

*Article*

# Mid-infrared Laser Spectroscopy Detection and Quantification of Explosives in Soils Using Multivariate Analysis and Artificial Intelligence

Leonardo C. Pacheco-Londoño <sup>1,2,3,\*</sup>, Eric Warren <sup>1</sup>, Nataly J. Galán-Freyle <sup>1,2,3</sup>, Reynaldo Villarreal-González <sup>3</sup>, Joaquín A. Aparicio-Bolaño <sup>1,4,5</sup>, María L. Ospina-Castro <sup>6</sup>, Wei-Chuan Shih <sup>7</sup> and Samuel P. Hernández-Rivera <sup>1,\*</sup>

<sup>1</sup> ALERT DHS Center of Excellence for Explosives Research, Department of Chemistry, University of Puerto Rico, Mayagüez, PR 00681, USA; [leonardo.pacheco@upr.edu](mailto:leonardo.pacheco@upr.edu) (L.P.); [nataly.galan@upr.edu](mailto:nataly.galan@upr.edu) (N.G.); [eric-warren@uiowa.edu](mailto:eric-warren@uiowa.edu) (E.W.)

<sup>2</sup> School of Basic and Biomedical Science, Universidad Simón Bolívar, Barranquilla 080002, Colombia. [leonardo.pacheco@unisimonbolivar.edu.co](mailto:leonardo.pacheco@unisimonbolivar.edu.co) (L.P.); [nataly.galan@unisimonbolivar.edu.co](mailto:nataly.galan@unisimonbolivar.edu.co) (N.G.)

<sup>3</sup> MacondoLab, Universidad Simón Bolívar, Barranquilla 080002, Colombia. [rvillarreal2@unisimonbolivar.edu.co](mailto:rvillarreal2@unisimonbolivar.edu.co) (R.V.)

<sup>4</sup> Department of Physics, University of Miami, Coral Gables, Florida 33124, USA; [jaa362@miami.edu](mailto:jaa362@miami.edu)

<sup>5</sup> Department of Physical Sciences, Miami Dade College Kendall Campus, Miami, Florida 33176, USA; [japaric2@mdc.edu](mailto:japaric2@mdc.edu)

<sup>6</sup> Grupo de Investigación Química Supramolecular Aplicada. Programa de Química, Universidad del Atlántico, Barranquilla 080001, Colombia; [mariaospina@mail.uniatlantico.edu.co](mailto:mariaospina@mail.uniatlantico.edu.co)

<sup>7</sup> Department of Electrical & Computer Engineering University of Houston, 4800 Calhoun Rd. Eng. Bldg. 1, Rm. N308, Houston TX 77204-4005, USA; [wshih@central.uh.edu](mailto:wshih@central.uh.edu)

\* Correspondence: [samuel.hernandez3@upr.edu](mailto:samuel.hernandez3@upr.edu); Tel.: 787-265-5404; 787-832-4040, x-5458, X-2638 (S.H.) [leonardo.pacheco@unisimonbolivar.edu.co](mailto:leonardo.pacheco@unisimonbolivar.edu.co); Tel.: +57-304-648-9549 (L.P.)

**Abstract:** A tunable quantum cascade laser (QCL) spectrometer was used to develop methods for detecting and quantifying high explosives (HE) in soil based on multivariate analysis (MVA) and artificial intelligence (AI). For quantification, mixes of 2,4-dinitrotoluene (2,4-DNT) with concentrations from 0% to 20% w/w were investigated using three types of soils: bentonite, synthetic soil, and natural soil. A Partial least squares regression model was generated for predicting 2,4-DNT concentrations. To increase its selectivity, the model was trained and evaluated using additional analytes as interferences, including other HEs such as PETN, RDX, and TNT and non-explosives such as benzoic acid and ibuprofen. For detection, mixes of different explosives in soils were used to implement two AI strategies. In the first strategy, the spectra of the samples were compared with those of soils recorded in a database to identify the most similar soils based on QCL spectroscopy. Next, a Classical Least Squares preprocessing (Pre-CLS) was applied to soils spectra selected from the database. The parameter obtained based on the sum of the weights of Pre-CLS was then used to generate a simple binary discrimination model for distinguishing between contaminated and uncontaminated soils, achieving an accuracy of 0.877. In the second AI strategy, the same parameter was added to a principal component matrix obtained from spectral data of samples and used to generate multi-classification models based on different machine learning algorithms. A Random Forest model worked best with 0.997 accuracy and allowing to distinguish between soils contaminated with DNT, TNT, or RDX and uncontaminated soils.

**Keywords:** quantum cascade laser; remote detection; partial least squares; high explosives; artificial intelligence; machine learning

---

## 1. Introduction



The intensive use of high explosives (HEs) in military operations and mining excavations has contributed to soil contamination. Providing that HEs and their decomposition products that are highly persistent, mutagenic, and classified as Group C human carcinogens threaten human health [1,2], research on the timely detection of HEs has continued to receive considerable attention over the past few years. The methods currently used to detect HEs include gas chromatography-mass spectroscopy (GC-MS), gas chromatography-chemiluminescence (GC-CL), ion mobility spectrometry (IMS) [3], immunosensors [4], electrophoresis [5], fluorescence [6], high-pressure liquid chromatography (HPLC) [7,8], HPLC/mass spectrometry [7], and photo-assisted electrochemical detection [9]. However, none of these methods provides the required speed or accuracy for *in-situ* detection of HEs in the presence of solid interfering materials.

Soil is considered to be a challenging matrix of organic compounds that interfere with HEs, making the detection of HEs in soils a difficult task [10-12]. While remote sensing has been applied to soils, the proposed system is complex [13]. Other studies conducted by the same research group involved the characterization [14-16], interaction [17-19], and detection [20,21] of HEs using Raman and Fourier-transform infrared (FT-IR) spectroscopy [22,23]. The detection of HEs was only marginally possible in all these cases providing that crystalline samples of explosives must be found within the solid matrix by microscopy to achieve the detection. The *in-situ* detection of HEs in real soils with high selectivity and sensitivity has not been reported yet. In this study, a portable, small, and easy to handle system employing mid-infrared (MIR) spectroscopy with a quantum cascade laser (QCL) source [24] was used for analyzing HEs in soils at a distance of 15 cm. Furthermore, multivariate analysis (MVA) and artificial intelligence (AI) techniques were employed to quantify and detect the analytes of interest in studied soil samples with complex matrices of organic compounds [25].

External cavity QCLs provide extensive wavelength tunability, justifying studies on threat chemicals [26-30] and biological compounds and microorganisms [31] with broad absorption features, including solids. QCLs were first demonstrated in 1994 by Faist *et al.* [32]. These devices offer multiple benefits such as room temperature operation, small sizes, long lifetimes, low energy consumption, long-term power stability, and fine-tuning of the output frequency [33]. Moreover, they provide an opportunity to devise portable systems for remote testing with excellent sensitivity, particularly on samples with low reflectivity or high absorptivity in the MIR range such as soils. Compared to the existing sources in the MIR range, QCL sources have a higher output power that allows remote sensing and analysis. The rapid detection of HEs on highly reflective substrates using this method was demonstrated and reported by other investigators [34-38]. This spectroscopic analysis can be achieved without contact and in a non-destructive way. A study conducted by Phillips *et al.* [39] demonstrated that the standoff detection of many important HEs, including trinitrotoluene (TNT), cyclotrimethylenetrinitramine (RDX), and Tetryl, occurs in the spectral fingerprint window from 800 to 1400  $\text{cm}^{-1}$  in the MIR region when samples are deposited on painted metal car doors. Others studies had demonstrated diverse approaches of QCL for rapid identification and characterization of explosives gases in view of routine requirements of security checks[40,41].

QCLs can provide significant benefits for public safety, particularly in locations such as airports, railways, bus stations, sports stadiums, and marathon sites. Therefore, this study is focused on the quantification and detection of HEs in soils [25] using a QCL source to enable remote sensing [42-53]. The limit of detection (LOD) was calculated for QCL spectroscopy based on the calibration curves of one HE in solid mixtures. This was achieved using the partial least squares (PLS) multivariate algorithm. Other figures of merit (FM) were also calculated. Furthermore, AI methods were employed to achieve HE detection. In the future, two main applications can be developed based on the outcomes of this research. First, the proposed methods can be used to detect explosives on ordinary, non-ideal, low-reflective, real-world solid matrices. Second, the methods can be employed to detect HEs on natural solid matrices; for example, they can be used to locate landmines in sites affected by war and conflicts throughout the world, both for military purposes and implementing humanitarian demining applications.

## 2. Materials and Methods

### 2.1 Reagents

The energetic materials used as analytes in this study were 2,4-dinitrotoluene (2,4-DNT or DNT for simplicity), pentaerythritol tetranitrate (PETN), RDX, and 2,4,6-TNT (or TNT for simplicity). These materials were synthesized in a laboratory, except DNT and TNT, which were purchased from Chem Service (West Chester, PA, USA). Other studied analytes were potassium bromide (KBr), benzoic acid (BA), and ibuprofen (IBP), which were purchased from Sigma-Aldrich (Millipore Sigma, Merck KGaA, St. Louis, MO, USA). Bentonite clay (BC) was also obtained from Sigma-Aldrich. Synthetic soil (SYN-S) samples were prepared from washed sea sand and alundum<sup>TM</sup> cement purchased from Fisher Scientific International (Thermo-Fisher Scientific, Pittsburgh, PA, USA), 75-300 mesh silica gel (Acros Organics, Thermo-Fisher Scientific), as well as montmorillonite, bentonite, and activated charcoal purchased from Sigma-Aldrich. One natural soil (NAT-S) sample was obtained from suburban sites located near the municipality of Mayagüez, PR, USA, with the coordinates 18°13'25.7"N and 67°07'51.2"W. Finally, another natural soil (NAT2-S) sample was obtained to evaluate the model for the coordinates 18°9'36"N and 67°6'40"W.

