GIS-driven Models of Soil Properties in the High Country of the South Island

Linda Lilburne[1], Allan Hewitt[2], Peter McIntosh[2] & Ian Lynn[1]

[1] Landcare Research
P. O. Box 69, Lincoln, New Zealand
Phone: +64 3 326 6700 Fax: +64 3 325 2418
Email: lilburnel@landcare.cri.nz

[2] Landcare Research
Private Bag 1930, Dunedin, New Zealand

Presented at the 10th Colloquium of the Spatial Information Research Centre, University of Otago, New Zealand, 16-19 November, 1998

Abstract

An extensive New Zealand-wide soil survey with sufficient detail for solving land use and environmental protection problems is not financially feasible. Therefore soil scientists need to develop new ways to extrapolate known information about soil. Integrating a conceptual soil-landscape model with a GIS has the potential to make the process of mapping soil data more efficient. This paper describes two GIS-driven soil models of topsoil carbon content in the high country of the South Island of New Zealand. Both models use topography and other information to produce maps of predictions of carbon content that vary continuously.

Keyword and phrases: soil-landscape model, Bayesian, soil carbon, spatial modelling**.**

1.0 Introduction

Soil survey can be an arduous and expensive exercise. Conventional soil maps, depicting the distribution of soil classes, are made by soil scientists who laboriously traverse the landscape with spade and auger, record soil observations, assign them to soil classes, and draw lines representing their spatial extent. Recently soil survey has all but lost its public funding base as resources have shifted from data acquisition to application. There is , however, a continuing need for more soil data. The soil exploration of New Zealand is incomplete and even in those areas already surveyed, solving land use problems often requires more detail, better quantification of soil data and sometimes new types of data.

Soil surveyors deal with three kinds of observations. The first are direct observations of the soil mantle gained by digging at points in the landscape. The second are direct observations of land surface features, for example, surface stoniness, slope angle and shape, and vegetation. The third are other observations derived from relevant spatial information such as geology maps and climate maps. Soil-landscape modelling uses observations of surface features and other relevant spatial information to interpolate and extrapolate the point observations of subsurface soil attributes. Soil-landscape models have mostly existed as conceptual models in the mind of the soil surveyor. They have usually not been written down. The surveyor walks across the landscape taking subsurface observations, and combines these with other observations and knowledge to make predictions. We call the models “eyZerience” and they usually go to the grave with the one who developed them (Hewitt, 1994).

The way forward for soil survey is to work more efficiently, making better use of field time, and squeezing more information out of the available field data and soil expertise by developing better soil-landscape models. Information science can help achieve efficiencies in several aspects of soil survey. One principle opportunity lies in the development of soil-landscape modelling (Webb, 1994) by making the models explicit and developing them as powerful soil survey tools. Model development also provides the opportunity of revising the data concepts held in traditional soil survey. In many parts of the landscape soil properties vary continuously, but it has been the usual practice to split these continua into soil classes. Soils are depicted on maps as sharp-edged polygons, labelled as soil taxonomic classes. This has been the convenient paradigm for traditional soil survey, but with geographic information system (GIS) technology we now have the opportunity to map single soil properties rather than taxonomic classes, and to treat them as continuously varying surfaces.

This paper presents two approaches in soil-landscape modelling of topsoil carbon content in the high country of the South Island. The paper demonstrates the benefits of integrating spatial information science with soil survey. The first approach uses a triangular irregular network (TIN) to interpolate in *topographic space*. The second uses a bayesian framework.

2.0 New Zealand's South Island High Country

Of the 6 million hectares of the hill and high country in the South Island of New Zealand, over 13% has an annual rainfall of <700 mm. This dry steep land is sensitive to human impact and climate change: some soils are eroded or prone to erosion, and the vegetation is prone to depletion through fire and overgrazing by farm stock and rabbits. The soils in these areas are among the least well characterised in New Zealand. Assessment and monitoring of these soil resources is important to determine the best land use options and to detect trends in soil health. Mapping the properties of these soils is also important nationally. For example, an accurate estimate of the carbon content of the extensive high country soils can influence the national carbon budget models used in greenhouse gas research.

