European Journal of Soil Science

European Journal of Soil Science, August 2011, **62**, 617–628 doi: 10.1111/j.1365-2389.2011.01352.x

Wavelet analysis of soil variation at nanometre- to micrometre-scales: an example of organic carbon content in a micro-aggregate

A. E. Milne[a], J. Lehmann[b], D. Solomon[b] & R. M. Lark[a]
[a]*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK, and* [b]*Department of Crop and Soil Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14853, USA*

Summary

This paper demonstrates the potential of wavelet analysis to investigate fine-scale spatial variation in soil without statistical assumptions that are generally implausible. We analysed the optical densities of different forms of carbon which were measured at intervals of 50 nm along a 16-μm transect on a soil micro-aggregate using near-edge X-ray fine-structure spectroscopy (NEXAFS). We found different patterns of scale-dependent variation between the carbon forms, which could be represented by pair-wise wavelet correlations at the different scales, and by principal components analysis of all the correlations at each scale. These results represent only one small soil micro-aggregate and are not presented as general findings about soil carbon, but they do indicate that fine-scale variation of soil carbon can be complex in ways that the wavelet analysis can accommodate but alternative spatial statistics such as variograms cannot. Among the patterns of variation that the analysis could identify were scale-dependent correlations of the different forms of carbon. In some cases, positive correlations were found at coarser scales and negative at the finest scales, suggesting a multi-scale pattern in which contrasting forms of carbon are deposited in common clumps but at finer scales either one or the other form dominates. Aromatic and carboxylic carbon varied jointly in this way. Other forms of carbon, such as carboxylic and aliphatic carbon, were strongly correlated at the finest scales but not the coarser scales. We found evidence for changes in the variance and correlation of forms of carbon along the transect, indicating that the spatial distribution of carbon at these fine scales may be very complex in ways that are inconsistent with the assumptions of geostatistics. This quantitative analysis of the spatial patterns of different soil components at micro-scales offers a basis for formulating and testing specific hypotheses on replicated samples.

Introduction

Micro-scale soil studies, methods and motivation

There is a growing interest in methods to examine the structure and composition of soil at nanometre- to micrometre-scale. These methods include X-ray microtomography to investigate soil physical structure (Feeney *et al.*, 2006), and methods to study the chemical composition of the soil including X-ray fluorescence (XRF, Isaure *et al.*, 2005), energy-dispersive spectroscopy (EDS, Coppin *et al.*, 2009), secondary ion mass spectroscopy (SIMS, De Rito *et al.*, 2005) and near-edge X-ray fine-structure spectroscopy combined with scanning transmission X-ray microscopy (NEXAFS-STXM, Lehmann *et al.*, 2008). We refer to these and similar technologies collectively as micro-scale methods.

Correspondence: A. E. Milne. E-mail: alice.milne@bbsrc.ac.uk

Received 12 February 2010; revised version accepted 21 December 2010

Young & Crawford (2004) argued that micro-scale methods probe the structure and composition of soil at spatial scales which are fundamental to soil functions, at least in so far as those functions depend on microbial processes. Some soil scientists have queried whether this is generally true. Fierer *et al.* (2009) argued that, for many geochemical functions of soil, the aggregate effects of processes over a heterogeneous assembly of micro-sites may be predictable without the need to measure the fine-scale variation. However, Jacobson *et al.* (2007) presented a case where the effects of a soil property on functions could not be understood without observing the fine-scale variation. They noted that apparent threshold concentrations of copper (measured on cores), above which microbial numbers and metabolism are affected in soil, are not consistent, varying from 5 to over 1000 mg kg^{-1}. They mapped the concentration of copper in a vineyard soil from Burgundy at resolutions of 0.3 mm and 20 μm using electron microprobes and synchrotron XRF spectroscopy. This showed that the distribution of copper was very patchy and therefore, while the overall concentration was large (417 mg kg^{-1}), micro-organisms would be exposed to very small concentrations in most of the soil volume.

The proposition that variation of soil structure and composition at fine scales must be understood if soil functions are to be predicted is therefore a general hypothesis that has been validated in some examples, and must be tested for other soil conditions and functions. As noted above, there are many measurement technologies to facilitate such studies. However, methods to analyse the resulting data, have received rather less attention. Process models have been used to analyse micro-scale data for specific purposes. A model of the soil as solid and voids can be used directly in Lattice–Boltzmann models to predict hydraulic properties by simulating fluid flows through the structure (Zhang *et al.*, 2005). Simple visualization of micro-scale soil data provides insight. Lehmann *et al.* (2008) published images that showed the spatial variation of the optical density of different forms of carbon in a soil micro-aggregate, and used cluster analysis on these data to highlight contrasting regions. By visual interpretation they could conclude that forms of carbon characteristic of microbes were found along pores, and that organic matter dominated by black carbon was clumped in particulates. However, visual interpretation did not allow them to express these conclusions in a quantitative way, which might provide a basis to compare variation between contrasting soils, or to identify scale-dependent associations between different forms of carbon quantitatively. This requires appropriate statistical methods.

The objective of this paper is to illustrate how methods for spatial analysis which have been applied to conventional soil data sets measured on cores at field scale or coarser can be used to identify hypothesized general patterns of scale-dependent variation and co-variation of soil properties at micro-scales, and so to generate more specific hypotheses about soil composition.

Micro-scale soil studies, spatial analysis of data

A measurement of a soil property corresponds to a finite volume of soil material. This volume of soil material (defined with respect to depth, shape and dimensions) is known as the support of the data. Consider a set of measurements on some support made on a regular sample array such as a grid or transect at intervals length x_0. The variations in these data occur at different scales. Components of the variation at different scales may be caused by different factors of soil formation. Micro-topography imposes variations over scales of a few tens of metres, differences between land-uses cause variations over a few hundred metres (field to field) and parent material differences cause variation at scales of a few kilometres. There are three general questions which soil scientists might ask about soil variation. These apply to micro-scale observations as well as to observations at field or landscape scale, and are as follows:

(i) At what dominant spatial scales does a soil property vary? It is useful to study how soil variation is partitioned between scales since this may indicate the relative importance of different factors of formation, and its implications for particular processes. Soil variation is commonly scale-dependent in the sense that distinct contributions to variation, of differing magnitude, occur at diverse scales and so the variation of a soil property depends on the support of the measurements and the extent of the sampled region. If there was no scale dependence over some scale interval (e.g. pore to core) then variation in this interval is effectively white noise.

(ii) At what spatial scales do two or more soil properties co-vary? Two soil variables might be influenced by common factors, in addition to particular factors specific to just one of them. It is therefore often found that the overall correlation between two soil properties masks underlying scale-dependent correlations.

(iii) Is the variation and co-variation of properties consistent across a region? The variation of a soil property, and its correlation with other properties, is not necessarily uniform. The dominant scales of variation and the overall variability of a soil property may change in space as different limiting or dominant factors come into play.

All three of these questions must be asked of soil variables measured at micro-scales. The central hypothesis of micro-scale research is that the behaviour of the soil at the support of the core must be explained by its variation at finer scales. This was exemplified by Jacobson *et al.* (2007), cited above. In terms of spatial analysis, we can say that the key feature of variation of copper concentration at within-core scales is the pronounced scale dependence. If there were no scale dependence within the core, then the existence of refugia for the microbiota, where key functions are not affected by copper, would be precluded. Similarly, strong correlation between concentrations of two metals in soil at a particular scale indicate a common factor in their distribution such as joint deposition on mineral surfaces or adsorption by particulate organic matter. It might also be expected that the variation or co-variation of soil properties at micro-scales is not uniform, and might change from the centre to the edge of micro-aggregates, or with increasing distance from the rhizosphere.

