Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Chemometrics tools for Advanced Spectroscopic Analyses

To cite this article: C Scatigno *et al* 2020 *J. Phys.: Conf. Ser.* **1548** 012030

View the article online for updates and enhancements.

You may also like

- Spectrophotometric applications of digital signal processing
Roman Z Morawski
- Effect of chemometrics to accuracy of NIR spectroscopy in predicting total soluble solid and hardness of dragon fruit
I W Budiastra and M R S Dzikri
- The Use of Principle Component Analysis in Type Classification of Air-dry Peat
N Chukhareva, A Zarubin and O Bulgakova

Breath Biopsy® OMNI®
The most advanced, complete solution for global breath biomarker analysis

TRANSFORM YOUR RESEARCH WORKFLOW

Expert Study Design & Management

Robust Breath Collection

Reliable Sample Processing & Analysis

In-depth Data Analysis

Specialist Data Interpretation

This content was downloaded from IP address 193.206.238.2 on 08/03/2023 at 11:05

Chemometrics tools for Advanced Spectroscopic Analyses

C Scatigno[1,2], R Senesi[3,4,5], G Festa[4] and C Andreani[3,4,5]

[1] Dipartimento di Scienze Fisiche e Tecnologie della materia, CNR, P.zzale Aldo Moro, 5 00185 Rome, Italy.
[2] Dipartimento di Scienze e Tecnologie Chimiche, Università degli Studi di Roma "Tor Vergata", V. le della Ricerca Scientifica, 1, 00133, Rome, Italy.
[3] Dipartimento di Fisica and Centro NAST, Università degli Studi di Roma "Tor Vergata", Via della Ricerca Scientifca 1, 00133, Rome, Italy.
[4] Centro Fermi - Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi", Piazza del Viminale 1, 00184, Rome, Italy.
[5] Istituto per i Processi Chimico-Fisici (IPCF), CNR. Viale Ferdinando Stagno d'Alcontres 37, Messina, Italy.

E-mail: claudia.scatigno@uniroma2.it

Abstract. Any system is described by several variables, often in the form of hidden information, able to describe and explain functional mechanisms for the majority of the processes which can be evaluated analytically only when we consider entire complex datasets. The relationship between those variables is the key to identify and quantify correlations among the parameters describing the data in a strictly model-free manner. In chemometrics one uses mathematical and statistical methods to improve the understanding of chemical information through the correlation of physical parameters or properties to analytical instrument data. This approach is currently used across chemistry, materials science, biology, with a growing impact is the field of spectroscopy. This paper presents the ability of chemometric technique applied to Advanced Spectroscopic Analyses, examples include spectroscopic data collected from both the High- resolution neutron Spectrometer TOSCA, operating at the ISIS pulsed Neutron and Muon Source (UK) and X-ray fluorescence (XRF) spectroscopy. This work demonstrates the high-resolution of the Principal Component Analysis (PCA) to a spectroscopic data-set dealing with the determination of marker bands from Inelastic Neutron Scattering (INS) spectra of a large data- set, the presence of a probably additional transition phase of one globular molecule and evidencing the metallic nature of the black/brownish inscriptions on daily-use textiles used in ancient Egypt. This study will pave the way for the analysis of multi-parametric, high-throughput INS data, now within reach using state-of-the-art chemical neutron spectrometers such as VESPA.

1. Introduction

Chemometrics is a rather new tool, widely used across chemistry, materials science, biology, and beyond [1]. It is becoming more common in high-throughput materials research with a growing impact in the spectroscopic field [2-5], to such an extent that chemometrics software is nowadays integrated with spectroscopic laboratory and process instrumentation as a standard [6]. Chemometrics basically reduces the dimension of the input data, large data-sets especially as a result of automated measurements, to describe the complete information with considerably fewer variables than were originally present. Principal Component Analysis (PCA) is one of the most widely used chemometric procedure for

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

multivariate data system. This approach is widely used as a discriminatory method to reveal superposition of molecular/elemental spectral band or emission lines significantly responsible for internal vibrational modes particularly intensives. In some way allows to benchmark the reference spectroscopic analysis, discriminating the characteristic vibrations. The PCA in a simple way identifies the variations in the spectral data and to interpret the data relative to the subset of the spectral variations. In this work, two different spectroscopic techniques and three different cases will be matter of enquire, highlighting the ability of the PCA technique.

