Identification of mixed substances using a random forest regressor to classify THz absorbance spectra

Arthur D. van Rheenen, Lars Aurdal, Helle Emilia Nystad, Magnus W. Haakestad

Arthur D. van Rheenen, Lars Aurdal, Helle Emilia Nystad, Magnus W. Haakestad, "Identification of mixed substances using a random forest regressor to classify THz absorbance spectra," Proc. SPIE 10800, Millimetre Wave and Terahertz Sensors and Technology XI, 1080009 (5 October 2018); doi: 10.1117/12.2325529

SPIE.

Event: SPIE Security + Defence, 2018, Berlin, Germany

Identification of mixed substances using a random forest regressor to classify THz absorbance spectra

Arthur D. van Rheenen[1*], Lars Aurdal[1], Helle Emilia Nystad[2], and Magnus W. Haakestad[1]
[1]Norwegian Defence Research Establishment (FFI), P. O. Box 25, NO-2027 Kjeller, Norway
[2]Haukeland University Hospital, Dept. of Clinical Engineering, P. O. Box 1400, NO-5021 Bergen, Norway
[*]arthur-d.vanrheenen@ffi.no

ABSTRACT

We report on the development and application of a random forest regressor that not only identifies but also estimates the relative concentrations of substances (one explosive and two simulants), both in one-substance and two-substance samples. Performance of the regressor is quantified using Receiver Operating Characteristics and the performance is contrasted with that of a simple Spectral Angle Mapping technique that worked well on single-substance samples [1-3].

Keywords: THz (differential) spectroscopy, spectral angle mapping, random forest regressor, receiver-operating characteristics

1. INTRODUCTION

Detection of transported illegal substances such as explosives and drugs remains a challenge for many customs authorities and a number of tools have been developed to counter the trafficking of such goods. Possibly, a combination of technologies needs to be used to improve the rate of success in such an endeavor. THz-technology has been shown, at least in the laboratory, to be able to detect a large number of chemical compounds.

Common for a large number of explosives and drugs is that they have spectral fingerprints in the 0.1 – 10 THz range, which many of the THz sources and detectors cover. An added advantage of the technology in these types of applications is the transparency of many common packaging materials, such as plastic, cardboard, and cloth, in this frequency range. This makes it possible to detect and identify concealed materials. The development of an imaging capability makes it possible to visualize hidden objects, shapes with different optical properties than their background, helping to identify suspicious objects, such as weapons or containers. By combining the spectroscopic and the imaging capability, detection of suspicious objects and subsequent substance identification seems in reach. However, there are a number of obstacles that have to be overcome before a reliable "identifier" is realized. Some of these difficulties, which are of course widely known, are listed in [1].

There is a vast body of literature [4-18] on identifying substances by comparing their spectra. Spectral features have been measured in numerous wavelength bands corresponding to the energies of the transitions of interest. The purpose of this work is to compare different schemes for comparing spectra and to find some objective measure to rank them. Equally important, we are interested in finding what the limitations are of the different schemes.

A general detection/identification scheme consists of transforming the measured raw data, a time-domain THz signal in our case, into a spectral characteristic, for instance the absorbance spectrum. This spectral characteristic has then to be compared, in some way, to known spectral characteristics. This comparison requires a measure of similarity or distance and a threshold so that a match or not-a-match may be declared. The performance of the detection/identification scheme is quantified by studying the Receiver Operating Characteristics (ROC) of the scheme: false positive rates and true positive rates are found as the threshold for declaration of detection/identification is swept from a minimum to a maximum value, zero to one, for instance.

To limit the scope of this work we consider only one spectral characteristic, the absorbance spectrum, and we did not investigate the effect of smoothing on the detection/identification performance of the schemes.

Millimetre Wave and Terahertz Sensors and Technology XI, edited by Neil A. Salmon, Frank Gumbmann, Proc. of SPIE Vol. 10800, 1080009 · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2325529

After describing the SAM and RFR schemes we briefly outline the specifics of ROC's, followed by a presentation of the THz transmission measurement set-up, samples used and procedure employed. Next, we describe the analysis we performed and present the results we obtained. These results are reiterated and discussed in the Summary section.

2. SPECTRAL ANGLE MAPPING (SAM)

The simplest and most intuitive method for comparing spectra is SAM. A measured spectrum is viewed as a vector and this vector's dot-product with library vectors (spectra) is calculated and then normalized by the lengths of the two vectors. In fact, one calculates the cosine of the angle between the vectors. Identical vectors point in the same direction, the angle between them is zero and the cosine equals 1. When the correlation between two vectors is small, their angle will be quite different from zero and the cosine significantly less than 1. The cosine of the angle is a direct correlation measure. When a threshold is defined then all correlations larger than the threshold will be declared a match and all others not-a-match. The only "intervention" is the choice of the threshold value.

3. RANDOM FOREST REGRESSOR (RFR)

RFR is a form of decision tree learning that produces a data classifier. By interrogating the data with questions that may be answered with a simple yes or no, the data is partitioned into finer and finer branches and ultimately into leaves, the classes. In a well-known example, one could ask: What is the chance of a particular passenger to survive the Titanic disaster? Is it possible to generate a list of successive queries to be asked of the data that will result in a correct classification: survivor or not? For instance, one could first divide the passengers into female and male, next divide them into over or under 20 years of age, then whether they travelled with family or not, and so on. In this example the solution is known: all the passengers on the Titanic are accounted for and both survivors and non-survivors can be grouped accordingly. The data may be used to train the classifier and then it could be used to predict the survival rates of future ship disasters. Rather than using a single decision tree one could consider using an ensemble of trees (a forest), each operating on a subset of the original input data, and then use a simple voting procedure to combine the results from the individual trees. In the case the predicted data are continuous (real) numbers, rather than discrete values, the classifier is called a regressor. A good introduction to the subject matter is Ref. 20.

