

# S-map – a new soil database for New Zealand

Linda Lilburne, Allan Hewitt, Trevor H. Webb and Sam Carrick

Landcare Research, P. O. Box 69, Lincoln, Canterbury, New Zealand. Email [lilburnel@landcareresearch.co.nz](mailto:lilburnel@landcareresearch.co.nz)

## Abstract

Soil information in New Zealand comprises maps and databases of varying accuracy and platforms. A project has been initiated to bring together this soil information in S-map – a new multi-layer soil database with national coverage. An important difference is that S-map is *designed* as a digital product and is thus not constrained by cartographic requirements. Rather than generalising polygons for aesthetic reasons, the best available information is now used regardless of its scale of origin. Also climate and topography are no longer used as soil descriptors. Key goals are to capture the fundamental soil inputs required by pedotransfer functions, and to record knowledge of uncertainty. This will then act as a platform for fusion with other environmental layers for spatial modelling.

S-map will contain data predominantly new to the existing national spatial soil database. Gaps will be filled and previous low-resolution data will be upgraded. The aim is to provide digital data layers of soil classes and attributes with resolution at least equivalent to 1:50 000. Work in the lowlands is proceeding using conventional soil survey techniques. In the uplands (where relief is sufficient to make effective use of the national 25-m resolution DEM) modelling approaches are used with the technique depending on data type and terrain.

There are some enormous challenges in this work including nationwide soil correlation, devising a common structure for lowland (good information) and upland (little information) soils, providing information about soil variability and uncertainty, and developing a model-base of pedotransfer functions that can be easily updated. This paper describes our approach and progress in solving these challenges.

## Key Words

Correlation, functional horizon, pedotransfer function, soil survey

## Introduction

New Zealand has a large accumulation of soil data in the form of soil maps acquired over the last 70 years. Most soil surveys have been undertaken prior to the advent of computer programs and simulation models, and soil data have largely been collected in the context of interpretation for land use suitability. In this context a qualitative estimate of such attributes as drainage class or soil depth was considered to be adequate to provide guidance on land-use issues. These data have served the agricultural community well but recent studies have shown that such qualitative data have been inappropriately used as input to simulation models (Fortin and Moon 1999). Modern users of soil information require more quantitative soil data that can be used in crop-growth and environmental-risk simulation models. Further problems with historical soil surveys are the generalised definition of soil series, the proliferation of soil series (many of which appear to be identical to one another), inconsistency between survey maps, and the lack of data on map unit composition.

Most soil maps in New Zealand have been made under guidelines defined in *Soil Survey Method* (Taylor and Pohlen 1970). Simple map units are defined as delineations where at least 85% of the area contains soils conforming to the definition of a single soil type. Compound map units comprise more than one named soil type. Unhappily, this standard of map purity has rarely been attained. Studies have shown that there is considerable variability in soil properties and soil types within soil map units on alluvial plains in New Zealand (Adams and Wilde 1976, Di and Kemp 1989, Karageorgis 1980, Webb *et al.* 2000). In the USA, where map units were based on the same principles of purity, Brubaker and Hallmark (1991) list a number of studies where taxonomic purity levels were found to be around 50% or less. In general, knowledge of soil variability gained in the course of soil surveys has not been captured on the resulting soil maps and reports. Most map units are simply described as being represented by a typical or modal profile, leaving the user with no information on soil variability. Rogowski and Wolf (1994) believed that, in relation to using soil data in models, the lack of data on variability was “perhaps the most serious limitation of the current soil survey process”.

The current national soil map is contained within *New Zealand Land Resource Inventory* (NZLRI) (Newsome 1992), which has recently been enhanced by the addition of 15 soil properties collectively known as the fundamental soil layers (FSL) (Wilde *et al.* 2000). The NZLRI was compiled at a scale of 1:63 360 from a range of soil maps, mostly pre-dating 1979. This required simplification of more-detailed survey polygons; consequently it does not contain the best available linework. In some areas the only source of information is the General Soil Survey maps of New Zealand (1:253 440 scale). Variability information is limited to identifying two component soils in any map unit. No confidence or reliability information is provided. Confidence and variability information is supplied with the FSLs but again the best knowledge is not available due to the use of predefined classes. Neither the NZLRI nor the FSLs have the flexibility to address the needs listed above.

