Published March 25, 2013

Pedology

High-Resolution 3-D Mapping of Soil Texture in Denmark

Kabindra Adhikari*
Rania Bou Kheir
Mette B. Greve
Dep. of Agroecology
Faculty of Science and Technology
Aarhus Univ.
Blichers Allé 20
P.O. Box 50, DK-8830 Tjele
Denmark

Peder K. Bøcher
Ecoinformatics and Biodiversity Group
Dep. of Bioscience
Aarhus Univ.
Ny Munkegade 114
DK-8000 Aarhus C
Denmark

Brendan P. Malone
Budiman Minasny
Alex B. McBratney
Dep. of Environmental Sciences
Faculty of Agriculture and Environment
The Univ. of Sydney, Biomedical
Building C81
1 Central Avenue
Eveleigh, NSW 2015
Australia

Mogens H. Greve
Dep. of Agroecology
Faculty of Science and Technology
Aarhus Univ.
Blichers Allé 20
P.O. Box 50, DK-8830 Tjele
Denmark

Soil texture which is spatially variable in nature, is an important soil physical property that governs most physical, chemical, biological, and hydrological processes in soils. Detailed information on soil texture variability both in vertical and lateral dimensions is crucial for proper crop and land management and environmental studies, especially in Denmark where mechanized agriculture covers two thirds of the land area. We modeled the continuous depth function of texture distribution from 1958 Danish soil profiles (up to a 2-m depth) using equal-area quadratic splines and predicted clay, silt, fine sand, and coarse sand content at six standard soil depths of GlobalSoilMap project (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm) via regression rules using the Cubist data mining tool. Seventeen environmental variables were used as predictors and their strength of prediction was also calculated. For example, in the prediction of silt content at 0 to 5 cm depth, factors that registered a higher level of importance included the soil map scored (90%), landscape types (54%), and landuse (27%), while factors with lower scores were direct insolation (17%) and slope aspect (14%). Model validation (20% of the data selected randomly) showed a higher prediction performance in the upper depth intervals but increasing prediction error in the lower depth intervals (e.g., R^2 = 0.54, RMSE = 33.7 g kg^{-1} for silt 0–5 cm and R^2 = 0.29, RMSE = 38.8 g kg^{-1} from 100–200 cm). Danish soils have a high sand content (mean values for clay, silt, fine sand, and coarse sand content for 0- to 5-cm depth were 79, 84, 324, and 316 g kg^{-1}, respectively). Northern parts of the country have a higher content of fine sand compared to the rest of the study area, whereas in the western part of the country there was little clay but a high coarse sand content at all soil depths. The eastern and central parts of the country are rich in clay, but due to leaching, surface soils are clay eluviated with subsequent accumulation at lower depths. We found equal-area quadratic splines and regression rules to be promising tools for soil profile harmonization and spatial prediction of texture properties at national extentacross Denmark.

Abbreviations: AIC, akaike information criteria; C, categorical data; DEM, digital elevation model; DSM, digital soil mapping; GIS, geographic information system; lidar, light detection and ranging; LSP, land surface parameters; ME, mean error; MFD, multiple-flow direction; MRVBF, multi-resolution index of valley bottom flatness; Q, quantitative data; RMSE, root mean square error; RNE, relative nugget effect; SAGA, system for automated geoscientific analyses; SD, standard deviation; TIN, triangular irregular network; TWI, topographic wetness index.

Soil, essentially a nonrenewable natural resource, is a dynamic and complex biomaterial (Young and Crawford, 2004), supports all terrestrial life on earth, and acts as a foundation for several environmental processes. Soil texture, which is spatially variable in nature (Burrough, 1993), is one of the most important physical properties, as it governs several physical, chemical, biological, and hydrological properties and processes in soils. Spatial variability in texture contributes to variations

Soil Sci. Soc. Am. J. 77:860–876
doi:10.2136/sssaj2012.0275
Received 28 Aug. 2012.
*Corresponding author (kabindra.adhikari@agrsci.dk).

in soil performance and crop yield (Warrick and Gardner, 1983; Tanji, 1996). It also affects soil moisture retention and availability (Crave and Gascuel-Odoux, 1997), soil aggregation, and risk of splash erosion (Luk, 1979). Therefore, a better understanding of such a vital resource is important for efficient farm management and environmental studies.

Denmark is one of the best mapped countries in the world in terms of its data coverage and availability. One of these data sets is a soil texture database holding more than one observation per km^2. Another data set is a digital elevation model with 1.6-m resolution, which is based on lidar scanning (light detection and ranging). Both data sets are freely available for the research communities. Although soil itself has received negligible political attention in Denmark, many scientific studies have addressed a range of soil quality aspects (Schjønning et al., 2009) where substantial work has been done to understand and manage Danish soils. Moreover, the country has had soil type-dependent restrictions on manure and fertilizer application from the early 1970s to protect soils and the environment.

One of the major outputs of previous work in Denmark is the national Danish soil map: The Danish Soil Classification, which is a choropleth map at 1:50,000 scale (Madsen et al., 1992) classifying the surface soil (0–20 cm) into eight textural classes. However, to cope with the present data-demanding modeling techniques for understanding the complex soil system and soil-landscape processes in Denmark (e.g., the Danish simulation model DAISY [Hansen et al., 1990] requires depth-wise information on soil particle size distribution), rigidly defined soil classes are not optimal. So, to overcome this problem, Greve et al. (2007) produced the first soil surface texture maps of Denmark at a spatial resolution of 250 m using an ordinary kriging approach. Although all national soil maps published in Denmark so far describe soil properties in the topsoil, there is a need, both in terms of geographical and feature space, for detailed and more accurate continuous texture information by depth; for example, detailed texture data is needed for mapping soils sensitive to the leaching of pesticides. Such a need is even more pronounced when about two thirds of the country is under continuous, intensive, mechanized cultivation resulting in serious threats to the Danish soils such as soil compaction (Schjønning et al., 2009). Moreover, to support the global initiative of the GlobalSoilMap project (Sanchez et al., 2009; Hartemink et al., 2010), which aims to map all soils in the world at a fine resolution (100-m grid size) and make the digital maps readily available for landusers to assist decision-making in a range of global issues like food production and hunger eradication, climate change and environmental degradation, mapping soil properties (including soil texture) in Denmark according to the project specifications is also necessary.

Soil texture has conventionally been mapped with polygons where each polygon illustrates a texture class. However, due to the presence of intra-polygon texture variability, there can be a large degree of uncertainty in the textural composition within the area labeled by a polygon (Heuvelink and Huisman, 2000). Thus, an alternative way to cope with this problem is to map different textural fractions numerically (van Meirvenne and van Cleemput, 2006) so that intra-polygon textural variability can be better assessed. Moreover, most soil texture maps produced today are continuous surface maps in two dimensions completely ignoring the fact that texture also varies with depth. Soil properties in soil profiles vary continuously with depth (e.g., Russell and Moore, 1968), and this variability could be modeled using different functions ranging from simple freehand curves connecting the midpoint attribute values of the horizons (Jenny, 1941) to advanced methods like exponential decay functions (Brewer, 1968; Moore et al., 1972; Minasny et al., 2006), linear regression and polynomials (Campbell et al., 1970), or equal-area splines of Ponce-Hernandez et al. (1986). However, Bishop et al. (1999) suggested a higher efficiency of the equal-area quadratic splines over other methods due to their mass preserving and polynomial nature of fitting while modeling the depth function of soil attribute, and it has been used in many soil mapping studies (Malone et al., 2009; Malone et al., 2011; Adhikari et al., 2012; Odgers et al., 2012).

Soil depth functions derived from the splines are point predictions and need to be upscaled to continuous soil attribute or soil class maps at different spatial scales according to the needs of end users, such as for food production, soil-water, gas emissions, and climate change studies. Such continuous maps are generated by a digital soil mapping (DSM) technique, a soil mapping process based on statistical models describing the relationship between soil properties, legacy soil information, and environmental variables influencing soil formation and distribution. A comprehensive overview of DSM and the environmental variables ('*scorpan*' factors) are well documented by McBratney et al. (2003). A more detailed explanation of DSM, its state-of-the-art and application in different environmental settings has also been compiled by Lagacherie et al. (2007) and Hartemink et al. (2008). The term '*scorpan*' is a mnemonic for soil-forming factors such as soil, climate, organisms, parent materials, age, and spatial position that are used for empirical prediction of soil properties over space. It is an extension of the famous soil-forming factors of Jenny (1941), *clorpt*, which was meant for quantitative modeling of soil genesis.

The relationship between '*scorpan*' factors and soil properties is a fundamental element of DSM. Several studies have shown such relationship (Odeh et al., 1994; McKenzie and Ryan, 1999; Grunwald, 2006; Minasny et al., 2008; Dobos and Hengl, 2009; Bou Kheir et al., 2010; Greve et al., 2012a; Minasny et al., 2013) where various statistical approaches were used to quantify it to spatially predict soil properties (Minasny et al., 2013). According to Grunwald et al. (2011), a more accurate relationship between the environmental variables and soil properties can be quantified using dense data where the density and scale of soil observations resemble more closely the spatial resolution of the environmental variables used. But, in practice, finding a precise relationship is always hindered by a number of logistic constraints such as scale mismatch between the variables used or the different support for soil measurements (McKenzie and Ryan, 1999). Nevertheless, the strength of such relationship determines the success of soil and environmental correlation which would be an additional benefit where the environmental variables are readily measurable and available throughout the study area (McKenzie and Ryan, 1999).

A wide range of '*scorpan*' factors could be integrated into the DSM to build prediction models. Although different modeling techniques are available, recent years have seen an increasing use of statistical methods which include decision trees (Venables and Ripley, 1994) and regression rule induction algorithms which can be implemented using the Cubist data mining tool (Quinlan, 1993) are gaining popularity as they do not require any data reduction steps and are able to explore complex, nonlinear soil-covariate relationships (Minasny and McBratney, 2008; Henderson et al., 2005; Tranter et al., 2007; Bui et al., 2006).

The main objectives of our study therefore are: (i) to model the depth function of soil texture in the Danish soil profiles, and (ii) to map the lateral and vertical distribution of soil texture at a spatial resolution of ≈30 m in Denmark. We expect that our outputs would improve and update the current Danish Soil Information System with new fine-resolution soil property maps that could be useful to end users and stakeholders.

MATERIALS AND METHODS

Study Area

Our study area is Denmark (Fig. 1), a northern European country located in Scandinavia with a land area of roughly 43,000 km^2. It lies between 54°33′35″N to 57°45′7″ N and 8°4′22″ E to 15°11′55″ E (excluding Greenland and Faroe Islands which are also part of the Kingdom of Denmark). Denmark has a temperate climate with winter and summer mean temperatures of 0°C and 16°C, respectively (Danmarks Meteorologiske Institut, 1998). Average annual precipitation varies from 500 mm in the east to 800 mm in the west of the country. During late autumn, winter, and early spring, precipitation normally exceeds evapotranspiration, leading to leaching through the soil. Due to multiple glaciations during the Weichselian geological stage, and the effect of late and postmarine transgressions, rather complex glacial land systems have developed, which have greatly influenced the formation and distribution of soils in Denmark (Schou, 1949). Soil types vary from coarse sand in the west to loamy soil in the eastern part of the country, while the northern parts consist of marine sediments mixed with fine sandy materials on post and late-glacial marine elements. The central and the eastern parts of the country consist of moraine landscapes from the last glaciation, while older and strongly-eroded landscapes are common in the west. The major crops grown across Denmark include wheat, maize, potato, and barley. Topographically, Denmark is a relatively flat country with a mean elevation of 32 m with the highest point about 172 m above msl.

