Multiscale Entropy-based Analysis of Soil Transect Data

A. M. Tarquis, N. R. A. Bird, A. P. Whitmore, M. C. Cartagena, and Yakov Pachepsky

A deeper understanding of the spatial variability of soil properties and the relationships between them is needed to scale up measured soil properties and to model soil processes. The object of this study was to describe the spatial scaling properties of a set of soil physical properties measured on a common 1024-m transect across arable fields at Silsoe in Bedfordshire, east-central England. Properties studied were volumetric water content (θ), total porosity (ϕ), pH, and N_2O flux. We applied entropy as a means of quantifying the scaling behavior of each transect. Finally, we examined the spatial intrascaling behavior of the correlations between θ and the other soil variables. Relative entropies and increments in relative entropy calculated for θ, ϕ, and pH showed maximum structure at the 128-m scale, while N_2O flux presented a more complex scale dependency at large and small scales. The intrascale-dependent correlation between θ and ϕ was negative at small scales up to 8 m. The rest of the intrascale-dependent correlation functions between θ with N_2O fluxes and pH were in agreement with previous studies. These techniques allow research on scale effects localized in scale and provide the information that is complementary to the information about scale dependencies found across a range of scales.

THE SPATIAl variability of soil properties and sediments is due to the combined action of physical, chemical, and biological processes that operate with different intensities and at different scales (Goovaerts, 1998; Bruland and Richardson, 2005). The significance of this variability has led scientists and practitioners to the realization of the need to quantify it. Statistics of soil or sediment properties have become essential components of data collection in vadose zone research (Hupet et al., 2004; Dyck et al., 2005; Pringle and Lark, 2006; Vereecken et al., 2007). The accumulation of such statistics has eventually led to the understanding that they change with the scale of sampling or description. Many data on soil and sediments are obtained from small samples and cores, monoliths, or small field plots, yet the goal is to reconstruct soil properties across fields, watersheds, and landforms, or to predict physical properties of pore surfaces and the structure of the pore space. The representation of processes and properties at a scale different from the one at which observations and property measurements are made is a pervasive problem in vadose zone hydrology.

Recently, fractal geometry has become an important source of scaling laws in soil hydrology. Fractal geometry focuses on geometric objects in which total length, area, or volume depends on the scale. Such objects exhibit similar geometric shapes when observations are made at different scales. They were termed fractals by Mandelbrot (1982), who suggested that fractals rather than regular geometric shapes like segments, arcs, circles, spheres, etc., are more appropriate to approximate irregular natural shapes that have hierarchies of ever-finer detail. This observation marked the beginning of the application of fractal geometry, which has become very popular during last 20 yr because of its promise to relate features of natural objects observed at different scales (Gimenez et al., 1997).

Fractal geometry characterizes and parameterizes scaling relationships across a range of scales. In theory, the wider the range of scales, the more reliable are the scaling parameters such as fractal dimensions or multifractal structure function. Depending on the application, the change in variability with scale may also be of interest for the cases in which changes in scale are not large. Fractal models are not meant for this type of analysis, and other tools of multiscale analysis have to be used. Ideally, they should allow one to parameterize the joint effect of small changes in

A.M. Tarquis, Dpto. de Matemática Aplicada a la Ingeniería Agronómica. E.T.S. Ingenieros Agrónomos, U.P.M., Madrid, Spain; N.R.A. Bird and A.P. Whitmore, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK; M.C. Cartagena, Dpto. de Química Agrícola. E.T.S. Ingenieros Agrónomos, U.P.M., Madrid, Spain; Ya. Pachepsky, USDA-ARS, Environmental Microbial Safety Lab., Beltsville, MD 20705. Received 22 Feb. 2007. Corresponding author (anamaria.tarquis@upm.es).

location and scale on variability. A search of such methods is currently underway.

It has been shown (Basseville et al., 1992; Kumar, 2003) that multiscale signal and image analysis of dyadic trees of scales can be used for the fusion of multiresolution data, downscaling, and efficient reconstruction of missing data, in particular soil moisture contents. Many analyses of the spatial structure of soil properties have been based on spatial cross-correlograms (Goovaerts, 1997), obtaining several parameters to estimate the significant spatial correlation between two variables (Kravchenko et al., 2002, 2003). This type of analysis is important when one of the variables is difficult or expensive to measure. Wavelet-based multiscale analysis has been successfully applied to analyze soil structure, salinity, and other soil properties (Lark et al., 2003, 2004; Zeleke and Si, 2005; Ding and Ding, 2006). Watershed analysis is yet another technique to perform multiscale analysis within a narrow range of scales (Sofou et al., 2001). The relative efficiency of these and other methods depends on the intended application.

