Contribution to D3.6: Open access harmonised maps

Holistic management practices, modelling and monitoring for European forest soils, HoliSoils

Project duration: 1.5.2021–31.10.2025

H2020 Grant Agreement No 101000289

Coordinator: Natural Resources Institute Finland (Luke)

Deliverable DX.X: Deliverable title (use short version if necessary)		
Brief description of your deliverable (75 – 100 words)		
Due date	October 2024	
Authors	Laura Poggio, Niels Batjes, Luis Calisto, Betony Colman, Ulan Turdukulov	
Date of publication	19 March 2025	
Dissemination level		
PU	Public, fully open e.g. web	X
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	
Nature of the Deliverable		
R	Document, report	x
DEM	Demonstration, pilot, prototype, plan design	
DEC	Websites, patents filing, market studies, press & media actions, videos etc.	
OTHER	Software, technical diagram etc.	
Ethics	Ethics deliverables	

This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 101000289.

Contents

1. Introduction

This report summarises the procedures for producing maps of soil properties for European Forest soils within the HoliSoils project. It briefly describes the input data, both observations and environmental layers, and the methodology. Subsequently, there is a short summary of the main results. The report, which is a contribution to HoliSoils Deliverable D3.6 "Open access harmonised maps" coordinated by Thünen Institute, ends with a short discussion and conclusion. The data availability section explains where the maps can be found and how they can be accessed.

2. Input data

Input data covers both the soil observations (i.e. point measurements) and the environmental layers used to produce the maps.

2.1 Observations

The soil observations used for mapping were queried from the WoSIS database, which was augmented with datasets provided in the context of the HoliSoils project :

- HoliSoils provided datasets (Wellbro*ck* et al., 2021) that include:
    - ICP Forest Level I and Level II (ICP Forests 2021),
    - A compilation of forest soil data for Poland (Brozek, et al. 2019)
    - French soil monitoring system (RQMS 2021), which considers multiple land covers besides forests, and
- Forest soil data for Europe held in WoSIS (Batjes, Calisto, and Sousa 2024).

The data shared in the framework of the HoliSoils projects, and their licences, are visualised in a dashboard, for details see Batjes and Calisto (2023). Out of these 7,200 'profiles' about 3,500 are under EU Corine 2018 land cover (see below). Complemented with the existing WoSIS forest 'profiles' this yielded about 27,825 profiles under Corine 2018 land classes for the mapping work (Figure 1).

Thus, to represent ~227 million ha forest soils in Europe, only observations within Corine forest land cover were considered. To select the observations, the following criteria were used:

- Belong to a dataset for which all observations were collected in forest (for example ICP Forest)
- Be in a forest land cover category as defined in the Corine land cover (European Union's Copernicus Land Monitoring Service information 2018) including: 3.1.1 Broad-leaved forest, 3.1.2 Coniferous forest and 3.1.3 Mixed forest.

