Book of abstracts

On-line version 27 June 2016

Contents

Local Organizing Committee Aarhus University

Chair:

Senior scientist Mogens H. Greve

Members:

Post doc Maria Knadel

Post doc Yi Peng

Academic employee Mette Balslev Greve

Secretary Karina Rysholt Christensen

Secretary Jytte Christensen

Scientific Committee

Responsible:

Mogens H. Greve, *Department of Agroecology, Aarhus University, Denmark*

Luboš Borůvka, *Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague*

Members of the scientific committee:

Alex McBratney, *Soil Science, The University of Sydney, Australia*

Allan Lilly, *Environmental and Biochemical Sciences, The James Hutton Institute, Aberdeen, Scotland*

A-Xing Zhu, *Department of Geography, University of Wisconsin, Madison, USA*

Budiman Minasny, *Department of Environmental Sciences, The University of Sydney, Australia*

Cristiano Ballabio, *Land Resource Management - Soil Action, Joint Research Centre, European Soil Data Centre (ESDAC)*

David G. Rossiter, *Department of Crop and Soil Sciences, Cornell University, Ithaca, New York, USA*

Ganlin Zhang, *Soil Ressource and Remote Sensing Applications, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China*

Kabindra Adhikari, *Department of Soil Science, University of Wisconcin, Madison, USA*

Maria Knadel, *Department of Agroecology, Aarhus University, Denmark*

S. Young Hong, *National Academy of Agricultural Sciences, Rural Development Administration, Suwon, Gyeonggi-Do, Republic of Korea*

Titia Mulder, *ULB – Biochemistry and Earth System Modelling unit, Brussels, Belgium*

Keynote: Climate modeling: the role of subsurface information in predicting climate change

Jens Hesselbjerg Christensen

Danish Meteorological Institute, Denmark

Climate models are comprehensive computer codes or system models with compartments of all relevant components of the climate system, handling atmospheric, oceanic and cryospheric processes as well as a complete description of the energy, moisture, carbon and other component fluxes. To this end interactions between the atmosphere and land surfaces are crucial and processes happening at a fine geographical scale may be essential to get right in order to provide credible information about the state of the system. In this talk, aspects of model resolution and the need for detailed soil information in a climate modeling context will be discussed and some examples of crucial model biases related to soil types and soil processes will be touched upon.

Session 1 Soil Spectroscopy
Chair: Yi Peng

Optimizing Model Development and Validation Procedures of Partial Least Squares for Spectral Based Prediction of Soil Properties

Nimrod Carmon[1, 2] and Eyal Ben Dor[2]

[1]Porter School for Environmental Studies, Tel-Aviv University, Tel-Aviv, Israel
[2]Department of Geography and Human Environment, Tel-Aviv University, Tel-Aviv, Israel

The benefits of rapid and easy assessment of soil's attributes by means of reflectance spectroscopy at both the VIS-NIR-SWIR and TIR regions have been extensively reported by many workers. The most common method to extract a spectral based model is by using a Partial-Least-Squares Regression (PLS-R) algorithm. Whereas this method is highly affected by the spectral and chemical preprocessing technique and by the pre selection of the calibration and validation samples in the analysis, its robustness has not been fully examined. In this study we close the gaps previously reported within the PLS-R modeling approach and report on an automated way to extract an optimal model from the PLS-R approach. This innovative tool was developed at TAU and termed PARACUDA®. Within PARACUDA®, the All Possibility Approach (APA) is applied, in which preprocessing algorithms are applied on the spectral data in all possible combinations before additional analytical modeling approaches are applied.

In general, the new system optimizes the extraction of a reliable model and reduces specific effects caused by different modeling methods. For this purpose, we have developed a semi-supervised sequential preprocessing module for spectral data and a normalization module for the modeled attribute. These unique transformation modules emphasize important features in the data, improving calibration capabilities. Furthermore, we have developed a novel approach for conducting iterative inner validation procedures for evaluating the performance of a population of models rather than reporting a single model's result. The validation module is based on a Latin Hypercube Sampling (LHs) algorithm for sophisticated and representative data grouping which eliminates the effects of sample variation between the groups. By performing multiple validation iterations and reporting each iteration's results, a fuller picture of the modeling potential is discovered and reported. As spectral assignment is a key issue in chemometrics, our system produces unique spectral assignment data for each modeled attribute, showing band dependencies across the spectrum. This system was tested on a well-known legacy spectral library of Israeli soils with 52 chemical and physical attributes measured for 91 soil samples. The results were extremely accurate and showed an average improvement of model performance by ~10%. The spectral assignment ability within the PARACUDA® machine provides the extracted model with a high level of confidence.

Memory Based Learning: A New Data Mining Approach to Model and Interpret Clay Diffuse Reflectance Spectra

Asa Gholizadeh[1], Mohammadmehdi Saberioon[2] and Luboš Borůvka[1]

[1] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Czech Republic
[2] Laboratory of Signal and Image Processing, Institute of Complex Systems, South Bohemia Research Center of Aquaculture and Biodiversity of Hydrocenoses, Faculty of Fisheries and Protection of Waters, University of South Bohemia in České Budějovice, Nové Hrady, Czech Republic

Successful estimation of spectrally active clay in soil with Visible and Near-Infrared (VNIR, 400-1200 nm) and Short-Wave-Infrared (SWIR, 1200-2500 nm) spectroscopy depends mostly on the selection of an appropriate data mining algorithm. The aims of this paper were: to compare different data mining algorithms including Partial Least Squares Regression (PLSR), which is the most common technique in soil spectroscopy, Support Vector Machine Regression (SVMR), Boosted Regression Trees (BRT), and Memory Based Learning (MBL) as a very new promising approach for estimating the content of clay, to explore whether these methods show differences regarding their ability to predict clay from VNIR/SWIR data and to evaluate the interpretability of the results. The dataset consisted of 264 samples from large brown coal mining dumpsites in the Czech Republic. Spectral readings were taken in the laboratory with a fibreoptic ASD FieldSpec III Pro FR spectroradiometer. Leave-one-out cross validation was applied to optimize and validate the models. Comparisons were made in terms of the coefficient of determination (R^2_{cv}) and the Root Mean Square Error of Prediction of Cross Validation ($RMSEP_{cv}$). Predictions of the clay by MBL outperformed the accuracy of the other algorithms. It produced the largest R^2_{cv} and smallest $RMSEP_{cv}$ values, followed by SVMR. Actually, the main goal of this work was to develop a suitable MBL approach for soil spectroscopy, it showed that MBL is a very promising approach to deal with complex clay VNIR/SWIR datasets. A systematic comparison like the one presented here is important as the nature of the target function has a strong influence on the performance of the different algorithms.

Utilization of the Internal Soil Standard (ISS) Method to Optimize Exchange of Soil Spectral Libraries

Eyal Ben Dor[1], Gila Notesco[1], Agustin Pimstein[1], Cindy Ong[2], Ian Lau[2], Veronika Kopackova[3], Jose Dematte[4] and Danilo Romero[41]

[1]The Remote Sensing Laboratory, Tel Aviv University
[2]CSIRO Perth, Western Australia
[3]Czech Geology Survey, Czech Republic
[4]Department of Soil Science, University of Sao Paulo, Brazil

In the last 5 years, extensive effort has been devoted to building, sharing and using soil spectral libraries on national, continental and global scales. It is a well-known fact in the world spectral community that the different protocols used, are based on the user preference and experience. Nevertheless, it seems that a wide range of factors can affect soil reflectance spectra when using different protocols (spectrometers, geometry, replications, sample preparation, external conditions and even user touch), resulting in subtle or strong alterations in wavelength location, peak absorption and spectral shape or albedo intensity. Although in soil, very weak spectral features are monitored for chemometric purposes, these factors have been found to alter the robust use of a selected spectral model for a wide range of spectrometers and users. Establishment of a common protocol that will be agreed upon by all users might reduce this problem, as well as enable better comparison between spectral libraries. Even so, it has been shown that a strict protocol can only reduce error by about 50%, while the other variations remain. In 2013, Pimstein et al. suggested using an internal standard to reduce the other factors' effects, thereby minimizing the spectral variations after using an agreed-upon protocol. The internal standard idea was adopted from the “wet chemistry” discipline where variations between instruments are reduced by using well-known solutions as standards. In 2015 this idea has been further developed by Ben Dor, Ong and Ian who found a location in Southern Australia where white sand dunes provided a sample to be used as an ideal internal standard for soil spectral measurements (entitled: Internal Soil Standard –ISS). Since then, the ISS method has been utilized by several users showing remarkable results. In this presentation we will provide the basic concept of the ISS method and show it performance along several laboratories worldwide. A focus will be given to construct soil spectral libraries from different sources that generate the measurements using different protocols, spectrometer makes, calibration schemes and users touch. The results show that the ISS concept reduces significantly the systematic variations between spectral readings from many sources allowing an optimal spectral data merging. Furthermore, the proxy analyses before and after applying the ISS correction show a significant improvement of the spectral model's accuracy using the ISS method. The final and important conclusion of this presentation is that using the ISS method as a routine stage at any laboratory with any protocol would allow better and accurate sharing, exchanging and utilizing of enlarging numbers of soil spectral libraries worldwide.

Session 2 Digital Soil Mapping of SOC

Chair: Alex McBratney

USDA CarbonScapes — A National Look at Carbon Landscapes and Meeting Place for the USDA Carbon Inventory and Modeling Community

James A. Thompson[1], Sharon W. Waltman[2], Kurt Donaldson[3], Maneesh Sharma[3], Frank Lafone[3], Mike Wilson[4], and Skye Wills[5]

[1] Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV, USA
[2] UDSA-NRCS-National Soil Survey Center Geospatial Research Unit, Morgantown, WV, USA
[3] West Virginia GIS Technical Center, West Virginia University, Morgantown, WV, USA
[4] USDA-NRCS-Soil Science Division, Lincoln, NE, USA
[5] UDSA-NRCS-National Soil Survey Center, Lincoln, NE, USA

Understanding the carbon cycle is one of the most difficult challenges facing scientists studying the global environment. Without an understanding of C sequestration potential across the landscape, it will be difficult to effectively manage our landscapes for future mitigation of atmospheric CO_2. Furthermore, policy actions to enhance C sinks and remove C sources can be prioritized only when their relative magnitudes are known. In response to the need for readily accessible and reliable data on terrestrial C stocks, the US Department of Agriculture (USDA) made a commitment to making USDA carbon-related information available to the public via a single website. A collaborative effort between the West Virginia GIS Technical Center (WVGISTC) and the Natural Resources Conservation Service (NRCS) National Soil Survey Center Geospatial Research Unit (NSSC-GRU) at West Virginia University led to the creation of an online application that allows a diverse set of clients to access data, models, and other tools related to regional and national inventories of C stocks in an understandable way. This web application, known as USDA CarbonScapes (carbonscapes.org), is organized into four major sections: ATLAS, DATA, EXPLORER, and LEARN. ATLAS summarizes reports for specific C pools in the landscape for C stock or C mass by a particular area (county, watershed, and major land resource area) within the conterminous United States. DATA catalogs and links users to USDA terrestrial C data and model resources published on Data.gov and other web sites. EXPLORER provides advanced users more in-depth tools to visualize and analyze C models and associated map layers as well as create maps, share geospatial links, and extract GIS data. LEARN provides useful links to easy to understand explanation of C in the terrestrial biosphere. The goal of USDA CarbonScapes is to provide a useful and easy to navigate web map application to educate and answer questions for stakeholders about USDA inventory, modeling, and mapping of terrestrial biosphere C across the landscape. USDA soil, forest and crop resource inventories are featured along with USDA C sequestration models used for conservation planning.

Coupling high resolution data and national baseline estimates for farm scale soil carbon auditing

C Hedley[1], P Roudier[1], B Malone[2], B Minansy[2] and AB McBratney[2]

[1]Landcare Research, Palmerston North, New Zealand
[2]Centre for Carbon, Water & Food, Faculty of Agriculture and Environment, The University of Sydney, Australia

There has been considerable effort by national governments to develop greenhouse gas inventories that include accounting for soil carbon stocks and stock changes, as a commitment to the reporting requirements of the United Nations Framework Convention on Climate Change (UNFCCC). In parallel, independent verification methods have been developed for soil carbon trading initiatives, and these typically target smaller private land holdings, to audit management strategies implemented to sequester carbon into soils. Here we present a method that disaggregates coarse resolution, national soil carbon maps to fine resolution, farm scale maps. The method takes advantage of high resolution sensor datalayers and the output maps can be used to stratify the farm and guide sampling for the purpose of assessing soil carbon change through time.

The disaggregation method used, named dissever, is a downscaling approach developed by Malone et al. (2012). In this study, it has been applied to national soil carbon models, using high resolution data available locally to disaggregate the national maps to the farm scale. An iterative regression process develops a quantitative relationship between the fine scale covariates and the coarse scale soil organic carbon map. An extension of the original dissever method enables different regression methods to be tested and compared (generalised additive models, linear models, MARS splines, random forests and Cubist). The method also uses bootstrapping on the final dissever model to derive prediction uncertainties.

The national model provides a baseline value for one case study site, a 160 ha pastoral farm in the Manawatu region of New Zealand, of 91.84±22.88 t C ha^{-1} (to 0.3 m soil depth). Disseveration then uses the fine gridded environmental covariate data to drive the downscaling procedure on a block by block basis, with mass preservation. The available fine scale covariates used are: digital elevation from airborne laser scanning (< 1m resolution) and derived terrain attributes, gamma-ray spectrometry (10 m resolution), electromagnetic sensor data (10 m resolution), and a legacy (polygon) soil map (1:6,000).

There is generally good agreement between the original data and the back transformed disaggregated data, with concordance correlation coefficient (CCC) > 0.8 for all regression methods tested. Performance indicators (CCC, RMSE, R^2, bias, SE) show that Cubist outperforms the other models with CCC of 1.00, RMSE of 0.15, R^2 of 1.00, bias of 0.00 and SE of 0.15.

The national model therefore provides a baseline value for any part of the country, and the disaggregation method can be used to guide local sampling through time, which is necessary to account for any management effects on soil organic carbon stocks.

Reference:

Malone, Brendan P, Alex B McBratney, Budiman Minasny, and Ichsani Wheeler. 2012. “A General Method for Downscaling Earth Resource Information.” Computers & Geosciences 41. Elsevier: 119–25.

Global space-time soil carbon assessment

José Padarian[1], Uta Stockman[1], Budiman Minasny[1] and Alex B. McBratney[1]

[1]*Faculty of Agriculture and Environment, The University of Sydney, New South Wales 2006, Australia*

Climate change including change in temperature, elevated CO_2 concentrations, increased rainfall variability, and altered land-use will have a great impact on soils. The influence of these factors will create a dynamic feedback between soil and the environment. Because of the inherently dynamic nature of soil change, addressing these questions requires empirical data over time. While there are efforts to collect soil data round the globe for estimation of soil carbon stocks, the results are still uncertain.

This paper produces a global space-time assessment of soil carbon dynamics in different biomes and ecozones of the world accounting for impacts of environmental factors. We utilised the world's prior investment in soil data infrastructure gathered over the past half-century and beyond to provide a comprehensive global space-time assessment of soil organic carbon (SOC) dynamics. We first applied the scorpan approach to reveal the effect of soil forming factors on SOC. Following this, we incorporated land cover change dynamics, generating a two-stage model capable of predicting SOC content given a land-use history. We present global SOC maps and meta-statistics by ecozones to illustrate the change in time of this critical soil property.

The outcome can be used for an analysis of the relative effects of climate change and soil management on soil carbon across the globe. These observations will lead to a design of optimised soil carbon change monitoring network.

A million dollar challenge to map peatlands in Indonesia

Budiman Minasny[1], Rudiyanto[2] and Budi Indra Seyiawan[2]

[1] Faculty of Agriculture and Environment, The University of Sydney, New South Wales 2006, Australia
[2] Department of Civil and Environmental Engineering, Bogor Agricultural University, Indonesia

The Indonesian government recently announced a million dollar prize competition to find a more accurate and faster way of mapping the extent and depth of Indonesia's peatlands. Never before, soil mapping can make one become a millionaire. Eager soil mappers around the world seek collaboration with Indonesian scientists to enter the competition.

This call is in response to last year's devastating peat fire disaster in Indonesia. Tropical peatlands provide important ecosystem functions essential for soil security: biomass production, water supply, carbon storage, and biodiversity conservation. Effort has begun on restoring degraded peatlands, but one of many obstacles faced by the restoration agency is the lack of accurate map that made it difficult to identify areas that need to be conserved.

This paper presents digital mapping techniques that can be used to map peat depth in Indonesia. First we will review some techniques that have been proposed in mapping peat depth, both in the tropical area and the northern hemisphere. These include Lidar and gamma radiometrics. We then evaluated the scorpan spatial function approach using field observations in combination with widely-available covariates. We assessed the efficiency and accuracy of DSM techniques in a peatland of 600,000 ha in Ogan Komering Ilir, South Sumatra. To begin with, we defined the mapping extent based on a peat hydrological unit, where a peatland is bounded by at least two rivers. We used the 30 m DEM from the SRTM as covariates. We formulated the spatial soil prediction functions using Cubist regression tree, and Quantile Regression Forests (QRF). The results showed that peat depth can be predicted accurately using elevation and distance from the nearest river. Both Cubist and QRF performed similarly in predicting peat depth with a RMSE less than 1m, however the 90% prediction confidence interval for QRF is wider than that of Cubist. We also calculated the carbon stock for the peatland using the estimated peat depth and C density measurements. Finally, we evaluated the use of stratification-based sampling to increase the efficiency of sampling and also covariate coverage.

Session 3 Digital Soil Mapping of SOC
Chair: Budiman Minasny

3D modelling of Scottish soil properties

Laura Poggio[1] and Alessandro Gimona[1]

[1] The James Hutton Institute, Craigiebuckler, Aberdeen (Scotland, UK)

Soil properties vary in space and time and the prediction of their variation is an important part of environmental modelling as it has to consider and communicate their inherent uncertainty. In this study the continuous vertical and lateral distributions of relevant soil properties in Scottish soils were modelled with a 3D-GAM+GS approach following globalsoilmap.net specifications. The values at each cell for each of the considered depth layers were defined using a hybrid GAM-geostatistical 3D model, combining the fitting of a GAM (Generalised Additive Models) to estimate the trend of the variable, using a 3D smoother with related covariates; and Gaussian simulations of the model residuals as spatial component to account for local details. A dataset of about 26,000 horizons (7,800 profiles) was used for this study. A validation set was randomly selected as 25% of the full dataset. Numerous covariates derived from globally available data, such as Digital Elevation Model and Remote Sensing derived information are considered. The results showed good validation metrics and an accurate reproduction of the spatial structure of the data for a range of soil properties, such as organic carbon, pH, texture and soil depth. The results have an out-of-sample RMSE between 10 to 15% of the observed range when taking into account the whole profile. The results mirror the morphology and the soil patterns with high organic soil on the west and agricultural areas on the east. The approach followed allows the assessment of the uncertainty of both the trend and of the residuals.

Mapping Soil Organic Carbon Stocks Using a General 3D Mapping Approach in the Northeast Tibetan Plateau, China

Ren-Min Yang[1] and Gan-Lin Zhang[1]

[1] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

In the Tibetan Plateau, alpine grassland is the most widespread ecosystem and plays an important role in the storage of SOC. Therefore, there is a need for accurate estimate of soil organic carbon (SOC) stocks for understanding the role of alpine soils in the global carbon cycle.

The study area is located in the northeast of the Tibetan Plateau (ca. 30000 km^2). We tested a method for mapping digitally the continuous distribution of the SOC stock in three dimensions. First, a step-wise exponential depth function was defined to describe SOC depth distribution, with four parameters. This depth function integrated the spatial distribution of the “mattic epipedon” which is a special surface horizon with intensive roots. Such topsoils rich in organic matter. It usually leads to a sharp decrease in SOC contents with depth. Consequently, a monotonic decreasing function may result in an unrealistic distribution of SOC in the mattic epipedon because of high SOC content in topsoils. Second, a combined model of classification and regression analysis in random forest was applied for mapping parameters of soil depth functions using environmental covariates across the study area. Third, SOC stocks were predicted by using soil depth functions at each location.