The objective of this work was to propose and apply two simple parameters to document scale-dependent changes in spatial variability and to test methods to find these parameters with data on soil properties along a transect. We used two such parameters—relative entropy and the intrascale correlation coefficient.

Multiscale Analysis

Our analysis was based on the scaling behavior of coarse-grained measures derived from data distributed on a geometric scale. In the context of this study, we had a set of positive soil property values x_i, sampled at equal intervals across a transect. A coarse-grained measure is defined by (Feder, 1989)

$$\mu_i(\delta) = \frac{\sum_{j=(i-1)\delta+1}^{i\delta} x_j}{\sum_{j=1}^{L} x_j} \qquad [1]$$

for values of $\delta = 2^k$ where $k = 1, \ldots, n$, and $L = 2^n$ is the total number of sampling points in the transect. Thus the measure $\mu_i(\delta)$ is created by placing a partition mesh of size δ on the transect and aggregating the values within each partition cell i.

The measure thus created forms the basis of multifractal analysis, used to characterize data sets when scaling symmetries are present in the data (Bird et al., 2006). Entropy (S) is defined as

$$S(\delta) = -\sum_{i=1}^{n(\delta)} \mu_i(\delta) \ln\left[\mu_i(\delta)\right] \qquad [2]$$

where $n(\delta)$ is the number of intervals in the transect of length δ and $S(\delta)$ is the entropy at scale or resolution δ. This is one of many resolution-dependent quantifications of heterogeneity that arise in multifractal analysis.

While this form of analysis is usually used to identify simple logarithmic scaling behavior, it has equal merit when such behavior is absent. Entropy evaluated at different resolutions then reveals the scale-dependent nature of heterogeneity in the data.

To be well placed to detect structure, especially when this is not pronounced, we used relative entropy, which is entropy expressed relative to that arising from a uniform and structureless measure (Bird et al., 2006). This relative entropy (E) is given by

$$E(\delta) = -\sum_i \mu_i(\delta) \log \mu_i(\delta) + \log \frac{\delta}{L} \qquad [3]$$

Entropy as a Measure of Multscale Heterogeneity

Plotting relative entropy against the resolution of observation δ reveals how structure in the measure evolves with increasing resolution, and by calculating increments, we may quantify structure at successively smaller scales.

Moving from scale or resolution 2δ to the finer resolution δ, a value of the measure $\mu_i(2\delta)$ is resolved into two adjacent values, $\mu_{2i-1}(\delta) = p_{i,1}\mu_i(2\delta)$ and $\mu_{2i}(\delta) = p_{i,2}\mu_i(2\delta)$, where both represent the distribution of the initial measure, $\mu_i(2\delta)$, in two intervals $p_{i,1}$ and $p_{i,2}$, the two percentages in which $\mu_i(2\delta)$ is dividing ($p_{i,1} + p_{i,2} = 1$). Thus we may rewrite Eq. [3] as

$$E(\delta) = -\sum_{i=1}^{L/\delta} \sum_{j=1}^{2} p_{i,j}\mu_i(2\delta) \log p_{i,j}\mu_i(2\delta) + \log \frac{\delta}{L} \qquad [4]$$

This, in turn, may be rewritten as

$$E(\delta) = E(2\delta) - \sum_{i=1}^{L/\delta} \left[\mu_i(2\delta) \sum_{j=1}^{2} p_{i,j} \log p_{i,j}\right] - \log 2 \qquad [5]$$

The incremental change in relative entropy then becomes

$$\Delta E(\delta) = E(2\delta) - E(\delta) = \sum_i \left[\mu_i(2\delta) \sum_{j=1}^{2} p_{i,j} \log p_{i,j}\right] + \log 2 \qquad [6]$$

The increment of E now describes the structure revealed in the data at scale δ. In particular, we have a sum of local entropies weighted by $\mu_i(2\delta)$, describing structure at scale δ revealed in a window of observation of size 2δ. Maximum structure, corresponding to $p_{i,1} = 1$, $\forall i$ yields $\Delta E(\delta) = \log 2$. No structure (local uniformity), corresponding to $p_{i,1} = p_{i,2} = 0.5$, $\forall i$ yields $\Delta E(\delta) = 0$. Thus we have

$$0 \le \Delta E(\delta) \le \log 2 \qquad [7]$$

and a record of $\Delta E(\delta)$ across scales δ provides a succinct record of scale-dependent structure within the original transect measure.

