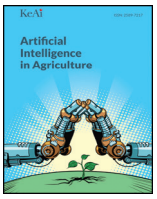
Artificial Intelligence in Agriculture 6 (2022) 230–241

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



Prediction of exchangeable potassium in soil through mid-infrared spectroscopy and deep learning: From prediction to explainability

Franck Albinet [a,b,*], Yi Peng [c], Tetsuya Eguchi [d,b], Erik Smolders [e], Gerd Dercon [b]

[a] Independent Researcher & Consultant, Guéthary, France
[b] Soil and Water Management & Crop Nutrition Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Seibersdorf, Austria
[c] Global Soil Partnership, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy
[d] Agricultural Radiation Research Center, Tohoku Agricultural Research Center, National Agriculture and Food Research Organization, Fukushima, Japan
[e] Soil and Water Management Unit, KU Leuven, Belgium

ARTICLE INFO

Article history:
Received 4 July 2022
Received in revised form 13 September 2022
Accepted 13 October 2022
Available online 18 October 2022

Keywords:
High-throughput soil characterization
Machine learning
Convolutional neural network
Agriculture
Nuclear emergency response
Remediation
Interpretability

ABSTRACT

The ability to characterize rapidly and repeatedly exchangeable potassium (K_{ex}) content in the soil is essential for optimizing remediation of radiocaesium contamination in agriculture. In this paper, we show how this can be now achieved using a Convolutional Neural Network (CNN) model trained on a large Mid-Infrared (MIR) soil spectral library (40,000 samples with K_{ex} determined with 1 M NH_4OAc, pH 7), compiled by the National Soil Survey Center of the United States Department of Agriculture. Using Partial Least Squares Regression as a baseline, we found that our implemented CNN leads to a significantly higher prediction performance of K_{ex} when a large amount of data is available (10000), increasing the coefficient of determination from 0.64 to 0.79, and reducing the Mean Absolute Percentage Error from 135% to 31%. Furthermore, in order to provide end-users with required interpretive keys, we implemented the GradientShap algorithm to identify the spectral regions considered important by the model for predicting K_{ex}. Used in the context of the implemented CNN on various Soil Taxonomy Orders, it allowed (i) to relate the important spectral features to domain knowledge and (ii) to demonstrate that including all Soil Taxonomy Orders in CNN-based modeling is beneficial as spectral features learned can be reused across different, sometimes underrepresented orders.

1. Introduction

Remediation of radioactive contamination of farmland requires accurate soil data. In case of nuclear emergencies affecting food and agriculture, exchangeable potassium (K_{ex}) plays a major role. As potassium competes with radiocaesium in soil-to-plant transfer, it can help reduce the crop uptake of this major fallout radionuclide. Information on potassium content in the soil is essential for optimizing remediation of radioactive contamination as happened in 1986 in Chernobyl or in 2011 in Fukushima (Komatsu et al., 2017). The ability to characterize K_{ex} rapidly and repeatedly will inform and facilitate decision making, and significantly reduce the cost of analyses needed for remediation over the years.

Traditional methods for characterizing soil properties, such as K_{ex}, through wet chemistry can provide high accuracy and precision, but it is often cost and time prohibitive (Viscarra Rossel et al., 2006). For several decades, soil scientists have been successfully using visible, near and mid-infrared spectroscopy techniques to estimate a wide range of soil physical, chemical, and biological properties (Viscarra Rossel et al., 2006; Frost et al., 2007; Shepherd and Walsh, 2007; Bellon-Maurel et al., 2010; Grinand et al., 2012; Peng et al., 2014; Torres Astorga et al., 2018; Dangal et al., 2019; Ng et al., 2022). MIRS is often more accurate and produces more robust calibrations than near-infrared reflectance spectroscopy (NIRS) when analyzing ground, dry soils (Viscarra Rossel et al., 2006; Ng et al., 2019b). However, models developed to estimate soil characteristics are often limited in scope, as only small and region-specific MIR spectral libraries of soils are available (Viscarra Rossel et al., 2008; de Sousa Mendes et al., 2022). The lack of standardization of wet chemistry methodologies (Viscarra Rossel and Bouma, 2016) coupled with situations of low data availability, or so-called low data regimes, also have long caused calibration models overfitting small and often noisy datasets (Torres Astorga et al., 2018). To avoid such overfitting often involves careful data pre-processing and dimensionality reduction (Viscarra Rossel et al., 2022), which are all pervading themes in MIRS soil data processing and statistical learning in general (Hastie et al., 2009).

* Corresponding author at: Independent Researcher & Consultant, Guéthary, France.
E-mail address: franckalbinet@gmail.com (F. Albinet).

https://doi.org/10.1016/j.aiia.2022.10.001

The situation of data scarcity is, however, changing rapidly today with the availability of large, and growing, high-quality and less region-specific libraries of MIR-scanned soil samples. A strong example is the spectral library maintained by the National Soil Survey Center (NSSC) Kellogg Soil Survey Laboratory (KSSL) from the United States Department of Agriculture (USDA-NRCS), which is available to the global scientific community.

The unprecedented volume and diversity of data now available, often released in accordance with open data policies, allows soil science researchers to increasingly shift their focus from traditional modeling techniques such as Partial Least Squares Regression (PLSR) to classes of Machine Learning (ML) models of higher complexity, such as Convolutional Neural Networks (CNN).

CNN is a class of deep neural networks commonly applied to data that has a grid-like topology, such as images, audio or time series. A notable characteristic of deep neural networks is their ability to learn and detect patterns in a hierarchical way. For instance, in the context of image recognition tasks, it has been shown that the first convolutional layers learn low-level features such as diagonal, horizontal and vertical edges while further layers learn higher-semantic components such as corners, circles or even complex images (Zeiler and Fergus, 2013). Harnessed recently in Mid-infrared soil spectroscopy, CNN has proved to be capable to learn, hence automate, spectra preprocessing steps and to relate complex absorption peak patterns to the presence of soil properties to be characterized and predicted (Ng et al., 2019b).

The performance gain of CNN compared to PLSR may not be significant for some parameters that are easy to predict such as soil organic carbon or calcium carbonate. However, recent studies showed that, under a large data regime, Convolutional Neural Networks (CNN) outperform PLSR on most soil property predictions (Wijewardane et al., 2018; Ng et al., 2019b; Shen and Viscarra Rossel, 2021; Shen et al., 2022; Ng et al., 2022). Further, performance of property prediction through PSLR can be often very low for some certain soil parameters (Ng et al., 2019b; Wijewardane et al., 2018), in particular when the relationship between the spectra and these properties is non-linear as for instance for K_{ex}. This inferior performance is known to be related to the fact that this soil property has no direct or strong spectral features to support the calibration model and is only inferred in relation to other spectrally active soil components (Janik and Skjemstad, 1995; Sanderman et al., 2020).

However, if such a relationship exists and can be learned, it is necessary to have classes of models such as CNN capable of (i) extracting in-depth information from the spectra for better calibration purposes and (ii) exploiting a large and diverse spectral library. To the best of our knowledge, the estimation of K_{ex} at a level of performance sufficient for nuclear remediation using a large MIR spectral library and a CNN algorithm has never been tested until KSSL database has been made available to the public recently.

Moreover, when intended to be used to derive new scientific insights, inform decision-making and policies during large-scale crisis responses, machine learning algorithms not only need to demonstrate enough predictive power, but must also equip end-users with interpretive keys allowing to understand on which basis, on which features (spectral regions) of the input data these predictions are made.

In theory, based on vibrational spectroscopy principles, and under controlled conditions with defined mixtures of soil constituents, characteristic infrared electromagnetic radiation absorption patterns can be attributed to functional groups and chemical bonds (Socrates, 2004). In practice, this is more challenging based on the measurement conditions and when analyzing soil constituents like K_{ex} that are in lower concentrations and only being observed by correlation with other spectrally active major soil components (Sanderman et al., 2020). Yet, the interpretation and the ability to assign spectral features to fundamental vibrations of specific soil properties is key (i) to develop more parsimonious models through the selection of specific wavenumbers, known as important, during the preprocessing, feature extraction and modeling phases (Viscarra Rossel and Lark, 2009; Mehmood et al., 2012; Ng et al., 2019a), (ii) to better understand model output, and provide insight into how these models can be further improved (Viscarra Rossel and Behrens, 2010), (iii) to potentially shed new lights on the current state of knowledge of the soil processes being modeled (Lundberg and Lee, 2017) and (iv) to foster end-users trust and facilitate adoption of machine learning-based technologies (Molnar et al., 2020).

So far, methods used to interpret the contribution of each input feature (i.e., the absorption of electromagnetic radiation at a particular wavenumber) to a model output are either model specific (Terhoeven-Urselmans et al., 2010; Viscarra Rossel et al., 2016; Clairotte et al., 2016; Wijewardane et al., 2018; Landre et al., 2018; Gomez et al., 2020) or have a high computational cost (Ng et al., 2019b). In addition, those methods average features' importance over all predictions whereas importance for individual or a group of predictions, for instance by Soil Taxonomy Order, is of crucial importance for both model developers and radiological environmental remediation domain experts (Molnar et al., 2020).

