ELSEVIER

Geoderma 100 (2001) 269–301

GEODERMA

www.elsevier.nl/locate/geoderma

Modelling soil variation: past, present, and future

G.B.M. Heuvelink [a,*], R. Webster [b]

[a] *Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands*

[b] *Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, UK*

Received 12 December 2000; received in revised form 16 February 2001; accepted 20 February 2001

Abstract

The soil mantles the land, except where there is bare rock or ice, and it varies more or less continuously. Many of its properties change continuously in time, too. We can measure the soil at only a finite number of places and times on small supports, and any statement concerning the soil at other places or times involves prediction. Variation in soil is also complex, so complex that no description of it can be complete, and so prediction is inevitably uncertain. Soil scientists should be able to quantify this uncertainty, and manage it. This means representing the variation by models that may be in part deterministic, but cannot be wholly so; they must have some random element to represent the unpredictable variation. Here we review three families of statistically based models of soil variation that are currently in use and trace their development since the mid-1960s. In particular, we consider classification and geostatistics for modelling the spatial variation, time series analysis and physically based approaches for modelling temporal variation, and space–time Kalman filtering for predicting soil conditions in space and time simultaneously. Each of these attaches to its predictions quantitative estimates of the prediction errors. Past, present and future research has been, is, and will be directed to the development of models that diminish these errors. A challenge for the future is to investigate approaches that merge process knowledge with measurements. For soil survey, this would be achieved by integration of pedogenetic knowledge and field observations through the use of data assimilation techniques, such as the space–time Kalman filter.

Keywords: Soil; Spatial variation; Temporal variation; Modelling; Soil classification; Geostatistics; Time series; Kalman filter

* Corresponding author.

E-mail addresses: g.b.m.heuvelink@frw.uva.nl (G.B.M. Heuvelink), richard.webster@bbsrc.ac.uk (R. Webster).

0016-7061/01/$ - see front matter
PII: S0016-7061(01)00025-8

1. Introduction

The soil varies from place to place, and many of its properties vary in time too. This is what makes the soil so fascinating. We place the variety we observe on record, and we seek explanation for it. Were the soil uniform, we should simply acknowledge the fact and switch our attention to something more interesting. On the practical level that same variety can be both a nuisance, in that we have to modify our land management to cope with it, and an asset in providing diverse habitats for a rich flora, both wild and cultivated, and resilience against stress and fluctuations in weather and climate.

Because the soil varies, sampling the soil at a finite number of places or points in time yields incomplete pictures. Often we need to predict between sampling points. For instance, we may need to interpolate in space to construct a map of a soil property. We may need to predict ahead in time to compare the effects of different scenarios or management practices. Prediction requires us to create models of the real world and apply them. In some cases, such models may be implicit only, as when a soil surveyor maps the soil with a conceptual model of soil variation in mind (Schelling, 1970; Dijkerman, 1974). Other soil scientists use models of soil variation more explicitly. For instance, computer-based simulation models are used to predict the movement of water and the leaching of solutes in the soil (Van Genuchten and Dalton, 1986; Addiscott,

Table 1
General properties of the statistical models of soil variation reviewed

Type of model	Temporal variation	Spatial variation	Primary source of information	Significant introduction to soil science
Soil classification	ignored	discrete representation	pedology	1960s
Soil geostatistics	ignored	continuous representation	observations	1980s
Classification merged with geostatistics	ignored	discrete and continuous representation	pedology and observations	1990s
Time series analysis	included	ignored	observations	1990s
State–space approach	included	ignored	physical laws and observationss	1990s
Space–time geostatistics	included	included	observations	1995 +
Spatial state–space approach	included	included	physical laws and observations	2000 +

1993). A geostatistician models soil properties as if they were realizations of random fields (Webster, 2000).

Regardless of how he or she predicts, a soil scientist assumes some kind of model of how the soil behaves in space and time. There may be huge differences between these models, but what they all have in common is that they involve some degree of uncertainty. Soil varies at all scales with great complexity, and there is no way that we can capture the full extent of its variation in a deterministic model. Statistical models recognize the associated uncertainty, and these models are the focus of this paper.

In this review we give a historic account of models of soil variation since the mid-1960s when computers sparked their quantitative development and assessment. We mention their successes in the past, we identify their merits and weaknesses in the present, and we speculate on future developments. We first discuss models of soil spatial variation, next address models of soil temporal variation, and last analyse models that treat soil spatio-temporal variation in an integrated fashion. Properties of the various models discussed are summarized in Table 1.

2. Soil spatial variation

Two principal approaches to representing spatial variation in soil can be distinguished. One, the earlier, has its roots in intuition overlaid with 19th century biological taxonomy and practice in geological survey, to which it was related. It partitions the soil into more or less discrete classes. The other sees soil as a suite of continuous variables and seeks to describe the way they vary over the land. It is necessarily quantitative and it embodies fairly advanced statistical theory. It also requires powerful computers, and so it had to await their coming.

It turns out that both approaches can be cast into a general model of spatial variation, each incorporating their specific assumptions. Recently, we have seen several attempts, with some success, to merge the two approaches into models that more realistically represent the real world.

Incidentally, we use the word ‘variation’ to describe the actuality rather than ‘variability’, which we reserve to mean the potential to vary, as in random processes. This accords with the distinction between ‘variate’ to signify a set of values of a ‘variable’, which in soil science is a property of the soil—see McBratney (1992), Burrough (1993a) or Webster (2001) for a fuller explanation and rationale.

2.1. Soil classification—the Dark Ages

Soil classification involves dividing the population of all soil into more or less discrete classes. In the spatial context, the classes are subdivisions of finite

circumscribed regions. The boundaries are sharp lines across which the observed differences are greater than elsewhere and within which the soil is relatively homogeneous. For any one region, the outcome is a map showing the region tessellated into spatial classes, which constitute a general-purpose classification. This usually is accompanied by a text in which the classes are described. There are thousands of examples.

Traditionally, such a map was made by the soil surveyor using his or her intuition and formal knowledge and understanding to collate all the information available on the soil and its relations with geology, geomorphology, vegetation, and land use. It involved observation in the field and latterly air photo interpretation. It required few expensive observations and laboratory analyses. The soil of the classes created was described from typical sites, which were chosen by the soil surveyors using judgement based again largely on intuition. Variation within the classes was acknowledged, but was described qualitatively and then usually in only vague terms. There was no formality, no quantitative expression of variation.

2.1.1. Statistical aspects of spatial classification and predicting from soil maps

By the 1960s, taxonomists were putting numerical limits on the discriminating criteria for consistency. This helped to codify description, but it did not obviously aid prediction with the uncertainties it entailed. Soil classification must be treated statistically to take into account the variation remaining within the classes and the consequent uncertainty.

Civil engineers provided the stimulus to put soil classification on a formal statistical basis. They wanted to build rural roads and airfields, often in remote areas, with the natural soil as the foundation. They could sample the soil, but only sparsely. Elsewhere they would have to predict. Morse and Thornburn (1961) saw a potential solution to the problem of prediction in soil maps made for agricultural purposes. They reasoned that if the variation within the classes was less than that in a region at large then using the class means of the properties of interest as predictors should be more precise than the regional means. Further, if they sampled with a sufficient element of randomness, and provided the data seemed to follow some theoretical distribution, ideally normal, then they could put their predictions on a probabilistic footing. And so they investigated the possibility. They took a soil map of a part of the Mid-West of the USA made by the US Department of Agriculture, sampled it with a stratified random design, treating the mapped classes as strata, and measured the amounts of sand, silt and clay, and Atterberg limits at several depths. They estimated the means and variances of the properties from the resulting data. The predictors would be the estimated stratum means, and the errors of prediction could be determined simply from the within-stratum variances. Kantey and Williams (1962) and Webster and Beckett (1970) took the same line of argument, but with greater purpose in that they made their own maps and tested them. All achieved

some success in that the variance within the classes was less than in the landscape as a whole, and so one could predict more precisely with the map than without it. This idea has been pursued many times since and is still used and investigated to this day (e.g. Oberthür et al., 1996; Brannon and Hajek, 2000).

The investigators' analysis was rooted in classical statistics, or design-based estimation, which Brus and De Gruijter (1997) have done much to promote recently. In this approach a soil property z takes values at an infinity of points $\boldsymbol{x} = \{x_1, x_2\}$ within a domain or region of interest $\mathscr{D}$. These values comprise the population. The region is divided into K spatial strata, $\mathscr{D}_k$, $k = 1, 2, \ldots, K$, that are mutually exclusive and exhaustive. In what follows, the terms 'class' and 'stratum' are used as synonyms. If the region is then sampled at random, then the value at any point in stratum $\mathscr{D}_k$ is predicted by the mean of the observations in $\mathscr{D}_k$. Class-based prediction can also be cast in a model-based setting (Heuvelink, 1996), and we do that here. We now treat the soil property as a random variable, and denote it by $Z(\boldsymbol{x})$. We use a capital Z to signify that the soil property is random. It satisfies the following model:

$$Z(\boldsymbol{x}) = \mu + \sum_{k=1}^{K} \delta_k(\boldsymbol{x})\,\alpha_k + \varepsilon(\boldsymbol{x}), \tag{1}$$

where μ is the global mean, α_k is the mean deviation of $Z(\boldsymbol{x})$ within stratum $\mathscr{D}_k$ from the global mean, and $\delta_k(\boldsymbol{x})$ is a binary function taking the value 1 for $\boldsymbol{x} \in \mathscr{D}_k$ and 0 elsewhere. The residual $\varepsilon(\boldsymbol{x})$ represents the within-class variation, which has zero mean and variance C_0, and which in addition is spatially uncorrelated (white noise). We term Eq. (1) the *discrete* model of spatial variation.

If the model is to have any merit, then the variance within the classes must be less than the total variance. Thus, it is desirable for C_0 to be a small proportion of the total, so that there are large jumps in $Z(\boldsymbol{x})$ at the class boundaries (Voltz and Webster, 1990). Many individual properties of the soil depend on soil type and so application of the discrete model of spatial variation, whereby the conventional soil map is used to partition the region, often proves to be successful (Beckett and Burrough, 1971; Van Kuilenburg et al., 1982; Wösten et al., 1985; Yost et al., 1993; Kern, 1994). This is illustrated in Fig. 1.

