Artificial intelligence and machine learning in soil spectroscopic modelling

Dr. Zefang Shen and Prof. Raphael Viscarra Rossel

Soil and Landscape Science
Curtin University, Australia

5th July 2023

Soil & Landscape Science Group
Curtin University

GLOBAL SOIL PARTNERSHIP

Food and Agriculture Organization of the United Nations

Outline

Diffuse reflectance spectroscopy

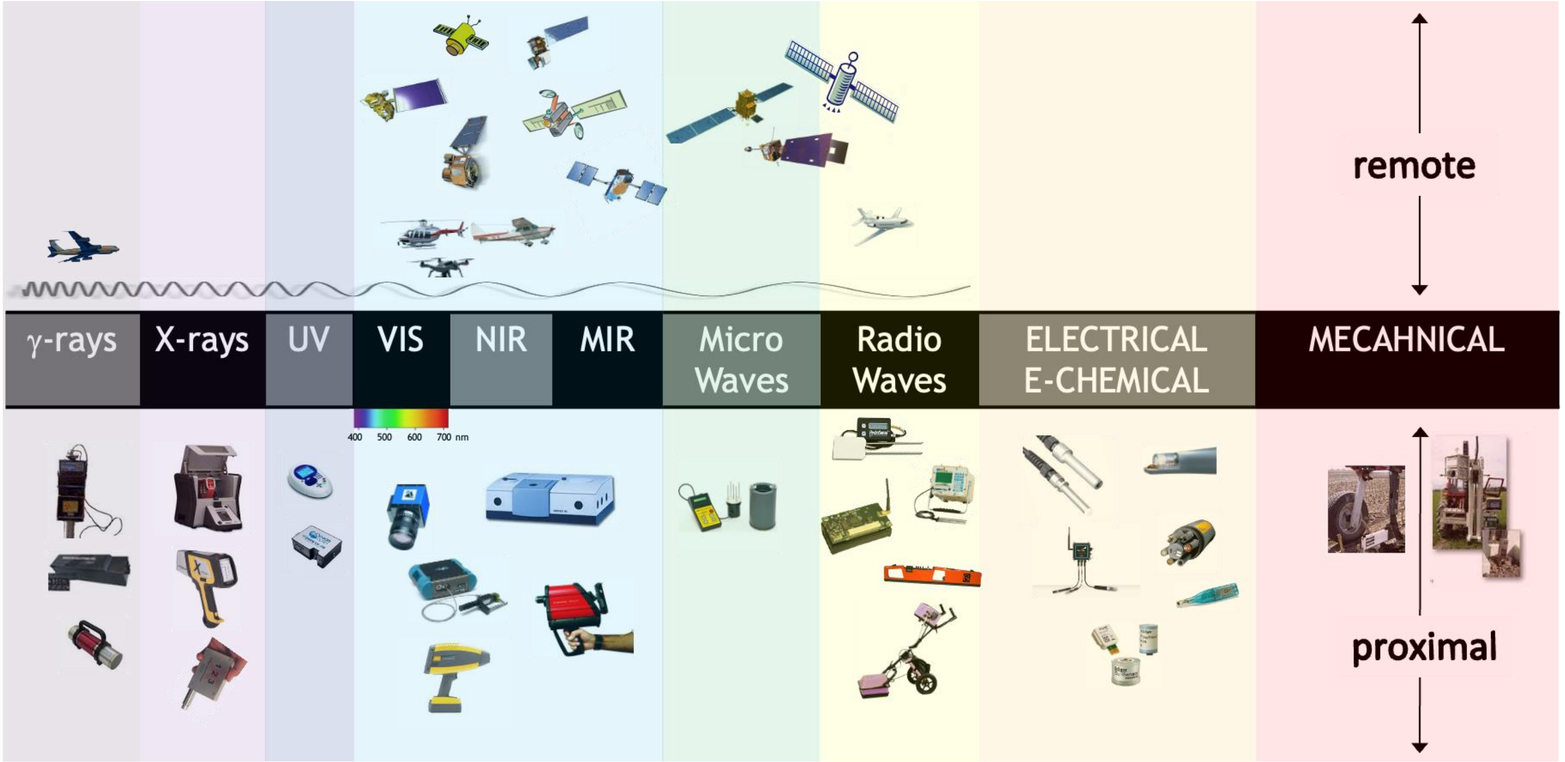
AI and machine learning

- Examples in soil spectroscopic modelling

Challenges

- Hyperparameter tuning
- Interpretation
- Localisation

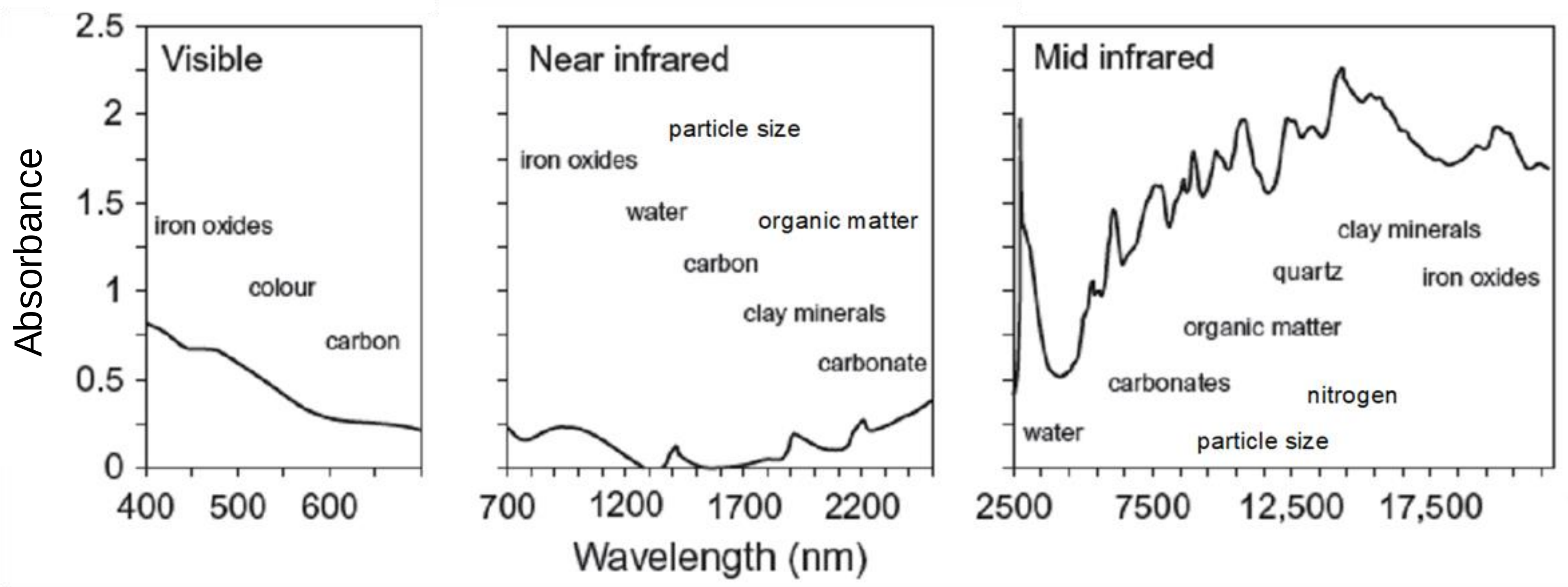
Spectroscopy



Sensors are becoming smaller, smarter, cheaper, faster, more energy efficient…

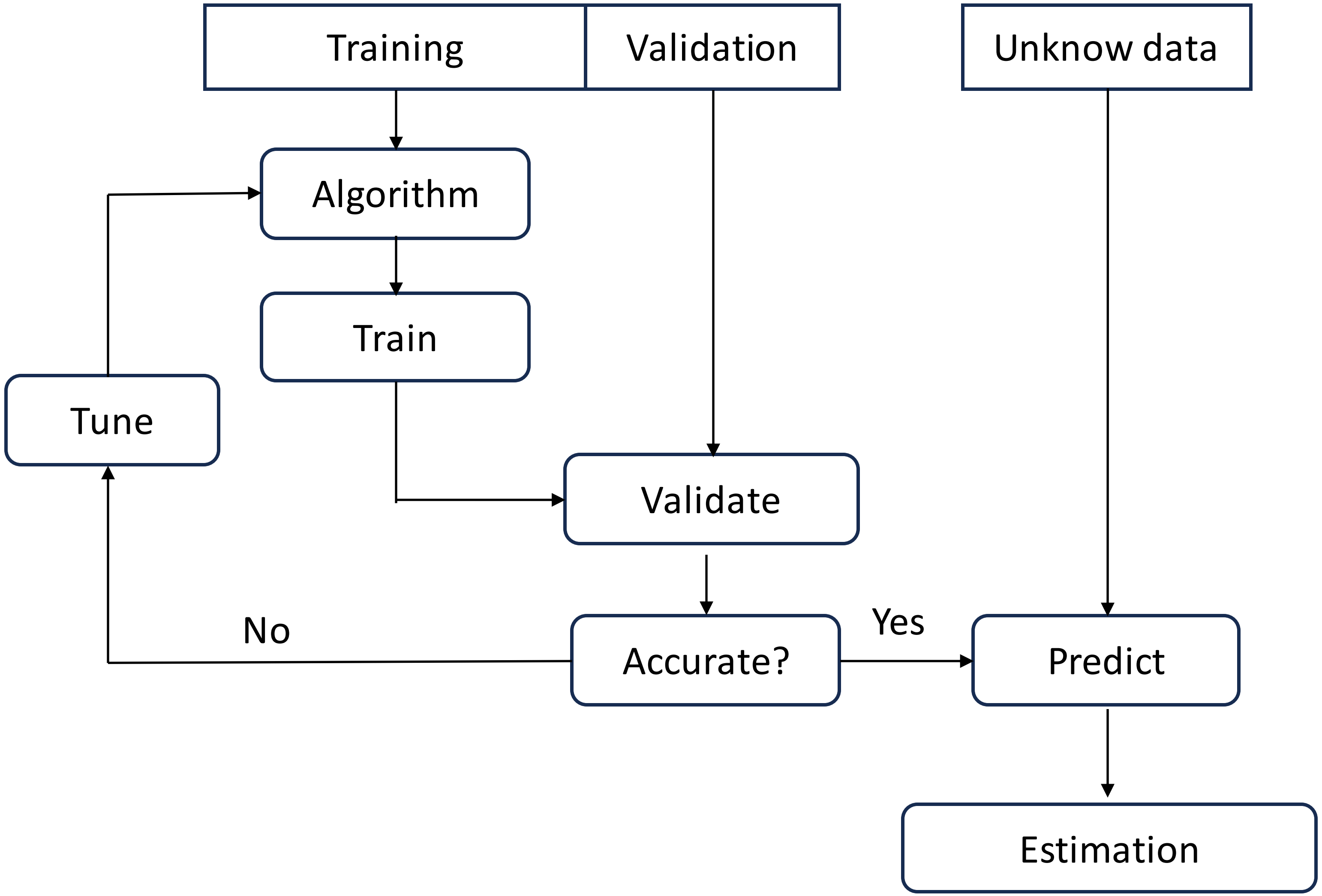
Soil diffuse reflectance spectroscopy

Spectra measure the composition of soil which determines soil properties



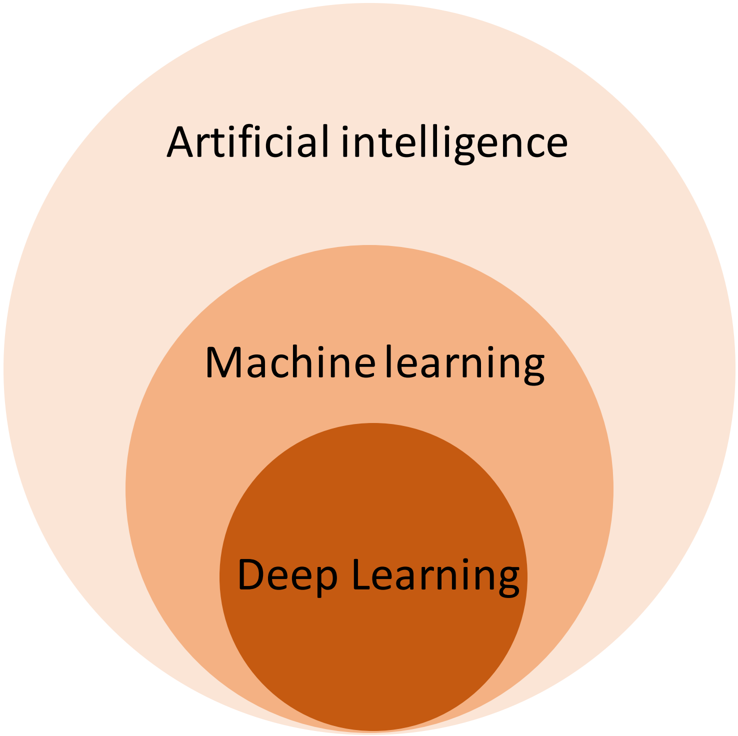
A single spectrum can provide information on the soil and its properties

Soil spectroscopic modelling - Multivariate regression



- A soil spectral library with spectra and corresponding soil properties.
- Statistical and machine learning can be used as the predictive models.
- Hyperparameters of the models need to be tuned for accurate predictive models.

