

Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset

Journal:	<i>European Journal of Soil Science</i>
Manuscript ID:	EJSS-060-13.R1
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	24-Aug-2013
Complete List of Authors:	Guerrero, César; Universidad Miguel Hernandez de Elche, Agroquímica y Medio Ambiente Stenberg, Bo; Swedish University of Agricultural Sciences, Soil and Environment Wetterlind, Johanna; SLU, Department of Soil and Environment Viscarra Rossel, Raphael; CSIRO, Land and Water Maestre, Fernando; Universidad Rey Juan Carlos, Departamento de Biología y Geología Mouazen, Abdul Mounem; National Soil Science Institute, Natural Resources Department Zornoza, Raul; Universidad Politécnica de Cartagena, Departamento de Ciencia y Tecnología Agraria Ruiz-Sinoga, Jose Damian; Universidad de Málaga, Departamento de Geografía Kuang, Boyan; Cranfield University, Natural Resources Department
Keywords:	SOC assessment, soil sensing, near infrared spectroscopy, spiking, extra-weighting

SCHOLARONE™
Manuscripts

Published by Wiley. This is the Author Accepted Manuscript.

This article may be used for personal use only. The final published version (version of record) is available online at [10.1111/ejss.12129](https://doi.org/10.1111/ejss.12129) . Please refer to any applicable publisher terms of use.

1 **Assessment of soil organic carbon at local scale with spiked NIR**
2 **calibrations: effects of selection and extra-weighting on the spiking**
3 **subset**

4 C. GUERRERO^A, B. STENBERG^B, J. WETTERLIND^B, R.A. VISCARRA ROSSEL^C, F.T.
5 MAESTRE^D, A.M. MOUAZEN^E, R. ZORNOZA^F, J.D. RUIZ-SINOGA^G & B. KUANG^E

6 *^aDepartamento de Agroquímica y Medio Ambiente, Universidad Miguel Hernández de*
7 *Elche, E-03202, Spain, ^bDepartment of Soil and Environment, SLU, Skara, Sweden,*
8 *^cCSIRO Land and Water, Canberra, Australia, ^dDepartamento de Biología y Geología,*
9 *Universidad Rey Juan Carlos, Spain, ^eDepartment of Environmental Science and*
10 *technology, Cranfield University, UK, ^fDepartamento de Ciencia y Tecnología Agraria,*
11 *Universidad Politécnica de Cartagena, Spain, and ^gDepartamento de Geografía,*
12 *Universidad de Málaga, Spain.*

13 **Correspondence:** César Guerrero. E-mail: cesar.guerrero@umh.es

14 **Running title:** Spiking and extra-weighting to improve soil organic carbon predictions
15 with NIR

16

17

18

19 **Summary**

20 Spiking is a useful approach to improve the accuracy of regional or national
21 spectroscopic calibrations when they are used to predict at local scales. To do this, a
22 small subset of local samples (spiking subset) is added to recalibrate the regional or
23 national calibration. If the spiking subset is small in comparison with the size of the
24 initial calibration set, then the spiking subset could have little noticeable effect and only
25 a small improvement can be expected. For these reasons, we hypothesised that the
26 accuracy of the spiked calibrations can be improved when the statistical relevance of the
27 spiking subset is given extra-weight. We also hypothesised that the spiking subset
28 selection and the initial calibration size were relevant, and could affect the accuracy of
29 the recalibrated models. To test these hypotheses, we evaluated different strategies to
30 select the best spiking subset, with and without extra-weighting, to spike three initial
31 calibrations of different sizes. These calibrations were used to predict the soil organic
32 carbon (SOC) content in samples from four target sites. Our results confirmed that
33 spiking improved the prediction accuracy of the initial calibrations. We observed
34 differences in accuracy depending on the spiking subset used. The best results were
35 obtained when the spiking subset contained local samples evenly distributed in the
36 spectral space, regardless of the initial calibration's characteristics. The accuracy was
37 significantly improved when the spiking subset was extra-weighted. For medium- and
38 large-sized initial calibrations, the improvement due to extra-weighting was larger than
39 that caused by the increase in spiking subset size. This result is interesting because
40 extra-weighting the spiking subset is an inexpensive task. Similar accuracies were
41 obtained using small- and large-sized initial calibrations, suggesting that incipient
42 spectral libraries could be useful if the spiking subset is properly selected and extra-
43 weighted. When small-sized spiking subsets were used, the predictions results were

44 more accurate than those obtained with ‘geographically local’ models. Overall, our
45 results indicate that we can minimise the efforts needed to effectively use near-infrared
46 (NIR) spectroscopy for SOC assessment at local scales.

47

48 **Keywords:** SOC assessment, soil sensing, near infrared spectroscopy, spiking,
49 extra-weighting.

50

51 **Introduction**

52 Using near-infrared (NIR) spectroscopy to estimate soil properties is rapid, non-
53 destructive and relatively inexpensive compared to conventional laboratory analyses,
54 particularly when processing many samples. For NIR spectra to be quantitatively useful,
55 we need to develop and use a soil spectral database or library to derive spectroscopic
56 models (calibrations) that relate the spectra to analytical data, e.g. soil organic carbon
57 (SOC). When assessing soil properties at a local scale, we can develop site-specific or
58 ‘geographically local’ calibrations (Wetterlind *et al.*, 2010) that are generally very
59 accurate because smaller areas tend to be less variable in terms of the dependent
60 variable (Stenberg *et al.*, 2010), and the samples used to develop the calibration and
61 those used for prediction share similar characteristics, such as mineralogy and organic
62 matter quality (Reeves *et al.*, 1999; Janik *et al.*, 2007; Guerrero *et al.*, 2010; Wetterlind
63 *et al.*, 2010). A disadvantage of these models is that they are only valid for the local
64 area, which could be an expensive strategy when evaluating multiple areas. Another
65 option is to use regional, national or global calibrations, but they should represent the
66 variability of the soils being analysed. This has caused a trend to develop larger-scale
67 calibrations with a very large number of samples to ensure that the local samples fall

68 within the model's domain (Shepherd & Walsh, 2002; Brown *et al.*, 2006; Viscarra
69 Rossel, 2009; Grinand *et al.*, 2012; Viscarra Rossel & Webster, 2012), although this
70 cannot be guaranteed because soils have such variable characteristics, even at a regional
71 scale. Furthermore, a set of samples comprising a large-scale calibration should be
72 considered heterogeneous, but the local samples could be considered as a homogeneous
73 set that is located in a small area of the overall calibration domain. This could be the
74 reason for inaccurate (biased) results observed by some authors when using regional
75 and national calibrations to make predictions at local scales (Brown *et al.*, 2005; Brown,
76 2007; Janik *et al.*, 2007; Christy, 2008; Sankey *et al.*, 2008; Guerrero *et al.*, 2010;
77 Stenberg *et al.*, 2010; Wetterlind & Stenberg, 2010), even when the local samples fall
78 within the model domain and are not recognised as outliers. This could also explain
79 why better results are obtained with local (spectrum-specific) models (Genot *et al.*,
80 2011; Gogé *et al.*, 2012), where a subset of library samples that are similar to the
81 unknown sample is used to construct the calibration (Pérez-Marín *et al.*, 2007).
82 However, local methods are expensive because a large spectral library is needed to find
83 sufficient similar samples for the calibrations.

84 Spiking is an alternative method proposed to improve the accuracy of regional or
85 national calibrations for use at local scales (Viscarra Rossel *et al.*, 2009; Guerrero *et al.*,
86 2010; Stenberg *et al.*, 2010; Wetterlind & Stenberg, 2010; Kuang & Mouazen, 2013).
87 Spiking—sometimes referred to as 'augmentation' (Brown *et al.*, 2006; Brown, 2007;
88 Sankey *et al.*, 2008) and other names—involves three main steps (Janik *et al.*, 2007).
89 First, analyse a few samples from the target site in the laboratory using the reference
90 method; then add these samples to the initial calibration matrix; and then recalibrate the
91 model. This procedure usually increases the accuracy of the predictions in the rest of the
92 samples from the target site (Brown *et al.*, 2005; Sankey *et al.*, 2008; Wetterlind &

93 Stenberg, 2010). The higher the number of local samples in the spiking subset, the
94 higher the accuracy in the prediction set (Brown, 2007; Guerrero *et al.*, 2010), but a
95 large spiking subset decreases the advantages of NIR spectroscopy as a quick and low-
96 cost analytical method. To increase the relative proportion of the spiking subset,
97 Guerrero *et al.* (2010) suggested decreasing the number of samples in the initial
98 calibration set because they obtained higher accuracies when small-sized calibrations
99 were spiked, where the spiking subset had a larger influence. However, the selection of
100 a small number of calibration samples can reduce the amount of important information
101 for modelling, and lead to less robust calibrations. For this reason, we proposed an
102 alternative approach to increase the relevance of the spiking subset in the NIR
103 calibrations. The approach is to increase the statistical weight of the spiking subset by
104 adding several copies of the subset to the calibration matrix. These extra-weighted
105 samples are more important than other samples used to form the statistical model
106 (Capron *et al.* 2005; Stork & Kowalski 1999), which forces the calibration to better fit
107 the extra-weighted samples. If these samples were similar to the overall prediction set,
108 the model should provide more accurate predictions. We also evaluated different
109 strategies to select the best spiking subset. Since each local sample is different to the
110 others, we hypothesised that the selection of a spiking subset would influence the
111 accuracy of the spiked models, and the selection would be more influential if fewer
112 samples were used for spiking.

113 The spiking approach tries to gain benefits from a previously developed or initial
114 large-scale calibration set. It is reasonable to assume that results obtained could be
115 affected by the characteristics of the initial calibration, as some authors observed
116 (Guerrero *et al.*, 2010; Wetterlind & Stenberg, 2010). For this reason, we included
117 different initial calibrations in this study and evaluated their influence on the spiking

118 process. Our first objective was to evaluate how local samples should be selected as a
119 spiking subset for optimal spiking. To do this, we compared thirteen different strategies
120 to select the samples for the spiking subset. Our second objective was to evaluate
121 whether an extra-weighted spiking subset increased the prediction accuracy. In addition,
122 we compared geographically local models that used three different sized spiking
123 subsets. We selected SOC as the soil property for prediction, and we used the
124 coefficient of determination (R^2), root mean square error of prediction (RMSEP),
125 standard error of prediction (SEP) and ratio of performance to deviance (RPD) to
126 evaluate the prediction performance for four different target sites.

127 **2. Material and methods**

128 *2.1. National samples and initial calibrations*

129 A national soil library ($n = 2836$) of soils from different sites across Spain
130 (predominantly southeastern Spain) was randomly split into three subsets. These subsets
131 were used to create three initial calibrations of different sizes, representing three
132 different stages or efforts to develop the spectral library: small (IC#1; $n = 192$), medium
133 (IC#2; $n = 365$) and large (IC#3; $n = 2279$). The soils in the soil library were collected
134 under forest and agricultural land uses. Most of these soils developed over sedimentary
135 (mostly calcareous) lithologies. The soil samples were air-dried and sieved (< 2 mm),
136 and the NIR spectra ($12\,000\text{--}3800\text{ cm}^{-1}$) were obtained by FT-NIR diffuse reflectance
137 spectroscopy (MPA, Bruker Optik GmbH, Germany). The scale of the spectra was
138 transformed to nanometers ($830\text{--}2630\text{ nm}$), and re-sampled to 1 nm resolution. The
139 SOC concentration (%) was determined using the Walkley & Black (1934) method. The
140 different initial calibrations, relating the SOC to the NIR spectra, were constructed

141 using partial least squares (PLS) regression (PLS-1 algorithm) (see section 2.6 for
142 details). Key characteristics of the initial calibrations are shown in Table 1.

