

1 **Rapid assessment of As and other elements in naturally-contaminated calcareous**
2 **soil through Hyperspectral VIS-NIR analysis**

3 F. Pallottino^{†a}, S.R. Stazi^{†*b}, A. D'Annibale^b, R. Marabottini^b, E. Allevato^b, F.
4 Antonucci^a, C. Costa^a, **M. C. Moscatelli^b**, P. Menesatti^a

5
6 ^a*Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA),*
7 *Centro di ricerca Ingegneria e Trasformazioni agroalimentari, Via della Pascolare 16,*
8 *00015 Monterotondo, Roma, Italy.*

9 ^b*DIBAF, University of Tuscia, Via S.C. De Lellis snc, 01100, Viterbo, Italy.*

10
11 dual-first authorship: the [†]authors equally contributed to the manuscript drafting

12
13 **Correspondence:* Silvia Rita Stazi

14 E-mail: srstazi@unitus.it

15 mobile: +393930071671

16 University of Tuscia - Via San Camillo de Lellis, snc -0110 Viterbo-Italy

17
18
19 Declarations of interest: none.

22

23 **Abstract**

24 Although arsenic (As) toxicity in soil vary depending on its chemical forms and
25 oxidation states, regulatory limits for this compartment rely on total As content.
26 Conventional methods of total As determination are expensive and time-consuming.
27 The development of predictive techniques might enable a speditive assessment of As
28 contamination in those scenarios, such as thermal spring sites, where exposure to the
29 metalloid poses a threat to human health. The objective of this study was to assess the
30 suitability of Visible Near Infrared spectrophotometry for predicting the total As content
31 in highly calcareous thermal spring soils and the same aim was pursued for those
32 elements (i.e. Al, Fe and Mn) the chemistry of which is tightly connected with that of
33 As. A Partial Least Square approach, including cross-validation and external
34 independent test, was used to relate the concentrations of the target elements to spectral
35 data. The most accurate prediction was found for As with Pearson's coefficient, RMSE,
36 RPD and SEP being equal to 0.94, 69.65, 2.9 and 66.99, respectively. Less accurate
37 predictions were found for Al ($r=0.88$; RMSE= 11014; RPD = 1.96; SEP=11014), Fe
38 ($r=0.93$; RMSE = 6921.1; RPD = 2.45; SEP=6462.4), and Mn ($r= 0.92$; RMSE =
39 542.01; RPD= 2.43; SEP=529.79).

40

41 **Keywords:** As polluted soil; Hyperspectral spectrophotometry; Visible Near Infrared;
42 hot springs; speditive soil analysis

43

44 **1. Introduction**

45 Arsenic (As) is a toxic element with carcinogenic capability in humans, causing cancer
46 of the skin, lungs, liver, and bladder (WHO, 2001). Its toxicity depends mainly on its
47 chemical form (organic or inorganic) and oxidation state (Bissen and Frimmel, 2003).
48 The widespread occurrence of As is due to both natural processes and anthropogenic
49 sources. In soil and related environments, the transformation and mobility of As is
50 governed by abiotic and biotic processes (Huang and Gobran, 2005; Violante et al.,
51 2008). Several studies demonstrated the close association of As with soil colloids, such
52 as clay minerals, Al, Fe and Mn (hydro)oxides and carbonates (Sadiq, 1997; Mehmood
53 et al., 2009). The adsorption capacities of these colloids are affected by pH and redox
54 conditions in soil and by dynamically changing factors, such as crystallinity, hydration,
55 isomorphic replacement and changes in cation coordination. For this reason, the
56 adsorption of As in soil has been defined as a complex and, sometimes, controversial
57 issue (Sadiq, 1997). In general, the surfaces of Al and Mn (hydro)oxides have been
58 shown to play a role in As adsorption only in acidic soils (Mehmood et al., 2009).
59 However, it has been shown that specific adsorption (ligand exchange mechanism) or
60 chemisorption processes might occur on surfaces of these constituents (Sadiq, 1997).
61 Moreover, chemisorption of As oxyanions on the surfaces soil colloids, especially those
62 of Fe oxide/hydroxides and carbonates, is deemed to be a usual mechanism for As solid
63 phase formation (Sadiq, 1997; Yang and Donahoe, 2007). Under alkaline conditions,
64 the desorption of arsenate from iron-oxide surfaces is favoured (Dixit and Hering, 2003;
65 Mamindy-Pajany et al., 2009) since at $\text{pH} > 8.0$ the negative net surface charge of iron
66 oxide repels negatively charged ions such as arsenate.

67 Goldberg and Glaubig (1988) reported that calcite decreased the mobility of As under
68 alkaline conditions via the incorporation of the metalloid into its lattice structure. The

69 affinity of As for calcite surface sites, at alkaline pH, was also confirmed by
70 Alexandratos et al. (2007) who showed the ability of the metalloid to form an inner-
71 sphere complex at the calcite surface.

