Analysis of complex soil variation using wavelets

R. Murray Lark

Silsoe Research Institute

1 Introduction

Soil scientists have made extensive use of geostatistical methods in order to study the spatial variations of soil. These have proved powerful for problems such as local estimation on point and block supports. However, the assumption of intrinsic stationarity, that is needed to define the variogram, is not plausible for the study of much soil variation particularly in complex landscapes with different parent materials, or for the study of variables arising from complex processes such as the evolution of trace gases from the soil. This is why there is interest in the wavelet transform as a tool to elucidate the variation of soil properties and processes without restrictive assumptions of spatial stationarity.

In this talk I show how wavelet transforms have been used to investigate changes in the variance of soil variables, changes in covariance, and the complex spatial variations of categorical soil variables in space.

2 Methods

I have used the maximum overlap discrete wavelet transform (MODWT) due to Percival and Guttorp (1994), and applying the adapted filters of Cohen et al (1993) in order to retain coefficients near the beginning and end of (1-D) data sets as proposed by Lark and Webster (2001). The MODWT transform is obtained by convolving the data with the dilated wavelet:

$$\psi_j(x) \;=\; \frac{1}{2^{\frac{i}{2}}}\psi\left(\frac{x}{2^j}\right). \tag{1}$$

for scale parameter 2^j, $j = 1, 2, \ldots$. This generates wavelet coefficients $d_{j,n}$.

From these wavelet coefficients we may obtain scale-specific measurements of variance

$$\breve{\sigma}^2_{u,j} = \frac{1}{2^j \breve{n}_j}\sum_{n=1}^{\breve{n}_j}\left(\breve{d}_{j,n}\right)^2, \tag{2}$$

and covariance for variables u and v.

$$\breve{C}_{u,v,j} = \frac{1}{2^j \breve{n}_j}\sum_{n=1}^{\breve{n}_j}\breve{d}^u_{j,n}\breve{d}^v_{j,n}. \tag{3}$$

I then followed Whitcher (1998) by searching for changes in the wavelet (co)variance at each scale by identifying breaks of slope in a normalized cumulative plot of the squared coefficients or product of the coefficients respectively. The evidence for a change was tested by Monte Carlo methods in each case (Lark et al, 2004).

The analysis of categorical soil variables is based on a proposal of Wang and Johnson (2002), itself based on ideas proposed by Stoffer et al. (1993) for stationary analysis of categorical time series. The key idea is to find some scaling that will map v onto a real-valued

sequence z with a continuous spectral density, so that

$$z(x_i) \in \boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_k\}^{\mathrm{T}} \in \Re^k \; \forall i. \tag{4}$$

The criterion for choosing the mapping in Stoffer's analysis was to maximize the spectral density of z across some frequency interval. In the analysis of Wang and Johnson (2002), which is based on the wavelet transform, the aim is to emphasize features of v that maximize the local variability as measured by a wavelet coefficient. A different mapping is therefore found in general for different parts of the sequence of values, and at different dilations of the wavelet.

I denote by $\mathbf{v}$ a sequence of values of a categorical variable v with k levels, $c_1, c_2, \ldots, c_k$. This may be mapped onto a set of indicator column vectors, $\mathbf{y}$, each of length k. Each of these column vectors is mapped onto a value in the numerical sequence $\mathbf{z}$ by computing

$$\mathbf{w}^{\mathrm{T}}\mathbf{y}, \; \text{where } \mathbf{w} \equiv \begin{bmatrix} w_1 \\ w_2 \\ . \\ . \\ w_k \end{bmatrix}, \tag{5}$$

so, if we denote by $\mathbf{Y}$ a $k \times n$ matrix with columns equivalent to a series of these indicator vectors $\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(n)$, then the real-valued sequence $\mathbf{z}$ is obtained by computing

$$\mathbf{z} = \mathbf{w}^{\mathrm{T}}\mathbf{Y}. \tag{6}$$

Now, if we are to perform a linear transform on $\mathbf{z}$, such as a wavelet transform, then a coefficient (in the case of the maximal overlap wavelet transform $d_{j,m}$ for some scale parameter 2^j and location m) is obtained by post-multiplying $\mathbf{z}$ by a column vector of n filter coefficients, $\mathbf{h}_{j,m}$ in the wavelet case, so

$$d_{j,m}^v = \mathbf{w}^{\mathrm{T}}\mathbf{Y}\mathbf{h}_{j,m}, \tag{7}$$

because the filtering steps are linear. We may therefore first compute the product of the indicator vectors and the transform filter coefficients to obtain

