THE NEW DIGITAL SOIL MAP OF SWEDEN – FOR FREE USE IN PRECISION AGRICULTURE

Piikki K. and Söderström M.

Swedish University of Agricultural Sciences (SLU), Skara, Sweden

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract

The Digital Soil Map of Sweden (DSMS) was finalized in 2015. The present paper describes the mapping strategy, the estimated uncertainty of the primary map layers and its potential use in precision agriculture. The DSMS is a geodatabase with information on the topsoil of the arable land in Sweden. The spatial resolution is 50 m × 50 m and it covers > 90% of the arable land of the country (~2.5 million ha). Non-agriculture land and areas with organic soil are excluded. Access to a number of national datasets made it possible to build the DSMS. Results from two soil sampling campaigns (~15 000 samples) were combined with remote sensing data (gamma radiation data from airborne radiometric scannings and a digital elevation model derived from airborne laser scanning) plus a Quaternary soil deposit map. Multivariate adaptive regression splines models were parameterized for clay content and sand content and the models were deployed for the 50 m square grid. A 10-fold bootstrap cross validation was performed to estimate the uncertainty of the map layers. The mean absolute errors (MAE) were 5.6 % for clay content and 10 % for sand content. In 75% of the validation samples the clay content MAE was < 8 % and the corresponding value for sand was 14 %. The DSMS data is provided for free by the Geological Survey of Sweden. Examples of applications developed based on the DSMS are a web application used by farmers to generate prescription files for variable-rate seeding and variable-rate liming based on the DSMS clay content information. Another potential use of the map data is to enhance the modelling of nutrient and pesticide leaching from agricultural land. We envisage a wide use of the DSMS by farmers, extension officers, private companies, researchers and expect it to expand with more data layers; the current database can be used as input to pedotransfer functions to derive secondary information.

.

Keywords. *Digital soil map, precision agriculture, public data, remote sensing.*

Introduction

In 2006, the Swedish government commissioned the Swedish Board of Agriculture to do a national soil sampling campaign. The method and the results from the sampling have been summarized in two reports (Djodjic 2015; Paulsson et al. 2015) and a regional digital soil mapping study was reported by Söderström et al. (2016). In this paper we present the Digital Soil Map of Sweden (DSMS; Fig. 1) which is a free map product derived from the abovementioned national soil sample database in combination with remote sensing data. The DSMS consists of primary map layers (topsoil fractions of clay and sand) as well as secondary map layers derived by deployment of pedotransfer functions (to date: silt and USDA/FAO soil texture classes). The precision agriculture (PA) community has expressed its considerable interest in the DSMS, since detailed texture information normally need to be produced by the farmers themselves through traditional soil sampling. Already before the maps were finalized, a private company developed a decision support system for variable-rate seeding and variable-rate application of lime for soil structure improvement, based on the DSMS clay content layer. The aim of this paper is to describe the primary map layers of this map product, how they were produced, their uncertainty, as well as their potential use.

