Improving the quantification of sediment source contributions using different mathematical models and spectral preprocessing techniques for individual or combined spectra of ultraviolet-visible, near-and middle-infrared spectroscopy

Tales Tiecher, Jean M Moura-Bueno, Laurent Caner, Jean P.G. Minella, Olivier Evrard, Rafael Ramon, Gabriela Naibo, Cláudia A.P. Barros, Yuri J.A.B. Silva, Fábio F Amorim, et al.

Improving the quantification of sediment source contributions using different mathematical models and spectral preprocessing techniques for individual or combined spectra of ultraviolet-visible, near- and middle-infrared spectroscopy

Tales Tiecher[a], Jean M Moura-Bueno[b], Laurent Caner[c], Jean PG Minella[b], Olivier Evrard[d], Rafael Ramon[e], Gabriela Naibo[e], Cláudia AP Barros[a], Yuri JAB Silva[f], Fábio F Amorim[g], Danilo S Rheinheimer[b]

[a] *Department of Soil Science, Universidade Federal do Rio Grande do Sul (UFRGS), Interdisciplinary Research Group on Environmental Biogeochemistry (IRGEB), Bento Gonçalves Ave. 7712, 91540-000 Porto Alegre, RS, Brazil*

[b] *Department of Soil Science, Universidade Federal de Santa Maria (UFSM), Roraima Ave. 1000, 97105-900 Santa Maria, RS, Brazil*

[c] *Université de Poitiers, IC2MP-HydrASA, UMR 7285, 7 rue Albert Turpain, B35, Poitiers, 86022, France*

[d] *Laboratoire des Sciences et de l'Environnement (LSCE-IPSL), UMR 8212 (CEA/CNRS/UVSQ), Université Paris-Saclay, CEA Saclay, Orme des Merisiers, 91 191 Gif-sur-Yvette Cedex, France*

[e] *Graduate Program in Soil Science, Federal University of Rio Grande do Sul, Bento Gonçalves Ave., 91540-000 Porto Alegre, RS, Brazil*

[f] *Agronomy Department, Universidade Federal do Piauí (UFPI), Planalto horizonte, Bom Jesus, Piauí 64900-000, Brazil*

[g] *Agronomy Department, Universidade Federal Rural do Pernambuco (UFRPE), Dom Manuel de Medeiros street, s/n - Dois Irmãos, 52171-900 Recife, PE, Brazil*

Abstract

In recent years, several sediment fingerprinting studies have used ultraviolet-visible (UV-Vis), near-infrared (NIR) and middle-infrared (MIR) spectroscopy as a low cost, non-destructive and fast alternative to obtain tracer properties to estimate sediment source contributions. For this purpose, *partial last square regression* (PLSR) has often been used to build predictive parametric models. However, spectra preprocessing and more robust and non-parametric models such as *support vector machines* (SVM) has gained little attention in these studies. Accordingly, the objectives of the current research were to evaluate (i) the accuracy of two multivariate methods (PLSR and SVM), (ii) the effect of eight spectra preprocessing techniques, and (iii) the effect of using the information contained in the UV-Vis, NIR and MIR regions considered either separately or in combination on sediment source apportionment. The estimated source contribution was then compared with contributions obtained by the conventional fingerprinting approach based on geochemical tracers. This study was carried out in the Arvorezinha catchment (1.23 km^2) located in southern Brazil. Forty soil samples were collected in three main potential source (cropland surface, unpaved roads and stream channels) and twenty-nine suspended sediment samples collected at the catchment outlet during nine rainfall-runoff events were used in this study. Both PLSR and SVM models showed a higher accuracy when calibrated and validated with the spectra submitted to spectral processing when compared to the direct use of the raw spectra. The best model results were obtained with PLSR and SVM mathematical models associated with the spectral preprocessing techniques 1st derivative Savitzky-Golay (SGD1), normalization (NOR) and combining NOR+SGD1 in the UV-Vis+NIR+MIR. The lowest errors were observed when the UV-Vis+NIR+MIR bands were combined due to the gain in information and, consequently, the increase in discriminant power achieved by the models. Despite the good accuracy of the models calibrated and validated with the mixtures, significant errors remain when results of source contributions are compared to those obtained with the conventional sediment fingerprinting technique based on geochemical tracers. Nevertheless, the magnitude of the contributions calculated by the spectroscopy and geochemical approaches remains very similar for all sources, especially when using the SVM-UV-Vis+NIR+MIR model. Therefore, spectroscopy proved to be a fast, cheap and accurate technique, offering an alternative to the conventional geochemical approach for discriminating sediment source contributions in agricultural catchments located in subtropical regions.

Keywords: soil erosion, alternative fingerprints, machine learning, Vis-NIR-MIR database, spectral preprocessing.

1. Introduction

The sustainable production of food, fiber and biofuel remains limited by soil erosion (Comino et al., 2015; Erkossa et al., 2015; Seutloali and Beckedahl, 2015; Taguas et al., 2015). The inadequate management of soils and the lack of runoff control exposes the soil to erosive agents, accelerating the processes of mobilization and transfer of sediments to the drainage network (Minella et al., 2014), along with the transportation of contaminants like pesticides (Magnusson et al., 2013; Yahia and Elsharkawy, 2014) and phosphorus (Dodd et al., 2014; Dodd and Sharpley, 2015; Poulenard et al., 2008). Erosion control and conservation of soil hydrological functionalities are essential to meet the demand in food production as well as to maintain the quality of water resources (Didoné et al., 2015; Merten et al., 2015). To better understand the occurrence of erosion processes at the catchment scale and to mitigate the problems arising from river overflow and excessive sediment production, it is first necessary to have quantitative information on the sources delivering sediment to the river systems.

To this end, the use of the sediment fingerprinting approach quantifies the contribution of non-point based sediment sources through the use of a variety of tracers, including geochemical properties and radionuclides (D'Haen et al., 2012; Davis and Fox, 2009; Haddadchi et al., 2013; Koiter et al., 2013; Walling and Woodward, 1995). However, the large-scale application of the sediment fingerprinting technique requires a large number of chemical analyses, which have a significant cost and which are relatively time consuming. However, the use of other soil and sediment characteristics may complement or provide a powerful alternative information in source identification, such as reflectance spectroscopy (Cooper et al., 2014).

Reflectance spectroscopy in the visible and infrared ranges provides an efficient, low-cost, fast and non-destructive method easily applicable to soil and sediment samples to quantify various physical, chemical and biological properties (McBratney et al., 2006; Viscarra Rossel et al., 2006). Poulenard *et al*. (2009) were pioneers in the use of diffuse infrared spectroscopy (Diffuse Reflectance Infrared Fourier Transform Spectroscopy - DRIFTS) to trace the origin of sediments in river catchments. The method was successfully used to discriminate and predict the respective contribution of surface and subsurface sources to sediment, as well as to discriminate soils developed on different lithology (Poulenard et al., 2012). Since then, several studies have been developed using the spectroscopic method to trace sediment sources (Table 1). Only 28 scientific articles had been published around the world after that of Poulenard et al., (2009) by early 2020, representing an average of about 2.5 scientific publications per year (Fig. 1). They include studies in the ultraviolet-visible (UV-Vis) and near-infrared (NIR) ranges in the French Alps (Legout et al., 2013) and the Southern France (Uber et al., 2019), Luxemburg (Martínez-Carreras et al., 2010c, 2010b, 2010a), Spain (Brosinsky et al., 2014a, 2014b), Ethiopia (Verheyen et al., 2014), South Africa (Pulley et al., 2018; Pulley and Rowntree, 2016), Argentina (Batistelli et al., 2018), the United Kingdom (Collins et al., 2014), Canada (Barthod et al., 2015; Boudreault et al., 2018; Liu et al., 2017), Brazil (Tiecher et al., 2016, 2015; Valente et al., 2020) and Iran (Nosrati et al., 2020). In the middle infrared region (MIR), sediment fingerprinting studies were carried out in France (Poulenard et al., 2012, 2009), Mexico (Evrard et al., 2013), the United Kingdom (Vercruysse and Grabowski, 2018), China (Liu et al., 2019), Brazil (Tiecher et al., 2017) and in a transnational river catchment covering part of Switzerland, France and Germany (Chapkanski et al., 2019). Many of these studies have shown a good agreement between the results obtained with the spectroscopic method and those provided by the

106 conventional approach based on geochemical and / or radionuclide properties (Evrard et
107 al., 2013; Legout et al., 2013; Martínez-Carreras et al., 2010c; Tiecher et al., 2015, 2016,
108 2017; Verheyen et al., 2014).

109 In addition to the more qualitative attempts that combine the use of discriminant analysis
110 with spectroscopy (Chapkanski et al., 2019), previous studies conducted to trace sediment
111 source contributions using spectroscopy (Table 1) can be divided into three main groups,
112 according to the way they used the spectroscopic information. The first group uses color
113 parameters extracted from the visible range (Barthod et al., 2015; Martínez-Carreras et al.,
114 2010c, 2010b, 2010a; Pulley et al., 2018; Pulley and Rowntree, 2016) and other spectral
115 characteristics *(spectral features* and *overtones)* (Brosinsky et al., 2014b, 2014a; Collins et
116 al., 2013, 2014) in an optimized mixed linear model, separately or in combination with
117 conventional geochemical tracers (Tiecher et al., 2015). The second group of studies uses
118 spectroscopic information to generate mathematical models using the least squares
119 method (PLSR) to estimate the concentrations of geochemical tracers, which in turn are
120 introduced in a mixed linear model optimized to estimate the contribution of sediment
121 sources (Vis-NIR-SWIR - Martínez-Carreras et al., 2010b).

122 The third group of studies directly uses the entire spectrum (Batistelli et al., 2018; Evrard et
123 al., 2013; Poulenard et al., 2012, 2009; Tiecher et al., 2017, 2016, 2015; Verheyen et al.,
124 2014) or the spectrum combined with color parameters extracted from the visible range
125 (Legout et al., 2013). They estimate the source contributions in sediment samples after
126 generating a model (*Partial Last Squares Regression* – PLSR) calibrated using artificial
127 mixtures combining potential sediment sources in variable proportions. To date, no study
128 has been conducted combining the bands of UV-Vis, NIR and / or MIR, although it may be

129 expected that predictions of soil properties may be improved using the combined range of
130 these regions of the electromagnetic spectrum (Knox et al., 2015; Reeves, 2010; Soriano-
131 Disla et al., 2014) due to the sum of distinct information that is added to the models
132 (Viscarra Rossel et al., 2006).

133 This third approach has the advantage of using all the spectral information directly in a
134 mathematical model calibrated using artificial mixtures. To achieve this goal, all the
135 previous studies used the Partial Least Squares Regression method (PLSR) (Table 1).
136 However, more robust and non-parametric models, such as Support Vector Machine
137 (SVM), could usefully be tested be tested for this type of application, as they were
138 successfully used to derive soil properties, including clay and organic carbon content (Dotto
139 et al., 2017; Lucà et al., 2017; Stevens et al., 2013; Viscarra Rossel and Behrens, 2010). This
140 multivariate method seeks to identify an interpolation function using the kernel function,
141 adjusting the calibration data until simultaneously minimizing the size of the coefficients
142 and the prediction errors, in which data with non-linear patterns can be better represented
143 by the calibrated model (Ivanciuc, 2007).

144 In spectroscopic fingerprinting studies, there is also a knowledge gap regarding the use of
145 spectrum pre-processing techniques in spectroscopic modeling (Table 1), despite the fact
146 that several soil studies demonstrated that this step is extremely important when
147 calibrating the models (Buddenbaum and Steffens, 2012; Dotto et al., 2017; Nawar et al.,
148 2016). Several spectrum pre-processing techniques can be used for this purpose, such as
149 smoothing, Savitzky-Golay with 1st or 2nd derivative using a first or second order
150 polynomial, standard normal variate, multiplicative scatter correction and normalization.
151 However, so far, very few fingerprinting sediment tracing studies have reported whether

152 they used any type of spectral pre-processing. Most of them used only pre-processing
153 (Brosinsky et al., 2014b, 2014a; Chapkanski et al., 2019; Tiecher et al., 2017, 2016, 2015),
154 except for Ni et al. (2019), who used four pre-processing techniques, although they did not
155 compare them to each other.

