

# Spatial Scaling for Digital Soil Mapping

## Brendan P. Malone\*

Faculty of Agriculture and Environment  
The University of Sydney  
Room 115  
Biomedical Building  
Australia Technology Park  
Eveleigh, NSW 2015, Australia

## Alex B. McBratney

Faculty of Agriculture and Environment  
The University of Sydney  
Room 105  
Biomedical Building  
Australia Technology Park  
Eveleigh, NSW 2015, Australia

## Budiman Minasny

Faculty of Agriculture and Environment  
The University of Sydney  
Room 108  
Biomedical Building  
Australia Technology Park  
Eveleigh, NSW 2015, Australia

We describe in this paper, a broad overview of spatial scale concepts and scaling procedures that are specifically relevant for digital soil mapping (DSM). Despite the recent growth and operational status of DSM, one existing and foreseeably growing issue for users of digital soil information is the inequality of spatial scales between what is required and what is actually available to adequately address soil-related questions posed from within and from outside the soil science community. In the absence of conducting new soil survey or not being able to acquire the original legacy soil information (soil point data) as a means of creating user-specified soil information products, spatial scaling provides a useful solution. Spatial scaling for DSM involves changes in map extent, grid-cell resolution, and prediction support. We review in this paper the different forms of spatial scaling, which are described in terms of changes to grid spacing and prediction support. Fine-gridding and coarse-gridding are operations where the grid spacing changes but support remains unchanged. Deconvolution and convolution are operations where the support always changes which may or may not involve changing the grid spacing. While dissection and conflation operations occur when the support and grid size are equal and both are then changed equally and simultaneously. Some possible and existing pedometric methods are described for implementation of each scaling process, as is an extended example for performing convolution where the support changes yet the resolution remains the same.

**Abbreviations:** AtoP, area-to-point; DSM, digital soil mapping; GCMs, general circulation models; P1, panel 1, P2, panel 2; P3, panel 3; P4, panel 4.

Humans are a dominating force on Earth (Crutzen, 2002). From anthropogenic forcing, we are seeing in many parts of the world, soils being degraded, or through intensive agriculture, have been stripped of vital nutrients to adequately support significant yields (Sanchez, 2010). The critical functions of soils—provisioning of food, fiber, and ecological support in an increasingly populous world—are threatened as a result. To address the manifold issues of soil degradation and nutrient depletion, a broad community of scientists, policy developers and land managers are increasingly turning to the soil science community for relevant and comprehensive information about the status of soils.

We are at a critical time where targeted and objective decision making is of the essence, which in turn needs to be complemented with quantitative modeling, monitoring, and measurement of particular soil services and functions. Digital soil mapping is currently experiencing a precipitous growth from a purely research endeavor to something akin to operational status (Grunwald et al., 2011). The reasons for this are not difficult to surmise—using a combination of sparsely populated legacy soil datasets and numerical inference, populating continuous spatially explicit

Soil Sci. Soc. Am. J. 77:890–902

doi:10.2136/sssaj2012.0419

Received 17 Dec. 2012.

\*Corresponding author ([brendan.malone@sydney.edu.au](mailto:brendan.malone@sydney.edu.au)).

© Soil Science Society of America, 5585 Guilford Rd., Madison WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

soil information databases can be achieved efficiently, and with quantifiable measures of quality or certainty (Lagacherie, 2008).

Currently the availability of digital soil information products encompasses a hierarchy of spatial scales which include global, continental, national, region, farm, and field extents (Grunwald et al., 2011). On a whole, we may not be experiencing a scarcity of comprehensive soil information; rather it is a question of whether the information that is available is relevant or compatible to meet the objectives of a given project or policy directive for a given mapping domain. The incompatibility largely stems from scale dissimilarity between what is required and what is available. Soil information may be available at one spatial scale, but may be required either at a finer or coarser scale and may even be required at a different support or volume (Papritz et al., 2005). For example, digital soil maps created from point support measurements (soil cores, pits, etc.) will generate point support maps which may not be of any use when a policy directive requires the support of predictions to be blocks or a specified land unit size. In the absence of new soil survey, to support the creation of tailored soil information products, we see there is substantial value and efficiency in using existing soil maps as the basis for implementing either upscaling or downscaling methods.

Concepts of spatial downscaling and upscaling have and will continue to have considerable traction in soil science. For example, Finke et al. (1998) detail the breadth of issues with many examples concerning spatial scale in the soil and water sciences domain. McBratney (1998) made some suggestions for a number of possible approaches for upscaling or downscaling soil information problems. Similarly, Bierkens et al. (2000) developed and presented a general framework in the form of a decision tree to detail processes and their models for solving various spatial-scaling problems. Issues of scale incompatibility are not unique to the soil science domain either. In climatology research, outputs of climate simulations from general circulation models (GCMs) cannot be directly used for hydrological impact studies of climate change because of a spatial scale mismatch (Bloschl, 2005). The grid resolution of GCMs is generally in the order of hundreds of kilometers. In contrast, the resolution at which inputs to hydrological impact models are needed is in the order of 10s or 100s of square meters. Practitioners in the remote sensing domain also, to understand the underlying geophysical process of some atmospheric or environmental variables, often use two of more instruments which measure the same processes, but measure it at different spatial supports (Nguyen et al., 2010). As a consequence, spatial-scaling methods are required to combine both information sources for making optimal inferences of the underlying process. Spatial scaling, as an operative procedure is essentially an inference of spatial processes at one resolution from data at another resolution; which in spatial statistics is often called the “change-of-support” problem (Cressie and Wikle, 2011). The change-of-support problem presents many statistical challenges (Heuvelink and Pebesma, 1999) and has been reviewed in Gotway and Young (2002) with other important contributions from Cressie (1996) and Fuentes and Raftery (2005).

The motivation for this paper is the idea that spatial soil information should be available as per the specifications of the user. One way to achieve these ends is via spatial-upscaling and -downscaling methods, for which, in terms specifically for DSM, are described herewith. First, some fundamental concepts of scale are detailed in the context of DSM. Second, we describe some (not exhaustive) existing pedometric techniques or processes that may be implemented for the spatial scaling of digital soil maps. Lastly, we describe some considerations and possible suggestions with regards to equating the uncertainties and validating the outputs that are generated from spatial-scaling procedures.

## THEORY

### The Digital Soil Map Model

The raster model is seen as a useful data structure in which to embed comprehensive and spatially explicit digital soil information, where each pixel or grid-cell (the single unit entity of a raster), which has a spatially explicit location, contains a value for a given target soil attribute (Hengl, 2006). Digital soil maps have three spatial-scale entities: extent, resolution, and support for which Western and Bloschl (1999) termed the scaling triplet when discussing spatial-scaling issues for hydrological modeling. Map extent is the areal expanse or coverage of a mapping domain, such that the map could be a soil map of the world, a country, a region, or a particular farm. Resolution is the grid-cell spacing or pixel size of the raster. A map made up of pixels which have dimensions of 10 by 10 m is a map with a resolution of 10 m. While support is likened to a volume or area. This could either be points—which have no defined area or volume, but generally consist of a soil core or a pit—or blocks (which have a measurable area and/or volume). Bishop et al. (2001) crystallized these fundamental concepts in their description of a generic soil map model. This model consists of a soil variable, which is estimated with some uncertainty. This variable is predicted onto grid-cell spacing,  $G$  that has a support  $B$  which could be a point or a block (Fig. 1).

As Bishop et al. (2001) described, the soil map model equates to a raster model when  $G$  is equal to  $B$  (block support 1 of Fig. 1). Because  $B$  has some definable dimensions or support, the value attributed to it represents an averaged value for that area or volume. When  $B$  is very small, the map model is essentially a grid of points; the support in this case is a point (point support of Fig. 1). While both examples described above have different soil map models, if they have the same  $G$ , they will have the same raster model. However, they are fundamentally different because they have different supports; the values attributed to the pixels mean different things and have different statistical properties. For further complexity,  $B$  may even be bigger than  $G$ , which is quite common in situations where block kriging is used (block support 2 of Fig. 1). A situation where this would be used, as discussed by Bishop et al. (2001), is where a map producer using block kriging may want a dense coverage of information (finely spaced  $G$ ), but to reduce the uncertainty of prediction, may choose a much larger  $B$  than  $G$ .