A special case occurs when $p_{i,1}$ and $p_{i,2}$ are independent of both i and δ. Then, from Eq. [6],

$$\Delta E(\delta) = \left(p_1 \log p_1 + p_2 \log p_2\right) + \log 2 \qquad [8]$$

is constant. This corresponds to a multifractal measure generated by a multicascade model (see Fig. 1), and entropy and relative entropy scale logarithmically as

$$S(\delta) = -D\log(\delta/L) \quad [9]$$

$$E(\delta) = (1-D)\log(\delta/L) \quad [10]$$

From Eq. [9], we can define D as the slope of entropy against δ, which is given here by

$$D = -\frac{(p_1 \log p_1 + p_2 \log p_2)}{\log 2} \quad [11]$$

Bivariate Analysis to Detect Scale-Dependent Correlations

Our bivariate analysis also exploits the scaling properties of coarse-grained measures, but now we cannot use entropy because this only provides a quantification of the degree of structure in the data. We consider two measures, $\mu_i(\delta)$ and $\beta_i(\delta)$, where $\beta_i(\delta)$ can be defined, based on a set of positive soil property values y_i, as

$$\beta_i(\delta) = \frac{\sum_{j=(i-1)\delta+1}^{i\delta} y_j}{\sum_{j=1}^{L} y_j} \quad [12]$$

First, we consider the measure $\mu_i(2\delta)$ at resolution 2δ. As discussed above, at resolution δ, each measure is resolved into two components, $p_{i,1}\mu_i(2\delta)$ and $p_{i,2}\mu_i(2\delta)$, where $p_{i,1} + p_{i,2} = 1$. Similarly for the second measure, we resolve $\beta_i(2\delta)$ into $q_{i,1}\beta_i(2\delta)$ and $q_{i,2}\beta_i(2\delta)$, where $p_{i,1}$ and $p_{i,2}$ represent the distribution of the initial measure, $\beta_i(2\delta)$, into two subintervals, and $\beta_{i,1} + \beta_{i,2} = 1$.

We now construct a sum of local covariances between the two measures, each covariance corresponding to a window of observation 2δ viewed with resolution δ:

$$C(\delta) = \sum_{i=1}^{L/2\delta}\sum_{j=1}^{2}\left[p_{i,j}\mu_i(2\delta) - \frac{\mu_i(2\delta)}{2}\right]\left[q_{i,j}\beta_i(2\delta) - \frac{\beta_i(2\delta)}{2}\right] \quad [13]$$

This simplifies to

$$C(\delta) = \sum_{i=1}^{L/2\delta}\mu_i(2\delta)\beta_i(2\delta)\left(p_{i,1}q_{i,1} + p_{i,2}q_{i,2} - \frac{1}{2}\right) \quad [14]$$

which may be written as

$$C(\delta) = \sum_{i=1}^{L/2\delta}\left[p_{i,1}\mu_i(2\delta)\right]\left[q_{i,1}\beta_i(2\delta)\right] + \left[p_{i,2}\mu_i(2\delta)\right]\left[q_{i,2}\beta_i(2\delta)\right] - \frac{1}{2}\sum_{i=1}^{L/2\delta}\mu_i(2\delta)\beta_i(2\delta) \quad [15a]$$

or

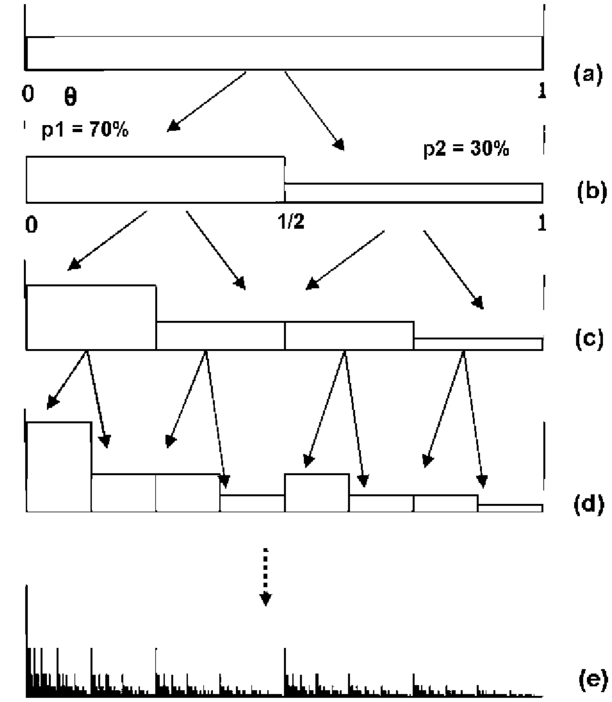

FIG. 1. Multiscale cascade model where p_1 and p_2, the two percentages into which the measure is divided ($p_1 + p_2 = 1$), are independent of the scale (δ).

$$C(\delta) = \sum_{i=1}^{L/\delta}\mu_i(\delta)\beta_i(\delta) - \frac{1}{2}\sum_{i=1}^{L/2\delta}\mu_i(2\delta)\beta_i(2\delta) \quad [15b]$$