156 In this context, the current research provides, to the best of our knowledge, the first
157 attempt to compare the outputs of different multivariate mathematical models (both
158 parametric and non-parametric) and preprocessing techniques of reflectance in UV-Vis, NIR
159 and MIR spectral bands used in combination or in isolation to predict the contribution of
160 sediment sources at the catchment scale. Accordingly, the objectives of the study were (i)
161 to evaluate the accuracy of two multivariate methods for sediment source apportionment
162 (PLSR and SVM), (ii) to evaluate the effect of eight spectra preprocessing techniques, and
163 (iii) to evaluate the effect of using the information contained in the UV-Vis, NIR and MIR
164 regions either separately or in combination. For this purpose, the estimated contribution
165 for each source using the spectroscopic models was also compared against the values
166 obtained by the conventional fingerprinting approach based on geochemical tracers

167

168 **2. Materials and methods**

169 **2.1. Study site**

170 The Arvorezinha catchment is located in the northeastern part of the Rio Grande do Sul
171 State, southern Brazil. Igneous rocks (basalts and rhyodacite) characterize the geology and
172 the altitude varies from 580 to 730 meters. The upper third of the catchment has an
173 undulating plateau relief with slopes up to 7%, and the middle and lower thirds of the
174 catchment have a much steeper topography with slope gradients often exceeding 15%. The

climate is classified as Cfb (subtropical super-humid with no dry season and warm summer) according to Köppen (Alvares et al., 2013). The mean annual precipitation for the last 15 years (2002-2016) is 1938 mm with a mean erosivity of 9344 MJ mm ha^{-1} h^{-1} yr^{-1} (Ramon, 2017). The main crop is tobacco grown in small farms. Corn, soybean, eucalyptus and native forests are also found in the catchment. The landscape is characterized by short, steep slopes with a strong hydrological connectivity between hillslopes and the drainage network. The soil classes found in the catchment are Acrisols, Cambisols and Leptosols (IUSS Working Group WRB, 2015). Inadequate soil management under agricultural land associated with limited water infiltration due absence of subsurface horizon or clayey B horizon favor the formation of runoff that controls erosion dynamics in cropland. In addition to erosion processes in cropland, the inadequate location and maintenance of unpaved roads generate preferential pathways for runoff concentration accelerating erosion, and converting these roads into significant potential sources of sediment. The catchment also shows signs of channel banks erosion due to the high flow energy observed during the events (flash flood) associated with the absence of riparian forest along several river sections. Further details on the catchment can be found in Tiecher et al. (2015).

Soil samples were collected from the three main potential sources of sediment, including (i) cropland surface, (ii) unpaved roads and (iii) stream channels. Cropland (n=20) and unpaved road (n=10) samples were taken using a non-metallic trowel from the uppermost layer (0–0.05 m). Stream channels (n=10) were sampled on exposed bank sites located along the main river channel network. Each sample was composed of at least 10 subsamples collected in the vicinity of the sampling point (within a radius of approximately 10 m). Source sampling sites were selected based on visible signs of soil erosion and hydrological connectivity as well as taking into account pedological variability. In order to

characterize the sediments transported in the drainage network, 29 suspended sediment samples were collected during nine rainfall events from October 2009 to July 2011 at the catchment outlet, where the flow and sediment concentration are also continuously monitored, which allows the calculation of liquid and solid discharges, as well as the catchment sediments yields. The sampled events cover the seasonal variability of land cover in the drainage area, as well as the variations of river flow conditions. For high magnitude events, several samples were collected to characterize the intra-event variability (rise, peak and recession of the hydrograph). Sediment concentrations during monitored events ranged from 300 to 2000 mg L^{-1}. To obtain a sufficient mass of suspended sediment for conducting all analyses, samples were collected using a portable continuous flow centrifuge (Alfie-500 Alfa Laval). All the source material and sediment samples were oven-dried at 50 °C, gently disaggregated using a pestle and mortar, and passed through a 63-µm mesh prior to laboratory analyses to investigate similar particle size-fractions for all the samples (suspended sediments and sediment sources).

2.2. Artificial mixtures of sediment sources

The samples of each potential sediment source were mixed in equal proportions in the laboratory to constitute a unique reference sample for the corresponding source. Then, those reference samples were mixed in 48 different weight proportions as presented in Fig. 2. These artificial mixtures containing different proportions of the three sources were then used to calibrate the multivariate mathematical models used to estimate the source contributions to the suspended sediment samples.

222 **2.3. Spectral analyses**

223 UV-Vis diffuse reflectance spectra of samples were recorded at room temperature from
224 200 to 800 nm with a 1-nm step using a Cary 5000 UV–Vis–NIR spectrophotometer (Varian,
225 Palo Alto, CA, USA). Samples were ground and loaded into a Harrick Praying Mantis diffuse
226 reflectance accessory that uses elliptical mirrors. $BaSO_4$ was used as a 100% reflectance
227 standard. Care was taken when adding the samples into the sample holder to avoid
228 differences in sample packing and surface smoothness.

229 Near infrared (NIR) spectra were recorded in the range 10000–4000 cm^{-1} (1000-2500 nm)
230 using a Nicolet 26700 FTIR spectrometer (Waltham, Massachusetts, USA) in diffuse
231 reflectance mode with an integrating sphere and a InGaAs detector with a resolution of 4
232 cm^{-1} and 100 readings per spectrum.

233 Mid infrared (MIR) spectra were obtained in the range 400–4000 cm^{-1} (2500-25000 nm)
234 using a Nicolet 510-FTIR (Thermo Electron Scientific, Madison, WI, USA) spectrometer in
235 diffuse reflectance mode with a resolution of 4 cm^{-1} and 100 readings per spectrum. A
236 direct current of air was used (dry and without CO_2) to eliminate CO_2 and water from the
237 spectrometer in order not to interfere with scanning and obtaining the spectra.

238 UV-Vis, NIR and MIR spectral data were subjected to spectral pre-processing to remove
239 physical variability due to light dispersion and to remove systematic variations of
240 instrumental and environmental conditions in order to emphasize the characteristics of
241 interest along the spectrum. Spectra without pre-processing constitute the "control
242 treatment" (RAW). Eight spectral processing techniques commonly employed in
243 chemometric studies (Rinnan et al., 2009) were tested to evaluate its effect on the
244 calibration of spectroscopic models.

Pre-processing includes: (i) smoothing (SMO) from a convolution function using a 25 nm mobile window, after initial testing to define the best search window; (ii) Savitzky-Golay (Savitzky and Golay, 1964) with 1st derivative using a first order polynomial (SGD1), with 25 nm search window, after initial testing to define the best search window; (iii) Savitzky-Golay (Savitzky and Golay, 1964) with 2nd derivative using a second order polynomial (SGD2), with 25 nm search window. The 1st and 2nd derivates calculate the change of reflectance in wavelength variation rate. This technique is widely used to remove baseline shifts and highlight spectral features of interest; (iv) varied normal standard deviation (SNV) is used to remove spectral data dispersion caused by noise and different particle sizes and consists of subtracting the mean and dividing it by the standard deviation (spectrum - mean/standard deviation) of each spectrum individually; (v) multiplicative scatter correction (MSC) is effective in minimizing baseline compensations and multiplicative effects; (vi) normalization (NOR) is the ratio of spectrum bands measured by standard deviation (NOR); (vii) combination of NOR+SGD1; (viii) combination of MSC+SGD1. These eight techniques can be divided into three groups according to the objective and the mathematical approach employed. The first group includes only the smoothing of the spectra, represented by the SMO. The second group is defined by the use of derivatives to remove baseline shifts and enhance spectral features, represented by SGD1 and SGD2. The third group corresponds to techniques for spectral data normalization and dispersion corrections such as SNV, MSC and NOR. All pre-processing was performed using the prospectr and clusterSim packages (Stevens and Ramirez-Lopez, 2020; Walesiak and Dudek, 2020) R software (R Core Team, 2020).

2.4. Spectroscopic model development: calibration and validation

The spectroscopic models were calibrated from the spectral signature of the 48 mixtures with different proportions of each sediment source (Figure 2). Spectral data were generated from the following regions of the electromagnetic spectrum: UV-Vis, NIR, MIR, UV-Vis+NIR, UV-Vis+MIR, NIR+MIR and UV-Vis+NIR-MIR. Two multivariate calibration methods were used to adjust the models, where the effect of the eight spectral pre-processes plus the raw spectrum (RAW) was tested. This totaled 126 spectroscopic models for each source, totaling 378 models (Figure 4).

Two multivariate methods with different approaches were selected to calibrate the spectroscopic models: (i) *Partial Least Squares Regression* (PLSR) (R *pls* package (Mevik et al., 2016)) parametric technique widely used in spectroscopic modeling (Angelopoulou et al., 2020; Dotto et al., 2018) and (ii) *Support Vector Machines* (SVM) (R *e1071* package (Meyer et al., 2019)), non-parametric technique. Methods with different approaches were selected due to the occurrence of linear and non-linear correlations between the organo-mineral components of soil/sediment and the spectral variables (Viscarra Rossel and Behrens, 2010). The PLSR model handles data sets containing many independent and highly correlated variables, such as UV-Vis-NIR-MIR spectral data. PLSR analysis reduces large data sets to a small number of uncorrelated orthogonal factors to minimize the sum of the squares of the predicted value errors (Varmuza and Filzmoser, 2009). The SVM model was used with the kernel function, which separates the calibration data into hyperplanes and seeks to establish correlations between the dependent and independent variables when these have non-linear behavior (Ivanciuc, 2007).

The parameterization in the calibration of the PLSR method was: method = 'pls', resampling method = ‘cross-validation 10 k-fold’, and number of components = ‘.ncomp = seq(1, 20, 1)’. For SVM the parameters were: method = 'svmLinear', resampling method = ‘cross-validation 10 k-fold’, and Kernel parameters = ‘Support Vector Machine with Linear Kernel’. Each model was calibrated with 70% of the samples (n= 34) and validated with 30% of the samples (n = 14). Both sets were randomly generated. To evaluate the accuracy of the models the following parameters were calculated: coefficient of determination (R^2) (Equation 1), bias (Equation 2) and mean square root of the prediction error (RMSE) (Equation 3).

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i) \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (3)$$

where: $\hat{y}$ = predicted value of each source; $\bar{y}$ = observed mean value of each source in the mixture; y = observed values of each source in the mixture; n = number of samples with i = 1, 2,..., n.

2.5 Estimation of sediment source contributions by spectroscopic models and independent validation comparing with the contribution results obtained with the conventional geochemical approach

The contribution of sediment sources was estimated by spectroscopic models from an independent set of spectral reflectance data, which were obtained from 29 suspended sediment samples. The predicted values were then compared with those obtained with the geochemical tracers (Tiecher et al., 2015) and the $bias_{sp}$ error statistics (Equation 2) and $RMSE_{sp}$ (Equation 3) of the proportion of sediment estimated for each source were calculated. This approach used the total concentration in various elements (Ag, As, Cr, Fe, Mo, and P) estimated by ICP-OES after microwave assisted digestion with concentrated HCl and HNO_3 in the 3:1 ratio (*aqua regia*). Detailed information regarding sediment sampling and the statistical procedure used in the conventional geochemical approach can be found in Tiecher et al. (2015).

Finally, data of quality of the models based on validation with artificial mixtures of sediment ($RMSE_v$), and compared with the sediment contribution values obtained with geochemical tracers ($RMSE_{sp}$) were entered in a conditional inference regression tree procedure to highlight the factors that most influenced the quality of the models.

3. Results and discussion

3.1. Model calibration performance

The results of the validation of the 378 models are presented in Figure 5. In general, there is variation in the predictive behavior of the models depending on the type of spectral pre-processing, spectral range, multivariate method and sediment source considered (Figure

5). One the one hand, the calibrated models with SGD1, NOR+SGD1 and NOR pre-processing achieved a greater accuracy in predictions, differing significantly from that obtained with other techniques (Figure 6a). Studies on organic carbon and soil clay content predictions have also shown that it is possible to increase the accuracy of spectroscopic models calibrated with SGD1 and NOR pre-processing (Dotto et al., 2018; Knox et al., 2015; Moura-Bueno et al., 2019; Pinheiro et al., 2017; Vasques et al., 2008). On the other hand, the models calibrated with preprocessed spectra using the SGD2, MSC, MSC+SGD1, SMO and SNV techniques presented the lowest performance among all spectral ranges and their combinations in the three sediment sources (Figure 6a). These results agree with the studies of Dotto et al. (2018) and Moura-Bueno et al. (2019) that showed a lower accuracy in organic carbon predictions of subtropical soils in southern Brazil when the SVM and PLSR models are calibrated with Vis-NIR spectra submitted to the MSC and SNV techniques, respectively.

The SDG2 technique stands out with a higher value of $RMSE_v$ observed in the predictions of the three sources, differing from the other pre-processes (Figure 6a). Moreover, the SDG2 technique was the only pre-processing that resulted in a $RMSE_v$ value higher than 15% and was also the only method with a $RMSE_v$ higher than that obtained for the spectra without any type of pre-processing (RAW). Possibly, the SGD2 technique may be eliminating features (predictor variables) of the spectra that are important for the prediction of sediment sources. Nevertheless, several studies have used the 2nd derivative technique to treat UV-Vis, NIR and MIR spectra (Tiecher et al., 2015, 2016, 2017) and MSC (Ni et al., 2019) and SNV (Chapkanski et al., 2019) techniques to treat MIR spectra, with the objective to derive the characteristics of interest along the spectrum and to improve the predictions

355 of the models of sediment source contributions. However, as observed in the current
356 research, these pre-processing methods do not appear to be the most promising.

357 There is still no consensus on how to define a priori the pre-processing techniques that will
358 produce predictive models with a greater accuracy. Each pre-processing method will
359 behave differently (Rinnan et al., 2009) depending on the set of soil samples considered, or
360 as in the current research, on the intra-source spectral variation. Thus, ideally, it remains
361 necessary to perform preliminary tests with different pre-processing methods, especially in
362 river catchments with contrasted geologies and soil types, and when considering greater
363 number of potential sediment sources and in larger basins, where the complexity is
364 greater. The content in organic matter, clay and iron oxides strongly influence the spectral
365 behavior of soils (Moura-Bueno et al., 2019; Viscarra Rossel and Behrens, 2010) and
366 sediments (Figure 3) (Tiecher et al., 2017, 2016, 2015). Therefore, in a context of great
367 variability between sources, as when comparing sources with different geology and
368 mineralogy (Poulenard et al., 2009), or surface and subsurface sources with contrasting
369 carbon contents (Evrard et al., 2013), pre-processing techniques may result in limited
370 improvement of predictive models. However, they may have a greater potential to improve
371 models in larger basins and in study areas characterized by sources showing homogeneous
372 carbon contents and mineralogical compositions.