The best available soil map for the South Island high country is the New Zealand Land Resource Inventory (LRI). The LRI was designed to show general fertility trends and for regional planning rather than for detailed resource assessment. Earlier maps are generally not linked to databases, or where they are (such as the LRI), soil properties are only discernible through a soil name. Soil description and analysis has only been done at a few sites, giving insufficient spatial coverage for producing accurate single-factor maps. The source scale of the soil data is 1:250 000, which is not of sufficient standard to quantitatively assess the soil resources of the high country on a local or regional scale.

We selected as our study area the 26 000-ha Benmore Range in South Canterbury, New Zealand, in the heart of the MacKenzie Basin. This range was chosen because (1) it is formed of greywacke rocks typical of much of the South Island ranges; (2) it contains soils representative of the drier land; (3) it has a large area with relatively easy access; (4) it has been the subject of several soil studies since 1978; (5) it includes an elevation range (400–1800 m) typical of farmed leasehold land; and (6) farmers were cooperative and supported the survey.

3.0 First Model: DTM-Based Interpolation

Our aim was to produce a "quantitative soil property distribution model", which when combined with a digital terrain model (DTM), would allow prediction of individual soil properties in any part of the dry high country landscape. The soil distribution model was derived from results of soil sampling at 72 sites and additional knowledge of soil processes. The soil distribution model is based on aspect and elevation, which are strong drivers of soil variation in hilly and steep land. A stratified sampling plan was used to obtain values for soil properties at key aspects and elevations. We used GPS to accurately identify the location of each sample site. The soil sample results are described and discussed in McIntosh *et al.* (1998).

The first version of the model simply categorised the range into areas of 36 combinations of elevation (low, moderate and high), aspect (north, east, south and west) and land system (dry, moist and cool, moist and warm), and assigned each combination the appropriate topsoil percent carbon value from Table 1. Values in the table are based on averages of two sites for each combination.

	Moist and warm				Dry				Moist and cool			
Elevation	W	N	E	S	W	N	E	S	W	N	E	S
Low	3.2	2.9	2.8	4.2	1.3	1.8	1.8	3.9	1.4	1.4	1.8	4.0
Moderate	3.6	2.4	2.9	5.1	2.4	3.0	2.4	5.6	2.0	1.1	5.8	6.7
High	1.8	1.8	3.6	5.4	2.9	1.9	2.9	7.4	1.3	1.7	3.4	9.3

Table 1: Percentage of carbon in Benmore Range topsoils Land system and Aspect

However, carbon varies continuously over the landscape, so a second version of the model was developed that could make better use of the available DTM (developed from 20-m contours). Table 1 was modified by using actual aspects and elevations of sampling sites, data from additional sampling sites, knowledge gained from previous studies in the area, and a more general understanding of the mechanisms of the processes involved (e.g., erosion, mineralisation, leaching). In particular, elevation trends were smoothed; carbon values were adjusted downwards to take into account estimated rock and scree cover; and a study by Hewitt (1995) was used to identify likely points on the compass where trends, of soil properties related to aspect, change sharply. For example, soil temperature decreased more sharply south of 70 and 250 degrees. There is a dramatic change in percent carbon between sunny and shady slopes. However, the boundary between sunny and shady is not a sharp line, so we opted for a “fuzzy” boundary that spanned 20 degrees (i.e., 60–80 and 240–260 degrees). In this way a model in *topographic space* was built such that given any aspect and elevation, a prediction is made of the topsoil percent carbon content. The model can be represented as a radar diagram where the predicted value of carbon can be read from the axes of the graph (Figure 1). For example, at a point facing due

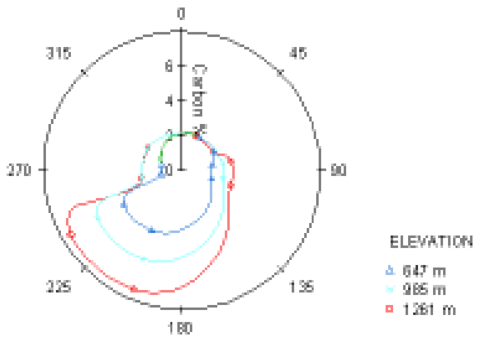

Figure 1: Polar graph showing predicted values of carbon content based upon aspect, for three elevations.

east at an elevation of 650 m, topsoil percent carbon is expected to be 1.75%. The value at, say, 110º and 1000 m is interpolated.

