Biosystems Engineering, Volume 114, Issue 3, March 2013, Pages 249–258

# Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture

**Boyan Kuang[1]; Abdul Mounem Mouazen[1, *]**

[1]*Environmental Technology and Science Department, School of Applied Sciences, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK.*

*E-mail of corresponding author:a.mouazen@cranfield.ac.uk*

## Abstract

This study was undertaken to evaluate the effects of moisture content (MC) and texture on the prediction of soil organic carbon (OC) and total nitrogen (TN) with visible and near infrared (vis-NIR) spectroscopy under laboratory and on-line measurement conditions. An AgroSpec spectrophotometer was used to develop calibration models of OC and TN using laboratory scanned spectra of fresh and processed soil samples collected from five fields on Silsoe Farm, UK. A previously developed on-line vis-NIR sensor was used to scan these fields. Based on residual prediction deviation (RPD), which is the standard deviation of the prediction set (S.D.) divided by the root mean square error of prediction (RMSEP), the validation of partial least squares (PLS) models of OC and TN prediction using on-line spectra was evaluated as very good (RPD = 2.01 - 2.24) and good to excellent (RPD = 1.86 - 2.58), respectively. A better accuracy was obtained with fresh soil samples for OC (RPD = 2.11 - 2.34) and TN (RPD = 1.91 - 2.64), whereas the best accuracy for OC (RPD = 2.66 - 3.39) and TN (RPD = 2.85 - 3.45) was obtained for processed soil samples. Results also showed that MC is the main factor that decreases measurement accuracy of both on-line and fresh samples, whilst the accuracy was greatest for soils of high clay content. It is

recommended that measurements of TN and OC under on-line and laboratory fresh soil conditions are made when soils are dry, particularly in fields with high clay content.

## Keywords

Visible, near infrared, spectroscopy, organic carbon, total nitrogen, soil, on-line, laboratory.

## 1. Introduction

Soil organic carbon (OC) is a key property for soil function, soil quality maintenance, plant nutrition supply and soil water holding capacity (Flessa *et al*., 2000). It is the major component of soil organic matter, which is important in all soil processes. Shortages of OC result in decline of soil quality and crop production over the field. There is a continuous cycling of OC in soils that is not uniform and depends mostly on land use and land management systems. Even small changes in OC stocks are associated with important $CO_2$ fluxes between terrestrial ecosystems and the atmosphere (Stevens *et al*., 2006). Soil OC is also an important component when considering soil resistance to wind and water erosion (Morgan, 2005). Nitrogen is an important nutrient for plants and sufficient N fertiliser application is critical for optimal plant growth and development. Nitrogen is a major component in chlorophyll and therefore essential for photosynthesis and crop protein. However, over-application of N fertilisers results not only in economic losses but also potential ground water contamination. Nitrate leaching from land to ground water and stream water causes depletion of soil minerals, acidifies soils and affects downstream freshwater and coastal marine ecosystems (Vitousek *et al*., 1997). Determination of within field variation in soil OC and TN at high resolution sampling might assist management of these properties in agronomic and environmental systems.

During the last two decades, visible and near infrared (vis-NIR) spectroscopy has proved to be a fast, cost-effective, non-destructive and relatively accurate alternative to the traditional laboratory analytical methods for measuring soil physical, chemical and biological properties (Shepherd and Walsh, 2002; Brown *et al*., 2006; Wetterlind *et al*., 2008; Mouazen *et al*., 2010). Although the application of vis-NIR spectroscopy has considerably reduced the labour and time requirements for the analysis, sample preparation for laboratory analysis including grinding and sieving is still tedious. Calibration models developed from processed (dried, ground and sieved) samples cannot be utilised for fresh soil samples and on-line measurements with vis-NIR spectroscopy, since these measurements are performed with unprocessed soils (Mouazen *et al*., 2005). Although MC can be successfully measured with vis-NIR spectroscopy, it is considered as one of the most critical factors that decreases the accuracy of measurement of other soil properties (Bogrekci and Lee, 2006; Minasny *et al*. 2011; Mouazen *et al*. 2006a; Sudduth and Hummel, 1993). Modifications for removing the influence of MC on the accuracy of vis-NIR measurement of other soil properties have been considered by the classification of samples into different MC groups (Mouazen *et al*. 2006a), adoption of an external parameter orthogonalisation (EPO) algorithm (Minasny *et al*. 2011), construction of processed soil spectra from raw spectra (Bogrekci and Lee, 2005) and drying of soil samples to remove MC (Ben-Dor and Banin, 1995; Chang *et al*., 2001). Most reports showed that the highest accuracy is to be expected when dried soil samples are used (Mouazen *et al*., 2006b; Tekin *et al*., 2011). Unfortunately, for fresh soil and on-line measurements, calibration models should be developed based on vis-NIR scanning of fresh soil samples.

Apart from soil MC, soil texture is the other main factor to affect accuracy of vis-NIR spectroscopy. Using processed soil samples in the laboratory, Stenberg (2010) concluded that predictions of OC were most inaccurate for soils with a high sand content. Cozzolino and

74 Moron (2006) found mixed results for different textures, with coefficients of determination in
75 calibration and standard errors in cross validation of 0.90 and 0.6, 0.92 and 0.4 and 0.96 and
76 2.1 for coarse-sand C, fine-sand C and clay + silt C, respectively. All the above literature
77 considered the single effect of MC or texture on the prediction of soil organic matter or OC.
78 The interaction effect of MC and texture fractions on prediction accuracy of OC and TN at
79 farm scale has not been studied so far. Accuracy of measurement under different
80 measurement conditions as affected by texture fractions and MC variation has not been
81 reported. Furthermore, to our knowledge, no reports studying these effects on models'
82 performance using spectra collected on-line can be found in the literature.
83 The objective of this paper is to understand and quantify the individual and interaction effects
84 of MC and soil texture on the performance of vis-NIR calibration models for the prediction of
85 soil OC and TN for laboratory scanned fresh soil and on-line field measurements. This has
86 the goal of improving the measurement performance of the vis-NIR sensor at farm scale.
87