Artificial intelligence and machine learning



…and also **SOIL SCIENCE**

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

ELSEVIER

Deep transfer learning of global spectra for local soil carbon monitoring

ez[b], Thorsten Behrens[c,d], Lei Cui[e], Mingxi Zhang[a], , Zhou Shi[g], Kenneth A Sudduth[h], Philipp Baumann[d], aphael A. Viscarra Rossel[a,*]

scientific reports

OPEN **Automated spectroscopic modelling with optimised convolutional neural networks**

Zefang Shen & R. A. Viscarra Rossel

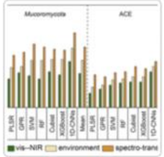
nature geoscience — ARTICLES

https://doi.org/10.1038/s41561-019-0373-z

Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls

[1,2], T. Behrens[3], Z. Luo[4], J. Baldock[5] and A. Richards[6]

Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions



Yuanyuan Yang[1], Zefang Shen[1], Andrew Bissett[2], and
[1]Soil and Landscape Science, School of Molecular and Life Sc
[2]CSIRO Oceans and Atmosphere, GPO BOX 1538, Hobart TAS

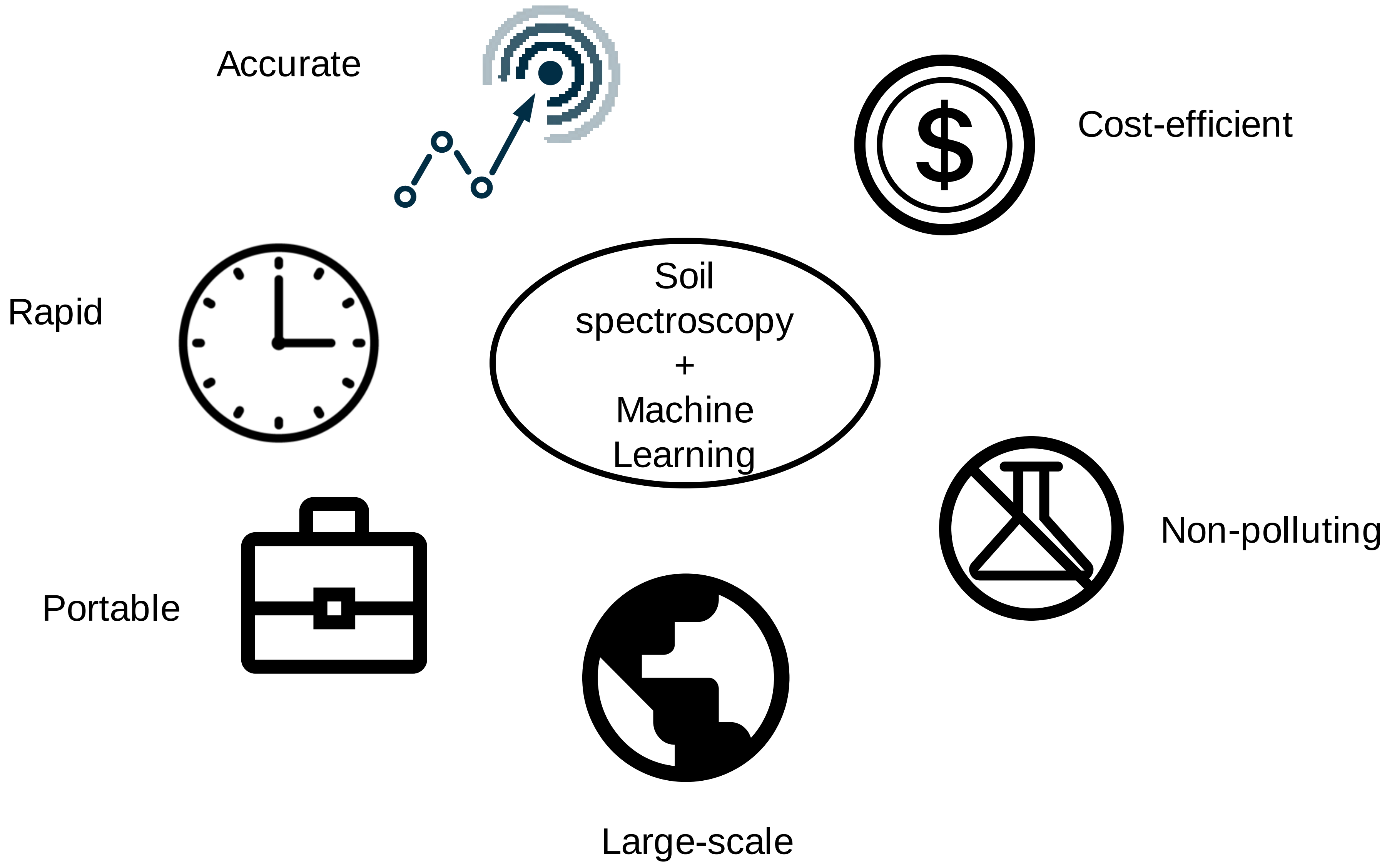
Correspondence: Raphael A. Viscarra Rossel (r.viscarra-rosse

SCIENTIFIC REPORTS

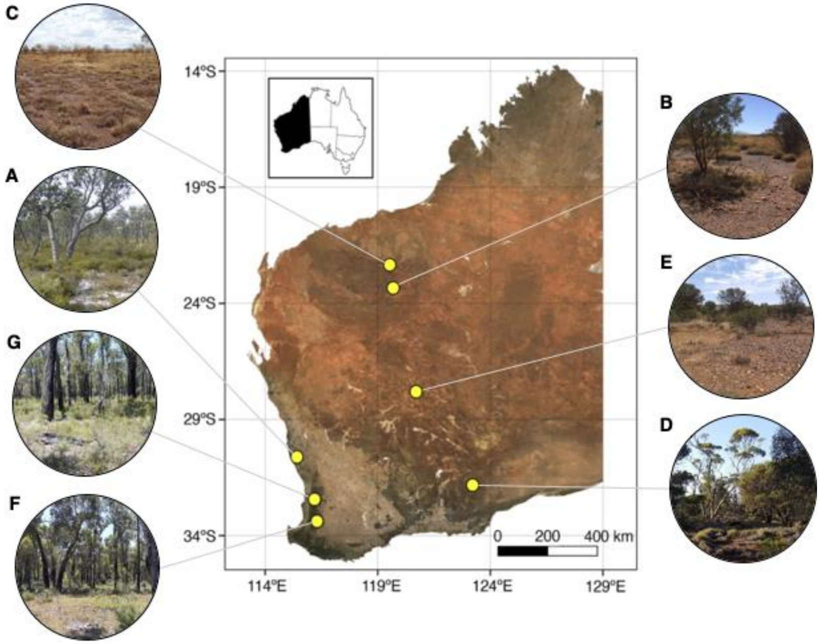
OPEN **Multi-scale digital soil mapping with deep learning**

Thorsten Behrens[1], Karsten Schmidt[1], Robert A. MacMillan[2] & Raphael A. Viscarra Rossel[3]

Soil spectroscopy with Machine Learning



Soil spectroscopy for mine site rehabilitation



56 sampling plots, 280 subplots.

SOIL, 8, 467–486, 2022
https://doi.org/10.5194/soil-8-467-2022

SOIL Open Access EGU

Miniaturised visible and near-infrared spectrometers for assessing soil health indicators in mine site rehabilitation

Zefang Shen[1], Haylee D'Agui[2], Lewis Walden[1], Mingxi Zhang[1], Tsoek Man Yiu[1], Kingsley Dixon[2], Paul Nevill[2,3], Adam Cross[2,4], Mohana Matangulu[1], Yang Hu[1], and Raphael A. Viscarra Rossel[1]

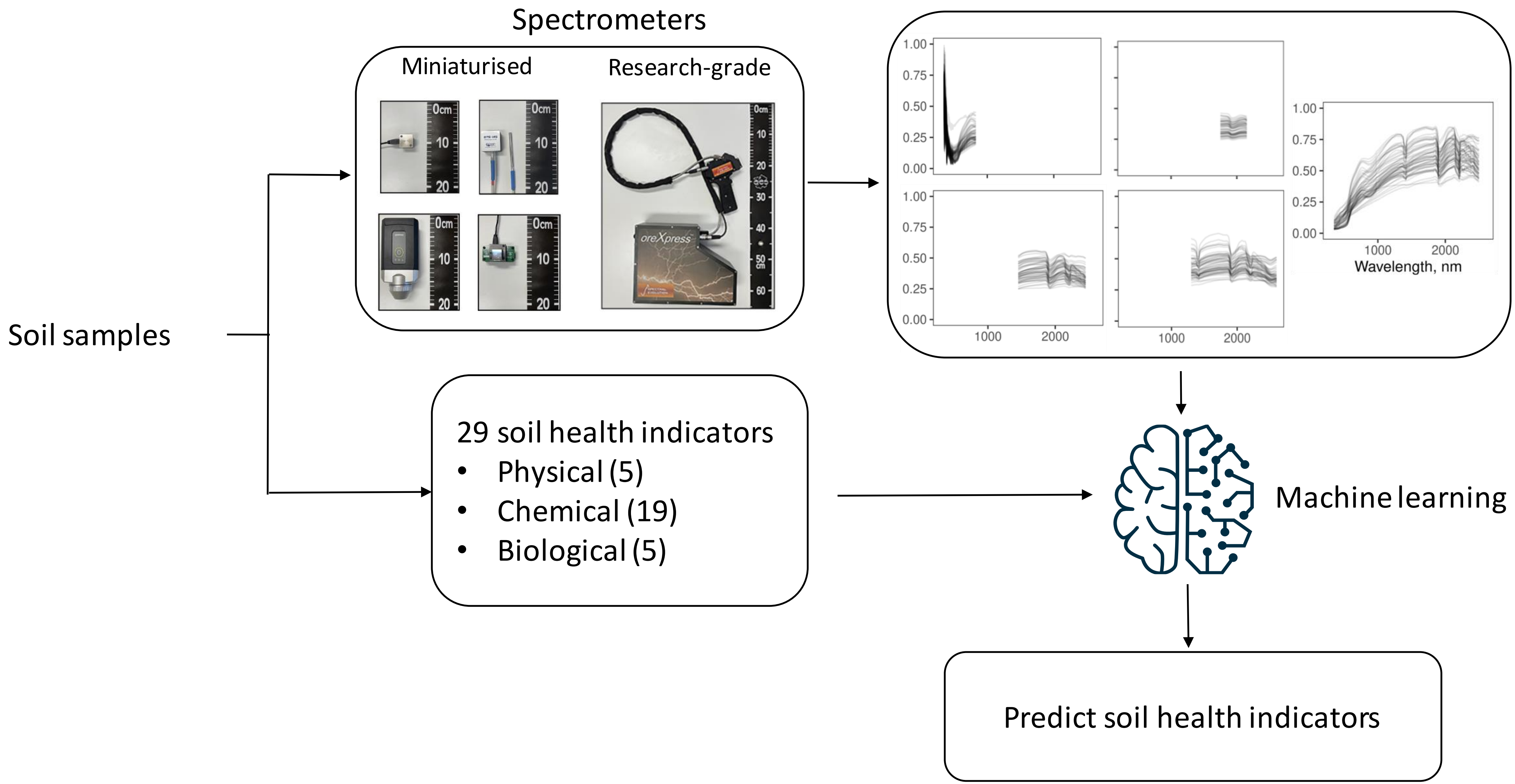
Physical property (5)	Unit	Mean	SD
Sand	%	51.44	12.38
Slit	%	24.10	12.22
Clay	%	24.46	17.12
Bulk density	g cm^{-3}	1.36	0.15
Electrical conductivity	dS m^{-1}	0.19	0.49

