



Monitoring organic carbon, total nitrogen, and pH for reclaimed soils using field reflectance spectroscopy

Authors: Sorenson, P.T., Small, C., Tappert, M.C., Quideau, S.A., Drozdowski, B., et al.

Source: Canadian Journal of Soil Science, 97(2) : 241-248

Published By: Canadian Science Publishing

URL: <https://doi.org/10.1139/cjss-2016-0116>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Monitoring organic carbon, total nitrogen, and pH for reclaimed soils using field reflectance spectroscopy

P.T. Sorenson, C. Small, M.C. Tappert, S.A. Quideau, B. Drozdowski, A. Underwood, and A. Janz

Updated online 9 August 2017: The license for this article has been changed to the CC BY 4.0 license. The PDF and HTML versions of the article have been modified accordingly.

Abstract: Assessing the success of soil reclamation programs can be costly and time-consuming due to the cost of traditional soil analytical techniques. One cost-effective tool that has been successfully used to efficiently analyze a range of soil parameters is reflectance spectroscopy. We used reflectance data to analyze natural and reclaimed soils in the field, examining three key soil parameters: soil organic carbon (SOC), total nitrogen (TN), and soil pH. Continuous wavelet transforms combined with machine learning models were used to predict these parameters. Based on the root mean square error (RMSE), R^2 value, and the ratio of performance to deviation (RPD), the Cubist model produced the most accurate models for SOC, TN, and pH. The RMSE, R^2 , and RPD values for SOC were 0.60%, 0.80, and 2.2, respectively. The TN model results were 0.05%, 0.81 and 2.5, and pH model results were 0.44, 0.69 and 1.8. Overall, the optimal model can be used to predict SOC and TN accurately, and improvements in the pH model are needed as pH values less than 6.5 were consistently overpredicted.

Key words: reflectance spectroscopy, reclamation, carbon, nitrogen.

Résumé : Le coût des techniques classiques d'analyse du sol rend parfois l'efficacité des programmes de restauration du sol aussi onéreuse que laborieuse à évaluer. La spectroscopie par réflectance est une technique rentable dont on s'est servi pour analyser efficacement un éventail de paramètres du sol. Les auteurs ont recouru à la réflectance pour analyser des sols naturels et restaurés sur le terrain, notamment les trois grands paramètres que sont la teneur en carbone organique, la concentration d'azote total et le pH. Ils ont prévu ces paramètres en combinant les transformées d'ondelettes continues et des modèles d'apprentissage machine. D'après l'écart-type, le coefficient R^2 et le ratio performance/écart (RPD), le modèle cubiste est celui qui donne les résultats les plus précis pour le carbone organique du sol, la concentration totale d'azote et le pH. L'écart-type, le coefficient R^2 et le RPD du carbone organique du sol s'établissent respectivement à 0,60 %, 0,80 et 2,2. Les résultats du modèle de l'azote total se chiffrent à 0,05 %, 0,81 et 2,5, tandis que ceux du modèle du pH sont de 0,44, 0,69 et 1,8. En règle générale, on peut se servir du modèle optimal pour prévoir avec précision la concentration de carbone organique et celle d'azote total dans le sol; il faudrait améliorer le modèle du pH, car il y a toujours surestimation des prévisions quand le pH est inférieur à 6,5. [Traduit par la Rédaction]

Mots-clés : spectroscopie par réflectance, restauration, carbone, azote.

Introduction

Conventional methodologies for assessing soils before and after disturbance are time-consuming and costly, requiring large volumes of samples to be sent to a commercial laboratory for analytical measurement. There is a need for technologies that can provide accurate data in a more timely and cost-effective manner. In Alberta,

the energy industry alone is responsible for soil disturbance associated with 470 000 wellsites, 39 000 other oil and gas facilities, and 24 open pit coal mines (Alberta Energy Regulator 2016). In addition to the facility footprint, there is additional soil disturbance due to borrow pits, transmission lines, pipelines, and access roads that are associated with these facilities.

Received 12 September 2016. Accepted 3 January 2017.

P.T. Sorenson and S.A. Quideau. Department of Renewable Resources, University of Alberta, 442 Earth Science Building, Edmonton, AB T6G 2E3, Canada.

C. Small, B. Drozdowski, and A. Underwood. Alberta Innovates — Technology Futures, 250 Karl Clark Road, Edmonton, AB T6N 1E4, Canada.

M.C. Tappert. TNT Mineral Science, Box 851, Gibsons, BC V0N 1V0, Canada.

A. Janz. Alberta Environment and Parks, 10th Floor, 9888 Jasper Avenue, Edmonton, AB T5J 5C6, Canada.

Corresponding author: Preston T. Sorenson (email: preston.sorenson@ualberta.ca).

