Odgers, N., Lilburne, L., Guo, J., Vickers, S., Webb, T., Carrick, S., Barringer, J., 2019. Methods for upscaling and downscaling S-map information: Provision of spatial soil information in various formats and scales. In: Nutrient loss mitigations for compliance in agriculture. (Eds L.D. Currie and C.L. Christensen). http://flrc.massey.ac.nz/publications.html. Occasional Report No. 32. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand. 15 pages.

METHODS FOR UPSCALING AND DOWNSCALING S-MAP INFORMATION: PROVISION OF SPATIAL SOIL INFORMATION IN VARIOUS FORMATS AND SCALES

Nathan Odgers, Linda Lilburne, Jing Guo, Shirley Vickers, Trevor Webb, Sam Carrick and James Barringer

Manaaki Whenua – Landcare Research, PO Box 69040, Lincoln 7640, New Zealand
Email: odgersn@landcareresearch.co.nz

Abstract

S-map is published at a nominal cartographic scale of 1:50,000. The size of map units drawn at this scale means that S-map is best suited for use in regional to sub-catchment scale applications. S-map map units can be too large to adequately depict the spatial variation of soil at the farm scale, whereas applications at the national scale may not require as much detail as is contained in S-map. This suggests that transformations could be made to S-map that provide soil information that is a better fit for these scales. Transforming S-map to fit farm-scale applications is a downscaling exercise, whereas transforming S-map to fit national-scale applications is an upscaling exercise. We identified a range of methods for performing downscaling and upscaling. We concluded that more contextual information is required to adequately downscale S-map than is currently available, so our main focus in this paper is on methods for upscaling S-map. In general, upscaling methods can be classified based on the order of operations performed: (i) geometric simplification followed by attribute simplification, versus (ii) attribute simplification followed by geometric simplification. We demonstrate an application of the first approach and its effect on water balance modelling, and an example of the second approach via the simplification of map unit components into soil property groups.

Introduction

The level of spatial and attribute detail of a soil map with a nominal scale of 1:50,000, like S-map, is probably most suitable for regional to sub-catchment scale applications (i.e. spatial extents of dimensions 10^4–10^5 m), such as land-use planning (Lilburne et al., 2012). Such a map may not be detailed enough to depict soil variability at the farm to paddock scale (10^2–10^3 m), depending on the resolution of farm management and the nature of the soil variability (Barringer et al., 2016; Carrick et al., 2014; Manderson and Palmer, 2006). Information about fine-scale soil variability is increasingly required for precision agriculture (Hedley et al., 2013). On the other hand, a map with a scale of 1:50,000 is probably too detailed for national-scale applications (10^6 m), where only a generalised overview of the main patterns or differences between soils may be required.

This has important implications. For example, a farmer may like to know precisely which soil sibling underlies his field out of the five that are defined in the S-map map unit that is mapped across it, but will require S-map to provide more detailed information about the spatial distribution of siblings within map units. On the other hand, a climate modeller working with 5 km-resolution raster climate data may decide that she would like raster soil information at a

comparable spatial resolution but is unsure how best to generalise the information she finds in S-map.

There are several ways to downscale and upscale S-map to provide soil spatial information that better suits these use cases.

Downscaling and upscaling methods

The soil map subdivides the landscape into a tessellation of polygons, where each polygon is linked to a soil map unit. S-map soil map units define a set of up to five soil siblings and their respective proportions. Several soil polygons can be assigned to a single soil map unit. In this section we describe a range of methods for downscaling and upscaling a polygon soil map such as S-map.

Downscaling

The purpose of downscaling a soil map is usually to produce a more detailed depiction of the spatial distribution of soil than is provided by the original soil map. One approach involves predicting the spatial distribution of the soil map units' components, in which case the downscaling process is known as spatial disaggregation. In recent years a substantial body of research into methods for performing spatial disaggregation has developed (e.g. Häring et al., 2012; Holmes et al., 2014; Kerry et al., 2012; Nauman and Thompson, 2014; Odgers et al., 2014; Sarmento et al., 2017; Subburayalu et al., 2014; Thompson et al., 2010; Vincent et al., 2018).

We suspect that the feasibility of spatially disaggregating S-map map units is currently limited. This is because little contextual information, such as the landscape position of map unit components, is currently available to inform the disaggregation. Furthermore, fine-resolution environmental covariates are not available everywhere because of patchy LiDAR coverage (for example), and robust relationships between key soil attributes and covariates in the New Zealand context are often not well understood. A preliminary investigation to see if slope information from a digital elevation model could be used to spatially locate the siblings was not successful.

Upscaling

The purpose of upscaling a soil map is to produce a more general depiction of the spatial distribution of soil across an area than is provided by the original soil map. A wider variety of methods are available to perform upscaling than to perform downscaling, and they may be classified in several ways. We distinguish two modes, *spatial simplification* and *attribute simplification*, and note that spatial simplification may be performed before attribute simplification or afterwards, but the upscaled results from these two approaches will not be the same.

Within each mode there are various simplification operations depending on whether the simplifications are focusing on vector or raster maps, taxonomic classes or soil properties, and categorical or numerical data. We briefly describe how each sequence of modes may be applied to both vector soil maps and raster soil maps, but due to space limitations we will restrict discussion to soil class maps. We note that numerical soil properties are readily upscaled by calculating numerical summary values such as the mean or range.

Spatial simplification followed by attribute simplification

Figure 1 illustrates the steps involved when upscaling is performed by spatial simplification followed by attribute simplification. The starting point is either a vector soil map or a raster soil map, and, as we describe below, the operations that are performed in steps 1 and 2 depend on the format.

