PREDICTIVE SOIL MAPPING SEMINAR:

Summary Report

Bonnie Drozdowski and Craig Aumann, InnoTech Alberta
Chris Powter, Enviro Q&A Services

INNOTECH ALBERTA RECLAMATION TEAM
250 KARL CLARK ROAD
EDMONTON, ALBERTA T6N 1E4
CANADA

MARCH 18, 2019

EXECUTIVE SUMMARY

Traditional soil mapping has been conducted in Alberta for decades and has provided valuable information to support industrial development plans and regulatory oversight. However, recent advances in remote sensing, modelling and data processing capabilities have produced opportunities to develop soil maps cheaper, faster, more effectively, and with greater reproducibility. Predictive Soil Mapping (PSM) is a suite of tools that allow for streamlining the entire soil mapping workflow – from sample design, to efficiencies in field data collection, field sampling, identification and numbering of samples from initial site ID, through lab processing to database entry, and finally to map production.

InnoTech Alberta has initiated a project to work collaboratively with numerous users of spatial soils data to develop the Forested Region of Alberta Soil Information System (FORASIS) utilizing existing data and predictive mapping techniques. For the forested regions of Alberta, the need for FORASIS exists because:

- Fit-for-purpose maps and GIS layers of soils information are not available
- The large quantity of soil information collected across the region is not available in a standardized digital database

The long term goal is that FORASIS be used as resource tool that fills these gaps and thereby to assists industry and government in making land management decisions at regional, local and/or operational scales.

A Seminar was held at the InnoTech Alberta facility in Edmonton on March 6, 2019 and was also accessible through a webinar to invited participants. A total of 148 people had registered to participate in the Seminar.

The objective of the seminar was to develop a collective understanding of the benefits and opportunities of Predictive Soil Mapping (PSM) as they relate to Alberta. It was not intended to be a high level training session in PSM; rather, it was intended to be an introduction to highlight the benefits and advantages offered by PSM in comparison to conventional soil mapping methods.

The speakers and high-level messages of the talks were:

- Dr. Robert (Bob) MacMillian – LandMapper Environmental Solutions Inc.
  - Provided a high-level overview of predictive soil mapping and how this mapping approach relates to conventional soil mapping approaches
- Dr. Markus Walsh – Agriculture and Food Security Center, Earth Institute, Columbia University
  - Described the value created to farmers by predictively mapping soil properties across Africa
  - How proximal soil sensors are being leveraged to efficiently collect soil property data across Africa
  - How crowd-sourcing has been utilized to develop maps of landcover and land-use

- Dr. Angela Bedard-Haughn – University of Saskatchewan
  - Overviewed the predictive mapping work being done in Saskatchewan and the factors driving this work
- Mr. Xiaoyuan Geng – Canadian Soil Inventory System (CanSIS)/Agriculture and Agri-Food Canada
  - Overviewed the current perspective on soil mapping at the national level
- Mr. Chuck Bulmer – BC Forests, Lands and natural Resource Operations
  - Described how predictive mapping is being used in parts of BC
- Dr. Tom Hengl – EnvirometriX / Open GeoHub
  - Provided a high-level overview on the state-of-the art practices for developing machine learning models to predict soil properties
  - Overviewed the costs of doing this kind of mapping
  - Described the opportunities available to Alberta as a result of all the high-resolution remote sensing data now available

CITATION

This report may be cited as:

Drozdowski, B., C. Aumann and C.B. Powter, 2019. Predictive Soil Mapping Seminar: Summary Report. InnoTech Alberta, Edmonton, Alberta. 188 pp.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

InnoTech Alberta is grateful for the financial, administrative and advertising support provided by Lisa Dryden of the Environmental Services Association of Alberta (ESAA) and Sebastian Dietrich of the Alberta Institute of Agrologists Soils Network (AIA). InnoTech Alberta really appreciates the time and effort taken by the speakers to participate in the Seminar – the event would not have been a success without their participation. In lieu of speaker gifts, the group donated to the Canadian charity "Scientists in Schools" which supports development of future scientists.

The authors thank Sarah Thacker, InnoTech Alberta, for taking notes during the sessions and monitoring the online input from the webinar participants. We would also like to thank all of the individuals who have contributed to this initiative thus far including the individuals who have served on our technical advisory committee (Len Leskiw, Bob MacMillian, Tony Brierley, Bruce Walker, Chris Powter, Lynette Esak, and Larry Nikiforuk) and the individuals from the environmental consulting companies (Golder Associates, Stantec, Paragon Soil and Environmental Consulting, and Matrix Solutions Inc.) who are interested in actively participating.

ACRONYMS USED IN THIS REPORT

AAFC	Agriculture and Agri-Food Canada
AfSIS	Africa Soil Information Service
AGRASID	Agricultural Region of Alberta Soil Inventory Database
AI	Artificial Intelligence
ALOS	Advanced Land Observing Satellite
API	Application Programming Interface
BMP	Best Management Practices
CanSIS	Canadian Soil Inventory System
CDEM	Canadian Digital Elevation Model
CLI	Canada Land Inventory
CLORPT	CLimate, Organisms (including humans), Relief, Parent material or geology, Time
CSM	Conventional Soil Map / Mapper / Mapping
DEM	Digital Elevation Model
DGG	Discrete Global Grid
DMP	Dry Matter Productivity
DSM	Digital Soil Map / Mapper / Mapping
DSS	Detailed Soil Survey
EGS	Ecological Goods and Services
EML	Ensemble Machine Learning
EthioSIS	Ethiopia Soil Information System
FORASIS	Forested Region of Alberta Soil Information System
FPAR	Fraction of Photosynthetically Active Radiation
FTIR	Fourier-transform Infrared (spectroscopy)
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GSIF	Global Soil Information Facilities
HPC	High-power Computing
ISRIC	International Soil Reference and Information Centre
LHS	Latin Hypercube Sampling
LiDAR	Light Detection and Ranging

LWBSF	Lake Winnipeg Basin Stewardship Fund
MIR	Mid-infrared
MLA	Machine Learning Algorithm
MODIS	Moderate Resolution Imaging Spectroradiometer
MRVBF	Multiresolution Index of Valley Bottom Flatness
ND	No Data
NDVI	Normalized Difference Vegetation Index
O&M	Observation and Measurement
OK	Ordinary Kriging
PCA	Principal Component Analysis
PEM	Predictive Ecosystem Map / Mapper / Mapping
PPR	Prairie Pothole Region
PSM	Predictive Soil Map / Mapper / Mapping
REML	Residual Maximum Likelihood
RF	Random Forest
RFsp	Random Forest spatial
RK	Regression-kriging
RoI	Region of Interest
RPA	Remotely Piloted Aircraft
RS	Remote Sensing
RSA	Remote Sensing Application
RT	Regression Tree
SAR	Synthetic-aperture Radar
SCORPAN	Soil, Climate, Organisms, Relief, Parent material, Age, N (spatial or geographic position)
SKSIS	Saskatchewan Soil Information System
SLC	Soil Landscapes of Canada
SMAPVEX	Soil Moisture Active Passive Validation Experiment
SNMZ	Soil Nutrient Management Zone
SOC	Soil Organic Carbon
SRTM	Shuttle Radar Thirty Metres
SVAECP	Soil Variability Analysis to Enhance Crop Production
SVM	Support Vector Machines

UAV	Unmanned Aerial / Autonomous Vehicle
UK	Universal Kriging
WG-DSM	Working Group on Digital Soil Mapping
WLS	Weighted Least Squares
XRF	X-ray Fluorescence

PREDICTIVE SOIL MAPPING SEMINAR: SUMMARY REPORT

BONNIE DROZDOWSKI, CRAIG AUMANN AND CHRIS POWTER

1.0 INTRODUCTION

Traditional soil mapping has been conducted in Alberta for decades and has provided valuable information to support industrial development plans and regulatory oversight. However, recent advances in remote sensing, modelling and data processing capabilities have produced opportunities to develop soil maps cheaper, faster, more effectively, and with greater reproducibility. Predictive Soil Mapping (PSM) is a suite of tools that can streamline the entire soil mapping workflow – from sample design, to efficiencies in field data collection, field sampling, identification and numbering of samples from initial site ID, through lab processing to database entry, and finally to map production.

InnoTech Alberta has initiated a project to work collaboratively with numerous users of spatial soils data to develop the Forested Region of Alberta Soil Information System (FORASIS) utilizing existing data and predictive mapping techniques. For the forested regions of Alberta, the need for FORASIS exists because:

- Fit-for-purpose maps and GIS layers of soils information are not available
- The large quantity of soil information collected across the region is not available in a standardized digital database

The long term goal is that FORASIS be used as resource tool that fills these gaps and thereby assists industry and government in making land management decisions at regional, local and/or operational scales.

A Predictive Soil Mapping Seminar was held at the InnoTech Alberta facility in Edmonton on March 6, 2019 and was also accessible through a webinar to invited participants. A total of 148 people had registered to participate in the Seminar (see Appendix A for the list of registrants).

1.1 WORKSHOP OBJECTIVES

The objective of the seminar was to develop a collective understanding of the benefits and opportunities of Predictive Soil Mapping (PSM) as they relate to Alberta. It was not intended to be a high level training session in PSM; rather, it was intended to be an introduction to highlight the benefits and advantages offered by PSM in comparison to conventional soil mapping methods. The key questions that the organizers wanted answered were seeking answer for include:

- At a high-level, how does PSM work?
- How good are PSM maps today? While "accuracy" has certainly improved with the new statistical techniques, are things "good enough" now, or are we still waiting for additional

technical developments so that things become "good enough"? That is, is PSM ready for commercial use to support the needs of government, industry and other end users?

- What are the risks of using predictive mapping systems in decision making?
- What opportunities do predictive maps open up that we couldn't do before?
- What does PSM cost generally?
- What are others around the world doing around predictive soil mapping? Are they having successes? And how are they defining "success"?
- What is the value proposition for potential funding agencies in any further work (cost, speed, regulatory acceptance), and what are the potential downsides to doing the work (e.g., a limited talent pool to carry out the work in a PSM world, etc.)?
- What will convince regulators to shift from status quo to a new way of doing things (accuracy, reproducibility, within the current range of natural variability they already accept, previously adopted in other fields they are familiar with, etc.)?
- What will the shift to PSM do to consulting business models, how can they defend the PSM results (to clients, regulators, and stakeholders), and what training / expertise will they need to hire?

1.2 Seminar and Report Structure

The Seminar consisted of six presentations given by five invited speakers, with opportunities to ask questions (see Appendix B for the Agenda and Appendix C for the presentation materials). The speakers included:

- Dr. Angela Bedard-Haughn – University of Saskatchewan (UofS)
  - Associate dean of research and graduate studies and professor at the U of S focusing on applied pedology. Angela is a key collaborator in the Saskatchewan Soil Information System (SKSIS) project. She provided an overview of the project and some lessons learned.
- Dr. Chuck Bulmer – BC Forests, Lands and Natural Resource Operations
  - Soil scientist with 25 years of experience evaluating rehabilitation efforts on sites disturbed by forestry and oil and gas development. His work has led to improved methods for evaluating soil compaction status and implications for site productivity. Recent work has included high resolution air photos to support soil conservation and monitoring, as well as improving BC's digital soil inventories using GIS technology and modeling to enhance the use of digital soil datasets in resource planning and operations.
- Mr. Xiaoyuan Geng – Canadian Soil Inventory System (CanSIS)/Agriculture and Agri-Food Canada
  - Scientist whom has been conducting R&D in the field of PSM for nearly a decade and is the lead on Soil Landscape Canada data development and renewal. He provided an update on national PSM activities.
- Dr. Tom Hengl – EnvirometriX / Open GeoHub
  - Over 20 years of experience as an environmental modeler, predictive soil mapper, data scientist and spatial analyst. Tom has published more than 50 journal articles and several textbooks in the fields of geo-information science, PSM and spatial statistics.

He currently leads production of the Global Open Land Data System LandGIS (https://landgis.opengeohub.org), which includes soil properties and classes. Tom reviewed the status of technologies for PSM and suggested possible development directions in Alberta.

- Dr. Robert (Bob) MacMillan – LandMapper Environmental Solutions Inc.
    - A retired environmental consultant with over 40 years of experience in creating, packaging, delivering and using environmental information on soils, ecosystems, landforms and hydrology. Bob spent 20 years working in public sector research and soil survey with Agriculture and Agri-Food Canada and the Alberta Research Council and then a second 20 years as a consultant offering services in predictive soil and ecological mapping. Since retiring, Bob has remained an active supporter, promoter, advocate, mentor and technical contributor to various projects to advance the science and technology of mapping soils and other ecosystem components.
- Dr. Markus Walsh – Agriculture and Food Security Center, Earth Institute, Columbia University
    - Over 25 years of experience in ecosystems and landscape ecology research in Africa. For the last decade he has focused on developing operational tools for diagnosis, mapping and monitoring with an emphasis on the application of IT and data science. Markus provided insights into the lessons learned from initiating a soil inventory system in Africa and the importance of collecting new data appropriately and efficiently.

The following sections provide the highlights of each presentation and a summary of the questions posed by the participants.

2.0 PRESENTATIONS AND DISCUSSIONS

2.1 SHIFTING FROM TRADITIONAL TO PREDICTIVE SOIL MAPPING (PSM) – DR. ROBERT (BOB) MACMILLAN

The Universal Model of Spatial Variation (close things are similar and far things are different) underpins PSM. The Universal Model of Soil Spatial Variation has three parts:

- Deterministic – can be captured by conceptual model (e.g., expert knowledge, or existing soil maps) or statistical models which are data-driven (e.g., elevation or other soil forming factors) and objective and provide a parameterization of these deterministic components.
- Stochastic – has spatial structure; analyses effect of distance by creating a variogram and interpolating via kriging (traditional mappers do this manually by drawing lines around similar soils or values)
- Error (noise – unstructured so can't interpolate or predict it).

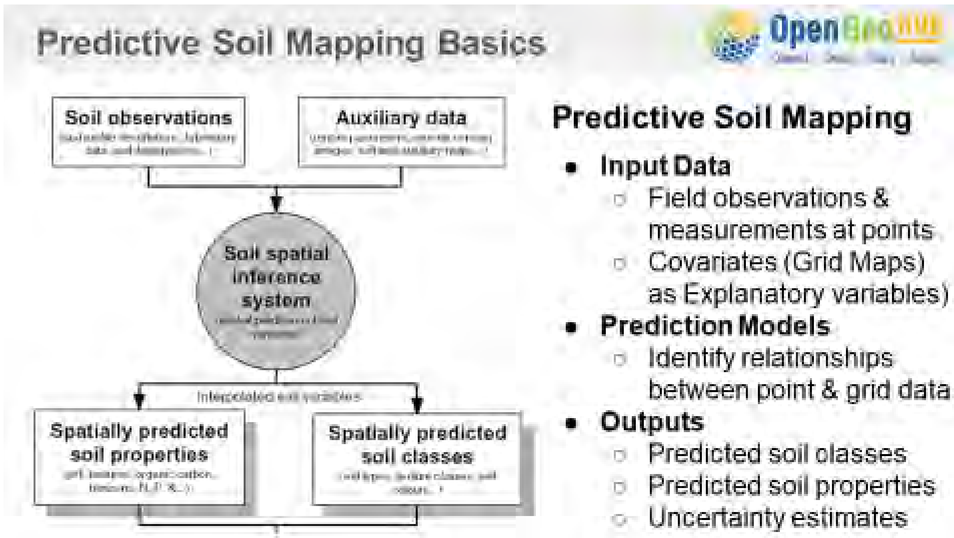
Predictive models range from linear regressions (simplest), non-linear regression, trees and random forest or other machine learning models (most complex). The output from machine learning models represents the mean of many models (ensemble).

It is important to remember that different environmental properties operate at different scales and therefore require data at different scales; these data must be brought to a common scale to be used in the models. There are different methods of scaling covariates, each with their own advantages and disadvantages.

Predictive soil mapping involves three components: inputs data, prediction models, and outputs. The key differences between conventional maps and predictive maps are shown in Table 1.

Table 1. Difference between conventional and predictive maps.

Conventional Map	Predictive Map
Based on expert judgement Subjective No two alike	Data driven Objective Reproducible
Static maps Very difficult to map changes	Minimize error Best fit to available data Multiple runs
Can't update easily One and done Have to return to field and redo to update	Easily updated Get new data Rerun models Get new maps



Question Period

Q: How do kriging maps work with cross validation, etc.?

A: Better in terms of spatial detail. One camp wants model-based interpretation and insight, while the other is only interested in the accuracy of the predictions. Machine learning can accomplish spatial kriging, so you don't need the regression kriging per se. Relatively easy to do cross validation with machine learning.

Q: Expand on Gaussian pyramid, what does it mean?

A: Take variables that operate across different spatial scales (e.g., climate, digital elevation model, coarse and fine satellite imagery) and resample these variables to the finest resolution. Machine learning models can determine which variables at which resolution work the best for the predictions one is trying to make.

2.2 LESSONS LEARNED: WHAT'S BEEN TRIED; PROBLEMS AND SOLUTIONS FOR UTILIZING PREDICTIVE SOIL MAPPING – DR. MARKUS WALSH

A workflow overview of the African Soil Information System (i_4^A Information for Agriculture) was provided along with an overview of the need for predictive mapping systems, lessons learned regarding methods, and processes implemented to increase efficiencies and reduce costs. Dr. Walsh provided the context for where they are currently monitoring various parameters within the photosynthetically active "region of interest" (RoI) in Africa to develop a process-level understanding about agricultural nutrient cycles at national and continental scales.

The AfSIS Information for Agriculture (i_4^A) workflow (which is a process for PSM) consists of the following steps:

1. Determining what information products end-users actually need. Revisit this step frequently!

2. Rapidly assess agricultural system distributions in a region of interest (RoI) with GeoSurvey. [GeoSurvey is a tool provided online at https://geosurvey.qed.ai/about/]
3. Apply spatially and temporally balanced sampling to identify field survey and experimental locations
4. Use MobileSurvey (which operates on hand-held phones) to log field observations, experiments and the associated physical samples. This creates a universal unique ID which allows sampling tracking throughout the entire process
5. Analyze all rock, soil, plant and livestock biomarker samples with high throughput lab methods. There are currently 17 spectral labs across Africa
6. Predict and monitor system state variables with machine learning and geostatistical models

Steps 1 to 5 apply to new data being collected while steps 1 and 6 apply to all samples (legacy data and new data). Several examples of applications were provided as well as links to access data and additional information. One excellent example of the above process is the Soil Nutrient Management Zones which provide fertilizer recommendations based on the maps generated. This information was then combined with nutrients measured from plant tissue and a comparison was made for nutrients in soils to nutrients in plants to assess variability in nutrient uptake based on the soil variability.

Strengths of the i_4^A system include: field data recorders (phone, tablet) for simple data entry; single QR coded samples that can be tracked from field to lab to database to map; simple tool to select regions of interest for sampling based on presence/absence information gleaned from remote sensing images; use of mid-IR and XRF spectrometers (measuring elemental profile of sodium to uranium on periodic table) – so they can obtain nutrients and soil pollutants, heavy metals, and soil geochemistry quickly and cheaply (a few dollars per sample). Hand-held devices used in the field can obtain similar information and are calibrated based lab-analyzed data. All data is publicly accessible.

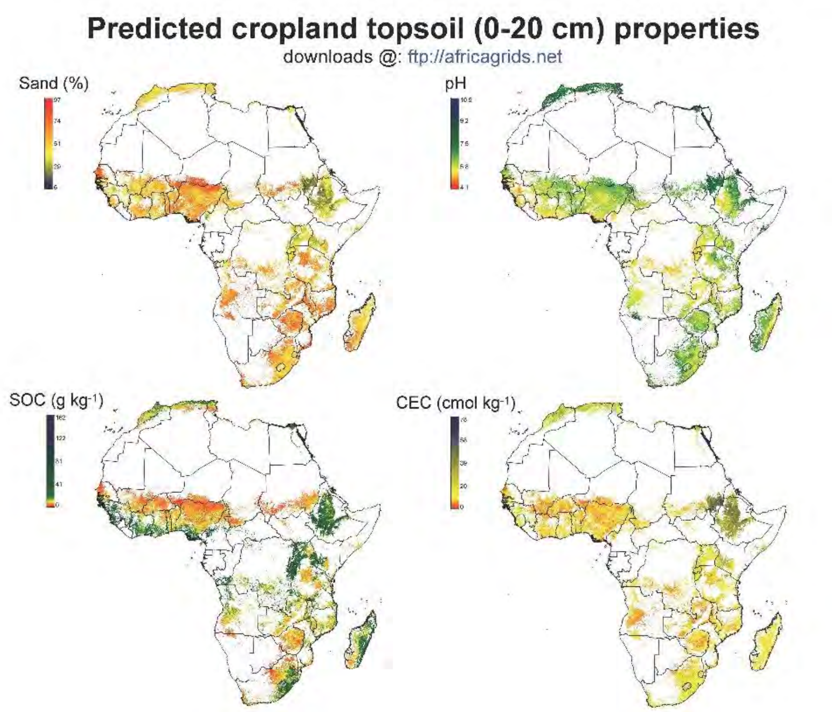
Challenges in monitoring and producing maps for large, remote areas such as many areas in Africa include: site accessibility, field expenses, power sources for labs, variation among reference laboratories in analytical methods and quality control, consistency of measurement methods over several years to decades, and geo-referencing for repeat sampling.

The products produced are not static maps, but really a monitoring system tool that allows them to generate predictions about the ecosystems they work over time as more data is collected, and forecast where changes are likely to happen.

Some key messages:

- It is important to collect new data to continue to input into the models so can monitor changes with time.
- Unique sample ID's assigned to samples right in the field decrease error substantially.
- Laboratory analysis methods and QA/QC need to be consistent. Costs can be reduced substantially be eliminating the need for wet chemistry analysis (or reducing the quantity of samples requiring it).

- Publicly funded data collection should mean that this data is publicly available (point observations, grids and models).



Question Period

Q: When using mobile field data collection, do you collect anything on paper? Ever lose data? If I want to do it, how can I avoid pitfalls?

A: Yes, can lose data. But, it is critical to collect data electronically. Like to be in and out of a site in <10 minutes to collect 20 cm deep soil sample; this would make paper collection difficult.

Q: How sensitive and how practical is handheld equipment? Concerns with colour, drying, homogenization – does that happen as well in the field?

A: Maybe not for all areas, but very promising. People working on calibrations to adjust for moisture content, homogenization issues, etc. MIR is the really promising technology.

Q: Things change quickly in Africa. In 5-10 years, cropland use could have changed (i.e., due to development). How do you capture these changes and prepare for these unpredicted changes?

A: We cannot remap all of a country or the entire continent every year. Need new GeoSurvey data and updated high resolution satellite information to investigate where land cover changes are happening most rapidly. Working with a group called Radiant Earth (specialize in providing scientists with high res, high quality, time stamped, satellite info) to try to answer these questions.

Q: Focus is monitoring crops, predict yields, crop nutrients. We use spectroscopy now. We could just use soil spectroscopy to predict yields and skip all the soil science. Thoughts?

A: Things should proceed in parallel. We want to understand soils, but also make the best crop predictions possible. We need both.

2.3 SASKATCHEWAN SOIL INFORMATION SYSTEM (SKSIS) – DR. ANGELA BEDARD-HAUGHN

Predictive soil mapping has tremendous potential for enhancing the quality of soil information everywhere, however, there is a need to identify methods that allow mappers to work with (and integrate) the strengths of legacy soil data, overcome the challenges of large spaces/coarse datasets, and create opportunities for soil maps to enhance land management. Saskatchewan Soil Information System (SKSIS) is the platform to make the data available (similar to Alberta's AGRASID).

Challenges have been identified with converting legacy data into raster format for current applications due to issues around format and scale. Predictive soil mapping provides the opportunity to utilize the data using disaggregation and machine learning techniques to remap soils at higher resolutions. Some key considerations in PSM is the quality and quantity of the input data. In general, the more data that goes into the model, the better the information that comes out in terms of the uncertainty associated with the prediction. Predictive soil mapping builds on the key soil forming factors described by Jenny (1994) and Ellis (1938 Soils of Manitoba). S = f(cl, o, r, p, t, h, gw). Due to the inclusion of the influence of ground water, Dr. Bedard-Haughn prefers to refer to Joe Ellis' soil forming factors which determine soil type as:

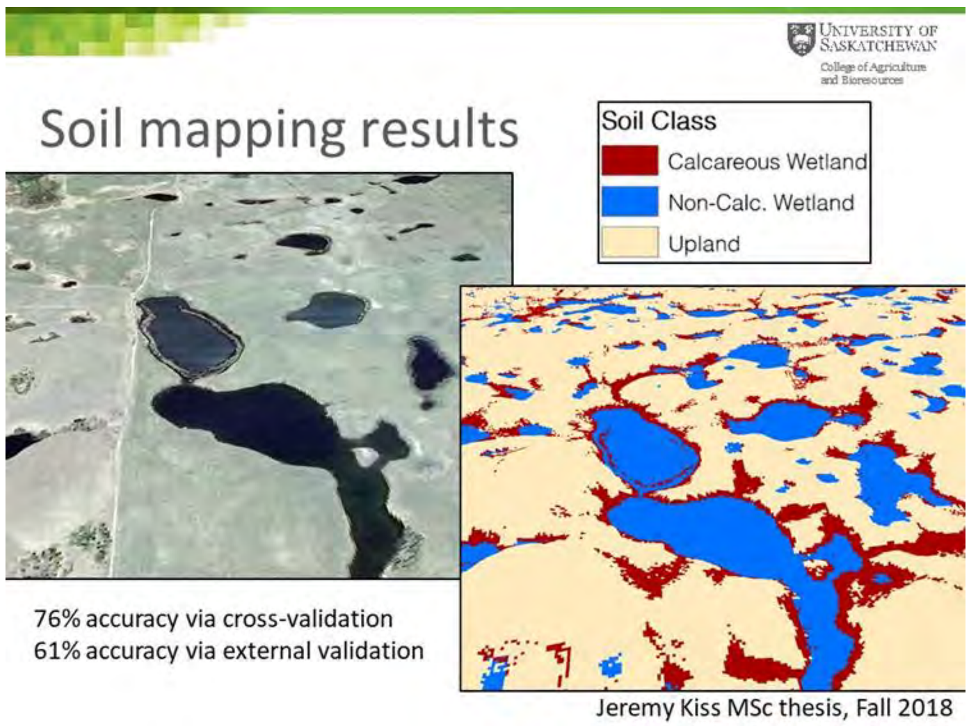
1. The climate, or the temperature and moisture within the soil;
2. The vegetation, which determines the type of organic matter added to the soil;
3. The parent material, or the geological deposits which determine the minerals on which the soil is formed, and in turn affect the texture, the water retention capacity, and the mineral reserve;
4. The position in which the soil is found in relationship to the topography;
5. The presence or absence of ground water within the soil profile;
6. The age or length of time the soil has been under the influence of its environment;
7. In the case of cultivated soils – the modifying effects of culture or the work of humans.

Conventional soil mapping was knowledge driven which requires one to understand the landscape and the conditions within which a soil has formed. To quantitatively predict soils with a data driven approach (as is the case with predictive soil models) it is essential to have high quality digital elevation models and input layers. It is important to recognize that just because a variable can be included doesn't automatically mean it should – focus on the covariates/predictor variables that are really necessary to interpret the output. It is also very important to know the limitations of the dataset.

Despite there being more soil information collected and precisely georeferenced via GPS than at any other point in history, predictive mappers struggle to have enough point data to carry out large-scale mapping efforts; still need to do a fair bit of field work to train the computer, so critics ask what the PSM advantage actually is. Need to be ready to address this question.

Key Messages

- Their approach to PSM is to build on the strengths, overcome the challenges, and take advantage of opportunities:
  - Strengths: Legacy data, conceptual model, computing power and creativity
  - Challenges: Spatial data resolution, obstacles to data sharing
  - Opportunities: Use emerging technologies and data sharing to enhance legacy data
- Errors within the DEM or co-variate data can cause compounded errors within predictive soil maps, thus QA/QC is essential
- More [quality] data that can be used in the model, the better the outputs
- Essential to identify WHY and WHAT. What is the desired outcome and why do you need/want to create predictive soil maps



Question Period

Q: Challenges with data sharing – any ideas for working with private entities sitting on a lot of soils data?

A: Having conversations with soil testing and agronomy companies currently. People fear how data might be used, trust issues. Soil testing folks most open to census based approach, where data shared more broadly in an aggregated form (i.e., in SKSIS), but primary, individual data would not be available to everyone. It's about providing the security people want before they will share the data. Agronomy people: variability in peoples' willingness to share. Comes down to the purpose of the data.

Comment: Data shared in aggregate could work in Alberta, through an AGRASID-type platform. Contaminants probably can't be shared or would only be available to certain users.

Comment: Issue of not having fine scale data. There is an opportunity in not having fine scale data. We all talk about bottom up mapping (start at fine scale and move up), but why not top down (i.e., from coarse to finer scales)? There are advantages in top down. And it gives context (upland ponds versus lowland ponds). Don't be overly committed to idea that best data is finest data.

A: Yes, in landscapes where our conceptual models are not very refined or we don't know a lot about the land. Agricultural regions: scale focused on the users that want the info and coarser info is not what they want/need. From top down, finer features can get missed. Should consider: why are you doing soil mapping and what is the desired outcome? That determines what data you are putting in. As their work is with Ag people, focus more on fine scale because that better meets end-use needs of clients.

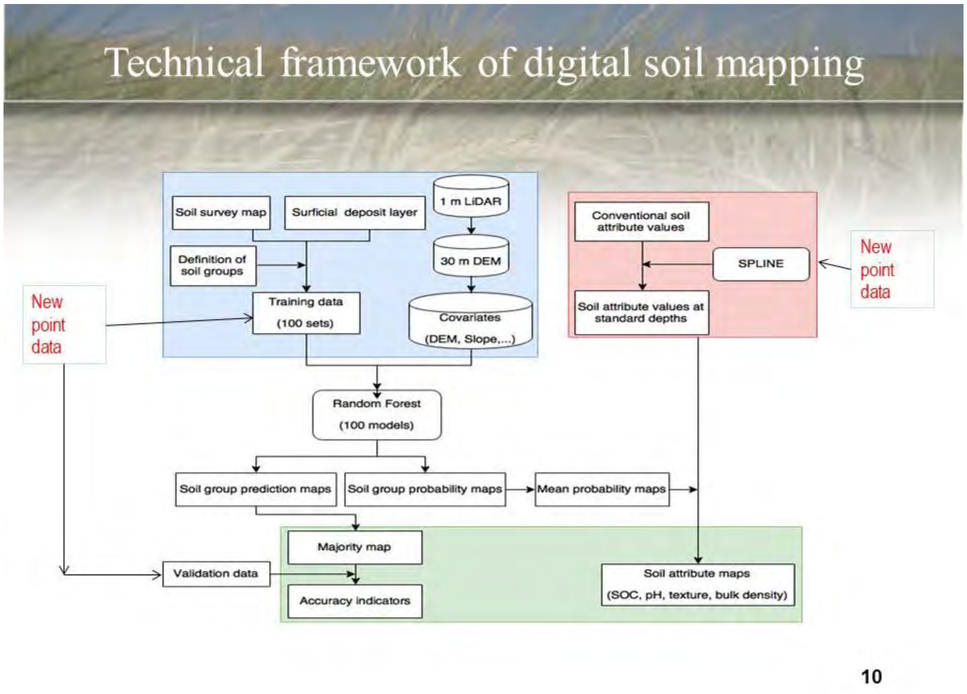
2.4 PREDICTIVE SOIL MAPPING — NATIONAL PERSPECTIVE – MR. XIOYUAN GENG

There are international, national and regional drivers for federal involvement in PSM. Soil and landscape data can be useful for a number of activities, including: Infrastructure, environment and watershed management, soil management, risk management; best management practices research, development and evaluation; precision agriculture, biodiversity etc.; and environmental goods and services measurement and assessment. Soil changes and information renewal is critical.

From Agriculture and Agri-Food Canada (AAFC)'s perspective, digital soil mapping is a cost-effective way to develop soil data, particularly at a higher resolution. AAFC is working towards a global and national database to develop higher resolution regional digital soil maps for a variety of applications, examples of which were provided. Particularly useful for land management decisions. Primarily focused on the agricultural regions of Canada with the intention of linking to precision agriculture, phenomics, resiliency metrics and soil microbiome initiatives. They will be advancing the national soil database structure by adding business driven attributes including soil total phosphorus and soil metagenomics.

Key Messages

- Data structure at different scales and resolutions causes different challenges in terms of its use in large national inventories
- Legacy soil data may need to be re-grouped to be classified by "drainage, slope, catena, etc." rather than "names"
- Need to be clear about what the desired outcome is
- There is an optimal number of samples required for the outcomes you are looking for
- We need to maximize the use of legacy soil data (by disaggregating, for example eco-district by eco-district, into vector based data)



2.5 PREDICTIVE MAPPING IN BEAUTIFUL BRITISH COLUMBIA – DR. CHUCK BULMER

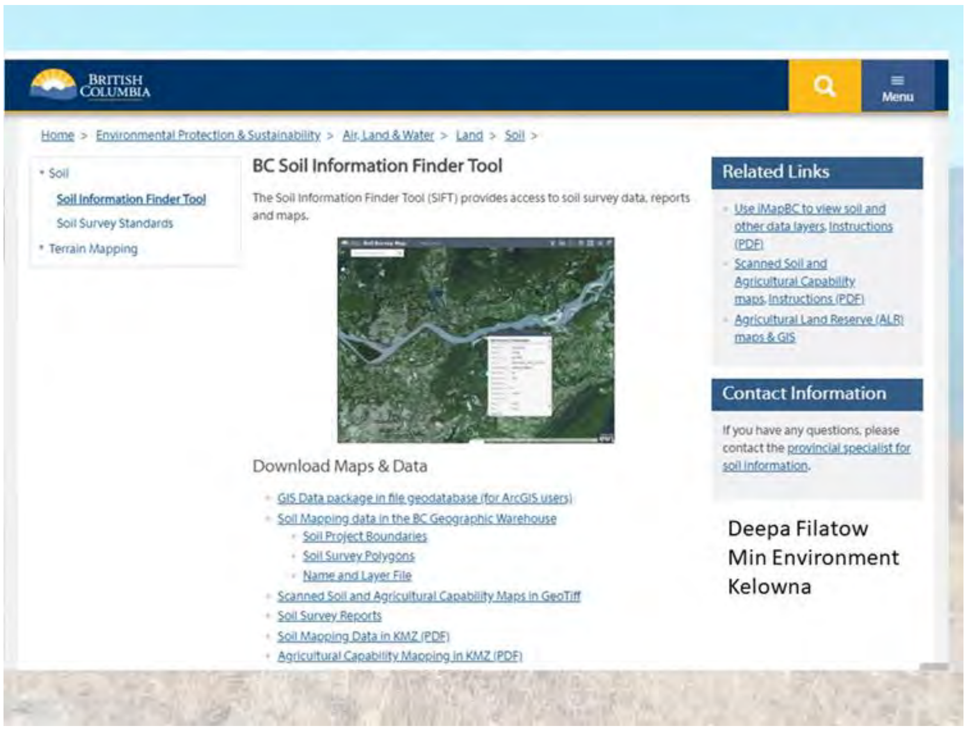
The BC Soil Information Finder Tool (SIFT) provides access to soil survey data, reports and maps, predominantly from the agricultural regions of the province. "Hectares BC" provides access to digital mapping data layers (raster data) for the province, including: soil landscapes, soil parent materials, land use/cover, and topography. A variety of predictive mapping exercises at various scales have been undertaken in various regions. A 250 m resolution map of soil classes has been developed (article in preparation) which is being used as an input for carbon mapping.

Successes and challenges include:

- Vector soil datasets are readily available for large parts of the province (SIFT)
- Need to define and organize existing data according to common rules and naming conventions for data and maps (e.g., is it colluvium or till)
- We've had some early success, and now trying to improve predictive maps
- Possible to represent common soil attributes directly without soil classes / names?
- BC has rugged topography
  - o lots of variation in relief factor of soil formation
  - o but parent material and vegetation and other factors often vary in 'unpredictable' ways
  - o sometimes finer resolution maps are no better than coarse resolution
- LiDAR has great potential (about 10 – 20 percent of BC has LiDAR)
- We're getting better and building capacity
- DSM is here to stay

Key Messages

- Combination of very specific and/or general attributes – easier to model and combine into complex interpretations (i.e., don't necessarily need soil names specifically – it is important to be clear about what soil attributes you need)
- Parent materials are easier to map than "great groups" however when you combine the two you are very close to a soil map even in the absence of other soil attributes



2.6 ENSEMBLE MACHINE LEARNING AS A FRAMEWORK FOR PSM – DR. TOM HENGL

The objective of the presentation was to provide an introduction to machine learning for soil mapping (soil properties, classes, class probabilities, etc.) and to discuss the new developments and benefits associated with Ensemble Machine Learning (which include: maximizing accuracy/data processing automation; modelling of non-linear complex relationships; combining geography/feature space patterns without too many strict statistical assumptions). Dr. Hengl decided to structure his talk based on the list of questions from Dr. Craig Aumann and included:

1. At a high level, how does PSM work?
2. How good are PSM maps today?
3. Is PSM ready for commercial use to support the needs of government and others?
4. What are the risks of using predictive maps in decision making?
5. What opportunities do predictive maps open up that we couldn't do before?
6. What does this generally cost?
7. What are others around the world doing around PSM? Are they having successes? And how are they defining success?

Hengl and MacMillan (2019) is a good reference for introductory and advanced information.

Digital soil mapping and predictive soil mapping are different (see presentation for links to good resources such as the USDA Natural Resources Conservation Service Soils website):

> ***Digital***: "The creation and the population of a geographically referenced soil database generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships."
>
> ***Predictive***: "The development of a numerical or statistical model of the relationship among environmental variables and soil properties, which is then applied to a geographic data base to create a predictive map."

Theoretically, PSM is largely based on CLORPT (Jenny, 1994). A conventional soil mapper (CSMer) knows how to draw polygon maps. The difference between a CSMer and a digital soil mapper (DSMer) is that a DSMer knows how to use GIS. The difference between a DSMer and PSMer is that a PSMer knows how to code (R, Python, Julia, etc.)! Modern PSM is largely based on using RS / proximal soil sensing / LiDAR and state-of-the-art Machine and Statistical Learning software (here focused on Open Source tools).

The objective of PSM is to produce optimal unbiased predictions of a mean value at some new location along with the uncertainty associated with the prediction, at the finest possible resolution. From the application point of view, the main application objective of soil mapping is to accurately predict response of a soil(-plant) ecosystem to various soil management strategies.

There are three main types of PSM projects:

- PSM projects in new, previously unmapped, areas — no point observations or samples currently exist.
- PSM projects using legacy points — sufficient point data to support PSM exist and are available, but no previous PSM modelling has been implemented for this area.
- PSM projects aimed at optimizing predictions and usability — previous PSM models have already been completed but previous results can still be improved / optimized.

The costs for the various types of PSM projects have been estimated based on a variety of factors. The costs are typically a function of the size of the area (spatial resolution), the total number of samples/points, the total number of variables and the distance that needs to be crossed to visit all points. Other projects can be developed with a different cost structure based on the intended outcomes. Another way of classifying PSM projects is by purpose:

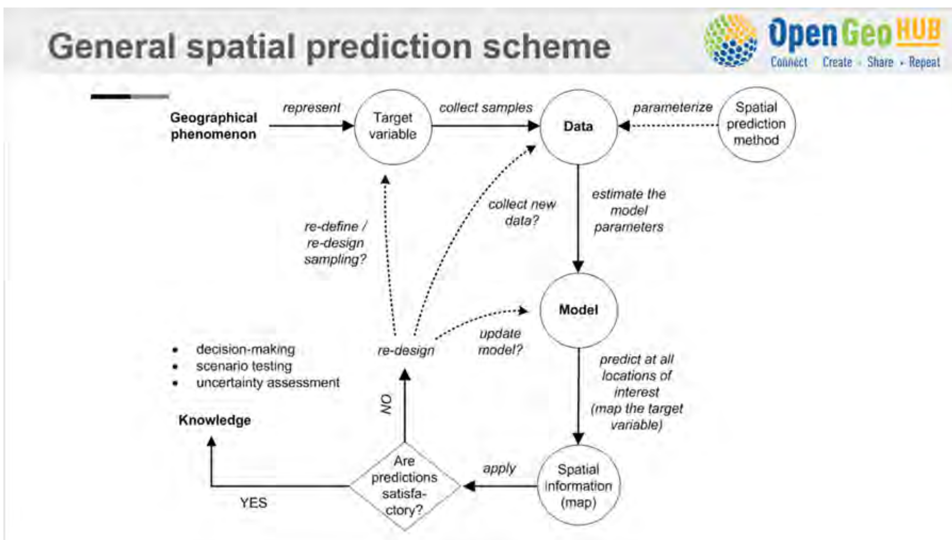
- PSM projects for the purpose of mapping static (stable) spatial patterns only.
- PSM projects for the purpose of one-time change detection (e.g. two time intervals).
- PSM projects for the purpose of monitoring soil conditions / status (continuous updates at regular intervals).

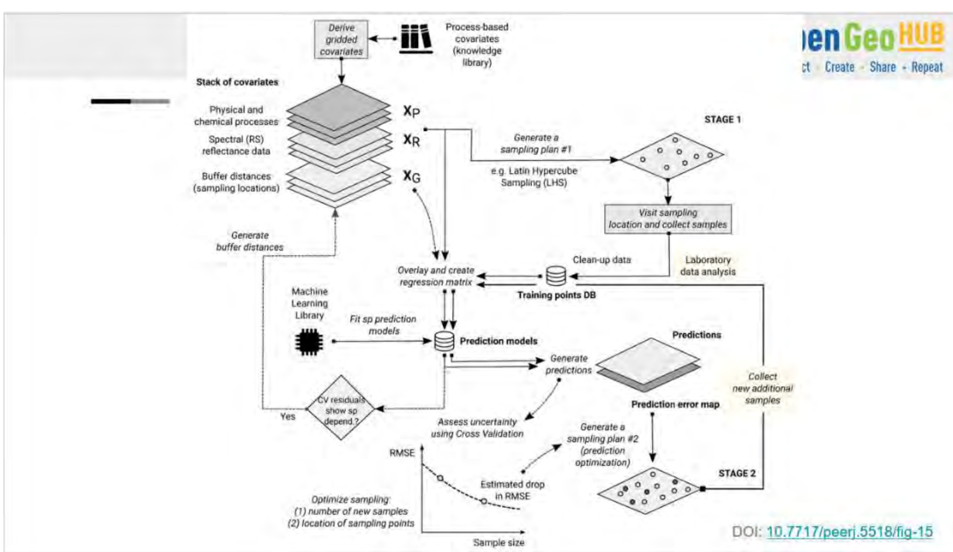
A good resource to evaluate the costs associated with PSM is Hengl et al. (2013).

One attractive framework to generate spatial predictions is the Ensemble Machine Learning (see e.g., SuperLearner, caret, mlr, h2o modeling frameworks for the R statistical language). By including geographical distance in the Machine Learning model, we can "elegantly" replace traditional geostatistics (kriging). Overall, it is scary how much we could automate generation of soil maps today, but that does NOT mean that: (a) we should stop sampling, (b) we do not need to learn/understand stats or soil science any more, or (c) that we need to stop coding.

Predictive soil mapping at a high level consists of two steps (which may be done iteratively in a second phase):

- Step 1: Observations coming in need to be verified with expert data and linked to co-variates.
- Step 2: The model needs to be calibrated (generally better if multiple models are being applied as is done with Ensemble Machine Learning). If the model and predictions are satisfactory, you can stop and use the information for the intended purposes. If the model is not providing satisfactory predictions you repeat these steps in what becomes Phase 2. These steps can continue with additional data, a different/updated model, and/or new/different covariates until satisfactory outcomes can be developed (see slides below).



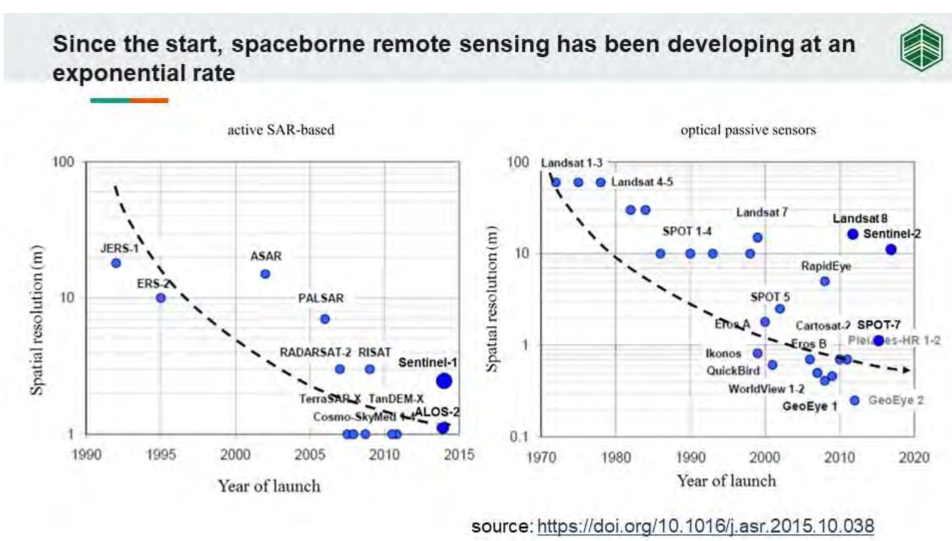


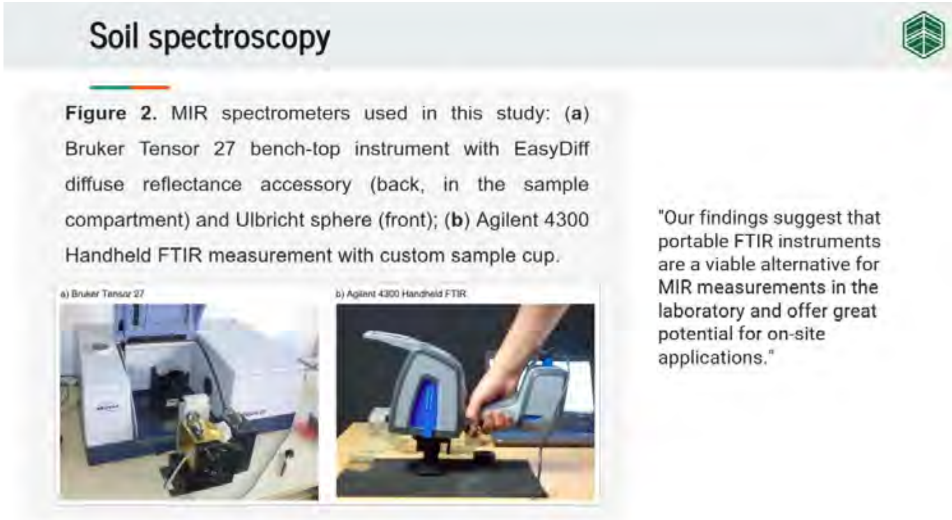
2.7 TECHNOLOGICAL AND METHODOLOGICAL ADVANCES IN SOIL MAPPING AND MONITORING AND OPPORTUNITIES FOR ALBERTA/CANADA – DR. TOM HENGL

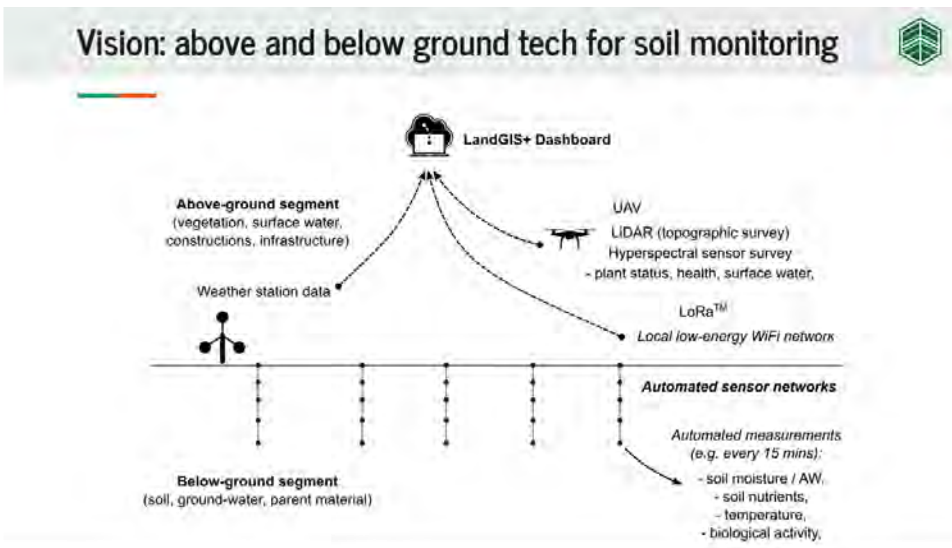
The presentation focused on advancing technologies, lessons learned from previous efforts, and which new application areas could benefit Alberta the most. The intent was to initiate discussion around how to leverage the current "big data" paradigm and how best to produce decision- and analyst-ready insight. Information was provided on the state-of-the-art remote sensing technologies and advancements in data resolution and availability. The impact of "free and open" data was highlighted. Examples were also provided where predictive data models did not accurately predict and the intended outcome could not be reached. Explanations for the un-successful examples included: high noise in the data, model parameters inaccurate, and cloud cover. A suite of advancements in new global data releases, "active"-type sensing and technologies for measuring soil properties were provided (Table 2).

Table 2. Promising technologies summary.

Monitoring	Processing
Public-access, high-res Space Sentinels, ALOS, Landsat 8, TANDEM	**Online geospatial data workflows** R-spatial (plumber), GeoTrellis DB, Cloud Optimized GeoTIFF
Commercial "NewSpace" Very fast system design/launch Data- or Intelligence-as-a-Service	**Permanent data storage** Digital data instantly accessible, unalterable, uncensorable
UAV fleets With LiDAR + radar + hyperspectral Data- or Intelligence-as-a-Service	**Scalable computing infrastructure** Diverse data storage/processing solutions as-a-Service
Integrated sensor networks (meteo, soil, acoustic ...)	







The following key conclusions were provided:

- More data available, but that does NOT mean easier mapping procedures (in fact we're in danger of being overwhelmed)
- Very promising new technologies for soil science: "IoT" sensor nets, LiDAR, Sentinel-1,2,5, ALOS
- Single sensor ↦ multi-sensor fusion;
  Single mission ↦ multidisciplinary projects;
  Single model ↦ Ensemble models;
- Machine Learning is increasingly available! Use it!

The following opportunities for Alberta (Canada) were noted:

- Large areas = leveraging RS is crucial (UAVs maybe not the best option for now?)

- Set-up automated sensor networks (weather, micro-weather, soil moisture, soil pollution...)
- On-demand web services: warning, spatial planning, scenario testing systems
- Next standard: hyper-resolution mapping 1–5 m
- Cloud / HPC solutions for geospatial processing GIS data (start with Open Source!)

Question Period

Q: My sense is there is a lot of noise in remote sensing. Can you suggest some of the ways you have dealt with that?

A: Take a time series, average that data for each month over 10 years (for example). So now you have a long term data set.

3.0 COMMON THEMES IN PRESENTATIONS

The following common themes were identified during the presentations:

1. Ask what your users want (information and scale), and ask frequently – PSM is about the user not the mapper!
2. As a visible end-product, the maps are important but the workflow, models and data underpinning the maps are the key to success.
3. To the extent possible, data and models should be public.
4. Once the first map is generated it creates opportunities for people to ask "what if …" and "could we …" questions that will spur new products and innovations.
5. Residual error maps allow for targeted sampling campaigns to focus on areas where there is high variability and/or missing data.
6. More data is better, but the right data in the right places is even better – use data that are relevant to the question being asked.
7. Legacy data still have value – incorporate them wherever possible.
8. Incorporating remote sensing tools and products will be the next big step in producing PSM.
9. Dynamic maps, showing trends over time, are the next big advance in PSM products.
10. PSM models are not black boxes – they are based on long-accepted soil-landscape principles going back to the 1940s.
11. Current technology offers opportunities to harness citizen science to vastly increase sampling capabilities.

4.0 APPLYING THE LEARNINGS TO FORASIS

After the presentations participants were given the chance to provide advice and comments relative to application of Predictive Soil Mapping in Alberta's Green Area (i.e., the FORASIS project). The majority of the discussion was centered on data availability and sharing. The following points were raised.

- Soils gets some money for research in BC, but ecology gets more (likely because of relationship to forestry). Potentially consider forest inventory work and forest fire planning in the green zone and forestry companies as end users.
- Soil moisture is a key issue.
- Question is, is it a lack of quality data or a lack of availability of data? Wonder how much info is sitting on consultant's desks. Industry could cut down on amount of sampling to save cost – would be a benefit of FORASIS, open data.
- Point field observations – where do they end up? Do they die after report submitted to regulatory body? Lab data – must have been in a digital format anyway. We collect the data for private companies, but then it is THEIR data. We don't own the data. So don't look at consultants for data.
- Needs to be a central data home for soil point data.
- So who has opportunity and ability to make it available? It would be good enough to know which companies how much data, in what area, etc. Maybe don't need all the data. Maybe then someone can offer to pay to see the actual data?
- Industry wants to reduce sampling cost. Having a standard in place for how data is reported would make it easy to access and use, potentially. Brings together private sector and public sector entity managing and maintaining that data.
- Could a soil mapping system look like the AB Wetland Inventory for mapping wetlands? Could that framework be applied to soils? Should be shared management approach.
- Environmental Site Assessment Repository (Alberta Environment and Parks) has reports that public can access, and would have soils info. Again, client, decides if that info goes into the repository. Landowners may also need to authorize accessing info from reports on their land. Site assessment data may not always be typical soil survey data.
- Seems like a failure to communicate. We should collect all point data. If government makes a regulation saying all data must be collected, then everyone has to do it. Doesn't all have to be made publicly available. But then that data could be used for things such as a predictive soil map. There is a new directive for renewable energy projects that gives AEP room to do this. If industry wants to reduce sampling cost, then it would mean moving in the direction of making more data available. Make data more accessible at a cost.
- Someone could just go out and start taking points. Show the way, show an example of the value of this PSM. This could trigger response to share data.

- I would like to return to original question. Who is the user of FORASIS? Could inform who will put in info, etc. It's not AGRASID extended into the Green Area. It can be used for pre-disturbance assessment, etc. – useful for consultants and such who need this sort of info for the Green Area.
- What would the disclaimer to the dataset look like? People want to be skeptical about predictive maps. But the same scrutiny is not applied to conventional maps. Conventional maps certainly have data errors, but there is a general sentiment that "my map is the best map" which prevents people from embracing new methods.
- What you put into the model may not have had QA/QC done. This is accounted for in the prediction error. Can do repeated sampling, nested sampling to reduce error. Can't report higher precision than what you can actually achieve.

5.0 ADDITIONAL RESOURCES

5.1 Documents

Aumann, C., 2017. Predictive Soil Mapping Pilot in NE Alberta. Prepared for Alberta Biodiversity Monitoring Institute, Edmonton, Alberta. 37 pp. plus appendices. https://ftp-public.abmi.ca/home/publications/documents/535_Aumann_2017_PredictiveSoilMapping_ABMI.pdf

Fawcett, M.D., W.L. Nikiforuk, R.L. McNeil and R.A. MacMillan, 1993. An Evaluation of the Extrapolatory Method of Soil Mapping. Environmental Research and Engineering Department, Alberta Research Council, Edmonton, Alberta. Alberta Research Council Open File Report 1993-09. 69 pp. https://ags.aer.ca/document/OFR/OFR_1993_09.pdf

Hengl , T., J. Mendes de Jesus, G.B.M. Heuvelink, M.R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M.N. Wright, X. Geng, B. Bauer-Marschallinger, M.A. Guevara, R. Vargas, R.A. MacMillan, N.H. Batjes, J.G.B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel and B. Kempen, 2017. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. PLoS ONE 12(2): e0169748. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169748

Hengl, T. and R.A. MacMillan, 2019. Predictive Soil Mapping with R. OpenGeoHub Foundation, Wageningen, The Netherlands. 370 pp. ISBN: 978-0-359-30635-0. https://envirometrix.github.io/PredictiveSoilMapping/index.html

Hengl, T., M. Nikolić and R.A. MacMillan, 2013. Mapping Efficiency and Information Content. International Journal of Applied Earth Observation and Geoinformation 22: 127-138.

Heung, B., 2017. Regional-Scale Digital Soil Mapping in British Columbia using Legacy Soil Survey Data and Machine-Learning Techniques. Ph.D. Thesis. Department of Geography, Simon Fraser University, Burnaby, British Columbia. 164 pp. http://summit.sfu.ca/system/files/iritems1/17377/etd10172_BHeung.pdf

Nadeau, L.B., C. Li and H. Hans, 2004. Ecosystem Mapping in the Lower Foothills Subregion of Alberta: Application of Fuzzy Logic. The Forestry Chronicle 80(3): 359-365. http://pubs.cif-ifc.org/doi/pdf/10.5558/tfc80359-3

5.2 Websites

AfSIS Database – https://qed.ai/afsisdb/

Agricultural Region of Alberta Soil Inventory Database (AGRASID) – https://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/sag10372

Alberta Data Partnerships (ADP) – http://abdatapartnerships.ca/

BC Soil Information Finder Tool (SIFT) – https://www2.gov.bc.ca/gov/content/environment/air-land-water/land/soil/soil-information-finder

Canadian Soil Information Service (CanSIS) – http://sis.agr.gc.ca/cansis/

Environmental Site Assessment Repository (ESAR) – http://www.esar.alberta.ca/esarmain.aspx

European Soil Data Centre (ESDAC) – DSM: Digital Soil Mapping – https://esdac.jrc.ec.europa.eu/projects/dsm-digital-soil-mapping

GeoDiscover Alberta – https://geodiscover.alberta.ca/geoportal/catalog/main/home.page

GeoSurvey – https://geosurvey.qed.ai/about/

GlobalSoilMap.org – https://www.globalsoilmap.net/

Hectares BC – https://www.hectaresbc.org/app/habc/HaBC.html

International Soil Reference and Information Centre (ISRIC) – https://www.isric.org/

OpenGeoHub – http://opengeohub.org/

Radiant Earth Foundation – https://www.radiant.earth/

Saskatchewan Soil Information System (SKSIS) – https://sksis.usask.ca/#/map

Soil$_X$ – http://www.soilx.ca/

US Department of Agriculture – Digital Soil Mapping – https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=stelprdb1254424

US Geological Survey – Digital Soil Mapping: High Resolution Maps for Modern Land Management Decisions – https://www.usgs.gov/centers/sbsc/science/digital-soil-mapping-high-resolution-maps-modern-land-management-decisions?qt-science_center_objects=0#qt-science_center_objects

APPENDIX A: LIST OF SEMINAR REGISTRANTS

The following people attended the March 6, 2019 Seminar at the InnoTech Alberta facility in Edmonton.

Name	Organization
Jaime Aguilar	
Murray Anderson	M.L. Anderson Advisory Services
Laurence Andriashek	Alberta Energy Regulator
Craig Aumann	InnoTech Alberta
David Bergstrom	Alberta Energy Regulator
Michael Bock	Government of Alberta
Scott Boorman	Paragon Soil and Environmental Consulting Inc.
Eldon Borgen	Paragon Soil and Environmental Consulting Inc.
Tony Brierley	
Christiane Brouwer	Golder
David Campbell	Forestry Corp
Chi Chen	Alberta Environment and Parks
Kimberly Cornish	Pachaterrae
Diana Dabrowa	Paragon Soil and Environmental Consulting Inc.
Kara Dallaire	Paragon Soil and Environmental Consulting Inc.
Sebastian Dietrich	University of Alberta
Konstantin Dlusskiy	Paragon Soil and Environmental Consulting Inc.
Becky Doherty	Forestry Corp
Lynette Esak	Esak Consulting Ltd.
Lane Feschuk	Paragon Soil and Environmental Consulting Inc.
Ted Furler	Stantec
Becca Gilroyed	Paragon Soil and Environmental Consulting Inc.
Claudia Gomez	Matrix Solutions
Reid Graham	Tree Time
Sanjay Gupta	

Name	Organization
Stephanie Hannem	Klohn Crippen Berger Ltd.
Jason House	Stantec
Javed Iqbal	Government of Alberta
Kyle Jones	Alberta Environment and Parks
Jahan Kariyeva	University of Alberta
Brian Lambert	Alberta Environment and Parks
Len Leskiw	Paragon Soil and Environmental Consulting Inc.
Barb Logan	Paragon Soil and Environmental Consulting Inc.
Derek Lonergan	
Ben Manshanden	Government of Alberta
Leon Marciak	Precedent Environmental Management Inc.
Jackie Maxwell	
Susan McGillivray	Alberta Environment and Parks
Larry Nikiforuk	SOIL-INFO LTD
Lekan Olatuyi	Alberta Environment and Parks
Claudia Palylyk	
Rashmi Pathak	ALS Global
Shane Patterson	Alberta Environment and Parks
Steven Pawley	Alberta Energy Regulator
Robbie Price	NorthWind Land Resources Inc.
Robert Proudfoot	
Jeff Reading	Pachaterrae
Steve Reed	SilvaComm
Thomas Romanowski	Sherritt
Jilene Sauve	Matrix Solutions
Ron Sawatzky	InnoTech Alberta
Colleen Shabada	SWAT Consulting Inc.
Preston Sorenson	Maapera Analytics Inc.
David Spiess	Alberta Agriculture and Forestry
Marcin Stanislawski	Golder

Name	Organization
Angela Taylor	Stantec
Sarah Thacker	InnoTech Alberta
Bruce Walker	
Don Watson	Alberta Environment and Parks
Marian Weber	InnoTech Alberta
Takele Zleke	Advisian

The following people registered to participate in the Seminar through a webinar broadcast.

Name	Organization [1]
Vaishalie Anand	Alberta Environment and Parks
Koreen Anderson	Alberta Health Services
Les Anderson	CCS Energy Services
Jose Armando Gomez	
Andrew Arnold	Summit Earth
Carlos Arregoces	Ecoventure Inc.
Christopher Bater	Government of Alberta
Kathryn Bessie	Tetra Tech Canada
Christopher Blackford	University of Toronto
Marla Bohm	Paragon Soil and Environmental Consulting Inc.
Atty Bressler	Wood PLC
Maxine Butler	
Luke Bye	Terex Environmental
Gayle Caltagirone	Tetra Tech Canada
Jennifer Canham	Weston Foundation
Christopher Chagumaira	
Nikki Chartrand	Secure Energy
Jeff Christisansen	Golder
Jill Clarke	
Helen Lynn Connally	

Name	Organization [1]
Burton Cosgrove	Integrated Environments (2006) Ltd.
Alexandra Dalton	
Michelle Dias	City of Calgary
Catrina Duffy	Solstice Canada Corp.
Leanne Erickson	Alberta Energy Regulator
Kahlie Forster	Stantec Consulting
William Gardin	
Yohannes Getachew	
Adam Gillespie	University of Guelph
Denise Gordon	Yukon Government
Travis Grant	
Tony Gregov	
Claire Gunoud	
Sarah Hall	
Sheldon Hann	
Lauren Harding	Matrix Solutions Inc.
Emily Herdman	Government of Alberta
Guillermo Hernandez Ramirez	University of Alberta
Brandon Heung	
Mitch Heynen	Yukon Government
Kirsten Horne	Terracon Geotechnique Ltd.
Tahmid Huq Easher	University of Guelph
Carolyn Inglis	
Stephanie Jaffray	
Ed Karpuk	Government of Alberta
Babak Kasraei	Simon Fraser University
Jaylene Kemp	
Karen Klimek	Agland Corp
Danny Lajoie	North Shore Environmental Consultants Inc.
Michael Lau	Volker Stevin

Name	Organization [1]
Dana Lee	McElhanney
Sarah Lepp	University of Guelph
Eduardo Loos	Vertex
Andrew Marcil	Jacobs
Hillard MacDonald	Core Geomatics
Joe Meaney	Suncor Energy
Tom Messier	
Symon Mezbahuddin	Government of Alberta
Brenda Nachtegaele	Strathcona County
Chris Newton	Ecoventure Inc.
Lindsay Oiffer	Matrix Solutions
Olusegun Oyewole	
Robert Palmer	
Karen Patz	
Colin Peters	Alberta Energy Regulator
Tyler Phillips	Teck
Carmella Pierce	Esker Consulting
Donald Poisson	
Matt Porter	Trent University
Clara Qualizza	
Fabrizio Re	SNC-Lavalin Environment Inc.
Mary Sarodub	
Daniel Saurette	Government of Ontario
Margaret Schmidt	Simon Fraser University
Kara Schwaebe	
Haylee Smysniuk	University of Saskatchewan
Karen Stals	Alberta Energy Regulator
Marcin Stanislawski	Golder
Taralee Stephenson	
Karen Trenholm-Boyle	

Name	Organization [1]
Deanna Van Muyen	Reclaimit
Stephanie Vickers	
Tara Wang	Alberta Energy Regulator
Troy Wawrinchuk	Keneco Environmental Services (2000) Inc.
Ryleigh-Raye Wolfe	MacKenzie County
Ian Young	SLR Consulting
Jin Zhang	Simon Fraser University

[1] In most cases, organization was interpreted by e-mail address.

APPENDIX B: WORKSHOP AGENDA

Time	Topic	Speaker
8:30 – 8:45 am	Registration and coffee	
8:45 – 8:55 am	Welcome, Introductions, and FORASIS Context	Bonnie Drozdowski
8:55 – 9:15 am	Shifting from traditional to predictive soil mapping (PSM)	Robert (Bob) MacMillian
9:15 – 10:30 am	Lessons Learned: what's been tried; problems and solutions for utilizing PSM	Markus Walsh
10:30 – 10:45 am	Break	
10:45 – 11:30 pm	Saskatchewan Soil Information System (SKSIS)	Angela Bedard-Haughn
11:30 – 12:00 pm	Predictive Soil Mapping – National Perspective	Xiaoyuan Geng
12:00– 12:15 pm	Predictive Mapping in Beautiful British Columbia	Chuck Bulmer
12:15 – 1:00 pm	Lunch (will be provided)	
1:00 – 2:15	Ensemble Machine Learning as a framework for PSM [the basics]	Tom Hengl
2:15 – 3:30	Technological and methodological advances in soil mapping and monitoring and opportunities for Alberta / Canada	Tom Hengl
3:30– 4:30	Closing Remarks, Questions, and Working Session	Craig Aumann and Chris Powter

APPENDIX C: PRESENTATIONS

The following sections contain the presentation materials provided by the speakers.

Shifting from traditional to predictive soil mapping (PSM) – Bob MacMillan

Lessons Learned: What's been Tried; Problems and Solutions for Utilizing PSM – Markus Walsh

Saskatchewan Soil Information System (SKSIS) – Angela Bedard-Haughn

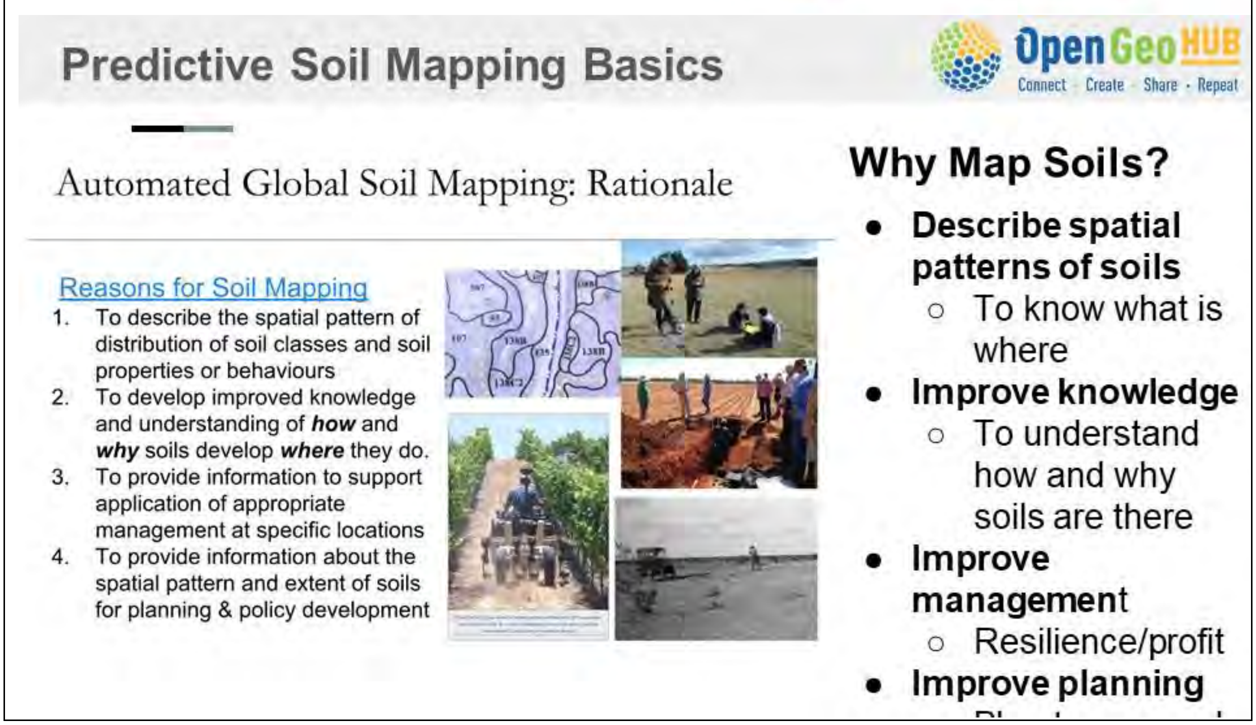
Predictive Soil Mapping—National Perspective – Xiaoyuan Geng

Predictive Mapping in Beautiful British Columbia – Chuck Bulmer

Ensemble Machine Learning as a Framework for PSM – Tom Hengl

Technological and Methodological Advances in Soil Sapping and Monitoring and Opportunities for Alberta / Canada – Tom Hengl

Shifting from traditional to predictive soil mapping (PSM) – Bob MacMillan



Predictive Soil Mapping Basics

OpenGeoHUB Connect · Create · Share · Repeat

Universal Model of Spatial Variation: A Unifying Scientific Theory

Applies to ALL spatial prediction
Matheron (1971)

$z(x,y) = f(x,y) + \varepsilon,$

Adapted to predict soil variation by:
Burgess and Webster (1980 a,b)
Webster and Burrough (1980)
Burrough (1986)
Webster and McBratney (1987)
Oliver (1989)

THEORY

The general statistical approach to prediction embodied in regionalized variable theory combines a deterministic component, such as that of trend surface analysis, with a stochastic one, so that the spatial variation in an attribute is expressed by

$$z(\mathbf{x}) = \sum a_k f_k(\mathbf{x}) + \varepsilon(\mathbf{x}), \qquad (3)$$

where **x** denotes the spatial coordinates in one, two or three dimensions, the f_k, $k = 0, 1, \ldots,$ are functions

Methods of Spatial Interpolation 155

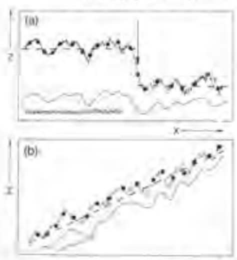


Fig. 8.10 The major components of spatial variation.
A 'structural' value that may change abruptly at a boundary (a), or vary with a constant trend (b).
Spatially correlated random variation $\varepsilon'(x)$
Spatially uncorrelated random variation ε'' (noise). Both $\varepsilon'(x)$ and ε'' are assumed to have uniform statistical properties over the whole of the sample area.
Sample sites, at which only a part of the true variation is observed.

