Issue 37, July 2015

Chair: Budiman Minasny

Vice Chair: Lin Yang

Editor: Jing Liu

ΠΕΔΟΜΕΤRON

Newsletter of the Pedometrics Commission of the IUSS

From the Chair

Greetings from Madison.

I have spent the last 2.5 months here at University of Wisconsin, Madison and I am also about to return home. Early last month we had the Digital Soil Morphometrics workshop. Pierre wrote a nice report about it in this issue. It is interesting that in the 1990s, soil scientists were encouraged not only to look and measure the profile, but there is much spatial variability that need to be captured. Now that we know much more about soil spatial variation and able to map it efficiently, we need to look back in the profile.

At the workshop, there are several presentations that looked at detailed sub-mm to microscale resolution of a soil profile using digital camera, infrared, and laser imaging. We observed fine scale variability within a profile and most properties vary smoothly with depth. Are the horizons that we observed a simplification of reality? Questions were also asked, how do we sample a profile? Do we just simply take measurement along at a fixed interval along a transect. How about the lateral variation of a profile?

My colleague Alfred takes the analogy of mapping a profile wall as mapping the landscape. Can spatial soil sampling theory be applied on a profile wall? Would you take random sample across a profile wall? We now rarely take sample across a transect to map the soil, so why do we still take transect sample in a profile? There is also discussion about what is the operational size of a pedon. Some said it is 1 m^2. Now that we are able to measure properties more efficiently using proximal sensors, it is time to go back to the profile and quantify more accurately the basic soil unit that we study.

The organisation of Pedometrics 2015 is on its way, and we got quite a good response And even before we start Pedometrics 2015, the planning for Pedometrics 2017 is underway. Gerard Heuvelink and colleagues kindly agreed to host the 2017 conference in Wageningen. The first pedometrics conference was held in Wageningen in 1992, and we will celebrate Pedometrics Silver Anniversary in 2017.

I hope to see you all in beautiful Cordoba in September.

Budi

Madison, July 2015.

Inside this issue

Pedometrics 2015

Thanks for submitting your abstracts. We have a good response and received more than 130 abstracts. We hope that we will have a great conference. If you have submitted an abstract, you should receive a notification email regarding the acceptance.

Student Awards

We are glad, for the first time, we are offering student awards to the conference, the awarded students will receive free registration to present their work (worth 300 Euro).

The pedometrics commission student award goes to:

- Helen Metcalfe, Rothamsted Research, UK. "Use Of An Unbalanced Nested Sampling Scheme To Reveal Scale-Dependent Variation In Soil Properties"
- Jason, Ackerson, texas A&M University, USA, "Continuous depth profiles of soil clay content from penetrometer-based in-situ visible near infrared spectroscopy"

And the IUSS also provides awards to:

- Boniface Massawe, from Tanzania, "Mapping Numerically Classified Soil Clusters Of Kilombero Valley Using Machine Learning"
- Benito Bonfatti, from Brazil, "A mechanistic model to predict soil depth in a plateau area of Rio Grande do Sul, Brazil"

Keynote Speakers:

- Gerard Heuvelink - Richard Webster Medal Winner 2014: The uncertain soil in an uncertain future
- Titia Mulder - Winner of Best Pedometrics Paper 2013: Pedometrics and large-extent digital soil mapping applications
- Phillippe Baveye - Kodak Professor, Rensselaer Polytechnic Institute: How research on microscale processes and ecosystem services leads to a fundamental rethinking of soil measurements
- Jaime Gomez-Hernandez - Professor of Hydrogeology, Universitat Politècnica de València: Ensemble Kalman filters for data assimilation in soil science
- Jed Kaplan – Professor and Team Leader Atmosphere, Regolith, Vegetation at University of Lausanne: How can pedometrics improve the limited representation of soils in global earth system models?
- Cristine Morgan – Professor of Soil Science, Texas A&M University: Relevance of Pedometrics to Global Soil Security
- Richard Webster - Rothamsted Research: From Beckett to Krige: how Pedometrics took off

Register Now

Register now for Pedometrics 2015 in Cordoba, September 15th-18th, 2015.

More info at: https://sites.google.com/site/pedometrics2015/home or www.pedometrics.org

Vote for 2014 Best Paper

D G Rossiter, Chairman Pedometrics Awards Committee

e-mail: dgr2@cornell.edu

Dear fellow Pedometricians,

The Pedometrics Awards committee for the best paper award (Grunwald, McBratney, Oliver, Rossiter, Yang) received a good response to our call for nominations, namely 18 interesting and relevant papers. These were scored by the committee and the top five are now presented for your reading pleasure and evaluation.

Although we received nominations for papers in eight journals, the five rated best by the committee were all from European Journal of Soil Science (2) and Geoderma (3). Two papers deal mainly with sampling, one with improving existing soil maps, one with 3D mapping of soil properties, and one on numerical soil classification.

There is a nice mix: geostatistics, sampling design, a pedometrics computation toolkit, spatial scaling, and numerical methods for spectroscopy. All are quite novel in their own way, and will surely stimulate and educate the reader – but of course many of you will have already read the papers when they appeared back in 2014.

The award will be presented at **Pedometrics 2015, 14-18 September in Córdoba (E).** Since many of you take some time off in August, please send in your votes by 31 July 2015. This gives you three months to read these excellent papers.

Please rank the papers in the "instant runoff" system: first choice, second choice… up till the last paper you are willing to vote for, i.e., the last paper that you think would deserve the award. Votes should then be sent to me (dgr2@cornell.edu) from a traceable e-mail address (to prevent over-voting). I will apply the "instant runoff" system[1] to determine the winner. A co-author may vote for her/his own paper(s).

The papers are listed here in order of DOI (so pedometrics is becoming bibliometric). (Papers published in Geoderma is Now OPEN ACCESS until end of July.)

1. Odgers, N. P., Sun, W., McBratney, A. B., Minasny, B., & Clifford, D. (2014). Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma*, *214–215*, 91–100. http://doi.org/10.1016/j.geoderma.2013.09.024

Abstract

Legacy soil maps typically consist of a tessellation of polygon soil map unit delineations where the map units consist of a defined assemblage of soil classes assumed to exist in more-or-less fixed proportions. There are several limitations in this kind of mapping approach that relate to the original intent of the soil survey, the effect of mapping scale, and the nature of soil polygon boundaries. Yet perhaps a more fundamental limitation is the fact that most of the time, the soil classes that comprise the soil map units are not mapped individually: in effect their spatial distributions are unknown beyond the qualitative indications given in the accompanying soil map unit report.

Spatial disaggregation of soil map units attempts to map the spatial distribution of the individual soil classes that comprise a legacy soil map. We developed an approach called "Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees" (DSMART). DSMART samples the polygons of a legacy soil map and uses classification trees to generate a number of realisations of the potential soil class distribution. The realisations are then used to estimate the probability of occurrence of the individual soil classes. These estimates are mapped as raster grids, which can overcome some of the limitations of mapping scale and polygon boundaries inherent in the original legacy soil map.

We demonstrate the DSMART approach on a legacy soil map from the former Dalrymple Shire in central Queensland, Australia. We mapped the estimated probability of occurrence of the 71 soil classes in the legacy soil map, as well as the most probable soil class, second-most-probable soil class and the degree of confusion between them as determined by a confusion index. Validation on 285 observed soil profiles indicated that for 48.4% of the validation profiles, the observed soil class was identified in the top three most probable soil classes.

2. Hughes, P. A., McBratney, A. B., Minasny, B., & Campbell, S. (2014). End members, end points and extragrades in numerical soil classification. Geoderma, 226–227, 365–375. http://doi.org/10.1016/j.geoderma.2014.03.010

Abstract

Soil classification has progressed with the introduction of computers in the mid 20th century to the point where algorithms can be used to organise soil information into clusters that correspond with soil classes. Algorithms such as fuzzy-k means perform well, but can be biased by extreme data. Fuzzy-k means with extragrades was devised to accommodate this problem but estimating the amount of extragrades can be challenging and can lead to dubious classifications. The idea of end members is discussed and it is concluded that end points, observations that represent the most extreme parts of the soil continuum, are useful in the identification of extragrades. We present and discuss a new clustering algorithm, akromeson which identifies extreme points in a given data set and converts them into pseudo clusters, which are then run concurrently with a semi-supervised fuzzy-k means algorithm. We constructed a synthetic data set in order to com-

pare this new method to fuzzy-k means and fuzzy-k means with extragrades. It was able to correctly fix the positions of the centroids, (which was beyond the capacity of fuzzy-k means), and correctly estimated which of the data were genuine extragrades, outperforming fuzzy-k means with extragrades. We then evaluated the performance of akromeson on a data set from the Edgeroi region of New South Wales, Australia. The algorithm identified an extreme cluster on the periphery of the data, and a method was determined on how to use this new method to routinely find clusters. The ability to efficiently cluster data may provide an added advantage to pedologists generally and to stakeholders when they are assessing land use practices, especially in regard to areas which exhibit extreme soil properties that require careful management, which this algorithm is capable of detecting .

3. Poggio, L., & Gimona, A. (2014). National scale 3D modelling of soil organic carbon stocks with uncertainty propagation - An example from Scotland. Geoderma, 232, 284–299. http://doi.org/10.1016/j.geoderma.2014.05.004

Abstract

The variation of soil properties down a profile is usually considered continuous. The aim of this study was to develop and test a methodology to model the continuous vertical and lateral distributions of SOC stocks in Scottish soils making explicit the modelling and spatial uncertainty of the results. A comparison with regression kriging and other depth function methods is provided to show that better performances can be achieved taking into account non-linear relationships between covariates and soil properties. The analysis was run for the whole of Scotland. The carbon stocks were calculated for each point, i.e. each horizon in each available profile. The stock value at each cell for each of the considered depth layers was defined using a hybrid GAM-geostatistical 3D model, combining: 1) the fitting of a GAM to estimate the trend of the variable, using a 3D smoother with related covariates; and 2) kriging or Gaussian simulations of GAM residuals as spatial component to account for local details. The use of GAM makes the approach flexible, because it is able to deal with both linear and non-linear relationships between soil properties and the considered covariates. The results confirmed that MODIS data are a useful source of information for DSM especially at national scale. When comparing the proposed approach with similar methods such as regression kriging, the results showed better agreement with the data in the validation set with a global R2 of 0.60. The median values obtained are comparable with the values reported from previous studies on stocks in Scotland using different methods. The uncertainty is large indicating a wide range of credible values for each pixel.

