# Aligning New Zealand digital soil mapping with the global soil mapping project

Allan Hewitt[A], James Barringer[A], Guy Forrester[A], Ian Lynn[A], Thomas Mayr[B], Stephen McNeil[A] and Trevor Webb[A]

[A]Landcare Research, Lincoln, New Zealand, Email hewitta@landcareresearch.co.nz
[B]National Soil Resources Institute, Cranfield University, UK, Email t.mayr@cranfield.ac.uk

**Abstract**
New Zealand is undertaking a new phase of soil mapping (S-map) with the goal of national soil map coverage at 1:50 000 scale. One S-map objective is to develop a national capability in digital soil mapping (DSM). Concurrent development of the global soil map (GSM) project provides an opportunity for the S-map project to align its work with the GSM initiative both to achieve an enhanced national soil map and contribute New Zealand data to the GSM project. Consequently we have applied GSM method guidelines nationally. We matched the generic GSM methods to 193 New Zealand soilscapes. Soilscapes with similar levels of legacy soil information were grouped into eight soilscape groups, and DSM methods applied to them. Four soilscape groups were in low relief land where intensification of land use generates high demand for new soil information. The other four method classes were in rolling, hilly and mountainous land. The most extensive method in low relief land requires establishment of better environmental covariate layers followed by soil map reference area extrapolation, and the most extensive method in high relief land involves sampling programmes and scorpan analysis. The proposed inference engines are Bayesian belief networks, SoLIM, and Random Forests.

**Key Words**
Digital soil mapping, soilscapes, soil mapping strategy, global soil map.

## Introduction
New Zealand (NZ) is undertaking a new phase of soil mapping (S-map project, Lilburne *et al*. 2004) to meet demands for soil information and especially provision of more accurate input data for a new generation of models for application in environmental policy development and land management. The S-map goal is for national soil map coverage at 1:50 000 scale. One of the S-map objectives is to develop a national capability in digital soil mapping (DSM). Concurrent development of the global soil map project provides an opportunity for the S-map project to align its work with the global soil map initiative to achieve an enhanced national soil map, to contribute NZ data to the global project, and to develop national capability in DSM. Consequently we have applied global soil map DSM method guidelines nationally. The purpose of this paper is to outline an adaption of the Minasny and McBratney (in press) global soil map method guidelines in developing a national DSM strategy. The global goal is to generate soil information as 90-m-resolution rasters. The NZ goal is to provide finer, 25-m-resolution rasters. The intention is to provide national DSM coverage using techniques and formats that will be compatible with global DSM coverage and from which global-resolution data for NZ may be easily extracted.

## Methods
*Soilscape definition and mapping*
The definition of NZ national soilscapes was explored by Hewitt *et al*. (in press) and completed in the work described in this paper (and available through www.landcareresearch.co.nz). We followed the soilscape definition of Lagacherie *et al*. (2001) where a soilscape is "a landscape unit including a limited number of soil classes that are geographically distributed according to an identifiable pattern". The quality of our national soilscape coverage will improve incrementally as DSM proceeds. First-approximation soilscapes are being used to plan DSM operations as in this paper. Subsequently it will be modified and improved using DSM results to derive a second-approximation coverage suitable for use as a more generalised representation of national soils. Hewitt *et al*. (in press) explored digital methods for generating soilscapes. They found that where they are available, legacy data and expert knowledge provided an efficient basis for generating a first approximation. Accordingly soilscapes for the South Island were based on the earlier map of "soil sets" for the South Island similar in concept to "land systems" (Soil Survey Staff 1968). Soilscapes for the North Island were based on a map of "erosion terrains" derived from the NZ Land Resource Inventory (NWASCO 1979). Although originally intended for erosion and sediment yield studies, the erosion terrains efficiently stratified soil patterns and rock types relevant to soilscapes mapping. For both islands, the soilscapes were

