# Pedometrics 2024
# Las Cruces


# PROGRAM
# AND
# ABSTRACTS

# Welcome to Pedometrics2024!

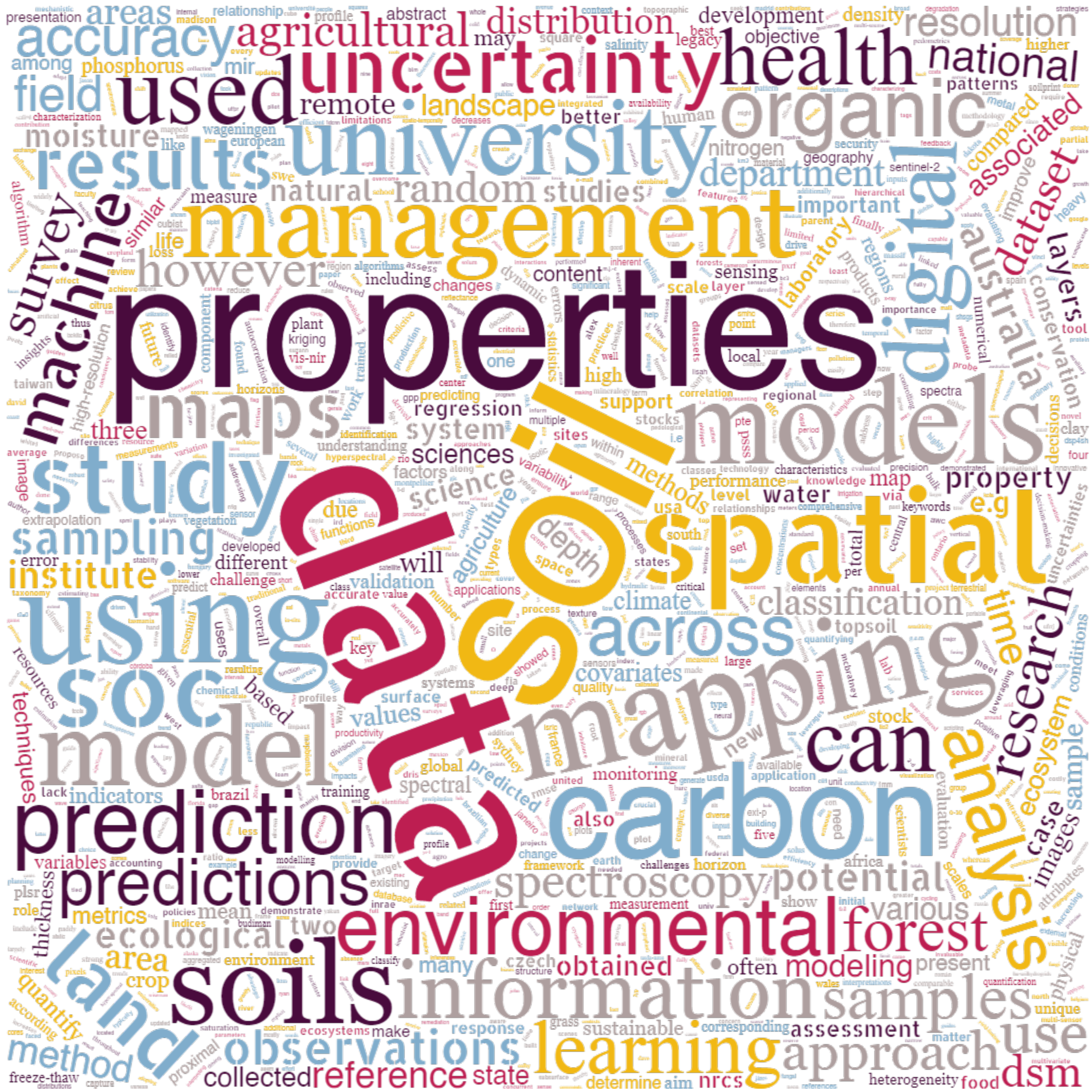
We are excited to host the bi-annual pedometrics meeting and we have designed the conference in a way that we think you will like. We have centered the conference around the "10 challenges for the future of Pedometrics" (Wadoux et al., 2021). These challenges capture broad trends in pedometrics research and we hope that this framework will encourage discussion and collaboration. We have dedicated one session to each challenge and have included ample time to discuss progress on each challenge. Additionally, we have made room for two field trips to visit and discuss the unique soils of arid lands and we have programed multiple engaging social events!

The conference will be held in the Corbett Student Union on the Campus of New Mexico State University, International Mall, Las Cruces, NM 88003

The Pedometrics 2024 Organizing Committee

Dr. Colby Brungard
Dr. Alexandre Wadoux
Dr. Shawn Salley

Alexandre M.J.-C. Wadoux, Gerard B.M. Heuvelink, R. Murray Lark, Philippe Lagacherie, Johan Bouma, Vera L. Mulder, Zamir Libohova, Lin Yang, Alex B. McBratney. Ten challenges for the future of pedometrics, Geoderma, Volume 401, 2021. https://doi.org/10.1016/j.geoderma.2021.115155.


properties
data
soil
spatial
mapping
models
uncertainty
management
university
health
used
organic
digital
carbon
prediction
predictions
soils
environmental
information
learning
approach
using
model
soc
study
results
accuracy
agricultural
distribution
national
resolution
field
machine
maps
analysis
research
forest
spectroscopy
potential
samples
observations
reference
land
method
australia
dataset
associated
landscape
department
random
studies
however
survey
sampling
institute
across
classification
time
can
use
dsm
ecological
area
areas
development
remote
natural
nitrogen
moisture
phosphorus
regression
system
sciences
science
depth
methods
property
water
climate
covariates
topsoil
mean
two
many
obtained
often
modeling
present
assessment
state
collected
freeze-thaw
heterogeneity
food

**Organizing Committee**

Colby W. Brungard, PhD, Associate Professor, New Mexico State University, USA

Shawn W. Salley, PhD, USDA-NRCS-National Soil Survey Center, USA

Alexandre Wadoux, PhD, Habil., Marie Skłodowska-Curie Fellow*,* National Research Institute for Agriculture, Food and Environment (INRAE), Montpellier, France

**Scientific Committee**

Alessandro Samuel-Rosa, Doctor, Federal University of Technology - Paraná

Alexandre Wadoux, INRAE, Montpelier, France

Benjamin Marchant, Statistics Lead, British Geological Survey

Brendan Malone, Senior Research Scientist, CSIRO

Bryant Scharenbroch, Associate Professor, University Wisconsin Stevens Point

Budiman Minasny, Professor, University of Sydney

David G. Rossiter, Guest Researcher, Wageningen University and ISRIC – World Soil Information

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

Gerard B.M. Heuvelink, Special professor in pedometrics and digital soil mapping, Wageningen University and ISRIC – World Soil Information

James Thompson, Professor, West Virginia University

Jason Ackerson, Research Soil Scientist, Soil Health Instutite

Jessica Philippe, Soil Scientist, USDA-NRCS, Soil & Plant Science Division

John Triantafilis, Portfolio Leader Managing Land & Water, Manaaki Whenua – Landcare Research

Jonathan Maynard, Research Soil Scientist, USDA-NRCS, Soil & Plant Science Division

Jonathan Sanderman, Senior Scientist and Carbon Program Director, Woodwell Climate Research Center

Kabindra Adhikari, Soil Scientist, USDA-Agricultural Research Service

Laura Poggio, Senior digital soil mapping and remote sensing expert, ISRIC - World Soil Information

László Pásztor, Director of Institute for Soil Sciences, HUN-REN Centre for Agricultural and Environmental Research

Lin Yang, Professor, Nanjing University

Luboš Borůvka, Professor, Czech University of Life Sciences

Marcos Angelini, International Consultant (Soil Information), Global Soil Partnership, FAO

Mario Guevara, Associate Professor, Centro de Geociencias - Universidad Nacional Autónoma de México

Nicolas Saby, INRAE - Info&Sols Unit, Orléans. France

Pierre Roudier, Senior Researcher, Manaaki Whenua – Landcare Research

Philippe Lagacherie, INRAE, Montpelier, France

Ruhollah Taghizadeh-Mehrjardi, Postdoc Researcher, University of Tübingen

Sabine Grunwald, Professor, University of Florida

Shawn Salley, Soil Scientist / Soil Geographer, USDA-NRCS, Soil & Plant Science Division

Stephen Roecker, GIS Specialist, USDA-NRCS, Soil & Plant Science Division

Stewart Wilson, Associate Professor, Cal Poly University

Taciara Zborowski Horst, Professor, Federal University of Technology – Paraná

Thomas Bishop, Professor, University of Sydney

Tiffany M. Allen, National Resource Soil Scientist for Federal Lands, USDA-NRCS, Soil & Plant Science Division

Travis Nauman, Research Soil Scientist, USDA-NRCS, Soil & Plant Science Division

Umakant Mishra, Principal Scientist, Sandia National Laboratories

Yakun Zhang, Research Scientist, University of Wisconsin-Madison

Zamir Libohova, Research Soil Scientist, USDA-Agricultural Research Service

## The 10 Challenges

| Challenge | Central Question | Keywords |
|---|---|---|
| How can we better understand soil formation? | | |
| 1 | Can we produce quantitative models of the complex short and long-term processes of soil formation which are predictive of the spatio-temporal variation of soil properties? | Soil change; Forecast; Quantitative models of pedogenetic processes; Time series; Dynamic mechanistic models; Soil-landscape evolution models; Quantifying soil genesis |
| 2 | Can we develop a quantitative and numerical global soil classification that unifies the existing systems and enables transfer between them? | Numerical soil classification; Soil taxonomy; Local and regional applications; Communication between classification systems; Similarities between soil profiles; Translation between systems; Near real-time soil classification |
| 3 | In what ways can we use data-driven models to learn about pedological processes? | Interpretation of complex models; Multi-scale drivers of soil variation; Functional relationships between covariates and soil data; Hypothesis discovery; Sensitivity analysis |
| How can we improve methods to obtain relevant soil data? | | |
| 4 | Can we measure soil properties more efficiently? | Soil sensing; Pedotransfer functions; Translation of qualitative soil information; Participatory approaches and citizen science; Sampling design; Measurement error; Multi-source data integration |
| 5 | Can we develop workable techniques to derive predictions of soil characteristics at scales appropriate for modelling and decision making, by up- and downscaling observations in 3D space and time? | Upscaling and downscaling; Sampling support; Change of support; Temporal scale issues in modelling change; Validation for change of support |
| 6 | Can we incorporate mechanistic pedological knowledge in digital soil mapping? | Pedological knowledge; Extrapolation; Qualitative soil information; Mechanistic modelling; State-space modelling; Uncertainty in mechanistic knowledge |
| How can we improve our ability to address demands by soil users? | | |
| 7 | How to recognize, quantify and map soil functionality? | Soil function and services; Citizen-observation of soil functions; Land evaluation; Multivariate mapping; Bio-physical models; Co-building of functions with end-users |
| 8 | Can we find ways to connect pedodiversity to soil biodiversity, and translate the connections to relevant soil services and soil management practices? | Pedodiversity; Pattern of soil biodiversity; Scaling issues in pedodiversity; Taxonomic distance; Hyper-variate data of soil biodiversity; Sensing for microscale biodiversity |
| 9 | Can we find ways to express the uncertainty of predictions of soil properties or class allocations which are meaningful to the users of those predictions? | Uncertainty quantification; Value of information; Risk assessment; Uncertainty and decision-making process; Communication of uncertainty; Decision theory and support scale |
| 10 | How to quantify soil contributions to ecosystem services with a framework enabling both local and regional soil management? | Ecosystem services; Local and regional soil management; Empirical land evaluation schemes; Soil health and security quantification; Soil contributions to realizing the SDG |

**Generalized Weekly Agenda**

Please refer to daily agenda for exact times

<table>
<tr><th>Time</th><th>Sunday</th><th>Monday</th><th>Tuesday</th><th>Wednesday</th><th>Thursday</th><th>Friday</th></tr>
<tr><td></td><td>4-Feb</td><td>5-Feb</td><td>6-Feb</td><td>7-Feb</td><td>8-Feb</td><td>9-Feb</td></tr>
<tr><td>7:30</td><td rowspan="2"></td><td></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="4"></td><td rowspan="4"></td></tr>
<tr><td>8:00</td><td rowspan="2">On-site registration and check in</td></tr>
<tr><td>8:30</td><td rowspan="8">Morning Workshops</td></tr>
<tr><td>9:00</td><td>Conference welcome</td><td rowspan="3">Challenge 4: Spectroscopy</td><td rowspan="3">Challenge 5</td></tr>
<tr><td>9:30</td><td rowspan="3">Challenge 1</td><td rowspan="4">Challenge 9</td><td rowspan="3">Challenge 10</td></tr>
<tr><td>10:00</td></tr>
<tr><td>10:30</td><td rowspan="3">Lunch</td><td rowspan="4">Challenge 7</td></tr>
<tr><td>11:00</td><td rowspan="2">Challenge 2</td><td rowspan="2">Reflection and future work</td></tr>
<tr><td>11:30</td><td rowspan="11">Field trip #2: White Sands National Park</td></tr>
<tr><td>12:00</td><td rowspan="2">Lunch</td><td rowspan="4">Challenge 4: Proximal Sensing</td><td>Conference Adjourns</td></tr>
<tr><td>12:30</td><td rowspan="2">Lunch on your own</td><td rowspan="9">Field Trip #1: Soils and Landscapes of Desert Basins and River Valleys</td><td rowspan="20"></td></tr>
<tr><td>13:00</td><td rowspan="3">Challenge 3</td></tr>
<tr><td>13:30</td><td rowspan="8">Afternoon Workshops</td></tr>
<tr><td>14:00</td><td rowspan="4">Challenge 4: Digital Soil Mapping</td></tr>
<tr><td>14:30</td><td rowspan="2">Pedometrics in Govn't, Scientific, and Commercial Organizations</td></tr>
<tr><td>15:00</td></tr>
<tr><td>15:30</td><td rowspan="15"></td></tr>
<tr><td>16:00</td><td rowspan="2"></td></tr>
<tr><td>16:30</td></tr>
<tr><td>17:00</td><td rowspan="4">Friendly Game of Soccer and/or Flag Football</td><td rowspan="11"></td><td rowspan="10">Conference Dinner, Awards, and Dancing</td></tr>
<tr><td>17:30</td><td></td></tr>
<tr><td>18:00</td><td rowspan="4">Welcome Reception</td></tr>
<tr><td>18:30</td></tr>
<tr><td>19:00</td><td rowspan="7"></td></tr>
<tr><td>19:30</td></tr>
<tr><td>20:00</td><td rowspan="5"></td></tr>
<tr><td>20:30</td></tr>
<tr><td>21:00</td></tr>
<tr><td>21:30</td></tr>
<tr><td>22:00</td><td></td></tr>
</table>


Figure 1. NMSU Campus Map

# Program

## Sunday 4th February

8:30-12:30 Containers for reproducible Digital Soil Mapping at different scales
Skeen Hall (N120)

8:30-12:30 Algorithms for Quantitative Pedology
-cancelled-

12:30-13:30 Lunch on your own

13:30-17:30 Assessment of spatial patterns of soil properties predictions
Skeen Hall (N120)

13:30-17:30 Deep learning for soil spectroscopy
Skeen Hall (N128)

18:00-20:00 Welcome Reception
Pete's Patio (Corbett Student Union)

## Monday 5th February

8:00-9:00 On-site registration and check in

9:00-9:10 Conference Welcome
Alexandre M.J-C. Wadoux - Pedometrics commission chair and organizing committee

9:10-9:20 The 10 Challenges
Shawn Salley – Pedometrics organizing committee

9:20-9:30 Conference Logistics
Colby Brungard – Pedometrics organizing committee

**Challenge 1. Can we produce quantitative models of the complex short and long-term processes of soil formation in the landscape which are predictive of the spatio-temporal variation of soil properties?**

**Moderator: Shawn Salley**

9:30-9:45 Keynote Challenge 1
Tom Vanwalleghem

9:45-9:55 1.1 Continental Monitoring Soil Property Changes Under Human Pressure Using Pedogenon Mapping
Quentin Styc

9:55-10:05 1.3 Century-Long Quantification of Soil Loss in Eastern South Dakota Agricultural Field
Eli Halverson

10:05-10:15 1.4 Modelling the effect of topographical position and precipitation on soil profile variation with Soilgen
Tom Vanwalleghem

10:15-10:30 Challenge 1 Discussion

10:30-10:50 Break

### Challenge 2. Can we develop a quantitative and numerical global soil classification that unifies the existing systems and enables transfer between them?

### Moderator: Shawn Salley

10:50-11:05 Keynote Challenge 2
Dylan Beaudette (Remote)

11:05-11:15 2.1 A global numerical classification of the soil surface layer
Alexandre M.J-C. Wadoux

11:15-11:25 2.3 Similarities among soil profiles in representative soil-landscapes plots and its implications for soil types discrimination-a case study in the three river's sources area in Qinghai Province, China.
Xia Zhao

11:25-11:30 2.4 What is isotic anyway? A Soil Taxonomy mineralogy class revisited
Ryan Hodges

11:30-11:45 Challenge 2 Discussion

11:45-13:00 Lunch

### Challenge 3. In what ways can we use data-driven models to learn about pedological processes?

### Moderator: Kabindra Adhikari

13:00-13:15 Keynote Challenge 3
Gerard Heuvelink

13:15-13:25 3.1 Assessing natural and human drivers on soil thickness variation using generalized additive models
Yakun Zhang

13:25-13:35 3.3 The determinants and regulation of surface soil bacterial and fungal biogeography in Australia
Alexandre Wadoux on behalf of Budiman Minasny

13:35-13:45 3.4 Biplots for understanding machine learning predictions in digital soil mapping
Gerard Heuvelink

13:45-14:00 Challenge 3 Discussion

14:00-14:20 Break

**Pedometrics in governmental, scientific, and commercial organizations**

**Moderator: Kabindra Adhikari**

14:20-14:30 P2 Spatio-temporal soil information based on open science and multidisciplinary collaboration
Taciara Zborowski Horst

14:30-14:40 P4 Leveraging Legacy Data: The Evolution of Mid-Infrared Spectroscopy at the NRCS Soil and Plant Science Division
Jonathan Maynard

14:40-14:50 P5 Enhancement of Soil Data in the U.S. Forest Service Forest Inventory and Analysis Program
John D. Shaw

14:50-15:00 P6 Scaling Carbon Stock Measurement for Carbon Markets
Sarah Coffman

15:00-15:10 P7 Commercial soil carbon accounting: challenges and opportunities for practicing pedometricians
Jason P. Ackerson

15:10-15:25 Discussion

15:25 Adjourn

# Tuesday 6th February

**Challenge 4. Can we measure soil properties more efficiently?**

9:00-9:15 Keynote Challenge 4
Sabine Grunwald

**Challenge 4. Soil Spectroscopy**

**Moderator: Alessandro Samuel Rosa**

9:15-9:25 4.1 Development of soil spectroscopy prediction models for the Western Highveld region, South Africa: Why we need local data.
Anru-Louis Kock

9:25-9:35 4.3 How can we be more assertive about soil spectroscopy predictions? The Open Soil Spectral Library study case
José Lucas Safanelli

9:35-9:45 4.4 Preserving Soil Data Privacy with SoilPrint: A Unique Soil Identification System for Soil Data Sharing
Tegbaru B. Gobezie

| | |
|---|---|
| 9:45-9:55 | 4.5 Using Vis-NIR, MIR, and pXRF spectra for predicting soil physical and chemical properties - A comprehensive review<br>Gafur Gozukara |
| 9:55-10:00 | 4.6 Spectral signature of soil horizons and soil orders in Wisconsin<br>Malithi Vidushika Weerasekara |
| 10:00-10:05 | 4.8 Mesoscale Soil Spatial Heterogeneity Charachterization Using Laser-Induced Breakdown Spectroscopy<br>Changwen Du |
| 10:05-10:10 | 4.10 Mapping soil particle fractions by training Digital Soil Mapping models with surrogate measurements obtained from laboratory and satellite Vis-NIR spectral data<br>Malithi Vidushika Weerasekara |
| 10:10-10:15 | 4.11 In-Situ Soil Spectroscopy Application for Extractable Phosphorus Prediction for Precision Agriculture Purposes<br>Katsutoshi Mizuta |
| 10:15-10:20 | 4.12 An objective test of the Open Soil Spectral Library service<br>Kanchan Grover |
| 10:20-10:40 | Soil Spectroscopy Discussion |
| 10:40-12:00 | Lunch |

**Challenge 4. Proximal Soil Sensing**

**Moderator: Taciara Zborowski Horst**

| | |
|---|---|
| 12:00-12:10 | 4.13 Quantitative soil profile observations<br>Alfred Hartemink |
| 12:10-12:20 | 4.14 Multi-sensor soil probe and machine learning modeling for predicting soil properties to revolutionize sustainable agriculture<br>Sabine Grunwald |
| 12:20-12:30 | 4.15 Spectral inference at the edge<br>José Padarian |
| 12:30-12:40 | 4.16 Going Deep: An assessment of artificial intelligence and deep learning techniques for image processing of soil surface and subsurface horizons<br>Perseveranca Mungofa |
| 12:40-12:45 | 4.17 Testing different combinations of proximal soil sensors for high-resolution mapping of key soil fertility properties<br>Jonas Schmidinger |
| 12:45-12:50 | 4.18 Evaluation of a novel, commercial, VisNIR probe for in-situ measurement of soil carbon stocks<br>Jason P Ackerson |

12:50-12:55 4.19 Predicting changes in soil nitrogen and phosphorus using nitrogen/phosphorus measurement sensors and machine learning
Jae E. Yang

12:55-13:00 4.20 Portable X-ray Fluorescence Spectrometry for Sensing Salinity and Sodicity in Glacial Northern Great Plains Soils
Adam Devlin

13:00-13:05 4.21 Effect of soil autocorrelational properties on regression model choice for mapping soil organic carbon in hyperspectral images
Shayan Kabiri

13:05-13:10 4.22 Application of computer vision semantic image segmentation and classification algorithms for processing of digital microscopic soil images acquired by a digital soil core sensor
Perseveranca Mungofa

13:10-13:15 4.23 Measurement of Soil Carbon Stocks In-Situ with Dual Wave Sensors
Kristopher Osterloh

13:15-13:20 4.24 Proposals for optimization in mapping electrical conductivity in sparse data through data fusion in irrigation zones: An application of spatial regression models
Hugo Rodrigues

13:20-13:40 Proximal Soil Sensing Discussion

13:40-14:10 Break

**Challenge 4. Digital Soil Mapping**

**Moderator: Richard Heck**

14:10-14:20 4.25 The benefits of using a reference sampling for mitigating the impact of legacy soil data errors on Digital Soil Mapping outputs.
Philippe Lagacherie

14:20-14:30 4.26 Seeking Validity in Soil Data
Stephen Roecker

14:30-14:40 4.27 Spatial pattern evaluation in comparing digital soil maps obtained with different methods: an important addition to pointwise metrics
Giulio Genova

14:40-14:50 4.28 Towards POLARIS v2: Improving Soil Properties Mapping Over the CONUS Using a New Hierarchical Geospatial Framework
Chengcheng (Emma) Xu

