Changes in the variance of a soil property along a transect, a comparison of a non-stationary linear mixed model and a wavelet transform

CrossMark

R.M. Lark

British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, UK

ABSTRACT

The wavelet transform and the linear mixed model with spectral tempering are two methods which have been used to analyse soil data without assumptions of stationarity in the variance. In this paper both methods are compared on a single data set on soil pH where marked changes in parent material are expected to result in non-stationary variability. The two methods both identified the dominant feature of the data, a reduction in the variance of pH over Chalk parent material, and also identified less pronounced effects of other parent material contrasts. However, there were differences between the results which can be attributed to (i) the wavelet transform's analysis on discrete scales, for which local features are resolved with scale-dependent resolution; (ii) differences between the partition of variation into, respectively, smooth or detail components of the wavelet analysis and fixed or random effects of the linear mixed model; (iii) the fact that the identification of changes in the variance is done sequentially for the wavelet transform and simultaneously in the linear mixed model.

© 2015 British Geological Survey, NERC. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The properties of soil depend on many factors, which vary at a range of spatial scales. As a result soil properties may show substantial spatial variability (Beckett and Webster, 1971), which requires statistical treatment. One statistical model of soil variation which has been widely used is the linear mixed model (LMM) (Lark et al., 2006) which is a generalization of the geostatistical model of regionalized random variables (Webster and Oliver, 2007). In the LMM we treat the variation of a soil property, z, in terms of fixed effects (categorical or continuous covariates), which represent factors that we can understand and measure. The LMM represents the remaining variation with random effects. There are two sets of random effects in a LMM, those which are spatially correlated because they are caused by factors which operate at spatial scales which can be resolved by the sampling used to obtain our data, and an uncorrelated white-noise component (called the nugget variation in geostatistics). The LMM is written

$$\mathbf{z} = \mathbf{X}\boldsymbol{\tau} + \mathbf{u} + \mathbf{e}, \tag{1}$$

where z is an n-vector containing n observations of variable z, $\mathbf{X}$ is a $n \times P$ matrix with n values of each of P fixed effects, $\boldsymbol{\tau}$ is a vector of fixed effects coefficients, $\mathbf{u}$ is a vector of correlated random variables and $\mathbf{e}$ is a vector of independent and identically distributed (iid) random

variables, the nugget component. In the LMM we treat the random effects as multivariate normal (after appropriate transformation, if necessary) and of mean zero, which means they are characterized by their $n \times n$ covariance matrix. In the case of the iid random effect, e, the covariance matrix is given by $\sigma_e^2 \mathbf{I}_n$ where σ_e^2 is the variance and $\mathbf{I}_n$ is an $n \times n$ identity matrix. In the case of $\mathbf{u}$ the covariance matrix, $\mathbf{C}$, has a more complex structure reflecting the spatial dependence between observations.

Parameters of the LMM must be estimated from data. This provides something of a challenge for the random effects because, treating them as a realization of a multivariate gaussian process means that we have just one realization from which to estimate the covariance parameters. This is solved by stationarity assumptions. In the LMM a common assumption is second order stationarity, whereby the covariance of u at locations $\mathbf{s}_i$ and $\mathbf{s}_j$ is a function only of the separation in space between them: $\mathbf{s}_i - \mathbf{s}_j$. In this study we simplify the model by considering only the distance between the locations, $h = |\mathbf{s}_i - \mathbf{s}_j|$, but directional dependence can be modelled too. One can therefore express $\mathbf{C}$ as the product of a correlation matrix, the entries of which depend only on the distances between the corresponding observations, and a constant variance, σ_u^2. The entries in the correlation matrix can be modelled most generally with a Matérn correlation function (Diggle and Ribeiro, 2007).

$$\rho(h|\nu, \phi) = \frac{(h/\phi)^\nu K_\nu(h/\phi)}{2^{\nu-1}\Gamma(\nu)} \tag{2}$$

E-mail address: mlark@nerc.ac.uk.

where $K_\nu(\cdot)$ denotes the modified Bessel function of the second kind of order ν, $\Gamma(\cdot)$ is the gamma function, ϕ is a distance parameter, and ν is a parameter which determines the smoothness of the spatial process. If $\nu = 0.5$ then the Matèrn function is equivalent to the widely-used exponential correlation, larger ν give smoother variation and smaller ν rougher variation. The random effects of the LMM are therefore fully characterized by the set of parameters $\theta = \{\phi, \nu, \sigma_u^2, \sigma_e^2\}$. Estimates of these parameters are best obtained by residual maximum likelihood (REML) as described by Lark et al. (2006).

Under the model outlined above the covariance of the random effects at any two locations depends only on the distance in space between them. This is a necessary assumption to make the model estimable, but it has a cost for pedological plausibility. Consider a transect from the levées of a small river, across braided sediment deposits onto gentle slopes covered with solifluicted material and onto local hilltops with loess over the underlying sandstone. If we examine the clay content of the soil on this transect we will observe trends in the mean, which may be accounted for with appropriate fixed effects in the LMM. However, the LMM also requires the assumption that the variation about this mean is homogeneous across the transect. This seems implausible, given the different processes (alluvial deposition, solifluction, aeolian deposition) causing textural variation, and the different scales at which they operate. The implausibility of the stationarity assumption may undermine the prediction error variances calculated for interpolated values of the clay content (Lark, 2009), more generally the parameters of the random effects do not represent soil variability anywhere on the transect.

The stationarity assumption is one reason why the LMM, as commonly implemented, may often give limited insight into soil variation. This is one reason why soil scientists considered an alternative analysis to examine scale-dependent variation in soil. This is the wavelet transform. The discrete wavelet transform (DWT) is discussed in more detail elsewhere (e.g. Lark and Webster, 1999) and in the theory section below. In short the DWT represents data by a set of coefficients that represent local variability at different spatial scales (discrete intervals of spatial frequency). In the context of the example transect above, wavelet coefficients at some scale may differ in magnitude from one part of the landscape to another, representing changes in the variability of the property of interest.

Recently, attention has been directed to the extension of the LMM to cases where the random effects have a non-stationary covariance. Of particular interest here is the development by Haskard and Lark (2009) of the spectral tempering method proposed by Pintore and Holmes (2004, 2005). This method allows one to model changes in the variance and autocorrelation of a variable as a function of location in space or some other covariate. This is done by considering the empirical spectrum of the data of interest, and modifying it locally to adjust the distribution of variance between spatial frequencies, as well as the absolute variance.

The LMM with spectral tempering and the DWT are different but complementary ways to represent spatial variation without assuming homogeneity of the variability. However, the two analyses have yet to be compared on a common data set. Such a comparison would be of interest. First, variations in the tempering parameter of the LMM with spectral tempering and variations in the relative magnitude of wavelet coefficients for different spatial scales must both reflect the particular spatial heterogeneity of variation of some soil property, and a direct comparison should be instructive about the methods and the insight that they can give into soil variation. Second, if the spectral tempering random effects model in the LMM can, at least in some circumstances, provide information on changes in soil variation comparable with those from the DWT then this could be useful for the interpretative analysis of irregularly sampled data on the soil, where the scope for wavelet analysis is limited (Milne and Lark, 2009).

In this paper I report the analysis of measurements of soil pH on a transect using both the LMM with spectral tempering and the DWT. The data set is selected as an example where distinct pedogenetic domains, with contrasting parent material, give rise to heterogeneous spatial variability. The spectral tempering model and the DWT-based analysis are compared.

2. Theory

2.1. Linear mixed model with spectral tempering

In the introduction I reviewed the LMM as commonly applied to soil variables. The non-stationary form of this model with spectral tempering starts from a stationary covariance matrix, $\mathbf{C}$, for the spatially correlated random term in the model, the random variable U. One may compute the n eigenvectors of $\mathbf{C}$, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ and corresponding eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_n$, as in a principal components analysis. This provides the basis for what is called the spectral decomposition of the covariance matrix,

$$\begin{aligned} \mathbf{C} &= \sum_{k=1}^{n} \mathbf{v}_k \lambda_k \mathbf{v}_k^{\mathrm{T}} \\ &= \mathbf{VDV}^{\mathrm{T}}, \end{aligned} \tag{3}$$

where the superscript 'T' denotes the transpose of a matrix or vector, the matrix $\mathbf{V}$ is $n \times n$ with the eigenvectors of $\mathbf{C}$ in its columns, and $\mathbf{D}$ is a matrix with zeros on all but the elements of the main diagonal, which contains the corresponding eigenvalues, ordered from the largest to the smallest.

The early and later eigenvalues correspond to low spatial frequencies (long-range variation) and to high frequencies (short-range variation) respectively (Haskard and Lark, 2009). The eigenvalues $\lambda_k, k = 1, 2, \dots, n$, therefore constitute an empirical spectrum which describes how the variance of U is partitioned between the spatial frequencies. The empirical spectrum is not obtained directly from data but from a stationary covariance function, and is therefore itself stationary. I refer to this as the pre-tempering spectrum. Tempering is a method to adjust the spectrum locally; for example by a relative increase in the early eigenvalues (low spatial frequencies) in these areas where the variable appears to be smoother than it is elsewhere. Pintore and Holmes (2004) proposed that this is achieved by raising the terms of the pre-tempering spectrum to some positive power η. Where $\eta > 1$ the low-frequency terms in the spectrum are increased relative to the others, while setting a local value of $\eta < 1$ has the opposite effect, which enhances the short-range variation. Of course, if $\eta = 1$ the spectrum is unchanged. The spectrum can be adapted locally by allowing η to vary spatially. This is possible if we can express η as a function of location in space $\eta(\mathbf{s})$. The joint value of η for any two locations is obtained as

$$\eta(\mathbf{s}_i, \mathbf{s}_j) = 0.5\eta(\mathbf{s}_i) + 0.5\eta(\mathbf{s}_j). \tag{4}$$

A modified covariance matrix of U, which is in general non-stationary, $\mathbf{C}^{\mathrm{NS}}$, can then be obtained from the spectral decomposition of the pre-tempering covariance matrix. The (i,j)th element of this matrix is

$$\mathbf{C}_{ij}^{\mathrm{NS}} = \sum_{k=1}^{n} [\mathbf{v}_k]_i \lambda_k^{\eta_{ij}} [\mathbf{v}_k]_j \tag{5}$$

where $[\mathbf{v}_k]_i$ denotes the ith element of $\mathbf{v}_k$, which corresponds to the ith location $\mathbf{s}_i$ and the term $\eta_{ij} = \eta(\mathbf{s}_i, \mathbf{s}_j)$ is obtained from Eq. (4). Haskard and Lark (2009) showed that $\mathbf{C}^{\mathrm{NS}}$ is positive definite for positive values of η and positive definite $\mathbf{C}$. Given this, it is possible for some data set and a set of fixed effects, to compute the residual log-likelihood for some set of parameters that specify $\mathbf{C}^{\mathrm{NS}}$, and so, by an appropriate numerical optimization, to find a set of parameter estimates that maximize this (or, equivalently, that minimize the negative residual log-likelihood).