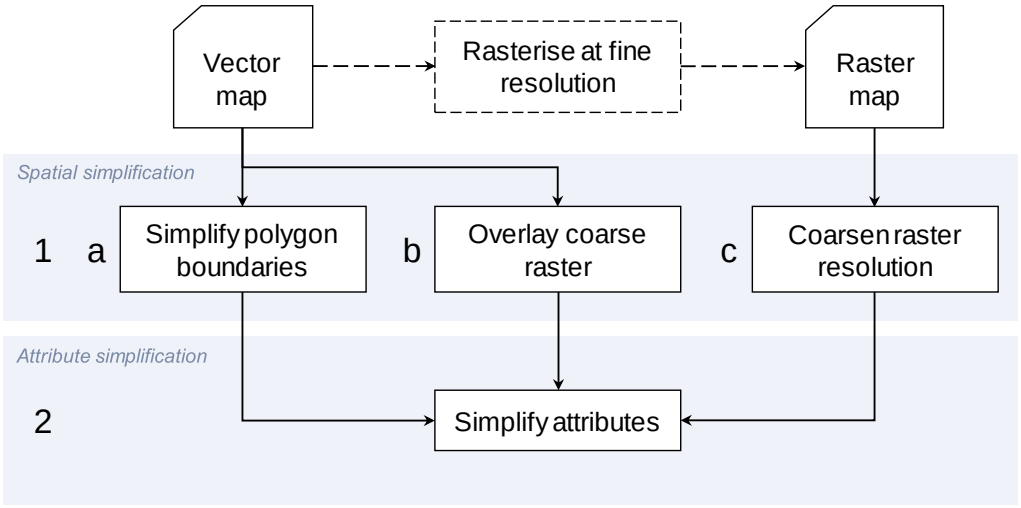

Figure 1. Illustration of the steps involved in an upscaling process where spatial simplification is performed before attribute simplification. Dashed lines represent optional steps.

For vector maps, step 1 involves the geometric simplification of the polygon boundaries in order to suit the coarser target map scale (method 1a in Figure 1). This is not a trivial task, because it raises questions about how much the boundaries will change, and how the changes affect the reliability of the generalised map (Valentine, 1981). Step 2 involves updating the attributes of the simplified geometry, as follows.

- Attribute the simplified map unit polygons with the dominant soil class in the original map unit or define a new, generalised map unit composition. This approach discards a lot of information, especially if the geometric simplification merges polygons from several map units.
- Retain information about all the components of the original map unit plus any other map units that were subsumed by it during the geometric simplification, and update the simplified map unit composition. This preserves the most information.

More options are available if the *target* format of the upscaling is a raster map. In method 1b of step 1 (Figure 1), a coarse raster grid is overlaid on the vector soil map; in method 1c a coarse raster grid is overlaid on a finer-resolution raster soil map.

The options for raster-based attribute simplification in step 2 are essentially the same whether the methods in Step 1b or 1c are employed.

- Attribute the coarse grid cells with the dominant soil class found among the polygons or finer grid cells that intersect with them.
- Retain information about all the components of the original map units that intersect the coarse grid cells.

Attribute simplification followed by spatial simplification

Figure 2 indicates the steps involved when upscaling is performed by attribute simplification followed by spatial simplification. The starting point is a vector soil map. The same set of operations could be performed on a raster soil map, but we believe there would be no difference between the upscaled raster soil map and a rasterised version of the upscaled vector soil map. In any case, step 1 involves the simplification of soil map unit attributes to inform re-labelling of the map units. Several methods are possible.

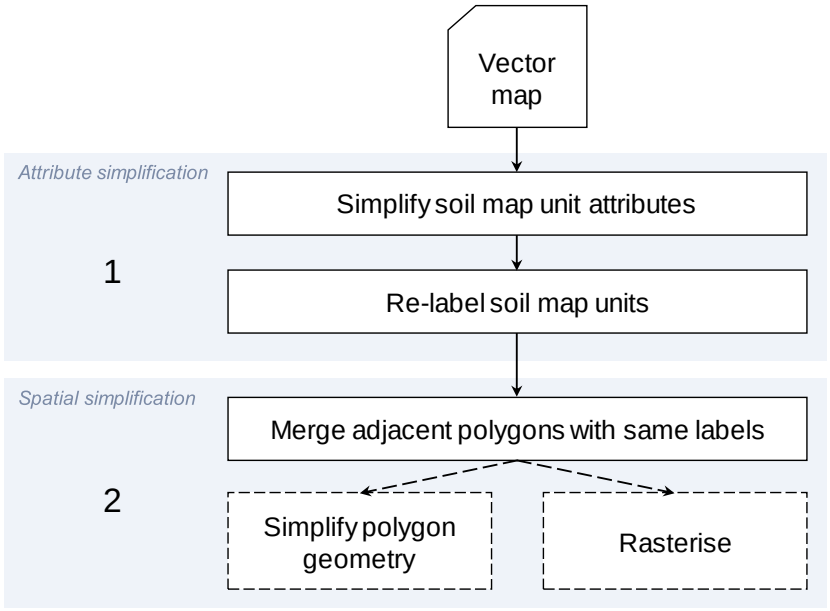

Figure 2. Illustration of steps involved in an upscaling process where attribute simplification is performed before spatial simplification. Dashed lines represent optional steps.

- Re-label the soil map units with their dominant component.
- Re-express the map unit composition in terms of a higher-level taxon (e.g. soil orders rather than soil siblings), then re-label soil map units with the dominant higher-level taxon.
- Simplify the soil map unit components by clustering them on functional properties in order to identify groups of "key soils" (Coucheney et al., 2018), then re-label the soil map units with the dominant key soil group.

Step 2 involves the spatial simplification of the map unit polygons based on their new labels. Adjacent polygons are merged if and only if they share the same simplified label. The result is a spatial simplification in the sense that the simplified map has fewer polygons than the original soil map, but the procedure does not generalise the level of detail in the polygon boundaries. As Figure 2 indicates, geometric simplification of polygon boundaries is an optional additional step.

Case studies

We present two case studies. The first case study is an example of attribute simplification followed by spatial simplification. The second case study is an example of spatial simplification followed by attribute simplification, and it illustrates the effect of upscaling S-map soil information on soil water balance modelling.

Case study 1: Key soil groups

The first case study is an example of upscaling by attribute simplification in order to generalise S-map siblings into a smaller number of key soil groups. We envisage a use case where a climate modeller cannot, for computer performance reasons, run her model on all the siblings in each climate grid cell but wishes to account for the soil variability by running her model on a handful of representative soils in each grid cell. If all the siblings in a grid cell could be aggregated into more general, functionally similar soil groups, she might be able to choose a set of representative soils more easily.