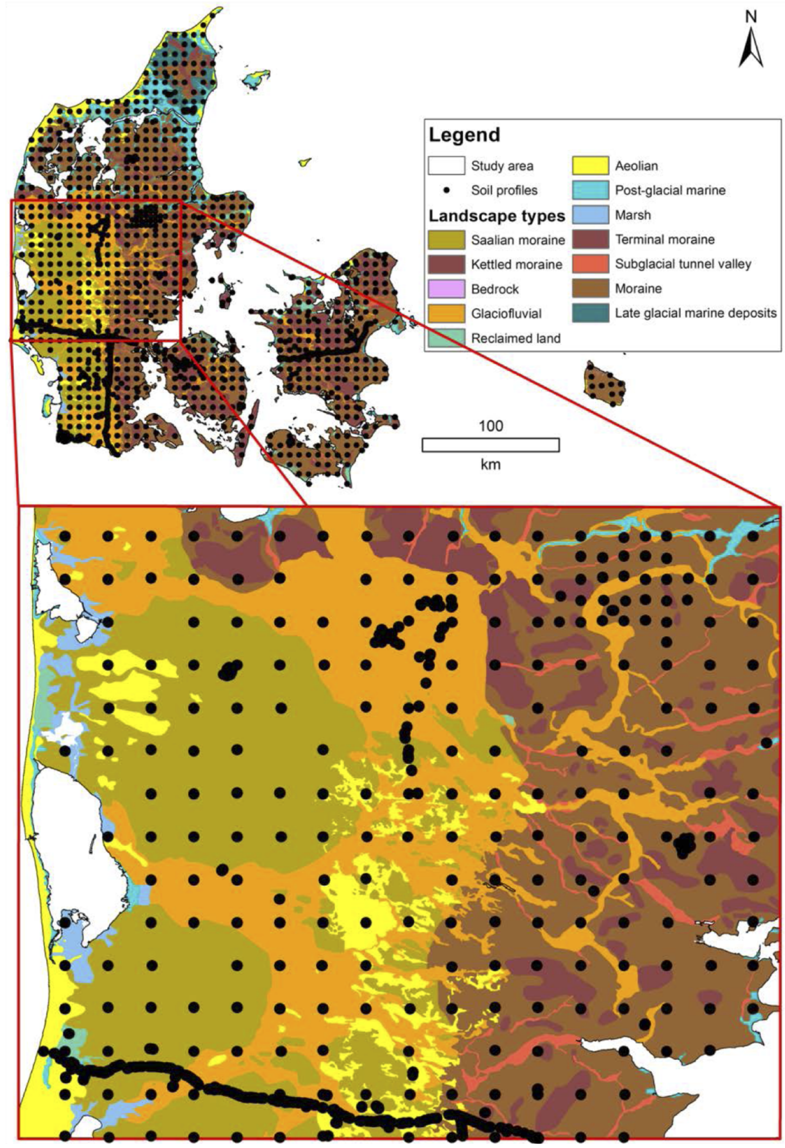

Fig. 1. Soil profile locations plotted on the landscape map of the study area clearly showing profiles along the gas pipeline (bottom), profiles in the regular grid, and specific research profiles as clusters.

Danish Soil Profiles and Texture Data

For our study, we used soil data from the Danish Soil Profile Database Service (Madsen et al., 1992), which has been a main source of soil profile data at national level in Denmark. It consists of soil texture data (and other properties) from soil profile horizons collected during the 1980s in connection with the following three activities: (i) national soil monitoring grids (7 km), (ii) natural gas pipeline trenches, and (iii) specific purpose profiles. Approximately 850 soil profiles from 7-km national grids, of which 663 were located on farmland, 106 in forests, and the rest on other landuses were investigated from 1987 to 1989 in the national nitrate study by the Danish Agricultural Advisory Service (Østergaard, 1990). A further 900 profiles were examined in conjunction with the installation of a natural gas pipeline across Denmark by Dansk Olie og Naturgas in the period 1981–1985 (Madsen and Jensen, 1985). Although soils along the pipelines were classified for each 25 m, approximately two to three profiles per kilometer were described in detail according to the genetic horizons (Madsen, 1983). Other studies have similarly investigated profiles in different parts of Denmark, which were also included in our study. All together, texture data from 1958 Danish soil profiles were gathered and used in our study. Most profiles considered in this study were extracted by digging a pit to 170 to 180 cm depth, followed by a further auger sample taken from a depth of 200 cm or 200 to 250 cm.

During soil sampling, samples were collected separately from each pedological horizon in each profile and taken to the laboratory for analysis. Samples were air-dried at room temperature before passing the samples through a 2-mm sieve, and the fine-earth fraction was analyzed for texture components in the lab using wet sieving and hydrometer methods. Four texture components or size fractions, namely clay (<0.002 mm), silt (0.002–0.02 mm), fine sand (0.02–0.2 mm), and coarse sand (0.2–2 mm) were obtained and were expressed in percentage mass (g 100 g^{-1}). These values were multiplied by 10 to convert the unit into g kg^{-1} as required in GlobalSoilMap project specifications.

Collection of Environmental Variables

Seventeen environmental variables were collected with full coverage of Denmark and used as predictors in mapping the texture components. These variables included a digital elevation model (DEM) based on airborne lidar data, a number of land surface parameters (LSP) derived from the DEM, and existing digital soil and geological data on a national extent in Denmark.

The DEM used in this study was first generated by the National Survey and Cadastre (2011) of the Danish Ministry of Environment using airborne lidar technology. The DEM was constructed as a raster, based on an interpolation of the point clouds using Delaunay Triangulation methods, and the resulting triangular irregular network (TIN) model was then converted to a raster of 1.6-m resolution. This high-resolution raster was subsequently coarsened by simple aggregation to 30.4-m resolution (1.6 m × 19 m = 30.4 m) to be used in our study. Although Greve et al. (2012b) found a better prediction performance at 24 m resolution when predicting soil clay content in Denmark using regression tree analysis, 30.4-m grid resolution was preferred in this study due to a longer processing time required with 24 m. The DEM was then hydrologically corrected by identifying and filling the sinks of ≤50-cm depths. This correction was crucial, especially when working with flow-related calculations because the sinks and peaks present on the surface, if not removed, might act as a local barrier for water flow and create a problem during drainage network extraction. The creation and preprocessing of the DEM was performed using the TerraStream algorithms (Danner et al., 2007). Multiple-flow direction (MFD) or FD8 algorithms (Freeman, 1991) were adopted to calculate all flow-related parameters, as this method considers the effect of divergent flow, which is important when modeling flow on slope gradients and gives much more realistic distributions of contributing area (Hengl and Reuters, 2009). After the correction, the following 12 LSP were extracted from the DEM using ArcGIS (ESRI, Redlands, WA) and SAGA GIS (SAGA User Group Association, Hamburg, Germany) programs: slope aspect, topographic wetness index (TWI), direct sunlight insolation, elevation, flow accumulation, mid-slope position, multi-resolution index of valley bottom flatness (MRVBF), SAGA wetness index, slope gradient, slope-length factor, vertical distance to channel network, and valley depth.

Another group of environmental data used included existing digital maps of landscape types, soil types, geology, landuse, and geo-regions on a national scale in Denmark. A map of landscape types, which represents the Danish landforms mostly referring to quaternary geological developments, was derived from the existing digital vector map at a scale of 1:100,000 (Madsen et al., 1992). The geological map represents the extent and types of parent materials from which the soils have been developed and was extracted from scanned and registered national geological maps of Denmark at a 1:25,000 cartographic scale (Danmarks Geologiske Undersøgelse, 1978). Similarly, a soil type map which was derived from the Danish Soil Classification represents different soil classes ranging from coarse sand to heavy clay and organic soils in Denmark. This map was generated using 36,000 topsoil point texture data (0–20 cm) from all over the country. The landuse map corresponds to land cover types derived from Corine Land Cover 2000 data for Denmark (Stjernholm and Kjeldgaard, 2004) showing cropped areas, forests, grasslands, urban areas, and water bodies. The map of geo-regions roughly divides Denmark into 10 distinct regions based on climatic and geographical settings (personal communication with M.H. Greve, 2012). All these data layers were brought to a common projection of ETRS1989 UTM32N with a cell size of 30.4 m for further analysis. Table 1 summarizes the covariate data used in this study with brief specifications.

Modeling Soil Texture Depth Function in Danish Soil Profiles

Equal-area quadratic splines were used to model the continuous depth function of soil texture distribution in the profiles. An equal-area quadratic spline consists of a series of local quadratic polynomials that join at 'knots' located at the soil horizon boundaries (Bishop et al., 1999). It passes through each soil horizon, maintaining the average value of the soil attributes, and is linear between and quadratic within the horizons giving a linear-quadratic smoothing spline. The areas above and below the fitted spline in any horizon are equal—hence the name equal-area (Ponce-Hernandez et al., 1986). In mathematical terms, the distribution of soil properties in a profile with the given depth of (x) is a continuous function of (x) which is unknown and must be estimated using horizon data of that profile. The spline function that models $f(x)$ consists of choosing the $f(x)$ that minimizes Eq. [1]:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\overline{f}_i\right)^2+\lambda\int_{x_0}^{x_n}\left[f'(x)\right]^2dx \qquad [1]$$

Table 1. Environmental covariates used to predict soil texture components.†

Variables	Data source/Brief description	Scale/Resolution	Type	Mean (Range)
Slope aspect	DEM/Direction of the steepest slope gradient from the North	30.4 m	Q	181.2 (0–360)
TWI	DEM/Calculates slope gradient and specific catchment area based Topographic wetness index. TWI = *ln (As/tan β)*: where *As* is upslope catchment area and β is the slope gradient (Moore et al., 1993)	30.4 m	Q	9.65 (4.0–32)
Direct Sunlight Insolation (1 yr)	DEM/Calculates potential incoming solar radiation (insolation) calculated in SAGA GIS (Boehner and Antonic, 2009)	30.4 m	Q	1269 (122–1707)
Elevation	DEM/lidar produced elevation of the land surface	30.4 m	Q	31.21 (0–172)
Flow Accumulation	DEM/Number of upslope cells	30.4 m	Q	61 (1–110908)
Mid-Slope Position	DEM/covers the warmer zones of slopes (Bendix, 2004)	30.4 m	Q	0.001 (0–1)
MRVBF	DEM/Identifies the depositional areas (Gallant and Dowling, 2003)	30.4 m	Q	4.26 (2.22–10.90)
SAGA Wetness Index	DEM/Same as TWI but it uses modified catchment area instead (Boehner et al., 2002)	30.4 m	Q	14.46 (6.9–19.1)
Slope gradient	DEM/maximum rate of change between cells and neighbors.	30.4 m	Q	1.63 (0–90)
Slope-length factor	DEM (calculates the slope-length as used by Universal Soil Loss Equation (Desmet and Govers, 1996)	30.4 m	Q	0.61 (0–54.50)
Vertical distance to channel network	DEM/calculates Vertical distance to channel network for each cell	30.4 m	Q	559.45 (0–10041.50)
Valley depth	DEM/relative position of the valley	30.4 m	Q	7.53 (0–90)
Geology	Scanned and registered geological map	1:25,000	C	37 classes
Geo-regions	Scanned geographical regions map	1:100,000	C	10 classes
Landscape types	Digital landscape elements	1:100,000	C	12 classes
Landuse	Corine land cover map	1:100, 000	C	31 classes
Soil types	Digital soil types' map	1:50,000	C	11 classes

