COMPARISON OF CALIBRATION MODELS DEVELOPED FOR A VISIBLE-NEAR INFRARED REAL-TIME SOIL SENSOR

B. S. N. Aliah
United Graduate School of Agriculture
Tokyo University of Agriculture and Technology
Tokyo, JAPAN

Mechanization and Automation Research Centre
Malaysian Agricultural Research and Development Institute (MARDI)
Kuala Lumpur, MALAYSIA

S. Shibusawa, M. Kodaira
Institute of Agriculture
Tokyo University of Agriculture and technology
Tokyo, JAPAN

K. Inoue
Graduate School of Agriculture
Tokyo University of Agriculture and Technology
Tokyo, JAPAN

ABSTRACT

The visible-near infrared (Vis-NIR) based real-time soil sensor (RTSS) is found to be a great tool for determining distribution of various soil properties for precision agriculture purposes. However, the developed calibration models applied on the collected spectra for prediction of soil properties were site-specific (local). This is found to be less practical since the RTSS needs to be calibrated separately for every field. General calibration approach is expected to minimize this limitation. This paper describes the feasibility of general calibration model developed from two types of paddy field and to compare the performance of the calibration models. For this purpose, Vis-NIR reflectance spectra of fresh soil were acquired at two fields (organic and inorganic paddy fields). Fresh soil samples were also collected from these two fields for analysis of moisture content (MC), organic matter (OM), total carbon (TC) and total nitrogen (TN) in the laboratory. Three calibration models were then developed for each soil properties using partial least square regression (PLSR) technique coupled with full cross-validation. The first model (CM1) was developed using dataset from organic field, second model (CM2) was from inorganic field and the third model (general model – CM3) was developed from combination of dataset from both fields. The performance of the three calibration models were compared based on the determination of coefficient (R_{val}^2), root mean square error of validation ($RMSE_{val}$) and residual prediction deviation (RPD). Results showed for MC and OM, CM3 produced highest prediction accuracy with R_{val}^2 of 0.90 and 0.95. For TC and TN, CM1 produced the highest accuracy. CM2 produced the lowest accuracy for all the soil properties. This result could be used as a step towards establishment a robust general calibration model for agriculture soil.

Keywords: calibration model, visible-near infrared, real-time soil sensor, organic, inorganic

INTRODUCTION

The visible-near infrared based real-time soil sensor (Vis-NIR RTSS) is found to be a great tool for describing distribution of various soil properties for precision agriculture purpose. It has been proven to be a rapid, inexpensive and relatively accurate tool for measuring soil properties. Furthermore, this sensing technology offers on-line measurement of soil properties at fine resolution sampling which may reduce the cost of producing map for precision agriculture application (Kodaira and Shibusawa, 2013).

At the heart of visible-near infrared (Vis-NIR) spectroscopy technique is the calibration model that relates reflectance spectra to soil properties measured by the laboratory analysis (reference value). In previous studies, researchers tended to develop local calibration model for each field they measured with Vis-NIR spectroscopy (Kodaira and Shibusawa, 2013, Mouazen et al., 2005). In other words, the developed calibration models applied on the collected spectra for prediction of soil properties were site-specific (local). This found to be less practical since the RTSS needs to be calibrated separately for every field. As consequences, the employment of RTSS for describing soil variability would become less time and cost effective because soil sampling and soil laboratory analysis need to be carried out every time when developing calibration model for every different field. General calibration approach is expected to minimize this drawback even though the accuracy of the model might less accurate but still good enough to be acceptable for farm management in precision agriculture application.

In order to establish a robust general calibration model, the Vis-NIR spectra and soil sample need be collected from a wide geographic range. Malley et al. (2004), Viscarra Rossel et al. (2006), and Stenberg et al. (2010) have drawn soils from a wide geographic range and thus do not directly address the use of reflectance spectroscopy to determine the soil properties for specific field. Estimation of soil properties in these studies has had varying degrees of success. The objective of this paper is therefore to describe the feasibility of general calibration model developed from two paddy fields that have different soil management (organic and inorganic) and to compare the performance of the calibration models. This study is a step towards developing a prediction models that could be applied to all Japan paddy field whatever the agricultural management history. We believe this is how predictive models are used in practice.