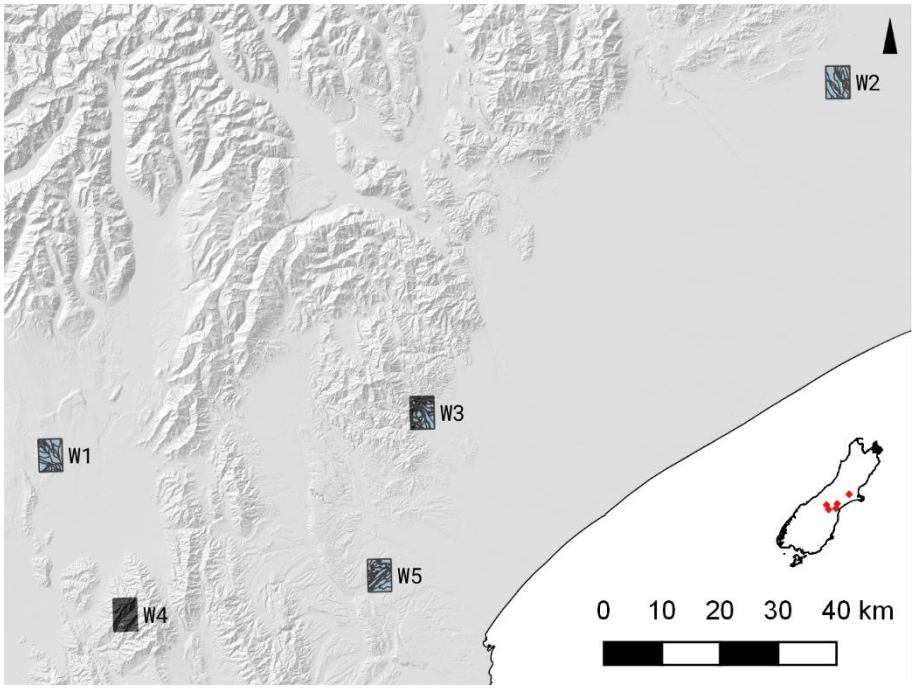

Figure 3. Location of five windows within which the key soil group upscaling was tested in case study 1.

Method

Study area: We tested the approach in five windows that represent common physiographies in the central South Island, New Zealand: coastal plains, plains–hill-country margins, hill country, and intermontane basin. The size and shape of the windows were based on NIWA's Virtual Climate Station Network (VCSN) grid cells[1] and were approximately 5 km × 5 km in size.

Soil information: Each window had spatially exhaustive S-map coverage. We represented the siblings in each window by derivatives of seven sibling- or family-level soil properties (Webb and Lilburne, 2011): drainage class, depth class, permeability class, texture class, topsoil stoniness class, presence or absence of pans, and profile available water class. All soil properties were represented as categorical variables. In addition, those properties whose states could be readily ordered (all except texture class and presence or absence of pans) were treated as ordinal categorical variables (Table 1).

Attribute simplification: The key soil group upscaling is a two-step procedure. Since all seven soil properties are represented as categorical variables, there can be only a limited number of combinations of their property states. We call the different combinations of soil property states soil property groups (SPGs). The first upscaling step involved identifying the SPGs present within each window and allocating each sibling to its SPG.

The SPGs vary in their degree of similarity to each other; for example, some are nearly identical except in one or two soil properties. Within each window we quantified the degree of similarity between SPGs using Podani's extension to Gower's distance, which accounts for ordinal categorical variables (Gower, 1971; Podani, 1999). The second upscaling step involved hierarchically clustering the SPGs within each window, using a complete-linkage algorithm, into more general groups that we call *supergroups*.

[1] https://www.niwa.co.nz/climate/our-services/virtual-climate-stations

Table 1. Soil properties used to represent S-map siblings. Most values are derivatives of class values published by Webb and Lilburne (2011).

Soil property	Values	Type
Drainage class	Poor or very poor, imperfect; well or moderately well	Ordinal
Depth class	Shallow or very shallow, moderately deep, deep	Ordinal
Permeability class	Slow, moderate, rapid	Ordinal
Texture class	(l+z)/c, (l+z)/s, c+c/*, l/k, l+z, p, s, s/(z+l+c), Tl/c, Tl/s, Tl/z, Ts/z, z/l, z/Tl	Categorical
Topsoil stoniness class	Stoneless or slightly stony, moderately stony, very stony	Ordinal
Presence or absence of pans	0 (pan absent), 1 (pan present)	Binary
Profile available water class	Low, moderate, high	Ordinal

To choose an appropriate number of supergroups within each window we tried to strike a balance between the number of supergroups and the median supergroup size: with too few supergroups the median number of SPGs in them is relatively large (too much generalisation), whereas with too many supergroups the median number of SPGs is relatively small (too little generalisation). Such a balance can be found at the elbow of the curve made when the number of supergroups is plotted against the median number of SPGs per supergroup (Figure 4).

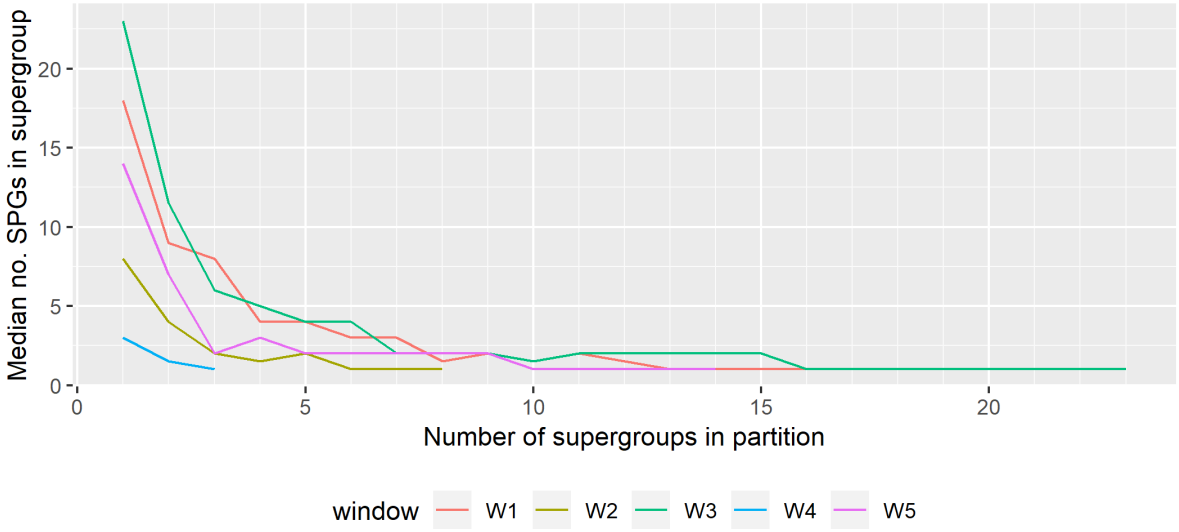

Figure 4. Relationship between the number of supergroups and the median number of SPGs in each supergroup, for each window.