The defined soil depth function provided a mean R^2 of 0.91 between the observed and fitted SOC content at calibration sites. Prediction models resulted in high prediction accuracy. The mean RMSE value of independent validation was 0.94 kg m^{-2}. By applying the predicted parameters of soil depth functions, we mapped the spatial distribution of SOC stocks across the study area. An average SOC stock in the mattic epipedon was estimated to be 4.99 kg m^{-2} in a mean depth of 14 cm. The average stock for the 0-30 cm layer was 5.54 kg m^{-2}, that of 6.11 kg m^{-2} for the 0-50 cm layer and 6.89 kg m^{-2} for the 0-100 cm layer. The amounts of SOC in the mattic epipedon, the upper 30 cm and 50 cm accounted for about 21 %, 80 % and 89 %, respectively, of the total SOC stock in the upper 1 m depth. By investigating the mattic epipedon, we were able to quantify the role of such an uppermost soil layer in storing SOC. Information on this layer is valuable for mapping the realistic distribution of SOC stocks in the Tibetan grasslands. Compared with previous estimates, our approach resulted in more reliable predictions.

High resolution digital soil organic carbon mapping in Western Greenland

Philipp Gries[1], Karsten Schmidt[1], Thorsten Behrens[1] and Thomas Scholten[1]

[1] University of Tuebingen, Chair of Soil Science and Geomorphology

Soils are an important carbon sink, which stores up to 2500 Pg carbon worldwide. Thereof, arctic soils contain the major portion of soil organic carbon (SOC). Being a sensible ecosystem, the Arctic is sensitive to climate change. Hence, thawing of permafrost-affected soils to greater depth and for longer periods increases the release of CO2 to the atmosphere, which queries soils as important carbon pool. There are several studies on predicting SOC content of arctic soils focusing on different depths, scales and regions. Most surveys base on soil or landscape units representing averaged point information. Being limited respecting intense natural and spatial variation of soils, these approaches are susceptible to errors of estimation. The spatial heterogeneity of soil properties, especially soil carbon, results from environmental factors varying highly in space and scale. Although, dealing with diversified soil properties, pixel based approaches are less widespread at pedological studies of the Arctic. This study comprises an area-wide prediction approach of SOC at two different study areas in Western Greenland using digital soil mapping (DSM) and data mining (DM). The objectives are (i) identifying major environmental factors controlling spatial variation of SOC in Western Greenland, (ii) creating high-resolution maps of relevant soil properties (iii), and to estimate carbon stocks. Being an ongoing study, intermediate results are presented.

We used two study areas (2 km² each) in Western Greenland (~67° N). The first area is four kilometers east to Sisimiut (Holsteinsborg) close to the seaside. The second area is located 20 kilometers east to Kangerlussuaq (Søndre Strømfjord) at Lake Aajuitsup Tasia close to the Greenland Ice Sheet (GrIS). Onshore oceanic conditions in Sisimiut turn into continental climate at GrIS showing less precipitation, more sunshine hours and increasing permafrost thickness. Input data of the DSM model includes soil data from fieldwork (bulk density, SOC, soil moisture, active layer thickness, vegetation cover). Environmental variables contain a wide range of local, complex and combined terrain attributes as well as remote sensed data (NDVI, soil moisture, surface roughness). Combinations of DSM and DM (random forest, support vector machines, artificial neural networks) are used to identify relationships between environmental variables and soil attributes for modeling high-resolution digital soil maps and calculating carbon stocks.

Intermediate results contain new digital elevation models created by aerial photos and Structure from Motion techniques. They have higher resolution (5x5 m) and are superior in comparison to measured altitudes by GPS ($R^2 = 0.83/0.76$, RMSE = 6.23/5.70) compared to existing ones ($R^2 = 0.77/0.62$, RMSE = 8.39/10.97). The topography, glacial and periglacial sediments of the study areas are comparable regarding the terrain variables aspect, profile curvature, flow direction, flow path length, Terrain Classification Index for Lowlands and Topographic Position Index. However, at the seaside, slope, plan curvature, flow accumulation, LS-Factor, valley depth, and relative mass balance are more than twice as high as close to GrIS. We divided both study areas into four landscape units using k-means cluster analysis on the terrain attributes listed. We used a random sampling design to select 10 sampling location per unit to calibrate respective DSM models. According to this concept, next steps will be collecting soil samples, laboratory and statistical analysis as well as high-resolution modeling of soil properties, especially SOC, in both study areas in Western Greenland.

Changes of soil organic matter stocks in Bavaria

Anna Kühnel[1], Martin Wiesmeier[1], Peter Spörlein[2], Bernd Schilling[2], Ingrid Kögel-Knabner[1]

[1]Lehrstuhl für Bodenkunde, Department für Ökologie und Ökosystemmanagement, Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt, Technische Universität München, Freising-Weihenstephan, 85350, Germany
[2]Bayerisches Landesamt für Umwelt (LfU), Hof, 95030, Germany

Climate change will have profound impacts on organic matter stocks and thus on the functionality of soils. The predicted rising temperatures in Bavaria might lead to an increased decomposition and release of soil carbon into the atmosphere, which would deteriorate a number of important soil functions. Information about the effect of rising temperatures on soils is however scarce.

In order to identify how organic carbon stocks in soils have already changed over the last 30 years, we used soil data from about 150 long term observation sites with constant management practises. The long term observation sites are homogeneously distributed over Bavaria and comprise forest, grassland and agricultural systems. These sites have been established in the middle of the 80’s and have been re-sampled approximately every 10 years.

In about two thirds of the analysed sites soil organic carbon changes could be detected. It is noteworthy, that there are sites with declining and sites with rising carbon stocks. The observed changes in organic carbon stocks will then be related to soil type, landuse, management, topographic and especially climatic data using random forest models. Thus, we can identify potential drivers of the soil organic carbon changes.

Soil organic carbon mapping at different scales in Madagascar

N. Ramifehiarivo[1], A. Andriamananjara[1], H. Razafimahatratra[2], T. Razafimbelo[1], M. Rabenarivo[1], A.Rasolohery[3], N. Ranaivoson[1], F. Seyler[4], A. Albrecht[5], F. Razafindrabe[6], M. Brossard[7] and H. Razakamanarivo[1]

[1] Laboratoire des RadioIsotopes, University of Antananarivo, Madagascar
[2] Ecole Supérieure des Sciences Agronomiques, University of Antananarivo, Madagascar
[3] Conservation International, Antananarivo, Madagascar
[4] IRD Institut de recherche pour le développement, UMR ESPACE-DEV Montpellier, France
[5] IRD Institut de recherche pour le développement, UMR Eco&Sols Montpellier, France
[6] Institut Géographique de Madagascar, Antananarivo, Madagascar
[7] IRD Institut de recherche pour le développement, UMR Eco&Sols Cayenne, France

The assessment of soil organic carbon is important for carbon sequestration known as the central driver in climate change mitigation and soil fertility. For Madagascar, advanced research on soil carbon mapping at multi-scales is required to improve the accuracy of the existing maps. This study aimed to (i) model the spatial distribution of Soil Organic Carbon (SOC) at local, regional and national scales from the relevant environmental variables, (ii) develop map of SOC at this three scales. For this, spatial models from the Random Forest algorithm were developed by using 13 pedoclimatic, topographic, and vegetation variables at local scale (Commune Rural of Didy, 1416 Km² of area) with 213 soil profiles (0-30 cm depth); at regional scale (Eastern humid ecoregion, 205 858 Km²)) with 800 soil profiles (0-30 cm depth); and national scale (587 000 Km²) with 1996 soil profiles (0-30 cm depth). The results showed that the prediction models at local scale had the best predictive capacity compared to regional and national scales, with a Root Mean Squared Error (RMSE) equal to 15.47 Mg.ha^{-1} for local scale, 24.55 Mg.ha^{-1} for regional scale, and 25.81 Mg.ha^{-1} for national scale. SOC variability could be explained by precipitation, aboveground biomass and land-use in local scale, and precipitation, altitude, and temperature for regional and national scale. The predicted SOC from fitted models ranged from 60 to 150 Mg.ha^{-1} at local scale, 60 to 130 Mg.ha^{-1} at regional scale, and 40 to 150 Mg.ha^{-1} at the national scale. This study improves the knowledge on the spatial distribution of SOC, by reducing the uncertainty related to up scaling process in Madagascar.

Session 4 Digital Soil Mapping
Chair: A-Xing Zhu

Predictive mapping of soil properties at high resolution by component wise gradient boosting from legacy data

Madlene Nussbaum[1], Andreas Papritz[1], Marielle Fraefel[2], Andri Baltensweiler[2], Lorenz Walthert[2], Armin Keller[3], Urs Grob[3] and Sanne Diek[4]

[1] Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland
[2] NABO, Agroscope, Zurich, Switzerland
[3] Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
[4] Remote Sensing Laboratories RSL, University of Zurich, Zurich, Switzerland

Accurate spatial information on soils is crucial for sustainable usage of the resource soil. Spatial planning, agriculture, forestry or natural hazards management need high resolution maps of potentials for particular soil functions (e. g. water storage, nutrient supply). Soil functions are derived from basic soil properties like soil organic matter, pH or soil texture. For many parts of Switzerland precise maps thereof are missing.

A wide range of statistical approaches (linear and additive models, external-drift kriging, Random Forest) was used for digital soil mapping in the past. If numerous environmental covariates (e. g. hyper-spectral remote sensing data or terrain attributes) are available the selection of the model with best predictive power is challenging. To handle these difficulties we used a gradient boosting approach that included categorical covariates, linear and smooth non-linear terms of continuous covariates. Besides numeric responses gradient boosting can predict binary or ordered categories (e. g. soil classification data) based on logistic or proportional odds models.

To explore the feasibility of the gradient boosting approach we mapped a wide range of basic soil properties for two agricultural study areas and one forested area in Switzerland. We predicted properties for top- and subsoil from legacy soil data on a fine-meshed grid (20 m mesh width). The gradient boosting approach could be applied to numerous responses with ease and successfully selected a small number of relevant covariates (10–30 out of 480).

Model performance – evaluated with independent validation data – depended on soil properties: Topsoil maps for pH, organic matter content or texture were more accurate compared to subsoil maps. For example, the spatial distribution of subsoil organic matter could not be captured due to very low subsoil organic matter content. For other soil properties (e. g. content of fragments > 2 mm) the model performance was independent from depth. Categorical information on waterlogging could be predicted with satisfactory accuracy. We compared the results to predictions derived from legacy soil maps. To account for the uncertainty we computed prediction intervals by a model-based bootstrap approach and validated externally the coverage of these intervals.

Mapping of soil moisture by deep learning based Cellular Automata

Xiao-Dong Song[1], Gan-Lin Zhang[1]*, Feng Liu[1], De-Cheng Li[1] and Yu-Guo Zhao[1]

[1]State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, 210008, China

Soil moisture content (SMC) is a key hydrological parameter in various application domains, and its spatio-temporal patterns are needed to address precise irrigation scheduling. The hybrid interaction of static and dynamic environmental parameters makes it particularly difficult to accurately and reliably model the distribution of SMC. In this research, a novel Macroscopic Cellular Automata (MCA) model was presented by integrating deep belief network (DBN) and applied to a prediction experiment over an irrigated corn field (22 square kilometers in the Zhangye artificial oasis, northwestern China). Recently deep learning has won numerous contests in machine learning and hence DBN, a breakthrough in deep learning, was trained to extract the transition functions for the simulation of the cell state changes. Static and dynamic predictors were prepared with regard to the complex hydrological processes. The widely used neural network, multi-layer perceptron (MLP) was utilized for comparison. The hybrid models were calibrated and validated on SMC data within four months, i.e. June to September 2012, which were automatically observed by a wireless sensor network. Compared with the MLP-MCA, DBN-MCA model led to a reduction in root mean squared error (RMSE) of 18%. Differences arose due to the propagating errors of variables, and difficulties of knowing soil properties and recording irrigation amount in practice. The sequential Gaussian simulation (sGs) was performed to assess the uncertainty of soil moisture patterns. Calculated with a threshold of SMC for each grid cell, local uncertainty of simulated results in the post processing suggested that the probability of SMC less than 25% will be in different areas at different time. Results showed that DBN-MCA model performs better than MLP-MCA and provides a powerful tool for predicting SMC in highly non-linear forms. Moreover, as modelling soil moisture using environmental information is gaining increasing popularity, DBN techniques could contribute to enhancing the calibration of MCA-based SMC models and hence provide an alternative approach for the SMC monitoring for irrigation systems based on canals.

Optimization of sampling configuration for digital soil mapping in a historical wine region, Hungary

Gábor Szatmári[1], Péter László[1], Katalin Takács[1], György Lukácsy[2], József Szabó[1] and László Pásztor[1]

[1] *Institute for Soil Science and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences*
[2] *Corvinus University of Budapest, Faculty of Horticulture, Department of Viticulture*

Tokaj is a historical region in Hungary for botritized dessert wine making. Recently the sustainable quality wine production in the region was targeted, which requires detailed soil surveys, digital soil mapping (DSM) procedures and uncertainty assessments using model-based (i.e. geostatistical) approaches. The objective of our work was to plan the sampling strategy considering predefined constraints for digital mapping of primary soil properties.

First of all a preliminary soil survey with a number of 200 observations has been carried out in 2014 to explore the relationship between the soil properties of interest and the auxiliary information (i.e. digital elevation model, geological map and satellite images) and to explore the spatial structures (i.e. variograms). In the second phase, the aim was to establish a sampling design especially for DSM purposes. This one can be considered as an optimization problem, where the spatial simulated annealing (SSA) was applied as an optimization algorithm. The regression kriging error variance (RKV) was used as a pre-quality measure, which requires the structure of the regression models and the variogram of the regression residuals. These were known from the preliminary survey. RKV's spatial average and RKV's spatial maximum were applied as optimization criterions, respectively. Along the optimization procedures a lot of constraints had to be considered, such as inaccessible areas for sampling and priority areas (where the mapping accuracy has to be higher). These have yielded a complex optimization problem, where a weighted optimization criterion was developed using the linear combination of RKVs and the weights of the priority areas (i.e. high, medium and low). Along the optimization procedures the sample size was systematically changed to yield a so-called "calibration curve", which was used to determine the sample size for the desired mapping accuracy. The final sampling configuration was evaluated using various statistical and point pattern analysis tools.

The aim of the recent paper is to present the challenges and the results of the second phase sampling strategy of the Tokaj wine region and the evaluation of the optimized sampling scheme.

Acknowledgement: Tokaj Kereskedőház Ltd. supported the project for the survey of the state of vineyards. Digital soil mapping was partly supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167).

Identifying soil variability using multi-year remote sensing for site-specific management

Yash Dang[1] and Ram Dalal[2]

[1] The University of Queensland, 203 Tor Street, Toowoomba Qld 4350, Australia

[2] Department of Science, IT and Innovation, 41 Boggo Road, Dutton Park, Qld 4001, Australia

Soils are highly variable, leading to variation in its potential productivity often over distances of only a few metres. A number of techniques/methodologies have been suggested to identify soil variability at paddock, farm and regional scales. It has been possible to identify sites that clearly manifest specific soil constraints at the site level or at local scale using grid soil sampling; however, this process is time-consuming and expensive. Yield mapping technologies collect high-resolution information, but they require some capital investment and, in our experience, can become an inconvenient distraction for growers at harvest. Multi-year spatial and temporal analysis of remotely sensed data can provide a rapid and accurate assessment of areas that are consistently low-yielding over several years, indicating the presence of at least one unknown soil constraint factor. We used historical mid-season normalized difference vegetation index (NDVI) generated from Landsat imagery at different wheat growth stages. The NDVI obtained at different wheat growth stages was compared to the actual yield map from a 74-ha paddock for wheat grown in 2007 (r^2 = 0.40-0.75) and 2010 (r^2 = 0.26-0.78). The NDVI at flowering best predicted the wheat grain yield. The regression function obtained was used to simulate wheat yield for rest of the years over 2000-13. We delineated three potential management classes (PMC) using k-means classification of the interpolated data of apparent electrical conductivity (EC_a), yield maps and multi-year remote sensing data. A minimum of four locations were selected within each PMC for soil sampling in 0-0.10 m then 0.20 m depth intervals to 1.5 m and analyzed for physico-chemical properties. The PMC identified by our analysis were: (i) a low-yielding class, LYC ($\bar{x}$=1.27 t/ha); (ii) a medium-yielding class, MYC ($\bar{x}$=1.99 t/ha); and, (iii) a high-yielding class, HYC ($\bar{x}$=3.36 t/ha). Soil Cl^- was significantly higher (>800 mg/kg) below 0.8 m depth in LYC and MYC compared with HYC. Exchangeable Mg percentage (EMgP) was significantly higher (>25%) in the LYC compared with MYC and HYC at all soil depths. The presence of an excessive EMgP can result in soil dispersion. The NO_3-N below 0.6 m depth was significantly higher in LYC and MYC compared with the HYC. In 2009, replicated strips were established in each PMC and 0 and 23 kg N/ha was applied across the field aligned in the direction of management operations while the rest of the field received 46 kg N/ha. Gypsum @2.5 t/ha was applied in LYC. No significant response to applied N was obtained in LYC; however, significant increases in wheat grain yield were obtained with increasing rates of N application in both MYC and HYC zones. Nitrogen requirement calculated using average wheat yield in different PMC, underlying soil NO_3-N and protein goal (11.5%), showed that LYC of the field (29 ha) had substantial unutilised NO_3-N in the soil profile from previous uniform N applications. The application of 46 kg N/ha results in a net wastage of 2.9 t urea as per farmer's uniform N management practice. Gypsum significantly increased wheat grain yield in the first and second wheat crop. The cummulative profit with gypsum application over two years was A$103/ha. The simulated yield mapping methodology offers an opportunity to identify within-field spatial variability using satellite imagery as a surrogate measure of biomass. However, the ability to successfully simulate crop yields at farm scale or regional scale requires wider evaluation across different soil types and climatic conditions.

An improved fuzzy logic slope-form system for predictive soil mapping of a landscape-scale area with strong relief conditions

Bui Le Vinh[1 2 3 4], Gerhard Clemens[2 3] and Karl Stahr[2]

[1]*Department of Land Management, Faculty of Land Management, Vietnam National University of Agriculture, Trau Quy, Gia Lam, Hanoi, Vietnam*
[2]*Institute of Soil Science and Land Evaluation (310), University of Hohenheim, 70593 Stuttgart, Germany*
[3]*The Uplands Program, Vietnamese-German Center, Technical University Hanoi, 1 Dai Co Viet, Hanoi, Vietnam*
[4]*International Center for Tropical Agriculture (CIAT), regional office for Asia, Institute of Agricultural Genetics, Km2 Pham Van Dong Street, Tu Liem, Hanoi, Vietnam*

This paper applies a fuzzy logic approach as a predictive soil mapping tool for a mountainous region of northwestern Vietnam with no soil information available. This approach uses Soil-Land Inference Model (SoLIM) software to study the relationship between soil and environment to predict spatial occurrences of soils and their quality. Purposive sampling strategy was used to obtain soil data from 110 profiles covering four major slope positions on a hillslope: crest, upper-, middle-, and foot slope. Influencing environmental parameters were identified and described from this sampling approach and are called descriptive knowledge. These parameters are parent material, slope inclincation, slope position, slope aspect, and elevation. A detailed system of 29 fuzzy slopeforms was constructed as another parameter for an assumption that slopeforms at this extreme relief area have certain influence on spatial distribution of soils and differences in soil quality Soil profiles were classified into reference soil groups (RSG) and subsoil unit according to WRB 2006. The occurrence of each of these classified RSGs and units is defined by distinctive combinations of the environmental parameters acquired during the conventional soil survey. The inference engine in SoLIM calculates and assigns a similarity value to a map pixel with based on the prescribed RSGs and subsoil units with values ranges from 0 (zero similarity) to 1.0 (maximum similarity). Similarity maps were produced for these RSGs and subsoil units and hardened to a RSG map and subsoil unit map, respectively. The soil quality map under this mapping approach was constructed in the same way. The validity of these products was checked with 50 field validation points. The RSG and subsoil unit maps showed good accuracy results of 76% and 72%, respectively. The results show that the fuzzy slopeforms in relation to the major slope positions and slope inclination do reveal occurrences of certain RSGs and units. The soil quality map shows a lower match with the validation points. The fuzzy slopeforms in this case do not show a significant correlation with the soil quality classes. Variations of soil quality depend more on slope inclination, slope positions, elevation. Land-use history was confirmed in several studies in the same are to be the major reason for soil fertility decline. However, it could not be quantified due to limitations in time and finance, which was probably the reason of having an average certainty result for the soil quality map.