S-map – a new spatial soil database for New Zealand – represents the response of Landcare Research in addressing this changing need for more quantitative soil information. This paper reports on progress to date in designing S-map. S-map builds upon existing spatial data yet makes some significant advances towards a soil database for the 21<sup>st</sup> century.

### **Principles of S-map**

S-map has 5 founding principles:

1. Describes soil (to a depth of 1 m) – soil should be separated from other environmental characteristics.
2. Digital format – thus releasing cartographic constraints on map unit depiction.
3. National soil correlation – thus reducing the number of defined soil series.
4. Incorporates knowledge of map unit variability and uncertainty.
5. Development of a soil database platform suited to modelling.

#### *1. Describes soil*

S-map focuses on soil information. Parent material is included as it is part of the soil classification. Other closely related environmental data, such as climate and topography-related variables (e.g. rainfall, slope), will no longer be part of the soil database, nor will they be used directly as soil classifiers. Even landform type will no longer be explicitly specified. It is expected S-map will be combined with these other environmental layers, as appropriate, to produce spatial layers of interest. For example, a simple model of water deficit would include rainfall, evapotranspiration, vegetation and soil variables.

#### *2. Digital format*

In the past cartographic constraints have dictated the level of spatial and thematic detail, i.e. polygon size and orientation, and label information. For example, long thin horizontal polygons were more acceptable than vertical ones as the former could more easily be labelled. The need for an aesthetically consistent hardcopy product meant that more detailed linework was spatially generalised, such that in some areas the national soil layer did not contain as much information, or was not as accurate, as the source soil survey. As S-map is primarily a digital product, it can now contain the best available soil information at any given location. However, surveys more detailed than 1:15 000 will not be included. The digital format also enables more flexibility in what information is conveyed in polygon labels.

#### *3. National soil correlation*

A third principle is to develop a national legend through a clear definition of soil taxonomic units, thus reducing the large number of soil series. Currently, soils with very similar properties have been recognised as unique soil series because of subtle differences in geomorphology, climate, parent materials or geography. The correlation of soils from widely separated geographic locations (such as North Island and South Island) may cause some confusion for local soil map users, as many soil series names will disappear. However, there are great benefits from this process. It will better support technology transfer and enable easier pooling of data for application of pedo-transfer functions to unsampled soils. A strict definition and application of the soil taxonomy criteria will also require the splitting or realigning of many current soil series, where the current series spans the new taxonomic criteria.

#### *4. Incorporate map unit variability and uncertainty*

To date little information has been provided on map unit variability or uncertainty in New Zealand. However, pedologists have accumulated much more knowledge of soil variability from considerable fieldwork over several decades. It is important to record this knowledge before it is lost due to retirement of key personnel. Important knowledge to record is classification reliability, and likely map unit composition (including proportions and alternatives). Furthermore, soil properties should be quantitatively characterised in a way that best describes what is known about each polygon. This necessitates a flexible structure that is not based on predefined ranges or depths. This structure must be the same for soils for which much is known (lowland soils) as for upland soils where little is known. The latter will be less precisely and accurately defined than the former.

Storing information about soil variability and uncertainty will allow environmental modellers to be more aware of the limitations of soil data.

#### *5. Development of a soil database platform*

Soil data need to be provided as a comprehensive platform for modelling of any environmental state, suitability or vulnerability that involves the soil. S-map will be designed so that it can be readily integrated with other key data sources. A strong link with the national soils database (NSD) is essential so that future pedological modellers can develop additional spatial layers of specific soil properties based on analyses in the NSD. It is expected that the combination of S-map and the NSD will significantly enhance the value of both databases. Finally, a modelling platform requires knowledge to be stored in formalised fields in a database. Use of textual or descriptive information as found in the NSD or in map legends can be very difficult to interpret automatically, and thus will be avoided.