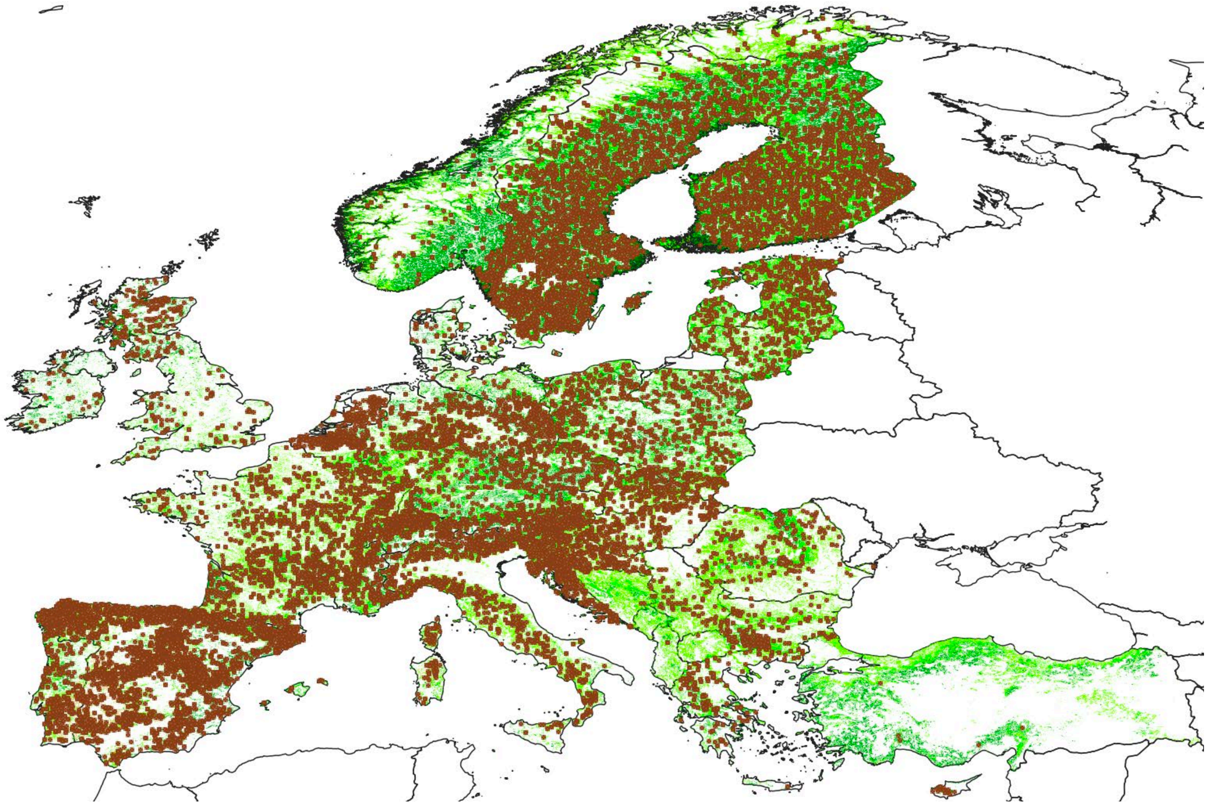


Figure 1. Distribution of forest soil 'profiles', in Corine 2018 land cover classes, considered in the HoliSoils mapping work.

It proved cumbersome to unambiguously 'flag' organic surface layers at the top of mineral forest soils in view of inconsistencies/differences of recording between countries and classification systems (e.g. Gobezie et al. 2024; Batjes et al. 2024). Alternatively, for organic (i.e. peat) soils the whole depth was considered. We are aware that this impacts the organic carbon stocks calculations, with possible effects on e.g. subsequent 'forest soil vulnerability' assessments for HoliSoils. Surface layers above mineral soil potentially are a large pool of organic carbon, especially in boreal and temperate forest (Ziche et al. 2019). This aspect will be investigated in further comparative studies.

2.1.1 Depth distribution

Most observations are available for the topsoil only (<60 cm). The table below is an example of the depth distribution of the observations for soil organic carbon (SOC). It is representative of the other properties as well as considered a best case scenario.

Depth interval (cm)	Number	Proportion
0-5	217	0.01
5-15	14074	0.71
15-30	5311	0.27
30-60	94	0.00
>60	None	None

2.1.2 Modelled properties

The following primary soil properties, as determined following discussions with HoliSoils partners, were modelled:

Property	Description	Units
Soc	Soil organic carbon content	g/kg
phh2o	pH in water	pH
nitrogen	Total nitrogen	g/kg
Bdod	Bulk density, oven dry	kg/dm3
Cfvo	Coarse fragments, volumetric	cm3/100cm3
Sic	Soil inorganic carbon content	g/kg
Sand	Sand	g/100g
Silt	Silt	g/100g
Clay	Clay	g/100g
Ocstk	Soil organic carbon stocks	kg/m2
Icstk	Soil inorganic carbon stocks	kg/m2
Nistk	Soil nitrogen stocks	kg/m2

2.2 Environmental layers as covariates

The following covariates were used:

- Sentinel1 composites band Synthetic Aperture Radar (SAR) instrument. The data were pre-processed, prepared, mosaicked and downloaded from Google Earth Engine (Gorelick et al. 2017). VH is a vertically transmitted and horizontally received SAR (Synthetic Aperture Radar) backscatter signal from S1. VV is a vertically transmitted and received SAR backscatter signal. The polarization ratio was calculated: (VH-VV)/(VV+VV). The overall median (2018 to 2020) was calculated and used as covariates.
- Sentinel 2 composites. The overall median for 2021 was calculated and used to derive the vegetation indices. In this project the following basic indices were used: NDVI: Normalized Difference Vegetation Index, NDWI: Normalized Difference Water Index, NDSI: Normalized Difference Snow Index.

- Geomorphological and terrain parameters derived from Copernicus Digital Elevation Model (ESA, 2024), including:
    - Elevation from sea level (m),
    - Slope as the steepest slope angle, calculated using the 'D8 method' (O'Callaghan & Mark, 1984),
    - Topographic wetness index (Sørensen et al., 2006) defined as the ln(a/tan(b))
    - Multi-resolution Valley Bottom Flatness (MRVBF) index (Gallant & Dowling, 2003).
- Climate data derived from ERA5. ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset. ERA5 provides aggregated values for each month for the following ERA5 climate reanalysis parameters: temperature_2m, total precipitation, runoff, total evaporation, surface net solar radiation. Monthly total precipitation values are given as monthly sums. All other parameters are provided as monthly averages.
- Land cover: The Copernicus Global Land Service (CGLS). The 2019 products (Buchhorn et al. 2020) with the proportional estimates for vegetation/ground cover for the land cover types were used in this project.
- Parent material: Information on parent material was derived from the European Soil Database v2.0 (Panagos et al. 2022). Both primary (Code for dominant parent material of the STU) and secondary (Code for secondary parent material of the STU) units where rasterized in binary (presence/absence) format.

3. Mapping approach

To generate soil properties maps for European forests, soil observations from multiple locations in Europe were analysed in relation to a set of environmental covariates. Random Forests (Breiman 2001), were used as the modelling technique, using the *ranger* package (Wright and Ziegler 2017), with the option *quantreg* to build Quantile Random Forests (QRF) (Meinshausen 2006) for spatial uncertainty assessment. With QRF, predictions generate a cumulative probability distribution of the soil property at each location, rather than a single average value from the ensemble of decision trees. These observations were split in 10 equal folds for cross-validation. Model tuning was conducted using a 10-fold cross-validation procedure applied to multiple combinations of hyper-parameters.