### 2.2 Sample preparation

Samples containing from 0% to 20% DNT w/w and a total mixture mass of approximately 0.20 g were prepared for the quantification analysis. The initial set of 10 samples and replicates entailed mixtures of several matrices with DNT. The matrices consisted of KBr, BC, SYN-S, and NAT-S. The SYN-S comprised 37% bentonite, 27% alundum<sup>TM</sup> cement, 16% montmorillonite, 10% silica gel, 8% washed sea sand, and 2% activated charcoal. The samples were prepared by grinding the HE into a fine powder using a mortar and pestle, followed by mixing in a mini vortex mixer for approximately 10 s at 3000 rpm. The mixed samples were ground again and mixed in the mini vortex mixer for a second time. Mixtures of KBr and NAT-S with lower concentrations of DNT, approximately from 0% to 3% by weight, were made and named KBr Low and NAT-S Low, respectively. Only these matrices were prepared and tested for comparing the variation of the results with the change of the concentration range. The preparation of these samples was similar to that of the others, except for using different DNT concentrations.

Mixtures containing other highly intrusive compounds such as explosives (TNT, RDX, and PETN) and non-explosives (BA and IBP) were prepared using the same preparation method with only one concentration per interference analytes, namely, 10% w/w. These samples were tested using the PLS model of DNT mixed with NAT-S. For the pattern recognition analysis based on AI, samples of mixtures containing from 0% to 20% of DNT, TNT, and RDX by weight in three soil types (BC and two NAT-S samples) were prepared.

### 2.3 Soil characterization

Soil samples from Mayagüez, PR (USA) were characterized using thermogravimetric analysis (TGA) to measure the water content, sand and clay percentages, and total organic matter (TOM) [54]. TOM values were determined by oxidation with hydrogen peroxide [55,56] and calcination [55]. The total dissolved solids levels were measured by dissolving soil in water, followed by its filtering and drying. The percent distributions of components in the Mayagüez-PR NAT-S samples are illustrated in the Supplementary Materials.

### 2.4 Data acquisition and QCL system

The background spectrum of a KBr substrate was obtained before measuring the QCL spectra of the samples. Providing the lack of MIR signals from the employed solid matrix, this background spectrum provided an excellent and smooth reference trace. The first samples were then placed in the wells of metal holders (1.1 cm in diameter and 3 mm deep). Duplicate spectra were collected at 10 different locations on the sample surfaces, resulting in a total of 20 spectra per sample. Sixteen

spectra were used for calibration and internal validation or cross-validation (CV), while the remaining four spectra were used for testing or external validation. This process was repeated for each concentration. The spectra were obtained in the reflectance mode at a distance of approximately 15 cm using a LaserScan™ MIR pre-dispersive spectrometer (Block Engineering, Marlborough, MA) equipped with three tunable MIR lasers having a tuning range from 990 to 1111  $\text{cm}^{-1}$ , 1111 to 1178  $\text{cm}^{-1}$ , and 1178 to 1600  $\text{cm}^{-1}$ , respectively, spectral linewidth was  $< 2 \text{ cm}^{-1}$ . The scan time was approximately 1.5 s for all spectral range. The average power typically varied between 0.5 and 10 mW across the entire tuning range of  $\sim 600 \text{ cm}^{-1}$  with 100:1 transverse electromagnetic ( $\text{TEM}_{\infty}$ ) polarization and a beam divergence of  $< 2.5 \text{ mrad}$  in the x-axis and  $< 5 \text{ mrad}$  in the y-axis. The spectrometer had a 7.6-cm diameter ZnSe lens, which was used to focus the MIR beam to collect the reflected light and focus the light onto a thermoelectrically cooled mercury–cadmium–telluride (MCT) detector. The wavelength accuracy and precision were  $0.5 \text{ cm}^{-1}$  and  $0.2 \text{ cm}^{-1}$ , respectively. The spectroscopic system worked best at a distance to the target of  $15 \pm 3 \text{ cm}$ , with each laser producing an elliptical spot with diameters of 4 and 2 mm in the same space at a distance of 15 cm due to the difference of beam divergence in the axes(the scheme of the system was reported in Galan-Freyle et al. [30])

## 2.5 Data quantification analysis

The OPUS™ data acquisition and analysis software (v. 4.2 and v. 6.0, Bruker Optics, Billerica, MA, USA) was used to perform multivariate data analysis. Calibration curves were generated from the PLS models of DNT in the studied solid mixtures, and the uncertainties and FM for this model were estimated. The accuracies of the multivariate calibration curves were evaluated using the root-mean-square error of estimation (RMSEE), root-mean-square error of CV (RMSECV), and root-mean-square error of prediction (RMSEP) for external validation. These parameters were used as criteria for evaluating the quality of the proposed method.

The linearity of the calibration curves was evaluated using the values of the coefficient of determination ( $R^2$ ), which indicates the percentage of variance present in the true component values reproduced by the PLS regression model. In contrast, the sensitivity (SEN) of multivariate methods can be estimated as the net analyte signal (NAS) [57] at unit concentration as follows [58]:

$$\text{SEN} = 1/\|\mathbf{b}\|, \quad (1)$$

where  $\mathbf{b}$  denotes the regression vector of the PLS model. The analytical sensitivity ( $\gamma$ ) can be defined similar to univariate calibrations [59] as

$$\gamma = \text{SEN}/\varepsilon, \quad (2)$$

where  $\varepsilon$  denotes instrumental noise. In the absence of interferences, the NAS would be equal to the intensity of the total analyte signal. Noise was measured by collecting 20 spectra of a blank (KBr) and calculating the average of the standard deviations for all wavenumbers.

The LOD was estimated using the following equation [60,61]:

$$\text{LOD} = \Delta(\alpha, \beta) * (\text{RMSEE} * \sqrt{1 + h_o}), \quad (3)$$

where  $h_o$  denotes the distance of the predicted sample to the mean of the calibration set at zero concentration and  $\Delta(\alpha, \beta)$  is a statistical parameter correlated with the  $\alpha$  and  $\beta$  probabilities of falsely stating the presence or absence of the analyte.  $\Delta(\alpha, \beta) = 3.3$  was used to compute the LOD values providing that the value for the degrees of freedom was  $> 25$ .

The residual prediction deviation or relative predictive determinant (RPD) was used to represent how the calibration model predicts a specific set; it can also be used to evaluate the performance of a model in absolute terms. The RPD values can be calculated as

$$\text{RPD} = \text{SD}/\text{SEP}, \quad (4)$$

$$\text{SEP} = \sqrt{((\sum(\text{Diff} - \text{Bias})^2)/(N - 1))}, \quad (5)$$

$$\text{Bias} = (\sum \text{Diff})/N, \quad (6)$$

where SD denotes the standard deviation, SEP is the standard error of prediction, N stands for the number of samples, and Diff denotes the difference in concentration values of the analytes between the predicted and reference sets.

## 2.6. Pattern recognition analysis

An algorithm for comparing different machine learning (ML) methods for classification was developed in Python 3 using the sklearn 3.2 library [62]. Ten ML methods for classification were employed: K-neighbors classifier, Support Vector Machine for classification (SVM), Decision Tree Classifier, Random Forest Classifier (RFC), AdaBoost Classifier (ABC), Gradient Boosting Classifier, Linear Discriminant Analysis, and Quadratic Discriminant Analysis. A basic description of each ML method used in this study is included in Supplementary Materials. Each classifier was trained using 53% of data, while the remaining 47% of data were reserved for testing. The method achieving the highest accuracy value and lowest log-loss value on the test data was considered to be the most efficient.

### 2.6.1. Artificial intelligence scheme

The spectra were preprocessed before they were evaluated by the ML methods. The preprocessing scheme employed in this study is shown in Figure 1. First, all spectra were normalized using vector normalization (VN) preprocessing. Next, the spectral data were reduced using Principal Component Analysis (PCA) from 3057 point to 20 PC. Separately, the pre-processed spectra were compared with those in a database of soils. Subsequently, a Classical Least Squares preprocessing (Pre-CLS) [36] was applied to the original spectra in order to extract a parameter that was proportional to the percentage of soil from the database spectra. Finally, PCA and the extracted parameter were used to generate the ML models.

Pre-CLS is based on a linear model of Classical Least Squares represented as

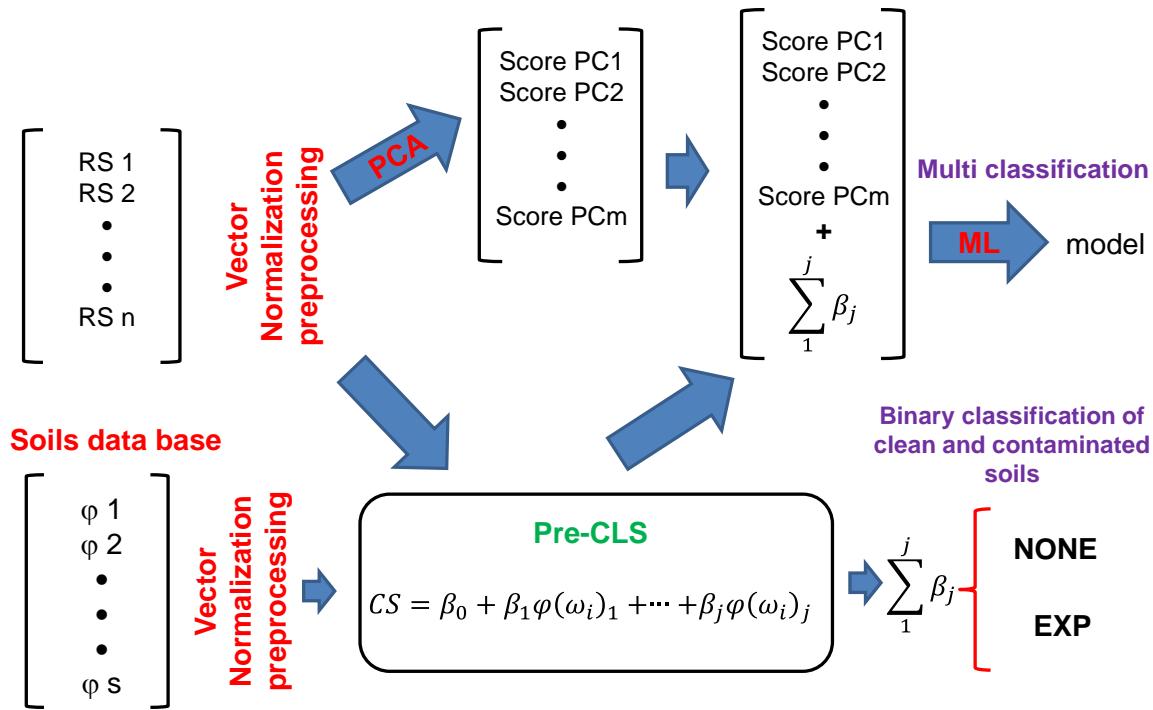
$$f(\varphi_j, \beta_j, \omega_i) = \beta_0 + \beta_1 \varphi(\omega_i)_1 + \cdots + \beta_j \varphi(\omega_i)_j, \quad (7)$$

where  $f(\varphi_j, \beta_j, \omega_i)$  denotes the normalized calculated spectrum (CS) of soil derived from a mixture of several normalized spectra of soils  $\varphi(\omega_i)_j$  recorded in the database of soils spectra (DBS);  $\omega_i$  denotes the wavenumber and  $\beta_j$  is a parameter indicating the fraction or proportion of  $\varphi(\omega_i)_j$  of a particular component in the CS. The model assumes that there are no binding interactions among the components in the mixture, which implies that the intensity contributions are additive. The  $\beta_j$  parameters can be calculated by finding the minimum of the square difference of the normalized intensity between the real spectrum and the CS. The minimum value of the sum of the squares of  $d_i(E)$  with respect to  $\beta_j$  can be found by equating the first-order partial derivatives with respect to  $\beta_j$  to zero and finding the  $\beta_j$  values. Since the model contains  $n$  parameters,  $n$  partial derivative equations are generated as follows:

$$\frac{\partial d^2}{\partial \beta_j} = -2 \sum_i d_i \frac{\partial f(\varphi_j, \beta_j, \omega_i)}{\partial \beta_j} = 0, \quad j = 1, 2, \dots, n. \quad (8)$$