#### **Design of S-map**

S-map soil data are polygon-based despite much attention in the literature to modelling of soil attribute surfaces (Burrough *et al.*, 1997, Moore *et al.*, 1993). We recognise the power of statistical interpolation techniques and their importance, and are designing S-map to underpin such techniques. A polygon base is required for the following reasons. First, point data are limited at a national scale and at a sampling density that will not allow generation of high-resolution surfaces by analysis of point data alone. Second, there is much useful polygon-based information contained within existing hard copy soil survey reports, and as expert knowledge, that needs to be made available for digital analysis. Third, the use of polygons operationally suits the description of many landforms in New Zealand that have relatively well-defined boundaries but contain high soil variability, for example the fans and terraces.

#### *Soil correlation*

The classification concept in S-map focuses on the practical performance or utility of the soil. Soils will be correlated across New Zealand according to “families” and “siblings”. Soils must be identified as belonging to one family and one sibling. A family is defined as any unique combination of subgroup (Hewitt 1998) and specified soilforms (Clayden and Webb 1994). The latter includes parent material, texture group and permeability (some parent material and texture categories have been aggregated). Where classification is uncertain, this is expressed for each level in the classification and the most likely alternatives can be specified.

Soil family names will generally be the most well known series name or the one with the largest coverage. However, the selected series name should also represent the mode of the family. The name is suffixed with “-ian” or “-ion” as per the *Soil Survey Method* (Soil Survey Staff 1951), e.g. Egmontion. The family will also be identified with a 4-character code. However, it is recognised that soil series names that are well known or described in the literature etc. are very important reference points for users of soil information. So a compromise position is proposed in which soil names that are important locally will still be able to be identified on maps and in databases. It is expected that minor soil names that are not used in any literature will disappear if the soil is correlated with another better-known soil.

Families are subdivided into siblings on the basis of any unique combination of classes of drainage, stoniness, depth, and texture (reluctantly we are allowing an option for additional undefined characteristics that can be justified for local needs). Map units can only be delineated according to these predefined classes. Siblings will be labelled with a unique code [soil family code][n] where n = 1, 2,...

The number has no meaning but “meaningful” labels can be generated on the fly as required from the database, e.g. [soil family code][depth code][texture code][drainage code].

Each sibling will be associated with a set of up to 5 functional horizons defined in terms of thickness, consistence and ped size (modified from Webb 2003).

#### *S-map property layers*

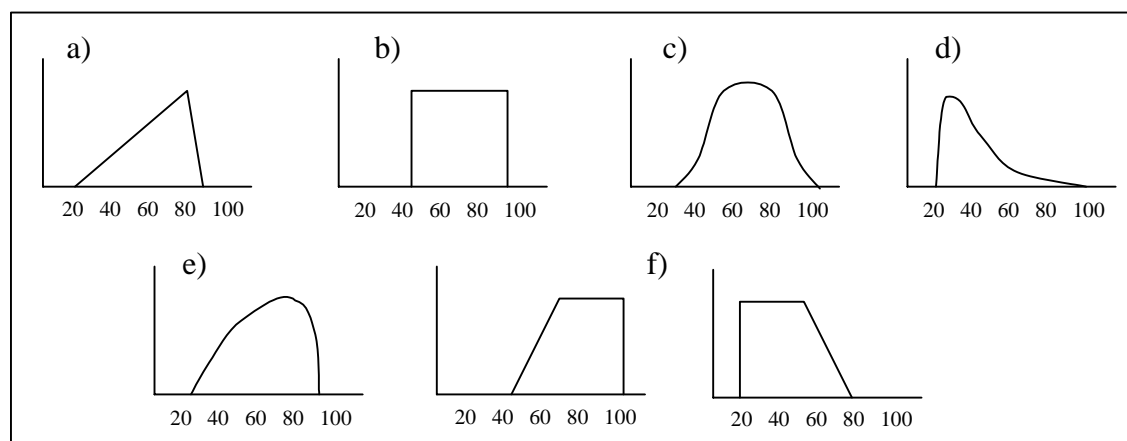
S-map thematic data are divided into 4 types: correlation properties, base properties, derived properties and interpretations (Table 1). Correlation properties (i.e. family and sibling criteria) are used to identify each soil in terms of the soil taxonomic class that best describes the modal characteristics of the soil. Base properties are described according to the known ranges of these attributes – as opposed to being constrained by predefined classification categories. These ranges might be smaller or wider than the predefined classes. Correlation and base properties are attributes in database tables that are linked to the soil polygons. Both are the result of expert knowledge and/or field observations so are manually specified. Slope is included only to allow existing slope data to be retained.