Biological property (5)	Unit	Mean	SD
CO_2 production	mg L^{-1}	30.41	29.22
Fungal richness	-	52.26	39.87
Fungal diversity	-	2.82	0.65
Bacterial richness	-	656.64	249.81
Bacterial diversity	-	5.58	0.60

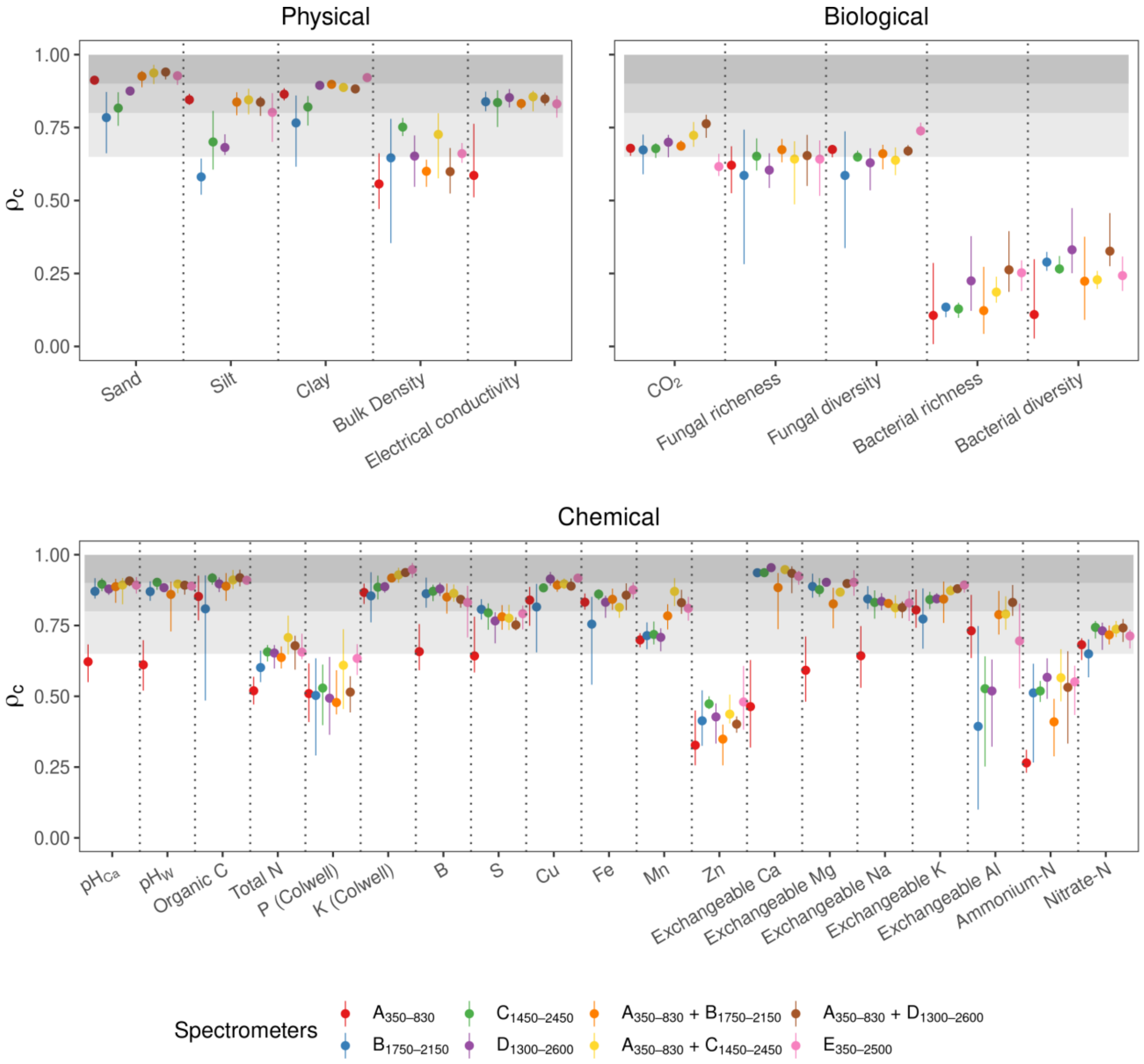
Chemical property (19)	Unit	Mean	SD
pH_{Ca}	-	5.74	1.20
pH_W	-	6.68	1.06
Organic C	%	1.21	1.07
Total N	mg kg^{-1}	14.52	23.50
P (Colwell)	mg kg^{-1}	4.53	2.62
K (Colwell)	mg kg^{-1}	191.71	145.15
B	mg kg^{-1}	0.73	1.04
S	mg kg^{-1}	60.13	225.00
Cu	mg kg^{-1}	0.84	0.72
Fe	mg kg^{-1}	18.24	16.89
Mn	mg kg^{-1}	13.94	15.64
Zn	mg kg^{-1}	0.56	0.35
Exchangeable Ca	meq 100 g^{-1}	6.09	5.67
Exchangeable Mg	meq 100 g^{-1}	1.91	1.81
Exchangeable Na	Meq 100 g^{-1}	0.65	1.53
Exchangeable K	meq 100 g^{-1}	0.38	0.31
Exchangeable Al	meq 100 g^{-1}	0.12	0.14
Ammonium N	meq 100 g^{-1}	3.39	3.27
Nitrate N	meq 100 g^{-1}	11 .13	21.66

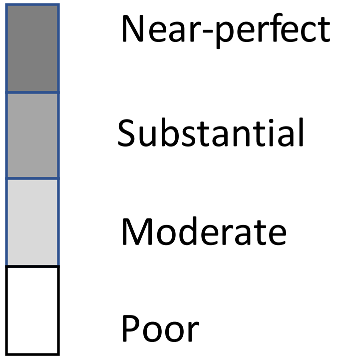
Shen et al., 2022

Soil spectroscopy for mine site rehabilitation



Soil spectroscopy for mine site rehabilitation





- The visible range spectrometer (A) accurately estimated soil texture (sand, silt, clay)
- The NIR spectrometers (B, C, D) estimated most of the soil properties with moderate or greater accuracy.
- Combining visible and NIR spectrometers produced more accurate estimates.
- Miniaturised spectrometers and their combinations estimated 24 out of 29 properties with moderate or greater accuracy.

Challenges in soil spectroscopic modelling

Hyperparameter tuning
Selection of hyperparameters are critical for the performance of machine learning models.

Interpretation
Machine learning models are usually considered as ‘black-boxes’ because they are difficult to interpret by humans.

Localisation
Models developed with large and diverse spectral datasets usually generalize poorly at more local scales.

Hyperparameter tunning

Grid search
Evaluate all possible hyperparameter configurations

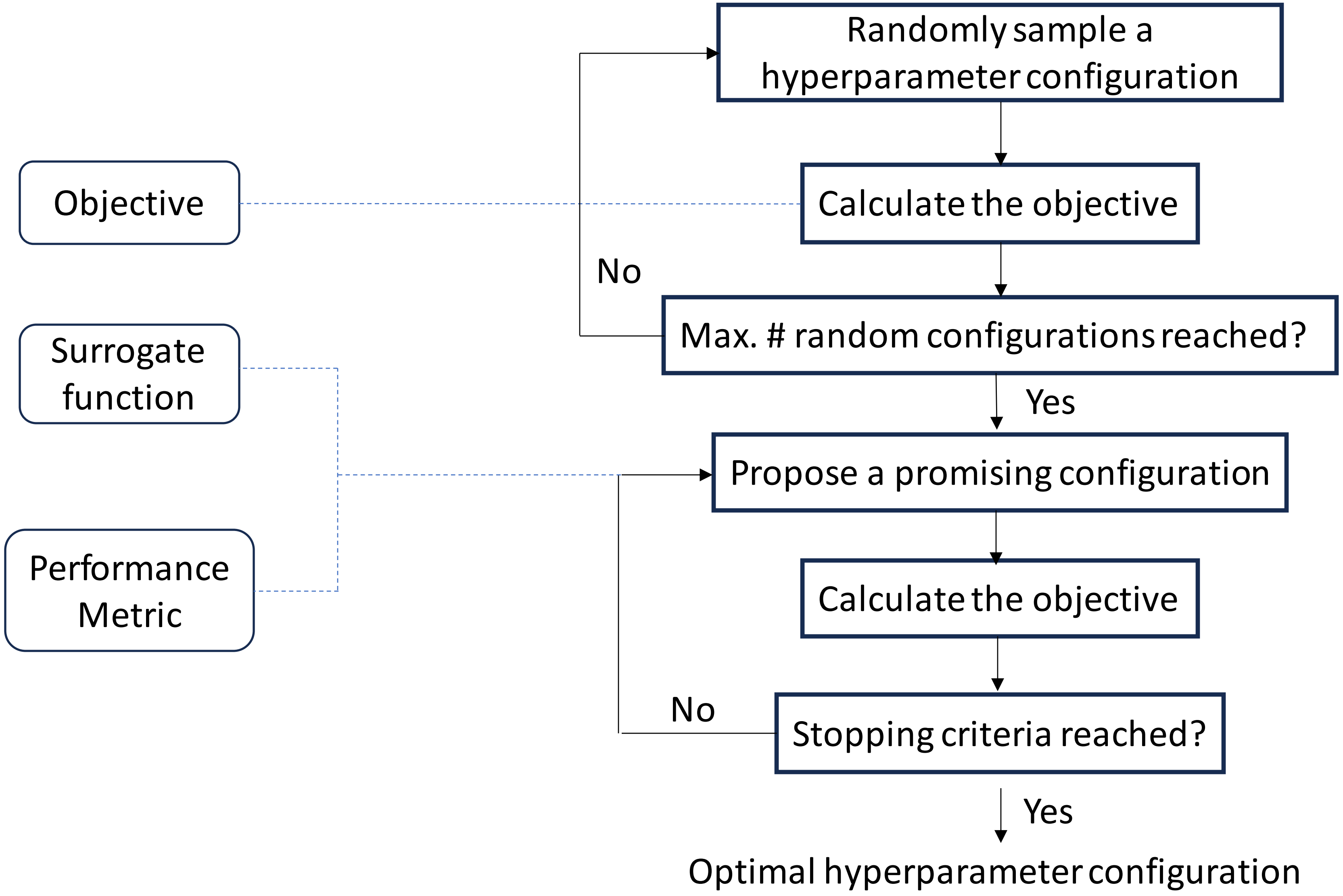
Random search
Evaluate randomly sampled hyperparameter configurations

Hyperparameter optimization, e.g. Bayesian optimization
Use information from the previous evaluations to guide the hyperparameter search.