† DEM, digital elevation model; TWI, topographic wetness index; lidar, light detection and ranging; MRVBF, multi-resolution index of valley bottom flatness; SAGA, system for automated geoscientific analyses; Q, quantitative data; C, categorical data

where y_i is the modeled estimate of measured texture data (y) from layer i, $\overline{f_i}$ is the mean from n soil layers and the depth of the boundaries of n layers as $x_0 < x_1, \ldots\ldots x_n$. The first term of Eq. [1] represents a fit to the data, while the second term measures the roughness of function $f(x)$ represented by its first derivative $f'(x)$, which is also a continuous and integrable. The parameter λ controls the trade-off between fidelity and roughness penalty and is also called a spline-smoothing parameter. A detailed explanation of spline model, its derivation, and use in modeling soil depth functions is given by Malone et al. (2009).

As the spline function basically depends on parameter λ, we tested seven λ values (0.00001, 0.0001, 0.001, 0.01, 0.1, 1, and 10) for all profiles, initially by fixing the maximum depth of 200 cm from the surface, and the λ with the lowest root mean square error (RMSE) or highest prediction capability was selected as the best λ to use for all the soil profiles. The spline function produced continuous soil texture values from the surface to a 200-cm depth. The continuous values of the spline were averaged within each six depth increment of 0 to 5, 5 to 15, 15 to 30, 30 to 60, 60 to 100, and 100 to 200 cm which correspond to the depths specified in the GlobalSoilMap project. Odgers et al. (2012) have applied a similar spline function to the soil components of a vector-based map to estimate soil organic carbon content for the whole United States at these six depths increments. These new texture values for the six standard depths acted as point data at different depths, which would be used for regression modeling. Specific to the case of Danish agricultural fields, where we assume a presence of a homogeneous plow layer (Ap horizon) in terms of texture distribution due to regular agricultural activities, we restricted our splines such that the predicted values for the plow depth are the same throughout. This was accomplished by introducing to the first horizon two 1-cm deep layers—one at the top and one at the bottom of that horizon—that forced the spline not to vary in the first horizon.

Regression Modeling and Spatial Prediction

To characterize the relationship between soil textural properties and '*scorpan*' factors, a regression analysis was performed using the Cubist 2.07 program (www.rulequest.com). Regression using Cubist works with condition-based rules where the output is a set of rules, and each rule has a specific multivariate linear model. Whenever a situation matches a rule's condition, the associated model is used to calculate the predicted value (Minasny and McBratney, 2008). Unlike regression trees, which predict a rigid value for each 'leaf,' regression rules build a multivariate linear function.

Before the regression analysis, the corresponding values of all the '*scorpan*' factors on soil profile locations were extracted and intersected with the spline-predicted texture values for the six standard depths. The data were randomly divided into two sets—80% for model calibration and 20% for validation. Once the model was built, the regression rules were applied to the covariate data compiled for all of Denmark. A program written in FORTRAN was used to generate raster grids from the prediction model. These grids represented texture distribution as a function of the covariates used. Moreover, the residual of prediction—which is the difference between measured and predicted texture values at the training profiles—was calculated, and its distribution over the study area was predicted using point kriging with local variograms using the Vesper program (Minasny et al., 2005). The kriged map of residuals was added to the corresponding regression output grid. This approach is analogous to regression kriging, which is a combination of a regression model with kriging of the regression residuals (Odeh et al., 1995; Hengl et al., 2007). To facilitate easy comparison of the predicted maps, we used a single color palette in all maps with an equal number of classes, the range of which was based on the existing Danish Soil Classification.

Similarly, uncertainty of prediction of soil texture fractions at each depth was assessed with prediction standard error maps. The general spatial texture distribution pattern of predicted maps was also compared with the existing maps of Greve et al. (2007) and Greve et al. (2012b).

The influence of different environmental variables in predicting texture fractions was also checked. The utility of each variable during the model building was derived from the model and expressed as relative importance (RI) of that variable in percentage. As RI shows the percentage usage of the variables in all the prediction rules, the total may not sum to 100% for a given model.

Assessment of Model Performance

Although there are better ways for validation of digital soil maps (e.g., Brus et al., 2011), the accuracy of model prediction in our study was evaluated on the 20% of profiles which were kept aside during regression analysis for the specific purposes of an external validation. The following prediction quality criteria were calculated:

1. Coefficient of determination (R^2). It shows the effectiveness of one variable in predicting another variable.

2. Mean error (ME) and Root mean square error (RMSE). The ME evaluates the tendency of the prediction model to make under or overprediction, whereas the RMSE measures the accuracy of prediction on a point by point basis. The difference between measured and predicted values for each validation point was calculated, and the mean was derived as ME. For RMSE, the differences were squared before summing them all, and the total of the squared difference is divided by the number of validation data and that output was square rooted. For an ideal model, ME should be closer to zero and RMSE as small as possible.

$$R^2 = \frac{\sum_{i=1}^{n}\left[z^*(x_i) - z(\bar{x})\right]^2}{\sum_{i=1}^{n}\left[z(x_i) - z(\bar{x})\right]^2} \quad [2]$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left[z(x_i) - z^*(x_i)\right] \quad [3]$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[z(x_i) - z^*(x_i)\right]^2} \quad [4]$$

where data points are denoted by (n), measured and predicted attribute values at i^{th} location as $z(x_i)$ and $z^*(x_i)$ with their corresponding means of $z(\bar{x})$ and $z^*(\bar{x})$, respectively.

Variogram Analysis

Spatial autocorrelation of spline-predicted texture components was also investigated using a variogram which describes the relationship between the lag or any integral multiple of the sampling interval and the semivariance (Goovaerts, 1997). An empirical variogram [$\gamma(h)$] can be calculated as

$$\gamma(h) = \frac{1}{2N(h)}\sum_{\alpha=1}^{N(h)}\left[z(x_\alpha + h) - z(x_\alpha)\right]^2 \quad [5]$$

where $N(h)$ denotes the point pairs separated by distance h, $z(x_\alpha)$, and $z(x_\alpha + h)$ are the measured values of location x and h distance away.

With the texture data from six standard depths, omnidirectional experimental variograms were calculated, and different theoretical variogram models were fitted using nonlinear least squares approximation. The selection of the best variogram model was based on Akaike Information Criteria (AIC) with its smallest value for the best fit (Akaike, 1973). Several soil variability studies in the past (e.g., McBratney and Webster, 1986; Yost et al., 1982) have successfully used these criteria for variogram model selection. After the fit, corresponding variogram parameters—i.e., nugget (C_0), partial sill (C_1), and a range of spatial dependence (A_1)—were extracted to examine the spatial variability pattern. Moreover, to check the contribution of nugget to the overall variance, relative nugget effect (RNE) was calculated according to Brooker (1986). Models showing a good spatial structure should have a relatively lower RNE.

$$RNE = \frac{C_0}{C_0 + C_1} \quad [6]$$

RESULTS

Continuous Depth Function of Texture Distribution in Danish Soil Profiles

Equal-area quadratic splines predicted the continuous function of soil texture distribution in the profiles where the highest prediction capability of the fitted splines was associated with the λ values of 0.1 for all four texture fractions. The spline-predicted texture content for the first three depths, i.e., 0 to 5, 5 to 15, and 15 to 30 cm, for the profiles from the agriculture fields remained more or less the same due to a 'slicing' pretreatment applied to the splines to take account of the homogeneity of the plow layer but were different for the profiles from nonagricultural fields, for example, from the forest. The mean clay content at 0 to 5 cm was 85 g kg^{-1}, which was similar to silt content at the same depth but increased to 104 g kg^{-1} at 30 to 60 cm where silt was only 77 g kg^{-1}. Mean fine sand and coarse sand contents were always higher than clay and silt content for all depths. Descriptive statistics of spline-predicted texture components from all the profiles in the study area at six depths are summarized in Table 2.

Clay showed the highest variability of all texture components for the top three depths (0–5, 5–15, and 15–30 cm) with corresponding coefficients of variation (*CVs*) of 85, 82, and 82%, followed by silt, coarse sand, and fine sand, whereas silt had the highest variability at the bottom three depths (30–60, 60–100, and 100–200 cm) with *CVs* of 85, 96, and 97%.

Figure 2 shows an example of a fitted spline to the measured clay, silt, fine sand, and coarse sand contents from a representative soil profile (Profile No. 1622, located in the moraine landscape) of the Danish soil profile database for which the average values of the spline-predicted texture fractions at six depths are given in Table 3.

For this profile, the average clay content for 0 to 5 cm was 52 g kg^{-1} and for 100 to 200 cm it was 23 g kg^{-1}. We observed in this profile a decrease in clay content from top to bottom, except at 60 to 100 cm depth where it suddenly increased to 73 g kg^{-1}. Silt followed a similar distribution to clay but with lower values for all depths. Silt content on the surface was 44 g kg^{-1} and at the bottom 5 g kg^{-1}. Similarly, the highest fine sand content was at the 60- to 100-cm depth and the lowest at the 100- to 200-cm depth. The distribution of the coarse sand fraction for all depths in the top 60 cm was almost similar at around 598 g kg^{-1}, and the lowest value of 524 g kg^{-1} was recorded at 60 to 100 cm. The highest coarse sand content of 813 g kg^{-1} was found below 100 cm in this profile.