Suppose that we have observations $z(\boldsymbol{x}_1)$, $z(\boldsymbol{x}_2), \ldots, z(\boldsymbol{x}_n)$ of $Z(\boldsymbol{x})$ that satisfy Eq. (1) with the global mean μ and α_k unknown. Assume that of these observations the first n_k are in class $\mathscr{D}_k$ and the remainder are not. It then makes sense to use the mean of the n_k observations to predict $Z(\boldsymbol{x})$ at any $\boldsymbol{x}_0$ within $\mathscr{D}_k$:

$$\hat{Z}(\boldsymbol{x}_0) = \frac{1}{n_k} \sum_{i=1}^{n_k} Z(\boldsymbol{x}_i). \tag{2}$$

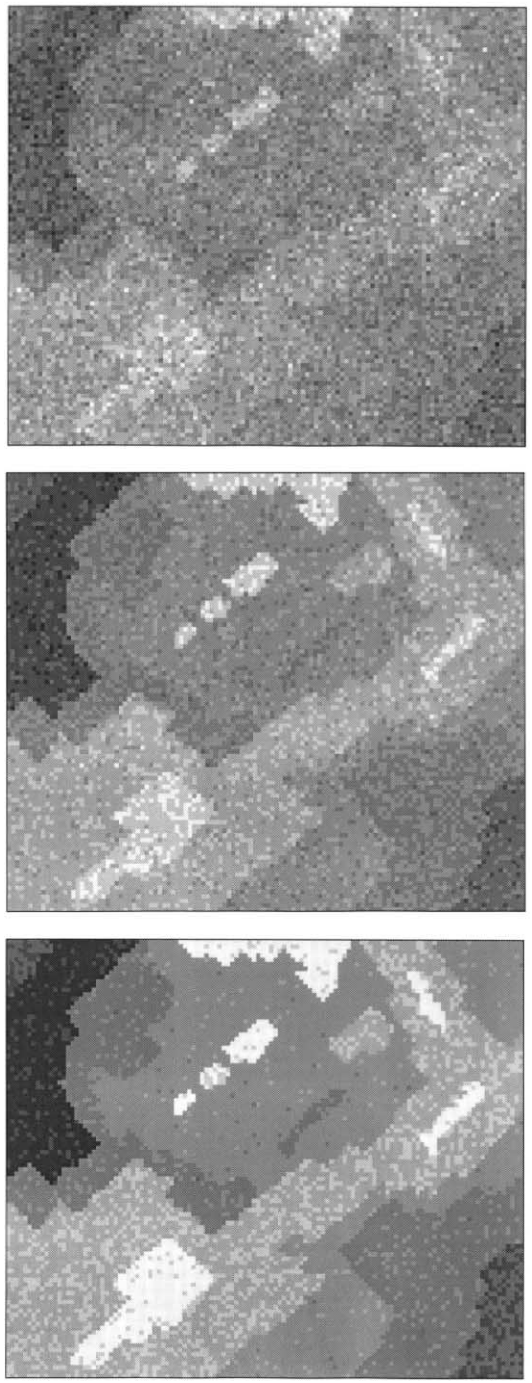

Fig. 1. Three realizations of a soil property satisfying Eq. (1). Left: small; middle: moderate; and right: large within-class spatial variation.

If we lack other information, then Eq. (2) is our best linear unbiased predictor for any unsampled place. The corresponding prediction error variance is given by

$$\sigma^2(\boldsymbol{x}_0) = \mathrm{Var}\left[Z(\boldsymbol{x}_0) - \hat{Z}(\boldsymbol{x}_0) \right] = C_0 + \frac{1}{n_k} C_0. \quad (3)$$

In words, the prediction error variance is composed of the residual variance in the population plus the estimation variance of the class mean. So, the uncertainty in our prediction arises from the inherent variation in the soil swollen by our uncertainty in our estimate of the mean. Evident from Eq. (3) is that we can diminish the uncertainty in the mean as much as we like—until it is negligible if we have sufficient resources—by increasing our sampling. In the limit, however, we are left with C_0, which no amount of sampling at this stage can affect. Prediction in the situation of the map to the right in Fig. 1 will yield large prediction error variances, no matter how large a sample we take. If we are to improve prediction any further, then we must change the classification so that each class encompasses less variation. Clearly, in the example situation of the map to the right in Fig. 1, this will not be possible, but in the real world one may hope for better classifications with smaller within-class variances.

Seen from a statistical point of view, this was the rationale for the immense effort devoted to soil classification and its refinement in the 1960s. The taxonomists attempted to improve the effectiveness of classifications by dividing soil into ever finer classes. Surveyors tried to apply in the field the refined classifications of the taxonomists, only to discover that the more narrowly were the classes defined the harder it was to identify them and the less 'pure' were their soil maps. They were forced to compromise. Further, in using data from records of classes defined on soil profile alone they added to the uncertainty of their spatial predictions the uncertainty arising from the 'impurity' of their mapped classes. Butler (1980), however, took a more robust attitude. As he saw it, they were going about their business in the wrong order. He made abundantly clear that successful classification of soil at the local level must represent what can be mapped at the chosen scale and is determined by the local landscape. Surveyors should therefore map the boundaries first, then see what they have captured and describe it. Here was no question of impurity; the map was definitive, and it remained only to sample its classes to provide statistical descriptions.

The matter was investigated at length and in detail by Beckett and co-workers Webster and Beckett (1968) had laid the foundations for evaluating soil maps by proposing the intra-class correlation as a basis for comparison. If the intra-class correlation were 1, then the classification would be perfect, there would be no variation left within classes, (i.e. $C_0 = 0$). At the other end of the scale, a

classification for which the intra-class correlation equals 0 would represent no advantage on no classification.

Beckett and Burrough (1971), in one of the most significant applications of the ideas, compared the effect of classification at several scales and intensities of observation in three distinct regions of England. Although they did observe improvements in some cases, overall, increasing the intensity of observation to refine the maps produced small benefits. The classification into soil series devised for publishing maps at 1:63 360 and delineated at that scale was almost as good as the more detailed delineations at the finer scales.

With hindsight, we can now guess the main reason: the within-class residual had a much shorter correlation range than the intervals between observations in these surveys. However, it was probably not the only reason. The authors did not take advantage of the larger scale to create finer subdivisions that might be mapped at the finer scales. Had they done so they might have found less variation within the classes.

Several other soil scientists pursued this line of investigation. Recently, Leenhardt et al. (1994) gave a full account of the logic and theoretical evaluation of the approach. In their paper, the authors focus in particular on optimizing the sample size and on how to handle preferential (representative) sampling statistically.

We might summarize conventional soil mapping, choropleth mapping to give it its technical name, as follows. It is quick and cheap, especially if there are close relations with physiography and land cover and use can be made of aerial photography, and it is likely to be moderately effective. Improving the efficacy, however, requires much more effort to recognize more subtle distinctions and delineate more intricate boundaries, and the rewards are likely to be small. Other shortcomings include weak correlation between some soil properties, especially chemical constituents and fractions, and the mapped classes, the representation of gradual change by abrupt boundaries, and the treatment of within-class variation as spatially uncorrelated (Webster and Cuanalo, 1975; Nortcliff, 1978; Campbell et al., 1989; Nettleton et al., 1991). Dealing with these requires a different view of the soil, one that recognizes the continuity rather than the discontinuity.

2.2. Age of enlightenment—soil geostatistics

Geostatistics was introduced into soil science some 20 years ago as an alternative to classification for representing spatial variation. Burgess and Webster (1980) were the first to use kriging, its practical application, in soil survey, and many other pedologists and environmental scientists have followed in their wake. There are now several excellent accounts of geostatistics and its underlying theory written for soil scientists (e.g. Goovaerts, 1997, 1999; Webster and Oliver, 2001), and so we repeat only the basic principles here.

Fundamental to geostatistics is the concept of a variable $Z(\boldsymbol{x})$ that is both random and spatially autocorrelated. In its simplest form, the model is

$$Z(\boldsymbol{x}) = \mu + \varepsilon(\boldsymbol{x}), \tag{4}$$

where μ is the mean, and $\varepsilon(\boldsymbol{x})$ is a random residual. Whereas in Eq. (1) we assumed the residuals to be uncorrelated, we now recognize explicitly spatial correlation in them, and we express it quantitatively by the semivariance:

$$\gamma(\boldsymbol{h}) = \frac{1}{2}\mathrm{E}\left[\{Z(\boldsymbol{x}) - Z(\boldsymbol{x}+\boldsymbol{h})\}^2\right] = \frac{1}{2}\mathrm{E}\left[\{\varepsilon(\boldsymbol{x}) - \varepsilon(\boldsymbol{x}+\boldsymbol{h})\}^2\right], \tag{5}$$

where E signifies expectation, and $\boldsymbol{h}$ is a vector separating two points $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$.

Eq. (4) is termed the *continuous* model of spatial variation. In contrast to the discrete model, there is no subdivision of $\mathscr{D}$.

In Eq. (5), one implicitly assumes that the semivariance depends only on the separation $\boldsymbol{h}$ and not on the positions $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$. This assumption is often made to facilitate its identification from observations of $Z(\boldsymbol{x})$. The function relating semivariance to the lag, $\boldsymbol{h}$, is the semivariogram, nowadays usually shortened to just the variogram. Spatial dependence manifests itself in the variogram typically by a monotonic increase from the origin with increasing lag distance. In other words, near points are more similar to one another on average than ones further apart. The variogram may reach an upper bound (its 'sill') at a finite distance (the 'range'), beyond which there is no longer spatial autocorrelation. Alternatively, $\gamma(\boldsymbol{h})$ may approach its maximum asymptotically. Both kinds of bounded variogram are characteristic of second-order stationary processes and have their equivalent autocovariance functions:

$$C(\boldsymbol{h}) = \mathrm{E}\left[(Z(\boldsymbol{x}) - \mu)(Z(\boldsymbol{x}+\boldsymbol{h}) - \mu)\right] = C(\mathbf{0}) - \gamma(\boldsymbol{h}), \tag{6}$$

where $C(\boldsymbol{h})$ denotes that autocovariance at lag $\boldsymbol{h}$, and $C(\mathbf{0})$, the autocovariance at lag zero, is the variance of the process.

Although the variogram of a continuous process must pass through the origin, we often find in practice that it appears to have a positive intercept on the ordinate. This is known as the 'nugget variance', a term from mining and signifying an uncorrelated component of variation such as that caused by rare gold nuggets in an ore. Random measurement errors also contribute to the nugget variance.

The variogram as expressed in Eq. (5) is that of the random process. We need it to predict $Z(\boldsymbol{x})$ between our sampling points, see below. We must therefore estimate it, and we usually do so by fitting a function to the experimental variogram, computed from the observations. Modern practice is to fit the variogram parameters by minimizing the sums of squares, suitably weighted, between the experimental values and those of the model, as recommended by Cressie (1993) and Lark (2000a). Other variogram estimation methods have been tried as well, such as maximum likelihood estimation (Lark, 2000b).

A soil scientist soon learns to recognize from the variogram the spatial characteristics of a soil property. Fig. 2 shows this with realizations of a variable satisfying Eq. (4) with three contrasting variograms. He or she should be wary, though. The variogram characterizes only the second-order moment of $Z(\boldsymbol{x})$; it does not capture higher-order moments. This is fine for normally distributed (Gaussian) processes, such as those shown in Fig. 2, where the higher-order moments are determined by the lower-order moments. But in the general situation, realizations of random processes can look very different from one another even when their variograms are the same.

