Soil and Tillage Research, Volume 128, April 2013, Pages 125–136

Effect of spiking strategy and ratio on calibration of on-line visible and near infrared soil sensor for measurement in European farms

Boyan Kuang & Abdul Mounem Mouazen*

Environmental Technology and Science Department, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom, e-mail of corresponding author: a.mouazen@cranfield.ac.uk

Tel: +44-01234750111. FAX: +44-01234752971

Abstract

A previously developed on-line visible and near infrared (vis-NIR) spectroscopy-based soil measurement system was implemented for the measurement of soil organic carbon (OC), total nitrogen (TN) and moisture content (MC) in three fields at three European farms. The on-line sensor platform was coupled with a mobile, fibre type, vis-NIR spectrophotometer (AgroSpec from tec5 Technology for Spectroscopy, Germany), with a measurement range of 305 - 2200 nm, to acquire soil spectra in diffuse reflectance mode. A general calibration set of 425 soil samples, spiked with different number of spectra from the three validation fields were used to establish calibration models for the studied soil properties using partial least squares (PLS) regression analysis. Different spiking strategies and spiking ratios were investigated and results revealed that the best prediction accuracy was obtained after 20% spiking ratio with samples whose spectra were measured in the laboratory. Evaluated by the values of residual prediction deviation (RPD), which is the ratio of standard deviation to root mean square error of prediction (RMSEP), the accuracy of the on-line measurement

was classified as excellent for MC (RPD = 2.76 - 3.96), good to very good for OC (RPD = 1.88 - 2.38) and good to excellent for TN (RPD = 1.96 - 2.52). Reducing the number of samples used for spiking resulted in deteriorating the prediction accuracy, although 1-2 samples per ha were found to provide good predictions. There was a distinguishable spatial similarity between the on-line and laboratory measured maps for all studied properties, although the full-data point maps provided more detailed information about the spatial variation. This confirms that the on-line vis-NIR soil sensor provides correct and detailed information about soil OC, TN and MC at high sampling resolutions.

Keywords: on-line measurement, soil properties, spiking, European farms, accuracy.

1. Introduction

One of the strategies used in precision agriculture (PA) is to minimize the production costs through enhanced efficiency, which can increase profits and conserve the environment. It is proved that PA is an environmental friendly strategy, in which farmers can vary the use of inputs to respond to variable soil and crop conditions within a field (Srinivasan, 2006). Conventional determination of soil spatial variability usually involves manual soil sampling, sample pre-treatment, laboratory reference analyses and mapping. This procedure is very expensive, time consuming and provides scattered measurement points. Therefore, the development of a fast, robust, cost effective and environmental friendly detecting method of the soil spatial variability is a preliminary task for the implementation of PA.

47 Visible and near infrared (vis-NIR) spectroscopy became recently a proven technique for
48 fast, cost effective and high resolution data collection on soil properties (Viscarra Rossel &
49 McBratney, 1998; Shepherd & Walsh, 2002; Mouazen et al, 2010; Stenberg et al., 2010;).
50 In comparison to non-mobile analysis, there is considerably less literature available about
51 on-line vis-NIR spectroscopy analysis of soil properties. Generally, both the laboratory and
52 *in situ* non-mobile vis-NIR methods provide better accuracy than the on-line method
53 (Kuang et al., 2012). This might be attributed to other factors influencing the latter method
54 only, such as noise associated with tractor vibration, sensor-to-soil distance variation
55 (Mouazen et al., 2009), stones and plant debris and difficulties of matching the position of
56 soil samples collected for validation with corresponding spectra collected from the same
57 position (Mouazen et al., 2007). A review on the current status of on-line vis-NIR
58 measurement systems confirms that only three systems are available today (Shibusawa et
59 al., 2001; Mouazen et al., 2005; Christy et al., 2008). The beginning of these systems dates
60 back to 1991, when Shonk et al., (1991) developed a system to measure soil organic matter
61 (SOM) and moisture content (MC), which utilised a single wavelength (660 nm) of light,
62 reporting a determination coefficient (R^2) of 0.83 for SOM. Shibusawa et al., (2001)
63 developed an on-line vis-NIR (400 – 1700 nm) sensor to predict organic carbon (OC), MC,
64 pH and NO3-N. Although this system is highly technically instrumented, it is rather
65 expensive. Christy et al., (2008) developed a prototype soil reflectance mapping unit
66 equipped with a vis-NIR spectrophotometer, which is commercially available in the market
67 today. The sapphire glass of the optical probe makes direct contact with soil and stones.
68 They have reported that OC can be successfully measured with a RMSEP of 3.0 gkg^{-1} in a
69 high OC variability area (standard deviation = 5.1 gkg^{-1} and range = 3.0-26.3 gkg^{-1}). A

