

# BLIND SOURCE SEPARATION IN RAMAN AND ATR-FTIR SPECTROSCOPY: A PROCESSING CASE STUDY

Maggioni G.M., Kocevskaja S., Rousseau R. W., Grover M. A.  
Georgia Institute of Technology  
Atlanta, GA 30332

## *Abstract Overview*

We have developed a blind source separation procedure to be applied on low-level nuclear waste processing, to identify the number of species in an aqueous mixture, label them with respect to a reference library, and determined their relative concentrations. We have tested our procedure against simulated and experimental data for a mixture of water plus five sodium ions using both Raman and ATR-FTIR measurements.

## *Keywords*

Spectroscopy, Blind Source Separation, Calibration.

## **Introduction**

Low-level radioactive waste at the Hanford Site in the USA is to be vitrified (i.e. the aqueous nuclear waste is transformed into a glass) to attain long-term, safe, and environmentally sustainable storage (Holt, 2018). To avoid deterioration of the equipment and to ensure stability of the process, operating conditions must remain within a tight range of values. Ultimately, the main variables to control (hence to monitor) the process and achieve safety and efficiency are the temperature, the species present in the feed, and their (relative) concentrations.

In the present work, we have developed a protocol to identify the ATR-FTIR and Raman spectra of pure species in a mixture and to compute a reliable estimate of their relative concentrations in a manner as calibration-free as possible, using Blind Source Separation (BSS) techniques (Griffin et al. 2016, Adar et al. 1997, Dunuwila et al. 1997). We have investigated the application of our protocol on a simulant of actual low-level radioactive waste, particularly in the presence of noise. We also discuss a method to determine the number of detectable species in the mixture relying on the data matrix alone.

## **Modeling**

The intensity of the absorbed (ATR-FTIR) or scattered (Raman) radiation in spectroscopy is related to the concentration of the active species. Typically, the intensity of each species follows a linear dependence on the concentration known as the Beer-Lambert law; the total intensity is obtained by combining the Beer-Lambert law with the linear superposition principle. When multiple spectra are collected, one obtains a linear system.

To avoid calibration, we propose a three-step procedure to analyze the spectroscopic data sets  $\mathbf{X}$ : 1) determination of the number of species from the data; 2) recovery of the spectra of the pure species 3) identification of the retrieved species against a library of references, which also enables estimation of the relative compositions through a one-point calibration.

The first step is based on the analysis of the singular values of the data matrix, obtained via the Singular Value Decomposition. The number of relevant species (i.e. the species that affect the final spectrum in a measurable manner) is equated to the effective rank,  $\tilde{n}_K$ , of  $\mathbf{X}$ , computed as the number of singular values necessary to reconstruct  $\mathbf{X}$  with a prescribed accuracy. The second step is the serial combination of two well-established blind separation methods (Hyvarinen 2000, Stoegbauer et al. 2009): Independent Component Analysis (ICA) and Multivariate Curve Resolution (MCR-ALS). The former is used to initialize the solution for the latter, whose solution is bounded to possess some physical properties such as non-negativity and mass-balance closure. Note that since measured spectra are corrupted by several unwanted phenomena, such as baseline drift, spikes, and -- more generally -- noise, pre-processing techniques are used prior to feeding the spectra to the algorithms. The third step consists in comparing the spectra identified with MCR-ALS with a library of pure components and calculating the molar fraction of each species.

## **Materials and Methods**

Low-level nuclear waste processed at Hanford typically is composed of more than 20 species. However,

water and a limited number of sodium (and potassium) salts account for most of its mass.

In our study, though, we substituted the real radioactive waste with non-radioactive simulant mixtures containing the relevant ions in proportions such that the chemical and physical properties are very similar to those of the actual material. Our simulant is formed by 5 sodium salts, namely sodium phosphate, sulfate, nitrite, carbonate, and nitrate (the Raman spectra of the pure species are reported in Figure 1). Note that the sodium ion is neither Raman- nor IR-active. Additionally, we note that the species have different Raman- and IR-activity, so that combining the two techniques provides a better reading of the system, for

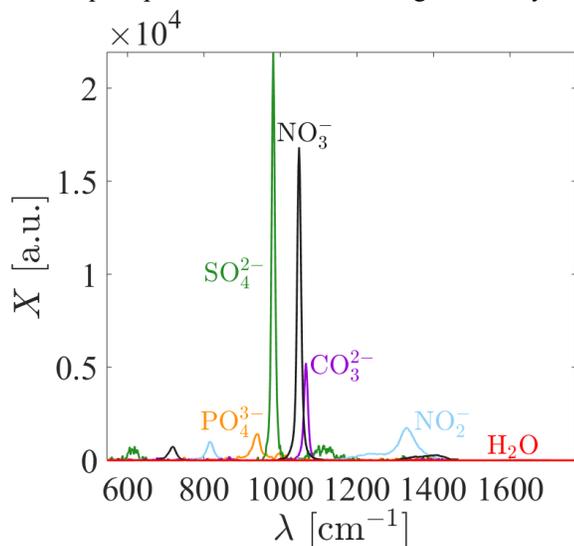


Figure 1. Raman spectra of the pure species.

instance in the case of phosphate ( $\text{PO}_4^{3-}$ ), which is weakly Raman-active but strongly IR-active. Second, these two measurement techniques are independent; hence, they can be used for mutual cross-validation for both species identification and composition estimates.

Measurements of a simulant solution at different concentrations were carried out at a constant temperature of 293 K, and Raman and IR spectra were collected simultaneously. All Raman spectra have been collected using a MarqMetrix All-in-one Raman System, with a nm laser, at 300 mW laser power, and 30-s integration time. ATR-FTIR spectra have been collected using a ReactIR ATR-FTIR technology from Mettler Toledo. The experimental measurements have been conducted in a 250-mL vessel and stirred at 400 rpm to ensure well-mixed conditions. The set of data used to run the algorithm comprise 17 samples, of which 14 had different compositions and 3 were pure water.

