Using homosoil to enrich sparsed soil data infrastructures globally

A.M. Nenkam (a,b); A.B. McBratney (a); B. Misnany (a); A.M.J-C Wadoux(a);
A.M. Whitbread (a,b); P.C.S. Traore (a,b)

a School of Life & Environmental Science, Sydney Institute of Agriculture, the University of Sydney, Australia
b International Crop Research Institute for the Semi-Arid Tropics (ICRISAT)

Many areas in the world suffer from relatively sparse soil data availability. This results in inefficient implementation of soil related studies and inadequate recommendations to improve soil management strategies. Commonly this problem is tackled by collecting new soil data which are used to update legacy soil surveys. In this paper, we use the homosoils to obtain new soil data for the area of interest. Homosoils are soils that can be geographically distant but with similar soil forming factors. We identify the homosoils for the area of interest by using distance metrics in the space spanned by the environmental covariates depicting soil formation. In a case study in Mali, we found that large areas in India, Southern Africa, Northern Australia, Brazil, Mexico share similar soil forming factors that the African Sahelian band in Africa. We consider these areas as homosoil and test the transferability of the digital soil data by calibrating a digital soil mapping models using data within the homosoil area. These models are applied in our area of interest in Mali where digital maps of soil properties are generated and validated using legacy data. The maps are also compared to publicly available gridded digital soil mapping products. Validation statistics show that the mapping model calibrated on the homosoils can effectively be applied in Mali. The soil maps produced with homosoils data have a similar pattern than the existing maps. The approach developed here shows that opportunity of transferring soil data across the globe to populate areas with relatively sparse soil data. The concept of homosoils is promising and we envision future applications such as transfer of agronomic and experimental results on soil management practices between areas considered as homosoils.

An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams

Alexandre M.J.-C Wadoux (a); Dennis J.J. Walvoort (b); Dick J. Brus (c)
a Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Australia
b Soil, Water and Landuse group, Wageningen Environmental Research, the Netherlands
c Biometris, Wageningen University & Research, the Netherlands

For many decades, soil scientists have produced spatial estimates of soil properties using statistical and non-statistical mapping models. Commonly in soil mapping studies the map quality is assessed through pairwise comparison of measured and predicted values of a soil property, from which statistical indices summarizing the quality of the entire map are computed. Often these indices are based on average error and correlation statistics. In this study, we recommend a more appropriate and effective method of map evaluation by means of Taylor and solar diagrams. Taylor and solar diagrams are summary diagrams exploiting the relationship between statistical indices to visualize differentiable aspects of map quality into a single plot. An important advantage over current map quality evaluation is that map quality can be assessed from the combined effect of a few statistical quantities, not just on the basis of a single index or list of indices. We illustrate the use of common statistical indices and their combination into summary diagrams with a simulation study and two applications on real-world soil data. In the simulation study nine maps with known statistical properties are produced and evaluated with tables and summary diagrams. In the first case study with soil data, change in the quality of a large-scale topsoil organic carbon map is tracked for a number of permutations in the mapping model parameters, whereas in the second case study several maps of topsoil organic carbon content for the same area, made by various statistical and non-statistical models, are compared and evaluated. We show that in all cases better insights in map quality are obtained with summary diagrams, instead of using a single index or an extensive list of indices. This underpins the importance of using integrated summary graphics to communicate on quantitative map quality so as to avoid excessive trust that a single map quality index may suggest.

Modelling measurement error in wet chemistry soil data with linear mixed-effects models

Cynthia van Leeuwen (a); Titia Mulder (b); Niels Batjes (a); Gerard Heuvelink(a,b)
a ISRIC - World Soil Information / Wageningen University;
b Wageningen University;

Wet chemistry measurements on soil properties are subjected to many error sources, such as the observer and lack of standardized methods. Therefore, soil data are imperfect, making us uncertain about the true value. Uncertainties in measurements of soil properties can propagate through pedometric models. Therefore, providing detailed uncertainty information about these data to potential users is important. In practice, uncertainty estimates are, unfortunately, rarely specified.
We aimed to quantify uncertainties in synthetic and real-world pHH2O and Total Organic Carbon (TOC) data. We assumed that measurement uncertainty can be represented by a normal distribution. A linear mixed-effects model was applied to estimate the parameters of the distribution, of both synthetic and real-world datasets. The model included ‘sample ID’ as a fixed effect, and ‘batch’ and ‘laboratory’ as random effects. The use of synthetic datasets allowed us to investigate how well the model parameters could be estimated given a specific experimental measurement design (EMD) with duplicate measurements, whereas the real-world case served to explore if the parameter estimates were still accurate for such unbalanced datasets.
For balanced datasets, using synthetic pH data for three hypothetical laboratories (two batches per laboratory), we imposed standard deviations (σ) of the random effects as σbatch=0.10, σlaboratory=0.24 and σresidual=0.2. These parameter values were accurately estimated from simulated synthetic pH data, although estimation errors increased as the size of the dataset decreased. Changes were made to the EMD by randomly removing 20%, 50% and 80% of the data, resulting in unbalanced datasets. The interquartile range (IQR) of the estimated σ for the batch effect and residual variance strongly increased with a larger percentage of removed data. However, the increase in IQR was larger for n=20 compared to, e.g., n=500. When comparing 0% and 80% randomly removed data, the IQR for the batch effect increased with 171%. Conversely, for n=500 an increase of only 15% was observed. The laboratory effect showed only a gentle increase in IQR.