PCA maps the intensity of the measured scattering function $S(Q,\omega)$ over the energy spectral range to a low-dimension space with the largest variability. In detail, the vibrational footprint of the bonding compound is studied, following the scattering of neutrons off molecular systems, with energy transfer from neutrons to bounded atoms [7]. Beyond this mechanism there is the Neutron Vibrational Spectroscopy (NVS) by exploiting Inelastic Neutron Scattering (INS) where the energy loss of the scattered neutrons is detected. In this work, INS data-set are investigating, taking into account 327 different analytes present into the ISIS database [8] to classify samples into various vibrational modes, individuating those more relevant into the inelastic scattering system. This represent a first attempt in which the Chemometrics and in particular PCA is used. Successively, the PCA is applied, for the first time, to explore phase behaviour in disordered molecular systems, precisely a globular molecule, namely t-butyl chloride (TBC) characterised by a rich phase diagram below room temperature, including a low-temperature ordered phase and two plastic-crystalline phases prior to melting. In the last case, the PCA will enable a quantitative analysis of correlations and differences in the proton-projected vibrational densities of states for these four different phases relative to less-abrupt changes of the INS response with temperature. This is of particular relevance in the context of next generation high throughput Neutron Vibrational Spectrometers, such as TOSCA [9] and VESPA- Vibrational Excitation Spectrometer with Pyrolytic-graphite Analyzers, under construction at the European Spallation Source in Lund (SE) [10]. Fast data generation and analysis on VESPA will require full integration with molecular modelling, parametric scanning and highly performing correlation analysis over large data-sets embedding multivariate data.

The third and final case regards the PCA's use to clarify the origin of metallic ink ancient Mediterranean cultures and in particular on the technology of writing on Egyptian textiles found on the tomb of Kha and his wife, Merit, was discovered in 1906 in the necropolis of Deir el-Medina (Luxor) [11].

2. Methods

The PCA successfully sorts a data-set through orthogonalization and the projection of the data onto a new coordinate system, where the coordinates are ordered hierarchically according to the variance of the projected data, without the need to invoke any specific hypothesis about the expected features [12]. It is also known as a projection method, because it takes information carried by the original variables and projects them onto a smaller number of latent variables called Principal Components (PC). Each PC explains a certain amount of the total information, namely explained variance: the identification of the primary directions of maximum variability in the data.

All the three spectroscopic cases are performed through Unscramble X version 9.5 (chemometric software package - Camo, Woodbridge, NJ, USA) [13].

2.1. INS database

Chemometrics has never been applied to INS techniques. Here, benchmarks able to determine which variables have an influence in the INS model is the aim of this first attempt analysis. 327 analytes recorded on first generation of INS beamline at ISIS (TXFA Time Focused Crystal Analyser) [14] were examined: three data matrices (327x1386, 327x800 and 327x275) were carried out to increase the chance of extracting relationships and correlations, where samples are the compounds present in the ISIS-INS data base [8] and the variables are the scattered intensity, which is directly proportional to the scattering function $S(Q,\omega)$, as a function of the energy transfer E (cm^{-1}).

2.2. T-butyl chloride

Disordered molecular systems, such as liquid and plastic phases, are still far from completely understood. The classic system t-butyl chloride has long been studied by means of quasielastic neutron scattering [15] and NMR [16, 17], at a limited number of discrete temperatures, in three different phases. There are transitions at 247K (from liquid to cubic phase I), 219K (from cubic to tetragonal phase II) and 183K (from tetragonal to an as yet unidentified solid phase III). Phase I is face-centred cubic, with non-random, but undetermined orientations [18], while Phase II is tetragonal, with head-to-tail alignment of the molecules along the C-Cl bond direction, but random orientation of the methyl groups around this axis. The original quasielastic neutron scattering studies were performed on IN5 at the ILL, at a wavelength of 7.8Å, forty years ago, and to our knowledge, no higher-resolution studies have been performed, even though this possibility is alluded to in [15]. More recently, t-butyl chloride has been studied on the TOSCA spectrometer at ISIS, over a very wide energy range and a large number of temperatures [19]. Here PCA will enable a quantitative analysis of correlations and differences in the proton-projected vibrational densities of states for these four different phases relative to less-abrupt changes of the INS response with temperature.