4. Brus, D. J. (2014). Statistical sampling approaches for soil monitoring. European Journal of Soil Science, 65(6), 779–791. http://doi.org/10.1111/ejss.12176

Abstract

This paper describes three statistical sampling approaches for regional soil monitoring, a design-based, a model-based and a hybrid approach. In the model-based approach a space-time model is exploited to predict global statistical parameters of interest such as the space-time mean. In the hybrid approach this model is a time-series model of the spatial means. In the design-based approach no model is used: estimates are model-free. Full design-based inference requires that both sampling locations and times are selected by probability sampling, whereas the hybrid approach requires probability sampling of locations only. In a case study on soil eutrophication and acidification, a rotational panel design was implemented with probability sampling of locations and non-probability sampling of times. The hybrid and model-based predictions of the space-time means and trend of the mean for pH and ammonium at three depths in the soil profile were very similar. For pH the standard errors of the space-time means were about equal, but for ammonium the full model-based predictor was more precise than the hybrid predictor. For soil monitoring I advocate the selection of sampling locations by probability sampling so that the statistical inference approach is flexible. Selecting locations by a self-weighting probability sampling design ensures that the model-based predictor is not affected by selection bias.

5. Lark, R. M., Rawlins, B. G., Robinson, D. A., Lebron, I., & Tye, A. M. (2014). Implications of short-range spatial variation of soil bulk density for adequate field-sampling protocols: methodology and results from two contrasting soils. European Journal of Soil Science, 65(6), 803–814. http://doi.org/10.1111/ejss.12178

Abstract

Soil bulk density (BD) is measured during soil monitoring. Because it is spatially variable, an appropriate sampling protocol is required. This paper shows how information on short-range variability can be used to quantify uncertainty of estimates of mean BD and soil organic carbon on a volumetric basis (SOCv) at a sampling site with different sampling intensities. We report results from two contrasting study areas, with mineral soil and with peat. More sites should be investigated to develop robust protocols for national-scale monitoring, but these results illustrate the methodology. A 20 × 20-m2 monitoring site was considered and sampling protocols were evaluated under geostatistical models of our two study areas. At sites with local soil variability comparable to our mineral soil, sampling at 16 points (4 × 4 square grid of interval 5 m) would achieve a root mean square error (RMSE) of the sample mean value of both BD and SOCv of less than 5% of the mean (topsoil and subsoil). Pedotransfer functions (PTFs) gave predictions of mean soil BD at a sample site, comparable to our study area on mineral soil, with similar precision to a single direct measurement of BD. On peat soils comparable to our second study area, the mean BD for the monitoring site at depth 0–50 cm would be estimated with

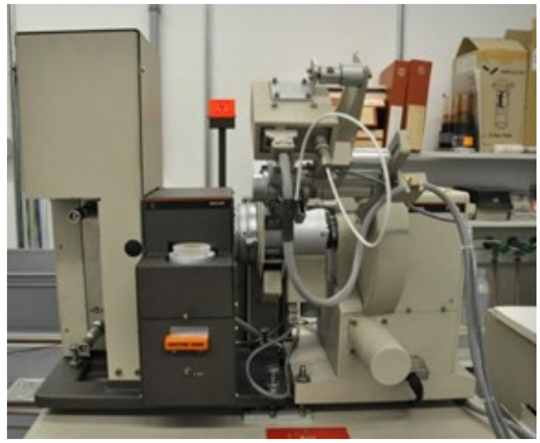
Coupling spectral deconvolution and regression tree analysis for quantifying mineral abundances

Titia Mulder,
Infosol Unit at INRA Centre d'Orléans

First of all, I want to thank you all for rewarding the work of my co-authors and me with the Pedometrics "Best paper award 2014" for the paper entitled Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1-2.4 µm) and regression tree analysis. This research was part of my PhD entitled "Spectroscopy-supported Digital Soil Mapping". Here, it was demonstrated that remote and proximal sensing can support soil mapping surveys in the sampling design, obtaining estimates of soil properties and mapping the spatial distribution of soil properties. The awarded paper showed that proximal sensing can be used to obtain soil property information, i.e. quantitative estimates of the dominant mineralogy of soil samples.

I was in the fortunate position to work together with several specialists in the field of mineralogy (Dr. Christian Mavris and Prof. Markus Egli, Department of Geography, University of Zürich, and Dr. Michael Plötze, ClayLab, Institute for Geotechnical Engineering, ETH Zürich; Switzerland), remote and proximal sensing (Prof. Michael Schaepman, Remote Sensing Laboratories, Department of Geography, University of Zürich, and Raymond Kokaly, U.S. Geological Survey, Denver, USA) and statistics (Dr. Sytze de Bruin, Laboratory for Geo-Information Science and Remote Sensing, Wageningen University, The Netherlands). All people devoted their time in order to collect, analyze and advance the research concerning quantifying mineral abundances within soil samples (Fig. 1).


(a)


(b)

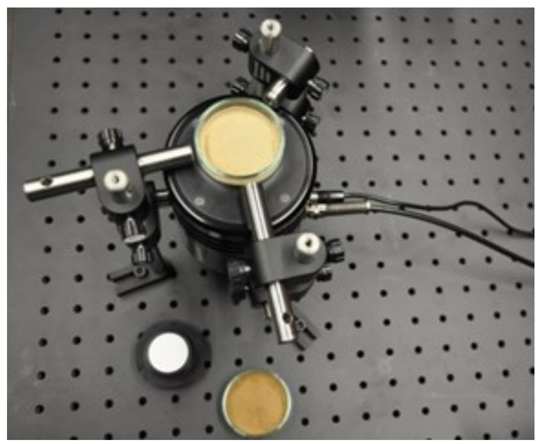

(c)

Figure 1: Data collection and analysis. a) Fieldwork Morocco: soil profile description; b) Laboratory measurements: X-ray diffraction; c) Laboratory measurements: Spectral measurements

Traditionally in environmental and geological studies, the characterization (and quantification) of soil mineralogy is typically achieved using X-ray diffraction (XRD). However, spectroscopy has proven to be an efficient alternative for the determination of various soil properties, including soil mineralogy. The combined application of both X-ray diffraction and Fourier Transform Mid Infrared (MIR) spectroscopy has been successfully used for the characterization of both parent material and soil clay forming processes (Mavris et al. 2011). The use of Visible Near Infrared and Shortwave Infrared (VNIR/SWIR) can provide critical structural information on soil minerals. In this paper we demonstrated its use for simultaneous quantification of mineral abundances from complex mixtures.

Detection of minerals having absorption features within the 0.350–2.500 µm VNIR spectral range have been successfully obtained using linear spectral unmixing techniques. However, these analyses were limited to estimating the main component within a sample having the most distinct absorption feature. Next, we employed the MICA-PRISM algorithm developed by R. Kokaly and his colleagues from the USGS (Clark et al., 2003). This allowed determining whether the sample originated from a calcite-rich or poor environment, in addition to the dominant mineral within a mixture (Mulder et al., 2012).

Through the course of our research it was confirmed that the reflectance spectra of our mixtures was indeed typically a complex result from the combinations of the spectral characteristics of the constituents. Depending on the composition, the abundance and the spatial arrangement of the minerals, the total reflectance resulting from the scattering of the minerals within the intimate mixture produces positional shifts, changes in intensity, disappearance of absorption features or changes in their shape (Clark et al., 1990). This is illustrated in Figure 2. The figure shows very clear absorption features for the spectral signatures of each mineral. However, when you mix them together in one sample most of these signatures disappear. Therefore, it was deemed necessary to employ a method that is better capable to deal with the non-linear scattering behavior of the samples.

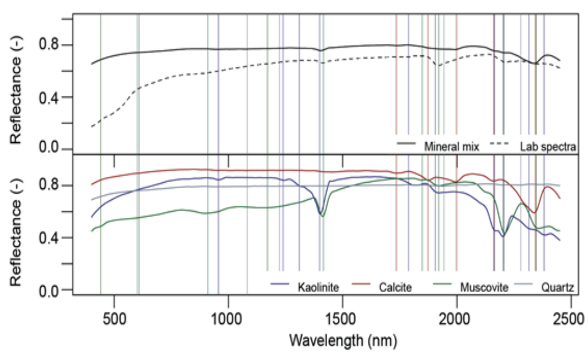


Figure 2: (top) Reflectance of the linear mixed and measured reflectance of sample 41, (bottom) Reflectance of the pure minerals (abundance > 5%) present in a sample. The vertical lines indicate absorption features of the various minerals (after Mulder et al. (2012))

Here, spectral deconvolution (SD) provided the answer to our research goal. SD involves modelling the total reflectance and the inference of absorption components within complex features by fitting (modified) Gaussian curves to the absorption features and absorption components (Pompilio et al., 2009). The concept is presented in Figure 3. The principle is fairly simple. One identifies the expected absorption wavelengths and at these positions Gaussians are fitted which together are an approximation of the original measured spectrum. Each Gaussian curve is described by a width, depth, asymmetry and saturation parameter. These parameters vary depending on the composition of a mixture but also on the quantity of each constituent. Previously, the parameters of the fitted Gaussians were being used in linear regression models to predict two minerals within a sample (Sunshine and Pieters, 1993). In our work we chose for regression tree analysis for prediction mineral abundances. Regression tree analysis allows dealing with nonlinearity and interactions between the EGO parameters which may improve the prediction accuracy.

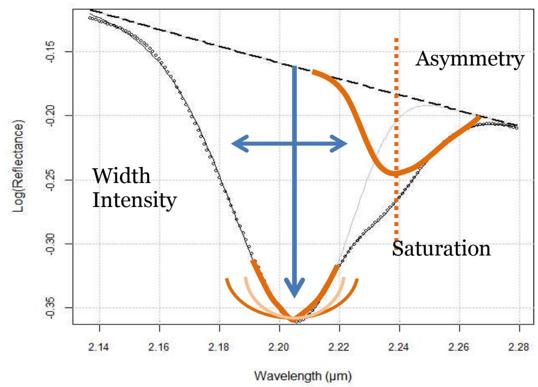


Figure 3: Illustration of spectral deconvolution.

The method was first tested on samples collected in the field. The first results were promising, however, we did not know if there was an influence of pollution coming from e.g. presence of small quantities of other minerals or traces of organic matter. Therefore, we prepared samples with known abundances of the dominant minerals found in the field samples, being kaolinite, dioctahedral mica, smectite, calcite and quartz. In addition, we also “polluted” some prepared samples with chlorite in order to test the robustness of the method.

