European Journal of Soil Science, June 2007, **58,** 822–831 doi: 10.1111/j.1365-2389.2006.00872.x

Inference about soil variability from the structure of the best wavelet packet basis

R. M. LARK
Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

Summary

The plausibility of the assumption that soil variation can be treated as a realization of a random spatial process that is stationary in the variance can break down in various ways. It is possible to test the assumption using methods based on the wavelet transform. To date these approaches have been applied using the discrete wavelet transform. A drawback of this approach is that it uses a partition of the spatial frequencies represented in the data into intervals (scales) that are arbitrarily defined in advance and are not necessarily suitable for the representation of the variation of the data in question. A solution to this problem is to identify the best basis for the data from a wavelet packet library. An interesting question is whether the structure of this best basis is in itself informative about the plausibility of the stationarity assumption. In this paper, I show that this is indeed the case. The best basis for a stationary random variable from some packet library is the basis on the maximum dilation of the mother wavelet, which gives the finest resolution in the frequency domain. I propose the ratio of the entropy cost functional for this basis to that of the empirical best basis as a measure of evidence against the null hypothesis of stationarity in the variance. Critical values of this statistic may be obtained by Monte Carlo methods. I demonstrate the method using data on the clay content of soil on a transect in central England. The null hypothesis of stationarity in the variance may be rejected. Tests for the uniformity of variance can then be applied to wavelet packets in the best basis. The dominant local feature that is reflected in this behaviour is the unique pattern of variation in alluvium around a drainage channel that crosses the transect. This variation contrasts with that seen at most positions on the transect, variation that arises from a more or less regular pattern of boundaries between contrasting Jurassic strata.

Introduction

> *…'the possibility of describing the world by means of Newtonian mechanics tells us nothing about the world: but what does tell us something about it is the precise way in which it is possible to describe it by these means. We are also told something about the world by the fact that it can be described more simply with one system of mechanics than with another.'*
>
> Wittgenstein, *Tractatus Logico-Philosophicus* 6.3.4.2

In the geostatistical study of the soil, we assume that a soil property, z, at some location, $\mathbf{x}$, is a realization of a random function $Z(\mathbf{x})$. We also assume that this random function has the property of intrinsic stationarity, such that the increment over a lag interval $\mathbf{h}$, $Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})$, has zero mean and a variance that depends only on $\mathbf{h}$ and not on $\mathbf{x}$. When the variability of the soil changes in space, the second part of this assumption is questionable.

Correspondence: R.M. Lark. E-mail: murray.lark@bbsrc.ac.uk
Received 5 April 2006; revised version accepted 10 August 2006

We should note that stationarity is a property of the random function $Z(\mathbf{x})$ and not of the data $z(\mathbf{x})$. Nonetheless, the plausibility of the assumption may be evaluated from data. Consider the case of a test of normality, such as we might apply to data from a random sample. We do not test the properties of the data: whether the frequency distribution of the data fits a normal function can be answered directly and almost always in the negative. Rather, we test the properties of an abstraction, which is the random variable from which we assume our data are drawn. With this proviso, we note that the plausibility of the assumption of stationarity might break down in different ways.

First, the variance of $Z(\mathbf{x})$ may appear to change with $\mathbf{x}$ (see Figure 1a for an idealized example). Voltz & Webster (1990) examined the variance of clay content within a moving window on a transect near Sandford in central England and found that it varied substantially. Lark (2005) examined CO_2 emissions from the soil on a transect across farmland, using the discrete wavelet transform (DWT), and found changes in their variability associated with topography, parent material and land use.

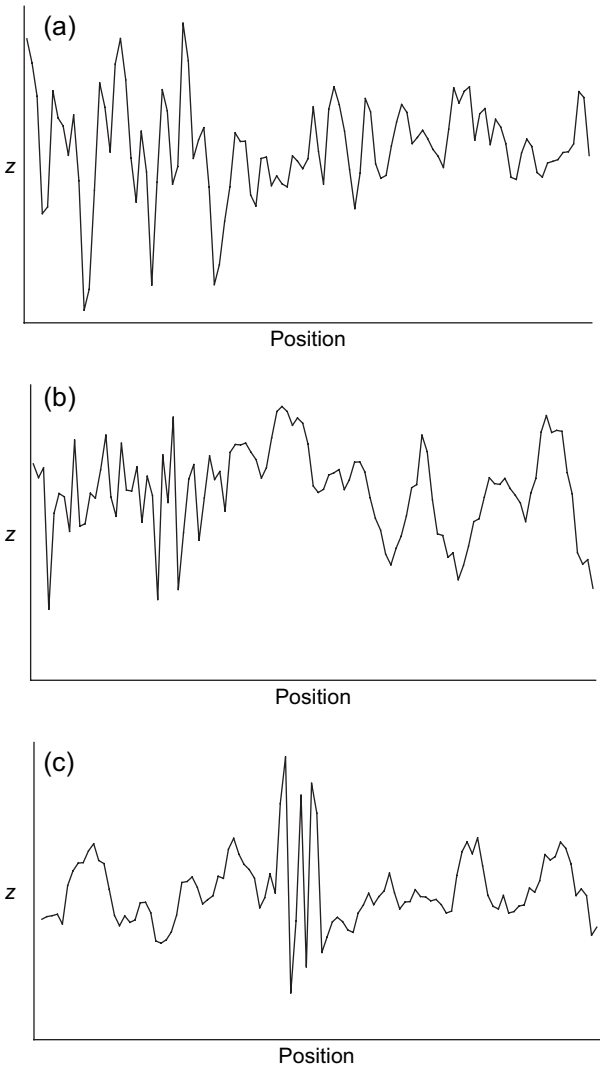

Figure 1 Three hypothetical transects on which the data appear inconsistent with an assumption of underlying stationarity: (a) there is a change in variance (but the autocorrelation remains unchanged); (b) there is a change in the autocorrelation function (but not in the variance); and (c) a realization of a stationary process is interrupted by a distinctive local pattern of variation.

Secondly, the distribution of variance between scales may change in space (see Figure 1b). This was found by Lark & Webster (1999) through wavelet analysis of the Sandford transect. Changes in the dominant spatial scale reflected stratigraphic variations on the transect. A change in the distribution of variance between scales is equivalent to a change in the autocorrelation function. These are the basic elements of non-stationarity in the variance. The soil might, of course, exhibit variation that cannot be regarded as a realization of a random process that is stationary in the variance because both the variance and the autocorrelation function appear to change.

Differences in the variance, the autocorrelation or both might be seen between broad regions (as illustrated in Figures 1a and b). By contrast, the soil might appear to vary in a manner that is inconsistent with the assumption of stationarity because of local intermittent changes in variance and/or autocorrelation (Figure 1c). We see marked variation at some scale in a few locations, but over most of our region there is little variation at this scale. Such variation will arise from particular local conditions such as rare wet patches, drainage channels, a clearing in woodland, locations where livestock congregate and anthropogenic features such as middens or pollution. The distinction between non-stationarity due to local intermittent effects and non-stationarity due to changes in variability between broad regions is a matter of scale, but it is important to remember that failures in the assumption of stationarity cannot always be handled by partitioning the study area into a few manageable regions. It is clear that intermittent variations will often be a feature of soil landscapes, given the many sources of soil variation (such as the examples given above), which have intermittent and local effects.