The value of  $\varphi(\omega_i)_j$  can be determined using a simple forward selection algorithm (FSA). According to this algorithm, an empty model is generated first and then variables (soils spectra from the database) are added one by one. At each next step, the  $E$  value shows the improvement of the model. The process is stopped when  $E$  cannot be further improved.

The criteria used to evaluate the performance of the classification models were recall, log-loss, precision, f1-score, weighted average, support, and accuracy. For binary classification, the recall of the positive class is also known as sensitivity, while the recall of the negative class is called specificity.



**Figure 1.** Machine learning pipeline for improving classification.

The log-loss function is used in (multinomial) logistic regression and its extensions such as neural networks. It is defined as the negative log-likelihood of the true labels given the predictions of a probabilistic classifier. The log-loss is defined only for two or more labels. For a single sample with the true label  $y_t$  in (0,1) and estimated probability  $y_p$  that  $y_t = 1$ , the log-loss is

$$-\log P(y_t|y_p) = -(y_t \log(y_p) + (1 - y_t) \log(1 - y_p)). \quad (9)$$

Precision is often called the positive predicted value and calculated as the ratio  $TP/(TP + FP)$ , where  $TP$  denotes the number of true positives and  $FP$  denotes the number of false positives. Intuitively, precision is the ability of a classifier to not label a negative sample as positive. The f1-score is also known as the balanced f-score or the f-measure. The f1-score can be defined as a weighted average of precision and recall, where the f1-score best value is 1 and the worst is 0. The relative contributions of precision and recall to the f1-score are equal. The f1-score in the multi-class and multi-label cases is the average of the f1-score for each class with weighting that depends on the average between precision and recall. The f1-score can be calculated as

$$\text{f1-score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}). \quad (10)$$

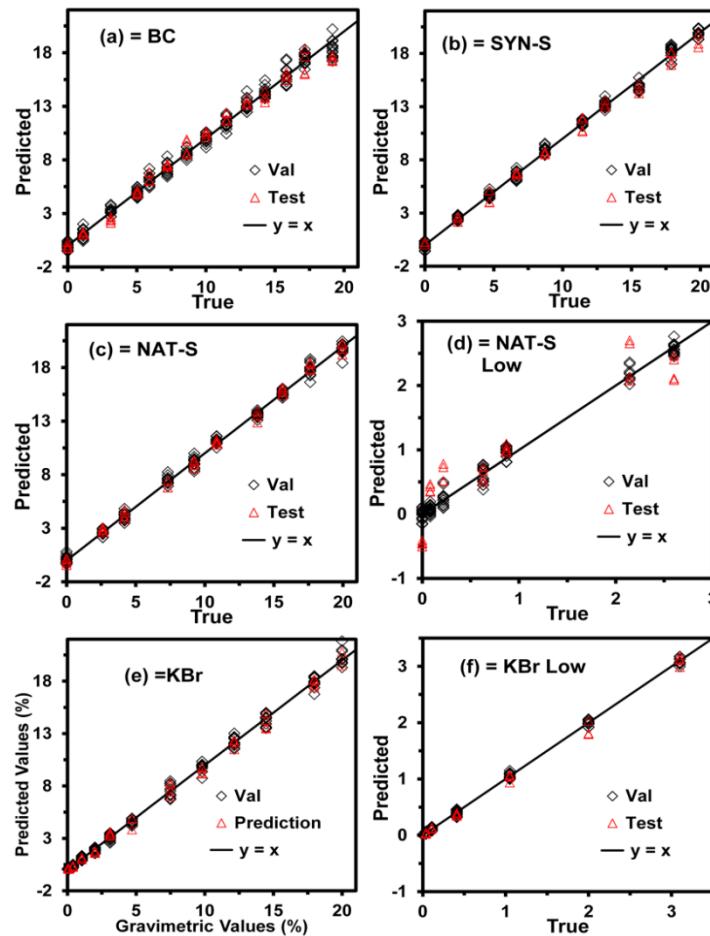
The support is the number of used records, i.e., the number of spectra for each class. The classification accuracy score in multi-label classification computes the accuracy subset: the set of labels predicted for a sample should exactly match the corresponding set of actual labels in an ideal case.

### 3. Results and Discussion

#### 3.1. Quantification analysis

All spectra were converted to the Kubelka–Munk function [63] because the reflectance values for all samples were very low. This transformation is appropriate for the diffuse reflectance spectra of powders [64] with R values < 60% [65]. The models were generated using this transformation. The graphs plotting the predicted versus true values for the samples used in CV and testing are shown in Figure 2, where the CV data are represented with open diamonds, while the test data are

represented with open triangles. The solid black line represents the ideal case or perfect model ( $y = x$ ), where the predicted values are equal to the true values throughout the entire data range.



**Figure 2.** Predicted values vs. real values of % DNT from PLS models on the matrices: (a) BC; (b) SYN-S; (c) NAT-S; and (d) NAT-S Low.

The data set employed in CV was generated using the leave-one-out CV (LOO-CV). It can be noticed from Figure 2 that the predicted values for the CV and test data sets closely follow the best performance line ( $y = x$ ). This alignment can also be confirmed by the low values of the RMSEE, RMSECV, and RMSEP listed in Table 1.

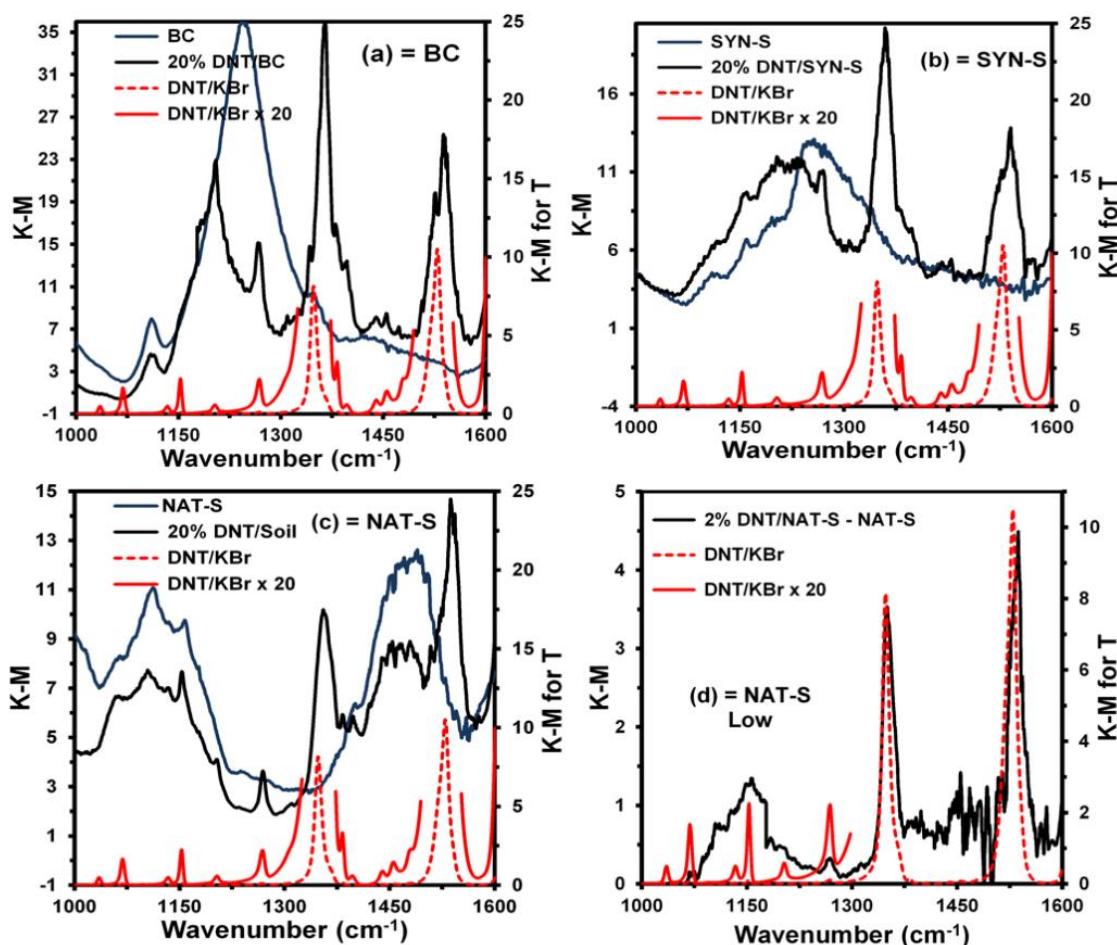
**Table 1.** Accuracy, bias, and number of loading (L) for the studied models and matrices

|           | RMSEE | RMSECV | RMSEP | Bias    | LV |
|-----------|-------|--------|-------|---------|----|
| BC        | 0.41  | 0.57   | 0.70  | -0.0007 | 11 |
| SYN-S     | 0.35  | 0.43   | 0.53  | -0.0069 | 7  |
| NAT-S     | 0.25  | 0.39   | 0.39  | 0.0010  | 10 |
| NAT-S Low | 0.08  | 0.10   | 0.34  | 0.0100  | 7  |
| KBr       | 0.18  | 0.32   | 0.41  | -0.0044 | 10 |
| KBr Low   | 0.02  | 0.03   | 0.08  | -0.0003 | 9  |

All models were generated using the complete spectral window (1000 to 1600  $\text{cm}^{-1}$ ). In this study, VN was used as a preprocessing step as it proved to be better than the other tested preprocessing steps except for KBr. Others preprocessing steps were proved such as mean centering (MC), linear offset subtraction, straight-line subtraction, minimum and maximum normalization, multiplicative scatter correction, first derivative, second derivative, and no preprocessing step.

