OISIN – Optimised Inventorying of Soil [Spatial] Information Networks

Sang, N., Aalders, I.,

Macaulay Land Use Research Institute, Aberdeen, AB15 8QH, UK.
n.sang@macaulay.ac.uk, i.aalders@macaulay.ac.uk

Abstract. Recognition of the need for 'evidence based policy', poses a challenge for science. It is almost inevitable that a system of monitoring points which provides a nationally representative picture for one variable will be less optimal for another. Determining a single set of monitoring points that is optimal (or at least sufficient) for both, is a time consuming and difficult task. For complex cross-compliant policy objectives, with many different requirements, it is (practically) impossible to achieve manually, particularly when information may be drawn from multiple existing monitoring networks. We present a methodology for automatically selecting a sub-sample from one or more existing point sample sets, optimized to represent the distributions of selected variables either within the original data sets or drawn from GIS layers on environmental context such as slope and land cover.

1 OISIN

Recent years have seen increasing recognition of the need for 'evidence based policy', through legislation such as the Water Framework Directive (EC 2000), European Landscape Convention (Council of Europe, 2003.) and the proposed European Soils Directive (EC 2004). This poses a challenge for science – to take into account how the behaviour of a system responds to complex circumstances in the real world. A process which is well understood under controlled or "laboratory" circumstances may change under different contexts of land cover, climate or human influences. Further more, policies often require integrated analysis of multiple processes. Understanding Carbon cycling might entail analysis of hydrological and soil processes simultaneously, for example.

This places great demands on the range and detail of data needed to support such research, because sufficient samples need to be taken to provide statistically significant information within each context of interest. Data also needs to be sufficiently extensive, since many of the challenges facing policy makers do not respect political boundaries. Much existing data (land cover maps, soil maps, water quality) was not collected with such analyses in mind, and so new monitoring systems need to be developed which allow research to isolate confounding factors, and are consistent across traditional boundaries. However, monitoring networks are expensive, so the scientific needs must also be balanced against available resources.

OISIN is an extension for ArcGIS (ESRI Inc., 2005.) being developed at the Macaulay Institute, Scotland, to aid researchers in developing such a cross-compliant monitoring network for soil. The system allows researchers to quickly access historical information about soil properties available from the National Soil Inventory of Scotland (Towers et al. 2006), in order to compare possible monitoring networks as to which would provide the most representative national picture. Other data sets, such as land cover, topography or socio-economic variables can be included and information for the proposed monitoring points automatically extracted to be compared with the national picture.

It is almost inevitable that a system of monitoring points which provides a nationally representative picture for one variable will be less optimal for another. Determining a single set of monitoring points that is optimal (or at least sufficient) for both, is a time consuming and difficult task. For complex cross-compliant policy objectives, with many different requirements, it is (practically) impossible to achieve manually – particularly if there is an element of negotiation between different interest groups as to which variables should receive most attention.

OISIN provides a means to reduce the number of options to be considered by a process of simulated annealing (Albright 2007, Ferreyraa 2002). This involves automatically selecting a possible monitoring network as a subset of the points in one or more existing soils databases. Selection can be completely random, within a certain density range, or systematically. Statistics for the subset are then compared with those of a global or training data set, for each variable in turn. The results may be weighted according to the importance of the variable, before an over all "goodness of fit" score is given. The process is repeated, with previous options being discarded when a better score is obtained. By this process the plethora of potential options is filtered down to the most representative option, or set of options, for consideration. Since the process is undertaken within a GIS, the optimisation process can also take into account the geometrical distribution of monitoring points ensuring it is not overly clustered.