Copyright remains with the author(s) or their institution(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Evaluating the success of soil reconstruction and reclamation requires measuring key soil parameters. Three soil parameters of particular relevance are soil organic carbon (SOC), total nitrogen (TN), and soil pH that affect a number of key functions (i.e., nutrient cycling, plant productivity, and soil structure development) important for the development of sites undergoing reclamation (Reeves 1997; Blake and Goulding 2002; Dessureault-Rompré et al. 2015). Research conducted on reclaimed mine topsoil in the Athabasca Oil Sands Region has shown that SOC and TN can change in response to disturbance and reclamation. SOC and TN have been shown to increase with time, following reclamation due to development of the forest floor (Sorenson et al. 2011). Similar findings were observed in agricultural mine soils, reclaimed forests, pasture, and hay land uses which tend to have lower SOC content relative to undisturbed soils (Indorante et al. 1981; Shrestha and Lal 2007). Soil pH can also be affected by soil disturbance. For example, in the Canadian prairies, soils underlain by carbonate-rich parent material, which are admixed with overlying topsoil during a disturbance, can lead to an elevated soil pH (Anderson and Cerkowniak 2010). Furthermore, in locations such as Alberta, where sulfur dust is a common by-product of the desulfurization of oil and natural gas, low pH can also result from the addition of elemental sulfur to soil or as a result of industrial emissions of sulfur dioxide (Alberta Energy Regulator 2011).

Unfortunately, obtaining laboratory data for SOC, TN, and pH can be costly and time-consuming, which can serve as a barrier to reclamation and monitoring programs. More affordable options include using field systems to collect reflectance spectroscopy data in a rapid and nondestructive manner. Reflectance spectroscopy has been used to predict SOC, TN, and pH in a number of studies (Ben-Dor and Banin 1995; Chang et al. 2001; McBratney et al. 2006; Bartholomeus et al. 2008; Gomez et al. 2008b; Rossel and Behrens 2010; Ge et al. 2014). SOC has been measured with error rates as low as 1.5 g kg⁻¹ on dried intact soil cores and on samples ranging from 0.1% to 10.4% TOC (McCarty et al. 2002; Doetterl et al. 2013). While work conducted in the Canadian provinces of Manitoba and Ontario has successfully shown that reflectance spectroscopy can be used to predict SOC and TN (Martin et al. 2002; Xie et al. 2011), there is a lack of information on the use of reflectance spectroscopy on Canadian soils and more specifically, Alberta soils.

The processing of soil spectral data has focused on using partial least squares regressions (e.g., Gomez et al. 2008a, 2008b; Xie et al. 2011; Kinoshita et al. 2012) and using signal processing techniques such as Savitzky–Golay smoothing, derivatives, and multiplicative scatter correction (e.g., He et al. 2007; Minasny et al. 2009; Doetterl et al. 2013; Nawar et al. 2016). Recently, two data processing techniques have been

applied to spectral data and have improved the accuracy of model predictions. Machine learning tools have been successfully applied to spectral data and generated more accurate predictions than partial least squares regression models. Specifically, multivariate adaptive regression splines, support vector machines, and Cubist models have tended to produce the most accurate predictions (Rossel and Behrens 2010; Doetterl et al. 2013; Nawar et al. 2016). Additionally, the application of an alternative signal processing method, the wavelet transform, has improved the use of spectral soil data to accurately predict different soil properties (Viscarra Rossel and Lark 2009; Rossel and Behrens 2010). Wavelet analyses can be discrete or continuous, while the continuous wavelet transform (CWT) can be highly redundant, it is directly comparable to the original spectra and therefore manual inspection of the wavelet results is easier (Rivard et al. 2008). Detailed information about the application of the CWT to spectral data is available in Rivard et al. (2008).

Overall, the main objective of this research was to determine if reflectance spectroscopy data collected in the field with a drill-rig mounted spectroscopy system could be used to measure SOC, TN, and pH. Specifically, this study focused on using machine learning models, which have been applied to soils in other regions (Rossel and Behrens 2010; Doetterl et al. 2013), along with wavelet transforms, and determining if they can be successfully applied to reflectance spectroscopy data from Canadian soils.

