Multivariate Analysis in Vibrational Spectroscopy of Highly Energetic Materials and Chemical Warfare Agents Simulants


Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/54104

1. Introduction

The detection of harmful materials in bulk and trace levels present in different matrices: gases/vapors, liquids, and solids is an important consideration for the development of sensors and standoff detection systems for use in National Defense and Security applications. Hazardous chemicals such as highly energetic materials (HEM), homemade explosives (HME), chemical and biological agents are classified as imminent threats, providing terrorists with ways to cause damage to civilians or troops. Chemical warfare agents (CWA) are usually classified as skin-damaging, nerve agents and toxins [1]. Examples of exposures have occurred since World War I with the development of chlorine, phosgene, cyanide and sulfur mustard which were also used in the Iran-Iraq war. In recent times, terrorist attempts involving chemical warfare agents have occurred all over the World as in the case of Sarin (Japan, 1994) and Ricin (London and Paris, 2003) [2]. Threat perception of chemical warfare agents has increased since September 11, 2001 [3,4]. Exposure to low levels of these chemical agents can cause respiratory problems, eye irritation, choking and blisters. LD₅₀ (mg/kg) values for Nerve agents include Tabun (0.08), Sarin (0.01), Soman (0.025) and VX (0.007) [5].

Most studies that have been published about detection of these compounds are based on spectroscopic and chromatographic (GC and HPLC) methodologies [6-11]. Vibrational spectroscopy has demonstrated to be valuable for the detection of HEM, HME, CWA and Simulants (CWAS) and Toxic Industrial Compounds (TIC). In particular, infrared spectroscopy (IRS) and Raman spectroscopy (RS) in various modalities have played unique roles in threat compounds detection [6,12-19]. IRS and RS can be employed for detection of
Explosives and Chemical Warfare Agents, as well as other chemical and biological threats in airports, in military environments, in government buildings and other public safety places. Raman Spectroscopy employs a non-invasive approach that provides high resolution and specificity. Some applications of IRS and RS includes lab based characterization studies as well as forensic field studies of organic and inorganic substances through their vibration signatures.

Optical Fiber Probes (OFP) have been employed in biomedical applications, in communications, in coupling instrumentation to sensing probes and other important modern applications. Moreover, their uses have been extended to excitation and detection of Raman and infrared signals [16, 20]. Fiber optics applications to Raman Spectroscopy can take advantage of a favorable excitation radiation distribution within the sample; allowing the use of higher laser power levels which, in turn, can yield an elevated signal-to-noise ratio (SNR) for a given experiment without increasing the risk of photo-damaging analytes [21, 22]. In 2011 Ramírez-Cedeño et al. utilized Optical Fiber Coupled Raman Spectroscopy (OFC-RS) to detect hazardous liquids concealed in commercial products [6]. They proved that an optical fiber coupled Raman probe was able to discriminate hazardous liquids inside consumer products from common drinks. Ellisson et al. (2007) have also reported drug and liquid explosives detection in concealed in colored plastic containers [23].

Recently infrared spectroscopy has shown progress in the use of more powerful IR sources, such as Quantum Cascade Lasers (QCL) by incorporating these devices in IR reflectance, IR transmission and even in IR microscopy applications [24]. QCL-based setups are being developed for in the field applications such as breath analysis, environmental research, airborne measurements, security applications, laser-based isotope ratio measurements, and many others. In particular, for security applications, optical methods are advantageous because of their capability for remote and standoff detection [14, 25, 26]. Due to improvements in QCL development, mid infrared lasers operating at room temperature with high output powers in the CW regime are commercially available and make it possible to set up a ruggedized system that allows sensing of explosives and others materials outside the laboratory and the ability to enter real world scenarios. With laser based standoff spectroscopy, the detection distance can be a few meters to tens of meters. Because of the inverse square dependence of light intensity, larger distances require high power, collimated light sources such as lasers. For homeland security applications such as detection of suicide bombers or improvised explosive devices, a distance of 50-100 m is generally sufficient.

In this chapter we illustrate the usefulness of incorporating powerful statistical routines to all traditional chemistry disciplines: Chemometrics is the application of statistical tools to plan, execute and analyze experiments in chemistry. To illustrate the power of Chemometrics techniques to analyze experiments in chemistry we have chosen three case studies, all involving identification, quantification, discrimination and classification of chemical threats in different matrices from vibrational spectroscopy multivariate data.

In the first case study a remote Raman detection study was performed for quantification of HEM such as pentaerythritol tetranitrate (PETN) present in different mixtures. The remote
measurements were carried out at 10 m by employing a frequency-doubled 532 nm Nd:YAG pulsed laser as excitation line, the quantification study was performed by using partial least squares regression analysis (PLS), Interval-PLS (iPLS) and Synergy-PLS (siPLS) as chemometrics tools to achieve the best correlation between the remote Raman signal and the concentration (%) of PETN explosive in a mixture with pharmaceutical compound.

In the second case study discussed, Optical Fiber Coupled Raman Spectroscopy (OFP-RS) was employed at 488 nm excitation wavelength for detection of a Chemical Warfare Agent Simulant (CWAS): triethyl phosphate (TEP) inside different commercial bottles: green-plastic, green-glass, clear-plastic, clear-glass, amber-glass and white plastic. Aqueous solutions were also used to discriminate on various bottle materials in commercial beverage products.

In a third case study a Fourier Transform infrared interferometer with MCT detector was used for recording vibrational infrared signals from nitroaromatic and peroxide explosives in the gas phase. Furthermore, a dispersive IR HEM detection system using a quantum cascade laser was used to record MIR spectral signals of 2,4,6-trinitrotoluene, pentaerythritol tetranitrate and cyclotrimethylenetriminitramine on travel baggage surfaces. Several models were generated with and without preprocessing throughout MIR spectrum.