The same model was fitted on real-world pH and TOC data, provided by the Wageningen Evaluating Programs for Analytical Laboratories (WEPAL). For measured pH, the model estimated σbatch=0.22, σlaboratory=0.17 and σresidual=0.19. Interestingly, σbatch was larger than σlaboratory. This could be explained by the absence of replicate measurements within each batch. To evaluate how accurate σ was estimated given the unbalanced dataset structure, we synthetically generated data, based on sample means and σ derived from the measured data, and fitted the model. The IQRs of the estimated σ from synthetic WEPAL data were 0.061 (batch), 0.062 (laboratory) and 0.059 (residual). The model fitted on the measured TOC data estimated σlaboratory=2.78% and σresidual=5.74%, while their IQRs using synthetic data were 0.59% and 2.40%, respectively.

Accuracy assessment of bare soil map of Hungary generated from mid-term Sentinel-2 data

János Mészáros; Tünde Takáts; Mátyás Árvai; Annamária Laborczi; Gábor Szatmári; László Pásztor
Institute for Soil Sciences, Centre for Agricultural Research, H-1022 Budapest, Hungary

As Earth observation (EO) data is increasing in volume, fast and reliable data-processing tools are also required especially for analyzing large areas with high spatial resolution. Google Earth Engine (GEE) platform provides wide sets of EO imagery and elevation data in a cloud-based processing environment. This research focused on i) the generation of bare soil map of Hungary and ii) the accuracy assessment of created soil maps representing soil texture (clay, sand, silt) and soil organic carbon content (SOC). In this study Copernicus Sentinel-2 optical images acquired on a mid-term time period between 2017 April and 2021 April were used to generate a median composite. Optical images were filtered for cloud coverage less than 50% and a cloud mask was also implemented on all remaining images. The threshold values for Normalized Difference Vegetation Index (NDVI) and Normalized Burn Ratio (NBR) indices were 0.3 and 0.1 respectively to differentiate bare soil pixels. We tested the prediction accuracy of bare soil composite supplemented by various environmental datasets as additional predictor variables in different scenarios: (i) using solely bare soil composite spectral data (ii) composite spectral data, elevation and its derived parameters (e.g. slope, aspect) (iii) composite data and spectral indices and (iv) all aforementioned data in fusion. For validation two types of datasets were used: i) the reference points of the Hungarian Soil Information and Monitoring System with a ten-fold cross-validation method and ii) the recently compiled national maps for soil texture and SOC parameters. Acknowledgement: Our research was supported by the Hungarian National Research, Development and Innovation Office (NKFIH; K-131820) and by the Scholarship of Human Resource Supporter (NTP-NFTÖ-20-B-0022).

Distinguish mineral and organic soil horizons in Histosols and Fluvisols by automated image analysis for DSM

Stefan Oechslin (a); Madlene Nussbaum (a); Peter Trachsel (b); Dylan Tatti (a); Stéphane Burgos (a)
a Bern University of Applied Sciences, School of Agricultural, Forest and Food Sciences HAFL (CH)
b Amt für Landwirtschaft und Natur, Abteilung Strukturverbesserungen und Produktion (CH)

In the lowlands of the Rhine Valley of St. Gallen in eastern Switzerland Histosols and Fluvisols are the dominant soil types. Due to artificial drainage large scale soil settlements are running. To examine these deep and heavily stratified soils, a drill vehicle was constructed with a soil sampler with a two-meter drill core. To efficiently describe and classify the samples, a digital camera with four bands (RGB, NIR) was mounted on the drill vehicle to imaging the whole drill core. For the classification of the local soil and the potential settlement risk in the area, the soil organic matter (SOM) is one of the most important descriptors. Therefore, this article shows the usage of the mount-ed camera to detect organic soil horizons. The results of an automated image analysis by the the ensemble learning method random forest based on the optical properties of the soil, were compared with expert data collected by naked eye on the drilling vehicle. A total of 800 drills with 32'000 images was used. The use of 15 different bands and indices, derived from the RGB image, showed good results, with an accuracy of 0.89 and a Cohen's Kappa of 0.71. The additional use of NIR data led to a higher bias. Due to the uncertainties in the estimation of SOM and the precision of the indicated depths in the expert data, the two methods can only partly be compared. It is assumed that a comparison with more accurate and precise data like multispectral data would show better accordance.
The generated point data was then used to model soil properties maps.