Grid search

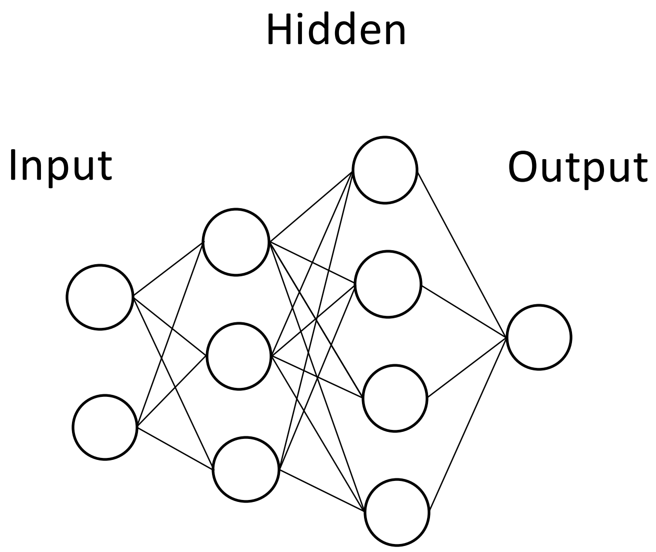
Algorithm	Number of Hyperparameters	Number of model evaluations
Partial least squared regression	1	~20
Cubist	2	~80
Convolutional Neural network	104	$> 2^{104} \sim 10^{31}$

Hyperparameter tunning – Bayesian optimisation

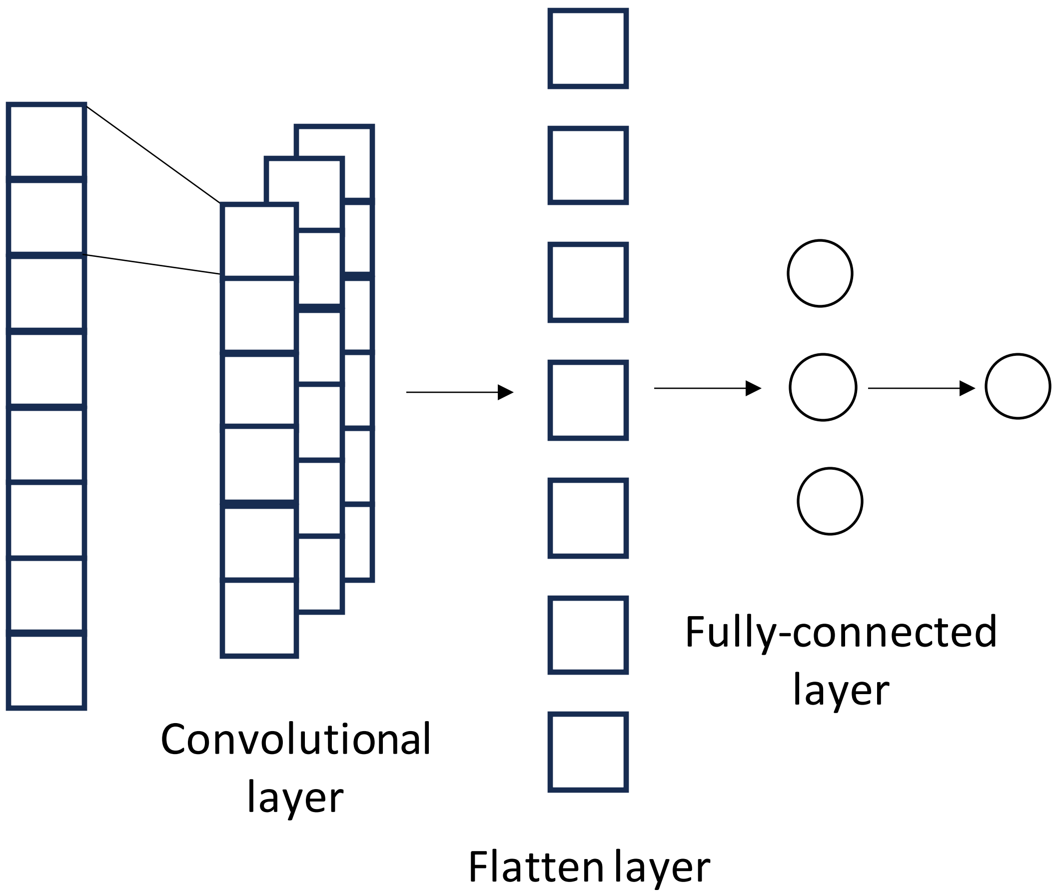


Hyperparameter tunning – Neural Networks

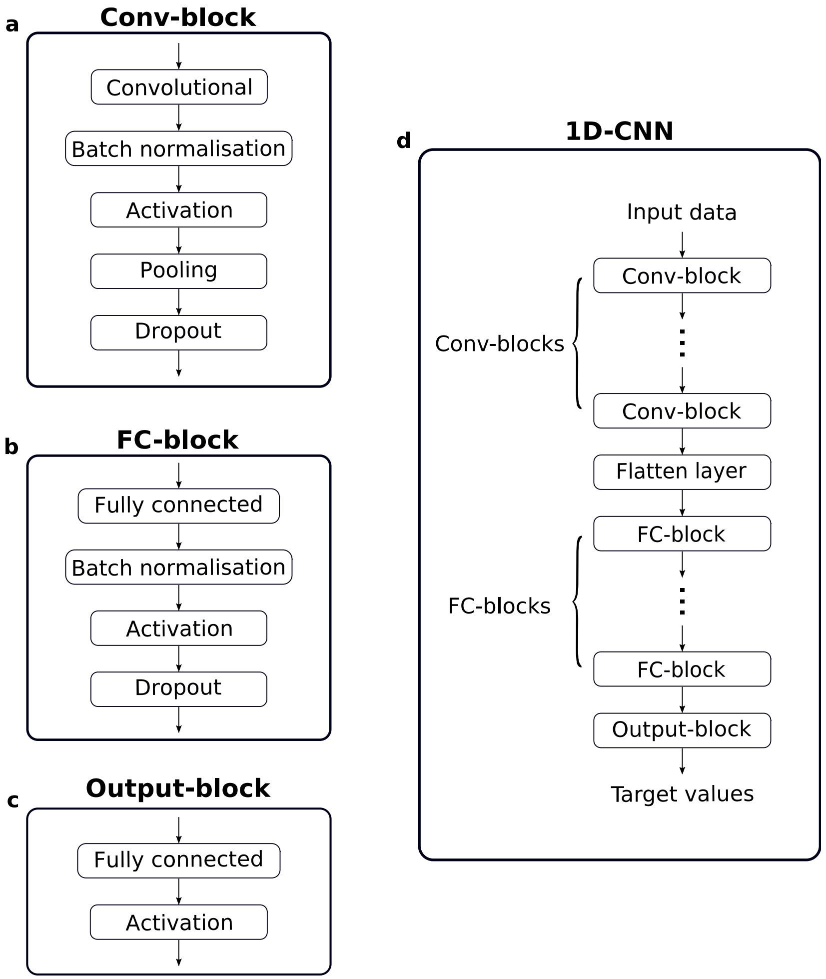
Artificial Neural Networks



One-dimensional convolutional neural networks

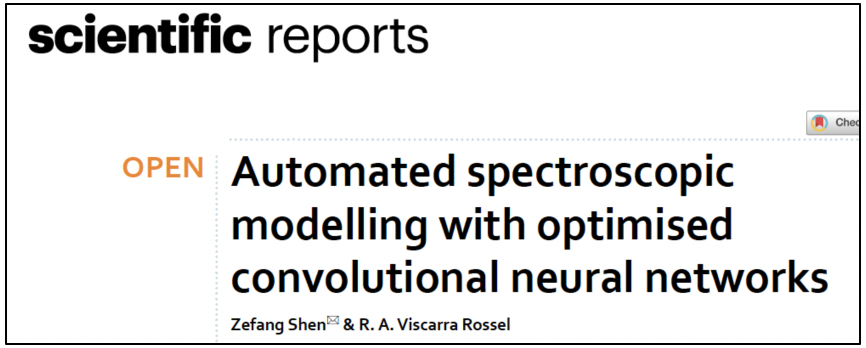


Hyperparameter tunning – Bayesian optimisation



Objective:
Mean RMSE from 10-fold cross-validation

Hyperparameters:
- # Conv-blocks
- # FC-blocks
- Hyperparameters in the Conv- and FC-blocks


scientific reports

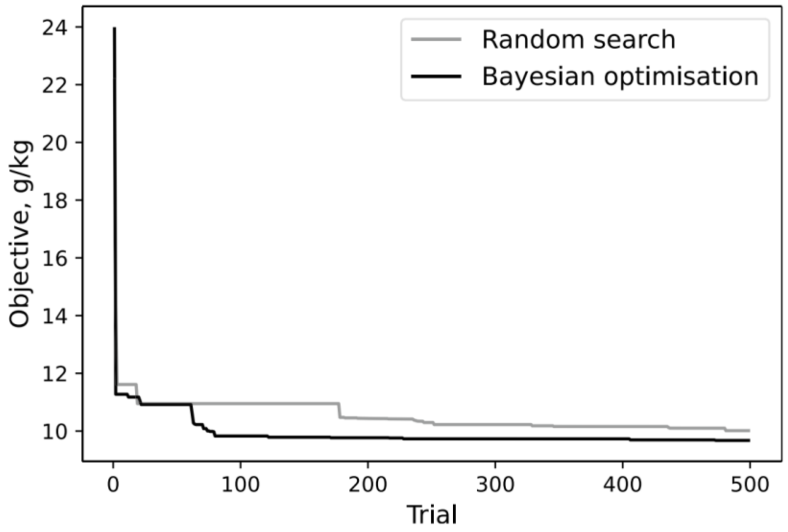
OPEN Automated spectroscopic modelling with optimised convolutional neural networks

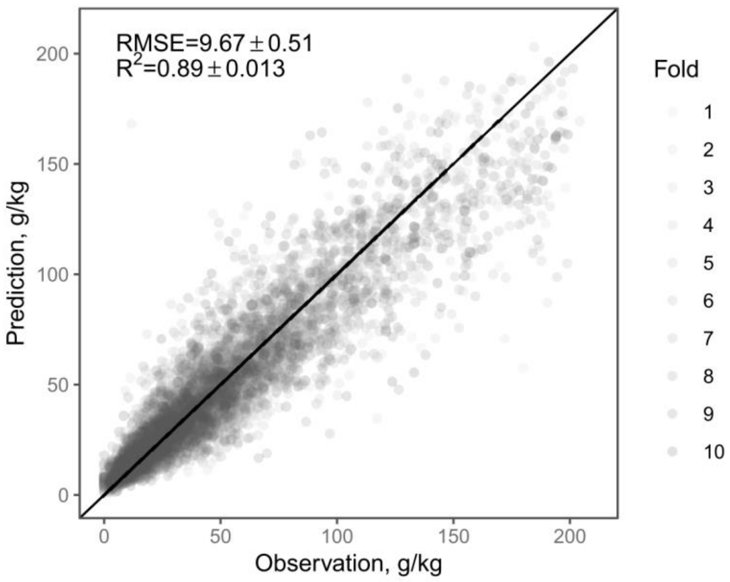
Zefang Shen & R. A. Viscarra Rossel

Building blocks are optimsed for best cross-validation performance using Bayesian optimisation.

Hyperparameter tunning – Bayesian optimisation

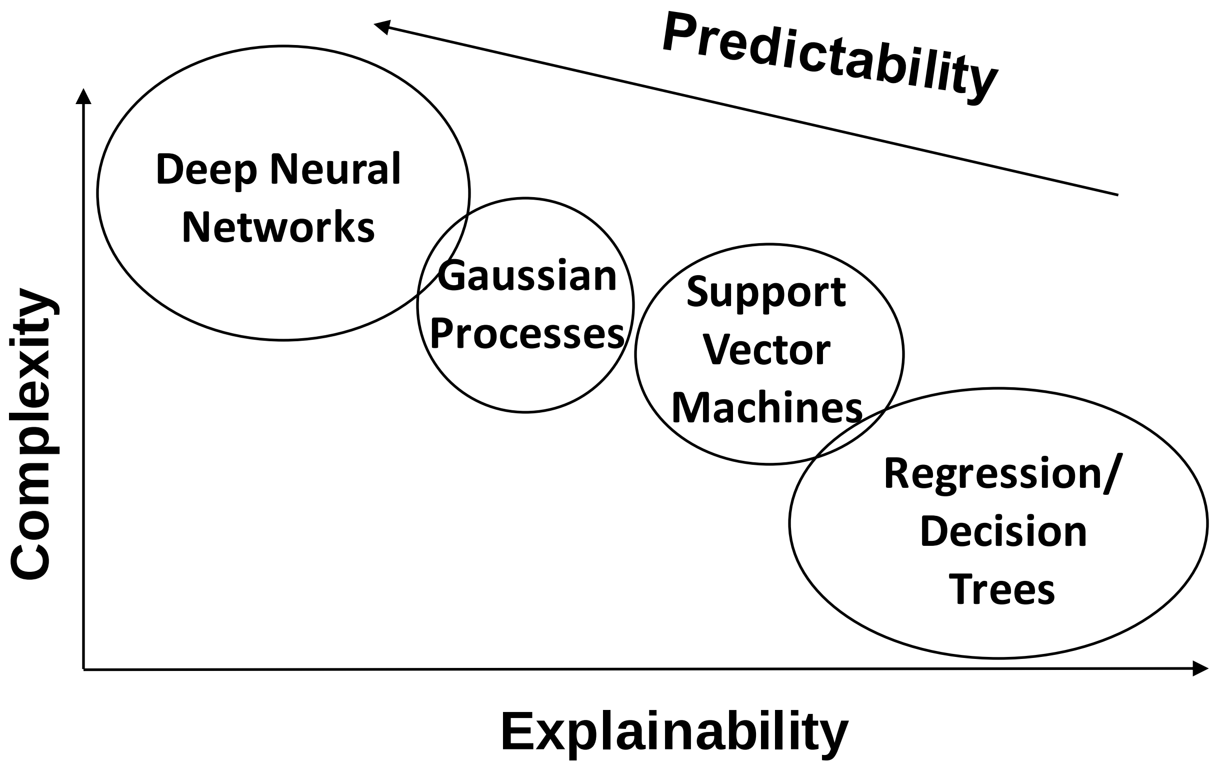
Optimal 1D-CNN on the LUCAS dataset.