Results and discussion

Table 2 indicates that the soil distribution as depicted by S-map is more complex in some windows than in others, despite the fact that the windows are all the same size. This is not purely a pedological phenomenon; it also depends on mapping methods and priorities, which vary over time and between mapping staff. The number of siblings in each window ranged from 7 to 40, the number of SPGs from 3 to 23, and the number of supergroups from 2 to 7. The key soil group upscaling produces supergroups whose members are more homogeneous in their soil properties compared to the whole set of siblings within a given window. For example,

Table **3** describes the siblings found in window W5 on the margin of the Canterbury Plains near Tīmaru. The window contains 21 siblings across five soil orders in the New Zealand Soil Classification (Hewitt, 2010). The siblings were generalised into 14 SPGs, which were then hierarchically clustered into five supergroups.

Table 2. Number of siblings, number of SPGs, and number of supergroups within each window.

Window	No. siblings	No. SPGs	No. supergroups
W1	20	18	7
W2	15	8	4
W3	40	23	7
W4	7	3	2
W5	21	14	5

Table 3. Siblings of window W5 and their SPG and supergroup allocation. Siblings in bold type are the spatially dominant sibling within their supergroup.

Sibling	NZSC	Area (ha)	Drainage[a]	Depth[b]	Permeability	Texture[c]	Topsoil stoniness[d]	Pan	PAW	SPG	Supergroup
River_1a.1	WW	12.4	w+mw	s+vs	rapid	s	vst	0	low	13	1
Rang_41a.1	RFT	31.1	w+mw	s+vs	rapid	s	mst	0	low	14	1
Rang_18b.2	RFT	64.4	w+mw	s+vs	rapid	s	vst	0	low	13	1
Ashb_12a.1	**WF**	**115.3**	**w+mw**	**s+vs**	**rapid**	**s**	**vst**	**0**	**low**	**13**	**1**
Cair_4a.1	EOMJ	9.7	i	d	slow	(l+z)/c	stl+sst	0	mod	11	2
Ytoh_1a.1	**PPJ**	**263.6**	**p+vp**	**d**	**slow**	**(l+z)/c**	**stl+sst**	**0**	**mod**	**4**	**2**
Timu_2a.1	PXM	9.2	i	md	slow	(l+z)/c	stl+sst	1	mod	10	3
Kaur_2a.2	PJT	13.7	w+mw	md	slow	l+z	stl+sst	1	mod	12	3
Timu_1a.2	PXM	94.7	i	md	slow	l+z	stl+sst	1	mod	8	3
Kelc_1a.1	PJM	123.1	i	md	slow	(l+z)/c	stl+sst	1	mod	10	3
Clar_1a.1	PPX	280.0	p+vp	md	slow	l+z	stl+sst	1	mod	5	3
Timu_1a.1	**PXM**	**543.0**	**i**	**md**	**slow**	**l+z**	**stl+sst**	**1**	**mod**	**8**	**3**
Lism_2a.1	BFP	11.7	w+mw	s+vs	moderate	l+z	mst	0	mod	2	4
Eyre_1a.1	ROW	71.8	w+mw	s+vs	moderate	l+z	stl+sst	0	mod	6	4
Raka_1a.1	RFW	90.9	w+mw	s+vs	moderate	l+z	stl+sst	0	mod	6	4
Eyre_3a.1	ROW	92.4	w+mw	s+vs	moderate	l+z	mst	0	mod	2	4
Raka_2a.1	**RFW**	**136.4**	**w+mw**	**s+vs**	**moderate**	**l+z**	**mst**	**0**	**mod**	**2**	**4**
Waka_1a.1	PIM	7.0	i	d	slow	l+z	stl+sst	0	high	3	5
Temp_9a.1	PIT	28.1	w+mw	md	slow	l+z	stl+sst	0	high	1	5
Temp_2a.1	PIT	33.2	w+mw	md	slow	l+z	stl+sst	0	mod	7	5
Fris_1a.1	**PJT**	**185.7**	**w+mw**	**d**	**moderate**	**l+z**	**stl+sst**	**0**	**mod**	**9**	**5**

[a]p+vp: poorly or very poorly drained; i: imperfectly drained; w+mw: well or moderately well drained
[b]s+vs: shallow or very shallow; md: moderately deep; d: deep
[c]l+z: loam or silt; (l+z)/c: loam or silt over clay; s: sand
[d]stl+sst: stoneless or slightly stony; mst: moderately stony; vst: very stony

In supergroups 1, 2 and 4, members varied on only one soil property; members of supergroup 3 varied on two soil properties, and members of supergroup 5 varied on four soil properties. When members varied on a soil property, they almost always expressed less than the full range of variation in the soil property compared to the whole set of siblings in the window. Only members of supergroup 3 expressed the full range of variation in a single soil property (the three property states of soil drainage class).

With respect to our use case, we must consider how the supergroups are represented to the end-user. The spatially dominant sibling member of a supergroup is one candidate to represent it, as is the spatially dominant SPG (since several siblings may be allocated to the same SPG). The dominant sibling may be a better choice since it may be a more recognisable concept to the end-user. The dominant sibling within each supergroup in window W5 is highlighted in bold type in

Table **3**.

Our climate modeller might run her model in window W5 with the five highlighted siblings that essentially represent well-drained very stony soils (supergroup 1), poorly drained soils (supergroup 2), imperfectly drained soils with a pan (supergroup 3), well-drained shallow soils (supergroup 4), and well-drained deep soils (supergroup 5).

The key soil group upscaling process is a consistent method for producing generalised groups of soils. Members of a group are expected to share functional similarities, but not necessarily taxonomic similarities. Since the groups are functional groups, the soil properties that determine function may vary depending on the application.

Case study 2: Water balance modelling

The second case study illustrates the effect of upscaling soil spatial information on the output of a soil-water balance model. Soil-water status is an important component of crop growth models and may also be used to schedule irrigation or forecast drought (Woodward et al., 2001). Soil properties are some of the key input parameters to soil-water balance models, so the spatial scale of the soil property information supplied to these models may affect the model output. This, in turn, may have implications for decision-making processes that are based on model predictions. We were interested in using the output of a soil-water balance model to examine the effect of changing the scale of soil information on predictions of crop water demand as an indicator of drought.

Method

Water balance model: We ran a daily water balance model called WATYIELD (Fahey et al., 2010) to produce daily estimates of soil-water content (SWC) across the part of New Zealand that has S-map coverage, for one simulation year that ran from 1 July 2004 to 30 June 2005.