The training set of The RFR (`RandomForestRegressor` function from the Python library `sklearn`, version 0.19.1) is based on spectral imaging of three samples: (i) 10% RDX in Teflon, 4-mm thick, (ii) 10% Lactose in Teflon, 4-mm thick, (iii) 10% Tartaric acid in Teflon, 4-mm thick. The 30-mm diameter samples were imaged in 2.3-mm steps resulting in about 110 spectra for each of the substances. This set of about 330 spectra is not sufficiently large to train the regressor, especially since the goal is not only identify but also quantify the relative content of active substances. We extended the training set with synthetic spectra. The synthetic spectra were generated by randomly selecting two numbers (weights) from the sequence 0.0, 0.1, 0.2, …, 1.0 and then randomly selecting two spectra from the set of measured data (three classes) augmented with no-known-substance spectra (4^{th} class) which are essentially low-pass-filtered random spectra. There are two rules: (i) the first two random numbers cannot be both equal to zero and (ii) the two spectra are taken from different classes. Synthetic spectra are formed by multiplying the first weight by the first spectrum, the second weight by the second spectrum, and summing the two terms. The resulting training set has 100.000 spectra.

The violin plot in Fig. 1 gives an impression of how well the RFR predicts the fraction of RDX present in the synthetic spectra. Plotted is the predicted fraction as a function of the actual fraction. The dashed line symbolizes the expected relationship and each symbol is formed as the distribution of predicted fractions

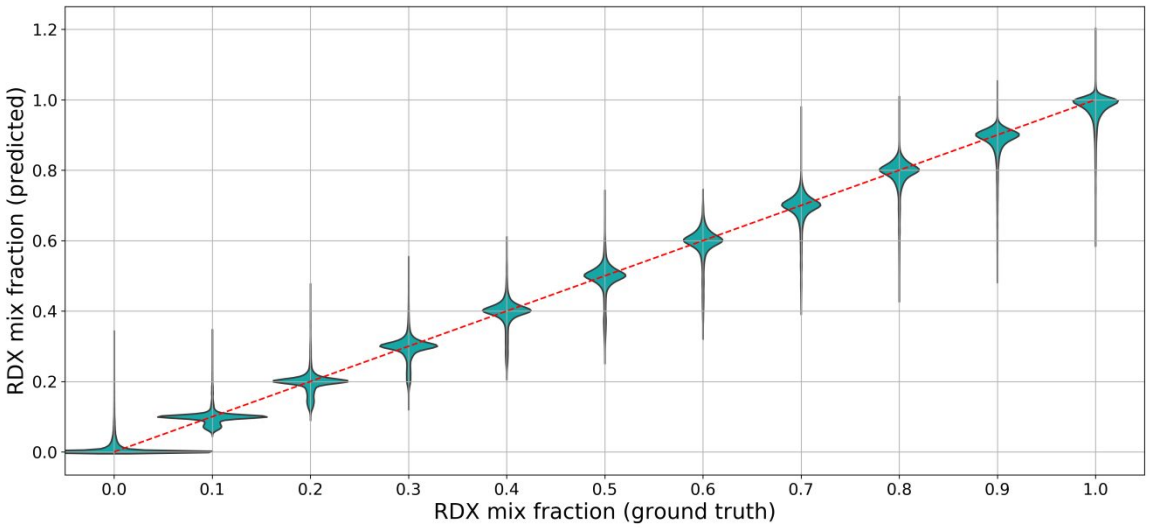

Figure 1. Relationship between predicted and actual fraction of RDX in the synthetic spectra. The symbols give an impression of the distribution of predicted values. The dashed line is the expected relationship.

4. RECEIVER OPERATOR CHRACTERISTICS (ROC)

ROC is a tool that is used to compare the performance of classifiers. A ROC is a plot of the true-positive rate (TPR) as a function of the false-positive rate (FPR) and is generated by sweeping the threshold value from the minimum to the maximum value and for each value counting the number of true and false positive matches that are found. For small threshold values, all data have larger correlation values than the threshold value and all true and false matches are detected, i. e. both the TPR and the FPR are 1. For maximum threshold value, all the data falls below the threshold and neither a false-positive nor a true positive is detected: TPR = FPR = 0. For intermediate values of the threshold, both true-positives and false-positive will be detected. For a good classifier the TPR > FPR and the ROC tends to bend towards the upper left-hand corner of the FPR-TPR space, where FPR = 0 and TPR = 1. The further the ROC is from the diagonal, the better the classifier is. This then provides a means to compare classifiers.

5. EXPERIMENT

The THz setup is based on a fiber-coupled time-domain spectroscopy system pumped by 100-fs pulses at 780 nm wavelength from a frequency-doubled Er-doped fiber laser [19]. THz images are acquired by mounting a sample holder on an x–y stage, which is scanned through the beam, with step size 5 mm for the training and reference data and 2 mm for the target data, while the transmitted THz waveform is captured. In this way a THz spectrum (after Fourier transform) is acquired for each stage position (pixel). A schematic of the setup is shown in Fig. 2. The distance between the emitter and detector modules is 31 cm and the sample holder has room for 3 x 3 sample pellets, with diameter 32 mm and thickness up to 4.2 mm. Fig. 1 (inset) shows the labeling of the sample positions. Teflon (25 µm average particle size) was used as a binder material, which was mixed with tartaric acid, lactose, or RDX and then pressed into pellets using a 2 ton press in two minutes. The top row of the sample holder (position 1–3) was used for reference measurements: a pure Teflon sample (4 mm thickness, position 1), no sample (position 2), and a metal plate (position 3). All measurements were performed in ambient air (21–26 ^{o}C, 10–50% relative humidity). The signal at each position of the x–y stage (pixel) was measured with a time window of 60 ps and a scan speed of 1 ps/s, with a sample rate of 32 Hz.