- Bayesian optimization produced the most accurate 1D-CNN.
- Bayesian optimization take much less trials to converge.
- Bayesian optimization can automatically discover an optimal 1D-CNN with best accuracy.

Interpretation of machine learning



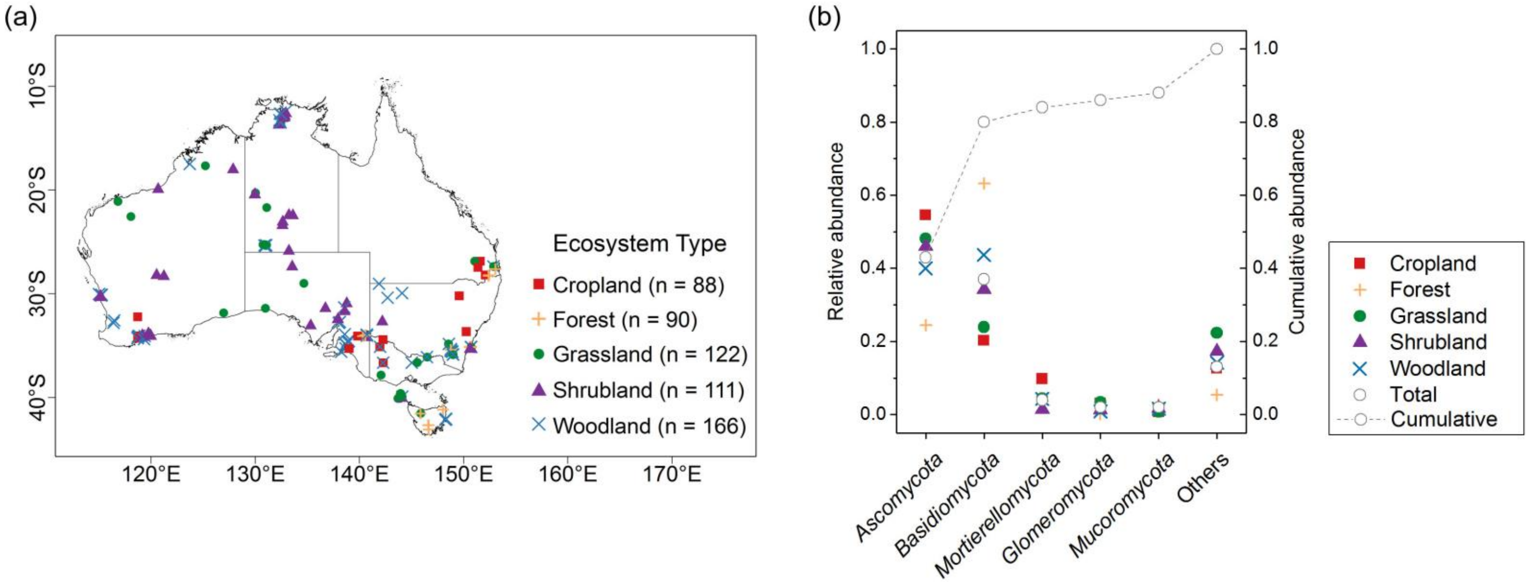
Explainable Artificial intelligence (XAI)
Tools and methods to help understand and interpret predictions made by your machine learning models

Examples:

- Perturbation-based feature importance
- SHapley Additive exPlanations (SHAP)

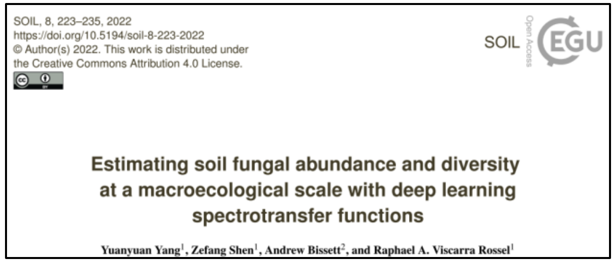
Understanding soil fungal abundance drivers

An example with perturbation-based XAI.



Models tested:

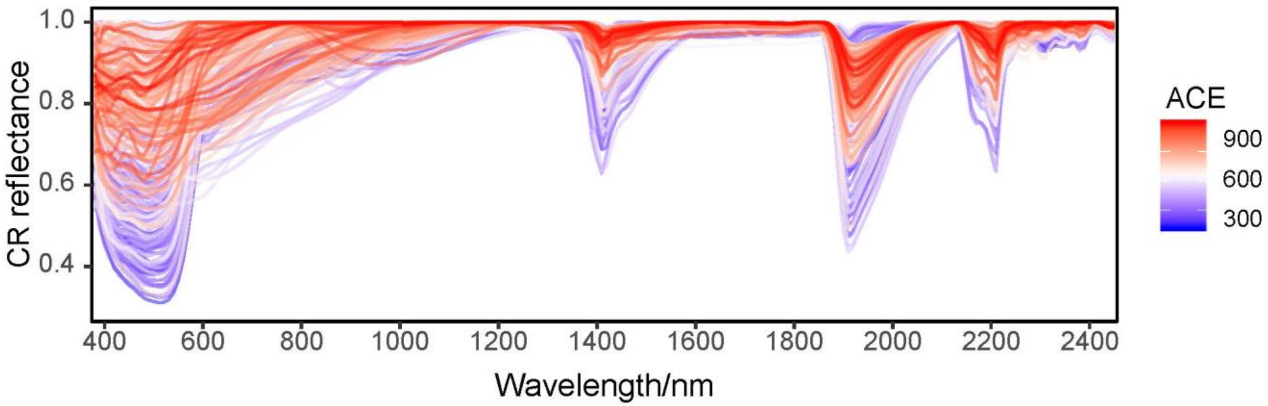
- Partial lest squared regression (PLSR)
- Support vector machines (SVM)
- Random forests (RF)
- eXtreme Gradient Boosting (XGBoost)
- 1D-CNN


SOIL, 8, 223–235, 2022
https://doi.org/10.5194/soil-8-223-2022
© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.

SOIL EGU

Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions

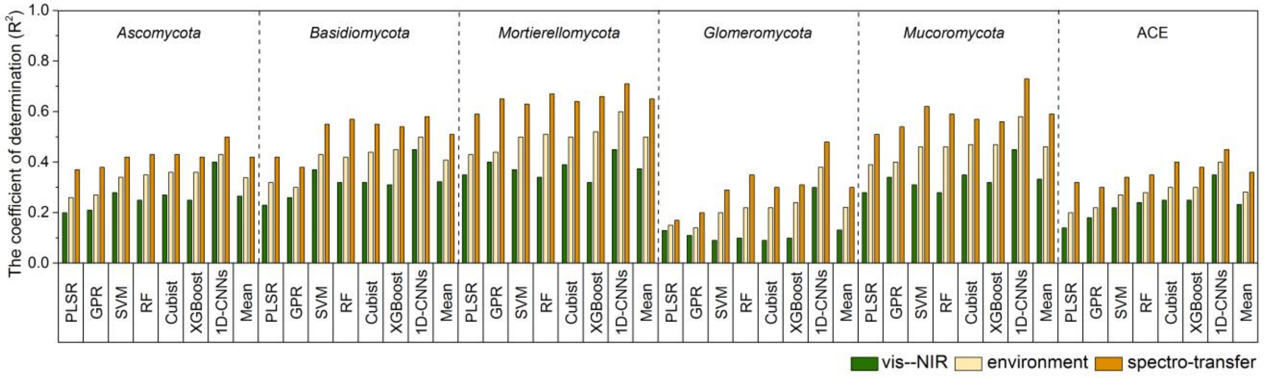
Yuanyuan Yang[1], Zefang Shen[1], Andrew Bissett[2], and Raphael A. Viscarra Rossel[1]



+

Climate
Terrain
Vegetation
Parent material

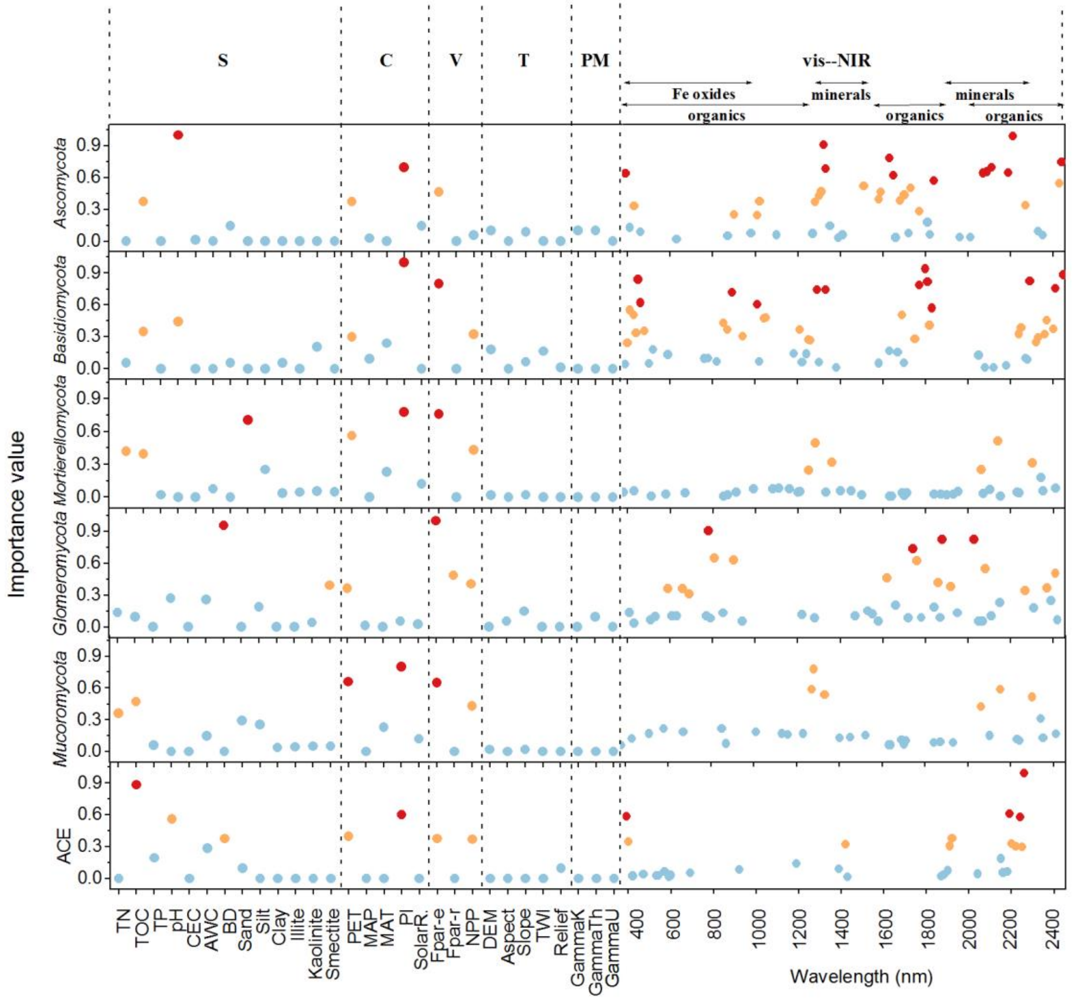
Understanding soil fungal abundancy drivers



1D-CNN produced the most accurate predictions for the fungal properties.