14:50-15:00 4.29 A metadata-focused harmonization workflow to generate high quality datasets for digital soil mapping and modeling: the Alaska Soil Data Bank project
Nicolas A Jelinski

| | |
|---|---|
| 15:00-15:10 | 4.30 3-D Mapping of Soil Moisture Holding Capacity with Soil Depth Functions and Machine Learning Algorithms in a Tropical Sub-Catchment in Tanzania<br>Jacob Kaingo |
| 15:10-15:15 | 4.31 Exploring extrapolation effects of random forest digital soil mapping: a case study in African countries<br>Gerard Heuvelink on behalf of Fatemeh Hateffard |
| 15:15-15:20 | 4.32 National scale mapping of soil organic carbon stocks in Taiwan<br>Chien-Hui Syu |
| 15:20-15:25 | 4.33 Digital mapping of Australian soil carbon stocks from inorganic carbon<br>Wartini Ng |
| 15:25-15:30 | 4.34 Evaluating the Performance of a Topsoil Organic Carbon Monitoring System at Continental Scale: Regional Validation in Wallonia, Belgium<br>Marmar Sabetizadeh |
| 15:30-15:35 | 4.35 Machine learning models do not provide higher accuracy models compared to ordinary kriging under high density soil observations<br>Chien-Hui Syu |
| 15:35-15:40 | 4.37 Digital Mapping Of Al, Fe2O3, Nb, Tio2 And W In Mineralized Laterites In The Brazilian Amazon<br>Niriele Bruno Rodrigues |
| 15:40-15:45 | 4.38 How can Google Earth Engine and Vis-NIR aid in the challenge of mapping alluvial soils in Tribal Nations<br>Marcelo Mancini |
| 15:45-15:50 | 4.39 Distribution of heavy metals in the soils of conterminous USA and implications for food and environmental safety<br>Kabindra Adhikari |
| 15:50-16:10 | Digital Soil Mapping Discussion |
| 16:10 | Adjourn |
| 17:00 – 19:00 | Friendly games of soccer and/or flag football<br>James B. Delamater Activity Center, 1600 Stewart St |

## Wednesday 7th February

**Challenge 5. Can we develop workable techniques to derive predictions of soil characteristics at scales appropriate for modelling and decision making, by up- and downscaling observations in 3D space and time?**

**Moderator: Kristopher Osterloh**

9:00-9:10 5.1 Gaussian process: A comparison with depth-harmonised approach - a case study of mapping soil constraints
Jie Wang

9:10-9:20 5.2 Modelling soil organic carbon stock in space and time at multiple scales: Case study from Hungary
Gábor Szatmári

9:20-9:30 5.3 Dealing with missingness, truncation, and censoring in multi-source data to map soil organic carbon stocks
Alessandro Samuel-Rosa

9:30-9:40 5.4 Leveraging Remote Sensing, Soil Properties, And Ai Technologies For Nowcasting/Forecasting Soil Moisture In 3D Space And Time
Sabine Grunwald

9:40-9:50 5.5 Fine-Resolution Near-Real-Time Soil Moisture Mapping in Tasmania through Transfer Learning
Jose Padarian on behalf of Budiman Minasny

9:50-9:55 5.6 Spatio-Temporal Mapping Of Soil Organic Carbon Stock In Brazil
Nícolas Augusto Rosin

9:55-10:00 5.7 Mapping of soil indicators at national scale in Lithuania using the Soil Data Cube and Artificial Intelligence-driven Earth Observation analysis
Nikiforos Samarinas

10:00-10:15 Challenge 5 Discussion

10:15-10:35 Break

**Challenge 7. How to recognize, quantify and map soil functionality?**

**Moderator: Alfred Hartemink**

10:35-10:50 Challenge 7 Keynote
Philippe Lagacherie

10:50-11:00 7.1 Quantifying the potential and current state of European soils functions
Alexandre M.J-C. Wadoux

11:00-11:10 7.2 Identifying hotspots of polluted forest soils in the Czech Republic: comparison of various pedometrical methods
Luboš Borůvka

11:10-11:20 7.3 3D Soil Hydraulic Database of Hungary at 100 m resolution (HU-SoilHydroGrids)
Gabor Szatmári on behalf of László Pásztor

11:20-11:30 7.4 Concurrent Electromagnetic Induction Sensing of Magnetic Susceptibility and Electrical Conductivity for the Field Delineation of Soil Drainage Class
Richard J Heck

11:30-11:45 Challenge 7 Discussion

11:45-12:05 Field Trip and Desert Project Overview Curtis Monger

12:05-12:30 Pick up box lunch and load busses

12:30-17:00 Field Tour #1: Soils and Landscapes of Desert Basins and River Valleys (Separate Guide)

# Thursday 8th February

**Challenge 9. Can we find ways to express the uncertainty of predictions of soil properties or class allocations which are meaningful to the users of those predictions?**

**Moderator: Jason Ackerson**

9:30-9:45 Keynote Challenge 9
A-Xing Zhu

9:45-9:55 9.1 Uncertainty of spatial averages and totals of soil property maps
Gerard Heuvelink

9:55-10:05 9.2 Quantifying Prediction Uncertainty Based on Third Law of Geography
A-Xing Zhu

10:05-10:15 9.4 Exploring land use planners' preferences about visualization of digital soil mapping products for informed decision-making under uncertainty
Léa Courteille

10:15-10:25 9.5 New evaluation criteria for digital soil mapping products from an user's point of view
Philippe Lagacherie

10:25-10:35 9.6 Evaluating On-Farm Functional Soil Variability: A Decision Support Framework
Jonathan Maynard

10:35-10:40 9.7 Using LandPKS algorithm to estimate the sensitivity of ecological site identification in response to uncertainties in soil observations
Pedro Martinez

10:40-10:45 9.8 Leveraging user feedback and normalized uncertainty maps to inform future updates to national soil property maps
Travis W Nauman

10:45-10:50 9.9 Landscape uncertainty for DSM at continental scale
David Rossiter on behalf of Laura Poggio

10:50-11:05 Challenge 9 Discussion

11:05-11:30 Pick up box lunch and load busses

11:30-17:00 Field Trip #2: Soils and Landscapes of White Sands National Park

17:00-22:00 Conference Dinner. New Mexico Farm and Ranch Museum

## Friday 9th February

**Challenge 10. How to quantify soil contributions to ecosystem services with a framework enabling both local and regional soil management?**

**Moderator: Anru-Louis Kock**

| | |
|---|---|
| 9:30-9:45 | Challenge 10 Keynote<br>Cristine Morgan |
| 9:45-9:55 | 10.1 Quantifying the contribution of topsoil depth to ecosystem productivity across ecosystems and climatic regions<br>Yakun Zhang |
| 9:55-10:05 | 10.2 Soil's Hidden Value: Mapping Available Water Capacity as a Component of Natural Capital in Australia<br>Nicolas Francos |
| 10:05-10:15 | 10.3 Producing and Utilizing a Digital Twin for a G.E.M Analysis to Improve Sustainable Farming<br>Daniel J. Rooney |
| 10:15-10:25 | 10.4 The challenges of using references to interpret soil health indicators<br>Cristine Morgan on behlaf of Daniel Liptzin |
| 10:25-10:30 | 10.5 Contextualizing soil health measurements from farm to continent<br>Nathaniel Looker |
| 10:30-10:35 | 10.6 Quantifying Soil Health Through an Efficient Set of Indicators and Management Indices<br>Minerva J. Dorantes |
| 10:35-10:40 | 10.7 Scaling soil health assessment in the Golden Horseshoe region of Ontario, Canada<br>Jennifer A. Bower |
| 10:40-10:45 | 10.8 Spatial modeling of dynamic soil properties in agricultural landscapes.<br>David Rossiter on behalf of Valentina Rubio |
| 10:45-10:50 | 10.9 Quantifying the Spatial Variability of Dynamic Soil Properties<br>Sage Reuter |
| 10:50-11:05 | Challenge 10 Discussion |
| 11:05-11:25 | Break |
| 11:25-12:10 | Reflection and Future Work: Attendee Discussion<br>Janis Boettinger |
| 12:10 | Conference Adjourns |

# Abstracts Monday 5th Feb


Figure 2. Moonrise over a Typic Haplocalcid

## 1.1
## Continental Monitoring Soil Property Changes Under Human Pressure Using Pedogenon Mapping

Quentin Styc[1]*, Mercedes Román Dobarco[2], Ho Jun Jang[1], Budiman Minasny[1], Alex McBratney[1].

[1]School of Life and Environmental Sciences, The University of Sydney, Biomedical Building C81, 1 Central Avenue, Australian Technology Park, Eveleigh, Sydney, NSW 2015, Australia

[2]Basque Centre for Climate Change (BC3) 48940 Leioa, Spain

Soil properties are susceptible to changes due to human activities, particularly agricultural management. Traditional methods of monitoring these changes often lack the precision and granularity required for comprehensive understanding.

This work uses the innovative approach of pedogenon mapping (Román Dobarco et al., 2021) over Australia, which leverages high-resolution environmental covariates as proxies of soil- forming factors, including relief, parent material, and climate. This method delineates 1370 pedogenons in Australia where soils share similar forming factors. To discern soil changes, we employed the concepts of genosoil and phenosoil. Genosoils represent soils evolving under natural conditions, such as woodlands and native vegetation, while phenosoils depict soils under human-induced pressures, like cropping areas and pastures. By integrating data estimating human activity impacts using the global Human Modification map (Theobald et al., 2020) and the Habitat Condition Assessment System map (Harwood et al., 2016b), we can distinguish between these soil types (genosoil or phenosoil) within a pedogenon (Román Dobarco et al., 2023). Zonal statistics were computed to highlight differences in soil pH and soil organic carbon from soil profiles observations between genosoils and phenosoils.

Our findings indicate discernible changes in these properties, underscoring the impact of human activities on soil evolution. Pedogenon mapping, combined with the genosoil and phenosoil concept, offers a nuanced and precise tool for monitoring soil property changes due to human pressures. This approach holds promise for future research on and policy- making in sustainable land management.

References

Dobarco, M. R., Campusano, J. P., McBratney, A. B., Malone, B., & Minasny, B. (2023). Genosoil and phenosoil mapping in continental Australia is essential for soil security. *Soil Security*, 100108.

Dobarco, M. R., McBratney, A., Minasny, B., & Malone, B. (2021). A modelling framework for pedogenon mapping. *Geoderma*, *393*, 115012.

Harwood, T. D., Donohue, R. J., Williams, K. J., Ferrier, S., McVicar, T. R., Newell, G., & White,

M. (2016). Habitat Condition Assessment System: a new way to assess the condition of natural habitats for terrestrial biodiversity across whole regions using remote sensing data. *Methods in Ecology and Evolution*, *7*(9), 1050-1059

Theobald, D. M., Kennedy, C., Chen, B., Oakleaf, J., Baruch-Mordo, S., & Kiesecker, J. (2020). Earth transformed: detailed mapping of global human modification from 1990 to 2017. *Earth System Science Data*, *12*(3), 1953-1972.

## 1.3
## Century-Long Quantification of Soil Loss in Eastern South Dakota Agricultural Fields

Eli Halverson and Dr. Kristopher Osterloh South Dakota State University

Agriculturally driven increases in soil loss remain a barrier to long term sustainable agro- ecosystems. It is difficult to accurately quantify soil loss over multidecade time periods due to a lack of useful legacy data. Utilizing historical soil survey descriptions of agricultural soils from the 1920's and 1950's in Southeastern South Dakota, we quantify soil loss over the last century. Although these are missing modern horizon nomenclature, they include marker features such as horizon depths, depth to carbonates, and depth to parent material. These descriptions were utilized to assess the approximate 100-year changes in soil horizon thickness to quantify the amount of soil lost over the 100-year period. Changes in depth to carbonates, horizon depths and boundaries, texture changes and contrast, and depth to parent material were used to quantify the range of soil loss and subsequent mixing of subsurface and surface soil horizons. The rates of soil loss were between 1.92-3.53mm/year (26.11-48.85Mg/ha/year) which is comparable to studies that utilized shorter timescales. This study highlights the utility of legacy soil datasets as well as the importance of long-term trends in pedological modeling.

## 1.4
## Modelling the effect of topographical position and precipitation on soil profile variation with Soilgen

Tom Vanwalleghem, Vanesa García-Gamero, Adolfo Peña, Andrea Román-Sánchez, Peter A. Finke

Department of Agronomy, University of Córdoba, Da Vinci building, Madrid km 396 Rd., 14071 Córdoba, Spain

Department of Rural Engineering, Civil Constructions and Engineering Projects, University ofCórdoba, Da Vinci building, Madrid km 396 Rd., 14071 Córdoba, Spain.

Department of Environment, University of Ghent, Coupure Links 653, 9000 Ghent, Belgium.

The sensitivity of chemical weathering to climatic and erosional forcing is well established at regional scales. However, soil formation is known to vary strongly along catenas where topography, hydrology, and vegetation cause differences in soil properties and possibly chemical weathering. This study applies the SoilGen model to evaluate the link between topographic position and hydrology with the chemical weathering of soil profiles on a north- south catena in southern Spain.

Pedogenesis was measured and simulated in seven selected locations over a 20000-year period. A good correspondence between simulated and measured chemical depletion fraction (CDF) was obtained (R2=0.47). An important variation in CDF values along the catena was observed, although the position along the catena alone, nor by the slope gradient, explained this variation well. However, the hydrological variables explained the observed trends beter. A positive trend between CDF data and soil moisture and infiltration and a negative trend with water residence time was found.

The model sensitivity was evaluated with a large precipitation gradient (200-1200 mm yr-1). While a marked depth gradient was obtained for CDF with precipitation up to 800 mm yr-1, a uniform depth distribution was obtained with precipitation above 800 mm yr-1. The basic pattern for the response of chemical weathering to

precipitation is a unimodal curve, with a maximum around a mean annual precipitation value of 800 mm yr-1. Interestingly, this corroborates similar findings on the relation of other soil properties to precipitation and should be explored in further research.

## 2.1
## A global numerical classification of the soil surface layer

Alexandre M.J.-C. Wadoux[1,2], Alex B. McBratney[2]

[1]LISAH, Université de Montpellier, INRAE, IRD, Institut Agro, Montpellier, France

[2]Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Australia

The quest for a global soil classification system has been a long-standing challenge in soil science. There currently exists two, seemingly disjoint, global soil classification systems, the USDA Soil Taxonomy and the World Reference Base for Soil Resources, and many regional and national systems. While both systems are acknowledged as international, there remain various examples of their shortcoming for accounting of topsoil features, local applications and communication with established regional classification systems. This calls for a numerical soil classification that addresses these discrepancies and achieves harmonization with existing national systems. In this paper, we report on the development of a layer classification system

-as opposed to the classification of soil profile entities, as a first step towards achieving a comprehensive global numerical soil classification not based on a priori defined classes. We implemented a modelling approach with a set of predicted key soil properties available globally for the soil surface layer with the same depth range of 0-5 cm. The set of properties were partitioned into a number of homogeneous and disjoint classes using the $k$-means clustering algorithm. Next, we investigated the pattern of variation of the clusters in association with the soil property map with principal component analysis. A three- component nomenclature system is derived in a transformed space of the class-specific centroids to account for the uneven distribution of the centroids in the principal component space. We show that it is possible to build a data-based objective numerical taxonomic classification of soil layers, and that existing sets of key soil properties, predicted separately, coalesce into identifiable clusters or classes and manifest discernible spatial and/or pedological patterns. This grouping of key soil properties to logical categories is a possible step to better define diagnostic horizon features and suggest new ones. The general-purpose map of soil surface layer classes of the world also has potential applications in assessing soil change and designing monitoring surveys.

## 2.3
## Similarities among soil profiles in representative soil-landscapes plots and its implications for soil types discrimination-a case study in the three river’s sources area in Qinghai Province, China.

XIA ZHAO

Presenter’s email address: zhaoxia@qhnu.edu.cn; zhaoxia-qh@163.com

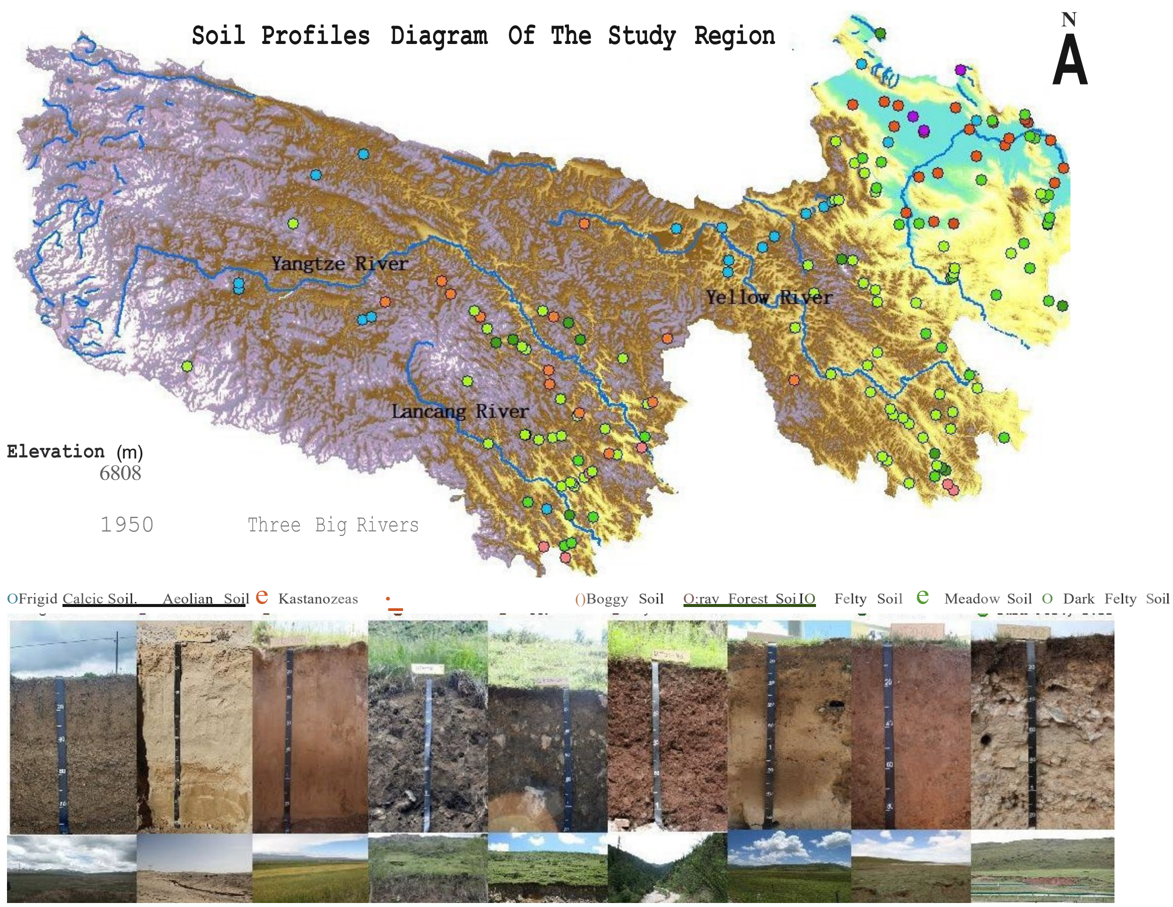
The assumption of intrinsic consistency among soil types-properties-landscapes is the scientific basis of soil type discrimination and soil attribute prediction, however, the criteria for judging these relationships are far from consistent across existing systems (e.g., soil taxonomy in the U.S., soil genetic classification system in China, and the legend units of world soil map), therefore, the question of developing a quantitative and numerical global soil classification that unifies the existing systems and enables transfer between them has been listed one of the big ten challenges that confront pedometrics.

The threes river’s (Yangtze River, Yellow River and Lantsang River) sources area in Qinghai Province owns

unique soil forming environments of Qinghai-Tibet Plateau, therefore some endemic soil types have been developed here, including frigid frozen soil and cold calcic soil that formed in high elevation frozen environment; felty soil, meadow soil and boggy soil that developed in Alpine wet and cold grassland environment; chernozem, kastanozems and aeolian soils that originated in Alpine arid desert environment.

We have identified several representative soil-landscape patterns over years of soil profiles survey (see the diagram below), we found that local people have more soil-landscape perceptions than we have imaged, and there are hot desires for digital soil mapping products on critical soil properties that restricting soil utilization, including effective soil thickness, gravel concentration both above surface and inside soil profiles, calcium carbonate content along soil profiles, ice contents and grass mat thickness etc.

So, the main idea of this work is to combine the quantitative rules of pedometric analysis with qualitative knowledge of local people by three steps, first is to calculate the similarities among soil profiles (including profile features and their environmental covariates) according to soil- landscape patterns, second is to transform the tactic knowledge of local people into qualitative rules, and finally integrate them into soil type identification criteria by referring to existing classifying system. The major outputs of this work will be a set of regional soil mapping products for both soil types and several important soil attributes that helpful to the rational utilization and protection of local soil resources.



## 2.4

## What is isotic anyway? A Soil Taxonomy mineralogy class revisited

**Ryan Hodges USDA, NRCS**

The concept of isotic soils and it's use in Soil Taxonomy at the family level was first introduced in 1996 to capture soils that did not meet the criteria to classify as having amorphic mineralogy or andic intergrades, but criteria exhibited soil properties akin to andic soils. These properties include having a colloidal fraction dominated by short-range-order mineralogy, higher than normal pH-dependent charge, and a greater ability to fix soluble P than other soils. Isotic soils key out at the family level under mineralogy class in sections C and D of Taxonomy, one step before mixed mineralogy. Some have argued that there is an apparent lack of interpretive value of the isotic versus the mixed mineralogy class, and that it is difficult to apply the class based on easily observed landscape/landform characteristics and correlation guides. The purpose of this study is to reassess not just the necessity of the isotic class, but to investigate the setting, context, and proxies for soil-forming factors associated with isotic versus mixed mineralogy taxa. In doing so, we will determine if there is any practical significance to both its use and the properties used to classify the isotic class. The KSSL laboratory data and other landscape data will be extracted, and various statistical analyses performed to compare differences in organic carbon, phosphate retention, and selective dissolution data between the amorphic, isotic, mixed mineralogy classes. Correlation analyses will be completed to determine degree of association of measured properties to those used in classifying isotic soils. Multivariate statistics will be used to determine soil properties and required thresholds that would best bifurcate soils into the isotic and mixed mineralogy classes. Results of our assessment will showcase which soil properties—both those that are and are not currently used to classify isotic soils—appropriately reflect the classification of soils into the isotic mineralogy class and what observed field correlation guides support the separation of isotic from mixed mineralogy, if any. Additionally, we will address how an improved taxa system at the junction between isotic and mixed mineralogy would increase their interpretive value on land use and management.

## 3.1

## Assessing natural and human drivers on soil thickness variation using generalized additive models

Yakun Zhang[1], Alfred E. Hartemink[1], Tom Vanwalleghem[2], Benito Roberto Bonfati[3], Steven Moen[4]

[1]University of Wisconsin-Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI, 53706, USA

[2]University of Córdoba, Department of Agronomy, Da Vinci building, Ctra. Madrid km 396, 14071 Córdoba, Spain

[3]State University of Minas Gerais, 700 Colorado Street, Passos, Minas Gerais, 37902-092, Brazil

[4]University of Wisconsin-Madison, Department of Statistics, Medical Sciences Center, 1300 University Avenue, Madison, WI, 53706, USA

Understanding changes in spatio-temporal paterns of soil thickness and their natural and anthropogenic driving factors are essential for earth system modeling and natural resource conservation. Here we compiled a long-term (1950–2018) and large-scale (conterminous USA) topsoil (A horizon, n=37,712) and solum (n=22,409) thickness data. We fited generalized additive models (GAMs) to quantify the spatial distributions of soil thickness and the nonlinear relationship between soil and environmental variables and A horizon and solum thickness. The GAMs resulted in an $R^2$ of 0.35 and 0.34 and Lin's concordance correlation coefficient of 0.52 and 0.52 for log-transformed A horizon and solum thickness respectively in the validation. The model coefficients of GAMs explained either a positive or negative contribution of each environmental factor on soil thickness variation. We found that climate was associated with the spatial distribution of soil thickness. The A horizon and solum thickness displayed a strong longitudinal pattern which was correlated with soil moisture (r=0.49) and temperature (r=0.74), respectively. Elevation influenced solum thickness via soil temperature at the national scale. When selected chronosequences in land resource regions to quantify their temporal variations using simple linear regressions. Temporal changes of the thickness varied across different land resource regions, which were affected by topography, land use, and erosional processes. These results clearly elucidated the factors controlling the soil production and erosion at the national and regional scales and identified regions that require conservation practices to reduce further topsoil loss.

