The International joint workshop of the IUSS Working Groups "Digital Soil Mapping" and "Global Soil Map" "Soil Mapping for a Sustainable future, 2023" Orléans, France, 7-9 février 2023.

Anne C Richer-De-Forges, Dominique Arrouays, Laura Poggio, Budiman Minasny

HAL Id: hal-04102347

https://hal.inrae.fr/hal-04102347

Submitted on 22 May 2023

ΠΕΔΟΜΕΤΡΟΝ

From the Chair

Welcome to the 47th issue of Pedometron.

We had a great start of the year with a highly successful digital soil mapping workshop in Orléans. The event attracted a large audience and provided an excellent platform for discussions on current practices and future directions. As a result of these discussions, two insightful contributions have been published in this latest edition of Pedometron.

In addition to the workshop, the Pedometrics Awards Committee has undergone significant changes to promote fairness in the attribution of the Best Paper Award. The new rules now require an extended awards committee, consisting of a diverse and large panel, to decide on the winner instead of an open vote system. A summary of these changes is available on the Pedometrics website.

The division is discussing the closure and reorganization of the working groups. A working group should not exceed a 8-year time period, beyond which it should be closed, converted, or merged into a commission. Division 1 has nine working groups, several of which are linked to pedometrics. Two have recently been closed, and several have exceeded the 8-year period and are being assessed. The chairs and vice-chairs of the working groups linked to pedometrics have recently been contacted by the division chair. We should take this as an opportunity to optimize the relevance of the working groups linked to pedometrics and to find new chairs and vice-chairs.

Looking ahead, the Pedometrics conference will be held in Las Cruces, New Mexico, from February 5-9, 2024. The event will be organized by Colby Brungard, Salley Shawn, and me, with the theme "Addressing the 10 Pedometrics Challenges". The conference will feature around 10 sessions covering all aspects of pedometrics, and several excursions and activities will be planned to allow participants to socialize and engage in informal discussions.

Finally, the IUSS Centennial will take place in Florence, Italy, from May 19-21, 2024. Session proposals are still being accepted until June 30, and I encourage everyone to submit a proposal so as to cover all aspects of pedometrics. The proposal should include a short description, a chair, and a co-chair.

This latest edition of Pedometron includes its regular features, such as a poem, a conference report for the DSM workshop, and a contribution from the workshop organizers. Additionally, Philippe shares his experience with local digital soil mapping and participative soil science, while the Sydney team contributes a piece on AI. The edition also includes a presentation of a new tool using INLA-SPDE for soil mapping.

Alexandre Wadoux
May 2023, Montpellier

In this issue

Delivered by

Chair
Alexandre Wadoux
Vice-chair
Simone Priori
Editor
Lei Zhang

Pedometrics2024

Addressing the 10 Pedometrics Challenges

Feb 5 – 9, 2024
New Mexico State University
Las Cruces, NM USA

Pre-conference workshops, field trips, active social programs, and post conference excursions are being planned!

Tentative Conference Program

Sunday	4-Feb-23	Pre-conference workshops
Monday	5-Feb-23	Main Conference Sessions
Tuesday	6-Feb-23	Main Conference Sessions, Soccer and Rugby
Wednesday	7-Feb-23	Main Conference Sessions and Field Trip
Thursday	8-Feb-23	Main Conference Sessions, Field Trip, and Conference Dinner
Friday	9-Feb-23	Main Conference Sessions
Saturday	10-Feb-23	Optional Excursions

Mulled Thoughts

By Alex McBratney

Predictive Soil Mapping

I haven't used the term predictive soil mapping very much. In fact, I have been a little critical of its use for what we have called digital soil mapping back in 2003. I think this this specific use of predictive soil mapping comes from language used in say ecological prediction and then transferred by a few over to soil science. I was indeed aware of this terminology when we coined digital soil mapping - a term to integrate everything that was happening in soil spatial prediction around 2002. The reason why I didn't run with predictive soil mapping was my mentor and supervisor Dick Webster had explained to me some 25 years earlier that all soil maps are predictive. It is true - whether they are digital, or not. On the conventional choropleth map, which in passing we should admit we haven't surpassed in their design and attractiveness, the various colours represent the predicted soil class at a quasi-infinity of points – based on soil observations at a few points and an understanding of the landscape. So, all soil mapping is really predictive soil mapping, whether it's digital or analogue.

Machine learning

There's a big to do about artificial intelligence and its subset machine learning, with the advent of ChatGPT which I think is a big breakthrough practically. (I'm too old fashioned to use it to write this. Would I have saved some time? Would have I said what I'm thinking?) We focus a lot these days on machine learning especially for our predictions – of course we worry about interpretability and parsimony (at least I do). We see a lot of over-optimistic applications where too few data are used to predict too much – spreading the information very meagrely. I think we do need to investigate other aspects of artificial intelligence besides machine learning – there is much to be learned and gained there.

Evolution

One of the algorithmic approaches to machine learning is the genetic algorithm which has been used in soil prediction, but not extensively. It was devised by those who were mimicking the evolutionary process in biology and it has turned out to be a successful approach in various applications. This raises a couple of points. First, we haven't really tackled soil dynamics, and in the longer term, evolution - with enough vim. I think it's time to turn the spotlight back on to that problem. Secondly, if we think of how soil evolves does that suggest an approach to machine learning? Learning by soil.

Local Digital Soil Mapping: Another way to climb our Everest

Philippe Lagacherie[1]

[1] UMR LISAH, INRAE, Institut Agro, IRD, 34000 Montpellier, France

Damien is an agronomist who is hired by a water authority to advise farmers for changing their agricultural practices in view of recovering the quality of a neighboring drinking water catchment. George is a forest owner who should select the different tree varieties he will plant next year after he will harvest the actual forest. Aurélie is a civil servant of a municipality who is in charge of the application on a new law aiming to preserve agricultural areas from urbanization growth.

I met recently people like Damien, Georges and Aurélie that asked me the same question: "where I can find the soil data that can support my decision?" This is a very simple question but also a very difficult one. Indeed, this amount to asking for soil data that would i) cover their whole area of interest, ii) be at a spatial resolution fine enough to allow decisions on their decision units (most often parcels) and iii) provide at each location the both accurate and relevant soil information that could effectively support their decisions.

As pedometricians, what can we say to Aurélie, Georges and Damien? Undoubtedly, our major contribution to fulfill their needs has been to product onto regular grids a set of predictions of soil properties with uncertainty estimations as specified and applied in the GlobalSoilMap project (Arrouays et al, 2014) and other connected projects (Global Soil Partnership, SoilGrids). Thanks to some pedometricians that went far beyond the production of scientific papers, Damien, Georges et Aurelie can now download DSM products that cover a whole country, continent or even the planet with spatial resolutions that look fairly compatible with the parcel scale. Unfortunately, recent studies from either pedometricians (Rossiter et al, 2022, Lemercier et al, 2022) or potential users (Djagba et al, 2022) have clearly shown that such DSM products were not able to accurately represent the local soil variations, contrary to what the high spatial resolution of these products may have suggested. Furthermore, it is unsure that Aurélie Georges and Damien can easily make decisions from only the primary soil properties addressed by the DSM products and feel comfortable in handling the uncertainty estimations provided by these products. Pedometricians have therefore many more steps to achieve to fully satisfy Damien, Georges and Aurelie's needs.

What is our way forward? The "business as usual" way would consist in go on enriching countrywide or worldwide centralized soil databases with the hope of producing updated DSM products that could be more accurate and complete than the former ones. I do not believe that this "top-down" approach is the way forward for satisfying our three local soil users. Indeed, every soil surveyor that is aware of the complexity of the soil cover can evaluate how huge would be the amount of soil data and covariates that would be required for capturing everywhere the local soil variations that Aurelie, Georges and Damien have to deal with. They would need to be very patient until many years… and meanwhile give up considering the soil when making their decisions.

In a recent prospective study on the future of soil mapping in France (Voltz et al, 2020), we proposed that the above evoked strategy could be complemented by studies focusing on the local soil variations to match more quickly the needs of local soil users. I am convinced that these studies should and can be conducted following local Digital Soil Mapping approaches especially adapted to small territories. The challenge is to make emerge a viable DSM activity conducted by DSM-trained soil surveyors that could answer with agility to the needs in soil information expressed by the users while respecting established standards of soil information and DSM methods.

Local Digital Soil mapping

Obviously, the development of such operational local digital soil mapping approaches can benefit from the pedometricians' past experiences that have been accumulated within the last two decades. However, two major inflexions are necessary from our current approach each of them being a methodological challenge: dealing with multi-source soil inputs ("multi-source DSM") and becoming a participative activity ("participative DSM"). In the following, I give some details on these two challenges.

Multi-source Digital Soil Mapping

Capturing the variability of soils at a local level need a large density of soil observations as input of our learning algorithms. We cannot only count on the legacy data stored in the current soil databases as we have done until now. We must be much more active and creative at the local level in using new soil data. As we are closer from the users, it will be certainly easier to convince them to invest locally in more soil analysis and observations. However, the costs of laboratory analysis may remain a strong limitation, even for motivated financer. Multi-source Digital Soil Mapping can be a response to this limitation. I call "multi-source Digital Soil Mapping" a DSM approach that would use spatial samplings of sites that would include both sites with real soil measurements and sites with surrogate soil measurements, such as qualitative soil observations, soil sensing data, soil maps extractions, farmers expert saying, etc., each of them having different levels of uncertainty and degrees of spatial coverages of the study area. Recently, we have set out toward this goal by applying a Random Forest algorithm on various learning datasets including both sites with measured soil properties and sites with soil properties estimated from soil sensing methods. Table 1 summarize the results we obtained.