Regression Rules for the Prediction of Soil Texture Components

During the regression analysis, the '*scorpan*' factors for soil texture prediction were selected automatically by the program based on their relative importance in the prediction. In most of the rules, soil types, geology, landscape types, and landuse were some of the covariates selected to set the rule conditions, and the LSP like TWI, SAGA wetness index, slope gradient, distance to channel network, elevation, MRVBF, slope-length factor, etc., were extensively used to build the linear models of prediction. For instance, Table 4 presents a list of the top covariates used in rule setting (Top 3) and building linear regression models (Top 5) with their relative importance in predicting soil silt content at six soil depths. For the 0- to 5-cm depth, 31% of the models used landuse for rule setting, while at the remaining depths little or even no contribution of landuse was observed, for example, at depths below 60 cm. The influence of geology on rule setting

Table 2. Descriptive statistics of spline predicted texture components from six depths over the study area (Number of observations = 1958).†

Texture fractions	Parameters	Depth, cm 0–5	5–15	15–30	30–60	60–100	100–200
Clay	Mean	85	86	92	104	111	109
	SD	72	70	76	89	96	96
	CV	85	82	82	86	86	88
	Skewness	1.89	1.83	1.70	1.37	0.98	1.11
	Kurtosis	6.30	6.08	4.72	2.59	0.93	1.86
Silt	Mean	86	87	84	77	74	74
	SD	65	65	64	65	70	72
	CV	76	74	76	85	96	97
	Skewness	0.93	0.93	1.04	1.12	1.24	1.10
	Kurtosis	1.65	1.66	1.68	1.96	1.99	1.24
Fine sand	Mean	343	346	347	344	337	332
	SD	173	167	163	147	186	199
	CV	51	48	47	51	55	60
	Skewness	−0.04	−0.42	0.24	0.13	0.22	0.34
	Kurtosis	−0.002	0.01	0.14	0.24	0.13	0.09
Coarse sand	Mean	369	375	384	400	404	395
	SD	230	226	256	248	270	280
	CV	62	60	59	62	67	71
	Skewness	0.09	0.09	0.17	0.27	0.38	0.45
	Kurtosis	−0.77	−0.72	−0.73	−0.84	−0.95	−0.95

† SD, standard deviation; CV, coefficient of variation.

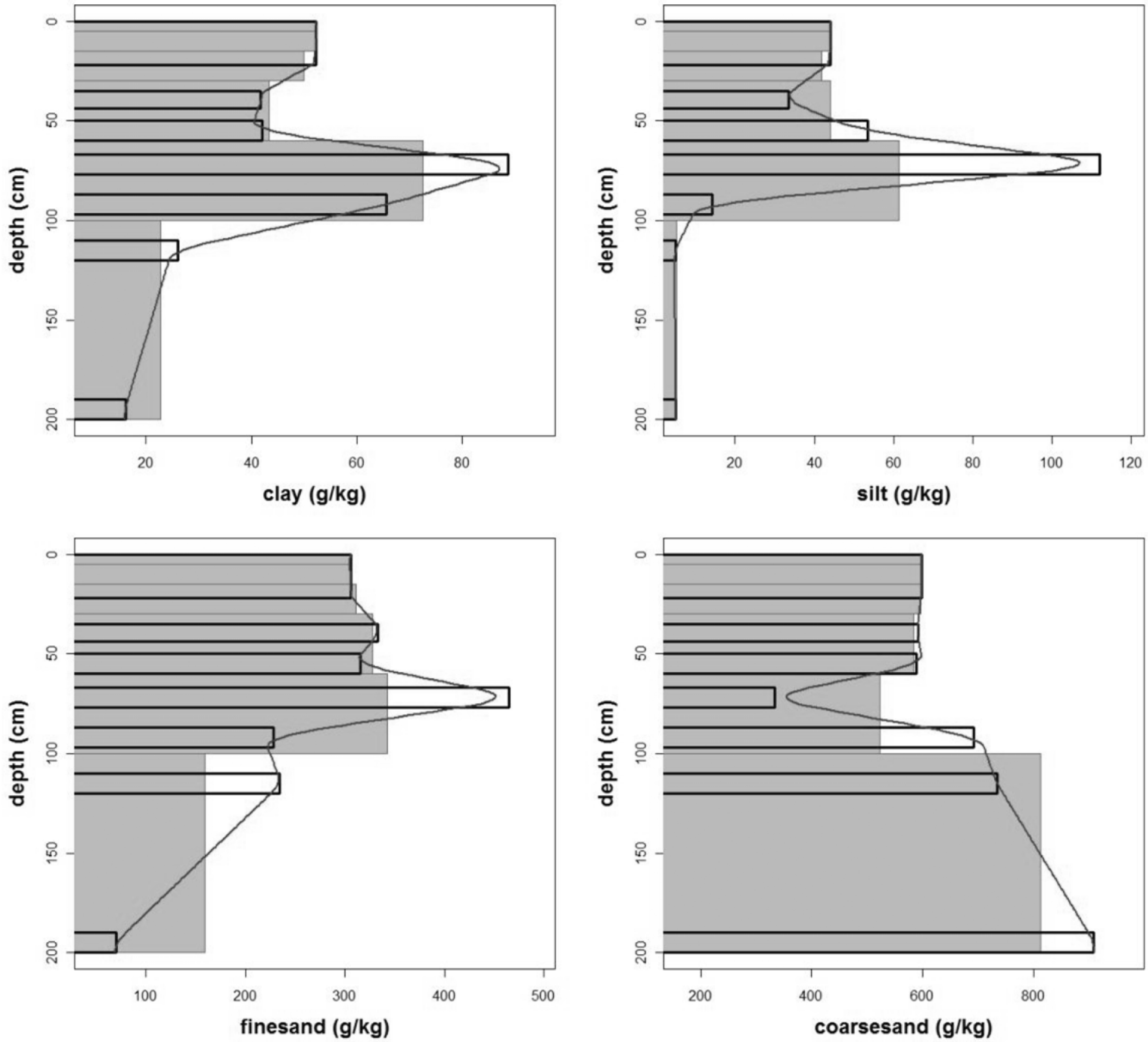

Fig. 2. Equal-area splines fitted to the soil texture data from Profile No. 1622. Dark horizontal bars are measured texture data at different depths, continuous vertical lines represent fitted splines ($\lambda = 0.1$), and gray horizontal bars represent averaged texture values at six global standard depths (i.e., 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm). Note that the *x* axes of these graphs have different scales.

appeared prominent for subsoil and increased with depth. Below 30 cm, geology contributed more than 50% of the rule setting. For the prediction of the rest of the fractions, a similar pattern of variable combination was also observed.

While predicting clay in the surface 0 to 30 cm, predictors like slope gradient, TWI, SAGA wetness index, and vertical distance to channel network scored >80% in importance, whereas geology, landscape, and vertical distance to channel network were more prevalent (>65%) at lower depths. Similarly, slope-length factor, elevation, and geology appeared to be more important for

Table 3. Average values of the spline predicted texture components in g/kg at six depths for Profile No. 1622 from the Danish soil profile database.

Texture components	Depths, cm					
	0–5	5–15	15–30	30–60	60–100	100–200
Clay	52	52	50	43	73	23
Silt	44	44	42	44	61	5
Fine sand	306	306	312	328	342	159
Coarse sand	598	598	597	585	524	813

modeling the distribution of both fine and coarse sand fractions at different soil depths. In almost all the rules, soil types, geology, and landscape were the three most important predictors for texture prediction.

In this way, texture components at different soil depths were predicted by a large number of regression rules produced during regression modeling, but in this paper we only described one of the 10 rules used to predict silt content at 0- to 5-cm depth as an example.

Rule 1: [169 cases, mean 25.93, range 0 to 163.27, estimated error 18.82]

If

landscape in {1, 2, 4, 8, 9, 10}
landuse in {2, 4, 7, 11, 17, 18, 23, 24}

then

silt from 0–5 (g kg^{-1}) = 75.28 – 1.60multi-resolution index of valley bottom flatness – 2.80SAGA wetness index – 3.10slope-length factor – 0.09elevation + 0.001 vertical distance to channel network – 0.50 slope gradient

Table 4. Top predictors selected by the cubist model and their relative importance in percentage while predicting soil silt content at six soil depths.†

Depths (cm)	Predictors	
	For rule setting (Top 3)	For model building (Top 5)
0–5	Soil types 90%, Landscape types 54%, Landuse 31%	Slope-length factor 100%, MRVBF 92%, Elevation 86%, SAGA Wetness Index 81%, Vertical distance to channel network 80%
5–15	Soil types 85%, Landscape types 56%, MRVBF 30%	Slope-length factor 100%, SAGA Wetness Index 81%, Vertical distance to channel network 79%, Elevation 73%, MRVBF 69%
15–30	Soil types 100%, Landscape types 26%, Geology 23%	MRVBF 90%, Slope gradient 84%, Valley depth 75%, SAGA Wetness Index 73%, Slope-length factor, Mild-slope position and Vertical distance to channel network 49%
30–60	Soil types 80%, Geology 51%, Landscape tyoes 39%	SAGA Wetness Index 100%, Slope-length factor 88%, Vertical distance to channel network 69%, MRVBF 61%, Elevation 42%
60–100	Soil types 82%, Vertical distance to channel network, Geology 50%, Geo-regions 31%	Slope-length factor, SAGA Wetness index and Slope gradient 91%, MRVBF 74%, Vertical distance to channel network 43%, Elevation 41%, Slope aspect 39%
100–200	Soil types 81%, Geo-regions 55%, Geology 51%	SAGA Wetness Index 85%, Slope-length factor 80%, Slope gradient 76%, Valley depth 44%, Elevation, MRVBF and Vertical distance to channel network 40%

† MRVBF, multi-resolution index of valley bottom flatness; SAGA; system for automated geoscientific analyses.

This rule was used to predict soil silt content at 0 to 5 cm depth in 169 out of 1566 training profiles. The model first set a condition and identified mostly sand rich and nonagricultural areas specified in aeolian, post-glacial and late glacial marine deposits, tunnel valleys, glacial flood plains, and salian moraine (as Classes 1, 2, 4, 8, 9, and 10) landscape types where landuse was found to be heath, mixed forest and shrub, broadleaved forest, conifers, natural grasslands, beach and dunes, marshy areas, and pasture in rotation (as Classes 2, 4, 7, 11, 17, 18, 23, and 24) and then used the corresponding linear model to predict soil silt content for the 0- to 5-cm depth where the covariates MRVBF, SAGA wetness index, slope-length factor, elevation, distance to channel network, and slope gradient were used. All the predictors, except distance to channel network, have negative coefficients, suggesting a decreasing silt content when increasing the values of the covariates. MRVBF, SAGA wetness index, and slope-length factor highly influenced the distribution of silt on the top 0 to 5 cm in those areas.

Model Validation

Table 5 summarizes different model assessment parameters with their corresponding values used to assess model predictive performances.

For all texture components, model predictions for the upper three depths were relatively better than the three lower depths with higher R^2 and lower RMSE values. Fine sand had the lowest R^2 values for almost all depths, whereas the highest R^2 were associated with silt for the upper three depths and with coarse sand content at the lower three depths. At 30 to 60 cm, R^2 values for clay, silt, fine sand, and coarse sand were 0.35, 0.42, 0.34 and 0.45, respectively. The highest RMSE values were associated with the coarse sand fraction, and silt content exhibited the lowest RMSE compared to other texture fractions at all depths.