Scientific Theory

- **For all spatial models**
  - Close things are similar
  - Far things are different
  - Similarity related to distance apart
- **Adapted for soils**
  - Universal model of ***soil*** variation
  - We see this later

Predictive Soil Mapping Basics

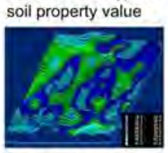
OpenGeoHUB Connect · Create · Share · Repeat

Universal Model of Soil Spatial Variation

A Unifying Framework for Automated Predictive Mapping of Soils

$$Z(s) = Z^*(s) + \varepsilon'(s) + \varepsilon'$$

Predicted soil type or soil property value

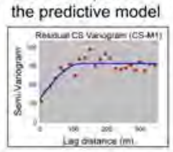


Predicted spatial pattern of some soil property or class including uncertainty of the estimate

Deterministic part of the predictive model

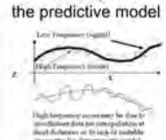
part of the variation that is predictable by means of some statistical or heuristic soil-landscape model

Stochastic part of the predictive model



part of the variation that shows spatial structure, can be modelled with a variogram

Pure Noise part of the predictive model



part of the variation that can't be predicted at the current scale with the available data and models

Theoretical Basis

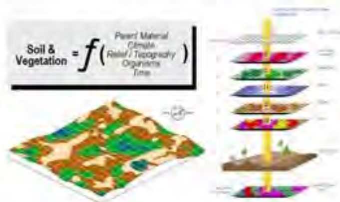
- **Deterministic part**
  - Part that can be explained
  - By some model
- **Stochastic part**
  - Part that has spatial structure
  - Can interpolate via kriging
- **Error part (noise)**
  - Unstructured
  - Can't interpolate

Predictive Soil Mapping Basics

Deterministic Part of Prediction Model: Z*(s)

Conceptual Models
Basis for manual soil mapping
Based on conceptual soil-landscape models
Empirical models developed by soil surveyors
Produce area-class polygon maps

Statistical Models
Basis for automated predictive soil mapping
CLORPT: relates soils/soil properties to observed environmental covariates
Explains spatial distribution of soils in terms of statistical relationships of soils to known soil forming factors as represented by covariates



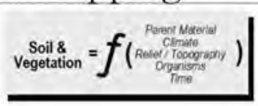
Theoretical Basis

- **Deterministic part**
  - Part that can be explained
- **Conceptual models**
  - Expert knowledge
  - Soil-landscape
  - Polygon maps
- **Statistical models**
  - Data driven
  - Objective
  - Soil-landscape too
  - Continuous maps

Predictive Soil Mapping Basics

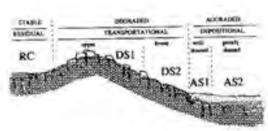
Theoretical Basis for Current Manual or "Conventional " Soil or Ecological Mapping

Jenny (1941)
CLORPT
Climate, Organisms, Relief, PM, Time (No space)



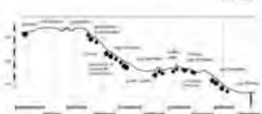
Simonson (1959)
Soil genesis conceptual process model
Additions, removals, translocations, transformations

Ruhe (1975)
Geomorphic process models
Erosional -Depositional surfaces, open/closed basins



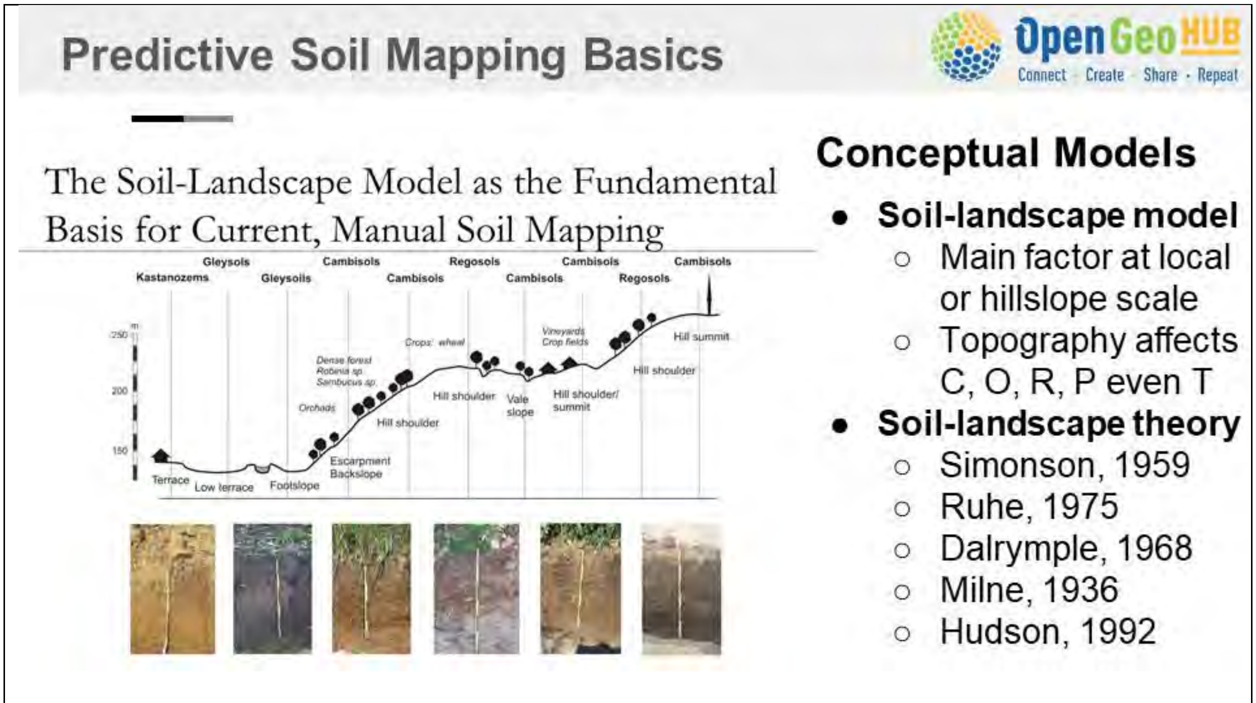
Dalrymple et al., (1968)
Hillslope (geomorphic) model
Nine unit hill slope model

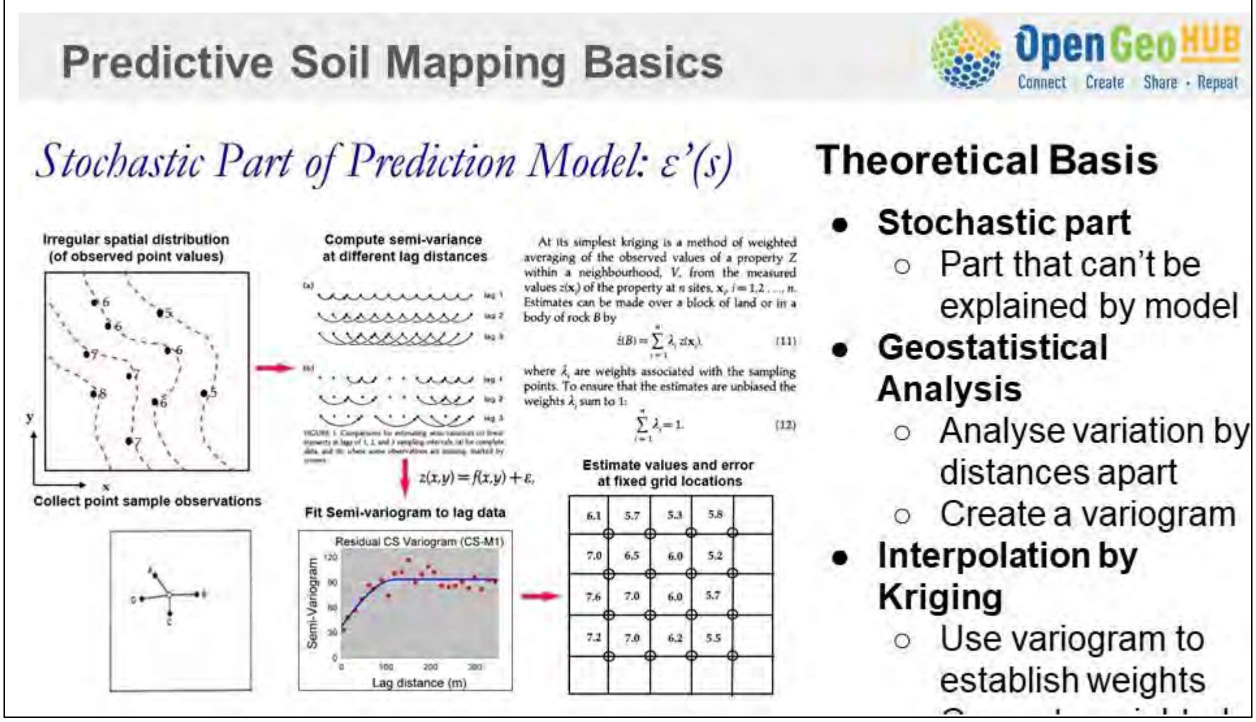
Milne (1936a, 1936b)
Soil-landscape conceptual model
Catena concept, toposequences



Conceptual Models

- **For soil mapping**
  - CLORPT (Jenny, 1941)
  - SCORPAN (McBratney, 2003)
- **Additional theory**
  - Simonson, 1959
  - Ruhe, 1975
  - Dalrymple, 1968
  - Milne, 1936
  - Hudson, 1992





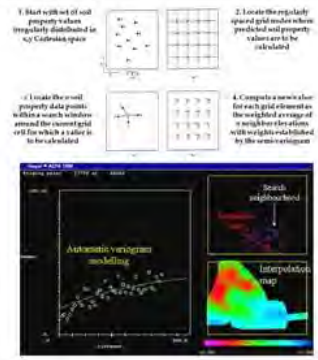
InnoTech ALBERTA

Predictive Soil Mapping Basics

Stochastic Part of Prediction Model: ε'(s)
Optimal Interpolation by Kriging

Geostatistical Estimation
Predict soil properties
Point or block kriging
Predict soil classes
Indicator kriging (0/1)
Soil class likelihood values
Predict error of estimate

Correct Deterministic Part
Error in deterministic part is computed
Compute residuals
If structure exists in error residuals
Then krige error & subtract from predictions



Geostatistics & Kriging

- **Kriging/spatial prediction**
  - Estimate value at point
  - Weighted average of all surrounding values
  - Weight is computed according to variogram
- **Manual Mappers**
  - Interpolate manually
  - "Dig it out"
  - Draw boundaries around similar soils or values

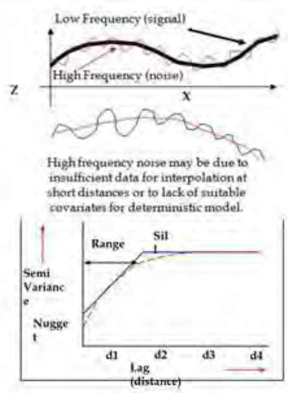
Predictive Soil Mapping Basics

Pure Noise Part of Prediction Model: ε'

Some variation is just not predictable
Have to be honest about this
Quantify it and report it (pure noise)

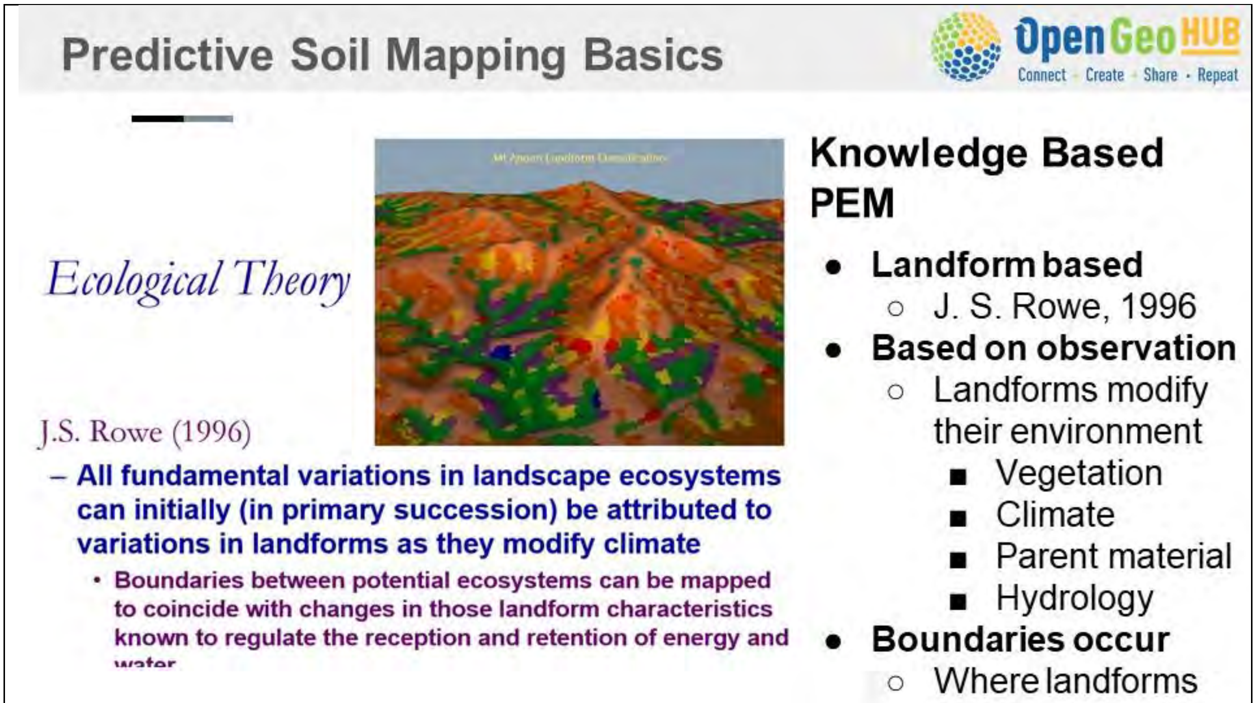
Deterministic prediction
Mental and statistical models
Not perfect –often lack suitable covariates to predict target variable, or not at proper resolution

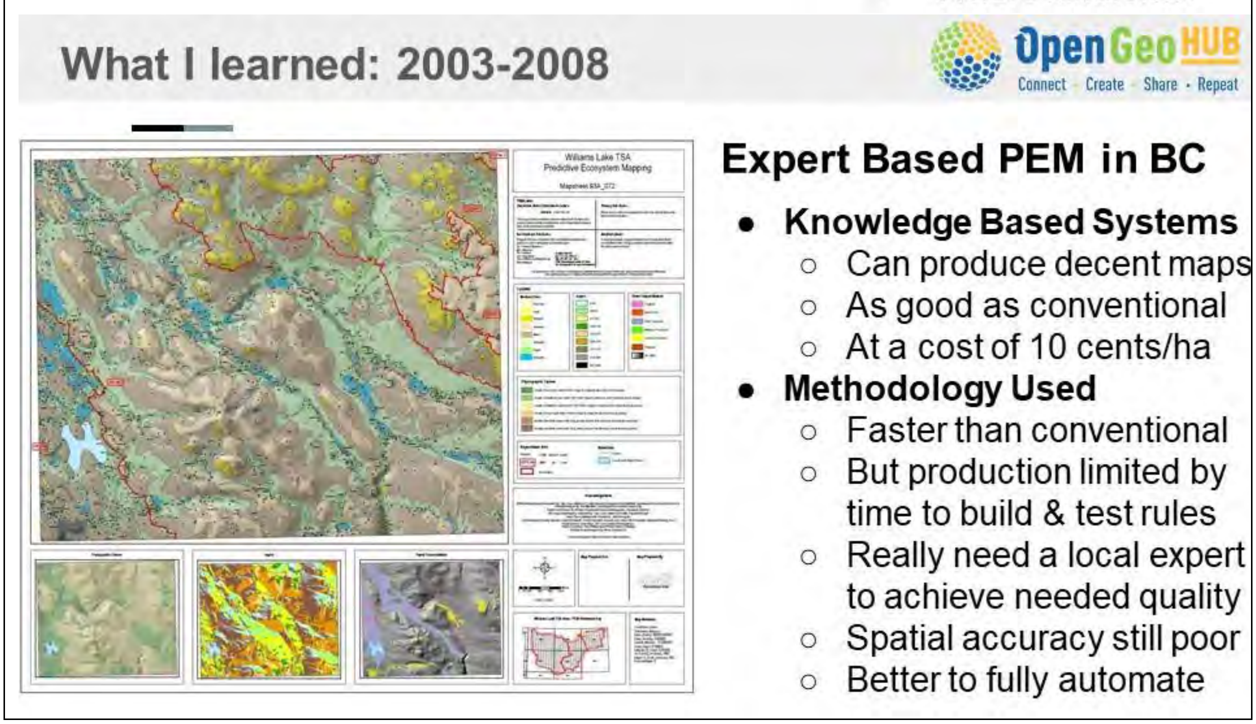
Geostatistical Prediction
Insufficient point input data
Can't predict at less than the smallest spacing of input point data



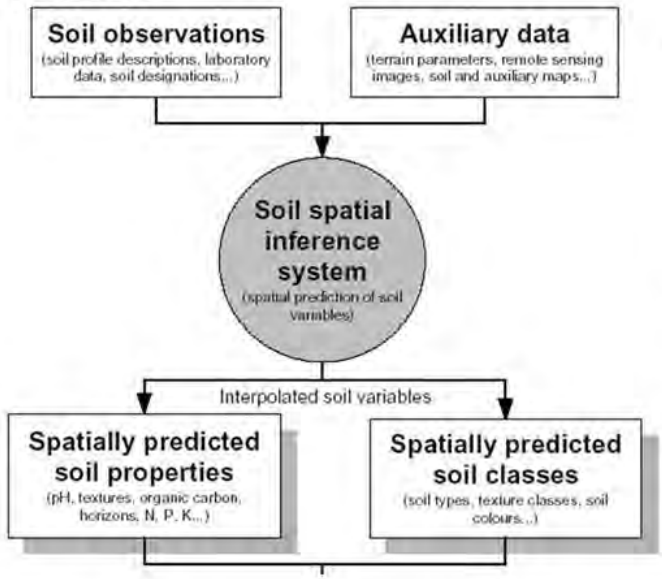
Theoretical Basis

- **Pure Noise part**
  - Part that can't be explained/predicted
- **May be deterministic**
  - But data or models can't capture it
- **May have structure**
  - But structure is shorter than spatial resolution of data
  - Don't have points close enough


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
Ecological Theory
J.S. Rowe (1996)
– All fundamental variations in landscape ecosystems can initially (in primary succession) be attributed to variations in landforms as they modify climate
• Boundaries between potential ecosystems can be mapped to coincide with changes in those landform characteristics known to regulate the reception and retention of energy and water
Knowledge Based PEM
● Landform based
○ J. S. Rowe, 1996
● Based on observation
○ Landforms modify their environment
■ Vegetation
■ Climate
■ Parent material
■ Hydrology
● Boundaries occur
○ Where landforms


What I learned: 2003-2008
OpenGeoHUB
Connect · Create · Share · Repeat
Williams Lake TSA
Predictive Ecosystem Mapping
Expert Based PEM in BC
● Knowledge Based Systems
○ Can produce decent maps
○ As good as conventional
○ At a cost of 10 cents/ha
● Methodology Used
○ Faster than conventional
○ But production limited by time to build & test rules
○ Really need a local expert to achieve needed quality
○ Spatial accuracy still poor
○ Better to fully automate

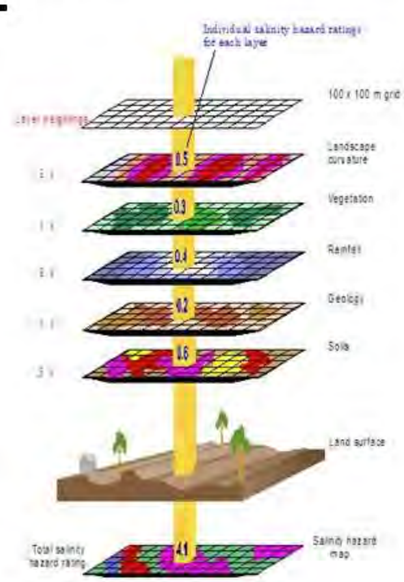
Predictive Soil Mapping Basics: Data Driven



Predictive Soil Mapping

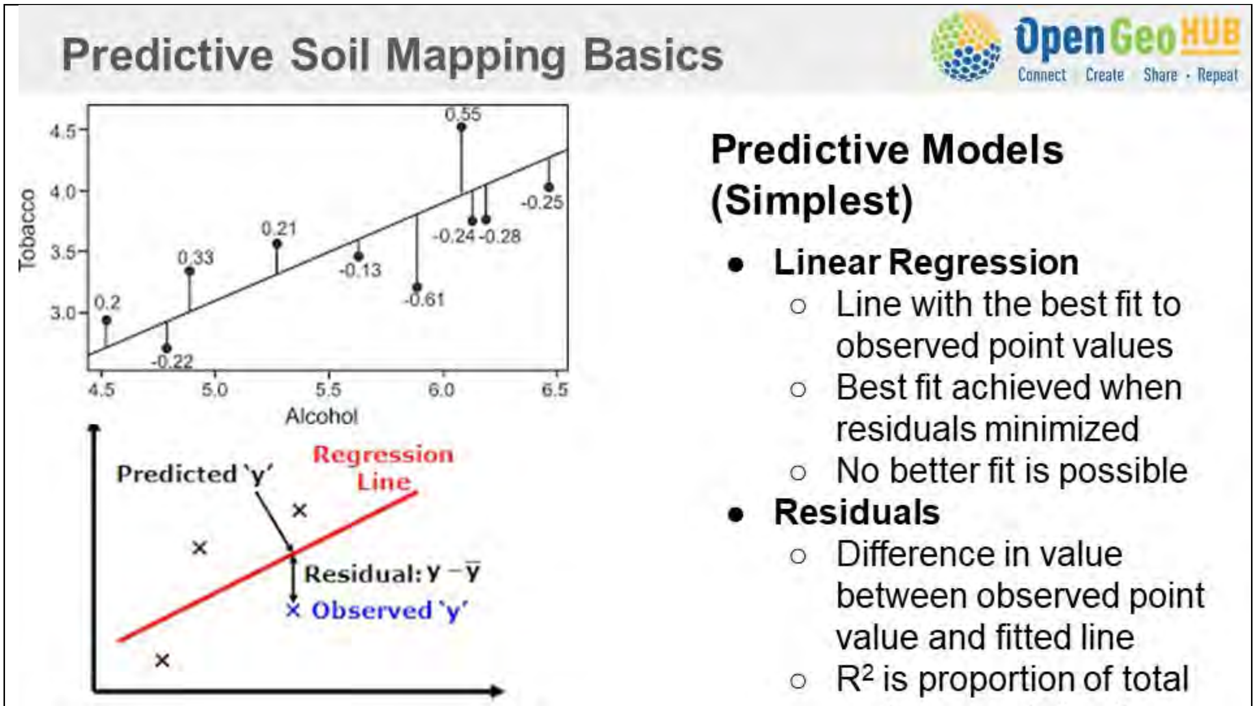
- **Input Data**
  - Field observations & measurements at points
  - Covariates (Grid Maps) as Explanatory variables
- **Prediction Models**
  - Identify relationships between point & grid data
- **Outputs**
  - Predicted soil classes
  - Predicted soil properties
  - Uncertainty estimates

Predictive Soil Mapping Basics

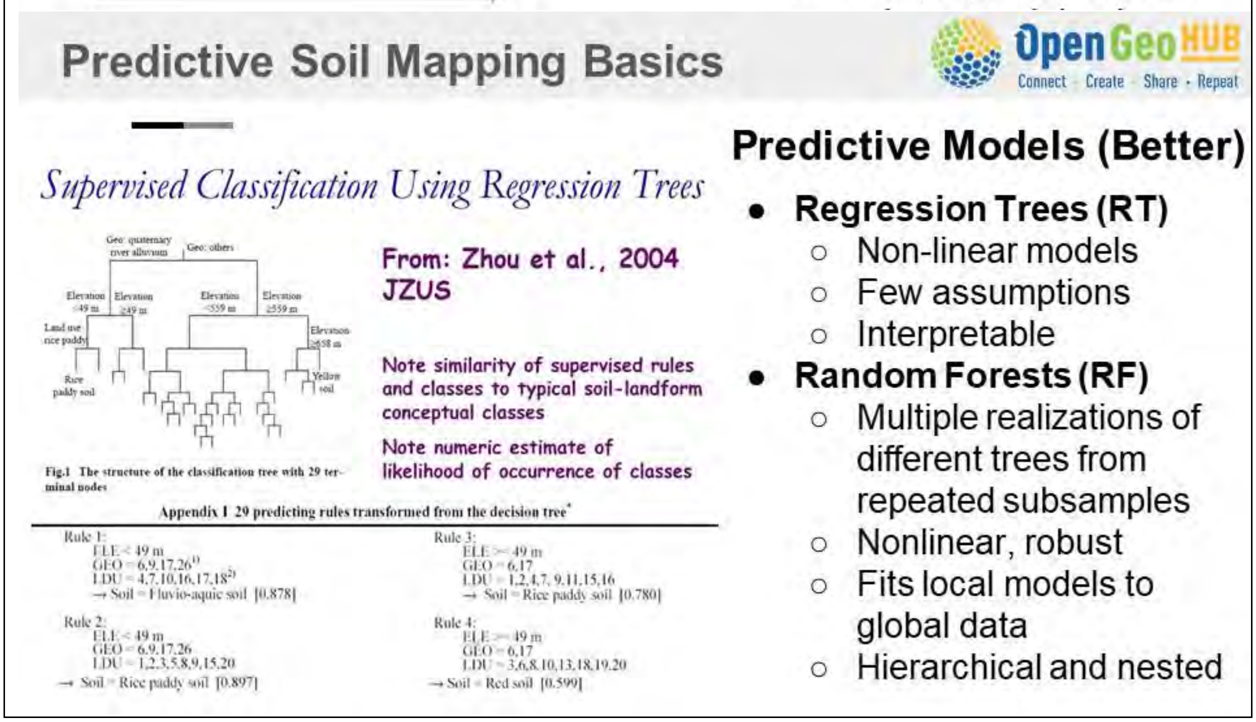


Predictive Model Input Data

- **Field Sample Data at Points**
  - Field observations & measurements at points
  - Provides the evidence needed to build models
  - Conceptual or statistical
- **Environmental Covariates**
  - Environmental variables usually as grid maps
  - Used to explain spatial patterns observed

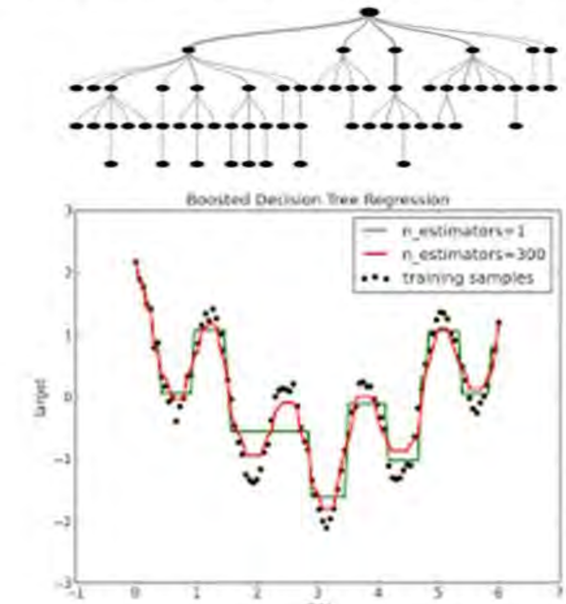


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
Tobacco
Alcohol
4.5
4.0
3.5
3.0
5.0
5.5
6.0
6.5
0.55
-0.25
0.21
-0.24 -0.28
0.33
-0.13
-0.61
0.2
-0.22
Predicted 'y'
Regression Line
Residual: Y − Ȳ
Observed 'y'
Predictive Models (Simplest)
Linear Regression
Line with the best fit to observed point values
Best fit achieved when residuals minimized
No better fit is possible
Residuals
Difference in value between observed point value and fitted line
R² is proportion of total



Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
Supervised Classification Using Regression Trees
Geo: quaternary river alluvium
Geo: others
Elevation <49 m
Elevation ≥49 m
Elevation <559 m
Elevation ≥559 m
Land use rice paddy
Elevation ≥658 m
Rice paddy soil
Yellow soil
Fig.1 The structure of the classification tree with 29 terminal nodes
From: Zhou et al., 2004 JZUS
Note similarity of supervised rules and classes to typical soil-landform conceptual classes
Note numeric estimate of likelihood of occurrence of classes
Appendix I 29 predicting rules transformed from the decision tree
Rule 1:
ELE < 49 m
GEO = 6,9,17,26
LDU = 4,7,10,16,17,18
→ Soil = Fluvio-aquic soil [0.878]
Rule 2:
ELE < 49 m
GEO = 6,9,17,26
LDU = 1,2,3,5,8,9,15,20
→ Soil = Rice paddy soil [0.897]
Rule 3:
ELE >= 49 m
GEO = 6,17
LDU = 1,2,4,7, 9,11,15,16
→ Soil = Rice paddy soil [0.780]
Rule 4:
ELE >= 49 m
GEO = 6,17
LDU = 3,6,8,10,13,18,19,20
→ Soil = Red soil [0.599]
Predictive Models (Better)
Regression Trees (RT)
Non-linear models
Few assumptions
Interpretable
Random Forests (RF)
Multiple realizations of different trees from repeated subsamples
Nonlinear, robust
Fits local models to global data
Hierarchical and nested

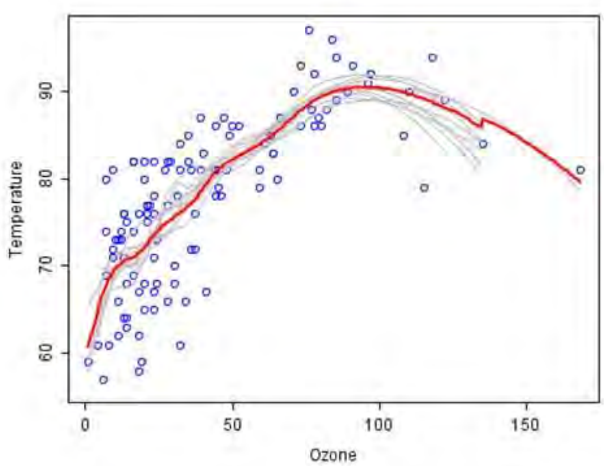
Predictive Soil Mapping Basics



Non-linear Regression (Best)

- **Piecewise linear fits (RF)**
  - Line segments with the best fit to observed values
  - Fit is achieved piecewise within tree branches

Predictive Soil Mapping Basics

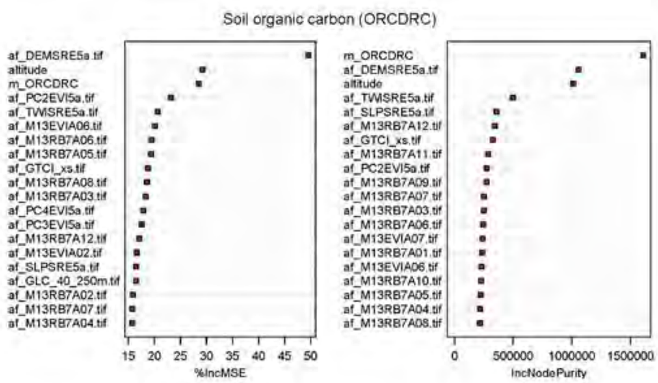


ML - Nonlinear Regression

- **Piecewise curve linear fits**
  - Curved line with the best fit to observed point values
  - Fit is achieved piecewise by Machine Learning for specific combinations of environmental conditions
  - Final fitted line represents mean of many models (ensemble)

Predictive Soil Mapping Basics

Open Geo HUB
Connect · Create · Share · Repeat



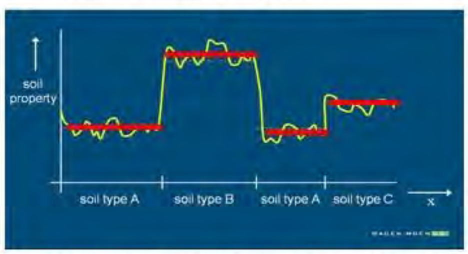
Random Forests Models

- **Importance Plots**
  - Permit interpretation
  - The most important covariates come to top
    - If you don't obscure by transforming (PCA)
  - Most important grid resolutions come to top
    - If you use a pyramid of different grid resolutions

Predictive Soil Mapping Basics

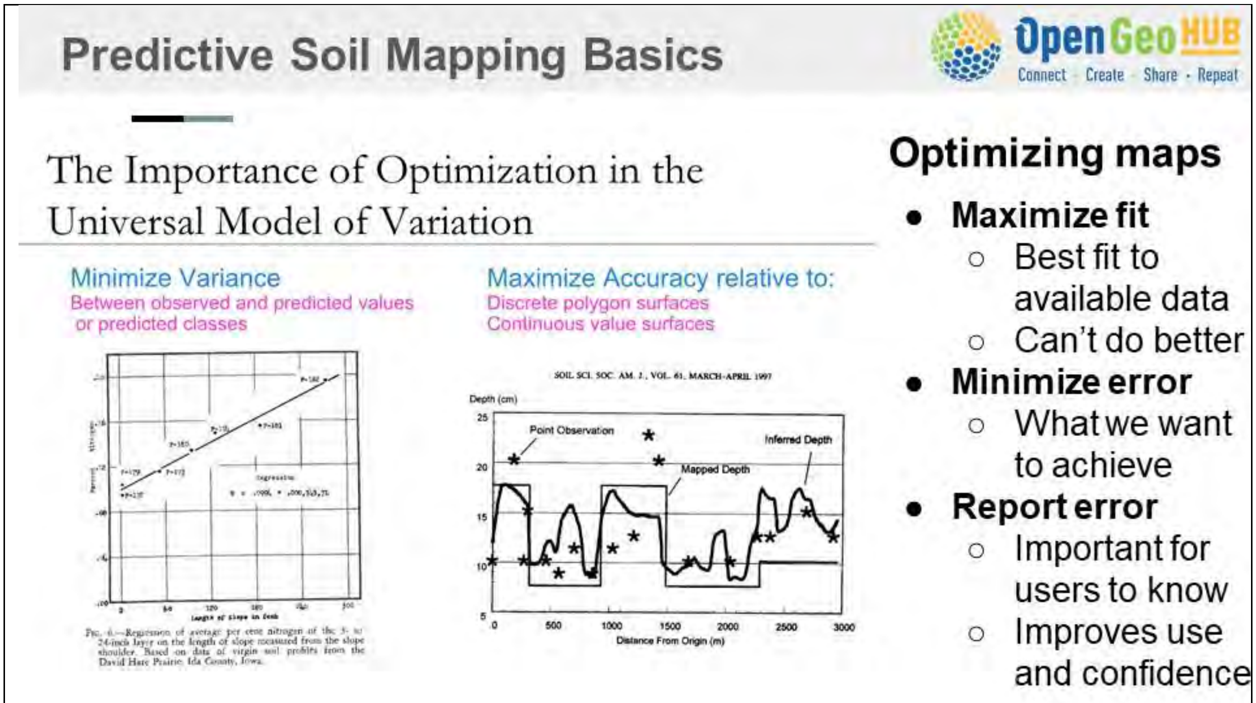
Open Geo HUB
Connect · Create · Share · Repeat

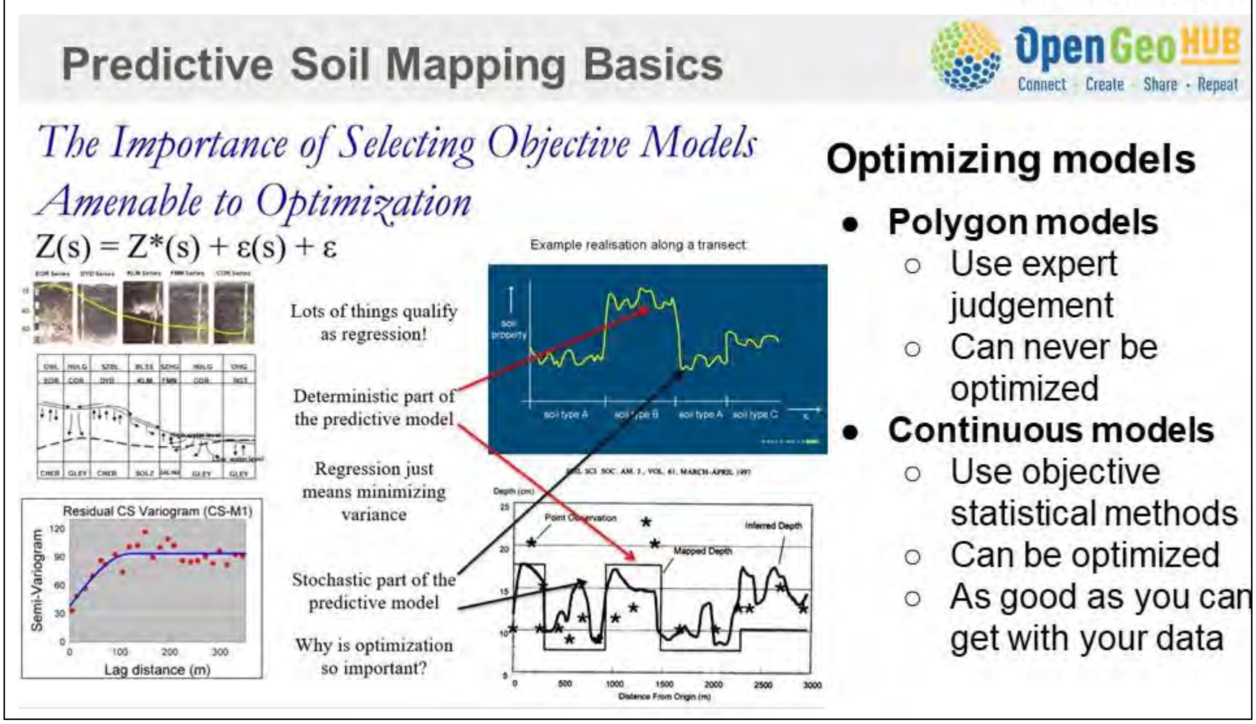
Example realisation along a transect:



Partitioning Variance

- **Polygon Maps**
  - Mean value across the whole polygon
  - Within polygon variance not described
  - Similar proportions of similar soils = similar mean polygon values
- **Continuous Maps**
  - Better capture local scale variation
  - But error remains


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
The Importance of Optimization in the Universal Model of Variation
Minimize Variance
Between observed and predicted values or predicted classes
Maximize Accuracy relative to:
Discrete polygon surfaces
Continuous value surfaces
SOIL SCI. SOC. AM. J., VOL. 61, MARCH-APRIL 1997
Depth (cm)
Point Observation
Inferred Depth
Mapped Depth
Distance From Origin (m)
Optimizing maps
Maximize fit
Best fit to available data
Can't do better
Minimize error
What we want to achieve
Report error
Important for users to know
Improves use and confidence


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
The Importance of Selecting Objective Models Amenable to Optimization
Z(s) = Z*(s) + ε(s) + ε
Lots of things qualify as regression!
Deterministic part of the predictive model
Regression just means minimizing variance
Stochastic part of the predictive model
Why is optimization so important?
Example realisation along a transect
soil property
soil type A
soil type B
soil type A
soil type C
Residual CS Variogram (CS-M1)
Semi-Variogram
Lag distance (m)
SOIL SCI. SOC. AM. J., VOL. 61, MARCH-APRIL 1997
Depth (cm)
Point Observation
Inferred Depth
Mapped Depth
Distance From Origin (m)
Optimizing models
Polygon models
Use expert judgement
Can never be optimized
Continuous models
Use objective statistical methods
Can be optimized
As good as you can get with your data

Predictive Soil Mapping Basics

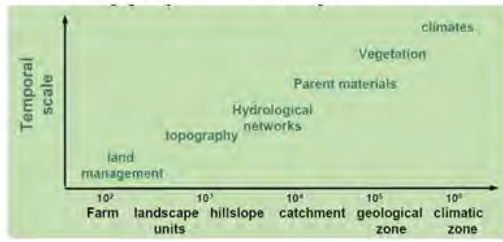
Resolution & Process Scales

Step 1: identifying process scales

Summary of semi-variogram analyses

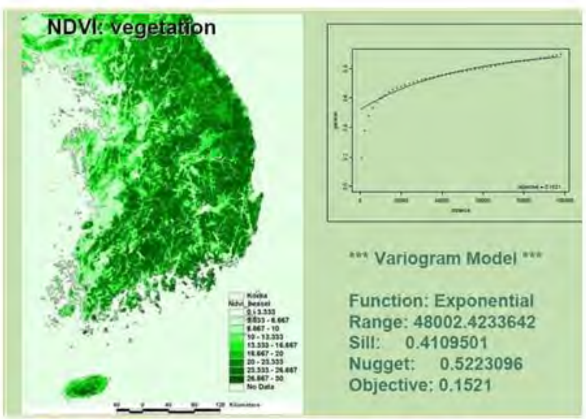
	Variables	Range(m)	sill	nuggett	model	objectives
70km	Rain	69,643.60	0.914	0.012	Gaussian	0.0232
	Temperature	71,217.74	0.499	0.042	Gaussian	0.0165
50km	Elevation	53,645.11	0.777	0.271	Exponential	0.0886
	NDVI	48,002.42	0.411	0.522	Exponential	0.1521
30km	Roughness	30,414.36	0.367	0.592	Exponential	0.142
	Slope	28,564.10	0.302	0.664	Exponential	0.0949
25km	Geology	24,605.21	0.793	0.266	Exponential	0.0088
0.6km	Curvature	620.85	1.066	0.000	Spherical	0.2039
	Contributing area	552.78	0.940	0.000	Exponential	0.1869
0.4km	Landuse	355.61	0.363	0.435	Exponential	0.121

- Soil processes operate over characteristic distances
  - Some are long range
  - Some are medium range
  - Some are short range

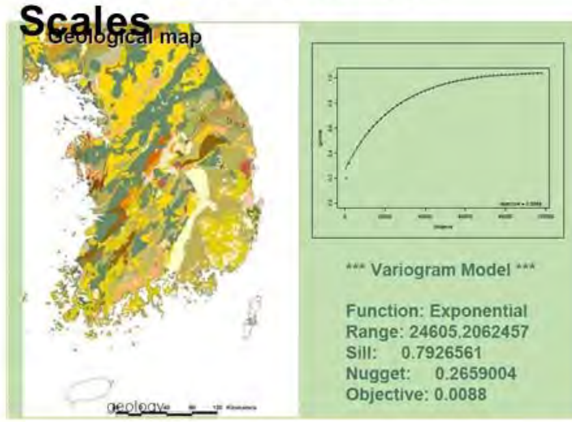


Predictive Soil Mapping Basics

Resolution & Process Scales



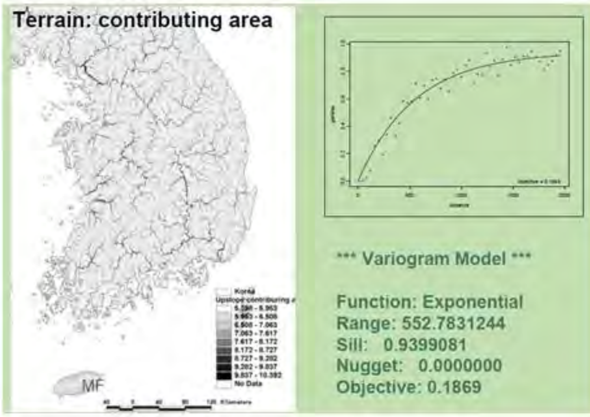
Vegetation: Long range

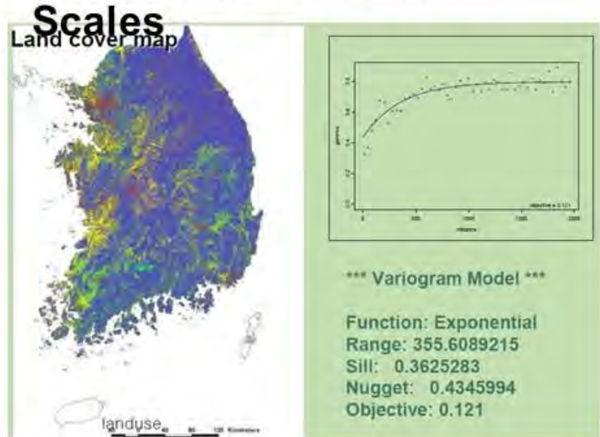


Geology: Medium range

Predictive Soil Mapping Basics

Resolution & Process Scales





Terrain:Medium-short range

Land Cover: Short range

Predictive Soil Mapping Basics

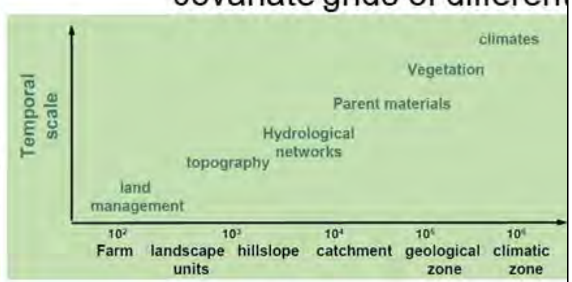
Different Process Scales

- Form a hierarchy and so need different resolution covariates
  - Best way is to produce a Gaussian Pyramid of covariate grids of different

Step 1: identifying process scales

Summary of semi-variogram analyses

	Variables	Range(m)	sill	nuggett	model	objectives
70km	Rain	69,643.60	0.914	0.012	Gaussian	0.0232
	Temperature	71,217.74	0.499	0.042	Gaussian	0.0165
50km	Elevation	53,645.11	0.777	0.271	Exponential	0.0886
	NDVI	48,002.42	0.411	0.522	Exponential	0.1521
30km	Roughness	30,414.36	0.367	0.592	Exponential	0.142
	Slope	28,564.10	0.302	0.664	Exponential	0.0949
25km	Geology	24,605.21	0.793	0.266	Exponential	0.0088
0.6km	Curvature	620.85	1.066	0.000	Spherical	0.2039
	Contributing area	552.78	0.940	0.000	Exponential	0.1869
0.4km	Landuse	355.61	0.363	0.435	Exponential	0.121



Predictive Soil Mapping Basics

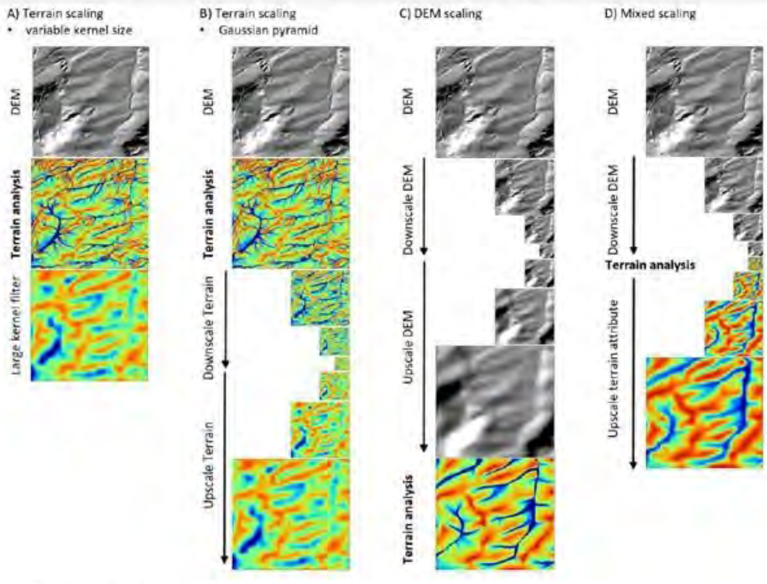
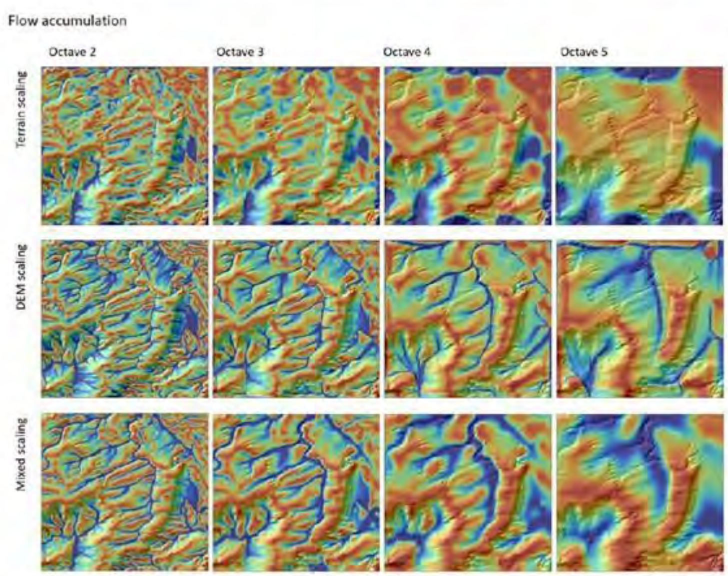


Fig. 1. Scaling approaches

Rescaling Covariates

- A) compute covariate
  - Apply expanding filters
- B) Compute covariate
  - Downscale covariate
  - Upscale covariate
- C) Downscale DEM
  - Upscale DEM
  - Compute covariate
- D) Downscale DEM
  - Compute covariate
  - Upscale covariate

Predictive Soil Mapping Basics



Rescaling Covariates

- B) Terrain Scaling
  - Compute covariate
  - Downscale covariate
  - Upscale covariate
- C) DEM Scaling
  - Downscale DEM
  - Upscale DEM
  - Compute covariate
- D) Mixed Scaling
  - Downscale DEM
  - Compute covariate
  - Upscale covariate

Predictive Soil Mapping Basics

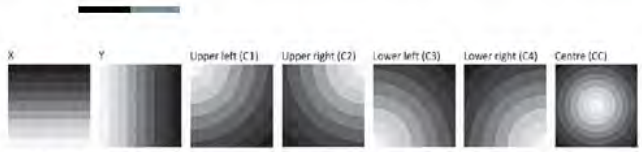
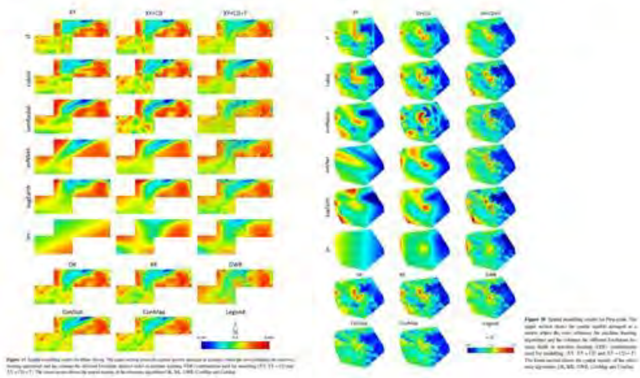
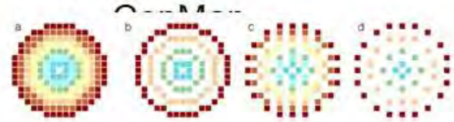


Figure 3 Visualization of the Euclidean distance fields in machine learning (EDM) predictors.



Locational Covariates

- **Euclidean Distance Fields**
  - X,Y location of point
  - Distances to sides and corners
  - Distances between points
- **Spatial context measures**

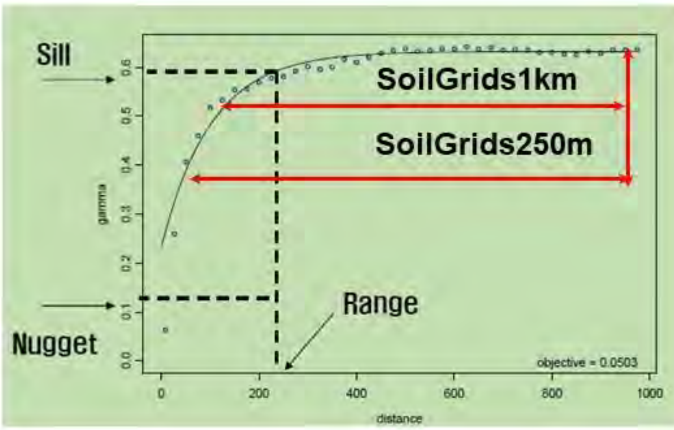


Predictive Soil Mapping Basics

Table 1. SoilGrids average prediction error for key soil properties based on 10–fold cross-validation. N = "*Number of samples used for training*", ME = "*Mean Error*", MAE = "*Mean Absolute Error*", RMSE = "*Root Mean Square Error*" and R-square = "*Coefficient of determination*" (amount of variation explained by the model). For variables with a skew distribution, such as organic carbon, coarse fragments and CEC, the error statistics are also provided on log-scale®.

Variable name	N	Min	Max	ME	MAE	RMSE	R-square	RMSE®	R-square®
Soil organic carbon (gravimetric)	605,054	0	520	-0.292	10.2	32.8	63.5 %	0.715	68.8 %
pH index (H_2O solution)	604,019	2.1	11.0	-0.002	0.4	0.5	83.4 %		
Sand content (gravimetric)	616,762	1 %	94 %	-0.037	9.0	13.1	78.6 %		
Silt content (gravimetric)	613,750	2 %	74 %	0.023	6.7	9.8	79.4 %		
Clay content (gravimetric)	625,159	2 %	68 %	-0.102	6.6	9.5	72.6 %		
Coarse fragments (volumetric)	303,139	0 %	89 %	-0.104	5.5	10.9	55.9 %	1.185	64.3 %
Bulk density (fine earth fraction)	140,596	250	2870	-1.574	108.3	164.7	75.8 %		
Cation-exchange capacity (fine earth fraction)	393,585	0	234	-0.071	5.5	10.3	64.5 %	0.483	67.0 %
Depth to bedrock (in cm)	1,580,798	0	125,000	-29	678	835	54.0 %	1.12	42.8 %

Predictive Soil Mapping Basics



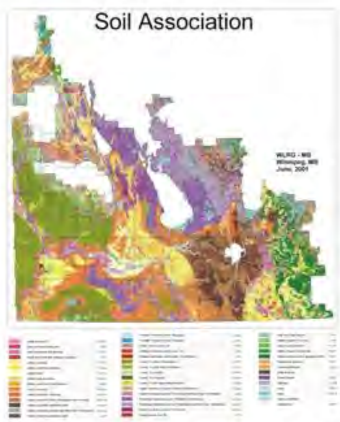
Back to the semi-variogram

- **Explains why SoilGrids works**
  - 1 km maps capture influence of climate/veg
  - 250 m maps capture influence of geology and regional physiography
  - Maps finer than 250 m capture the influence of topography and land use
- **Explains why local maps**

Predictive Soil Mapping Basics

Conventional vs. Pedometric Mapping

Conventional Mapping
Subjective and empirical
Only tries to implement the deterministic part of UMSV (using mental models)
Impossible to replicate
Too many possible outcomes
Impossible to automate
Too many manual decisions
Impossible to optimize
Unlimited number of possible realizations
Can't maximize fit to data



Conventional Maps

- **Expert judgement**
  - Subjective
  - No two alike
- **Static maps**
  - Very difficult to map changes
- **Can't update easily**
  - One and done
  - Have to return to field and

Predictive Soil Mapping Basics

Conventional vs. Pedometric Mapping

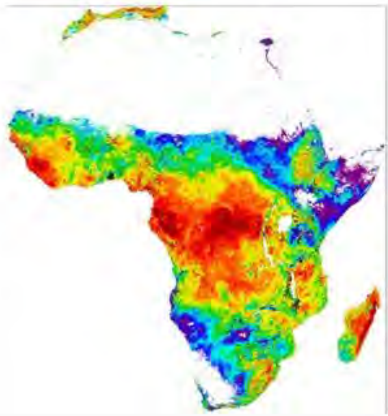
Pedometric Mapping

Objective and systematic
Implements both parts Deterministic & Stochastic

Completely reproducible
Models can be re-run

Can be automated
Fully automated workflows

Can be optimized
Multiple models can be run to identify best fit to data
Objective identification of best fit



Predictive Maps

- **Data driven**
  - Objective
  - Reproducible
- **Minimize error**
  - Best fit to available data
  - Multiple runs
- **Easily updated**
  - Get new data
  - Rerun models
  - Get new maps

Predictive Soil Mapping Basics

Rationale Against Conventional Mapping

Essentially its a Religion Now

Based on belief and faith
Subjective assessments/mental models

Can only map static features
Can't handle time, management

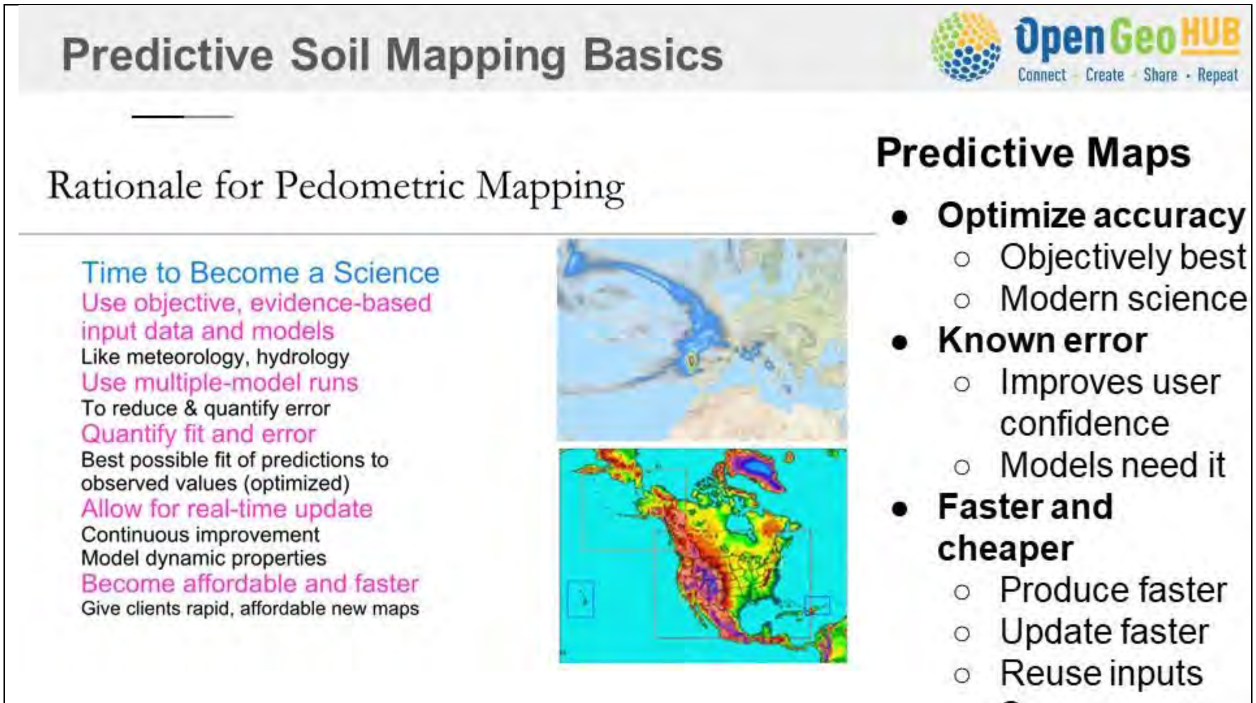
Can't objectively optimize
Impossible to guarantee best fit of predictions to observed values

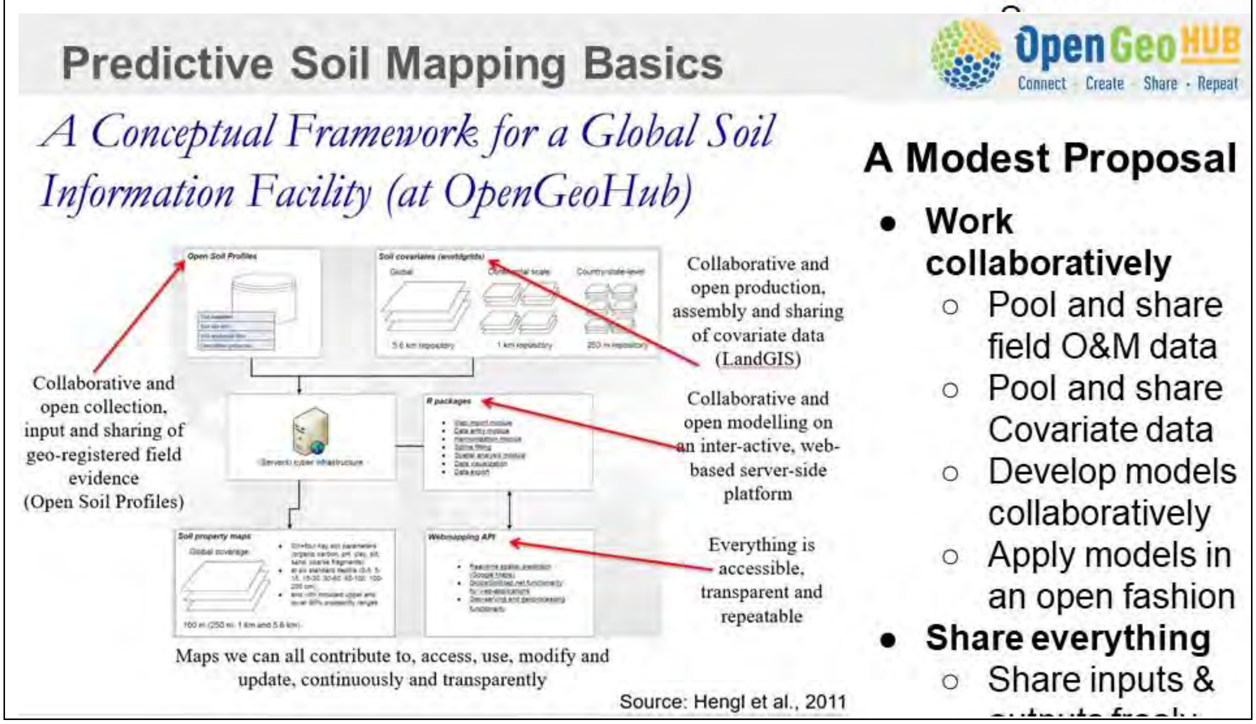
Very difficult to update
Takes years to re-map an area
Can't reuse/recycle existing data

Its too slow and too expensive
Can no longer sell to clients at current costs

Conventional Maps

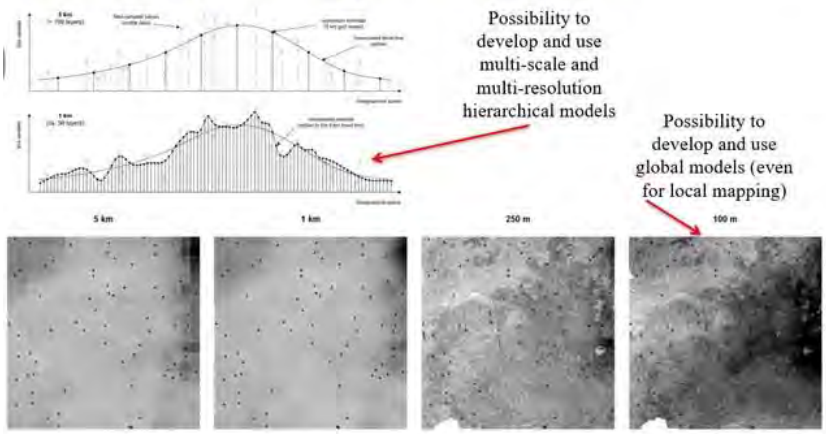
- **Time to redesign**
  - Make data driven
  - Make objective
- **Time to speed up**
  - Map faster
  - Update faster
  - Distribute faster
- **Time to map change**
  - Changes in time
  - Changes due to management


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
Rationale for Pedometric Mapping
Time to Become a Science
Use objective, evidence-based input data and models
Like meteorology, hydrology
Use multiple-model runs
To reduce & quantify error
Quantify fit and error
Best possible fit of predictions to observed values (optimized)
Allow for real-time update
Continuous improvement
Model dynamic properties
Become affordable and faster
Give clients rapid, affordable new maps
Predictive Maps
Optimize accuracy
Objectively best
Modern science
Known error
Improves user confidence
Models need it
Faster and cheaper
Produce faster
Update faster
Reuse inputs


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
A Conceptual Framework for a Global Soil Information Facility (at OpenGeoHub)
Collaborative and open production, assembly and sharing of covariate data (LandGIS)
Collaborative and open collection, input and sharing of geo-registered field evidence (Open Soil Profiles)
Collaborative and open modelling on an inter-active, web-based server-side platform
Everything is accessible, transparent and repeatable
Maps we can all contribute to, access, use, modify and update, continuously and transparently
Source: Hengl et al., 2011
A Modest Proposal
Work collaboratively
Pool and share field O&M data
Pool and share Covariate data
Develop models collaboratively
Apply models in an open fashion
Share everything
Share inputs &

Predictive Soil Mapping Basics

The Future: Global to Local, Multi-Scale Modeling of Soil Properties or Classes



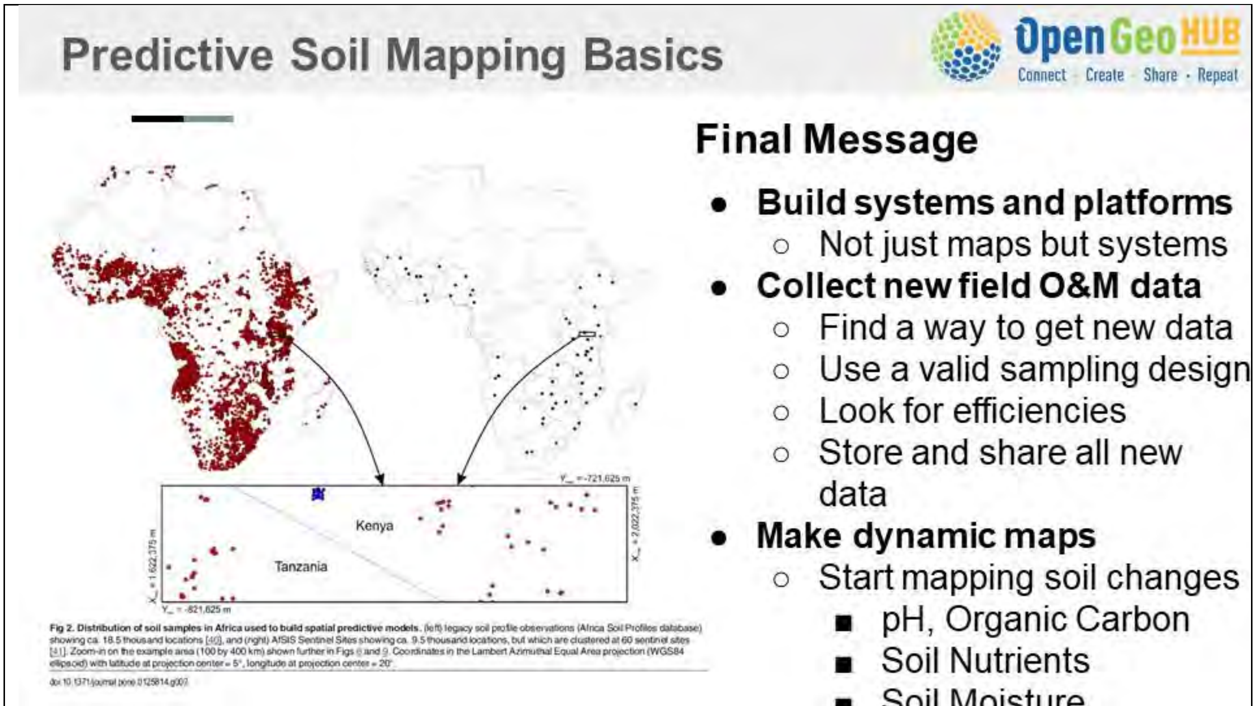
Source: Hengl et al., 2011

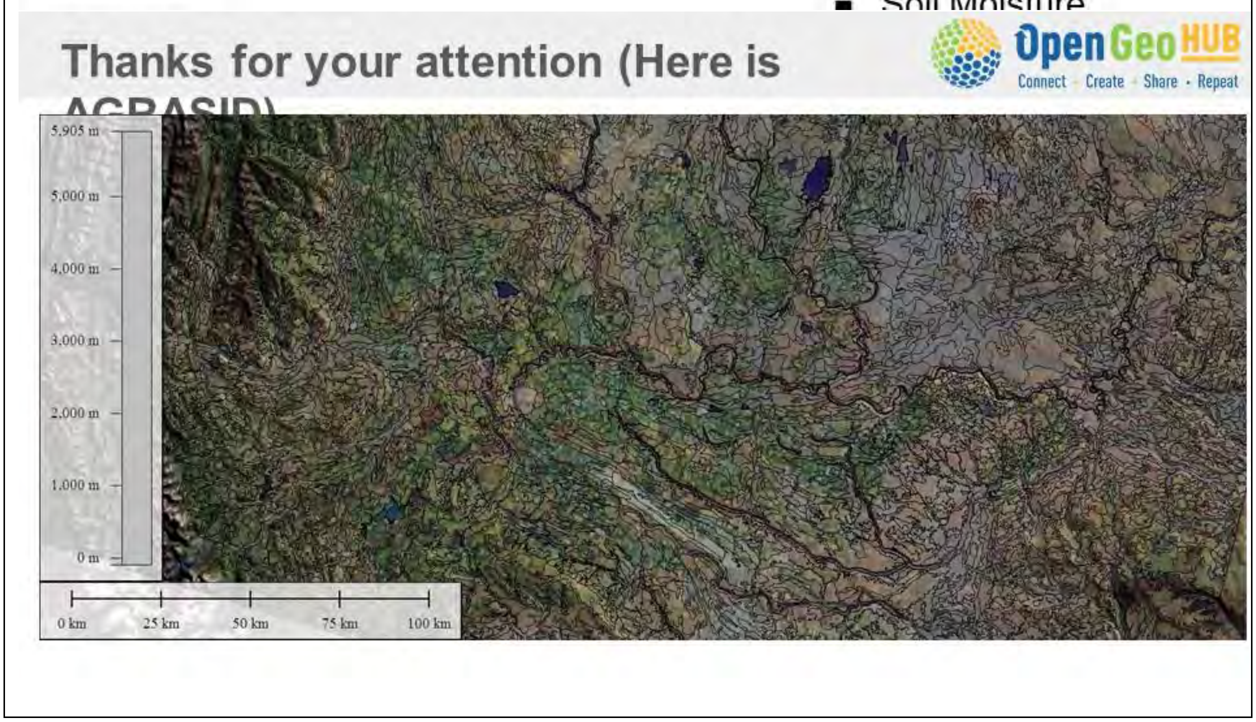
Future Possibilities

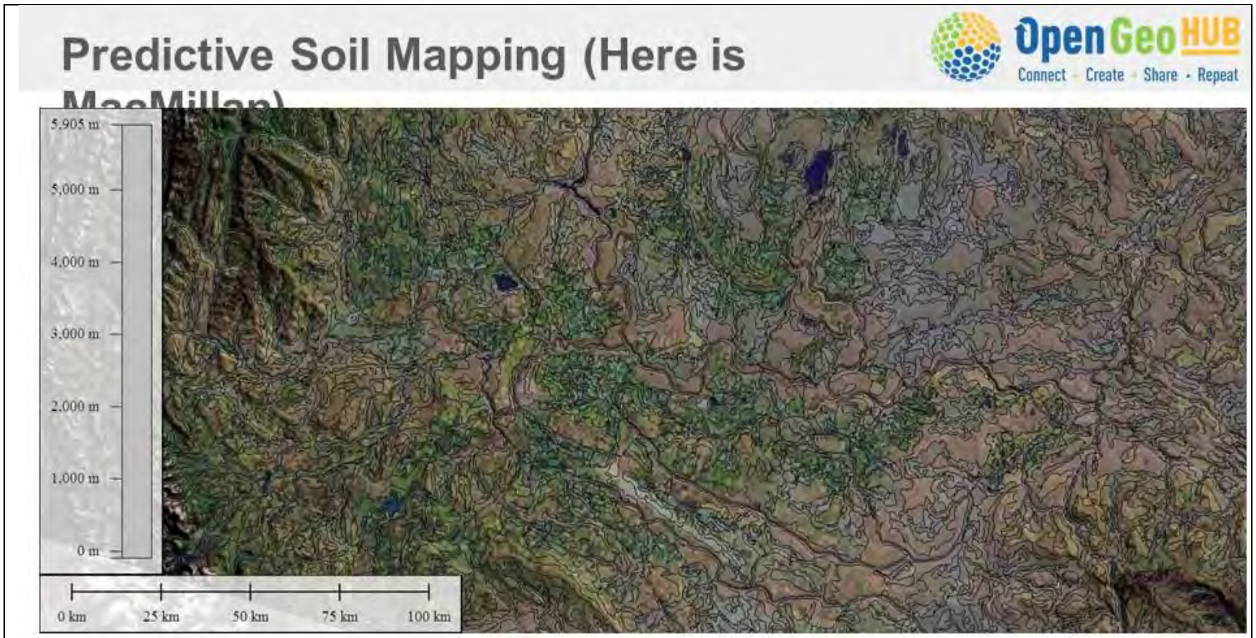
- **Global to local maps**
  - Global maps establish priors
  - Provide context
  - Local maps provide detail
- **Local to global data**
  - Local detail and data can improve global maps
  - Feed local data

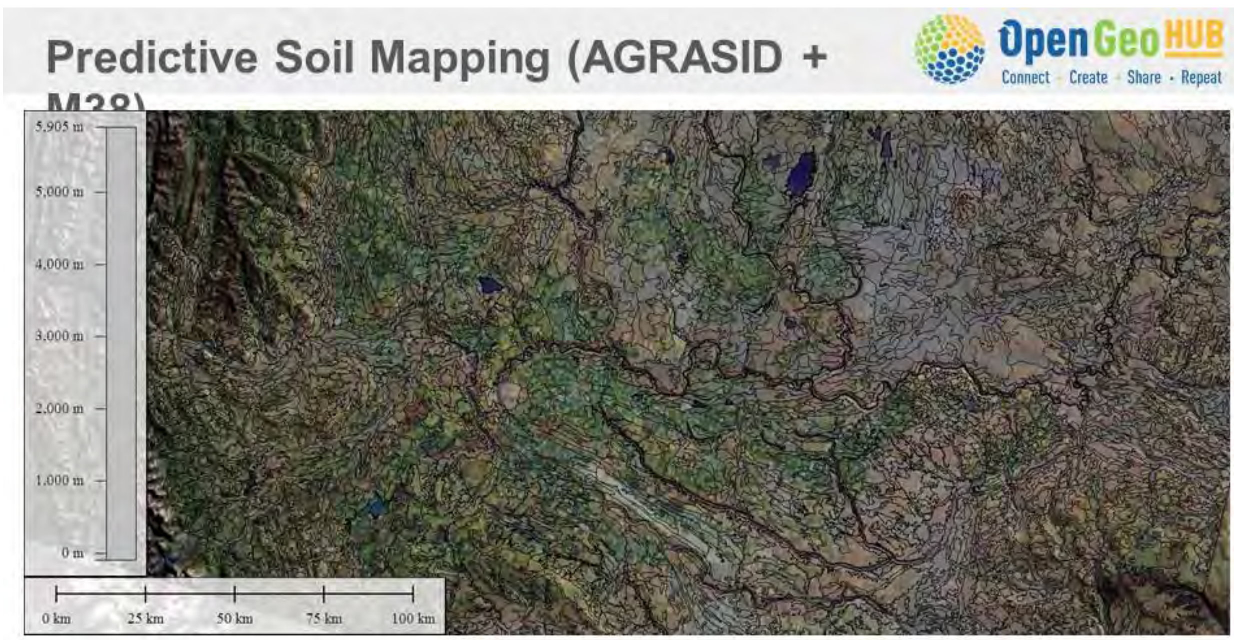
Predictive Soil Mapping: Lessons Learned

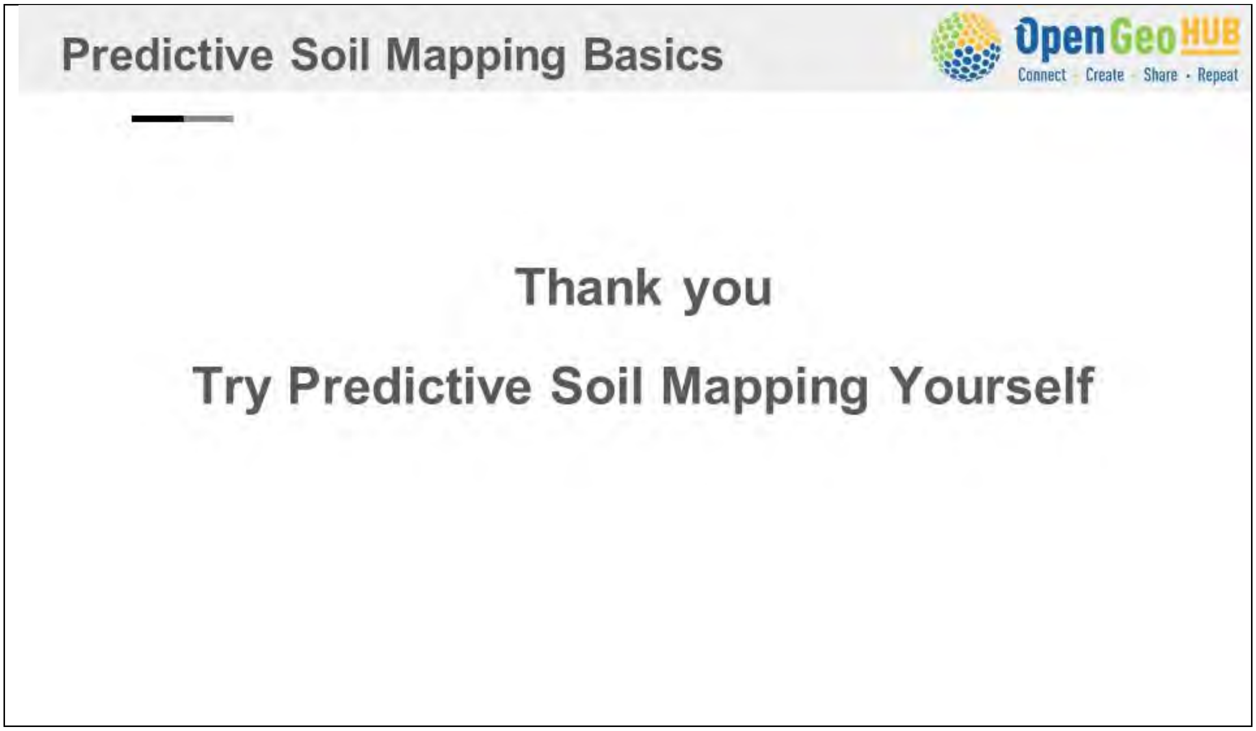
- **Traditional soil mappers mostly <u>got it right!</u>:**
  - Globally, soils nest within spatial hierarchies defined by combinations of **C**limate, **V**egetation, **R**elief, **P**arent Material and **T**ime (Jenny's **CLORPT** model and **scorpan** model)
  - Locally the distribution of soil classes is mainly controlled by topography (as per the soil-landscape model).
- **Mostly, predictive soil mapping seems to be confirming and validating past empirical tacit knowledge**
  - But quantitatively and objectively
  - With known and reported accuracy and spatially explicit error
  - Faster, cheaper and more reproducible/updatable
  - And with new opportunities (change in time, depth, land use)


Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
Kenya
Tanzania
Y = -621,625 m
Y = -721,625 m
X = 1,622,375 m
X = 2,022,375 m
Fig 2. Distribution of soil samples in Africa used to build spatial predictive models. (left) legacy soil profile observations (Africa Soil Profiles database) showing ca. 18.5 thousand locations [40], and (right) AfSIS Sentinel Sites showing ca. 9.5 thousand locations, but which are clustered at 60 sentinel sites [41]. Zoom-in on the example area (100 by 400 km) shown further in Figs and 9. Coordinates in the Lambert Azimuthal Equal Area projection (WGS84 ellipsoid) with latitude at projection center = 5°, longitude at projection center = 20°.
doi:10.1371/journal.pone.0125814.g007
Final Message
Build systems and platforms
Not just maps but systems
Collect new field O&M data
Find a way to get new data
Use a valid sampling design
Look for efficiencies
Store and share all new data
Make dynamic maps
Start mapping soil changes
pH, Organic Carbon
Soil Nutrients
Soil Moisture

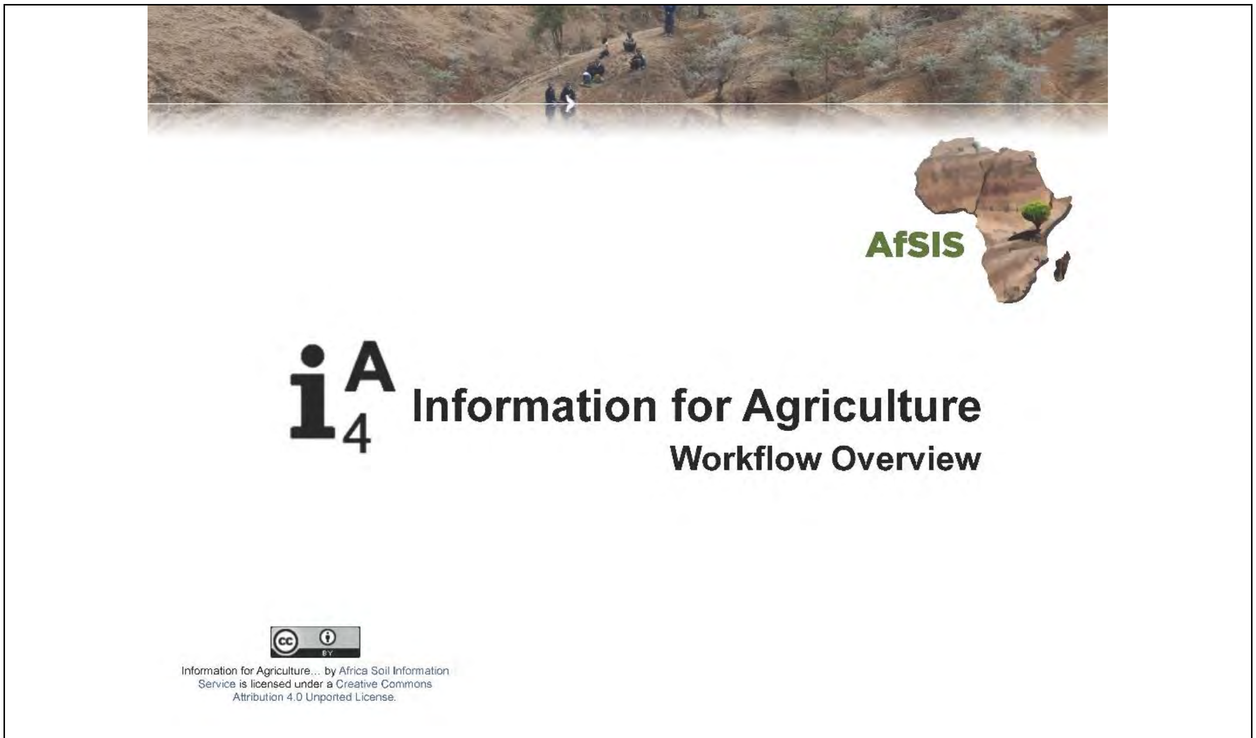

Thanks for your attention (Here is
OpenGeoHUB
Connect · Create · Share · Repeat
5,905 m
5,000 m
4,000 m
3,000 m
2,000 m
1,000 m
0 m
0 km
25 km
50 km
75 km
100 km

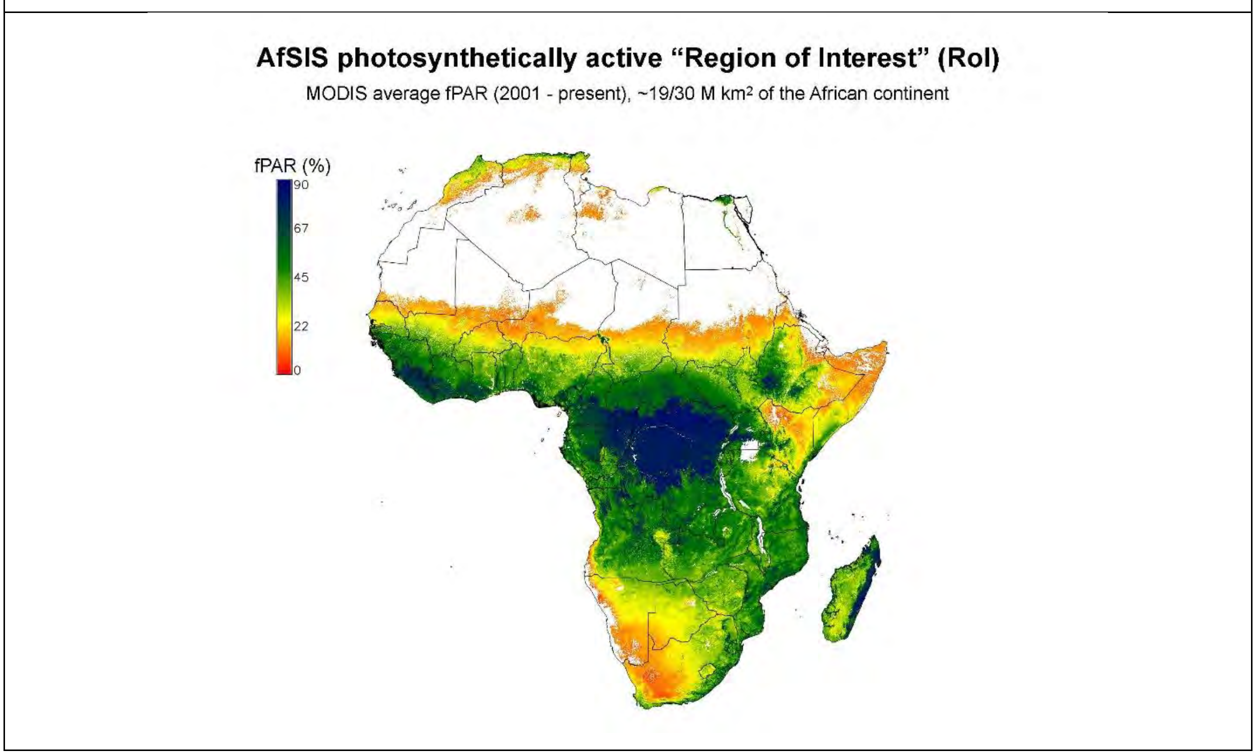

Predictive Soil Mapping (Here is MacMillan)
OpenGeoHUB
Connect · Create · Share · Repeat
5,905 m
5,000 m
4,000 m
3,000 m
2,000 m
1,000 m
0 m
0 km
25 km
50 km
75 km
100 km


Predictive Soil Mapping (AGRASID + M38)
OpenGeoHUB
Connect · Create · Share · Repeat
5,905 m
5,000 m
4,000 m
3,000 m
2,000 m
1,000 m
0 m
0 km
25 km
50 km
75 km
100 km

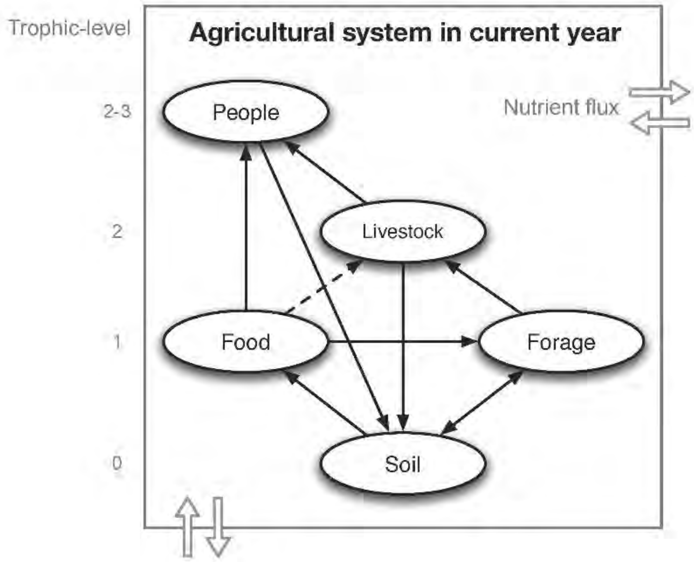

Predictive Soil Mapping Basics
OpenGeoHUB
Connect · Create · Share · Repeat
Thank you
Try Predictive Soil Mapping Yourself

Lessons Learned: What's been tried; Problems and Solutions for Utilizing PSM – Markus Walsh

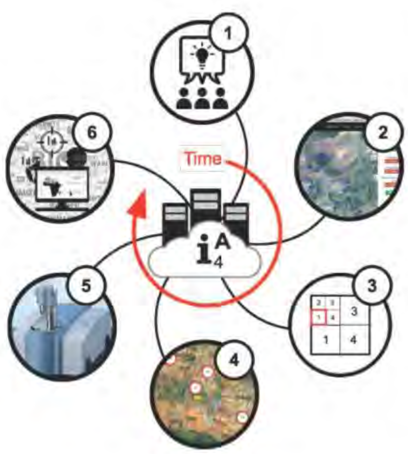




Process-level understanding about agricultural nutrient cycles at national & continental scales

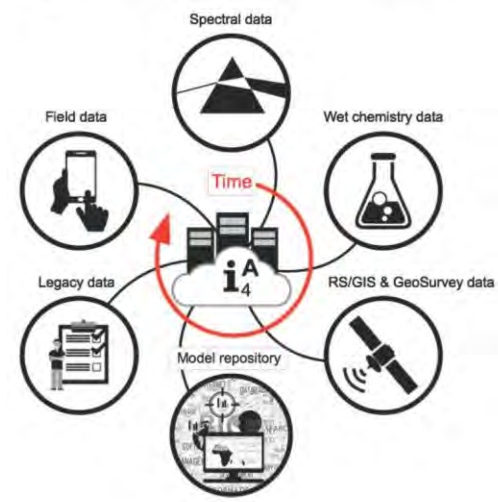


i_4^A Workflow



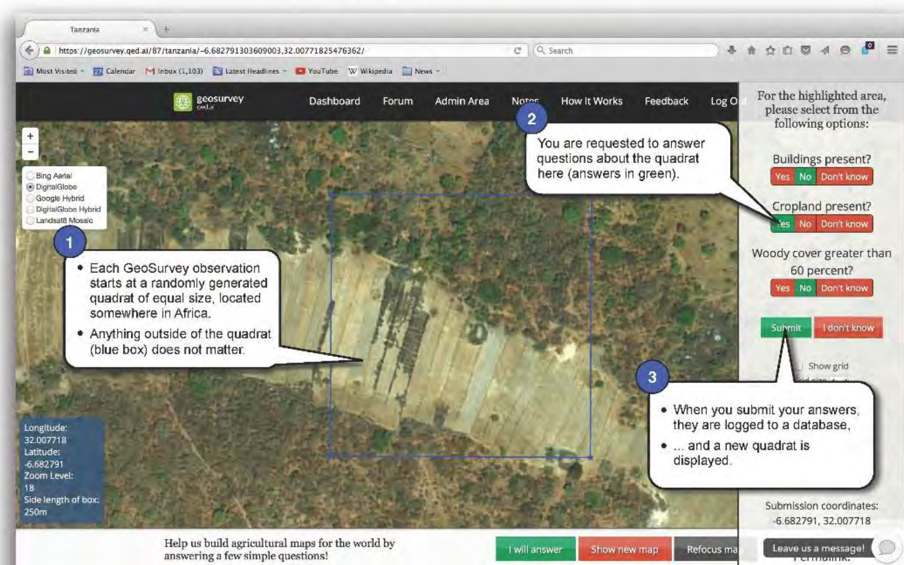
1. Specify information products with end users. Revisit this step frequently!
2. Rapidly assess agricultural system distributions in a region of interest (RoI) with GeoSurvey.
3. Apply spatially & temporally balanced sampling to identify field survey and experimental locations.
4. Use MobileSurvey to log field observations, experiments and the associated physical samples.
5. Analyze all rock, soil, plant and livestock biomarker samples with high throughput lab methods.
6. Predict and monitor system state variables with machine learning and geostatistical models.

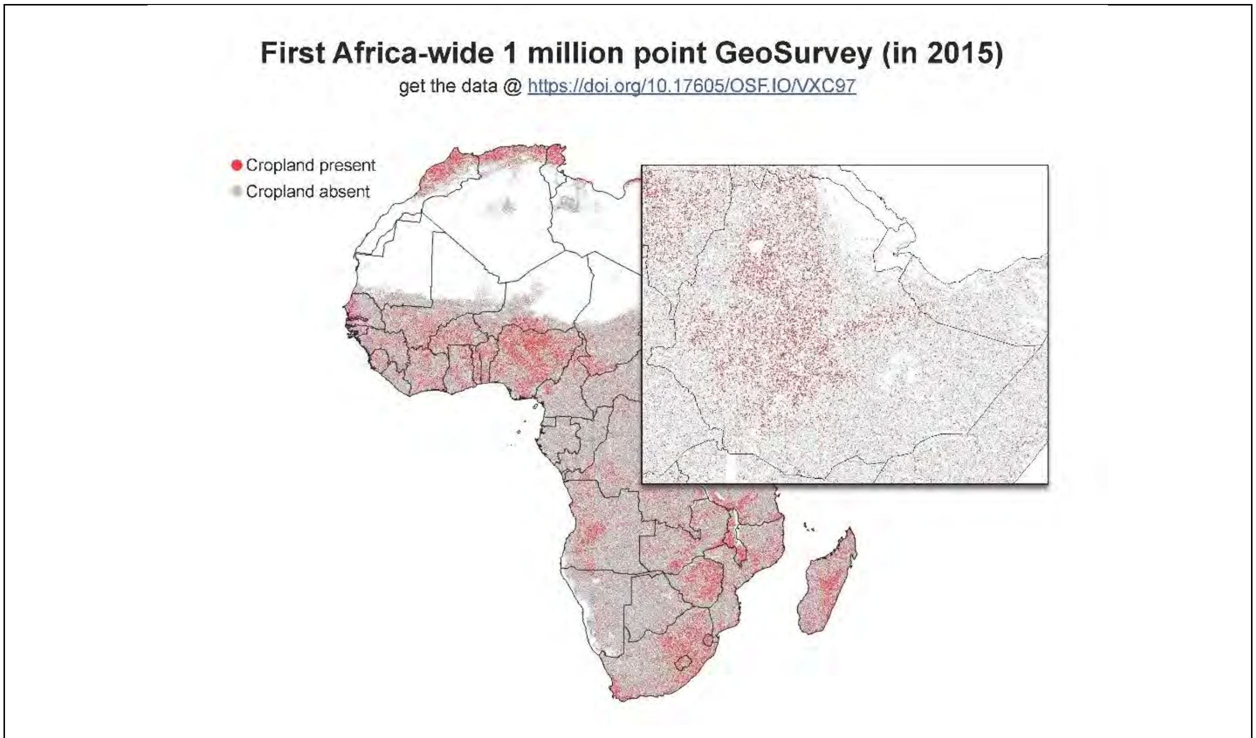
Public data and models

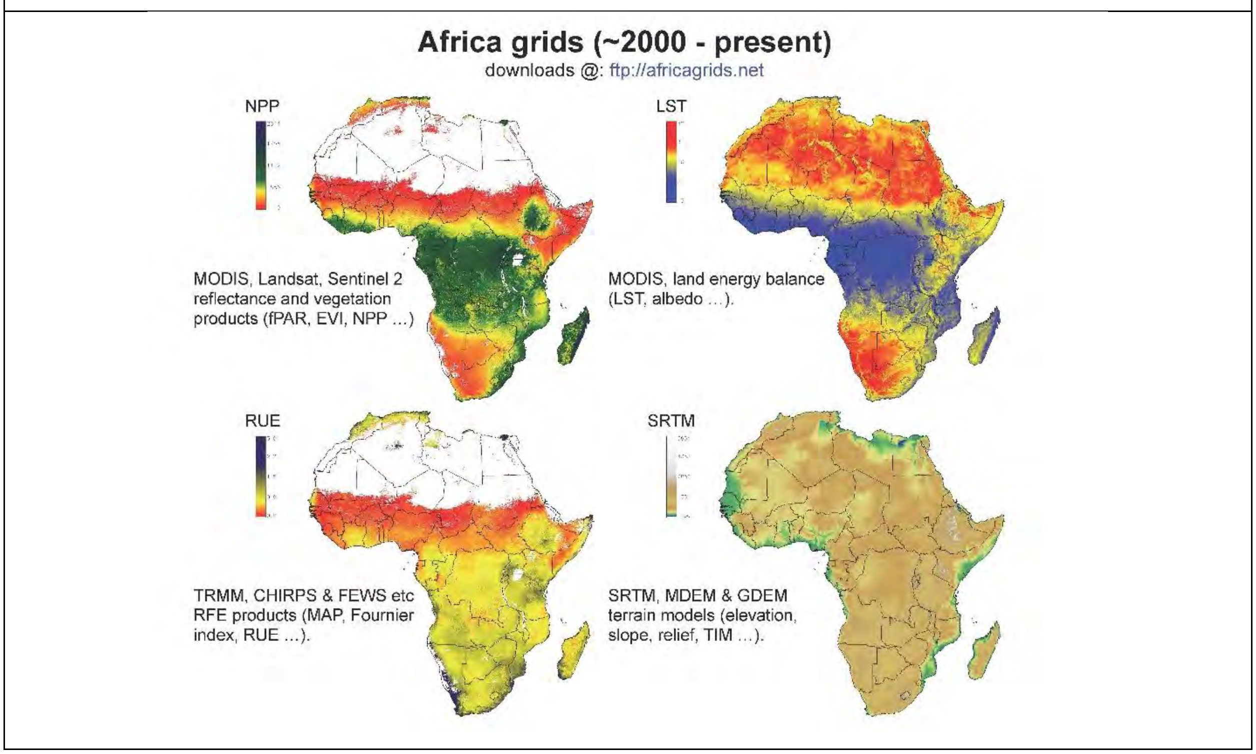


Basic GeoSurvey

sign up and help at: https://geosurvey.qed.ai

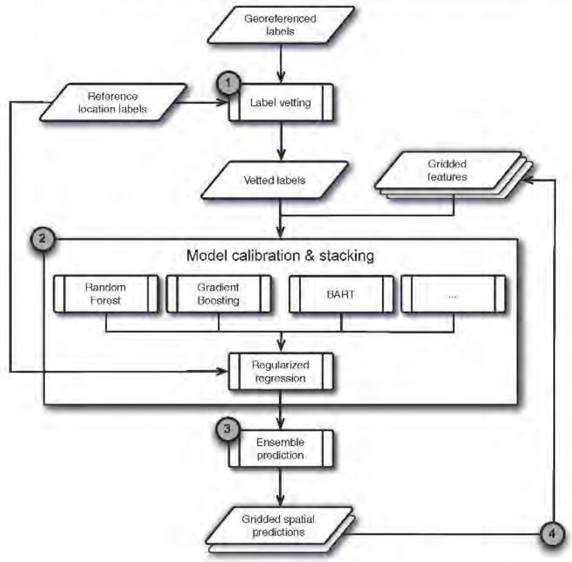



First Africa-wide 1 million point GeoSurvey (in 2015)
get the data @ https://doi.org/10.17605/OSF.IO/VXC97
Cropland present
Cropland absent


Africa grids (~2000 - present)
downloads @: ftp://africagrids.net
NPP
LST
MODIS, Landsat, Sentinel 2 reflectance and vegetation products (fPAR, EVI, NPP ...)
MODIS, land energy balance (LST, albedo ...).
RUE
SRTM
TRMM, CHIRPS & FEWS etc RFE products (MAP, Fournier index, RUE ...).
SRTM, MDEM & GDEM terrain models (elevation, slope, relief, TIM ...).

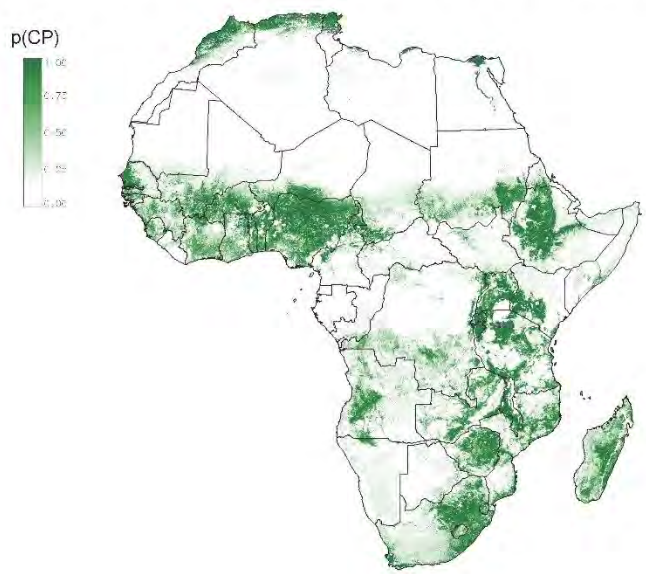
GeoSurvey: ensemble prediction workflow

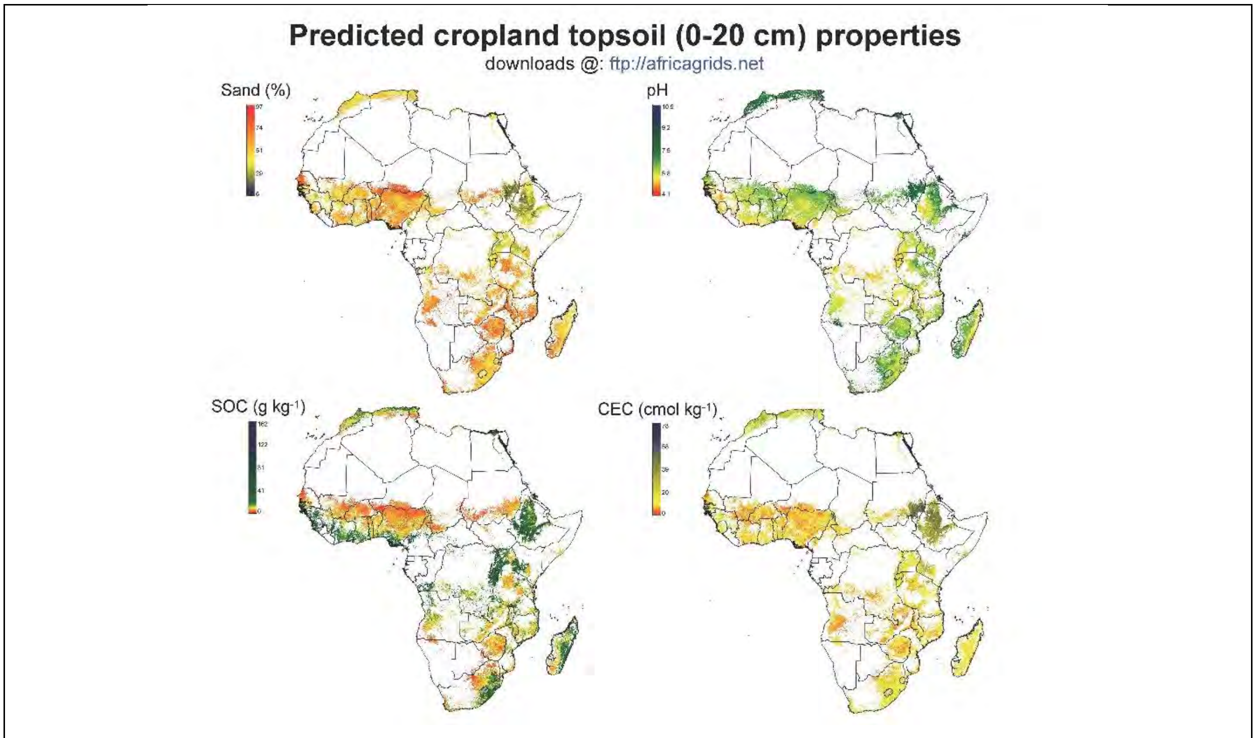
see examples @ https://doi.org/10.17605/OSF.IO/VXC97

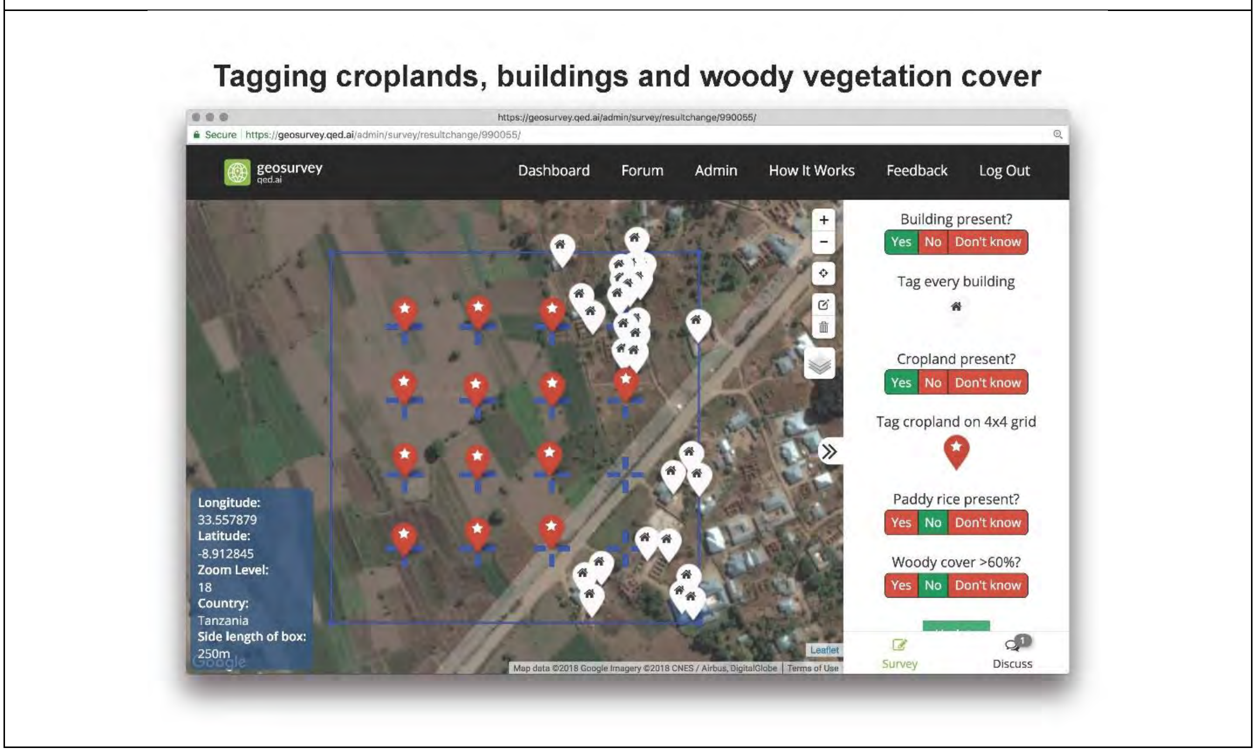


GeoSurvey probability results 2015: Croplands present?

2015 cropland area estimate ~6.7 / 30M km² of the African continent



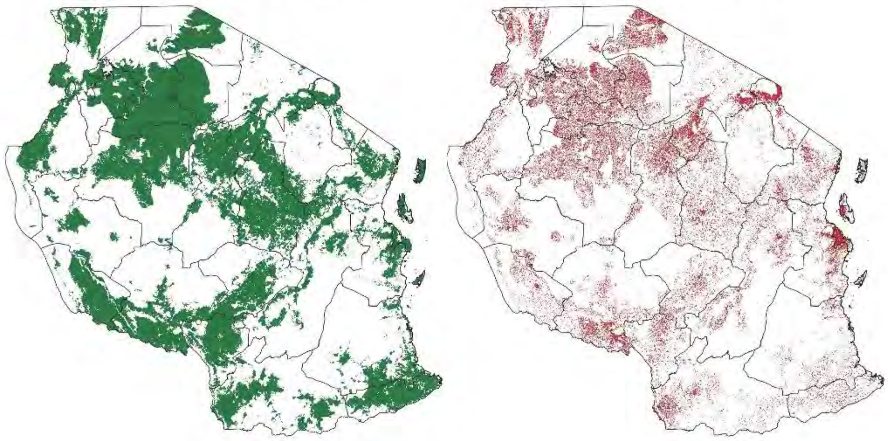

Predicted cropland topsoil (0-20 cm) properties
downloads @: ftp://africagrids.net
Sand (%)
pH
SOC (g kg-1)
CEC (cmol kg-1)


Tagging croplands, buildings and woody vegetation cover
geosurvey
Dashboard
Forum
Admin
How It Works
Feedback
Log Out
Building present?
Yes No Don't know
Tag every building
Cropland present?
Yes No Don't know
Tag cropland on 4x4 grid
Paddy rice present?
Yes No Don't know
Woody cover >60%?
Yes No Don't know
Longitude:
33.557879
Latitude:
-8.912845
Zoom Level:
18
Country:
Tanzania
Side length of box:
250m
Survey
Discuss

Tanzania GeoSurvey: ensemble predictions 2018

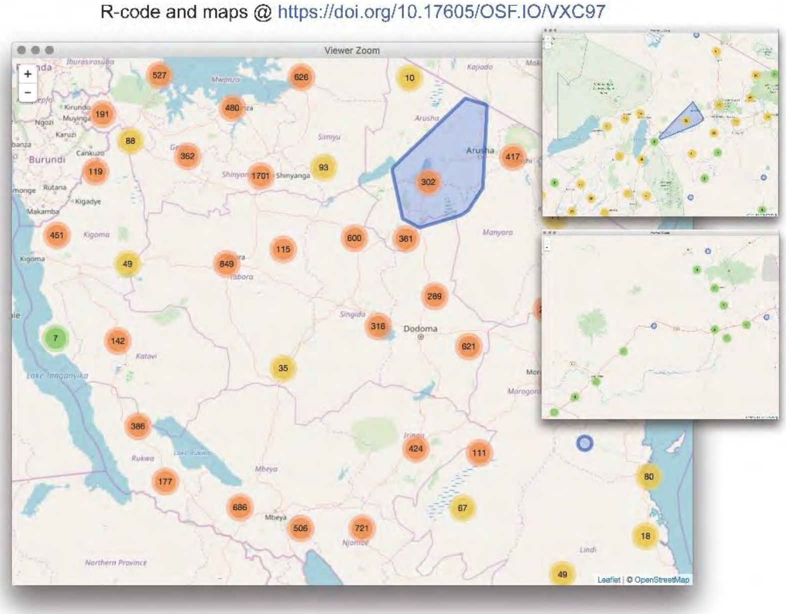
Data, R-code and maps @ https://doi.org/10.17605/OSF.IO/VXC97

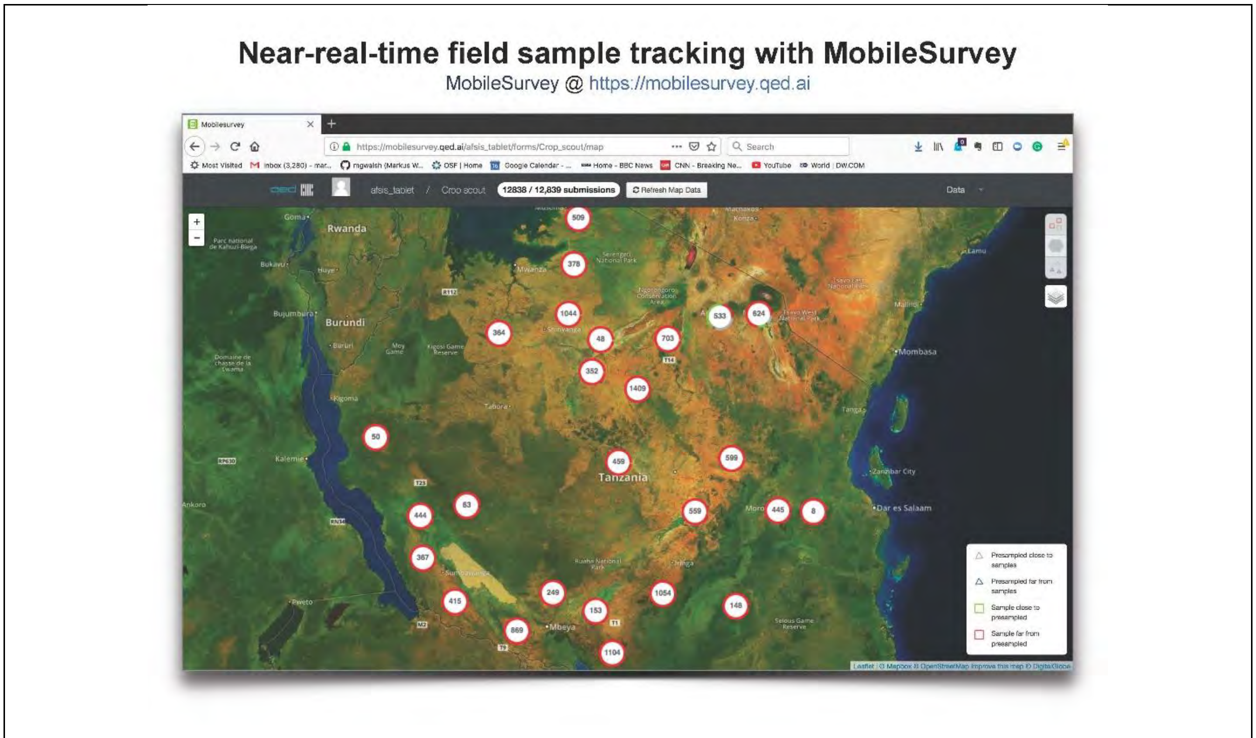
Cropland mask (~34.2 Mha) Settlement mask (~11.5 Mha)

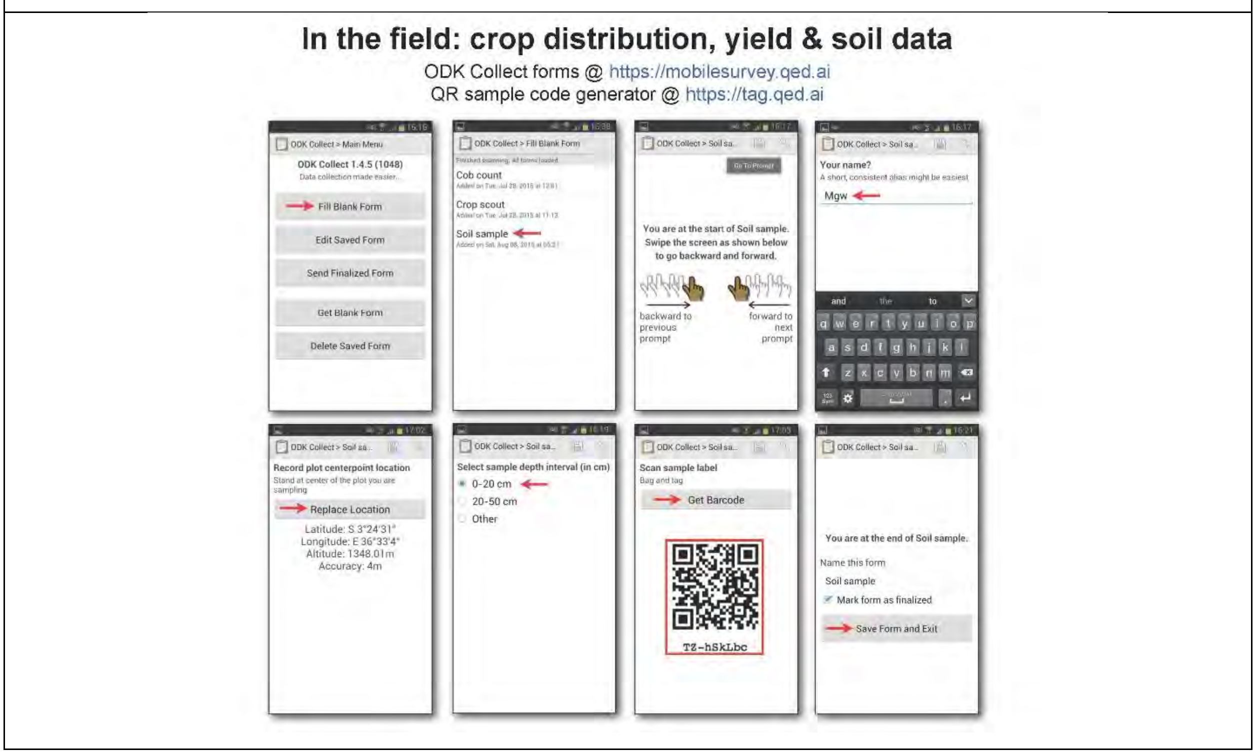


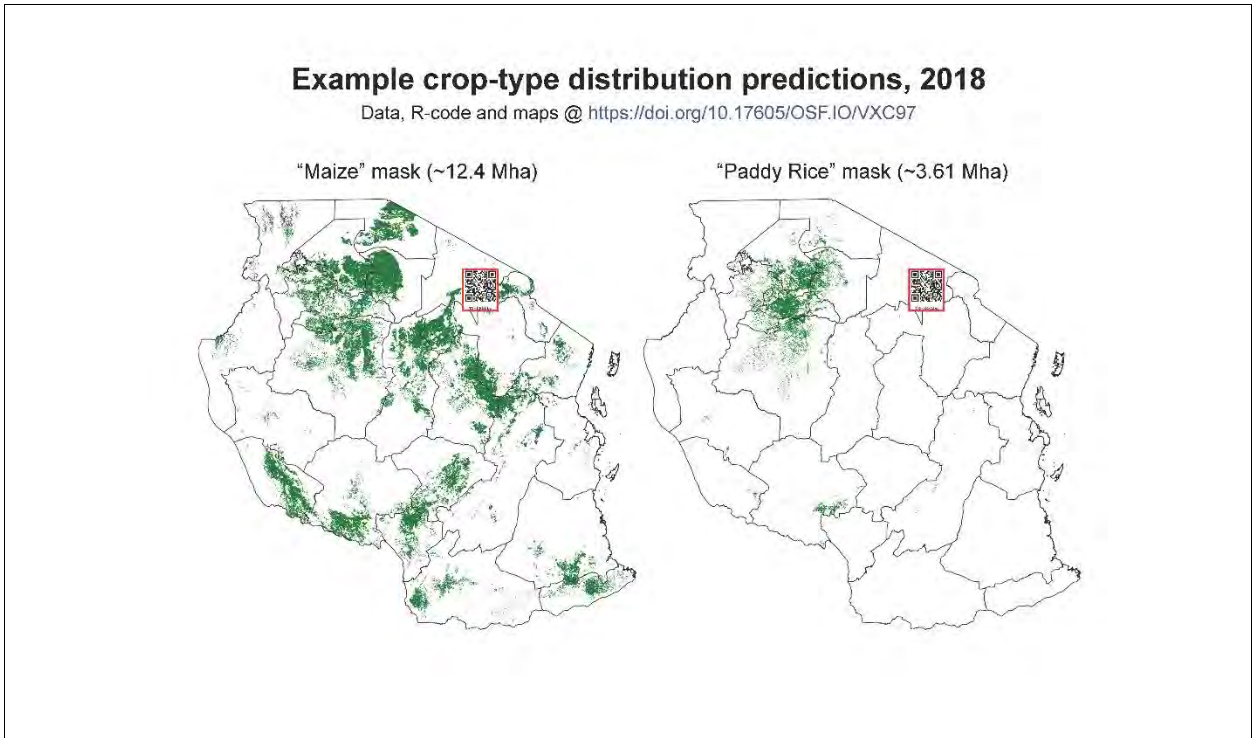
Spatially balanced sampling frames

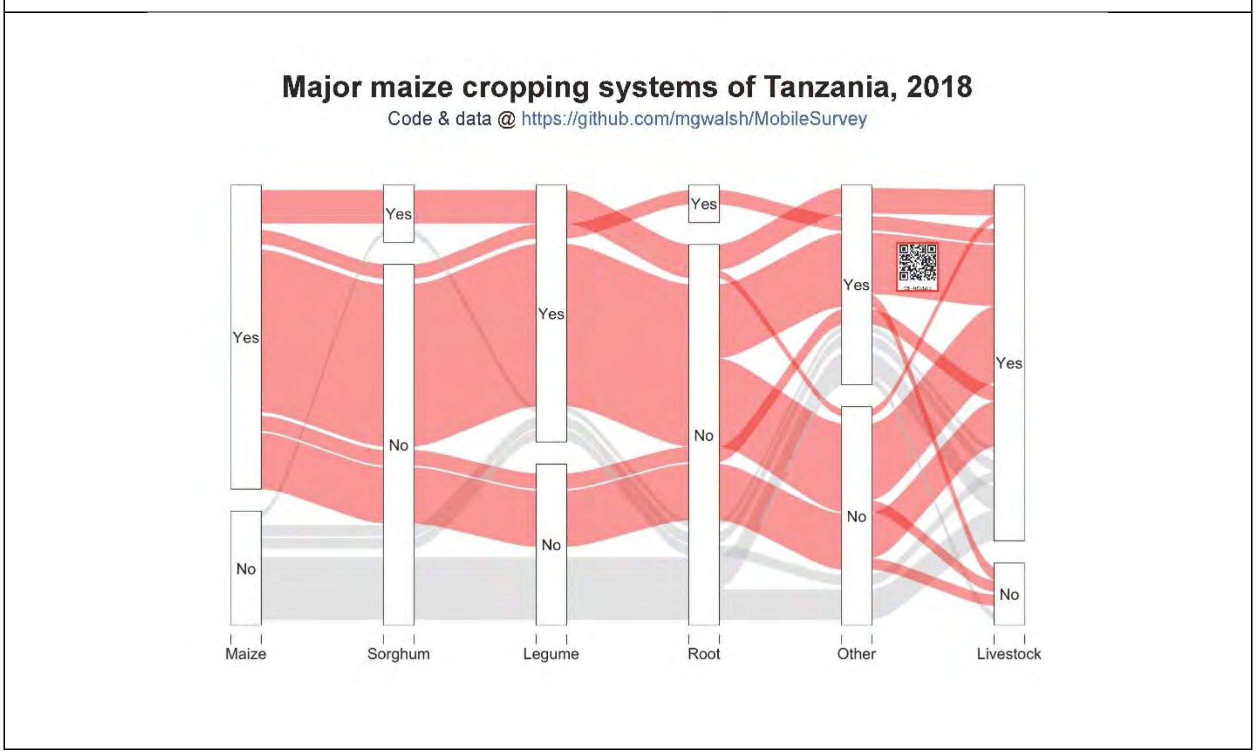
R-code and maps @ https://doi.org/10.17605/OSF.IO/VXC97




Near-real-time field sample tracking with MobileSurvey
MobileSurvey @ https://mobilesurvey.qed.ai
12838 / 12,839 submissions
Refresh Map Data
Rwanda
Burundi
Tanzania
Mombasa
Dar es Salaam
Zanzibar City
Presampled close to samples
Presampled far from samples
Sample close to presampled
Sample far from presampled


In the field: crop distribution, yield & soil data
ODK Collect forms @ https://mobilesurvey.qed.ai
QR sample code generator @ https://tag.qed.ai
ODK Collect > Main Menu
ODK Collect 1.4.5 (1048)
Fill Blank Form
Edit Saved Form
Send Finalized Form
Get Blank Form
Delete Saved Form
ODK Collect > Fill Blank Form
Cob count
Crop scout
Soil sample
You are at the start of Soil sample. Swipe the screen as shown below to go backward and forward.
backward to previous prompt
forward to next prompt
Your name?
A short, consistent alias might be easiest
Mgw
Record plot centerpoint location
Replace Location
Latitude: S 3°24'31"
Longitude: E 36°33'4"
Altitude: 1348.01m
Accuracy: 4m
Select sample depth interval (in cm)
0-20 cm
20-50 cm
Other
Scan sample label
Get Barcode
TZ-hSkLbc
You are at the end of Soil sample.
Name this form
Soil sample
Mark form as finalized
Save Form and Exit


Example crop-type distribution predictions, 2018
Data, R-code and maps @ https://doi.org/10.17605/OSF.IO/VXC97
"Maize" mask (~12.4 Mha)
"Paddy Rice" mask (~3.61 Mha)


Major maize cropping systems of Tanzania, 2018
Code & data @ https://github.com/mgwalsh/MobileSurvey
Yes
No
Yes
No
Yes
No
Yes
No
Yes
No
Yes
No
Maize
Sorghum
Legume
Root
Other
Livestock

Spectral measurements of rock, soil, plant & livestock ionomes

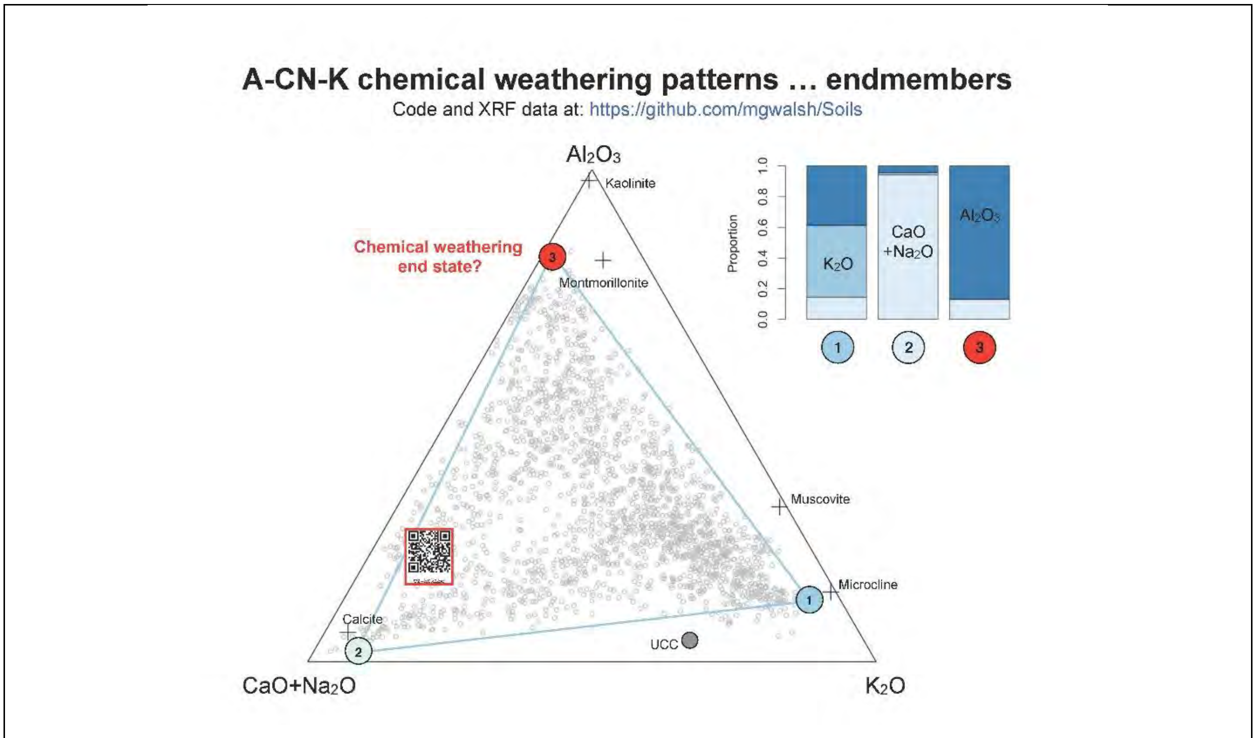
Africa data @ https://registry.opendata.aws/afsis/

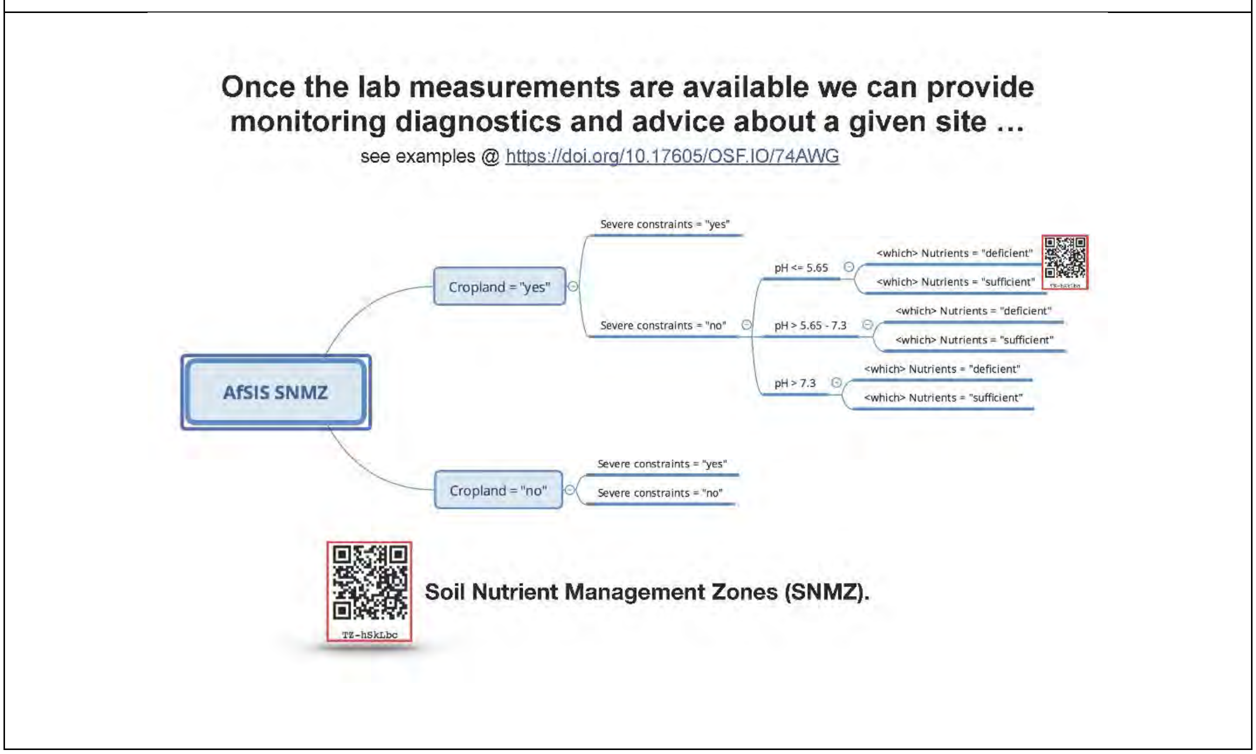
data analysis examples @ https://doi.org/10.17605/OSF.IO/74AWG

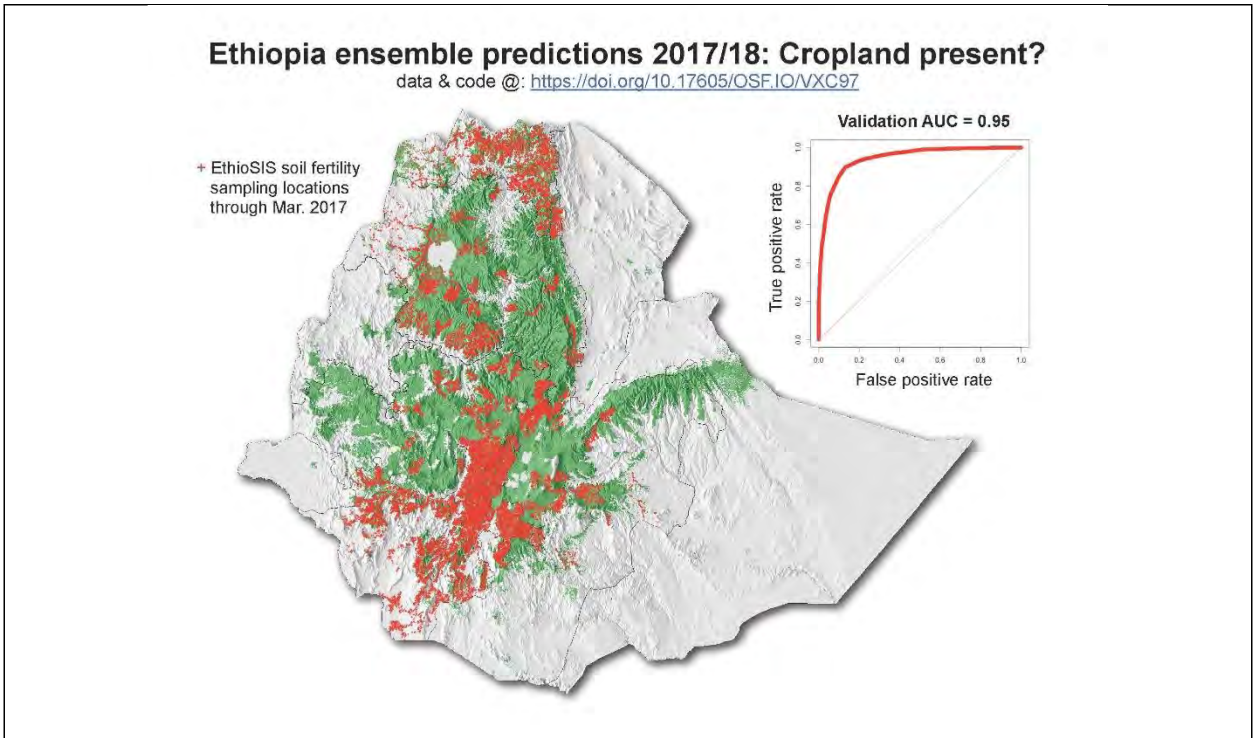
Mid-infrared spectrometer

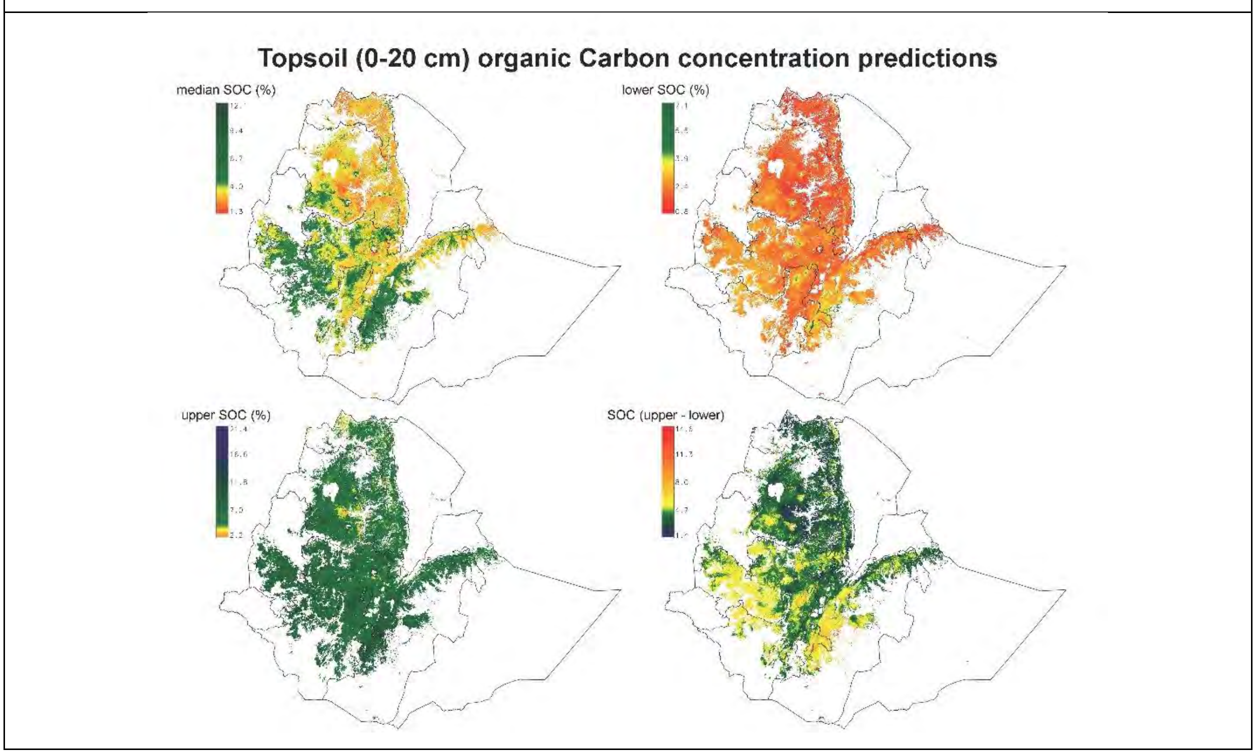
XRF spectrometer

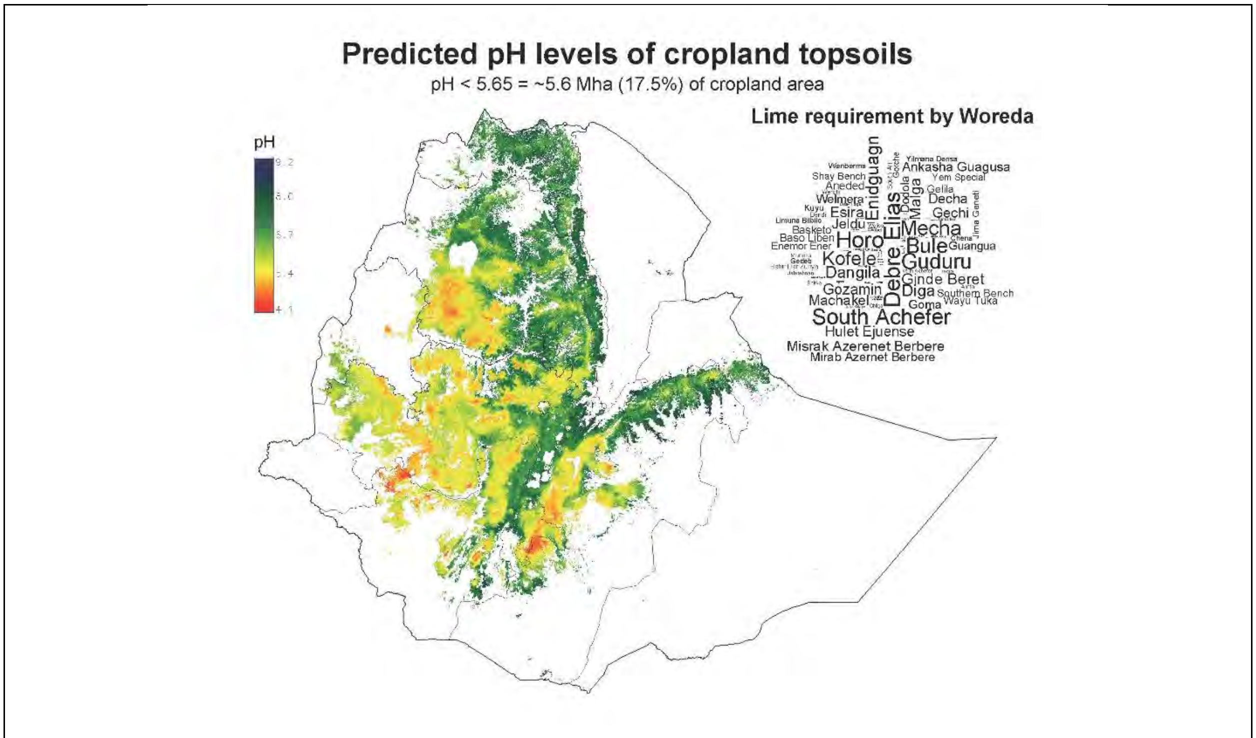
The benefit is that you can run everything reliably and cost effectively in remote, off-grid areas …


A-CN-K chemical weathering patterns … endmembers
Code and XRF data at: https://github.com/mgwalsh/Soils
Al2O3
Kaolinite
Chemical weathering end state?
Montmorillonite
Muscovite
Microcline
Calcite
UCC
CaO+Na2O
K2O
Proportion
K2O
CaO +Na2O
Al2O3

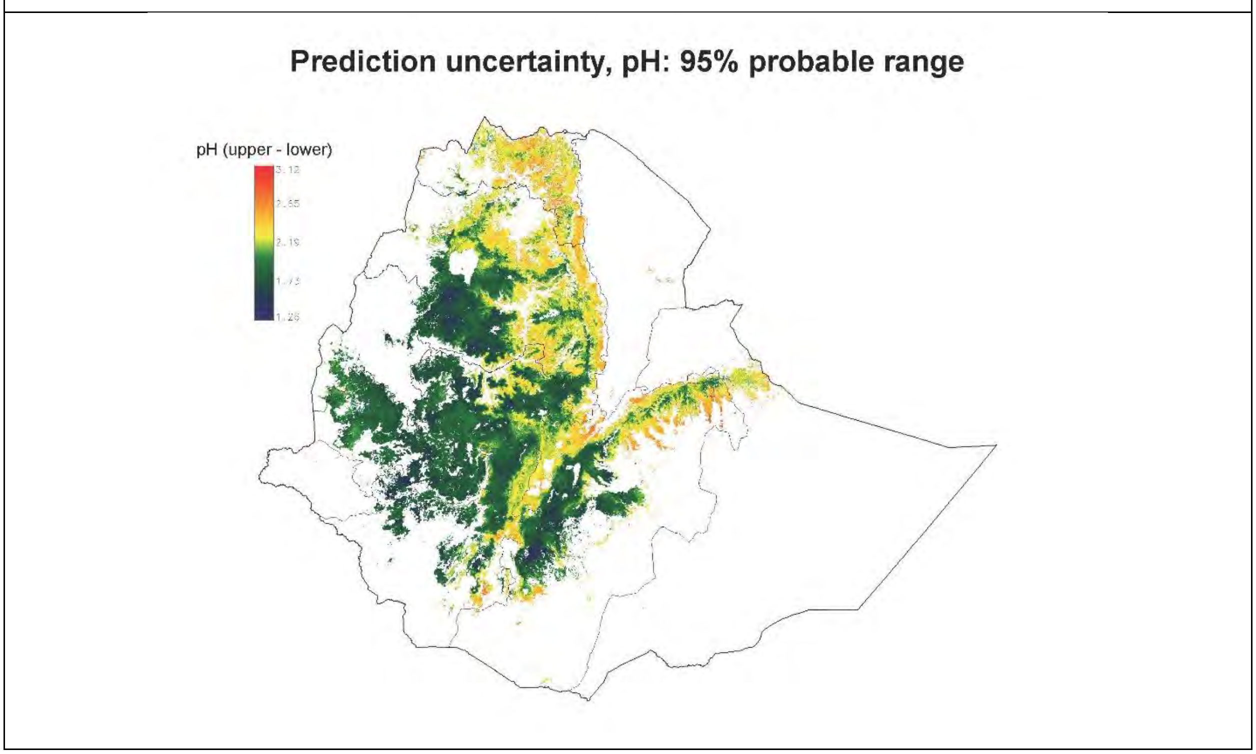

Once the lab measurements are available we can provide monitoring diagnostics and advice about a given site …
see examples @ https://doi.org/10.17605/OSF.IO/74AWG
AfSIS SNMZ
Cropland = "yes"
Cropland = "no"
Severe constraints = "yes"
Severe constraints = "no"
pH <= 5.65
pH > 5.65 - 7.3
pH > 7.3
<which> Nutrients = "deficient"
<which> Nutrients = "sufficient"
Soil Nutrient Management Zones (SNMZ).
TZ-hSkLbc


Ethiopia ensemble predictions 2017/18: Cropland present?
data & code @: https://doi.org/10.17605/OSF.IO/VXC97
+ EthioSIS soil fertility sampling locations through Mar. 2017
Validation AUC = 0.95
True positive rate
False positive rate

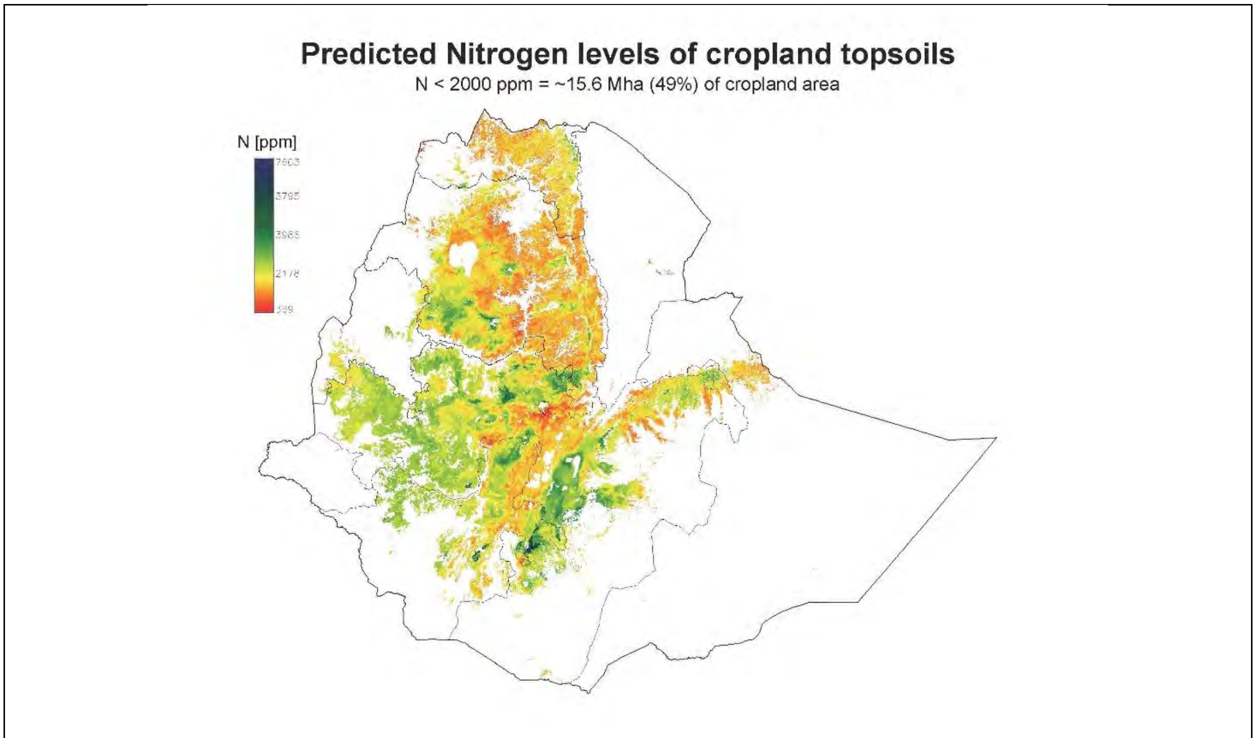

Topsoil (0-20 cm) organic Carbon concentration predictions
median SOC (%)
lower SOC (%)
upper SOC (%)
SOC (upper - lower)

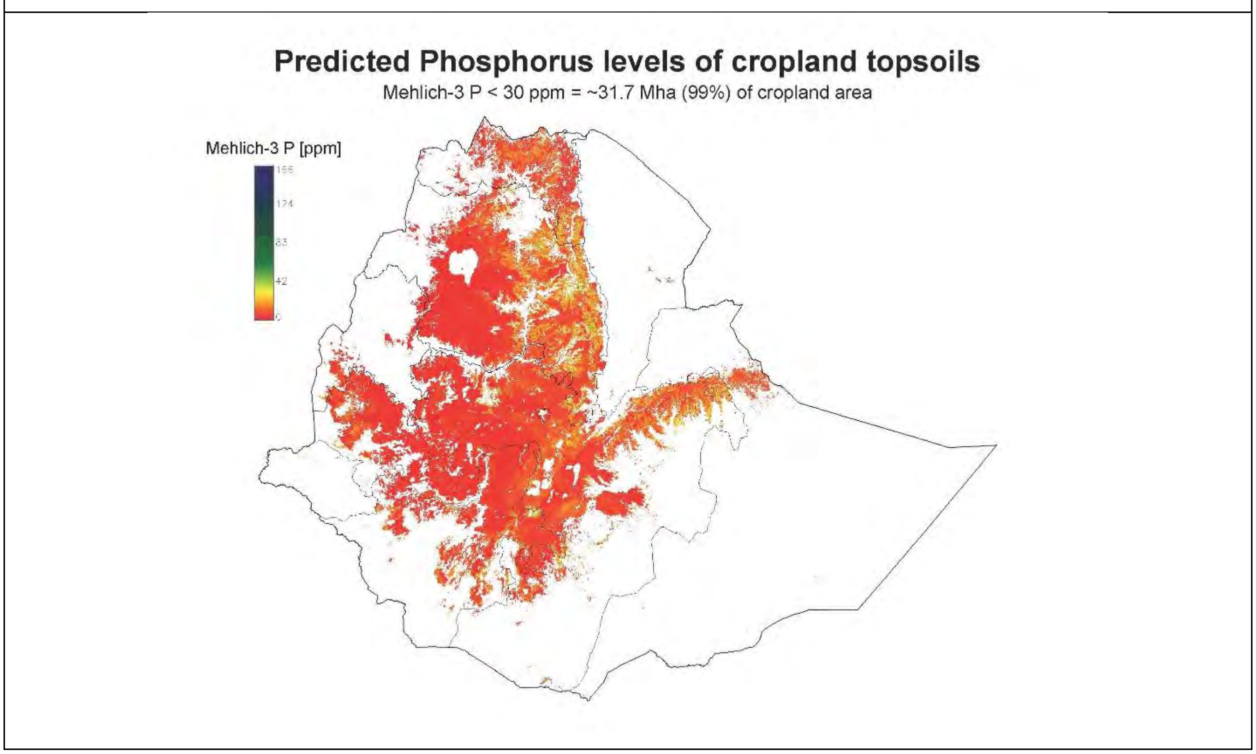


Predicted pH levels of cropland topsoils
pH < 5.65 = ~5.6 Mha (17.5%) of cropland area
Lime requirement by Woreda
pH
Enidguagn
Debre Elias
Horo
Mecha
Bule
Guduru
Kofele
Dangila
Gozamin
Machakel
Diga
Ginde Beret
South Achefer
Hulet Ejuense
Misrak Azerenet Berbere
Mirab Azernet Berbere
Ankasha Guagusa
Decha
Gechi
Esira
Jeldu
Welmera
Aneded
Shay Bench
Malga
Guangua
Goma
Wayu Tuka
Southern Bench
Basketo
Baso Liben
Enemor Ener
Yem Special



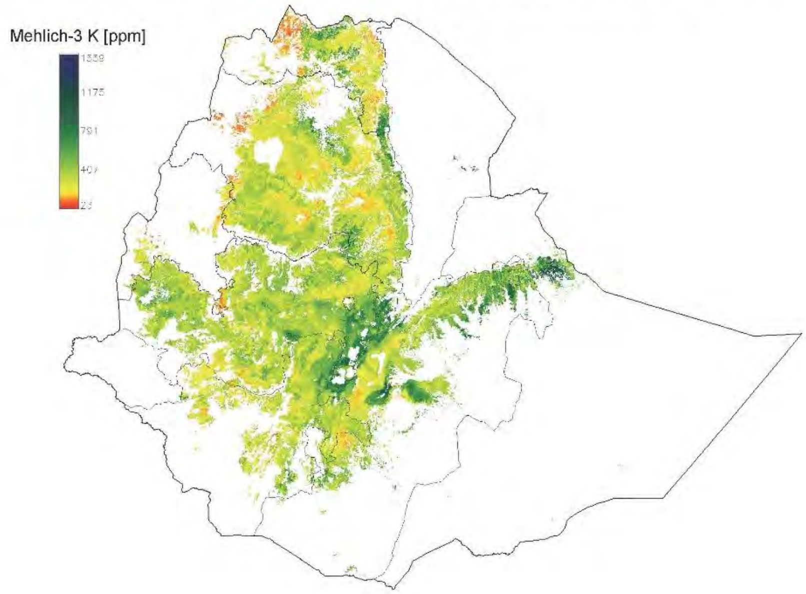
Prediction uncertainty, pH: 95% probable range
pH (upper - lower)


Predicted Nitrogen levels of cropland topsoils
N < 2000 ppm = ~15.6 Mha (49%) of cropland area
N [ppm]


Predicted Phosphorus levels of cropland topsoils
Mehlich-3 P < 30 ppm = ~31.7 Mha (99%) of cropland area
Mehlich-3 P [ppm]

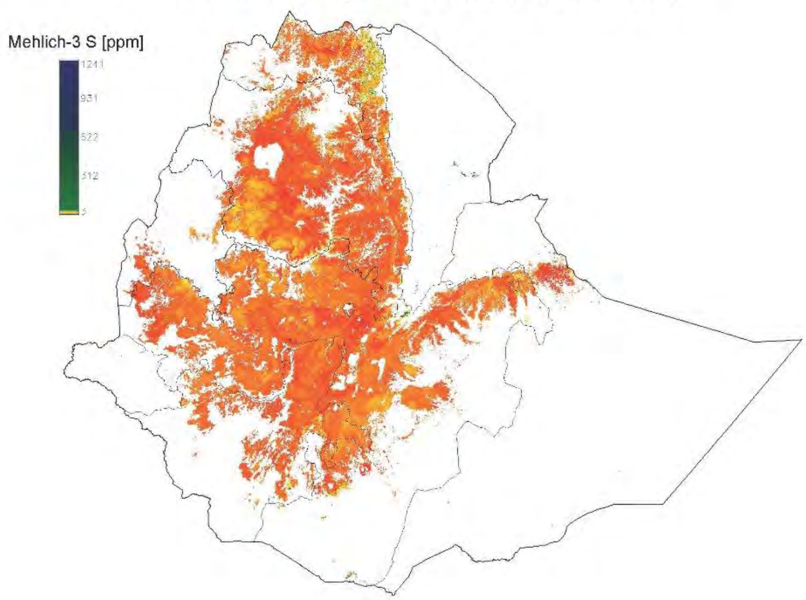
Predicted Potassium levels of cropland topsoils

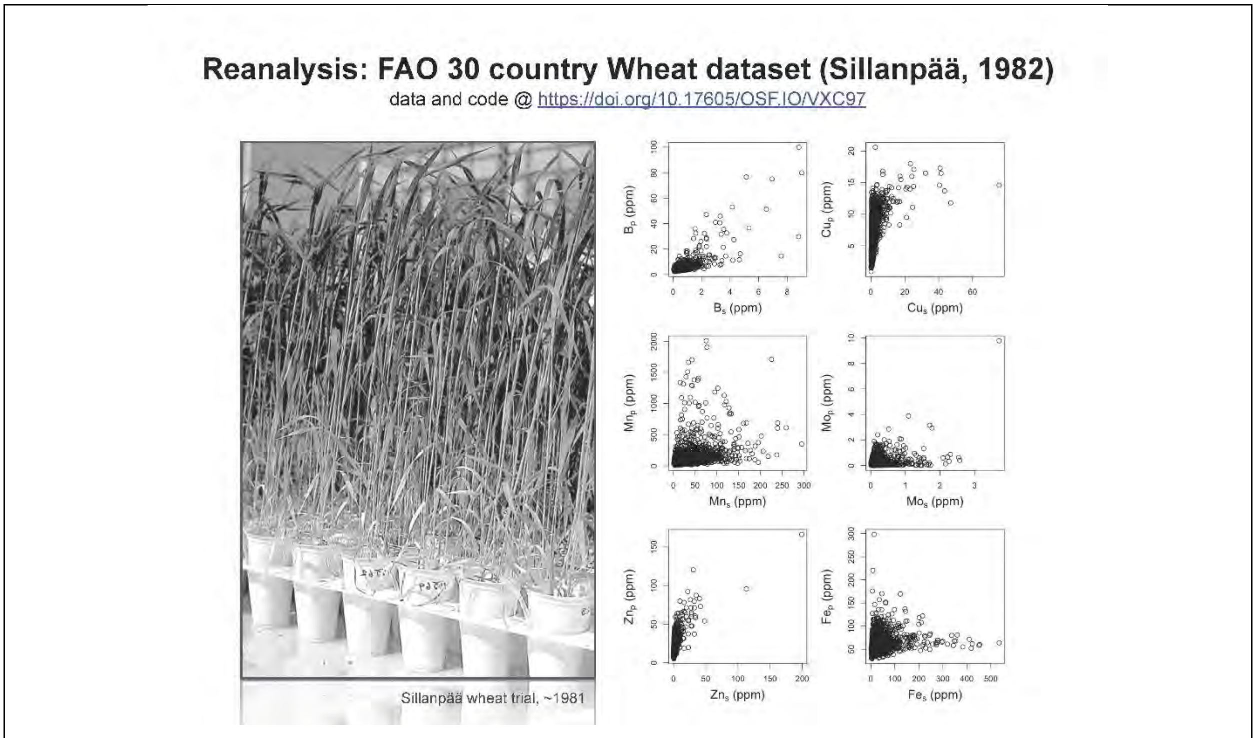
Mehlich-3 K < 190 ppm = ~3.1 Mha (~9.6%) of cropland area

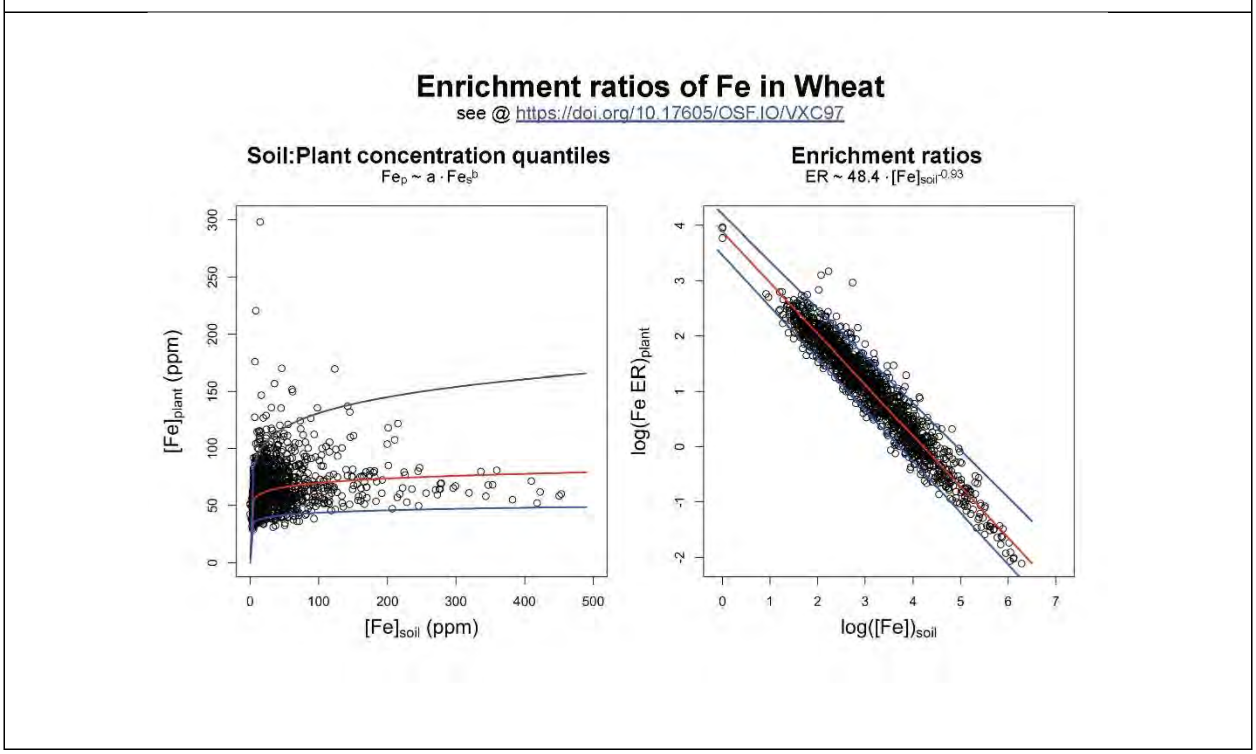


Predicted Sulfur levels of cropland topsoils

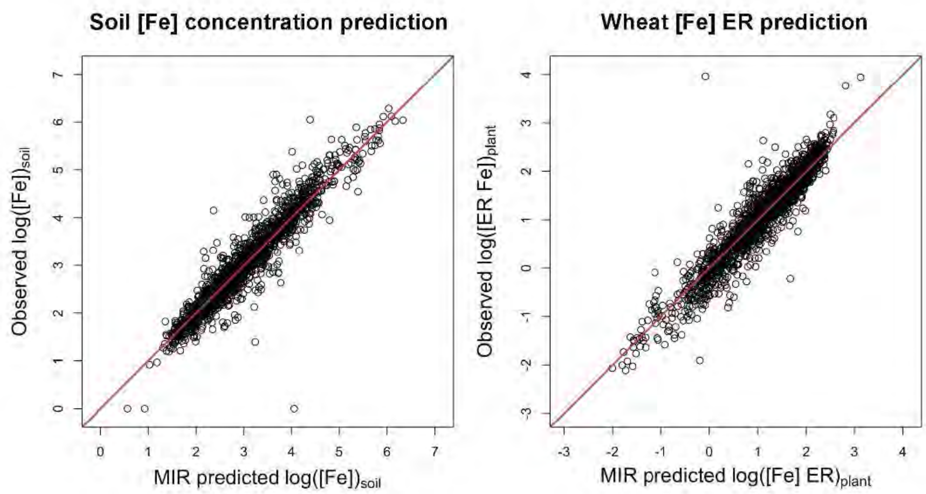
Mehlich-3 S < 20 ppm = ~31.6 Mha (99%) of cropland area




Reanalysis: FAO 30 country Wheat dataset (Sillanpää, 1982)
data and code @ https://doi.org/10.17605/OSF.IO/VXC97
Sillanpää wheat trial, ~1981
Bp (ppm)
Bs (ppm)
Cup (ppm)
Cus (ppm)
Mnp (ppm)
Mns (ppm)
Mop (ppm)
Mos (ppm)
Znp (ppm)
Zns (ppm)
Fep (ppm)
Fes (ppm)

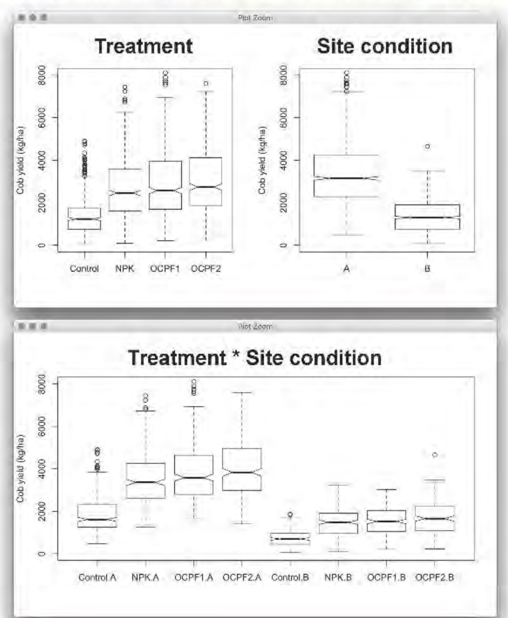

Enrichment ratios of Fe in Wheat
see @ https://doi.org/10.17605/OSF.IO/VXC97
Soil:Plant concentration quantiles
Fep ~ a · Fes^b
[Fe]plant (ppm)
[Fe]soil (ppm)
Enrichment ratios
ER ~ 48.4 · [Fe]soil^-0.93
log(Fe ER)plant
log([Fe])soil

Topsoil Fe and enrichment ratios in wheat predicted from soil MIR data



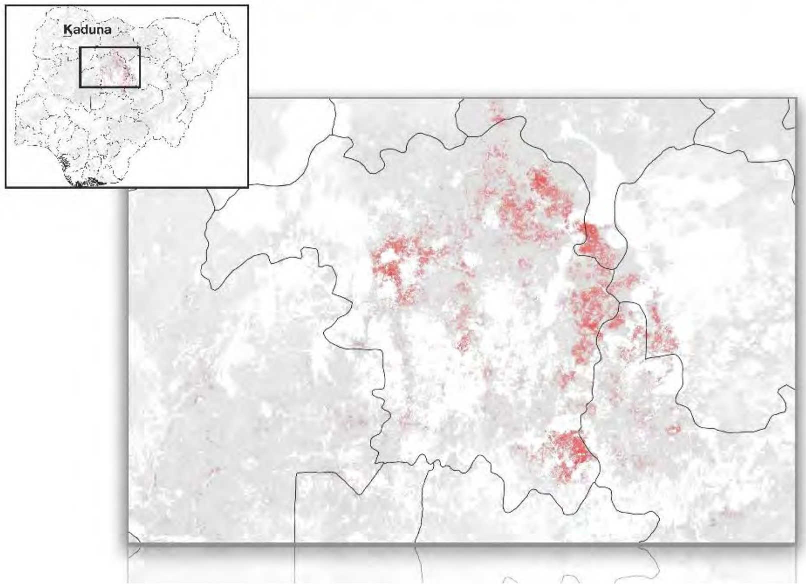
Special constraints: e.g. central Nigeria site indices:

https://github.com/mgwalsh/TRM/blob/master/OCP_trial_data.R

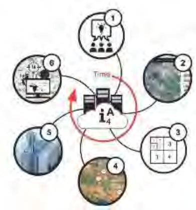


Predicted site condition indices of croplands in central Nigeria

data & code @ https://doi.org/10.17605/OSF.IO/2K5PJ



Main soil / plant / livestock monitoring challenges?