In this publication, our first hypothesis is that the ability to characterize K_{ex} in soil at a sufficient level of performance for remediation is conditioned by the availability of a large amount of data and the use of high capacity model classes such as CNN. Our second hypothesis is that CNN-based soil property prediction benefits from merging MIRS information from a wide range of soil types (Soil Taxonomy Orders), even when data are not equally distributed across all orders. By merging, soil orders with a low amount of data in the KSSL dataset also benefit from the CNN leaning process, reusing, when possible, spectral features learned on well represented soil orders. Such a feature would provide a major opportunity to address the lack of data on the fate of radionuclides in poorly explored environments such as arid, tropical and monsoonal climates.

Therefore, in this study, our first aim is to assess the ability of CNN models to predict K_{ex} in soil when trained on a large MIRS spectral library with a sufficient level of accuracy for the remediation of radioactive contamination of agricultural land. To articulate the role of the data regime and model capacity in this analysis, PLSR is used as a baseline. Our second aim is to investigate, through the analysis of model's performance across Soil Taxonomy Orders, how domain experts can exploit CNN interpretability algorithms to gain new insights into the soil processes being modeled and build confidence in the decision-making process during nuclear remediation.

2. Materials and methods

2.1. Modeling pipeline overview

Table 1 summarizes the entire data pre-processing, analysis and modeling pipeline developed in this study, and highlights in particular the differences between PLSR and CNN models. The following sections provide full details of each individual task.

To ensure reproducibility, the source code of the modeling pipeline and data to replicate this work are available at the following link: https://fr.anckalbi.net/mirzai.

Python version 3.6, NumPy (https://numpy.org), Scikit-Learn (https://scikit-learn.org) and PyTorch (https://pytorch.org) have been used for the data pre-processing, analysis and modeling parts as well as Captum (https://captum.ai/) to address the interpretability needs. A Google Colab Pro computing environment has been used with the following specification: 1 GPU Tesla P100 16GB, 2 Intel(R) Xeon(R) CPU @ 2.30GHz cores and 26GB RAM. This democratized access to the sufficient computational resources is also a key ingredient of reproducibility research.

2.2. Data selection, preprocessing

The ability to predict and characterize K_{ex} is key to optimize the remediation of radioactive contamination in agriculture. For this

Table 1
Overview of K_{ex} characterization pipeline using Mid-infrared spectroscopy, and Partial Least Squares and Convolutional Neural Network models: from data selection, pre-processing, analysis, models validation and optimization to interpretability[a].

Phase	Task		Models	
			PLSR	CNN
I. Data	Selection		$K_{ex} > 0.12$ in cmol(+).kg^{-1} and all soil orders	
	Preprocessing	Spectra	SNV	
			Savitzky-Golay CO_2 region removal	
		K_{ex}	Log_{10} transformation	
II. Modeling	Evaluation scheme		Train (80%), Validation (10%) and Test (10%) random splits repeated 20 times	
	Metrics		R^2, LCCC, RMSE, MAPE, Bias	
	Hyperparameters		Number of components	Number of epochs
	Regularization		Early stopping	
				Dropout
III. Interpretability				GradientShap

[a] With PLSR: Partial Least Squares Regression; CNN: Convolutional Neural Network; SNV: Standard Normal Variate; R^2: coefficient of determination; LCCC: Lin's concordance correlation coefficient; RMSE: Root Mean Square Error; Bias: Mean Error and MAPE: Mean Absolute Percentage Error.

study, we focused on a selection of the Mid-Infrared Spectral database made available to the Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture by the Kellogg Soil Survey Laboratory (KSSL) from the United States Department of Agriculture (USDA-NRCS). The Mid-Infrared spectra were acquired on air-dried and ground soil samples using a Bruker Vertex 70 FTIR spectrometer with an HTS-XT high-throughput accessory (Dangal et al., 2019). In total 50,494 samples were identified and selected from the dataset, on which K_{ex} was measured (as of January 2022). The mid-infrared spectra (from wavenumbers 4000 to 600 cm^{-1} at resolution of 2–1764 wavenumbers in total) were systematically centered and normalized, for both the baseline (PLSR) and CNN models, using the Standard Normal Variate transform (SNV) (Barnes et al., 1989). When used to train the baseline model, a Savitsky-Golay transformation (window length of 11, polynomial order of 1 and first derivative) was further applied and the CO_2 spectral region was removed to avoid influence from this artifact independent from the soil analysis (from wavenumbers 2269 to 2389 cm^{-1}). The CNN will safely ignore the CO_2 spectral region and learn the first derivative (finite difference) operator if it is deemed necessary to K_{ex} prediction, hence, both pre-treatments were not applied in that case.

In the dataset involved in this study, K_{ex} content in the soil has been measured using the NH_4OAc pH 7 extraction method (in cmol(+).kg^{-1}). Table 2 provides the summary statistics of the empirical distribution of these measurements and their log_{10}-transformation before any data selection (50,494 measurements in total).

Figure 1 shows both the number of samples and the distribution of K_{ex} concentration by Soil Taxonomy Orders (Baillie, 2001) before any selection. We observe that certain soil orders are largely under-represented although their inclusion in the majority "Undefined" class remains theoretically possible. However, the relative distribution of the defined soil orders, from Mollisols to Oxisols, is representative of soil types present in the geographical area covered by the KSSL dataset (United States mainly) (Dangal et al., 2019).

The method detection limit (MDL) for K_{ex} content (target variable), through the NH_4OAc pH 7 extraction method, used for the KSSL database, is 0.06 cmol(+).kg^{-1} while its Practical Detection Limit (PDL) is 0.30 cmol(+).kg^{-1} (5 times of MDL), set higher than the PDL in pure chemistry (approximately 3.3 times of MDL) because of more complex matrices in soil extracts (Klesta and Bartz, 1996). However, the appropriate K_{ex} content to ensure radiocaesium concentration of 99% of brown rice below the target level was 0.14 cmol(+).kg^{-1} in Fukushima area. To include this appropriate K_{ex} content (0.14 cmol (+).kg^{-1}), the detection limit used in this study is closer to the MDL and a threshold of 0.12 cmol(+).kg^{-1} was applied, i.e. two times the MDL. About 20% of data had values below that threshold and the remaining 80% of the data, i.e. 40,132 samples, was log_{10}-transformed to avoid weighting the loss functions accounting for the heteroscedasticity of the prediction residuals otherwise.

2.3. Modeling

2.3.1. Evaluation scheme

For all experiments conducted in this study, the data were randomly split in the following proportions: 80% for training, 10% for validation and 10% for testing. To optimize model hyperparameters, an "early stopping" technique was used. The number of components or epochs for the PLSR and CNN model respectively was increased so long as the loss on validation continued to decrease. When the loss either leveled off or increased during three consecutives increases of the hyperparameters, the training was stopped. Models were then trained, validated and tested on 20 different versions (random seeds) of this splitting scheme (i) to assess their robustness and (ii) to mitigate the effect of relatively small sample sizes for certain Soil Taxonomy Orders such as Oxisols, Vertisols, Gelisols and Histosols. For each of these different 20 random splits, both models were exposed to the same train, validation and test (unseen) data. The mean performance as well as their standard deviation were reported.

Table 2
Summary statistics of measured K_{ex} in cmol(+).kg^{-1}. Q1, Q2 and Q3 are respectively the First, Second (Median) and Third quartiles.

	Min	Max	Mean	Standard Deviation	Quartiles			Skewness	Kurtosis
					Q1	Q2	Q3		
K_{ex}	0	32.33	0.66	1.06	0.15	0.36	0.74	7.1	103
$log_{10}(K_{ex})$[a]	−2.34	1.51	−0.43	0.49	−0.75	−0.42	−0.11	−0.1	0.02

[a] For K_{ex} values >0 only.

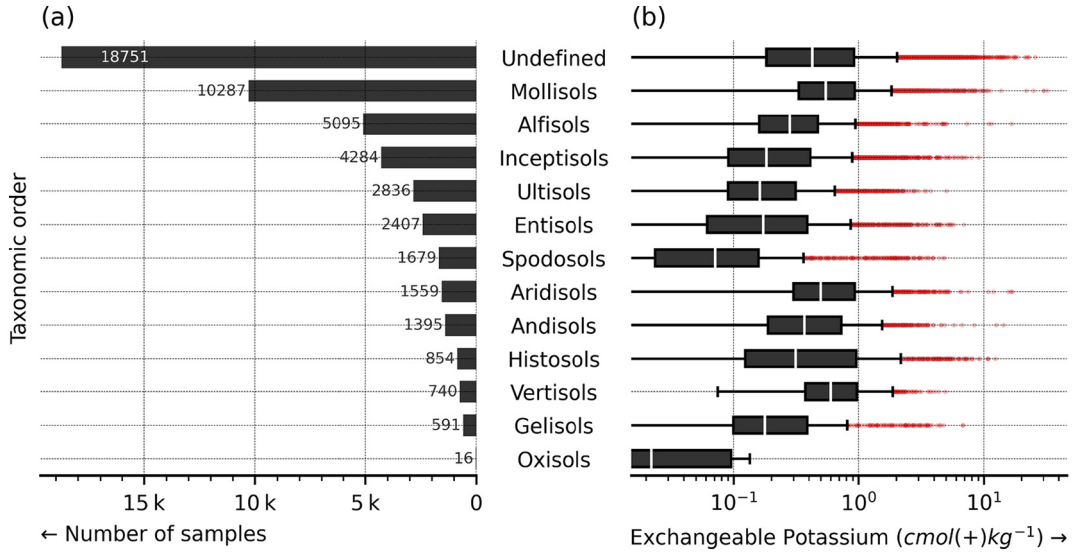


Fig. 1. (a) Number of samples per Soil Taxonomy Orders (note that most of them are "Undefined") (b) Boxplot of K_{ex}. The box extends from the first to third quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data. Flier points (red) are those past the end of the whiskers. No data selection is applied. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To assess the ability of the PLSR and CNN models to leverage an increasing data regime, PLSR and CNN models were also trained on random subsamples of increasing sizes of the data and their performance on the test sets portrayed through the so-called "Learning curve" graph (Viering and Loog, 2021).