Linear interpolation between points in our model of topographic space involved non-Cartesian 3-dimensional interpolation that was non-trivial. This was solved by creating a series (one for each land system) of triangular irregular networks (TINs) based on aspect and elevation (or topographic space). In other words, the *x* and *y* axes of each TIN were aspect and elevation, and the *z* axis was carbon content. For each pixel of the DTM, the appropriate TIN is used to predict the carbon content thus generating a layer of topsoil percent carbon for the whole mountain range (Figure 2). The whole process has been captured in an AML‘ macro (used in Arc/Info‘) so that, should the numbers making up the model be revised, the layer can be easily regenerated. Similarly, layers for other soil properties (e.g., pH, phosphate retention) can be readily generated once the numbers for each property at key aspects and elevations have been defined.

Thirty-eight additional sample sites were used to test the model. Half were randomly selected points, the rest were on selected combinations of key aspects and elevations. An initial analysis that classified carbon into three classes showed that the sites on selected key aspect/elevation combinations were reasonably well modelled (58% correctly classified). Random sites, however, were 47% correctly classified.

4.0 Second Model: Bayesian Soil Landscape Model

The second soil-landscape model was developed using Expector. This is a knowledge-based soil-attribute mapping package that mimics soil surveyor thought (Corner *et al.*, 1997). It is based on bayesian