143 2.2. Target sites

144 We selected four independent target sites from four regions with spectral characteristics
145 that differed from each other and from those observed in the initial calibrations
146 (Figure 1; Appendix 1). Each target site is a relatively small area of dense sampling,
147 from several hectares to a few square kilometres in size. A different number of local
148 samples were collected at each target site (Table 2). One site was located in Sweden
149 (TS1), two in Spain (TS2, TS3) and one in the United Kingdom (TS4). As with the
150 initial calibration samples, the soil samples from the target sites were air-dried and
151 sieved (< 2 mm), and the NIR spectra and SOC content were obtained. Most of the
152 spectra were collected using a FT-NIR (MPA, Bruker Optik GmbH, Germany), except
153 the TS1 samples, which were scanned using a vis-NIR (ASD FieldSpec Pro Fr, USA).
154 The scale of the FT-NIR spectra was transformed from cm^{-1} to nanometers, and re-
155 sampled to 1 nm. For details about FT-NIR and vis-NIR scanning, see Guerrero *et al.*
156 (2010) and Wetterlind & Stenberg (2010), respectively.

157 2.3. Calibration types

158 Different types of calibrations relating SOC and NIR spectra were obtained using PLS
159 as a regression method (see section 2.6), and were used to predict the SOC contents in
160 the target site samples.

161 Initial calibrations: three different-sized initial calibrations (IC#1, IC#2 and IC#3,
162 described in section 2.1) that did not contain any samples from the target sites;
163 referred to as unspiked initial calibrations (Figure 2a; section 2.6).

164 Spiked calibrations: the three initial calibrations modified by adding a spiking subset
165 (n = 8) (Figure 2b). We used 13 different spiking subsets to spike each of the initial
166 calibrations (see section 2.4). In each initial calibration, we obtained 13 subtypes of
167 spiked calibrations, and we repeated this procedure for each of the four target sites.

168 Spiked calibrations with extra-weighting: in each of the different spiked calibrations,
169 the spiking subset was extra-weighted. To do this, we added 24 copies of each
170 spiking subset sample to the calibration set (Figure 2c), and then recalibrated the
171 model (see section 2.6). Each of the eight spiking subset samples appears 25 times
172 in the calibration matrix, becoming 24 times more influential than the soil library
173 samples because we have modified their leverage (Stork & Kowalski, 1999). We
174 selected 24 copies because the leverage of the target site samples followed an
175 asymptotic pattern after the addition of 15–20 copies (data not shown).

176 *2.4. Strategies to select the spiking subset from the target site samples*

177 For each target site, we used 13 strategies to select the different types of spiking subsets.
178 We hypothesised that each strategy had different advantages. The strategies were
179 designed and grouped on the basis of (i) the SOC values of target site samples, (ii) the
180 spectral characteristics of the target site samples and (iii) the spectral relationships
181 between the initial calibrations and the target site samples using the Mahalanobis
182 distance values. The first group of five strategies was designed on the basis of the SOC
183 content of target site samples. These strategies have a strictly theoretical value for
184 interpreting some results because the SOC contents of the target site samples would be
185 unknown in a real scenario, and thus these strategies would not be useful in practice.

186 Strategy 1 (OC low): select eight target site samples with the lowest SOC values (left
187 tail of SOC histogram). Samples with low SOC contents will show more clearly the

188 spectral features of the inorganic constituents, which are the most important factors
189 impeding the use of a calibration from one site to another. Moreover, these samples
190 could be useful to correct the bias in target site samples with low SOC contents.

191 Strategy 2 (OC high): select eight target site samples with the highest SOC values (right
192 tail of SOC histogram). These samples mask the inorganic spectral features, and
193 clearly show the SOC spectral features in the local samples. Moreover, these
194 samples can be useful to correct bias in target site samples with high SOC contents.

195 Strategy 3 (OC tails): select four samples with the lowest SOC values (from the left tail
196 of the SOC histogram) and four with the highest SOC values (from the right tail).
197 These samples can be useful to correct bias because the low and high SOC contents
198 are well established. Since low and high values are well described, the offset should
199 be also corrected.

200 Strategy 4 (OC centre): select eight target site samples with SOC values around the
201 median SOC value of the set.

202 Strategy 5 (OC distrib): select eight target site samples at regular intervals over the
203 entire range of SOC values (samples evenly distributed across the SOC values).
204 These samples should also be adequate for bias and offset correction.

205 To apply the three strategies in the second group, we performed a principal
206 component analysis (PCA) of the target site samples (NIR spectra pre-processed with
207 Savitzsky–Golay first derivative). The scores of the first, second and third principal
208 components (i.e. the first three) are represented in a scatter-plot.

209 Strategy 6 (PC periph): select eight target site samples located at the periphery of the
210 principal component spectral space defined by the first three principal components.

211 Strategy 7 (PC centre): select eight target site samples located at the centre of the
212 principal component spectral space defined by the first three principal components.

213 These are the most similar samples to the mean spectrum of the target site spectra.

214 Strategy 8 (PC distrib): select eight target site samples evenly distributed across the
215 principal component spectral space defined by the first three principal components.

216 This is the most intuitive strategy to uniformly cover the spectral diversity. This
217 selection was made using the ‘Automatic selection subset’ option in OPUS
218 (version 6.5 software; BrukerOptik GmbH, Ettlingen, Germany), which selects
219 samples in a similar fashion to the Kennard–Stone algorithm (Kennard & Stone,
220 1969).

221 The third group of five strategies was based on the Mahalanobis distance values of
222 the target site samples. The Mahalanobis distance values were calculated with respect to
223 the unspiked initial calibrations. Each target site sample had a different Mahalanobis
224 distance depending on the initial calibration used (i.e. IC#1, IC#2 or IC#3).

225 Strategy 9 (MD low): select eight target site samples with the lowest Mahalanobis
226 distance values (left tail of Mahalanobis distance histogram). These target site
227 samples are the closest to the initial calibration samples and the overall target site
228 samples, and could become a ‘bridge’ between both sets.

229 Strategy 10 (MD high): select eight target site samples with the highest Mahalanobis
230 distance values (right tail of Mahalanobis distance histogram). These samples are
231 the first recognised as outliers. In some schemes of calibration maintenance
232 (Shepherd & Walsh; 2002), it has been suggested the addition of this type of
233 samples when calibrations must be updated. These target site samples are the most
234 effective decreasing the Mahalanobis distance of the overall target site set (Capron
235 *et al.*, 2005).

236 Strategy 11 (MD tails): select four target site samples with the lowest Mahalanobis
237 values and four with the highest Mahalanobis distance values.

238 Strategy 12 (MD centre): select eight target site samples with Mahalanobis distance
239 values around the median Mahalanobis distance value.

240 Strategy 13 (MD distrib): select eight target site samples at regular intervals over the
241 entire range of Mahalanobis distance values (samples evenly distributed across the
242 Mahalanobis distance values).

243 *2.5 Experimental design and statistical analysis*

244 For this study, a repeated measures factorial design was established. The between-
245 subject factors were ‘initial calibration’, with three levels (i.e. three initial calibrations
246 of different sizes, IC#1, IC#2 and IC#3) and ‘strategy’, with 13 levels (i.e. 13 spiking
247 subset selection strategies). The within-subject factor was ‘extra-weighting’, with two
248 levels (i.e. without and with extra-weighting). For each combination of factors, we
249 calculated the R^2 , RMSEP, SEP and RPD to compare the actual SOC content of the
250 target site samples with the SOC predicted by the different calibrations. This design was
251 applied separately to the four target sites. The prediction performance parameters
252 obtained in each target site were considered as replicates. We used RMSEP to inform us
253 about accuracy and SEP about precision. The RPD (the ratio between the standard
254 deviation of the prediction set and the RMSEP) allowed us to compare the accuracy
255 obtained in prediction sets with different standard deviations.

256 The differences in RMSEP, SEP and RPD were analysed using a repeated measures
257 ANOVA. We excluded the strategies based on the SOC values (strategies 1–5) from the
258 statistical analysis because they are not useful in practice. In this way, the repeated
259 measures ANOVA was performed using eight levels of spiking subset selection strategy

260 and three levels of initial calibration as the between-subject factors, and two levels of
261 extra-weighting as the within-subject factor. Homocedasticity and normality was
262 checked using Levene and Kolmogorov–Smirnov tests, respectively; the original
263 variables were transformed to meet with the ANOVA assumptions when appropriate.
264 The R^2 was excluded from this statistical analysis because it did not meet the
265 assumptions. The assumption of sphericity was not violated when using the Mauchly's
266 test of sphericity. The software IBM SPSS Statistics version 20 (IBM, Armonk, NY)
267 was used for statistical analyses. We also obtained predictions using the unspiked initial
268 calibrations, but these results were not included in the statistical analysis.

269 *2.6. Development of calibrations with PLS-regression*

270 The models relating the NIR spectra with the SOC contents in soils were obtained with
271 PLS-regression (PLS-1 algorithm; OPUS version 6.5 software; BrukerOptik GmbH,
272 Ettlingen, Germany). We selected the number of PLS-vectors through leave-one-out
273 cross-validation. Before calibration, the SOC contents were transformed by the square
274 root but predicted SOC data were back-transformed before we compared them with
275 actual SOC and calculated the prediction performance parameters. NIR-spectra were
276 transformed by the first derivative (Savitzsky–Golay, 25 points). The number of PLS-
277 vectors in the spiked calibrations was set to the same number as in the corresponding
278 initial calibration. In TS1, we used the spectral range 1000–2500 nm to meet a common
279 range with a similar noise to the spectra collected with the FT-NIR instrument.

280 *2.7. Additional comparisons: extra-weighting effect versus the increase of the spiking* 281 *subsets size and versus geographically local models*

282 These comparisons were made only with spiking subsets selected by the ‘PC distrib’
283 strategy, which was one of the most effective selection strategies in terms of increasing

284 accuracy. We compared the extra-weighting effect against the increase of the spiking
285 subsets size. To do this, we spiked the three initial calibrations with 8, 16 and
286 32 spiking subset samples selected by the 'PC distrib' strategy. Similar to the procedure
287 described in section 2.3, we obtained spiked calibrations by adding 24 copies of the
288 spiking subset (denoted as EW_24). For each target site, we used these calibrations to
289 predict the SOC contents in the target site samples. In all cases, the 32 spiking subset
290 samples were not used in the RMSEP computation, to allow a fair comparison of
291 accuracy regardless of the size of the spiking subset. The RMSEP values were analysed
292 with a repeated measures ANOVA, where two levels of extra-weighting (with and
293 without extra-weighted) acted as the within-subject factor, and three levels of the
294 spiking subsets size (8, 16 and 32 samples) acted as the between-subject factor. Due to
295 the large differences between the sizes of the initial calibrations, we also used a
296 different approach to calculate the number of copies to add, which was the ratio
297 between the initial calibration size and the spiking subset size. In this way, more copies
298 are added when the initial calibration size is larger or when the spiking subset size is
299 smaller. The extra-weighting effect obtained using the initial calibration-to-spiking
300 subset ratio (denoted as EW_ratio) was evaluated using repeated measures ANOVA, as
301 for the EW_24 approach. The data used in these statistical analyses did not violate the
302 ANOVA assumptions (homocedasticity and normality) or the condition of sphericity.
303 For each target site, three geographically local or site-specific models were constructed
304 using the 8, 16 and 32 spiking subsets selected by the 'PC distrib' strategy.