Table 1: Gains of accuracy of soil property predictions obtained by complementing soil measurements by surrogate soil inputs

RÉFÉRENCES	LOCAL STUDY AREAS	SOIL PRO-PERTY	NUMBER OF MEASURED SITES	SURROGATE SITES			GAINS OF ACCURACY (DMEC)
				Number	source of estimates	Uncertainty (MEC*)	
Zare et al., 2021	Maharlu Lake Watershed, Iran (450 km^2)	Topsoil EC	120	210	EM38	0.63	-0.02
Styc et al., 2021	Bouillargues (16 km^2)	Available Water Capacity	69	2781	Qualitative observa-tions	N.C.	+0.13
Weerasekara et al. (in prep)	Berambadi Watershed, India (80 km^2)	Topsoil clay%	90	185	Vis-NIR lab spectra	0.89	+0.03
Weerasekara et al. (in prep)	Berambadi Watershed, India (80 km^2)	Topsoil clay%	90	3000	Sentinel 2 bands	0.74	+0.09

*MEC is the model efficiency coefficient evaluating the ratio of explained variance by the model.

At first sight, we could deduce from these results that adding a lot of surrogate sites ("Number of surrogate sites" column), even less precisely characterized ("surrogate site uncertainty" column), would be more efficient ("DMEC" column) than adding few well-characterized sites. However, these results are highly site-specific in that they depend from the spatial structure of the local soil variations. Indeed, the importance of adding many sites should be maximal in regions where dense spatial samplings are required to capture soil variations occurring within short distances. However more experiments are required to fully determine the respective impacts of the number and the uncertainty of the surrogate data that are added to the soil measurements.

On the other hand, we may assume that our current learning algorithms that consider equally each individual case are not optimal when using such heteroclite leaning samples. We already observed that in the Maharlu Lake watershed (Zare et al, 2021). Indeed, a regression co-kriging approach that considered the surrogate sites (EM38 measurements) as soft data largely outperformed a classical approach that ignored the differences between soil measurements and estimations from soil sensing (DMEC = 0.09). However, it is often uneasy to fit the linear model of co-regionalization required to apply co-kriging and it is nearly impossible to do so if too many different surrogate data have to be considered. More manageable DSM models dealing with multi-source soil datasets are therefore to be applied in the future. Pioneer works from Wadoux et al (2018) or van Leeuwen et al. (2023) bring the hope to get these models soon.

Local Digital Soil mapping

Participative Digital Soil Mapping

Like soil surveying, Digital Soil Mapping has been up to now conducted by a restricted community of "happy few" having weak, if not no interactions with outsiders. Because local Digital Soil mapping consider territories of modest size, we get the opportunity to involve inhabitants of these territories in the DSM process, making Digital Soil Mapping become a participatory activity. Rossiter et al. (2015) well anticipated the possible contributions of citizens to Digital Soil Mapping. However, to my knowledge, there have not been yet any published paper describing an application of their proposals. Without repeating what David and his colleagues wrote, I just want to briefly focus on two groups of citizens with which we are working now in view to improve our local DSM products, namely farmers and local decision makers.

Figure 1: Participative workshops on local soil typologies: Gopalapura, Southern India (left) and Auterive, Southwestern France (right)

Farmers have a specific knowledge on soils as it was observed since a long time by the ethnopedologists (Barrera Bassols & Zink, 2003). The challenge is to elicit this knowledge for using it as a new input for the local mapping of soils. We got on the road to this goal by co-building local soil typologies with farmers and soil scientists during participative workshops (Lagacherie et al, 2021, Lagacherie et al, 2023) and ask to the farmers to locate the resulting soil types on a map (Lagacherie et al, 2023). More research is required to relate this farmer knowledge to the classical soil descriptions in view of retrieving from this knowledge a valuable input for local digital Soil Mapping approaches.

Local decision makers are people that, like Damien, Georges and Aurélie, will be the daily users of the DSM products. Up to now, these DSM products are stored in geographical databases under a predefined form that cannot be modified. Why not enabling local decision makers to find by themselves the best form of Digital Soil Mapping products for making their decisions? This would mean e. g. i) enabling them to find the most suitable and comfortable uncertainty representation, ii) letting them performing their own Digital Soil Assessment by enabling the update of the soil function satisfaction criteria according to their local knowledge or iii) enabling a collective negotiation about a consensual soil quality map by playing with soil function weights that define the respective importance of each soil function in this specific area. Such new requirements require to deeply transform the way DSM products are diffused to users, moving to less static and more interactive geographical representations. The contributions of specialists in computer geography and geo-visualization will be necessary to achieve that.

Local Digital Soil mapping

The two above-evoked examples of participative DSM bring hope of improvements of our products. Beside and perhaps more importantly, they will certainly ensure a better appropriation of these products by their potential users along the time. Indeed, I often saw during my career potentially useful and relevant soil maps that fell into oblivion simply because too few people were involved in their production and no one could advocate for their usefulness once their initial orderer have left. I believe this has less chance to occur if a large number of people is involved in the process of DSM production. Furthermore, this should raise significantly the connectivity of local people to the soil resource.

Damien, Georges and Aurélie and many others across the world are now knocking at our door. This reveals undeniably an increasing interest for soils that we are all happy to see. Satisfying these demands with acceptable delay and costs is our today Everest. Local Digital Soil Mapping is one possible way to climb this Everest. Let's go.

References

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

Barrera-Bassols, N.; Zinck, J.A., 2003. Ethnopedology: a worldwide view on the soil knowledge of local people. Geoderma 111, 171–195.

Djagba, J.F.; Johnson, J.-M.; Saito, K., 2022. Can soil fertility properties in rice fields in sub-Saharan Africa be predicted by digital soil information? A case study of AfSoilGrids250m. Geoderma Reg. 30, e00563.

Lagacherie, P. ; Laroche, B.; Le Gouée, P. ; Rigou, L., 2021. Une démarche de co-construction de la connaissance pédologique par les agriculteurs et les pédologues. L'exemple de l'aire d'alimentation de captage de la Gimone (82). Agron. Environ. Société 11.

Lagacherie, P.; Dharumarajan, S.; Vasundhara, S. (2003) Exploring the contribution of the farmer soil knowledge to local digital soil mapping. Oral communication + abstract in Soil Mapping for a Sustainable Future, 2nd joint workshop of the IUSS Working groups Digital Soil Mapping and Global Soil map. Orléans, France.

Leeuwen van, C.C.E.; Mulder, V.L.; Batjes, N.H.; Heuvelink, G.B.M. The effect of uncertain calibration and validation soil data on the prediction accuracy of pedotransfer functions. Oral communication + abstract in Soil Mapping for a Sustainable Future, 2nd joint workshop of the IUSS Working groups Digital Soil Mapping and Global Soil map. Orléans, France.

Lemercier, B. ; Lagacherie, P. ; Amelin, J. ; Pichelin, P. ; Richer-de-Forges, A.C. ; Arrouays, D., 2022. Multiscale evaluations of global , national and regional digital soil mapping products in France. Geoderma 425.

Rossiter, D.G.; Liu, J.; Carlisle, S.; Zhu, A.-X., 2015. Can citizen science assist digital soil mapping? Geoderma 259–260, 71–80.

Rossiter, D.G.; Poggio, L.; Beaudette, D.; Libohova, Z., 2022. How well does digital soil mapping represent soil geography? An investigation from the USA. SOIL 8, 559–586.

Styc, Q.; Gontard, F.; Lagacherie, P., 2021. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. Geoderma Reg. 24, e00353.

Voltz, M. ; Arrouays, D. ; Bispo, A. ; Lagacherie, P. ; Laroche, B. ; Lemercier, B. ; Richer-de-Forges, A.C. ; Sauter, J., 2020. Geoderma Regional. Possible futures of soil-mapping in France. Geoderma Reg. 23, e00334.

Wadoux, A.M.J.; Padarian, J.; Minasny, B., 2018. Multi-source data integration for soil mapping using deep learning. SOIL Discuss., https://doi.org/10.5194/soil-2018-39 .EGU.

Weerasekara, M.; Lagacherie, P.; Dharumarajan, S.; Gomez, C. (in prep). Exploring the use of surrogate data as new measurements for training DSM approach: case of lab Vis-NIR spectral data and Sentinel-2 data for soil particles mapping.

Zare, S.; Abtahi, A.; Rashid, S.; Shamsi, F.; Lagacherie, P., 2021. Combining laboratory measurements and proximal soil sensing data in digital soil mapping approaches. Catena 207, 105702.

The Pedometrics Awards Committee 2022-2026

The Pedometrics Commission has three awards to recognize the best papers in the field, up-and-coming talents, and pedometricians that significantly contributed to the advancement of pedometrics as a scientific discipline. These are:

1. The Best Paper in Pedometrics (yearly),
2. The Richard Webster medal (every 4 years), and
3. The Margaret Oliver Award for Early-Career Pedometricians (every 2 years).

The Pedometrics Awards Committee is comprised of previous Richard Webster medal winners and early career scientist. They are chosen as to result in a fair gender balance and geographical spread.

The Medals are awarded by the Award committee and the Best Paper Award is awarded by the Award committee with help from the Extended Award committee. Medals and Awards committee members are appointed for a four -year term, with a possible extension to a second four-year term. New members are proposed by past members and through an open call. Appointment and approval of new members is made at the business meeting of the Pedometrics conference. The Award committee is empowered to replace a member who resign and also exceptionally to appoint new members if the business meeting cannot take place. The Chair of the Award committee is usually the last winner of the Webster Medal.