373 The lowest variation and value in the $RMSE_v$ for the unpaved roads source in all spectral
374 ranges and combinations (Figure 6c) compared to the other potential sediment sources are
375 likely associated with the spectral behavior of this source. Among the three sources
376 considered, unpaved roads are the most depleted in organic matter and the coarsest grain-
377 sized (Tiecher et al., 2019). Furthemore, it is composed of subsoil material that is richer in

2:1 clay mineral (Tiecher et al., 2016). In addition, qualitatively, organic matter from unpaved roads contains a higher proportion of alkyl-benzene, and a lower proportion of polysaccharides and amino acids, compared to cropland and stream channels, which in turn are sources that show organic matter of similar quality (Tiecher et al., 2015). The lower organic matter content of the unpaved roads strongly limits masking effects on spectral features related to iron oxides between 450-850 nm. The influence of organic matter on the shape and albedo of the spectral curve along the entire UV-Vis+NIR spectrum, with an emphasis on specific regions of Fe oxides, was reported in studies conducted in Brazil (Dalmolin et al., 2005; Galvao and Vitorello, 1998; Moura-Bueno et al., 2019) and worldwide (Ben-Dor, 1997). This results in lower spectral variation (Viscarra Rossel and Behrens, 2010) compared with samples with higher organic matter content, such as cropland and stream channels. This explains the higher $RMSE_v$ variation in the UV-Vis range among different pre-processing methods observed for these two sources compared to the unpaved road source (Figure 5a). Consequently, the calibrated models for unpaved roads achieve a greater accuracy in estimates. In addition, the unpaved road source is also richer in hematite oxides of Fe, with prominent features compared to crop fields and stream channels (Tiecher et al., 2015). This shows that the composition of each source affects the spectral behavior of sediment differently in each spectral range (Figure 3) and, consequently, this affects in turn the predictive models calibrated for each source (Batistelli et al., 2018). These results are also in line with those observed in soil spectroscopic modeling, in which variations in the organo-mineral composition of the soil were shown to strongly influence the performance of predictive models (Araújo et al., 2014; Moura-Bueno et al., 2019; Wijewardane et al., 2016) with an emphasis on parametric multivariate methods (Lucà et al., 2017; Ramirez-Lopez et al., 2013).

By comparing the spectral ranges, it is possible to note that the $RMSE_v$ has decreased in the following order: UV-Vis > NIR > MIR (Fig. 6b). Good calibration results using MIR were also observed in previous sediment tracing studies (Collins et al., 2014; Evrard et al., 2013; Ni et al., 2019; Poulenard et al., 2012). This may be related to an increased sensitivity and an improved identification of functional groups of organic matter by MIR compared to UV-Vis and NIR (Viscarra Rossel and Behrens, 2010). Studies show that spectroscopic models calibrated with MIR region spectra have the potential to discriminate different fractions of organic matter (Knox et al., 2015) and sediment sources (Evrard et al., 2013; Ni et al., 2019). This also explains why, in the current research, the models calibrated with MIR spectra and their combinations presented the lowest variation in the $RMSE_v$ of predictions among all pre-processing techniques (Figures 5c, 5e, 5f and g), in particular for the models calibrated with the SVM method.

Our results show that there was no significant gain in terms of $RMSE_v$ when combining the UV-Vis+NIR, UV-Vis+MIR and NIR+MIR bands, but that $RMSE_v$ decreases significantly when the three bands of the electromagnetic spectrum were combined in a single model (UV-Vis+NIR+MIR) (Figure 6b). The same result has been observed in studies performed to estimate the concentration of elements in soil samples (Knox et al., 2015; Reeves, 2010; Soriano-Disla et al., 2014). The increase observed in $RMSE_v$ for models calibrated with UV-Vis range (Figure 6b) is a consequence of the lower UV-Vis spectral range (200-800 nm), which does not allow distinguishing organic matter components (Viscarra Rossel and Behrens, 2010), associated with the absence of NIR range, which is more sensitive to detect clay minerals. This implies a reduction in the predictive capacity of models calibrated with UV-Vis only or with the absence of the NIR band. This behavior is observed for models calibrated with UV-Vis+MIR range, where the absence of NIR range increased the

prediction error ($RMSE_v$) of the estimates (Figure 6b). Although the $RMSE_v$ of the models combining Vis-UV+MIR are intermediate to the $RMSE_v$ of the models calibrated using these spectral ranges separately (i.e. Vis-UV and MIR alone), it was expected that the errors would be smaller when combining them. A possible explanation for this result may be the overfitting caused by the high number of parameters and a relatively low number of artificial mixture samples used for calibrating the models (48 in total). Therefore, the low number of artificial mixture samples used to calibrate the models can be a limitation of this study, especially when combining different spectral ranges. Further efforts in future studies should evaluate the effect of the number of artificial mixture samples on the overfitting of spectroscopic models.

In summary, the explanation for this difference is associated with the interaction of UV-Vis, NIR and MIR wavelength bands, which are bands basically related to color, particle size, type of minerals and organic matter, their chemical bonds and functional groups (Viscarra Rossel and Behrens, 2010), resulting in gain of explanatory information of the spectral variation of the data (Viscarra Rossel et al., 2006). Therefore, as sediment often consists of a heterogeneous mixture, there is a greater variation of these factors, resulting in a larger amount of functional groups detected on their surface and, therefore, the combination of several wavelengths, for example, UV-Vis+NIR+MIR, enhances the performance of spectroscopic models to discriminate between contrasted sediment sources.

In general, most spectroscopic models had $RMSE_v$ values below 15% for the validation, and in some cases, values below 5% were found (Figure 5). The models calibrated to MIR, UV-Vis+NIR, UV-Vis+MIR and NIR+MIR spectra achieved $RMSE_v$ values lower than 10% in the three sediment sources (Figure 6b), and those calibrated for the UV-Vis+NIR+MIR spectral

449 range, presented values lower than 7% for crop fields and stream channels and equal to 5%
450 for unpaved roads. These values are lower than the error of 15%, which is often considered
451 to provide the acceptable level for sediment fingerprinting studies (Collins and Walling,
452 2002). This shows that the use of the spectroscopy technique employing the most
453 appropriate spectral processing approaches, spectral bands and multivariate methods has
454 a great potential for discriminating the contribution of sediment sources.

455 The lowest $RMSE_v$ values in the predictions for both multivariate methods (PLSR and SVM)
456 were observed for unpaved roads, which differed significantly from the other potential
457 sources (Figure 6c). The models calibrated with the SVM method had a better performance
458 compared to the PLSR, especially for the crop fields source (Figure 6c). This is explained by
459 the greater ability of the SVM method to model non-linear relationships (Ivanciuc, 2007;
460 Viscarra Rossel and Behrens, 2010) compared to the PLSR method. In this case, the crop
461 fields source presents a heterogeneous organo-mineral composition, mainly regarding the
462 quality of organic matter. In this scenario, non-linear relationships between the organo-
463 mineral components and the spectral variables of the crop fields source predominate. In
464 this case, SVM method is able to establish explanatory correlations of data variance,
465 resulting in more accurate estimates. This also explains why the lowest variations in
466 predictions are observed for the SVM method, which is shown to provide a more stable
467 method for the modeling of sediment sources, with an emphasis on sources with greater
468 variations in organo-mineral composition.

469

470 **3.2. Comparing sediment source predictions by spectroscopy models with those obtained**
471 **with the conventional geochemical approach**

The estimated sediment contribution of each source obtained from the best calibrated models for the UV-Vis, NIR, MIR spectral ranges and their combinations is shown in Figure 7. In general, the PLSR and SVM models calibrated with spectra of the three spectral bands and their combinations submitted to SGD1, NOR, and NOR+SGD1 pre-processing have achieved a greater accuracy in predicting sediment sources when compared to estimates derived from the geochemical method tested by Tiecher et al. (2015) (Appendix 1). The models calibrated with spectra processed by the SGD2 technique showed the lowest predictive performance among all approaches (Appendix 1). These results corroborate those observed during the calibration and validation of the models (Figs. 5 and 6).

In all modeling approaches, estimates of PLSR and SVM models clearly indicate that the main contribution of sediment is supplied by the crop fields source, followed by stream channels and unpaved roads (Figure 7), corroborating the results estimated by the geochemical method (Tiecher et al., 2015). Good agreement is observed between the contribution of the sources obtained with the geochemical tracer approach and that estimated by the PLSR and SVM models when combining the UV-Vis+NIR+MIR spectral ranges, with the crop fields source contributions amounting to 57%, 62% and 55%, respectively, followed by the unpaved roads sources, with 23%, 24% and 19%, and stream channel 20%, 20% and 21%, respectively (Figure 7g).

In general, the highest $RMSE_{sp}$ values were observed for crop fields and stream channel source estimates and the lowest for unpaved roads (Appendix 1). The predictions of crop fields and stream channel sources by the models generated with the SVM method were better than with PLSR. For the unpaved roads source, the SVM and PLSR methods presented very close $RMSE_{sp}$ values. For example, for the SVM method, the model with the

most accurate estimates was SVM-SGD1-UV-Vis+NIR+MIR, with $RMSE_{sp}$ of 17.7, 16.4 and 13.9%, and $bias_{sp}$ of 2.0, -0.7 and -2.7% for crop fields, stream channel and unpaved roads, respectively (Figure 8; Appendix 1). Accordingly, an overestimation of the crop fields contribution to sediment was found, such as an underestimation of the stream channel and thee unpaved roads source contributions. Despite the good accuracy of the models calibrated and validated with the mixtures of proportions of each source of sediment (Figure 8), there is still a considerable error compared to predictions made with fingerprinting based on geochemical tracers. This can be partly attributed to the low number of sediment samples evaluated (*n* = 29). Future studies should use a larger number of sediment samples in order to better understand this relationship. Nevertheless, the magnitude of the contributions calculated by spectroscopy and geochemical approach is very similar for each source (Figure 8).

For the PLSR method, the PLSR-NOR-UV-Vis+NIR+MIR model was more accurate, with higher $RMSE_{sp}$ for the crop fields (20.0%) and stream channel (18.7%) sources and also larger under- ($bias_{sp}$ = -4.5%) and overestimations ($bias_{sp}$ = 3.6%), respectively, and the lower $RMSE_{sp}$ (14.5%) and $bias_{sp}$ (-0.1%) were observed for unpaved roads source. The highest prediction errors observed for crop fields and stream channel sources are in accordance with the findings of Tiecher et al. (2015), who used the PLSR method and NIR spectral range to estimate sediment sources in the same study area. According to the authors, the three sources is enriched in 1:1 kaolinite type minerals. However, an increase in the abundance of 2:1 clay minerals may be observed in the following order: crop fields > stream channel >> unpaved roads, and an increase in the abundance of quartz may be found in reverse order. In addition, crop fields and stream channels also have a higher organic matter content than unpaved roads (Tiecher et al., 2016). Therefore, the closer

519 mineral composition and higher organic matter content of the crop fields and stream
520 channels provide higher spectral variations, resulting in higher prediction errors (20.0 and
521 15.8% for crop fields and stream channel, respectively) than those obtained for the
522 unpaved roads (15.5%) for the PLSR-NOR-UV-Vis+NIR+MIR model (Appendix 1).
523 Furthermore, this variation in the organo-mineral composition of the sources has a greater
524 effect on the prediction errors when looking at the estimates of the calibrated PLSR models
525 for the separate spectral ranges such as UV-Vis which showed $RMSE_{sp}$ of 23. 7, 26.6 and
526 15.7% for crop fields, stream channel and unpaved roads, respectively, for the PLSR-NOR-
527 SGD1 model; and MIR with $RMSE_{sp}$ 25.8, 25.7 and 18.6% for crop fields, stream channel
528 and unpaved roads, respectively, for the PLSR-SGD1 model. This is because the UV-Vis and
529 MIR bands are less sensitive to clay mineral types (Viscarra Rossel and Behrens, 2010) than
530 the NIR spectral band, which identifies these constituents more clearly. Therefore, for the
531 PLSR-SGD1-NIR model, no major differences in $RMSE_{sp}$ values (23.8, 16.2 and 15.8% for
532 crop fields, stream channel and unpaved roads, respectively) were observed with respect
533 to the PLSR-UV-Vis+NIR+MIR.