## 3.3
## The determinants and regulation of surface soil bacterial and fungal biogeography in Australia

Peipei Xue[1], Budiman Minasny[1], Alexandre M.J.-C. Wadoux[1,2] , Mercedes Román Dobarco[1], Alex McBratney[1]

[1]The University of Sydney, Sydney, New South Wales, Australia

Soil microorganisms are highly abundant and diverse, playing crucial roles in nutrient cycling, carbon sequestration, and soil structure regulation. However, understanding their distribution and the environmental factors influencing them at a continental scale remain a challenge due to their high density, diversity, and the need for molecular techniques to study them comprehensively. In this study, we investigate the determinants of soil bacterial and fungal distribution of the soil surface horizon across Australia using a comprehensive dataset of more than 1000 soil samples from diverse bioregions and soil types.

Our findings highlight that the interplay between soil properties and climate factors stands out as the primary driver of microbial distribution at the continental level. Principal coordinate analysis reinforces the notion that soils sharing similar characteristics tend to exhibit similarity in the composition of bacterial and fungal communities.

Leveraging these insights, we developed digital soil mapping models that establish associations between observed microbial abundances and environmental variables, allowing us to create continental maps of soil bacteria and fungi. These maps unveil microbial hotspots, such as the eastern coast, southeastern coast, and western coast, which are dominated by Proteobacteria and Acidobacteria. In the case of fungi, precipitation emerges as a dominant influence, with Ascomycota prevailing in the drier? central region. The detailed maps also indicate that some of the microbial hotspots are located in areas with high human pressure which may be vulnerable to change.

The map can be instrumental for regional soil biodiversity assessments and for monitoring how microbial communities respond to global environmental changes.

## 3.4
## Biplots for understanding machine learning predictions in digital soil mapping

Stephan van der Westhuizen[1,2], Gerard B.M., Heuvelink[2] Sugnet Lubbe[1], Catherine E. Clarke[3]

[1] Department of Statistics and Actuarial Sciences, Stellenbosch University, South Africa

[2] Soil, Geography and Landscape, Wageningen University, the Netherlands

[3] Department of Soil Science, Stellenbosch University, South Africa

Keywords: biplot, explainable machine learning, random forest, soil organic carbon, South Africa, XAI

In digital soil mapping, machine learning models are often preferred to traditional statistical methods such as geostatistical models. The reason for this is that machine learning models can effectively capture complex relationships between soil properties and environmental covariates, leading to more accurate soil maps compared to traditional models. However, unlike traditional models, a notable drawback of machine learning models is that they are often referred to as "black-box" methods due to their limited ability to provide comprehensive interpretations for their predictions.

Explainable machine learning, a rapidly growing field in machine learning literature, focuses on model-specific or model agnostic methods designed to understand predictions made by machine learning models either locally or globally. Popular model-agnostic methods include partial dependence plots (used for global interpretations), independent conditional expectation (local) curves, and Shapley values (local and global). These methods assume independence between covariates which is a very restrictive assumption. For cases where covariates are dependent, an alternative approach is the Accumulated Local Effect plot, which however is limited to depicting one or two covariates at a time.
Another disadvantage of the above-mentioned methods is that no readily available goodness-of-fit metric is available.

In this paper we propose the use of principal coordinate analysis biplots as a model-agnostic visualization approach for understanding the predictions made by a machine learning model for digital soil mapping. A biplot is a powerful visualization tool that is often used to seek patterns in multivariate data. A biplot would allow a user to investigate machine learning model predictions locally and globally, and does not require any assumptions (e.g., independence between covariates) about the data. There is also no limit to the number of covariates that can be viewed at a time.

Furthermore, an analytically derived goodness-of-fit metric is provided which allows the user to evaluate the accuracy of the approximation. We present examples from a case study in South Africa in which soil organic carbon is mapped with a random forest model. Our findings show that biplots can provide meaningful interpretations for the soil organic carbon predictions and its relation with covariates where other explainable machine learning methods fail.

## P2
## Spatio-temporal soil information based on open science and multidisciplinary collaboration

Taciara Zborowski Horst[1], Alessandro Samuel-Rosa[1]

[1] taciaraz@utfpr.edu.br; alessandrorosa@utpf.edu.br; Federal University of Technology-Paraná (UTFPR) - Brazil

Digital soil mapping has played an essential role in generating global soil information at various scales. Mapping dynamic soil properties across large areas, however, has not been addressed due to its greater data requirements to encompass both spatial and temporal coverage. In this presentation, we will showcase how MapBiomas, a network formed by NGOs, universities, and technology startups, is addressing these challenges to deliver annual updates of soil property maps for Brazil. Employing large amounts of open field and satellite data, cloud computing, and machine learning, MapBiomas produced a series of maps of topsoil organic carbon

(SOC) stocks of the Brazilian territory covering the period 1985-2021 with temporal and spatial resolutions of one year and 30 m, respectively. The series will be updated annually to show how the spatio-temporal dynamics of SOC stocks are linked to land cover and use, climate, and soil management. Such updates are only feasible due to the way the network is organized. First, MapBiomas covers the mapping of various themes, with working groups updating every year the series of annual maps of land cover and use, fire scars, water surface, and infrastructure among others. Second, a multidisciplinary team of experts is based at institutions across the Brazilian biomes. These experts interpret, choose, and process the existing environmental data proxies to best represent the processes linked to SOC changes in their biomes. Third, a community effort to retrieve, curate, standardize, and harmonize any existing field soil data collected by public and private organizations, making it immediately made available to the wider community via the Brazilian soil data repository (SoilData). Finally, the network maintains a transparent and accessible framework, with a regular agenda of outreach activities, encouraging open access to data, code, and results, thereby fostering data sharing, reproducibility, and reuse. Fostering community engagement and adopting open practices is key to enable the participation of the general public in providing data and assessing the quality of the data products (citizen science). It is our understanding that only through strong multidisciplinary networking, integrating the production of soil information in a broader framework, and adopting open practices, the community will be able to address demands by decision makers and soil managers.
Keywords: MapBiomas; Open data; Land assessment; Soil organic carbon stock.

## P4
## Leveraging Legacy Data: The Evolution of Mid-Infrared Spectroscopy at the NRCS Soil and Plant Science Division

Authors: Jonathan J. Maynard1*, Rich Ferguson1, Scarlett Murphy1, Cathy Seybold2, Andrea Williams1, Travis Waiser3, Dave Hoover1, and David Lindbo4

1 USDA-NRCS, Soil and Plant Science Division (SPSD), National Soil Survey Center 2USDA-NRCS, SPSD, Soil Services and Information
3USDA-NRCS, SPSD, South Central Region 4USDA-NRCS, SPSD, Headquarters
*Presenting author

Soil mid-infrared spectroscopy (MIR) has emerged as a highly effective and efficient technique for predicting key soil properties (e.g., clay, organic carbon, cation exchange capacity (CEC), etc.). With a background of evidence supporting its potential in soil survey applications, the Natural Resources Conservation Service (NRCS) Soil Plant and Science Division (SPSD) Kellogg Soil Survey Laboratory (KSSL) initiated the use of MIR as an additional method for internal laboratory quality control in 2011. The success of this initial work prompted consideration for expanding MIR technology to NRCS field offices, leading to an initial pilot study in the Central Plains of the United States in 2017. To facilitate the transfer of MIR calibrations from KSSL to suitably equipped field offices, careful consideration was given to selecting appropriate sample processing and scanning technologies. This strategic approach ensured optimal model transfer and laid the foundation for successful implementation of MIR technology beyond the pilot stage, with current deployment of MIR technology in field offices stretching from Alaska to Puerto Rico. The KSSL has greatly expanded its MIR spectral data collection to capture broad compositional variability in both the United States and abroad, resulting in the establishment of the world's largest open-source MIR spectral library, comprising spectra for over 84,000 soil samples (as of 2023, with many more to scan). By exploiting its vast archive of diverse soil samples and reference data, NRCS has also facilitated progressive, organized, and cooperative research and development with the global spectroscopy community.

## P5

## Enhancement of Soil Data in the U.S. Forest Service Forest Inventory and Analysis Program

John D. Shaw[1,*] and Colby Brungard[2]

[1]USDA Forest Service, Rocky Mountain research Station, [2]New Mexico State University, Dept. of Plant and Environmental Sciences, *john.d.shaw@usda.gov

The Forest Inventory and Analysis (FIA) program has been collecting data and reporting on the status and trends of the forests of the United States for almost 90 years. Over the decades, the FIA program has diversified in response to new questions and data needs. In the late 1990s, FIA implemented a suite of "indicator" protocols, which were transferred to FIA from an EPA program that was designed to monitor forest health. One component was soil sampling to 20cm, with the common chemical analyses being done for mineral soil and primarily C and N done for forest floor samples. In comparison to FIA data collected for vegetation, data produced by the soil sampling protocol has been underutilized. This was due to many factors, among which were the low sampling intensity (~1 soil plot per 44,500 ha) and the (unrealistic ?) expectation that the protocol would permit the evaluation of soil property changes on a 5-year plot revisit cycle. Given the high cost, low value return on the data, and impending budget shortages, collection of soil data ceased in most states between 2005 and 2007. However, the Rocky Mountain Research Station FIA program (RMRS-FIA), which covers AZ, CO, ID, MT, NV, NM, UT, and WY, took the opposite approach and revised its soil protocol to move toward the full sampling intensity (~1 plot per 2500 ha) that is used for forest vegetation and other attributes. In addition, RMRS-FIA, in partnership with New Mexico State University, has started to extend FIA soil data to make it more useful. For example, texture is now determined by particle size analysis, compared to only texture-by-feel done previously. Dry-end moisture release curves are being developed for all FIA mineral samples collected since 2000. In addition, the existing laboratory data and stored samples are being used to develop a spectral library for near-infrared analysis. The existing and additional soil data, combined with the higher sampling intensity and the fact that every soil sample comes from a location with continuous vegetation monitoring, should provide a rich dataset for analysis of forest soil properties in the Mountain West states.

## P6

## Scaling Carbon Stock Measurement for Carbon Markets

Sarah Coffman and Marissa Wiseman

Yard Stick is actively working to meet the growing need for accurate and scalable carbon stock measurements for agricultural carbon markets. We want to enable measurement at a scale of millions of acres. This presentation will present Yard Stick's path to scalability through our processes, software, and patented VisNIR technology. We will discuss how Yard Stick is creating efficiency at every step of the customer life cycle (scope, sample plan design, field work, lab data, and data return). Software solutions for sample plan design, field work execution, lab data review and stock reports will be described and displayed. The unique role that the Yard Stick VisNIR handheld probe plays in scaling C stock measurements will be intertwined throughout the content.

## P7

# Commercial soil carbon accounting: challenges and opportunities for practicing pedometricians

Jason Ackerson, Ayush Guwali, Faye Smith, Matt Duncan, Cristine LS Morgan

Recent interest in soil carbon sequestration as a tool to mitigate climate change has spurred the creation of a soil carbon accounting industry. Registration bodies such as Verra and The Climate Action Reserve have developed methodologies for commercial soil carbon developers to generate tradable soil carbon offsets and businesses have made substantial investments in developing and selling soil carbon offsets. This emerging soil carbon industry provides challenges and opportunities for pedometricins to work in conjunction with commercial partners to accurately quantify soil carbon sequestration on spatial large scales. In this presentation, we will discuss the partnership between The Soil Health Institute, a non-profit scientific organization, and Truterra a for-profit carbon developer. We will highlight key learnings and insights from this partnership including: 1) challenges in sample design and implementation with commercial partners, 2) limitations of pedometric tools in real-world applications, and 3) opportunities of novel pedometrics research.

# Abstracts Tuesday 6th Feb

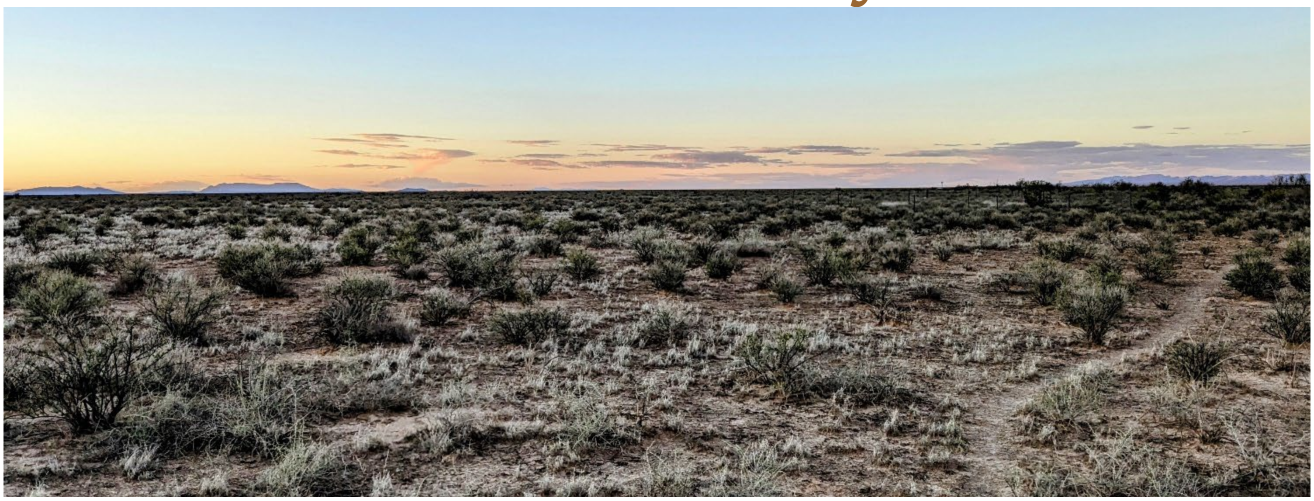


Figure 3. Sunset on the Jornada Experimental Ranch

### 4.1

### Development of soil spectroscopy prediction models for the Western Highveld region, South Africa: Why we need local data.

Anru-Louis Kock a, George Van Zijl a and Dimakatso Ramphisa a

a Unit for Environmental Sciences and Management, North-West University, Potchefstroom, North West, South Africa.

Precision agriculture practises must overcome a special set of obstacles due to the spatial heterogeneity of South African soils. Traditional soil analysis methods are costly, restricting data availability for precision applications. Soil spectroscopy has recently been mentioned as a cheaper, more time effective alternative, but it requires accurate calibration algorithms, using local samples. This study evaluates the potential of Mid-Infrared (MIR) reflectance spectroscopy to predict exchangeable base cations, pH (KCl), and phosphorus (Bray-1 P) in cultivated soils. Mid infrared spectra were obtained from samples, alongside measurements of pH (KCl), NH4Oac extractable exchangeable base cations, and Bray-1 P. Calibration algorithms were created using the Cubist, Partial Least Squares Regression (PLSR), and Random Forest (RF) machine learning algorithms to develop local prediction models. Additionally, a subset of spectra was also submitted to the newly developed global soil spectral database with prediction models – Open Soil spectral Library (OSSL) to obtain its predictions based on the spectra. The results demonstrate promising outcomes for local predictions at regional scale.
Accurate predictions for pH, calcium (Ca), and magnesium (Mg), with ratio of performance to inter- quartile distance (RPIQ) values surpassing 2.13 were achieved. However, predictions for phosphorus (P), potassium (K), and sodium (Na) did not meet reliability requirements. The results from the OSSL predictions were consistently less accurate than the local models with the OSSL model overpredicting on all soil properties except pH (KCl) of which there is no prediction model. RPIQ values for all soil properties predicted with the OSSL models were < 1. This indicates the importance of contextual specificity when developing predictive tools for local sites as the prediction models calibrated with local samples outperformed global prediction models for the aforementioned soil properties. This regional focus enhances the accuracy of predictions, aligning them more closely with the unique characteristics of South African croplands. By prioritizing regional precision models, this work contributes to the evolution of agriculture in the North West province and, more

broadly, to the development of precision agriculture in South Africa.

## 4.3

## How can we be more assertive about soil spectroscopy predictions? The Open Soil Spectral Library study case

José Lucas Safanelli1, Jonathan Sanderman1, Robert Minarik2, Leandro Parente2, Tomislav Hengl2, Dellena Bloom3, Katherine Todd-Brown3

1Woodwell Climate Research Center, Falmouth, MA, USA 2OpenGeoHub Foundation, Wageningen, the Netherlands 3University of Florida, Gainesville, FL, USA

Diffuse reflectance spectroscopy is a technology that has been extensively investigated for estimating soil properties due to its appealing characteristics of being rapid and cost-effective. A research problem that is still under investigation in the field of soil spectroscopy is how we can be more assertive about the quality of the spectrally-based predictions and how they impact user applications. With the development of the Open Soil Spectral Library (OSSL) as part of the Soil Spectroscopy for Global Good initiative (SS4GG), we faced this challenge and proposed two interrelated solutions based on recent advances reported in the literature. The first is the uncertainty estimation via conformal prediction, a method that has gained attention in recent years due to its intuitive yet robust derivation of uncertainty for a predefined error probability.

The second is more common in the chemometrics field and helps to flag potential outliers or underrepresented samples respective to the trained model, usually referred to as control chart, but here defined as trustworthiness flag. By using principal component analysis for controlling multicollinearity and for dimensionality reduction of the spectra, we calculate the residual unexplained variance (q-statistics) of new samples and compare it with a critical value estimated from the training set as part of the trustworthiness flag. We also implement uncertainty estimation via conformal prediction by leveraging the 10-fold cross validation predictions from an internal evaluation, resulting in two separate models for calculating prediction intervals: response and error models. This study will describe both methods implemented in the OSSL Engine with detailed results of their strengths and limitations.

Keywords: chemometrics, uncertainty, conformal prediction, trustworthiness, q-statistics

## 4.4

## Preserving Soil Data Privacy with SoilPrint: A Unique Soil Identification System for Soil Data Sharing

Tegbaru B. Gobezie[1]* and Asim Biswas[1]

[1.] School of Environmental Sciences, University of Guelph, Guelph, ON, Canada

*Presenting author

Soil is an indispensable resource with critical implications for various fields such as agriculture, environmental science, climate change, hydrology, ecology, and geoscience. Accurate and accessible soil data is crucial for making informed decisions. However, the sharing and harmonization of soil data present significant challenges, particularly due to the lack of a comprehensive identification system that ensures privacy, ownership, and stewardship in a federated data sharing framework. Moreover, the inherent heterogeneity of soil properties across space and time complicates the establishment of connections between soil profiles and their corresponding properties at specific locations. To overcome these challenges, we propose a novel and persistent soil data identifier called SoilPrint, which can be likened to a fingerprint.

SoilPrint utilizes a mathematical algorithm that effectively integrates the properties of soil profile layers (SPLP) with Geohash, offering an efficient solution. By incorporating SoilPrint, the process of data federation becomes seamless within a secure and distributed ledger, eliminating the need for complex data mapping or alignment. This approach ensures data privacy and ownership throughout the sharing process, addressing concerns associated with data management. To demonstrate the practical application of SoilPrint, we present a case study using soil data from Ontario, Canada. The results underscore the unique identification capabilities of SoilPrint for soil profiles and their associated properties, making it a promising tool for soil data management. SoilPrint facilitates data tracking, reuse, and analysis, enhancing the efficiency and effectiveness of soil-related research and decision-making processes.

Key Words: Soil data, Unique Identification, SoilPrint, Federated

## 4.5

## Using Vis-NIR, MIR, and pXRF spectra for predicting soil physical and chemical properties - A comprehensive review

Gafur Gozukara[1]*, Alfred E. Hartemink[2]*, Yakun Zhang[2], Jingyi Huang[2],

[1]Eskisehir Osmangazi University, Department of Soil Science and Plant Nutrition, Eskisehir, 26160, Türkiye. [2]University of Wisconsin-Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI, 53706, USA.

*Corresponding e-mail: ggozukara@ogu.edu.tr, hartemink@wisc.edu

We reviewed 305 published papers that used Vis-NIR, MIR, and pXRF spectra based for predicting soil properties. The objectives of this review were to compare the prediction accuracy using the extracted coefficient of determination (R2) values of Vis-NIR, MIR, and pXRF spectra, and to understand which factors impact characterization and prediction accuracy. The results demonstrated that spectral prediction papers increased exponentially from 2001 to 2022, and that much work has been conducted in China, USA, and Brazil. Approximately 44% of papers focused on the prediction of SOC using Vis-NIR spectra. The partial least square regression was most widely used. Many papers focused on the prediction performance in the topsoil (<40 cm) and Alfisols, Inceptisols, and Entisols using Vis-NIR, MIR, and pXRF spectra. The prediction accuracy of all soil properties was affected by soil type, depth, horizon, preprocessing methods, spectral range, and type of the prediction models (i.e., machine and deep learning). We recommend MIR spectra to obtain the highest prediction accuracy for sand, clay, total nitrogen (TN), total carbon (TC), inorganic carbon (SIC), organic carbon (SOC), organic matter (SOM), cation exchange capacity (CEC), and pH.

Keywords: Proximal soil sensors, soil spectral information, predictive models, soil pedogenesis

## 4.6

## Spectral signature of soil horizons and soil orders in Wisconsin

Malithi V. Weerasekara, Alfred E. Hartemink, Yakun Zhang

University of Wisconsin–Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI 53706, USA

We used MIR spectra to classify soil horizons and soil orders from a dataset comprising 99 pedons and 321 samples collected from genetic horizons across five soil orders (Alfisols, Entisols, Mollisols, Spodosols, and Histosols). The MIR spectra (4000 to 600 cm-1) and soil properties (soil organic carbon, pH, texture, Fe, Si) were measured. We used random forest model to classify five master horizons (O, A, E, B, C), three B horizons (Bs, Bt, Bw) and soil orders. Soil master horizons and B horizons had prediction accuracies of 0.83 and 0.94, respectively, while soil orders had an accuracy of 0.79 in the validation. Absorption peaks of MIR are a result of fundamental molecular vibrations, which can distinguish soils with different organic and mineral

compounds. Organic soils exhibited unique absorption characteristics distinct from those of mineral soils at 3,695 cm-1 and 3,620 cm-1. The random forest model accurately distinguished the O horizon with a precision of 100%. In addition, the spectra of Bs horizons and topsoil (average of O and A horizon) of Spodosols were comparable to the O horizon and made them easily identifiable using the spectral curve. The distinctions between soil horizons, soil orders and their spectral features were related to the soil physical and chemical properties. The C horizons had the highest sand content (mean = 86%) and stronger absorption peaks in 2000 – 1,650 cm-1. Spodosols and Entisols, which have high sand content, displayed these peaks, which enabled distinguishing them from other soil orders. The C horizon had the highest pH (mean 6.1) and showed a spectral peak at 2,517cm-1 representing $CaCO_3$ availability whereas the E horizon which has the lowest pH (mean = 5.3) showed the lowest absorbance at the same spectral range. Our results show the potential of soil MIR spectra for accurate soil horizon and order delineation, particularly for distinguishingly different soils.

