Application of Mid-Infrared Spectroscopy to the prediction and specification of pesticide sorption: a promising and cost-effective tool

Jeanne Dollinger[1], Jeanne-Chantal Thoisy[2], Cécile Gomez[1], Anatja Samouelian[1]*

(1): UMR LISAH, Université Montpellier, INRAE, IRD, Institut Agro, AgroParisTech, Montpellier 34060, France
(2): UMR ECOSYS, Université Paris-Saclay, INRAE, AgroParisTech, Palaiseau 91120, France.
(*): corresponding author: jeanne.dollinger@inrae.fr

Abstract: The cocktail of pesticides sprayed to protect crops generates a miscellaneous and generalized contamination of water bodies. Sorption, especially on soils, regulates the spreading and persistence of these contaminants. Fine resolution sorption data and knowledge of its drivers are needed to manage this contamination. The aim of this study is to investigate the potential of Mid-Infrared spectroscopy (MIR) to predict and specify the adsorption and desorption of a diversity of pesticides. We constituted a set of 37 soils from French mainland and West Indies covering large ranges of texture, organic carbon, minerals and pH. We measured the adsorption and desorption coefficients of glyphosate, 2,4-dichlorophenoxyacetic acid (2,4-D) and difenoconazole and acquired MIR Lab spectra for these soils. We developed Partial Least Square Regression (PLSR) models for the prediction of the sorption coefficients from the MIR spectra. We further identified the most influencing spectral bands and related these to putative organic and mineral functional groups. The prediction performance of the PLSR models was generally high for the adsorption coefficients Kd_{ads} ($0.4 < R^2 < 0.9$ & RPIQ > 1.8). It was contrasted for the desorption coefficients and related to the magnitude of the desorption hysteresis. The most significant spectral bands in the PLSR differ according to the pesticides indicating contrasted interactions with mineral and organic functional groups. Glyphosate interacts primarily with polar mineral groups (OH) and difenoconazole with hydrophobic organic groups (CH_2, C=C, COO^-, C-O, C-O-C). 2,4-D has both positive and negative interactions with these groups. Finally, this work suggests that MIR combined with PLSR is a promising and cost-effective tool. It allows both the prediction of adsorption and desorption parameters and the specification of these mechanisms for a diversity of pesticides including polar active ingredients.

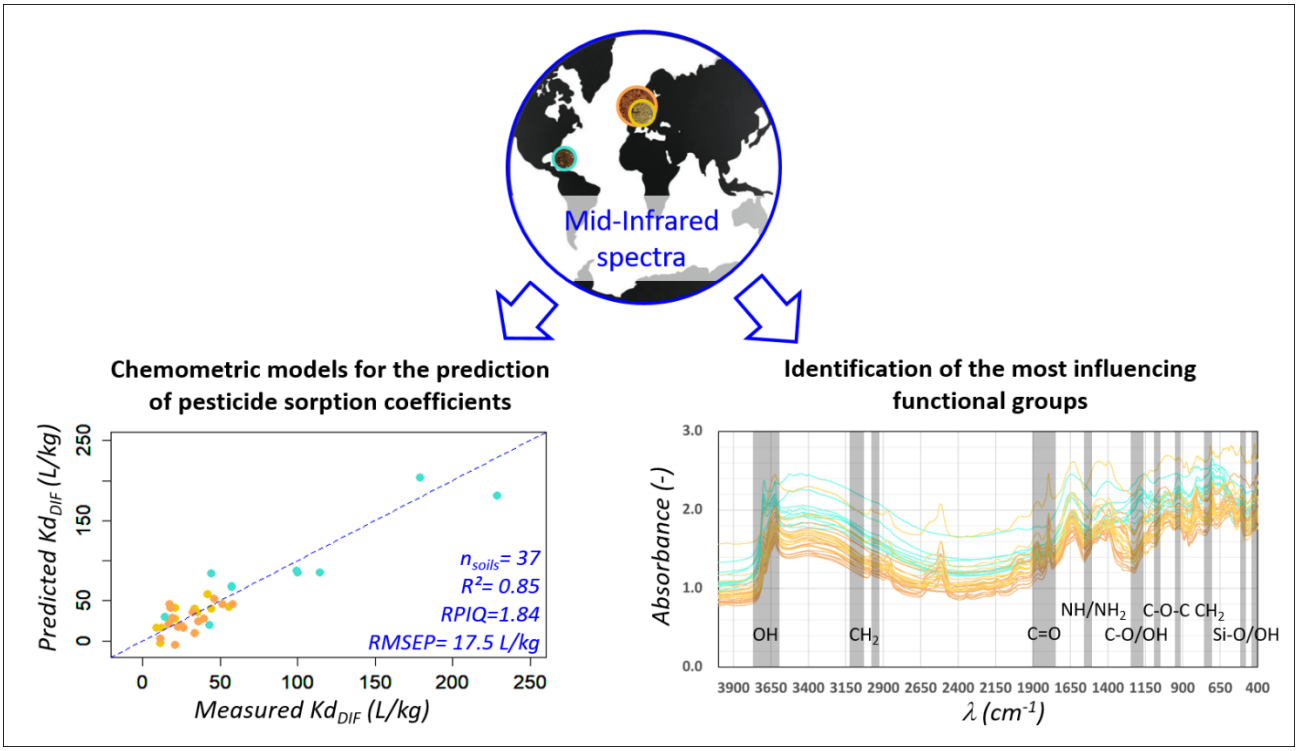
Keywords: FT-IR; MIR; Partial Least Square Regression; Pesticides; Soil; Adsorption; Desorption.

Highlights

- We used MIR spectroscopy combined with PLSR to predict and specify pesticide sorption

- The prediction performance is good for the adsorption coefficients Kd_{ads} (RPIQ > 1.8)
- Low desorption hysteresis is challenging the performance of the PLSR models
- Both mineral and organic groups are involved in the sorption of the three pesticides
- Functional groups influencing sorption coefficients differ for the three pesticides

Graphical abstract:



Introduction

The annual use of about three millions tons of synthetic pesticides for the protection of crops from pests and weeds (De et al., 2014; Sabzevari and Hofman, 2022; Sharma et al., 2019) has generated a globalized contamination of terrestrial and freshwater ecosystems (Malla et al., 2021; Pietrzak et al., 2019; Sabzevari and Hofman, 2022; Sharma et al., 2019; Tang et al., 2021). This contamination is miscellaneous as hundreds of pesticide active ingredients are commercialized worldwide, each having contrasted physico-chemical properties, environmental behavior, toxicity and ecotoxicity (PPDB, 2023; Sabzevari and Hofman, 2022). The pesticide cocktail varies locally with crop type, target pest or weeds, climate and regulations. At the watershed scale, the mix potentially contain tens to over hundred active ingredients (BNV-D, 2022; Sabzevari and Hofman, 2022; Sharma et al., 2019).

The offsite transport of pesticides from croplands to surrounding ecosystems is regulated mainly by sorption mechanisms (Farenhorst, 2006; Kookana et al., 2014; Tang et al., 2012). Sorption also influences their persistence as it modulates their bioavailability to degrading microorganisms (García-Delgado et al., 2020; Kookana et al., 2014). Sorption itself is modulated by soil properties including soil organic carbon (SOC) content and nature, texture, pH, minerals or cation exchange capacity (CEC) with varying influence according to the pesticide physico-chemical properties (García-Delgado et al., 2020; Kah and Brown, 2007; Kookana et al., 2014; Novotny et al., 2020; Weber et al., 2004). Sorption coefficients are potentially very variable both at local and global scales depending on the variation in soil properties (Hermansen et al., 2020; Paradelo et al., 2016; Umali et al., 2012).

The current challenge is to gain insight into the sorption mechanisms of a wide range of pesticides to identify and design suitable mitigation measures while generating fine-resolution sorption data for accurate parametrization of the risk assessment tools (models/indicators). Yet, conventional laboratory methods for measuring sorption coefficients are extremely time-consuming and expensive (Forouzangohar et al., 2009). Therefore, there is a need to develop rapid and cost-effective methodologies for both predicting and specifying sorption mechanisms for a large range of pesticides and soil types (Dagès et al., 2023; Gatel et al., 2019; Singh et al., 2016).

Pedotransfer functions based on SOC, texture and pH or, less frequently, on CEC and metal oxides are the historical approach for predicting and specifying pesticide sorption (Boivin et al., 2005; Dollinger et al., 2015; Kah and Brown, 2007; Kodešová et al., 2011; Weber et al., 2004). The limited number and covariation of the soil properties considered can hinder their ability to predict and specify sorption mechanisms. Recent approaches combining SOC characterization by nuclear magnetic resonance (NMR) or metabolomics and chemometrics provide good predictive performance for a range of pesticides whose sorption depends primarily on SOC content and nature (Dollinger et al., 2023; Kookana et al., 2014). However, they do not account for the influence of mineral constituents particularly known to influence the sorption of polar pesticides (Kah and Brown, 2007). These approaches are also rather time consuming and expensive (Dollinger et al., 2015; Kookana et al., 2014).

Infrared spectroscopy combined with chemometrics has been featured as a rapid and cost-effective alternative to traditional laboratory methods for the estimation of numerous soil properties (Barra et al., 2021; Ng et al., 2022; Seybold et al., 2019). Both the Visible-Near-Infrared (Vis-NIR, 400 – 2500 nm) and Mid-Infrared (MIR, 4000 – 400 cm^{-1}) spectral ranges were successfully used for the estimation of pesticide adsorption coefficients (Ding et al., 2011; Forouzangohar et al., 2009; Hermansen et al., 2020; Paradelo et al., 2016; Parolo et al., 2017; Shan et al., 2020; Singh et al., 2016; Umali et al., 2012). MIR generally outperform NIR for the prediction of numerous soil properties including pesticide adsorption

coefficients (Forouzangohar et al., 2009; Ng et al., 2022; Seybold et al., 2019). However, MIR was applied only for two hydrophobic pesticides, diuron and chlorpyrifos (Forouzangohar et al., 2009; Parolo et al., 2017; Umali et al., 2012). In addition, there was, to our knowledge, no attempt to estimate the pesticide desorption coefficients neither from NIR nor from the MIR spectral domain.

MIR spectroscopy is able to discriminate various mineral and organic functional groups potentially involved in the pesticide sorption mechanisms (Forouzangohar et al., 2009; Ng et al., 2022; Parolo et al., 2017). Indeed, absorption in the MIR region results from fundamental vibration of minerals and organic functional groups, whereas the Vis-NIR region is dominated by broad and overlapping peaks from overtones and combination of these fundamental vibrations (Lohumi et al., 2015; Paradelo et al., 2016; Seybold et al., 2019). These surface functional groups are active binding sites for a range of contaminants including pesticides with influence varying according to the structure and elemental composition of the pesticides (García-Delgado et al., 2020; Khalid et al., 2020; Novotny et al., 2020).

Given the diversity of mineral and organic components of soil having a specific signal in the MIR region, we hypothesized that the estimation of sorption coefficients from MIR spectroscopy combined with chemometrics can be extended to a larger range of pesticides including polar and hydrophobic active ingredients. We also hypothesized that the specificity of the signal from these functional groups in the MIR region is a precious and under exploited opportunity to specify pesticide adsorption and desorption mechanisms. Accordingly, the aims of the study are 1) to evaluate the performance of MIR spectroscopy combined with partial least square regression (PLSR) for the prediction of soil adsorption and desorption coefficients of a range of polar to hydrophobic pesticides and 2) to identify functional groups involved in the sorption mechanisms of these pesticides.

Three pesticides among the most used worldwide for a variety of crops, including cereals, orchards or vineyards (Matich et al., 2019; Sharma et al., 2019) and covering an extended range of hydrophobicity, solubility and surface charges were selected for this study. These are: glyphosate, a hydrophilic broad-spectrum post-emergence herbicide; 2,4-dichlorophenoxyacetic acid (2,4-D), a hydrophilic selective post-emergence herbicide and difenoconazole, a hydrophobic systemic fungicide.

