

1 Application of Non-Linear Statistical Tools to a Novel Microwave Dipole Antenna
2 Moisture Soil Sensor

3

4 A. Berardinelli^{a*}, G. Luciani^{b,c}, M. Crescentini^{b,d}, A. Romani^{b,d}, M. Tartagni^{b,d}, L. Ragni^{a,e}

5

6

7 ^aDepartment of Agricultural and Food Sciences, Alma Mater Studiorum, University of Bologna, P.zza Goidanich 60,
8 47521 Cesena (FC), Italy

9 ^bDepartment of Electrical, Electronic and Information Engineering "Guglielmo Marconi" - University of Bologna, Via
10 Venezia 52, Cesena 47521, Italy

11 ^cCenter for Industrial Research on ICT (CIRI ICT), University of Bologna, via Venezia 52, 47521 Cesena (FC), Italy.

12 ^dAdvanced Research Centre on Electronic Systems (ARCES), University of Bologna, 40123 Bologna, Italy.

13 ^eInter-Departmental Centre for Agri-Food Industrial Research, Alma Mater Studiorum, University of Bologna, P.zza
14 Goidanich 60, 47521 Cesena (FC), Italy

15 *Corresponding author.

16

17 *Abstract*

18

19 In this paper we will show the boosting performance of nonlinear machine learning techniques
20 applied to a novel soil moisture sensing approach. A probe consisting in a transmitting and a
21 receiving dipole antenna was set up to indirectly assess the moisture content (%) of three different
22 types of soils (silty clay loam, river sand and lightweight expanded clay aggregate, LECA). Gain
23 and phase signals acquired in the 1.0 GHz – 2.7 GHz frequency range were used to build predictive
24 models based on linear PLS regression and on nonlinear Kernel-based orthogonal projections to
25 latent structures (K-OPLS) algorithms. K-OPLS algorithm slightly increased the accuracy of the
26 models built on the gain response on specific kind of soils with respect to classical linear PLS.
27 However, the predictability increases significantly in the case where the models are built starting
28 from a matrix containing all the considered soil samples (silty clay loam + river sand + LECA)

29 achieving $R^2 = 0.971$ (RMSE = 1.4 %) when using K-OPLS non-linear model with respect to $R^2 =$
30 0.513 (RMSE = 6.1 %) obtained using linear PLS. Therefore, K-OPLS algorithm appears to give a
31 significant improvement to modelling data where nonlinear behaviours occur.

32

33 Keywords. Dielectric spectroscopy; soil moisture content; linear and nonlinear multivariate data
34 analysis; PLS; K-OPLS.

35

36

37 **1. Introduction**

38

39 The spectroscopic and time-domain analyses of the interaction between the electromagnetic wave
40 and the agricultural soil is, at date, widely explored methods for the indirect assessment of its water
41 content [1]. The acquired waveforms appeared to contain information related to different soil
42 physico-chemical properties and the quantitative estimation accuracy is affected by two main
43 factors as the used techniques and the statistical tools [2]. Examples of these techniques are Visible,
44 Near and Infrared sensors [2,4], Theta probes, measuring apparent impedance at 100 MHz [5] and
45 the Time Domain Reflectometry (TDR) [6], based on the analysis of the propagation time of the
46 electromagnetic wave through a coaxial cable to a probe immersed in the medium (20 kHz - 1.5
47 GHz), a function of the soil dielectric permittivity.

48 Powerful multivariate data analysis tools able to relate two data matrix X (spectra acquired from
49 several samples) and Y (the analytical properties), have played a big role in the development of the
50 techniques [7].

51 Originated around 1975, the widespread linear multivariate Partial Least Squares (PLS) regression
52 is considered a standard procedure in chemometrics and it has been shown to be potential for
53 extracting useful information starting from highly linearly correlated data coming from
54 bioengineering indirect measurements. The tool uses a two-block quantitative PLS model based on
55 a latent variable decomposition of X and Y variables keeping most of the variance of the

56 explanatory variables. It is well known that PLS regression has proven to be extremely useful in
57 situation when the number of observed variables is much higher than the numbers of acquired
58 samples, typical situation with spectral data [8].