simpler design to the one of Shibusawa et al., (2001) without sapphire window optical configuration was developed by Mouazen (2006). So far, the system provided variable degrees of success for the measurement of MC, total nitrogen (TN), total carbon (TC), pH and available P in different soils in Belgium and northern France (Mouazen et al., 2005; Mouazen et al., 2007; Mouazen et al., 2009). More recently, employing the on-line system developed by Christy et al. (2008), Bricklemyer et al. (2010) reported on the on-line measurement for soil OC and clay content. Although authors did not report quantitative estimation of accuracy for OC, they calculated a standard error of prediction (SEP) of 3.4 gkg^{-1} and a RPD value of 1.4 for clay content. Using the same system, coupled with topography and aerial photograph data, Munoz & Kravchenko (2011) reported low to moderate accuracy of soil OC measurement with R^2 and RMSEP ranged from 0.44 to 0.66 and from 1.41 to 1.51 gkg^{-1}, respectively. Applying this vis-NIR system in combination with electrical conductivity (EC) and temperature sensors in a Danish field, Knadel et al. (2011) obtained moderate prediction accuracy (RMSEP = 59.4 gkg^{-1} and RPD = 2.3) for soil OC. However, the vis-NIR sensor alone only achieved a relatively low accuracy (RMSEP = 59.8 gkg^{-1}, RPD = 1.9). This brief review reveals that the existing on-line sensors do not provide sufficiently accuracy and stability to recommend them for site specific application of different inputs. The variable degrees of performance of these on-line sensors might be attributed to the fluctuation in model performance, with the majority of them established for field-scale analysis (e.g. Shibusawa et al., 2001; Mouazen et al., 2005; Christy et al., 2008; Munoz & Kravchenko, 2011), or for regional- and country-scale analysis (e.g. Mouazen et al., 2007; Mouazen et al., 2009). Some studies suggested spiking local samples into the general calibration models can improve prediction accuracy under

93 laboratory or *in situ* measurement conditions (Shepherd & Walsh, 2002; Janik et al., 2007;
94 Waiser et al. 2007; Sankey et al. 2008; Viscarra Rossel et al. 2008; Minasny et al. 2009;
95 Wetterlind and Stenberg, 2010; Guerrero et al. 2010). As yet, no general calibration
96 procedure that included spiking strategy have been reported for establishing calibration
97 models of on-line sensors, which has led to stable performance of on-line sensors, at the
98 lowest model production cost.
99 This paper aims at reporting on a methodology for the calibration of a vis-NIR on-line
100 measurement system (Mouazen, 2006), including spiking concept for automatic data
101 collection of OC, TN and MC at farm scale in three European farms.

102

103 **2. Material and methods**

104

105 ***2.1 Soil samples and laboratory analyses***
106

107 A total of 425 soil samples were used as a general dataset to be used to establish calibration
108 models for the measurement of OC, TN and MC in European soils (Table 1). They were
109 collected from four different farms in Europe, namely Mespol Medlov, A.S. (Czech
110 Republic, http://farmsubsidy.org/CZ), Wimex (Germany, http://www.wimex-online.de),
111 Bramstrup Estate (Denmark, http://www.bramstrup.dk) and Silsoe Farm (The UK). Bulked
112 samples from 16 cores were collected from the upper soil layer (0-30 cm) in the spring of
113 2008 (Czech Republic and Germany), spring of 2009 (Denmark) and summer of 2009 (UK)
114 and covered diverse soil conditions (Kuang and Mouazen, 2011). A total of 128, 97 and 48

115 soil samples were, respectively, collected from seven fields in Mespol Medlov, A.S. farm,
116 six fields in Bramstrup Estate farm and two fields in Silsoe farm. Further 152 soil samples
117 were collected from four sub-areas belong to Wimex farm, with four samples from two
118 fields at Reppichau, 50 samples from nine fields at RAG, 20 samples from ten fields at
119 Aken and 54 samples from fifteen fields at the Wulfen farm.

120 Apart from these soil samples, another 113 soil samples were collected from three selected
121 fields, where on-line measurement was carried out, namely, in Mespol Medlov,
122 Bramstrump Estate and Silsoe farms (Table 2). These were considered as validation fields.
123 They were collected during the on-line measurement from the bottom of trenches at a depth
124 of 15 cm. The validation field in Silsoe farm is 500 m away from the nearest fields, where
125 samples used for general data set were collected, whereas the validation fields in Mespol
126 Medlov and Bramstrump Estate farms are 2.5 km 3 km far from the nearest fields,
127 respectively. A principal component analysis (PCA), performed on 425 samples of general
128 calibration dataset and 113 validation data set shows clear separation between validation
129 samples of each individual farm from the general dataset samples (Fig. 1).