72 The development of rapid techniques for assessing As contamination in soil, such as
73 hyperspectral spectrophotometry, would have significant advantages over conventional
74 methods. The ultimate benefits of these alternative approaches would be to reduce
75 analytical costs and introduce a rapid and easy to use measurement technique. Inorganic
76 toxic elements in soils are absorbed or bound by spectrally active constituents (Song et
77 al., 2012) and the visible–near infrared (VIS-NIR) region has been reported to enable a
78 reliable determination of some of them (Kemper and Sommer, 2002; Wu et al., 2005).
79 The spectral changes in this region are primarily due to iron-containing minerals, such
80 as, for instance, hematite and goethite (Wu et al., 2005). Moreover, the presence of
81 various vibrational features of the soil mineral components [e.g., Fe–oxygen (O), Al–O,
82 Si–O, and hydroxide (OH)] have been shown to contribute to the VIS-NIR spectral
83 characterization of soils. Although As is spectrally neutral within the VIS-NIR region,
84 the spectral signatures of As-binding minerals were successfully used for the indirect
85 estimation of the metalloid in remote sensing applications (Choe et al, 2008). On the
86 same basis, the As content in soil was predicted with good accuracy by the use of VIS-
87 NIR spectral reflectance (Stazi et al., 2014); in that case, however, the soil had a clay
88 loam texture and was artificially spiked with As.

89 To date, the diffuse VIS-NIR spectral reflectance method has not been applied to the
90 prediction of As in naturally-contaminated and highly calcareous soils. This is the case
91 of thermal spring sites of volcanic origin characterized frequently by the geogenic
92 accumulation of As. As a matter of fact, several areas around the world are

93 characterized by the widespread presence of As-contaminated geothermal systems
94 (López et al., 2012; Ormachea et al., 2015; Bundschuh and Maity, 2015). Moderate to
95 high levels of As, in fact, are common in soils around thermal spring sites which are
96 attended by people for recreational purposes. Thus, to address exposure risks to As of
97 the attendees of these sites, it is important to monitor frequently its concentration
98 through the use of speditive approaches, such as proximal sensing techniques. In
99 addition to ingestion, in fact, the exposure pathways of As also involve dermal contact
100 (Kamunda et al., 2016) and the hazards derived from short-term exposure to this
101 metalloid in contaminated soil has been reported (White, 1999).

102 For these reasons, this study was aimed at assessing the feasibility of VIS-NIR image
103 based reflectance technique in predicting the concentration of As in highly calcareous
104 soils collected from the immediate proximity of the thermal springs of the Viterbo area
105 (Baiocchi et al., 2012). Moreover, the prediction ability of the technique was also
106 assessed for Al, Fe and Mn concentrations, taking into consideration their
107 aforementioned interactions with As in soil. To this aim, hyperspectral VIS-NIR
108 spectrophotometry was combined with multivariate statistical analysis, based on partial
109 least squares (PLS) regression.

110

111 **2. Materials and Methods**

112 *2.1 Study area and soil sampling*

113 This study was conducted in the Bullicame thermal springs area (Viterbo, Central Italy).
114 Viterbo is a naturally As-rich area, as a result of the geochemical mobilization of this
115 metalloid through hydrothermal processes that lead to the up-flow of thermal waters
116 with an As concentrations up to 300 $\mu\text{g L}^{-1}$. The area is characterized by sandy-loamy
117 soils classified as Haplic Calcisol (FAO WRB, 1998) where the presence of this

118 metalloids is due to a primary enrichment from volcanic parent material and to a
119 secondary enrichment caused by the rising of deep fluids generated by the hydrothermal
120 geological activity. Soil samples were collected, at 0-20 cm depth, in 5 different sites
121 namely S1, S2, S3, S4 and S5 from the closest to farthest distance (till 350 m) from the
122 water springs in order to have an As concentration gradient. In particular, after the
123 removal of the turf, 9 soil cores (approx. 500 g of soil's fresh weight) were collected
124 with the aid of a soil corer from each site and each one was transferred to a labelled bag;
125 as a consequence, 45 samples were obtained in total. All the samples underwent air-
126 drying at room temperature for 4 days prior to sieving (2-mm) and, finally, crushing
127 with a planetary ball mill (RETSCH PM 100, RETSCH GmbH, Haan, Germany). Each
128 soil sample was split into two aliquots, one of which destined to chemical determination
129 of soil physico-chemical properties and the other one to spectral measurements.

130

131 *2.2 Soil chemical analysis*

132 Total organic carbon (TOC) and total nitrogen (TN) were determined using an
133 elemental analyzer (Thermo Soil NC Flash EA1112). Total carbonate content was
134 measured by the Calcimetry method. Active CaCO_3 was determined with 0.1 M NH_4 -
135 oxalate (Drouineau, 1942).

136 Active and exchangeable acidity were measured on sieved soil suspended either in a
137 solution of deionised water (active) or in 1 N KCl (exchangeable), respectively, in 1:2.5
138 (w/v) ratio. The pH was measured in the supernatant with a pH meter (pH 211, Hanna
139 Instruments). All analyses were performed on the original samples.