59 However, non linear behaviours are very frequent in biosystems, such as the light absorbance in
60 milk, dependent on fat content [9], or the dielectric permittivity in microwave region, dependent on
61 the soil moisture [10], just to cite a couple of examples. Samples variability and level of complexity
62 of the matrices together with temperature fluctuations and interactions between sensor and product
63 can negatively affect the robustness of PLS models and cause non linear behaviours as shown in
64 different works conducted on quantitative assessment of fruits chemical properties through NIR
65 spectral measurements [11-13]. Agricultural soil is a complex heterogeneous matrix characterised
66 by organic (humus and different particulate residues) inorganic mineral fractions (proportions of
67 sand, silt and clay particles), moisture and air [14]. Conversely, multivariate regression models
68 based on non linear machine learning tools have shown significant improvements in the accuracy of
69 the prediction of different physical and chemical properties of this complex matrix [4,7,15].

70 In order to improve robustness of PLS models in presence of non linearity, a considerable number
71 of methods integrating non linear features within the linear PLS algorithm have been proposed.
72 Quadratic PLS [16], smooth bivariate spline function [17], Neural Network PLS [18], Radial Basis
73 Function (RBF) neural networks [19], and Kernel PLS (KPLS) [20] are some examples of the
74 proposed machine learning implementation in PLS modelling. In KPLS the original X variables are
75 transformed into a high-dimensional feature space by a non linear mapping. In this feature space a
76 linear relationship can be appreciated and the PLS algorithm can then be performed; the feature
77 space is defined after selecting a kernel function providing a similarity measure between pairs of
78 spectra [21]. The accuracy of the KPLS algorithms were tested with images analysis generated by
79 an airborne scanner with nine wavelength bands (from 500 to 10487 nm) [22], with UV-visible and
80 FT-IR spectra for the prediction of different mixtures contents [23] and with NIR spectra for the

81 prediction of apple sugar content [24] and for a rapid screening of water samples containing
82 malathion [24].

83 Our approach is substantially different with respect to the above-mentioned electro-magnetic
84 techniques. Differently from TDR, it is based on spectra analysis in the frequency domain instead of
85 the time-domain. Then, in contrast with commonly used IR spectra techniques, we perform a
86 spectral analysis of transfer functions involving microwaves. This ensures a better interaction with
87 soil in terms of depth of penetration and also utilizes higher information content given by the phase.
88 Finally, we use non-linear machine learning tools to boost the statistical inference of data.

89 A new probe in the dielectric sensors panorama characterised by a transmitting and a receiving
90 dipole antenna was set up for the indirect assessment of the moisture content of different types of
91 soils: silty clay loam soil, river sand and lightweight expanded clay aggregate (LECA). This
92 innovative probe requires the previously drilling of the soil and then the inserting of the probe. With
93 respect to traditional TRD probes it could less suffer for incomplete adherence of the soil to the
94 sensor. In fact, the sensing is performed in a large portion of the volume surrounding the probe and
95 any interference, such as air, can be removed by the powerful statistical analysis. Therefore, the
96 information contained in both gain and phase signals acquired in the 1.0 GHz – 2.7 GHz frequency
97 range, will be processed by using the Kernel-based orthogonal projections to latent structures (K-
98 OPLS, an implementation of KPLS with a solution able to separate structured noise). Predictive
99 models of the moisture content will be built starting from data sets characterised by the same soil
100 type or starting from data sets containing all the analysed soil types.