305 **3. Results**

306 *3.1. Effect of spiking (without extra-weighted)*

307 The predictions obtained with the unspiked initial calibrations for each target site were
308 inaccurate, with large prediction errors (Figure 3). For the 12 cases (three initial
309 calibrations applied to four target sites), the RPD values ranged from < 0.10 to 1.44,
310 which clearly indicated poor predictions. Figure 4 shows the R^2 , RMSEP, SEP and RPD
311 values obtained with the unspiked and spiked calibrations, where each value shown is
312 the mean value of those obtained for the four target sites. The unspiked IC#1 provided
313 very low quality predictions, with $R^2 = 0.33 \pm 0.34$ (mean \pm standard deviation) and
314 $RPD = 0.52 \pm 0.21$ (Figure 4a). Once spiked, we observed a drastic and positive change
315 in all the parameters related to the quality of predictions (Figure 4a), and bias was
316 substantially decreased. There were differences in accuracy for the spiked calibrations
317 depending on the strategy used to select the spiking subset. For example, the RMSEP
318 values obtained with the IC#1 spiked using the 'OC low'(worst) and 'PC distrib'(best)
319 strategies were $0.70 \pm 0.16\%$ and $0.37 \pm 0.15\%$ SOC, respectively, both of which were
320 clearly better than the RMSEP for the unspiked IC#1 of $1.86 \pm 1.77\%$ SOC (Figure 4a).
321 Similarly, spiking of IC#2 (Figure 4b) caused a noticeable improvement in prediction
322 accuracy, mostly due to improvement of bias. Interestingly, the worst ('OC low') and
323 best ('PC distrib') strategies for IC#2 were the same as those observed for IC#1. A
324 substantial improvement in accuracy was also obtained when IC#3 was spiked, due to a
325 strong decrease in bias (Figure 4c). In this case, the worst and best strategies (in terms
326 of accuracy) were not the same as for IC#1 and IC#2. In general, the best accuracies
327 were obtained using IC#1 (the calibration with the smallest size) and the worst
328 accuracies were obtained with IC#3 (the calibration with the largest size). To illustrate
329 the effect of spiking with different spiking subsets, individual results for the four target
330 sites obtained with the 'MD centre' and 'PC distrib' selection strategies are shown in
331 Figure 3.

332 *3.2. Effect of extra-weighting on the spiking subset selection strategies*

333 The addition of several copies of the spiking subset (i.e. extra-weighting) in the spiked
334 calibrations caused a significant improvement ($P < 0.001$) in the RMSEP, SEP and RPD
335 (Table 3). The effect of extra-weighting on these parameters was similar across the
336 spiking subset selection strategies (extra-weighting \times strategy, $P > 0.05$; Table 3), and
337 also similar in the three different initial calibrations evaluated (extra-weighting \times initial
338 calibration, $P > 0.05$; Table 3), although the extra-weighting effect on the R^2 was
339 greater in IC#3 (Figure 4).

340 We observed that accuracy differed depending on the strategy used to select the
341 spiking subset (Figure 4). Indeed, all the parameters evaluated showed significant
342 differences across the strategies (Table 3). The differences between strategies were
343 similar in the three initial calibrations evaluated, as suggested by the non-significant
344 interaction between the 'strategy' and the 'initial calibration' ($P > 0.05$; Table 3). In two
345 strategies ('OC low' and 'OC high'), extra-weighting had a negative effect through an
346 increase in bias (Figure 4). The 'OC low' strategy was worst for IC#2 and IC#3, and
347 second worst for IC#1. When extra-weighting was applied, 'PC distrib' was the best
348 performing strategy in the three initial calibrations, and clearly improved the accuracy
349 due to decrease in bias, but also due to a decrease in SEP (Figure 3 & Figure 4). In IC#1
350 and IC#2, the combined use of the spiking subset ('PC distrib') and extra-weighting
351 increased the RPD by 1.5 units compared to the unspiked initial calibrations, allowing
352 RPD values to exceed 2 (Figure 4). The results obtained with the 'MD centre' and 'PC
353 distrib' strategies (without and with extra-weighting) for each target site illustrate the
354 extra-weighting effects (Figure 3).

355 *3.3. Increase of spiking subsets size versus extra-weighting, and comparison with*
356 *geographically local models*

357 We compared the effects of increasing the spiking subset size with extra-weighting for
358 the ‘PC distrib’ selection strategy. There was a positive effect on the accuracy when the
359 spiking subset size was increased (Figure 5), although this effect was not significant
360 ($P > 0.05$; Table 4). Regardless of the spiking subset size, there was a significant
361 improvement in the accuracy when the spiking subsets were extra-weighted ($P < 0.001$,
362 Table 4). These results were similar for the two approaches followed to select the
363 number of copies to add for extra-weighted (Table 4, Figure 5). It is worth highlighting
364 that in IC#2 and IC#3, the improvement of the accuracy due to extra-weighting was
365 clearly higher than the duplication of the spiking subset size (Figure 5), and even higher
366 than the quadruplication of the spiking subset size in IC#3 (Figure 5). The extra-
367 weighting effect in IC#1 was smaller because spiking was enough to cause the
368 saturation of the improvement, mainly due to its smaller size. When the spiking subset
369 was not extra-weighted (black bars in Figure 5), the best results were obtained with the
370 small-sized initial calibration (IC#1), and results obtained with IC#2 and IC#3 were less
371 accurate than those obtained with the geographically local models. Once the spiking
372 subset was extra-weighted, the differences between initial calibrations practically
373 disappeared, especially when the number of copies added was selected according to the
374 ratio of the initial calibration to the spiking subset (EW_ratio; light grey bars in
375 Figure 5). When this approach was used for extra-weighting (EW_ratio), the spiked
376 initial calibrations were more accurate than the geographically local models. When a
377 large number of local samples (32) were considered as spiking subset size ($SS = 32$),
378 and also as ‘n’ of the geographically local models ($n = 32$), scarce differences between
379 both approaches were observed, except for the reduced robustness obtained with the
380 geographically local models (Figure 5).

381

382 4. Discussion

383 4.1. Effect of spiking

384 The predictions obtained using the unspiked initial calibrations had a low accuracy. The
385 bias was the main problem, representing more than 50% of the error, as some authors
386 observed (e.g. Bellon-Maurel & McBratney, 2011). These results were expected, and
387 clearly demonstrate how we cannot safely used calibrations do not cover the
388 characteristics of the target sites. As for any model, the spectroscopic calibrations are
389 valid only for samples with similar characteristics as those used in the calibration
390 (Viscarra Rossel *et al.*, 2008). For these reasons, there is a trend to develop large
391 spectral libraries (Shepherd & Walsh, 2002; Brown *et al.*, 2006; Viscarra Rossel, 2009;
392 Grinand *et al.*, 2012; Viscarra Rossel & Webster, 2012). But the accuracy of the
393 calibrations improved drastically when only eight local samples were added to spike the
394 initial calibrations. Once the calibrations contained relevant information for the target
395 site, the predictions became more accurate. The improved accuracy was mostly due to
396 the decrease in bias, in accordance with previous studies (e.g. Stork & Kowalski, 1999;
397 Bricklemeyer & Brown, 2010; Guerrero *et al.*, 2010; Stenberg *et al.*, 2010; Wetterlind &
398 Stenberg, 2010), but also by an improvement in precision. Many factors affect soil
399 genesis, and soils present an extraordinary variation in composition and characteristics
400 compared with other environmental materials. This makes it difficult to construct a
401 calibration containing the immense variation found in soils, even at a regional scale
402 (Sudduth & Hummel, 1996; Sankey *et al.*, 2008; Minasny *et al.*, 2009; Reeves & Smith,
403 2009). In this way, a large calibration does not guarantee accurate predictions. In fact,
404 several authors observed inaccurate predictions when calibrations were used in samples
405 from independent sites (Christy, 2008; D'Acqui *et al.*, 2010; Wetterlind & Stenberg,
406 2010; Bellon-Maurel & McBratney, 2011). Thus, trying to include all the soil's

407 variation is an immense and probably unnecessary effort. Spiking could be an attractive
408 and economical alternative, avoiding the need for large spectral libraries, since we
409 observed the best results when the small-sized initial calibration was spiked. As
410 Guerrero *et al.* (2010) observed, the new information added (i.e. the spiking subset) was
411 more influential on a small-sized initial calibration than on a large-sized one, which
412 explains why better predictions were obtained after spiking the small-sized initial
413 calibration (IC#1).

414 4.2. *Effects of extra-weighting on the spiking subset selection strategies*

415 To directly increase the significance or relevance of the added information, several
416 copies of the spiking subset were included in the spiked initial calibrations. The addition
417 of several copies increased their weight and influence on the model (Stork & Kowalski,
418 1999). Under these circumstances, the calibration was forced to fit preferentially to
419 these samples. Consequently, if the extra-weighted samples are representative of the
420 overall prediction set (i.e. the target site), then the calibration must provide reliable
421 predictions for that set. Indeed, extra-weighting caused a significant improvement
422 ($P < 0.001$) on all the parameters related to the quality of predictions. It is interesting to
423 highlight that the effects on the precision (SEP) and accuracy (RMSEP) were similar for
424 the three initial calibrations evaluated, suggesting a robustness of that pattern, since the
425 three initial calibrations were different to each other. So, extra-weighting is a simple,
426 fast and inexpensive task that we recommend when spiking calibrations. The extra-
427 weighting caused a strong decrease in the leverage of the spiking subset (Stork &
428 Kowalski, 1999; Capron *et al.*, 2005). Consequently, the extra-weighting could be
429 considered as a manipulation of the spectral space, since it causes a displacement of the
430 calibration centroid toward the extra-weighted samples. In this sense, the extra-
431 weighting is a frequent approach used in samples that are added for updating

432 calibrations to new conditions, especially when their number is relatively low in
433 comparison with the overall calibration set (Stork & Kowalski, 1999), as in our
434 scenarios (especially in IC#2 and IC#3).

435 The improvement in the RMSEP, SEP and RPD was dependent on the strategy used
436 to select the spiking subset, as Capron *et al.* (2005) also observed. The differences
437 found between strategies were similar in the three initial calibrations used, as revealed
438 by the non-significant interaction ($P > 0.05$) between the ‘strategy’ and ‘initial
439 calibration’ factors. These results suggest that the effects exerted by the added samples
440 (spiking subset) are not totally controlled by the characteristics of the initial calibration.
441 The soil samples within a local set are different from each other, the information
442 provided by each sample is different (Naes, 1987; Isaksson & Naes, 1990; Shetty *et al.*,
443 2012), and consequently, the improvement in the accuracy of the spiked calibration
444 should also vary. In this sense, using an inadequate spiking subset could be one of the
445 reasons explaining why some authors have found a scarce effect of spiking
446 (Brickley & Brown, 2010; Guerrero *et al.*, 2010). Thus, the identification of a
447 successful strategy to select the most adequate spiking subset is clearly relevant. For
448 these reasons, we evaluated strategies aimed to cover a wide range of different types of
449 spiking subset. Since large bias values have been the most common problem observed
450 (Stork & Kowalski, 1999; Janik *et al.*, 2007; Bellon-Maurel & McBratney, 2011), we
451 suspected that using a spiking subset containing strategic SOC values could be adequate
452 to improve the bias, and consequently the accuracy. In fact, we observed that the ‘OC
453 tails’ and ‘OC distrib’ selection strategies offered better predictions than the ‘OC
454 centre’, ‘OC high’ and ‘OC low’ strategies, since they were adding information in
455 several strategic spaces related with the bias, slope and offset. But it is important to note

456 that the strategies based on the SOC values are not useful in practice, and they were
457 included in the experiment for conceptual evaluation and comparison.