534 The difference between the multivariate methods is related to the statistical approach
535 followed in each method (PLSR - parametric and SVM - non-parametric). In this case, the
536 estimate of the contribution of the crop fields source obtained by the SVM-UV-
537 Vis+NIR+MIR model presented a value very similar to that observed by the geochemical
538 method (Figure 7g), indicating that this non-parametric model is more robust for the
539 spectroscopic modeling of this sediment source. Among the three sources considered, crop
540 fields had greater spectral variation (Tiecher et al., 2015) and, therefore, concomitant
541 occurrence of linear and nonlinear correlations between spectral variables and sediment.
542 In this scenario, non-parametric methods present better adjustments in the models,

543 especially SVM, which uses the kernel mathematical function to establish relationships
544 between the dependent and independent variables, in which the model seeks to identify
545 an interpolation function between the variables and creates support vectors (Ivanciuc,
546 2007). Studies have observed a similar behavior for the spectroscopic modeling of organic
547 carbon content (Lucà et al., 2017; Viscarra Rossel and Behrens, 2010), and also that of
548 exchangeable clay and calcium (Ramirez-Lopez et al., 2013) in soil samples with high
549 spectral variations.

550 The results show that there is a difference in the performance of the calibrated
551 spectroscopic models with each spectral range and their combinations. For example,
552 models calibrated only with spectrum in the UV-Vis range have the highest error for both
553 PLSR and SVM, with $RMSE_{sp}$ values > 22% for crop fields; $RMSE_{sp}$ > 26% for stream channels
554 and $RMSE_{sp}$ > 20% for unpaved roads (Appendix 1). By contrast, the lowest errors were
555 achieved in the UV-Vis+NIR+MIR ranges, where $RMSE_{sp}$ values were ~18% for crop fields
556 and ~16% for stream channels and unpaved roads. This shows that when using narrower
557 spectrum bands there is a loss of information and consequently a loss of discriminating
558 power of the models. However, the models that combined the three UV-Vis+NIR+MIR
559 spectral bands (Figure 7g) achieved a greater accuracy due to the better discrimination of
560 the inherent compositional characteristics of each source, as all major components that
561 may influence the spectral behavior (organic matter, clays and oxides) are taken into
562 account in the spectra (Knox et al., 2015; Reeves, 2010; Viscarra Rossel et al., 2006).

563 Other differences in error metrics with respect to spectral ranges are observed by the
564 higher predictive capability of UV-Vis+NIR+MIR (Appendix 1) compared to models using
565 only UV-Vis+NIR range. This is in accordance with the findings of Reeves (2010) and Knox et

al. (2015) for soil organic carbon estimation. Furthermore, a study conducted by Bellon-Maurel and McBratney (2011) and Knox et al. (2015) showed that models developed to quantify organic carbon content using MIR region data only produce slightly better results than the UV-Vis+NIR region. However, this was not observed in the current research, where models using only MIR showed a lower performance compared to that obtained with UV-Vis+NIR (Appendix 1). It should be noted that in this study we are modeling sediment source contributions, which is very different from obtaining spectroscopic estimates of elemental concentrations. Sediments consist of a mixture of particles with different contents and types of clays, Fe oxides and organic matter. In this case, UV-Vis and NIR spectral ranges have potential to discriminate between contrasted contents and types of clays and Fe oxides (Viscarra Rossel and Behrens, 2010). They are therefore important for discriminating between contrasted sediment sources, as already reported in the literature (Collins et al., 2014; Legout et al., 2013; Pulley and Rowntree, 2016; Tiecher et al., 2016).

Moreover, the estimates obtained by the models in the UV-Vis and MIR spectral ranges (Figure 7a, 7c, respectively) showed a greater dispersion between the PLSR and SVM models and the three sediment sources. The same behavior is observed for the combination UV-Vis+MIR (Figure 7e), particularly for crop fields. This may be attributed to the interaction of UV-Vis and MIR wavelengths, which are bands related to the content and type of Fe oxides, and functional groups of organic matter, respectively (Viscarra Rossel and Behrens, 2010). As sediment is a rather heterogeneous mixture, mainly supplied by crop fields, there is a greater compositional variation in these samples, which in this case, the UV-Vis and MIR spectra are unable to capture. This observation interferes in the correlations between sediment and spectral bands and, therefore, in the predictive power

590 of models. By contrast, the NIR spectral band resulted in estimates very close to those
591 obtained by the geochemical method, with emphasis on the crop fields and unpaved roads
592 sources (Figure 7b). The explanation for this is due to the wavelengths corresponding to
593 the NIR region being able to jointly identify particle size, type of minerals and organic
594 matter, and supertons of chemical bonds and functional groups (Viscarra Rossel and
595 Behrens, 2010). Furthermore, the combinations of spectral bands with NIR (such as
596 NIR+MIR and UV-Vis+NIR+MIR) present the same tendency as that observed for NIR, i.e.,
597 lower amplitude in estimates (Figure 7f, 7g). This shows that the NIR spectral range is the
598 most important region of the electromagnetic spectrum for building spectroscopic models
599 for estimating sediment source contributions. This confirms previous findings obtained in
600 studies conducted at different locations around the world, which showed the good
601 performance of models that use data derived from the NIR spectral range to estimate
602 sediment source contributions (Collins et al., 2014; Tiecher et al., 2016).

603

604 **3.3. Assessing the quality of the models**

605 It is important to note that in all models tested here, the contribution of each source is
606 estimated independently, i.e. each model estimates the proportion of a source
607 independently of the other two sources. Therefore, the sum of the estimates generated for
608 each source can provide a good indicator of model quality. In this case, it is understood
609 that models with the sum of contributions from sediment sources closer to 100% are
610 better (Legout et al., 2013). Figure 8 shows this comparison, where it is observed that
611 under and overestimation (ranging from 90 to 132% - Figure 9a) of the sum of the sources
612 for some approaches occur. Regarding the spectral pre-processing techniques,

613 overestimation occurs for the models calibrated for both multivariate methods (PLSR and
614 SVM) with raw spectra (RAW) and submitted to SMO, SGD2, SNV, MSC and NOR techniques
615 (Figure 9a). Moreover, a slight underestimation is observed for MSC+SGD1. It is noted that
616 RAW and SMO spectra presented the highest overestimates and showed significant
617 differences compared to the others, indicating that the absence of preprocessing of
618 spectral data and/or only the smoothing provided less accurate estimates of source
619 proportions. It is also possible to note that the NOR+SGD1 and SGD1 techniques presented
620 the lowest variations in the sum of the sediment source contributions (Figure 8a). This
621 corroborates the best performance observed for these models during the validation step
622 (Figure 6a). The models that reached values closer to 100% were PLSR-SGD1 (100.6%) and
623 SVM-SGD1 (100.1%) (Figure 9a).

624 Regarding the spectral ranges, it is noted that the models calibrated with the combinations
625 NIR+MIR and UV-Vis+NIR+MIR presented the values closest to 100%, differing significantly
626 from the others (Figure 8b). The model calibrated with the SVM method and the
627 combination of UV-Vis+NIR+MIR presented the values closest to 100% (100.1%) (Figure
628 8b). Therefore, the results observed in Figure 8 corroborate those discussed in the section
629 dealing with the accuracy of the model validation and in the estimates of the sediment
630 source contributions. In addition, future studies should address the use of spectral variable
631 selection algorithms. This strategy has shown the potential to improve spectroscopic
632 estimates (Xiaobo et al., 2010; Gomes et al., 2013; Hong et al., 2020). Additionally, research
633 employing the use of two-dimensional correlation (Hong et al., 2018) to identify regions or
634 bands most correlated with different sediment sources can be a promising approach.

635 Finally, the decision tree analysis shows that the quality of calibration of spectroscopic
636 models depends primarily on the spectral preprocessing technique, and secondarily on the
637 spectral band, and that the sediment source has little or no influence (Figure 10b). It is
638 evident that pre-processing with SGD2 and UV-Vis spectral band always result in higher
639 $RMSE_v$ values. However, when comparing the quality of the models based on the estimates
640 of sediment sources obtained with geochemical tracers ($RMSE_{sp}$, Figure 10b), the spectral
641 band and sediment source is of greater importance, and NIR range or its combination with
642 the other spectral ranges result in contributions that are more similar to those obtained
643 with the geochemical approach.

644 **4. Conclusions**

645 The current research demonstrated the great potential to improve the estimation of the
646 sediment source contributions using spectroscopy when using adequate spectral pre-
647 processing technique, multivariate method, and spectral range. In general, the non-
648 parametric support vector machine (SVM) model was more robust than the partial last
649 square regression (PLSR), especially to estimate the contribution of sediment sources with
650 high organo-mineral variations, such as the crop fields source. For both models tested
651 (PLSR and SVM), a better performance was obtained using Savitzky-Golay spectral pre-
652 processing techniques with 1st derivative (SGD1), normalization (NOR) and combining
653 NOR+SGD1. Furthermore, it was verified that the combination of the three spectral ranges
654 of the electromagnetic spectrum tested (UV-Vis, NIR and MIR) enhanced the performance
655 of the spectroscopic models, resulting in lower errors in the predictions of the sediment
656 source contributions. This is due to the sum of different information contained in each
657 spectral range related to the organic and mineral composition of each sediment source.

Despite the good accuracy of the models calibrated and validated with the mixtures, significant errors remain when comparing sediment source contributions c to the results obtained with the conventional sediment fingerprinting method based on geochemical tracers. Nevertheless, the magnitude of the contributions calculated by spectroscopy and geochemical approaches remains very similar for all sources. Efforts should be done in future studies to validate these findings in larger catchments as well as in sites where more potential sediment sources may supply material to the river systems.

Acknowledgements

The first author thanks the National Council for Scientific and Technological Development (CNPq) for the research grant 311788/2019-0.

References

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorol. Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Angelopoulou, T., Balafoutis, A., Zalidis, G., Bochtis, D., 2020. From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review. Sustainability 12, 443. https://doi.org/10.3390/su12020443

Araújo, S.R., Wetterlind, J., Demattê, J.A.M., Stenberg, B., 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. Eur. J. Soil Sci. 65, 718–729.

https://doi.org/10.1111/ejss.12165

Barthod, L.R.M., Liu, K., Lobb, D.A., Owens, P.N., Martínez-Carreras, N., Koiter, A.J., Petticrew, E.L., McCullough, G.K., Liu, C., Gaspar, L., 2015. Selecting Color-based Tracers and Classifying Sediment Sources in the Assessment of Sediment Dynamics Using Sediment Source Fingerprinting. J. Environ. Qual. 44, 1605. https://doi.org/10.2134/jeq2015.01.0043

Batistelli, M., Martínez Bilesio, A.R., García-Reiriz, A.G., 2018. Development of a fast and inexpensive method for detecting the main sediment sources in a river basin. Microchem. J. 142, 208–218. https://doi.org/10.1016/j.microc.2018.06.040

Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol. Biochem. 43, 1398–1410. https://doi.org/10.1016/j.soilbio.2011.02.019

Ben-Dor, E., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. Remote Sens. Environ. 61, 1–15. https://doi.org/10.1016/S0034-4257(96)00120-4

Boudreault, M., Koiter, A.J., Lobb, D.A., Liu, K., Benoy, G., Owens, P.N., Danielescu, S., Li, S., 2018. Using colour, shape and radionuclide fingerprints to identify sources of sediment in an agricultural watershed in Atlantic Canada. Can. Water Resour. J. / Rev. Can. des ressources hydriques 43, 347–365. https://doi.org/10.1080/07011784.2018.1451781

Brosinsky, A., Foerster, S., Segl, K., Kaufmann, H., 2014a. Spectral fingerprinting: sediment source discrimination and contribution modelling of artificial mixtures based on VNIR-SWIR spectral properties. J. Soils Sediments 14, 1949–1964. https://doi.org/10.1007/s11368-014-0925-1

Brosinsky, A., Foerster, S., Segl, K., López-Tarazón, J.A., Piqué, G., Bronstert, A., 2014b. Spectral fingerprinting: characterizing suspended sediment sources by the use of VNIR-SWIR spectral information. J. Soils Sediments 14, 1965–1981. https://doi.org/10.1007/s11368-014-0927-z

705 Buddenbaum, H., Steffens, M., 2012. The effects of spectral pretreatments on chemometric
706 analyses of soil profiles using laboratory imaging spectroscopy. Appl. Environ. Soil Sci. 2012.
707 https://doi.org/10.1155/2012/274903

708 Chapkanski, S., Ertlen, D., Rambeau, C., Schmitt, L., 2019. Provenance discrimination of fine
709 sediments by mid-infrared spectroscopy: Calibration and application to fluvial
710 palaeo-environmental reconstruction. Sedimentology. https://doi.org/10.1111/sed.12678

711 Collins, A.. L., Walling, D.. E., 2002. Selecting fingerprint properties for discriminating potential
712 suspended sediment sources in river basins. J. Hydrol. 261, 218–244.
713 https://doi.org/10.1016/S0022-1694(02)00011-2

714 Collins, A.L., Williams, L.J., Zhang, Y.S., Marius, M., Dungait, J.A.J., Smallman, D.J., Dixon, E.R.,
715 Stringfellow, A., Sear, D.A., Jones, J.I., Naden, P.S., 2014. Sources of sediment-bound organic
716 matter infiltrating spawning gravels during the incubation and emergence life stages of
717 salmonids. Agric. Ecosyst. Environ. 196, 76–93. https://doi.org/10.1016/j.agee.2014.06.018