Materials and Methods

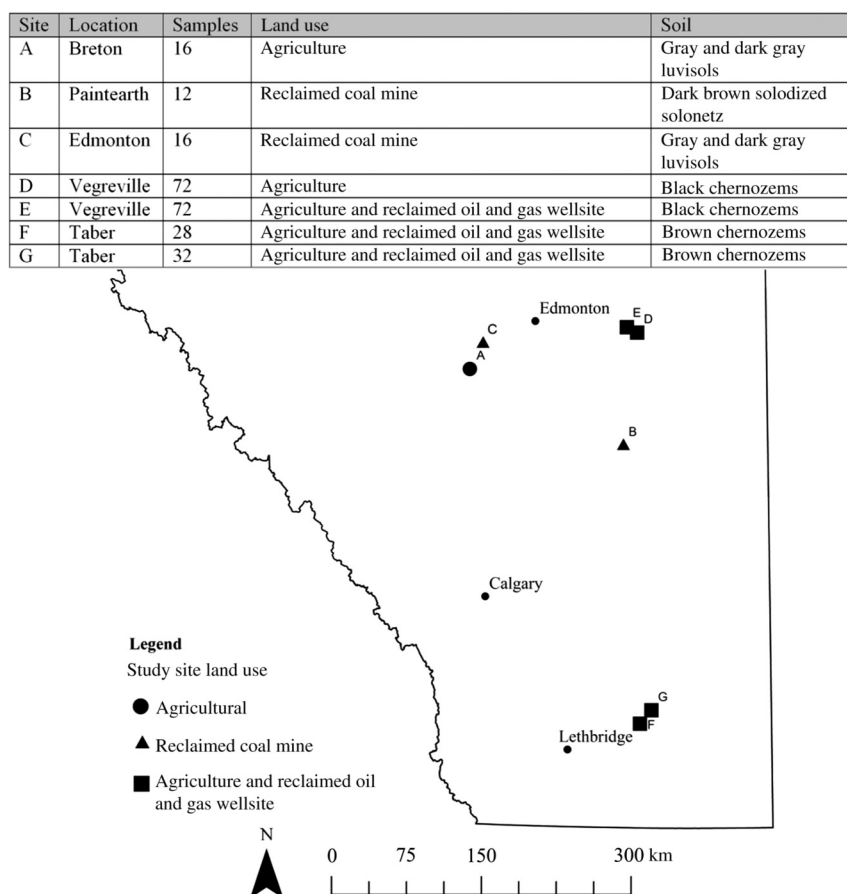
Studied soils

Sampling locations were selected to include a range of soil types, with variations in SOC, TN, and pH. Samples were collected from natural reference locations and from two common land disturbances in Alberta. Soil samples were collected from agricultural soils, reclaimed oil and gas wellsites, and reclaimed coal mines at seven sites across central and southern Alberta, Canada; 144 were collected from agricultural soil and 104 were collected from reclaimed soils (Fig. 1). Agricultural soil samples were collected at a set of permanent agricultural plots established in 1929 that investigate farming practices near Breton, AB; at two separate locations near Vegreville, AB; and at two locations near Taber, AB. Reclaimed soils were collected from two reclaimed coal mines, one near Paintearth, AB and the other near Edmonton, AB. The remaining soil samples were collected from reclaimed oil and gas wellsites located near Vegreville and Taber, AB.

Spectral measurements

Spectral measurements were collected in-situ using a P4000 drill rig mounted spectrometer (Veris® Technologies, Salina, KS, USA). The P4000 contact

Fig. 1. Location of study sites sampled using the Veris® spectrometer P4000 probe in Alberta, Canada. Source: North American State Province Boundaries from Esri, TomTom.



probe has a sapphire window and an integrated light source. Detectors for the P4000 include a 3648 element Toshiba TCD1304AP Linear CCD Array and a 256 element InGaAs Linear Image Sensor G9206-02. The P4000 has a response range of 350–2200 nm, 384 wavelength bands, and a spectral resolution of 8 nm. Reflectance values were calculated by adjusting the soil radiance based on the radiance measurement of a black and a white reference panel provided for regular calibration of the P4000. Reference readings were taken when initializing the equipment and then every 10 min using the integrated light source. Spectral measurements were collected every 0.15 m continuously throughout each borehole. Each borehole was completed to 1 m below ground surface. Samples for the training datasets were collected from discrete 15 or 30 cm increments from the borehole at 3 or 4 sample locations. Samples were collected from A, B, and C horizons in the profile. If samples were collected from a 30 cm interval, the two spectra were averaged for that training data point. If spectra collected at a borehole did not have a corresponding laboratory data point, then it was not included in the analysis.

Laboratory analyses

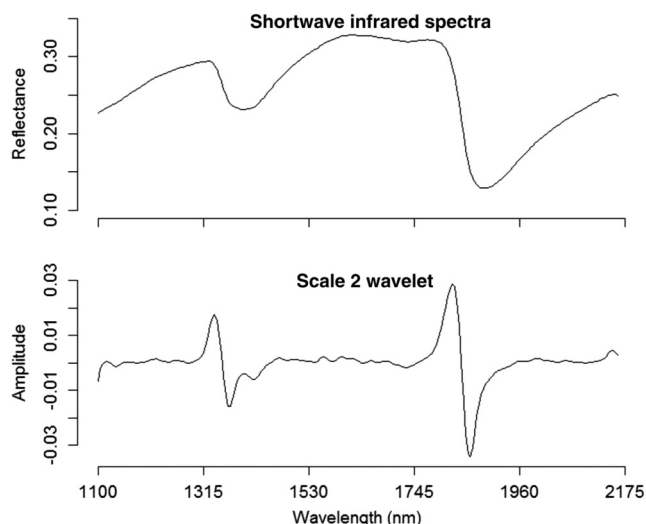
Laboratory analyses were run on samples collected from the boreholes to generate a training dataset. Soil samples were oven dried and then ground with a ball mill prior to laboratory analysis. Total carbon (TC) and TN were determined by dry combustion according to the methods described in Nelson and Sommers (1996) and Bremner (1996). Samples were analyzed for TC and nitrogen on an Elementar Vario Max CNS Analyzer (Elementar Americas Inc., Mt. Laurel, NJ, USA). Soil samples were analyzed for pH according to the saturated paste method (McLean 1982). In total, 248 samples were analyzed for TC, 164 samples were run for TN, and 232 samples were analyzed for pH.

