Food and Agriculture Organization of the United Nations

3rd Workshop of the International Network of Black Soils

Status and methodology of the GBSmap development

Yusuf Yigini & Marcos Angelini

13-14 December 2021

BLACK SOILS INTERNATIONAL NETWORK

GLOBAL SOIL PARTNERSHIP

Overview

GSP - Soil Data and Information

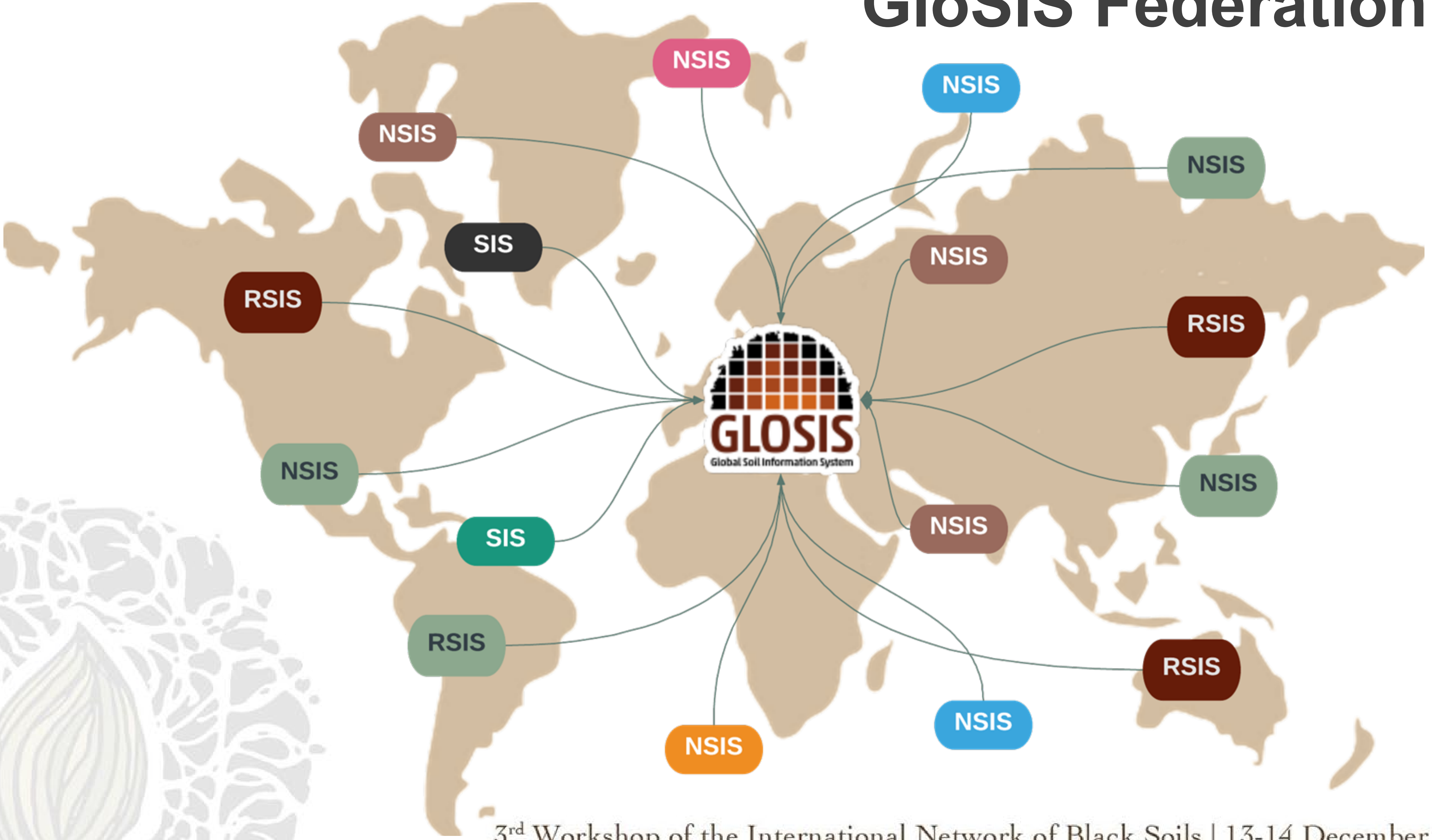
- GloSIS - Global Soil Information System
- GloSIS Global and Country Driven Global Datatsets
- GSOCmap, GSOCseq, GSASmap…

GBSmap

- Methodology
- Result
- Q&A
- Way forward
- Conclusion


GLOSIS
Global Soil Information System
INSII
INTERNATIONAL NETWORK
OF SOIL INFORMATION INSTITUTIONS

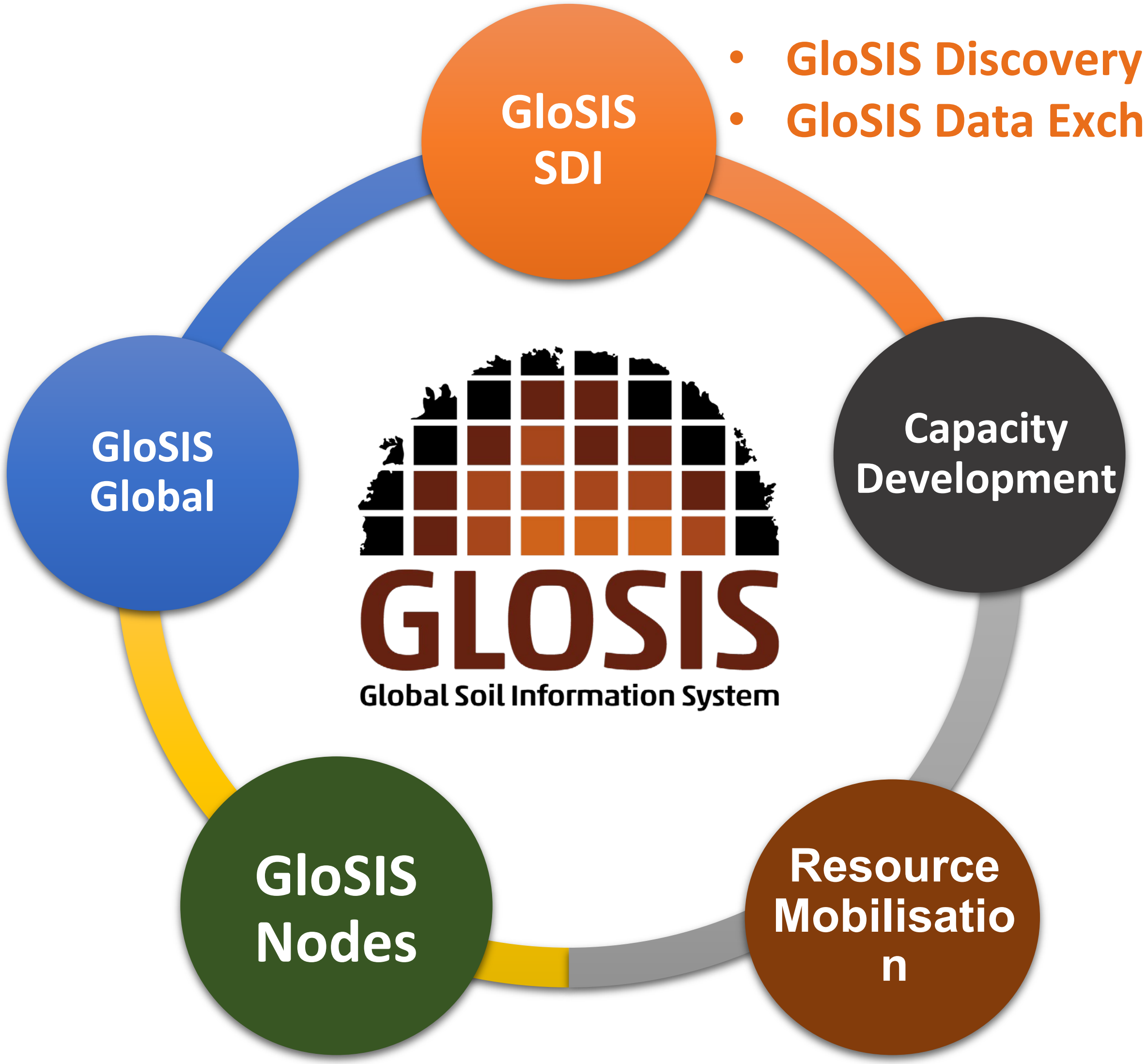
GloSIS Federation



GLOBAL SOIL PARTNERSHIP

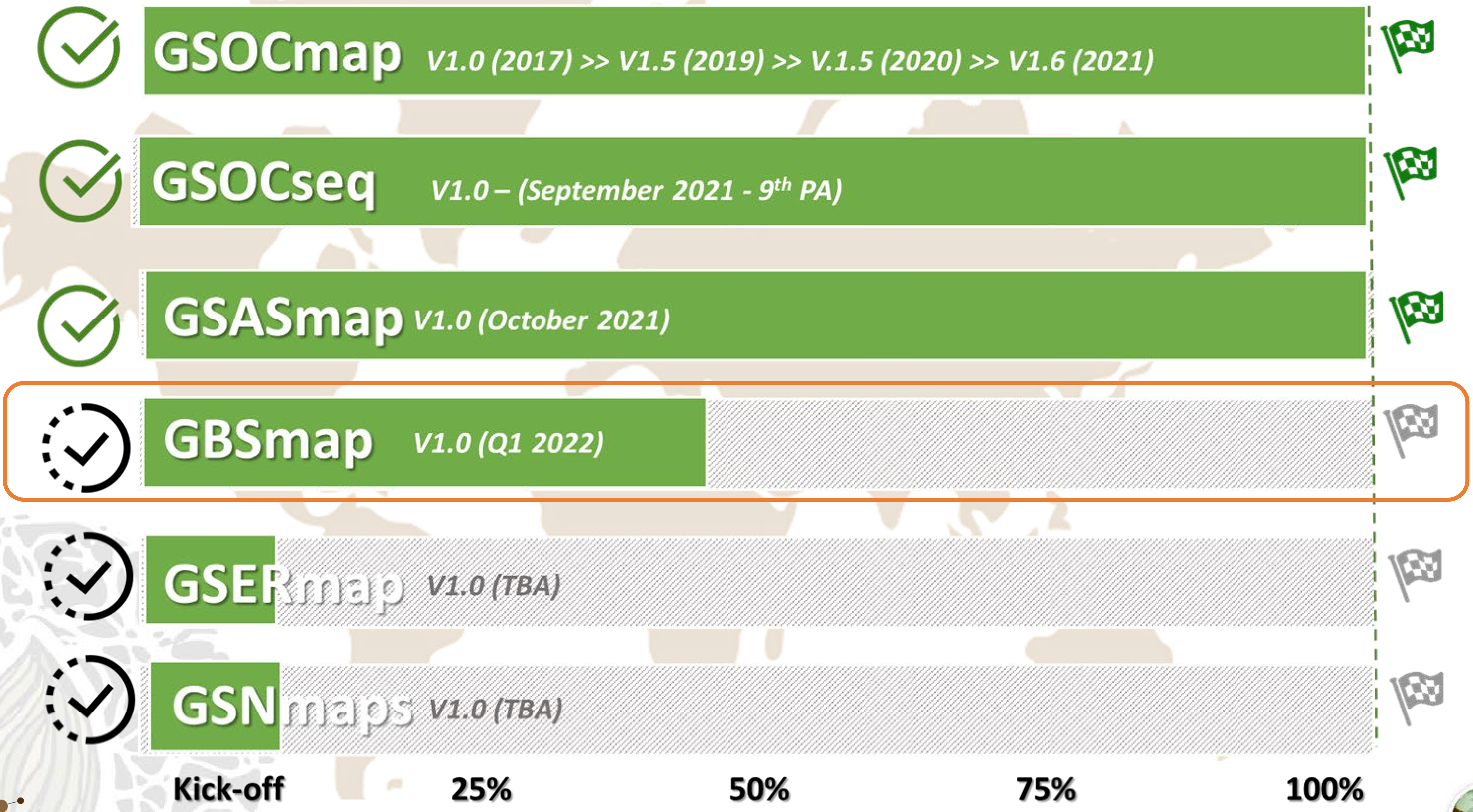
- GloSIS Discovery Hub
- GloSIS Data Exchange

- GSOCmap
- GSSmap
- GSOCseq
- GBSmap
- GSERmap
- ...