101

102 **2. Materials and methods**

103

104 *2.1 Probe and acquisition chain*

105

106 The probe, designed to be inserted in the soil, assembles a transmitting (TX) and a receiving (RX)
107 dipole antenna, spaced 50 mm, located in a 170 mm long PVC sealed pipe, with outer and inner
108 diameter of 16 and 13 mm, respectively. Both TX and RX antennas consists of a $\frac{1}{4}$ of ring per pole.
109 The dipole was mounted on a made of nylon ring and placed in the pipe rotated by 90° one with
110 respect to the other to avoid direct coupling of the EM signal from transmitting to receiving
111 antenna. A layout of the probe containing the dipoles is shown in Fig. 1a together with the
112 particulars of the dipole antenna (b) and the probe inserted in the soil (c). The above described
113 prototype was designed for moisture determination in the soil layer pertaining the secondary tillage.
114 A longer probe, containing an array of antennas, suitably spaced, could be constructed for in depth
115 stratified moisture assessment (Fig. 1d). The TX antenna was connected to a sweeper oscillator
116 (HP8350B combined with the HP83592B plug in), by means of a power divider. The signal from
117 the other output of the divider and that coming from the RX antenna were connected to a gain and
118 phase comparator (Analog Devices AD8302) through a 20 dB attenuator. The outputs of the
119 comparator give a measurement of both gain over a ± 30 dB range, scaled to 30 mV/dB, and of
120 phase between signals over a 0° – 180° range, scaled to 10 mV/degree. The output of the comparator
121 was connected to a sampling board (National instrument, DAQ USB-4431) with 24 bit of resolution
122 and sampling frequency from 1 kS/s to 102 kS/s. The board was connected to the PC. LabVIEW
123 software was used to display the spectrum and decimate the sampling frequency for reducing the
124 number of data. A layout of the instrumental chain was depicted in Fig. 2. The sinusoidal oscillation
125 (13 dBm) was linearly swept from a frequency of 1.0 GHz to 2.7 GHz in 60 s.

126

127

128 *2.2 Soil samples*

129

130 Waveform acquisition was conducted on three different soil samples: silty clay loam soil (collected
131 from Romagna region agricultural soil, Italy), river sand (Bacchi S.P.A., Italy), and lightweight

132 expanded clay aggregate (LECA) (Laterlite, Italy). According to USDA textural classification [26],
133 the chosen materials are characterised by very distinct physical properties (textural classes). Silty
134 clay loam soil is made of particles with the following size (s) distribution: $s < 0.002$ mm (34%);
135 $0.002 \leq s < 0.050$ mm (49%); $0.050 \leq s \leq 2$ mm (17%). Density of the silty clay soil (at 6% of
136 moisture) is 1137 kg/m^3 . River sand, classified as “sand”, consisting of grain with maximum size of
137 0.6 mm (density at 0.1% is 1371 kg/m^3). LECA consists of granules with dimensions from 4 to 10
138 mm (density at 0.2% of moisture is 380 kg/m^3) During measurements, silty clay loam samples were
139 characterised by clods of size suitable for seeding operations.

140 For each soil type, five different hydration levels (moisture contents, %) were considered (standard
141 error in brackets): 5.7 (0.02) %, 9.9(0.04)%, 15.9(0.21)%, 22.3(0.17)% and 27(0.50)% for silty clay
142 loam soil; 0.1(0.003)%, 4.8 (0.04)%, 9.6% (0.17), 14.5(0.41)% and 18.9 (0.38)% for river sand; 0.2
143 (0.01)%, 5.4(0.13)%, 13.1(0.24)%, 19.7(0.61)% and 28%(0.42) for LECA. The lowest level sample
144 was assessed by a thermo-gravimetric method [27]. Starting from this sample, the remaining ones
145 were produced by adding a specific amount of deionised water. The highest level to produce was
146 chosen taking into account the field capacity, evaluated on each type of soil according to Cihlar and
147 Ulaby [28]. To allow the water diffusion, the samples were hermetically isolated for about 48 hours
148 before the dielectric acquisitions.

149

150 *2.3 Experimental set up*

151

152 Each soil sample (type \times hydration level) was placed (constant volume of 17600 cm^3) in a plastic
153 cylindrical container ($20 \text{ cm} \times 14.4 \text{ cm}$) (Fig. 1). Spectral acquisitions were conducted at constant
154 temperature of $22^\circ\text{C} (\pm 1^\circ\text{C})$; all the spectra were sampled at 100 Hz from 1.0 GHz to 2.7 GHz (260
155 kHz of resolution). For each moisture level, nine different acquisitions were conducted. These
156 acquisitions were obtained by rotating the container with respect to its axis each time by an
157 approximately constant angle (40°). Having the temperature a noticeable effect on the complex

158 permittivity in the microwave range, a test was carried out on a silty clay loam with 16% of
159 moisture at 2 ° C and 22 ° C. Five measurement repetitions for each temperature were carried out
160 for this test.