Attempts have been made to characterize the scale-dependent variation of micro-scale data sets by computing their variograms. Feeney *et al.* (2006) computed variograms of X-ray tomographic data to characterize soil structure and Nunan *et al.* (2006) used them to draw inferences about impacts of particular treatments on structure. Geostatistical methods have also been used to describe the spatial distribution of microbial colonies identified in images of soil thin sections (Nunan *et al.*, 2003). However, the variogram is used in geostatistics primarily to model the spatial covariance structure of a variable and predict it at unsampled sites. It is a blunt instrument for description of complex spatial variation as explained and illustrated by Lark (2010). This might account for the finding of Nunan *et al.* (2006) that the variograms of porosity in soils at micro-scales under strongly contrasting treatments were rather similar.

Furthermore, the computation of the variogram requires a stationarity assumption, such that the variance of the difference between two locations depends only on their separation in space, and not on their location. This assumption prejudges the third general question about spatial variation identified above. When we use variograms it is required that the variation does not change systematically across the region of interest. For this reason, Lark & Webster (1999) proposed that the wavelet transform is used for spatial analysis of soil variables, so as to provide quantitative information on scale-dependent spatial variation without relying on any implausible assumptions of stationarity. We therefore propose that wavelet transforms are most appropriate for addressing questions about micro-scale data.

The scope of this paper is as follows. We present a case study, the analysis of the NEXAFS-STXM data on forms of carbon in a soil micro-aggregate from Kenya first published by Lehmann *et al.* (2008). We recognize that this analysis does not allow generalizations beyond the micro-aggregate in question about the spatial distribution of soil carbon. Our aim rather is to show how wavelet analysis allows us to test general hypotheses about soil variation at micro-scale, framed in the terms introduced above, and to generate from the analysis more specific hypotheses about spatial variation which can then be tested given the information on scale dependence that the analysis has provided.

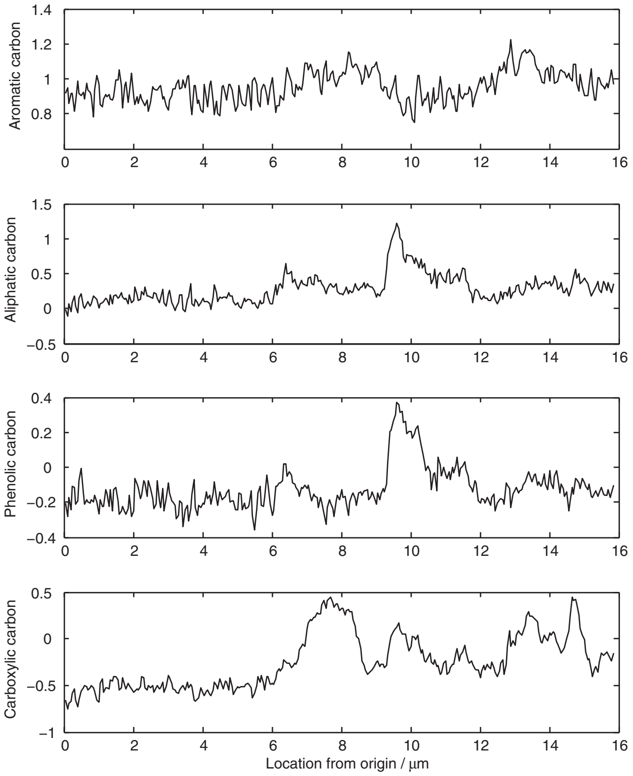

Figure 1 Optical densities for four forms of carbon on a 16-µm transect.

Methods

Data

The data that we use, from a soil in Nandi Forest, western Kenya, were first published by Lehmann *et al.* (2008), who presented the technical details. The basic data were measurements, on a sectioned micro-aggregate of approximately 30 × 40 µm, of transmitted X-ray intensity over increments of ionization energy, for pixels of length 50 nm. Characteristic ratios of intensity at different energies were computed. These correspond to the optical density of the soil with respect to total carbon content and the separate contents of aromatic, phenolic, aliphatic and carboxyl forms of carbon. Because the most powerful inferences based on the wavelet transform are restricted to one dimension, we extracted four transects from the image for more detailed analysis. Since all the transects yielded similar results, we focus on a single example in this paper. Figure 1 shows optical densities for the different forms of carbon along the transect.

Hypothesized patterns of spatial variation and co-variation for forms of carbon

We hypothesize five general patterns of spatial variation and co-variation which might be expected for forms of soil carbon at micro-scales and then use the wavelet analyses to identify these in the data. The optical density for a particular form of soil carbon recorded for a pixel depends on the composition of the soil carbon in the corresponding part of the sectioned micro-aggregate. This carbon might be derived from various sources such as detritus (including more or less decomposed litter, dead roots, faecal matter from mesofauna), exudates, sloughed-off root material, black carbon, known to be significant in the soil at Nandi (Nguyen *et al.*, 2008). Some of this material will have been altered in its composition *in situ*, perhaps by microbial activity. As a result, the forms of carbon at any particular site will be due to inherited properties (such as characteristic components of black carbon) and local conditions for microbial or abiotic transformations. The first hypothesized general pattern is concerned with scale-dependent variation of the different forms of carbon. Hypothetical patterns (ii)–(iv) concern scale-dependent correlation between the different forms of carbon. Finally, (v), we consider the uniformity of spatial variability from one part of the soil to another.

(i) Any form of carbon in the soil will show scale-dependent spatial variation at micro-scales because the distribution in space of the carbon at these scales will reflect the ‘clumped’ distribution of its sources and the variation at different scales of conditions for biotic and abiotic transformations of carbon which are controlled, in part, by the physical structure of the soil.

(ii) At the coarsest spatial scales within a micro-aggregate, we expect the different forms of carbon to show positively correlated patterns of variation. This might reflect the existence of ‘organic-rich’ patches which contrast with voids, large mineral grains or clusters of flocculated clay.

(iii) At intermediate scales some of the forms of carbon will show positively correlated spatial variations because they tend to be inherited jointly from a particular source (e.g. a fragment of black carbon).

(iv) At the finest scales we expect negative correlations between forms of carbon if the spatial resolution is sufficiently fine that the region corresponding to a particular pixel is, usually, dominated by a single form.

(v) The strength of the correlations or the particular patterns of scale-dependent correlation hypothesized above (particularly under (ii) and (iii)) might change from one part of the soil to another because of variations in soil architecture (such as the distribution of pore volumes) or the dominant sources of organic matter. Similarly the scale-dependent variance might change in space.

The identification, or otherwise, of these hypothesized patterns and their association with specific spatial scales by an appropriate spatial analysis would show the potential of wavelet analysis to address more specific questions about the micro-scale distribution of soil carbon which are relevant to the prediction of soil functions (e.g. are more recalcitrant and labile forms of carbon found together, or in contrasting patches?) and to understanding its origins and development.