458 The calibrations spiked with samples evenly distributed in the principal component
459 spectral space ('PC distrib') gave better predictions than those spiked with samples
460 evenly distributed along the concentration values ('OC distrib'). Both strategies select
461 different local samples because the SOC content is not uniquely responsible for the
462 spectral variation within a target site. Compared to texture and mineralogy composition,
463 SOC typically has a fairly small influence on spectra (Stenberg *et al.*, 1995; Islam *et al.*,
464 2005; Stenberg *et al.*, 2010). This result is interesting since only the spectral
465 information is available in a real situation (Kusumo *et al.*, 2008; Mora & Schimleck,
466 2008). The predictions obtained with calibrations spiked with a spiking subset selected
467 using the 'PC centre' strategy were less accurate than those selected with 'PC periph'.
468 The samples selected with the 'PC centre' strategy are those more similar to the mean
469 spectrum of the target site. In contrast, those selected with 'PC periph' are more
470 dissimilar to the mean spectrum, but they represent greater diversity. The strategies that
471 included most of the spectral diversity were 'PC distrib' and 'PC periph', and they were
472 two successful strategies, especially the latter. Indeed, there are several methods for
473 optimal sample selection based on spectral characteristics (Naes, 1987; Puchwein, 1988;
474 Isaksson & Naes, 1990; Shenk & Westerhaus, 1991; Kusumo *et al.*, 2008) but two of
475 the most commonly used are the Kennard–Stone algorithm (Kennard & Stone, 1969;
476 Mora & Schimleck, 2008; Shetty *et al.*, 2012), which covers the experimental region
477 uniformly (as in 'PC distrib'), and the D-optimal procedure (Olsson *et al.*, 2004;
478 Rodionova & Pomerantsev, 2007; Brandmaier *et al.*, 2012), which selects objects
479 located on the periphery (most extreme) of the experimental region (as in 'PC periph').

480 There were scarce differences between the selections made using the Mahalanobis
481 distance. The values of Mahalanobis distance were extremely high, and all the local
482 samples were always classified as outliers. Consequently, these sets are not sensitive to
483 the Mahalanobis distance criterion. This criterion would probably be relevant when
484 samples from the target sites are more similar to those comprising the initial calibration
485 (Puchwein, 1988; Capron *et al.*, 2005).

486 4.3. Increase of spiking subset size versus extra-weighting, and comparison with 487 geographically local models

488 When the 'PC distrib' strategy was used to select the spiking subset, extra-weighting
489 was preferred over the increase in spiking subset size. This was a very interesting result,
490 since extra-weighting caused a significant improvement in accuracy without any
491 analytical effort. In contrast, the increase of the spiking subset size implies efforts in
492 terms of time and money, and the improvement of the RMSEP was not statistically
493 significant. The non-significant improvement of the RMSEP was probably due to the
494 high efficiency of the 'PC distrib' strategy to select the most representative samples.
495 Consequently, a further addition of samples would prove scarcely useful, since the new
496 added samples would be redundant (in comparison with the first ones selected). These
497 results agree with those obtained by other authors (Naes, 1987; Puchwein, 1988;
498 Isaksson & Næs, 1990; Capron *et al.*, 2005; D'Acqui *et al.*, 2010; Grinand *et al.*, 2012;
499 Shetty *et al.*, 2012), where only a small subset of samples properly selected can offer a
500 similar accuracy than a larger set. In this context, extra-weighting the spiking subset is
501 an efficient approach, which can avoid the need of large-sized spiking subsets.

502 The influence of spiking was greater in the small-sized initial calibrations than in
503 the large-sized ones (Guerrero *et al.*, 2010). When the extra-weighting was made using
504 the same number of copies regardless of the initial calibration size (EW_24), this

505 pattern was still present, but clearly to a lesser degree. When the extra-weighting was
506 based on the initial calibration to spiking subset ratio (EW_ratio), more copies were
507 included in the large-sized initial calibration (IC#3) than in the smaller-sized initial
508 calibrations (IC#1 and IC#2). However, even under these conditions, the results
509 obtained for the three initial calibrations were similar. This result was very interesting
510 because it suggests that small-sized initial calibrations could offer a similar accuracy
511 than large-sized initial calibrations. Consequently, this approach can be considered as a
512 strong alternative to the need to develop large spectral libraries. In addition, in those
513 circumstances where only a few local samples can be analysed by the reference method
514 (i.e. 8–16 samples), this approach offered more accurate results than the geographically
515 local (or site-specific) models. When a larger number of local samples were analysed
516 (32 local samples), small differences in accuracy were observed between both
517 approaches, although the geographically local models were less robust, indicating the
518 difficulty to develop consistent spectroscopic calibrations when the number of samples
519 is low.

520 More studies are needed to evaluate if extra-weighting can outperform local models
521 (spectrum-specific models), where a dedicated model is calibrated for an individual
522 unknown sample (Pérez-Marín *et al.*, 2007), or other approaches where a partition of
523 the spectral information is used (Viscarra Rossel & Webster, 2012). It is interesting to
524 highlight that local methods (spectrum-specific) can be used only when the spectral
525 library contains similar samples to the target site samples, which is not the case for sets
526 evaluated in this paper. In contrast, spiking with a properly selected spiking subset,
527 together with extra-weighting, can overcome this problem, allowing the extrapolation of
528 the initial calibrations applicability.

529

530 Conclusions

531 The addition of a small spiking subset (eight local samples) to spike the calibrations
532 improved the accuracy of the SOC predictions. There were, however, important
533 differences in accuracy, which were dependent on the strategy used to select the spiking
534 subset. The best results were obtained when the calibrations were spiked with local
535 samples that were evenly distributed across the space defined by the first three principal
536 components (spiking subset selected with the 'PC distrib' strategy). In addition, extra-
537 weighting was an effective way to improve the accuracy of the spiked calibrations.
538 Extra-weighting of the spiking subset accentuates the spiking effect, giving an
539 acceptable level of accuracy when predictions of SOC are needed at local scale, and
540 when using small-sized spiking subsets. Large-sized calibrations are probably not
541 needed when these approaches are considered, since similar results were obtained with
542 the small- and large-sized calibrations, and it suggests that incipient spectral libraries
543 could be useful if they are properly spiked and extra-weighted. Consequently, extra-
544 weighting is a simple, fast and inexpensive task that we highly recommend when
545 calibrations are spiked, and can avoid the need to develop geographically local models.
546 Overall, our results indicate that the efforts needed to use NIR spectroscopy for SOC
547 assessment at local scales can be minimised.

548 Acknowledgements

549 This work was part of a research project (Ref. CGL2011-27001) sponsored by the
550 Spanish Government Ministerio de Economía y Competitividad, and C. Guerrero
551 gratefully acknowledges this financial support. C. Guerrero acknowledges the Spanish
552 Government Ministerio de Educación for a travel grant (ref. JC2011-0342).
553 F.T. Maestre was supported by the European Research Council through the European

554 Commission's Seventh Framework Programme (FP7/2007–2013) under grant
555 agreement no. 242658 (BIOCOM).

556

For Peer Review

557 **References**

- 558 Bellon-Maurel, V. & McBratney, A. 2011. Near-infrared (NIR) and mid-infrared (MIR)
559 spectroscopic techniques for assessing the amount of carbon stock in soils – Critical
560 review and research perspectives. *Soil Biology and Biochemistry*, **43**, 1398–1410.
- 561 Brandmaier, S., Sahlin, U., Tetko, I.V. & Oberg, T. 2012. PLS-Optimal: A stepwise D-
562 Optimal design based on latent variables. *Journal of Chemical Information and*
563 *Modeling*, **52**, 975–983.
- 564 Bricklemyer, R.S. & Brown, D.J. 2010. On-the-go VisNIR: Potential and limitations for
565 mapping soil clay and organic carbon. *Computers and Electronics in Agriculture*,
566 **70**, 209–216.
- 567 Brown, D.J. 2007. Using a global VNIR soil-spectral library for local soil
568 characterization and landscape modelling in a 2nd-order Uganda watershed.
569 *Geoderma*, **140**, 444–453.
- 570 Brown, D.J., Bricklemyer, R.S. & Millar, P.R. 2005. Validation requirements for
571 diffuse reflectance soil characterization models with a case study of VNIR soil C
572 prediction in Montana. *Geoderma*, **129**, 251–267.
- 573 Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D. & Reinsch, T.G. 2006. Global
574 soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, **132**,
575 273–290.
- 576 Capron, X., Walczak, B., de Noord, O.E. & Massart, D.L. 2005. Selection and
577 weighting of samples in multivariate regression model updating. *Chemometrics and*
578 *Intelligent Laboratory Systems*, **76**, 205–214.

- 579 Christy, C.D. 2008. Real-time measurement of soil attributes using on-the-go near
580 infrared reflectance spectroscopy. *Computers and Electronics in Agriculture*, **61**,
581 10–19.
- 582 D’Acqui, L.P., Pucci, A. & Janik, L.J. 2010. Soil properties prediction of western
583 Mediterranean islands with similar climatic environments by means of mid-infrared
584 diffuse reflectance spectroscopy. *European Journal of Soil Science*, **61**, 865–876.
- 585 Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, Y. & Dardenne, P. 2011. Near
586 infrared reflectance spectroscopy for estimating soil characteristics valuable in the
587 diagnosis of soil fertility. *Journal of Near Infrared Spectroscopy*, **19**, 117–138.
- 588 Gogé, F., Joffre, R., Jolivet, C., Ross, I. & Ranjard, L. 2012. Optimization criteria in
589 sample selection step of local regression for quantitative analysis of large soil NIRS
590 database. *Chemometrics and Intelligent Laboratory Systems*, **110**, 168–176.
- 591 Grinand, C., Barthès, B.G., Brunet, D., Kouakoua, E., Arrouays, D., Jolivet, C., Caria,
592 G. & Bernoux, M. 2012. Prediction of soil organic and inorganic carbon contents at
593 a national scale (France) using mid-infrared reflectance spectroscopy (MIRS).
594 *European Journal of Soil Science*, **63**, 141–151.
- 595 Guerrero, C., Zornoza, R., Gómez, I. & Mataix-Beneyto, J. 2010. Spiking of NIR
596 regional models using samples from target sites: Effect of model size on prediction
597 accuracy. *Geoderma*, **158**, 66–77.
- 598 Isaksson, T. & Naes, T. 1990. Selection of samples for calibration in near-infrared
599 spectroscopy. Part II: selection based on spectral measurements. *Applied*
600 *Spectroscopy*, **44**, 1152–1158.

