ΠΕΔΟΜΕΤRΟΝ

Newsletter of the Pedometrics Commission of the IUSS

Issue 41, December 2017

Chair: Budiman Minasny   Vice Chair: Lin Yang   Editor: Jing Liu

From the Chair

Happy World Soil Day 2017 and Welcome to the 41st issue of Pedometron.

We had a lavish 25th Anniversary celebration in Wageningen in June. We were treated with lots of great science and also amazing cultural and recreational feasts. Needless to say, it was a great conference. Thanks to Gerard and the organizing committee for putting the fabulous conference.

At the conference, we learnt that Pedometrics had a humble start and now we have made an impact in soil science. The research started from analyzing data of small experimental fields and now we have studies that model the world. FAO recently released a cookbook on Soil Organic Carbon mapping. Much of this cookbook deals with procedures that we introduce in operationalizing DSM and GlobalSoilMap. We are proud that our research is now adopted worldwide.

At the conference, Alex and Jaap also outlined challenges in Pedometrics: It is mostly supply driven, not demand driven, and whether results of pedometrical applications really used by land users? And finally Soil is mostly absent in Earth system science. We will strive to work on these issues.

This issue of Pedometron brings you highlights of the conference, various exciting news and a back to the past article. We also have a couple of discussion topics. The first is raised by David on the better use of term “evaluation” rather than “validation”. We got responses from the Master Pedometricians. Following on, I asked the question to Expert Digital Soil Mappers, Do finer-scale covariates give us better maps?

Enjoy reading. I hope to see you in RIO next year.

Budiman Minasny

In This Issue

Pedometrics 25th Anniversary, 26th June - 1st July2017.

Report on the 25th Anniversary Pedometrics Conference

By Grant Campbell

At the end of June, I attended my first ever international PhD conference in Wageningen, the Netherlands. This was an extra special conference for me because I was a part of the conference's 25th anniversary and it provided me with a great opportunity to meet fellow PhD students and experts in my research field of Digital Soil Mapping (DSM) in an international environment. It also provided me with an opportunity to present my latest work I've been doing on DSM of soil properties across Great Britain. During the week there were a lot of things on offer and I will explain my experiences in this post below. According to the book, there were 5 workshops, 6 keynote speakers, 130 oral talks and 120 poster presentations: a LOT of science! The conference was attended by 260 participants from 51 countries.

Day 1: After an initial welcome from Gerard Heuvelink to the conference, we were treated to two excellent keynote presentations from speakers Alfred Hartemink who discussed how the pedon is at the core of Digital Soil Morphometrics and Laura Poggio who looked into fusing data and expert knowledge for Digital Soil Assessment. I personally found both presentations thought provoking and very insightful.

The day then moved into some parallel sessions where attendees could go along to a range of oral presentation sessions on themes centered on soil monitoring, uncertainty in soil data and predictions & geostatistics, data fusion in soil mapping and modelling, proximal soil sensing and Digital Soil Morphometrics.

Report on Pedometrics Conference by Grant Campbell

Day 2: The next day kick started with two keynote presentations from Tom Orton on sample support in connection to soils data and Jianghao Wang on what Pedometrics can learn from big data.

After lunch, we were then introduced to a few presentations at a very exciting and interesting 25th anniversary celebration session. The session began with the Alex McBratney and Jaap De Gruitjer talking about the start of Pedometrics and the journey it has undertaken to date. After the duo completed their presentation, Murray Lark stepped up to discuss the past, present and future methods of mathematical models in Pedometrics – another very entertaining session.

The session was then suspended for a coffee break featuring excellent 25th anniversary cake and a quiz centered on soils!

Day 3: The penultimate day of the Pedometrics Conference began with a talk from Mike Kirkby from the University of Leeds who looked into implementing a conceptual model of physical and chemical soil profile evolution. This was another engaging talk and made us think about the soils more specifically that what we may not have thought about on a regular basis.

The Conference dinner was a rather interesting evening which took place in a Zoo…Yes, you may laugh, but the busses took us to the Burgers Zoo just outside Wageningen for the meal at night! A rather interesting and unusual choice of location but one that worked extremely well! After a great evening of food, drink and terrible dancing, the conference rolled onto the final day for me on the Friday.

Day 4: The final day, after an excellent night before, began with a keynote talk from Rob Beens and Peter van Erp talk who were discussing the marketability and business concept for fast and reliable soil sensing. This was very different from the other talks and I found it particularly interesting indeed.

The session was ended by an excellent final keynote from another one of the founders of DSM research Budiman Minasny who looked at unearthing soil change with what he called 'dirty' data! It was a very thought provoking talk which certainly got me thinking about how to deal with data that may be incomplete or not ideal to work with.

Winners

A short mention to congratulate the students who won awards at the Pedometrics Conference notably Wartini Ng from The University of Sydney for Best Oral Presentation entitled: Rapid sensing of petroleum-contaminated soils with mid infrared spectrometers and Alexandre M.J-C. Wadoux from Wageningen University for Best Poster Presentation entitled: Modelling the soil information content of mid-infrared spectra at European scale.

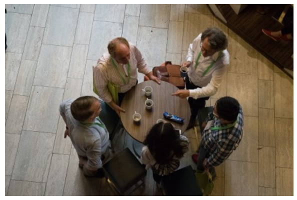
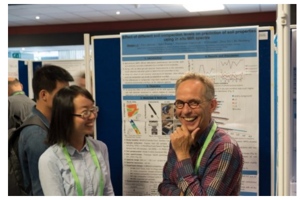
Reflections

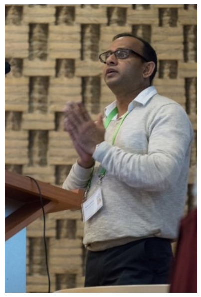
As the conference closed, I reflected on one of the best experiences I have ever had since I started my PhD. Being my first international conference made it extra special and I would encourage all students during the PhD's to try to attend a conference internationally. It is definitely a different experience from attending a conference at national level as you can mix it up with different colleagues and networks from different countries and get a sense of the different philosophies that people are implementing in their own research. The Pedometrics Conference was an invaluable experience for me and I'll certainly take a lot of the feedback, connections and thoughts on board going forward for the rest of my own PhD.

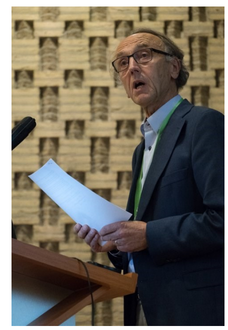
For photos and further information see www.pedometrics2017.org

Grant Campbell is a PhD student based at The James Hutton institute and Cranfield University. Grant's PhD aims to improve the scale of Scotland and England and Wales soils data by using Digital Soil Mapping (DSM) and modelling algorithms for a range of stakeholders.

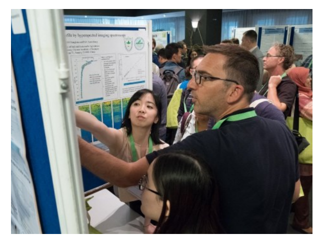
Photo Gallery

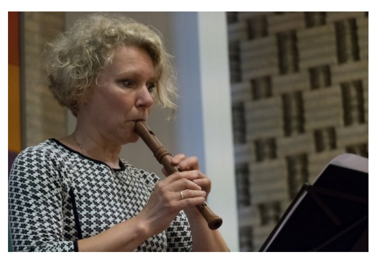



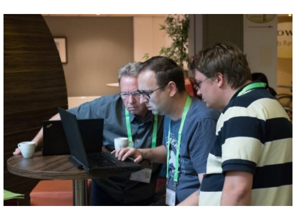













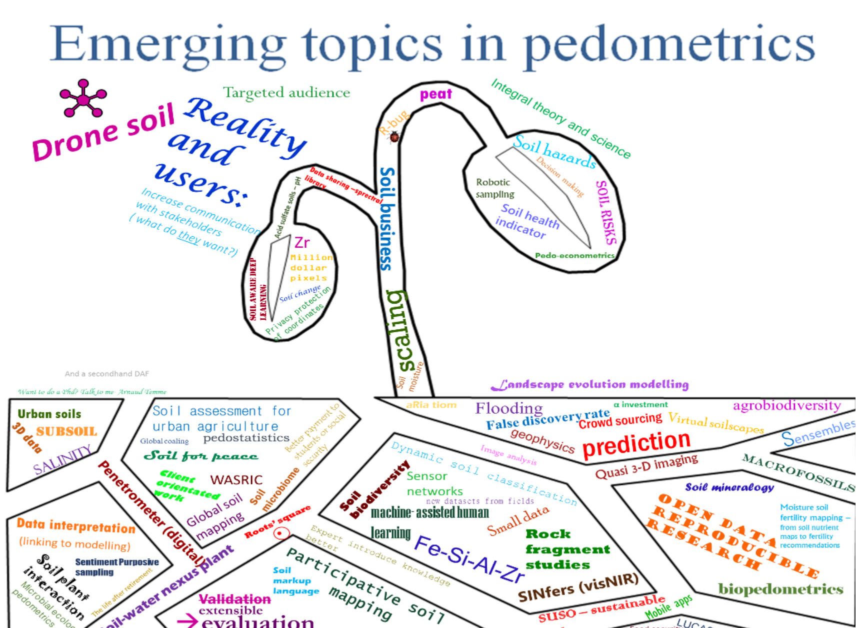
Photos courtesy of Guy Ackermans

Impressions of the 25th anniversary conference of Pedometrics in Wageningen 2017

by Florence Scherer – Wageningen UR

I had not heard much about Pedometrics before the conference in Wageningen. In fact, I had never been to any scientific conference before. I had no idea of what would await me in the Hof van Wageningen except for endlessly long presentations, given in a scientific language that I probably could not understand and ending in discussions that would not come to a point. At least that was the impression I got after a short research on the internet. In one article, the author compared scientists on a conference to students on a school yard, trying to impress each other, improve their own reputation, making new friends and finding out about the weaknesses of the other person's research[1].

Nevertheless, I was excited to see for myself what a conference is all about. I felt heartily welcomed from the first day on, even though I was still an undergraduate from a non-pedometric field of study, just there to observe the events. Soon I realized that I was not the only non-pedometrician. Interviewing participants on their reason to work in the pedometrics field, I was confronted several times with the answer: "I'm not a real pedometrician!". There were statisticians, a soil physicist, who claimed to be the total opposite of a pedometrician, ecologists, mathematicians and a researcher that explained he's working on the context of soils in the landscape by means of statistical methods, but he is not a real pedometrician either. It were mainly the scientists that regularly participated in the Pedometrics conferences that would call themselves "a pedometrician". Maybe it is not about the field that you do research in, that define a pedometrician, but how you are engaging with the community.



