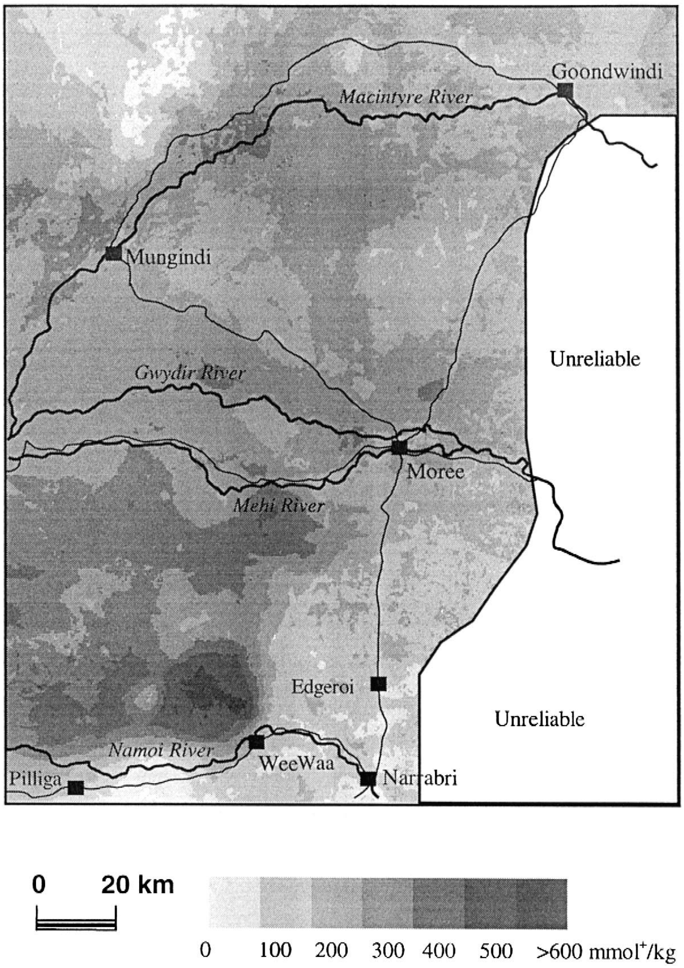

Fig. 11. Predicted CEC across the three valleys, using regression kriging (RK).

regional planning and for the elucidation of areas of low CEC, which need further investigation.

5. General conclusion

Application of each of the pedometric techniques depends on the purpose, resolution and setup of the survey as the ultimate use of soil survey information determines the accuracy required. Different techniques produce different error of interpolation. For example, the accuracy of soil information required for field-scale is clearly different from that which is required for total catchment management. Other examples of applications of pedometric methods in analysing soil survey information for various purposes at the national or continental scale exist in literature (e.g., Wendland et al., 1998; Stoorvogel and Smaling, 1998). In all the cases, since the purposes are different, the risk of taken wrong decisions due to survey error is also different. Therefore, the pedometric techniques described above cannot just be applied to any situation without consideration of the specific needs and appropriateness of the inherent assumptions of the techniques. Table 7 presents a summary of examples as a guide for selecting the best pedometric technique, given the purpose of the soil survey, the precision required, the scale of the survey and the final pixel resolutions of the resulting thematic maps. These techniques need to be incorporated into the mainstream GIS packages for appropriate (geo)statistical analysis prior to GIS operations for land/soil quality analysis, soil contamination and pollution studies, various decision maps (including precision agriculture) and total environmental management.

From all these examples and new methods being generated, it appears that:

- Pedometrics and pedology are growing closer together.
- Fuzzy sets and pedodiversty analyses offer new techniques.
- There is a range of new data sources, and proximal sensing is one of those that are being developed.
- Hybrid methods of CLORPT and geostatistics offer powerful spatial prediction methods, especially up to catchment and regional extent.

Finally, although all these techniques allow better use of existing qualitative and quantitative soil information there is a danger that new data are not generated to test models. We shall always need more new soil data, and this should be part and parcel of quantitative studies.

References

Ahmed, S., DeMarsily, G., 1987. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water Resour. Res. 23, 1717–1737.

Ameskamp, M., 1997. Three-dimensional rule-based continuous soil modelling. Uplish. PhD. Dissertation, Christian-Albrechts University, Kiel.

Bardossy, A., Duckstein, L., 1995. Fuzzy Rule-Based Modelling with Applications to Geophysical, Biological and Engineering Systems. CRC Press, New York.