- 601 Islam, K., McBratney, A. & Singh, B. 2005. Rapid estimation of soil variability from
602 the convex hull biplot area of topsoil ultra-violet, visible and near-infrared diffuse
603 reflectance spectra. *Geoderma*, **128**, 249–257.
- 604 Janik, L.J., Skjemstad, J.O., Sheperd, K.D. & Spouncer, L.R. 2007. The prediction of
605 soil carbon fractions using mid-infrared-partial least square analysis. *Australian*
606 *Journal of Soil Research*, **45**, 73–81.
- 607 Kennard, R.W. & Stone, L.A. 1969. Computer aided design of experiments.
608 *Technometrics*, **11**, 137–148.
- 609 Kuang, B. & Mouazen, A.M. 2013. Effect of spiking strategy and ratio on calibration of
610 on-line visible and near infrared soil sensor for measurement in European farms.
611 *Soil & tillage Research*, **128**, 125–136.
- 612 Kusumo, B.H., Hedley, C.B., Hedley, M.J., Hueni, A., Tuohy, M.P. & Arnold, G.C.
613 2008. The use of diffuse reflectance spectroscopy for in situ carbon and nitrogen
614 analysis of pastoral soils. *Australian Journal of Soil Research*, **46**, 623–635.
- 615 Minasny, B., Tranter, A.B., Brough, D.M. & Murphy, B.W. 2009. Regional
616 transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil
617 chemical properties. *Geoderma*, **153**, 155–162.
- 618 Mora, C.R. & Schimleck, L.R. 2008. On the selection of samples for multivariate
619 regression analysis: application to near-infrared (NIR) calibration models for the
620 prediction of pulp yield in *Eucalyptus nitens*. *Canadian Journal of Forest Research*,
621 **38**, 2626–2634.
- 622 Naes, T. 1987. The design of calibration in near infrared reflectance analysis by
623 clustering. *Journal of Chemometrics*, **1**, 121–134.

- 624 Olsson, I-M., Gottfries, J. & Wold, S. 2004. D-optimal onion designs in statistical
625 molecular design. *Chemometrics and Intelligent Laboratory Systems*, **73**, 37–46.
- 626 Pérez-Martín, D., Garrido-Varo, A. & Guerrero, J.E. 2007. Non-linear regression
627 method in NIRS quantitative analysis. *Talanta*, **72**, 28–42.
- 628 Puchwein, G. 1998. Selection of calibration samples for near-infrared spectrometry by
629 factor analysis of spectra. *Analytical Chemistry*, **60**, 569–573.
- 630 Reeves III, J.B., McCarty, G.W. & Meisinger, J.J. 1999. Near infrared reflectance the
631 analysis of agricultural soils. *Journal of Near Infrared Spectroscopy*, **7**, 179–193.
- 632 Reeves III, J.B. & Smith, D.B. 2009. The potential of mid- and near-infrared diffuse
633 reflectance spectroscopy for determining major- and trace-element concentrations
634 in soils from a geochemical survey of North America. *Applied Geochemistry*, **24**,
635 1472–1481.
- 636 Rodionova, O.Y. & Pomerantsev, A.L. 2008. Subset selection strategy. *Journal of*
637 *Chemometrics*, **22**, 674–685.
- 638 Sankey, J.B., Brown, D.J., Bernard, M.L. & Lawrence, R.L. 2008. Comparing local vs.
639 global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS)
640 calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma*,
641 **148**, 149–158.
- 642 Shenk, J.S. & Westerhaus, M.O. 1991. Population definition, sample selection, and
643 calibration procedures for near-infrared reflectance spectroscopy. *Crop Science*, **31**,
644 469–474.
- 645 Shepherd, K.D. & Walsh, M.G. 2002. Development of reflectance spectral libraries for
646 characterization of soil properties. *Soil Science Society of America Journal*, **66**,
647 988–998.

- 648 Shetty, N., Rinnan, A. & Gislum, R. 2012. Selection of representative calibration
649 sample sets for near-infrared reflectance spectroscopy for predict nitrogen
650 concentration in grasses. *Chemometrics and Intelligent Laboratory Systems*, **111**,
651 59–65.
- 652 Stenberg, B., Nordkvist, E. & Salomonsson, L. 1995. Use of near infrared reflectance
653 spectra of soils for objective selection of samples. *Soil Science*, **159**, 109–114.
- 654 Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M. & Wetterlind, J. 2010. Visible and
655 near infrared spectroscopy in soil science. *Advances in Agronomy*, **107**, 163–215.
- 656 Stork, C.L. & Kowalski, B.R. 1999. Weighting schemes for updating regression models
657 – a theoretical approach. *Chemometrics and Intelligent Laboratory Systems*, **48**,
658 151–166.
- 659 Sudduth, K.A. & Hummel, J.W. 1996. Geographical operating range evaluation of a
660 NIR soil sensor. *Transactions of the ASAE*, **39**, 1599–1604.
- 661 Viscarra Rossel, R.A. & Webster, R. 2012. Predicting soil properties from the
662 Australian soil visible-near infrared spectroscopic database. *European Journal of*
663 *Soil Science*, **63**, 848–860.
- 664 Viscarra Rossel, R. 2009. The Soil Spectroscopy Group and the development of a
665 global soil spectral library. *NIR News*, **20**, 14–15.
- 666 Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A. & McBratney A.B. 2008. Using a
667 legacy soil sample to develop a mid-IR spectral library. *Australian Journal of Soil*
668 *Research*, **46**, 1–16.
- 669 Viscarra Rossel, R.A., Cattle, S.R., Ortega, A. & Fouad, Y. 2009. In situ measurements
670 of soil colour, mineral composition and clay content by vis-NIR spectroscopy.
671 *Geoderma*, **150**, 253–266.

672 Walkley, A. & Black, I.A. 1934. An examination of the Degtjareff method for
673 determining soil O.M. and a proposed modification of the chromic acid titration
674 method. *Soil Science*, **37**, 29–38.

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

678 Wetterlind, J., Stenberg, B. & Söderström, M. 2010. Increased sample point density in
679 farm soil mapping by local calibration of visible and near infrared prediction
680 models. *Geoderma*, **156**, 152–160.

681

682

For Peer Review

683 **FIGURE CAPTIONS**

684

685 **Figure 1** Projections of the NIR spectra from the target sites (TS) into the principal
686 component space defined by the first two principal components, in each initial
687 calibration (IC). Grey stars denote the national samples of the initial calibrations and
688 black dots denote target site samples.

689

690 **Figure 2** Schematic description of the experimental setup: a) initial calibration (IC)
691 unspiked, constructed only with national samples (NS); b) initial calibration spiked with
692 a spiking subset (SS) selected by strategy #1; c) initial calibration spiked with spiking
693 subset selected by strategy #1, where an extra-weighting was applied to the spiking
694 subset. This scheme only shows one of the 13 strategies of spiking subset selection and
695 one of the three initial calibrations. This scheme was used with four different target sites
696 (TS). Dashed and double lines denote spiking and the use of the calibration for
697 obtaining predictions (\hat{y}), respectively.

698

699 **Figure 3a** Representative illustration of predictions obtained in each target site (TS)
700 with the different calibrations conducted. Left: predictions obtained with the unspiked
701 IC#1 (white stars; dotted line). Centre: predictions obtained with IC#1 spiked with the
702 spiking subset selected with the 'MD centre' strategy (white circles, dashed line) and
703 spiking subset extra-weighted (EW) (black circles, solid line). Right: predictions
704 obtained with IC#1 spiked with the spiking subset selected with the 'PC distrib' strategy
705 (white circles, dashed line) and spiking subset extra-weighted (black circles, solid line).

706

707 **Figure 3b** Representative illustration of predictions obtained in each target site (TS)
708 with the different calibrations conducted. Left: predictions obtained with the unspiked
709 IC#2 (white stars; dotted line). Centre: predictions obtained with IC#2 spiked with the
710 spiking subset selected with the 'MD centre' strategy (white circles, dashed line) and
711 spiking subset extra-weighted (EW) (black circles, solid line). Right: predictions
712 obtained with IC#2 spiked with the spiking subset selected with the 'PC distrib' strategy
713 (white circles, dashed line) and spiking subset extra-weighted (black circles, solid line).

714

715 **Figure 3c** Representative illustration of predictions obtained in each target site (TS)
716 with the different calibrations conducted. Left: predictions obtained with the unspiked
717 IC#3 (white stars; dotted line). Centre: predictions obtained with IC#3 spiked with the
718 spiking subset selected with the 'MD centre' strategy (white circles, dashed line) and
719 spiking subset extra-weighted (EW) (black circles, solid line). Right: predictions
720 obtained with IC#3 spiked with the spiking subset selected with the 'PC distrib' strategy
721 (white circles, dashed line) and spiking subset extra-weighted (black circles, solid line).

722

723 **Figure 4** Predictions obtained with unspiked and spiked calibrations (without and with
724 extra-weight) using the 13 different strategies to select the spiking subset. Strategies in
725 spiked calibrations (with and without extra-weighting) are arranged by RMSEP. a)
726 IC#1; b) IC#2; c) IC#3. In all cases, $n = 4$ (from the four target sites studied). The two
727 horizontal dark grey lines are displaying values of $RMSEP = 0.4\%$ soil organic carbon
728 (SOC) and $RMSEP = 0.8\%$ SOC to facilitate visual comparisons.