Hierarchical approach of incorporating legacy information into the digital soil mapping process to provide soil property information at a scale where people make decisions

Phillip R. Owens[1], Jenette Ashtekar[1] , Minerva Dorantes[1] , Mayesse DaSilva[2] and Zamir Libohova[3]

[1] Purdue University, Department of Agronomy, 915 W. State Street, West Lafayette IN 47907
[2] Center for International Agriculture in the Tropics (CIAT), Cali, Colombia
[3] USDA-NRCS-NSSC, Federal Building, Room 152, 100 Centennial Mall North Lincoln, NE 68508-3866

Text Soil maps are basic infrastructure for nations. The variable function of soils controls agriculture production, flooding, landslides, road construction and water storage/recharge to aquifers. These are only a few of the many society-based needs for understanding this natural resource. There is a great need to understand the general function of soils; however, there is a greater need to understand soil function at the scale where people make management decisions. The task of creating maps at a resolution that highlights soil variability at a field scale is difficult using soil point data alone because the required input is rarely available. Fortunately, there is information in legacy maps and other sources that can build upon the accumulated knowledge of past generations of field scientist that provided boundary conditions for predictions. Utilizing maps of surface geomorphology describes the parent material information that can infer the soil texture, pH, base saturation, clay mineralogy and other properties that are necessary to predict. At the scale of a geomorphic surface, the parent material properties control the possibilities of soil properties. Within a scale of a field, topography is the greatest driver of soil variability because of water redistribution over geologic time. Water is the energy that drives the differentiation of similar parent material that results in variations of soil properties. This paper will present the hierarchal process used to map soils by utilizing regional legacy data to constrain the predictions so that finer resolution topographical information can identify the natural patterns that developed based on topographical differences. The patterns can be predicted and data can be tied to patterns to make soil property predictions. This paper will present examples from the USA, El Salvador and Kenya where this process is being utilized to map at the resolution of the digital elevation model. The process is being adopted by governments and industry to make decisions regarding the function of soil.

Knowledge discovery from samples using the partial dependence of random forest under fuzzy logic

Canying Zeng[1], A-Xing Zhu[1, 2, 3] and Lin Yang[2]

[1]*Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, 1 Wenyuan Road, Nanjing, Jiangsu 210023, China*
[2]*State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China*
[3]*Department of Geography, University of Wisconsin-Madison, Madison, USA*

The knowledge of relationships between soil and its environmental factors is very important for soil mapping. Numerous methods have been developed to discover soil-environment knowledge, such as, decision tree, artificial neural network, SoLIM (Soil Landscape Inference Model), and etc. In those methods, using fuzzy memberships to express the soil-landscape knowledge have several advantages: 1) it can better reflect the space gradient information and continuously of soil since fuzzy membership describes how similarity between a local soil and the typical case of the given soil type will change as environmental conditions change; 2) it is beneficial to accumulation and transplant knowledge. The fuzzy membership function is mainly constructed from experts, existing soil maps. However, most areas do not have soil experts or soil map, and the accuracy of knowledge extracted from soil map highly depends on the quality of the using map. In general, soil samples also imply the knowledge of relationships between soil types and their underlying environmental conditions. The excisting knowledge discovery method from soil samples requires the representative field points making its application restricted (Yang et al., 2013).

In this paper, we presented a method to construct fuzzy membership function from partial dependence of environmental variables generated using random forest (RF) based on soil samples. Partial dependence is the dependence of a variable on the class probability (when used for random forest classification) or response (when random forest used for regression) (Friedman, 2001). When applying RF on soil classification, one can obtain partial dependence curves for all variables for each soil classes. Partial dependence functions (curves) can measure the dependence of a certain soil type on an environment variable. The stronger the partial dependence of a value of some variable is, the higher probability of the soil existing in this value of the variable. Therefore, the partial dependence curve contains the knowledge of relationships between soil type and environmental conditions. The fuzzy membership deverived from partial dependences were finally used for soil type inference under the SoLIM framework to predict soil types in Heshan study area. In order to test how the training samples, especially the representativeness (typicality) of the training samples, impact the extracted knowledge and the predicted mapping results, two types of training samples were set. Case 1: the first type of training samples were all the representative samples. To compare with the previous study (Yang et al., 2013), which constructed fuzzy membership functions using descriptive knowledge generated from the representative samples in the same study area, we used the same training samples (Thirty three

representative soil samples) and the same fifty validation samples including regular sampling, subjective sampling, and transect sampling (Yang et al., 2013). Case 2: the second type of training samples was generated by splitting the total sample set randomly without assessment of the representativeness of the samples. This is because it is hard to know the representativeness of soil samples in most cases. In this study, we randomly split the 83 sample in to a training dataset and a validation dataset using the proportion of 2:1 (55 samples for training and 28 samples for validation). And to avoid the randomness of the splitting result, the split was conducted six times, so that six groups of different training and validation data sets were generated for this study area. Random forest was also adopted to compare with the knowledge based soil mapping method.

Results showed that the accuracy of predicted map using the knowledge derived from partial dependence based on representative samples were much higher than using random forest (78% vs 60%). The main reason is that the random forest method needs the full coverage in environmental conditions of the soil sample. But the representative samples only can represent the typical environmental conditions for each soil type. The soil type predictions in one pixel using random forest only have two alternatives within a certain environmental conditions: either yes or no. Thus when the samples are typical, it is difficult for random forest to capture the gradient information. However, the knowledge extracted from partial dependence in random forest is not only can capture the typical environmental conditions but also can express the graduation information. Therefore, it is more suitable by using the knowledge discovery method to predict soil type than random forest when the samples are typical. In addition, compared with the accuracy, 78% of case 1 in our study is slightly higher than the highest accuracy of Yang et al. (2013) with value of 76%. The results of case 2 showed that training samples impact the mapping accuracies greatly. In overall, the knowledge-based method generated more stable mapping accuracy results (ranges from 68% to 75% for group 1-5) than random forest(ranges from 58% to 82% for group 1-5) when training samples changed. If the accuracy of random forest was too low, such as group 6 with accruacy of 50%, the knowledge derived from the partial dependence will have derivation. In such condition, it is difficult to further improve the results.

It can be concluded that the knowledge discover method proposed in our study is more suitable than random forest when using the representative soil samples. The random forest could produce very high accuracy if the training samples covered most of the feature spaces which including the typical samples and the transitional samples. Otherwise, the knowledge discovery method based on the partial dependence can be an alternative method to improve it.

References

Friedman, J., 2001. Greedy function approximation: the gradient boosting machine. *Ann. of Stat*.

Yang, L., Zhu, A.X., Qi, F., Qin, C.Z., Li, B., Pei, T., 2013. An integrative hierarchical stepwise sampling strategy for spatial sampling and its application in digital soil mapping. *International Journal of Geographical Information Science* 27, 1-23.

Keynote: Flood-risk screening using detailed and massive terrain data

Lars Arge

Center for Massive Data Algorithmics, Department of Computer Science, Aarhus University

Improvements of mapping technologies such as LiDAR have resulted in a major increase in both the amount and quality of terrain data being acquired. Detailed terrain data is useful in a number of applications, such as for example when modelling how water flow on the terrain surface in order to estimate areas that are in risk of flooding during extreme rain events. However, the size of the data has also reveal scalability problems with existing terrain processing applications. Often these problems are a result of the underlying algorithms not taking the hierarchical structure of memory systems into account - especially the large difference in the access time of main memory and disk. In this presentation, we will discuss these problems further and describe how development and use of algorithms that minimize the number of accesses to disk, so-called I/O-efficient algorithms, has led to large practical runtime improvements in many flood-risk screening applications. The presentation will also include a demonstration of an online flood risk analysis tool based on these algorithms.

Session 5 Digital Soil Mapping of large areas

Chair: A-Xing Zhu

Mapping chemical and physical properties of soils at regional scale on the Swiss Plateau by robust external drift kriging from legacy soil data

Andreas Papritz[1], Madlene Nussbaum[1], Marielle Fraefel[2], Andri Baltensweiler[2], Lorenz Walthert[2], Armin Keller[3], Urs Grob[3] and Sanne Diek[4]

[1]Department of Environmental Systems Sciences, ETH Zurich, Zurich, Switzerland
[2]NABO, Agroscope, Zurich, Switzerland
[3]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
[4]Remote Sensing Laboratories RSL, University of Zurich, Zurich, Switzerland

Switzerland still lacks accurate, spatially highly resolved information on the properties of its soils in many parts of the country. Large-scale, polygon-based soil maps, sometimes of limited scope, are available for only about a third of the area used for intensive agriculture. Assessment of soil functionality for sustainable management of the soil resource and for better conservation of soils in spatial planning is seriously hampered by the lack of spatial soil information. We currently explore in the in a project of the Swiss National Research Programme "Sustainable Use of Soil as a Resource"(NRP 68) whether and how digital soil mapping can generate the required spatial soil information from legacy soil data and a comprehensive set of environmental covariates with sufficient accuracy. Apart from Random Forests and boosted geo-additive models robust external drift kriging is used for spatial statistical modeling. Unlike machine learning methods, the geostatistical approach exploits residual auto-correlation when computing spatial predictions, but parameter estimation and model building is clearly more cumbersome than with the former approaches. In the presentation, we shall present selected results of the geostatistical analyses, compare them with the result of the machine learning approaches and discuss the pros and cons of the various approaches.

The new digital soil map of Sweden –method and applications

Piikki K.[1], Söderström M.[1], Stadig H[2]. Sohlenius G.[3] and Rodhe, L.[3]

[1]*Swedish University of Agricultural Sciences (SLU), Skara, Sweden*
[2]*Hushållningssällskapet Skaraborg, Skara, Sweden*
[2]*Geological Survey of Sweden (SGU), Uppsala, Sweden*

The digital soil map of Sweden (DSMS) is a new public digital soil map that provides information on topsoil properties in farmland soil. The map resolution is 50 m × 50 m and covers virtually all arable land in Sweden (> 90%). The map layers produced to date describe the topsoil fractions of clay and sand. Secondary attributes calculated from these are the fraction of silt and the FAO soil texture classification. The calculated values have an uncertainty that varies in different regions and at different scales. For the entire map standard errors at the raster cell level are 5.6 % clay and 11.3 % sand. The accuracy is higher for field averages.

Although the original purpose of DSMS was to form a basis for improved calculations of national phosphorous loads to water bodies, it has gained interest from farmers and the private companies in the agricultural sector. The clay content map is by many farmers judged good enough to be used as decision support for e.g. variable seeding rate within fields. This is useful when the sowing conditions differ across a field. If a uniform rate is used, the farmer normally adjusts it to be high enough at the parts with high clay content, with the consequence that the canopy becomes too dense on sandier parts where the germination rate is higher. The effect can be increased risk for fungal infections and lodging. A variable seeding rate based on clay content and the farmer's own experience can remedy such problems.

The DSMS primary layers were derived from a combination of gamma radiation data (isotopes ^{232}Th and ^{40}K) from airborne scanning (SGU), a digital elevation model obtained by laser-scanning of the terrain (National Land Survey of Sweden) and a Quaternary Deposit map (SGU). The predictor data were recalculated to a 50 m × 50 m point grid. Prediction models for the fractions of clay and sand were parameterized using a national soil sample dataset (Swedish Board of Agriculture and the Swedish Environmental Protection Agency). The dataset included values on topsoil texture from ~15 000 positioned samples. The sampling design was a 1 km square grid with a small random displacement of the sampling locations. The fractions of clay and sand were determined by the sedimentation method. The prediction models (multivariate adaptive regression splines without interactions) were applied on the point grid and the method was validated at the sampling locations (~50 m^2 support) by a 10-fold segmented validation strategy.

The DSMS is a living map product that will be continuously updated as new data and improved algorithms become available. Additional layers can be added by applying pedotransfer functions to the primary layers. Our next challenges will be to make the map interactive (i.e. to develop the possibility to locally improve the information by use of own samples) and to find a functional method for mapping soil organic carbon at the extent and resolution of the DSMS.

Automated soil mapping based on Machine Learning: towards a soil data revolution

T. Hengl[1]

[1] ISRIC - World Soil Information P.O. Box 353, 6700 AJ Wageningen, The Netherlands

Automated soil mapping as a data-driven approach to soil mapping with little or no human interaction, commonly based on using optimal (where possible) statistical methods that employ relationships between target soil variables (sampled in the field and geolocated) and covariate layers, primarily coming from remote sensing data. With the rapid evolution of Machine Learning methods, it is becoming possible to automate most of the processing steps required to produce gridded maps from point data. Author will discuss especially use of tree-based ML techniques (random forest, gradient boosting and neural networks) for fitting models for numeric and factor type soil variables, and with the goal of complete soil mapping automation. The results show that ML techniques have a significant predictive potential for mapping various physical, chemical soil properties and classes – both locally and globally. Automating all processes of soil mapping will likely lead to a new soil data revolution: number of maps such system generates might start growing exponentially and so will their price drop exponentially. One limitation of automated soil mapping, however, is that the number of accurately georeferenced locations of reliable soil observations (particularly with analytical data) is often not sufficient to completely capture and describe all significant patterns of soil variation in an area. There may be too few sampled points and the exact location of recorded point data may not be well recorded. Data-driven soil mapping is field-data demanding and collecting field data can require significant expenditures of time, effort and money. Another limitation of the automated soil mapping is that there may be no obvious relationship between observed patterns of soil variation and the available environmental covariates. This may occur when a soil property of interest does strongly covary with some mappable environmental covariate (e.g. soil clay content with airborne radiometric data) but data for that environmental covariate are not available for an area. An effect of automating soil mapping via Machine Learning, because it is completely data driven, is that the next generation soil mappers / soil modelers, will have to put much more effort in optimizing sampling (or at least producing unbiased sampling), and much more effort in producing harmonized, consistent and complete training data sets. Data will become the heart of the system. ML will not put soil mappers out of work, but it might change the focus of the field from model building and comparison, to interpretation of results and optimization of computing.

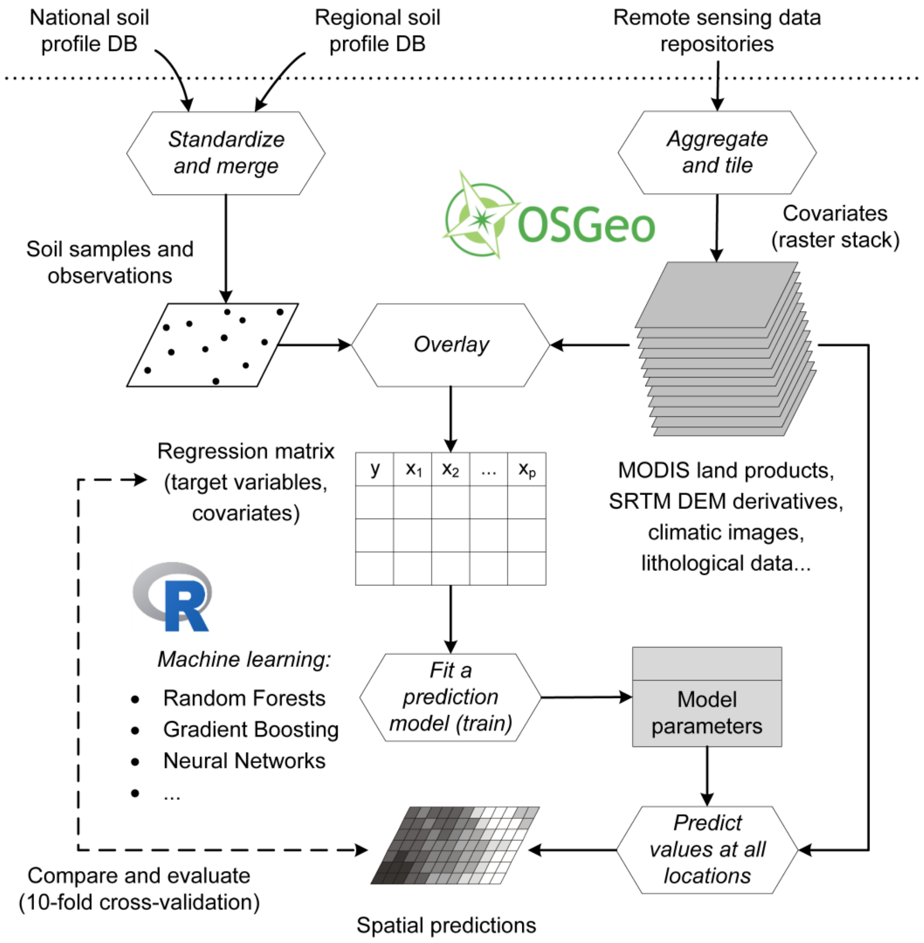


Figure: A generic framework for automate soil mapping used to produce SoilGrids (global predictions of soil properties and classes).

Using Remote Sensing and Ancillary Data Mapping Heavy Metals in Qatar Soils

Yi Peng[1], Rania Bou Kheir[1†], Kabindra Adhikari[3], Radosław Malinowski[1], Mette B. Greve[1], Maria Knadel[1], Basem Shomar[2] and Mogens H. Greve[1]

[1] *Aarhus University, Faculty of Science and Technology, Department of Agroecology, Blichers Allé 20, P.O. Box 50, DK-8830 Tjele, Denmark*
[2] *Qatar Environment & Energy Research Institute (QEERI), P.O. Box 5825, Doha, Qatar*
[3] *Department of Soil Science, FD Hole Soils Lab, University of Wisconsin−Madison, Madison, WI 53706, USA*

After decades of mining and industrialization in Qatar, it is important to estimate its impact on soil pollution with heavy metals. In the last ten years, the potential of using remote sensing (RS) techniques (aerial or satellite data) for mapping heavy metal in soils have received some attention. The study utilized 300 topsoil (0-30cm) samples, multi-spectral images (Landsat 8), spectral indices and environmental variables to model and map the spatial distribution of arsenic (As), chromium (Cr), nickel (Ni), copper (Cu), lead (Pb), and zinc (Zn) in Qatar soils. The prediction model used condition-based rules generated in the Cubist tool. All calibration models showed good predictive capabilities for the metals based on RMSE, R^2, RPD and RPIQ. Of all the prediction results, Cu had the highest R^2 and RPD value (0.74 and 1.88 mg kg^{-1}, respectively), followed by As > Pb > Cr > Zn > Ni. This study found that all the models only chose images from January and February as predictors, which indicates that images from these two months are important for soil heavy metal monitoring in arid soils, due to the climate and the vegetation cover during this season. Various RS spectral indices also had a high attribute usage for modeling the target elements The Tasseled Cap transformation and biophysical composition index (BCI) were the most frequently used indices. The Tasseled Cap transformation ranked in the top 10 predictors for As, Zn, Pb and Cu, while the BCI index ranked in the top 10 predictors for Cr, Ni, Zn, and Pb. Topsoil maps of the six heavy metals were generated The maps can be used to prioritize the choice of remediation measures and can be applied to other arid areas of similar environmental/socio-economic conditions and pollution causes.

Comparison of a soil texture map synthetized from GlobalSoilMap standard layers with a goal specifically compiled product

Annamária Laborczi[1], Gábor Szatmári[1], Katalin Takács[1] and László Pásztor[1]

[1] *Department of Environmental Informatics, Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences*

There are increasing demands nowadays on spatial soil information in order to support environmental related and land use management decisions. The demands often relate to physical soil properties, among which one of the most informative is particle size distribution. A few of them can be satisfied by the sand, silt, and clay fraction maps compiled according to the GlobalSoilMap (GSM) specifications.

Soil texture class information represents direct input of numerous agro-meteorological and hydrological models. Soil texture classes (e. g. according to USDA) can be unambiguously derived from the three fraction data, texture map can be compiled based on the proper fraction maps. In numerous cases, however, the required soil information refers to depth intervals different from those defined by GSM specifications. The model inputs frequently require maps representing soil features of 0-30 cm depth, which is covered by three consecutive depth intervals according to GSM specifications: 0-5 cm, 5-15 cm, 15-30 cm. Becoming GSM the most detailed freely available spatial soil data source, the common model users (e. g. meteorologists, agronomists, or hydrologists) would produce input map from (the weighted mean of) these three GSM layers. Having the basic soil data and the proper knowledge, a soil texture map targeting directly the 0-30 cm layer could be independently compiled. The two different approaches would not produce identical results. What extent the differences could be? We intended to investigate these issues.