2. Description of methodologies

Explosives compounds employed in the studies illustrated were pentaerythritol tetranitrate (PETN) cyclotrimethylenetriminitramine (RDX), triacetone triperoxide (TATP), 2,4-dinitrotoluene (2,4-DNT), 2,4,6-trinitrotoluene (TNT) were synthesized in the laboratory according to methods described by Urbanski [27] and pharmaceutical active compound: acetaminophen (APAP) was purchased from Aldrich-Sigma Chemical Co. (Milwaukee, WI). Powder mixtures were prepared employing the both compounds (PETN and APAP) mentioned above and the compositions of PETN in the mixtures was varied from 1 to 34% w/w. Components of the mixtures were carefully weighed and mixed using an agate mortar ensuring homogeneity throughout the sample with a total weight of 200 mg.

![Figure 1](image1.png)

**Figure 1.** Commercial bottles (glass and plastic) used to TEP detection using Optical Fiber Coupled Raman spectroscopy.
A chemical warfare agent simulant: (CWAS) triethyl phosphate 99% (TEP) from Acros Organics (New Jersey, USA) was used to demonstrate the detection and quantification capabilities of this kind of compound by Optical Fibers Coupled Raman spectroscopy. Several formulations were made with the stimulant in their commercial containers: Snapple® Kiwi-Strawberry (Snapple Group USA), Pepsi (PepsiCo Inc., USA), Mountain Dew® (PepsiCo Inc., USA), Heineken® (Mendez & Company, PR), Mott’s apple juice® (Mott’s LLP., Rye Brook, NY), Leche Suiza® Low Fat Grade A (Suiza Dairy Corp., Puerto Rico) and Malta India® (PR Brewing Co., Mayaguez, PR). The containers can be seen in Figure 1.

For gas phase measurements of 2,4-DNT and TATP using infrared spectroscopy vapors were collected in a gas cell (10 cm long and 3.5 cm diameter) by slowly heating the sample to generate vapors. Finally, traces of PETN, TNT and RDX were placed on travel baggage surface with size of 1 in². Figure 2 shows a view of how PETN was deposited on travel baggage substrates at different surface concentration. Figure 3 shows a summary the experiments carried out using vibrational spectroscopy techniques such as infrared and Raman. This figure shows the modalities used, the target tested and the chemometrics models utilized.

For detection of PETN mixed with APAP, Raman spectra of mixtures were acquired by employing a Remote Raman system. This system has been described in detail previously [16]. The prototype was modified using a Headwall Photonics Raman Explorer™ spectrograph with optical layout for 532 nm (Headwall Photonics, Inc.) instead of the Andor spectrometer. The remote Raman system consisted of a MEADE ETX-125 Maksutov-Cassegrain telescope (125 mm clear aperture, 1900 mm focal length f/15). The reflecting collector was coupled to the Raman spectrometer with non-imaging, 200 µm diameter optical fiber (model SR-OPT-8024, Andor Technology, Belfast, Northern Ireland). A frequency-doubled 532 nm Nd:YAG pulsed laser system (Quanta Ray INDI Series, Newport-Spectra Physics, Mountain View, CA) was used as the excitation source. The
maximum energy/pulse of the laser at 532 nm was 25 mJ, and it operated at a repetition rate of 10 Hz. The pulse width was approximately 5-8 ns, and the beam divergence was less than 0.5 mrad. A gateable, intensified CCD detector (iStar™ ICCD camera, Model DH-720i-25F-03, Andor Technology, Belfast, Northern Ireland) was used as the photon detector. Andor Technology Solis™ software for spectroscopic, imaging and time-resolved studies was used for spectral data acquisition and processing from the intensified and gated CCD detector. Using this software, the data could be acquired in both imaging and spectroscopic modes. In this experiments, each mixture was placed into a stainless steel sample holder of 0.6 cm in diameter where 30 ft·lb of pressure was applied to generate a tablet. Remote Raman spectra of mixtures were measured at a target at telescope standoff distance of 10 m, in the Raman Shift region 450-3000 cm⁻¹ using an pulsed laser operating at 532 nm excitation line with a constant energy of 25mJ/pulse (at head) and 100 pulses were applied to achieve spectra with good Signal to Noise. A total of 10 spectra were collected for each mixture acquiring around 56 spectra in the specific Raman shift range.

Figure 3. General scheme of experiments using IR and Raman vibrational spectroscopy.

In the detection of TEP chemical warfare simulant in commercial bottles, Raman experiments were performed with a custom built setup (Figure 4) using the strong blue excitation line from an argon ion laser INNOVA 310/8 from Coherent, Inc. at 488.0 nm. The first strand of optical fiber (non-imaging, 600 µm diameter, model AL 1217, Ocean Optics, Inc.) as well as the second (200 µm diameter, model SR-OPT 8024, Andor Technologies Inc.) were used to couple the Raman probe to which a set of laser line filter (to clean satellite lines) and Semrock RazorEdge™ edge filter was used to filter the Rayleigh scattered light. An Andor Technologies spectrograph: Shamrock SR-303i (aperture: f/4; focal length: 303 mm; wavelength resolution: 0.1 nm or 4.2 cm⁻¹ at the excitation wavelength) equipped with a 1200 grooves/mm grating was used to analyze the Stokes scattered light. A high performance, back thin illuminated CCD camera (Andor Technologies model # DU970N-
UVB) with quantum efficiencies of 90% (200 cm\(^{-1}\)) to 95% (3600 cm\(^{-1}\)) served as light detector (Figure 4) and calibration was performed using cyclohexane. Detection was performed with a light source in order to be carried out in the high-background environment conditions. In this experiment, six different bottle materials: clear-glass, green-glass, brown-glass, clear-plastic, white-plastic and green-plastic were used to measure the amount of simulant (triethyl phosphate) within the container. Mixtures ranged from 0% to 100% (v/v) of simulant and water, simulant and commercial beverage product were analyzed.