2.3. Inks on Linen Egyptian textiles

It is argued that the inks are related to the family of iron gall inks, whose introduction is commonly attributed to the third century BCE. Inks from 15th century BCE Egyptian textiles are investigated by combining non-invasive techniques, including ultraviolet (UV) reflected imaging, near-infrared reflectography (NIRR), X-ray fluorescence (XRF) spectroscopy, Raman spectroscopy and prompt-gamma-activation-analysis (PGAA). PCA has been used on XRF spectroscopy data (where 15 elements are detected) of 143 measurements, organised in a matrix which is subdivided in three submatrices (41x15, 20x15, 82x15) corresponding to three different categories: unmarked linen areas on tunics (37 measurement points), unmarked linen areas on loincloths (24 measurement points) an on Inscriptions (82 measurement points). The different detected elements associated with the XRF peaks or the peaks' intensities (integrated counts) were used to determine the correlations among the detected chemical elements and spot outliers. The algorithm used is the singular value decomposition (SVD), with the cross-validation method, and mean centre data as a model input to obtain a total of three components which contribute to the largest variance in elemental composition without reference to prior knowledge. Furthermore, no weighting corrections (i.e. standard deviation) are applied to the variables; one runs the risk of infecting noise in wavelengths with little information.

3. Results and discussion

3.1. INS database

PCA brings to light the INS vibrations of the phenyl ring: four vibrational modes are observed at 391, 425, 466 and 520 cm^{-1} [20] in particular, of the torsional mode of occluded anilinium [21] (black rectangle – red line PCA, **Fig. 1**), characteristic of aromatic molecule such as benzylamine ($C_6H_5NH_2$). The torsional mode τ(NH2) in these molecules is strong and a sharp line in the INS spectrum appears at 380 cm^{-1}. Methyl and ethyl torsion prevail in the INS-model vibration that found into the data: the sensitivity, in inelastic neutron scattering, of the positions of the CH_3 torsional modes is demonstrated [22, 23]. The ethyl torsion are located in the region 76-176 cm^{-1} [24, 25], whilst methyl torsions are located in the region 238-374 cm^{-1}. THF compounds (C_4H_8O) give relevant vibrational frequencies below 1500 cm^{-1} and above 400 cm^{-1} (harmonic motions), where the pseudorotational mode becomes a large-amplitude ring-deformation vibration [26]: the lattice modes are easily distinguished from the intramolecular ones by the fact that they are less sensitive to a temperature rise in the solid phase (effect of the Debye-Waller factor). Deformation modes of the of the NH_4^+ ion are observed in the range 1200- 2400 cm^{-1}.