I want to understand better how soil variation departs from what we might expect to see under assumptions of stationarity in the variance. This requires quantitative methods to assess the evidence against this assumption. Because we assume that our soil data are generated from a random process and we know that there may be substantial fluctuations between different realizations of such a process, we require inferential methods that allow us to weigh the evidence against the assumptions of stationarity given the inherent randomness of our underlying model and the uncertainty due to sampling a process at discrete points. In previous work, the DWT has been used for this purpose (Lark & Webster, 2001; Lark, 2005). These methods test for changes in variance at particular scales, and they can be used to identify both changes in the overall variance of a process and in the distribution of this variance between scales.

A problem with this approach is that the spatial scales of the DWT analysis are somewhat arbitrary. A spatial scale in the DWT is an interval of spatial frequencies. A set of data on the soil, sampled regularly on a grid or transect of interval x_0 units, contains information on spatial frequencies in the interval $\left[0, \frac{1}{2x_0}\right]$. The spatial scales of the DWT are defined by a scale parameter, 2^j for $j = 1, 2, \ldots, j_{\max}$. The corresponding frequency intervals are

$$\left[\frac{1}{2^{(j+1)}x_0}, \frac{1}{2^j x_0}\right], \quad \text{for } j = 1, 2, \ldots, j_{\max},$$

and a final $(j_{\max} + 1)$ th frequency interval corresponds to the variation in the interval $\left[0, \frac{1}{2^{(j_{\max}+1)}x_0}\right]$. This is called the smooth component of the DWT and, because its basis function is a scaling function not a wavelet function, we do not compute a corresponding wavelet variance component (Lark & Webster, 1999).

Note that these frequency intervals depend only on the sampling interval x_0 and are otherwise predetermined. As noted by Lark (in press), this results in relatively poor resolution in the frequency domain for the highest spatial frequencies (the top half of the interval of spatial frequencies corresponds to the finest scale, with scale parameter 2^1). However, the spatial resolution of the corresponding features is good. This particular compromise between the resolution in the spatial domain and in the frequency domain might not be best suited to the analysis of a particular data set. If, for example, there were marked intermittent features of variation over short intervals, these might largely contribute to the variation in the finest scale, but this might be poorly resolved because of the breadth of the corresponding interval of spatial frequencies. For this reason, it makes sense to follow Gabbanini *et al.* (2004) and use the discrete wavelet packet transform (DWPT) for the analysis of variation by spatial scale. Gabbanini *et al.* (2004) conducted tests for uniformity of variance on coefficients from a set of wavelet packets that divided $\left[0, \frac{1}{2x_0}\right]$ into intervals of equal width. An alternative approach would be to select the basis of wavelet packets that is best suited to represent the data by some criterion (Coifman & Wickerhauser, 1992), not necessarily with uniform frequency intervals, and to analyse the corresponding coefficients.

In a previous paper (Lark, in press) I discussed the DWPT. A wavelet packet, like a wavelet in the DWT, is a basis function that can represent local variation over some interval of spatial frequencies. Given a particular basic wavelet function, one can define a large library of wavelet packets and then divide the frequency interval $\left[0, \frac{1}{2x_0}\right]$ into sets of discrete, non-overlapping wavelet packets in various different ways. One can then find a best basis for a particular data set, best in the sense that it represents a partition of the interval of spatial frequencies that is in some sense optimal for representing the data. A common criterion for identifying the best basis is an entropy measure that will find a basis for which many of the wavelet coefficients (which correspond to a fixed frequency interval, but are computed for different locations in space) are effectively zero, and most of the variation is concentrated in relatively few coefficients that correspond to local intermittent components of variation. This is clearly relevant to the identification of intermittent non-stationarity because, under stationarity, the coefficients for any packet are expected to be uniform, while intermittent features can be represented by a few wavelet coefficients, so a 'sparse' wavelet representation is possible. Given this discussion, it may seem sensible to identify the best wavelet packet basis for a data set and then apply tests for stationarity to the resulting wavelet coefficients.

A caveat is needed here. Algorithms to find the best basis are efficient at identifying any irregularities in data that can be described on a sparse wavelet basis. Inference about non-stationarity on the resulting packets is conditioned on their identification and may therefore be biased. For this reason, we ideally require some prior inference (in general terms) that the data cannot be regarded as a realization of a stationary random function. The methods for change detection can then be applied to the coefficients on the best basis to identify where these changes in variance occur. In this paper, I advance the hypothesis that the structure of a best basis provides insight into the variability of a set of data and that, by comparing the representation of our data on their best basis with their representation on the best basis under a null hypothesis of stationarity, we can make precisely the quantitative inference about the weight of evidence against this null hypothesis that is needed.

The remainder of this paper provides some more background on the DWPT and best bases. A test is then developed to evaluate evidence against a null hypothesis of stationarity in the variance from the structure of the best basis. This is then applied to simulated data and to real data on the soil.

Theory

The discrete wavelet packet transform (DWPT) and best basis

A detailed account of the DWPT is given by Lark (in press), and here I recapitulate only the essential principles. A wavelet transform decomposes a sequence of values into wavelet coefficients, each of which describes the variation of the data at some spatial scale and within a local, scale-dependent, neighbourhood. One obtains these coefficients, in effect, by processing the data with two filters, a wavelet function filter $\mathbf{h}$ and a scaling function filter $\mathbf{g}$. The former returns wavelet coefficients and the latter returns scaling coefficients, a smoothed representation of the data from which the components represented by the wavelet coefficients have been removed. The wavelet function has a scale parameter, $2^j x_0$, that is increased in the dyadic sequence $j = 1, 2, \ldots$. Increasing the scale parameter dilates the wavelet, so that the corresponding coefficients describe the local variation at a coarser scale of generalization. The wavelets are translated (along a transect or across a grid) in scale-dependent steps.

In practice, this dilation and translation is achieved in the pyramid algorithm. The data are convolved with a single wavelet filter and a single scaling function filter, and alternate elements in the output are subsampled. These filtering and subsampling operations are denoted by $\mathbf{h}\downarrow\frac{1}{2}$ and $\mathbf{g}\downarrow\frac{1}{2}$, respectively. When we apply operation $\mathbf{h}\downarrow\frac{1}{2}$ to the output from the $(j-1)$th operation $\mathbf{g}\downarrow\frac{1}{2}$, the result is equivalent to a filtering of the data with the jth dilation of the wavelet function translated in steps of $2^j x_0$.