In the procedure proposed by Pintore and Holmes (2004) the pre-tempering covariance matrix is the estimated covariance matrix for the correlated term in a LMM for the data under an assumption of second-order stationarity, the covariance matrix for this term in the non-stationary model is obtained with Eq. (5) and a stationary nugget variance is also estimated. Haskard and Lark (2009) proposed a modification of this procedure to make the tempering parameters η independent of the units of measurement of the variable, and to allow the variance of U to be modelled as a non-stationary parameter independently of local variations in the autocorrelation. They proposed that the non-stationary covariance matrix, $\mathbf{C}^{NS}$, obtained by applying tempering parameters obtained with some function $\eta(\mathbf{s})$ to the pre-tempering covariance matrix be converted to a correlation matrix, $\mathbf{B}$. This correlation matrix is then rescaled to a non-stationary covariance matrix $\mathbf{C}_R^{NS}$. In this paper I did this with the following expression

$$\mathbf{C}_R^{NS} = \text{diag}(\,\boldsymbol{\sigma}\,)\mathbf{B}\,\text{diag}(\boldsymbol{\sigma}\,), \tag{6}$$

where $\text{diag}(\mathbf{a})$ for some n-vector $\mathbf{a}$ denotes an $n \times n$ matrix with the elements of $\mathbf{a}$ on the main diagonal and zeroes in all off-diagonal elements and $\boldsymbol{\sigma}$ is a vector, length n, the elements of which are standard deviations of U at each sample location, $\sigma_i, i = 1, 2, \dots n$. One may express σ_i as a function of location, $\sigma(\mathbf{s})$, analogous to $\eta(\mathbf{s})$.

Recalling the LMM for a soil variable, z, in Eq. (1), under the non-stationary model

$$u \sim \mathcal{N}\left\{\,0_n, \mathbf{C}_R^{NS}\right\},$$

and

$$e \sim \mathcal{N}\{\,0_n, \text{diag}(\boldsymbol{\gamma})\mathbf{I}_n \text{diag}(\boldsymbol{\gamma}\,)\},$$

with $\mathbf{C}_R^{NS}$ calculated as in Eqs. (5) and (6) and $\boldsymbol{\gamma}$ is a vector, length n, the elements of which are standard deviations of $\mathbf{e}$ at each sample location, $\gamma_i, i = 1, 2, \dots n$. One may express γ_i as a function of location, $\gamma(\mathbf{s})$, analogous to $\sigma(\mathbf{s})$ and $\eta(\mathbf{s})$.

We denote by ϑ the set of parameters which characterize the random effects of the non-stationary LMM, this comprises parameters of the functions $\eta(\mathbf{s})$, $\sigma(\mathbf{s})$ and $\gamma(\mathbf{s})$ in addition to the stationary correlation parameters that specify the pre-tempering covariance matrix, ν and η when a Matérn correlation is used. As noted above, Pintore and Holmes (2004) originally proposed that these parameters are fixed at the values from the best-fitting stationary model for the data, but Haskard and Lark (2010) found that a better fit (as measured by the maximized residual log-likelihood) might be obtained with a pre-tempering covariance matrix with different parameters and suggested that a profiling approach be used to find the best values of ν and η.

2.2. Maximal overlap discrete wavelet transform

The wavelet transform has been used by soil scientists to investigate complex spatial variation (e.g. Lark and Webster, 1999; Lark et al., 2004; Zeleke and Si, 2007; Milne et al., 2013; Biswas, 2014). For details of the wavelet transform the reader is referred to the text by Percival and Walden (2000) and to the paper by Lark and Webster (1999). In summary, the wavelet transform is an analysis of a variable on vector basis functions. One can think of the basis functions as basic mathematical building blocks which can be combined, with appropriate coefficients, to represent any real variable of finite variance. In the wavelet transform one finds the coefficients required to represent a variable of interest on some preselected set of wavelet basis functions.

A wavelet function can be represented by the mother wavelet $\psi(x)$. Many such functions have been defined, but they have certain common properties. A wavelet integrates to zero mean but has square norm one, indicating that it oscillates around zero. A wavelet is compact, the

oscillations of a wavelet damp rapidly to zero, and so a wavelet takes non-zero values only over a relatively narrow window, its support. This means that the coefficient for a particular wavelet in the analysis of some variable depends only on the values of that variable within the narrow support. This has two implications. First, a wavelet coefficient provides information on local rather than global variability. This is why it is of interest for the analysis of non-stationary variation. Second, to complete the analysis of a variable over some finite distance one will require coefficients for wavelets with compact supports in different locations. The support of a wavelet depends on M, its number of *vanishing moments* where

$$\int x^c \psi(x)\ dx = 0 \tag{7}$$

for $c = 0, 1, 2, \dots, M - 1$. Within a family of wavelet functions, increasing M makes the wavelet smoother and increases the support.

A wavelet analysis is based on wavelet bases which are obtained by *dilating* and *translating* the mother wavelet. The dilation of a wavelet function increases the width of its support, and also reduces the frequency of the components of the variable which are represented by its coefficients. A series of dilations of the wavelet therefore provide wavelet bases which will represent different spatial scales of variation. The *translation* of a wavelet changes the location of its support.

Here we are interested in the analysis of discretely sampled data. The discrete wavelet basis functions for some mother wavelet are given by

$$\psi_{m,n}(x) \quad = \quad \frac{1}{\sqrt{2^m}}\psi\left(\frac{x-n2^m}{2^m}\right), \tag{8}$$

which is the basis function for the mth dilation of the wavelet and its nth translation. Note that the dilations of the discrete wavelet transform are in the dyadic sequence $2^1, 2^2, 2^3, \dots$ The wavelet coefficients from all wavelets in the mth dilation are said to represent variation at scale $2^m x_0$ where x_0 is the basic sample interval. To avoid confusion with cartographic scale I refer to scales with large m as 'coarse' scales. A scale corresponds to a frequency interval (with some leakage between adjacent frequencies), increasing m reduces the frequency. Note also that the translation of the wavelet takes place in steps of 2^m, i.e. larger steps at coarser scales.

A wavelet coefficient is obtained as the inner product of a sequence of data values, $f(x)$ with the particular dilation and translation of the wavelet:

$$D_{m,n} \quad = \quad \left\langle f, \psi_{m,n}\right\rangle. \tag{9}$$

One practical difficulty that arises from this definition of the wavelet is the computation of coefficients at locations near the start and end of the sequence of values where components in the dilated wavelet function overlap the ends of the data. A standard procedure is to pad the ends of the data, the best way to do this for our purpose is by reflection at each end of the sequence (see Percival and Walden, 2000).

The wavelet transform can be inverted, reconstituting the original data. If one sets all wavelet coefficients to zero, except the set that correspond to some dilation, $m = j$, then compute the inverse transform then one obtains an additive component of the data with variations at frequencies in an interval corresponding to scale $2^j x_0$. This is called the 'detail' component of the data at this scale. If one computed all the detail components for $m = 1, 2, \dots, j$ and subtracted these from the data one would be left with what is called a 'smooth' component for the jth scale. The set of detail components for some set of dilations of the wavelet, and the residual smooth component are known as a *multiresolution analysis* (MRA) of the data (Mallat, 1989).

The Discrete Wavelet Transform (DWT) partitions the variance of a sequence of observations over all scales (see Daubechies, 1992). On

this basis one can define the variance for a particular scale, $2^j x_0$ by

$$\hat{\sigma}_j^2 = \frac{1}{2^j n_j} \sum_{n=1}^{n_j} D_{j,n}^2 \tag{10}$$

where n_j is the number of DWT coefficients at the jth scale. This is known as the sample wavelet variance (Percival, 1995).

Recall that the wavelet coefficients $D_{j,n}$ for the jth scale are obtained by translations of the dilated wavelet in steps of 2^j sample intervals. This is equivalent to filtering the signal with the dilated wavelet and subsampling the output at interval 2^j. A consequence of this is that all wavelets in the basis are orthogonal to all others. However, Percival and Guttorp (1994) showed that the wavelet variance could be more efficiently estimated by retaining all output from the filtering of the signal with the dilated wavelet. This is called the maximal overlap discrete wavelet transform (MODWT). The wavelet functions for the jth scale in the MODWT basis are translated in unit steps rather than steps of 2^j. The kth MODWT coefficient for the jth scale is denoted $d_{j,k}$. The MODWT wavelet variance is computed as

$$\breve{\sigma}_j^2 = \frac{1}{2^j \breve{n}_j} \sum_{k=1}^{\breve{n}_j} d_{j,k}^2 \tag{11}$$

where $\breve{n}_j$ is the number of MODWT coefficients at the jth scale. These coefficients are not orthogonal, that is to say they do not provide $\breve{n}_j$ independent pieces of information about the variance at this scale. In this study I followed Percival and Walden (2000) and used n_j as the equivalent degrees of freedom of $\breve{\sigma}_j^2$ to compute confidence intervals.

The wavelet variance is a global property of a variable, the particular value of wavelet transforms is that they are based on local coefficients. One can use the wavelet coefficients for some scale to examine the evidence that variance at this scale is not spatially homogeneous. This was developed by Whitcher et al. (2000), see also Percival and Walden (2000). The method has been applied to examine spatial variation of soils (Lark and Webster, 2001; Milne et al., 2013), and I outline it below.