The assessment of the models was further performed in two different contexts that we call in this study "global" and "local". In the global setting, the models were trained on all data and performance was assessed on the entire test set and by Soil Taxonomy Orders. To provide further insights into the CNN training process, the validation loss by Soil Taxonomy Orders was also monitored as the number of epochs increased during training. In the local setting, the models were trained by Soil Taxonomy Orders separately, and tested on the Soil Taxonomy Orders they were trained on. Only performance on Mollisols, Vertisols and Gelisols are reported (i) to highlight the role of distinctive data regimes (from high to low), and (ii) to assess the ability of the CNN model to "re-use" or transfer spectral features learned across different Soil Taxonomy Orders, in particularly between well and non-represented ones in the MIRS KSSL dataset.

2.3.2. Metrics

To evaluate both the PSLR and the CNN models on training, validation and test datasets, the following metrics were used: (i) coefficient of determination (R^2) and (ii) Lin's concordance correlation coefficient (LCCC). LCCC measures the linear correlation between the predicted and true values but penalized by a bias term allowing to assess the agreement between the observed and predicted values with respect to the 1:1 line (Lawrence and Lin, 1989). These two metrics were calculated on the $\log_{10}$-transformed version of the true and predicted values of K_{ex}. In addition, the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) were calculated on the original scale of the target variable to facilitate interpretation and provide two complementary perspectives on the performance of the models. Indeed, given the skewed empirical distribution of the target variable, RMSE and MAPE tend to overweight errors made respectively for large and small K_{ex} values. Bias was also calculated on the original scale and reported.

The equations of the above-mentioned metrics are explained below:

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2} \quad (1)$$

$$LCCC = \frac{2r\sigma_{\hat{y}}\sigma_y}{\sigma_y^2 + \sigma_{\hat{y}}^2 + \left(\bar{y} - \bar{\hat{y}}\right)} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (3)$$

$$MAPE = \frac{100}{m}\sum_{i=1}^{m}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (4)$$

$$Bias = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i) \quad (5)$$

where m is the number of instances; i the index of an individual instance; y the observed value; $\hat{y}$ the predicted one; $\bar{y}$ the mean of the observed value; $\bar{\hat{y}}$ the mean of the predicted value; r the Pearson's correlation coefficient between the observed versus predicted values; σ_y the standard deviation of observed values and $\sigma_{\hat{y}}$ the standard deviation of predicted ones.

2.3.3. Architecture of convolutional neural network models

PLSR was used as a baseline in this study, to highlight its behavior in increasing data regimes.

The design of our CNN was inspired by the VGG16 architecture (Simonyan and Zisserman, 2014) stacking uniform convolution-pooling layers with small kernel sizes (i) to promote automatic extraction of hierarchical features, (ii) to control the growth in the number of learned parameters and (iii) to dispose of sufficient model capacity to predict K_{ex} in large data regime.

Fig. 2 provides an overview of the architecture of our Convolutional Neural Network consisting of two main components:

- A features learner composed of stacked blocks, each consisting of a convolutional, batch normalization, non-linear activation and average pooling layers, with each consecutive block seeing its depth (number of filters) doubled (up to 512 in the last block).
- A task learner that (i) flattens the learnt features maps, (ii) performs dimensionality reduction of final features into a 20-dimensional latent space using a fully connected layer and finally (iii) applies a non-linear activation and a single neuron linear layers (with no further activation function) predicting the soil property of interest. A Dropout and Batch Normalization layers were also added.

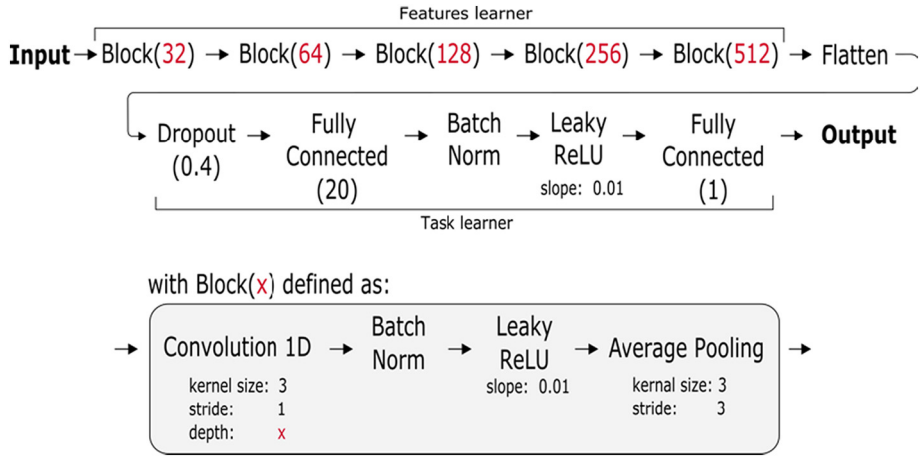


Fig. 2. Convolutional Neural Network architecture used in this study composed of (i) 5 stacked Convolutional-Pooling blocks with constant and small kernel size and increasing depth (feature learner) and (ii) a task learner.

The model used in our study has 162,401 trainable parameters in total and was regularized using early stopping (after 3 epochs without validation loss decrease) to guarantee its generalization property and prevent overfitting.

While we relied on sensible defaults with regards to the network architecture and did not perform extensive hyperparameters tuning, we used cyclic learning rates (Smith, 2017). The motivation behind cyclic learning rates is that, by alternatively decreasing and increasing the learning rate, we give to the optimizer the opportunity to explore, visit multiple local minima of the loss function and hopefully prevent the optimizer from getting trapped in a local minimum. In this study, we use an Adam optimizer (Kingma and Ba, 2014) with cyclic learning rates between 3×10^{-5} and 1×10^{-3}. A Mean Squared Error loss was used to optimize the CNN and the determination of the min and max learning rates was decided based on the structure of the loss over 8 learning epochs (Smith, 2017).

2.4. Interpretability

In this study, we implemented the GradientShap (Lundberg and Lee, 2017; Sundararajan et al., 2017; Smilkov et al., 2017; Kokhlikyan et al., 2020) model interpretability algorithms to identify the spectral features regarded as important by the learning algorithm for the prediction task being performed. The importance of spectral features can be computed model-wide or for an individual or group of instances with marginal additional computational burden. In our effort to relate this information to domain knowledge, feature importance was calculated for the overall CNN models based on (i) the full dataset, used in this study, and (ii) the Soil Taxonomy Orders separately. The correlation of the GradientShap values of the spectral features for predicting K_{ex} content by Soil Taxonomy Orders was also assessed, allowing understanding their similarities. This information on correlation does allow not only increasing interpretability but also helps to understand how the learned spectral features are reused across soil orders.

The GradientShap algorithm takes the gradients of model output with respect to the input features as it could be done in a linear regression setting. However, the non-linear behavior of CNN models introduces masking effects (Shrikumar et al., 2019) making such an approach less suitable. Instead, to remedy this possible lack of sensitivity, the GradientShap algorithm was integrated in our study with the gradient along the path between several baselines (in our study, 100 spectra randomly sampled from the training set), so allowing to better approximate the calculation of the SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) value.

3. Results

3.1. Exchangeable potassium prediction models in a growing data regime

The learning curves and model performance of both PLSR and CNN models for the prediction of exchangeable potassium ($\log_{10}(K_{ex})$), with increasing number of training examples and using all Soil Taxonomy Orders, are shown in Fig. 3.

Up to a dataset of 5000 samples, both PLSR and CNN models perform poorly with an R^2 below 0.65 and a high level of variance across different random splits. However, a data regime of 10,000 samples appears to be an inflection point. While the PLSR model levels off at approximately an R^2 of 0.64, the performance of the CNN keeps increasing up to an R^2 of about 0.79 as the size of the dataset increases.

3.2. Exchangeable potassium prediction model training and evaluation

3.2.1. Global setting

In this setup, the PLSR and CNN models were trained on all data and performance was assessed on the entire test set and by Soil Taxonomy

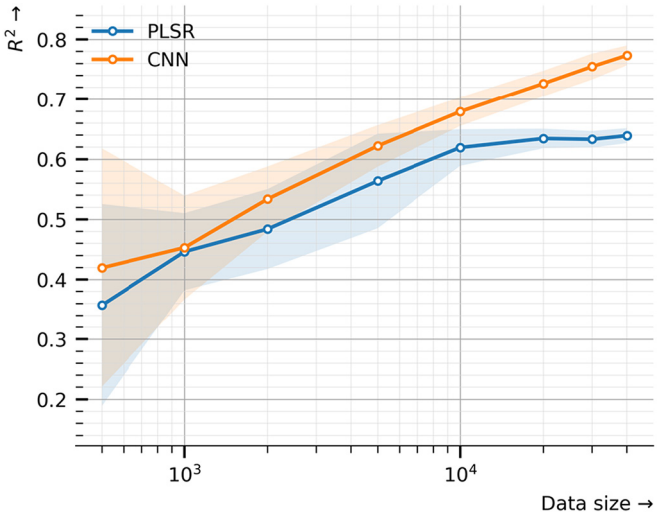


Fig. 3. Learning curves of PLSR and CNN models trained to predict K_{ex} in a growing data regime using all Soil Taxonomy Orders. The mean and the standard deviation of the coefficient of determination R^2 calculated on 20 different test splits as the size of the dataset increases are reported.

