Modeling regional and vertical variation effects when estimating soil nitrogen from loss on ignition

A. Leon[A], M. S. Cresser[B] and R. Leon[C]

[A]National Institute for Agro-Environmental Sciences, 3-1-3 Kannondai, Tsukuba, 305-8604, Japan, Email leon@affrc.go.jp
[B]Environment Department, The University of York, Heslington, York, YO10 5DD, UK, Email msc@york.ac.uk
[C]National Graduate Institute for Policy Studies (GRIPS), 7-22-1 Roppongi, Minato-ku, Tokyo, 106-8677, Japan

Abstract
Knowledge of soil total N stocks is potentially a useful basis for environmental planning and management and in the design of pollution abatement strategies. However, a shortage of data for soil total N deep in the soil profiles makes calculation of the total N pool in soils difficult. We have examined how regional and vertical variations are best taken into account when loss on ignition (LOI) is used to predict missing values of soil total N as measured by the semi-micro Kjeldahl method. Even though statistical tests indicated significant spatial variations, prediction error was little reduced by taking the spatial information into account (i.e. a 4.5% reduction), except for B and C horizon soils, where the reduction was larger by taking account of depth (25.0% and 14.3%). Therefore, one can predict total N (%) across a range of soils by multiplying LOI (%) by 0.020. This approach is useful when one uses published data with much more abundant LOI data for soil profiles than directly determined total N data. However, this can only be used to obtain a quick and approximate estimate of total N.

Key Words
Bayesian information criterion, cross-validation, soil parent material.

Introduction
It is a challenge for sustainable land and water resource management to avoid practices or nitrogen (N) pollution loads that result in adverse changes in the soil total N pool, while meeting the demands of society, such as increase in agricultural output. For this purpose, it becomes vital to understand quantitatively what factors govern total N pool in the soil. However, the calculation of soil total N pool often is made difficult by a shortage of data for total N in available soil survey data (Batjes 1996), especially deeper in soil profiles. To circumvent this problem, regression equations based on the relationships between total N, as measured for example by the semi-micro Kjeldahl method, and LOI may be used. Missing N values may then be imputed when LOI data are available. While the C *vs* LOI relationships have been analysed by many studies (e.g. Ball 1964 among others), total N *vs* LOI relationships seem to have been investigated only by Jackson (1958) and Craft *et al.* (1991). Little attempt has been made to assess regional and vertical variations in soil total N *vs* LOI relationships. Furthermore, none of these studies examined whether taking account of these factors improves total N prediction. Differences in mineral types and relative amount of clay minerals might change the way in which temperature affects LOI values (e.g. Ball 1964). In addition, the soil C:N ratio is likely to change with depth (e.g. Vejre *et al.* 2003). Consequently, the equations that describe total N *vs* LOI relationships might vary among parent materials or depth in a soil profile. However, even if these factors affect the relationships, failure to include them into the regression equation used for prediction might not significantly affect total N out-of-sample prediction.

Therefore, the aims of this paper were (1) to test the hypothesis that the relationships between total N and LOI change with soil layer position (depth) in the soil profile, with soil parent material, and/or with type of soil horizon and (2) to check whether exploiting this spatial information helps to improve out-of-sample predictions of total N. Although LOI might overestimate organic matter components in soil because of carbon dioxide driven from carbonates and water released from soils containing large amounts of clay (e.g. Robinson 1949), we decided to investigate the use of LOI for predictive purposes because LOI data are much more abundant for soil profiles than directly determined total N data.

Materials and methods
Data resources for calibration: We used readily available soil survey data taken from the Country round Aberdeen, Inverurie and Fraserburgh (Glentworth and Muir 1963). The area is in north-east Scotland, to the north and west of Aberdeen, and covers 1617.3 km^2. Soils in this area developed on a diverse range of parent

1 – 6 August 2010, Brisbane, Australia. Published on DVD.

materials (Table 1). A total of 160 soil profiles had data recorded for LOI and total N Loss on ignition was measured for horizons from the surface to 1.04 m on average, whereas total N was measured from the surface to 278 mm on average. Twenty three soil profiles were natural (minimally managed) profiles and the others were cultivated soils. The average depths for LFH (n=12), A (n=60), B (n=124) and Ap (n=149) horizon soils were 76, 102, 102, 163 mm, respectively. There was no directly determined total N for the C horizon soils. Total N in the entire calibration (n=345) ranged between 0.02 and 1.98 (mean 0.27, standard deviation: SD 0.32). Loss on ignition ranged between 0.08 and 95.0 (mean 11.5, SD 14.7).