Derived properties are spatial layers (vector or raster) that may or may not result from the combination of other spatial layers with the soil polygon layer. Derived layers can be generated automatically as their derivation must be recorded in a model. This model can take any form including a rulebase, a statistical metric based on measured data, pedotransfer function, or a lookup table. The requirement for a model will allow these layers to be readily updated with advances in data or knowledge of soil distribution, and very importantly, they will record the rationale behind the information. Interpreted layers are also generated from a model and represent knowledge about the environmental risks, resource capacity or sustainability of various land management practices.

**Table 1 The four types of thematic data**

Type	Description	Attributes	
Correlation	Taxonomic class (predefined) of a soil polygon	<b>NZSC</b> <ul style="list-style-type: none"> <li>• Order</li> <li>• Group</li> <li>• Subgroup</li> </ul>	} family
		<b>Soilform</b> <ul style="list-style-type: none"> <li>• Parent material</li> <li>• Rock class</li> <li>• Texture group</li> <li>• Permeability</li> </ul>	
Base	Specified manually using expert knowledge and/or data	<b>Other</b> <ul style="list-style-type: none"> <li>• Depth class</li> <li>• Stoniness class</li> <li>• Texture</li> <li>• Drainage</li> </ul>	} sibling
		<b>Miscellaneous</b>	
		<b>Accessory attributes:</b> <ul style="list-style-type: none"> <li>• Slope (optional)</li> <li>• Up to 5 functional horizons</li> </ul>	
		<b>Depth</b> <ul style="list-style-type: none"> <li>• Rooting</li> <li>• Diggability</li> <li>• To slowly permeable layer</li> </ul>	
		<b>Rooting barrier</b> For each functional horizon <ul style="list-style-type: none"> <li>• Thickness</li> <li>• Stoniness</li> <li>• Clay content</li> <li>• Sand content</li> </ul>	

Derived	Generated from a model or pedotransfer function that predicts the attribute	Profile Available Water (PAW) Preferential flow For each functional horizon <ul style="list-style-type: none"> <li>• Field capacity</li> <li>• Wilting point</li> <li>• Total porosity</li> <li>• Macroporosity</li> <li>• Bulk density</li> <li>• Total carbon</li> <li>• Total nitrogen</li> <li>• P (H<sub>2</sub>SO<sub>4</sub>)</li> <li>• Ca</li> <li>• CEC</li> <li>• pH</li> <li>• P retention</li> </ul>
Interpreted	A risk, vulnerability, suitability or other environmental or resource capacity map	e.g. soil compaction risk, leaching risk, best management practices for dairy effluent discharge, crop suitability, biosecurity risk, biodiversity restoration zone



### Variability and uncertainty

Known soil variability within polygons is primarily described using combinations of siblings, and probability distribution functions (pdfs) of each soil property. Polygons can contain up to 5 siblings to describe the soils that are likely to be present in a polygon. The likely proportion of each sibling is given (see Table 2 for default proportions). For example, a map unit might be identified as being a TMPT8+EYRE6 (70:30), i.e. 70% is a Templetonion sibling and 30% is an Eyrion sibling. Siblings estimated as making up less than 10% of a map unit will not be listed.