Let us turn to the prediction for which we have computed the variogram. Given point observations $z(\boldsymbol{x}_i)$, $i = 1,2,\ldots,n$, we can predict $Z(\boldsymbol{x}_0)$ at an unobserved place $\boldsymbol{x}_0$ as the linear weighted sum:

$$\hat{Z}(\boldsymbol{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i), \tag{7}$$

in which the λ_i are weights. This is much the same as Eq. (2); the difference is that whereas the weights in Eq. (2) were either $1/n_k$ or 0, now they may take values from a continuous range. As before, the weights are chosen such that they minimize the prediction error variance, provided they sum to 1 and thereby ensure that prediction is unbiased. The differences between the models (1) and (4) cause differences in the weights obtained.

The prediction error variance is

$$\sigma^2(\boldsymbol{x}_0) = 2\sum_{i=1}^{n} \lambda_i \gamma(\boldsymbol{x}_i - \boldsymbol{x}_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\boldsymbol{x}_i - \boldsymbol{x}_j). \tag{8}$$

It represents the uncertainty in the prediction. Where there is spatial dependence, the semivariance tends to increase with increasing distance from observations. So, in general, the error decreases with the density of data and not just with the total number of them, as under the discrete model of spatial variation.

Kriging minimizes the prediction error variance *if the model is correct*. Similarly, the prediction is unbiased in terms of the model. If the model is wrong, which it always is to some extent, then the prediction is not optimal. However, the predictions themselves are little affected by the choice of model, provided it is reasonable, and this is one of the strengths of kriging—it is robust in this sense. The variances can be seriously affected, however.

The introduction of kriging to soil science resulted in a major shift in attention. Since 1980, kriging has found many and diverse applications, as in pollution, fertility, trace element deficiencies and salinity. In many instances, the need has been to map these. Also, geostatistics has not stood still during that time. There have been numerous developments and refinements, which have been essentially extensions to the basic principle. Some deal with non-normality

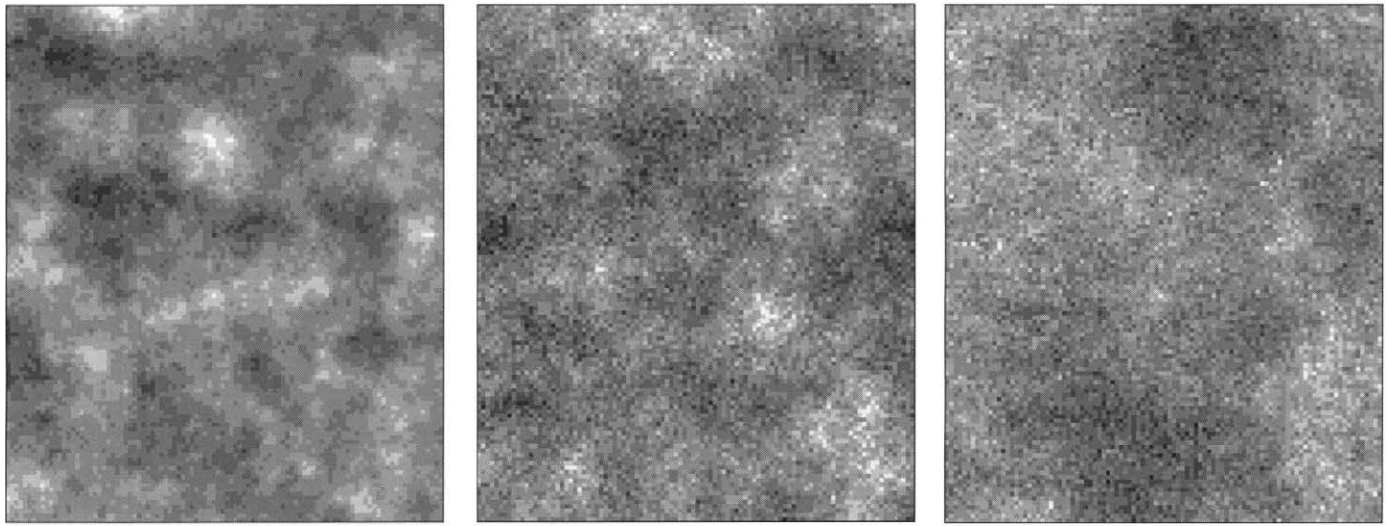

Fig. 2. Three realizations on a 100×120 rectangle of a stationary Gaussian random process satisfying Eq. (4) with a spherical variogram, $\gamma(h) = c_0 + c \times \mathrm{sph}(a)$. Left: zero nugget ($c_0 = 0$) and small range ($a = 20$). Middle: 30% nugget variance, i.e. $c_0 = 0.3(c_0 + c)$, and small range ($a = 20$). Right: 30% nugget variance and moderate range ($a = 60$).

(lognormal kriging, disjunctive kriging, indicator kriging), others address non-stationarity, i.e. varying trend or drift (universal kriging, kriging with external drift, irf$k$ kriging, and stratified kriging).

One elaboration worth specific mention is block kriging; this allows predictions for areas larger than the supports on which the observations were made. The 'support' in this context is the size of the body of material, its length, breadth and thickness, and its orientation. Block kriging is useful where measurements are made on small cores of soil yet predictions are required for large swaths of land, as when a farmer applies fertilizer in bands typically 24 m wide. It requires only little modification to the kriging equations, see for example Webster and Oliver (2001). Due to an averaging-out effect, block kriging variances are typically much smaller than the corresponding point kriging variances. However, this in itself is not a valid reason for preferring block kriging over point kriging. The choice between point and block kriging should be based solely on whether one is interested in values on the same supports as the data or in block averages. These variables are not the same, and neither are their variograms. Indeed, the values that a soil property takes and its spatial pattern can be strongly influenced by the support. This is illustrated in Fig. 3, which shows that increasing the support size decreases the amount of spatial variation in the data. This is due to a spatial averaging. Fig. 3 also shows that different support sizes may yield different spatial patterns. The pattern on the point support (left map in Fig. 3) is dominated by the short-range nugget effect, whereas the spatial pattern of the 1×1 block map (middle map) reflects a combined effect of the spatially correlated part of the stochastic residual and the long-range trend. The spatial pattern of the 10×10 block support map (right map) is influenced almost entirely by the long-range trend.

Much more has been written about the support and how it affects the accuracy of predictions in soil science (Heuvelink and Pebesma, 1999; Bierkens et al., 2000; Lark, 2000c), but we cannot pursue the subject further here.

One development that has become popular in the last decade in soil science is stochastic simulation of spatially distributed soil properties. Its aim is to compute realizations of the random process, given the variogram and usually given the data too. Such simulated realizations preserve the correlation structure, whereas kriging smoothes, i.e. loses variance. A map of a realization looks like reality. The process is swift on a modern computer, and so we can see many such possible maps in a little time. Conditional simulation, whereby we incorporate the data, enables us to get a feel for the uncertainty in our picture of reality and allows us to carry out an uncertainty analysis; perhaps this is its main purpose. There is a price, namely a doubled prediction error variance, but then optimal prediction is not the aim of the simulation. Incidentally, the means of repeated conditional simulations converges to the kriging predictions as the number of simulations increases, as does the variance in the simulated realizations converge to the kriging variance.

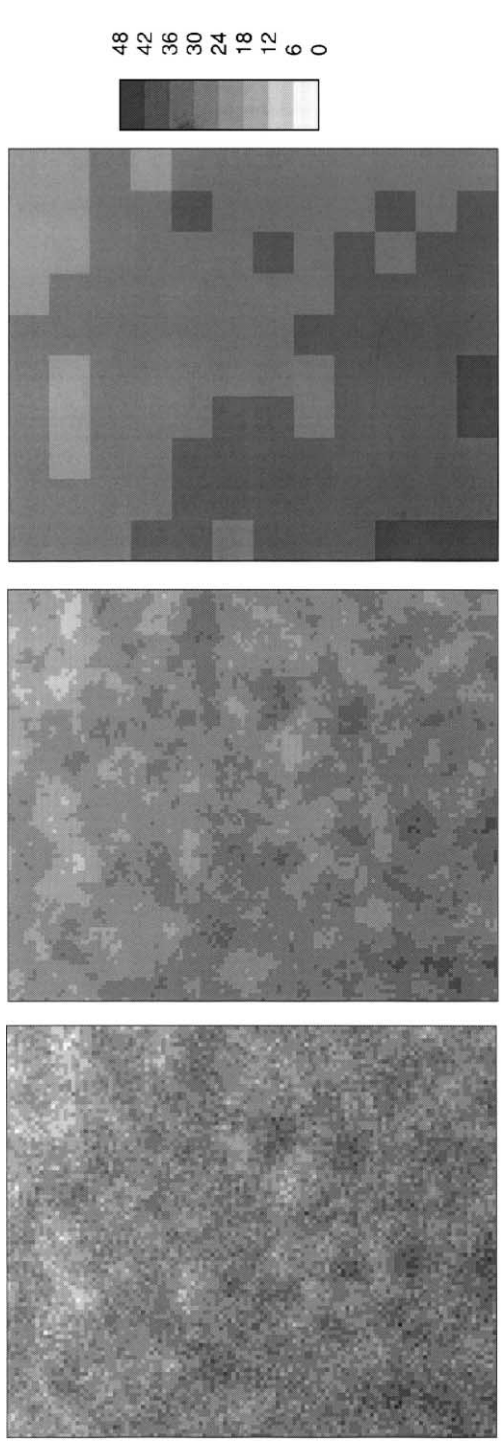

Fig. 3. A realization on a 100×120 rectangle of a random function satisfying Eq. (4) with $\mu = 32 - 0.05x - 0.10y$ and $\gamma(h) = 10 + 10 \times \text{sph}(10)$, viewed at three different supports. Left: point support. Middle: block support with blocks the size of the pixels (i.e. 1×1 units). Right: block support with blocks the size of 100 pixels (i.e. 10×10 units).

2.3. Merging classification with geostatistics

Both classification and geostatistics have their disadvantages. Neither can handle situations in which both abrupt and gradual spatial variation are present in the same region. Once soil scientists had overcome their kriging 'euphoria' (in the early 1990s), they started to realize that abandoning conventional classification outright was not so sensible after all. Ordinary kriging also had disadvantages, such as its inability to deal with sharp boundaries, though no one so far has tried to adapt Marechal's (1984) technique for handling geological faults, as far as we know.

The renewed recognition for classification led to studies in which the performance of the discrete and continuous model were compared. These studies date mainly from the 1990s (Yost et al., 1993; Brus et al., 1996; Utset et al., 2000), although the first of its kind was done earlier (Van Kuilenburg et al., 1982).