When considering the upper three depths and lower three depths as two different sections (Section I representing the topsoil, while Section II is for the subsoil), model assessment parameters (e.g., R^2) within each section were similar and comparable but different between the sections. Clay and silt were always better predicted compared to fine and coarse sand contents in

Table 5. Evaluation of model performances using different model assessment criteria calculated from the validation data set and their corresponding values for four texture components at different soil depths.†

Texture fractions	Parameters	Depth, cm					
		0–5	5–15	15–30	30–60	60–100	100–200
Clay	R^2	0.54	0.50	0.55	0.35	0.39	0.26
	ME	2.2	4.2	1.8	−5.5	1.7	0.1
	RMSE	37.0	37.8	37.4	64.4	50.6	51.8
Silt	R^2	0.54	0.55	0.54	0.42	0.35	0.29
	ME	0.05	1.1	1.8	−0.3	−2.0	−1.3
	RMSE	33.7	33.2	33.4	36.5	41.8	38.8
Fine sand	R^2	0.40	0.44	0.43	0.34	0.37	0.27
	ME	3.2	2.8	−2.0	7.0	7.6	12.1
	RMSE	96.3	87.1	95.0	132.5	88.5	107.2
Coarse sand	R^2	0.50	0.51	0.48	0.45	0.42	0.37
	ME	13.7	8.0	6.8	20.7	12.1	27.8
	RMSE	122.5	121.5	130.1	148.6	160.1	163.7

† R^2, coefficient of determination; ME, mean error; RMSE, root mean square error.

Section I, but predictability of these components decreased rapidly for Section II. For Section II, sand fractions had higher R^2 values, except for the fine sand content at 100 to 200 cm, which had the lowest R^2 value of 0.27. In general, R^2 values declined with depth with increasing RMSE values. This might be due to a higher texture variability in the deeper layers and could be assessed by variogram analysis. However, it could also be attributed to the fact that most of the environmental variables shown in Table 1 are based on land surface characteristics (DEM derivatives, landscape types, land use), such that they have a greater association with surficial soil properties to some extent.

Variogram Analysis

The variogram parameters of texture components for six standard depths are given in Table 6. Of the different theoretical variogram models fitted to the omnidirectional experimental variogram, the lowest AIC values were associated with a spherical model so this was selected as the best model to characterize spatial autocorrelation of texture distribution. As we move from the surface into the soil, we observe an increasing semivariance (i.e., $C_0 + C_1$) with a decreasing range of spatial dependence for all texture components. This confirms the existence of a higher textural variability at lower depths. However, with increasing depth the contribution of nugget to the total variance also increased as seen with clay and silt. But for coarse and fine sand fractions, RNE first decreased up to 60 cm and subsequently started increasing. At depths lower than 60 cm, the RNE for clay and silt always remained higher than that for fine and coarse sand components. Similarly, for the first two depths, fine and coarse sand were associated with a higher RNE than clay and silt from the same depths.

Mapping Soil Texture Components

The predicted maps of clay, silt, fine sand, and coarse sand contents from different soil depths in Denmark are displayed in Fig. 3 to 6. Across Denmark, the average clay content of the top layer (0–5 cm) was 79 g kg^{-1} with a standard deviation of 49 g kg^{-1}, but at 60- to 100-cm depth the clay content increased to 107 g kg^{-1}. Silt content, which was higher than clay for the top 0 to 5 cm and 5 to 15 cm, fell from 84 g kg^{-1} to 78 g kg^{-1} from 0- to 5-cm to 60- to 100-cm depth. The lowest silt content was recorded at the 100- to 200-cm depth. Similarly, mean fine sand and coarse sand contents for the 5- to 15-cm depth were almost identical (at about 362 g kg^{-1}), but below that depth, coarse sand contents were always higher than fine sand. Coarse sand content was high (about 385 g kg^{-1}) for the bottom 100- to 200-cm layer. When looking at the spatial distribution of clay and silt across Denmark at the different soil depths, higher values were observed toward the center and in the eastern part, and lower values in the west and north for almost all depths. In contrast, most soils in the western region of the country were rich

Table 6. Variogram parameters and the best fitted model for the spline predicted texture components at different soil depths (Number of observations = 1958).†

Texture fractions	Parameters	Depth, cm 0–5	5–15	15–30	30–60	60–100	100–200
Clay	AIC	257	250	247	237	214	225
	Model	Spherical	Spherical	Spherical	Spherical	Spherical	Spherical
	C_0	2389	2399	3179	5623	7178	7220
	C_1	2510	2397	2385	1910	1707	1749
	A_1	74.40	72.80	65.30	66.72	59.62	35.00
	RNE	0.48	0.50	0.57	0.74	0.80	0.80
Silt	AIC	221	215	218	206	215	229
	Model	Spherical	Spherical	Spherical	Spherical	Spherical	Spherical
	C_0	2209	2102	2125	2354	2847	2749
	C_1	1939	1925	1892	1490	1310	1379
	A_1	97.13	96.77	93.78	90.21	80.4	75.71
	RNE	0.53	0.52	0.52	0.61	0.68	0.66
Fine sand	AIC	290	286	289.2	297	303	305
	Model	Spherical	Spherical	Spherical	Spherical	Spherical	Spherical
	C_0	16008	14366	12449	14072	18146	23449
	C_1	11650	11386	12702	14755	14460	13355
	A_1	73.62	72.35	68.17	67.90	72.25	76.90
	RNE	0.57	0.55	0.49	0.48	0.55	0.63
Coarse sand	AIC	282	278.2	278.1	285	294	289.7
	Model	Spherical	Spherical	Spherical	Spherical	Spherical	Spherical
	C_0	35127	32962	32160	36494	43890	49052
	C_1	16051	15705	16912	23055	26682	26536
	A_1	36.93	37.77	44.95	45.14	38.32	34.40
	RNE	0.68	0.67	0.65	0.61	0.62	0.64

† AIC, akaike information criteria; C_0, nugget (g/kg); C_1, partial sill (g/kg); A_1, range (km); RNE, relative nugget effect.

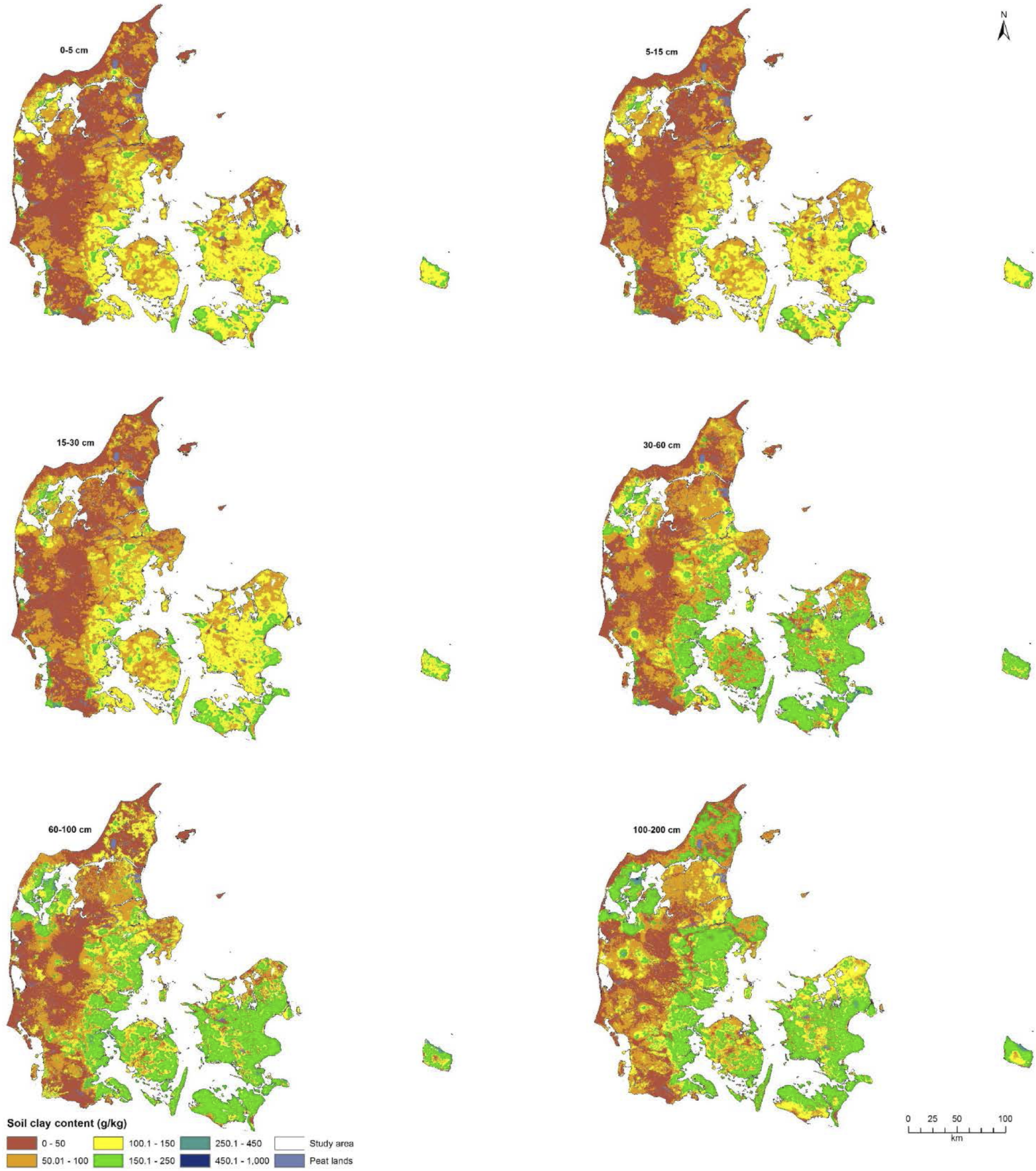

Fig. 3. Predicted clay content of Danish soils at six depths.

in coarse sand at all depths, and northern parts were rich in fine sand. Moreover, patches of coarse-textured soils developed on the inland aeolian deposits were also found in the west. Similarly, soils that developed on the late and postglacial marine materials as in the northern part of the country were relatively poor in clay content but rich in fine sands. Highly organic soils (peats) with little or no clay material were also prevalent in these areas.

