Extracting task-specific soil and landscape information from soil survey

Michael J. Grundy[1]

CSIRO, 306 Carmody Rd, Brisbane, Qld 4067, Australia. Tel: +61 7 3241 2909; Fax: +61 7 3214 2308; E-mail: Mike.Grundy@csiro.au

1. Abstract

Soil and terrain information is vital for watershed management which depends increasingly on results from simulation models. These models require input data relating to watershed response at a scale relevant to watershed management – the information must be *task-specific*. Often, the only source of soil information across a watershed is from general-purpose soil surveys. Unfortunately soil maps, and even soil survey reports, do not often provide task-specific information in a readily accessible form. However, there is hidden value because new approaches allow us to substantially improve the utility of soil survey information.

Task-specific information for watershed management includes landscape attributes (e.g. nature and dynamics of vegetated cover, terrain parameters including those relating to fluvial geomorphology, land use characteristics and the presence of anthropomorphic structures) and soil attributes (e.g. soil erodibility, soil thickness, soil permeability, pertinent soil chemistry and the vertical and horizontal variation of these features). Deriving task-specific information for watershed management from soil survey requires innovative use of existing data (legacy data). It is also forcing a re-evaluation of how we do soil surveys. This paper describes a suite of approaches for the re-use of soil survey information in Queensland, Australia and the development of new approaches designed to broaden the impact of new survey.

The utility of legacy data for watershed management has been increased by the development of:

1. Flexible data systems for soil and landscape data relating to soil attributes, soil profiles, sites, and map units;
2. Integration of vector and raster capabilities within geographic information systems (GISs);
3. Application of exploratory and confirmatory statistical methods for spatial modelling (e.g. data mining methods, fuzzy modelling and geostatistics) and the use of new forms of remotely sensed data, especially gamma radiometric spectroscopy (Wilford 2008); and
4. Supplementary data gathering based on an analysis of regions or landscape types where existing sampling is sparse and as a consequence, predictions have large uncertainty.

Our experience has led to the redesign of soil survey procedures and to an increasing recognition of the long term value of soil data sets when they are part of a coherent program of data improvement.

2. Introduction

Modelling is now central to watershed management. The capacity to explore options and predict the outcomes of management depends on a suite of models representing crop and land management systems, rainfall and runoff, sediment transport and water quality. These models gain predictive power when they are linked to temporal remote sensing. Given the need to understand past, present and future changes in the nature of watershed processes and the need to interpolate between observations, these models have become essential both in the management and monitoring of condition and trend.

The models require data which fit the conceptualisation of the model (e.g. the scale of measurement must match the scale of the relevant process) and, as importantly, fit the scale at which the watershed will be modelled and managed. This paper refers to this as "task-specific" information.

In many parts of the world, apart from remote sensing, general climate surfaces, generalised geology maps and digital elevation data, the only biophysical information available across watersheds comes from soil surveys. Typically, these surveys were designed to meet "pre-computer" information use with soil type or interpreted maps and accompanying reports – they were not designed to inform simulation modelling or watershed management. There is substantial value, however, in many of these surveys and when used carefully, they yield useful data for watershed modelling.

This paper illustrates this potential with several recent examples from Queensland, Australia, and draws some conclusions from the experience.

3. The models and the data

Models now commonplace in watershed management range from point-based one-dimensional models for water balances, forest growth and cropping systems (e.g. APSIM – http://www.apsim.info/apsim/what-is-apsim.asp) through to various forms of distributed sediment and water quality models. Australian examples of the latter include the E2 model (Argent et al. 2005) and SEDNET (Prosser et al. 2001). The soil data required by these models varies but often include complex parameters for water storage, soil erosion, sediment transport and the connections between soils, hillslopes and surface or groundwater.

4. The nature of survey information

Conventional surveys have not often provided estimates of the input parameters for models. The surveys were more concerned with conceptual models of soil–landscape relationships that link observable soil features and soil types with genetic and environmental factors in general accord with Jenny's (1941,1981) factors of soil formation. The linkage is necessary because it allows the placing of soil bodies within a spatial pattern. However, the models developed during soil survey have rarely been explicit or predictive (McSweeney et. al., 1994). In summary, conventional soil survey typically produces the following types of information.

Descriptive soil sites

A survey involves the collection of descriptive data often used for the internal purposes of survey (e.g. specification of soil classes and preparation of map legends). These data are often spatially dense but they often only comprise descriptions of soil morphology and without associated laboratory analysis, they are difficult to use for other purposes.

Analytical sites

Sites which correspond to the central concept of mapped soil classes are often analysed for a range of soil chemical, physical and mineralogical attributes. These data are then, either explicitly or implicitly, linked to the soil class and mapping units showing the distribution of these classes. The validity of this linkage is rarely tested.