Characterization of podzolic soils using digital morphometrics in a subtropical subalpine forest of Taiwan

Yin-Chung Huang; Zeng-Yei Hseu
Department of Agricultural Chemistry, National Taiwan University

With the advances in the field of digital soil morphometrics, soils can be rapidly characterized by proximal sensors and their pedological features can be further predicted by multivariate analysis and machine learning algorithms. However, the direct linkage between soil attribute and sensor signal is difficult to elucidate with such approach. Thus, the study aimed to build up relationship between soil properties and elemental composition by portable X-ray fluorescence spectrometry (pXRF). Podzolic soils are potential materials for this approach because of remarkable migration of elements through profile. Four podzolic soil profiles (SL1, SL2, SL3 and SL4) from the subalpine forest in central Taiwan were performed for the measurements. The color in pedons SL1 and SL2 met the criterion of spodic horizon (7.5YR 5/6), however no albic horizons existed in the pedons. Although albic horizons were found in pedons SL3 and SL4, their B horizons only clearly accumulated illuvial clay. All horizon samples were subjected to pXRF analysis and principle component analysis (PCA). Elements Fe, Al, Cr, Mn, As, Ni, Rb and Bi were accumulated in the B horizons and corresponding to the accumulation of spodic materials. Significant and positive linear correlations existed between clay and the elements Fe and Rb. Additionally, the first principle component of PCA explained 47.4% of total variance and was useful to separate horizon samples. For next step, the data will be combined with visible-near infrared spectroscopy to elucidate the direct linkage between proximal sensor signals and soil properties.

Five approaches to evaluate map accuracy: 3D soil pH maps at 25m resolution for the Netherlands

Anatol Helfenstein (a); Vera Leatitia Mulder (a); Gerard B.M. Heuvelink (a); Joop P. Okx (b)

a Soil Geography and Landscape Group, Wageningen University, The Netherlands
b Soil, Water and Land Use Team, Wageningen Environmental Research, The Netherlands

Since the establishment of Digital Soil Mapping (DSM) as a research field, the main focus has been on implementing new methods to improve the predictive performance of soil maps. However, there is still an ongoing debate over the best way to evaluate the accuracy of soil maps. This is essential for soil maps to be adopted by a broader community, for future research guidance and to ensure that they are used correctly. We introduce a high-resolution 3D soil modelling and mapping platform for the Netherlands using Quantile Regression Forest (QRF) models that includes different ways to assess map quality. The objective is to compare 5 different approaches of evaluating mapping accuracy using (i) out-of-bag (OOB) observations, (ii) a 10-fold cross-validation grouped by location, (iii) an independent validation set, (iv) a stratified random sampling design separated for each depth layer of the independent validation set and (v) spatially explicit thresholds of the 90% prediction interval (PI) using the GlobalSoilMap (GSM) specifications for Tier 4 products. We provide soil pH [KCl] prediction maps for the Netherlands at 25m resolution at any desired depth between 0m and 2m.

QRF models were calibrated using 15338 soil observations between 0m to 2m depth from 4230 locations and 195 covariates. For approaches i-iv, mean and quantile prediction accuracy were evaluated based on the mean error (ME), root mean squared error (RMSE), model efficiency coefficient (MEC) and the prediction interval coverage probability (PICP). The independent validation dataset for approaches iii and iv consisted of 5677 observations between 0m to 2m depth from 1151 locations. In approach iv, we were able to spatially validate each depth layer separately by adjusting the ME, RMSE and MEC metrics based on strata weights and computing their respective 95% confidence intervals. For approach v, we used the 90% PI to categorize each 25 m pixel into A, AA and AAA quality.

Accuracy plots and metrics for approach i were overly optimistic (ME = 0.01 pH, RMSE = 0.48 pH, MEC = 0.87). When comparing approach ii and iii, 10-fold cross-validation indicated less bias (ME = 0 vs. 0.12 pH), a lower RMSE (0.71 vs. 0.80 pH) and higher fit (MEC = 0.72 vs. 0.67). Spatial validation metrics in approach iv resulted in ME = 0.05 to 0.16 pH, RMSE = 0.70 to 0.79 pH and MEC = 0.73 to 0.82 depending on the depth layer. PICP results revealed that QRF prediction uncertainty is over-estimated. Our Tier 4 GSM maps revealed that approximately 90% of the Netherlands was categorized as A quality or below, while less than 10% and 1% were qualified as AA and AAA, respectively.
Our results show that accuracy evaluation methods that do not take into account a sampling design are not necessarily indicative of the map quality. Furthermore, the performance of our models using accuracy metrics compared to thresholds using 90% PI raises concerns over the achievability of Tier 4 GSM specifications on a national scale.

Soil Mapping Based on Globally Optimal Decision Trees

Arseniy Zhogolev

V.V. Dokuchaev Soil Science Institute

Digital soil mapping can be provided based on the explicit rules of soil delineation. These rules were formulated by the soil-mapping experts for delineation of soil units while the process of traditional soil mapping. In addition to soil mapping, the rules can be used for expert testing of the notional consistency of soil maps, soil trend prediction, soil geography investigations, and other applications. We propose an approach that imitates traditional soil mapping by constructing compact globally optimal decision trees (EVTREE) of soil mapping rules for the covariates of traditionally used soil formation factor maps. We evaluated our approach by regional-scale soil mapping at a test site in the Belgorod region of Russia. The notional consistency and compactness of the decision trees created by EVTREE were found to be suitable for expert-based analysis and improvement. With a large sample set, the accuracy of the predictions was slightly lower for EVTREE (59%) than for CART (67%) and much lower than for Random Forest (87%). With smaller sample sets of 1785 and 1000 points, EVTREE produced comparable or more accurate predictions and much more accurate models of soil geography than CART or Random Forest.