Table 1. Area covered by each of the 12 Associations used in the present study, in km^2 and as a proportion of the total area of Scotland (data from the Macaulay Institute for Soil Research 1984).

Associations	Parent material	Area (km^2)	Area (%)
Boyndie/ Corby	Fluvio-glacil sands and gravels derived from acid rocks	2377	3.08
Collieston	fluviatile deposits derived from Old Red Sandstone sediments together with gravel layers	153	0.20
Countesswells	till derived from granites and granitic rocks	4435	5.75
Foudland	till derived from weakly metamorphosed rocks	2508	3.25
Insch	till derived from basic igneous rocks	518	0.67
Cuminestown/ Ordley	till derived from Old Red Sandstone sediments and argillaceous schists	166	0.22
Peterhead	till derived from sedimentary rocks of Old Red Sandstone sediments	138	0.18
Strichen	till derived from quartz-mica-schist	6151	7.98
Tarves	till derived from acid and basic rocks	1595	2.07
Tipperty	lacustrine sediments derived from Old Red Sandstones	51	0.07

Independent validation data: Data for model validation came from both within and outside the geographical area used for calibration. The data were originally produced by the Soil Survey of Scotland in Aberdeen, and made available from the Macaulay Institute (J. Gauld, pers. comm.). Total N and LOI (n=1007) ranged between 0.01 and 3.52 (mean 0.42, SD 0.57) and between 0.50 and 98.3 (mean 16.7, SD 24.2), respectively. Analytical methods for both validation and calibration data sets: Loss on ignition was measured on 5 to 10 g of finely-ground samples (2-mm sieved). Ground and sieved air-dried samples were weighed and oven-dried at 105 ºC for a minimum of 3 hours and reweighed. Then they were ignited at 800-900 ºC for 30 minutes, which is higher than commonly used for ignition. Therefore, the extent of overestimation might be larger. Soil N was determined by the semi-micro-Kjeldahl method (Glentworth and Muir 1963).

Statistical methods: We considered influences of three types of spatially variable information (i.e. layer position, parent material and soil horizon) on total N - LOI relationships. Four sets of dummy variables were used to represent the spatial information: 1 dummy to stand for information on the upper layer, 11 dummies for parent material, 4 dummies for type of soil horizon, and 23 dummies for soil layer and parent material at the same time (i.e. this set includes a dummy variable for each combination of parent material and soil layer). With these dummies, we considered nine ways of representing spatial information in N - LOI relationships, as shown in Table 2. All of these included a function of LOI (*f*(LOI)). One of these models did not include information on spatial variation (i.e. it included only *f*(LOI)). Four other models represented spatial information by including only a set of intercept dummies, whereas another four included, in addition to one set of intercept dummies, interactions between *f*(LOI) and the dummies. In each model that took spatial information into account, an F-test was carried out to analyse whether the spatial regressors were statistically significant.

Table 2. Description of models.

	Regressors	*k* in Linear
Model 1	*f(LOI)*	2
Model 2	*f(LOI)*, D_U and interactions between *f(LOI)* and D_U	4
Model 3	*f(LOI)* and D_U	3
Model 4	*f(LOI)*, D_P and interactions between D_P and *f(LOI)*	24
Model 5	*f(LOI)* and D_P	13
Model 6	*f(LOI)*, D_{UP} and interactions between D_{UP} and *f(LOI)*	48
Model 7	*f(LOI)* and D_{UP}	25
Model 8	*f(LOI)*, D_S and interactions between D_S and *f(LOI)*	10
Model 9	*f(LOI)* and D_S	6