arranged in a hierarchy of six levels: level 1, land province – major climate, geologic terrains and landscape units; level 2, land region – major physiographic units; level 3, lithology – major rock and cover material types; level 4, climate; level 5, altitude; and level 6, slope and landforms. For the national-scale DSM strategy we used soilscapes at level 5 for the South Island and soilscapes at level 3 for the North Island. Climate (level 4) and altitude (level 5) were not used for the North Island because these factors were less variable and were of less significance than in the South Island. Level 6 slope and landform attributes were not used because they stratified finer scale variations considered more relevant for local rather than national planning. This provided 193 soilscapes for analysis nationally (52 for the North Island and 141 for the South Island).

*Data available for DSM in NZ*
Point data are mainly limited to analysed pedons of the National Soils Database (NSD). Data quality is high but the number of sampled sites is less than 3000. These tend to be clustered in former study areas. There are few areas where soil survey auger observations are available in digital form. Although point data are sparse, there is good remaining pedologist expertise. DSM methods that are able to incorporate expert knowledge are therefore favoured.

Environmental covariates of national extent include climate layers (Leathwick *et al.* 2002), a national digital elevation model at 25-m and 15-m resolution based on national-extent 1:50 000 scale topographic elevation maps (Barringer *et al*. 2008), and nearly completed 1:250 000 scale of geological coverage update (Nathan 1993). National-extent remote sensing imagery has been collated by the EcoStat project (Dymond and Shepherd 2004) and projects in support of Kyoto Protocol compliance.

A*llocation of DSM methods to soilscapes*
A major distinction is made between "lowlands", comprising plains or basins with flat to easy rolling slopes, and "uplands", comprising rolling, hilly, plateau and steep land. Uplands have sufficient relief to make effective use of the national 25-m-resolution DEM for soil–landscape modelling. Lowlands relief is insufficient for effective use of the DEM. There is also a lack in the lowlands of good environmental covariates of sufficient resolution to support DSM techniques.

Based on Minasny and McBratney (in press) we assigned the following land and soil information attributes to all soilscapes: (1) lowland or upland, (2) area of legacy soil maps, (3) soil survey quality classification assignment for all legacy soil maps in five classes based on age and map scale, (4) number of national soil database analysed profile sites, and (5) available expert knowledge either in the form of reports, soil–landscape models, or living people. We grouped the soilscapes with similar attributes into eight soilscape groups, and for each of these we assigned generic DSM methods.

*Port Hills trials for operational DSM methods*
A study area in the Port Hills adjacent to Christchurch City was chosen to test three possible inference engines; a Bayesian belief network, SoLIM (Zhu *et al.* 2001), and a classification using the Random Forests method (Breiman 2001). Belief networks enable excellent depiction and exploration of soil–landscape models and are therefore good for capturing expert knowledge. It is time-consuming to set up the networks, and they require categorical training data and can only predict categorical data. SoLIM is also well able to capture expert data. It can accept both categorical and point data but is limited to categorical data outputs. Random Forests is very fast and able to predict both categorical (i.e. classification) and numerical data (i.e. regression). It is more able than the other methods to handle large learning and processing databases, often with little or no built-in knowledge of the data under study or the relationship between data elements. Its main disadvantage, common with all general machine-learning methods (Hastie *et al*. 2001), is that it is a black box with no direct opportunity for expert knowledge, and it is often difficult to extract the reasoning for the derived relationships.

The legacy 1:25 000 scale Port Hills soil map and report is a high quality soil survey of hilly and steep land. The area is complex and presents a challenging test area for DSM and has been used as a test-bed for developing concepts of soil–landscape models (Webb 1994). Environmental covariates were derived from geology and climate DEM layers. The three inference engines were evaluated by comparing output soil classes to the original soil survey using confusion matrices. Judgement focused on the mapping of soil classes that were least likely to be wrongly mapped in the legacy soil survey, for example, the contrast between shallow soils derived from volcanic rock on shoulder landforms verses deep colluvium soils on steep talus cone aprons.

