



**HAL**  
open science

## Analysis of soda-lime glasses using non-negative matrix factor deconvolution of Raman spectra

William Woelffel, Corinne Claireaux, Michael J. Toplis, Ekaterina Burov, Etienne Barthel, Abhay Shukla, Johan Biscaras, Marie-Hélène Chopinet, Emmanuelle Guillard

### ► To cite this version:

William Woelffel, Corinne Claireaux, Michael J. Toplis, Ekaterina Burov, Etienne Barthel, et al.. Analysis of soda-lime glasses using non-negative matrix factor deconvolution of Raman spectra. *Journal of Non-Crystalline Solids*, 2015, 428, pp.121-131. 10.1016/j.jnoncrysol.2015.08.016 . hal-01191795

**HAL Id: hal-01191795**

**<https://hal.sorbonne-universite.fr/hal-01191795>**

Submitted on 2 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Analysis of soda-lime glasses using Non-negative Matrix Factor deconvolution of Raman spectra

William Woelffel<sup>a</sup>, Corinne Claireaux<sup>a</sup>, Michael J. Toplis<sup>b</sup>, Ekaterina Burov<sup>a</sup>, Étienne Barthel<sup>a,e,1</sup>, Abhay Shukla<sup>d</sup>, Johan Biscaras<sup>d</sup>, Marie-Hélène Chopinet<sup>a</sup>, Emmanuelle Guillard<sup>a</sup>

<sup>a</sup>Surface du Verre et Interfaces (UMR 125), CNRS/Saint-Gobain Recherche, 39 quai Lucien Lefranc, 93300 Aubervilliers, France

<sup>b</sup>Institut de Recherche en Astrophysique et Planétologie (UMR 5277), Observatoire Midi-Pyrénées, 14 avenue Belin, Toulouse, 31400, France

<sup>c</sup>Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie (UMR 7590), CNRS/UPMC, 4 place Jussieu, 75005 Paris, France

<sup>d</sup>École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI) ParisTech, PSL Research University, Sciences et Ingénierie de la Matière Molle, CNRS UMR 7615, 10 rue Vauquelin, F-75231 Paris Cedex 05, France

<sup>e</sup>Sorbonne-Universités, UPMC Univ. Paris 06, SIMM, 10, Rue Vauquelin, F-75231 Paris Cedex 05, France.

---

## Abstract

Novel statistical analysis and machine learning algorithms are proposed for the deconvolution and interpretation of Raman spectra of silicate glasses in the Na<sub>2</sub>O-CaO-SiO<sub>2</sub> system. Raman spectra are acquired along diffusion profiles of three pairs of glasses centered around an average composition of 69.9 wt.% SiO<sub>2</sub>, 12.7 wt.% CaO, 16.8 wt.% Na<sub>2</sub>O. The shape changes of the Raman spectra across the compositional domain are analyzed using a combination of principal component analysis (PCA) and sparse non-negative matrix factorization (NMF). This procedure yields components accounting for the observed changes, as well as their mixing proportions, without any direct prior assumption as to their actual shape, number or position. These methods are applied separately to the *Q* band (wavenumbers in the range 850–1400 cm<sup>-1</sup>), the main band (200–850 cm<sup>-1</sup>) and to the whole spectra (200–1400 cm<sup>-1</sup>). Compositional profiles obtained by Electron Probe Micro-Analysis (EPMA) are then used to relate spectral components to structural entities. Spectral components extracted from a *Q* band analysis and a complete spectral analysis show significant similarities both in terms of shape of the components and their mixing proportions. This result implies a link between *Q<sup>n</sup>* species and the shift of the medium-range network features in the main band of the Raman spectra. To illustrate the possibilities of the method, a linear regression model is used to relate the proportions of spectral components derived from the Raman spectra to chemical composition. This model can be used to determine the composition of different glasses inside the investigated compositional domain with reasonable accuracy.

**Keywords:** Soda-lime glass, Raman spectroscopy, EPMA, diffusion, chemical analysis, deconvolution

---

## 1. Introduction

Raman spectroscopy is a widely used technique for the study of the structure of silicate glasses and melts [1, 2]. Various structural units of the silicate network have Raman-active vibrational modes, and rapid acquisition times enable *in situ* studies of melt structure at high temperature [3–10] or phase transformations [11]. Furthermore, Raman spectroscopy is appreciated for the fact that it is non-destructive and can be used in fields such as cultural heritage [12–14]. It also has the advantage that the volume excited by the laser is small

(~1 μm<sup>3</sup>) allowing the study of spatial heterogeneities of composition or structure, *e.g.* resulting from ionic exchange [15], laser-irradiation [16] or densification following micro-indentation [17, 18].

Despite such successes, application of Raman spectroscopy to amorphous silicates is hampered by the fact that there is no simple *a priori* way to relate the shape of broad Raman bands to specific network vibrations and structures, and that the intensity of a Raman band is not a quantitative measure of species abundance, in contrast to other widely used techniques such as Nuclear Magnetic Resonance (NMR) [19]. A comparison with crystalline structures and the evolution of spectra with chemical composition has been used to interpret the positions of broad Raman bands in terms of struc-

---

Email addresses: william.woelffel@saint-gobain.com  
(William Woelffel), corinne.claireaux@saint-gobain.com  
(Corinne Claireaux)

tural entities. This approach has been particularly applied to the spectral range 850–1400  $\text{cm}^{-1}$ , where silicate tetrahedra surrounded by a variable number of non-bridging oxygens (the  $Q^n$  species defined by the NMR community, where  $n$  is the number of bridging oxygens) are expected to resonate [20, 21]. Indeed, a large body of the silicate literature is devoted to discussions of how to interpret Raman spectra in terms of  $Q^n$  species, for different kinds of network modifiers [4, 20, 21] or in the presence of other network formers [22–25]. At lower wavenumber, silicate glasses also show significant broad intensity in the range (200–850  $\text{cm}^{-1}$ ) typically referred to as the main band. These resonances are associated with T–O–T bending (where T is a tetrahedral network forming cation), although the exact relationship to the structural entities that resonate in the high wavenumber  $Q$  band (850–1300  $\text{cm}^{-1}$ ) is not generally considered [26]. Although theoretical models have been used to provide insight into the principal features of Raman spectra of amorphous silicates [27, 28] and amorphous silica [29], this approach does not currently have the ability to quantitatively interpret signal shape and intensity of multicomponent glasses. Historically, a common proxy for band shapes consists in fitting Raman spectra with a given number of Gaussians of variable positions, widths and intensities [13, 14, 30–32]. Inside the  $Q$  band, such Gaussians are attributed to specific  $Q^n$  units and environments. Using such an approach, a reasonable correlation has been found in lithium silicates [32] between the ratios of Gaussian bands fitted to Raman spectra, and the ratio of  $Q^n$  units determined from  $^{29}\text{Si}$  NMR. However, broad and complex band shapes often require the use of a large number of Gaussian components [10, 33, 34] for a correct fit. Furthermore, Gaussian-fitting methods typically do not impose band positions, which may thus vary between spectra, limiting the interpretation in terms of well defined  $Q^n$  units. Finally, even in simple binary silicates the number of Gaussian components required to fit the Raman spectra is greater than the number of potential  $Q^n$  species, leading to ambiguity and controversy concerning spectral assignments [6, 7, 33]. In the light of these difficulties, application of statistical methods has emerged as an alternative approach for the decomposition of spectra into several components (using techniques inspired by those used for blind source separation). Such methods typically make fewer assumptions concerning the shape of partial spectral components, and hence require a smaller number of adjustable parameters than Gaussian-based methods. For example, principal component analysis [35] decomposes a signal into orthogonal components and is one of the most pop-

ular methods [36], followed by independent component analysis [37] for statistically-independent signals. Non-negative matrix factorization (NMF) [38] has recently gained attention since it enables a spectroscopic interpretation of the non-negative extracted features. NMF has been used and improved [39] for fluorescence spectroscopy [40],  $^1\text{H}$ -NMR spectroscopy [41] or hyperspectral imaging [42]. Recent studies [43–46] have applied NMF-inspired techniques, coupled with chemical constraints, to shed light on the speciation of glasses and melts in binary alkali and alkali-earth silicates, an approach that leads to the isolation of spectral components in the  $Q$  band that can be assigned to the spectral signature of the different  $Q^n$  species.

In this paper, we extend previous work in binary systems [43–46] to the ternary system of sodium and calcium-bearing silicates. Inter-diffusion experiments of melts with different initial compositions are used to generate a range of compositions for which a large number of Raman spectra may be acquired. A novel deconvolution algorithm based on NMF is developed and applied to this comprehensive dataset. The shapes of the different components identified by the algorithm for our ternary system are presented and the correlations between the high-frequency envelope ( $Q$  band) and the main band are discussed. A chemical model that relates Raman spectra to chemical composition is proposed and tested against compositions lying outside the training dataset.