Figure 2 shows how the DWPT operates. The top segment in this figure represents the N raw data. These contain variation over the range of spatial frequencies bounded by zero (the data average) and $\frac{1}{2x_0}$, the Nyquist frequency, which is the highest frequency that can be resolved for a given sampling interval. The wavelet filter operator $\mathbf{h}\downarrow\frac{1}{2}$ generates $N/2$ wavelet coefficients, and the scaling function filter generates the same number of scaling coefficients. These represent the variation in

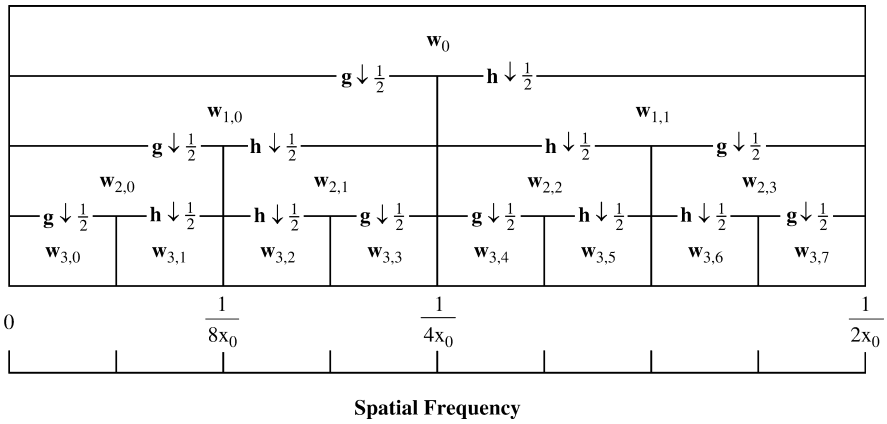

Figure 2 The discrete wavelet packet transform (DWPT) as an even partition of the spatial frequency range by successive passes of scaling function and wavelet function filters.

the two frequency intervals divided at $\frac{1}{4x_0}$. Note that these are nominal frequency bands and that, in practice, there is some leakage. Figure 2 shows how successive passes of these operators over the output from previous operators (in effect, dilation of the wavelet) generates an even subdivision of the interval $\left[0, \frac{1}{2x_0}\right]$ into frequency bands. As we use more dilations, the frequency resolution is improved. We denote by k the number of dilations considered in any case ($k = 3$ in Figure 2), which yields 2^k wavelet packets at the finest frequency resolution: $\mathbf{w}_{k,1}, \mathbf{w}_{k,1} \ldots, \mathbf{w}_{k,2^k-1}$. As was noted previously (Lark, in press), the cost is a reduction in the spatial resolution of the analysis. We may distinguish finer features in the frequency space, but our uncertainty about their location on our transect increases.

We might therefore obtain a better representation of a particular data set by selecting representations other than those obtained on all wavelet packets obtained by the full dilation, $\mathbf{w}_{3,0}, \mathbf{w}_{3,1}, \ldots, \mathbf{w}_{3,7}$ in Figure 2. The packets $\mathbf{w}_{3,0}$, $\mathbf{w}_{3,1}$, $\mathbf{w}_{2,1}$, $\mathbf{w}_{1,1}$ are one such possible basis (which corresponds to the standard DWT). Any non-overlapping set of packets that covers the full frequency interval $\left[0, \frac{1}{2x_0}\right]$ will constitute an orthogonal basis for our data (provided that the basic wavelet function has certain properties). That is to say, the coefficients correspond to independent additive components of the data. Thus, $\mathbf{w}_{2,0}$, $\mathbf{w}_{3,2}$, $\mathbf{w}_{3,3}$, $\mathbf{w}_{1,1}$ constitutes one orthogonal basis with relatively good spatial resolution at the highest and lower frequencies, but better frequency resolution, at the cost of poorer spatial resolution, in the intermediate frequency interval $\left[\frac{1}{8x_0}, \frac{1}{4x_0}\right]$.

Previously, I presented the procedure of Coifman & Wickerhauser (1992) for selecting the best basis on which to represent a particular set of data. I denote some basis by $\mathcal{C}$, a set of index doublets (j, n) that thereby denote a set of wavelet packets. I denote by $\mathcal{N}$ the full set of wavelet packets in the 'library', such as all those represented in Figure 2, and by $\mathcal{C} \tilde{\subset} \mathcal{N}$ I denote that the basis $\mathcal{C}$ consists of a non-overlapping subset of those in $\mathcal{N}$, which gives complete coverage of the frequency interval $\left[0, \frac{1}{2x_0}\right]$. Under the procedure of Coifman & Wickerhauser (1992), the best basis is $\mathcal{C}_B$ where

$$\sum_{(j,n)\in\mathcal{C}_B} M(\mathbf{w}_{j,n}) = \min_{\mathcal{C}} \left[\sum_{(j,n)\in\mathcal{C}} M(\mathbf{w}_{j,n}); \mathcal{C} \tilde{\subset} \mathcal{N} \right], \qquad (1)$$

where

$$M(\mathbf{w}_{j,n}) = \sum_{x=1}^{N_j} m\left(\left|w_{j,n,x}\right|\right), \qquad (2)$$

$m(\ )$ is a cost functional and $w_{j,n,x}$ is the xth coefficient for wavelet packet $\mathbf{w}_{j,n}$.

Coifman & Wickerhauser (1992) proposed an entropy cost functional:

$$\begin{aligned} m\left(\left|w_{j,n,x}\right|\right) &= -\left(\frac{w_{j,n,x}}{\|\mathbf{w}_0\|}\right)^2 \log\left(\frac{w_{j,n,x}}{\|\mathbf{w}_0\|}\right)^2, \ \forall\, w_{j,n,x} \neq 0, \\ &= 0, \ \text{otherwise}, \end{aligned} \qquad (3)$$

where $\|\mathbf{w}_0\|$ is the sum of squared data values. This cost functional is of interest to us because it will favour a basis on which much of the variation in the data can be concentrated in a few coefficients. That is to say, the favoured basis is one that can best resolve sparse features in the data and so return just a few large coefficients that correspond to these features, while many coefficients are effectively zero. For this reason, we would expect the best basis for a realization of a stationary process to differ from the best basis for a data set that contains intermittent features, and one way to test the evidence that the data are a realization of a stationary process (stationary in not containing intermittent variations) would be to compare the values of the cost functionals $\sum_{(j,n)\in\mathcal{C}_B} M(\mathbf{w}_{j,n})$ with $\sum_{(j,n)\in\mathcal{C}_B^S} M(\mathbf{w}_{j,n})$, where $\mathcal{C}_B^S$ is the set of doublets that constitute the best basis under stationarity. In the next section, I consider the question of what basis is best for a realization of a stationary random function.

What is the best basis for a realization of a stationary process?