Measurements were taken on two sets of samples: one used to train the RFR and one to judge the performance of the RFR. For training the six remaining slots in the sample holder contained the samples listed in Table 1 (columns 2 and 3), whereas the samples used to generate the unknown target set are listed in the last two columns of Table 1.

Table 1. Samples used to generate training set, columns 2 and 3, and samples used to verify RFR performance.

	Samples for training set		Samples for verification	
Position	Active compound	Thickness (mm)	Active compound	Thickness (mm)
4	Tartaric acid 10% (unground)	4	RDX 5%, Lactose 5%	4
5	Tartaric acid 10%	1	Tartaric acid 5%, RDX 5%	4
6	Tartaric acid 5%	4	Lactose 5%, Tartaric acid 5%	4
7	RDX 10%	4	RDX 10%	4
8	Lactose 10%	4	Lactose 10%	4
9	Tartaric acid 10%	4	Tartaric acid 10%	4

Figure 3(left) shows reference spectra for an open beam (air) and blocked beam (noise). All spectra were calculated from the time-domain signals by calculating the Fourier transform (FFT). The reference spectra indicate a bandwidth of about 2.5 THz and a peak signal-to-noise ratio (SNR) of about 60 dB. Figure 3(right) shows the absorbance for samples containing RDX (10%, 3.5 mm thickness), tartaric acid (10%, 4.0 mm thickness), and lactose (10%, 4.2 mm thickness). These samples were used as reference samples in the spectral library. The part of the spectrum spanning the frequency range 0.1 to 1.5 THz was used in the correlation calculations, as the SNR is high in this frequency range (SNR > 40 dB for an open beam). Although there are several water vapor absorption lines in this wavelength range [2], we did not perform any numerical removal of water lines in the data processing. The location of the water lines was used to verify the calibration of the frequency axis in our measurements.

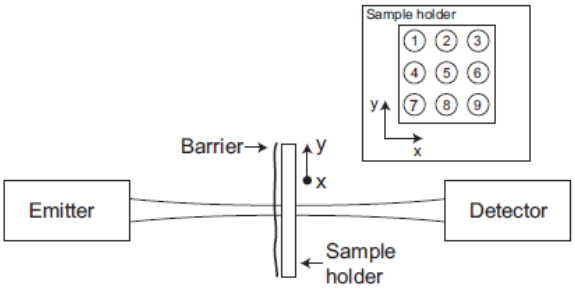

Figure 2. Experimental setup. Two fiber-coupled photoconductive antennas act as emitter and detector modules, which are separated by 31 cm. A sample holder, with transverse dimensions 15 x 15 cm, is scanned through the THz beam. Inset: Sample holder with labeled sample positions.

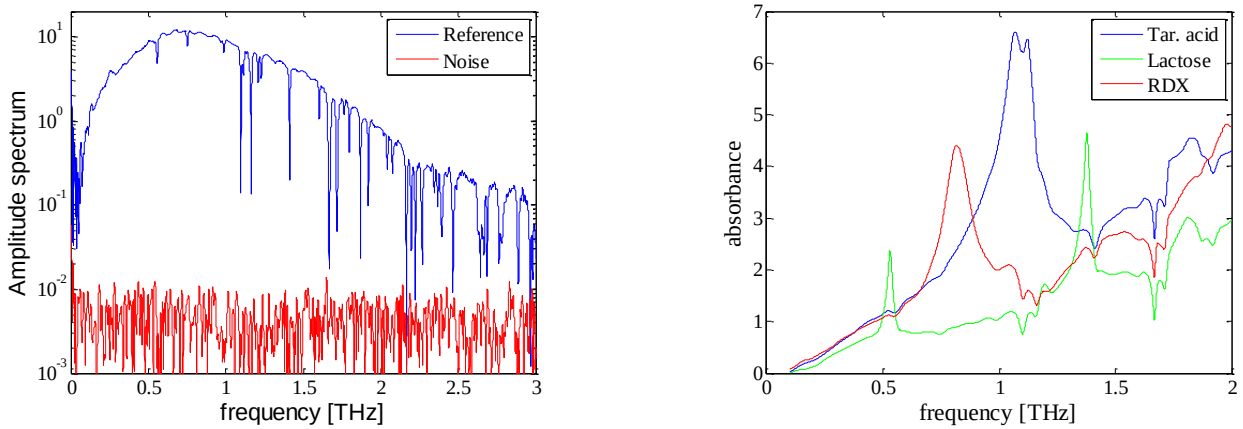

Figure 3. Left: Reference spectra of air and noise (blocked beam). The maximum power SNR > 60 dB and the many water absorption lines are clearly visible. Right: Absorbance of the three pure substances considered in this study: tartaric acid (blue), lactose (green), and RDX (red).

Spectra are not corrected for background scattering effects.

6. ANALYSIS RESULTS

6.1 Spectral Angle Mapping

As an example of the SAM analysis we show in Figure 4 (left) the correlation (SAM) of the target image spectra with the RDX reference spectrum. To find the location of the samples in the image we use the relative time delay of the THz pulse. Since the samples (4 mm thick) are significantly thinner than the sample holder (Teflon, 25 mm), the THz pulse through the samples is much less delayed than through the holder, providing a simple and error free method to separate sample spectra from other spectra.