The current committee members are:

Award committee 2022-2026

- Budiman Minasny, The University of Sydney, Australia (Chairman)
- Gerard Heuvelink, Wageningen University & Research and ISRIC – World Soil Information, the Netherlands
- Murray Lark, University of Nottingham, UK
- Laura Poggio, ISRIC – World Soil Information, the Netherlands
- Lydia Mumbi Chabala, The University of Zambia, Zambia

Extended Award committee 2022-2026

- Tom Orton, The University of Queensland, Australia
- Colby Brungard, New Mexico State University, USA
- Zamir Libohova, USDA-ARS Dale Bumpers Small Farms Research Center, USA
- Ruhollah Taghizadeh-Mehrjardi, University of Tübingen, Tübingen, Germany
- Joulia Meshalkina, Moscow Lomonosov State University and Russian State Agrarian University, Russia
- Alice Milne, Rothamsted Research, UK
- Yuxin Ma, Manaaki Whenua – Landcare Research, New Zealand
- Taciara Zborowski Horst, Federal University of Technology – Parana, Brazil
- Bhabani S. Das, Indian Institute of Technology Kharagpur, India
- Nisha Bao, Northeastern University, China

Best Paper in Pedometrics 2021 – The Winner

Every year the pedometrics commission gives an award to recognize the best paper in the field. The best paper award is awarded by the by the Award committee with help from the Extended Award committee. See the committee members on the website: http://pedometrics.org/awards/.

We had five nominees for Best Paper in Pedometrics in 2021 (in first authors' alphabetical order):

- Bennett, J.M., Roberton, S.D., Ghahramani, A. and McKenzie, D.C., 2021. Operationalising soil security by making soil data useful: Digital soil mapping, assessment and return-on-investment. Soil Security, 4, p.100010. https://doi.org/10.1016/j.soisec.2021.100010

- Dobarco, M.R., McBratney, A., Minasny, B. and Malone, B., 2021. A modelling framework for pedogenon mapping. Geoderma, 393, p.115012. https://doi.org/10.1016/j.geoderma.2021.115012

- Kasraei, B., Heung, B., Saurette, D.D., Schmidt, M.G., Bulmer, C.E. and Bethel, W., 2021. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. Environmental Modelling & Software, 144, p.105139. https://doi.org/10.1016/j.envsoft.2021.105139

- Styc, Q., Gontard, F. and Lagacherie, P., 2021. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. Geoderma Regional, 24, p.e00353. https://doi.org/10.1016/j.geodrs.2020.e00353

- Wadoux, A.M.C. and Brus, D.J., 2021. How to compare sampling designs for mapping?. European Journal of Soil Science, 72(1), pp.35-46. https://doi.org/10.1111/ejss.12962

Based on the vote of the award committee members, the winner of the 2021 Best Pedometrics paper is

Bennett, J.M., Roberton, S.D., Ghahramani, A. and McKenzie, D.C., 2021. Operationalising soil security by making soil data useful: Digital soil mapping, assessment and return-on-investment. Soil Security, 4, p.100010. https://doi.org/10.1016/j.soisec.2021.100010

Abstract

Farmer decision making about soil constraint management in eastern Australia has not changed greatly since the 1970s and major yield gaps exist. Limited capital tended to be used for purchase and clearing of native vegetation on new land, rather than improvement of land already cleared. Incentives did not exist for natural capital outcomes. But farmers and their advisers now have access to approaches that comprehensively measure and digitally map key soil factors throughout the root-zone of crops, map grain yields and compare the results with modelled data to provide an estimate of yield gaps and financial losses. To encourage environmental stewardship, it is likely that farmers will have the opportunity to be rewarded for environmental services through payable credits and/or offsets via commercial environmental markets. Success of such evaluations and markets cannot be assured without clear demonstration of the economic value of on-farm soil data. An important initial step is to consider the cost of comprehensive and accurate soil assessment and management as an up-front capital investment, rather than the tradition of regarding it as an annual cost. This study demonstrates that soil constraints potentially can be regarded from now on as economic opportunities through the use of accurate soil testing and variable rate management. Productivity improvements raise annual returns because of increases in grain yield, and also can boost land values following successful soil assessment and amelioration. Theory associated with the global Soil Security Framework is shown to be compatible with the ROI approach developed here.

Congratulations to the authors. The award document will be given during the next Pedometrics conference.

Budiman Minasny and the Awards committee.

AI for Pedometrics?

Budiman Minasny [1], Alex. McBratney [1], Josè Padarian [1]

[1] Sydney Institute of Agriculture, The University of Sydney, Australia.

"Artificial intelligence is the science of programming computers to mimic the kinds of human behavior that require intelligence, judgment, and experience. There are three major areas of emphasis in artificial intelligence. One is natural language interfaces, which involves programming computers to recognize, understand, and respond to words spoken or written in common language. Knowledge-based systems usually referred to as expert systems, comprise the third major area of artificial intelligence." (The other area mentioned in the article is pattern recognition).

We see more and more soil science applications of machine learning on a daily basis but is this artificial intelligence? Soil science has longed for the application of Artificial Intelligence (AI) in research for a long time. Indeed the above paragraph is what Don Holt wrote in 1987 in "Potentials for Artificial Intelligence and Supercomputers in Soil Science" as part of the book on the Future Developments in Soil Science research. He envisioned AI to revolutionise agriculture and soil science.

Although that article only received 4 citations, Dan Holt's prediction has almost come true, we now have an AI with a natural language interface! But can it do pedometrics?

AI for soil science

Don Holt was Director of the Illinois Agricultural Experiment Station and like many is the '70s and '80s had been developing crop simulation models for production forecasting. In 1985, he wrote in Science on "Computers in production agriculture" (which only has 50 citations). He foresaw the development of an expert system where a developer can successively add sophistication from simple rules. He foresaw the use of supercomputers to design and implement complex finite element and finite difference models of soils and plants. Do we have those?

Another example is by McCracken and Cate (1986) who attempted an expert system with a simple rule-based system for the USDA soil taxonomy. The study found challenges in defining the inputs and rules. However, they thought, unpedometrically, that numerical classification would be problematic (on the sheer amount of data). They suggested applying the principles of cognitive and measurement theory, within the context of expert systems, could improve soil taxonomic systems. Later Dale et al. (1989) from Australia suggested inference procedures and a user interface of expert systems to allow the prediction of soil information.

In 2003, we proposed a soil inference system as an AI to predict all soil properties from any given information about a soil using pedotransfer functions as a knowledge base. Frustratingly, it has taken several iterations and almost 20 years to realise it as a software, and is still a work in progress, but shows promise. The advent of soil chemometric spectroscopy has helped.

Nowadays, with the proliferation of machine learning use in soil prediction, AI is often confused with machine learning (ML), but ML is a subset of AI, but not everything related to AI falls in the ML category (e.g. expert systems). For some of us, ML only becomes AI when the prediction rules are fashioned together into an inference system.

Chatbot

Around Christmas time, there was much chat(ter) about ChatGPT, an AI which can answer any queries. Can we use ChatGPT for pedometrics research? Here are a couple of examples…

AI for Pedometrics?

Write a code…

One of the most remarkable skills of chatGPT is that it can write code. You can tell it to write code for particular equations, e.g. "write a code to calculate the concordance correlation coefficient". Or "write a code for estimating the semi-variogram". (You should try and see the answer, it will take you hours to figure that out).

For a more practical and fashionable application, ask "show an example of calculating Shap (ley) values for a random forest model."

Pedometricians working on coding can find ChatGPT useful to explain how a piece of code works, simplify your codes, debug your codes, translate from one programming language to another, etc.
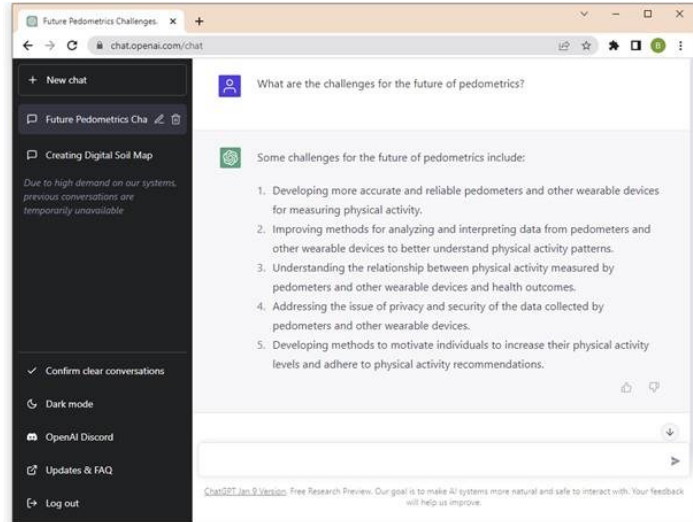
Write …

You can ask ChatGPT to do some mundane task, e.g. "write a letter to the editor of a journal complaining about the rejection of my paper."

What is…

You can ask ChatGPT any question. It is useful for mathematical questions, e.g. "What is the first derivative of the van Genuchten function?"

For more specific soil science questions, don't trust the answers completely. Ask "what are examples of AI application in soil science" , and the examples provided are just regression models (but aren't machine learning models just fancy hyper-parameterised regression models?)

The one we found interesting is 'what is scorpan model in digital soil mapping?'