130 Around 200 g of soil from each sample was kept deep frozen (-18 °C) until analysis. After
131 careful mixing, half of each sample from Czech Republic, Germany and Denmark was sent
132 to Leibniz Centre for Agricultural Landscape Research (ZALF) in Germany for soil
133 chemical analyses for TN and OC and the second half was sent to Cranfield University for
134 optical measurement and data analysis. Samples collected from the UK and 113 soil
135 samples collected during the on-line measurement in 2010, were subjected to both optical
136 and chemical measurement at Cranfield University. Sample statistics of laboratory

137 reference measurements is summarised in Table 3 for the calibration and validation
138 datasets.

139

140 ***2.2 On-line measurement***

141

142 The on-line measurement system designed and developed by Mouazen (2006) was used
143 (Fig. 2) to measure three fields in the three European farms. It consists of a subsoiler, which
144 penetrates the soil to the required depth, making a trench, whose bottom is smoothened by
145 the downwards forces acting on the subsoiler. The optical probe is housed in a steel lens
146 holder. This is attached to the backside of the subsoiler chisel in order to acquire soil
147 spectral reflectance data from the smooth bottom of the trench. The subsoiler was
148 retrofitted with the optical unit and attached to a frame. This was mounted onto the three
149 point linkage of the tractor (Mouazen et al., 2005a). An AgroSpec mobile, fibre type, vis-
150 NIR spectrophotometer (Tec5 Technology for Spectroscopy, Germany) with a
151 measurement range of 305-2200 nm was used to measure soil spectra in diffuse reflectance
152 mode. The spectrometer was an IP 66 model, protected for harsh working environments. A
153 deferential global positioning system (DGPS) (EZ-Guide 250, Trimble, USA) was used to
154 record the position of on-line measured spectra with sub-meter accuracy. A Panasonic
155 semi-rugged laptop was used for data logging and communication. The spectrometer
156 system, laptop and DGPS were powered by the tractor battery.
157 Three fields were measured in summer 2010, namely, in Mespol Medlov, A.S. farm in
158 Czech Republic, Bramstrup Estate farm in Denmark and Silsoe farm in the UK. In each

159 field, blocks of 150 m by 150 m, covering about 2 ha of land were measured. Each
160 measured line was 150 m long with 10 m intervals between adjacent transects (Fig. 3). The
161 travel speed of the tractor was around 2 km/h and the measurement depth was set at 15 cm.
162 During the measurement at each line, 2 or 3 soil samples were collected from the bottom of
163 the trench and the sampling positions were carefully recorded with a DGPS. Each of those
164 samples was equally divided into two parts. One half used to carry out the laboratory
165 reference measurements of soil OC, TN and MC and the other half used for optical
166 scanning.
167
168 ***2.3 Laboratory reference analyses***
169
170 Soil OC and TN were measured by a TrusSpecCNS spectrometer (LECO Corporation, St.
171 Joseph, MI, USA), using the Dumas combustion method. Soil MC was determined by oven
172 drying of the soil samples at 105 ºC for 24 h. The results of the textural analysis of the three
173 on-line measured fields were based on a mixed sample from each field. Each sample was
174 subjected to wet sieving and a hydrometer test in order to determine the particle size
175 distribution. Texture classes were determined according to the United State Department of
176 Agriculture (USDA) classification system (Table 2).
177
178 ***2.4 Optical measurement***
179

Each soil sample was put into a glass container and mixed well. Big stones and plant residue were removed (Mouazen et al., 2007). Soil from each sample was placed into three Petri dishes, which were 2 cm deep and 2 cm in diameter. The soil in the Petri dish was shaken and pressed gently before levelling with a spatula. A smooth soil surface ensures maximum light reflection and a high signal to noise ratio (Mouazen et al., 2007). The soil samples were scanned by the same AgroSpec portable spectrophotometer (Tec5 Technology for Spectroscopy, Germany), used during the on-line measurement. A 100 % white reference was used before scanning. A total of 10 scans were collected from each cup and these were averaged in one spectrum.

2.5 Sample pre-treatment and development of calibration models

2.5.1 Pre-treatment of spectra

The spectral range of the soil spectra was first reduced to 371 - 2150 nm to eliminate the noise at both edges of each spectrum. The number of wavelengths was then reduced by averaging three successive wavelengths in the visible range, and 15 points in the NIR range. The Savitzky-Golay smoothing, maximum normalisation and first derivation (Martens & Naes, 1989) were successively implemented using Unscrambler 7.8 software (Camo Inc.; Oslo, Norway). The pre-treated spectra and the laboratory chemical measurement values were used to develop the calibration models.