When applying VN, the average intensity is calculated first, and then this value is subtracted from the spectrum. Next, the sum of the squared intensities is calculated, and the spectrum is divided by the square root of this sum. Models for DNT in KBr were generated to evaluate the detection in the absence of interferences: one at high concentrations (0% to 20%) and another at low concentrations (0% to 3%). The errors for these models are listed in Table 1. The most effective preprocessing method for the KBr models was MC. This result suggests that VN is a suitable preprocessing step only when interferences from the matrix are present; in a model free from the interfering matrix (KBr), applying VN for generating samples with low concentrations had the same signal intensity as samples with high concentrations. This is reflected in the bad prediction and high uncertainty for samples with low concentrations (see Supplemental Materials).

The spectra of the neat matrices, matrices with 20% DNT, and the corresponding standard reference spectra of DNT in KBr for transmittance are shown in Figures 3(a)–3(c), respectively. Figure 3(d) shows the resulting spectrum for the matrix with 2% DNT in NAT-S minus the spectrum of the neat matrix and corresponding standard reference spectra of DNT in KBr for transmittance. For clarity, DNT reference spectra are represented with the red, dotted lines. From these graphs, it is evident that DNT signals are observed on the background matrix signals in all cases, including 2% DNT in NAT-S.



**Figure 3.** DNT spectrum in the investigated soil matrices: (a) BC; (b) SYN-S; (c) NAT-S, and (c) NAT-S Low.

### 3.1.1 Figures of merit

The accuracies of the models were determined from the values of the RMSEE, RMSECV, and RMSEP. The results for each model (labeled by matrix) are listed in Table 1. The precision was inferred through the values of the relative standard deviation (RSD). The RSD values for the prediction of the same sample in the same site were calculated for different concentrations and their

average to measure the repeatability (RSDr). The RSD values for the prediction of the same sample in various locations on the sample surfaces were calculated for different concentrations and their average to measure the mixture homogeneities (RSDh). The RSD values for different samples at the same concentration were calculated for different concentrations and their average to measure the reproducibility (RSDrd). From these values, which are listed in Table 2, it can be concluded that the technique has excellent repeatability, while the samples have excellent homogeneity. However, only good reproducibility was determined for the models.

Another FM used for measuring precision was the RPD. As previously mentioned, the RPD is the ratio of the variation in the validation samples and the size of probable errors occurring during predictions. The RPD values are also listed in Table 2.

**Table 2.** Relative standard deviation (RSD) and relative predictive determinant (RPD) values for the studied models

|                  | RSDr | RSDh | RSDrd | RPD-CV | RPD-test |
|------------------|------|------|-------|--------|----------|
| <b>BC</b>        | 2.4  | 5.5  | 12.7  | 9.7    | 11.3     |
| <b>SYN-S</b>     | 1.5  | 5.3  | 9.1   | 9.2    | 16.2     |
| <b>NAT-S</b>     | 2.2  | 4.2  | 4.8   | 16.1   | 27.6     |
| <b>NAT-S Low</b> | 3.2  | 13.8 | 32.9  | 9.1    | 12.6     |
| <b>KBr</b>       | 3.4  | 3.6  | 4.1   | 19.8   | 36.5     |
| <b>KBr Low</b>   | 5.7  | 6.2  | 11.9  | 27.5   | 50.1     |

The best models are the ones with the lowest RMSEP values and the highest RPD values. The values obtained for the RPD were excellent. It has been suggested that models with RPD values greater than 5 can be considered suitable for quality control, models with RPD values greater than 6.5 can be used for process monitoring, while models with RPD values greater than 8 can be used in any application [66]. All the models tested in this study have RPD values higher than 8. This result indicates that the proposed technique can be used for the direct analysis of explosives in soils. At the same time, the detection of explosives and their respective effects over different soil types have not been studied in detail so far. We plan to expand the present study to include reliable models over various types of soils as part of our future work.

In the context of determining HEs, sensitivity refers to the ability of a method or instrument to detect an analyte at a specified concentration. In contrast, the sensitivity of an analytical method as defined in this study is the capability of the method to discriminate between small differences in concentrations or masses of a target analyte. The sensitivity of each studied method was calculated according to equations (1) and (2), and the results of these calculations are presented in Table 3. The SEN value calculated for a model with VN preprocessing should be interpreted differently from the one without preprocessing or employing other preprocessing steps. Thus, the sensitivity values were derived from two types of models: models with VN as the preprocessing step and models without preprocessing steps. When using VN preprocessing, the spectra were normalized such that the sum of squares of intensity is equal to one. In other words, the spectra are converted to unit vectors. The SEN parameters were calculated from the magnitude of the vector calibration functions **b**, which, in turn, were derived from the spectra and their respective concentrations of standards. VN directly affects the magnitude  $\|\mathbf{b}\|$  and value of SEN as a consequence. However, a better parameter for sensitivity can be obtained by calculating  $\gamma$  because this parameter is only affected by instrumental noises. Noise was measured by collecting 20 spectra of a blank (target) and calculating an average of the standard deviations for all wavenumbers. The resulting noise for the models was different when considering 20 normalized spectra with VN (see Table 3). Otherwise, the  $\gamma$ -values calculated from the two types of models were very similar, which is an indication that  $\gamma$  is not affected by VN preprocessing or any other types of preprocessing.

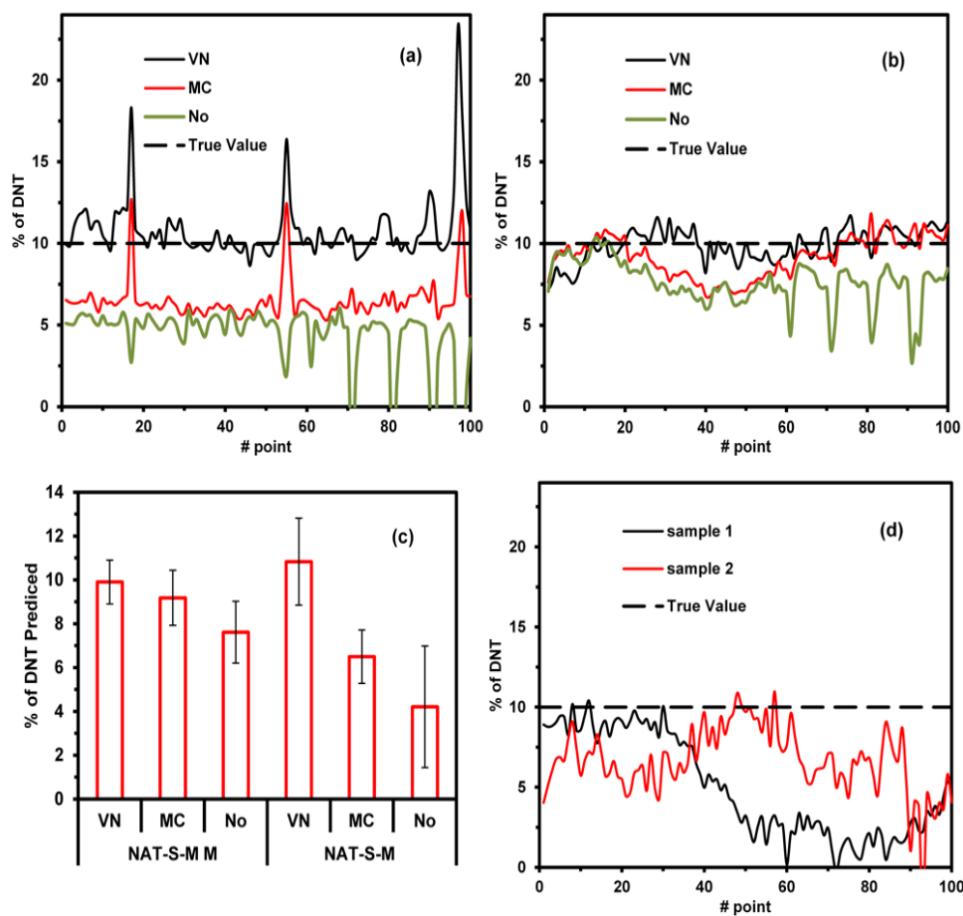
**Table 3.** Sensitivity, analytical sensitivity, limit of detection (LOD), and noise models with and without vector normalization (VN) preprocessing

| VN               |      |       |          |               |
|------------------|------|-------|----------|---------------|
|                  | LOD  | SEN   | $\gamma$ | $\gamma^{-1}$ |
| BC               | 1.4  | 0.013 | 43       | 0.023         |
| SYN-S            | 1.2  | 0.016 | 53       | 0.019         |
| NAT-S            | 0.8  | 0.010 | 32       | 0.031         |
| NAT-S Low        | 0.3  | 0.087 | 289      | 0.003         |
| Noise = 0.0003   |      |       |          |               |
| No preprocessing |      |       |          |               |
|                  | LOD  | SEN   | $\gamma$ | $\gamma^{-1}$ |
| BC               | 2.3  | 1.48  | 39       | 0.025         |
| SYN-S            | 1.4  | 2.25  | 60       | 0.017         |
| NAT-S            | 1.0  | 1.36  | 36       | 0.028         |
| NAT-S Low        | 0.6  | 20.63 | 552      | 0.002         |
| KBr              | 0.6  | 0.51  | 14       | 0.072         |
| KBr Low          | 0.07 | 1.26  | 34       | 0.029         |
| Noise = 0.037    |      |       |          |               |

The inverse of  $\gamma$  (denoted as  $\gamma^{-1}$ ) provides an estimation of the minimum concentration difference (resultant) discernible by the model considering the instrumental noise as the only source of error. In the case of NAT-S Low,  $\gamma$  was 0.003% for the model with VN preprocessing and 0.002% for the model without preprocessing, with the difference being statistically insignificant. It is not possible to make a comparison of the sensitivities between the modes with matrix and without matrix because the magnitude of  $\|\mathbf{b}\|$  depends on the number of signals and number of latent variables (LVs) (see Supplemental Materials). For the models of spectra with many signals (BC, SYN-S, and NAT-S), the magnitude of  $\|\mathbf{b}\|$  is higher than that for the models with low intensity signals (KBr models). A better FM for this comparison is the LOD. For the models with the matrix (BC, SYN-S, and NAT-S), the LOD values are close to that of the model without the matrix (KBr models). Curves for the low concentrations of DNT in NAT-S were generated to determine whether the employed concentration range influenced the LOD values of the curves. In these cases, the LOD decreased from 0.8% to 0.3%. DNT concentrations between 0.3% and 0.8% in the NAT-S Low model can be quantified with higher uncertainties and higher probabilities of false positives and missed detections because RSD should be between 10% to 33%.