Defining the following functions

$$X_{\mu\beta}(\delta) \equiv \sum_i \mu_i(\delta)\beta_i(\delta) \quad [16a]$$

$$X_{\mu\mu}(\delta) \equiv \sum_i \mu_i(\delta)\mu_i(\delta) \quad [16b]$$

$$X_{\beta\beta}(\delta) \equiv \sum_i \beta_i(\delta)\beta_i(\delta) \quad [16c]$$

we finally write

$$C(\delta) = X_{\mu\beta}(\delta) - \frac{1}{2}X_{\mu\beta}(2\delta) \quad [17]$$

From this we define the intrascale-dependent correlation function:

$$R(\delta) = \frac{2X_{\mu\beta}(\delta) - X_{\mu\beta}(2\delta)}{\sqrt{\left[2X_{\mu\mu}(\delta) - X_{\mu\mu}(2\delta)\right]\left[2X_{\beta\beta}(\delta) - X_{\beta\beta}(2\delta)\right]}} \quad [18]$$

This function provides us with a way of recording correlations at different scales δ, based on coarse graining the measure. This function is equivalent to a Haar wavelet correlation function, which arises from the simplest form of wavelet analysis using the Haar wavelet function (see, e.g., Percival and Walden, 2000).

Equation [18] seems to be quite close in form to the cross-correlogram function; however, a cross-correlogram will define the correlation existing between $\mu_i(\delta)$ and $\beta_i(\delta)$ values or between $\mu_i(2\delta)$ and $\beta_i(2\delta)$ values separated by a lag distance δ, 2δ, 3δ, ..., etc., but not the correlation at different scales δ.

We applied these analyses to a set of transect data recording intrascale-dependent variation of soil properties.

Materials And Methods

Case Study

The data used here were collected in a survey on a transect across arable fields at Silsoe in Bedfordshire, east-central England. The data have previously been described by Lark et al. (2004). The first sample point on the transect was at UK Ordnance Survey (OS) coordinates 508570, 235605, and the soil was sampled at 256 locations at 4-m intervals on a line running on a bearing of 188° relative to UK OS grid north. The data selected from this survey for analysis here were porosity (ϕ), volumetric water content (θ), pH, and N_2O flux.

The values of all these variables are shown in Fig. 2. The mean, standard deviation, and asymmetry (skewness) of the four variables (measures) are described in Table 1.

Relative Entropy and Bivariate Analysis

Relative entropy and increments in relative entropy were calculated using Eq. [4] and [6], respectively, for each soil variable.

Entropy and relative entropy were calculated by selecting the first point of the transect as the origin for the partition mesh that was used to coarse grain the transect data. Other origins could be chosen, yielding different values for entropy, but thus would require an assumption of spatial periodicity beyond the endpoints of the transect to allow the partition mesh to extend beyond these endpoints.

Values for intrascale correlation $R(\delta)$ were calculated using Eq. [18], using the combinations of θ with ϕ, pH, and N_2O. Confidence limits for this $R(\delta)$ were computed using Fisher's z transforms (Piegorsch and Bailer, 2005):

$$z(\delta) = 0.5L\frac{1+R(\delta)}{1-R(\delta)} \qquad [19]$$

The transformed estimate of $R(\delta)$ is approximately normal, with a sample variance of $1/(n-3)$, where the correlation is derived from n independent observations. Therefore, the confidence limits were calculated as usual on $z(\delta)$ for $\alpha = 0.05$ and then transformed into $R(\delta)$ limits by

$$R(\delta) = \frac{\exp[2z(\delta)]-1}{\exp[2z(\delta)]+1} \qquad [20]$$

In our case, we followed the work of Whitcher (1998), considering that the number of independent observations was

$$n = \frac{N}{2^j} \qquad [21]$$

with $j = 1, 2, ..., 6$ and $N = 256$.

Table 1. Statistical description of the variables studied: volumetric water content (θ), total porosity (ϕ), pH, and N_2O flux.

Measure	θ	ϕ	pH	N_2O flux
	$m^3\ m^{-3}$			$\mu g\ kg^{-1}\ d^{-1}$
Average	0.38	0.57	7.20	54.61
SD	0.05	0.06	1.00	54.52
Min.	0.23	0.36	5.20	0.00
Max.	0.54	0.71	8.30	319.00
Skewness	0.04	−0.85		1.59