## 2. Experimental methods

### 2.1. Glasses preparation

Three initial glass compositions were prepared from industrial grade sand, limestone, and sodium carbonate. Reagent grade  $\text{Na}_2\text{SO}_4$  was used as a fining agent. Small amounts of coloring agents  $\text{CoO}$  and  $\text{FeCr}_2\text{O}_4$  were added to certain glasses to provide a quick means of identification after diffusion experiments. The different glass compositions are summarized in table 1. Glasses used for diffusion experiments are centered around an average composition of 69.9 wt%  $\text{SiO}_2$ , 12.7 wt%  $\text{CaO}$ , 16.8 wt%  $\text{Na}_2\text{O}$ , close to common window glass. Trace amounts (0.1 wt%) of  $\text{Al}_2\text{O}_3$  and  $\text{MgO}$  are also present due to the use of industrial grade sand and limestone. Three other test glasses (compositions given in table 1) were also prepared.

The raw materials were dried at 110°C for over 24 h, then mixed with a Turbula shaker-mixer. 1-kg batches used to prepare glasses for diffusion experiments were melted in a Pt-Rh Joule-heated crucible at 1450°C and

stirred for 2 h with a Pt-Rh rod. The melts were poured onto an iron plate, air-quenched and annealed at 550°C for 2 h. The compositions were also checked by electron microprobe (EPMA) on fragments of the resulting glasses, and were found to be homogeneous (standard deviation below 0.2 wt%).

Table 1: Glasses compositions from X-ray fluorescence analysis in wt%. Standard deviations from repeated measurements on a standard glass are as follows: 0.02 wt.% CaO, 0.07 wt.% Na<sub>2</sub>O, 0.01 wt.% Al<sub>2</sub>O<sub>3</sub>, 0.02 wt.% MgO, 0.002 wt.% Fe<sub>2</sub>O<sub>3</sub>, 0.0002 wt.% Cr<sub>2</sub>O<sub>3</sub>.

Name	SiO <sub>2</sub>	CaO	Na <sub>2</sub> O	Al <sub>2</sub> O <sub>3</sub>	MgO	Fe <sub>2</sub> O <sub>3</sub>	Cr <sub>2</sub> O <sub>3</sub>
S	75.4	10.1	14.2	0.1	0.1	0.0	0.0
C	67.6	17.9	14.0	0.1	0.1	0.0	0.0
N	66.7	10.1	22.3	0.2	0.2	0.1	0.1
TEST1	73.4	10.4	15.6	0.0	0.0	0.0	0.0
TEST2	66.1	16.3	17.2	0.0	0.0	0.0	0.0
TEST3	72.7	13.7	13.3	0.0	0.0	0.0	0.0

## 2.2. Diffusion profiles

20 × 20 × 5 mm slices were cut from previously prepared glasses and one surface (one side) for each glass slice was polished to 40 μm. Slices were stacked so that the lower-density one was on top of the other to limit convection. Densities at 1000°C were predicted using the model of [47]. The space around glass slices was filled with silica sand to limit liquid movement (see Fig. 1). This setup accommodates thermal expansion while reducing mass transfer acceleration due to flow of the softening glass.

Samples were introduced in a preheated muffle furnace at 1000°C for 1 h. This temperature-time range was selected such that compositional profiles generated by diffusion were on the order of a couple of millimeters. At the end of the experiment the samples were quenched in air to avoid crystallization. This quench typically fractured the glass. Samples were subsequently annealed for 1 h at 600°C and cooled inside the furnace overnight. Slices of the glass stack were cut perpendicular to the initial interface, the presence of coloring agents helping to identify the region of variable composition and/or avoid areas where complex convection had taken place. Samples were polished to optical quality before further analysis. Each diffusion couple is named after its two endmember glasses, *e.g.* the diffusion sample extracted from S glass (silica-rich) over C glass (lime-rich) is referred to as SC, and allows for measurements over a continuous range of compositions from one endmember to the other.

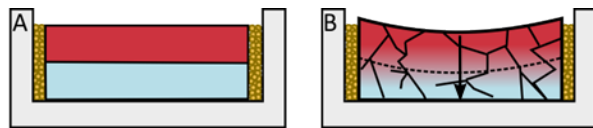


Figure 1: Diffusion profile experiments. (a) Two endmember glass slabs, schematically represented as red and blue, are stacked in a crucible. They are surrounded by silica sand, depicted as yellow dots, to prevent them from flowing. (b) Sample after diffusion and annealing: the dashed line depicts the original interface between the glass slices, the solid lines represent the fractures that often result from air-quenching the sample. The curved shape of the free surface is due to the slight creep, in spite of the silica sand. Diffusion between the endmember glasses is represented by the arrow.

Composition measurements along the diffusion profiles were made on a Cameca SX 100 EPMA with a 15 keV, 1 nA electron beam during 1 min for Si and Ca analysis and 3 × 20 s counting time for Na analysis.

## 2.3. Raman spectroscopy

Raman spectra along the diffusion profiles were collected with a confocal Jobin-Yvon Xplora system in back-scatter geometry, focusing a 532 nm laser on the sample through a ×100 microscope objective. Nominal power of the laser is about 15 mW. The spectra were collected with 3 acquisitions of 1 min each.

Raman spectra of the test glasses were collected with a confocal ThermoFischer DXR system in back-scatter geometry, using a 532 nm laser with 10 mW of power through a 25 μm slit and a ×50 objective. The spectra were collected with 10 acquisitions of 30 s each.

In order to test the reproducibility between the two spectrometers, Raman spectra from the diffusion samples were also collected on the ThermoFisher Raman spectrometer. No differences were found with the spectra previously recorded on the Xplora spectrometer.

For all 169 Raman spectra collected, the background was corrected for by subtracting the minimum of the spectrum over the 200–1400 cm<sup>-1</sup> spectral region. A temperature-frequency correction [1, 48] was then applied. Finally, all spectra were normalized to unit area over the same spectral range. In the following this procedure is referred to as the cleaning stage.

## 3. Deconvolution method

The Raman spectrum of a silicate glass can be described as the sum of the partial Raman spectra (PRS), *i.e.* the sum of its Raman-active structural species. For example, the signal in the 850–1300 cm<sup>-1</sup> spectral range is often interpreted as the sum of the  $Q^n$  stretching modes [21]. Similarly, the so-called main band

at lower frequencies can be thought of as the sum of medium-range structural species (such as rings of tetrahedra), bending vibrations of the tetrahedra and inter-tetrahedral linkages [22, 49, 50]. Several attempts have been made to predict localized vibrations through normal mode analysis on tetrahedral clusters [20] or molecular dynamics [28].

As discussed above, the most widespread method to rationalize the Raman spectra of silicate glasses is to fit a sum of Gaussian functions, whose positions, standard deviations and amplitudes are optimized for each spectrum. Furthermore, this approach is typically restricted to the  $Q$  band spectral range. Each Gaussian is then attributed to a  $Q^n$  species, or sometimes to a  $Q^n$  species with a specific surrounding environment [10]. Ideally, for spectra collected on glasses with similar compositions one would expect a fixed number of Gaussians with similar Raman shifts but of variable proportion. However, consideration of the literature shows that different authors fit spectra in different ways depending on the criterion used for selecting the number of Gaussians, leading to ambiguity in structural interpretation (e.g. close to  $\text{NaAlSiO}_4$  in the system  $\text{Na}_2\text{O}-\text{Al}_2\text{O}_3-\text{SiO}_2$  [6, 7, 33]).

The goal of the deconvolution procedure proposed here is to obtain a unique set of partial Raman spectra related to the Raman-active species in the overall set of experimental spectra, and to determine the corresponding weights of those endmembers in individual spectra. No assumption is made concerning the shape of the PRS, nor their number. The only assumption is that  $Q^n$  units undergo no significant distortion, and thus have the same PRS over the entire compositional range. This requirement is met thanks to the restricted composition range studied. Our deconvolution algorithm shares many similarities with the approach of Malfait and collaborators [43–45]. The deconvolution procedure, summarized in Fig. 2 and described in the following, is fully automatic and is designed to have a minimum number of arbitrary parameters. Our implementation relies on the Python scientific stack [51–53]. The reader interested in the chemical and physical results may safely skip the following mathematical description to read Section 4.