**Figure 4.** Experiment setup for detection of chemical warfare simulant in commercial bottles.

**Figure 5.** Schematic diagram of experimental setup used in the IR detection. Traces of explosives in gas phase were dragged by air flow and transported to a gas cell for detection.
The IR equipment used for the experiments was a model IFS 66v/S interferometer (Bruker Optics, Billerica, MA). For the experiments described, the system was equipped with a DTGS detector and a potassium bromide (KBr) beamsplitter. A spectroscopic measurement averaged 20 scans at a resolution of 4 cm\(^{-1}\) using OPUS Version 4.2 software in the range of 7500 – 400 cm\(^{-1}\). Target chemicals were TATP, 2,4-DNT and 2,4,6-TNT. Second, an EM-27 Open Path FT-IR interferometer (Bruker Optics, Billerica, MA) was used to obtain the IR spectral information of gas phase TATP samples with a (TE) cooled MCT detector. Third: Quantum Cascade Laser (QCL) based dispersive IR spectrometer LaserTune™ (Block Engineering, Marlborough, MA) was used to obtain the IR spectral information of TATP samples. An Agilent 6890 gas chromatograph (GC) coupled to an Agilent 5893 mass selective detector (MSD) with a capillary column: HP-5 MS 5% phenyl methyl siloxane, Length: 30.0 m, 250.0 µm in diameter and 0.25 µm of film thickness was used for detecting the presence of explosive TATP, 2,4-DNT and 2,4,6-TNT in the gas phase.

Figure 5 shows a schematic diagram of the experimental setup used in the investigation. Samples of 100-300 µg/cm\(^2\) of explosives were placed on the bottom of a 500 mL Erlenmeyer flask at the position labeled (1) in Figure 5. A flow of dry air (1-16 mL/s) at temperatures of 0-38°C was used to transfer the solid explosives to the gas phase. The measurements as function of temperature were done in two forms, scanning the range of temperature and fixed temperature point measurements. Traces of explosives in gas phase were dragged from the surface by the air flow and transported to an IR gas cell for detection. Spectra were recorded using the instrument first at 4 cm\(^{-1}\) of resolution and 25 scans were used for the experiments. A total of 1089 spectra of air with 2,4-DNT, 1194 spectra of air with TATP and 2200 spectra of air were recorded to generate the models. On the OP EM-27 active mode experiments were carried out at lab temperature (25 oC) at 30 scans and 4 cm\(^{-1}\) resolution. The spectral range used was from 700 to 1600 cm\(^{-1}\). Experiments using QCL LaserTune™ active mode experiments were carried out at lab temperature of 20 °C at 1 spectrum and 4 cm\(^{-1}\) resolution. The spectral range used was from 830 to 1430 cm\(^{-1}\). The presence of TATP in air was determined by GC-MS, and the concentration of 2,4-DNT in air for different flow conditions and temperatures were calculated by calibration curves from GC-µECD. Finally, explosive traces (PETN, TNT and RDX) were deposited on travel baggage of 1 in\(^2\) size of and analyzed with the QCL spectrometer.

3. Chemometrics to vibrational spectral analysis

The automation and computerization of laboratories have been carried out with various important consequences. One of them is the rapid acquisition of large amounts of data. However, it is well known that acquiring such large amount of data is far from to providing appropriate answers quicker. Obtaining vibrational spectroscopy multivariate data is not synonymous with possessing vibrational information. The later must be interpreted and placed in context to convert it into useful information for the user. Chemometrics is the field of chemistry that provides the user with the required tools to enable that capability.

A great deal of chemometrics tools have been developed and tested. The most used of these tools to identify, quantify and classify data sets are those that make use of principal
components analysis (PCA), partial least squares (PLS), discriminant analysis (DA), and their combined usage: PLS-DA and hierarchical cluster analysis (HCA). PCA transforms a set of variables into fewer variables (called factors, rank, dimensions or principal components) which contain most of the information (variance) of the initial data set [28-31].

Figure 6. Simplified scheme for a PCA analysis.

The PCA algorithm seeks to save the information from a large number of variables in a small number of uncorrelated components, with minimal loss of information. The main reasons for performing a PCA are: reduction of the number of variables to fewer dimensions that contain as much information as possible and have uncorrelated dimensions (used to avoid multi-collinearity in multiple regressions, among other things). An important method for qualitative analysis of spectral data is principal component analysis. PCA is a method for the investigation of the variation within a multivariable data set. The first step in PCA is to subtract the average value or spectrum from the entire data set, this is called mean centering. The largest source of variation in the data set is called principal component PC1. The 2nd largest source of variation in the data, which is independent of PC1, is called PC2. Principal components form a set of orthogonal vectors. For each one of the data points, the projection of the data point onto the P1 or P2 vector is called a score value. Plots of sample score values for different principal components, typically P1 versus P2 are called score plots. Score plots provide important information about how different samples are related. Principal component plots, also called loading plots, provide information about how different variables are related to each other. In practical cases, PCA uses a single X matrix which is represented by the infrared spectra. PCA is a purely qualitative analysis (does not give a quantitative value that establishes how different are a spectral dataset) to visualize if there is variability between a set of IR spectra. PCA can thus also be used to detect the presence of outliers. Figure 6 shows a simplified PCA scheme [28-29].