Comparison was one possibility, but merging the discrete and continuous models of spatial variation was a more challenging task. To bridge the gap between the two, several models have been proposed. Rogowski and Wolf (1994) proposed unweighed averaging of the predictions from the discrete and the continuous model. Heuvelink and Bierkens (1992) weighed these predictions using weighting factors derived from the prediction error variances. These are essentially heuristic approaches, though. Another method was to stratify the area and adopt the continuous model separately in each stratum (Stein et al., 1988; Voltz and Webster, 1990). The disadvantage of this is that it excludes any spatial autocorrelation across the boundaries. One solution to this problem is to replace crisp boundaries by gradual transition zones (Boucneau et al., 1998). Heuvelink (1996) suggested the mixed model of spatial variation, in which the soil property is treated as the sum of a global mean, a class-dependent deviation from the mean, and a spatially correlated residual. Prediction with this model boils down to kriging with an external drift (Delhomme, 1978), which in this case is a classification. Its main advantage is that it performs well over the whole range of spatial variation, from exclusively discrete realities, such as in Fig. 1, to exclusively continuous realities, such as in Fig. 2 (Heuvelink and Huisman, 2000). Goovaerts and Journel (1995) used an indicator approach to integrate soil map information into modelling the spatial variation of continuous soil properties.

One other approach that aims to escape from the abrupt boundary assumption of conventional soil classification is based on fuzzy set theory. This approach led to a method known as 'continuous classification' (Burrough, 1989; De Gruijter et al., 1997). The key property of continuous classification is that objects need not belong to one class only. Instead, an object can be a member of multiple classes, with class-specific membership values that sum to 1. Application to soil mapping then means that a gradual transition from one soil type to

another is represented by the gradual decrease and increase of membership values, as one moves across the boundaries between soil strata. Strictly speaking, the model underlying continuous classification is not a statistical model. It does not assess the accuracy of its predictions. Fuzzy set theory deals with possibilities instead of probabilities, although the subtle distinction between the two is a constant source of confusion (Fisher, 1994). One other source of confusion is that fuzzy set theory and statistics are both used for representing the uncertainty of soil information (McBratney, 1992).

Improvements to approaches that aim to combine soil survey information with continuous observations are foreseen. Particularly promising in this respect seems to be a related research area known as 'landscape-guided' soil mapping. Here landform and environmental attributes such as parent material, local relief, upstream area and vegetation serve as additional information in the kriging. Several recent examples explore this line (McKenzie and Austin, 1993; Odeh et al., 1994; Bourenanne et al., 1996; Lark, 1999; McKenzie and Ryan, 1999; Thomas et al., 1999; Lagacherie and Voltz, 2000). The success of the method depends on the strength of the relations between soil and landform. These relations are often found to be weak, but through a wise choice of plausible explanatory variables the approach might lead to substantial progress in the future.

3. Soil temporal variation

Many properties of the soil vary in time. Water content and potential, strength, and concentration of solutes come most immediately to mind. In fact, one might argue that every soil property changes in time. It is just that some change so slowly that for practical purposes we treat them as invariant in the time with which we are concerned. For instance, particle-size distribution is unlikely to change much over tens of years, but when we consider thousands of years (as we should when thinking of soil genesis) we can no longer assume it to be constant.

The models that people have used for temporal variation differ from those for variation in space, although there are similarities too, as we shall see. As it happens, the soil scientists who have tackled the two have hardly communicated with one another. This is both remarkable and unfortunate, because they produce ideas that should be useful to one another.

One of the main reasons that models of spatial and temporal variation are tackled differently is that the time dimension is different from the dimensions of space. It is more than just one dimension against two or three. Time has a direction, it moves forward only, processes take place in a sequence, and prediction is interesting only for the future, though occasionally we may wish to interpolate if we lose observations or have data that are too sparse. Also, the

laws of continuity and of the conservation of mass and energy must be obeyed in time, and they are so well known that it would be foolish to ignore them. Similar information cannot be used in a purely spatial setting. Thus, one big class of suitable models relies heavily on knowledge of process in time. Before dealing with these models, however, we discuss a second class that does bear relation to those of spatial variation discussed in the previous section. These are models stemming from time series analysis.

3.1. Time series analysis

If we were to measure a variable, such as the water content or matrix suction of the soil, at a certain place at regular intervals in time, then we should obtain a series of observations $z(t_i)$, $i = 1,2,\ldots,n$. In some situations this discrete time series of observations would be a sufficient description. But to predict $z(t)$ at some future time, or perhaps interpolate between time points when our measuring device failed, we should need a model characterizing the dynamic behaviour of $z(t)$. One of the earliest examples of which we know is that of Forbes (1846), who fitted a deterministic sine wave to data on soil temperature over several years. In time series analysis, which has its foundation in engineering, an empirical statistical approach is used. In the simplest case, the measured series is treated as the realization of a stationary random process $Z(t)$ satisfying

$$Z(t) = \mu + \varepsilon(t), \tag{9}$$

in which μ is the global mean as before and $\varepsilon(t)$ is a temporally autocorrelated random residual with a mean of zero and variance characterized by its autocovariance function $C(s) = \mathrm{Cov}[\varepsilon(t), \varepsilon(t+s)]$, where s denotes the lag. Note the resemblance here with geostatistics discussed in Section 2.2; the main difference is that now the lag is one-dimensional in time, whereas previously it was a vector in two and potentially three dimensions. Note also that the statistical properties of $Z(t)$ again depend on the support. The averaging effect mentioned in Section 2.2 is as valid in the temporal domain as in the spatial domain. Thus, the temporal support with which a soil variable is observed must be chosen wisely. For instance, measuring and modelling the nitrate concentration of the topsoil on an hourly support implies that short-term factors such as weather will have an effect on the concentration and should therefore be incorporated in the modelling. The effect of weather will be much less strong when the chosen support is the mean annual nitrate concentration.

Estimates of the autocovariances are typically plotted against the time lag, and these graphs give as much insight into the temporal behaviour as do the variograms in a spatial context. In soil science applications, it is more common simply to present graphs of the soil variable against time (Comegna and Vitale, 1993; Renschler et al., 1999; Van Es et al., 1999; Wendroth et al., 1999a; Knotters and Bierkens, 2001).

Direct fitting of (authorized) functions to plotted autocovariances is legitimate, but time-series analysts often take a somewhat different route to obtain the autocovariance function. They assume that the random process belongs to a specific class of processes, such as the class of auto-regressive moving-average (ARMA) processes. The simplest model of this class is that of a first-order autoregression:

$$Z(t) = \mu + a\{Z(t-1) - \mu\} + \varepsilon(t), \tag{10}$$

where a is a regression parameter, and $\varepsilon(t)$ is now an uncorrelated random residual. In other words, the temporal autocorrelation of $Z(t)$ is defined through the recurrence relation in Eq. (10).

Under an explicit model such as Eq. (10), the form of the autocovariance function is completely determined. For the first-order autoregression, the autocovariance function is approximately exponential, and exactly exponential when we replace the stochastic difference Eq. (10) by a stochastic differential equation.

Although such explicit modelling is rare in geostatistics, it has been done (Baker, 1984; Matérn, 1986). The problem is two-dimensional, of course, and so the model must recognize this. The natural spatial analogue of Eq. (10) is the first order spatial autoregressive model (Zhang and Yang, 1996; Heuvelink, 1998). It yields a variogram incorporating a Bessel function, specifically the modified Bessel function of the second kind, K_1, which derives from diffusion in two dimensions (Whittle, 1954; King and Smith, 1988). Indirect modelling of the variogram through an explicit model such as Eq. (10) ensures that the variogram obtained makes sense physically. This is by no means guaranteed with conventional variogram fitting. For instance, the most commonly used spherical variogram has properties that do not make much physical sense (Stein, 1999, p. 52).

We can predict future values of $Z(t)$ or interpolate between times optimally, given observations, based on either Eq. (9) or an explicit model formulation such as Eq. (10). The first leads effectively to kriging in one dimension; there is no fundamental difference. The second formulation leads to a somewhat different solution method, but in the end, the predictions and prediction error variances are the same as those obtained using kriging (Chirlin and Wood, 1982). This is because both use the same optimality criterion (i.e. they seek the best linear unbiased predictor).

In time-series analysis there are many extensions to the basic model, such as non-constant trends, seasonal effects (such as the deterministic sine wave fitted to soil temperature by Forbes (1846)), higher-order autoregressive terms, moving averages, and so on. Time-series analysis also leads conveniently into the frequency domain by Fourier transformation. Webster (1977) and Webster and Oliver (2001) have adapted it to space to find the frequency, and equivalently wavelength, of what appears to be periodic fluctuation in soil across a landscape.

3.2. Approaches based on physical processes

We now discuss models that describe dynamic behaviour based on physical laws such as that of conservation of mass. This is not the place to discuss these approaches at length. However, a brief description enables us to incorporate these essentially deterministic approaches into our statistical framework for modelling soil variation.

Physical–deterministic modelling of soil processes has a long history. Darcy's law and the laws of conservation of mass and energy were already used as points of departure for mathematical model building in the mid-1960s (Srivastava and Abrol, 1967) and even earlier, but the widespread use of these numerically intensive methods had to await the modern computer. We now have many implementations of models that can simulate processes such as the movement of water and solutes in the soil (e.g. Van Genuchten and Dalton, 1986; Petach et al., 1991). Likewise, we have ones to describe quantitatively the weathering and acidification of the soil, nitrification, and so on. These models, which we consider here under the heading of temporal models but which are strictly speaking spatio-temporal because they also take the vertical spatial dimension into account, share with the time-series models above the main goal of predicting the future conditions of the soil. Some of them have been successful, which explains their current popularity. Nevertheless, there are problems with calibration and validation, and these problems are particularly delicate because the models are often developed at a scale much finer than the scale of practical applications (Addiscott, 1993; Heuvelink and Pebesma, 1999). Also, no matter how elaborate the model becomes, there are always substantial deviations between the model's predictions and independent observations. Some of this discrepancy is caused by error in the model itself. The discrepancies between the model and reality are too complex either to understand or to model explicitly. They can be incorporated by our adding a stochastic residual. The soil variable $Z(t)$ then satisfies

$$Z(t) = f(t) + \varepsilon(t), \tag{11}$$

in which $f(t)$ represents the physical–deterministic part of the model, and $\varepsilon(t)$ is the stochastic component. Soil scientists have rarely taken this line, though we know of examples in soil hydrology (Bierkens, 1998; Russo, 1998).