## Results and Discussion

First we must determine the number of species, which is done by inspecting the singular values. This inspection suggests to select  $\tilde{n}_K = 6$  for both Raman and IR. Note that setting  $\tilde{n}_K > 6$  does not produce the reconstruction of the original dataset.

After determining the number of species, we have proceeded with steps two and three of our procedure. The results of the mole fraction estimates are reported in Figures 2 for the anions, for IR (right) and Raman (left) respectively.

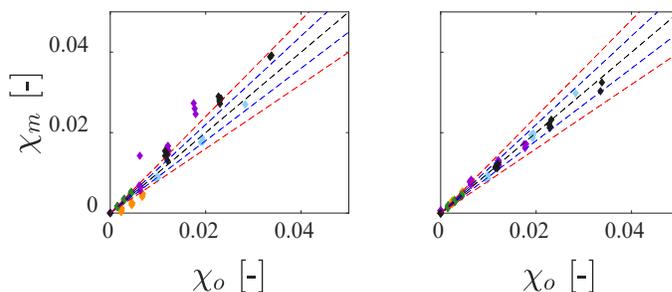


Figure 2. The modeled mole fractions (subscript “m”) versus the experimental values (subscript “o”). The blue and red dashed lines indicates deviations of 10 and 20% from the actual values, respectively.

Overall, the IR estimates seem better than the Raman. Indeed, the sources identified by form blind techniques have correlation coefficients larger than 0.95 with their associated spectra of pure components, indicating a good performance of the algorithm. On the contrary, the procedure for Raman data cannot identify any source related to phosphate; additionally, two sources are identified that, although distinct, are both correlated with the nitrate anion. However, the mole fractions estimated from Raman data are also close to their actual values.

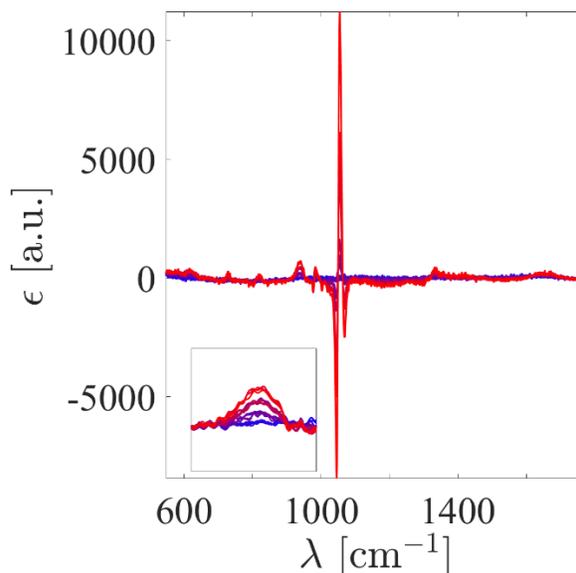


Figure 3. The residual between the Raman data and the reconstructed signal. The inset show a magnification of the characteristic region of phosphate.

An analysis of the residual, given by the element-wise difference between the measured Raman spectra and their reconstructed version, shows that most of error with a pattern is located in two regions: one corresponds to the peak of phosphate ( $880 \text{ cm}^{-1}$ ), the other corresponds to the

peaks of nitrate ( $1649\text{ cm}^{-1}$ ) and partly to carbonate ( $1669\text{ cm}^{-1}$ ). The fact that phosphate is not recognized can be explained in terms of the signal-to-noise ratio, SNR. In the Raman spectra, the average SNR is about 40 dB, but the local value in the phosphate region is only about 10 dB; i.e. the noise obscures the signal of phosphate to the point that the algorithm does not capture it. As for the deviations in the nitrate-carbonate region, the presence of a negative peak (corresponding to the location of the reference pure nitrate) associated with an almost equally intense positive peak (corresponding to the spurious source also correlated to nitrate and associated to the presence of nitrite in the system) suggests a violation to Beer-Lambert law, i.e. a nonlinear interaction between nitrate and nitrite.

## Final Remarks

We investigate the performance of a three-step blind procedure on a simulant mixture of five ionic species in water. The procedure also aims to determine the relative concentration of the species in the mixture, based on a one-point calibration approach. Our study shows that, if the assumptions of linearity in concentration, and linear superposition holds, the procedure is effective (IR data). Even in presence of mild nonlinearities, and under noisy conditions, our procedure can identify the most relevant species and yields estimates of the molar fractions close to the actual values.

## Acknowledgments

Financial support from the Consortium for Risk Evaluation with Stakeholder Participation (CRESP) is gratefully acknowledged.

## References

- Holt, M. Civilian nuclear waste disposal, *Congressional Research Service 2018*.
- Griffin, D. J.; Grover, M. A.; Kawajiri, Y.; Rousseau, R. W. (2014) Robust multicomponent IR-to-concentration model regression. *Chem. Eng. Sci.* 116, 77–90.
- Adar, F.; Geiger, R.; Noonan, J. (1997) Raman spectroscopy for process/quality control. *Appl. Spectrosc. Rev.* 32, 45–101.
- Dunuwila, D. D.; Berglund, K. A. (1997) ATR FTIR spectroscopy for in situ measurement of supersaturation. *J. Cryst. Growth*, 179, 185–193
- Hyvarinen, A. (2012). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984), 20110534–20110534.
- Stoegbauer, H.; Kraskov, A.; Astakhov, S. A.; Grassberger, P. (2004) Least-dependent-component analysis based on mutual information. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 70, 1–17.