Cross validation showed that the prepared samples of kaolinite, dioctahedral mica, smectite and calcite were predicted with a root mean square error (RMSE) less than 9 wt%. For the field samples, the RMSE was less than 8 wt% for calcite, dioctahedral mica and kaolinite abundances (Figure. 4). Smectite could not be well predicted, which was attributed to spectral variation of the cations within the dioctahedral layered smectites. Substitution of part of the quartz by chlorite at the prediction phase hardly affected the accuracy of the predicted mineral content; this suggests that the method is robust in handling the omission of minerals during the training phase. The degree of expression of absorption components was different between the field sample and the laboratory mixtures. This demonstrates that the method should be calibrated and trained on local samples. Concluding, our method allows the simultaneous quantification of more than two minerals within a complex mixture and thereby enhances the perspectives of spectral analysis for mineral abundances.

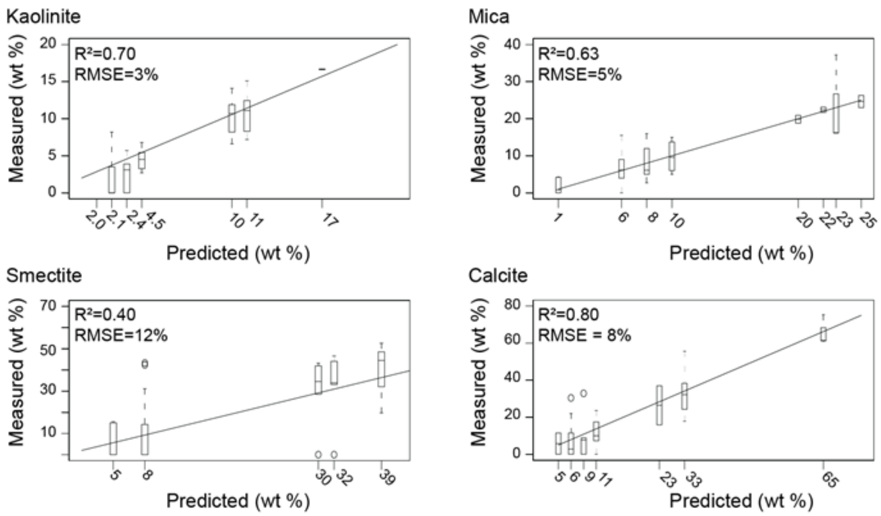


Figure 4: Predicted relative mineral content (wt%) from the field experiment compared to the values obtained from the XRD analysis. The predicted mineralogy is presented by boxplots of the samples which were assigned to the terminal nodes of the regression tree (after Mulder et al., 2013)

References

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High Spectral Resolution Reflectance Spectroscopy of Minerals. Journal of Geophysical Research 95(B8), 12653-12680.

Clark, R.N., Swayze, G.A., Livo, K.E., Kokaly, R.F., Sutley, S.J., Dalton, J.B., McDougal, R.R., Gent, C.A., 2003. Imaging spectroscopy: Earth and planetary remote sensing with the USGS

Mavris, C., Plötze, M., Mirabella, A., Giaccai, D., Valboa, G., Egli, M., 2011. Clay mineral evolution along a soil chronosequence in an Alpine proglacial area. Geoderma 165(1), 106-117.

Mulder, V.L., Plötze, M., de Bruin, S., Schaepman, M.E., Mavris, C., Kokaly, R., Egli, M., (2013). Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1-2.4 μm) and regression tree analysis. Geoderma, (207) 279-290. DOI: 10.1016/j.geoderma.2013.05.011

Mulder, V.L., de Bruin, S., Schaepman, M.E., 2012. Retrieval of composite mineralogy by VNIR spectroscopy. In: B. Minasny (Ed.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012. CRC Press, Sydney, Australia, pp. 488.

Pompilio, L., Pedrazzi, G., Sgavetti, M., Cloutis, E.A., Craig, M.A., Roush, T.L., 2009. Exponential Gaussian approach for spectral modeling: The EGO algorithm I. Band saturation. Icarus 201(2), 781-794.

Sunshine, J.M., & Pieters, C.M. (1993). Estimating Modal Abundances from the spectra of natural and laboratory pyroxene mixtures using the Modified Gaussian Model. Journal of Geophysical Research, 98, 9075-9087

Titia holds MSc degrees in Soil Science – Land Science and Geo-Information Science & Remote Sensing from Wageningen University (The Netherlands). In 2013, she obtained her PhD from Wageningen University, in close cooperation with Zürich University, with a thesis on spectroscopy-supported digital soil mapping. Titia is a post-doctoral researcher within the Infosol Unit at INRA Centre d'Orléans and received an AgreenSkills' fellowship for performing her research. Her research interests are in large-scale modelling and mapping of natural resources, thereby integrating data mining techniques, geostatistics and remote sensing.

The gate's locked! I can't get to the exact sampling spot … can I sample nearby?

Colby Brungard[1] & Jamin Johanson[2],

[1]*Utah State University*

[2]*Dover-Foxcroft MLRA Soil Survey Office in Dover-Foxcroft ME. USA*

Conditioned Latin hypercube sampling (cLHS, Minasny and McBratney, 2006) identifies physical sampling locations that are optimized to represent the multivariate distribution of input environmental covariates. As long as it is possible to visit the exact physical location of each sampling point, the sampling campaign will capture the variability of input covariates. However; In the event of restricted access (e.g., a locked gate) we need to know where to move the sampling location without jeopardizing the representativeness of the original cLHS point. This is a particular problem in areas where access is difficult and soil surveyors have traveled long distances. In this event, deciding not to sample at the location is not a viable option.

When sampling in a rugged and remote area of the southern Philippines, Thomas et al. (2012) noted that the "main frustration with cLHS sampling is that it's black-box nature" leaves the "surveyor unsure why the cLHS has chosen any given site". This "prevents the soil surveyor from choosing alternative sites in the field when cLHS sites prove impossible to reach for practical reasons". Thomas et al. (2012) suggest landscape stratification, fuzzy clustering of the covariates, and cLHS of the resulting fuzzy memberships.

Kidd et al. (2015) faced similar operational constraints in Tasmania. Pre-defined cLHS sample locations, even with contingency sites, proved difficult to implement in the field, with a variety of access issues making sampling slow and difficult. As an alternative they used a 'relaxed' sample design by sampling from fuzzy k-means covariate clusters. A map of clusters provided to soil sampling staff allowed difficult sites to be relocated within the same cluster, maintaining stratification. They found that this relaxed approach still adequately represented the covariate distribution while providing greater flexibility to field sampling staff. Clifford et al. (2014) modified this idea to flexible Latin hypercube sampling, which involves an optimization process for selecting accessible sites in a region while still maintaining the LHS criteria. The flexible sampling algorithm produces an ordered list of alternative sites close to the primary target site .

While the Flexible LHS is sound, it involves a complex computational procedure. An easier pragmatic option would be to calculate a similarity measure to the original cLHS within a given area around each cLHS point. Areas with high similarity to the original cLHS could then be identified and the sample location moved to these areas if needed .

To test this idea, we used cLHS to identify 100 sampling locations for a coupled soil-vegetation study over approximately 91,646 km^2 in the northeastern USA (Figure. 1). Input environmental covariates were elevation, slope, aspect, land cover type, and catena group. Catena groups were created from existing soil maps. Latitude and longitude were also included as covariates because this study covered a large area and exhibited strong climate and vegetation trends. All covariates had a 30 m resolution.

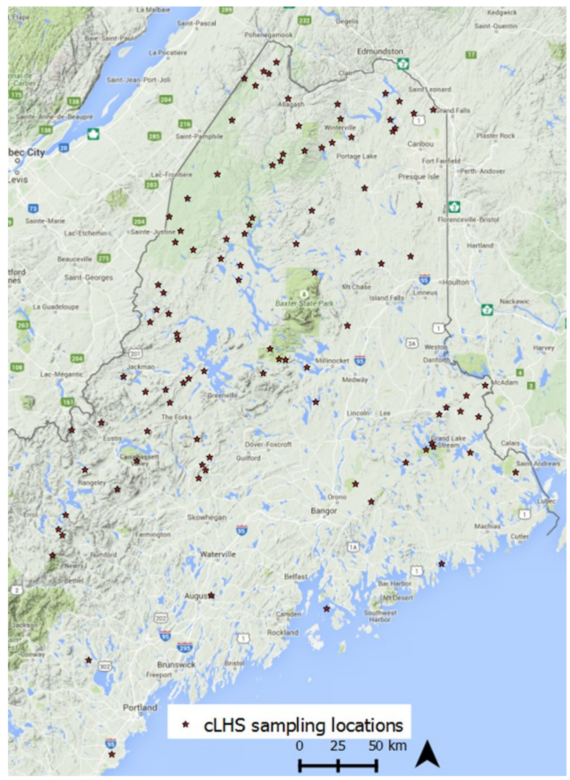


Figure 1. Study area in northeastern USA showing locations of 100 cLHS sampling locations.

We applied a 1 km buffer (2 km diameter) around each cLHS point, extracted the covariate values inside each buffer, calculated Gower's dissimilarity index (Gower, 1971, Mallavan et al., 2010) between the cLHS point and the covariate values, and converted dissimilarity to similarity by subtracting from 1. Gower's dissimilarity index was chosen to calculate similarity because it can handle both categorical and continuous values.