$$\mathbf{u}_{j,m} = \mathbf{Y}\mathbf{h}_{j,m}, \tag{8}$$

so

$$d_{j,m}^v = \mathbf{w}^{\mathrm{T}}\mathbf{u}_{j,m}. \tag{9}$$

The contribution made to the wavelet variance from coefficient $d_{j,m}$ is

$$\frac{\left(d_{j,m}^v\right)^2}{2^j} = \frac{1}{2^j}\left|\mathbf{w}^{\mathrm{T}}\mathbf{u}_{j,m}\right|^2 = \frac{1}{2^j}\mathbf{w}^{\mathrm{T}}\mathbf{u}_{j,m}\bar{\mathbf{u}}_{j,m}^{\mathrm{T}}\mathbf{w}, \tag{10}$$

where $\bar{\mathbf{u}}^{\mathrm{T}}$ denotes the conjugate transpose, equivalent to the simple transpose here since our wavelets are all real-valued.

Wang and Johnson (2002) proposed that we find a weight matrix $\mathbf{w}$ such that the wavelet variance component in Equation (10) is maximized subject to the constraint that $||\mathbf{w}||^2 = 1.0$. They show that this is achieved if $\mathbf{w}$ is proportional to the eigenvector corresponding to the largest eigenvalue of $\mathbf{u}_{j,m}\bar{\mathbf{u}}^{\mathrm{T}}$. In practice we can find the largest eigenvalue and the corresponding normalized eigenvector, $\tilde{\mathbf{w}}$. The former is the local (maximized) component of the wavelet

variance and the latter is the local weight matrix for mapping the categorical variable to the continuous numerical sequence.

A problem remains in that $\mathbf{u}$, as defined in Equation (8), is $k \times 1$, so $\mathbf{u}_{j,m}\bar{\mathbf{u}}^{\mathrm{T}}$ is of rank 1 and therefore singular. Wang and Johnson (2002) propose that n_j matrices $\mathbf{u}_{j,m}$ are computed for contiguous points $m, m + 1, \ldots, m + n_j - 1$ from the MODWT wavelet coefficients. These are then averaged to form a new vector $\breve{\mathbf{u}}_{j,m}$ and the largest eigenvalue and corresponding standardized eigenvector are extracted from $\breve{\mathbf{u}}_{j,m}\breve{\mathbf{u}}_{j,m}^{\mathrm{T}}$. These are then the wavelet variance and weight vector respectively for the segment of n_j units. Johnson (pers. comm.) confirmed that they set n_j to 2^j.

3   Examples

I shall present example analyses. The first set are drawn from a recent study on the spatial variability of trace gas emissions from the soil. Analysis of the wavelet coefficients shows how the variability of rates of carbon dioxide emission from the soil changes in response to parent material and topography. Analysis of the coefficients for nitrous oxide emission rate and those for certain properties of the soil reveals both scale-dependence, some soil factors are correlated with the emission rates at fine spatial scales and others at coarse scales, and location dependence — the correlation between a soil property and the rate of gas emission at some scale is not uniform in space.

The second analysis is of a categorical variable. This is a three-level classification of topography on a transect over Gilgai terrain on the Bland Plain in Eastern Australia. The wavelet analysis shows changes in the variability of the topography and transient features. Analysis of continuous soil variables measured at the same location shows similar changes in variability.

4   Conclusions

The wavelet transform, in particular the MODWT and wavelet transforms of categorical variables, have much to offer in the analysis of complex environmental variation. They allow us to identify key non-stationary features of variation such as transients or changes in the variance or correlation of variables. This information may be used both to gain insight into the underlying processes, and for problems such as planning sampling or validating process models.

5   Acknowledgements

References

Cohen, A., Daubechies, I. and Vial, P. 1993. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, **1**, 54–81.

Lark, R.M. and 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., Milne, A.E., Addiscott, T.M., Goulding, K.W.T., Webster, C.P. and 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**, (In Press).

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

Stoffer, D.S., Tyler D.E. and McDougall, A.J. 1993. Spectral-Analysis for Categorical Time-Series - Scaling and the Spectral Envelope. *Biometrika*, **80**, 611–622

Wang W. and Johnson D.H. 2002. Computing linear transforms of symbolic signals. *IEEE Transactions on Signal Processing*, **50**, 628–634.

Whitcher, B.J. 1998. *Assessing nonstationary time series using wavelets*. Ph.D. Thesis, University of Washington, Seattle WA.
(www.cgd.ucar.edu/∼whitcher/papers/thesis.ps.gz).