## 4.8
## Mesoscale Soil Spatial Heterogeneity Characterization Using Laser-Induced Breakdown Spectroscopy

Fei Ma[1], Changwen Du[1,2,*], Zhou Jianmin[1]

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

[2]University of Chinese Academy of Sciences, Beijing 100049, China

Soil heterogeneity studies increasingly require approaches to describe soil composition and features from macroscale to microscale with more visualization methods. In this study, the laser induced breakdown spectroscopy (LIBS) technique was used to generate microscale ablation craters to investigate soil variation in various soil samples at the mesoscale level. The LIBS spectra of four types of agricultural soils pellets (Fluvo-aquic soil, paddy soil, red soil, and black soil in four provinces in China) having different soil organic matter contents were collected and there were 52 shot spots on the surface of each sample, respectively. The elements Si, Al, Ca, Cu, Ti, Mo, Fe, Ba, Mg, Na, Li, K, H, and O were identified; however, their emission line intensities varied in different soil types. The first three principal component analysis (PCA) loading values and scores reflected the correlation of elements in each soil sample and the first 10 positions that contained most PC1 scores were marked separately. The results showed that the Fe, Ti, Al, Mo, and O loading values were positive with soil organic matter (SOM) content, whereas Ca and Na were negative in Fluvo-aquic soil samples; Al, Mo, Ti, Li, and O were positive with SOM contents, while Ca and Mg showed the opposite changes in paddy soil samples. Fe, Mo, and Ti decreased with the decrease in the SOM contents in red soil samples. Ti, Al, Mg, Ca, Fe, and K showed strong correlations loading values in the black soil samples. Finally, the Red-Green-Blue composite displayed visualized soil heterogeneity maps. The soil sample maps indicated high SOM content in the Fluvo-aquic soil, paddy soil, and red soil were with higher variability. For the Fluvo-aquic soil with medium SOM content, Ca was abundant on the pellet surface. Ti, Mo, and Cu were richer on the surface of paddy soil with medium SOM content. Al, Ca, Ti, and Na were abundant on black soil pellet with low SOM content. Moreover, the red soil types displayed highest heterogeneity in all four types. These results may help extend the utilization of spectral techniques for soil heterogeneity at the mesoscale level for various soil types.

Keyword: Laser induced breakdown spectroscopy, Mesoscale soil heterogeneity, Principal component analysis, Kriging interpolation, RGB composite

## 4.10
## Mapping soil particle fractions by training Digital Soil Mapping models with surrogate measurements obtained from laboratory and satellite Vis-NIR spectral

## data.

Weerasekara[1,2] M., Lagacherie[2] P., Dharumarajan[3] S., C. Gomez[2,4]

[1]University of Wisconsin-Madison, Department of Soil Science, 1525 Observatory Drive, Madison, WI 53706, USA
[2]LISAH, University Montpellier, IRD, INRAE, Institut Agro Montpellier, 34060 Occitanie Montpellier, France
[3]ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre Bangalore-560024, India
[4]Indo-French Cell for Water Sciences, IRD, Indian Institute of Science, Bangalore-560012, India Digital Soil

Mapping (DSM) is an excellent method for spatial soil prediction and is well suited for the growing appetite in spatial soil information. Due to the costly expense of acquiring conventional soil data, a lack of soil inputs is a limiting factor in Digital Soil Mapping (DSM). Laboratory Visible and near-infrared (Vis- NIR, 400–2500 nm) spectroscopic data and Sentinel-2 (S2) satellite data are promising alternatives for predicting soil properties, and these surrogate data can enhance spatial samplings. In this study, both laboratory Vis-NIR data and S2 data were utilized independently to predict soil data and then complement conventional soil data for the optimization of DSM models. Subsequently, these models underwent testing for sand, silt, and clay mapping. The results demonstrate that adopting few soil properties data predicted with high accuracy by laboratory Vis-NIR spectra as surrogate data did not significantly improve the DSM model performance. Conversely, using many soil properties data predicted with modest accuracy by S2 data improved the DSM model performances for sand and clay. None of the DSM models could predict silt accurately due to its low variation across the study area. S2 based model for predicting sand content performed best with $R^2$val, RMSEval and MEC of 0.62, 9.35 % and 0.61, respectively. Finally, the limited spatial density of soil properties data predicted by laboratory Vis- NIR data hindered local spatial variation capture, while soil properties data predicted by Sentinel- 2 data significantly improved predictions despite their larger uncertainty. High-density soil dataset improved performance, resulting in markedly more accurate results. The abundance of S2 data at high frequencies holds the potential to propel the DSM community toward its objective of refining existing soil maps.

## 4.11
## In-Situ Soil Spectroscopy Application for Extractable Phosphorus Prediction for Precision Agriculture Purposes

Katsutoshi Mizuta Ohio Wesleyan University

Soil degradation resulting from excessive phosphorus fertilizer applications poses a significant threat to food security. To optimize fertilizer application, it is crucial to understand the spatial heterogeneity of soil phosphorus in farm fields. However, traditional soil sampling and chemical analysis methods are expensive, labor-intensive, and time- consuming, often leading to over or under-application of fertilizers by farmers.
As an alternative, researchers have explored spectroscopy technology. Some studies explored the use of soil spectrum to predict extractable soil phosphorus (Ext-P); however, many of the predictive models developed are still driven by data collected from a small scale (<20ha) field unlike soil carbon studies. More evidence is necessary to determine the necessary prediction accuracy for achieving the environmentally, agronomically, and economically optimal site-specific fertilizer application. Therefore, the objective of this paper is to establish spectral prediction models for soil Ext-P and assess their usefulness in phosphorus management. Sampling locations were determined using a one-acre and four-acre grid method, resulting in 513 and 144 locations, respectively. Soil samples were collected from seven Oregon farm fields in 2022 fall and analyzed for Ext-P using the Mehlich-3 extraction method. Each sample was scanned using visible-near- infrared spectroscopy in the spectral range of 350–2500nm. Satellite imagery from previous years was also collected to define yield-based management zones (MZ) within the fields. The accuracy of soil spectroscopy in predicting Ext-P was evaluated by comparing it with laboratory-measured Ext-P. Geospatial models were used to generate maps based on both laboratory-measured and spectrum-based predicted Ext-P, which were then overlaid with the MZ map. Preliminary results indicate that the soil spectroscopy approach has the potential to effectively identifies yield-limiting zones associated with Ext-P. Detailed nutrient mapping can assist farmers

in reducing excessive fertilizer use, saving costs, and mitigating P-related environmental pollution. Further research is needed to explore calibration models that improve prediction accuracy and optimize farmers productivity by using the spectral models specifically tailored for the yield-limiting zones in their fields.

## 4.12
## An objective test of the Open Soil Spectral Library service

*Kanchan Grover and Colby Brungard*

*New Mexico State University, Las Cruces, NM, USA*

The promise of soil spectroscopy is the replacement of laborious and costly wet chemistry analytical methods with rapid and low-cost spectral analysis regardless of the soil type or geographic location. Early research demonstrated the ability of soil spectroscopy to produce accurate and precise analytical results if the spectra under investigation closely matched spectra from a geographically-specific spectral library. This led to the development of large spectral libraries and the development of internet-based spectral modeling capabilities to handle large and diverse spectral libraries. The Open Soil Spectral Library (OSSL) is an example of an on-line spectral modeling service built on top of large soil spectral libraries. This research investigates the accuracy of the OSSL for predicting analytical results from the North American Proficiency Testing (NAPT) program. NAPT furnishes analytical laboratories with blind and double-blind soil surface samples from different soil types across the conterminous USA. Each participating laboratory returns their analytical results to the NAPT program which then publishes the average test results. Thus, NAPT represents an exceptionally robust, and geographically diverse dataset of analytical results.
This research used 325 NAPT soil samples to evaluate the precision and accuracy of OSSL Midinfrared (MIR) tools for predicting 27 soil properties: Total Nitrogen, Total, Organic, and Inorganic Carbon, Cation Exchange Capacity (displacement) Clay, Sand, Silt, Electrical Conductivity (1:2), pH (1:1), pH (1:2) 0.01M CaCl2, and Extractable K, Ca, Mg, Na, Al, Fe, Mn, Fe, Cu, and B

NAPT soil spectra were measured using a Bruker Invenio-R HTS-XT l and uploaded to OSSL engine v1.2. Results show varying degrees of prediction accuracy across different soil properties. Total carbon, total nitrogen (N), total sand as well as cation exchange capacity (CEC), were predicted with high accuracy. Silt, and calcium extractable using ammonium acetate were moderately well-predicted, demonstrating an acceptable level of accuracy. However, soil pH (both in water and CaCl2), Carbonates, and Clay showed a marginal level of acceptability. The Mehlich 3 extractable elements: K, Mg, Fe, Cu, Ca, B, Na, Mn Al were not predicted well. Likewise, Ammonium acetate extractable elements; Na, K, Mg were poorly predicted. Electrical conductivity (1:2) and Aluminum extractable KCl as well as phosphorus extractable Bray 1, did not predict well.

## 4.13
## Quantitative Soil Profile Observations

Alfred E. Hartemink, Jingyi Huang & Yakun Zhang

University of Wisconsin-Madison, Department of Soil Science, 1525 Observatory Drive, Madison Wi 53711, USA

Understanding the earth's soil mantel has advanced through countless observations, some with depth, many across large areas. Soil observations have been fed into theory and models, whereas some theory has been formed from observations. The triangular relationship between soils, landform, and land use sometimes across large areas (climatic or parent material gradients) is often a driver for contemporary soil studies. Much attention is given to sampling schemes, analytical techniques such as spectral pedology, and spatial prediction methods that have matured in the last few years. Observations with depth and across a soil profile wall have advanced from 1-D to 2-D using proximal sensing techniques and through image analysis methods. Soil

horizon delineation methods have improved using cluster analysis of proximally sensed data sourced from vis-NIR or XRF spectroscopy, or digital images. Also, 1-D continuous depth functions and 2-D soil profile maps have been generated to visualize and quantify soil profile variations.
Sampling schemes have been established for quantifying soil profile variation and reducing sampling effort. Sensor data fusion methods can quantify the soil profile attributes, and guidelines should be developed for routine quantitative and analytical methods of soil profile observations.

## 4.14
## Multi-sensor soil probe and machine learning modeling for predicting soil properties to revolutionize sustainable agriculture

Sabine Grunwald et al.

This study introduces a data-driven, multi-sensor digital soil mapping approach to assess soil health for precision agricultural management. Our ATV-mounted Digital Soil Core (DSC) system contains seven different sensors including sleeve friction, tip force, dielectric permittivity, electrical resistivity, soil imagery, acoustics, and visible and near-infrared spectroscopy. These sensors have been integrated into a penetrometer system developed by LandScan to sense soil characteristics at high spatial resolution (mm scale) along in-situ soil profiles up to a depth of 120 cm. The sensor data collected with the DSC are integrated into a data cube providing vertical high-density knowledge associated with physical-physical- chemical-biological soil conditions. In contrast, soil samples derived from soil cores for lab-based soil analytics are bound by substantially coarser spatial resolution and multiple compounding errors. We investigated the effects of mismatched scale between high-resolution proximal sensor data and coarser resolution soil lab measurements to develop soil prediction models. Our case study was conducted in central California in soils used for almond production. We collected multi-sensor data with the DSC and spatially co-located soil cores that were sliced into narrow horizons for lab-based soil measurements (for example soil organic carbon, texture, B, Ca, Cu, Zn, pH). Partial Least Squares Regression (PLSR) cross- validation was used to compare results testing four data integration methods. Method A reduced the high- resolution sensor data to discrete values paired with horizon-based soil lab measurements. Method B used stochastic distributions of sensor data paired with horizon-based soil lab measurements. Method C allocated the same soil analytical data to each one of the high-resolution multi-sensor data within a horizon. Method D linked the high-density multi-sensor soil data directly to crop responses (crop performance and behavior metrics) bypassing costly laboratory soil analysis. Overall, the soil models derived from Method C outperformed Method A and B. Soil predictions derived using Method D were most cost-effective and practical to assess soil-crop relationships and is well suited for industrial-scale precision agriculture applications. Method D represents a paradigm shift from conventional methods of soil property prediction using laboratory or other subjective and error-prone approaches and is not directly comparable to the other methods.

## 4.15
## Spectral inference at the edge

José Padarian*, Budiman Minasny, Alex McBratney

Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, New South Wales, Australia

The use of soil infrared spectroscopy has been considered as a viable technique to acquire soil information for soil monitoring, in an effort to complement or replace soil laboratory analyses. The combination of large soil spectral databases and advances in modelling techniques have made predictive spectral modelling one of the main applications of machine learning in soil sciences. Notably, advanced machine learning models and training techniques based on neural networks have shown great potential, con- siderably outperforming traditional methods such as partial least-squares regression, cubist and random forest. In recent years, developments in sensor manufacturing have led to the availability of more portable and accessible near infrared spectrometers which have shown potential to predict several soil properties, including organic carbon. These instruments are usually part of a closed infrastructure with complex interactions with remote servers, associated data privacy issues, model development restrained by vendors, programmed obsolescence, etc. Inference "at the edge" allows the use of models directly stored in low-power con- sumption devices, providing extra portability. Combined with an open-source software ecosystem, it solves the aforementioned problems. Here, we present some of the of the challenges of using deep learning soil spectral models in low-power hardware, including techniques to reduce model size without affecting performance, improve latency and reduce power consumption.

## 4.16
## Going Deep: An assessment of artificial intelligence and deep learning techniques for image processing of soil surface and subsurface horizons.

Perseveranca Mungofa[1], Stephan Mantel[2], Jorge Samuel Mendes de Jesus[2], Sabine Grunwald[1], Arnold Schumann[1]

University of Florida, Soil, Water, and Ecosystem Sciences[1]. International Soil Reference and Information Centre (ISRIC) – World Soil Information

Traditionally, analysis and modeling of soil properties has received considerable attention, mostly accounting for the surface horizon due to the agricultural importance of soils. However, a number of research studies have showcased the importance of soil information at down to 500 cm for soil carbon accounting and other relevant variables for agricultural and environmental sustainability. Known challenges in soil assessment are the adequate sampling density and frequency, while considering the complexity of the labor-intense standard laboratory methods.

In this research, a new approach for analysis of soil samples is proposed using digital images of soils at the profile scale at distinct horizon and profile depths. Artificial intelligence (AI) and deep learning (DL) algorithms were used as a method for feature extraction and data analysis. Two datasets of soil images were used: low resolution RGB imagery acquired from various photo galleries of the United States Department of Agriculture (USDA) and the internet (N = 397); and high resolution RGB imagery from the International Soil Reference and Information Centre (ISRIC) World Soil Reference Monolith Collection (N=853). Samples were analyzed using AI DL algorithms, including image generative models for image data augmentation (Generative Adversarial Networks (GANs), and Diffusion-based models); image segmentation models for horizon and feature segmentation; convolutional neural networks for image classification and prediction of soil properties with linear regression (CNN-regression).

## 4.17
## Testing different combinations of proximal soil sensors for high-resolution mapping of key soil fertility properties

Jonas Schmidinger Jonas.Schmidinger@uni-osnabrueck.de Universität Osnabrück

Up to now, there is no single proximal soil sensor (PSS) which is capable of predicting multiple key soil properties with high accuracy. Fusing multiple PSSs can address this issue, as different sensors ideally provide complementary information about soil characteristics. However, not all PSSs benefit from synergetic effects when fused. The accuracy of predictions may even deteriorate when the data from the combined PSSs mainly consists of redundant information or when the PSSs fail to generate predictors that are meaningfully related to the target soil property. For this reason, we aim to identify robust and capable PSS combinations for predicting multiple key soil properties. In a case study, eight state-of-the- art PSSs were deployed along a two-hectare transect of an agricultural field: near-infrared spectroscopy, laser-induced breakdown spectroscopy, Raman spectroscopy, apparent electrical conductivity, gamma- ray spectroscopy, capacitive soil moisture, ionselective electrodes for pH and X-ray fluorescence spectroscopy. Using a cubist model, we exhaustively tested the predictive capability of every possible PSS combination when fusing one to five different PSSs. Additionally, we investigated how fusing bare soil multi-spectral remote sensing (RS) data from Sentinel-2 affects different PSS combinations. The target soil properties were pH, soil moisture content, soil organic carbon and plant available potassium, magnesium as well as phosphorus. In an analysis using the root-mean-square error (RMSE) and Nash– Sutcliffe model efficiency coefficient (MEC) for evaluation, we observed that fusing more PSSs considerably increased prediction accuracy over the given set of target soil properties. The magnitude of the improvement varied among the different target soil properties. While some PSSs were moderately related to many different soil properties, other PSSs exhibited strong sensitivity to one specific soil property. Overall, there was no PSS combination that clearly outperformed the other set of combinations as various sets of PSSs led to rather accurate predictions. We also observed that the addition of RS mostly had a positive influence when using only a single- or a few PSSs. Yet, not all PSSs benefited from synergetic effects with RS data.

## 4.18
## Evaluation of a novel, commercial, VisNIR probe for in-situ measurement of soil carbon stocks.

Jason Ackerson, Ayush Guwali, Marissa Wiseman, Chris Tolles, Kevin Meisner, Cristine Morgan.

Visible near infrared spectroscopy has been used for many years to measure soil properties including soil carbon content. While several researchers have utilized *in situ* spectroscopy for soil analysis, existing VisNIR spectrometers have several shortcomings for application in the field. Existing commercial or custom *in situ* VisNIR probes require custom foreoptics and heavy, hydraulic soil sampling machines to operate. In this study, we evaluate the effectiveness of a commercial, handheld VisNIR probe developed by Yardstick PBC. The probe overcomes many of the shortcomings of previous *in situ* VisNIR tools in that it is easily deployed by a single person without the necessity of heavy equipment and can rapidly collect high-resolution VisNIR data. This study demonstrates the viability of the Yardstick probe to measure soil carbon stocks accurately and cost-effectively providing new capability for high-resolution soil measurement and monitoring.

## 4.19
## Predicting changes in soil nitrogen and phosphorus using nitrogen/phosphorus measurement sensors and machine learning

Kyoung Jae Lim, Jae E. Yang*

Kangwon National University, Chuncheon 24341, Republic of Korea

In order to measure the soil properties contained within the soil, each component is chemically separated from the soil and measured. This method has the disadvantage of requiring time and manpower to measure nitrogen and phosphorus in the soil. Recently, various research using sensors has been conducted. However, research on sensors that measure nitrogen and phosphorus in soil is insufficient. In addition, these sensors do not display the type or unit of nitrogen and phosphorus being measured, limitation the in measuring accurate phosphorus and nitrogen values. In this study, to overcome these limitations, the sensor was calibrated using soil and nitrogen standard solutions from the test site. During the sensor calibrating process, a regression equation was calculated. Based on the regression equation, changes in nitrogen and phosphorus in the test field from August to September 2022 were analyzed. Furthermore, the previously calculated regression equation was learned through machine learning, which has recently been used worldwide in various fields such as data regression analysis, image recognition, and natural language processing.

The machine learning algorithms used for learning in this study are decision tree (DT), random forest (RF), gradient boost (GB), extreme gradient boost (XGB), deep neural network (DNN), and long short- term memory (LSTM). Changes in nitrogen and phosphorus at the test site were predicted using the developed machine learning model. The predictive results showed a high correlation, confirming that it is possible to predict nitrogen and phosphorus in the soil using sensors in the field

## 4.20
## Portable X-ray Fluorescence Spectrometry for Sensing Salinity and Sodicity in Glacial Northern Great Plains Soils.

Adam Devlin, Kristopher Osterloh, PhD, South Dakota State University, adam.devlin@sdstate.edu

Saline and sodic soils are an increasing concern across the Northern Great Plains (NGP) due to overlapping factors of climate change and land management that are~~a~~ drawing geologically derived salts to the land surface. Traditional laboratory assessments such as electrical conductivity (EC) and sodium absorption ratio (SAR) can be time consumptive and expensive.

Importantly, they do not discriminate the type of salts causing salinity or dispersion. This has led to the desire for more rapid, accurate measurement alternatives. Portable X-ray fluorescence spectrometry (PXRF) may be a viable proximal sensing alternative, as it is able to provide accurate elemental data in minutes under field or laboratory conditions and can directly quantify

salinity-associated elements like Ca, Mg, and S. PXRF paired with predictive models has proven to be useful for a diverse range of soil applications such as prediction of taxa, parent material, horizonation, texture, cation exchange capacity, fertility, contamination, and salinity. This study assessed the viability of PXRF elemental data from lab-prepped glacial till soils for predicting EC. Multiple liner regression ($R^2$ = 0.6833, RMSE = 0.5836), random forest ($R^2$ = 0.8619, RMSE = 0.3881), and cubist ($R^2$ = 0.8736, RMSE = 0.3814) models were then developed through 10-fold cross validation of a 33- element suite.

## 4.21
## Effect of soil autocorrelational properties on regression model choice for mapping soil organic carbon in hyperspectral images

S. Kabiri[1], S.M. O'Rourke[1]

[1]UCD School of Biosystems & Food Engineering, University College Dublin, Belfield, Dublin, Ireland.

Modelling and mapping soil organic carbon (SOC) content and other soil properties from high- resolution hyperspectral images presents the opportunity to study carbon distribution in the soil profile under different management scenarios. Existing studies on soil hyperspectral imaging often use advanced machine learning methods selected to capture non-linear relationships, however issues specific to soil data such as inherent autocorrelation of SOC and other soil properties are not accounted for. In this study the procedure for regression model selection for SOC modelling in soil core hyperspectral images was investigated. Nine intact 1 m soil cores and their corresponding pressed pellets were scanned by a short-wave infrared (SWIR) hyperspectral sensor, and reference SOC was measured for each 10 cm depth. Standard cross- validation and two spatial cross-validation methods were used to determine which one of three regression methods, Partial Least Squares Regression (PLSR), Gaussian Process Regression (GPR) and Neural Network Regression (NNR) was suitable for modelling soil organic carbon from hyperspectral data. All three models achieved equal performance for standard cross validation and core-out validation ($R^2$~0.95), but GPR failed ($R^2$=0) and NNR performed worse ($R^2$~0.49) than PLSR ($R^2$=0.64) for depth-out validation. This highlights that generalization of SOC models for soil cores with hyperspectral images can be impacted by autocorrelation of SOC along the depth axis. The modelling exercise was repeated to model SOC for scanned soil pellets and whilst the results remained the same for standard and core-out validation, for depth- out validation GPR failed ($R^2$=0), and NNR's performance deteriorated ($R^2$=0.32), but the performance for PLSR improved ($R^2$ = 0.78). These results show that the autocorrelation of SOC along soil depth might be captured by soil texture, and the relationship between SOC and soil texture is partially neutralized in the pelleting process. The overall results of this study demonstrate that PLSR is a superior regression technique when the autocorrelation of SOC is considered and is more likely to capture actual chemical properties in soil cores compared to GPR and NNR regression models.

## 4.22
## Application of computer vision semantic image segmentation and classification algorithms for processing of digital microscopic soil images acquired by a digital soil core sensor

Authors: Perseverança Mungofa[1], Daniel Rooney[2], Nicolas Guries[2], Woody Wallace[2], Arnold Schuman[1], and Sabine Grunwald[1]

University of Florida, Soil, Water, and Ecosystem Sciences[1]. LandScan, Davis, California[2] Optimization of sampling protocols and laboratory analysis for soil characterization is still a work in progress in the field of quantitative soil science with emerging technologies to improve *in-situ* analysis of soil properties. The recent advancements in computer vision and computing bring together a synergistic potential for image processing of soil samples and make inferences on soil physical and chemical properties from digital images.

The objective of this study was to develop deep learning machine vision application for *in-situ* imagery segmentation and classification of soil pore space and particle density for accurate estimation of soil physical properties from digital images. Digital microscopic images of soil samples were collected from a 37 acres almond grove with coarse-loamy soils in California, USA. The soil images were extracted from video frames of soil profiles to a depth of 100 cm, at a depth increment of 1 cm, with an image resolution of 1920 x 1080 pixels, and a spatial resolution of 3 microns with a field of view of 2.3 x 1.2 mm. A pretrained semantic image segmentation model – DeepLabV3+ was calibrated for 300 iterations using a total of 630 images, 80% for training and 20% for validation. The test inferences were performed on an external dataset consisting of 400 images. The input data consisted of binary segmentation masks, generated using ImageJ image processing software.