140 The elements under study were extracted with a microwave-assisted digestion. Nitric
141 acid (69% HNO_3) was of SupraPur grade (Merck, Darmstadt, Germany). High-purity
142 water (18 M Ω cm) from a Milli-Q water purification system (Millipore, Bedford, USA)

143 was used for dilution of the standards, for preparing samples throughout the chemical
144 process, and for final rinsing of the acid-cleaned vessels, glasses, and plastic utensils.

145 The external calibration solutions were prepared from standard certified multi-elemental
146 solutions (CaPurAn, CPA chem, Bulgaria) and Milli-Q water containing 3% HNO₃.
147 As₂O₃ standard was purchased from CaPurAn. Yttrium was used as the internal
148 standard.

149 The purity of the plasma torch argon was greater than 99.99%. The accuracy of the
150 determinations was assessed using the trace metals loamy sand 3 standard reference
151 material (CRM034–Fluka).

152

153 *2.3 Digestion procedure and Al, Fe, As and Mn quantification in soil*

154 **The soil aliquot destined to metal(loid) analysis was subjected to coning and quartering**
155 **and the resulting material (approx. 20 g) was digested in duplicate.** Each sample (500
156 mg) was transferred directly into a PTFA microwave-closed vessel and the SupraPur
157 grade nitric acid (HNO₃ 69%) (Merck, Darmstadt, Germany) solution (10 ml) was
158 added to each vessel. The heating program was as follows: from 25 to 165 °C in 10 min,
159 hold at 165 °C for 2 min; increase from 165 °C to 180 °C in 6 min and, finally, hold at
160 180 °C for 10 min. After the digestion procedure and subsequent cooling, the digested
161 samples were diluted to a final volume of 50 ml with High-purity water (18 MΩ cm)
162 from a Milli-Q water purification system (Millipore, Bedford, USA). Blanks were
163 prepared for each batch of samples. The acid digestion of the soils samples was
164 performed using a commercial high-pressure laboratory microwave oven (Mars plus
165 CEM, Italy) operating at an energy output of 1800 W.

166 For elements determination, an inductively coupled plasma optical emission
167 spectrometer (ICP OES) with an axially viewed configuration (Optima 8000DV, Perkin
168 Elmer) and equipped with a Scott nebulizer, was used. For the detection of the elements
169 under study, the following emission lines were chosen: Al (237.31 nm), As (193.7 nm),
170 Fe (238.2 nm) and Mn (257.6 nm). The external calibration solutions were prepared
171 from standard certified multi-elemental solutions (CaPurAn, CPA Chem, Bulgaria) and
172 Milli-Q water containing 3% HNO₃. As₂O₃ standard was purchased from CaPurAn.
173 Yttrium was used as the internal standard. The purity of the plasma torch argon was
174 greater than 99.99%. The accuracy of the determinations was assessed using the trace
175 metals loamy sand 3 standard reference material (CRM034–Fluka). **The 2**
176 **aforementioned mineralized solutions related to each sampling point were analysed in**
177 **duplicate thus resulting in 4 ICP-OES determinations per sampling point.**

178 A sequential extraction of soil was done in order to highlight the distribution of As with
179 reference to its association with solid phases. A specific procedure for calcareous soils
180 was carried out according to Matera et al. (2003), with minor modification. The
181 extraction scheme released the following eight fractions: water-soluble As, As soluble
182 in MgCl₂, As bound to carbonates, As bound to Mn-oxides, As bound to amorphous Fe
183 oxides, As bound to crystalline Fe oxides, As bound to organic matter and sulphides,
184 Residual As.

185

186 *2.4 VIS-NIR Spectral analysis*

187 Each soil sample was placed and replicated into two borosilicate optical-glass Petri
188 dishes (Duraplan®) and then displaced against a black background. Since punctual
189 spectral analysis is extremely less informative than image analysis, **a HSI**
190 **(Hyperspectral Imaging) scanner was adopted** to acquire a whole hyperspectral image of

191 the two Petri dishes, belonging to each sampling site. These instruments are highly
192 informative generating huge amount of data per acquisition. Each scan produces the so-
193 called data hypercube were, the x and y axes represent the pixel position values within
194 the acquired image, and the third axis (λ) represents the spectral values obtained step by
195 step (5 nm in this case) following the instrument spectral resolution. Indeed, two
196 circular ROI (Region of Interest) were extracted and the resulting pixels were averaged
197 step by step obtaining a 2D matrix. This approach produces extremely stable results in
198 comparison with conventional punctual systems that suffer from the variability owing to
199 the acquired small point.