2. Material and Methods

2.1. Chemicals

The three selected pesticides cover an extended range of physico-chemical properties as described here after. Glyphosate has a very high aqueous solubility (100 g/L), low hydrophobicity (logP -6.28) and is a zwitterion under pH 10.2 (PPDB, 2023). 2,4-D also has a very high aqueous solubility (24 g/L), low

hydrophobicity (logP -0.82) but is negatively charged under environmental pH ranges (PPDB, 2023). Difenoconazole has a low aqueous solubility (15 mg/L), high hydrophobicity (logP 4.36) and is uncharged under environmental pH ranges (PPDB, 2023).

Non-labeled glyphosate, 2,4-D and difenoconazole were supplied by Merck and 14C-labeled pesticides by ISOBIO (Fleurus, Belgium). Merck supplied sodium azide (NaN_3) and calcium chloride ($CaCl_2$). All the chemicals used were HPLC grade.

2.2. Soil sampling and characterization

We constituted a set of 37 soil samples, from three locations in the French mainland and overseas territories, with the purpose of covering a wide range of soil types with contrasted physico-chemical properties. Ten soils were collected on the Basse-Terre Island, Guadeloupe, French West Indies (WI). These soils belong to a tropical toposequence of volcanic ash soils and are differentiated according to the age of the volcanic deposit. The other soils were collected in two vineyard catchments in southern France characterized by a Mediterranean climate, the Roujan and Rieutor watersheds (FR-RO and FR-RI, respectively). These are only a few kilometers apart but characterized by contrasted soils due to variations of underground rocks and pedogenesis processes. Some of the FR-RO soils were sampled in un-cropped areas of the site such as fallows, hedgerows, grass strips or ditches to diversify the type and content of organic carbon. Eleven and 16 samples were collected from FR-RO and FR-RI, respectively.

The texture, organic carbon content (OC), pH_{H2O}, cationic exchange capacity (CEC) and calcium carbonates ($CaCO_3$) were measured at the INRAE LAS laboratory (Arras, France) for both FR-RO and FR-RI soils and at the Cirad US 49 laboratory (Montpellier, France) for WI soils (specific habilitation for analyzing overseas soils). The standardized methods used were the same at both laboratories. These were NF ISO 11277, NF ISO 14235, NF ISO 23470, NF ISO 10693 and NF ISO 10390 for the texture, OC, CEC, $CaCO_3$ and pH, respectively. These properties are displayed in Figure 1.

2.3. Measurement of the sorption coefficients

Both adsorption and desorption isotherms were characterized for the 37 soils. The adsorption batch test procedure was designed following the OECD guidelines n°106 (OECD, 2000) and is described in details in Dollinger et al. (2023). Briefly, 14C-labelled glyphosate, 2,4-D and difenoconazole were used for the experiments. The concentration of the solutions used were 5, 10, 50, 100 and 1000 µg/L. The background electrolyte was composed of 200 mg/L NaN_3 for glyphosate and 200 mg/L NaN_3 plus 0.01M

153 $CaCl_2$ for the other pesticides. The solid-to-liquid ratio for all batches was 1:10 (g/mL). Soils were
154 equilibrated for 24h with the pesticide in glass tubes at a shaking speed of 150 rpm. The tubes were
155 then centrifuged at 3000 rpm (1770 g) for 10 min, and the supernatant was sampled and analyzed by
156 liquid scintillation counting (LSC). Following the adsorption phase, a five-step desorption was
157 performed as detailed in Dollinger et al., (2023). The adsorption and desorption batches were all
158 conducted in triplicates.

159 Both linear (Equation 1) and Freundlich (Equation 2) models were fitted to the adsorption isotherms.
160 Given the excellent linearity of the adsorption isotherms ($0.91<n_{ads}<1.01$), only the linear adsorption
161 coefficients Kd_{ads} are used for the rest of the study. However, the desorption isotherms are non-linear.
162 Therefore, the Freundlich Kf_{des} and n_{des} coefficients are used. The adsorption being linear, n_{des} provides
163 an estimation of the desorption hysteresis that is considered significant when $H < 0.70$ ($H = n_{des}/n_{ads}$)
164 (Barriuso et al., 1994).

165 $$Cs = Kd * Caq \quad \text{(Equation 1)}$$

166 $$Cs = Kf * Caq^{n} \quad \text{(Equation 2)}$$

167 Where Caq is the concentration in the aqueous phase at equilibrium (µg/L), Kd the linear sorption
168 coefficient (L/Kg), K_f ([µg/kg]/[µg/L]^n) and n (-) are the Freundlich coefficients and Cs the concentration
169 in the soil (µg/Kg). For Kd, Kf and n, the subscript "ads" is used when the models (Equation 1 or 2) are
170 fitted to an adsorption isotherm and "des" when the models are fitted to a desorption isotherm.

171

172 **2.4. Acquisition and preprocessing of the MIR spectra**

173 The soils were ground and sieved at 200 µm. Prior to the acquisition of the MIR (mid-infrared) spectra,
174 the soils were dried for four days at 35°C. All the soil spectra were scanned by a Nicolet Is10 (Thermo
175 Scientific) spectrophotometer equipped with a DRIFT (Diffuse reflectance mid-infrared Fourier
176 transform) accessory. The MIR spectra were recorded as the mean of 32 scans from 4000 to 400 cm^{-1}
177 and the spectral resolution is 1 cm^{-1}. Absorbance spectra are obtained using equation 3:

178 $$Absorbance = \log(1/Reflectance) \quad \text{(Equation 3)}$$

179 Two of the most popular spectral pre-treatments, Savitsky-Golay (STG) and Standard Normal Variate
180 (SNV) were tested individually (Barra et al., 2021; Ng et al., 2022; Shan et al., 2020). The "signal" (signal
181 developers, 2013) and "prospectr" (Stevens and Ramirez-Lopez, 2022) packages of the R software (R
182 Core Team, 2023) were used to apply the STG and SNV pre-treatments, respectively.

183

2.5. Chemometric prediction and specification of the pesticide sorption coefficients

2.5.1 Partial Least Square Regression models

Partial least square regression (PLSR) was performed with the pls package (Liland et al., 2022) of R software (R Core Team, 2023) to establish predictive models for Kd_{ads}, Kf_{des} and n_{des} of the three pesticides. The optimal number of latent variables (LV) was determined for each predictive model using the leave-one-out cross-validation (LOO CV) method. This LV number is considered optimal when the Root Mean Squared Error in cross-validation ($RMSE_{CV}$) is the lowest. The maximum number of latent variables was set to 20.

The raw, the SNV- and STG-treated spectra were used as predictive variables, respectively. We compared their performance for all of the sorption coefficients. The R^2, $RMSE_{CV}$ from leave-one-out cross validation and the ratio of performance to interquartile distance (RPIQ) were used to evaluate the performance of the PLSR models. RPIQ was used instead of ratio of performance to deviation (RPD) as, except for Kd_{ads}(GLY), the distributions of the sorption coefficients are non-normal (P-values <0.01 for Shapiro tests) (Bellon-Maurel et al., 2010). In addition, we compared the error of the spectral prediction to the uncertainty (sd) of the Kd_{ads} measure using a Prediction Accuracy index PAi as defined in Equation 4. For PAi ≤ 1, the error of spectral prediction (predicted Kd_{ads} – measured Kd_{ads}) is higher than the sd of the Kd_{ads} measure calculated from the batch triplicates (section 2.3). The sd of the Kd_{ads} measure represents 8-10 % of the average Kd_{ads}. PAi ≤ 1 therefore indicates that the PLSR model is very accurate.

$$PAi = abs[Kd_{ads}(predicted)-Kd_{ads}(measured)] / sd(Kd_{ads}(measured)) \quad \text{(Equation 4)}$$

Where PAi is the accuracy criteria (-), Kd_{ads}(predicted) the predicted adsorption coefficient (L/Kg), Kd_{ads}(measured) is the measured adsorption coefficient (L/Kg) and sd(Kd_{ads}(measured) the standard deviation of the measured Kd_{ads} (L/Kg).

For 2,4-D the adsorption on the FR-RO and FR-RI soils was very weak and the desorption elevate. Therefore, accurate measurement of Kf_{des} and n_{des} for these soils was not possible or very uncertain. Given the limited number of accurate calibration values of Kf_{des}(2,4-D) and n_{des}(2,4-D), we chose not to build PLSR models.

2.5.2 Significant wavelengths identification and assignment

In order to gain insight into the sorption mechanisms, the most significant wavelengths were identified from each PLSR model. A hybrid wavelength point selection method combining Variable Importance in the Projection (VIP) and regression coefficients (RC) was implemented (Fu et al., 2022; Lohumi et al.,

2015; Wang et al., 2022). For each PLSR model, the RC used for the significant wavelength identification step were those from the last factor (LV), that integrates all the spectral features used in the regression (Forouzangohar et al., 2009). The threshold value was set to 1 for VIP and to the standard deviation (sd) value for RC. All wavelengths with VIP> 1 & RC>sd were considered significant.

A review of the literature was conducted to relate the significant wavelengths to functional groups and to their putative origin (Table 1).

3. Results

3.1. Soil properties and pesticide sorption

The set of soils selected for this study covers most of the texture classes from the USDA textural classification (Fig. 1a) and an extensive range of SOC (0.46–6.50%), pH_{H2O} (4.63–8.68) and CEC (5.99–48.50 cmol/kg) (Fig. 1b, c & d). The WI soils are characterized by elevate SOC, clay and CEC as well as low pH (Fig. 1 & Fig. S1, supplementary material). The FR-RO and FR-RI are distinct from the WI soils and more diverse. They have in general higher pH values and coarser textures but contrasted calcium carbonate and SOC contents (Fig. 1 & Fig. S1).

The absorption in the MIR region (4000–400 cm^{-1}) gives further indications about the relative proportion of diverse functional groups that vibrate at specific wavelengths (Table 1). The clustering of the MIR spectra by PCA also highlights the differences between the WI and FR-RO/FR-RI soils and the greatest diversity of FR-RO/FR-RI soils compared to WI soils (Fig. S2). The shape of the MIR spectra differs among the soils especially for the 3700–3500 cm^{-1} (clay minerals), 3000–2800 cm^{-1} (hydrophobic SOM), 2640–2440 cm^{-1} (hydrophobic SOM and carbonates), 2300–1760 cm^{-1} (hydrophilic SOM) and 1560–400 cm^{-1} (hydrophobic and hydrophilic SOM with mineral overtones) spectral bands (Fig. 8 & Table 1). This indicates that both the mineral and the organic fractions of these soils are contrasted.

Figure 2 displays the correlation matrix between soil physico-chemical properties and pesticide sorption coefficients. It shows that SOC significantly (p-value < 0.05) and strongly correlates with all of the adsorption and desorption coefficients except Kd_{ads}(GLY) and n_{des}(2,4-D). CEC exhibits equivalent correlations with the sorption coefficients than SOC, which is not surprising given their strong covariation. pH also correlates with all sorption coefficients except n_{des}(2,4-D). It is the lone of these four soil properties that correlate with Kd_{ads}(GLY) for this set of soils. Clay content correlate with the

Kf_{des} of the three pesticides and with Kd_{ads}(2,4-D) and n_{des}(DIF). n_{des}(2,4-D) doesn't correlate with any of these soil properties.

In accordance with the ranges of soil properties and their relative influence on the sorption of the selected pesticides, the measured adsorption and desorption coefficients cover several orders of magnitude (Fig. 3). The sorption behavior is also contrasted among the three pesticides. Glyphosate has a moderate to high adsorption (Kd_{ads} 3.2–28.8 L/kg) and a very strong desorption hysteresis (Kf_{des} 263–4844 ([µg/kg]/[µg/L]^n) & n_{des} 0.04–0.25). Difenoconazole has a very high adsorption (Kd_{ads} 8.5–228.5 L/kg) and a strong desorption hysteresis (Kf_{des} 140–4116 ([µg/kg]/[µg/L]^n) & n_{des} 0.03–0.65). Last, exception made for the WI soils (Kd_{ads} 1.5–7.1 L/kg, Kf_{des} 189–624 ([µg/kg]/[µg/L]^n) & n_{des} 0.11–0.40), 2,4-D is weakly adsorbed (Kd_{ads} 0.02–0.75 L/kg) and has moderate to no desorption hysteresis (Kf_{des} 0–6 ([µg/kg]/[µg/L]^n) & n_{des} 0.03–1.55).