88 ## 2. Material and methods

89

90 *2.1. Soil samples*

91 A total of 174 soil samples were used in this study. They were collected in summer 2009
92 from five fields on Silsoe Farm, UK (*Fig. 1*) with variable soil texture and cropping (Table 1).
93 The soils in these five fields are from the same mother material and have been subjected to
94 similar farm management practices. Soil samples were collected from the surface layer at the
95 bottom of a 15 cm deep trench, opened by a tractor-drawn subsoiler during on-line
96 measurement. Around 200 g soil was taken from each soil sample and placed into a tightly
97 sealed plastic bag to hold field moisture. The soil samples were then transported to the soil
98 laboratory in Cranfield University, where they were stored deep frozen (-18 °C) until analysis.

After defrosting and thorough mixing, each soil sample was divided into two equal parts. One part was used for MC measurement by oven drying at 105 °C for 24 hours. The fresh soil (wet and unprocessed) was scanned with a vis-NIR spectrophotometer before drying. This scanning was designated as laboratory scanning of fresh soil samples. The other part of the soil was dried at 45°C and sieved with a 2 mm sieve, before it was scanned with a vis-NIR spectrophotometer. This was designated as laboratory scanning of processed soil samples. This part was also used to determine soil OC, TN and texture by standard laboratory analyses.

### *2.2. Laboratory analyses*

Laboratory analysis of OC and TN was carried out with a TrusSpecCNS spectrometer (LECO Corporation, St. Joseph, MI, USA) using the Dumas combustion method. Soil texture was measured with sieving and sedimentation methods, according to BS 7755 Section 5.4 (BSI, 1998). The particle size distribution (PSD) analysis of each field was based on a mixed sample collected from each field. Soil texture was classified according to the United State Department of Agriculture (USDA) classification system.

### *2.3. On-line measurement*

The on-line vis-NIR sensor designed and developed by Mouazen (2006) was used (Fig. 2) to carry out the field measurements. It consists of a subsoiler that penetrates the soil to the required depth, making a trench, whose bottom is smoothened due to the downwards forces acting on the subsoiler (Mouazen *et al.*, 2005). The optical probe, housed in a steel lens

holder, was attached to the rear of the subsoiler chisel to acquire soil spectra in reflectance mode from the smooth bottom of the trench. The subsoiler, retrofitted with the optical unit, was attached to a frame that was mounted onto the three point hitch of the tractor (Mouazen *et al*., 2005). An AgroSpec mobile, fibre type, vis-NIR spectrophotometer (tec5 Technology for Spectroscopy, Germany) with a measurement range of 305-2200 nm was used to measure soil spectra in diffuse reflectance mode. The spectrophotometer was IP66 (ingress protection) protected for harsh working environments. Although this spectrophotometer does not cover the entire wavelength range in the NIR region, it was selected in this study as it uses diode array detectors, which have been proven to be stable under on-line measurement conditions (Mouazen *et al*., 2009). A 20 W halogen lamp was used as a light source. A 100 % white reference was used before scanning, and this was repeated every 30 min. A differential global positioning system (DGPS) (EZ-Guide 250, Trimble, USA) was used to record the position of the on-line measured spectra with sub-metre accuracy. A Panasonic semi-rugged laptop was used for data logging and communication. The spectrometer system, laptop and DGPS were powered by the tractor battery. A New Holland T5000 tractor with 100 Ah battery was used. The total power consumption for all electrical parts of the on-line vis-NIR sensor was around 60 W.

The on-line sensor was used to measure five fields on Silsoe Farm in the UK (*Fig. 1*) in summer 2009, namely Avenue, Orchard, Ivy Ground, Shoeground and Copse fields. In each field, blocks of 150 m by 200 m, covering 3 ha of land were measured. Each measured line was 200 m long with 10 m intervals between adjacent transects. The forward speed of the tractor was around 2 km $h^{-1}$ and the measurement depth was set at 150 mm. During each line measurement, two or three soil samples were collected from the bottom of the trench and the sampling positions were carefully recorded.

147 *2.4. Optical measurement*

148

149 Scanning of soil samples took place under fresh (unprocessed) and processed (dried and
150 sieved) conditions. Samples were placed in glass containers and mixed well, with big stones
151 and plant residues excluded. Then each soil sample was placed into three Petri dishes which
152 were 2 cm deep and 2 cm in radius. The soil in the Petri cup was shaken and pressed gently
153 before levelling with a spatula. A smooth soil surface ensures maximum light reflection and
154 high signal to noise ratio (Mouazen *et al*., 2005). The soil samples were scanned in diffuse
155 reflectance mode by the same vis-NIR spectrophotometer (AgroSpec from tec5 Technology
156 for Spectroscopy, Germany). A total of 10 scans were collected from each triplicate, and
157 these were averaged in one spectrum.