Bell, J.C., Butler, C.A., Thompson, J.A., 1995. Soil-terrain modeling for site-specific agricultural management. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Site-Specific Management for Agricultural Systems. ASA-CSSA-SSSA, Madison, WI, pp. 209–227.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York.

Bouma, J., Hoosbeek, M.R., 1998. Obtaining soil and land quality indicators using research chains and geostatistical methods. Nutrient Cycling Agroecsyst. 50, 35–50.

Bouma, J., Batjes, N.H., Groot, J.J.R., 1998. Exploring land quality effects on world food supply. Geoderma 86, 24–42.
Bourgault, G., Journel, A.G., Rhoades, J.D., Corwin, D.L., Lesch, S.M., 1997. Geostatistical analysis of a soil data set. Adv. Agron. 58, 241–292.
Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties: I. The semivariogram and punctual kriging. J. Soil Sci. 31, 315–331.
Burrough, P.A., 1986. Principles of Geographical Information Systems for Land Resources Asesessment. Clarendon Press, Oxford.
Burrough, P.A., 1989. Fuzzy mathematics methods for soil survey and land evaluation. J. Soil Sci. 40, 447–492.
Burrough, P.A., MacMillian, R.A., van Deusen, W., 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. J. Soil Sci. 43, 193–210.
Chang, L., Burrough, P.A., 1987. Fuzzy reasoning: a new quantitative aid for land evaluation. Soil Surv. Land Eval. 7, 69–80.
Cook, S.E., Corner, R.J., Groves, P.R., Grealish, G.J., 1996. Use of gamma radiometric data for soil mapping. Aust. J. Soil Res. 34, 183–194.
Corwin, D.L., Rhoades, J.D., 1990. Establishing soil EC-depth relations from electromagnetic induction measurements. Commun. Soil Sci. Plant Anal. 21, 861–901.
Curran, P.J., 1998. Remote sensing technologies for global ecological applications: an extended abstract. 9th Australasian Remote Sensing and Photogrammetric Conference, July 20–24, 1998, University of New South Wales, Sydney, Australia.
De Gruijter, J.J., 1977. Numerical Classification of Soils and its Application in Soil Survey. Agricultural Research Report No. 855. Pudoc, Wageningen.
De Gruijter, J.J., Mcbratney, A.B., 1988. A modified fuzzy *k*-means method for predictive classification. In: Bock, H.H. (Ed.), Classification and Related Methods of Data Analysis. Elsevier, Amsterdam, pp. 97–104.
Fuzzy Sets in Soil Science. In: De Gruijter, J.J., McBratney, A.B., McSweeney, K. (Eds.), Geoderma vol. 77 pp. 83–339.
Dobermann, A., Oberthur, T., 1997. Fuzzy mapping of soil fertility — a case study on irrigated riceland in the Philippines. Geoderma 77, 317–339.
Doran, J.W., Parkin, T.B., 1994. Defining and assessing soil quality. In: Doran, J.W., Coleman, D.C., Bezdicek, D.F., Stewart, B.A. (Eds.), Defining Soil Quality for a Sustainable Environment. SSSA Special Publication vol. 35 pp. 3–22, USA.
ESRI, 1997. Understand GIS: The Arc/Info Method. Environmental Systems Research Institute, Redlands, CA.
FAO, 1976. A framework for land evaluation. Soils Bull. 32, FAO, Rome.
Finke, P.A., Bouma, J., Hoosbeek, M.R., 1989. Soil and Water Quality at Different Scales. Developments in Plant and Soil Sciences vol. 80 Kluwer Academic Publishing, p. 324.
Fisher, R.A., 1954. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.
Furley, P.A., 1971. Relationships between slope form and soil properties developed over chalk parent materials. In: Brusden, D. (Ed.), Institute of British Geographers Special Publication vol. 3 pp. 141–164.
Geovariances, 1997. ISATIS, Avon Cedex, France.
Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil landscape modelling and spatial prediction of soil attributes. Int. J. Geogr. Info. Syst. 9, 421–432.
Goovaerts, P., 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. J. Soil Sci. 43, 597–619.
Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford Univ. Press, New York.

Hendricks Frassen, H.J.W.M., van Eijnsbergen, A.C., Stein, A., 1997. Use of spatial prediction techniques and fuzzy classification for mapping soil pollutants. Geoderma 77, 243–262.

Hole, F.D., Hironaka, M., 1960. An experiment in ordination of some soil profiles. Soil Sci. Soc. Am. Proc. 24, 309–312.