Wavelet analyses

For spatial analysis, we use the maximal overlap discrete wavelet transform (MODWT) proposed by Percival & Guttorp (1994) and analyses described by Percival & Walden (2000) and Lark & Webster (2001). Lark & Webster (1999, 2001) give a general introduction to wavelet transforms. In summary, in a wavelet transform we partition the variation of a signal into additive components that correspond to different intervals of spatial scale from fine (short-range) to coarse (long-range). The particular advantage of wavelet transforms over others that achieve a comparable analysis (such as the Fourier transform) is that the wavelet transform produces a set of local coefficients for each scale so that if the variability of the signal at some scale is not uniform (such as from the centre towards the edge of a micro-aggregate) or shows local intermittent features (such as near the edges of pores) then this can be identified. The wavelet analyses described below therefore allow us to study the scale dependence of the variation and co-variation of the different forms of carbon on the transect and to look for changes in the variability along the transect. They therefore allow us to test all five hypotheses.

Wavelet scales. The finest scale interval that we can study in data from a transect with measurements at intervals x_0 has a lower bound of $2x_0$ (corresponding to the Nyquist frequency $1/2x_0$). The scales in a wavelet transform are labelled with the lower bound, in the MODWT the successive scales are 2^1x_0, 2^2x_0, $2^3x_0, \ldots$, so in the study reported here where $x_0 = 50$ nm, the MODWT represents the data by sets of coefficients with scale labels 0.1 (which by the convention refers to the scale interval 0.1–0.2 μm), 0.2, 0.4, 0.8, 1.6, 3.2 and >6.4 μm. The first six of these scales are 'detail' coefficients, extracted from the signal by successive dilations of a high-pass wavelet filter, and the seventh is a 'scaling' coefficient, which represents the filtered version of the signal when all the details are subtracted.

Wavelet coefficients, variances and co-variances. Let us denote the ith MODWT coefficient of variable u along a one-dimensional transect at scale $2^j x_0$ by $d^u_{i,j}$. This coefficient is obtained by multiplying the elements of the jth dilation of the basic wavelet filter, translated to location i, with the corresponding values in the signal and summing the products. Because a wavelet has 'compact support', which means that it tends rapidly to zero away from location i, the resulting coefficient responds only to local variation. This wavelet coefficient therefore represents the local variation of the data at a particular scale interval. Because of the properties of the basic wavelet function (see Lark & Webster, 1999) the wavelet coefficients at scale j can be back-transformed to yield a scale-specific 'detail' component of the original data. If we have N data where $2^k \le N < 2^{k+1}$ then we can compute $k - 1$ such detail components and a residual 'smooth' component. The set of detail components for scales $j = 1, 2, \ldots, k - 1$ along with the 'smooth' component comprise a multi-resolution analysis (MRA) of the data into components of different scale which, if summed, reconstitute the original data.

We may estimate a scale-specific component of the variance of our data on variable u, the wavelet variance $\sigma^2_{u,j}$, from the N_j wavelet coefficients at this scale. The wavelet variance therefore indicates how much of the variation of the variable of interest is associated with the particular scale interval. The estimator is given by Percival (1995)

$$\widehat{\sigma}^2_{u,j} = \frac{1}{N_j 2^j} \sum_{i=1}^{N_j} \left(d^u_{i,j}\right)^2 . \tag{1}$$

In geostatistical analysis, we assume that the variance of a soil property is uniform in space, but this is not generally plausible. However, since wavelet coefficients describe local variation, it is possible that a wavelet analysis need not make this same assumption. Whitcher *et al.* (2000) proposed a procedure to identify changes in the wavelet variance components on a one-dimensional signal and Lark & Webster (2001) applied this procedure to data on soil. Lark (2006) presents a somewhat refined version of the procedure, which we use here.

In similar fashion, when a MODWT has been performed on two variables u and v, we may estimate the scale-specific component of the covariance, the wavelet covariance $C^{u,v}_j$,

$$\widehat{C}^{u,v}_j = \frac{1}{N_j 2^j} \sum_{i=1}^{N_j} d^u_{i,j} d^v_{i,j} . \tag{2}$$

As with a conventional covariance, the interpretation is easier on standardization to a correlation in the interval $[-1, 1]$, so the wavelet correlation for the jth scale is estimated by

$$\widehat{\rho}_j^{u,v} = \frac{\widehat{C}_j^{u,v}}{\sqrt{\hat{\sigma}_{u,j}^2 \hat{\sigma}_{v,j}^2}}. \tag{3}$$

This measures the strength of the linear relationship between the components of variation of the two variables associated with the specific scale. As shown by Lark & Webster (2001) and Lark *et al.* (2004), the wavelet correlation often reveals relationships between variables at some scales that are obscured in the overall correlation over all scales.

As with the wavelet variance we may identify locations in space at which the wavelet correlation at some scale changes significantly. Lark *et al.* (2004) describe how this is done, in summary we search for candidate change points in the wavelet covariance, and then apply an appropriate test to the difference between the wavelet correlations either side of the change point. This is potentially interesting in the analysis of soil data, since changes in the correlation between variables across a region, of any size, may indicate changes in the dominant factors that drive some soil process as Lark *et al.* (2004) exemplify in a study at landscape scale.

We may define $\mathbf{R}_j$ as the wavelet correlation matrix for a set of variables at scale $2^j x_0$, such that

$$\left[\mathbf{R}_j\right]_{u,v} = \rho_j^{u,v}. \tag{4}$$

In this study, we computed MRAs for the optical densities on our transect across the micro-aggregate, associated with aromatic, aliphatic, phenolic and carboxylic carbon. We computed wavelet variances, co-variances and correlations, and undertook tests to identify change points in the variance and correlations following Lark *et al.* (2004), where it is also explained how confidence intervals may be computed for the wavelet correlations. We then computed the principal components of the wavelet correlation matrices for the forms of carbon for scales 0.1 to 3.2 μm. Principal components analysis is widely used in soil science to summarize the variation of a set of variables by identifying uncorrelated linear combinations of those variables (principal components) which account for as much as possible of the overall variation of the original data set. Webster (2001) shows how a plot of the correlation of each variable with the principal components allows one quickly to visualize the similarity or contrast between the spatial patterns for the forms of carbon at each scale, and we did that here. We also computed principal components for the four corresponding smooth components from the multi-resolution analysis to summarize the joint variation of the forms of carbon at scales >6.4 μm.

Note that any wavelet transform requires a procedure to generate coefficients near the start and end of the signal where the filter overlaps the edges. Here we used a standard procedure of 'reflecting' the signal at the ends, and we followed Milne *et al.* (2009) to remove those coefficients that would bias estimates of the variance and covariance associated with the particular scales.

Results

Figure 2 shows the MRAs for the different forms of carbon. Each stacked line corresponds to the variations of the optical density for a particular form of carbon which corresponds to a scale of the wavelet transform, with the finest scale at the bottom and the smooth component that corresponds to the coarsest scale at the top. The stacked lines are separated for clarity so there is no vertical scale on the graphs, but in each case a vertical grey bar is shown which indicates the magnitude of the vertical variations in units of optical density. If the components are added together they result in the original data. For all forms of carbon there is a general increase in optical density along the transect, which is most pronounced in the case of the carboxyl carbon. These trends are shown in the smooth components which are the top line in each of the multi-resolution analyses in Figure 2. The detail components at finer scales all have a mean of zero, and illustrate fluctuations about the trend at these different scales.