2.5.2 Spiking

The purpose of spiking is to introduce variability of a new set of data into a general calibration dataset. Three different spiking strategies of soil samples collected from the target validation fields were selected. These included - no spiking of soil samples, - spiking with laboratory non-mobile scanned spectra and - spiking with on-line (mobile) spectra recorded during the on-line measurement. When spiking was considered with the latter 2 cases, 20% of a randomly selected set of samples (21 samples from the three fields) from the total 113 soil samples collected during the on-line measurement were spiked into the general data set of 425 samples, collected from the 4 European farms (see also 2.1). Furthermore, smaller ratios of spiking samples of 5% (8 samples) and 10% (12 samples) were randomly selected from the laboratory measured spectra only, and were used as spiking sets to envisage the influence of different spiking ratios (sample numbers) on the on-line prediction accuracy. By adopting these spiking strategies and ratios we obtained five calibration sets (Tables 4), namely, without spiking, 5%, 10% and 20% spiking ratios of laboratory measured spectra and 20% spiking ratio of on-line measured spectra.

2.5.3 Development of calibration models

The five calibration sets were subjected to a partial least squares (PLS) regression analysis with the leave-one-out cross validation using an Unscrambler 7.8 software (Camo Inc.; Oslo, Norway), which resulted in five groups of PLS models for each soil property (Table 4). For all five spiking strategies and ratios the same on-line data set of 80 %, was used as

the validation set for the three soil properties in each field. This resulted in three validation sets of 39, 30 and 23 samples in Czech Republic, Denmark and UK validation fields, respectively (Table 5). These three data sets were used to validate the on-line measurement and to develop maps comparing the laboratory reference with the corresponding on-line measured values.

The number of latent variables for a model was determined by examining a plot of the leave-one-out cross-validation residual variance against the number of latent variables obtained from the PLS. The latent variable of the first minimum value of residual variance was selected. Outliers were detected using the residual sample variance plot after PLS. Samples located far from the zero line of residual variance were considered outliers and excluded from the cross-validation sample set.

The prediction performance of the PLS models was evaluated by means of root mean square error of prediction (RMSEP) and the rresidual prediction deviation (RPD), which is the ratio of standard deviation (SD) of the measured values to RMSEP. Viscarra Rossel et al. (2006) classified RPD values as follows: RPD<1.0 indicates very poor model/predictions and their use is not recommended; RPD between 1.0 and 1.4 indicates poor model/predictions where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates fair model/predictions which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good model/predictions where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates very good, quantitative model/predictions, and RPD>2.5 indicates excellent model/predictions. This classification system was adopted in this study.

2.6 Development of soil maps

Two types of maps were developed, namely, comparison and full-data points maps. The comparison maps were developed to compare on-line predicted with laboratory reference measurement of a soil property based on randomly selected points in the field. This comparison also included maps of predicted values based on laboratory scanned spectra of the same randomly selected samples. The full-point maps were developed using all on-line predicted data points (~1500 - 2000 point per ha). ArcGis 10 (ESRI, USA) software was used to generate the comparison maps, using the inverse distance weighing (IDW) interpolation methods. To produce the latter maps, Vesper 1.6 software, developed by Australian Centre for Precision Agriculture, was used to develop semivariogram models for OC, TN and MC using the entire field on-line data. Based on semivariogram parameters and kriging interpolation method, ArcGis 10 (ESRI, USA) was used to produce the full-data point maps.

A flow diagram illustrating the different steps taken into account in this study, from the collection of soil sample to the development of different maps is shown in Fig. (4). This Figure illustrates clearly that the vis–NIR calibration models were developed using soil spectra scanned under non-mobile laboratory environment, except for spiking with 20% on-line measured spectra.

3. Results and discussion

3.1 Accuracy of calibration models in cross-validation

Based on spiking strategies and ratio of samples spiked in the general calibration set, the accuracy of calibration models achieved based on cross-validation proves promising results (Table 4). Examining RPD values reveals that the models with 20 % spiking ratio of laboratory-scanned spectra were the most accurate with RPD values of 2.82, 2.61 and 4.28 for OC, TN and MC, respectively. Similar results were reported by Kuang & Mouazen (2011) but for a larger range (350-2500 nm) vis-NIR spectrophotometer. Furthermore, RPD values of all models were equal or larger than 2.5. According to Viscarra Rossel et al. (2006), an RPD values above 2.5 indicates excellent prediction results. Results also show that models without spiking provide smaller accuracy (larger RMSEP and smaller RPD values) than models with spiking for all spiking strategies and ratios (Table 4), except for models with 20% spiking ratio of on-line samples for OC and TN.