Polygon map

Soil maps are usually a collection of polygons (i.e. the map units) coded and shaded to match a legend which lists the dominant soil type, the landscape context and in most cases interpretations relevant to the expected use (e.g. land suitability for a range of land uses). In recent decades, these polygons form a GIS layer.

Map legend

These normally list characteristic combinations of soil (soil profile class), landform (landform pattern and element), regolith type and geology. The legends capture part of the conceptual soil-landscape model used by the surveyor during air photo interpretation and field survey.

Survey reports

These vary substantially in content and utility. They often outline the interpretations of landscape processes and the link to the classification system – this is useful for estimating soil attributes for simulation models.

Important assumptions in conventional soil survey provide additional limitations. A soil map and its implicit, underlying conceptual models use classification (profile classes or local taxonomic groups) to divide the multi-dimensional attribute space into classes designed to carry the array of attributes. The classes then are tagged to map units; these place the classes in space (the aim is "pure" units at fine scales; explicit combinations of classes at broader scales). The implied assumption is that attributes tend to be coincident in the class and in the polygon – rarely-tested hypotheses. Further, distinctions between soil bodies (the classes and the polygons), are believed to be good surrogates for distinctions between land qualities and land use limitations, so that soil map units can be useful for the application or prescription of land management practices. Any understanding of the shades of soil and landscape change that the surveyor develops is lost in classes, groups and polygons.

Soil survey information is thus difficult to use well in watershed modelling – and apply the richness of the landscape understanding that the surveyor developed. A common method, for example, is to further characterize some of the (few) detailed analytical sites to derive attributes such as available water holding capacity, rooting depth, hydraulic conductivity and surface erodibility. These are then modelled and the results distributed across the landscape through transfer by analogy (Thomas et al. 1995).

With an understanding of the nature of the information and armed with a new array of analytical tools and pertinent ancillary data, we can do substantially better than this – by re-engineering survey data.

5. Deriving task specific information

In Queensland, the soil survey and modelling community are increasingly working together – and a major focus has been on improving the spatial performance of modelling by deriving task-specific information from survey data. The approaches have developed into a coherent program of data improvement. Some of the developmental elements of the program are listed below and will be illustrated in the oral presentation.

5.1 Data systems and data re-interpretation – the first step

The fundamental tool for the reuse or enhanced use of soil survey information is a comprehensive and spatially enabled data system. The SALI system in Queensland has sites, polygon and classification seamlessly embedded with a GIS system (Metadata: http://www.nrm.qld.gov.au/asdd/qsii2/ANZQL0132000109.html). These data types are cross-referenced in the system so that, for example, a soil profile class can be statistically described based on the population of sites which are identified within the class, and polygons can be linked to sites enclosed within the polygon, to similar sites elsewhere which have been tagged as representative of a class or to the wider population of similarly classified sites.

This system has provided the flexibility to extract as much information value as possible from the standard suite of survey products. At a program level, the new system has been essential to meet the requirements of the Australian Soil Resource Information System (http://www.asris.csiro.au/index_other.html; Brough et al, 2006) which is building a hierarchical national soil attribute system. With the use of pedotransfer functions, polygons across the state have been tagged with a dominant attribute value and estimated uncertainty. Such data underpin watershed modelling projects such as the water quality modelling using EMSS in the Maroochy watershed (Searle et al. 1995).

5.2 Fixing common errors with new approaches

It is not unusual for historic soil data to be poorly registered geodetically or, due to limitations in aerial imagery or field operations, to have errors in reflecting the soil – landscape concepts intended in the mapping. Simple corrections are possible with physiographic data derived from digital elevation models and classified satellite imagery. Much of Queensland has land-systems mapping and these have been systematically corrected with simple GIS adjustment: overlay analysis, ancillary data and land-systems concepts.

A more interesting and powerful approach used gamma radiometric spectroscopy (Wilford 2008) both to improve historic land-system mapping and to increase the utility for land evaluation (Wilson and Philip 1999). Initially, the land system concept models derived from the mapping and reports were used to stratify and classify airborne gamma radiometrics Thorium and Potassium signals within dominant land-systems for the study area in north-west Queensland. This process allowed definition of broad landscape types and showed some spatial relationships between land systems. Through a field process, these classes were refined to define major geological variations between land systems, reinforcing and in many cases, improving the mapping. The data were particularly valuable in upland erosional areas where radiometric signals are dominated by exposed country rock; in distinguishing differences in basalt flows which are hard to map using conventional methods and which have significance in assessing soil nutrient variations; and in delineating with K radiometric signals, more recent, active flood-out and depositional areas in visually uniform clay plains. The land system mapping was significantly improved – within the original landscape concepts.