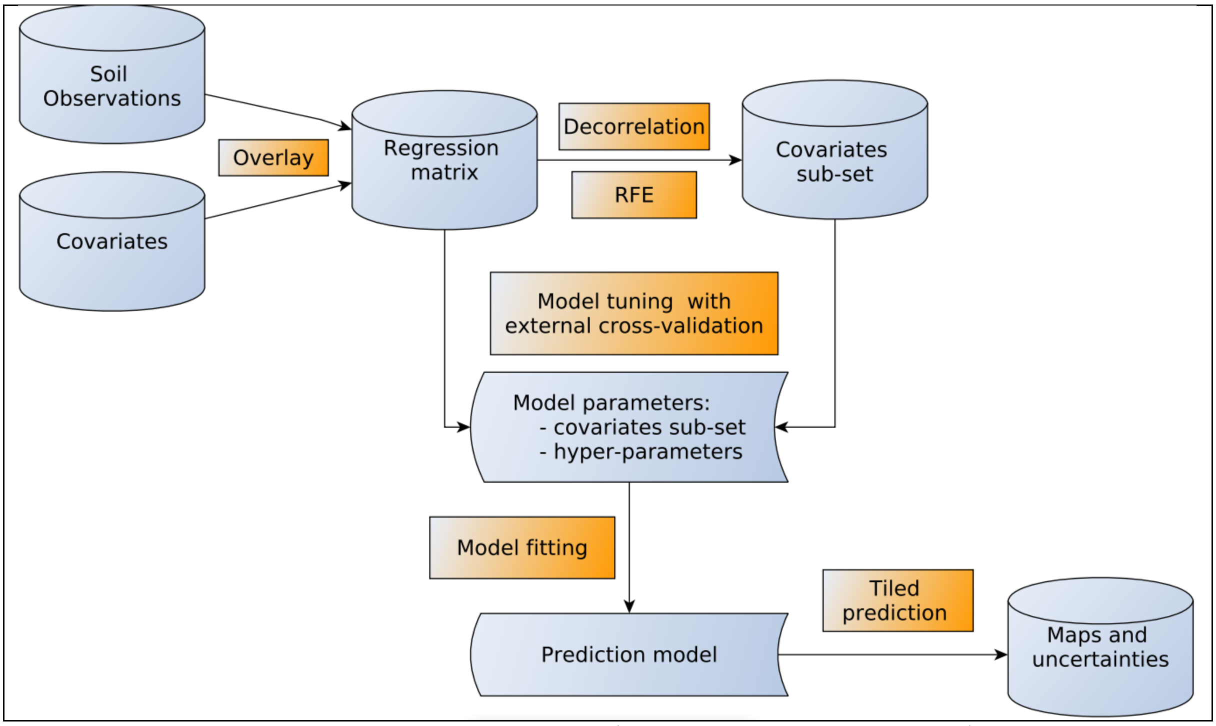


Figure 2. Workflow for the digital soil mapping (Source: Poggio et al., 2021).

Models were trained with different combinations of the number of decision trees (*ntree* parameter) and numbers of covariates (*mtry* parameter). Each of the resulting combinations of *ntree* and *mtry* parameters was used to train a different model with observations from nine folds. Predictions were then assessed on the remaining fold using root mean squared error (RMSE) and model efficiency coefficient (MEC)(Janssen and Heuberger, 1995). The model evaluation was based on the performance metrics of the selected hyper-parameters combination. The final model was fitted with all available observations, the covariates and the hyperparameters selected in the previous steps.

This approach provides the mean predictions, but also the quantiles. We chose the 5th and 95th quantiles according to the *globalsoilmap* specifications (Arrouays et al. 2014). The uncertainty was defined as:

$$uncertainty = \frac{|Q0.95 - Q0.05|}{Q0.50} * 10$$

This approach builds on the methods described in Poggio et al. (2021).