Statistical analysis

Data processing

The analysis focused on two spectral ranges: (1) 700 and 950 nm and (2) 1097 and 2183 nm. These spectral ranges were selected as the near infrared (NIR) and short-wave infrared (SWIR) regions have spectral features associated with organic carbon. Data were trimmed at the end of the NIR and the beginning of the SWIR region

Fig. 2. Example shortwave infrared spectral signature and Scale 2 Wavelet for an Ah horizon sample collected near Vegreville, Alberta. Scale 2 Wavelet is a second order Gaussian Wavelet Transform. The soil sample contains 4.14% soil organic carbon, and 0.36% total nitrogen and has a pH of 6.6.



due to a low signal to noise ratio associated with detector switch between the NIR and SWIR regions. The end of the SWIR region was also removed due to noise associated with being at the edge of the detector range. Each band in both of these wavelength ranges was used in the model development. Each spectrum was processed using CWT (e.g., Rivard et al. 2008; Tappert et al. 2015; Scafutto et al. 2016) using the *wmtsa* package in R (Percival et al. 2016). The CWT outputs were calculated for each spectral range using an eight-scale second-order Gaussian transform. Between 700 and 950 nm, Scale 4 was deemed to best capture the spectral absorption features relating to composition and used to develop the statistical model, and between 1097 and 2183 nm, Scale 2 was selected (Fig. 2).

Model development

The spectral data were calibrated against SOC, TN, and soil pH calibration datasets. All model development was conducted using the *caret* package (Kuhn et al. 2016) in R (R Core Team 2016), with models previously used to predict soil parameters using spectroscopic data (Minasny and McBratney 2008; Rossel and Behrens 2010; Doetterl et al. 2013). Models included in the analysis are multivariate adaptive regression splines (Milborrow 2016), artificial neural networks (Venables and Ripley 2002), radial basis support vector machines (Karatzoglou et al. 2004), partial least squares regression (Mevik et al. 2015), random forests (Liaw and Wiener 2002), and Cubist models (Kuhn et al. 2014). All models were initially run with the default values defined by the *caret* package, with the final model

parameters automatically selected by the *caret* package during model training (Kuhn et al. 2016). Model parameters are automatically selected by the *caret* package during a leave-one-out cross-validation, with the model parameters producing the smallest root mean square error (RMSE) of cross-validation being selected.

Model evaluation

Model accuracy was evaluated based on the RMSE, R^2 , and ratio of performance to deviation (RPD). The RPD value is the ratio of the standard deviation to the RMSE. Generally, RPD values greater than 2 indicate that a model can be accurately used for prediction, values between 1.4 and 2 are satisfactory, but improvement would be valuable, and values less than 1.4 indicate the model has no prediction capability (Chang et al. 2001). The best model calibration was determined based on the model that minimized the RMSE value. Models were created using the combined reclaimed and natural soil data. The cross-validation results for the natural and reclaimed sites were separated to investigate how well the model performed in both contexts.

Results

Soil organic carbon (SOC)

SOC concentrations ranged from 0.25% to 6.14% with an average concentration of 1.93% in the natural background soils (Table 1). Soil organic carbon concentrations in the reclaimed soils ranged from 0.43% to 4.40% and had an average concentration of 1.63% (Table 1). The Cubist model produced the lowest RMSE value (0.60%) (Fig. 3a), followed by the random forest model (0.62%), multivariate adaptive regression splines (0.66%), support vector machines (0.67%), partial least squares regression (0.90%), and finally the artificial neural nets (1.56%) model (Table 2). The RPD values for the Cubist (2.2), random forest (2.1), multivariate adaptive regression splines (2.0), and support vector machines (2.0) models indicate that these models can accurately be used for prediction. The partial least squares regression (1.5) produced satisfactory RPD values, whereas the artificial neural nets (0.9) model produced results that cannot be used for accurate prediction of SOC.