Figure 3 shows the corresponding wavelet variances. All forms of carbon show scale dependence on this transect and the wavelet variance increases with scale for most forms of carbon. The scale dependence is least marked for the aromatic carbon which shows a distinct peak at the 0.2-μm scale, but it also has its largest component of variance at the coarsest scale detail component (3.2 μm). The multi-resolution analysis in Figure 2 shows that this variation is associated with two peaks in optical density about 8 and 13 μm along the transect respectively. The carboxylic carbon shows a more typical scale dependence with the wavelet variance increasing with scale. The smooth component shows a peak in optical density near 8 μm along the transect, and there are associated fluctuations in the density at the 3.2- and 1.6-μm scales. In short, all forms of carbon show a pronounced scale dependence in their spatial variation on this transect, with sources of variation at all the scales considered. This is consistent with our general hypothesis (i) in which the spatial variation of soil carbon has its origins in clumped distribution of the initial organic inputs to the soil and scale-dependent variation in the factors that affect subsequent transformations of the material.

The locations at which significant changes in the wavelet variance were found are shown as vertical black bars on the MRA figures (Figure 2). Note that the variance of each form of carbon changed at one or more spatial scales. In the case of aromatic carbon there is a significant reduction in the variance at the finest scale (0.1 μm) at about 6 μm along the transect. Otherwise the variation of this form of carbon appears to be homogeneous. The variance of aliphatic carbon is spatially uniform, except for scales 0.4 and 0.8 μm where there is an increase in variance near 8.5 μm on the transect, due largely to the large fluctuations in these scales near the peak in the overall trend. Similarly, carboxyl carbon shows a significant increase in variance at the 0.8-μm scale near the pronounced increase in the overall trend of optical

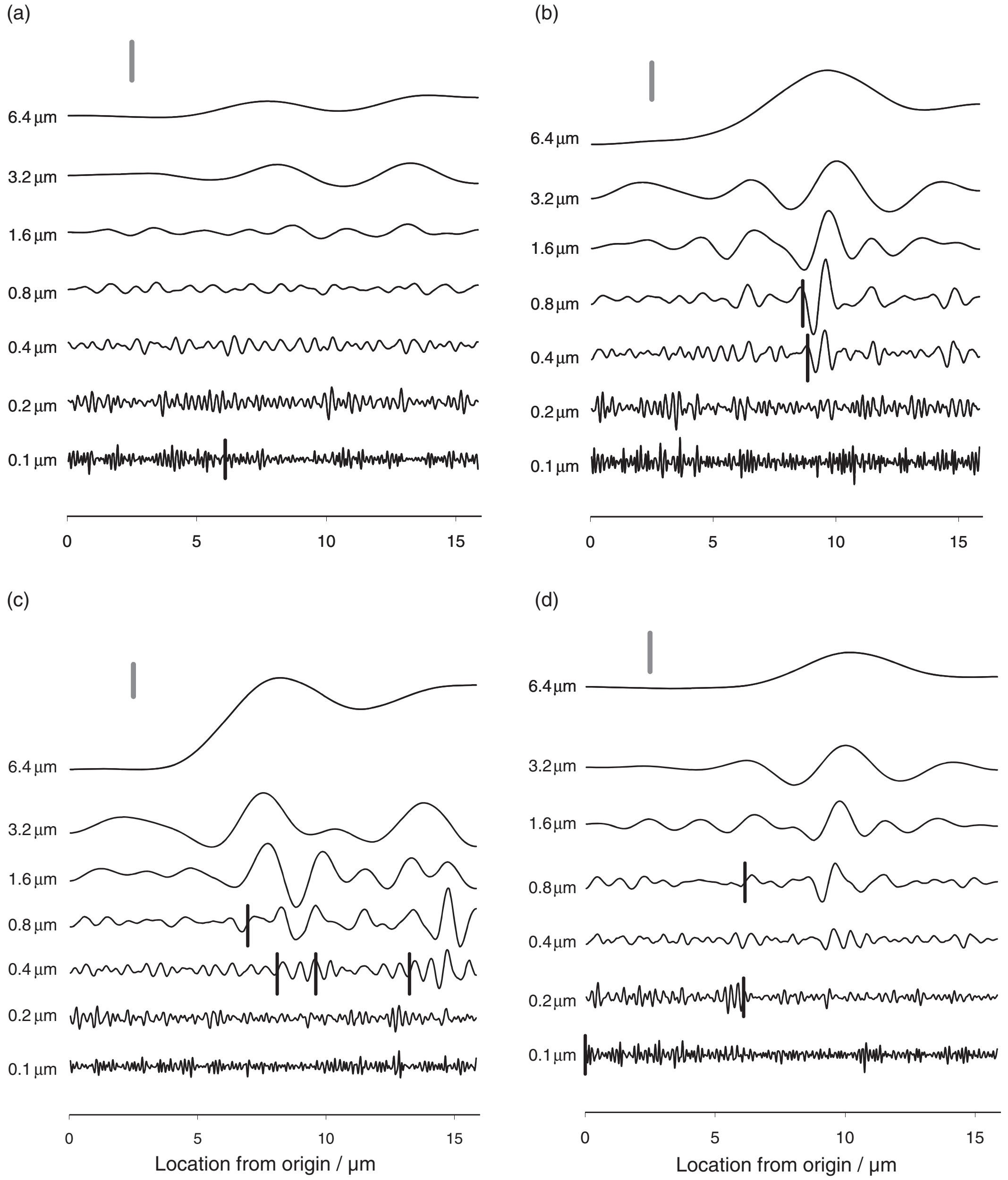

Figure 2 The MRA of optical densities of (a) aromatic carbon (b) aliphatic carbon (c) carboxylic carbon and (d) phenolic carbon. Significant changes in variation are shown by the black vertical lines on the MRA components. The thick grey vertical line represents 0.2 optical density units. Each component is labelled by the lower bound of the corresponding scale interval.

density, and there are two regions of elevated variance at the 0.4-µm scale in this region where the mean density of the carboxyl carbon is larger. There is similarly an increase in the variance of phenolic carbon at the 0.8-µm scale at 6 µm on the transect, and there is a peak in the smooth component near 10 µm on the transect where the larger fluctuations are seen. However, there is a marked reduction in the variance of the 0.2-µm scale component at 6 µm on the transect. This non-uniformity of the variance in forms of carbon at different scales confirms our hypothesis (v) with respect to the variances of single variables. This indicates why the variogram of these variables would not fully account for their spatial variation.