*Principal component analysis.* PCA is commonly used to reduce the dimensionality of data sets. PCA of an  $n$ -observation set of  $p$  features yields up to  $m = \min(n, p)$  principal components, ranked by decreasing order of explained variance. In our case,  $n = 169$  is the number of Raman spectra collected on diffusion samples, and  $p = 715$  is the spectral sampling, i.e. the number of wavenumber values at which the signal inten-

sity is measured. Since  $n \ll p$ ,  $m = n$ . Out of the  $m$  principal components, only a small number  $k$  represent relevant information, while the rest only convey noise. The choice of  $k$  is often based on the decreasing explained variance ratio. However, this is not a robust criterion when  $n \ll p$  as here. Moreover, principal component analysis yields orthogonal components, which necessarily have both positive and negative parts if the components are overlapping. Potential negative values make PCA components unsuitable for further spectroscopic interpretation. Independent component analysis (ICA), whose purpose is to unmix  $k$  linearly independent sources from  $n \geq k$  observations, would seem well-suited for deconvolution. However, the first step of ICA is to decorrelate components, which is rather contradicting with the expected correlation between components due to the chemistry. Besides, ICA is not guaranteed to find non-negative components or mixing weights.

*Choice of the number of components.* PCA can nevertheless be used as a first step in order to determine  $k$ , the number of components to look for, without making any assumption about the glass structure. For this purpose the collection of  $n$  spectra is stacked in an  $n \times p$  matrix  $\Sigma$ , where  $p$  is the number of spectral features. It is decomposed into components ranked by decreasing singular values  $\lambda_i$ . These  $\lambda_i$  are used to determine the value of  $k$  since they are an indication of how much “information” is carried by the corresponding component. Choosing too low a value for  $k$  leads to loss of information but a small level of noise, while too high a value includes noise and unnecessary information. A compromise between information and simplification must be found. Malinowski developed [54] a statistical approach to assess the amount of error introduced by the addition of an increasing number of components. His empirical indicator function [55] IND is computed from the singular values:

$$\forall r \in \{1, \dots, n-1\}, \quad \text{IND}(r) = \frac{\text{RE}(r)}{(n-r)^2} \quad (1)$$

where

$$\text{RE}(r) = \sqrt{\frac{\sum_{i=r+1}^n \lambda_i}{p(n-r)}}. \quad (2)$$

The number of components to consider  $k$  is that at which IND (Fig. 3) has its minimum value.

*Non-negative matrix factorization.* NMF is applied to deconvolute  $\Sigma$  into two non-negative matrices  $\mathbf{W}$  of

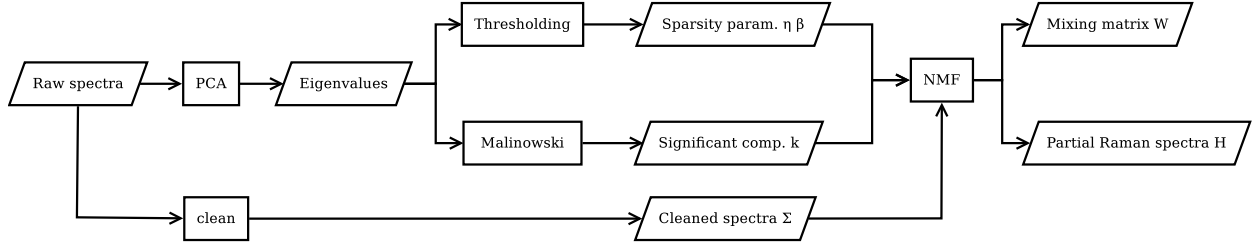


Figure 2: Deconvolution flowchart

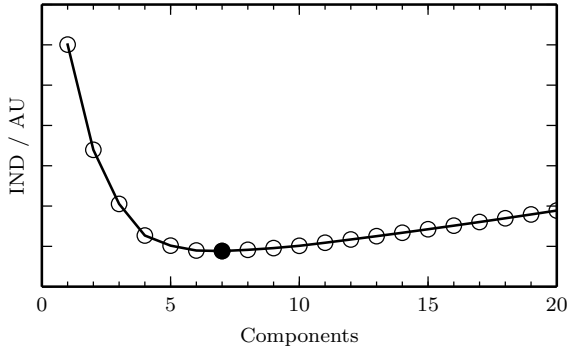


Figure 3: Malinowski's Indicator function for the PCA of the complete set of spectra over the whole spectra range

shape  $n \times k$  and  $\mathbf{H}$  of shape  $k \times p$  so that

$$\Sigma = \mathbf{W}\mathbf{H} + \epsilon, \quad (3)$$

where  $\epsilon$  is the residual error.  $\mathbf{H}$  contains up to  $k$  non-null lines, which are the partial Raman spectra.  $\mathbf{W}$  is the mixing matrix. Figure 4 illustrates this decomposition. Any spectrum from the  $\Sigma$  set can thus be written as a weighted linear combination of the  $k$  partial Raman spectra.

$$\forall i \in \{1, \dots, n\}, \Sigma_i = \sum_{j=1}^k w_{ij} \mathbf{H}_j \quad (4)$$

Factorization is obtained by alternatively minimizing the cost function

$$f(\mathbf{W}, \mathbf{H}) = \|\Sigma - \mathbf{W}\mathbf{H}\|_F^2 \quad (5)$$

for  $\mathbf{W}$  and then for  $\mathbf{H}$  [56], while keeping both matrices positive.  $\|\cdot\|_F$  denotes the Frobenius norm of the matrix, defined for a  $n \times m$  matrix  $\mathbf{A}$  by

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2 = \sum_{i=1}^n \|\mathbf{A}_i\|_2^2. \quad (6)$$

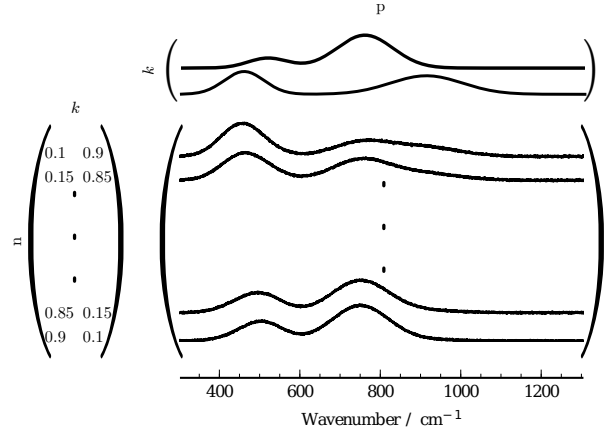


Figure 4: Non-negative matrix factorization using  $k = 2$  factors. Each spectrum in the collection of  $n$  spectra can be decomposed into a weighted sum of  $k = 2$  PRS. Weights are in the left-hand side  $n \times k$  matrix. PRS are in the top  $k \times p$  matrix. For instance, the second spectrum of the collection is equal to 0.15 times the first PRS plus 0.85 times the second PRS.

This alternative non-negative least-square algorithm (ANLS) is not, as such, a convex optimization problem, and the solution to this factorization problem is not unique [57]. In general, there is an infinite number of solutions corresponding to local minima of  $f$ . To achieve reproducible results, a deterministic initialization is therefore performed with non-negative double singular value decomposition [58], readily available in the scikit-learn Python package.

To enable a physical and chemical interpretation, components should have signal over as narrow as possible a spectral range. In other words, each line of the component matrix  $\mathbf{H}$  should have as many zeros as possible. This is referred to as sparsity: a matrix is said to be sparse when it is primarily filled with zeros. Consequently, a sparsity constraint is applied to  $\mathbf{H}$ . This constraint is achieved by modifying the cost function (5) minimized by the ANLS to

$$f(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\Sigma - \mathbf{W}\mathbf{H}\|_F^2 + \eta \|\mathbf{W}\|_F^2 + \beta \sum_{i=1}^n \|\mathbf{H}_i\|_1^2. \quad (7)$$

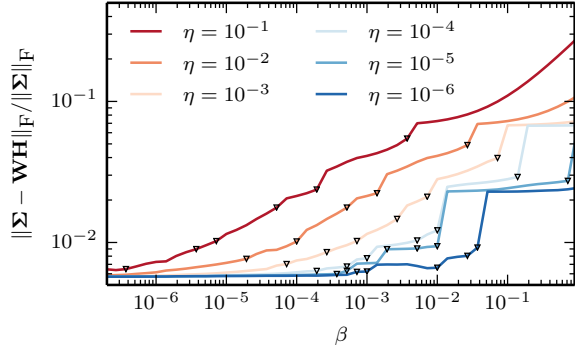


Figure 5: Deconvolution error as a function of  $\beta$  for several values of  $\eta$ . Smaller  $\beta$  means smaller sparsity constraint. Black triangles represent the appearing of a new non-null component.