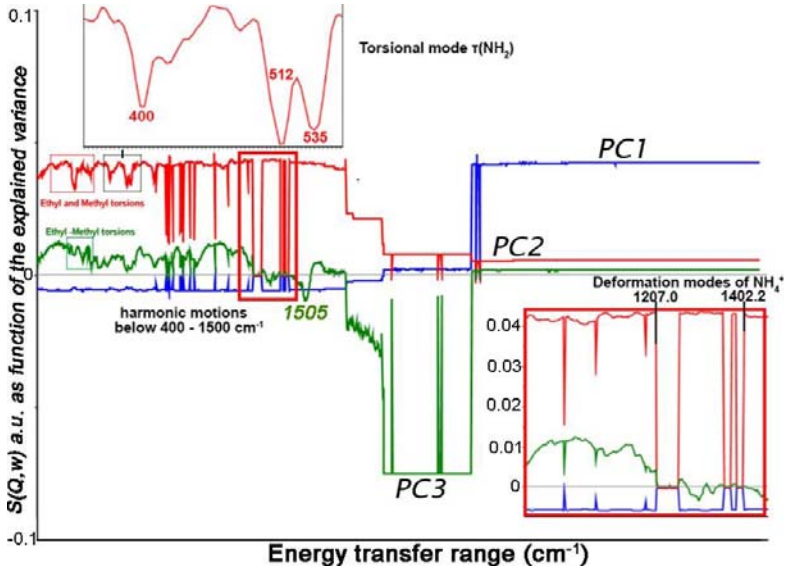

Figure 1. Loadings plot (relationships between the variables) of the variance explained by three PCs per a total of 88% (PC1 blue line, PC2 red line, PC3 green line). Size of the dataset: {327x1386}. The samples (row, 327) represent the analytes present into the database. On the x-axe, the En.tranf. range considered (16-4000 cm^{-1}), variables (column, 1386 n. of bins).

3.2. T-butyl Chloride

A PCA of the whole E-T dataset indicates that the transition at 183K may be more complicated, and that there may even be a 4th transition at ~178K (**Fig. 2**). While TOSCA has a minimum measurable energy transfer of around 2.5cm^{-1}, there is evidence in both the heights and widths of the quasielastic line that something interesting is going on between 170K and 190K (**Fig. 3**), in the low-temperature phase III: there is quasielastic scattering in Phase III, below 175K, even though it has not previously been resolved in spectroscopic measurements.

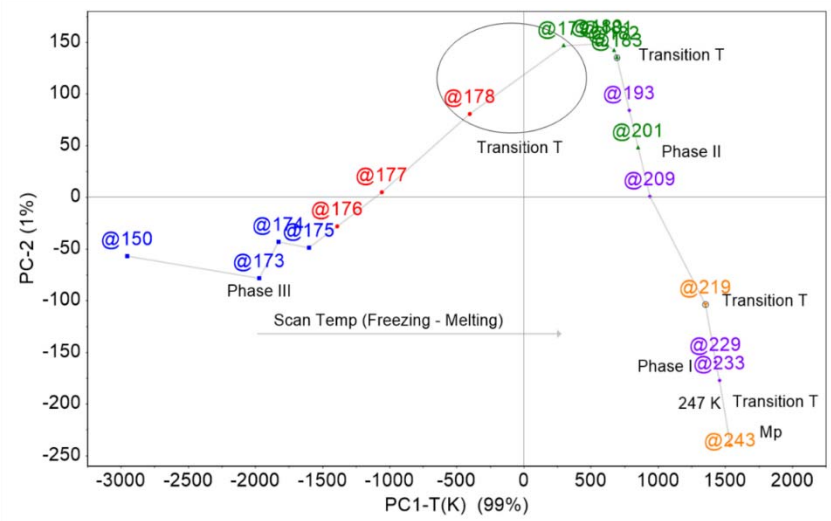

Figure 2. Variance correlation and data distribution as function of scan temperature and energy transfer range (cm^{-1}) of T Butyl Chloride investigated by INS on TOSCA. The PCA and sample grouping where both scan temperature and phase transitions are reported. The PC1 (x-axis) describes 99% of variance correlation, attributable to the rising T (from freezing to melting point "Mp". The runs where connected by solid grey line). The PCA shows 5 clusters that are grouped according their phase, crystalline and amorphous state of the molecule. The black circle shows the probable additional transition temperature (red group – 178 K), during the changeover from the solid phase (phase III – blue group, 173 K) and the transition phase (green group, 183 K) to the phase II (violet group).

3.3. Inks on Linen Egyptian textiles

PCA carried out on the whole set of XRF data, has showed a homogeneous elemental composition on the inscriptions. It has been possible to define also a direction of the release of ink from the nib (that increases from left to right). The Fe is statistically correlated with Mn, demonstrating that the inks used for the inscriptions on the textiles of Kha are likely based on ochre, a precursor of the iron gall inks. For more detail please see [11].