## Results

*Soilscape groups*

Table 1 summarises the soilscape groups, their definitions, and derived DSM methods modified from Minasny and McBratney (in press).

**Table 1. Soilscape groups and proposed DSM methods based on criteria modified from Minasny and McBratney (in press). Low = lowland, Up = upland, H = high, L = low, qual = quality, pt = point, NI = North Island, SI = South Island, ECovar = environmental covariates, Extrap = extrapolation**

| Soilscape group | Soilscape group definitions | DSM Methods | Area ($km^2$) NI | Area ($km^2$) SI |
|---|---|---|---|---|
| Group 1 | Low, H qual full cover map, good pt data | ECovar + Spatial disaggregation | - | 14788 |
| Group 2 | Low, H qual ref. area maps, sparse pt data | ECovar + Map Extrap | 24 559 | - |
| Group 3 | Low, L qual, ref. area maps, sparse pt data | ECovar + Training map + Extrap | 7 249 | 11647 |
| Group 4 | Low, L qual surveys, no pt data, | ECovar + Training map + Extrap | 57 | 1106 |
| Group 5 | Up, H qual ref. area maps, sparse pt data, | Map Extrap | 5 755 | 4383 |
| Group 6 | Up, L qual ref. area maps, sparse pt data, | Sampling + Scorpan | 23 673 | 46435 |
| Group 7 | Up, L qual ref. area maps, sparse pt data, | Sampling + Scorpan | 11 912 | |
| Group 8 | Up, no useful surveys no pt data | Sampling + Scorpan | 40 677 | 3264 |

Group 1 includes the alluvial outwash and loess-blanketed lowlands of the Southland and Canterbury plains. Soil map polygons incorporate associations and complexes of well-defined soil classes. Development of good quality environmental covariates is needed to spatially disaggregate these polygons. The NSD point data sites are not likely to be sufficient for application of scorpan kriging (McBratney *et al*. 2003) but they do provide good datasets for development of pedotransfer functions. Group 2 is the dominant lowland area that includes a wide range of alluvium, ash, loess, and sand dune soilscapes. High quality soil surveys have patchy distribution but provide good reference areas for spatial extrapolation. As for group 1, the development of spatial covariates is necessary to enable map extrapolation and spatial disaggregation of polygons. Groups 3 and 4 include land similar to group 2 but reference soil maps are of low quality. The groups have either sparse or no point data. A practical approach for both groups would be to choose and map reference training areas and then extrapolate to the full soilscape areas by the methods of group 2. If the primary application of the NSD point data is for the development of pedotransfer functions then the distinction between point data coverage between groups is not important because all point data would be pooled across all groups to develop functions. The areas of validation for these functions would most likely be independent of the soilscape groups.

In upland areas existing environmental covariates are generally suitable for DSM. Development of 5-m-resolution DEMs by ALOS Prism and radar imagery is proceeding to provide better land element discrimination in lower relief rolling and hill land. DEM derivatives are powerful soil predictors of soils in NZ because the NZ landscape and soil cover is predominantly of late-Pleistocene or Holocene in age and soils are closely related to landform position. Group 5 includes soilscapes with high coverage of quality reference area soil surveys that will provide a good basis for map extrapolation. Groups 6, 7 and 8 have either poor or unsuitable legacy soil surveys and sparse or no data. Much of the land is of low priority and a significant proportion has poor accessibility due to rugged terrain. Consequently mapping must be based on sampling programmes and scorpan modelling. Sampling strategies must take into account access.

Most of the soilscape classes are mapped as several delineations. Because of this a legacy soil map reference area may need to be extrapolated into non-contiguous unmapped areas. Feature space analysis may assist in confirming the integrity of soilscape delineations. Extrapolations will need to be validated.