MATERIALS AND METHODS

Experimental Site

Japan country consists of four main islands which are Honshu, Hokkaido, Kyushu and Shikoku. The first field experiment was conducted at an organic paddy field at Matsuyama City of Ehime Prefecture in Shikoku Island, Japan (33º8’N, 132º8’E) (Fig. 1). This site comprises a number of small paddy fields where organic farm management is implemented. Field no. 437 (58.3m x 21.7m) was selected for this study. The experiment was conducted after paddy harvesting in autumn 2012. The average, maximum and minimum temperature of the day were 20.8, 26.3 and 14.5 ℃, respectively. The soil texture of the field was described according to three depths as follows: 52.82% sand, 24.71% silt and 22.47% clay at a depth of 10cm, 54.55% sand, 21.02% silt and 24.43% clay at a depth of 15cm, 66.29% sand, 11.82% silt and 21.89% clay at a depth of 20cm.

The second field experiment was carried out at an inorganic paddy farm at Yamatsuri City of Fukushima Prefecture in Honshu Island (36º52’N, 140º25’E) (Fig.1). The average, maximum and minimum temperatures of the day were 7.9, 10.9 and 4.9 ℃, respectively. Inorganic farm management was implemented at this site. The experiment was conducted after the paddy harvesting season in early winter 2013. Two fields selected for this study were Field 2 (0.27 ha) and Field 5 (0.41 ha). The soil textures of the two fields were described as follows: 66.0% sand, 13.6% silt and 20.4% clay of field 2 and 62.1% sand, 16.2% silt and 21.7% clay of field 5.

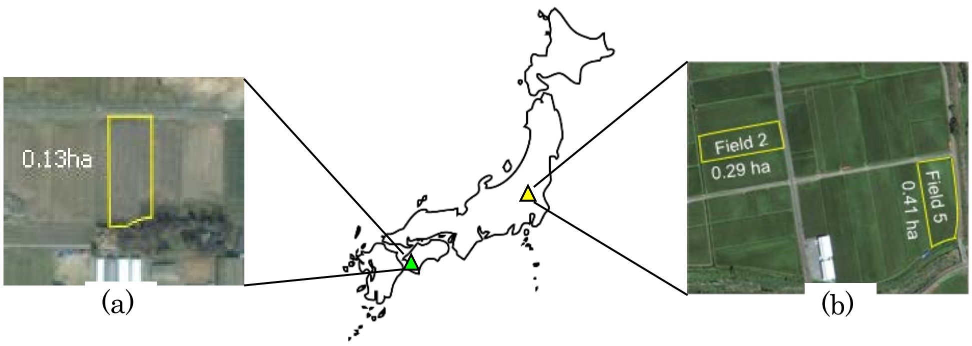


Fig. 1. Location of the experimental site (a) Matsuyama in Shikoku Island and (b) Fukushima in Honshu Island

Real-time Soil Sensor

The RTSS used for this study was SAS1000, SHIBUYA MACHINERY Co., Ltd. (Fig. 2). It comprises of sensor unit’s housing, a touch panel, soil penetrator and the housing for the probes. The sensor unit’s housing consists of a personal computer, differential global positioning system (DGPS) receiver, 150-W Al-coated tungsten halogen lamp as a light source and two spectrophotometers. The first spectrophotometer is for Vis (350 to 1100 nm), has a 256-linear diode array while the second spectrophotometer is for NIR (950 to 1700 nm), has a 128-pixel linear diode array of multiplexed InGaAs. In the probe housing, two optical fibers were used to guide the light from the light source (halogen lamp) and illuminate the underground soil surface with an area of about 50 mm in diameter. The underground soil Vis-NIR reflectance spectra were then collected through an additional optical fiber probes to the two spectrophotometers. The probe housing is also equipped with a micro CCD camera to capture, record and display the images of uniform soil surfaces while the RTSS running across the field. The saved images were then used to detect outlier in the calibration and prediction process. In addition, a laser line marker located close to the optical fiber was used to monitor distance variations between the soil surface and the micro optical devices.