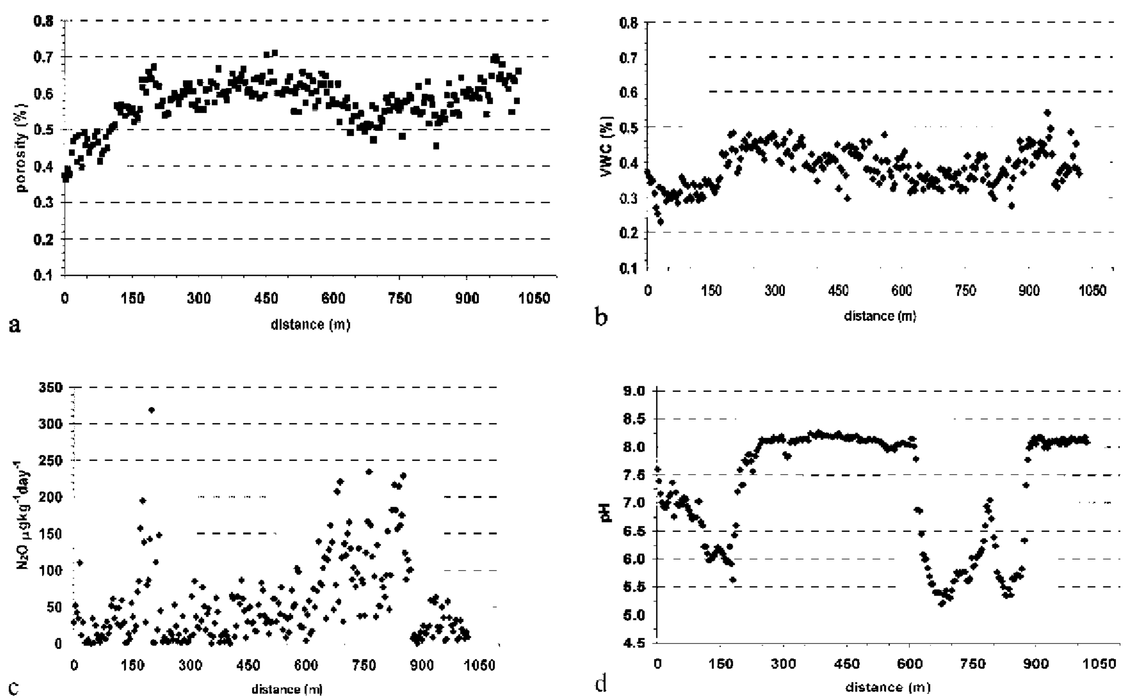

Fig. 2. Original data of the soil variables: (a) total porosity, (b) volumetric water content (VWC), (c) N2O flux, and (d) pH.