The resulting model had a training accuracy of 91% and loss of 6.2%, and a validation accuracy of 92% with validation loss of 16.7%. The model was then used to mask out the porous space from the soil images to develop a two-dimensional soil porosity index and subsequent estimation of soil physical properties such as color (CIE-L*a*b* color coordinates, and extract, hue, saturation, and value), entropy, fractal dimension, and lacunarity. The outputs from the segmented images were compared with original images, showing visible improvements of pre-processed images with inference time to automatically segment and process images of less than 100 milliseconds.

## 4.23
## Measurement of Soil Carbon Stocks In-Situ with Dual Wave Sensors

Kristopher Osterloh

South Dakota State University

Soil organic carbon (SOC) stocks are an important measurement for soil health, monitoring carbon sequestration, and soil productivity. As a dynamic soil property, SOC stocks need to be regularly monitored and can be highly variable across even small landscapes. SOC measurements are expensive and consumptive, and it remains a challenge among soil scientists to accurately measure SOC and bulk density in the field. In this study we assess the use of a dual-wave sensor fitter with a force meter to measure in- situ SOC stocks. This sensor, mounted to a hydraulic soil probe, can be used to non- destructively measure soil carbon at a high spatial resolution (2 cm depth increments). If effective and reliable, this technology will allow for an increased access to soil carbon monitoring, especially to marginalize land managers who don't have access to traditional soil carbon monitoring due to economic hurdles.

## 4.24
## Proposals for optimization in mapping electrical conductivity in sparse data through data fusion in irrigation zones: An application of spatial regression models

Hugo Rodrigues[1]; Wagner Tassinari[1]; Marcos Bacis Ceddia[1]; Gustavo Mattos Vasques[2]; Niriele Bruno Rodrigues[1]; Matheus Leal Neves[1]; Ronaldo Pereira de Oliveira[2]; Silvio Roberto Lucena Tavares[2]

[1] Universidade Federal Rural do Rio de Janeiro, Km 07, BR-465, Seropédica, Rio de Janeiro, Brazil;
[2] Embrapa Solos, Rua Jardim Botânico, n° 1024, Jardim Botânico, Rio de Janeiro, Brazil .
wtassinari@gmail.com; rodrigues.machado.hugo@gmail.com; marcosceddia@gmail.com; nirielebr@yahoo.com.br; gustavo.vasques@embrapa.br; ronald o.oliveira@embrapa.br; silvio.tavares@embrapa.br

* Corresponding author: rodrigues.machado.hugo@gmail.com

The scientific field of precision agriculture employs increasingly innovative techniques to optimize inputs, maximize profitability, and reduce environmental impact. Therefore, obtaining a high number of soil samples is a challenge to make precision agriculture viable. However, there is a *trade-off* between the amount of data needed and the time and resources spent to obtain this data compared to the accuracy of the maps produced with more or fewer points. In the present work, the research was based on a dataset of apparent electrical conductivity (ECa) containing 3906 points distributed along 26 transects with spacing between each of up to 40 meters, measured by the proximal soil sensor EM38- MK2, for a grain-producing area of 72 ha in São Paulo - Brazil. Then, a second dataset was simulated, showing only four transects and, at the end, with only 162 CEa points. We took as reference the map of CEa via ordinary kriging from the grid with 26 transects, and then the ECa was mapped from kriging with external drift and geographically weighted regression. These last two methods allow the increment of auxiliary variables, such as those obtained by remote sensors that present spatial resolution compatible with the pivot scale, such as data from the Landsat-8, Aster, and Sentinel-2 satellites, as well as ten terrain covariates derived from the Alos Palsar digital elevation model. Finally, each map was evaluated for accuracy using external validation using 400 previously selected ECa points. The three methods were submitted to a k-means clustering algorithm to define three management zones for irrigation purposes, and each management zone map was checked for its efficiency based on analysis of variance from soil texture data obtained from clay samples measured at a depth of 0 – 10 cm of soil in a grid of 72 points, i.e., 1 point per hectare. The best mapping method using sparse grids was the kriging method with external drift, and it was also the one that presented the most significant potential for defining management zones for irrigation when compared to the reference map.

## 4.25
## The benefits of using a reference sampling for mitigating the impact of legacy soil data errors on Digital Soil Mapping outputs.

Philippe Lagacherie1, Maider Arregui2 and David Fages1

LISAH, Univ. Montpellier, IRD, INRAE, Institut Agro Montpellier, France BRL Exploitation, Nimes, France

Most of the Digital Soil Mapping products now available across the globe have been developed from the deposits of punctual soil observations inherited from several decades of soil survey activity. By using these legacy data as inputs for calibrating our DSM models, we implicitly make the assumption that these legacy soil data are accurate and therefore do not affect significantly our DSM products. However, this assumption has

never been tested.

In this study, we compared, for six topsoil properties, three Digital Soil Mapping models that were calibrated from different datasets obtained at same locations: i) recent soil analyses performed by a certified soil laboratory (“reference soil data”), ii) soil analysis performed between 1955 and 1992(“legacy soil data”) and iii) soil property values obtained by applying a regression function that estimated the former from the latter (“corrected legacy soil data”).

The comparisons between the reference data and the legacy data revealed that the latter had large overall errors (MSEs between 30 % and 377% of the total variances) and large biases (absolute values of MEs between 16% and 62% of the means). However, biases could be corrected by linear functions calibrated onto the reference sampling data, which in turn reduced the overall errors (from -15% to -87
%).

The evaluations of soil predictions provided by the Digital Soil mapping models showed that the biases affecting the legacy input data were largely propagated to the soil predictions (absolute values of MEs between 18% and 62% of the means). Substantial decreases of predicted vs observed correlations were also observed for the best predicted soil properties by the reference model ($R^2$ decreases between 0.06 and 0.18). However, the soil predictions obtained from the DSM models using corrected legacy soil data were unbiased whatever the soil proper>es and exhibited only moderate decreases of predicted vs observed correlations (R2 decreases between 0 and 0.07) except for Clay ($R^2$ decreases of 0.19).

This study highlights the need to better control the quality of the legacy soil data used in Digital Soil Mapping and to account for this source of uncertainty in the DSM models.

## 4.26
## Seeking Validity in Soil Data

Stephen Roecker[1], John Hammerly [1], Andrew Brown [1], Jay Skovlin [1], Dylan Beaudette [1] [1] USDA – NRCS

Soil and Plant Science Division

The term validation has multiple meanings depending on which other words it is paired with. Often when Pedometricians hear the term, they think of model validation, which occurs toward the end of the model building process. However, another critical form of validation is data validation. Data validation in contrast to model validation ensures the training data used to build a model or analyze data meets certain assumptions. If the data does not meet the expected assumptions, it could have serious impacts on the scientific results, ranging from added uncertainty to invalid conclusions. This issue is particularly relevant when analyzing legacy data. In such cases, the analyst often did not participate in the original data collection and therefore may lack an intimate knowledge of the nuances (e.g., problems) within the data. The use of data validations can identify problematic observations or models. The types of validations that can be applied to data range from ensuring the data adheres to a particular format (e.g., pH values range from 0 to 14 or labels match one of the categories in a lookup table) to logical checks that compare related data elements for internal consistency. Many obvious checks can be automated, but others require manual inspection by domain experts. The following presentation will demonstrate several common data validations and methods to apply them.

## 4.27
## Spatial pattern evaluation in comparing digital soil maps obtained with different methods: an important addition to pointwise metrics

Gilio Genova [1], Laura Poggio[1], David Rossiter [1,2]

[1] ISRIC- World Soil information, Wageningen, the Netherlands

[2] Section of Soil & Crop Sciences, New York State College of Agriculture and Life Sciences, 233 Emerson Hall, Cornell University, Ithaca NY 14853 USA

Digital Soil Mapping (DSM) is a useful tool to generate soil properties maps. Machine Learning (ML) algorithms have been widely used in DSM. Most applications focused on using covariates values at the soil observation location, and evaluations of the map accuracy were usually pointwise. In this study, we used Convolutional Neural Networks (CNNs), an ML model capable of incorporating contextual information around each point, using covariates values as images (patch) centered around soil observations. We assessed the outcomes of both the commonly-used pointwise metrics but also the spatial patterns of the generated maps. As a control we also compared the CNN maps to maps made by Random Forest (RF). The models were trained on a global dataset comprising 110,000 topsoil samples, employing 40 environmental covariates as predictors for soil properties including pH, Soil Organic Carbon, Sand, Silt, and Clay concentrations. To evaluate spatial patterns, we checked the range and magnitude of spatial autocorrelation and computed diverse landscape metrics commonly used in landscape ecology. Our findings reveal that CNN's pointwise predictive accuracy is comparable to that of the RF model. However, the spatial patterns generated by these two models, as well as CNN with different patch sizes, exhibit significant disparities. Relying solely on pointwise statistics is not sufficient to provide a comprehensive view of a DSM model, as spatial patterns are intricately linked to soil geography and land use potential.
This study underscores the value of accounting for spatial patterns in DSM, suggesting that a consistent and reliable methodology is needed to quantify the differences in spatial patterns and interpret those differences linking with landscape and pedological information.

## 4.28
## Towards POLARIS v2: Improving Soil Properties Mapping Over the CONUS Using a New Hierarchical Geospatial Framework

Chengcheng (Emma) Xu, Nathaniel Chaney

Department of Civil and Environmental Engineering, Duke University

Accurate CONUS-wide soil properties maps are essential for hydrological modeling, climate change research, and sustainability studies. They act as vital inputs for large-scale Earth system models. Although the existing POLARIS dataset provides soil properties information across the CONUS, it exhibits relatively high uncertainties due to algorithmic weaknesses and the constrained utilization of abundant in-situ soil data. The new POLARIS framework (POLARIS v2) aims to enhance soil properties predictions by addressing these limitations. First, we incorporate novel soil covariates to improve the data-driven model's understanding of intricate relationships between soil properties and environmental covariates. Second, a hierarchical geospatial framework is implemented to address soil type imbalance, thus improving the overall model accuracy. Third, uncertainties are quantified and reduced by integrating more soil survey data through regression kriging in the process of estimating soil properties.

This work utilizes the USGS Watershed Boundary Dataset Hydrologic Unit Code 8 (HUC8) subbasins as modeling units, similar to a moving window method. It leverages the inherent similarity of environmental characteristics within each HUC8 domain before making soil classifications. Subsequently, a Hierarchical Random Forest approach is applied to classify soil types according to the USDA soil taxonomic system. This method utilizes the hierarchical structure of soil taxonomies, effectively addressing the imbalance of soil types and producing more plausible classification results. To enhance soil classification accuracy further, this framework incorporates advanced remote sensing data, including GOES 16/17 Land surface temperature. In predicting soil properties, the maps of soil classes are integrated with a harmonized soil properties database, yielding preliminary soil properties maps. Employing regression kriging and leveraging a wealth of in-situ soil observations further enhance the overall model performance. This work will provide soil physical and hydraulic properties at a 30-m resolution over the CONUS, demonstrating a significant enhancement in the predictive performance of soil properties.

## 4.29
## A metadata-focused harmonization workflow to generate high quality datasets for digital soil mapping and modeling: the Alaska Soil Data Bank project

Jelinski, N. [1], Brungard, C. [2], Grunwald, S. [3], Ainuddin, I. [1], Macander, M[4]., Ives, S. [4], D'Amore, D[5]., Nawrocki, T. [6], Frances, B. [5], Foss, J. [5], Sousa, M. [7]

[1]University of Minnesota [2] New Mexico State University [3]University of Florida, Soil, Water, and Ecosystem Sciences [4]ABR, Inc., [5]US. Forest Service, [6]University of Alaska, [7]USDA Soil Survey

Data harmonization efforts in soil science have typically relied on two distinct strategies: templating - where contributors or database curators faithfully transcribe information from the original data source into a prescribed format; and scripting - where unique code is written by database curators for each contributed dataset to convert it into harmonized form. Scripting provides flexibility, preserves the original data format, and facilitates schema updates. Traditional scripting workflows are reasonable for data compilation efforts that harmonize a limited number of large data sources but may become cumbersome when data contributions come from a large number of diverse sources. We present a new approach to the scripting workflow that relies on metadata curation and field metadata tagging as the primary method for harmonizing a wide array of data sources. Data curators append field metadata tags, which relate to a controlled vocabulary supported by a data dictionary or ontology. Subsequently, a comprehensive script mines field metadata tags to produce a harmonized dataset. Because this workflow focuses on controlling the quality and completeness of hierarchical metadata (data source, data sets, and data fields), it has the advantages of 1) preserving the original data formats, 2) ensuring deep, high quality metadata, and 3) requiring a single, flexible harmonization script instead of numerous, data source-specific scripts. This workflow is currently being implemented within the GEMS platform of the University of Minnesota Supercomputing Institute as part of the Alaska Soil Data Bank project to support digital soil mapping efforts in Alaska. However, this workflow is generalizable and easily adaptable to other data platforms and soil harmonization projects.

## 4.30
## 3-D Mapping of Soil Moisture Holding Capacity with Soil Depth Functions and Machine Learning Algorithms in a Tropical Sub-Catchment in Tanzania

Jacob Kaingo, Siza D. Tumbo, Boniface Mbilinyi, Godfrey Taulya Sokoine University of Agriculture TAMISEMI

Department of Crop and Animal Production, Mountains of the Moon University, Fort Portal, International Institute of Tropical Agriculture, Makerere University, Kampala

Soil moisture holding capacity (SMHC) is highly variable and greatly influences agricultural productivity. Machine learning algorithms and soil depth functions (SDF) offer means for accurate and detailed characterization of lateral and vertical variability of SMHC. This study examined the application of machine learning algorithms and soil depth functions for 3-dimensional mapping of SMHC. Soil samples were taken from 100 points through a stratified random sampling design at 3 depths (15 cm, 45 cm, and 75 cm). Spatial ancillary data was subjected to principal component analysis as covariates for SMHC prediction. Equal-area quadratic spline soil depth functions were fitted to model continuous vertical distribution of SMHC data. Random forests (RF) and Cubist decision trees (CBT) machine learning algorithms were trained on SDF fitted data to predict SMHC with principal components of spatial covariates as predictors. Validation was performed with mean error (ME) and root mean square error (RMSE) and R2 as indices. Computations were performed in R-software. Prediction accuracy was good with RMSEs ranging between 0.011-0.015 cm-3cm-3 and R2 between 36 - 81.4 %. Random forests had better accuracy than the CDTs. An RF-CDT ensemble improves

prediction accuracy. Observed results could be due to finer resolution of mapping covariates and learning ability of algorithms.

## 4.31
## Exploring extrapolation effects of random forest digital soil mapping: a case study in African countries

Fatemeh Hateffard1, Luc Steinbuch2, Gerard B. M. Heuvelink2,3

Department of Landscape Protection and Environmental Geography, University of Debrecen, Egyetem ter 1, H-4032 Debrecen, Hungary

Soil Geography and Landscape group, Wageningen University and Research, Wageningen, the Netherlands

ISRIC – World Soil Information, Wageningen, the Netherlands

Developing comprehensive global and national spatial soil information systems with high resolution faces the challenge of having sufficient sampling density in all regions within the area of interest. Due to a lack of extensive resources to acquire new soil samples, it is not uncommon that in practice we make use of spatial extrapolation: using soil data from one area to predict in other areas sharing similar soil-forming factors. Extrapolation across geographical space frequently leads to extrapolation in feature space, posing a significant risk to the accuracy of predictions. This study aimed to explore the extrapolation effect of the random forest algorithm to predict soil properties in four African countries. Topsoil data (0-20 cm) for organic carbon, clay and pH were extracted from the ISRIC Africa Soil Profiles database. The study comprised eight experiments in which soil data from either one or three countries were used as donor areas to make predictions for the other countries acting as recipient areas. Similarities between donor and recipient areas were identified by four measures of extrapolation, including similarity in soil types, homosoil, dissimilarity index by area of applicability (AOA) and quantile regression forest prediction interval width (QRF-PIW). The objective was to determine whether these measures generally agree with each other and to identify which one had the strongest correlation with validation metrics. The cross- validation results of the RF trained model for donor countries were satisfactory. However, when a model was extrapolated and was validated with data from the recipient area, the results were poor, highlighting extrapolation risks. A positive correlation was found between soil type similarity, homosoil, and the dissimilarity index by AOA, whereas a negative correlation was observed between the dissimilarity index by AOA and the QRF-PIW. No strong correlation was observed between the extrapolation measures and validation metrics. Soil type and homosoil showed a stronger correlation with validation metrics compared to AOA and QRF-PIW, which was disappointing given the expected higher correlation due to AOA and QRF relying on training data, covariates, and calibrated models. The results showed that further research and more case studies are needed to assess the effects of extrapolation of DSM models.

Keywords: Spatial extrapolation, DSM challenges, soil similarities, prediction accuracy

## 4.32
## National scale mapping of soil organic carbon stocks in Taiwan

Chien-Hui Syu1 *, Bo-Jiun Yang1, Budiman Minasny2, Wartini Ng2, Tsang-Sen Liu1 1Division of Agricultural Chemistry, Taiwan Agricultural Research Institute, Taichung, Taiwan

2School of Life and Environmental Sciences, The University of Sydney, New South Wales, Australia

*Corresponding email: chsyu@tari.gov.tw

Soil organic carbon (SOC) is important for nutrients retention, aggregate structure stabilization, water holding capacity and crop productivity, which also play a crucial role in ecosystem service such as climate regulation. Therefore, accurately predicting the spatial distribution of SOC is important for estimating the carbon stock in

soils.

The objectives of this study are to use digital soil mapping (DSM) to generate the spatial distribution baseline map of SOC stock in the topsoil (0-30 cm) in Taiwan

(36,000 km2), and to compare the differences in those among different landcover (paddy, upland, orchard, forest and other). About thirty thousand soil samples were used in this study, which collected from 2008 to 2020, the sampling density was higher than 0.8 sample km-2. Soil depth, organic carbon content, bulk density and coarse fragment of topsoil (0-30 cm) were used to calculate the SOC stocks, and two machine learning models cubist and random forest (RF) approach were used for modeling and mapping the SOC stocks with the help of several environmental variables. The results showed that random forest (RF) model had better prediction performance ($R2 = 0.35$, $RMSE = 1.45$), compared with cubist model ($R2 = 0.32$, RMSE

$= 1.50$), and the spatial distribution of SOC was mainly influenced by topographic and climatic variables such as mean annual temperature and elevation. The RF model predicted average SOC stock of the forest soils (4.74 kg m-2) in this study area is

higher than the other landcover types, and the total SOC stock in Taiwan is about 143 Mt. This map is the first national baseline map of SOC stock using the DSM technique for Taiwan at 20 m resolution, which provides valuable information to policymakers for evaluating the future SOC stock change.

## 4.33
## Digital mapping of Australian soil carbon stocks from inorganic carbon

Wartini Ng [A] *, Mercedes Román Dobarco [A,B], Budiman Minasny [A], Alex B. McBratney [A]

[A] School of Life and Environmental Sciences and Sydney Institute of Agriculture, Faculty of Science, The University of Sydney, New South Wales 2015, Australia

[B] BC3 Basque Centre for Climate Change, Bilbao, Bizkaia, Spain

Soil carbon stocks plays an important role in the global carbon cycle and climate change as carbon sink. In arid and semi-arid regions, like Australia, the carbon content from soil inorganic carbon could potentially dominate those of organic carbon fraction. However, currently there is a lack of clear understanding on its magnitude compared to its organic counterpart. This study aims to determine the soil inorganic carbon content and stock using quantile regression forests mixture model of classification and regression models for six global soil map depth intervals: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm 60–100 cm, and 100–200 cm at 90 m x 90 m resolution. The models utilised a compilation of environmental covariates and inorganic carbon content related data from pH (n=41,590), effervescence (n=15,105) and soil inorganic carbon measurements (n=5,776). The elevated concentration of soil inorganic carbon is consistent with the distribution of calcareous soils, and mainly accumulates in the lower depth. Despite that the carbon stock from inorganic carbon is half of those in organic carbon in the upper 1 m depth; in the lower depth interval of 1–2 m, it is three times larger. This study provides a baseline measure of soil as a carbon sink in forms of carbonates within Australia. To mitigate climate change, sustainable land management should be implemented so that the soil can remain to be carbon sink.

# 4.34
# Evaluating the Performance of a Topsoil Organic Carbon Monitoring System at Continental Scale: Regional Validation in Wallonia, Belgium

Marmar Sabetizadeh[1], Adrián Sanz[2], María Julia Yagüe Ballester[2], Nikolaos Tsakiridis[3], Nikolaos Tziolas[4], Uta Heiden[5], Pablo d'Angelo[5], Paul Karlshoefe[5], Peter Schwind[5], Rupert Muller[5], Laura Poggio[6], Giulio Genova[6], Sabine Chabrillat[7], Robert Milewski[7], Asmaa Abdelbaki[7], Asa Gholizadeh[8], Daniel Žížala[8], Eyal Ben-Dor[9], Bar Efrati[9], Bas van Wesemael[1]

[1] Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
[2] GMV - Remote Sensing and Geospatial Analytics Division, Isaac Newton 11, P.T.M. Tres Cantos, E- 28760 Madrid, Spain
[3] Laboratory of Remote Sensing, Spectroscopy, and GIS, Department of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[4] Department of Soil, Water and Ecosystem Sciences, Institute of Food and Agricultural Sciences, University of Florida, 2685 State Rd 29N, Immokalee, FL 34142, USA
[5] German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Oberpfaffenhofen, 82234 Wessling, Germany
[6] ISRIC - World Soil Information, Droevendaalsesteeg 3, 6708 PB Wageningen (Building 101), the Netherlands
[7] Section Remote Sensing, Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
[8] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamycka 129, Prague, 16500, Czech Republic
[9] The Remote Sensing Laboratory, Tel Aviv University, Zelig 10, Tel Aviv 69978, Israel

The Worldsoils monitoring system for topsoil SOC combines Earth Observation (EO) data with reference data from the European LUCAS soil archive and computational models. The system generates reflectance composites using time series satellite imagery, applying a moving time window of three years from 2018 to 2022. These composites, featuring a 50-meter spatial resolution across Europe, are then categorized into either bare soil (cropland) or permanently vegetated pixels (grassland). For areas of bare soil, SOC predictions are performed using multi-input 1-D convolutional neural networks that utilize Sentinel-2 spectral bands. Meanwhile, for vegetated areas, SOC predictions are created with a digital soil mapping method using quantile random forest and incorporating environmental predictors and spectral composites. Bare and vegetated soil SOC predictions are then combined interpolating the values at the edges.
To verify Worldsoils System output, three European national entities collaborated. In Belgium, external data for validation is sourced from regional geo-referenced SOC data. For validation, the predicted SOC value for each sample's corresponding pixel in the geo-referenced dataset is extracted. The selected samples coincide with the acquisition period for the Sentinel-2 composite. The number of validation samples exceeds10,000 for each year, showing a mean SOC content of 17.4 to 18.0 g kg$^{-1}$ and a standard deviation of 10.6 and 11.6 g kg$^{-1}$. The system shows a tendency to overestimate values above 80 g kg$^{-1}$. Performance metrics are evaluated for both croplands and grasslands by contrasting the observed SOC with predicted content. Overall, based on project's aims, the model's performance is adequate, with an $R^2$ value around 0.5 and a Ratio of Performance to Deviation (RPD) of about 1.4. The Root Mean Square Error (RMSE) is relatively high, at 7.6 to 8.4 g kg$^{-1}$, largely due to less accurate predictions for pixels with SOC contents exceeding 25 g kg$^{-}$

[1]. The bias in SOC predictions is minimal, ranging from -0.35 to 0.125 g $kg^{-1}$. The accuracy of the prediction system allows detecting the effect of regenerative agriculture in regions with similar pedo-climatic conditions. However, the monitoring period was insufficient to reveal differences in SOC content over time within the same regions.