The best basis for a realization of a random spatial process will depend on the criterion for judging the best basis, and the

statistical properties of the process. Given these, our best basis will have the best compromise between spatial resolution and frequency resolution over different frequency intervals. When the assumption of stationarity in the variance is compromised, because of intermittent sources of variation, we are concerned with detecting sparse features in the spatial variability of our variable of interest, with frequency intervals in the resulting basis for which many of the coefficients can be set to zero because the variation at this scale is localized and intermittent. From the discussion in the preceding section, this suggests that the entropy criterion for the best basis may be appropriate. I conjecture that, under intrinsic stationarity, with no intermittent sources of variation, there should be no possibility of reducing the entropy cost functional of a wavelet packet basis by resolving spatially sparse variation, and so the best basis from a packet with k dilations of the wavelet should be provided by the packets $\mathbf{w}_{k,0}, \mathbf{w}_{k,1}, \ldots, \mathbf{w}_{k,2^k-1}$. This is the basis that gives the best resolution in the frequency domain, and the poorest spatial resolution. There are other criteria for the definition of a best basis, i.e. other cost functionals, $m(\ )$, that can be used in Equation (2) and these might define a different best basis under stationarity in the variance. I do not consider these criteria in the present paper.

Consider a wavelet packet $\mathbf{w}_{j,n}$, for a variable z with coefficients $w_{j,n,t}$, $t = 1, 2, \ldots$. The variance of the packet coefficients is denoted $v_{j,n}$, and $v_{j,n} = \mathrm{E}[w^2_{j,n,t}|j, n]$ for any packet for which $\mathrm{E}[w_{j,n,t}|j, n] = 0$, which is the case for any packet $\mathbf{w}_{j,n}$, $n \neq 0$, and holds in general if $\mathrm{E}[z] = 0$. The wavelet packet variance is an additive component of the variance of the data (Walden & Contreras Cristan, 1998):

$$\sigma^2_{j,n} = \frac{1}{2^j}\mathrm{E}\left[w^2_{j,n,t}\middle|j, n\right] = \frac{v_{j,n}}{2^j}. \tag{4}$$

Following Krim & Pesquet (1995), we can compute the expected entropy cost functional for $\mathbf{w}_{j,n}$ assuming for the moment that $\|z\| = 1$. The expected value of the cost functional, given the variance of the packet coefficients (provided that the variable is normal and the process is stationary), is

$$\mathrm{E}\left[m(w_{j,n,t})\middle|j, n\right] = -v_{j,n}\left(\kappa + \ln v_{j,n}\right), \tag{5}$$

where $\kappa = 2 - \ln 2 - \gamma$ and γ is Euler's constant:

$$\gamma = \lim_{n\to\infty}\sum_{l=1}^{n}\frac{1}{l} - \ln n \approx 0.5772.$$

Consider two adjacent packets, $\mathbf{w}_{j,n}$ and $\mathbf{w}_{j,n+1}$. There is one corresponding packet at the $j-1$th dilation: $\mathbf{w}_{j-1,n'}$. The wavelet packet variances are also related:

$$\sigma^2_{j,n} + \sigma^2_{j,n+1} = \sigma^2_{j-1,n'}, \tag{6}$$

so, given that $v_{j,n} = \mathrm{E}[w^2_{j,n,t}|j, n]$ and from Equation (4),

$$\frac{v_{j,n}}{2^j} + \frac{v_{j,n+1}}{2^j} = \frac{v_{j-1,n'}}{2^{j-1}}. \tag{7}$$

Now the cost functional that we work with is

$$\sum_{t=0}^{N_j-1} m(w_{j,n,t}),$$

of which the expected value, given the assumptions behind Equation (5), is

$$-N_j v_{j,n}(\kappa + \ln v_{j,n}). \tag{8}$$

By the same token, the expected cost functional for packet $\mathbf{w}_{j-1,n'}$ is

$$-N_{j-1}v_{j-1,n'}(\kappa + \ln v_{j-1,n'}).$$

If there are N_0 data, then $N_j = \frac{N_0}{2^j}$, so Equation (8) can be written as

$$-\frac{N_0 v_{j,n}}{2^j}(\kappa + \ln v_{j,n}). \tag{9}$$

In the best basis calculation, the cost functional for the parent packet, $\mathbf{w}_{j-1,n'}$,

$$-\frac{N_0}{2^{j-1}}v_{j-1,n'}(\kappa + \ln v_{j-1,n'})$$

is compared with the sum of the functionals for the two daughter packets, which is

$$\begin{aligned} &-\frac{N_0 v_{j,n}}{2^j}(\kappa + \ln v_{j,n}) - \frac{N_0 v_{j,n+1}}{2^j}(\kappa + \ln v_{j,n+1}) \\ &\quad = -\tfrac{N_0}{2^j}\left\{v_{j,n}(\kappa + \ln v_{j,n}) + v_{j,n+1}(\kappa + \ln v_{j,n+1})\right\}. \end{aligned} \tag{10}$$

If the sum of the entropies for the daughter packets is less than or equal to that of the parent packet, the parent packet will not appear in the best basis (see Percival & Walden, 2000, pp. 225–226). Under my conjecture, the best basis from a packet library (obtained when we consider up to k dilations of the basic wavelet function) for some z that is a realization of a stationary process will consist of all the packets corresponding to the largest dilation, $\mathbf{w}_{k,0}, \mathbf{w}_{k,1}, \ldots, \mathbf{w}_{k,2^k-1}$ and so, for any j and n,

$$\begin{aligned} &-\frac{N_0}{2^j}\left\{v_{j,n}(\kappa + \ln v_{j,n}) + v_{j,n+1}(\kappa + \ln v_{j,n+1})\right\} \\ &\quad \leq -\frac{N_0}{2^{j-1}}v_{j-1,n'}(\kappa + \ln v_{j-1,n'}), \end{aligned}$$

so

$$\begin{aligned} &-\left\{v_{j,n}(\kappa + \ln v_{j,n}) + v_{j,n+1}(\kappa + \ln v_{j,n+1})\right\} \\ &\quad \leq -2v_{j-1,n'}(\kappa + \ln v_{j-1,n'}). \end{aligned} \tag{11}$$

This inequality can be shown to hold for all cases where the wavelet packet variances are non-zero (see the Appendix).

Weighing the evidence for non-sparse stationarity

I denote by $\mathcal{E}_{\mathcal{C}}$ the value of the entropy cost functional for some particular data set on the wavelet packet basis denoted by $\mathcal{C}$. We may compute, for any data set and specified packet library, the entropies $\mathcal{E}_{\mathcal{C}_B}$ and $\mathcal{E}_{\mathcal{C}_B^S}$. I propose the ratio $\mathcal{R} = \mathcal{E}_{\mathcal{C}_B^S}/\mathcal{E}_{\mathcal{C}_B}$

as a measure of the evidence for the assumption of non-sparse stationarity. The ratio has a minimum value of 1; larger values indicate that some local sparsity allows a more parsimonious representation of the data than is possible on the basis $\mathcal{C}_{\mathcal{B}}^{S}$. Because fluctuations in a realization of a stationary random function will lead to variation in the values of wavelet coefficients for any given packet, it is necessary to assess the variation in the ratio $\mathcal{R}$ that can be attributed to random variations in the realizations of a stationary process so that a null hypothesis of non-sparse stationarity can be assessed for real data. I propose the following procedure.