- Product Oriented Training
- DSM Training
- Product Guidelines
- Technical Manuals
- Tools/Software

- Regional, National, Institutional Soil Information Systems


GSOCmap V1.0 (2017) >> V1.5 (2019) >> V.1.5 (2020) >> V1.6 (2021)
GSOCseq V1.0 – (September 2021 - 9th PA)
GSASmap V1.0 (October 2021)
GBSmap V1.0 (Q1 2022)
GSERmap V1.0 (TBA)
GSNmaps V1.0 (TBA)
Kick-off
25%
50%
75%
100%

GSP Capacity Development Programme

140+ Countries

1100+ Experts

50+ Sessions

8 Regions

GSP Capacity Development Programme

GBSmap Training
9-11 December 2020
45 Participants
19 Countries

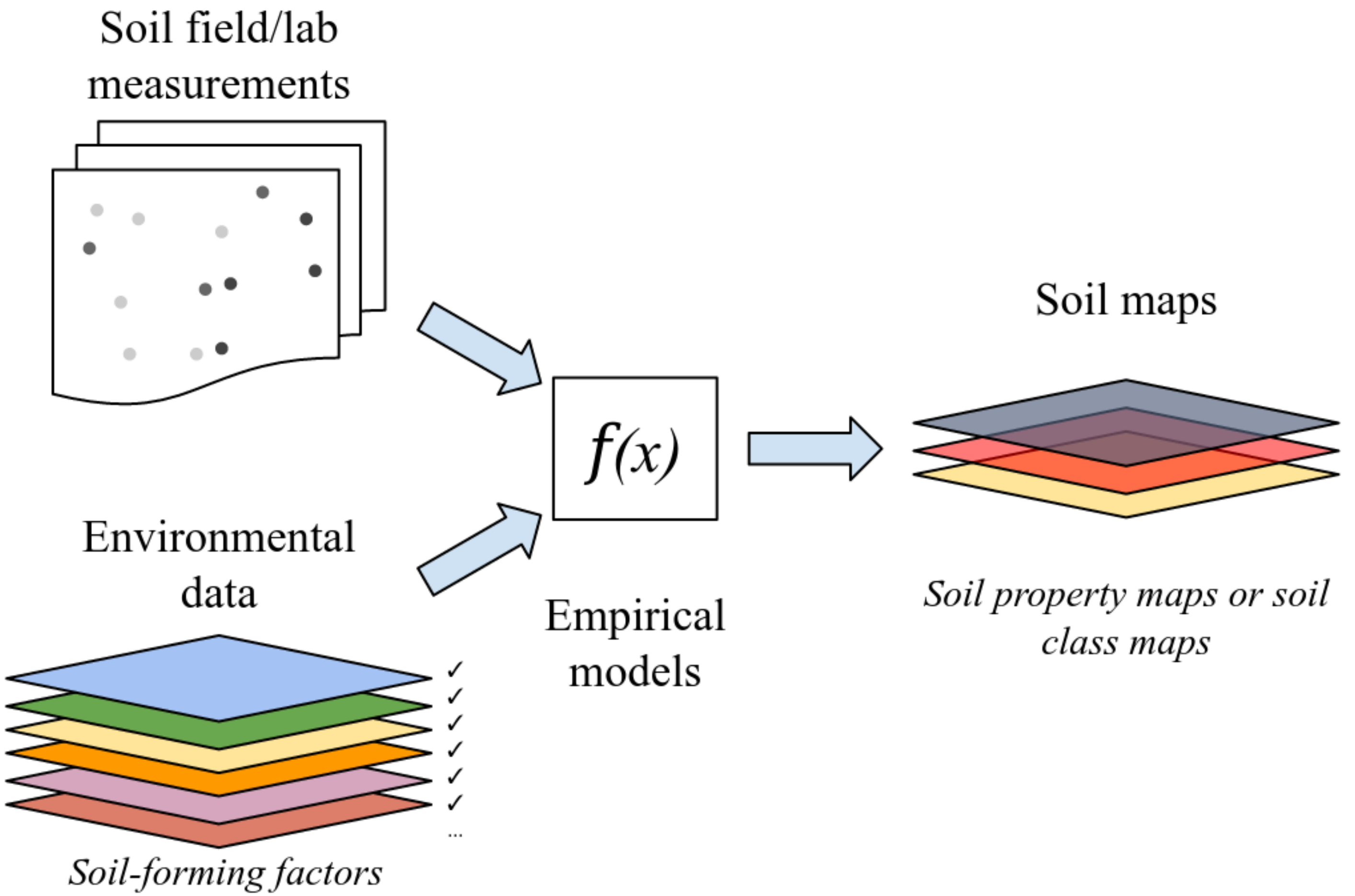
Country summision

14 countries contributed to the second chapter of this report

1. Argentina
2. Brazil
3. Bulgaria
4. Canada
5. Colombia
6. Japan
7. Mexico
8. Mongolia
9. Poland
10. Syrian Arab Republic
11. Slovakia
12. Ukraine
13. Uruguay
14. United States of America

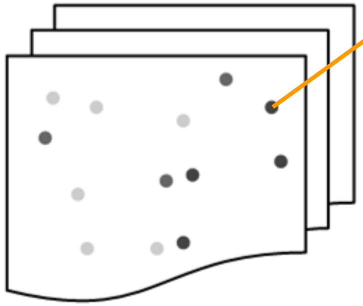
Methodology for the GBSmap

Digital soil mapping framework

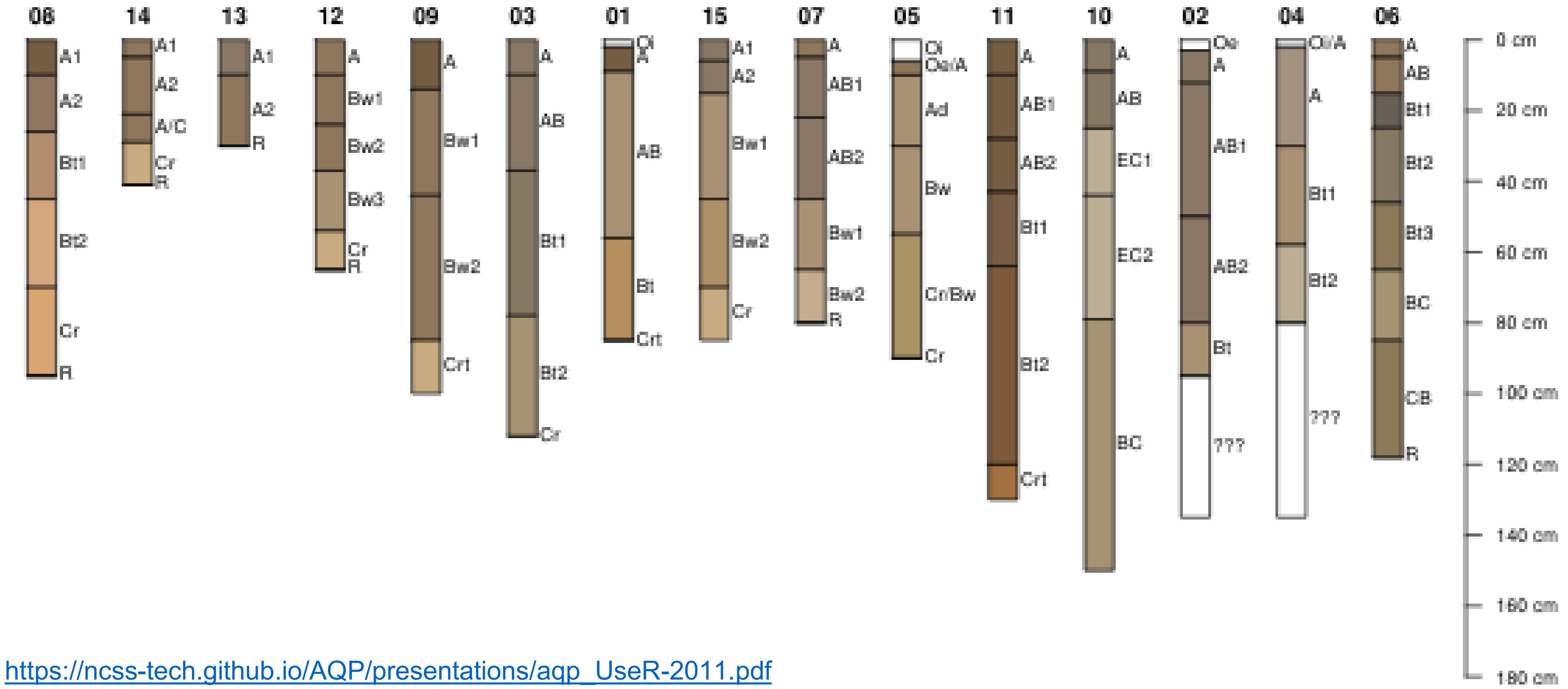


Soil data

Soil field/lab measurements

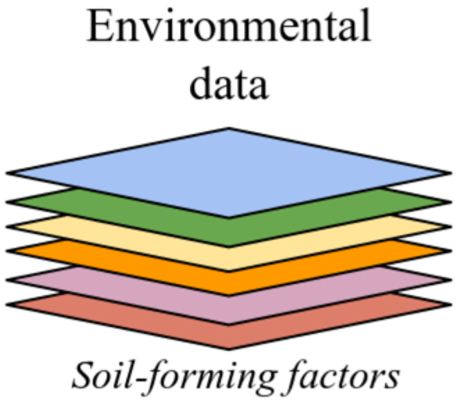


(latitude, longitude, value@soilDepth or class)



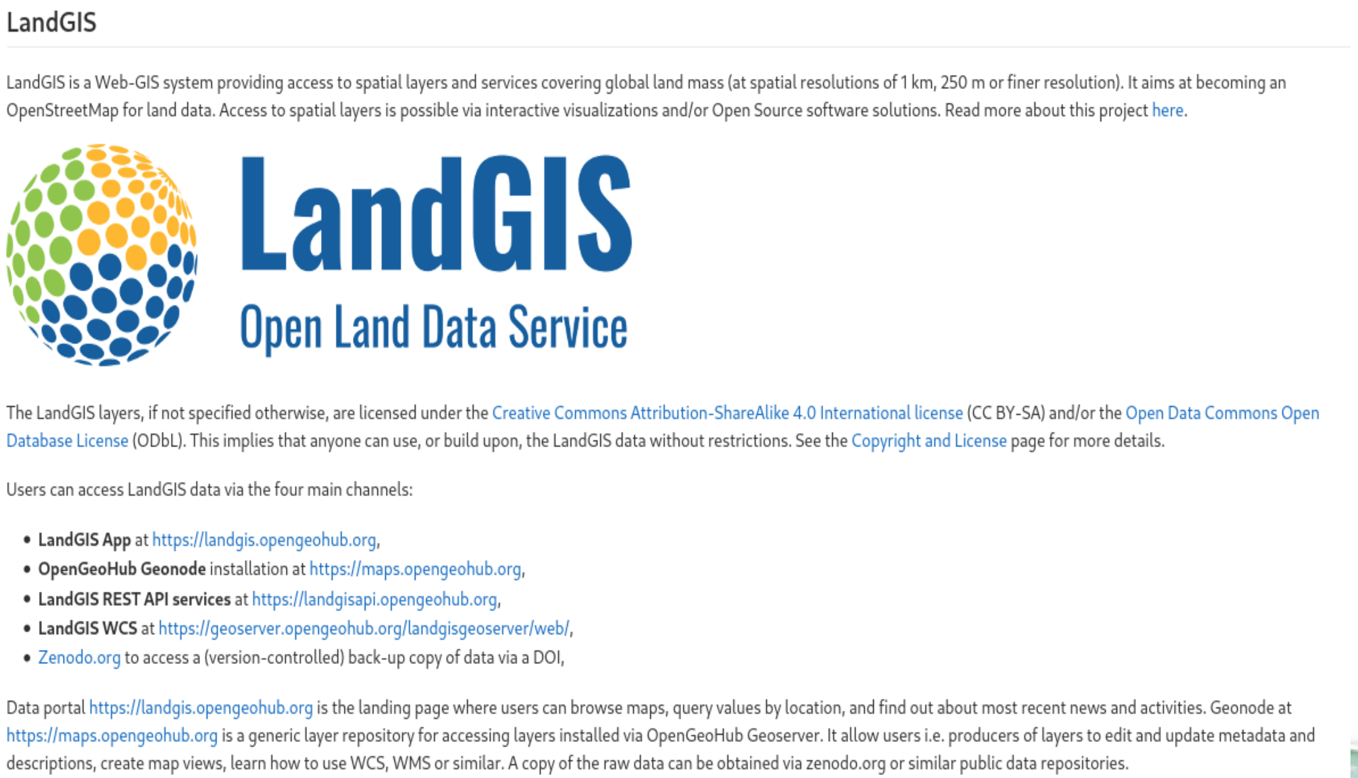
https://ncss-tech.github.io/AQP/presentations/aqp_UseR-2011.pdf