Keywords: Digital soil mapping, Earth Observation Data, Computational Models, Sentinel-2 Imagery.

## 4.35
## Machine learning models do not provide higher accuracy models compared to ordinary kriging under high density soil observations

Chien-Hui Syu1 *, Budiman Minasny2, Bo-Jiun Yang1

1Division of Agricultural Chemistry, Taiwan Agricultural Research Institute, Taichung, Taiwan 2School of Life and Environmental Sciences, The University of Sydney, New South Wales, Australia

*Corresponding email: chsyu@tari.gov.tw

The impact of the quantity of training data on the accuracy of machine learning models to predict soil properties has been extensively examined. Research findings consistently indicate that machine learning models tend to yield optimal results when trained on a substantial volume of data, often surpassing the performance of ordinary kriging. However, these investigations have predominantly relied on pre-existing, sparsely sampled datasets. Notably, no comprehensive studies have explored the influence of high soil sample density observations (more than 1 sample/km2) on machine learning model performance, primarily due to the scarcity of real-world data in this regard. In the current study, we leveraged data from the Taiwanese soil survey, where the sample density amounted to one observation per 250 meters on a grid or approximately four observations per square kilometer. The study area was located in central Taiwan. Our data was divided into two sets: a random quarter subset for testing (n = 1389, equating to roughly one sample per square kilometer), and the remaining data (n = 5553, approximately three samples per square kilometer) designated as the training dataset. We conducted surface soil organic carbon (SOC) stock predictions at a spatial resolution of 20 meters by 20 meters, employing Random Forests, and compared the results with those obtained through Random Forests kriging and ordinary kriging.

Systematically, we downsized the training dataset from 5553 samples to 130 samples across the study area. Notably, the testing data demonstrated that a reduction in the number of samples led to an exponential decrease in the Root Mean Square Error (RMSE). While marginal distinctions were observed, it consistently emerged that ordinary kriging outperformed both Random Forests and Random Forests kriging. This outcome suggests that the density of soil observations plays a more pivotal role than the choice of machine learning models, implying that the available covariates may not suffice to capture localized soil variations adequately.

## 4.37
## Digital Mapping of Al, Fe2O3, Nb, TiO2 and W in Mineralized Laterites in the Brazilian Amazon

RODRIGUES, Niriele Bruno[1]; SILVA, Júlio Cesar Lopes da [2]; PINHEIRO, Helena Saraiva Koenow [3]; CARVALHO JUNIOR, Waldir 4

1 UFRRJ, nirielebr@ufrrj.br; 2 UFRJ, jlopes@geologia.ufrj.br; 3 UFRRJ, lenask@gmail.com;4 waldir.carvalho@embrapa.br

The combination of remote sensing and topographical data associated with machine learning (ML) models, especially for digital geological and pedological mapping, has contributed to the identification of areas with economic potential for mineral prospecting. The Amazon contains a thick crust of laterite (>200 m), where the carbonation processes of the siderite have produced a goethite/hematite crust. In this sense the goal was to

identify patterns in rock alteration types and target mineralization, especially in areas that are difficult to access. In order to achieve the goal different ML models were tested (Multivariate Adaptive Regression Spline (MARS), Radial Support Vector Machine (svmRadial) and Random Forest (RF)) to predict the spatial distribution of Al, $Fe_2O_3$, Nb, $TiO_2$ and W contents in Morro dos Seis Lagos, Brazilian Amazon. The input dataset gathers geochemical data from 341 samples (soil, sediments, and rock materials) with morphometric covariates and spectral indices from remote sensing data, obtained by combining satellite bands from Sentinel-2A, Sentinel-1A and Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER).
The most important covariates for each mineral compound and each model were selected using the Recursive Feature Elimination (RFE) algorithm. The results obtained showed better performance for the prediction of Al ($R^2$ =0.25), $Fe_2O_3$ ($R^2$ =0.36), W ($R^2$ =0.35), $TiO_2$ ($R^2$=0.15), using the RF model, while the MARS model presented better performance to predict Nb content ($R^2$ =0.10). The RFE algorithm highlighted the relevance of the covariates Elevation, Real Surface Area, LS-factor, Saga Wetness Index, Multiresolution Index of Valley Bottom Flatness (MRVBF), Topographic wetness index and Ferrous Iron. In this context, it was found that the characteristics of the local relief played a more significant role in understanding the spatial variation of mineral compounds, given the greater influence of morphometric covariates to predict the different elements and compounds.
Keywords: Pedometrics. Machine-learning; Poorly accessible areas.

## 4.38
## How can Google Earth Engine and Vis-NIR aid in the challenge of mapping alluvial soils in Tribal Nations

Marcelo Mancini[1,2]; Edwin Winzler[3]; Joshua Blackstock[3]; Phillip R. Owens[3]; Zamir Libohova[3]; David Miller[1]; Sérgio H. G. Silva[1]; Nilton Curi[1]; Amanda Ashworth[4]

[1]Department of Crop, Soil, and Environmental Sciences, Univ. of Arkansas, 465 Agriculture Building, Fayetteville, AR 72701, USA

[2]Department of Soil Science, Federal Univ. of Lavras, 1001 Doutor Silvio Menicucci Ave., Lavras, Minas Gerais 37200-900, Brazil

[3]USDA-ARS, Dale Bumpers Small Farms Research Center, 6883 AR-23, Booneville, AR 72927, USA

[4]USDA-ARS, Poultry Production and Product Safety Research Unit, Univ. of Arkansas, O-303 Poultry Science Center, Fayetteville, AR 72701, USA

Alluvial soils have intricate spatial distributions stemming from continuous deposition of sediments, land use changes and management. Soils from alluvial floodplains in the Colorado River Indian Tribes (CRIT) are no exception and soil data in Tribal Nations are scarce for spatial mapping, land use and management planning. An initiative is undertaken by tribes and researchers to gather data and support Native American agrarian communities in the CRIT. Here we explore how Google Earth Engine (GEE) and Visible and Near- infrared Reflectance Spectroscopy (Vis-NIR) can be used to tackle the challenge of mapping a highly variable landscape with yet limited data (n=137). Proximal sensing is more accurate than satellite-derived reflectance (Vis-NIR) but can only offer point-data and hence requires interpolation, adding more uncertainty to results. Conversely, reflectance data from satellites cover almost all globe and are publicly available; however, they have a lower signal-to-noise ratio compared to proximally sensed reflectance.
Here we investigate the usage of remotely and proximally sensed data in predicting macronutrient content via machine learning models that were trained with: i) Sentinel-2 bands from the day of sampling (May 15, 2023); ii) pixel-based statistics (medians and inter-quantile ranges) of Sentinel-2 scenes between 2018 and 2023 (555 scenes, 9 bands) that had vegetation and clouds masked out using GEE; iii) Vis-NIR bands. Most macronutrients could not be accurately predicted by the approaches. Best results were attained for K by models trained with pixel-based statistics of masked bands (train $R^2$=0.73; test $R^2$=0.44) and for Mg using Vis-NIR bands (0.62; 0.37). The worst results came from using unmasked bands from the sampling date.

Models trained with pixel-based statistics achieved results comparable to those using Vis-NIR data. In other words, statistics calculated from masked satellite data were found to highlight geomorphological persistent patterns in the landscape, which were more correlated to the spatial distribution of macronutrients than the unmasked scene. Masking vegetation and clouds using GEE provided spatiotemporal data capable of offering performance comparable to proximal sensing. Statistics of spatiotemporal satellite data and proximal sensing can support digital soil mapping endeavors in alluvial soils and help improve crop management in Tribal Nations.

Keywords: Soil fertility; digital soil mapping; machine learning; proximal sensing; remote sensing; Sentinel- 2.

## 4.39
## Distribution of heavy metals in the soils of conterminous USA and implications for food and environmental safety

Kabindra Adhikari[1], Marcelo Mancini[2,3], Zamir Libohova[4], Joshua Blackstock[4], Edwin Winzeler[4], Phillip R. Owens[4], Sérgio H. G. Silva[3], Nilton Curi[3]

1USDA-ARS, Grassland Soil and Water Research Laboratory, Temple, TX 76502, USA

2University of Arkansas, Dept. of Crop, Soil, and Environmental Sciences, Fayetteville, AR 72701, USA

3Federal University of Lavras, Dept. of Soil Science, 37200-900 Lavras, Minas Gerais, Brazil

4USDA-ARS, Dale Bumpers Small Farms Research Center, Booneville, AR 72927, USA

Management of sites contaminated with heavy metals requires precise information on their spatial distribution. This study aimed to predict and map the distribution of Cd, Cu, Ni, Pb, and Zn across the conterminous USA using point observations, environmental variables, and Histogram-based Gradient Boosting (HGB) modeling. Nearly 9200 surficial soil observations from three data sources: the Soil Geochemistry Spatial Database (n=1150), the Geochemical and Mineralogical Survey of Soils (n=4857), and the Holmgren Dataset (n=3400), and 28 covariates representing climate, topography, soils, and environmental hot-spots were compiled. Model performance was evaluated on 20% test data using R2, ρc, and RMSE indices. Prediction uncertainty was calculated as the difference between the estimated 95% and 5% quantiles provided by HGB. The model explained up to 50% of the variance in the data with RMSE between 0.16 (Cu) and 23.4 mg kg-1 (Zn), respectively. High Pb concentrations were observed near urban areas. Peak concentrations of all metals were found in the Mississippi River Valley. Cu, Ni, and Zn concentrations were higher on the West Coast; Cd concentrations were higher in the central USA. Clay, pH, evapotranspiration, temperature, and precipitation were among the model's top five important variables. The combined use of point observations and environmental variables coupled with machine learning provided reliable predictions and updated maps of heavy metals distribution in the soils of the conterminous USA. These maps would support monitoring and policies for managing the environmental and human impacts of heavy metals. The methodology could be applied to similar areas and conditions worldwide.

Keywords: soil contamination, soil chemistry, digital soil mapping, metal pollution, prediction uncertainty

# Abstracts Wednesday 7th Feb

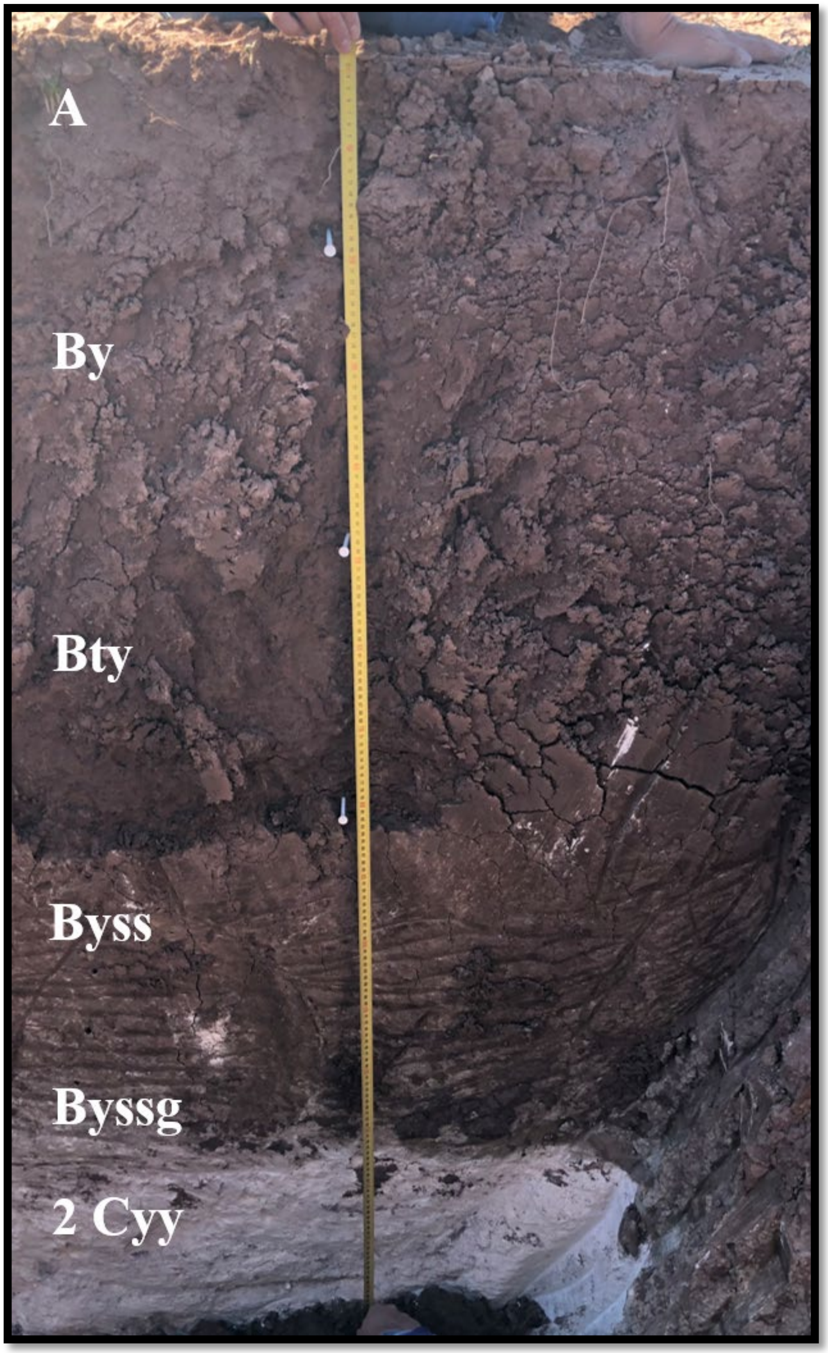



Figure 4. Fine Vertic Argigypsid in a pluvial lake bed

## 5.1
## Gaussian process: A comparison with depth-harmonised approach - a case study of mapping soil constraints

Jie Wang[*1], Patrick Filippi[1], Sebastian Haan[2], Bret Whelan[1], Thomas FA Bishop[1] [1]Sydney Institute of Agriculture, School of Life and Environmental Sciences, The University

of Sydney, Sydney, NSW 2006, Australia

[2]Sydney Informatics Hub, The University of Sydney, Sydney, NSW 2006, Australia Correspondence: jie.wang@sydney.edu.au

Soil constraints play a pivotal role in shaping potential crop yields and the overall profitability of farming endeavours. The distribution of these constraints within the soil profile exhibits a continuous and spatially varied pattern. However, effectively modelling and mapping these constraints encounter a challenge due to uneven sampling of specific depths. Precisely identifying the exact depth at which a constraint becomes evident is a critical objective. To address this concern, a spline-then-model approach can be employed, which involves the harmonization of soil profile data followed by the fitting of a spatial model. Nonetheless, this technique fails to consider the uncertainty associated with the values derived from the depth functions as it typically assumes these values to be error-free.

To surmount this limitation, this study introduces Gaussian process (GP), a method comprising two key components: a mean function and a kernel that characterizes residual variability. The primary objective is to contrast GP with the commonly used spline-then-model approach. This comparative analysis was carried out on a case study farm located in northern New South Wales, Australia, where a three- dimensional (3D) mapping of soil pH and electrical conductivity (EC) was performed. Embracing the GP methodology offers more than just point predictions; it yields an entire probability distribution of predictions. This distribution empowers the quantification of prediction uncertainty at various points. Furthermore, GP enables the estimation of average constraint values for soil volumes rather than point support, resulting in a notable 70% reduction in uncertainty on average. This capability to assess uncertainty holds particular significance in the context of decision-making and risk assessment, as it equips us with the information needed to make informed choices based on our confidence level in the predictions.
Integrating volume-based predictions enhances the precision and credibility of soil mapping, thereby enabling more effective land management strategies and resource allocation.

## 5.2
## Modelling soil organic carbon stock in space and time at multiple scales: Case study from Hungary

Gábor Szatmári[1]*, Annamária Laborczi[1], Katalin Takács[1], János Mészáros[1], András Benő[1], Zsófia Bakacsi[1], Béla Pirkó[1], Sándor Koós[1], László Pásztor[1]

[1] Institute for Soil Sciences, HUN-REN Centre for Agricultural Research, Budapest, Hungary

*E-mail: szatmari@rissac.hu

Soil organic carbon (SOC) plays a crucial role in addressing various environmental issues and challenges (e.g. climate change, land degradation, food security, water security). Therefore, spatially and, more often, spatio-temporally explicit information on SOC stock is required for a number of national and international initiatives focusing on, for example, mitigating climate change, achieving land degradation neutrality, etc. However, the spatial scale and time period for which information on SOC stock is needed may vary widely from application to application, which could pose a real challenge.

The Hungarian Soil Information and Monitoring System (SIMS), which has been in operation since 1992, collects information on SOC content every three years at 1236 monitoring sites in Hungary. The SOC stock data derived from SIMS together with spatio-temporally exhaustive environmental covariates formed the basis of this research, with the aim of building a space-time model for SOC stock that allows its prediction at different supports in space and time.

To model the space-time variability of SOC stocks, a combination of machine learning and space-time geostatistics was applied. Random forest was used to model the spatio-temporally varying trend component, while space-time geostatistics was used to model the spatio-temporally correlated stochastic component. The latter is the key to a reliable quantification of the prediction uncertainty at a support larger than the support of the observations, as it is important to take the space-time correlation of the interpolation errors into account. After building the space-time model, SOC stock was predicted at various spatial supports (e.g. point support, square blocks with different sizes) and the change in SOC stock was predicted for different time periods (e.g. 1 year, 3 year, 5 year).

The aim of this presentation is to outline the methodology we used, to highlight some methodological challenges we faced, to present the resulting predictions and maps, and finally, but importantly, to discuss the experience in a wider context.

**Acknowledgements:** This research was funded by the National Research, Development and Innovation Office (NKFIH; grant number: K-131820) and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

## 5.3
## Dealing with missingness, truncation, and censoring in multi-source data to map soil organic carbon stocks

Alessandro Samuel-Rosa[1], Taciara Zborowski Horst[1]

[1] Federal University of Technology-Paraná (UTFPR) <alessandrorosa@utfpr.edu.br>
<taciaraz@utfpr.edu.br>

MapBiomas is a network of universities, research centers, NGOs, and tech startups producing time series of land information across the Brazilian territory. Soil organic carbon (SOC) stock (0-30 cm; 30- 100 cm) is one of the themes mapped by the network. The approach consists of using point data from thousands of soil samples to train machine-learning algorithms that are later used to make predictions in space and time using hundreds of spatially exhaustive covariates. Like any initiative concerned with mapping soil properties over large territorial extensions, MapBiomas has to gather training point soil data from multiple sources. Thus, a key step is the preprocessing of these training data to achieve consistency and completeness. As the data was originally produced for uses other than mapping SOC stocks in space and time, missingness, truncation, and censoring are common features. Data on soil bulk density and volume of coarse fragments and roots are usually missing. Agricultural experiments and the like generally produce data on SOC content only for the first 10 or 20 cm of the topsoil. Several soil surveys only sample (augering) a layer of about 20 cm of the A and B horizons necessary to classify the soil up to the second level of the Brazilian classification system. Data on the total soil depth is rarely recorded, even when the soil is shallow (<100 cm). In this presentation, we will show our approach for imputing data on key soil properties for computing SOC stocks, such as bulk density and volume of coarse fragments. The approach is based on training imputation algorithms that can explicitly handle missingness even in the auxiliary variables. We will also show how natural splines and survival models are being employed to model soil-depth functions. These soil-depth functions are used to map profile soil data to a common vertical support (0-30 cm and 30-100 cm), fill gaps between sampling layers and horizons, and extend any topsoil data on SOC stock down to the lowermost depth limits of 30 and 100 cm. Various examples will be presented using real-world data obtained from the Brazilian soil data repository (SoilData, https://soildata.mapbiomas.org).

**Keywords:** Legacy data; Imputation algorithms; Natural splines; Survival models; Pedotransfer functions

## 5.4
## Leveraging Remote Sensing, Soil Properties, and AI Technologies for Nowcasting/Forecasting Soil Moisture in 3D Space and Time

*Saman Rabiei[1], Ebrahim Babaeian[1], **Sabine Grunwald**[1]*

[1]Soil, Water and Ecosystem Sciences Department, University of Florida, Gainesville, FL

Accurate real-time and future state of Soil Moisture (SM) is vital for a range of applications, including hydrologic modeling, weather forecasting, and enhancing water management in agricultural fields. However, current satellite observations such as National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) products, come with inherent limitations in providing high-resolution SM estimates across multiple soil layers. Insights are offered by SMAP, but its limitations include infrequent data updates (1-2 days) and large pixel sizes (9- and 36 km). Additionally, its products, covering surface (0-5 cm) and profile-averaged (100 cm) SM, are not applicable for decision-making in finer-scale areas. In response, a novel framework introduced in this study for estimating high-resolution SM at multiple layers of the soil profile (0-100 cm) by integrating SMAP's SM products with an array of geodata, including high- resolution soil physical properties, meteorological variables (precipitation), surface reflectance data from satellite remote sensing observations (short wave infrared and vegetation index), topographic characteristics (slope, curvature, and compound topographic index), and ground- reference measurements. The high-resolution (100 m) physical soil attributes maps provided by the Natural Resource Conservation Services (NRCS) Soil Landscapes of the United States (SOLUS) dataset, and SMAP SM product into a Convolutional Neural Network (CNN) – Long Short-Term Memory (LSTM) deep learning model. This enables the complex and non-linear relationships between SM and soil physical properties to be defined for producing high-resolution 'real-time' SM nowcasts and forecasts, revolutionizing the precision of SM estimation in multiple soil layers. In this research, the accuracy of the models is validated against ground reference data from the U.S. Climate Reference Network (CRN) and the Soil Climate Analysis Network (SCAN). Our approach is supported by an extensive, multi-source, multi-scale, dataset and cutting-edge AI techniques, providing an invaluable tool for understanding and managing SM dynamics in soil profiles which is essential for irrigation planning and precision agricultural applications.
How to cite: Rabiei, S., Babaeian, E, Grunwald, S. 2024. Leveraging Remote Sensing, Soil Properties, and Ai Technologies for Nowcasting/Forecasting Soil Moisture In 3D Space and Time. the Pedometric 2024: addressing the 10 challenges, New Mexico State University, Las Cruces, NM USA. February 5-9, 2024

## 5.5
## Fine-Resolution Near-Real-Time Soil Moisture Mapping in Tasmania through Transfer Learning

Marliana Widyastuti[1], José Padarian[1], Budiman Minasny[1], Mathew Webb[2], Darren Kidd[2]

[1] School of Life an Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia.

[2] Environment, Heritage & Land Division, Department of Natural Resources and Environment Tasmania, Prospect, Tasmania, Australia

Mapping the dynamics of soil moisture is crucial for water resource management, agriculture, and climate studies, but it poses challenges due to its spatial and temporal variability. Although current remote sensing products offer fine temporal resolution for global soil moisture, their spatial resolution remains coarse. This study's primary objective was to map daily soil moisture across Tasmania, Australia, at 80-meter resolution, with a limited training dataset. We explored three modeling strategies: models calibrated using an Australian dataset, models calibrated using the Tasmanian dataset, and a transfer learning approach that leveraged the knowledge gained from Australian models and applied it to the Tasmanian data. Our models used the Soil Moisture Active

Passive (SMAP) dataset combined with weather data, elevation maps, land cover information, and multilevel soil properties maps, to generate daily soil moisture estimates for both surface (0-30cm) and subsurface (30-60cm) layers.

