

The potential of near infrared laboratory based hyperspectral imaging for the determination of geochemical and carbon properties in Irish soils

Sharon O'Rourke* and Nicholas Holden

UCD Bioresources Research Centre/Biosystems Engineering, University College Dublin, Dublin 4, Ireland.

*Corresponding author: sharon.orourke@ucd.ie

Introduction

The pending European Union Soil Framework Directive has outlined a number of main threats to soil quality, of which loss of organic matter and soil contamination are particularly pertinent for Irish soils.¹ In response to the *Thematic Strategy on Soil Protection* [COM(2006)231], a national soil database (NSDB) has been compiled in Ireland to provide a referenced baseline of soil organic carbon (OC) and geochemistry. Ongoing monitoring is required in national remediation strategies for degraded sites and to detect any further decline in soil OC. Hyperspectral imaging, estimated to be more cost-effective than other spectroscopic techniques for soil OC, due to its capacity to process more samples per day,² was employed in this study to investigate the potential of national scale calibrations for the determination of soil OC and geochemical properties.

Soil OC has been predicted from the regional³ to the continental scale⁴ with similar accuracy as traditional soil analysis. To date however, a subset of elements (As, Cd, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Sb, S and Zn) has only been estimated where concentrations were expected to be elevated, in suburban agricultural soils,⁵ mining soils,⁶ industrial soils⁷ and freshwater sediments,⁸ and accuracy has been variable between studies. Mid-infrared and near infrared (NIR) spectroscopies have previously been investigated for the determination of 46 major and trace elements in a continental scale soil survey of predominantly agricultural soils.⁹ Extreme sample diversity at this scale, where soils sampled were essentially unique in terms of parent material and land-use, resulted in poor calibrations, even for soil OC. For this paper, a higher sample resolution than that used for the continental soil survey was deployed, and the data offer scope for an evaluation of the prediction of soil geochemical and carbon properties on a national scale. Soil OC, which was expected to produce a good prediction, provides a good comparison for the success of geochemical prediction at this scale.

Materials and Methods

National soil database samples ($n = 1310$) were collected at fixed locations on a predetermined grid of 10×10 km squares using a Dutch auger (0–10 cm). Four hundred samples were selected from the NSDB archive in proportion to soil type and land use representation. Soils from the land-use 'bog' were unavailable and samples representing $<1\%$ (<13 samples; lithosols and rendzinas) were excluded. In total, 7 soil types, 4 land use types and 8 parent rock materials were included. Soils were analysed for the following elements: Al, As, Ba, Ca, Cd, Ce, Co, Cr, Cu, Fe, Ga, Ge, Hg, La, Li, Mg, Mn, Mo, Na, Nb, Ni, P, Pb, Rb, S, Sb, Sc, Se, Sn, Sr, Ta, Ti, Th, Tl, U, V, W, Y and Z by autologous conditioned plasma or inductively coupled plasma mass spectrometry. Soil OC was determined by two different methods, the Walkley-Black method¹⁰ (for the survey of South East Ireland) amended with a 1.16 correction factor, and dry combustion.¹¹

Hyperspectral measurements were obtained using 2 pushbroom line-scanning hyperspectral imaging systems (Desktop Spectral Scanner, DV Optics, Padova, Italy) operating in the visible (VIS; 400 to 1000 nm) and near infrared (NIR; 880 to 1720 nm). Approximately 30g of soil (dried and <2 mm) was scanned in an open polypropylene Petri dish (50 mm diameter) at 5 and 7 nm resolution in the VIS and NIR. The hyperspectral image was treated as a 2 dimensional spectral dataset by selecting a 'region of interest' of 178 and 72 pixels² to represent mean spectra in the VIS and NIR. Due to noise, approximately 50 nm was cropped from the ends of each spectrum.

Samples were partitioned into calibration (75%) and validation (25%) sets by random selection whereby every 4th sample was selected for validation and selections were maintained for each soil property and spectral range. Principal component analysis was used for exploratory analysis. Reflectance values were transformed by Log (1/R) and additional preprocessing techniques tested were Savitzky-Golay smoothing, standard normal variate and multiple scatter correction. Partial least square regression models were developed for each soil property in the VIS, NIR and combined VIS-NIR, using full leave-one-out cross validation. The best ranking models were selected based on the highest coefficient of determination (R^2),

Reference paper as:

S. O'Rourke and N. Holden (2012). The potential of near infrared laboratory based hyperspectral imaging for the determination of geochemical and carbon properties in Irish soils, in: *Proceedings of the 15th International Conference on Near Infrared Spectroscopy*, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 290-293.

lowest root mean square error (RMSE) and highest ratio of predicted deviation (RPD). The complete dataset was then stratified by (i) soil organic carbon content (mineral soils had <15% OC; organic soils had >15% OC), (ii) soil type, (iii) land use type and (iv) parent rock material.