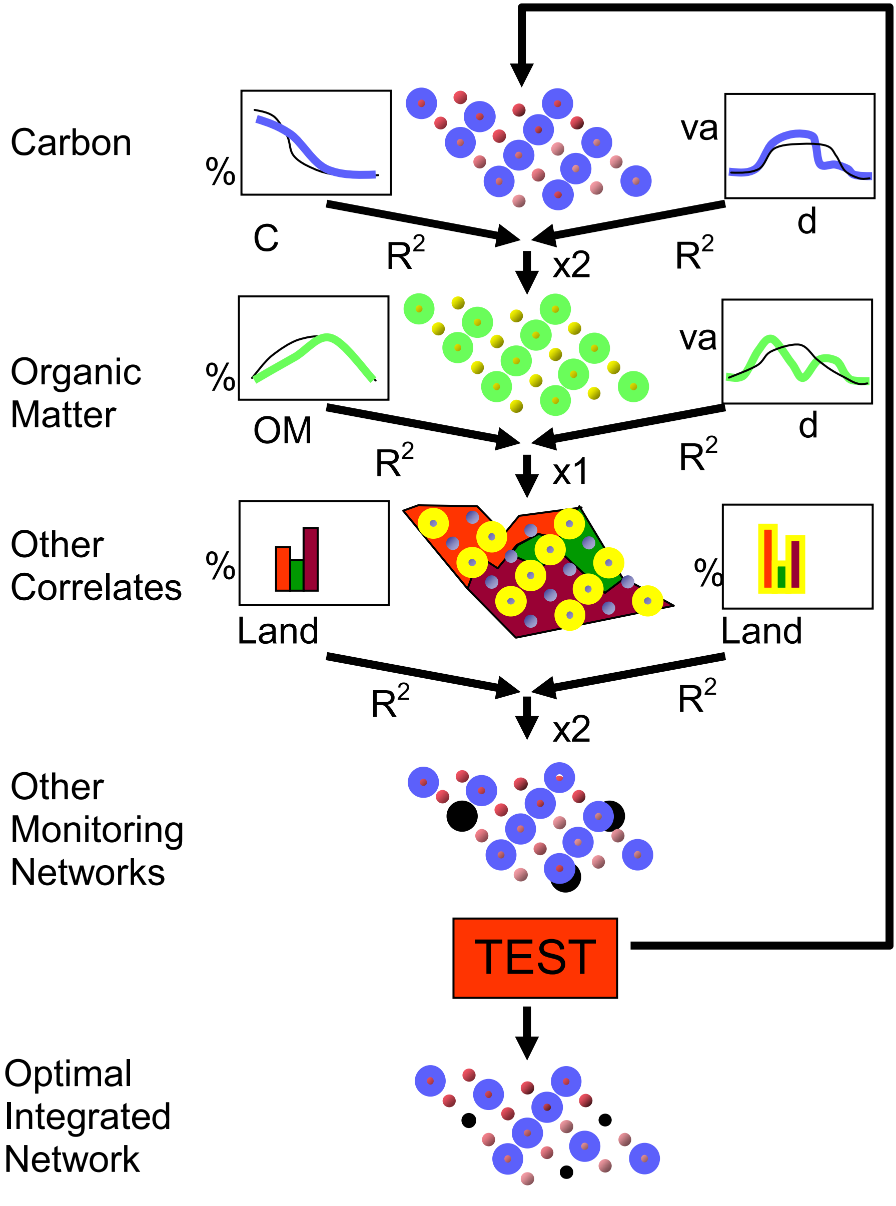

Fig. 1. The distribution of values of Carbon (C) and Organic Matter (OM) for the subset are compared with the full dataset and the R2 difference in trend calculated. A Semi-variogram provides an indication of spatial distribution of values in each case. The proportion of samples taken under different land covers is considered. Samples sufficiently close to points from other monitoring networks are considered "free" data, and the sampling effort re-allocated elsewhere. The statistical similarity of the sample to the global data set, is weighted for each variable according to importance in the monitoring scheme. Once minimum criterions are met, the best sample set is suggested.

In what is believed to be a new departure for sample optimisation systems, the ability to integrate different monitoring schemes has been included. By providing the option to select points from other monitoring networks, duplication of effort between monitoring agencies can be avoided, and data made more compatible. For example a monitoring scheme for Scotland might try to include as many points from the suggested EU monitoring Grid as possible, but not visit sites near to those for which soil data is already available from statutory water monitoring.

References

1. **Albright, B**., 2007, An Introduction to Simulated Annealing, The College Mathematics Journal, Volume 38, Number 1, January 2007, pp. 37-42(6) Mathematical Association of America.
2. **Council of Europe**, 2003. *Presentation of the European Landscape Convention*. Strasbourg, Council of Europe.
3. **EC,** 2004, DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing a framework for the protection of soil and amending Directive 2004/35/EC
4. **EC,** 2000, Directive of the European Parliament and of the Council establishing a framework for the Community action in the field of water policy 2000/60/EC
5. **ESRI Inc.,** 2005. ArcGIS 8.2. ESRI Inc., Redlands, CA.
6. **Ferreyraa, R.,** Apezteguíab, H., Serenob, R., Jonesa, J., 2002, Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing, Geoderma , Volume 110, Issues 3-4 , 265-289.
7. W. Towers, I.C. Grieve, G. Hudson, C.D. Campbell, A.Lilly, D.A. Davidson, J.R. Bacon, S.J. Langan and D.W. Hopkins, 2006, ISBN 0 7559 6250 5, http://www.scotland.gov.uk/Publications/2006/09/21115639/0