$\|\cdot\|_1$  is the  $\ell_1$ -norm, defined for a vector  $\mathbf{V}$  as

$$\|\mathbf{V}\|_1 = \sum_{i=1}^n |v_i|. \quad (8)$$

The sparsity constraint is induced by the use of the  $\ell_1$ -norm on the lines of  $\mathbf{H}$  [59]. Its intensity is determined by the value of the two arbitrary parameters  $\eta$  and  $\beta$ . The higher the value of  $\beta$ , the sparser the resulting partial Raman spectra, and the fewer non-null partial Raman spectra are found in  $\mathbf{H}$ . Uninteresting solutions can arise from non-sparse solutions by downscaling  $\mathbf{H}$  while simultaneously multiplying  $\mathbf{W}$  by the inverse factor. This is prevented by the  $\|\mathbf{W}\|_F^2$  term in  $f$ . The value of  $\eta$  is adjusted accordingly to maintain a balance between  $\mathbf{W}$  and  $\mathbf{H}$ . The influence of  $\eta$  and  $\beta$  on the reconstruction error  $\epsilon$  is plotted on Figure 5.

Regardless of the value of  $\eta$ , when almost no sparsity constraint is applied (very low  $\beta$ ), the residual error reaches a plateau corresponding to the experimental noise. However, for a given value of  $\eta$ , the smaller the sparsity constraint, the more PRS are needed, up to  $k$  PRS when the constraint is fully relaxed. Thus,  $\eta$  and  $\beta$  can be coupled and the choice of both can be reduced to one arbitrary decision.  $\eta = 10^{-4}$  is chosen to keep  $\|\epsilon\|_F$  low while avoiding division errors from too low values. In line with [43], an explained variance ratio threshold of 99.5 % in the PCA was targeted. The number of components necessary to reach this variance ratio is thus obtained from PCA, and the value of  $\beta$  is found according to Fig. 5.

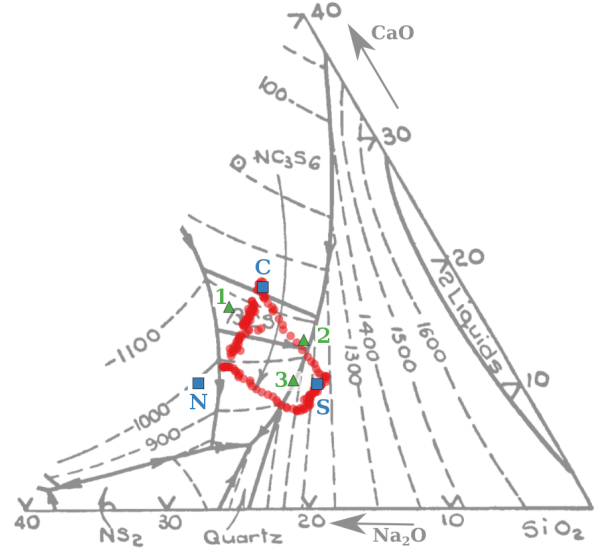


Figure 6: EPMA composition interpolation at the position of the Raman spectra in red circles, endmember glasses in blue squares, test glasses in green triangles, represented on the  $\text{SiO}_2$ - $\text{CaO}$ - $\text{Na}_2\text{O}$  ternary phase diagram. All compositions are in wt%.

## 4. Results

### 4.1. Analysis of diffusion profiles

221, 170 and 255 electron microprobe analyses were considered on the NC, SC and SN samples respectively. Spatial resolution of analyses was adjusted to the perceived steepness of the composition gradient. Composition profiles were then interpolated, and composition was later computed at the position where Raman spectra were collected. These compositions are represented on the  $\text{SiO}_2$ - $\text{CaO}$ - $\text{Na}_2\text{O}$  phase diagram [60] in Fig. 6.

The curved shape of the diffusion lines between the N and S endmembers is noteworthy and involves uphill diffusion, a phenomenon that still raises interest [61]. We also note that diffusion profiles do not extend all the way to the sodium-bearing endmember glass. This is due to the fact that Raman spectra were not collected all the way into the original endmember glass because it was difficult to assess where the diffusion profile terminated, even with the presence of coloring agents.

### 4.2. Raman spectra

56, 41 and 72 spectra were manually collected on the NC (Fig. 7c), SC (Fig. 7a) and SN (Fig. 7b) samples respectively, with higher spatial sampling where signal variation is steeper. All spectra show a broad asymmetric peak in the  $540$ – $650 \text{ cm}^{-1}$  region, and another one in the  $900$ – $1200 \text{ cm}^{-1}$  region.

SC and SN samples show significant spectral changes in both the main and  $Q$  bands, in response to the exchange between the network former  $\text{SiO}_2$  and the network modifiers  $\text{CaO}$  and  $\text{Na}_2\text{O}$ . The broad main band peak becomes wider and shifts from  $\sim 545 \text{ cm}^{-1}$  to  $\sim 570 \text{ cm}^{-1}$  with increasing addition of network modifiers. Low wavenumber ( $950\text{--}1000 \text{ cm}^{-1}$ ) signal increases in the  $Q$  peak at the expense of intensity in the range  $1150\text{--}1200 \text{ cm}^{-1}$ . NC exchange seems to have almost no impact on the main band, and only subtle impact on the  $Q$  band.

All spectra are stacked and cleaned into the  $\Sigma$  matrix (as explained in 2.3). The slope of the lower wavenumber part of the spectra may be due to the choice of background removal described in 2.3 and it is questionable whether it conveys relevant structural information.

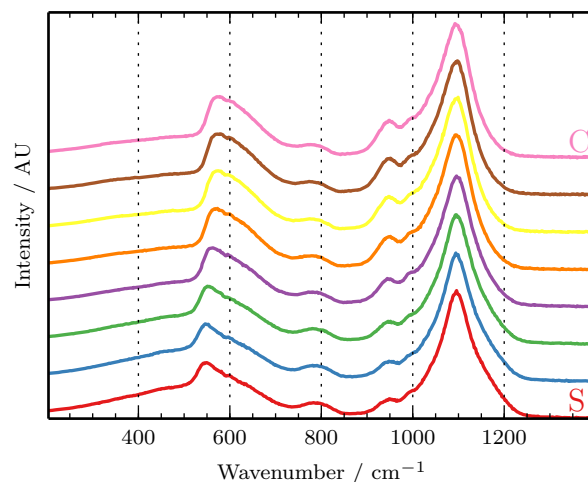
#### 4.3. $Q$ band deconvolution

Spectra are split into the main band and the  $Q$  band at a wavenumber of  $850 \text{ cm}^{-1}$ . At first, deconvolution is carried out on the  $Q$  band only, that is, spectral data below  $850 \text{ cm}^{-1}$  are discarded before area normalization.  $\Sigma$  is deconvoluted using the procedure sketched out in Fig. 2.  $k = 7$  components are allowed, with values of  $\eta$  and  $\beta$  deduced from thresholding as described in Section 3.

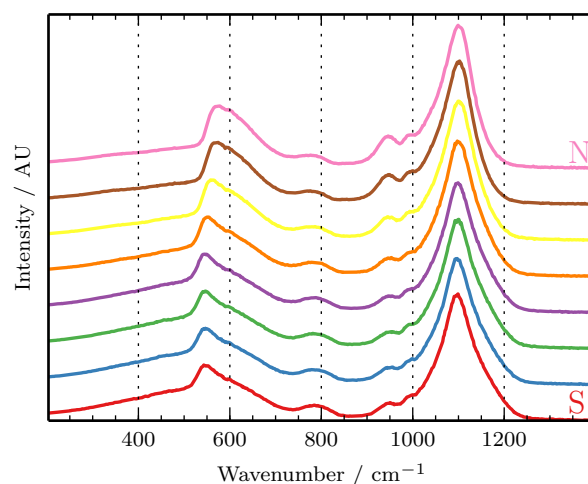
$Q$  band deconvolution yields four PRS, named  $Q^a$  to  $Q^d$  and plotted in Fig. 8. Since PRS and their relative weights for each spectrum are both determined simultaneously by sparse NMF, intensities of the PRS *cannot be directly related to the composition*. Instead, the *area ratios* of the PRS for a spectrum bear the actual information.

$Q^a$  has a broad peak maximum at  $1096 \text{ cm}^{-1}$ .  $Q^b$  has a narrower peak, with maximum intensity at  $1163 \text{ cm}^{-1}$ .  $Q^d$  is bimodal with a main peak at  $948 \text{ cm}^{-1}$  and a less intense broad plateau in the range  $1000\text{--}1050 \text{ cm}^{-1}$ .  $Q^c$  has an almost symmetric peak centered at  $1118 \text{ cm}^{-1}$  and smaller features in the same range as  $Q^d$ . Figure 9 illustrates the reconstitution of two representative  $Q$  band spectra, demonstrating the overall quality of fit.