The SCORPAN model, or Soil COregionalization and Regression PArred with kriging and Neural Networks, is a digital soil mapping method that combines the strengths of both geostatistical and machine learning techniques. The SCORPAN model uses kriging, a geostatistical interpolation technique, to predict soil properties at unsampled locations based on measurements at sampled locations, and then uses a neural network to improve the predictions by modeling the relationships between soil properties and environmental variables.

The SCORPAN model is composed of four main steps:

1. Co-regionalization: This step uses multiple regression kriging to model the relationships between soil properties and environmental variables. It estimates the spatial variation of the soil properties and environmental variables simultaneously, and generates a set of co-kriged maps.

2. Training of neural network: In this step, the co-kriged maps generated in step 1 are used as inputs to a neural network, which is trained to predict soil properties based on environmental variables.

3. Testing of neural network: The neural network is tested using a validation dataset to assess its performance.

11

AI for Pedometrics?

4. Mapping of soil properties: Finally, the neural network is used to predict soil properties at unsampled locations based on environmental variables, and the predictions are combined with the co-kriged maps from step 1 to produce a final map of the soil properties.

FYI, that is an incorrect answer. ChatGPT cannot do your DSM research.

ChatGPT can do simple tasks, it has a broad knowledge base and can write general text, but don't depend on it for your research. Creativity is an important part of research, an AI cannot do that.

Coauthorship

Recently Budi posted on Twitter the answer from a question "write a code on the prediction of soil carbon from NIR spectra". Someone replied that we need to give ChatGPT authorship. Now there are some papers which have listed ChatGPT as a co-author.. Such as in an editorial in Nurse Education and Practice and a preprint.

So we asked ChatGPT, "can you be a co-author of a paper?"

And it answered:

"As an artificial intelligence, I am not able to be a co-author of a paper since authorship is typically reserved for living individuals who have made significant contributions to the research and writing of a paper. However, my capabilities can be used to assist researchers in generating text, analyzing data, or providing insights. But ultimately, it's the researcher who should be credited for the work."

Those authors didn't get consent from ChatGPT as co-author!

While you can ask ChatGPT to write about a topic, we don't think it is a good idea to ask ChatGPT to write a paper for you!

Maybe one day, we can customise ChatGPT to use all soil science papers as its knowledge base and have a real AI for soil science. For now, it is fun tool!

Finally, you can ask ChatGPT to write an abstract of your paper, here's the abstract of this article written by ChatGPT:

The application of artificial intelligence in soil science, or "Pedometrics," has been a topic of interest for many researchers in the field. In 1987, Don Holt wrote about the potential uses of AI and supercomputers in soil science research and predicted that it would revolutionize agriculture. However, the actual implementation of AI in soil science research has been slow due to challenges in defining inputs and rules. Despite this, recent advancements in machine learning have led to an increase in the use of AI in soil prediction. This article explores the potential use of AI, specifically through natural language interfaces and expert systems, in pedometrics research. Additionally, the article highlights the potential use of AI language model, ChatGPT, for tasks such as writing code and performing mundane tasks in pedometrics research.

References

Dale, M.B., McBratney, A.B. and Russell, J.S., 1989. On the role of expert systems and numerical taxonomy in soil classification. Journal of Soil Science, 40(2), pp.223-234.

Holt, D., 1987. Potentials for Artificial Intelligence and Supercomputers in Soil Science. Future Developments in Soil Science Research, pp.459-468.

McBratney, A.B., Minasny, B., Cattle, S.R. and Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma, 109(1-2), pp.41-73.

The International joint workshop of the IUSS Working Groups "Digital Soil Mapping" and "Global Soil Map" Soil Mapping for a Sustainable future, 2023
Orléans, France, 7-9 février 2023

Anne C. Richer-de-Forges[1], Dominique Arrouays[1], Laura Poggio[2], Budiman Minasny[3]

[1] *INRAE, Info&Sols, Orléans, France*
[2] *ISRIC- World Soil Information, Wageningen, The Netherlands*
[3] *The University of Sydney, Everleigh, NSW, Australia*

CONTEXT

The STUDIUM Institute for Advanced Research Studies (Orléans, France) organized from February 7 to 9, 2023, in Orléans an international workshop on Digital Mapping of Soils and their properties at all scales. This conference, entitled "Soil mapping for a Sustainable Future" was co-organized by the INRAE Info&Sols Unit of Orléans, Centre Val de Loire, France.

It was the second international workshop bringing together the "Digital Soil Mapping" (DSM) and "GlobalSoilMap" (GSM) working groups of the Commission 1.5 Pedometrics if the International Union of Soil Sciences (IUSS), after a first conference organized in Santiago (Chile) in 2019. The main scientific coordinators of this conference were Dominique Arrouays, Anne Richer-de-Forges (INRAE) and Laura Poggio (ISRIC, The Netherlands), D. Arrouays and L. Poggio being respectively chairs of the working groups mentioned above.

Looking back at the workshop

The conference brought together around 150 representatives from 30 different countries. The objectives were to assess the advancement of the projects, advocate for the products and their utilization, converse about the technical and scientific aspects, and inspire new collaborators to come on board. A significant challenge is to disseminate information to both soil science experts and user communities, thereby expanding participation to all organizations interested in employing these digital soil data for addressing global challenges across various scales.

SCIENTIFIC CONTENT AND ORGANIZATION

Venue

The STUDIUM Val de Loire d'Orléans organized from February 7 to 9, 2023 in Orléans, France, a world conference on Digital Soil Mapping at all scales and World Mapping of Soil Properties and entitled "Soil Mapping for a Sustainable Future". This conference was held in the auditorium of the Musée des Beaux-Arts (1 Rue Fernand Rabier, Orléans) for the oral presentations and in the premises of the STUDIUM (7 rue Dupanloup, Orléans), for the poster sessions, breaks and lunches).

Scientific Committee

A scientific committee of 12 highly qualified members from 8 organizations was formed to define the structure of the conference and its major themes and to select the presentations.

Chairs

Dr. Ir. Dominique Arrouays, INRAE, Info&Sols, Orléans, France.

Dr. Laura Poggio, ISRIC-World Soil Information, Wageningen, The Netherlands.

Members

Dr. Kabindra Adhikari, United States Department of Agriculture-Agriculture Research Service - USA

Dr. Songchao Chen, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou - China.

Dr. Subramanian Dharumarajan, ICAR-NBSS&LUP, Bengalore, - India.

Prof. Dr. Gerard Heuvelink Wageningen University & ISRIC - World Soil Information - The Netherlands.

Dr. Zamir Libohova, United States Department of Agriculture-Agriculture Research Service - USA.

Prof. Dr. Budiman Minasny, University of Sydney - Eveleigh, NSW, Australia.

Prof. Dr. Vera Leatitia (Titia) Mulder, Wageningen University, Soil Geography and Landscape group - The Netherlands.

Anne Richer-de-Forges, INRAE, Info&Sols, Orléans, France.

Dr. Pierre Roudier, Manaaki Whenua - Landcare Research, Palmerston North - New Zealand

Dr. Alexandre Wadoux, University of Sydney - Eveleigh, NSW, Australia.

The presentations were splitted into several categories:

- Invited keynotes: 20 minutes;
- Long talks: 15 minutes;
- Talks: 10 minutes;
- Short talks: 5 minutes;
- Posters: displayed throughout the conference and benefiting from a dedicated 2-hour session.

Looking back at the workshop

All the sessions were moderated by world-renowned scientists bringing together the main institutions responsible for soil mapping at local, national, continental and global scales.

The conference brought together approximately 150 representatives from 30 countries. The main objectives were to take stock of the progress of the projects, promote the products and their use, discuss the technical and scientific aspects, and encourage new partners to join. A major challenge has been to inform soil science specialists and user communities and thus broaden participation to all entities interested in using these digital soil data in the context of major planetary issues. (climate change, food security, water supply, protection of biodiversity, maintenance of soil resources, human health, etc.) or more local (protection of surface and groundwater, sustainable land development, urbanization and zero land-take objective, etc.).

Integration with the history of the existing situation

The INRAE Info&Sols Unit is the structure that sets up the national information system for the soils of France and the evolution of their qualities, under the patronage of the GIS Sol which brings together the ministries in charge of the environment and agriculture, Ademe, IGN, IRD, OFB and INRAE. The programs developed within the framework of the GIS Soil in metropolitan France and overseas are the subject of international recognition which has materialized in particular by the chair of the WG Global Soil Map of the commission 1.5 of the IUSS.

- The organization of the first "GlobalSoilMap" world conference in Orléans in 2013 and the publication of an associated book.

- The co-organization of the world symposium "Global Soil Security" in 2016 in Paris and the publication of an associated book.

- The co-organization of the second "GlobalSoilMap" world conference in 2017 in Moscow and the publication of an associated book.

- The co-organization of the global symposium "Digital soil mapping and GlobalSoilMap– Scientific advances and the operational use of digital soil mapping to address global environmental challenges" in Santiago de Chile in 2019 and the publication of a special issue of the journal " Geoderma Regional" in 2021.

- Obtaining a funding from the STUDIUM for a research consortium on this topic, bringing together researchers from INRAE (FR), Wageningen Research University (NL), Univ. Sydney (AU), United-States Department of Agriculture (USA), LandCare Research (NZ) and ISRIC-World Soil Information (NL).

Despite a difficult period linked to sanitary conditions, this consortium has produced a considerable body of knowledge and has contributed to the expansion of a very large number of international collaborations which have been summarized in one of the posters presented at this conference.