Key findings from this study revealed transfer learning demonstrated significant performance improvements, reducing errors by up to 45% and increasing correlation values by 50%, compared to models trained solely using Tasmanian data. In addition, the LSTM (Long Short Term Memory) enhances the transfer learning achieving the highest overall performance, with average root mean square error (RMSE) of 0.07, and a correlation coefficient of 0.70. These fine-resolution soil moisture maps accurately captured both spatial and temporal variations, reflecting the distinct seasonal changes in Tasmania's landscape. The soil moisture models captured the drying of agricultural soils in Tasmania due to the El niño season since the beginning of 2023. The model is live, provides real-time predictions of daily soil moisture levels and weather data, offering valuable insights for land managers and farmers to optimize soil water management for crop production and environmental monitoring.

## 5.6
## Spatio-Temporal mapping of soil organic carbon stock in Brazil

Nícolas Augusto Rosin[a], José A. M. Demattê[a], Raul Roberto Poppiel[a], Jorge Tadeu Fim Rosas[a], Heidy Soledad Rodriguez-Albarracín[a] and Fernando Yutaro Makino[a]

[d]*Department of Soil Science, Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Piracicaba, São Paulo, zip code: 13418-900, Brazil. narosin@usp.br; jamdemat@usp.br; jorge.fimrosas@usp.br; hsrodrigueza@usp.br; fer.yut@usp.br*

Digital mapping of soil information is key for soil health and food security. The soil organic carbon (SOC) is a key soil attribute, having direct relation with physical, chemical, and biological properties. The soil could be sink or drain of C to atmosphere depending on management and the loss of C reduces the soils functions. We aimed to map the spatio- temporal distribution of the SOC stock in Brazilian soils for superficial layer based in remote sensing covariates and machine learning. We obtained a temporal database with soil observation and environmental covariates (static and dynamic) for Brazilian territory. The 0- 20cm layer SOC stock was mapped with high resolution (30m) from 1984 to 2023 for 5-year periods by digital soil mapping framework. Terrain attributes (static), vegetation indices (dynamic) land use and land cover (LULC) (dynamic) and a soil-vegetation image (dynamic) were used as covariates. A unique Random Forest model was calibrated and used to predict the SOC stock by use of dynamic covariates of each period. The SOC stock predictive model reached $R^2$ of 0.86, RMSE of 14.77 ton $ha^{-1}$ and RPIQ of 1.63. The most important covariates were some terrain attributes and LULC. The Brazilian soils in superficial layer had 33.42 Gt of SOC in the first period (1984-1998) and now have 32.95 Gt of SOC in the last period (2019-2023), which represents a loss of 0.47 Gt of C (-1.43%). The losses were of 0.40 Gt (- 2.36%) in Amazon, of 0.04 Gt (-0.39%) in Cerrado, of 0.03 Gt (-3.50%) in Pampa and of Mata Atlântica 0.03 Gt (-0.61%). In the other hand, the soils from Caatinga (0.01 Gt / +0.39 %) and Pantanal (0.01 Gt / +1.38%) gained SOC. The losses of SOC are associated manly with LULC changes and the gains with several factors. The use multitemporal machine learning model based in remote sensing covariates is an efficient way to access the SOC stock in the past and present. These SOC stock maps at detailed scale for the Brazilian territory can serve as subsidy for public policies for low C agriculture and climate change mitigation.

## 5.7
## Mapping of soil indicators at national scale in Lithuania using the Soil Data Cube and Artificial Intelligence-driven Earth Observation analysis

Nikiforos Samarinas[1,2,*], Nikolaos L. Tsakiridis[2], Eleni Kalopesa[2], and George C. Zalidis[1,2] Interbalkan

[1]Environment Center, 18 Loutron Str., 57200 Lagadas, Greece

[2]Laboratory of Remote Sensing, Spectroscopy, and GIS, Department of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

*smnikiforos@topo.auth.gr

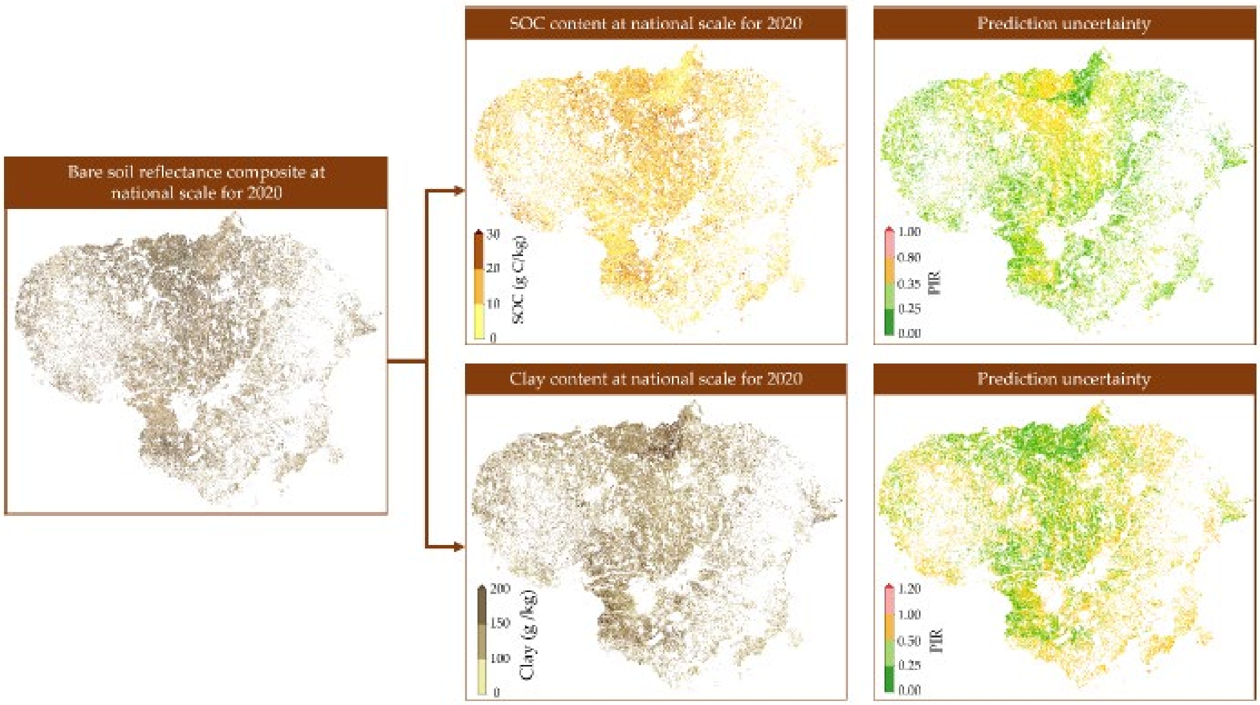
This study addresses the pressing need for evidence-based conservation recommendations in policy- making by advancing soil health monitoring through multidimensional Earth Observation-driven approaches. Existing readily available soil maps suffer from coarse spatial resolution (>200m) and outdated information, rendering them inadequate for fulfilling the requirements of both farmers and policies like the Common Agricultural Policy of the European Union.

To bridge this gap, we present a novel approach utilizing the Soil Data Cube, a custom self-hosted tool built on the open datacube initiative. This innovative methodology generates annual soil thematic maps for Lithuania's entire agricultural area, focusing on critical indicators such as exposed soil, Soil Organic Carbon (SOC), and clay content. Our approach leverages a diverse set of Earth Observation data sources, including a time series of Copernicus Sentinel-2 satellite imagery (2018 – 2022), the Land Use/Cover Area frame statistical Survey topsoil database, the European Integrated Administration and Control System, and state-of-the-art Artificial Intelligence architectures. This enables not only enhanced prediction accuracy but also a notable spatial resolution of 10 meters, allowing for precise discrimination within the parcel.

Our study evaluated five different prediction models, with the Convolutional Neural Network model emerging as the best performer, achieving an R-squared metric of 0.51 for SOC and 0.57 for clay content. Importantly, our model predictions are accompanied by prediction uncertainties based on the PIR formula, offering valuable insights for model interpretation and stability.

The application of our model and the final predictions of soil indicators relied on national scale bare soil reflectance composite layers. These were generated through a pixel-based composite approach, overlaying annual bare soil maps and using a combination of various vegetation indices and filters such as NDVI, NBR2, and ESA's scene classification layer.

The findings of this research provide significant contributions to the production of high-resolution soil thematic maps at large scales. This advancement in soil health monitoring supports more efficient and sustainable soil management, thereby facilitating evidence-based policy decisions. These insights will be invaluable to both policy-makers and the agri-food private sector in their conservation efforts.



Bare soil reflectance composite at national scale for 2020
SOC content at national scale for 2020
Prediction uncertainty
SOC (g C/kg)
30
20
10
0
PIR
1.00
0.80
0.35
0.25
0.00
Clay content at national scale for 2020
Prediction uncertainty
Clay (g /kg)
200
150
100
0
PIR
1.20
1.00
0.50
0.25
0.00

## 7.1
## Quantifying the potential and current state of European soils functions

Alexandre M.J.-C. Wadoux[1], Marko Debeljak[2], Philippe Lagacherie[1], Rachel E. Creamer[3]

[1]*LISAH, Université de Montpellier, INRAE, IRD, Institut Agro, Montpellier, France* [2]*Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia* [3]*Soil Biology Group, Wageningen University & Research, Wageningen, The Netherland*

Soils sustain a number of functions playing a key role in ecosystem functioning and providing a multitude of services to human society. While all soils are multifunctional, the supply of soil functions and their interactions differ spatially with land use type, soil characteristics, climate and management. In this presentation we will explore the quantification of the current state of soil multifunctionality, but also the potential – that is, the maximum that a soil can offer based on inherent soil indicators not affected by management practices. We quantify five functions of major importance to European soils and relevant to achieve the objectives defined by the Mission Board for Soil Health and Food: 1) primary productivity, 2) water purification and regulation, 3) carbon storage and climate regulation, 4) nutrient cycling and 5) provision of habitat for biodiversity. We built a decision support system model with a hierarchical structure. The model takes as input a simplified set of indicators related to dynamic and stable soil properties, as well as to climate and local information such as management practices, and returns qualitative aggregated attributes representing the soil functions fulfilment. Thresholds for the soil functions fulfilment are obtained by expert knowledge and vary across European environmental zones, whereas the potential is obtained through simulations for change in management practices. The model is tested on a large European topsoil dataset in cropland and grassland.

## 7.2
## Identifying hotspots of polluted forest soils in the Czech Republic: comparison of various pedometrical methods

Luboš Borůvka[1], Radim Vašát[1], Vít Šrámek[2], Kateřina Neudertová-Hellebrandová[2], Věra Fadrhonsová[2], Vincent Yaw Oppong Sarkodie[1], Lenka Pavlů[1], Václav Tejnecký[1], Radek Novotný[2]

[1]Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Czech Republic; e-mail: boruvka@af.czu.cz

[2]Forestry and Game Management Research Institute, Jíloviště-Strnady, Czech Republic

Forest floors, i.e. the superficial organic horizons of forest soils, represent an important pool of potentially toxic elements (PTE) accumulated there during long term atmospheric deposition. The PTE like Pb can be immobilized in these organic horizons and do not represent an actual environmental risk. However, they can be mobilized by organic matter decomposition for example after deforestation, which can present a potential risk. The aim of this contribution was to identify the major hotspots of forest floor pollution with PTE in the Czech Republic using various approaches and pedometrical methods and to compare their results.

We used data from the aggregated database of forest soils of the Czech Republic, containing standardized soil properties compiled from several national-scale soil surveys done in the years 2000- 2020. There are data on the content and stock of Cd, Pb and Zn in forest floor of more than 4000 locations. For assessment of polluted sites, we used reference values of PTE content and stock in forest floor set up in a previous project for several categories of forest stands defined by forest vegetation zones (governed by altitude) and tree species composition (coniferous vs. deciduous and mixed). For the pollution hotspot identification, we used several approaches: 1) indicator kriging based on exceeding the reference values in the database; 2) ordinary kriging of PTE values; 3) prediction of PTE values using random forest; the predicted values in 2) and 3) were consequently compared to the reference values. The results of the methods are evaluated and compared and their advantages and disadvantages are highlighted. In addition, the hotspots determined based on the PTE contents were compared to the hotspots determined from the PTE stocks. The results will enable the assessment of potential risk of forest soil pollution and an adjustment of forest management in the identified pollution hotspots.

Acknowledgement: This contribution was supported by the Ministry of Agriculture of the Czech Republic, project No. QK22020217, and by the Technology Agency of the Czech Republic, project No. SS06010148.

## 7.3
## 3D Soil Hydraulic Database of Hungary at 100 m resolution (HU-SoilHydroGrids)

*Brigitta Szabó[a,b], János Mészáros[a,b], Annamária Laborczi[a,b], Katalin Takács[a,b], Gábor Szatmári[a,b], Zsófia Bakacsi[a,b], András Makó[a,b], László Pásztor[a,b]*

[a]Institute for Soil Sciences, HUN-REN Centre for Agricultural Research, Hungary
[b]National Laboratory for Water Science and Water Safety, Hungary

**Keywords:** 3D soil hydraulic maps, machine learning, multi-layered gridded information, pedotransfer functions, soil hydraulic conductivity, soil water retention, van Genuchten parameters

Spatially detailed quantitative data regarding soil hydraulic properties is in high demand for a range of modeling applications. EU-SoilHydroGrids has demonstrated its utility at the European level, contributing to ecological forecasts, geological and hydrological hazard evaluations, and agri- environmental modeling, among other studies. Building on this continental precedent, a comparable but larger-scale, national 3D soil hydraulic database, known as HU-SoilHydroGrids, has been developed for Hungary with several enhancements in its elaboration process.

(i) Pedotransfer functions (PTFs) were developed using advanced machine learning techniques, both independently and as part of ensemble models.
(ii) These models were trained using the national soil hydrophysical dataset called MARTHA (acronym for Hungarian Detailed Soil Hydrophysical Database), ensuring the derivation of region-specific PTFs.
(iii) The set of predictors utilized in the PTFs was augmented by additional environmental variables with comprehensive spatial coverage, including DEM-derived geomorphometric indices, climatic parameters, OE provided surface reflectance and derived data products, LULC.
(iv) To spatially apply the resulting models, 100 m resolution information on primary soil properties was obtained from DOSoReMI.hu (Digital Optimized Soil Related Maps and Spatial Information in Hungary).
(v) Finally, based on a detailed accuracy assessment, the spatial predictions (map products) were complemented with co-layers representing the 5% and 95% quantiles.

HU-SoilHydroGrids provides nationwide information on the most frequently required soil hydraulic properties (water content at saturation, field capacity and wilting point, saturated hydraulic conductivity and van Genuchten parameters for the description of the moisture retention curve) at a spatial resolution of 100 meters, up to 2 meters soil depth for six GSM standard layers. In comparison to EU-SoilHydroGrids, the description of soil moisture retention curves and hydraulic conductivity has significantly reduced squared error in the case of HU-SoilHydroGrids.
HU-SoilHydroGrids opens up possibilities for countrywide applications and research studies to analyze environmental problems. The further development of this dataset will be directed by its integration into

environmental models and their subsequent practical application.

**Acknowledgement:** This work was carried out within the framework of the Széchenyi Plan Plus program with the support of the RRF 2.3.1 21 2022 00008 project and the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA).

## 7.4
## Concurrent Electromagnetic Induction Sensing of Magnetic Susceptibility Electrical Conductivity for the Field Delineation of Soil Drainage Class

Richard J. Heck, Farzad Shirzaditabar, Adam Gillespie, Asim Biswas, Daniel Saurete & Mike Catalano

RJH, FS, AG & AB - School of Environmental Sciences, University of Guelph, 50 Stone Road East, GUELPH, Ontario, N1G 2W1, Canada.

DS – Ontario Ministry of Agriculture, Food and Rural Affair, 1 Stone Road West, Guelph, Ontario, N1G 4Y1, Canada.

MC - Geonics Limited, 745 Meyerside Drive, Unit 8, MISSISSAUGA, Ontario, L5T1C6, Canada.

Traditional approaches to assessing soil drainage class, at the field level, have frequently considered the presence/absence of hydromorphological features, including redox concentrations, redox depletions (traditionally referred to as mottling), and reduced matrices, resulting from gleization processes, as well as surficial accumulations of organic matter. These are still fundamental to many systems of soil taxonomy. The profile and landscape distribution of redoximorphic features is ultimately linked to aquic conditions, such as endosaturation (gleyic), episaturation (stagnic) or anthric saturation. Characterization of these features typically requires an invasive observation of the soil profile, by pit or by auger, neither of which are conducive to comprehensive soilscape surveys. The relationship between soil redoximorphism and magnetic susceptibility (MS) has been widely studied. Electromagnetic induction (EMI) techniques are now often used to map the apparently electrical conductivity (EC) of soil, with interpretations focusing mostly on soil salinity and moisture dynamics. Though EMI measurements of apparent MS are common in other geosciences and archeology, their applications to pedology have been quite limited. Our current research is focused on the integration of concurrent EMI surveys, inversion modelling and spatial interpolation, of apparent MS and EC, in conjunction with soil landform quantification, to improve the delineation of drainage class in agricultural fields.

# Abstracts Thursday 8th Feb


Figure 5. Typic Calciargid formed in alluvial parent materials

## 9.1
## Uncertainty of spatial averages and totals of soil property maps

Gerard B.M. Heuvelink[1] and Alexandre M.J.-C. Wadoux[2]

[1] Department of Statistics and Actuarial Sciences, Stellenbosch University, South Africa

*[2] LISAH, Université de Montpellier, INRAE, IRD, Institut Agro, Montpellier, France*

Digital soil mappers take pride in routinely quantifying the uncertainty of maps produced, by computing quantiles of the predictive distributions and prediction intervals. Quantification of the prediction uncertainty, derived from the kriging variance in geostatistical mapping, or through methods like quantile regression forest in machine learning, is well-established. However, this uncertainty pertains to point support predictions, i.e. prediction that have the same spatial support as the observations used for model training. Yet, many users seek information about spatial averages or totals of soil properties, such as the mean clay content in a field or the total soil organic carbon stock in a region. While deriving predictions of spatial averages and totals from point predictions is straightforward, determining the associated uncertainty is challenging, due to spatial autocorrelation of prediction errors. Block kriging addresses this in geostatistical modelling, but for soil property maps created using machine learning algorithms, the solution is less obvious.

In this presentation, we propose a new model-based approach that sidesteps the numerical complexity of block kriging, making it feasible for large-scale studies employing machine learning for soil mapping. Our approach uses Monte Carlo integration to derive uncertainty of spatial averages or totals from point support prediction errors. In a first case study, we employed block kriging and show that uncertainty in predicted topsoil organic carbon in France decreases as the spatial support increases. We illustrate the broad applicability of the Monte Carlo integration method with a non-soil example in a second case study. We estimated the uncertainty of spatial aggregates from a machine learning map of above-ground biomass in Western Africa, finding it to be small due to weak spatial autocorrelation of standardized map errors.

This work introduces a scalable method that is of key importance to studies that aim to evaluate the statistical significance of predicted differences in aggregated soil properties and other environmental variables between regions or over time.

## 9.2
## Quantifying Prediction Uncertainty Based on Third Law of Geography A-Xing Zhu[1,2]

[1]Department of Geography University of Wisconsin-Madison

[2]School of Geography Nanjing Normal University

It is widely known that for a given sample set, not every location in the study area is represented by the sample set at the same level. Thus, it is unavoidable that uncertainty associated with the prediction of soil properties varies from location to location in digital soil mapping. This paper presents a theoretical basis for quantifying this varying prediction uncertainty. The basis is the premise of Third Law of Geography, geographic similarity, that is, “the more similar the geographic configurations between two locations, the more similar of the target geographic attribute”. This basis is then used to measure geographic similarity between the location of prediction and the sample set. Prediction uncertainty is then inversely related to this similarity, that is the higher the similarity of the location to the set of samples the lower the prediction uncertainty value.  Two case studies in digital soil mapping were conducted with one to illustrate the effectiveness of this idea in quantifying prediction uncertainty and the other to demonstrate the utility of this idea in assessing sample quality. The results showed that the quantified uncertainty does reflect the quality of the prediction well and is useful in improving sampling efficiency. The results also demonstrated that the idea was successful in improving sample quality.

Keywords: Geographic Similarity, Third Law of Geography, Digital Soil Mapping, Prediction Uncertainty.

## 9.4
## Exploring land use planners' preferences about visualization of digital soil mapping products for informed decision-making under uncertainty

Léa Courteille[1], Philippe Lagacherie[1], Nadia Boukhelifa[2], Evelyne Lutton[2] & Léa Tardieu[3] LISAH, Univ. Montpellier, IRD, INRAE, Institut Agro Montpellier, France

MIA-Paris-Saclay, UMR 518 AgroParisTech, INRAE, Université Paris-Saclay, Palaiseau, France

TETIS, Univ. Montpellier, INRAE, CIRAD, France **Corresponding author email:** lea.courteille@inrae.fr

Soil multifunctionality maps are required to inform planning decisions. These are generally provided to users in the form of fine-resolution raster maps, which have two major drawbacks: (1) they are usually associated with very high uncertainties and (2) the information is not represented at the scale of the spatial objects on which the decisions are based, which are much larger than a pixel (plot, water catchment) *[Vaysse et al., 2017]*. Our hypothesis is that by aggregating the pixels of a raster into homogeneous areas, we can adapt the level of spatial detail to the needs of decision-makers, while limiting uncertainty, and thus facilitate information retrieval and decision-making.

In order to test this hypothesis, we started from a soil potential multifunctionality index (SPMI) map developed for the coastal plain of the Occitanie Region *[Angelini et al., 2023]*. The initial SPMI estimations available at 25m resolution were increasingly spatially aggregated using an agglomerative spatial clustering algorithm *[Carvalho et al., 2009]*, which iteratively groups neighbouring pixels having similar values of predicted SPMI. The uncertainty of these aggregated maps was expressed either via a separate map or via a hatch pattern. More or less aggregated maps with different uncertainty visualizations were then submitted to users via an online survey. The first stage of the survey, which put the user in the shoes of a land-planner, allowed us to assess the quality of the produced maps as decision supports and to which extent users take uncertainty values into account. In the second stage, several pairs of maps were submitted to the user, and each time, they were asked to select the most intelligible one. This pairwise comparison data served as input to compute the Elo ranking of each map *[Elo et al., 1978, Langlois et al., 2022]*, which was used as a proxy for the intelligibility of the produced maps. This step helped us deduce which characteristics (in terms of uncertainty representation, level of aggregation, etc.) make a map more meaningful for end-users.

References:

Vaysse, K., Heuvelink, G. B., & Lagacherie, P. (2017). Spatial aggregation of soil property predictions in support of local land management. *Soil Use and Management*, *33*(2), 299-310.

Angelini, M. E., Heuvelink, G. M., & Lagacherie, P. (2023). A multivariate approach for mapping a soil quality index and its uncertainty in southern France. *European Journal of Soil Science*, *74*(2), e13345.

Carvalho, A. X. Y., Albuquerque, P. H. M., de Almeida Junior, G. R., & Guimaraes, R. D. (2009). Spatial hierarchical clustering. *Revista Brasileira de Biometria*, *27*(3), 411-442.

Elo, A. E., & Sloan, S. (1978). The rating of chessplayers: Past and present. *(No Title)*.

Langlois, J., Guilhaumon, F., Baletaud, F., Casajus, N., Braga, C. D. A., Fleuré, V., ... & Mouquet, N. (2022). The aesthetic value of reef fishes is globally mismatched to their conservation priorities. *Plos Biology*, *20*(6), e3001640.

## 9.5
## New evaluation criteria for digital soil mapping products from an user's point of view

Philippe Lagacherie and Léa Courteille (INRAE LISAH Montpellier)

Until now, soil mapping products (digital or otherwise) have mainly been evaluated on the basis of the agreement between predicted and actual values of soil properties on unvisited point sites ("accuracy"). At a time when soils are increasingly taken into account in territorial decisions, we argue that this evaluation from the data producer's point of view should be complemented by criteria that would take better account of the end-user's point of view.