Logistics:

- Difficulty of implementing “representative” sampling designs at scale … There are a lot of places in Africa that we simply can’t get into
- Difficulties in accessing private or remote land areas in otherwise accessible countries
- Field expenses limiting the use of large sample / experiment numbers.

Cumulative measurement errors due to variations in:

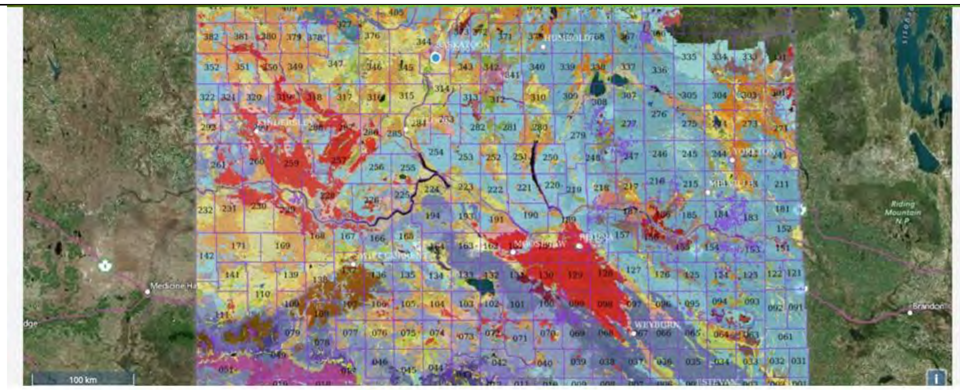
- Geo-referencing for repeat sampling & sampling depth inaccuracies
- Methods for measuring soil bulk density (or cumulative soil mass)
- Electricity for soil crushing and sieving, sub-sampling
- Variation among reference laboratories in analytical methods and quality control
- Consistency of measurement methods over several years to several decades.

AfSIS

Thank you for your attention and please feel free to contact me @:

markusgwalsh@gmail.com

Saskatchewan Soil Information System (SKSIS) – Angela Bedard-Haughn



UNIVERSITY OF SASKATCHEWAN

Predictive Soil Mapping – the SK experience *(and the role of SKSIS)*

Angela Bedard-Haughn, PhD, PAg

Soil Science, College of Agriculture and Bioresources, University of Saskatchewan

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

Strengths, challenges, opportunities

- Predictive soil mapping has tremendous potential for enhancing the quality of soil information everywhere
- We are trying to identify methods that allow us to work with the *strengths* of our legacy data, the *challenges* of our large spaces/coarse datasets and the *opportunities* for soil maps that enhance land management
- Objective for today: Share some examples of our experiences that are relevant to Alberta's PSM goals.

University of Saskatchewan
College of Agriculture and Bioresources

STRENGTH: Legacy Survey Data

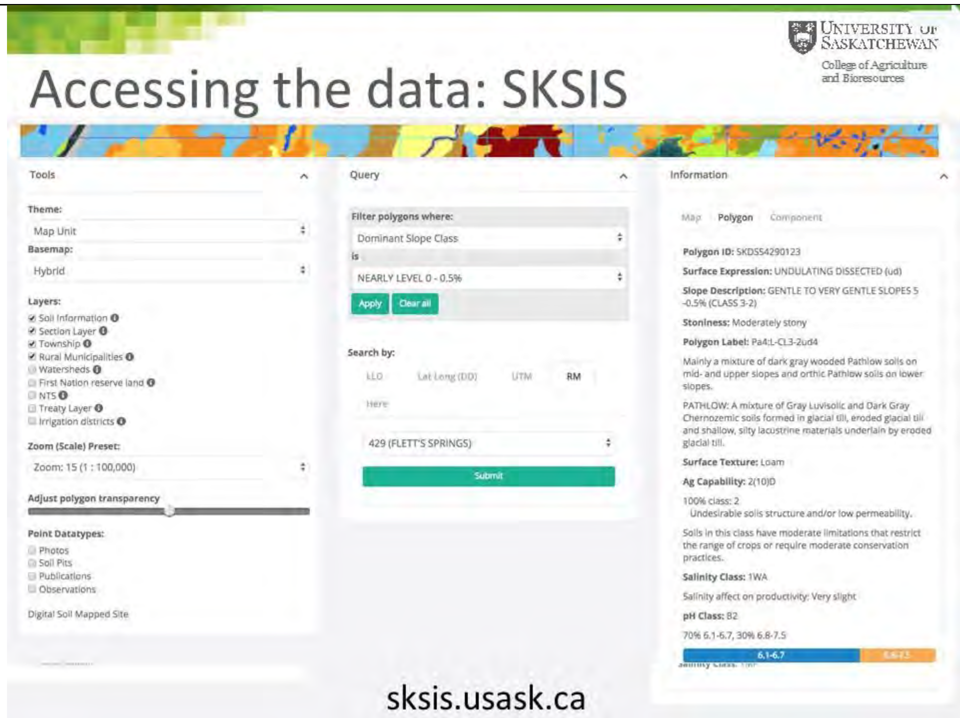
- What?
  - Map soil and vegetation zones
  - The original **big soils data**
- Why?
  - Determine agriculture and forest capability, irrigation potential, agri-environmental rating, GHG
- How?
  - Canadian system started in 1945
  - Glory years: 1950-95

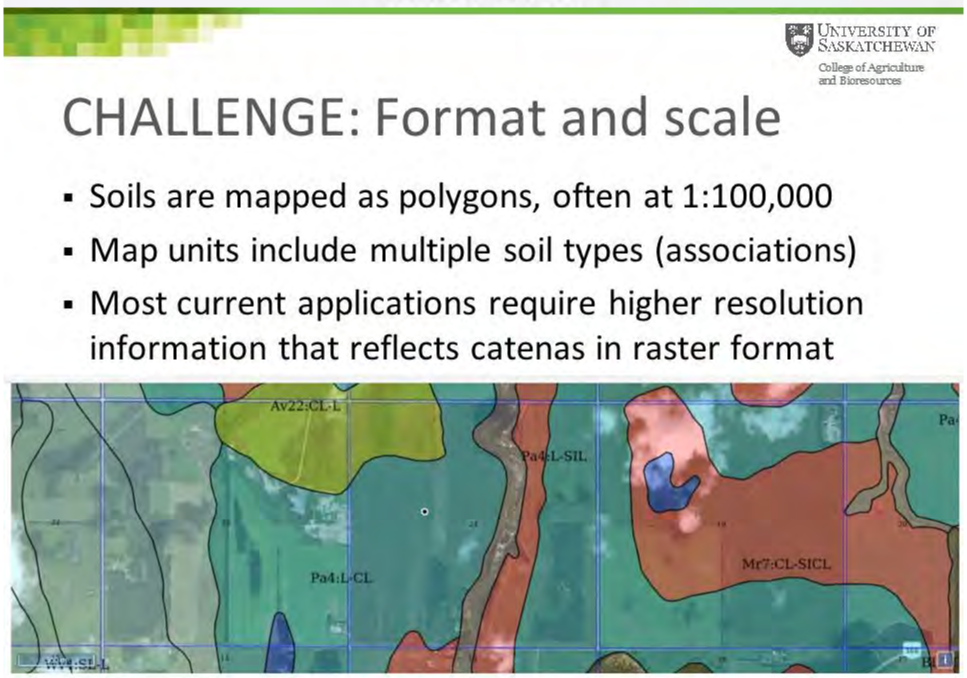
University of Saskatchewan
College of Agriculture and Bioresources

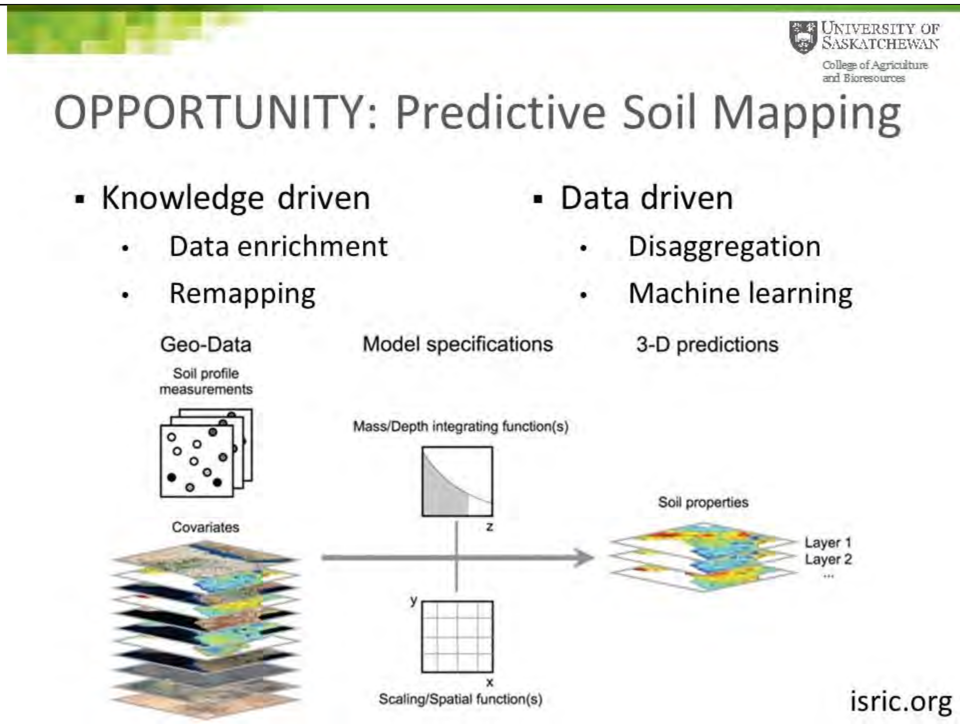
STRENGTH: The Data Legacy

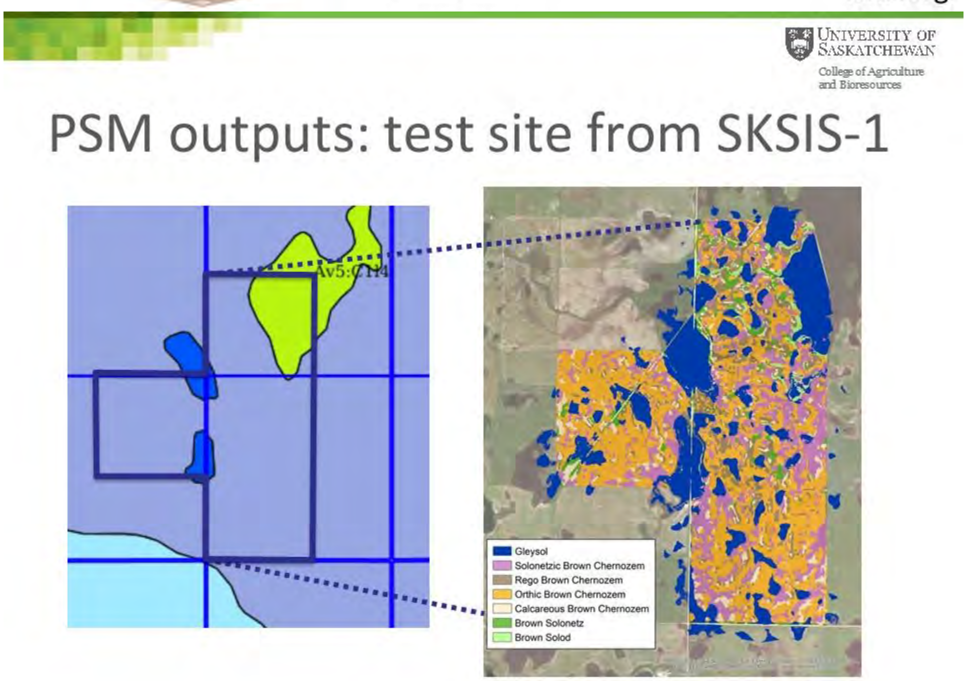
PDFs: sis.agr.gc.ca/cansis/publications/surveys/index.html

GIS: sis.agr.gc.ca/cansis/nsdb/index.html


University of Saskatchewan
College of Agriculture and Bioresources
Accessing the data: SKSIS
Tools
Theme:
Map Unit
Basemap:
Hybrid
Layers:
Soil Information
Section Layer
Township
Rural Municipalities
Watersheds
First Nation reserve land
NTS
Treaty Layer
Irrigation districts
Zoom (Scale) Preset:
Zoom: 15 (1 : 100,000)
Adjust polygon transparency
Point Datatypes:
Photos
Soil Pits
Publications
Observations
Digital Soil Mapped Site
Query
Filter polygons where:
Dominant Slope Class
is
NEARLY LEVEL 0 - 0.5%
Apply
Clear all
Search by:
LLD
Lat Long (DD)
UTM
RM
429 (FLETT'S SPRINGS)
Submit
Information
Map
Polygon
Component
Polygon ID: SKDS54290123
Surface Expression: UNDULATING DISSECTED (ud)
Slope Description: GENTLE TO VERY GENTLE SLOPES 5 -0.5% (CLASS 3-2)
Stoniness: Moderately stony
Polygon Label: Pa4:L-CL3-2ud4
Mainly a mixture of dark gray wooded Pathlow soils on mid- and upper slopes and orthic Pathlow soils on lower slopes.
PATHLOW: A mixture of Gray Luvisolic and Dark Gray Chernozemic soils formed in glacial till, eroded glacial till and shallow, silty lacustrine materials underlain by eroded glacial till.
Surface Texture: Loam
Ag Capability: 2(10)D
100% class: 2
Undesirable soils structure and/or low permeability.
Soils in this class have moderate limitations that restrict the range of crops or require moderate conservation practices.
Salinity Class: 1WA
Salinity affect on productivity: Very slight
pH Class: B2
70% 6.1-6.7, 30% 6.8-7.5
6.1-6.7
sksis.usask.ca


University of Saskatchewan
College of Agriculture and Bioresources
CHALLENGE: Format and scale
▪ Soils are mapped as polygons, often at 1:100,000
▪ Map units include multiple soil types (associations)
▪ Most current applications require higher resolution information that reflects catenas in raster format
Av22:CL-L
Pa4:L-SIL
Pa4:L-CL
Mr7:CL-SICL


University of Saskatchewan
College of Agriculture and Bioresources
OPPORTUNITY: Predictive Soil Mapping
▪ Knowledge driven
• Data enrichment
• Remapping
▪ Data driven
• Disaggregation
• Machine learning
Geo-Data
Soil profile measurements
Covariates
Model specifications
Mass/Depth integrating function(s)
z
y
x
Scaling/Spatial function(s)
3-D predictions
Soil properties
Layer 1
Layer 2
...
isric.org


University of Saskatchewan
College of Agriculture and Bioresources
PSM outputs: test site from SKSIS-1
Gleysol
Solonetzic Brown Chernozem
Rego Brown Chernozem
Orthic Brown Chernozem
Calcareous Brown Chernozem
Brown Solonetz
Brown Solod

University of Saskatchewan
College of Agriculture and Bioresources

Key considerations in PSM

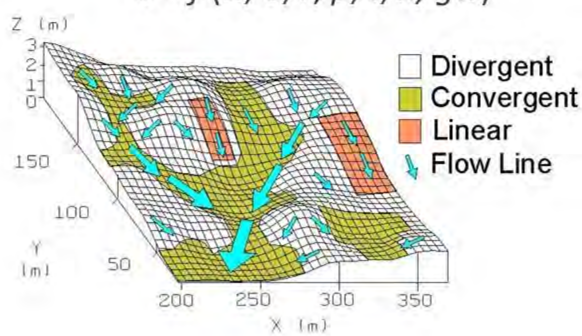
- Quality and quantity of input data (aka: can't make something from nothing)
  - Decent DEM and covariate data?
  - Sufficient training point data?
  - Evaluation dataset available?
- More data in makes better info out: quantify uncertainty
- Outputs of applications can be added to database: errors can become magnified without adequate QA/QC

University of Saskatchewan
College of Agriculture and Bioresources

STRENGTH: Conceptual model

- Building on the legacy of our soil surveyors, we know the key soil-forming factors (Jenny & Ellis) and how they act/interact across the landscape

$$S = f\,(cl, o, r, p, t, h, gw)$$



University of Saskatchewan
College of Agriculture and Bioresources

Joe Ellis' soil-forming factors

Factors which Determine Soil Type:

The factors that determine soil type (or group) are:

(1) The climate, or the temperature and moisture within the soil;

(2) The vegetation, which determines the type of organic matter added to the soil;

(3) The parent material, or the geological deposits which determine the minerals on which the soil is formed, and in turn affect the texture, the water retention capacity, and the mineral reserve;

(4) The position in which the soil is found in relationship to the topography;

(5) The presence or absence of ground water within the soil profile;

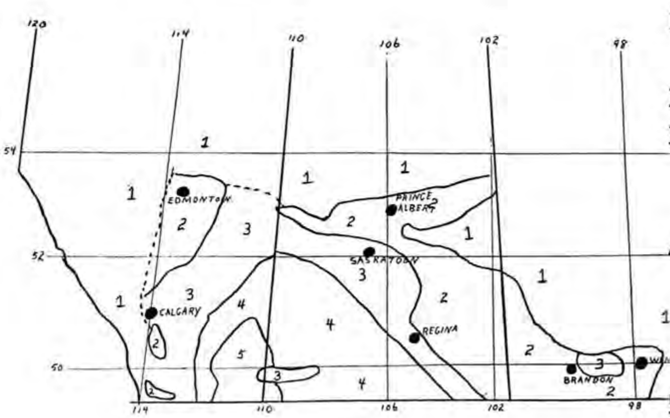
(6) The age or length of time the soil has been under the influence of its environment; and

(7) In the case of cultivated soils—the modifying effects of culture or the work of man.

Ellis 1938 Soils of Manitoba

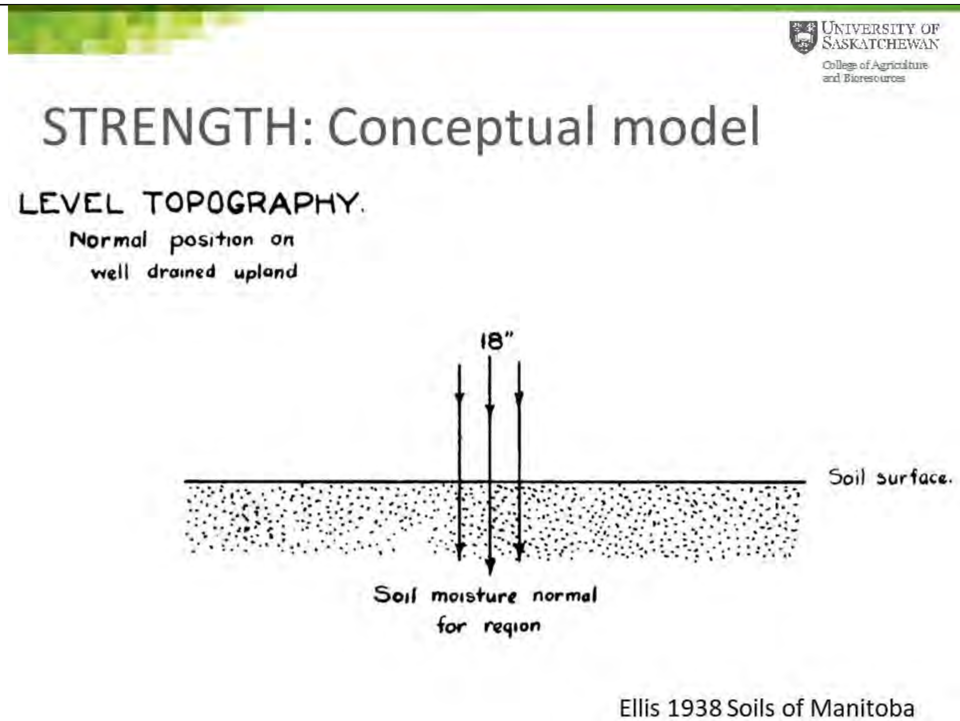
University of Saskatchewan
College of Agriculture and Bioresources

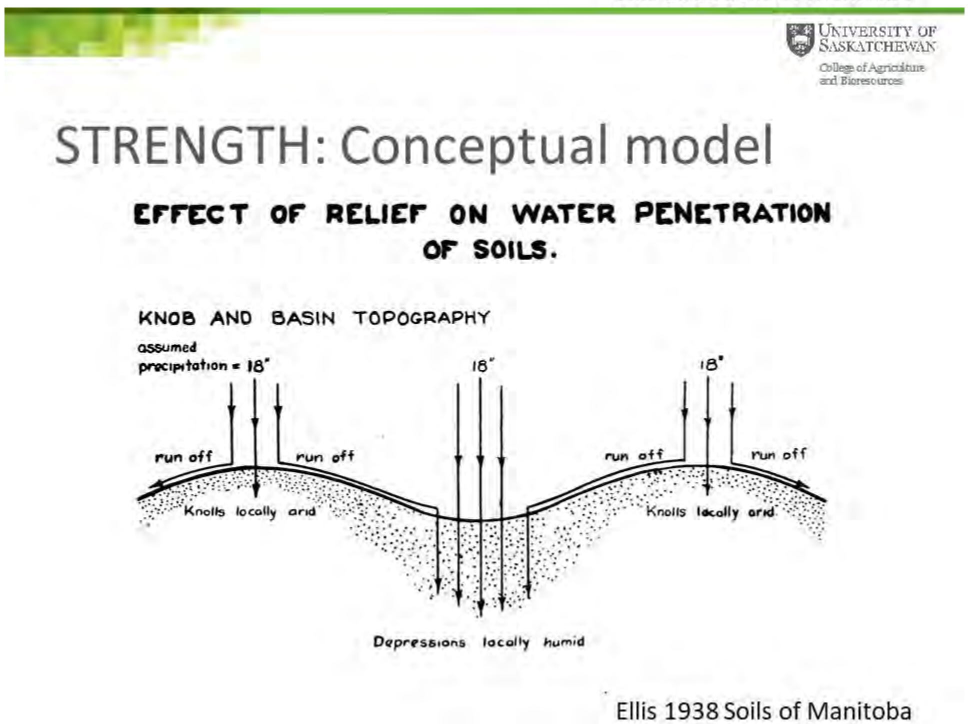
Soil and Vegetation Zones ca. 1928



1. Greyish timber soils
2. Black park lands
3. Dark brown or chestnut brown prairie-plains
4. Brown or chocolate plains (short and medium grasses)
5. Greyish brown, short grass plains

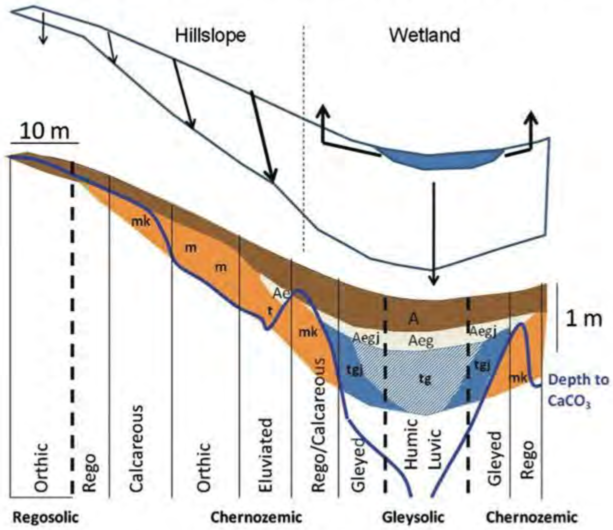
Joel, 1928


UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources
STRENGTH: Conceptual model
LEVEL TOPOGRAPHY.
Normal position on well drained upland
18"
Soil surface.
Soil moisture normal for region
Ellis 1938 Soils of Manitoba


UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources
STRENGTH: Conceptual model
EFFECT OF RELIEF ON WATER PENETRATION OF SOILS.
KNOB AND BASIN TOPOGRAPHY
assumed precipitation = 18"
18"
18"
run off
run off
run off
run off
Knolls locally arid
Knolls locally arid
Depressions locally humid
Ellis 1938 Soils of Manitoba

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

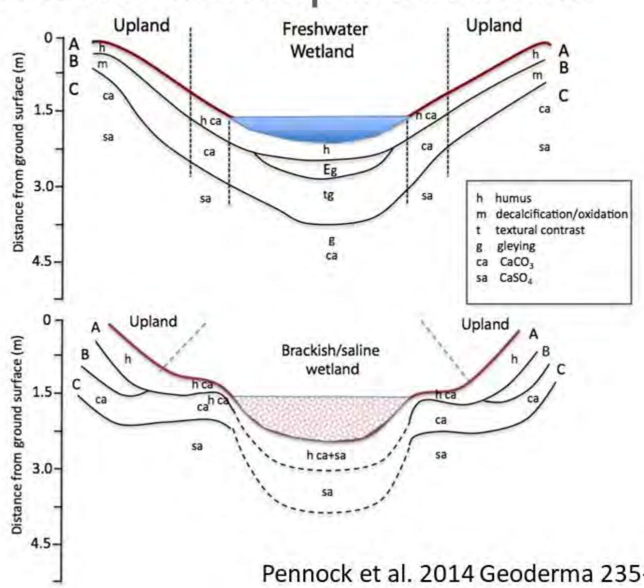
STRENGTH: Conceptual model



Pennock et al. 2011 CJSS 91: 719-747

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

STRENGTH: Conceptual model



Pennock et al. 2014 Geoderma 235-236: 199-211

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

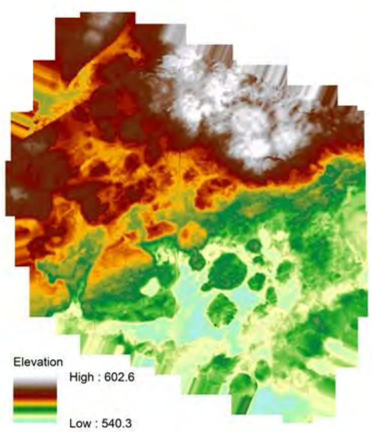
CHALLENGE: Conceptual model to quantitative model

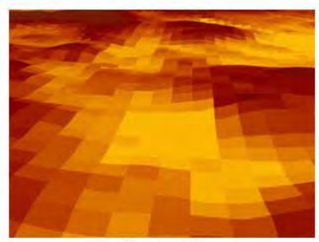
- It's one thing to understand the landscape, another to be able to quantitatively predict it
- Knowledge driven vs. data driven: the former makes intuitive sense, but can be limited in the complexity of "real landscapes"
    - Scott's folly…
- To make the link, need great covariates, which require high quality digital surface models/DEMs, input layers

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

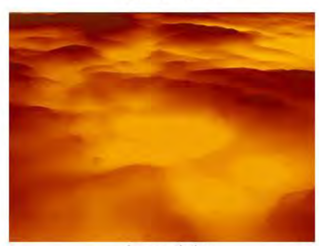
Digital Elevation Model (DEM)

- Grid-based landscape model
- Every cell has an elevation value





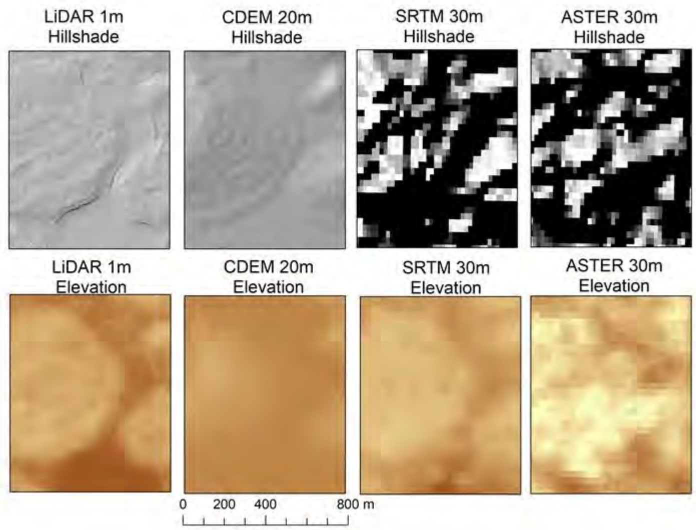
30 m resolution



1 m resolution

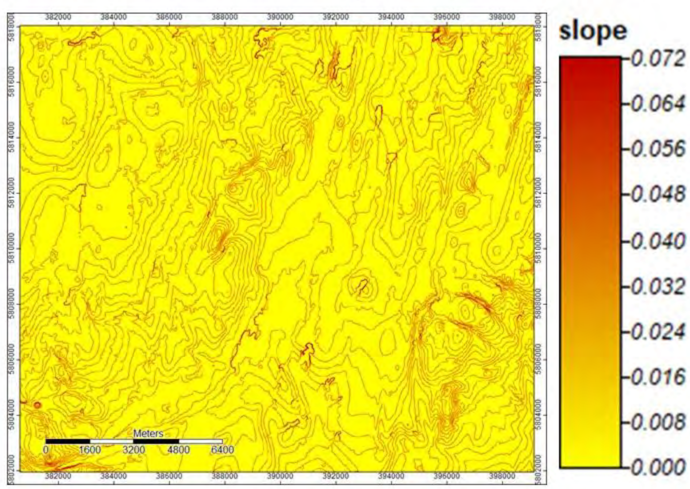
University of Saskatchewan
College of Agriculture and Bioresources

Know the limits of your dataset



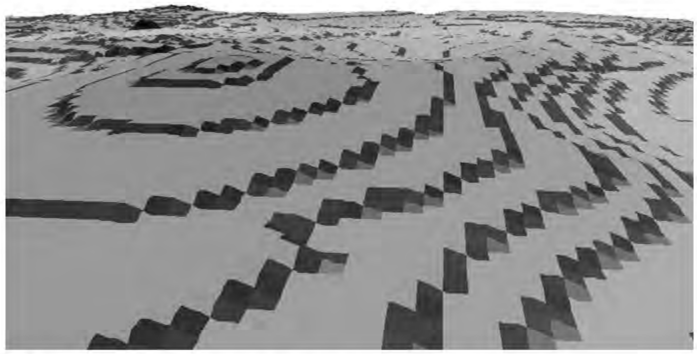
University of Saskatchewan
College of Agriculture and Bioresources

Know the limits of your dataset



UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

Wetness Index – interpret this…



FYI – this is not a terraced field...

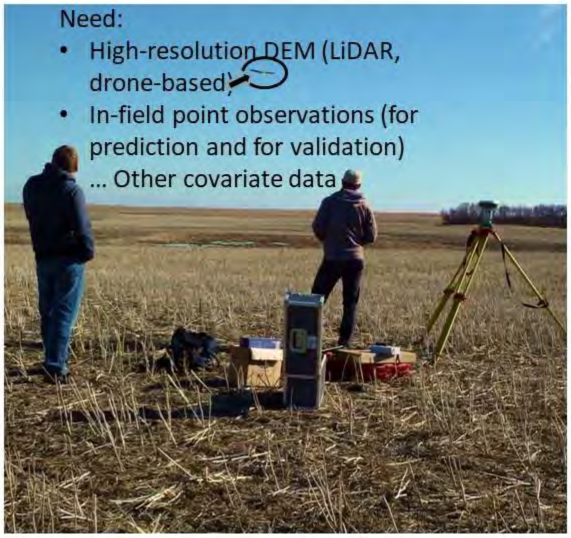
UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

OPPORTUNITY: We have the technology!

- Geographic information systems (commercial & open)
- Digital elevation models (DEM): LiDAR, UAV-derived
- Remote sensing
- Environmental datasets: vegetation, climate, hydrology, geology, etc.
- **More data** never been easier to collect!

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

OPPORTUNITY: DIY DEM



UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

STRENGTH: Computing power

- The mapping we can now do on an average laptop dwarfs the capabilities of our legacy surveyors
- Once we have a quality DEM, we can use that information, in combination with other information layers (remotely-sensed and otherwise) to develop covariates or predictor variables
- These predictor variables are combined with point data to “train” the computer about what points go with what variables: machine learning at its essence

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

CHALLENGE 1: Sensible predictor variables

- Just because something *can* be included doesn't automatically mean it *should*
- How many covariates/predictor variables are really necessary?
    - Can they be used to interpret the output?
    - How do we incorporate that conceptual model into the quantitative analyses?

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

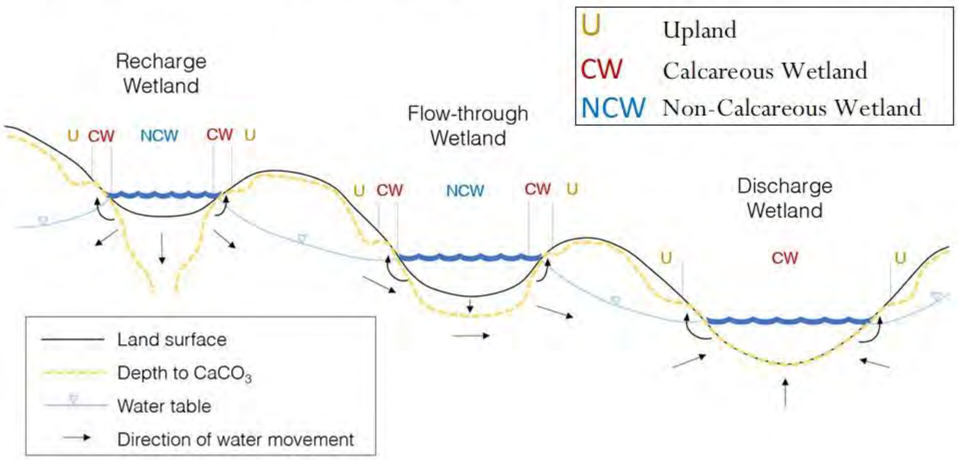
Example: PSM in the PPR (Kiss thesis)

- Available LiDAR-derived digital elevation models (DEMs)
- Characteristics representative of the PPR:
    - Glacial till parent material
    - Undulating to hummocky topography
- Span 3 major soil zones

InnoTech ALBERTA

University of Saskatchewan
College of Agriculture and Bioresources

Example: PSM in the PPR (Kiss thesis)

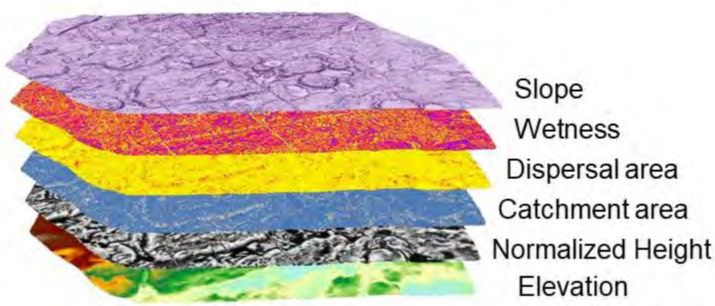


Jeremy Kiss MSc thesis, Fall 2018

University of Saskatchewan
College of Agriculture and Bioresources

Example: PSM in the PPR (Kiss thesis)

- 32 predictor variables
- All derived from the DEM
- Measures of slope, aspect, curvature, wetness index, etc.



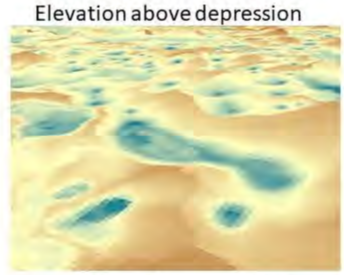
Jeremy Kiss MSc thesis, Fall 2018

OPPORTUNITY 1: Knowledge-Based Predictor Variables

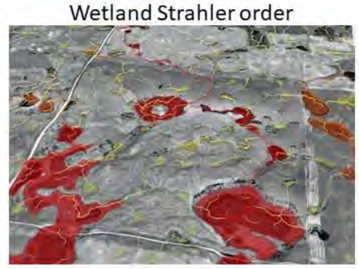
UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

- Several predictor variables were developed/selected to reflect the morphology of the PPR:
  - Elevation above depression bottom
  - Elevation relative to depression spillover elevation
  - Wetland Strahler order
  - Depression catchment area
  - Depression depth

Elevation above depression

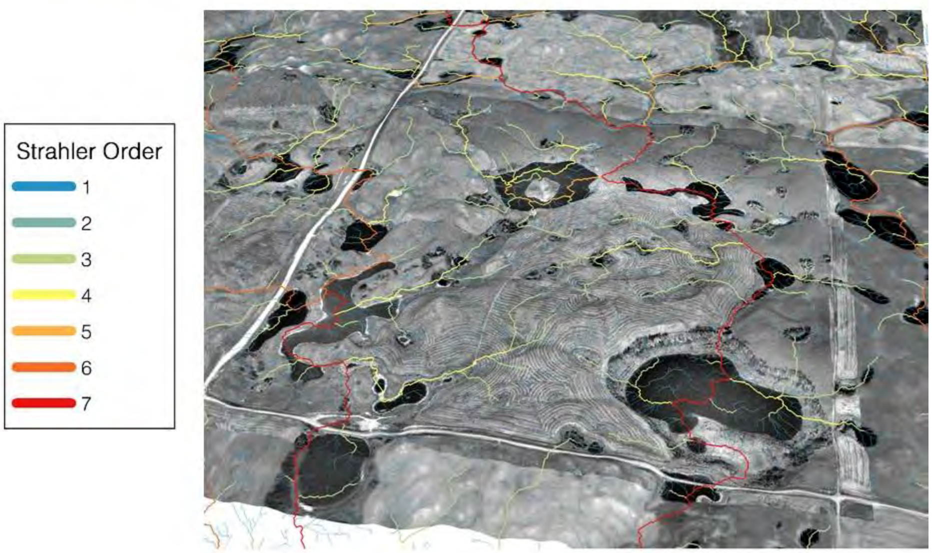


Wetland Strahler order

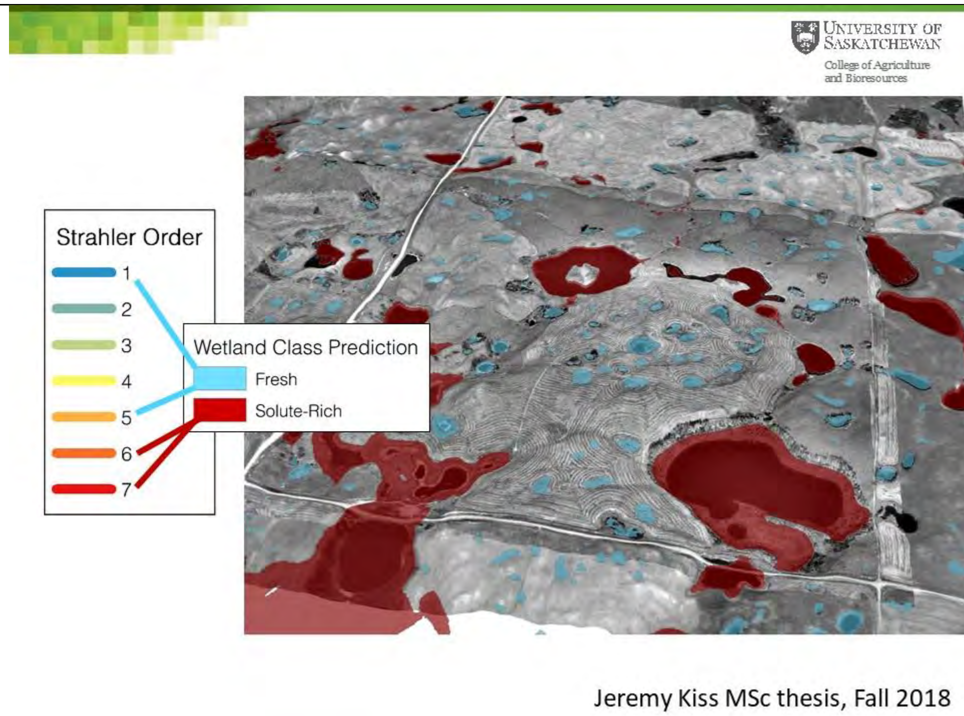


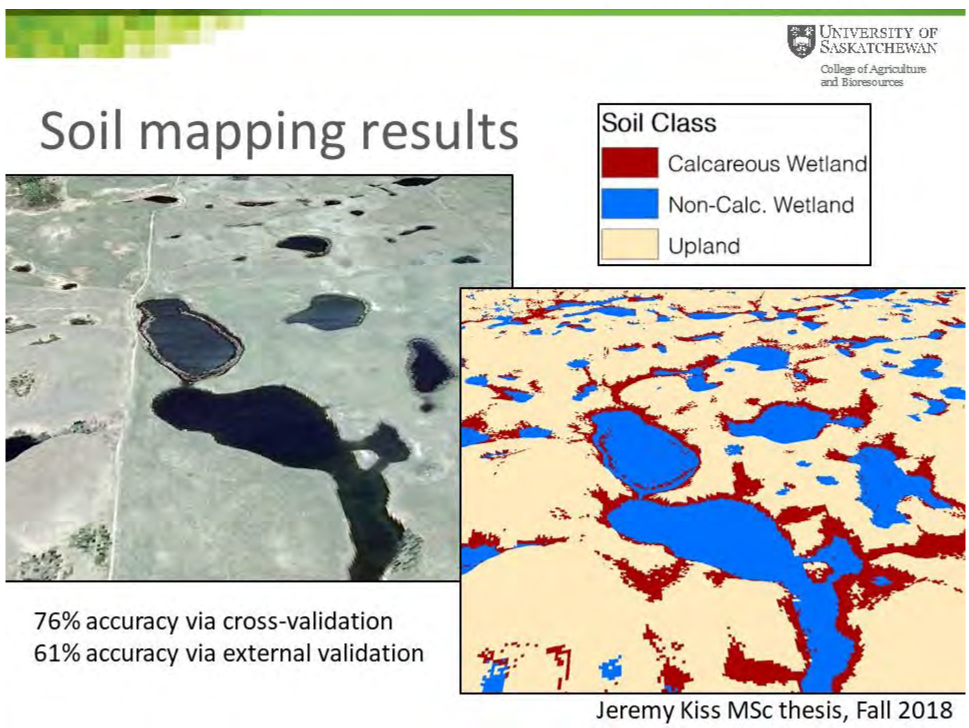
Jeremy Kiss MSc thesis, Fall 2018

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources



Jeremy Kiss MSc thesis, Fall 2018


University of Saskatchewan
College of Agriculture and Bioresources
Strahler Order
1
2
3
4
5
6
7
Wetland Class Prediction
Fresh
Solute-Rich
Jeremy Kiss MSc thesis, Fall 2018

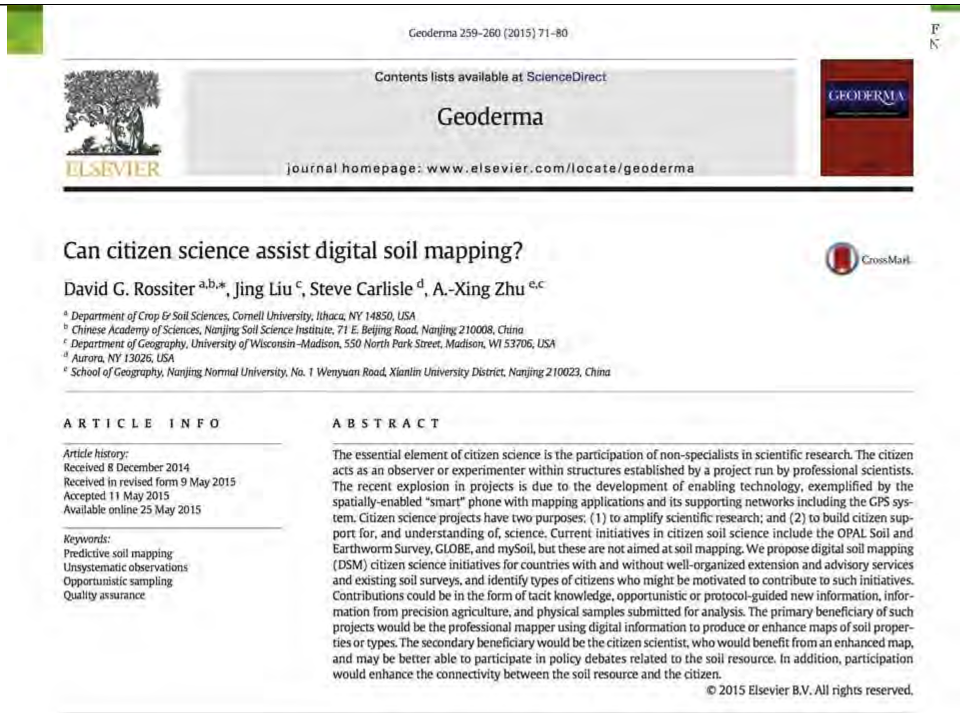

University of Saskatchewan
College of Agriculture and Bioresources
Soil mapping results
Soil Class
Calcareous Wetland
Non-Calc. Wetland
Upland
76% accuracy via cross-validation
61% accuracy via external validation
Jeremy Kiss MSc thesis, Fall 2018

CHALLENGE 2: Lack of point data

- Despite there being more soil information being collected and precisely georeferenced via GPS than at any other point in history, predictive mappers struggle to have enough point data to carry out large-scale mapping efforts
  - AAFC legacy data in some areas (i.e., SK, but others, too) only referenced to the quarter-section
  - "Pure" polygons not a solution for the Prairies
  - Still need to do a fair bit of field work to train the computer – critics ask what the PSM advantage actually is… (there's good answers, but need to be ready for the question)

OPPORTUNITY 2: Data sharing

- With the decentralization of soils enthusiasts, more opportunity than ever to expand our soil database through data sharing: "crowdsourcing"
- Also have the opportunity to broaden the type(s) of data/information
- Many hands (smartphones?) make light work!


Geoderma 259-260 (2015) 71-80

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

ELSEVIER

Can citizen science assist digital soil mapping?

David G. Rossiter [a,b,*], Jing Liu [c], Steve Carlisle [d], A.-Xing Zhu [e,c]

[a] Department of Crop & Soil Sciences, Cornell University, Ithaca, NY 14850, USA
[b] Chinese Academy of Sciences, Nanjing Soil Science Institute, 71 E. Beijing Road, Nanjing 210008, China
[c] Department of Geography, University of Wisconsin-Madison, 550 North Park Street, Madison, WI 53706, USA
[d] Aurora, NY 13026, USA
[e] School of Geography, Nanjing Normal University, No. 1 Wenyuan Road, Xianlin University District, Nanjing 210023, China

ARTICLE INFO

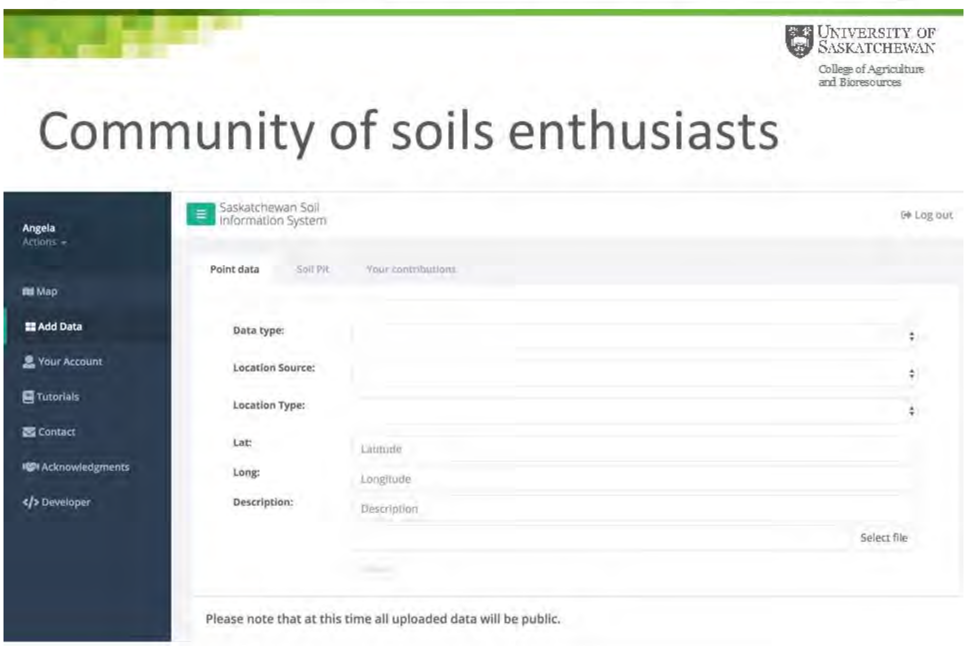
Article history:
Received 8 December 2014
Received in revised form 9 May 2015
Accepted 11 May 2015
Available online 25 May 2015

Keywords:
Predictive soil mapping
Unsystematic observations
Opportunistic sampling
Quality assurance

ABSTRACT

The essential element of citizen science is the participation of non-specialists in scientific research. The citizen acts as an observer or experimenter within structures established by a project run by professional scientists. The recent explosion in projects is due to the development of enabling technology, exemplified by the spatially-enabled "smart" phone with mapping applications and its supporting networks including the GPS system. Citizen science projects have two purposes: (1) to amplify scientific research; and (2) to build citizen support for, and understanding of, science. Current initiatives in citizen soil science include the OPAL Soil and Earthworm Survey, GLOBE, and mySoil, but these are not aimed at soil mapping. We propose digital soil mapping (DSM) citizen science initiatives for countries with and without well-organized extension and advisory services and existing soil surveys, and identify types of citizens who might be motivated to contribute to such initiatives. Contributions could be in the form of tacit knowledge, opportunistic or protocol-guided new information, information from precision agriculture, and physical samples submitted for analysis. The primary beneficiary of such projects would be the professional mapper using digital information to produce or enhance maps of soil properties or types. The secondary beneficiary would be the citizen scientist, who would benefit from an enhanced map, and may be better able to participate in policy debates related to the soil resource. In addition, participation would enhance the connectivity between the soil resource and the citizen.

© 2015 Elsevier B.V. All rights reserved.



UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

OPPORTUNITY: Act local

- What can each of us do right now?
  - Government: enhance access to existing data
  - Private: think about what data can be shared to enhance quality of existing data
  - Students: make sure you understand the whole data life cycle, from the use of legacy survey data to the collection, integration, modeling, and *sharing* of new data!
- We have the legacy of the soil surveyors who came before us – we need to continue the tradition for those who come after…

UNIVERSITY OF SASKATCHEWAN
College of Agriculture and Bioresources

Build on the strengths, overcome the challenges, take advantage of opportunities

- STRENGTHS: Legacy data, conceptual model, computing power and creativity
- CHALLENGES: Spatial data resolution, obstacles to data sharing
- OPPORTUNITIES: Use emerging technologies and data sharing to enhance legacy data

With thanks to…

CORE TEAM

- J. Kiss
- B. Heung
- M. Bentham
- P. Krug
- K. Walters
- T. Jamsrandorj
- R. Deters
- D. Cerkowniak

COLLABORATORS

- Don Campbell
  Western Sales
- J. Schoenau
- T. Nerbas
- H. de Gooijer
- D. Anderson
- B. McConkey
- L. Henry

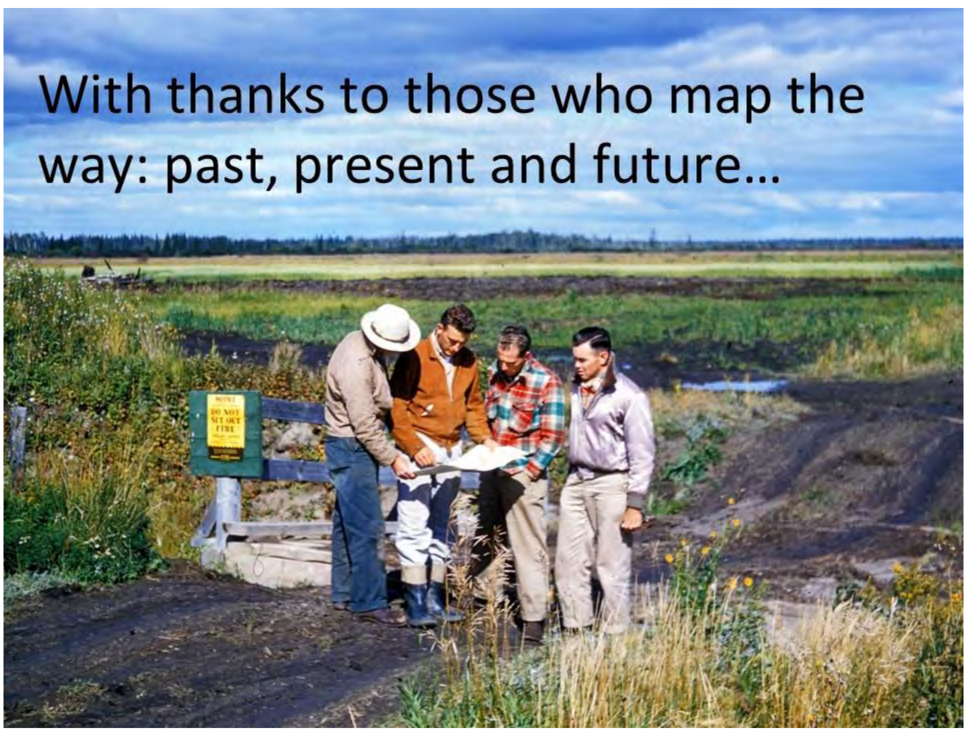
University of Saskatchewan
College of Agriculture and Bioresources

Government of Saskatchewan
Growing Forward 2
A federal-provincial-territorial initiative

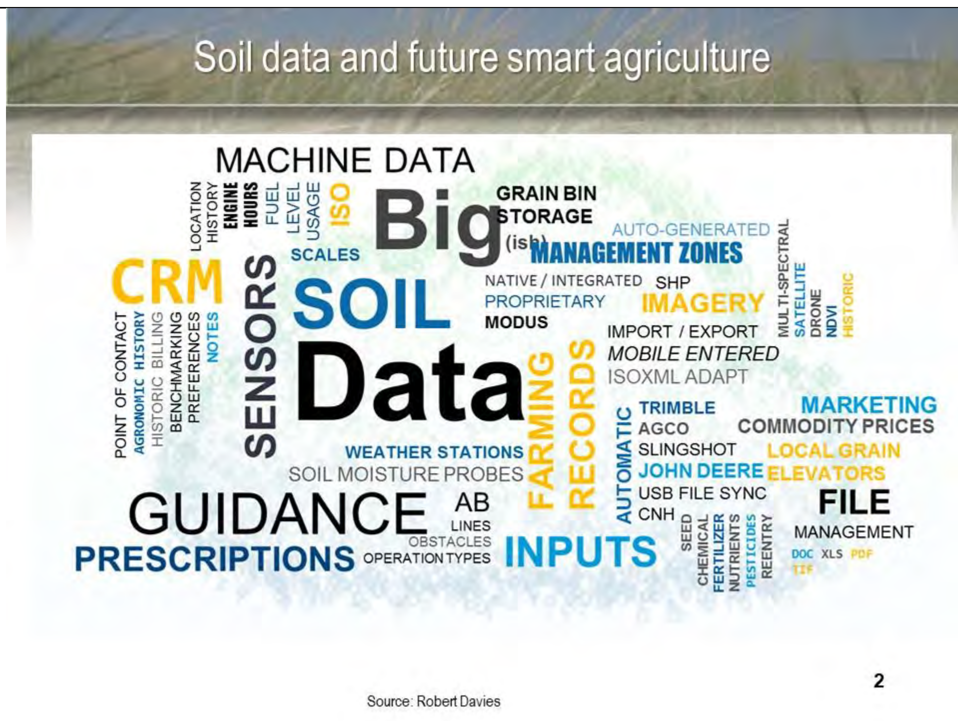
SaskCanola
Saskatchewan Canola Development Commission

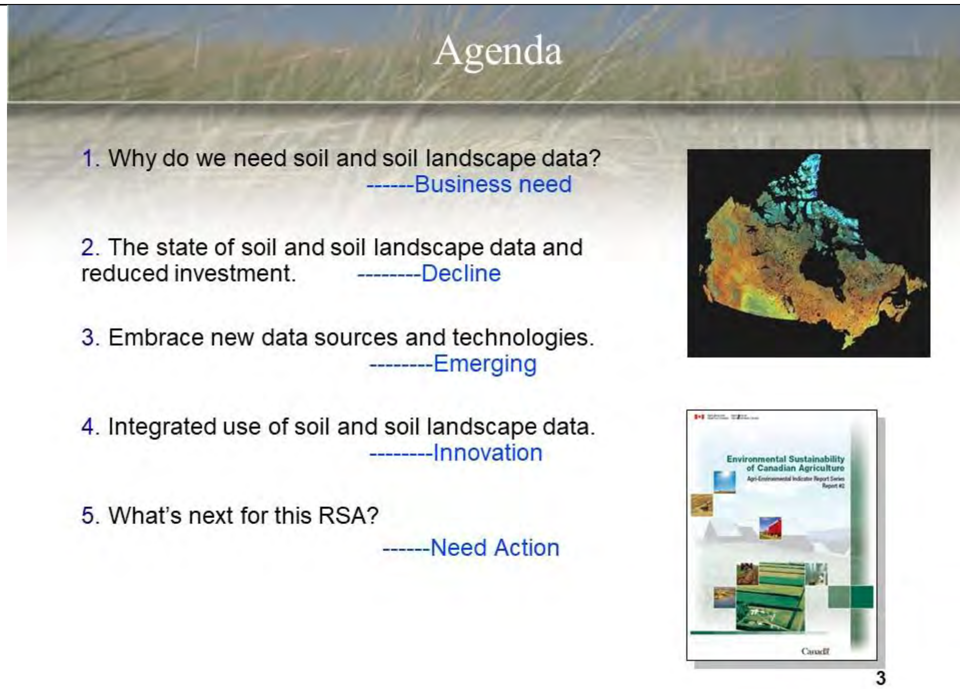
Saskatchewan Pulse Growers

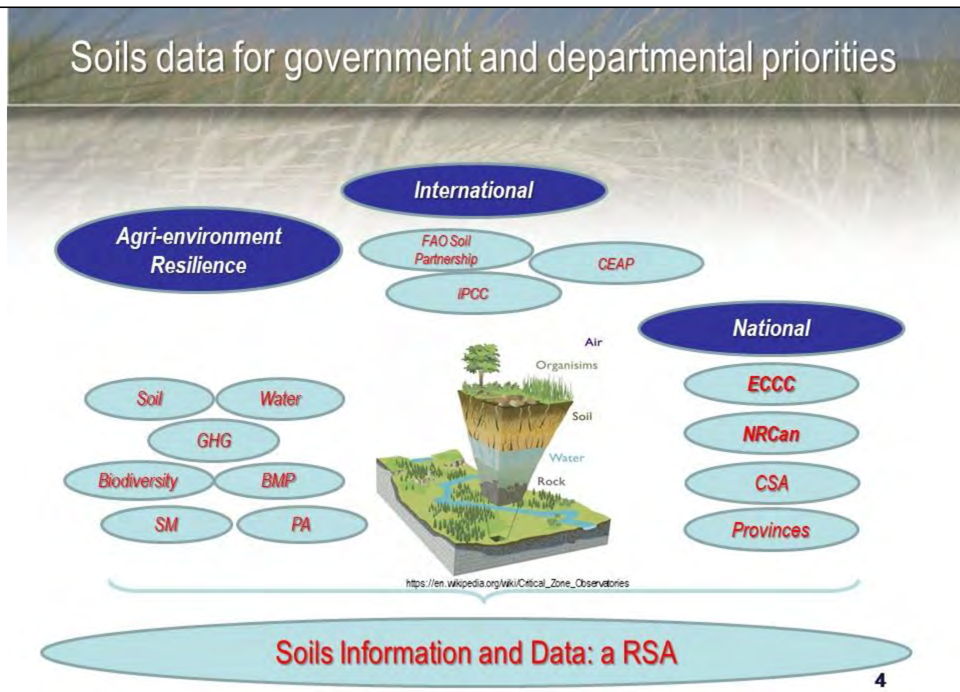
Western Grains Research Foundation



Predictive Soil Mapping—National Perspective – Xiaoyuan Geng

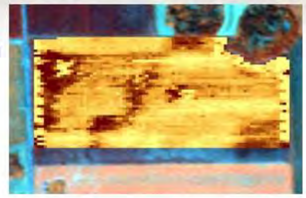


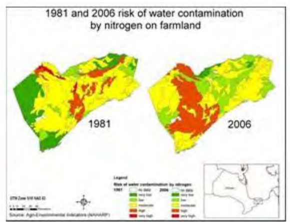

Agenda
1. Why do we need soil and soil landscape data?
------Business need
2. The state of soil and soil landscape data and reduced investment. --------Decline
3. Embrace new data sources and technologies.
--------Emerging
4. Integrated use of soil and soil landscape data.
--------Innovation
5. What's next for this RSA?
------Need Action
Environmental Sustainability of Canadian Agriculture
3

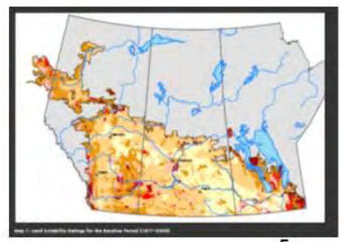

Soils data for government and departmental priorities
International
FAO Soil Partnership
CEAP
IPCC
Agri-environment Resilience
Soil
Water
GHG
Biodiversity
BMP
SM
PA
Air
Organisims
Soil
Water
Rock
National
ECCC
NRCan
CSA
Provinces
https://en.wikipedia.org/wiki/Critical_Zone_Observatories
Soils Information and Data: a RSA
4

The needs for soil and soil landscape data

- Infrastructure, environment and watershed management, SM, risk management
- BMPs research, development and evaluation
- Precision agriculture, biodiversity etc.
- EGS measurement and assessment



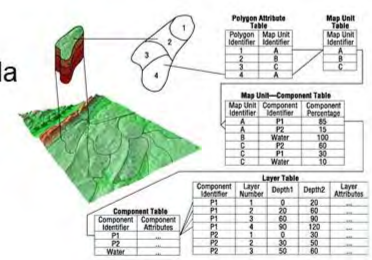




5

State of national soil and soil landscape data

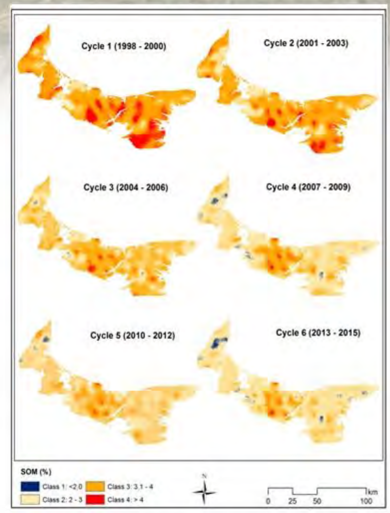
- National Ecological Framework (ECO)
- Soil Landscapes of Canada (SLC)
- Canada Land Inventory (CLI)
- Detailed Soil Surveys (DSS)
- Site (pedon) data
- Soil Classification System for Canada
- National soil carbon database
- http://sis.agr.gc.ca/cansis



6

Soil changes and information renewal is critical

These pillars at the eastbound rest stop on I-80 near Adair in western Iowa are designed to look like giant soil cores showing soil lost over time in Iowa. This display illustrates a 14-inch topsoil depth in 1850, compared to 5.5 inches in 2000. [Photo by Christopher Gannon]