Network of co-authors of the 100 publications produced by the consortium funded by LE STUDIUM

© Anne Richer-de-Forges, INRAE. Source: GLAD-SOILMAP consortium publications (list available at https://www6.inrae.fr/gladsoilmap-consortium/ Deliverables/Publications. *Figure from the communication:* Arrouays, D., Richer-de-Forges, A.C., Minasny, B., Poggio, L., Libohova Z., Mulder, V.-L., Roudier, P., Martin, M.P., Lagacherie, P., Martelet, G. Nehlig, P., Bourrennane H. 2023. An overview of the Consortium GLAD-SOILMAP supported by the LE STUDIUM Loire Valley Institute for advanced research studies (France).Joining efforts between sub-national, national, continental and global scale digital soil mapping of soils, soil properties and soil functions.Soil mapping for a Sustainable Future Conference, Orleans, France 7-9 Feb. 2023.

Looking back at the workshop

The difficult health period left a residue of operation of this consortium at the STUDIUM (mainly funding for trips and stays which had to be cancelled). LE STUDIUM has therefore proposed to allocate this balance in the form of a contribution to the simultaneous organization of an international symposium on this theme. In order to balance the budget, fundings were also obtained from the French Ministries in charge of Agriculture and of Environment, from INRAE, from the city of Orléans, and from the French Soil Science Society. A funding of 2,400$ was also asked to the IUSS stimulus fund, to help 4 low income countries (LIC) participants, and received a favorable response. Unfortunately, all the LIC people interested could not get their visa or the funding complement from their laboratories.

MAIN OUTCOMES OF THE WORKSHOP

Reaching a large and various audience

Despite some difficulties due to visa problems, the participation of countries was numerous, with in particular the presence of large emerging countries such as India and Brazil. Thirty countries from different continents actively participated in this conference. About a third of the participants were of French origin, which corresponds to the average proportion observed in this type of conference. It should be noted that the participants were also from various backgrounds, mostly scientific, but that the audience also included private companies, representatives of public authorities at different territorial and national levels and NGOs.
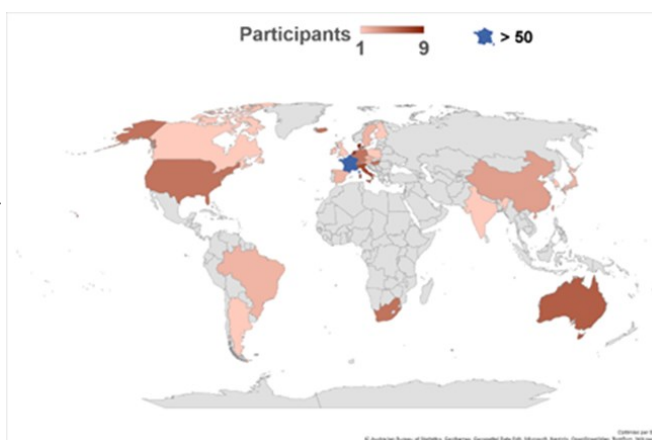
Geographical origin of the participants
© Anne C. Richer-de-Forges (INRAE Info&Sols, France) using the file of registrations at the conference

Group photo of the participants
© Sacha Desbourdes (INRAE Info&Sols, France) - In front of the hotel Dupanloup, Orléans, France

Looking back at the workshop

Take stock of progress

During the first two sessions, progress in geographical and thematic coverage of digital soil mapping worldwide was reviewed. Several countries have emerged as pioneers in this field, including Australia, the United States, France, Hungary, Belgium, Denmark, the Netherlands, and South Korea. Furthermore, there are significant initiatives in progress in large nations like China, India, Brazil, and across continents such as Europe, Africa, and South America. Many other countries have also made substantial strides, such as Italy, Austria, Germany, Canada, South Africa, the Czech Republic, and Argentina. Digital soil maps at varying resolutions have been compared and occasionally utilized together to more effectively leverage soil evolution processes expressed at different scales.

Unite scientific communities

The significant turnout at the conference is indicative of the growing interest in the research themes. However, due to the current international security and health situation and certain strikes, we observed a slight reduction in the expected number of participants, which was around 180. Nevertheless, the discussions were highly productive, relaxed, and engaging, with a notable level of scientific expertise and vision displayed by world leaders in the field. The discussions were focused on substantive issues, without getting lost in purely technical aspects, which is noteworthy for a community where mathematical and statistical components are particularly significant. Additionally, the breaks and poster sessions provided an opportunity to expand the scientific community further and directly communicate with potential users. All participants commended the positive atmosphere of the event, and the quality of its organization.

Awards

Awards were given to the best oral presentation, the best oral presentation by a student, the best poster, the best student poster, and the most innovative idea.

Awards
© Maurine Villiers Le STUDIUM Institute for advanced research studies, Orléans, France

Main advances and methodological issues

The primary methodological advancements and challenges discussed were as follows:

- Novel modelling approaches and their ability to enhance predictive power and process explanation.
- Combination or fusion of different models.
- Emergence of new variables, especially those obtained from high-resolution remote sensing data (spatial, spectral, and temporal).

- Methodologies used for selecting and choosing co-variables.
- Incorporating uncertain data in machine learning models.
- Various strategies for handling data density and spatial structure during learning.
- Different validation strategies to consider.
- Integration of soil process knowledge in spatial modeling, and vice versa.
- Developing robust and low-data-intensive validation techniques.
- Risk of over-adjustment of models.
- Estimation and propagation of uncertainties throughout the processing chain in models and decision support tools.

In general, the presentations demonstrated a shift from merely mapping soil and its properties to mapping ecosystem functions and services. This transition is vital to extend digital soil mapping beyond experts and generate input data for models used by other disciplines, such as hydrology, economics, agronomy, climatology, and more. Additionally, this shift can produce operational tools on various spatial scales and timescales, ranging from managing cultural interventions to extreme climate events, projections, and long-term forecasts related to global changes that affect the planet as a whole.

It is indeed crucial to better explain how to use DSM and to better communicate the consequences of uncertainties to end users involved in the sustainable management of soils, their conservation, the optimization of their services, and the assessment of risks. Another essential discussion focused on information dissemination media, which should not be unique but should be adapted to the needs of the final end-users.

As much as it is possible to produce maps of "pixels" (or more exactly of "voxels") of increasingly fine resolution, these products will not necessarily be those expected by the majority of users and a very local prediction will inevitably be marred by a very great error. This involves reversing our production logic, starting from the request and producing the map that best meets the needs of a user in terms of spatial support, extent of the map, the question asked, and acceptable consequences of uncertainties.

It is also a question of not promising the "moon" but of agreeing to say if it is reasonable to produce such a prediction over a given area or if the state of our data or our knowledge does not yet allow us to do it.

Finally, a better understanding of the various sources of prediction uncertainties appeared essential, whether it concerns the uncertainty linked to the initial measurements on the soil (which are most often considered to be the truth), those on the co-variables (remote sensing data, etc.), those linked to the parameterization of the models, to both spatial and statistical coverage of the variables of interest, or errors inherent in the very structure of the models.

This path should make it possible to make better progress on the priorities to be implemented to improve predictions.

Other practical prospects discussed

- Putting the presentations online subject to the agreement of the authors.
- Posting of a selection of photographs of the event after agreement of the authors of the photographs and the right to the image of the person (if individual or small group).
- Public report in the Information Bulletin of the International Union of Soil Sciences (this text).
- Publication of a special issue of Geoderma open access with reduced fees (call ongoing).
- NBSSS-LUP proposal for a next conference in India, Bangalore, in 2024.
- Redesign and elections for the renewal of the IUSS working groups to be organized before the 2024 conference.

COMMUNICATION-RELATED COMMITMENTS TOWARDS THE PROJECT FUNDERS

All the partners who have supported or proposed to support the project are each described in one full page (after their approval) in the final booklet. Their logos appear on all documents, posters and on the website. They were publicly thanked during the introductory and concluding sessions.

Conference report

Report on the DSM conference in Orléans, France, 2023

By Léa Courteille

From February 7th to 9th in Orléans took place the "Soil Mapping for a Sustainable Future" conference, or – to be more precise – the 2nd joint Workshop of the IUSS Working Groups "Digital Soil Mapping" and "Global Soil Map". As the first joint workshop had taken place in 2019 in Santiago de Chile, almost 4 years ago, everybody was very happy to meet in-person again. As for me, it was the first international conference I attended in my short career.

On the first day, the speakers of sessions 1 and 2 gave us a good overview of the diversity of digital soil mapping products, in terms of scales, mapped variables, trained models, methods used to improve accuracy, gather data and/or deal with the lack of it, collaborations with other stakeholders.

The second day was a busy one! Session 3 dealt with how additional data sources can be used to derive covariates and improve model performances, session 4 with sampling techniques and uncertainty quantification, and session 5 with incorporating pedological knowledge into DSM techniques. On this day many talks made me understand concepts that, as I realised, I had never truly grasped from the literature.

The poster session was held in the morning, and for me it was the opportunity to have many insightful conversations. I had no poster but I put a few maps I generated on display to illustrate what I am currently working on:

aggregating pixels into homogeneous zones on DSM maps so that it is easier for decision-makers to retrieve information and its associated uncertainty.

In the evening, we went on a guided tour around the old city of Orléans, and afterwards we all met at a very nice restaurant, where I don't know which of the delicious food or the photobooth, was the biggest hit.

On the last day, talks were focused on assessing soil functions and services, and the presented research was very much on the operational side. I had the chance to give a short talk during this session, and to get some thought-provoking questions and feedback afterwards.