Partial least squares (PLS) regression is a quantitative spectral decomposition technique that is closely related to PCA regression. The importance of PLS is that it is used to design and build robust calibration models for multivariate quantitative analysis. PLS actually uses the concentration information during the decomposition process. This causes spectra containing higher constituent concentrations to be weighted more heavily than those with lower concentrations. The main idea of PLS is to get as much concentration information as possible
into the first few loading vectors (number of component, factors, ranks or principal components). PLS regression consists of two fundamental steps. First, to transform the $X$ predictive matrix (spectra) of order $n \times p$ ($n =$ number of samples and $p =$ number of variables: cm$^{-1}$ or nm), in an matrix of components or latent variables uncorrelated, $T = (T_1, ..., T_p)$ of order $n \times p$, called PLS components, using the $Y$ response vector (concentrations) of order $n \times 1$; this contrasts with the principal component analysis in which the components are obtained using only the $X$ predictive matrix. Second, to calculate the estimated regression model using the $Y$ response original vector as predictive, PLS components. The dimensionality reduction can be applied directly on the components as they are orthogonal. The number of PC required for the regression analysis must be much smaller than the number of predictors. There is a number of ways of expressing these, a convenient one being (Eq.1 and 2) [29]:

$$X = TP + E$$  
\[ (1) \]

$$c = Tq + f$$  
\[ (2) \]

Figure 7 illustrates a simplified scheme for PLS: $X$ represents the experimental measurements (e.g. spectra) and $c$ (or $Y$) the concentrations. The first equation above appears similar to that of PCA, but the scores matrix also models the concentrations, and the vector $q$ has some analogy to a loadings vector. The matrix $T$ is common to both equations. $E$ is an error matrix for the $X$ block and $f$ an error vector for the $c$ block. The scores are orthogonal, but the loadings ($P$) are not orthogonal, unlike in PCA, and usually they are not normalized.

Chemometrics techniques have improved the last years in order to save time and computational resources in different models to be used without compromising the quality of results. In 2000 Norgaard and co-workers [32,33], developed different algorithms useful in Chemometrics field called interval partial least squares (iPLS) and this tool was presented for use on NIR spectral data. Recently, this new graphically oriented local modeling procedure has been implemented in many areas of research such as petrochemicals, pharmaceutical and beverage industry [34-36].
The principle of the iPLS is to optimize the predictive capability of PLS regression models and to support in interpretation. This algorithm which develops local PLS models on equidistant subintervals of the full-spectrum region. Its major objective is to provide an overall perspective of the significant information in different spectral subdivisions, thereby focusing on important spectral regions and removing interferences from other regions. The sensitivity of the PLS algorithm to noisy variables is highlighted by the informative iPLS plots [32]. For synergy interval PLS (siPLS), the basic principle of this algorithm is the same as iPLS first, it is to split the data set into a number of intervals (variable-wise), next, to develop PLS regression models for all possible combinations of two, three or four intervals. Thereafter, RMSECV is calculated for every combination of intervals. The combination of intervals with the lowest root mean square error of cross-validation (RMSECV) is selected.

Finally, cluster analysis is the name given to a set of techniques that seeks to determine the structural characteristics of multivariate data sets by dividing the data into groups, clusters, or hierarchies. For cluster analysis, each sample is treated as a point in an n-dimensional measurement space. The coordinate axes of this space are defined by the measurements used to characterize the samples. Cluster analysis assesses the similarity between samples by measuring the distances between the points in the measurement space. Samples that are similar will lie close to one another, whereas dissimilar samples are distant from each other [28].

In this chapter, remote Raman detection experiments were performed to quantify HEM such as PETN present in mixtures with non-HEM. The remote measurements were carried out at 10 m employing a frequency-doubled 532 nm Nd:YAG pulsed laser as excitation source. The quantification study was performed by using PLS, iPLS and siPLS as chemometrics tools to achieve the best correlation between the remote Raman signal and the concentration (%) of PETN explosive in a mixture with pharmaceutical compound. Discrimination of chemical warfare agent simulant (CWAS) TEP concealed within commercial beverage bottles using Optical Fiber Coupled Raman Spectroscopy with the use of different chemometrics techniques such as PLS, PLS-DA. Finally infrared spectroscopic information analysis using Chemometrics was designed and implemented in the detection of HEM: 2,4-DNT, TATP, PETN and RDX, present at trace level on surfaces and in air were analyzed by Chemometrics Enhanced Vibrational Spectroscopy.

4. Multivariate Detection and Quantification from Vibrational Spectra

4.1. Remote raman experiments

Different preprocessing methods such as vector normalization (VN), mean centering (MC), auto scaling (AS), multiple scattering correction (MSC), standard normal variate (SNV) and first and second derivatives have been developed to improve a good multivariate quantification. The 56 remote Raman spectra taken from PETN detection in mixes with APAP were randomly split into two groups: a first group with the 70% of the data for calibration and cross validation (training set) and a second group for external validation (test set) formed by the remaining 30% of the data. The quantitative model was performed
by using chemometrics tools, such as PLS, iPLS and siPLS. The PLS program used was from PLS-ToolBox™ (Eigenvector Research Inc.) for use with MatLab™. The iPLS and siPLS algorithms used in this work were carried out by employing iToolbox™, (downloaded from http://www.models.kvl.dk). The performance of the final PLS, iPLS and siPLS models were evaluated according to the root mean square error of cross-validation (RMSECV), a leave-one-sample-out cross-validation method and the predictive ability of the models were assessed by the root mean square error of prediction (RMSEP) and the correlation coefficient (R). In general for all the PLS models RMSECV were calculated as follows:

\[
RMSECV = \sqrt{\frac{\sum_{i=1}^{n_{cal}}(c_i - \bar{c}_i)^2}{n_{cal}}}
\]  
(3)