Emerging Topics in Pedometrics

The more conference days passed by, the more I could grab the idea of the pedometrics workfield. The empty poster "Emerging topics in Pedometrics", that I hung up on the first day of the conference, was filled with colorful terms and small sketches on the last day. It shows the different ideas of participants about the future of their field. The listing ranged from technical improvements to mentioning the big challenges that the world is facing, such as "climate change" or "food security". "Reality and users", "biopedometrics" and "machine assisted human learning" are just some of the ideas that appear on the poster. Next to that several colleagues made appeals about working more interdisciplinary.

Participants had very different reasons to join the conference. It ranged from joining the workshops and excursion, networking, and making international acquaintances, learning some new tricks on working with R, meeting former colleagues, representing the researcher's country and its projects, to getting feedback on their work and finding links between their own field of work and pedometrics. But also visiting the Netherlands was a simple reason to come, or to

[1] " Was auf Tagungen wirklich geschieht", Axel Bojanowski, June 2017, spiegel.de
http://www.spiegel.de/wissenschaft/mensch/regeln-fuer-konferenzen-was-wirklich-auf-tagungen-geschieht-a-1146826.html

talk to Alex McBratney before he retires. The conference seemed to be an inspiration for everyone, for some people it was next to growing their network in this field also a chance to challenge themselves. It seemed to me that most people left the conference positively surprised how much more they gained from the conference than they expected. Several times I overheard people saying: “this has been the best conference in a long time.” I was surprised, how fast my idea of a boring conference changed and how enthusiastic and motivated I left after the week. As a matter of fact, this conference, like any other conferences, comprised sitting in a conference hall for a week and listening to up to 20 presentations a day, some more monotonous than others and most of them talked about terms that I never heard before. It was tiring from time to time, and I admit sometimes it was just the coffee that kept me in focus while I was hopping from room to room to hear talk after talk.

I was impressed how the participants could keep up the spirit and ask relevant questions to the presenter and give feedback till the very last day. Conferences are the kind of events that scientists work for all year. To get results and write papers and then get feedback and some fresh thoughts on their topic. This is one part of a successful conference, to have space for discussions, presentations, feedback and questions. The other part is having activities outside of the scientific context, where the organizing team put a great effort in. I do not know whether it was the daily morning runs through the city or the many other activities that created the warm and friendly atmosphere but all of it played a role in connecting the participants.

Pedometrics 2017 was characterized by its positive competition and its rather small size, where people had enough time to talk to each other. In my opinion, the feeling of being a community grew each day and had its peak at the conference dinner on Thursday night in the Burgers’ Zoo. Next to great food, awarding young pedometricians, good conversations, people enjoyed themselves on the dancefloor where a live band was playing. The following day, a lady reflected that on a conference there is always something new to learn, might it be new dance moves or the difference between the term validation and evaluation. In the end, I realized a scientific conference isn’t just important for the individual participant to share and gain knowledge but also to create a general feeling of being united in the work. Conferences create and define the community of pedometricians, that work together and motivate each other. In the keynotes, the look into the future of Pedometrics always came along with reminders to find a framework, a purpose of your research to bring the next step in the discovery of new methods. From other people I heard that in the future it is important that pedometricians focus on the translation of information into decisions for policy makers or companies, others talked about big data and citizen science, tackling uncertainty, increasing the amount of data or learning to handle data scarcity. I am excited to see what will come in the next years for the Pedometrics community.

My Pedometrics 2017

By Colby Brungard, New Mexico State University

It was my privilege and delight to attend Pedometrics 2017 in Wageningen, NL. As this was the 25th anniversary of Pedometrics, there was a celebratory atmosphere that permeated the entire conference which, for me, amplified the exciting nature of Pedometrics research. After researching and teaching as the lone Pedometrician in my institution, I found it quite refreshing to associate with a large (and ever growing) group of Pedometricians, who are working on so very many fascinating aspects of digital soil mapping, digital soil morphometrics, soil and landscape evolution modeling, proximal soil sensing, and soil monitoring.

In addition to the excellent scientific program, the delightful social activities were a highlight of this meeting. I participated in several activities, though the rugby match, cycling tour to the Doorwerth Castle, and closing social were particularly memorable. On the evening of Tuesday June 27th approximately 50 pedometricians gathered on the Duivendaal field near the conference venue for games of Rugby and football (soccer to the Americans!). I participated in the Rugby game with absolutely no idea of how the game was played. With John Triantafilis calling instructions, the match was attempted and an exciting time was had by all. Rather surprisingly, no one was seriously injured throughout the match.

The closing social was held at the Royal Burgers' Zoo in Arnhem. Although held in a zoo, this closing social was not one to miss! After a light hors d'oeuvre we walked and visited through bush, desert (reminiscent of the American southwest), ocean, and tropical rainforest environments. Throughout the dinner I enjoyed visiting with many old and new friends capped off with a delicious dinner. Following the dinner many participated in the after-dinner dance. It was a lot of fun to rock, disco, jive, strut, samba, and conga with pedometricians young and old from around the globe!

Pedometrics 2019

Pedometrics 2019 will be in Guelph, Ontario, Canada. Hosted by Asim Biswas and colleagues, at University of Guelph with a theme: "Connecting Existing to the New Data and Methods for Process-based Ecosystem Modelling"

The proposed dates are June 4-7, 2019. The host promises an exciting adventure in the City of Guelph, and also a field trip to Niagara Falls and Niagara Peninsula.

Best Papers in Pedometrics 2016

D G Rossiter, Chairman Pedometrics Awards Committee

e-mail: dgr2@cornell.edu

Thirty votes were received. Unlike the 2015 award, this time there were no clear favourites, and in fact two papers received the most the first place votes. Seven of the eight papers received at least one first-place vote, and the voters had quite a range of rankings, with quite some controversy (expressed to me in e-mails with the votes) about what is pedometrics and what are research vs. review. In any case the papers were all excellent and of high importance, either in terms of their methods, their concepts, or their results. After re-assigning votes through seven rounds, they both ended up with 14 votes. For the first time we have a split victory; in DOI order:

- Viscarra Rossel, R.A., T. Behrens et al. (2016). A global spectral library to characterize the world's soil. Earth-Science Reviews 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012
- Poggio, L., Gimona, A., Spezia, L., & Brewer, M. J. (2016). Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. Geoderma, 277, 69–82. https://doi.org/10.1016/j.geoderma.2016.04.026

The Viscarra Rossel et al. paper was appreciated for the global coverage and integration of diverse sources of information into a consistent product that can be used for pedometric research. As these authors say, "We hope that this work will reinvigorate our community's discussion towards larger, more coordinated collaborations. We also hope that use of the database will deepen our understanding of soil so that we might sustainably manage it and extend the research outcomes of the soil, earth and environmental sciences towards applications that we have not yet dreamed of."

The Poggio et al. paper was appreciated for the introduction of a cutting-edge geostatistical technique to the digital soil mapping toolbox. As the authors say "The Bayesian framework using INLA offers a viable alternative to existing methods for digital soil mapping, with comparable validation results, important computational gains, good assessment of uncertainty and potential for integrated modelling uncertainty propagation."

Who says your vote doesn't count? If one more person had voted for either of these papers higher than the other, that paper would have won.

Thanks to all who nominated papers and who voted after having read these excellent papers. For those who have not read all the papers, I encourage you to find time to do so; you will be informed of the state-of-the-art in Pedometrics and stimulated in your own work. This award, along with the 2015 award, was presented at Pedometrics 2017 in Wageningen.

Congratulations to

Sabine Grunwald

For Being elected as Fellows of the Soil Science Society of America. Sabine has been an active member of the Pedometrics commission. She was the vice-chair of the commission 2002-2006.

Johan Bouma

Emeritus Professor of Soil Science, Wageningen University, the Netherlands, will receive the Dokuchaev Award of the IUSS at the World Congress of Soil Science in Rio de Janeiro, Brazil, in August 2018. Prof. Bouma is known for the functional characterization of soils, emphasizing use of data from soil surveys to improve soil physical characterization of soils. He is Fellow of the Soil Science Society of America and also the recipient of its Presidential Award. In 2017, he received the Alexander von Humboldt Medal of the European Geosciences Union. He is an Honorary Member of the International Union of Soil Sciences. He was the first soil scientist to be elected as a member of the Royal Dutch Academy of Sciences, Arts, and Letters.

Geoderma Best Paper 2016

The Geoderma Best Paper Award is chosen every year and started with 2013 publications. The Geoderma Editorial Board is asked each year to nominate papers from the previous year, and a short-list is then created for voting by the editorial Board. Free Articles online until October 2018.

Best review of 2016:

Linking soils to ecosystem services — A global review authored by Kabindra Adhikari & Alfred Hartemink.

Best original paper of 2016:

Farm-scale soil carbon auditing authored by Jaap de Gruijter, Alex McBratney, Budiman Minasny, Ichsani Wheeler, Brendan Malone & Uta Stockmann

CREA Best Paper

The Council for Agricultural Research and Economics of Italy (CREA) awarded the best 2016 paper published to Sergio Saia, a researcher at CREA.

Schillaci, C., Lombardo, L., Saia, S., Fantappiè, M., Märker, M., Acutis, M., 2017. Modelling the topsoil carbon stock of agricultural lands with the Stochastic Gradient Treeboost in a semi-arid Mediterranean region. Geoderma 286, 35–45. doi:10.1016/j.geoderma.2016.10.

The paper mapped the topsoil C stock of the Sicilian agricultural area showing the amount of carbon accumulated and the role of soil in carbon capture. It also provides guidance to encourage measures to increase carbon accumulation in agricultural soils.