1 Standardize the data to an average of zero and norm $\|z\| = 1$.

2 Compute the ratio $\mathcal{R}$ for the selected packet library.

3 Estimate the variogram of the experimental data and fit a model.

4 Compute a sampling distribution of $\mathcal{R}$ under the null hypothesis by multiple iterations of the following steps.

a Generate an unconditioned realization of a stationary random function with the variogram obtained in step 3 above at the locations of the sample points. Because the realization is unconditioned, its variation will not resemble the original data, so large and small values may be found in different parts of the transect. However, because the realization is computed from the variogram obtained in step 2, its spatial variability will resemble the stationary content of the original data. Our simulated data are a realization of a stationary random function, without intermittent variability or other non-stationary features.

b Standardize the simulated data to an average of zero and norm $\|z\| = 1$.

c Compute the ratio $\mathcal{R}$ for these data given the specified packet library.

5 Use the empirical distribution function of $\mathcal{R}$ from step 4 as an approximation to the sampling distribution under the null hypothesis.

Percival & Walden (2000) observe that the best basis for a set of data is sensitive to shifts in the origin. That is, if we repeat an analysis after advancing the start point of the transform by one place, we may find that a different best basis is selected. To make the above procedure insensitive to shifts in the origin, I propose to follow Percival & Walden (2000) and compute the best basis for each of $2^k - 1$ analyses in which the start of the transform (on the symmetrically extended data) is advanced by one place on each occasion. The best basis criterion is computed from only the coefficients that correspond to the original data values, and that shift for which the entropy criterion for the best basis was the smallest over all shifts is identified. This basis is then selected and the entropy criterion, $\mathcal{E}_{\mathcal{C}_{\mathcal{B}}}$, is noted. Also noted is the value of the entropy criterion for the best basis assuming non-sparse stationarity, $\mathcal{E}_{\mathcal{C}_{\mathcal{B}}^{S}}$, (i.e. on all packets $\mathbf{w}_{k,0}, \mathbf{w}_{k,1}, \ldots, \mathbf{w}_{k,2^k-1}$) for the same shift that yields the best basis. This 'shift-cycling' procedure should be used at steps 2 and 4c in the procedure outlined above.

Case studies

Simulations

To demonstrate the approach presented above in circumstances when the non-stationary content of the data is known and controlled, I conducted a simulation study.

First, I computed an empirical distribution function (edf) of $\mathcal{R}$ for data at 256 locations on a regular transect. I undertook 10 000 iterations of step 4 in the process described in the previous section. I specified an exponential variogram for a standard normal random variable:

$$\gamma(h) = 1 - \exp(-h/30), \tag{12}$$

where h is the lag distance (units are transect intervals). The mean of the random variable was zero. The 'shift-cycling' procedure described in the previous section was followed, with $k = 3$ giving a maximum of eight wavelet packets $\mathbf{w}_{3,0}, \mathbf{w}_{3,1}, \ldots, \mathbf{w}_{3,7}$.

Next, I repeated the procedure; but, after each realization of the stationary random variable was generated, the simulated values at locations $i, i+1, \ldots, i+29$, for some randomly selected i where $1 \leq i \leq 227$, were replaced with independent random values from a normal distribution of mean zero and variance 1. The resulting data are therefore drawn from a process that shows intermittent local variation that contrasts with that over most of the transect' the variance is unchanged but the autocorrelation is different. The edf of $\mathcal{R}$ for this random variable was generated by 10 000 iterations of the simulation procedure. I then repeated this twice, with the variance of the intermittent spatially uncorrelated component of variation set to 2 and 5, respectively.

Figure 3 shows the resulting edfs of $\mathcal{R}$. Note that, in all cases, the edfs of the non-stationary random variables are to the right of the edf for the stationary process and that increasing the variance of the intermittent component of variation increases the expected value of $\mathcal{R}$. Note that 61, 72 and 80% of the values of $\mathcal{R}$ for the intermittent processes (variances of the intermittent components 1, 2 and 5, respectively) exceeded the 95th percentile of the edf of $\mathcal{R}$ for the realizations of the stationary random variable.

These simulations show that the proposed criterion $\mathcal{R}$ is indeed sensitive to the presence of intermittent non-stationary components of variation, when these correspond to a change in the autocorrelation only (when the variance of the intermittent component is 1) or to a change in both the autocorrelation and the variance.

Topsoil clay content from the Sandford transect

I applied this approach to the data on clay content of the topsoil at the first 256 locations on the Sandford transect of Webster & Cuanalo (1975), who describe the data in detail. Table 1 lists the geological units at each site. I used the first 256 sites on the

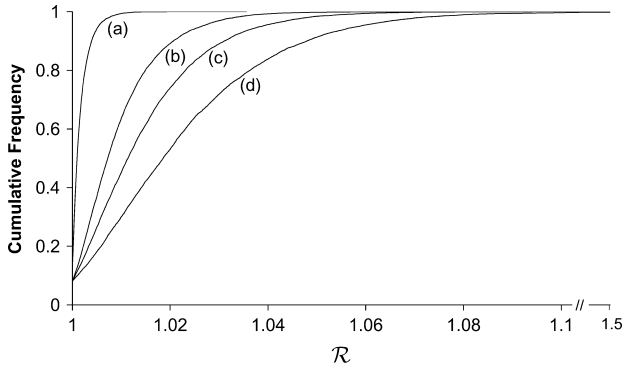

Figure 3 Empirical distribution functions of $\mathcal{R}$ for: (a) simulations of a spatially autocorrelated stationary random variable of variance 1; and simulations of the same stationary random variable interrupted by a spatially uncorrelated segment of variance (b) 1, (c) 2 and (d) 5.

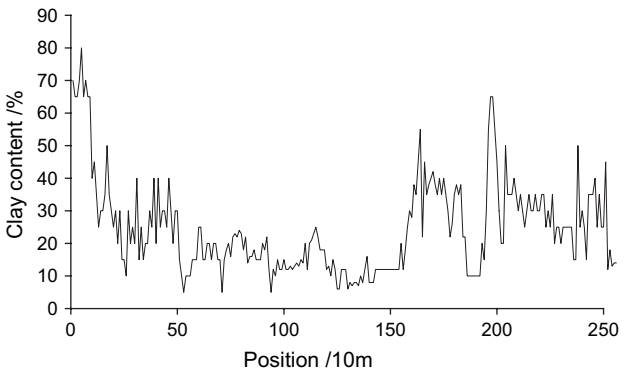

Figure 4 Topsoil clay content at first 256 locations on the Sandford transect.

transect for convenience (there is an integer power of two data points), but the method used below could be extended to any number of data.

Figure 4 shows the raw data, and Figure 5 shows their experimental variogram. An exponential model was fitted to the estimates:

$$\gamma(h) = 38.9 + 149\{1 - \exp(-h/167.5)\}, \qquad (13)$$

where h is the lag distance in metres.