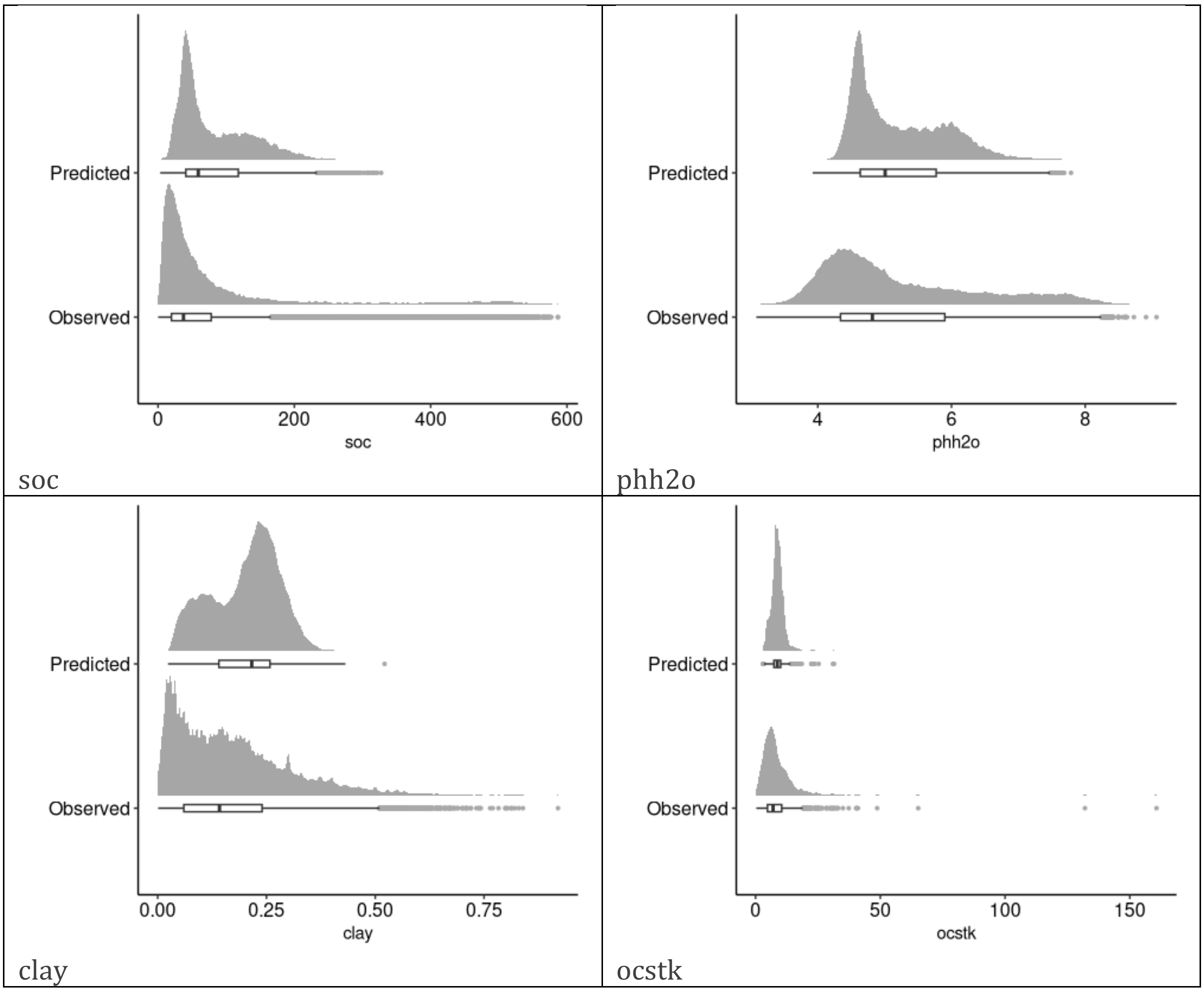
4. Summary of modelling results

4.1 Topsoil only

This section describes the results of the topsoil only modelling. Topsoil was defined here as 0-30 cm, in accordance with IPCC (2016) standards, and values were averaged using a weighted average by the thickness of the layer (Poggio et al. 2021).

	bdod	Cfvo	Clay	icstk	nistk	nitrogen	ocstk	pH	sand	sic	silt	soc
MEC	0.21	0.23	0.16	0.34	0.16	0.07	0.05	0.43	0.25	0.26	0.19	0.17
RMSE	0.35	15.02	0.12	3.93	0.38	4.73	7.78	0.88	0.21	9.73	0.16	12.88