718 Collins, A.L., Williams, L.J., Zhang, Y.S., Marius, M., Dungait, J.A.J., Smallman, D.J., Dixon, E.R.,
719 Stringfellow, A., Sear, D.A., Jones, J.I., Naden, P.S., 2013. Catchment source contributions to
720 the sediment-bound organic matter degrading salmonid spawning gravels in a lowland river,
721 southern England. Sci. Total Environ. 456–457, 181–195.
722 https://doi.org/10.1016/j.scitotenv.2013.03.093

723 Comino, J.R., Brings, C., Lassu, T., Iserloh, T., Senciales, J.M., Martínez Murillo, J.F., Ruiz Sinoga, J.D.,
724 Seeger, M., Ries, J.B., 2015. Rainfall and human activity impacts on soil losses and rill erosion
725 in vineyards (Ruwer Valley, Germany). Solid Earth 6, 823–837. https://doi.org/10.5194/se-6-
726 823-2015

727 Cooper, R.J., Rawlins, B.G., L??z??, B., Krueger, T., Hiscock, K.M., Lézé, B., Krueger, T., Hiscock, K.M.,
728 2014. Combining two filter paper-based analytical methods to monitor temporal variations in
729 the geochemical properties of fluvial suspended particulate matter. Hydrol. Process. 28, 4042–

730 4056. https://doi.org/10.1002/hyp.9945

731 D'Haen, K., Verstraeten, G., Degryse, P., 2012. Fingerprinting historical fluvial sediment fluxes. Prog.
732 Phys. Geogr. 36, 154–186. https://doi.org/10.1177/0309133311432581

733 Dalmolin, R.S.D., Gonçalves, C.N., Klamt, E., Dick, D.P., 2005. Relação entre os constituintes do solo
734 e seu comportamento espectral. Ciência Rural 35, 481–489. https://doi.org/10.1590/S0103-
735 84782005000200042

736 Davis, C.M., Fox, J.F., 2009. Sediment Fingerprinting: Review of the Method and Future
737 Improvements for Allocating Nonpoint Source Pollution. J. Environ. Eng. 135, 490–504.
738 https://doi.org/10.1061/(ASCE)0733-9372(2009)135:7(490)

739 Didoné, E.J., Minella, J.P.G., Merten, G.H., 2015. Quantifying soil erosion and sediment yield in a
740 catchment in southern Brazil and implications for land conservation. J. Soils Sediments 15,
741 2334–2346. https://doi.org/10.1007/s11368-015-1160-0

742 Dodd, R.J., McDowell, R.W., Condron, L.M., 2014. Is tillage an effective method to decrease
743 phosphorus loss from phosphorus enriched pastoral soils? Soil Tillage Res. 135, 1–8.
744 https://doi.org/10.1016/j.still.2013.08.015

745 Dodd, R.J., Sharpley, A.N., 2015. Recognizing the role of soil organic phosphorus in soil fertility and
746 water quality. Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2015.10.001

747 Dotto, A.C., Dalmolin, R.S.D., Grunwald, S., ten Caten, A., Pereira Filho, W., 2017. Two
748 preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR
749 spectroscopy. Soil Tillage Res. 172, 59–68. https://doi.org/10.1016/j.still.2017.05.008

750 Dotto, A.C., Dalmolin, R.S.D., ten Caten, A., Grunwald, S., 2018. A systematic study on the
751 application of scatter-corrective and spectral-derivative preprocessing for multivariate
752 prediction of soil organic carbon by Vis-NIR spectra. Geoderma 314, 262–274.
753 https://doi.org/10.1016/j.geoderma.2017.11.006

754 Erkossa, T., Wudneh, A., Desalegn, B., Taye, G., 2015. Linking soil erosion to on-site financial cost:
755 Lessons from watersheds in the Blue Nile basin. Solid Earth 6, 765–774.
756 https://doi.org/10.5194/se-6-765-2015

757 Evrard, O., Poulenard, J., Némery, J., Ayrault, S., Gratiot, N., Duvert, C., Prat, C., Lefèvre, I., Bonté,
758 P., Esteves, M., 2013. Tracing sediment sources in a tropical highland catchment of central
759 Mexico by using conventional and alternative fingerprinting methods. Hydrol. Process. 27,
760 911–922. https://doi.org/10.1002/hyp.9421

761 Galvao, L.S., Vitorello, I., 1998. Role of organic matter in obliterating the effects of iron on spectral
762 reflectance and colour of Brazilian tropical soils. Int. J. Remote Sens. 19, 1969–1979.
763 https://doi.org/10.1080/014311698215090

764 Haddadchi, A., Ryder, D.S., Evrard, O., OLLEY, J., 2013. Sediment fingerprinting in fluvial systems:
765 review of tracers, sediment sources and mixing models. Int. J. Sediment Res. 28, 560–578.
766 https://doi.org/10.1016/S1001-6279(14)60013-5

767 IUSS Working Group WRB, 2015. World reference base for soil resources 2014, update 2015
768 International soil classification system for naming soils and creating legends for soil maps.
769 World Soil Resources Reports No. 106. FAO, Rome.

770 Ivanciuc, O., 2007. Applications of Support Vector Machines in Chemistry, in: Lipkowitz, K.B.,
771 Cundari, T.R. (Eds.), Reviews in Computational Chemistry. Wiley-VCH, Weinheim, pp. 291–400.
772 https://doi.org/10.1002/9780470116449.ch6

773 Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B., Harris, W.G., 2015. Modelling
774 soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy.
775 Geoderma 239–240, 229–239. https://doi.org/10.1016/j.geoderma.2014.10.019

776 Koiter, A.J., Owens, P.N., Petticrew, E.L., Lobb, D.A., 2013. The behavioural characteristics of
777 sediment properties and their implications for sediment fingerprinting as an approach for
778 identifying sediment sources in river basins. Earth-Science Rev. 125, 24–42.

779 https://doi.org/10.1016/j.earscirev.2013.05.009

780 Legout, C., Poulenard, J., Nemery, J., Navratil, O., Grangeon, T., Evrard, O., Esteves, M., 2013.
781 Quantifying suspended sediment sources during runoff events in headwater catchments using
782 spectrocolorimetry. J. Soils Sediments 13, 1478–1492. https://doi.org/10.1007/s11368-013-
783 0728-9

784 Liu, C., Li, Z., Berhe, A.A., Zeng, G., Xiao, H., Liu, L., Wang, D., Peng, H., 2019. Chemical
785 characterization and source identification of organic matter in eroded sediments: Role of land
786 use and erosion intensity. Chem. Geol. 506, 97–112.
787 https://doi.org/10.1016/j.chemgeo.2018.12.040

788 Liu, K., Lobb, D.A., Miller, J., Owens, P., Caron, M., 2017. Determining sources of fine-grained
789 sediment for a reach of the Lower Little Bow River, Alberta, using a colour-based sediment
790 fingerprinting approach. Can. J. Soil Sci. 98, CJSS-2016-0131. https://doi.org/10.1139/CJSS-
791 2016-0131

792 Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G., Buttafuoco, G., 2017. Effect of calibration set
793 size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma 288, 175–
794 183. https://doi.org/10.1016/j.geoderma.2016.11.015

795 Magnusson, M., Heimann, K., Ridd, M., Negri, A.P., 2013. Pesticide contamination and phytotoxicity
796 of sediment interstitial water to tropical benthic microalgae. Water Res. 47, 5211–21.
797 https://doi.org/10.1016/j.watres.2013.06.003

798 Martínez-Carreras, N., Krein, A., Gallart, F., Iffly, J.F., Pfister, L., Hoffmann, L., Owens, P.N., 2010a.
799 Assessment of different colour parameters for discriminating potential suspended sediment
800 sources and provenance: A multi-scale study in Luxembourg. Geomorphology 118, 118–129.
801 https://doi.org/10.1016/j.geomorph.2009.12.013

802 Martínez-Carreras, N., Krein, A., Udelhoven, T., Gallart, F., Iffly, J.F., Hoffmann, L., Pfister, L.,
803 Walling, D.E., 2010b. A rapid spectral-reflectance-based fingerprinting approach for

804 documenting suspended sediment sources during storm runoff events. J. Soils Sediments 10,
805 400–413. https://doi.org/10.1007/s11368-009-0162-1

806 Martínez-Carreras, N., Udelhoven, T., Krein, A., Gallart, F., Iffly, J.F., Ziebel, J., Hoffmann, L., Pfister,
807 L., Walling, D.E., 2010c. The use of sediment colour measured by diffuse reflectance
808 spectrometry to determine sediment sources: Application to the Attert River catchment
809 (Luxembourg). J. Hydrol. 382, 49–63. https://doi.org/10.1016/j.jhydrol.2009.12.017

810 McBratney, A.B., Minasny, B., Viscarra Rossel, R., 2006. Spectral soil analysis and inference systems:
811 A powerful combination for solving the soil data crisis. Geoderma 136, 272–278.
812 https://doi.org/10.1016/j.geoderma.2006.03.051

813 Merten, G.H., Araújo, A.G., Biscaia, R.C.M., Barbosa, G.M.C., Conte, O., 2015. No-till surface runoff
814 and soil losses in southern Brazil. Soil Tillage Res. 152, 85–93.
815 https://doi.org/10.1016/j.still.2015.03.014

816 Mevik, B.-H., Wehrens, R., Liland, K.H., 2016. Partial Least Squares and Principal Component
817 Regression. Packag. R CRAN.

818 Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., Lin, C.-C., 2019. Misc
819 Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU
820 Wien. R Packag. version 1.7-3.

821 Minella, J.P.G., Walling, D.E., Merten, G.H., 2014. Establishing a sediment budget for a small
822 agricultural catchment in southern Brazil, to support the development of effective sediment
823 management strategies. J. Hydrol. 519, 2189–2201.
824 https://doi.org/10.1016/j.jhydrol.2014.10.013

825 Moura-Bueno, J.M., Dalmolin, R.S.D., ten Caten, A., Dotto, A.C., Demattê, J.A.M., 2019.
826 Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more
827 accurate soil organic carbon predictions. Geoderma 337, 565–581.
828 https://doi.org/10.1016/j.geoderma.2018.10.015

829 Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M., 2016. Estimating the soil clay content
830 and organic matter by means of different calibration methods of vis-NIR diffuse reflectance
831 spectroscopy. Soil Tillage Res. 155, 510–522. https://doi.org/10.1016/j.still.2015.07.021

832 Ni, L.S., Fang, N.F., Shi, Z.H., Tan, W.F., 2019. Mid-infrared spectroscopy tracing of channel erosion
833 in highly erosive catchments on the Chinese Loess Plateau. Sci. Total Environ.
834 https://doi.org/10.1016/j.scitotenv.2019.06.116

835 Nosrati, K., Akbari-mahdiabad, M., Ayoubi, S., Degos, E., Koubansky, A., 2020. Storm dust source
836 fingerprinting for different particle size fractions using colour and magnetic susceptibility and
837 a Bayesian un-mixing model.

838 Pinheiro, É., Ceddia, M., Clingensmith, C., Grunwald, S., Vasques, G., 2017. Prediction of Soil
839 Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance
840 Spectroscopy in the Central Amazon. Remote Sens. 9, 293. https://doi.org/10.3390/rs9040293

841 Poulenard, J., Dorioz, J.-M., Elsass, F., 2008. Analytical Electron-Microscopy Fractionation of Fine
842 and Colloidal Particulate-Phosphorus in Riverbed and Suspended Sediments. Aquat.
843 Geochemistry 14, 193–210. https://doi.org/10.1007/s10498-008-9032-5

844 Poulenard, J., Legout, C., Némery, J., Bramorski, J., Navratil, O., Douchin, A., Fanget, B., Perrette, Y.,
845 Evrard, O., Esteves, M., 2012. Tracing sediment sources during floods using Diffuse
846 Reflectance Infrared Fourier Transform Spectrometry (DRIFTS): A case study in a highly erosive
847 mountainous catchment (Southern French Alps). J. Hydrol. 414–415, 452–462.
848 https://doi.org/10.1016/j.jhydrol.2011.11.022

849 Poulenard, J., Perrette, Y., Fanget, B., Quetin, P., Trevisan, D., Dorioz, J.M., 2009. Infrared
850 spectroscopy tracing of sediment sources in a small rural watershed (French Alps). Sci. Total
851 Environ. 407, 2808–2819. https://doi.org/10.1016/j.scitotenv.2008.12.049

852 Pulley, S., Rowntree, K., 2016. The use of an ordinary colour scanner to fingerprint sediment
853 sources in the South African Karoo. J. Environ. Manage. 165, 253–262.

https://doi.org/10.1016/j.jenvman.2015.09.037

Pulley, S., Van der Waal, B., Rowntree, K., Collins, A.L., 2018. Colour as reliable tracer to identify the sources of historically deposited flood bench sediment in the Transkei, South Africa: A comparison with mineral magnetic tracers before and after hydrogen peroxide pre-treatment. CATENA 160, 242–251. https://doi.org/10.1016/j.catena.2017.09.018

R Core Team, 2020. R: A Language and Environment for Statistical Computing [WWW Document]. URL https://www.r-project.org/

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013. The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma 195–196, 268–279. https://doi.org/10.1016/j.geoderma.2012.12.014

Ramon, R., 2017. Kinetic energy measurement of rainfall and defining a pluvial index to estimate erosivity in Arvorezinha/RS. Federal University of Santa Maria.