D_U consists of 1 dummy for the upper layer, D_P of 11 dummies for parent material, D_S of 4 dummies for soil horizon and D_{UP} of 23 dummies for each combination of soil layer and parent material. k is the number of regressors including the intercept constant. The equations derived in this study were used subsequently to predict total N values for the validation data set. In order to compare the prediction accuracy, the root mean square error (RMSE) and leave-one-out cross validation analyses (Geisser 1975) were carried out. For the latter analysis, as an additional assessment of the prediction accuracy of each equation, the validation and calibration datasets were pooled. In both standard validation and cross validation analyses, predictions for total N were truncated at 0% or 100% whenever model predictions lay outside the interval (0-100%). We calculated RMSE using sample weights for the validation sample (Deaton 1997) to get an unbiased estimation of the overall average error for Scottish soils. The weight for each observation depended on the Soil Association to which it belonged and was equal to the area in km^2 of each Soil Association in Scotland (Table 1). With regard to the RMSE analysis, as noted before, there were no directly measured total N data for C horizon soils in the calibration data set, which is a common situation in many studies, since data on C horizons is less available. Because of this, when making predictions for C horizon using models 8 and 9 (Table 2), a natural approach would be to predict total N for C horizon in the same way as one would predict it for B horizon soils, since it is the nearest horizon. This was the approach taken in the RMSE analysis. Note that this difficulty did not arise in the cross validation analysis, as in this case the sample did include data from C horizons.

Results and discussion

Bayesian information criterion (BIC, Schwarz 1978) preferred models that took type of soil horizon into account (Table 3). That is, models 9 and 8 are preferred. The estimated coefficients derived in Model 8 are shown in Table 4. This result was also supported when calibration and validation data were merged. Furthermore, it was found that in all models, except for model 5, the spatial regressors were jointly significant at a 1% significance level (Table 3). Moreover, if the calibration and validation data were pooled, the spatial regressors were jointly significant at a 1% level or less in all models (except for model 5). The slope of a simple linear equation derived in this study was 0.020 (Table 4). This slope is similar to that previously determined by Jackson (1958) for terrestrial soils (0.022-0.03) and by Craft et al. (1991) for marsh soils. However, the results of the F-test showed that the equations derived in this study differed significantly from their studies at a 1% level.

Table 3. The Bayesian information criterion (BIC) values† and the results of F-test checking the significance of spatial regressors.

	Linear equation (n=345)								
	Model1	Model 2	Model 3	Model 4	Model 5	Model6	Model 7	Model 8	Model 9
BIC	276	317	296	235	248	228	240	395	350
p-values	-	0.000	0.000	0.003	0.633	0.000	0.000	0.000	0.000
	Linear equation when calibration and validation data were combined (n=1352)								
BIC	140	160	163	132	111	90	108	181	174
p-values	-	0.000	0.000	0.000	0.039	0.000	0.000	0.000	0.000

†a model with the largest BIC is most preferable.

Table 4. The equation preferred by Bayesian information criterion (BIC) and the simple linear equation.

Equation preferred by BIC (Model 8) †

Total N%=0.049+0.016×LOI +(-0.082+0.009×LOI)$D_{A\ horizon}$+ (-0.002+0.008×LOI)$D_{Ap\ horizon}$ + (-0.040+0.001×LOI) $D_{B\ horizon}$

s.e. (0.053) (0.001) (0.054) (0.002) (0.063) (0.004) (0.054) (0.002)

The simple linear equation (Model 1)

Total N%=0.038+0.020×LOI

s.e. (0.011) (0.001)

†Note that the set of potential explanatory variables does not include a dummy for LFH horizon in Model 8. This omission is necessary to avoid a problem of perfect multi-collinearity.