158

159 *2.5. Data pre-treatment and establishment of calibration models*

160

161 The data set contained averaged spectra for each sample and was subjected to noise cut to
162 remove the noisy part of spectra on both sides, reducing the wavelength range to 371 - 1900
163 nm. A 3-point (wavelength) average was applied in the visible wavelength range and 10-point
164 average was applied in the infrared wavelength range to reduce the number of wavelengths
165 and smooth the spectra. This was followed successively by Savitzky-Golay (S-G) smoothing,
166 maximum normalisation and first derivative with S-G method. A 2:2 smoothing was first
167 carried out to remove noise from the measured spectra. Normalisation is typically used to get
168 all data to approximately the same scale, or to get a more even distribution of the variances
169 and the average values. The maximum normalisation method adopted in this study is a

normalisation that “polarises” the spectra. The peaks of all spectra with positive values scale to +1, while spectra with negative values scale to -1. Since all the soil spectra in this study had positive values, the peaks of these spectra scaled to +1 (Mouazen, et al., 2006a). The maximum normalisation led to better results for the measurement of OC and TN, as compared to the other pre-treatment options tested including mean and peak normalisation. Spectra were then subjected to Savitzky-Golay first derivative (Martens and Naes, 1989). This method enables the first or higher-order derivatives, including a smoothing factor, to be computed which determines how many adjacent variables will be used to estimate the polynomial approximation used for derivatives. A second order polynomial approximation was selected. The same pre-treatment was used for both OC and TN models, and was carried out using Unscrambler 7.8 software (Camo Inc.; Oslo, Norway).

The pre-treated soil spectra of processed and fresh samples and the laboratory chemical measurement of OC and TN were used to develop calibration models of OC and TN. Out of the total 174 samples, 60 % of samples collected from each field were used for the development of calibration models (104 samples), whereas the remaining 40 % of samples from each field were used as prediction set (70 samples) (Table 2). The calibration spectra were modelled using a partial least squares regression (PLSR) with leave-one-out cross validation using Unscrambler 7.8 software (Camo Inc.; Oslo, Norway). The resulting accuracy is reported in Table 3 for the laboratory processed and fresh sample and on-line measurements. A flow diagram explaining different steps performed from on-line field measurement to model validation with three prediction sets is shown in Fig. (3).

*2.6. Statistical evaluation of model calibration and prediction*

193 Both root mean square error of prediction (RMSEP) and residual prediction deviation (RPD),
194 which is the ratio of the standard deviation of prediction set (S.D.) to RMSEP, were used to
195 compare the prediction accuracy of OC and TN models for processed soils, fresh soils and for
196 on-line measurement. Viscarra Rossel *et al*. (2006) classified RPD values as follows:
197 RPD<1.0 indicates very poor model/predictions and their use is not recommended; RPD
198 between 1.0 and 1.4 indicates poor model/predictions where only high and low values are
199 distinguishable; RPD between 1.4 and 1.8 indicates fair model/predictions which may be
200 used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good
201 model/predictions where quantitative predictions are possible; RPD between 2.0 and 2.5
202 indicates very good, quantitative model/predictions, and RPD>2.5 indicates excellent
203 model/predictions. This classification system was adopted in this study. Furthermore, to
204 evaluate how significant are the separate effects of MC and texture fractions on the prediction
205 accuracy of OC and TN, one way univariate analysis of variance (ANOVA) of between-
206 subjects effects was carried out considering RPD and RMSEP for the prediction set as
207 accuracy indicators. The interaction effects between MC and soil texture fractions were
208 evaluated with two-way ANOVA. Both ANOVA analyses were performed using Office 2007
209 (Microsoft, WD, USA)
210
211 **3. Results and discussion**
212
213 *3.1. Analysis of soil spectra*
214
215 To distinguish and group soil spectra from each field, principal component analysis (PCA)
216 was carried out using raw soil spectra collected in the laboratory from fresh soil samples. The

first two principal components (PCs) accounted for 97 % of total variance with the first PC accounting for 92 % of total variance (*Fig. 4a*). Three groups of soil samples can be observed, which are separated along principal component 1 (PC1), namely, sandy (Avenue and Shoeground), loam (Orchard) and clay (Ivy Ground and Copse) groups. However, separation between two pairs of fields (e.g. Avenue and Shoeground, and Ivy Ground and Copse) from the same group can also be observed, with minimal overlapping of samples. These separations occur along PC2, and might be explained by the different MC in these fields. This is supported by the spectral pattern PC2 (*Fig. 4b*) obtained from the PCA, where the water absorption band (at 950, 1450 and 1950 nm) can be clearly observed as peaks on spectral pattern 2. This implies that there is soil moisture and texture information in the vis-NIR soil spectra, which will potentially influence the soil vis-NIR measurement of soil properties.

To analyse the soil MC and soil texture effects on the soil vis-NIR spectra, the average spectra of each field under fresh and processed conditions are shown in *Fig. 5*. This Figure explains differences in MC and texture of each field, which is reflected in and in line with the PCA plot shown in *Fig. 4a*. It can be clearly observed that the processed soil absorbs less light energy (larger reflectance), when compared to fresh soils from the same field. Also, the clay soils of Ivy Ground and Copse fields have smaller reflectance (larger apparent absorption) in the NIR spectral range than those of loamy soils (Orchard field) and sandy soils (Avenue and Shoeground fields). Due to the interaction effect of MC and texture, average reflectance in the vis-NIR range differs between processed and fresh soil samples (Fig. 5).

### *3.2. Prediction accuracy of models with fresh soil samples*

241 Tables 3 summarises the RMSEP and RPD values of PLS in cross-validation and prediction
242 using spectra of fresh soil samples from the prediction set for each field. According to RPD
243 limits of accuracy proposed by Viscarra Rossel et al. (2006), the accuracy of the prediction of
244 OC in the prediction set is evaluated as very good with RPD values ranging from 2.11-2.34 in
245 all five fields. The accuracy for TN prediction is evaluated as good to excellent with larger
246 range of RPD of 1.91-2.68.