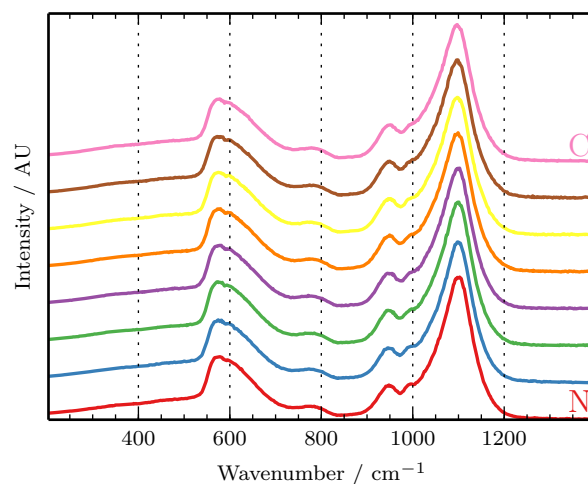
The shapes and positions of the spectral features of the  $Q$  band PRS provide certain constraints on their structural origins, but more direct information can be extracted from the evolution of their area ratios along the diffusion profiles as a function of chemistry. A detailed analysis of the evolution of the PRS area ratios (Figure 9) along the diffusion profiles is outside the scope of this article, and shall be presented in the future. However, correlation between EPMA measurements and PRS area ratios (table 2) sheds light on the



(a) SC sample



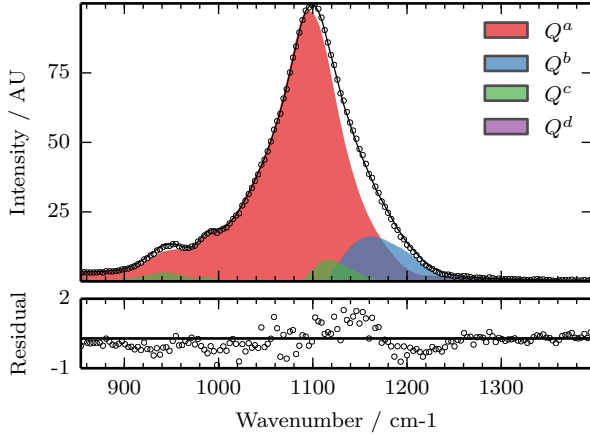
(b) SN sample



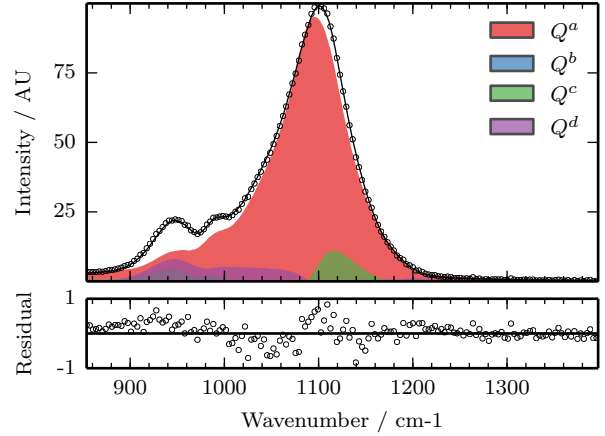
(c) NC sample

7 Figure 7: Sampling of the spectra collected along the diffusion profiles. Total number of spectra collected and used in the deconvolution process is respectively 41, 72, and 56.





(a) S side. EPMA analysis (wt%): 76.0 SiO<sub>2</sub>, 8.5 CaO, 15.5 Na<sub>2</sub>O. Area ratios: 85.7 %  $Q^a$ , 10.3 %  $Q^b$ , 3.9 %  $Q^c$ , .0 %  $Q^d$ .



(b) N side. EPMA analysis (wt%): 68.9 SiO<sub>2</sub>, 11.5 CaO, 19.6 Na<sub>2</sub>O. Area ratios: 86.6 %  $Q^a$ , 0.2 %  $Q^b$ , 5.8 %  $Q^c$ , 7.4 %  $Q^d$ .

Figure 9: Reconstitution of the  $Q$  band of two spectra from the SN sample. Circles are measured spectra, continuous lines are reconstitutions from summing weighted PRS. Residuals are obtained by subtracting reconstitutions from the experimental data.

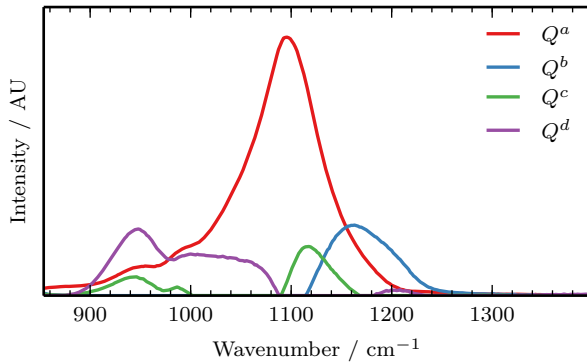


Figure 8:  $Q$  band (850–1400 cm<sup>-1</sup>) deconvolution PRS as returned by the NMF.

species represented by the PRS.  $Q^d$  is very strongly correlated to the CaO content of the glass, while being very strongly negatively correlated to its SiO<sub>2</sub> content. Such a relationship clearly indicates that  $Q^d$  is associated with a depolymerized unit in the glass structure, since its area ratio decreases with the addition of a network former, and increases with addition of a depolymerizing agent.  $Q^d$  is also bimodal, and in this respect we note that bimodality has been predicted for the Raman spectrum of  $Q^2$  units [27], and observed in sodium silicate glasses at room temperature [44]. We therefore assign this PRS to  $Q^2$  that we infer are dominated by Ca-bearing entities (table 2). The converse is true for  $Q^b$ , which is very strongly correlated to SiO<sub>2</sub> content, while being

significantly negatively correlated to network modifier content.  $Q^b$  can consequently be linked to  $Q^4$  tetrahedra.  $Q^a$  and  $Q^c$  are close to each other, and represent the largest part of the areas of the spectra. Both are strongly correlated to the addition of network modifiers.  $Q^a$  is negatively correlated to CaO but is not significantly correlated to Na<sub>2</sub>O, whereas  $Q^c$  is. However,  $Q^a$  accounts for much more signal than  $Q^c$ . Mathematically,  $Q^a$  can be assigned to the “average” or “common” component of the glasses studied here. In this respect, we note that this “average” composition is very close to the stoichiometry of  $Q^3$  species, *i.e.* at 67 mol% SiO<sub>2</sub> (Table 1).  $Q^c$  is thus interpreted as the “deviation” from the average in the presence of more Na<sup>+</sup>,  $Q_{Na}^3$ . Further consideration of the nature of  $Q^a$  is found in part 5. These four PRS are remarkably close, in their positions and interpretations, to the deconvolutions existing in the literature for similar glasses based upon comparison with crystalline analogs [26, 62].

Four components are enough for a successful deconvolution of 169 spectra, as illustrated by the small residual levels on Fig 9. However, area ratios cannot be directly interpreted in terms of speciation information although this would be possible with independent <sup>29</sup>Si NMR calibration.

#### 4.4. Full spectral range deconvolution

After successful analysis of the  $Q$  band, five PRS are obtained from the deconvolution of the whole spectral range, named from  $F^a$  to  $F^e$ . Resulting partial Raman spectra are plotted in Fig. 10 and examples of fits

Table 2: Correlation between  $Q$  band area ratios and EPMA measurements along diffusion profiles. Correlation coefficients with absolute value above 0.5 are colored according to their value. A blue to red “cool to warm” color map is used.

	$Q^a$	$Q^b$	$Q^c$	$Q^d$
SiO <sub>2</sub>	0.428	0.969	0.119	-0.931
CaO	-0.719	-0.724	-0.624	0.962
Na <sub>2</sub> O	0.434	-0.520	0.828	-0.057

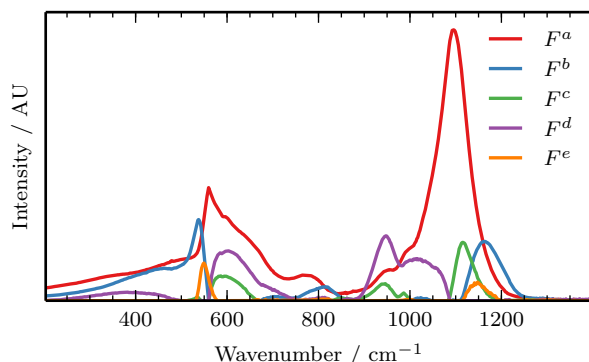


Figure 10: Full spectra range deconvolution partial Raman spectra as returned by the NMF.

to glasses from the SN sample are illustrated in Figure 11. Five PRS are sufficient to account for all spectra with very low levels of residual information (Fig. 11). It should come as no surprise that more components are needed to reproduce spectra over a wider range of wavenumber. None of the PRS are restricted to a single narrow spectral range in spite of the sparsity constraint, but none of them conveys pure noise, nor even significant levels of noise. This result thus demonstrates that resonances in the main band are not independent from those in the  $Q$  band, as discussed below. Further confirmation of this idea is provided by deconvolution of the main band alone (in an analogous way to the  $Q$  band described above). That exercise produces four PRS that show significant similarity to the main band region of Fig. 10, indicating that the features derived from the treatment of the whole spectral range are robust features associated with well-defined resonances. In the light of this fact, we concentrate on the PRS derived from the full spectral range, highlighting the correlations between the main and  $Q$  bands.