Soil information: The key soil parameters that WATYIELD requires are a soil profile's total available water (TAW) and readily available water (RAW), in millimetres. TAW is the difference between the SWC at field capacity and at permanent wilting point, whereas RAW is the difference between the SWC at field capacity and at the trigger point for crop water stress (Allen et al., 1998), which in New Zealand is typically the SWC at 100 kPa suction. The trigger point, SWC_{TP}, is the soil-water content below which a crop can no longer extract enough water to sustain optimal plant growth.

The TAW and RAW were supplied to WATYIELD as raster layers, where each cell of the raster contained the soil profile's TAW or RAW value, respectively. We tested the effect of upscaling the soil spatial information by rasterising New Zealand's S-map coverage at a range

of grid resolutions: 100 m, 1000 m, 5000 m, and 20,000 m. Each raster cell received the TAW and RAW of the dominant soil sibling in the cell at the given raster resolution, or a null value if the cell was outside the S-map coverage.

In addition, we also created two hypothetical soil profiles: the first soil had a TAW of 150 mm and an RAW of 75 mm (the "nominal" soil), and the second soil had a TAW of 133 mm and an RAW of 53 mm (the "average" soil). The first soil is the hypothetical soil profile that NIWA uses in its water balance simulations, and the second soil represents the average of all soils in S-map. The TAW and RAW of these soils were held constant and were gridded across New Zealand at a 5000 m resolution.

Climate information: The key climate parameters WATYIELD requires are precipitation, in millimetres, and a reference crop evapotranspiration, $\mathrm{ET_0}$, in millimetres. We obtained daily precipitation rasters and daily potential evapotranspiration rasters for the simulation year from NIWA (Tait et al., 2016). These data are gridded on NIWA's VCSN grid at a spatial resolution of approximately 5000 m.

Potential evapotranspiration deficit calculations: We used the simulated daily soil water content values from WATYIELD to estimate the daily actual evapotranspiration (AET), in millimetres of water, of a notional generic pasture crop. Under optimal conditions, when there is sufficient soil water for crop growth, the potential evapotranspiration (PET) of a crop, in millimetres, may be estimated as a proportion of a reference evapotranspiration:

$$PET_C = K_C * \ (ET_0 - E_{CAN})$$

where $\mathrm{ET_0}$ is the evapotranspiration, in millimetres of water, of a reference surface with a grass reference crop, E_{CAN} is the evaporation, in millimetres, of free water from the reference crop canopy (Neitsch et al., 2011), and K_{C} is a dimensionless crop coefficient that adjusts the reference evapotranspiration for a specific crop (Allen et al., 1998). We set the K_{C} of the generic pasture crop to 1.0, after Allen et al. (1998).

In reality, optimal conditions for plant growth do not always exist. WATYIELD assumes that the crop draws water from the soil at the same rate as the AET when the crop is not under water stress; however, as the soil dries, the soil water content eventually decreases to a point where the crop can no longer extract enough water for optimal plant growth. This point is called the trigger point (TP) for crop water stress, and the soil water content at this point ($\mathrm{SWC_{TP}}$), in millimetres, is calculated as (Allen et al., 1998):

$$\mathrm{SWC_{TP}} = \mathrm{TAW} - \mathrm{RAW}$$

When the SWC is less than $\mathrm{SWC_{TP}}$ and the crop is under water stress, AET begins to decrease at a rate that is described by a dimensionless water stress coefficient, K_{S} (Allen et al., 1998):

$$K_{\mathrm{S}} = \mathrm{SWC}_i/(\mathrm{TAW} - \mathrm{RAW})$$

Accounting for crop water stress, the AET of the crop on day i of the simulation year ($i = 1, 2, \ldots, 365$) is therefore calculated as follows (Allen et al., 1998):

$$\mathrm{AET}_i = \begin{cases} K_{\mathrm{C}} * \ (\mathrm{ET_0} - E_{\mathrm{CAN}}) = \mathrm{PET_C}, & \mathrm{SWC}_i \geq \mathrm{SWC_{TP}} \\ K_{\mathrm{C}} * K_{\mathrm{S}} * \ (\mathrm{ET_0} - E_{\mathrm{CAN}}), & \mathrm{SWC}_i < \mathrm{SWC_{TP}} \end{cases}$$

Finally, the crop water demand (CWD) is the difference between a crop's AET_i and its $\mathrm{PET_C}$; in other words it is the additional amount of soil water, in millimetres, that a crop needs on day i in order to achieve an optimal growth rate:

$$\mathrm{CWD}_i = \mathrm{PET_C} - \mathrm{AET}_i$$

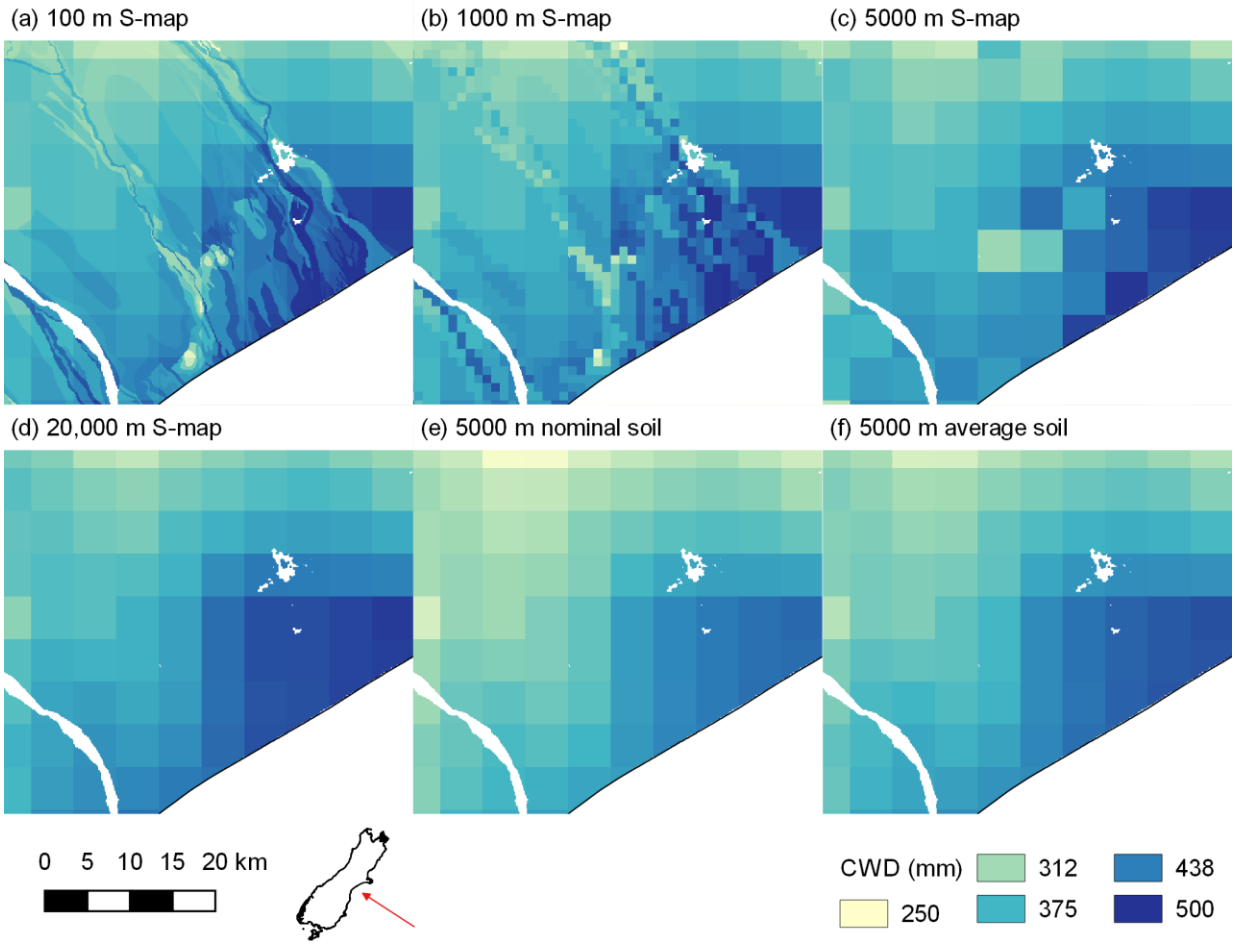

Figure 5. Annual CWD as derived from WATYIELD output, based on TAW and RAW at a range of spatial scales

If $\mathrm{CWD}_i > 0$ mm then the crop is under water stress on day *i*, and indicates that the soil is in drought. If the daily CWD is summed over a period of time, the sum indicates the total amount of water the crop needed over the period in order to maintain optimal growth.

Model runs: In total we made six runs of WATYIELD using the six sets of TAW and RAW: one run for each of the four S-map upscalings and for each of the two hypothetical soils. This generated six sets of daily SWC rasters. For each daily SWC raster, we computed CWD_i using the method described above. Finally, we reclassified the CWD_i rasters to a binary representation in order to depict instance of drought: in the reclassified rasters, values of 0 indicated CWD of 0 mm (no drought); values of 1 indicated CWD greater than 0 mm (drought).

Common grid: To properly compare the effect of upscaling the soil information, the 5000 m climate rasters, the 1000 m, 5000 m, and 20,000 m S-map TAW and RAW rasters, and the 5000 m resolution hypothetical soil TAW and RAW rasters were all resampled using a nearestneighbour algorithm to the 100 m raster grid, and all simulations were performed on this common grid. Therefore the only information that varied between WATYIELD runs for each 100 m grid cell was the soil TAW and RAW information.

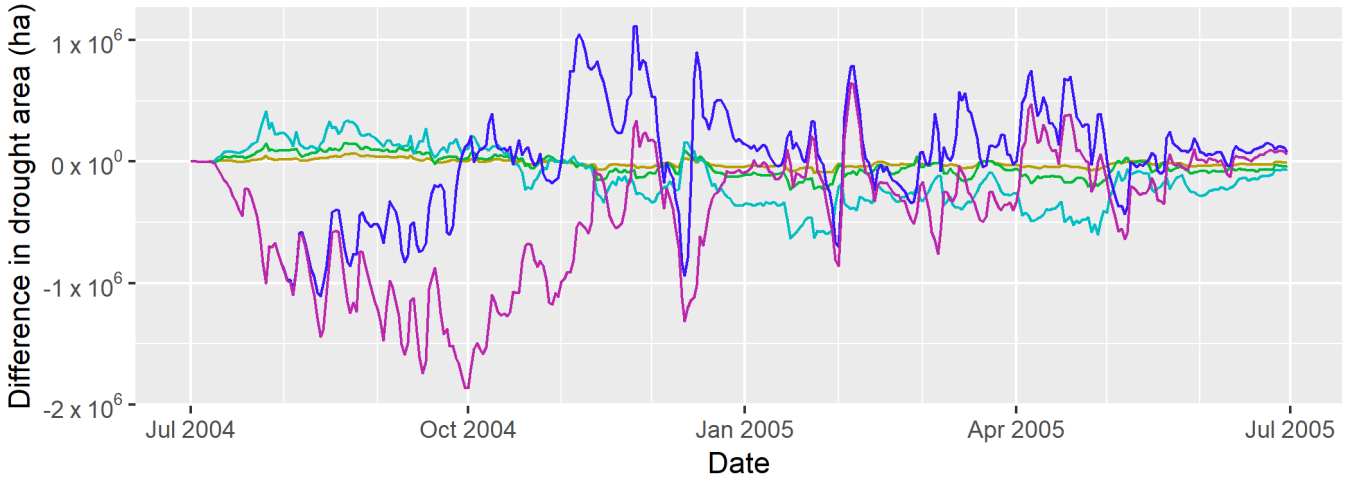

Figure 6. Time series of the difference in the area in drought between the predictions based on the 100 m-resolution S-map TAW and RAW, and the predictions based on the other scales. Positive values are where the estimate at the other scale is larger than the estimate at the 100 m resolution and vice-versa.

Results and discussion

For each of the six WATYIELD runs we summed the daily CWD and drought rasters to estimate the annual CWD and area in drought. The effect of changing the scale of soil information on annual CWD is depicted in Figure 5 for a small area of Canterbury. It is easy to see the scaling effect between panels (a) and (b); the effect is obscured in panels (c) to (f) because the climate rasters also had a resolution of 5000 m. The climate effect explains why we don't see an imprint of 20,000 m × 20,000 m blocky artefacts in panel (f). However, the effect of using actual soil information compared to using a single nominal or average soil is still evident when compared on the same scale, where the scaled S-map data in panel (c) shows a clear difference in CWD pattern, compared to panels (e) and (f) which used a respective nominal and average soil.