Table 3
Comparison of PLSR and CNN models performance[a] in predicting K_{ex} when trained on all Soil Taxonomy Orders and tested on individual Soil Taxonomy Orders separately. Figures reported are averages and standard deviations of metrics computed on test sets over 20 different train/validation/test splits. N is the number of samples. Performances on Oxisols are not reported as the size of its test set contains often a single sample.

			Metrics			
Orders	N	Models	R^2	LCCC	RMSE	MAPE
All	4032	PLSR	0.64 ± 0.01	0.780 ± 0.005	1.06 ± 0.45	135.0 ± 3.3
		CNN	0.79 ± 0.08	0.884 ± 0.005	0.60 ± 0.16	30.9 ± 0.9
Undefined	1553	PLSR	0.65 ± 0.01	0.788 ± 0.007	1.15 ± 0.52	153.0 ± 7.2
		CNN	0.81 ± 0.01	0.897 ± 0.006	0.76 ± 0.28	31.6 ± 1.3
Mollisols	977	PLSR	0.60 ± 0.02	0.742 ± 0.012	0.77 ± 0.20	94.3 ± 3.1
		CNN	0.78 ± 0.02	0.868 ± 0.014	0.44 ± 0.07	27.4 ± 1.2
Alfisols	422	PLSR	0.54 ± 0.05	0.728 ± 0.029	0.41 ± 0.18	82.6 ± 4.4
		CNN	0.69 ± 0.03	0.822 ± 0.018	0.38 ± 0.17	27.3 ± 1.3
Inceptisols	289	PLSR	0.54 ± 0.04	0.715 ± 0.021	0.56 ± 0.12	117.6 ± 6.8
		CNN	0.72 ± 0.03	0.838 ± 0.023	0.40 ± 0.07	35.0 ± 3.1
Ultisols	192	PLSR	0.30 ± 0.08	0.581 ± 0.056	0.32 ± 0.05	80.9 ± 9.0
		CNN	0.62 ± 0.05	0.763 ± 0.038	0.26 ± 0.06	32.5 ± 2.4
Entisols	165	PLSR	0.54 ± 0.10	0.734 ± 0.051	0.44 ± 0.09	124.3 ± 15.2
		CNN	0.77 ± 0.05	0.875 ± 0.033	0.32 ± 0.05	30.8 ± 3.4
Aridisols	163	PLSR	0.45 ± 0.16	0.686 ± 0.067	2.13 ± 2.74	145.9 ± 56.2
		CNN	0.69 ± 0.04	0.822 ± 0.022	0.66 ± 0.32	35.1 ± 3.1
Andisols	133	PLSR	0.59 ± 0.03	0.729 ± 0.028	0.59 ± 0.29	106.4 ± 11.3
		CNN	0.74 ± 0.04	0.856 ± 0.026	0.48 ± 0.15	32.6 ± 3.3
Vertisols	95	PLSR	0.55 ± 0.10	0.732 ± 0.062	0.39 ± 0.10	98.7 ± 12.3
		CNN	0.75 ± 0.06	0.857 ± 0.037	0.27 ± 0.06	26.9 ± 3.3
Histosols	80	PLSR	0.64 ± 0.09	0.778 ± 0.055	1.16 ± 0.38	243.1 ± 46.1
		CNN	0.76 ± 0.06	0.870 ± 0.029	0.87 ± 0.27	45.4 ± 7.4
Spodosols	64	PLSR	0.70 ± 0.08	0.824 ± 0.043	0.48 ± 0.13	157.1 ± 19.7
		CNN	0.78 ± 0.05	0.880 ± 0.025	0.41 ± 0.11	37.2 ± 4.7
Gelisols	61	PLSR	0.63 ± 0.10	0.767 ± 0.060	0.65 ± 0.14	193.8 ± 31.4
		CNN	0.75 ± 0.08	0.860 ± 0.047	0.59 ± 0.14	47.3 ± 9.9

[a] N is the number of samples in the test set. The coefficient of determination R^2 and Lin's concordance correlation coefficient (LCCC) are unitless; the Root Mean Square Error (RMSE) is expressed in cmol(+).kg^{-1}. The Mean Absolute Percentage Error (MAPE) is expressed in %. Both the mean and standard deviation of the metrics calculated over the 20 different random splits are reported.

Orders. Table 3. reports their mean and standard deviation performances on the test set. We recall that the models were trained, validated and tested on 20 different random train, validation and test splits.

Results confirm that the CNN model outperforms PLSR in the context of a large data regime, reaching in average an R^2 of 0.79 and a MAPE of 30.9% whereas PLSR levels off at an R^2 of 0.64 and a MAPE of 135% on the overall test set. Furthermore, the predictions calculated by the CNN over 20 different random splits show little variance with a standard deviation of 0.08 and 0.9% for R^2 and MAPE respectively. With such prediction performance, the CNN model is more suitable for the remediation of radioactive contamination.

Moreover, when evaluated by Soil Taxonomy Orders, the average MAPE ranges from 26.9% for Vertisols to 47.3% for Vertisols. On Mollisols, the most represented order in the KSSL dataset, the average MAPE is 27.4%. The variability of the metrics increases as sample size of the test set decreases as it is the case for Gelisols with a standard deviation of 9.9% for its MAPE.

The average bias values for CNN predictions evaluated on all data is 0.0047 cmol(+).kg^{-1}. When evaluated on individual Soil Taxonomy Orders, bias values range from −0.04 to 0.02 cmol(+).kg^{-1} for less represented soil orders such as Gelisols or Histosols. As for the PLSR, the average bias values predictions evaluated on all data is −0.0002 cmol(+).kg^{-1}. When evaluated on individual Soil Taxonomy Orders, bias values range from −0.09 to 0.04 cmol(+).kg-^{1} respectively for Gelisols and Ultisols.

The scatterplots of observed vs predicted K_{ex} on a single train/test split shown in Fig. 4 confirm the higher overall agreement of the CNN model between the observed and predicted values with respect to the 1:1 line.

As confirmed by Fig. 4, the reported RMSE and MAPE metrics also provide complementary perspectives on the errors performed, as overweighting respectively errors made on higher and lower ranges of predicted K_{ex}. Indeed, these scatterplots clearly highlight a higher dispersion when predicting with PLSR, in particular for levels of K_{ex} lower than 1 cmol(+).kg^{-1} (once back-transformed in their original scale). For higher levels (higher than 1 cmol(+).kg^{-1}), the PLSR model shows a bias, almost systematically underestimating K_{ex}.

To see the learning process of the CNN in action, the validation loss by Soil Taxonomy Orders is reported in Fig. 5.

As indicated by the dynamic of the loss evaluated on the entire validation set (All), 200 training epochs were required to meet the "early stopping" criteria (stop after 3 epochs without validation loss decrease). Below this curve, the first group of validation curves including Andisols, Alfisols, Mollisols and Vertisols shows a continuous decrease. As opposed, above this curve, the second group of validation curves including Inceptisols, Aridisols, Ultisols, Spodosols and Gelisols has leveled off long before the end of the training, indicating a risk of overfitting. The validation curve for Histosols still shows a steady decrease up to approximately 180 training epochs but with higher MSE values.

3.2.2. Local setting

Figure 6 shows the comparison of the test performances (R^2 and MAPE) for both PLSR and CNN models on Mollisols, Gelisols and Vertisols Soil Taxonomy Orders when trained (i) on all orders included (global) and (ii) on Mollisols, Gelisols or Vertisols separately (local).

This experiment highlights clearly that, at least when predicting K_{ex}, stratifying the modeling approach by Soil Taxonomy Orders was of little or no interest for the PLSR model and substantially detrimental for the CNN model. Additionally, high standard deviations were reported when training and testing on Gelisols due to the small sample size, which consequently confirms the relevance to systematically report result on multiple random splits.

3.3. Assignment of spectral bands

The GradientShap values averaged over the whole test set or by Soil Taxonomy Orders is shown in Fig. 7 along with the mean spectrum for each class.