3.2. Performances of MIR-PLSR models for pesticide sorption coefficients prediction

Figure 6 displays the scatter plots of the Kd_{ads} measured *versus* predicted by PLSR for glyphosate, 2,4-D and difenoconazole. Figure 7 shows the scatter plots of the desorption coefficients Kf_{des} and n_{des} measured *versus* predicted by PLSR for glyphosate and difenoconazole. The performance criteria (R2, RMSE, RPIQ) and the number of latent variable (LV) are also display on Figures 6 and 7. Given the high resolution of the raw spectra, the STG pre-treatment didn't significantly improve the prediction accuracy of the PLSR for any of the predicted coefficient. SNV outperformed the raw signal only for Kd_{ads}(GLY), Kf_{des}(DIF) and n_{des}(DIF). Therefore, the predictive models were established using the raw spectra except for Kd_{ads}(GLY), Kf_{des}(DIF) and n_{des}(DIF) that were based on the SNV-treated signals. The number of latent variables varied from 4 to 10 for Kd_{ads}, from 3 to 8 for Kf_{des} and from 2 to 7 for n_{des}.

The predictive performance, featured by the R^2 (-), RPIQ (-) and $RSME_{CV}$ (L/kg) values, varies across the range of coefficients and pesticides considered. For the adsorption coefficients Kd_{ads}, the goodness of fit, featured by the RPIQ values, is good and equivalent for the three pesticides. The R^2 of glyphosate is half that of 2,4-D or difenoconazole. However, the PAi, that compares the error of prediction to the standard deviation of the Kd_{ads} measure, is lower for glyphosate than for 2,4-D and difenoconazole. This is due to higher sd of the Kd_{ads} measures for glyphosate than for 2,4-D and difenoconazole. Furthermore, despite similar R^2 and RPIQ values for 2,4-D and difenoconazole, the predictive performance is lower for 2,4-D. Indeed, the PAi is high and about ten time higher for the FR-RO & FR-RI soils compared to the WI soils for 2,4-D (Fig. 6).

For the desorption coefficients, Kf_{des} & n_{des}, the predictive performance is good for difenoconazole and fair for glyphosate (Fig. 7). The performance of the PLSR for the prediction of n_{des} is generally weaker than that of Kf_{des}. This, along with the poor correlation of n_{des} with SOC, texture, CEC or pH (Fig. 2), suggests that other characteristics of the systems (e.g. solid-liquid ratio, amount of pesticide in the system etc.) influence this coefficient (Dollinger et al., 2015; Wauchope et al., 2002).

3.3. Functional groups involved in the pesticide sorption mechanisms

The wavelength selection method (section 2.5.2) identifies the most significant spectral bands in the predictive PLSR models. The significant spectral bands for the prediction of Kd_{ads} are displayed in Figure 4. The significant spectral bands for the prediction of Kf_{des} and n_{des} are displayed in Figure 5 and Figure S3 (supplementary material), respectively. Functional groups vibrating at these wavelengths are deemed to play a prime order role in the adsorption or desorption of the tested pesticide. For most of the significant spectral bands, putative functional groups could be assigned (Table 1). The RC value gives further indication about the nature of the interaction with the functional groups. Some functional groups are positively correlated to the sorption parameters, which suggests that these are active biding sites. However, other are negatively correlated indicating a repulsive tendency. Figures 8 synthesizes the significant spectral bands for all predicted sorption coefficients as well as the nature of the correlation (positive or negative) and the assigned putative functional groups.

It is interesting to note that, for the three tested pesticides, the functional groups interfering in the adsorption and desorption mechanisms differ. However, for all of the three pesticides, both organic and mineral functional groups are involved (Fig. 8 & Table 1).

The adsorption of glyphosate is positively correlated to OH groups from kaolinite and aluminum oxides as well as organic NH/NH_2 and C=C groups and negatively to COO- and OH groups from goethite or phenols (Fig. 8 & Table 1). As a contrast, the adsorption of difenoconazole is negatively correlated to OH groups from kaolinite & aluminum oxides and C=O from hydrophilic soil organic matter (SOM) and positively correlated to CH_2, C=C, COO-, C-O, C-O-C and OH from hydrophobic SOM (Fig. 8). For 2,4-D, both attractive and repulsive interactions are evidenced with the mineral (kaolinite/gibbsite) OH and CO_3^{2-} functional groups. Kd_{ads} from 2,4-D is also positively correlated to, OH and C-O from hydrophilic SOM and negatively to C=O, C=C, NH/NH_2, COO-.

Some of the significant wavelengths are common to the Kd_{ads} and Kf_{des} PLSR models (Fig. 8). When these wavelengths are positively correlated to both the Kd_{ads} and Kf_{des} in the PLSR models, it could indicate a strong and faintly reversible bond with the corresponding functional group (Fig. 8 & Table 1). It is the case for example for glyphosate with the mineral OH group and for difenoconazole with

311 the C-O and organic OH groups. In contrast, some of the wavelengths are negatively correlated to both
312 the Kd_{ads} and Kf_{des} in the PLSR models, indicating weak bonds or repulsive influence. This includes the
313 COO- group for glyphosate; mineral OH, Si-O and the C=O groups for difenconazole. Other functional
314 groups are positively correlate to the Kd_{ads} and negatively to the Kf_{des} suggesting strong but reversible
315 bonds. The desorption hysteresis appears to be influenced both by the mineral fraction of soil and by
316 the hydrophilic fraction of SOM (Fig. 8). There are indeed few significant wavelengths for n_{des} in the
317 3500 – 2000 cm^{-1} range and many in the 3700 - 3550 and 1900 - 400 cm^{-1}.

318

319 **4. Discussion**

320

321 **4.1. Extent of the sorption coefficient datasets**

322 The extensive ranges of measured Kd_{ads} (Fig. 3) are in accordance with the great variability of the soil
323 physico-chemical properties (Fig. 1) and their reported influence on the adsorption of these pesticides
324 (Dollinger et al., 2015; Wang et al., 2020; Weber et al., 2004; Werner et al., 2013). The measured Kd_{ads}
325 values of 2,4-D and difenoconazole exceed the ranges of Kd_{ads} reported in the literature (0.3-1.9 L/kg
326 for 2,4-D & 2-99 L/kg for DIF) (Godeau et al., 2021; PPDB, 2023; Wang et al., 2020; Werner et al., 2013).
327 The Kd_{ads} values of glyphosate are in the low-medium range of values reported in the literature (0.8-
328 510 L/kg) (Dollinger et al., 2015; Gurson et al., 2019; Hermansen et al., 2020; Paradelo et al., 2016).
329 Higher Kd values reported for glyphosate were measured with $CaCl_2$ as background electrolyte. We
330 chose not to add $CaCl_2$ to the glyphosate solutions as it significantly and artificially increases the Kd_{ads}
331 values (Cruz et al., 2007; de Jonge and Wollesen de Jonge, 1999; Dollinger et al., 2015).

332 Few information about the ranges and drivers of pesticide desorption is available in the literature. This
333 is partly due to the elevate time and costs needed to acquire these parameters. The desorption
334 hysteresis is actually not represented in the risk assessment tools. However, even if these coefficients
335 can't be easily implemented in the risk assessment tools, they help assess the uncertainty of their
336 outputs and the long-term efficiency of the mitigation measures. Indeed, desorption controls the
337 remobilization of pesticides after spraying during the successive runoff events. If the adsorption is
338 significant (Kd_{ads}>1 L/kg) and the desorption hysteretic (H<0.7), the models/indicators overestimate
339 the risk of dispersion.

340

341 **4.2. Performance of the MIR-PLSR approach for predicting pesticide sorption**

342 The extended sorption coefficient ranges are ideal for testing the global performance of the MIR-PLSR
343 approach. Although the dataset used to calibrate the PLSR models is limited compared to applications
344 of Infrared spectroscopy for the prediction of primary soil properties (SOC, texture etc.) (Barra et al.,
345 2021; Ng et al., 2022; Seybold et al., 2019), it is quite elevate in the case of pesticide sorption
346 coefficients (Bengtsson et al., 2007; Paradelo et al., 2016; Parolo et al., 2017; Umali et al., 2012).

347 For Kd_{ads}, the performance of the PLSR is good for the three pesticides (Fig. 6). However, the differences
348 between the WI and FR-RO/FR-RI soils (Fig. 1, Fig S1 & S2) induce a bimodal distribution of Kd_{ads}(2,4-
349 D), and, to a minor extent, of Kd_{ads}(DIF). This challenges the prediction performance of Kd_{ads}(2,4-D) for
350 the FR-RO/FR-RI soils. However, it has limited influence for the prediction of Kd_{ads}(DIF). Contrastingly,
351 the distribution of the Kd_{ads}(GLY) is normal. While the calibration of the approach with this set of very
352 contrasted soils provides good estimations of the Kd_{ads}, site-specific calibration could modulate the
353 prediction accuracy depending on the local variability of soil properties. This is evidenced by two
354 studies estimating glyphosate Kd_{ads} with NIR-PLSR, one at the scale of New Zealand (Hermansen et al.,
355 2020) and the second at the field scale (Paradelo et al., 2016). The $RMSE_{CV}$ and CV at the field scale
356 (Paradelo et al., 2016) were about two times lower than at the country scale (Hermansen et al., 2020).

357 For the desorption coefficients, Kf_{des} & n_{des}, the performance of the PLSR models seems related to the
358 magnitude of the desorption hysteresis. The goodness of fit is good for difenoconazole and fair for
359 glyphosate (Fig. 7). The low adsorption of 2,4-D (Kd_{ads} <1L/kg) and its low to null desorption hysteresis
360 on the FR-RO/FR-RI soils challenged the measure of Kf_{des}(2,4-D) & n_{des}(2,4-D) for these 27 soils. Given
361 the poor accuracy of the Kf_{des}(2,4-D) & n_{des}(2,4-D) measures for these soils, the dataset was too
362 restricted to establish PLSR models. Last, n_{des} seems to be less influenced by the range of functional
363 groups captured by the MIR spectroscopy.

364 For the three pesticides, the coefficient of variation of the Kd_{ads} measures is 7-8% (calculated from the
365 batch replicates (see section 2.3). For the average Kd_{ads} values, it represents a disparity of 1.4 L/kg for
366 glyphosate, 3.8 L/kg for difenoconazole and 0.09 L/kg for 2,4-D. The $RMSE_{CV}$ of the MIR-PLSR models
367 are 3 to 8 times higher than these experimental uncertainties but are quite low compared to the Kd_{ads}
368 ranges (Fig. 3). PAi further evaluates this difference between the prediction and measure accuracies
369 for each parameter value (Fig. 6). PAi increase from glyphosate<difenoconazole<2,4-D. It is generally
370 lower for the highest Kd_{ads} values of the distributions.

371 The prediction uncertainties of MIRS-PLSR are also lower than those of traditional estimation methods
372 such as the Koc (PPDB, 2023; Wauchope et al., 2002) and pedotransfer functions (Boivin et al., 2005;
373 Dollinger et al., 2015; Weber et al., 2004). It is also lower than PLSR combined with metabolomics that
374 was tested on the same set of soils in a companion study (Dollinger et al., 2023). For polar pesticides,

and especially for glyphosate, higher R^2 were reported for NIR-PLSR (Hermansen et al., 2020; Paradelo et al., 2016). Indeed, polar pesticides are primarily influenced by mineral constituents of soils (Dollinger et al., 2015; Kah and Brown, 2007) that have strong signals in the Vis-NIR range (Hermansen et al., 2020; Paradelo et al., 2016). However, the coefficients of variation (CV) calculated by dividing the $RMSE_{CV}$ by the mean Kd_{ads}, is equivalent to that of our study (Hermansen et al., 2020).