Despite these statistical results, however, RMSE in the standard validation analysis and cross-validation exercises was not smaller in models that accounted for spatial information (Table 5). Cross-validation excises showed RMSE was reduced by only 4.5% by taking account of the spatial information (this was also found when f(LOI) was quadratic). However, RMSE in cross-validation exercises showed that prediction errors in B and C horizon soils were reduced by 25.0% and 14.3%, respectively, when information on type of soil horizon was used (model 8). This result suggests the importance of taking account of type of soil horizon, which can be attributed to change in soil C:N ratio with depth (e.g. Vejre *et al.* 2003). Furthermore, LOI is overestimated more when LOI values are smaller than when LOI values are larger (Howard 1965). One of the sources of variation not accounted for in our approach is amount of inorganic nitrogen. Organic matter as measured by LOI was used to predict total N analysed by the semi-micro Kjeldahl method, since most of nitrogen is organically bound. However, inorganic N cannot be predicted with LOI. The amounts of inorganic nitrogen will differ with several factors, such as pH (Ste-Marie and Pare 1999), or temperature, moisture (Sierra 1997; Agehara and Warncke 2005).

Table 5. The root mean square error and cross validation values for each model†.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model6	Model 7	Model 8	Model 9
					RMSE (linear equation) (n=1007)				
All	0.26	0.28	0.26	0.31	0.27	0.27	0.26	0.30	0.29
					Cross validation (linear equation)				
All	0.22	0.21	0.21	0.21	0.22	0.21	0.21	0.21	0.21

†Note that with regard to the RMSE analysis, when making predictions for C horizon using models 8 and 9, total N was predicted in the same way as one would predict it for B horizon soils, since there was no directly determined total N for this horizon. On the other hand this difficulty did not arise in the cross validation analysis, as in this case the sample did include data from C horizons

Conclusions

Failure to include regional and vertical variations in total N measured by the semi-micro Kjeldahl method *vs* LOI relationships into the regression equation did not significantly affect total N prediction. However, for B and C horizon soils it appears probable that a significant influence of soil horizon on LOI *vs* total N exists. Accordingly, one can predict total N across a range of soils by multiplying LOI by 0.020 for surface horizons. This approach is useful when one uses published data with much more abundant LOI data for soil profiles than directly determined total N data. However, this can only be used to get a quick and approximate estimate of total N.

Acknowledgement

The authors are indebted to Dr. Jimmy Gauld at The Macaulay Land Use Research Institute, Aberdeen, for provision of data and helpful advice.

References

Agehara S, Warncke DD (2005) Soil Moisture and temperature effects on nitrogen release from organic nitrogen sources, *Soil Science Society of America Journal* **69**, 1844-1855.

Ball DF (1964) Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils. *Journal of Soil Science* **15**, 84-92.

Batjes NH (1996) Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science* **47**, 151-163.

Craft CB, Seneca ED, Broome SW (1991) Loss on ignition and Kjeldahl digestion for estimating organic carbon and total nitrogen in estuarine marsh soils: calibration with dry combustion. *Estuaries* **14**, 175-179.

Deaton A (1997) ‘The Analysis of Household Surveys: A Microeconometric Approach to Development Policy’ (Johns Hopkins University Press: London).

Geisser S (1975) The predictive sample reuse method with applications. *Journal of The American Statistical Association* **70**, 320-328.

Glentworth R and Muir JW (1963) ‘The Soils round Aberdeen Inverurie and Fraserburgh’ (H.M.S.O.: Edinburgh).

Howard PJA (1965) The carbon-organic matter factor in various soil types. *Oikos* **15**, 229-236.

Jackson ML (1958) ‘Soil Chemical Analysis’ (Prentice-Hall: Englewood Cliffs, N.J.)

Macaulay Institute for Soil Research (1984) ‘Organization and Methods of the 1:250 000 Soil Survey of Scotland’ (The Macaulay Land Use Research Institute: Aberdeen).

Robinson GW (1949) ‘Soils, Their Origin, Constitution, and Classification’. (Wiley : New York).

Schwarz G (1978) Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.

Sierra J (1997) Temperature and soil moisture dependence of N mineralization in intact soil cores. *Soil Biology & Biochemistry* a**29**, 1557-1563.

Ste-Marie C, D Paré (1999) Soil, pH and N availability effects on net nitrification in the forest floors of a range of boreal forest stands. *Soil Biology & Biochemistry* **31**, 1579-1589.

Vejre H, Callesen I, Vesterdal L, Raulund-Rasmussen K (2003) Carbon and nitrogen in Danish Forest Soils – contents and distribution determined by soil order. *Soil Science Society of America Journal* **67**, 335-343.