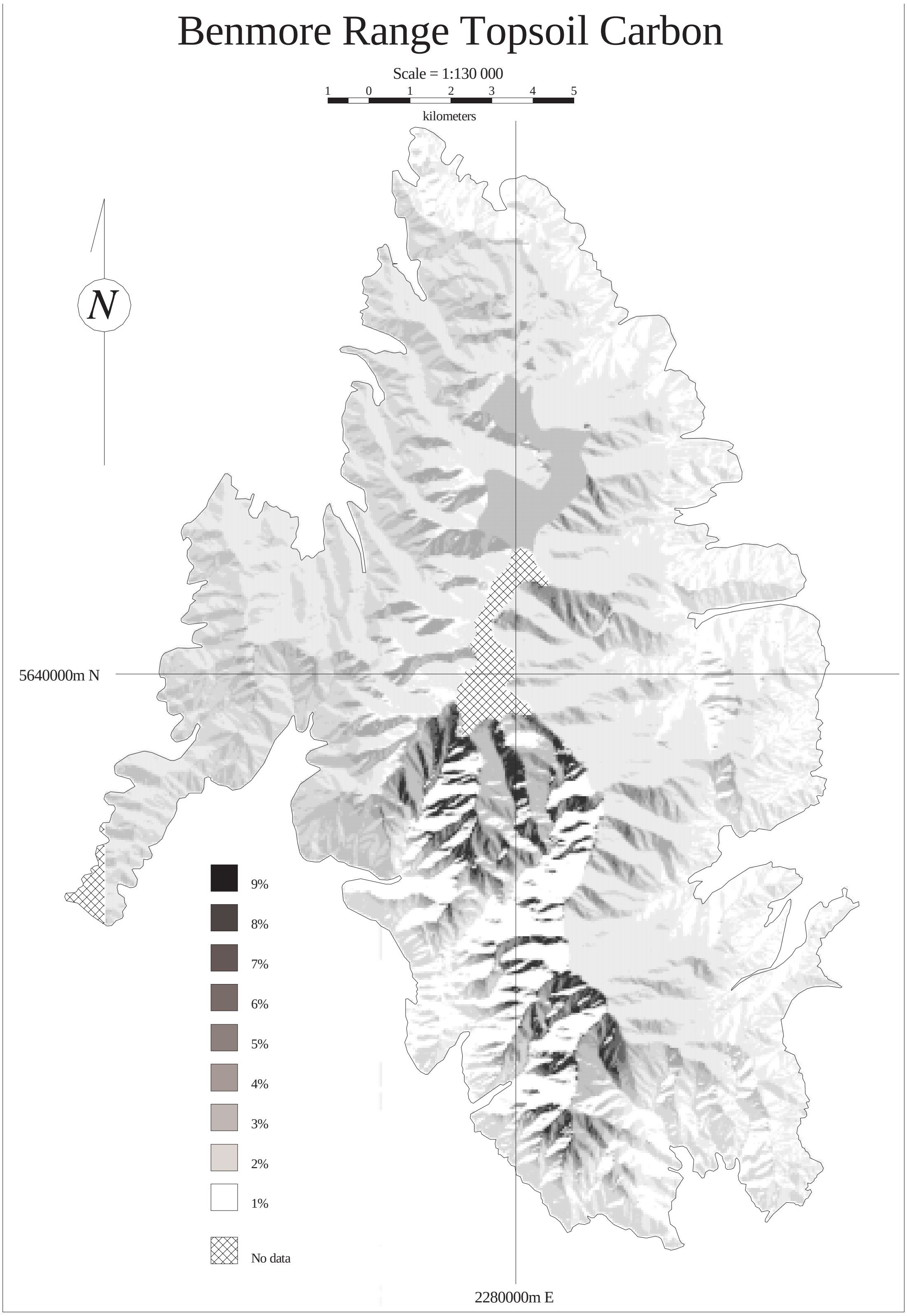

Figure 2: Map of predicted levels of carbon content

analysis and is able to utilise the three kinds of observational data gathered by the soil surveyor as well as the experiential knowledge carried as conceptual soil-landscape models. It is also able to incorporate knowledge of sampling bias and map purity.

In Expector the soil attribute to be mapped is chosen and its expected range is divided into classes. Each class is termed a hypothesis class. The ultimate output is in the form of a raster map — one for each hypothesis class. Expector generates maps of the whole area showing the probabilities of occurrence of each hypothesis class. Pixels are assigned the probability that the mapped soil attribute lies within the appropriate hypothesis class.

Three kinds of information are required in the use of Expector (Corner and Moore, 1997). The first is data on values of the attribute sampled in the mapping area. These values are treated as "prior knowledge" of the map-wide probability of the hypothesis classes. If no data is recorded, the analysis can still proceed with prior knowledge, which is based solely on what the mapper might expect from experience. However, it is preferable that prior knowledge is based on an unbiased sample of attribute values.

The second kind of information is in the form of spatially represented evidence layers that influence the hypothesis to some degree. The evidence layers must be conditionally independent. For each evidence layer, values are grouped into classes that are appropriate for making statements about the relationships between the evidence and hypothesis classes. The map-wide probabilities for each evidence class are constrained by the prior probabilities (prior knowledge) of the evidence classes.

The third kind of information is an expression of map purity for each evidence layer. Map purity is the probability that areas are incorrectly mapped as another class. For example, in an evidence layer containing aspect, a pixel with a southerly aspect has a moderate probability of being incorrectly classified as south-west but a low probability of being incorrectly classified as north.

We tested Expector on a major ridge on the eastern side of the Benmore Range, at Glencairn Station. The tussock grassland vegetation is depleted on lower slopes and on slopes exposed to the north-west, but there is moderate tussock cover elsewhere. Previous work by McIntosh *et al.*(1996) indicated a range of soil health conditions. Our goal was to map areas that had poor soil health due to low carbon levels, and areas most at risk of poor soil health due to marginal carbon levels.

In our test we divided topsoil carbon into four classes (Table 2). Topsoil carbon was sampled at 100 sites in the mapping area and map-wide probabilities of the hypothesis classes calculated.

Five evidence layers were chosen that could be used to predict carbon content. They were elevation, solar radiation, contour curvature, soil order, and vegetation class. Elevation, solar radiation and contour curvature were derived respectively from 5-m, 25-m and 10-m interpolations of a 20-m contour dataset. The solar radiation was inferred from daily accumulated sun illumination at the equinox (Kumar *et al.*,1997). The soil class was derived from a soil map of Glencairn Station. The vegetation map was derived from aerial photo interpretation together with 100 observations of ground cover at the soil carbon sampling points.