4.3. Potential of the MIR-PLSR approach for specifying sorption mechanisms

In addition to its good performance for predicting the sorption coefficients of contrasted pesticides, the MIR-PLSR approach also provides useful information to specify the underlying sorption mechanisms. Indeed, the laboratory characterization of soils constituents (SOC, texture, pH, CEC, metal oxides, clay minerals etc.) to identify the drivers of sorption mechanisms, by establishing correlation with the Kd_{ads}, is extremely time consuming and expensive (Boivin et al., 2005; Kah and Brown, 2007; Ng et al., 2022; Seybold et al., 2019; Weber et al., 2004; Werner et al., 2013). Moreover, as displayed in Figure 2, correlation among these soil constituents can mask their relative influence on pesticide sorption. This pedotransfer function approach also relies on the diversity of the soil constituents measured.

Therefore, approaches predicting sorption coefficients based on indirect characterization of soil constituents constitute interesting alternatives for the specification of these mechanisms. Approaches combining chemometrics with NMR or metabolomics provide, but are restricted to, detailed information about the influence of the amount and nature of SOC (Dollinger et al., 2023; García-Delgado et al., 2020; Kookana et al., 2014). In contrast, infrared spectroscopy, especially in the MIR region, provide detailed information about the nature of both mineral and organic fractions and soil properties ensued from these (Ng et al., 2022; Seybold et al., 2019).

The foremost influence of OH groups from kaolinite and iron/aluminum oxides and the poor influence of organic functional groups on the adsorption of glyphosate (Fig. 8, Table 1) are in accordance with the literature. Indeed, many studies evidenced significant correlations between its Kd_{ads} and clay minerals, iron/aluminum oxides, CEC and pH, while the reported influence of SOM is secondary and contrasted (De Gerónimo and Aparicio, 2022; Dollinger et al., 2015; Hermansen et al., 2020). In terms of binding mechanisms, glyphosate has been reported to form strong bonds by ligand exchange on the broken edges of layer silicates, poorly ordered silicates or iron- and aluminum oxides (Borggaard and Gimsing, 2008; Dollinger et al., 2015; Ololade et al., 2014). Other possible binding mechanisms include the formation of complexes between glyphosate and the soil-exchanged polyvalent cations or the formation of HS–Me–glyphosate complexes in which Me is a trivalent or divalent metal cation and HS

are humic substances (Dollinger et al., 2015). The pH influences the charges of both glyphosate and of the soil.

For 2,4-D, the influence of both mineral and organic functional groups (Fig. 4) is in accordance with its reported ability to bind both to mineral and organic constituents of soil with a strong influence of pH (Benoit et al., 1996; Werner et al., 2013). Kd_{ads} measured on isolated soil constituents suggest limited interactions of 2,4-D with quartz, calcite, kaolinite and montmorillonite (Kd_{ads} 0-0.05 L/kg), significant interactions with iron/aluminum oxides (Kd_{ads} 0.4-460 L/kg) and with humic acids (Kd_{ads} 14-60 L/kg) (Werner et al., 2013).

There is no meta-analysis or pedotransfer function approach reported for difenoconazole in the literature. As for other highly hydrophobic pesticides, its sorption is supposed to be influenced mainly by the organic fraction of soils (Weber et al., 2004). Influence of pH has also been reported by Wang et al. (2020) in a very small dataset. Although this reported influence of pH needs further evidences, it could corroborate the negative correlation with the Si-O & OH groups from minerals and C=O from hydrophilic SOM (section 3.3). This is also in accordance with the correlation with pH and SOC displayed in Figure 2.

Amendments with raw or treated organic wastes such as compost, digestate or biochar is a popular mitigation measure for limiting the dispersion of pesticides in croplands (Briceño et al., 2007; Dollinger et al., 2022; García-Delgado et al., 2020; Khalid et al., 2020). The knowledge of the functional groups interfering in the sorption of a range of pesticides could be very useful for evaluating the efficacy of this mitigation measure (García-Delgado et al., 2020). MIR spectra of organic amendments could give indications about their relative potential to retain the contaminants of concern.

Conclusion

MIR-PLSR is a promising tool to predict the adsorption and desorption coefficients of polar and nonpolar pesticides for soils having contrasted physico-chemical properties. The prediction performance is good for the adsorption coefficients of the three pesticides. It is also good for the desorption coefficients of pesticides exhibiting strong desorption hysteresis such as glyphosate and difenoconazole. The establishment of the PLSR models requires a calibration step to integrate to the variability of soil properties from the investigated pedo-climatic contexts. This can be time-consuming depending on the number of targeted pesticides. Yet once this is achieved, a single MIR spectrum can provide estimations for both adsorption and desorption coefficients for the whole range of pesticides tested. Therefore, it is beneficial in terms of risk assessment to diversify the range of pesticides

evaluated and to refine the resolution of the sorption parametrization in the risk assessment tools. The approach was tested for a very diverse set of soils, but its local precision related to agricultural practices and pedomorphologic characteristics of landscapes remains to be evaluated. MIR spectroscopy is very rapid, non-destructive and cost-effective technique. The specificity of the signal for a diversity of mineral and organic functional groups in the MIR region help to gain insight into the sorption mechanisms and the soil constituents involved. This can ease the a priori evaluation of mitigation measures efficacy.

Acknowledgement

The authors would like to thank David Fages for his help with the soil sampling. We also thank Sandrine Negro and Manon Lagacherie for their help with the measurement of the sorption coefficients. Last we warmly thank Pauline Campan for providing the sorption coefficients and soil physico-chemical properties of the WI soils.

Statements & Declarations

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contributions

Jeanne Dollinger contributed a vast majority of the study conception and design with the help of Cécile Gomez and Anatja Samouelian. Jeanne Dollinger performed the measurement of the sorption coefficients for the FR-RO and FR-RI soils. Data on the WI soils were provided by Anatja Samouelian. The MIR spectra were acquired by Jeanne-Chantal Thoisy. Jeanne Dollinger and Cécile Gomez performed the chemometric analyses. The first draft of the manuscript was written by Jeanne Dollinger, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

The study was funded by a young scientist starting grant from the AgroEcoSystem department of INRAE.

Ethical Approval

Not applicable

Consent to Participate

Not applicable

Consent to Publish

Not applicable

References

Allo, M., Todoroff, P., Jameux, M., Stern, M., Paulin, L., Albrecht, A., 2020. Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. CATENA 189, 104452. https://doi.org/10.1016/j.catena.2020.104452

Artz, R.R.E., Chapman, S.J., Robertson, A.H.J., Potts, J.M., Laggoun-Défarge, F., Gogo, S., Comont, L., Disnar, J.-R., Francez, A.-J., 2008. FTIR spectroscopy can predict organic matter quality in regenerating cutover peatlands. Soil Biol. Biochem. 40, 515.

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. TrAC Trends Anal. Chem. 135, 116166. https://doi.org/10.1016/j.trac.2020.116166

Barriuso, E., Laird, D.A., Koskinen, W.C., Dowdy, R.H., 1994. Atrazine Desorption from Smectites. Soil Sci. Soc. Am. J. 58, 1632–1638. https://doi.org/10.2136/sssaj1994.03615995005800060008x

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends Anal. Chem. 29, 1073–1081. https://doi.org/10.1016/j.trac.2010.05.006

Bengtsson, S., Berglöf, T., Kylin, H., 2007. Near Infrared Reflectance Spectroscopy as a Tool to Predict Pesticide Sorption in Soil. Bull. Environ. Contam. Toxicol. 78, 295–298. https://doi.org/10.1007/s00128-007-9167-x

Benoit, P., Barriuso, E., Houot, S., Calvet, R., 1996. Influence of the nature of soil organic matter on the sorption-desorption of 4-chlorophenol, 2,4-dichlorophenol and the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D). Eur. J. Soil Sci. 47, 567–578. https://doi.org/10.1111/j.1365-2389.1996.tb01856.x

BNV-D, 2022. Données sur les ventes de produites phytopharmaceutiques.

Boivin, A., Cherrier, R., Schiavon, M., 2005. A comparison of five pesticides adsorption and desorption processes in thirteen contrasting field soils. Chemosphere 61, 668–676. https://doi.org/10.1016/j.chemosphere.2005.03.024

Borggaard, O.K., Gimsing, A.L., 2008. Fate of glyphosate in soil and the possibility of leaching to ground and surface waters: a review. Pest Manag. Sci. 64, 441–456. https://doi.org/10.1002/ps.1512

Briceño, G., Palma, G., Durán, N., 2007. Influence of Organic Amendment on the Biodegradation and Movement of Pesticides. Crit. Rev. Environ. Sci. Technol. 37, 233–271. https://doi.org/10.1080/10643380600987406

Cocozza, C., D'Orazio, V., Miano, T.M., Shotyk, W., 2003. Characterization of solid and aqueous phases of a peat bog profile using molecular fluorescence spectroscopy, ESR and FT-IR, and comparison with physical properties. Org. Geochem. 34, 49–60. https://doi.org/10.1016/S0146-6380(02)00208-5

Cox, R.J., Peterson, H.L., Young, J., Cusik, C., Espinoza, E.O., 2000. The forensic analysis of soil organic by FTIR. Forensic Sci. Int. 108, 107–116. https://doi.org/10.1016/S0379-0738(99)00203-0

Cruz, L.H. da, Santana, H. de, Zaia, C.T.B.V., Zaia, D.A.M., 2007. Adsorption of glyphosate on clays and soils from Paraná State: effect of pH and competitive adsorption of phosphate. Braz. Arch. Biol. Technol. 50, 385–394. https://doi.org/10.1590/S1516-89132007000300004

Dagès, C., Voltz, M., Bailly, J.-S., Crevoisier, D., Dollinger, J., Margoum, C., 2023. PITCH: A model simulating the transfer and retention of pesticides in infiltrating ditches and channel networks for management design purposes. Sci. Total Environ. 891, 164602. https://doi.org/10.1016/j.scitotenv.2023.164602

De, A., Bose, R., Kumar, A., Mozumdar, S., 2014. Worldwide Pesticide Use, in: De, A., Bose, R., Kumar, A., Mozumdar, S. (Eds.), Targeted Delivery of Pesticides Using Biodegradable Polymeric Nanoparticles, SpringerBriefs in Molecular Science. Springer India, New Delhi, pp. 5–6. https://doi.org/10.1007/978-81-322-1689-6_2

De Gerónimo, E., Aparicio, V.C., 2022. Changes in soil pH and addition of inorganic phosphate affect glyphosate adsorption in agricultural soil. Eur. J. Soil Sci. 73, e13188. https://doi.org/10.1111/ejss.13188

de Jonge, H., Wollesen de Jonge, L., 1999. Influence of pH and solution composition on the sorption of glyphosate and prochloraz to a sandy loam soil. Chemosphere 39, 753–763. https://doi.org/10.1016/S0045-6535(99)00011-9

Ding, Q., Wu, H.L., Xu, Y., Guo, L.J., Liu, K., Gao, H.M., Yang, H., 2011. Impact of low molecular weight organic acids and dissolved organic matter on sorption and mobility of isoproturon in two soils. J. Hazard. Mater. 190, 823–832. https://doi.org/10.1016/j.jhazmat.2011.04.003

Dollinger, J., Bourdat-Deschamps, M., Pot, V., Serre, V., Bernet, N., Deslarue, G., Montes, M., Capowiez, L., Michel, E., 2022. Leaching and degradation of S-Metolachlor in undisturbed soil cores amended with organic wastes. Environ. Sci. Pollut. Res. 29, 20098–20111. https://doi.org/10.1007/s11356-021-17204-z