The effect of spatially scaling the soil information is clearest when the results are viewed at a regional extent or finer, because climate is the main driver of the spatial distribution at the national extent.

We were also interested in the effect of the soil information scaling on the daily predictions. To examine this effect, we estimated the total daily CWD and drought area for each simulation day by summing the daily CWD and drought incidence rasters, respectively. We did this separately for each of the six WATYIELD runs. We then subtracted the daily CWD and drought area of the simulations based on the 100 m S-map upscaling from the corresponding daily values of the other five WATYIELD runs. This let us examine differences in the temporal trend in CWD and area in drought across scales, relative to the predictions made using the 100 m-resolution S-map TAW and RAW. Time series of the differences in drought area are plotted in Figure 6.

Another method of illustrating the differences in CWD and drought area between scales is to plot time series of the cumulative sum of the daily total CWD and drought area estimates. The cumulative time series for CWD is plotted in Figure 7.

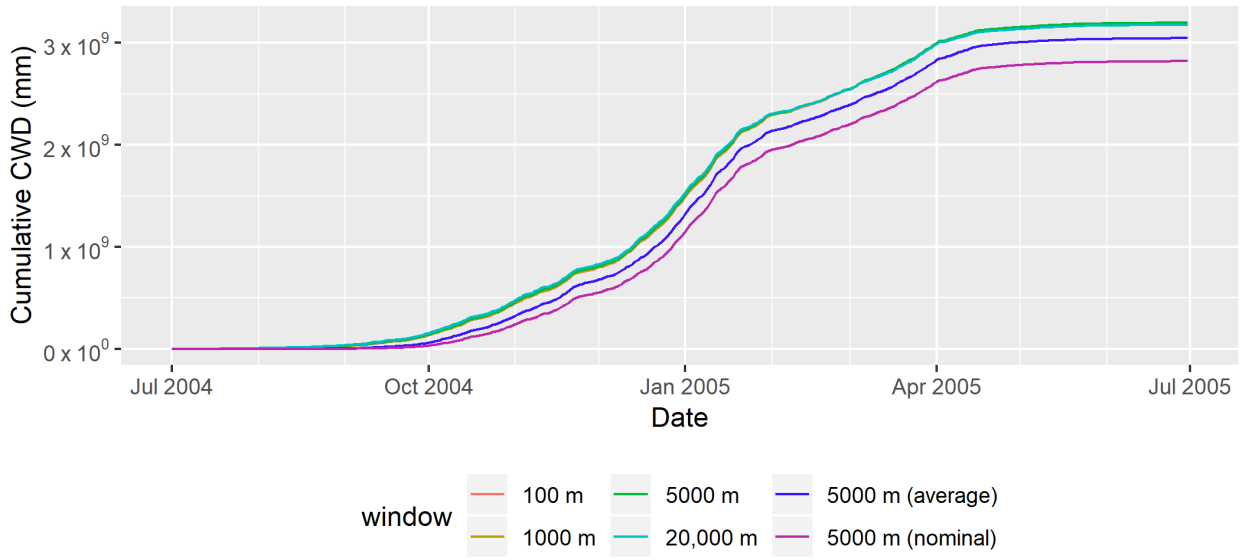

Figure 7. Cumulative sum of CWD for simulations based on each set of soil maps. For each simulation, the daily value that is accumulated is the sum of the day's CWD raster.

The dominant S-map sibling changes from place to place, but can also change at a given location depending on the raster resolution. We observe that upscaling S-map to coarser and coarser scales tends to remove areas of soils with higher TAW and RAW because soils with lower TAW and RAW happen to be spatially dominant. For example, the mean TAW of the 100 m S-map was 133 mm, whereas the mean TAW of the 20,000 m S-map was 113 mm. Climate also has a spatially variable effect, and all these effects together influence the behaviour of the simulated variables over time. Some areas go into drought as early as early July, although according to Figure 6 the total area in drought is largest during summer and early autumn. In the first three and a half months of the simulation year, simulations based on the coarsest S-map scaling typically show the greatest area of drought on any given day; this relationship holds until about the start of October, after which they frequently show the smallest daily area of drought.

What causes these relationships? As we observed above, the coarser S-map upscalings tend to have more soils with lower TAW and RAW. As evapotranspiration occurs, their SWC decreases and their trigger points are reached sooner than in the soils with higher TAW and RAW that are more common in the finer S-map upscalings. Later in the simulation year SWC has decreased to the point that even many of the higher TAW and RAW soils in the finer upscalings are in drought. Precipitation leads to an increase in the SWC, which is often enough to bring many of the lower TAW/RAW soils out of drought, but is often insufficient to raise the water content of the higher TAW/RAW soils above their trigger points.

These relationships also help to explain why areas of average and nominal soils do not begin to go into drought until August: they initially hold more water than many of the S-map soils. Drought takes longer to accumulate in simulations based on the nominal soil than in simulations based on the average soil because the nominal soil has a larger RAW and a lower SWC_{TP}. Simulations based on the nominal soil underestimate the annual accumulated CWD compared to simulations based on S-map and the average soil for the same reason (Figure 7).

It is interesting that while there are differences in the area in drought from day to day between the simulations based on the four S-map upscalings, Figure 7 indicates that the annual accumulated CWD (i.e. the cumulative value at 30 June 2005, the final day of the simulation

year) varies relatively little between these simulations, compared to simulations based on the average and nominal soils. There is some variance, but the curves for the simulations based on the S-map upscalings essentially overlap at the end of the simulation year. The largest annual CWD was 3.2×10^9 mm and was based on the simulation that used the 100 m S-map.

The results suggest that the use of spatially variable soil information in water balance models leads to more realistic summary estimates of drought indicators compared to model estimates based on spatially constant soil information, which underestimated annual summaries of CWD. This is clear in Figures 6 and 7 where the simulations at 5000 m resolution based on a single nominal or average soil produced quite different CWD to that based on scaling of S-map data. Within the group of simulations that were based on the S-map upscalings, the degree of upscaling made a relatively small difference to annual summaries of predictions.

The degree of upscaling is more important in applications that require spatial estimates of the drought indicators, because coarser upscalings convey less spatial information. For national- and some regional-extent applications, a fairly coarse upscaling, such as our 20,000 m resolution upscaling, may be adequate. That is to say, the level of spatial information lost at this scale, relative to the original S-map, may be acceptable given the intended use of the predictions—such as to ease a visualisation or computational burden. On the other hand for other regional- and finer-extent applications, a finer upscaling may be appropriate.

