S-map Blueprint (v0.2)

What is S-Map?

A new multi-layer soil database, with national coverage, that incorporates the digital soil surveys. S-map is a digital product that comprises the best available data (1:15,000 and above) for any given area. New mapping will be done at 1:50,000 scale. In order to have national consistency, mapping criteria need to be specified, and a national legend devised. As a result some linework may be simplified (adjacent polygons merged). This will be minimised by having the option of an additional mapping criterion, so that information is not lost.

As S-map is a digital product, there will be no hardcopy outputs apart from promotional maps. S-map should be made available over the web. A full national legend could be produced in a booklet form. If RCs require hardcopy 1:50,000 maps (as per Otago Regional Council) then they must pay for the necessary generalisation and cartography.

Different approaches will be used to map lowland and upland soils but the same database structure will be used. The latter are defined as those areas where a DEM can be used for soil prediction. It is expected that upland soils will be less accurately mapped. Climate related properties (e.g., water deficit), age (e.g., terrace surface) and topography (e.g., hill/terrace/fan) will not be directly incorporated into S-map.

The Purpose of S-map

The main objectives of S-map are to:

- provide the best available soil information for any given location,
- separate soil information from other environmental variables e.g., climate.
- reduce the number of defined soils through correlation so that ‘more can be said about fewer soils’,
- record knowledge of soil variation and uncertainty,
- develop a platform that can be readily integrated with other databases (especially the NSD) for modelling soil properties, and environmental state or vulnerability.

S-map Correlation Design

The overall classification concept in S-map is the practical performance of the soil or its utility (as opposed to genetic theory). Soils will be correlated across New Zealand according to ‘families’ and ‘siblings’. Soils must be identified as belonging to one family and one sibling. This will allow more to be said about ‘fewer’ soils as information available for one soil can be applied to other soils in the same family/sibling. A family is any unique combination of the NZSC soil form classification according to the rules specified in Appendix A. A sibling is any unique combination of classes of drainage, stoniness, depth, texture, and an optional phase-like option. Each sibling will be associated with a single set of functional horizons or building blocks that describes the soil profile of the sibling. 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][slope code].

All soils with the same soilform as per the categories in Appendix A must have the same national legend soil family name. Criteria of landtype, depth, stoniness, texture, drainage, slope and miscellaneous variant, can all be used to identify and delineate siblings where appropriate. So adjacent polygons might have the same family label but different sibling attributes (e.g., different stone content). Soil polygons that have been identified as having more than one soil form yet have the same soil series name must be reclassified, e.g., Manawatu includes RFW, RFAW, RFT, RFMW polygons. Three of these will need to be correlated into different families and possibly given another series name if there is no local equivalent. Or they might be reclassified to have the same NZSC classification but some mapunits might have a different phase (e.g., those previously classified as RFMW might now be a RFW sibling with a mottled phase).

Similarly, where a soil series as currently classified spans more than one family and sibling, it can be reclassified. For example some of the polygons classified as being Becks will need to be spatially relabelled as belonging to a Gley family/sibling rather than a Semiarid one like most of the Becks mapunits. Alternatively, polygons can be associated with up to 5 siblings, which can be from the same family or different ones. So a mapunit can be associated with two siblings that cover the depth range of a series that spans two depth classes. This approach also allows complexes, associations and inclusions to be explicitly specified.

Soil family names will generally be the most leading or well known series name or the one with the most coverage. The selected series name will represent the mode of the family. The name is suffixed with "–ian" or "-ion" as per the Soil Survey Method, 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. So a compromise position is proposed in which soil names that are important locally could be lost in the national legend but an additional attribute associated with each polygon will contain the locally recognised series name. 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. Plots can be produced with mapunits labelled either by their soil family name or by their local series name. Similarly clicking a mapunit on the web will display both the family name and any local name.