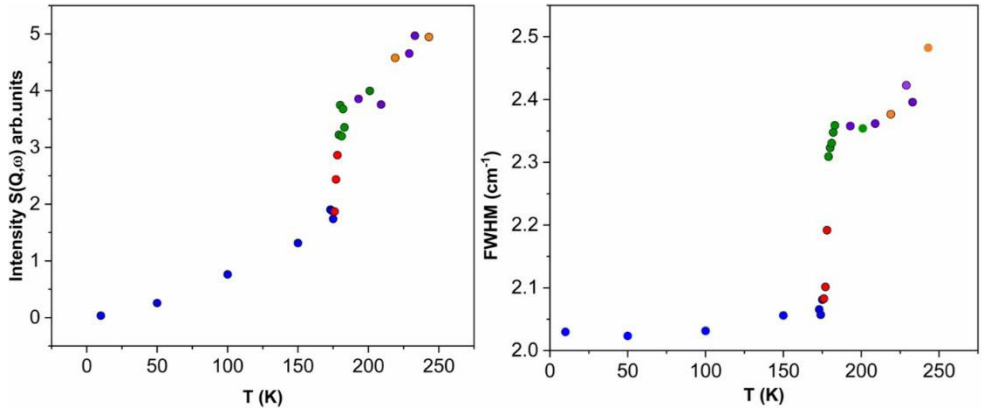

Figure 3. Right panel: Plot of the fitted quasielastic linewidth, as measured on TOSCA, which has a resolution (elastic line) of approximately 2.41 cm^{-1}. No correction has been made for the instrumental resolution. No correction has been made for the instrumental resolution. Left panel: Plot of peak intensity vs temperature (K) of the quasielastic line.

4. Conclusions

Chemometrics successfully discriminates the vibrational molecular modes. Here, interesting results on INS confirmed the ability of this powerful technique. The first data treatment attempt on INS database has pointed out the potential of the multivariate method in the drafting of database, identifying the specific vibrational modes and selected bands for future benchmarks, which could be selected on the basis on what range is to investigate. The preliminary results on t-butyl chloride gave good chances to repeat the original IN5 experiment, at a range of temperatures around those transitions, using the specific and more accurate spectrometers, more sensible to the nature of the diffusional motions, seeking to extend the study to much lower frequencies (energies), on time scales longer than those studied to date, especially in the low-temperature phase III.

Acknowledgments

The authors gratefully acknowledge the financial support of Consiglio Nazionale delle Ricerche (CNR) within CNR-STFC Agreement 2014-2020 (N 3420), concerning collaboration in scientific research at the ISIS pulsed neutron and muon source (UK) and of Science and Technology Facilities Council STFC).

References

[1] Workman Jr, J. 2019 Advances in Chemometrics for the Analytical Sciences.

[2] Scatigno C, Prieto-Taboada N, García-Florentino C, de Vallejuelo S F, Maguregui M and Madariaga J M 2017 Combination of in situ spectroscopy and chemometric techniques to discriminate different types of Roman bricks and the influence of microclimate environment *Environ.Sci.Pollut R* **25** 1-15

[3] Scatigno C, Prieto-Taboada N, Martinez M P, Conte A and Madariaga J 2018 A non-invasive spectroscopic study to evaluate both technological features and conservation state of two types of ancient Roman coloured bricks *Spectrochim.Acta A* **204** 55-63

[4] Lim J, Nam J, Shin H, Park J, Song H, Kang M, Lim K and Choi Y 2019 Identification of newly emerging influenza viruses by detecting the virally infected cells based on surface enhanced Raman spectroscopy (SERS) and principal component analysis (PCA) *Anal.Chem.*