Define the normalized cumulative sum of squared MODWT coefficients for scale j at the kth coefficient by

$$S_k = \frac{\sum_{m=1}^{k} d_{j,m}}{\sum_{m=1}^{\breve{n}_j} j_{j,m}}, \quad k = 1, 2, \dots, \breve{n}_j - 1. \tag{12}$$

Under the null hypothesis whereby the variance of our variable at the jth scale is spatially uniform we expect S_k to increase linearly with k. Consider an alternative where the variance changes at location k'. In this instance one expects the graph of S_k against k to be approximately a bilinear piecewise function with the breakpoint at k'. The null hypothesis is tested by computing a statistic B which finds a candidate change point at some location and measures how far the graph of S_k departs from the expected linear function under the null hypothesis. Define

B^+ and B^- as

$$B^+ \equiv \max_{1 \leq k \leq \breve{n}_j - 1} \left(\frac{k}{\breve{n}_j - 1} - S_k \right), \tag{13}$$

and

$$B^- \equiv \max_{1 \leq k \leq \breve{n}_j - 1} \left(S_k - \frac{k-1}{\breve{n}_j - 1} \right); \tag{14}$$

then

$$B \equiv \max \left[B^+, B^- \right]. \tag{15}$$

In this study I followed Percival and Walden (2000) by using Monte Carlo simulation to find the distribution of B under the null hypothesis for some j and $\breve{n}_j$ when $\breve{n}_j < 128$ and otherwise using a large-sample result from Inclán and Tiao (1994) to compute the P-value. In the approach of Whitcher et al. (2000) one tests the first candidate change point in a sequence of observations. If the null hypothesis is rejected then the change point is accepted, and one may then use the same procedure to test the homogeneity of variance of the two segments of the original sequence, divided at the accepted change point.

3. Materials and methods

3.1. Data collection

The data that are used in this study are on the pH of topsoil (depth 0–15 cm) on a transect of just over 7.5 km length in eastern England sampled at regular intervals of 29.45 m. The data are from a study reported by Haskard et al. (2010a) and by Milne et al. (2013). The study was primarily concerned with the emission of N_2O from the collected cores, but soil pH was determined, and in both cited papers the relationship between pH and aspects of the variability of N_2O-emission rate was examined, but the pH data themselves have not previously been analysed.

The first sample point was at 508,329, 237,450 on the UK Ordnance Survey grid (Warren Wood, near Silsoe, Bedfordshire), and the transect was aligned approximately North–South on a line of bearing 173.5 degrees from grid north. The soils of the landscape traversed by this transect are described by King (1969) and the principal soils are described in Table 1. The parent material is identified from the descriptions of King (1969) and with reference to the British Geological Survey mapping at 1:50 000 (British Geological Survey, 1992). There are five soil associations. The Cottenham Association overlies the Lower Greensand Group, a Cretaceous Sandstone. The soils of the Wicken Association are formed in drift over the Gault Clay. The drift deposits in this area show marked and complex variation, ranging from heavy Chalky Boulder Clay, with large pH, to lighter-textured acidic material (Sandy Loams) influenced by the Lower Greensand (King, 1969). The three associations in the southern section of the transect are all formed over Chalk from below the base of the scarp and

Table 1
Principal soils on the transect.

Sampling positions	Parent material	Soil association (King, 1969)	WRB groups (IUSS working group WRB, 2006)
1–70	Mixed drift over Lower Greensand Group	Cottenham	Arenosols, Cambisols
71–171	Mixed drift over Gault Formation	Wicken	Cambisols, Luvisols and Acrisols
172–235	Chalky colluvium over West Melbury Marly Chalk	Burwell	Gleyic and Rendzic Leptosols
236–254	Chalk (West Melbury Marly Chalk and harder lithological units[a])	Icknield	Rendzic Leptosols
255,256	Boulder Clay over White Chalk Subgroup	Oak	Gleyic Luvisols

[a] Totternhoe Stone Member, Zig Zag Chalk Formation and Melbourn Rock Member.

just onto the top of the dipslope. The Burwell Association is formed over the soft, argillaceous, West Melbury Marly Chalk Formation below the scarp. The Icknield Association is found at the base of the scarp over the West Melbury Marly Chalk and over somewhat harder units of the Grey Chalk Subgroup (Zig Zag Chalk Formation and Totternhoe Stone Member) and the White Chalk Subgroup (Melbourn Rock Member) which form the scarp itself. The last two sample locations are from soil formed in Boulder Clay over the White Chalk.

There was varied land use on the transect as shown in Table 2. Predominantly broadleaf woodland occurred over both Lower Greensand and Gault Clay. In addition to some grass and shrubby paddock, field margins and waste ground, and set-aside, most of the land on the transect was in arable use, either cultivated prior to drilling of spring-sown crops or with emergent winter-sown cereals or oilseed rape. These variations in land use could be expected to contribute to variation of soil pH. Soils over the Lower Greensand, and some over the Gault Clay, may be limed routinely if they are used for arable production, which may enhance contrasts in pH with soil under woodland or rough vegetation over the same parent material. Woodland was not found over the Chalk on the transect, otherwise there is no evidence of a trend in land use from north to south with a mixture of arable, grazed and uncultivated land over all parent materials. While land use might contribute to local variations of pH, there is no reason to expect it to contribute to any non-homogeneity of the variance of pH.

All sampling was completed in a single day (5th February 2008). Soil pH was determined from one core of length 150 mm and diameter 44 mm collected at each sample site. After air-drying soil pH was determined for a 10-g subsample of soil mixed with 25 ml of deionized water. The five distinct soil associations can be expected to differ with respect to soil pH. In general the Cottenham Association is expected to have the most acid soils because of the influence of the underlying Lower Greensand. Larger pHs may be found in soils of the Wicken Association, because of the presence of Chalky Boulder Clay, but pH is expected to be variable over this association because of the variability of the drift. The Burwell and Icknield associations are all formed over chalk, albeit groups with differing lithology, so the pH of the soil can be expected to be controlled by the presence of free calcium carbonate from the underlying solid rock, particularly on the scarp slope where the soils are

Table 2
Land use on the landscape-scale transect.

Sampling positions	Land use
1–21	Woodland
22–34	Winter-sown crop
35–36	Woodland
37–51	Cultivated for spring-sown crop
52–52	Uncultivated field margin
53–54	Woodland
55–58	Uncultivated waste ground
59–63	Sports field
64–70	Cultivated for spring-sown crop
71–97	Set-aside
98–108	Woodland
109–123	Winter-sown crop
124	Uncultivated field margin
125–130	Winter-sown crop
131–150	Uncultivated grass and scrub
151–152	Animal paddock
153–154	Uncultivated field margin
155–172	Winter-sown crop
173–178	Uncultivated field margin
179–181	Animal paddock
182–187	Uncultivated grass and scrub
188–238	Winter-sown crop
239–239	Uncultivated field margin
240–248	Winter-sown crop
249–256	Uncultivated field margin

relatively shallow. It is likely that the soils of the Oak Association have pH values in excess of 7 because the Boulder Clay typically contains some chalk fragments, but the presence of drift over the White Chalk Subgroup may reduce the influence of the solid geology on the pH of these soils.

3.2. Data analysis

Given the marked variations in parent material it can be expected that the variability of soil pH on this transect is complex and not homogeneous. The analysis by the DWT with the detection of change points in the variance at different scales is a natural way to represent this complexity. A corresponding approach based on the LMM with spectral tempering is to specify change points, the location of which are themselves variance parameters of the model, and separate values of the correlated and nugget variances and the tempering parameter, η, for the segments defined by these change points. I describe these analyses below.

3.2.1. Linear mixed model

It is expected that the variability of pH will change along the transect, but also the mean which should increase from the soils over cretaceous sandstone in the north to soils over chalk in the south. I chose to use cubic B-splines as fixed effects in the model to describe this trend. Cubic splines have been used before to represent soil variation (e.g. Voltz and Webster, 1990). The cubic B-spline basis functions were obtained with the bs procedure from the splines package for the R platform (R Core Team, 2014). The number of spline basis functions was chosen in an exploratory analysis by considering increasing numbers, in all cases with the interior knots at equal intervals) and examining the distribution of the residuals. The objective was to find the model with the smallest number of spline basis functions for which the assumption of normal residuals seemed plausible. A fixed effects model with a constant term and 10 spline basis functions was selected on this basis. Note that I used the default option in the bs procedure under which the basis does not include a constant, so the constant is an additional fixed effects parameter.

The LMM with spectral tempering and N change-points has a total of $4N + 5$ variance parameters. These are:

1. the two parameters, ν and ϕ, of the stationary autocorrelation function — Eq. (2) — used to define the pre-tempering spectrum;
2. the $N + 1$ values of the tempering parameter, one value for each segment of the transect;
3. the $N + 1$ values of the nugget variance, c_0, one for each segment;
4. the $N + 1$ values of the correlated variance, c_1, one for each segment;
5. the locations of the N change-points themselves.

As noted above, Haskard and Lark (2010) recommended that the first two parameters, ν and ϕ, are not fixed at values estimated in a stationary LMM for the data. They suggested that a profiling method is used. This is computationally demanding, so in this study I used the following procedure. First, LMMs with stationary variance parameters were estimated. I estimated five such models with the parameter ν fixed, respectively, at values 0.1, 0.5, 1.0, 1.5 and 2.0, following the recommendation of Diggle and Ribeiro (2007), and the other parameters allowed to vary. The model with the smallest negative residual log-likelihood was selected. I then fixed ν at this value and computed the negative log-residual likelihood profile for the LMM with spectral tempering, fixing the ϕ parameter at each of a set of values in turn, including the estimated value for the stationary model, and finding the maximum residual likelihood estimates of the other parameters subject to that constraint. The value of the ϕ parameter was selected for which the negative log-residual likelihood was smallest. In addition, I computed the values of Akaike's information criterion (AIC) for each likelihood on the profile (Akaike, 1973). If the negative residual

log-likelihood is ℓ for some model with P parameters then AIC takes the value A.

$$A = 2P + 2\ell, \qquad (16)$$

Note that the positive sign appears in this expression because ℓ is the negative residual log-likelihood. The AIC is a measure of the goodness of fit of a model which penalizes improvements in likelihood which increase the number of parameters. By selecting from a set of models the one with smallest AIC one minimizes the expected information loss through the selection decision (Verbeke and Molenberghs, 2000).