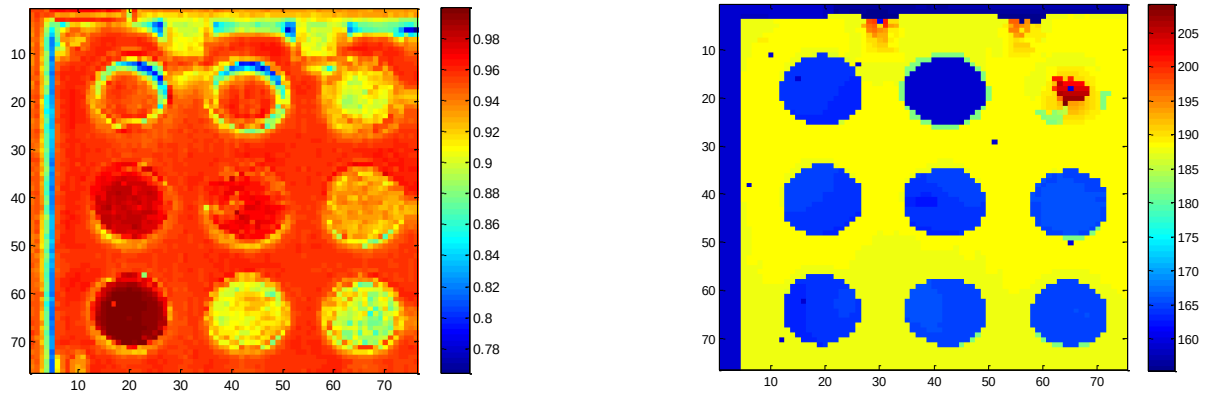

Figure 4. Left: Spectral correlation of target image with RDX. The color bar indicates the correlation scale. Right: Relative time delay of the peak of the THz pulse. The color bar indicates the delay in ps. The delay clearly separates the samples (blue) from the rest of the image and is used to identify the true sample pixels.

Looking at the left panel of Fig. 4 we observe that the best correlation is obtained for the pixels in sample position 7 (10% RDX) followed by the pixels in sample positions 4 (5% RDX + 5% Lactose) and 5 (5% RDX + 5% Tartaric acid). This result is of course anticipated: best correlation with the purest sample and less, but still significant, correlation with the mixed samples. We observe also that the contrast with the background (sample holder) is relatively small and it seems there is correlation also with spectra corresponding to sample position 6, which contains a mixture of Lactose and Tartaric acid only. Clearly this result points towards problems with high false alarm rates. Since we did not correct the spectra for background spectra, a significant part of the correlation comes from the background contribution. This is also the reason we experimented with the derivative of the spectrum in earlier work [3], yielding larger contrasts.

The correlation results for Lactose and tartaric acid are shown in the left and right panels of Fig. 5, respectively. Similar remarks may be made about those results.

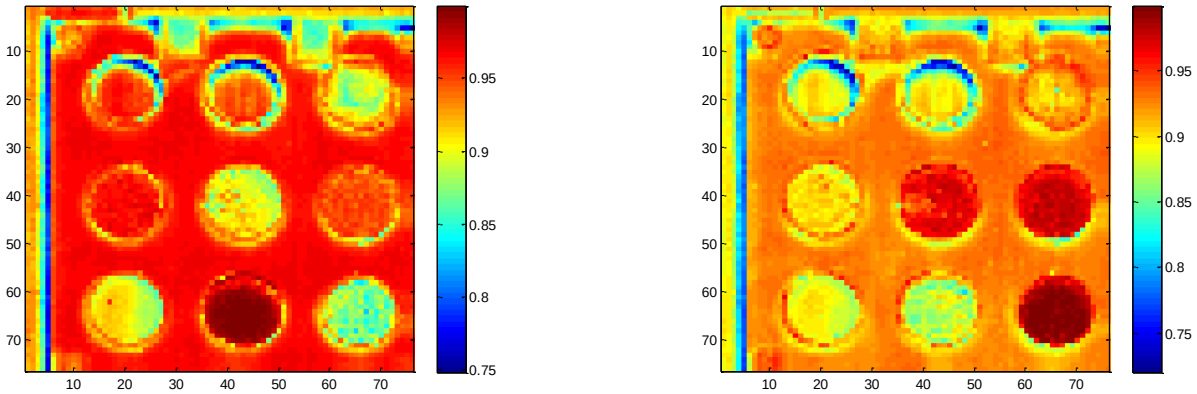

Figure 5. Spectral correlation of target image with Lactose (left) and Tartaric acid (right). The color bar indicates the correlation scale.

In an attempt to quantify these observations we plotted the average, over the true sample pixels, as well as the standard deviation, of the correlation between the six samples and the three compounds. The results are plotted in Fig. 6. For all three substances one can define a demarcation between samples that contain the target compound and samples that do not, however, the separation between the two groupings is not large. Another complication is the significant correlation of the compound spectra with the background (sample holder) spectra, whose average value is indicated by the solid lines in the panels