Fig. 2. Real-time Soil Sensor SAS1000

Spectra Acquisition and Soil Sampling

For the first field experiment (organic farm, Matsuyama), the Vis-NIR reflectance spectra at range of 350 to 1700 nm were acquired at three depths: 0.10 m, 0.15 m and 0.20 m from the soil surface by adjusting the gage wheels of the RTSS. At each depth, the tractor attached with the RTSS was travelled on 4 transects at spacing of 5 m and at speed of 0.25 ms^{-1}. When the RTSS was running on the track, the soil penetrator tip with a flat plane edge ensured uniform soil cuts and the soil flattener following behind formed a trench with a uniform underground surface. The Vis-NIR reflectance spectra of the underground soil were acquired automatically from the bottom of the trench at every 4 s and this resulted in the Vis-NIR reflectance spectra being sampled at distance of every 1 m.

At each depth, two sets of twenty soil samples were subsequently collected at the bottom of the trench (5 samples at each transects) in-line with the RTSS's tracks. In total, there were 2 sets of 60 soil samples collected. However, due to the RTSS encountering obstacle at one point at a depth of 0.20 m, invalid spectra were acquired at that particular point. Hence, the soil sample corresponding to that single point was omitted from each set. Finally, only 59 soil samples of each set were collected from Matsuyama field.

For the second field experiment (inorganic farm, Fukushima), the Vis-NIR soil reflectance spectra were acquired at 0.15 m depth on two fields

(Field 2 and Field 5). The similar spectra range as in the first experiment was acquired at every 3 s while the tractor travelling at speed of 0.28 ms^{-1}. This resulted in the Vis-NIR reflectance spectra being sampled at every approximately 0.84 m distance

Two sets of 63 and 67 fresh soil samples were then collected from Field 2 and Field 5 respectively along the RTSS's tracks for analysis of moisture content (MC), organic matter (OM) and total carbon (TC) in laboratory. In total, there were two sets of 130 soil samples collected.

The first sets of soil samples collected from both fields were sent to Tokyo University of Agriculture and Technology Laboratory (TUAT) to determine the amount of MC and OM by applying oven-dried and loss ignition combustion method respectively. The second sets were sent to Agriculture Product Chemical Research Laboratory (APCRL) to determine the amount of TC using Tyurin's method and TN using Kjeldahl method.

Spectra Pre-treatment and Calibration Model Development

Prior to the development of calibration models, all collected underground Vis-NIR soil reflectance spectra were converted to absorbance using the Beer-Lambert's Law. The absorbance spectra were then converted to 5-nm-interval data by the interpolation method using Data Monitor Software (Shibuya Seiki Co., Ltd.). The original absorbance spectra range was from 350 to 1700 nm. To enhance weak signals and remove background noises, the absorbance spectra were pre-treated using second derivative Savitzky and Golay method. Moreover, both edges of the wavelengths were removed as these parts of the spectra were unstable and rich in noise. Hence, the spectra range of 500 to 1600 nm was further used for developing calibration models.

The calibration models were developed by applying the partial least square regression (PLSR) technique coupled with full-cross validation to establish the relationship between the amount of soil properties obtained by chemical analysis (reference values) with the pre-treated Vis-NIR soil absorbance spectra. These were performed using the Unscrambler X10.2 software. For each of the soil property, three calibration models were developed. The first model (CM1) was developed using the dataset (spectra and reference values) from Matsuyama field, the second model (CM2) was developed using the dataset from Fukushima field and the third model (CM3) combined the dataset from both fields. In the calibration process, outliers were detected by checking the residual sample variance plot after the PLSR. Individual samples located far from the zero line of residual variance on the validation views were considered to be outliers and excluded from the analysis. The performance of the three calibration models for each of the soil property was then assessed based on the value of coefficient of determination (R_{val}^2), root means square error of prediction ($RMSE_{val}$) and residual prediction deviation (RPD) produced from the PLSR analysis. RPD is given by the ratio of standard deviation (SD) of the reference dataset to the root mean square error of full cross-validation ($RMSE_{val}$)

The best calibration model that possesses maximum R_{val}^2 and RPD but minimum $RMSE_{val}$ in the regression analysis for each soil property was then used to provide quantitative prediction and mapping of the respective soil properties using ArcGIS Ver10.0 software. The soil maps were interpolated using the inverse distance weighing (IDW) method.