Where \(c_i\) and \(c_p\) are the experimental and predicted concentration, respectively, of the ith calibration sample when situated in a left out segment, \(n_{cal}\) is the number of calibration samples in the training set. The number of PLS components included in the model is selected according to the lowest RMSECV. This procedure is repeated for each of the preprocessed spectra. For the test set, the root mean square error of prediction (RMSEP) is calculated as follows:

\[
RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{test}}(c_i - \bar{c}_i)^2}{n_{test}}}
\]  
(4)

The best model with the overall lowest RMSECV will be selected as final model. Correlation coefficients between the predicted and the true concentration are calculated for both the calibration and the test set, which are calculated as follows from Equation 5, where \(\bar{c}_i\) is the mean of the experimental measurement results for all samples in the train and test sets.

\[
R = \sqrt{1 - \frac{\sum_{i=1}^{n_{cal}}(c_i - \bar{c}_i)^2}{\sum_{i=1}^{n_{cal}}(c_i - \bar{c})^2}}
\]  
(5)

The implementation of new methodologies for enhanced detection of hazardous compounds such as explosives is always attractive for many countries principally for defense and security applications. Terrorist employ different ways to pose threats and make illegal acts against military and civilian people. According to this situation our study is focused on detection of explosives present in mixture prepared intentionally with a pharmaceutical product by employing remote Raman detection and chemometrics tools. Remote Raman spectra of PETN, APAP in mixtures of them are illustrated in Figure 8. The results show that mean centering (MC) pre-processing method was the most successful method for correcting background and was selected for construction of further models because they presented small improvement in RMSEC.

The full spectrum was split in 20 independent intervals and the RMSECV values for PLS models constructed with different intervals is shown in Figure 9. Models with no intervals were better than PLS models with all variables (dotted in line) and the intervals 6 (1185.2-1328.9 cm\(^{-1}\)), 9 (1619.8-1755.4 cm\(^{-1}\)), and 19 (2878-2988.4 cm\(^{-1}\)), presented the lowest RMSECV values where more variability exists. These values are shown in Table 1. The number of
latent variables required for the models obtained using different intervals is the numbers shown inside the rectangles.

**Figure 8.** Remote Raman spectra of PETN, APAP and mixture of them, collected at 10 m of target to collector distance employing 532 nm laser with 100 pulses of 25 mJ/pulse.

**Figure 9.** RMSECV values for PLS models obtained for the 20 different intervals (bars) used in iPLS models. The horizontal dotted line represents the RMSECV value for the PLS model with all variables. Numbers inside the rectangles are the optimal number of latent variables.
Multivariate Analysis in Vibrational Spectroscopy of Highly Energetic Materials and Chemical Warfare Agents Simulants

Table 1. Full-cross-validated PLS, iPLS, and siPLS models for prediction of PETN in the range 1.0–34.0% Remote Raman spectra. All models are based on MC data.

In synergy interval-PLS (siPLS) model calibration, the number of intervals was also optimized according RMSECV values. Table 1 shows the results of siPLS model calibration when the spectra were split into different number of intervals. The optimum siPLS model was obtained with the combination of 3 intervals (3, 9 and 19) and 7 PLS components. The lowest RMSECV was 1.4, compared with RMSECV values obtained for PLS model with all variables and iPLS models. According to the statistical results illustrated in Table 1, it is important to establish that iPLS or siPLS models with 4 or more intervals (data not shown) including intervals 10-18 were explored. These intervals correspond to noisy areas which were not eliminated in order that the models could choose the spectral region of larger variability. The capability of prediction of siPLS models was better when compared to the other models. As shown in the correlation plots in Figure 10, there is a good relationship between the True and Predicted concentration (%) for PETN, with RCV values of 0.993. This can also be appreciated by the good prediction of the test set of samples with values of RMSEP of 1.8% for the corresponding explosives. The final model separated the vibrational spectra into 20 intervals, 7 latent variables were used and the intervals number 3, 9 and 19 were combined. The selected intervals included regions of 724.2 - 876.7 cm⁻¹, 1619.8-1755.4 cm⁻¹ and 2878 - 2988.4 cm⁻¹. The first Raman shift region correspond to NO₂ scissoring mode and O-N stretching band; the second region is relevant for NO₂ asymmetric stretching mode and C=O stretching band; the third region represents the C-H stretching mode [37-39].

![Figure 10. Predicted vs. True PETN concentration for siPLS model using 3, 9, and 19, intervals and 7 latent variables.](image-url)
4.2. Optical fiber probe raman spectroscopy experiments

In the optical fibers coupled Raman spectrum of TEP, shown in Figure 11, the CWAS has characteristics peaks at 733 cm\(^{-1}\) (PO\(_3\) symmetric stretching mode), 813 cm\(^{-1}\) (PO\(_3\) asymmetric stretch), 1032 and 1098 cm\(^{-1}\) (C–O stretch), 1162 cm\(^{-1}\) (CH\(_3\) rocking) and 1279 cm\(^{-1}\) (P–O symmetric stretch) [6]. Mixtures of TEP with commercial liquids were measured in their corresponding commercials bottles. TEP concentration varied from 0 to 100 (%v/v). In Figure 12, TEP Raman spectra are shown for different bottle materials. At all concentrations, the TEP characteristic peaks could be distinguished within the different types of materials of the container with the exception of brown glass and white plastic. These two bottle materials had lower transmittance in the 200 to 1400 cm\(^{-1}\) region and TEP characteristic peaks in the 2700 to 3200 cm\(^{-1}\) region. UV-VIS spectra (data not shown) show the increased absorbance in bottle materials such as white plastic and amber glass (Malta\textsuperscript{TM}). This confirms nature of the low intensity Raman peaks in the region (200-1400 cm\(^{-1}\)) shown in Figure 12. When light scatters turbid materials, such as amber glass or white plastic, the material is absorbing or blocking the light when compared to clear glass and clear plastic. Thickness of the bottle material and coloration also play a role in absorbance and transmission. The high intensity peak at 2300 cm\(^{-1}\) corresponds to the background light (mercury vapor from fluorescent lamps). This peak is shown with higher intensity in Raman spectra of brown glass and white plastic in comparison to the rest of spectra due to the increase in integration time for these two bottle materials. All bottle materials were subject to background light in order to simulate real-time conditions found in military, airport and other environments where a light source is involved. This analysis is based on increased absorbance shown in the UV-VIS Spectra for different bottle materials (data not shown).