247 Successful vis-NIR calibration models with fresh soil samples have been reported for OC
248 (Fystro *et al*., 2002; Mouazen *et al*., 2010; Kuang *et al*., 2011) with variable accuracy (RPD =
249 1.3 - 4.95, RMSEP = 2.9 - 14 g $kg^{-1}$). However, the RMSEP values obtained in this study
250 (RMSEP = 1.23 – 1.83 g $kg^{-1}$) are considerably lower than those reported in the literature,
251 suggesting a better prediction accuracy. Other studies have reported successful calibration of
252 vis-NIR spectroscopy for TN with fresh soil samples (Chang *et al*., 2005; Cohen *et al*., 2005;
253 Mouazen *et al*., 2006b; Awiti *et al*., 2008) with high accuracy (RPD = 2.1 - 3.88 and RMSEP
254 = 0.2 - 0.6 g $kg^{-1}$), comparable to those obtained in the current study (Table 3). Using fresh
255 soil samples, Fystro (2002) has reported less accuracy for the prediction of TN than was
256 obtained in the current study, although their RMSE values were based on cross-validation
257 procedure.

258

259 *3.3. On-line prediction accuracy of models with fresh soil samples*

260

261 As soil organic matter (SOM) or OC is essential for soil management and carbon
262 sequestration, these have been the main properties considered for on-line vis-NIR
263 measurement (Shonk *et al*., 1991; Shibusawa *et al*., 2001; Hummel *et al*., 2001; Christy *et al*.,
264 2008; Mouazen *et al*., 2007; Bricklemyer *et al*., 2010). Although there is significant spectral

265 information for carbon in both vis and NIR spectral ranges, only few moderately successful
266 cases for on-line measurement have been reported so far (Hummel *et al*., 2001; Christy *et al*.,
267 2008). This might be attributed to the affecting factors during on-line measurement, including
268 noise associated with tractor vibration, sensor-to-soil distance variation (Mouazen *et al*.,
269 2009), stones and plant roots, and difficulties in matching the position of soil samples
270 collected for validation with corresponding spectra collected from the same position
271 (Mouazen *et al*., 2007). Compared to the on-line sensing of SOM or OC, there have been
272 even fewer studies of TN that can be found in the literature (Mouazen *et al*., 2007; Christy *et*
273 *al*., 2008). Only Christy *et al*., (2008) achieved high accuracy for TN (coefficient of
274 determination ($R^2$) = 0.86) for a field scale calibration. This was not confirmed to be stable
275 for different fields, though it is in the current study where on-line validation is classified as
276 good to very good in the five measured fields (Table 3). Actually, there is high correlation
277 between OC and TN, because N content in soil is almost entirely dependent on the organic
278 matter content, and the overall TN:OC ratio is 1:10 (Martin *et al*., 2002).
279
280 *3.4. Prediction accuracy with models of processed soil samples*
281
282 Overall, after MC removal, the accuracy of OC and TN prediction is considerably
283 improved compared to that for the fresh soil samples and on-line predictions, not only in the
284 cross-validation, but also in prediction sets of the five measured fields. Therefore, with
285 processed soil samples, the models of OC and TN are classified as giving excellent accuracy
286 (Table 3). This confirms that soil MC decreases the accuracy of prediction of OC and TN
287 with vis-NIR spectroscopy. The measurement of soil OC can be classified as excellent for
288 both cross-validation (RPD = 3.36) and prediction sets of 5 measured fields (RPD = 2.66 -

289 3.39, RMSEP = 0.76 – 1.60 g $kg^{-1}$), which is better accuracy than achieved in previous
290 studies (Dunn *et al.*, 2002; Fidencio *et al*., 2002; Shepherd & Walsh. 2002; Chang et al., 2005;
291 Brown et al., 2006). The measurement accuracy of TN is even better, which can also be
292 classified as excellent for cross-validation (RPD = 3.84) and prediction sets of five measured
293 fields (RPD = 2.85 - 3.45, RMSEP = 0.08 - 0.1 g $kg^{-1}$), comparable to those obtained in
294 previous studies (Dalal and Henry, 1986; Vagen *et al*., 2006; Guerrero *et al*., 2010).

295

296 *3.5. Interactions of soil moisture content and texture on prediction accuracy*

297

298 *Figures 5* and *6* illustrate the effect of soil MC and texture (in terms of soil clay content) on
299 the measurement accuracy of soil OC and TN, respectively. For fresh soil condition, the
300 increase of field clay content and MC result in a decrease of measurement accuracy of OC
301 and TN in terms of RPD. Although the influence of MC on prediction accuracy of OC and
302 TN with vis-NIR spectroscopy is in line with other studies (Chang *et al*., 2005; Mouazen *et*
303 *al*., 2006b; Tekin *et al*., 2011), clay content has been found to contribute to enhance
304 measurement accuracy (Stenberg, 2010). However, results shown in *Figs (6 & 7)* suggest that
305 clay fraction has the same negative effect as MC on prediction accuracy. When MC is
306 removed by considering processed samples, it becomes clear that clay content plays a
307 positive role on the accuracy of vis-NIR spectroscopy (*Figs 6 & 7*). The illusion that with
308 increase in clay content, the prediction accuracy of OC and TN deteriorates when using fresh
309 soil samples, can be explained by the fact that clay can hold larger amount of water than sand
310 due to the large water holding capacity and plasticity index of clay. When processed soil
311 samples are used the highest accuracy for OC and TN prediction, measured as RPD, is
312 obtained for field with the highest clay content (Copse field), whereas the lowest RPD values