Results and Discussion

All 40 geochemical properties were positively skewed except for Ge, and many were leptokurtic (i.e. narrowly distributed), which is consistent with the findings of the NSDB.¹² By the subsampling scheme employed, soils from contaminated sites were not fully captured, in particular Pb, which only represented 4% of the NSDB concentration range. On a national scale no geochemical model produced an RPD classification better than poor (>1.5) based on the calibration dataset of 300 samples. In general, the VIS models yielded lower R^2 and RPD values, while the NIR and combined VIS-NIR model results were very similar. On the principle of parsimony, further analysis was focused only on the NIR spectral region as combining the VIS region added no value. The best models achieved overall were for Al ($R^2 = 0.64$), Ge ($R^2 = 0.70$), S ($R^2 = 0.63$), Sc ($R^2 = 0.55$), Th ($R^2 = 0.66$) and Ti ($R^2 = 0.55$) at the threshold between poor and good model status. The inherent variation in both mineral and organic parent material resulting from mineralogy, texture and organic matter variation was probably too large to capture the weak signal from the geochemical elements at a national scale with the sample numbers used here. In contrast, soil OC achieved a good model calibration (RPD = 2.01, R^2 of 0.75, RMSE of 3.66 g.kg⁻¹), comparable to results achieved for other countries.¹³ However, prediction accuracy declined upon independent validation (Figure 1).

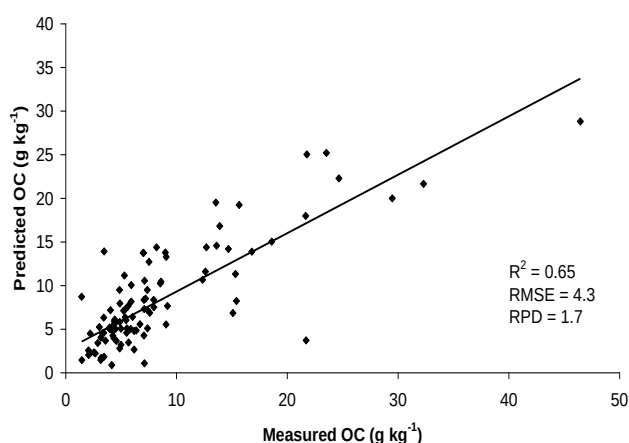


Figure 1. Soil organic carbon validated ($n = 100$) for a national scale model.

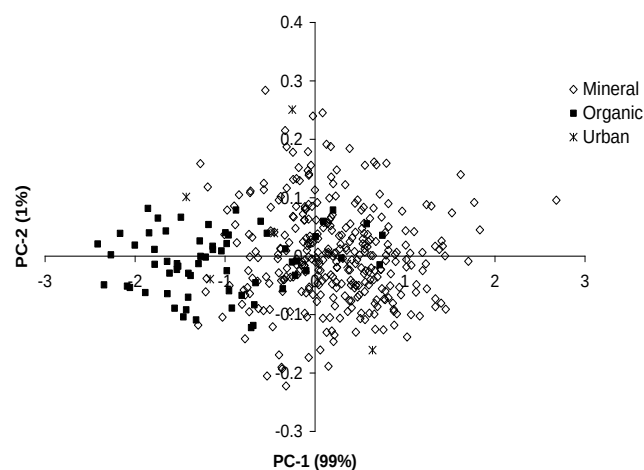


Figure 2. Score plot of principal components for the whole dataset ($n = 400$) showing stratification of soils into mineral, organic and urban soil types.

The stratification of samples based on soil OC threshold was required¹² to obtain meaningful correlation and cluster analysis of geochemical properties in the NSDB. Stratification into mineral ($n = 331$) and organic ($n = 63$) soils (6 urban soils were excluded) did not result in spectrally distinct groups. Classification of mineral and organic soils is usually performed considering the whole soil profile, and top soils with soil OC >15% are possible in mineral soils. Only minor increases in RPD were achieved for mineral soils, but models were not improved enough to produce good predictions. This suggests that the spectral noise arising from the parent material masks the geochemical signal more than the organic matter signal.

Stratifying the dataset by soil, land-use or rock type succeeded in achieving good to excellent predictions. In the case of parent rock material, the stratification broke up the mineral noise into coherent groupings which has had the effect of allowing the geochemical signal to emerge. The soil association stratification has had a similar effect because there was quite a strong relationship between parent material and soil type. The influence of land-use was less obvious as its major contribution to the signal was organic matter and anthropogenic inputs. Unfortunately, no comparison can be made to the US Soil Survey⁹, as the large variability encountered in 220 sites sampled across two transects (north-to-south and east-to-west) of the USA and Canada did not permit similar grouping by soil type or land-use.