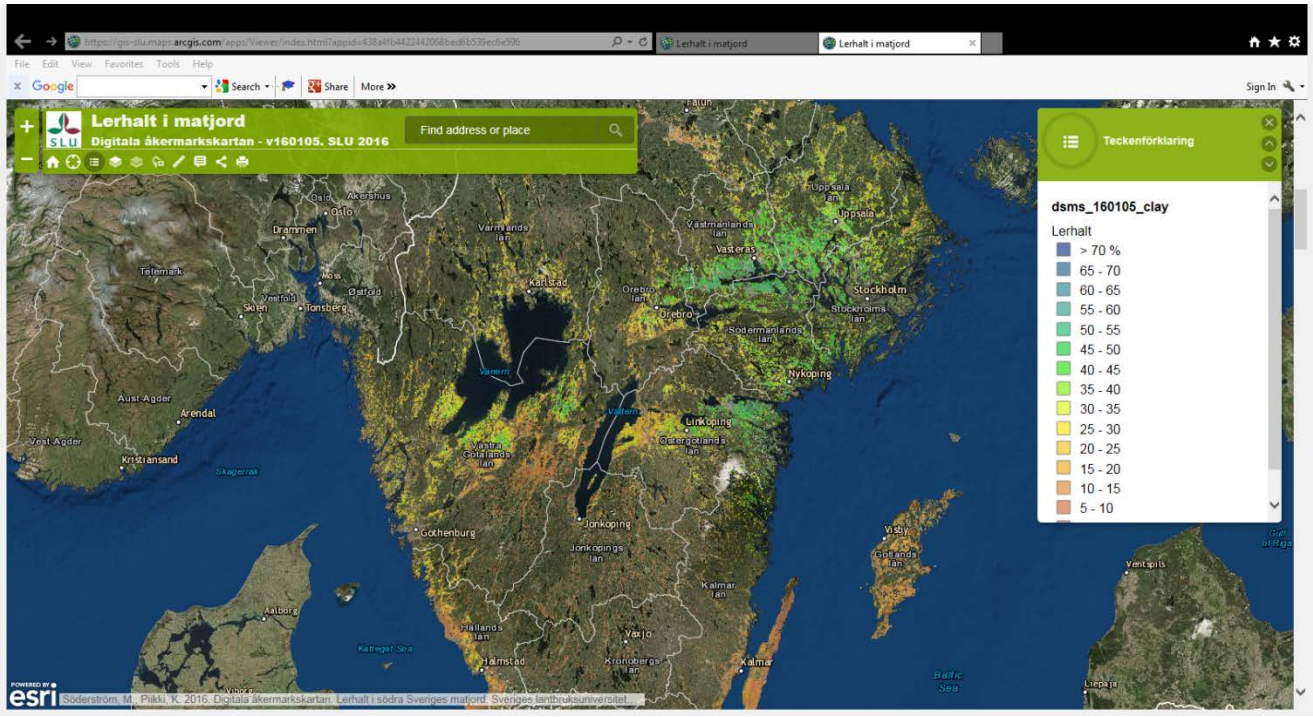


Fig 1. The Digital Soil Map of Sweden covers some 2.5 million hectares of arable land and is available for free. A map viewer of clay content is available on ArcGIS Online (http://arcg.is/238oOXf). Further information is found at www.precisionsskolan.se.

Materials and methods

Principles of digital soil mapping

In recent years, digital soil mapping (DSM) has become an established discipline (Minasny & McBratney 2016). In DSM, reference soil analyzes are used to calibrate (often multiple) exhaustive datasets to derive a geodatabase of soil properties, e.g. in the form of a digital raster map. The exhaustive data are often derived by sensor measurements, either measured from close range, < 2 m from the soil, so called *proximal sensing* (Viscarra Rossel et al. 2011), or from a much longer range, from aircrafts, satellites or drones (*remote sensing*). Also legacy map data can be used. The calibration is done by parameterizing multivariate prediction models that translate the exhaustive data (the predictors) the soil properties measured on the soil samples (the responses).

Data management

To handle the large amounts of data, and to create a homogeneous dataset, the entire map area was split into 23 subregions (Fig. 2). These were handled separately and the final maps were merged. Texture was not modelled in areas with soils classified as organic in the Quaternary geology map 1:50-100k by the Geological Survey of Sweden (SGU) or classified as wetlands in the Swedish Board of Agriculture's data base of agricultural land (known as the Block data base) in 2013. These areas were instead classified as organic soil in the final map. The Block data base was also used as a mask for all calculations, which means that areas not classified as farmland was not mapped.