In our work we compared Hungary's soil texture maps compiled using the same reference and auxiliary data and inference methods but for differing layer distribution. We produced the 0-30 cm clay, silt and sand map as well as the maps for the three GSM layers (0-5 cm, 5-15 cm, 15-30 cm). Maps of sand, silt and clay percentage were computed through regression kriging (RK) applying Additive Log-Ratio (alr) transformation. In addition to the Hungarian Soil Information and Monitoring System as our basic data, digital elevation model and its derived components, soil physical property maps, remotely sensed images, land use -, geological-, as well as meteorological data were applied as auxiliary elements. We thoroughly compared the synthetized and the directly computed maps and evaluated their differences in order to call attention to the possible risks.

Acknowledgement: Our work was supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167).

DOSoReMI.hu; results of a project for functional, DSM based renewal of Hungary's national soil spatial data infrastructure

László Pásztor[1], Annamária Laborczi[1], Katalin Takács[1], Gábor Szatmári[1], Gábor Illés[2], Endre Dobos[3], Zsófia Bakacsi[1] and József Szabó[1]

[1]Institute for Soil Science and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences
[2]National Agricultural Research and Innovation Centre, Forest Research Institute
[3]Department of Physical Geography and Environmental Sciences, University of Miskolc

Due to former soil surveys and mapping activities significant amount of soil information has accumulated in Hungary. In traditional soil mapping the creation of a new map was troublesome and laborious. As a consequence robust maps were elaborated and rather the demands were fitted to the available map products. Until recently spatial soil information demands have been serviced with the available datasets either in their actual form or after certain specific and often enforced, thematic and spatial inference. Considerable imperfection may occur in the accuracy and reliability of the map products, since there might be significant discrepancies between the available data and the expected information. The DOSoReMI.hu (Digital, Optimized, Soil Related Maps and Information in Hungary) project was started intentionally for the renewal of the national soil spatial infrastructure in Hungary. During our activities we have significantly extended the potential, how soil information requirements could be satisfied. Soil property, soil type as well as functional soil maps were targeted. The set of the applied digital soil mapping techniques has been gradually broadened incorporating and eventually integrating geostatistical, data mining and GIS tools.

Soil property maps have been compiled partly according to GSM.net specifications, partly by slightly or more strictly changing some of their predefined parameters (depth intervals, pixel size, property etc.) according to the specific demands on the final products. Data mining classification methods were applied for two purposes: (i) for the compilation of categorical (like genetic soil type and soil texture class) maps and (ii) for the understanding of specific soil-landscape models involved in existing soil maps, and for the post-formalization of survey/compilation rules. The elaborated primary maps were further processed, since even DOSoReMI.hu intended to take steps for the regionalization of higher level soil information (processes, functions, and services) involving crop models in the spatial modelling. The framework of DOSoReMI.hu also provides opportunity for the elaboration of goal specific soil maps, with the prescription of the parameters (thematic, resolution, accuracy, reliability etc.) characterizing the map product. As a result unique digital soil map products (in a more general meaning) were elaborated regionalizing specific soil (related) features, which were never mapped before, even nationally with high (~1 ha) spatial resolution. Our paper intends to present shortly the mapping processes themselves, the resulted new national maps and some conclusions drawn from the experiences.

Acknowledgement: Our work was supported by the Hungarian National Scientific Research Foundation (OTKA) under Grant K105167 and AGRARKLÍMA.2 VKSZ_12-1-2013-0034.

Session 6 Digital Soil Mapping of large areas

Chair: Lubos Burovka

Designing soil monitoring schemes for large areas based on digital soil mapping products

Alex. B. McBratney[1], Jaap De Gruijter[1] and Budiman Minasny[1]

[1]Centre for Carbon, Water & Food, Faculty of Agriculture and Environment, The University of Sydney, New South Wales 2006, Australia

We recognise that soil varies across space and time, and thus we need a proper sampling methodology to obtain reliable data that enable the estimation of some statistical parameter or spatial predictions of soil properties over an area and time periods. Sampling is constrained by the financial and available resources, thus an efficient sampling strategy is sought for applications in soil survey for mapping, and for establishing sites for monitoring networks.

Digital soil maps have been produced at continental and regional extents. These maps of soil properties along with their uncertainty can be used to establish a network for soil monitoring sites.

In this article, we used of a design-based sampling methodology, called Ospats which optimised spatial stratification and allocation for stratified random sampling of points. Ospats uses a grid of points from digital soil maps with uncertain predictions of the target variable. We demonstrate its application for establishing a network of monitoring sites in New South Wales, Australia with an area of 809,000 km2 based on a GlobalSoilMap product.

This soil monitoring network will have the ability to identify areas which are approaching a critical limit in soil conditions. These critical limits may be considered tipping points or resilience thresholds and as such the network acts as an early warning system. For example, the critical limit corresponds to an organic carbon content where soil physical degradation will occur resulting in destruction of soil aggregates, and increasing susceptibility to crusting and erosion.

Legacy soil survey data mining for digital soil mapping in Prince Edward Island (PEI) province in Canada

Xiaoyuan Geng[1], Juanxia He[1], Yefang Jiang[2] and Bert VandenBygaart[1]

[1] Science and Technology Branch, Agriculture & Agri-Food Canada, ON, Canada

[2] Science and Technology Branch, Agriculture & Agri-Food Canada, PEI, Canada

When field soil point or pedon data is lacking, the extracted or mined information from legacy soil surveys has been one of the key knowledge sources for machine learning-based digital soil mapping (DSM). In Canada where nationwide soil point data is sparse, detailed soil surveys from 1:10,000 to 1:20,000 exist in some cultivated regions. These detailed soil surveys can be used to extract point soil information or knowledge for DSM. For example, the 1:20,000 detailed soil survey data in PEI contains many single soil component polygons. In this study, the detailed soil survey of PEI is used for soil information or knowledge mining. With the mined soil knowledge, this study is aimed to investigate the validity of applying inference logicsgenerated for the whole area of interest to one subset or tile of the area at a time, and to investigate effective methods of legacy soil survey data mining for machine learning-based DSM.

In this study, it is assumed that any location within each of the single component polygons of the soil survey can be used to represent a spatial location of the associated soil component or type for that polygon. For DSM of a 50m grid size using Random Forest inference algorithms in PEI, stratified random sampling with minimum 50m spacing was conducted. The sampled soil points with associated soil types were further used to extract associated covariate values. The covariates included surficial geological material map, multi-scaled landscape derivatives including elevation difference ratio, slope gradient, plenary and profile curvatures, and relative hillslope position. Along with the selected covariate values 70% of the sampled points were used to generate classification logics, the remaining 30% of the data were retained for validation use. The covariate importance was ranked when all the collected covariate data layers were used. Among the selected covariate data layers, the surficial geological materials at 50m resolution was ranked as the most important covariate followed by elevation difference ratio at 50m resolution, slope at 150m and slope at 250m resolution.

All test runs applying inference logics of the study area to various subsets of the area showed above 99% consistency. So when computational capacity becomes a constraint, tile-based data structure and computing design can be used. The prediction accuracy using Random Forest algorithms was around 22%. This partially indicates that the accuracy of the mined point soil data was low in the first place. There is a need to test the hypothesis that any geospatially stratified location within each of those single component polygons of the soil survey can be used to represent a spatial location of the associated soil component or type of that polygon. The geospatial stratification logics will be constructed based on the association between soil development and the geospatial covariate. Further study is being conducted to test this hypothesis.

Applications of DSM for erosion modelling with uncertainty propagation

Laura Poggio[1] and Alessandro Gimona[1]

[1]*The James Hutton Institute, Craigiebuckler, Aberdeen (Scotland, UK)*

Soils play a crucial role in the ecosystem functioning and the modelling of their contributions is fundamental especially in the context of climate change. In this work we used continuous 3D soil information derived from digital soil mapping (DSM) approaches to map sediment erosion and deposition patterns due to rainfall. The test area covers the whole of mainland Scotland, excluding the Northern Islands. The topsoil information from over 7,800 profiles was interpolated using a hybrid Generalised Additive Models method for a range of soil properties such as organic matter, texture, soil depth and peat information (presence, degradation, depth). The same method was used to interpolate climatic data and management information. Remote sensing data were integrated in the process and land use data included. The uncertainty was accounted and propagated across the whole process. The Scottish test case is particularly important to highlight the differences in roles between mineral and organic soils. Therefore the assessment was adapted to organic and mineral soils. The results and intermediate steps were compared with available continental scale results, with lower erosion raters modelled using national data. The results show the importance of the use of DSM approaches for modeling soils and ecosystem functions and assess uncertainty propagation.

A Systematic Approach to Building a Sustainable Digital Soil Map

Minerva Dorantes[1], Phillip R. Owens[1], Jenette Ashtekar[1], Mayesse DaSilva[2] and Zamir Libohova[3]

[1]*Purdue University, Department of Agronomy, 915 W. State Street, West Lafayette IN 47907*
[2]*Center for International Agriculture in the Tropics (CIAT), Cali, Colombia*
[3]*USDA-NRCS-NSSC, Federal Building, Room 152, 100 Centennial Mall North Lincoln, NE 68508-3866*

Understanding soil variability at the field level is important for making decisions about crop management, water conservation, road construction, and for natural hazard mitigation among other things. This information is especially valuable in regions of the world along the equator that are strongly affected by climate change and suffer from prolonged drought in conjunction with excessive rainfall that exacerbates soil erosion. In these areas, access to high-resolution soil information is limited or non-existent and soil point data alone cannot capture the variability at the resolution necessary. By combining spatial legacy data which represents the regional variability in soil processes with local soil and terrain analysis, a functional soil property map can be produced. Geomorphic maps, geologic surveys, soil surveys, climate data, and regional topography are used to identify distinct soil environments in which the processes that govern soil variability are similar. Utilizing the derivatives of the digital elevation model and landform classification, each unique soil environment is further stratified by the local topography. This process results in a new soil class map where each class represents a set of unique soil forming factors. The class map is reviewed by local soil scientists and adjusted through field observations. Soil property values are interpolated within each class and thus predictions are made that follow the patterns of soil variability. The class and property maps can easily be updated as more legacy data becomes available or new soil point data is collected; however, they work well with limited point data. A project is underway to build local capacity in digital soil mapping for scientists in Southern Mexico, El Salvador, Guatemala, Nicaragua, and Honduras, through hands-on training to develop country-wide soil property maps. This paper will outline the methods used to prepare a soil property map using this full-spectrum approach and present results for the pilot area in El Salvador as well as describe how to expand the model to the rest of the country.

Managing water by managing soils: The importance of soil information for the sustainable development of Central America

Axel Schmidt[1], Phillip R. Owens[2], Zamir Libohova[3], Jenette Ashtekar[2] and Minerva Dorantes[2]

[1]*Catholic Relief Services (CRS), Mesoamerican Agriculture, Soils and Water Program, La Molina, Lima, Perú*
[2]*Purdue University, Department of Agronomy, 915 W. State Street, West Lafayette IN 47907, USA*
[3]*USDA-NRCS-NSSC, Federal Building, Room 152, 100 Centennial Mall North Lincoln, NE 68508-3866, USA*

Climate change is impacting the amount and the distribution of precipitation worldwide. Soils, one of the biggest terrestrial water reservoirs, and their management play a crucial role on water resource management especially within rain fed agricultural systems for mitigating the impacts of climate change. Managing water by managing soils to maximize water productivity in rain fed agricultural systems is challenging, particularly for developing countries facing poverty, poor infrastructure and ongoing land degradation. In Central America these problems are exacerbated and becoming more challenging because 75% of the land resources are already affected by degradation. The majority of decisions related to crop & livestock production and soil-water management, especially in rain fed systems, are being made at the small farm level that cumulatively impact entire landscapes and regions. Therefore, soil information that is functional, simultaneously location specific and scalable, is needed for improved and informed decision making at different levels (farm to regional). Soil information needs to be gathered in relatively short time, at low cost, be easily accessible and easy to use and interpret. Low cost monitoring tools are also required to assess soil health over time across larger landscapes in order to mitigate negative impacts on soil functions. Such detailed and timely information will support and influence informed decisions on technical assistance, development programs & projects, public and private investments, and international cooperation for poverty reduction, disaster risk reduction and human resource development. The emerging of high resolution digital data available worldwide at relatively lower costs combined with new technologies may offer a promise to provide such detailed soil information that is dynamic and continuously up to date. Ultimately, the goal is to make soil science relevant by effectively bridging scientific knowledge with natural resource management decisions in developing nations.

GlobalSoilMap developments in France

V.L. Mulder[1 2*], M. Lacoste[3], A.C. Richer-de-Forges[1] and M.P. Martin D. Arrouays[1]

[1]INRA, Unité Infosol, US 1106, CS 40001, Ardon, 45075, Orléans cedex 2, France
[2]Department of Geoscience, Environment & Society, Université Libre de Bruxelles, Brussels, Belgium
[3]INRA, UR 0272 Science du sol, Centre de recherche d'Orléans, CS 40001 Ardon, 45075 Orléans cedex 2, France

This work presents some *GlobalSoilMap* (GSM) products for France. We developed procedures for mapping the primary soil properties (clay, silt, sand, coarse elements, pH, soil organic carbon (SOC), cation exchange capacity (CEC) and soil depth) using a data-mining technique and a straightforward method for estimating the 90% confidence intervals (CIs).

The most accurate models were obtained for pH, sand and silt. Next, CEC, clay and SOC were found reasonably accurate in their prediction while coarse elements and soil depth where the least accurate of all models. Overall, all models were considered robust; important indicators for this were 1) the small difference in model diagnostics between the calibration and cross-validation set, 2) the unbiased mean predictions, 3) the reduced spatial structure, as observed from the prediction residuals and 4) the similar performance compared to other developed *GlobalSoilMap* products. Nevertheless, the confidence intervals (CIs) were rather wide for all soil properties. The median predictions became less reliable with increasing depth, as indicated by the increase of CIs with depth. In addition, model accuracy and the corresponding CIs varied depending on the soil variable of interest, soil depth and geographic location. These findings indicated that the CIs are as informative as the model diagnostics.

For SOC we compared our results with other global estimates and came to the conclusion that the French national product is more reliable than more global ones.

Concluding, the presented work resulted in reasonable accurate predictions for the majority of the soil properties. End users can employ the products for different purposes, as was demonstrated with some practical examples. The mapping routine is flexible for cloud-computing and provides ample opportunity to be further developed when desired by its users. This allows regional and international GSM partners with fewer resources to develop their own products or, otherwise, to improve the current routine and work together towards a robust high-resolution digital soil map of the world.

Session 7 Digital Soil Mapping
Chair: Gerard B.M. Heuvelink

Predicting soil depth using survival analysis models

Quentin Styc[1] and Philippe Lagacherie[2]

[1] INRA, UMR LISAH, Montpellier, France

Soil depth is a very important soil property for agronomists and environmental scientists and has been included as such in the GlobalSoilMap specifications (Arrouays et al, 2014). The Digital Soil Mapping applications that deal with soil depth are all confronted with the problem of the “right-censored” input soil data. This means that a significant proportion of the observed sites that serve as input data for Digital Soil Mapping models are characterized by a maximal depth of observation that is smaller than the real soil depth. The information on soil depth at such sites is not an exact value but an inequality (soil depth ⩾ site observation depth).

Such situations are frequently encountered in survival analysis that aim to predict the expected duration of time until one or more events happen, such as death in biological organisms and failure in mechanical systems. This very active branch of statistics has produced a lot of statistical inference models for dealing with right-censored data. By doing an analogy between time at “depth” or “failure” and soil depth (“depth at soil end”), such models could be applied to the digital mapping of soil depth.

In this paper, we apply two survival analysis models, namely Cox Regression (Andersen and Gill, 1982) and Random Survival Forest (Ishwaran et al, 2008) for predicting soil depths in two study areas of southern France with contrasted extents and soil observation densities (Languedoc Roussillon Region,and Peyne Watershed). The soil depth prediction results provided by these two models are evaluated over independent validation sets and compared with classical DSM models using as input i) the sites with a known exact soil depth values only or ii) the whole set of sites with the replacement of the censored soil depth by a « common sense » estimation.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes, a large sample study. Annals of Statistics 10, 1100-1120.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca Santos, M.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.-L., 2014. Chapter Three – GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties, in: Advances in Agronomy. pp. 93–134.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. Ann. Appl. Stat. 2, 841–860

An error budget for digital soil mapping using proximally sensed EM induction and remotely sensed γ-ray spectrometer data

Jingyi Huang[1], Thomas F.A. Bishop[2] and John Triantafilis[1*]

[1] School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney, NSW 2052, Australia

[2] Faculty of Agriculture and Environment, The University of Sydney, Sydney, NSW 2006, Australia

The cation exchange capacity (CEC) of soil is widely used for agricultural assessment because it is a measure of fertility and an indicator of structural stability. However, measurement of CEC is time consuming. Whilst geostatistical methods have been used, a large number of samples must be collected. Using pedometric methods and specifically coupling easy-to-measure ancillary data with CEC have improved efficiency in spatial prediction. The evaluation of mapping uncertainty has not been considered, however. In this study, we use an error budget procedure to quantify the relative contributions that model, input and covariate error make to prediction error of a digital map of CEC using gamma-ray (γ-ray) spectrometry and apparent electrical conductivity (EC_a) data. The error budget uses empirical best linear unbiased prediction (E-BLUP) and conditional simulation to produce numerous realizations of the data and their underlying errors. Linear mixed models (LMM) estimated by residual maximum likelihood (REML) is used to create the prediction models. Results show that the combined error of model (5.07 cmol(+)/kg) and input error (12.88 cmol(+)/kg) is approximately 12.93 cmol(+)/kg, which is twice as large as the standard deviation of CEC (6.8 cmol(+)/kg). The individual covariate errors caused by the γ-ray (9.64 cmol(+)/kg) and EM error (8.55 cmol(+)/kg) are also large. To overcome the former, pre-processing techniques to improve the quality of the γ-ray data could be considered. In terms of the EM error, this could be reduced by the use of a smaller sampling interval and in particular near the edges of the study area and also at Pedoderm boundaries.

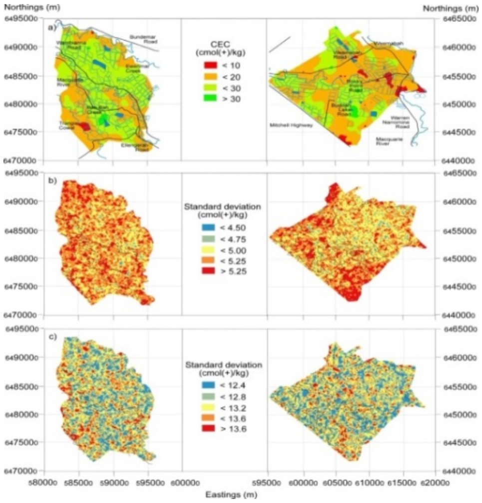


Figure 1 Spatial distributions of a) predicted CEC using E-BLUP; and spatial distribution of standard deviation of predicted CEC due to b) model error and c) input error across Trangie and Warren study areas. Note: Boundaries of soil pedoderm components are marked as black lines.

Can citizen science be used to assist Digital Soil Mapping?

David G. ROSSITER[1,2], Liu JING[3], Steve CARLISLE[4] and A-Xing ZHU[2,3]

[1] *Department of Crop & Soil Sciences, Cornell University, Ithaca NY 14850 USA*
[2] *School of Geography, Nanjing Normal University, No. 1 Wenyuan Road, Xianlin University District, Nanjing 210023, China*
[3] *Department of Geography, University of Wisconsin--Madison, 550 North Park Street, Madison, Wisconsin 53706 USA*
[4] *Aurora NY 13026 USA*

The essential element of citizen science is the participation of non-specialists in scientific research. The citizen acts as an observer or experimenter within structures established by a project run by professional scientists. The recent explosion in projects is due to the development of enabling technology, exemplified by the spatially-enabled "smart" phone with mapping applications and its supporting networks including the GPS system. Citizen science projects have three purposes: (1) to amplify scientific research; (2) to build citizen support for, and understanding of, science; (3) to help set the research agenda. Current initiatives in citizen soil science include the OPAL Soil and Earthworm Survey, GLOBE, and mySoil, none of which are aimed at assisting digital soil mapping. We present a framework for citizen science initiatives to support or enhance digital soil mapping for countries with and without well-organized extension and advisory services and existing soil surveys. We classify the types of citizens who might be motivated to contribute to such initiatives, and review their possible motivations and contributions, including tacit knowledge, opportunistic or protocol-guided new information, information from precision agriculture, and physical samples submitted for analysis. The primary beneficiary of such projects would be the professional mapper using digital information to produce or enhance maps of soil properties or types. The secondary beneficiary would be the citizen scientist, who would benefit from an enhanced map, and may be better able to participate in policy debates related to the soil resource. In addition, participation would enhance the connectivity between the soil resource and the citizen.