Environmental covariates



- Climate
- Relief and terrain attributes
- Geology
- Land use and land cover
- Legacy maps
- etc.

https://gitlab.com/openlandmap/global-layers



LandGIS

LandGIS is a Web-GIS system providing access to spatial layers and services covering global land mass (at spatial resolutions of 1 km, 250 m or finer resolution). It aims at becoming an OpenStreetMap for land data. Access to spatial layers is possible via interactive visualizations and/or Open Source software solutions. Read more about this project here.

LandGIS
Open Land Data Service

The LandGIS layers, if not specified otherwise, are licensed under the Creative Commons Attribution-ShareAlike 4.0 International license (CC BY-SA) and/or the Open Data Commons Open Database License (ODbL). This implies that anyone can use, or build upon, the LandGIS data without restrictions. See the Copyright and License page for more details.

Users can access LandGIS data via the four main channels:

- **LandGIS App** at https://landgis.opengeohub.org,
- **OpenGeoHub Geonode** installation at https://maps.opengeohub.org,
- **LandGIS REST API services** at https://landgisapi.opengeohub.org,
- **LandGIS WCS** at https://geoserver.opengeohub.org/landgisgeoserver/web/,
- Zenodo.org to access a (version-controlled) back-up copy of data via a DOI,

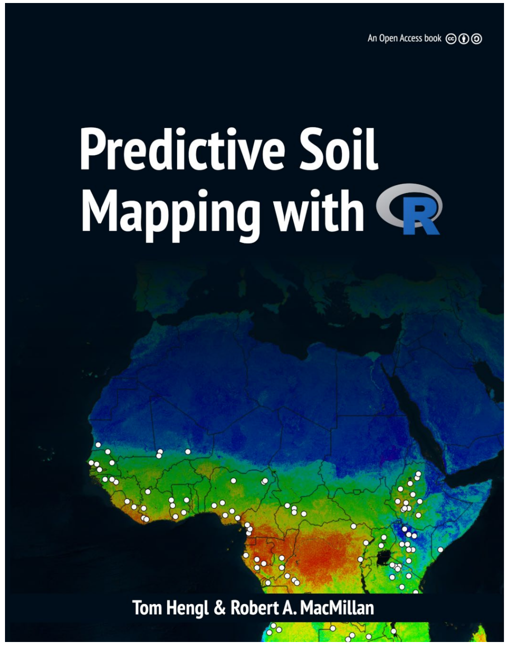
Data portal https://landgis.opengeohub.org is the landing page where users can browse maps, query values by location, and find out about most recent news and activities. Geonode at https://maps.opengeohub.org is a generic layer repository for accessing layers installed via OpenGeoHub Geoserver. It allow users i.e. producers of layers to edit and update metadata and descriptions, create map views, learn how to use WCS, WMS or similar. A copy of the raw data can be obtained via zenodo.org or similar public data repositories.

Empirical models

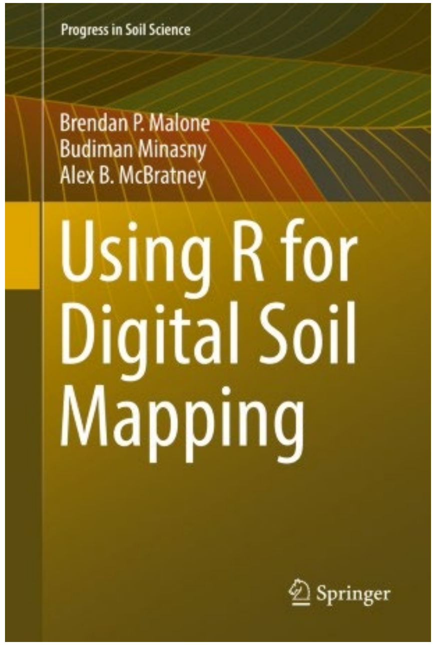
$f(x)$

- Multiple linear regression
- Multivariate regression
- Partial least square
- Decision trees
- Random forest
- Artificial neural networks
- Geostatistics (kriging)
- Regression-kriging

https://soilmapper.org/

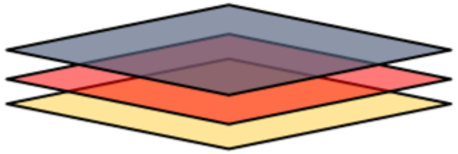


https://link.springer.com/book/10.1007%2F978-3-319-44327-0



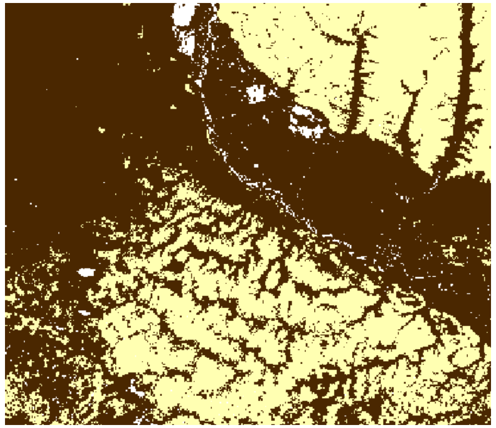
Resulting maps

Soil maps



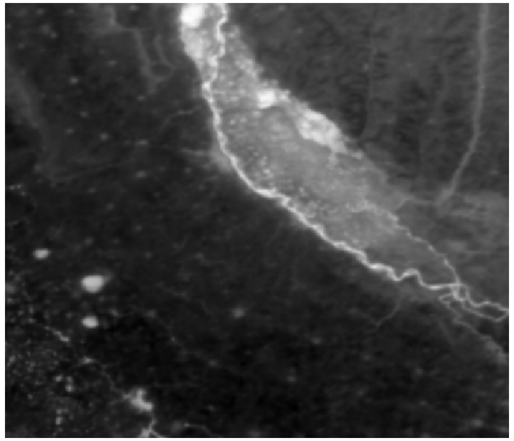
Soil property maps or soil class maps

Categorical map

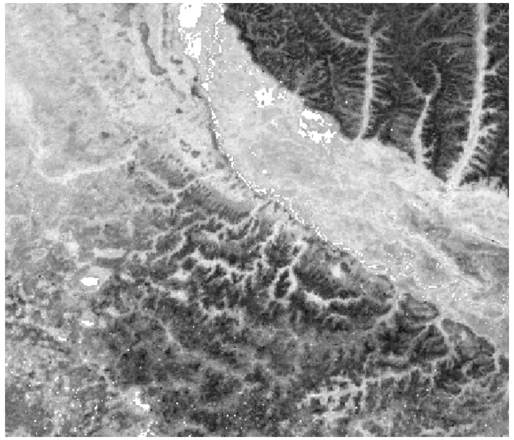


uncertainty map

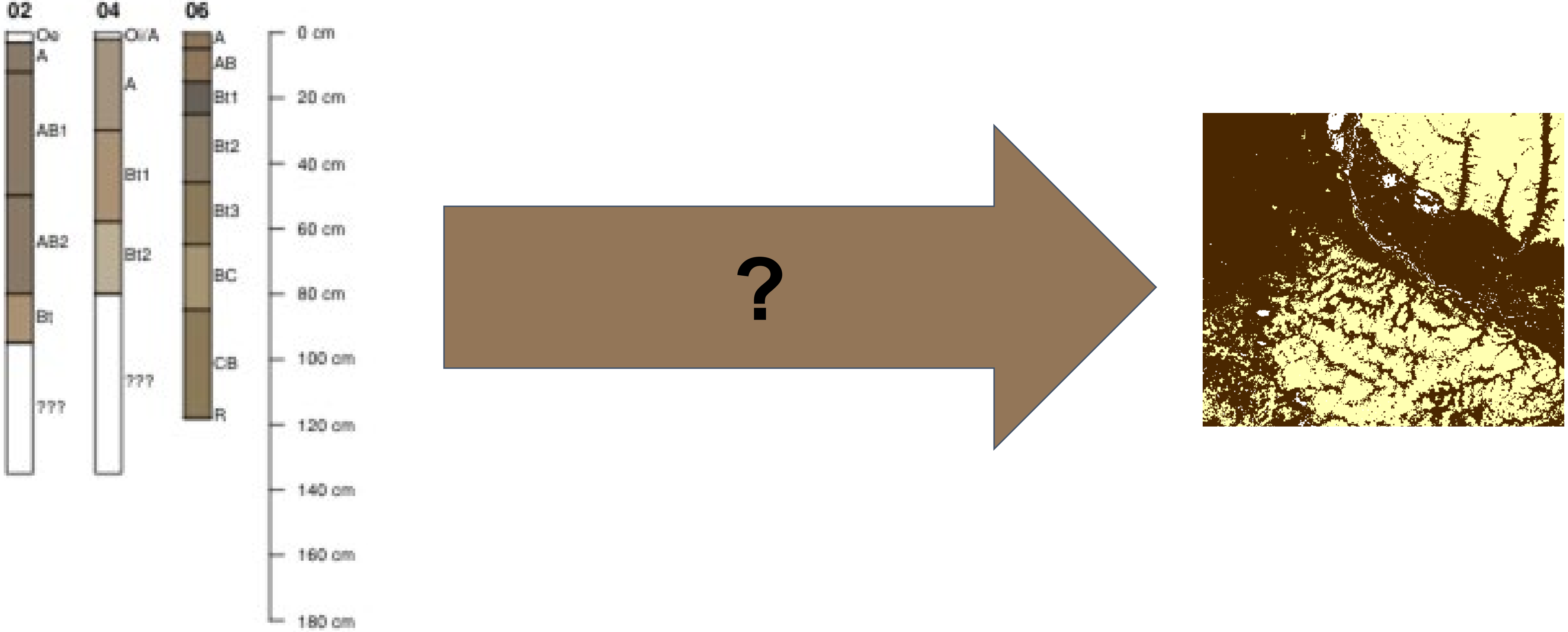
standard deviation map



probability map



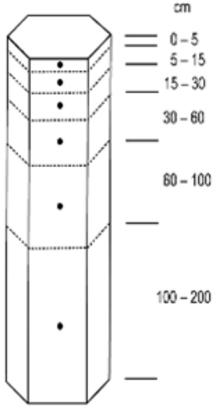
From soil profiles to Black Soil map



GLOBAL SOIL PARTNERSHIP

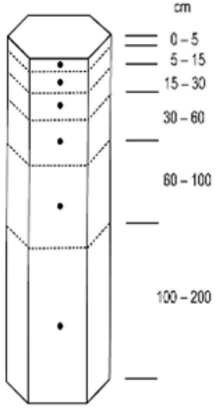
What is a black soil?