Sibling attribute classes are for correlation purposes only. They represent the modal properties of the sibling. All of these correlation attributes will also be made available as soil property layers in which uncertainty is specified. The use of predefined classes and fixed depths will be minimised, allowing soils to be accurately described within limits of pedological knowledge.

S-map attributes

S-map attributes are divided into 4 types: correlation properties, base properties, derived properties and interpretations (Table 1). Correlation properties are used to identify each soil in terms the soil taxonomic class that best describes the modal characteristics of the soil. Base properties describe the soil as completely as knowledge allows. For example, quantitative properties are specified directly using a continuous range as opposed to being constrained by predefined classes.

Correlation and base properties are attributes in PAT tables that are linked to the soil polygons. Both are the result of expert knowledge so are manually specified. 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 is recorded in a model. This model can take any form including a rulebase, a statistical formula, pedotransfer function, or a lookup table. 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 or attribute data

Type	Description	Attributes
Correlation	Modal class of a soil Uses taxonomic (predefined) classes	Order Group Subgroup Parent material Rock Class Texture Group Permeability Depth class Stoniness class Texture Landtype? Drainage Miscellaneous
		Accessory attributes: Slope (optional) Up to 5 functional horizons (FH)
Base	Specified manually using expert and/or data.	Depth • rooting • diggability • to slow permeable layer Rooting barrier Thickness (for each FH) Stoniness (for each FH) Clay content (for each FH) Sand content (for each FH)
Derived	Generated from a model or pedotransfer function that predicts the attribute.	PAW Field capacity (for each FH) Wilting point (for each FH) Aeration (for each FH) Macroporosity (for each FH) Preferential flow?? Bulk density (for each FH) Total carbon (for each FH) Total nitrogen (for each FH) P (H_2SO_4) (for each FH) Ca (for each FH) CEC (for each FH) pH (for each FH) P retention (for each FH)
Interpreted	A risk, vulnerability, suitability or other environmental or resource capacity map	Soil compaction risk? ?

Developing S-map

A shared national correlation file will be used. At first this will be a spreadsheet based on Sam's Southland/Otago work. Pedologists will test using a PDA device for auger and simple profile descriptions initially. A small camera lens can also be mounted on a PDA.

The approach used to map the lowlands will be based on the TopoSouth and GrowOtago work where existing surveys are brought together in a common format and additional field work and mapping done to fill in the gaps. An innovative approach will be used to map the uplands area – typically pastoral / forestry hill country. Jochen Schmidt's approach divides up the landscape into key land elements, each of which can then be linked to correlation classes and estimates of soil properties. Alternatively a logistic regression approach may be useful in some upland areas? The intention is to use the same database structure of family and siblings for both lowland and uplands.

Specifying Uncertainty

Known soil variability should be primarily described using combinations of siblings, and pdfs. Possible errors or inaccuracy should be primarily described using the 'alternative' fields.

Mapunits can be associated with up to 5 siblings to describe the soils that are likely to be present in a polygon. The probability of each sibling being present is given (see table 2 for default probability). So a mapunit might be identified as being a TMPT24+EYRE6 (70:30) for example. Siblings making up less than 10% of a mapunit should not be listed.

Table 2

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

Variability information will be attached to mapunits (soil symbols) 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 mapunit are either less or more precise than the sibling default.

Most pdfs will be specified using a triangular distribution (min, mode and max) (Fig. 1a) or a uniform distribution (min, max) (Fig. 1b). Alternatively, a normal (Fig. 1c), lognormal (Fig. 1d) or a beta function (Fig. 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 Fig. 1f.