Evaluation of digital soil mapping and soil landscapes relationships in the tropical monsoon sub-climate in Jember District, East Java, Indonesia.

D Cahyana; B Barus; Darmawan; B Mulyanto; and Y Sulaeman

a Department of Soil Science and Land Resource, IPB University, Bogor, Indonesia;
b Indonesian Center for Agricultural Land Resources Research and Development, Bogor, Indonesia

Digital soil mapping studies have been developed and applied as a cost-effective alternative technique to conventional mapping. Most studies in predicting soil classes evaluated the accuracy of the digital map based on the percentage of correctly classified pixels or other global statistical measures. There is a lack of spatial comparison of the actual soil-landscape relationships produced by the digital maps. This study examined the use of digital elevation models and their derivatives in predicting the spatial diversity of soil types in the Jember Regency, East Java Province, Indonesia. We evaluated the accuracy of the map using a fuzzy set-map comparison, and in addition, we evaluated the soil-landscape relationship of the digital maps. The study used 783 training data in the form of polygons in combination with a suite of covariates representing topography, organisms, and soil moisture. The prediction was carried out using machine learning techniques in the form of K-nearest neighbors (KNN), random forest (RF), and Decision Tree (DT). The predicted soil map was evaluated using the fuzzy logic technique to determine the estimation power of the model and the soil-landscape relationship. The study results showed that the global fuzzy matching comparison of the predicted map with the reference map was between 0.299-0.389. A high local fuzzy inference value of 0.687-0.842, was found in the lowland landscapes near the coast in the south and highlands in the mountains in the north. Three digital maps using three covariate combinations generated by the random forest model show that the dominant soil in the lowlands near the estuary and the coast was Typic Endoaquepts, then shifted upstream to Typic Epiaquepts. At the top of the mountain, the soil was predicted as Typic Hapludands and at the bottom was Andic Dystudepts. This pattern showed that the digital map could detect the relationship between soils types formed in local environmental conditions. Evaluating the soil-landscape relationships of digital soil maps can reconcile pedology and digital soil mapping.

Spatio-temporal modelling of soil organic carbon stock in Hungary to support land degradation neutrality assessment

Annamária Laborczi; Gábor Szatmári; János Mészáros; Sándor Koós; Béla Pirkó; László Pásztor
Institute for Soil Sciences, Centre for Agricultural Research

‘Strategic objective 1’ of the United Nations Convention to Combat Desertification (UNCCD) aims to improve conditions of affected ecosystems, combat desertification/land degradation, promote sustainable land management, and contribute to land degradation neutrality. The indicator ‘Proportion of land that is degraded over total land area’ (SO1) is compiled from three sub-indicators: ‘Trends in land cover’ (SO1-1), ‘Trends in land productivity or functioning of the land’ (SO1-2), ‘Trends in carbon stocks above and below ground’ (SO1-3).
Soil organic carbon (SOC) stock can be adopted as the metric of SO1-3, until globally accepted methods for estimating the total terrestrial system carbon stocks will be elaborated. SOC can be considered as one of the most important properties of soil, which shows not just spatial but temporal variability. According to our previous results in the topic, UNCCD default data of SOC stock for Hungary is strongly recommended to be replaced with country specific estimation of SOC stock.
SOC stock maps were compiled in the framework of DOSoReMI.hu (Digital, Optimized, Soil Related Maps and Information in Hungary) initiative, predicted by proper digital soil mapping (DSM) method. Reference soil data were derived from a countrywide monitoring system. The selection of environmental covariates was based on the SCORPAN model. The elaborated SOC stock mapping methodology have two components: (1) point support modelling, where SOC stock is computed at the level of soil profile, and (2) spatial modelling (quantile regression forest), where spatial prediction and uncertainty quantification are carried out using the computed SOC stock values.
We analyzed how SOC stock changed between 1998 and 2016. Nationwide SOC stock predictions were compiled for the years 1998, 2010, 2013, and 2016. For the intermediate years, we do not recommend to calculate SOC stock values, because we have no information on the dynamics of change in the intervening years. Based on the 1998 SOC stock prediction, we compiled a SOC stock map for 2018, using only land use conversion factors, according to the default data conversion values.
According to the elaborated scheme during the respective period, significant changes cannot be detected, only tendentious SOC stock changes appear. Based on our results, we recommend to use spatially predicted layers for all years when data are available, rather than calculating SOC stock change based on land use conversion factors.

Acknowledgment: Our research was supported by the Hungarian National Research, Development and Innovation Office (NKFIH; K-131820) and by the Premium Postdoctoral Scholarship of the Hungarian Academy of Sciences (PREMIUM-2019-390) (Gábor Szatmári).