The resulting output was a raster where cell values inside each buffer area are a measure of similarity to the original cLHS point

We anticipate this output will be useful in identifying areas to

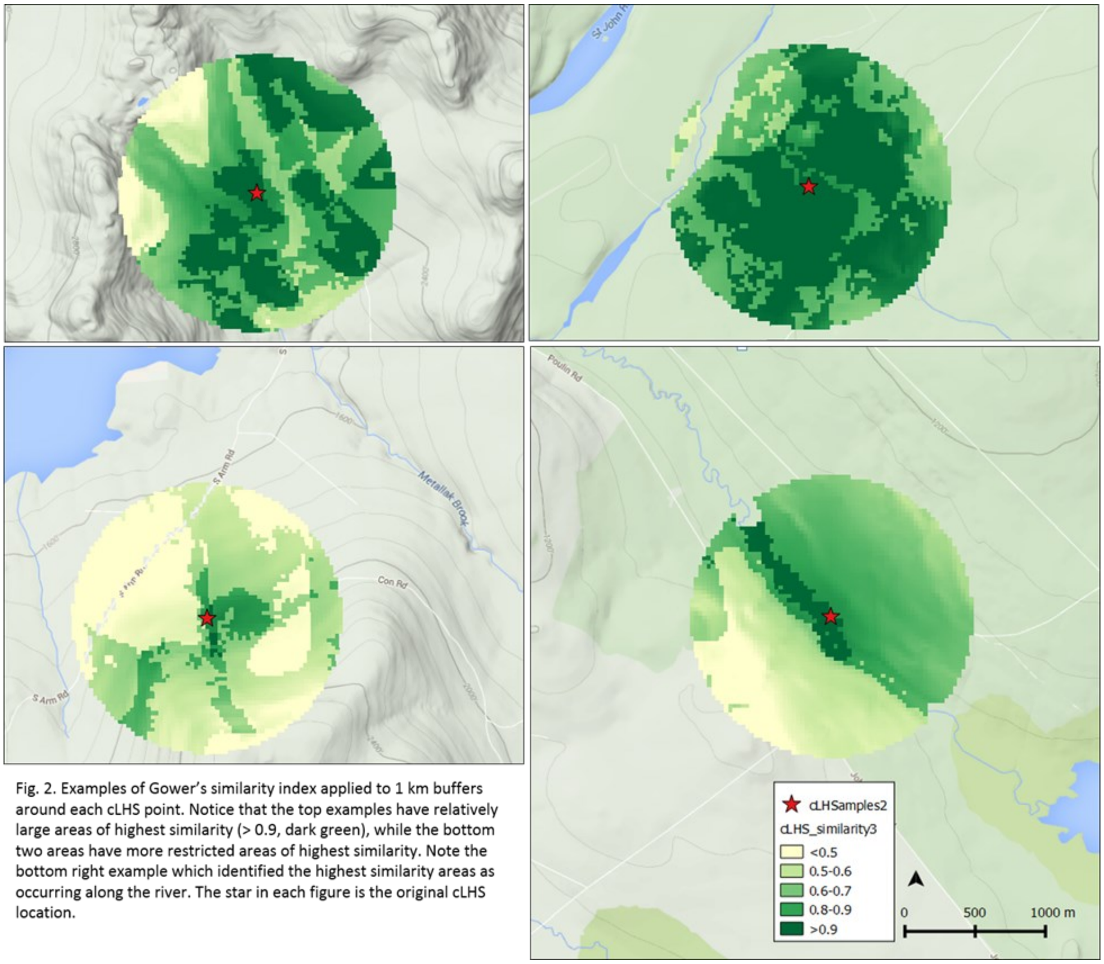


Fig. 2. Examples of Gower's similarity index applied to 1 km buffers around each cLHS point. Notice that the top examples have relatively large areas of highest similarity (> 0.9, dark green), while the bottom two areas have more restricted areas of highest similarity. Note the bottom right example which identified the highest similarity areas as occurring along the river. The star in each figure is the original cLHS location.

Figure 2. Cell values potentially range from 0 (not similar) to 1 (perfectly similar), but in practice the lower limit, at least for our covariates and buffer area size, was ~ 0.25.

which a soil surveyor could move a sampling location and still remain relatively similar to the original cLHS point. Such decisions would entail choosing an area (cell) that is accessible and that has a high enough similarity to the original cLHS point. We heuristically suggest a distance > 0.9 as a value for deciding if an area is similar enough, but this could be investigated further by randomly sampling within similarity deciles and plotting the resulting covariate distribution against the original cLHS covariate distribution.

As the decision to move a sampling location must often be made in the field, we have found the GarmingCustomMap QGIS addon (https://hub.qgis.org/projects/garmincustommap) useful for loading maps onto a GPS, and it may be possible to load the cLHS similarity index raster onto a Garmin GPS with custom map capabilities. This would then allow the soil surveyor to rapidly identify similar locations in the field in the event of restricted access.

Complete and commented, but by no means polished, R code to implement Gower's similarity index on cLHS points can be obtained by emailing Colby Brungard at envsoilco@gmail.com. Data necessary to run this code includes: the covariates used as input to cLHS, the actual cLHS points, and the desired buffer size.

References:

Clifford, D., Payne, J.E., Pringle, M.J., Searle, R., Butler, N., 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. Comput. Geosci. 67, 62–68. doi:10.1016/j.cageo.2014.03.005

Gower, J.C., 1971. A General Coefficient of Similarity and Some of Its Properties. Biometrics 27, 857–871. doi:10.2307/2528823

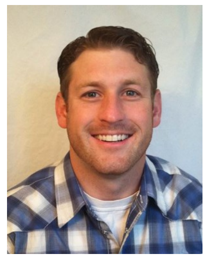
Kidd, D., Malone, B., McBratney, A., Minasny, B., Webb, M., 2015. Operational sampling challenges to digital soil mapping in Tasmania, Australia. Geoderma Reg. 4, 1–10. doi:10.1016/j.geodrs.2014.11.002

Mallavan, B.P., Minasny, B., McBratney, A.B., 2010. Homosoil, a Methodology for Quantitative Extrapolation of Soil Information Across the Globe, in: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Springer Netherlands, pp. 137–150.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

Thomas, M., Odgers, N.P., Ringrose-Voase, A., Grealish, G., Glover, M., Dowling, T., 2012. Soil survey design for management-scale digital soil mapping in a mountainous southern Philippine catchment, in: Minasny, B., Malone, B.P., McBratney, A. (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney Australia, pp. 233–238.

Colby Brungard is a researcher at Utah State University, in Logan, UT. USA



Jamin Johanson is an NRCS -Ecological Site Specialist at the Dover-Foxcroft MLRA Soil Survey Office in Dover-Foxcroft ME. USA.

GlobalSoilMap got an Honourable Mention

GlobalSoilMap project was recently awarded an **Honorable Mention** by the 2015 International Data Rescue Award in Geosciences. Dominique Arrouays and Johan Leenaars submitted the application with many inputs from researchers around the world. Dominique, the science coordinator of GlobaSoilMap, said that although we did not win he is pleased with a global recognition, and demonstrated that this project has gathered support from various institutions around the world and some good examples have been achieved.

Pedometricians’ Favourite Equations

17 Equations that Changed the World is a book written by Ian Stewart in 2013. The author listed 17 most important mathematical equations that he thought have been a driving force behind nearly every aspect of our lives. Pedometricians have frequently used at least 8 of the equations in their works (can you recognise them?). Starting this issue, we asked some pedometricians on their favourite equations and why they love them.

17 Equations That Changed the World

by Ian Stewart

1.	**Pythagoras’s Theorem**	$a^2 + b^2 = c^2$	Pythagoras,530 BC
2.	**Logarithms**	$\log xy = \log x + \log y$	John Napier, 1610
3.	**Calculus**	$\frac{\mathrm{d}f}{\mathrm{d}t} = \lim_{h \to 0} = \frac{f(t+h) - f(t)}{h}$	Newton, 1668
4.	**Law of Gravity**	$F = G\frac{m_1 m_2}{r^2}$	Newton, 1687
5.	**The Square Root of Minus One**	$i^2 = -1$	Euler, 1750
6.	**Euler’s Formula for Polyhedra**	$V - E + F = 2$	Euler, 1751
7.	**Normal Distribution**	$\Phi(x) = \frac{1}{\sqrt{2\pi}\rho} e^{\frac{(x-\mu)^2}{2\rho^2}}$	C.F. Gauss, 1810
8.	**Wave Equation**	$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$	J. d’Almbert, 1746
9.	**Fourier Transform**	$f(\omega) = \int_{\infty}^{\infty} f(x) e^{-2\pi i x \omega} \mathrm{d}x$	J. Fourier, 1822
10.	**Navier-Stokes Equation**	$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v}\right) = -\nabla p + \nabla \cdot \mathbf{T} + \mathbf{f}$	C. Navier, G. Stokes, 1845
11.	**Maxwell’s Equations**	$\nabla \cdot \mathbf{E} = 0$ $\quad \nabla \cdot \mathbf{H} = 0$ $\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}$ $\quad \nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial E}{\partial t}$	J.C. Maxwell, 1865
12.	**Second Law of Thermodynamics**	$\mathrm{d}S \geq 0$	L. Boltzmann, 1874
13.	**Relativity**	$E = mc^2$	Einstein, 1905
14.	**Schrodinger’s Equation**	$i\hbar \frac{\partial}{\partial t}\Psi = H\Psi$	E. Schrodinger, 1927
15.	**Information Theory**	$H = -\sum p(x) \log p(x)$	C. Shannon, 1949
16.	**Chaos Theory**	$x_{t+1} = kx_t(1 - x_t)$	Robert May, 1975
17.	**Black-Scholes Equation**	$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} - rV = 0$	F. Black, M. Scholes, 1990

Gerard Heuvelink

When Budi asked what is my favourite equation, I realised that I had never really thought about it. I did recall that the favourite equation of many mathematicians is said to be Euler's identity:

$$e^{i\pi} + 1 = 0$$

It is their favourite because it unites the five most important mathematical symbols. Euler's identity certainly is elegant and intriguing (although much of the intrigue is lost when realising that by definition $e^{ix} = \cos(x)+i\sin(x)$) but it has no practical use, and therefore I decided to look for another equation that could be my favourite. I ended up with the (simple) kriging variance equation:

$$\sigma_{SK}^2 = C(0) - \sum_{i=1}^{n} \lambda_i \cdot C(h_i)$$

I chose it because it is much criticised by geostatisticians. It needs someone to stand up for it, and I am happy to do so, because it is criticised for the very same reason that I like it so much.

The criticism is that the kriging variance does not depend on the data values. This, of course, is only partly true because the data values have been used to derive the covariance function (i.e. variogram), but the variogram and the spatial data configuration are indeed all that is needed to calculate it. Is that very wrong? No it is not, because it is a direct consequence of the geostatistical model that is assumed. Edzer Pebesma and I did numerical experiments that confirmed that under common assumptions the interpolation error does not depend on local spatial variation, in spite of what one would expect intuitively. If you do not like it, change your model, but do not throw away the kriging variance while keeping the kriging prediction. Sometimes mathematics gives us a present, and this is one of such cases: because the kriging variance does not depend on the data values, we can optimise spatial sampling designs and evaluate the propagation of interpolation errors through environmental models prior to collecting the data. It is a godsend that should be relished, not condemned!