After the last meal, we had a wrap-up afternoon there were general discussions on DSM (mainly focused on data management, data sharing) and on the future of the working group, and the awards were given. I also must not omit the second musical performance of Dominique Arrouays, definitely one of the highlights of the conference!

On the whole, I really enjoyed this workshop: I have got to know many people that I look forward to seeing again in the future, and learned a lot from both the talks and the conversations I had. In the end, I was lucky that this conference was held at the right moment of my PhD: late enough for me to understand most of the concepts and issues that were discussed, and early enough so that I still have 2 years left to concretize the ideas and suggestions I got.

Thanks to Le Studium and all the organisers for this great event!

DUBIETY

Pedogenesis

The old question

Creeps or jerks?

Recurring spasmodic equilibria

Punctuated stasis

To amnesia

A process supermesh

Cal in the pedometricians

Armed alogorithmically

To unriddle it

Or to reconstruct it

With their guile

Artifice and incertitude

By David van der Linden

inlabru : Convenient fitting of Bayesian digital soil mapping models using INLA-SPDE

Nicolas Saby[1], and Thomas Opitz[2]

4/24/23

[1] INRAE, Info&Sols, Orléans, France
[2] INRAE, BioSP, Avignon, France

Introduction

Pedometricians are nowadays big fans and heavy users of Machine Learning (ML) approaches, with on the top the widely used random forest algorithm, see for example L. Poggio et al. (2021). These algorithms are indeed particularly well adapted to the management of large data sets for mapping soil properties on large geographic areas in a wide range of situations. The techniques are based on classification and regression algorithms, but they do not take account of spatial correlations in residuals (Heuvelink and Webster 2022). This trend towards heavy use of ML tools seems to be accompanied by a diminished use of geostatistical techniques that often require more computer resources but also profound statistical skills to construct and fine-tune models. In many applications, prediction is performed in several steps (*eg* regression or any other machine learning prediction in step 1, followed by spatial kriging of the residuals in step 2), but then an accurate assessment of the prediction uncertainties is difficult since uncertainties from the first step must be propagated through to the second step.

In this paper, we propose to solve these issues by using the fully Bayesian estimation framework based on the integrated nested Laplace approximation (INLA,(Rue, Martino, and Chopin 2009)), combined with the so-called stochastic partial differential equation approach (SPDE, Lindgren, Rue, and Lindström 2011) that provides numerically convenient representations of Gaussian processes over continuous space. Over the last decade, the INLA method has become the most popular tool in spatial statistics for estimating a wide variety of Generalized Additive Mixed Models (i.e., Generalized Additive Models with random effects) in a Bayesian setting. It is a relatively easy-to-use alternative to traditional Markov chain Monte Carlo methods since it provides off-the-shelf implementation of fast and accurate deterministic approximations of posterior inferences for a large class of models. INLA with SPDE is a powerful combination to handle very large spatial datasets. Models are formulated as Bayesian hierarchical models where covariate effects and Gaussian processes can be additively included in a latent process (that is not directly observed), whereas the probability distribution of observations can be of different types (continuous such as Gaussian, skew-Gaussian, Gamma, extreme-value, or discrete such as Poisson, binomial, negative binomial), and the latent Gaussian process is embedded into a key parameter of the probability distribution, such as the mean.

INLA-SPDE was already introduced by Poggio et al. (2016) or Huang et al. (2017) to the pedometrics community. However, wider use of this approach by the community was probably hindered by the complexity of the `INLA` R package. Recently, the `inlabru` R package (Yuan et al. 2017), an add-on package to `INLA` originally developed with a strong focus on point process models for discrete data in ecology, has integrated a range of functions to help implement INLA-SPDE models in a more convenient way through a more ergonomic interface. We propose here to illustrate how this package works by using a simple and classical regression kriging approach as an example.

What's new in R?

Set-up

Load packages

We use here the set of R packages given in the list below.

The latest version of R (eg >4.2) should be installed on your computer for using the `inlabru` package. The classical soil dataset for the Meuse area that we use here is available in the `gstat`package.

```r
library(INLA)
library(inlabru)
library(dplyr)
library(tmap)
library(gstat) # for the meuse data
library(tmap)
library(ggplot2)
```

The `inlabru` method is a convenient wrapper for the `INLA::inla` function and provides multiple enhancements, such as an improved integration of spatial object classes of type `sp` in R, more convenient syntax for defining the structure of the model, convenient functions to perform Bayesian prediction using simulations from the estimated posterior model, and estimation facilities for certain model structures that are not possible with the classical `INLA` package.

Point data and rasters

We use the open data `meuse` from the `gstat` package.

```r
data(meuse)
data(meuse.grid)

str(meuse)

'data.frame':   155 obs. of  14 variables:
 $ x      : num  181072 181025 181165 181298 181307 ...
 $ y      : num  333611 333558 333537 333484 333330 ...
 $ cadmium: num  11.7 8.6 6.5 2.6 2.8 3 3.2 2.8 2.4 1.6 ...
 $ copper : num  85 81 68 81 48 61 31 29 37 24 ...
 $ lead   : num  299 277 199 116 117 137 132 150 133 80 ...
 $ zinc   : num  1022 1141 640 257 269 ...
 $ elev   : num  7.91 6.98 7.8 7.66 7.48 ...
 $ dist   : num  0.00136 0.01222 0.10303 0.19009 0.27709 ...
 $ om     : num  13.6 14 13 8 8.7 7.8 9.2 9.5 10.6 6.3 ...
 $ ffreq  : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
 $ soil   : Factor w/ 3 levels "1","2","3": 1 1 1 2 2 2 2 2 1 1 2 ...
 $ lime   : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
 $ landuse: Factor w/ 15 levels "Aa","Ab","Ag",..: 4 4 4 11 4 11 4 2 2 15 ...
 $ dist.m : num  50 30 150 270 380 470 240 120 240 420 ...

str(meuse.grid)

'data.frame':   3103 obs. of  7 variables:
 $ x      : num  181180 181140 181180 181220 181100 ...
 $ y      : num  333740 333700 333700 333700 333660 ...
 $ part.a: num  1 1 1 1 1 1 1 1 1 1 ...
 $ part.b: num  0 0 0 0 0 0 0 0 0 0 ...
 $ dist  : num  0 0 0.0122 0.0435 0 ...
```

```
$ soil  : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
$ ffreq : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
```

The first action is to create sp objects:

- a SpatialPointsDataFrame corresponding to the regression matrix and,

- the prediction grid, here already provided in the meuse.grid-object along with covariates.

```
coordinates(meuse) <- c('x','y')

coordinates(meuse.grid) <- c("x","y")
gridded(meuse.grid) = TRUE
```

Fully Bayesian DSM approach

The hierarchical DSM model

We construct a hierarchical model for the soil property $z(s)$ at spatial locations s in the setting of prediction in the presence of exhaustively observed ancillary information (i.e., one or several covariates). We will assume the following linkage between model components and observations, where we denote the latent process by $\eta(s)$. In this paper, z will correspond to the organic matter provided in the variable om.

$$\eta(s) \sim \text{Intercept} + \underbrace{\sum_i \beta_i\, z_i(s)}_{\text{Covariates}} + \underbrace{W(s)}_{\text{Spatial Gaussian field}}$$

The spatial field $W(s)$ captures autocorrelation not explained by the covariates. The latent process $\eta(s)$ will then be used in the observation likelihood, which is here chosen, as a Gamma distribution since soil carbon is a non-negative variable and is known to present a relatively heavy-tailed distribution.

We use the | notation to indicate conditioning of the property at the left side of | on the parameters given at the right side of |. This leads to the following hierarchical formulation for the observations,

$$z(s)|(\eta(s),\theta) \sim \Gamma\big(\exp\big(\eta(s_i)\big),\tau\big),$$

where the Gamma distribution is parametrized in a way such that $\exp\big(\eta(s_i)\big)$ is its mean and τ is a precision parameter related to the variance around the mean. The hyperparameter vector is $\theta = (\theta_\eta, \tau)$ with θ_η the hyperparameters controlling the linear predictors η.

Moreover, different observations $z(s)$ are conditionally independent given the latent process $\eta(s)$ and the hyperparameters in θ controlling it. Therefore, the precision parameter τ controls how smoothly the observations are dispersed around the latent log-Gaussian mean $\exp\big(\eta(s)\big)$.

Construction of the mesh for the SPDE model

INLA and inlabru use a space triangulation method to estimate spatial Gaussian effects with a Matérn covariance function. The latent spatial Gaussian random field is computed at the mesh nodes by solving a Stochastic Partial Differential Equation (SPDE), while elsewhere it is computed by linear interpolation of the values at the mesh nodes. The mesh definition usually presents a trade-off between a high resolution to capture variability of the spatial effect at fine spatial scale and a lower number of nodes that usually comes with faster and potentially also numerically more stable calculations. Many applications already

come with a regular grid used to discretize space, such as the `meuse.grid` object here, but often it still makes sense to choose different nodes for the space triangulation used to represent the Gaussian field $W(s)$, especially in cases where the resolution of the grid from the data is too high for being handled directly by INLA. Below, we present how to build a mesh where the construction of the mesh nodes is initialized using the set of coordinates of the calibration sites. This makes sense since can be useful to have a mesh that is relatively denser in areas with many calibration sites (where data provide more information).

First, we create a matrix `xyMesh` with coordinates of the sites. Next, we define the boundaries of the domain used for computing the spatial latent effect with the SPDE approach. Generally, it is a good strategy to compute an internal boundary (delimiting the study area where we want to predict) and an external boundary (providing an extension zone around the study area that is necessary to avoid strong boundary effects from the SPDE) with different resolutions. The purpose of the extension zone is to push the outer boundary away from the study area, and we can set a lower mesh resolution in this extension zone where we do not want to predict the soil property.