GlobalSoilMap 2017 Conference
Moscow, Russia

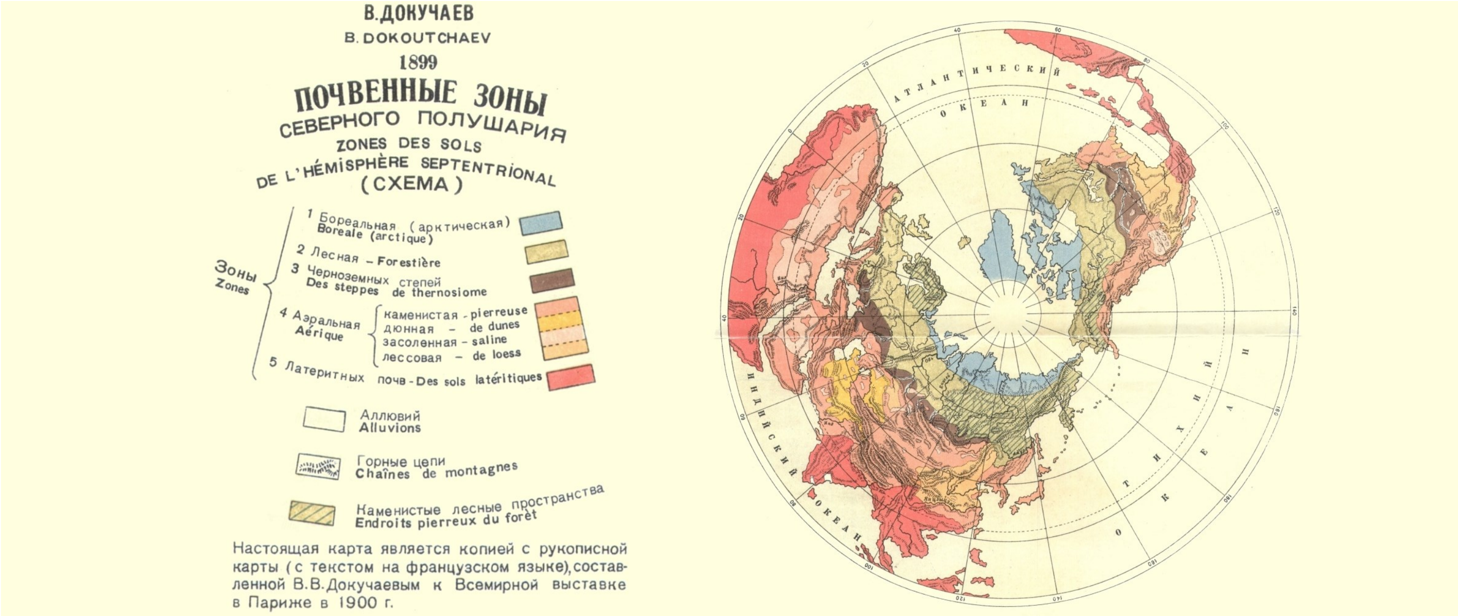
José Padarian

On the 4th— 6th July of 2017, the Agrarian and Technological Institute of People's Friendship University of Russia (RUDN) hosted the Second GlobalSoilMap conference. The conference was co hosted by the V. V. Dokuchaev Soil Science Institute, a world renowned organisation. RUDN is one of the leading universities of Russia, having a history of uniting students of different countries. Currently, RUDN has more than 145 are represented, making it the most multicultural university of the world.

The most obvious outcomes from GlobalSoilMap are the status and advances of the national soil products. In this edition, we were happy to hear outcomes from many countries including Australia, Chile, China, Hungary, USA, Mexico, Russia, Scotland, and Argentina. Even when Digital Soil Mapping (DSM) method is not yet applied at a national extent, the results of regional studies are important for the development of the discipline. That is the case for Indonesia, Cameroon, Brazil, India, Australia, Nigeria, Mexico and Turkey, with applications ranging from supporting government policies to coffee beans production.

Not everything is about the final product. DSM has a strong theory behind it, which is reinforced by methods and the data that were utilised in prediction. The active GlobalSoilMap community constantly updates the theory, methods and types of data we use. In this conference, different techniques were explored, including Bayesian modelling, boolean logic, and knowledge discovery. The use of pedotransfer functions to map soil properties was also presented. In terms of data for mapping, we see the use of EM, VIS-NIR-SWIR spectroscopies, and also different types of traditional soil maps.

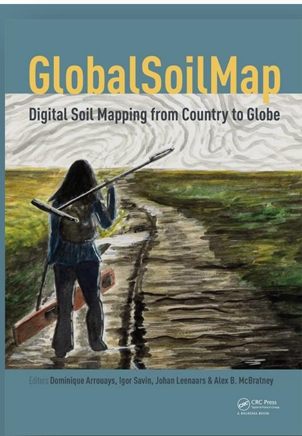
More Photos & Videos at http://globalsoilmap2017.ru



One of the recurrent discussions is the comparison between traditional and digital soil maps. It was a healthy discussion, which is important to revisit from time to time, in order to clarify not only their differences, but also their similarities in terms of conceptualising the soil. They are usually seen as rivals, but in reality they co-exist in the in spectrum between modelling and management.

Many passionate young students and researchers joined this conference, communicating their ideas and findings either as oral or poster presentations. It is encouraging to see the interest of new generations, probably attracted by the use of new technologies like satellite imagery and coding. The latter generated another important but less polemic discussion. Coding and technology do not replace soil knowledge. They, in conjunction with statistics and other skills/disciplines, are the base of DSM and we should aim to master all of them in order to correctly generate and evaluate GlobalSoilMap products.

All in all, it was a very nice conference, and we look forward to see you all at the next DSM conference to be held in Chile, in December 2018.



GlobalSoilMap:
Digital Soil Mapping from Country to Globe

Edited by: Dominique Arrouays, Igor Savin, Johan Leenaars, Alex B. McBratney

Contains contributions that were presented at the 2nd GlobalSoilMap conference, held 4-6 July 2017 in Moscow, Russian Federation. These contributions demonstrate new developments in the GlobalSoilMap project and digital soil mapping technology in many parts of the world, with special focus on former USSR countries. GlobalSoilMap: Digital Soil Mapping from Country to Globe aims to stimulate capacity building and new incentives to develop full GlobalSoilMap products in all parts of the world.

More Photos & Videos at http://globalsoilmap2017.ru

In Memoriam

Ivan Rodrigo Orjuela Osorio

Ivan was a PhD student from the Universidad Nacional de Colombia and worked on soil carbon prediction by spectroscopy. Ivan passed away on the evening of Friday 30th June 2017, shortly after a tragic traffic accident in Wageningen. He was an enthusiastic participant in Pedometrics 2017 and received excited feedback for his conference presentation. He was making plans with all the new scientific contacts he made at the conference. We are deeply saddened by this loss and the IUSS and Pedometrics community offer our sincere condolences to his family and friends.

Lotfi Zadeh

Prof Lotfi Zadeh, the father of Fuzzy Sets and Fuzzy Logic, passed away at the age of 96, September 2017. Undoubtedly, fuzzy sets and fuzzy clustering have played a major role in pedometrics research. Fuzzy systems allow pedometricians to deal with imprecision and uncertainty in soil measurement and modelling. Quoting Albert Einstein "As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality"

Barry Rawlins

Dr Barry Rawlins from British Geological Survey passed away on Friday 29th September 2017, two years after being diagnosed with cancer. Barry led the Sustainable Soils team at the British Geological Survey. He is known to Pedometrics and Digital Soil Mapping his varied collaborations, mapping soil carbon across the UK, and national and regional scale soil inventory and monitoring. More recently he worked on soil properties at finer scales, developing approaches to the measurement of aggregate stability and investigating the distribution of soil organic carbon within the physical architecture of the soil. He will be missed for his energy and commitment to science, his passionate belief in an evidence-based approach to environmental problems, and his personal warmth and generosity.

RIO 2018

The 21st World Congress of Soil Science will be in Rio De Janeiro, Brazil, 12-17 August 2018. https://www.21wcss.org/

ABSTRACT submission is now open and the deadline has been extended to January 20, 2018.

The Pedometrics commission and associated working groups will convene several Symposia please submit your work and see you in Rio!

1.5.1 - **Global soil carbon modeling**: This symposium is to bring together scientist both from the soil science and global change community involved in global soil carbon modelling.

1.5.2 - **Crucial techniques for the critical zone: Soil morphometrics, monitoring & modelling**: This co-organized workshop focus on the cross-pollinations between the research areas of soil evolution modelling, sampling and monitoring and morphometric methods.

1.5.3 - **Reconciling pedometrics and pedology**: This symposium will bring together scientists both from pedometrics and pedology to create a synergy on advancing soil resource management.

Interdivision Symposium: Pedodiversity and biodiversity: We will bring closer the relationships between pedodiversity and soil biodiversity. Different approaches to describe and estimate soil diversity, effects of biodiversity on pedodiversity, and vice versa will be discussed.

WG - Digital soil mapping: **Progress in digital soil mapping**: Recent advances in Digital Soil Mapping techniques

WG - Digital Soil Morphometrics: **Soil imaging and image analysis at multiple scales**

WG - GlobalSoilMap: **GlobalSoilMap: progress and challenges**: The state of the art of GlobalSoilMap. The symposium will review the specifications and the state of progress of GlobalSoilMap products delivery.

WG - **Proximal soil sensing (PSS)**: State-of-the-art soil sensing technologies; modern statistical methods for analyzing soil sensor data; methods for multi-sensor data fusion; and methods for sampling and fine resolution digital soil mapping using sensor data

WG - Soil Monitoring: **Soil monitoring evolving tools and challenges**: The objectives will be to provide a review of the state of the art related to Soil Monitoring. Special attention will be paid to the issues related to the use of new techniques or proxies for monitoring, to recent advances in cost effective sampling designs and harmonization issues.

WG - Universal Soil Classification: **Progress for the development of a Universal Soil Classification System**: The symposium will provide an overview of the history and progress of the development of a Universal Soil Classification System.

Election of Chair & Vice Chair

Election of chairs and vice chairs of the International Union of Soil Sciences (IUSS) divisions and commissions has begun. If you are a member of your national soil science society, you should have receive an email from your national society. Please vote though you national society, Voting closes on 31 December.

The candidates for the Pedometrics commission are:

Chair: Vera Leatitia MULDER (The Netherlands)

Assistant professor at the Soil Geography and Landscape group, Wageningen University, The Netherlands. She has a background in remote sensing and soil science. She was awarded the prestigious Agreenskills postdoc and best paper awards (from Pedometrics and Geoderma). She has contributed to international research projects e-SOTER, GlobalSoilMap and SATEX.

Her main research interest involves spatial, temporal and soil-landscape modelling for understanding large-extent soil and ecosystem dynamics. In her current work, she focusses on soil carbon dynamics and climate change adaptation, contributing to the '4 per 1000 Initiative – Soils for Food Security and Climate'. URL: https://www.researchgate.net/profile/VL_Mulder

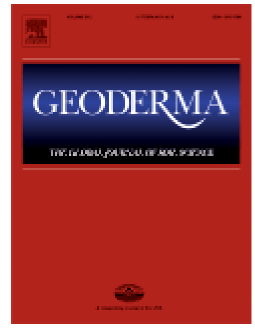
Chair: Thomas BISHOP (Australia)

Associate Professor at the University of Sydney where he teaches soil, water and data science. He completed hisy PhD at the University of Sydney and spent time overseas in the United Kingdom and USA as a postdoc before returning to his current position. His main research interest is in predicting the variation of soil properties in space and time using geospatial data, statistical methods and mechanistic models. His current focus is on soil carbon and moisture with a particular interest in predict at multiple spatial resolutions. URL: http://sydney.edu.au/science/people/thomas.bishop.php

Vice chair: Nicolas SABY (France)

Nicolas Saby obtained his PhD from AGROCAMPUS OUEST in 2008 with a thesis on digital soil mapping and soil monitoring. Nicolas currently works at INRA Infosol as a manager for the French programs of soil survey and Monitoring. He has 17 years of experience in modelling, quantifying and mapping the spatial and temporal variation and uncertainty of soils. He works across the discipline on optimal design of soil surveys, and the spatial and temporal variation of soil properties. He was awarded the Silver medal from the French Academy of Agriculture and Best Paper in Environmental Microbiology (2009) and Pedometrics (2010). URL: https://www.researchgate.net/profile/Nicolas_Saby

Pedometrics Special Issue



The long-waited Advances in pedometrics: Special issue of Pedometrics 2015, Cordoba is finally being published in Geoderma Vol. 311. http://www.sciencedirect.com/science/journal/00167061/311?sdc=1

This issue contains a selection of papers from the 2015 Pedometrics conference hosted at the University of Córdoba, Spain, between 14 and 18 September 2015. The conference presented many fascinating presentations, and we can only select a handful of papers showing the latest development in pedometrics.