3.2 Validation of on-line measurement

The same spectra pre-treatment was used for the on-line collected spectra as that used for the development of calibration models. The on-line collected spectra after spectra pre-treatment were used to predict MC, TN and OC using calibration models developed in advance, as explained above. The laboratory reference measurement values were compared with the on-line predicted concentration values at the same positions. Table 5 summarises the accuracy of the on-line measurement for studied soil properties based on different spiking strategies (no spiking, spiking with laboratory or on-line collected spectra) and spiking ratios. Table 5 reveals that RPD values were above 2 for all soil properties in all

294 fields, except for 3 cases. Out of these cases, RPD in one case was 1.98 for OC in the field
295 in Denmark. The other two cases (OC and TN in the Danish and the UK fields,
296 respectively) are for models developed based on 20% spiking ratio with on-line collected
297 spectra. Adopting Viscarra Rossel et al. (2006) classification system for the prediction
298 accuracy reveals that the on-line prediction of OC is good/very good and good/excellent for
299 TN and is excellent for MC, for all modelling strategies without and with spiking (Table 5).
300 As SOM or OC are essential for soil management and carbon sequestration, they were the
301 main properties considered for on-line vis-NIR measurement (Shonk et al., 1991; Hummel
302 et al., 2001; Shibusawa et al., 2001; Mouazen et al., 2007; Christy et al., 2008; Bricklemyer
303 et al., 2010). Although there are some significant absorbance peaks associated with C=O,
304 C–H + C–H and C–H + C–C overtones and combinations in the NIR spectral ranges, only
305 few moderate successful cases for on-line measurement have been reported so far (Hummel
306 et al., 2001; Christy et al., 2008). This might be attributed to the several affecting factors
307 during on-line measurement (Mouazen et al., 2007; Mouazen et al., 2009; Stenberg et al.,
308 2010). Compared to the on-line sensing of OC, there are even fewer studies on TN can be
309 found in the literature (Christy et al., 2008; Mouazen et al., 2007). Only Christy et al.,
310 (2008) archived high accuracy for TN (R^2 = 0.86) for a field scale calibration. This was not
311 confirmed stable for different fields, as reported in the current study where on-line
312 validation is classified as good to excellent in the three measured fields (Table 5). Actually,
313 there is high correlation between OC and TN, because the N content in soil is almost
314 entirely dependent on SOM content, and the overall TN:OC ratio is 1:10 (Martin et al.,
315 2002). This high correlation explains the successful measurement of TN, although no direct
316 spectral response exists for TN.

317 Soil MC is the most successfully on-line measured soil property and higher accuracies as
318 compared to OC and TN were reported in the literature (Hummel et al., 2001; Mouazen et
319 al., 2005; Christy et al., 2008). This is due to significant O-H absorbance peaks at 950,
320 1450 and 1950 nm overtones. Although the range of MC was relative narrow in the current
321 study (SD = 12.8 - 23.4 gkg^{-1}) (Table 5), the RMSEP for on-line prediction after 20%
322 spiking with laboratory scanned spectra is very small (3.23 gkg^{-1} - 7.40 gkg^{-1}), proving the
323 excellent performance of the sensor for on-line sensing of MC.
324
325 ***3.3 Soil maps***
326
327 ***3.3.1 Comparison maps***
328
329 Figure 5 compares maps of on-line spectra-based predicted values (e.g. models of 20%
330 spiking ratio with laboratory scanned samples), laboratory spectra-based predicted values
331 and laboratory reference measured values for OC (5a), TN (5b) and MC (5c) in the field of
332 Bramstrup Estate farm in Denmark, shown as an example. In order to allow for meaningful
333 comparisons between reference and on-line measured maps, the same number of classes (7
334 classes) was considered for all maps with each class has identical range in the three maps
335 (Mouazen et al., 2007). A comparison between maps of measured and predicted soil
336 properties investigated shows large spatial similarity, with high and low zones match
337 almost perfectly. No spatial differences can be observed between validation maps
338 developed with on-line measured spectra and the corresponding maps developed with

laboratory (non-mobile) measured soil spectra. This proves the high quality of on-line measured spectra, which reflects the sensor stability and robustness during on-line measurement. Figure 6 shows the error maps between laboratory reference measurement and on-line prediction values of OC, TN and MC in the Denmark field. It can be observed that the largest error in the three maps occur at the boarder of each field, which might be attributed to errors associated with placing or removing the optical probe in the soil, respectively, at the start or end of some measured lines. Another reason might be the irregularity of the soil surface at the field circumference.