To test the NAT-S model, a map of three new types of samples was analyzed, and a map of  $10 \times 10 \text{ mm}^2$  (100 points) for each sample was generated. Two samples comprised the same soil as that used in the NAT-S models and were contaminated with 10% of DNT. The first sample consisted of a simple mixture (NAT-S-M). The second sample involved mixing and macerating the components (NAT-S-MM). The concentrations were predicted using the NAT-S model with three preprocessing methods: VN, MC, and no preprocessing. The map for the NAT-S model with VN is included in Supplementary Materials, while the predictions for the 100 points for each map using different preprocessing methods is shown in Figures 4(a) and 4(b). VN preprocessing applied to both samples (NAT-S-M and NAT-S-MM) resulted in better results, with the predicted values being close to the true value on average (10% DNT, see Figure 4(c)). MC preprocessing worked better in NAT-S-MM than NAT-S-M. This indicates that while the macerated process homogenized the size of the particles, MC was not able to compensate for the difference in the particle size. In contrast, the VN preprocessing method was able to compensate for these differences. The prediction of DNT in NAT-S-M shows peaks of high DNT concentrations (>10% DNT). This is because the particles of DNT were not

homogenized. The models with no preprocessing provided bad predictions due to the difference in the baseline of the spectra. The above discussion demonstrates that VN preprocessing corrects the spectral variation due to changes in the particle size. To demonstrate the change in spectrum with the particle size, the tested soil was sieving for three particle sizes ( $d$ ):  $d > 0.85$  mm,  $0.85 \text{ mm} > d > 0.25$  mm, and  $d < 0.25$  mm. One hundred spectra were acquired at various locations on the sample surface for each value of  $d$ , and the average and standard deviation of the spectra were determined (see Supplementary Materials). The background offset spectral decreases with  $d$ , whereas the standard deviation increases with  $d$ ; however, this pattern is not consistent throughout the spectra. It is higher in the  $1000\text{--}1200\text{-cm}^{-1}$  and  $1400\text{--}1600\text{-cm}^{-1}$  regions. This can be explained by the fact that MC is not able to correct the background offset spectral completely in contrast to VN; VN is better because it scales the spectrum to unit vectors, whereas MC only changes the baseline.



**Figure 4.** Predictions for 100 pt. maps after applying VN, mean centering (MC), and no preprocessing: (a) NAT-S-M; (b) NAT-S-MM; (c) Averages for the predictions of NAT-S-M and NAT-S-MM (c) NAT2-S with VN preprocessing.

In the third sample, another type of natural soil (NAT2-S) was used to evaluate the NAT-S model. Two sample of NAT2-S was contaminated with 10% of DNT and mixed. Mappings of  $10 \times 10 \text{ mm}^2$  (100 points) were generated with the %DNT predicted from the NAT-S model using VN preprocessing. The Mappings are present in Supplementary Materials, while predictions for each point are illustrated in Figure 4(d). It can be noticed from the figure that the predictions were lower than the true values. This indicates that the NAT-S model quantifies below the true value of 10% DNT; however it is capable of predicting the existence of an explosive in a soil. This indicates that the technique should be used in known soil in order to have good quantification.

To challenge the NAT-S model with other interferences, mixtures of other analytes in soil were prepared with a concentration of 10% (median of the model). The analytes used as interferences were BA, IBP, PETN, RDX, and TNT. BA and IBP do not have nitro groups but have aromatic rings and thus many common signals with DNT in the range of 1000 to 1600 cm<sup>-1</sup>. PETN and RDX are nitro aliphatic explosives. TNT is very similar to DNT but considered to be a more challenging interference. Predictions for these samples were generated using the calibration curves for DNT/NAT-S. While the predicted values should have been zero or close to zero because the samples did not contain DNT, the average predicted concentrations were 8.8% (BA), 3.1% (IBP), -8.0% (PETN), 2.3% (RDX), and 25.8% (TNT) (see Table 4). The objective was to measure the model's capability of discriminating these interferences.

**Table 4.** Prediction of %DNT from the NAT-S models

|             | NAT-S | NAT-S OPT-Val |
|-------------|-------|---------------|
| <b>BA</b>   | 8.8   | 0.02          |
| <b>IBP</b>  | 3.8   | -0.02         |
| <b>PENT</b> | -8.0  | 0.00          |
| <b>RDX</b>  | 2.3   | 0.02          |
| <b>TNT</b>  | 25.8  | 0.11          |

To improve the model with respect to the recognition of interferences, an optimization procedure was applied. This procedure involved implementing optimization of the most significant regions of the spectra after using various preprocessing steps. During optimization, the spectral region was divided into equal spectral sub-regions. Then, the optimum combination of sub-regions was determined by starting with ten sub-regions and successively excluding one sub-region. This procedure continued until the values of the cross-validation errors did not improve further. The RMSECV values were calculated for each combination of the preprocessing steps, and the models with the lowest values of errors were selected. Twenty spectra of each interference were introduced together with the validation spectra set off with 0% of DNT true value. Then, optimization was performed by minimizing the RMSECV value. This was labeled as optimization of the validation set (OPT-Val).

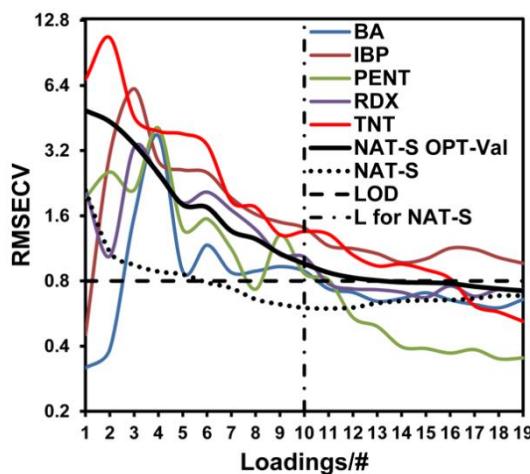
**Table 5.** Performance indicators parameters for the NAT-S models

|                           | NAT-S  | NAT-S OPT-Val |
|---------------------------|--------|---------------|
| <b>R<sup>2</sup> cal</b>  | 99.87  | 99.65         |
| <b>R<sup>2</sup> val</b>  | 99.61  | 98.92         |
| <b>R<sup>2</sup> test</b> | 99.63  | 98.07         |
| <b>RMSEE</b>              | 0.245% | 0.425%        |
| <b>RMSECV</b>             | 0.39%  | 0.72%         |
| <b>RMSEP</b>              | 0.39%  | 0.88%         |
| <b>Bias</b>               | 0.001  | -0.0016       |
| <b>L</b>                  | 10     | 19            |
| <b>ho</b>                 | 0.06   | 0.021         |
| <b>LOD</b>                | 0.80%  | 1.4%          |
| <b>RPD</b>                | 16.1   | 9.61          |

In OPT-Val, all interferences were stabilized with the correct rejections. The parameters for the NAT-S OPT-Val model were approximately similar to those of the NAT-S model except for the

number of loadings ( $L$ ). In particular, nine loadings were added to the NAT-S OPT-Val model to stabilize the interferences (Table 5). In the optimization process, the entire region was selected, and the best pretreatment was found to be VN. Therefore, the only difference between the two models was that the optimized model had more LVs. This procedure was optimum for the elimination of the interferences; however, it assumed that the interferences do not interact with the analyte or the matrix. As part of our future work, samples of DNT in soil with interferences interacting with DNT can be added to the model to remove these potential errors.

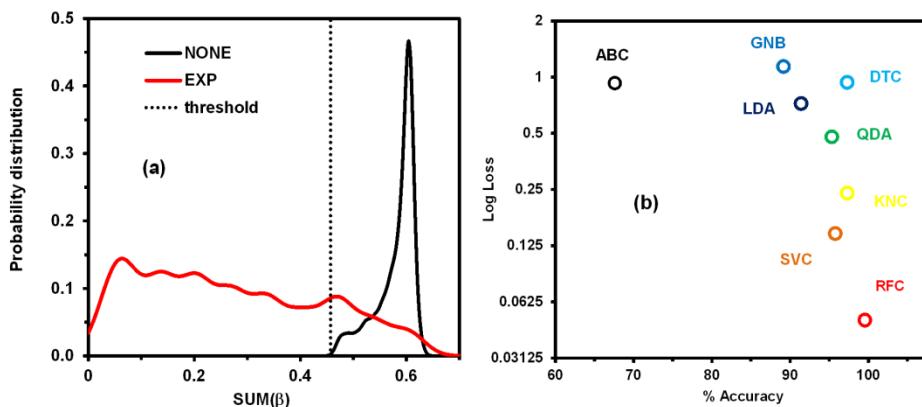
For the NAT-S OPT-Val model, 19 loadings were required to obtain a good RMSECV value. Figure 5 shows the dependence of the RMSECV values for the prediction of each interference as a function of the number of loadings. The NAT-S OPT-Val and NAT-S models were generated to determine how the RMSECV values decrease with the number of loadings and compare the two models. The NAT-S model showed the minimum RMSECV at  $L = 10$ . Adding more loadings resulted in worse RMSECV values. For the NAT-S OPT-Val model, the minimum RMSECV was achieved with 19 loadings. Nevertheless, at 10 loadings, the RMSECV values for other interferences were also sufficiently close to the RMSECV value for the model. For that reason, TNT and IBP interferences did not allow the RMSECV value of the NAT-S OPT-Val model to approach the RMSECV value of the NAT-S model.