*Port Hills trial*

Map outputs from the three inference engines tested on the Port Hills data were compared with the legacy soil map. The three methods had comparable performance in recognising major soil contrasts. However, Random Forests had superior performance overall. A likely explanation for this is that the Random Forests method involved data mining of soil map units. The other two methods, however, were more influenced by expert knowledge of the relationships of soil taxonomic units to what were considered key environmental covariates. Use of the legacy soil map as the standard may not then be a fair basis for comparison and suggests the need for an independent field sample based approach for a more accurate assessment.

**Conclusions**

- The global soil map has potential to provide opportunities for linkage with the international DSM community that that will help national as well as international goals.
- Development of better environmental covariates, particularly in the lowlands, is of high priority.
- As development of DSM capability is important there is need for the mapping team to gain experience by testing methods with each of the soilscape group areas.
- Because DSM techniques continue to advance it is likely that current techniques chosen will be superseded. Our choice therefore has to be provisional.
- The results of this analysis provide a basis for costing the DSM effort required in NZ.

**References**

Barringer JRF, Hewitt AE, Lynn IH, Schmidt J (2008) National mapping of landform elements in support of S-Map, a New Zealand soils database. In 'Advances in Digital Terrain Analysis'. (Eds Q Zhou, B Lees, G Tang) pp. 443–458. (Springer).

Breiman L (2001) Random Forests. *Machine Learning* **45**, 5–32.

Dymond JR, Shepherd, JD (2004) The spatial distribution of indigenous forest and it composition in the Wellington Region, New Zealand, from ETM+ satellite imagery. *Remote Sensing of Environment* **90**, 116–125.

Hastie T, Tibshirani R, Friendman J (2001) 'The Elements of Statistical Learning.' (Springer: New York).

Hewitt AE, Barringer JRF, Forrester GJ, McNeill SJ (in press) Soilscapes basis for digital soil mapping in New Zealand. In 'Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Third Global Workshop on Digital Soil Mapping in Logan, Utah, USA (DSM 2008)'. (Eds AE Hartemink, A McBratney, J Botenger) (Springer).

Lagacherie P, Robbez-Masson JM, Nguyen-The N, Barthès JP (2001) Mapping of reference area representativity using a mathematical soilscape distance. *Geoderma* **101**, 105–118.

Leathwick J, Morgan F, Wilson G, Rutledge D, McLeod M, Johnstone K (2002) 'Land Environments of New Zealand: A Technical Guide.' (Ministry for the Environment: Wellington).

Lilburne L, Hewitt AE, Webb TH, Carrick S (2004) S-map: a new soil database for New Zealand. In 'SuperSoil 2004 : 3rd Australian New Zealand Soils Conference" University of Sydney, Australia Sydney. P [unpaged]. Available: www.regional.org.au/au/asssi

McBratney AB, Minasny B, Mendonca Santos ML (2003) On digital soil mapping. *Geoderma* **117**, 3–52.

Minasny B, McBratney AB. (in press) Methodologies for global soil mapping. In: 'Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Third Global Workshop on Digital Soil Mapping in Logan, Utah, USA (DSM 2008)' (Eds AE Hartemink, A McBratney, J Botenger) (Springer).

Nathan S (1993) Revising the 1:250 000 Geological Map of New Zealand: a discussion paper. *Institute of Geological & Nuclear Sciences Scientific Report 93/*26.

NWASCO (1979) 'Our Land Resources.' Bulletin accompanying NZLRI worksheets. (National Water and Soil Conservation Organisation: Wellington).

Soil Survey Staff (1968) General survey of the soils of the South Island, New Zealand. *Soil Bureau Bulletin* **27**. (DSIR: Wellington). 404 p., including maps.

Webb TH (Ed.) (1994) Soil-landscape modelling in New Zealand : proceedings of a workshop held at Aokautere, New Zealand, 8–9 February 1993. *Landcare Research Science Series* **5**. (Manaaki Whenua Press: Lincoln).

Zhu AX, Hudson B, Burt J, Lubich K, Simonson D (2001) Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal* **65**, 1463–1472.