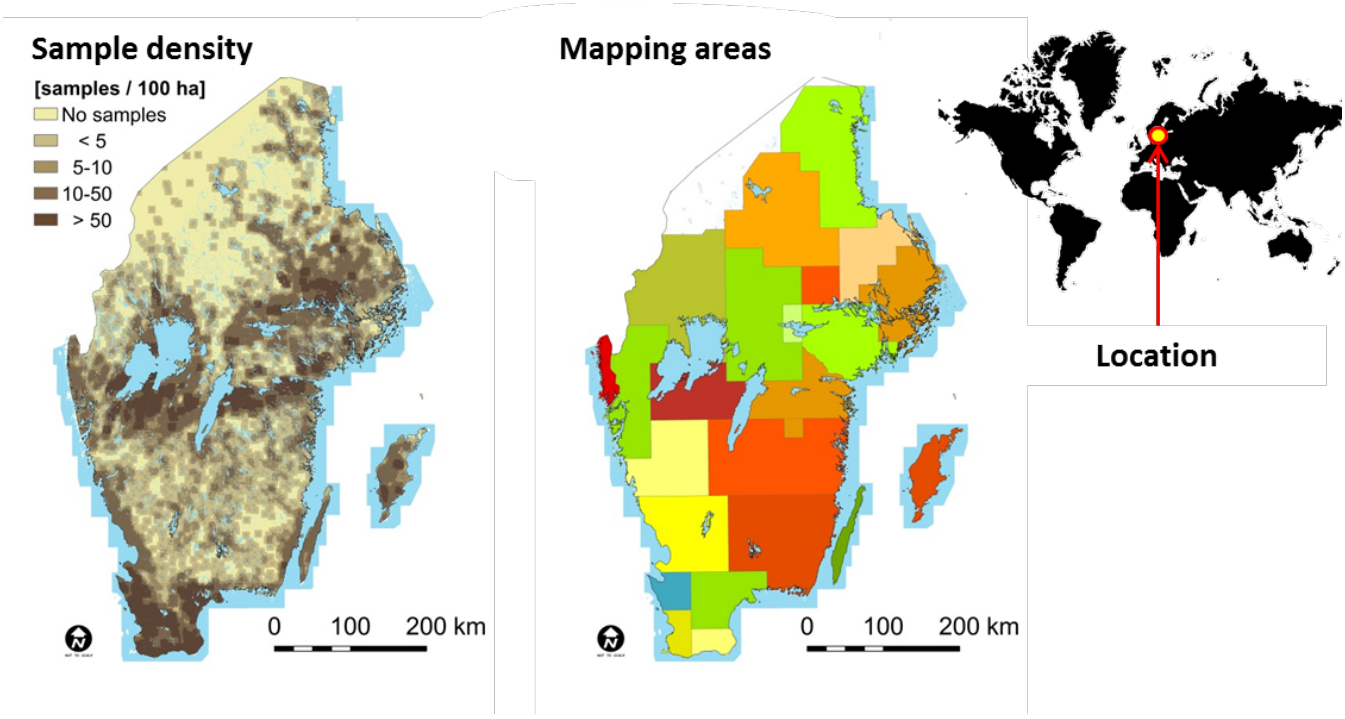


Fig 2. Maps of southern Sweden displaying soil sample density and mapping areas. The entire region was split into 23 subareas that were modelled separately and subsequently merged.

Predictor data

Quaternary Deposit Map

The Quaternary deposit (QD) maps of SGU were used as predictor data for topsoil sand and clay contents. It was necessary to first reduce the number of classes in the map. Two reduced

classifications were therefore made; one of seven classes (clay, sand, silt, till, till clay, organic soils, and other) and one with only three classes (clay, organic soils, and other). In addition, the polygon geometry was simplified to avoid erroneous classification due to uncertainty in boundaries between different soil classes; polygons smaller than 1 ha were merged with the neighboring polygon with the largest area.

Gamma Radiation Survey Data

Gamma radiation data from SGU airborne geophysical surveys were also used as predictor data. The dataset consisted of point registrations every 17 m along parallel flight lines which were separated by about 200 m (400 m in a few small areas). The measurements were made at two different altitudes, 30 m during the period 1968 to 1980, and 60 m after 1980. The footprint (measured area) of each measurement can basically be assumed to have a radius that is about 4 times the altitude (IAEA, 2003).

The isotopes recorded were ^{40}K, ^{232}Th and ^{238}U, although the latter was not used because of considerable noise in the measured data in some areas that caused overfitted non-robust prediction models. Measurement locations that were >300 m from agricultural land or <100 m from a lake or watercourse were removed. In addition, measurements of ^{40}K <0% or ^{232} Th <1 ppm were omitted. To avoid problems caused by outliers ^{40}K >4.5% was set to 4.5% and Th >30 ppm was set to 30 ppm. In some cases manual level adjustment / filtration of gamma radiation data was required.

The point data were interpolated to a 50-m square grid through ordinary 2 × 2 block kriging. We used exponential or spherical models with a nugget and the search radius was set to a maximum of 300 m regardless of range. In each of four circle sectors maximum 5 data points were used for interpolation.

Digital Elevation Model

Laser-scanned elevation data (Grid 2+) from the Swedish National Land Survey was used to calculate a number of terrain attributes that were used as predictors. First, the digital elevation model (DEM) was resampled to a resolution of 10 m × 10 m (original DEM: 2 m × 2 m). The resampled DEM was then used to calculate the landform (local topography) in three different scales, as the raster cell elevation compared to the neighborhood (5 ha, 50 ha, and 500 ha) elevation (Fig. 3).

Reference soil sample data

The reference soil sample data consisted of a combination of two national soil sampling series. They were spatially complementary to each other and formed together a 1 km square grid with a small random displacement of each sample. The sample density distribution is shown in Fig. 2. One soil sampling campaign was commissioned by the Swedish Environmental Protection Agency and has been described by Eriksson et al. (2010). The other, much more extensive, sampling was financed by the Swedish Board of Agriculture and has been described by Paulsson et al. (2015). In total there were about 15,000 samples available to calibrate prediction models. The samples were taken in the topsoil (upper 0.2 m depth). Seven to ten cores formed a circular support area of 3-5 m radius. The clay content was determined on dried (35-40 °C), milled and sieved (2 mm) samples by the sedimentation method (Gee and Bauder 1986) at the Eurofins Laboratory, Kristianstad, Sweden.