729

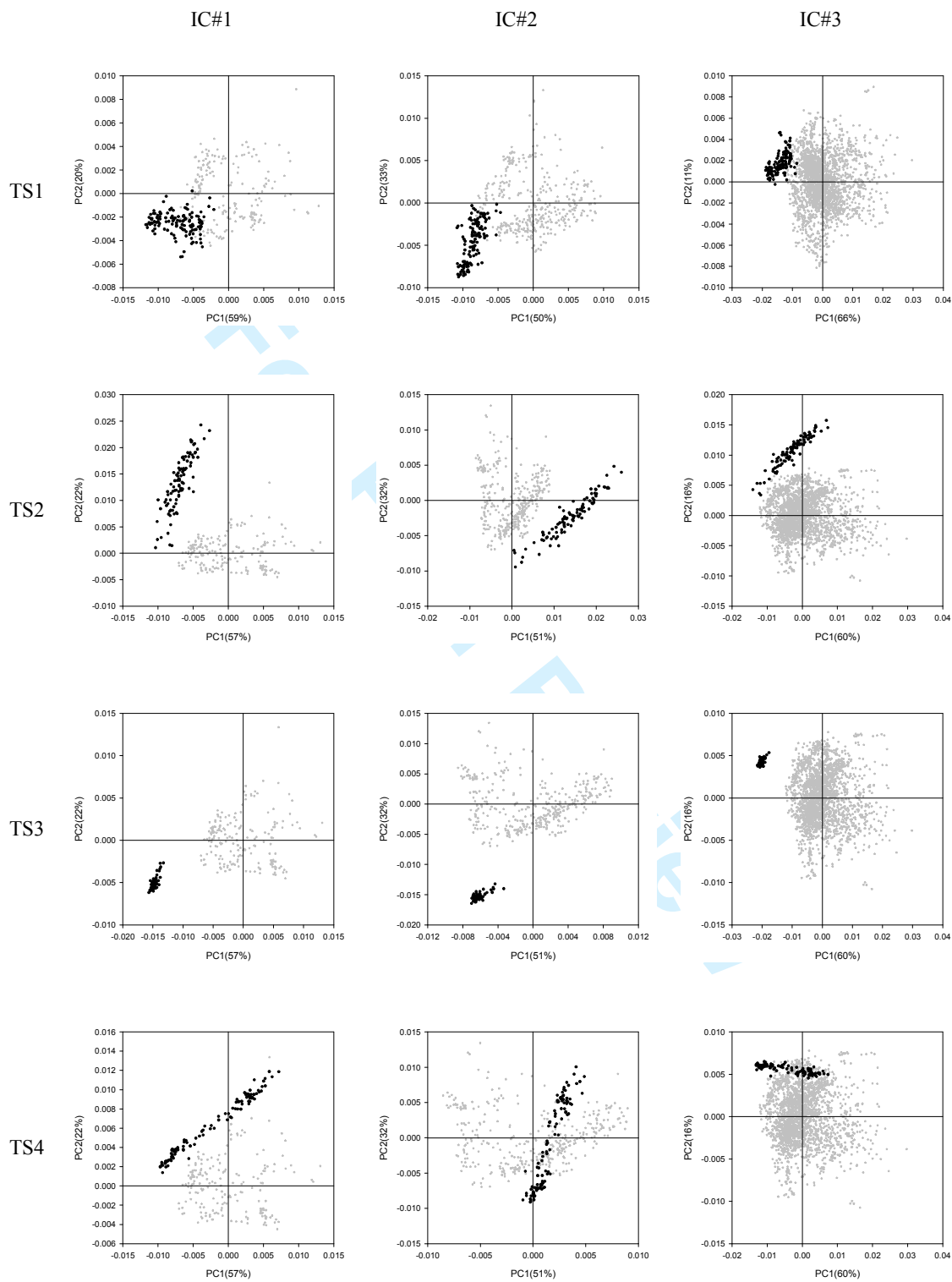
730

731 **Figure 5** Values of the root mean square error of prediction (RMSEP) obtained with the
732 three initial calibrations (IC) spiked with a spiking subset (SS) of size 8 (SS8), 16
733 (SS16) and 32 (SS32), without extra-weight (black bars), and with extra-weight (EW;
734 grey bars). Dark-grey bars are used when 24 copies of the spiking subset were added for
735 extra-weighting (EW_24), and light-grey bars are used when the numbers of copies
736 were added in proportion of the initial calibration to spiking subset ratio (EW_ratio).
737 White bars and horizontal lines were used to show the RMSEP obtained with
738 geographically local models, constructed uniquely with 8 (horizontal dotted line), 16
739 (horizontal dashed line) or 32 local samples (horizontal solid line). In all the cases, the
740 local samples were selected by the 'PC distrib' strategy. In all the cases $n = 4$ (from four
741 target sites). The error bars are denoting one standard deviation.