Topsoil carbon (%)	**Carbon class**	**Significance of the topsoil carbon classes**	**Map-wide probability of topsoil carbon classes.**
0 – 1.9	1	Very low carbon indicating poor soil health	0.4
2.0 – 2.9	2	Low carbon indicating soils at risk of poor soil health	0.18
3.0 – 4.0	3	Moderate carbon	0.23
> 4.0+	4	High carbon	0.19

Table 2: The Expector hypothesis expressed as classes of topsoil carbon, and the prior knowledge of the map-wide probability of those classes analysed from 100 soil samples in the map area.

Class	Elevation (m)	Change in probability of (%) carbon hypothesis class compared to the marginal probability			
		0 – 2.0	2.0 – 2.5	2.5 – 4.0	4.0 – 100
High	>1000	-23	0	+6	+17
Moderate	600 – 1000	-4	-1	-1	17
Low	< 600	+13	+1	-2	-12

Table 3. Evidence layer classes for elevation and their relationship to the hypothesis classes.

The classes of one of the evidence layers (elevation), and an indication of their relationship to the hypothesis classes are shown in Table 3. A "marginal probability" is where there is no knowledge of the relationship between an evidence class and a hypothesis class. Knowledge is expressed by modifying the marginal probabilities. In Table 3, an indication of the relationship between an evidence class and a hypothesis classes is indicated by stating the change in the assigned joint probability compared with the marginal probability. A decrease expresses knowledge that the carbon class is less likely to be represented in that evidence class, and an increase expresses a greater likelihood of such representation.

Data prepared in Expector on joint probabilities and map purity for each evidence layer is passed to the host GIS (Arc/Info‘), which combines the probability estimates from all evidence layers to produce a suite of maps showing the variable probabilities of occurrence of each of the hypothesis classes. Figure 3 shows the probability of each of the four classes of carbon, where white is low probability and black is high probability.

We retained 20% of the original Glencairn data set for validity testing. In the first run 50 % of predicted

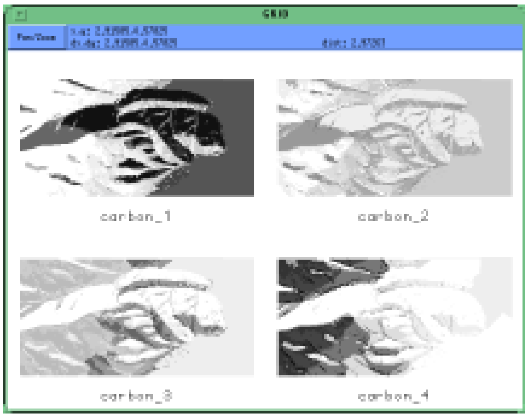

Figure 3: Probability of each class of carbon.

Carbon Class	**1**	**2**	**3**	**4**
1	7	0	1	0
2	2	0	0	0
3	2	0	1	3
4	0	0	2	2
% correctly classified	63.6	-	25	40

Table 4: Frequencies of ground truth class v predicted class

probabilities agreed with the data set (Table 4). The columns refer to the predicted class, i.e., the class with the highest probability. The rows refer to the ground-truth, i.e., the class the site belongs to.

Few sites (15%) were grossly misclassified: two carbon class 3 sites were misclassified as class 1, and one carbon class 1 site was misclassified as class 3.

5.0 Discussion

Both of the above studies quantify a conceptual soil-landscape model in order to make predictions of carbon content of topsoil from limited observations. The first model combined results from 72 soil pits and topography (aspect and elevation) to make predictions over the landscape. Expert knowledge was used to refine the relationships between topography and carbon. However, this unstructured knowledge does not lend itself to being automated or stored for later revision. The interpolation problem was solved by building a series of TINs representing topographic space. While this is appropriate for two dimensions (i.e., elevation and aspect), the same technique could not be used to interpolate in a 3-dimensional (or higher) space. While the model certainly appears to provide considerable detail of topsoil carbon content, it is not yet clear how many pits are required to achieve an acceptable level of accuracy, nor whether the two topographic variables are sufficient to explain the variability of carbon.