Adaptation of public digital soil maps for practical use in agriculture

Söderström M[1,2], Piikki K.[1,2] and Cordingley J.[3]

[1]*International Center for Tropical Agriculture (CIAT), Nairobi, Kenya*
[2]*Swedish University of Agricultural Sciences (SLU), Skara, Sweden*
[3]*Crop Nutrition Laboratory Services Ltd., Nairobi, Kenya*

There is currently a strong –and welcome– trend to make digital soil maps public. One example is the AfSoilGrids250m[1], a soil property map covering the African continent. We have evaluated the suitability of this map for use at local scale (small farms 0.5-1 ha) and regional scale (Sectors ~50 km^2 size) in Rwanda, and examined how this kind of geographic database, describing general patterns, can be adapted for use at a more detailed level. Comparison with 900 soil samples analyzed for soil organic carbon (SOC) showed a clear discrepancy between the values of the AfsoilGrids250 data and the observed values, also at the sector level. The resolution of AfSoilGrids250m is $250 \times 250\ \text{m}^2$ and the present result confirms that the cell size of a raster map does not reflect its uncertainty. In order to promote accurate use (or rather prevent inadvertent misuse) of published soil data, the digital soil mapping community must help users assess whether the maps are appropriate for their use. We recommend that local soil data is used to assess if such a map can be used for a specific application (e.g. agricultural management support), if not, it may be possible to improve it by e.g. regression kriging. In this study, the mean absolute error for sector averages of SOC in Rwanda was reduced from 11.3 g C kg^{-1} (if only the continental dataset was used) to 4.5 g C kg^{-1} when as few as 100 national soil observations were used to locally to improve the continental dataset (maps shown in Figure 1). We suggest further studies on approaches for local improvement of global and continental datasets and call for innovative ideas on how map uncertainties can be made accessible and understandable to general users.

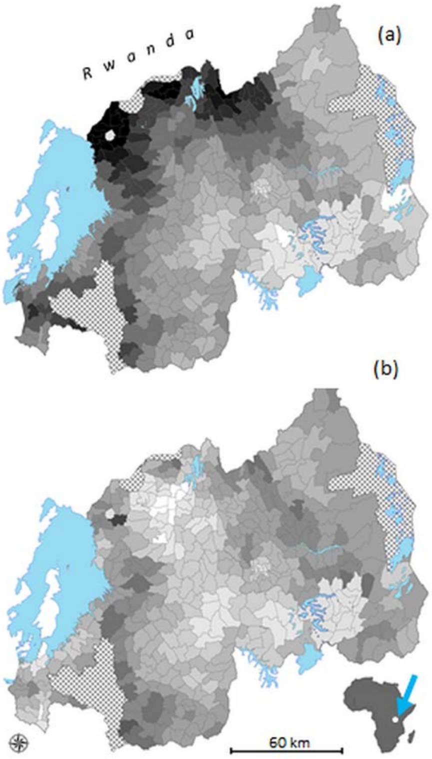


Figure 1. *a) Setcor average values of topsoil organic carbom (SOC) computed from the AfSoilGrids250m and b) Same map locally adapted by 100 soil samples. Dark color = high SOC. Light color = low SOC. Modified from Söderström et al. (2016)*[2].

---

[1] Hengl T. et al. 2015 *PLoS ONE* 10: e0125814.

[2] Söderström et al 2016. *South African Journal of Plant and Soil*. In press.

Ensemble of topsoil texture predictions for Region Centre (France)

Román Dobarco, M.[1], Lagacherie, P.[2], Ciampalini, R.[3], Arrouays, D.[1] and Saby, N.P.A.[1]

[1] *Unité Infosol, US1106, INRA, Orléans, France*
[2] *UMR LISAH, INRA Montpellier, France*
[3] *School of Earth and Ocean Sciences, Cardiff University, United Kingdom*

The expansion of digital soil mapping (DSM) in the world has led to the production of multiple maps of soil properties at regional, national, and global level. In the case of France, soil profile and horizon data used to calibrate the models often proceed from the national programs for spatial inventory and soil monitoring carried out by the association of scientific interest in soil (Groupement d'intérêt scientifique Sol, Gis Sol), but data from different programs are rarely used together. Several spatial predictions are then available for the same geographical areas and soil properties but with different qualities, which normally depend on the spatial resolution and type of support of the reference data, and the resolution of the predictor variables. Moreover, regional and national maps can be combined with maps at global or continental scale produced by international DSM initiatives. The objective of this study was to explore the possibilities for combining spatial predictions fitted with data from different Gis Sol programs. In particular, the aim was to improve the accuracy of predictions for topsoil texture in Region Centre (France). The three primary maps for topsoil texture were respectively: i) a regional texture map which follows the GlobalSoilMap specifications and used 2487 soil profiles from the French soil database (IGCS) in a regression-cokriging approach (Ciampalini et al., 2014), ii) a regional topsoil texture map that resulted from the disaggregation of areal data from the French soil test database (BDAT) applying regression and area-to-point cokriging (Román Dobarco et al., *in press*), iii) a topsoil texture map elaborated with regression and the LUCAS topsoil survey for the extent of Europe (Ballabio et al., 2016). The methods used for averaging the predictions were the Granger-Ramanathan (G-R) and the Bates-Granger (B-G), or variance weighed. The averaging coefficients for the G-R model were fitted with independent georeferenced observations from the BDAT. Data from the systematic random grid (16 km x 16 km) of the French soil monitoring network was used for independent validation. After applying these averaging methods to our case of study we discussed some of their limitations: i) the G-R method provides unbiased predictions relative to the reference values, therefore the calibration data should be a representative and unbiased sample of the study area's texture, ii) although it is becoming a more widespread practice, uncertainty estimates are not always included in the outputs of the DSM, hindering the application of the B-G method. Model ensemble methods may be an inexpensive and relatively easy way to improve prediction accuracy, but more so, to join efforts from different research groups and combine soil datasets requiring different treatment of the data.

Mapping of skidding track soil compaction and displacement from high resolution LIDAR-based DEM's in broadleaved forest on poorly drained soils

Ingeborg Callesen[1], Stinna Filsø Susgaard[1], Jesper Riis Christiansen[1] and Lene Fischer[1]

[1] *Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark*

Tree skidding and forwarding operations with machinery in forests traditionally take place in the winter and early spring when poorly drained soils are frequently near water saturation due to the winter precipitation surplus in Denmark. In recent years, also felling has been mechanized in middle-aged and older broadleaved stands causing traffic in such stands to increase. Soil deformation and rutting from machine operations increase with soil water content and especially so on loamy and loamy clay soils. Poorly drained soils with a loamy or loamy clay texture class on flat terrain are predominant in SE Denmark with parent materials that can be e.g. lacustrine deposits or lodgement till. Groundwater influenced soils also occur in terrain depressions in e.g. kame and kettle topography (landscape formed by dead-ice) near marginal moraines. Although the planning of operations may take soil wetness into consideration, machine operations on wet soils often cannot be avoided and should be monitored. The soil displacement on tracks can be used for mapping rutted and compacted soils with reduced hydraulic conductivity. New high resolution (40 cm) LIDAR-based DEMs were used to identify skid tracks in 24 forest stands and quantify the length and area per hectare. The benefits of monitoring compacted and rutted areas in forests in relation to the ecological functioning e.g. tree growth and non-CO_2 greenhouse gas emissions are discussed in the paper. So far, the mapping procedure of tracks cannot be automated.

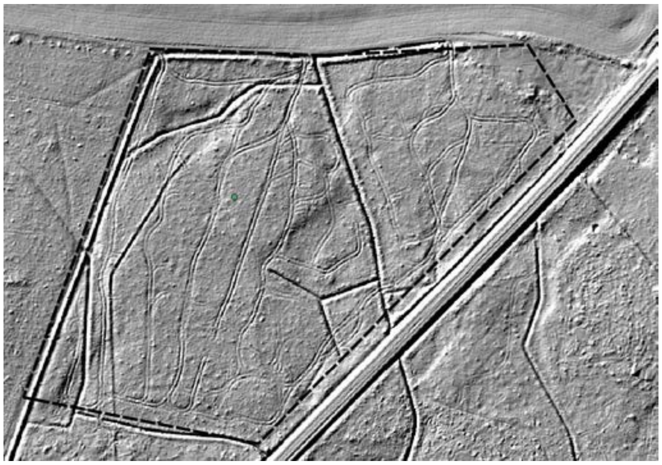


Figure 1 Example of hill-shade visualization of Lidar DEM with 40 cm resolution where skid tracks can be seen as parallel lines within a 90 year old oak stand.

Scope to map soil management units at the district level from remotely sensed γ-ray spectrometry and proximal sensed EM induction data

Y. Z. Jing[1], J. Huang[1], R. Banks[2] and J.Triantafilis[1,*]

[1] *School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney, NSW 2052, Australia*
[2] *SoilFutures Consulting Pty Ltd, Gunnedah, NSW 2380, Australia*

It is important for farmers to understand how their individual fields differ in soil type and how similar their soil is to that which surrounds their farm. The traditional method involves classifying soil into pre-existing classes using morphological observations. To accelerate the process, easier and cheaper to measure ancillary variables are used to add value. However, this process of digital soil mapping (DSM) is still affected by the soil classifications used to identify soil types. In this study, we use a different DSM approach where ancillary data act as surrogates for soil morphological data with soil types identified by numerical clustering of remote and proximal sensed data collected across a farming district near Gunnedah, Australia. The remote data was obtained from an air-borne gamma-ray (γ-ray) spectrometry survey; potassium (K), thorium (Th), uranium (U) and total counts (TC). Proximal sensed data was collected using an EM38 (i.e. EM38h and EM38v). Using fuzzy *k*-means (FKM) we cluster the data, with the fuzzy performance index (FPI) and normalised classification entropy (NCE) indicating that when using the Diagonal metric and a fuzziness exponent (ϕ) of 1.6 a local minima exists when $k = 5$. Using measured topsoil (0–0.30m) and subsoil (0.9–1.2 m) physical (e.g., clay) and chemical (e.g., CEC) properties we find that $k = 5$ was optimal given mean squared prediction error (i.e., $\sigma^2_{p,C}$) was a minimum for as compared to $k = 2$–8. The results were consistent with major soil landscape units identified using a traditional approach. The DSM was unsuccessful in identifying smaller units; however the inclusion of elevation data might be better able to discern these owing to the resistant nature of the geology and higher elevations which they occupy. The approach also highlighted subtle differences in physical and chemical properties in the more agriculturally productive area. This research has implications for providing fast, accurate and meaningful DSM for farmers, especially on large districts scales where traditional methods prove to be restrictive in terms of time and cost.

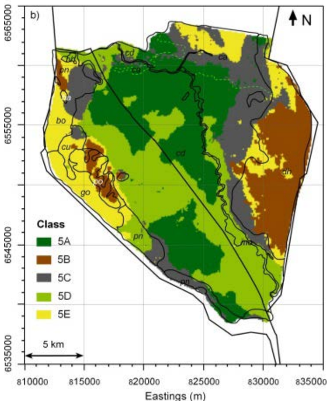


Figure 1 Spatial distribution of fuzzy k-means (FKM) classes (*k*) when $k = 5$ as determined using Diagonal metric with $\phi = 1.6$.□

Digital land resource mapping to address information and capacity shortages in developing countries

AJ Ringrose-Voase[1], M Thomas[2], GJ Grealish[2], MTF Wong[3], A Mercado[4], GP Nilo[5], MR Glover[1] and TI Dowling[1]

[1] CSIRO, GPO Box 1700, Canberra ACT 2601 Australia, [2] CSIRO, Adelaide, Australia, [3] CSIRO, Perth, Australia, [4] ICRAF, Claveria, Philippines, [5] Bureau of Soil and Water Management, Quezon City, Philippines

Like many developing countries, the Philippines has a shortage of land resource information at sufficiently fine scales for land use planning, especially in more marginal areas such as in upland catchments, where land use pressures have caused degradation. The country is also short in capacity to acquire such information with a declining number of soil surveyors and limited soil analytical laboratories. The information is required for evidence-based land use planning to allow agricultural intensificiation or diversification and improve farmer livelihoods, whilst maintaining the resource base. At the same time planning is required to protect catchment functions, such as provision of water, mitigation of flash flooding and prevention of landslides. Such information also allows extension services to tailor advice to the soil problems found at different locations. Digital land resource mapping (DLRM) can address these shortages because it can make more efficient use of the limited soil survey resources by devolving many tasks to local, non-expert teams. A project in the Cabulig catchment of Northern Mindanao (220 km²) has developed operational protocols for a DLRM framework based on four 'pillars'.

A statistically-based sampling strategy not only ensures unbiassed coverage of the range of landscape positions, but also removes the need for expert judgement in site selection in the field. Instead seed sites were selected from landform classes derived from a 20 m DEM using a stratified random approach and located by GPS. At each seed site a toposequence was sampled using additional sites positioned along random transects up to 500 m away. This clustering made the sampling logistics more efficient. 47 transects were sampled with a total of 236 sites over a 7 week field campaign.

A simplified site protocol concentrated on sample collection rather than description. Samples for mid-infrared spectroscopy (MIR) were collected from fixed depths at all sites. Bag samples were only collected at seed sites. The site and profile description form was simplified by including only features relevant in the catchment. This decreased the time required at each site and removed the need for an expert pedologist in the field. Quality control can be carried out later by experts using 'chip tray' samples and photography.

Rapid soil analysis by MIR was used to estimate soil attributes for all layers at every site after developing local calibrations using conventional laboratory analysis of the bag samples. Direct estimation of such attributes at all sites eliminates the need to infer them from soil classification. Finally, a range of soil attributes was mapped using statistical spatial prediction by regression tree modelling with Cubist. Accompanying estimates of uncertainty alerts decision-makers to areas of poor quality information. Overall, this framework can enable more efficient use of scarce pedological expertise and laboratory facilities as commonly experienced in developing countries. An added benefit is that the local teams acquire soil literacy and can help with on-going interpretation and application of the survey results.

Session 8 Digital Soil Mapping
Chair: Tom Hengl

Combining uncertainties from the feature domain and spatial domain for digital soil mapping: where to sample more?

A-Xing Zhu[1,2,3], Yan Li[4] and Zhou Shi[4]

[1] Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA
[2]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China
[3] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.
[4] Institute of Land Science and Property Management, School of Public Affairs, Zhejiang University, Hangzhou 310058, China

This paper presents an uncertainty directed sampling method for designing additional samples to complement the existing soil samples for digital soil mapping by combining the uncertainty from the feature domain (the domain of relationships with environmental covariates) and that from the spatial domain (domain of spatial autocorrelation). The method consists of three steps: 1) Selection of ranked list of additional sample locations based on uncertainty from the feature domain using a method called 'individual predictive soil mapping' (iPSM) ; 2) Selection of ranked list of additional sample locations based on uncertainty from the spatial domain using ordinary kriging; 3)Determination of a final ranked list based on both uncertainties by merging the ranked lists from 1) and 2). To evaluate the effectiveness of this method, the three respective lists were used in mapping soil organic matter (SOM) in a 299.14km^2 study area near by Fuyang city in northwest region of Zhejiang Province, China. The respective mapping accuracies from these lists were calculated and used to assess the effectiveness of the method. The case study showed that compared with additional sampling scheme based on the uncertainty either only from feature domain or only from spatial domain, the root-mean-squared-error (RMSE) of the additional sampling scheme using the final list based on both uncertainties were the smallest, ranging from 0.429 to 0.540 and the agreement coefficient (AC) were the largest, ranging from 0.845 to 0.895. These results suggest that sampling based on two uncertainties is better than the ones either based on uncertainty from feature domain or from spatial domain. The work suggests that integrating the uncertainty from both the spatial and feature domains is useful in finding additional samples which effectively complement the existing samples for digital soil mapping.

Digital soil mapping using data with different accuracy levels

G.B.M. Heuvelink[1], T. Hengl[1], B. Kempen[1] and J.G.B. Leenaars[1]

[1]ISRIC World Soil Information, Wageningen, The Netherlands

Digital soil mapping combines soil point observations with maps of environmental covariates to predict the soil at unobserved locations. Soil observations are key because these determine the regression models that link the soil to the covariates and are used for spatial interpolation with kriging. In using these methods we often assume that the soil observations are without error, while in fact we know that they are not. Both systematic and random measurement errors are present in field and lab observations, and only rarely will these errors be negligible compared to soil spatial variation. In the near future digital soil mapping projects will also increasingly make use of crowd-sourced data and pseudo-points, which yields informative additional data at cheap or zero cost, but the accuracy associated with these data will usually be less than that of institutional data. In this work we extend regression kriging to the case in which each individual observation can have a different measurement error variance. As a result observations with small measurement error variance carry more weight than observations with large measurement error variance, both in regression modelling and kriging. The effect of systematic measurement error is also accounted for, by representing the (unknown) systematic error as a zero-mean random error that is equal for all observations from the same source. The statistical methodology is straightforward but requires adaptations of existing software implementations. Extension from linear regression to machine-learning regression methods, such as artificial neural networks and random forests, is also less obvious. We use a digital soil mapping application using data from the Africa Soil Profiles database to map soil properties for a region in sub-Saharan Africa. We compare prediction maps and prediction error variance maps with those obtained when measurement error is ignored. Results show marked differences and indicate that measurement errors should not be ignored, particularly when there are large differences in accuracy levels between observations within the conditioning dataset.

Accurate digital mapping of endemic soils

Colby Brungard[1] and Budiman Minasny[2]

[1]Department of Plants, Soils, and Climate, Utah State University, United States
[2]Faculty of Agriculture & Environment, The University of Sydney, NSW 2006, Australia

Accurate modeling of endemic soils is required for many environmental and economic applications such as assessing rare plant distribution (Baker et al., 2016, Damschen et al., 2012), identifying vineyard suitability (Malone et al., 2014), or applying targeted land management practices (e.g., wetland restoration). However; rare or endemic soil classes (i.e., soil classes with few observations) are difficult to accurately predict regardless of the statistical/machine learning model used (Brungard et al., 2015; Hengl et al., 2007; Taghizadeh-Mehrjard et al., 2012). This is because the imbalanced distribution of most soil class datasets (some soil classes are highly under-represented compared to other soil classes) makes many conventional machine learning methods less effective (Wang and Yao, 2012).

Imbalanced class datasets are difficult to accurately model for conceptual and practical reasons. Conceptually, soil classes are defined by pedon scale soil properties, while predictive modeling relies on environmental covariates. This mismatch between defining and predictive variables likely leads to poor predictive accuracy because variability between soil classes is not directly tied to the variables used to produce predictions and leads to classes that are difficult to separate in covariate space. Carré and McBratney (2005) combined pedon scale soil properties and environmental covariates to produce Terrons; combined soil-landscape management units. This approach significantly improved predictive accuracy of all classes, but the resulting classes were no longer specifically soil taxonomic classes.

Closely related to this idea, is the difficulty associated with defining an endemic soil. Minority soil classes do not automatically indicate an endemic soil. Truly endemic soils only occur in geographical locations where a combination of soil forming processes interacts to form unique soils. Minority soil classes may also result from data collection issues including biased sampling, or the inclusion of small areas of otherwise common soils when study area boundaries are arbitrary. Further complicating the identification of endemic soils is the fact that soil classes are somewhat arbitrary divisions of n-dimensional soil property space. Divisions between soil classes can be large or small and methods to identify when a soil is “different enough” (in both soil property and covariate space) to be endemic are unclear.