Dollinger, J., Dagès, C., Voltz, M., 2015. Glyphosate sorption to soils and sediments predicted by pedotransfer functions. Environ. Chem. Lett. 13, 293–307. https://doi.org/10.1007/s10311-015-0515-5

Dollinger, J., Pétriacq, P., Flandin, A., Samouelian, A., 2023. Soil metabolomics: A powerful tool for predicting and specifying pesticide sorption. Chemosphere 139302. https://doi.org/10.1016/j.chemosphere.2023.139302

Ellerbrock, R.H., Kaiser, M., 2005. Stability and composition of different soluble soil organic matter fractions–evidence from δ13C and FTIR signatures. Geoderma 128, 28–37. https://doi.org/10.1016/j.geoderma.2004.12.025

Farenhorst, A., 2006. Importance of Soil Organic Matter Fractions in Soil-Landscape and Regional Assessments of Pesticide Sorption and Leaching in Soil. Soil Sci. Soc. Am. J. 70, 1005–1012. https://doi.org/10.2136/sssaj2005.0158

Forouzangohar, M., Cozzolino, D., Kookana, R.S., Smernik, R.J., Forrester, S.T., Chittleborough, D.J., 2009. Direct Comparison between Visible Near- and Mid-Infrared Spectroscopy for Describing Diuron Sorption in Soils. Environ. Sci. Technol. 43, 4049–4055. https://doi.org/10.1021/es8029945

Fu, J., Yu, H.-D., Chen, Z., Yun, Y.-H., 2022. A review on hybrid strategy-based wavelength selection methods in analysis of near-infrared spectral data. Infrared Phys. Technol. 125, 104231. https://doi.org/10.1016/j.infrared.2022.104231

García-Delgado, C., Marín-Benito, J.M., Sánchez-Martín, M.J., Rodríguez-Cruz, M.S., 2020. Organic carbon nature determines the capacity of organic amendments to adsorb pesticides in soil. J. Hazard. Mater. 390, 122162. https://doi.org/10.1016/j.jhazmat.2020.122162

Gatel, L., Lauvernet, C., Carluer, N., Weill, S., Tournebize, J., Paniconi, C., 2019. Global evaluation and sensitivity analysis of a physically based flow and reactive transport model on a laboratory experiment. Environ. Model. Softw. 113, 73–83. https://doi.org/10.1016/j.envsoft.2018.12.006

Godeau, C., Morin-Crini, N., Staelens, J.-N., Martel, B., Rocchi, S., Chanet, G., Fourmentin, M., Crini, G., 2021. Adsorption of a triazole antifungal agent, difenoconazole, on soils from a cereal farm: Protective effect of hemp felt. Environ. Technol. Innov. 22, 101394. https://doi.org/10.1016/j.eti.2021.101394

Gurson, A.P., Ozbay, I., Ozbay, B., Akyol, G., Akyol, N.H., 2019. Mobility of 2,4-Dichlorophenoxyacetic Acid, Glyphosate, and Metribuzine Herbicides in Terra Rossa-Amended Soil: Multiple Approaches with Experimental and Mathematical Modeling Studies. Water. Air. Soil Pollut. 230, 220. https://doi.org/10.1007/s11270-019-4266-y

Hermansen, C., Norgaard, T., Wollesen de Jonge, L., Moldrup, P., Müller, K., Knadel, M., 2020. Predicting glyphosate sorption across New Zealand pastoral soils using basic soil properties or Vis–NIR spectroscopy. Geoderma 360, 114009. https://doi.org/10.1016/j.geoderma.2019.114009

Kah, M., Brown, C.D., 2007. Prediction of the Adsorption of Ionizable Pesticides in Soils. J. Agric. Food Chem. 55, 2312–2322. https://doi.org/10.1021/jf063048q

Khalid, S., Shahid, M., Murtaza, B., Bibi, I., Natasha, Asif Naeem, M., Niazi, N.K., 2020. A critical review of different factors governing the fate of pesticides in soil under biochar application. Sci. Total Environ. 711, 134645. https://doi.org/10.1016/j.scitotenv.2019.134645

Kodešová, R., Kočárek, M., Kodeš, V., Drábek, O., Kozák, J., Hejtmánková, K., 2011. Pesticide adsorption in relation to soil properties and soil type distribution in regional scale. J. Hazard. Mater. 186, 540–550. https://doi.org/10.1016/j.jhazmat.2010.11.040

Kookana, R.S., Ahmad, R., Farenhorst, A., 2014. Sorption of Pesticides and its Dependence on Soil Properties: Chemometrics Approach for Estimating Sorption, in: Non-First Order Degradation and Time-Dependent Sorption of Organic Chemicals in Soil, ACS Symposium Series. American Chemical Society, pp. 221–240. https://doi.org/10.1021/bk-2014-1174.ch012

Liland, K., Mevik, B., Wehrens, R., 2022. pls: Partial Least Squares and Principal Component Regression. R package version 2.8-1.

Lohumi, S., Lee, S., Cho, B.-K., 2015. Optimal variable selection for Fourier transform infrared spectroscopic analysis of starch-adulterated garlic powder. Sens. Actuators B Chem. 216, 622–628. https://doi.org/10.1016/j.snb.2015.04.060

Madari, B.E., Reeves, J.B., Machado, P.L.O.A., Guimarães, C.M., Torres, E., McCarty, G.W., 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. Geoderma 136, 245–259. https://doi.org/10.1016/j.geoderma.2006.03.026

Malla, M.A., Gupta, S., Dubey, A., Kumar, A., Yadav, S., 2021. Chapter 7 - Contamination of groundwater resources by pesticides, in: Ahamad, A., Siddiqui, S.I., Singh, P. (Eds.), Contamination of Water. Academic Press, pp. 99–107. https://doi.org/10.1016/B978-0-12-824058-8.00023-2

Matich, E.K., Chavez Soria, N.G., Aga, D.S., Atilla-Gokcumen, G.E., 2019. Applications of metabolomics in assessing ecological effects of emerging contaminants and pollutants on plants. J. Hazard. Mater. 373, 527–535. https://doi.org/10.1016/j.jhazmat.2019.02.084

Ng, W., Minasny, B., Jeon, S.H., McBratney, A., 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Secur. 6, 100043. https://doi.org/10.1016/j.soisec.2022.100043

Niemeyer, J., Chen, Y., Bollag, J.-M., 1992. Characterization of Humic Acids, Composts, and Peat by Diffuse Reflectance Fourier-Transform Infrared Spectroscopy. Soil Sci. Soc. Am. J. 56, 135–140. https://doi.org/10.2136/sssaj1992.03615995005600010021x

Novotny, E.H., Turetta, A.P.D., Resende, M.F., Rebello, C.M., 2020. The quality of soil organic matter, accessed by 13C solid state nuclear magnetic resonance, is just as important as its content

concerning pesticide sorption. Environ. Pollut. 266, 115298. https://doi.org/10.1016/j.envpol.2020.115298
Ololade, I.A., Oladoja, N.A., Oloye, F.F., Alomaja, F., Akerele, D.D., Iwaye, J., Aikpokpodion, P., 2014. Sorption of Glyphosate on Soil Components: The Roles of Metal Oxides and Organic Materials. Soil Sediment Contam. Int. J. 23, 571–585. https://doi.org/10.1080/15320383.2014.846900
Paradelo, M., Hermansen, C., Knadel, M., Moldrup, P., Greve, M.H., de Jonge, L.W., 2016. Field-Scale Predictions of Soil Contaminant Sorption Using Visible–Near Infrared Spectroscopy. J. Infrared Spectrosc. 24, 281–291. https://doi.org/10.1255/jnirs.1228
Parolo, M.E., Savini, M.C., Loewy, R.M., 2017. Characterization of soil organic matter by FT-IR spectroscopy and its relationship with chlorpyrifos sorption. J. Environ. Manage. 196, 316–322. https://doi.org/10.1016/j.jenvman.2017.03.018
Pietrzak, D., Kania, J., Malina, G., Kmiecik, E., Wątor, K., 2019. Pesticides from the EU First and Second Watch Lists in the Water Environment. CLEAN – Soil Air Water 47, 1800376. https://doi.org/10.1002/clen.201800376
PPDB, 2023. Pesticide Properties Database [WWW Document]. URL http://sitem.herts.ac.uk/aeru/ppdb/en/search.htm (accessed 1.6.23).
R Core Team, 2023. : A language and environment for statistical computing. R Foundation for Statistical Computing.
Sabzevari, S., Hofman, J., 2022. A worldwide review of currently used pesticides' monitoring in agricultural soils. Sci. Total Environ. 812, 152344. https://doi.org/10.1016/j.scitotenv.2021.152344
Seybold, C.A., Ferguson, R., Wysocki, D., Bailey, S., Anderson, J., Nester, B., Schoeneberger, P., Wills, S., Libohova, Z., Hoover, D., Thomas, P., 2019. Application of Mid-Infrared Spectroscopy in Soil Survey. Soil Sci. Soc. Am. J. 83, 1746–1759. https://doi.org/10.2136/sssaj2019.06.0205
Shan, R., Chen, Y., Meng, L., Li, H., Zhao, Z., Gao, M., Sun, X., 2020. Rapid prediction of atrazine sorption in soil using visible near-infrared spectroscopy. Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 224, 117455. https://doi.org/10.1016/j.saa.2019.117455
Sharma, A., Kumar, V., Shahzad, B., Tanveer, M., Sidhu, G.P.S., Handa, N., Kohli, S.K., Yadav, P., Bali, A.S., Parihar, R.D., Dar, O.I., Singh, K., Jasrotia, S., Bakshi, P., Ramakrishnan, M., Kumar, S., Bhardwaj, R., Thukral, A.K., 2019. Worldwide pesticide usage and its impacts on ecosystem. SN Appl. Sci. 1, 1446. https://doi.org/10.1007/s42452-019-1485-1
signal developers, 2013. signal: Signal processing.
Simkovic, I., Dlapa, P., Doerr, S.H., Mataix-Solera, J., Sasinkova, V., 2008. Thermal destruction of soil water repellency and associated changes to soil organic matter as observed by FTIR spectroscopy. CATENA, Fire Effects on Soil Properties 74, 205–211. https://doi.org/10.1016/j.catena.2008.03.003
Singh, B., Farenhorst, A., McQueen, R., Malley, D.F., 2016. Near-Infrared Spectroscopy as a Tool for Generating Sorption Input Parameters for Pesticide Fate Modeling. Soil Sci. Soc. Am. J. 80, 604–612. https://doi.org/10.2136/sssaj2015.03.0118
Stevens, A., Ramirez-Lopez, L., 2022. An introduction to the prospectr package. R package Vignette R package version 0.2.6.
Tang, F.H.M., Lenzen, M., McBratney, A., Maggi, F., 2021. Risk of pesticide pollution at the global scale. Nat. Geosci. 14, 206–210. https://doi.org/10.1038/s41561-021-00712-5
Tang, X., Zhu, B., Katou, H., 2012. A review of rapid transport of pesticides from sloping farmland to surface waters: Processes and mitigation strategies. J. Environ. Sci. 24, 351–361. https://doi.org/10.1016/S1001-0742(11)60753-5
Umali, B.P., Oliver, D.P., Ostendorf, B., Forrester, S., Chittleborough, D.J., Hutson, J.L., Kookana, R.S., 2012. Spatial distribution of diuron sorption affinity as affected by soil, terrain and management practices in an intensively managed apple orchard. J. Hazard. Mater. 217–218, 398–405. https://doi.org/10.1016/j.jhazmat.2012.03.050

Verchot, L.V., Dutaur, L., Shepherd, K.D., Albrecht, A., 2011. Organic matter stabilization in soil aggregates: Understanding the biogeochemical mechanisms that determine the fate of carbon inputs in soils. Geoderma 161, 182–193. https://doi.org/10.1016/j.geoderma.2010.12.017