Reeves, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? Geoderma 158, 3–14. https://doi.org/10.1016/j.geoderma.2009.04.005

Rinnan, Åsmund, Berg, F. van den, Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28, 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Seutloali, K.E., Beckedahl, H.R., 2015. Understanding the factors influencing rill erosion on roadcuts in the south eastern region of South Africa. Solid Earth 6, 633–641. https://doi.org/10.5194/se-6-633-2015

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev. 49, 139–186.

878 https://doi.org/10.1080/05704928.2013.811081

879 Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of Soil
880 Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy.
881 PLoS One 8, e66409. https://doi.org/10.1371/journal.pone.0066409

882 Stevens, A., Ramirez-Lopez, L., 2020. An introduction to the prospectr package [WWW Document].
883 URL https://cran.r-project.org/web/packages/prospectr/index.html

884 Taguas, E.V., Guzmán, E., Guzmán, G., Vanwalleghem, T., Gómez, J.A., 2015. Characteristics and
885 importance of rill and gully erosion: a case study in a small catchment of a marginal olive
886 grove. Cuad. Investig. Geográfica 41, 107. https://doi.org/10.18172/cig.2644

887 Tiecher, T., Caner, L., Minella, J.P.G., Bender, M.A., dos Santos, D.R., 2016. Tracing sediment sources
888 in a subtropical rural catchment of southern Brazil by using geochemical tracers and near-
889 infrared spectroscopy. Soil Tillage Res. 155, 478–491.
890 https://doi.org/10.1016/j.still.2015.03.001

891 Tiecher, T., Caner, L., Minella, J.P.G., Evrard, O., Mondamert, L., Labanowski, J., Rheinheimer, D. dos
892 S., 2017. Tracing Sediment Sources Using Mid-infrared Spectroscopy in Arvorezinha
893 Catchment, Southern Brazil. L. Degrad. Dev. 28, 1603–1614. https://doi.org/10.1002/ldr.2690

894 Tiecher, T., Caner, L., Minella, J.P.G., Santos, D.R. dos, 2015. Combining visible-based-color
895 parameters and geochemical tracers to improve sediment source discrimination and
896 apportionment. Sci. Total Environ. 527–528, 135–149.
897 https://doi.org/10.1016/j.scitotenv.2015.04.103

898 Tiecher, T., Ramon, R., Laceby, J.P., Evrard, O., Minella, J.P.G., 2019. Potential of phosphorus
899 fractions to trace sediment sources in a rural catchment of Southern Brazil: Comparison with
900 the conventional approach based on elemental geochemistry. Geoderma 337, 1067–1076.
901 https://doi.org/10.1016/j.geoderma.2018.11.011

902 Uber, M., Legout, C., Nord, G., Crouzet, C., Demory, F., Poulenard, J., 2019. Comparing alternative
903 tracing measurements and mixing models to fingerprint suspended sediment sources in a
904 mesoscale Mediterranean catchment. J. Soils Sediments 19, 3255–3273.
905 https://doi.org/10.1007/s11368-019-02270-1

906 Valente, M.L., Reichert, J.M., Legout, C., Tiecher, T., Cavalcante, R.B.L., Evrard, O., 2020.
907 Quantification of sediment source contributions in two paired catchments of the Brazilian
908 Pampa using conventional and alternative fingerprinting approaches. Hydrol. Process.
909 https://doi.org/10.1002/hyp.13768

910 Varmuza, K., Filzmoser, P., 2009. Introduction to Multivariate Statistical Analysis in Chemometrics.
911 CRC Press.

912 Vasques, G.M., Grunwald, S., Sickman, J.O., 2008. Comparison of multivariate methods for
913 inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146, 14–25.
914 https://doi.org/10.1016/j.geoderma.2008.04.007

915 Vercruysse, K., Grabowski, R.C., 2018. Using source-specific models to test the impact of sediment
916 source classification on sediment fingerprinting. Hydrol. Process. 32, 3402–3415.
917 https://doi.org/10.1002/hyp.13269

918 Verheyen, D., Diels, J., Kissi, E., Poesen, J., 2014. The use of visible and near-infrared reflectance
919 measurements for identifying the source of suspended sediment in rivers and comparison
920 with geochemical fingerprinting. J. Soils Sediments 14, 1869–1885.
921 https://doi.org/10.1007/s11368-014-0938-9

922 Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse
923 reflectance spectra. Geoderma 158, 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

924 Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible,
925 near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous
926 assessment of various soil properties. Geoderma 131, 59–75.

927 https://doi.org/10.1016/j.geoderma.2005.03.007

928 Walling, D.E., Woodward, J.C., 1995. Tracing sources of suspended sediment in river basins: a case
929 study of the River Culm, Devon, UK. Mar. Freshw. Res. 46, 327–336.
930 https://doi.org/10.1071/MF9950327

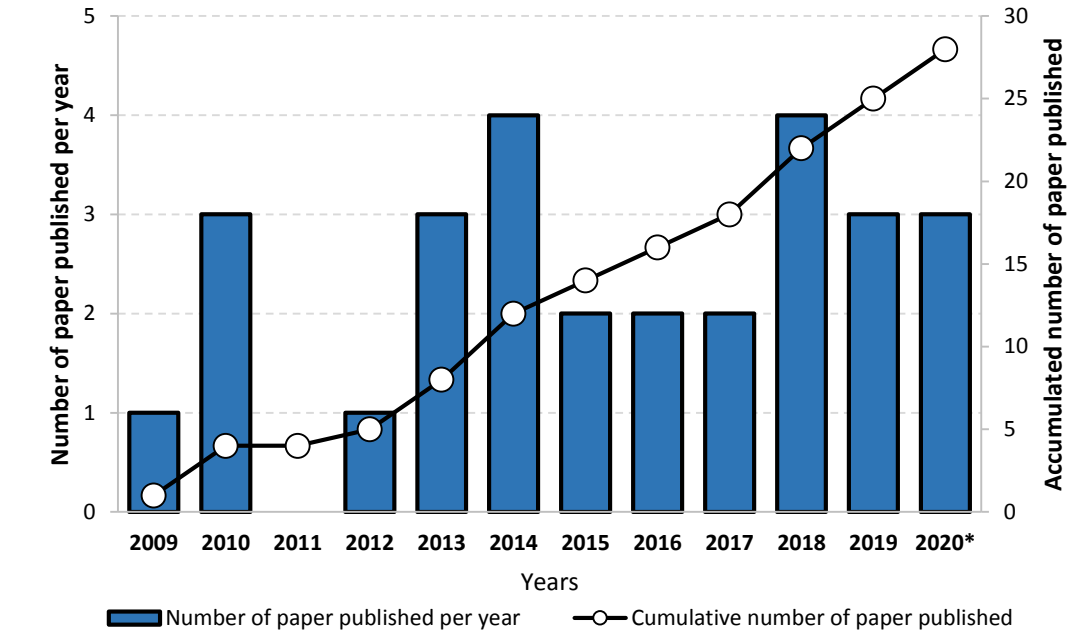
931 Walesiak M, Dudek A. 2020. The Choice of Variable Normalization Method in Cluster Analysis. In
932 Soliman KS (ed.), Education Excellence and Innovation Management: A 2025 Vision to Sustain
933 Economic Development During Global Challenges, 325-340.

934 Wijewardane, N.K., Ge, Y., Wills, S., Loecke, T., 2016. Prediction of Soil Carbon in the Conterminous
935 United States: Visible and Near Infrared Reflectance Spectroscopy Analysis of the Rapid
936 Carbon Assessment Project. Soil Sci. Soc. Am. J. 80, 973–982.
937 https://doi.org/10.2136/sssaj2016.02.0052

938 Yahia, D., Elsharkawy, E.E., 2014. Multi pesticide and PCB residues in Nile tilapia and catfish in
939 Assiut city, Egypt. Sci. Total Environ. 466–467, 306–314.
940 https://doi.org/10.1016/j.scitotenv.2013.07.002

941

Figure



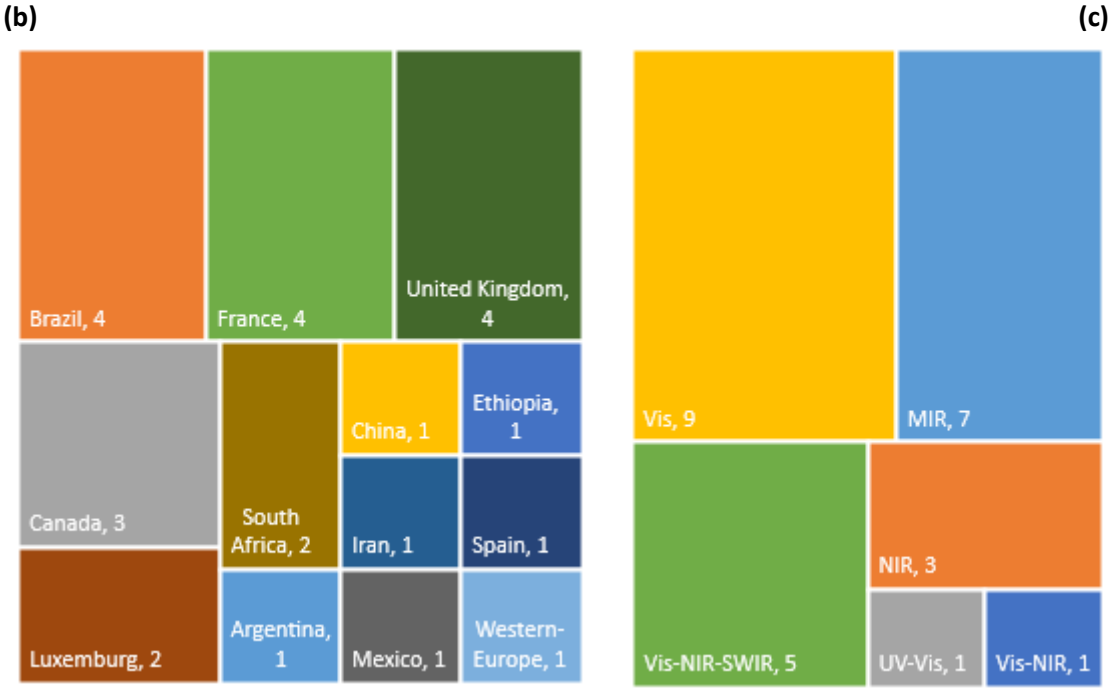


Figure 1.
Number of published scientific articles per year and cumulative number of published articles using spectroscopy to trace sediment sources for the period 2009-2020 (a), relative distribution of spectroscopic fingerprinting studies by country (b) and spectral range (c). *Until 15th June 2020. UV, ultraviolet. Vis, visible. NIR, near infrared. SWIR, short-wave infrared. MIR, mid infrared.

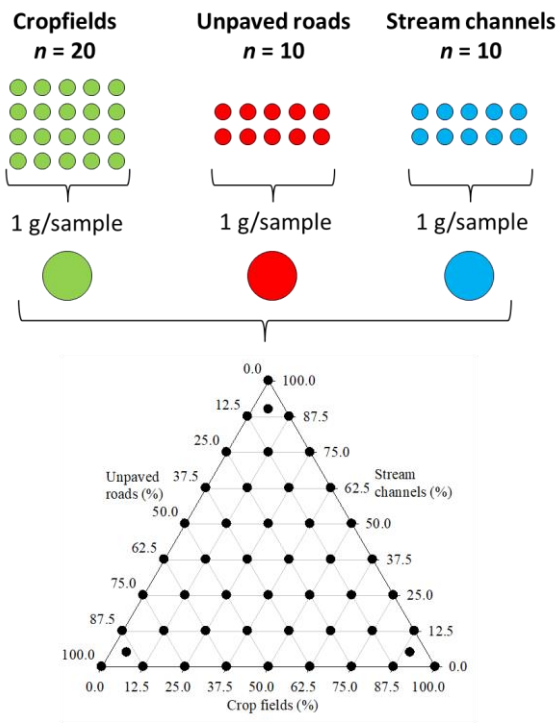


Figure 2.

Ternary diagram with the position of the experimental mixtures prepared for the calibration and validation of the spectroscopic-models.