How well does Predictive Soil Mapping represent soil geography? Preliminary results of an investigation from the USA

David G. Rossiter (a); Laura Poggio (a); Zamir Libohoba (b); Dylan Beaudette (b)
a ISRIC - World Soil Information (NL)
b NRCS (USA)

We developed methods to evaluate the spatial patterns -- not just the statistical evaluations -- of the geographic distribution of soil properties in the USA, as shown in gridded maps produced by global, national, regional and local Predictive Soil Mapping (PSM), and compared them to spatial patterns known from detailed field survey. gNATSGO served as the best field estimate, and these were confronted with SoilGrids v2.0, POLARIS soil properties, and the Soil Properties and Class 100m Grids of the United States at a common 250m nominal resolution. We also compared gSSURGO with POLARIS at 30m nominal resolution. We selected four areas well-known to the authors (central NY, coastal plain NC, southwestern IN, Sierra foothills CA) and a range of soil properties. Preliminary results show large discrepancies. In this talk we will show a few of these and discuss their possible causes. We will also show methods to quantify the spatial patterns and differences between them.

Modeling the effect of urban heat island on the spatial-temporal variation of soil microbial respiration in Moscow megapolis

Vasenev V. (a); Varentsov M. (b); Konstantinov P. (b); Romzaykina O. (c); Kanareykina I. (c); Dvornikov Y. (c)

a Wageningen University
b RUDN University, Lomonosov Moscow State University
c RUDN University

Soil respiration is among the predominant C outflows to atmosphere. The contribution of microbial respiration ranges from 40 to 90% for different biomes and therefore makes a substantial contribution to CO2 emissions from soils. Soil temperature is a key abiotic factor driving microbial respiration, its temporal dynamics and spatial variability. Although the relationship between temperature and microbial respiration are classic and were analyzed in numerous studies, the experience for urban soils is lacking. Urban heat island is a mesoclimatic anomaly resulting in the increase of air and soil temperatures in the central and densely built-up areas in comparison to suburbs and natural references. The research aimed to integrate mesoclimatic modeling to digital soil mapping of soil organic carbon to project the effect of urban heat island on soil microbial respiration in Moscow megapolis. Anomalies in air and soil temperatures in Moscow megapolis were modeled by COSMO-CLM parameterized by TERRA_URB. In result, the 500 m grid maps of soil temperature and soil moisture were obtained. The 30 m grid maps of soil C stocks and contents in 0-10, 10-30 and 30-50 cm were obtained based on the DSM approach. Relationships between microbial respiration in soils with different pH and C contents was calculated based on the laboratory experiments with the set-up soil temperatures (10, 22, 30 and 40 C) and moisture (0.3 and 0.7 from field capacity) and approximated by non-linear regression equation. This equation was used to integrate the modelled maps of soil temperature and moisture and the map of C content in 0-10 cm to analyze spatial variability of microbial respiration during the period May-October 2019. The general pattern was the following: higher respiration was shown for New Moscow and natural areas in the Moscow suburbs, whereas the average respiration in the central part was low. However, in the central parts few green areas and roadsides where high C contents coincided with temperature anomalies were the hotspots of soil respiration. Moreover, the temperature coefficient Q10 which indicates the potential temperature-driven increase in soil respiration were also the highest in these areas. Therefore, green zones and urban lawns in the central part of Moscow are considered the potential source of green house gases emission, which biodegradation potential is strengthened by urban heat island effect.

Multivariate digital soil mapping to support soil quality index mapping in southern France

Angelini, M.E. (a); Lagacherie, P. (b); Heuvelink, G.B.M. (c); Rabot, E. (d); Guiresse, M. (e)

a Soil Institute, INTA, Hurlingham, Argentina;
b LISAH, Univ Montpellier, INRAE, IRD, Montpellier SupAgro, Montpellier, France;
c ISRIC – World Soil Information, Wageningen, The Netherlands;
d INRA, UR0272 Science du Sol, Orléans, France;
e EcoLab, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France

Pedometricians have spent a lot of effort on mapping soil types and basic soil properties. However, end-users typically need a more elaborate soil quality index (SQI) for land management. Soil quality indices are typically derived from multiple individual soil properties, by evaluating whether specific criteria are met. If this is based on individually mapped soil properties then an important problem is that cross-correlations between soil properties are ignored. This makes it impossible to quantify the uncertainties associated with the mapped indices.

Our objective was to map a soil quality index over a 12 125 km2 study region located along the French Mediterranean coast to help urban planners preserve soils of highest quality. The index considered the ability of soils to fulfill four functions: 1) production of a physical habitat for plant growth; 2) production of a chemical habitat for plant growth; 3) retention and transfer of water and pollutants, and 4) carbon sequestration, under five land use scenarios: 1) annual crop; 2) perennial crop; 3) pastures; 4) forest; and 5) shrubland. Each soil function fulfillment for a given scenario was represented by a categorical map defined from a set of conditions involving basic soil properties (CEC, organic carbon, clay, silt, sand, pH, soil depth, and coarse fragments), and was represented by a 0/1 value. The final soil quality index was the sum of these values.

A regression co-kriging model was developed that, first, mapped separately the basic soil properties from legacy soil data and spatial soil covariates using a Random Forest algorithm and, then, interpolated the residuals using cokriging and the linear model of coregionalization (LMC). Both correlations between different soil properties and between the same soil property for different depth layers were accounted for. The mapping uncertainties of soil properties were propagated by calculating the soil quality index over 300 stochastic simulations of soil properties derived from the LMCs. The final soil quality index and its associated uncertainty were estimated respectively by the mean and standard deviation across the 300 simulations. All resulting maps were at 25 m spatial resolution. For validation we used a cross-validation approach repeated 20 times.