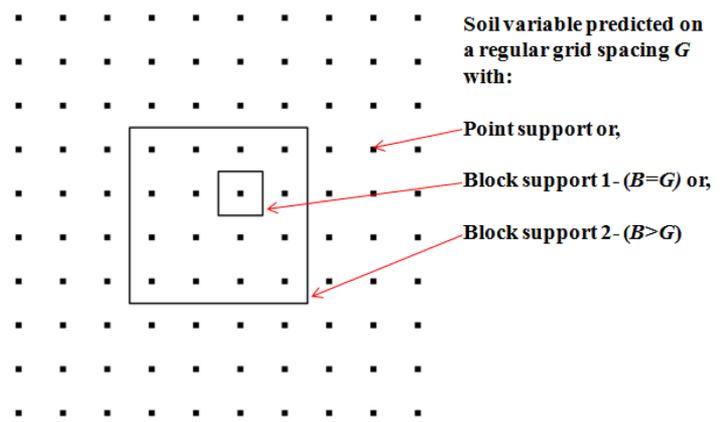
## Spatial Scaling of Digital Soil Maps

The problem is this: A soil map is acquired from a given source (source map) which could be an online repository, a colleague, or a map archive etc. The spatial-scale specifications (while in the same mapping extent) of the source map are mismatched to the desires of the map user. The constraints to efficiently solving this problem are that we cannot use the data which were used to make the source map (because it is not available) nor can we go out into the field to collect new data (it may be too expensive or impractical to do so). Therefore one option is to implement some form of spatial upscaling or downscaling, which is dependent on the specifications of the desired product.

First, adjustments to map extent can and usually are coupled with increasing or decreasing the grid-cell spacing to up-scale or downscale, respectively (McBratney, 1998). Yet upscaling and downscaling as descriptive procedures for spatial scaling in reference to the generic soil map model, may be too general in meaning. For example, spatial downscaling a source point support soil map that has 1-km grid-cell resolution to destination point support map that has 90-m grid-cells is quite different— from a practical and geostatistical point of view (Dungan et al., 2002)— to downscaling to 90-m grid-cell resolution when the block size is defined as 90 by 90 m. The first situation could be crudely referred to as a point-to-point spatial-scaling procedure, while the second is a point-to-block. Aggregation and disaggregation also have an equivalent meaning to upscaling and downscaling in soil science (Bierkens et al., 2000). Yet these terms are also used frequently to describe procedures for combining or separating traditional soil map class/units, respectively. Therefore, in this paper we describe spatial scaling for DSM in reference to the four contrived digital soil maps represented on Fig. 2. All four maps have the same spatial extent. Panel 1 (P1) and Panel 2 (P2) are the same raster model, but different soil map models. The grid spacing is the same but in P1, the support is a point, while in P2 the support is a block where  $B$  has dimensions equal to  $G$ . The situation is the same when comparing Panel 3 (P3) with Panel 4 (P4). Obviously P1 has finer grid spacing than P3 and it should be assumed that in the hierarchy of spatial scales P1 is below P3, meaning a smaller spatial scale and in reality may or may not have a smaller extent. It is also to be assumed that P1 and P2 exist on the same level of the spatial-scale hierarchy.

The spatial-scaling categories are summarized in Table 1 and is to be interpreted by deciding first which soil map model suits the source map with a corresponding row selection. This is followed by a column selection of the digital soil map model that is the desired scale destination of the new map. The row and column coordinate pair then refer to the nature of scaling required to perform the process. We believe that all methods of scaling for DSM can be summarized by three main categories:

1. Fine gridding and coarse gridding: These are situations where  $G$  changes but  $B$  remains unchanged. Examples of fine gridding are situations where spatial scaling requires moving from P3 to

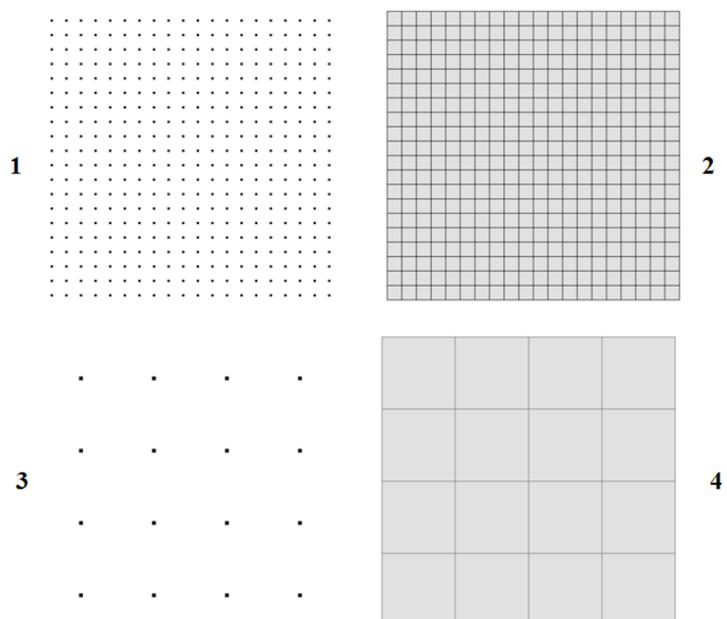


**Fig. 1. Generic soil map model. Support of predictions is point when block  $B$  is very small. Block support prediction occurs when  $B$  has some defined areal value. Block support 1 is when  $B$  equal grid spacing  $G$ . Block support 2 is when  $B$  is greater than  $G$ .  $B$  may be larger than grid spacing. (adapted from Bishop et al. 2001).**

P1 (Fig. 2; P3→P1). While coarse-gridding situations involves P1→P3 spatial scaling.

2. Disseveration and conflation: These are situations when  $B$  and  $G$  are equal and both are changed equally and simultaneously. Examples of disseveration are situations where a P4→P2 spatial scaling is required. While conflation situations involve P2→P4 spatial scaling.

3. Deconvolution and convolution: These are situations where  $B$  always changes which may or may not involve changing  $G$ . However when both  $B$  and  $G$  are equal and changed simultaneously, the changes *are not* equally applied. Convolution processes always involve an increase in  $B$ , with the examples being P1→P2, P1→P4, P3→P2, P3→P4 spatial-scaling operations. While



**Fig. 2. Exemplar soil map models. Panel 1 (P1) and Panel 2 (P2) have the same grid spacing yet P2 is on block support (where block size is equal to the grid spacing), P1 is on point support. Similarly for Panel 3 (P3) and Panel 4 (P4) except the grid spacing is larger.**

**Table 1. Coordinate table of scaling processes based on attributes of source map and scale attributes of destination map.**

Source	Support and (resolution)	Destination			
		Points (fine)	Blocks (fine)	Points (coarse)	Blocks (coarse)
	Points (fine)		Convolution	Coarse gridding	Convolution
	Blocks (fine)	Deconvolution		Deconvolution	Conflation
	Points (coarse)	Fine gridding	Convolution		Convolution
	Blocks (coarse)	Deconvolution	Disseveration	Deconvolution	

deconvolution always involves a decrease in  $B$  such as  $P2 \rightarrow P1$ ,  $P2 \rightarrow P3$ ,  $P4 \rightarrow P1$  or  $P4 \rightarrow P3$  spatial scaling.

### Pedometric Spatial-scaling Methods for Digital Soil Mapping

When considering the hierarchy of spatial scales recognized for soil (Hoosbeek and Bryant, 1992), the  $i$ -levels of interest in the DSM would be global ( $i+6$ ), continental ( $i+5$ ), region ( $i+4$ ), watershed ( $i+3$ ), farm ( $i+2$ ) and field ( $i+1$ ) extents. With these spatial extents in mind, this section will explore in more detail, some pertinent concepts and subsequent pedometric procedures for performing soil spatial scaling (for DSM). For reference, Table 2 details a number of real examples from the soil mapping literature where implementation of different spatial-scaling techniques examples have occurred. We summarize each study by stating the type of spatial scaling that was implemented, based on the categories we have described; the area of the mapping domain (extent); the resolution and support of the source and destination soil map information; and the target variable which underwent spatial scaling. For most of the examples, spatial scaling was not explicitly stated nor the intended purpose or focus of their investigations. Yet the methods they describe—either to map soil, or use soil information that underwent some sort of spatial scaling as a means of investigating other environmental phenomena—exemplify a variety of different pedometric approaches for implementing spatial scaling.

#### Fine Gridding and Coarse Gridding

The most common spatial-scaling methods encountered in DSM would be either fine gridding or coarse gridding. These are spatial-scaling methods where the grid-spacing  $G$  changes without any change of the support, that is,  $B$  remains constant. Fine gridding is a downscaling problem. Alternatively, coarse gridding is an upscaling problem.