3.2.1. The state-space approach

One important benefit of adding a stochastic residual to a deterministic model is that it allows us to decide rationally how predictions from the model may be corrected with independent observations. Consider Eq. (11) again, which in many cases may be formulated as

$$Z(t) = g\{Z(t-1)\} + \varepsilon(t). \tag{12}$$

This acknowledges that the state at time t depends on the state at time $t-1$. Eq. (12) is known as the state equation. In addition, we have a measurement equation:

$$Y(t) = h\{Z(t)\} + \eta(t), \tag{13}$$

where $Y(t)$ is the measurement, and $\eta(t)$ is the error in it. Note that the first term on the right hand side of Eq. (13) reads $h\{Z(t)\}$ instead of just $Z(t)$. This is useful to allow indirect or partial measurements of the state $Z(t)$. For instance, we may determine the soil water flux, denoted by $Z(t)$, indirectly from

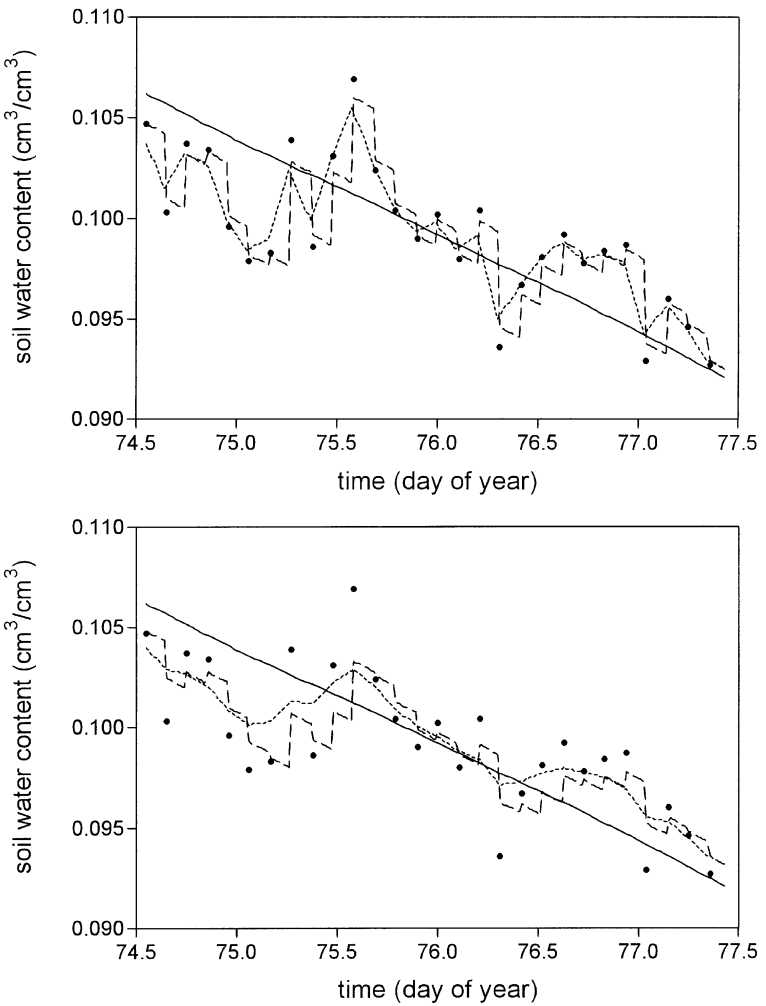

Fig. 4. Soil water content at a certain depth in a column of soil. The solid line is the prediction from the physical–deterministic model without conditioning to observations; the dots are observations. The dashed line is the Kalman filter, the dotted one is the Kalman smoother. Top: small measurement error. Bottom: large measurement error.

measurements of the soil water potential. Also, missing observations may be represented by assigning the null function to $h\{Z(t)\}$.

If we know the probability distribution of $Z(t-1)$, then we can use Eq. (12) to predict $Z(t)$. The accuracy of the prediction will be affected by the variance in $Z(t-1)$ and by the variance in the system noise $\varepsilon(t)$. Now, since we also know the accuracy of the observation $Y(t)$ (if we have any) through Eq. (13), we can update the original prediction with the observation and do so in a way that optimally weighs the uncorrected prediction with the observation. This is the essence of Kalman filtering (Kalman, 1960). Eqs. (12) and (13) constitute what is known as the 'state-space' approach. Optimal prediction in this setting, using the Kalman filter or related techniques, is nowadays also often termed 'data assimilation'. Starting from an initial situation $Z(t_0)$ and associated variance $P(t_0)$, the procedure moves forward one time step at the time. The computations involved are fairly straightforward in the usual situation where $g\{Z(t)\}$ and $h\{Z(t)\}$ are assumed linear and where the noise terms are normally distributed (Jazwinski, 1970). There are ways to relax these assumptions, but they typically yield approximate solutions only (Schweppe, 1973).

Kalman introduced his filter for signal processing, and it has since been used in various scientific fields. Despite its attractive properties, the applications in soil science remain few and are biased towards soil hydrology (Aboitiz et al., 1986; Bierkens, 1998; Galantowicz et al., 1999; Wendroth et al., 1999b; Hoeben and Troch, 2000). The state-space approach has also been used to identify the parameters of the physical–deterministic model (Katul et al., 1993; Cahill et al., 1999).

The strength of the state-space approach lies in its ability to merge the formulations of physical-deterministic models with observations. Generalizations of Eqs. (12) and (13) to include input (sources and sinks) from the outside world can be easily incorporated. Also, the extension from filtering to smoothing (using future observations to predict the present) is fairly straightforward. Kalman smoothing has much resemblance with (time series) interpolation, but the advantage is that it uses both process knowledge and observations (past, present and future) to predict the present.

Fig. 4 demonstrates the principle of the Kalman filter and smoother applied to an experiment in which the water content is measured in a column of soil.

4. Soil spatio-temporal variation

Few soil scientists have attempted to analyse variation of soil in space and time simultaneously, and those who have done have concentrated on the comparison of spatial and temporal variation (Goovaerts and Chiang, 1993; Van Es, 1993; Comegna and Basile, 1994; Mohanty et al., 1998; Cain et al., 1999; Rockström et al., 1999; Van Es et al., 1999; Wendroth et al., 1999a; Chevalier et

al., 2000). One notable exception is in the work of Comegna and Vitale (1993), who modelled the variation of the soil water status simultaneously in space and time using a spatial generalization of the classical autoregressive model (which shows much resemblance with the first-order spatial autoregressive model mentioned in Section 3.1). Indeed, it is only recently that geostatisticians more generally have started to extend their spatial models to the time domain. Physically based space–time modelling has often been restricted to purely deterministic approaches.

4.1. Space–time geostatistics

Conventional geostatistics deals with variation exclusively in space. Soil is dynamic, however, and several of its attributes that vary spatially, such as its water content and pressure, heat flow, and solute concentration, vary in time also. Soil scientists have recognized this, and they are now attempting to model its behaviour in space and time simultaneously and to apply spatio-temporal kriging (Papritz and Flühler, 1994; Bogaert and Christakos, 1997; Heuvelink et al., 1997; Hoosbeek et al., 1998).

The extension of kriging to the space–time domain is not straightforward. Incorporating time is more than just adding a third (or fourth) dimension. This is because the behaviour of an attribute over time is often very different from its behaviour in space, as was already pointed out in Section 3. Consequently, suitable geostatistical models characterizing the spatio-temporal behaviour must take these differences into account. Much of the current research in spatio-temporal geostatistics is directed to the design, structural analysis, and application of such models (Kyriakidis and Journel, 1999). The general model is as follows:

$$Z(\boldsymbol{x},t) = m(\boldsymbol{x},t) + \varepsilon(\boldsymbol{x},t). \tag{14}$$

Here the term $m(\boldsymbol{x},t)$ is a deterministic trend (in the simplest case a constant, i.e. equal to a global mean μ), and $\varepsilon(\boldsymbol{x},t)$ is a zero-mean, second-order stationary residual with autocovariance function $C(\boldsymbol{h},s) = \mathrm{Cov}[\varepsilon(\boldsymbol{x},t),\varepsilon(\boldsymbol{x}+\boldsymbol{h}, t+s)]$. Because of the difference in behaviour in space and time, the residual $\varepsilon(\boldsymbol{x},t)$ will usually have space–time anisotropies. This makes identification of $C(\boldsymbol{h},s)$ from point observations of $Z(\boldsymbol{x},t)$ difficult, especially because a further simplification by strict separation of variability in space from that in time through $C(\boldsymbol{h},s) = C_1(\boldsymbol{h}) + C_2(s)$ yields unrealistic results (Rouhani and Myers, 1990; Dimitrakopoulos and Luo, 1994; Heuvelink et al., 1997).

In the usual geostatistical setting we find that we have three-dimensional models with zonal and geometric anisotropies. Geometric anisotropies arise because a unit in time is not the same as a unit in space. An anisotropy ratio is therefore needed to match the temporal range of the variogram with its spatial range. Zonal anisotropies will arise because the amount of variation in time may be smaller or larger than that in space. Temporal persistence of spatial patterns

implies that spatial variation is larger than temporal variation, spatial persistence of temporal patterns implies the opposite. Vachaud et al. (1985), Goovaerts and Chiang (1993), Comegna and Basile (1994) and Heuvelink et al. (1997) have recorded temporal persistence of spatial patterns of mineralizable nitrogen, soil water content, and other soil properties. Fig. 5 shows space and time slices taken from a realization of a random process with spatial persistence of temporal patterns. Fig. 6 shows space and time slices from a realization of a random process with temporal persistence of spatial patterns.

Modelling the space–time variogram is likely to be difficult, and we need to think what models and fitting techniques we should use for the purpose. If we can do that, however, we should be able to krige in the usual way to obtain optimal predictions of the soil property at points in space and time from observations nearby in both. So in this sense there is no big difference, and existing routines will be applicable.

An important disadvantage of the purely geostatistical approaches to space–time modelling is that it is not easy to include physically based knowledge in the model. For this we need to rewrite the model expressed by Eq. (14) into state-space form, just as we did in the previous section where we addressed temporal soil variation only.

4.2. The spatio-temporal Kalman filter

We introduced the state-space approach in Section 3.2 where we limited it to time only; but there is a natural extension to the combination of space and time. Let us rewrite Eqs. (12) and (13) as

$$\mathbf{Z}(t) = \mathbf{A}(t)\mathbf{Z}(t-1) + \boldsymbol{\varepsilon}(t) \tag{15}$$

and

$$\mathbf{Y}(t) = \mathbf{B}(t)\mathbf{Z}(t) + \boldsymbol{\eta}(t), \tag{16}$$

where we have assumed the linear form. In Eqs. (15) and (16) $\mathbf{Z}(t)$ is a (very large) random vector representing the soil variable in all layers at all locations in the (discretized) spatial domain. We denote it with the bold capital $\mathbf{Z}$ to indicate that it is random; it is not a matrix. Likewise $\mathbf{Y}(t)$ is also a random vector. Matrix $\mathbf{A}(t)$ describes how the state at time t depends on the state at time $t-1$; lateral or vertical flow is represented by assigning non-zero values to the off-diagonal elements of $\mathbf{A}(t)$. Matrix $\mathbf{B}(t)$ will usually be a diagonal or even the identity matrix (measurements are rarely influenced by the surrounding), although there are exceptions. For instance, measurements made at a larger spatial support, such as when soil water content is measured over the whole soil profile (one may think of a soil column that is placed on a balance), may be represented by assigning equal 'weights' to the corresponding cells in the matrix $\mathbf{B}(t)$. Temporal correlation in $\mathbf{Z}(t)$ is introduced through the recurrence relation