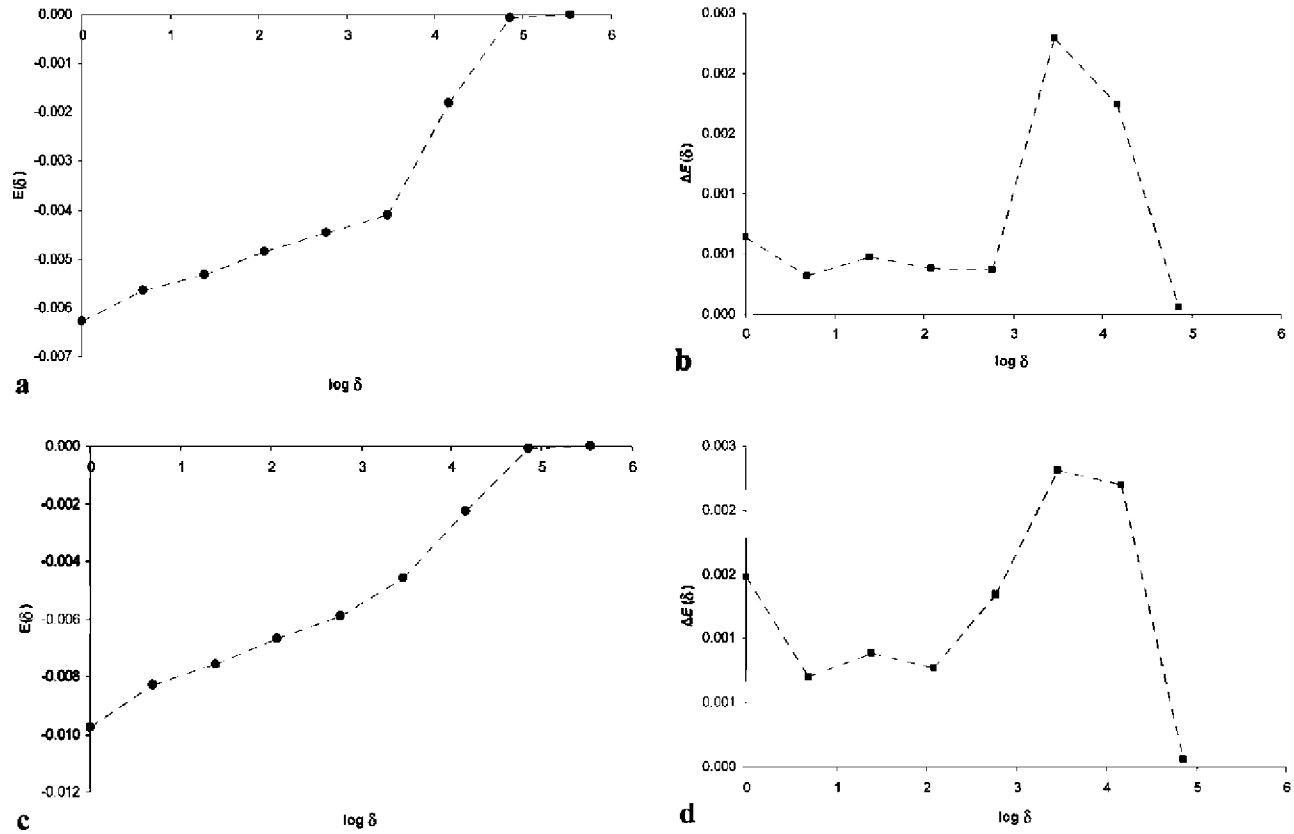

FIG. 3. Entropy study: (a) relative entropy, $E(\delta)$, of porosity, (b) increment of relative entropy, $\Delta E(\delta)$, of porosity, (c) relative entropy of volumetric water content, and (d) increment of relative entropy of volumetric water content.

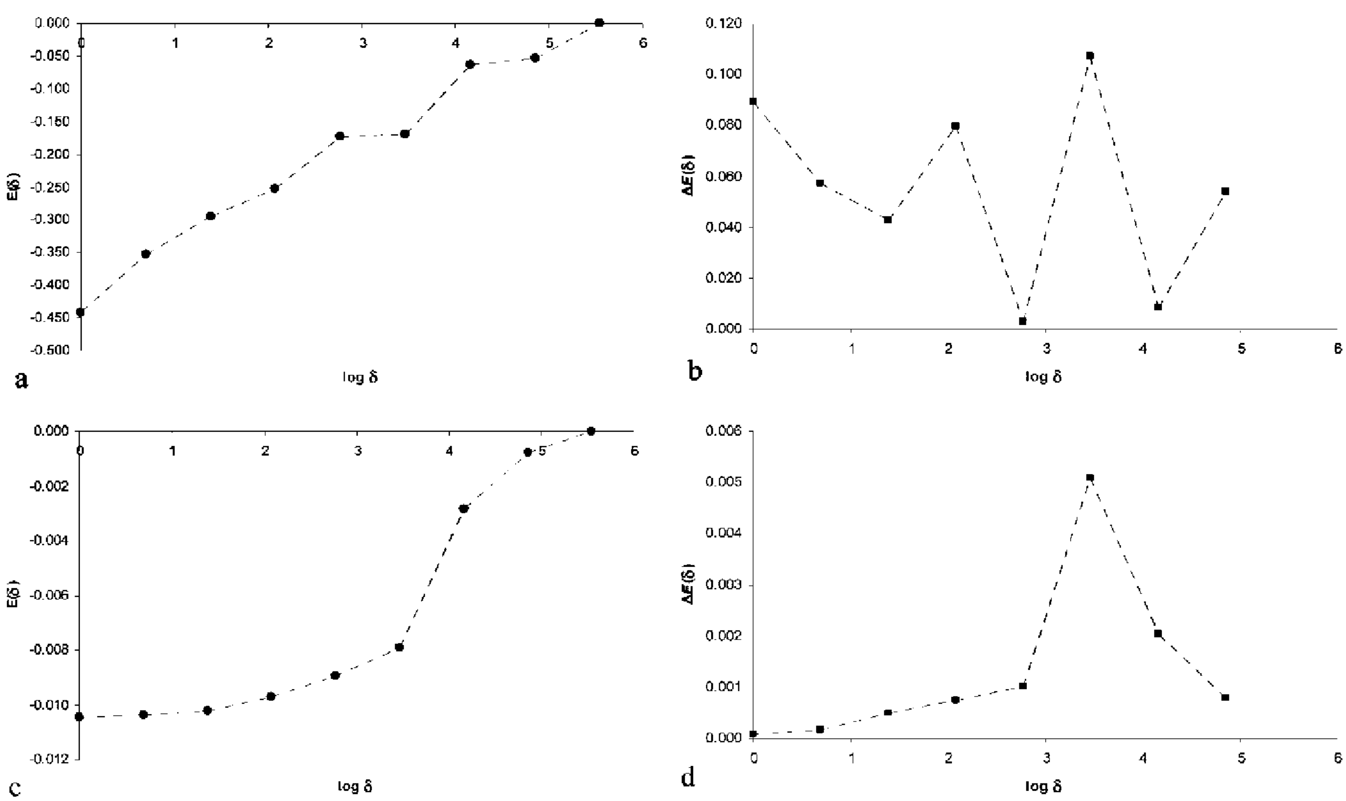

FIG. 4. Entropy study: (a) relative entropy, $E(\delta)$, of N_2O flux, (b) increment of relative entropy, $\Delta E(\delta)$, of N_2O flux, (c) relative entropy of pH, and (d) increment of relative entropy of pH.