Fine gridding requires some form of point interpolation or spatial prediction. A stochastic process such as ordinary punctual (point) kriging may be used to interpolate onto the finely resolved grid nodes. See Isaaks and Srivastava (1989) for more theoretical details of ordinary kriging. Assuming the mean is unknown, the values at the interpolated point locations (fine resolution points) are treated as random variables and are estimated from surrounding point predictions at the coarser scale. The ordinary punctual kriging predictor is:

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i z(x_i) \quad [1]$$

where  $\hat{Z}(x_0)$  is the value of the target variable at unvisited location  $x_0$  which is predicted from a weighted linear combination of  $N$  number of neighboring point observations  $z(x_i)$  at the coarser scale with weights  $\lambda_i$ . To ensure an unbiased estimate, the weights from the vector  $\lambda$  (which is of length  $N$ ) are made to sum to 1, and are obtained by solving the ordinary kriging system:

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \mathbf{A}^{-1} \mathbf{s} \quad [2]$$

where  $\mu$  (which is of length  $\lambda$ ) is a Lagrange multiplier necessary to solve the system,  $\mathbf{A}$  is a matrix with semi-variances between the data points at the coarse resolution and has the structure:

$$\mathbf{A} = \begin{bmatrix} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \dots & \gamma(x_1, x_N) & 1 \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \dots & \gamma(x_2, x_N) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \gamma(x_N, x_1) & \gamma(x_N, x_2) & \dots & \gamma(x_N, x_N) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \quad [3]$$

where  $\gamma$  is semi-variance, obtained from the fitted variogram of the attribute of interest. The length- $N$  vector  $\mathbf{s}$  contains the semi-variances between the coarse-scaled points and the fine-scaled point  $x_0$  and has the structure:

$$\mathbf{s} = \begin{bmatrix} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \vdots \\ \gamma(x_N, x_0) \\ 1 \end{bmatrix} \quad [4]$$

The kriging variance is equated as:

$$\hat{\sigma}^2(x_0) = \mathbf{s}^T \lambda \quad [5]$$

For DSM, the value of  $z(x_i)$  will often be uncertain, such that the exact value will not be known. These uncertainties could be measurement or prediction errors and need to be accounted for in the kriging system. Kriging with uncertain data was introduced by Delhomme (1978) and the method requires some modification of the standard ordinary kriging equations. However, to do this we need to assume: (i) The errors are uncorrelated; (ii) the errors are not correlated with the target variable; and (iii) the variance of the errors is a known quantity and varies from point-to-point (Delhomme, 1978). Under these assumptions, following the formulations from Christensen (2011) the semi-variance elements ( $i, j$ ) of the matrix  $\mathbf{A}$  can be modified on the off-diagonals, that is, where  $i \neq j$  to:

**Table 2. Some examples from the soil literature that have implemented spatial scaling for soil mapping purposes. These studies are summarised based on the spatial scaling categories: Fine-gridding, course-gridding, conflation, deconvolution, disseveration, conflation. Note that for description of spatial support, a value of Block (90m) means a block size of 90 by 90m etc. Source is the existing soil map or information requiring scaling, while Destination is the map or information resultant from scaling.**

Spatial scaling	Location	Extent		Resolution		Support		Target attribute	Author
		Source	Destination	Source	Destination	Source	Destination		
		— Km <sup>2</sup> —		— m —					
Convolution (P1→P2, P4)	Moree, NSW	0.17	0.17	10	10	Point	Blocks (2, 5, 10, 20, 40, 60, 80, 100 m)	Soil potassium concentration	Bishop et al. (2001)
Deconvolution (P4→P1)	Swiss Jura	14.5	14.5	Areal polygons	25	Areal polygons	Point	Soil cobalt concentration	Goovaerts (2011)
Deconvolution (P4→P3)	Northern Ireland	14000	14,000	Areal polygons	Survey sites	Areal polygons	Point	Soil carbon concentration	Kerry et al. (2012)
Fine-gridding (P3→P1)	Hanford, CA	0.32	0.32	25	10	Point	Point	Soil electrical conductivity	Lesch et al. (1995)
Disseveration (P4→P2)	Edgeroi, NSW	1500	1,500	1000	90	Block (1000m)	Block (90 m)	Soil carbon stock	Malone et al. (2012)
Convolution (P1→P4)	Edgeroi, NSW	1500	1,500	90	1000	Point	Block (90 m)	Soil carbon stock	Malone et al. (2012)
Fine-gridding (P3→P1)	Netherlands	25	25	1000	1	Point	Point	Soil zinc concentration	Stein et al. (2001)
Disseveration (P4→P2)	Languedoc-Roussillon region, southern France	65	65	90	15	Block (90m)	Block (15m)	Vegetative evapotranspiration	Taylor et al. (2013)
Conflation (P2→P4)	Southwest Oklahoma	603	603	800	1600; 12,800	Block (800m)	Blocks (1600; 12,800 m)	Soil Moisture	Tsegaye et al. (2003)
Coarse-gridding (P1→P3)	western Oregon	98	98	25	100, 250, 500, 1000	Point	Point	Landcover	Turner et al. (2000)
Conflation (P2→P4)	United States: Independent study areas in Georgia, Indiana, and Washington	1394	1,394	30	60, 120, 210, 240, 480, 960, 1920	Block (30m)	Blocks (60, 120, 210, 240, 480, 960, 1920 m)	Soil type	Lynn Usey et al. (2004)
Convolution (P3→P2)	Chelif Valley, Algeria	380	380	250	100	Point	Block (100 m)	Soil electrical conductivity	Walter et al. (2001)

$$(i, j) \text{ element of } \mathbf{A} = \gamma^*(x_i, x_j) + \frac{\sigma^2(x_i) + \sigma^2(x_j)}{2} \quad [6]$$

where  $\sigma^2$  is the measurement error variance. The  $\gamma^*$  is a bias adjusted semi variance which is detailed further on. The  $i$ th element of the  $\mathbf{s}$  vector is modified to:

$$i\text{th of element of } \mathbf{s} = \gamma^*(x_i, x_0) + \frac{\sigma^2(x_i)}{2} \quad [7]$$

Note that for Eq. [6] and [7] the measurement error variances are halved because we are dealing with semi-variances. Also, bear in mind these adjustments are only valid when the target variable is not correlated with the error variances. While more problematic, Christensen (2011) formulates a method for dealing with correlated error variances for kriging with uncertain data which is based on using variance-stabilizing transformations as proposed by Box and Cox (1964).

Bias adjustment of the semi-variances essentially means correcting the variogram for the measurement errors. This is done by calculating the spatial average of the error variances  $\left[ \frac{1}{N} \sum_{i=1}^N \sigma^2(x_i) \right]$ , then subtracting this average from the variogram (Christensen, 2011). In practice, a semi-variogram is fitted to all  $z(x_i)$  from which the variogram parameters of the nugget, partial sill and range, denoted as  $c_Z$ ,  $v_Z$  and  $r_Z$ , respectively, are obtained. To correct the variogram we simply subtract  $\frac{1}{N} \sum_{i=1}^N \sigma^2(x_i)$  from  $c_Z$ . In general the range and partial sill are unaffected by the bias correction (Christensen, 2011). Occasionally the estimated nugget may be less than the averaged measurement error variance. When  $c_Z - \frac{1}{N} \sum_{i=1}^N \sigma^2(x_i) < 0$ , the nugget for the adjusted semi-variogram can be set to zero and the adjusted partial sill can be set to  $v_Z + c_Z - \frac{1}{N} \sum_{i=1}^N \sigma^2(x_i)$  so that the sill (sum of the nugget plus partial sill) is still reduced by  $\frac{1}{N} \sum_{i=1}^N \sigma^2(x_i)$  (Christensen 2011).

Other forms of fine gridding may involve the use of fine-scaled environmental covariate data, such as that derived from a digital elevation model or some remote sensing platform. Using the coarser-scaled target variable estimates as pseudo-observations, and the fine-scaled covariates as predictors, a deterministic empirical or data mining approach could be implemented. McBratney et al. (2000) provides an extensive review of such *scorpan*-based spatial soil prediction methods. Combining both deterministic and stochastic processes through a regression-kriging approach is also another viable option; the method of which is detailed by Odeh et al. (1995). There is a good logical consistency in transfer between the hierarchies of spatial scales using available covariate information; we know that the variation of soil properties depends on factors such as parent material, climate, land use, and topography. These factors all operate at different scales and therefore influence soil processes and soil variation at different spatial scales (Addiscott, 1993).