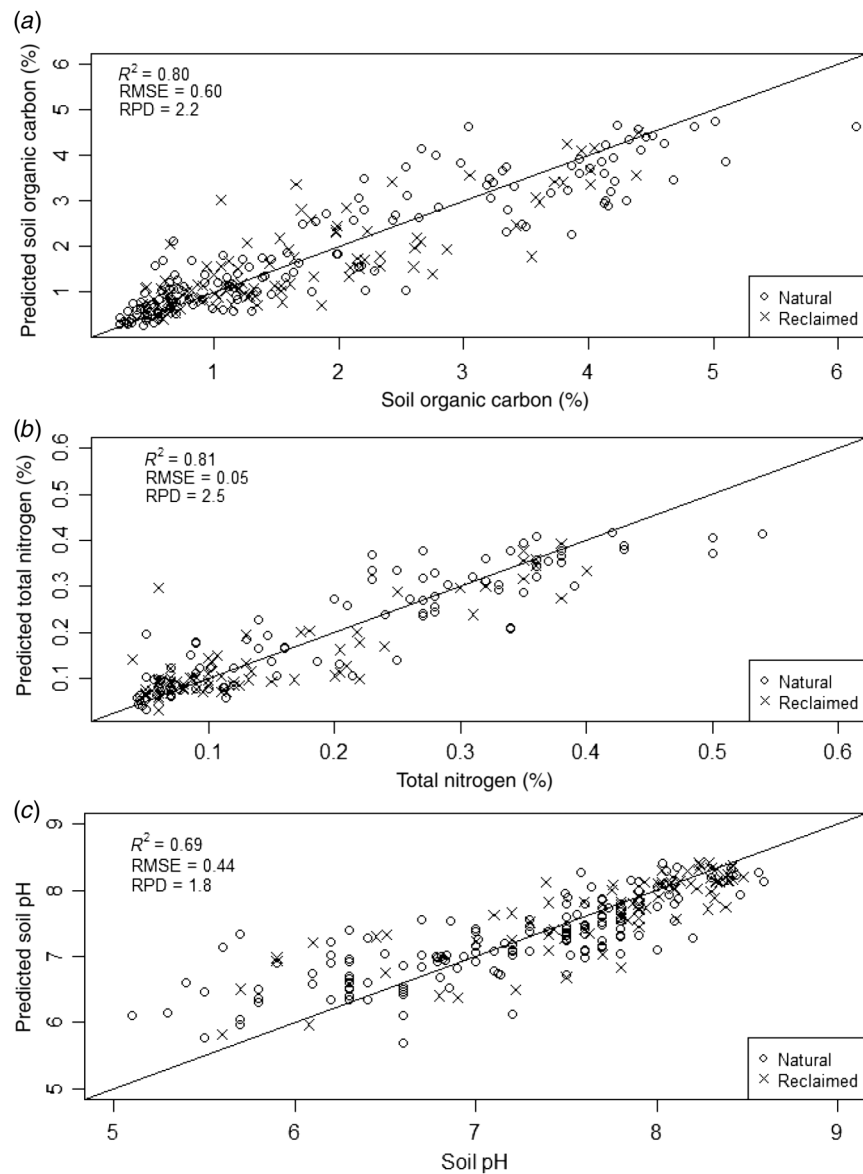
Soil nitrogen

Soil nitrogen concentrations ranged from 0.04% to 0.54% in the natural background soils and from 0.43% to 0.40% in the reclaimed soils (Table 1). Natural background soils on average contained 0.21% TN and the reclaimed soils had an average TN content of 0.15% (Table 3). The Cubist model produced the lowest RMSE value (0.05%) (Fig. 3b), followed by multivariate adaptive regression splines (0.06%), random forest (0.06%), support vector machines (0.06%), partial least squares regression (0.07%), and artificial neural nets model (0.12%). The RPD values for the Cubist (2.5), multivariate adaptive

Table 1. Soil organic carbon (SOC), total nitrogen (TN), and soil pH statistics for natural and reclaimed soils analyzed in this study.

Property	Natural			Reclaimed		
	Mean	Range	SD	Mean	Range	SD
SOC (%)	1.93	0.25–6.14	1.49	1.63	0.43–4.40	1.06
TN (%)	0.21	0.04–0.54	0.13	0.15	0.04–0.40	0.10
pH	7.1	5.1–8.6	0.80	7.7	5.6–8.5	0.68

Note: A total of 248 samples were analyzed for SOC, 164 samples for TN, and 232 samples for soil pH.

Fig. 3. Cross-validated predicted versus observed (a) soil organic carbon (%), (b) total nitrogen (%), and (c) soil pH content for natural and reclaimed soils as produced by the Cubist model. The solid line indicates the 1:1 line to illustrate deviations between predicted and measured data.

regression splines (2.1), random forest (2.1), and support vector machines (2.1) models indicate that they can accurately be used for prediction. The partial least squares

regression (1.8) model was found to satisfactory based on the RPD value. However, the artificial neural nets (1.0) model did not accurately predict TN.

Table 2. Cross-validation results for soil organic carbon prediction using reflectance spectroscopy data.

Model	Cross-validation results		
	RMSE	R ²	RPD
Multivariate adaptive regression splines	0.66	0.76	2.0
Artificial neural nets	1.56	0.01	0.9
Support vector machines	0.67	0.75	2.0
Partial least squares regression	0.90	0.54	1.5
Random forest	0.62	0.78	2.1
Cubist	0.60	0.80	2.2

Note: Model results are evaluated based on the root mean square error (RMSE), R², and the ratio of performance to deviation (RPD).

Table 3. Cross-validation results for total nitrogen prediction using reflectance spectroscopy data.

Model	Cross-validation results		
	RMSE	R ²	RPD
Multivariate adaptive regression splines	0.06	0.75	2.1
Artificial neural nets	0.12	0.20	1.0
Support vector machines	0.06	0.78	2.1
Partial least squares regression	0.07	0.67	1.8
Random forest	0.06	0.78	2.1
Cubist	0.05	0.81	2.5

Note: Model results are evaluated based on the root mean square error (RMSE), R², and the ratio of performance to deviation (RPD).

Soil pH

Soil pH in the natural background soils ranged from 5.1 to 8.6, with an average pH of 7.1 (Table 1). The pH in the reclaimed soils ranged from 5.6 to 8.5 with an average pH of 7.7. The Cubist model produced the lowest RMSE value (0.44) (Fig. 3c), followed by random forest (0.47) and support vector machines model (0.51) (Table 4). Multivariate adaptive regression splines were next (0.54), followed by partial least squares regression (0.68) and artificial neural networks (6.39). No model had an RPD value above 2.0. The cubist (1.8), random forest (1.7), support vector machines (1.6), and multivariate adaptive regression splines models (1.5) had RPD values greater than 1.4 indicating that they could satisfactorily be used for prediction. The partial least squares regression (1.2) and artificial neural nets (0.1) models could not be successfully used for prediction of soil pH.