313 is obtained for field with the lowest clay content (Shoeground field). Moreover, as the soil
314 clay content increases, the prediction of soil OC in the field improved. Stenberg (2010)
315 reported an increase in prediction error of OC with increase in sand content. The author
316 proposed that, due to a small surface area in sand as compared to clay, the organic matter is
317 more visible. This can be considered as a clay-free soil with a normal small amount of OC
318 being more or less black, while a clay soil with the same amount of OC is much brighter
319 (when processed). This would be consistent with a general over-prediction of OC in very
320 sandy soils.
321 Tables 4 and 5 summarise the output of ANOVA, to evaluate how significant the effects of
322 MC and soil texture fractions and the interaction between them are on the prediction of OC
323 and TN under both fresh soil conditions (Table 4) and on-line (Table 5) measurement
324 conditions. In terms of measurement accuracy evaluated as RPD, $R^2$ and RMSEP, MC, clay
325 content, silt content and sand content all have significant effects on the prediction of OC
326 ($p<0.05$), with MC has the most significant influence ($p = 0.0001$ regarding RPD). Tekin et al.
327 (2011) found the effect of MC on the prediction of soil OC to be significant when this was
328 tested on a data set with mixed texture samples collected from the UK and Turkey, which
329 supports the finding of the current study. Further analysis shows that the interaction effects
330 between MC and texture are significant for all accuracy standards for both on-line and fresh
331 soil samples conditions.
332

333 ## 4. Conclusions

334

335 This study was carried out to understand and quantify the individual and interaction effects of
336 moisture content (MC) and texture fractions on accuracy of organic carbon (OC) and total
337 nitrogen (TN) predictions using laboratory-scanned visible and near infrared (vis-NIR)

spectra of processed and fresh soil samples and on-line collected spectra. Results obtained lead to the following conclusions:

1. Soil MC, sand and silt fractions have negative influences on measurement accuracy, whereas clay content has a positive effect. Soil MC, silt, clay and sand fractions play significant role for the prediction of soil OC and TN with the vis-NIR spectroscopy, with MC having the most significant influence on both on-line and fresh soil sample measurement conditions.
2. The interaction effects of MC with soil texture were found to be significant for both on-line and fresh soil sample measurements, though this was less significant than individual effects.
3. A better accuracy of vis-NIR spectroscopy of soil OC and TN is expected for processed clayey than for fresh sandy samples. The worst accuracy is expected in the highest clay content soils under the wettest conditions. When measurement is needed under both mobile (on-line) or non-mobile conditions, it is recommended to access the field under driest possible conditions, particularly in clay fields. Accessing clay fields under wet condition may result in the highest expected error, as water holding capacity of clay is high. Larger error is to be expected in this case as compared to that in fields with light soils.

## Acknowledgements

This project was part of the collaborative research project FutureFarm. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 212117.

## References


Awiti, A.O., Markus, G.W., Kinyamario, J. (2008). Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma*, 143, 73–84.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D. & Reinsch, T.G. (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, 132, 273-290.

Bogrekci, I & Lee, W. S. (2005). Improving phosphorus sensing by eliminating soil particle size effect in spectral measurement. *Transactions of the American Society of Agricultural and Biological Engineers*, 48, 1971-1978.

Bogrekci, I & Lee, W. S. (2006). Effects of soil moisture content on absorbance spectra of sandy soils in sensing phosphorus concentrations using UV-VIS-NIR spectroscopy. *Transactions of the American Society of Agricultural and Biological Engineer*, 49,1175-1180.

Ben-Dor, E. & Banin, A. (1995). Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal, 59, 364-372.

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

British Standard Institute (BSI). (1998) "British Standard BS 7755 Section 5.4:1998, Determination of particle size distribution in mineral soil material – Method by sieving and sedimentation", BSI 389 Chiswick High Road, London W4 4AL, UK.

Chang, C. W. Laird, D. A. Mausbach, M. J. & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Science Society of America Journal, 65, 480-490.

Chang, G. W. Laird, D. A. & Hurburgh, G. R. (2005). Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. *Soil Science,* 170, 244–255.

Christy, C. D. (2008). Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Comput Electron Agric*, 61,10–19.

Cohen, M.J., Prenger, J.P. & DeBusk, W.F. (2005). Visible-near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. *Journal of Environmental Quality*, 34, 1422-1434.

Cozzolino, D. Moron, A. (2006). The potential of near-infrared reflectance spectroscopy to analyze soil chemical and physical characteristics. *Journal of Agricultural Science* , 140,65-71.

Dalal, R. C. & Henry, R. J. (1986). Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Sci. Soc. Amer. J,* 50, 120-123.

Dunn, B.W., Beecher, H.G., Batten, G.D. & Ciavarella, S. (2002). The potential of near-infrared reflectance spectroscopy for soil analysis—A case study from the Riverine Plain of south- eastern Australia. *Australian Journal of Experimental Agriculture*, 42, 607–614.

Fidencio, P.H., Poppi, R.J., & Andrade, J.C. (2002). Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Analytica Chimica Acta*, 453, 125-134.

Flessa H, Ludwig B, Heil B, and Merbach, W. (2000). The origin of soil organic C, dissolved organic C and respiration in a long-term maize experiment in Halle, determined by 13C natural abundance. *J. Plant Nutr. Soil Sci,* 163, 157-163.

Fystro, G. (2002). The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. *Plant Soil*, 246, 139–149.

Guerrero, C., Zornoza, R., Gómez, I. & Mataix-Beneyto, J. 2010. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. Geoderma, **158**, 66-77.

Hummel, J. W. Sudduth, K. A. Hollinger, S. E. (2001). Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Comput. Electron. Agric*, 32, 149–165.

Kuang, B. & Mouazen, A.M. (2011). Calibration of a visible and near infrared spectroscopy for soil analysis at field scales across three European farms. *European Journal of Soil Sciences*, 62, 629–636.

Martens, H., Naes, T. Multivariate Calibration. (2nd edition). (1989). John Wiley & Sons, Ltd., Chichester, United Kingdom.