**Table 2. Proposed default proportions for the number of siblings in a map unit**

No. of siblings	Default proportion
1	100
2	60 : 40
3	50 : 30 : 20
4	40 : 30 : 20 : 10
5	40 : 20 : 15 : 15 : 10

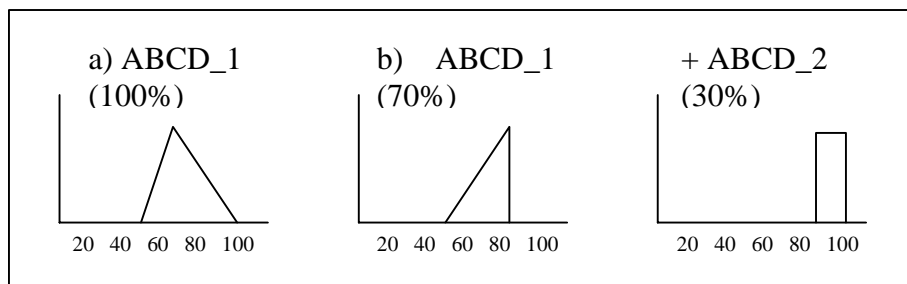
Variability information will be attached to map units in the form of probability distribution functions (quantitative) or alternatives (qualitative). Each sibling will have a default pdf for each property/functional horizon. Some of these will be based on the taxonomic definition of the sibling. These

defaults should be overridden wherever there is knowledge that the properties of a polygon, group of polygons, or map unit are either less or more precise than the sibling default.

Best available expert knowledge and observed or measured data will be used to specify the pdfs. Most pdfs will be specified using a triangular distribution (min, mode and max) or a uniform distribution (min, max) (Figure 1a, b). Alternatively, a normal (Figure 1c), lognormal (Figure 1d) or a beta function (Figure 1e) can be used. An irregular pdf in the form of a trapezoid is also possible – although this can take only the forms shown in Figure 1f.

**Figure 1. Standard probability functions that can be used to describe variability of a soil property. A) triangular, b) uniform, c) normal, d) lognormal, e) beta and f) irregular.**

Each pdf will describe the likely distribution of attribute values within the sibling part of the given map unit. For example, a map unit that is defined as being an ABCD\_1 sibling might have a triangular depth pdf (cm) of [45, 60, 100] (Figure 2a). A map unit that is defined as being ABCD\_1 (70%) + ABCD\_2 (30%), where the siblings are defined as having different depths (moderately deep, i.e. 45–90 cm, and deep (>90 cm) respectively), would have two sibling pdfs of [45, 90, 90] and [90, 100] (Figure 2b). In the former the pdf expresses the soil expert's belief that the map unit is likely to have a very small proportion of deep soils although this has not been explicitly specified as it has in the latter case, where the deep soil has been identified as being present with a 30% probability.



**Figure 2. Example of depth pdfs associated with a) a single sibling map unit, and b) a double sibling map unit.**

S-map has provided for the description of soil variability but there is still the issue of how certain or accurate this information is. S-map allows estimates of uncertainty and error to be recorded in the following ways:

1. Confidence of taxonomic classification in each correlation property
2. Possible alternative taxonomic classifications
3. Confidence in the base property data
4. Reliability of the spatial linework.

The confidence ratings indicate whether the correlation classification (or base property pdf) is based on a reasonable number of auger and pit observations, is poorly sampled, or not sampled at all.

Reliability or confidence in spatial linework is initially based on a survey confidence rating. Survey confidence is determined by landscape predictability and survey quality, i.e. standard observation density and linework/registration quality. For example, in an “A” survey it can be expected that in a single symbol soil polygon, the soil will be found in at least 80% of the polygon, whereas in a “B” survey the soil may only be found in 60% of the polygon.