Prediction accuracy for natural and reclaimed soils

The accuracy of the model predictions varied slightly between the natural background and reclaimed soils for SOC, TN, and pH values (Table 5). While the RMSE value for natural soil carbon (0.60) was slightly higher

Table 4. Cross-validation results for soil pH prediction using reflectance spectroscopy data.

Model	Cross-validation results		
	RMSE	R ²	RPD
Multivariate adaptive regression splines	0.54	0.58	1.5
Artificial neural nets	6.39	0.01	0.1
Support vector machines	0.51	0.60	1.6
Partial least squares regression	0.68	0.31	1.2
Random forest	0.47	0.67	1.7
Cubist	0.44	0.69	1.8

Note: Model results are evaluated based on the root mean square error (RMSE), R², and the ratio of performance to deviation (RPD).

Table 5. Cross-validation results from the Cubist model for samples collected from natural and reclaimed soils.

Parameter	Cubist model cross-validation results		
	RMSE	R ²	RPD
Natural soil carbon	0.60	0.84	2.5
Reclaimed soil carbon	0.59	0.70	1.8
Natural soil nitrogen	0.06	0.82	2.3
Reclaimed soil nitrogen	0.05	0.76	2.0
Natural soil pH	0.48	0.62	1.6
Reclaimed soil pH	0.39	0.68	1.8

Note: Model results are evaluated based on the root mean square error (RMSE), R², and the ratio of performance to deviation (RPD).

than the RMSE value for reclaimed soil carbon (0.59), the predicative capability for carbon on natural soils was slightly higher when compared with reclaimed soils. The same pattern was also observed for nitrogen. Soil pH model results for reclaimed soils had a lower RMSE value and a higher RPD compared with the natural soils.

Discussion

Overall, our results indicate that there is potential for using field spectroscopy to monitor reclaimed soil properties. Specifically, our models could be used to predict SOC and TN concentrations from field collected spectra based on the criteria outlined in Chang et al. (2001). For soil pH, the model is adequate for prediction at pH values above 6.5, but improvements are required to accurately predict lower pH values as there was consistent over prediction of pH at values less than 6.5 (Fig. 3c). The pH model may be improved by using a spectrometer that collects spectral data to 2500 nm, as key carbonate features are found near 2336 nm (Ben-Dor and Banin 1990).

For our study, model results were comparable for natural and reclaimed soils for TN and soil pH (Table 5). The reclaimed SOC samples had a lower RPD value than

the natural samples, and the success of the model on reclaimed soil samples could use improvement. Modelling results can likely be improved by obtaining more data from reclaimed sites at SOC concentrations above 4%, as the reclaimed samples had less data at these concentrations.

The results of the Cubist models for predicting SOC and TN under field conditions were satisfactory and compared favourably with results from studies using dried and homogenized soil samples in the laboratory (Chang et al. 2001; Martin et al. 2002; Xie et al. 2011). Our results compare favourably with a study in Australia that used a portable spectrometer to measure SOC contents in the field, as we were able to obtain a RPD value of 2.2 compared with 1.92 in the study in Australia (Gomez et al. 2008b). An important point about predictive modelling of reflectance spectroscopy data is that a model developed with more homogenous training datasets tend to be more accurate with lower RMSE values, but also less transferable to new sites. Therefore, it is important to consider the RPD value when assessing results and not just the RMSE. Overall, this study using a drilling rig mounted spectrometer probe had comparable results with other laboratory and field spectroscopy studies.

Additional work is needed to investigate methods that improve modelling results using field collected spectra. Variation in soil moisture in the field can have a substantial effect on overall reflectance. Other researchers have successfully used external parameter orthogonalization (EPO) to correct spectra based on soil moisture variation (Minasny et al. 2009). Initially developed to remove the effects of temperature when analyzing fruit for sugar contents with reflectance spectroscopy (Roger et al. 2003), EPO corrects the spectra based on spectral areas affected by soil moisture and then projects the spectra orthogonal to the soil moisture variation (Minasny et al. 2009). Further work is needed to compare results using low-scale wavelets and EPO corrections to remove the effects of soil moisture from reflectance spectra. Modelling results may also be improved using a detector that allows data collection up to 2500 nm, as SOC has important spectral features between 2200 and 2500 nm (Rossel and Behrens 2010).

Improving the system to simultaneously measure carbonates, salinity and bulk density would increase the benefit of these tools through the dual measurement of soil chemical and physical parameters in the field. The measurement of carbonates using reflectance spectroscopy is well established (Ben-Dor and Banin 1990; Lagacherie et al. 2008; Gomez et al. 2012), and does not represent a significant technical challenge. This parameter could help to inform reclamation practitioners of admixing, since the presence of carbonate-rich material from parent materials is common in reconstructed soils in the Canadian Prairies (Anderson and Cerkowski 2010). The Veris® P4000 Spectrophotometer Probe has dipole EC contacts and a load cell force sensor integrated into the probe, suitable for measuring soil bulk density.

Determining how accurately dipole EC contacts and load cell force sensors can also be used to measure soil salinity and bulk density in situ is necessary before deploying this system for reclamation monitoring.