Martin, P.D., Malley, D.F., Manning, G. & Fuller, L. (2002). Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. Canadian Journal of Soil Science, 82, 413-422.

Minasny, B. McBratney, A. B. Bellon-Maurel, V. Roger, J. Gobrecht, A. Ferrand, L. Joalland, S. (2011). Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma, 167-168, 118-124.

Morgan, R.P.C. (2005). Soil Erosion and Conservation. Third edition, Blackwell publishing, Oxford, UK. 303 p.

Mouazen, A.M., De Baerdemaeker, J., Ramon, H. (2005). Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. Soil & Tillage Research, 80, 171-183.

435 Mouazen, A. M., Karoui, R., De Baerdemaeker, J., & Ramon, H. 2006a. Characterization of
436 soil water content using measured visible and near infrared spectra. *Soil Science Society of*
437 *America Journal*, 70, 1295-1302.
438 Mouazen, A. M., De Baerdemaeker, J. & Ramon, H. 2006b. Effect of wavelength range on
439 the measurement accuracy of some selected soil properties using visual-near infrared
440 spectroscopy. *Journal of Near Infrared Spectroscopy*, 14, 189-199.
441 Mouazen A. M. (2006). Soil Survey Device. International publication published under the
442 patent cooperation treaty (PCT). World Intellectual Property Organization, International
443 Bureau. International Publication Number: WO2006/015463;
444 PCT/BE2005/000129; IPC: G01N21/00; G01N21/00.
445 Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., and Ramon, H. (2007). On-line
446 measurement of some selected soil properties using a VIS-NIR sensor. *Soil and Tillage*
447 *Research*, 93, 13-27
448 Mouazen, A. M., Maleki, M. R., Cockx, L., Van Meirvenne, M., Van Holm, L. H. J., Merckx,
449 R., De Baerdemaeker, J., Ramon, H. (2009). Optimum three-point linkage set up for
450 improving the quality of soil spectra and the accuracy of soil phosphorus measured using
451 an on-line visible and near infrared sensor. *Soil and Tillage Research*, 103, 144-152.
452 Mouazen, A.M., Kuang, B., De Baerdemaeker, J. & Ramon, H. (2010). Comparison between
453 principal component, partial least squares and artificial neural network analyses for
454 accuracy of measurement of selected soil properties with visible and near infrared
455 spectroscopy. *Geoderma*, 158, 23-31.
456 Shepherd, K.D. & Walsh, M.G. (2002). Development of reflectance spectral libraries for
457 characterization of soil properties. *Soil Science Society of America Journal*, 66, 988-998.

458 Shibusawa, S. Made Anom, S.W. Sato, H. P. Sasao, A. (2001). Soil mapping using the real-
459 time soil spectrometer. In ‘‘ECPA 2001’’ (G. Gerenier and S. Blackmore, Eds.),Vol. 2, pp.
460 485–490. agro Montpellier, Montpellier, France.
461 Shonk, J. L, Gaultney L. D, Schulze D. G, Scoyoc, G. E. V. (1991). Spectroscopic sensing of
462 soil organic matter content. *Trans. ASAE*, 34: 1978–1984.
463 Stevens, A. Wesemael, B. Vandenschrick, G. Toure, S. & Tychon, B. (2006). Detection of
464 carbon stock change in agricultural soils using spectroscopic techniques. *Soil Sci. Soc. Am.*
465 *J.* 70, 844–850.
466 Stenberg, B. (2010). Effects of soil sample pre-treatments and standardised rewetting as
467 interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon.
468 Geoderma, 158(1-2), 15-22.
469 Sudduth, K. A. and Hummel, J. W. (1993). Soil organic matter, CEC, and moisture sensing
470 with a prototype NIR spectrophotometer. *Transactions of the ASAE*, 36, 1571-1582.
471 Tekin, Y. Tumsavas, Z. Mouazen, A. M. (2011). Effect of moisture content on prediction of
472 organic carbon and pH using visible and near infrared spectroscopy. *Soil Science Society*
473 *of America Journal.* doi:10.2136/sssaj2011.0021.
474 Vagen Tor-G, Keith D. Shepherd, Markus G. Walsh. (2006). Sensing landscape level change
475 in soil fertility following deforestation and conversion in the highlands of Madagascar
476 using Vis-NIR spectroscopy. Geoderma, 133, 281–294.
477 Vitousek, P. M. Aber, J. D. Howarth, R. W. Likens, G. E. Matson, P. A. Schindler, D. W.
478 Schlesinger, W. H. Tilman, D. G. (1997). Human alteration of the global nitrogen cycle:
479 sources and consequences. *Ecological application*, 7, 737-750.
480 Viscarra Rossel, R. A., McGlynn, R.N., McBratney, A. B. (2006). Determining the
481 composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance
482 spectroscopy. Geoderma, 137, 70-82.

Wetterlind, J., Stenberg, B. & Söderström, M. (2008). The use of near infrared (NIR) spectroscopy to improve soil mapping at the farm scale. *Precision Agriculture.* 9, 57-69.