Coarse gridding is an upscaling problem and is popularly practiced within Geographical Information Science (GIS) environments through such operations as re-sampling fine-gridded data to a coarser resolution. Nearest-neighbor samplings, in addition to averaging and smoothing spline-type operations, are

popular re-sampling methods. One must be careful that, in the context of coarse gridding,  $B$  remains constant in the scaling procedure. Therefore in the context of the soil map model, regardless of whether an averaging or smoothing spline re-sampling procedure is used, the upscaled soil information product will still be on point support. Effectively, coarse gridding is analogous to throwing some data away (which without good reason is generally undesirable). The purpose of this may be because of a computer memory saving reason; or to align a series of different spatial maps to a common resolution; or that a particular map at a fine scale is difficult to interpret and by performing coarse gridding, the map becomes more general (and interpretable).

## Deconvolution and Convolution

Manipulations of scale that involve changing the support coupled with or without changing the grid spacing involve either deconvolution or convolution. Convolution is an upscaling problem because all situations entail increasing the support of the predictions, for example, point-to-block operations. Deconvolution is a downscaling problem where always the support size is decreased, that is, block-to-point operations. For both convolution and deconvolution, changing the support is always performed, but changing the grid spacing is not always necessary.

## Convolution Problems

There are a few different forms of convolution. First, there is P1→P4 spatial scaling. Here the grid spacing increases in addition to an increase in the support of the predictions. Because each block has many point observations, convolution could involve averaging the point observations contained within each block or pixel (Bierkens et al., 2000), such that:

$$\bar{Z}_H = \frac{1}{N_H} \sum_{i=1}^{N_H} z(x_i) \quad [8]$$

where the prediction  $Z$  with support  $H$  is obtained as an average of all  $z(x_i)$  within  $H$ . The variance is then computed as:

$$\hat{\sigma}^2(\bar{Z}_H) = \frac{1}{N_H - 1} \sum_{i=1}^{N_H} [z(x_i) - \bar{Z}_H]^2 \quad [9]$$

It is necessary to indicate the variance so as to derive a confidence interval about the block average, because the estimate is based only on a limited (not exhaustive) number of points. However the derivation of a confidence interval is based on the assumption that the  $N$  points are independent and that the sample mean follows a normal distribution. Brus and de Gruijter (1997) state that independence can be created through randomization of the point locations. For P1→P4 processes the distribution of the points will not be randomly distributed; they will in fact be regularly spaced points. This means that the spatial coverage of points within each block may be useful in the practical sense of deriving a meaningful block average, yet the suitability of this method from a statistical view is not optimal. While not optimal, the suitability of implementing this particular P1→P4 process will rely on having many (e.g., >50) points within each block.

When there are not a sufficient number of points within a block, ordinary block kriging could be used (Burgess and Webster, 1980). It is such that block kriging rather than punctual kriging computes the mean value of a target variable in a region  $V$  of area  $H$  that centers on a point at  $x_0$ . The block kriging estimator is defined as:

$$\hat{Z}_H(x_0) = \sum_{i=1}^N \lambda_i z(x_i) \quad [10]$$

The predictor  $Z$  with support  $H$  is obtained from a weighted linear combination of  $N$  neighboring point observations  $z(x_i)$ . The weights ( $\lambda_i$ ) are obtained by solving the block kriging system which is the same as that for ordinary punctual kriging. In these cases fine-scaled point observations will be upscaled to the area of  $H$ , which will be set to the dimensions of the pixels. However, one difference between point and block kriging is the nature of the  $s$  vector, such that:

$$s = \begin{bmatrix} \bar{\gamma}(x_1, x_0) \\ \bar{\gamma}(x_2, x_0) \\ \vdots \\ \bar{\gamma}(x_N, x_0) \\ 1 \end{bmatrix} \quad [11]$$

where  $\bar{\gamma}$  is the average semi-variance between  $z(x_i)$  and  $x_0$  which is the block and is the integral:

$$\bar{\gamma}(x_i, H) = \frac{1}{|H|} \int_H \gamma(x_i, x) dx \quad [12]$$

where  $\gamma(x_i, x)$  denotes the semi-variance between the point  $x_i$  and a point  $x$  inside the block. The block kriging variance is equated as:

$$\hat{\sigma}^2(x_0) = s^T \lambda - \bar{\gamma}(H, H) \quad [13]$$

where  $\bar{\gamma}(H, H)$  is the within-block averaged semi-variance value.

Different convolution problems are those that involve P1→P2 or P3→P4 scaling. Consider the situation where a digital soil map may be available at point support where each pixel value represents a single point within the areal extent of the pixel (usually the central node). Without additional sampling, it may be necessary to know what the average of the target variable is across the entire area of each pixel. In this situation, a change of support is required, yet a change in the grid spacing is not applied. We propose that this type of problem could be solved via block kriging. Such that, to increase the support of a point map all that is required is to set the block size  $H$  equal to the grid spacing. An example of this is detailed in the following section.

### Block Kriging Example of Panel 1→Panel 2 Processes

The  $\gamma$  radiometric signal of thorium was collected using a proximal sensing device across fields of a particular farm.

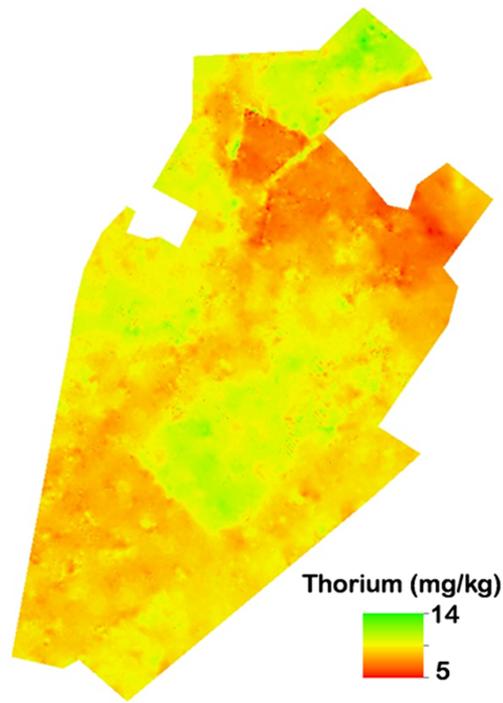


Fig. 3. Five-meter point support thorium concentration (mg/kg) map.

Observation of thorium concentration in mg/kg was made onto a regular grid of points with 5 m spacing (point support map), and is shown on Fig. 3. Each independent observation also has some quantitative value of the measurement error given as a variance which is on average 1.4 mg/kg. This measurement error is due to the instrumentation, meaning that the errors are spatially uncorrelated, and are not correlated with the target variable concentration.

The aim of the example is to create new maps on block support with resolutions of 20, 50, and 80 m, meaning that the block sizes are 20 by 20 m, 50 by 50 m, and 80 by 80 m, respectively. The reason why three increasingly larger resolutions are used is for comparative purposes in assessing the quality the outputs of the P1→P2 procedure.

Using the 5-m point map, a simple way to generate block support maps at these desired resolutions and supports is to average all the observations within each block. This particular procedure is in fact a P1→P4 process (described previously). Because there is some uncertainty about the 5-m spaced observations, we may arithmetically determine the average estimate of thorium concentration in each block as:

$$\bar{Z}_H = \frac{\sum_{i=1}^{N_H} \left( \frac{z_i}{s_i^2} \right)}{\sum_{i=1}^{N_H} \left( \frac{1}{s_i^2} \right)} \quad [14]$$

where  $\bar{Z}_H$  is the weighted averaged value of a block,  $N_H$  is the number of point observations of the target variable  $z_j$  within each block, and with measurement error variance  $s_i^2$ . The variance of  $\bar{Z}_H$  can be estimated by:

$$\text{var}(\overline{Z}_H) = \frac{1}{\sum_{i=1}^{N_H} \frac{1}{s_i^2}} \times \frac{1}{(N_H - 1)} \sum_{i=1}^{N_H} \frac{(z_i - \overline{Z}_H)^2}{s_i^2} \quad [15]$$

Figure 4 shows the block support maps in P1a, P1b and P1c. The spatial average of the  $\text{var}(\overline{Z}_H)$  was  $2 \times 10^{-3}$ ,  $7 \times 10^{-4}$ , and  $5 \times 10^{-4}$  for the 20-, 50-, and 80-m maps, respectively. We could consider these maps as the “true” block support maps and use

them to compare with the outputs of the P1→P2 process, which is described now.