Alex McBratney

I guess most people will say it is the variogram equation, but my favourite equation is the fuzzy k means (with extragrades) allocation equations.

$$m_c = \frac{d_c^{-2/(\Phi-1)}}{\left\{\sum_{j=1}^{k} d_j^{-2/(\Phi-1)} + (((1-\alpha)/\alpha)\sum_{j=1}^{k} d_j^{-2})^{-1/(\Phi-1)}\right\}} \qquad c = 1,\ldots,k,$$

$$m_* = \frac{(((1-\alpha)/\alpha)\sum_{j=1}^{k} d_j^{-2})^{-1/(\Phi-1)}}{\left\{\sum_{j=1}^{k} d_j^{-2/(\Phi-1)} + (((1-\alpha)/\alpha)\sum_{j=1}^{k} d_j^{-2})^{-1/(\Phi-1)}\right\}}$$

The above equation allows the polythetic allocation of new soil data into any existing soil classes.

The first part of this equation calculates the memberships, m, in each of the k "regular" soil classes and the second part calculates the membership, m_*, in the extragrade class. It has 2 parameters Φ, the degree of fuzziness and α the degree of outliers (extragrades) which can be estimated from the data. Evaluation of these equations requires the calculation of the Mahalanobis distance, d, from each centroid to the new individual.

We are currently using this formulation to set up a Universal Soil Classification System with (some k=300) centroids derived from the 'great group' taxa of Soil Taxonomy, WRB and other national systems. The centroids in the new system are described by the depth functions (as fixed depth slices) of 22 soil properties, so some 400 properties in all – the Mahalanobis' distance recognises the correlations between the properties.

Hopefully field allocation of soil profiles using this equation will be a fast calculation rather than the traditional thumbing through of pages in an elaborate key.

Reference:

McBratney, A. B. (1994). Allocation of new individuals to continuous soil classes. Australian Journal of Soil Research, 32(4), 623-633.

Tom Bishop

Standard error of the mean = σ/ sqrt(n)

I love this equation because it is so simple and illustrates much about the issues we face in estimating a parameter about a population.

We have the numerator with a measure of variation so with more variation we have a larger std. error but to counter this is the denominator we have the number of observations so with more observations we can get smaller std. errors. It exemplifies the tension or balance we as Pedometricians have to deal with – working in a world of variation which we can better explain or predict when we have more observations.

What soil is it?

In June, we asked what soil is behind this famous scene in North by Northwest. Here are a couple of video clips that may help you identify the soil. Video 1 and Video 2.

So here is the answer and winner(s).

The scene from North by Northwest, where Thorhnhill getting off at the ‘Prairie Stop on Route 41’ and chased by a cropduster, was meant to be in Northern Indiana. However it was shot in California at Wasco, near Bakersfield along Garces Highway in Kern County, California. There is an article about it.

The first guess by Aitor García Tomillo from UDC Spain as an alluvial (Entisol) soil from the Wasco Series. Rachel Downward from Lincoln, New Zealand also guessed it as Wasco, Coarse-Loamy, Mixed, Superactive, Nonacid, Thermic Typic Torriorthents.

But the most likely soil series is Nahrub rightly identified by David Rossiter, who said “If I had only looked at the photos I would still have guessed a semi-aridic moisture regimes, clayey but no obvious gilgai, dark epipedon but probably not thick enough to be mollic, maybe some salt content.” Minerva Dorantes from Purdue University also guessed it correctly. Fine, smectitic, calcareous, thermic Vertic Torriorthents.

Well how did they know it? The key is Wikipedia and IMDB actually has the geographical coordinates (35°45′39″N 119°33′41″W), so that is a giveaway. The map unit identified it as 70% Nahrub and 20% Lethent (Typic Natrargids). It is also close to Garces series (Typic Natrargids).

Honourable mention to Marco Angelini who used his pedological observation “I expected to find calcic feature in that soil. Looking at the second video, 40 second, you can see a blocky structure with carbonates as well. “Homayoun Fathollahzadeh recognised the look as Inceptisol, probably with calcic horizon and more dry seasons.

But no one commented that the corn looks fake. As the article said: Wasco High students "planted" cornstalks to simulate a cornfield.

Pedometrics 2017

Pedometrics will be celebrating its Silver Jubilee in 2017. The first pedometrics conference was held in Wageningen in September 1992. Planning for Pedometrics 2017 is underway to celebrate 25th anniversary at its birthplace. The Organising Committee (Gerard Heuvelink, Dick Brus, Bas Kempen, Ichsani Wheeler, Harm Bartholomeus, David Marcellis and Jetse Stoorvogel) will be organising the conference from 26 June to 2 July 2017 at Hof van Wageningen which is the same location used in 1992! This will also be a joint conference with four of its Working Groups:

1. Digital Soil Mapping
2. Proximal Soil Sensing
3. Soil Monitoring
4. Modelling of Soil and Landscape Evolution

Mark on your calendar now for this once in a lifetime important event!

It's the accuracy, stupid

Gerard Heuvelink,

Wageningen University and ISRIC World Soil Information

I wrote about the confusion between spatial resolution and accuracy in Pedometron 36, but I did not get to discuss another accuracy-related issue that I also want to get off my chest. This relates to so-called "Frankenstein" maps. The term "Frankenstein" map, coined by my colleague and friend Tom Hengl, refers to a soil map in which administrative or otherwise artificial boundaries are clearly visible. Even though soil science tells us that such boundaries are not real, I will argue in this article that in many cases it is perfectly all right to produce or use a Frankenstein map. In fact, we should be glad that our maps have Frankenstein features.

Two nice examples of Frankenstein maps are given in Figure 1. I came to these examples with the help of Marc van Meirvenne (Ghent University) and Folkert de Vries (Alterra), but there are many more examples and I guess that most of you will have come across many other examples during your pedometrics career. Perhaps you even created some!

Frankenstein maps are not aesthetically appealing, but that does not mean that they are useless and should be avoided at all costs. The real issue is, as always, the ACCURACY of the map. If I were to choose between two maps, one of which a Frankenstein map and another free from artefacts, while knowing that the Frankenstein map is more accurate than the other, I would use the Frankenstein map.

Consider Figure 2 on the next page, which addresses a synthetic case of the spatial distribution of a soil property for a square study area. Assume that the study area is cut in two halves by a straight-line country border. The soil does not care about country borders, so no border effect can be detected in the map shown in the top left panel. Next imagine that the soil property was mapped twice: first in a global ("continental") approach, second in a local ("country") approach. The global approach (top centre panel) also ignores the country border, so it has no border artefacts. The local map (top right panel) was obtained by gluing two country maps, and the Frankenstein effect is clear. However, similar to what would occur in real-world practice, the mapping investment per unit area for the global map was lower than that for the local maps. Hence the global map is less accurate, as can be seen in the bottom panel of Figure 2. On average, the local map (red line) is closer to the true value (black line) than the global map (blue line). The visual impression is confirmed by calculating the root mean squared error, which is 0.70 for the global map, while it is 0.51 for the local map. Which of the two

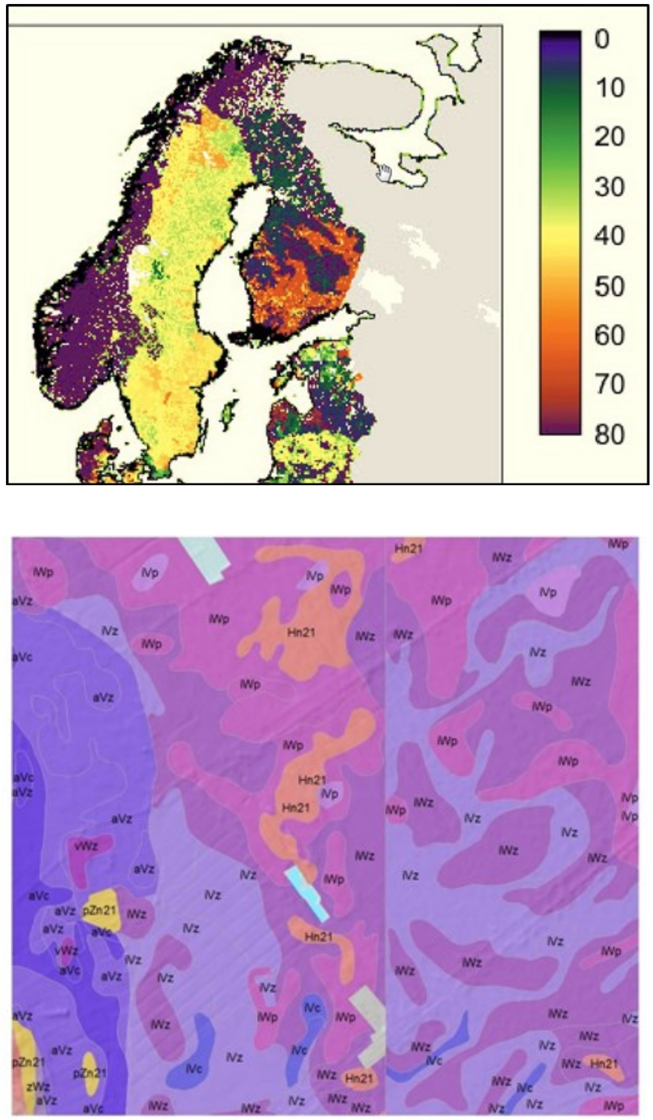


Figure 1. Examples of Frankenstein maps. Left: zoom-in on Scandinavia of the European sand content map as published in the European Soil Atlas (http://eusoils.jrc.ec.europa.eu/projects/Soil_Atlas/Index.html). Sweden sticks out, presumably because the Swedish government has a large-scale programme in place to increase the sand content of its soils. Right: zoom-in on the Dutch soil map, where a vertical straight line marks the border between two glued map sheets.

maps do you prefer?

For those interested, let me detail the technicalities of the synthetic example (I am also happy to share the R code, just send me an email): I simulated a reality on a 400 by 400 grid using unconditional sequential Gaussian simulation, with constant mean equal to 5 and a spherical variogram with a nugget of 0.1, a sill of 1.1, and a range of 100. Next I took a simple random sample of 100 observations and created the global prediction map using ordinary kriging (using the same variogram). For the local prediction maps, I first took a simple random sample of 1000 observations, split it into two subsets depending on whether the observation locations were to the left or right of the country border, and used ordinary kriging again to interpolate for each country separately[1].