161

162 2.4 Data analysis

163

164 For each type of soil and for both gain and phase spectra, data were arranged in a 45 (samples) ×
165 6536 (independent variables, gain or phase) X matrix and in a 45 (samples) × 1 (dependent variable,
166 moisture content %) vector Y column. Six (3 soil types × gain or phase) predictive models of the
167 moisture content (%) were obtained and discussed by means of PLS regression analysis of gain or
168 phase signals (MatlabR2016b, Statistics and Machine Learning Toolbox). *N*-fold cross validation
169 (random segments with about 14% of the samples in each segment) was used to validate
170 multivariate models and accuracy was expressed in terms of coefficient of determination R^2 and
171 Root Mean Square Error (RMSE). In *N*-fold cross validation, models are subsequently built from
172 the remaining samples (*n*-1 segments). According to Esbensen [29], *N*-fold cross validation can be
173 considered a powerful solution when there is a relative number of objects in the training data set.

174 The optimal number of PLS components was selected by analysing the validation residual variance
175 [29].

176 In order to improve the accuracy of the PLS prediction in presence of possible non linear
177 behaviours, the Kernel-based Orthogonal Projections to Latent Structures (K-OPLS) (K-OPLS
178 package for MATLAB, <http://kopls.sourceforge.net/>) was explored. The algorithm constructs a
179 regression model for predicting the dependent variable Y (moisture content, %) by using the
180 information contained in a Kernel matrix.

181 Orthogonal projection to latent structures (O-PLS) can improve the model interpretation and reduce
182 its complexity by removing systematic variability in X that is not correlated with Y or, in
183 mathematical terms, that is orthogonal to Y [30]. O-PLS can be considered as a combination of PLS

184 algorithm with a pre-processing tool and relative models are often characterised, for a given
185 accuracy, by a lower number of PLS components [31].

186 The transformation to higher dimensional spaces is performed by using a kernel function $k(x,y)$.

187 Particularly, the K-OPLS models were fitted using the Gaussian kernel function:

188

$$189 \quad k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$$

190

191 where σ parameter was selected according to a traditional approach by performing a grid search
192 over a defined parameter interval; at each setting, n -fold cross validation (random segments with
193 about 14% of the samples in each) was used to identify the parameter value yielding the error
194 minimisation [32]. N -fold cross validation (random segments with about 14% of the samples in
195 each segment) and errors plot were also used to identify the optimal number of predictive PLS
196 components and Y-orthogonal components.

197 The potentiality of PLS and K-OPLS algorithms were also explored with the aim of building
198 predictive models of soil moisture content independently from the soil type. To this purpose, two
199 135 (silty clay loam soil + river sand + LECA samples) \times 6536 (independent variables, gain or
200 phase) X matrices and two 135 (silty clay loam soil + river sand + LECA samples) \times 1 (dependent
201 variable, moisture content %) vector Y columns were processed.

202

203 **3. Results**

204

205 As an example, the acquired waveforms (average of 9 acquisitions for each hydration level) are
206 reported in Figure 3 in the 1.10 – 1.36 GHz frequency range for both gain and phase signals. For all
207 the analysed soil types, different moisture contents appeared to produce differences in the gain and
208 phase spectrum. As can be observed for the gain spectra, river sand gain are characterised by
209 significant variations and by a behaviour that can be evidently non linear. The waveform acquired at
210 temperature of 2°C and 22°C for a silty clay soil with 16 % of moisture were shown in Figure 4.

211 From this figure, shifts of the waveforms are observable from 1.18 GHz to 1.28 GHz and from 1.30
212 GHz, for gain and phase respectively. More evident differences were found in other spectral regions
213 (but not used for statistical analysis). All these difference can be attributed to the dependence of the
214 loss factor and dielectric constant on the temperature, but they are plausibly influenced by the
215 complex water-soil chemical-physical interactions. On the contrary, several spectral regions
216 appeared influenced by moisture content but not (or minimally) by the temperature, so the moisture
217 prediction algorithms could take advantage of this feature.

218 Main results of the PLS and K-OPLS regressions are summarised in Table 1 for gain and phase
219 models characterised by the same soil type and for those built starting from all the three analysed
220 soils. The models accuracy is described in terms of R^2 and RMSE obtained by performing n -fold
221 cross validations in the 1.10 – 1.36 GHz frequency range, resulted to produce, in general, the best
222 performances.