With this in mind, we introduce three additional qualitative criteria for evaluating soil mapping products from the user's point of view: relevance, integrity and intelligibility. We will illustrate the application of these criteria to a range of currently available soil mapping products. We will review the advances in pedometrics that have made it possible to better meet these criteria. Finally, we will highlight the methodological issues that still need to be resolved to fully satisfy these user-oriented criteria.

## 9.6
## Evaluating On-Farm Functional Soil Variability: A Decision Support Framework

Jonathan J. Maynard, USDA NRCS; Dylan Beaudette, USDA NRCS; Shawn Salley, USDA NRCS; and Jeffrey Herrick, USDA ARS

Sustainable land management depends on access to soil information that can directly inform decision-making. Yet, the site-specific accuracy of soil information and thus its appropriateness for directing management actions is often unclear. In the U.S., there are two main sources of uncertainty associated with soil map information: (1) spatial uncertainty of soil classes assigned to a soil map unit, and (2) the uncertainty of soil property values for a given soil class (i.e., low-representative-high). Land managers are faced with the challenge of understanding when these sources of uncertainty matter (i.e., measurable impact on management outcomes) and, if they do, what additional sources of information (e.g., field observation) can be collected to minimize uncertainty and improve the accuracy of the soil map information used for decision making. We introduce a decision support framework to assess the importance of soil variability in land management. This framework evaluates soil functional variability using simulated soil profile realizations from SSURGO soil data, with the number of simulations for each soil type proportional to its map coverage. Each simulated soil profile realization is derived from the joint probability distribution of the SSURGO data, allowing for the creation of probabilistic soil property distributions for each soil property of interest and to propagate soil property uncertainties. To illustrate this framework, we use plant available water storage (PAWS) in the top 50 cm as our functional indicator. Our presentation will show how this framework facilitates the evaluation of site-specific soil functional variability. This framework offers a robust solution to the challenges posed by uncertainties in soil information through systematically evaluating the functional impact of these uncertainties; thus, fostering more informed and effective land management decisions.

## 9.7
## Using the LandPKS algorithm to estimate the sensitivity of ecological site identification in response to uncertainties in soil observations

Pedro Martinez, USDA-ARS, Jornada Experimental Range

Ecological site information allows land managers to make informed management decisions that ensure rangelands are used within the bounds of their land potential. To provide information on ecological sites across public lands, the Bureau of Land Management (BLM) launched, in 2011, the Assessment, Inventory, and Monitoring (AIM) terrestrial strategy with the purpose of characterizing key ecosystem processes following standard soil, vegetation, and geomorphological protocols. Although the AIM terrestrial strategy delivered a large dataset with over 50,000 monitoring plots and 120,000 soil observations, it is still unknown the level of error (sensitivity) in ecological site identification in response to uncertainties in soil and geomorphic descriptions in the field. A better understanding of the level of error in the AIM dataset would ensure that data users (e.g., ecologists, rangeland managers, soil scientists, etc.) are aware of potential limitations in the dataset and can inform training of future field data collectors. Here, we compare observer-identified ecological sites in the AIM dataset, expert-reviewed identification, and the ecological sites predicted using the Land Potential Knowledge System (LandPKS) soil matching algorithm which leverages information from national databases (e.g., SSURGO and STATSGO2) along with site-specific characteristics, such as GPS coordinates, slope, and soil observations (e.g., soil texture and rock fragment volume). Our findings provide insights into the uncertainties in ecological site identification and can be used to improve future ecological site identification by field observers in the United States.

## 9.8
## Leveraging user feedback and normalized uncertainty maps to inform future updates to national soil property maps

Travis Nauman[1], Suzann Kienast-Brown[1], Dave White[1], Colby Brungard[2], Stephen Roecker[1], Jessica Philippe[1], James Thompson[3]

[1] *USDA – NRCS Soil and Plant Science Division*

[2] *New Mexico State University*

[3] *West Virginia University*

The United States National Cooperative Soil Survey (NCSS) is producing a suite of digital soil property maps entitled Soil Landscapes of the United States (SOLUS) that will be regularly updated. The first iteration of SOLUS is 100-meter maps of 20 soil properties commonly used for modeling and soil survey interpretations. The NCSS has numerous standards for reviewing soil survey projects, but SOLUS products required a new approach for internal review to ensure transparency and quality for end users. To enable easy and consistent feedback, online Google Earth Engine applications were created with custom visualizations for all properties at seven depths. A relative prediction interval (RPI) map is also provided along with a rendering of property estimates from weighted average maps of the best available NCSS soil survey data (i.e., Gridded National Soil Survey Geographic Database [gNATSGO]) for the same property and depth. Users could click on the map and view all the property predictions and uncertainty metrics (prediction intervals and RPI) for any point. Users were prompted to provide feedback that could be general in nature or tied to a specific coordinate via a Google form. Responses ranged from very specific location-based feedback that cited uncertainty to broad statements of subjective approval or disapproval of the maps. The majority of the feedback included objective criteria that can be used to both (i) inform users of the data and (ii) inform strategies to update SOLUS to improve the quality. Initial results show that critical comments gathered in the review to correspond well to areas with high RPI values, indicating more model uncertainty and suggesting that RPI can help direct future update efforts. A synthesis of these comments and their relationship to uncertainty will be used to develop methods to improve the future versions of SOLUS. This presentation will summarize the SOLUS internal review synthesis in order to prompt a discussion around how to incorporate user feedback into digital soil maps.

## 9.9
## Landscape uncertainty for DSM at continental scale

Laura Poggio, David Rossiter, Giulio Genova, Bas Kempen, Luis Calisto, Niels Batjes

Landscape uncertainty in the context of digital soil mapping (DSM) refers to the inherent variability and uncertainty associated with soil properties across a landscape. Soil properties can vary significantly across a landscape due to natural factors such as topography, parent material, climate, and vegetation, as well as anthropogenic factors including land use and land management practices. The quality, quantity and spatial distribution of soil observations and environmental covariates can affect the level of uncertainty of soil mapping products.

DSM studies commonly assess prediction uncertainty using various approaches, including multiple simulations or quantile random forests. However, this does not encompass all the potential elements that could be used to characterize the uncertainty of a DSM product. These other elements include positional accuracy of the training points and resolution of the covariate layers (with the magnitude of this effect related to the level of spatial autocorrelation in the covariate space), area of applicability (i.e., the area in covariate space where the model learns about relationships based on the training data) and the landscape heterogeneity both in the landscape itself and in covariate space.

In this study we present initial results on how to integrate the elements mentioned above in an assessment of DSM uncertainty at continental scale. The test case is Europe with input observations with high positional accuracy and observations with a 1 km positional accuracy. We use a covariates space that covers the soil forming factors according to the scorpan model. We characterize the spatial heterogeneity of the landscape and the covariates space using commonly-used landscape metrics. The results imply some practical reflections on how to integrate all the above elements to identify regions where the confidence in the predictions is higher and the resulting uncertainty is lower.


Figure 6. Calcic Petrocalcid on Lower La Mesa surface near Las Cruces, NM

## 10.1
## Quantifying the contribution of topsoil depth to ecosystem productivity across ecosystems and climatic regions

Yakun Zhang[1]*, Ankur R. Desai[2], Jingfeng Xiao[3], Alfred E. Hartemink[1]

[1] University of Wisconsin–Madison, Department of Soil Science, FD Hole Soils Lab, 1525 Observatory Drive, Madison, WI 53706, USA

[2] University of Wisconsin–Madison, Department of Atmospheric and Oceanic Sciences, 1225 W Dayton St, Madison, WI 53706, USA

[3] Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH 03824, USA

Terrestrial ecosystem productivity is essential for global food security and promoting carbon sequestration. Understanding the controlling mechanisms of soil properties on ecosystem productivity is essential for sustaining productivity and increasing resilience under a changing climate. Here we investigate the control of topsoil depth (e.g., A horizons) on long-term ecosystem productivity. We used nationwide observations (n=2,401) of topsoil depth and multiple scaled datasets of gross primary productivity (GPP) for five ecosystems (cropland, forest, grassland, pasture, shrubland) over 36 years (1986–2021) across the conterminous USA. We first investigated the relationship between topsoil depth and GPP across five ecosystems and climatic regions using simple linear regression. We found that the topsoil depth-GPP relationship is primarily associated with water availability, which is particularly significant in arid regions under grassland, shrubland, and cropland (r=0.37, 0.32, 0.15, respectively). Then we selected 103 pairs of relatively shallow and deep topsoils while holding other variables (climate, vegetation, parent material, soil type) constant and conducted pairwise comparisons and linear mixed- effects models. Results showed that the positive control of topsoil depth on GPP occurred primarily in cropland (0.73) and shrubland (0.75). The GPP difference between deep and shallow topsoils was small and not statistically significant. Structural equation modeling was used to investigate the contributions of topsoil depth and other soil and environmental factors on GPP, and we found that the contribution of topsoil on GPP (coefficients: 0.09–0.33) was similar to that of heat (coefficients: 0.06–0.39) but less than that of water (coefficients: 0.07– 0.87). The resilience of ecosystem productivity to climate extremes was further evaluated using annual GPP and climate data over 36 years. Deeper topsoils increased stability and decreased the variability of GPP under climate extremes in most ecosystems, especially in shrubland and grassland. We conclude that the conservation of topsoil in arid regions and improvements of soil depth representation and moisture-retention mechanisms are critical for carbon-sequestration ecosystem services under a changing climate. These findings and relationships should also be included in Earth system models.

## 10.2
## "Soil's Hidden Value: Mapping Available Water Capacity as a Component of Natural Capital in Australia"

Nicolas Francos*, Alex McBratney, Damien Field, Budiman Minasny, Julio Pachon, Anilkumar Hunakunti, Jose Padarian, Wartini Ng, Sandra Evangelista, Peipei Xue, Thomas O'Donoghue

Sydney Institute of Agriculture & School of Life & Environmental Sciences, The University of Sydney, NSW 2006, Australia
*Corresponding author: nicolas.francos@mail.sydney.edu.au

Soil natural capital has been commonly considered difficult to quantify due to the many benefits that it provides. This task is challenging as many of them may overlap generating double accounting conflicts. Therefore, the soil capital value, an important dimension of soil security, may well have been undervalued.

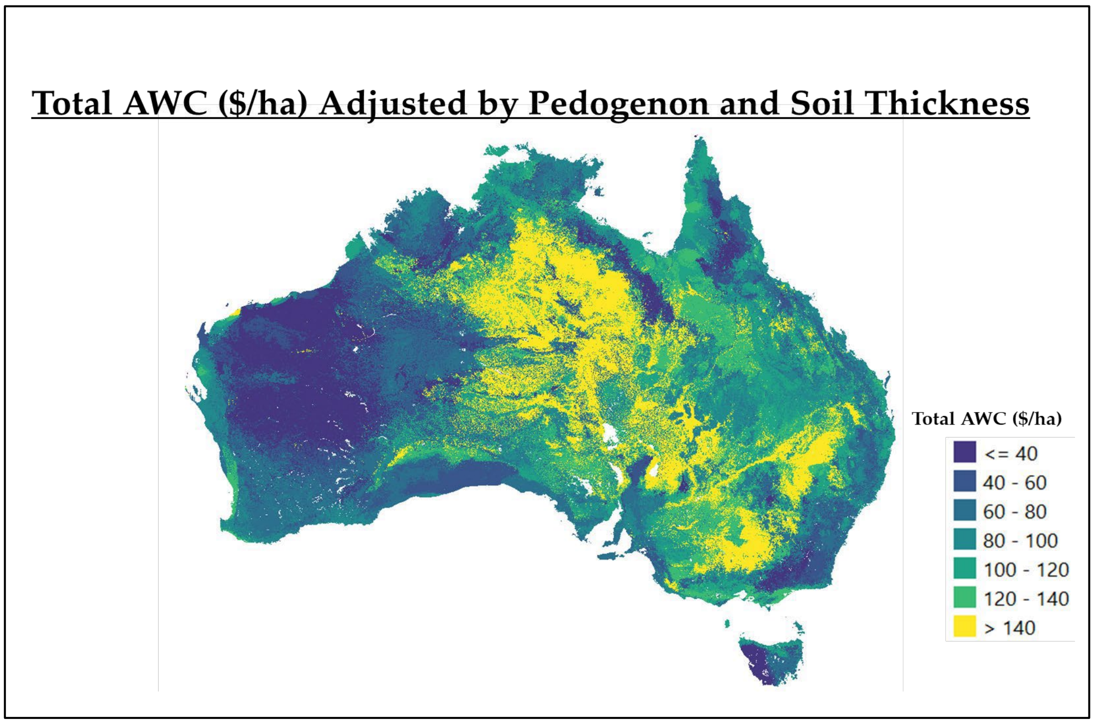
Available water capacity (AWC) is a critical property that serves as a soil security metric in all the soil security dimensions as shown in the proposal published by Evangelista et al., (2023). In that work, different soil roles were linked to AWC as it can be economically estimated. Between these roles we can mention food production, energy securing, climate balance and soil remediation. Although indirectly, AWC affects the implementation of water conservation practices and policies.

In this study we mapped the AWC for the whole of Australia on a pixel-by-pixel basis (83.3 meters) using legacy data from the Soil and Landscape of Australia (SLGA). To this end, we made careful integration of a soil thickness layer with other 6 AWC layers representing gradually increasing depths. The total AWC in megalitres (ML) units was transformed to monetary values considering the median value (80 AUD$/ML) for the 2022-2023 Southern Murray–Darling Basin which is the main irrigation water source in Australia.

These values were then associated with 1370 classes each of which are considered homogeneous prior to the European settlement in Australia (≈year 1750). This classification represents similar soil formation factors as it was based on climate, vegetation, relief, parent material, and time related covariates.

These classes are termed "pedogenons", and the mean value per class was used to represent each pedogenic zone for a more robust analysis and to reduce possible inaccuracies.

This analysis was applied to 991,394,526 pixels representing 83.33 metre resolution. Considering that the maximum soil thickness considered is 2 metres, this study estimated the total AWC soil natural capital in Australia as ≈ AUD$ 63,610,071,777 (≈USD$ 42,406,714,518). After filtering water bodies and pixels that did not match the examined layers, this analysis represents an area of 6,857,936 km2 out of
≈7,692.024 km2 which is the total area of Australia.



## 10.3
## Producing and Utilizing a Digital Twin for a G.E.M Analysis to Improve Sustainable Farming

Daniel J. Rooney et al.

A digital twin is a virtual representation of a system and its components and is a useful tool for baselining, monitoring, modeling, and decision support in a growing number of industries. The digital twin is only as powerful as the accuracy and level of its characterization. In agricultural systems, particularly those involving intensely managed crops, adequate spatial, statistical, and information resolution is critical for the characterization and subsequent analysis of both the plants and the soil environment. A unique and powerful suite of technologies (Platform for Discovery) has been developed by LandScan that produces a digital twin for agriculture and enables a G.E.M. analysis (crop performance and behavior is a function of genetics, environment, and management). A G.E.M. enables quantifiable and objective decision support for industrial-scale sustainable farming by optimizing irrigation, nutrition, and soil health while maximizing production. A site is initially mapped by a remote sensing platform that obtains high-resolution photogrammetry fused with spectral and thermal imagery collected at various times throughout the growing season. Each plant is converted into a virtual, digital representation where it can be classified, baselined, and monitored. Multiple representative locations within each crop class are identified for the targeting of the characterization of the soil environment. A tool for digitally characterizing the soil in-situ is deployed that contains 7 sensors capable of obtaining a continuous vertical profile to over 120cm in about a minute. The sensors include imagery, visible and near-infrared diffuse reflectance spectroscopy, acoustical, dielectric permittivity, electrical resistivity, tip force and sleeve friction. Because genetics and management are known, the result is a fully digital, objective, repeatable, and transferable method of creating a G.E.M. for any cropping system. Numerous ML/AI techniques are used to analyze the relationships between the G.E.M. attributes to generate practical management decisions on several almond orchards in central California. These case studies will be presented. One technique applies a multivariate frontier analysis to determine the maximum plant performance and behavior associated with one or more limiting soil factors so the grower can understand which factors can be practically managed and the opportunity cost/benefit to manage them.

## 10.4
## The challenges of using references to interpret soil health indicators

Daniel Liptzin

Soil Health Institute

In general, soil health is assessed by measuring soil health indicators. However, in order to quantify soil health, the local conditions need to be taken into account. That is, a value of soil organic carbon of 1% would be interpreted very differently depending on the location it came from. Our understanding of what soils should look like in a particular location is still largely based on Jenny's state factor model.

Human disturbance to soils has added a layer of complexity as management in agricultural systems has often changed the soils dramatically. Concepts like the genoform/phenoform and soil capability and condition are among the more recent ways to understand soil characteristics and functioning. In a soil health context, the Soil Health Institute is developing a benchmarks approach to provide insight into interpreting the current state of soils in various cropping systems. This approach depends on finding reference sites with perennial vegetation on similar soils to the cropped systems. While this approach is appropriate conceptually for evaluating soil health, there are challenges in practice. For example, in irrigated cropping systems, should the reference soils also be irrigated. What about cropping systems, like dairy forage, that continuously receive manure? Must the references also receive manure? Can the reference system have a perennial crop such as grass hay where most of the biomass is removed? Are some soil health indicators more or less sensitive to the type of reference? The choice of the appropriate reference may depend on the cropping system of interest, but it is essential to carefully consider the choice of reference for quantifying and interpreting soil health

## 10.5
## Contextualizing soil health measurements from farm to continent

Nathaniel Looker[1], Dianna Bagnall[1], Jennifer Bower[1], Jason Ackerson[1], Cristine L. S. Morgan[1] [1]Soil Health Institute

Space-for-time soil health surveys improve the relevance of soil health measurements for land managers by providing evidence of the range of measurement values that soils can achieve across pedologic contexts and management systems. Designing sampling schemes for soil health surveys presents a tradeoff between specificity and generalizability: constraining sampling to a narrow range of inherent soil properties and site characteristics reduces the variance in measurements driven by factors other than management but limits the area over which inferences are applicable. We introduce soil health sampling groups (SHSGs) as a means of balancing specificity and generalizability when conducting soil health surveys at scale.

Defined as unique combinations of soil surface particle size class and drainage class within geographic regions, SHSGs can be mapped using traditional soil surveys or digital soil models. SHSGs are easily substratified to allocate sampling effort across within-region gradients in environmental or anthropogenic factors (e.g., climate or accelerated erosion, respectively). We demonstrate how the SHSG framework facilitates prioritization of soils for sampling and identify the most important SHSGs (by area) per land use and crop across the United States and Canada. In addition to guiding soil sampling design, SHSGs can be used to communicate the importance of pedodiversity to non-academic stakeholders.

## 10.6
## Quantifying Soil Health Through an Efficient Set of Indicators and Management Indices

Minerva J. Dorantes Soil Health Institute

Providing farmers and growers a way of assessing and monitoring their progress towards improving soil health is crucial for adopting and maintaining sustainable practices. Quantifying and interpreting soil health, however, is challenging due to the absence of standardized metrics, laboratory costs, and the need for a systematic approach to compare management practices. In this study, we compared soil

health indicators in conventional and soil health production systems against perennial reference systems across Iowa. An optimized soil sampling scheme was developed to collect 250 samples from farmer participant crop fields and perennial sites. Soil organic carbon, potential carbon mineralization,

aggregate stability, and available water holding capacity were recorded for each site. Additionally,

detailed management history was collected and translated into indices reflecting soil disturbance, living roots, and soil armor. The analysis accounted for the effect of management practices, inherent soil properties, and topographic factors on soil health. Results emphasize the necessity of accounting for inherent soil variability when collecting samples and evaluating soil health. They also underscore the

significance of using management metrics to assess soil health.

## 10.7
## Scaling soil health assessment in the Golden Horseshoe region of Ontario, Canada

Jenny Bower, Cristine Morgan

Soil health assessment can be used to guide management decisions to ensure that soil maintains essential ecosystem functions. Efforts to accurately evaluate soil health are complicated by inherent soil properties and management and can be spatially limited. The goal of this work was to measure soil health using a stratified sampling design, determine the principal drivers of soil health, and test the relevance of this work in the study region and beyond. In the spring of 2023, 124 sites representing grain and oilseed farms and perennial reference sites were sampled on similar soils in the Golden Horseshoe region of Ontario, Canada. Enrollment in the program was voluntary, with specific soils and management systems prioritized. Three groups were sampled, representing high tillage frequency and no cover crops, minimum/no-till with or without cover crops, and untilled reference sites under perennial vegetation. Soil organic carbon stocks, respiration, and aggregate stability were measured at each site, and management data was collected. Significant differences in indicators were detected between all three management groups, the magnitude of which varied according to management.

Results suggest that management and texture are dominant factors influencing soil health. Management data from this study is compared with regional management information to test the applicability of our results to the total landscape. The scalability of our work beyond the province in areas with similar soil- forming factors is investigated using datasets at multiple scales. This will enable us to fine-tune our collection of soil health observations to support decision making across scales.

## 10.8
## Spatial modeling of dynamic soil properties in agricultural landscapes.

Valentina Rubio Joseph Amsili Andrew McDonald David Rossiter Harold van Es

Soil properties impact soils' ability to function and provide ecosystem services. Evaluating soil functionality (soil health; SH) involves measuring a comprehensive set of soil properties that may vary over time and space due to interactions among inherent and baseline soil properties,
current and historical land use, and management strategies. Soil inventories have traditionally focused on static properties, but soil functioning is increasingly defined by dynamic properties that are impacted by anthropogenic processes. Machine learning offers a promising option for modeling and mapping dynamic soil properties by integrating inherent and dynamic properties with remotely sensed data of its main drivers. However, the limited availability of land use and management data can pose challenges to these SH evaluations. The primary objectives of this study were to: 1) Assess the key drivers of dynamic soil characteristics through SH indicators across New York State; 2) Establish relationships between climate, inherent and baseline soil properties, and land use in relation to SH indicators; 3) Develop data-driven models for predicting and mapping dynamic soil properties at the regional scale; 4) Utilize the generated models to estimate the impacts of hypothetical regional land use change scenarios on dynamic soil properties. We evaluated a range of physical and biological properties (water holding capacity, wet aggregate stability, organic matter, soil protein, respiration, and active carbon) using 1,456 samples voluntarily submitted to the Cornell Soil Health Laboratory. To assess anthropogenic impacts, six-year USDA Crop-specific Land Cover data were used to identify land-use systems and crop and pasture frequencies, which were combined with mid and short- term NDVI values. Our approach proved to be a valuable strategy for modeling and mapping dynamic soil properties, with an average out-of-bag R2 value of 0.58. Anthropogenic processes explained approximately 42% of the variations in dynamic properties. The geospatial application of machine learning models provided valuable insights into their variability and drivers, which can support policies and management interventions. While land use changes might have minor mean effects on dynamic soil properties over a region, understanding the spatial variations in changes allows solutions to be targeted to sites where higher benefits are anticipated.

## 10.9
## Quantifying the Spatial Variability of Dynamic Soil Properties

Sage Reuter

South Dakota State University

Dynamic soil properties (DSPs) are a great tool for monitoring land-use changes and soil health metrics. The value of the information that they provide is used by the NRCS and other soil researchers to monitor and better understand the long-term effects of land management decisions. These studies capture the temporal variability of DSPs, but spatial variability is limited due to time and budged restraints. This project examines the variation with in-field sampling vs between field sampling to create a framework that will help soil scientists determine the decision-making process for sampling methods and locations in a statistically valid way. Using the framework and multi-source data integration, we can establish a more robust understanding of how single point data represents field scale management decisions.