Practically, soil classes are difficult to predict because of limitations associated with existing machine learning methods. A number of solutions exist for overcoming the practical limitations of machine learning and increasing the predictive accuracy of the minority classes (i.e., endemic soils) while still retaining the predictive accuracy of the majority class(s). These include decreasing the number of classes, increasing the number of observations in minority classes, or applying a weighting/cost scheme.

Decreasing the number of classes by either grouping minority classes with similar soil classes or including minority classes in an ‘other’ soil class is the most common approach to class reduction. Soil class reduction has also been approached through class decomposition – converting multi-class data into several one-vs-all sub-datasets (Wang and Yao, 2012). However, class decomposition aggravates imbalanced distributions and can cause potential classification errors when the predictive models are combined as the resulting predictive models are not built with full data knowledge (Wang and Yao, 2012). Although soil class reduction generally improves predictive accuracy, it does so at the expense of removing minority, possibly endemic, soil classes.

Increasing the number of observations in minority classes is a promising approach, though increasing the number of observations by field sampling may be difficult given financial and logistical constraints, and because it is likely difficult to identify *a priori* which classes will need to be more intensively sampled. Increasing observation numbers through synthetic sampling is likely a more feasible option. Synthetic samples are observations not obtained by actual field sampling. Shi et al. (2004) used case-based reasoning to create synthetic samples based on soil-surveyor knowledge. Another option to generate synthetic samples is through oversampling of minority classes, or under-sampling the majority class (Chawla et al., 2002). Minority class oversampling may be preferred over majority under-sampling (Abdi and Hashemi, 2016; Wang and Yao, 2012). Minority oversampling generates synthetic samples by duplicating samples or by generating samples similar to real samples until class numbers are equal between classes (Chawla et al., 2002).

Applying a cost scheme for improving prediction accuracy has been suggested as a useful method for improving classification accuracy of minority classes without resorting to class decomposition or synthetic sampling (Wang and Yao, 2012). Taxonomic distance (Minasny and McBratney, 2007) could be used as a cost matrix for implementing a cost scheme when predicting soil taxonomic classes. Such an approach may be ideal if taxonomic distance is used to define endemic soil classes.

In summary, accurate predictive models of endemic soil classes are needed. Efforts to produce such predictions must first find a method to identify endemic soils and then implement practical solutions to overcome machine learning limitations. Such solutions include decreasing the number of classes, synthetic sampling to increase the number of observations in minority classes, and through the incorporation of a taxonomic distance.

Potential to map depth-specific soil organic matter content across an olive grove using quasi-2d and quasi-3d inversion of DUALEM-21 data

J. Huang[1], A. Pedrera-Parrilla[2], K. Vanderlinden[2], E.V. Taguas[3], J.A. Gómez[4] and J.Triantafilis[1,*]

[1] *School of Biological, Earth and Environmental Sciences, UNSW Australia, Sydney, NSW 2052, Australia*
[2] *IFAPA, Centro Las Torres-Tomejil. Ctra. Sevilla-Cazalla km 12.2, 41200 Alcalá del Río (Sevilla), Spain*
[3] *Departamento de Ingeniería Rural, Universidad de Córdoba. Campus de Rabanales, Edificio da Vinci. Ctra. Madrid km 396, 14071 Córdoba, Spain*
[4] *Instituto de Agricultura Sostenible, CSIC. Avda. Menéndez Pidal s/n, 14080 Córdoba, Spain*

To conserve the soil, land management strategies have to be implemented that minimize erosion and enhance at the same time rain water conservation to safeguard crop yields. An indicator of soil quality and the success of soil management will be organic matter content (OM - %). There is therefore a need to measure, map, manage and monitor its content. Proximal sensors such as electromagnetic (EM) induction instruments may be useful in mapping this because the apparent soil electrical conductivity (EC_a – mS/m) is related to clay, salinity and mineralogy, which influence organic matter content. In this research we collect data from a single frequency and multiple-coil DUALEM-21 along a transect and across an olive grove in the ''La Manga'' catchment in Setenil de las Bodegas in the southwest of Spain. We inverted the data using EM4Soil software and developed calibrations between estimates of true electrical conductivity (σ – mS/m) with depth against measured OM % using the quasi-2d algorithm along a single transect. We did this by using a fitted linear regression model and by varying the forward modeling algorithm (cumulative function and full solution), inversion algorithm (S1 and S2) and damping factor (λ) to determine a suitable electromagnetic conductivity image (EMCI) for 2-d and 3-d mapping. Our results along a detailed transect showed promise and suggest differences in OM content could be mapped down a topographic sequence. We applied this calibration to a quasi-3d model across the entire olive grove and to predict OM at various depths. The results across the olive grove were compromised in some locations and within geomorphological complex locations in the landscape, such as near the erosion gully where frequent erosion and deposition occurs. We conclude that better results may have been achieved if more detailed ECa data collection was undertaken in and around the gully and also across a larger extent.

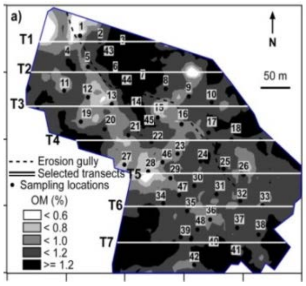


Figure 1 Spatial distribution of predicted organic matter content (OM – %) at the depth of 0–0.1 m across the olive grove using a quasi-3d inversion-based modelling.

Digital Soil Assessment of Landscape-Scale Forest Restoration Using a Species Distribution Model

Adrienne C. Nottingham[1], James A. Thompson[1] and Michael P. Strager[2]

[1] *Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV, USA*
[2] *Division of Resource Management, West Virginia University, Morgantown, WV, USA*

Red spruce (*Picea rubens*) ecosystems in the higher elevations of the Central Appalachians of the eastern United States are the focus of ongoing restoration efforts due to the valuable ecosystem services these forests provide. Historic documentation suggests that much of the Central Appalachians were covered with red spruce forests, but extensive logging and wildfires during the late 1800s and early 1900s reduced the extent of red spruce forests to approximately 10% of its previous extent. However, because of the pedoecological linkages between red spruce cover and the underlying soils, recent research has established that spodic materials still present in the soil offer insight into the historic extent and distribution of red spruce cover and, therefore, can be used to target areas where restoration efforts are expected to be most successful.

MaxEnt is a species distribution model (SDM) that has been used widely in the fields of ecology and conservation biology. MaxEnt implements maximum entropy theory to model the distribution of species from various environmental covariates using presence-only records by minimizing the relative entropy between probability densities in covariate space. This research examines the potential for using MaxEnt to map soil properties, specifically spodic intensity. Spodic intensity is an index that quantifies the degree of development of spodic materials, and it has previously been used to both inform and guide red spruce restoration in the Central Appalachians. A suite of environmental variables in conjunction with known occurrences of sites with high spodic intensity ratings was used with MaxEnt to model the fundamental niche for podzolization—and, thereby, red spruce habitat—as a function of the presence or absence of spodic materials. Model outputs were validated using a separate group of presence points independently sampled and withheld from the model.

Mapping sustaining soils in prehispanic Western Mexico. Archaeopedology as a tool for understanding ancient agriculture

Antoine Dorison[1] and Christina Siebe[2]

[1]University of Paris 1 - Panthéon Sorbonne – CNRS, UMR 8096 ArchAm, Archaeology of the Americas
[2]Geology Institute, National Autonomous University of Mexico (UNAM)

Soil is the main medium through which archaeologists try to understand the past, using the artefacts it contains and the principles of stratigraphy. However, in spite of its obvious usefulness as a time capsule, it rarely constitutes an accurately studied aspect of past societies land occupation and soil science is too often reduced to its very basics in archaeological research. Moreover, soils are extremely meaningful in traditional farming societies and thus constitute a key element in understanding their economy. We argue that pedological mapping is therefore a relevant tool in understanding the settlement patterns of archaeological sites sustained by agriculture, for soils nourishing potential was well known to ancient human groups.

In this paper, we combine pedological and archaeological field surveys and computed geographic data (LiDAR, DEM, Satellite imagery...) to generate more comprehensive maps of past land occupation, assess its capacity to provide food and discuss the economical strategies set up by prehispanic societies.

Our study took place in the sub-tropical region of Zacapu, in the northern margin of the Michoacán-Guanajuato volcanic field (MGVC). At the beginning of the 13th century A.D., the area experienced a major economical reorganization as a result of an unprecedented urban phenomenon. It is characterised by the construction of four cities more than 30 ha wide each and able to accommodate a cumulative population of 16 000 to 20 000 inhabitants. The cities were less than 5 km away from each other, which induced a dramatic demographic pressure in the area. However, sudden abandonment occurred around 1450 A.D.

What is even more striking is their location on top of Holocene lava flows where pedogenesis has hardly begun – and sometimes has not begun at all. This volcanic formation, locally known as the *Malpaís of Zacapu* is unsuitable for agricultural practices. It covers some 50 km² west of a lacustrine plain with high organic matter content soils and is surrounded on its other sides by Pleistocene volcanic landforms presenting rich soils formed on volcanic ash deposits. The habitat density and the highly limiting nature of the recent lava flows prevented city-dwellers from cultivating inside the urban settlements so that they necessarily had to rely on this varied but often complicated environment to sustain their society. By mapping soils and contemporaneous archaeological remains around the cities, we intent to offer a better understanding of prehispanic land use and the impact of the urban phenomenon.

Mapping of Functional Soil Classes across Scales

Jenette M. Ashtekar[1], Phillip R. Owens[1], Minerva Dorantes[1], Mayesse DaSilva[2] and Zamir Libohova[3]

[1] *Purdue University, Department of Agronomy, 915 W. State Street, West Lafayette, IN*
[2] *Center for International Agriculture in the Tropics (CIAT), Cali, Colombia*
[3] *USDA-NRCS-NSSC, Federal Building, Room 152, 100 Centennial Mall North Lincoln, NE*

Soil taxonomy has been the paradigm for soil classification since the initiation of soil survey. Historically soils were delineated along taxonomic lines following procedures that could identify soil boundaries and taxonomic classes without the aid of computer technology and laboratory analysis. With the introduction of personal computing came the adoption of digital methods for soil mapping and though these innovations spurred the development of continuous property mapping and fuzzy classification, the paradigm of taxonomic based soil classes has failed to evolve. By combining historic spatial data layers, such as geomorphological surveys, with new tools like digital elevation models, terrain algorithms, satellite imagery, and global climate datasets, it is now possible to develop new systems for soil classification and delineation tangential to soil taxonomy. Functional soil mapping is a new way of delineating soil types based not on their taxonomy, which describes morphological differences between classes, but instead by grouping soils on the landscape as a function of unique soil forming environments (i.e., soil forming factors). Areas on the landscape with similar soil forming factors should produce soils that function similarly. This type of soil classification is particularly useful for land management at varying scales for a variety of purposes. Functional soil mapping is currently being undertaken to map large portions of Central America (Honduras, Nicaragua, El Salvador, Guatemala), Southern Mexico, and Colombia at resolutions ranging from 10 to 30 m utilizing a combination of legacy mapping, climatic datasets, DEM derived topographic attributes, and terrain based classification algorithms. Functional soil mapping is also being performed at the field level in the Midwestern United states at 5 to 10 m resolutions for the purposed of precision agricultural management. This paper will outline methods for functional soil mapping at multiple scales and present functional soil map results for areas of El Salvador as well as field level functional maps for farm fields in Indiana, USA.

Keynote: Recent Advances in the Image Spectroscopy for Geo Spatial Information of Soils

Eyal Ben Dor

The Remote Sensing Laboratory (RSL) , Department of Geography, Tel Aviv University (TAU), Tel Aviv, Israel. bendor@pots.tau.ac.il

Combining soil science and Imaging spectroscopy (IS) is a promising tool for studying soil properties.. Going from point to image spectrometry is not only a journey from micro to macro scales, but also a long stage where problems such as dealing with data having a low signal-to-noise level, contamination of the atmosphere, large data sets, the BRDF effect, crust and more are often encountered. In this presentation we provide an up-to-date overview of some of the case studies that have used IS technology for soil science applications. Besides a brief discussion on the advantages and disadvantages of IS for studying soils, the following cases are comprehensively discussed: soil degradation, soil mapping and classification, soil genesis and formation, soil contamination, soil water content, and soil swelling. We review these case studies and suggest that the IS data be provided to the end users as real reflectance and not as raw data and with better signal-to-noise ratios than presently exist. These limitations serve as a barrier that impedes potential end-users, inhibiting researchers from trying this technique for their needs. A review of the forthcoming IS missions (air and space) and their affiliation to soil mapping is provided especially for open mines close by area. Also a brief description on the effort done worldwide to gather many potential users into one group is reported. The paper ends with a general call to extend the utilization of the IS technique into soil science and applications.

Session 9 Remote Sensing and Soil Spectroscopy
Chair: Eyal Ben Dor

Geographically closest resampling strategy for soil organic carbon and clay content prediction

Yi Peng[1], Maria Knadel[1], Mette B. Greve[1] and Mogens H. Greve[1]

[1] *Aarhus University, Faculty of Science and Technology, Department of Agroecology, Blichers Allé 20, P.O. Box 50, DK-8830 Tjele, Denmark*

Danish soil visible-near infrared (vis-NIR) spectral library has been developed as one of the largest database in soil science community. This library is based on Danish national soil database (7km national grid sampling). It has been proved that this database is able to predict soil attributes such as soil organic carbon (SOC) and clay content at field scale using geographical closest resampling strategy in Denmark. However, this strategy was applied to predict soils from one field only. For the future application it is necessary to test this approach on all types of Danish soils and evaluate the predictive capabilities of the library at a national scale. In the present study, partial least squares regression was used to develop models and predict top SOC (490 samples) and clay contents (442 samples) for each 7km grid sampling point at national scale. In resampling and modelling process, each target sample was predicted by a specific model which was calibrated using geographically closest soil spectra. In order to find the most optimal number of samples for calibration, the geographically closest 20, 30, 40, and 50 sampling points (each point includes soil samples up to 6 horizons) were selected and used in models development. Models were evaluated on the basis of root mean square error (RMSE), R^2, ratio of performance to deviation (RPD) and the ratio of performance of interquartile distance (RPIQ). For both SOC and clay predictions the best results were obtained based on 40 geographically closest sampling points (7km grid solution) SOC prediction resulted in R^2: 0.76; RMSE: 4.02 %; RPD: 1.59; RPIQ: 0.35. The results for clay prediction were also successful (R^2: 0.84; RMSE: 2.36 %; RPD: 2.35; RPIQ: 2.88). For SOC predictions, over 90% of soil samples were well predicted in comparison with the uncertainties of traditional laboratory analysis. However, due to the fact that all of the organic soils (48 samples SOC >7%) originating from the wetland or forest areas, the SOC predictions of organic soils were generally under estimated and not acceptable. For clay prediction, of all 442 predicted samples, only 12 predictions were not acceptable, and these 12 predictions can be explained by geological reasons. We concluded that geographical closest resampling strategy is a promising and efficient method to create sub-models from the Danish spectral library and predict unknown Danish soil samples. The accuracy of SOC model development was depended on landuse/landcover, and the clay content model development was strongly affected by soil parent material and landscape.

Identification of soil classes based on vis-NIR reflectance spectra using depth harmonization and machine learning techniques

Xianli Xie[1], Decheng Li[1], Rong Zeng[1], Xianzhang Pan[1] and Ganlin Zhang[1]

[1]State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, 71 East Beijing Road, 210008 Nanjing, China

There is a large potential of soil visible-near infrared (vis–NIR) reflectance spectroscopy used to identify soil classes for timely updating soil surveys with low cost. In this study, our objective was to evaluate the feasibility of using soil vis-NIR reflectance spectra for identification of soil classes from the Chinese Soil Taxonomy (CST) which is a quantitative classification system and been promoted in China. Harmonized depth data of soil reflectance was generated at continuous and standard depths. Machine learning technique was used to discover the spectral patterns of soil classes and makes predictions from the discovered patterns.

We studied 134 soil profiles collected from the soil series survey in Anhui Province (29°41′N to 34°38′N, 114°54′E to 119°37′E), China. Due to the broad variety of soil classes and the highly imbalanced distribution of profiles in soil classes, we combined the original classes and reclassified these profiles into 3 orders, 4 suborders, 11 groups in CST. The vis-NIR reflectance of soil was measured in the 350 - 2500 nm range. For each soil profile, the raw reflectance data of genetic horizons were interpolated to a maximum depth of 120 cm with 10 cm interval by fitting an equal-area spline function. The generated reflectance data at 10 cm interval of depths were transformed to the first derivatives with Savitzky-Golay smoothing. The complete dataset was divided into a training dataset (75% of profiles) and a testing dataset (25% of profiles) by stratified random sampling at the group level. Principal Component Analysis (PCA) was applied to summarize the spectral data in the training set. The scores of extracted 6 principal components were submitted to machine learning algorithms to recognize the spectral patterns of soil horizons from soil classes of CST. Three different machine learning algorithms, k-nearest neighbor (KNN), boosted decision-tree C5.0 (Boosted C50), and random forest (RF), were applied. The soil class of a profile was allocated based on the classes of its horizons. Table 1 shows the accuracy of each machine learning algorithm for predicting soil classes of profiles in the testing set. The soil classes with small number of soil profiles were generally poorly predicted, and tended to be misclassified into similar soil classes.

Table 1. Accuracy for prediction of soil classes at three levels. KNN: K-nearest neighbor; Boosted C50: Boosted decision-tree C5.0; RF: Random forest.

	KNN		Boosted C50		RF	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
Order level	0.84	0.73	0.81	0.68	0.81	0.68
Suborder level	0.78	0.67	0.84	0.76	0.81	0.71
Group level	0.58	0.52	0.62	0.56	0.66	0.60

How reflectance spectroscopy can assist in soil classification?

Rong Zeng[1,2], Gan-Lin Zhang [1,2*], Fan Yang[1,2], De-Cheng Li[1] and Yu-Guo Zhao[1]

[1] *State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, PR China*
[2] *University of the Chinese Academy of Sciences, Beijing 100049, PR China*

Soil class maps are missing or coarse in scales in many areas, because properly classifying a soil profile requires laboratory determination of many soil properties, which is time and money consuming. The advantage of reflectance spectroscopy and the development of spectra libraries can provide a relatively less-cost solution toward solving this problem. Reflectance spectroscopy has demonstrated its ability to rapidly predict soil physio-chemical properties, however the prediction accuracy varies among soil properties, due to the intrinsic spectra characteristics of the target property, the type of regression models, the calibration data size, the specificity of the study area etc. If the predicted properties are used as reference for allocating a soil profile to a class in a known classification system, how reliable is the classification results remains a question and this issue has not been explored by previous research. In addition, all current soil classification systems are semi-quantitative systems, in which the soil with properties falling within the same set of range is categorized to the same class. Thus from the view of a soil classifier, they are more concerned about whether the properties can be predicted within the required range or threshold rather than the statistical number of prediction accuracy. Thus, the target of this research is to explore whether the soil properties predicted by reflectance spectroscopy can be used to correctly allocate soil profiles into soil taxa at different categorical levels (order, suborder, group and subgroup). 230 soil profiles sampled in Heihe river basin, the northwestern part of China, were classified into 8 orders, 12 suborders, 25 groups and 53 subgroups according to Chinese Soil Taxonomy (CST). Ten soil properties were predicted by spectra using full cross-validated PLSR modelling. The SOC was predicted with the highest accuracy (RPD=3.50), followed by prediction of silt and sand (RPD close to 2.0). The prediction accuracy for BD, pH, clay and CaSO4 ranged from 1.41 to 1.92, meeting the category of acceptable accuracy ($1.4<\text{RPD}<2.0$), while $CaCO_3$ equivalents, salt and EC were poorly predicted ($\text{RPD}<1.4$). Based on these predicted soil properties, the overall allocation accuracy for order, suborder, group and subgroup level is 97.0%, 97.0%, 88.7% and 72.2% respectively. For order level, all seven profiles wrongly classified is due to underestimation or overestimation of EC, i.e. misclassifying Aridosols as Halosols. The misallocations at group level is mainly due to underestimation of $CaCO_3$ equivalents or overestimation of EC, making Calci-Orthic Aridosols misallocated as Hapli-Orthic Aridosols or Sali-Orthic Aridosols. For subgroup level, the misallocation is due to that the predicted $CaCO_3$ equivalent failed to discriminate the dignostic features of Calcic horizon, Calcic evidence and Calcaric property. In summary, misallocations are mainly caused by soil properteis predicted with $\text{RPD}<1.4$ and RPD values can serve as an indicator for the reliability of the classification results. Thus we proposed the use of predicted soil properties by reflectance spectroscopy as one data source for soil class allocation, but the uncertainty propagated by prediction error should be evaluated.