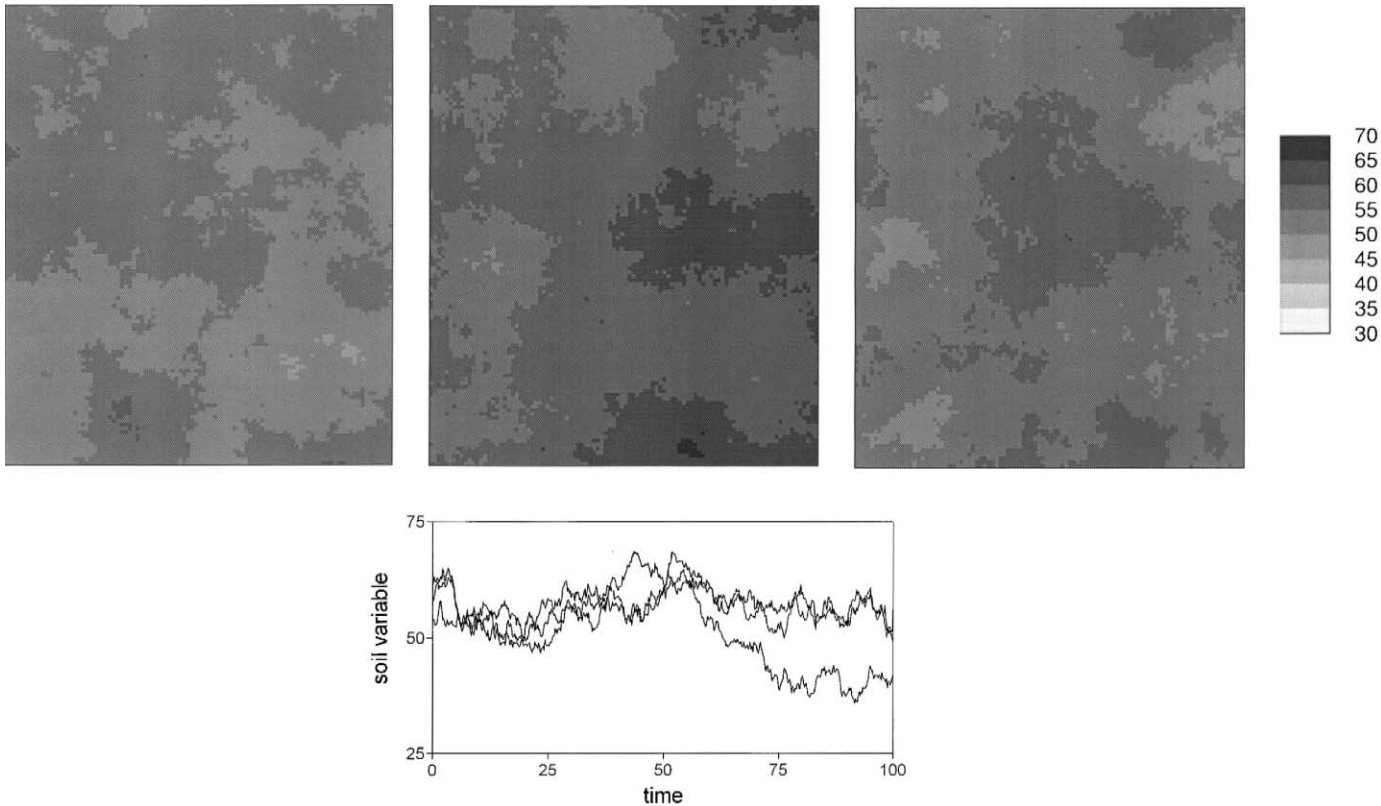

Fig. 5. Time and space slices taken from a realization of a $2D+$time stationary Gaussian random process exhibiting spatial persistence of temporal patterns. Mean $\mu = 50$, the variogram used is a spherical model with zero nugget and range 40 (space and time). Temporal sill equals 60, spatial sill 6. Realization obtained for a 100×120 spatial rectangle $\times [0,100]$ time interval. Time slices taken at $t = 20$ (top left), $t = 60$ (top middle) and $t = 70$ (top right). Space slices (bottom) taken at $\boldsymbol{x} = \{40,70\}$, $\boldsymbol{x} = \{50,60\}$ and $\boldsymbol{x} = \{90,110\}$.

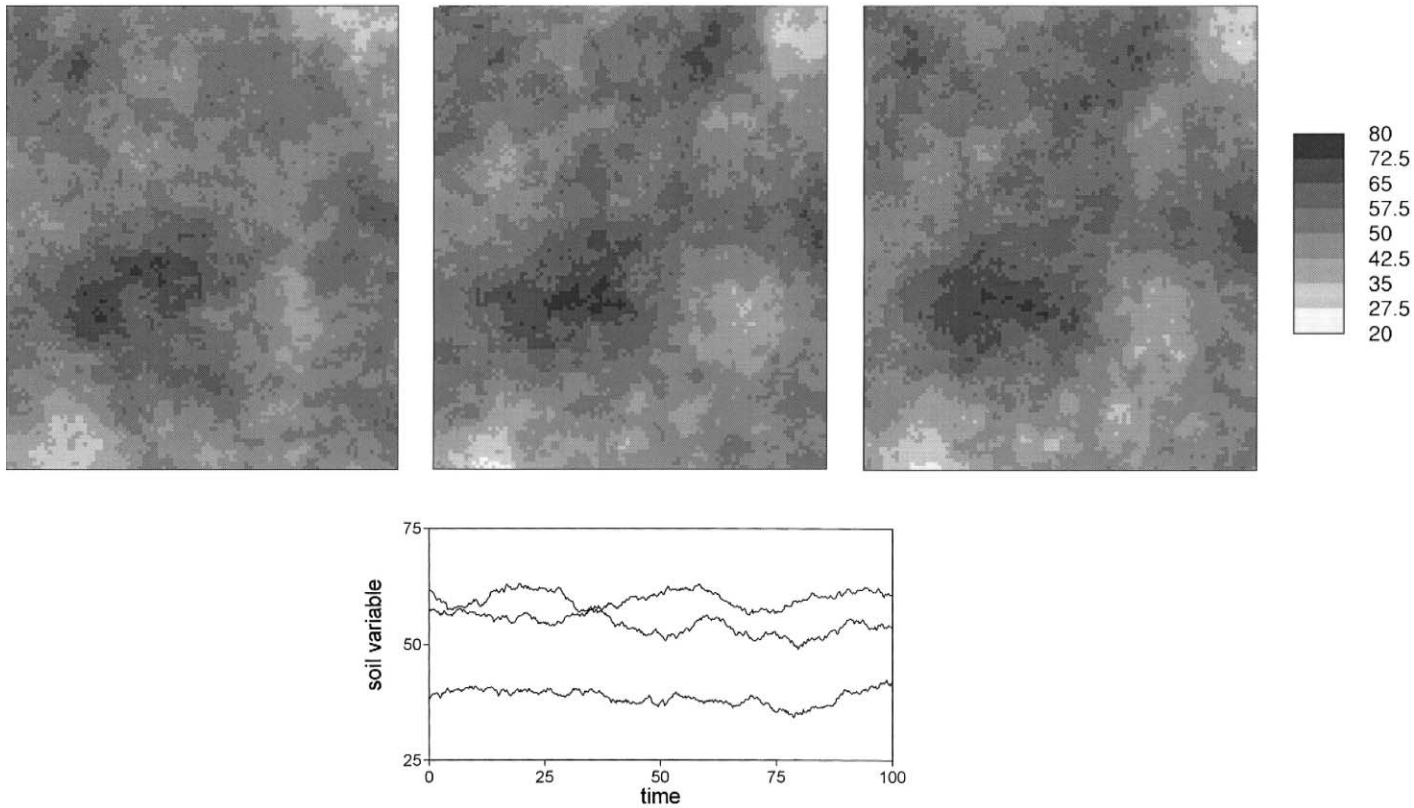

Fig. 6. Time and space slices taken from a realization of a $2D+$time stationary Gaussian random process exhibiting temporal persistence of spatial patterns. Mean $\mu = 50$, the variogram used is a spherical model with zero nugget and range 40 (space and time). Spatial sill equals 60, temporal sill 6. Realization obtained for a 100×120 spatial rectangle $\times$ [0,100] time interval. Time slices taken at $t = 20$ (top left), $t = 60$ (top middle) and $t = 70$ (top right). Space slices (bottom) taken at $\boldsymbol{x} = \{40,70\}$, $\boldsymbol{x} = \{50,60\}$ and $\boldsymbol{x} = \{90,110\}$.

in Eq. (15), spatial correlation through non-zero correlations between components of the noise vector $\boldsymbol{\varepsilon}(t)$. In Eq. (15) the law of continuity can be easily incorporated by placing constraints on the coefficients in $\mathbf{A}(t)$, i.e. by requiring the elements of rows of $\mathbf{A}(t)$ to sum to 1.

Just as before, under the model defined by Eqs. (15) and (16) and a few additional assumptions, the prediction of $\mathbf{Z}(t)$ at an unobserved point in space and time is obtained from the Kalman filter or smoother. Now, of course, it is in a space–time context, but the space is hidden in the very large vector. So the spatial extension can be dealt with at the expense of a strong increase of the dimensionality (a 100×100 two-dimensional grid yields a vector of length 10 000, so the variance–covariance matrix is $10\,000 \times 10\,000$). This causes numerical problems, which is an important reason why we have seen so little of the spatio-temporal Kalman filter so far. But computers are becoming ever faster, and much work is being done to find efficient routines that solve the problem at least approximately (Huang and Cressie, 1996; Verlaan, 1998; Zhang et al., 1999). Many of these techniques exploit the sparsity of the system matrices or make approximations to facilitate computations.

Or and Hanks (1992) used Kalman filtering in combination with kriging to reduce the computational load. Morkoc et al. (1985) used the state-space approach for describing variation in soil, but their work has been followed up only very recently (Bierkens et al., 2001; Knotters and Bierkens, 2001).

A continuous time space analogue of Eqs. (15) and (16) may also be formulated. This results in a similar solution to the filtering problem, although the mathematics behind it is more advanced.

One major difficulty with the state-space approach to space–time prediction is to identify the parameters. The matrices in Eqs. (15) and (16) contain many parameters that must be estimated from observations. The same holds for the variance–covariance matrix of the system and measurement noise (or even the full joint distribution). Although this problem has been rigorously addressed in stochastic systems theory (Jazwinski, 1970), current procedures are numerically intensive and will usually involve subjective choices to reduce the number of estimable parameters.

4.3. Soil-landscape evolution modelling

Spatio-temporal Kalman filtering is presented above for merging knowledge of processes with observations for modelling variation in soil properties in space and time. What comes first to mind are processes that take place over fairly short times, such as water and solute movement. But there is nothing against using the same approach for longer times, such as we might postulate for the development of the soil landscape.

We started out discourse by reviewing soil classification. We noted that in many instances pedologists had an underlying notion (a mental model) of how

the soil formed from its parent material in the particular climate over time. This required an appreciation of how geology and climate interact with tectonics and geomorphology, by erosion, sedimentation, surface and subsurface flow, infiltration, weathering, organic matter accumulation, etc., and thus form the soil. The knowledge and understanding used in soil survey to map the soil are undoubtedly subjective and qualitative. We put them aside and turned our attention to geostatistics, though we tried to merge the two. But now we have added the temporal dimension, and this opens up possibilities.