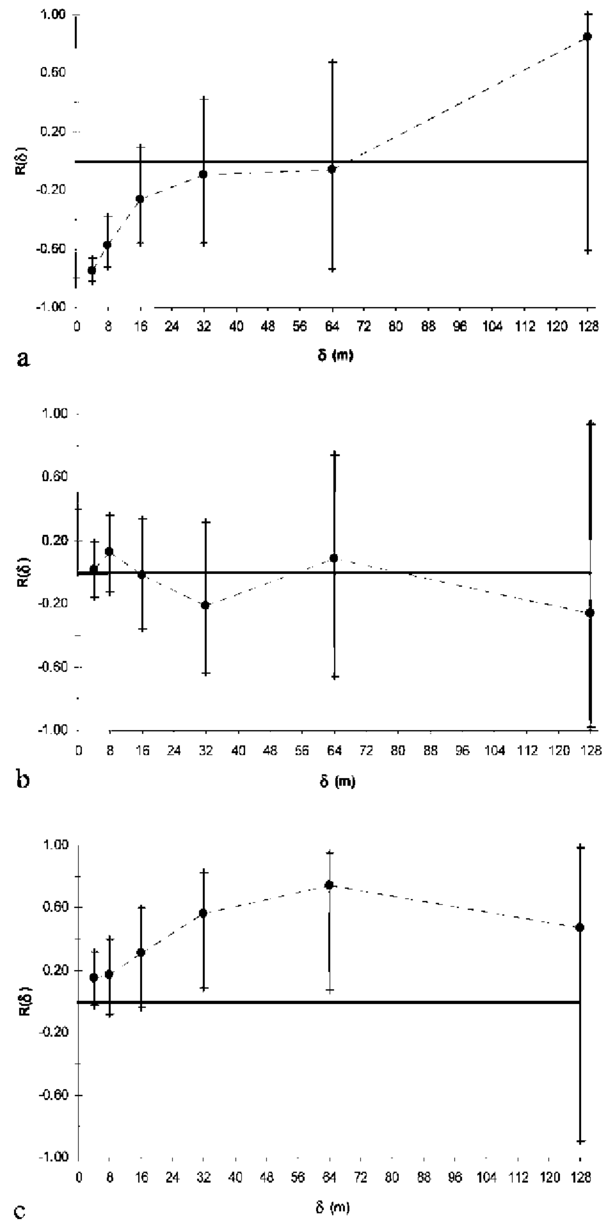

Fig. 5. Bivariate analysis through intrascale correlation function, $R(\delta)$, of volumetric water content and: (a) porosity, (b) N_2O flux, and (c) pH. The plus marks are the upper and lower limits of the confidence interval (95% confidence level).

Results and Discussion

Relative Entropy

Relative entropies and increments in relative entropy calculated for the four variables are shown in Fig. 3 and 4. Volumetric water content, ϕ, and pH show similar scaling trends, with maximum structure revealed at scale $\delta = 32$, corresponding to 128 m in the transect. This coarse-scale structure corresponds well with the wavelet analysis of the same transect by Lark et al. (2004), who attributed structure in the data at these scales to changes in underlying parent material. The N_2O data reveal a different and more complex scale dependency with structure, apparent at both large and small scales. This is not unexpected for a soil process with complex dependencies across a range of different soil properties.

Bivariate Analysis to Detect Scale-Dependent Correlations

The intrascale-dependent correlation functions are shown in Fig. 5. Figure 5a reveals that θ and ϕ are negatively correlated at fine scales up to $\delta = 2$, corresponding to 8 m. This we may attribute to the presence or absence of cracks and other air-filled macroporosity in otherwise similar media, which has opposing effects on the values of the two variables. At larger scales, the two variables become uncorrelated as we aggregate the data values.

Figure 5b shows no significant correlations between θ and N_2O emissions at any scale. Again, this is not surprising given the complexity of the denitrification process and its dependency on a range of soil properties.

Figure 5c shows positive correlations at intermediate scales between θ and pH, suggesting that both variables are responding in a like manner to changes in underlying parent material.

Conclusions

During recent years, the concepts of fractals and multifractal measures have been increasingly applied in the analysis of spatial variability of processes and properties in soil. In terms of modeling, it is important to characterize the multiscale heterogeneity of soil properties in a useful way, not only restricted to the study of multifractal behavior.