RESULTS AND DISCUSSION

Soil Compositions and Spectra Properties

The statistic result (Table 1) of the laboratory soil chemical analysis shows that the Fukushima soil possesses higher content of all the soil properties especially MC with the mean value is 32.94 %. However, the variability of Matsuyama field is higher than Fukushima field based on the coefficient of variation (CV) of both fields. By merging the both fields' data, the variability of the soil properties is higher than the just Fukushima field. However for Matsuyama field, the merging data increased the variability of only MC and OM but not TC and TN. The higher variability of Matsuyama field might be due to the soil samples were collected at three different depths unlike Fukushima, where the samples were collected at just single depth (0.15 m).

The mean 2nd derivatives of absorbance spectra for the two fields are depicted in Fig. 2. The spectral data were analyzed by principal component analysis (PCA). The two-dimensional scatter plot of PCA gives information about patterns among the samples (Fig. 3). The closer together samples were in the scatter plot, the more similar they were in composition as reflected in their spectra. The majority of samples were from the Fukushima field and these formed a fairly tight group based on their spectral properties. The samples from the Matsuyama field tended to separate from the Fukushima group along the PC-1 axis. From this PCA scatter plot also, the discrimination on the spectra for different depths of Matsuyama field can be observed.

Table 1. Statistical results of soil chemical analysis on soil properties in calibration dataset

Calibration Dataset	Statistics	MC (%)	OM (%)	TC (%)	TN (%)
Matsuyama (CM1)	N	59	59	59	59
	Mean	18.25	5.21	1.23	0.13
	Min	11.77	3.85	0.50	0.07
	Max	23.63	6.30	1.94	0.18
	SD	2.47	0.61	0.42	0.03
	CV	13.53	11.75	33.79	23.50
Fukushima (CM2)	N	130	130	130	130
	Mean	32.94	7.65	1.65	0.15
	Min	26.36	6.11	1.13	0.11
	Max	40.12	9.13	2.10	0.19
	SD	3.06	0.61	0.23	0.02
	CV	9.28	7.94	13.80	12.87
Matsuyama and Fukushima (CM3)	N	189	189	189	189
	Mean	28.35	6.89	1.52	0.15
	Min	11.77	3.85	0.50	0.07
	Max	40.12	9.13	2.10	0.19
	SD	7.40	1.29	0.35	0.03
	CV	26.12	18.67	23.40	17.63

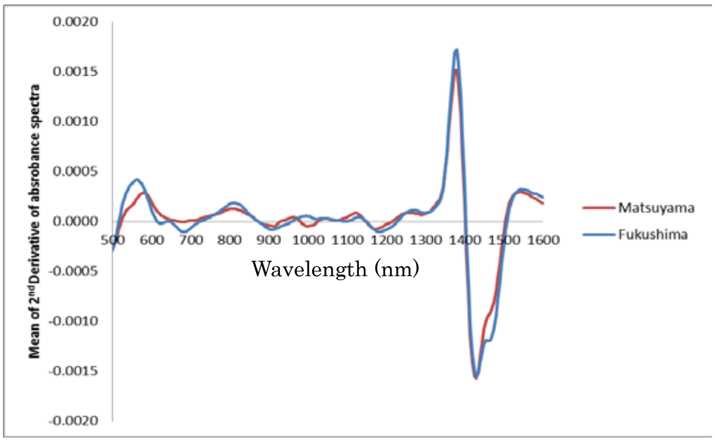
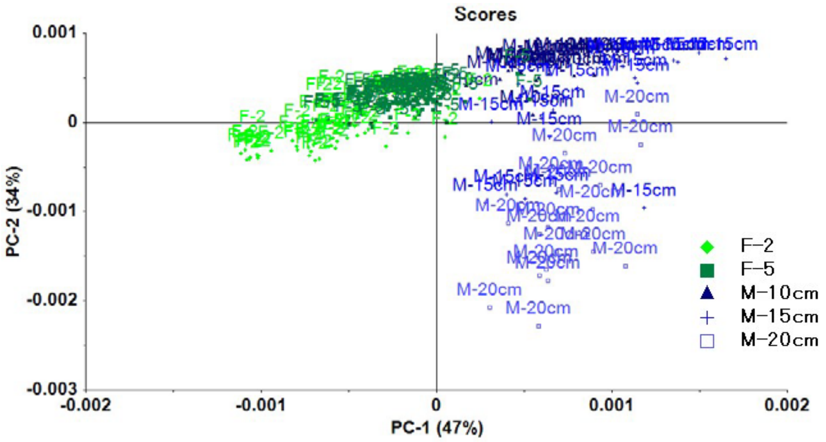