**Figure 5.** Dependence of the root-mean-square error of cross-validation (RMSECV) on the loading ( $L$ ) for the considered interferences: the NAT-S OPT-Val and NAT-S models.

### 3.2. Pattern recognition analysis

All sample spectra were normalized using VN preprocessing. Then, they were evaluated using Pre-CLS. The soil database used for the evaluation initially contained QCL reflectance spectra of four different natural soils, sand, and clay standards. It was expanded using linear combinations of sand and clay by applying the concept of Self-Simulated Learning Artificial Intelligence (SSLAI). SSLAI consists of an artificial intelligence which uses minimal information to develop the machine learning models. In this case, other types of soils are simulated using those that are available and these new simulated soils are used for the analysis. FSA algorithm for Pre-CLS was applied for spectra of the samples and one spectrum of the database was added to the model (see equation 7) for most spectra. Next, the sum of  $\beta_j$  (SUM( $\beta$ )) parameters was calculated without including the bias ( $\beta_0$ ), and a simple binary discrimination model was generated and evaluated (see Table 6) using this parameter to distinguish between contaminated and clean soils.



**Figure 6.** (a) Probability distribution of  $\text{SUM}(\beta)$  for a simple binary discrimination model. (b) Comparison of the tested ML methods in terms of their log-loss and accuracy.

Figure 6(a) shows the probability distribution of  $\text{SUM}(\beta)$  for the two classes of the binary model. In this figure was observed a broad distribution of  $\text{SUM}(\beta)$  for the samples contaminated by explosives (EXP), it is due to the range of concentrations. Also, when samples EXP has the highest concentration of explosives, the value  $\text{SUM}(\beta)$  approaches 0 because it has the lowest proportion of soil. On the other hand, in the NONE sample, the distribution is sharpest and the  $\text{SUM}(\beta)$  is highest.

The evaluation criteria (precision, recall, f1-score, and accuracy) and the confusion matrix for the model are shown in Table 6. This model has high precision in predicting clean soils (NONE), whereas it is moderate in predicting contaminated soils (EXP). This is because soils with low explosive concentrations are predicted as clean, increasing the number of false negatives. This happens at low concentrations (<2%), where the model does not work well.

**Table 6.** Confusion matrix and evaluation criteria for the binary discrimination model using  $\text{SUM}(\beta)$  parameters.

|      | Precision | Recall | F1-score | Support | Accuracy | Matrix of confusion |      |
|------|-----------|--------|----------|---------|----------|---------------------|------|
|      | Model     |        |          |         |          | EXP                 | NONE |
| EXP  | 0.768     | 1.000  | 0.869    | 1126    | 0.877    | 865                 | 261  |
| NONE | 1.000     | 0.793  | 0.884    | 997     |          | 0                   | 997  |

Following the scheme illustrated in Figure 1, PCA was applied to the normalized data providing 20 principal components accounting for 99.99% of variance. These components were normalized using an auto-scale preprocessing, and the  $\text{SUM}(\beta)$  was added as an additional principal component. Then, various ML methods were used to generate multi-classification models to discriminate between soils contaminated with three types of explosives (DNT, TNT, and RDX) and clean soil (NONE). These models were evaluated using accuracy and log-loss calculated over the test data (Figure 6(b)). The best method was RFC. This method constructs a multitude of decision trees during the training process and outputs the class that is the mode of the classes of the individual trees. The evaluation criteria (precision, recall, f1-score, and accuracy) and the confusion matrix calculated for the RFC model over the test data are listed in Table 7. The test data were classified almost entirely. Other strategies such as the generation of models without  $\text{SUM}(\beta)$  were also evaluated (data not shown); however, these provided poor results with a high number of false positives, i.e., many samples of clean soil were classified as containing explosives. This was resolved by adding the  $\text{SUM}(\beta)$  parameter.

**Table 7.** Confusion matrix and evaluation criteria for the Random Forest Classification model

| precision | recall | f-score | support | accuracy | confusion matrix |
|-----------|--------|---------|---------|----------|------------------|
|-----------|--------|---------|---------|----------|------------------|

|             | <b>Model</b> |       |       |     | <b>DNT</b> | <b>NONE</b> | <b>RDX</b> | <b>TNT</b> |
|-------------|--------------|-------|-------|-----|------------|-------------|------------|------------|
| <b>DNT</b>  | 0.997        | 0.997 | 0.997 | 347 | 0.997      | 346         | 1          | 0          |
| <b>NONE</b> | 1.000        | 0.996 | 0.998 | 564 |            | 0           | 564        | 0          |
| <b>RDX</b>  | 0.993        | 1.000 | 0.997 | 149 |            | 1           | 0          | 148        |
| <b>TNT</b>  | 0.990        | 1.000 | 0.995 | 100 |            | 0           | 1          | 0          |
|             | <b>Test</b>  |       |       |     | <b>DNT</b> | <b>NONE</b> | <b>RDX</b> | <b>TNT</b> |
| <b>DNT</b>  | 1.000        | 0.997 | 0.998 | 289 | 0.996      | 289         | 0          | 0          |
| <b>NONE</b> | 0.993        | 1.000 | 0.997 | 433 |            | 0           | 430        | 1          |
| <b>RDX</b>  | 0.993        | 0.993 | 0.993 | 139 |            | 1           | 0          | 138        |
| <b>TNT</b>  | 1.000        | 0.981 | 0.990 | 102 |            | 0           | 0          | 102        |

## 5. Conclusions

This paper presented methods for quantifying DNT and detecting HEs in natural and synthetic soil matrices. The remote detection of HEs at a distance of 15 cm was achieved using multiple benefits of QCL spectroscopy, providing the evidence that it can be used in direct field applications. The obtained LOD values were adequate for the intended application. This study demonstrated that it is possible to obtain good detection in the MIR range using diffuse reflection in the back-reflection mode for matrices containing components with low reflectivity such as soils. The remote detection of HEs in soils is now much more viable due to the high-power QCLs that are commercially available.

In the presented experiments, DNT signals could be observed in the spectra of solid mixtures with complex matrices, enabling detection of explosives and quantification of DNT. For quantifying DNT in soil matrices, PLS models were useful for generating calibration curves. The best-performing models were obtained using VN preprocessing for all spectra. This preprocessing procedure proved to be suitable for samples, in which signals from analytes spectra overlapped with signals from matrices (natural soil, synthetic soil and bentonite) spectra and relative intensity differences between spectra were small. In models that involved mixtures of explosives with KBr, VN preprocessing is not suitable because the KBr matrix does not have interfering signals in the spectroscopic window of interest. Therefore, this preprocessing step was not required. Hence, it can be concluded that VN is a good preprocessing step for soil samples which present in highly interfering matrices such as others explosives (TNT, RDX, and PETN) and non-explosives (BA and IBP). The strategy of adding samples of other analytes as possible interferences produced more selective models, making them more robust. The interferences used were organic compounds with many signals in the studied region. The interferences included explosives such as TNT, PETN, and RDX, as well as organic compounds such as BA and IBP. The NAT-S model was able to evade all interferences when it was trained with samples contaminated with these compounds. For the model to be accurate for other soils and possible interferences, it should be re-trained with corresponding samples.

Furthermore, an AI based in ML pipeline scheme (see Figure 1) was applied to obtain an RFC model for HEMs detection in soil matrices. This scheme included Pre-CLS assisted by SSLAI and ML methods.

The expansion of the soil database generated by the linear combinations of QCL reflectance spectra of four different soils (natural soils, sand, and clay standards) allowed the application of Self-Simulated Learning Artificial Intelligence (SSLAI) concept. At the same time, it was possible to couple a simple forward selection algorithm (FSA) that facilitated the compression of the model when using simple multivariate analyses; CLS as additional preprocessing and PCA to reduce the number of spectral variables. Therefore, the combination of PCA, Pre-CLS, and SSLAI is an innovative ML pipeline that allows enriching the prediction of the model. Almost all tested samples of the studied soils and explosives (DNT, RDX and TNT) were correctly classified by the model (see Table 7), providing an accuracy score of 0.996.

**Supplementary Materials:** The following supplementary materials are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1): Table S1: Composition percent of soil (NAT-S) samples; Figure S1: KM Spectra of 20%- and 3%-DNT from the NAT-S sample; Figure S2: Vector Normalization of KM Spectra of 20%- and 3%-DNT from the NAT-S sample; Figure S3: PLS models for DNT in KBr using VN prepossessing; Figure S4: a) **b** regression vector for models without preprocessing and with one loading, b) **b** regression vector for models with vector normalization preprocessing and with one loading, c) **b** regression vector for models without preprocessing and with eight loadings, d) **b** regression vector for models with vector normalization preprocessing and with eight loadings; Figure S5: Plot of the number of loading vs. analytical sensitivity for PLS models for DNT in KBr and soil with VN preprocessing and without preprocessing; Figure S6: Map of % of DNT from the NAT-S model (VN) of the NAT-S-M sample; Figure S7: Map of % of DNT from the NAT-S model (VN) of the NAT-S-MM sample; Figure S8: Map of % of DNT from the NAT-S model (VN) of the NAT2-S sample; Figure S9: Map of % of DNT from the NAT-S model (VN) of the NAT2-S sample; Figure S10: QCL average spectra for different particle sizes of soil samples; Figure S11: QCL standard deviation spectra for different sizes of soil samples; Figure S12: QCL average spectrum for each spectrum in differently sized soil samples; Table S13: Basic description of ML methods for classification.

**Author Contributions:** Conceptualization, Leonardo Pacheco-Londoño; Formal analysis, Maria Ospina-Castro, Wei-Chuan Shih and Samuel Hernandez-Rivera; Investigation, Eric Warren; Methodology, Leonardo Pacheco-Londoño, Eric Warren, Nataly Galan-Freyle, Joaquin Aparicio-Bolaño and Maria Ospina-Castro; Project administration, Samuel Hernandez-Rivera; Resources, Samuel Hernandez-Rivera; Software, Reynaldo Villarreal-Gonzalez; Validation, Reynaldo Villarreal-Gonzalez; Writing – original draft, Leonardo Pacheco-Londoño, Nataly Galan-Freyle and Samuel Hernandez-Rivera; Writing – review & editing, Leonardo Pacheco-Londoño, Nataly Galan-Freyle and Samuel Hernandez-Rivera..