Before further analysis, I then standardized the data to an average of zero and norm $\|z\| = 1$. Any wavelet analysis requires a strategy for padding the data to satisfy the wavelet filters where they overlap the ends. I followed the procedure that I reported previously (Lark, in press) and extended the data by symmetrical reflection at the ends of the transect. I used a wavelet packet library generated by up to $k = 3$ dilations of the Daubechies (1988) extremal phase wavelet with two vanishing moments. I selected this wavelet because it is orthogonal with a very compact support. I set $k = 3$, generating eight wavelet packets at the finest frequency resolution, on the judgement that more than eight packets for a data set this size would give little additional insight at the cost of poorer spatial resolution.

Figure 6 shows the decomposition of the data on all packets $\mathbf{w}_{3,0}, \mathbf{w}_{3,1}, \ldots, \mathbf{w}_{3,7}$ produced by three dilations of the wavelet (Figure 6a) and the decomposition on the best basis selected by the shift-cycling procedure (Figure 6b). Figure 7 illustrates the best basis. Note that this is, in fact, the DWT, $\mathbf{w}_{3,0}$, $\mathbf{w}_{3,1}$, $\mathbf{w}_{2,1}$, $\mathbf{w}_{1,1}$.

The entropy for the best basis over all shifts was $\mathcal{E}_{C_B} = 3.58$. The entropy $\mathcal{E}_{C_B^s}$ for the corresponding shift was 3.69, giving a ratio $\mathcal{R} = 1.03$.

Table 1 Geological segments on the first 256 positions of the Sanford transect

Segment	Stratigraphic class	Transect positions
1	Sharp's Hill Beds (Clay)	1–8
2	Great Oolite (Limestone)	9–50
3	Lower Estuarine Beds (Silt)	51–65
4	Chipping Norton Limestone (Limestone)	66–76
5	Lower Estuarine Beds (Silt)	77–86
6	Chipping Norton Limestone (Sand)	87–107
7	Chipping Norton Limestone (Limestone)	108–114
8	Chipping Norton Limestone (Sand)	115–158
9	Upper Lias (Clay)	159–182
10	Alluvium, Pleistocene and recent (Silt and Clay)	183–202
11	Upper Lias (Clay)	203–218
12	Chipping Norton Limestone (Sand)	219–234
13	Upper Lias (Clay)	235–256

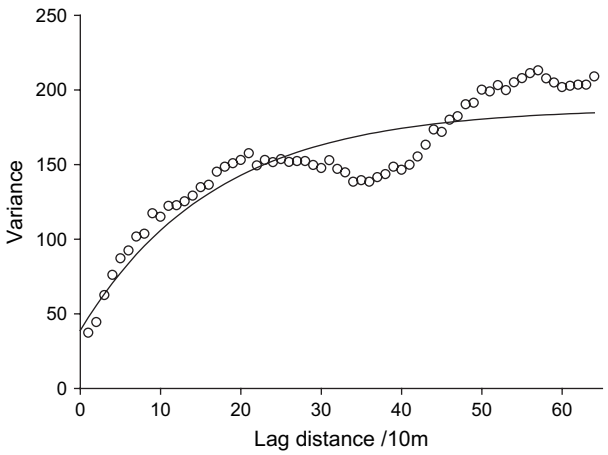

Figure 5 Experimental variogram (symbols) and fitted exponential model (line). The model is given in Equation (13).

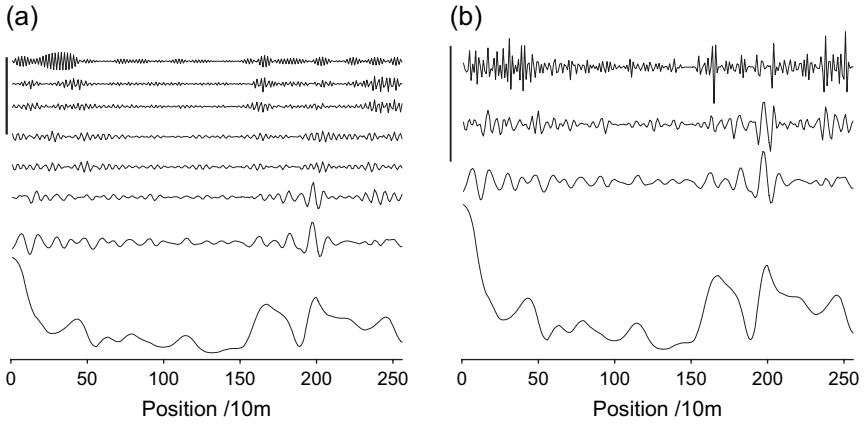

Figure 6 Multiresolution analysis of topsoil clay content showing components of variation on: (a) all basis packets $\mathbf{w}_{3,0}, \mathbf{w}_{3,1}, \ldots, \mathbf{w}_{3,7}$ from the bottom to the top; and (b) the best basis with increasingly fine scales from the bottom to the top. In both cases, the components are separated vertically for ease of plotting, so the absolute vertical position of any graph is arbitrary. However, for the graphs on each basis, the vertical scale for each component is the same. The vertical scale bar in each case corresponds to a variation in soil clay content of 50% of the mineral fraction by mass.

I then generated a sampling distribution for this ratio under the null hypothesis as described in the previous section. Unconditioned realizations of a random function with the variogram specified in Equation (17) were computed at step 4a. Figure 8 shows one such realization. I undertook 10 000 iterations of steps 4a–c and the edf of the resulting values of $\mathcal{R}$ was plotted. This is presented in Figure 9.

The approximate P value corresponding to our value of $\mathcal{R}$ can be read off the graph in Figure 9. This shows that we can reject the null hypothesis of an underlying non-sparse stationary random function with $P < 0.004$.

As a final stage in the analysis, I then tested the homogeneity of the wavelet variance for each packet with a non-zero frequency bound in the best basis (i.e. packets $\mathbf{w}_{3,1}$, $\mathbf{w}_{2,1}$ and $\mathbf{w}_{1,1}$). The procedure is described in detail elsewhere (Lark & Webster, 2001; Lark, 2005). In outline, the squared wavelet packet coefficients $w^2_{j,n,t}$, $t = 1, 2, \ldots$, are components of the variance for the data for the scale (frequency interval) associated with wavelet packet $\mathbf{w}_{j,n}$. Under the assumption of stationarity in the variance, the components $w^2_{j,n,t}$ have a homogeneous distribution over all t, and so the accumulated sum of squared coefficients,

$$S_k^{j,n} = \sum_{t=1}^{k} w^2_{j,n,t}, \tag{14}$$

should increase linearly with k. The procedure looks for breaks of slope in the graph of $S_k^{j,n}$ against k, and it tests these for significance against a sampling distribution obtained by Monte Carlo simulation.

Significant changes in variance were found for all three packets, thus dividing the transect into different regions of contrasting variance at each scale. These regions are shown in Figure 10.