$F^a$  has one sharp peak in the main band at  $560\text{ cm}^{-1}$ , and a smaller, broader one at  $776\text{ cm}^{-1}$ . Its most prominent feature is in the  $Q$  band, with a broad, almost symmetric peak at  $1096\text{ cm}^{-1}$ , very similar to  $Q^a$ .  $F^b$  has

two peaks of comparable intensities, a sharp one in the main band at  $538\text{ cm}^{-1}$  and a broad one in the  $Q$  band at  $1163\text{ cm}^{-1}$ . The latter bears a striking resemblance to  $Q^b$ .  $F^c$  has one broad peak at  $587\text{ cm}^{-1}$ , one weaker peak at  $946\text{ cm}^{-1}$  and one other peak in the  $Q$  band at  $1116\text{ cm}^{-1}$ . Therefore, it looks similar to  $Q^c$ .  $F^d$  has a bimodal signal at  $948\text{ cm}^{-1}$  and  $1013\text{ cm}^{-1}$ , akin to  $Q^d$ , in the lower part of the  $Q$  band, as well as a broad asymmetric peak at  $603\text{ cm}^{-1}$ . Finally, the new component  $F^e$  has a sharp peak at  $550\text{ cm}^{-1}$  and another one at  $1148\text{ cm}^{-1}$ , and cannot be compared to any of the components from the  $Q$  band deconvolution.

## 5. Discussion

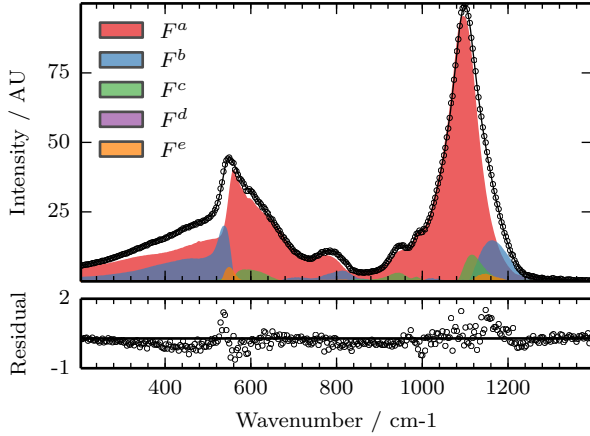
### 5.1. General considerations

When treating Raman spectra of silicate glasses, Herzog *et al.* [46] point out common deconvolution pitfalls such as:

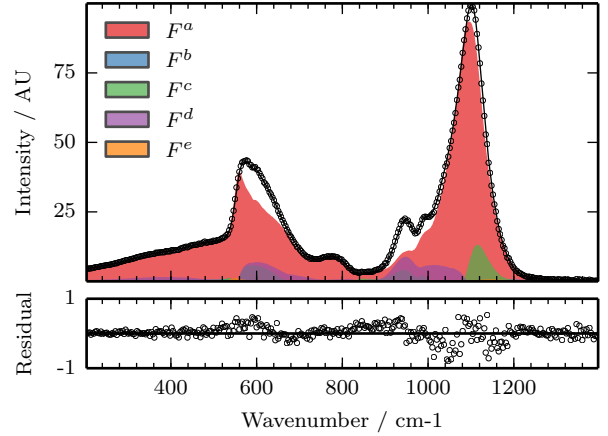
1. improper rank analysis, that often leads researchers to arbitrarily pick the rank.
2. wishful use of linear algebra tools to obtain components, in particular if a non-negativity constraint is not used. As a result, components may contain a significant number of negative points. Such an outcome is sometimes overlooked, even though it makes the extracted components unsuitable for interpretation as PRS.
3. poor noise robustness.

Our methods avoid these pit-falls through (1) the use of PCA and Malinowski’s IND function, (2) the alternative non-negative least square approach (ANLS) at the heart of the NMF and (3) an experimental design that enables us to acquire a set of numerous high-quality Raman spectra. Indeed, samples are stable at room temperature and can be probed for as long as necessary to obtain a low level of noise in the Raman spectra. Furthermore, thanks to the scale difference between the diffusion length (1 cm) and the EPMA/Raman probing spatial resolutions (1-20  $\mu\text{m}$ ) our experimental approach generates a high number of samples of variable composition over a spatially limited area. In addition, only a limited number of high temperature experiments is required, making diffusion an efficient way to generate compositional variations in situations where other parameters such as volatility (see Malfait *et al.* [44]) are low.

Compared to the classic Gaussian deconvolution, this method has the additional advantage of not requiring any previous knowledge of the glass structure to decide how many components should be used to fit, nor



(a) S side. EPMA analysis (wt%): 76.0 SiO<sub>2</sub>, 8.5 CaO, 15.5 Na<sub>2</sub>O. Area ratios: 80.0 %  $F^a$ , 15.2 %  $F^b$ , 3.8 %  $F^c$ , 0.0 %  $F^d$ , 1.0 %  $F^e$



(b) N side. EPMA analysis (wt%): 68.9 SiO<sub>2</sub>, 11.5 CaO, 19.6 Na<sub>2</sub>O. Area ratios: 85.2 %  $F^a$ , 0.5 %  $F^b$ , 5.8 %  $F^c$ , 8.4%  $F^d$ , 0.2 %  $F^e$

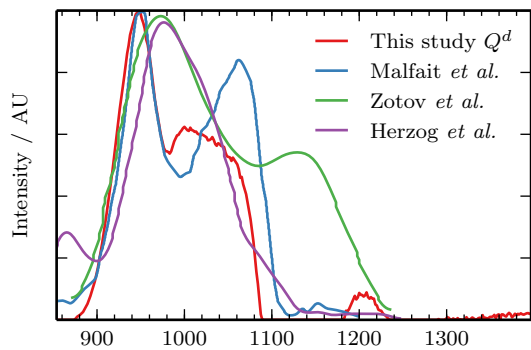
Figure 11: Reconstitution of two spectra from the SN sample. Circles are measured spectra, continuous lines are reconstitution from summing weighted PRS. Residuals are obtained by subtracting reconstitutions from the experimental data.

where and how wide they may be. On the down side, the purely mathematical nature of the approach does not make it possible to optimize the deconvolution based on *a priori* knowledge of the structure. It does, however, drastically reduce the number of free parameters of the model, which boils down to a variance threshold and the value of  $\eta$ . Fewer free parameters, and the sparsity constraint applied to the NMF, enhance robustness and reliability, while offering the possibility to compare PRS derived making no *a priori* assumptions with the results of alternative deconvolution strategies such as Gaussian fitting.

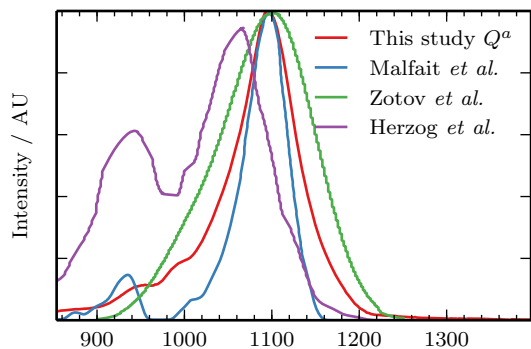
### 5.2. Endmember PRS, structural entities, and their variation as a function of composition

First of all, it is of interest to compare our PRS with those derived from experimental or theoretical studies of binary silicates, at least in the high wavenumber range ( $Q$  band) where this has been previously studied [27, 43, 46], as illustrated in Fig. 12. For example, based upon the arguments made in section 4.3, chemical correlation and shape comparison lead us to assign PRS  $Q^a$  and  $Q^c$  to the presence of  $Q^3$  species in our glasses. Indeed, Fig. 12b shows a remarkable agreement both in position and shape between Malfait's deconvolution of sodium silicate glasses, Zotov's VDOS calculations for NS<sub>4</sub> glass and  $Q^a$ , in spite of the presence of CaO in this study. In addition,  $Q^d$  agrees very well with the PRS that Malfait assigned to  $Q^2$  (Fig. 12a)