**Table 1**

**Information about test fields on Silsoe Farm, UK**

| *Field* | *Area, ha* | *Crop* | *Sample Nr* | *Sand, %* | *Silt, %* | *Clay, %* | *Texture* | *MC, %* |
|---|---|---|---|---|---|---|---|---|
| Avenue | 3 | wheat | 28 | 61.87 | 20.06 | 18.07 | Sandy loam | 13.66 |
| Orchard | 2 | wheat | 26 | 40.11 | 27.38 | 32.51 | Clay loam | 17.84 |
| Ivy | 3 | soybean | 40 | 21.14 | 27.17 | 51.69 | Clay | 25.05 |
| Shoeground | 4 | wheat | 40 | 64.98 | 20.93 | 14.09 | Sandy loam | 13.43 |
| Copse | 3 | wheat | 40 | 14.55 | 27.84 | 57.61 | Clay | 26.88 |

522 **Table 2**

523 **Sample statistics of organic carbon (OC) and total nitrogen (TN) for calibration set and**
524 **prediction set of individual field samples and of all 174 samples collected from five fields**
525 **on Silsoe farm, UK**

| | *Calibration set* | | | | | | | | | *Prediction set* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Field | OC, g kg$^{-1}$ | | | | | TN, g kg$^{-1}$ | | | | OC, g kg$^{-1}$ | | | | | TN, g kg$^{-1}$ | | | |
| | No | Min | Max | Mean | SD | Min | Max | Mean | SD | No | Min | Max | Mean | SD | Min | Max | Mean | SD |
| Avenue | 17 | 9.4 | 22 | 17 | 4.2 | 0.9 | 1.9 | 1.5 | 0.35 | 11 | 13 | 18 | 16 | 3.8 | 1.2 | 1.7 | 1.5 | 0.31 |
| Orchard | 15 | 16 | 22 | 20 | 3.9 | 1.6 | 2.1 | 1.9 | 0.27 | 11 | 17 | 25 | 20 | 3.6 | 1.7 | 2.0 | 1.9 | 0.26 |
| Ivy | 24 | 24 | 35 | 28 | 3.2 | 2.4 | 3.1 | 2.7 | 0.33 | 16 | 27 | 34 | 30 | 2.8 | 2.2 | 3.1 | 2.8 | 0.28 |
| Shoeground | 24 | 12 | 20 | 15 | 4.4 | 1.1 | 1.9 | 1.4 | 0.25 | 16 | 12 | 20 | 16 | 4.3 | 1.2 | 1.9 | 1.5 | 0.27 |
| Copse | 24 | 22 | 29 | 25 | 2.8 | 2.2 | 2.7 | 2.5 | 0.34 | 16 | 20 | 30 | 26 | 2.6 | 1.9 | 2.6 | 2.4 | 0.31 |
| Overall | 104 | 9.4 | 35 | 21 | 7.9 | 0.9 | 31 | 2 | 0.76 | 70 | 13 | 34 | 22 | 7.4 | 1.2 | 3.1 | 2.2 | 0.72 |

526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549

**Table 3**

**Validation results of partial least squares (PLS) regression in cross validation (CV) and in prediction (P) for the prediction of soil organic carbon (OC) and total nitrogen (TN) based on spectra of processed and fresh soil sample and on-line field measurements**

| *Measurement* | *Field* | *OC* | | | *TN* | | |
|---|---|---|---|---|---|---|---|
| | | RMSEP, g $kg^{-1}$ | RPD | SD | RMSEP, g $kg^{-1}$ | RPD | SD |
| CV – Fresh | All | 3.34 | 2.36 | 7.9 | 0.28 | 2.71 | 0.76 |
| CV – Processed | All | 2.04 | 3.36 | 7.9 | 0.19 | 3.84 | 0.76 |
| P – Fresh | Avenue | 1.69 | 2.24 | 3.8 | 0.11 | 2.64 | 0.31 |
| | Orchard | 1.66 | 2.16 | 3.6 | 0.11 | 2.45 | 0.26 |
| | Ivy | 1.30 | 2.15 | 2.8 | 0.11 | 2.17 | 0.28 |
| | Shoeground | 1.83 | 2.34 | 4.3 | 0.16 | 2.68 | 0.27 |
| | Copse | 1.23 | 2.11 | 2.6 | 0.16 | 1.91 | 0.31 |
| P - On-line | Avenue | 1.74 | 2.18 | 3.8 | 0.12 | 2.51 | 0.31 |
| | Orchard | 1.75 | 2.05 | 3.6 | 0.11 | 2.36 | 0.26 |
| | Ivy | 1.40 | 2.03 | 2.8 | 0.13 | 2.08 | 0.28 |
| | Shoeground | 1.90 | 2.24 | 4.3 | 0.10 | 2.58 | 0.27 |
| | Copse | 1.29 | 2.01 | 2.6 | 0.16 | 1.86 | 0.31 |
| P – Processed | Avenue | 1.32 | 2.86 | 3.8 | 0.1 | 2.96 | 0.31 |
| | Orchard | 1.16 | 3.08 | 3.6 | 0.08 | 3.15 | 0.26 |
| | Ivy | 0.88 | 3.17 | 2.8 | 0.08 | 3.27 | 0.28 |
| | Shoeground | 1.60 | 2.66 | 4.3 | 0.09 | 2.85 | 0.27 |
| | Copse | 0.76 | 3.39 | 2.6 | 0.08 | 3.45 | 0.31 |

RMSEP root mean square error of prediction.

RPD residual prediction deviation (Standard deviation (SD) / RMSEP).