![Figure 11](image_url)  
**Figure 11.** Raman vibrational spectrum of TEP excited at 488 nm.

Calibration models were performed with PLS regression model to distinguish between the samples that contain TEP in aqueous solutions compared to the solutions with TEP and the commercial product. In Figure 13, eight PLS regression models are chosen in order to show the marked difference between the best and the worst regression model, each of these with
and without pre-processing steps. Since integration times were normalized for each bottle, Limits of Detection were similar with the exception of amber/brown glass (Malta™). The aqueous solutions show a better $R^2$ values than the mixtures with the commercial product. For clear glass (Snapple™), the $R^2$ value is 0.9925 for aqueous solutions compared to 0.9747 for the mixtures with the commercial product. The $R^2$ values for Malta™ in aqueous solutions showed a significant increase with optimization (0.4193 without preprocessing and 0.9508 with optimization). However, optimization with Malta™ shows a lower $R^2$ value 0.7646 compared to 0.8047 without preprocessing.

Figure 12. Raman spectra of Triethyl Phosphate in various types of bottle materials

It is clear that the $R^2$ values increase in PLS regression models for aqueous solutions since water does not present strong signatures in Raman Spectroscopy. Every other PLS regression model (green plastic, green glass, clear plastic, clear glass, and white plastic) in aqueous and beverage solution presented nearly similar limits of detection. Each of these limits improved with their respective preprocessing step (vector normalization, standard normal variate and mean centering). With the help of integration time for each bottle material, normalization was achieved with the limits of detection and root-mean-squared error cross-validation (RMSECV). These values were found as acceptable in an average between the best models of approximately 2.5%. The Limits of Detection for PLS methods were estimated using the equation 6 [40]:

$$\text{Limits of Detection} = \frac{\text{RMSECV}}{\sqrt{n}}$$
Root mean squared error of calibration (RMSEC) was obtained from the square fit errors $[(c_{\text{predicted}} - c_{\text{true}})^2]^{1/2}$ where the sum extends to all samples of the calibration set. The degrees of freedom were then calculated as $\nu = n - F - 1$ where $F$ is the number of latent variables and $n$ is the number of samples in the set. The distance of the predicted sample from zero concentration to the calibration set’s mean is the leverage $h_0$. Ultimately, $\Delta(\alpha,\beta,\nu)$ corresponds to a statistical parameter that notices the $\alpha$ and $\beta$ probabilities of falsely stating presence/absence of the chemical warfare agent stimulant. Since $\nu \geq 25$, we used $\alpha = 3.4$ for the LOD. LOQ values as per Eq. 7 were studied at a concentration with a Relative Standard Deviation (RSD) of 15% as stated by Felipe-Sotelo et al. [40]:

$$LOQ = 100 \times \text{RMSEC} \times (1 + h_0)^{0.5} \times \text{RSD}$$  \hspace{1cm} (7)
TEP with commercial products and with less data (5, 30, 70 and 0 %v/v) due to limited time. Snapple has lower limits of detection which is favorable for detection of chemical warfare stimulants in commercial bottles made out of various materials. When comparing limits of detection for aqueous solutions versus solutions with commercial beverage product inside commercial bottles, limits of detection are considerably lower. R² prediction values were higher in aqueous solutions since water does not present significant Raman signal. Limits of detection were found as low as 1 percent for white plastic. Optimization also improves or lowers the limits of detection as shown in Figure 13.

Table 2 shows Limits of Detection and Quantification (LOD and LOQ respectively) for various commercial beverage bottle solutions with TEP for the best models. Preprocessing options include Vector Normalization (V.N.), No preprocessing (N/A), Mean Centering (M.C.), Constant Offset Normalization (C.O.N.), First Derivative (F.D.) and Multiplicative Scatter Correction (M.S.C). Higher limits of detection and quantification for amber glass and clear plastic were presented due to their dark coloration in bottle material (amber) and commercial beverage product (Pepsi and Malta). An unexpected low value for limits of detection and quantification for white plastic was observed. This may be due to the low amount of trials (5 instead of 10 for 5, 30, 50 and 70 (%v/v of TEP) as was done with other bottle materials due to the high integration times for this material. Even though TEP, a surfactant agent, did not present a homogeneous solution with milk, integration times were normalized in order to obtain a better model of a clear linear regression with an R² value of 0.9987 and excellent limits of detection of 0.01(1%).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LOD (%)</td>
<td>3</td>
<td>1</td>
<td>26</td>
<td>4</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>LOQ (%)</td>
<td>8</td>
<td>3</td>
<td>77</td>
<td>11</td>
<td>66</td>
<td>9</td>
</tr>
<tr>
<td>Preprocess</td>
<td>V.N.</td>
<td>V.N.</td>
<td>N/A</td>
<td>M.C. + V.N.</td>
<td>C.O.N.</td>
<td>M.S.C. + M.C.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AQUEOUS MIXTURES</th>
<th>LOD (%)</th>
<th>LOQ (%)</th>
<th>Preprocess</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11</td>
<td>33</td>
<td>V.N.</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>48</td>
<td>F.D. + V.N.</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>22</td>
<td>V.N.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25</td>
<td>V.N.</td>
</tr>
</tbody>
</table>

Table 2. Limits of Detection and Quantification for the PLS models of TEP in commercial beverage bottles and aqueous mixtures along with their respective preprocessing methods.