1st category BS

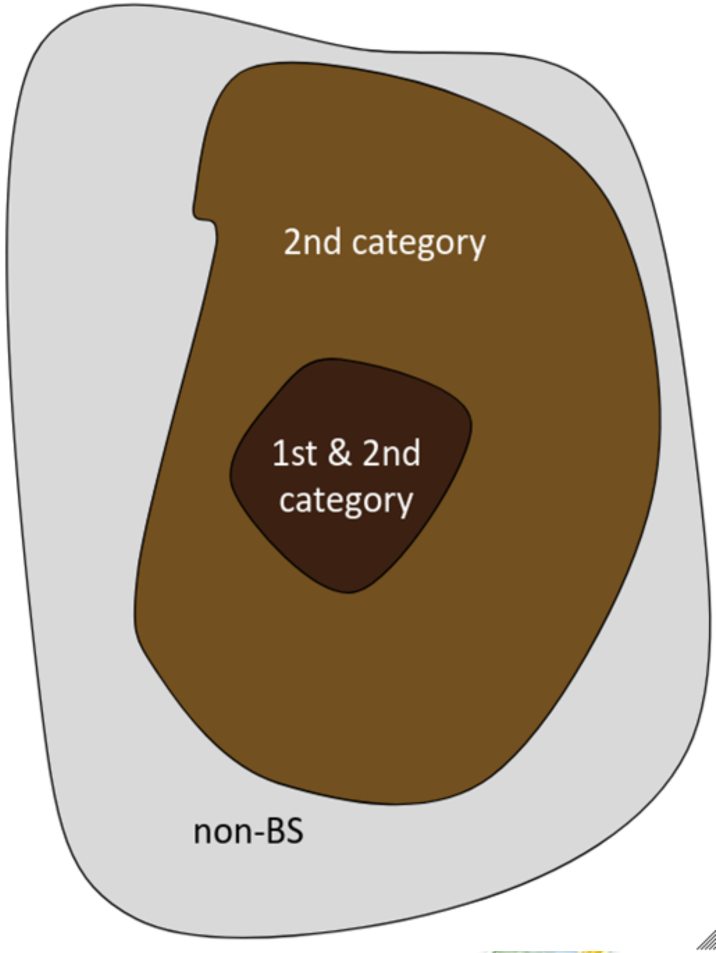


- Soil depth ≥ 25cm
- OC ≥ 1.2% (0.6% tropics)
- Munsell colour chroma ≤ 3 value ≤ 3 moist or value ≤ 5 dry
- CEC ≥ 25 $cmol_c.kg^{-1}$
- Base sat. ≥ 50%

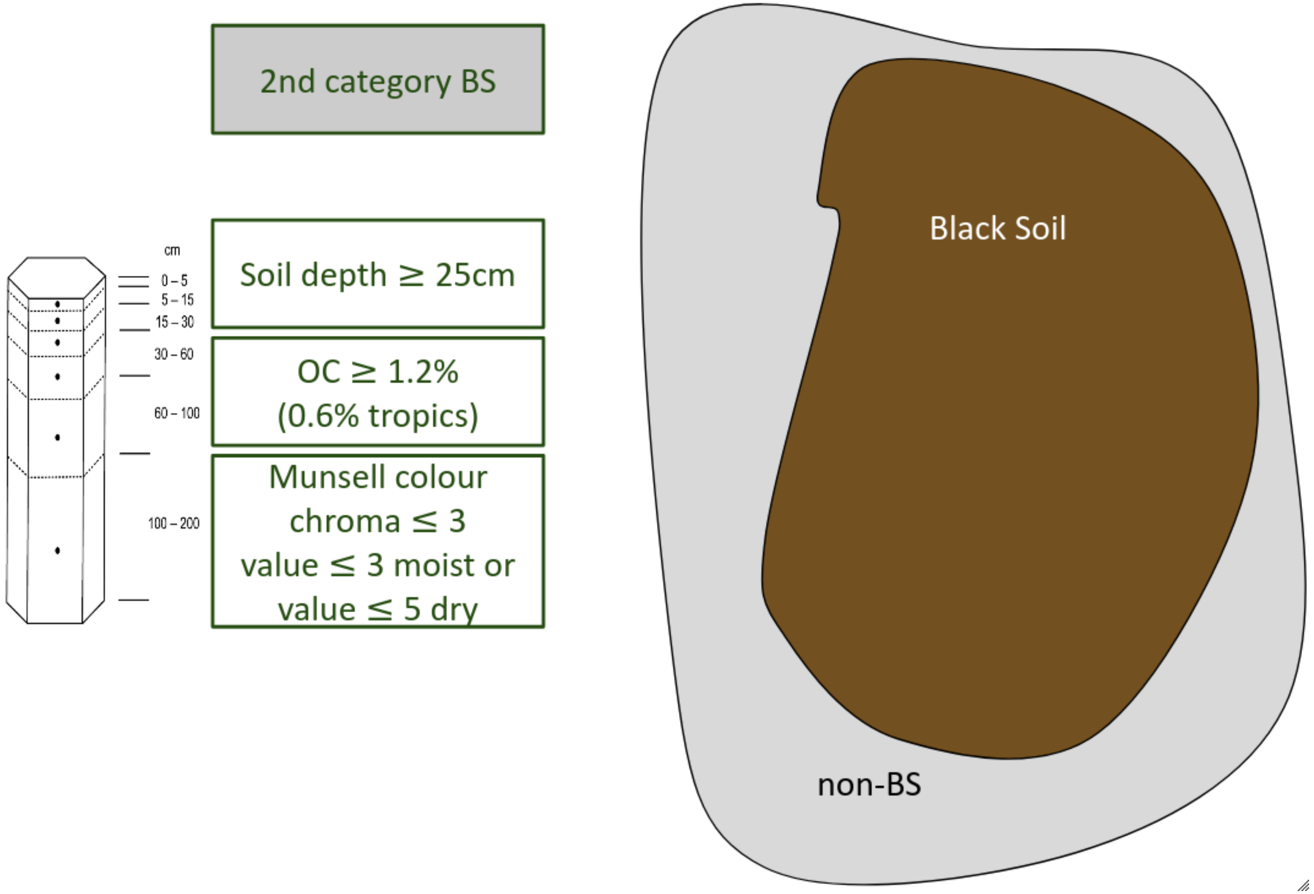
2nd category BS



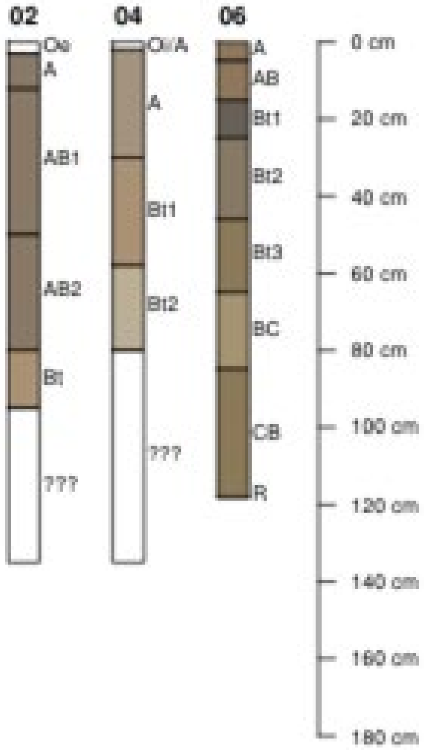
- Soil depth ≥ 25cm
- OC ≥ 1.2% (0.6% tropics)
- Munsell colour chroma ≤ 3 value ≤ 3 moist or value ≤ 5 dry



What is a black soil? -adaptation-



Mapping trajectories

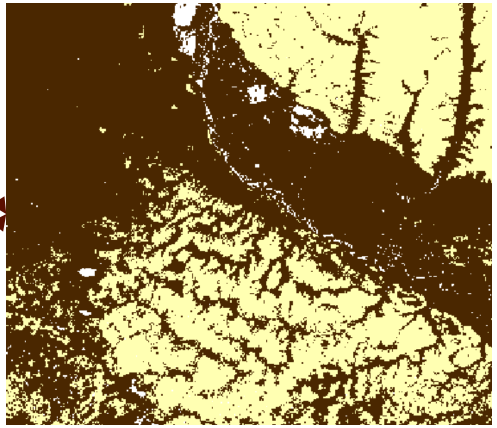


GSASmap

Modelling soil properties → **Merging maps**

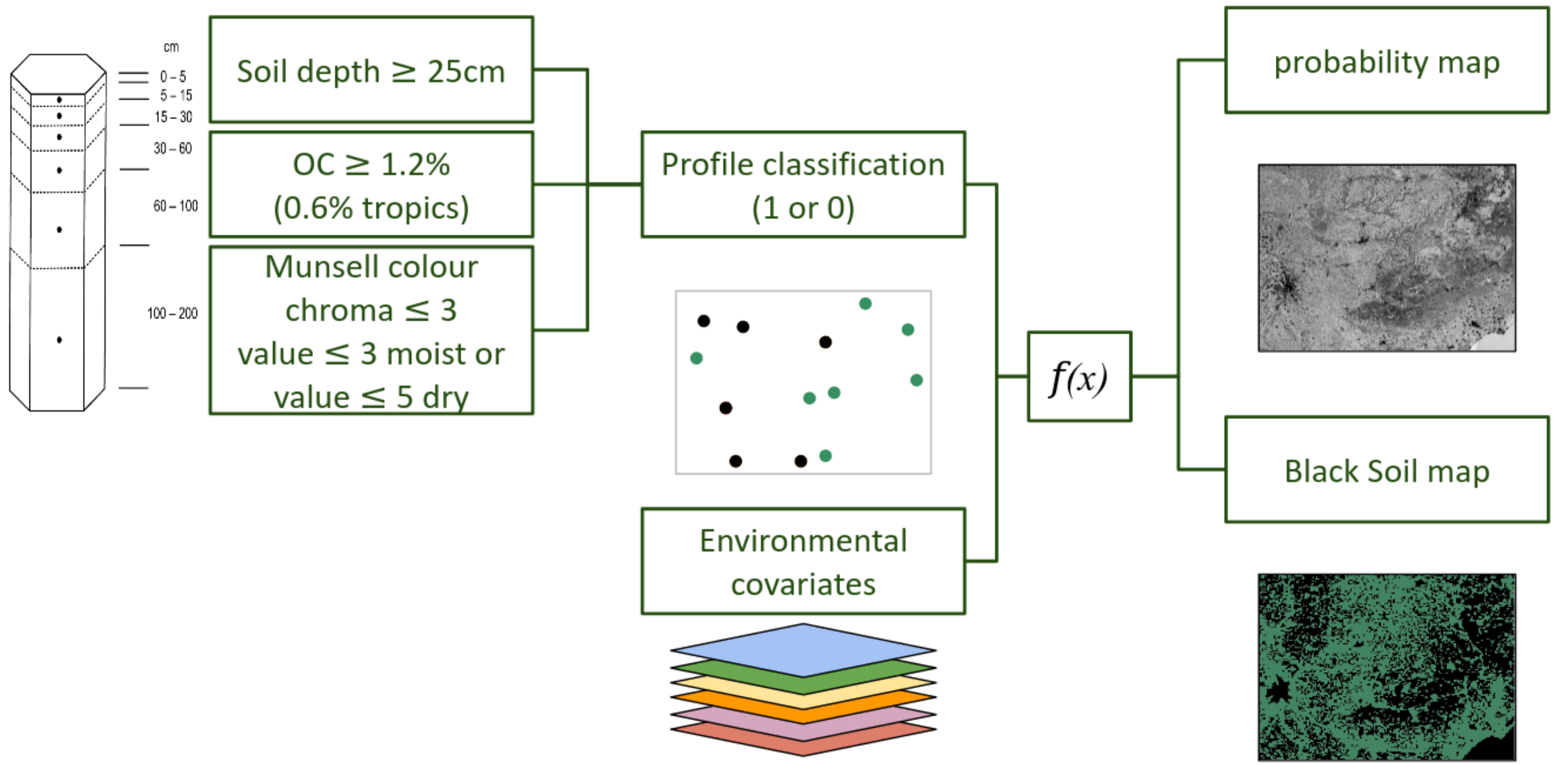
Convert profiles to classes → **Modelling the class BS**

GBSmap



GLOBAL SOIL PARTNERSHIP

Mapping Black Soils



Accuracy measurement.