Fig. 3. The mean of 2nd Derivative of absorbance spectra of Matsuyama and Fukushima soil



F-2 : Fukushima Field 2, F-5 : Fukushima Field 5, M-10cm : Matsuyama at depth of 10cm, M-15cm : Matsuyama at depth of 15cm, M-20cm : Matsuyama at depth of 20cm

Fig. 4. Score plot of the 189 samples on the first two principal components explaining the variance in the Vis-NIR spectral data.

Comparison of Calibration Models

The PLSR results of the calibration and validation were obtained as shown in Table 2. Based on the determination of coefficient (R_{val}^2) and root mean square error of validation ($RMSE_{val}$), CM3 that combined datasets of soil from both fields are resulted in the highest accuracy for MC and OM with the R_{val}^2 are 0.90 and 0.95 respectively. For TC and TN, CM1 (dataset from Matsuyama only) produced highest model accuracy with R_{val}^2 and $RMSE_{val}$ are 0.88 and 1.38 for TC, 0.85 and 0.26 for TN. The lowest model accuracy for all the soil properties is CM2.

According to Chang (2001) Values of RPD larger than 2 are considered excellent, between 1.4 and 2 are good and below 1.4 are unreliable. Referring to this classification, CM1 and CM3 showed excellent calibration models for all the soil properties. CM2 was only regarded as excellent calibration model for TN and good for other soil properties. The scatter plots of the CM3 models are depicted in Fig. 5.

Results from this study show that the combination of the calibration dataset from two fields of different soil management gave a wider range of dataset and resulted in better prediction. The low accuracy of CM2 is expected due to the small variation of Fukushima soil properties. This is consistent with a study by Sudduth (2010) who found that when variation in the parameter of interest was small, generally poor estimations of OM was obtained at the field scale. Furthermore, as reported by Bricklemyer (2011), field moisture content is one of the factors that can reduce the accuracy of Vis-NIR method. This might be another reason for low accuracy of CM2 model as the Fukushima soil is high in moisture content (Table 1). Moreover, Terhoeven-Urselman (2010) noted that the calibrations might have been stronger if soil reference analysis (laboratory soil analysis) had been done in a short period of time and at the same time for all large number of samples. For laboratory soil analysis of Fukushima soil for the reference values, the large samples of Fukushima (130 samples) were not analyzed at the same time in a short period of time. Unlike Matsuyama, the soil analysis was performed for all the 59 samples on the same day and in a short period of time.

Table 2. Summary of Partial Least Square Regression (PLSR)

Soil Properties	Calibration Model[a]	N[b]	Calibration			Validation			SD	RPD
			R_{cal}	R_{cal}^2	$RMSE_{cal}$	R_{val}	R_{val}^2	$RMSE_{val}$		
MC [%]	CM1	53	0.97	0.94	0.97	0.94	0.88	1.40	3.95	2.82
	CM2	117	0.88	0.77	1.38	0.85	0.73	1.52	2.89	1.90
	CM3	170	0.96	0.93	1.30	0.95	0.90	1.55	4.77	3.08
OM [%]	CM1	53	0.93	0.87	0.22	0.91	0.82	0.26	0.61	2.35
	CM2	117	0.91	0.82	0.24	0.87	0.77	0.28	0.58	2.07
	CM3	170	0.98	0.97	0.22	0.97	0.95	0.28	1.28	4.57
TC [%]	CM1	53	0.95	0.91	0.13	0.94	0.88	0.15	0.43	2.87
	CM2	117	0.91	0.83	0.09	0.85	0.72	0.12	0.22	1.83
	CM3	170	0.95	0.90	0.11	0.93	0.87	0.13	0.36	2.77
TN [%]	CM1	53	0.96	0.92	0.01	0.92	0.85	0.01	0.03	3.00
	CM2	117	0.91	0.82	0.01	0.85	0.73	0.01	0.02	2.00
	CM3	170	0.93	0.87	0.01	0.92	0.84	0.01	0.03	3.00