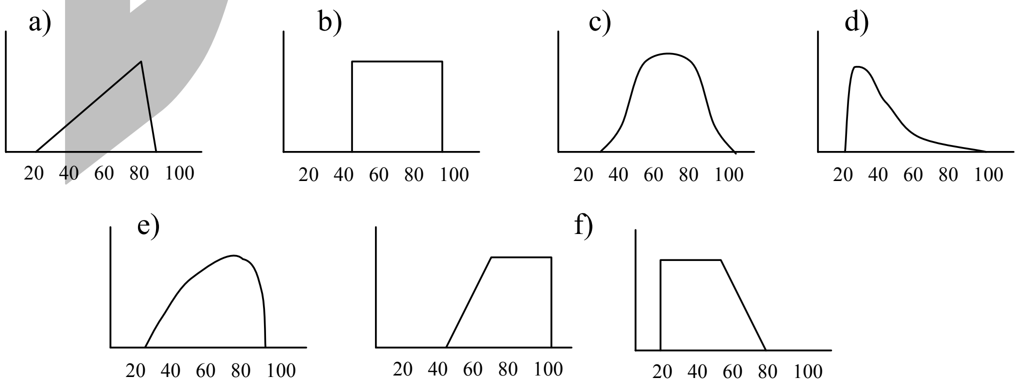

Figure 1 Standard probability functions that can be used to describe variability of a soil property. A) triangular, b) uniform, c) normal, d) lognormal, e) beta and f) irregular.

Each pdf will describe the likely distribution of attribute values within the sibling part of the given mapunit. For example, a mapunit that is defined as being an ABCD_1 sibling might have a triangular depth pdf (cms) of [45, 60, 100] (fig. 2a). A mapunit that is defined as being ABCD_1 (70%) + ABCD_2 (30%), where the siblings are defined as having different depths (mod. deep and deep respectively), would have two sibling pdfs of [45, 90, 90] and [90, 100] (fig. 2b). In the former the pdf expresses the soil expert's belief that the mapunit 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. Here the deep soil has been identified as being present with a 30% probability.

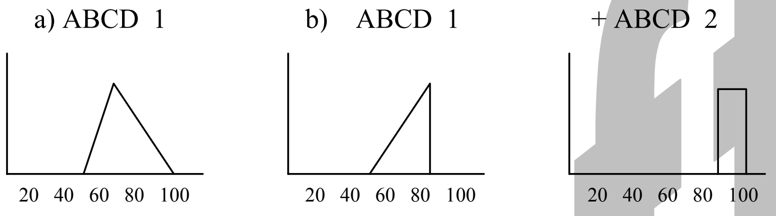

Figure 2 Example of depth pdfs associated with a a) single sibling map unit, and b) a double sibling map unit.

The uncertainty of each correlation attribute will be specified used the codes in Table 3. Possible alternative values should be provided where appropriate. Alternatives should generally be used to describe uncertainty or possible error, rather than variability (which is described using multiple siblings).

Table 3 Uncertainty codes

Code	Description
1	Confident that classification is correct based on an acceptable level of augur and pit observations for scale of study.
2	Reasonably confident that classification is correct or that it is one of the listed alternatives based on an acceptable level of augur and pit observations for scale of study
3	Classification is probably reliable based on inference
4	Classification is not reliable but is at least based on limited or historical data
5	Classification is unreliable: there is no data

This source/reliability code (table 4) should be specified for all base properties.

Table 4 Source of estimate

Code	Description	Note
M1	Observed or measured	Unbiased sampling to meet suitable confidence limits (or not)
M2	Observed or measured	With at least 5 good quality observations or measurements widely distributed
M3	Observed or measured	Limited sampling, (biased, localised or incomplete)
E	Estimated	Estimated by comparison with related soils
G	Unreliable estimate	

Polygon confidence or reliability is initially based on a survey confidence rating. Later this may be refined to sibling- or mapunit-level where appropriate. Survey confidence is determined by landscape predictability and survey quality i.e., std observation density and linework/registration quality. Confidence classes might be:

Table 5 Reliability of linework code

Confidence Code	Uncertainty of the specified probability of the sibling
A	± 20%
B	± 40%
C	Unknown

For example, in an A survey it can be expected that in a single symbol soil polygon, it is expected that the soil will be found in at least 80% of the polygon; in a double symbol polygon it is likely that the first soil will cover 48 – 72% of the polygon (60±20%), and the second soil will cover 32 – 48% of the polygon (40±20%).