		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

Global accuracy and ~~Kappa~~

http://scaryscientist.blogspot.com/2016/03/confusion-matrix.html

Deliverables

- Categorical BS map
- Probability BS map
- The map shall be produced at regular fixed horizontal dimensions of 30 by 30 arc-seconds grid (approximately only 1x1km at the equator).

Training session

December 11th 2020

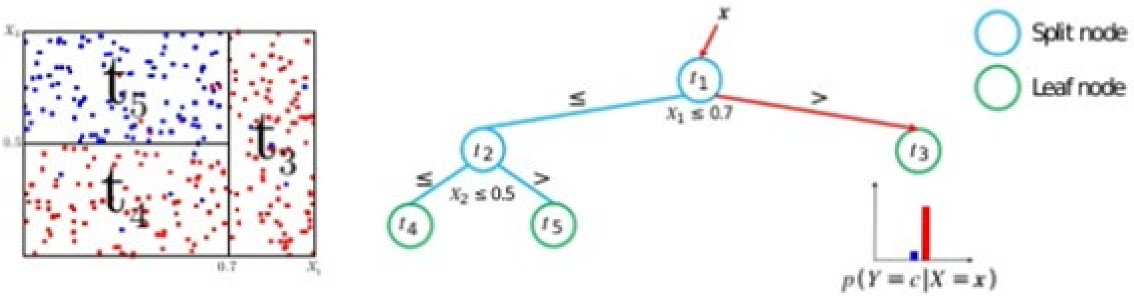
- Introduction to caret and randomForest packages
- Issues and solutions of class imbalances
- Modelling, validation and prediction: some concepts
- Analysis of results
- Questions, presentation of other study cases

Random Forest model: main features

- Regression and classification model
- Based on decision tree models
- it does not assume normal distribution of our dependent variable
- Extreme values will be hard to predict
- Model performance depend on hyperparameters that have to be tuned: mtry (number of features randomly selected) and ntree (number of trees)

randomForest documentation
https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Decision trees



$t \in \varphi$: nodes of the tree φ
X_t : split variable at t
$v_t \in \mathbb{R}$: split threshold at t
$\varphi(\mathbf{x}) = \arg\max_{c \in \mathcal{Y}} p(Y = c|X = \mathbf{x})$

https://www.slideshare.net/glouppe/understanding-random-forests-from-theory-to-practice

Caret package

Classification And REgression Training (CARET).

The package contains tools for:

- data splitting
- pre-processing
- feature selection
- model tuning using resampling
- variable importance estimation

https://topepo.github.io/caret/index.html

```
Define sets of model parameter values to evaluate
for each parameter set do
    for each resampling iteration do
        Hold–out specific samples
        [Optional] Pre–process the data
        Fit the model on the remainder
        Predict the hold–out samples
    end
    Calculate the average performance across hold–out predictions
end
Determine the optimal parameter set
Fit the final model to all the training data using the optimal parameter set
```

Class imbalances

- In **classification** problems, a disparity in the frequencies of the observed classes can have a significant negative impact on model fitting.

https://topepo.github.io/caret/subsampling-for-class-imbalances.html

https://medium.com/x8-the-ai-community/solving-class-imbalance-problem-in-cnn-9c7a5231c478

Modelling and validation

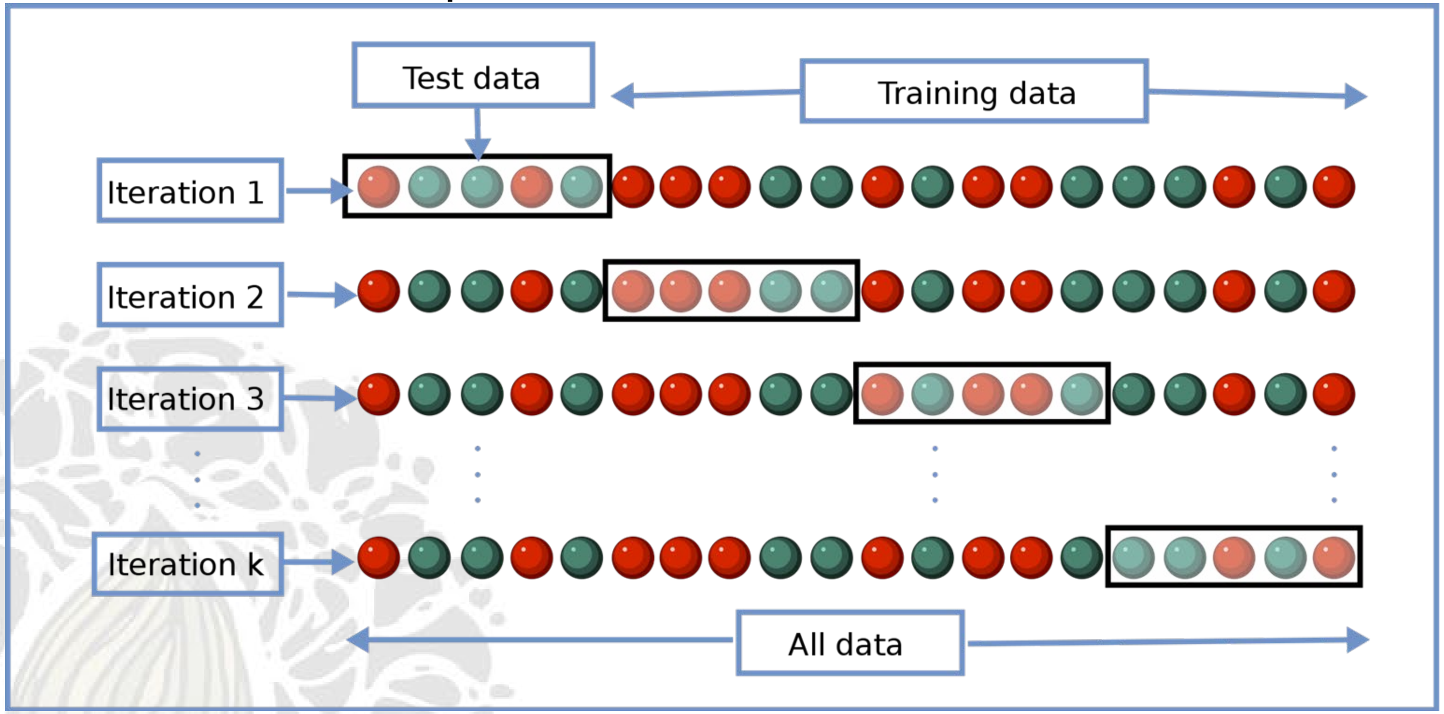
1. Modelling a random forest model needs to tune hyperparameters. We test different values for these parameters.

$100 < ntree < 500$ $\qquad$ $mtry \cong sqrt(m)$

1. Ideally, we should repeat downsampling >50 times for repeating balancing

Modelling and validation

1. Validation will be done by 10 times repeated 10-fold cross-validation

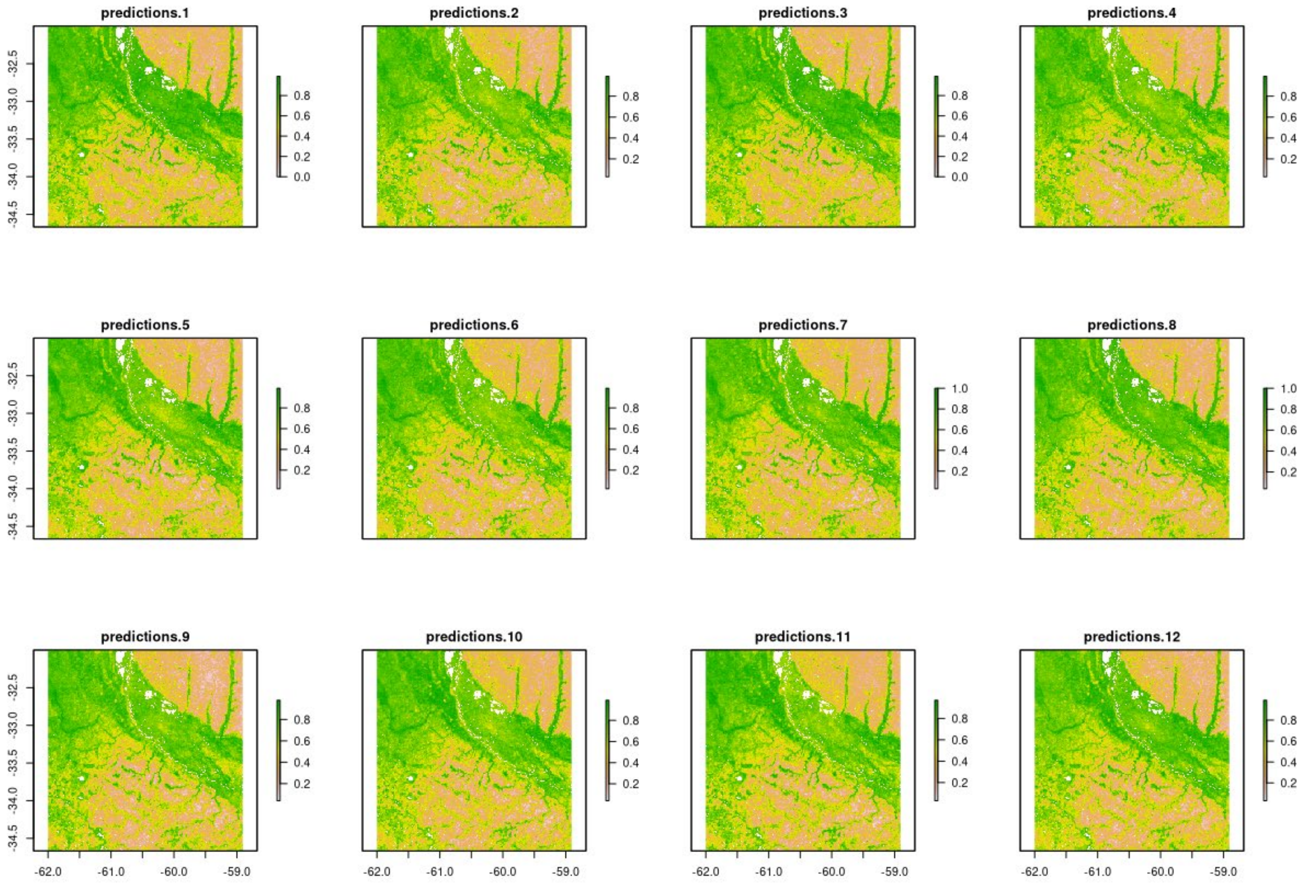


		Predicted Class	
		No	Yes
Observed Class	No	TN	FP
	Yes	FN	TP

Global accuracy

Final results

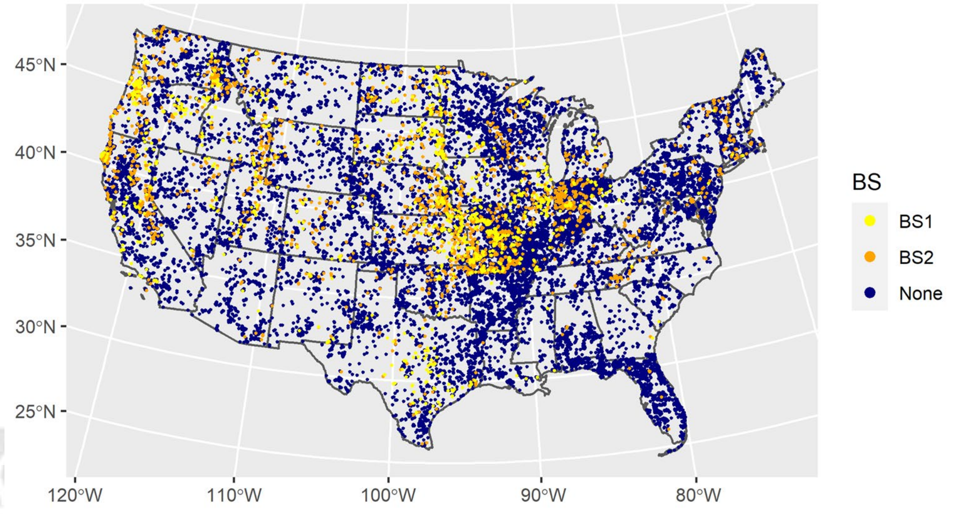
1. Downsampling may remove sensitive samples. By repeating downsampling and prediction we can create several maps that will me merged at the end by estimating the mean and standard deviation of them.