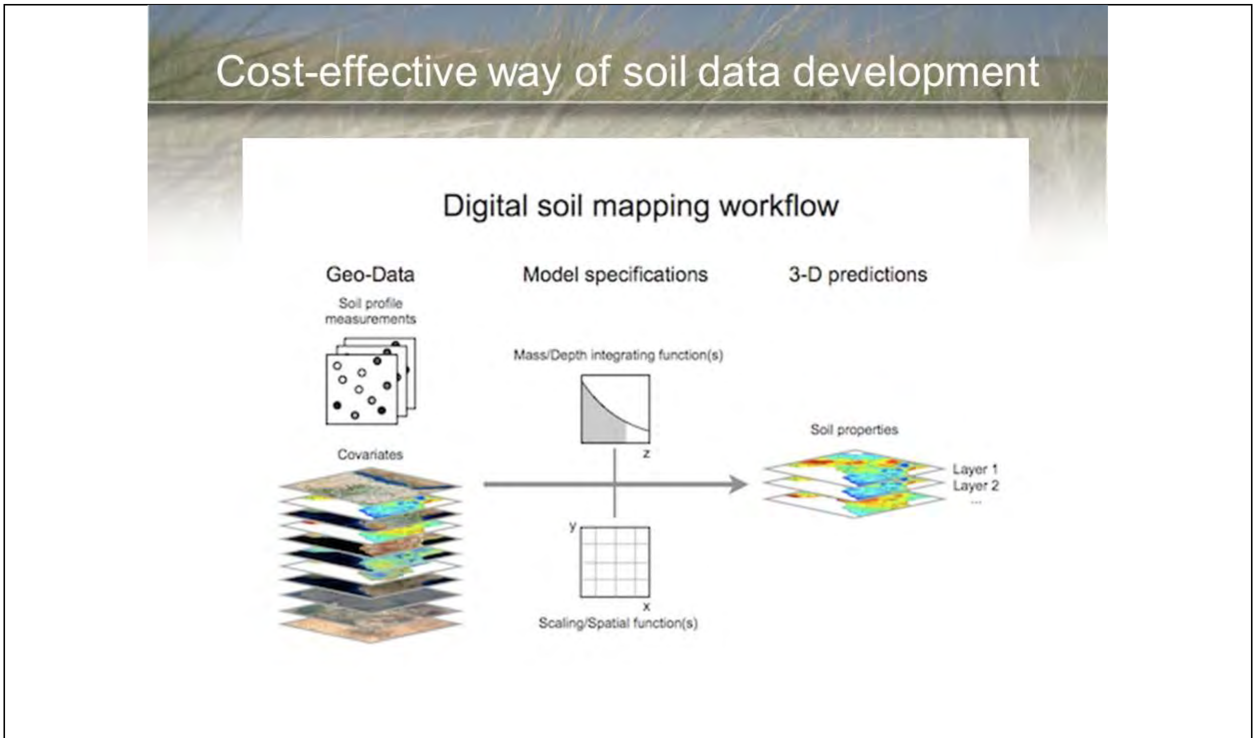


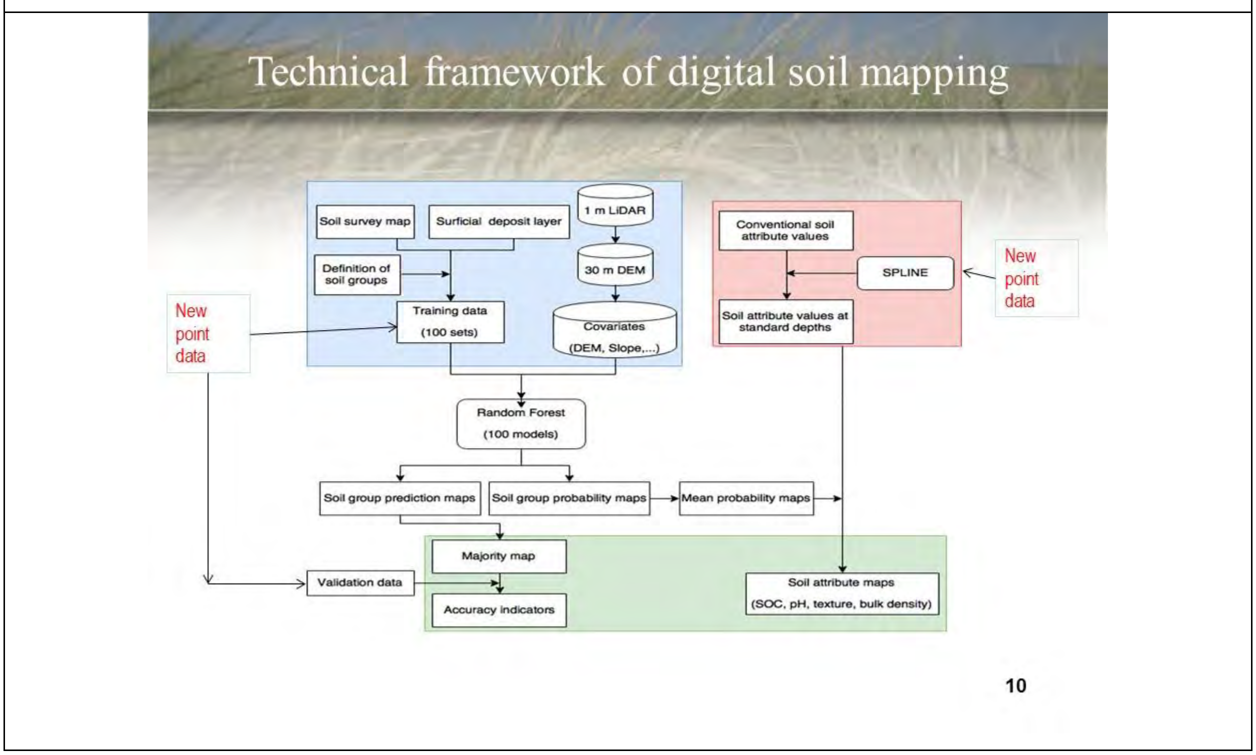
Nyiraneza, J., Thompson, B., Geng, X., He, J., Jiang, Y., Fillmore, S., and Stiles, K., 2017. Changes in soil organic matter over 18 years in Prince Edward Island, Canada. Can. J. of Soil Science, 97(4): 745-756, https://doi.org/10.1139/cjss-2017-0033

7

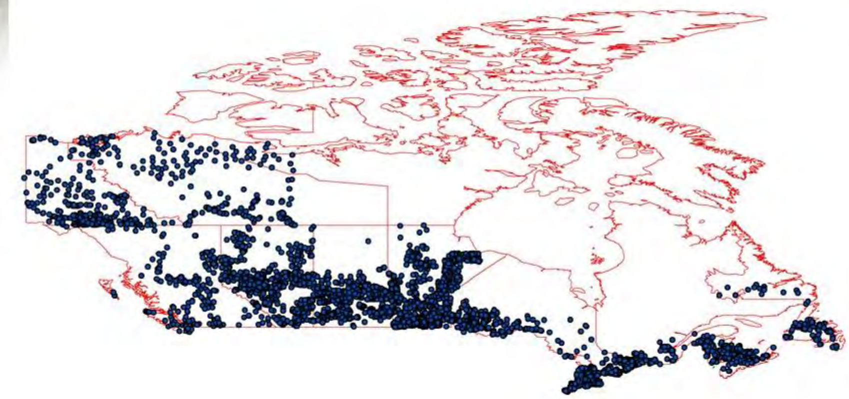
Opportunity
Innovation
Investment

8


Cost-effective way of soil data development
Digital soil mapping workflow
Geo-Data
Model specifications
3-D predictions
Soil profile measurements
Mass/Depth integrating function(s)
Covariates
Soil properties
Layer 1
Layer 2
z
y
x
Scaling/Spatial function(s)


Technical framework of digital soil mapping
Soil survey map
Surficial deposit layer
1 m LiDAR
30 m DEM
Definition of soil groups
Training data (100 sets)
Covariates (DEM, Slope,...)
New point data
Conventional soil attribute values
SPLINE
Soil attribute values at standard depths
New point data
Random Forest (100 models)
Soil group prediction maps
Soil group probability maps
Mean probability maps
Majority map
Validation data
Accuracy indicators
Soil attribute maps (SOC, pH, texture, bulk density)
10

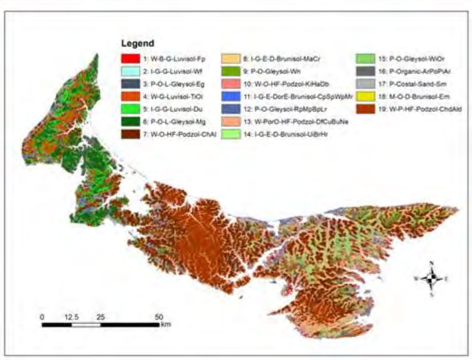
Global and National Data Development

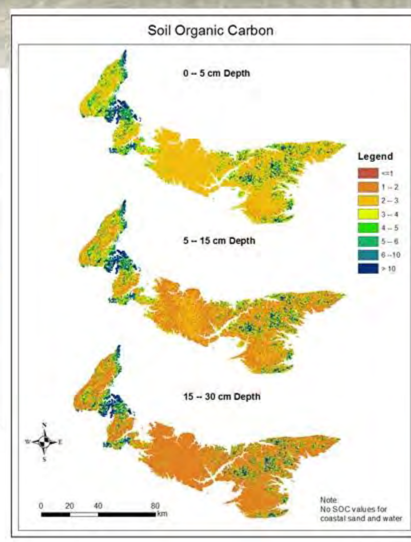


Hengl, T., J. M. Jesus, G. B . M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagoti, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N.H. Batjes, J.G.B. Leenaars, I. Wheeler, S. Mantel, B. Kempen, 2016. SoilGrids250m: global gridded soil information based on Machine Learning

11

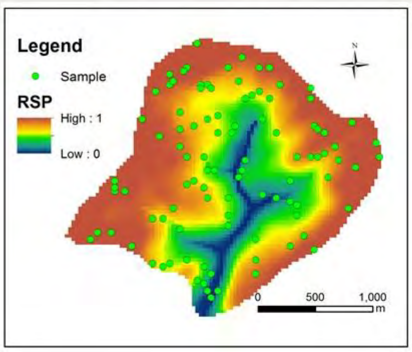
Regional higher resolution digital soil mapping

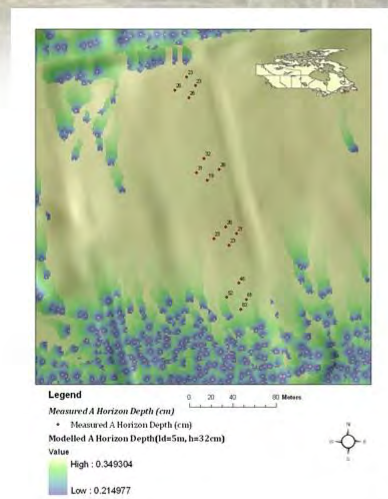




12

DSM at watershed and field scale

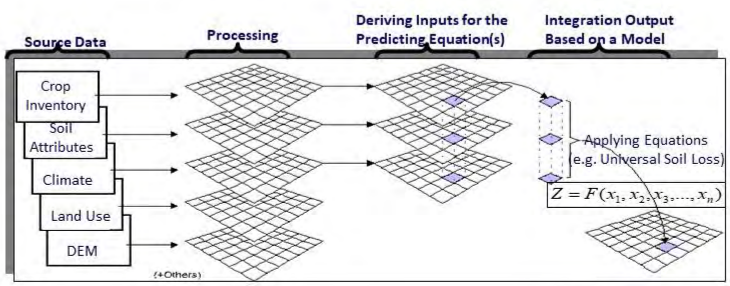




13

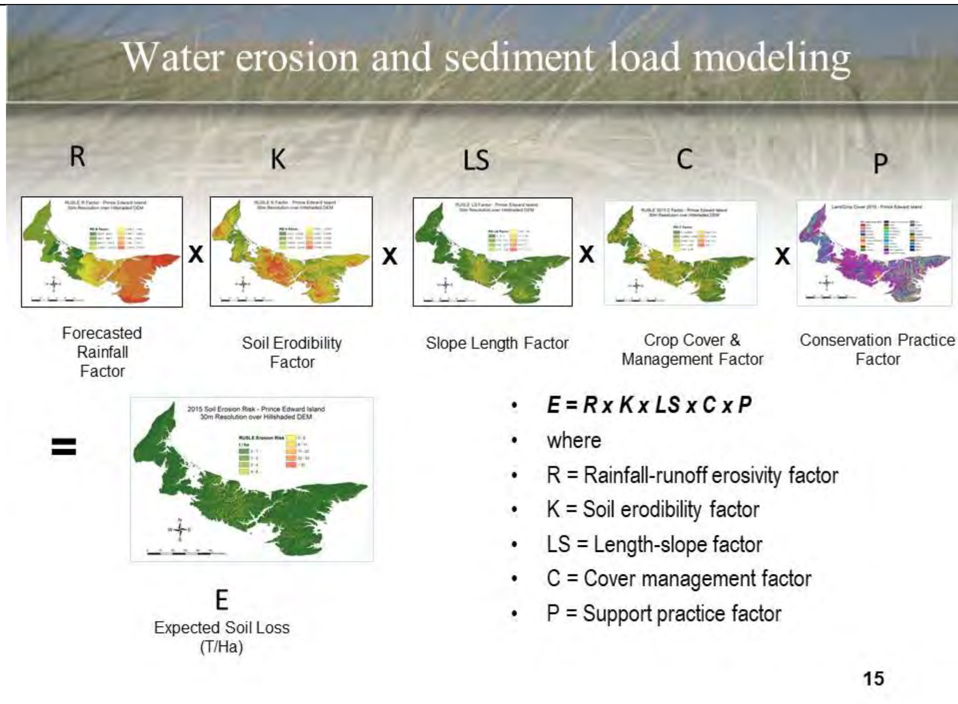
E

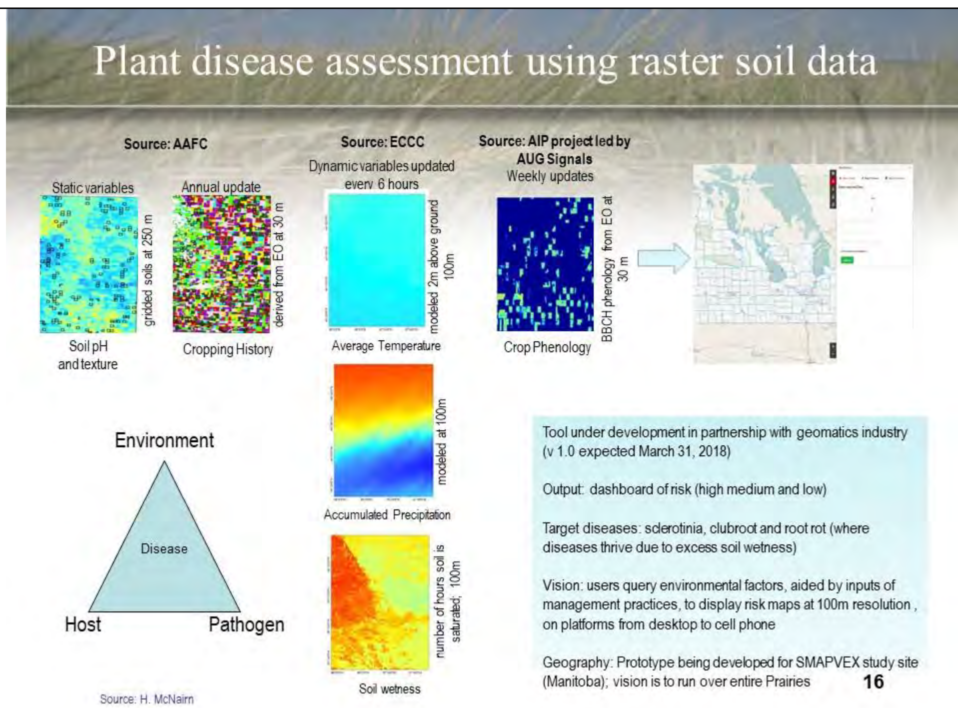
Integrated use of soil and soil landscape data

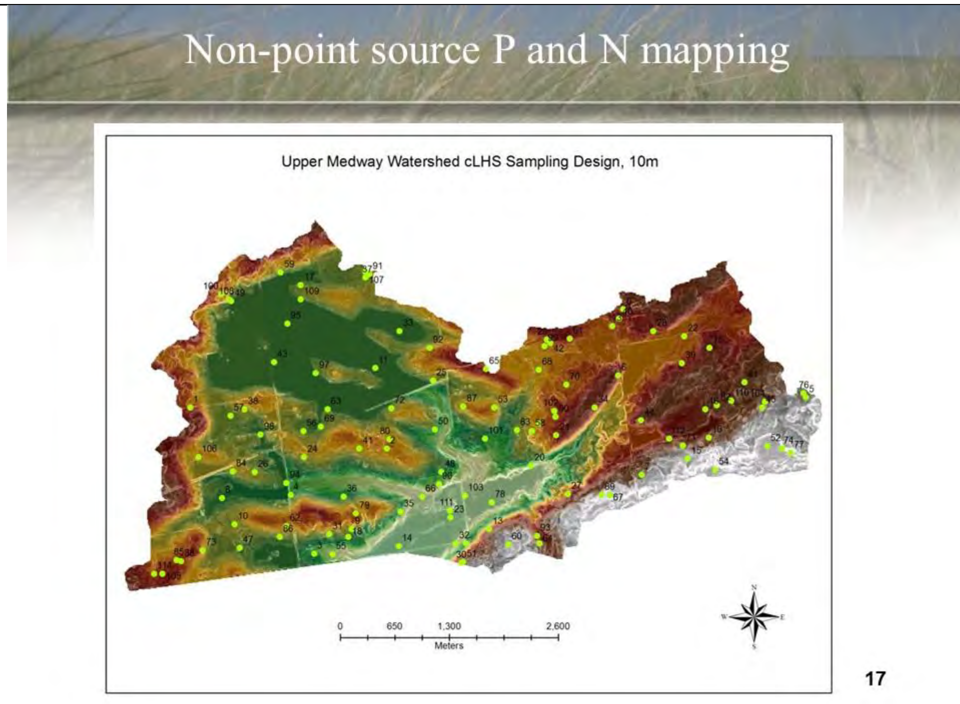


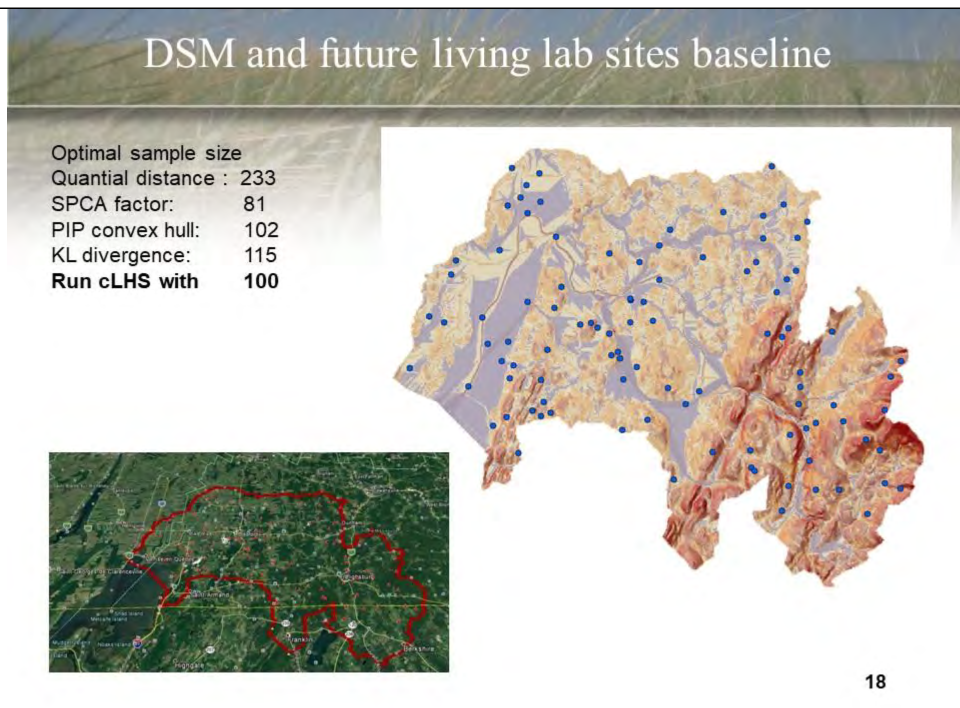
Source: B. Daneshfar

14


Water erosion and sediment load modeling
R
K
LS
C
P
x
x
x
x
Forecasted Rainfall Factor
Soil Erodibility Factor
Slope Length Factor
Crop Cover & Management Factor
Conservation Practice Factor
=
E
Expected Soil Loss (T/Ha)
• E = R x K x LS x C x P
• where
• R = Rainfall-runoff erosivity factor
• K = Soil erodibility factor
• LS = Length-slope factor
• C = Cover management factor
• P = Support practice factor
15


Plant disease assessment using raster soil data
Source: AAFC
Static variables
gridded soils at 250 m
Soil pH and texture
Annual update
derived from EO at 30 m
Cropping History
Source: ECCC
Dynamic variables updated every 6 hours
modeled 2m above ground 100m
Average Temperature
modeled at 100m
Accumulated Precipitation
number of hours soil is saturated; 100m
Soil wetness
Source: AIP project led by AUG Signals
Weekly updates
BBCH phenology from EO at 30 m
Crop Phenology
Environment
Disease
Host
Pathogen
Source: H. McNairn
Tool under development in partnership with geomatics industry (v 1.0 expected March 31, 2018)
Output: dashboard of risk (high medium and low)
Target diseases: sclerotinia, clubroot and root rot (where diseases thrive due to excess soil wetness)
Vision: users query environmental factors, aided by inputs of management practices, to display risk maps at 100m resolution, on platforms from desktop to cell phone
Geography: Prototype being developed for SMAPVEX study site (Manitoba); vision is to run over entire Prairies
16


Non-point source P and N mapping
Upper Medway Watershed cLHS Sampling Design, 10m
0
650
1,300
2,600
Meters
17


DSM and future living lab sites baseline
Optimal sample size
Quantial distance : 233
SPCA factor: 81
PIP convex hull: 102
KL divergence: 115
Run cLHS with 100
18

What is next?

- Advance current Canadian portion of the peer-reviewed global 250m digital soil data
- Develop robotic computing framework for finer resolution (between 10m to 100m grid size) soil and soil landscape data production across the agriculture regions of Canada.
- Customize soil and soil landscape data production and provision in priority regions including Lake Erie basin, Lake Champlain basin, PEI, and anticipated Living Lab sites.
- Link to Precision Agriculture, Phenomics, Resiliency Metrics, and soil microbiome initiatives.
- Advance national soil database structure by adding business driven attributes including soil total Phosphorus and soil Metagenomics

19

Data mining via disaggregation by ecodistrict

The fundamental assumption is that topography has a significant influence on the spatial arrangement of soil types and that these relationships can be used for disaggregation.

Within an Ecodistrict, climate will have less influence on spatial variabilities of soil. Topography, parent materials and vegetation are the dominant soil development co-variables.

The relationships between soil distribution and the covariant are not always linear. Similarity or integrated indices may be more suitable to model and inference the nonlinear relationship.

20

Stepwise Approach with Ecodistrict Unit

- Soil association and catena
  Group soils by association
  Knowledge mining by catena
- Soil taxonomy and local names
  Group soils by sub-group
- Soil similarity and fuzzy c-mean
  Similarity with fuzzy membership
  Supervised inference logics
- Manual and automation

istrict_sics_2_sl_cmp

OID	SL	CMP	SOIL_CODE	MODIFIER	LU	SLOPE	LOCSF	KIND	DRAINAGE	ORDER3	S_GROUP3	G_GROUP3	SOILKEY	ECODISTRIC	SL_V2_2
300	809001	1	MAB		A	B	H	M	W	CH	O	BC	ABMAB###	809	482215
301	809001	2	HND		A	C	H	M	W	CH	O	DBC	ABHND###	809	482215
302	809001	3	NUT		A	C	H	M	W	CH	R	DBC	ABNUT###	809	482215
303	809001	4	HDY		A	A	U	M	W	SZ	B	SO	ABHDY###	809	482215
304	809001	5	ZGW		N	A	U	M	P	GL	O	HG	ABZGW###	809	482215
305	809002	1	HND		A	C	H	M	W	CH	O	DBC	ABHND###	809	482114
306	809002	2	NUT		A	C	H	M	W	CH	R	DBC	ABNUT###	809	482114
307	809002	3	BFD		A	A	U	M	W	SZ	DB	SO	ABBFD###	809	482114
308	809002	4	ZGW		N	A	U	M	P	GL	O	HG	ABZGW###	809	482114
309	809003	1	LFE		A	B	H	M	W	CH	E	DBC	ABLFE###	809	482113
310	809003	2	BFD		A	A	U	M	W	SZ	DB	SO	ABBFD###	809	482113
311	809003	3	ZGW		N	A	U	M	P	GL	O	HG	ABZGW###	809	482113
312	809004	1	MAB		A	C	H	M	W	CH	O	BC	ABMAB###	809	482112
313	809004	2	HDY		A	A	U	M	W	SZ	B	SO	ABHDY###	809	482112
314	809004	3	ZGW		N	A	U	M	P	GL	O	HG	ABZGW###	809	482112
315	809005	1	HDY		A	B	H	M	W	SZ	B	SO	ABHDY###	809	482111
316	809005	2	ROL		A	B	H	M	W	CH	SZ	BC	ABROL###	809	482111
317	809005	3	ZGW		N	A	U	M	P	GL	O	HG	ABZGW###	809	482111
318	809006	1	MAB		A	C	H	M	W	CH	O	BC	ABMAB###	809	482108
319	809006	2	TVS		A	C	H	M	W	CH	CA	BC	ABTVS###	809	482108
320	809006	3	CFD		A	A	U	M	W	CH	O	BC	ABCFD###	809	482108
321	809006	4	ZGW		N	A	U	M	P	GL	O	HG	ABZGW###	809	482108

21

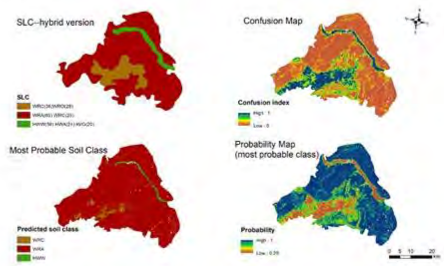
District 810:Unique soils and supporting information

SL_from_(	CMP_fron	SOIL_COD	LU_from_	PERCENT_	SLOPE_fro	DRAINAGI	MDEP1_fr	ORDER3	S_GROUP3	G_GROUP3	pm_code	pm_notes	pm_reclass
780003	1	RAA	A	57	A	W	LACU	VE	O	V	88	Well drained, fine/very fine, weakly/strongly calcareous lacustrine	88
780003	2	SUA	A	25	A	W	LACU	CH	O	DBC	88	Well drained, fine/very fine, weakly/strongly calcareous lacustrine	88
780003	3	EWA	A	7	A	W	LACU	CH	O	DBC	70	Well drained, medium/medium fine, weakly/strongly calcareous lacustrine	70
780003	4	TUA	A	5	A	W	LACU	SZ	DB	SZ	88	Well drained, fine/very fine, weakly/strongly calcareous lacustrine	88
780003	5	GXX	N	4	A	P	UNDM	GL	-	-	672	poorly drained, medium/medium fine, weakly/strongly calcareous Undifferentiated mineral	672
780003	6	RWW	N	2	B	W	UNDM	RG	-	-	72	Well drained, medium/medium fine, weakly/strongly calcareous Undifferentiated mineral	72
770007	1	BRA	A	35	A	W	FLLC	CH	O	DBC	70	Well drained, medium/medium fine, weakly/strongly calcareous fluviolacustrine	70
770007	3	STA	A	15	A	W	LACU	CH	O	DBC	70	Well drained, medium/medium fine, weakly/strongly calcareous lacustrine	70
770007	5	BRA	A	5	A	W	FLLC	CH	O	DBC	70	Well drained, medium/medium fine, weakly/strongly calcareous fluviolacustrine	70
770007	6	WRA	A	5	A	W	TILL	CH	O	DBC	66	Well drained, medium/medium fine, weakly/strongly calcareous glacial till	66
810001	3	WRC	A	15	B	W	TILL	CH	CA	DBC	66	Well drained, medium/medium fine, weakly/strongly calcareous glacial till	66
770008	5	AQC	A	8	A	R	FLUV	CH	CA	DBC	46	Well drained, moderately coarse, weakly/strongly calcareous fluvial	10
770008	6	HWW	N	8	D	W	UNDM	RG	-	-	72	Well drained, medium/medium fine, weakly/strongly calcareous Undifferentiated mineral	72
770008	7	BRA	A	7	A	W	FLLC	CH	O	DBC	70	Well drained, medium/medium fine, weakly/strongly calcareous fluviolacustrine	70
770008	9	BGA	A	5	A	VR	FLUV	CH	O	DBC	28	Well drained, very coarse, weakly/strongly calcareous fluvial	28
770008	10	VRA	N	5	A	VR	FLEO	RG	O	R	21	Well drained, coarse, medium/strongly acidic fluvioeolian	21
810002	2	HWA	N	24	D	W	UNDM	CH	-	-	72	Well drained, medium/medium fine, weakly/strongly calcareous Undifferentiated mineral	72
810002	3	AVG	N	20	A	P	FLUV	GL	-	-	665	poorly drained, medium/medium fine, weakly/strongly calcareous fluvial	665
810003	2	WRO	A	28	C	W	TILL	RG	O	R	66	Well drained, medium/medium fine, weakly/strongly calcareous glacial till	66
810003	5	WRE	A	9	C	W	TILL	CH	E	DBC	66	Well drained, medium/medium fine, weakly/strongly calcareous glacial till	66
813001	1	HRA	A	35	B	W	TILL	CH	O	BC	66	Well drained, medium/medium fine, weakly/strongly calcareous glacial till	66
770010	1	APA	N	41	D	VR	FLEO	RG	O	R	21	Well drained, coarse, medium/strongly acidic fluvioeolian	21
770010	3	HTA	A	12	A	R	FLUV	CH	O	BC	46	Well drained, moderately coarse, weakly/strongly calcareous fluvial	10
770010	4	AQA	A	9	A	R	FLUV	CH	O	DBC	46	Well drained, moderately coarse, weakly/strongly calcareous fluvial	10
770010	6	CHA	A	6	B	VR	FLUV	CH	O	BC	28	Well drained, very coarse, weakly/strongly calcareous fluvial	28
813002	2	VAA	A	15	B	W	LACU	CH	O	BC	70	Well drained, medium/medium fine, weakly/strongly calcareous lacustrine	70
813002	5	WWA	A	13	A	W	LACU	CH	O	BC	88	Well drained, fine/very fine, weakly/strongly calcareous lacustrine	88
813002	6	ADA	A	8	C	W	TILL	CH	O	BC	66	Well drained, medium/medium fine, weakly/strongly calcareous glacial till	66
813002	7	BYA	A	6	A	W	FLLC	CH	O	BC	70	Well drained, medium/medium fine, weakly/strongly calcareous fluviolacustrine	70
813005	4	FXA	A	15	A	W	LACU	CH	O	BC	70	Well drained, medium/medium fine, weakly/strongly calcareous lacustrine	70

22

Would DSMART Like Tool Work?

1 Prepare database of soil polygons, covariates
2 Take *m* random samples from each polygon
3 Assign a soil class to each sampling point
4 Build C4.5 decision tree with sampling points
5 Apply decision tree to map grid
6 Count number of times each grid cell is allocated to each soil type
7 Estimate probabilities of occurrence from counts
8 Generate maps



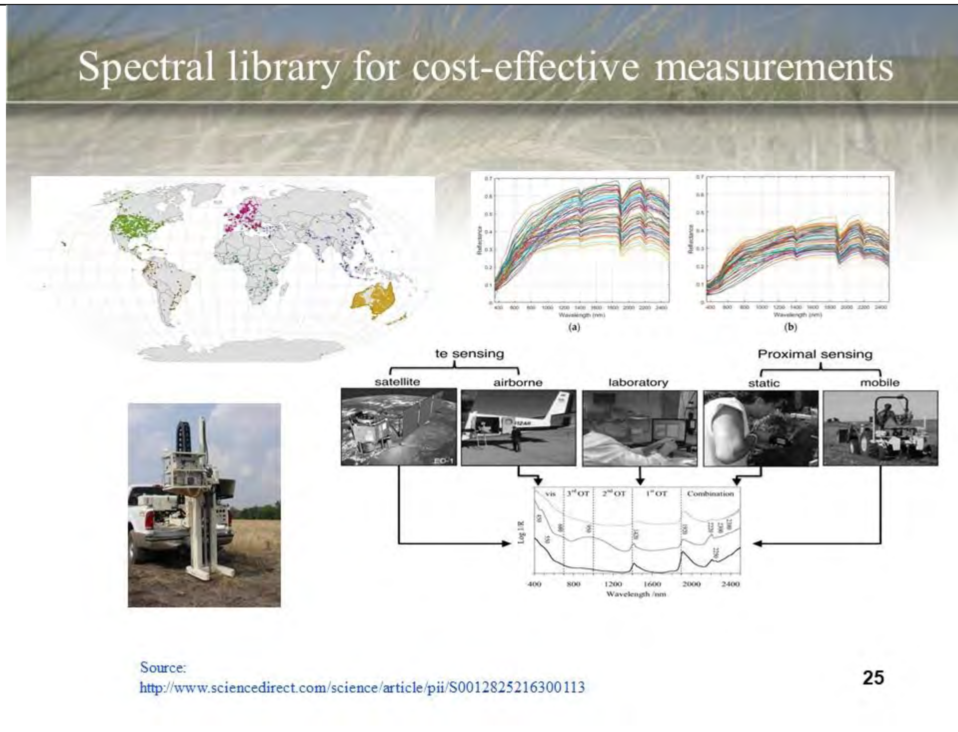
23

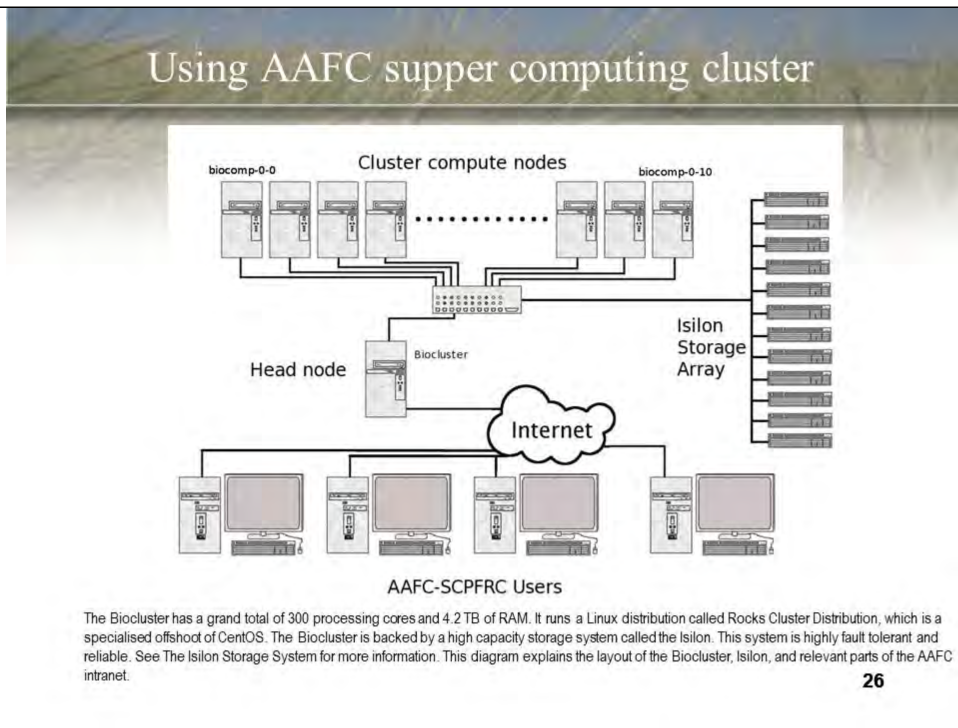
Extending national soil data collections:

Profiling soil microbiome using high-throughput sequencing

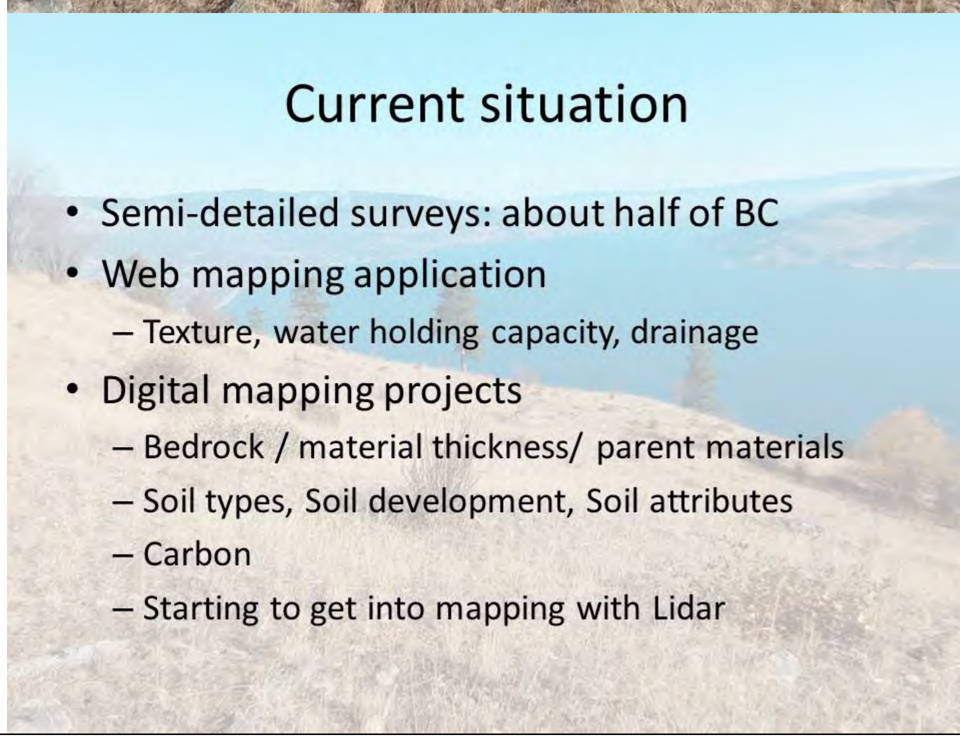
24

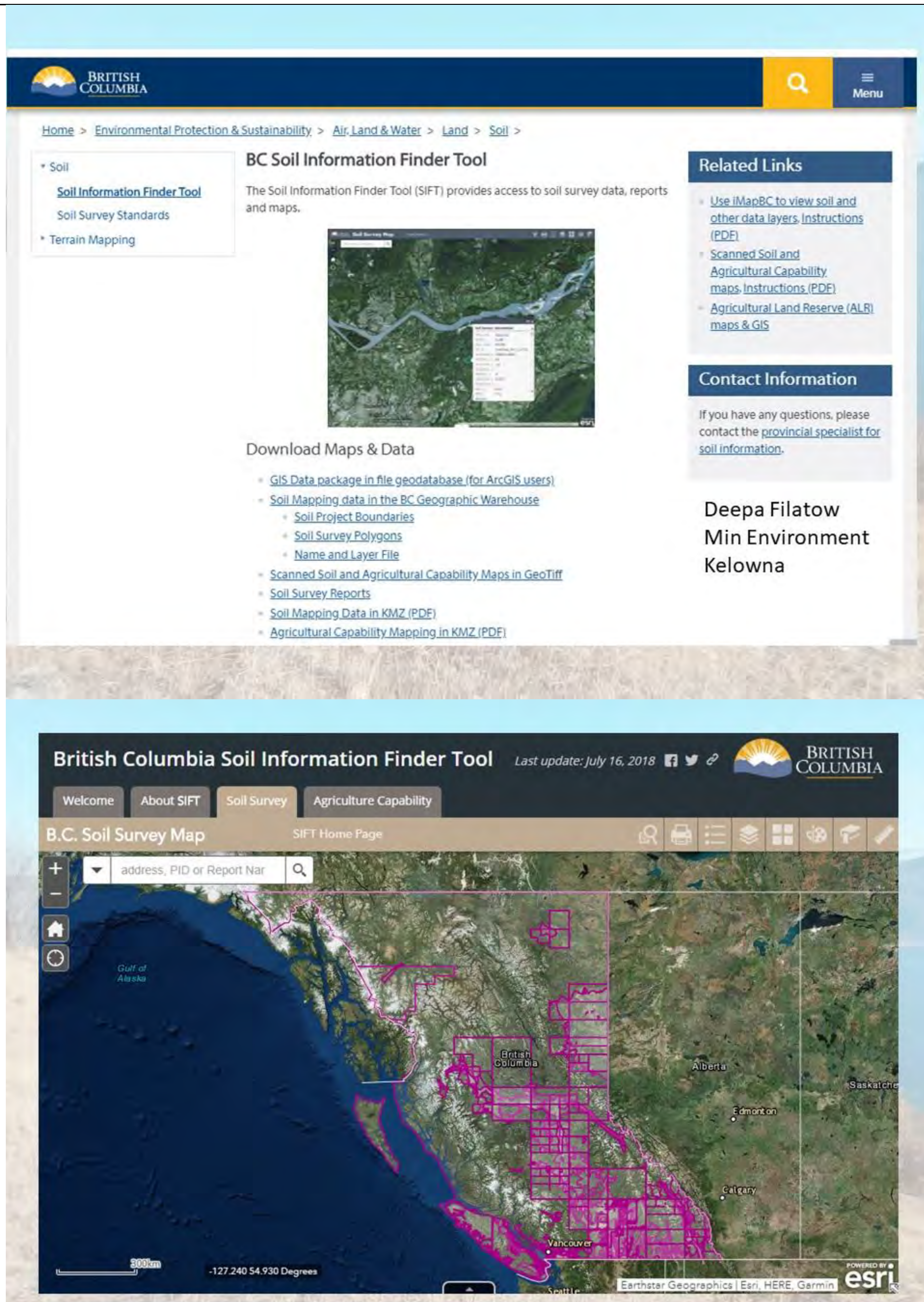
Robotics and miniaturization Source: W. Chen


Spectral library for cost-effective measurements
(a)
(b)
te sensing
Proximal sensing
satellite
airborne
laboratory
static
mobile
Wavelength /nm
Source:
http://www.sciencedirect.com/science/article/pii/S0012825216300113
25

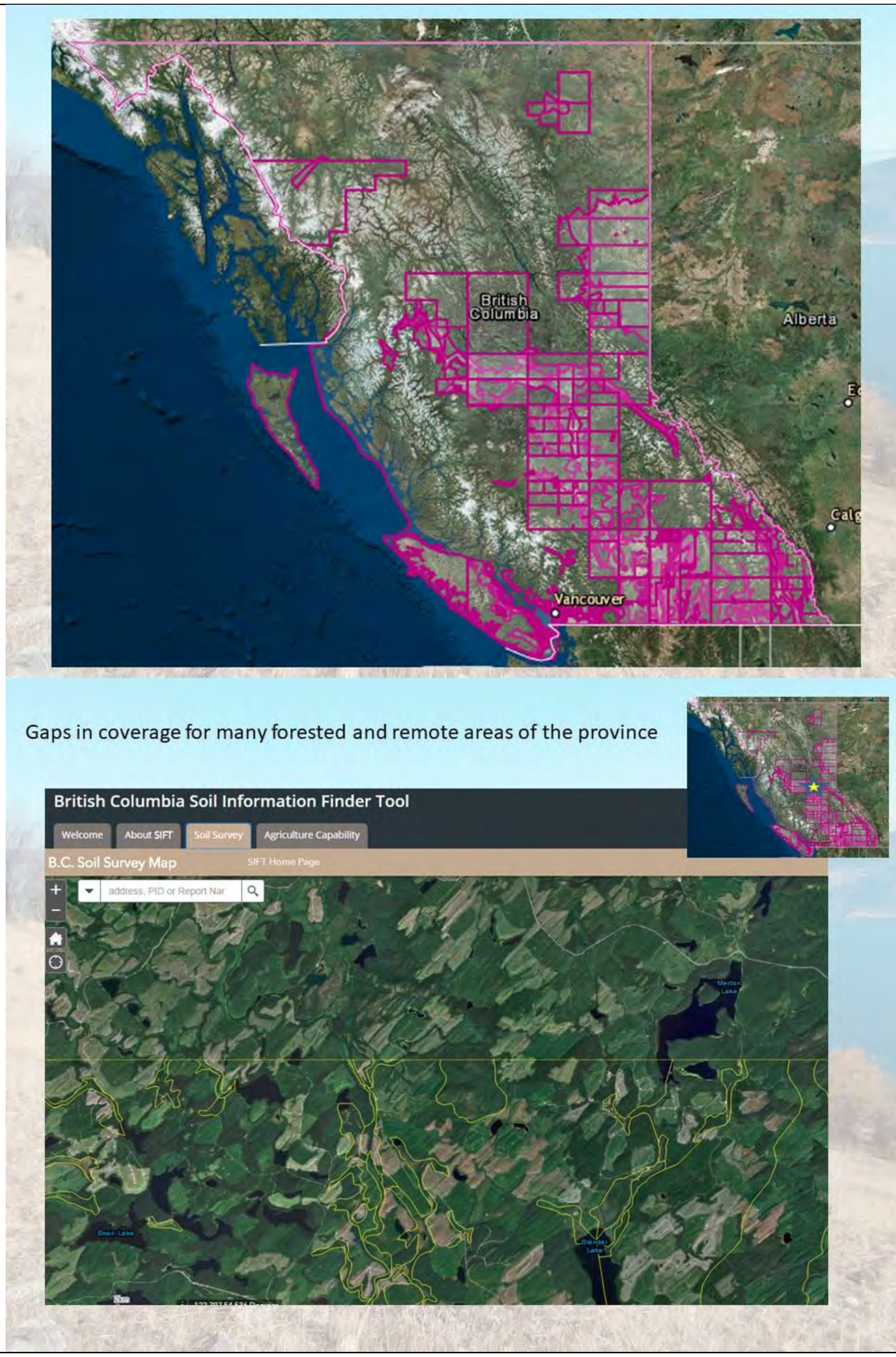

Using AAFC supper computing cluster
Cluster compute nodes
biocomp-0-0
biocomp-0-10
Isilon
Storage
Array
Head node
Biocluster
Internet
AAFC-SCPFRC Users
The Biocluster has a grand total of 300 processing cores and 4.2 TB of RAM. It runs a Linux distribution called Rocks Cluster Distribution, which is a specialised offshoot of CentOS. The Biocluster is backed by a high capacity storage system called the Isilon. This system is highly fault tolerant and reliable. See The Isilon Storage System for more information. This diagram explains the layout of the Biocluster, Isilon, and relevant parts of the AAFC intranet.
26

Predictive Mapping in Beautiful British Columbia – Chuck Bulmer

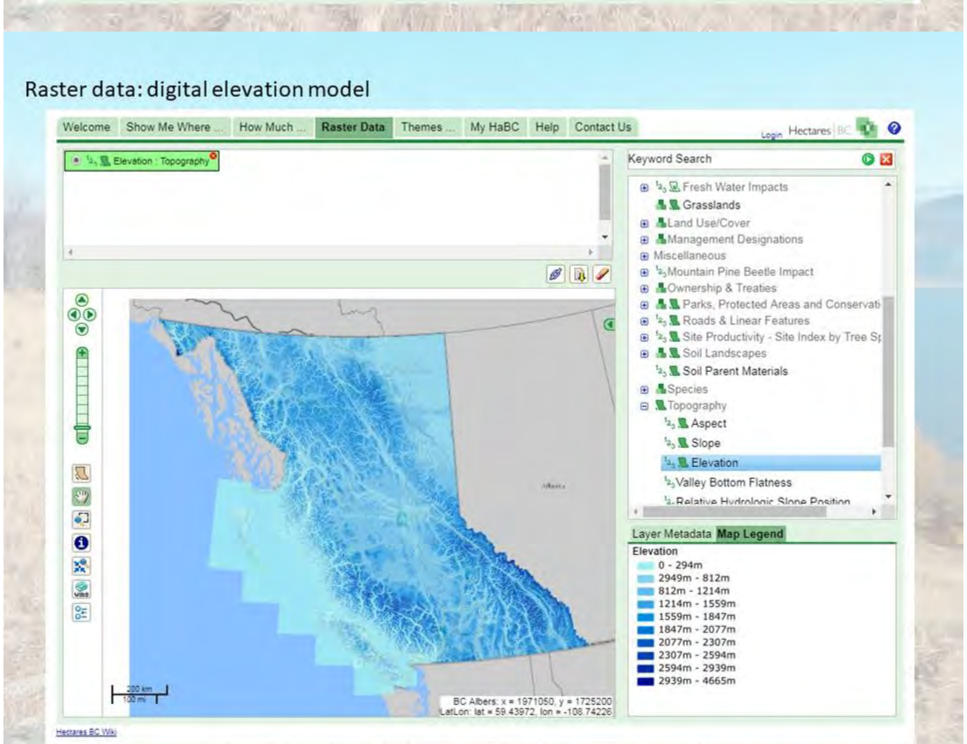


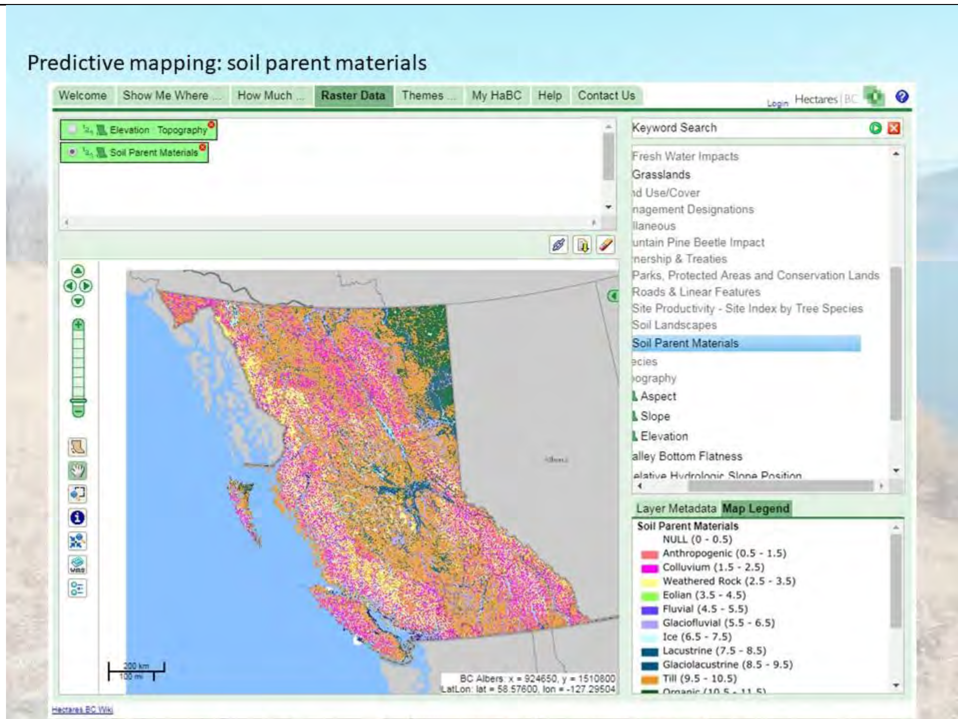


https://governmentofbc.maps.arcgis.com/apps/MapSeries/index.html?appid=cc25e43525c5471ca7b13d639bbcd7aa


British Columbia
Alberta
Vancouver
Gaps in coverage for many forested and remote areas of the province
British Columbia Soil Information Finder Tool
Welcome
About SIFT
Soil Survey
Agriculture Capability
B.C. Soil Survey Map
SIFT Home Page


Raster data available on hectaresbc.org
Welcome
Show Me Where ...
How Much ...
Raster Data
Themes ...
My HaBC
Help
Contact Us
Login
Hectares BC
Welcome to
Hectares | BC
Hints
Use the Keyword Search box (top right) to find Data Layers directly without having to expand long lists.
Recent Updates
December 1, 2017
Updated/Added: ROS North east Pilot 2016, Natural Gas Quality, PNG Tenure, Consolidated Cutblocks, Timber Supply Area (TSA), UWR, WHA, Parks and Protected Areas, Conservation Lands, Road Length, Distance To Road, WMA, Mineral, Placer, Coal Tenures, OGMA, Special Protected Area, Management Areas, VRI, BEC, Enhanced Wetland Classification, Calving Area Predictive Map Data
Migrated application to new server
more...
British Columbia
The Best Place on Earth
Nature Conservancy Canada
Fraser Salmon & Watersheds Program
GeoConnections
GéoConnexions
Fisheries and Oceans Canada
Pêches et Océans Canada
Parks Canada
Parcs Canada


Raster data: digital elevation model
Welcome
Show Me Where ...
How Much ...
Raster Data
Themes ...
My HaBC
Help
Contact Us
Login
Hectares BC
Elevation : Topography
Keyword Search
Fresh Water Impacts
Grasslands
Land Use/Cover
Management Designations
Miscellaneous
Mountain Pine Beetle Impact
Ownership & Treaties
Parks, Protected Areas and Conservati
Roads & Linear Features
Site Productivity - Site Index by Tree Sp
Soil Landscapes
Soil Parent Materials
Species
Topography
Aspect
Slope
Elevation
Valley Bottom Flatness
Relative Hydrologic Slope Position
Layer Metadata
Map Legend
Elevation
0 - 294m
2949m - 812m
812m - 1214m
1214m - 1559m
1559m - 1847m
1847m - 2077m
2077m - 2307m
2307m - 2594m
2594m - 2939m
2939m - 4665m
200 km
100 mi
BC Albers: x = 1971050, y = 1725200
LatLon: lat = 59.43972, lon = -108.74226
Hectares BC Wiki



Predictive mapping: soil parent materials

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Predictive soil parent material mapping at a regional-scale: A Random Forest approach

Brandon Heung [a], Chuck E. Bulmer [b], Margaret G. Schmidt [a,*]

[a] Soil Science Lab, Department of Geography, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada
[b] British Columbia Ministry of Forests Lands and Natural Resources Operations, Forest Sciences Section, Vernon, BC, V1B 2C7, Canada

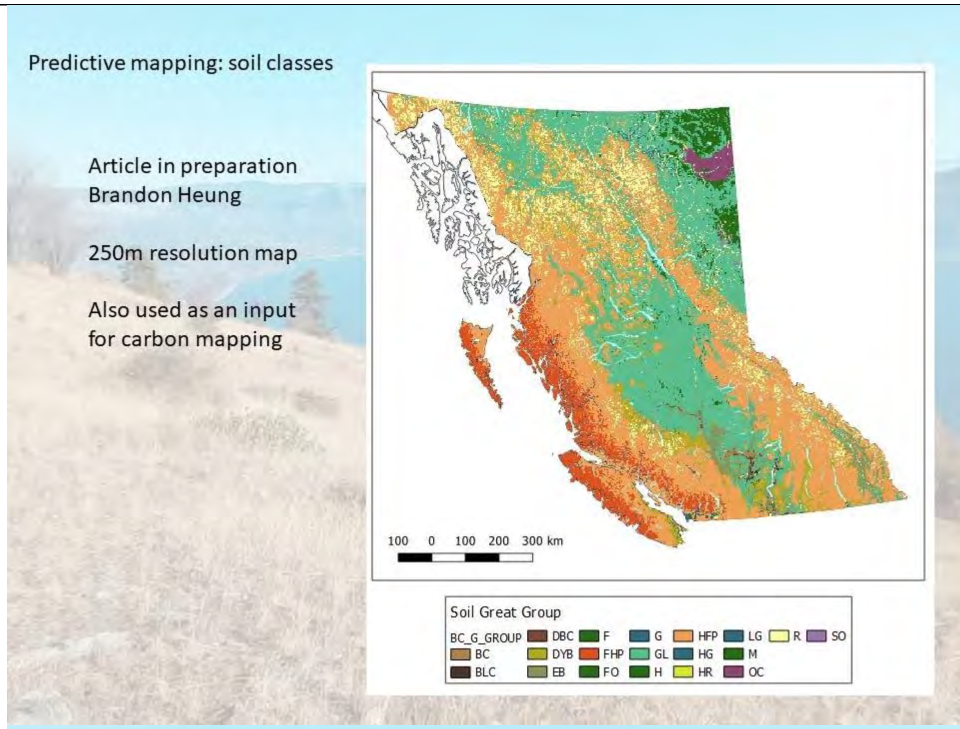
ARTICLE INFO

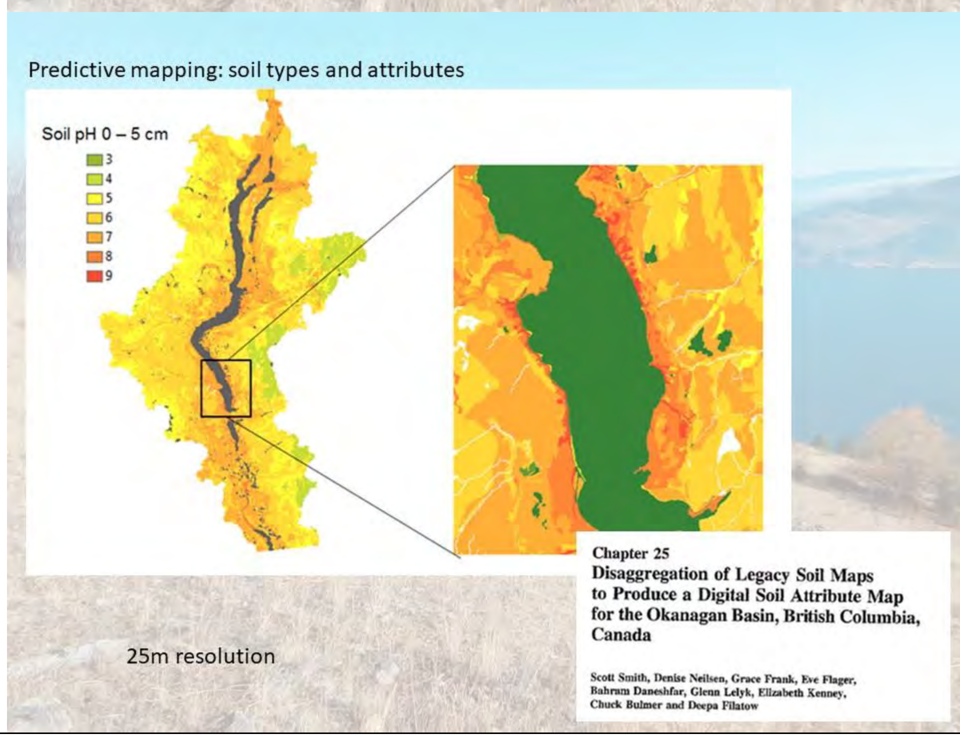
Article history:
Received 1 June 2013
Received in revised form 15 September 2013
Accepted 16 September 2013
Available online 11 October 2013

Keywords:
Soil parent material
Digital soil mapping
Knowledge discovery
Random forest
Environmental correlation
Machine learning

ABSTRACT

In this study, we evaluate the application of a Random Forest (RF) classifier as a tool for understanding and predicting the complex hierarchical relationships between soil parent material and topography using a digital elevation model (DEM) and conventional soil survey maps. Single-component soil polygons from conventional soil survey maps of the Langley–Vancouver Map Area, British Columbia (Canada), were used to generate randomized training points for 9 parent material classes. Each point was intersected with values from 27 topographic indices derived from a 100 m DEM. RF's m_{try} parameter was optimized using multiple replicates of 5-fold cross validation and parent material predictions were made for the region. Predictive parent material maps were validated through comparisons with legacy soil survey maps and 307 field points. Results show that predictions made by a non-optimized RF resulted in a kappa index of 89.6% when validated with legacy soil survey data from single-component polygons and a kappa index of 79.5% when validated with field data. Variable reduction and m_{try} optimization resulted in minimal improvements in RF predictions. Our results demonstrate the effectiveness of RF as a machine learning and data mining approach; however, the need for reliable training data was highlighted by less reliable results for polygon disaggregation in portions of the map where fewer training data points could be established.


Predictive mapping: soil classes
Article in preparation
Brandon Heung
250m resolution map
Also used as an input for carbon mapping
100 0 100 200 300 km
Soil Great Group
BC_G_GROUP
DBC
F
G
HFP
LG
R
SO
BC
DYB
FHP
GL
HG
M
BLC
EB
FO
H
HR
OC


Predictive mapping: soil types and attributes
Soil pH 0 – 5 cm
3
4
5
6
7
8
9
Chapter 25
Disaggregation of Legacy Soil Maps to Produce a Digital Soil Attribute Map for the Okanagan Basin, British Columbia, Canada
Scott Smith, Denise Neilsen, Grace Frank, Eve Flager, Bahram Daneshfar, Glenn Lelyk, Elizabeth Kenney, Chuck Bulmer and Deepa Filatow
25m resolution

Predictive mapping: bedrock and soil thickness

Geoderma 282 (2016) 59–69

Contents lists available at ScienceDirect

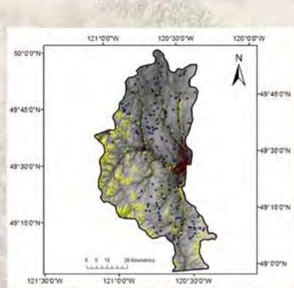
Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Modelling soil thickness in the critical zone for Southern British Columbia

Christopher Scarpone [a], Margaret G. Schmidt [a,*], Chuck E. Bulmer [b], Anders Knudby [c]

[a] Soil Science Lab, Department of Geography, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
[b] Department of Geography, University of Ottawa, 75 Laurier Ave E, Ottawa, Ontario K1N 6N5, Canada
[c] British Columbia Ministry of Forests Lands and Natural Resources Operations, Forest Sciences Section, Vernon, BC V1B 2C7, Canada



Geomorphology
Volume 285, 15 May 2017, Pages 214-224

Semi-automated classification of exposed bedrock cover in British Columbia's Southern Mountains using a Random Forest approach

Christopher Scarpone, Margaret G. Schmidt, Chuck E. Bulmer, Anders Knudby

Show more

https://doi.org/10.1016/j.geomorph.2017.02.013

Get rights and content

Predictive mapping: soil organic carbon

Environmental Research Letters

LETTER

Large, climate-sensitive soil carbon stocks mapped with pedology-informed machine learning in the North Pacific coastal temperate rainforest

Gavin McNicol [1], Chuck Bulmer [2], David D'Amore [3], Paul Sanborn [4,5], Sari Saunders [6], Ian Giesbrecht [6,7], Santiago Gonzalez Arriola [5], Allison Bidlack [3], David Butman [8] and Brian Buma [9]

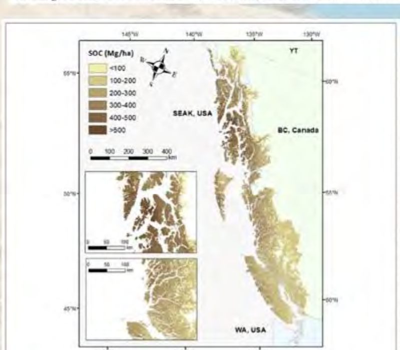


Figure 4. Soil organic carbon stock predictions to 1 m (Mg C ha^{-1}) at 90.3 m resolution for small NPCTR watersheds across BC and SE Alaska.

Used 1300 point s
Some from AK and BC
Almost 1000 BEC pts had soil horizons
$r^2 = .32$
Largest errors associated with spots in the landscape with very high C stock

Predictive mapping: soil classes and attributes near Kamloops

Jin Zhang

290 sample points

Soil class; pH; texture; C

Soilclass rf models OK

Attributes were lower

Predictive mapping: soil classes and attributes /carbon with LiDAR near Kamloops

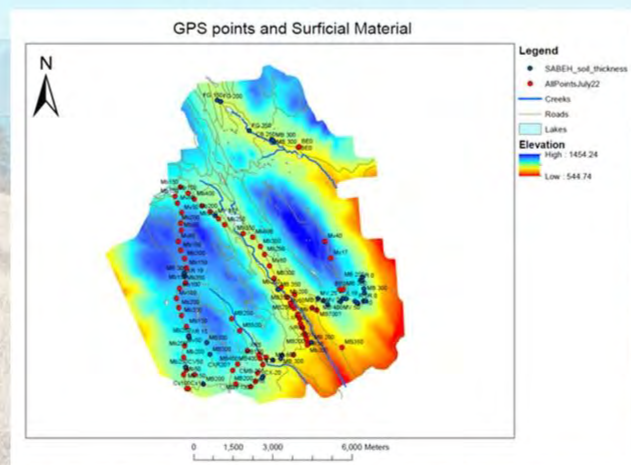
Babak Kasraei
William Bethel

200 points 30 cm

50 road cuts 100cm

Quick observations

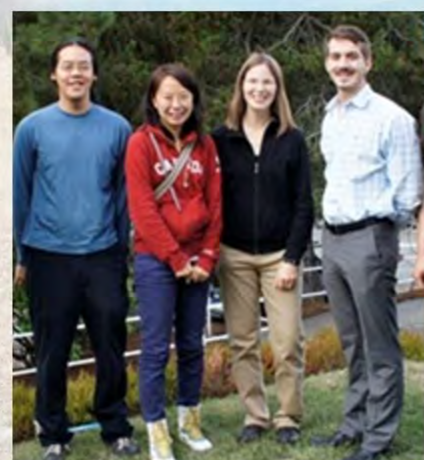
Rf regressions for carbonate depth; material thickness, and coarse frags have low R^2



Predictive mapping successes and challenges

- Vector soil datasets are readily available for large parts of the province (SIFT)
- Now we need to better incorporate the DSM products into web mapping
- We've had some early success, and now we are trying to improve our predictive maps
  - Possibly model attributes directly without soil classes / names ?
- BC has rugged topography
  - lots of variation in relief factor of soil formation
  - but parent material, veg and other factors often vary in 'unpredictable' ways
  - sometimes finer resolution maps are no better than coarse resolution
  - LiDAR has great potential (about 10 -20 percent of BC has LiDAR)
- We're getting better and building capacity
- DSM is here to stay

The team…



Ensemble Machine Learning as a Framework for PSM – Tom Hengl

These slides are periodically updated at the following link –
https://urldefense.proofpoint.com/v2/url?u=https-3A__docs.google.com_presentation_d_1QstVEF4pVKrntGCmRd9OWIRUSgO0SZfRXgkWoyjfP94_edit-3Fusp-3Dsharing&d=DwID-g&c=Nx4cNT2ND5PKJMUw1Yiqvt5PYVXmyHyY3_zNPmJl6pU&r=Eh0WZEy_sZ0BIi5Pzh7zGn81ISx1CpSbaYWJwZLAAGXAm2j72_VH5SiSG3Ku4GWQ&m=QXDQubowTMU4-pnSyOb5aqFK9N8zii4DXay-DWYvTu0&s=PS2zCbbiFR49iSewOQu3239mcOCqOJC1Q3vYVp5mMIU&e=

OpenGeoHUB
Connect · Create · Share · Repeat

Ensemble Machine Learning as a framework for PSM

March 6th 2019, InnoTech Alberta

tom.hengl@opengeohub.org
@tom_hengl
thengl
https://opengeohub.org / https://envirometrix.net

About me

OpenGeoHUB
Connect · Create · Share · Repeat

Co-founder of **OpenGeoHub foundation** (https://opengeohub.org) and co-founder / technical director at **EnvirometriX Ltd** (https://envirometrix.net).

My special interests:

- greener / cleaner / healthier planet, open science (education), open source, democracy...

My special skills:

- **Machine Learning on spatial data**, geocomputing, importing and processing (Large) environmental data, soil mapping, organizing and motivating groups to do complex problem solving, sharing...

This talk

- Introduction to **Predictive Soil Mapping**
- A gentle introduction to Machine Learning for soil mapping (soil properties, classes, probabilities etc)
  - the "black-box" problems,
- Some newest developments (**Ensemble Machine Learning**) and some obvious benefits:
  - maximizing accuracy / data processing automation,
  - modelling of non-linear complex relationships,
  - combining geography / feature space patterns without too many strict statistical assumptions,

Questions I (Craig Aumann)

- **Q1: At a high-level, how does PSM work?**
- **Q2: How good are PSM maps today?** While "accuracy" has certainly improved with the new statistical/ML techniques, are things "good enough" now, or are we still waiting for additional technical developments so that things become "good enough"?
- **Q3: Is PSM ready for commercial use to support the needs of government and others?**

Questions II (Craig Aumann)

- **Q4: What are the risks of using predictive mappings in decision making?**
- **Q5: What opportunities do predictive maps open up that we couldn't do before?**
- **Q6: What does this costs generally?**
- **Q7: What are others around the world doing around predictive soil mapping? Are they having successes? And how are they defining "success"?**

Evolution

- ⇓ 1999–2003: Kriging (in 1999 I was very excited about it),
- ⇓ 2001–2007: Regression-kriging (main topic of my PhD thesis!),
- ⇓ 2014–...: Random forest (Africa 250m),
- ⇓ 2016–...: Random forest + XGboost + SVM (global 250m),
- ⇓ 2018–...: Random forest spatial (RFsp),
- ⇓ 2018–...: Ensemble Machine Learning spatial (EMLsp),
- ⇓ 2019–...: Global spatiotemporal EMLsp,

Predictive Soil Mapping with R by T. (Tom) Hengl and R.A. (Bob) MacMillan

build passing

The text and code for the forthcoming Open Access book: **Predictive Soil Mapping with R**. Pull requests and general comments are welcome.

This book is continuously updated and contributions are welcome (pull requests and general comments are welcome) provided that some minimum requirements are met. For news and updates please refer to the github issues. Older versions of the book can be found at: https://doi.org/10.5281/zenodo.2561034

Cite this as:

- Hengl, T., MacMillan, R.A., (2019). **Predictive Soil Mapping with R**. OpenGeoHub foundation, Wageningen, the Netherlands, 370 pages, www.soilmapper.org, ISBN: 978-0-359-30635-0.

Hard copies of this book can be ordered from www.lulu.com. By purchasing a hard copy of this book from Lulu you donate $12 to the OpenGeoHub foundation.

6 Machine Learning Algorithms for soil mapping

Edited by: T. Hengl

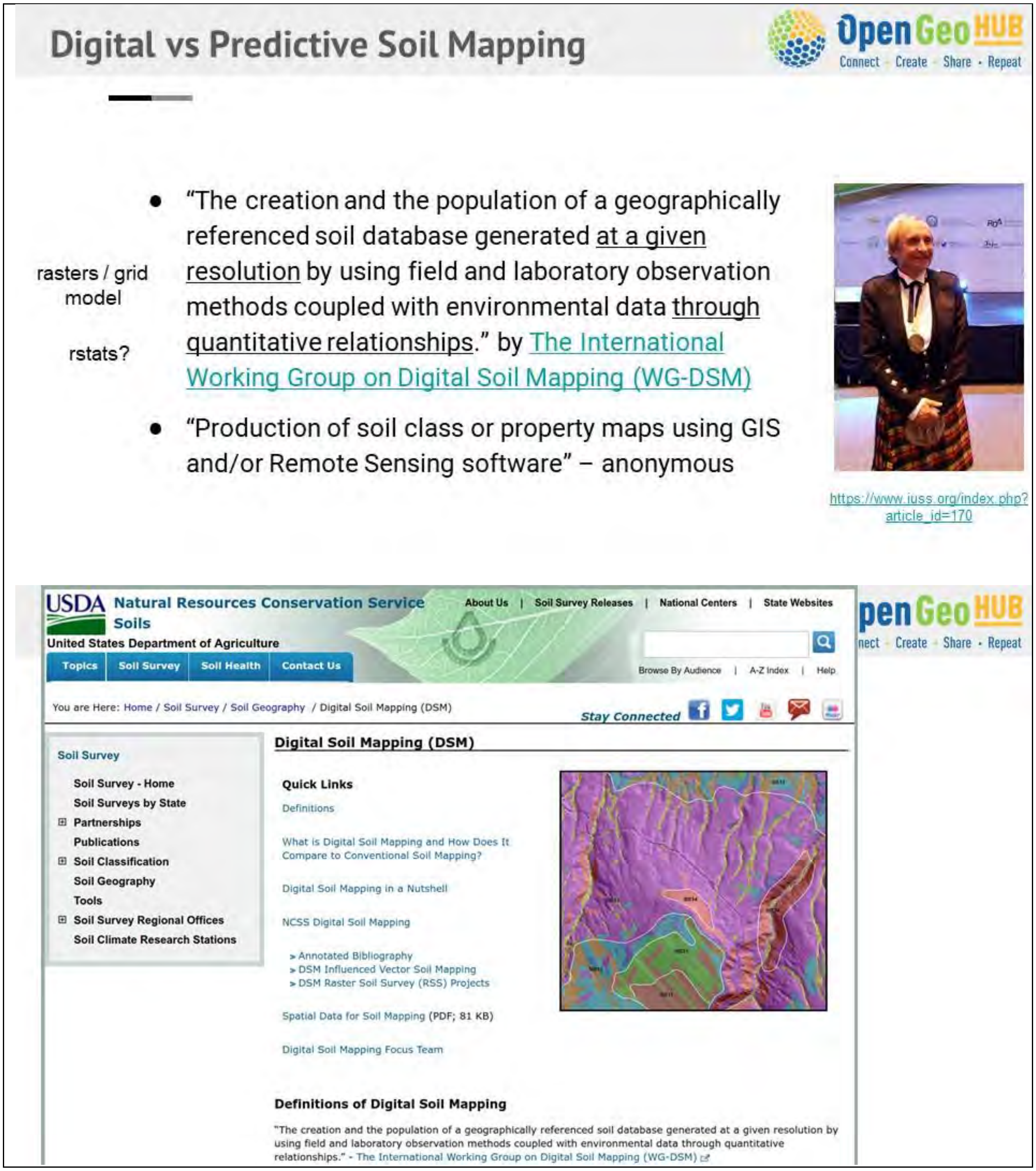
6.1 Spatial prediction of soil properties and classes using MLA's

This chapter reviews some common Machine learning algorithms (MLA's) that have demonstrated potential for soil mapping projects i.e. for generating spatial predictions (Brungard et al. 2015; Heung et al. 2016; Thorsten Behrens et al. 2018). In this tutorial we especially focus on using tree-based algorithms such as random forest, gradient boosting and Cubist. For a more in-depth overview of machine learning algorithms used in statistics refer to the CRAN Task View on Machine Learning & Statistical Learning. Some other examples of how MLA's can be used to fit Pedo-Transfer-Functions can be found in section 3.9.