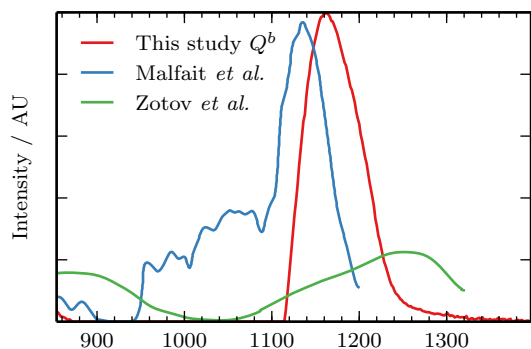
since its main peak is at the same position and its minor peak ends close to that of [44]. On the other hand, the small number of glass samples prepared by Herzog *et al.* in the lime-bearing system may explain that there is a clear shift between their partial Raman spectra and the PRS derived in this study. We also note that the  $Q^4$  component determined by Zotov *et al.* is significantly different from that determined in the present study, although our  $Q^4$  component is in good agreement with that derived experimentally by Malfait *et al.* [43, 46]. This discrepancy between calculated and experimental results may thus indicate that the limited number of atoms used in the simulation of the NS<sub>4</sub> glass, was not sufficient for accurate representation of  $Q^4$  units. Further support for our  $Q$  band assignments can be found in the comparison with composition. For example, using our assignments, an increase in depolymerized  $Q^2$  and  $Q^3$  species at the expense of  $Q^4$  tetrahedra is observed when moving away from SiO<sub>2</sub> in the SN and SC samples, as expected. For compositions characterized by NC exchange at constant bulk polymerization state, our data indicate that replacement of Na<sub>2</sub>O by CaO displaces the equilibrium  $2Q^3 \rightleftharpoons Q^2 + Q^4$  to the right, a result that is perfectly consistent with the fact that Ca<sup>2+</sup> is known to cluster non-bridging oxygens in  $Q^2$  rather than in  $Q^3$  [63]. This interpretation thus provides additional support for the idea that band  $Q^d$  is dominated by  $Q^2$  species associated with Ca, as indicated by the fact that its weight is positively correlated to CaO and negatively to SiO<sub>2</sub>, while not being correlated to Na<sub>2</sub>O



(a)  $Q^2$  PRS



(b)  $Q^3$  PRS



(c)  $Q^4$  PRS

Figure 12: Comparison of our  $Q$  band PRS with PRS extracted thanks to similar processing of sodium silicate glasses (Malfait *et al.* [44]), calcium silicate glasses (Herzog *et al.* [46]) and VDOS simulation of sodium silicate glasses (Zotov *et al.* [27])

(table 2). To summarize our results concerning the  $Q$  bands, we note that only 4 PRS are sufficient to produce high quality reconstructions of the spectra in this range of wavenumber. In particular, there is no need for shifts in the positions of the peaks to account for spectral information, in agreement with our basic assumption. Our PRS also respect the classic order where structural entities of increasing degree of polymerization give rise to intensity at higher wavenumber. However, we note that our PRS do not have simple shapes and are clearly not Gaussian, as expected given that they integrate all the vibrational information for a given Raman-active species. Thus,  $Q_{\text{avg}}^3$ ,  $Q_{\text{Na}}^3$  and  $Q^2$  contain symmetric and antisymmetric stretching bands.

A second issue of interest is the correlations between intensity in the  $Q$  band and that in the main band. Indeed, the close relationship between the high wavenumber part of the  $F^n$  partial spectra and the  $Q^n$  PRS obtained from a separate deconvolution provides strong evidence that a change in network polymerization impacts Si–O–Si bond vibrations that account for intensity in the main band. Such a correlation has been suggested in [64], in which the sensitivity of the main band as a function of network polymerization is also discussed. In detail, our PRS of the entire spectra indicate that the changes in glass structure associated with the presence of different  $Q$  species give rise to changes in intensity in the main band that may be relatively localized (*e.g.*  $F^b$ ). However, we recall the implicit assumption in our approach, that vibrational modes for a given structural entity do not shift in wavenumber with composition. Such shifts are not expected for spatially localized vibrations, such as those that give rise to intensity in the  $Q$  band, but the situation is less clear for inter-tetrahedral vibrations that occur in the main band. Studies combining IR, Raman and hyper-Raman spectroscopies have shown that the vibrations responsible for the main band in a-SiO<sub>2</sub> [65, 66] are better accounted for by Si–O–Si bending rather than tetrahedral  $E$ -bending. The former modes are sensitive to angle distribution, as exemplified by silica densification. Further work on Ca- and Na-aluminosilicates [67] show how the in-phase vibrations of network-modifying cations are responsible for part of the main Raman band. Consequently, they are likely to be affected by changes in overall network polymerization. Previous work on alkali silicates [26, 68] have either shown, in the case of  $Q^3$  species, or anticipated, in the case of  $Q^4$  and  $Q^2$  species, that bands shifts in the 400–850 cm<sup>-1</sup> range are as a matter of fact expected: as polymerization decreases, the mean Si–O bond length decreases, which does not change Si–NBO stretching modes but does im-

compact less localized vibrations like O–Si–NBO modes. We note that, should shifts appear, they would be deconvoluted as series of components, increasing then decreasing one after the other. This is not the case here: our data indicate that since compositional ranges are modest, main-band vibrations remain relatively constant and are strongly correlated to variations at higher wavenumber.

In terms of structural interpretation including the main band, we note that  $F^a$  accounts for over 80 % of the signal. Given the position of its peak in the  $Q$  band and its area ratio, it would appear to convey a  $Q^3$  Raman signal like  $Q^a$ . However, as noted above, the main PRS bears the overall response of an “average” glass network; the other PRS bear the deviation from it.  $F^c$  and  $F^e$  also exhibit peaks in both the main and the  $Q$  ranges. Their position, close to that of  $F^a$  peaks, suggests that they represent closely related species, or could be viewed as adjustment variables used by the NMF to account for the global band shifts. Based upon comparison with the detailed  $Q$  band deconvolution (section 4.3), the high wavenumber peak in  $F^b$  hints that it may convey  $Q^4$  signal because it appears at the highest wavenumbers.  $F^d$ 's bimodal signal in the lower part of the  $Q$  band points toward  $Q^2$  tetrahedra.  $F^e$  may represent the Si–O–Si network modification due to composition changes. However, more detailed interpretation of its shape and position is prevented by our limited compositional range.

### 5.3. Chemical regression

As noted above, the absolute intensity of PRS area ratios cannot be interpreted directly. Furthermore, our incomplete knowledge of the structural entities giving rise to the PRS also limits our capacity to quantitatively link them to composition. However, the strong linear correlation coefficients found in table 2 do suggest a clear link between PRS area ratios and glass chemistry. It is therefore of interest to explore a multilinear regression between Raman data and composition, with the aim of quantifying glass chemistry from Raman spectra, at least in the compositional range studied here. For a given spectrum  $\Sigma_i$ , PRS areas  $a_{ij}$  are obtained by integrating over the wavenumber range  $[\sigma_{\min}, \sigma_{\max}]$  the values of the partial spectra  $\mathbf{H}_j$  weighted by the corresponding  $w_{ij}$ :

$$a_{ij} = w_{ij} \int_{\sigma_{\min}}^{\sigma_{\max}} \mathbf{H}_j(\sigma) d\sigma \quad (9)$$

PRS area ratios  $\mathbf{R}$  are then derived:

$$r_{ij} = \frac{a_{ij}}{\sum_{l=1}^k a_{il}} \quad (10)$$

Since area ratios and EPMA measurements both add up to 1, dimensionality degenerates into a 3D (space of components) to 2D (space of oxides) regression. It is therefore not strictly necessary to regress over all the area ratios to all the oxides. However, for the sake of clarity and chemical sense, a multilinear regression is carried out from all 4 area ratios to all three oxides. Robustness is tested by regressing over all  $n$  data subsets obtained by removing one record. The following results stem from averaging the results of this leave-one-out procedure. Standard deviations are given in brackets.

$$\text{SiO}_2 \text{ (wt\%)} = 70.38(0.10)r_i^a + 124.26(0.62)r_i^b + 76.33(0.81)r_i^c + 52.40(0.57)r_i^d \quad (11)$$

$$\text{CaO (wt\%)} = 5.41(0.08)r_i^a + 36.02(0.48)r_i^b + 1.76(0.63)r_i^c + 85.06(0.48)r_i^d \quad (12)$$

$$\text{Na}_2\text{O (wt\%)} = 24.22(0.10)r_i^a - 60.28(0.59)r_i^b + 21.91(0.75)r_i^c - 37.46(0.56)r_i^d \quad (13)$$

This linear model <sup>1</sup> yields an RMS  $\ell_2$  distance between real and regressed composition of 0.88 wt% with a standard deviation of 0.50 wt%. Higher precision is obtained for CaO (RMS 0.43 wt%) and Na<sub>2</sub>O (RMS 0.49 wt%) than for SiO<sub>2</sub> (RMS 0.59 wt%). However, the error is of the order of magnitude of the EPMA measurements for Na<sub>2</sub>O in such a glass due to the migration of Na<sup>+</sup> under the electron beam.