Ever since its introduction, Jenny's (1941) formulation of the State Factor Model (SFM) of soil formation has generated much discussion. This model states that the soil is a function of five soil-forming factors, namely climate, organisms, relief, parent material and time. If such a model were to become quantitatively operational, then we should have a very powerful tool to aid and test our understanding. But can the SFM be made operational? Kline (1973) noted that the equations derived from Jenny's expression 'would be of unprecedented complexity, and probably completely intractable'. Time has not stood still, however, and soil scientists now take a more optimistic view of the feasibility of building simulation models of pedogenesis (Hoosbeek and Bryant, 1992; Burrough, 1993b), and we already witness some first rudimentary implementations (Minasny and McBratney, 1999).

The problems of successfully implementing the SFM can hardly be underestimated. Parts of the SFM are highly non-linear, and there is immense scope for positive feedback; so small variations in the initial state of the soil persist and grow over time. This, together with the fact that the initial state is largely unknown, means that we shall never know exactly how a given site was perturbed over the course of tens of thousands of years (Phillips, 1998). Implementing the SFM to predict the soil will inevitably induce large uncertainties. Predictions from the SFM must therefore be conditioned on observations. The state-space approach provides the framework for this.

When framed in terms of fluxes and exchanges of energy and matter, the SFM is mathematically formulated in terms of differential equations. Discretizing these differential equations and adding noise to the system yields equations of the type (15) (but then probably the non-linear form). Adding noise is sensible because, although we have some idea of the structural equations describing soil formation, our knowledge of the parameters and inputs is so incomplete. So there is much uncertainty, which should be conveyed explicitly. Further, we shall want to use our observations to improve our predictions, and a measurement equation of the type (16) is defined in addition.

If the SFM can be cast in a state-space form, then data assimilation techniques (i.e. the Kalman filter) may be used to predict the soil. In this way, information from the SFM will be optimally combined with the observations. When we are very uncertain about the SFM, the noise in the system will be large. As a result, the prediction will rely heavily on the observations. The result

will be conventional geostatistics. When our observations are scarce or contain large measurement errors, we shall rely more on our knowledge of soil formation as contained in the state equation.

5. Conclusions: a look into the crystal ball

There is much diversity in the approaches to modelling soil variation. This is in part because there are so many sources of information. At the landscape scale and over times of millenia and longer, there are our theories or hypotheses of soil genesis, which state how geology and climate are believed to cause variation in soil. At the scale of the individual field and over times of minutes, hours, days and months, we may use laws of physical processes, expressed quantitatively in terms of Richards' equation and Fick's law, for example. Last but not least, there are the observations. Over the years these sources of information have been weighed differently in models of soil and its behaviour.

In the 1960s soil variation was described largely qualitatively and was modelled on an understanding of soil genesis, again almost entirely qualitative. When that approach was found wanting pedologists switched to a purely empirical statistical approach, namely geostatistics that had proved so effective in mining. This approach, which relies almost solely on the information contained in the observations, has become increasingly refined. It has its analogue in time series analysis, but this has not been rigorously explored.

For the future, we think that soil scientists should use more of the collective knowledge of processes (i.e. SFM) when mapping soil. In fact, we are not alone in this view (Hoosbeek and Bryant, 1992; Burrough, 1993b). By using data assimilation techniques such as the Kalman filter and smoother, we may be able to predict soil conditions more accurately, because these techniques make use of all sources of information, and weigh them in accordance with their accuracy. Reduction of prediction error variance will be the most important advantage from a practitioner's point of view. However, from a scientific point of view incorporating the SFM is first and foremost justified because it may aid in *explaining* soil variation. In the words of Amundson (1998): it provides '. . . a blue print from which we may begin to unravel the mysteries of our surroundings'.

The new approach is, we believe, promising, and it may advance the environmental correlation approaches to soil-landscape modelling discussed in Section 2.3. The current procedures suffer from using terrain attributes of the present time to predict the soil, whereas the soil was formed in the past when the environment was very different. One solution to this problem seems to be to adopt a dynamic approach, as described in Section 4.3.

Incorporating knowledge of processes into statistical models is intuitively desirable, but the parameterization is difficult. When advocating this we have to

grapple with the well-known problems of identifying the parameters as mechanistic or functional, and there is a limit on the extent to which one can go. Perhaps in many cases one simply has to stick to purely empirical models that rely on observations only. One major problem is to quantify processes that took place in the past yet of which little is known quantitatively. But it is worth attempting a solution. It will require a joint effort of scientists with varied backgrounds: they will have to integrate geology, geomorphology, hydrology, vegetation science, and soil science. There are loops: soil affects vegetation, vegetation affects soil; geomorphology affects soil, soil affects geomorphology; and so on.

So, there is much to be done and many problems to be resolved to make things work. The problems are not only computational, we have also to transform conceptual models into mathematical formulae, we have to deal with the great deal of uncertainty about past conditions (we do not have detailed digital terrain models or land cover maps of 1000 years ago, for example), we have the non-linearities and the non-stationarities. Our work as designers and builders of models that can adequately describe the complex variation of the soil is far from finished.

Acknowledgements

We thank Ate Visser for providing the data for Fig. 4. We thank Alfred Hartemink, Alex McBratney and Judith Snepvangers for valuable comments on a draft of our script.

References

Aboitiz, M., Labadie, J.W., Heermann, D.F., 1986. Stochastic soil moisture estimation and forecasting for irrigated fields. Water Resources Research 22, 180–190.

Addiscott, T.M., 1993. Simulation modelling and soil behaviour. Geoderma 60, 15–40.

Amundson, R., 1998. In Discussion of: J.D. Phillips. On the relations between complex systems and the factorial model of soil formation (with Discussion). Geoderma 86, 25–27.

Baker, R., 1984. Modeling soil variability as a random field. Journal of the International Association of Mathematical Geology 16, 435–448.

Beckett, P.H.T., Burrough, P.A., 1971. The relation between cost and utility in soil survey: IV. Comparison of the utilities of soil maps produced by different survey procedures, and to different scales. Journal of Soil Science 22, 460–480.

Bierkens, M.F.P., 1998. Modeling water table fluctuations by means of a stochastic differential equation. Water Resources Research 34, 2485–2499.

Bierkens, M.F.P., Finke, P.A., De Willigen, P., 2000. Upscaling and Downscaling Methods for Environmental Research. Kluwer Academic Publishers, Dordrecht.

Bierkens, M.F.P., Knotters, M., Hoogland, T., 2001. Space–time modeling of water table depth using a regionalized time series model and the Kalman filter. Water Resources Research (in press).

Bogaert, P., Christakos, G., 1997. Stochastic analysis of spatiotemporal solute content measurements using a regression model. Stochastic Hydrology and Hydraulics 11, 267–295.
Boucneau, G., Van Meirvenne, M., Thas, O., Hofman, G., 1998. Integrating properties of soil map delineations into ordinary kriging. European Journal of Soil Science 49, 213–229.
Bourenanne, H., King, D., Chéry, P., Bruand, A., 1996. Improving the kriging of a soil variable using slope gradient as external drift. European Journal of Soil Science 47, 473–483.
Brannon, G.R., Hajek, B.F., 2000. Update and recorrelation of soil surveys using GIS and statistical analysis. Soil Science Society of America Journal 64, 679–680.
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, 1–59.
Brus, D.J., De Gruijter, J.J., Marsman, B.A., Visschers, R., Bregt, A.K., 1996. The performance of spatial interpolation methods and choropleth maps to estimate properties at points: a soil survey case study. Environmetrics 7, 1–16.
Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties: I. The semi-variogram and punctual kriging. Journal of Soil Science 31, 315–331.
Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40, 477–492.
Burrough, P.A., 1993a. Soil variability: a late 20th century view. Soils and Fertilizers 56, 529–562.
Burrough, P.A., 1993b. The technologic paradox in soil survey: new methods and techniques of data capture and handling. ITC Journal 1, 15–22.
Butler, B.E., 1980. Soil Classification for Soil Survey. Oxford Univ. Press, Oxford.
Cahill, A.T., Ungaro, F., Parlange, M.B., Mata, M., Nielsen, D.R., 1999. Combined spatial and Kalman filter estimation of optimal soil hydraulic properties. Water Resources Research 35, 1079–1088.
Cain, M.L., Subler, S., Evans, J.P., Fortin, M.-J., 1999. Sampling spatial and temporal variation in soil nitrogen availability. Oecologia 118, 397–404.
Campbell, D.J., Kinniburgh, D.G., Beckett, P.H.T., 1989. The soil solution chemistry of some Oxfordshire soils: temporal and spatial variability. Journal of Soil Science 40, 321–339.
Chevalier, T., Voltz, M., Blanchart, E., Chotte, J.L., Eschenbrenner, V., Mahieu, M., Albrecht, A., 2000. Spatial and temporal changes of soil C after establishment of a pasture on a long-term cultivated vertisol (Martinique). Geoderma 94, 43–58.
Chirlin, G.R., Wood, E.F., 1982. On the relationship between kriging and state estimation. Water Resources Research 18, 432–438.
Comegna, V., Basile, A., 1994. Temporal stability of spatial patterns of soil water storage in a cultivated Vesuvian soil. Geoderma 62, 299–310.
Comegna, V., Vitale, C., 1993. Space–time analysis of water status in a volcanic Vesuvian soil. Geoderma 60, 135–158.
Cressie, N.A.C., 1993. Statistics for Spatial Data (revised edition). Wiley, New York.
De Gruijter, J.J., Walvoort, D.J.J., Van Gaans, P.F.M., 1997. Continuous soil-maps—a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77, 169–195.
Delhomme, J.-P., 1978. Kriging in the hydrosciences. Advances in Water Resources 1, 251–266.
Dijkerman, J.C., 1974. Pedology as a science: the role of data, models and theories in the study of natural soil systems. Geoderma 11, 73–93.
Dimitrakopoulos, R., Luo, X., 1994. Spatiotemporal modelling: covariances and ordinary kriging systems. In: Dimitrakopoulos, R. (Ed.), Geostatistics for the Next Century. Kluwer Academic Publishers, Dordrecht, pp. 88–93.
Fisher, P.F., 1994. Probable and fuzzy models of the viewshed operation. In: Worboys, M.F. (Ed.), Innovations in GIS 1. Taylor and Francis, London, pp. 161–175.

Forbes, J.D., 1846. Account of some experiments on the temperature of the earth at different depths and in different soils near Edinburgh. Transactions of the Royal Society of Edinburgh 16, 189–236.

Galantowicz, J.F., Entekhabi, D., Njoku, E.G., 1999. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. IEEE Transactions on Geoscience and Remote Sensing 37, 1860–1870.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford Univ. Press, New York.

Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89, 1–45.

Goovaerts, P., Chiang, C.N., 1993. Temporal persistence of spatial patterns for mineralizable nitrogen and selected soil properties. Soil Science Society of America Journal 57, 372–381.

Goovaerts, P., Journel, A.G., 1995. Integrating soil map information in modelling the spatial variation in continuous soil properties. European Journal of Soil Science 46, 397–414.