This procedure was repeated to fit LMM models with spectral tempering with the number of change points, $N = 1, 2$, until increasing the number of change points increased the value of AIC for the selected value of ϕ from the profile. I then fixed the value of ϕ at this value and computed residual log-likelihood profiles for the discrete values of ν examined for the stationary model, and with different numbers of change points to see whether an improved fit could be obtained with a value of ν different to the one selected for the stationary model.

Values of ν and ϕ selected were those for which the value of AIC was smallest in the two sets of profile residual log-likelihoods described above. In addition to this non-stationary LMM for the whole data set I fitted two additional models with the same number of change points. First, the maximal model (separate variances and tempering parameter values for each segment) was fitted to a subset of 150 observations formed by simple random sampling without replacement. This was to examine the effect of using a sparser and irregularly sampled data set. Second, I considered the possibility that the variability of pH is stationary in the autocorrelation but not in the variance by fitting a LMM in which only the variances of the correlated and nugget components differ between the segments. In this case I used the pre-tempering spectrum selected for the non-stationary model and estimated a constant value of η to produce the stationary autocorrelation matrix, this allowed me to use the same code as used for the fully non-stationary models.

3.2.2. Wavelet transform

The maximal overlap discrete wavelet transform, as described in the theory section, was implemented using the waveslim package for the R platform (R Core Team, 2014; Whitcher, 2015). I wrote separate code in R, using the modwt function with the option boundary = "reflection", to implement the detection and testing of candidate change points, using either Monte Carlo simulation with 5000 realizations or the method of Inclán and Tiao (1994) as described above. In the Monte Carlo case one evaluates an empirical P-value for acceptance or rejection of the null hypothesis from the number, M' out of M simulations for which the B-statistic exceeded the observed value for the actual data. I followed Percival and Walden (2000) by rejecting the null hypothesis with confidence level α if

$$\left(\frac{M'}{M} - \alpha\right)^2 > \frac{4M'(M - M')}{M^3}. \qquad (17)$$

This is more robust than considering only the empirical P-value, M'/M, by accounting for the sampling error.

A reviewer drew attention to the fact that changes in variance near the ends of the transect may be detected with less power than those near the middle. This has received no attention in the literature on wavelet transforms, but is clear in the original paper on the cumulative sum of squares method by Brown et al. (1975) where it can be seen that the value of B corresponding to an increase in variance may be necessarily restricted below the threshold for some P-value in the first half of a signal, and, similarly, the value corresponding to a reduction in variance may be restricted in the second half of the signal. To avoid this problem any candidate change point for which the null hypothesis of uniform variance was accepted in these conditions (possible variance increase in the first half of the transect or segment of the transect, or variance reduction in the second half) was re-tested after reversing the order of
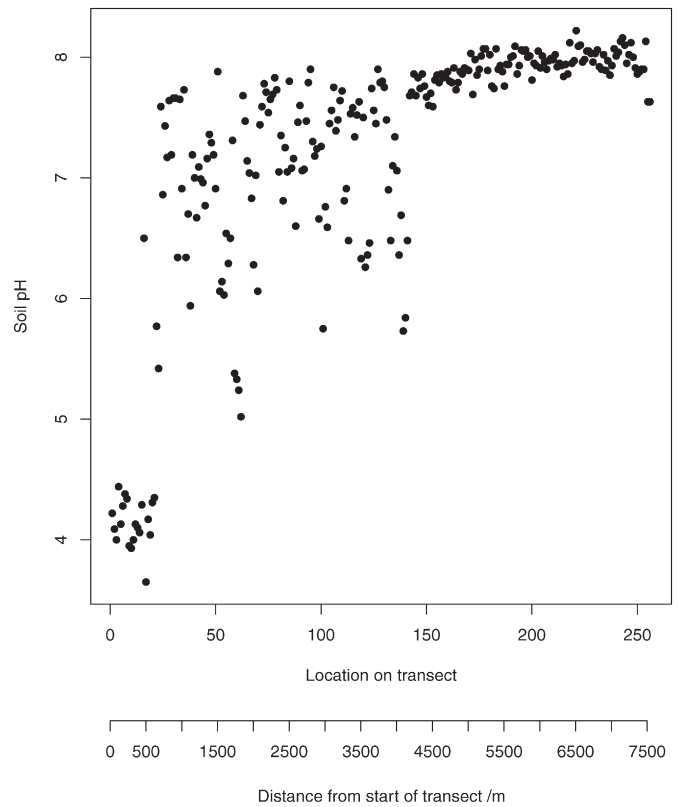

Fig. 1. Soil pH measured at locations along the transect.

data values. In no case was there any substantial change in P-values or any change in inference about candidate change points.

In these analyses I used Daubechies's extremal phase wavelet with 2 vanishing moments (Daubechies, 1988). This wavelet is specified by the option wf = "d4" in the waveslim package, the number in d4 refers to the number of filter coefficients which is twice the number of vanishing moments. This was preferred over smoother wavelets with more vanishing moments, since we do not necessarily expect smooth variation in soil pH, and limiting the number of vanishing moments makes the basis of the wavelet more compact and so improves the spatial resolution of changes in the variance. However, it was also preferred over the Haar wavelet, with just one vanishing moment, since this has poorer frequency resolution, (Percival and Walden, 2000). The MODWT, using reflection to deal with end effects, was specified to obtain multiresolution analyses using the mra function in waveslim with the options method = "modwt", boundary = "reflection".

4. Results

Fig. 1 shows soil pH plotted against position on the transect, with distance from start of the transect in metres also shown. In Table 3 are shown statistics of residuals from the exploratory fit of models with increasing number of B-spline basis functions. The skewness decreases

Table 3
Summary statistics for residuals of exploratory spline models.

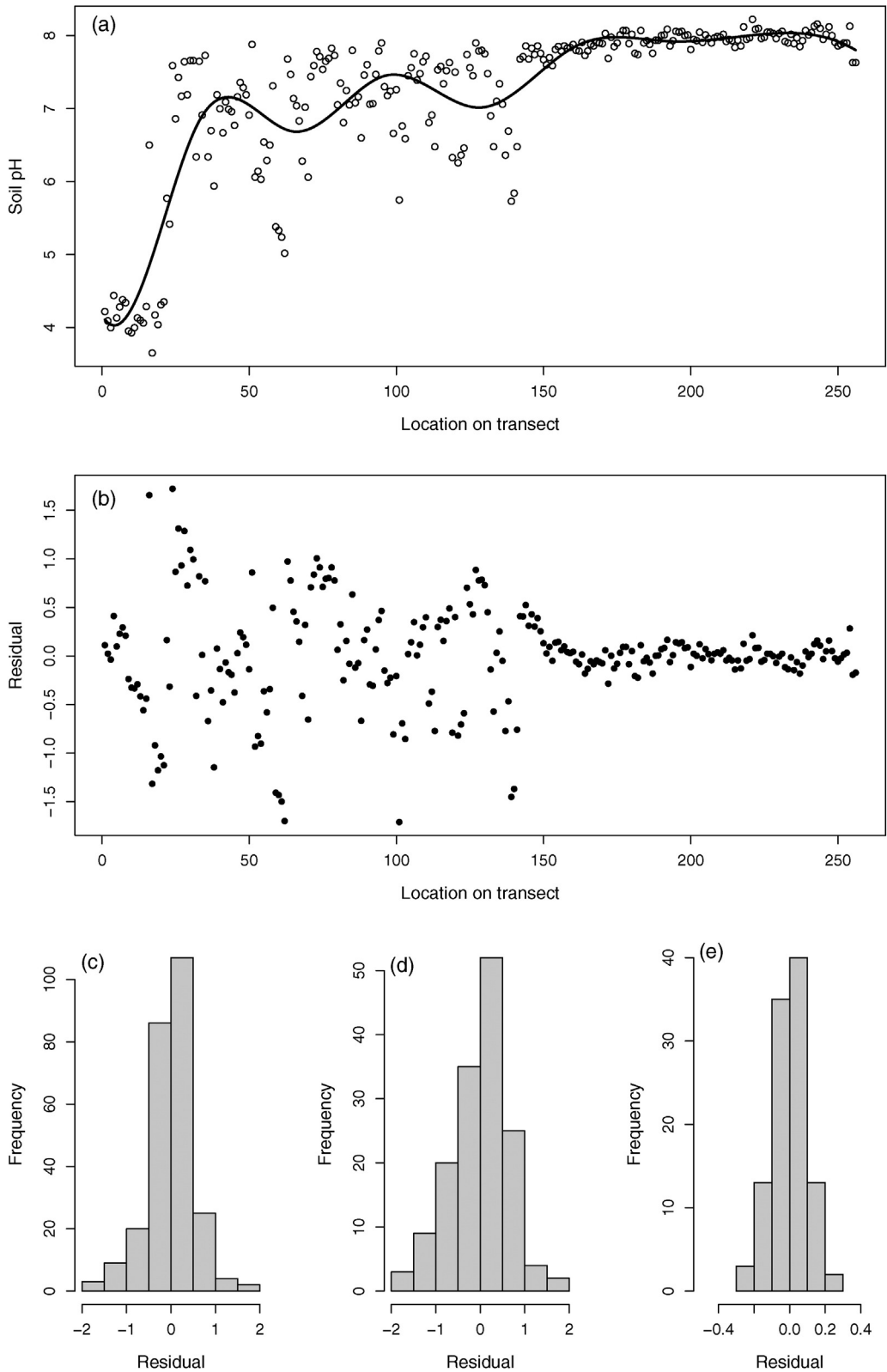| Number of spline basis functions	Median	Standard deviation	Skewness	Kurtosis
5	0.05	0.57	−0.75	2.13
6	0.05	0.57	−0.77	1.98
7	0.05	0.57	−0.80	1.92
8	0.03	0.57	−0.74	2.08
9	0.04	0.56	−0.62	2.00
10	0.02	0.52	−0.31	1.73

Fig. 2. (a) Fitted spline (10 basis functions) (b) residuals from the fitted model plotted against location (c) histogram of residuals (d) histogram of residuals for the first 150 locations and (e) histogram of residuals from location 151 to 256.

markedly when 10 basis functions are used, Fig. 2 shows this fitted model, the residuals and their histogram. As the plot of residuals shows a marked change in their variability near location 150, separate histograms are also shown for the first 150 residuals on the transect and the remaining ones. An assumption of normality of the residuals

looks plausible so this spline basis was selected as the fixed effect for all linear mixed models. Table 4 shows results for a LMM with these fixed effects and stationary random effects with the ν autocorrelation parameter fixed at different values. The value of ν was fixed at 1.5 for the pre-tempering model for the initial set of non-stationary LMM as

Table 4
Estimated stationary variance parameters for models with different fixed values of ν.