This special issue begins by two papers oriented at generating soil maps over large areas, using the latest geostatistical techniques. Pásztor et al. (2017) use traditional and newly tested DSM classification methods to obtain a national soil-type map for Hungary. An et al. (2017) focus on the important question of how to select representative samples from multi-source samples for mapping soil organic matter content in Anhui province, China.

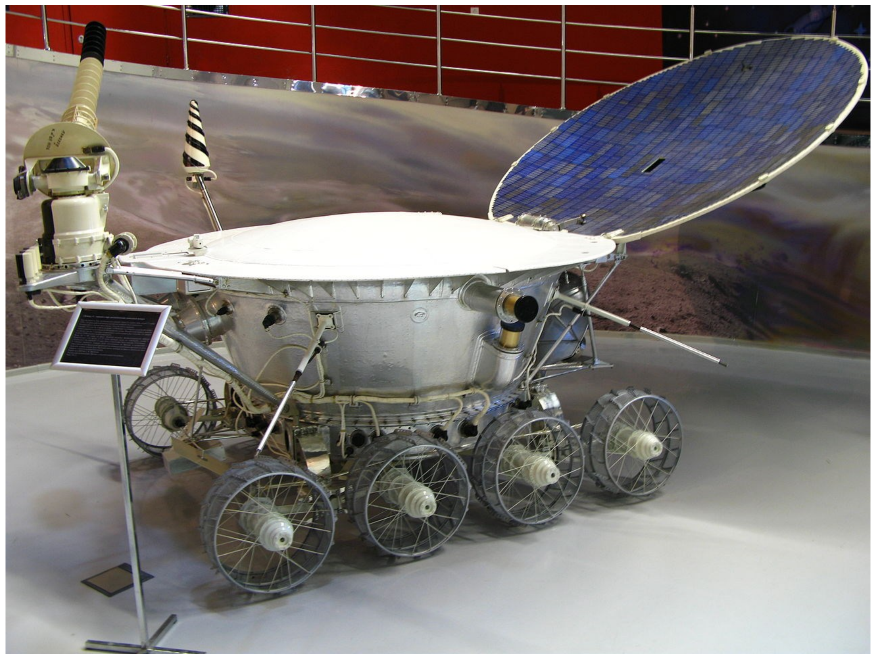
The next study uses artificial neural networks to predict the occurrence of acid sulfate soils in Jutland, Denmark (Beucher et al., 2017). A correct identification of areas where such soils occur is critical for preventing environmental contamination with acid leachates into watercourses and estuaries. Another soil contamination study by Fernández et al. (2017) evaluated diffuse pollution from heavy metals in the area of Asturias, a rough mountaineous region in the north of Spain. The paper by Vincent et al. (2017) applies spatial disaggregation to predict soil information for the region of Brittany at a finer resolution than the original information. The study by Massawe et al. (2017) also addresses the problem of generating accurate soil information, especially in areas with little or no observations. Their study is in the Kilombero Valley of Tanzania, an area of important agricultural expansion. One important conclusion of their work was that 1 arc SRTM and 12 m WordDEM data resulted in the same predictions of soil taxa, encouraging the use of the freely available SRTM data. Finally, a study by Valverde Arias et al. (2017) works on how to define an optimal map of homogeneous soil areas for its use in agricultural insurance schemes. Given that the importance of insurances in agriculture, especially in the light of future climate change, this study in the coastal region of Ecuador presents an interesting example.

Finally, three studies were after a detailed understanding of soil properties and contribute to our fundamental understanding of soil properties, from the field scale to the microscopic scale. Román-Sánchez et al. (2017) present a catchment-scale study on the variability of soil carbon stocks in SW Spain and show how solar radiation and NDVI are key controls. Their study explores the detailed variability of not only soil carbon, but also bulk density and stoniness. The microscopic complexity of soil has recently been shown to have important implications for understanding different processes, ranging from soil carbon protection, over soil microbiology to soil physics. Torre et al. (2017) apply a novel multifractal analysis to process X-ray microtomography imagery. Martín-Sotoca et al. (2017) also use fractal concepts to propose a new segmentation method. They also propose a novel method to create 3D synthetic soil images, called truncated Multifractal Method, an application especially interested for modelling – for example of soil hydrology or biology, where such data is often needed.

Lunokhod 1, There's Nothing New under the Moon

Proximal Soil Sensing from a long way away, and a long time ago

Alex McBratney & Budiman Minasny



Lunokhod 1 (Луноход, moon walker in Russian) was the first astro-remote-controlled rovers landed on the Moon by the USSR. It was carried to the moon by the Luna 17 spacecraft, landed on November 17, 1970 in the northwestern part of Mare Imbrium, and operated until September 29, 1971.

Lunokhod 1 is an eight-wheel rover with a mass of 756 kg, and has four panoramic cameras. A report by the RAND Corporation authored by Simon Kassel in 1971 described the capabilities and sensors of Lunokhod 1, obtained from the USSR literature. The report stated its sensory equipment comprised narrow-band television systems, mechanical probes, a chemical analysis system, X-ray and cosmic ray detectors, and assorted temperature and pressure gauges.

Specifically, on page 9, the report stated the Lunokhod 1 conducted chemical analysis of the lunar soil using an X-Ray Fluorescent instrument referred as RIFMA (Roentgen Isotopic Fluorescence Method of Analysis):

"The RIFMA spectrometer consists of an isotopic source which irradiates the investigated sector of the path and thus induces X-ray emission from atoms of various elements constituting the lunar soil. The X-ray receiver is a system of specially developed proportional counters with characteristic filters. The information on the energy spectrum and intensity of X-rays is recorded by a 64-channel amplitude analyzer." This is essentially a mobile XRF.

In addition, Lunokhod 1 also performed mechanical analysis of the lunar soil with PrOP (Prinoro Ochenki Prokhodimosti – Cross-Country Capability Evaluation Instrument):

"Of all the scientific experiments, the greatest amount of information was obtained from measuring the bearing strength of lunar soil and the mechanics of interaction between the soil and the vehicle's wheels. The soil-testing program employs two types of devices. One is a penetrometer which measures the strength of soil and operates only when the vehicle is at rest. "

The PrOP penetrometer has a cone of 60 degree and 5 cm base. The other device measures the traction and operates only when the vehicle is in motion by means of a ninth wheel, which rolls without slipping. The penetrometer is a conical paddle driven into the soil and rotated. A set of sensors continuously measures the force required to rotate the paddle, its rotation angle, and depth of penetration. This operation is performed regularly every 15 to 30 m of the path. The mechanical properties of the soil are also studied from the image of the wheel tracks transmitted by television. The bearing capacity of the soil varied from 0.2 to 1.1 kg/cm^2 (on moon it is 3—18 kPa while on earth it is equivalent to 20—108 kPa). Its shear strength is 0.02 to 0.09 kg/cm^2 (or 320-1500 Pa).

A digital terrain model was produced based on the analysis of image stereopairs.

Within 11 months, Lunokhod 1 travelled 10.5 km and performed 25 lunar geochemical soil analyses and measured penetration resistance 500 different locations.

The average elemental concentration of lunar soil from Lunokhod 1 analysis is: Si: 20%, Fe: 12%, Ca: 8%, Al: 7%, K: 1%.

We are still not sure why this rover technology is not yet widely-available for earth.

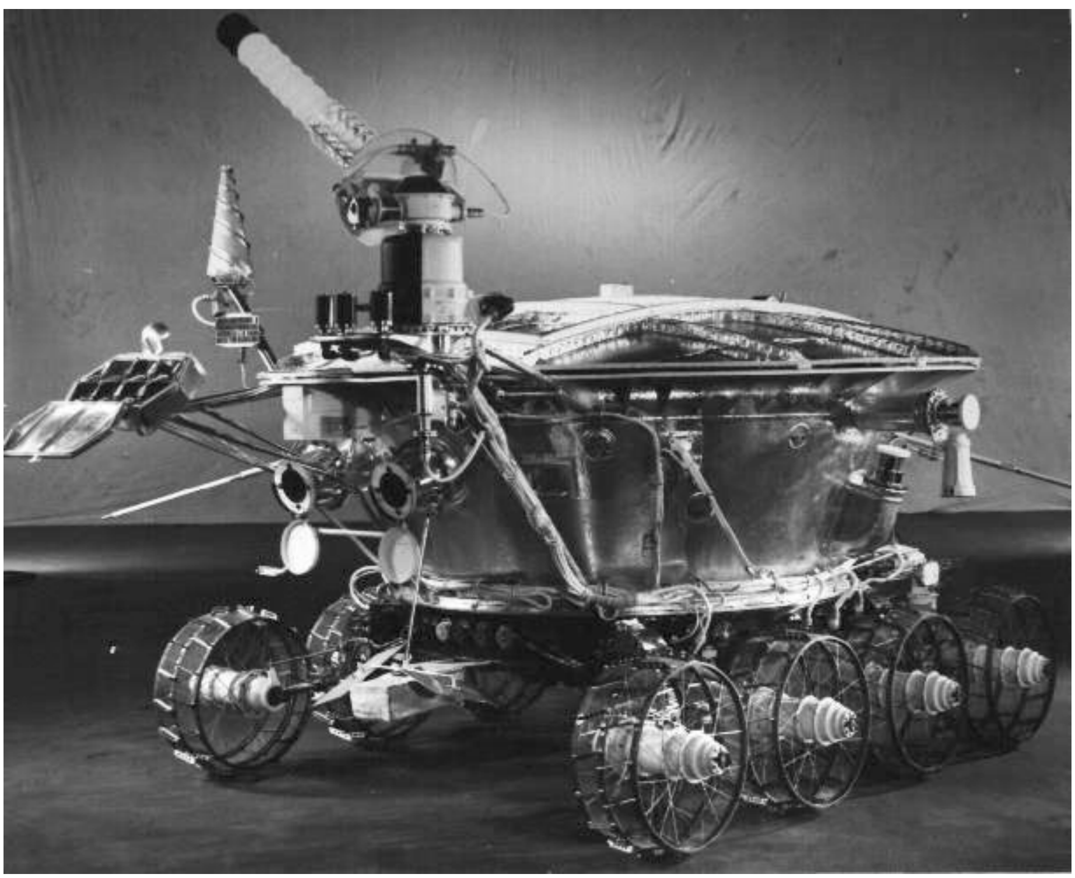
References

Basilevsky, A.T., Abdrakhimov, A.M., Head, J.W., Pieters, C.M., Wu, Y. and Xiao, L., 2015. Geologic characteristics of the Luna 17/Lunokhod 1 and Chang'E-3/Yutu landing sites, northwest Mare Imbrium of the Moon. Planetary and Space Science, 117, pp.385-400.