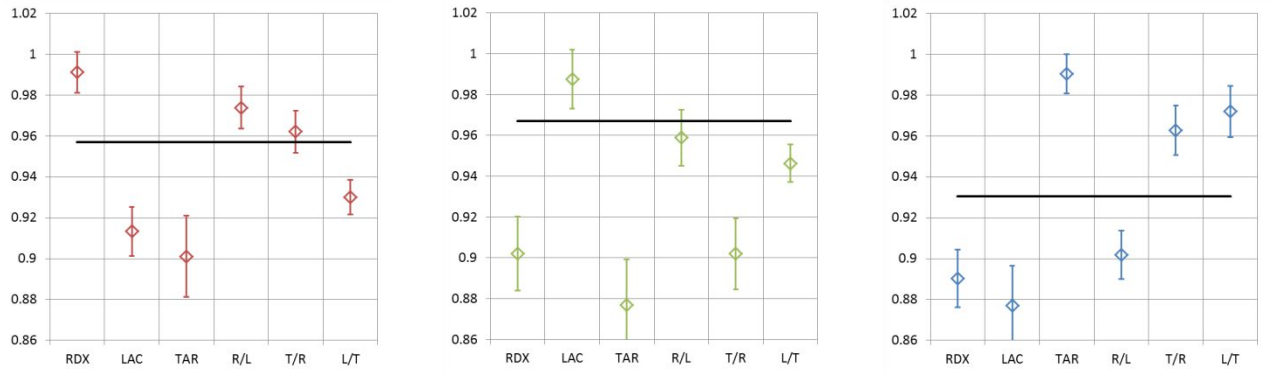

Figure 6. Average (over sample) correlation for each of the six samples in the holder with the three substances: RDX (left), Lactose (middle), and Tartaric acid (right). The line indicates the correlation of the background with the substance. R/L, T/R, and L/T refer to the mixed samples RDX/Lactose, Tartaric acid/RDX, and Lactose/Tartaric acid.

From an operational point of view, ROC's are of great interest because they describe the relationship between the true positive rates and the true negative rates of identification. For the true positive rates one considers only the pixels that "contain" the substance of interest and for the false positive rates one considers the pixels that do NOT "contain" the substance of interest. In the case of SAM, one then sweeps the threshold from -1 to +1 (the range of correlation values) in small steps and counts how many pixels have a correlation that is larger than the threshold value for these two pixel sets and then divide by the respective total number of pixels in each set.

As a starting point we first considered the bottom third of the image, "containing" only the unmixed samples. As the ROC in Fig. 7 (left) shows the curves tend to crowd into the upper left-hand corner, where the true positive rate is high and the false positive rate is low, the ideal situation. Especially RDX and Tartaric acid are easily detected correctly. Lactose detection seems more difficult.

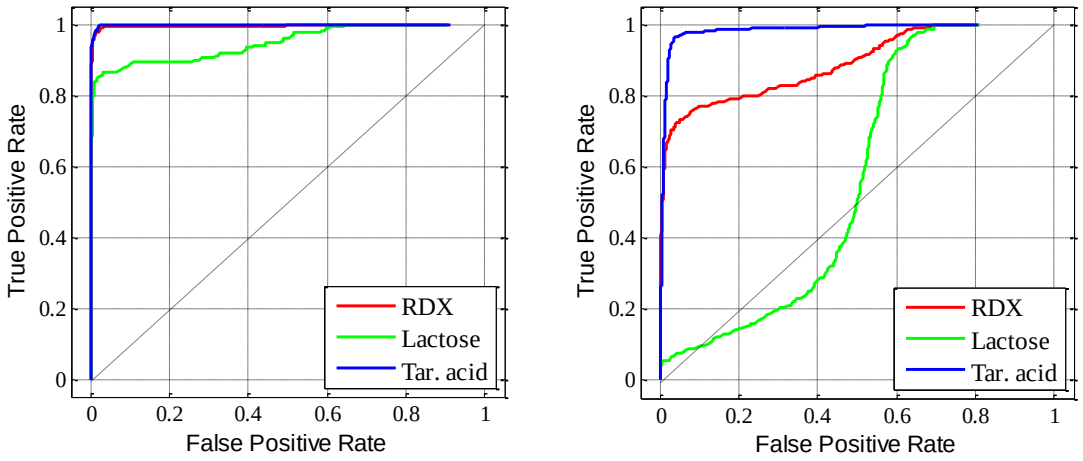

Figure 7. Receiver Operating Characteristics for recognition of RDX, Lactose, and Tartaric acid in (left) unmixed samples and (right) mixed samples obtained using SAM.

Next, we considered the middle third of the image, the one that "contains" the mixed sample only. The ROC for this part of the image is shown in Fig. 7 (right). Although Tartaric acid still can be detected easily without many false positives, the situation for the other two substances is worse, with reliable detection of Lactose practically impossible in the mixed samples considered here.

6.2 Random Forest Regressor

The result of applying the trained RFR to the unknown target image is shown in Fig. 8. The color in the image corresponds to the relative content of the active compound in the sample, full scale (red) corresponding to 14%. The unmixed samples (bottom row in each image) are easily identified and their relative content of the active compound, 10%, correctly indicated. The same holds true for the mixed samples, middle row, where the content is correctly estimated at about 5%.