[a]Calibration Model. CM1: Matsuyama model, CM2: Fukushima model, CM3: Fukushima and Matsuyama model, [b]Number of samples used in the model, SD : standard deviation, RPD : residual prediction deviation

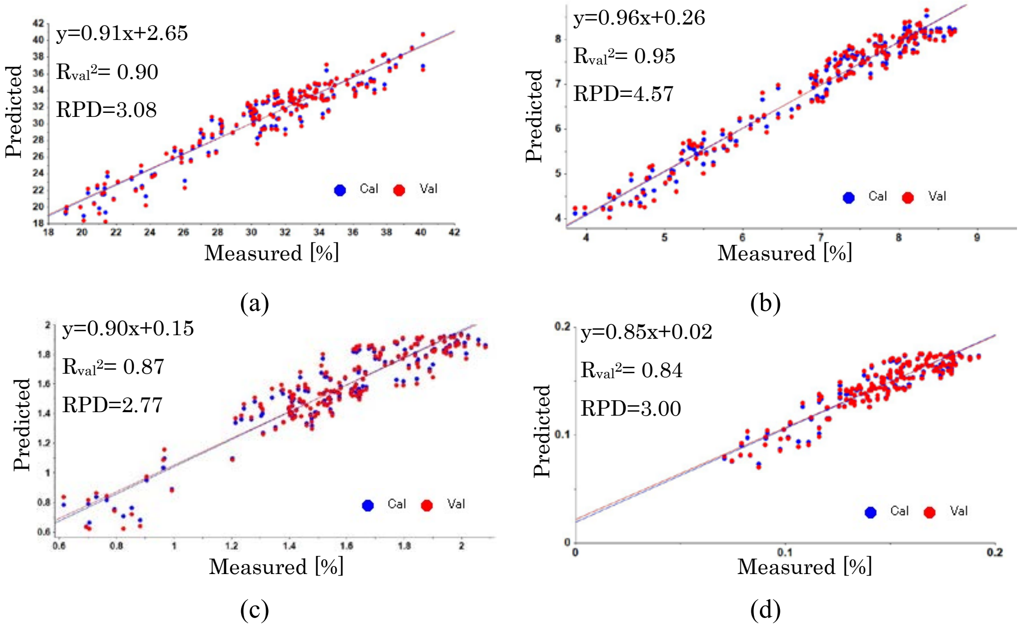


Fig. 5. Scatter plot of measured values versus Vis–NIR predicted values of CM3 using partial least squares regression (PLSR) coupled with full cross-validation datasets for: (a) MC, (b) OM, (c) TC and (d) TN

Soil Properties Mapping

The RTSS also acquired other spectra in between the sampling points on transects of Matsuyama and Fukushima fields. For Matsuyama field, the number of spectra acquired was 145 at depth of 0.10 m, 269 at depth of 0.15 m and 377 spectra at depth of 0.20 m. For Fukushima field, 723 spectra were acquired at Field 2 and 1715 at Field 5. All these unknown spectra were predicted using calibration model CM3 to determine the amount of the soil properties. The predicted values were then used to generate soil map using ArcGIS 10 (ESRI, USA) mapping software. The inverse distance weighting (IDW) method was used for the spatial interpolation. In order to allow for useful comparisons between measured and predicted maps, the same number of classes for both maps was used for every soil properties. Fig. 6 compares between maps of laboratory measured and predicted of MC and OM of Fukushima Field (Field 2 and Field 5), taken as an example. A comparison between maps of measured and predicted MC and OM shows similarity, which was also achieved for TC and TN (maps are not shown).