First, using the 5-m point support map, coarse gridding (P1→P3) was performed to generate new maps. For example, to create the 20-m point support map we sampled the 5-m map at grid nodes every 20 m apart and so on. Block kriging based on locally fitted variograms of the nearest 200 sample point support observations is used to create the desired block support maps where the support  $H$  is set to the same size as the map resolution. However, because there is uncertainty about the true values of all  $z_i$  expressed as prediction or measurement error variances, there is a need to modify the standard ordinary block kriging equations. We can assume the error variances are independent of  $z_i$  and carry out what was described for punctual kriging with uncertain data, by modifying the  $A$  matrix and  $s$  vector accordingly which are formulated in Eq. [6] and [7], respectively. Similarly the variogram of  $z_i$  is adjusted to correct for the bias due to measurement errors where the spatial average of the error variances ( $\frac{1}{N} \sum_{i=1}^N s_i^2$ ) is subtracted from the variogram (Christensen, 2011). Because block kriging is being used, the  $i$ th elements of the  $s$  vector is modified to:

$$i\text{th of element of } s = \overline{\gamma}(x_i, x_0) + \frac{s_i^2}{2} \quad [16]$$

The block kriging variance is equated as in Eq. [13].

Panel 2 of Fig. 4 shows the maps that resulted from block kriging with uncertain data for each of the three resolutions and supports. For a comparative exercise, block kriging without including the uncertainties using the standard ordinary block kriging equations was also performed and the maps are shown in P3 of Fig. 4. While quite similar to the true block support maps, including the measurement error variances into the kriging equations resulted in smoother representations of thorium concentration at each of the three supports. The spatial average of the kriging variances from kriging with the uncertain data were  $3 \times 10^{-2}$ ,  $6 \times 10^{-2}$ , and  $9 \times 10^{-2}$  for the 20-, 50-, and 80-m maps, respectively. The spatial averages of the kriging variances when not including the error variances was  $2 \times 10^{-3}$ ,  $3 \times 10^{-2}$ , and  $5 \times 10^{-2}$  for the 20-, 50-, and 80-m

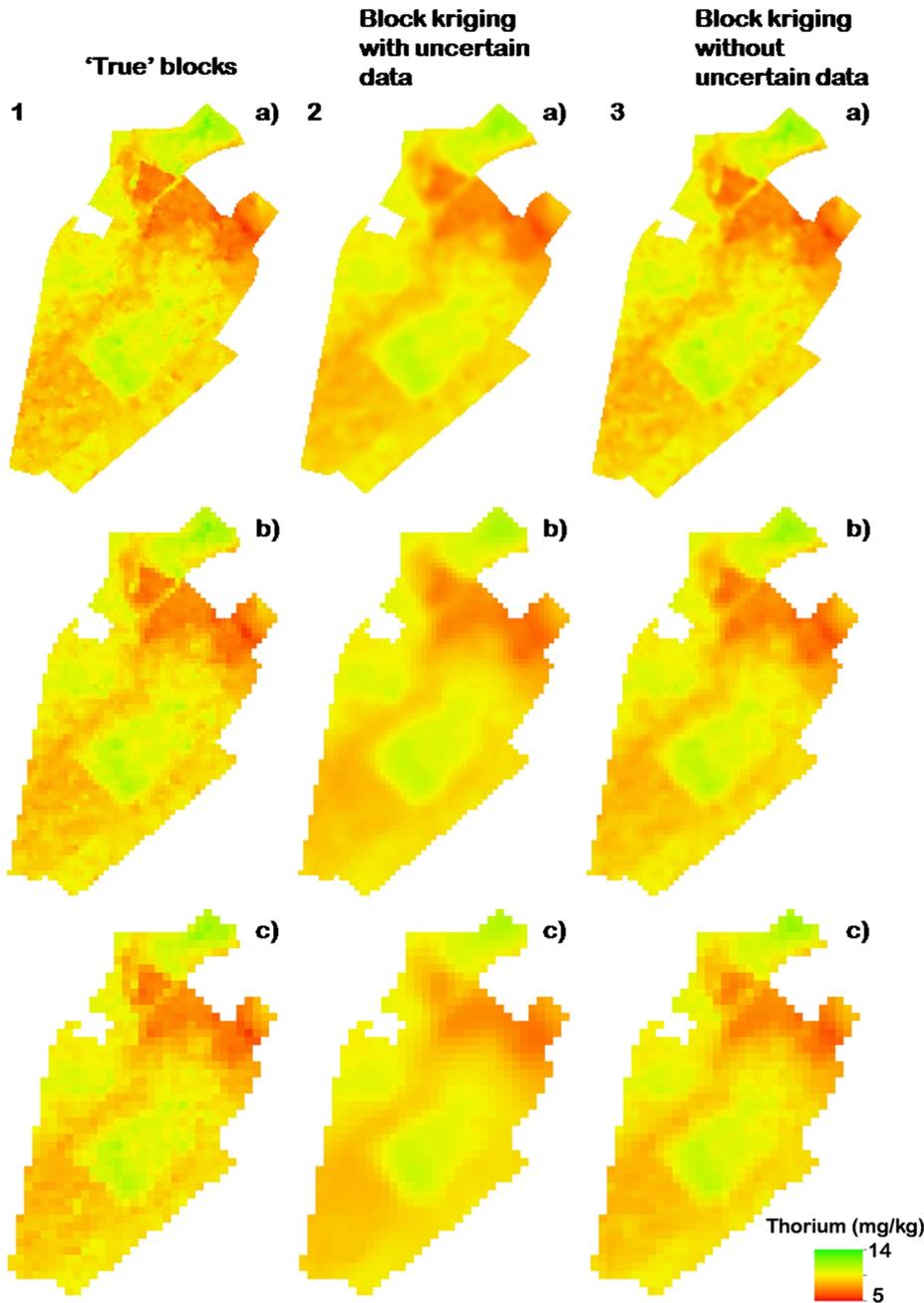


Fig. 4. Block support maps of thorium concentration where support size equals grid-cell size (resolution)- (a) 20 m, (b) 50 m, and (c) 80 m. Panel 1 (P1): true blocks created directly from 5-m point support map with a P1→Panel 4 (P4) process-weighted averaging. Panel 2 (P2): P1→P2 process (Block kriging with uncertain data). Panel 3: P1→P2 process (Block kriging without including uncertainties).

maps, respectively. Essentially what these results represent is that for P1→P2 processes, the uncertainty increases with increasing resolution and support size. Logical also is the fact that the uncertainties are higher when kriging is performed using uncertain data compared with when the data are assumed to be without error. In any case the spatial averages of the kriging variances from both methods were higher than found for the true blocks, which is to be expected.

The plots in Fig. 5 illustrate the similarity of the maps resulting from kriging with uncertain data with the true block maps. With 20-m blocks, concordance (Lin, 1989) was quantified as 0.97 while the root mean square error of prediction (RMSE) was found to be 0.25. This indicates a high degree of similarity between the true and predicted map. By not including the measurement error variances however, both the true and predicted maps are very close to identical where a concordance of 0.99 and RMSE of 0.13 was quantified. Similarly with the 50-m blocks, the map resulting from kriging with uncertain data is a very good representation of the true block map (concordance = 0.95, RMSE = 0.29). With 80-m blocks the concordance was found to be 0.93 and RMSE was 0.31. At these two supports (50 and 80 m), kriging without uncertain data resulted concordance of 0.98 between the predicted block support maps and the true block map, while the RMSE was 0.17 and 0.20, respectively.

From this example of a P1→P2 process, block kriging tends to work better when smaller supports and resolutions are used. This is because more data close to the location where a prediction is to be made is available. This type of phenomena has previously been reported in Costanza and Maxwell (1994). When uncertain data are used for kriging, the resulting maps will be smoother than when they are not. Empirically from this example, this is because more weighting (from the kriging weights) is assigned to points further away from the location where a prediction is to be made.