Carrier, W.D., Olhoeft, G.R. and Mendell, W., 1991. Physical properties of the lunar surface. Lunar sourcebook, pp.475-594.

Kassel,S., 1971. Lunokhod-1 Soviet Lunar. Surface Vehicle. A Report prepared for Advanced Research Projects Agency.

Ulivi, P., 2004. Lunar Exploration: Human Pioneers and Robotic Surveyors. Springer Science & Business Media.



Maps and models are never valid, but they can be evaluated

D G Rossiter

david.rossiter@wur.nl; d.g.rossiter@cornell.edu

At Pedometrics 2017 I began my invited talk on "Past, present, & future of information technology in pedometrics" with a slide explaining why I think pedometricians should not use the term "model validation", but rather call this process "model evaluation". That slide provoked quite some comment, so I thought it would be useful to explain the reasoning for this proposed semantic shift.

This is by no means my idea. It originated more than 20 years ago in two seminal articles written by the historian of science Naomi Oreskes (1994, 1998)[1], presently Professor of the History of Science at Harvard University; these are well worth reading to get a deeper philosophical and practical justification than I give here. I was introduced to these articles and the use of "model evaluation" about five years ago by colleague Janneke Ettema at the University of Twente, an atmospheric scientist faced, like pedometricians, with modelling natural systems. I was convinced and have tried ever since to refer to "model evaluation". As I will explain, this is not only more correct, especially in communication with decision makers, but also opens up possibilities for deeper discussions of our models than are possible with the term "model validation".

Oreskes' context was the application of numerical simulation models of natural systems for predictions, these to be used to direct public policy, and the communication to decision makers of the uncertainty in the information obtained from these models. Pedometrics is active in this sort of modeling, see for example the Vereecken et al. (2016) review of soil process modelling; (these authors consistently use the term "validation"). However, we are rarely directly in the line of fire of the public and decision makers, especially in our papers written for our peers. Still, our soil maps are used for contentious issues such as soil pollution, wetland regulation, and agricultural subsidies. Our pedotransfer functions are used in hydrologic models for flood hazard mapping and surface and groundwater pollution maps. Our models for soil hydrology, soil survey interpretations and land valuation have legal force in some contexts.

The core of Oreskes' sophisticated argument, as I rephrase it, is that the word "valid" in English and any language where this word derives from the Latin "validus" (strong, powerful, effective), the relevant definition in the OED is: "Of arguments, proofs, assertions, etc.: Well founded and fully applicable to the particular matter or circumstances; sound and to the point; against which no objection can fairly be brought." In normal speech we use phrases such as "valid argument", meaning that it is correct. A related meaning is used in "valid passport", i.e., the passport can be used to travel between political entities. Thus, the usual sense of "valid" is "true, not false", a binary concept. If I take my birth certificate to the local administration to establish my identity, they require that it be "validated" by the responsible official where it was issued. It is then a "valid", and can be used to establish my identity. Oreskes: "[validity] denotes the establishment of legitimacy, typically given in terms of contracts, arguments, and methods." In pedometrics we can speak of a "valid" method for a laboratory determination or sampling design, but only in the sense that the method has been established as a standard.

---

[1] Naomi Oreskes has written several semi-technical and popular books on the history of science and science policy, well worth a place on your nightstand. I especially enjoyed her edited compilation of first-person accounts on how the theory of plate tectonics was proposed, rejected, confirmed and accepted: *Oreskes, N. (Ed.). (2001). Plate tectonics: an insider's history of the modern theory of the Earth. Boulder, Colo.: Westview Press*. This story is told much more briefly in *Oreskes, N. (2013). Earth science: How plate tectonics clicked. Nature News, 501(7465), 27. DOI: 10.1038/501027a*. She has also dealt with how unscrupulous scientists may play-for-pay: Oreskes, N., & Conway, E. M. (2010). Merchants of doubt. New York: Bloomsbury Press.

Maps and models are never valid, but they can be evaluated

The public, including decision makers, thus equate talk of a "valid" model with one that is correct and true. What is at stake is scientific credibility. The term "valid" is only positive, and the public is rightly suspicious of scientists who over-sell their models either because these scientists cynically feel the public does not understand nuances, or because they consider that the policy is too important to be left to the public, or because they ignore Niels Bohr (quoting Robert Storm Petersen, Danish cartoonist): "Prediction is very difficult, especially if it's about the future". Oreskes provides a devastating analysis of all three faults in the infamous Club of Rome "The Limits to Growth" models from the early 1970s. The spectacular failure of their predictions set back serious discussions of what are in fact the limits to growth.

In pedometrics or any kind of statistical, process, or simulation modelling we never claim that the results are valid in this sense. Indeed, that is why we compute so-called "validation statistics" and hold out "validation samples" on which we compute these. But the result of these computations is some "degree of validity", which is a contradiction in terms. The statement that the RMSE, RPD etc. have certain values when the model is applied to independent observations is certainly useful information, but even an RMSE of zero does not prove a model is valid in the wider sense, and as the public uses the term. That is, the model may not be a correct simplified representation of reality, even if the match with observations is good. The classic example here is the Ptolemeic model of the heavens, which, once calibrated, had excellent agreement with observations as known prior to 1610 but failed when Galileo showed the impossibility of the model to explain the phases of Venus.

Pedometric models and their outputs have many sources of uncertainty. Some sources are more or less easily quantified, for example, uncertainty of calibration data due to sampling error and measurement procedures. We can account for these by simulation and sensitivity analysis. Uncertainty in model form can be quantified by comparing the results of various plausible model forms on the same data. These are good examples of quantitative evaluation. But some uncertainty cannot be measured; in particular, are all relevant processes or factors included in the model? This is of primary importance for evaluation of predictive power in new contexts, which are especially of interest in public policy discussions. A recent and controversial example are models of possible C sequestration in soils used to promote the "4 pour mille" concept (Minasny et al. 2017).

What are we really doing in the so-called "validation" step of modeling? Answer: we are evaluating the model. That is, we are determining to what degree the model is useful in our problem, whether the model form and assumptions are justified, whether the data at hand are sufficient to give a reliable answer, how far the model can be extrapolated in space and time. The OED sense of "evaluation" here is "the action ... estimating the force of probabilities, evidence, etc.". Another sense "the action of appraising or valuing", in our case appraising the value of the model to our problem. For example, how useful is a predictive map made by digital soil mapping techniques? What level of certainty can the map user expect? Is this sufficient for the map user's proposed applications?

On the positive side, the term "evaluation" opens up the discussion to more than statistical measures of output agreement with independent observations. We can now discuss the model form, model assumptions, modelling choices, selection of evaluation criteria, selection of evaluation sample etc. -- i.e., we are "evaluating" the entire modeling process, a much richer implication than the word "validating" can provide. Oreskes: "Quality can be evaluated in several ways: on the basis of the underlying scientific principles, on the basis of the quantity and quality of input parameters, and on the ability of the model to reproduce independent empirical data. All of these things can be discussed, but none of them should be discussed in either/or terms." So what we now refer to as "validation statistics" can be better described as "agreement of model output with independent observations". This then forces us to describe the plan by which independent observations were selected, and what population they represent.

Oreskes identifies model flaws of four kinds: theoretical, empirical (imprecise or limited measurements), parametrical and temporal. All of these can lead to disagreement of model output with reality. But which are causing this? Theoretical flaws are due to our poor understanding of processes, and these are quite difficult to identify. The empirical flaw is of course dominant in pedometric studies -- our observations are a tiny fraction of the population. Parametrical flaws arise with model simplification, e.g., assuming isotropy and second-order stationarity when computing an empirical variogram from limited observations. Temporal flaws, which pedometricians might revise to spatio-temporal flaws, arise when extrapolating into the future or unobserved regions. There is no way to know that conditions will be the same as those that produced the observations on which the calibrated model is based.

Maps and models are never valid, but they can be evaluated

Consider ordinary kriging. Theoretical flaws refer to the theory of random fields on which kriging is based -- is this an accurate representation of local soil spatial variability? Is the random field second-order stationary? Further, even supposing we have such a random field, does the selected variogram model form correctly represent its structure? Empirical flaws are due to the sparse sample on which we calibrate a variogram model. Even given a proper variogram model, the parametrical flaw is the fitting of variogram parameters from the limited sample. Finally, the spatio-temporal flaw is in assuming the fitted model can be extrapolated. This does not imply that maps made with ordinary kriging are not useful. The so-called "validation statistics" do give some idea of predictive accuracy within the context in which the original study was done, and if these are satisfactory we have some basis for using the resulting map. This is a reasonable "evaluation" criterion for the model.

What then do we call the various terms that have been traditionally used in pedometrics papers?

1. "Model validation" becomes "model evaluation", and should be described in the broader sense explained above, as "fitness for use" and “appropriateness of modelling approach”.

2. "Validation dataset" (or "observations") becomes "independent observations [not used in modelling]". Another possibility is "Assessment dataset" (or "observations"). This is a short way to contrast this with "calibration dataset": Calibration vs. (quantitative) assessment.

3. The process of numerically comparing model predictions to independent observations is one part of model evaluation, and should be referred to as "predictive accuracy assessment". This supposes that the assessment dataset is representative of the target population. Note that this need not be the population sampled for calibration, if the interest is in extrapolation.

4. "Validation statistics" becomes "agreement between model and observations".

5. The term "cross-validation" is retained, because (1) "cross-evaluation" sounds awkward, (2) it is limited to a specific method of using the data to evaluate the model, and (4) is rarely used in communication with decision-makers. However, the terms "to cross-validate" or "was cross-validated" should not be used, instead something like "model output and observations were compared with statistics from 10-fold cross-validation ..."