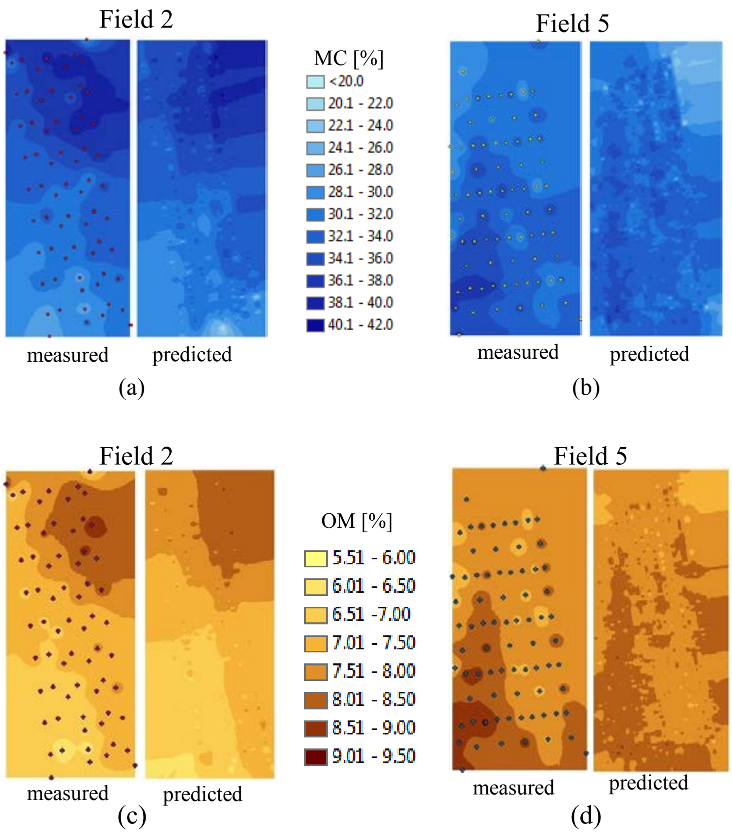


Fig. 6. Comparison of measured and predicted map of MC and OM for Field 2 (a and c) and Field 5 (b and d)

CONCLUSIONS

Three spectroscopic calibration models based on Vis-NIR underground reflectance spectra of two different fields (organic and inorganic farm) have been developed. The third calibration model (CM3-general model) that incorporated soil from both fields improved the accuracy of CM2 model for all soil properties and CM1 model for MC and OM. The generalization of the model has good potential for minimizing the repetitiveness of developing calibration model every time for every different field. However, the calibration sample selection methods need to be optimized which covering as much of the soil variation as possible within all samples and within the calibration samples. Moreover, validation using independent samples is required in additional study. Thus, incorporation of more soil samples from various types of cultivation fields at other region of Japan is necessary in future studies as to improve the generality and robustness of the calibration model that could be applied to all Japan soils whatever the agricultural management background. Even though samples used for model validation were not independent of those used for calibration, the result from this study nevertheless could be used as a step towards establishment of a robust general calibration model for various type of agriculture soil.

REFERENCES

Bricklemyer, R. S. and Brown, D. J. 2010. On-the-go VisNIR: Potential and limitations for mapping soil clay and organic. Computers and Electronics in Agriculture 70 : p. 209-216

Kodaira, M. and S. Shibusawa. 2013. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. Geoderma, 199: p. 64-79

Malley, D. F., Martin, P. D. and Ben-Dor, E. 2004. Application in analysis of soils. p. 729–784. In C.A. Roberts et al. (ed.) Near infrared spectroscopy in agriculture. Agron. Monogr. 44. ASA, CSSA, and SSSA, Madison, WI.

Mouazen, A. M., De Baerdemaeker, J., Ramon, H., 2005. Towards development of on-line soil moisture content sensor using a fibretype NIR spectrophotometer. Soil Till. Res. 80 (1–2): p. 171–183

Stenberg, B., Viscarra Rossel, R. A. Mouazen, A. M. and Wetterlind, J. 2010. Visible and near infrared spectroscopy in soil science. Advances in Agronomy. 107: p.163–215. doi:10.1016/S0065-2113(10)07005-7

Sudduth, K .A., N. R. Kitchen, E. J. Sadler, S. T. Drummond, and D. B. Myers. 2010. VNIR spectroscopy estimates of within-field variability in soil properties. p. 153–163. In R.A. Viscarra Rossel et al. (ed.) Proximal soil sensing. Springer Science + Business Media, Dordrecht, the Netherlands.

Terhoeven-Urselman, T., Vagen, T., Spaargaren, O. and Shepherd, K. D. 2010. Prediction of Soil Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library. Soil Science Society of America. 74. (5): p. 1792 - 1799

Viscarra Rossel, R. A., D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131: p. 59–75. doi:10.1016/j.geoderma.2005.03.007