200 In detail, for the spectral analysis was used a Spectral Scanner (DV Optics, Padua, Italy)
201 able to acquire whole images with spectral range from 420 to 950 nm (spectral
202 resolution equal to 5 nm) as previously reported by Menesatti et al. (2010). The VIS-
203 NIR spectrophotometer utilized is made of four components: 1) a sample transportation
204 platter; 2) a collimated illumination system (Fiber-lite, Dolan-Jenner, Mass., USA)
205 equipped with a 150 W halogen lamp and illuminating through an optical fiber opening
206 measuring 200 x 2 mm (LxW) with an angle of 45° with respect to the transportation
207 plate to minimize light divergence; 3) an imaging spectrographs (ImSpec V10; Specim
208 Ltd, Oulu, Finland) paired with a standard C-mount zoom interchangeable lens; 4)
209 digital camera (charge coupled device, CCD, monochrome camera, Toshiba-Teli
210 CS8310BC). Each frame acquired contains, as explained above, a line of pixels in one
211 dimension (which represents the spatial axis) plus the spectral pixels in the other
212 dimension (which represents the spectral axis), thus, providing full spectral information
213 for each pixel line. The reconstruction of the entire image (hyperspectral) of each
214 sample is carried out by the software that allows a line by line scanning of the sample
215 and an “ex post” reconstruction of the sample. The resolution of the line was equal to

216 700 pixels by 10 bits. This system was used in a laboratory shielded against light to
217 minimize ambient noise. The spectral values acquired were expressed in terms of
218 relative reflectance calibrating the system with white and black references.

219 For each hyperspectral image, a Region of Interest (ROI) of near 70 x 70 pixels (4900
220 pixels) was selected by a trained operator in order to determine the average VIS-NIR
221 spectral reflectance. This was extracted from the central part of each sample image to
222 avoid any interfering effects (e.g. shadows) of the Petri dish.

223

224 *2.5 Statistical analysis*

225 Data were subjected to analysis of variance followed by multiple pair-wise comparisons
226 by the post-hoc Tukey test. Data were also mean centered and unit variance-scaled (soft
227 scaling) and then subjected to principal component analysis (PCA) by the use of the
228 Modde 13.0 software package (Umetrics, Umea, Sweden). The possible presence of
229 either moderate or strong outliers in scores was assessed by the squared prediction
230 errors of residuals and Hotelling (T^2) of t-scores, respectively (MacGregor and Kourti
231 1995). At the variable level, variable power (VP), defined as the explained standard
232 deviation, was calculated by equation:

$$233 \quad VP = 1 - \frac{SV_j}{SV_j^0} \quad (1)$$

234 where SV_j is the residual standard deviation of the j^{th} variable and SV_j^0 is its initial
235 standard deviation, which is equal to unity for all variables after soft scaling.

236 A partial least squares (PLS) regression (Wold et al., 2001; Antonucci et al., 2011;
237 Febbi et al., 2015) approach was adopted to build prediction models. The employed
238 protocol involved (1) extraction of raw spectra (X block); (2) digitization of the dataset
239 of the concentrations of the target elements determined as described in Subsection 2.3

240 (Y block); 3) data fusion of the two datasets (X and Y blocks) in a single one for each
241 regression model to be extracted; (4) partitioning of the sample set relying on joint X
242 and Y blocks (SPXY) as described by Harrop Galvao and collaborators (2005); (5) split
243 of the datasets into two subsets, one of which referred to as model set (MS) for the
244 model (75%) and the other one (test set, TS) for the independent validation test (25%);
245 (6) cross-validation of the models, (7) application of pre-processing algorithms to both
246 X and Y; (8) application of PLS (modelling and testing) and, finally, (9) calculation of
247 the following efficiency parameters of prediction: bias error, standard error of prevision
248 (SEP), correlation coefficient (r), root mean square error (RMSE), ratio of percentage
249 deviation (RPD) according to the the RPD classification proposed by Williams (1987)
250 and the Variable Importance in Projection (VIP) score.

251 All the statistical analyses were performed by using the software packages (MATLAB
252 R 2013a, Mathworks, Natick, Mass., USA) and PLS Toolbox 6.5.4 (Eigenvector
253 Research Inc., Wenatchee, Wash., USA). Several combinations of different pre-
254 processing types and Latent Vectors were tested by a routine written in MATLAB with
255 an ensuing production of a large number of models. Among these alternative models,
256 for each of the target element (As, Al, Fe and Mn), the best one was selected on the
257 basis of maximization of r and RPD (calculated to RMSE of test subset) with
258 concomitant minimization of SEP and bias error for prediction of their respective
259 concentrations. The VIP scores were taken into consideration in order to estimate the
260 relevance of each variable in the projection made by the PLS model and, therefore, for
261 the variable selection (Taiti et al., 2014). A variable with a VIP score with value close to
262 or greater than 1 can be considered important in a specific model (Chong and Jun 2005).