223 For classical linear PLS models, highest R^2 values emerged for phase acquisitions: 0.980 (RMSE =
224 1.1 %), 0.983 (RMSE = 0.9 %), and 0.981 (RMSE = 1.4 %) respectively for silty clay loam, river
225 sand, and LECA soils. Phase spectra appeared to be better linearly correlated to the dielectric
226 characteristics for all the three soil samples. Gain acquisitions on these soils are respectively
227 characterised by R^2 values of 0.970 (RMSE = 1.4 %), 0.960 (RMSE = 1.4 %), and 0.975 (RMSE =
228 1.6 %).

229 Passing from PLS to K-OPLS results, a significant improvement in the regression can be observed
230 in a single case, for river sand gain model ($R^2 = 0.988$, RMSE = 0.7 %). In this case, the K-OPLS
231 algorithm seemed to better interpret nonlinear variability observed among the five moisture levels.
232 As expected, all K-OPLS models are characterised by only one predictive PLS component. One
233 predictive PLS component appeared to be sufficient to discriminate river sand samples according to
234 the moisture content as shown in Figure 5a plotting the K-OPLS predictive score vector T_p against
235 the first Y-orthogonal score vector T_o . The hydration discriminatory direction is clearly described

236 by the predictive PLS component while systematic and linearly independent variations are modelled
237 by the first Y-orthogonal ones.

238 Predicted versus observed values of the moisture content (%) for river sand gain spectra obtained
239 from K-OPLS is given in Figure 5b (*N*-cross validation).

240 Respect to classical linear PLS regressions, K-OPLS phase models appeared to be characterised by
241 a significant lower accuracy for all soil types: 0.867 (RMSE = 2.3 %), 0.851 (RMSE = 2.0 %), and
242 0.946 (RMSE = 2.2 %) respectively for silty clay loam, river sand, and LECA soils.

243 For predictive models built starting from matrices including all soil types samples (silty clay loam +
244 river sand + LECA), the nonlinear multivariate approach K-OPLS clearly better interprets the
245 information correlated to the moisture content, respect to PLS regressions. Furthermore, the K-
246 OPLS algorithm seems to overcome problems related to systematic variations due to differences
247 among soil types. In *N*-fold cross validation, R^2 values of 0.971 (RMSE = 1.4 %) and 0.909 (RMSE
248 = 1.6 %) were observed respectively for gain and phase acquisitions. Predicted versus observed
249 moisture (%) are shown in Figure 6 for both gain and phase K-OPLS models. By performing PLS
250 regression analysis, significantly lower R^2 values were calculated: 0.513 (RMSE = 6.1 %) and
251 0.553 (RMSE = 5.8 %) respectively for gain and phase acquisitions (Fig. 7). By plotting scores
252 vectors (first component Tp-Up score plot including all soil types samples) obtained from both PLS
253 and K-OPLS gain models, a linear correlation can be better appreciated for K-OPLS plot respect to
254 PLS ones (Fig. 8).

255 The indirect prediction of the moisture content independently from the soil physical and chemical
256 characteristics it is not easily obtainable by traditional linear models. An example is provided by
257 Yin et al. [33] where a combination of 4 different soils can produce an R^2 value of 0.642 (RMSE up
258 to 9.26 % according to the soil type and for a range of 0 – 52 %) starting from a NIR reflectance
259 sensor. R^2 value of about 0.973 was also shown by Zanetti et al. [34] by using the apparent
260 dielectric constant (K_a) obtained from TDR waveforms and a combination of different physical

261 properties as input variables (bulk density, sand, silt, clay, and organic matter content) of ANN
262 models.

263

264 **4. Conclusions**

265

266 A novel approach for soil moisture sensing based on non-linear machine learning tools applied to
267 microwave spectra has been presented. A cylindrical dipole antenna probe operating in the low
268 frequency microwave region, together with gain and phase spectral data processed by linear and
269 nonlinear PLS statistical tools, shown to be a promising technique for soil moisture determination.
270 Validation R^2 values for basic PLS were from 0.960 (RMSE = 1.4 %) to 0.983 (RMSE = 0.9%),
271 depending on kind of soil and used spectra (gain or phase). Advanced K-OPLS algorithm allows to
272 greatly improve the prediction accuracy independently on the kind of soil ($R^2 = 0.971$, RMSE =
273 1.4%, gain data). The current probe could be developed for moisture determination at several depths
274 by equipping the sensor with an array of dipole antennas. Temperature can have a crucial influence
275 on the measured waveforms, so the calibration dataset will have to take into account this parameter.