I illustrate these proposals with some revisions of phrases from a recent paper on which I am co-author (Zeng et al. 2016):
Original: "First, due to the low density of samples in this study area, we used leave-one-out cross-validation to evaluate the results"
Revised: "First, due to the low density of samples in this study area, we used leave-one-out cross-validation to evaluate the predictive accuracy of the results"

Original: "...each sample was validated individually on the basis of the calibration set compiled from the remaining dataset"
Revised: "...each observation was compared to its prediction made from the model calibrated on the basis of the other observations"

Original: "*n* is the total number of validation samples"
Revised: "*n* is the total number of observations not used in model calibration"

Original: "...the leave-one-out cross validation method was used to validate the mapping results."
Revised: "...the leave-one-out cross validation method was used to assess the relative accuracy of the mapping results."

Evaluations of this proposal are welcome!

References:

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263, 641--646. https://doi.org/10.1126/science.263.5147.641

Oreskes, N. (1998). Evaluation (not validation) of quantitative models. Environmental Health Perspectives, 106(Suppl 6), 1453--1460.

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., & Winowiecki, L. (2017). Soil carbon 4 per mille. Geoderma, 292, 59-86. https://doi.org/10.1016/j.geoderma.2017.01.002

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., & Young, I. M. (2016). Modeling Soil Processes: Review, Key Challenges, and New Perspectives. Vadose Zone Journal, 15(5). https://doi.org/10.2136/vzj2015.09.0131

Zeng, C., Zhu, A.-X., Liu, F., Yang, L., Rossiter, D. G., Liu, J., & Wang, D. (2017). The impact of rainfall magnitude on the performance of digital soil mapping over low-relief areas using a land surface dynamic feedback method. Ecological Indicators, 72, 297-309. https://doi.org/10.1016/j.ecolind.2016.08.023

Careful writing should make that clear, adoption of a half-baked linguistic rule will not

Murray Lark

I am entirely unconvinced by the argument that David has put forward, and I think it quite wrong. First, though, I do not think we are in any fundamental disagreement about what the process of model validation does and does not do. Tom Addiscott, formerly leader of the soil process modellers at Rothamsted, strongly advocated the use of the word “validate” to describe the process of testing model predictions against data, as opposed to “verify”. He argued this on the basis that “verify” means to establish truth, whereas “validate” means to establish validity which means “strength, or fitness for purpose”.

Why is this an important distinction? In modern English the word “true” can be affirmed or denied only in respect of a proposition, some statement of fact. For example “there is no largest prime number”, “there will be earthquakes in East Acton tomorrow”, or “the pH of the topsoil at location x is 6.5”. These propositions can be verified, the first one by mathematical proof, and the second two by empirical observation. If I visit x and measure the soil pH and get a value of 6.2, then the prediction was false. However, the pedometrician would not conclude that the prediction was meaningless on this basis. She knows that any prediction has attendant uncertainty, and her assessment of it will take account of this, comparing, for example, the magnitude of the prediction error with observed natural or analytical variation. She might also observe whether the prediction, while not precise, is useful to the manager. Will it lead to a bad decision? In short, she assesses the usefulness of the prediction, and the evidence that it encodes information about the variable of interest. That is what we mean by its “strength, applicability, well-foundedness”, which are synonyms of “validity” in the Shorter Oxford English Dictionary’s definition of “valid”.

Now David’s argument, in my opinion, takes a wrong turn where, having observed that the English “valid” derives from Latin "validus" (strong, powerful, effective), he segues into the quite unfounded statement that

In normal speech we use phrases such as "valid argument", meaning that it is correct.

Well I must disagree. A valid argument is one that is correctly constructed. It is a formal property of the argument. If someone offers an argument, the conclusion of which is certainly true, it does not follow that the argument is valid. For example

Some bears are brown
Grizzlies are bears
Grizzlies are brown

is not a valid argument, although it has a true conclusion and true premises. We can test the conclusion in the field, but the validity of the argument is tested by the logician. In modern logic we turn the argument into symbols and manipulate these to test validity. This process would show that the structure of the argument is not sound. Consider another argument

All bears are brown
Polar-bears are bears
Polar-bears are brown

This is a valid argument, but its conclusion is untrue. The classical logician would identify the argument as an example of Aristotle’s valid syllogism in BARBARA, if the premises are correct then the conclusion must be. The zoologist notes that the first premise is wrong, which is why the valid argument can lead to a false conclusion, as it does in this case.

So I reject the idea that, in English, a “valid argument” means an argument with a true conclusion. It does not mean that at all. Rather it means an argument which is sound and fit for purpose. Exactly what we want to establish about a method for pedometrical prediction.

David goes on to say

A related meaning is used in "valid passport", i.e., the passport can be used to travel between political entities. Thus, the usual sense of "valid" is "true, not false", a binary concept.

I think this argument a complete non-sequitur. A passport is valid if it has not been withdrawn by the issuing authority, if it has not expired, if it has not been defaced etc. That makes it fit for purpose, “valid” in the sense of The Oxford English Dictionary, and Tom Addiscott’s “model validation”. We cannot say that a passport is “true” or “false”. What about a forged passport? Well here again a bit of logic, in particular Bertrand Russell’s logic, will help us. Grammatically the expressions “Dutch Passport”, “Blue Passport” and “Forged Passport” are the same, an adjective qualifies a noun. In the first two cases the adjective tells us what kind of passport we are dealing with. In the third case it actually tells us that the object in question is not a passport (a document issued by a competent authority), but is rather a document run up by some shady character in exchange for a handful of used ten pound notes. Russell would have recognized this as an example of a case in which the structure of ordinary language obscures the logical structure of an expression, leading us into muddles. I respectfully suggest that David has ended up in such a muddle, because the only sense that can be given to the expression “valid passport”, is “fit for purpose”, back to Tom Addiscott and the OED.

When we have a pedometrical model we have a statistical prediction, a conditional expectation (or some other moment of the conditional distribution of the variable of interest) and one or more measurements of the uncertainty of that quantity, treated as a prediction of an unknown variable. When we validate that model we check whether the prediction, as compared with the observation, gives us grounds to regard the model as fit for purpose, and “doing what it says on the tin”. There are various statistics which we might use to help with this, and some of them, pace David’s argument, will indeed indicate degrees of validity. A prediction set with an RMSE of 1.0 comes from a model with a greater degree of validity (strength, fitness for purpose) than a prediction set with an RMSE of 10, although both may be useful if the standard deviation of the variable in question is 100. The standardized square prediction error, or coverage probabilities of the prediction, tell us whether the quantification of the uncertainty in the predictions seems to be sound (which might not be the case if, for example, the variogram has been influenced by outlying data).

The Shorter Oxford English Dictionary gives a particular sense of the verb “to validate”. It reads thus

[in] Computing etc. …. confirm or test the suitability of (a system, program, etc.)

I cannot think of a better way to describe what we do in pedometrics when we validate predictions; and I, for one, will fight to preserve such a valuable word.

I very much hope that editors of soil science journals and others will not take David’s recommendation on board. They should, however, insist that authors are always clear and explicit about what they are doing when they validate predictions, and the claims that are based on such validations. We are not claiming to show that predictions are “true” (even where careful analysis could attach meaning to such a statement), we are testing the system used to generate the predictions and quantifying their fitness for purpose. Careful writing should make that clear, adoption of a half-baked linguistic rule will not.

We can avoid confusion and misunderstanding if we make the change

Gerard Heuvelink

Agreement on the meaning of terms is crucial to science. If different people interpret terms differently then this may cause confusion and obstruct scientific progress. One of the terms that is regularly debated is ‘validation’, because its meaning is not the same to all. For instance, in my experience soil physicists and soil hydrologists interpret “the model has been validated” as “the model has been proven suitable”, while my interpretation and that of most pedometricians of “the model has been validated” is “the model predictions were put to the test and compared with independent observations”. So even if a model does a very poor job, has a large Mean Squared Error and low Amount of Variance Explained, it would still be considered validated.

David reminds us of the work of Naomi Oreskes that defines validation as “the establishment of legitimacy” and states that a valid model “does not contain known or detectable flaws and is internally consistent”. This definition agrees fairly well with the interpretation of soil physicists and soil hydrologists and refutes that of us, pedometricians. Even closer to the interpretation of soil physicists and soil hydrologists comes the definition of Rykiel (1996), who defines model validation as “a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model”.

So should we abandon the term ‘validation’ and replace it with ‘evaluation’ when we refer to the process of comparing model predictions with independent observations? I started writing this comment having in mind that my answer to this question is ‘no’ (because replacing validation with evaluation is not a solution in all cases, because making the change in practice may prove to be very difficult, and because we pedometricians know very well what we mean by validation so why bother?), but the more I think about it the more I agree with David. It looks as if our interpretation of model validation is quite different from that used in mainstream science. This is clearly undesirable because we do not live on an island and need to communicate well with other (soil) scientists. We can avoid confusion and misunderstanding if we make the change.

So David has my full support. From now on I will try to use evaluation instead of validation whenever appropriate, although I realise that this may be difficult because it is oh so easy to fall back into old habits!

Gerard Heuvelink

Reference

Rykiel, E.J. (1996), Testing ecological models: the meaning of validation. Ecological Modelling 90, 229-244.

A problem well-stated is a problem half solved

Philippe Lagacherie

David Rossiter provides us with a bright and very convincing argument for replacing “validation“ by “evaluation” in the DSM papers. This could be seen as some kind of a Don Quichotte’s enterprise since everybody in the DSM community has used validation for a long time now and it is never easy to change habits, especially the bad ones. Yet, it is far to be a vain fight. Remember Kettering’s quote saying “a problem well-stated is a problem half solved”. Furthermore, as David points out, “validation” is misleading for users of our soil maps which are not familiar with our jargon. Therefore, I thank David for rowing upstream and I invite my colleagues rowing with him.

That being said, I would like to take an opportunity to stress that model validation – or model evaluation – remains in 2017 a third-class passenger - if not the clandestine one - of our way towards an operational DSM. Although valuable contributions have been brought in the past to set the theoretical framework of model evaluation, we are still lacking in ‘valid’ and realistic solutions that could be really applied on real study areas. The current DSM literature does not help us a lot in dealing with some basic questions like determining how large an ‘assessment dataset’ must be, deciding whether a difference of 0.02 in R^2 means a difference in performances between two models or not, or evaluating how much the uncertainty is underestimated by the often-unescapable cross-validation method. A more holistic approach of model evaluation that would include knowledges on expected soil patterns and on limitations of

A valid map is a map that is made by a valid model

Dick Brus

Introduction

First of all I would like to congratulate David Rossiter with his nicely written, thought provoking article on map validation and evaluation. The article forced me to gather my thoughts. This comment is the result of this struggle, I hope it is of use for others .