Although the final map was pedologically meaningful, its performance in terms of amount of variance explained (AVE) was low. The simulations were able to reproduce the width of the observed distribution, although the shapes of the distributions differed considerably. However, we envisage some ways for improvement, such as changing soil property thresholds that are hard to predict, creating probability maps instead of the mean from simulations, and changing the prediction support from point to area.

Digital soil mapping of soil available water capacity from legacy soil data: a test in an irrigated perimeter in Southern France

Quentin Styc (a) ; François Gontard (b); Philippe Lagacherie (a)
a LISAH, Univ Montpellier, INRAE, IRD, Institut Agro, Montpellier, France ;
b BRL Exploitation, Nîmes, France ;

In spite of the recent effort to digitize the soil data legacy from the past decades, an important amount of those, mainly in physical display regrouping both profile and auger hole soil observations, remained unexploited. Digital soil mapping (DSM) studies has shown that legacy data can be used to produce soil property maps. However, those studies only focus on the production of soil map at both regional and national scale, using a small part of the information available in a selected territory which, therefore, limit their accuracy.

The aim of this study was to test if an approach of DSM for soil available water capacity (SAWC) in an irrigated perimeter of the land management company of “Bas Rhône et du Languedoc” (BRL) (6,636 km2, Languedoc plain, Occitanie region) is feasible. From 1957 to 1992, the soil prospection of this perimeter was realized at a high-density sampling (3.8 soil profiles/ km2 and 30.6 auger holes/km2). The production of a high-resolution SAWC map for an irrigation company such as BRL is essential, at the same time, to anticipate and optimize the water consumptions of irrigating farmers but also, to plan the required evolution of the water distribution network.

The test was set in the city of Bouillargues (Gard department) for mapping SAWC at several depths (30 cm, 60 cm, 100 cm and maximal soil observation depth). We used the quantile regression forest algorithm calibrated with 69 soil profiles and 2781 auger hole of the study area and validated by using a 10-fold cross validation iterated 20 times. The uncertainty was as well estimated by a propagation error using first-order Taylor analysis, required for the SAWC since it evolves several uncertainties from each of the SAWC components.

The results of this study show that using soil profiles and auger holes from legacy soil harvesting campaign highly improve both the spatial resolution and the accuracy of SAWC maps (up to 70 % of the variance explained) and its associated uncertainty. However, the characterization errors of the SAWC, especially on auger holes, remain a major limiting factor. Those results confirm the feasibility of a DSM approach to enhance the legacy dataset which remains undigitized, to answer to the local farmer and to the land management requirements. Nevertheless, an important effort should be considered to set an automation in digitizing the legacy dataset in order to reduce the cost of it.

Global mapping of volumetric water content at 10, 33 and 1500 kPa using the WoSIS database

Maria Eliza Turek(a,b); Gerard Heuvelink(a,c); Niels H. Batjes(a); Luis M. de Sousa(a); Robson A. Armindo(d); Quirijn de Jong van Lier(e); Laura Poggio(a)

a ISRIC - World Soil Information, Wageningen, The Netherlands
b Federal University of Paraná, Curitiba, PR, Brazil
c Wageningen University & Research, Wageningen, The Netherlands
d Federal University of Lavras, Lavras, MG, Brazil
e University of São Paulo, Piracicaba, SP, Brazil

Soil water content is a key property for modelling the water balance in hydrological, eco-hydrological and agro-hydrological models. Currently available global maps of soil water content are mostly derived from pedotransfer functions (PTFs) applied to maps of other basic soil properties. We developed global maps of the volumetric water content at 10, 33 and 1500 kPa by direct mapping using measured soil water content data extracted from the WoSIS Soil Profile Database together with data estimated using a random forest pedotransfer function based on silt, clay, organic carbon, and soil pH, as well as environmental covariates describing vegetation, terrain morphology, climate, geology, and hydrology. The preparation of the input soil water content data consisted of the verification of available volumetric water content data and conversion of gravimetric to volumetric data using measured and estimated bulk density. In total we used 7292, 33 192 and 42 016 soil water content observations at 10, 33 and 1500 kPa, respectively, and complemented the dataset with 436 108 estimated observations at each tension. Around 200 covariates were prepared as candidate predictors for global mapping. After covariates de-correlation, selection, model tuning and cross-validation and final model fitting for 3D spatial prediction, global results were presented with uncertainty quantification. Tenfold cross-validation yielded a Root Mean Square Error (RMSE) of 6.380, 7.112 and 6.485 10-2 cm3cm-3, and a Model Efficiency Coefficient of 0.430, 0.386, and 0.471, respectively, for 10, 33 and 1500 kPa. The results were also compared to three published global maps of water retention to evaluate differences between direct and indirect mapping approaches. Evaluation points were obtained by sampling the maps at an equally distributed hexagonal grid. Direct mapping was superior to all three indirect mapping approaches for 33 and 1500 kPa, while for 10 kPa it performed worse than one of the three indirect mapping approaches. This may be due to the limited number of soil water content observations available at 10 kPa. Directly developing global maps of soil water content, with associated uncertainty, is a novel approach for this type of properties, and contributes to improving global soil data availability and quality.