When comparing predicted texture fractions at different soil depths, soils from 0 to 15 cm or 0 to 30 cm contained less clay than the underlying horizons for almost the entire area of Denmark. The highest clay content was found at a depth of 60 to 100 cm. A similar trend of clay distribution was noticed in a small area including an island in the north-western part of Denmark and on Bornhølm (the easternmost island of the country). Silt content, on the other hand, decreased with depth throughout the country. Although some patches in the eastern part had medium silt contents (150–250 g kg^{-1}), for example at depths of 30 to 60 cm or at 100 to 200 cm, average silt content for those layers was about 77 g kg^{-1}. The distribution of fine sand in the profiles was found to be relatively even for all depths, whereas coarse sand

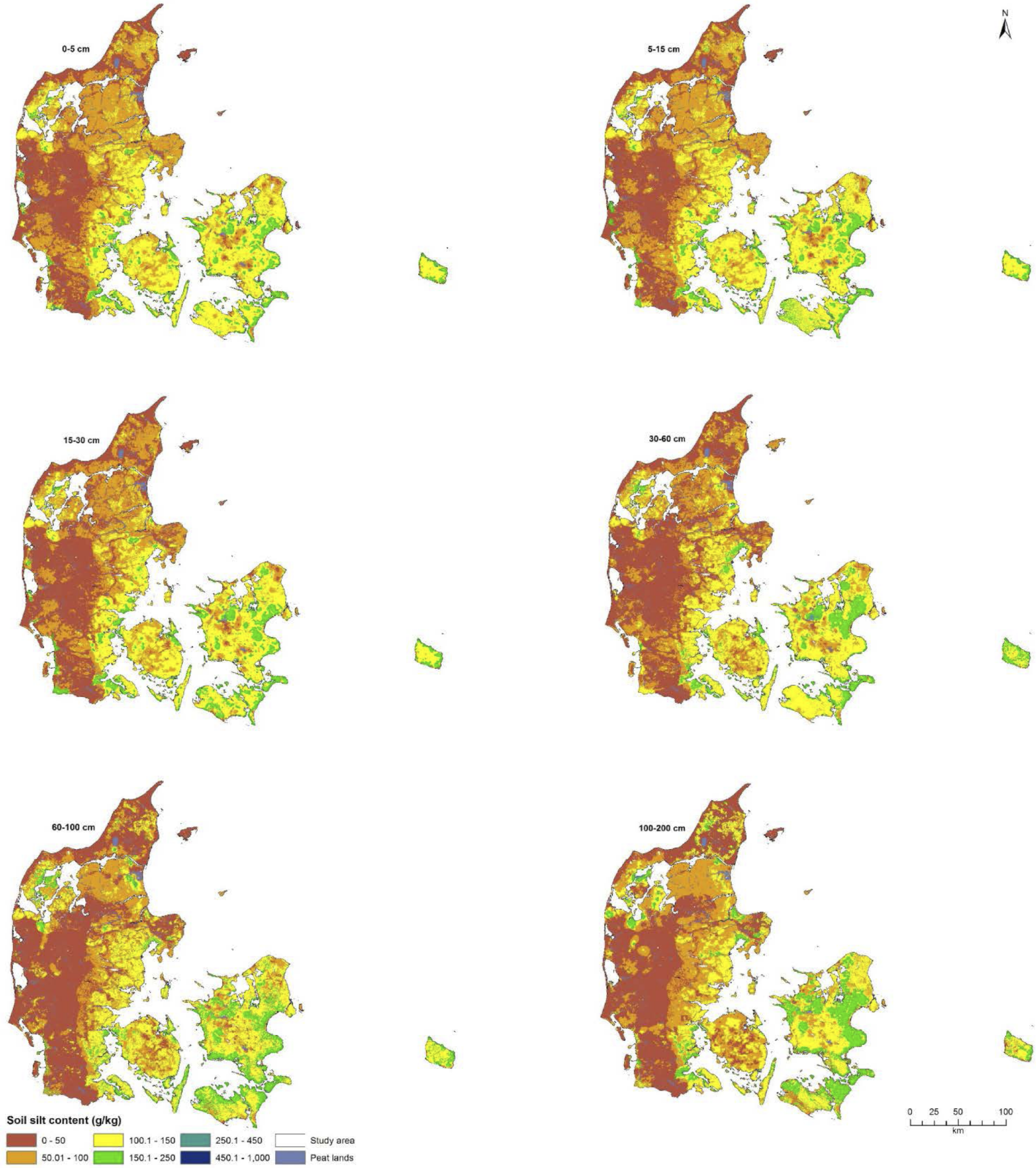

Fig. 4. Predicted silt content of Danish soils at six depths.

content increased with depth. However, for the top two depths, the fine sand content always exceeded the average value of the coarse sand content.

The map in Fig. 7 shows an example of the standard error of prediction for the clay content at 15 to 30 cm soil depth. Darker areas in the southwest corner of the study area were associated with a high level of error compared to the rest of the area.

We found a similar spatial distribution pattern (e.g., clay content) in general when compared to the existing maps of Greve et al. (2007) and Greve et al. (2012b), although the input data, measurement units, and prediction methods were different. Both authors used high density (>40,000 points) texture data (in g/100 g) from 0- to 30-cm depth: Greve et al. (2007) used stratified ordinary kriging, whereas Greve et al. (2012b) applied regression tree modeling for prediction.

DISCUSSION

Our study has utilized the available soil and environmental data from different sources to map the vertical distribution of soil texture components in Denmark at a national extent using

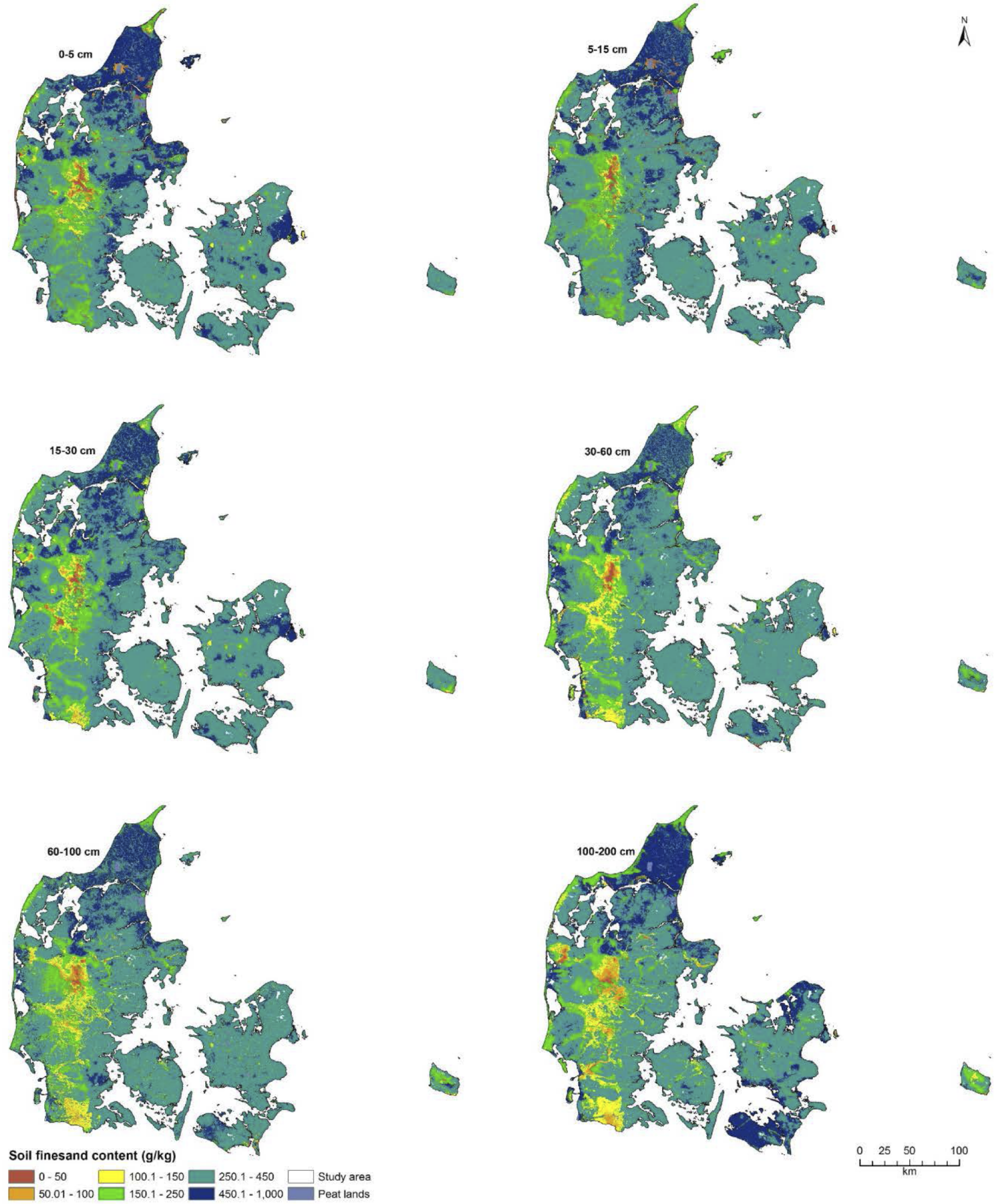

Fig. 5. Predicted fine sand content of Danish soils at six depths.

regression rules based DSM. Although the availability of high-density soil data is not always a primary requirement in DSM, good quality data generally reduces mapping uncertainty and the outputs turn out to be more reliable. As suggested by Minasny and McBratney (2007), improvement in the prediction of soil properties not only depends on the use of more sophisticated statistical methods but also on the use of good-quality data. In that sense, Denmark could be considered as one of the best countries for mapping soil properties through national-extent DSM. Based on the profile observations, Denmark has been found to be rich in sands where surface soils are mostly clay eluviated. To model the continuous depth function of soil texture fractions from Danish soil profiles, equal-area quadratic splines were found to be promising. Such a continuous texture data to 200-cm depth could be useful for studying and modeling several landscape properties and processes in the critical zone. Specifically for Danish agricultural fields, we expect the plow layer to have a similar texture distribution throughout due to continuous agriculture activities with frequent mixing of the topsoil. So, slicing the first horizon before spline fitting was a good choice because it managed to maintain the textural homogeneity in the plow layer. In this way we were able to capture the influence of agricultural

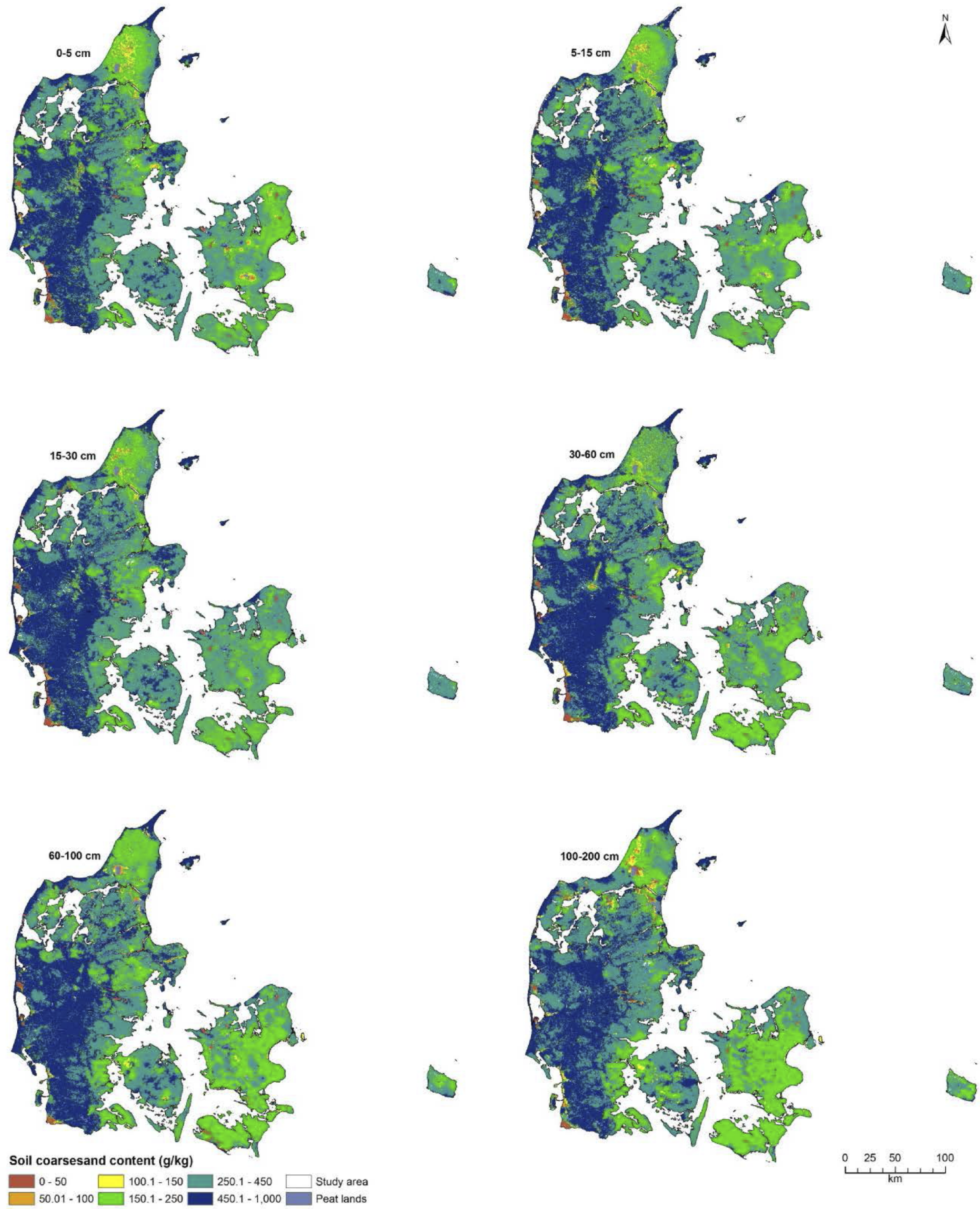

Fig. 6. Predicted coarse sand content of Danish soils at six depths.

operations in the distribution of texture components within the plow layer in Danish fields.