Perturbation-based Feature Importance

1. Perturb a predictor of interest.
2. Predict on validation data.
3. Calculate the changes in accuracy (R^2).



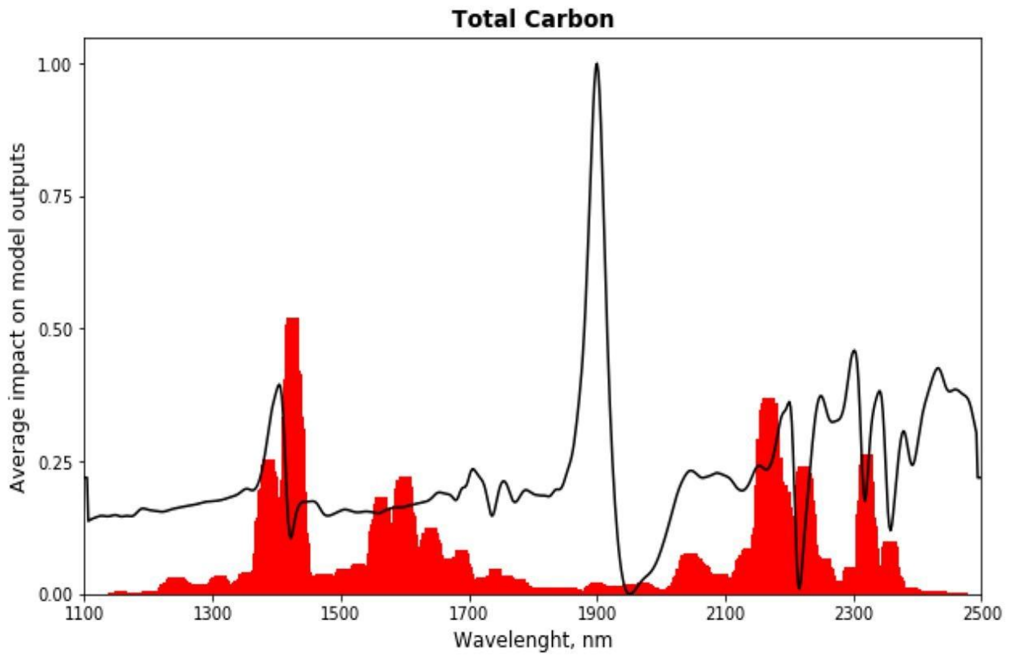
Interpreting soil spectroscopic modeling with SHAP values

SHAP values

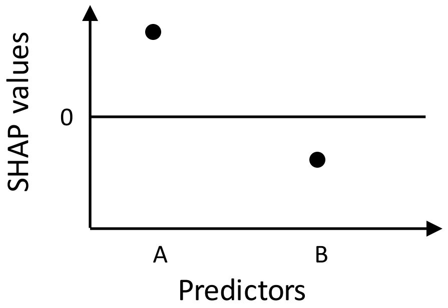
- SHAP values are based on cooperative game theory and derive from Shapley values.
- Determine the contribution of each feature for a prediction.
- A background dataset is used as a reference for calculating the expected value.

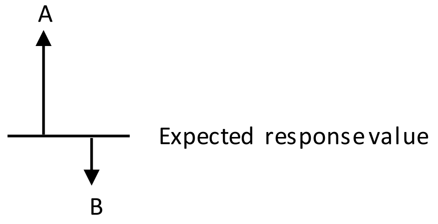
Understanding SHPA values

- Positive SHAP values indicate that the feature increases the prediction compared to the expected value
- Negative values indicate the opposite.



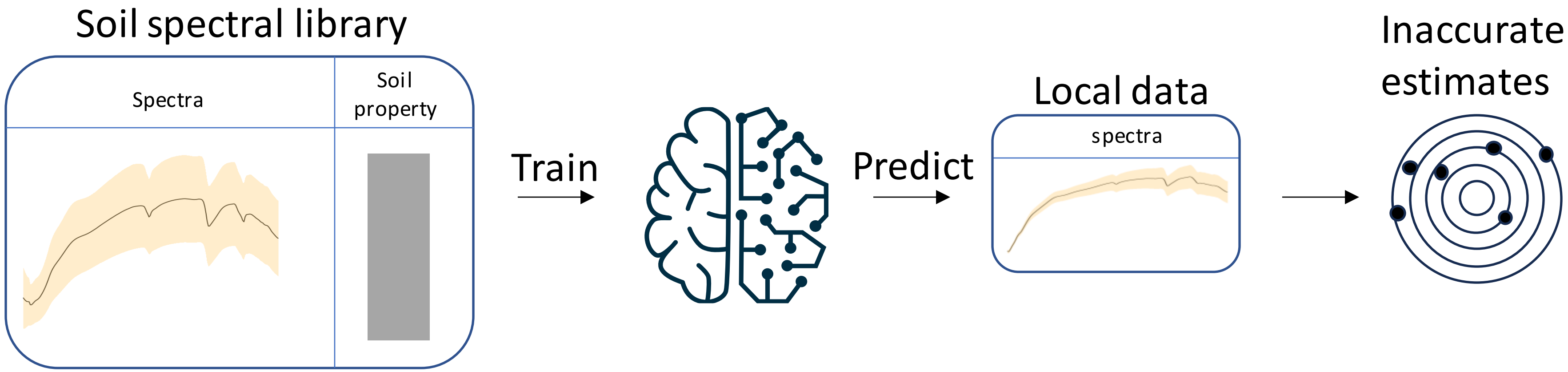
Haghi et al., 2021





Localising soil spectroscopic modelling

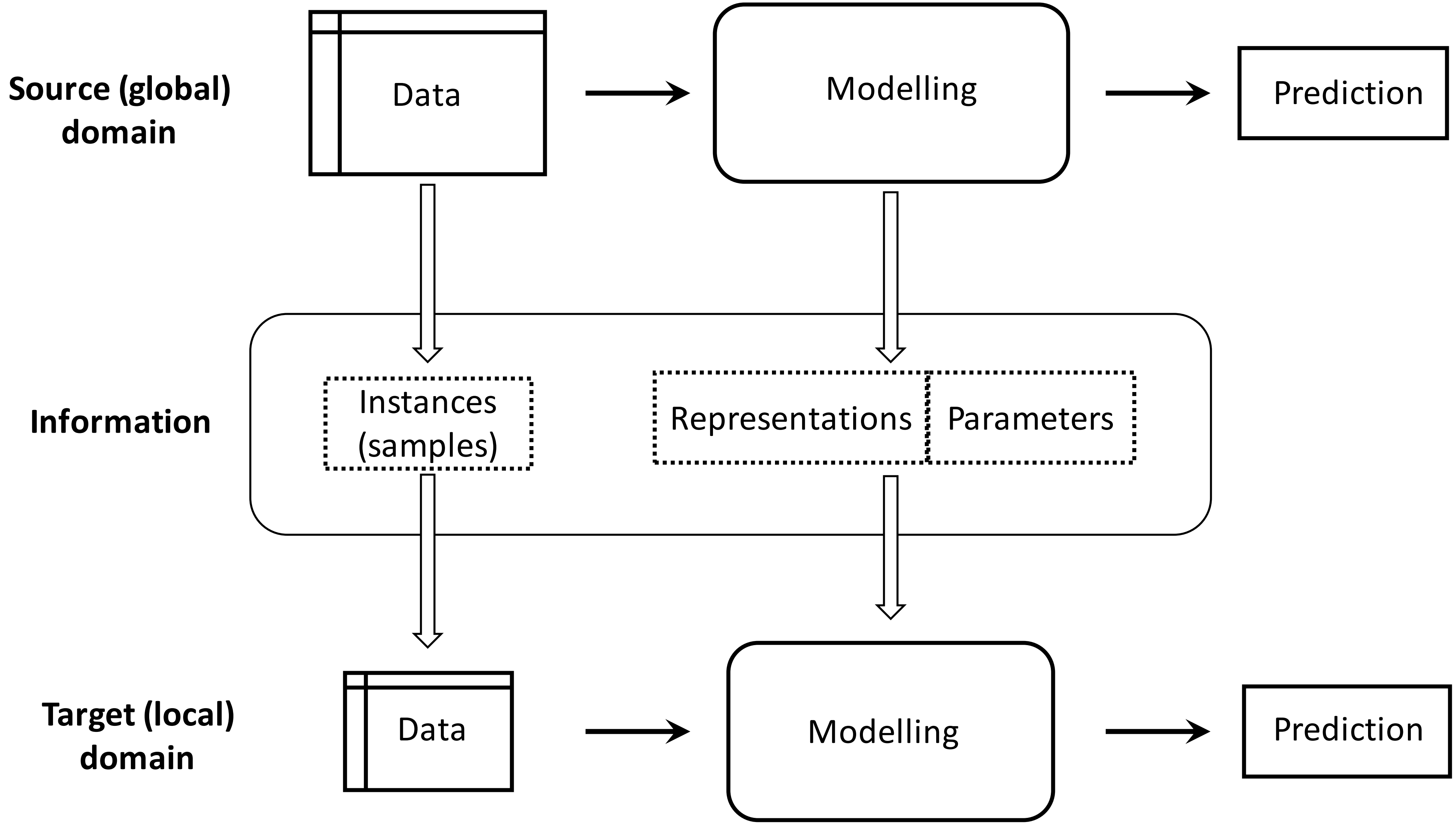
"Global" models built with large and diverse datasets do not generalise well on more homogeneous "local" data.



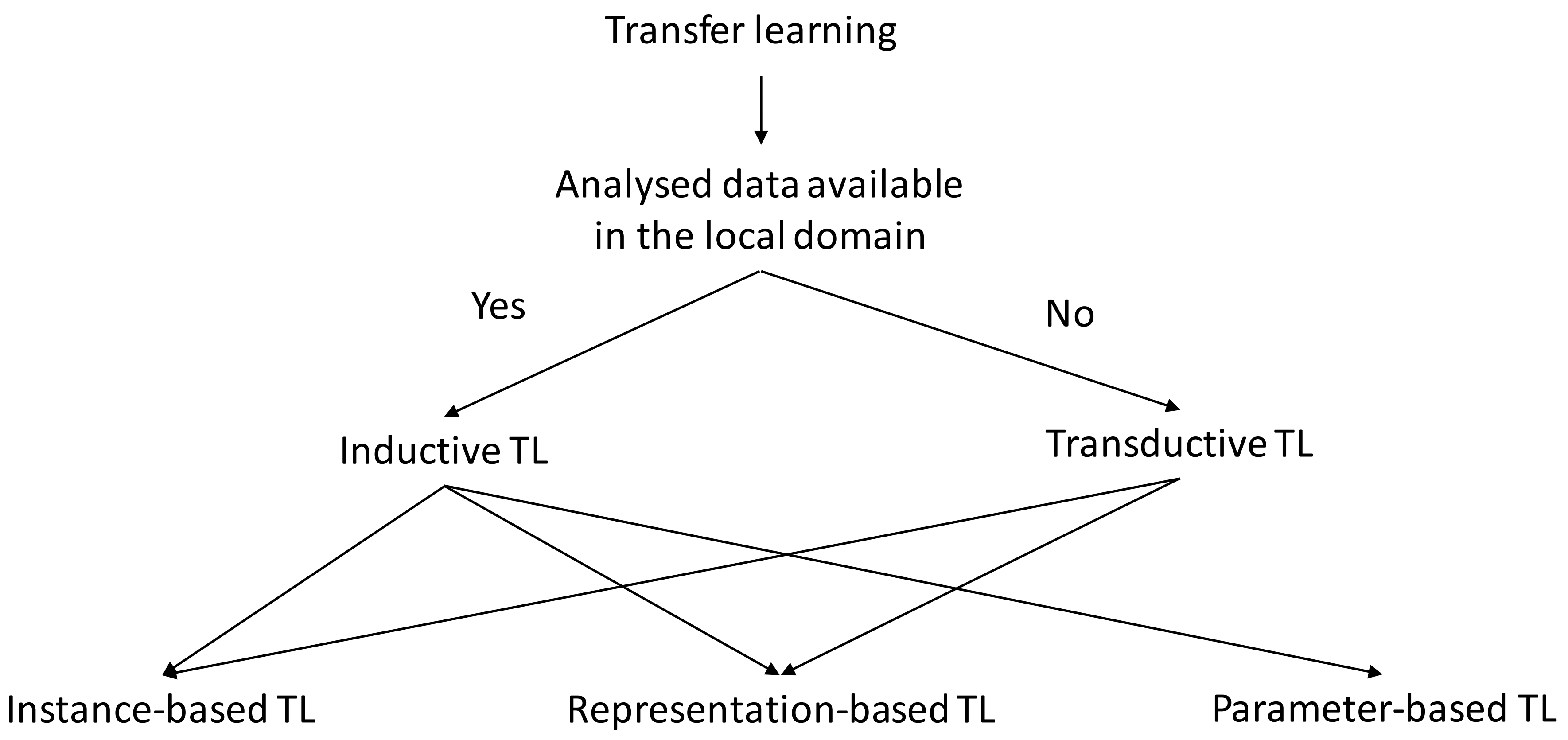
Localisation methods

- **Spiking**
  - Augment the local data with the soil spectral library
- **Conditional filtering**
  - Filter the soil spectral library with pedologic, geographic, land use information etc.
  - Augment the local data with the filtered library
- **Distance-based deterministic search**
  - Use distance metrics (e.g. Mahalanobis distance) to select spectral neighbours in the soil spectral library.
  - Develop spectroscopic modelling on the selected neighbours and predict on local data.
- **Data-driven heuristic search**
  - Generate subsets from the soil spectral library
  - Evaluate the subsets on the analysed local data
  - Augment the local data with the best subset
- **Reusing feature representations**
  - Train a neural network on the soil spectral library
  - Freese the early layers and re-train the neural network on local data