The effect of upscaling methods depends a lot on the nature of the spatial variation of the soil and the linearity or otherwise of the model that is using the soil information. For example, if soil variation followed a pattern where the dominant soil was generally a deep soil and the secondary soils were generally shallow, then CWD would be underestimated. The effect might be exacerbated if the model is strongly non-linear. If, however, the spatial pattern alternated between shallow and deep soils being dominant, then the effect of upscaling is likely to be small. Model non-linearity would only be important if soils with a non-linear response were not well represented after the spatial upscaling operation.

Conclusion

Upscaled soil information can ease a computational or visualisation burden. The appropriateness of upscaled soil data produced by a given method will depend on one's purpose. For some regional or national questions, relatively coarse data may be fit for purpose. Other questions will require more detailed data. The fitness for purpose of a given scale of soil information for a specific purpose should always be tested.

Acknowledgements

We would like to thank NIWA for provision of the climate data we used in the water balance modelling.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56. Food and Agriculture Organisation of the United Nations, Rome.

Barringer, J.R.F., Lilburne, L., Carrick, S., Webb, T., Snow, V., 2016. What difference does detailed soil mapping information make? A Canterbury case study, in: Currie, L.D., Singh, R. (Eds.), Integrated Nutrient and Water Management for Sustainable Farming, Occasional Report. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand.

Carrick, S., Hainsworth, S., Lilburne, L., Fraser, S., 2014. S-map @ the farm-scale? Towards a national protocol for soil mapping for fam nutrient budgets, in: Currie, L.D., Christensen, C.L. (Eds.), Nutrient Management for the Farm, Catchment and Community, Occasional Report. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand.

Coucheney, E., Eckersten, H., Hoffmann, H., Jansson, P.-E., Gaiser, T., Ewert, F., Lewan, E., 2018. Key functional soil types explain data aggregation effects on simulated yield, soil carbon, drainage and nitrogen leaching at a regional scale. Geoderma 318, 167–181. doi:10.1016/j.geoderma.2017.11.025

Fahey, B., Ekanayake, J., Jackson, R., Fenemor, A., Davie, T., Rowe, L., 2010. Using the WATYIELD water balance model to predict catchment water yields and low flows. Journal of Hydrology (New Zealand) 49, 35–58.

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27, 857–874. doi:10.2307/2528823

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. Geoderma 185–186, 37–47. doi:10.1016/j.geoderma.2012.04.001

Hedley, C.B., Roudier, P., Yule, I.J., Ekanayake, J., Bradbury, S., 2013. Soil water status and water table depth modelling using electromagnetic surveys for precision irrigation scheduling. Geoderma 199, 22–29. doi:10.1016/j.geoderma.2012.07.018

Hewitt, A.E., 2010. New Zealand Soil Classification, 3rd ed. Landcare Research Science Series. Manaaki Whenua Press, Lincoln, New Zealand. doi:10.7931/DL1-LRSS-1-2010

Holmes, K.W., Odgers, N.P., Griffin, E.A., van Gool, D., 2014. Spatial disaggregation of conventional soil mapping across Western Australia using DSMART, in: Arrouays, D., McKenzie, N.J., Hempel, J.W., Richer de Forges, A.C., McBratney, A.B. (Eds.), GlobalSoilMap: Basis of the Global Spatial Soil Information System. Taylor & Francis, London, United Kingdom, pp. 273–279.

Kerry, R., Goovaerts, P., Rawlins, B.G., Marchant, B.P., 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. Geoderma 170, 347–358. doi:10.1016/j.geoderma.2011.10.007

Lilburne, L.R., Hewitt, A.E., Webb, T.W., 2012. Soil and informatics science combine to develop S-map: A new generation soil information system for New Zealand. Geoderma 170, 232–238. doi:10.1016/j.geoderma.2011.11.012

Manderson, A., Palmer, A., 2006. Soil information for agricultural decision making: a New Zealand perspective. Soil Use and Management 22, 393–400. doi:10.1111/j.1475-2743.2006.00048.x

Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213, 385–399. doi:10.1016/j.geoderma.2013.08.024

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute Technical Report TR-406. Texas Water Resources Institute, College Station, Texas.

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. Geoderma 214, 91–100. doi:10.1016/j.geoderma.2013.09.024

Podani, J., 1999. Extending Gower's general coefficient of similarity to ordinal characters. Taxon 48, 331–340. doi:10.2307/1224438

Sarmento, E.C., Giasson, E., Weber, E.J., Flores, C.A., Hasenack, H., 2017. Disaggregating conventional soil maps with limited descriptive data: A knowledge-based approach in Serra Gaúcha, Brazil. Geoderma Regional 8, 12–23. doi:10.1016/j.geodrs.2016.12.004

Subburayalu, S., Jenhani, I., Slater, B.K., 2014. Disaggregation of component soil series using possibilistic decision trees from an Ohio County soil survey map. Geoderma 213, 334–345. doi:10.1016/j.geoderma.2013.08.018

Tait, A., Sood, A., Mullan, B., Stuart, S., Bodeker, G., Kremser, S., Lewis, J., 2016. Updated Climate Change Projections for New Zealand for Use in Impact Studies. Synthesis Report RA1.

Thompson, J.A., Prescott, T., Moore, A.C., Bell, J., Kautz, D.R., Hempel, J.W., Waltman, S.W., Perry, C.H., 2010. Regional approach to soil property mapping using legacy data and spatial disaggregation techniques. Presented at the 19th World Congress of Soil Science, Brisbane, Queensland, pp. 17–20.

Valentine, K.W.G., 1981. How soil map units and delineations change with survey intensity and map scale. Canadian Journal of Soil Science 61, 535–551. doi:10.4141/cjss81-063

Vincent, S., Lemercier, B., Berthier, L., Walter, C., 2018. Spatial disaggregation of complex soil map units at the regional scale based on soil-landscape relationships. Geoderma 311, 130–142. doi:10.1016/j.geoderma.2016.06.006

Webb, T.H., Lilburne, L.R., 2011. Criteria for defining the soil family and soil sibling: The fourth and fifth categories of the New Zealand Soil Classification. Landcare Research Science Series. Manaaki Whenua Press, Lincoln, New Zealand. doi:10.7931/DL1-LRSS-3-2011

Woodward, S.J.R., Barker, D.J., Zyskowski, R.F., 2001. A practical model for predicting soil water deficit in New Zealand pastures. New Zealand Journal of Agricultural Research 44, 91–109. doi:10.1080/00288233.2001.9513464