#### *Procedure*

Two approaches will be used to develop S-map – these differ according to the terrain:

- Lowlands, dominantly flat to rolling land. Landforms are of such low relief that digital elevation models (DEM: based on current 20-m contour data) cannot be used for soil-landscape modelling. Soil mapping uses conventional methods, based on air-photo interpretation and free survey techniques.
- Uplands, dominantly hill and mountain terrain. Relief allows application of soil-landscape modelling based on DEM and other spatial information. The actual modelling used will depend on the land system, the sampling cost and availability of data. The predominant technique will be to derive soil distribution rules from available data, literature and new sampling, and apply these to modelled

landform land elements. These are generated by analysing a DEM to separate spurs, noses and ridges from back and side slopes (Schmidt and Hewitt 2004).

All polygons will be mapped (either from existing survey maps or newly mapped) and associated with one or more siblings. Strict application of the family and sibling criteria will ensure national consistency. All correlation and base properties will be specified by a pedologist when the polygon is mapped, as will the relevant uncertainty information. Initially mapping will focus on Northland, Wairarapa, Motueka, and Canterbury. Some models for derived and interpretative layers have been developed. These will be built into S-map later.

### **Discussion**

S-map is a significant step in the process moving from the conceptual taxonomic definitions of the previous system (the basis of the current national soil map, the NZLRI), to the more clearly defined taxa in the new NZSC (Hewitt 1998). This more quantitative classification system enables more confidence in prediction of a number of soil properties. However, this translation will require existing series to be merged and split to match the new taxonomic criteria. For example, some of the map units classified as being part of the Becks series will need to be relabelled as belonging to a Gley family/sibling rather than a Semiarid one like most of the Becks map units. The work that has been done so far in correlating soils between islands has already identified that several minor modifications to the NZSC will be needed, for example a new parent material class to separate layered gravels from non-layered stony soils.

Another major advance of S-map from the NZLRI is that S-map will contain the best available data. Users who require national-scale soil data (i.e. the NZLRI) cannot easily access detailed soil survey information (e.g. Part Paparua County (Cox 1978)), or more recent regional surveys (e.g. Plains and Downs: Webb, in prep.). These surveys contain considerably more spatial detail than the NZLRI. The NZLRI will still be available for those for whom cartographic consistency is more important than accuracy. A much-improved ability to record what is known about spatial variability and uncertainty is also a key advance from the NZLRI. Pedological advances include the nationwide correlation and application of NZSC, and use of functional horizons.

S-map has been designed to be a platform for future modelling. By linking every polygon to a family, sibling, and set of functional horizons, modellers will be better able to predict other soil properties, especially once the NSD profiles are similarly classified. Base properties are ones that are very important inputs for modelling but are not themselves easily predicted, hence are derived from expert knowledge of map units. The functional horizons will be the main carrier of a number of soil physical data. Some physical data are poorly correlated with functional horizons and will need to be estimated from other attributes. For example, field capacity and wilting point will be estimated from bulk density, clay, sand, and carbon content. Soil chemical properties are best predicted using a subset of attributes used in the definition of siblings plus other features such as land use and climate.

Incorporating uncertainty into S-map appreciably complicates the database structure as every soil property will vary within polygons, and all estimates will be uncertain. So for one soil value, up to 5 additional values may be needed to describe the variability and uncertainty (up to 4 parameters for the pdf and one reliability indicator). Furthermore, these variability estimates will vary spatially. For example, map units that originate from a detailed survey and from the general island-wide soil survey might have the same sibling classification but very different pdfs.

A model database will be required to manage the various models that are used to generate the derived and interpreted layers. Each model needs to be clearly documented as to its strengths and limitations. For example, scale is an important issue, as a regional-scale model of carbon, for example, could easily look quite different to a national model. An essential part of the documentation will be model error information (e.g. goodness of fit, standard error estimates, and a spatial layer of error).