3.3.2 Full-data point maps

The semivarigram parameters of the three properties in the Denmark field are summarised in Table 6. Maps show high spatial variability of the three soil properties, although the field area is remarkably small of 2 ha (Fig. 7). This high variability encourages the need for on-line soil sensor for the characterisation of within field spatial variability of soil properties, as zones with different levels of concentration should be managed differently in PA, particularly for site specific fertilisation. In particular, high similarity between OC and TN maps can be observed, which can be attributed to the high correlation exist between OC and TN in the soil. Comparing between laboratory reference analyses maps produced based on few measurement points (Figure 5) with corresponding full-data point maps based all on on-line measurement points (Figure 7), more detailed characterisation of within field

variation of the three soil properties can be observed in the latter maps. This detailed variation is essential for precise management of input applied to the soil-crop system.

3.4 Effect of spiking strategy and spiked sample number

Comparing the effect of different spiking strategies and ratios, revealed that the best accuracy of on-line prediction of the three soil properties is achieved with models developed after 20% spiking ratio with laboratory scanned spectra. However, it can be clearly seen that even the smallest 5% spiking ratio with laboratory scanned spectra can improve the on-line prediction for all properties in all fields, comparing to on-line prediction using models without spiking. Furthermore, the prediction accuracy decreases as the number of spiked samples decreases and vice versa. However, for spiking with on-line measured spectra (e.g. 20% spiking ratio), the prediction accuracies are smaller than the corresponding accuracies obtained from models without spiking. This may suggest that spiking with on-line collected spectra is not the best strategy to follow. This might be attributed to difficulties associated with precision of matching sampling position of on-line collected spectra with corresponding soil sample used for laboratory reference analysis, which is an issue highlighted by Mouazen et al. (2007).

In order to investigate the effect of sample number per hectare spiked into the general calibration set on prediction accuracy, variations in RPD values obtained from model validation of on-line set versus sample number spiked in the general data set (425 samples)

381 is shown in Fig. (8). Clear increases in RPD values with spiked sample number per ha can
382 be observed. However, the degree of increase differs according to soil property considered,
383 with larger increases is observed for OC and TN, and smaller increases for MC. This is also
384 affected by the field measured. On the basis of average values of the RPD of the three
385 fields, an increase in spiked sample number from 1/1.5 to 3.5/4.5 per ha, leads to an
386 average increase in % RPD of 9.1% and 11.1% for OC and TN, respectively, whereas an
387 increase of 10.4% is observed for MC. This quantitative evaluation of % difference of RPD
388 values may suggest that spiking of laboratory scanned spectra with a sampling rate of 1 to 2
389 samples per ha is sufficient to obtain accurate on-line prediction of soil properties. Figure
390 (8) proves that spiking with 1-2 sample per ha results in a very good model performance for
391 OC (RPD = 2.13-2.32) and TN (RPD = 2.15-2.27) and excellent model performance for
392 MC (RPD = 2.91-3.62). However, this number of samples is only valid for the degree of
393 spatial variation similar to those encountered in the three validation fields of this study.
394 This number might increase or decrease for other study cases, if the spatial variation is
395 larger or smaller, respectively, as compared to those of the fields of the current study. This
396 sampling rate is almost identical to the sampling rate considered for conventional analysis
397 of soil properties, adopted today by laboratories to provide fertilisation recommendations
398 (Mouazen et al., 2007 & 2009). This reveals that with only slight decrease in accuracy, the
399 consumable cost of on-line measurement of OC, TN and MC would be approximately
400 identical to that of laboratory conventional methods, after excluding the costs of on-line
401 sensor and general dataset. However, traditional laboratories provide fertilisation
402 recommendations based on 1 sample per ha, where the spatial variation in soil properties is
403 ignored. Furthermore, the on-line measurement system enables the collection of high