MEC : model efficiency coefficient; RMSE : root mean squared error



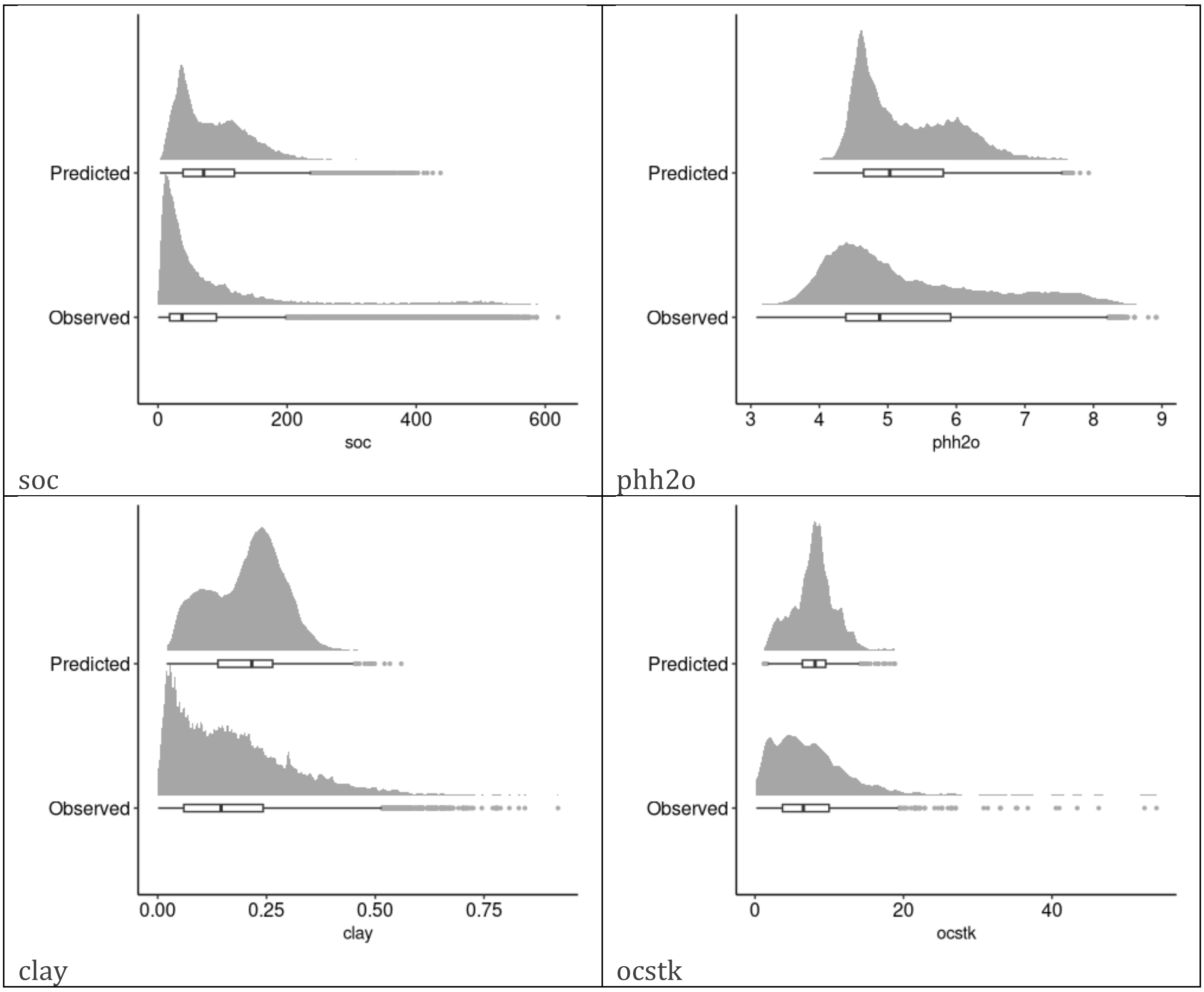
4.2 Soil to 60 cm thick

This section describes the results of the 3D modelling for three layers (0-20, 20-40 and 40 to 60 cm). Values were averaged using a weighted average by the thickness of the layer

(Poggio et al. 2021). For the purpose of this report, the cross-validation statistics are for the whole profile and not per layer.

	Bdod	cfvo	clay	icstk	nistk	nitrogen	ocstk	pH	sand	sic	silt	soc
MEC	0.49	0.35	0.08	-0.06	0.18	0.35	0.15	0.41	0.27	0.26	0.24	0.43
RMSE	0.30	17.71	0.14	11.74	0.36	4.66	5.49	0.92	0.23	11.46	0.16	10.72

MEC : model efficiency coefficient; RMSE : root mean squared error



5. Discussion and conclusion

As indicated earlier, relatively few forest soil data were shared in the context of the HoliSoils project, and many of these only considered the upper soil layers (<60 cm). This had direct implications on the possible accuracy of the predictions.