Modeling

Multivariate adaptive regression splines (MARSplines) modelling is a data mining regression technique that is used to parameterize the relationships between the exhaustive predictor variables (X) and the response variables (Y; Hastie et al. 2009). In essence, a MARSplines model consists of a set of piecewise linear regression functions, each defined above or below a specified threshold of the predictor variable MARSplines models can be additive models or allow for interactions among the basis functions. A benefit of this type of model is that it is possible to combine quantitative and qualitative predictor data.

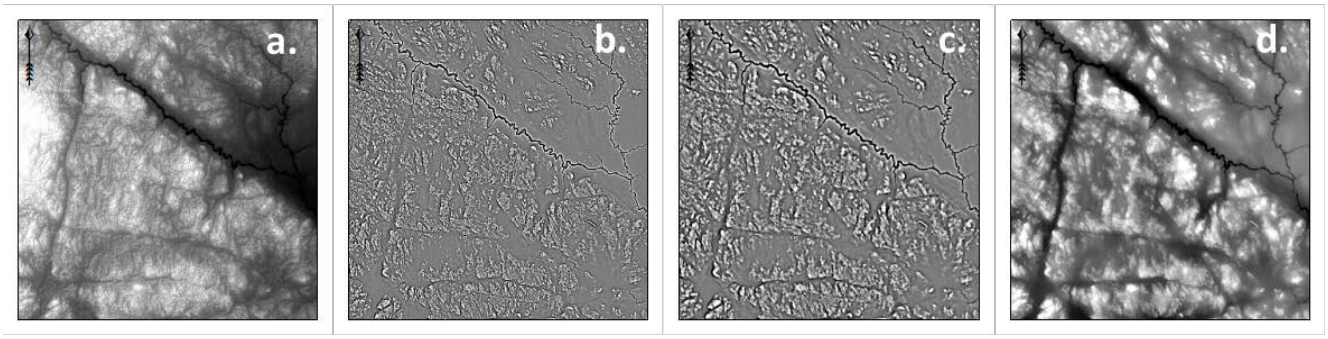


Fig 5. Predictor data derived from the digital elevation model: a) elevation above sea level, b-d) relative topography values calculated for different neighborhood sizes, b) 5 ha, c) 50 ha and d) 500 ha. Black indicate a (local) valley and white show (local) hills

In summary, eight predictor variables were used:

- Two continuous predictors from the airborne gamma surveys: Activity concentrations of ^{232}Th and ^{40}K.
- Four continuous predictors from the DEM: Elevation and landform in three scales.
- Two categorical predictors from the QD map: seven soil classes and three soil classes.

All predictor data were calculated onto a 50 m square grid. Grid cell values were then extracted for the 15,000 soil sample locations, which mean that there was one grid dataset with all predictor data as attributes and one reference point dataset with both predictor data and response data as attributes. For each of the 23 modeling areas (Fig. 2.), two MARSplines prediction models were calibrated, one for clay content and one for sand content. The models were deployed for the 50 m × 50 m grids and the grids were merged to full coverage (and these full extent rasters are the primary rasters of the DSMS). The secondary layers USDA/FAO texture classes and silt content were calculated from the primary layers. To avoid non-realistic values in individual grid cells, predicted values <0 were set to 0 and predicted values greater than the maximum value in the calibration point dataset were set to the maximum in the point dataset.

To validate the modelling procedure, a tenfold bootstrap cross-validation was performed as follows: a random tenth was removed from the reference soil sample dataset and a model was parameterized using the remaining 90% of the data. The parameterized model was then deployed for the withheld data subset. This procedure was repeated 10 times. The predicted clay and sand content values were compared to the clay and sand contents determined in the laboratory and the mean absolute error (MAE) of the predicted values, and the coefficient of determination r^2 for a linear regression between predicted and measured values were calculated.The average MAE and r^2 was calculated for 25 km square areas

Softwares

Most of the management of geographic data was done in the ArcGIS 10 (ESRI Inc., Redlands, CA, USA) with the program extensions Spatial Analyst and Geostatistical Analyst. The statistical analyses and the modeling were carried out in R and the Earth Package (R Core Team, 2015).