Wang, F., Cao, D., Shi, L., He, S., Li, X., Fang, H., Yu, Y., 2020. Competitive Adsorption and Mobility of Propiconazole and Difenoconazole on Five Different Soils. Bull. Environ. Contam. Toxicol. 105, 927–933. https://doi.org/10.1007/s00128-020-03034-1

Wang, Z., Wu, Q., Kamruzzaman, M., 2022. Portable NIR spectroscopy and PLS based variable selection for adulteration detection in quinoa flour. Food Control 138, 108970. https://doi.org/10.1016/j.foodcont.2022.108970

Wauchope, R.D., Yeh, S., Linders, J.B.H.J., Kloskowski, R., Tanaka, K., Rubin, B., Katayama, A., Kördel, W., Gerstl, Z., Lane, M., Unsworth, J.B., 2002. Pesticide soil sorption parameters: theory, measurement, uses, limitations and reliability. Pest Manag. Sci. 58, 419–445. https://doi.org/10.1002/ps.489

Weber, J.B., Wilkerson, G.G., Reinhardt, C.F., 2004. Calculating pesticide sorption coefficients (Kd) using selected soil properties. Chemosphere 55, 157–166. https://doi.org/10.1016/j.chemosphere.2003.10.049

Wen, J., Li, Z., Huang, B., Luo, N., Huang, M., Yang, R., Zhang, Q., Zhai, X., Zeng, G., 2018. The complexation of rhizosphere and nonrhizosphere soil organic matter with chromium: Using elemental analysis combined with FTIR spectroscopy. Ecotoxicol. Environ. Saf. 154, 52–58. https://doi.org/10.1016/j.ecoenv.2018.02.014

Werner, D., Garratt, J.A., Pigott, G., 2013. Sorption of 2,4-D and other phenoxy herbicides to soil, organic matter, and minerals. J. Soils Sediments 13, 129–139. https://doi.org/10.1007/s11368-012-0589-7

Xiao, K., Abbt-Braun, G., Horn, H., 2020. Changes in the characteristics of dissolved organic matter during sludge treatment: A critical review. Water Res. 187, 116441. https://doi.org/10.1016/j.watres.2020.116441

Yeasmin, S., Singh, B., Johnston, C.T., Sparks, D.L., 2017. Evaluation of pre-treatment procedures for improved interpretation of mid infrared spectra of soil organic matter. Geoderma, 5th International Symposium on Soil Organic Matter 2015 304, 83–92. https://doi.org/10.1016/j.geoderma.2016.04.008

Table 1: Peak assignments for the selected MIR spectral bands

Spectral band (cm^{-1})	*Functional group*	*Putative origin*	*Influenced sorption coefficient*	*References*
3700-365O	OH (*vibration/stretching*)	Kaolinite	Kd_{ads}(GLY/2,4-D/DIF) Kf_{des}(GLY/DIF) n_{des}(GLY)	*Allo et al., 2020; Parolo et al., 2017; Simkovic et al., 2008; Yeasmin et al., 2017*
3625-3620	Al_2OH (*stretching*)	Aluminum oxide	Kd_{ads}(GLY/DIF) Kf_{des}(DIF) n_{des}(DIF)	*Allo et al., 2020; Yeasmin et al., 2017*
3540-3515	OH (*vibration/stretching*)	Gibbsite	Kd_{ads}(2,4-D) Kf_{des}(DIF) n_{des}(DIF)	*Allo et al., 2020; Madari et al., 2006; Yeasmin et al., 2017*
3450-3400	OH (*vibration/stretching*)	Goethite/hematite	Kd_{ads}(GLY) Kf_{des}(DIF)	*Allo et al., 2020; Yeasmin et al., 2017*
3391	OH (*vibration/stretching*)	Gibbsite	-	*Allo et al., 2020; Yeasmin et al., 2017*
3500-3200	OH (*vibration/stretching*)	Phenols, humic substances, lignin, possible overtones with phyllosilicate minerals	Kd_{ads}(GLY/2,4-D) Kf_{des}(DIF)	*Cox et al., 2000; Parolo et al., 2017; Wen et al., 2018*
3000-2800	CH, CH_2 (*stretching vibration*)	Fats, wax, lipids, humic substances	Kd_{ads}(DIF)	*Cocozza et al., 2003; Madari et al., 2006; Niemeyer et al., 1992; Verchot et al., 2011; Wen et al., 2018*
2800-2260	CH	Benzene rings =C-H	-	*Wen et al., 2018*
2520-2500	CO_3^{2-} (*vibration*)	Minerals	Kd_{ads}(2,4-D) Kf_{des}(DIF) n_{des}(GLY)	*Parolo et al., 2017*
1920-1840	C=O (*stretching*)	Carboxilic acids	Kd_{ads}(2,4-D/DIF) Kf_{des}(GLY/DIF) n_{des}(DIF)	*Wen et al., 2018*
1820-1760	C=O (*stretching*)	Hydrophylic SOM	Kd_{ads}(2,4-D/DIF) Kf_{des}(GLY/DIF) n_{des}(DIF)	*Simkovic et al., 2008; Verchot et al., 2011*

1700-1540	C=O / COO- / C=C	Aromatic organic matter	Kd_{ads}(DIF) n_{des}(GLY /DIF)	Ellerbrock and Kaiser, 2005; Madari et al., 2006; Simkovic et al., 2008; Wen et al., 2018
1650-1600	C=C (*stretching*)	Lignin, aromatic or aliphatic carboxylates	Kd_{ads}(GLY/2,4-D) n_{des}(GLY/DIF)	*Cocozza et al., 2003; Madari et al., 2006; Niemeyer et al., 1992; Parolo et al., 2017; Verchot et al., 2011*
1600-1500	N-H / NH_2 (*bending vibration*)	Proteins	Kd_{ads}(GLY/2,4-D/DIF) Kf_{des}(GLY) n_{des}(DIF)	*Artz et al., 2008; Madari et al., 2006; Parolo et al., 2017; Simkovic et al., 2008*
1450-1430	CO_3^{2-} (*vibration*)	Calcite and minerals of the calcite and dolomite groups	Kd_{ads}(2,4-D) Kf_{des}(GLY)	*Parolo et al., 2017*
1430-1330	COO- (*stretching*)	Carboxylate/Carboxylic structures (humic acids)	Kd_{ads}(GLY/2,4-D) Kf_{des}(GLY/DIF) n_{des}(GLY /DIF)	*Artz et al., 2008; Madari et al., 2006; Parolo et al., 2017; Wen et al., 2018*
1300-1200	C–O *(stretching)* / OH *(deformations)*	Alcohols, ethers, phenols, carboxylic acids and esters	Kd_{ads}(2,4-D/DIF) Kf_{des}(GLY/DIF) n_{des}(GLY/DIF)	*Madari et al., 2006; Parolo et al., 2017; Verchot et al., 2011*
1100-1000	C-O-C	Cellulose, polysaccharides	Kd_{ads}(DIF) Kf_{des}(GLY) n_{des}(GLY)	*Ellerbrock and Kaiser, 2005; Verchot et al., 2011; Wen et al., 2018*
835	Aromatic CH	Lignin	n_{des}(GLY)	*Artz et al., 2008; Madari et al., 2006*
730-720	CH, CH_2	Long chain (>C4) alkanes	Kd_{ads}(DIF)	*Artz et al., 2008; Madari et al., 2006*
900-400	Si-O (*stretching*) / OH *(bending)*	Minerals	Kd_{ads}(GLY/2,4-D/DIF) Kf_{des}(GLY/DIF) n_{des}(GLY/DIF)	*Parolo et al., 2017; Xiao et al., 2020*

703

704

705

Figures

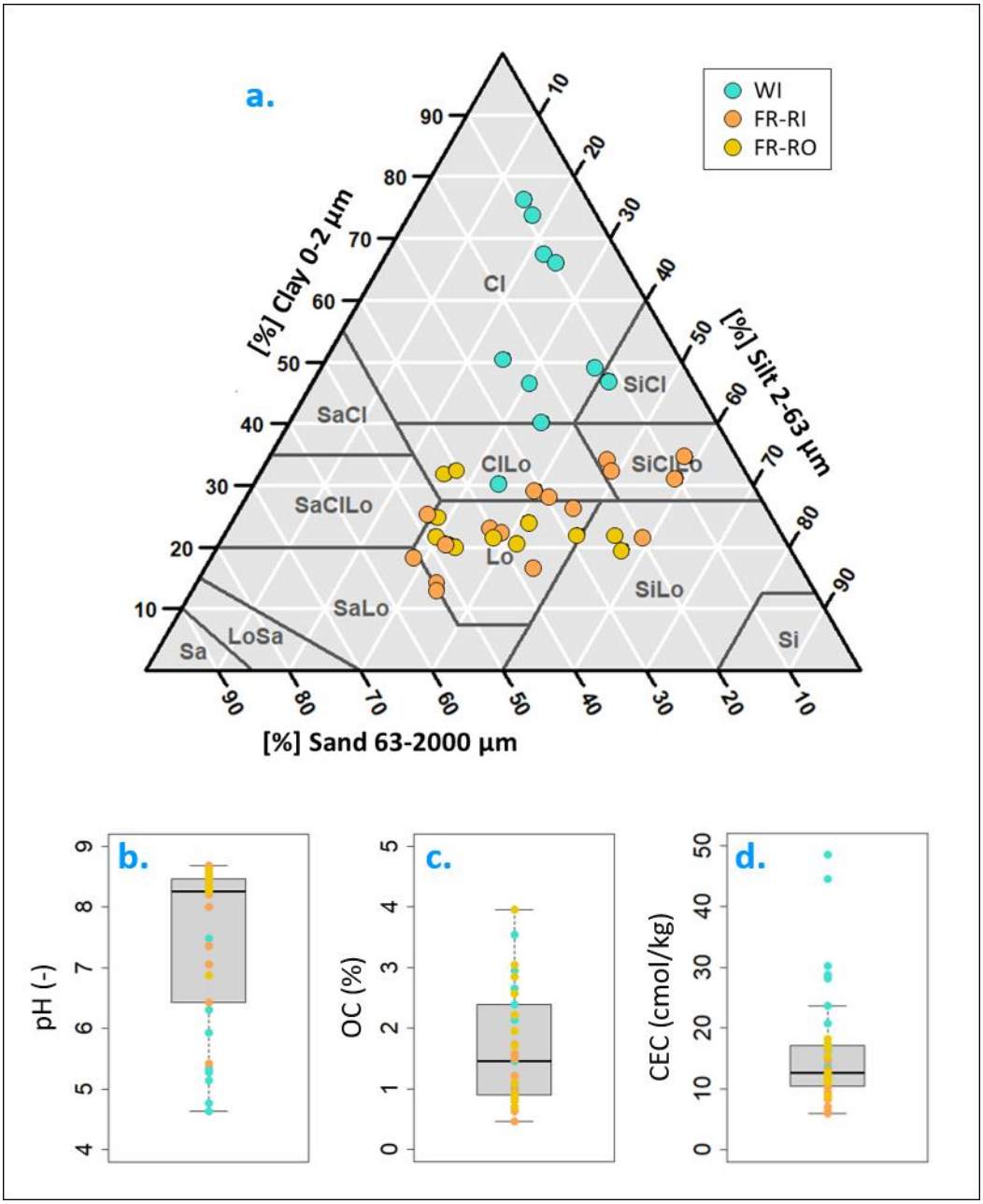
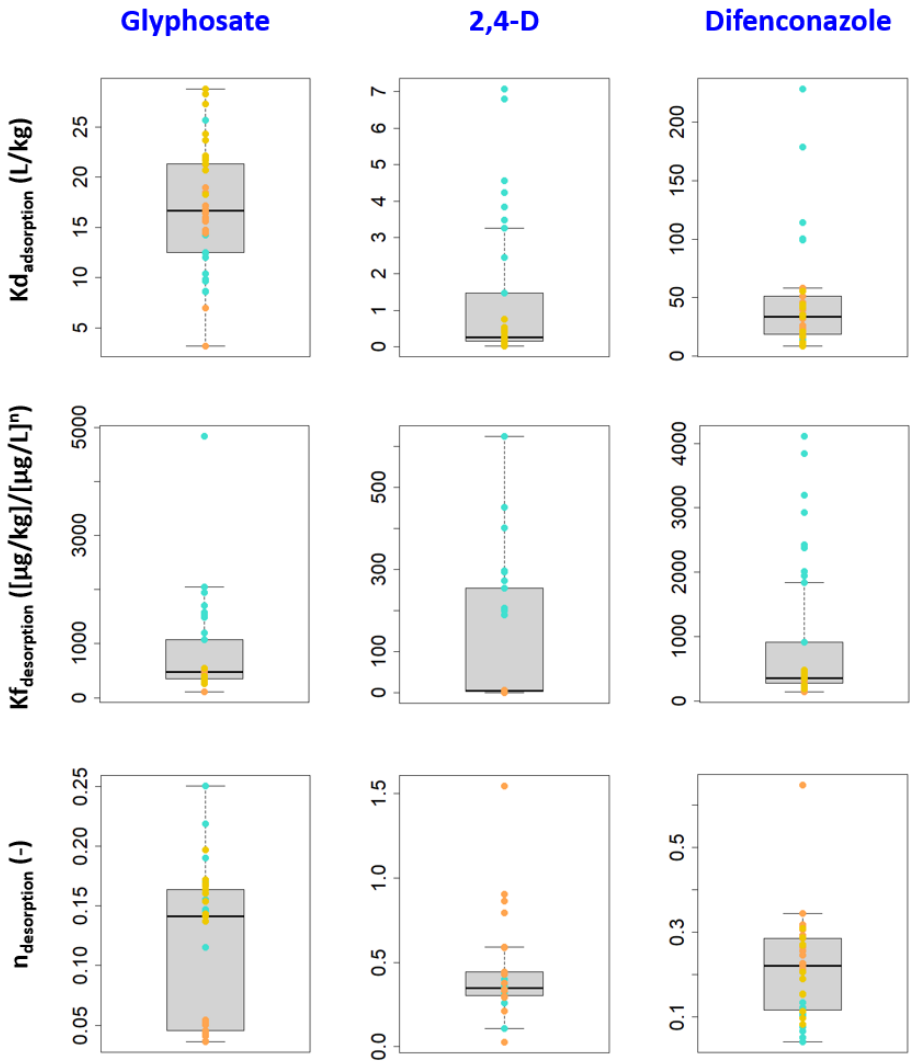