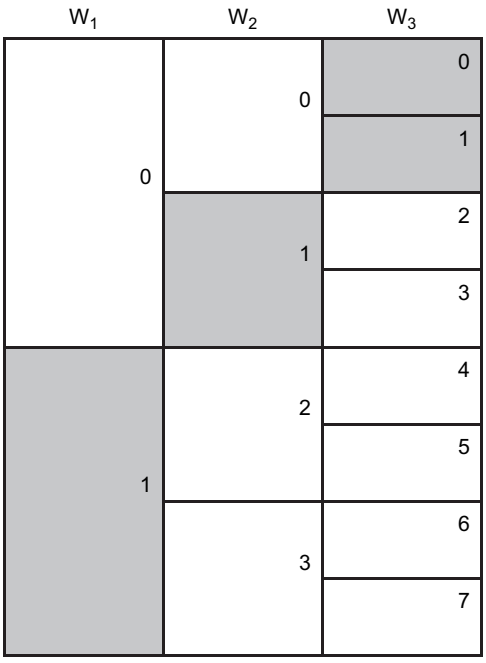

Figure 7 Wavelet packet table for all packets to $j = 3$. The dilation is given at the top of each column of the table and the packet number is given in the cell. The shaded cells constitute the best basis (entropy criterion) for the data on topsoil clay content.

Discussion

The best basis for these data is not what we should expect for a realization of a non-sparse stationary random function, and the ratio of the entropy of the wavelet packet coefficients

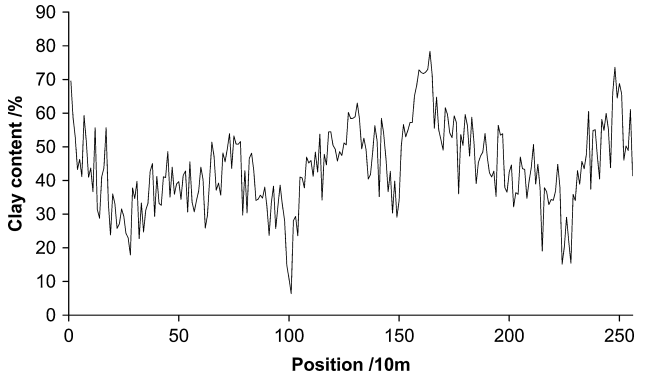

Figure 8 One realization, at the sample sites on the Sandford transect, of a stationary random variable with the same variogram as the data on topsoil clay content.

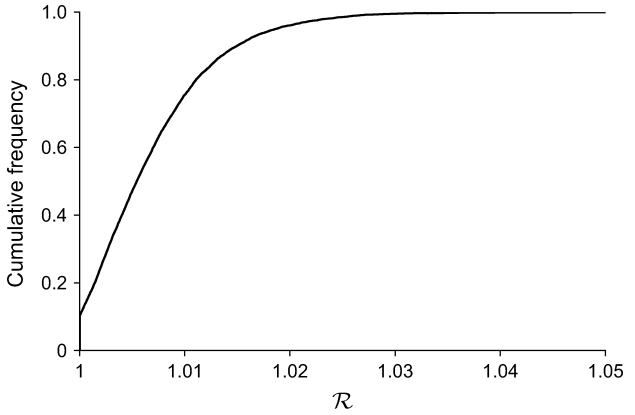

Figure 9 Empirical cumulative frequency distribution for $\mathcal{R} = \mathcal{E}_{C_B^S} / \mathcal{E}_{C_B}$ under a null hypothesis of a stationary random function.

expected for a sparse process to the entropy of the best basis for these data is large enough to reject the null hypothesis of an underlying non-sparse stationary random function with $P < 0.004$. The best basis achieves better spatial resolution of the higher frequency components of the data than a best basis under non-sparse stationarity. As we would expect, a subsequent test for the homogeneity of the variance of the underlying wavelet coefficients for each packet identifies changes, i.e. there are regions of the transect with elevated variance that this test can identify.

When we examine the multiresolution component of the data obtained for the wavelet packet corresponding to highest spatial frequencies ($\mathbf{w}_{1,1}$), we see that there are some large local fluctuations in the clay content within this frequency band, which give rise to large local variance components for this scale (Figure 10). Some (up to around position 50 on the transect) fall over the Great Oolitic limestone (where, on average, the clay content is less than elsewhere). This variation may represent the effects of the complex depositional history of the Oolite around Oxford as described by Arkell (1947). The others are all over the Upper Lias (where, on average, the clay content is large), so these local effects would seem to have some relation to local soil-forming factors.

In wavelet packets $\mathbf{w}_{3,1}$ and $\mathbf{w}_{2,1}$, there are some marked local fluctuations, particularly over the Oolite and the final outcrop of the Upper Lias, but the dominant feature is around position 200 where there is a very pronounced variation in the clay content of the soil and the largest local variance components are seen (Figure 10). From Table 1, we see that, in this region, the transect crossed a mixture of Pleistocene and recent alluvial deposits, which Webster & Cuanalo (1975) explain flanked a small stream. These alluvial deposits of variable texture constitute a local and singular source of variation in the clay content of the topsoil, which that is not found anywhere else on the transect (where the variation is determined by a more or less regular pattern of boundaries between Jurassic outcrops and which contribute largely to the variation captured by wavelet packet $\mathbf{w}_{3,0}$). Such a feature is a good example of a local source of sparse variation that is incompatible with assumptions of an underlying random function of uniform variance across the transect. It is encouraging that our inferential approach based on wavelet packet bases has identified this particular feature.

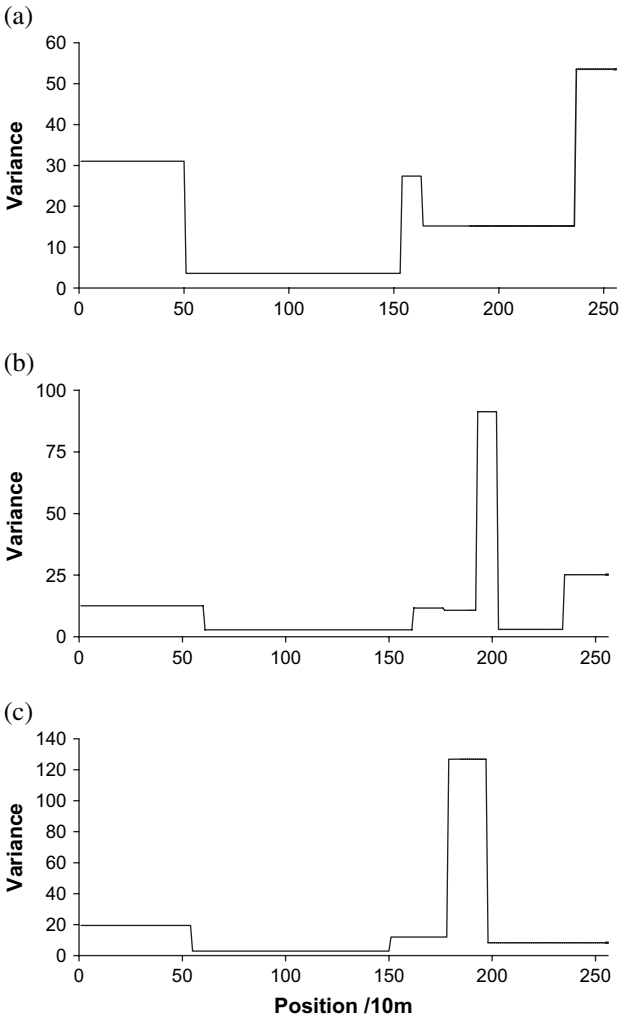

Figure 10 Partition of the Sandford transect by change points of the wavelet variance for three wavelet packets: (a) $\mathbf{w}_{1,1}$, (b) $\mathbf{w}_{2,1}$, and (c) $\mathbf{w}_{3,1}$.