Ibanez, J.J., De-Alba, S., Lobo, A., Zucarello, V., 1998. Pedodiversity and global soil patterns at coarse scales (with discussion). Geoderma 83, 171–192.

Isbell, R.F., 1996. The Australian Soil Classification. Australian Soil and Land Survey Handbook. CSIRO Publishing, Melbourne.

Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-Hill, New York, 281 pp.

Kandel, A., 1986. Fuzzy Mathematical Techniques with Applications. Addison-Wesley, Massachusetts.

Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. Comparison of Kriging with for Spatia Interpolation of Horizon Depth with Censored Observations.

Lesch, S.M., Strauss, D.J., Rhoades, J.D., 1995. Spatial prediction of soil salinity using electromagnetic induction techniques: 1. Statistical prediction models: a comparison of multiple linear regression and cokriging. Water Resour. Res. 31, 373–386.

Matheron, G., 1965. Les Variable Regionalisées et leur Estimation. Masson, Paris.

Matheron, G., 1969. Le krigeage universel. Cahiers du Centre de Morphologie Mathematique vol. 1 Ecole des Mines de Paris, Fontainebleau.

McBratney, A.B., 1992. On variation, uncertainty, and informatics in environmental soil management. Aust. J. Soil Res. 30, 913–935.

McBratney, A.B., Odeh, I.O.A., 1997. Applications of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77, 85–113.

McBratney, A.B., Pringle, M.J., 1997. Spatial variability in soil-implications for precision agriculture. In: Stafford, J.V. (Ed.), Precision Agriculture 1997. BIOS Scientific Publ., Oxford, UK, pp. 3–33.

McBratney, A.B., Webster, R., 1981. Spatial dependence and classification of soil along a transect in north-east Scotland. Geoderma 26, 63–82.

McBratney, A.B., Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy. J. Soil Sci. 34, 137–162.

McBratney, A.B., Hart, G.A., McGarry, D., 1991. The use of region partitioning to improve the representation of geostatistically mapped soil attributes. J. Soil Sci. 42, 513–532.

McBratney, A.B., de Gruijter, J.J., Brus, D.J., 1992. Spacial prediction and mapping of continuous soil classes. Geoderma 54, 39–64.

McGarry, D., Ward, W.T., McBratney, A.B., 1989. Soil Studies in the Lower Namoi: Methods and Data. 1. The Edgeroi Data Set. CSIRO, Australia.

Minasny, B., McBratney, A.B., Whelan, B.M., 1999. VESPER Version 1.0. Australian Centre for Precision Agriculture, McMillan Building A05. The University of Sydney, NSW 2006. http://www.usyd.edu.au/su/agric/acpa.

Moore, A.W., Russell, J.S., 1967. Comparison of coefficients and grouping procedures in numerical analysis of soil trace element data. Geoderma 1, 139–158.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57, 443–452.

Myers, D.E., 1982. Matrix formulation of cokriging. Math. Geol. 14, 249–257.

Odeh, I.O.A., McBratney, A.B., 2000. Using AVHRR imageries for spatial prediction of clay content in the lower Namoi Valley of Eastern Australia. Geoderma, this issue.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Fuzzy-c-means and kriging for mapping soil as a continuous system. Soil Sci. Soc. Am. J. 56, 1848–1854.

Odeh, I.O.A., Gessler, P.E., McKenzie, N.J., McBratney, A.B., 1994. Using attributes derived