742

743

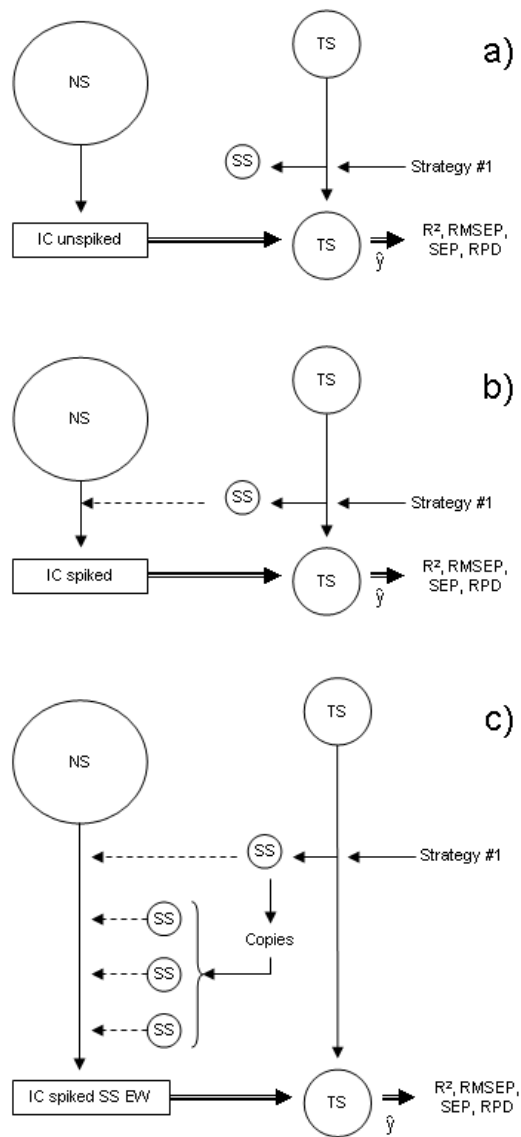
744



745 **Figure 1**

746

747



748

749 **Figure 2**

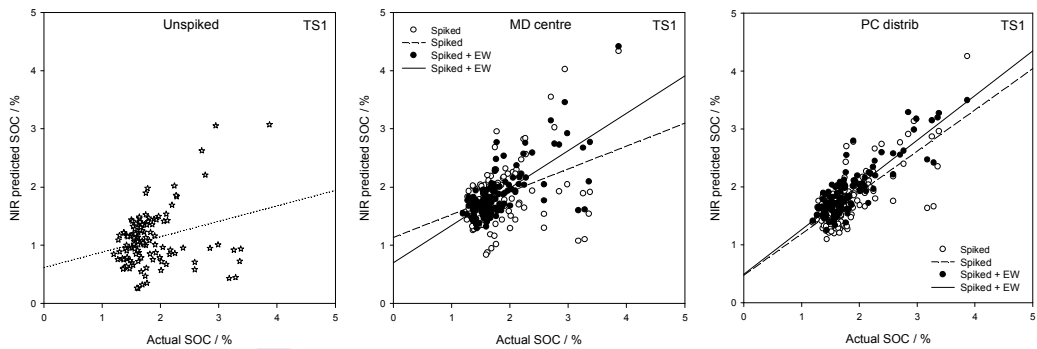
750

751

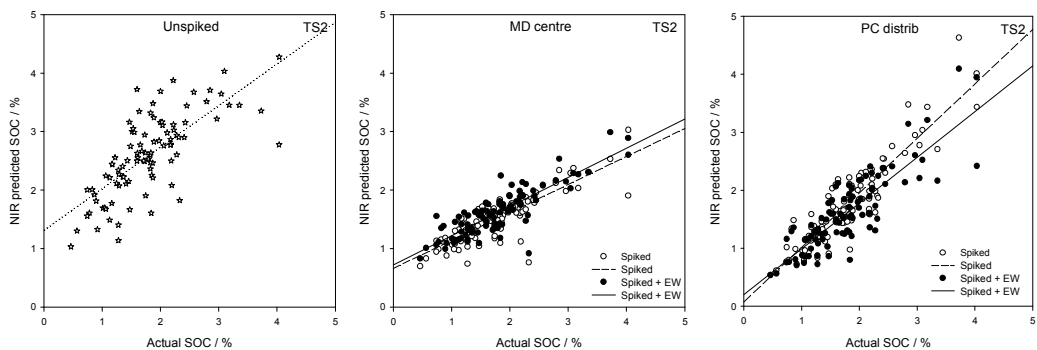
752

753

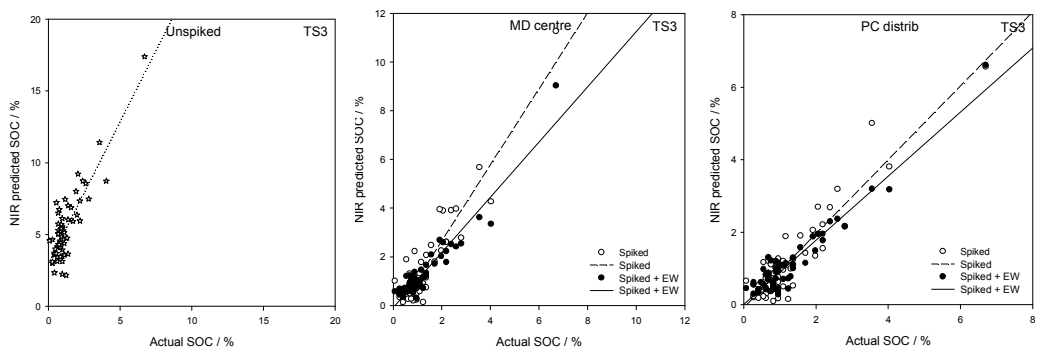
754



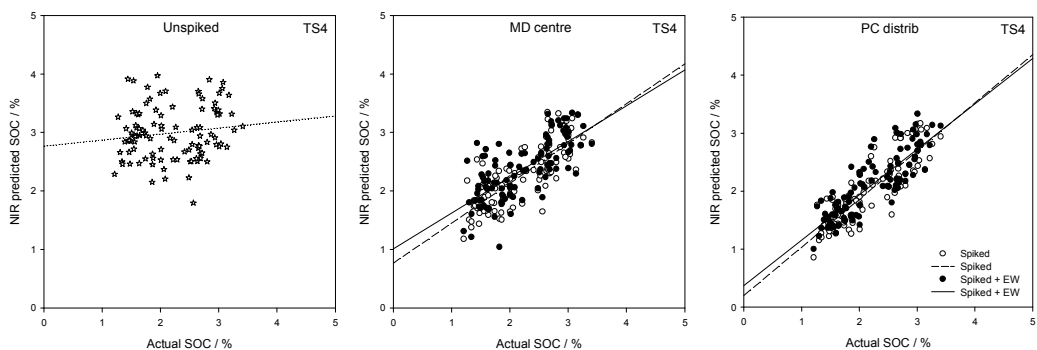
755



756



757

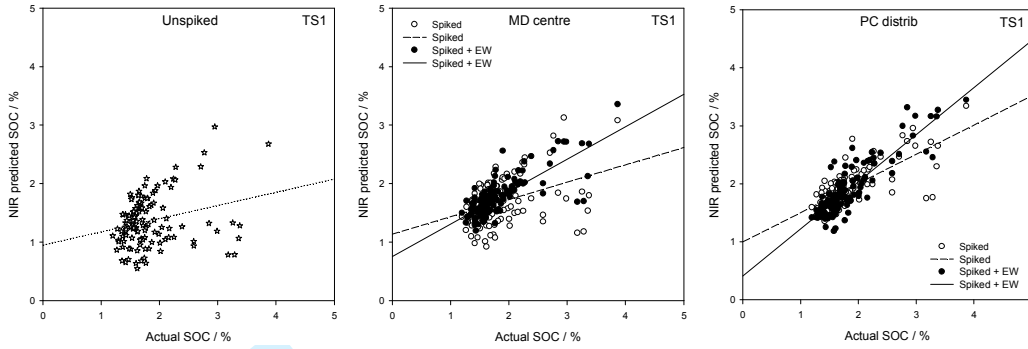


758

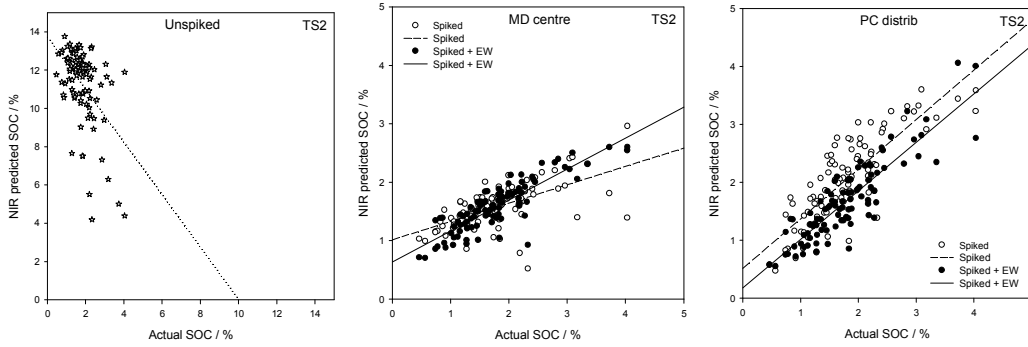
759 **Figure 3a**

760

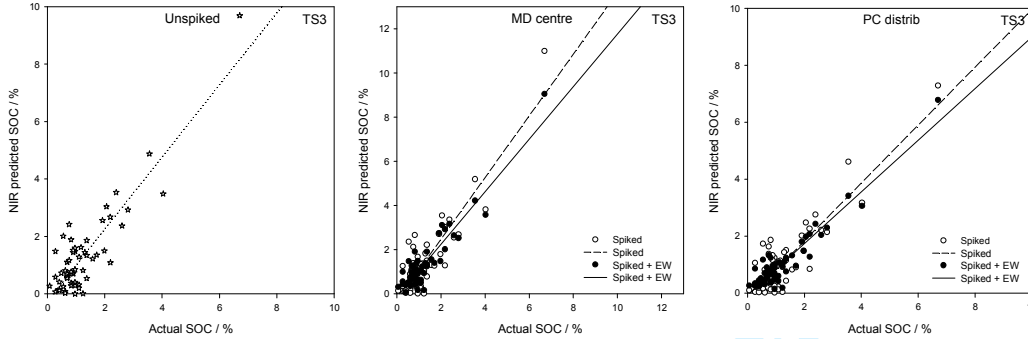
761



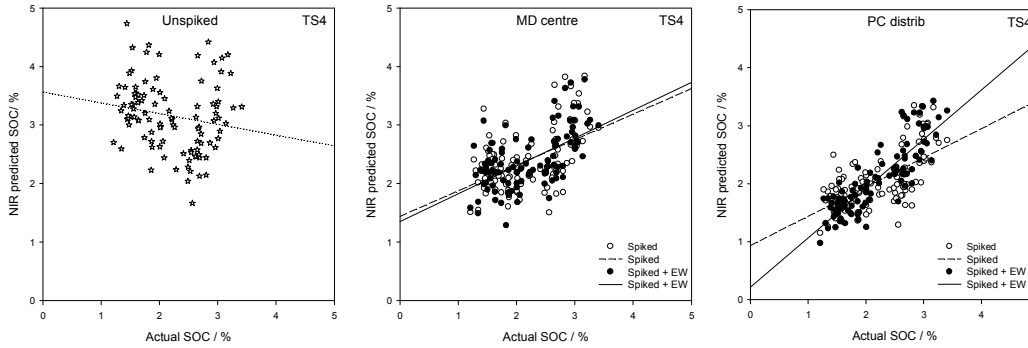
762



763



764

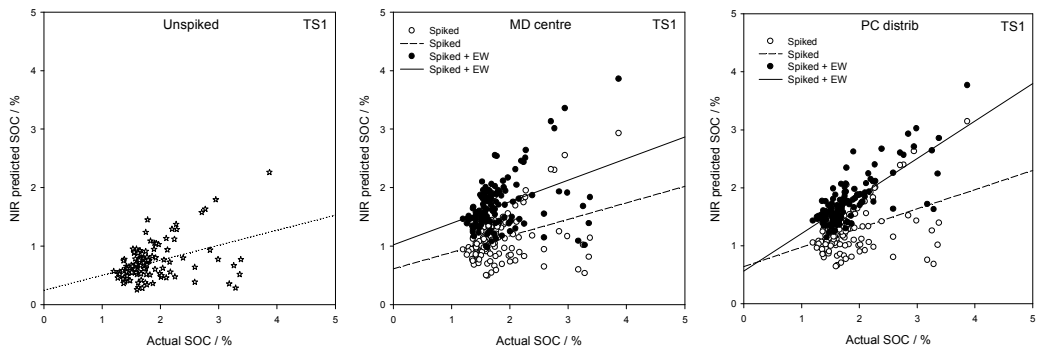


765

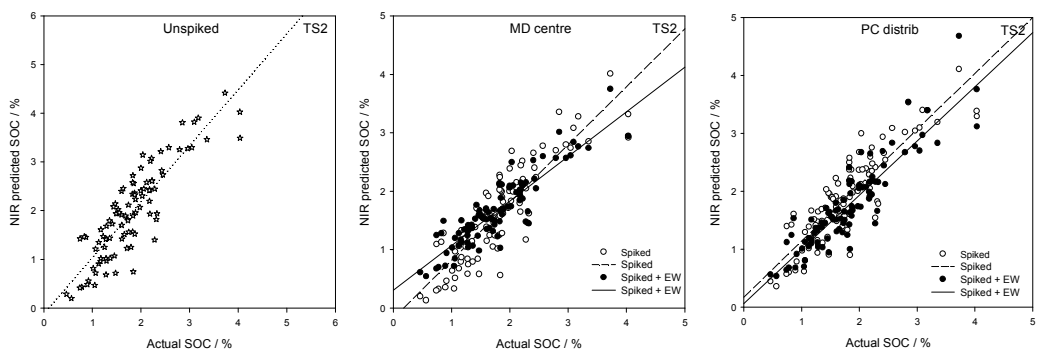
766 **Figure 3b**

767

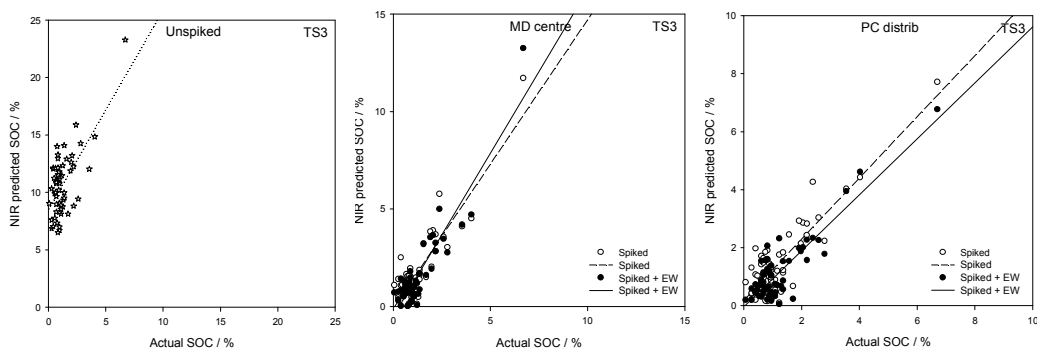
768



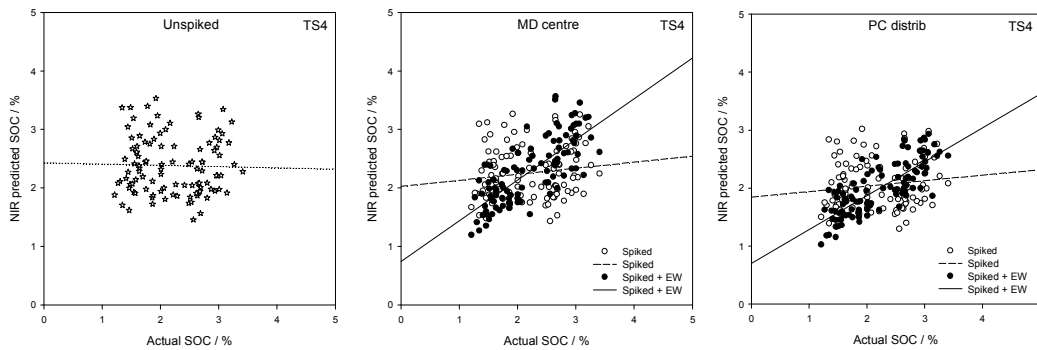
769



770



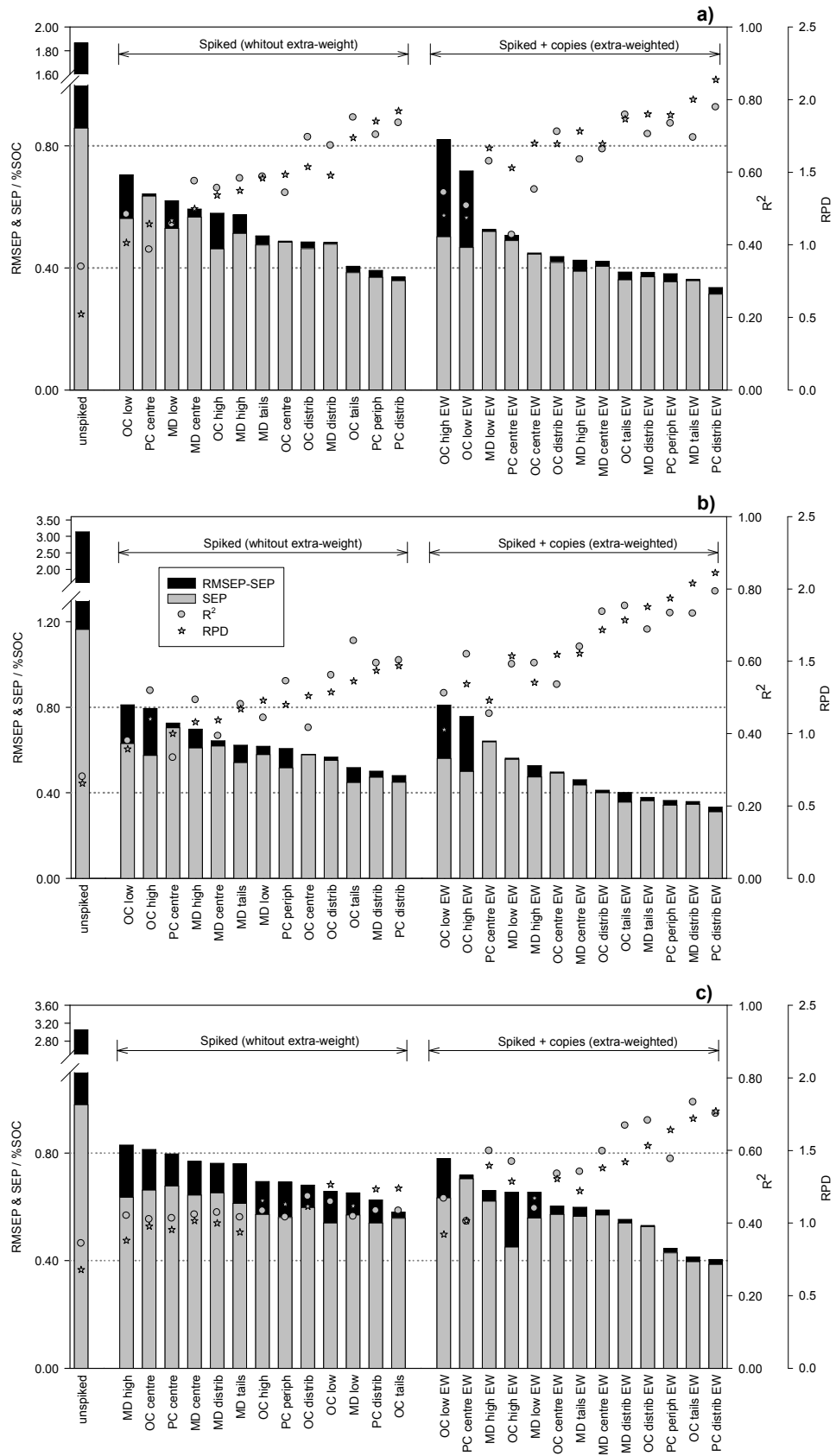
771



772

773 **Figure 3c**

774



775

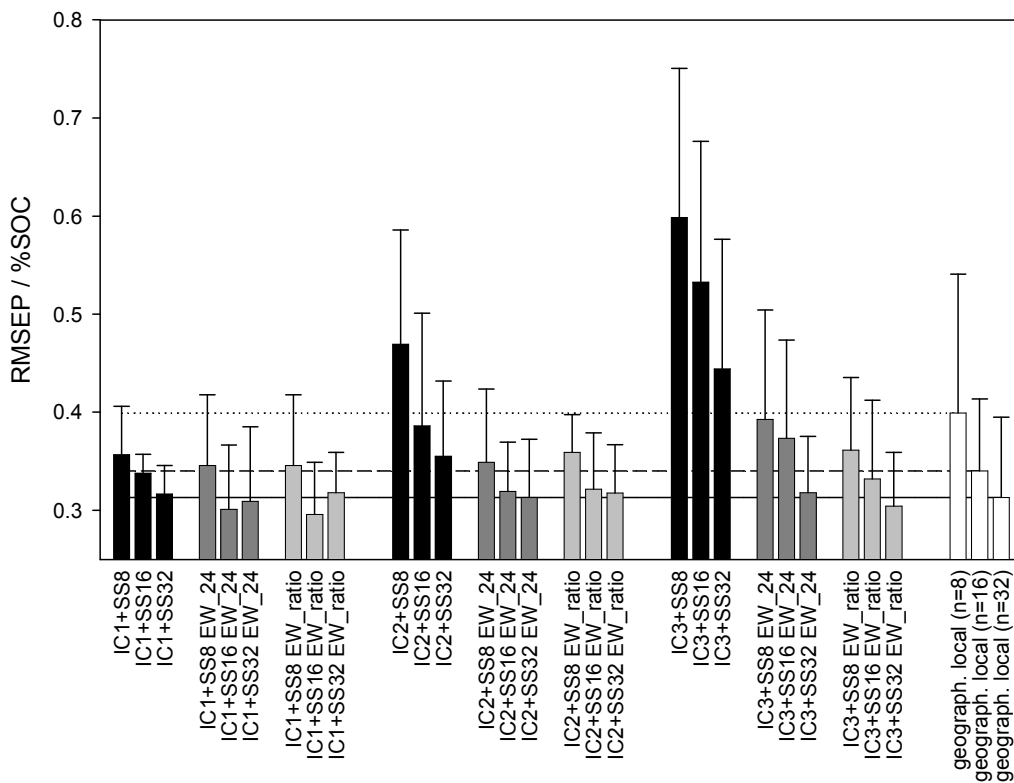
776

777

778

Figure 4

779



780

781 **Figure 5**

782

783

784

Review

785 TABLES

786

787 **Table 1** Characteristics of the three subsets used for the development of the different Initial
788 Calibrations (ICs), and the coefficient of determination (R^2) and root mean square error (RMSE)
789 obtained in the cross-validations (RMSECV). All the results refer to soil organic carbon (in %).
790

	IC #1	IC #2	IC #3
n	192	365	2279
Minimum	0.32	0.32	0.10
Maximum	8.97	14.49	14.62
Mean	2.35	5.07	1.54
Standard deviation	1.87	3.59	2.14
Skewness	1.05	0.41	3.20
R^2	0.95	0.96	0.93
RMSECV	0.40	0.67	0.54

791

792

793 **Table 2** Characteristics of the four target sites used. Data refer in all cases to soil organic carbon (SOC; %).

794

	Target site 1	Target site 2	Target site 3	Target site 4
Coordinates	55°41'N, 13°19'E	38°32'N, 0°49'W	37°09'N, 2°35'W	52°00'N, 0°26'W
Site (country)	Sjöstorp (Sweden)	Sax (Spain)	Gergal (Spain)	Silsoe (UK)
Parent material	Sandy till (25%) and sedimentary clay with elements of chalk (75%)	Gypsum	Mica schists	Mudstone
Method SOC	LOI ^a (900°C)	Elemental Analyser	Walkley & Black	LOI (900°C)
Spectral range / nm	1000-2500	834-2650	834-2650	834-2650
n	125	95	60	104
Minimum	1.20	0.47	0.07	1.21
Maximum	3.87	4.04	6.70	3.41
Mean	1.83	1.80	1.23	2.20
Standard deviation	0.50	0.71	1.05	0.60

795 ^a LOI: loss on ignition

796

797

798 **Table 3** Results of the repeated measures ANOVA to evaluate the effects of extra-weighting, initial calibration and strategy on the different prediction
799 performance parameters: root mean square error of prediction (RMSEP), standard error of prediction (SEP) and ratio of performance to deviance (RPD).

800

Variable	Source	Sum of squares	Degrees of freedom	Mean square	F	P	
RMSEP ^a	Between-subjects	Initial Calibration (IC)	0.605	2	0.302	11.84	0.0000
	Between-subjects	Strategy	0.701	7	0.100	3.918	0.0011
	Between-subjects	IC × Strategy	0.078	14	0.005	0.220	0.9985
	Between-subjects	Error	1.840	72	0.025		