Localisation as a Transfer Learning problem



Classification of Transfer Learning



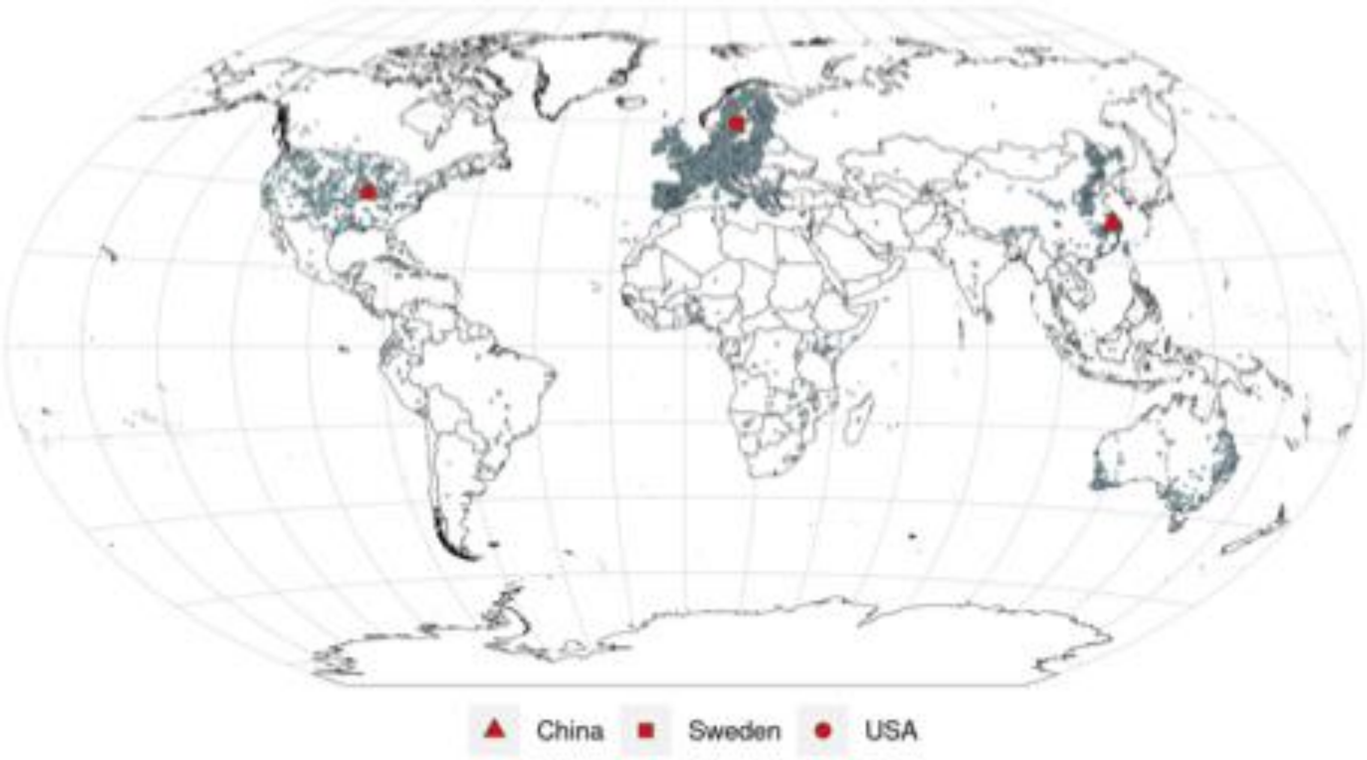
- Spiking
- Conditional filtering
- Distance-based deterministic search
- Data-driven heuristic search

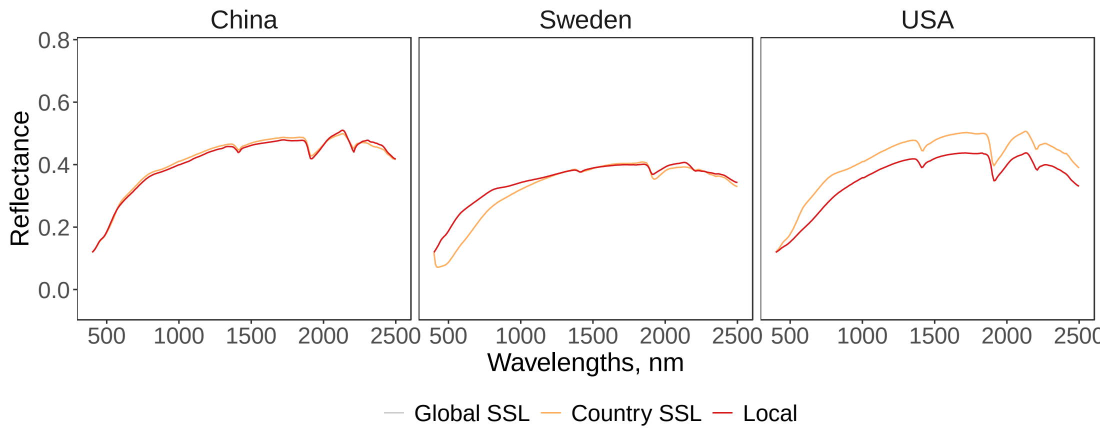
- Reusing feature representations

Localisation with Deep Transfer Learning (DTL)

Local soil organic carbon (SOC) modelling
Deep transfer learning for localising spectroscopic estimates of soil organic carbon at the farm-scale with a global soil spectral library (SSL).

	Dataset	# samples
SSL	Global	50,422
	China	5,183
	Sweden	2,319
	USA	4,155
Local	China	135
	Sweden	108
	USA	216



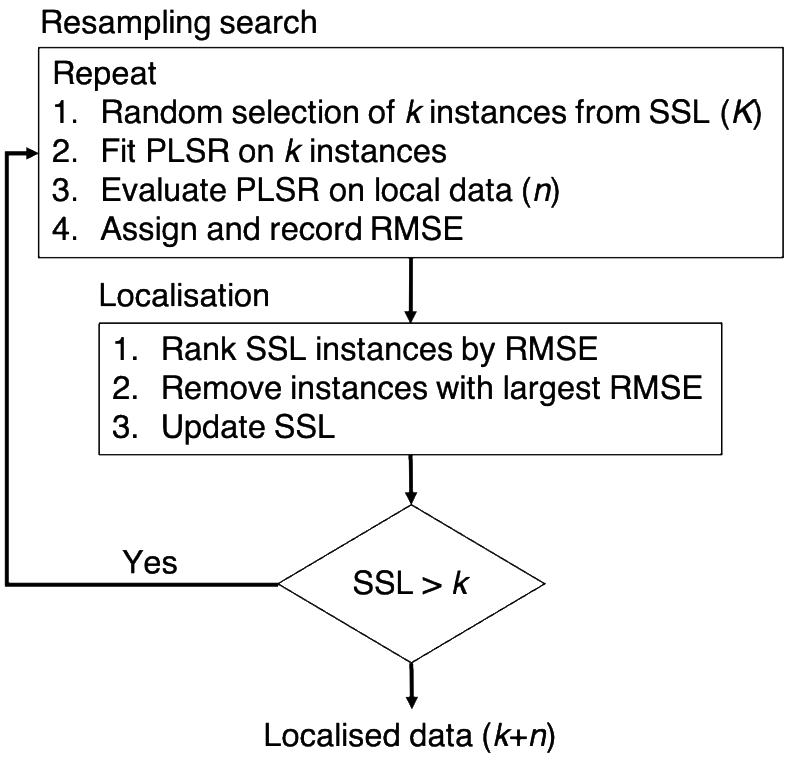


Shen et al. (2022), Viscarra Rossel et al. (2016).

Localisation with Deep Transfer Learning (DTL)

Transferring instances

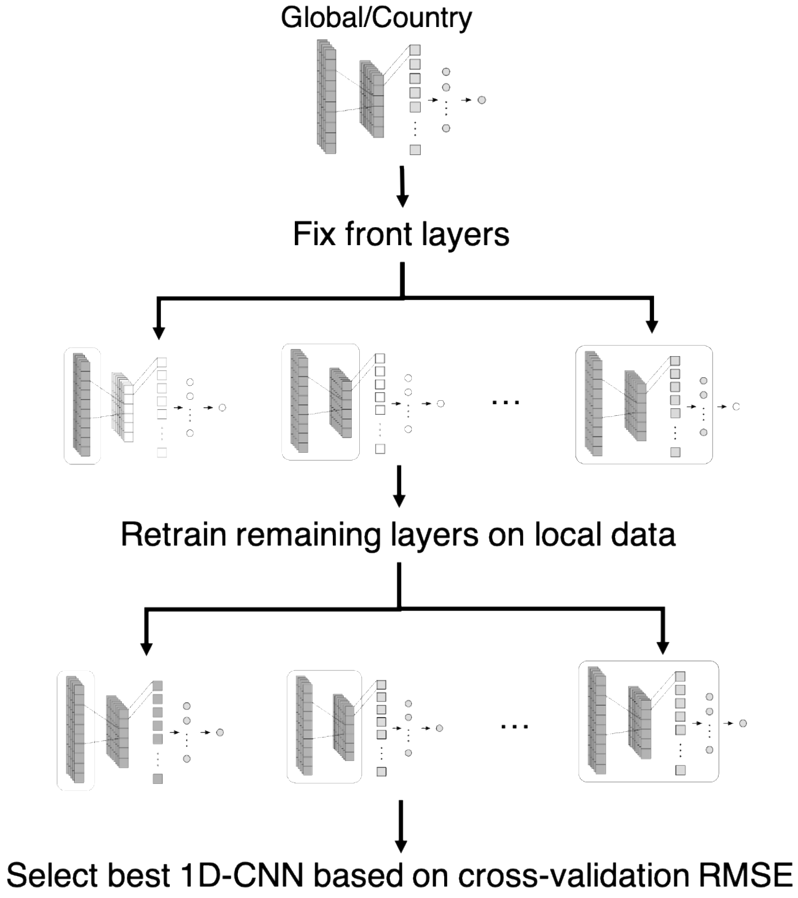
RS-LOCAL-v2.0



PLSR—partial least square regression;
RMSE—root means squared error.