### Further Convolution Problems

Convolution problems could also involve situations where one requires

a process for scaling from P3→P2. The purpose for these processes may be that in addition to requiring point predictions to be expressed on an areal support, the target variable information is needed at a finer resolution to what is currently available. It is possible to achieve this directly through such methods as ordinary block kriging or universal block kriging. Because there is a need to describe the variation of a target variable at a finer resolution, ordinary block kriging would suit in situations where no

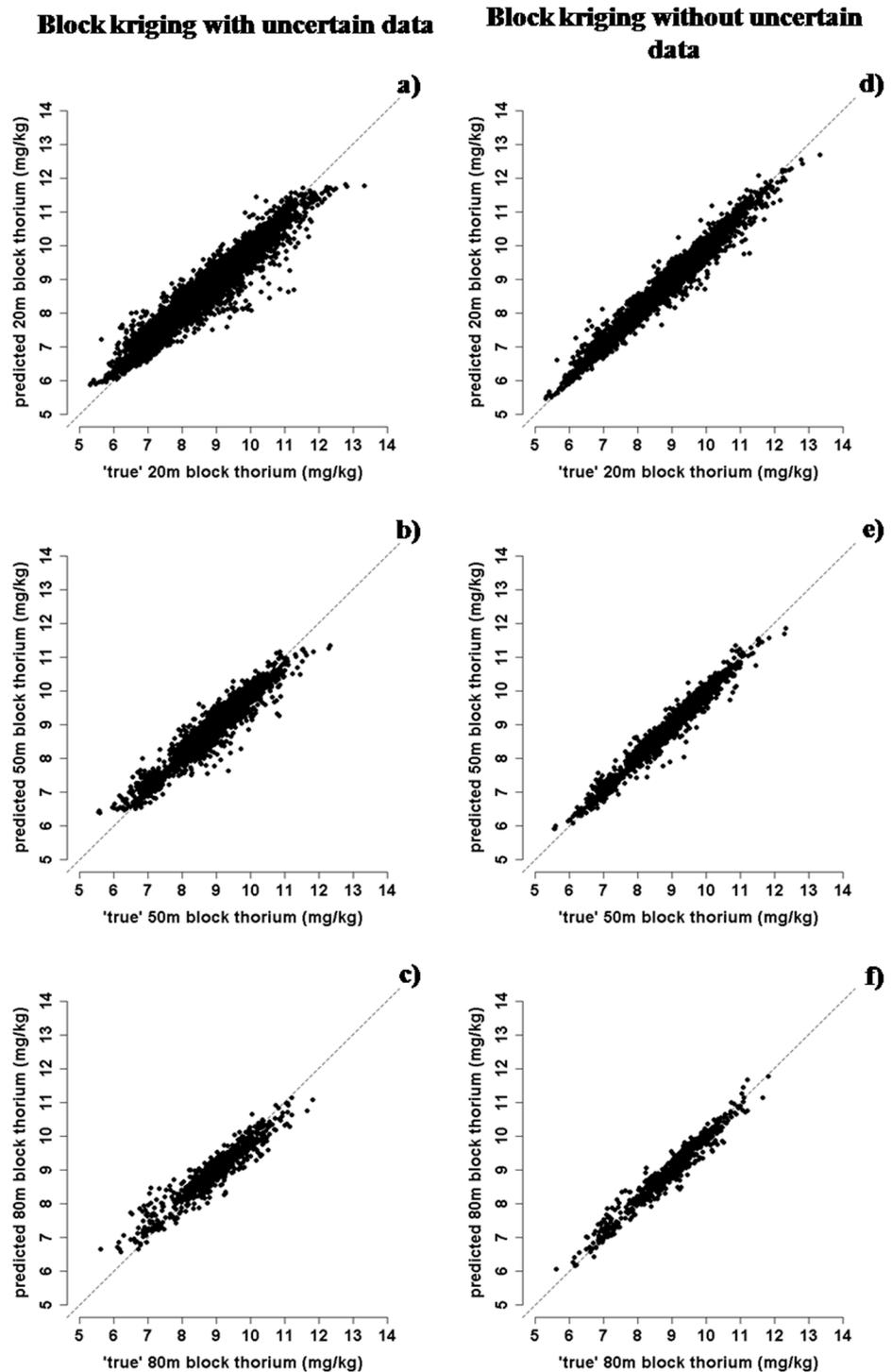


Fig. 5. Comparisons between true block maps with maps from block kriging with uncertain data (a) 20 m, (b) 50 m, and (c) 80 m. Comparisons between true block maps with maps from block kriging (d) 20 m, (e) 50 m, and (f) 80 m.

available covariate information is available. A preferable alternative is where covariate information is available, for which universal block kriging or kriging with external drift would be suited. Universal kriging may be described as some spatial process which comprises both stochastic and deterministic components and represented by the general model:

$$Z(x) = \sum_{k=0}^K a_k f_k(x) + \varepsilon(x) \quad [17]$$

The deterministic component is represented in the above equation by a set of functions (usually first or second order polynomials),  $f_k(x)$ ,  $k = 0, 1, \dots, K$ , and unknown coefficients  $a_k$  which need to be estimated based on the relationship between the target variable and covariates. The  $\varepsilon(x)$  term is the stochastic field with zero mean. A block universal kriging estimate of a target variable centred by a point  $x_0$  based on  $N$  point observations at neighbouring sites is:

$$Z_H(x_0) = \sum_{k=0}^K \sum_{i=0}^N a_k \lambda_i f_k(x_i) \quad [18]$$

where  $\lambda_i$  are the kriging weights. More detail regarding universal kriging can be found in Webster and Oliver (2001).

## Deconvolution Problems

Deconvolution is a downscaling problem which involves a decrease in spatial support such as acquiring point estimates from areal information. This type of procedure is not uncommon in soil map disaggregation exercises where a map producer will require some method to discretize points within polygons before generating soil attribute maps (Goovaerts, 2011). Area-to-point (AtoP) kriging (Kyriakidis, 2004) is one, and a natural candidate for implementing deconvolution. Area-to-point kriging is essentially the counterpart of block kriging in that point estimates are obtained from areal (block) measurements. In the case of digital soil map deconvolution, each pixel is a block where the pixel value is some spatially averaged estimate of the target variable. The idea of AtoP kriging for deconvolution is therefore to use this areal information to discretize point estimates on a regular grid spacing as defined by the map producer. The AtoP kriging estimate for any given point  $x_0$  is expressed as:

$$Z_{AtoP}(x_0) = \sum_{k=1}^K \lambda_k Z(v_k) \quad [19]$$

where  $K$  is the number of areal data  $Z(v_k)$  encapsulating and surrounding the point  $x_0$ . Generally, areal data are chosen according to adjacency rules, such that the encapsulating areal datum and all its adjacent areal data are used for prediction (Goovaerts, 2010). A key property of AtoP kriging is that it preserves the mass-balance or pycnophylactic property of the areal data. The mass-balance property means that the average of all discretized points within each  $v_k$  returns the areal value of  $Z(v_k)$ . However, the constraint imposed on Eq. [19] is that the same  $K$  areal data are used for prediction at each location within the block

$v_k$  where the point estimations are required (Goovaerts, 2011). Furthermore, because areal or block estimates are used to derive point predictions, there is a requirement to know the point support variogram model. Obviously this is not available but can be evaluated in two steps: (i) compute and model the variogram of the areal data and (ii) deconvolute the block-support model to derive the point support variogram. Goovaerts (2008) proposed an iterative deconvolution procedure that seeks the point support model that, once regularized, is the closest to the model fit of the areal data.

## Conflation and Disseveration

Scaling problems where both the source and destination maps both have some sort of areal support may be conditionally approached with either conflation or disseveration. Conflation and disseveration procedures deal strictly with processes where the support and the grid spacing are equal and both are changed equally and simultaneously. In accordance with Fig. 2 conflation processes require spatial scaling from P2→P4. A conflation process would be performed where given a large project area extent, a map producer requires regional predictions of a target variable using available fine-scaled areal estimates such as those derived for farm extents. Conflation here involves the averaging of the fine-scaled areal observations within each coarse-scaled block. With this upscaling procedure, while the overall mean of the target variable across the same map extents will remain unchanged, the overall variance will decline as the block and grid spacing simultaneously increase. The decline in variance will, with increasing resolution, result in the creation of homogenous maps.

Disseveration is analogous to P4→P2 scaling. Here the requirement is that in addition to needing a method for estimating the variation of the target variable at a fine resolution (given that only the value at the coarse resolution is known) there is a need to maintain the pycnophylactic property whereby the target variable value given for each coarse grid-cell equals the average of all target variable values at the fine scale in each coarse grid-cell. This additional requirement of mass preservation is the explicit difference between downscaling methods that involve simply fine gridding which is essentially a points-to-points procedure and those which involve disseveration which is a block-to-block procedure.