***Figure 1**: **Physico-chemical properties of the soils.** This set of 37 soils includes soils sampled in Guadeloupe in the French West Indies (WI) and two catchments from southern France (FR-RO and FR-RI). The figure displays the texture range (a.), the pH range (b.), the soil organic fraction range (c.) and the cationic exchange capacity range (d.). The texture (a) is classified according to the USDA classification. The letters in the texture triangle refer to the texture class (e.g. ClLo is "Clay Loam").*

Clay												
0.42***	**pH**											
0.29***	0.21**	**CEC**										
0.06	0.09	0.68***	**SOC**									
0.09	0.51***	0.02	0.006	**Kd_{ads} (GLY)**								
0.59***	0.20**	0.43***	0.23**	0.01	**Kf_{des} (GLY)**							
0.08	0.21**	0.19**	0.18**	0.0001	0.007	**n_{des} (GLY)**						
0.34***	0.56***	0.70***	0.48***	0.21**	0.30***	0.29***	**Kd_{ads} (2,4-D)**					
0.41***	0.33**	0.85***	0.75***	0.01	0.63***	0.45***	0.78***	**Kf_{des} (2,4-D)**				
0.12	0.04	0.09	0.05	0.0002	0.13	0.05	0.07	0.12	**n_{des} (2,4-D)**			
0.07	0.21**	0.77***	0.73***	0.05	0.30***	0.07	0.63***	0.65***	0.03	**Kd_{ads} (DIF)**		
0.36***	0.43***	0.81***	0.55***	0.13*	0.50***	0.20**	0.84***	0.89***	0.10	0.77***	**Kf_{des} (DIF)**	
0.19**	0.19**	0.32***	0.42***	0.00009	0.23**	0.23**	0.30***	0.47***	0.004	0.29***	0.40***	**n_{des} (DIF)**

Figure 2: Correlation matrix between the soil properties and the pesticide sorption coefficients. *P-values 0 *** 0.001 ** 0.01 * 0.05. The size of the circles is proportional to the R^2 value.*



718

719 ***Figure 3**: **Measured sorption coefficients**. The figure shows the distributions of adsorption (Kd_{ads}) and*
720 *desorption (Kf_{des} & n_{des}) coefficients measured for the WI soils (turquoise dots), the FR-RO (gold dots)*
721 *and FR-RI (orange dots) soils for the pesticides glyphosate, 2,4-D and difenoconazole.*

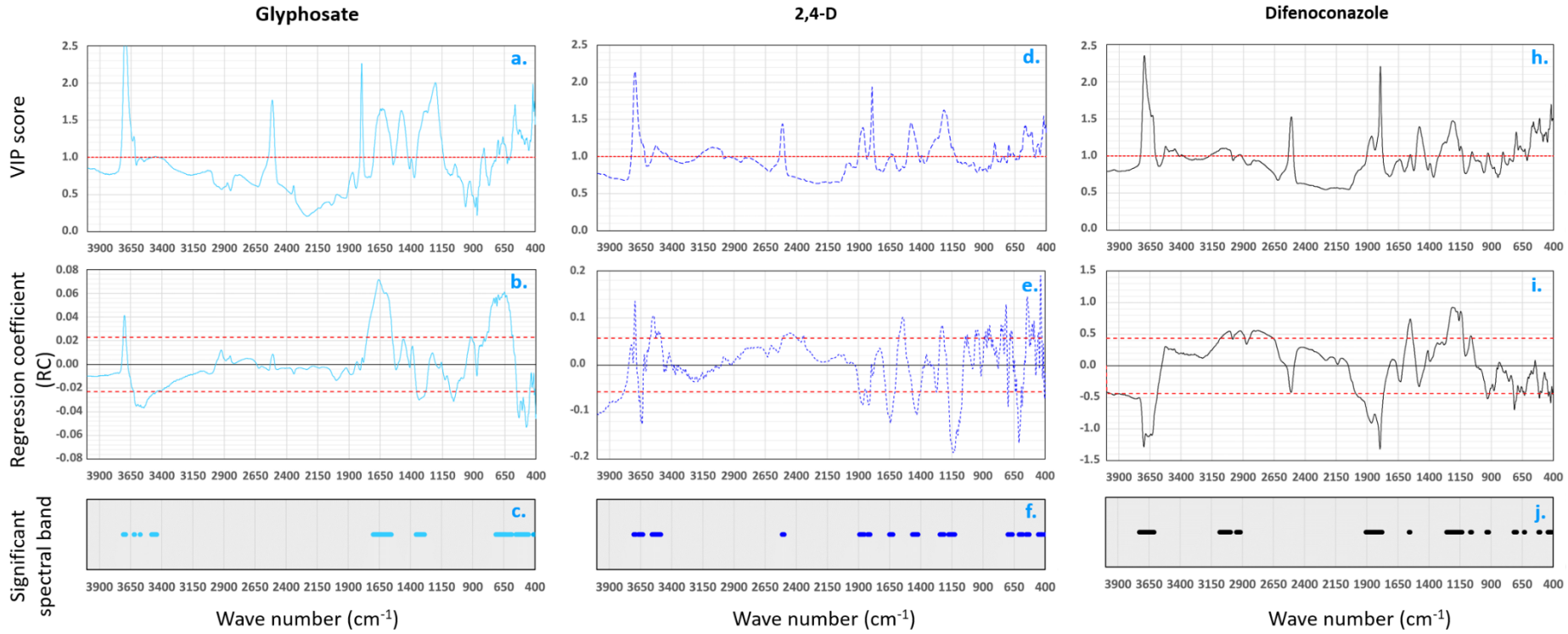
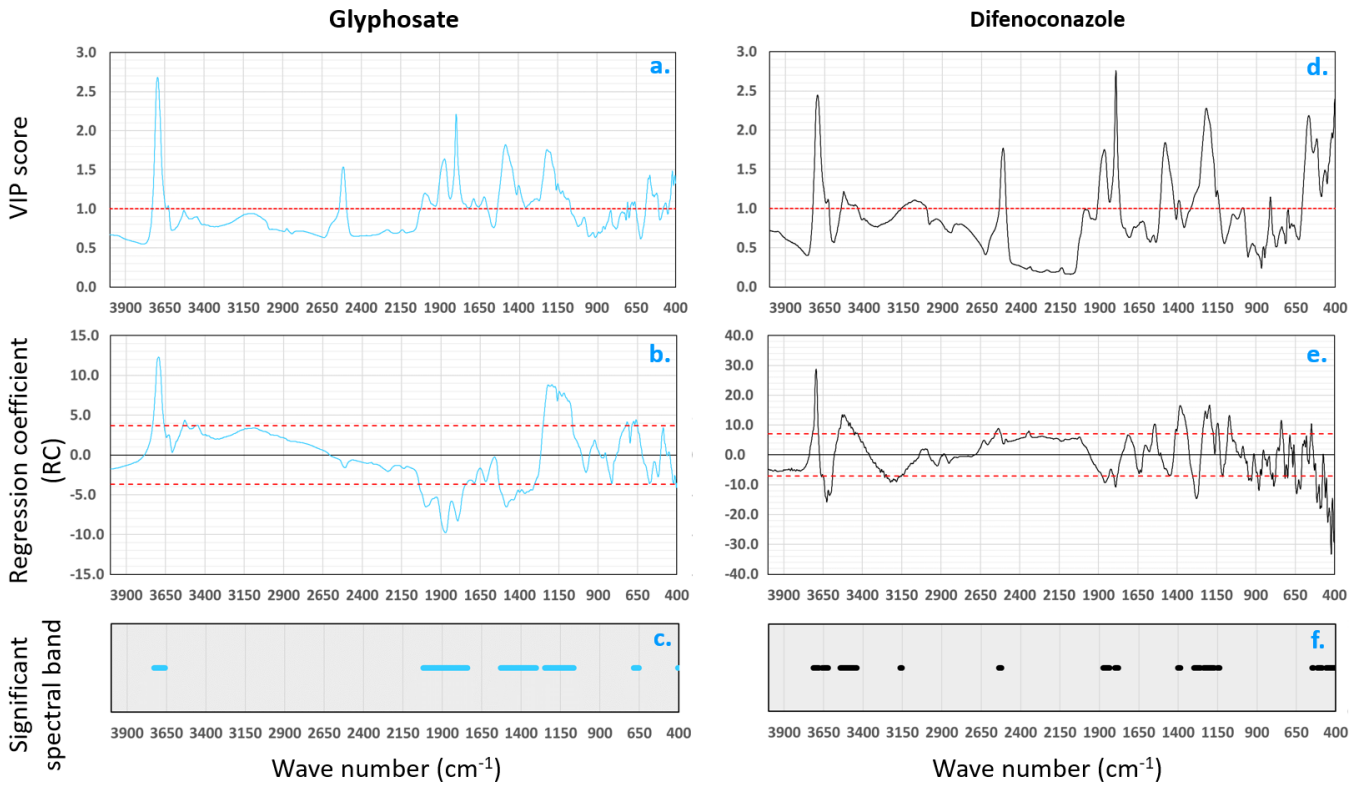


Figure 4: Identification of the discriminant spectral bands in the PLSR for the estimation of the adsorption coefficients (Kd_{ads}). *The wavelength selection is based on the regression coefficients (>sd) and on the VIP score (>1). The variable importance in the projection (VIP) scores are displayed on top for Kd_{ads}(GLY) (**a.**), Kd_{ads}(2,4-D) (d.) and Kd_{ads}(DIF) (h.). The regression coefficients (RC) are plotted in the middle for Kd_{ads}(GLY) (**b.**) Kd_{ads}(2,4-D) (**e.**) and Kd_{ads}(DIF) (**i.**). The red dashed lines represent the significance level above which the wavelengths are considered significant (RC>sd & VIP>1). The selected spectral bands are materialized as vertical dashes; turquoise for glyphosate (**c.**), blue for 2,4-D (**f.**) and black for difenoconazole (**j.**).*