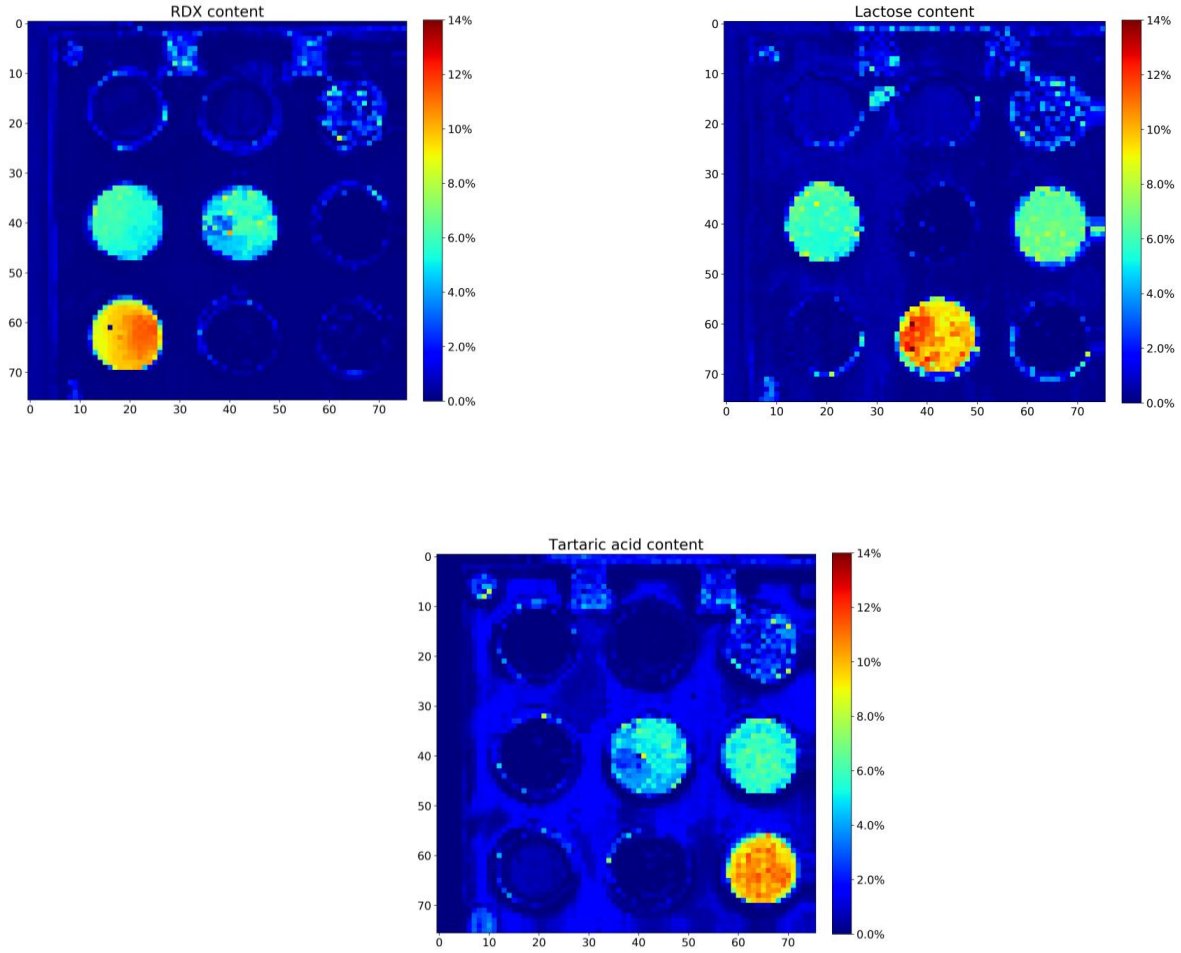

Figure 8. Estimation of the active-compound content in the sample. Full scale of the color bar corresponds to a content of 14%.

As for the SAM procedure we calculate the average value of the estimated active-compound fraction, as well as its standard deviation, for each of sets of sample pixels. These values are plotted in Fig. 9.

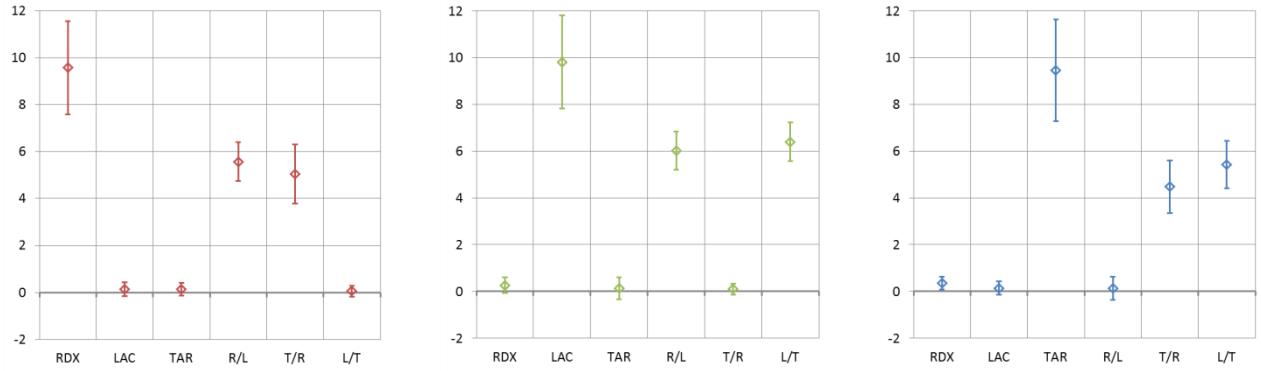

Figure 9. Average, over the sample pixels, predicted fraction (%) of active compound present in sample: left – RDX, middle – Lactose, right – Tartaric acid. R/L, T/R, and L/T refer to the mixed samples RDX/Lactose, Tartaric acid/RDX, and Lactose/Tartaric acid.

The unmixed samples contain 10% of the active compound, whereas the mixed samples contain 5% of the active compound. We observe that the RFR generally predicts the content within the margin of error for both RDX and Tartaric acid, but slightly over-predicts the Lactose content in the mixed samples. Note that our RFR predicts practically zero-content for samples that do not contain that compound. Comparing Fig. 9 with Fig. 5 shows that the RFR output has a much larger dynamic range, allowing for a clearer classification. This observation is reinforced by looking at the ROCs for this output.