### **Conclusions**

Designing a soil database to meet modern demands for soil information that in part relies upon historical survey data is a challenging task. This task covers issues that range from national-level correlation, reconciling raster modelling techniques with expert knowledge of soil survey polygons, and devising a

structure that can be used for both well- and little-known areas. Another key difficulty is finding the balance between recording estimates of variability and uncertainty of as many key soil properties as possible, yet not setting too onerous a task for the pedologists or confusion for the users of soil data. We hope the result will be a flexible and comprehensive database that can provide quantitative information as required by modellers, and the best-available soil data (whether measured, estimated or interpreted) for use by land managers and policy analysts.

### Acknowledgements

Funding for this work was provided by Foundation for Research Science and Technology, contract CO9X0306. We are grateful to Peter Almond, Hugh Wilde and Ian Lynn for their comments on the manuscript.

### References

- Adams JA, Wilde RH (1976) Variability within a soil mapping unit mapped at the soil type level in the Wanganui district. *New Zealand Journal of Agricultural Research*, **19**, 165-176.
- Brubaker SC, Hallmark CT (1991) A comparison of statistical methods for evaluating map unit composition. In 'Spatial variabilities of soils and landforms'. pp. 73-88. (Soil Science Society of America: Madison, Wisconsin)
- Burrough PA, van Gaans PFN, Hootsman R (1997) Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma*, **77**, 115-135.
- Clayden B, Webb TH (1994) Criteria for defining the soilform - the fourth category of the New Zealand soil classification, Landcare Research Science Series No. 3. (Landcare Research: Lincoln, NZ)
- Cox JE (1978) Soils and agriculture of part Paparua County, NZ Soil Bureau Bulletin 34. (DSIR: Wellington)
- Di HJ, Kemp RA (1989) Variation in soil physical properties between and within morphologically defined series taxonomic units. *Australian Journal of Soil Research*, **27**, 259-273.
- Fortin M-C, Moon DE (1999) Errors associated with the use of soil survey data for estimating plant-available water at a regional scale. *Agronomy Journal*, **91**, 984-990.
- Hewitt AE (1998) 'New Zealand soil classification.' (Manaaki Whenua Press: Lincoln, NZ)
- Karageorgis D (1980) Soil variability and related crop productivity within a sample area of the Templeton soil mapping unit. (University of Canterbury: Christchurch, NZ)
- Moore ID, Gessler PE, Nielsen GA, Peterson GA (1993) Soil attribute prediction using terrain analysis. *Soil Science Soc. Am. J.*, **57**, 443-452.
- Newsome PFJ (1992) 'New Zealand land resource inventory Arc/Info data manual.' (Landcare Research: Lincoln, NZ)
- Rogowski AS, Wolf JK (1994) Incorporating variability into soil map unit delineations. *Soil Science Society of America Journal* **58**, 163-174.
- Schmidt J, Hewitt A (2004) Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma*, **121**, 243-256.
- Soil Survey Staff (1951) 'Soil survey manual 18.' (Soil Conservation Service, U.S. Department of Agriculture: Washington DC)
- Taylor NH, Pohlen IJ (1970) 'Soil survey method: a New Zealand handbook for the field study of soils.' NZ Soil Bureau Bulletin 25. (DSIR: Wellington)
- Webb TH (2003) Identification of functional horizons to predict physical properties for soils from alluvium in Canterbury, New Zealand. *Australian Journal of Soil Research*, **41**, 1005-1019.
- Webb TH, Claydon JJ, Harris SR (2000) Quantifying variability of soil physical properties within map units to address modern land-use issues on the Canterbury Plains, New Zealand. *Australian Journal of Soil Research*, **38**, 1115-1129.
- Wilde RH, Willoughby EJ, Hewitt AE (2000) 'Data manual for the national soils database spatial extension.' (Landcare Research: Lincoln, NZ)