ν (fixed)	c_0	c_1	ϕ/m	ℓ
0.1	0.016	0.214	249.9	−53.97
0.5	0.100	0.156	94.9	−60.97
1.0	0.115	0.183	68.2	−63.59
1.5	0.122	0.200	54.2	−64.29
2.0	0.123	0.171	40.4	−64.10

this value had the smallest negative log residual likelihood (nlrl). This autocorrelated process is smoother than one with an exponential correlation ($\nu = 0.5$) but less smooth than a gaussian correlation ($\nu = \infty$).

Fig. 3 (left column) shows the profile nlrl (top) and AIC (bottom) for LMM fitted by spectral tempering with one to five change points, and with ν fixed at 1.5 and the ϕ parameter fixed at different values for the pre-tempered spectrum. When there are three or more change points there is a clear minimum in the profile with ϕ for the pre-tempering spectrum equal to 1472 m. Adding change points reduces the minimum nlrl as expected, but the AIC was smallest for a model with $N = 4$ so this was selected. The right column in Fig. 3 shows corresponding nlrl and AIC values for models with different numbers of change points and the pre-tempered spectrum based on parameters $\phi = 1472$ m and ν fixed at values from 0.5 to 2.0. This shows that the initial selection of $\nu = 1.5$ was robust. With $N = 4$ the minimum nlrl/AIC
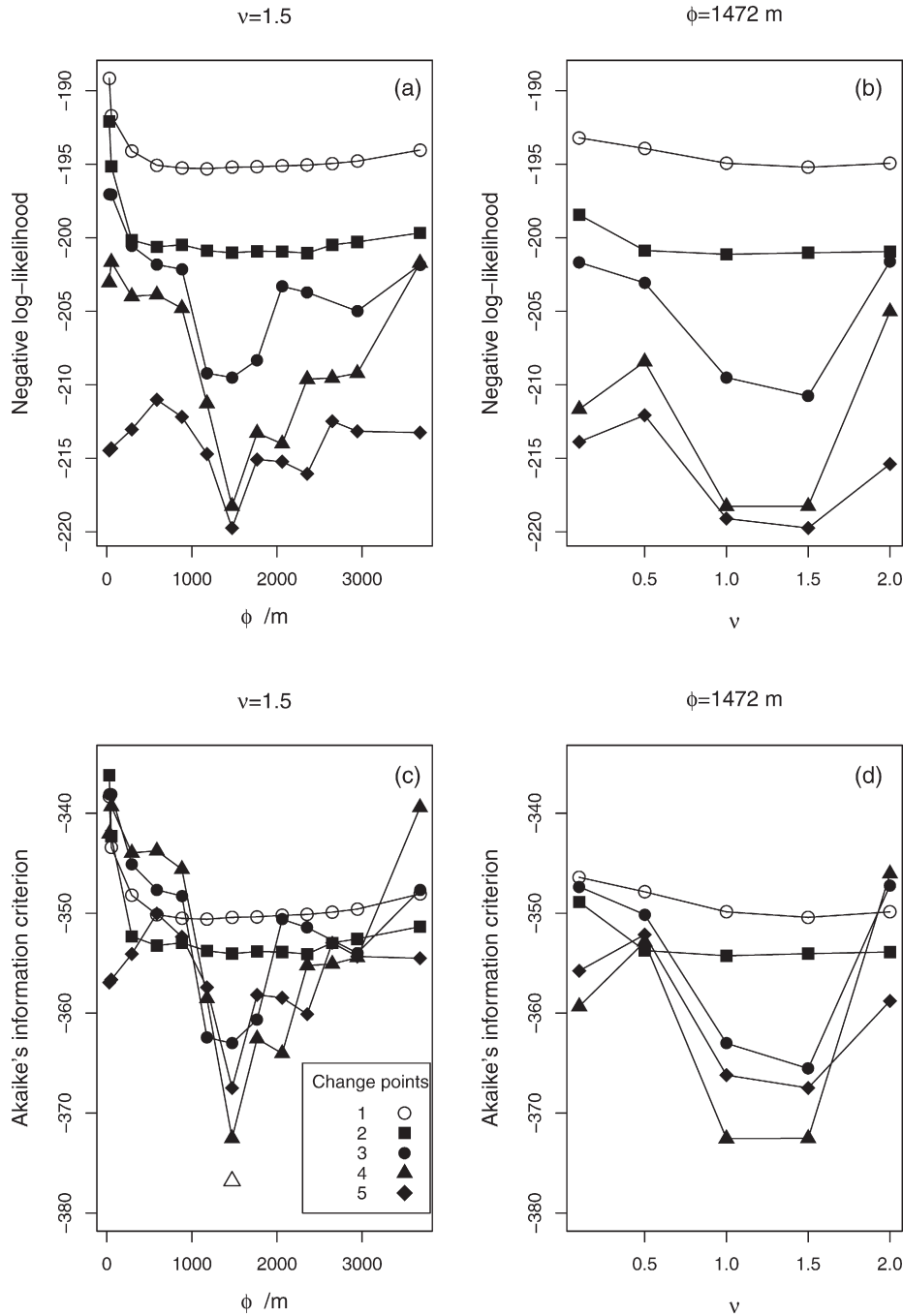

Fig. 3. Results for fitting non-stationary linear mixed models with 1–5 change points. The symbol corresponding to each number of change points is shown in the legend to Fig. 3c. The open triangle corresponds to a model with five change points but a stationary autocorrelation. Top left (a) Negative residual log-likelihood (ℓ) for different values of ϕ and $\nu = 1.5$. Top right (b) Negative residual log-likelihood (ℓ) for different values of ν and $\phi = 1472$ m. Bottom left (c) Values of AIC corresponding to points in (a). Bottom right (d) Values of AIC corresponding to points in (b).

Table 5

Parameters of non-stationary linear mixed model identified from profile on parameter ϕ with $\nu = 1.5$. The first line of the table shows, for comparison, the parameters of the selected stationary model (from Table 4) with $N = 0$. Note that the terms in column 'Start' are the $N+1$ first locations (transect positions from 1–256) of the segments formed by N change points. The model denoted 4* has 4 change points but a common value of the η parameter over all segments (i.e. stationary autocorrelation with non-stationary variances). The model denoted 4** has four change points and non-stationary autocorrelation and variances, and was fitted to a random subset of 150 observations. Note that A, the value of AIC, is not given for model 4** because it is comparable only between models fitted to the same data.

N	P	ν	ϕ	ℓ	A	Start	η	c_1	c_0
0	15	1.5	54.2	−64.3	−98.6	–	–	0.200	0.122
1	20	1.5	1177.6	−195.3	−350.6	1	0.372	0.826	0.170
						142	0.365	0.008	0.007
2	24	1.5	1472.0	−201.0	−354.0	1	0.347	1.037	0.244
						71	0.350	0.749	0.072
						142	0.333	0.006	0.008
3	28	1.5	1472.0	−210.8	−365.6	1	0.646	2.343	0.019
						16	0.242	1.163	0.129
						65	0.274	0.463	0.057
						142	0.956	1.828	0.008
4	32	1.5	1472.0	−218.3	−372.6	1	0.390	0.275	0.008
						16	0.369	1.603	0.327
						71	0.374	0.851	0.091
						142	0.354	0.008	0.006
						254	0.255	0.015	0.042
5	36	1.5	1472.0	−219.7	−367.4	1	0.659	1.406	0.032
						12	0.025	0.755	0.063
						64	0.372	1.700	0.173
						108	0.434	1.594	0.077
						143	1.518	0.430	0.029
						250	0.528	0.005	0.027
4*	28	1.5	1472.0	−216.4	−376.8	1	0.344	0.411	0.001
						15	_"_	1.536	0.279
						71	_"_	0.755	0.065
						142	_"_	0.009	0.006
						254	_"_	0.009	0.049
4**	32	1.5	1472.0	−335.2		1	0.596	0.719	0.302
						7	0.370	0.598	0.527
						62	0.078	1.635	0.527
						142	0.732	0.386	0.007
						169	1.233	0.244	0.007

Fig. 4. Marginal residuals of the non-stationary LMM showing segments determined by change points from (numbered) the non-stationary LMM and (letters) wavelet change-point detection. Note that the region in grey shows the range of change points identified from the wavelet coefficients near to 4270 m (location 145 on the transect).

was obtained with ν equal to 1.0 or 1.5, with three or for change points the minimum was at $\nu = 1.5$. Table 5 shows estimated parameters for LMM with $N = 1, 2, 3, 4$, and the values of the change points identified by REML. The model indicated by $N = 4^*$ in Table 5 has four change points but η is fixed over all segments (i.e. the autocorrelation is stationary) and only the variances differ. Note that the AIC is smallest for this than for any other model for the full data set, it is also indicated by an open triangle on the bottom left plot in Fig. 3. This indicated that, while a model with non-stationary variance is needed for these data, only the variance components and not the autocorrelation need be treated as non-stationary.

The change points for the selected model were at the 15th, 71st, 142nd and 254th location (first location in the second to fifth segments respectively). These are illustrated in Fig. 4 on a plot of the marginal residuals from the fixed effect model, where the segments are labelled numerically 1–5. The second change point corresponds exactly to the geological boundary between the Lower Greensand Group and the Gault Clay (see Table 1). Both the nugget variance, c_0, and the correlated variance, c_1, become smaller at this boundary, possibly reflecting how the Chalky Boulder Clay may constrain the scope for variation in pH where it occurs over the Gault Clay, and the potential for very small pH values over the Lower Greensand. Note that this change point also appears in the LMMs with varying values of η and with two and four change points; in the model with 3 change points one appears at the 65th location, less than 200 m from the mapped position of the boundary.