404 number of data points (around 1500 - 2000 readings per ha), with an average of 2 points per
405 meter travel distance. Assuming that laboratory methods of soil analysis do not prone to
406 measurement error, which is not the case in practice, the detailed information about the
407 spatial variation in soil properties obtained with the on-line sensor, even with a smaller
408 accuracy than that of the laboratory methods can be considered a pronounced advancement
409 in the field of proximal soil sensing. In addition to the high accuracy achieved for the
410 measurement of the three properties, the performance illustrates stability (Table 5) among
411 the three fields, which is the ultimate objective of the vis-NIR calibration of soil properties,
412 prone to fluctuation in performance among different sites.
413 Considering the wide geographical and texture variation of the three measured fields (Table
414 2), the concept of general data set spiked with a small number of samples (1-2 samples per
415 ha) from each new measured field seems to be a successfully calibration procedure for on-
416 line vis-NIR measurement of soil OC, TN and MC. Generally, the general model concept
417 proposed by Mouazen et al (2007) is based on a sufficient number of soil samples to
418 account for the soil spatial variability in a new target site, where the prediction will be
419 carried out (Viscarra Rossel et al., 2008; Guerrero et al., 2010). Spiking local (target site)
420 soil samples into global or continental models for non-mobile calibrations proved to be an
421 efficient way to improve the prediction accuracy of target field for some soil constituents
422 (Shepherd & Walsh, 2002; Janik et al., 2007; Waiser et al. 2007; Guerrero et al., 2010;
423 Minasny et al. 2009; Viscarra Rossel et al. 2009; Wetterlind and Stenberg, 2010). The
424 successful implementation of spiking general calibration models with field specific samples
425 confirms the sample spiking technique is successfully applied for on-line vis-NIR

426 spectroscopy sensing, similar to that of laboratory and *in situ* non-mobile vis-NIR
427 spectroscopy.
428

429 **4. Conclusions**

430

431 This paper reports on the performance of vis-NIR spectroscopy-based calibration models
432 spiked with different sampling strategies and ratios for the on-line prediction of soil TN,
433 OC and MC. The results reported allow the following conclusions to be drawn:

434 1- The on-line measurement system is robust that enables the collection of data on
435 several soil properties, simultaneously.
436 2- This system enables the collection of large data points per field (around 1500-2000
437 readings per ha). The large number of points will open new possibilities for the
438 management of soil-water-plant system, which was not achievable so far.
439 3- The accuracy of on-line prediction of OC, TN and MC was classified as good to
440 excellent prediction performance with RPD values range between 1.88 and 3.96.
441 4- The concept of general calibration models, spiked with 1-2 samples per ha from
442 new measured fields is a successful procedure for the calibration of the on-line vis-
443 NIR sensor. However, a smaller or larger number of spiking samples might be
444 required for fields with a smaller or larger spatial variability, respectively.

445 Further work is being undertaken to improve the prediction accuracy of the system by
446 accounting for the effect of MC and texture on the result obtained. The concept of spiking

447 of general calibration models needs to be tested for other soil properties than those reported
448 in this study.

449

450 **Acknowledgement**

451

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

455

456 **References:**

457

458 Bricklemyer, R. S., Brown, D. J. 2010. On-the-go VisNIR: Potential and limitations for
459 mapping soil clay and organic carbon. Computers and Electronics in Agriculture, 70,
460 209-216.
461 Christy, C. D. 2008. Real-time measurement of soil attributes using on-the-go near infrared
462 reflectance spectroscopy. Comput. Electron. Agric. 61, 10–19.
463 Guerrero, C., Zornoza, R., Gómez, I. & Mataix-Beneyto, J. 2010. Spiking of NIR regional
464 models using samples from target sites: Effect of model size on prediction accuracy.
465 Geoderma, 158, 66-77.

466 Hummel, J. W., Sudduth, K. A., and Hollinger, S. E. 2001. Soil moisture and organic
467 matter prediction of surface and subsurface soils using an NIR soil sensor.
468 Comput.Electron. Agric., 32, 149–165.
469 Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R. 2007. The prediction of soil
470 carbon fractions using mid-infrared-partial least squares analysis. Australian Journal of
471 Soil Science, 45, 73–81.
472 Knadel, M., Thomsen, A., Greve, M. H. 2011. Multisensor on-the-go mapping of soil
473 organic carbon content. Soil Sci. Soc. Am. J. 75:1799–1806.

474 Kuang, B., Mahmood, H. S., Quraishi, Z., Hoogmoed, W. B., Mouazen, A. M., van Henten,
475 E. J., 2012. Sensing soil properties in the laboratory, *in situ* and on-line – a review.
476 Advances in Agronomy. Volume 114.
477 Kuang, B & Mouazen, A. M. 2011. Calibration of visible and near infrared spectroscopy
478 for soil analysis at the field scale on three European farms. Eur. J. Soil Sci., 62, 629-636
479 Martens, H., Naes, T., 1989. Multivariate Calibration, 2_{nd} edition. John Wiley & Sons, Ltd.,
480 Chichester, United Kingdom.
481 Martin, P.D., Malley, D.F., Manning, G. & Fuller, L. 2002. Determination of soil organic
482 carbon and nitrogen at the field level using near-infrared spectroscopy. Canadian
483 Journal of Soil Science, **82**, 413-422.