Not all models stratified by soil, land-use or rock type displayed improved predictive performance, which may reflect the diversity of the sample group. 72 out of 600 stratified models returned good to excellent model estimates. The elements that produced excellent models are indicated for each stratification type in

Table 1 (if good models were included, then the number of elements would increase by a third). Overall, stratification resulted in excellent calibrations, particularly for elements grouped in the poor metals (5 out of 7 elements) and transition metals (9 out of 18 elements), and less so for alkali, alkali earth and rare earth metals. Excellent models were predicted for soils overlying limestone, sandstone and shale parent materials, predominantly peat and podzol soils, and forest and rough grass land-use types. Most notably, when stratified by land-use type, no excellent predictions were produced for grassland soils, the largest land-use type in Ireland.

Of the contaminant metals studied by spectroscopy to date,^{5,6,7,8} As, Cd, Cu, Fe and Mn produced excellent models despite not capturing the full concentration range in the NSDB, whilst others reached good model status (Cr, Hg, S and Zn). Concentration range did not dictate which geochemical properties were well modelled as both Sb and Ni, representing 83 and 50% of the NSDB concentration range, or 0 to 4.4 mg.kg⁻¹ and 1.7 to 176.0 mg.kg⁻¹, respectively, failed to produce good models.

Table 1. The geochemical properties that produced excellent prediction (RPD >2.5) when stratified by soil, land-use or parent rock material type.

Geochemical property	Stratification type		
	Parent rock material	Soil type	Land use type
Al			Forest
As		Podzol	Rough grass
Ca	Impure limestone	Grey Brown Podzolic	
Cd	Pure limestone		
Ce		Podzol	
Co	Sandstone	Podzol	
Cu		Podzol	
Fe	Sandstone	Podzol	
Ga			Forest
Ge		Podzol	
Mn	Sandstone, Shale	Acid brown Earth, Peat, Podzol	Rough grass
P		Podzol	
Pb			Forest
Rb	Impure limestone	Grey Brown Podzolic, Peat	Forest
Sn	Impure limestone		Forest
Ta		Peat	
Ti		Peat	
U		Acid brown Earth	
W		Podzol	Tillage

Conclusion

This study evaluated the potential of predicting baseline soil geochemistry from spectral data. It found that reliable predictions were not achieved for geochemical properties derived from a national scale dataset. This was because the geochemical signal was in all cases too weak compared to the diversity of the mineral component of the soil. Stratification by soil, land-use and parent rock material all permitted good to excellent models to be generated, demonstrating that it is necessary to develop soil and parent type specific models and that a generalised national scale model will not be appropriate. Efforts should be concentrated on validating the geochemical properties which produced excellent status models to reduce the need for conventional analysis in soil quality monitoring.

Acknowledgements

This project is funded by the Environmental Protection Agency, 2007-CD-1-S1, funded by the National Development Plan (2007-2013).

References

1. R.E. Creamer, F. Brennan, O. Fenton, G. Healy, S.T.J. Lalor, G.L. Lanigan, J.T. Regan and B.S. Griffiths, *Soil Use Manage.* **26**, 198-211 (2010).
2. S.M. O'Rourke and N.H. Holden, *Soil Use Manage.* **27**(2), 143-155 (2011).
3. D.J. Brown, R.S. Bricklemeyer and P.R. Miller, *Geoderma* **129**, 251-267 (2005).
4. K.D. Shepherd and M.G. Walsh, *Soil Sci. Soc. Am. J.* **66**, 988-998 (2002).
5. Y. Wu, J. Chen, X. Wu, Q. Tian, J. Ji and Z. Qin, *Appl. Geochem.* **20**, 1051-1059 (2005).
6. T. Kemper and S. Sommer, *Environ. Sci. Technol.* **36**, 2742-2747 (2002).

Reference paper as:

S. O'Rourke and N. Holden (2012). The potential of near infrared laboratory based hyperspectral imaging for the determination of geochemical and carbon properties in Irish soils, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 290-293.

7. G. Siebielec, G.W. McCarty, T.I. Stuczynski and J.B. Reeves III, *J. Environ. Qual.* **33**, 2056-2069 (2004).
8. D.F. Malley and P.C. Williams, *Environ. Sci. Technol.* **31**, 3461-3467 (1997).
9. J.B. Reeves III and D.B. Smith, *Appl. Geochem.* **24**, 1472-1481 (2009).
10. A.Walkley and I.A. Black, *Soil Sci.* **37**, 29-37 (1934).
11. J.S. Bailey and A.F. Wright, *Comm. Soil Sci. Plant Anal.* **32**, 3243-3258 (2001)
12. D. Fay, G. Kramers, C. Zhang, D. McGrath and E. Grennan, *Soil Geochemical Atlas of Ireland*, Teagasc and the Environmental Protection Agency, Ireland (2007).
13. G.M. Vasques, S. Grunwald and J.O. Sickman. *Geoderma* **146**, 14-25 (2008).

Reference paper as:

S. O'Rourke and N. Holden (2012). The potential of near infrared laboratory based hyperspectral imaging for the determination of geochemical and carbon properties in Irish soils, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 290-293.