276

277 **References**

278

279 [1] N. Romano, Soil moisture at local scale: measurements and simulations, *J. Hydrol.* 516 (2014)
280 6-20.

281

282 [2] S.L. S.U., D.N. Singh, M. Shojaei Baghini, A critical review of soil moisture measurement,
283 *Measurement* 54 (2014) 92-105.

284

- 285 [3] R.A. Viscarra Rossel, D.J.J. Walvoort, A.B. McBratney, L.J. Janik, J.O. Skjemstad, Visible,
286 near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous
287 assessment of various soil properties, *Geoderma* 131 (2006) 59-75.
- 288
- 289 [4] A. Morellos, X-E. Pantazi, D. Moshou, T. Alexandridis, R. Whetton, G. Tziotzios, J.
290 Wiebensohn, R. Bill, A.B. Mouazen, Machine learning based prediction of soil total nitrogen,
291 organic carbon and moisture content by using VIS-NIR spectroscopy, *Biosyst. Eng.* 152 (2016)
292 104-116.
- 293
- 294 [5] D.A. Robinson, C.M.K. Gardner, J.D. Cooper, Measurement of relative permittivity in sandy
295 soils using TDR, capacitance and theta probes: comparison, including the effects of bulk soil
296 electrical conductivity, *J. Hydrol.* 223 (1999) 198-211.
- 297
- 298 [6] D.A. Robinson, S.B. Jones, J.M. Wraith, D. Or, S.P. Friedman, A review of advances in
299 dielectric and electrical conductivity measurement in soils using time domain reflectometry, *Vadose*
300 *Zone J.* 2 (2003) 444-475.
- 301
- 302 [7] A.M. Mouazen, B. Kuang, J. De Baerdemaeker, H. Ramon, Comparison among principal
303 component, partial least squares and back propagation neural network analyses for accuracy of
304 measurement of selected soil properties with visible near infrared spectroscopy, *Geoderma* 158
305 (2010) 23-31.
- 306
- 307 [8] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometr.*
308 *Intell. Lab. Syst.* 58 (2001) 109-130.
- 309

- 310 [9] L. Ragni, E. Iaccheri, C. Cevoli, A. Berardinelli, Spectral-sensitive pulsed photometry to predict
311 the fat content of commercialized milk, *J. Food Eng.* 171 (2016), 95-101.
312
- 313 [10] A. Szyplowska, J. Szerement, A. Lewandowski, M. Kafarski, A. Wilczek, W. Skierucha,
314 Impact of soil salinity on the relation between soil moisture and dielectric permittivity, in:
315 Proceedings of the 12th International Conference on Electromagnetic Wave Interaction with Water
316 and Moist Substances, Lublin, Poland 4-7 June 2018, pp. 14-16.
317
- 318 [11] A. Peirs, J. Tirry, B. Verlinden, P. Darius, B.M. Nicolai, Effect of biological variability on the
319 robustness of NIR models for soluble solids content of apples, *Postharvest Biol. Technol.* 28 (2003)
320 269-280.
321
- 322 [12] A. Peirs, N. Scheerlinck, B.M. Nicolai, Temperature compensation for near infrared
323 reflectance measurement of apple fruit soluble solids contents, *Postharvest Biol. Technol.* 30 (2003)
324 233-248.
325
- 326 [13] F. Chauchard, R. Cogdill, S. Roussel, J.M. Roger, V. Bellon-Maurel, Application of LS-SVM
327 to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for
328 acidity prediction in grapes, *Chemometr. Intell. Lab. Syst.* 71 (2004) 141-150.
329
- 330 [14] J.O. Skjemstad, P. Clarke, A. Golchin, J.M. Oades, Characterization of soil organic matter by
331 solid-state ¹³C NMR spectroscopy, in: G. Gadish, K.E. Giller (Eds.), *Driven by nature: plant litter*
332 *quality and decomposition*, CAB International, Wellington (UK), 1997, pp. 253-271.
333