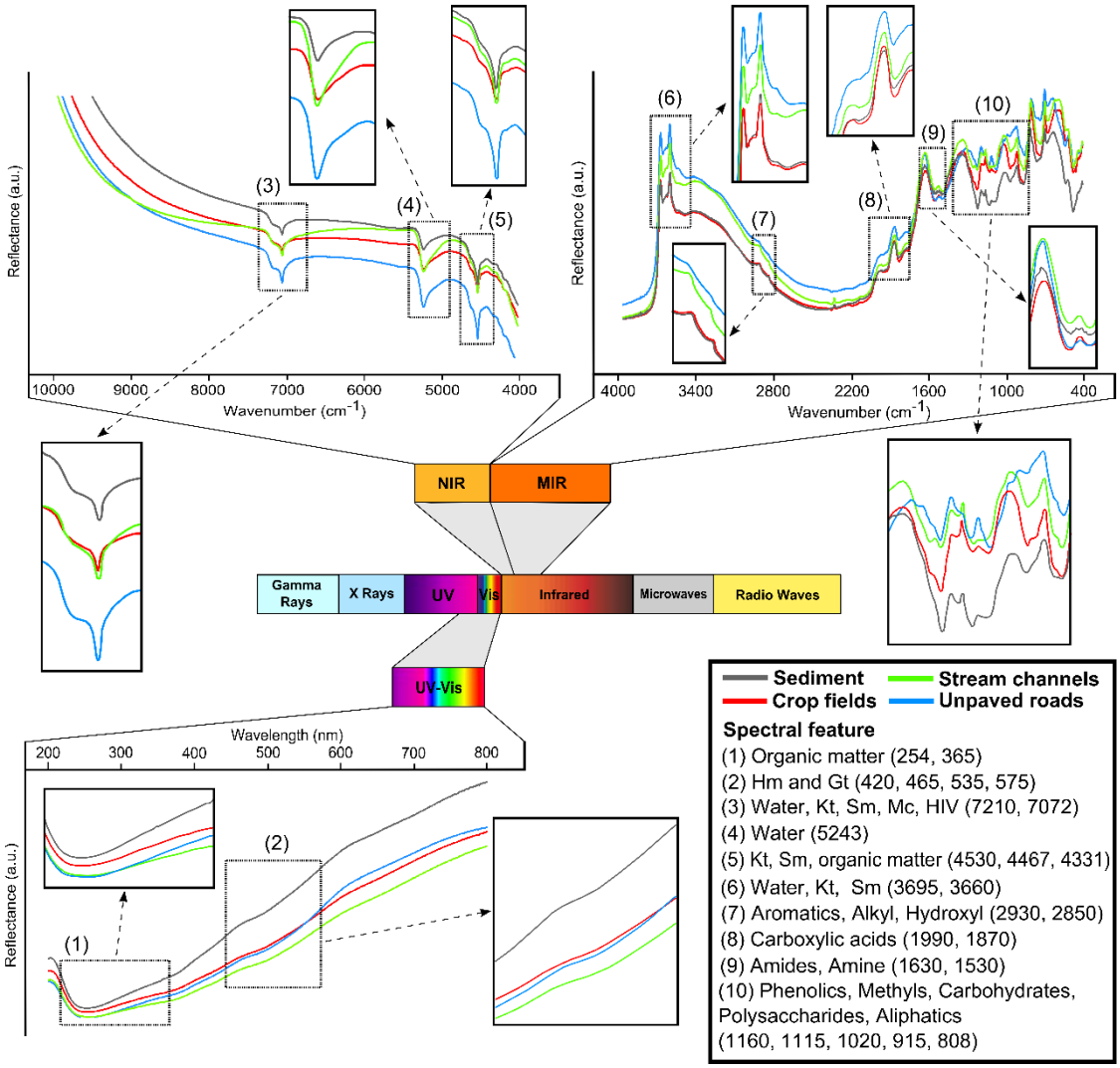


Figure 3.

Characterization of the main spectral features found in the UV-Vis, NIR and MIR ranges for the suspended sediment and potential sediment sources in the Arvorezinha catchment.

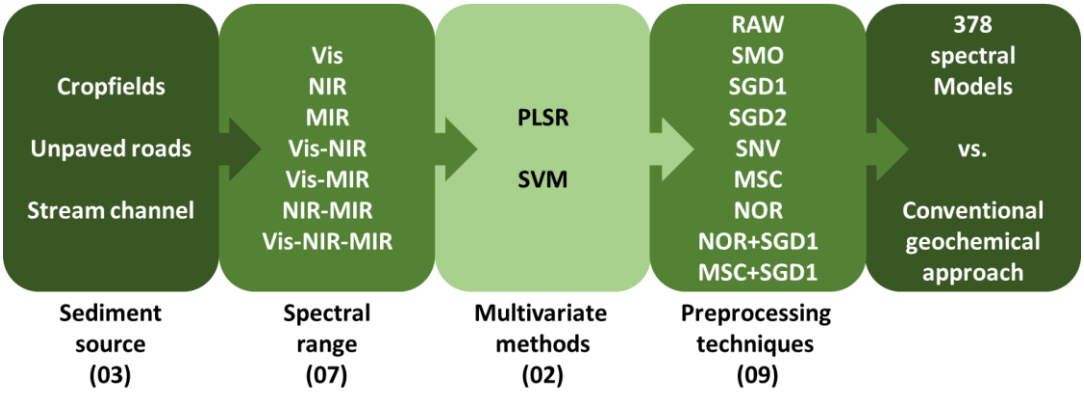


Figure 4.

Schematic construction of predictive models of the spectral signature of sediment sources.

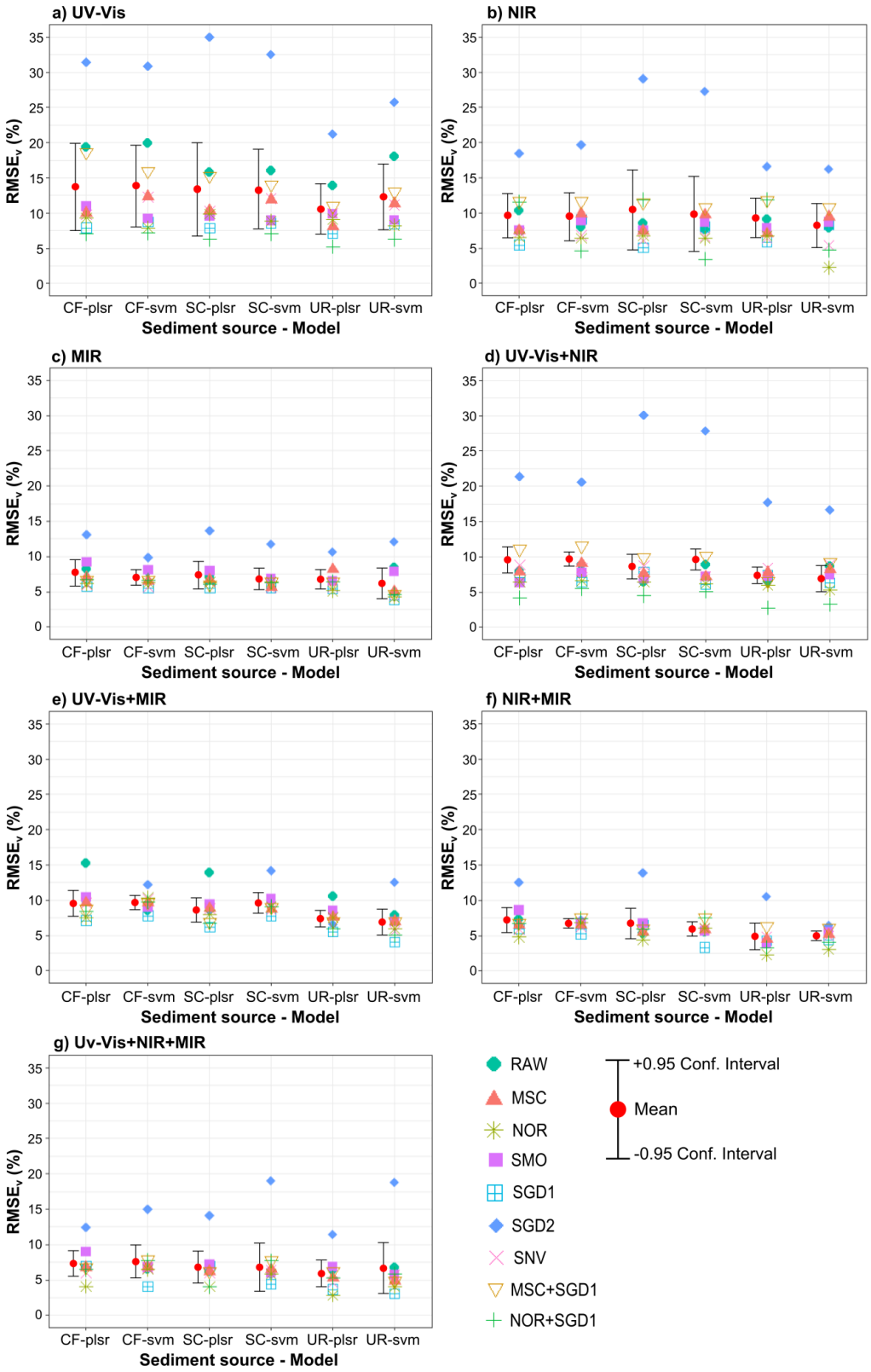
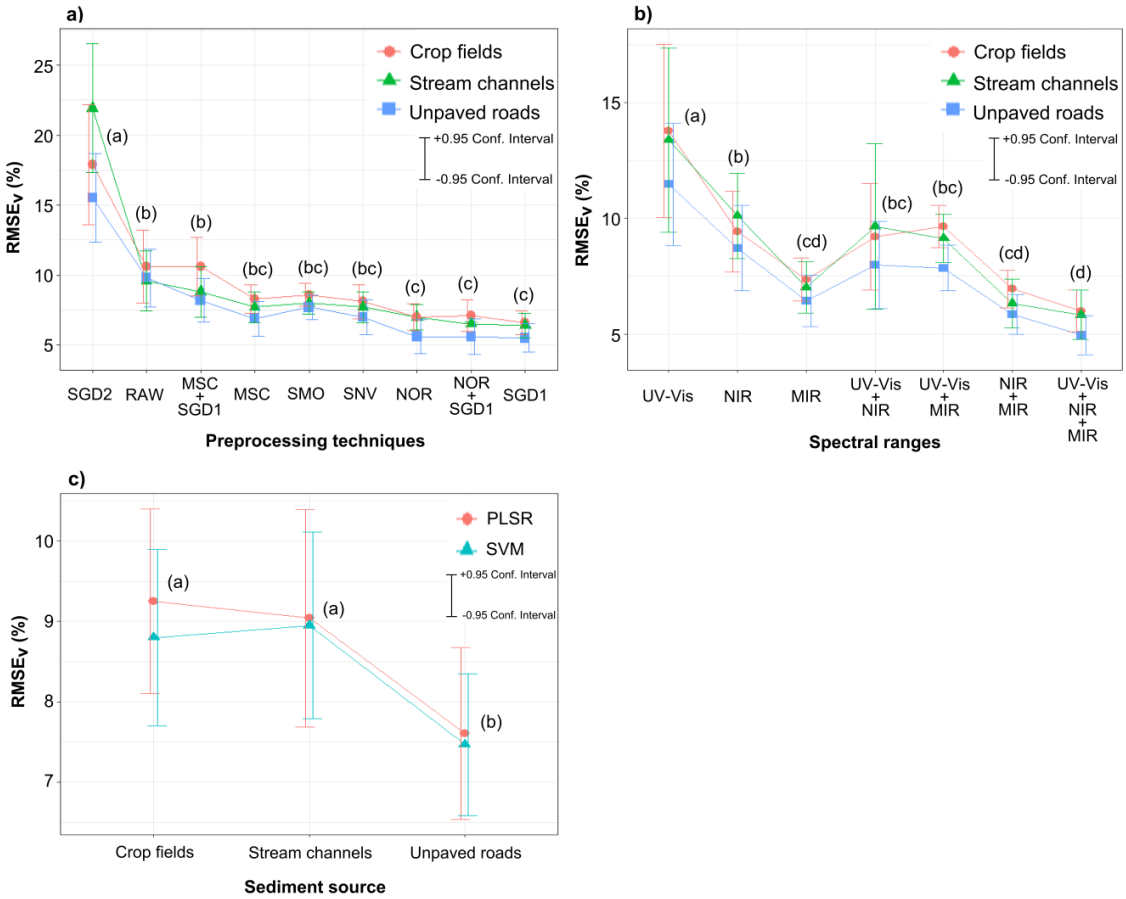


Figure 5.

Performance in validating Partial Last Square Regression (PLSR) and Support Vector Machine (SVM) prediction models from raw spectral data and combined with the eight spectral pre-processing techniques for the sediment sources including stream channels (SC), unpaved roads (UR) and surface of crop fields (CF). RAW - raw spectral; SMO - smoothing; SGD1 - Savitzky-Golay with 1st derivative; SGD2 - Savitzky-Golay with 2nd derivative; SNV - varied standard deviation correction; MSC - multiplicative scatter correction; NOR - normalization by standard deviation.



28

29 **Figure 6.**

30 Mean values of the $RMSE_v$ error statistic of the validation of the prediction models of the three
31 sediment sources in relation to (a) eight spectral pre-processing techniques; (b) spectral bands
32 and their combinations; (c) Partial Last Square Regression - PLSR and Support Vector Machine -
33 SVM. Means followed by the same letter do not differ by the Tukey's test at $p<0.05$.