An example of where disseveration would be enacted would be in a situation where regional estimates of a target variable (at block support) are available only at a coarse resolution and there is a requirement to generate estimates of this property to a farm or even field extent. It is therefore quite reasonable to expect that downscaling here also involves a reduction of the areal extent in addition to reduction of the spatial resolution.

Methods for disseveration include AtoP kriging as discussed previously. Alternatively, when covariate information is available, iterative processes of model fitting and adjustment (in an attempt to optimize the downscaling by preserving the mass balance) may be used and for which one procedure is described in Malone et al. (2012).

## Further Solutions for Scaling Problems

The spatial-scaling processes described so far are specific for a given problem. For example, block kriging for  $P1 \rightarrow P4$  processes (convolution), point kriging for  $P3 \rightarrow P1$  processes (fine gridding), AtOP kriging for  $P4 \rightarrow P1$  processes (deconvolution) and disseveration for  $P4 \rightarrow P2$  processes. Recently however, Gotway and Young (2007) introduced a generalized geostatistical framework that in addition to solving the problems of scaling described in this paper can also be implemented for other problems that cannot be visualized by using the contrived soil maps on Fig. 2. For example, deconvolution problems that are block-to-block processes or problems that involve overlapping supports (which are described as side-scaling problems). The idea of Gotway and Young (2007) is that data of any kind of support whether it be point or block is  $Z(\mathbf{B}) = Z(B_1), \dots, Z(B_n)$  and prediction of  $Z(A)$  is of interest. The volumes  $A$  and  $\mathbf{B}$  can be general which allows for several different types of scaling problems. For example, block kriging is a special case of this method when  $A$  is a volume or area and  $B_i$  are points. If  $A$  is a point and  $B_i$  is a volume or area where  $A$  is nested within  $B_i$ , the problem becomes one of deconvolution and the principals of mass balance are preserved. Gotway and Young (2007) detail the statistical inference of this framework and its demonstration of use is also given in Young and Gotway (2007). Obvious advantages of this framework are its versatility for solving a range of scaling problems with one method, negating the requirement for resorting to specific solutions for a given problem. Furthermore, measures of uncertainty can be obtained for the predictions.

## Validation of Soil Information Products Generated from Scaling Methods

Some comment is necessary on steps to assess the validity of outputs generated from spatial-scaling source digital soil maps. This is particularly important in the context that the digital soil information may be used for decision making, management, or modeling purposes.

It is well established that the spatial prediction of soil is inherently difficult. Consequently, the source maps will have some level of uncertainty attributed to it. In kind, these uncertainties will also propagate through to the destination map, which is added to the uncertainty associated with the actual spatial-scaling procedure. More often than not the uncertainties are very rarely included with a source map, which is not ideal. When uncertainties are available, incorporating them into the spatial-scaling process is desirable, for example, kriging with uncertain data (Delhomme 1978). Furthermore, as described for disseveration, the method or program presented by Malone et al. (2012)—*dissever*—allows one to incorporate uncertainties of the source map into the process for creating the destination map.

More generally, an implied and over-arching assumption of the spatial-scaling methods discussed is that the behavior of soil at large scales is explained by the average of the soil behavior at fine scales. This may or may not be upheld in reality or may only be relevant at a specific range of scales (Addiscott and Mirza,

1998). Intuitively the specific range of spatial scales maybe those relevant for DSM, that is,  $i+1$  to  $i+6$ . Grunwald et al. (2011), citing deYoung et al. (2008), does explain however that nonlinear dynamics and alternate states are well known in ecological systems, yet they have been poorly investigated in the soil science domain. It is beyond the scope of this paper, but a pragmatic way to investigate the variance of soil properties at different scales, as Pettitt and McBratney (1993) suggests, is by performing nested sampling which will additionally help recognize the existence of natural hierarchies. Lark (2005) also described the value of nested sampling for understanding soil processes at different scales. Understanding the dynamics of soil processes better at different spatial scales will obviously complement efforts when scaling of existing soil information is required.

Ultimately, the outputs from spatial scaling will require some form of validation to assess their quality. When kriging operations are performed, the kriging prediction error provides a quantitative and spatially explicit measure of the uncertainty. Otherwise, internal validations from diagnostic measures such as the coefficient of determination or the RMSE, among others, provide some way of assessing the validity of outputs. This would be the case if a *scorpan* model was used for a fine-gridding operation; similarly for disseveration with covariate information. However these internal validations may be susceptible to bias (Brus et al., 2011) and the kriging prediction variances as exemplified in the example of convolution ( $P1 \rightarrow P2$ ) will underestimate or oversimplify the true prediction uncertainty.

Brus et al. (2011) propose that to unbiasedly estimate map quality, one needs to collect additional samples from the mapping domain of interest. The recommendation is that a design-based sampling strategy involving probability sampling be implemented. There are associated costs required to implement a sampling for validation program, but more importantly there are some things to consider in terms of prediction support. For example, when the support of observations is a point, the external validation is not a technically difficult exercise. One just needs to come up with an appropriate design and subsequent sampling configuration of  $N$  number of points. However, validation of digital soil maps with some sort of areal support is less straightforward to implement. Lagacherie et al. (2012) pragmatically proposed that block averages can be validated by first performing a validation (here cross-validation) of a model using the same data on point support. The obvious advantage of this is that no new data need to be collected. However, probability sampling is more optimal. Implementation would require a sample collection at a limited number of point locations within randomly selected validation supports. The average of the soil variable at these locations would be assumed as representative of the entire support unit. The question is then, how many samples are required from each block support unit? It seems intuitively attractive to sample twice with each unit, but it could be argued that more than this is necessary. Therefore further research and examples are recommended to determine an optimal, efficient, and general scheme for validating block support maps.

## CONCLUSIONS

Tailoring digital soil information products to the specifications of the end-user is likely to become the norm into the future. One important challenge in the soil mapping community is how can we use existing soil maps that are not necessarily ideal in terms of scale representation. Therefore, it is timely to open the discussion in the soil mapping community, first of some general spatial scale concepts relevant for DSM. We did this in terms of the generic soil map model previously introduced by Bishop et al. (2001). We then categorized three main spatial-scaling types that vary in terms of changes in resolution and/or prediction support. Fine gridding and coarse gridding are operations where the grid spacing changes but support remains unchanged. Deconvolution and convolution are situations where the support always changes, which may or may not involve changing the grid spacing. Disseveration and conflation operations occur when the support and grid size are equal and both are then changed equally and simultaneously. We have not attempted to describe the full suite of pedometric methods that may be used to perform each type of spatial scaling. Yet to initiate further discussion we describe a few likely candidates for each. Some immediate challenges exist in terms of quantification of uncertainties and validation (particularly of block support maps) of the maps or outputs resultant from spatial scaling.