6.1.1 Loading the packages and data

We start by loading all required packages:

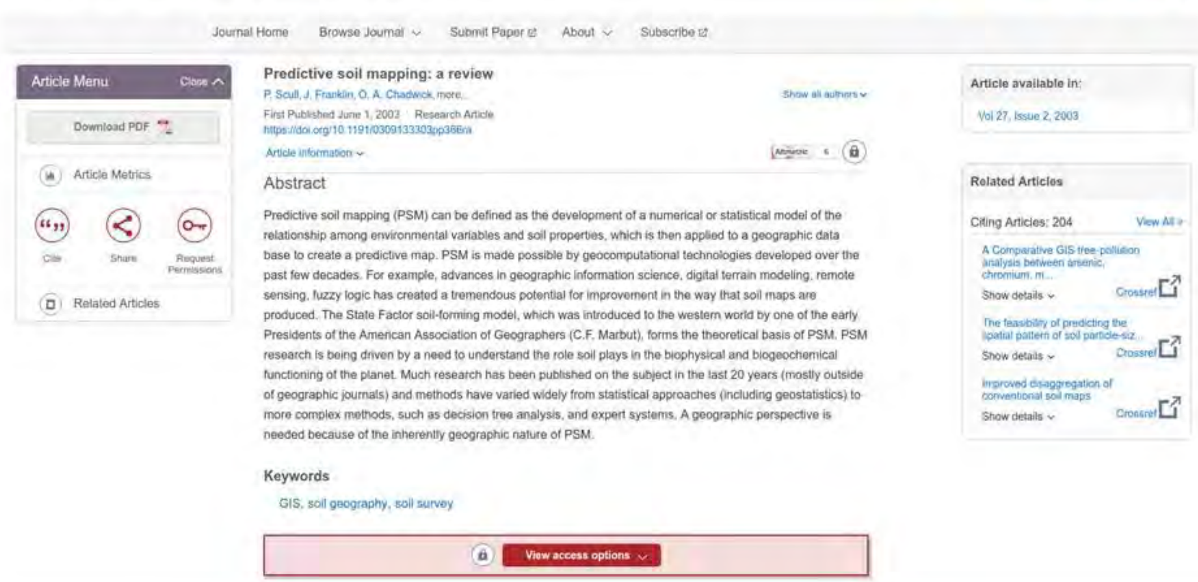
```
library(plotKML)
#> plotKML version 0.5-9 (2019-02-04)
#> URL: http://plotkml.r-forge.r-project.org/
library(sp)
library(randomForest)
```



Digital vs Predictive Soil Mapping
OpenGeoHUB
Connect · Create · Share · Repeat
rasters / grid model
rstats?
"The creation and the population of a geographically referenced soil database generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships." by The International Working Group on Digital Soil Mapping (WG-DSM)
"Production of soil class or property maps using GIS and/or Remote Sensing software" – anonymous
https://www.iuss.org/index.php?article_id=170
USDA Natural Resources Conservation Service
Soils
United States Department of Agriculture
About Us | Soil Survey Releases | National Centers | State Websites
Topics
Soil Survey
Soil Health
Contact Us
Browse By Audience | A-Z Index | Help
You are Here: Home / Soil Survey / Soil Geography / Digital Soil Mapping (DSM)
Stay Connected
Soil Survey
Soil Survey - Home
Soil Surveys by State
Partnerships
Publications
Soil Classification
Soil Geography
Tools
Soil Survey Regional Offices
Soil Climate Research Stations
Digital Soil Mapping (DSM)
Quick Links
Definitions
What is Digital Soil Mapping and How Does It Compare to Conventional Soil Mapping?
Digital Soil Mapping in a Nutshell
NCSS Digital Soil Mapping
> Annotated Bibliography
> DSM Influenced Vector Soil Mapping
> DSM Raster Soil Survey (RSS) Projects
Spatial Data for Soil Mapping (PDF; 81 KB)
Digital Soil Mapping Focus Team
Definitions of Digital Soil Mapping
"The creation and the population of a geographically referenced soil database generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships." - The International Working Group on Digital Soil Mapping (WG-DSM)

Predictive Soil Mapping

OpenGeoHUB Connect · Create · Share · Repeat



Journal Home Browse Journal Submit Paper About Subscribe

Article Menu Close

Download PDF

Article Metrics

Cite Share Request Permissions

Related Articles

Predictive soil mapping: a review

P. Scull, J. Franklin, O. A. Chadwick, more... Show all authors

First Published June 1, 2003 Research Article

https://doi.org/10.1191/0309133303pp366ra

Article information

Abstract

Predictive soil mapping (PSM) can be defined as the development of a numerical or statistical model of the relationship among environmental variables and soil properties, which is then applied to a geographic data base to create a predictive map. PSM is made possible by geocomputational technologies developed over the past few decades. For example, advances in geographic information science, digital terrain modeling, remote sensing, fuzzy logic has created a tremendous potential for improvement in the way that soil maps are produced. The State Factor soil-forming model, which was introduced to the western world by one of the early Presidents of the American Association of Geographers (C.F. Marbut), forms the theoretical basis of PSM. PSM research is being driven by a need to understand the role soil plays in the biophysical and biogeochemical functioning of the planet. Much research has been published on the subject in the last 20 years (mostly outside of geographic journals) and methods have varied widely from statistical approaches (including geostatistics) to more complex methods, such as decision tree analysis, and expert systems. A geographic perspective is needed because of the inherently geographic nature of PSM.

Keywords

GIS, soil geography, soil survey

View access options

Article available in:

Vol 27, Issue 2, 2003

Related Articles

Citing Articles: 204 View All

A Comparative GIS tree-pollution analysis between arsenic, chromium, m...

Show details Crossref

The feasibility of predicting the spatial pattern of soil particle-siz...

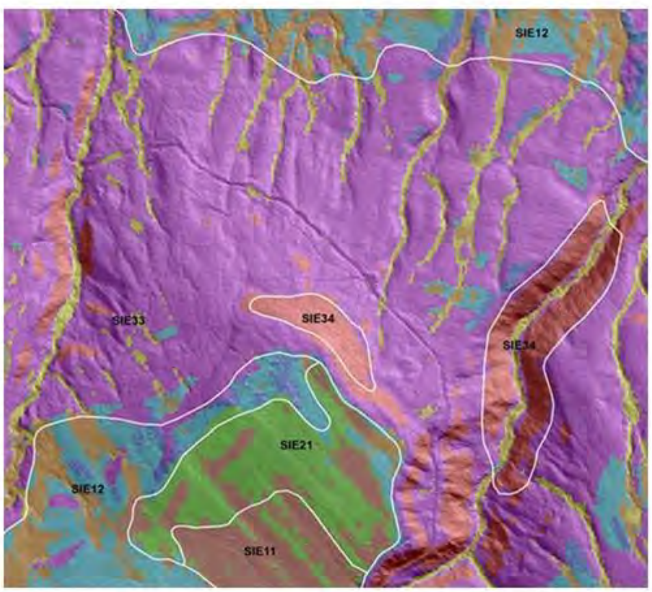
Show details Crossref

Improved disaggregation of conventional soil maps

Show details Crossref

SMUs vs grids

OpenGeoHUB Connect · Create · Share · Repeat



In conventional mapping, the primary question is "Where is the boundary between two soil types / polypedons?"

In Predictive Soil Mapping, **the question is what is the most objective (statistical) estimate of a value of some soil property / probability of some feature given the training data (points)**.

CLORPT

- Theoretically speaking, PSM is largely based on CLORPT.
- where **cl** stands for climate, **o** for organisms (including humans), **r** is relief, **p** is parent material or geology and **t** is time (originally presented by Jenny; 1994).

PSM basic concepts

- In statistical terms, **the main objective of soil mapping is to describe the spatial complexity of soils**, then represent this complexity using maps, summary measures, mathematical models and simulations. From the application point of view, the main application objective of soil mapping is **to accurately predict response of a soil(-plant) ecosystem to various soil management strategies**.
- The objective of PSM is **to produce optimal unbiased predictions of a mean value at some new location along with the uncertainty associated with the prediction, at the finest possible resolution**.

Q6

Q6: What does PSM costs generally?

Standard soil survey costs

Mapping scale also directly determines the survey costs. Burrough et al., Bie and Ulph, and Bie et al. postulated in early 70s that the survey costs are a direct function of the mapping scale:

$$\log \left\{ \begin{matrix} \text{cost per km}^2 \\ \text{or} \\ \text{man} - \text{days per km}^2 \end{matrix} \right\} = a + b \cdot \log(\text{map scale}) \quad (5)$$

This model typically explains >75% of the survey costs (Burrough et al., 1971). Based on Eq. (5), we have fitted a linear model to the empirical table data from e.g. Legros (2004, p.75), to produce the following model:

$$\hat{X} = \exp\left[19.0825 - 1.6232 \cdot \log(\text{SN})\right] \qquad (6)$$

where X is the minimum cost/ha in Euros (based on es-

In 2019, it depends on multiple things

Three main types of PSM projects:

A. **PSM projects in new, previously unmapped, areas** — no point observations or samples currently exist.
B. **PSM projects using legacy points** — sufficient point data to support PSM exist and are available, but no previous PSM modelling has been implemented for this area.
C. **PSM projects aimed at optimizing predictions and usability** — Previous PSM models have already been completed but previous results can still be improved / optimized.

PSM type A

Total costs = FC + [VC + PC + DC]

where FC are fixed costs, VC are variable cost based on the number of variables, PC are variable cost based on the number of points, and DC are the variable costs based on the total distance between all points.

Total costs are a function of:

- the size of the area / spatial resolution,
- total number of samples/points,
- total number of variables, and
- distance that needs to be crossed to visit all points (includes distance between study sites),

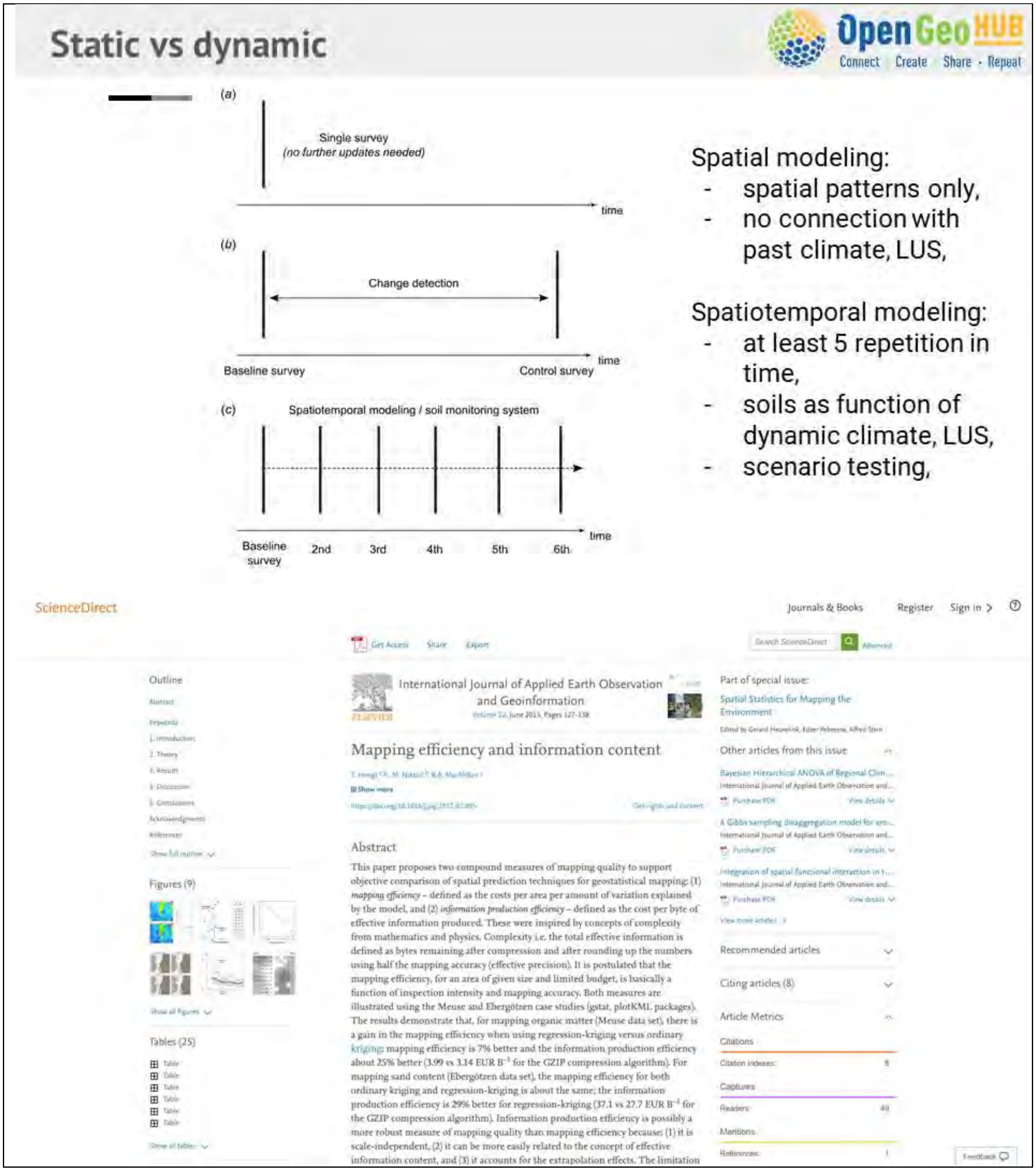
Costs structure

1. **Initiation costs** (feasibility study, implementation plan, basic infrastructure hardware and software),
2. **Field work and laboratory analysis costs**:
   a. per sample, per day of field work, per area, per hour of computing, per 1GiB of data,
3. **Data analysis / data mining (programming and computing) costs**
4. **Data distribution and serving costs:**
   a. back-end, front-end development costs: web hosting and web-services,
5. **Consulting, design, PR and quality control costs**

Another look at the costs structure

PSM projects can also be classified into:

1. PSM projects for the purpose of mapping static (stable) **spatial patterns only**.
2. PSM projects for the purpose of **one-time change detection** (e.g. two time intervals).
3. PSM projects for the purpose of **monitoring soil conditions / status** (continuous updates at regular intervals).


Static vs dynamic
OpenGeoHUB
Connect · Create · Share · Repeat
(a)
Single survey
(no further updates needed)
time
(b)
Change detection
time
Baseline survey
Control survey
(c)
Spatiotemporal modeling / soil monitoring system
time
Baseline survey
2nd
3rd
4th
5th
6th
Spatial modeling:
- spatial patterns only,
- no connection with past climate, LUS,
Spatiotemporal modeling:
- at least 5 repetition in time,
- soils as function of dynamic climate, LUS,
- scenario testing,
ScienceDirect
Journals & Books
Register
Sign in
Get Access
Share
Export
Outline
Figures (9)
Tables (25)
International Journal of Applied Earth Observation and Geoinformation
Part of special issue:
Spatial Statistics for Mapping the Environment
Mapping efficiency and information content
Show more
Abstract
This paper proposes two compound measures of mapping quality to support objective comparison of spatial prediction techniques for geostatistical mapping: (1) mapping efficiency – defined as the costs per area per amount of variation explained by the model, and (2) information production efficiency – defined as the cost per byte of effective information produced. These were inspired by concepts of complexity from mathematics and physics. Complexity i.e. the total effective information is defined as bytes remaining after compression and after rounding up the numbers using half the mapping accuracy (effective precision). It is postulated that the mapping efficiency, for an area of given size and limited budget, is basically a function of inspection intensity and mapping accuracy. Both measures are illustrated using the Meuse and Ebergötzen case studies (gstat, plotKML packages). The results demonstrate that, for mapping organic matter (Meuse data set), there is a gain in the mapping efficiency when using regression-kriging versus ordinary kriging: mapping efficiency is 7% better and the information production efficiency about 25% better (3.99 vs 3.14 EUR B^{-1} for the GZIP compression algorithm). For mapping sand content (Ebergötzen data set), the mapping efficiency for both ordinary kriging and regression-kriging is about the same; the information production efficiency is 29% better for regression-kriging (37.1 vs 27.7 EUR B^{-1} for the GZIP compression algorithm). Information production efficiency is possibly a more robust measure of mapping quality than mapping efficiency because: (1) it is scale-independent, (2) it can be more easily related to the concept of effective information content, and (3) it accounts for the extrapolation effects. The limitation
Other articles from this issue
Recommended articles
Citing articles (8)
Article Metrics
Citations
Captures
Mentions
Feedback

Information production efficiency

Second equivalent measure of mapping efficiency is the *information production efficiency* (IPE):

$$\Upsilon = \frac{X}{\text{gzip}} \qquad [\text{EURB}^{-1}] \tag{17}$$

where gzip is the size of data (in Bytes) left after compression and after reformatting the values to match the effective precision (based on Eq. (13)). This measure is scale independent as the area is not included in the equation. The requirement for an accurate estimate of the information production efficiency is that both a reliable estimate of the model uncertainty (extrapolation

Numeric resolution

Hengl et al. / *Special Issue SS2011 (2012) 1–16* 9

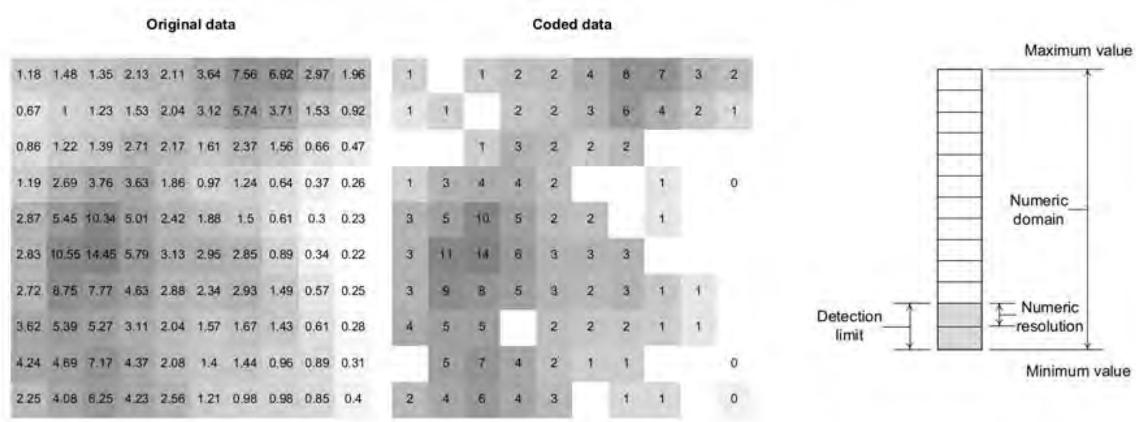
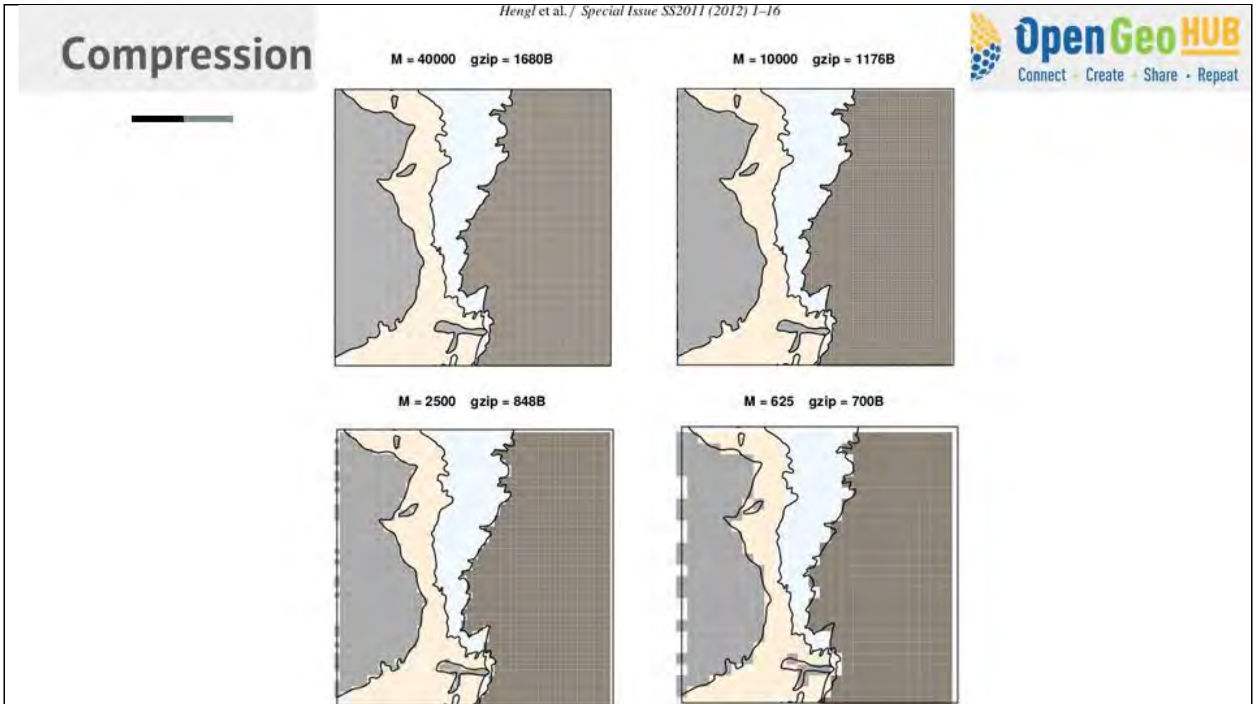
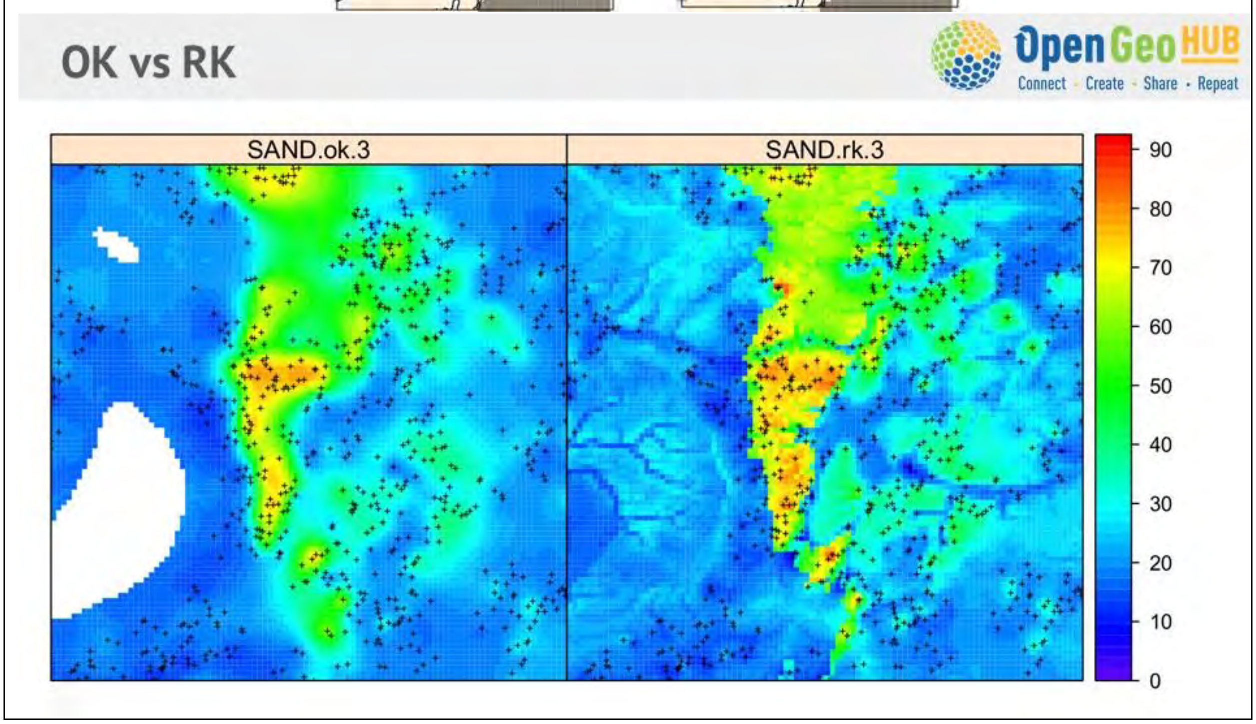
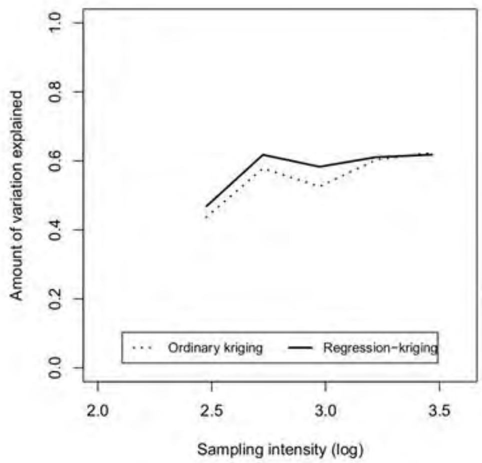


Figure 5: Simulated example of conversion of original predicted values (left) to rounded numbers/integers using effective precision (right). The removed pixels indicate extrapolation areas (mapping error greater than the total global variance). See further examples with real data sets in Figs. 6 and 7.


Hengl et al. / Special Issue SS2011 (2012) 1–16
Compression
OpenGeoHUB
Connect · Create · Share · Repeat
M = 40000 gzip = 1680B
M = 10000 gzip = 1176B
M = 2500 gzip = 848B
M = 625 gzip = 700B


OK vs RK
OpenGeoHUB
Connect · Create · Share · Repeat
SAND.ok.3
SAND.rk.3
90
80
70
60
50
40
30
20
10
0

Predictogram



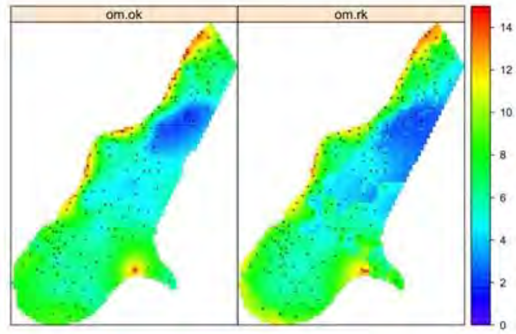
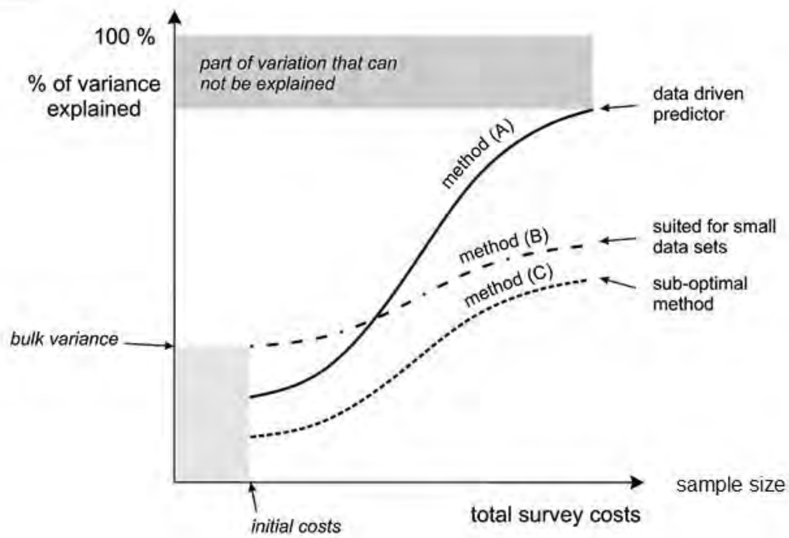


Figure 6: Comparison of predictions for organic matter for the Meuse data set. Estimated mapping efficiency for the map on the left 14.0 EUR km^{-2} %$^{-1}$ and 13.1 EUR km^{-2} %$^{-1}$ for the map on the right.

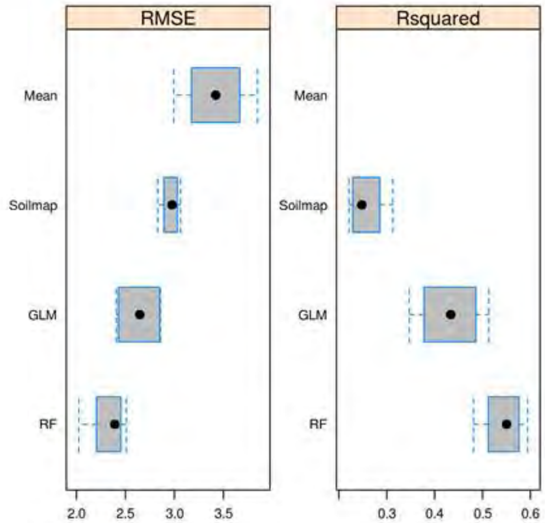
Table 1: File sizes, information production efficiency for OK and RK and their ratio.

Compression algorithm	OK file size (*Bytes*)	RK file size (*Bytes*)	OK IPE (EUR B^{-1})	RK IPE (EUR B^{-1})	IPE ratio –
GZIP	907	1152	3.99	3.14	1.27
BZIP2	681	889	5.32	4.08	1.31
XZ	808	976	4.48	3.71	1.21
RAR	850	1055	4.26	3.43	1.24

Predictogram general scheme



Predictogram: performance comparison



Back then: 4 young DSMers

A conventional soil mapper (CSMer) knows how to draw polygon maps.

The difference between a CSMer and a digital soil mapper (DSMer) is that a DSMer knows how to use GIS.

The difference between a DSMer and PSMer is that **a PSMer knows how to code** (R, Python, Julia etc)!

SM grades

Open Geo HUB — Connect · Create · Share · Repeat

We recognize four classes of advanced soil mapping methods (B, C, D and E in Fig. 1.7) which all belong to a continuum of *digital soil mapping* methods (Malone, Minasny, and McBratney 2016; McBratney, Minasny, and Stockmann 2018). We promote in this book specifically the Class E soil mapping approach i.e. which we refer to as the *predictive* and/or *automated soil mapping*.

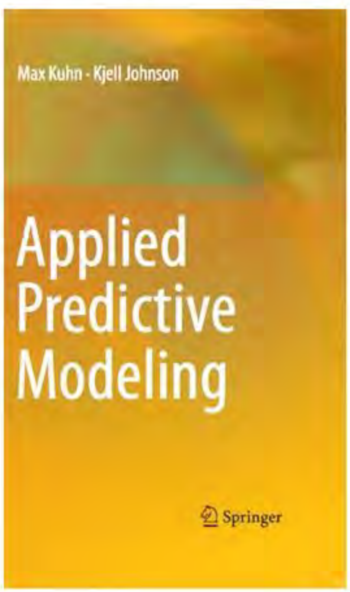
		Predictions produced using optimal statistical models	*Objective assessment of (spatial) uncertainty*	*Predictions can be fully automated*	*Predictions based primarily on RS data*
Automated soil mapping ←	*Class E*	✓	✓	✓	✓
	Class D	×	✓	✓	✓
	Class C	×	×	✓	✓
	Class B	×	×	×	✓
Conventional soil mapping ←	*Class A*	×	×	×	×

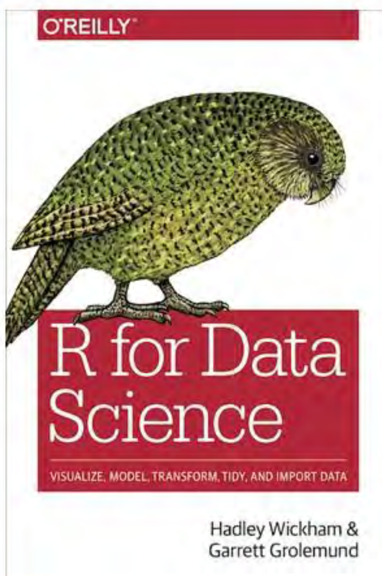
Figure 1.7: A classification of approaches to soil mapping: from purely expert driven (Class A), to various types of digital soil mapping including fully automated soil mapping (Class E).

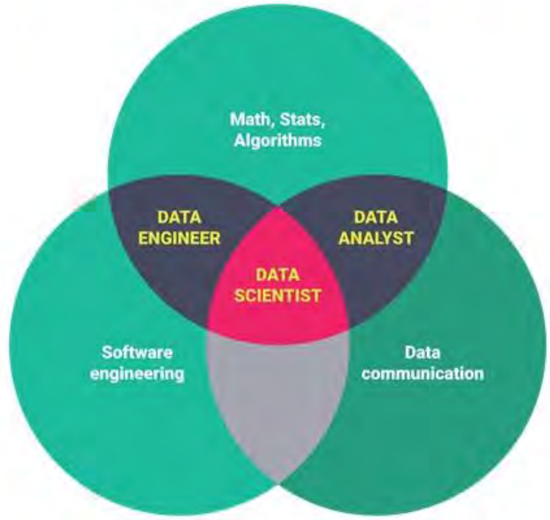
Some key advantages of the pedometric (statistical) approach to soil mapping are that it is objective, systematic, repeatable, updatable and represents an optimal expression of statistically validated understanding of soil-environmental relationships in terms of the currently available data.

Applied Predictive Modeling

Open Geo HUB — Connect · Create · Share · Repeat







To summarize

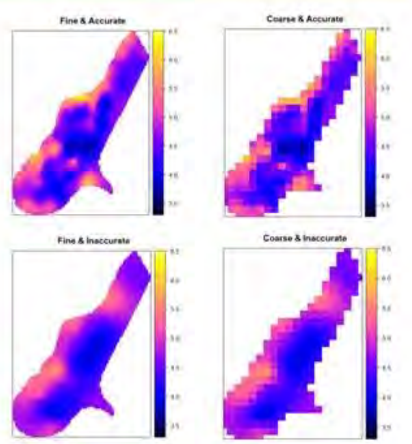
- The objective of PSM is to produce optimal unbiased predictions of a mean value at some new location along with the uncertainty associated with the prediction, at the finest possible resolution (or to explain and describe complexity of soils in a dynamic landscape).
- The main application objective of soil mapping is to accurately predict response of a soil(-plant) ecosystem to various soil management strategies.
- The difference between a PSMer and DSMer is that a PSMer knows GIS but also knows how to code (**PSMer = Data scientist working on predictive soil mapping**).

Technical specifications of a PSM project

- Targeted soil variables (soil properties and classes) and their complexity,
- **Spatial resolution**,
- Standard depths / depth intervals,
- Revisit periods (monitoring),
- Required accuracy levels (accuracy standards),
- Distribution channels (web-GIS? data services?),
- Commercialization options?

It's the accuracy stupid!

It's the accuracy, stupid



Which of these four maps do you prefer? Clearly, this will be the high-resolution, high-accuracy map. I can also imagine that in some cases the high-resolution, low-accuracy map will be considered second-best, although most users should benefit more from the low-resolution high-accuracy map. Coarse resolution and low accuracy is the least attractive.

It's the accuracy, stupid

by Gerard Heuvelink

More than twelve years ago I was asked to advise in a debate between a Dutch research institute and one of their clients. The institute had prepared groundwater table maps but the client was not satisfied. The maps had been made using regression kriging and were very different from maps derived using conventional survey. This was one of the client's concerns, but the client also pointed to various places on the map where the spatial patterns and predictions were very wrong. The institute argued that given the available budget, one could not expect more accurate maps. They also mentioned that the client could have known about this because the uncertainty of the spatial predictions had been quantified by means of separate standard deviation maps. The client then took a look at the uncertainty maps and noticed that the prediction intervals were indeed very wide. The client had expected much more accurate maps, and one of the reasons why they expected high thematic accuracy was that in the contract the institute had agreed to deliver the maps at scale 1:10,000. With this, the institute merely referred to the cartographic scale at which printed maps would be completed, while the client interpreted it as a sign of high accuracy. They were used to maps at scale 1:10,000 being much more accurate than maps at scale 1:25,000 or 1:50,000.

In conventional soil mapping, map scale is indeed in effect a measure of map accuracy, because mapping at fine scales involves many more man hours and borings per unit area than mapping at coarse scales. As long as everybody knows and agrees about this it is okay, although it would obviously be better to simply quantify the accuracy of a map directly (for instance, it could have avoided the misunderstanding between the research institute and client mentioned above). It may be implicit that a conventional soil map at scale 1:10,000 has higher (thematic) accuracy than a 1:50,000 scale map, but exactly how accurate and how much more accurate is rarely quantified. So even in conventional soil mapping it is not ideal to interpret map scale as a synonym for map accuracy, but in digital soil mapping it is even more so.

We all know that cartographic scale is no longer relevant in the digital world because we can freely zoom in and out on a digital screen. Indeed, it would be pure coincidence if 1 cm on the screen agrees with 100 m in the real world when a 1:10,000 soil map is displayed on the screen. In digital soil mapping, we therefore rarely refer to the 'scale' of a map (although some hardliners still do!) and use the term 'resolution' instead. That is all very fine, as long as we do not make the same mistake again by confusing resolution with accuracy. Resolution and accuracy are very different things. For instance, a high-resolution soil property map may well have very poor accuracy because of high short-distance spatial variation combined with low sampling density. Also, a low-resolution map may be very accurate because the mapped soil property has little to no short-distance spatial variation, while the sampling density is high and/or there is strong correlation with available covariates.

Success of a PSM project

It is NOT only about thematic accuracy!
Success of soil maps is typically a product of multiple things:

= relevance for decision making / spatial planning × level of detail (spatial accuracy) × thematic (attribute) accuracy × completeness × consistency × accessibility × price

= RELEVANCE × LA × TA × COMP × CONS × ACCESS × PRICE

http://www.pedometrics.org/Pedometron/Pedometron38.pdf

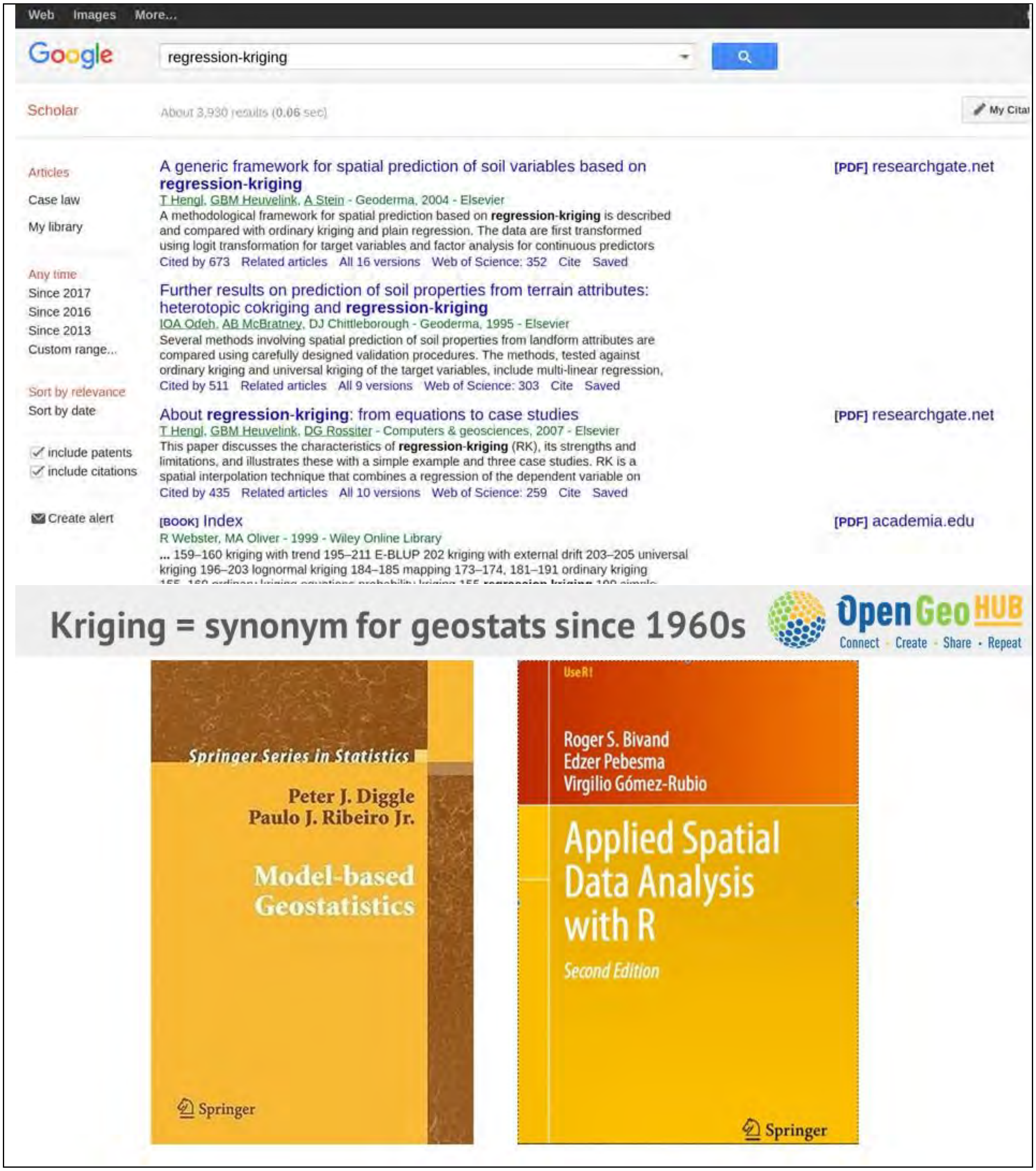
Regression-kriging

5.2.5 Regression-kriging (generic model)

Ignoring the assumptions about the cross-correlation between the trend and residual components, we can extend the regression-kriging model and use any type of (non-linear) regression to predict values (e.g. regression trees, artificial neural networks and other machine learning models), calculate residuals at observation locations, fit a variogram for these residuals, interpolate the residuals using ordinary or simple kriging, and add the result to the predicted regression part. This means that RK can, in general, be formulated as:

$$\text{prediction} = \begin{array}{c}\text{trend predicted}\\ \text{using regression}\end{array} + \begin{array}{c}\text{residual predicted}\\ \text{using kriging}\end{array} \quad (5.15)$$

Again, statistical inference and prediction is relatively simple if the stochastic residual, or a transformation thereof, may be assumed normally distributed. Error of the regression-kriging model is likewise a sum of the regression and the kriging model errors.


Web Images More...
Google
regression-kriging
Scholar
About 3,930 results (0.06 sec)
My Citat
Articles
Case law
My library
Any time
Since 2017
Since 2016
Since 2013
Custom range...
Sort by relevance
Sort by date
include patents
include citations
Create alert
A generic framework for spatial prediction of soil variables based on regression-kriging
T Hengl, GBM Heuvelink, A Stein - Geoderma, 2004 - Elsevier
A methodological framework for spatial prediction based on regression-kriging is described and compared with ordinary kriging and plain regression. The data are first transformed using logit transformation for target variables and factor analysis for continuous predictors
Cited by 673 Related articles All 16 versions Web of Science: 352 Cite Saved
[PDF] researchgate.net
Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging
IOA Odeh, AB McBratney, DJ Chittleborough - Geoderma, 1995 - Elsevier
Several methods involving spatial prediction of soil properties from landform attributes are compared using carefully designed validation procedures. The methods, tested against ordinary kriging and universal kriging of the target variables, include multi-linear regression,
Cited by 511 Related articles All 9 versions Web of Science: 303 Cite Saved
About regression-kriging: from equations to case studies
T Hengl, GBM Heuvelink, DG Rossiter - Computers & geosciences, 2007 - Elsevier
This paper discusses the characteristics of regression-kriging (RK), its strengths and limitations, and illustrates these with a simple example and three case studies. RK is a spatial interpolation technique that combines a regression of the dependent variable on
Cited by 435 Related articles All 10 versions Web of Science: 259 Cite Saved
[PDF] researchgate.net
[BOOK] Index
R Webster, MA Oliver - 1999 - Wiley Online Library
... 159–160 kriging with trend 195–211 E-BLUP 202 kriging with external drift 203–205 universal kriging 196–203 lognormal kriging 184–185 mapping 173–174, 181–191 ordinary kriging
[PDF] academia.edu
Kriging = synonym for geostats since 1960s
OpenGeoHUB
Connect · Create · Share · Repeat
Springer Series in Statistics
Peter J. Diggle
Paulo J. Ribeiro Jr.
Model-based Geostatistics
Springer
UseR!
Roger S. Bivand
Edzer Pebesma
Virgilio Gómez-Rubio
Applied Spatial Data Analysis with R
Second Edition
Springer

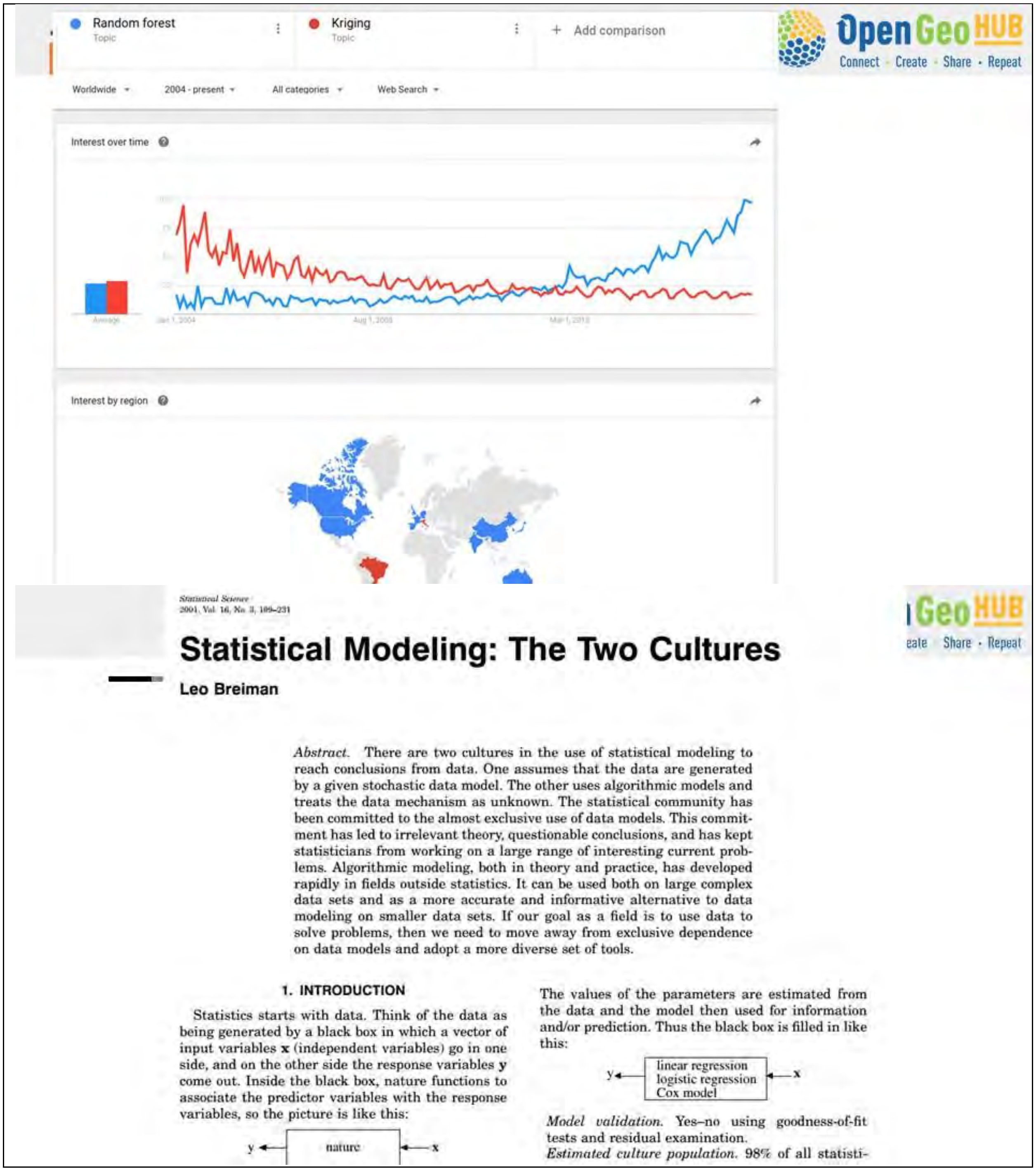
Kriging: standard steps

1. Determine distribution of the target variable and appropriate transformation (normal, log-normal, zero-inflated, Gamma, Poissonic ...)
2. Fit variogram (WLS, REML, ...), deal with multicolinearity (PCA?), non-stationary properties, support size, mixed effects...
3. Predict (mean values and uncertainty)
4. Validate predictions (mapping accuracy)

Variogram modeling and kriging

```
R> zinc.vgm <- likfit(zinc.geo, lambda = 0,
ini=c(var(log1p(zinc.geo$data)), 500), cov.model
= "exponential")

R> zinc.ok <- krige.conv(zinc.geo, locations =
locs, krige = krige.control(obj.m = zinc.vgm))
```

krige.conv: model with constant mean
krige.conv: performing the Box-Cox data transformation
krige.conv: back-transforming the predicted mean and variance
krige.conv: Kriging performed using global neighbourhood



Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

Abstract. There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

1. INTRODUCTION

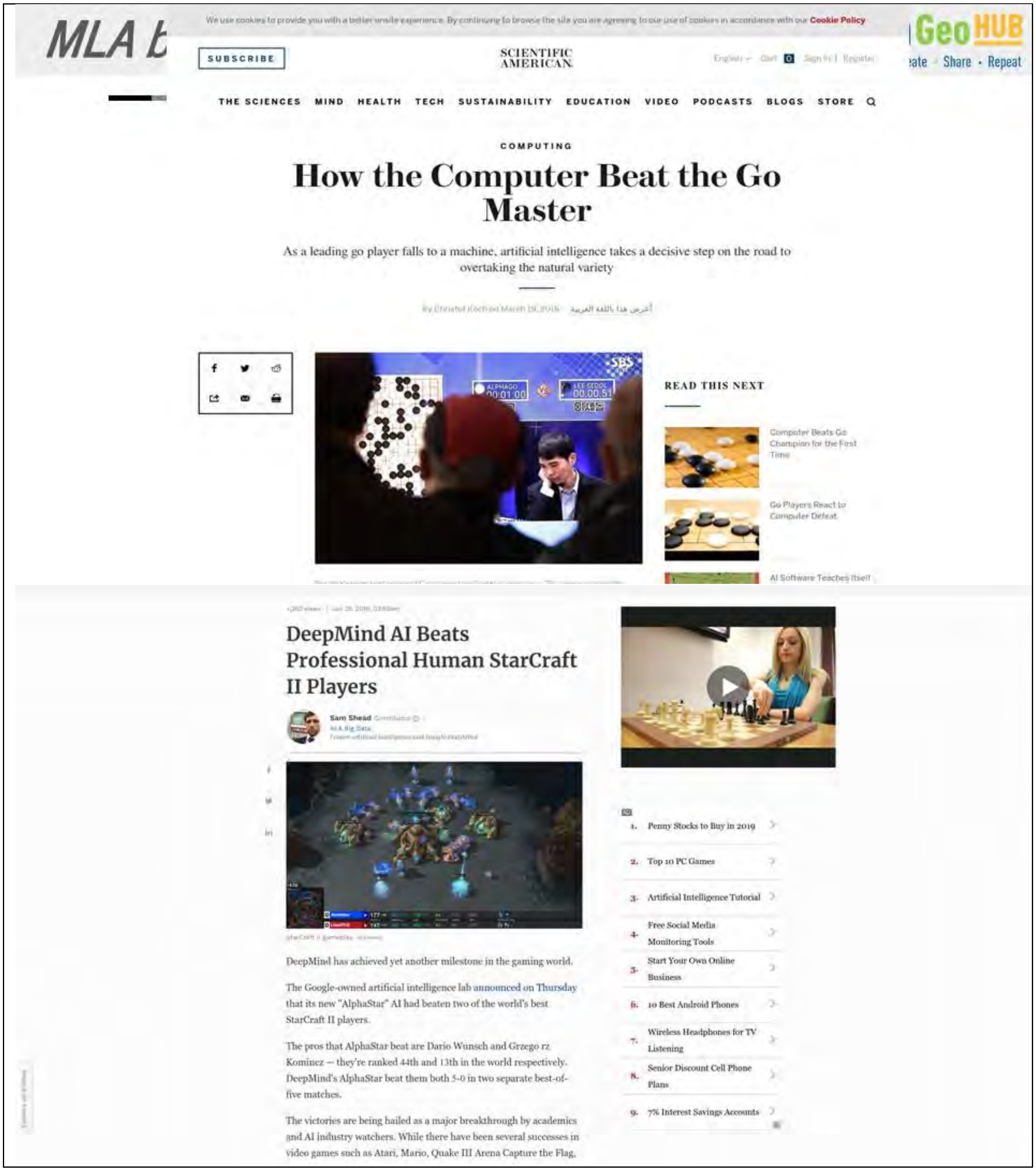
Statistics starts with data. Think of the data as being generated by a black box in which a vector of input variables **x** (independent variables) go in one side, and on the other side the response variables **y** come out. Inside the black box, nature functions to associate the predictor variables with the response variables, so the picture is like this:

y ← nature ← x

The values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:

y ← linear regression / logistic regression / Cox model ← x

Model validation. Yes–no using goodness-of-fit tests and residual examination.
Estimated culture population. 98% of all statisti-


MLA
GeoHUB
ate · Share · Repeat
We use cookies to provide you with a better onsite experience. By continuing to browse the site you are agreeing to our use of cookies in accordance with our Cookie Policy
SUBSCRIBE
SCIENTIFIC AMERICAN
THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE
COMPUTING
How the Computer Beat the Go Master
As a leading go player falls to a machine, artificial intelligence takes a decisive step on the road to overtaking the natural variety
By Christof Koch on March 19, 2016
أعرض هذا باللغة العربية
ALPHAGO
LEE SEDOL
SBS
READ THIS NEXT
Computer Beats Go Champion for the First Time
Go Players React to Computer Defeat
AI Software Teaches Itself
DeepMind AI Beats Professional Human StarCraft II Players
Sam Shead
AI & Big Data
DeepMind has achieved yet another milestone in the gaming world.
The Google-owned artificial intelligence lab announced on Thursday that its new "AlphaStar" AI had beaten two of the world's best StarCraft II players.
The pros that AlphaStar beat are Dario Wunsch and Grzegorz Komincz — they're ranked 44th and 13th in the world respectively. DeepMind's AlphaStar beat them both 5-0 in two separate best-of-five matches.
The victories are being hailed as a major breakthrough by academics and AI industry watchers. While there have been several successes in video games such as Atari, Mario, Quake III Arena Capture the Flag,
1. Penny Stocks to Buy in 2019
2. Top 10 PC Games
3. Artificial Intelligence Tutorial
4. Free Social Media Monitoring Tools
5. Start Your Own Online Business
6. 10 Best Android Phones
7. Wireless Headphones for TV Listening
8. Senior Discount Cell Phone Plans
9. 7% Interest Savings Accounts

MIT Technology Review

Log in / Register Search

Topics+ Top Stories Magazine Business Reports More+

OpenGeoHUB
Connect · Create · Share · Repeat

Computing

The Extraordinary Link Between Deep Neural Networks and the Nature of the Universe

Nobody understands why deep neural networks are so good at solving complex problems. Now physicists say the secret is buried in the laws of physics.

by Emerging Technology from the arXiv September 9, 2016

In the last couple of years, deep learning techniques have transformed the world of artificial intelligence. One by one, the abilities and techniques that humans once imagined were uniquely our own have begun to fall to the onslaught of ever more powerful machines. Deep neural networks are now better than humans at tasks such as face recognition and object recognition. They've mastered the ancient game of Go and thrashed the best human players.

But there is a problem. There is no mathematical reason why networks arranged in layers should be so good at these challenges. Mathematicians are flummoxed. Despite the huge success of deep

Reichstein et al.

OpenGeoHUB
Connect · Create · Share · Repeat

PERSPECTIVE

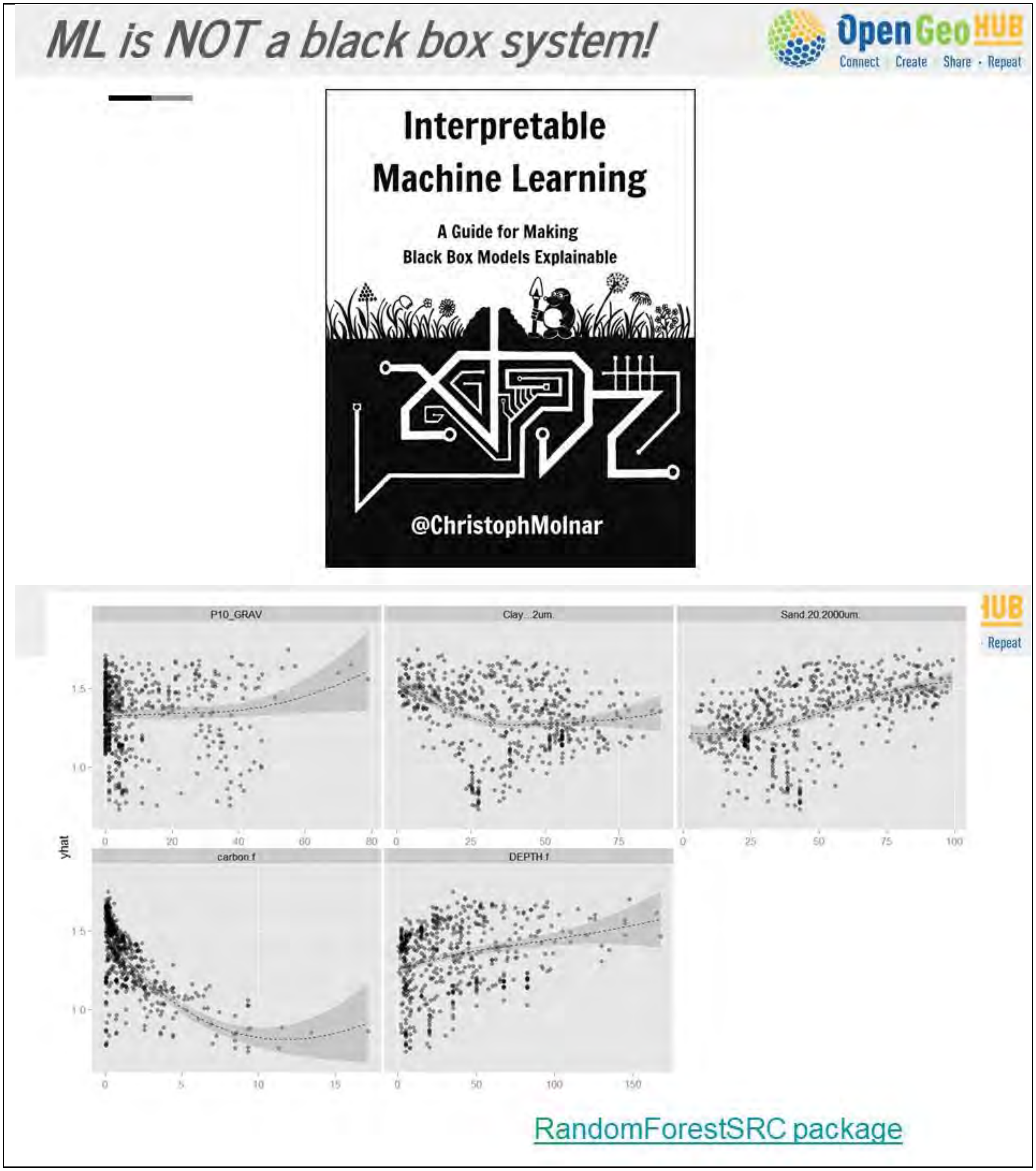
https://doi.org/10.1038/s41586-019-0912-1

Deep learning and process understanding for data-driven Earth system science

Markus Reichstein[1,2*], Gustau Camps-Valls[3], Bjorn Stevens[4], Martin Jung[1], Joachim Denzler[2,5], Nuno Carvalhais[1,6] & Prabhat[7]

Machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data, but current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Here, rather than amending classical machine learning, we argue that these contextual cues should be used as part of deep learning (an approach that is able to extract spatio-temporal features automatically) to gain further process understanding of Earth system science problems, improving the predictive ability of seasonal forecasting and modelling of long-range spatial connections across multiple timescales, for example. The next step will be a hybrid modelling approach, coupling physical process models with the versatility of data-driven machine learning.

https://www.nature.com/articles/s41586-019-0912-1


ML is NOT a black box system!
OpenGeoHUB
Connect · Create · Share · Repeat
Interpretable
Machine Learning
A Guide for Making
Black Box Models Explainable
@ChristophMolnar
P10_GRAV
Clay..2um
Sand.20.2000um
carbon.f
DEPTH.f
yhat
RandomForestSRC package

MLA is interesting for PSM, however...

Spatial auto-correlation, especially if still existent in the cross-validation residuals, indicates that the predictions are maybe biased, and this is suboptimal.

To account for this, we use Random Forest (as implemented in the ranger package) in combination with geographical distances to sampling locations to fit models and predict values.

Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables

Tomislav Hengl, Madlene Nussbaum, Marvin N. Wright, Gerard B.M. Heuvelink, Benedikt Gräler

Published August 29, 2018

PubMed 30186691

Author and article information

Abstract

Random forest and similar Machine Learning techniques are already used to generate spatial predictions, but spatial location of points (geography) is often ignored in the modeling process. Spatial auto-correlation, especially if still existent in the cross-validation residuals, indicates that the predictions are maybe biased, and this is suboptimal. This paper presents a random forest for spatial predictions framework (RFsp) where buffer distances from observation points are used as explanatory variables, thus incorporating geographical proximity effects into the prediction process. The RFsp framework is illustrated with examples that use textbook datasets and apply spatial and spatio-temporal prediction to numeric, binary, categorical, multivariate and spatiotemporal variables. Performance of the RFsp framework is compared with the state-of-the-art kriging techniques using fivefold cross-validation with refitting. The results show that RFsp can obtain equally accurate and unbiased predictions as different versions of kriging.

5 Total citations
5 Recent citations
n/a Field Citation Ratio
n/a Relative Citation Ratio

I published in PeerJ and it is very fast, has good editors, has consistently given good quality and rigorous reviews of my work, and produces visually appealing manuscripts.

Matthew Jackson

Download

Content Alert

Tools & info

Outline

UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR MEDIZINISCHE BIOMETRIE UND STATISTIK
IMBS
HUB
Repeat
Home
Main Menu
Home
Staff
How to find us
Weblinks
Research
Awards
Projects
Publications
Software
Employment Ad
Courses
Biometrical Consulting Service
Lectures
Leselunch
Seminar
Student Research Project
Veranstaltungen
Search
Search this site:
Search
Marvin N. Wright, M.Sc.
Institut für Medizinische Biometrie und Statistik
Universität zu Lübeck
Universitätsklinikum Schleswig-Holstein, Campus Lübeck
Ratzeburger Allee 160
23562 Lübeck
fon: 0451-500 2782
fax: 0451-500 2999
wright@imbs.uni-luebeck.de
Publications
Wright, M. N., Ziegler, A. & König, I. R. (2016). Do little interactions get lost in dark random forests? BMC Bioinformatics 17:145. http://dx.doi.org/10.1186/s12859-016-0995-8.
Wright, M. N. & Ziegler, A. (2016). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, in press. http://arxiv.org/abs/1508.04409.
Schirmer, J. H., Wright, M. N., Vonthein, R., Herrmann, K., Nolle. B., Both, M., Henes, F., Arlt, A., Gross, W. L., Schinke, S., Reinhold-Keller, E., Moosig, F. & Holle, J. U. (2016). Clinical presentation and long-term outcome of 144 patients with microscopic polyangiitis in a monocentric German cohort. Rheumatology (Oxford) 55:71-79. http://dx.doi.org/10.1093/rheumatology/kev286.
Wright, M. N. & Ziegler, A. (2015). Multiple censored data in dentistry: A new statistical model for analyzing lesion size in randomized controlled trials. Biometrical Journal 57:384-394. http://dx.doi.org/10.1002/bimj.201400118.
Paulick, C., Wright, M. N., Verleger, R. & Keller, K. (2014). Decomposition of 3-way arrays: A comparison of different PARAFAC algorithms. Chemometrics and Intelligent Laboratory Systems 137:97-109. http://dx.doi.org/10.1016/j.chemolab.2014.06.009.
Technical Reports
Schmid, M., Wright, M. N. & Ziegler, A. (2015). On the Use of Harrell's C for Node Splitting in Random Survival Forests. Technical Report. http://arxiv.org/abs/1507.03092.
README.md
RFsp — Random Forest for spatial data (R tutorial)
Hengl, T., Nussbaum, M., and Wright, M.N.
Installing and loading packages
Spatial prediction 2D continuous variable using buffer distances
Spatial prediction 2D variable with covariates
Spatial prediction of binomial variable
Spatial prediction of categorical variable
Spatial prediction of variables with extreme values
Weighted RFsp
Spatial prediction of multivariate problems
Prediction of spatio-temporal variable
References
https://github.com/thengl/GeoMLA
Abstract: This tutorial explains how to use Random Forest to generate spatial and spatiotemporal predictions (i.e. to make maps from point observations using Random Forest). Spatial auto-correlation, especially if still existent in the cross-validation residuals, indicates that the predictions are maybe biased, and this is suboptimal. To account for this, we use Random Forest (as implemented in the ranger package) in combination with geographical distances to sampling locations to fit models and

Solution: RFsp

$$Y(\mathbf{s}) = f(\mathbf{X}_G, \mathbf{X}_R, \mathbf{X}_P) \tag{18}$$

where $\mathbf{X}_G$ are covariates accounting for geographical proximity and spatial relations between observations

$$\mathbf{X}_G = (d_{p1}, d_{p2}, \ldots, d_{pN}) \tag{19}$$

where d_{pi} is the buffer distance (or any other complex proximity upslope/downslope distance, as explained in the next section) to the observed location pi from $\mathbf{s}$ and N is the total number of training points. $\mathbf{X}_R$ are surface reflectance covariates, i.e. usually spectral bands of remote sensing images, and $\mathbf{X}_P$ are process-based covariates.

Geographical distances (proximity)

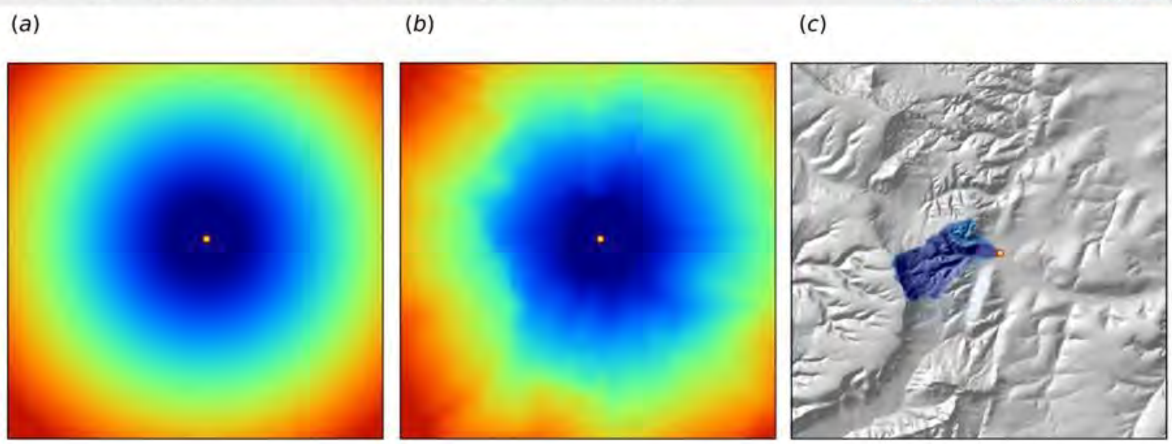
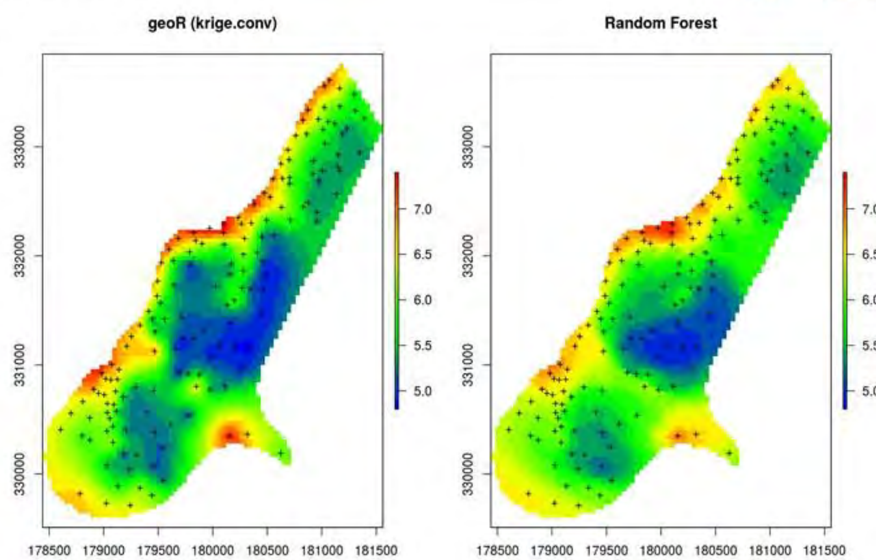


Figure 2. Examples of distance maps to some location in space (yellow dot) based on different derivation algorithms: (a) simple Euclidean distances, (b) complex speed-based distances based on the gdistance package and Digital Elevation Model (DEM) (van Etten, 2017), and (c) upslope area derived based on the DEM in SAGA GIS (Conrad et al., 2015). Case study: Ebergötzen (Böhner et al., 2006).

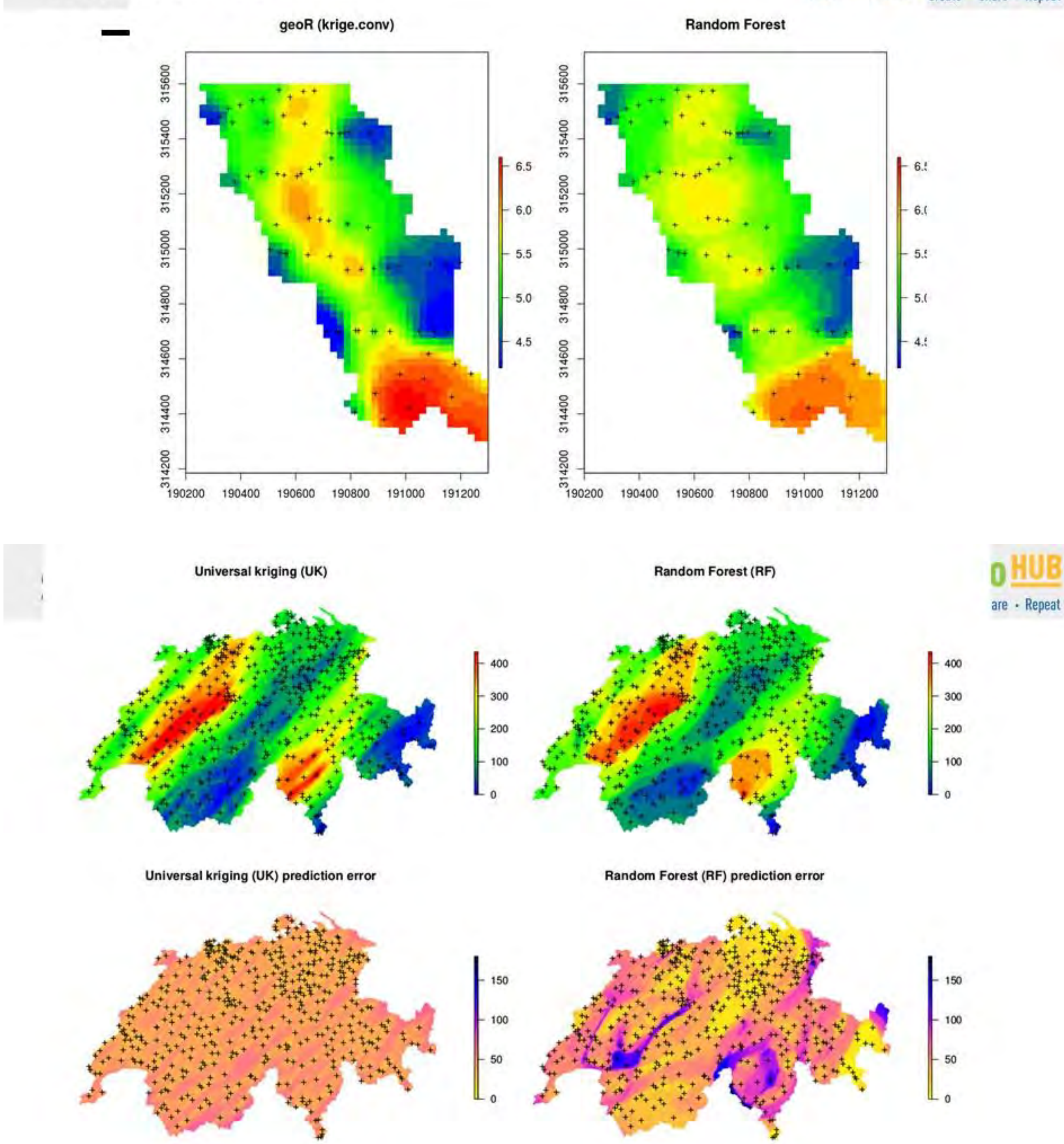
Machine Learning for interpolation

```
R> grid.dist0 <- buffer.dist(meuse["zinc"],
meuse.grid[1], as.factor(1:nrow(meuse)))

R> ov.zinc <- over(meuse["zinc"], grid.dist0)

R> m.zinc <- ranger(as.formula(paste("zinc ~",
paste(names(grid.dist0), collapse="+")),
cbind(meuse@data["zinc"], ov.zinc))

R> zinc.rfd <- predict(m.zinc, grid.dist0@data)
```

Meuse data set



Geul data set



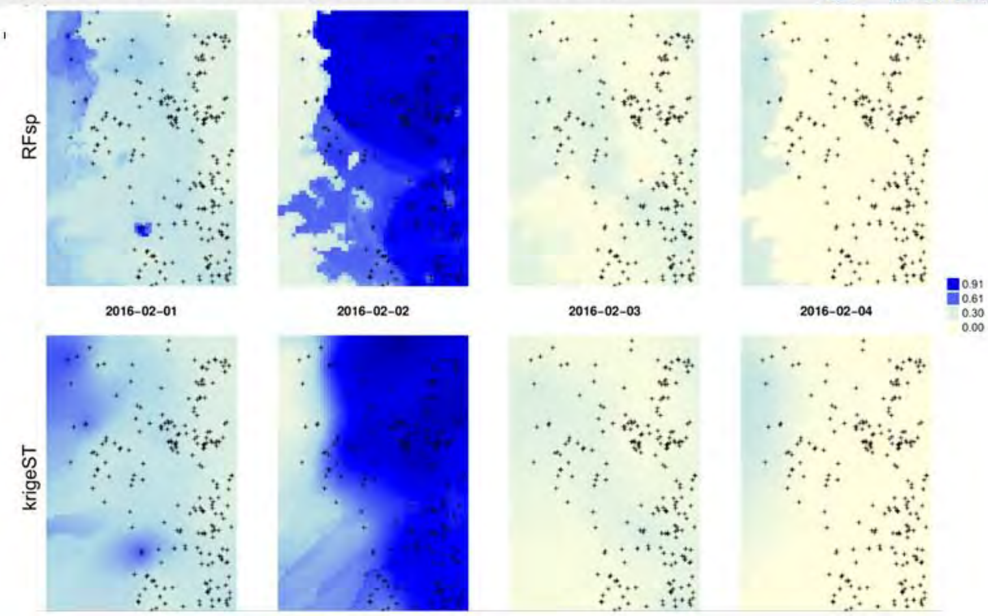
Spatiotemporal models

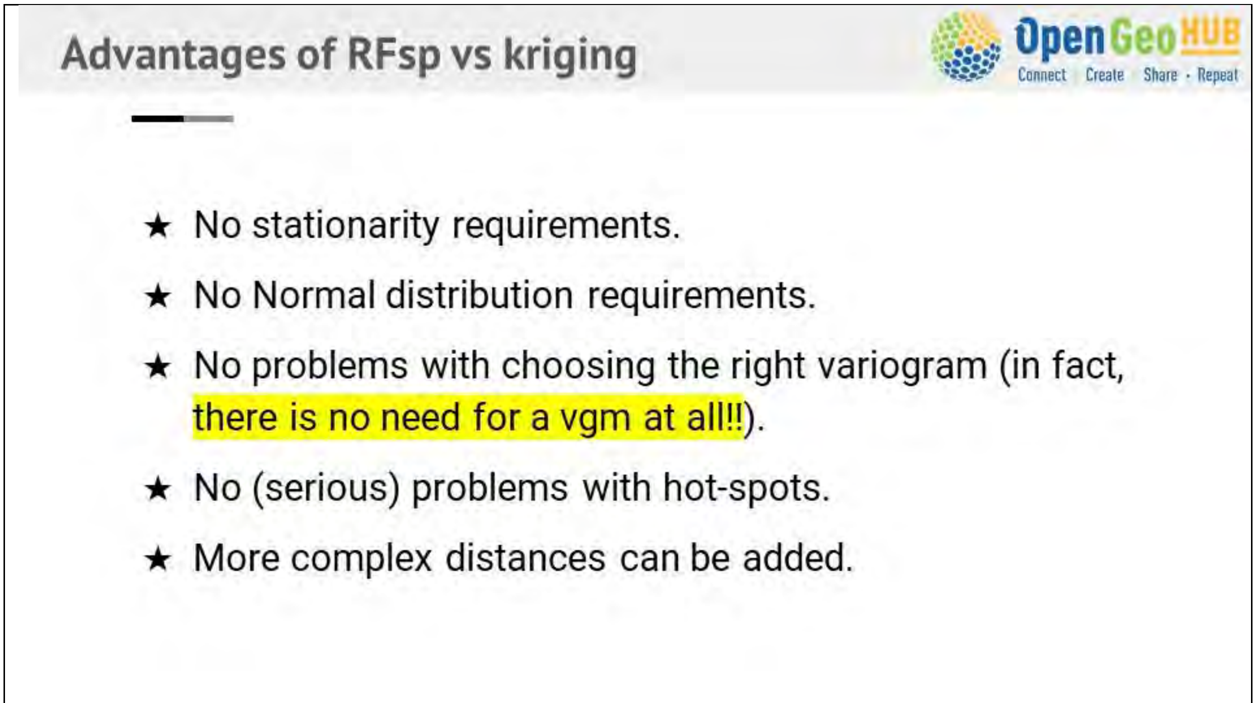
OpenGeoHUB Connect · Create · Share · Repeat

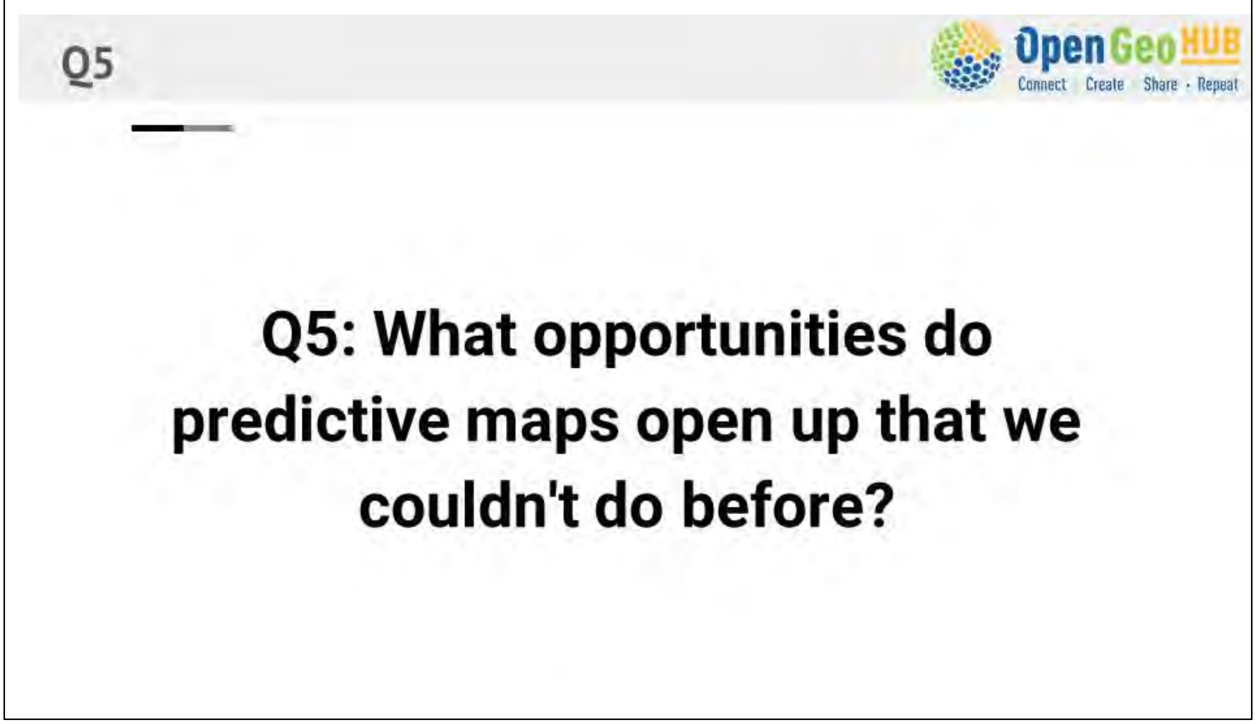
```
R> fmP <- as.formula(paste("PRCP ~ cdate + doy +
elev_1km + PRISM_prec +", dnP))
R> fmP

PRCP ~ elev_1km + PRISM_prec + cdate + doy +
layer.1 + … + layer.N

R> m1.prec <- ranger(formula = fmP, data =
rm.prec, mtry=212)
```

Daily precipitation (spatiotemporal)

OpenGeoHUB Connect · Create · Share · Repeat




Advantages of RFsp vs kriging
OpenGeoHUB
Connect · Create · Share · Repeat
★ No stationarity requirements.
★ No Normal distribution requirements.
★ No problems with choosing the right variogram (in fact, there is no need for a vgm at all!!).
★ No (serious) problems with hot-spots.
★ More complex distances can be added.