The `INLA::inla.mesh.2d` function creates a triangle mesh based on initial point locations, user-specified or automatically calculated boundaries, and parameters controlling the mesh structure, in particular the `cutoff` parameter. This tuning parameter sets the minimum length of edges between two nodes and allows us to keep the number of nodes at most moderately high and to avoid instabilities in computations related to the covariance structure that could arise because of very high Gaussian correlations at nodes that are very close in space. More information is provided here: https://rpubs.com/jafet089/886687

```
cutoffValue = 50 # in meter

xyMesh <- rbind(coordinates(meuse)) # transform into matrix

max.edge = diff(range(xyMesh[,1]))/(3*5)
bound.outer = diff(range(range(xyMesh[,1])))/3

bndint <- inla.nonconvex.hull(meuse, convex=-.05)
bndext <- inla.nonconvex.hull(meuse, convex=-.3)

# Use of inla.mesh.2d
mesh = inla.mesh.2d(loc=xyMesh,
                    boundary = list(int = bndint,
                                    out = bndext),
                    max.edge = c(1,3)*max.edge,
                    cutoff = cutoffValue,
                    crs = meuse@proj4string@projargs)
ggplot() +
  gg(mesh) +
  gg(meuse) +
  coord_equal()
```
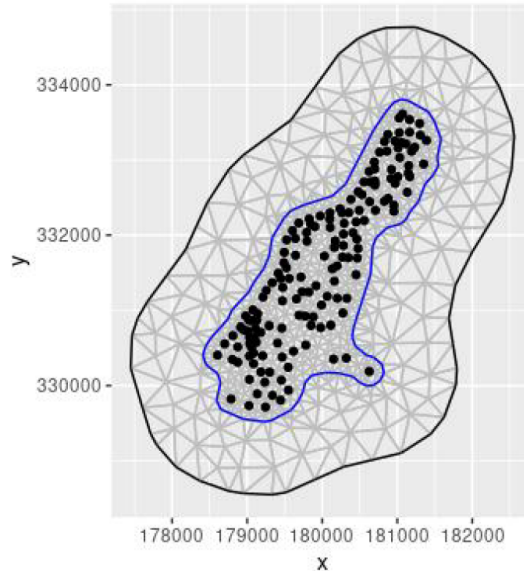
Defining the spatial Gaussian random field $W(s)$

We choose the Matérn covariance function for the Gaussian random field because it can be easily used within `INLA` through the SPDE approach that provides convenient numerical representations for estimation with large numbers of observations and up to several thousand mesh nodes. The Matérn covariance in `INLA` depends on three parameters: - a fractional order parameter *alpha* in the SPDE linked to the smoothness of the solution (which has to be fixed by the user), - a standard deviation parameter *sigma* and, - a spatial correlation parameter known as the *range*.

The α parameter must be fixed by the user, and we here choose $\alpha = 2$ (which is also the default value in the INLA package) corresponding to a Matérn regularity parameter of 1. We specify the other two parameters in our model by selecting a penalized complexity prior using the `INLA::inla.spde2.pcmatern` function. For more details, please read the introduction to spatial models with `INLA` in chapter 7 at <https://becarioprecario.bitbucket.io/inla-gitbook/ch-spatial.html>.

```
matern <-
  INLA::inla.spde2.pcmatern(mesh,
                     alpha = 2,
                     prior.sigma = c(1, 0.5),# P(sigma > 1) = 0.5
                     prior.range = c(10000, 0.9)  # P(range < 10000 m) = 0.9
  )
```

Specifying the hierarchical model

We then specify the model components in the `cmp` object using the convenient `inlabru` approach. In this example implementation, we include the following latent effects: two fixed effects (an intercept, and a linear relationship with the covariate corresponding to the distance to the river), and the Gaussian random field as a random effect.

```
cmp <- om ~
  field(coordinates, model = matern) +
```

```
  Intercept(1) +
  dist(dist, model = 'linear' )
```

Finally, we fit the hierarchical model to the data using the `bru` function of the `inlabru` package. This function requires the model components defined earlier (`cmp`), the dataset (`meuse`), the mesh (`mesh`) where the model will be evaluated, and several options to control the INLA algorithm.

For handling the uncertainty stemming from the prior distributions of the three hyperparameters in θ, we use the `eb` strategy as it is much quicker to compute but a bit less accurate. This `empirical Bayes` approach sets the hyperparameters to their `maximum a posteriori` for some of the calculations performed during the estimation algorithm, that is, it uses a mechanism similar to frequentist inference techniques for handling the hyperparameters.

```
fit <- inlabru::bru(
  components = cmp,
  data = meuse,
  family = "gamma",
  domain = list(coordinates = mesh),
  options = list(
    control.inla = list(int.strategy = "eb"),
    verbose = FALSE
  )
)
```

The summary of the fitted model gives the posterior estimates of fixed effects (intercept and distance to the Meuse river) and of hyperparameters (standard deviation and correlation range of the spatial field, and precision parameter of the Gamma distribution).

```
summary(fit)

inlabru version: 2.7.0
INLA version: 22.12.16
Components:
field: main = spde(coordinates)
Intercept: main = linear(1)
dist: main = linear(dist)
Likelihoods:
  Family: 'gamma'
    Data class: 'SpatialPointsDataFrame'
    Predictor: om ~ .
Time used:
    Pre = 1.39, Running = 1.11, Post = 0.0576, Total = 2.56
Fixed effects:
            mean     sd 0.025quant 0.5quant 0.975quant    mode kld
Intercept  2.349 0.223      1.912    2.349      2.785   2.349   0
dist      -1.311 0.390     -2.076   -1.311     -0.547  -1.311   0

Random effects:
  Name    Model
    field SPDE2 model

Model hyperparameters:
                                               mean       sd 0.025quant
Precision parameter for the Gamma observations  14.227    2.487      9.969
Range for field                               1760.208  831.851    781.670
Stdev for field                                  0.555    0.175      0.318
```

```
                                              0.5quant 0.975quant    mode
Precision parameter for the Gamma observations   14.01     19.738   13.58
Range for field                                1558.31   3944.193 1227.82
Stdev for field                                   0.52      0.996    0.45


Deviance Information Criterion (DIC) ...............: 664.90
Deviance Information Criterion (DIC, saturated) ....: 203.62
Effective number of parameters .....................: 47.08


Watanabe-Akaike information criterion (WAIC) ...: 661.10
Effective number of parameters .................: 35.86


Marginal log-Likelihood:  -366.62
 is computed
Posterior summaries for the linear predictor and the fitted values are computed
(Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```

Spatial predictions

Next, we use the fit to predict the field on a regular lattice, and we therefore generate a set of results using 100 realizations from the posterior distribution of the model. The approach of using posterior simulation for prediction allows us to appropriately represent the uncertainties in the predictions, and we can choose very flexibly for which parameters and properties we would like to provide predictions. In the predictor formula, we use the exp function to take into account the log-link between the mean of the Gamma distribution of the om variable and our linear predictor. This approach provides predictions of the mean surface of the Gamma distribution.

```
pred <- predict(
  fit,
  n.samples = 100,
  meuse.grid,
  ~ exp(field + Intercept + dist) ,
  num.threads = 2
)
```

Internally, the predict function draws samples from the posterior distribution and then combines them to provide the requested predictions. It is also very simple to perform the sampling step directly to obtain the posterior samples using the generate function. For illustration, we here we draw 5 samples and select the first one.

```
samp <- generate(fit,
                 meuse.grid,
                 ~ exp(field + Intercept + dist) ,
                 n.samples = 5
)

str(samp)

 num [1:3103, 1:5] 14.3 16.6 14.2 11.8 17.7 ...

pred$sample <- samp[, 1]
```
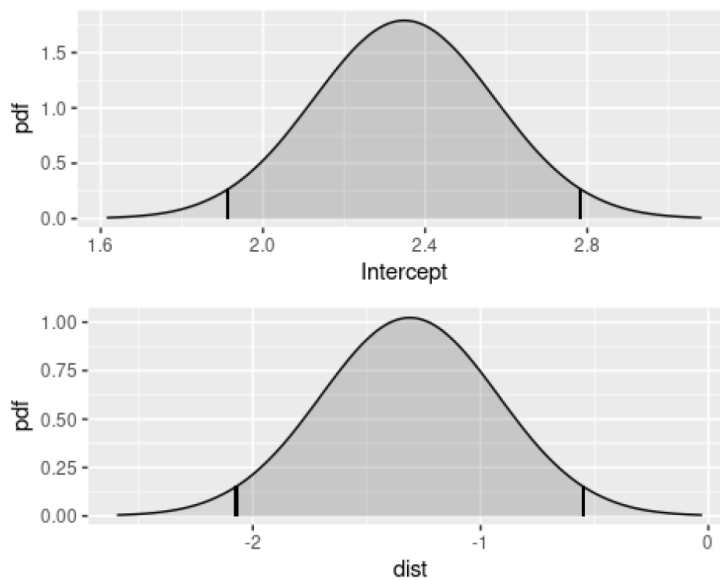
Plotting results

The different effects

We can plot the posterior densities for the latent effect `Intercept` and the distance `dist` to the Meuse river.

To this end we will use the `inlabru::plot()` function,

```
p1 <- plot(fit, "Intercept")
p2 <- plot(fit, "dist")
multiplot(p1, p2)
```


As the credibility interval of the `dist` coefficient does not contain 0, we can conclude that the covariate effect of `dist` is significant.

We can also plot the posterior distribution of the parameters of the Gaussian field, here given as correlation range and variance.