Results

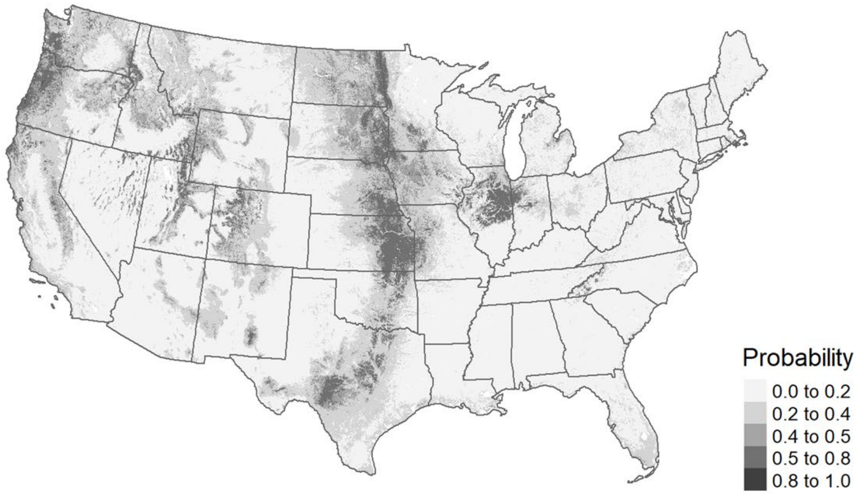
United States of America

Black Soil Category	FALSE	TRUE	SCD	RaCA	Total
BS1	34,286	1,139	33,716	1,709	35,425
BS2	31,331	4,364	33,928	1,767	35,695



Category 1

Category 2

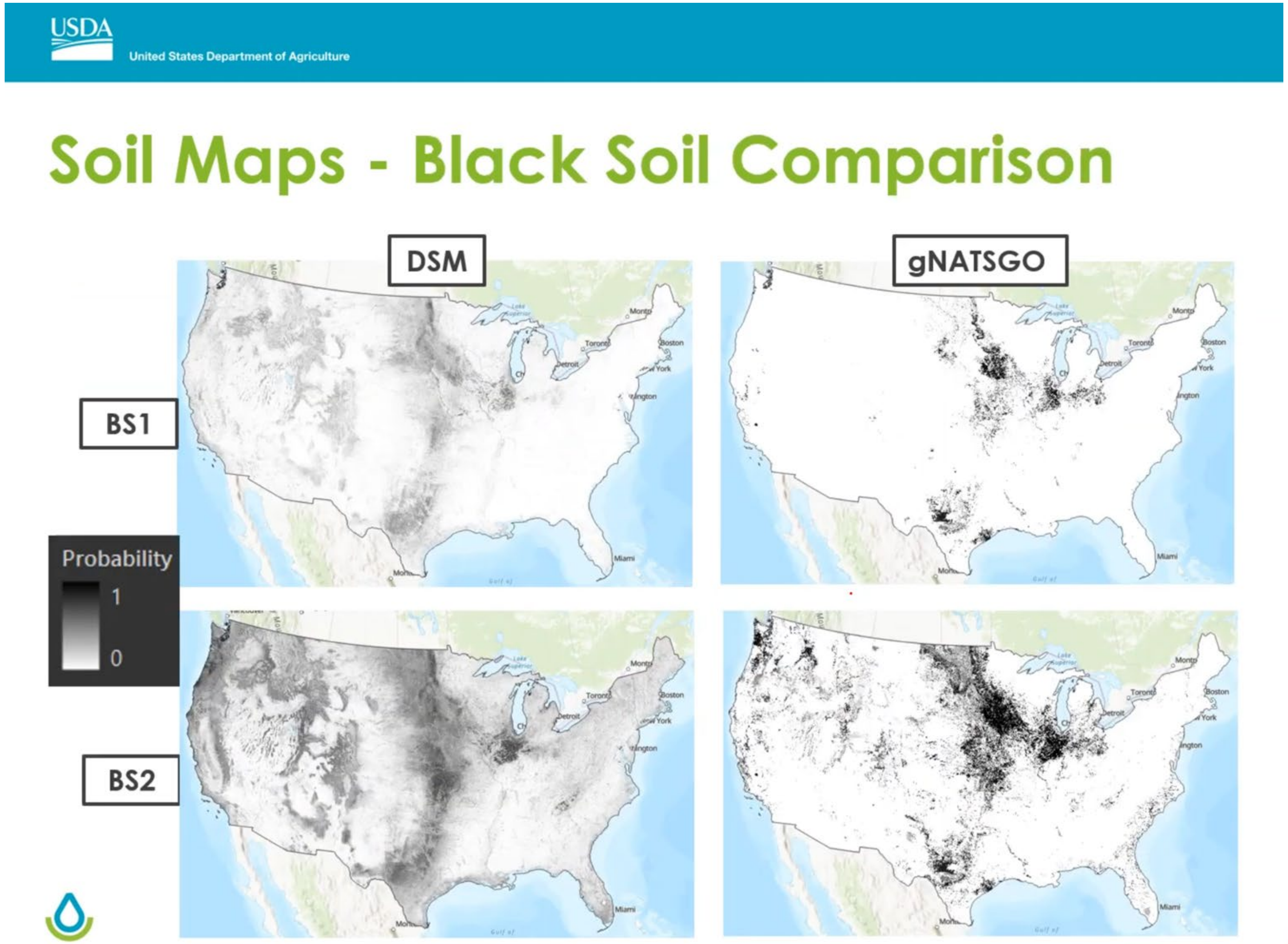


- They produced maps of BS category 1 (BS1) and BS category 2 (BS2)
- Applied downsampling

GLOBAL SOIL PARTNERSHIP

United States of America

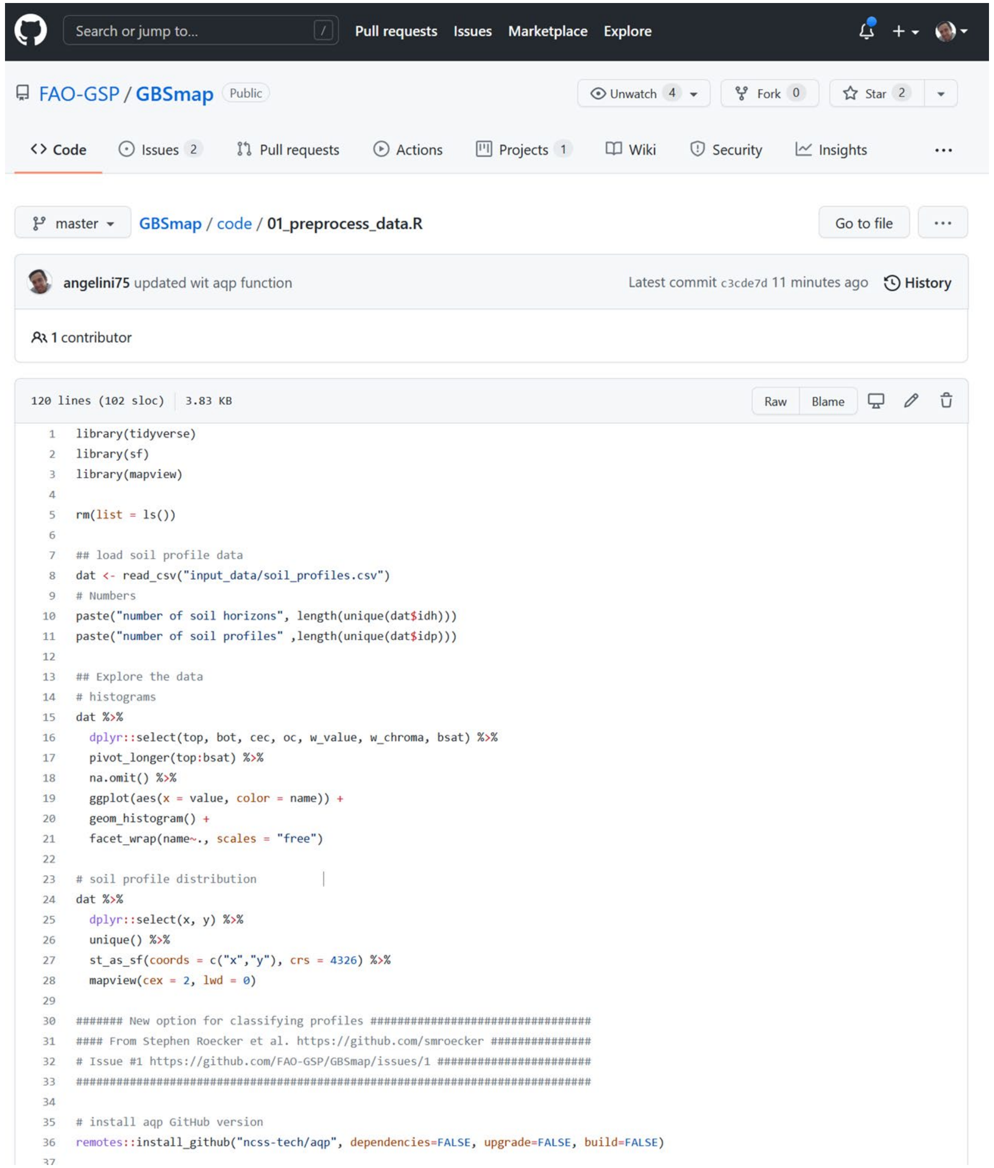
https://youtu.be/GBgCp3i-s-g



United States of America

- Stephan Roecker et al. (USDA-NRCS) also contributed to improve the code proposed by us.

https://github.com/FAO-GSP/GBSmap/blob/master/code/01_preprocess_data.R



GLOBAL SOIL PARTNERSHIP

United States of America

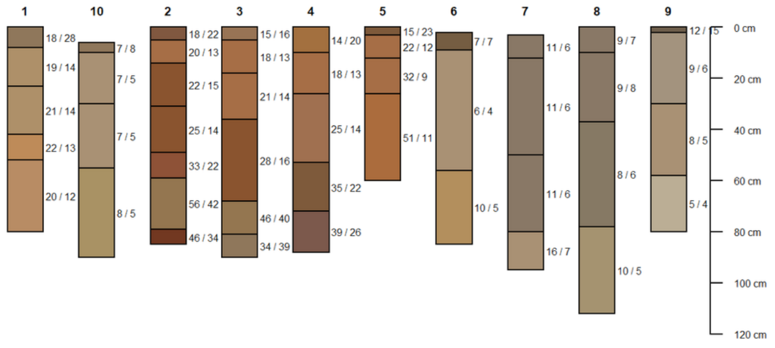
- Algorithm for quantitative pedology (aqp) package

https://ncss-tech.github.io/aqp/docs/index.html

allocate() https://ncss-tech.github.io/aqp/docs/reference/allocate.html

```
allocate(
  ...,
  to = c("FAO Salt Severity", "FAO Black Soil", "ST Diagnostic Features"),
  droplevels = TRUE
)
```

```
plotSPC(sp3)
```




```
allocate(
  sp3,
  to = 'FAO Black Soil',
  OC = 'tc',
  m_chroma = 'chroma',
  m_value = 'value',
  d_value = 'value',
  CEC = 'cec',
  BS = 'bs'
)
#>    peiid   BS1   BS2
#> 1      1 FALSE FALSE
#> 2     10 FALSE FALSE
#> 3      2 FALSE FALSE
#> 4      3 FALSE FALSE
#> 5      4 FALSE FALSE
#> 6      5 FALSE FALSE
#> 7      6 FALSE FALSE
#> 8      7 FALSE FALSE
#> 9      8 FALSE FALSE
#> 10     9 FALSE FALSE
```