Mapping soil carbon sequestration across Argentina and Mexico using Roth C

Veronica Reinoso (a); Franco Daniel Frolla (b);, Sol Ortiz (a); Marcos Esteban Angelini (c), Areli Cerón (a); Marcelo Javier Beltrán (c); Verónica Bunge (a); Guillermo Ezequiel Peralta (d); Luciano Elias Di Paolo (d); Darío Martín Rodríguez (c); Guillermo Andrés Schulz (c); Juan Velazquez (a); Carla Pascale Medina (d); Mario Guevara (e)

a Secretaría de Agricultura y Desarrollo Rural, México
b Instituto Nacional de Tecnología Agropecuaria - Bordenave, Argentina
c Instituto Nacional de Tecnología Agropecuaria - Soil Institute, Argentina
d Global Soil Partnership Secretariat - FAO, Rome
e Centro de Geociencias, Universidad Nacional Autónoma de México

There is a pressing need to support sustainable soil management initiatives to mitigate climate change. A large challenge is to identify areas with the potential for soil organic carbon (SOC) sequestration with a national perspective.We predict SOC and SOC sequestration rates for the years 2020- 2040 across Argentina and Mexico. The Roth C model is adapted using global and national sources of information across 1km grids. Our methodological framework follows the specifications recently suggested by the Global Soil Partnership of the FAO for the development of the Global Soil Organic Carbon Sequestration Map (to be launched in 2021). This framework includes the prediction of SOC in the future (20 years) considering a business as usual model and three hypothetical scenarios of annual plant carbon inputs (by 5, 15 and 20%) associated with the potential implementation of sustainable soil management practices. We observe in Argentina and Mexico that agricultural systems are currently a source of CO2 rather than net sink. According to the Roth C model, to achieve SOC neutrality in Argentina and Mexico and change the situation from being a sink of SOC rather than a source, it is necessary to implement a SOC sequestration scenario with a rate of over 10% of SOC gain in the next 20 years. These results suggest that achieving a positive SOC balance in the future would require more time than 20 years, as increasing SOC by 10% during this period of time can be easily achieved under humid temperate regions but unrealistic for semiarid regions. Yet, our results also suggest that the adoption of sustainable soil management practices could reduce SOC emissions from soils to the atmosphere significantly. Over 50% of SOC emissions predicted by the Roth C model can be prevented across both countries by adopting sustainable soil management practices across agricultural land. Thus, our results contribute as a benchmark to assess the impact of global environmental change on SOC sequestration rates with a national perspective. These results provide the means to spatially observe trends of SOC to identify intervention areas that require sustainable soil management practices to prevent SOC losses. This coordinated effort and collective learning between Argentina and Mexico will boost the capacity of both countries to enable the monitoring of SOC and identify intervention areas to support SOC conservation efforts.

Spatiotemporal modeling of soil organic carbon at coarse and high spatial resolution: a framework for long-term monitoring of soils

Tomislav Hengl; Leandro L. Parente
OpenGeoHub foundation, Wageningen

We are developing methods to model and predict soil organic carbon (content and density in kg/m3) using spatiotemporal machine learning. For the coarse resolution data (5-km spatial resolution) we use a time-series of monthly land cover, NDVI, climatic, land use (HYDE dataset) images covering the period 1982-2016. For the fine-resolution data we develop methods to model soil organic carbon dynamics for Europe for the period 2000-2020; as covariates we use cloud-free and gap-filled Landsat seasonal quantiles of reflectances (P25, P50 and P75), NDVI, night-light images and terrain variables. For spatiotemporal machine learning we use an Ensemble framework (mlr3) combining Random Forest (ranger), classification trees (Cubist), Lasso (glmnet), and gradient boosting (mboost). The results show promising cross-validation accuracy (spatial 5-fold CV) with R-square close to 0.6 for the global data and 0.71 for fine-resolution data. Overall, rainfall and temperature (MODIS LST) come highest on the variable importance list, while from the Landsat products NDVI and green band come as the most important. Spatiotemporal modeling of soil organic carbon for longer periods seem to be currently limited due to two main reasons: (1) limited training datasets exists where repeated measurements are available so that changes in SOC can be potentially correlated to changes in land cover / reflectances, (2) covariates that help explain past soil formation processes (for up to 500 yrs ago) are difficult to prepare and use.

Using advanced Bayesian methods to predict soil properties

Tabassom Sedighi; Jacqueline Hannam
Centre for Environmental and Agricultural Information, School of Water, Energy and Environment, Cranfield University, Cranfield, BEDS, UK.

