Using the NAG library to help analyse the variation in data from the environment

Alice Milne[1] & Murray Lark[2]
[1]*Rothamsted Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*
[2]*British Geological Survey, Keyworth, Nottingham NG12 5GG*

To understand the world around us, scientists measure numerous properties of the environment. Their aim is often to quantify a resource or assess how it changes over space or time. For example, a soil scientist might be interested in how the clay content of soil varies over a landscape, and a hydrologist might want to know how the nitrate concentration in a river varies over time. At its simplest, the variance statistic quantifies this. Scientists are also often interested in the correlation between two variables as this can, in some cases, provide insight into how one variable responds to changes in another.

Typically we should expect both variance and correlation to change with the scale of the measurements and the locations at which they were made. This is because there are likely to be different processes affecting a given variable at different scales, and these processes might change across space or time. The Discrete Wavelet Transform (DWT) and the Maximum Overlap DWT (MODWT) are signal processing methods that can be used to partition the variance (and correlation) in a series of measure according to given scale intervals and locations (see Percival and Walden, 2000). This type of analysis can help us to understand the processes that affect the variable of interest.

We wrote a program to compute wavelet coefficients and from these estimate wavelet variance and wavelet correlation at several scales. We wanted to compute the confidence intervals for these estimates and used the chi square routine from chapter G01 of the NAG Library (Numerical Algorithms Group, www.nag.com) for this purpose. The program also looks for significant changes in the wavelet variances and wavelet correlations across the series of data. Changes are deemed significant when they are larger than we might expect from a stationary series of data with the same auto covariance structure. This analysis uses the NAG library random number generator and LU decomposition algorithms, from the pseudorandom number generation suite in chapter G05, to generate simulated data with a given covariance structures. We also make use of NAG routines to compute correlation and covariance matrices, in the correlation and regression analysis chapter (G02) and routines from the time series analysis chapter (G13). We found the NAG Library functions simple to use, reliable and effective.

Our program has been used to analyse data from several applications, including our investigation on how soil properties influence nitrous oxide emissions from agricultural land. Nitrous oxide is one of the most potent greenhouse gases, and we wanted to understand how soil properties influence emissions at different spatial scales. We took soil cores from 256 locations along a 7-km transect in Bedfordshire. We measured emissions of nitrous oxide, and the contents of nitrate and water in the soil (among other variables). We used our program to calculate wavelet variances for each of the variables and the wavelet correlations between the nitrous oxide emissions and other soil properties. We also considered the spatial uniformity of the correlations between soil properties and emission rates across the landscape. The analysis revealed a complex pattern of scale dependence. There was a pronounced correlation between emission rates and a function of the water-filled pore space, seen only at landscape scales. Emission rates were strongly correlated with the nitrate content of the soil at intermediate and coarsest scales. The wavelet analysis showed that these correlations were not spatially uniform. The correlation between nitrate concentration and emission rates at the finest landscape scale was not significant in the northern part of the transect corresponding primarily to soils over the Lower Greensand, but these variables were significantly correlated at this scale over other parent materials. Our findings showed that, at the landscape scale, nitrate content and water-filled pore space are crucial soil properties for predicting nitrous oxide emissions and should therefore be incorporated into process models and emission factors for inventory (see Milne et al., 2011).

(fig I – Murray Lark and Ruth Skilton taking soil samples from a field at Silsoe, Bedfordshire.)

References

Numerical Algorithms Group (www.nag.com), NAG Library Manual, Oxford.

Milne, A.E., Haskard, K.A., Webster, C.P., Truan, I.A., Goulding, K.W.T. & Lark, R.M. *Wavelet analysis of the correlations between soil properties and potential nitrous oxide emission at farm and landscape scales*, European Journal of Soil Science, 62, 467–478.

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

Acknowledgements Our work was done under Rothamsted Research's programme in Computational and Mathematical Biology, which is funded by the Biotechnology and Biological Sciences Research Council (BBSRC).