Results

Descriptive statistics

Descriptive statistics of the lab analyzed and the predicted clay and sand content values are found in Fig. 4. The clay content showed a somewhat right-skewed distribution with a median of 21 % clay while the sand content showed a bimodal distribution and a median value of 39 % sand. The frequency distributions of the predicted values (Fig. 4c-d) were similar to those of the observed values (Fig. 4a-b).

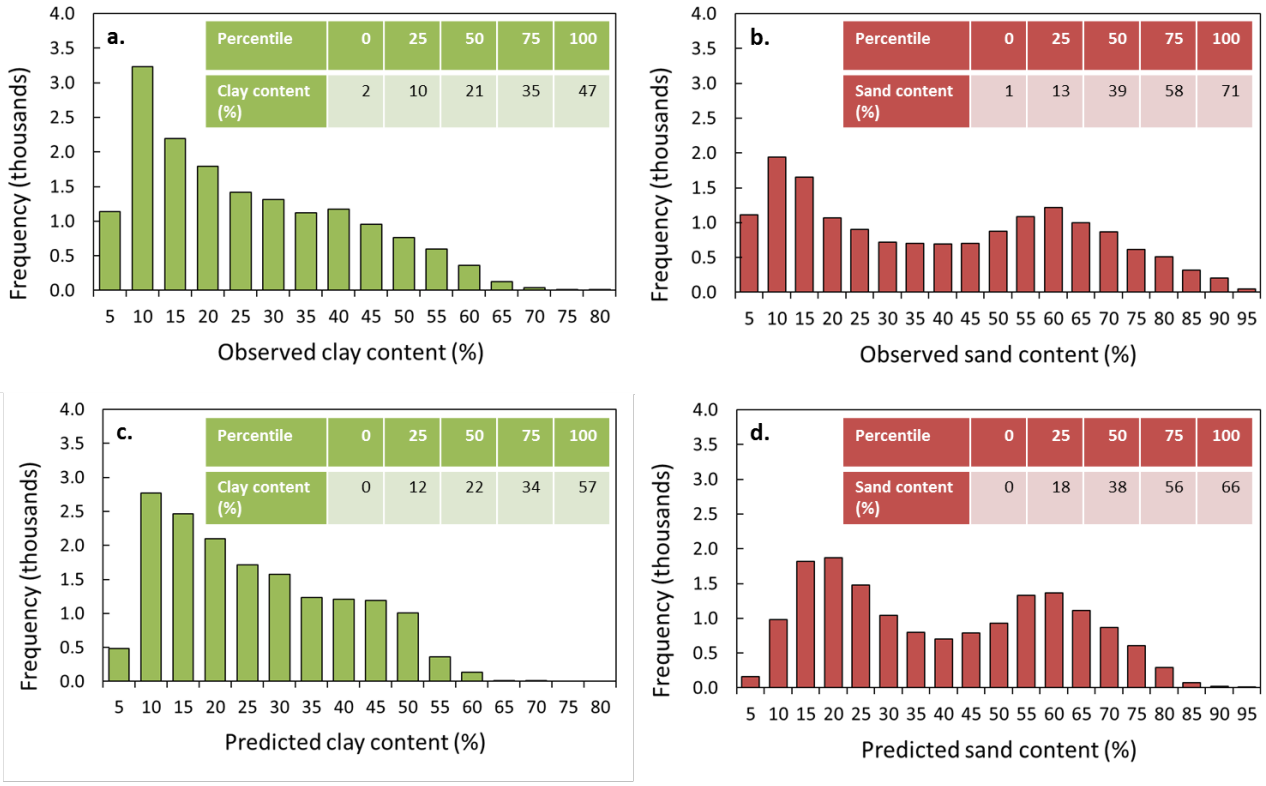


Fig 4. Histograms and statistics of the observed (a-b) and predicted (c-d) clay and sand contents based on the point validation dataset obtained by the tenfold bootstrap cross-validation.

Map layers

The mapping procedure was validated for each response variable. The final predicted maps of clay content and sand content are displayed in Fig.5. The large-scale variation patterns and the distribution of arable land can be inferred from these maps (the non-arable land is shown in gray color).

