GI6.5 EGU23-5501
vHall ESSI/GI/NP

Predicting cesium dynamics and uptake using Mid-Infrared Spectroscopy

Ayane Kan[1,3], Maria Heiling[1], Arsenio Toloza[1], Franck Albinet[2], Takuro Shinano[3], and Gerd Dercon[1]

[1]Soil and Water Management & Crop Nutrition Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Seibersdorf, Austria
[2]Independent Researcher & Consultant, Guéthary, France, [3]Graduate School of Agriculture, Hokkaido University, Hokkaido, Japan

Introduction

The release of radiocaesium radionuclides (RCs) has affected food safety after the accidents at Fukushima Daiichi nuclear power plants. ^{137}Cs, in particular, is of major concern due to its relatively long half-life and easy absorption by plants.

Potassium (K) fertilization has successfully reduced Cs uptake by plants. However, the relation between the transfer from soil to crop and exchangeable potassium (K_{ex}) differs depending on the soil. Hence, understanding the Cs dynamics in the soil is important.

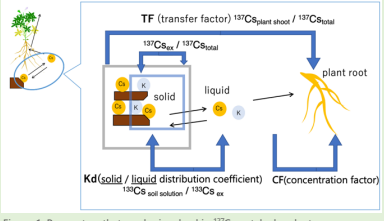


Figure 1. Parameters that can be involved in ^{137}Cs uptake by plants

- the solid/liquid distribution coefficient (Kd) of Cs
- the ratio of exchangeable ^{137}Cs versus total ^{137}Cs

could be involved in the determination of the risk of ^{137}Cs uptake by plants, whereas the methods to obtain this soil information are often time-consuming.

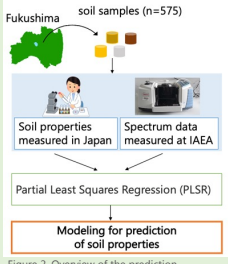
Therefore, Mid-Infrared Spectroscopy (MIRS) may be utilized for the rapid determination of these parameters. However, the prediction of them has yet to be assessed.

Objective

To assess whether MIRS can predict Cs-related parameters in the soil such as Kd, $^{137}Cs_{ex}$/$^{137}Cs_{total}$, K_{ex} and other parameters that may influence on the behaviour of RCs in the soil, such as total carbon.

Materials & Methods

In total, about 1685 soil samples were collected from agriculture fields in the Fukushima Prefecture in Japan from 2013 to 2019. So far, 575 out of 1685 soil samples were analyzed.



Spectra data was obtained with four replicates in each soil sample.

Prediction by PLSR (Partial Least Squares Regression)

- Kd
- $^{137}Cs_{ex}$/$^{137}Cs_{total}$
- K_{ex}
- Total carbon

Figure 2. Overview of the prediction

Table 1. Modeling pipeline overview

Phase	Model : PLSR
Data Selection	log $_{10}$ transformation, Average of spectra in each soil sample (n=4)
Preprocessing (Spectra)	1. CO_2 spectral region removal 2. Standard normal variance (to be systematically centered and normalized) 3. Savizky-Golay filtering (noise removal) 4. Outlier detection and removal (Mahalanobis distance)
Evaluation scheme	Training (75%) and Test (25%) *split randomly 20 times (The number of principle components was decided from RMSEP calculated by 10-fold cross validation on the training dataset)
Metrics	RMSE, R^2 (Coefficient of determination)

Results & Discussion

The PLSR methods provided a relatively high accuracy score for the prediction of soil total carbon, where the R^2 (coefficient of determination) of the test dataset was 0.82±0.04 (Table 2). However, due to the limitation of the available dataset of Kd, the validation scores of Kd of ^{133}Cs were around 0.2.

For the ratio of ^{137}Cs, the R^2 value was 0.36±0.07 (Figure 3) and the accuracy had been improved with the addition of the dataset, indicating that the estimation accuracy of the ratio potentially increases. The variable importance on projection (Figure 4) indicated that the region that can be related to clay mineralogy, quartz, and organic matter was considered influential in estimation of $^{137}Cs_{ex}$ / $^{137}Cs_{total}$.
The accuracy of the estimation of K_{ex} was not high, probably because of the narrow range of the dataset and the complicated dynamics of K in Andosol.

Table 2. PLSR model performance in predicting soil properties. Figures reported are averages and standard deviations of metrics calculated on test datasets split 20 times randomly. N is the number of soil samples in the whole dataset. Note that the unit and the RMSE are reported as the information of original (untransformed) data, while R^2 score is reported as the information of the transformed data.

				Metrics					
Soil properties	N	nc	unit	RMSE			R^2		
Total carbon	567	6	%	0.54	±	0.03	0.82	±	0.04
Kd	162	5	L/kg	31.6	±	12.2	0.13	±	0.09
$^{137}Cs_{ex}$ / $^{137}Cs_{total}$	554	10	-	0.04	±	0.003	0.36	±	0.07
K_{ex}	573	10	mg/100gDW	19.7	±	2.3	0.16	±	0.08

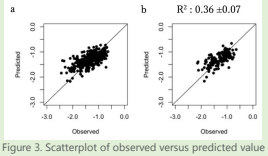


Figure 3. Scatterplot of observed versus predicted value of log $_{10}$ ($^{137}Cs_{ex}$ / $^{137}Cs_{total}$). The left-hand side plot (a) shows the observed against predicted values for the training dataset, while the right-hand side plot (b) shows the observations against predictions for the test dataset. The dataset was split by randomly and one of them are shown.

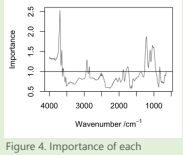


Figure 4. Importance of each wavenumber on projection of the principal components on the estimation for log $_{10}$ ($^{137}Cs_{ex}$ / $^{137}Cs_{total}$).