Figure 4 shows the pair-wise correlations by scale among the forms of carbon, and Figure 5 represents the principal components analysis of the wavelet correlation matrices and the correlation matrix of the smooth components. The plot in Figure 5 for the smooth component (scales larger than 6.4 µm) shows that these components are clustered together and all strongly correlated with the first principal component. This indicates that the smooth components show a similar spatial pattern at the coarsest scale,

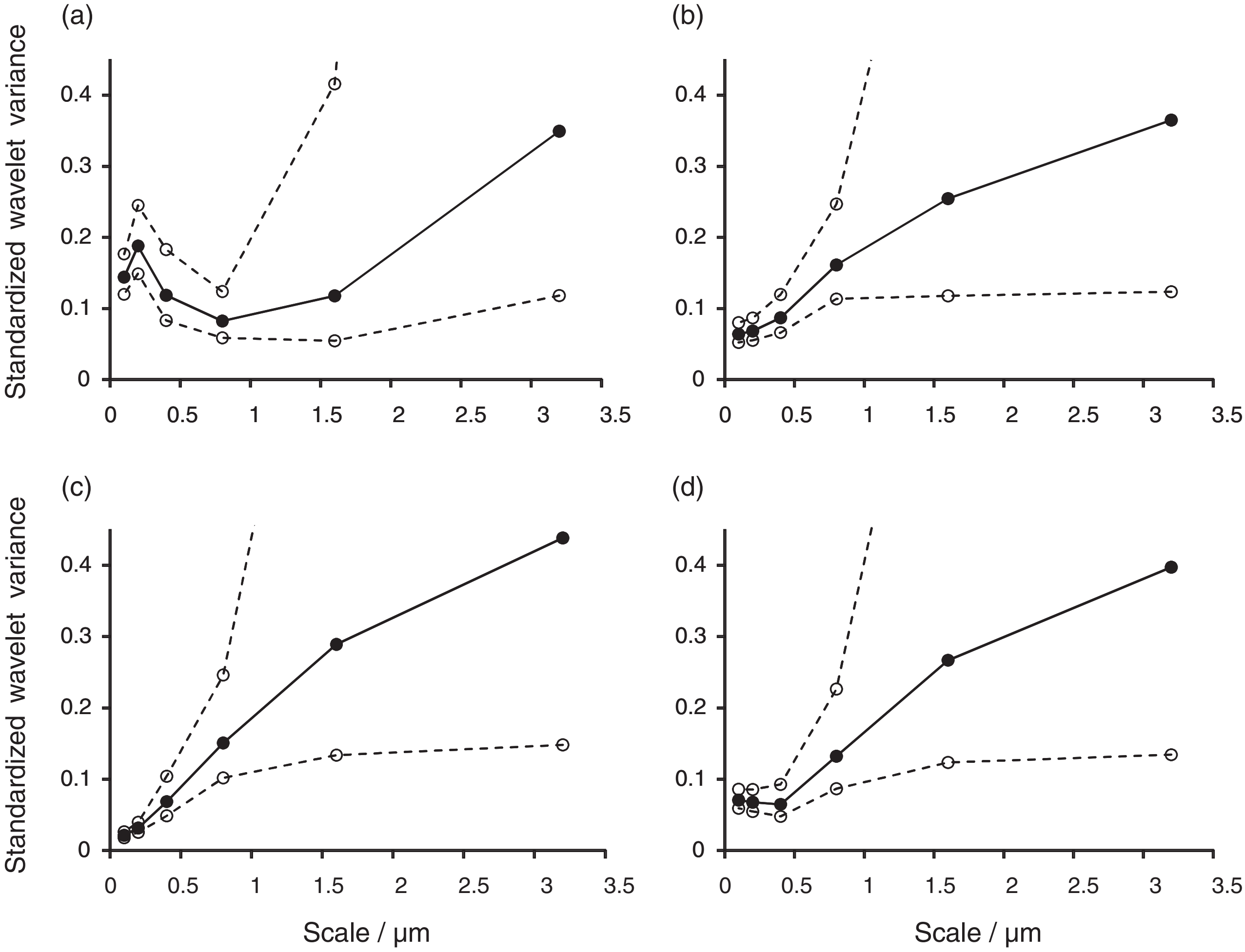

Figure 3 Standardized wavelet variances (wavelet variance / total wavelet variance) by scale for four forms of carbon, with 95% confidence intervals shown by dotted lines: (a) aromatic carbon; (b) aliphatic carbon; (c) carboxylic carbon; (d) phenolic carbon.

which is consistent with our hypothesis (ii), that at coarser scales the soil tends to be locally enriched in all forms of carbon or to have small concentrations of all forms of carbon. This implies that, at least for this micro-aggregate, processes that cause spatial differentiation of the forms of carbon operate at scales finer than 6.4 µm and cannot be understood unless we observe the soil at these scales. At the finer scales there are more complex patterns of variation. Notably the four forms are all furthest apart from each other in the plot in Figure 5 for the finest scale, 0.1 µm. This shows that the components at this scale are rather weakly correlated with each other. The plots show that aliphatic and phenolic carbon are closely related to one another across the scales, although this begins to break down at finer scales. At the two coarsest scales (3.2–6.4, and >6.4 µm) carboxyl carbon is most closely related to aromatic carbon but at finer scales it is more closely related to aliphatic and phenolic carbon. Aromatic carbon is distinctly different from aliphatic and phenolic carbon, always appearing in a different quadrant to these two groups in the plots in Figure 5.

In summary, there are four general patterns of spatially-dependent correlation among the forms of carbon:

(i) There are significant positive correlations at all scales, the strongest at the coarsest scale (e.g. between aliphatic and phenolic carbon), this is consistent with patterns (ii) and (iii) hypothesized above.
(ii) There are weak (<0.5) but significant positive correlations at fine scales, becoming weaker and non-significant as scale increases (such as between carboxylic carbon and aliphatic/phenolic carbon).
(iii) There are (significant) weak negative correlations at fine scales, becoming (non-significant) more strongly negative at coarser scales (such as between aromatic carbon and aliphatic/phenolic). This suggests forms that are not generally found together, possibly because they develop under contrasting conditions.
(iv) There are weak negative (significant) correlations at fine scale and positive (some significant) correlations at coarse