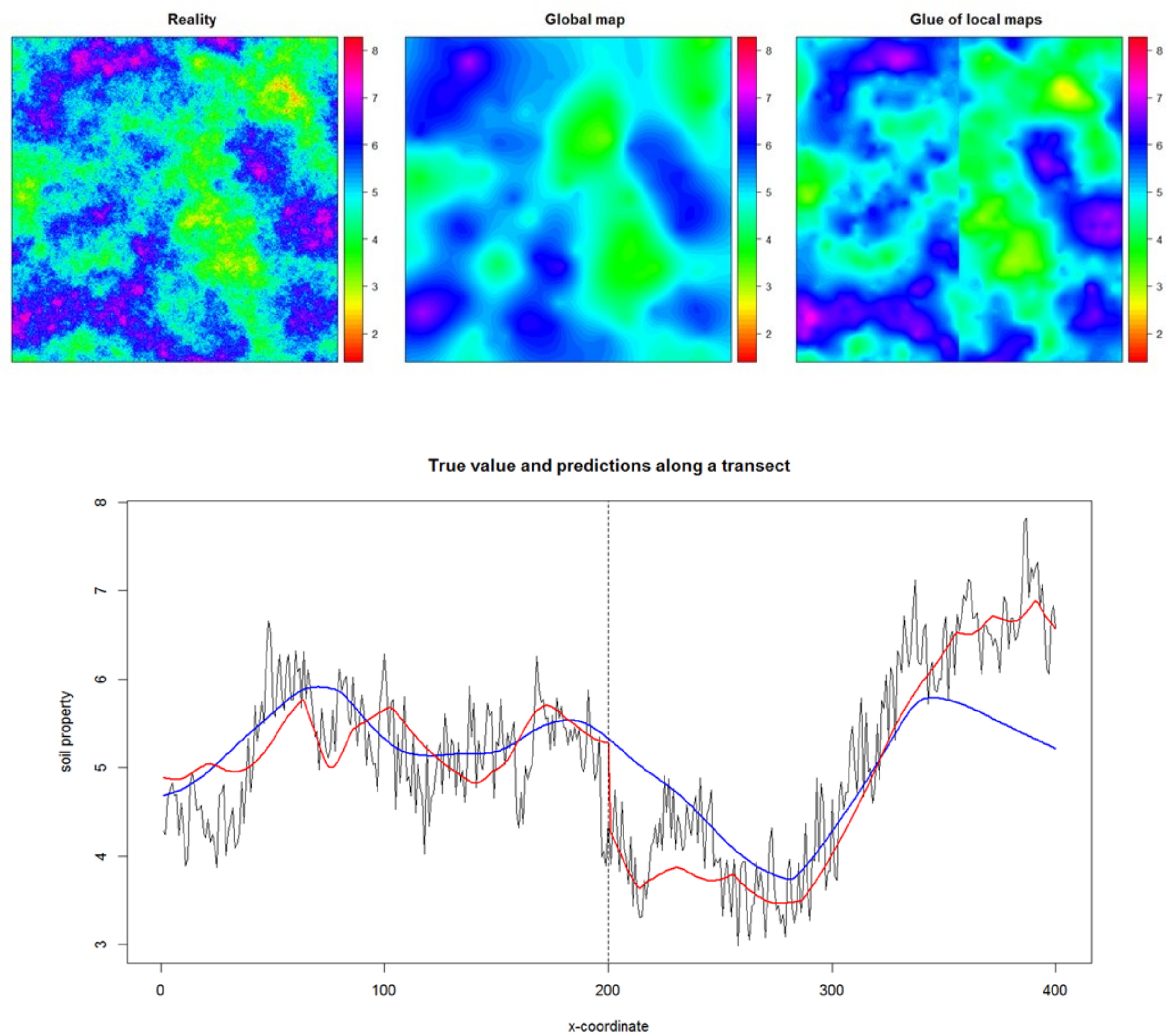


Figure 2. Synthetic example evaluating the performance of a global and local map of a soil property for an area divided in two by a “country” border. Top left: true (unknown) soil property. Top centre: “global” map, obtained by ignoring the country border and low investment (i.e. low sampling density) per unit area. Top right: glue of two local maps, one for each country, with higher investment per unit area. Bottom: true value (black line) and predictions across a horizontal transect across the area. Blue line shows the global prediction, red line the local prediction. Note the Frankenstein effect in the local map when crossing the border (discrete jump in red line at vertical dashed line)

[1] The observant reader will notice that in an ideal world the most accurate (and Frankenstein–free!) map would be obtained if the two countries worked together, merged their datasets and jointly made a map for both countries, but alas such solutions are not easily achieved in the real world.

The synthetic example shown in Figure 2 is not a far-fetched, exotic case that would never occur in real life. Indeed digital soil mappers are frequently confronted with this problem, when they need to decide on which covariates to include and which not. Some covariates may have significant predictive power but have artefacts as well. If these are included, Frankenstein effects will occur. A nice example are the SoilGrids1km maps (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105992), which provide global predictions of soil properties at standard depths. These maps were made with regression kriging, with lots of covariates taken into account, some of which are categorical and have discrete spatial boundaries that not always occur at 'natural' places. Figure 3 shows a zoom-in on parts of Israel and neighbouring countries, where mapped topsoil bulk density and pH clearly show Frankenstein features. In this case, these are likely caused by the Harmonised World Soil Database map and/or the Global Lithological Map, that were both used as categorical covariates. Still, it makes perfect sense to include these maps as covariates, because the resulting soil property maps are likely more accurate than maps that would not have taken these covariates into account.

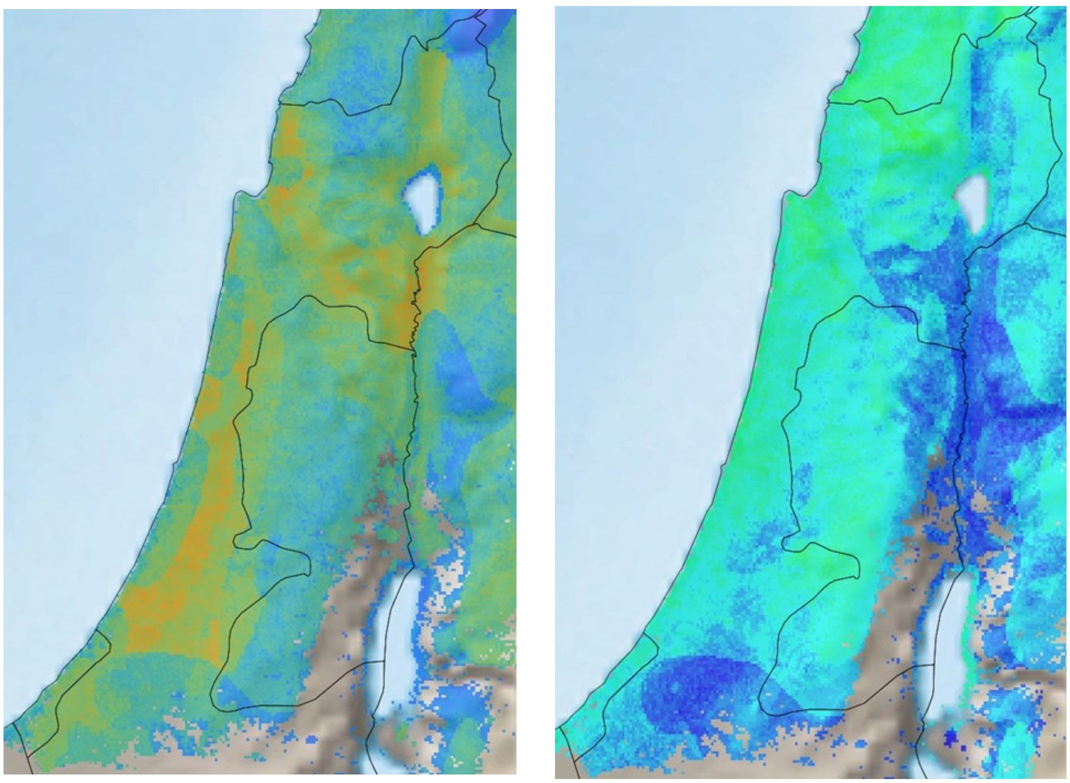
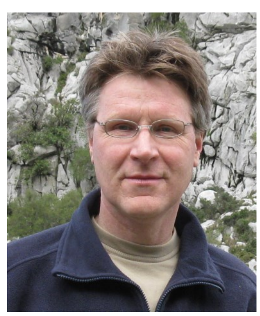


Figure 3. Maps of the topsoil bulk density (left) and pH (right) of parts of Israel and neighbouring countries as predicted by SoilGrids (www.soilgrids.org, accessed 2 July 2015).

So, Frankenstein maps are perfectly acceptable, because maps should be judged on the accuracy with which they represent the real world, not on how pretty they look. In fact, I would argue that it is not a bad thing that a map has artefacts, because these constantly remind us of the fact that the map is only an approximation of the real world. Many users tend to think that maps are perfect, and it is our task to explain them that they are not. Frankenstein features provide a very powerful and convincing method to do just that.

Let me end with the wise words of Peter Burrough, from his 1998 GIS book with Rachael McDonnell (page 220), that are still very valid: "The quality of GIS products is often judged by the visual appearance of the end-product ..... uncertainties and errors are intrinsic to spatial data and need to be addressed properly, not swept away under the carpet of fancy graphics displays".



Gerard Heuvelink holds a position as associate professor in Geostatistics with the Soil Geography and Landscape group of Wageningen University (since 2003) and senior researcher Pedometrics and Digital Soil Mapping with ISRIC World Soil Information (since 2011). His research interests include geostatistics, spatial uncertainty aAnalysis, statistics, pedometrics, and sampling design optimization.

A Hydrologic and Land Surface Modeler's Plea

Nathaniel W. Chaney,
Princeton University

We need higher resolution soil data in our global environmental models, and we need it today. Let me explain: since the inception of numerical weather forecasting and climate modeling in the 1950's, it has been a persistent challenge to accurately represent the spatial characteristics of the land surface water, energy, and carbon cycles. In the early 1990's, the land schemes of climate models were commonly run over the globe at spatial resolutions ranging between 100 and 500 km. Over time, the continual increase in high performance computing (HPC) resources has allowed the community to increase the spatial resolution of these models, which now range globally from 5 to 25 kilometers. Although an improvement, the land and hydrologic modeling communities argue that this is still insufficient, as many important land surface processes need to be represented at the field and hillslope scales (~100 meters) [Wood et al., 2011; Bierkens et al., 2014]. Thus, the modeling community is currently developing "hyperresolution" environmental models. But as we attempt to solve the puzzle of increasing the resolution of our models, we are realizing that there are missing pieces—one of them being the availability of global high-resolution and high-quality soil data. To put things into perspective, current global models continue to rely on outdated continental soil datasets (e.g., Harmonized World Soil Database) that provide information at spatial scales ranging from 25 to 500 km – very far from the sub-100 meter goal.