263

264 3. Results and Discussion

265

266 3.1 Soil analysis

267 The main chemical properties of the soils collected from the area of the thermal springs
268 near Viterbo are reported in Table 1. The soils were slightly/moderately alkaline with
269 slight albeit significant differences among some samples. TOC contents were very low,
270 ranging from 3.4 to 16.8 g kg⁻¹. The contents of total CaCO₃, which was the
271 predominant constituent in the soils under study, varied widely from 135 to 744 g kg⁻¹,
272 depending on the sampling site. Total As contents in the five soils under study were by
273 far outside of the concentration range (1-20 mg kg⁻¹) of typical uncontaminated soils set
274 for this metalloid for agricultural soils (García-Sánchez et al., 2010) and the soil
275 guidelines (25 mg kg⁻¹) set by the Canadian Council of Ministers of the Environment
276 (CCME, 1999/2002) for protection of human health. In fact, with the only exception of
277 the S5 soil, the As content of which was 78.6 mg kg⁻¹, the remaining soils exhibited
278 concentrations higher than 150 mg kg⁻¹, as shown in Table 1. However, the labile forms
279 of As are mainly responsible for its toxic and harmful effects on both environment and
280 human health. Table 1 shows that among the relatively mobile As fractions, the
281 abundances of which were expressed in percent terms with respect to the total As
282 content, the dominant fraction was that associated with the carbonates (easily
283 exchangeable) and ranged from 46.5 to 72.5% for the S5 and S4 soils, respectively. The
284 cumulative abundances of all the relatively mobile fractions of As ranged from 65.4 to
285 83.6%, indicating that the metalloid had high bioavailability and mobility in the tested
286 soils. These results confirm the potential hazard to human health due to the exposure to
287 As in this area.

288 In order to show similarity/dissimilarity relationships among the soils collected from the
289 five sites of the thermal spring area and to gain insights into the effect of variables on

290 their distribution, an unsupervised multivariate approach, based on principal component
291 analysis, was used. Around 71% of total variance was explained by the first two
292 components. A neat separation of S2 and S5 soils from S1, S3 and S4 ones was evident
293 along the first component which explained 55.7 % of total variance (Fig. 1). The most
294 influential variables in driving this separation were CaCO₃, Fe, Al and Mn contents
295 along with real and potential acidity. The scores of the S1 soil were located in the right
296 hand quadrant which was diagonally opposite to loadings of Mn, Al, Fe and TOC; as a
297 matter of fact, the S1 soil exhibited the lowest contents of these parameters. The scores
298 of both S2 and S5 soils, which were the least As-polluted ones, instead, located, in the
299 lower left quadrant, which was diagonally opposite to the loadings of total As (Fig.1).
300 Based on data shown in Table 1, and irrespective of the sampling site, Table 2 shows
301 the correlation matrix among the physico-chemical variables of the soils under study.
302 For the elements under study, Fe, Mn and Al showed negative and highly significant
303 correlations ($P < 0.001$) with CaCO₃ and to a lesser extent with active carbonates.
304 Positive and highly significant correlations ($P < 0.001$) were also observed among the
305 same elements. Conversely, total As was positively correlated with CaCO₃ ($P = 0.004$)
306 and potential acidity ($P = 0.009$) and negatively correlated with TOC ($P = 0.005$).

307

308 *3.2 Vis-NIR reflectance spectrophotometry*

309 The concentrations of the elements under study (i.e., As, Al, Mn, Fe), used for the
310 model's construction, are shown in Table S1.

311 Fig. 2 shows that the average Vis/NIR reflectance spectrum of these soils was
312 characterized by the presence of weak and broad absorption bands in the visible region
313 while the signals tended to be narrower and sharper in the near-IR region. The spectra of
314 the soils under study resembled those reported in other investigations (Wu et al., 2005;

315 Choe et al., 2008) which suggested that these broad bands in the 400-900 nm range are
316 mainly associated with iron, e.g. Fe-O, in iron-bearing minerals (Ben-Dor et al., 1999).
317 Although metal(loid)s are spectrally featureless within the Vis-NIR region, their binding
318 to spectrally active components of soil, such as clay, metal (hydro)oxides and organic
319 matter, can induce spectral variations in these components (Ben-Dor et al., 1999). For
320 instance, Choe and collaborators (2008) found that the combination of different heavy
321 metals on the mineral surface promoted changes in the spectral features of oxygen and
322 lattice OH, the extents of which were found to be linked to actual concentrations of
323 heavy metals. For this reason, a variety of reflectance spectroscopy-based applications,
324 relying on the spectral signatures of minerals able to bind/sorb metal(loids), have been
325 developed either for the indirect detection (Song et al., 2012; Stazi et al., 2014; Wu et
326 al., 2005) or the mapping of metal dispersion (Choe et al., 2008).

327 However, as previously noted, the reflectance spectra of soils are generally broad and
328 non-specific, due to the presence of overlapping peaks of soil constituents.
329 Consequently, chemometric methods are often required to analyse the spectral
330 characteristics of soil and to relate them to metal concentrations (Song et al., 2012). In
331 this respect, principal component regression (PCR) (Pirie et al., 2005), support vector
332 machine (Stazi et al., 2014) and partial least-squares (PLS) regression (Moros et al.,
333 2009) have been exploited to establish models aimed at estimating the concentrations of
334 metal(loid)s in soils or sediments.