	Within-subjects	Extra-weighting (EW)	0.668	1	0.668	81.90	0.0000
	Within-subjects	EW × IC	0.015	2	0.007	0.956	0.3890
	Within-subjects	EW × Strategy	0.045	7	0.006	0.794	0.5940
	Within-subjects	EW × IC × Strategy	0.085	14	0.006	0.751	0.7165
	Within-subjects	Error (EW)	0.587	72	0.008		
	SEP ^b	Between-subjects	IC	1.872	2	0.936	6.593
Between-subjects		Strategy	3.760	7	0.537	3.782	0.0015
Between-subjects		IC × Strategy	0.420	14	0.030	0.211	0.9988
Between-subjects		Error	10.22	72	0.142		
Within-subjects		EW	2.125	1	2.125	60.76	0.0000
Within-subjects		EW × IC	0.126	2	0.063	1.801	0.1725
Within-subjects		EW × Strategy	0.235	7	0.033	0.959	0.4673
Within-subjects		EW × IC × Strategy	0.306	14	0.021	0.626	0.8346
Within-subjects		Error (EW)	2.518	72	0.035		
RPD ^b		Between-subjects	IC	3.209	2	1.604	7.372
	Between-subjects	Strategy	3.716	7	0.531	2.439	0.0266
	Between-subjects	IC × Strategy	0.417	14	0.029	0.137	0.9999
	Between-subjects	Error	15.67	72	0.217		
	Within-subjects	EW	3.543	1	3.543	81.90	0.0000
	Within-subjects	EW × IC	0.082	2	0.041	0.956	0.3890
	Within-subjects	EW × Strategy	0.240	7	0.034	0.794	0.5940
	Within-subjects	EW × IC × Strategy	0.454	14	0.032	0.751	0.7165
	Within-subjects	Error (EW)	3.114	72	0.043		

801 ^a Log transformed802 ^b Ln transformed

803

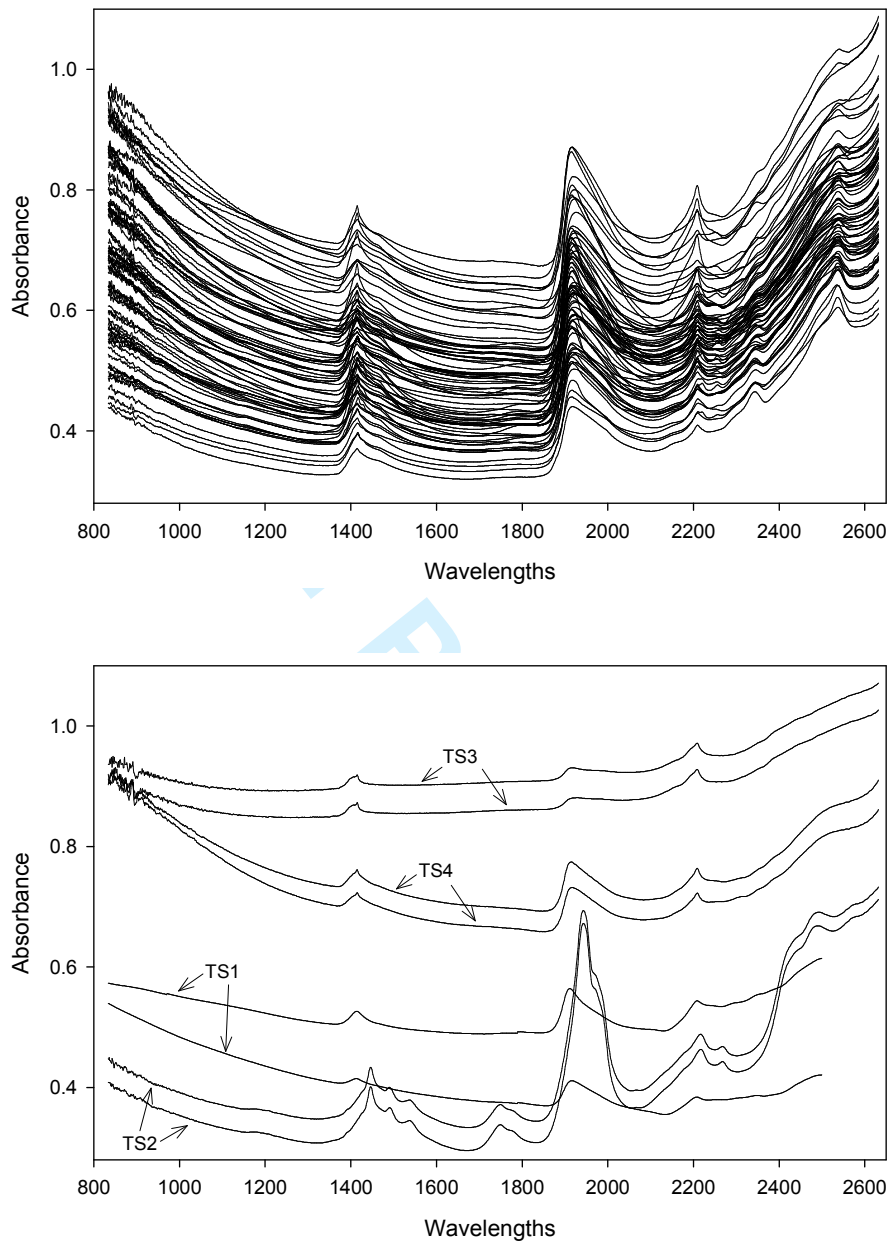
804 **Table 4.** Results of the repeated measures ANOVAs to evaluate the effects of the spiking subset size (SS-size), and those of the extra-weighting (EW) on the
 805 root mean square error of prediction (RMSEP) obtained with spiked calibrations. (a) Results obtained when 24 copies were used for EW (EW_24). (b)
 806 Results obtained when the number of copies to add for EW was equal to the ratio between the IC size and the SS size (EW_ratio).

807

		Source	Sum of squares	Degrees of freedom	Mean square	F	<i>P</i>
(a)	Between-subjects	SS-size	0.0696	2	0.0348	2.328	0.1133
	Between-subjects	Error	0.4936	33	0.0149		
	Within-subjects	EW_24	0.1341	1	0.1341	21.28	0.0000
	Within-subjects	EW_24 × SS-size	0.0087	2	0.0043	0.695	0.5058
	Within-subjects	Error	0.2079	33	0.0063		
(b)	Between-subjects	SS-size	0.0649	2	0.0324	3.117	0.0575
	Between-subjects	Error	0.3437	33	0.0104		
	Within-subjects	EW_ratio	0.1578	1	0.1578	18.45	0.0001
	Within-subjects	EW_ratio × SS-size	0.0119	2	0.0059	0.695	0.5058
	Within-subjects	Error	0.2821	33	0.0085		

808

809



Supplementary content: **Appendix I.** Representative NIR spectra of the national samples included in the initial calibrations (top), and two representative NIR spectra of each of the four target sites (bottom).

2014-01-17

Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset

Guerrero, C.

Wiley

Guerrero, C. et al. (2014) Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset. *European Journal of Soil Science*, Volume 65, Issue 2, pp. 248-263

<http://dx.doi.org/10.1111/ejss.12129>

Downloaded from Cranfield Library Services E-Repository