Derived Property Layers

A number of attribute layers of key soil properties will be created. These may or may not relate directly to soil polygons therefore they are separate layers. Some properties are profile-related, others will be available at multiple depths (to a max of 1 m) or horizons. A soil can be characterised by up to 5 functional horizons. Table 6 shows details of the base and derived soil property layers.

Property	Control section	Definition	Aggregation level	Model	Inputs
Stones	Functional horizons	0-100%	Symbol	Base data (polygon lookup)	
Depth (dig-ability)	Profile	0-100 cm	Symbol	Base data (polygon lookup)	
Rooting Depth	Profile	cm to root barrier	Symbol	Base data (polygon lookup)	
Depth to slow permeable layer	Profile	0-100cm	Symbol	Base data (polygon lookup)	
Rooting barrier	Profile	Categorical: Appendix A	Symbol	Base data (polygon lookup)	
Thickness	Functional horizons	Cm	Symbol	Base data (polygon lookup)	
Texture profile	Profile	Categorical: Table 6	Symbol	Base data (polygon lookup)	
Clay	Functional horizons	%	Symbol	Base data (polygon lookup)	
Sand	Functional horizons	%	Symbol	Base data (polygon lookup)	
Slope	?	Categorical: Appendix A	Polygon?	Base data (polygon lookup)	
PAWC	Profile	% or mm	Functional horizons?	field data, or Giltrap funtions i.e., profile inputs, or profile inputs + C and Ox. Al, or profile inputs + BD data??	Horizon NZSC group Depth Silt and clay Ped size [Bulk den, C, Al .ox]
RAW Or PRAW?	Profile	% or mm	Functional horizons?	Giltrap?? New Model R^2=.316 and sd 3% Note – 3 alternatives, only field data inputs, or field + C and Ox. Al, or with BD data)	Clay, silt Depth Ped size NZSC [Bulk den, C, Al .ox]
Field Capacity	Functional horizons				
Wilting Point	Functional horizons				
Macroporo	Functional	% min ??		Giltrap?	Sand

sity	horizons			New model, R^2=.53, SD 5%	Ped size and type
Bulk density	Functional horizons			*Giltrap? (%Carbon, BD Class, Oxalate Al, Desig Sub Class) or f(Subgroup, Depth, Main Horizon Class*	Horizon notation NZSC order or group Depth Ped type Silt and sand [carbon, BD Oxalate Al if available]
K_{40}					
KSat					
Total Carbon	Functional horizons			Convert the CMS fixed depth model of Total Carbon. BUT note that CMS is a national model – a regional scale model might be better.	
Total Nitrogen	Functional horizons			f(carbon, landuse) Assume c:n ratio from landuse?	
P (H2SO4)	Functional horizons	Total or plant measurable?		Robbie's model?	
Ca	Functional horizons			f(pm, rainfall)	
CEC	Functional horizons			F(carbon, ?)	
PH	Functional horizons				
P retention	Functional horizons		Sibling level		
Clay mineralogy	*2 tiers*			*F(NSD data, NZSC)*	

Base layers will be the result of a manual process of assigning attributes to soil polygons, but all other soil property layers (derived and interpreted) will be derived from a model of some sort that can be coded. This will allow layers to be readily updated with advances in models or data, and very importantly to record the rationale behind the layer. In a manual approach this knowledge is lost. Initially the base layers will be driven by soil symbol (or sibling), then later by polygon if more information is available for some of the polygons. Note that the simplest default automated model is to populate the layers according to a simple family level (or subgroup) weighted mean of NSD data.

Note that regolith properties will probably require new observations and a manual model too.

Issue: A model management system needs to be set up, with minimum documentation standards.