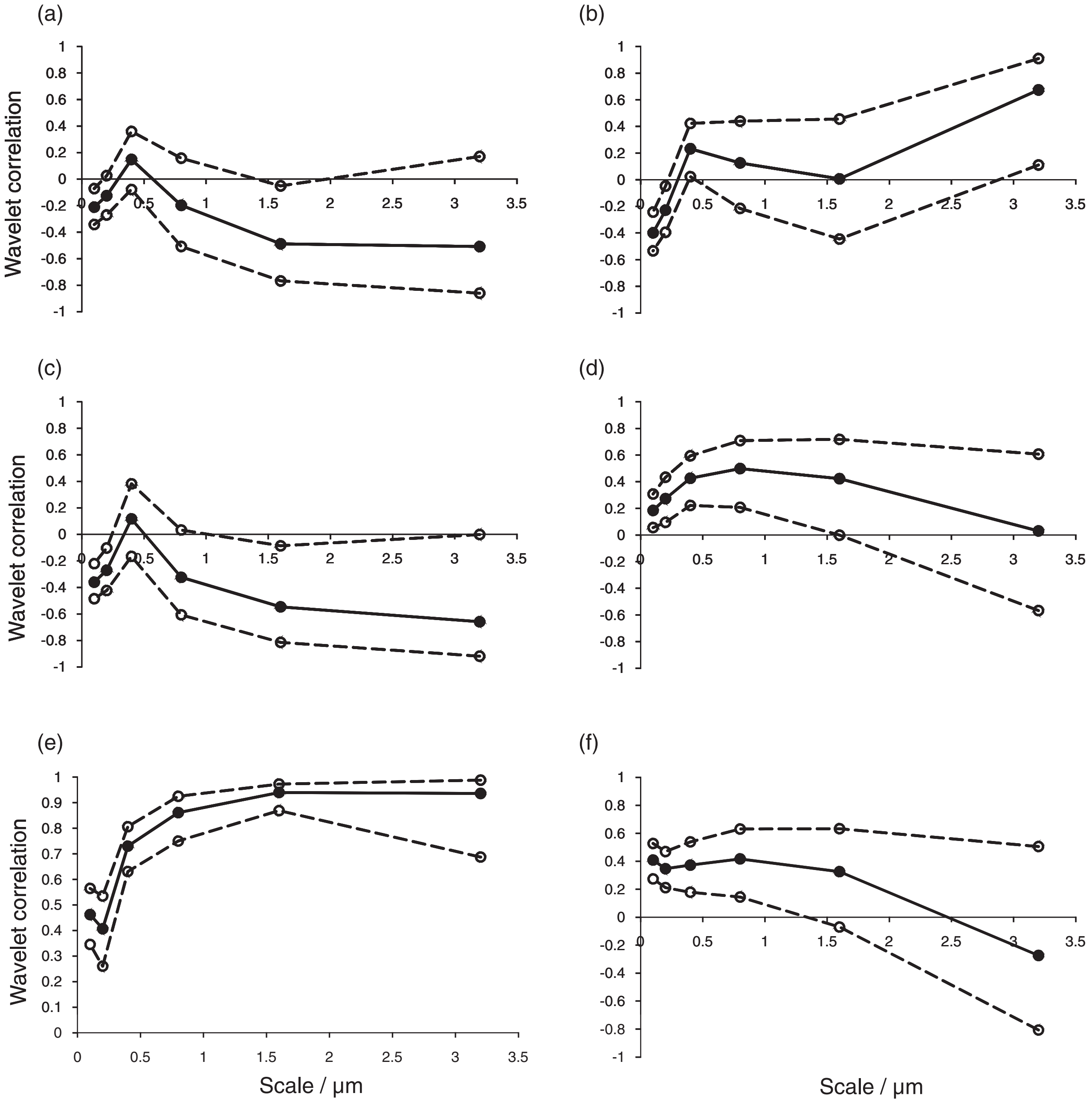

Figure 4 Wavelet correlations between the forms of carbon, with 95% confidence intervals shown by dotted lines: (a) aromatic and aliphatic carbon; (b) aromatic and carboxylic carbon; (c) aromatic and phenolic carbon; (d) aliphatic and carboxylic carbon; (e) aliphatic and phenolic carbon; (f) carboxylic and phenolic carbon.

scales (such as carboxylic and aromatic carbon). This suggests that two forms of carbon may both be found in large clusters (such as filling pores) but when these clusters are examined closely particular forms of carbon are locally dominant. This is consistent with the fourth pattern hypothesized above.

These general patterns of pair-wise scale-dependent correlation among the forms of carbon (Figure 4) and the overall pattern revealed in the principal components analysis on all four forms (Figure 5), exemplify the expected patterns of scale dependent co-variation hypothesized earlier. There are clear scale dependencies in the correlation among all the forms of carbon. The principal components analysis for the smooth component of the MRA shows that there are similar patterns of variation for all forms at the coarsest scales, consistent with pattern (ii) and suggesting that there is a general spatial pattern of enrichment in carbon content which is common to all forms. We also see the hypothesized negative

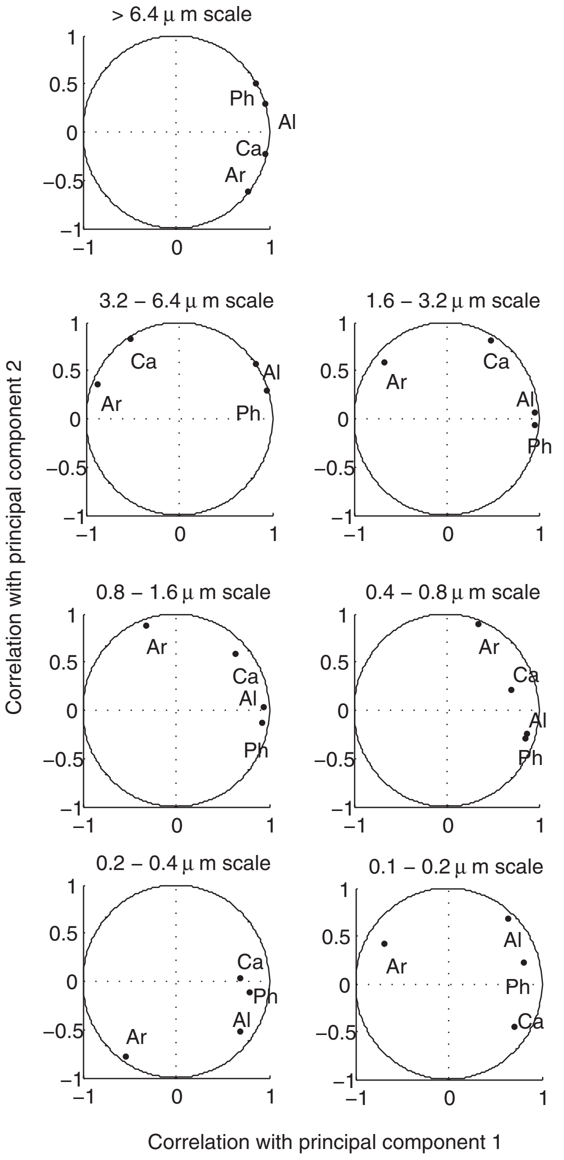

Figure 5 Correlations between the first two principal components of the wavelet correlation matrices (scales 0.1–3.2 μm) and the corresponding detail components for four forms of carbon. The unit circle shows the space of permissible values. In the case of scale > 6.4 μm the principal components were obtained from the smooth components of the MRA.

correlations between some forms at the finest scales (three out of six of the pair-wise correlations in Figure 4). At intermediate scales we see the significant positive correlations between particular forms of carbon proposed in pattern (iii), notably in the correlations between aliphatic and carboxylic carbon, and carboxylic and phenolic carbon, suggesting common factors (original organic material and processes, or both) controlling the distribution of these forms.

As is shown in Figure 6, there is a point about half way along this transect where the correlations of carboxylic carbon with phenolic and aromatic forms change at the 1.6 μm scale. In the first part of this transect carboxylic and aromatic carbon are significantly and positively correlated and in the second half they are not correlated (overall the wavelet correlation of these two forms at this scale is very close to zero). In the first part phenolic and carboxylic carbon are not correlated but they are strongly and significantly (positively) correlated in the second part. This exemplifies co-variation of soil components at micro-scale which is not consistent with assumptions of stationarity; analysis of such data with cross-variograms would fail to identify important aspects of their joint spatial variation.

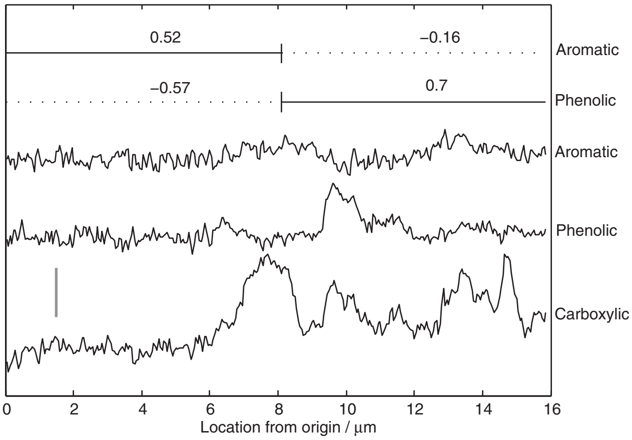

Figure 6 The bottom three lines show stacked plots of the optical densities of carboxylic, aromatic and phenolic carbon along the transect. The top two straight lines illustrate the location of the significant change in wavelet correlation at scale 1.6 μm between aromatic and carboxylic forms (top line) and phenolic and carboxylic forms (second line). The values of wavelet correlation are shown above each line. The dotted line indicates where correlations were not significantly different from zero. The vertical grey bar represents 0.5 optical density units for the three lower graphs.