Q5
OpenGeoHUB
Connect · Create · Share · Repeat
Q5: What opportunities do predictive maps open up that we couldn't do before?

New opportunities

OpenGeoHUB Connect · Create · Share · Repeat

- ★ Spacetime modeling of soil dynamics,
- ★ Data processing automation,
- ★ Global, regional, local and multiscale models,
- ★ ML models on very large data sets,

All these leads to (a) significant increases in accuracy, (b) significant reduction of costs, (c) new areas of application (scenario testing, econometrics, new crediting systems...)

RFsp problems still to solve

OpenGeoHUB Connect · Create · Share · Repeat

1. Extrapolation problems (quality of spatial sampling)
2. Computation intensity **very high**
3. Over-fitting problems (spatial de-clustering)
4. Produce geostatistical simulations, co-kriging etc.

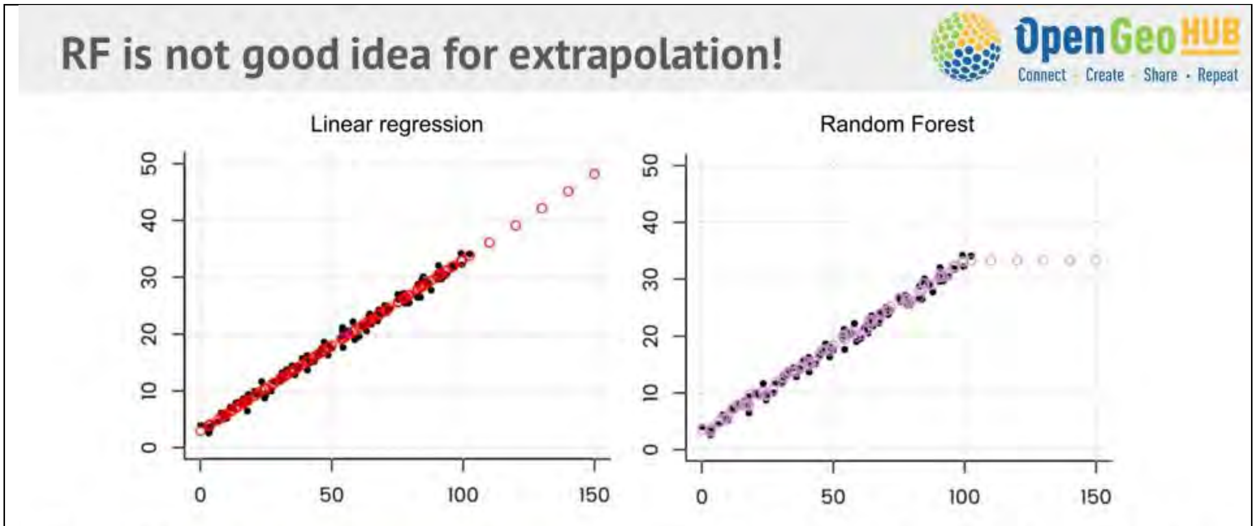
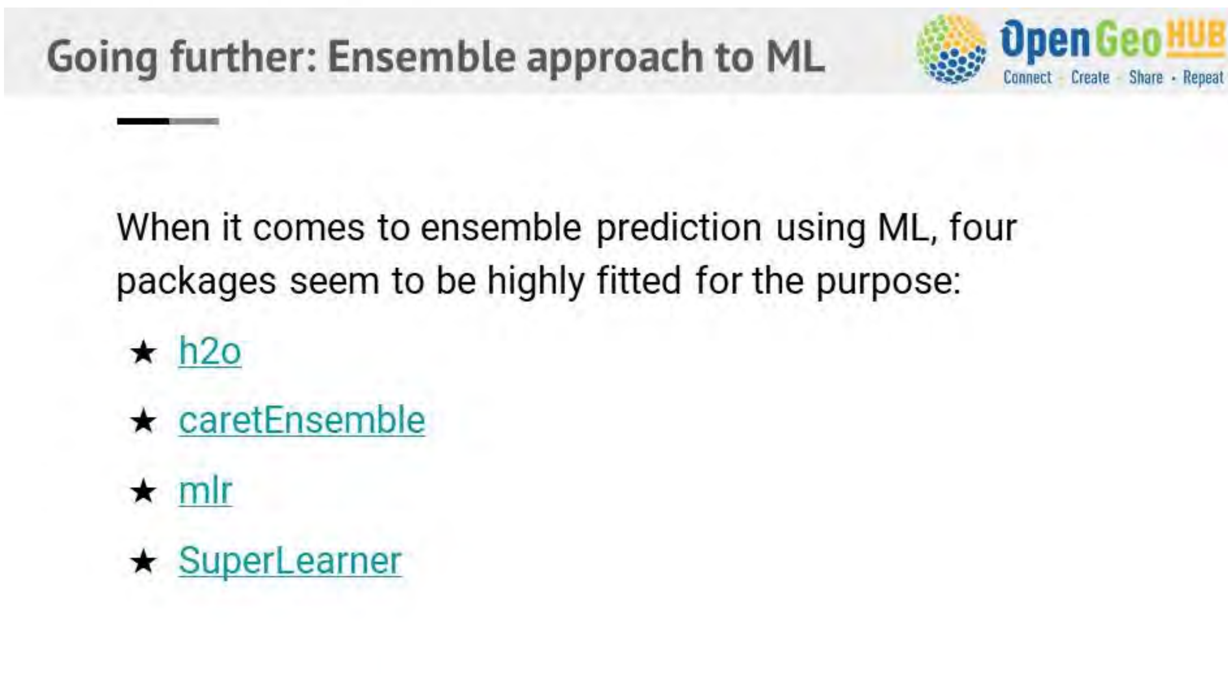
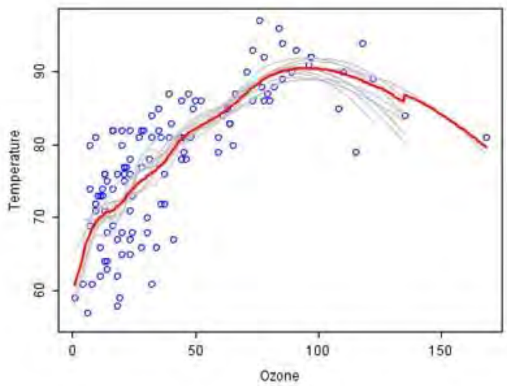
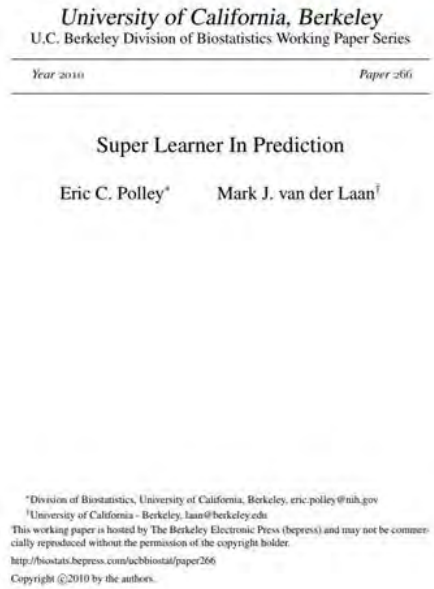


Figure 14. Illustration of the extrapolation problem of Random Forest based on the code examples from Peter Ellis (`http://freerangestats.info`). Even though Random Forest is more generic than linear regression and can be used also to fit complex non-linear problems, it can lead to completely nonsensical predictions if applied to extrapolation domains.



Ensemble methods





University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2010 *Paper 266*

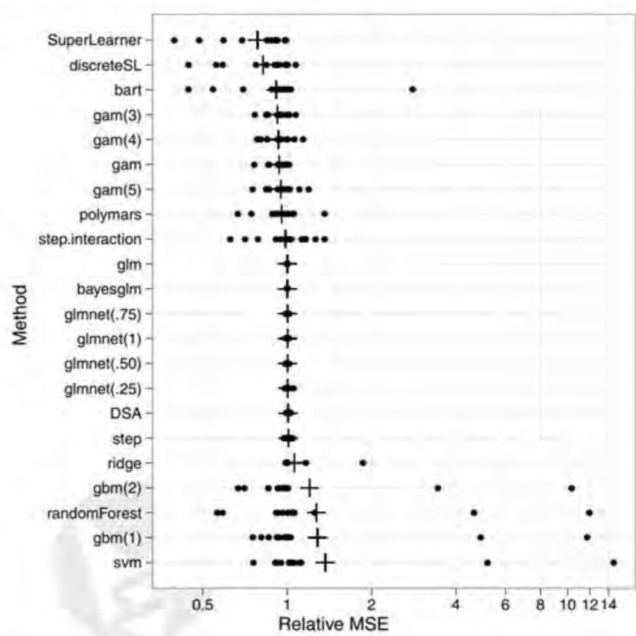
Super Learner In Prediction

Eric C. Polley* Mark J. van der Laan†

*Division of Biostatistics, University of California, Berkeley, eric.polley@nih.gov

†University of California - Berkeley, laan@berkeley.edu

Figure 3: 10-fold cross-validated relative mean squared error compared to glm across 13 real datasets. Sorted by the geometric mean, denoted with the plus (+) sign.



"Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to **decrease variance** (bagging), **bias** (boosting), or **improve predictions** (stacking)." https://blog.statsbot.co/ensemble-learning-d1dcd548e936

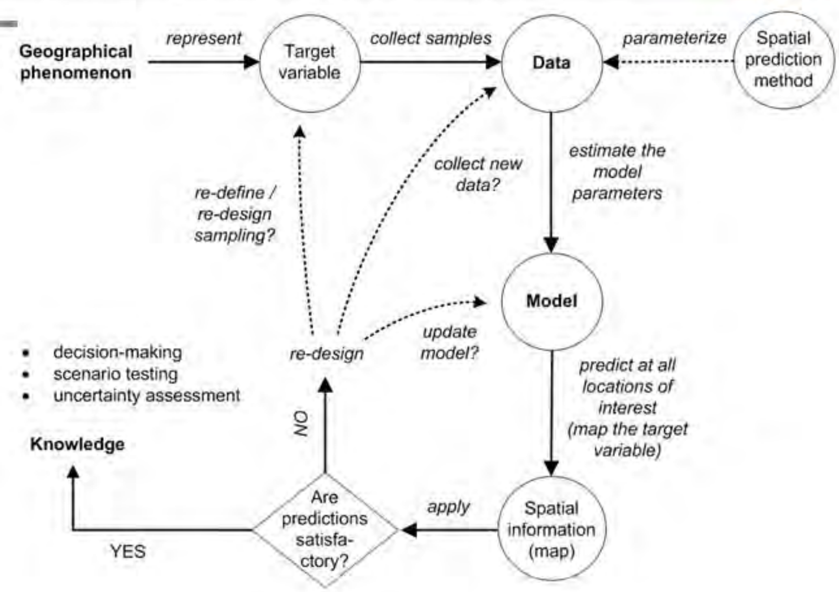
This however comes at costs:

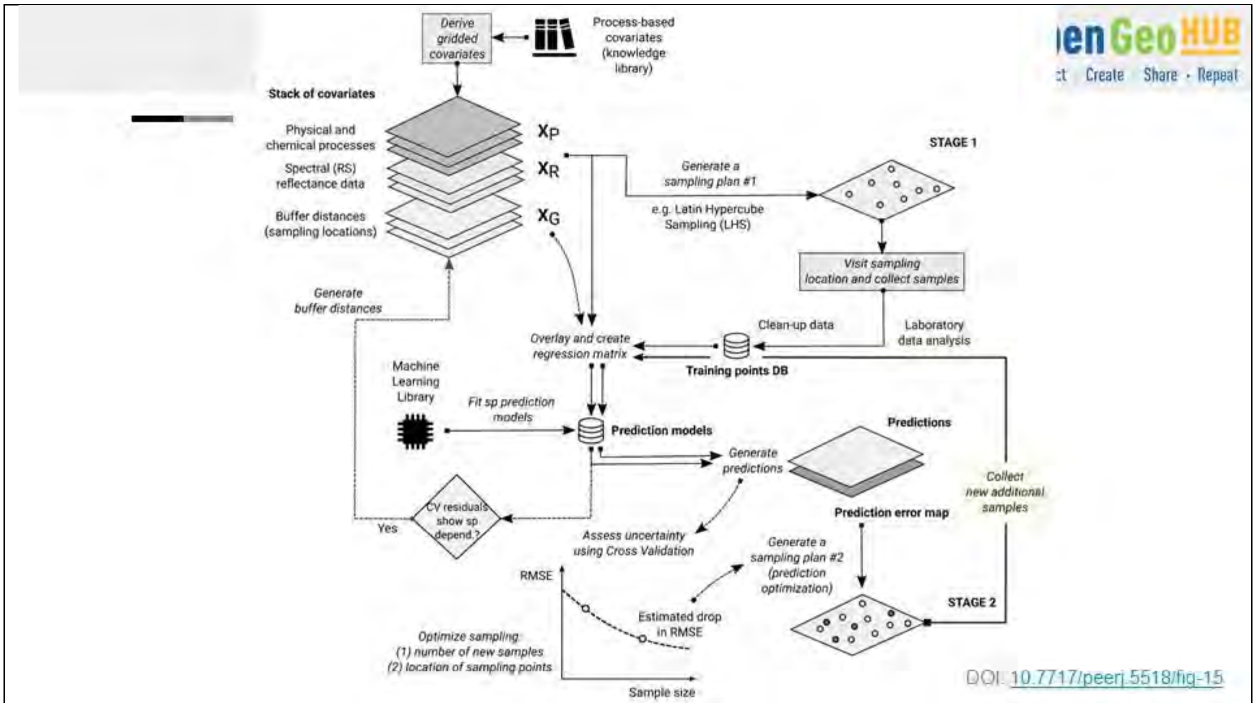
- higher computational load,
- higher RAM requirements,

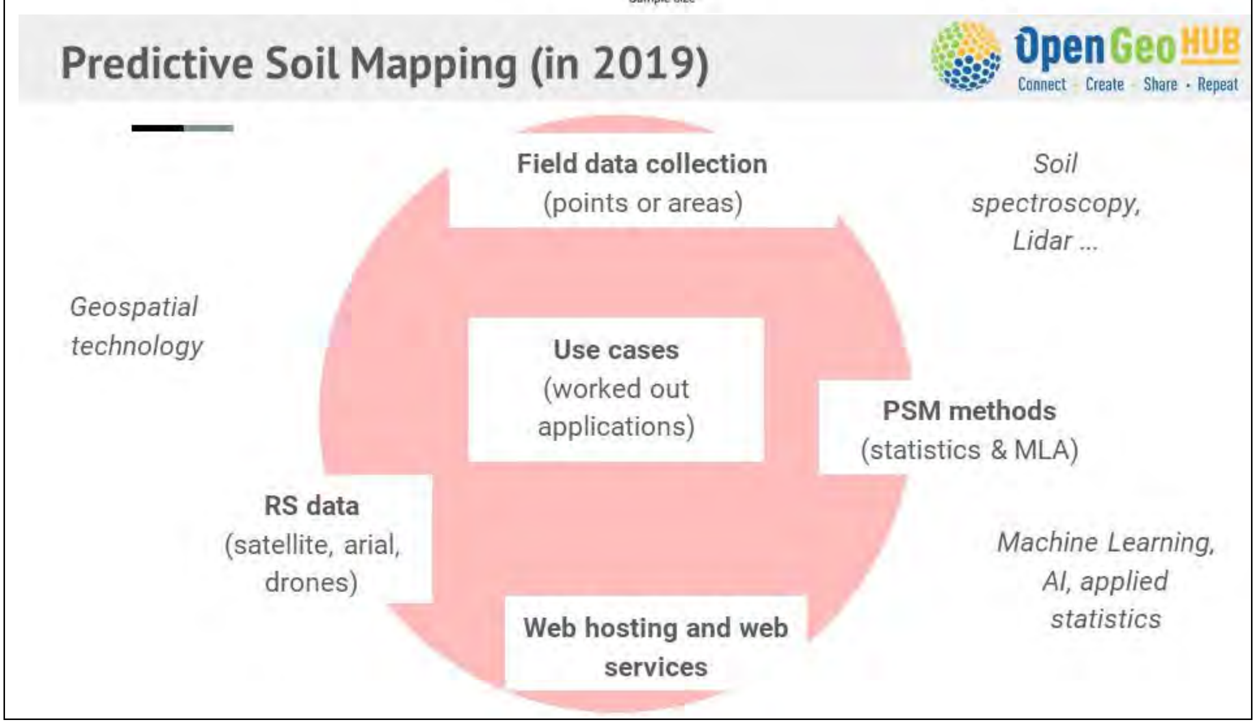
Q1

Q1: At a high-level, how does PSM work?

General spatial prediction scheme




Derive gridded covariates
Process-based covariates (knowledge library)
Stack of covariates
Physical and chemical processes
Spectral (RS) reflectance data
Buffer distances (sampling locations)
XP
XR
XG
STAGE 1
Generate a sampling plan #1
e.g. Latin Hypercube Sampling (LHS)
Visit sampling location and collect samples
Generate buffer distances
Clean-up data
Laboratory data analysis
Overlay and create regression matrix
Training points DB
Machine Learning Library
Fit sp prediction models
Prediction models
Predictions
Generate predictions
Collect new additional samples
Prediction error map
CV residuals show sp depend.?
Yes
Assess uncertainty using Cross Validation
Generate a sampling plan #2 (prediction optimization)
RMSE
STAGE 2
Estimated drop in RMSE
Optimize sampling: (1) number of new samples (2) location of sampling points
Sample size
DOI: 10.7717/peerj.5518/fig-15


Predictive Soil Mapping (in 2019)
OpenGeoHUB
Connect · Create · Share · Repeat
Field data collection (points or areas)
Soil spectroscopy, Lidar ...
Geospatial technology
Use cases (worked out applications)
PSM methods (statistics & MLA)
RS data (satellite, arial, drones)
Machine Learning, AI, applied statistics
Web hosting and web services

Conclusions I

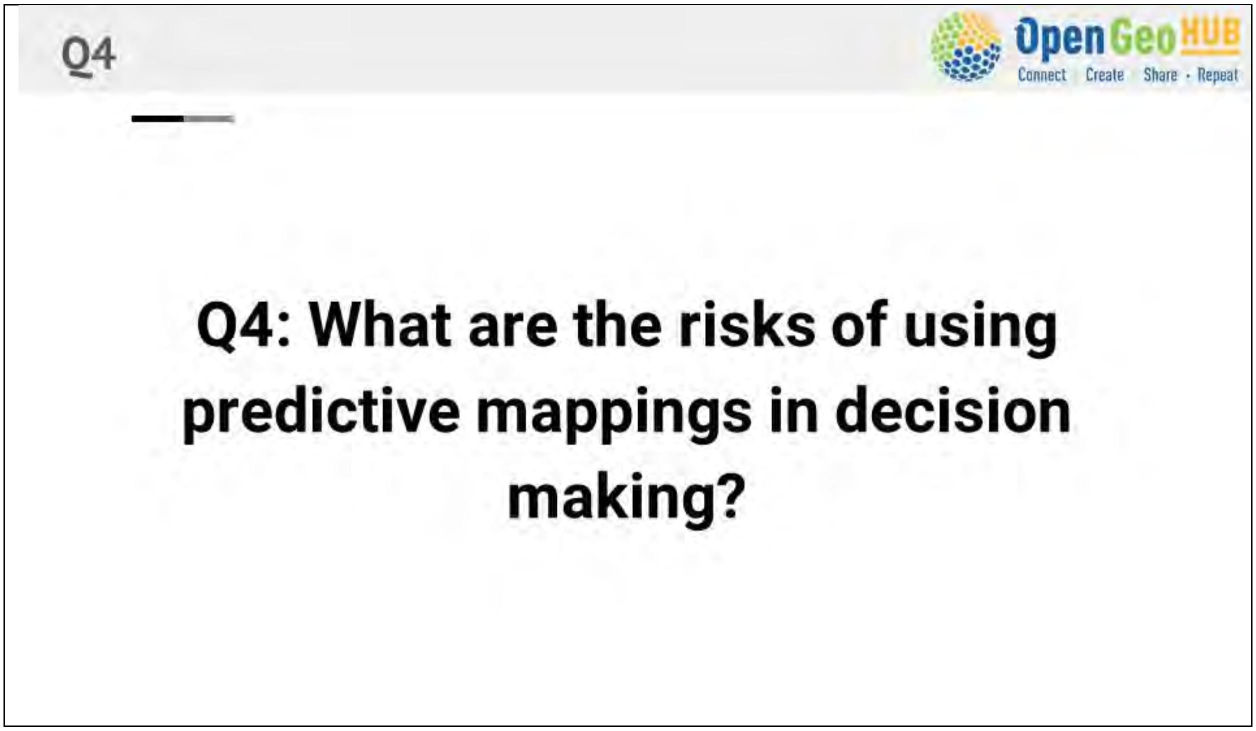
OpenGeoHUB Connect · Create · Share · Repeat

1. The principal difference between a DSMer and PSMer is in the advance coding capacities (**do you speak R, Python, Scala, Java and/or Julia?**).
2. Once you become a soil **Data Scientist** / **geocomputation expert**, you should also be able to make better predictions of soil properties and classes.
3. Modern PSM is largely based on using **RS / proximal soil sensing / LiDAR** and **state-of-the-art Machine and Statistical Learning software** (here focused on Open Source).

Conclusions II

OpenGeoHUB Connect · Create · Share · Repeat

4. One attractive framework to generate spatial predictions is the **Ensemble Machine Learning** (see e.g. SuperLearner, caret, mlr, h2o modeling frameworks).
5. By including geographical distance in the Machine Learning, we can be used to "elegantly" replace traditional geostatistics (kriging). **EMLsp** is the <u>kriging killer</u>!
6. It is <u>scary</u> how much we could automate generation of soil maps today, but that does **NOT** means that: (a) we should stop sampling, (b) we do not need to learn/understand stats or soil science any more, (c) that we need to stop coding.

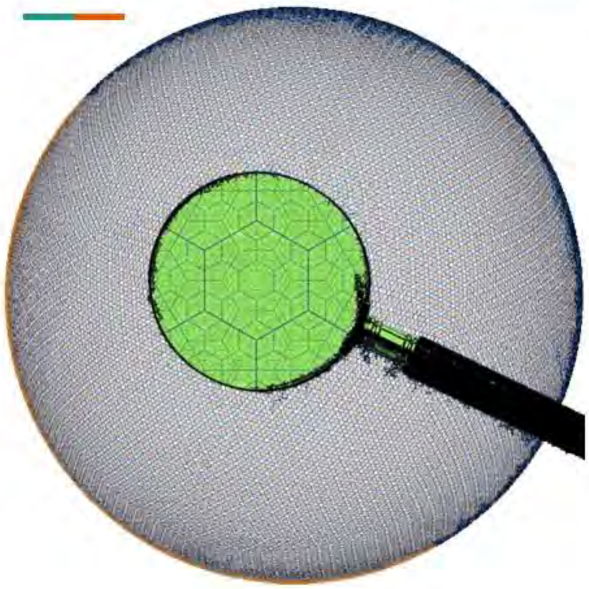

Q4
OpenGeoHUB
Connect · Create · Share · Repeat
Q4: What are the risks of using predictive mappings in decision making?

Technological and methodological advances in soil mapping and monitoring and opportunities for Alberta / Canada – Tom Hengl

These slides are periodically updated at the following link –
https://urldefense.proofpoint.com/v2/url?u=https-3A__docs.google.com_presentation_d_1tXStQEhDb4m0CYNNL4RJR6tEoTku2oJiGhvk3okyQVY_edit-3Fusp-3Dsharing&d=DwID-g&c=Nx4cNT2ND5PKJMUw1Yiqvt5PYVXmyHyY3_zNPmJl6pU&r=Eh0WZEy_sZ0BIi5Pzh7zGn81ISx1CpSbaYWJwZLAAGXAm2j72_VH5SiSG3Ku4GWQ&m=QXDQubowTMU4-pnSyOb5aqFK9N8zii4DXay-DWYvTu0&s=9dIFUdIX4JvofmBfvO-YM7vRFcOtdJ0Crhkh4N4LTwM&e=


Research, Innovation and Consultancy
EnvirometriX
March 6th 2019, InnoTech Alberta
Technological and methodological advances in soil mapping and monitoring and opportunities for Alberta / Canada
tom.hengl@opengeohub.org
@tom_hengl
thengl
https://opengeohub.org / https://envirometrix.net
tom@environmental.systems
@TomNCollins
Great to be in Canada!
1. Past: Birthplace of GIS (Roger Tomlinson, https://ucgis.org/roger-tomlinson)
2. Present: leading Geospatial data library GDAL (Frank Warmerdam, https://en.wikipedia.org/wiki/GDAL)
3. Future: Geodesic Discrete Global Grids?? (Perry Peterson, https://globalgridsystems.com/)

Pyxis (now Global Grid Systems, Inc.)

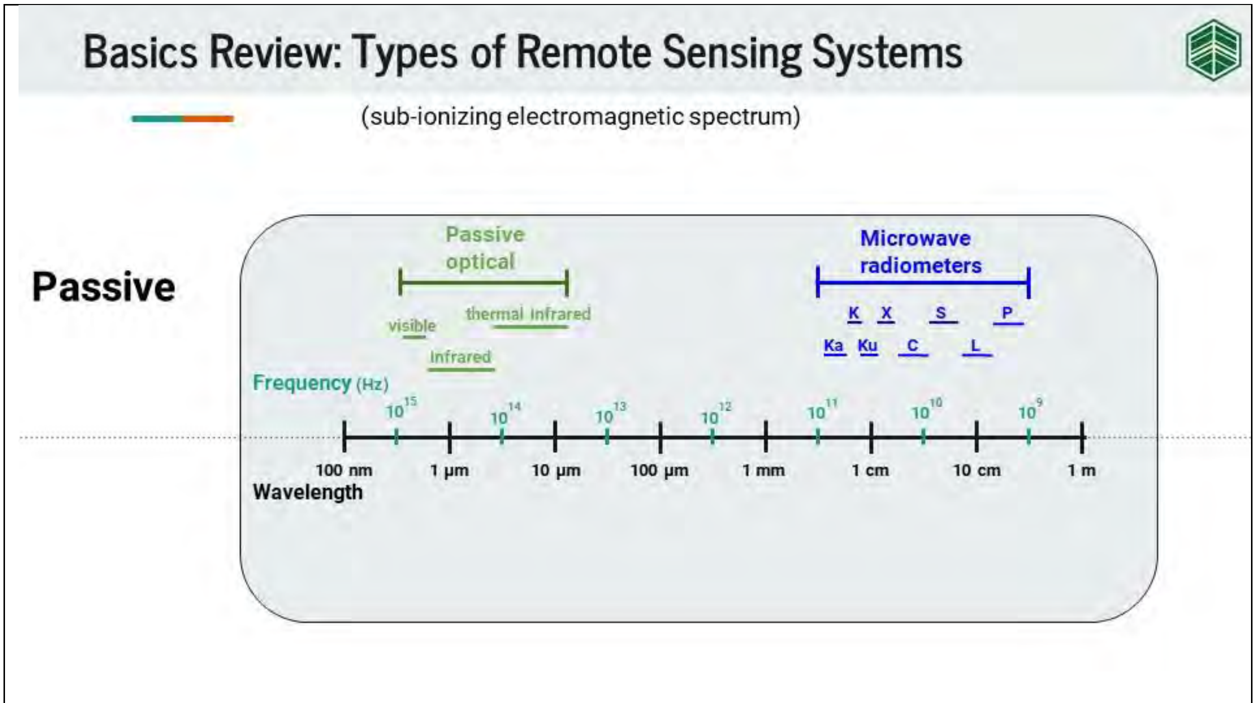


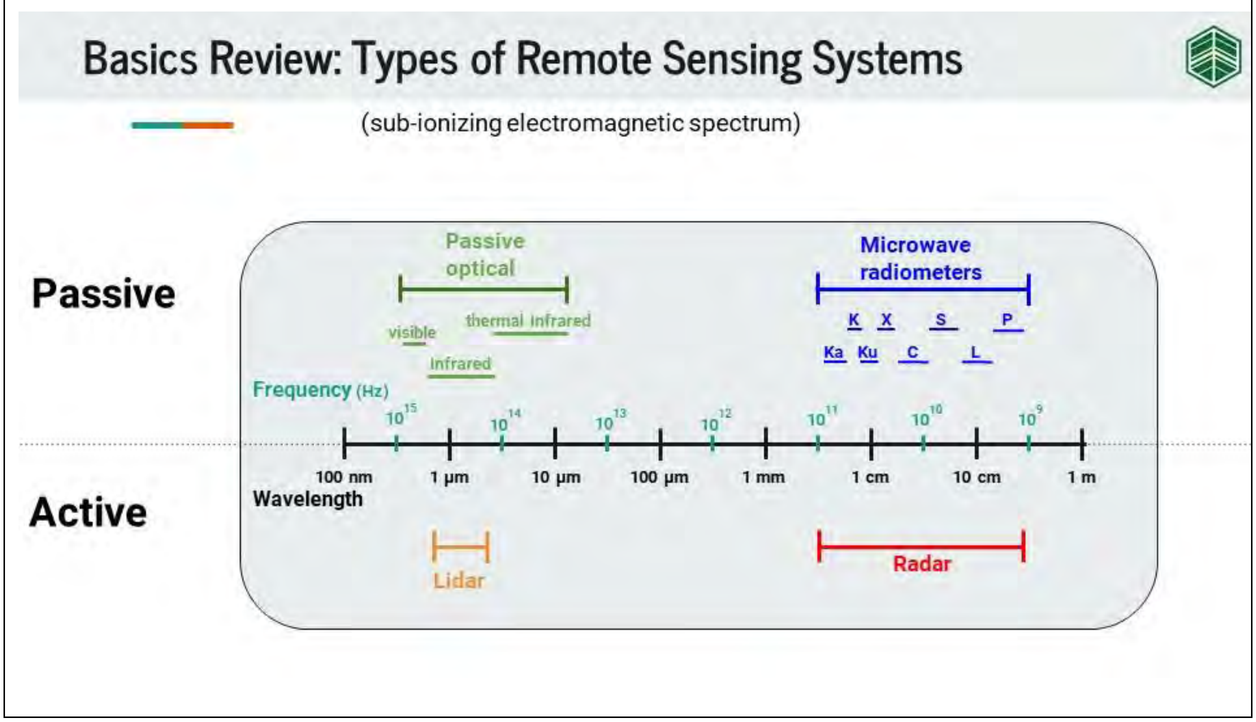
"A Discrete Global Grid (DGG) is a mosaic which covers the entire Earth's surface. Mathematically it is a space partitioning: it consists of a set of non-empty regions that form a partition of the Earth's surface.[1] In a usual grid-modeling strategy, to simplify position calculations, each region is represented by a point, abstracting the grid as a set of region-points. Each region or region-point in the grid is called a cell."

https://en.wikipedia.org/wiki/Discrete_Global_Grid

This talk

1. Advancing technologies (hardware and software) in 2019, and what to expect 2020+.
2. Lessons learned from previous efforts (to keep in mind).
3. Then, discuss:
    a. Which new application areas could benefit most?
    b. How must we prepare for the new 'big data' paradigm?
    c. How best to produce decision- and analyst-ready insight?


Basics Review: Types of Remote Sensing Systems
(sub-ionizing electromagnetic spectrum)
Passive
Passive optical
visible
infrared
thermal infrared
Microwave radiometers
K X S P
Ka Ku C L
Frequency (Hz)
10^15 10^14 10^13 10^12 10^11 10^10 10^9
100 nm 1 µm 10 µm 100 µm 1 mm 1 cm 10 cm 1 m
Wavelength


Basics Review: Types of Remote Sensing Systems
(sub-ionizing electromagnetic spectrum)
Passive
Passive optical
visible
infrared
thermal infrared
Microwave radiometers
K X S P
Ka Ku C L
Frequency (Hz)
10^15 10^14 10^13 10^12 10^11 10^10 10^9
100 nm 1 µm 10 µm 100 µm 1 mm 1 cm 10 cm 1 m
Wavelength
Active
Lidar
Radar

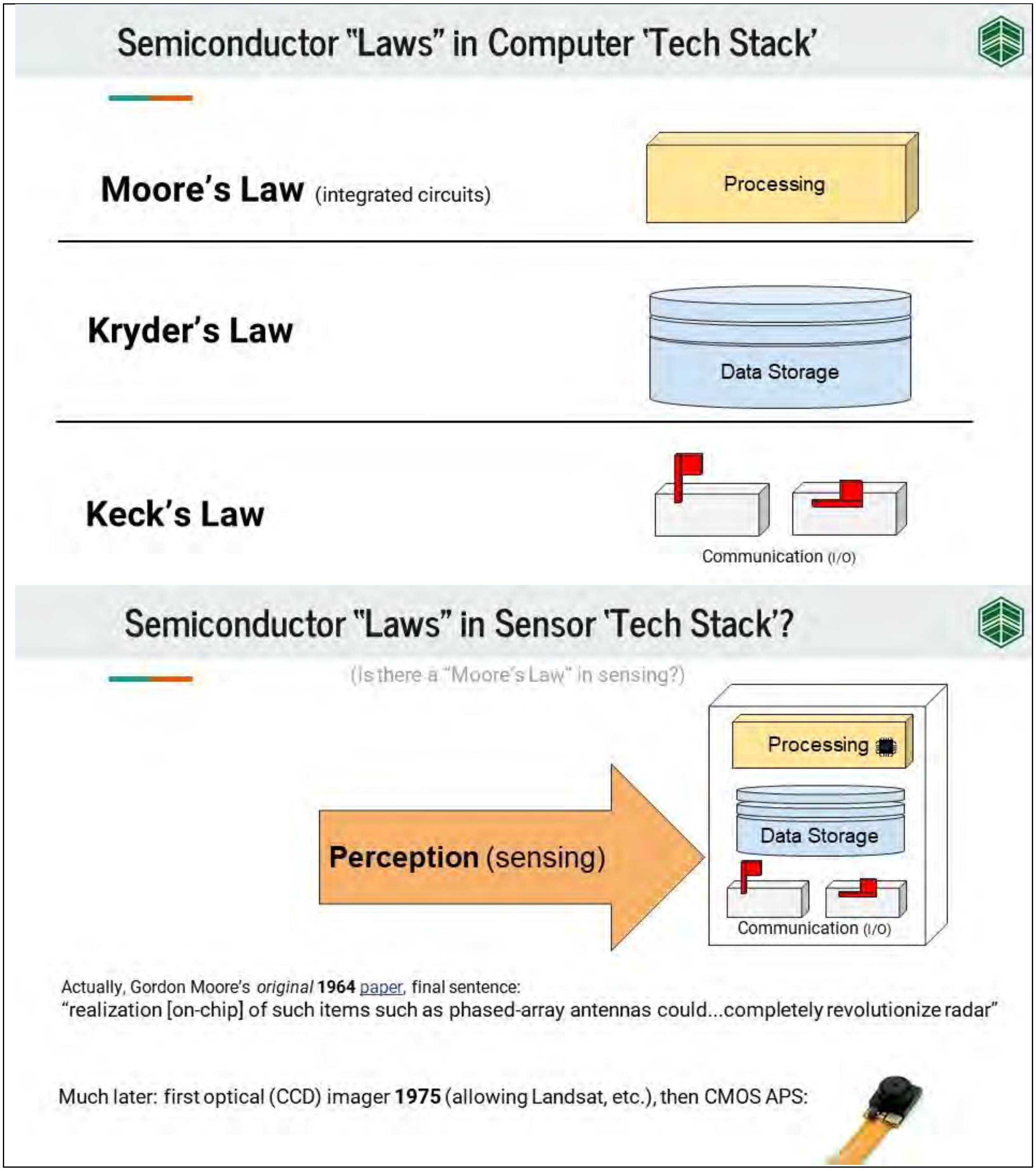
Progress in sensing following clear trends:

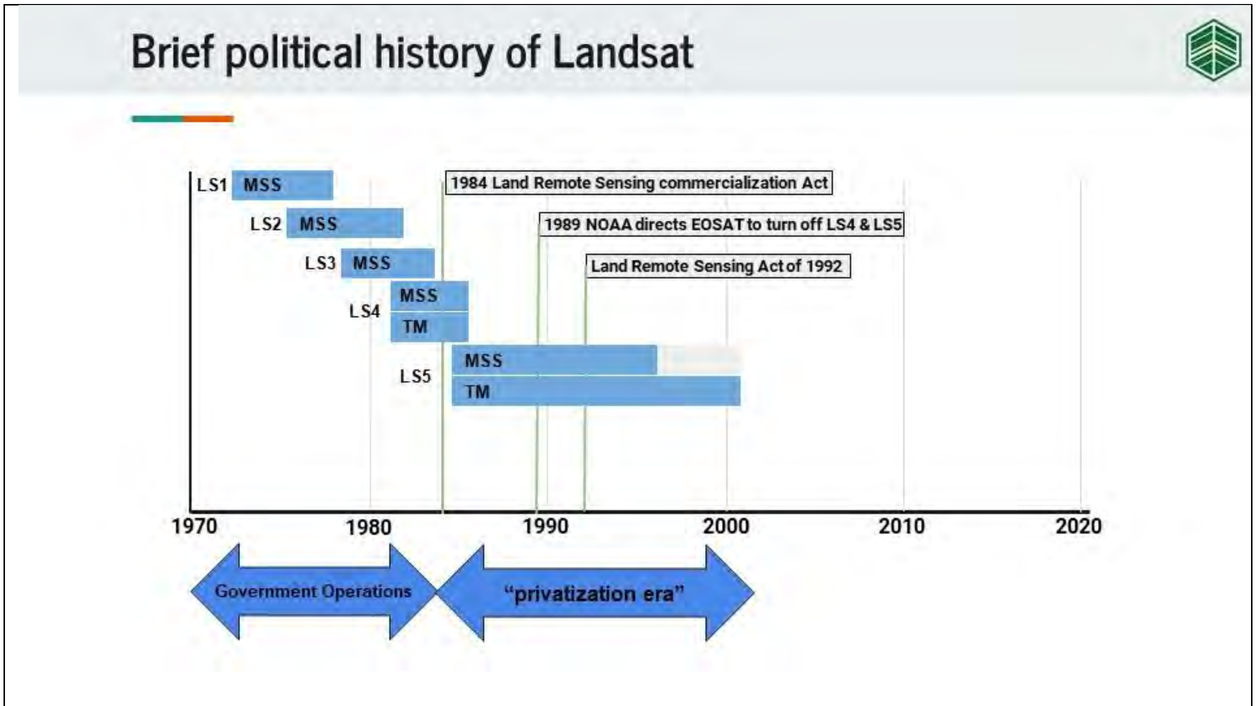
1. finer and finer resolution
   - Spatial (>1km to <1m, both optical/radar)
   - Temporal (weekly revisit, to near-continuous?!)
   - Spectral (optical, microwave spectral coverage)
2. Lower cost / more insight
   - Much public data/science becoming 'open'
   - Strong private competition and activity
3. New advances closing gaps
   - Active (lidar/radar), multi-sensor fusion, etc.
   - New airborne/surface systems

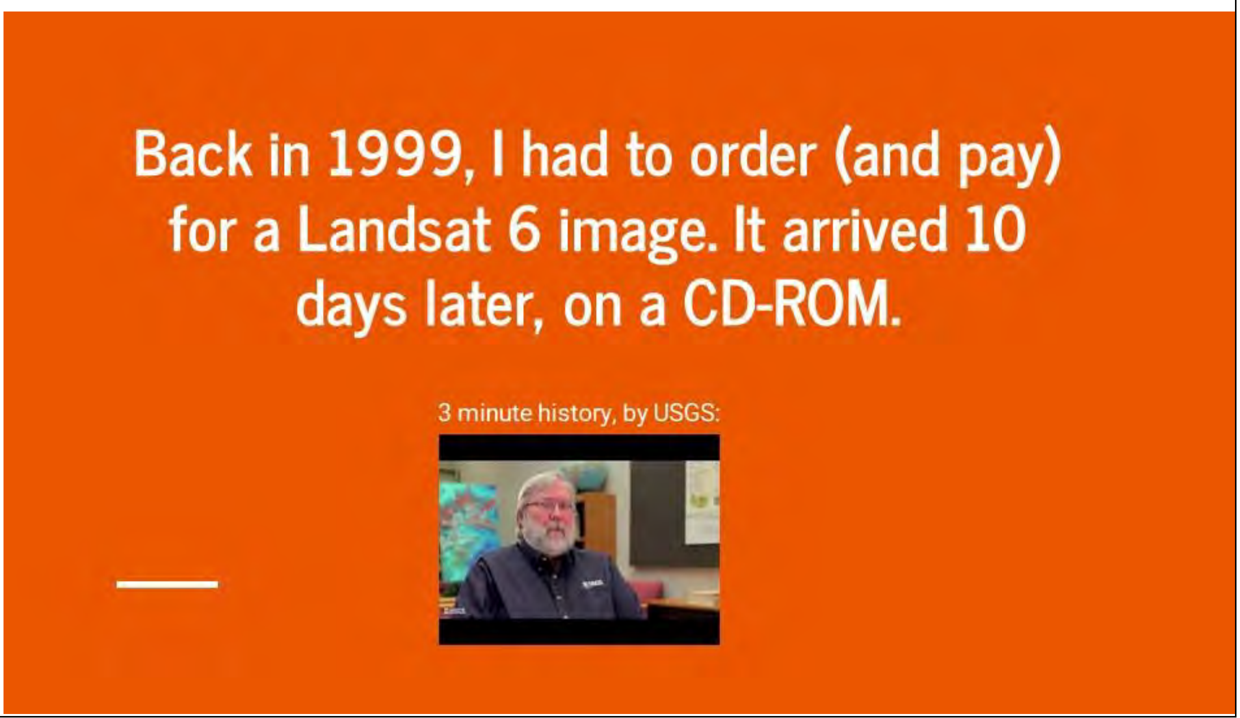
Advances in remote sensing are *accelerating.*

What are the underlying drivers?

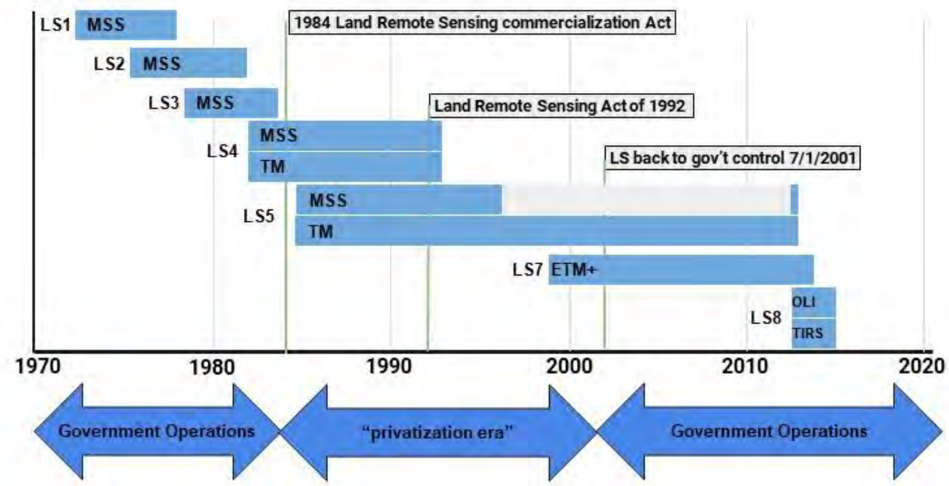
What implications are most important?


Semiconductor "Laws" in Computer 'Tech Stack'
Moore's Law (integrated circuits)
Processing
Kryder's Law
Data Storage
Keck's Law
Communication (I/O)
Semiconductor "Laws" in Sensor 'Tech Stack'?
(Is there a "Moore's Law" in sensing?)
Processing
Data Storage
Communication (I/O)
Perception (sensing)
Actually, Gordon Moore's original 1964 paper, final sentence:
"realization [on-chip] of such items such as phased-array antennas could...completely revolutionize radar"
Much later: first optical (CCD) imager 1975 (allowing Landsat, etc.), then CMOS APS:


Brief political history of Landsat
LS1 MSS
LS2 MSS
LS3 MSS
LS4 MSS TM
LS5 MSS TM
1984 Land Remote Sensing commercialization Act
1989 NOAA directs EOSAT to turn off LS4 & LS5
Land Remote Sensing Act of 1992
1970
1980
1990
2000
2010
2020
Government Operations
"privatization era"

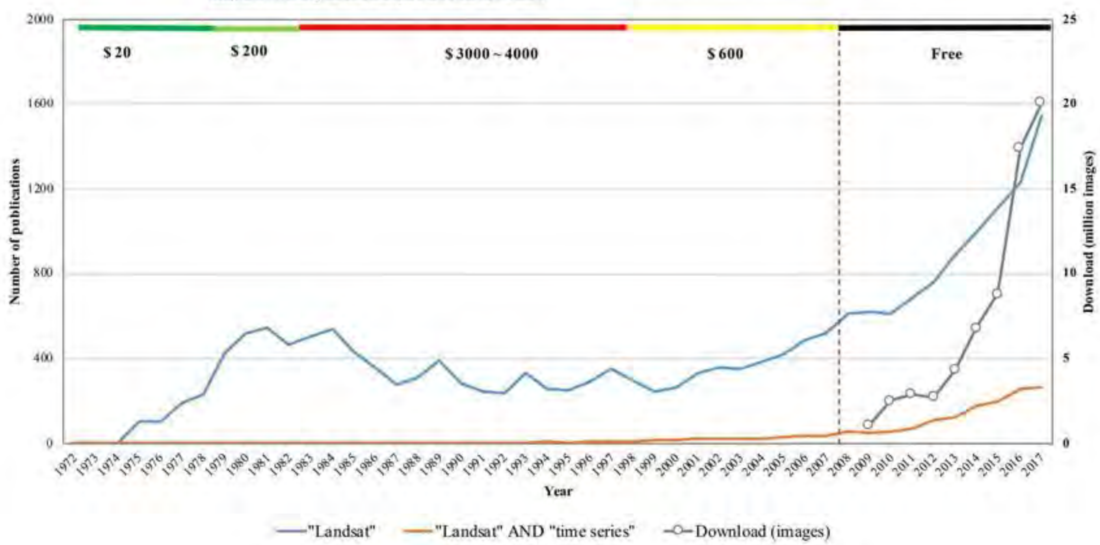

Back in 1999, I had to order (and pay) for a Landsat 6 image. It arrived 10 days later, on a CD-ROM.
3 minute history, by USGS:

Brief political history of Landsat



Data cost collapses to $0

Zhu, Wulder, et al. (2019). *Benefits of the free and open Landsat data policy.* https://doi.org/10.1016/j.rse.2019.02.016

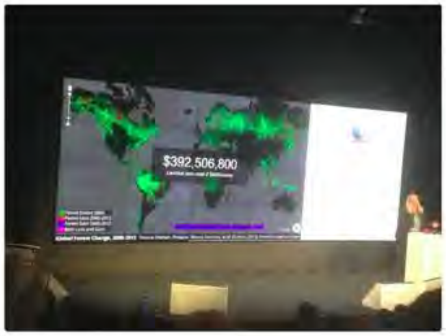


Impact of "Free and Open"

Ollie Wearn
@olliewearn

Follow

Pre-2008, Landsat wasn't free. It would have cost $400 million dollars to buy the Landsat scenes for the Hansen et al. 2013 forest cover study! Impossible. Kudos to NASA/USGS for their foresight 👊 #EEUS18



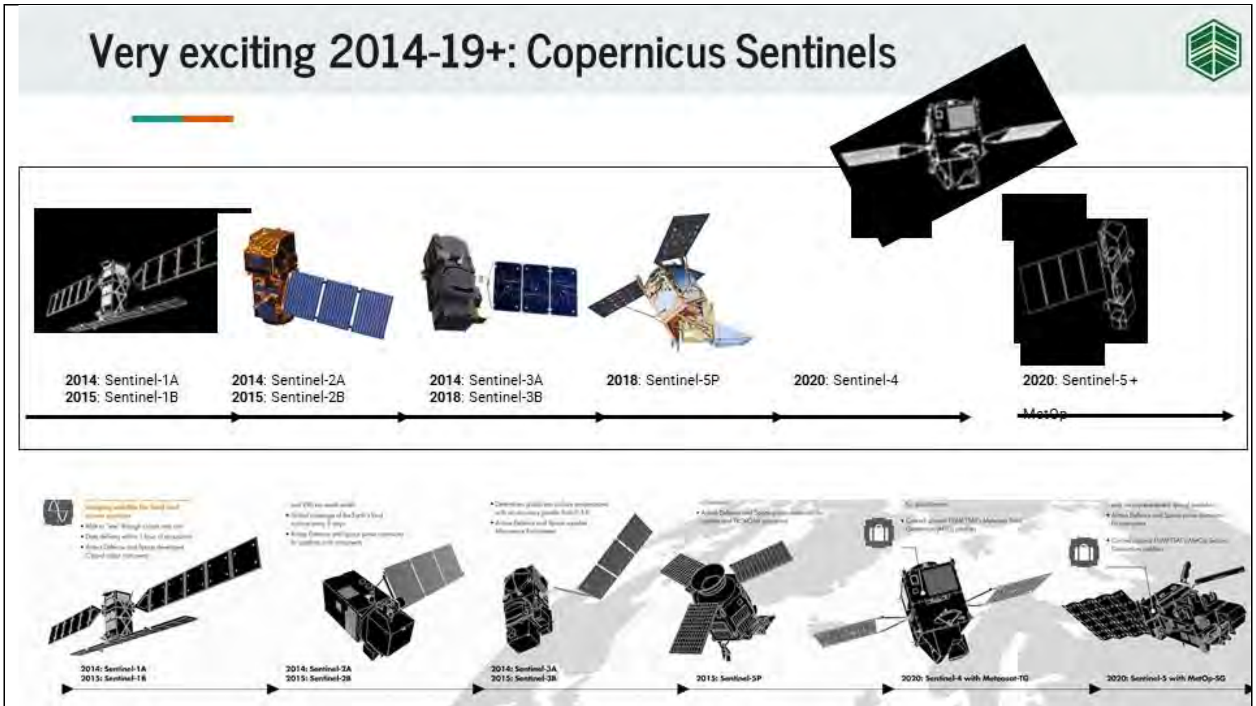
10:53 AM - 12 Jun 2018

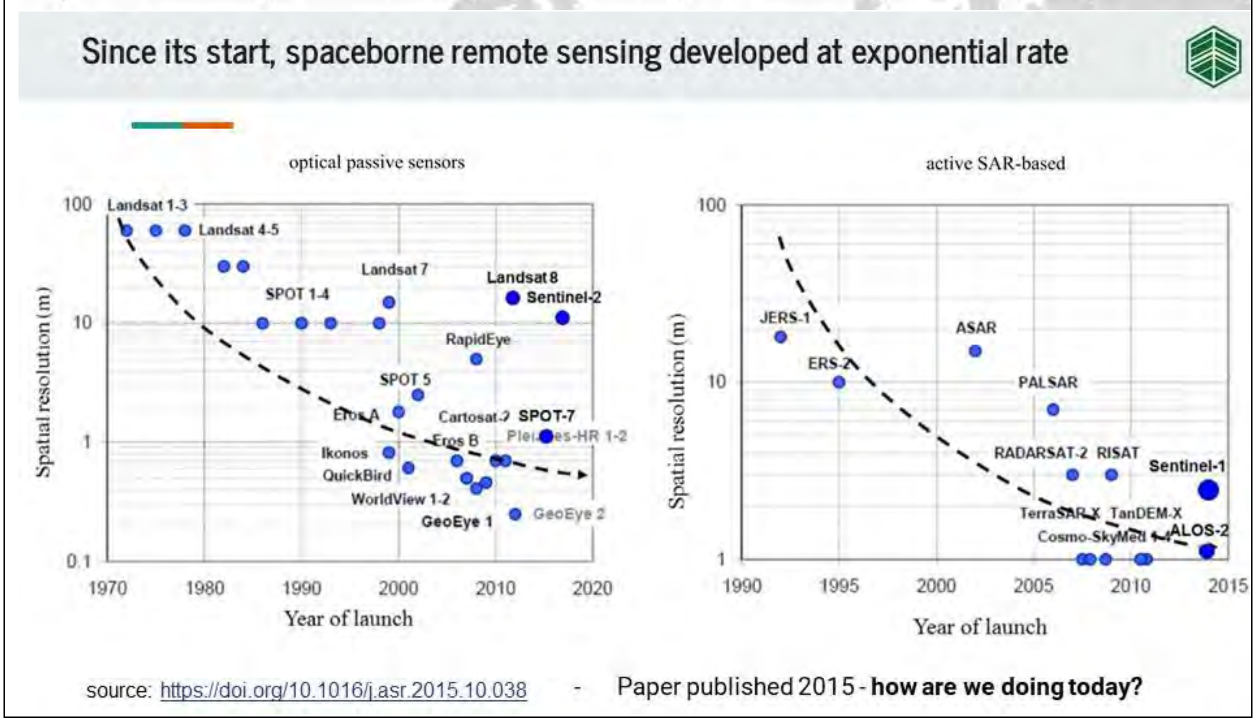
We should be grateful to those in the USA, EU, Canada, Japan for pushing to release RS data publicly / being the trend-setters...

European Space Agency - Sentinel Data Policy

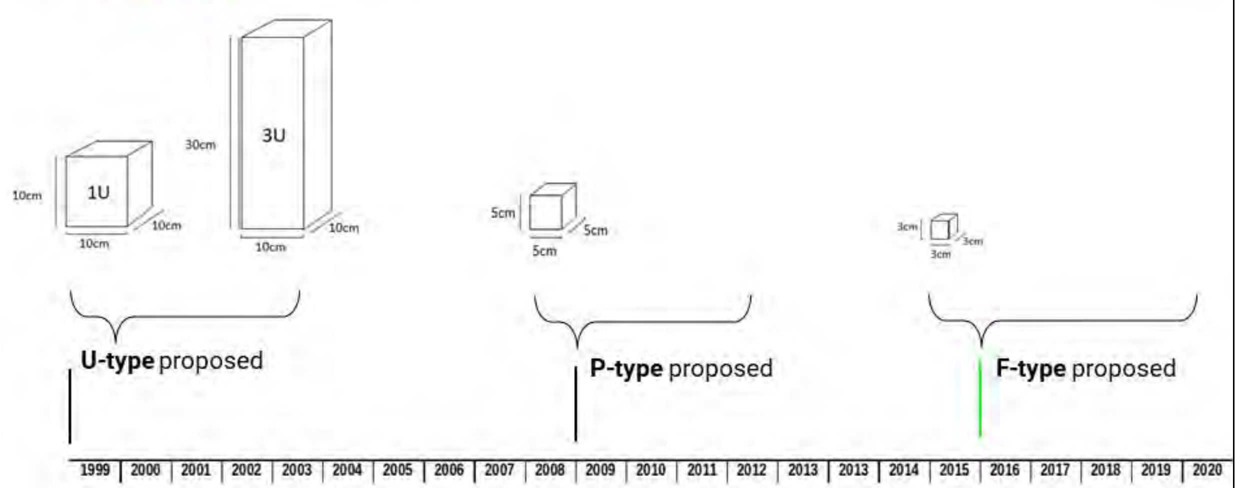
Old data policy: research and commercial use different license.

Sentinel Data Policy since Nov. 2013: *Free, Full, and Open*


Very exciting 2014-19+: Copernicus Sentinels
2014: Sentinel-1A
2015: Sentinel-1B
2014: Sentinel-2A
2015: Sentinel-2B
2014: Sentinel-3A
2018: Sentinel-3B
2018: Sentinel-5P
2020: Sentinel-4
2020: Sentinel-5+
MetOp


Since its start, spaceborne remote sensing developed at exponential rate
optical passive sensors
Spatial resolution (m)
Year of launch
Landsat 1-3
Landsat 4-5
Landsat 7
Landsat 8
Sentinel-2
SPOT 1-4
RapidEye
SPOT 5
Eros A
Cartosat-2
SPOT-7
Eros B
Pleiades-HR 1-2
Ikonos
QuickBird
WorldView 1-2
GeoEye 1
GeoEye 2
active SAR-based
JERS-1
ERS-2
ASAR
PALSAR
RADARSAT-2
RISAT
Sentinel-1
TerraSAR-X
TanDEM-X
Cosmo-SkyMed 1-4
ALOS-2
source: https://doi.org/10.1016/j.asr.2015.10.038 - Paper published 2015 - how are we doing today?

Tiny satellites: "CubeSats"



Theme 4: Daily VHR images in near future

Mike Wulder is on the Landsat Science Team. Also Canadian Forest Service.

But things don't always go smoothly.

Some not-so-successful examples:

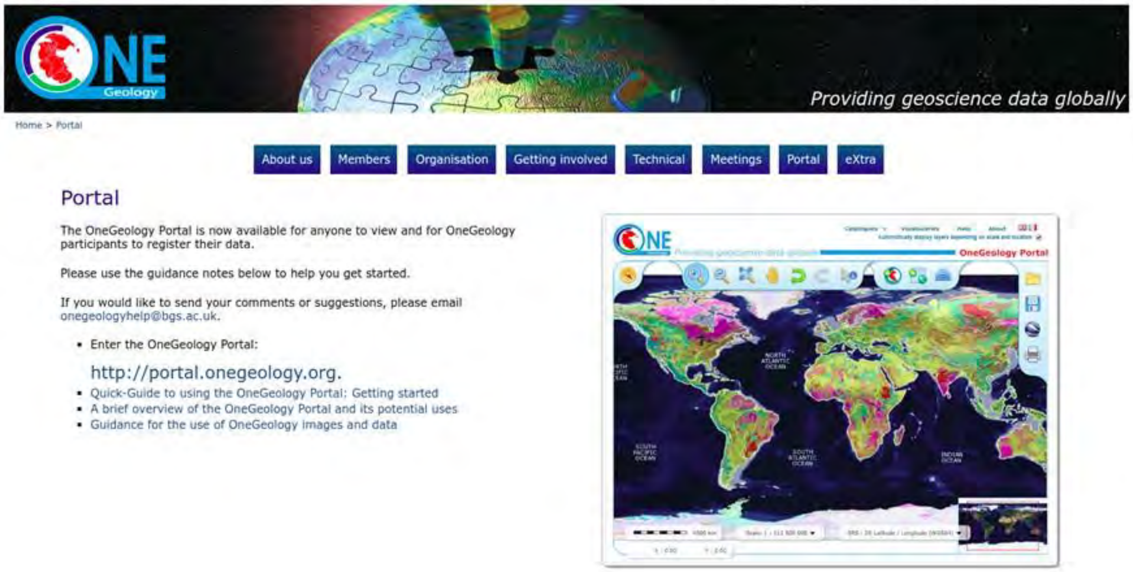
Example 1: Dry matter productivity in kg/ha/day

Derived differences in DMP from MODIS and/or Copernicus do not match available deforestation maps (e.g. from the GlobalForestWatch).

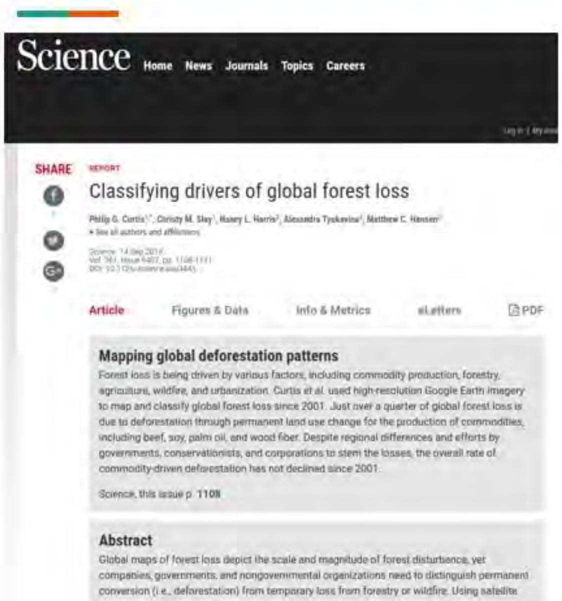
Possible explanations:

- *high noise in data?*
- *model parameters for some areas (tropics) inaccurate?*
- *cloud cover problems?*

Example 2: OneGeology

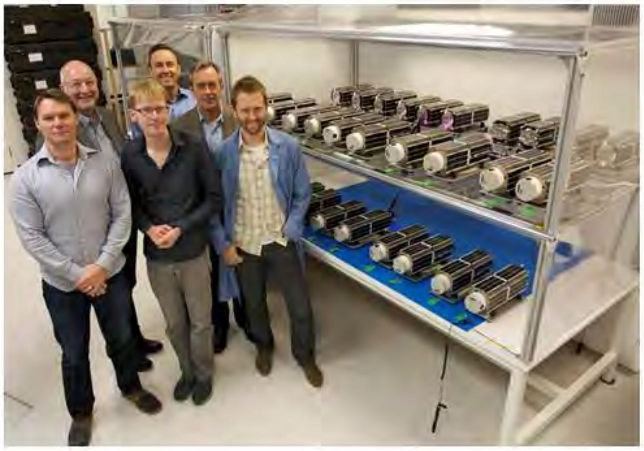


Impact: so far, more data hasn't stopped deforestation



"Despite regional differences and efforts by governments, conservationists, and corporations to stem the losses, **the overall rate of commodity-driven deforestation has not declined since 2001."**

Data explosion - new technical and ethical questions



In Feb. 2017 Planet Labs, Inc. (AKA Google) claimed to collect a 50 Terapixel image every day - whole earth at 5m resolution.

In 2019, at least 8 competitors, proposing to collect much more!

Are we in danger to be overwhelmed with data we do not need?

These companies sell only to those who pay! Ethics, inequality issue?

Not clear if applied work will be easier

"The rich "network effects" made possible by chained Web services, GRID computing, sensor webs, geospatial semantics, and online catalogs for data, services and schemas hold great promise, but there is no guarantee that this promise will be fulfilled.

The question is, can we find the institutional will - in academia and government - to make changes that enable societies around the world to make the most of these new tools?"

Prof. Mike Jackson, David Schell and Prof. D.R. Fraser Taylor:
https://www.directionsmag.com/article/2366

nature > correspondence > article

a natureresearch journal

nature

CORRESPONDENCE · 26 FEBRUARY 2019

Soil pollution — speed up global mapping

Deyi Hou & Yong Sik Ok

Too few countries are investing in national surveys of soil pollution. A global map is urgently needed, not least to prevent international trading of contaminated produce and the migration of persistent organic pollutants across borders. We urge all member states at next month's fourth session of the UN Environment Assembly (UNEA) to speed up their assessments.

A global map of soil pollution will also guide policymakers on protecting soils; inform chemical and waste management (see Y. Geng et al. Nature 565, 153–155; 2019); prevent further pollution by identifying sources and controlling polluter behaviour; and reduce risks to public health and the environment.

The World Health Organization and the United Nations Food and Agriculture Organization are among those required by the UNEA since December 2017 to report on the extent of global soil pollution, monitor future trends and identify associated risks and impacts. The results will be presented at the UNEA's fifth session in 2020. Many hurdles must be overcome before a global assessment can be made. Collaboration between developing and developed nations in allocating technical and financial resources is a priority.

PDF version

RELATED ARTICLES

See more letters to the editor

SUBJECTS

Environmental sciences · Policy

Without global analysis and global monitoring systems, it is difficult to appreciate how much is done locally.

Challenges for Alberta

Environmental Reviews

Home · About Us · Journals · Books · Compilations · Open Access · Authors · Librarians · Societies

Review

Air synthesis review: polycyclic aromatic compounds in the oil sands region

Abstract · PDF (6939 K) · PDF-Plus (2810 K) · Figures · Tables · References · Citing articles

Browse the journal

- List of issues
- e-First articles
- Just-IN articles

The following topics have been identified as new science priorities:

(*i*) improving emissions reporting
(*ii*) improving information on the ambient concentrations, long-range transport, and...release (with co-contaminants) from different types of mining activities;
(*iii*) further optimizing...assessing PAC deposition (e.g., snow sampling, lichens, passive ambient sampling) spatially across the [Oil Sands] and downwind regions;
(*iv*) designing projects that integrate monitoring efforts with source attribution models and ecosystem health studies to improve understanding of sources, receptors, and effects;
(*v*) further optimizing natural deposition archives (e.g., sediment, peat, tree rings) and advanced forensic techniques (e.g., isotope analysis, marker compounds) to provide better understanding...over space and time;
(*vi*) conducting process research to improve...assessing exposure to wildlife and humans; and
(*vii*) developing tools and integrated strategies for assessing cumulative risk to wildlife and humans by accounting for the toxicity of the mixture of chemicals..."

We are now in 2019,

What are the most exciting new global data releases?

What new application areas will result from these data?

Theme 1: Copernicus land products

Global, improved 100 m resolution land cover classification product

Early 2019?


Theme 2: ALOS DSM product
Home| Site Map| Contact Us| Japanese
Advanced Land Observing Satellite DAICHI-2・DAICHI ALOS-2・ALOS
ALOS Research and Application Project of EORC, JAXA
Image Library
ALOS-2 Project
Calibration&Validation
ALOS Observation Strategy
Use DATA
Dataset
Conference & Workshop
Kyoto & Carbon Initiative
Research Announcement
Home > Dataset > ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)"
ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)"
AW3D30 Version 2.1
Black color means no-data area due to masked out by cloud/snow cover.
Please click the map to zoom in/out in another browser.
Best elevation product at 30 m at the moment?


Theme 2: ALOS RADAR, forest map. products
Advanced Land Observing Satellite DAICHI-2・DAICHI ALOS-2・ALOS
ALOS Research and Application Project of EORC, JAXA
Image Library
ALOS-2 Project
Calibration&Validation
ALOS Observation Strategy
Use DATA
Dataset
Conference & Workshop
Kyoto & Carbon Initiative
Research Announcement
Dataset Homepage
This page shows the research products processed in EORC, JAXA. These will be used in many application fields to contribute the Earth science.
K&C Mosaic Homepage
The reference dataset for the "ALOS Kyoto and Carbon Initiative Project".
1) PALSAR 50m Orthorectified Mosaic Product.
2) PALSAR 500m Browse Mosaic Product.
Global PALSAR-2/PALSAR/JERS-1 Mosaic and Forest/Non-Forest map
Global 25m-resolution mosaic using ALOS-2 PALSAR-2, ALOS PALSAR, and JERS-1 SAR images and global forest/non-forest map.
Old Version
PALSAR 10m Mosaic Data
PALSAR mosaic dataset with 10m spatial resolution.
1) Porto Velho, Rondonia, Brazil
2) Riau, Indonesia
3) Congo River
Global Forest / Non-forest map
The global forest and non-forest area distribution in 2007 and 2009 by PALSAR.
Already available for download!

Theme 3: TANDEM-x forest/non-forest map at 50 m

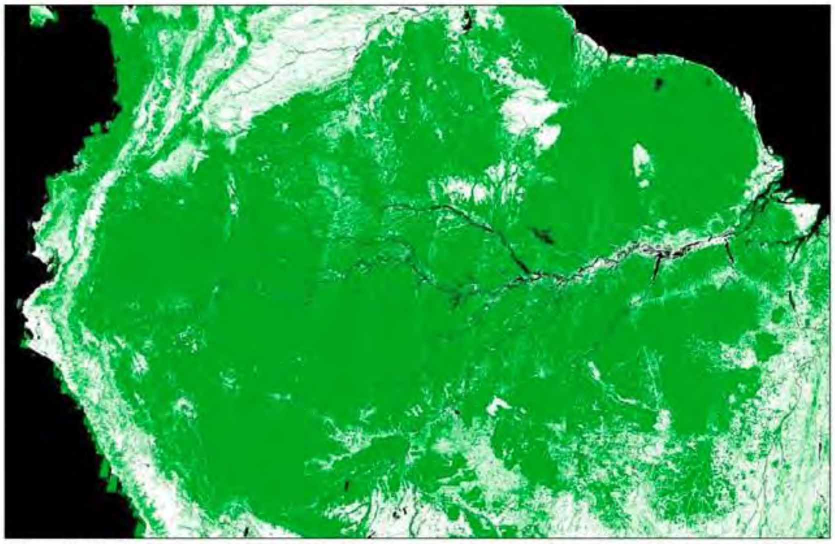


Figure 17: TanDEM-X Forest/Non-Forest Map example over the Amazon Rainforest (image credit: DLR)

Forest/Non-forest map of the world at very high accuracy and 50 m resolution (coming out soon)

Theme 3: TANDEM-x global DEM at 100 m

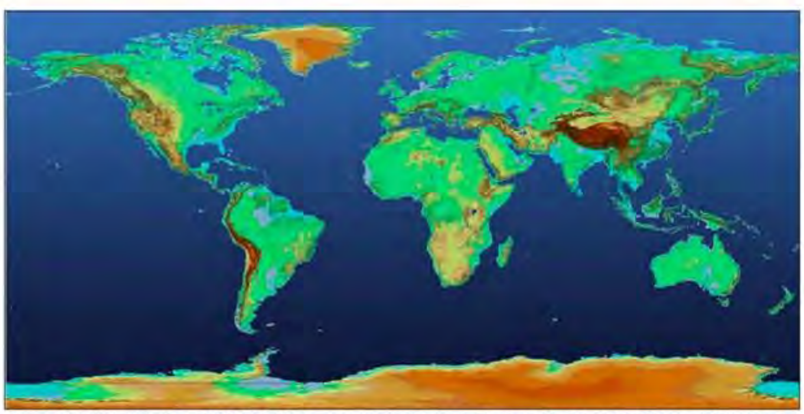


Figure 19: The global TanDEM-X DEM is a consistent data set covering all land surfaces at unprecedented absolute height accuracy of about 1m at a horizontal sampling of 12 m by 12 m. Between 2011 and 2014 at least two acquisitions have been collected by the bistatic TanDEM-X SAR interferometer, mountainous areas have covered up to six times (image credit: DLR)

Global land surface model at 100 m OPEN DATA LICENSE? (release date: end of 2018?)