335 Among them, the PLS regression approach was adopted in this study, due also its ability
336 to handle multi-collinearity which is intrinsic to reflectance spectra (Summers et al.,
337 2011). Although PLS regression is reportedly able to handle several response variables
338 at a time, individual PLS models were established for each of the target elements. **Table**
339 **3** shows the results of the best PLS models for the prediction of Al, As, Fe and Mn

340 concentrations through VIS-NIR spectrophotometric analysis which were selected on
341 the basis of concomitant maximization of both r and RPD and minimization of SEP and
342 bias. All the best PLS model (Al, As, Fe and Mn) predictions were achieved by using
343 autoscaling as the pre-processing for the Y-block.

344 The best model of Al prediction was reached with 7 latent vectors (LV) yielding r
345 values equal to 0.92 and 0.88 for the training- and the test-set, respectively, while the
346 respective RPD values were 2.62 and 1.96. The error parameters (*i.e.*, SEP, RMSE)
347 were very high in both models and tests.

348 The best PLS model for As prediction was achieved with 5 LV and yielded high r and
349 RPD values for the model (0.96 and 3.5, respectively); the values of the same
350 parameters were 0.94 and 2.9, respectively when the model was validated on the test-
351 set. In addition, the error parameters were low, as shown in Table 3. On the basis of
352 both r and RPD values, the As prediction capacity of the established model was better
353 than that reported in other investigations (Song et al., 2012; and Stazi et al., 2014).

354 As for the Fe and Mn predictions, the r and RPD model values were equal to 0.94 and
355 0.94 and to 3 and 3.02 respectively. The r and RPD test values were equal to 0.93 and
356 0.92 and to 2.45 (very good) and 2.43 (very good).

357 **Fig. 3** shows the scatter plots of the predicted versus observed concentrations of the
358 elements under study and the relative r values. From the distribution of data around the
359 ideal line represented by the bisector of the diagram, it can be noted that the Al model
360 led to a slight over-prediction and under-prediction in the lower and upper concentration
361 range of the observed values, respectively. In the case of As, instead, the model
362 appeared to overestimate predictions in the upper range of the observed values. The
363 very close distribution of points around the bisector for Fe and Mn indicated that their

364 respective models estimated correctly the respective concentrations over the entire
365 range.

366 **Fig. 2** shows the averaged spectral reflectance curve VIS-NIR (nm) and, for each
367 element, the vertical grey bars underline the most relevant variables in the VIS-NIR
368 spectral range, identified on the basis of the greater than one rule of VIP scores. For
369 each element, the relevant spectral bands can be summarized as follows: Al, 450-470,
370 495-510, 590-615, 720-755, 855-880, 890-895 and 910-945 nm; As, 445-515, 585-640,
371 710-780, 790-795, 870-880 and 890-950 nm; Fe 450-465, 490-515, 590-615, 720-755,
372 855-880, 890-895, 910-930 and 940-950 nm; Mn 450-505, 595-615, 730, 745, 755,
373 865-880, 890-895, 910-930 and 940-945 nm.

374

375 **4. Conclusions**

376 This study showed that VIS-NIR reflectance spectroscopy combined with PLS
377 regression was used successfully in predicting the As concentration in naturally-
378 contaminated calcareous soils. The same approach was also useful in predicting the
379 concentration of elements (i.e., Fe, Mn and Al) involved reportedly in interactions with
380 the metalloid in soil. Among them, the most accurate predictions regarded both Fe and
381 Mn. These results suggest that Vis-NIR reflectance spectroscopy methods have great
382 potential for site-specific soil monitoring in high-risk areas destined to recreational
383 purposes, such as thermal spring sites.

384

385 **Acknowledgements**

386 The authors wish to thank Prof. E. Barberis (University of Torino, Department of
387 Agricultural, Forestry and Food Sciences, Torino, Italy) for coordinating the project.

388

389 **Funding**

390 The research was supported by the Italian Ministry of University and Research,

391 Research Projects of National Interest PRIN 2010–2011 (2010JBNLJ7-004).

392

393 **References**

- 394 Alexandratos, V.G., Elzinga, E.J., Reeder, R.J., 2007. Arsenate uptake by calcite:
395 macroscopic and spectroscopic characterization of adsorption and incorporation
396 mechanisms. *Geochim. Cosmochim. Acta* 71, 4172–4187.
- 397 Antonucci, F., Pallottino, F., Costa, C., Rimatori, V., Giorgi, S., Papetti, P., Menesatti,
398 P., 2011. Development of a rapid soil water content detection technique using active
399 infrared thermal methods for in-field applications. *Sensors* 11, 10114-10128.
- 400 Baiocchi, A., Lotti, F., Piscopo, V., 2012. Conceptual hydrogeological model and
401 groundwater resource estimation in a complex hydrothermal area: the case of the
402 Viterbo geothermal area (Central Italy). *J. Water Res. Protect.* 4, 231–247.
- 403 Ben-Dor, E., Irons, J.R., Epema, G. F., 1999. Soil reflectance. In: Rencz A.N. (Ed.),
404 Remote sensing for the earth sciences: Manual of remote sensing, John Wiley &
405 Sons, New York: pp. 111–188.
- 406 Bissen, M. Frimmel, F.H., 2003. Arsenic-a review. Part I: Occurrence, toxicity,
407 speciation, mobility. *Acta Hydrochim. Hydrobiol.* 31, 9–18.
- 408 Bundschuh, J., Maity J.P., 2015. Geothermal arsenic: occurrence, mobility and
409 environmental implications. *Renew. Sust. Energ. Rev.* 2015, 42, 1214–22.
- 410 CCME (Canadian Council of Ministers of the Environment). Canadian Environmental
411 quality guidelines. MB, Canada: Winnipeg; 1999/2002.
- 412 Choe, E., van der Meer, F., van Ruitenbeek, F., van der Werff, H., de Smeth, B., Kim,
413 K.W., 2008. Mapping of heavy metal pollution in stream sediments using combined
414 geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study
415 of the Rodalquilar mining area, SE Spain. *Remote Sens. Environ.* 112, 3222-3233.