5.3 Refining attribute estimation with ancillary data – soil depth prediction

Soil depth is a key task specific parameter – and one which is unsatisfactorily predicted in conventional soil survey. Moore et al. (1993) pioneered the use of digital terrain data and identified the potential to distribute soil attributes within map units where attribute values are largely a function of terrain. In many landscapes the compound topographic index (CTI) derived from a digital elevation model (Gessler *et al*. 1995) usefully separates erosional from depositional facets. Across a large area of south-east Queensland, Claridge et al. (2000) used the index to distribute soil depth within mapped polygons. Because the frequency distribution of the CTI within each zone varies in different parts of the landscape, the approach taken was to treat the mid value (or mean) of the depth index as the mode for the distribution of depth within each unit, and the (extended) minimum and maximum values as the limits to the distribution of depth in the polygon – essentially two linear depth curves meeting at the mode. The resulting surface substantially altered and improved the spatial prediction of soil properties with the forest model 3PG.

5.4 Estimating attributes by finding the soil – landscape concepts

Agroforestry has become an increasing part of the mosaic of land use in many agricultural areas. In a recent integrated modelling study, the model *3PGSpatial* was used to evaluate the potential for agroforestry in south-east Queensland. The parameters required were not present in conventional survey data so several soil attributes (both morphological and chemical) were derived from the maps and site data by using statistical methods. The soil landscape models present in the data were extracted using decision trees and generalised linear models. Two study areas were investigated which allowed the comparison of statistical results; they differed in size and pedogenetic environments – and in the success of the method. Core digital environmental data sets included a 25 m Digital Elevation Model (DEM) and derivatives, geology coverages, LANDSAT TM, airborne gamma ray spectrometry and climate indices that were topographically adjusted. The study produced soil attribute maps with measured statistical confidence. Good predictions (r^2>0.5) were obtained for CEC, clay content, total P and plant available P and acceptable predictions for a range of other parameters. These attributes were then used as inputs to forest models to test the improvement in the spatial resolution. The enhanced data produced different and finer spatial predictions of forest suitability compared with more conventional use of soil information and expert advice. Within the forest productivity model *3PGSpatial*, the prediction of spatial variation of key forest productivity variables such as average Diameter at Breast Height (DBH) and Leaf Area Index (LAI) were significantly improved. The use of the high spatial resolution soil water holding capacity layer also appeared to simulate the distribution of soil water within the watershed more accurately than broad scale soil data (Coops 2001).

5.5 Building on the survey – supplementary sampling for attribute prediction

The value of a survey can be enhanced and task-specific information more reliably derived if the survey data is strategically improved by the collection of new information. This approach was used by Claridge and Grundy (2004) to improve the attribute estimation of landscapes in the lower Balonne area of Queensland. Existing soil site data were available but had been collected to produce a conventional map; attribute data were required to model changing management practices to prevent salinity development. The approach used was to determine the extent to which the existing sites covered the attribute space for soil factors by matching with terrain and gamma radiometric data. It is commonly observed that conventional survey has strong sampling biases related to the original survey purpose – that was the case here. A sampling scheme was then established to cover the inadequately sampled attribute space. The enhanced soil site data were then used within a tree-based regression analysis to derive attribute maps for the area – with estimates of uncertainty – and to feed into modelling for salinity risk and land management options (Grundy and Macaulay 2004).

6. The drive to re-interpret and the imperative for survey change

There are numerous other examples of the extended use of Queensland soil survey data. For example, in other studies, sites collected for an acid sulfate soils survey, have been re-analysed using kriging to produce surfaces of soil depth and fuzzy modelling has been used with ancillary terrain data to distribute land zone values within complex land system polygons.

Across many land resource agencies, there is as much effort in re-using soil survey data for new purposes as there was in using the data for the original purpose and the benefits are likely to be much larger and significant (ACIL 1996). The range of techniques and the power of the new data generated continue to grow. Rossel et al. (2008) demonstrate sampling from existing soil archives with spectroscopic scanning to capture the attribute ranges within the samples and derive rapid estimates of soil properties well beyond the scope of the existing survey. There is increasing investment in a new generation of soil inference or transfer functions (Tranter et al. 2007) and national survey databases are being re-engineered to increase the utility of the data (ASRIS in Australia, S-Map in New Zealand - Lilburne et al. 2004).

This history suggests a strong need for a change in the methodology of survey. The confluence of new methodologies, substantially expanded computation capacity and a wide range of new uses for and demands on soil information now make this possible and necessary. McKenzie et al. (2008) make the case for the change and provide the guidelines to achieve it.

7. References

ACIL Economics and Policy Ltd (1996) The development of an economic framework and instruments for assessing the benefits and costs of land resource assessment in Australia. Final report to the Australian Collaborative Land Evaluation Program. http://www.clw.csiro.au/aclep/Other/Acil(s).pdf

Argent, R. M., Grayson, R. B., Podger, G. M., Rahman, J. M., Seaton, S. P., Perraud, J.-M., 2005. E2 - A flexible framework for watershed modelling. In: Zerger, A. and Argent, R. M., (Eds.), MODSIM 05 International Congress on Modelling and Simulation: Melbourne, Modelling and Simulation Society of Australia, p. 594-600.