Model validation

Validation is used both for models and for maps: model validation and map validation. A map can also be seen as a model, but here I use model for a deterministic or statistical equation. Let me start with model validation, more specific validation of statistical models. In my basic statistics classes I teach the students that in linear regression modelling several assumptions are made:

- a linear relation between response variable and independent variables (predictors)
- the data are independent
- the variance is constant
- if we want confidence intervals of estimated means and or prediction intervals for unobserved units in the population, we also assume a normal distribution for the residuals

I stress that we always should check the validity of these assumptions, for instance by making scatter plots of residuals against fitted values to check the linearity and the constant variance assumptions, making a Q-Q plot of the residuals to check the assumption of a normal distribution, and estimating an experimental variogram of the residuals to check the assumption of independent data. I also point them to formal test such as the Moran’s I test for spatial autocorrelation of the residulas, the Shapiro-Wilk test for normality and the Levene’s test for constant variance within groups. If none of the modelling assumptions is clearly violated, we treat the model as a valid model. It is like statistical testing of an hypothesis, the null hypothesis being that the model is valid. Unless there is clear evidence or theory against the null hypothesis, it is not rejected, and we take the model as a valid model.

Map validation

If there is no evidence that one or more assumptions are violated, we can use the model to predict for unobserved units in the population, i.e. mapping. What is usually done in map validation is computing map quality indices such as the mean error (ME), mean squared error (MSE), standardized squared prediction error (SSE) *et cetera* for continuous soil maps, and the overall, user’s and producer’s accuracies for categorical maps. The closer the map quality indices are to their ideal values, the more this supports the validity of the model. So when the aim of estimating the map quality indices is to check the validity of the model underlying the map, these quality indices serve as validation statistics, and the estimation of these statistics can be named *model* validation. An accuracy plot (Goovaerts, 2001) and a variogram of the prediction errors are examples of graphs that are also tailored at checking the validity of the modelling assumptions. If we define a valid *map* as a map that has been made by a valid *model*, the estimation of map quality indices and graphs *tailored at checking the validity of the model* can also be named *map* validation.

David Rossiter states that ‘even an RMSE of zero does not prove the model is valid’. David illustrates the problem with the nice example of the Ptolemeic model of the theory of the heavens. Even if the match between predictions and observations is perfect, this does not prove the validity of the model. This is entirely analogous to statistical hypothesis testing. The null hypothesis (in our case ‘the model is valid’) simply cannot be proven. We can only reject it, so that we conclude that it is very unlikely that the model is valid (very likely that it is invalid), or not reject it. In the latter case we have not proven that the model is valid, we only have not enough evidence against it. Making this clear does not convince me that the estimation of map quality indices such as the RMSE cannot be tailored at map validation, and therefore should not be named map validation, but preferably map evaluation.

Valid estimates of map quality indices

I would like to stress here that the map quality indices ME, MSE, SSE, overall, user’s and producer’s accuracies et cetera should be defined in terms of population parameters, not as sample averages. For instance, the ME should be defined as the average of the errors over all N units in the population. A subset of these units (sample) is used to estimate this population ME. We are uncertain about the population ME, and it is important to quantify our uncertainty, for instance by estimating its standard error. We can then statistically test the null hypothesis ‘the population ME equals zero’. If this hypothesis is not rejected, this supports the correctness, validity of the model underlying the map; if it is rejected than we have evidence that the model and map is not valid.

As argued in our ‘Sampling for validation’ paper (Brus et al., 2011) the validation sample can best be selected by additional probability sampling, so that the map quality indices can be estimated by design-based inference. Design-based estimates are model-free, no modelling assumptions are made, and as a consequence no objections can be made against the estimated map quality indices and their standard errors. In other words, with a design-based sampling strategy we obtain valid estimates of the map quality indices. The quality of these estimates does not depend on the quality of a model, simply because there is no such model. This is of great importance when these map quality estimates are used as validation statistics, i.e. for checking the validity of the model underlying the map.

In case of non-probability sampling, we need a spatial model of the prediction or classification errors (multivariate distribution of the errors). Several modelling assumptions must be made, such as about stationarity of the mean and of the variance of the errors, and about the covariance of the errors. These assumptions make the estimates prone to discussions. Knotters and Brus (2013) describe an example of how different modelling assumptions about classification errors lead to largely different estimates of the map quality indices. So with non-probability sampling and model-based inference I would not qualify the estimates of the map quality indices and their standard errors a priori as valid estimates. The validity of the modelling assumptions must be carefully checked: is it realistic to assume that the errors are independent, and is it realistic that the mean and or variance of the errors are constant throughout the area? For this reason model-based estimates of map quality indices are less suitable for model (map) validation.

Conclusion

- I define a valid map as a map that is made by a valid model.
- I see map validation as statistical testing of a hypothesis about the validity of the model that is used to construct the map
- The null hypothesis is ‘the model is valid’. This null hypothesis cannot be proven. It can be rejected, in which case we conclude that it is very unlikely that the model is valid, or not be rejected, in which case our conclusion is that we do not have evidence that it is not valid.
- The hypothesis can be tested using map quality indices such as the population ME, MSE and SSE as test statistics. The map quality indices then serve as validation statistics.
- When used as validation statistics, the map quality indices can best be estimated by a design-based sampling strategy, involving probability sampling and design based estimation. No model of the prediction or classification errors is used, which guarantees the validity of the estimated map quality indices.

References

Brus, D. J., Kempen, B., and Heuvelink, G. B. M. (2011). Sampling for validation of digital soil maps. European Journal of Soil Science, 62(3):394–407.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. Geoderma, 103:3–26.

Knotters, M. and Brus, D. J. (2013). Purposive versus random sampling for map validation: a case study on ecotope maps of floodplains in the Netherlands. Ecohydrology, 6:425– 434.

What’s in a name?

Budiman Minasny

A rose by any other name would smell as sweet… according to a famous author.

I have the benefit of seeing all comments beforehand, but I will still throw in my comments. I agree mostly with Murray that changing a name is not a solution. We still do the same thing, either good or bad validation. Having a new terminology that will be used by a handful of pedometricians would create further confusion.

Hastie et al.’s *Elements of Statistical Learning* (2nd Edition) talk about model selection and model assessment. Model selection: estimating the performance of different models in order to choose the best one. And Model assessment: having chosen a final model, estimating its prediction error (generalization error) on new data. They further added that:

“*If we are in a data-rich situation, the best approach for both problems is to randomly divide the dataset into three parts: a training set, a validation set, and a test set. The training set is used to fit the models; the validation set is used to estimate prediction error for model selection; the test set is used for assessment of the generalization error of the final chosen model*.”

Most of the time in soil data, we only have a limited number of samples, and thus we cannot divide our data into 3 sets as above. Thus, we do validation or cross-validation. Having a new term “evaluation” does not solve this problem.

Following on Philippe’s clandestine concern, I would like to point out some of the bad habits we like to do on the use of goodness of fit measures:

(1) We like to compare our R^2 or RMSE to other papers to justify our method or results are better. E.g. Our model validation of organic carbon content in such field in Australia has an R^2 of 0.50, which is much higher than the results of Murray (2014) in Scotland who only reported $R^2 = 0.30$. Clearly you can’t compare that! Unless you are comparing the same field or the same set of data.

(2) We like to justify our validity of goodness of fit by some kind of made-up standard. E.g. Our model has RPD > 2 which means it is accurate and shows a good prediction. Not Necessarily True! There is no basis for such classification, it depends on the data and it is a relative measure.

I would rather that we (including myself) make sure these bad habits are not to be repeated in papers. As Naomi Oreskes in 1994 said “Models can only be evaluated in relative terms”. And the quote I like most from Oreskes is “Verification and validation of numerical models of natural systems is impossible.”

Do finer resolution covariates produce more accurate DSM models?

I pose this question to a number of Digital Soil Mappers. It seems that in this digital era, we always want higher resolution covariates, but in several publications and our own work we found that better relationships, most of the time, come from coarser-scale relationships. It is a paradox!

Why do we always want higher resolution covariates?

Alessandro Samuel-Rosa

I remember attending some soil science meetings in 2011 and 2012. The soil scientists were talking about using digital soil mapping techniques to produce soil maps for Brazil. At that time, there was a consensus that this task could only be accomplished if up-to-date, higher resolution satellite images and digital elevation models were available for the entire country. I was a newbie in the field of digital soil mapping, but for some reason this argument sound unrealistic to me. Landsat images at 30-m spatial resolution were already available for the entire Brazilian territory. A 30-m spatial resolution digital elevation model covering the country had also just been released. In addition, the Brazilian government acquired RapidEye imagery also covering the entire Brazilian territory (5 m spatial resolution). Looking at this scenario, I thought that we already had the necessary data for moderate to large scale digital soil mapping in Brazil. So why were soil scientists interested in higher and higher resolution covariates? Am I wrong?

The study that we developed – and many many other studies (e.g. using LiDAR in a watershed in US, DEM sampled at various resolutions in Ireland) – showed that soil properties and environmental covariates commonly show stronger correlations when we use an aggregated or coarser version of the covariates. So I guess I am correct. But the reason why this happens still remained unclear to me. It also remained unclear to me the reason why, despite existing evidences, using higher resolution covariates continued to be one of the primary targets of most digital soil mapping projects. This is the paradox mentioned by Budiman.

Digital soil mapping is a field that heavily builds on remote sensing of the Earth's surface as a source of covariates. In remote sensing, we generally collect data to represent the visible features of the surface of our planet, for example, the relief and the vegetation (or land use and occupation). When we want to map a forest with higher accuracy, we need more densely sampled data, that is, higher resolution satellite images so that we can "see" its individual constituents (trees). The same applies to the relief: if we want a "perfect" representation of it, then we have to move away from coarse resolution digital elevation models, that smooth the surface, and aim at collecting Terabytes of data using, for example, laser scanning techniques. By having finer resolution satellite images we generally have more accurate classification of vegetation and land use. In the same way, hydrological models generally produce more accurate results with finer resolution digital elevation models.

Do finer resolution covariates produce more accurate DSM models?

The use of covariates in digital soil mapping comes from the understanding that soil is the product of forming processes. Assuming that this holds, by deduction, we expect that the finer the resolution of the covariates, the more accurately it will represent the environmental conditions, thus producing more accurate digital soil maps. Actually, this rationale is used by all soil mapping schools. In traditional mapping, the cartographic scale of the map is closely related to its accuracy, both thematic and position. Because many of the polygon-based soil maps have not been validated, the cartographic scale is commonly taken as an indicator of the accuracy of the soil map. Based on that argument, in digital soil mapping, we expect a positive relationship between spatial resolution and map accuracy.

Note that I didn't mention about sampling density. This is because we often determine the resolution based on the available covariates. The reasons for these are two (I think): the first was described in the previous paragraphs; the second is driven by an external factor.