```
spde.range.W0 <- spde.posterior(fit, "field", what = "range")
spde.logvar.W0 <- spde.posterior(fit, "field", what = "variance")
range.plot.W0 <- plot(spde.range.W0)
var.plot.W0 <- plot(spde.logvar.W0)
multiplot(range.plot.W0, var.plot.W0)
```
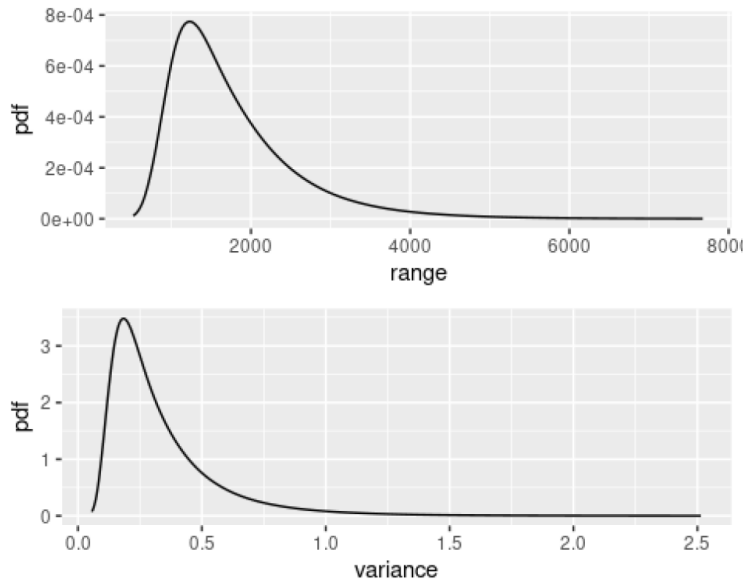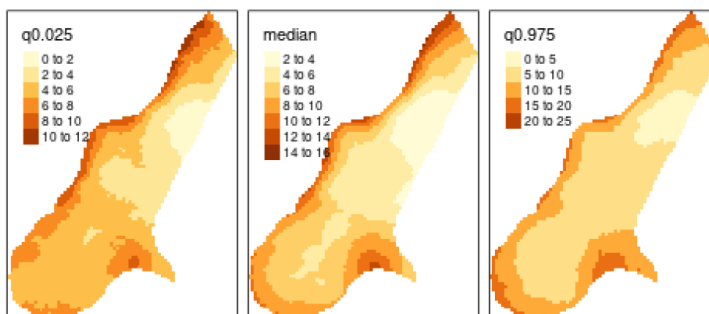
Spatial predictions with uncertainty bounds

We can plot the spatial surfaces corresponding to the pointwise median and to the lower and upper bounds of a 95% credible interval as follows (assuming that the predicted intensity is in the object pred).

```r
tm_shape(pred) +
  tm_raster(
    c("q0.025","median","q0.975")
    )
```
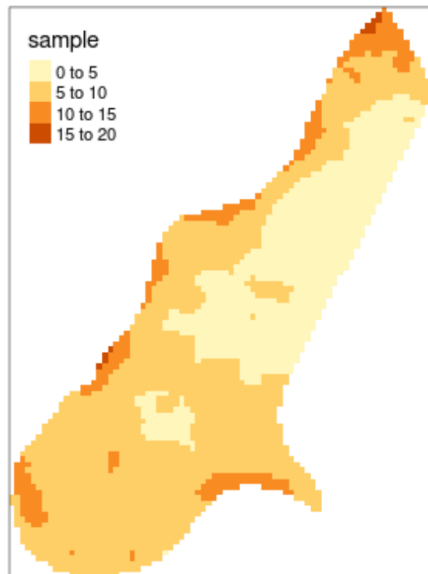
One realization of the posterior distribution

The first sample that we have drawn from the posterior distribution can be mapped as follows.

```
tm_shape(pred) + tm_raster("sample")
```


### The maps of the random and the fixed effects

Next, we plot the 2 effects of the model:

- the spatial Gaussian random field $W(s)$,

- the combination of the two fixed effects.

```
pred <- predict(
  fit,
  n.samples = 100,
  meuse.grid,
  ~ field  ,
  num.threads = 2
)

fixed <- predict(
  fit,
  n.samples = 100,
  meuse.grid,
  ~  exp(Intercept + dist)  ,
  num.threads = 2
)

pred$FixedEffect <- fixed$median
pred$W <- pred$median
```

```
tm_shape(pred) +
  tm_raster(c("W","FixedEffect"))
```
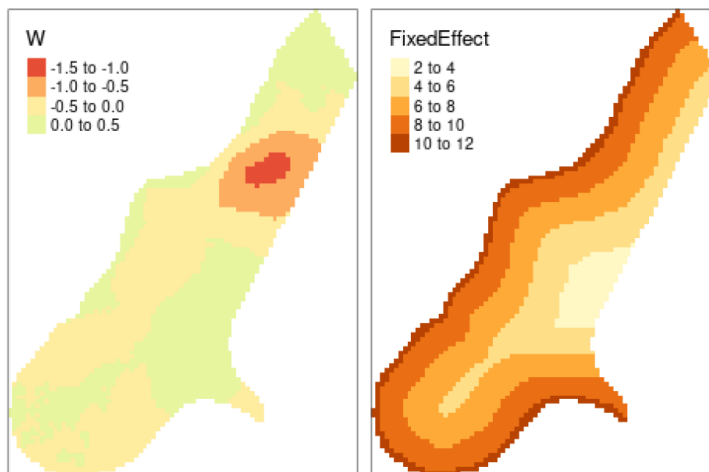

Final remarks

Heuvelink and Webster (2022) listed a set of challenges for pedometricians and spatial statisticians to strengthen the role of spatial statistics and to fully exploit its modern cutting-edge tools. While not being able to solve all of them, we are convinced that fully Bayesian modelling using `INLA` with its numerically highly efficient implementation can provide satisfactory answers to some of these challenges, in particular regarding improved uncertainty quantification, the change of the support, and the incorporation of attribute and positional measurement uncertainty.

The goal of inlabru is to further facilitate spatial modeling using integrated nested Laplace approximation via the R-INLA package. The recent developments made available through `inlabru` allow for an even more convenient construction of Bayesian spatial models of soil properties along with precise uncertainty assessments through the INLA-SPDE approach. Various types of model components can be specified based on various predictors used as inputs, and internally these components are represented through a set of latent Gaussian variables. The predictors are specified via general R expressions. Most of the technical details of the implementation are hidden and handled internally by `inlabru`.

The user can choose a likelihood family, such as `gaussian`, `poisson` or `binomial`, from a long list of possible choices. The default family is `gaussian`. A list of possible alternatives can be seen by typing `names(inla.models()$likelihood)`. Therefore, it is possible to fit a wide range of models allowing to tackle a great diversity of data types and problems in soil science. Here we used the `gamma` family to impose nonnegativity and cope with heavy tails.

In their studies, Poggio et al. (2016) and Huang et al. (2017) reported that INLA-SPDE became quite slow when estimating the posterior marginal distributions of the environmental variables when datasets were large. When the number of observations is huge, it is important hat one can improve the performance of the high-dimensional matrix computations conducted in INLA by using the PARDISO solver library. It is

already fully included in the standard INLA installation but has to be activated through a licence key. To activate it (note that it is free for non commercial uses), go to https://www.pardiso-project.org/r-inla/#license to obtain the license, which will take you at most several minutes. Also, you can type inla.pardiso() at the R command line for viewing the (very simple) instructions on how to enable the PARDISO sparse library. Moreover, additional methodological developments have become available in the latest INLA versions for handling especially data-rich models to achieve even faster inference, improved numerical stability and scalability (Van Niekerk et al. 2023).

Code availability

The code is also available on github : https://github.com/nsaby/pedometron042023
More codes are available here: https://inlabru-org.github.io/inlabru/articles/web/random_fields_2d.html

References

Heuvelink, Gerard B. M., and Richard Webster. 2022. "Spatial Statistics and Soil Mapping: A Blossoming Partnership Under Pressure." Spatial Statistics 50: 100639. https://doi.org/https://doi.org/10.1016/j.spasta.2022.100639.

Huang, Malone, J. 2017. "Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping." Science of The Total Environment 609: 621--632.

Lindgren, Finn, Håvard Rue, and Johan Lindström. 2011. "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (4): 423–98.

Poggio, Laura, Alessandro Gimona, Luigi Spezia, and Mark J Brewer. 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.

Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty." SOIL 7 (1): 217–40. https://doi.org/10.5194/soil-7-217-2021.

Rue, Håvard, Sara Martino, and Nicolas Chopin. 2009. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." Journal of the Royal Statistical Society: Series b (Statistical Methodology) 71 (2): 319–92.

Van Niekerk, Janet, Elias Krainski, Denis Rustand, and Håvard Rue. 2023. "A New Avenue for Bayesian Inference with INLA." Computational Statistics & Data Analysis 181: 107692. https://doi.org/https://doi.org/10.1016/j.csda.2023.107692.

Yuan, Y., F. E. Bachl, F. Lindgren, D. L. Brochers, J. B. Illian, S. T. Buckland, H. Rue, and T. Gerrodette. 2017. "Point Process Models for Spatio-Temporal Distance Sampling Data from a Large-Scale Survey of Blue Whales." https://arxiv.org/abs/1604.06013.