In order to generate the ROCs we have to transform the data to make them fit the typical ROC mold. The predicted values are distributed around the target values: 5% for the two-substance samples and 10% for the one-substance samples. In principle the range of predicted values has no bounds, contrary to the SAM case where the range of correlation values is limited to [-1, +1]. In the RFR case we have two target values. By taking the absolute difference between predicted and target value and subtracting it from the data range has an upper bound, +1.

We cannot simply let a threshold value run from -1 to +1 in small steps, as we did in the SAM case. In principle the threshold range is not bounded. The predicted values for the unmixed samples are expected to lie around 10%, but some prediction may result in values larger than 10 %. In this case the distance between the predicted and actual fractions is of interest: an over-prediction by 1 %-point is equally “bad” as an under-prediction by 1%-point. We map the data by looking at the absolute value of the difference between predicted and actual value and subtract this from one. At least this gives an upper bound of +1 for the data. With a bit of trying we find a sufficiently low starting point for the threshold value sweep. The results are presented in Fig. 10, in a similar way to the data in Fig. 7.

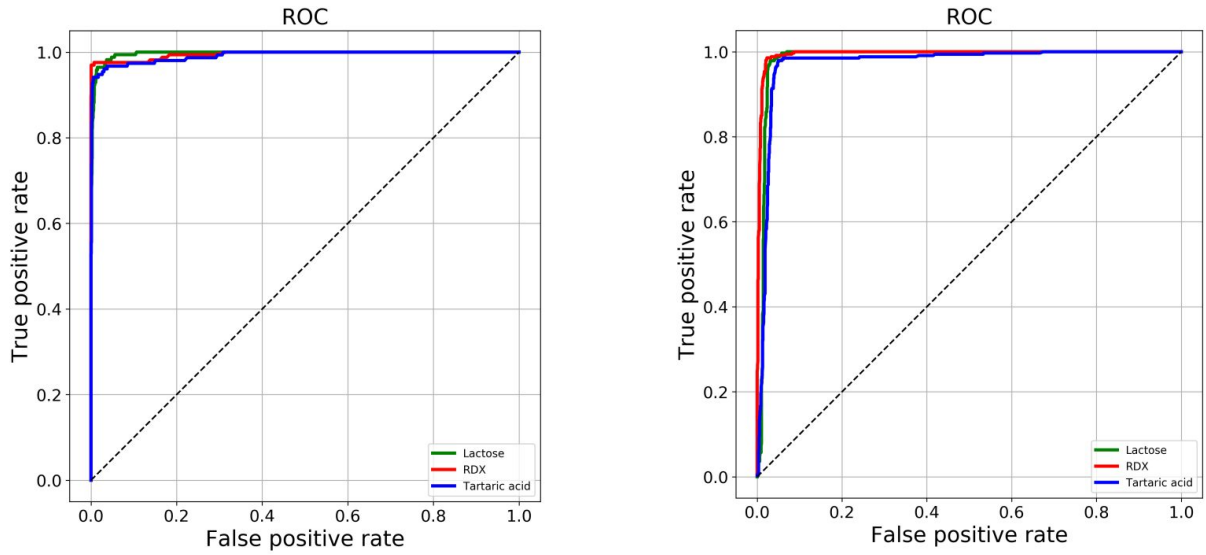

Figure 10. Receiver Operating Characteristics for recognition of RDX, Lactose, and Tartaric acid in (left) unmixed samples and (right) mixed samples obtained using RFR.

The characteristics in this case are more ideal than the ones obtained using SAM (compare with Fig. 7). Most significantly we do not observe the strong deterioration of the performance in regard to the quantification of the lactose content in mixed samples. With RFR we obtain a much more robust classification. In addition, the relative content of the substances in the samples may be estimated

7. SUMMARY

There are many approaches that could be useful to identify substances according to their measured spectra. In this study, we make use of the fact that a number of interesting substances have spectral features in the THz frequency domain. This technology may then be used to scan mail, parcels, and even human beings for possible illegal transportation of controlled substances. In security applications especially, but also in most other applications, false alarm rates must be balanced with throughput, sensitivity, and specificity. These requirements imply automated detection, with software identifying spectral matches and hence the sought-after substances. In this work we looked for a classifier that not only could identify the substances of interest but also estimate the relative amount of the substance that is present, even in mixed samples.

In previous work we compared the performance of SAM and PCA and investigated spectral smoothing strategies as well as which spectral characteristic to use. But that work was limited to samples that contained only one substance. Here we consider samples that also consist of mixes of two substances. A preliminary investigation using SAM on these samples showed that especially the detection of Lactose in the mixed samples proved difficult: high false alarm rates. Despite the obvious advantages of the simplicity of the approach, one needs only single copies of library spectra and the method is transparent, the poor results require a different approach.

Looking for a different approach we opted for a RFR, a technique which uses an ensemble of decision trees in which the data is sorted and binned in consecutive steps. The technique requires a data set for training purposes, a data set that could be very large. Fortunately, it is possible to train the regressor with synthetic data, consisting of mixes of a more limited set of measured spectra.

After training the performance of the regressor was validated by feeding it with spectra of both unmixed and mixed samples, new data that was not part of the training data. As we showed, not only were the substances identified, the relative substance contents were estimated with success. The improvement of the RFR over SAM is most clearly

demonstrated by the ROCs (Fig. 7 for SAM and Fig. 10 for RFR). Where false alarm rates, especially for detection of Lactose in mixed sample, caused problems for SAM, they were significantly reduced in the RFR approach.
Looking for a reason as to why SAM has problems identifying Lactose in the mixed samples we speculate that since the two spectral features are rather narrow, compared to those for RDX and Tartaric acid, they may not contribute significantly enough to the correlation, whereas RFR is trained to look exactly for those features. Possibly, a different spectral characteristic, such as the derivative, which we experimented with in previous work but only on single substance samples, in conjunction with SAM could yield better results. In the derivative, the contribution of the background scattering in the spectrum, to the correlation is reduced, increasing the relative contribution of the absorption peaks to the correlation, and hence making Lactose more "visible" in the spectrum.
We realize that the scope of this study is limited to only three substances of interest whose spectra do not show significant overlap of spectral features, but the results are very encouraging.