from digital elevation models for spatial prediction of soil properties. Resource Technology '94, Melbourne, Australia, pp. 451–463.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–225.
Odeh, I.O.A., McBratney, A.B., Slater, B.K., 1996. Predicting soil properties from ancillary information: non-spatial methods compared with geostatistical and combined methods. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong '96 vol. 2 Kluwer Academic Publishing, Netherlands, pp. 1008–1019.
Pham, T.D., 1997. Grade estimation using fuzzy-set algorithms. Math. Geol. 29, 291–305.
Renger, M., 1971. Die Ermittlung der Porengrößenverteilung aus der Körnung, dem Gehalt an organischer Substanz und der Langerungsdichte. Z. Pflanzenernähr. Bodenkd 130, 53–67.
Rhoades, J.D., 1992. Instrumental field methods of salinity appraisal. In: Topp, G.C., Reynolds, W.D., Green, R.E. (Eds.), Advances in the Measurement of Soil Physical Properties: Bringing Theory into Practise. Soil Science Society of America, Madison, WI, USA, pp. 299–336.
Rhoades, J.D., Corwin, D.L., 1981. Determining soil electrical conductivity-depth relations using an inductive electromagnetic soil conductivity meter. Soil Sci. Soc. Am. J. 45, 255–260.
Ruhe, R.V., Walker, P.H., 1968. Hillslope models and soil formation: I. Open systems. Transactions of the 9th Congress vol. 4 International Society Soil Science, Adelaide, pp. 551–560.
Samara, J.S., Singh, V.P., 1990. Spatial dependence of soil reclamation. Soil Technol. 3, 153–165.
Shatar, T.M., McBratney, A.B., 1999. Empirical modeling of relationships between sorghum yield and soil properties. Precis. Agric. 1, 249–276.
Stannard, M.E., Kelly, I.D., 1977. The Irrigation Potential of the Lower Namoi Valley. Water Resources Commision, New South Wales.
Statistical Sciences, 1995. S-Plus Guide to Statistical and Mathematical Analysis. Statsci, Seattle.
Stein, A., van Dooremolen, W., Bouma, J., Bregt, A.K., 1988. Co-kriging point data on moisture deficit. Soil Sci. Soc. Am. J. 52, 1418–1423.
Stein, A., Bouma, J., Mulders, M.A., Weterings, M.H.W., 1989. Using cokriging in variability studies to predict physical land qualities of a level terrace. Soil Technol. 2, 385–402.
Stoorvogel, J.J., Smaling, E.M.A., 1998. Research on soil fertility decline in tropical environments: integration of spatial scales. Nutr. Cycling Agroecosyst. 50, 151–158.
Sudduth, K.A., Drummond, S.T., Birrell, S.J., Kitchen, N.R., 1997. Spatial modeling of crop yield using soil and topographic data. In: Stafford, J.V. (Ed.), Precision Agriculture '97: Spatial Variability in Soil and Crop vol. 1 BIOS Scientific Publishers, Oxford, pp. 439–447.
Tomer, M.D., Anderson, J.L., Lamb, J.A., 1995. Landscape analysis of soil and crop data using regression. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Site-Specific Management for Agricultural Systems. ASA-CSSA-SSSA, Madison, WI, pp. 273–284.
Triantafilis, J., McBratney, A.B., 1998. Development of a mobile electromagnetic sensing system for soil salinity assessment in irrigated cotton fields, Proceedings of the the 9th Australian Cotton Conference, August 12–14 1998, Broadbeach, Gold Coast, Queensland, Australia ACGRA.
Venables, W.N., Ripley, B.D., 1994. Modern Applied Statistics with S-Plus. Verlag, New York.
Vereecken, H., Maes, J., Feyen, J., Daruis, P., 1989. Estimating the soil moisture retention characteristic from texture, bul density, and carbon content. Soil Sci. 148, 389–403.
Wackernagel, H., 1988. Geostatistical techniques for interpreting multivariate spatial information. In: Chung, C.F. (Ed.), Quantitative Analysis of Mineral and Energy Resources. Reidel, Dordrecht, pp. 393–409.
Wackernagel, H. et al., 1994. Multivariate spatial statistics. Geoderma 62, 83–92.
Wackernagel, H., 1995. Multivariate Geostatistics. Springer, Berlin, p. 256.

Ward, W.T., personal communication. Soils and landscapes near Narrabri and Edgeroi, New South Wales, with data analysis using fuzzy *k*-means. CSIRO Division of Soils Divisional Report.

Webster, R., 1994. The development of pedometrics. Geoderma 62, 1–15.

Webster, R., Burrough, P.A., 1974. Multiple discriminant analysis in soil survey. J. Soil Sci. 25, 120–134.

Wendland, F., Bach, M., 1998. The influence of nitrate reduction strategies on temporal development of the nitrate pollution of soil and groundwater throughout Germany: a regional differentiation case study. Nutr. Cycling Agroecosyst. 50, 167–180.

Whelan, B.M., McBratney, A.B., 1999. Prediction uncertainty and implications for digital map resolution. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Precision Agriculture, Proceedings of the 4th International Conference on Precision Agriculture. ASA/CSSA/SSSA, Madison, WI, pp. 1185–1196.

Wilding, L.P., 1985. Spatial variability: its documentation, accommodation and implication to soil surveys. In: Nielsen, D.R., Bouma, J. (Eds.), Soil Spatial Variability. Proceedings of a Workshop of ISSS and the SSSA, Les Vegas USA. Nov. 30–Dec. 1, 1984.

Zeverbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Processes Landforms 12, 47–56.