## REFERENCES

- Addiscott, T.M. 1993. Simulation modelling and soil behaviour. *Geoderma* 60:15–40. doi:10.1016/0016-7061(93)90016-E
- Addiscott, T.M., and N.A. Mirza. 1998. New paradigms for modelling mass transfers in soils. *Soil Tillage Res.* 47:105–109. doi:10.1016/S0167-1987(98)00081-6
- Bierkens, M.F.P., P.A. Finke, and P. de Willigen. 2000. Upscaling and downscaling methods for environmental research. Kluwer, Academic Publ., Dordrecht, the Netherlands.
- Bishop, T.F.A., A.B. McBratney, and B.M. Whelan. 2001. Measuring the quality of digital soil maps using information criteria. *Geoderma* 103:95–111. doi:10.1016/S0016-7061(01)00071-4
- Bloschl, G. 2005. Statistical upscaling and downscaling in hydrology. In: M.G. Anderson and J.J. McDonnell, editors, *Encyclopaedia of hydrological sciences*. John Wiley & Sons, Chichester, West Sussex, England.
- Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. *J. R. Stat. Soc. Series B Stat. Methodol.* 26:211–252.
- Brus, D.J., and J.J. deGrujter. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* 80:1–44. doi:10.1016/S0016-7061(97)00072-4
- Brus, D.J., B. Kempen, and G.B.M. Heuvelink. 2011. Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* 62:394–407. doi:10.1111/j.1365-2389.2011.01364.x
- Burgess, T.M., and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. 2. Block kriging. *J. Soil Sci.* 31:333–341. doi:10.1111/j.1365-2389.1980.tb02085.x
- Christensen, W.F. 2011. Filtered kriging for spatial data with heterogeneous measurement error variances. *Biometrics* 67:947–957. doi:10.1111/j.1541-0420.2011.01563.x
- Costanza, R., and T. Maxwell. 1994. Resolution and predictability: An approach to the scaling problem. *Landscape Ecol.* 9:47–57. doi:10.1007/BF00135078
- Cressie, N. 1996. Change of support and the modifiable areal unit problem. *Geographical Systems* 3:159–180.
- Cressie, N., and C.K. Wikle. 2011. *Statistics for spatio-temporal data*. John Wiley & Sons, Hoboken, NJ.
- Crutzen, P.J. 2002. Geology of mankind. *Nature (London)* 415:23. doi:10.1038/415023a
- Delhomme, J.P. 1978. Kriging in the hydrosocieties. *Adv. Water Resour.* 1:251–266. doi:10.1016/0309-1708(78)90039-8
- deYoung, B., M. Barange, G. Beaugrand, R. Harris, R.I. Perry, M. Scheffer, and F. Werner. 2008. Regime shifts in marine ecosystems: Detection, prediction and management. *Trends in Ecology & Evolution* 23:402–409. doi:10.1016/j.tree.2008.03.008
- Dungan, J.L., J.N. Perry, M.R.T. Dale, P. Legendre, S. Citron-Pousty, M.J. Fortin et al. 2002. A balanced view of scale in spatial statistical analysis. *Ecography* 25:626–640. doi:10.1034/j.1600-0587.2002.250510.x
- Finke, P.A., J. Bouma, and M.R.E. Hoosbeek. 1998. *Soil and water quality at different scales*. Kluwer, Dordrecht, the Netherlands.
- Fuentes, M., and A.E. Raftery. 2005. Model evaluation and spatial interpolation by Bayesian combinations of observations with outputs from numerical models. *Biometrics* 61:36–45. doi:10.1111/j.0006-341X.2005.030821.x
- Goovaerts, P. 2008. Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences* 40:101–128. doi:10.1007/s11004-007-9129-1
- Goovaerts, P. 2010. Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. *Mathematical Geosciences* 42:535–554. doi:10.1007/s11004-010-9286-5
- Goovaerts, P. 2011. A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. *Eur. J. Soil Sci.* 62:371–380. doi:10.1111/j.1365-2389.2011.01368.x
- Gotway, C.A., and L.J. Young. 2002. Combining incompatible spatial data. *J. Am. Stat. Assoc.* 97:632–648. doi:10.1198/016214502760047140
- Gotway, C.A., and L.J. Young. 2007. A geostatistical approach to linking geographically aggregated data from different sources. *J. Comput. Graph. Statist.* 16:115–135. doi:10.1198/106186007x179257
- Grunwald, S., J.A. Thompson, and J.L. Boettinger. 2011. Digital soil mapping and modelling at continental scales: Finding solutions for global issues. *Soil Sci. Soc. Am. J.* 75:1201–1213. doi:10.2136/sssaj2011.0025
- Hengl, T. 2006. Finding the right pixel size. *Comp. Geosci.* 32:1283–1298. doi:10.1016/j.cageo.2005.11.008
- Heuvelink, G.B.M. and E.J. Pebesma. 1999. Spatial aggregation and soil process modelling. *Geoderma* 89:47–65. doi:10.1016/S0016-7061(98)00077-9
- Hoosbeek, M.R., and R.B. Bryant. 1992. Towards the quantitative modelling of pedogenesis-A review. *Geoderma* 55:183–210. doi:10.1016/0016-7061(92)90083-J
- Isaaks, E.H., and R.M. Srivastava. 1989. *An introduction to applied geostatistics*. Oxford Univ. Press, New York.
- Kerry, R., P. Goovaerts, B.G. Rawlins, and B.P. Marchant. 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
- Kyriakidis, P.C. 2004. A geostatistical framework for area-to-point spatial interpolation. *Geogr. Anal.* 36:259–289. doi:10.1111/j.1538-4632.2004.tb01135.x
- Lagacherie, P. 2008. Digital soil mapping: A state of the art. In: A.E. Hartemink, A.B. McBratney, and M.L. Mendonça Santos, editors, *Digital soil mapping with limited data*. Springer Sci., Australia, p. 3–14.
- Lagacherie, P., J.S. Bailly, P. Monestiez, and C. Gomez. 2012. Using scattered hyperspectral imagery data to map the soil properties of a region. *Eur. J. Soil Sci.* 63:110–119. doi:10.1111/j.1365-2389.2011.01409.x
- Lark, R.M. 2005. Exploring scale-dependent correlation of soil properties by nested sampling. *Eur. J. Soil Sci.* 56:307–317. doi:10.1111/j.1365-2389.2004.00672.x
- Lesch, S.M., D.J. Strauss, and J.D. Rhoades. 1995. Spatial prediction of soil salinity using electromagnetic induction techniques. 1. Statistical prediction models- a comparison of multiple linear regression and co-kriging. *Water Resour. Res.* 31:373–386. doi:10.1029/94wr02179. doi:10.1029/94WR02179
- Lin, L.I. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268. doi:10.2307/2532051
- Malone, B.P., A.B. McBratney, B. Minasny, and I. Wheeler. 2012. A general method for downscaling earth resource information. *Comput. Geosci.* 41:119–125. doi:10.1016/j.cageo.2011.08.021
- McBratney, A.B. 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutr. Cycl. Agroecosyst.* 50:51–62. doi:10.1023/a:1009778500412. doi:10.1023/A:1009778500412
- McBratney, A.B., I.O.A. Odeh, T.F.A. Bishop, M.S. Dunbar, and T.M. Shatar. 2000. An overview of pedometric techniques for use in soil survey.

- Geoderma 97:293–327. doi:10.1016/S0016-7061(00)00043-4
- Nguyen, H., N. Cressie, and A. Braverman. 2010. Spatial statistical data fusion for remote-sensing application. Technical Rep. 849. Dep. of Statistics, The Ohio State Univ., Columbus.
- Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes- heterotopic cokriging and regression kriging. *Geoderma* 67:215–226. doi:10.1016/0016-7061(95)00007-B
- Papritz, A., C. Herzig, F. Borer, and R. Bono. 2005. Modelling the spatial distribution of copper in the soils around a smelter in north-eastern Switzerland. In: P. Renard, H. Demougeot-Renard, and R. Froidevaux, editors, *Geostatistics for environmental applications*. Springer-Verlag, Berlin, Heidelberg, p. 343–354.
- Pettitt, A.N., and A.B. McBratney. 1993. Sampling designs for estimating spatial variance components. *J. R. Stat. Soc. Ser. C Appl. Stat.* 42:185–209.
- Sanchez, P.A. 2010. Tripling crop yields in tropical Africa. *Nat. Geosci.* 3:299–300. doi:10.1038/ngeo853
- Stein, A., J. Riley, and N. Halberg. 2001. Issues of scale for environmental indicators. *Agric. Ecosyst. Environ.* 87:215–232. doi:10.1016/S0167-8809(01)00280-8
- Taylor, J.A., F. Jacob, M. Galleguillos, L. Prevot, N. Guix, and P. Lagacherie. 2013. The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping). *Geoderma* 193:83–93. doi:10.1016/j.geoderma.2012.09.009
- Tsegaye, T.D., W.L. Crosson, C.A. Laymon, M.P. Schamschula, and A.B. Johnson. 2003. Application of neural network-based spatial disaggregation scheme for addressing scaling of soil moisture. In: Y. Pachepsky, D.E. Radcliffe, and H. Magdi Selim, editors, *Scaling Methods in Soil Physics*. CRC Press, USA.
- Turner, D.P., W.B. Cohen, and R.E. Kennedy. 2000. Alternative spatial resolutions and estimation of carbon flux over a managed forest landscape in Western Oregon. *Landscape Ecol.* 15:441–452. doi:10.1023/a:1008116300063. doi:10.1023/A:1008116300063
- Walter, C., A.B. McBratney, A. Douaoui, and B. Minasny. 2001. Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. *Aust. J. Soil Res.* 39:259–272. doi:10.1071/sr99114. doi:10.1071/SR99114
- Webster, R., and M.A. Oliver. 2001. *Geostatistics for environmental scientists*. John Wiley & Sons, West Sussex, England.
- Western, A.W., and G. Bloschl. 1999. On the spatial scaling of soil moisture. *J. Hydrol.* 217:203–224. doi:10.1016/S0022-1694(98)00232-7
- Young, L.J., and C.A. Gorway. 2007. Linking spatial data from different sources: The effects of change of support. *Stochastic Environ. Res. Risk Assess.* 21:589–600. doi:10.1007/s00477-007-0136-z