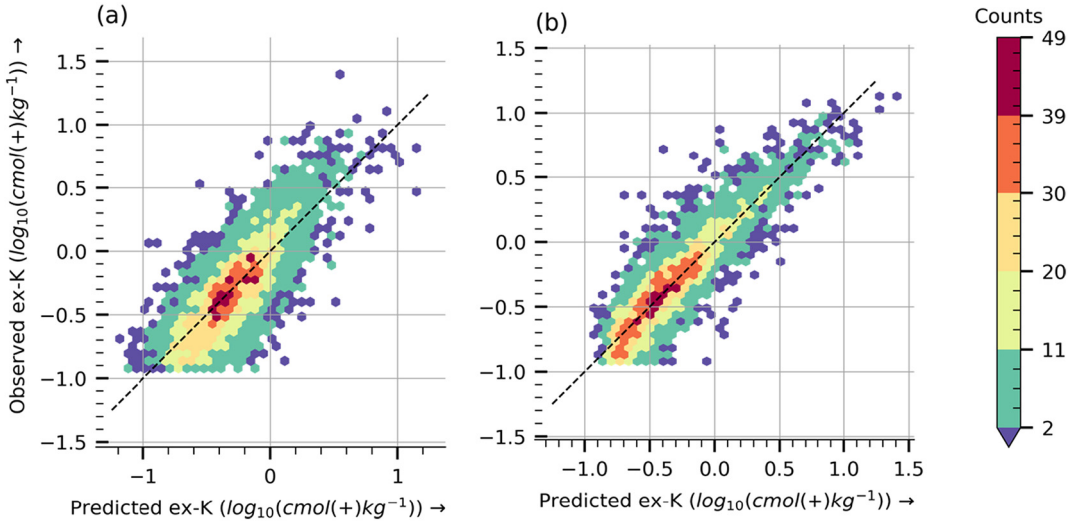


Fig. 4. Observed vs. Predicted K_{ex} with (a) PLSR and (b) CNN using all Soil Taxonomy Orders.

It is important to recall that when the SNV (Standard Normal Variate) preprocessing step is applied upstream of the modeling pipeline, MIR spectra are made translation and scale invariant. Hence, machine learning algorithms exploit only deformation/distortion patterns to provide estimates. GradientShap values indicate the effect of local deformations of the spectra on predicted values.

A first analysis of the GradientShap values shows that the spectral region around wavenumbers 3700–3400 cm^{-1} is considered by the CNN as important to predict K_{ex} for most soil orders. Several studies (Viscarra Rossel et al., 2006; Le Guillou et al., 2015) have highlighted the role played by soil mineralogical characteristics in this spectral region with the presence of absorption peaks of soil constituents such as calcite, kaolinite, 2:1 type clay minerals, and the features of free water. Importance of mineralogy was also implied by regions 1200–970 cm^{-1}, assigned to Si–O stretching (Farmer, 1974). Soil organic matter may also contribute potassium retention in soil orders with high organic matter content such as Histosols, Gelisols and Andisols as implied by the importance of the 3000–2800 cm^{-1} spectral region (Le Guillou et al., 2015).

To further compare GradientShap values computed for each Soil Taxonomy Orders, a correlation matrix is shown in Fig. 8. The assessment of GradientShap values similarity allows to postulate the presence of spectral features reused across Soil Taxonomy Orders and will support our analysis of the behavior of the CNN model at different data regimes.

A detailed analysis of the correlation levels is conducted below. To only consider the general trends at this stage, a high level of correlation is observed between the GradientShap values calculated for Spodosols, Entisols and the Undefined category. The GradientShap values calculated on Mollisols, the most represented Soil Taxonomy Orders in the KSSL database, are highly correlated positively with Vertisols and negatively with Spodosols and Entisols.

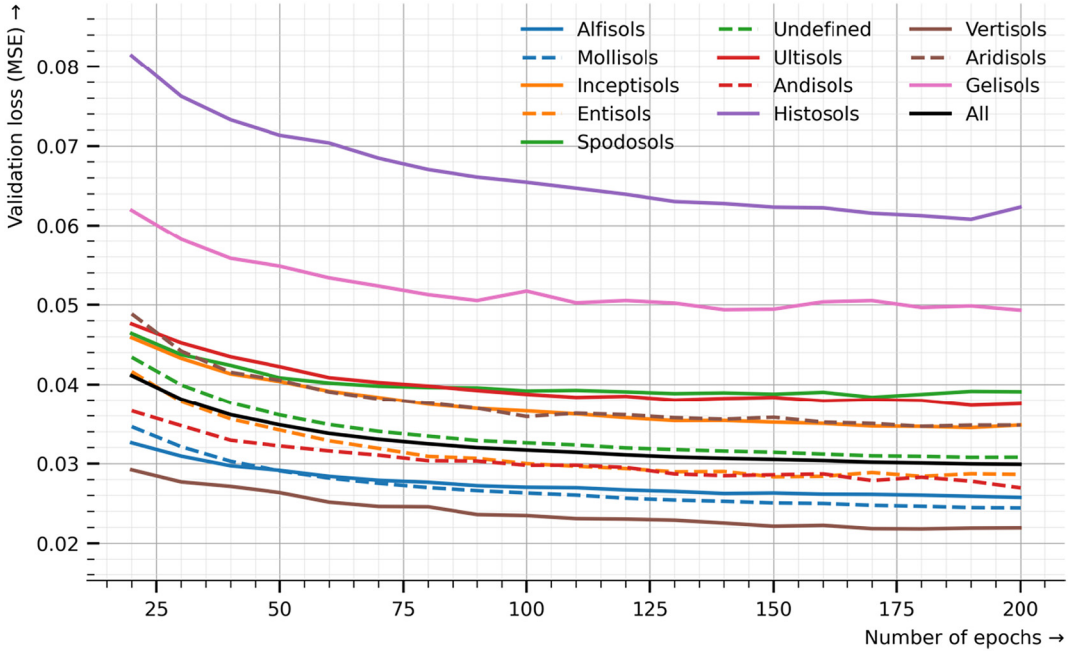


Fig. 5. Validation loss (MSE: Mean Squared Error) of the CNN during the training process (200 epochs) calculated on the entire validation set (All) and by Soil Taxonomy Orders. The different curves represent the mean MSE over 20 different random splits.

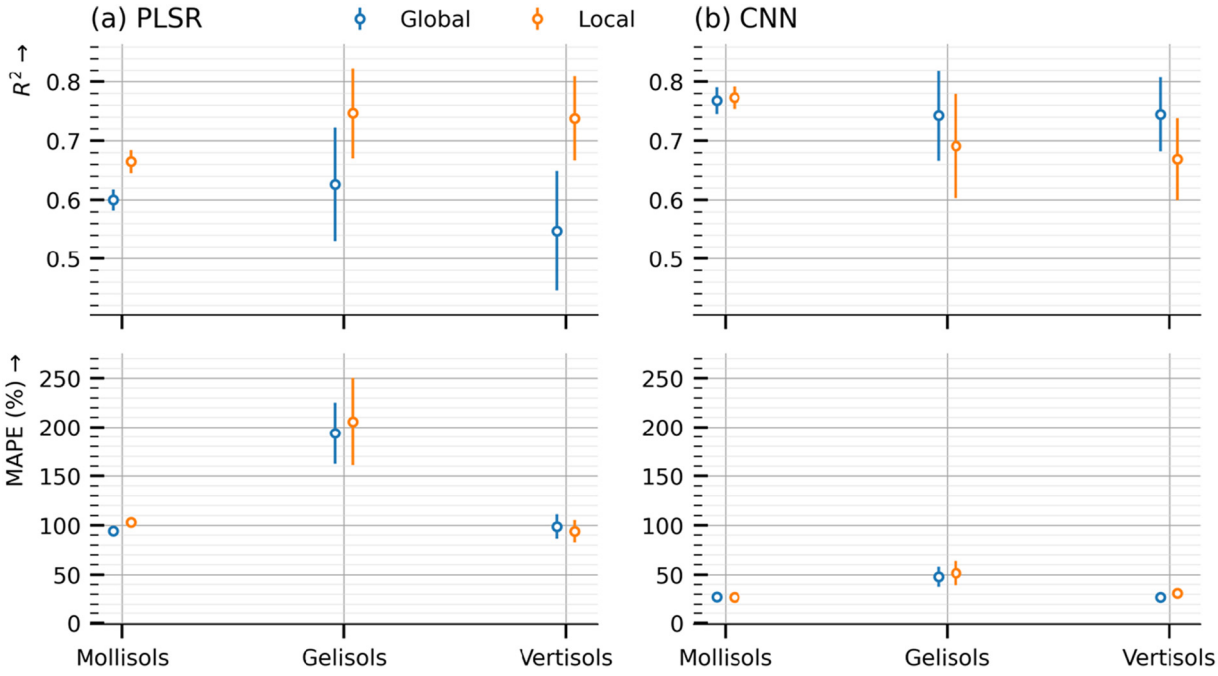


Fig. 6. Comparison of test performances, coefficient of determination R^2 and Mean Absolute Percentage Error (MAPE) for both PLSR (a) and CNN models (b) on Mollisols, Gelisols and Vertisols Soil Taxonomy Orders when trained (i) on all orders included (Global) and (ii) on Mollisols, Gelisols or Vertisols separately. Metrics reported were computed on 20 different random splits.

On the other hand, the GradientShap values computed for Andisols, Aridisols and Ultisols show little correlation with other orders and appear to be distinctive. Finally, the GradientShap values of Gelisols and Histosols correlate strongly with each other but show a low correlation with the other orders, suggesting a distinctive cluster.

4. Discussion

4.1. Performance analysis

Soils with high radiocaesium transfer risk to crops require K_{ex} of around 1 cmol(+).kg^{-1} to bottom out radiocaesium transfer to crops (Hirayama et al., 2018). Hence, in the context of the remediation of radioactive contamination, the ability to accurately characterize levels of K_{ex} lower than 1 cmol(+).kg^{-1} is of critical importance. At such lower potassium concentrations, the available potassium is not enough to avoid uptake of radioactive elements, such as radiocaesium, typically present in case of a fallout linked to a nuclear emergency. Hence, the much lower MAPE errors of the CNN (31%) over the PLSR (135%) models constitutes a major opportunity for this application domain.