Acknowledgements

We thank Alberta Environment and Parks [formally the Alberta Environmental Monitoring and Evaluation Agency (AEMERA)] and Alberta Innovates-Technology Futures for providing financial assistance, soil samples, and soil proximal soil data for the project. We also acknowledge the National Science and Engineering Council (NSERC) for providing financial support for this project with a Discovery Grant to S.A. Quideau (no. 462352-2014).

References

- Alberta Energy Regulator. 2011. Directive for monitoring the impact of sulphur dust on soils. Edmonton, AB. [Online]. Available: <http://aep.alberta.ca/lands-forests/land-industrial/inspections-and-compliance/directive-for-monitoring-the-impact-of-sulphur-dust-on-soils/default.aspx>.
- Alberta Energy Regulator. 2016. Statistical reports. [Online]. Available: <https://www.aer.ca/data-and-publications/statistical-reports>.
- Anderson, D., and Cerkowski, D. 2010. Soil formation in the Canadian prairie region. *Prairie Soils Crop*. 3: 57–64.
- Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W.B., and Spaargaren, O.S.P. 2008. Spectral reflectance based indices for soil organic carbon quantification. *Geoderma*, 145: 28–36. doi:10.1016/j.geoderma.2008.01.010.
- Ben-Dor, E., and Banin, A. 1990. Near-infrared reflectance analysis of carbonate concentration in soils. *Appl. Spectrosc.* 44: 1064–1069. doi:10.1366/0003702904086821.
- Ben-Dor, E., and Banin, A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci. Soc. Am. J.* 59: 364. doi:10.2136/sssaj1995.03615995005900020014x.
- Blake, L., and Goulding, K.W.T. 2002. Effects of atmospheric deposition, soil pH and acidification on heavy metal contents in soils and vegetation of semi-natural ecosystems at Rothamsted Experimental Station, UK. *Plant Soil*, 240: 235–251. doi:10.1023/A:1015731530498.
- Bremner, J.M. 1996. Nitrogen — total. Pages 1085–1123 in D.L. Sparks, A. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnston, and M.E. Summer, eds. *Methods of soil analysis part 3 — chemical methods*. American Society of Agronomy, Madison, WI.
- Chang, C.-W., Laird, D.A., Mausbach, M.J., and Hurburgh, C.R. 2001. Near-infrared reflectance spectroscopy — principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65: 480–490. doi:10.2136/sssaj2001.652480x.
- Dessureault-Rompré, J., Zebarth, B.J., Burton, D.L., and Georgallas, A. 2015. Predicting soil nitrogen supply from soil properties. *Can. J. Soil Sci.* 95: 63–75. doi:10.4141/cjss-2014-057.
- Doetterl, S., Stevens, A., Van Oost, K., and van Wesemael, B. 2013. Soil organic carbon assessment at high vertical resolution using closed-tube sampling and vis-NIR spectroscopy. *Soil Sci. Soc. Am. J.* 77: 1430. doi:10.2136/sssaj2012.0410n.
- Ge, Y., Morgan, C.L.S., and Ackerson, J.P. 2014. Vis-NIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma*, 221–222: 61–69. doi:10.1016/j.geoderma.2014.01.011.