416 Chong, I.G., Jun, C.H., 2005. Performance of some variable selection methods when
417 multicollinearity is present. *Chemometr. Intell. Lab. Sys*, 78, 103–112.

418 Dixit, S., Hering, J.G., 2003. Comparison of Arsenic(V) and Arsenic(III) sorption onto
419 iron Oxide minerals: Implications for Arsenic Mobility. *Environ. Sci. Technol.* 37,
420 4182–4189.

421 **Drouineau, G., 1942. Dosage rapide du calcaire actif du sol; nouvelles donnees sur la**
422 **separation et la nature des fractions calcaires. *Ann. Agron.* 12, 441-450.**

423 FAO (Food & Agriculture Organization for the United Nations). World reference base
424 for soil resources. Vol. 3, 1998.

425 Febbi, P., Menesatti, P., Costa, C., Pari, L., Cecchini, M., 2015. Automated
426 determination of poplar chip size distribution based on combined image and
427 multivariate analyses. *Biomass Bioen.* 73, 1-10.

428 García-Sánchez, A., Alonso-Rojo, P., Santos-Francés, F. 2010. Distribution and
429 mobility of arsenic in soils of a mining area (Western Spain). *Sci. Tot. Environ.*
430 408, 4194-4201.

431 Goldberg, S., Glaubig, R.A., 1988. Anion sorption on a calcareous, montmorillonitic
432 soil arsenic. *Soil Sci. Soc. Am. J.* 52, 1297–1300.

433 Harrop Galvao, R.K., Ugulino Araujo, M.C., Emidio Jose, G., Coelho Pontes, M.J.,
434 Cirino Silva, E., Bezerra Saldanha, T.C., 2005. A method for calibration and
435 validation subset partitioning. *Talanta* 67, 736-740.

436 Huang, P.M., Gobran, G.R. 2005. *Biogeochemistry of Trace Elements in the*
437 *Rhizosphere*, Elsevier Science, Amsterdam.

438 Kamunda, C., Mathuthu, M., Madhuku, M., 2016. Health risk assessment of heavy
439 metals in soils from Witwatersrand gold mining basin, South Africa. *Int. J. Environ.*
440 *Res. Publ. Health* 13, 663.

441 Kemper, T., Sommer, S., 2002. Estimate of heavy metal contamination in soils after a
442 mining accident using reflectance spectroscopy. *Environ. Sci. Technol.* 36, 2742–
443 2747.

444 López, D.L., Bundschuh, J., Birkle, P., Armienta, M.A., Cumbal, L., Sracek, O.,
445 Cornejo, L., Ormachea, M. M., 2012. Arsenic in volcanic geothermal fluids of
446 Latin America. *Sci. Total Environ.* 429, 57–75.

447 MacGregor, J., Kourti, T., 1995. Statistical process control of multivariate processes.
448 *Control Eng. Pract.* 3, 403–414.

449 Mamindy-Pajany, Y., Hurel, C., Marmier, N., Roméo, M., 2009. Arsenic adsorption
450 onto hematite and goethite. *Compt. Rend, Chimie* 12, 876-881.

451 Matera, V., Le Hecho, I., Laboudigue, A., Thomas, P., Tellier, S., Astruc, M., 2003. A
452 methodological approach for the identification of arsenic bearing phases in polluted
453 soils. *Environ. Pollut.* 126, 51-64.

454 Mehmood, A., Hayat, R., Wasim, M., Akhtar, M. S., 2009. Mechanisms of arsenic
455 adsorption in calcareous soils. *J. Agric. Biol. Sci.* 1, 59-65.

456 Menesatti, P., Antonucci, F., Pallottino, F., Rocuzzo, G., Allegra, M., Stagno, F., &
457 Intrigliolo, F., 2010. Estimation of plant nutritional status by Vis-NIR
458 spectrophotometric analysis on orange leaves [*Citrus sinensis* (L) Osbeck cv
459 Tarocco]. *Biosys. Eng.* 105, 448-454.