Relative entropy and the intrascale correlation coefficient were used in this work. Both parameters have general applicability and do not require the presence of scaling symmetries or any other prior assumptions as to the structure of the data.

The proposed approach provides information about space and scale dependencies that are localized both in space and in scale. It provides information that is complementary to the information about scale dependencies found across a range of scales. Space- and scale-localized features are also revealed with the wavelet analysis. Establishing a relationship between these two localization methods presents an interesting avenue for the further research.

Acknowledgments

We are grateful to Dr. Murray Lark for providing the data and for the useful comments that he has given to this study. This work has been partially supported by CM-UPM under Project R05/11261 and INIA-RTA04-111-C3. Rothamsted Research receives grant-in-aid from the UK Biotechnology and Biological Sciences Research Council.

References

Basseville, M., A. Benveniste, K.C. Chou, S.A. Golden, R. Nikoukhah, and A.S. Willsky. 1992. Modeling and estimation of multiresolution stochastic processes. IEEE Trans. Inf. Theory 38:766–784.

Bird, N., M.C. Díaz, A. Saa, and A.M. Tarquis. 2006. A review of fractal and multifractal analysis of soil pore-scale images. J. Hydrol. 322:211–219.

Bruland, G.L., and C.J. Richardson. 2005. Spatial variability of soil properties in created, restored, and paired natural wetlands. Soil Sci. Soc. Am. J. 69:273–284.

Ding, Q., and W. Ding. 2006. Stress wavelets: Multi-scale and multi-resolution assessment of soil structure by the drop-shatter method. Soil Tillage Res. 88:168–179.

Dyck, M.F., R.G. Kachanoski, and E. de Jong. 2005. Spatial variability of long-term chloride transport under semiarid conditions: Pedon scale. Vadose Zone J. 4:915–923.

Feder, J. 1989. Fractals. Plenum Press, New York.

Gimenez, D., E. Perfect, W.J. Rawls, and Ya.A. Pachepsky. 1997. Fractal models for predicting soil hydraulic properties: A review. Eng. Geol. 48:161–183.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.

Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biol. Fertil. Soils 27:315–334.

Hupet, F., J.C. van Dam, and M. Vanclooster. 2004. Impact of within-field variability in soil hydraulic properties on transpiration fluxes and crop yields: A numerical study. Vadose Zone J. 3:1367–1379.

Kravchenko, A.N., R. Omonode, G.A. Bollero, and D.G. Bullock. 2002. Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity. Soil Sci. Soc. Am. J. 66:235–243.

Kravchenko, A.N., K.D. Thelen, D.G. Bullock, and N.R. Miller. 2003. Relationship among crop grain yield, topography, and soil electrical conductivity studied with cross-correlograms. Agron. J. 95:1132–1139.

Kumar, P. 2003. Multiple scale conditional simulation. p. 179–192. *In* Ya.A. Pachepsky et al. (ed.) Scaling methods in soil physics. CRC Press, Boca Raton, FL.

Lark, R.M., S.R. Kaffka, and D.L. Corwin. 2003. Multiresolution analysis of data on electrical conductivity of soil using wavelets. J. Hydrol. 272:276–290.

Lark, R.M., A.E. Milne, T.M. Addiscott, K.W.T. Goulding, C.P. Webster, and S. O'Flaherty. 2004. Scale- and location-dependent correlation of nitrous oxide emissions with soil properties: An analysis using wavelets. Eur. J. Soil Sci. 55:611–627.

Mandelbrot, B.B. 1982. The fractal geometry of nature. W.H. Freeman, New York.

Percival, D.B., and A.T. Walden. 2000. Wavelet methods for time series analysis. Cambridge Univ. Press, Cambridge, UK.

Piegorsch, W.W., and A.J. Bailer. 2005. Analyzing environmental data. John Wiley & Sons, Chichester, UK.

Pringle, M.J., and R.M. Lark. 2006. Spatial analysis of model error, illustrated by soil carbon dioxide emissions. Vadose Zone J. 5:168–183.

Sofou, A., C. Tzafestas, and P. Maragos. 2001. Segmentation of soil section images using connected operators. p. 1087–1090. *In* Proc. IEEE Int. Conf. Image Processing (ICIP-2001), Thessaloniki, Greece. 7–10 Oct. 2001. IEEE, New York.

Vereecken, H., R. Kasteel, J. Vanderborght, and T. Harter. 2007. Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: A review. Vadose Zone J. 6:1–28.

Whitcher, B.J. 1998. Assessing nonstationary time series using wavelets. Ph.D. diss. Univ. of Washington, Seattle (Diss. Abstr. 9907961).

Zeleke, T.B., and B.C. Si. 2005. Scaling relationships between saturated hydraulic conductivity and soil physical properties. Soil Sci. Soc. Am. J. 69:1691–1702.