Comparisons with soil organic carbon (SOC) maps generated in the framework of several other EU-scale projects, such as EJP, CUP4SOIL and ESA World Soils, indicated that there are still quite some differences in the estimated properties as well as in their spatial distribution.

The work showed the importance of having sufficient quality-assessed, harmonised profile data when mapping and modelling across countries at continental scale to support decision making towards climate and sustainability goals. Possibly, data fusion of the different SOC maps could lead to better and more reliable estimates in the future.

6. Data availability

Ultimately, all maps produced by HoliSoils will be made available via the Thünen Institute Scientific Data Repository (TISDAR), which will provide improved visualisation and querying options. Meanwhile, the soil property maps can be accessed via web-services in WMS format from the ISRIC website.

The following webpage provides details on links, services and citation of the soil property maps:

https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/b3d7c844-cbee-4b0f-8431-3a9373f5a59a

7. Acknowledgements

This report was prepared at ISRIC in conjunction with the HoliSoils project, with funding from the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101000289).

8. References

Arrouays, Dominique, Michael G. Grundy, Alfred E. Hartemink, Jonathan W. Hempel, Gerard B. M. Heuvelink, S. Young Hong, Philippe Lagacherie, et al. 2014. "Chapter Three - GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties." In, edited by Donald L. Sparks, 125:93–134. Advances in Agronomy. Academic Press. https://doi.org/http://dx.doi.org/10.1016/B978-0-12-800137-0.00003-0

Batjes, N. H., L. Calisto, and L. M. de Sousa. 2024. "Providing Quality-Assessed and Standardised Soil Data to Support Global Mapping and Modelling (WoSIS Snapshot 2023)." *Earth System Science Data* 16 (10): 4735–65. https://doi.org/10.5194/essd-16-4735-2024

Batjes N. H. and L. Calisto. 2023. D3.4 Open access harmonised soil database. Holistic management practices, modelling and monitoring for European forest soils, HoliSoils (EU H2020 Grant Agreement No 10100028), p.13 with dashboard. https://holisoils.eu/wp-content/uploads/2024/03/Deliverable-3.4-Open-access-harmonised-soil-database.pdf

Becker, Francois, and Bhaskar J Choudhury. 1988. "Relative Sensitivity of Normalized Difference Vegetation Index (NDVI) and Microwave Polarization Difference Index (MPDI) for Vegetation and Desertification Monitoring." *Remote Sensing of Environment* 24 (2): 297–311. https://doi.org/10.1016/0034-4257(88)90031-4

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. http://dx.doi.org/10.1023/A:1010933404324

Brozek S, Lasota J, Gruba P, Wanic T, Zwydak M, Blonska E, and Pacanowki P. 2019. *AtlasSiedlisk Lesnych Polski (Atlas of Polish Forest Habitats).* Centrum Informacyjne LaszówPanstowych, Warsaw (Poland). https://www.lasy.gov.pl/pl/publikacje/copy_of_gospodarka-lesna/urzadzanie/atlassiedlisk-lesnych-4.pdf/@@download/file/Atlas siedlisk leśnych.pdf

Buchhorn, Marcel, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, Luc Bertels, and Bruno Smets. 2020. "Copernicus Global Land Cover Layers - Collection 2." *Remote Sensing* 12 (6). https://doi.org/10.3390/rs12061044

Copernicus Climate Change Service (C3S). 2017. "ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS)." https://cds.climate.copernicus.eu/

European Space Agency (2024). *Copernicus Global Digital Elevation Model*. https://doi.org/10.5270/ESA-c5d3d65

European Union's Copernicus Land Monitoring Service information. 2018. "CORINE Land Cover 2018." https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac

Gherboudj, Imen, Ramata Magagi, Aaron A. Berg, and Brenda Toth. 2011. "Soil Moisture Retrieval over Agricultural Fields from Multi-Polarized and Multi-Angular RADARSAT-2 SAR Data." *Remote Sensing of Environment* 115 (1): 33–43. https://doi.org/10.1016/j.rse.2010.07.011