Of the different '*scorpan*' factors used in our prediction, higher predictive abilities were associated with geology, landscape, slope gradient, and other factors such as TWI, SAGA wetness index, and MRVBF and valley depth which correspond to erosion and deposition phenomena, but with a little contribution from factors such as slope aspect and direct insolation. These results are in good agreement with the recent findings of Greve et al. (2012a) who suggested a higher relative importance of geology and landscape, little importance of climatic variables, and extremely low importance of LSP such as slope aspect when predicting surface soil texture fractions on a national extent in Denmark. A higher RI for TWI (RI 91%) and SAGA wetness index (RI 81%) in predicting surface clay content where the soil type was sandy clay, was justifiable as greater values of these indices help to identify moist areas in a landscape (Rodhe and Seibert, 1999) where more clay could be expected.

Not surprisingly, based on the RI, the soil types in our prediction showed good predictive ability because it, first, was based

Fig. 7. Map of prediction uncertainty for clay content at a depth of 15 to 30 cm.

on high-density soil data and, second, several other parameters such as slope gradient that define the limits to agricultural machinery use, geological origin of the surface materials, drainage condition, and local expert knowledge were all considered while compiling it (Madsen et al., 1992). For these reasons, the soil types probably became the best instrument for defining the texture-landscape relationship during regression modeling for texture prediction. Similarly, the importance of landuse for surface texture prediction and geology in the deeper horizons also seemed obvious because geology corresponded to parent material and its influence is more pronounced at deeper horizons where soils are relatively younger than the soils of the surface. A higher influence of landuse in surface texture prediction is conditioned by the decade long agricultural practices which homogenized the top soils in major parts of Danish lands.

The range of R^2 values associated with our predictions were 0.20 to 0.55, which are comparable to many other soil quantitative prediction studies (e.g., Minasny et al., 2006, R^2 of 0.5; Stoorvogel et al., 2009, R^2 of 0.08–0.23; Malone et al., 2009, R^2 of 0.08–0.55). In DSM, the inherent large variation in the soil property of interest and the environment in which predictions are made could be the reason behind the range of R^2 values obtained in the studies listed above. On top of that, specific to our study, the distribution of soil profiles was not uniform as >50% of the profiles were not from the grid points. This could probably explain the range of R^2 we obtained. Moreover, due to the higher variability with soil depth as suggested by the variogram analysis, relatively lower prediction accuracies were expected there. In general, model performances decreased with depth for all properties. For the variogram parameters, lower prediction performances with depth might be due to the higher textural variability in deeper soil layers, which our prediction models were not completely able to capture. Unlike in the topsoil where the soils are regularly mixed, soils in the less disturbed deeper layers are more heterogeneous. Moreover, extension of toungs (tonguing effect) with surface material into the subsoil—a common pedogenetic phenomena in the Danish soil profiles—could also be linked to the high textural variability in deeper soil layers.

When comparing the predictive performance between the two sections, that is, topsoil and subsoil, it showed a poorer predictive performance for layers in Section II than in Section I. This might be due to a possible difference in soil development and distribution settings between the two sections influencing model prediction capabilities accordingly. Comparable values of variogram parameters in three upper layers (the plow layer) again confirmed the expected higher homogeneity of the topsoil due to the frequent soil mixing by agricultural activities, which is common practice in Denmark.

Our predicted maps showed the presence of sandy soils in the south-west and fine-textured clayey material in the eastern and central parts of the country for all depths. This corresponded well with the findings by Greve et al. (2007) of a similar texture distribution pattern in Denmark. The pattern was also consistent with the extent of Danish soil regions based on pedological development. Soils in the eastern and central Denmark were developed on a loamy or clayey parent material, whereas soils in the west were developed on sandy material (Madsen et al., 1992). This may also be linked to the geological activities in the last glaciation where thick ice covered the soils of central and eastern parts of Denmark, where the basal loamy till was deposited under the ice. In contrast to this, the western area of the country mainly consist of old and strongly eroded landforms with low-relief glacio-fluvial plains and sand deposits from which finer soil particles have been transported away. Schou (1949) has explained how these rather complex Danish landforms were formed during the late and postglacial transgressions with special emphasis on the development of sand-rich large glacial flood plains in western Denmark. On top of that, the presence of sand dunes and patchy inland sand deposits due to aeolian activities has probably increased the amount of sandy material in the west.

As Danish soils are highly leached (Jensen and Jensen, 1999), the higher clay content that we found in lower soil horizons, mostly in the central and eastern part of the country might be due to an eluviation of clay from the upper horizons through pedogenesis. With a fairly homogenous plow layer, the maps show only a small difference in the textural distribution in the top 0 to 5, 5 to 15, and 15 to 30 cm layers. The small difference may show the influence of, first, nonagricultural profiles where splines performed differently than in a disturbed soil depth and, second, of variable soil and '*scorpan*' factor interactions, which could be specific to the local environment.

CONCLUSIONS AND RECOMMENDATION

We conclude that equal-area quadratic splines appear to be a promising tool to predict the vertical distribution of soil texture in the Danish soil profiles. Fixing the splines for the plow depth was very logical, because of the expected rather homogeneous texture distribution in that layer, which is specific to Danish conditions. The new soil database derived from splines could be useful, for example, in nitrate leaching or soil-moisture related studies in the profiles as the continuous data could be translated into the texture data for any required depth interval within the spline. Moreover, fine-resolution texture property maps produced in this study could be useful for the many soil and environmental scientists and land managers in Denmark. Some of the applications include, soil compaction studies, mapping risk areas for pesticide leaching or ground water pollution, soil quality assessments and soil erosion modeling.

As the rules generated by the Cubist data mining tool were capable of finding complex relationships between soil texture and environmental covariates, we suggest using regression rules where we highly recommend using soil, landscape, and geological information for any future soil mapping activities in Denmark. We also believe that the predictability of the model could possibly be improved by incorporating other variables that explain textural variability in Danish soils, such as remote sensing indices or γ radiometric data. In the future, we envisage applying the same method to map topsoil texture properties using an existing dense topsoil point data (>40,000 soil samples). The possible usefulness of such maps to predict corresponding texture fractions in the subsoil horizons will also be explored.

ACKNOWLEDGMENTS

This work is a part of SINKS project (2009-2012) funded by the Danish Ministry of Climate and Energy. The authors also thank the editor and anonymous reviewers for their help in improving the quality of the manuscript.

REFERENCES

Adhikari, K., R. Bou Kheir, M.B. Greve, P.K. Bøcher, B.P. Malone, B. Minasny, A.B. McBratney, and M.H. Greve. 2012. Progress towards GlobalSoilMap.net soil database of Denmark. In: B. Minasny, B.P. Malone, and A.B. McBratney, editors, Digital soil assessment and beyond. CRC Press, Boca Raton, FL. p. 445–451.

Akaike, H. 1973. Information theory and an extension of maximum likelihood principle. In: B.N. Petrov and F. Csaki, editors, Second International Symposium on Information Theory, Tsahkadsor, Armenia. 2–8 Sept. 1971. Akademia Kiado, Budapest. p. 267–281.

Bendix, J. 2004. Geländeklimatologie. Gebrüder Borntraeger, Berlin.

Bishop, T.F.A., A.B. McBratney, and G.M. Laslett. 1999. Modeling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45. doi:10.1016/S0016-7061(99)00003-8

Boehner, J., and O. Antonic. 2009. Land surface parameters specific to topo-climatology. In: T. Hengl and H.I. Reuter, editors, Geomorphometry: Concepts, software, applications. Elsevier, New York. p. 195–226.

Boehner, J., R. Koethe, O. Conrad, J. Gross, A. Ringeler, and T. Selige. 2002. Soil regionalization by means of terrain analysis and process parameterization. In: E. Micheli, F. Nachtergaele, and L. Montanarella, editors, Soil Classification 2001. Eur. Soil Bur., Res. Rep. No. 7, EUR 20398 EN, Luxembourg. p. 213–222.

Bou Kheir, R., M.H. Greve, P.K. Bøcher, M.B. Greve, R. Larsen, and K. McCloy. 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. J. Environ. Manage. 91:1150–1160. doi:10.1016/j.jenvman.2010.01.001

Brewer, R. 1968. Clay illuviation as a factor in particle-size differentiation in soil profiles. In: Transactions of the 9th international congress of soil science. Angus and Robertson, London. p. 89–499.

Brooker, P.I. 1986. A parametric study of robustness of Kriging variance as a function of range and relative Nugget Effect for a spherical semivariogram. Math. Geol. 18(5):477–488. doi:10.1007/BF00897500

Brus, D.J., B. Kempen, and G.B.M. Heuvelink. 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62(3):394–407. doi:10.1111/j.1365-2389.2011.01364.x

Bui, E.N., B.L. Henderson, and K. Viergever. 2006. Knowledge discovery from models of soil properties developed through data mining. Ecol. Modell. 191(3–4):431–446. doi:10.1016/j.ecolmodel.2005.05.021

Burrough, P.A. 1993. Fractals and geostatistical methods in landscape studies. In: N. Lam and L. de Cola, editors, Fractals in geography. Prentice Hall, Englewood Cliffs, NJ. p. 87–112.