Early this year, I participated in new meetings to prepare for a project for producing soil maps of Brazil. An early draft of the project was evaluated by the representatives of a funding agency. One of the main critique was that the project included field work to collect more soil samples. It was then requested that the project should include heavy use of technology, mainly high resolution products of remote sensing, so that the project would be considered for evaluation. The representatives of the funding agency seems to have found it very archaic that we still have to go to the field, open pits and collect soil samples in this 21st Century Era of Technology. This episode shows that we are commonly required to prepare research projects that include the use of new technologies irrespective of their utility. In digital soil mapping, this generally means using finer and finer resolution covariates. This can also be judged in publishing: we can expect that it would be easier to publish a research paper in a high impact journal if we use the newest, highest resolution DEM or technologies.

In summary, I understand that these are the key driving forces that make us always want higher resolution covariates. It is obvious that we really want soil maps to be the most accurate – an R^2 of 0.50 always is a stone in the shoe, specially when we have to explain to others – specially to our colleagues of the soil chemistry lab – that this is a very good result.

Connected multi-scale digital soil mapping, rather than producing finest grain soil map

Brendan Malone

Indeed I think most people within reason will prefer to use fine scale covariates if they are available. And if they are free, all the better. I don't really know what people were using prior to the 3 arc sec SRTM DEM but there was certainly significant enthusiasm in digital soil mapping when this information was freely available. Similarly with the products derived from the various Landsat missions and other sources. Projects such as the GlobalSoilMap and others that have been inspired from that, have largely been to the benefit of the 3 sec STRM DEM and other covariates. Now that the 1 sec DEM is also freely available together with other finer grain variables, I accept there is heightened enthusiasm that better maps can be produced than those that were derived from the 3 sec product.

Whether we interpret better as meaning more accurate or better as in more detail I guess is the question Budiman has proposed. My feeling is that if all other things remain unchanged, using finer grain covariates will not necessarily make more accurate maps. It's possible to happen of course, but more likely the produced mapping will remain as accurate or perhaps even a little less accurate.

My reason for this is that without additional sampling there is a higher likelihood that the existing sampling will not capture the variation that is contained in the fine grained covariates. This will be particularly pronounced in heterogeneous environments. However, in a floodplain environment (which we assume to be topographically homogenous) for example; models using fine grain terrain covariates are not going to be any more accurate than if relatively coarser grain terrain covariates are used. I think there is a potential for accuracy to actually decrease because terrain analysis in these environments using a fine grain DEM often leads to artefacts, highlighting or accentuating features in the landscape that don't really exist in the field. This could lead to problematic parameter estimation resulting in either model under-fitting or over-fitting.

Naturally there are other issues to consider. For example, the target variable to be mapped and the associated correlations with other variables used as proxies for soil forming factors. Even if we permit additional soil sampling to capture the variation in the covariates, the change in scale of the covariates may not be matched by that of the target variable. For example, we might expect more short range variation in target variables such as carbon, nitrogen and pH compared to others such as soil texture or CEC etc. Effectively the model response to changes in covariate grain size are not going to be universal. Such that we would expect correlations to change between covariates and target variable with change in grain size and this could lead to either model accuracy improvement or reductions. Moreover, the response in model to change in grain size will be different for each covariate used, together with the interaction between covariates (which one might expect to be non-linear with changes in grain size).

To address this, I could imagine a model to have a mix of resolutions and scales of covariates. To put this into context; at a farm scale, for example across a flat landscape, I could imagine using a 30m DEM such as from the SRTM, in combination with 5-10m proximally sensed gamma radiometrics and EMI data to model soil variation. The question then is, what should the resolution of the soil map be if a mix of grain sizes are used for the covariates? Using a mix of grain sizes and scales of covariates is not a new thing, but there is probably limited examples of it being used in the operational sense of digital soil mapping. I guess the next question is how one would go about sampling the landscape given covariates of mixed grain sizes? As we know presently, algorithms such as conditioned Latin Hypercube sampling allow one to capture variation in a given mapping domain given a stack of rasters that are aligned to the same resolution and extent. Perhaps if sampling is to be optimised for a particular target variable, then a mix of grain sizes could be used for different landscape variables. More thought will be required to realise this of course, in addition to making allowance for optimising on the basis of multiple target variables instead of a single one.

In summary I think it is far from axiomatic to suggest using finer grained covariates will result in more accurate maps. I think it is largely dependent on the target variable, and the density and coverage of sampling to cover a presumably more heterogeneous environment. There will be situations where going fine grain will make improvements, and places where no improvement can be made. I guess for projects that want to achieve finer grain outputs to what has already been achieved, they will

Why use a fine paintbrush when a broad brush will do?

Jack Hannam

In a survey asking European stakeholders "what improvements to soil data would you like?" the top answer was "finer resolution data". In a paper published a while ago we explored Budiman's question "do finer scale covariates produce more accurate models?" TL;DR : No, not always We found that using a finer scale DEM for the job of predicting soil classes did not really improve the model in some areas. So apparently everyone wants soil mapping to the max but hi-res covariates don't always give us more info.

Why use a fine paintbrush when a broad brush will do?

Budiman's question arrived when I was painting my apartment hallway (I was banned from using the munsell colour chart for choosing the paint colour but it was probably close to 2.5Y 8/2 for anyone interested). Choosing the right tools for the job is key here. A small paintbrush was needed for the more complex painting of fancy woodwork. If I used the same tool for the homogenous expanse of walls, I would still be painting them now. Clearly a roller was much more effective for this task. Similarly in the landscape we know that there are very homogenous and very heterogeneous (and all the things in between) patterns of soils and soil properties, and these vary at different spatial and temporal scales. The problem is we usually don't have all the necessary information (like I did with the hallway) to make the decision of whether to bother with the fine resolution stuff. In some cases it is not really worth going to the trouble of mapping at fine scale resolution because there really is no meaningful change (Discuss! - perhaps for another edition of Pedometron?), in the soil type or soil property of interest at this new improved resolution. The two classic block diagrams in the figures illustrate these comparisons in soil landscape complexity. Sometimes 'traditional' soil maps are very similar at different scales. It's not the resolution of the mapping that is a problem, or that there is a mistake or that the surveyor was lazy, rather it is just a boring soil landscape without much variation.

What do we mean by fine resolution covariates?

We use the SCORPAN factors for our covariates but in reality many of the soils or soil properties we are trying to represent are not controlled by the finer scale datasets that are available. Most of these are DTM products, which is OK if the soil property or type varies as a response to terrain attributes but is not useful when other soil forming factors also explain the spatial variation (such as parent material, although this is not independent of terrain). We might not have very detailed geological maps to match our DTM but can use proxy measurements such as geophysics to fill the gap. But these can only partially explain the covariate at a finer scale. The proxy will only account for some key attributes of the parent material and may not necessarily identify all the geological variation that is contributing to the soil type or property change. In this case the coarser data may have better overall representation of the covariate space that accounts for the variability of soil within that particular landscape.

What of covariates anyway?

Pedogenesis takes time. Many of our DSM covariates to not take this into account and we have all talked about static and dynamic DSM in various Pedometrics meetings over the years (and also links to soil and landscape modelling). But we are still not there yet. Our covariates often represent a snapshot in pedogenic time. Sometimes the covariates themselves each represent a different timescale. In this case the covariates are decoupled from the interactions between the soil forming factors that we are trying to represent and the resulting soil conditions that we observe today. For example we use long term (30 year average) climate datasets but these may have nothing to do with past climates that have determined the soil characteristics that are still evident in the soil profile. A finer scale dataset will not help with this issue and potentially lead to worse predictions than a coarse scale dataset. Similarly how can we represent the effects of short term (decadal) soil management in our land use data? And what of soil properties at depth? Many of the covariate datasets (e.g. from satellites) represent interaction with the surface layers and we still rely on the legacy soil profile datasets to interpolate in 3D.

So to answer the question do fine scale covariates produce more accurate DSM models? No, yes sometimes and don't know. Still more to discuss!

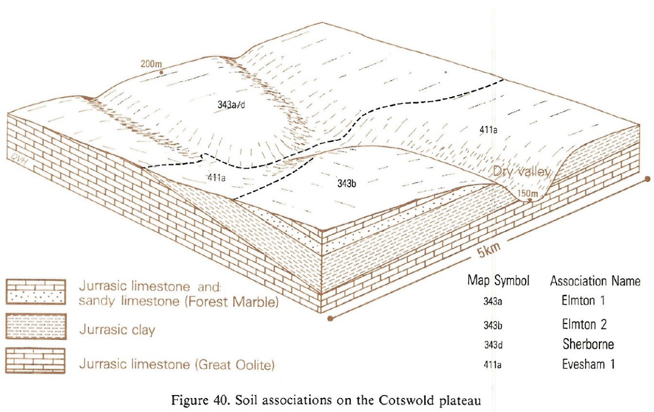


Figure 1. Homogenous or boring soil landscape in England. Probably doesn't require fine scale covariates to improve DSM models © Cranfield University

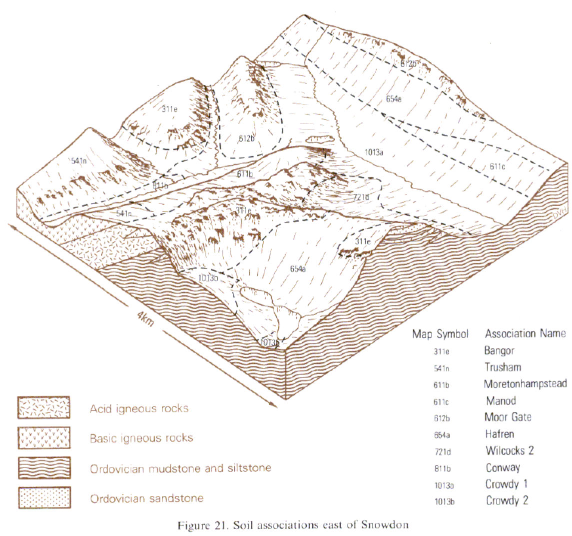


Figure 2. Heterogeneous or more exciting soil landscape in Wales. A good candidate for fine scale covariates to improve DSM models © Cranfield University

Soil Numerology

The UN declared World Soil Day as 5th December in honour of the Late H.M. King Bhumibol Adulyadej of Thailand, who officially sanctioned the event.

Recently, my colleague from Korea told me that they celebrate Soil Day on the 11th November. So why 11th November?

He explained that it comes from the Chinese character of soil "土" (Du), which is made up of two characters: 十 (Shí = ten) and 一 (yī = one). Thus the Chinese numerology for soil = 11.

According to the Latin numerology, the word “soil” has a number of:
s=1, o=6, i=9, l=3, Total=19, 1+9=10, 1+0=1

So indeed soil is always number one.