Gobezie T.B., Daniel S., Scott S.D, Daggupati P., Bedard-Haughn A. and Biswas A. 2024. Application of Multidimensional Soil Data Harmonization to Develop the Ontario Soil Information System (OSIS). *Canadian Journal of Soil Science* 105:1-15. https://doi.org/10.1139/cjss-2024-0070

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment*. https://doi.org/http://dx.doi.org/10.1016/j.rse.2017.06.031

ICP Forests. 2021. *ICP Forests Monitoring Manual. Part x: Sampling and Analysis of Soil.* https://storage.ning.com/topology/rest/1.0/file/get/9995584862?profile=original

IPCC. 2006. *IPCC Guidelines for National Greenhouse Gas Inventories Volume 4: Agriculture, Forestry and other Land Use*, IPCC National Greenhouse Gas Inventories Programme, Hayama (JP). https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html

Janssen, P. H. M., and P. S. C. Heuberger. 1995. "Calibration of Process-Oriented Models." *Ecological Modelling* 83 (1): 55–66. https://doi.org/https://doi.org/10.1016/0304-3800(95)00084-9

Meinshausen, Nicolai. 2006. "Quantile Regression Forests." *Journal of Machine Learning Research* 7 (35): 983–99. http://jmlr.org/papers/v7/meinshausen06a.html

O'Callaghan, J. F., and D. M. Mark. 1984. "The Extraction of Drainage Networks from Digital Elevation Data." *Computer Vision Graphics Image Processing* 28: 323–44.

Panagos, Panos, Marc Van Liedekerke, Pasquale Borrelli, Julia Köninger, Cristiano Ballabio, Alberto Orgiazzi, Emanuele Lugato, et al. 2022. "European Soil Data Centre 2.0: Soil Data and Knowledge in Support of the EU Policies." *European Journal of Soil Science* 73 (6): e13315. https://doi.org/10.1111/ejss.13315

Patel, Parul, Hari Shanker Srivastava, Sushma Panigrahy, and Jai Singh Parihar. 2006. "Comparative Evaluation of the Sensitivity of Multi?polarized Multi?frequency SAR Backscatter to Plant Density." *International Journal of Remote Sensing* 27 (2): 293–305. https://doi.org/10.1080/01431160500214050

Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Riberio, and D. Rossiter. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty." *SOIL*, 1–37. https://doi.org/10.5194/soil-2020-65

RQMS. 2021. *Analyses Physico-Chimiques Des Sites Du Réseau de Mesures de La Qualité desSols (RMQS) Du Territoire Métropolitain Pour La 1ère Campagne (2000-2009), Avec Coordonnées Théoriques.* Edited by INRAE. https://dx.doi.org/10.15454/QSXKGA

Sorensen, R., U. Zinko, and J. Seibert. 2006. "On the Calculation of the Topographic Wetness Index: Evaluation of Different Methods Based on Field Observations." *Hydrology and Earth System Sciences* 10: 101–112. https://doi.org/10.5194/hess-10-101-2006

Wellbrock N, Lehtonen A, Mäkipää R, Salmivaara A and Nieminen TM 2021. *D3.1 List of confirmed data & data providers*, 3 p. https://holisoils.eu/deliverables/

Woodhouse, I. 2005. *Introduction to Microwave Remote Sensing*. CRC Press, 400 p. https://doi.org/10.1201/9781315272573

Wright, Marvin N., and Andreas Ziegler. 2017. "Ranger: A Fast Implementation of Random Forests for High Dimensional Data in c++ and r." *Journal of Statistical Software* 77 (1): 1–17. https://doi.org/10.18637/jss.v077.i01

Ziche, Daniel, Erik Grüneberg, Lutz Hilbrig, Juliane Höhle, Thomas Kompa, Jari Liski, Anna Repo, and Nicole Wellbrock. 2019. "Comparing Soil Inventory with Modelling: Carbon Balance in Central European Forest Soils Varies Among Forest Types." *Science of The Total Environment* 647: 1573–85. https://doi.org/10.1016/j.scitotenv.2018.07.327