- 334 [15] B. Stenberg, R.A. Viscarra Rossel, A.M. Mouazen, J. Wetterlind, Visible, and near infrared
335 spectroscopy in soil science, in: D. L. Sparks (Ed.), *Advances in Agronomy*, Academic Press,
336 Burlington, 2010, pp. 163-215.
- 337
- 338 [16] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Non-linear PLS modelling, *Chemometr. Intell.*
339 *Lab. Syst.* 7 (1989) 53-65.
- 340
- 341 [17] S. Wold, Non-linear Partial Least Squares Modelling II. Spline Inner Function, *Chemometr.*
342 *Intell. Lab. Syst.* 14 (1992) 71-84.
- 343
- 344 [18] G. Baffi, E.B. Martin, A.J. Morris, Non-linear projection to latent structures revisited (the
345 neural network PLS algorithm), *Comput. Chem. Engin.* 23 (1999) 1293-1307.
- 346
- 347 [19] D.J.H. Wilson, G.W. Irwin, G. Lightbody, Nonlinear PLS modelling Using Radial Basis
348 Functions, American Control Conference, Albuquerque, New Mexico, June 4-6,1997.
- 349
- 350 [20] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in Reproducing Kernel Hilbert
351 Space, *J. Mach. Learn. Res.* 2 (2001) 97-123.
- 352
- 353 [21] R. Rosipal, Kernel partial least squares for nonlinear regression and discrimination, *Neural*
354 *Netw. World* 13 (2003) 291-300.
- 355
- 356 [22] F. Lindgreen, P. Geladi, S. Wold, Kernel-based PLS regression; Cross-validation and
357 applications to spectral data, *J. Chemom.* 8 (1994) 337-389.
- 358

- 359 [23] S. Rännar, P. Geladi, F. Lindgren, S. Wold, A PLS kernel algorithm for data set with many
360 variables and few objects. Part II: Cross validation, missing data and examples, *J. Chemom.* 9
361 (1995) 459-470.
- 362
- 363 [24] B.M. Nicolai, K.I. Theron, J. Lammertyn, Kernel PLS regression on wavelet transformed NIR
364 spectra for prediction of sugar content of apple, *Chemometr. Intell. Lab. Syst.* 85 (2007) 243-252.
- 365
- 366 [25] C. Gu, B. Xiang, Y. Su, J. Xu, Near-Infrared spectroscopy coupled with Kernel Partial Least
367 Squares-Discriminant Analysis for rapid screening water containing malathion, *Am. J. Analyt.*
368 *Chem.* 4 (2013) 111-116.
- 369
- 370 [26] Soil Survey Staff, Soil survey manual, Soil Conservation Service U.S. Department of
371 Agriculture Handbook, 18, 1993.
- 372
- 373 [27] ASTM D2216, Standard Test Methods for Laboratory Determination of Moisture (Moisture)
374 Content of Soil, ASTM International, West Conshohocken, PA, 2008.
- 375
- 376 [28] J. Cihlar, F.T. Ulaby, Dielectric properties of soils as a function of moisture content, Kansas
377 Univ. Center for Research, Inc. 1974, NASA-CR-141868.
- 378
- 379 [29] K. H. Esbensen, *Multivariate Data Analysis – in practice. An introduction to multivariate data*
380 *analysis and experimental design*, 5th edition, CAMO Software AS, 1994.
- 381
- 382 [30] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), *J. Chemometr.* 16
383 (2002) 119-128.
- 384

- 385 [31] D.J. Biagioni, D.P. Astling, P. Graf, M.F. Davis, Orthogonal projection to latent structures
386 solution properties or chemometrics and systems biology data, *J. Chemom.* 25 (2011) 514-525.
387
- 388 [32] M. Bylesjö, M. Rantalainen, J.K. Nicholson, E. Holmes, J. Trygg, K-OPLS package: Kernel-
389 based orthogonal projections to latent structures for prediction and interpretation in feature space.
390 *BMC Bioinformatics*, 9 (2008) 106.
391
- 392 [33] Z. Yin, T. Lei, Q. Yan, Z. Chen, Y. Dong, A near-infrared reflectance sensor for soil surface
393 moisture measurement, *Comput. Electron. Agric.* 99 (2013) 101-107.
394
- 395 [34] S.S. Zanetti, R.A. Cecílio, V.H. Silva, E.G. Alves, General calibration of TDR to assess the
396 moisture of tropical soils using artificial neural networks, *J. Hydrol.* 530 (2015) 657-666.
397
398