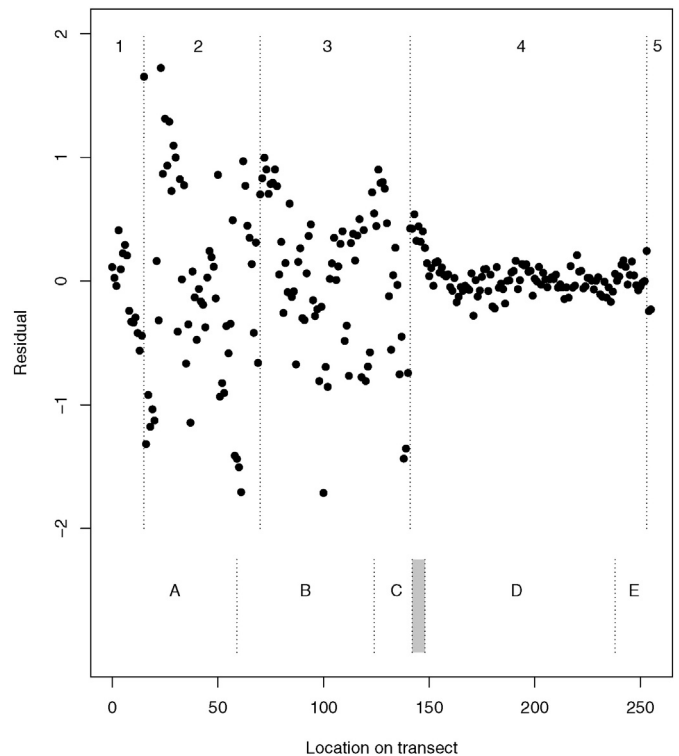The third change point in the selected model was somewhat north of the mapped boundary between soils formed over the Gault Clay and those formed over Chalk units. However, this change point is very apparent on plots of the residuals (see Fig. 4). The discrepancy with the mapped boundary could reflect boundary uncertainty. It could also reflect superficial influences of the Chalk on soils formed over the Gault Clay, either by transport of Chalky material in the Boulder Clay from where the Chalk is in outcrop just to the south, or by recent processes of colluviation. At this point the variation in pH becomes markedly smaller than it is to the north, presumably because of the constraint that free calcium carbonate imposes on possible soil pH values (see discussion). Note that this change point appears consistently in all models fitted to the full data set (although it is at location 143 rather than 142 in the model with five change points).

The first change point in the selected model, at the 15th sample site, at which the variance of pH increases, occurs in broadleaf woodland. There is no obvious relation to geology, the first two segments of the transect both lie over the Lower Greensand. The last three locations in the transect form a segment. Note that the last two are over Chalky Boulder Clay and have a notably smaller pH than neighbouring samples. The Chalky Boulder Clay, as the name indicates, is not decalcified, and the pH of the last three observations are all larger than 7, but there seems to be more scope for variation of pH in soils formed over this superficial material than there is for soils formed directly over the Chalk. However, care must be taken not to over-interpret an effect that only involves the last three observations. The nugget variance in the final segment is larger than for the segment immediately to the north over Chalk, and this may just reflect some outlying observations rather than an underlying non-stationarity that would have been reinforced by additional observations on Boulder Clay over the Chalk.

There are some differences in the change points for the model fitted to the subset of 150 data. Note that the change point at location 142, with a marked reduction in the variance, but the other change points are some way from those found for the full data set. It is not surprising that the removal of a substantial number of observations should affect

the estimated parameters, including the locations at which changes in variance are indicated. However, this analysis shows how the LMM can be used to detect non-stationary features in irregularly sampled data which would be harder to examine with a wavelet transformation.

Fig. 5 shows three autocorrelation functions for LMM fitted to these data. The dotted line shows the autocorrelation for the stationary LMM (parameters in Table 4 with $\nu = 1.5$). The dashed line shows the pre-tempering autocorrelation with parameters for the selected model (four change points, constant η, different variances). The solid line shows the autocorrelation for the non-stationary LMM, recall that only the variances changed between segments of the transect in the selected model. While, in principle, the autocorrelations and variances of the random effects of a spatial mixed model can be treated independently of each other as stationary or non-stationary, it is of interest that when the assumption of stationarity in the variance was relaxed the stationary autocorrelation model showed spatial dependence over a longer effective range. This shows that a stationary spatial model may give a misleading impression of the spatial dependence of a variable even when it is stationary if the assumption of a stationary variance is not justified.

Fig. 6 shows the overall wavelet variances for the first four dilations of the selected wavelet function. These are the four scales equivalent to 2, 4, 8 and 16 times the sampling interval (i.e. approximately 60, 120, 240 and 480 m). Fig. 7 shows the corresponding components of the MRA of the data. These show the marked reduction in variance at all scales near the 142nd location. The significant change points are listed in Table 6. Fig. 8 shows the diagnostic plot of the S_k statistic for the second scale (120 m) with the candidate change point, and the empirical distribution of the B-statistic under the null hypothesis, with the value of the actual B statistic indicated by a symbol on the axis of the plot. The locations of significant change points are indicated by vertical lines on the plot of the MRA (Fig. 7), and also on Fig. 4 where the vertical dotted lines at the bottom of the graph delineate segments, labelled with letters that differ with respect to wavelet variance at one or more scales. As noted above there was a change in variance near the 142nd location at all scales, the grey bar on Fig. 4 indicates the interval over which these change points occurred. Fig. 9 shows the separate wavelet variances computed from the MODWT wavelet coefficients at each scale within each segment separately.
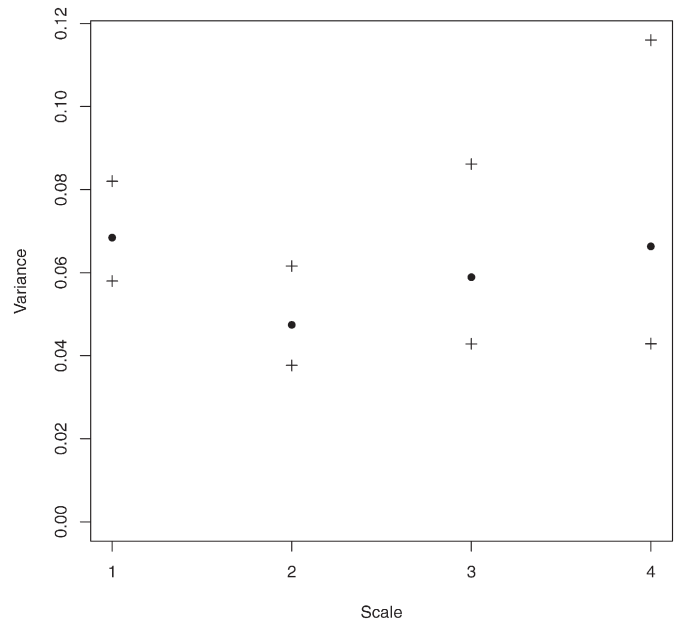

Fig. 6. Wavelet variances for pH data estimated from MODWT coefficients of four dilations of Daubechies's extremal phase wavelet, 2 vanishing moments. Also shown are 95% confidence intervals.

The change in wavelet variance between segment C and D is seen at all four scales and is consistent with the change point in the LMM with
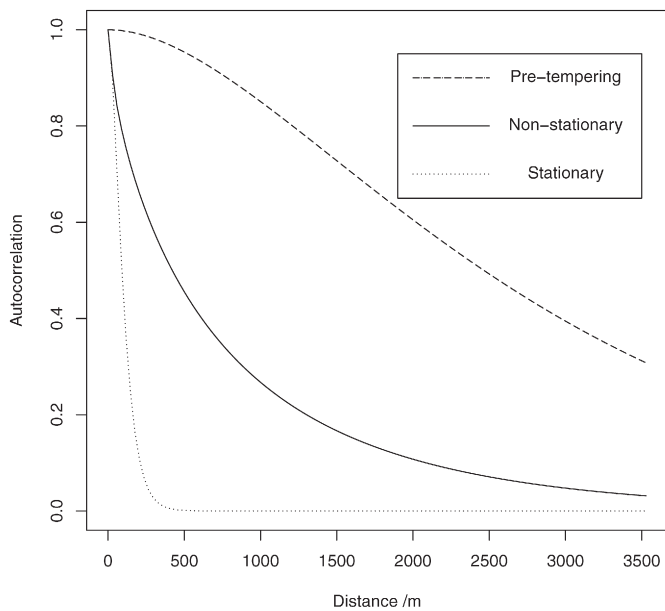

Fig. 5. Correlation functions from LMM for the autocorrelated random variable U. The 'pre-tempered' autocorrelation is the function that provided the basis for a tempered spectrum with maximum likelihood.
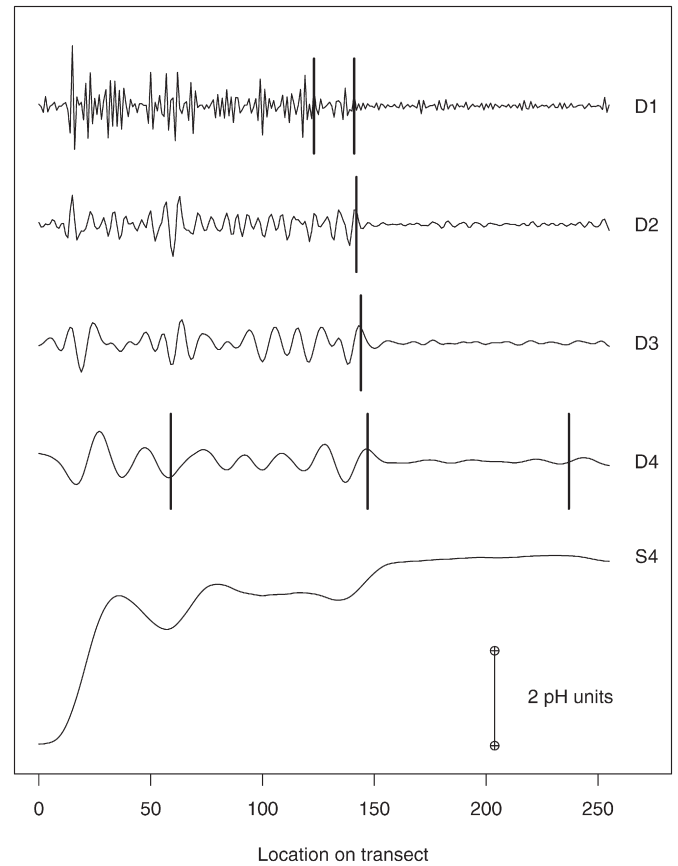

Fig. 7. Detail components and smooth component for MRA of pH data by MODWT with Daubechies's extremal phase wavelet, 2 vanishing moments. The vertical symbol at the bottom of the plot shows 2 pH units on the vertical scale. Vertical lines show significant change points in the wavelet variance at each scale.