484 Minasny, B., McBratney, A. B., Pichon, L., Sun, W., and Short, M. G. 2009. Evaluating
485 near infrared spectroscopy for field prediction of soil properties. Australian Journal of
486 Soil Research, 47, 664–673

487 Mouazen, A.M., De Baerdemaeker, J., Ramon, H. 2005. Towards development of on-line
488 soil moisture content sensor using a fibre-type NIR spectrophotometer. Soil & Tillage
489 Research, 80, 171-183.
490 Mouazen, A.M., 2006. Soil Survey Device. International publication published under the
491 patent cooperation treaty (PCT). World Intellectual Property Organization, International
492 Bureau. International Publication Number: WO2006/015463;
493 PCT/BE2005/000129; IPC: G01N21/00; G01N21/00.
494 Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., and Ramon, H. 2007. On-line
495 measurement of some selected soil properties using a VIS-NIR sensor. Soil and Tillage
496 Research, 93, 13-27
497 Mouazen, A.M., Maleki, M.R., Cockx, L., Van Meirvenne, M., Van Holm, L.H.J., Merckx,
498 R., De Baerdemaeker, J., & Ramon, H. 2009. Optimum three-point link set up for
499 optimal quality of soil spectra collected during on-line measurement. Soil & Tillage
500 Research, 103, 144-152.
501 Mouazen, A.M., Kuang, B., De Baerdemaeker, J. & Ramon, H., 2010. Comparison
502 between principal component, partial least squares and artificial neural network analyses
503 for accuracy of measurement of selected soil properties with visible and near infrared
504 spectroscopy. Geoderma, 158, 23-31.
505 Munoz, J. D., Kravchenko, A. 2011. Soil Carbon mapping using on-the-go near infrared
506 spectroscopy, topography and aerial photographs. Geoderma, 166**,** 102-110.
507 Sankey, J. B., Brown, D. J., Bernard, M. L., Lawrence, R. L. 2008. Comparing local vs.
508 global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS)

509 calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma, 148,
510 149-158.
511 Shibusawa, S., Made Anom, S. W., Sato, H. P., and Sasao, A. 2001. Soil mapping using the
512 real-time soil spectrometer. In ‘‘ECPA 2001’’ (G. Gerenier and S. Blackmore,
513 Eds.),Vol. 2, pp. 485–490. agro Montpellier, Montpellier, France.
514 Shepherd, K.D. & Walsh, M.G. 2002. Development of reflectance spectral libraries for
515 characterization of soil properties. Soil Science Society of America Journal, 66**,** 988-
516 998.
517 Shonk, J. L., Gaultney, L. D., Schulze, D. G., and Scoyoc, G. E. V. 1991. Spectroscopic
518 sensing of soil organic matter content. Trans. ASAE., 34, 1978–1984.
519 Srinivasan, A. 2006. Handbook of Pricision Agriculture: Principles and Applications. The
520 Haworth Press, Oxford (Chapter 1).
521 Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M. & Wetterlind, J. 2010. Visible and
522 near infrared spectroscopy in soil science. Advances in Agronomy, 107, 163-215.
523 Viscarra Rossel, R.A. & McBratney, A.B. 1998. Laboratory evaluation of a proximal
524 sensing technique for simultaneous measurement of soil clay and water content.
525 Geoderma, 85, 19-39.
526 Viscarra Rossel, R.V., McGlyn R.N., McBratney, A.B., 2006. Determining the composition
527 of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy.
528 Geoderma 146, 403-411.
529 Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A., McBratney, A.B. 2008. Using a legacy soil
530 Sample to develop a mid-IR spectral library. Australian Journal of Soil Research, 46, 1

Viscarra Rossel, R. A., Cattle , S. R. Ortega, A. Fouad, Y. 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma, 150, 253–266.

Waiser, T. H., Morgan, C. L. S., Brown, D. J., Hallmark, C. K. 2007. In Situ Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy. Soil Sci. Soc. Am. J, 71,389–396.

Wetterlind, J. and Stenberg, B. 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. European Journal of Soil Science, 61, 823–843.



https://dspace.org/

Effect of spiking strategy and ratio on calibration of on-line visible and near infrared soil sensor for measurement in European farms

Kuang, Boyan Y.; Mouazen, Abdul Mounem

2013-04-02T00:00:00Z
Elsevier Science B.V., Amsterdam.
http://dx.doi.org/10.1016/j.still.2012.11.006
Kuang B, Mouazen AM. (2013) Effect of spiking strategy and ratio on calibration of on-line visible and near infrared soil sensor for measurement in European farms. Soil & Tillage Research. Volume 128, April 2013, pp. 125-136

Downloaded from a DSpace repository. https://dspace.org/