Heuvelink, G.B.M., 1996. Identification of field attribute error under different models of spatial variation. International Journal of GIS 10, 921–935.

Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modelling with GIS. Taylor and Francis, London.

Heuvelink, G.B.M., Bierkens, M.F.P., 1992. Combining soil maps with interpolations from point observations to predict quantitative soil properties. Geoderma 55, 1–15.

Heuvelink, G.B.M., Huisman, J.A., 2000. Choosing between abrupt and gradual spatial variation? In: Mowrer, H.T., Congalton, R.G. (Eds.), Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing. Ann Arbor Press, Chelsea, MI, pp. 111–117.

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

Heuvelink, G.B.M., Musters, P., Pebesma, E.J., 1997. Spatio-temporal kriging of soil water content. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong '96. Kluwer Academic Publishers, Dordrecht, pp. 1020–1030.

Hoeben, R., Troch, P.A., 2000. Assimilation of active microwave observation data for soil moisture profile estimation. Water Resources Research 36, 2805–2819.

Hoosbeek, M.R., Bryant, R.B., 1992. Towards the quantitative modeling of pedogenesis—a review. Geoderma 55, 183–210.

Hoosbeek, M.R., Van Es, H.M., Stein, A. (Eds.), 1998. Modeling Spatial and Temporal Variability as a Function of Scale. Geoderma 85, nos. 2–3.

Huang, H.C., Cressie, N., 1996. Spatiotemporal prediction of snow water equivalent using the Kalman filter. Computational Statistics and Data Analysis 22, 159–175.

Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.

Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Transactions of the American Society of Mechanical Engineers: Journal of Basic Engineering 82D, 35–45.

Kantey, B.A., Williams, A.A.B., 1962. The use of soil engineering maps for road projects. Transactions of the South African Institution of Civil Engineers 4, 149–159.

Katul, G.G., Wendroth, O., Parlange, M.B., Puente, C.E., Folegatti, M.V., Nielsen, D.R., 1993. Estimation of in situ hydraulic conductivity function from nonlinear filtering theory. Water Resources Research 29, 1063–1070.

Kern, J.S., 1994. Spatial patterns of soil organic carbon in the contiguous United States. Soil Science Society of America Journal 58, 439–455.

King, P.R., Smith, P.J., 1988. Generation of correlated properties in heterogeneous porous media. Mathematical Geology 20, 863–877.

Kline, J.R., 1973. Mathematical simulation of soil–plant relationships and soil genesis. Soil Science 115, 240–249.
Knotters, M., Bierkens, M.F.P., 2001. Predicting water table depths in space and time using a regionalised time series model. Geoderma, in press.
Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space–time models: a review. Mathematical Geology 31, 651–684.
Lagacherie, P., Voltz, M., 2000. Predicting soil properties over a region using sample information from a mapped reference area and digital elevation data: a conditional probability approach. Geoderma 97, 187–208.
Lark, R.M., 1999. Soil–landform relationships at within-field scales: an investigation using continuous classification. Geoderma 92, 141–165.
Lark, R.M., 2000a. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science 51, 137–157.
Lark, R.M., 2000b. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. European Journal of Soil Science 51, 717–728.
Lark, R.M., 2000c. A geostatistical extension of the sectioning procedure for disaggregating soil information to the scale of functional models of soil processes. Geoderma 95, 89–112.
Leenhardt, D., Voltz, M., Bornand, M., Webster, R., 1994. Evaluating soil maps for prediction of soil water properties. European Journal of Soil Science 45, 293–301.
Marechal, A., 1984. Kriging seismic data in presence of faults. In: Verly, G., David, M., Journel, A.G., Marechal, A. (Eds.), Geostatistics for Natural Resources Characterization, Part 1. D. Reidel, Dordrecht, pp. 271–294.
Matérn, B., 1986. Spatial Variation. 2nd edn. Lecture Notes in Statistics, vol. 36, Springer, Berlin.
McBratney, A.B., 1992. On variation, uncertainty and informatics in environmental soil management. Australian Journal of Soil Research 30, 913–935.
McKenzie, N.J., Austin, M.P., 1993. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. Geoderma 57, 329–355.
McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.
Minasny, B., McBratney, A.B., 1999. A rudimentary mechanistic model for soil production and landscape development. Geoderma 90, 3–21.
Mohanty, B.P., Shouse, P.J., Van Genuchten, M.T., 1998. Spatio-temporal dynamics of water and heat in a field soil. Soil and Tillage Research 47, 133–143.
Morkoc, F., Biggar, J.W., Nielsen, D.R., Rolston, D.E., 1985. Analysis of soil water content using state-space approach. Soil Science Society of America Journal 49, 798–803.
Morse, R.K., Thornburn, T.H., 1961. Reliability of soil maps. Proceedings of the 5th International Conference on Soil Mechanics and Foundation Engineering. I. Dunod, Paris, pp. 259–262.
Nettleton, W.D., Brasher, B.R., Borst, G., 1991. The taxadjunct problem. Soil Science Society of America Journal 55, 421–427.
Nortcliff, S., 1978. Soil variability and reconnaissance soil mapping: a statistical study in Norfolk. Journal of Soil Science 29, 403–418.
Oberthür, T., Dobermann, A., Neue, H.U., 1996. How good is a reconnaisance soil map for agronomic purposes? Soil Use and Management 12, 33–43.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63, 197–214.
Or, D., Hanks, R.J., 1992. Spatial and temporal soil water estimation considering soil variability and evapotranspiration uncertainty. Water Resources Research 28, 803–814.
Papritz, A., Flühler, H., 1994. Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. Geoderma 62, 29–43.
Petach, M.C., Wagenet, R.J., DeGloria, S.D., 1991. Regional water flow and pesticide leaching using simulations with spatially distributed data. Geoderma 48, 245–269.

Phillips, J.D., 1998. On the relations between complex systems and the factorial model of soil formation (with Discussion). Geoderma 86, 1–42.
Renschler, C.S., Mannaerts, C., Diekkrüger, B., 1999. Evaluating spatial and temporal variability in soil erosion risk-rainfall erosivity and soil loss ratios in Andalusia, Spain. Catena 34, 209–225.
Rockström, J., Barron, J., Brouwer, J., Galle, S., De Rouw, A., 1999. On-farm spatial and temporal variability of soil and water in pearl millet cultivation. Soil Science Society of America Journal 63, 1308–1319.
Rogowski, A.S., Wolf, J.K., 1994. Incorporating variability into soil map unit delineations. Soil Science Society of America Journal 58, 163–174.
Rouhani, S., Myers, D.E., 1990. Problems in space–time kriging of geohydrological data. Mathematical Geology 22, 611–623.
Russo, D., 1998. Stochastic analysis of flow and transport in unsaturated heterogeneous porous formation: effects of variability in water saturation. Water Resources Research 34, 569–581.
Schelling, J., 1970. Soil genesis, soil classification and soil survey. Geoderma 4, 165–193.
Schweppe, F.C., 1973. Uncertain Dynamical Systems. Prentice-Hall, Englewood Cliffs, NJ.
Srivastava, R.C., Abrol, I.P., 1967. Non-linear thermodynamics of irreversible processes and transport of material and energy through soil systems. Geoderma 1, 69–75.
Stein, M., 1999. Interpolation of Spatial Data. Some Theory for Kriging. Springer, New York.
Stein, A., Hoogerwerf, M., Bouma, J., 1988. Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. Geoderma 43, 163–177.
Thomas, A.L., King, D., Dambrine, E., Couturier, A., Roque, J., 1999. Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges mountains (Northeastern France). Geoderma 90, 291–305.
Utset, A., López, T., Díaz, M., 2000. A comparison of soil maps, kriging and a combined method for spatially predicting bulk density and field capacity of ferrasols in the Havana-Matanzas Plain. Geoderma 96, 199–213.
Vachaud, G., Passerat De Silans, A., Balabanis, P., Vauclin, M., 1985. Temporal stability of spatially measured soil water probability density function. Soil Science Society of America Journal 49, 822–828.
Van Es, H.M., 1993. Evaluation of temporal, spatial, and tillage-induced variability for parametrization of soil infiltration. Geoderma 60, 187–199.
Van Es, H.M., Ogden, C.B., Hill, R.L., Schindelbeck, R.R., Tsegaye, T., 1999. Integrated assessment of space, time, and management-related variability of soil hydraulic properties. Soil Science Society of America Journal 63, 1599–1608.
Van Genuchten, M.T., Dalton, F.N., 1986. Models for simulating salt movement in aggregated field soils. Geoderma 38, 165–183.
Van Kuilenburg, J., De Gruijter, J.J., Marsman, B.A., Bouma, J., 1982. Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units. Geoderma 27, 311–325.
Verlaan, M., 1998. Efficient Kalman filtering algorithms for hydrodynamic models. PhD Thesis, Delft University of Technology, Delft.
Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. Journal of Soil Science 31, 505–524.
Webster, R., 1977. Spectral analysis of gilgai soil. Australian Journal of Soil Research 15, 191–204.
Webster, R., 2000. Is soil variation random? Geoderma 97, 149–163.
Webster, R., 2001. Statistics to support soil research and their presentation. European Journal of Soil Science 52, in press.
Webster, R., Beckett, P.H.T., 1968. Quality and usefulness of soil maps. Nature (London) 219, 680–682.

Webster, R., Beckett, P.H.T., 1970. Terrain classification and evaluation using air photography: a review of recent work at Oxford. Photogrammetria 26, 51–75.

Webster, R., Cuanalo de la C., H.E., 1975. Soil transect correlograms of North Oxfordshire and their interpretation. Journal of Soil Science 26, 176–194.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists. Wiley, Chichester.

Wendroth, O., Pohl, W., Koszinski, S., Rogasik, H., Ritsema, C.J., Nielsen, D.R., 1999a. Spatio-temporal patterns and covariance structures of soil water status in two Northeast-German field sites. Journal of Hydrology 215, 38–58.

Wendroth, O., Rogasik, H., Koszinski, S., Ritsema, C.J., Dekker, L.W., Nielsen, D.R., 1999b. State-space prediction of field-scale soil water content time series in a sandy loam. Soil and Tillage Research 50, 85–93.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434–449.

Wösten, J.H.M., Bouma, J., Stoffelsen, G.H., 1985. The use of soil survey data for regional soil water simulation models. Soil Science Society of America Journal 49, 1238–1244.

Yost, R.S., Loague, K., Green, R.E., 1993. Reducing variance in soil organic carbon estimates: soil classification and geostatistical approaches. Geoderma 57, 247–262.

Zhang, R., Yang, J., 1996. Iterative solution of a stochastic differential equation: an efficient method for simulating soil variability. Geoderma 72, 75–88.

Zhang, X.F., Heemink, A.W., Janssen, L.H.J.M., Janssen, P.H.M., Sauter, F.J., 1999. A computationally efficient Kalman smoother for the evaluation of the CH_4 budget in Europe. Applied Mathematical Modelling 23, 109–129.