**Funding:** This research received no external funding.

**Acknowledgments:** The authors thank the students of the 2013 class of Detection and Analysis of Explosives (CHEM 5175), Chemical Department, University of Puerto Rico Mayaguez Campus: Alexa Gonzalez-Rosario, Angel A. Lugo-Lugo, Julmarie Ramos-Ramirez, Felix Rojas-Roig, and Glorimar Zeno-Rosario, for their contributions in obtaining soil samples and performing preliminary experiments. This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 2013-ST-061-ED0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Frische, T.; Höper, H. Soil microbial parameters and luminescent bacteria assays as indicators for in situ bioremediation of TNT-contaminated soils. *Chemosphere* **2003**, *50*, 415-427, doi:[http://dx.doi.org/10.1016/S0045-6535\(02\)00603-3](http://dx.doi.org/10.1016/S0045-6535(02)00603-3).
2. Correa-Torres, S.N.; Pacheco-Londono, L.C.; Espinosa-Fuentes, E.A.; Rodriguez, L.; Souto-Bachiller, F.A.; Hernandez-Rivera, S.P. TNT removal from culture media by three commonly available wild plants growing in the Caribbean. *Journal of Environmental Monitoring* **2012**, *14*, 30-33, doi:10.1039/C1EM10602C.
3. Hildenbrand, J.; Herbst, J.; Wöllensteiner, J.; Lambrecht, A. Explosive detection using infrared laser spectroscopy. *Proc. SPIE* **2009**, *7222*, 72220B-72220B-72212.
4. Narang, U.; Gauger, P.R.; Ligler, F.S. A Displacement Flow Immunosensor for Explosive Detection Using Microcapillaries. *Analytical Chemistry* **1997**, *69*, 2779-2785, doi:10.1021/ac970153d.
5. Hilmi, A.; Luong, J.H.T. Micromachined Electrophoresis Chips with Electrochemical Detectors for Analysis of Explosive Compounds in Soil and Groundwater. *Environmental Science & Technology* **2000**, *34*, 3046-3050, doi:10.1021/es000888r.

6. Kumar, S.; Venkatramaiah, N.; Patil, S. Fluoranthene Based Derivatives for Detection of Trace Explosive Nitroaromatics. *The Journal of Physical Chemistry C* **2013**, *117*, 7236-7245, doi:10.1021/jp3121148.
7. Sheremata, T.W.; Halasz, A.; Paquet, L.; Thiboutot, S.; Ampleman, G.; Hawari, J. The Fate of the Cyclic Nitramine Explosive RDX in Natural Soil. *Environmental Science & Technology* **2001**, *35*, 1037-1040, doi:10.1021/es001389t.
8. Larson, S.L.; Martin, W.A.; Escalon, B.L.; Thompson, M. Dissolution, Sorption, and Kinetics Involved in Systems Containing Explosives, Water, and Soil. *Environmental Science & Technology* **2008**, *42*, 786-792, doi:10.1021/es0717360.
9. Marple, R.L.; LaCourse, W.R. Application of Photoassisted Electrochemical Detection to Explosive-Containing Environmental Samples. *Anal. Chem.* **2005**, *77*, 6709-6714, doi:10.1021/ac050886v.
10. Gallagher, N.B.; Blake, T.A.; Gassman, P.L. Application of extended inverse scatter correction to mid-infrared reflectance spectra of soil. *Journal of Chemometrics* **2005**, *19*, 271-281, doi:10.1002/cem.929.
11. Forouzangohar, M.; Kookana, R.S.; Forrester, S.T.; Smernik, R.J.; Chittleborough, D.J. Midinfrared Spectroscopy and Chemometrics to Predict Diuron Sorption Coefficients in Soils. *Environmental Science & Technology* **2008**, *42*, 3283-3288, doi:10.1021/es702750d.
12. Gallagher, N.B.; Gassman, P.L.; Blake, T.A. Strategies for Detecting Organic Liquids on Soils Using Mid-Infrared Reflection Spectroscopy. *Environmental Science & Technology* **2008**, *42*, 5700-5705, doi:10.1021/es8005404.
13. ACS Applied Materials & InterfacesMukherjee, A.; Von der Porten, S.; Patel, C.K.N. Standoff detection of explosive substances at distances of up to 150 m. *Appl. Opt.* **2010**, *49*, 2072-2078, doi:10.1364/AO.49.002072.
14. Hernández, M.D.; Santiago, I.; Padilla, I.Y. Macro-sorption of 2,4-dinitrotoluene onto sandy and clay soils. *Proc. SPIE* **2006**, 6217, 621736-621736-621738.
15. Baez, B.; Correa, S.N.; Hernandez-Rivera, S.P.; de Jesus, M.; Castro, M.E.; Mina, N.; Briano, J.G. Transport of explosives I: TNT in soil and its equilibrium vapor. *Proc. SPIE* **2004**, 5415, 1389-1399.
16. Torres, A.; Padilla, I.; Hwang, S. Physical modeling of 2,4-DNT gaseous diffusion through unsaturated soil. *Proc. SPIE* **2007**, 6553, 65531Q-65531Q-65512.
17. Herrera-Sandoval, G.M.; Ballesteros, L.M.; Mina, N.; Briano, J.; Castro, M.E.; Hernandez-Rivera, S.P. Raman signatures of TNT in contact with sand particles. *Proc. SPIE* **2005**, 5794, 1245-1253.
18. Blanco, A.; Mina, N.; Castro, M.E.; Castillo-Chara, J.; Hernandez-Rivera, S.P. Effect of environmental conditions on the spectroscopic signature of DNT in sand. *Proc. SPIE* **2005**, 5794, 1281-1289.
19. Ballesteros, L.M.; Herrera, G.M.; Castro, M.E.; Briano, J.; Mina, N.; Hernandez-Rivera, S.P. Spectroscopic signatures of PETN in contact with sand particles. *Proc. SPIE* **2005**, 5794, 1254-1262.
20. Hernandez-Rivera, S.P.; Manrique-Bastidas, C.A.; Blanco, A.; Primera, O.M.; Pacheco, L.C.; Castillo-Chara, J.; Castro, M.E.; Mina, N. Spectroscopic characterization of nitroaromatic landmine signature explosives. *Proc. SPIE* **2004**, 5415, 474-485.
21. Osorio, C.; Gomez, L.M.; Hernandez, S.P.; Castro, M.E. Time-of-flight mass spectroscopy measurements of TNT and RDX on soil surfaces. *Proc. SPIE* **2005**, 5794, 803-811.
22. Manrique-Bastidas, C.A.; Mina, N.; Castro, M.E.; Hernandez-Rivera, S.P. Raman microspectroscopy and FTIR crystallization studies of 2,4,6-TNT in soil. *Proc. SPIE* **2005**, 5794, 1358-1365.
23. Blanco, A.; Pacheco-Londoño, L.C.; Peña-Quevedo, A.J.; Hernández-Rivera, S.P. UV Raman detection of 2,4-DNT in contact with sand particles. *Proc. SPIE* **2006**, 6217, 621737-621737-621710.

24. Galán-Freyle, N.J.; Pacheco-Londoño, L.C.; Figueroa-Navedo, A.M.; Hernandez-Rivera, S.P. Standoff Detection of Highly Energetic Materials Using Laser-Induced Thermal Excitation of Infrared Emission. *Appl. Spectrosc.* **2015**, *69*, 535-544, doi:10.1366/14-07501.
25. Galán-Freyle, N.J.; Ospina-Castro, M.L.; Medina-González, A.R.; Villarreal-González, R.; Hernández-Rivera, S.P.; Pacheco-Londoño, L.C. Artificial Intelligence Assisted Mid-Infrared Laser Spectroscopy In Situ Detection of Petroleum in Soils. *Applied Sciences* **2020**, *10*, 1319.
26. Rüther, A.; Pfeifer, M.; Lórenz-Fonfría, V.A.; Lüdeke, S. pH Titration Monitored by Quantum Cascade Laser-Based Vibrational Circular Dichroism. *The Journal of Physical Chemistry B* **2014**, *118*, 3941-3949, doi:10.1021/jp4122886.
27. Lüdeke, S.; Pfeifer, M.; Fischer, P. Quantum-Cascade Laser-Based Vibrational Circular Dichroism. *Journal of the American Chemical Society* **2011**, *133*, 5704-5707, doi:10.1021/ja200539d.
28. Shi, Q.; Nelson, D.D.; McManus, J.B.; Zahniser, M.S.; Parrish, M.E.; Baren, R.E.; Shafer, K.H.; Harward, C.N. Quantum Cascade Infrared Laser Spectroscopy for Real-Time Cigarette Smoke Analysis. *Anal. Chem.* **2003**, *75*, 5180-5190, doi:10.1021/ac034217y.
29. Wörle, K.; Seichter, F.; Wilk, A.; Armacost, C.; Day, T.; Godejohann, M.; Wachter, U.; Vogt, J.; Radermacher, P.; Mizaikoff, B. Breath Analysis with Broadly Tunable Quantum Cascade Lasers. *Analytical Chemistry* **2013**, *85*, 2697-2702, doi:10.1021/ac3030703.
30. Galán-Freyle, N.J.; Pacheco-Londoño, L.C.; Román-Ospino, A.D.; Hernandez-Rivera, S.P. Applications of Quantum Cascade Laser Spectroscopy in the Analysis of Pharmaceutical Formulations. *Appl. Spectrosc.* **2016**, *70*, 1511-1519.
31. Padilla-Jiménez, A.C.; Ortiz-Rivera, W.; Rios-Velazquez, C.; Vazquez-Ayala, I.; Hernández-Rivera, S.P. Detection and discrimination of microorganisms on various substrates with quantum cascade laser spectroscopy. *OPTICE* **2014**, *53*, 061611-061611, doi:10.1117/1.OE.53.6.061611.
32. Faist, J.; Capasso, F.; Sivco, D.L.; Sirtori, C.; Hutchinson, A.L.; Cho, A.Y. Quantum Cascade Laser. *Science* **1994**, *264*, 553-556, doi:10.1126/science.264.5158.553.
33. Hvozdara, L.; Pennington, N.; Kraft, M.; Karlowatz, M.; Mizaikoff, B. Quantum cascade lasers for mid-infrared spectroscopy. *Vib. Spectrosc.* **2002**, *30*, 53-58, doi:[http://dx.doi.org/10.1016/S0924-2031\(02\)00038-3](http://dx.doi.org/10.1016/S0924-2031(02)00038-3).
34. Ruiz-Caballero, J.L.; Blanco-Riveiro, L.A.; Ramirez-Marrero, I.A.; Perez-Almodovar, L.A.; Colon-Mercado, A.M.; Castro-Suarez, J.R.; Pacheco-Londoño, L.C.; Hernandez-Rivera, S.P. Enhanced RDX Detection Studies on Various Types of Substrates via Tunable Quantum Cascade Laser Spectrometer Coupled with Grazing Angle Probe. *IOP Conference Series: Materials Science and Engineering* **2019**, *519*, 012007, doi:10.1088/1757-899x/519/1/012007.
35. Pacheco-Londoño, L.C.; Galán-Freyle, N.J.; Figueroa-Navedo, A.M.; Infante-Castillo, R.; Ruiz-Caballero, J.L.; Hernández-Rivera, S.P. Quantum cascade laser back-reflection spectroscopy at grazing-angle incidence using the fast Fourier transform as a data preprocessing algorithm. *Journal of Chemometrics* **2019**, *33*, e3167, doi:10.1002/cem.3167.
36. Pacheco-Londoño, L.C.; Aparicio-Bolaño, J.A.; Galán-Freyle, N.J.; Román-Ospino, A.D.; Ruiz-Caballero, J.L.; Hernández-Rivera, S.P. Classical Least Squares-Assisted Mid-Infrared (MIR) Laser Spectroscopy Detection of High Explosives on Fabrics. *Appl. Spectrosc.* **2019**, *73*, 17-29, doi:10.1177/0003702818780414.