Using field spectroscopy in the VNIR-SWIR spectral region for predicting hydrophobicity level of undisturbed soils

Amihai Granot[1] and Eyal Ben Dor[1]

[1]*Department of Geography and Human Environment, Tel-Aviv University, Tel-Aviv, Israel*

Hydrophobicity is an important soil property for both surface and sub-surface environmental and hydrological management. Traditional soil hydrophobicity quantification methods include Water Drop Penetration Time (WDPT) and Contact Angle (CA) among others, all of which are time consuming, expensive and require specialized personnel. This study focuses on evaluating the potential of reflectance spectroscopy for rapid assessment of undisturbed soil hydrophobicity in situ, using field and airborne hyperspectral sensors. For this purpose, we have developed a novel field spectral measurement attachment device termed “SoilPro®” which enables constant environmental conditions and extremely accurate spectral readings. The “SoilPro®” assembly was subjected to rigorous laboratory and field testing on multiple spectrometers which resulted in remarkable agreement with spectral measurements obtained under the CSIRO ISS protocol. “SoilPro®” is currently pending for a US patent and provisional application has been approved (S/N 62/278, 471, 14 January 2016).

In order to develop a robust spectral-hydrophobicity prediction model for implementation of the rapid hydrophobicity assessment system, a representative dataset, containing 228 soil samples with varying soil water repellency values, was generated. Aside from spectral measurements, this database contains both WDPT and CA attributes for every soil. These samples originated from an orchard plot research area with sandy soil (Rhodoxeralf according to the USDA) characterized as heavily affected by hydrophobicity. We have obtained a prediction model using our in-house spectral data mining machine termed “Paracuda®” which employs the all-possibilities-approach (APA) concept and a partial least squares regression (PLS-R) modeling algorithm. Using “Paracuda®” we have managed to gain a robust model with R^2=0.92 and RPD=3.28 for the “test” group in an internal validation procedure for CA modeling.

Our next step in this research focuses on applying in-situ spectral measurements using the “SoilPro®” device along with the spectral hydrophobicity library for adjusting and improving the accuracy of the developed model for field data. This will pave the road for acquiring a fully functional rapid hydrophobicity in-situ assessment system with significant advantages over the expensive and time consuming traditional quantification methods.

Multi-temporal composites of airborne imaging spectroscopy data for the use in digital soil mapping

Sanne Diek[1], Rogier de Jong[1] and Michael Schaepman[1]

[1]Remote Sensing Laboratories RSL, University of Zurich, Zurich, Switzerland

The demand for spatial soil information has been growing with the use of global and regional models, which often require full coverage soil information (Mulder et al., 2011). This drives a growing need for low-cost, standardised and large-scale spatial soil data. Remote sensing is a promising data source for this purpose, and soil spectroscopy has already shown assuring results. However, soil spectroscopy is limited to the thin upper layer of the soil (topsoil) and most research is based on laboratory-derived spectra or on field studies in semi-arid areas. Difficulties with surface coverage, variation in soil moisture and management are eluded in this way (Ben-Dor et al., 2009). The presence of crop rotation in agricultural areas offers an opportunity to deal with vegetation coverage in temperate climates. Therefore, we explore the use of multi-temporal soil spectroscopy data to increase the area of bare soils in a heterogeneous agricultural landscape. with study sites in Switzerland.

The Airborne Prism Experiment (APEX) sensor provides high-resolution data, both spectrally (300 bands in the 400-2500 nm VNIR domain) and spatially (2 m) (Schaepman et al., 2015). APEX data was collected in September 2013, May 2014 and May 2015. Bare soils were selected using various vegetation, water and built-up masks. The multi-temporal image was created based on a linear regression per wavelength for the overlapping bare soil pixels towards the year with most bare soil pixels (2014).

Using a multi-temporal composite resulted in twice more bare soil area compared to a single-date flight. This greatly improves the data available for full-coverage mapping. Difference indicators (root mean square difference, difference angle (Price, 1994) and R^2) and summary statistics (mean and standard deviation) show that reflectance values of the multi-temporal image are very similar to the single-date of 2014. Spatial distribution of the correction factors and analysis of the spatial variability, however, indicate that short term factors like variation in soil moisture and land management are largely influencing the result of the multi-temporal image. More research for the quantitative effects of these factors is necessary to correct for these. Nevertheless, using the crop rotation sequences offers a methodology to increase the area of bare soils in temperate climates, which bring the broad application of soil spectroscopy a bit closer.

References:

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using Imaging Spectroscopy to study soil properties. Remote Sensing of Environment 113(Supplement 1), S38-S55.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping - A review. Geoderma 162(1-2), 1-19.

Price, J.C., 1994. How unique are spectral signatures? Remote Sensing of Environment 49(3), 181-186.

Schaepman, M.E., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F.D., Laurent, V., Popp, C., Seidel, F.C., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., De Vos, L., Meuleman, K., Meynart, R., Schläpfer, D., Kneubühler, M., Itten, K.I., 2015. Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX). Remote Sensing of Environment 158(0), 207-219.

Seasonal Changes of Tilled Soil Surface as Information Factor for Efficient Soil Mapping Using Remote Sensing Data

I. Savin[1,2], E. Prudnikova[1], N. Vasilyeva[1] and A. Bairamov[2]

[1]V.V. Dokuchaev Soil Science Institute, Moscow, Russia
[2]Agrarian-Technological Institute of PFUR, Moscow, Russia

The state of the open surface of tilled soils strongly varies during the year and depends on properties of the upper soil horizon, meteorological conditions and specific character of its usage. The change in the tilled soil surface has a significant effect on its reflective properties, and consequently on the possibilities to use remote sensing data for soil mapping. For the maximal extraction of information about the spatial heterogeneity of soil properties the satellite data acquisition in temperate climate must be carried out before the first treatment of tilled soil surface in spring.
Based upon a comprehensive analysis of results obtained in the course of field observations on 3 test sites on the European part of Russia, it seems reasonable to make the following conclusions:

1. The bare surface of arable soils reveals significant changes (изменения в смысле внешний вид, форма, шероховатость?) Может in time, being affected by atmospheric precipitation, wind and thawing water in the spring. At this surface a crust is formed, then it is cracked and covered by water-stable soil aggregates resistant to action of raindrops. The manifestation degree and the color of the above structural particles are predetermined by the soil properties and the substance composition (Fig.1).

2. The satellite imagery of the transformed soil surface seems to be more informative for decoding of the soil properties than the soil surface before transformation. The soil surface transformation reveals changes not only in the sunlight reflected by the soil surface but also in its spectral composition.

3. The bare surface of arable soils reaches the maximum transformation in late autumn before the snow cover as well as in spring before the first soil treatment. These dates prove to be preferable for remote surveying the bare soil surface with the view of decoding some soil properties. The optimum time is the spring due to the better weather conditions and clearly expressed transformation of the soil surface.

4. The peculiar dynamics of the bare soil surface should be taken into consideration in using remote sensing data for a better organization of monitoring over the soil cover and mapping of some soil properties especially in the case of using the multi-band and hyper-spectral images.

The research has been conducted with the financial support from Russian Science Foundation (project 1516-30007) and RFBR (project 15-04-04717).

Figure 1 Examples of open soil surface transformation for Dark-gray forest soil (left), Soddy podzolic soil (center), and Podzolized chernozem (right)

Mapping of land covers in South Greenland using very high resolution satellite imagery for SOC upscaling

Menaka Chellasamy[1], Mateja Ogric[1], Mogens Humlekrog Greve[1] and René Larsen

[1] *Department of Agroecology, Aarhus University, 8830 Tjele, Denmark*

Over recent decades, artic warming has been triggering changes on climate, vegetation and environmental conditions which in turn showing a positive effect on the natural environment of Greenland. Increasing the development of agriculture in the island is the interest of many researchers in recent years. This calls for mapping Soil Organic Carbon (SOC) stocks in the study area to predict how suitable and fertile the Greenlandic soil is for further development of agriculture. Hence, this study targets to map different land covers at very high resolution in order to upscale the SOC stocks measured at field scale. WorldView2 (WV2) multispectral imagery will be used for identifying the land covers in the study area. Object-based & Support Vector Machine (SVM) classifier will be used to obtain the detailed land cover map. In addition to the training and test samples collected in the field, Unmanned Aerial Vehicle (UAV)-based imagery will be used to collect the samples for classification and to validate the map resulted from the study. The main contribution of this study will be: 1) investigating the performance of several image features (such as spectral, texture and indices) from WV2 for mapping different land covers 2) identifying the appropriate scale for object-based classification 3) quantifying different land covers in the study area with good classification accuracy and 4) potential usage of UAV imagery for validating the land cover map.

Session 10 Digital Soil Class Mapping

Chair: James A. Thompson

Mapping Drainage Classes in Denmark by Means of Decision Tree Classification

Anders Bjørn Møller[1], Bo Vangsø Iversen[1], Amélie Beucher[1] and Mogens Humlekrog Greve[1]

[1] Department of Agroecology, Aarhus University

Soil drainage, understood as the removal of excess water from the soil matrix, plays an integral role in the environment and land use. It affects plant growth, soil biota, the leaching of nutrients and other solutes, the release of greenhouse gasses and the risk of soil compaction and surface erosion. Insufficient soil drainage is a major agricultural concern and can be caused either by a high groundwater table or the stagnation of water in the soil matrix. In Denmark soil drainage is described by five drainage classes from 1 (i.e. very well drained soils) to 5 (i.e. very poorly drained soils). Despite efforts to map the soil texture and taxonomic classes of Denmark, soil drainage classes have not been mapped yet.

Decision Tree Classification has achieved promising accuracies, but has not previously been used to map drainage classes in areas larger than a few hundred square kilometers.

In this study the soil drainage classes of Denmark were mapped by means of Decision Tree Classification based on 1702 soil profiles. 31 environmental predictors were used, including soil and terrain parameters, spectral indices obtained from satellite images and information about land use and cropping history of the areas in question.

As drainage classes are an ordered series, an approach was tested wherein misclassifications were assigned costs equal to the deviation from the actual drainage class. Moreover, an approach was tested, wherein predictors were selected based on the change in performance caused by their removal from the dataset. Two ensemble techniques, boosting and bagging, were applied. Performance was tested on an independent validation set constituting one third of the original data.

With the boosting approach, the optimal performance was obtained using all the available predictors and undifferentiated costs for misclassifications. The best performance was achieved after 61 boosting trials with an overall accuracy of 50.8% and mean absolute error of 0.67, using the numeric value of the drainage classes. Bagging produces variable results, and therefore the process was repeated 30 times in each experiment. The best performance was obtained using all predictors and variable costs for misclassification, with accuracy reaching 52.0% and the mean absolute error reduced to 0.55. The best performance was observed in moraine landscapes and reclaimed areas, while the worst performance was observed on glacial flood plains and on aeolian deposits. The accuracy was above 40% in all landscape types.

The five best predictors, on average used nn more than 50% of the training cases over the trials, were the same for the best boosting model and the best bagging model: land use, geology, the slope gradient to the nearest waterbody, the extent of wetlands, and the clay content at a depth of 100 to 200 centimeters.

The results underline the usefulness of ensemble techniques and the viability of regarding drainage classes as an ordered series.

Using regionalization maps in Digital Soil Mapping

Zhogolev A.V. [1]

[1] *V.V. Dokuchaev Soil Science Institute*

In conventional soil mapping, soil-landscape relationships are studied separately for different soil or lithologic-geomorphological zones. Expert-based studying using regionalization maps allows to take into consideration local features of the mapped area. Regionalization maps have not been used in digital soil mapping yet. By the example of a key site, the comparison of digital soil mapping with and without using regionalization maps was performed. A digital soil map using regionalization map was more accurate and it has shown more accurate soil-landscape relationships.

In conventional soil mapping, soil-landscape relationships are studied separately for different soil or lithologic-geomorphological zones. Using regionalization maps allows to build more appropriate rules for soil mapping taking into account the local features of various areas. Currently regionalization maps have not been used in digital soil mapping and soil-landscape relationships have been established for the entire study area as a whole. This research provides the comparison of digital soil maps using or without using regionalization maps.

Zaokskaya part of the Moscow region was chosen as a key site. Digital soil maps were built by establishing soil-landscape relationships with the help of Classification and Regression Trees method for conventional soil map and maps of soil formation factors. For the first digital soil map soil-landscape relationships were built for the entire study area as a whole, but for the second separate classification tree for each landscape zone was build. DEM SRTM, quaternary geological map and vegetation map were used as maps of soil-forming factors. Spatial data analysis was performed in ILWIS and R. Accuracy of both digital soil maps were assessed by comparing to conventional soil map. For the digital soil map created without using regionalization map the area of match in soil units was 62%, and for the map built using regionalization map - 66%. The digital soil map based on regionalization reflects the soil cover more adequately: areas of Chernozems are properly confined mainly to the southern part of the study area. Also several large patches of Cambic Greyic Phaeozems were properly mapped at watershed areas in the South of study area which matches the conventional soil map.

As a result, it was showed that the use of regionalization map for digital soil mapping has allowed to find out substantiated soil-landscape relationships which take into consideration local features of different areas and to map the soil cover more accurately.

The research has been conducted with the financial support from Russian Science Foundation (project 15-16-30007).

Multinomial Logistic Regression with soil diagnostic features and land surface parameters for soil mapping of Latium (Central Italy)

Marchetti Alessandro[1], Napoli Rosario[1], Rivieccio Rosa[1] and Piccini Chiara[1]

[1] CREA Council for Agricultural Research and Economics - Research Centre for the Soil-Plant System

One of the main challenges in traditional soil mapping is the identification of land components (LCs) - suitable combinations of morphology, lithology and land use - that represents a fundamental step in the definition of soil typological units. LCs are traditionally used by pedologists to correlate different soils and to identify the relationships between soil and geography. The recognition of the various soil characteristics for LCs definition is usually performed considering for each feature a number of classes (e.g. slope classes), defined a priori at a national scale. Such classes by nature tend to generalize and to flatten the actual local variability. Moreover, when dealing with large areas taking into account a very fragmented layout is very difficult. To reduce subjectivity in interpretation, the choice of features defining soil associated LCs should be performed in an automatic or semi-automatic way. Digital Soil Mapping techniques can help us to overcome these limits. In this work we propose a procedure to define the soil LCs for the territory of Latium administrative Region (Central Italy), starting from a dataset of about 1,500 fully described and analyzed soil profiles and associated land surface parameters. Some measured soil diagnostic characteristics - depth, internal drainage, topsoil and subsoil texture, gravel, stoniness, organic carbon content, cation exchange capacity, calcium carbonate content, bulk density - were used together with auxiliary information derived from prior thematic maps (geology, land use, pedoclimate), and from Digital Elevation Model (geomorphometric parameters). Indices derived from remotely sensed high resolution images were also introduced to improve the estimate in areas with uniform land characteristics (i.e. flat alluvial valleys and coastal plains). Classes for the measured characteristics were defined on the basis of their frequency distribution and of the WRB classification diagnostic thresholds. The covariates were chosen by a stepwise regression, and a Principal Component Analysis was performed to avoid multicollinearity. Principal components were used as predictors to build a raster layer for each considered soil characteristic by Multinomial Logistic Regression. From all these layers, a final map of LCs was derived by means of map algebra operations. The final map represents the basis to map Soil Typological Units for the whole Latium Region.

Exploring effects of sampling approaches and quantities of training samples on updating conventional soil maps performance

Xueqi Liu[1], Axing Zhu[1,2,3] and Lin Yang[2]

[1]*School of Geographical Science, Nanjing Normal University, Jiangsu 210000, China*
[2]*State Key Laboratory of Environment and Resources Information System, Institute of Geographical Sciences and Resources Research, Chinese Academy of Sciences, Beijing 100101, China*
[3]*Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA*

The conventional soil maps imply expert knowledge of soil-environmental relationship. Data mining methods can extract that knowledge for updating conventional soil maps. Relative researches have drawn growing attention. Generally, the updating procedure is to select training samples from conventional soil maps, and then utilizing data mining model to obtain the soil-environmental relationship for the prediction of soil spatial distribution. Therefore, the sampling approaches and quantities of training data significantly influence the effectiveness of exploring soil-environmental relationship and the exactness of inferring soil maps. However, recent researches often focused more on the parameters adjustment of data mining methods. Limited studies paid attention to the quality of training data, particularly even less conducted a comparison between different selection and quantities of training samples.

This study evaluated 4 training sample selection methods: (1) the number of each soil type samples, based on their areas, was determined as a classification proportion; and the sample points were selected from typical pixels of each soil type (Qi, 2004) (typical-level sampling); (2) the number of each soil type samples, based on their areas, was determined as a classification proportion; and the sample points of each soil type were randomly selected from all pixels (random-level sampling); (3) the soil type sample number was determined as an area weighted proportion, and the sample points were selected from typical pixels of each soil type (typical-weighted sampling); (4) the soil type sample number was determined as an area weighted proportion; and within each soil type, the sample points were randomly selected from all pixels (area-weighted sampling). Meanwhile, this research applied three different quantities of training data on each sampling method. Sample numbers were 335, 670 and 1005 respectively. In order to test the effectiveness of different sampling methods and sample numbers, each sampling scheme was repeated 100 times. Random forest model based on the generated training samples was used to infer the soil spatial distribution in Raffelson watershed, Wisconsin, United States. The inferring maps were validated using 92 independent field points. The accuracy of the conventional soil map was 69.6%.

The result indicated that when the sample numbers were 335, the ratio of updating conventional soil map for each sampling method were 79.0% (typical-level sampling), 71.8% (random-level sampling), 54.2% (typical-weighted sampling) and 63.6% (area-weighted sampling). Moreover, the ratio of updating conventional soil maps has been significantly improved as quantities of training data increased. On average, the typical-level sampling approach for selecting training samples is the most effective. Therefore, we concluded that the generating training sample methods based on classification were superior to area-weighted sampling and typical-weighted sampling. With low quantities of training samples, typical-level sampling had better performance than random-level sampling. But the difference between choosing typical pixels and all pixels have gradually slight influence on the precision of the inferring maps due to the increasing quantities of training samples. This study may contribute to appropriately generate training sample for updating conventional soil maps and digital soil mapping.

Reference:

[1] Qi F. Knowledge Discovery from Area-Class Resource Maps: Data Preprocessing for Noise Reduction [J]. Transactions in Gis, 2004, 8(3): 297-308.

Session 11 Digital Soil Mapping and Environmental Covariate

Chair: Zamir Libohova

Artificial Neural Networks for soil drainage class mapping in Denmark

Amélie Beucher[1], Anders Bjørn Møller[1] and Mogens Humlekrog Greve[1]

[1] Aarhus University, Department of Agroecology, 8830 Tjele, Denmark

Soil drainage constitutes a substantial factor impacting plant growth and various biophysical processes, such as nutrient cycling and greenhouse gas fluxes. Consequently, soil drainage maps represent crucial tools for crop, forest and environmental management purposes. Extensive field surveys being time- and resource-consuming, alternative modeling techniques have been previously used for mapping soil drainage classes. In particular, Artificial Neural Networks were tested over relatively large areas in Canada (Zhao et al., 2008, 2013). In Denmark, soil drainage classes, defined from very well drained soils (i.e. class 1) to very poorly drained soils (i.e. class 5), have not been mapped yet.

In this study, Artificial Neural Networks are assessed for mapping soil drainage classes in Denmark (c. 43,000 km^2). About 1,700 existing soil observations and 31 environmental variables, including soil and terrain parameters, as well as spectral indices derived from satellite images, were utilized as input data within the models. A preliminary assessment using all environmental variables yielded an overall prediction accuracy of 50.2%, based on a 30% hold-back validation data. The mean absolute error reached 0.62, using the numeric value of the drainage classes. Furthermore, models based on various network topologies will be tested for different combinations of environmental variables. Artificial Neural Networks demonstrate promising predictive classification abilities for mapping soil drainage classes over large areas.