**Table 4**

**Results of univariate analyses of variance (ANOVA) performed on the prediction set to evaluate individual and interaction effects of moisture content (MC) and texture on prediction accuracy of organic carbon (OC) and total nitrogen (TN) evaluated as residual prediction deviation (RPD), and root mean square error of prediction (RMSEP) for fresh soil samples**

| | | *OC* | | | *TN* | | |
|---|---|---|---|---|---|---|---|
| | Source | MS | F -ratio | p-value | MS | F -ratio | p-value |
| RPD | MC | 737.19 | 37.10 | 0.0002 | 722.67 | 36.28 | 0.0003 |
| | Clay | 2655.92 | 13.98 | 0.005 | 2163.83 | 8.455 | 0.027 |
| | sand | 3672.97 | 13.96 | 0.005 | 2628.28 | 13.83 | 0.005 |
| | silt | 1262.92 | 171.48 | 0.008 | 1243.89 | 167.78 | 0.009 |
| Interaction | MC*Texture | 1107.96 | 4.65 | 0.01 | 1098.59 | 4.6 | 0.01 |
| $R^2$ | MC | 768.57 | 33.87 | 0.0002 | 894.36 | 43.74 | 0.0001 |
| | Clay | 2647.24 | 16.54 | 0.003 | 3251.32 | 16.23 | 0.006 |
| | sand | 3823.16 | 14.44 | 0.005 | 4675.2 | 15.22 | 0.004 |
| | Silt | 1325.95 | 180.93 | 0.008 | 1123.26 | 212.63 | 0.05 |
| Interaction | MC*Texture | 1124.77 | 4.34 | 0.01 | 1234.12 | 5.12 | 0.009 |
| RMSEP | MC | 794.77 | 39.94 | 0.0002 | 925.63 | 46.60 | 0.0001 |
| | Clay | 2764.23 | 14.54 | 0.005 | 3003.98 | 15.81 | 0.004 |
| | sand | 3800.16 | 14.44 | 0.005 | 4080.4 | 15.51 | 0.004 |
| | Silt | 1337.95 | 180.93 | 0.03 | 1506.26 | 204.63 | 0.05 |
| Interaction | MC*Texture | 1144.77 | 4.80 | 0.009 | 1226.69 | 5.15 | 0.007 |

MS: mean square
$R^2$: coefficient of determination

**Table 5**

**Results of univariate analyses of variance (ANOVA) performed on the prediction set to evaluate individual and interaction effects of moisture content (MC) and texture on prediction accuracy of organic carbon (OC) and total nitrogen (TN) evaluated as residual prediction deviation (RPD), and root mean square error of prediction (RMSEP) for on-line measurement**

| *Effect* | | *OC* | | | *TN* | | |
|---|---|---|---|---|---|---|---|
| | Source | MS | F -ratio | p-value | MS | F -ratio | p-value |
| RPD | MC | 745.63 | 37.53 | 0.0003 | 730.51 | 36.69 | 0.0003 |
| | Clay | 2671.91 | 14.06 | 0.006 | 2643.22 | 13.91 | 0.005 |
| | sand | 3691.77 | 14.03 | 0.005 | 3658.03 | 13.90 | 0.006 |
| | silt | 1273.96 | 172.96 | 0.02 | 1254.17 | 169.34 | 0.05 |
| Interaction | MC*Texture | 1113.39 | 4.67 | 0.01 | 1103.65 | 4.63 | 0.01 |
| $R^2$ | MC | 723.12 | 32.33 | 0.0001 | 834.23 | 43.74 | 0.0001 |
| | Clay | 2234.24 | 12.54 | 0.004 | 3654.11 | 16.21 | 0.003 |
| | sand | 3673.16 | 14.44 | 0.006 | 4945.2 | 15.67 | 0.006 |
| | Silt | 1123.95 | 183.93 | 0.009 | 1247.26 | 212.38 | 0.04 |
| Interaction | MC*Texture | 1865.77 | 4.34 | 0.01 | 1234.12 | 5.67 | 0.007 |
| RMSEP | MC | 788.18 | 39.61 | 0.0002 | 926.21 | 46.63 | 0.0001 |
| | Clay | 2751.94 | 14.48 | 0.005 | 3005.02 | 15.81 | 0.004 |
| | sand | 3785.74 | 14.39 | 0.005 | 4081.61 | 15.51 | 0.004 |
| | Silt | 1329.40 | 179.80 | 0.04 | 1507.00 | 204.73 | 0.01 |
| Interaction | MC*Texture | 1140.594 | 4.78 | 0.009 | 1140.59 | 4.78 | 0.009 |

MS: mean square

$R^2$: coefficient of determination

593 **Figure captions**

594

595 *Figure 1. Location of five fields on Silsoe Farm, UK measured in summer 2009 with the on-*
596 *line visible and near infrared (vis-NIR) sensor*

597 *Figure 2. On-line visible and near infrared (vis-NIR) sensor developed by Mouazen (2006)*

598 *Figure 3. A flow diagram explaining different steps performed during the study*

599 *Figure 4. (a) Principal component analysis (PCA) similarity maps determined by principal*
600 *components 1 (PC1) and 2 (PC2) for fresh soil samples from Avenue (solid fill round),*
601 *Orchard (square ), Ivy Ground (triangle), Shoeground (cross) and Copse (no fill round)*
602 *fields; (b) spectral patterns 1 (-) and 2 (---) obtained from PCA on the raw visible and near*
603 *infrared (vis-NIR) soil spectra of fresh samples collected from the five fields*

604 *Figure 5. The combined effect of moisture content (MC) and texture on soil spectra of*
605 *processed (top) and fresh (bottom) samples of Avenue (Av), Shoeground (Sh), Orchad (Or),*
606 *Ivy (Iv) and Copse (Co) fields*

607 *Figure 6. Mean values of residual prediction deviation (RPD) of three replicates for the*
608 *prediction of soil organic carbon (OC) in five fields based on spectra of prediction sets*
609 *collected in the field with the on-line (white) sensor, and in the laboratory for fresh (grey),*
610 *and processed soil samples (black). The triangles indicate the moisture content of the fresh*
611 *samples from each field*

612 *Figure 7. Mean values of residual prediction deviation (RPD) of three replicates for the*
613 *prediction of soil total nitrogen (TN) in five fields based on spectra of prediction sets*
614 *collected in the field with the on-line (white) sensor, and in the laboratory for fresh (grey),*

*and processed soil samples (black). The triangles indicate the moisture content of the fresh samples from each field*