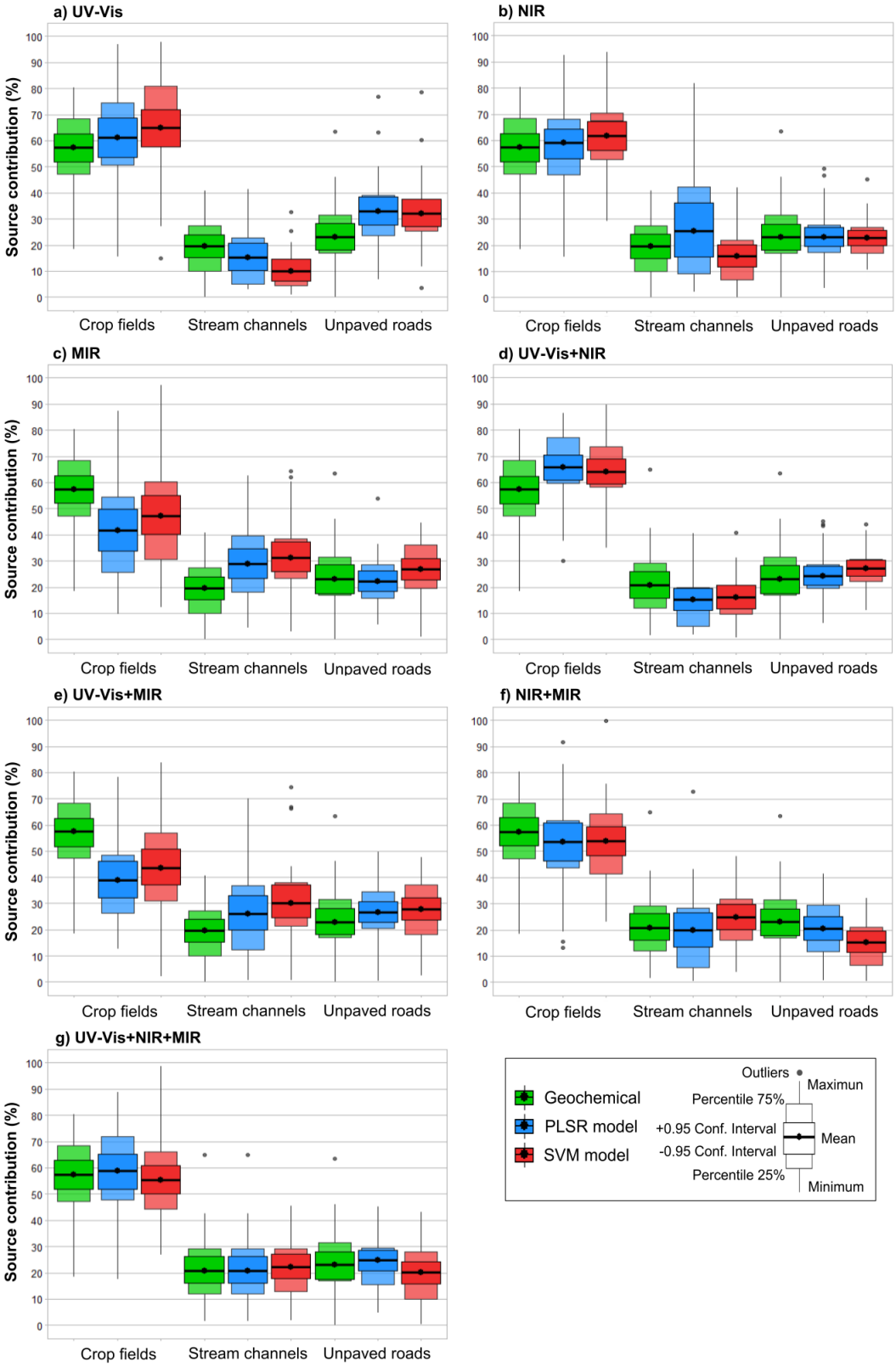


Figure 7.

Boxplot of the contribution of sediment sources estimated by the different approaches for the 29 sediment samples. The estimates are derived from the models that presented the highest accuracy among all processing and spectral range combinations for the PLSR and SVM methods (UV-Vis = SGD1, NIR = NOR+SGD1, MIR = SGD1, UV-Vis+NIR = NOR+SGD1, UV-Vis+MIR = SGD1, NIR+MIR = NOR+SGD1, UV-Vis+NIR+MIR = SGD1).

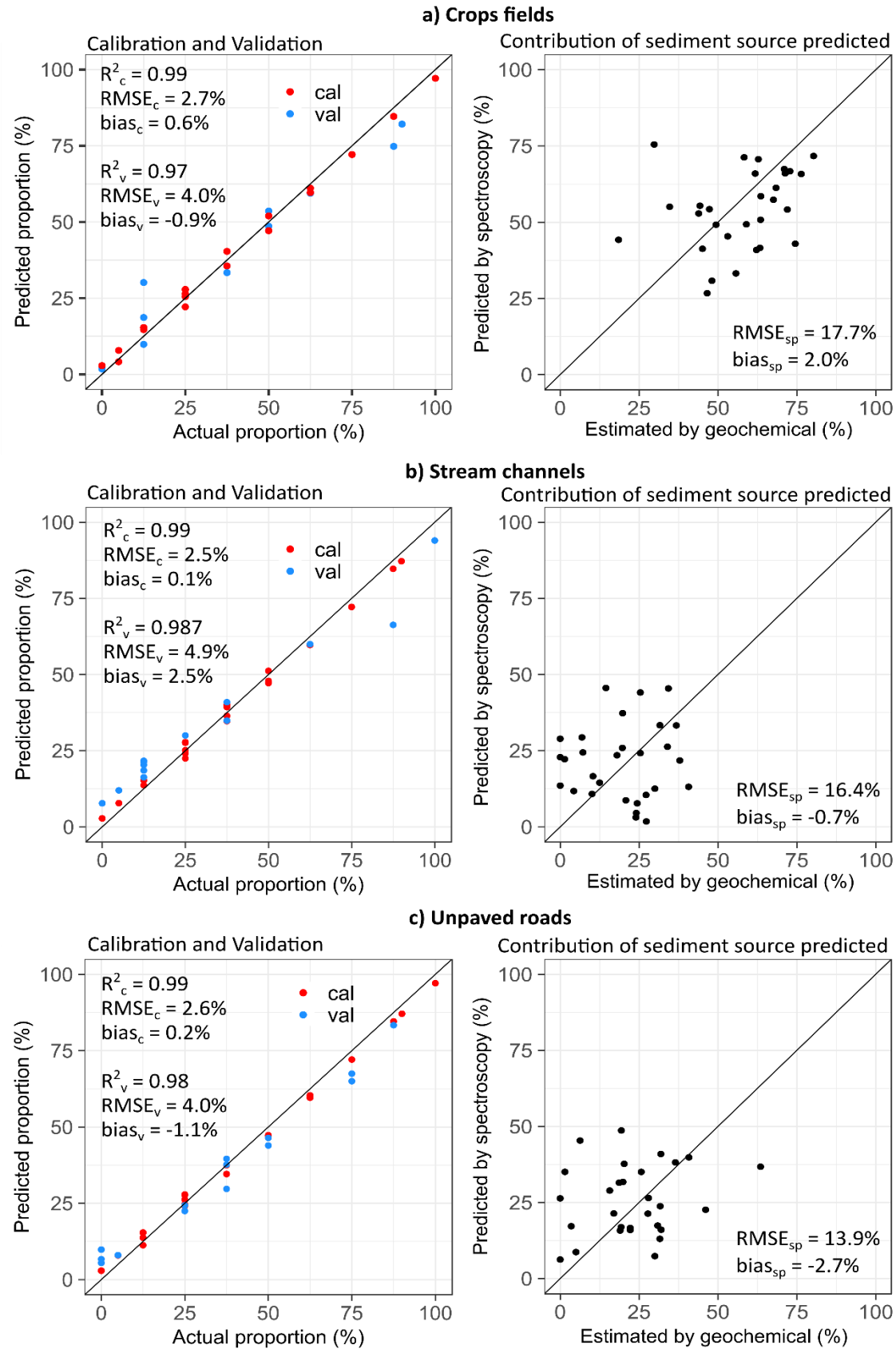


Figure 8.

Scatter plot of the validation and calibration of the best model (SVM-SGD1-UV-Vis + NIR + MIR) for mixtures the three sediment sources and the predicted contribution values by the spectra and observed by the geochemical method.

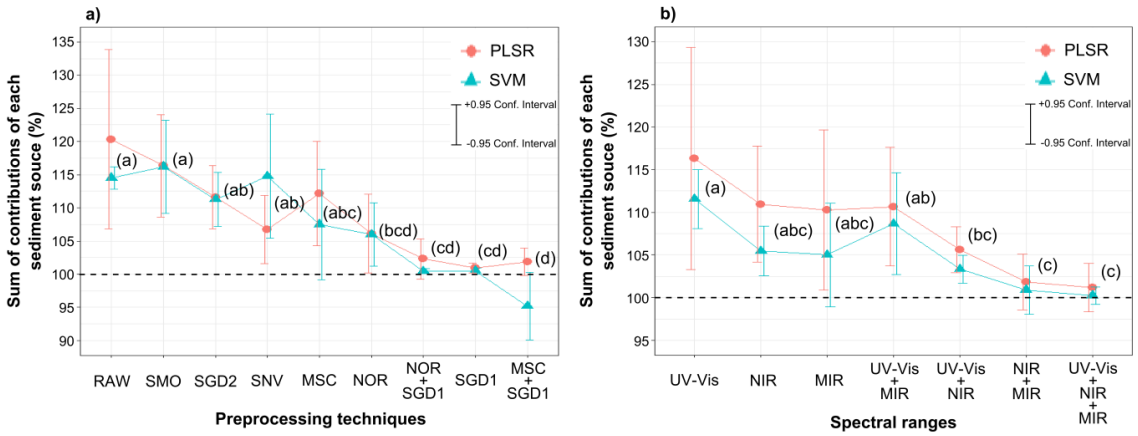


Figure 9.

Sum of sediment source contributions comparing Partial Last Square Regression - PLSR and Support Vector Machine -SVM multivariate methods for each pre-processing technique (a) and for each spectral range and their respective combinations (b). Means followed by the same letter do not differ according to the Tukey's test at $p<0.05$. The dotted line represents 100%.

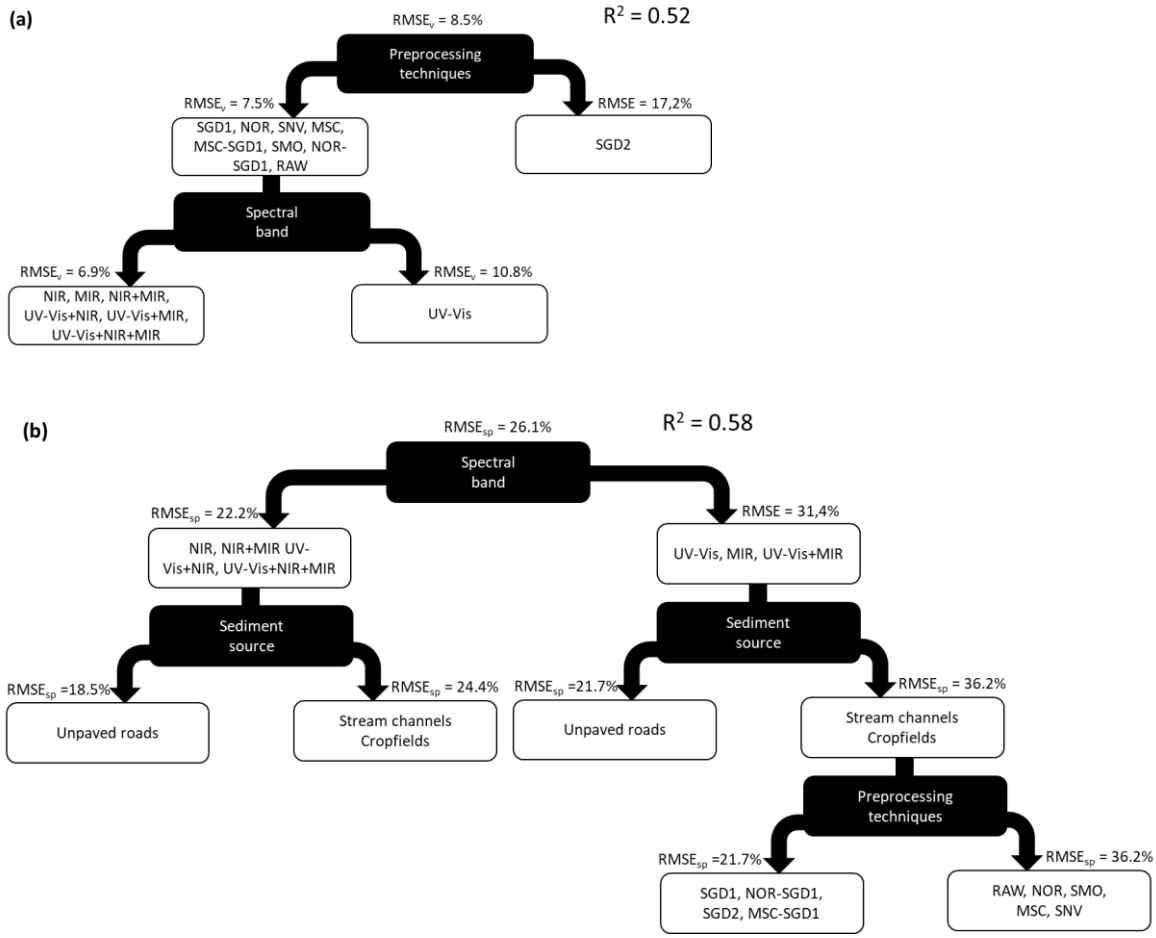


Figure 10.

Conditional inference tree analysis evaluating the factors that most affect the quality of the models based on validation with artificial mixtures of sediment (RMSEv - a), and compared with the sediment contribution values obtained with geochemical tracers (RMSEsp -b).