4.3. Gas phase infrared spectroscopy experiments

Multivariate calibration methods such as Partial Least Squares (PLS) models can be formulated as a regression equation [41, 42]. The equation in metrical form is \( Y = XB \), where
**B** is computed as $B = W(PTW)^{-1} QT$ and $W$ is the matrix of weights of $X$, $Q$ is the loadings matrix of $Y$, and $P$ is the $X$ loadings matrix. In this study, the $Y$ matrix represents the dependent variables but this is changed from continuous to discrete variation, and contains information about different classes of objects [43-45], it is a simple two states function: 1 represents the condition for the presence of explosive in the sample and 0 stands for the absence of explosive in the sample analyzed. By these means it will be possible to decide if an explosive substance is present or not in a sample. The values originating from the analysis: wavenumber range or parts of spectra are the independent variables ($X$ matrix). In this study the loading vectors or number of component ($B$ matrix) were used for independent variables in the DA.

TATP and 2,4-DNT in air were detected using FTIR spectroscopy. At trace levels, the vibrational signatures are not easily perceptible. Vibrational signatures of explosive can be confused with vibrations arising from the background air components. Thus the first task was to determine the possible interference of the two spectra. Figures 14a and 14b show the spectra of flowing gas that contains TATP and 2,4-DNT traces. The characteristic infrared signals of TAPT at 1200 cm$^{-1}$ and at 1550 cm$^{-1}$ for 2,4-DNT can be observed in Figure 14 which confirms the presence of these compounds in air. Linear Combination Analysis in the form of Partial Least Squares (PLS) was calculated for all FTIR spectra (7500-600 cm$^{-1}$). Two and four vectors were required to find the perturbation produced by TATP and 2,4-DNT respectively, on the normal flowing air IR spectrum. The discriminating function used was a two position switch type function: On – Off (Yes/No). The nomenclature in the DA was for classification of samples in terms of “Disc-1=TATP present” or “Disc-0 = TATP not present” in air, for TATP; and “Disc-1 = 2,4-DNT present” or “Disc-0 = 2,4-DNT not present” in air, for 2,4-DNT. The results were presented in the form of histogram, where the y-axis is the frequency and x-axis is the discrimination function. Also the prediction of new sample was present in this form, (Figures 15 and 16) in these graphs, the improvement of models, when vectors are added successively is observed.

Figure 14. FTIR vibrational spectra of gas explosive in air: a. TATP and b. 2,4-DNT traces.

The best discriminant function was selected based on statistical significance ($p$) and the percentage of cases correctly classified (PCCC) [46]. The validation was done by internal jackknifing validation and external validation. Internal validation: in this method, each spectrum was successively removed from the data set, and then it was discriminated from a
new model built from the remaining spectra. This procedure was done for each one of the spectra in the data set, and the predicted discriminations were then compared with the experimental observations. The generated percentage of cases correctly classified is called the cross-percentage of cases correctly classified (PCCCC). External validation: before making the model, 100 spectra of air, 100 spectra air with TATP and 50 spectra air with 2,4-DNT were taken from the data set randomly. These spectra were analyzed by the validation model.

![Model with 1 vector](image1.png)
![Prediction of new sample with model of 1 vector](image2.png)

![Model with 2 vectors](image3.png)
![Prediction of new sample with model of 2 vectors](image4.png)

**Figure 15.** Histogram for discrimination models of TATP and external validation.

For the PCA analysis of TATP in air spectra were recorded using the EM-27 and the LaserScope™ instruments. A total of 60 spectra were recorded from clean air and 120 spectra from air with TATP present using EM-27 and 35 spectra were recorded from clean air and 30 spectra from TATP present in air using LaserScope™. All PCA analysis including any preprocessing in the spectral data were run using PLS-Toolbox software. PCA runs were made with the raw data and using different preprocessing treatments. The preprocessing treatments used were: auto scale, smoothing, SNV-standard normal variation, Mean center, auto scale + 1st derivate, auto scale + 2nd derivate, mean center + 1st derivate, mean center + 2nd derivative, MSC-multiplicative scattering correction. The algorithm used to carryout smoothing and derivatives was that of Savitzky-Golay (every 11, 17, 21 and 31 points). During the PCA runs it was not necessary to eliminate spectral data.

The infrared data from clean air and air with TATP were run together in the PCA model for each instrument used. Figure 17 shows the Scores plots for the PCA obtained. Figure 17a shows the first two principal components from spectral data using the EM-27 FTIR spectrometer with a Globar source Figure 17b shows the first two principal component analyses from spectral data of TATP detection from LaserScope™ spectrometer using
Quantum Cascade Laser Source. The best results achieved for both PCA models (illustrated in Figures 17a and 17b) were using raw data. Both results allowed classifying gas phase TATP explosive from clean air. In Figure 17 can be noticed that PC1 tends to relate the differences between the IR dataset two.

**Figure 16.** Histogram for discrimination models of 2,4-DNT and external validation.
Multivariate Analysis in Vibrational Spectroscopy of Highly Energetic Materials and Chemical Warfare Agents Simulants

Figure 17. Score plots for the PCA, presented as (a) PC2 vs. PC1 for TATP detection from EM-27 FTIR spectrometer using Globar source and (b) PC2 vs. PC1 for TATP detection from LaserScope™ spectrometer using QCL source.