MirrorSAR (6 m?)

Theme 3: TANDEM-x canopy height data

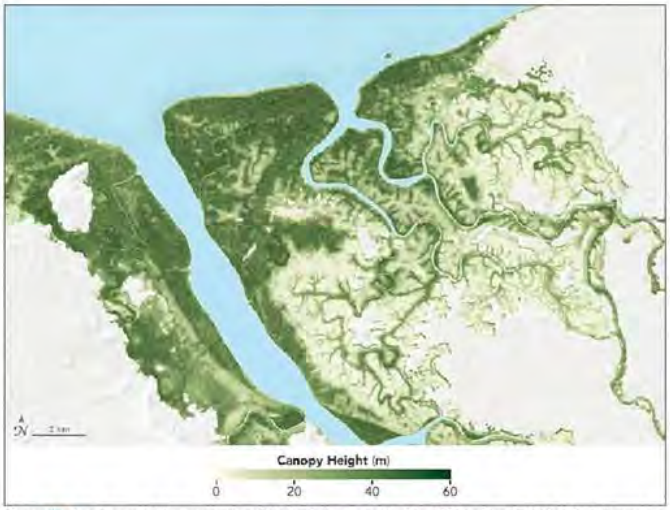
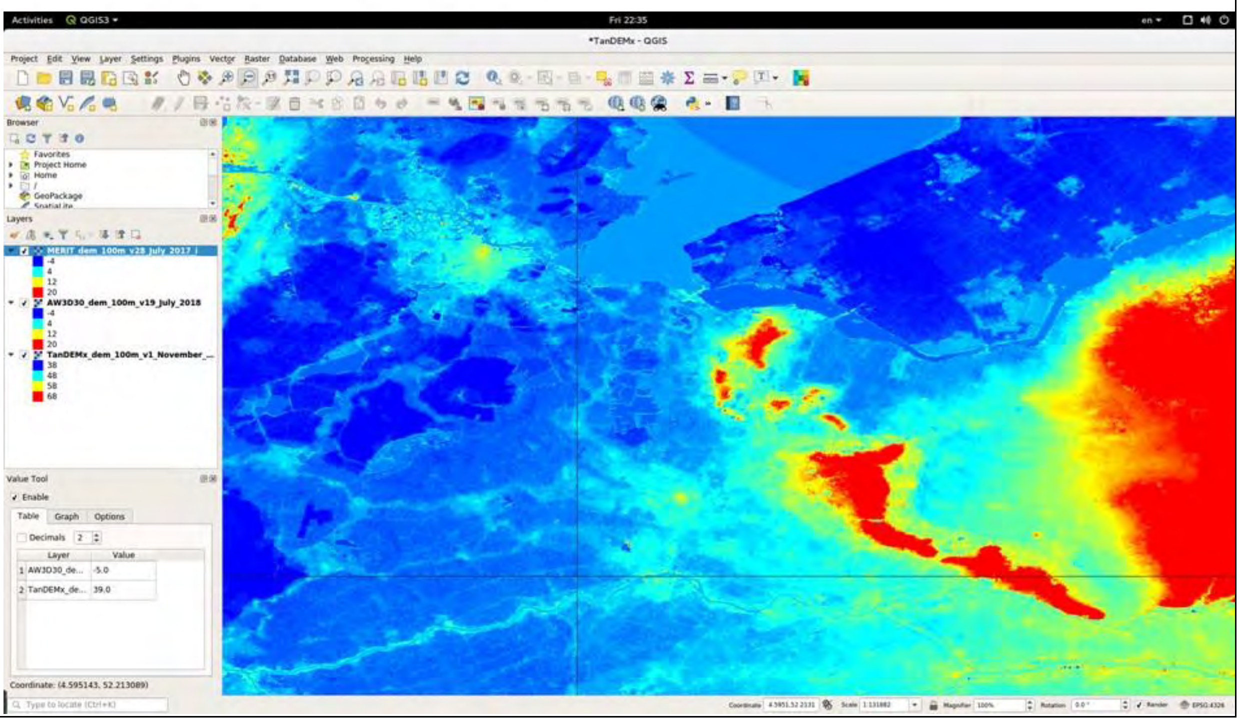
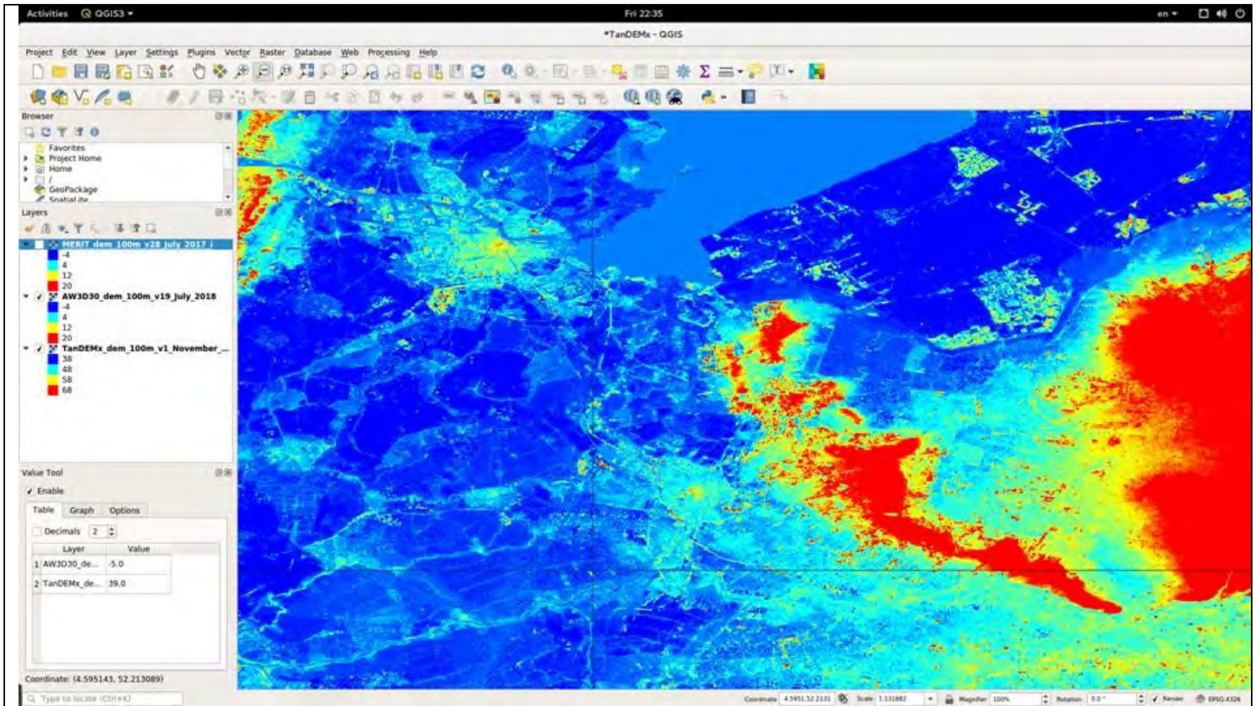


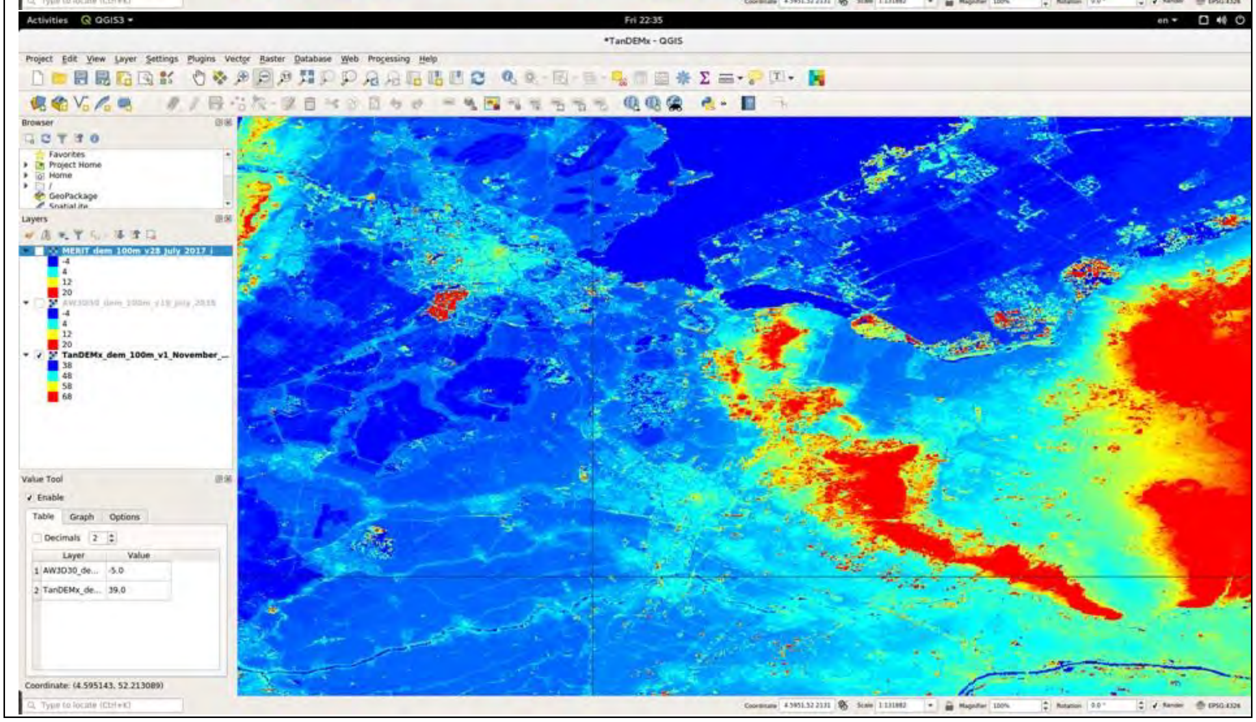
Figure 23: Detail map of the mangrove parks in Gabon, acquired in 2015 by TanDEM-X of DLR (image credit: NASA Earth Observatory, maps by Joshua Stevens, using canopy height data courtesy of SeungKuk Lee/NASA GSFC/NASA Carbon Monitoring Systems. Story by Kathryn Hansen)

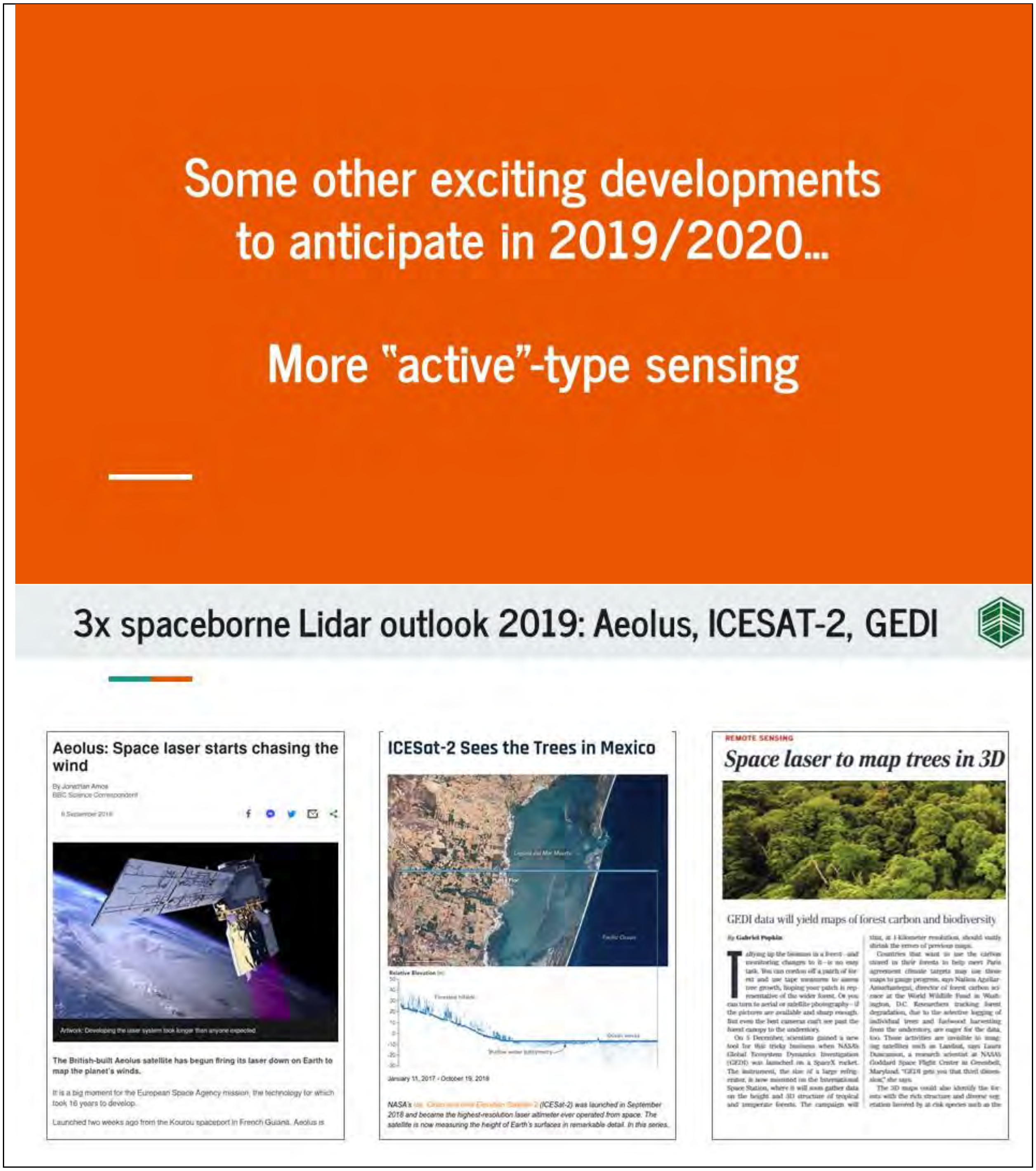
High vertical accuracy, detailed canopy height

(from 2019, every year? global volume changes?)

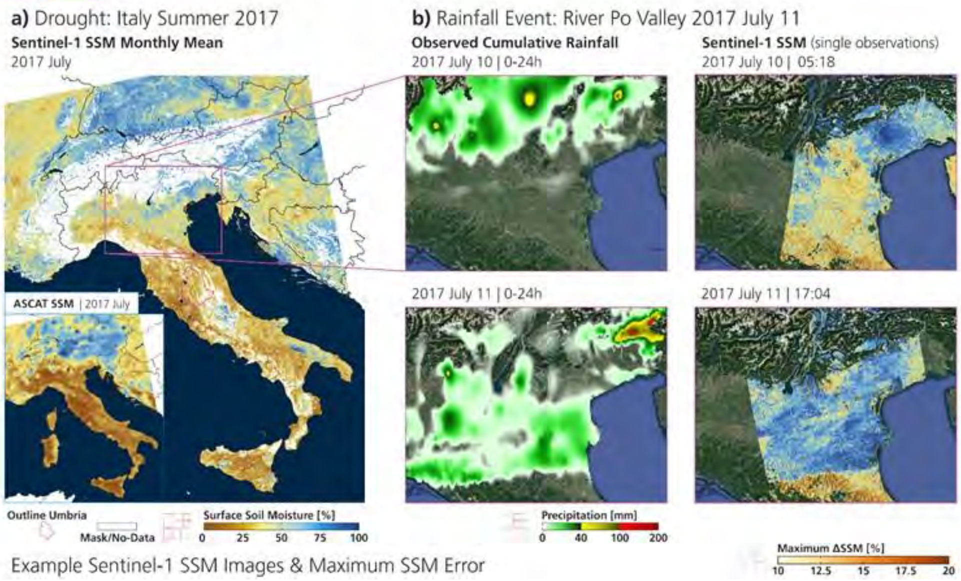



Activities
QGIS3
Fri 22:35
*TanDEMx - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Browser
Favorites
Project Home
Home
GeoPackage
Layers
MERIT dem 100m v28 July 2017
-4
4
12
20
AW3D30_dem_100m_v19_July_2018
-4
4
12
20
TanDEMx_dem_100m_v1_November_...
38
48
58
68
Value Tool
Enable
Table
Graph
Options
Decimals 2
Layer
Value
1 AW3D30_de... -5.0
2 TanDEMx_de... 39.0
Coordinate: (4.595143, 52.213089)

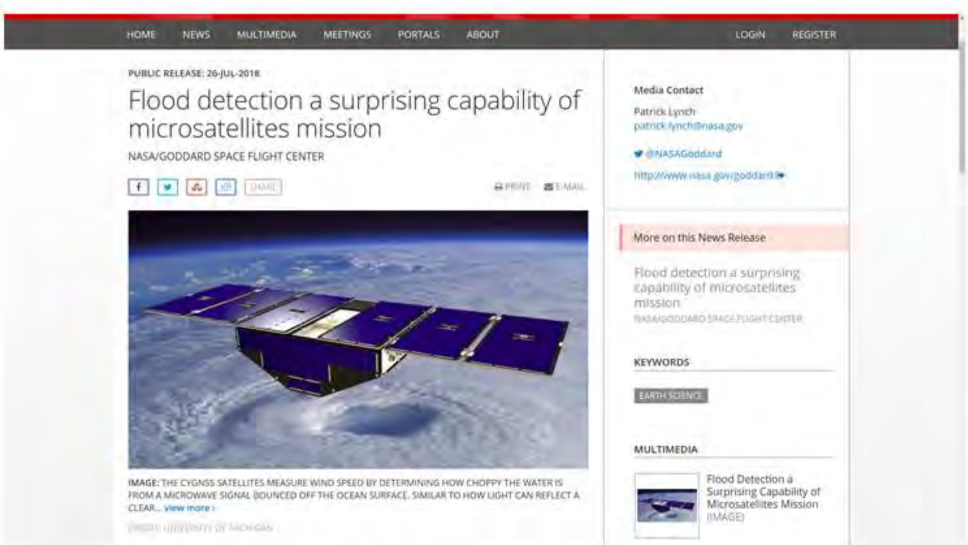

Activities
QGIS3
Fri 22:35
*TanDEMx - QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Browser
Favorites
Project Home
Home
GeoPackage
Layers
MERIT dem 100m v28 July 2017
-4
4
12
20
-4
4
12
20
TanDEMx_dem_100m_v1_November_...
38
48
58
68
Value Tool
Enable
Table
Graph
Options
Decimals 2
Layer
Value
1 AW3D30_de... -5.0
2 TanDEMx_de... 39.0
Coordinate: (4.595143, 52.213089)

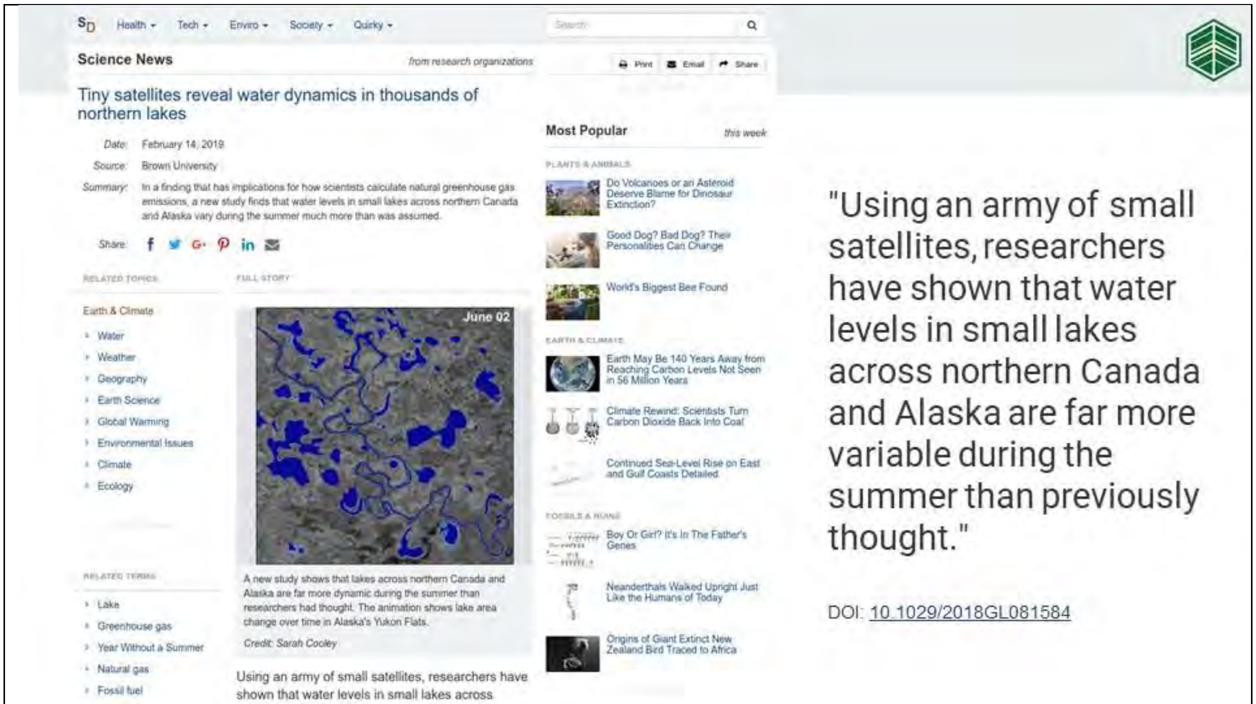

Some other exciting developments to anticipate in 2019/2020…
More "active"-type sensing
3x spaceborne Lidar outlook 2019: Aeolus, ICESAT-2, GEDI
Aeolus: Space laser starts chasing the wind
By Jonathan Amos
BBC Science Correspondent
Artwork: Developing the laser system took longer than anyone expected
The British-built Aeolus satellite has begun firing its laser down on Earth to map the planet's winds.
It is a big moment for the European Space Agency mission, the technology for which took 16 years to develop.
Launched two weeks ago from the Kourou spaceport in French Guiana, Aeolus is
ICESat-2 Sees the Trees in Mexico
January 11, 2017 - October 19, 2018
NASA's Ice, Cloud and land Elevation Satellite-2 (ICESat-2) was launched in September 2018 and became the highest-resolution laser altimeter ever operated from space. The satellite is now measuring the height of Earth's surfaces in remarkable detail. In this series,
REMOTE SENSING
Space laser to map trees in 3D
GEDI data will yield maps of forest carbon and biodiversity
By Gabriel Popkin

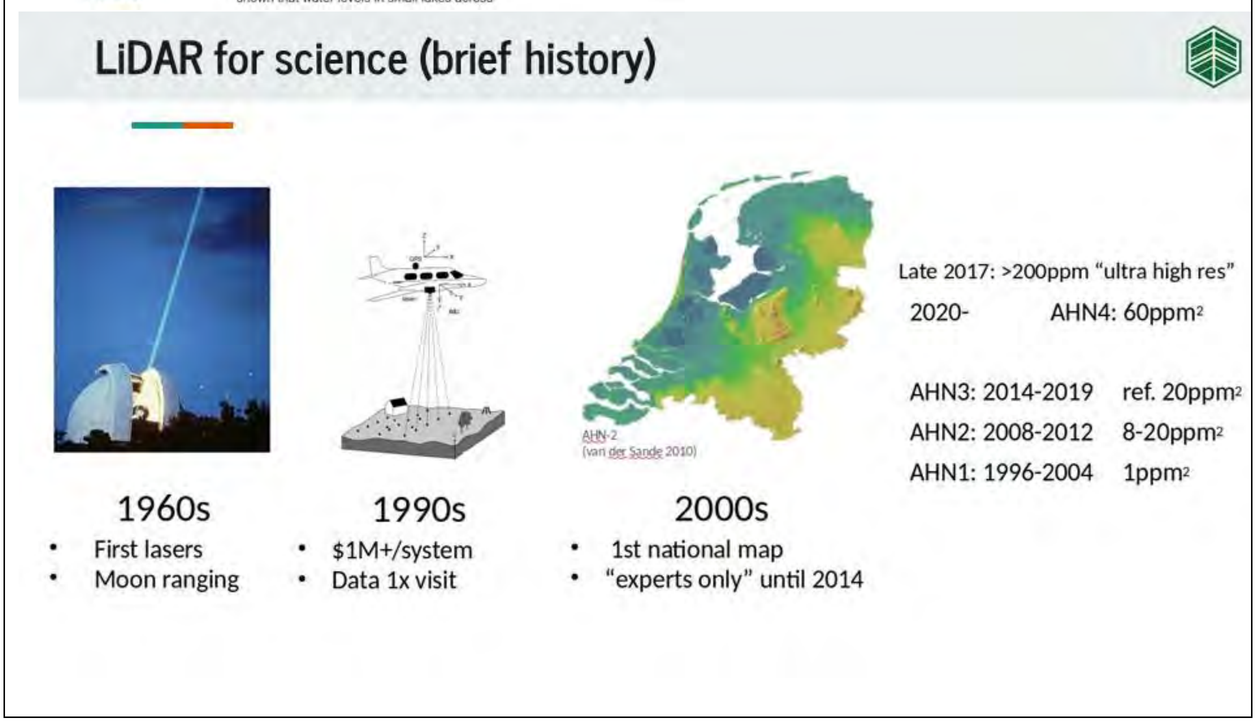
Sentinel 1 - Monitoring soil moisture Italy

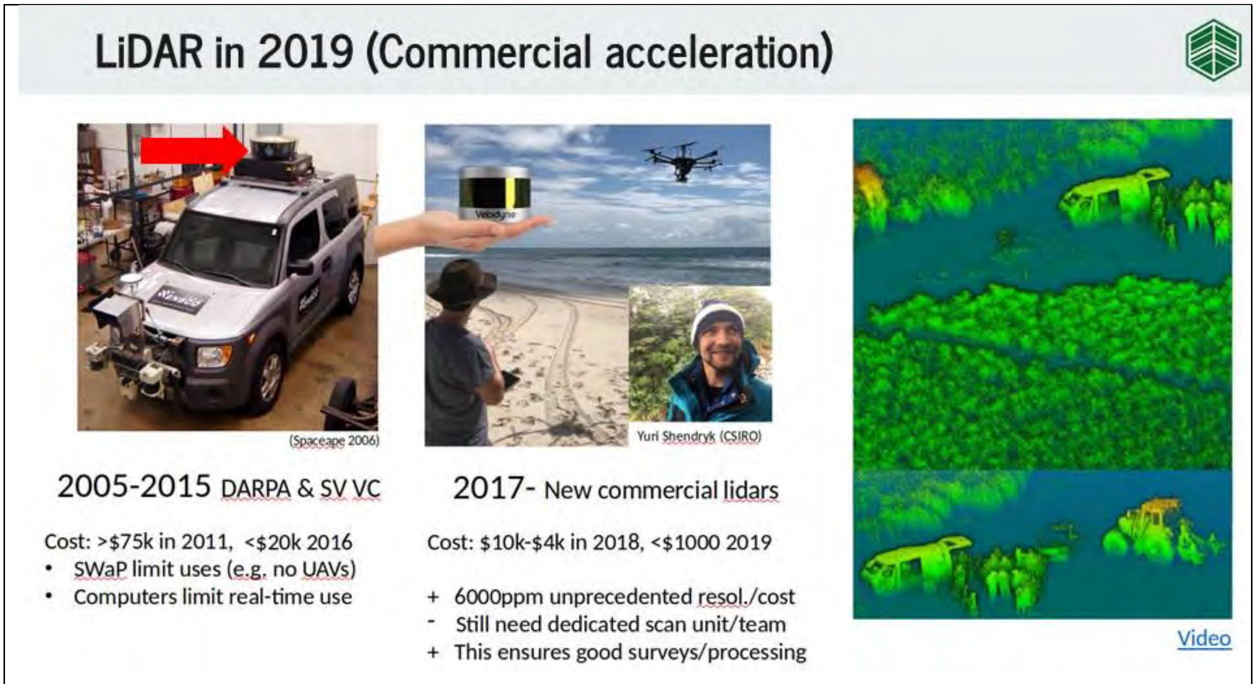


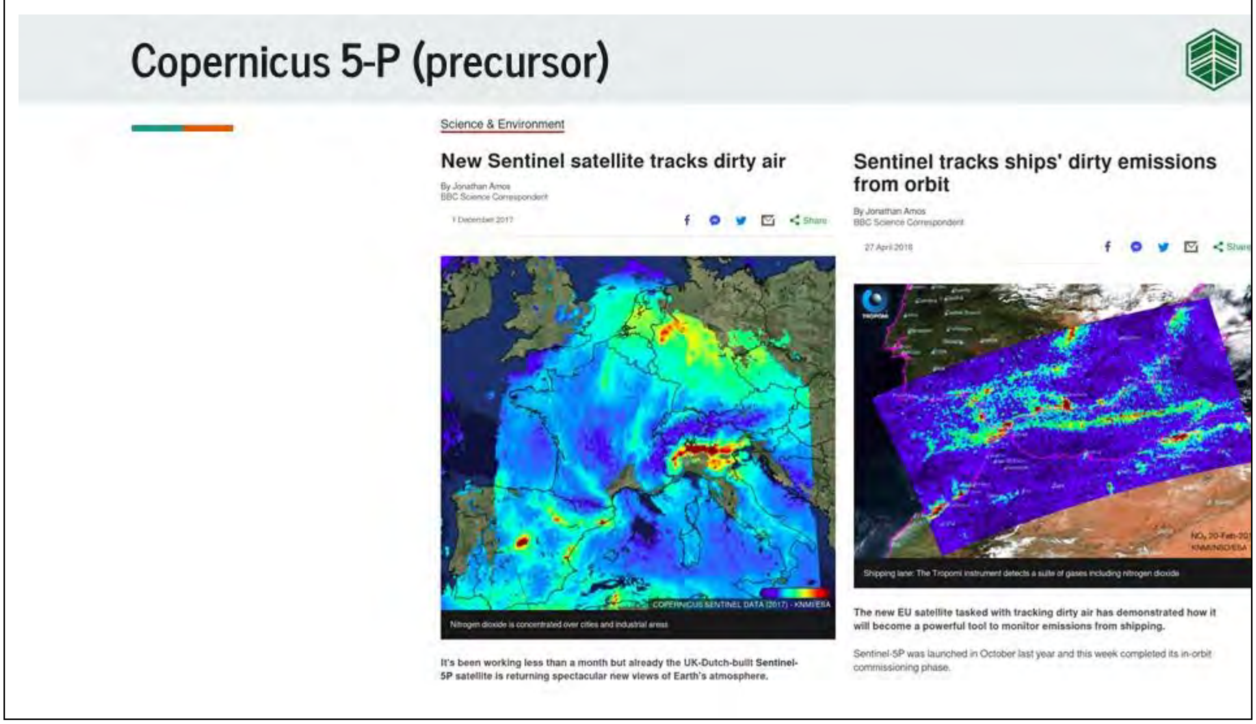
CYGNSS - L-band bistatic "passive radar" using GPS




Health
Tech
Enviro
Society
Quirky
Science News
from research organizations
Print
Email
Share
Tiny satellites reveal water dynamics in thousands of northern lakes
Date: February 14, 2019
Source: Brown University
Summary: In a finding that has implications for how scientists calculate natural greenhouse gas emissions, a new study finds that water levels in small lakes across northern Canada and Alaska vary during the summer much more than was assumed.
Share:
RELATED TOPICS
Earth & Climate
Water
Weather
Geography
Earth Science
Global Warming
Environmental Issues
Climate
Ecology
FULL STORY
June 02
A new study shows that lakes across northern Canada and Alaska are far more dynamic during the summer than researchers had thought. The animation shows lake area change over time in Alaska's Yukon Flats.
Credit: Sarah Cooley
RELATED TERMS
Lake
Greenhouse gas
Year Without a Summer
Natural gas
Fossil fuel
Using an army of small satellites, researchers have shown that water levels in small lakes across
Most Popular
this week
Do Volcanoes or an Asteroid Deserve Blame for Dinosaur Extinction?
Good Dog? Bad Dog? Their Personalities Can Change
World's Biggest Bee Found
Earth May Be 140 Years Away from Reaching Carbon Levels Not Seen in 56 Million Years
Climate Rewind: Scientists Turn Carbon Dioxide Back Into Coal
Continued Sea-Level Rise on East and Gulf Coasts Detailed
Boy Or Girl? It's In The Father's Genes
Neanderthals Walked Upright Just Like the Humans of Today
Origins of Giant Extinct New Zealand Bird Traced to Africa
"Using an army of small satellites, researchers have shown that water levels in small lakes across northern Canada and Alaska are far more variable during the summer than previously thought."
DOI: 10.1029/2018GL081584


LiDAR for science (brief history)
AHN-2
(van der Sande 2010)
Late 2017: >200ppm "ultra high res"
2020- AHN4: 60ppm2
AHN3: 2014-2019 ref. 20ppm2
AHN2: 2008-2012 8-20ppm2
AHN1: 1996-2004 1ppm2
1960s
First lasers
Moon ranging
1990s
$1M+/system
Data 1x visit
2000s
1st national map
"experts only" until 2014


LiDAR in 2019 (Commercial acceleration)
(Spaceape 2006)
Velodyne
Yuri Shendryk (CSIRO)
2005-2015 DARPA & SV VC
Cost: >$75k in 2011, <$20k 2016
• SWaP limit uses (e.g. no UAVs)
• Computers limit real-time use
2017- New commercial lidars
Cost: $10k-$4k in 2018, <$1000 2019
+ 6000ppm unprecedented resol./cost
- Still need dedicated scan unit/team
+ This ensures good surveys/processing
Video


Copernicus 5-P (precursor)
Science & Environment
New Sentinel satellite tracks dirty air
By Jonathan Amos
BBC Science Correspondent
Nitrogen dioxide is concentrated over cities and industrial areas
It's been working less than a month but already the UK-Dutch-built Sentinel-5P satellite is returning spectacular new views of Earth's atmosphere.
Sentinel tracks ships' dirty emissions from orbit
By Jonathan Amos
BBC Science Correspondent
Shipping lane: The Tropomi instrument detects a suite of gases including nitrogen dioxide
The new EU satellite tasked with tracking dirty air has demonstrated how it will become a powerful tool to monitor emissions from shipping.
Sentinel-5P was launched in October last year and this week completed its in-orbit commissioning phase.

Monitoring soil moisture

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles

Bernhard Bauer-Marschallinger, *Member, IEEE*, Vahid Freeman, Senmao Cao, Christoph Paulik, Stefan Schaufler, Tobias Stachl, Sara Modanesi, Christian Massari, Luca Ciabatta, Luca Brocca, and Wolfgang Wagner, *Senior Member, IEEE*

Abstract—**Soil moisture is a key environmental variable, important to, e.g., farmers, meteorologists, and disaster management units. Here, we present a method to retrieve surface soil moisture (SSM) from the Sentinel-1 (S-1) satellites, which carry C-band Synthetic Aperture Radar (CSAR) sensors that provide the richest freely available SAR data source so far, unprecedented in accuracy and coverage. Our SSM retrieval method, adapting well-established change detection algorithms, builds the first globally deployable soil moisture observation data set with 1-km resolution. This paper provides an algorithm formulation to be operated in data cube architectures and high-performance computing environments. It includes the novel dynamic Gaussian upscaling method for spatial upscaling of SAR imagery, harnessing its field-scale information and successfully mitigating effects from the SAR's high signal complexity. Also, a new regression-based approach for estimating the radar slope is defined, coping with Sentinel-1's inhomogeneity in spatial coverage. We employ the S-1 SSM algorithm on a 3-year S-1 data cube over Italy, obtaining a consistent set of model parameters and product masks, unperturbed by coverage discontinuities. An evaluation of therefrom generated S-1 SSM data, involving a 1-km soil water balance model over Umbria, yields high agreement over plains and agricultural areas, with low agreement over forests and strong topography. While positive biases during the growing season are detected, the excellent capability to capture small-scale soil moisture changes as from rainfall or irrigation is evident. The S-1 SSM is currently in preparation toward operational product dissemination in the Copernicus Global Land Service.**

I. Introduction

THE Sentinel-1 satellites have been scanning Earth's surface using high-resolution radar sensors since 2014 with unprecedented spatiotemporal coverage. They carry a Synthetic Aperture Radar (SAR) system and deliver information on surface properties independent of daylight and cloud cover, operating in C-band (CSAR, at 5.405 GHz). It is a mission of the European earth observation program *Copernicus* with two identical spacecrafts, Sentinel-1A (S-1A) launched in April 2014 and Sentinel-1B (S-1B) in April 2016. Together with the multispectral sensors of the Copernicus constellation on-board Sentinel-2 [1] and Sentinel-3 [2], the Sentinel-1 SAR mission [3] is a pillar of current and future earth observation.

High expectations have been raised by the Sentinel-1 mission. The two Sentinel-1 satellites share the same orbit 180° apart and follow a strict acquisition scenario with a 12-day repeat cycle, featuring stable viewing angles, and thus predictable viewing geometry and spatial coverage. This setup enables via SAR interferometry the detection of millimeter-scale vertical surface deformations within 6 days. In its main sensing mode over land, the Interferometric Wide (IW) swath mode, the Sentinel-1 SAR offers a ground range detected (GRD) resolution of 20 m × 22 m at revisit frequency of 1.5–4 days over Europe (sensed from different orbits).

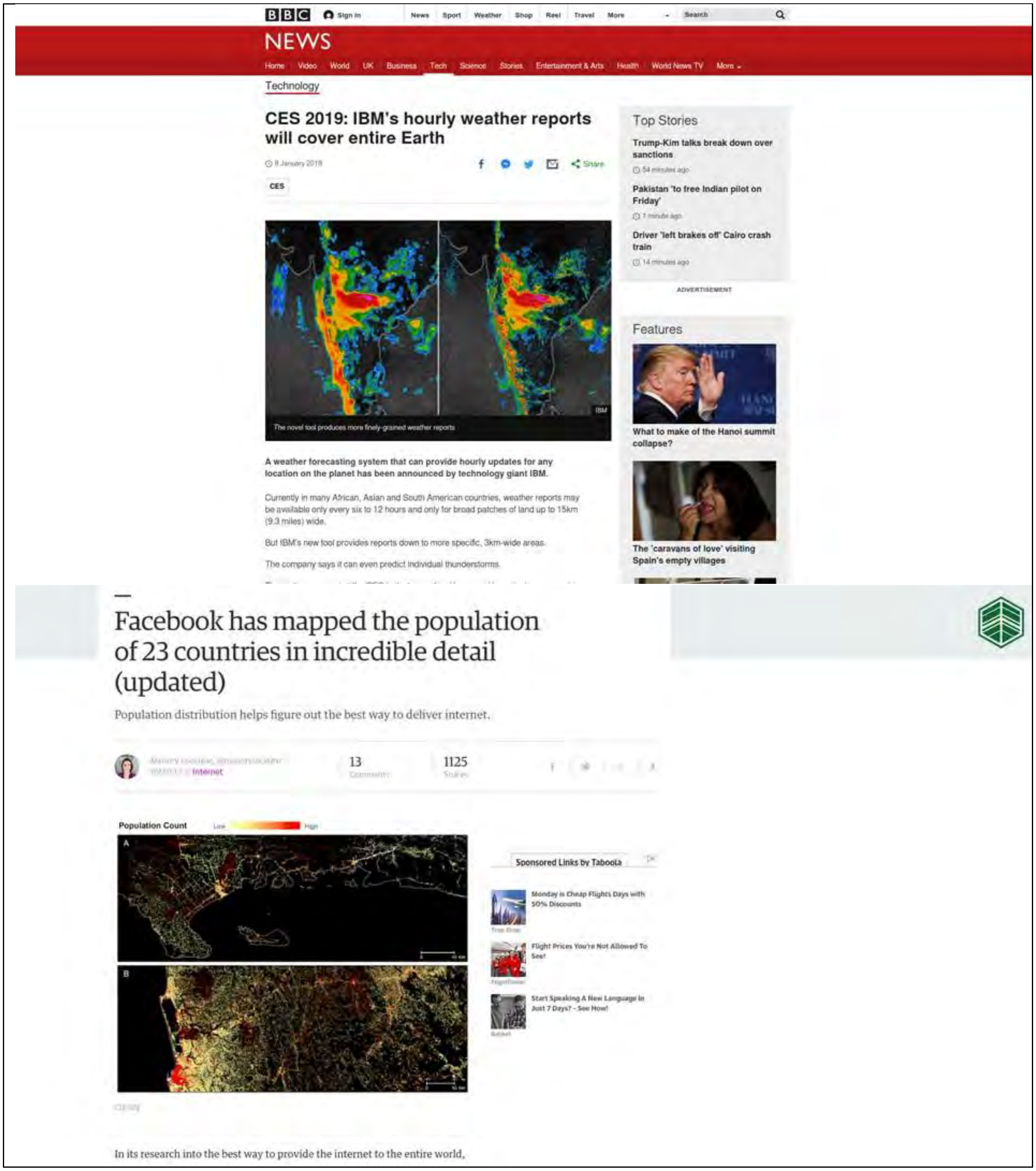
untitled

Radar Sounding Through the Earth's Ionosphere at 45 MHz

Anthony Freeman, *Fellow, IEEE*, Xiaoqing Pi, and Essam Heggy

Abstract—**Radar sounding from aircraft or ground-coupled radars has long provided scientists with a powerful technique to sound through ice layers to retrieve local depth and layering structure. More recently, it has been used to detect shallow aquifers in warm, dry, and desert regions. At Mars, a long-wavelength radar sounding from low orbit altitudes has produced global maps that reveal the presence of ice layering at all latitudes and glacial deposits on the flanks of volcanoes. Until now, sounding from the earth orbit at wavelengths long enough to penetrate ice sheets and arid sand was thought to be infeasible, because of the electromagnetic properties of the ionosphere. In this paper, we show that a radar sounding at frequencies as low as 45 MHz is, in fact, theoretically possible under viewing conditions that occur often enough to be practical. This conclusion opens up a previously unutilized portion of the electromagnetic spectrum for large-scale, spaceborne remote sensing of subsurface features on the earth.**

Index Terms—**Ionospheric propagation, radar sounding, very**


BBC Sign in News Sport Weather Shop Reel Travel More Search

NEWS

Home Video World UK Business Tech Science Stories Entertainment & Arts Health World News TV More

Technology

CES 2019: IBM's hourly weather reports will cover entire Earth

8 January 2019

CES

The novel tool produces more finely-grained weather reports

A weather forecasting system that can provide hourly updates for any location on the planet has been announced by technology giant IBM.

Currently in many African, Asian and South American countries, weather reports may be available only every six to 12 hours and only for broad patches of land up to 15km (9.3 miles) wide.

But IBM's new tool provides reports down to more specific, 3km-wide areas.

The company says it can even predict individual thunderstorms.

Top Stories

Trump-Kim talks break down over sanctions

Pakistan 'to free Indian pilot on Friday'

Driver 'left brakes off' Cairo crash train

ADVERTISEMENT

Features

What to make of the Hanoi summit collapse?

The 'caravans of love' visiting Spain's empty villages

Facebook has mapped the population of 23 countries in incredible detail (updated)

Population distribution helps figure out the best way to deliver internet.

Internet

13 Comments

1125 Shares

Population Count Low High

A

B

Sponsored Links by Taboola

Monday is Cheap Flights Days with 50% Discounts

Flight Prices You're Not Allowed To See!

Start Speaking A New Language in Just 7 Days? - See How!

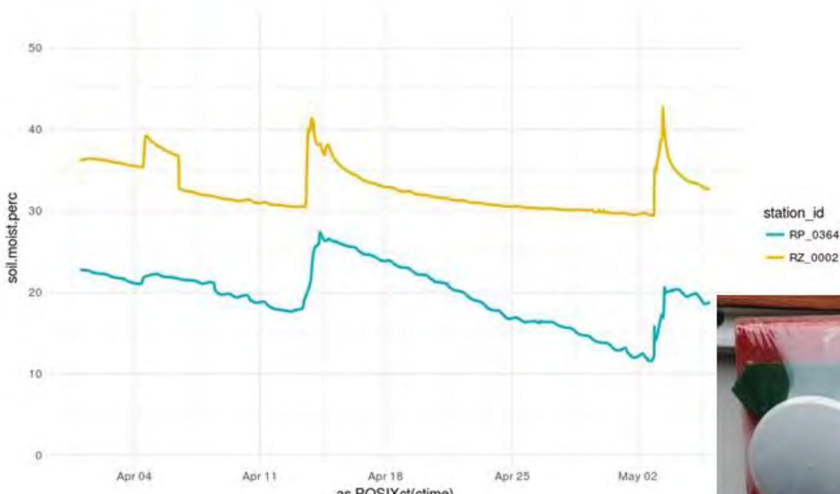
In its research into the best way to provide the internet to the entire world,

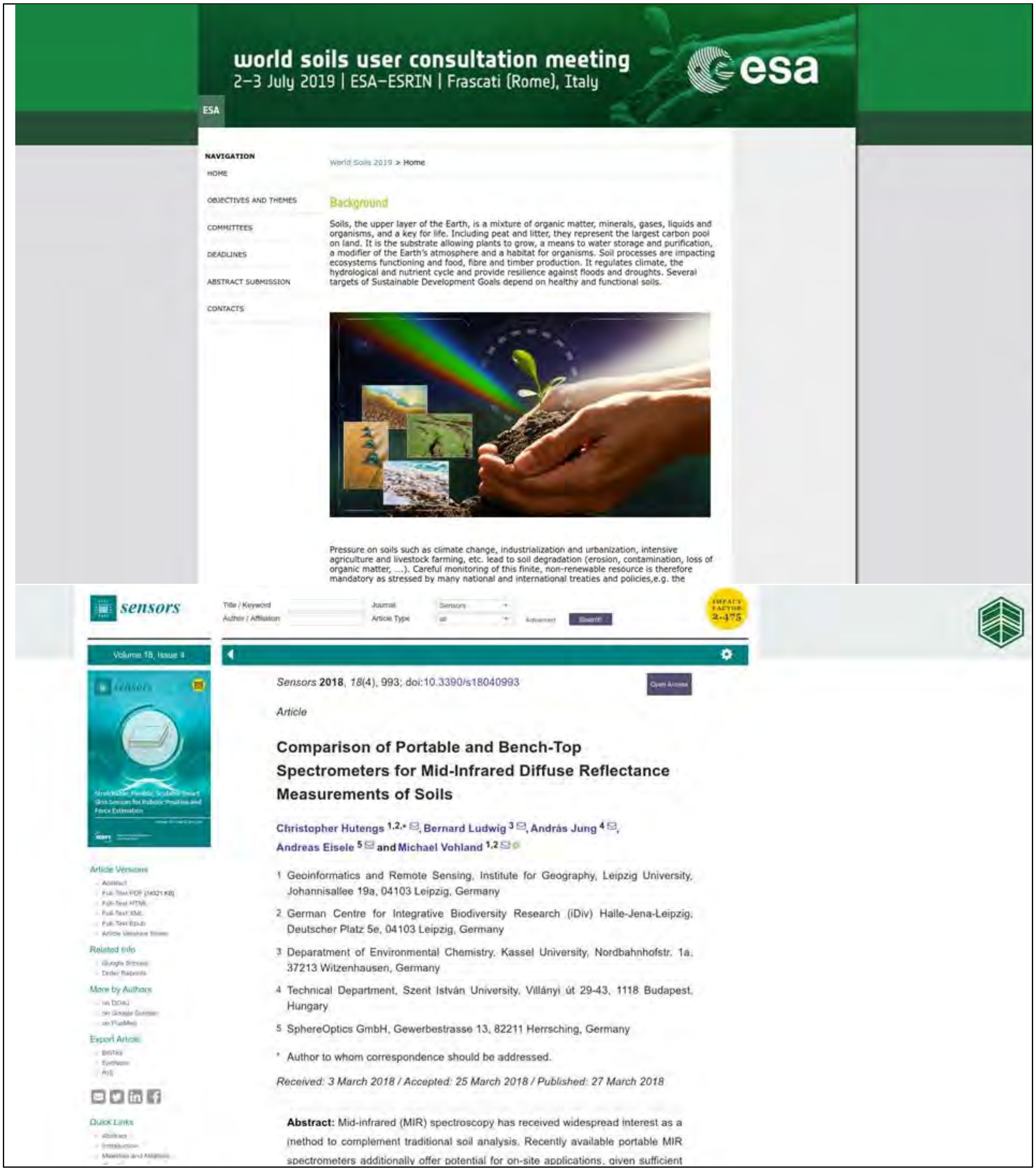
Machine Learning in Open Source

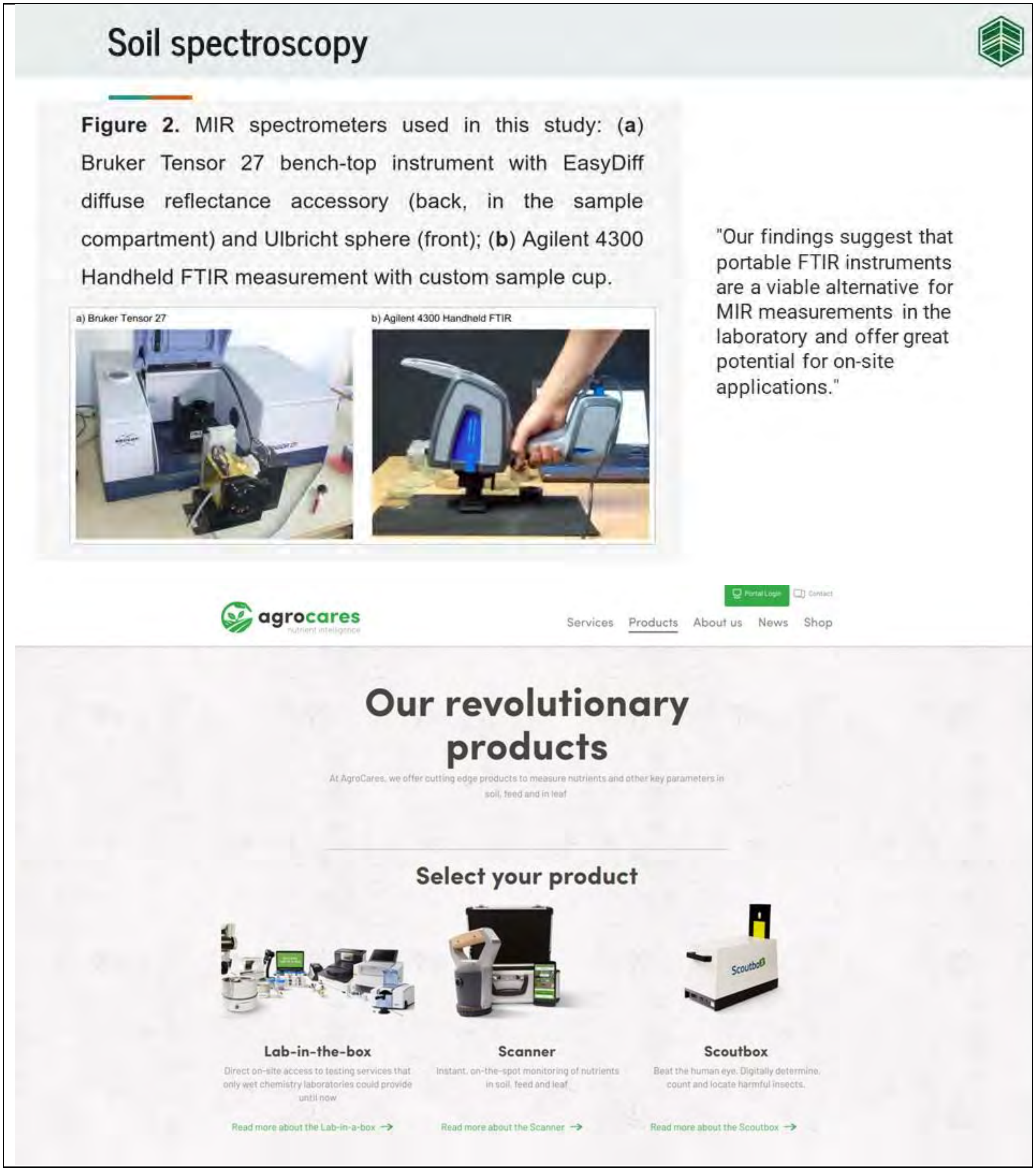
Some popular libs:

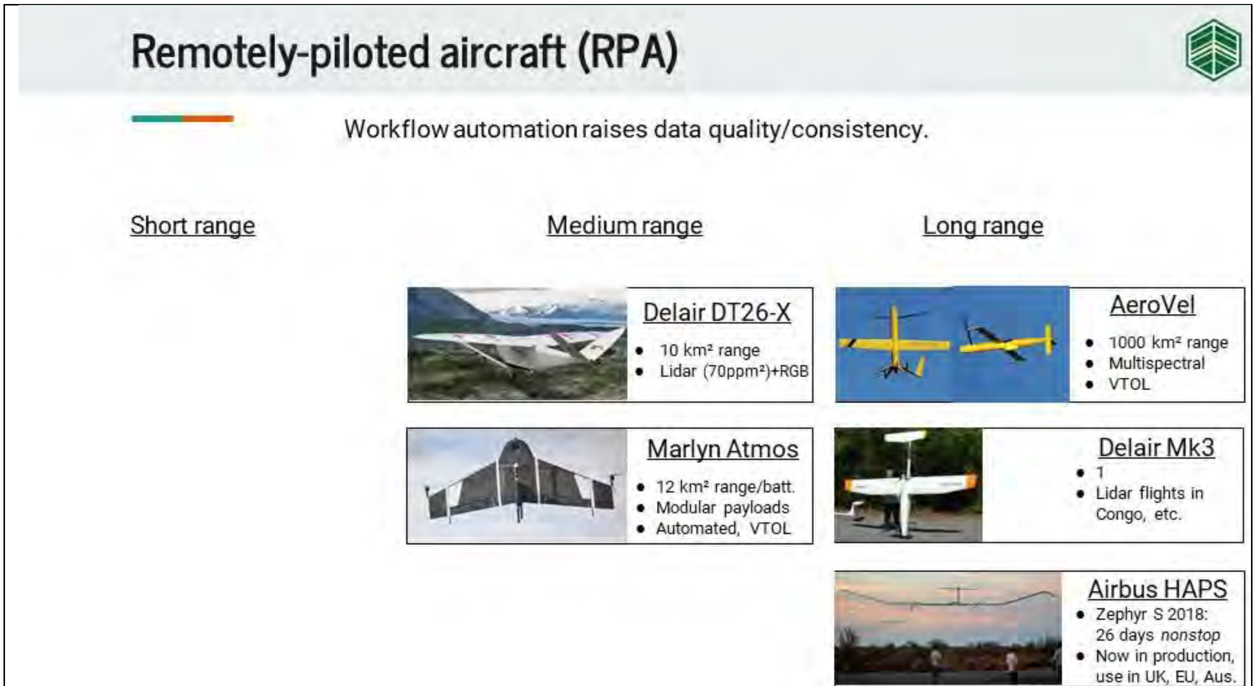
- **TensorFlow**
- **Keras**
- **PyTorch**
- **h2o**
- **MCT**
- **MXNet**
- **ranger**
- **XGBoost**
- **liquidSVM**

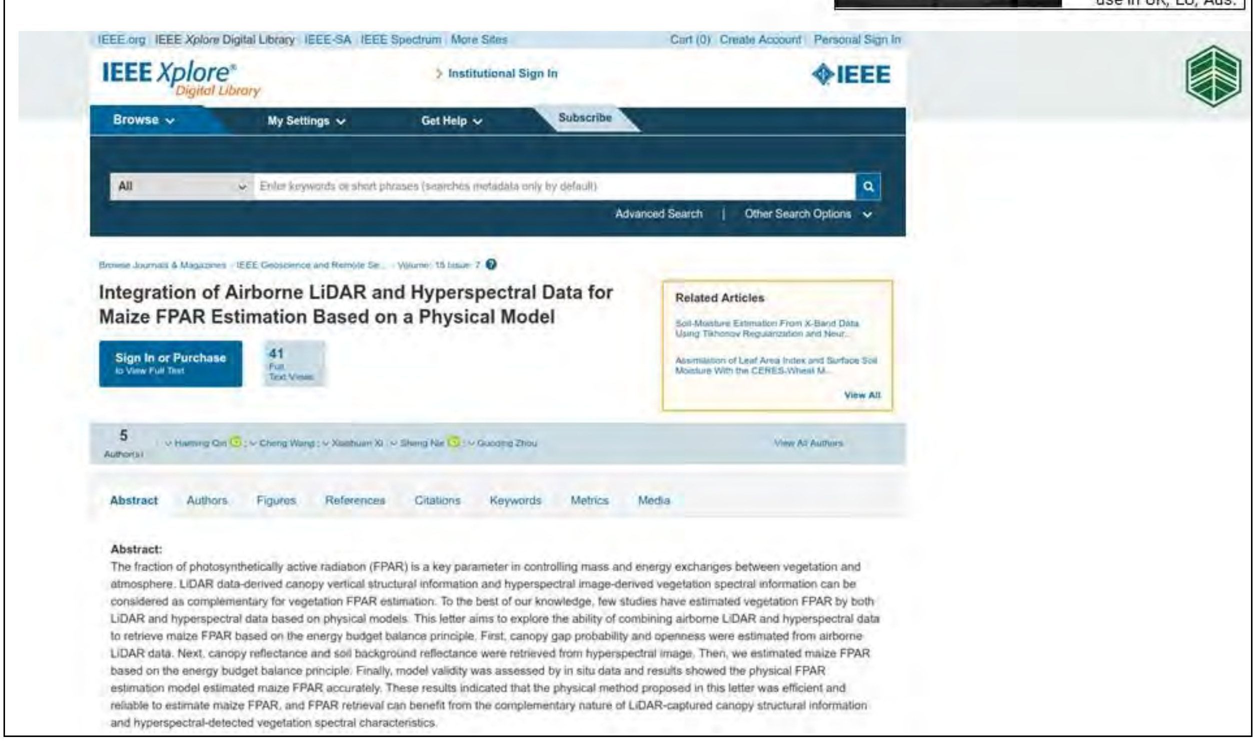
80 EUR microclimatic station




world soils user consultation meeting
2–3 July 2019 | ESA–ESRIN | Frascati (Rome), Italy
esa
ESA
NAVIGATION
HOME
OBJECTIVES AND THEMES
COMMITTEES
DEADLINES
ABSTRACT SUBMISSION
CONTACTS
World Soils 2019 > Home
Background
Soils, the upper layer of the Earth, is a mixture of organic matter, minerals, gases, liquids and organisms, and a key for life. Including peat and litter, they represent the largest carbon pool on land. It is the substrate allowing plants to grow, a means to water storage and purification, a modifier of the Earth's atmosphere and a habitat for organisms. Soil processes are impacting ecosystems functioning and food, fibre and timber production. It regulates climate, the hydrological and nutrient cycle and provide resilience against floods and droughts. Several targets of Sustainable Development Goals depend on healthy and functional soils.
Pressure on soils such as climate change, industrialization and urbanization, intensive agriculture and livestock farming, etc. lead to soil degradation (erosion, contamination, loss of organic matter, ...). Careful monitoring of this finite, non-renewable resource is therefore mandatory as stressed by many national and international treaties and policies,e.g. the
sensors
Title / Keyword
Author / Affiliation
Journal
Article Type
Search
IMPACT FACTOR 2.475
Volume 18, Issue 4
Sensors 2018, 18(4), 993; doi:10.3390/s18040993
Open Access
Article
Comparison of Portable and Bench-Top Spectrometers for Mid-Infrared Diffuse Reflectance Measurements of Soils
Christopher Hutengs 1,2,*, Bernard Ludwig 3, András Jung 4, Andreas Eisele 5 and Michael Vohland 1,2
1 Geoinformatics and Remote Sensing, Institute for Geography, Leipzig University, Johannisallee 19a, 04103 Leipzig, Germany
2 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
3 Deparatment of Environmental Chemistry, Kassel University, Nordbahnhofstr. 1a, 37213 Witzenhausen, Germany
4 Technical Department, Szent István University, Villányi út 29-43, 1118 Budapest, Hungary
5 SphereOptics GmbH, Gewerbestrasse 13, 82211 Herrsching, Germany
* Author to whom correspondence should be addressed.
Received: 3 March 2018 / Accepted: 25 March 2018 / Published: 27 March 2018
Article Versions
Related Info
More by Authors
Export Article
Quick Links
Abstract: Mid-infrared (MIR) spectroscopy has received widespread interest as a method to complement traditional soil analysis. Recently available portable MIR spectrometers additionally offer potential for on-site applications, given sufficient


Soil spectroscopy
Figure 2. MIR spectrometers used in this study: (**a**) Bruker Tensor 27 bench-top instrument with EasyDiff diffuse reflectance accessory (back, in the sample compartment) and Ulbricht sphere (front); (**b**) Agilent 4300 Handheld FTIR measurement with custom sample cup.
a) Bruker Tensor 27
b) Agilent 4300 Handheld FTIR
"Our findings suggest that portable FTIR instruments are a viable alternative for MIR measurements in the laboratory and offer great potential for on-site applications."
agrocares
Services
Products
About us
News
Shop
Our revolutionary products
At AgroCares, we offer cutting edge products to measure nutrients and other key parameters in soil, feed and in leaf
Select your product
Lab-in-the-box
Direct on-site access to testing services that only wet chemistry laboratories could provide until now
Read more about the Lab-in-a-box →
Scanner
Instant, on-the-spot monitoring of nutrients in soil, feed and leaf
Read more about the Scanner →
Scoutbox
Beat the human eye. Digitally determine, count and locate harmful insects.
Read more about the Scoutbox →





Browse Journals & Magazines > IEEE Geoscience and Remote Se... > Volume: 15 Issue: 7

Integration of Airborne LiDAR and Hyperspectral Data for Maize FPAR Estimation Based on a Physical Model

Sign In or Purchase to View Full Text

41 Full Text Views

Related Articles

Soil-Moisture Estimation From X-Band Data Using Tikhonov Regularization and Neur...

Assimilation of Leaf Area Index and Surface Soil Moisture With the CERES-Wheat M...

View All

5 Author(s) Haiming Qin; Cheng Wang; Xiaohuan Xi; Sheng Nie; Guoqing Zhou View All Authors

Abstract | Authors | Figures | References | Citations | Keywords | Metrics | Media

Abstract:
The fraction of photosynthetically active radiation (FPAR) is a key parameter in controlling mass and energy exchanges between vegetation and atmosphere. LiDAR data-derived canopy vertical structural information and hyperspectral image-derived vegetation spectral information can be considered as complementary for vegetation FPAR estimation. To the best of our knowledge, few studies have estimated vegetation FPAR by both LiDAR and hyperspectral data based on physical models. This letter aims to explore the ability of combining airborne LiDAR and hyperspectral data to retrieve maize FPAR based on the energy budget balance principle. First, canopy gap probability and openness were estimated from airborne LiDAR data. Next, canopy reflectance and soil background reflectance were retrieved from hyperspectral image. Then, we estimated maize FPAR based on the energy budget balance principle. Finally, model validity was assessed by in situ data and results showed the physical FPAR estimation model estimated maize FPAR accurately. These results indicated that the physical method proposed in this letter was efficient and reliable to estimate maize FPAR, and FPAR retrieval can benefit from the complementary nature of LiDAR-captured canopy structural information and hyperspectral-detected vegetation spectral characteristics.

Innovation in information processing as well

Switch to data processing part of talk - divider here?

Eric van Rees Follow
Writer and editor. Interested in all things geospatial.
Sep 2 · 5 min read

How Geospatial Technology Changed over Time

Geospatial technology has changed significantly over the years. It started in a scientific environment and has gotten a wider adoption over time by others. From a technological perspective, geospatial technology went from the desktop to server and now incorporates everything from desktop, mobile, server, the cloud and now the blockchain.

Desktop GIS

cloud GIS, APIs and open source

Cloud computing services

UAV, LIDAR...

Data Science revolution

Blockchain, the next frontier?
(https://demo.soar.earth/#)

Immutable Data Storage

ltdp

In 2006, the European Space Agency (ESA) initiated a coordination action to share among all the European (and **Canadian**) stakeholders a **common approach to the long term preservation of Earth Observation space data**.

In 2007, [they agreed a] Long Term Data Preservation strategy targeting the preservation of all European (incl. **Canada**) EO space data for an **unlimited time-span ensuring and facilitating their accessibility and usability [in a] cooperative and harmonized collective approach**.

Source (ESA)

Current research: Blockchain-based "global, permanent hard drive"

Distributed data storage: uncensorable "pay once, store *forever*" capability.

Contributions possible by anyone online - citizens/scientists, companies, governments.

Data archived 'append-only' in *real-time*, then analysis of this and publish methods released with encryption key!

Promising technologies summarized

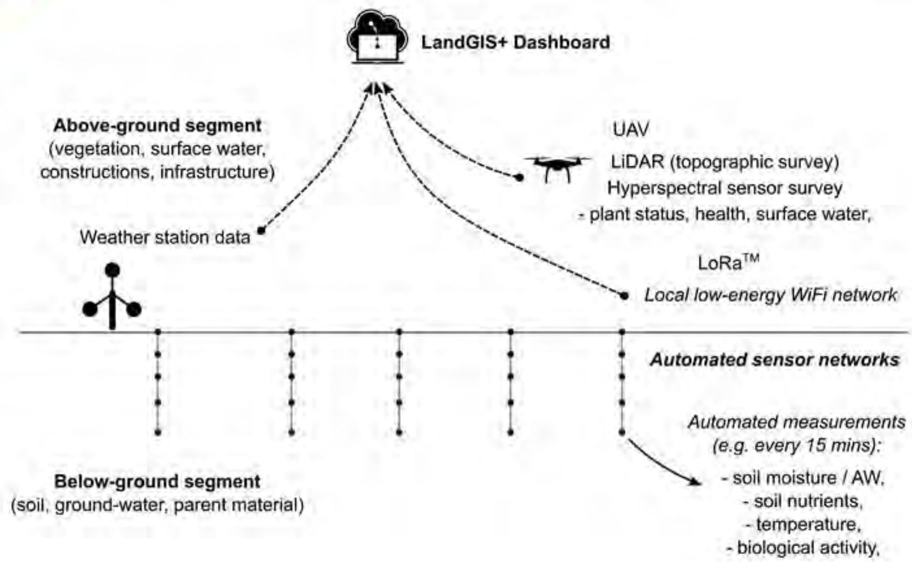
Monitoring

- **Public-access, high-res Space**
  Sentinels, ALOS, Landsat 8, TANDEM ...
- **Commercial "NewSpace"**
  Very fast system design/launch
  Data- or Intelligence-as-a-Service
- **UAV fleets**
  With lidar + radar + hyperspec,
  Data- or Intelligence-as-a-Service
- **Integrated sensor networks**
  (meteo, soil, acoustic ...)

Processing

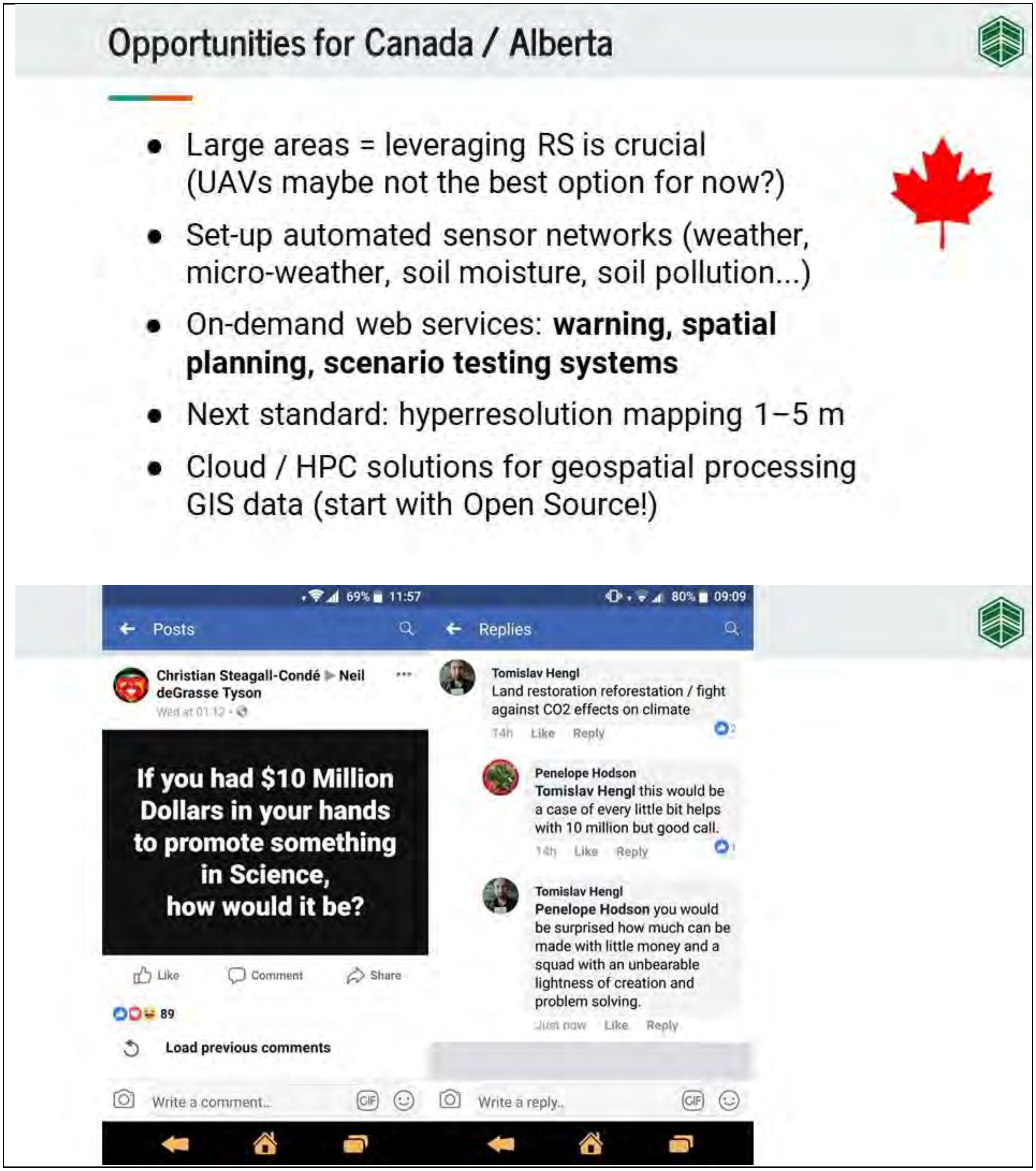
- **Online geospatial data workflows**
  R-spatial (plumber), GeoTrellis DB, Cloud Optimized GeoTIFF
- **Scalable compute infrastructure**
  Diverse data storage/processing solutions as-a-Service.
- **Permanent data storage**
  Digital data instantly accessible, unalterable, uncensorable.

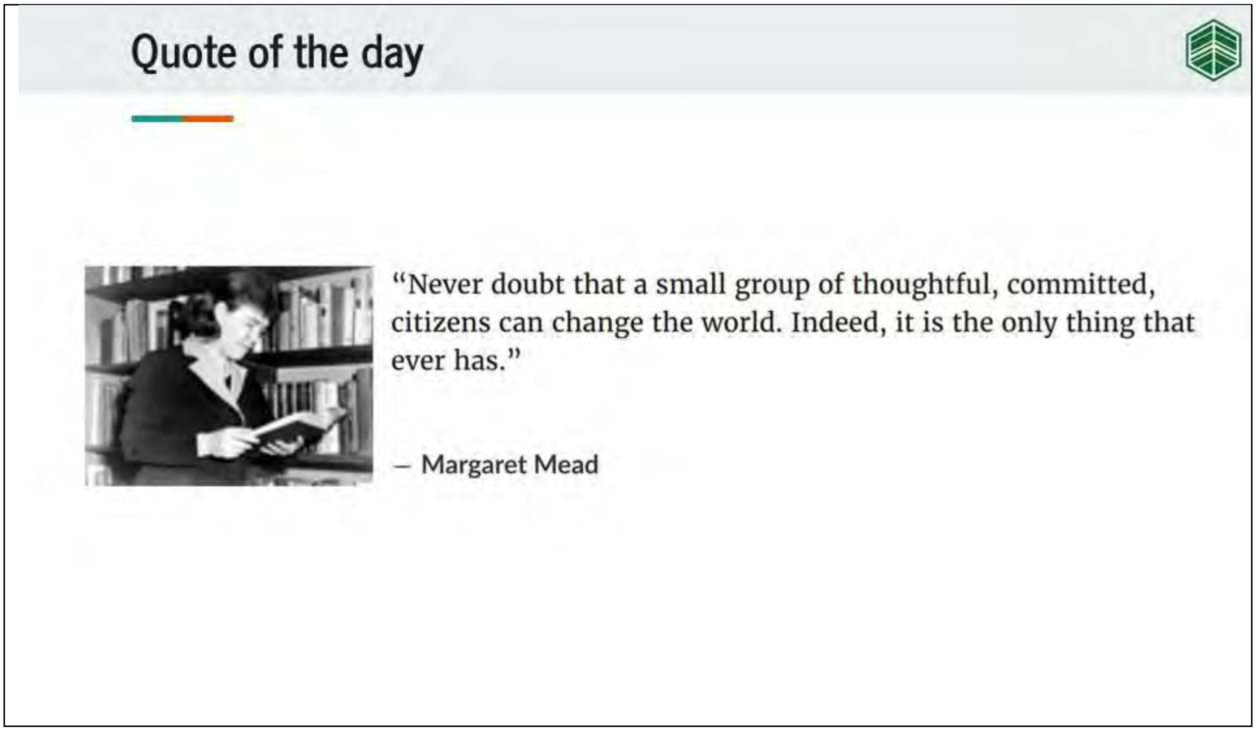
Vision: above and below ground tech for soil monitoring



Summary points I

1. More data available, but that does NOT mean easier mapping procedures (in fact we're **in danger of being overwhelmed**).
2. Very promising new technologies for soil science: "IoT" sensor nets, LiDAR, Sentinel-1,2,5, ALOS
3. Single sensor ↦ **multi-sensor fusion**;
   Single mission ↦ **multidisciplinary projects**;
   Single model ↦ **Ensemble models**;
4. Machine Learning is increasingly available! use it!


Opportunities for Canada / Alberta
Large areas = leveraging RS is crucial (UAVs maybe not the best option for now?)
Set-up automated sensor networks (weather, micro-weather, soil moisture, soil pollution...)
On-demand web services: **warning, spatial planning, scenario testing systems**
Next standard: hyperresolution mapping 1–5 m
Cloud / HPC solutions for geospatial processing GIS data (start with Open Source!)
69% 11:57
Posts
Christian Steagall-Condé ▶ Neil deGrasse Tyson
Wed at 01:12
If you had $10 Million Dollars in your hands to promote something in Science, how would it be?
Like
Comment
Share
89
Load previous comments
Write a comment...
80% 09:09
Replies
Tomislav Hengl
Land restoration reforestation / fight against CO2 effects on climate
14h Like Reply
Penelope Hodson
Tomislav Hengl this would be a case of every little bit helps with 10 million but good call.
14h Like Reply
Tomislav Hengl
Penelope Hodson you would be surprised how much can be made with little money and a squad with an unbearable lightness of creation and problem solving.
Just now Like Reply
Write a reply...


Quote of the day
"Never doubt that a small group of thoughtful, committed, citizens can change the world. Indeed, it is the only thing that ever has."
— Margaret Mead