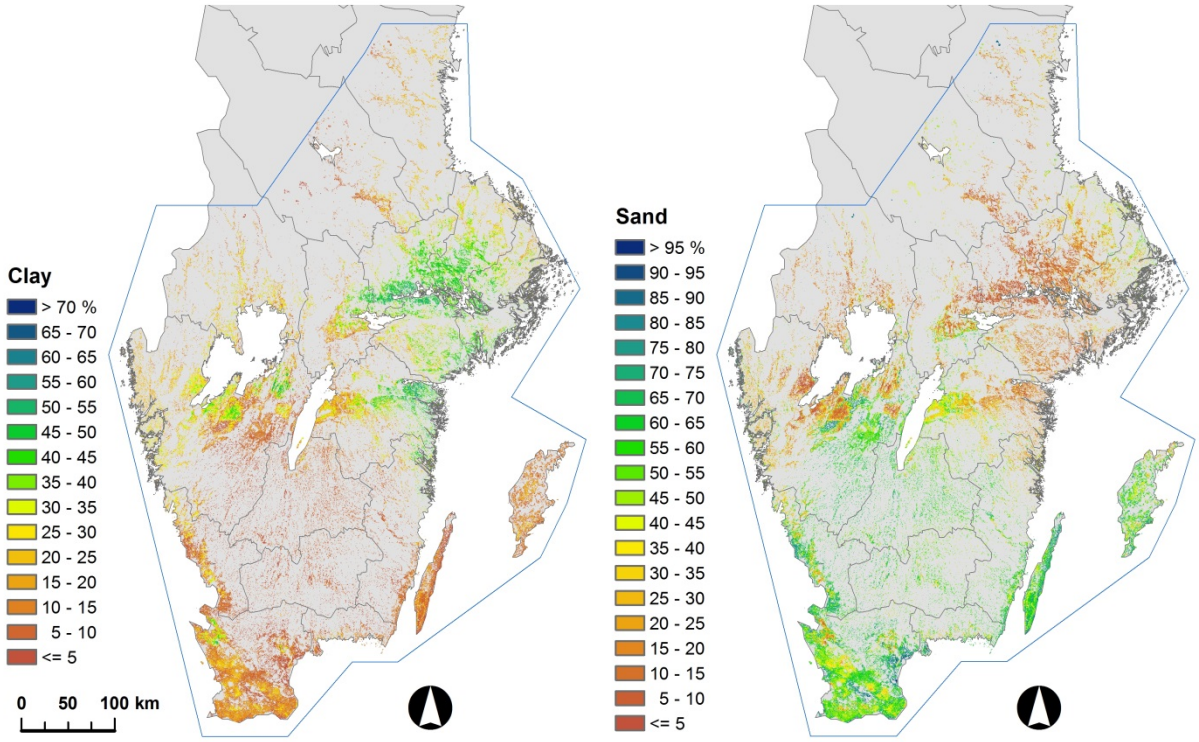
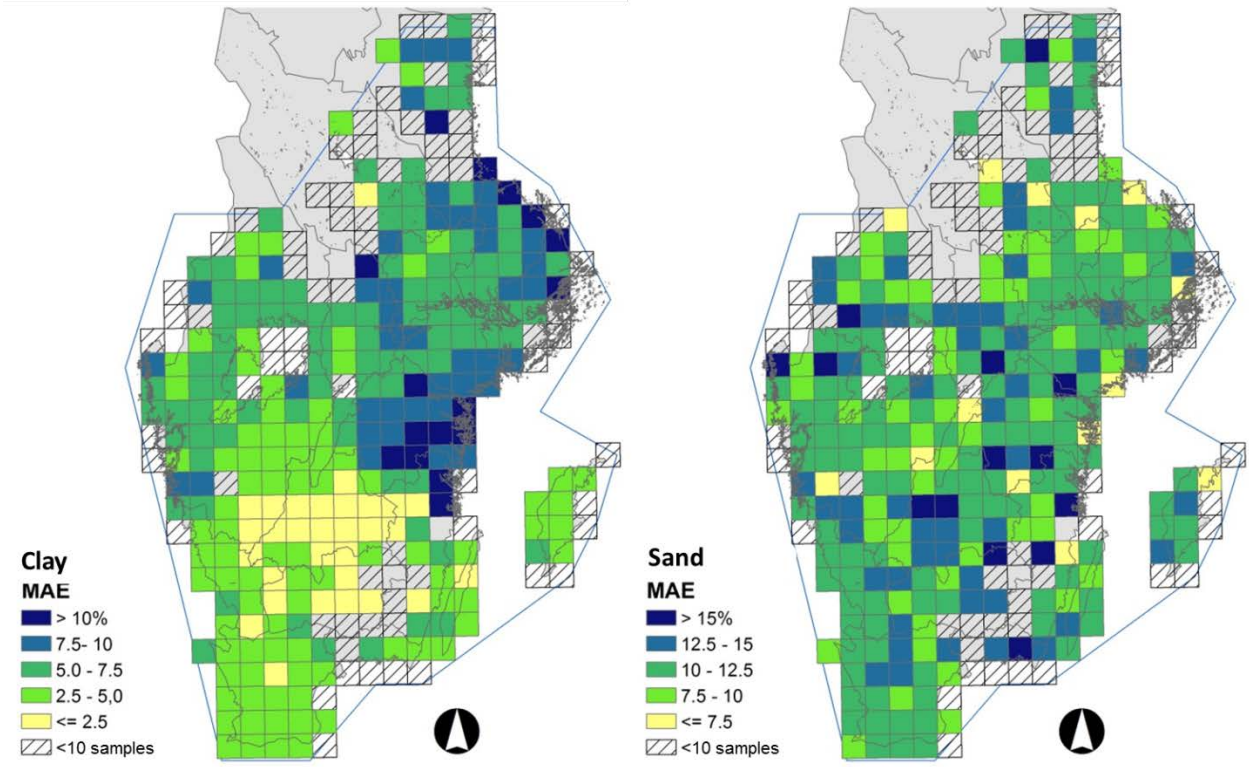