[5] Chatterjee S, Singh M, Biswal B P, Sinha U K, Patbhaje S and Sarkar A 2019 Application of laser-induced breakdown spectroscopy (LIBS) coupled with PCA for rapid classification of soil samples in geothermal areas *Anal.Bioanal.Chem* 1-12
[6] Geladi P 2003 Chemometrics in spectroscopy. Part 1. Classical chemometrics **58** 767-82
[7] Price D L and Fernandez-Alonso F 2013 An introduction to neutron scattering **44** 1-136
[8] http://wwwisis2.isis.rl.ac.uk/INSdatabase/Theindex.asp
[9] S. F. Parker, C. J. Carlile, T. Pike, J. Tomkinson, R. J. Newport, C. Andreani, F. P. Ricci, F. Sacchetti, M. Zoppi 1998 TOSCA: a world class inelastic neutron spectrometer *Physica* **B 241** 154-156
[10] Zanetti M, Bellissima S, del Rosso L, Masi F, Chowdhury M, De Bonis A, Di Fresco L, Scatigno C, Armstrong J, Rudić S, Parker S F, Hartl M, Colognesi D, Senesi R, Andreani C, Gorini G and Fernandez-Alonso F 2019 Neutronic developments on TOSCA and VESPA: Progress to date *Physica B: Condensed Matter*
[11] Festa G, Christiansen T, Turina V, Borla M, Kelleher J, Arcidiacono L, Cartechini L, Ponterio R, Scatigno C and Senesi R 2019 Egyptian metallic inks on textiles from the 15 th century BCE unravelled by non-invasive techniques and chemometric analysis *Scientific reports* **9** 7310
[12] Hamill J M, Zhao X, Mészáros G, Bryce M and Arenz M 2018 Fast data sorting with modified principal component analysis to distinguish unique single molecular break junction trajectories *Phys.Rev.Lett.* **120** 016601
[13] https://www.camo.com/unscrambler/
[14] Delaplane, R. G., Küppers, H., Noreland, J., and Parker, S. F. 2002 Dynamics of the water molecule in potassium oxalate monohydrate as studied by single crystal inelastic neutron scattering. *Applied Physics A*, **74**, 1366-1367.
[15] Frost J, Leadbetter A J and Richardson R 1980 Molecular crystals and liquid crystals: new results for t-butyl chloride **290** 567-82
[16] Rudman R and Post B 1968 Polymorphism of the crystalline methylchloromethane compounds **5** 95-110
[17] Silver L and Rudman R 1970 Polymorphism of the crystalline methylchloromethane compounds. III. Differential scanning calorimetric study *J.Phys.Chem.* **74** 3134-9
[18] Schwartz R S, Post B and Fankuchen I 1951 X-Ray Investigation of t-Butyl Chloride and t-Butyl Bromide1 *J.Am.Chem.Soc.* **73** 4490-1
[19] Scatigno C 2018 Principal Component Analysis of Phase Behaviour on the Upgraded TOSCA.
[20] Powell A and Dickens P 1990 An incoherent inelastic neutron scattering study of NH 4 TaWO *Applied Physics A*, **51** 226-30
[21] Herzog-Cance M, Jones D J, El Mejjad R, Rozière J and Tomkinson J 1992 Study of ion exchange and intercalation of organic bases in layered substrates by vibrational spectroscopy. Inelastic neutron scattering, infrared and Raman spectroscopies of aniline-inserted α-and γ-zirconium hydrogen phosphates *J. of the Chemical Society, Faraday Transactions* **88** 2275-81
[22] Hudson M J, Workman A D, Jones D J, Bonnet B and Tomkinson J 1991 Study of ion exchange and intercalation of organic bases in layered substrates by vibrational spectroscopy. Part 1.—Inelastic neutron scattering, infrared and Raman spectroscopies of ammonium-and tetramethylammonium-exchanged α-tin hydrogen phosphate **1** 923-8
[23] Ratcliffe C I and Waddington T C 1976 Torsional modes in multimethylammonium halides *J. of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **72** 1935-1956
[24] Pal M, Agarwal A, Khandelwal D and Bist H 1984 Dynamic disorder of (CH3) 4N in (CH3) 4NX (X= Cl, Br and I) as studied by Raman spectroscopy *J.Mol.Struct.* **112** 309-16
[25] Brand H, Curtiss L, Iton L, Trouw F and Brun T 1994 Theoretical and inelastic neutron-scattering studies of tetraethylammonium cation as a molecular sieve template *J.Phys.Chem.* **98** 1293-301
[26] Ferraro J and Walker A 1965 Comparison of the infrared spectra (4000–70 cm^{-1}) of several hydrated and anhydrous salts of transition metals *J.Chem.Phys.* **42** 1278-85