Campbell, N.A., M.J. Mulcahy, and W.M. McArthur. 1970. Numerical classification of soil profiles on the basis of field morphological properties. Aust. J. Soil Res. 8:43–58. doi:10.1071/SR9700043

Crave, A., and C. Gascuel-Odoux. 1997. The influence of topography on time and space distribution of soil surface water content. Hydrol. Processes 11:203–210. doi:10.1002/(SICI)1099-1085(199702)11:2<203::AID-HYP432>3.0.CO;2-K

Danmarks Geologiske Undersøgelse. 1978. Foreløbige geologogiske kort (1:25,000) over Danmark. DGU Serie A,3. Geol. Surv. of Denmark and Greenland, Copenhagen.

Danmarks Meteorologiske Institut. 1998. Danmarks Klima 1997. Danmarks Meteorol. Inst., Copenhagen.

Danner, A., K. Yi, Th. Mølhave, P.K. Agarwal, L. Arge, and H. Mitasova. 2007. TerraStream: From elevation data to watershed hierarchies. In: Proceedings of the 15th annual ACM International Symposium on Advances in Geographic Information Systems. ACM, New York. p. 212–219.</conf>

Desmet, P.J.J., and G. Govers. 1996. A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. J. Soil Water Conserv. 51(5):427–433.

Dobos, E., and T. Hengl. 2009. Soil mapping applications. In: T. Hengl and H.I. Reuter, editors, Geomorphometry, concepts, software and applications. Elsevier, New York. p. 461–479.

Freeman, T.G. 1991. Calculating catchment area with divergent flow based on a regular grid. Comput. Geosci. 17(3):413–422. doi:10.1016/0098-3004(91)90048-I

Gallant, J.C., and T.I. Dowling. 2003. A multi-resolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39(12):1347–1359. doi:10.1029/2002WR001426

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.

Greve, M.H., R. Bou Kheir, M.B. Greve, and P.K. Bøcher. 2012a. Quantifying the ability of environmental parameters to predict soil texture fractions using regression-tree model with GIS and LIDAR data: The case study of Denmark. Ecol. Indic. 18:1–10. doi:10.1016/j.ecolind.2011.10.006

Greve, M.H., R. Bou Kheir, M.B. Greve, and P.K. Bøcher. 2012b. Using digital elevation models as an environmental predictor for soil clay contents. Soil Sci. Soc. Am. J. 76:2116–2127. doi:10.2136/sssaj2010.0354

Greve, M.H., M.B. Greve, P.K. Bøcher, T. Balstgrom, H. Breuning-Madsen, and L. Krogh. 2007. Generating a Danish raster-based topsoil property map combining choropleth maps and point information. Dan. J. Geogr. 107(2):1–12.

Grunwald, S., editor. 2006. Environmental soil-landscape modeling: Geographic information technologies and pedometrics. CRC Press, Boca Raton, FL.

Grunwald, S., J.A. Thompson, and J.L. Boettinger. 2011. Digital soil mapping and modeling at continental scales: Finding solutions for global issues. Soil Sci. Soc. Am. J. 75:1201–1213. doi:10.2136/sssaj2011.0025

Hansen, S., N.E. Jensen, and H. Svendsen. 1990. DAISY: A soil plant atmosphere system model. NPO-Res. Rep. A10. Natl. Agency for Environ. Prot., Copenhagen, Denmark.

Hartemink, A.E., A.B. McBratney, and M.L. Mendonça-Santos, editors. 2008. Digital soil mapping with limited data. Springer, Dordrecht, the Netherlands.

Hartemink, A.E., et al. 2010. GlobalSoilMap.net: A new digital soil map of the world. In: J.L. Boettinger et al., editors, Digital soil mapping: Bridging research, environmental application, and operation. Springer, Dordrecht, the Netherlands. p. 423–427.

Henderson, B.L., E.N. Bui, C.J. Moran, and D.A.P. Simon. 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124:383–398. doi:10.1016/j.geoderma.2004.06.007

Hengl, T., G.B.M. Heuvelink, and D.G. Rossiter. 2007. About regression-kriging: From equations to case studies. Comput. Geosci. 33(10):1301–1315. doi:10.1016/j.cageo.2007.05.001

Hengl, T., and H.I. Reuters, editors. 2009. Geomorphometry: Concepts, software, applications. Developments in Soil Science. Elsevier, New York.

Heuvelink, G.B.M., and J.A. Huisman. 2000. Choosing between abrupt and gradual spatial variation. In: H.T. Mowrer and R.G. Congalton, editors, Quantifying spatial uncertainty in natural resources. Ann Arbor Press, Chelsea, MI. p. 111–117.

Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. McGraw-Hill, New York. p. 281.

Jensen, H.E., and S.E. Jensen. 1999. Jordfysik og jordbrugsmeteorologi Det fysiske miljø for plantevækst. DSR Forlag, Copenhagen.

Lagacherie, P., A.B. McBratney, and M. Volz, editors. 2007. Digital soil mapping: An introductory perspective. Elsevier, Amsterdam, the Netherlands.

Luk, S.H. 1979. Effect of soil properties on erosion by wash and splash. Earth Surf. Processes 4(3):241–255. doi:10.1002/esp.3290040305

Madsen, H.B. 1983. A pedological soil classification system for Danish soils. Pedologie 33(2):171–197.

Madsen, H.B., and N.H. Jensen. 1985. Pedological soil databases in Denmark. Geogr. Tidsskr. 85:1–8.

Madsen, H.B., A.B. Nørr, and K.A. Holst. 1992. The Danish soil classification: Atlas over Denmark. R. Dan. Geogr. Soc.. Copenhagen, Denmark.

Malone, B.P., A.B. McBratney, and B. Minasny. 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160:614–626. doi:10.1016/j.geoderma.2010.11.013

Malone, B.P., A.B. McBratney, B. Minasny, and G.M. Laslett. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154:138–152. doi:10.1016/j.geoderma.2009.10.007

McBratney, A.B., M.L. Mendonça-Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3–52. doi:10.1016/S0016-7061(03)00223-4

McBratney, A.B., and R. Webster. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J. Soil Sci. 37:617–639. doi:10.1111/j.1365-2389.1986.tb00392.x

McKenzie, N.J., and P.J. Ryan. 1999. Spatial prediction of topsoil properties using environmental correlation. Geoderma 89:67–94. doi:10.1016/S0016-7061(98)00137-2

Minasny, B., and A.B. McBratney. 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140(4):324–336. doi:10.1016/j.geoderma.2007.04.028

Minasny, B., and A.B. McBratney. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intell. Lab. Syst. 94:72–74. doi:10.1016/j.chemolab.2008.06.003

Minasny, B., A.B. McBratney, and R.M. Lark. 2008. Digital soil mapping technologies for countries with sparse data infrastructures. In: A.E. Hartemink, A.B. McBratney, and M.I. Mendonça-Santos, editors, Digital soil mapping with limited data. Springer, London. p. 15–30.

Minasny, B., A.B. McBratney, B.P. Malone, and I. Wheeler. 2013. Digital mapping of soil carbon. Adv. Agron. 118:1–47. doi:10.1016/B978-0-12-405942-9.00001-3

Minasny, B., A.B. McBratney, M.L. Mendonça-Santos, I.O.A. Odeh, and B. Guyon. 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Aust. J. Soil Res. 44:233–244. doi:10.1071/SR05136

Minasny, B., A.B. McBratney, and B.M. Whelan. 2005. VESPER version 1.62. Australian Centre for Precision Agriculture. Univ. of Sydney, Sydney, NSW, Australia.

Moore, A.W., J.S. Russell, and W.T. Ward. 1972. Numerical analysis of soils: A comparison of three soil profile models with field classifications. J. Soil Sci. 23:193–209. doi:10.1111/j.1365-2389.1972.tb01653.x

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57:443–452. doi:10.2136/sssaj1993.03615995005700020026x

National Survey and Cadastre. 2011. Produktspecifikation. Danmarks Højdemodel, DHM/Terræn. Data Version 1.0. Natl. Surv. and Cadastre of Den., Copenhagen.

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

Odeha, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63:197–214. doi:10.1016/0016-7061(94)90063-9

Odgers, N.P., Z. Libohova, and J.A. Thompson. 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189–190:153–163. doi:10.1016/j.geoderma.2012.05.026

Østergaard, H.S. 1990. Kvadratnettet for nitratundersøgelser i Denmark 1986–89. Landbrugets Rådgivningscenter, Landskontoret for Planteavl, Skejby, Århus, Denmark.

Ponce-Hernandez, R., F.H.C. Marriott, and P.H.T. Beckett. 1986. An improved method for reconstructing a soil profile from analyses of a small number of samples. J. Soil Sci. 37:455–467. doi:10.1111/j.1365-2389.1986.tb00377.x

Quinlan, J.R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann, San Mateo, CA.

Rodhe, A., and J. Seibert. 1999. Wetland occurrence in relation to topography: A test of topographic indices as moisture indicators. Agric. For. Meteorol. 98–99:325–340. doi:10.1016/S0168-1923(99)00104-5

Russell, J.S., and A.W. Moore. 1968. Comparison of different depths weightings in the numerical analysis of anisotropic soil profile data. In: Transactions of the 9th international congress of soil science. Angus and Robertson, London. p. 205–213.

Sanchez, P.A., et al. 2009. Digital soil map of the world. Science 325(5941):680–681. doi:10.1126/science.1175084

Schjønning, P., G. Heckrath, and B.T. Christensen. 2009. Threats to soil quality in Denmark: A review of existing knowledge in the context of the EU Soil Thematic Strategy. DJF Report Plant science no. 143. Aarhus Univ., Det Jordbrugsvidenskabelige Fakultet, Aarhus, Denmark.

Schou, A. 1949. Atlas of Denmark I: The landscapes. R. Dan. Geogr. Soc., Copenhagen.

Stjernholm, M., and A. Kjeldgaard. 2004. CORINE landcover update in Denmark. Final report. Natl. Environ. Res. Inst., Aarhus, Denmark.

Stoorvogel, J.J., B. Kempen, G.B.M. Heuvelink, and S. de Bruin. 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. Geoderma 149:161–170. doi:10.1016/j.geoderma.2008.11.039

Tanji, K.K. 1996. Agricultural salinity assessment and management. ASCE, New York.

Tranter, G., B. Minasny, A.B. McBratney, B. Murphy, N.J. Mckenzie, M. Grundy, and D. Brough. 2007. Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use Manage. 23(4):437–443. doi:10.1111/j.1475-2743.2007.00092.x

van Meirvenne, M., and I. van Cleemput. 2006. Pedometrical techniques for soil texture mapping at different scales. In: S. Grunwald, editor, Environmental soil-landscape modeling-geographic information technologies and pedometrics. CRC Press, Boca Raton, FL. p. 323–341.

Venables, W.M., and B.D. Ripley. 1994. Modern applied statistics with S-plus. Springer, New York.

Warrick, A.W., and W.R. Gardner. 1983. Crop yield as affected by spatial variations of soil and irrigation. Water Resour. Res. 19:181–186. doi:10.1029/WR019i001p00181

Yost, R.S., G. Uehara, and R.L. Fox. 1982. Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. Soil Sci. Soc. Am. J. 46:1028–1032. doi:10.2136/sssaj1982.03615995004600050028x

Young, I.M., and J.W. Crawford. 2004. Interactions and self-organization in the soil-microbe complex. Science 304:1634–1637. doi:10.1126/science.1097394