Table 6
Results of tests of homogeneity of wavelet variance, using MODWT and Daubechies's extremal phase wavelet with 2 vanishing moments. The reported P-values are obtained by the method of Inclan–Tiao or by Monte Carlo simulation, following Percival and Walden (2000). In the latter case the null hypothesis can be rejected at 0.05, 0.01 or 0.001 respectively if the empirical P-value is less than 0.044, 0.007 or 0.0004 respectively, indicated by *, ** or *** in the table. P-values shown in italic are those for which the candidate change point was retested with the order of values reversed as described in the text.

Segment	Candidate change point	B	P-value		Method of inference[a]
Scale 1					
256	124	0.445	≪0.001	***	IT
123	63	0.198	*0.078*		MC
256	142	0.526	<0.001	***	MC
141	133	0.367	*0.117*		MC
256	253	0.146	0.338		MC
Scale 2					
256	143	0.424	<0.001	***	MC
142	64	0.169	0.182		MC
256	241	0.660	0.328		MC
Scale 3					
256	145	0.423	<0.001	***	MC
144	27	0.106	0.860		MC
256	244	0.087	0.976		MC
Scale 4					
256	60	0.421	0.001	**	MC
59	30	0.326	*0.241*		MC
256	148	0.516	<0.001	***	MC
147	132	0.358	0.116		MC
256	238	0.457	0.016	*	MC
237	217	0.267	0.354		MC
256	249	0.458	0.129		MC

[a] Inclan–Tiao (IT) or Monte Carlo (MC).

spectral tempering. Both the wavelet transform and the LMM show this to be the dominant non-stationary feature in the soil variation.

The change from segment A to B in the wavelet analysis, with a reduction in variance at the coarsest scale (480 m) is at location 60, about 330 m from the boundary between the Lower Greensand and the Gault clay, which is a change point in the selected LMM. This change can be given an interpretation comparable with the change at location 71 for the LMM, with a smaller variance (at the 480-m scale) south of the change point over the Gault Clay. The wavelet variances in Fig. 9 for these segments show that the variances at the three finer scales are all similar in segments A, B and C.

Other change points indicated by the MODWT do not so clearly correspond to change points in the LMM. No change point (or rejected candidate change point) in the wavelet analysis corresponds to the change at the 15th location in the selected LMM. Similarly, there is no counterpart in the LMM to the change in the wavelet coefficients between segments B and C. The change-point between segments D and E is someway north of the change-point in the LMM at the Chalky Boulder Clay over the White Chalk Group, and is a reduction in variance at a coarse scale, whereas the LMM showed an increase in variance for the last three locations over the Chalky Boulder Clay.

The selected LMM has a stationary autocorrelation function with changes only in the variance. It is interesting to compare the sets of wavelet variances in Fig. 9 in the light of this. For all segments other than D the wavelet variance at the finest scale is the largest, and there is a general decline in variance at coarser scales. This is consistent with a stationary autocorrelation, but the wavelet variances for segment D, all of similar order, with the maximum at the third scale (240 m) indicates that the variation at these scales is smoother than in the other segments.
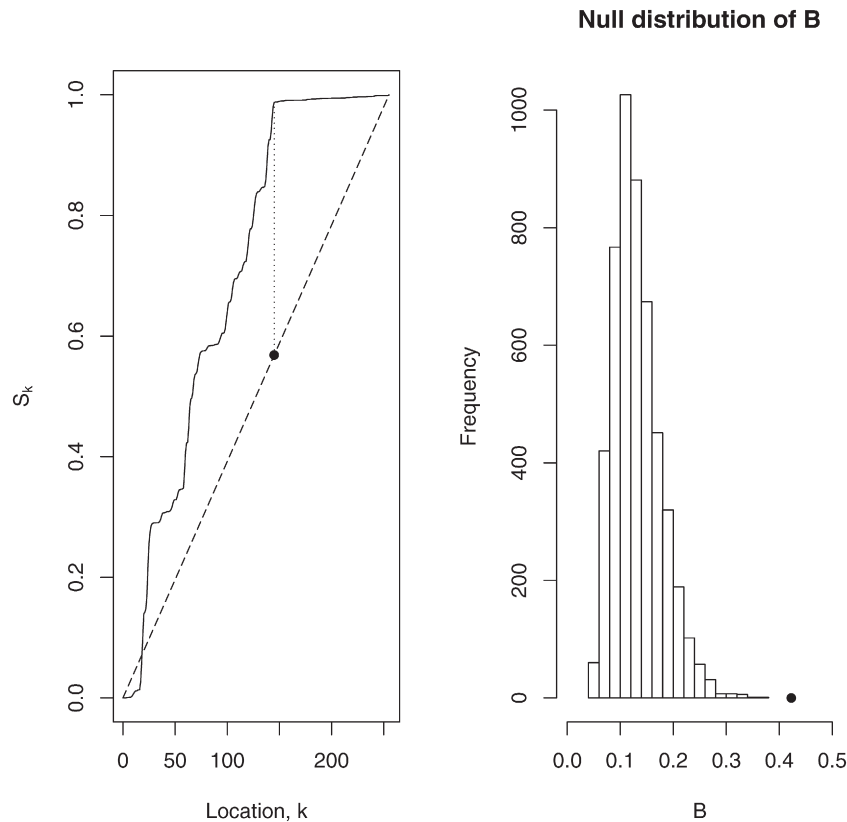
Null distribution of B

Fig. 8. Detection of the change point for the second detail component. On the left is plotted (solid line) the value of S_k against k (location on the transect). The dashed line shows the expected value of S_k for a homogeneous variable. The solid symbol indicates the location of the candidate change point, and the length of the dotted line is the value of the B statistic. On the left is shown the distribution of the corresponding B statistic for 5000 realizations of a homogeneous random variable of length 256. The solid symbol shows the value of the statistic for the data on soil pH.
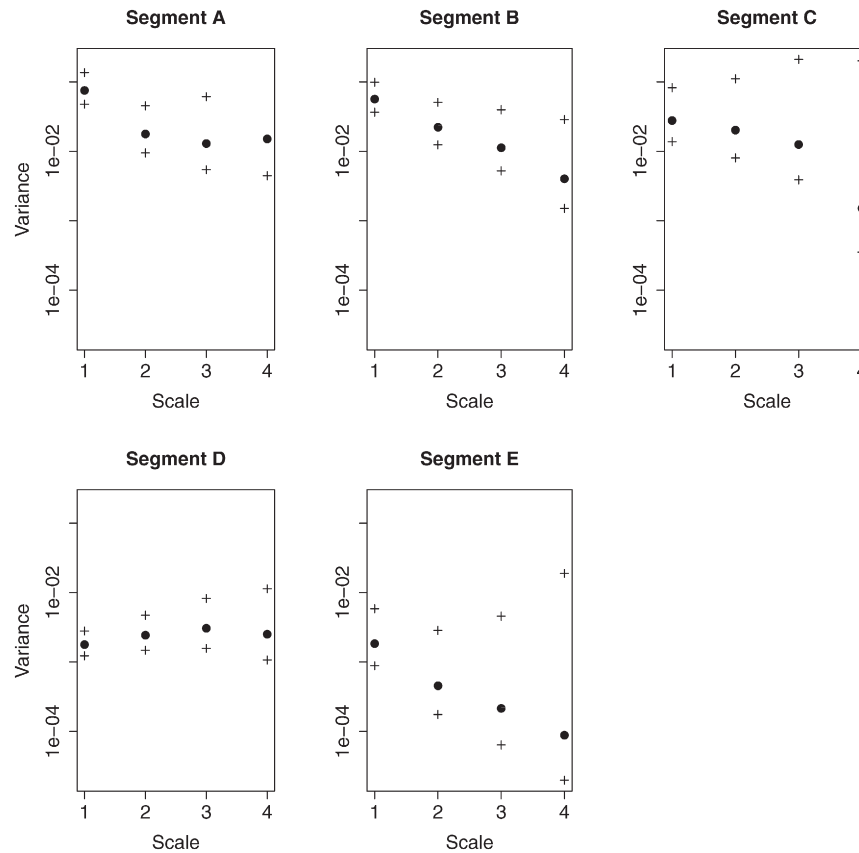
Fig. 9. Wavelet variances with confidence intervals for five segments of the transect defined by change points identified by wavelet analysis. Note that variances are shown on a log scale.

5. Discussion

The LMM with spectral tempering provides a basis for modelling non-stationary variation, and the inferences from likelihood using the AIC allows us to justify the number of change points in the finally-selected model. In the data set examined here the model parameters provide insight into the sources of soil variation. In particular the model shows how differences between the parent materials with respect to the constraints on possible soil pH (narrowest over the Chalk and widest over the Lower Greensand) give rise to non-stationarity of the variance in soil pH, although the autocorrelation appears to be stationary. The difference in variances of pH between the parent materials make sense in terms of basic soil chemistry. In calcareous soils without large concentrations of sodium the pH is determined by the equilibrium of the calcium carbonate–calcium bicarbonate–carbon dioxide system. The pH depends on the partial pressure of carbon dioxide but can be expected to be buffered between 7.5 and 8.5 (Russell, 1973). This is consistent with the narrow range of soil pH over the Chalk on this transect. A much wider range of soil pH is seen in the soil over the other parent materials which ranges from the Lower Greensand under woodland with pH as small as 4, and pH approaching 8, presumably due to the presence of free calcium carbonate, in soils formed from Chalky Boulder Clay. In principle land use will influence soil pH, but there is no evidence here of trends in land use which might give rise to non-stationarity in the variance of pH.

It is interesting to compare the autocorrelation functions in Fig. 5. When stationary variances are assumed there is a much shorter range of autocorrelation than when non-stationary variances are modelled. This may be because large contrasts over short distances between or within particularly variable parts of the transect dominate the stationary model. This would have implications for the reliability of any decisions made about further sampling in comparable environments based on the stationary LMM.