For higher K_{ex} concentration (higher than 1 cmol(+).kg^{-1}), as indicated in Fig. 4, a clear bias can be observed for the PLSR model. In addition, the PLSR model appears to be dominated by majority of measurements available in the dataset within the 0.3 to 1 cmol(+).kg^{-1} range (−0.5 to 0 when $\log_{10}$-transformed). In contrast, the CNN model has clearly superior predictive power in this range, despite the lack of representativeness of this type of data in the training set.

As mentioned, the early stopping technique was used to train our models. In the case of the CNN, the training was stopped after 3 epochs without validation loss decrease to avoid overfitting. However, Fig. 5 showing the validation curves by Soil Taxonomy Orders allows for a more in-depth and differentiated analysis. Two groups of validation curves were identified, one group of curves showing a steady decrease throughout the training period and another group whose curves often level off or start increasing again by the end of the training period. Our interpretation is that the overall validation loss decrease was mainly driven by Mollisols and Alfisols orders, the two most represented orders in the KSSL dataset. More training epochs would have conducted most likely to even better performance on those orders. However, the group including Gelisols, Spodosols and Ultisols validation losses stopped decreasing after 100 epochs (halfway through the training period). We are overfitting those soil orders. Hence, a more conservative approach would have been to stop training at 100 epochs.

4.2. The capacity to leverage growing data regimes

A question that often arises is how many samples are needed to use Deep Learning models. The analysis of the Learning curves shown in Fig. 3 allows to answer this question for our specific application domain. The learning curves of both the PLSR and CNN models for K_{ex} prediction highlight the importance of the data regime with respect to model capacity. In this study, the CNN model is found to be superior to the PLSR for the prediction of K_{ex} when the data regime exceeds about 10,000 samples, where the performance of the PLSR model reaches a plateau at an R^2 of about 0.65 while the performance of the CNN continues to increase until it reaches an R^2 close to 0.8. The learning curves show also high standard deviation in low data regimes. This is an artifact of the evaluation scheme used. While train, validation, test random splits in the respective proportions: 80%, 10% and 10% are relevant in high data regimes and commonly used, in lower data regime a k-fold evaluation scheme or at least train, validation, test random splits in the respective proportions: 60%, 20% and 20% would have been more suitable. However, this limitation does not affect our rationale as (i) the size of the splits, sometimes insufficient, is reflected in the standard deviation of the metrics reported and (ii) our experiment focuses on the trend in the highest and growing data regimes.

4.3. CNN interpretability and domain knowledge

Fig. 6 shows a superior performance of the CNN model in predicting K_{ex} for Vertisols when trained on all orders than when

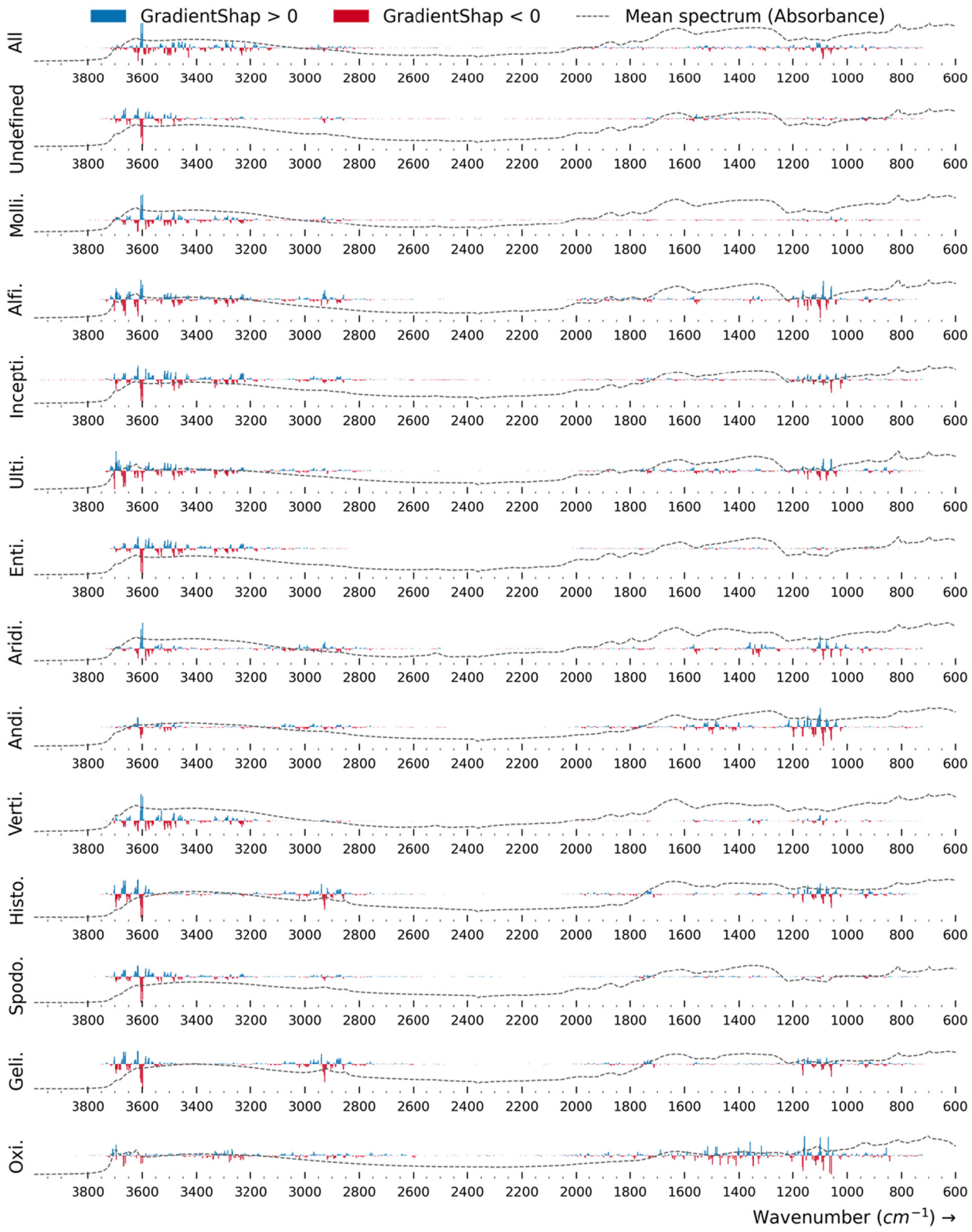


Fig. 7. GradientShap values for CNN based K_{ex} (log10-transformed) prediction by Soil Taxonomy Orders (Oxi.: Oxisols, Geli.: Gelisols, Spodo.: Spodosols, Histo.: Histosols, Verti: Vertisols, Andi.: Andisols, Aridi.: Aridisols, Enti.: Entisols, Ulti.: Ultisols, Incepti.: Inceptisols, Alfi.: Alfisols, Molli.: Mollisols). A positive (blue) value indicates that distorting the spectrum upwards would lead to an increase of predicted K_{ex}; a negative (red) value indicates that distorting the spectrum upwards would lead to a decrease of predicted value, or equivalently distorting the spectrum downwards would lead to an increase of predicted value.

trained on Vertisols only. A further analysis of computed GradientShap values on different Soil Taxonomy Orders and in particular their levels of correlation as shown in Fig. 8 provides us with additional interpretation keys.

Table 4 is a synthesis of our analysis whose objectives are (i) to identify situations where spectral features are reused/transferred between different soil orders, (ii) to hypothesize the predominance of certain soil orders in the "Undefined" class and last (iii) to further analyze

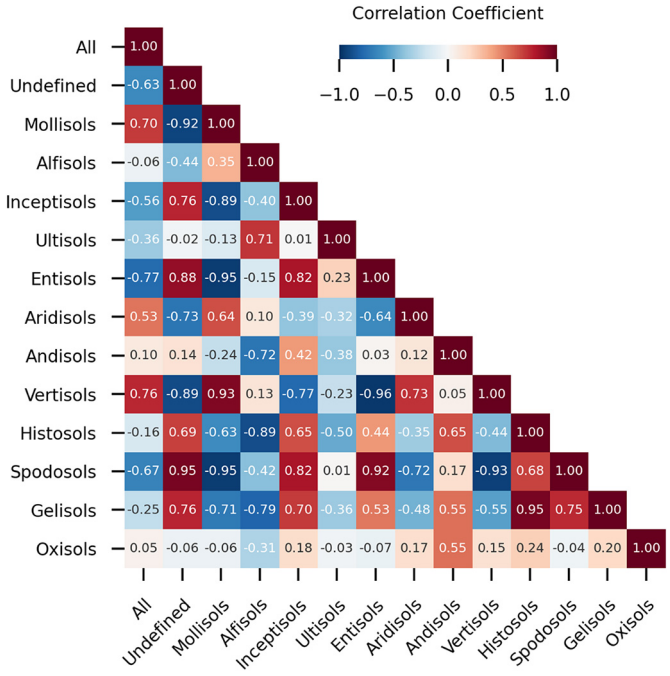


Fig. 8. Correlation matrix of GradientShap values computed based on the CNN model prediction of K_{ex} (log_{10}-transformed) by Soil Taxonomy Orders.

spectral features considered as important by the CNN model when predicting K_{ex}.