730

731 ***Figure 5: Identification of the discriminant spectral bands in the PLSR for the estimation of the desorption coefficients (Kf_{dess}).*** *The wavelength selection is*
732 *based on the regression coefficients (RC>sd) and on the variable importance in the projection (VIP) score (>1). The figure displays the VIP scores on top for*
733 *Kf_{des}(GLY) (**a.**) and Kf_{des}(DIF) (**d.**). The regression coefficients are plotted in the middle for Kf_{des}(GLY) (**b.**) and Kf_{des}(DIF) (**e.**). The red dashed lines represent the*
734 *significance level above which the wavelengths are considered significant (RC>sd & VIP>1). The selected spectral bands are materialized as vertical dashes;*
735 *turquoise for glyphosate (**c.**) and black for difenoconazole (**f.**).*

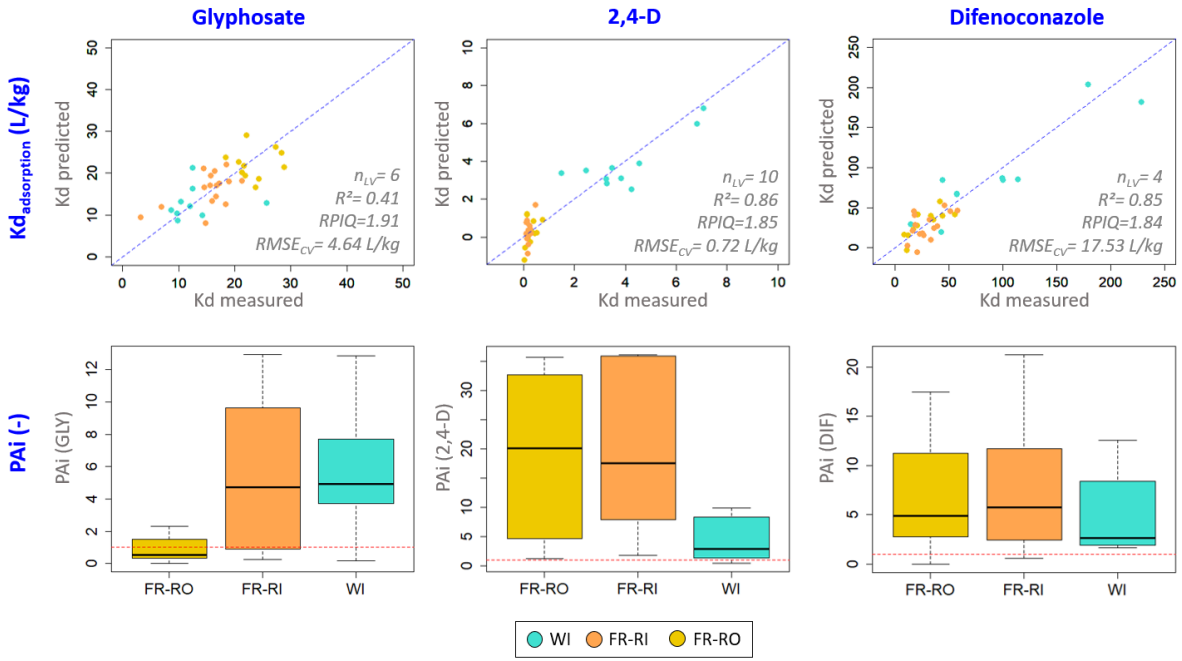


***Figure 6**: **Performance of the PLSR models for the estimation of the adsorption coefficients**. On top the plots represent the predicted vs the measured coefficients and their position relative to the 1:1 line (dashed blue line) for the WI soils in turquoise, the FR-RI soils in orange and the FR-RO soils in gold. The performance criteria (R^2, RPIQ & $RMSE_{CV}$) as well as the number of latent variables (n_{LV}) in the PLSR are indicated for each model. For Kd_{ads}(GLY) the SNV-treated spectra yielded higher accuracy and were used for the PLSR. For the other coefficients, the raw spectra were used for the PLSR. On the bottom, the graphs display the distribution of the PAi index that compares the precision of the model to the accuracy of the Kd_{ads} measure. The red dashed line indicates if the model is more precise (PAi <1) or less precise (PAi >1) than the standard deviation of the measure.*

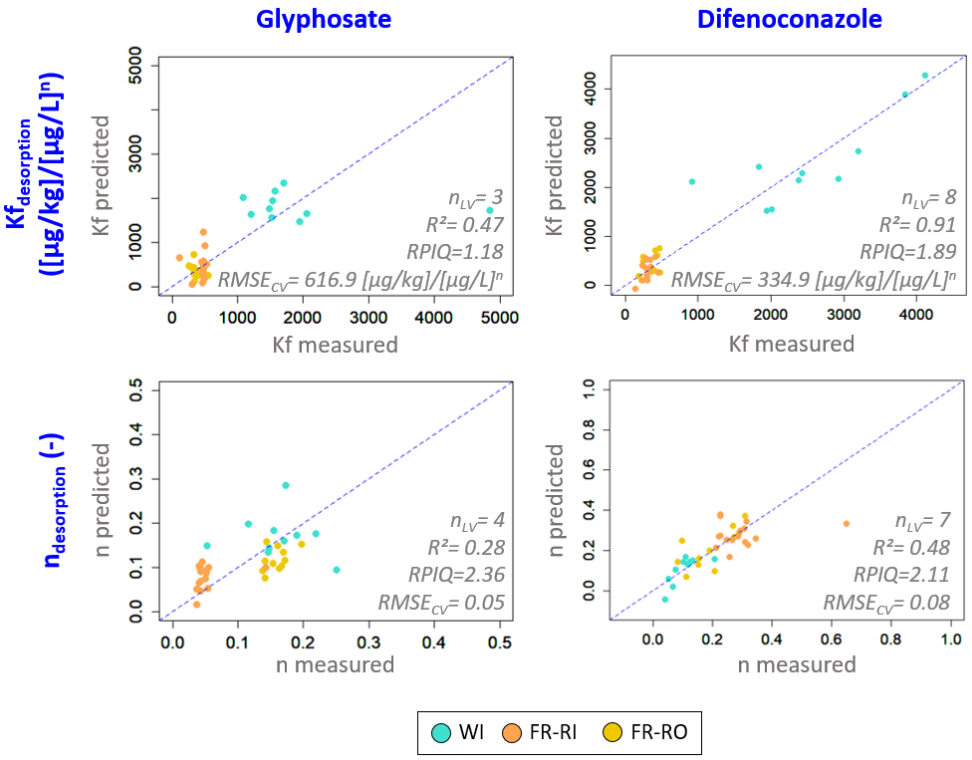


***Figure 7*: Performance of the PLSR models for the estimation of the desorption coefficients.** *The plots represent the predicted vs the measured coefficients and their position relative to the 1:1 line (dashed blue line) for the WI soils in turquoise, the FR-RI soils in orange and the FR-RO soils in gold. The performance criteria (R^2, RPIQ & $RMSE_{CV}$) as well as the number of latent variables (n_{LV}) in the PLSR are indicated for each model. For Kf_{des}(DIF) and n_{des}(DIF) the SNV-treated spectra yielded higher accuracy and were used for the PLSR. For the other coefficients, the raw spectra were used for the PLSR.*

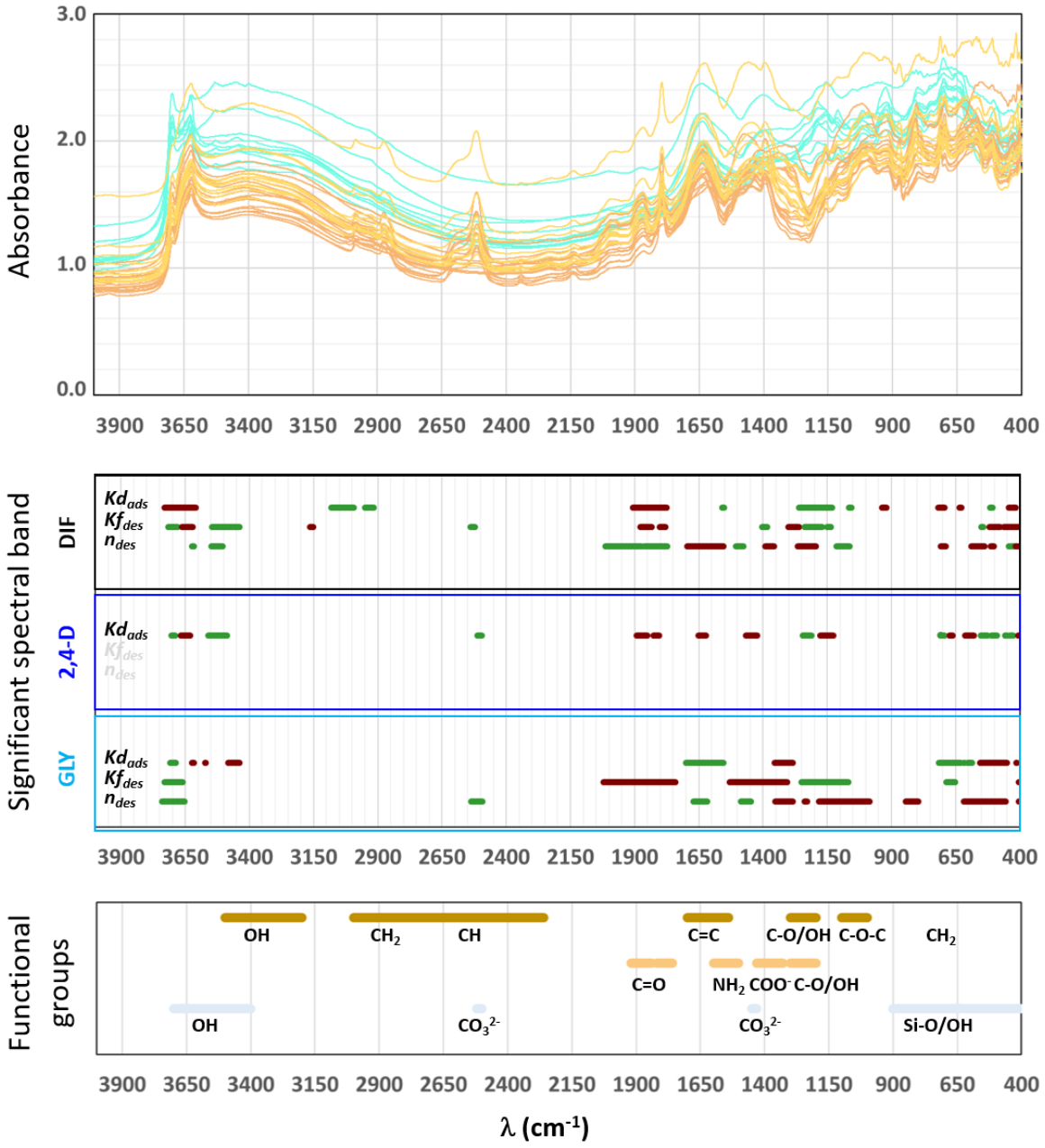


***Figure 8**: Synthesis of the significant wavelengths in the PLSR models for the prediction of the adsorption (Kd_{ads}) and desorption (Kf_{des}, n_{des}) coefficients of glyphosate (GLY), 2,4-D and difenoconazole (DIF). The significant spectral bands are selected from both VIP and RC values. The green and red bands correspond to positive and negative RC values, respectively. Putative functional groups assigned to the significant spectral bands are reminded at the bottom. These are classified according to their putative origin; mineral constituents (grey), hydrophilic SOM (ocher) and hydrophobic SOM (brown).*