Fig 5. Predicted clay content and sand content of the Digital Soil Map of Sweden (DSMS).



Figur 6. The geographical distribution of the magnitude of errors.

Map accuracy

The MAE for each 25 km square grid cell is presented in Fig. 6. Grid cells with fewer than 10 validation samples were omitted from the MAE calculation. The overall MAE = 5.6% clay (for the entire map area) and the overall r^2 = 0.76 for the linear regression between the predicted and the measured clay content. For sand the overall MAE = 10% sand and the overall r^2 = 0.70. As the errors showed a very right skewed distribution the median absolute errors were considerably smaller than the MAEs (Fig. 7). The errors were generally higher in areas with a higher clay or sand content (not shown) and higher for sand than for clay, which is expected considering the fact that there was on average a larger fraction of sand than of clay in the soil (Figure 5 a-b).

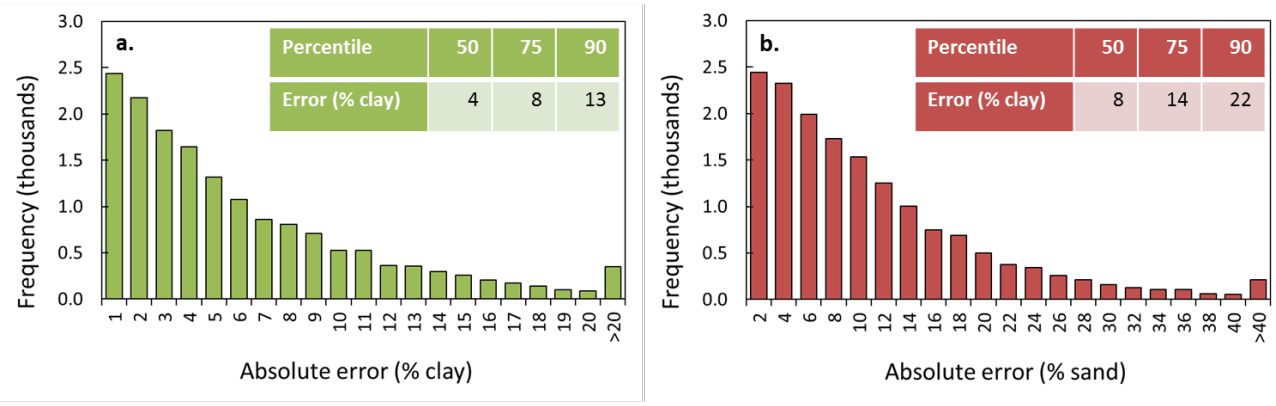


Fig 7. Histograms and statistics of the absolute errors in the point validation dataset obtained based on the tenfold bootstrap cross-validation.

Discussion

There are several possible usages for DSMS data in addition to enhanced modeling of phosphorus leaching from agricultural soil, which was the originally intended use of this map product. The high-resolution clay content maps can be used to target control measures against phosphorus leaching to parts of fields where they would be most efficient. Examples of such control measures are structural liming (which is normally applied on soils with > 15 % clay), construction of lime filled ditches and installation of phosphorus ponds. This does not mean that the clay content is the only thing that should be considered in decisions about actions

The DSMS can also be a valuable basis for those who want to vary inputs to their fields. In a test version of a web service (Fig. 8) the DSMS clay content layer serve as a basis for generating variable-rate prescription files. A popular use is to vary the seeding density. Normally the farmer chooses a seeding rate that guarantees an optimal canopy density in the clayey parts of the field. With a uniform seeding rate that will most often cause a too dense canopy in sandier field areas, where the germination conditions are often better. A variable seeding rate based on clay content and the farmer's own experience is expected to remedy such problems.