As summarized in Table 4, CNN model performance seems to be driven by two competing forces, namely the data regime and the distinctive nature of the GradientShap values computed for each Soil Taxonomy Orders, as characterized by their level of correlation over all wavenumbers or specific spectral regions. This analysis scheme is employed below to highlight the potential role played by the transfer and reuse of spectral features learned by the CNN for each type of soil. The different scenarios will then be interpreted in light of domain expertise.

In general, the pattern of higher level of performance as the data regime increases is consistent with the summary of Table 4. However there are exceptions.

For instance, the case of Vertisols is instructive, where, despite the low data regime, very good level of performance is reported. The analysis of the correlation of its GradientShap values shows that it is (negatively) highly correlated with Mollisols and Entisols, both soil orders very well represented in the KSSL database and with Spodosols. The 2:1 type clay minerals (e.g., micas, vermiculites, and smectites) determine soil's ability to retain potassium in most cases (Sawhney, 1970; Sharpley, 1989; Parfitt, 1992; Coulombe et al., 1996). In general, Vertisols have high clay content with relatively simple clay mineralogy dominated by smectite (Coulombe et al., 1996). Spodosols has sand to loam texture (McKeague et al., 1983), whereas the texture of Vertisols is, in definition, heavier than clay loam (Coulombe et al., 1996; Baillie, 2001; Soil Survey Staff, 2014). In addition, the major 2:1 type clay minerals, if any, in surface soil of Spodosols are irregular interstratified mica-vermiculite-smectites, not pure smectites (McKeague et al., 1983). In Entisols, due to the low degree of mineral weathering, soil texture and clay mineralogy are diverse reflecting the nature of parent materials (Allen and Fanning, 1983; Grossman, 1983). Fig. 7 highlights this contrasted clay mineralogy with symmetric GradientShap values of Vertisols and Entisols and Spodosols in the spectral region 3700–3400 cm^{-1}. Furthermore, a high level of correlation is observed between Vertisols and Mollisols (0.93). In the United States, smectites are also common in Mollisols (Fenton, 1983; Nettleton and Peterson, 1983). The similar smectitic nature of Vertisols and Mollisols may have also driven the unexpected high level of performance for Vertisols, suggesting a reuse of spectral features learned on Mollisols, very well represented in the KSSL database.

In contrast, the level of performance for Histosols and Gelisols is the lowest. Their GradientShap values are highly correlated with each other (0.95) but not with the other orders. Indeed, they show unique spectral features importance in the 3000–2800 cm^{-1} region, related to high organic matter content that control the cation exchange capacity of the soil and, hence K_{ex} (Le Guillou et al., 2015). Both their low data regime and unique GradientShap values might explain the poor performance.

The level of performance on Spodosols is medium despite the low data regime. Most of the Spodosols samples were discarded during the data selection step as below the detection level. Again, a high level of correlation of its GradientShap values is observed with the Undefined class, Mollisols and Entisols. Given these high levels of correlation, however, we would have expected a level of performance comparable to that achieved for Vertisols. This situation requires further investigation.

Table 4
Synthesis of relationships between level of performance, data regime and correlation of GradientShap values by Soil Taxonomy Orders and from the perspective of a CNN predicting K_{ex}. Soil Taxonomy Orders are ranked in ascending order of performance.

Level of performance[a]	Order	Data regime[b]	Highly correlated with[c]
Low (>40%)	Histosols	Low	Gelisols (0.95)
	Gelisols	Low	Histosols (0.95)
Medium (30%–40%)	Spodosols	Low	Undefined (0.95)
			Mollisols (−0.95)
			Entisols (0.92)
			Vertisols (−0.93)
	Ultisols	Medium	None
	Inceptisols	High	None
	Aridisols	Medium	None
	Entisols	Medium	Mollisols (−0.95)
			Vertisols (−0.96)
	Andisols	Medium	None
High (<30%)	Alfisols	High	None
	Mollisols	High	Vertisols (0.93)
			Entisols (−0.95)
			Spodosols (−0.95)
	Vertisols	Low	Mollisols (0.93)
			Entisols (−0.96)
			Spodosols (−0.93)

[a] Based on Mean Absolute Percentage Error (MAPE) on the test set as repored in Table 3.
[b] Based on the number of samples in the test set as reported in Table 3 (column N): Low below 100 samples, 100 < Medium <200 and High for higher number of samples otherwise.
[c] Correlation coefficients higher than 0.9 are considered only.

The level of performance on Aridisols is also medium despite its medium data regime and distinctive GradientShap values. Its GradientShap values profile shows high values in the 1400–1200 cm^{-1} spectral region. According to Johnston and Aochi, 1996, absorbance peaks in that spectral region relates to organic matter content. The calculated GradientShap values of orders with high organic matter content such as Histosols and Gelisols do not highlight such spectral region as important. The interpretation of this paradox requires further investigations.

The predominance of the "Undefined" Soil Taxonomy class requires us to take precautions in the case-by-case analysis of model performance by Soil Taxonomy Orders. A closer look at the GradientShap values shows a high positive correlation with Entisols (0.88), Spodosols (0.95), Gelisols (0.76), Inceptisols (0.76) might suggest the presence of these classes in the "Undefined" class. However, the interpretation of the GradientShap correlation values remains difficult given the potentially complex mixture of very different soil types included in this class.

Finally, in line with our interpretative framework, the level of performance on Andisols is medium (33%) and high (27%) for Alfisols with respectively medium and high data regimes. Andisols are of volcanic origin and have a distinct mineralogy with more sesquioxides. Their GradientShap values are also distinctive with all wavenumbers between 3700 and 2900 cm^{-1} and 1400–900 cm^{-1} considered important by the CNN to predict K_{ex}.

Whether it is preferable to stratify MIRS modeling by Soil Taxonomy Orders rather than including them all together has been a long-standing debate in the Mid-infrared soil spectroscopy research community (Torres Astorga et al., 2018). Our study proposes a methodology to inform this decision in the light of the spectral features learned and reused by the prediction algorithm. As a first analysis, we have shown that the quality of the prediction of K_{ex} is on the one hand indexed on the availability of a large data volume and on the other hand on model classes capable to leverage it. A more detailed analysis of these regimes by Soil Taxonomy Orders, however, revealed that the predictions of K_{ex} for some underrepresented classes such as Vertisols is possible. Hence, we have shown that the pooling of all Soil Taxonomy Orders is beneficial in this circumstance but for instance not, in the case for Histosols and Gelisols. Such an analysis is instructive and would be interesting to conduct also in the context of prediction of other soil analytes.

5. Conclusion

The approach and methodology developed in this paper and its future developments constitute a major opportunity for the remediation of radioactive contamination in agriculture. They open the possibility of high throughput characterization of K_{ex} under different soil conditions. The level of predictive performance achieved now allows precise monitoring of the effects of the remediation action, especially in the range of K_{ex} values influencing the dynamics of root uptake. Moreover, the proposed approach allows to equip the experts of the domain, in our case soil scientists, with tools for introspection and interpretation of the model to better understand the modeled phenomenon.

Beyond the progress made in this work, there are several avenues for future research. First, the potential of Transfer Learning as a major opportunity to cut down on the important cost of spectral data acquisition and leverage pre-trained state of the art Deep Learning architectures should be further considered. Second, in this study we have focused on Mid-Infrared (MIR) data. However, Near-Infrared (NIR) data might also offer the opportunity to characterize K_{ex} at an even higher throughput given the availability of low-cost portable spectrometers. The question of its level of performance remains open and justifies further investigations. Third and last, MIR and NIR spectroscopy allows local or "point" estimation of K_{ex}. However in the context of the remediation of farmland affected by nuclear accidents, the possibility of a swift characterization of K_{ex} in soil at landscape scale represents a major challenge. The question of determining the way that remote sensing and soil spectroscopy data can be used together in this context remains open and needs further exploration.

Author contributions

F.A and G.D designed the study. F.A developed the Machine and Deep Learning pipelines and analyzed the results from the Machine/Deep Learning perspective. T.E, G.D analyzed the result from radioecology and remediation of radioactive contaminated agricultural land perspectives, E.S evaluated the significance of the text from an environmental perspective while Y·P brought his extensive knowledge of soil spectroscopy.

CRediT authorship contribution statement

Franck Albinet: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft. **Yi Peng:** Methodology, Writing – review & editing. **Tetsuya Eguchi:** Writing – review & editing. **Erik Smolders:** Writing – review & editing. **Gerd Dercon:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was carried out in the context of the IAEA funded Coordinated Research Project (CRP D1.50.19) titled "Remediation of Radioactive Contaminated Agricultural Land", under IAEA Technical Contract n°23685. This work is a direct outcome from the collaboration between IAEA and The Food and Agriculture Organization (FAO) of the United Nations. We thank the support from the Global Soil Partnership of the FAO and its Global Soil Laboratory Network initiative on soil spectroscopy (GLOSOLAN-Spec) which is actively supporting countries to develop their capacities on soil spectroscopy. We also thank Richard Ferguson from Kellogg Soil Survey Laboratory for providing access to the USDA MIR soil spectra library and the required training sessions for its operation. Finally the authors are thankful for the comments by the reviewers of this paper and particularly to Raphael *V*iscarra Rossel.

References

Allen, B.L., Fanning, D.S., 1983. Composition and soil genesis. Developments in Soil Science. Elsevier, pp. 141–192 https://doi.org/10.1016/S0166-2481(08)70601-9.