The wavelet transform also allows us to identify locations on a transect at which the variance of a property changes. In this case study, where both methods were used on a single data set we can see consistencies between the two analyses. The clearest common feature is the pronounced change in variance near location 142 where the effect of Chalk in the parent material on the scope for variation of pH is seen. There is a change point identified at this location in the LMM, and change points near this location for all scales of the wavelet transform. There is some variation in the exact locations of the changes in wavelet variance, but this should not be surprising as the spatial resolution of the wavelet coefficients declines with increasing scale (Percival and Walden, 2000).

The LMM identified a change in the variance coincident with the boundary between soils formed over the Lower Greensand and those formed over the Gault Clay. The variance was smaller over the latter soils. There was a change in the wavelet variance for the fourth dilation (scale 480 m) at location 60, about 330 m from this soil boundary, with a reduction in wavelet variance. The difference in location of these change points could be due to various factors.

First, as noted above, the spatial resolution of the wavelet coefficients is scale-dependent and decreases with scale (while the frequency resolution increases). In the LMM fitted here the autocorrelation is stationary so the change point indicates a change in variance over all spatial scales represented by the random effects of the model. So, while the change points are likely both to be explained by the soil boundary, the wavelet analysis is considering variation at particular scales separately, and the spatial resolution of local features such as change points is poorer at the coarser scales.

Second, in the wavelet analysis we look for changes in variation at scales represented by the detail components of the MRA, the variation not represented by the smooth component. In the LMM we look for non-stationarity in the variance of the random effects in the model, the variation not accounted for by fixed effects (here the spline

model). If one compares the smooth component in the MRA (Fig. 7) with the spline function from the LMM (Fig. 2), then one can see that, though broadly similar, there are some differences. In particular there are fluctuations in the spline function between locations 50 and 150 that are not seen in the smooth component of the MRA, and which probably contribute largely to the variation at the 480-m scale. For this reason the wavelet analysis and the LMM are not examining exactly the same variation, with the differences likely to be greatest at the coarsest scale of the wavelet analysis.

Finally, it should be recalled that all change points in the non-stationary LMM are found in a single optimization. In the case of the wavelet analysis the change points are found in sequence, so that the position of a change point in the wavelet analysis is based on a contrast in variance between two segments of the transect, one or both of which may subsequently be subdivided by further change points. Given these factors it should not be surprising that, while consistent features can be seen in the two analyses of this data set, there are also differences.

The wavelet transform has advantages over the LMM. The algorithm is fast to execute, and the analyses of the wavelet coefficients for change detection, even with Monte Carlo inference for the distribution of the *B* statistic in the null case, are faster than maximization of the residual likelihood of the LMM. However, the wavelet transform does require the selection of an appropriate wavelet function, on which the inferences are conditioned, and, as noted above, there is a pay-off between frequency and spatial resolution which depends on scale.

Estimation of the parameters of the LMM is computationally intensive, in particular the profiling procedure to find a good set of parameters for the pre-tempering spectrum is time-consuming, but the LMM framework does allow for inference using the log residual likelihood, via Akaike's information criterion. Furthermore, the LMM procedure can be applied to irregularly-sampled data, as seen here, in a more straightforward way than can be done with the wavelet transform (Milne and Lark, 2009). However, the interpretation of the LMM with spectral tempering for markedly irregularly sampled data may not be straightforward. When sampling density changes systematically one component of the spectral decomposition (eigenvector) may represent rather different frequencies at different locations, for example (Haskard and Lark, 2009), which may reduce the sensitivity with which changes in the covariance can be modelled by tempering. The properties of the spectral decomposition and tempered empirical spectrum under irregular sampling, and the implications for the spectral tempering method, require further investigation.

This study has compared the two methods for the analysis of non-stationary variation of soil properties in a case where discrete changes may be expected at locations where the transect traverses boundaries between distinctive soil parent materials. The wavelet transform is particularly well-suited to such a case. However, it is conceivable that a non-stationary property may show continuous variation in space, in which case discrete change points are not appropriate. The LMM with spectral tempering has greater flexibility to deal with such variation, by treating the tempering parameter as some function of a continuous covariate or allowing it to show a spatial trend, perhaps using a B-spline basis in each case (Haskard et al., 2010a). It should also be noted that the LMM can be readily extended to the two-dimensional case when more general models are available for the tempering parameter (e.g. Haskard et al., 2010b).

6. Conclusions

I have presented the first comparison between a wavelet analysis and an analysis with a non-stationary LMM of a common soil data set. The two analyses both gave consistent results on broad changes in the variability of soil pH along a transect which are not consistent with assumptions of stationarity in the variance. These changes were attributable to differences in soil parent material, which constrain the ranges of possible soil pH. The largest change, induced by the appearance of

Chalk parent materials in one section of the transect, was identified by both analyses. Differences between the analyses can be attributed to the way in which the wavelet analysis examines different scales separately, to the scale-dependent spatial resolution of the wavelet analysis, to the fact that the variation attributed to the smooth component of the wavelet MRA and the fixed effects of the LMM may not be exactly the same and to the way that the wavelet analysis identifies change points sequentially rather than simultaneously. Both analyses allow for formal testing of the hypothesis of stationarity, and the complexity of the non-stationary model. The main advantage of the wavelet transform is its computational efficiency. The LMM with spectral tempering is computationally demanding. However, the LMM with spectral tempering allows for a more flexible family of models to specify the non-stationary variation of the random component than the discrete change points identified by a wavelet analysis. It can also be more readily extended to two dimensions, and to deal with irregularly sampled data.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademia Kiado, Budapest, pp. 267–281.
Beckett, P.H.T., Webster, R., 1971. Soil variability: a review. Soils Fertil. 34, 1–15.
Biswas, A., 2014. Landscape characteristics influence the spatial pattern of soil water storage: similarity over times and at depths. Catena 116, 68–77.
British Geological Survey, 1992. 1:50 000 Series, England and Wales Sheet 220, Leighton Buzzard, Solid and Drift Geology. British Geological Survey, Keyworth.
Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. J. R. Stat. Soc. B 37, 149–192.
Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets. Commun. Pure Appl. Math. 41, 909–996.
Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
Diggle, P.J., Ribeiro, P.J., 2007. Model-based Geostatistics. Springer, New York.
Haskard, K.A., Lark, R.M., 2009. Modelling non-stationary variance of soil properties by tempering an empirical spectrum. Geoderma 153, 18–28.
Haskard, K.A., Lark, R.M., 2010. Spectral tempering to model non-stationary variation of soil properties: sensitivity to the initial stationary model. Geoderma 159, 350–357.
Haskard, K.A., Welham, S.J., Lark, R.M., 2010a. A linear mixed model with spectral tempering of the variance parameters for nitrous oxide emission rates from soil across an agricultural landscape. Geoderma 159, 358–370.
Haskard, K.A., Rawlins, B.G., Lark, R.M., 2010b. A linear mixed model, with non-stationary mean and covariance, for soil potassium based on gamma radiometry. Biogeosciences 7, 2081–2089.
Inclán, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. J. Am. Stat. Assoc. 427, 913–923.
IUSS Working Group WRB, 2006. World Reference Base for Soil Resources 2006, 2nd. World Soil Resources Report No 103 edn. FAO, Rome.
King, D.W., 1969. Soils of the Luton and Bedford district. Special Survey No 1Soil Survey of England and Wales. Lawes Agricultural Trust, Harpenden.
Lark, R.M., 2009. Kriging a soil variable with a simple non-stationary variance model. J. Agric. Biol. Environ. Stat. 14, 301–321.
Lark, R.M., Webster, R., 1999. Analysis and elucidation of soil variation using wavelets. Eur. J. Soil Sci. 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. Eur. J. Soil Sci. 52, 547–562.
Lark, R.M., Milne, A.E., Addiscott, T.M., Goulding, K.W.T., Webster, C.P., O'Flaherty, S., 2004. Analysing spatially intermittent variation of nitrous oxide emissions from soil with wavelets and the implications for sampling. Eur. J. Soil Sci. 55, 611–627.
Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Eur. J. Soil Sci. 57, 787–799.
Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693.

Milne, A.E., Lark, R.M., 2009. Wavelet transforms applied to irregularly sampled soil data. Math. Geosci. 41, 661–678.

Milne, A.E., Haskard, K.A., Webster, C.P., Truan, I.A., Goulding, K.W.T., Lark, R.M., 2013. Wavelet analysis of the variability of nitrous oxide emissions from soil at decametre to kilometre scales. J. Environ. Qual. 42, 1070–1079.

Percival, D.P., 1995. On estimation of the wavelet variance. Biometrika 82, 619–631.

Percival, D.B., Guttorp, P., 1994. Long-memory processes, the Allan variance and wavelets. In: Foufoula-Georgiou, E., Kumar, P. (Eds.), Wavelets in Geophysics. Academic Press, New York, pp. 325–344.

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

Pintore, A., Holmes, C.C., 2004. Spatially adaptive non-stationary covariance functions via spatially adaptive spectra. Technical Report. Department of Statistics, University of Oxford, Oxford, UK (Available at: http://www.stats.ox.ac.uk/~cholmes/Reports/spectral_tempering.pdf).

Pintore, A., Holmes, C.C., 2005. A dimension-reducing approach for spectral tempering using empirical orthogonal functions. In: Leuangthong, O., Deutsch, C.V. (Eds.), Geostatistics Banff 2004. Springer, Dordrecht, pp. 1007–1015.

R Core Team, 2014. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (URL http://www.R-project.org/).

Russell, E.W., 1973. Soil Conditions and Plant Growth. 10th ed. Longman, London.

Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer, New York.

Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J. Soil Sci. 41, 473–490.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. second ed. John Wiley & Sons, Chichester.

Whitcher, B., 2015. Waveslim: Basic Wavelet Routines for One-, Two- and Three-dimensional Signal Processing. R Package Version 1.7.5. (Available at: https://CRAN.R-project.org/web/packages/waveslim/waveslim.pdf).

Whitcher, B., Guttorp, P., Percival, D.B., 2000. Multiscale detection and location of multiple variance changes in the presence of long memory. J. Stat. Comput. Simul. 68, 65–88.

Zeleke, T.B., Si, B., 2007. Wavelet-based multifractal analysis of field scale variability in soil water retention. Water Resour. Res. 43, W07446.