The plausibility of an assumption of underlying stationarity is undermined when we identify such sparse features in the variation of soil properties. The practical implications for geostatistical analysis require attention. We might choose to use robust estimators of the variogram (Lark, 2000), so that our variogram estimates are not unduly influenced by local features of variation. When such features are strongly associated with singular phenomena in the landscape (such as alluvial deposits), we might remove these for a separate analysis.

So far statisticians have been interested in best basis algorithms to achieve efficient compression of data. This is the first study, of which I am aware, where it has been shown that the best

basis (entropy criterion) can be used for quantitative inference about the variability of a spatial process. In principle, there is no reason why the analysis presented in this paper should not be extended to two dimensions in a straightforward way. However, suitable data (of adequate size on a regular grid) on soil properties are rare, and the approach might be most usefully applied to the exploration of large data sets from sensors (e.g. satellite images of the land surface, or micrographs of the soil).

Acknowledgments

I am grateful to Dr H. E. Cuanalo de la Cerda who collected the soil data and to Professor R. Webster for making them available. I am also grateful to Dr B. Marchant for assistance with the derivative in Equation (A5) and to Professor R. Webster and Drs A. Milne, R. Corstanje and M. Pringle for comments on a first draft of this paper. Two anonymous referees made helpful observations. This research was funded by the Biotechnology and Biological Sciences Research Council through its core grant to Rothamsted Research.

References

Arkell, W.J. 1947. *The Geology of Oxford*. Oxford University Press, Oxford.

Coifman, R.R. & Wickerhauser, M.V. 1992. Entropy-based algorithms for best basis selection. *IEEE Transactions of Information Theory*, **38,** 713–718.

Daubechies, I. 1988. Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics*, **41,** 909–996.

Gabbanini, F., Vannucci, M., Bartoli, G. & Moro, A. 2004. Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi Dome. *Journal of Computational and Graphical Statistics*, **13,** 639–658.

Krim, H. & Pesquet, J.-C. 1995. On the statistics of best bases criteria. In: *Lecture Notes in Statistics, Number 103: Wavelets and Statistics* (eds A. Antoniadis and G. Oppenheim), pp. 193–207. Springer-Verlag, New York.

Lark, R.M. 2005. Analysis of Complex Soil Variation Using Wavelets. In: *Geographic Information Technologies for Environmental Soil Landscape Modeling* (ed. S. Grunwald), pp. 343–371. CRC Press, New York.

Lark, R.M. In press. The representation of complex soil variation on wavelet packet bases. *European Journal of Soil Science* doi: 10.1111/j.1365-2389.2005.00779.x

Lark, R.M. & Webster, R. 1999. Analysis and elucidation of soil variation using wavelets. *European Journal of Soil Science*, **50,** 185–206.

Lark, R.M. & Webster, R. 2001. Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet transform. *European Journal of Soil Science*, **52,** 547–562.

Percival, D.B. & Walden, A.T. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge.

Voltz, M. & Webster, R. 1990. A comparison of Kriging, cubic splines and classification for predicting soil properties from sample data. *Journal of Soil Science*, **41,** 473–490.

Walden, A.T. & Contreras Cristan, A. 1998. The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. *Proceedings of the Royal Society of London, A*, **454,** 2243–2266.

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.

Appendix: Proof of the inequality (11)

The inequality to be proven is

$$-\left\{v_{j,n}(\kappa+\ln v_{j,n})+v_{j,n+1}(\kappa+\ln v_{j,n+1})\right\} \leq -2v_{j-1,n'}(\kappa+\ln v_{j-1,n'}).$$

For simplicity, we write $\alpha \equiv v_{j,n}$ and $\beta \equiv v_{j,n+1}$. From Equation (7), we can write

$$\frac{v_{j-1,n'}}{2^{j-1}}=\frac{\alpha}{2^j}+\frac{\beta}{2^j},$$

so

$$v_{j-1,n'}=\frac{\alpha+\beta}{2}. \tag{A1}$$

Substituting into the inequality (11) gives

$$-\{\alpha(\kappa+\ln\alpha)+\beta(\kappa+\ln\beta)\} \leq -(\alpha+\beta)\left(\kappa+\ln\frac{\alpha+\beta}{2}\right)$$

and so

$$\{\alpha(\kappa+\ln\alpha)+\beta(\kappa+\ln\beta)\}-(\alpha+\beta)\left(\kappa+\ln\frac{\alpha+\beta}{2}\right)\geq 0. \tag{A2}$$

If this is expanded and tidied up, we obtain

$$\alpha\left(\ln\alpha-\ln\frac{\alpha+\beta}{2}\right)+\beta\left(\ln\beta-\ln\frac{\alpha+\beta}{2}\right)\geq 0, \tag{A3}$$

which on rearrangement yields

$$\left(\frac{2\alpha}{\alpha+\beta}\right)^{\alpha}\left(\frac{2\beta}{\alpha+\beta}\right)^{\beta}\geq 1. \tag{A4}$$

Note that, if all variances are positive, i.e. $\alpha>0$ and $\beta>0$, the left-hand side (LHS) of the inequality is equal to 1 for any $\alpha=\beta$.
Now

$$\frac{\partial}{\partial\beta}\left(\frac{2\alpha}{\alpha+\beta}\right)^{\alpha}\left(\frac{2\beta}{\alpha+\beta}\right)^{\beta}=\left[\ln\left(\frac{2\beta}{\alpha+\beta}\right)\right]\left(\frac{2\beta}{\alpha+\beta}\right)^{\beta}\left(\frac{2\alpha}{\alpha+\beta}\right)^{\alpha}. \tag{A5}$$

It can be seen that this is defined for any $\alpha>0$ and $\beta>0$ and is zero when $\alpha=\beta$. It can also be seen that, if $\alpha>0$, $\beta>0$ and $\alpha\neq\beta$, the sign of the derivative depends only on the first term in square brackets. This will be negative when $\beta<\alpha$ and positive when $\beta>\alpha$. The same argument applies if we consider the derivative with respect to α. From this, it is clear that the expression is at a minimum with respect to both α and β at any $\alpha=\beta$, where it takes the value 1, and so inequality (A4) holds and therefore inequality (11) holds too.