Baillie, I.C., 2001. Soil Survey Staff 1999, Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys, Agricultural Handbook 436. Natural Resources Conservation Service, USDA, Washington DC, USA, p. 869.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–777.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Prediction of soil attributes by NIR spectroscopy. A critical review of chemometric indicators commonly used for assessing the quality of the prediction. Trends Anal. Chem. 29, 1073–1081.

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N.P., Bernoux, M., Barthès, B.G., 2016. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. Geoderma 276, 41–52.

Coulombe, C.E., Wilding, L.P., Dixon, J.B., 1996. Overview of Vertisols: characteristics and impacts on society. Adv. Agron. 57, 289–375.

Dangal, S.R.S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Syst. 3, 11. https://doi.org/10.3390/soilsystems3010011.

Farmer, V.C., 1974. Vibrational Spectroscopy in Mineral Chemistry.

Fenton, T.E., 1983. Chapter 4 Mollisols. In: (Eds.), Developments in Soil Science. Elsevier, p. 125 https://doi.org/10.1016/S0166-2481(08)70615-9.

Frost, R.L., Palmer, S.J., Reddy, B.J., 2007. Near-infrared and mid-IR spectroscopy of selected humite minerals. Vib. Spectrosc. 44, 154–161.

Gomez, C., Chevallier, T., Moulin, P., Bouferra, I., Hmaidi, K., Arrouays, D., Jolivet, C., Barthès, B.G., 2020. Prediction of soil organic and inorganic carbon concentrations

in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. Geoderma 375, 114469.
Grinand, C., Barthes, B.G., Brunet, D., Kouakoua, E., Arrouays, D., Jolivet, C., Caria, G., Bernoux, M., 2012. Prediction of soil organic and inorganic carbon contents at a national scale (France) using mid-infrared reflectance spectroscopy (MIRS). Eur. J. Soil Sci. 63, 141–151.
Grossman, R.B., 1983. Chapter 2 Entisols. Developments in Soil Science. Elsevier, pp. 55–90 https://doi.org/10.1016/S0166-2481(08)70613-5.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
Hirayama, T., Takeuchi, M., Nakayama, H., Nihei, N., 2018. Effects of decreasing radiocesium transfer from the soil to soybean plants and changing the seed nutrient composition by the increased application of potassium fertilizer. Bull. Fukushima Agric. Tech. Cent. 9, 1–10.
Janik, L.J., Skjemstad, J.O., 1995. Characterization and analysis of soils using mid-infrared partial least-squares. 2. Correlations with some laboratory data. Soil Res. 33, 637–650.
Johnston, C.T., Aochi, Y.O., 1996. Fourier transform infrared and Raman spectroscopy. Methods Soil Anal. Part 3 Chem. Methods 5, 269–321.
Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. ArXiv (Prepr. ArXiv14126980).
Klesta, E.J., Bartz, J.K., 1996. Quality assurance and quality control. Methods Soil Anal. Part 3 Chem. Methods 5, 19–48.
Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Alsallakh, B., Reynolds, J., Melnikov, A., Kliushkina, N., Araya, C., Yan, S., Reblitz-Richardson, O., 2020. Captum: A Unified and Generic Model Interpretability Library for PyTorch.
Komatsu, M., Hirai, K., Nagakura, J., Noguchi, K., 2017. Potassium fertilisation reduces radiocesium uptake by Japanese cypress seedlings grown in a stand contaminated by the Fukushima Daiichi nuclear accident. Sci. Rep. 7, 1–10.
Landre, A., Saby, N.P.A., Barthès, B.G., Ratié, C., Guerin, A., Etayo, A., Minasny, B., Bardy, M., Meunier, J.-D., Cornu, S., 2018. Prediction of total silicon concentrations in French soils using pedotransfer functions from mid-infrared spectrum and pedological attributes. Geoderma 331, 70–80.
Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 255–268.
Le Guillou, F., Wetterlind, W., Viscarra Rossel, R.A., Hicks, W., Grundy, M., Tuomi, S., 2015. How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? Soil Res. 53, 913–921.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774.
McKeague, J.A., DeConinck, F., Franzmeier, D.P., 1983. Chapter 6 Spodosols. Developments in Soil Science. Elsevier, pp. 217–252 https://doi.org/10.1016/S0166-2481(08)70617-2.
Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in partial least squares regression. Chemom. Intell. Lab. Syst. 118, 62–69.
Molnar, C., Casalicchio, G., Bischl, B., 2020. Interpretable machine learning–a brief history, state-of-the-art and challenges. ArXiv (Prepr. ArXiv201009337).
Nettleton, W.D., Peterson, F.F., 1983. Chapter 5 Aridisols. Developments in Soil Science. Elsevier, pp. 165–215 https://doi.org/10.1016/S0166-2481(08)70616-0.
Ng, W., Minasny, B., Malone, B.P., Sarathjith, M.C., Das, B.S., 2019a. Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling. Comput. Electron. Agric. 158, 201–210.
Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019b. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma 352, 251–267.
Ng, W., Minasny, B., Jeon, S.H., McBratney, A., 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Secur. 100043.
Parfitt, R.L., 1992. Potassium-calcium exchange in some New Zealand soils. Soil Res. 30, 145–158.
Peng, Y., Knadel, M., Gislum, R., Schelde, K., Thomsen, A., Greve, M.H., 2014. Quantification of SOC and clay content using visible near-infrared reflectance–mid-infrared reflectance spectroscopy with Jack-knifing partial least squares regression. Soil Sci. 179, 325–332.
Sanderman, J., Savage, K., Dangal, S.R., 2020. Mid-infrared spectroscopy for prediction of soil health indicators in the United States. Soil Sci. Soc. Am. J. 84, 251–261.
Sawhney, B.L., 1970. Potassium and cesium ion selectivity in relation to clay mineral structure. Clay Clay Miner. 18, 47–52.
Sharpley, A.N., 1989. Relationship between soil potassium forms and mineralogy. Soil Sci. Soc. Am. J. 53, 1023–1028.
Shen, Z., Viscarra Rossel, R.A., 2021. Automated spectroscopic modelling with optimised convolutional neural networks. Sci. Rep. 11, 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 J. Photogramm. Remote Sens. 188, 190–200. https://doi.org/10.1016/j.isprsjprs.2022.04.009.
Shepherd, K.D., Walsh, M.G., 2007. Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. J. Infrared Spectrosc. 15, 1–19.
Shrikumar, A., Greenside, P., Kundaje, A., 2019. Learning important features through propagating activation differences. ArXiv (170402685 Cs).
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. ArXiv (Prepr. ArXiv14091556).
Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M., 2017. SmoothGrad: Removing Noise by Adding Noise. Doi: 10.48550/arXiv.1706.03825.
Smith, L.N., 2017. Cyclical learning rates for training neural networks. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 464–472.
Socrates, G., 2004. Infrared and Raman Characteristic Group Frequencies: Tables and Charts. John Wiley & Sons.
Soil Survey Staff, S, 2014. Keys to Soil Taxonomy. 12th edn. Washington. DC Nat. Resour. Conserv. Serv. U. S. Dep. Agric. Sch.
de Sousa Mendes, W., Demattê, J.A., Rosin, N.A., da Silva Terra, F., Poppiel, R.R., Urbina-Salazar, D.F., Boechat, C.L., Silva, E.B., Curi, N., Silva, S.H.G., 2022. The Brazilian soil mid-infrared spectral library: the power of the fundamental range. Geoderma 415, 115776.
Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. ArXiv (170301365 Cs).
Terhoeven-Urselmans, T., Vagen, T.-G., Spaargaren, O., Shepherd, K.D., 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. Soil Sci. Soc. Am. J. 74, 1792–1799.
Torres Astorga, R., de los Santos Villalobos, S., Velasco, H., Domínguez-Quintero, O., Pereira Cardoso, R., Meigikos dos Anjos, R., Diawara, Y., Dercon, G., Mabit, L., 2018. Exploring innovative techniques for identifying geochemical elements as fingerprints of sediment sources in an agricultural catchment of Argentina affected by soil erosion. Environ. Sci. Pollut. Res. 25, 20868–20879.
Viering, T., Loog, M., 2021. The Shape of Learning Curves: A Review.
Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54.
Viscarra Rossel, R.A., Bouma, J., 2016. Soil sensing: a new paradigm for agriculture. Agric. Syst. 148, 71–74.
Viscarra Rossel, R.A., Lark, R.M., 2009. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. Eur. J. Soil Sci. 60, 453–464.
Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.
Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A., McBratney, A.B., 2008. Using a legacy soil sample to develop a mid-IR spectral library. Soil Res. 46, 1–16.
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., 2016. A global spectral library to characterize the world's soil. Earth-Sci. Rev. 155, 198–230.
Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Ge, Y., Gomez, C., Guerrero, C., Peng, Y., Ramirez-Lopez, L., 2022. Diffuse reflectance spectroscopy for estimating soil properties: a technology for the 21st century. European Journal of Soil Science 73 (4).
Wijewardane, N.K., Ge, Y., Wills, S., Libohova, Z., 2018. Predicting physical and chemical properties of US soils with a mid-infrared reflectance spectral library. Soil Sci. Soc. Am. J. 82, 722–731.
Zeiler, M.D., Fergus, R., 2013. Visualizing and Understanding Convolutional Networks.