Discussion

The wavelet analysis of these data on forms of carbon in a single micro-aggregate has enabled us to make more specific and quantified statements about the spatial distribution of carbon in this micro-aggregate than the visual interpretation facilitated by the analyses and figures presented by Lehmann *et al.* (2008). Specifically, we have shown how all forms of carbon exhibit scale-dependent variation along this transect. On the basis of these results we might propose more specific hypotheses which could be tested by applying the wavelet methodology demonstrated here to a larger set of micro-aggregates. For example, the scale dependence of the aromatic carbon is rather different to the other forms. In the other carbon forms, there is an increase in wavelet variance with scale, for aromatic carbon the finest scales (0.1 and 0.2 μm) are relatively more important than the 0.4–1.6 μm scale but the largest components of variation us still seen at the coarsest (3.2 μm) scale. One might propose the hypothesis that this distinctive scale dependence is characteristic of aromatic forms of carbon in soils such as this one which have relatively large inputs of black carbon. This could be tested by undertaking this wavelet analysis on similar transects selected, according to appropriate sampling schemes, on sectioned micro-aggregates sampled from a range of soils with contrasting amounts of black carbon.

The forms of carbon on this transect also exhibited contrasting patterns of scale-dependent correlation, which might be expected to arise from common sources or related processes of transformation in the soil. For example, it has been observed in NEXAFS spectra of cells in microbial biofilms (Lawrence *et al.*, 2003) and of isolated bacteria and fungi (Liang *et al.*, 2006) that carboxylic and aliphatic carbon forms are both characteristic of microbial carbon. In this micro-aggregate, these forms of carbon are correlated significantly but weakly at scales below 2 μm, that is scales corresponding to the size of a typical soil bacterial cell (Guzev & Zvyagintsev, 2003). We might therefore advance the specific hypothesis that a strong correlation of these forms is an indicator of microbial carbon, and test this in further studies on sections with differing microbial biomass. Similarly, we observe that aromatic carbon forms are typical of black carbon and lignin. Carboxylic carbon is also found at the surface of micrometer-sized particles of black carbon due to oxidation. This may explain why these two carbon forms are positively correlated at the coarsest scales, but show contrasting patterns of variation at finer scales. This specific hypothesis could be tested by comparing the wavelet correlations of these carbon forms on sections from soils with contrasting contents of black carbon. Such hypotheses could not be tested without the quantitative methods demonstrated in this paper.

The same approach might be used in future work to examine hypotheses about the interaction of organic and mineral fractions of the soil. There is evidence that carboxylic carbon is associated with clay surfaces (Oades, 1988), and the hypothesis that such interactions determine the distribution of carboxylic carbon at fine scales could be tested by wavelet correlation of NEXAFS-derived optical densities for mineral elements such as Fe, Al and Si in addition to forms of carbon (Lehmann & Solomon, 2010).

The change in the correlation of carboxylic carbon with phenolic and aromatic forms shown in Figure 6 is consistent with our expectation that the spatial variation of soils at fine scales will not, in general, be consistent with assumptions of stationarity. More specific hypotheses of this form could be advanced. If organo-mineral interactions are as important as proposed by Kleber *et al.* (2007), then changes in correlation between forms of carbon might be associated with transitions from void to mineral regions of the soil which NEXAFS data on key mineral elements would show.

This paper has revealed the complexity of multi-scale co-variation of forms of carbon at micro-scales in one section from one micro-aggregate. It shows the potential of the approach, and suggests that appropriate spatial analyses of such data might help to explain the origins and properties of soil organic carbon. However, there is clearly a substantial challenge if we are to use information on soil variation at micro-scales to explain soil behaviour at the plot, field or landscape scales at which practical questions about management of soil carbon and its impact on other environmental systems must be addressed. We propose that the key to this problem is to develop suitable sampling schemes by which the measurements on micro-scale-support can be analysed over a range of scales from micro-aggregate to landscape. This might be achieved by the combination of wavelet analyses from multiple sets of sections to identify different patterns of micro-scale variation in soil carbon at micro-aggregate scale, and then to extend these to the coarser scales. Another approach would be to use spatially nested sampling (Lark, 2005). This is a procedure to estimate components of variance or covariance that are associated with different spatial scales. Nested sampling could be undertaken on sections that they were obtained by nested sampling from micro-aggregates. These micro-aggregates could be selected by spatially nested sampling from within sample cores. The cores could be collected on a sampling scheme spatially nested on scales up to the catchment or beyond. This would provide a basis for a unified analysis of the variation of forms of carbon over the scale range from landscape to nanometres. Using the extensions of nested sampling and analysis proposed by Lark (2005), the scale-dependent correlations that the current paper has demonstrated could be analysed over the full scale range, and by following Lark & Corstanje (2009) one could undertake such analyses without assuming stationarity of the variance components. This analysis would allow direct tests of some of the exemplar specific hypotheses advanced in this discussion, allowing us, for example, to compare the co-variation of particular carbon forms in soils which have received contrasting inputs of black carbon.

Trumbore & Czimczik (2008) call for more process-level insight into soil carbon and its fate. One key question is what confers stability on soil organic carbon. Whether the answer is pore-filling or organo-mineral interactions or both can be addressed only by observing spatial organization at the sub-micrometre scale (Lehmann *et al.*, 2007). We have shown that this organization can be characterized by wavelet analysis and would caution against the use of geostatistical models of spatial variation which require simpler assumptions about spatial variation.

Finally, we observe that NEXAFS technology can be used to examine constituents of the soil other than carbon. We have already referred to mineral components that might interact with carbon, but the distribution of metals in soil, and their co-variation with other components could contribute to our understanding of the behaviour of key pollutants, as Jacobson *et al.* (2007) illustrated. Wavelet analysis might be similarly useful to study these variables. Indeed as we try to understand the joint variation of larger numbers of variables the need for quantitative methods such as those developed and demonstrated here will become more pressing.

Conclusions

Wavelet analysis of micro-scale data on forms of carbon allowed quantitative conclusions about their joint variation at different scales which go beyond the previous visual interpretation of Lehmann *et al.* (2008). In particular, the wavelet analysis allows quantitative statements to be made about scale-dependent variation and identifies patterns of scale-dependent joint variation of particular variables in the soil, again at specific and quantitatively defined spatial scales. This is essential for further work if various soils are to be compared rigorously. Our expectations of scale-dependent variance and co-variances of the forms of carbon in a micro-aggregate and evidence that this cannot be assumed to reflect an underlying stationary process were confirmed. In particular, we have noticed patterns of correlation consistent with broad contrasts in the origins of soil carbon, local aggregations of carbon with characteristic composition and forms of carbon which show different distributions at fine scales (smaller than microbial cells). While our particular results refer only to one small micro-aggregate they show that the spatial co-variation of forms of soil carbon at these scales exhibits a complexity which cannot generally be accounted for by simple summary statistics or the geostatistical methods that have been applied to other micro-scale data sets. Wavelet analysis provides a basis to test hypotheses about micro-scale soil variation in a rigorous way.