8. ACKNOWLEDGEMENT

The authors acknowledge support in part from the Norwegian Customs Administration for this work.

REFERENCES

[1] van Rheenen, Arthur D. and Haakestad, Magnus W., "Robust identification of concealed dangerous substances by spectral correlation of Terahertz transmission images", IEEE Transactions on Terahertz Science and Technology, vol. 5, pp. DOI: 10.1109/TTHZ.2015.2400224, March (2015).
[2] Nystad, H. E., Haakestad, M. W., and van Rheenen, A.D., "Robust identification of concealed dangerous substances using THz imaging spectroscopy", Proc. SPIE 9483, 29 (2015)
[3] van Rheenen, Arthur D. and Haakestad, Magnus W., "Terahertz Imaging Spectroscopy - Towards Robust Identification of Concealed Dangerous Substances, presented at IRMMW & THz, Tucson, September (2014).
[4] van Exter, M., Fattinger, C., and Grischkowsky, D., "Terahertz time-domain spectroscopy of water vapor," Optics Letters, vol. 14, pp. 1128–1130, Oct. (1989).
[5] Platte, F., and Heise, M., "Substance identification based on transmission THz spectra using library search", J. Molecular Structure Volume: 1073 Special Issue: SI Pages: 3-9 (2014).
[6] There are many tutorials on PCA available on the internet, as an examples we mention L. I. Smith, (2002) (http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf), and J. Shlens (2014) (http://arxiv.org/pdf/1404.1100.pdf)
[7] Chan, W. L., Deibel, J., and Mittleman, D. M., "Imaging with terahertz radiation," Rep. Prog. Phys., vol. 70, pp. 1325–1379, Jul. (2007).
[8] Brigada, D. and Zhang, X.-C., "Chemical identification with information-weighted Terahertz spectrometry," IEEE Transactions on Terahertz Science and Technology, vol. 2, pp. 107–112, Jan. (2012).
[9] Tonouchi, M., "Cutting-edge terahertz technology," Nature Photonics, vol. 1, pp. 97–105, Feb. (2007).
[10] El Haddad, J., Bousquet, B., Canioni, L., and Mounaix, P., "Review in terahertz spectral analysis," TRAC - Trends in analytical chemistry, vol. 44, pp. 98–105, (2013).
[11] Fischer, B., Hoffmann, M., Helm, H., Modjesch, G, and Uhd Jepsen, P., "Chemical recognition in terahertz time-domain spectroscopy and imaging," Semicond. Sci. Technol., vol. 20, pp. S246–S253, (2005).
[12] Shen, Y. C., Lo, T., Taday, P. F., Cole, B. E., Tribe, W. R., and Kemp, M. C., "Detection and identification of explosives using terahertz pulsed spectroscopic imaging," Applied Physics Letters, vol. 86, paper no. 241116, (2005).
[13] Chen, J., Chen, Y., Zhao, H., Bastiaans, G. J., and Zhang, X.-C., "Absorption coefficients of selected explosives and related compounds in the range of 0.1-2.8 THz," Optics Express, vol. 15, pp. 12 060–12 067, Sep. (2007).
[14] Ellrich, F., Torosyan, G., Wohnsiedler, S., Bachtler, S., Hachimi, A, Jonuscheit, J., Beigang, R, Platte, F, Nalpantidis, K., Sprenger, T., and Hübsch, D., "Chemometric tools for analysing terahertz fingerprints in a postscanner," in 37th Int. Conf. on Infrared, Millimeter, and THz Waves, Wollongong, Australia, Sep. (2012).
[15] Wu, H., Heilweil, E. J., Hussain, A. S., and Khan, M. A., "Process analytical technology (pat): Quantification approaches in terahertz spectroscopy for pharmaceutical application," Journal of Pharmaceutical Sciences, vol. 97, pp. 970–984, Feb. (2008).

[16] Watanabe, Y., Kawase, K., Ikari, T., Ito, H., Ishikawa, Y., and Minamide, H., "Component analysis of chemical mixtures using terahertz spectroscopic imaging," Optics Communications, vol. 234, pp. 125–129, (2004).
[17] Kemp, M. C., "Explosives detection by terahertz spectroscopy–a bridge too far?" IEEE Transactions on Terahertz Science and Technology, vol. 1, pp. 282–292, Sep. (2011).
[18] van Rheenen, A. D. and Haakestad, M. W., "Detection and identification of explosives hidden under barrier materials - what are the THz technology challenges?" Proc. SPIE, vol. 8017, p. 801719, (2011).
[19] Ellrich, F., Weinland, T., Theuer, M, Jonuscheit, J., and Beigang, R., "Fibercoupled Terahertz spectroscopy system," Techn. Messen, vol. 75, pp. 14–22, (2008).
[20] Hastie, T., Tibshirani, R., and Friedman, J., "The Elements of Statistical Learning, Springer Series in Statistics, Springer, New York (2001).
[21] Nystad, Helle E., "Comparison of Principal Component Analysis and Spectral Angle Mapping for Identification of Materials in Terahertz Transmission Measurements", Master's thesis, Norwegian University of Technology and Science, January (2015).