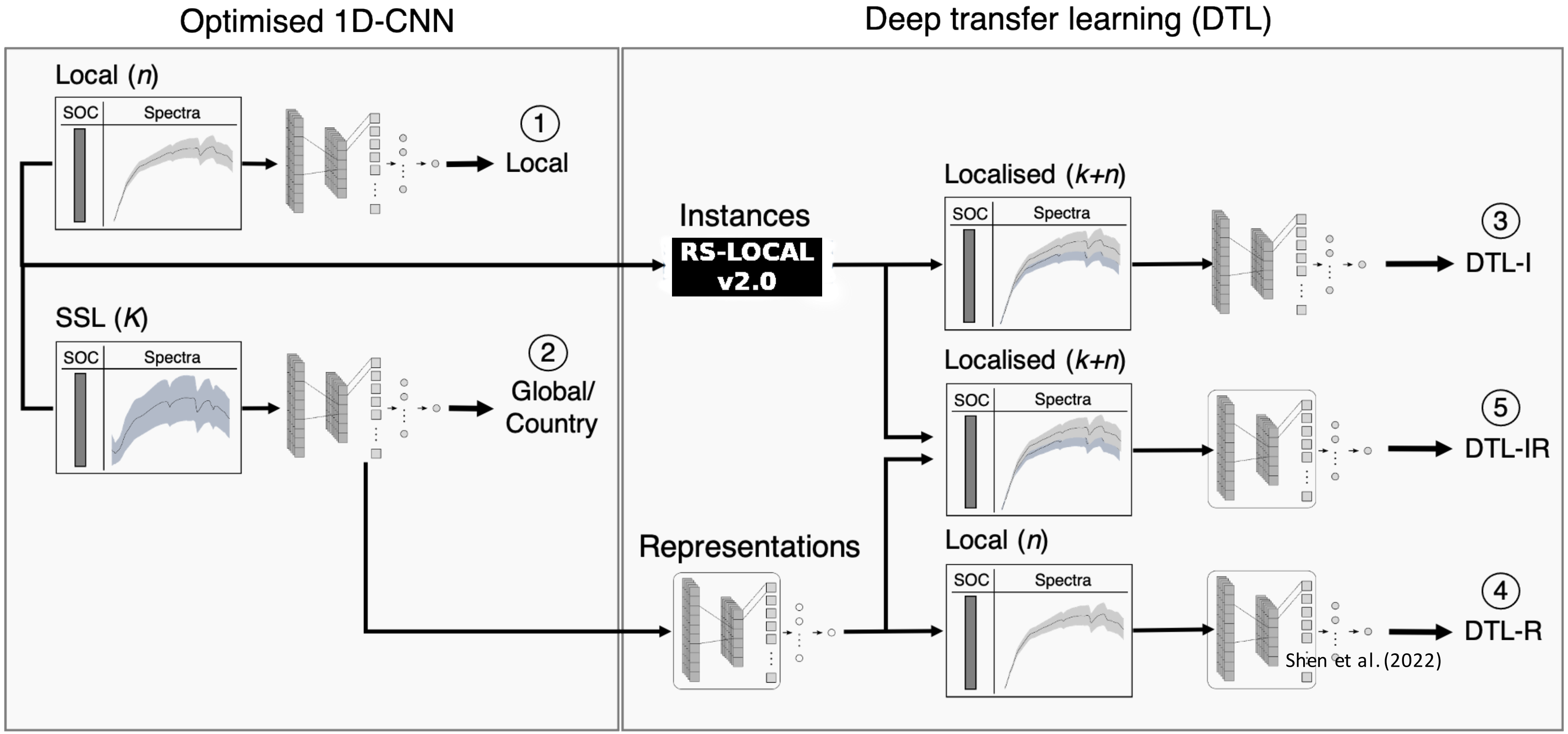
RS-Local-v2.0 selects relevant samples from SSL to augment local data for modelling.

1D-CNN and transferring representations



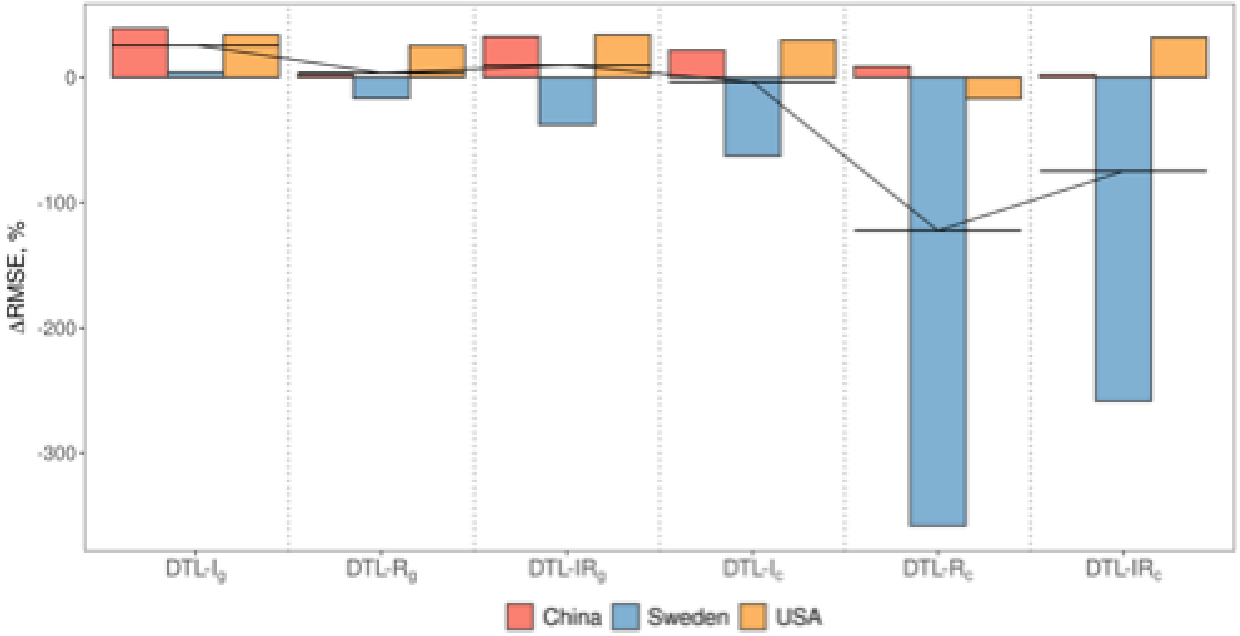
Lobsey et al., (2017), Shen and Viscarra Rossel (2021), Shen et al. (2022)

Localisation with Deep Transfer Learning (DTL)



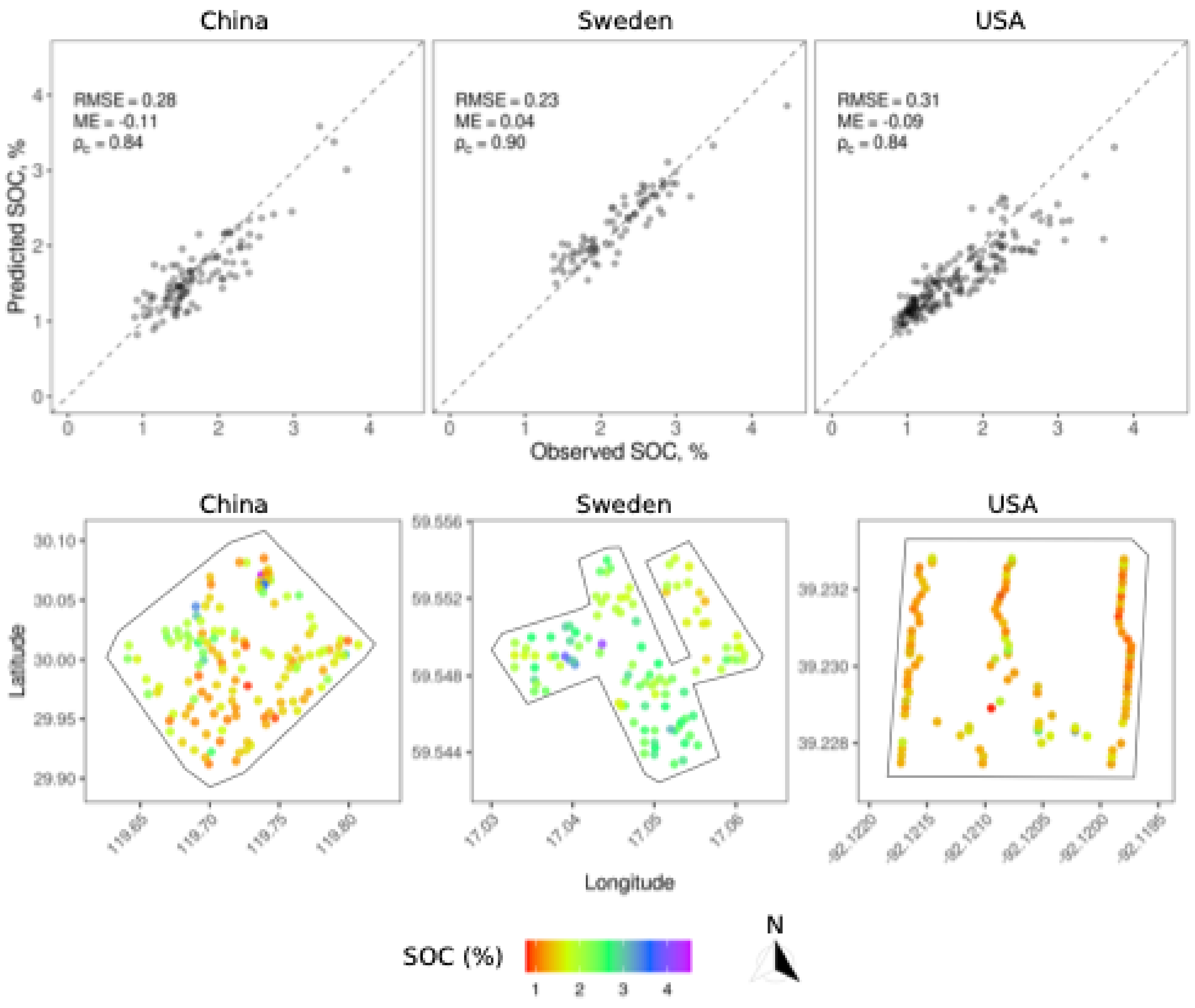
① Local: 1D-CNN developed on local data (*n*=30).
② Global/Country: 1D-CNN developed on Global/Country SSL(s)
③ DTL-I: Deep transfer learning of Instances
④ DTL-R: Deep transfer learning of Representations
⑤ DTL-IR: Deep transfer learning of Instances and Representations.

Localisation with Deep Transfer Learning (DTL)



$$\Delta RMSE = (RMSE_{Local} - RMSE_{Other}) \times 100$$

- DTL-I from global SSL improved local SOC prediction accuracy by 25.8% on average.
- DTL-R and DTL-IR did not show consistent improvement.



ME—Mean error;

ρ_c—Concordance correlation coefficient.

Take-home messages

- Soil spectroscopy is cost-effective for predicting soil properties.
- AI and machine learning are commonly used for soil spectroscopic modelling.
- Hyperparameter tunning is critical for obtaining accurate predictive models.
- Explainable Artificial intelligence (XAI) can help derive understanding of machine learning-based soil spectroscopic modelling.
- Transfer learning is the key for accurate local modelling with large spectral libraries.

References

- Shen, Z. and Viscarra Rossel, R.A., 2021. Automated spectroscopic modelling with optimised convolutional neural networks. Scientific Reports, 11(1), pp.1-12.
- Shen, Z., Ramirez-Lopez, L., Behrens, T., Cui, L., Zhang, M., Walden, L., Wetterlind, J., Shi, Z., Sudduth, K.A., Baumann, P., Song, Y., Catambay, K., Viscarra Rossel, R.A., 2022. Deep transfer learning of global spectra for local soil carbon monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, pp.190-200.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V. and Aïchi, H., 2016. A global spectral library to characterize the world's soil. Earth-Science Reviews, 155, pp.198-230.
- Lobsey, C.R., Viscarra Rossel, R.A., Roudier, P. and Hedley, C.B., 2017. RS-Local data-mines information from spectral libraries to improve local calibrations. European Journal of Soil Science, 68(6), pp.840-852.
- Yang, Y., Shen, Z., Bissett, A. and Viscarra Rossel, R.A., 2022. Estimating soil fungal abundance and diversity at a macroecological scale with deep learning spectrotransfer functions. Soil, 8(1), pp.223-235.
- Shen, Z., D'Agui, H., Walden, L., Zhang, M., Yiu, T.M., Dixon, K., Nevill, P., Cross, A., Matangulu, M., Hu, Y. and Viscarra Rossel, R.A., 2022. Miniaturised visible and near-infrared spectrometers for assessing soil health indicators in mine site rehabilitation. Soil, 8(2), pp.467-486.
- Haghi, R.K., Pérez-Fernández, E. and Robertson, A.H.J., 2021. Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: A comparison of near infrared and mid-infrared spectroscopy. *Geoderma*, *396*, p.115071.

Thank you!

Dr. Zefang Shen

Research Associate

Soil & Landscape Science

Curtin University, Australia

Email: zefang.shen@curtin.edu.au

Soil & Landscape Science Group

Curtin University