The second model allowed us to structure expert knowledge into the form of joint distribution functions between layers of evidence and the predicted attribute of interest — in this case carbon content. Expector allows the user to choose the balance between the information provided by an unbiassed sample dataset and expert knowledge, by overriding the probabilities calculated using the sample dataset. There is however no mechanism for storing the reasons why probabilities were changed to aid revision at a later stage. The resulting probabilities intuitively seem to provide more information than a simple prediction, including an indication of the uncertainty of the prediction. These probabilities could also be seen as fuzzy membership vectors and used in fuzzy logic (Buick and Lilburne, 1995; Zhu, 1997). However, the developed model is only suitable for relatively fine-scale modelling, in that some of the evidence layers are not readily available over large areas. In particular, vegetation and soil type were obtained through time-consuming aerial photography and a manual analysis of landforms (McIntosh *et al.* (1995)).

Both models incorporated mathematical functions and a GIS to extend New Zealand's soil data. The spade, auger and experience were required to derive the models, but these models combined with GIS technology then allowed predictions to be made over a larger area. Modelling soil attribute predictions also allows a continuous representation of carbon to be created.

6.0 Future Plans

Both models will be tested further and refined by obtaining data from other areas. The importance of each of the input variables at different scales of interest e.g., field (ridge), farm (part of range), regional (multiple ranges), needs to be explored further. We plan to use Expector to model the whole of Benmore Range. We also intend to use it to map classes of risk of soil compaction by dairy cows in the Waikato. GIS-driven soil-landscape models will surely be an important part of soil mapping in the next decade.

Acknowledgement

Our thanks go to Trevor Webb, James Barringer and Christine Bezar for their useful comments. The DTM and solar radiation layers were created by James Barringer. This research was funded by the Foundation for Science, Research and Technology under research contracts C09811 and C09626. Land Information New Zealand (Digital License Number TD0098872/92)

References

Buick, R. D. and L. R. Lilburne (1995) An object-oriented spatial decision support system (SDSS) for planning more sustainable land use on high country farms. *Proceedings of AURISA/SIRC '95*, Massey, 26-28 April 1995, pp. 311-322.

Corner, R. J., S. E. Cook and G. A. Moore (1997) Expector: a knowledge based soil attribute mapping method. ACLEP newsletter, June, pp. 9-11.CSIRO Land and Water, Canberra.

Corner, R. J. and G. Moore (1997) The Expector soil attribute mapping method: software for quantitative evaluation. CSIRO Land and Water, Canberra.

Hewitt, A.E. (1994) Introduction to soil-landscape models. In Webb, T.H. (editor) Soil-Landscape Modelling in New Zealand. Landcare Research Science series No.5. pp. 6-11.

Hewitt, A.E. (1995) Soils of the Conroy Land System, Central Otago. Landcare Research Science series No.13.

Kumar, L., A. K. Skidmore, and E. Knowles. (1997). Modelling topographic variation in solar radiation in a GIS environment. *International journal of geographical information systems*, 11(5), p. 475-497.

McIntosh, P. D., I. Lynn, L. Lilburne, M. Kingsbury and K. Giddens (1998) Is the soil carbon resource of dry South Island mountain ranges being accurately assessed. *Soil News* 46(1), pp. 13-15.

McIntosh, P. D., I. Lynn and T. H. Webb (1995) Soil map of Glencairn Station, Omarama 1:25 000, Landcare Research, Lincoln, NZ.

McIntosh, P.D., G. Ogle, R. Patterson, B. Aubrey, J. Morriss, K. Giddens (1996) Changes of soil nutrients and sustainability of pastoralism on grazed hilly and steep land, South Island, New Zealand. *Journal of Range Management 49*, pp. 361-367.

Webb, T.H. (Ed) (1994) Soil-Landscape Modelling in New Zealand, Landcare Research Science series No.5. Manaaki Whenua Press, Lincoln.

Zhu, A.-X. (1997) A similarity model for representing soil spatial information. *Geoderma* 77, pp. 217-242.