Fig. 8. A first version of a free-to-use web application for precision agriculture based on the Digital Soil Map of Sweden (DSMS) clay content (Dataväxt AB, Grästorp, Sweden). The seeding (or other input) rate can be adjusted according to the clay content information. The interface is interactive and the farmer can set the seeding rate for each of five clay content classes derived for the chose fields. Yellow = low clay content, brown = high clay content. The application is available at Markdata.se

The DSMS data will be made available for public use, free of charge, by SGU. All data are provided as raster map layers in tif-format with 50 m pixels. Uncertainty estimations are provided for the 25 km grid. During the development process, we also developed an algorithm to make the map interactive (Söderström et al. 2016) and we aim to develop an interface, through which users will be able to add available soil analyses and thereby improve the map locally. The DSMS geodatabase (predictors, primary map layers secondary map layers and reference soil sample data) that was built during the project can be used for improved estimates of a variety of other soil properties, in particular those related to texture, either directly from prepared predictor datasets, or by applying pedotransfer functions. Söderström et al. (2016) tested this for buffering capacity and target-pH, two spatial data inputs needed to adjust soil pH.

Summary

MARSplines models of topsoil clay content and topsoil sand content were parameterized with data from airborne gamma radiation scannings, a laser-scanned digital elevation model and a Quaternary deposits map as predictors. The models were deployed onto a 50-m square grid across 2.5 million hectares of arable soil in Sweden. The uncertainties of the soil attribute predictions were summarized for 25 $\times$ 25 km^2 areas. The overall median absolute errors were 4% units of clay and 8 % units of sand content. There are several possible uses for this type of free, high-resolution, digital soil maps. In practical precision agriculture, the clay content map layer has so far been used for determination of variable seed rates and in support of liming for adjusting soil pH, as well as liming for better soil structure – the latter being efficient for reducing leaching of phosphorus in certain soils. In addition to precision agricultural applications, detailed soil textural data is important in various general modelling efforts concerning leaching of pesticides and nutrients in watersheds and larger regions. The DSMS geodatabase is a structure that can be continuously developed by adding more data and applying pedotransfer functions to derive texture-related soil information.

Acknowledgements

The Swedish National Space Board (SNSB) financed the method development and the Geological Survey of Sweden (SGU) financed the national deployment.

References

Djodjic, F. (2015). *Jordartsfördelning och växtnäringstillstånd i svensk åkermark – sammanställning av resultat från Jordbruksverkets nationella jordartskartering*. SLU, Vatten och miljö, Rapport 2015:11. http://pub.epsilon.slu.se/12545/. Accessed 23 April 2016. (In Swedish)

Eriksson, J., Mattsson, L. & Söderström, M. (2010). *Tillståndet i svensk åkermark och gröda*. Naturvårdsverket, Rapport 6349. https://www.naturvardsverket.se/Documents/publikationer/978-91-620-6349-8.pdf . Accessed 23 April 2016. (In Swedish)

Gee, G. W., Bauder, J. W., & Klute, A. (1986). *Particle-size analysis. Methods of soil analysis*. Part 1. Physical and mineralogical methods, 383-411.

R Core Team (2015). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 23 April 2016.

Paulsson, R., Djodjic, F., Carlsson Ross, C.& Hjerpe, K. (2015). *Nationell jordartskartering, Matjordens egenskaper i åkermarken*. Jordbruksverket, Rapport 2015:19. http://webbutiken.jordbruksverket.se/sv/artiklar/ra1519.html. Accessed 23. April 2016. (In Swedish).

Hastie, T., Tibshirani, R.& Friedman, J. (2009). *The elements of statistical learning: data mining inference and prediction*. 2nd edition. New York: Springer Series in Statistics.

IAEA (2003). *Guidelines for radioelement mapping using gamma ray spectrometry data*. http://pub.iaea.org/mtcd/publications/pdf/te_1363_web.pdf . Accessed 23 April 2016. (In Swedish)

Söderström, M., Sohlenius, G., Rodhe, L., & Piikki, K. (2016). Adaptation of regional digital soil mapping for precision agriculture. *Precision Agriculture*, 1-20, doi: 10.1007/s11119-016-9439-8