37. Pacheco-Londoño, L.C.; Ruiz-Caballero, J.L.; Ramírez-Cedeño, M.L.; Infante-Castillo, R.; Gálán-Freyle, N.J.; Hernández-Rivera, S.P. Surface Persistence of Trace Level Deposits of Highly Energetic Materials. *Molecules* **2019**, *24*, 3494.
38. C. Pacheco-Londoño, L.; R. Castro-Suarez, J.; J. Galán-Freyle, N.; M. Figueroa-Navedo, A.; L. Ruiz-Caballero, J.; Infante-Castillo, R.; P. Hernández-Rivera, S. Mid-Infrared Laser Spectroscopy Applications I: Detection of Traces of High Explosives on Reflective and Matte Substrates. In *Infrared Spectroscopy - Principles, Advances, and Applications*, 2019; 10.5772/intechopen.81923.
39. Phillips, M.C.; Bernacki, B.E. Hyperspectral microscopy of explosives particles using an external cavity quantum cascade laser. *OPTICE* **2012**, *52*, 061302-061302, doi:10.1117/1.OE.52.6.061302.
40. Pacheco-Londoño, L.C.; Castro-Suarez, J.R.; Hernández-Rivera, S.P. Detection of Nitroaromatic and Peroxide Explosives in Air Using Infrared Spectroscopy: QCL and FTIR. *Advances in Optical Technologies* **2013**, *2013*, 532670, doi:10.1155/2013/532670.
41. Sirkeli, V.P.; Yilmazoglu, O.; Preu, S.; Küppers, F.; Hartnagel, H.L. Proposal for a Monolithic Broadband Terahertz Quantum Cascade Laser Array Tailored to Detection of Explosive Materials. *Sensor Letters* **2018**, *16*, 1-7, doi:10.1166/sl.2018.3919.
42. Pettersson, A.; Wallin, S.; Östmark, H.; Ehlerding, A.; Johansson, I.; Nordberg, M.; Ellis, H.; Al-Khalili, A. Explosives standoff detection using Raman spectroscopy: from bulk towards trace detection. *Proc. SPIE* **2010**, *7664*, 76641K-76641K-76612.
43. Yang, C.S.C.; Brown, E.E.; Hommerich, U.; Jin, F.; Trivedi, S.B.; Samuels, A.C.; Snyder, A.P. Long-Wave, Infrared Laser-Induced Breakdown (LIBS) Spectroscopy Emissions from Energetic Materials. *Appl. Spectrosc* **2012**, *66*, 1397-1402.
44. Misra, A.K.; Sharma, S.K.; Acosta, T.E.; Porter, J.N.; Bates, D.E. Single-Pulse Standoff Raman Detection of Chemicals from 120 m Distance During Daytime. *Appl. Spectrosc.* **2012**, *66*, 1279-1285.
45. Gottfried, J.L.; De Lucia, F.C.; Munson, C.A.; Mizolek, A.W. Standoff Detection of Chemical and Biological Threats Using Laser-Induced Breakdown Spectroscopy. *Appl. Spectrosc* **2008**, *62*, 353-363.
46. Castro-Suarez, J.R.; Pacheco-Londoño, L.C.; Vélez-Reyes, M.; Diem, M.; Tague, T.J.; Hernandez-Rivera, S.P. FT-IR Standoff Detection of Thermally Excited Emissions of Trinitrotoluene (TNT) Deposited on Aluminum Substrates. *Appl. Spectrosc* **2013**, *67*, 181-186.
47. Carter, J.C.; Angel, S.M.; Lawrence-Snyder, M.; Scaffidi, J.; Whipple, R.E.; Reynolds, J.G. Standoff Detection of High Explosive Materials at 50 Meters in Ambient Light Conditions Using a Small Raman Instrument. *Appl. Spectrosc.* **2005**, *59*, 769-775.
48. Averett, L.A.; Griffiths, P.R. Mid-Infrared Diffuse Reflection of a Strongly Absorbing Analyte on Non-Absorbing and Absorbing Matrices. Part II: Thin Liquid Layers on Powdered Substrates. *Appl. Spectrosc.* **2008**, *62*, 383-388.
49. Pacheco-Londoño, L.; Ortiz-Rivera, W.; Primera-Pedrozo, O.; Hernández-Rivera, S. Vibrational spectroscopy standoff detection of explosives. *Anal. Bioanal. Chem.* **2009**, *395*, 323-335, doi:10.1007/s00216-009-2954-y.
50. Van Neste, C.W.; Senesac, L.R.; Thundat, T. Standoff Spectroscopy of Surface Adsorbed Chemicals. *Anal. Chem* **2009**, *81*, 1952-1956, doi:10.1021/ac802364e.
51. Moros, J.; Lorenzo, J.A.; Lucena, P.; Miguel Tobarria, L.; Laserna, J.J. Simultaneous Raman Spectroscopy–Laser-Induced Breakdown Spectroscopy for Instant Standoff Analysis of Explosives Using a Mobile Integrated Sensor Platform. *Anal. Chem* **2010**, *82*, 1389-1400, doi:10.1021/ac902470v.

52. Moros, J.; Laserna, J.J. New Raman–Laser-Induced Breakdown Spectroscopy Identity of Explosives Using Parametric Data Fusion on an Integrated Sensing Platform. *Anal. Chem.* **2011**, *83*, 6275–6285, doi:10.1021/ac2009433.
53. Ortiz-Rivera, W.; Pacheco-Londoño, L.; Castro-Suarez, J.; Felix-Rivera, H.; Hernandez-Rivera, S. *Vibrational spectroscopy standoff detection of threat chemicals*; SPIE: 2011; Vol. 8031.
54. Miyazawa, M.; Pavan, M.A.; de Oliveira, E.L.; Ionashiro, M.; Silva, A.K. Gravimetric determination of soil organic matter. *Braz. Arch. Biol. Technol.* **2000**, *43*, 475–478.
55. Weiner, E.R. *Applications of environmental chemistry: a practical guide for environmental professionals*; Lewis Pub.: Boca Raton, 2000.
56. Yanjun, C.; Achari, G.; Langford, C.H. Protocols for the analysis of transformer oil and its degradation in soil by hydrogen peroxide. *Can. J. Civ. Eng.* **2009**, *36*, 1547–1557, doi:10.1139/L09-065.
57. Lorber, A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.* **1986**, *58*, 1167–1172, doi:10.1021/ac00297a042.
58. Ferreira, M.H.; Braga, J.W.B.; Sena, M.M. Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. *Microelectron. J.* **2013**, *109*, 158–164, doi:<http://dx.doi.org/10.1016/j.microc.2012.03.008>.
59. Olivieri, A.C.; Faber, N.M.; Ferré, J.; Boqué, R.; Kalivas, J.H.; Mark, H. Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report). *Pure Appl. Chem.* **2006**, *78*, 633.
60. Felipe-Sotelo, M.; Cal-Prieto, M.J.; Ferre, J.; Boque, R.; Andrade, J.M.; Carlosena, A. Linear PLS regression to cope with interferences of major concomitants in the determination of antimony by ETAAS. *J. Anal. At. Spectrom.* **2006**, *21*, 61–68, doi:10.1039/B506783A.
61. Galan-Freyle, N.J.; Figueroa-Navedo, A.M.; Pacheco-Londoño, Y.C.; Ortiz-Rivera, W.; Pacheco-Londoño, L.C.; Hernández-Rivera, S.P. Chemometrics-enhanced fiber optic Raman detection, discrimination and quantification of chemical agents simulants concealed in commercial bottles. *Analytical Chemistry Research* **2014**, *2*, 15–22.
62. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J., et al. API design for machine learning software: experiences from the scikit-learn project. In Proceedings of European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases, Prague, Czech Republic, 2013-09-23.
63. Kubelka, P. New Contributions to the Optics of Intensely Light-Scattering Materials. Part I. *J. Opt. Soc. Am.* **1948**, *38*, 448–448, doi:10.1364/JOSA.38.000448.
64. Bull, C.R. Compensation for particle size effects in near infrared reflectance. *Analyst* **1991**, *116*, 781–786, doi:10.1039/AN9911600781.
65. Sirita, J.; Phanichphant, S.; Meunier, F.C. Quantitative Analysis of Adsorbate Concentrations by Diffuse Reflectance FT-IR. *Anal. Chem.* **2007**, *79*, 3912–3918, doi:10.1021/ac0702802.
66. Igne, B.; Hurlburgh, C.R. Local chemometrics for samples and variables: optimizing calibration and standardization processes. *J. Chemometrics* **2010**, *24*, 75–86, doi:10.1002/cem.1274.