Other hand, the loadings plot were analyzed too to support the results from PCA with the finality of knowing which spectral signals cause differences between the dataset. Figure 18 shows the PC1 loading from Figure 17b, in this it can be seen that the spectral features are equal to infrared vibrational signal of reference TATP. Some signal recording can be tentatively assigned according to B. Brauer and J. Oxley as [47,48]: 891.8 cm$^{-1}$ to O–C–O and Me–C–Me sym str, and Me–C–O asym str; 946 cm$^{-1}$ to C-O str; 1197.6 cm$^{-1}$ to O–C–O and Me–C–Me asym str, Me–C–O sym str; 1205 cm$^{-1}$ to O–C–O and Me–C–Me sym str and finally 1234 cm$^{-1}$ to C-C str.

Figure 18. Figure 18. Loading plot for PC1 from PCA for TATP detection from LaserScope™ spectrometer using QCL source.

4.4. Quantum cascade laser based ir reflectance experiments

OPUS 6.0 Software (Bruker Optics, Billerica, MA, USA) was used to analyze the data obtained. Four spectra were obtained for each sample. The spectra were carried out using as backgrounds: substrate without explosive. PLS was applied to the data using different preprocessing treatments: raw data, auto scale. Mean center, auto scale + 1st derivative, auto scale + 2nd derivative, mean center + 1st derivative, mean center + 2nd derivative. The
spectral range was 1000-1600 cm\(^{-1}\). PLS shown below was that best results obtained. Figure 19 shows PLS plots of RDX deposited on TB. The best result was achieved using the spectral region of 1000-160 cm\(^{-1}\) and using mean centering as preprocessing. A total of 10 latent variables or factors were necessary to obtain a \(R^2\) and RMSECV equal to 0.9915 and 2.32 \(\mu\)g/cm\(^2\), respectively.

Figure 19. PLS of RDX on travel baggage (TB) as substrate.

Figure 20 shows PLS of PETN deposited on TB. The best resulted was achieved using the spectral region of 1000-1600 cm\(^{-1}\) and using mean centering as preprocessing treatment. A total of 10 latent variables or factors were necessary to obtain a \(R^2\) and RMSECV equal to 0.9994 and 1.82 \(\mu\)g/cm\(^2\), respectively.

Figure 20. PLS of PETN on Travel Baggage as substrate.
5. Conclusion

Raman and infrared vibrational techniques were used for the detection of highly energetic materials and chemical warfare agents simulants in different matrices such as pharmaceutical mix, commercials bottles and travel baggage. The analysis of the spectral data allows emphasizing certain results. Satisfactory results were found for the quantification of explosives with good values of $R^2_{cv}$, RMSECV. Reliable predictions obtained by remote sensing based on Raman spectroscopy at remote distance of 10 m employing 532 nm laser as excitation source. Remote Raman system using the appropriate chemometrics tools such as PLS, iPLS and siPLS promises to be a reliable technique for finding the existence of highly energetic material such as PETN deliberately hidden in matrices with similar chemical structures.

Partial Least Squares (PLS) calibration models reported limits of detection very low for white plastic in commercial beverage bottle solutions which was the best model. Due to the bottle material and commercial beverage product coloration, Malta was the worst model with reported limits of detection more elevated. Limits of Detection and Quantification for commercial bottles were compared in aqueous and mixtures. It is observed that limits of Detection were significantly lower for mixtures of TEP with the commercial product. Integration times were the same for both aqueous and commercial beverage bottle solutions (each normalized with respect to bottle material, color and thickness). Water does not transmit significant Raman signal, which would make limits of detection lower for aqueous solutions. However, commercial beverage bottles mixtures showed lower limits of detection than aqueous solution since the beverage solutions inside each bottle showed significant Raman signal and, therefore, increasing CWAS presence in the spectra.

PLS-DA model and discriminant analysis was done to detect TATP and 2,4-DNT traces in fluid air. The region of 600 to 7500 cm$^{-1}$ was highly significant in the discrimination with $p < 0.00001$ and 100 % discrimination for two vectors for TATP and four vectors for 2,4-DNT. These results show the ability of the Chemometrics methods to discriminate between vapor phase explosive (2,4-DNT) and air.

Results obtained from principal component analysis to determine the presence of peroxides explosives such as TATP when they are in gas phase mixed with air shown be useful for distinction between TATP vapors and air. The principal component analysis from infrared spectral data used little PC for predict the variability of the spectral data, being the first two PC more important. PC1 loadings confirm the results from the PCA because it contained features from TATP spectrum. Other hand, the PLS model were shown chemometrics tool for quantify explosive such as RDX and PETN on substrate of the real world such as travel baggage.

In general, vibrational spectroscopy systems designed based on this work should be useful for National Defense and Security applications, for screening hazardous liquids in government installations, seaports and in public installations to improve defense against terrorist attacks.
Author details


ALERT DHS Center of Excellence, Department of Chemistry
University of Puerto Rico-Mayagüez, Mayagüez, PR

Acknowledgement

Support from the U.S. Department of Homeland Security under Award Number 2008-ST-061-ED0001 is also acknowledged. However, the views and conclusions contained in this document are those of the authors and should not be considered a representation of the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

This contribution was supported by the U.S. Department of Defense, Proposal Number: 58949-PH-REP, Agreement Number: W911NF-11-1-0152. The authors also acknowledge contributions from Dr. Richard T. Hammond from Army Research Office, DOD.

6. References


[34] Delfino I., Camerlingo C., Portaccio M., Della Ventura B., Mita L., Mita D.G., Lepore M. Visible micro-Raman spectroscopy for determining glucose content in beverage industry. Food Chemistry 2011;127(2) 735–742.


[40] Felipe M., Cal 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(1) 61-68.