In 2013, as I was working on the development of a field-scale hydrologic model over the contiguous United States (CONUS), I began to address this problem by exploring methods to spatially downscale existing legacy soil data. Thankfully, the extensive literature on digital soil mapping saved me a lot of time. Having worked with the Soil Survey Geographic (SSURGO) database over the United States, I was familiar—or at least I thought I

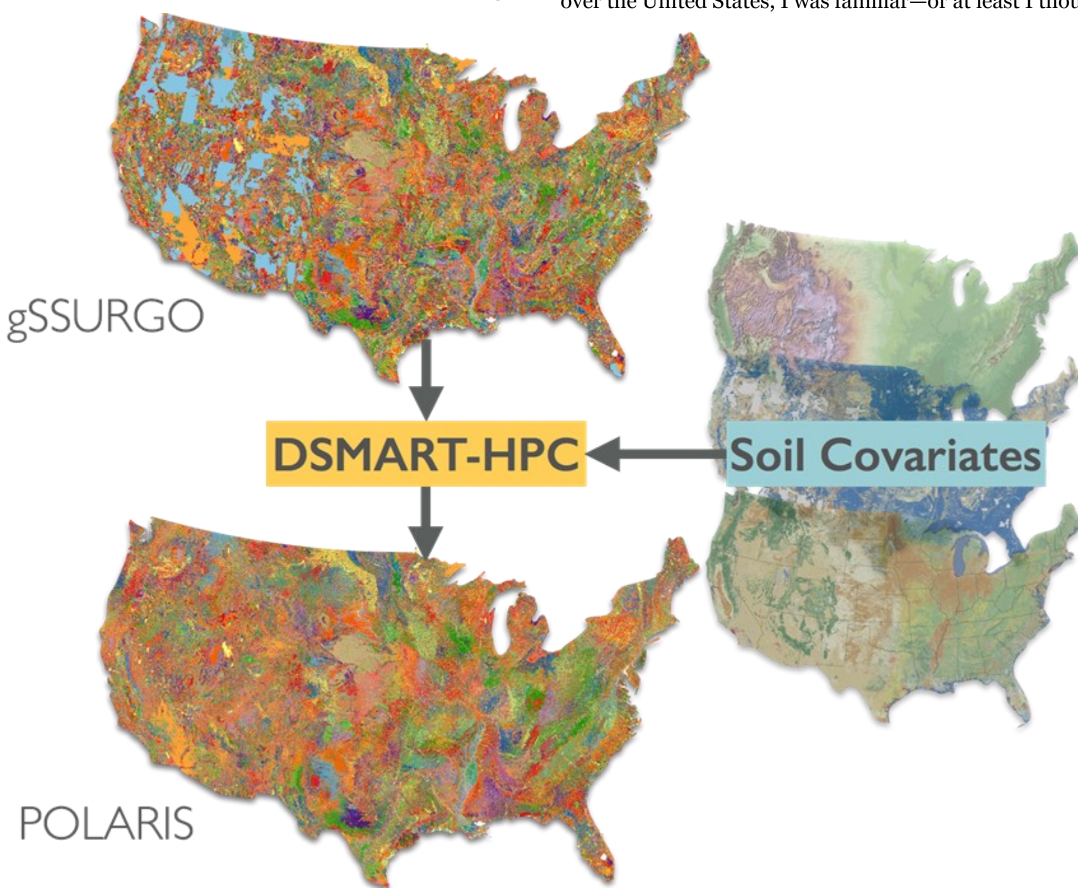


Figure 1. Spatial disaggregation, harmonization, and gap filling of SSURGO over CONUS using DSMART-HPC and high-resolution environmental soil covariates to produce the POLARIS dataset.

was—with the strengths and weaknesses of legacy soil data (e.g., political boundary discontinuities, unsurveyed areas, and varying effective spatial resolutions). I was immediately drawn towards the emerging spatial disaggregation techniques. As I visited Alex McBratney's group in Sydney in 2014, I quickly resolved to implement DSMART—a spatial disaggregation and harmonization algorithm [Odgers et al., 2014] over the state of Illinois in the United States. Unfortunately, what I learned was that these types of algorithms do not scale well beyond relatively small legacy soil datasets and medium size domains. Needless to say, these algorithms are computationally prohibitive when trying to disaggregate extensive legacy soil databases over continental extents. The computational power and storage required for these tasks are far beyond what is possible on a state-of-the-art personal desktop, and it will remain unachievable for decades.

However, if we are able to combine the power of tens of thousands of desktops to work together, we can be creative and apply these sophisticated spatial disaggregation algorithms over continental extents at very high spatial resolutions. Instead of waiting 20 years, this can be done today. This is the basic idea behind high performance computing (HPC). Dedicated HPC centers consist of tens to hundreds of thousands of computing cores that interact and work in parallel to complete a user-defined task. These resources keep becoming more powerful and more accessible to the scientific community (e.g., Google Earth Engine). For example, the Blue Waters supercomputer in Illinois has roughly 600,000 computing cores and 500 petabytes of storage. In simple terms, this machine can perform in one hour what it would take a single-core personal desktop to run in 68 years. Furthermore, its storage capacity is massive—it can store roughly 10% of all words ever spoken by humankind.

Given my background in HPC and access to the Blue Waters supercomputer, Alex McBratney and I saw the opportunity to extend DSMART to a parallelized framework. In collaboration with Alex McBratney's group and Jon Hempel at the National Cooperative Soil Survey (NCSS) in the United States, we devised DSMART-HPC, an extension of the DSMART algorithm that can be run on a supercomputer to spatially disaggregate large legacy soil databases over entire continents. As a proof of concept, we applied it over CONUS to spatially disaggregate and harmonize the SSURGO database. To do this, DSMART-HPC splits CONUS into 12,500 overlapping blocks and then the DSMART algorithm is applied independently on each block. Using Blue Waters, this allowed for a reduction in the time to complete the task from 500,000 hours (~57 years) to under 5 hours. The resulting dataset is POLARIS—Probabilistic Remapping of SSURGO—a spatially complete soil class database (with uncertainties) over the entire CONUS at a 30-meter spatial resolution. Figure 1 shows the difference between the maps of the most probable soil class of the original SSURGO product and the corresponding most probable soil class in the POLARIS dataset. Notice the ability of the algorithm to infill the missing regions in the Mountain West while maintaining the general spatial structure of the original SSURGO database.

All this being said, I am the first to acknowledge that the POLARIS dataset has multiple weaknesses (e.g., at times placing valley -dominant soils on ridges) that will need to be addressed as we move forward. However, that is not the point. The main purpose behind this work is to provide a proof of concept of what is possible in digital soil mapping with high performance computing resources that are available to the scientific community. If this is not motivation enough, let me whet your appetite: creating POLARIS only required 500,000 core hours, but allocations on existing supercomputers can range from millions to hundreds of millions of cores hours. In other words, an algorithm such as DSMART-HPC could be applied today over the globe at a 30-meter spatial resolution.

To conclude, let me return to my original point. The land surface and hydrologic modeling communities are already taking advantage of Big Data and HPC resources to increase the spatial resolution of environmental models over the globe. However, these models are only as good as the data that is fed into them; soil information has long been ignored and the community is quickly realizing the importance of accurately representing the vertical structure and horizontal spatial distribution of soils in these models, especially as we continue to move towards finer spatial resolutions. To make progress, we need the data today and not tomorrow. To make this possible, I urge the DSM community to embrace high performance computing. From my (limited) experience, many of the existing algorithms can be easily adapted to run on existing supercomputers to make this goal a reality. I, for one, will use the data from day one.

Nathaniel Chaney is a postdoctoral research associate in the Program of Atmospheric and Oceanic Sciences at Princeton University and at the Geophysical Fluid Dynamics Laboratory. His research interests revolve around high-resolution land surface and hydrologic modeling. He received a B.A. in Applied Mathematics and Earth and Planetary Science from U.C. Berkeley in 2010, and a Ph.D. in Civil and Environmental Engineering from Princeton University in 2015.

Noteworthy Articles

The Mathematical Shape of Things to Come

An article in Quanta Magazine highlights the new Data Driven research. In pedometrics and digital soil mapping we dealt a lot with big data, and in many ways we draw conclusions and gain understanding on soil based on empirical relationships. Science writer Jennifer Oullette wrote: Today's big data is noisy, unstructured, and dynamic rather than static. It may also be corrupted or incomplete. Researchers need new mathematical tools in order to glean useful information from the data sets. The article presented an example on topological data analysis (TDA) that reduce large, raw data sets of many dimensions to a compressed representation in lower dimensions without sacrificing the most relevant topological properties. Are we ready to use this kind of approach for large soil data such as hyperspectral imaging?

Sleeping Beauties

Sleeping beauty in a scientific publication is a paper that goes unnoticed for a long time, and suddenly attracts many citations. The term was proposed by Anthony van Raan in 2003. And last month the concept was rediscovered by a group in Indiana University and published in PNAS and made a lot of fuss in the media. The authors listed the top 15 Sleeping Beauties which is a bit unexpected, considering some of the papers are already well known. Soil scientists will quickly recognise the top sleeping beauty paper authored by Freundlich which was apparently awakened in 2002. There is also Langmuir, and Washburn (or the capillary rise equation). Pedometricians will recognise the most commonly used measure of linear correlation by Pearson (1901), which was apparently awakened in 2002 as well.

Do you know of any Sleeping beauty papers in Soil Science?

Balanced Sampling for Soil Survey

Dick Brus, our soil sampling expert, recently proposed the use of Balanced Sampling for Soil Survey. Pedometricians usually optimized their sampling strategy to cover the geographical space, feature (covariate) space or both. Now Dick found that there is Balanced Smapling which is commonly used in socioeconomic study has not been noticed in soil science. It has several advantages, one of them is that Latin hypercube sampling is a special case of a balanced sampling. In addition, if balanced sampling is applied, the inclusion probabilities of the samples can be calculated and optimised. Thus samples obtained by this design can be used both in design-based or model-based statistical inference. Dick illustrated its application with examples in soil surveys. The paper is published online in the upcoming September issue of Geoderma.