Brough DM, Claridge J and Grundy MJ (2006). Soil and landscape attributes: The creation of a soil and landscape information system for Queensland. Queensland Department of Natural Resources, Mines & Water, Brisbane, Australia. QNRM06186

Claridge, J and Grundy, MJ (2004) Prediction of fine scale soil attributes and deep drainage, 4pp, Paper 671, In: 13th International Soil Conservation Organisation Conference – Brisbane, July 2004

Claridge, J., Williams, K.J. and Storey, R., (2000) Creation of the South-East Queensland depth index rescaled using CTI. Enhanced Resource Assessment 2000 - 05. Queensland Department of Natural Resources and Mines, Brisbane.

Coops N (2001) Report to Queensland Department of Natural Resources. Completion of Selected Tasks of RIRDC Project Number: QDN3A. Estimating the productivity of forest systems in Southeast Queensland. CSIRO Forestry and Forest Products. Victoria. Australia.

Gessler, P.E., Moore, I.D., McKenzie, N.J. and Ryan, P.J. (1995). Soil-landscape modelling and spatial prediction of soil attributes. International Journal of GIS, **9**(4):421-432.

Grundy, MJ and Macaulay, S (2004). Land Management. In: Chamberlain T and Wilkinson K (eds) Salinity investigations using airborne geophysics in the Lower Balonne area, Southern Queensland. Department of Natural Resources and Mines, Queensland. QNRM04255. pp175-206

Jenny H (1941) 'Factors of soil formation.' (McGraw-Hill: New York).

Jenny H (1980) 'The soil resource: origin and behaviour.' (Springer-Verlaag: New York).

Lilburne, L. Hewitt, A. Webb, TH and Carrick, S (2004) S-map – a new soil database for New Zealand In: Singh, B (2004). Supersoil 2004: Proceedings of the 3rd Australian New Zealand Soils Conference, University of Sydney, Australia, 5 – 9 December 2004. www.regional.org.au/au/asssi/supersoil2004

Littleboy, M., Silburn, D. M., Freebairn, D. M., Woodruff, D. R, and Hammer, G. L. (1989). PERFECT: A computer simulation model of productivity erosion runoff functions to evaluate conservation techniques. Queensland Department of Primary Industries Publication QB89005.

McKenzie, NJ, Grundy, MJ, Webster, R and Ringrose-Voase, A (2008) Guidelines for Surveying Soil and Land Resources, Second Edition, Australian Soil and Land Survey Handbooks Series. CSIRO publishing, ISBN: 9780643090910 576 pages

McSweeney, K., Slater, B.K., Hammer, R.D., Bell, J.C., Gessler, P.E. and Peterson, G.W. 1994. Towards a new framework for modeling the soil-landscape continuum. Soil Science Society of America, 677:127-145.

Moore, I. D., Gallant, J.C., Guerra, L., and Kalma, J.D. 1993. Modelling the spatial variability of hydrological processes using GIS. International Conference on Application of Geographic Information Systems in Hydrology and Water Resources Management. Austria.

Prosser, I. P., Rutherfurd, I. D., Olley, J. M., Young, W. J., Wallbrink, P. J. and Moran, C. J. (2001). Large-scale patterns of erosion and sediment transport in river networks, with examples from Australia. Marine and Freshwater Research 52(1): 81-99.

Rossel RAV, Jeon YS, Odeh IOA, McBratney AB (2008) Using a legacy soil sample to develop a mid-IR spectral library. Australian Journal of Soil Research **46**, 1–16

Searle, R., Oliver, P.,Sallaway, M., Griffith, G. and Weber, T. (2005) Implementation Of A Water Quality Model By The Community – Some Lessons. In Zerger, A. and Argent, R.M. (eds) MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 1168-1174. ISBN: 0-9758400-2-9.

Thomas EC, Gardner EA, Littleboy M, Shields P (1995) The cropping systems model PERFECT as a quantitative tool in land evaluation: an example for wheat cropping in the Maranoa area of Queensland. Australian Journal of Soil Research 33, 535–554.

Tranter, G., Minasny, B., McBratney, A., Murphy, B., McKenzie, N., Grundy, M., and Brough, D. (2007). Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management, 23 (no. 4): 437-443

Wilson, P.L. & Philip, 1999. S.R. "The Use of Airborne Gamma Spectrometry (Radiometrics) and Digital Elevation Models for Soils Mapping in North Queensland" Proceedings of NARGIS 99 - North Australian Remote Sensing and GIS Conference.