Conclusion

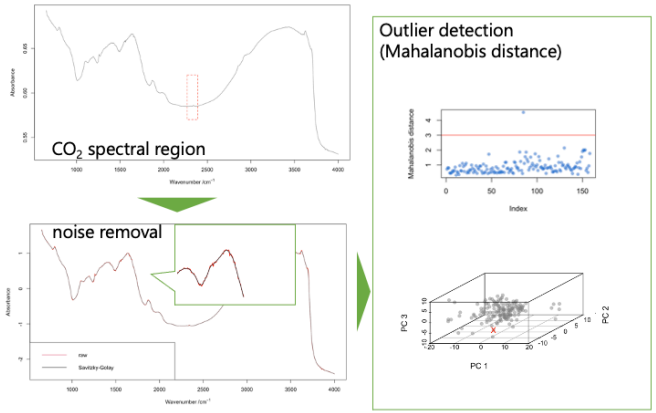
The prediction of one of Cs- related parameters, $^{137}Cs_{ex}$ / $^{137}Cs_{total}$ in soil, using MIRS was properly performed although it still requires further investigation. Estimation of Kd provided low accuracy scores due to a limited number of samples and low values of original Kd data. Prediction of $^{137}Cs_{ex}$ / $^{137}Cs_{total}$ can be improved by the addition of a dataset.

Modeling pipeline overview

Phase	Task	Model : PLSR
1.Data	Selection	Average of four replicates
	Preprocessing (Spectra)	1. CO_2 spectral region removal (2268 ~ 2390 cm^{-1}) 2. Standard normal variance (to be systematically centered and normalized) 3. Savizky-Golay filtering (noise removal) 4. Outlier detection and removal(Mahalanobis distance, H distance)
2.Modeling	Evaluation scheme	Train (75%) and Test (25%) (The number of principle components were decided from RMSEP calculated by 10-fold cross validation on the calibration dataset)
	Metrics	R^2 (Coefficient of determination) ME, RMSE, r^2 (square of Pearson's r correlation coefficient), rhoC, RPD, RPIQ

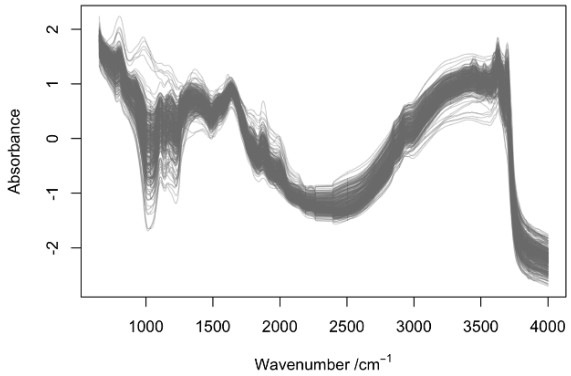
1

Pre-processing



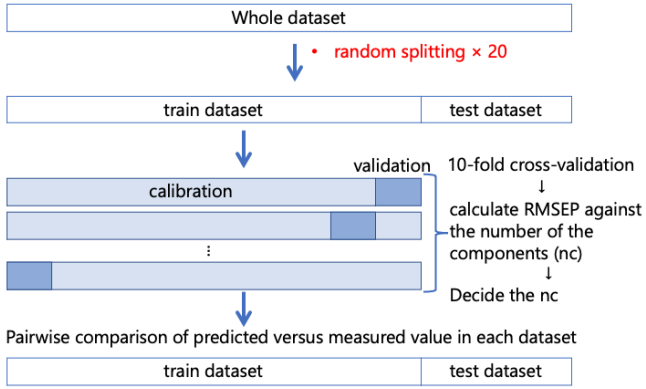
2

The spectra data after pre-processing (example)



3

Data splitting



4

Estimating soil carbon from Spectra collected using all data, not average

Phase	Task	Model : PLSR
1.Data	Selection	Spectra data collected for three months (removed data whose value = 0),four replicates (not average), log $_{10}$ transformation
	Preprocessing (Spectra)	1. CO_2 spectral region removal (2268 ~ 2390 cm^{-1}) 2. Standard normal variance (reduce variation) 3. Savizky-Golay filtering (noise removal) 4. Outlier detection and removal(Mahalanobis distance, H distance) -> removed ID: 212d,2002b,195d,195c,195b,218a,277c,178a -> removed ID: 195a, 195b, 195c, 277c (for Kd estimation)
2.Modeling	Evaluation scheme	Remove rows in which the target soil property is NA Train (75%) and Test (25%) using Kennard-Stone or random *20 (The number of principle components were decided from RMSEP calculated by 10-fold cross validation on the calibration dataset)
	Metrics	R2 (Coefficient of determination) ME, RMSE, r2, rhoC, RPD, RPIQ

5

Estimating soil carbon from Spectra collected using all data, not average

Table. PLSR model performance in predicting soil properties. Figures reported are averages and standard deviations of metrics calculated on test datasets split 20 times randomly. n is the number of measurements (4 replicates per one soil sample) in the whole dataset. Note that the unit is reported as the information of original (untransformed) data, while RMSE and R2 score are reported as the information of the transformed data.

				Metrics				
Soil properties	n	nc	unit	RMSE			R^2	
Total carbon	2266	6	%	0.08	±	0.00	0.81	± 0.01
Kd	628	7	L/kg	0.40	±	0.02	0.21	± 0.07
$^{137}Cs_{ex}$ / $^{137}Cs_{total}$	2214	10	-	0.27	±	0.01	0.39	± 0.03
K_{ex}	2290	10	mg/100gDW	0.24	±	0.01	0.22	± 0.03

6