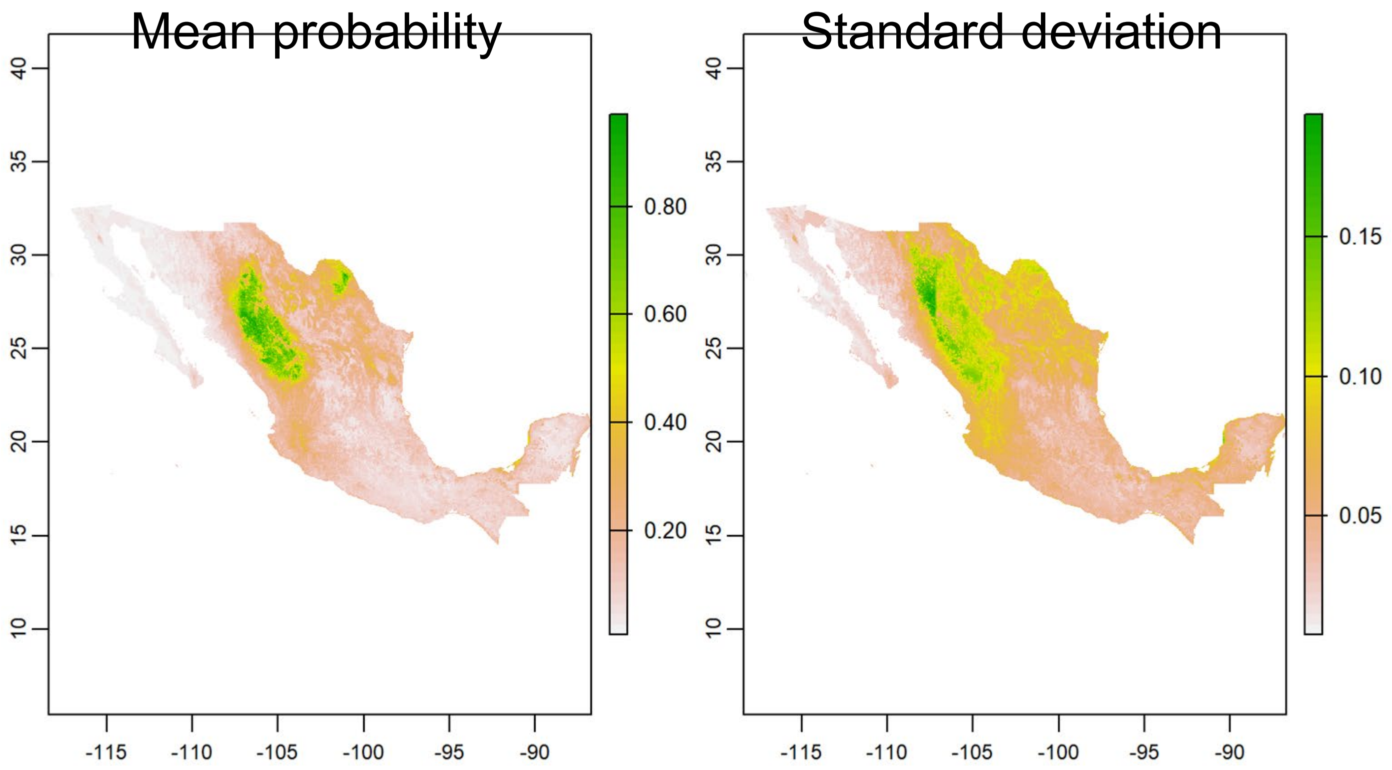

GLOBAL SOIL PARTNERSHIP

Other countries that followed DSM standard method

- Argentina
- Brazil
- Colombia
- Poland
- Ukraine
- Uruguay

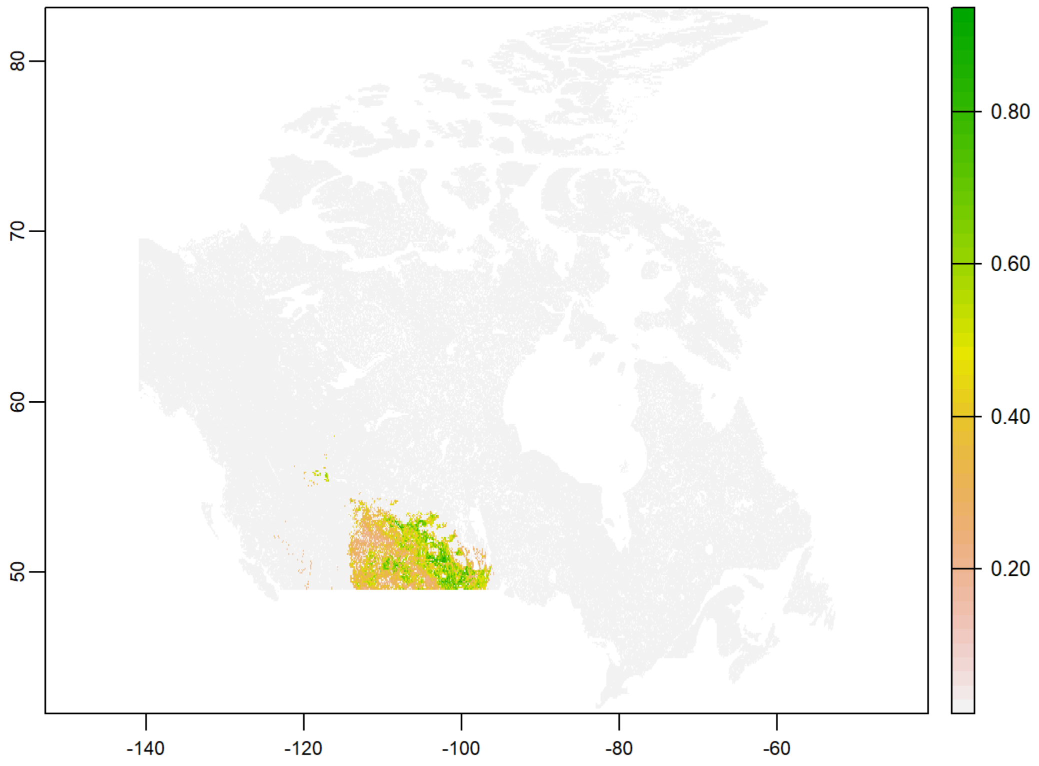
México

- 4400 samples non-BS vs 20 BS
- Downsampling would reduce training dataset too much
- We increased 20 BS to 100, and downsampled non-BS to 100 samples
- Sample balancing, calibration and prediction was repeated 100 times
- Cross-validation was not valid because upsampling introduced non independent data (on going work)
- 16 PCA layers were used as covariates



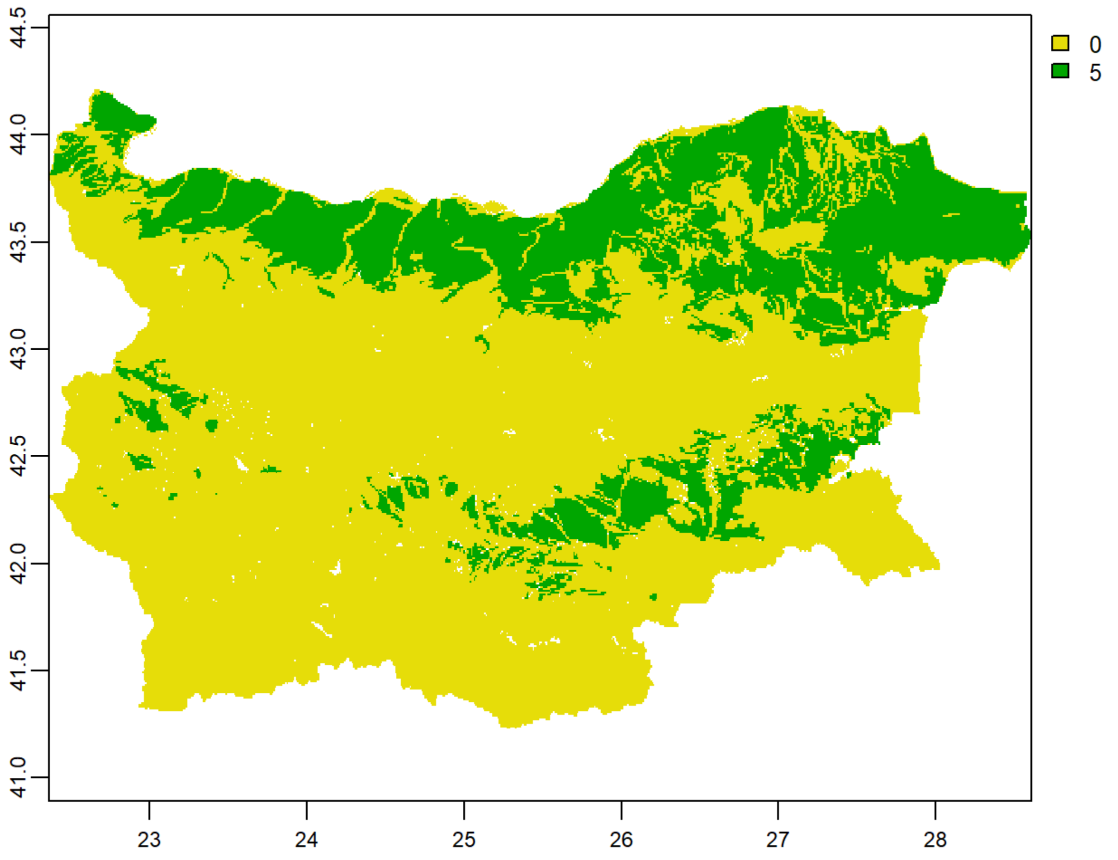
Canada

- Based on soil profiles with taxonomic classes
- DSM of soil clases
- Probability of classes related to BS



