

Soil spectroscopy data

When using these data, please cite:

Mainka, M.; Bakker, L.; Maier, A.; Ruethers, J.; Bakke Westergaard, K.; Fior, S.; Frossard, A.; Griepentrog, M.; Van de Broek, M.; Wilken, F, Dötterl, S.; Alexander, J.; Magnabosco, C. (2025): Arctic Greening Database. v0.1. GFZ Data Services. <https://doi.org/10.5880/fidgeo.2025.074>

Introduction

We used quantitative soil spectroscopy to predict soil properties wherever possible. In the following, we provide more information on our modelling workflow, sample preparation, and model fits of the properties predicted in the tables *soil_organic_matter*, *soil_chem* and *soil_phys*.

Table 1. General information

project_name	ETH+ Arctic Greening
creation_date	2025-04-12
last_modification_date	2025-05-13
sampling_period	n/a
creator	Moritz Mainka
contact_1	moritz.mainka@usys.ethz.ch
contact_2	sdoetterl@usys.ethz.ch
number_samples	n/a
processing_status	n/a
upload_status	n/a

Sample preparation

Prior to scanning, soil samples were sieved to < 2 mm, oven-dried at 30 °C and ground to a homogeneous texture using a Roller-mill (similar to Arnold and Schepers 2004).

Spectrometer information

Samples were scanned on a Fourier Transform Infrared (FT-IR) Spectrometer with a High Throughput Screening Extension (HTS-XT) using a 24-well plate (Bruker Optics, Vertex 70, Germany). Samples were scanned in triplicates in the mid-infrared spectral range between 7500 - 600 cm⁻¹ (~ 1330 – 16670 nm). Each sample was scanned 32 times and then averaged. Normalization was done using a gold background (Infragold NIR-MIR Reflectance Coating, Labsphere) with correction for atmospheric CO₂ and H₂O using the OPUS Spectrometer software (Bruker Optics, Germany). On each plate a project-specific reference soil was measured as well

as two laboratory standards at the beginning and end of each measurement day (namely the CSIRO standards “Lucky Bay 2” and “Wylie Bay”).

Preprocessing and calibration sampling

All subsequent spectral processing was performed in R Version 4.4.2 (R Core Team 2024) using *prospectr* v0.2.7 (Stevens and Ramirez-Lopez 2024), *opusreader2* v0.6.3 (Baumann, Knecht, Roudier 2024), *resemble* v2.2.3 (Ramirez-Lopez et al. 2024), and *proximater* (R package under development used with the permission of L. Ramirez-Lopez).

First, spectra were resampled to a resolution of 2 cm^{-1} and cut to $3997 - 600\text{ cm}^{-1}$ to correct for light scattering. Triplicate scans were averaged. Second, spectral signals were preprocessed using a Savitzky-Golay filter, using the first derivative with a first-order polynomial approximation and a window size of 23 (Savitzky and Golay 1964). Finally, the preprocessed spectra were analyzed regarding their dissimilarity using a Principal Component Analysis (PCA). The centered scores of the first 8 principal components were used to identify samples that best cover the variability within the spectral signal. For this purpose, the k-means sampling algorithm was used to select approximately 20 % of the total sample size for wet chemistry analyses (calibration samples). Complementary samples were selected manually as the k-means algorithm tends to not sample extreme values.

Predictive modelling

Prior to building the model, the resampled and averaged spectra were preprocessed to minimize noise and amplify the signal. For this purpose, a second-order Savitzky-Golay filtering (window size = 25 cm^{-1}) and the first derivative calculated. A repeated 5-fold cross-validation (for pH: 6-fold) was used to estimate the prediction error within the calibration samples. The partial least squares regression (PLSR) used different numbers of components depending on the available number of calibration samples.

If needed, outliers were carefully removed in the process of fitting the model. Finally, the respective model was used to predict the soil property of interest based on the available spectral information. The model validity depended on the soil properties. For certain soil properties (see Table 2 for details), some calibration samples were excluded from building the model which limits the validity of the model. In the respective columns in the data tables, these samples have NA values.

Table 2: Model metrics used to predict specific soil properties. Root mean squared error (RMSE) has the same unit as indicated for each soil property. The model validity column indicates to which extent the model validity is limited. If certain types of samples were excluded from building the model, the model validity is restricted.

Soil property	Unit	Calibration samples	RMSE	R ²	Model validity
Total C	%	76	1.59	0.96	Mineral + organic
Total N	%	64	0.15	0.94	Mineral + organic
d15N	‰	97	4.35	0.74	Organic (excl. litter)
pH	-	111	0.36	0.77	Mineral + organic
CEC_eff	cmol _c kg ⁻¹	52	3.29	0.93	Mineral + organic
Silt + clay	%	48	11.61	0.73	Mineral + organic
Fe_DCB	g kg ⁻¹	53	2.28	0.77	Mineral
Mn_AO	g kg ⁻¹	53	0.08	0.81	Mineral
Ca_XRF	µg g ⁻¹	47	2399.35	0.95	Mineral + organic (excl. settlement samples)
Mg_XRF	µg g ⁻¹	57	2347.46	0.84	Mineral + organic
K_XRF	µg g ⁻¹	57	2862.34	0.87	Mineral + organic
Fe_XRF	µg g ⁻¹	50	4636.41	0.76	Mineral + organic (excl. settlement sites)
Al_XRF	µg g ⁻¹	57	8126.89	0.92	Mineral + organic
Si_XRF	µg g ⁻¹	56	15996.78	0.96	Mineral + organic

References

- Arnold, S. L., and J. S. Schepers. "A simple roller-mill grinding procedure for plant and soil samples." *Communications in Soil Science and Plant Analysis* 35, no. 3-4 (2004): 537-545. <https://doi.org/10.1081/CSS-120029730>
- Baumann P, Knecht T, Roudier P (2024). `_opusreader2`: Read Spectroscopic Data from Bruker OPUS Binary Files_. R package version 0.6.3, <https://github.com/spectral-cockpit/opusreader2>
- R Core Team (2024). `_R`: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ramirez-Lopez, L., and Stevens, A., and Viscarra Rossel, R., and Shen, Z., and Wadoux, A., and Breure, T. (2024). `resemble`: Regression and similarity evaluation for memory-based learning in spectral chemometrics. R package Vignette R package version 2.2.3
- Savitzky, Abraham, and Marcel JE Golay. "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry* 36, no. 8 (1964): 1627-1639. <https://doi.org/10.1021/ac60214a047>
- Stevens, A., and Ramirez-Lopez, L. (2024). An introduction to the `prospectr` package. R package Vignette R package version 0.2.7.