- Gomez, C., Lagacherie, P., and Coulouma, G. 2008a. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, **148**: 141–148. doi:10.1016/j.geoderma.2008.09.016.
- Gomez, C., Lagacherie, P., and Coulouma, G. 2012. Regional predictions of eight common soil properties and their spatial structures from hyperspectral vis-NIR data. *Geoderma*, **189–190**: 176–185. doi:10.1016/j.geoderma.2012.05.023.
- Gomez, C., Viscarra Rossel, R.A., and McBratney, A.B. 2008b. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: an Australian case study. *Geoderma*, **146**: 403–411. doi:10.1016/j.geoderma.2008.06.011.
- He, Y., Huang, M., García, A., Hernández, A., and Song, H. 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. *Comput. Electron. Agric.* **58**: 144–153. doi:10.1016/j.compag.2007.03.011.
- Indorante, S.J., Jansen, I.J., and Boast, C.W. 1981. Surface mining and reclamation: Initial changes in soil character. *J. Soil Water Conserv.* **36**: 347–351.
- Karatzoglou, A., Smola, A., Hornik, K., and Seileis, A. 2004. kernlab — an S4 Package for Kernel methods in R. *J. Stat. Softw.* **11**: 1–20. [Online]. Available: <http://www.jstatsoft.org/v11/i09/>.
- Kinoshita, R., Moebius-Clune, B.N., van Es, H.M., Hively, W.D., and Bilgili, A.V. 2012. Strategies for soil quality assessment using visible and near-infrared reflectance spectroscopy in a Western Kenya chronosequence. *Soil Sci. Soc. Am. J.* **76**: 1776. doi:10.2136/sssaj2011.0307.
- Kuhn, M., Weston, S., Keefer, C., and Coulter, N. 2014. Cubist: rule- and instance-based regression modeling. [Online]. Available: <https://cran.r-project.org/package=Cubist>.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., and Candan, C. 2016. caret: classification and regression training. [Online]. Available: <https://cran.r-project.org/package=caret>.
- Lagacherie, P., Baret, F., Feret, J.B., Madeira Netto, J., and Robbez-Masson, J.M. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sens. Environ.* **112**: 825–835. doi:10.1016/j.rse.2007.06.014.
- Liaw, A., and Wiener, M. 2002. Classification and regression by random forest. *R News*, **2**: 18–22.
- Martin, P.D., Malley, D.F., Manning, G., and Fuller, L. 2002. Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. *Can. J. Soil Sci.* **82**: 413–422. doi:10.4141/S01-054.
- McBratney, A.B., Minasny, B., and Viscarra Rossel, R. 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. *Geoderma*, **136**: 272–278. doi:10.1016/j.geoderma.2006.03.051.
- McCarty, G.W., Reeves, J.B., Reeves, V.B., Follett, R.F., and Kimble, J.M. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* **66**: 640–646.
- McLean, E.O. 1982. Soil pH and lime requirement. Pages 199–225 in A. Page, R.H. Miller, and D.R. Keeney, eds. *Methods of soil analysis part 2, chemical and microbiological properties*. Second edition. American Society of Agronomy, Madison, WI.
- Mevik, B.H., Wehrens, R., and Liland, K.H. 2015. pls: partial least squares and principal component regression. [Online]. Available: <https://cran.r-project.org/package=pls>.
- Milborrow, S. 2016. earth: multivariate adaptive regression splines. [Online]. Available: <https://cran.r-project.org/package=earth>.
- Minasny, B., and McBratney, A.B. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemom. Intell. Lab. Syst.* **94**: 72–79. doi:10.1016/j.chemolab.2008.06.003.
- Minasny, B., McBratney, A.B., Pichon, L., Sun, W., and Short, M.G. 2009. Evaluating near infrared spectroscopy for field prediction of soil properties. *Aust. J. Soil Res.* **47**: 664–673. doi:10.1071/SR09005.
- Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., and Mouazen, A.M. 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res.* **155**: 510–522. doi:10.1016/j.still.2015.07.021.
- Nelson, D.W., and Sommers, L.E. 1996. Total carbon, organic carbon, and organic matter. Pages 961–1011 in D.L. Sparks, A. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnston, and M.E. Sumner, eds. *Methods of soil analysis, part 3 — chemical methods*. Soil Science Society of America, Madison, WI, USA.
- Percival, D.B., Walden, A.T., William, A., Percival, D., and Constantine, M.W. 2016. Wavelet methods for time series analysis. [Online]. Available: <https://cran.r-project.org/package=wmts>.
- R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: <https://www.r-project.org/>.
- Reeves, D.W. 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil Tillage Res.* **43**: 131–167. doi:10.1016/S0167-1987(97)00038-X.
- Rivard, B., Feng, J., Gallie, A., and Sanchez-Azofeifa, A. 2008. Continuous wavelets for the improved use of spectral libraries and hyperspectral data. *Remote Sens. Environ.* **112**: 2850–2862. doi:10.1016/j.rse.2008.01.016.
- Roger, J.M., Chauchard, F., and Bellon-Maurel, V. 2003. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom. Intell. Lab. Syst.* **66**: 191–204. doi:10.1016/S0169-7439(03)00051-0.
- Rossel, R.A.V., and Behrens, T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, **158**: 46–54. doi:10.1016/j.geoderma.2009.12.025.
- Scafutto, R.D.M., de Souza Filho, C.R., and Rivard, B. 2016. Characterization of mineral substrates impregnated with crude oils using proximal infrared hyperspectral imaging. *Remote Sens. Environ.* **179**: 116–130. doi:10.1016/j.rse.2016.03.033.
- Shrestha, R.K., and Lal, R. 2007. Soil carbon and nitrogen in 28-year-old land uses in reclaimed coal mine soils of Ohio. *J. Environ. Qual.* **36**: 1775–1783. doi:10.2134/jeq2007.0071. PMID:17965380.
- Sorenson, P.T., Quideau, S.A., MacKenzie, M.D., Landhäusser, S.M., and Oh, S.W. 2011. Forest floor development and biochemical properties in reconstructed boreal forest soils. *Appl. Soil Ecol.* **49**: 139–147. doi:10.1016/j.apsoil.2011.06.006.
- Tappert, M.C., Rivard, B., Fulop, A., Rogge, D., Feng, J., Tappert, R., and Stalder, R. 2015. Characterizing kimberlite dilution by crustal rocks at the Snap Lake diamond mine (Northwest Territories, Canada) using SWIR (1.90–2.36 μm) and LWIR (8.1–11.1 μm) hyperspectral imagery collected from drill core. *Econ. Geol.* **110**: 1375–1387. doi:10.2113/econgeo.110.6.1375.
- Venables, W.N., and Ripley, B.D. 2002. *Modern applied statistics*. 4th ed. Springer, New York, NY, USA.
- Viscarra Rossel, R.A., and Lark, R.M. 2009. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *Eur. J. Soil Sci.* **60**: 453–464. doi:10.1111/j.1365-2389.2009.01121.x.
- Xie, H.T., Yang, X.M., Drury, C.F., Yang, J.Y., and Zhang, X.D. 2011. Predicting soil organic carbon and total nitrogen using mid- and near-infrared spectra for Brookston clay loam soil in Southwestern Ontario, Canada. *Can. J. Soil Sci.* **91**: 53–63. doi:10.4141/cjss10029.