References:

Zhao, Z. Y., Ashraf, M. I., Meng, F.-R., 2013. Model prediction of soil drainage classes over a large area using a limited number of field samples: A case study in the province of Nova Scotia, Canada. Canadian Journal of Soil Science 93(1), 73-83.

Zhao, Z. Y., Chow, T. L., Yang, Q., Rees, H. W., Benoy, G., Xing, Z. S., Meng, F. R., 2008. Model prediction of soil drainage classes based on digital elevation model parameters and soil attributes from coarse resolution soil maps. Canadian Journal of Soil Science 88(5), 787-799.

Development of environmental covariates for mapping soil properties over an alluvial plain

Feng Liu[1], Gan-Lin Zhang[1], Xiao-Dong Song[1] and Yu-Guo Zhao[1]

[1]*State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China*

In low relief areas such as plains, easily observed soil formative environmental factors generally do not spatially co-vary with soil conditions to the level that they can be used effectively in predictive soil mapping. This study presented an approach to developing environmental covariates for soil properties mapping over such areas. The development of the covariates was mainly based on remotely sensed dynamic feedback patterns of land surface to the input of solar radiation at different temporal scales. The patterns are defined as the changes of land surface energy which are captured by MODIS sensor on a time series. The energy status of land surface is represented by land surface temperature. Methods such as the Fourier transform and polynomial fitting functions were used to depict the land surface dynamic feedback patterns. The resulting parameters were used to characterize the patterns and construct a set of covariates. Then an environmental similarity-based spatial prediction method was developed for applying the constructed covariates for soil properties mapping. The approach was demonstrated to mapping soil texture and soil organic carbon concentration over a 28,000 km^2 alluvial plain located in the Ejin County of western Inner Mongolia, China. The results showed that the developed environmental covariates have significant correlations with soil properties and can be effectively used for the soil properties mapping. This suggests that the proposed approach is effective for improving the accuracy of soil mapping over low relief areas. (This work is supported by the National Natural Science Foundation of China (41130530, 91325301, 41571212 and 41401237) and the Project of the ISSCAS “One-Three-Five” Strategic Planning & Frontier Sciences (No. ISSASIP1622).)

Selection of principal stand factors as predictors for digital mapping of potentially toxic element contents in forest soils

Luboš Borůvka[1], Radim Vašát[1], Václav Tejnecký[1], Vít Šrámek[2], Milan Sáňka[3], Jarmila Čechmánková[4], Karel Němeček[1] and Vít Penížek[1]

[1] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague 6 - Suchdol, CZ-165 21, Czech Republic, E-mail: boruvka@af.czu.cz
[2] Department of Forest Ecology,Forestry and Game Management Research Institute, Strnady 136, Jíloviště, CZ-252 02, Czech Republic
[3] Research Centre for Toxic Compounds in the Environment (RECETOX),Faculty of Science, Masaryk University Brno, Kamenice 753/5, pavillion A29, Brno CZ-625 00, Czech Republic
[4] Department of Soil Hygiene, Research Institute for Soil and Water Conservation, Žabovřeská 250, Praha 5 - Zbraslav CZ-156 27, Czech Republic

Presence and excessive amounts of potentially toxic elements (PTE) in forest soils can lead to detrimental effects on the functioning of the whole forest ecosystems. An attempt was made to develop models for spatial prediction of PTE (As, Cd, Pb, and Zn) in forest soils of the Czech Republic. The objective of this contribution is to analyze the importance of principal stand factors as predictors for the models specific for each element and depth, and to interpret their importance in the PTE spatial distribution control.

Artificial neural networks, random forests, and multivariate adaptive regression splines were used as model types. Various combinations of data on stand factors (topography – digital terrain model, geology, land use, vegetation, forest typology, legacy soil information) were used as the inputs for model calibration, together with measured PTE content in 3 soil depths on 120 sites all over the Czech Republic,. The models were validated using an independent dataset. For the best models for each element and depth, relative weights of the predictors were also recorded from the models.

As PTE content in soils is influenced by anthropogenic activities, the reliability of models based purely on natural stand factors are generally less reliable compared to similar models e.g. of soil organic carbon distribution. The performance of the models for PTE spatial distribution improved, especially locally, when the relative position to the principal sources of pollutants (industrial areas) were added as predictors. Data on land use, vegetation, and forest typology were more important predictors in models for the surface soil layers (like forest floor) due to the principal input of PTE to soil from atmospheric deposition. For deeper mineral horizons, the influence of geology was stronger. Relief data were important for all depths.

We can conclude that even the data, whose spatial distribution is strongly influenced by human activity, can be spatially predicted with a reasonable accuracy. However, appropriate selection of model types and best predictors is a crucial issue.

Acid sulfate soil mapping in Denmark using legacy data and LiDAR-based derivatives

Amélie Beucher[1], Kabindra Adhikari[2], Henrik Breuning-Madsen[3], Mette Balslev Greve[1], Peter Österholm[4], Sören Fröjdö[4], Niels Henrik Jensen[5] and Mogens Humlekrog Greve[1]

[1] Aarhus University, Department of Agroecology, 8830 Tjele, Denmark
[2] University of Wisconsin-Madison, Department of Soil Science, 53706 Madison, USA
[3] University of Copenhagen, Department of Geosciences and Natural Resource Management, 1350 Copenhagen, Denmark
[4] Åbo Akademi University, Geology and Mineralogy, 20500 Åbo, Finland
[5] Roskilde University, Department of Science and Environment, 4000 Roskilde, Denmark

Acid sulfate soils leach considerable amounts of acidity and metals into watercourses, often causing severe ecological damage. Covering extensive coastal areas all over the world, they represent an important environmental issue. Mapping thus constitutes a fundamental step in the management of environmental risks related to these soils. In particular, Danish legislation proscribes drainage of areas classified as potential acid sulfate soils without prior permission from environmental authorities. In Denmark, potential acid sulfate soils were previously surveyed through conventional mapping. In this study, probability maps for potential acid sulfate soil occurrence will be constructed for the wetlands located in Jutland, Denmark (c. 6,500 km^2), using Artificial Neural Networks. More than 8,000 existing soil observations and 18 environmental predictors, including geology, landscape type, land use and terrain parameters, are available as input data. Prediction models based on different training algorithms and network topologies will be assessed. Various selections of soil observations and combinations of environmental predictors will also be tested in order to estimate the optimal amount of soil profiles required to yield satisfactory predictions, as well as the best predictors.

Predicting Soil Processes: Digital Soil Mapping as a platform for bridging scale discrepancies between measurements and predictions

Zamir Libohova[1], Phillip R. Owens[2], Philip Schoneberger[1], Jenette Ashtekar[2], Minerva Dorantes[2], Skye Wills[1] and Hans E. Winzeler[2]

[1]*UDSA-NRCS-National Soil Survey Center, Lincoln, NE, USA*
[2]*Purdue University, Department of Agronomy, West Lafayette, IN, USA*

Inferring soil processes and predicting soil properties have been at the center of soil science since its inception, whether at microscale, horizon, pedon, hillslope or catchment scales. The advancement of soil knowledge throughout centuries has deepen our understanding of soil development with several models developed to describe soil spatially and temporally. There are new opportunities and challenges for soil science due to 1) the emergence of high resolution spatial-temporal resolution data at various scales; 2) higher accuracy demands on natural phenomenon and processes; and 3) the combination of increased storage capacity and computing process capabilities. One of the most formidable challenges for soil scientists is the bridging scales, especially between field measurement data and predictions. We argue that Digital Soil Mapping (DSM) offers a platform for bridging these scale discrepancies and this research will provide two examples for soil depth and saturated hydraulic conductivity (K_{sat}) spatial predictions for distributed hydrological modelling. Soil depth was measured at 95 sites that were stratified by three major slope positions (summit, backslope and toeslope) but randomly allocated within each slope position for a HUC 12 watershed approximately 35 km^2 in size. Among various interpolation methods regression kriging had the highest R^2 = 0.26 while kriging and random forest had the lowest 0.10 and 0.08, respectively with an RMSE = 10 cm for all three models. Approximately 50% of soil depth variability was explained using a soil landscape model derived from terrain attribute analysis utilizing traditional soil surveys knowledge, but with a higher RMSE = 0.33 cm for a soil depth range between 20-200 cm. However, the spatial distribution of soil depth based on soil landscape model derived from the combination of a traditional soil survey and tacit knowledge more accurately represented the soil erosion processes at hillslope scale. The Ksat was measured at a small catchment approximately 5 ha on summit, backslope and toeslope based on major soil horizons using (i) constant head permeameters representing point scale measurements; (ii) wells and (iii) piezometers, representing pedon scale measurements. Additionally, Ksat was measured from the flume discharge representing the small catchment scale. The Ksat variability decreased from point measurement to hillslope to catchment scale, however the mean values were comparable among all measurement methods. Using DSM approaches and the tacit knowledge from the existing soil data and soil landscape models, the Ksat values were up-scaled from the small catchment to the HUC 12 large watershed using DSM techniques. Other soil hydraulic properties including soil depth were up-scaled in the same manner and used as input for daily stream flow predictions based on the Distributed Hydrology Soil Vegetation Model (DHSVM). Nash-Sutcliffe (NS) model efficiency (similiar to R^2) of 0.52 indicated overall a good performance of the model without any calibration. However, NS varied from 0.32 during dry season to 0.72 for winter season. DSM in these studies provided a platform to assess various approaches, from geostatistics to terrain attribute analysis based on soil landscape models, for creating spatial soil data input for distributed hydrological modeling.

Session 12 Digital Soil Mapping
Chair: Mogens H. Greve

Providing spatial SOC estimates for complex and remote soil-landscapes of scarce data availability and structure

Mareike Ließ[1]

[1] *Helmholtz Centre for Environmental Research – UFZ, Department of Soil Physics*

Little is known about the soil organic carbon (SOC) stocks of remote tropical mountain areas. The complex soil-landscapes of difficult accessibility pose a challenge to spatial analysis and prediction: Spatial estimates are difficult to obtain while (1) interpolation methods fail due to the available low soil information density and their inappropriate spatial distribution, and (2) predictor-response correlations concerning the predictors obtained from the only available sources to adjust a SCORPAN model – DEM and Landsat images – are weak. To improve the expected poor spatial prediction results, various regression modelling aspects were tested. Five machine learning algorithms with model tuning via grid search were compared to optimise the model structure: random forest, artificial neural network, multivariate adaptive regression splines, boosted regression trees and support vector machines. Different spatial predictor settings and three predictor selection strategies were used to optimise the predictor input. All model aspects were compared and tested via five simultaneous runs of a tenfold cross-validation. Predictor selection and model tuning improved the models' predictive performance in all five machine learning algorithms. In predictor selection, choosing predictors due to their individual performance was vanquished by the two procedures which accounted for predictor interaction. The boosted regression tree algorithm resulted in the overall best prediction result. SOC stocks ranged between 0.2 to 17.7 kg m^{-2}, displaying a huge variability with diffuse insolation and curvatures of different setting guiding the spatial pattern. The tested approach provides a good means to optain spatial soil information in complex and remote soil-landscapes of scarce data availability and structure.

Classification and mapping of soil pH depth function groups for Denmark

Kabindra Adhikari[1], Budiman Minasny[2], José Padarian[2] and Mogens H. Greve[3]

[1]University of Wisconsin−Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI 53706, United States
[2]The University of Sydney, Faculty of Agriculture & Environment, 2006 NSW, Sydney, Australia
[3]Aarhus University, Department of Agro-ecology, Blichers Allé 20, Tjele 8830, Denmark

Influences of soil pH on soil quality and crop productivity in agricultural soils is well known. The depth functions of soil pH also can infer pedological processes and anthropogenic influences. This study modeled the depth functions of 1934 soil profile pH measurements (1:2.5 soil-water ratio) and mapped its spatial distribution for entire Denmark at 30-m grid spacing. The pH depth function was modeled with equal-area spline function, and predictions at unsampled location were made by rule-based regression kriging that employed soil, terrain, land use, and climate data as factors of pH variations. Model validation on 25% unused data showed a R^2 range of 0.51 (100-200 cm) to 0.57 (15-30 cm), and RMSE of 0.82 to 0.56 for the same depth intervals. Geology, soil types, land use, precipitation and slope were among the key variables affecting soil pH distribution. Average predicted soil pH on the surface (0-5 cm) was 6.27 (SD ±0.91), and that for the bottom depth (100-200 cm) was 6.41 (SD ±1.21). Comparison of these results with the Global SoilGrids revealed that the latter always had lower pH with a narrow range throughout the profile. In addition SoilGrids cannot model soil pH depth functions meaningfully. Based on the pH distribution pattern in soil profiles, seven classes were identified, and were mapped. Range of pH values for those classes were-pH >7 throughout; 5.5 < pH <7 throughout; pH <=5.5 throughout; 5.5 < pH <7 on 0-30, and increase with depth; pH <=5.5 on top and increase with depth; 5.5 < pH <7 on 0-30 and decrease until 60-100, but increase in 100-200; and 5.5 < pH <7 on 0-30 and decrease with depth. These depth functions correspond to known soil forming factors and illustrate the effect of anthropogenic influences.

Object-oriented digital soil mapping for the support of Delineation of Areas with Natural Constraints in Hungary

Katalin Takács[1], József Szabó, Zsófia Bakacsi[1], Tibor Tóth[1], Gábor Szatmári[1], Annamária Laborczi[1] and László Pásztor[1]

[1]Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences

In Hungary a significant amount of soil data is available in different databases or soil information systems, however there are frequent discrepancies between the available and the expected soil information. The tasks of regional planning increasingly demand unusual or more complex information about soils, which cannot be fully satisfied by formerly elaborated spatial soil information systems, therefore the soil data of the previous surveys should be reinterpreted and reprocessed to meet the demands of regional planning activities.

It is a recent task in Hungary to designate Areas with Natural Constraints according to the common European regulation. For the EU member states an objective, science based common criteria system was compiled, which includes criteria referring to climate, soil and topography. Soil related criteria, which refer to basic properties – like pH or rooting depth – can be easily mapped by the data of the most commonly used soil information systems. But there are certain criteria, where the situation is more difficult: there are no direct information in the available datasets, which could be used in proper spatial inference to identify areas concerned and/or they were never targeted to be mapped. Resolving this problem the integration of more data sources and/or the reprocessing of soil survey data are needed.

In this study we present three examples how the requirements of delineation of Areas with Natural Constraint according to the common European biophysical criteria were fulfilled by specific digital soil mapping products based on the reinterpretation of former soil survey information. The criterion maps of soil sandiness, vertic properties and soil salinity were complied. Soil data were fusioned from the most comprehensive Hungarian soil information systems: the Digital Kreybig Soil Information System (DKTIR), the Hungarian Soil Information and Monitopring System (SIMS) and the Hungarian Detailed Soil Hydrophysical Database (MARTHA). The maps were created by applying regression kriging method and by the aid of digital elevation information, satellite images, climatic, geologic and land cover maps. The compiled new maps can satisfy the needs of designation of Areas with Natural Constraints.

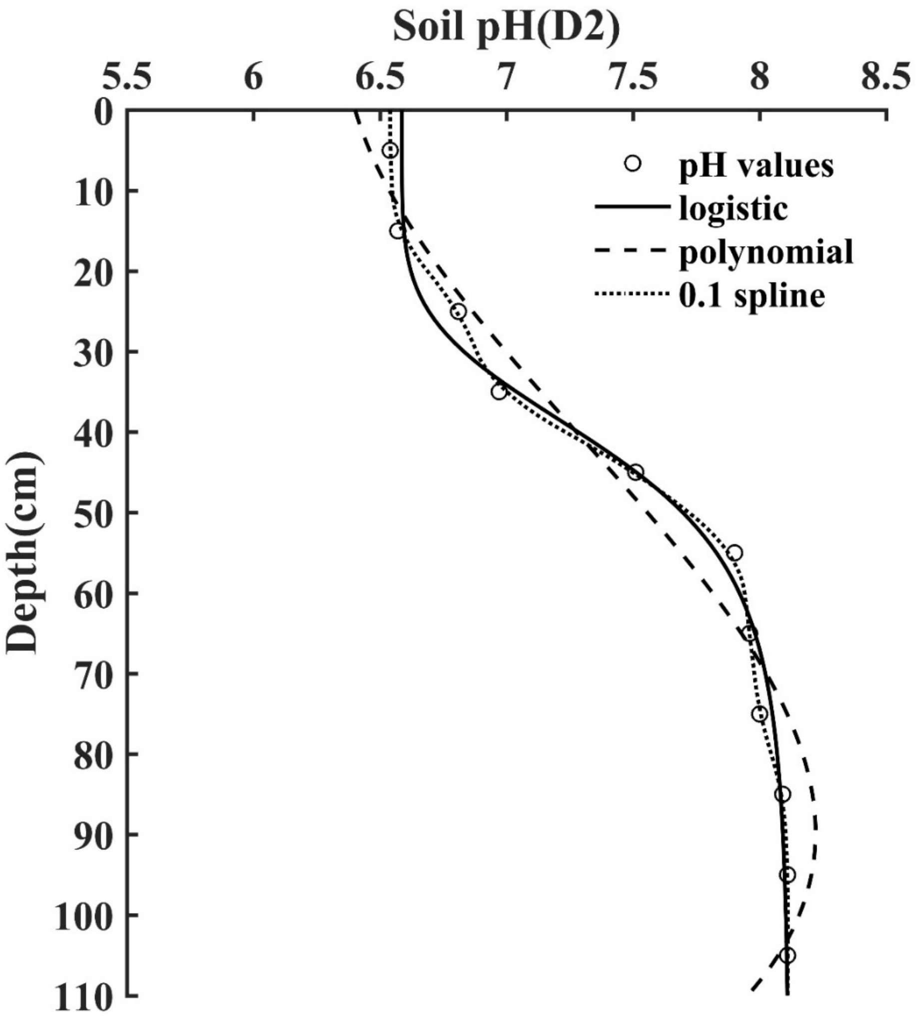
Object-oriented digital soil mapping is proposed to be also applied in other tasks posed by regional planning – Delineation of Areas with Excellent Productivity, spatial concerns of irrigation, drought or inland excess water strategy –, if the target soil property could be clearly defined.

A new pH depth function for agricultural soils

Asim Biswas[1, 2] and Yakun Zhang[1]

[1] Department of Natural Resource Sciences, McGill University, 21111 Lakeshore Road, Ste-Anne-de-Bellevue, Quebec, Canada H9X3V9
[2] School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada N1G2W1

Soil pH controls the availability of the majority of plant nutrients, if not all, and determines the growth environment for plant roots. Sample collection for multiple subsurface layers and subsequent measurement of pH are challenging and time intensive. Some models (depth functions) that are individually fitted to soil profile data have been used to predict soil attribute to continuous depths but lack the generality. Moreover, these functions often lack physical explanation. This research proposes a new depth function to predict pH for a whole soil profile. Soil properties including pH are often similar within the plough layer from mixing during tillage and other agricultural operations. Similarly, soil pH below the root zone tends to be very uniform due to least disturbance, leaving a transition zone from the bottom of the tillage layer to the bottom of the root zone with variable pH. Keeping this physical condition in mind, a closed form equation (depth function) was developed similar to a logistic curve. The function has 4 parameters (with an option to be two) including 1) minimum pH, 2) maximum pH, 3) hillslope parameter representing steepness of the curve that is determined by the length of the root zone, and 4) inflection point representing almost the midpoint of the transition zone or root zone. A total of 32 soil cores down to about 1.1 m depths were collected from an agricultural field of Macdonald farm, McGill University. The sub-samples were taken at every 10 cm and analyzed for pH in the laboratory in soil: water suspension. The lab measured pH was used to test the fitting performance of the logistic function. Additionally, a global dataset with 432 profiles with various soil classes, drainage types, land use, and altitude was also used to test the generality of the new function. The performance of this function was compared with the results of the commonly used 3rd order polynomial regression function and the equal-area quadratic spline function. Improved performance of the new function with better physical explanation showed promise in predicting soil pH at depths over the commonly used ones (Fig. 1). The Spline function had the highest accuracy but lacked a fixed trend. The Polynomial function had good accuracy and displayed a non-monotonous trend, which can also be used as a substitute for some profiles with complex variability.


Soil pH(D2)
5.5
6
6.5
7
7.5
8
8.5
0
10
20
30
40
50
60
70
80
90
100
110
Depth(cm)
pH values
logistic
polynomial
0.1 spline