Optimal spatial stratification using digital soil maps

Jaap de Gruijter and colleagues rcently developed a new method to optimize spatial stratification and allocation for stratified random sampling of points. The method uses a grid of points with uncertain predictions of the target variable (for example a digital soil map). The objective function of the sampling stratification is defined by generalized distances between pairs of grid points, determined by the difference between the predictions, the variances of the prediction errors, and their covariance as a function of the geographical distance. The authors used an iterative reallocation algorithm to minimise this objective function. The resulting stratifications represent solutions on a continuous scale between two extremes: for errorless predictions, a stratification close to those by the cum-root-f method, and for entirely uninformative, a compact geographical stratification based only on the locations of the grid points. The authors illustrated it with a soil survey example. The study was published in the Journal of Survey Statistics and Methodology.

Block correlation and the spatial resolution of soil property maps made by kriging

Murray Lark recently proposed a new measure of the uncertainty of block kriging predictions. This is in response to the different resolution issues in digital soil maps. Block correlation is the expected correlation of the block prediction with the value that it estimates: the spatial mean of the target variable across the block. This correlation can be computed if the variogram and disposition of sample points are known. Using a hypothetical and real examples, Murray showed that he can calculate the block correlation as a function of block length and grid spacing. This paper is published in Geoderma.

Can citizen science assist digital soil mapping?

David Rossiter and colleagues pose this question in a recent article published in Geoderma.

While there are few citizen soil science projects, include the OPAL Soil and Earthworm Survey, GLOBE, and mySoil, but these are not aimed at soil mapping. The authors proposed digital soil mapping citizen science initiatives for countries with and without well-organized extension and advisory services and existing soil surveys, and identify types of citizens who might be motivated to contribute to such initiatives. Contributions could be in the form of tacit knowledge, opportunistic or protocol-guided new information, information from precision agriculture, and physical samples submitted for analysis.

The authors concluded that “the potential of citizen-provided information to accelerate and improve DSM projects seems too great to ignore. The benefits of citizen science DSM-related projects may go beyond the immediate aim of improving soil maps. In the longer term such projects would likely enhance the “connectivity” between soil and citizen.”

The 4th global workshop on Proximal Soil Sensing (Hangzhou, China, 12-15 May 2015)

Marc Van Meirvenne, Chairman of the WG-PSS

Since its establishment in June 2008, the Working Group on Proximal Soil Sensing organises biannually a global workshop, and this year the event took place at the Zhejiang University in the city of Hangzhou, relatively close to Shanghai, China. The overall theme was “Sensing soil conditions and functions”. The organisation was in the hands of Professor Zhou Shi of the Institute of Agricultural Remote Sensing and Information System of the College of Environmental and Resource Sciences of the Zhejiang University.

The workshop was a success. There were 112 registered participants from 16 countries and all inhabitable continents, and as usually the majority of participants (87) came from the organising country. The workshop covered three days, two days of plenary and poster sessions and one day with a field trip to experimental fields and the surrounding areas of Hangzhou. The two days of plenary sessions were subdivided according to different themes, each with a keynote speaker. These were (in chronological order):

- Raphael Viscarra Rossel “Baseline estimates of organic carbon and uncertainty by proximal soil sensing and soil spectroscopy”(CSIRO, Australia),
- Minzan Li “Development of soil nutrient sensors with spectroscopy” (China Agricultural University, China),
- Immo Trinks “State-of-the-art geophysical archaeological prospection and virtual archaeology” (Ludwig Boltzmann Institute for Archaeological Prospection, Austria),
- Richard Webster “Field sampling for proximal soil sensing), Advances in Field Spectroscopy for Soil Analyses” (Rothamsted Research, UK),
- Abdul Mouazen “Advances in Field Spectroscopy for Soil Analyses” (Cranfield University, UK).

The major attention of the workshop went to soil spectroscopy, reflecting the intensively conducted research on this promising technology. The major focus of application was agricultural, but other functions of soil, such as environmental health and protection of the buried cultural heritage, were also covered. There is a clear trend towards multi-signal and multi-sensor systems and a widening of soil sensor applications.

One session was devoted to a plenary discussion on the future planning and strategy of our working group. During this session it was agreed to join the Pedometrics-2017 initiative for a joint conference of the commission and all its working groups in Wageningen in The Netherlands, between 26 June and 2 July 2017. Given the frequently organised conferences and workshop by the commission on Pedometrics and its four working group, often with overlapping themes, it might be good to explore further such initiatives for jointly organised activities. It was also decided to expand the focus of our working group by seeking joint activities with other organisations dealing with soil sensing applications such as archaeology, civil engineering, environmental sanitation and protection, natural capital assessment, image processing… Finally, it was discussed which outcome of the workshop was to be preferred. The choice fell on a special issue of a scientific journal and it was decided to contact “Biosystems Engineering”. Dr. A. Mouazen agreed to act as the main supervisor of this special issue.

The workshop was followed by an extra day with two hands-on courses: “Sensors as data source and data acquisition methods” and “Data processing using R software”.

This 4th workshop of our working group was excellently and efficiently organised by the team of Prof. Shi. We are very grateful to Prof Shi and his collaborators!

Inaugural Workshop on Digital Soil Morphometrics

Pierre Roudier,
Research Division of Soils and Landscapes, New Zealand

The inaugural Global Workshop on Digital Soil Morphometrics was held in Madison (Wisconsin, USA), from the 1st to the 4th June 2015. The event, organised locally by Alfred Hartemink and Budiman Minasny at the wonderful University of Wisconsin campus. The workshop was held in the frame of the International Year of Soil, and run under the auspices of the International Union of Soil Science (IUSS) --- in particular of the Working Group on Digital Soil Morphometrics which is led by Alfred. This Working Group is the most recent one within the IUSS, and it is great to see that despite being still at an emerging state, this group is extremely dynamic and energetic.

Digital Soil Morphometrics (Alex McBratney suggested the acronym "DSMorph" to mark the distinction with Digital Soil Mapping) is a relatively newcomer in the soil science world, and has been formalised by Alfred and Budiman in their seminal 2014 paper (Hartemink and Minasny, 2014). It is defined as "the application of tools and techniques for measuring, mapping and quantifying soil profile attributes and deriving continuous depth functions". It is acknowledging the emergence of new tools and quantitative techniques that can be used in soil profile descriptions --- a domain where techniques and toolkits have been quite stable for the past 60 years.

The workshop certainly was a great illustration of the dynamism of this emerging discipline. Around 70 participants from all around the world converged to Madison: Africa (Tanzania), Asia, (Taiwan), Europe (Belgium, Germany, Hungary, UK), North America (USA, Canada), Oceania (Australia, New Zealand), and South America (Brazil). 35 oral presentations, in addition to 7 keynote presentations, were given over two and a half days on the campus. But of course, that was after the very first day of the workshop, which very fittingly was spent in the field. Led by Birl Lowery, with the support of the other colleagues from the University of Madison, we had a fantastic tour of the diverse landscapes and soils of Wisconsin. The first stop was at the West Madison Agricultural Research Station, where we had the opportunity to appreciate (and quantify!) the variety of soils that can occur at quite a short scale: 2 soil pits were dug about 50 meters apart. The large width of exposed soil was also a good reminder of the short-scale, horizontal variations that occur when sampling soils. The second stop was on a farm in the Central Sands region. After a well-earned lunch stop, we had the opportunity to inspect the soil pits using a variety of sensors: portable XRF, portable Vis-NIR, penetrometer... completing of course the more traditional tools of the pedologist. The final stop of the day was at Devil's Lake State Park, where everyone enjoyed the local brews and cheeses that make the pride of Wisconsin, followed by a very scenic walk to wash these down.

Back on the University Campus, day 2 started with a keynote from Alex McBratney, who kicked off the Workshop with an introductory perspective to Digital Soil Morphometrics. The Workshop was split into a range of thematic sessions, spanned over 2 and a half days:

- Prediction of soil properties on the soil profile itself
- Imaging techniques on the soil profile
- Soil depth functions
- The role of Digital Soil Morphometrics in Digital Soil Mapping (DSMorph for DSM!)

Amongst the keynote presentations, the presentations of Erika Michéli and Jon Hempel, who are both involved in the Universal Soil Classification, did a very good job at putting the workshop in perspective with the latest developments in soil classification techniques. Erika and Jon demonstrated how Digital Soil Morphometrics techniques are central, through the concept of depth functions, to the emerging Universal Soil Classification system, and to a variety of other soil information products. The question of how to derive such depth functions was at the core of Budiman Minasny's keynote talk: Budi mentioned in particular the role of high resolution scans of the soil profile to do so. The development of the concept of depth functions for soil classification is also generating the need to create a collaborative depth functions library.

A range of presentations gave an overview of the sensing techniques that can be useful in Digital Soil Morphometrics. Daniel Hirmas and Brian Slater, for example, showed interesting approaches to capture the profile's surface variations at very high resolution, and derive a variety of indicators related to soil structure and soil texture. In his keynote, Markus Steffens also demonstrated the use of a hyperspectral camera on the surface of soil profiles, allowing to derive predictions of soil attributes at a very fine scale. Taking a different approach, another interesting keynote was provided by Matt Aitkenhead: Matt showed how a simple digital camera from a smartphone can be combined with environmental data to put soil organic carbon estimates directly in the hands of the farmer or casual user. Different users, different technologies, and different techniques. Darrell Schulze presented a very promising approach to the visualisation of soil structure using computer graphics, and procedural modelling. While Darrell's project is focusing on the visualisation of soil structure, several of us noted that those tools could be investigated to measure soil structure. Finally, Jose Dematte's presentations underlined the value of these tools as a data provider for digital soil mapping products.

The Workshop concluded on an excellent discussion session led by Alex McBratney, and concluded by Alfred Hartemink, who announced that the next Workshop on Digital Soil Morphometrics will be organised in Aberdeen (Scotland) by Matt Aitkenhead and his team at the James Hutton Institute. On light of the week of discussion in Wisconsin, and I certainly hope to be able to take part to the next workshop.

Finally, I would address a warm word of congratulations to Alfred and Budiman for the organisation and leadership: I think the Workshop was a great success, and I enjoyed very much every aspect of it. I also want to thank and congratulate the very helpful team from the Department of Soil Science who helped to organise and run the conference (Birl Lowery, Bill Bland, Jenna Grauer-Gray, Jenifer Yost, Kabindra Adhikari, Luis Reyes-Rojas, and Benito Bonfatti). Photos from the workshop is available here.

References

Hartemink, Alfred E., and Budiman Minasny. "Towards digital soil morphometrics." Geoderma 230 (2014): 305-317.