In this research, Bayesian methods (Bayesian networks and Gaussian process) are presented for predicting UK soil properties (carbon, PH, and texture) in complex and open soil systems. Bayesian methods are used as simulation tools for effective processing of the complex system outcomes by probability propagation methods. This permits evaluation and potential intervention in complex soil systems and determines the dependencies between different variables. We have used Bayesian methods to identify key factors in predicting UK soil properties. Then, we explored the relationships between different key factors such as geographical, environmental or climate and their roles individually in predicting soil properties, particularly to identify those which have the highest impact. The proposed Bayesian methods were also used to calculate the effectiveness of the interventions where the uncertainties associated with the causal relationships at the same time. This approach works with data from a variety of sources and handles a mix of subjective and objective data and can incorporate variables that differ across the contexts. By predicting the soil property values which were evaluated by the domain experts and limited experimental information, the proposed Bayesian methods have proven to be the promising decision support tools to identify the critical factors in changing soil property values and the valuable tools for predicting them when physical sampling is not available.

Artificial Neural Networks for Global Soil Mapping

Giulio Genova (a,b,c) ;Luis M. de Sousa (a) ;Laura Poggio (a)
a ISRIC - World Soil Information - Wageningen (NL)
b Free University of Bolzano, Faculty of Science and Technology, Bolzano/Bozen, Italy
c Eurac Research, Institute for Alpine Environment, Bolzano/Bozen, Italy;

Reliable soil maps are crucial to face several challenges such as reducing soil erosion, climate change adaptation and mitigation, ensuring food and water security, and biodiversity conservation planning. Digital Soil Mapping (DSM) relies on algorithms like Random Forest (RF) to produce maps of key soil properties. Artificial Neural Networks (ANN) have not been yet fully exploited in DSM but recent studies show their potential benefits in improving the accuracy of the predicted maps. In this study, we tested different ANN architectures on a global top-soil dataset of ca. 110.000 samples, comparing the results for the different architectures with the more traditional approach of RF. We predicted soil pH, Soil Organic Carbon, Sand, Silt, and Clay at a global scale. We selected 23 environmental covariates from a pool of over 400 to represent the most important pedogenetic factors. We tried simpler architectures (single input – single target) using point observations for one target variable with corresponding raster cell values for spatially explicit environmental covariates. We also used more complex architectures (multi input - multi target) to incorporate contextual information surrounding an observation (convolutional) and multiple target variables. Using the contextual information around the point with an image of 3x3 pixels (patch size 3) performed best for most of the target variables. Bigger patch sizes create a smoothing effect on the map patterns, thus substantially affect the mapping results. The overall prediction accuracy of the ANN was comparable with the RF model. ANN is a promising, new approach for Global Digital Soil Mapping but further research is needed to improve performance in order to be a viable alternative to other machine learning techniques.

How useful can mid-IR spectra be for direct applications in soil survey and classification?

Fabrício da Silva Terra; José Alexandre Melo Demattê; Ingrid Horák-Terra
Federal University of Jequitinhonha and Mucuri Valleys; University of São Paulo; Federal University of Jequitinhonha and Mucuri Valleys;

How applicable can mid-IR reflectance spectroscopy be in soil survey, classification, and mapping, but not simply as input data for soil properties quantification? We tried to answer this question based on the hypothesis that soil samples and profiles can be efficiently characterized and distinguished only by their mid-IR spectral behaviors. So, our objectives were to cluster soil samples according to their mineralogical, chemical, and physical attributes, and to cluster soil profiles according to their pedogenetic processes and differences of these attributes in depth. A soil database with 1,259 samples of surface and subsurface horizons from 458 soil profiles was used. The predominant soil classes are Rhodic Ferralsol (40.51%), Rhodic Nitisol (15.58%), and Rhodic Acrisol/Lixisol (12.18%). Other classes has frequency lower than 10%. The Nicolet 6700 Fourier Transform Infrared spectrometer equipped with Smart Diffuse Reflectance was employed to scan the dry soil samples (≤ 200 μm). Reflectance spectra were converted into absorbance and were correlated with those normalized soil attributes to identify and quantify their effects along the mid-IR range. Mean-centered principal components analysis was applied to the absorbance spectra, and its scores were clustered by the Fuzzy K-means algorithm in order to distinguish the soil samples. For soil profiles clustering, the same fuzzy methodology was applied, however, the weighted average of the scores was calculated taking into account the thickness and numbers of horizons from each one. The suitable number of clusters and scores for both procedures was assessed as follows: partition coefficient, partition entropy, modified partition coefficient, mean silhouette index, fuzzy silhouette index, Xie and Beni index, and membership lower than 0.5. The high number of absorptions in mid-IR enabled a proper characterization of first-order soil properties related to the spectra. It could be observed to sand, silt, and clay contents due to silica/quartz and 2:1 and 1:1 minerals features, to organic content due to organic compounds features, and to oxides (sulfuric acid digestion) due to features of phyllosilicate clay minerals and Fe and Al oxides. For soil samples, 8 clusters were obtained according differences in particle size distribution, chemical properties, organic content, and weathering levels (mineralogy), corroborated by the values of the following properties for each cluster: clay, organic carbon, cation exchange capacity, pH, clay activity, and kr index. For soil profiles, 24 clusters were obtained grouping soils based on pedogenetic processes, evolution, mineralogy, and number of horizons. Ferralitization and lessivage/argiluviation were the main processes driving spectral distinctions, but some clustering mistakes occurred. Answering our question, yes, mid-IR reflectance spectroscopy is a useful and reliable tool for pattern recognition of soils and direct applications in pedological assessments.