Acknowledgements

The contribution to this work by AEM and RML was undertaken as part of Rothamsted Research's programme in Mathematical and Computational Biology which is funded by the United Kingdom Biotechnology and Biological Sciences Research Council (BBSRC). D.S. and J.L. acknowledge funding from the Agriculture and Food Research Initiative Competitive Grant of the National Institute of Food and Agriculture (2008-35107-04511), and from the Geobiology and Low-Temperature Geochemistry Program of the National Science Foundation (EAR-0819689). We are grateful to Dr Saran Sohi, Professor Phil Brookes, Dr Simona Hapca and Dr Wilfred Otten for helpful discussions of this work and comments on the manuscript.

References

Coppin, F., Chabroullet, C. & Martin-Garin, A. 2009. Selenite interactions with some particulate organic and mineral fractions isolated from a natural grassland soil. *European Journal of Soil Science*, **60**, 369–376.

De Rito, C.M., Pumphrey, G.M. & Madsen, E.L. 2005. Use of field-based stable isotope probing to identify adapted populations and track carbon flow through a phenol-degrading soil microbial community. *Applied and Environmental Microbiology*, **71**, 7858–7865.

Feeney, D.S., Crawford, J.W., Daniell, T., Hallett, P.D., Nunan, N., Ritz, K. *et al*. 2006. Three-dimensional microorganization of the soil-root-microbe system. *Microbial Ecology*, **52**, 151–158.

Fierer, N., Grandy, A.S., Six, J. & Paul, E.A. 2009. Searching for unifying principles in soil ecology. *Soil Biology & Biochemistry*, **41**, 2249–2256.

Guzev, V.S. & Zvyagintsev, D.G. 2003. The biometric analysis of bacterial cells in soil. *Microbiology*, **72**, 187–192.

Isaure, M.P., Manceau, A., Geoffroy, N., Laboudigue, A., Tamura, N. & Marcus, M.A. 2005. Zinc mobility and speciation in soil covered by contaminated dredged sediment using micrometer-scale and bulk-averaging X-ray fluorescence, absorption and diffraction techniques. *Geochimica et Cosmochimica Acta*, **69**, 1173–1198.

Jacobson, A.R., Dousset, S., Andreux, F. & Baveye, P.C. 2007. Electron microprobe and synchrotron x-ray fluorescence mapping of the heterogeneous distribution of copper in high-copper vineyard soils. *Environmental Science and Technology*, **41**, 6343–6349.

Kleber, M., Sollins, P. & Surron, R. 2007. A conceptual model of organo-mineral interactions in soils: self-assembly of organic molecular fragments into zonal structures on mineral surfaces. *Biogeochemistry*, **85**, 9–24.

Lark, R.M. 2005. Exploring scale-dependent correlation of soil properties by nested sampling. *European Journal of Soil Science*, **56**, 307–317.

Lark, R.M. 2006. Analysis of complex soil variation using wavelets. In: *Environmental Soil Landscape Modeling* (ed S. Grunwald), pp. 343–371. CRC Press, New York.

Lark, R.M. 2010. Two contrasting spatial processes with a common variogram: inference about spatial models from higher-order statistics. *European Journal of Soil Science*, **61**, 479–492.

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.

Lark, R.M. & Corstanje, R. 2009. Non-homogeneity of variance components from spatially nested sampling of the soil. *European Journal of Soil Science*, **60**, 443–452.

Lark, R.M., Milne, A.E., Addiscott, T.M., Goulding, K.W.T, Webster, C.P. & O'Flaherty, S. 2004. Scale- and location-dependent correlation of nitrous oxide emissions with soil properties: an analysis using wavelets. *European Journal of Soil Science*, **55**, 601–610.

Lawrence, J.R., Swerhone, G.D., Leppard, G.G., Araki, T., Zhang, X., West, M.M. *et al*. 2003. Scanning transmission x-ray, laser scanning, and transmission electron microscopy mapping of the exopolymeric matrix of microbial biofilms. *Applied Environmental Microbiology*, **69**, 5543–5444.

Lehmann, J. & Solomon, D. 2010. Organic carbon chemistry in soils observed by synchrotron-based spectroscopy. In: *Synchroton-based Techniques in Soils and Sediment* (eds B. Singh & M. Gräfe), pp. 289–312. Elsevier, Amsterdam.

Lehmann, J., Kinyangi, J. & Solomon, D. 2007. Organic matter stabilization in soil micro-aggregates: implications from spatial heterogeneity of organic carbon contents and carbon forms. *Biogeochemistry*, **85**, 45–57.

Lehmann, J., Solomon, D., Kinyangi, J., Dathe, L., Wirick, S. & Jacobsen, C. 2008. Spatial complexity of soil organic matter forms at nanometre scales. *Nature Geoscience*, **1**, 238–242.

Liang, B., Lehmann, J., Solomon, D., Kinyangi, J., Grossman, J., O'Neill, B. *et al*. 2006. Black carbon increases cation exchange capacity in soils. *Soil Science Society of America Journal*, **70**, 1719–1730.

Milne, A.E., Macleod, C.J.A., Haygarth, P.M., Hawkins, J.M.B. & Lark, R.M. 2009. The Wavelet Packet transform: a technique for investigating temporal variation of river water solutes. *Journal of Hydrology*, **379**, 1–19.

Nguyen, B., Lehmann, J., Kinyangi, J., Smernik, R., Riha, S.J. & Engelhard, M.H. 2008. Long-term black carbon dynamics in cultivated soil. *Biogeochemistry*, **89**, 295–308.

Nunan, N., Wu, K.J., Young, I.M., Crawford, J.W. & Ritz, K. 2003. Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. *FEMS Microbiology Ecology*, **44**, 203–215.

Nunan, N., Ritz, K., Rivers, M., Feeney, D.S. & Young, I.M. 2006. Investigating microbial micro-habitat structure using X-ray computed tomography. *Geoderma*, **133**, 398–407.

Oades, J.M. 1988. The retention of organic matter in soils. *Biogeochemistry*, **5**, 35–70.

Percival, D.B. 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: *Wavelets in Geophysics* (eds E. Foufoula-Georgiou & P. Kumar), pp. 325–344. Academic Press, New York.

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

Trumbore, S. & Czimczik, C. 2008. An uncertain future for soil carbon. *Science*, **321**, 1455–1456.

Webster, R. 2001. Statistics to support soil research and their presentation. *Journal of Soil Science*, **52**, 331–340.

Whitcher, B.J., Guttorp, P. & Percival, D.B. 2000. Multiscale detection and location of multiple variance changes in the presence of long memory. *Journal of Statistical Computation and Simulation*, **68**, 65–88.

Young, I.M. & Crawford, J.W. 2004. Interactions and self-organization in the soil–microbe complex. *Science*, **304**, 1634–1637.

Zhang, X.X., Deeks, L.K., Bengough, A.G., Crawford, J.W. & Young, I.M. 2005. Determination of soil hydraulic conductivity with the lattice Boltzmann method and soil thin-section technique. *Journal of Hydrology*, **306**, 59–70.