460 Moros, J., de Vallejuelo, S.F.O., Gredilla, A., de Diego, A., Madariaga, J.M., Garrigues,
461 S., de la Guardia, M., 2009. Use of reflectance infrared spectroscopy for monitoring
462 the metal content of the estuarine sediments of the Nerbioi-Ibaizabal River
463 (Metropolitan Bilbao, Bay of Biscay, Basque Country). *Environ. Sci. Technol.* 43,
464 9314–9320.

465 Ormachea, M. M., Bhattacharya, P., Sracek, O., Ramos Ramos O., Quintanilla, A. J.,
466 Bundschuh, J., Jyoti, P. M., 2015. Arsenic and other trace elements in thermal
467 springs and in cold waters from drinking water wells on the Bolivian Altiplano. *J.*
468 *South Am. Earth Sci.* 60,10–20.

469 Pirie, A., Singh, B., Islam, K., 2005. Ultra-violet, visible, near-infrared, and mid-
470 infrared diffuse reflectance spectroscopic techniques to predict several soil
471 properties. *Austral. J. Soil Res.* 43, 713–721.

472 Sadiq, M., 1997. Arsenic chemistry in soils: an overview of thermodynamic predictions
473 and field observations. *Water, Air, and Soil Pollut.* 93, 117-136.

474 Song, Y., Li, F., Yang, Z., Ayoko, G.A., Frost, R.L., Junfeng, J., 2012. Diffuse
475 reflectance spectroscopy for monitoring potentially toxic elements in the
476 agricultural soils of Changjiang River Delta, China. *Appl. Clay Sci.* 64, 75–83.

477 Stazi, S.R., Antonucci, F., Pallottino, F., Costa, C., Marabottini, R., Petruccioli, M.,
478 Menesatti, P., 2014. Hyperspectral Visible–Near Infrared Determination of Arsenic
479 Concentration in Soil. *Commun. Soil Sci. Plant Anal.* 45, 2911-2920.

480 Summers, D., Lewis, M., Ostendorf, B., Chittleborough, D., 2011. Visible near-infrared
481 reflectance spectroscopy as a predictive indicator of soil properties. *Ecol. Indic.* 11,
482 123–131.

483 Taiti, C., Costa, C., Menesatti, P., Comparini, D., Bazihizina, N., Azzarello, E., Masi, E.
484 Mancuso, S., 2014. Class-modelling approach to PTR-TOFMS data: A peppers
485 case study. *J. Sci. Food Agric.* 95, 1757–1763.

486 Violante, A., Gaudio, S.D., Pigna, M., Pucci, M., Amalfitano, C., 2008. Sorption and
487 desorption of arsenic by soil minerals and soils in the presence of nutrients and
488 organics, 39-69. In: Huang, Q., Huang, P.M., Violante, A. (Eds.), *Soil mineral*

489 microbe-organic interactions: Theories and applications, Springer Verlag, Berlin
490 Heidelberg, 1-353.

491 White, J.W., 1999. Hazards of short-term exposure to arsenic-contaminated soil.
492 Washington State Department of Health, Office of Environmental Health
493 Assessment Services.

494 WHO 2001. Environmental Health Criteria 224: Arsenic Compounds. Second edition,
495 World Health Organization, Geneva.

496 Williams, P.C., 1987. Variables affecting near-infrared reflectance spectroscopic
497 analysis. In: Williams and K. Norris (Eds.), Near-Infrared Technology in the
498 Agricultural and Food Industries, American Association of Cereal Chemists, St
499 Paul, Minnesota, pp. 143-167.

500 Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of
501 chemometrics. *Chemometr. Intell. Lab. Syst.* 58, 109-130.

502 Wu, Y., Chen, J., Wu, X., Tian, Q., Ji, J., Qin, Z., 2005. Possibilities of reflectance
503 spectroscopy for the assessment of contaminant elements in suburban soils. *Appl.*
504 *Geochem.* 20, 1051–1059.

505 Yang, L., Donahoe, R. J., 2007. The form, distribution and mobility of arsenic in soils
506 contaminated by arsenic trioxide, at sites in southeast USA. *Appl. Geochem.* 22,
507 320-341.

508

509 **Figure Captions**

510

511 **Fig. 1.** PCA correlation biplot of soils (filled circles) of the Bullicame thermal spring
512 area and physico-chemical variables (shadowed triangles). The Arabic numbers at the
513 end of each score label refer to the identification of sampling sites (Si). The percentage
514 of variance explained by the first and second component amounted to 55.7% and 14.9%,
515 respectively.

516

517 **Fig. 2.** The averaged spectral reflectance curve in the VIS-NIR region and the VIP
518 scores. The grey bars outline those spectral variables considered more important in the
519 model (VIP scores ≥ 1).

520

521 **Fig. 3.** Scatter plots of observed versus predicted values extracted from Partial Least
522 Squares best models of Al, As, Fe and Mn concentrations. The dashed line in each
523 diagram represents the bisector and the inset contains the Pearson correlation coefficient
524 (r).

525