Bulgaria

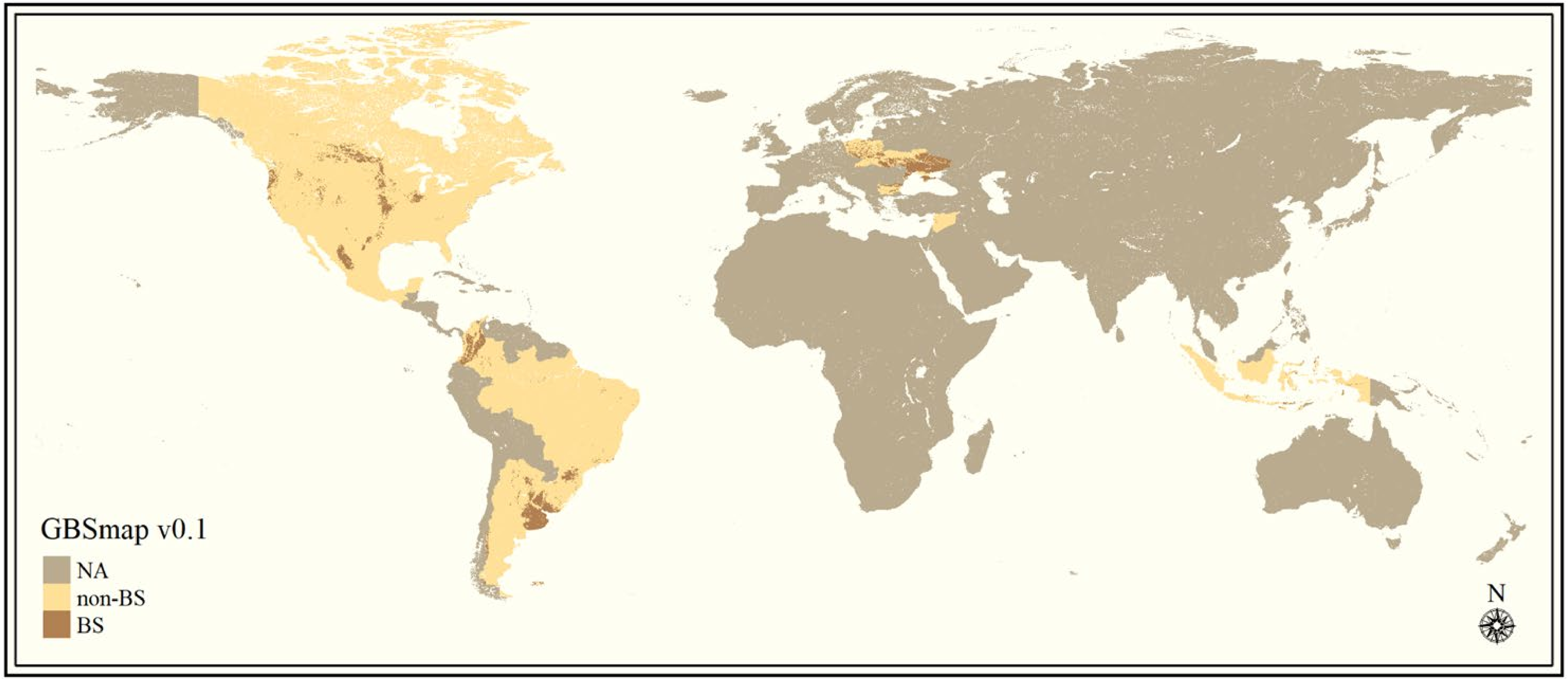
- BS map derived from soil-type maps



Other countries that followed soil-type based method

- Slovakia
- Indonesia
- Thailand (under review)
- Syrian Arab Republic

Global Black Soil map (GBSmap) v0.1



Ways forward for improvement

What is the main constraints for producing/improving the map of your country?

- Working load
- Lack of soil profiles
- Soil profiles are in paper support
- Lack of capacities
- Other

Ways forward for improvement

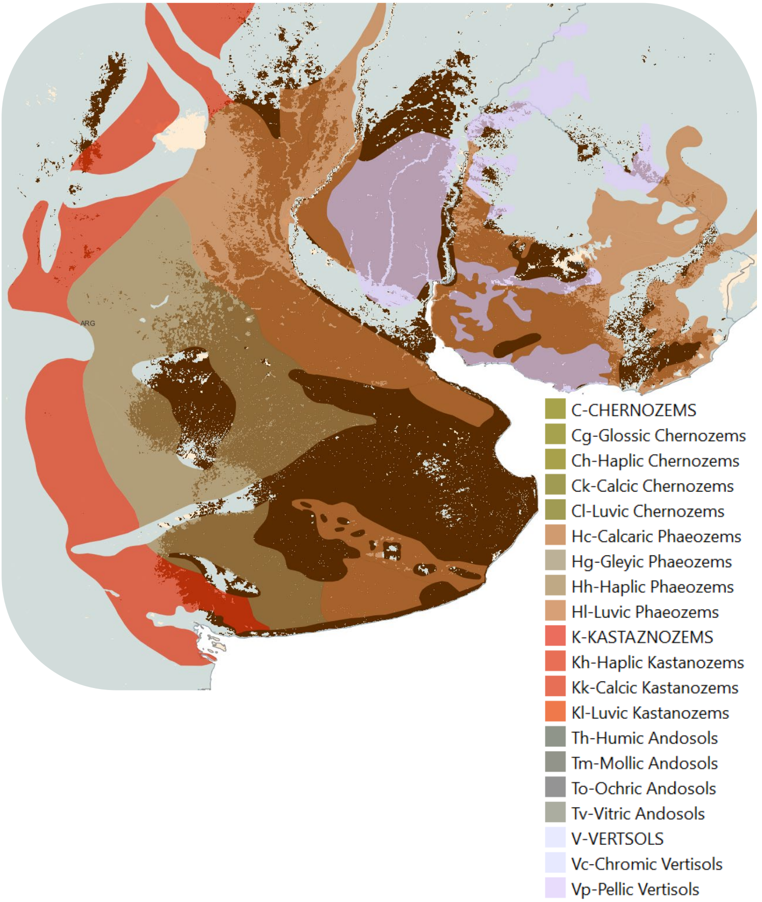
In what extent do you agree with this statements?

- Maps represented well what we consider black soils
- Maps overestimated the black soils area
- Maps underestimated the black soils area

Ways forward for improvement

Do BS categories represent what they were created for?

- Yes
- Maybe yes
- Maybe not
- No



Ways forward for improvement

What type of map do you think is more appropiate for general purposes?

- Probability maps (continuous values)
- Categorical maps (0 or 1)
- Both of them

Alternative ways for getting soil color

ELSEVIER

Geoderma
Volume 379, 1 December 2020, 114556

A soil colour map of China

Feng Liu [a, 1], David G. Rossiter [a, d, e] [2], Gan-Lin Zhang [a, c, b, f] [3], De-Cheng Li [a]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.geoderma.2020.114556   Get rights and content

Highlights

- A gridded soil colour map of China at 1 km resolution and 9 depths was produced.
- Predictive soil mapping from 4600 profile descriptions and 40 covariates.
- Regional patterns and details of soil colour are clearly shown.
- Random forest models did not predict less common colours.
- Maps can be used for education or linked to soil processes and classification.

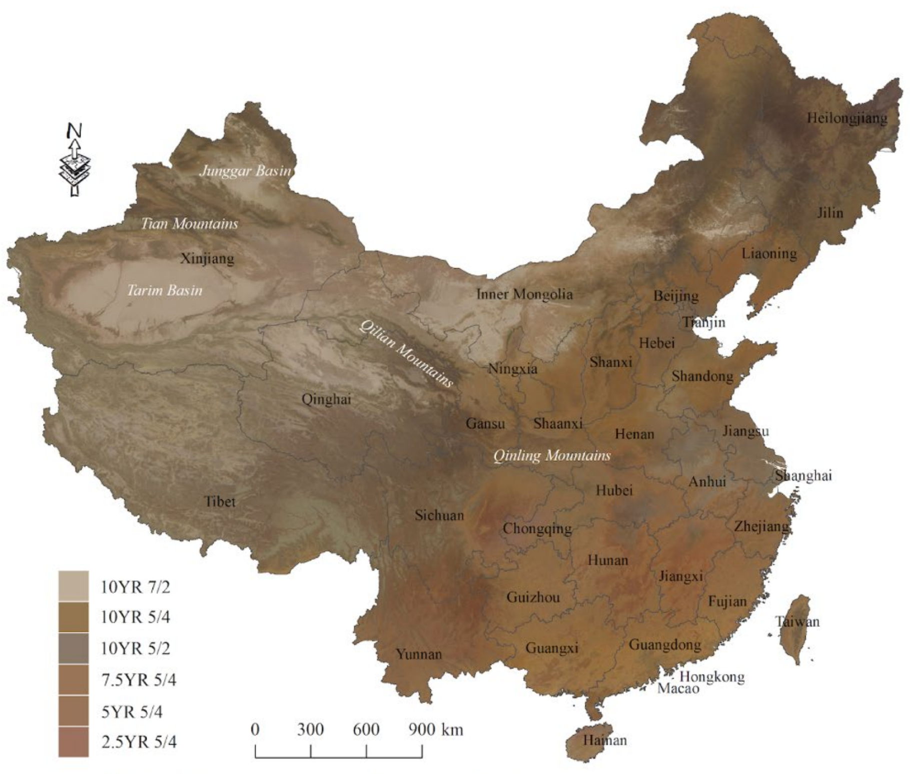


Fig. 4. Predicted moist colours at 25 cm depth. Colour bar shows some representative colours.

Can we extrapolate soil color models?

GLOBAL SOIL PARTNERSHIP

Alternative ways for getting soil color

SOIL SCIENCE AND PLANT NUTRITION
2019, VOL. 65, NO. 6, 605-615
https://doi.org/10.1080/00380768.2019.1676624

Taylor & Francis
Taylor & Francis Group

Comparison of visual and instrumental measurements of soil color with different low-cost colorimeters

Naoki Moritsuka[a], Kensuke Kawamura[b], Yasuhiro Tsujimoto[b], Michel Rabenarivo[c], Andry Andriamananjara[c], Tovohery Rakotoson[c], and Tantely Razafimbelo[c]

[a] Graduate School of Agriculture, Kyoto University, Kyoto, Japan [b] Japan International Research Center for Agricultural Sciences, Ibaraki, Japan [c] Laboratory of Radioisotopes, University of Antananarivo, Antananarivo, Madagascar

ABSTRACT

Soil color has been conventionally measured using a Munsell soil color chart. Recently launched colorimeters can also measure the color of an object at a reasonable cost. This study was undertaken to evaluate to what extent such low-cost colorimeters (< 500 USD) can be useful for soil color analysis in the laboratory as compared with conventional colorimeters (> 3000 USD) and a Munsell soil color chart (about 200 USD). Sixty two air-dried soil samples collected from rice fields in Madagascar were subjected to two pretreatments for homogenization (2-mm sieving or additional hand-grinding) and instrumental analysis using four low-cost colorimeters (CS-10, Cube, Nix Pro, and Color Muse) and four conventional colorimeters (SPAD-503, CR-20, CR-400, and CR-410). The color of 2-mm sieved samples was also measured visually using a color chart. The effects of pretreatments and the analytical conditions were evaluated by the repeatability and stability of the measurement, the comparability of the soil color data obtained, and the time required for analysis. Overall, instrumental measurement was much more repeatable than visual observation. Both the repeatability and stability of the low-cost colorimeters tended to be lower than those of the conventional colorimeters. Among the low-cost colorimeters examined, soil color parameters (L*, a*, and b*) obtained with Nix Pro were most comparable ($r > 0.97$ for all parameters of 2-mm sieved samples) with those obtained with SPAD-503, which is an instrument designed specifically for soil color analysis. Additional hand-grinding pretreatment improved the repeatability of the instrumental analysis and reduced the subsample weight to two grams. However, this pretreatment also increased the L* value (lightness) of the samples, decreased the comparability with the data from 2-mm sieved samples, and prolonged the time required to complete the whole analysis. Among the various methods tested, 2-mm sieving of air-dried samples followed by the color measurement with Nix Pro several times per sample was considered the most cost-effective approach for measuring soil color in the laboratory.

ARTICLE HISTORY
Received 29 March 2019
Accepted 1 October 2019

KEY WORDS
Cost effectiveness, grinding pretreatment, nondestructive analysis, repeatability, soil color

GLOSOLAN could consider soil color to be measured

Conclusions

- Black soil categories (1 and 2) overlap one to each other, what might be confusing
- Missing data is the main constraints for mapping BS
- Soil-type map disaggregation might be an alternative approach
- We need to involve more countries in the mapping process
- Sharing data between countries is a sensitive issue, but can improve national maps

The Global Soil Partnership (GSP) is a globally recognized mechanism established in 2012. Our mission is to position soils in the Global Agenda through collective action. Our key objectives are to promote Sustainable Soil Management (SSM) and improve soil governance to guarantee healthy and productive soils, and support the provision of essential ecosystem services towards food security and improved nutrition, climate change adaptation and mitigation, and sustainable development.

Thank you all for your kind attention

yusuf.yigini@fao.org, marcos.angelini@fao.org