Three test glasses were used to validate the regression inside and outside the learning domain (see Fig. 6). Raman spectra acquired on the test glasses are projected on the 4 previously determined  $Q$  band partial spectra using non-negative least squares. This procedure yields, for each test glass  $i$ , four weights  $w_i^a$  to  $w_i^d$ , which are converted to areas  $a_i^a$  to  $a_i^d$  according to eq. (9). Area ratios  $r_i^a$  to  $r_i^d$  are computed according to eq. (10), and are then used in eq. (11)-(13). Results are presented in Table 3 and compared to compositions obtained from wet chemical analysis. A good agreement between the actual and regressed compositions is found.

Regardless of the exact nature of the vibrational units responsible for the components, this linear regression model provides a simple and chemically justified tool to analyze soda-lime glasses in the glassmaking zone of interest. In detail, SiO<sub>2</sub> content is predicted to rise with increasing  $Q^4$  content, as expected.  $Q^3$  should be less

<sup>1</sup>Coefficients of the linear model do not necessarily have the same signs as the correlation table (Table 2), since correlation coefficients compare fluctuations around mean values, while the linear model links components fractions to oxides without subtracting the mean.

Table 3: Comparison of wet chemical analysis and Raman to chemistry regression. All amounts are in wt%

	TEST1			TEST2			TEST3		
	Chemistry	Raman	Discrepancy	Chemistry	Raman	Discrepancy	Chemistry	Raman	Discrepancy
SiO <sub>2</sub>	73.4	74.4	1.0	66.1	67.6	1.5	72.7	72.4	-0.3
CaO	10.4	12.1	1.7	16.3	18.2	1.9	13.7	15.3	1.6
Na <sub>2</sub> O	15.6	13.5	-2.1	17.2	14.3	-2.9	13.3	12.2	-1.1
Total	99.4	100.0	0.6	99.6	100.1	0.5	99.7	99.9	0.2

influential than  $Q^4$ , and  $Q^2$  even less than  $Q^3$ . Such behavior is indeed found in eq. (11). Eq. (12) describes a similarly logical evolution for the CaO proportion. Increase in  $Q^2$  content, and, to a lesser extent,  $Q^4$  content, would reflect a rise in CaO content.  $Q^c$  was assigned to a Na-rich  $Q^3$  and  $Q^a$  to an “average” background glass. It should come as no surprise that they have little influence in this equation. Slightly less understandable is the predicted negative impact of  $Q^4$  and  $Q^2$  on Na<sub>2</sub>O content. Although  $Q^4$  tetrahedra bear no sodium, its increase does not *per se* mean a decrease in Na<sub>2</sub>O content. The preference of Na for  $Q^3$  non-bridging oxygen may account partially for the negative coefficient with respect to  $Q^2$  content.

The accuracy of the regression is limited, since the mixing constraint  $\text{SiO}_2 + \text{CaO} + \text{Na}_2\text{O} = 100 \text{ wt}\%$  is not taken into account. This may result in regressed composition  $> 100 \%$  as exemplified by TEST2 glass in table 3. The apparent underestimation of Na<sub>2</sub>O and the overestimation of CaO seem to be coincidental. A systematic deviation would result in a non-zero average discrepancy, which is not found. Despite these shortcomings, chemical regression from Raman spectra provides a precise compositional estimate for glasses inside the regression domain (TEST1 and TEST3, inside the red domain in Fig. 6). A fair accuracy is obtained outside the domain (TEST2), though improvement is needed. A wider regression domain, or a non-linear mixing model may bring improvement in the future.

## Conclusion

The evolution of 169 Raman spectra collected on 3 diffusion couples of ternary SiO<sub>2</sub>-CaO-Na<sub>2</sub>O glasses has been studied using a novel deconvolution algorithm. Without using any prior knowledge of glass structure, the algorithm yields results in agreement with existing studies. Contrary to the classic Gaussian deconvolution, this method hardly requires any knowledge of the glass structure since it determines the number of components to seek from the data set, and can be applied to the full spectral range as well as to a specific one. Deconvo-

luting the high wavenumber  $Q$  band results in 4 partial Raman spectra; the evolution of their area ratios is correlated to compositional profiles along the diffusion lines. High absolute values of correlation coefficients enable a multilinear regression from PRS area ratios to oxide contents. These three regression equations provide a precise means of analysis within the training data set (standard deviation from real composition 0.5 wt%). They have been tested with three glasses inside and outside the regression domain, showing reasonable precision. At the present time, our deconvolution algorithm does not clearly discriminate Na- and Ca-compensated NBOs, although this may be possible in the future if glasses with a marked Na/Ca imbalance are included in the diffusion dataset. Expanding the compositional range would be helpful to better separate the different vibration entities in different partial Raman spectra. Further work, including especially <sup>29</sup>Si MAS-NMR, could bring more significance into the PRS in terms of structure. Extending the approach further, addition of other network-forming oxides such as Al<sub>2</sub>O<sub>3</sub> or B<sub>2</sub>O<sub>3</sub> may be more difficult to deal with if their vibrations overlap with those of SiO<sub>4</sub>, but the methods presented here provide the theoretical basis for the construction of models capable of calculating glass chemistry from Raman spectra.

## Acknowledgments

Sabrina Simoès do Nascimento and Pauline Bouhier are acknowledged for their patience and care in preparing the glasses. The authors also thank Odile Majérus and Matthieu Schwartz for their support in exploring the Raman spectrometer market, and fruitful discussions. Our gratitude also goes to Vera Pukhkaya and Isabelle Tannou for their feedback on the usability and limits of the Python implementation. This research was supported by the CNRS and Saint-Gobain Recherche.

## References

- [1] P. F. McMillan, G. H. Wolf, Vibrational spectroscopy of silicate liquids, *Reviews in Mineralogy and Geochemistry* 32 (1) (1995)

- 247–315.
- [2] D. R. Neuville, D. de Ligny, G. S. Henderson, Advances in Raman spectroscopy applied to earth and material sciences, *Rev. Mineral. Geochim.* 78 (1) (2014) 509–541.
  - [3] P. F. McMillan, B. T. Poe, P. Gillet, B. Reynard, A study of SiO<sub>2</sub> glass and supercooled liquid to 1950 K via high-temperature Raman spectroscopy, *Geochim. Cosmochim. Acta* 58 (17) (1994) 3653–3664.
  - [4] J. D. Frantz, B. O. Mysen, Raman spectra and structure of BaO-SiO<sub>2</sub>SrO-SiO<sub>2</sub> and CaO-SiO<sub>2</sub> melts to 1600°C, *Chem. Geol.* 121 (1) (1995) 155–176.
  - [5] I. Daniel, P. Gillet, B. T. Poe, P. F. McMillan, In-situ high-temperature Raman spectroscopic studies of aluminosilicate liquids, *Phys. Chem. Miner.* 22 (2) (1995) 74–86.
  - [6] D. R. Neuville, B. O. Mysen, Role of aluminium in the silicate network: In situ, high-temperature study of glasses and melts on the join SiO<sub>2</sub>-NaAlO<sub>2</sub>, *Geochim. Cosmochim. Acta* 60 (10) (1996) 1727–1737. doi:10.1016/0016-7037(96)00049-X.
  - [7] B. O. Mysen, Structure and properties of magmatic liquids: from haplobasalt to haploandesite, *Geochim. Cosmochim. Acta* 63 (1) (1999) 95–112. doi:10.1016/S0016-7037(98)00273-7.
  - [8] M. Toplis, B. Reynard, Temperature and time-dependent changes of structure in phosphorus containing aluminosilicate liquids and glasses: in situ Raman spectroscopy at high temperature, *J. Non-Cryst. Solids* 263 (2000) 123–131.
  - [9] Q. Zhao, M. Guerette, G. Scannell, L. Huang, In-situ high temperature Raman and Brillouin light scattering studies of sodium silicate glasses, *J. Non-Cryst. Solids* 358 (24) (2012) 3418–3426.
  - [10] O. N. Koroleva, V. N. Anfilogov, A. Shatskiy, K. D. Litasov, Structure of Na<sub>2</sub>O-SiO<sub>2</sub> melt as a function of composition: In situ Raman spectroscopic study, *J. Non-Cryst. Solids* 375 (0) (2013) 62–68. doi:10.1016/j.jnoncrysol.2013.04.060.
  - [11] P. Richert, B. Mysen, D. Andraut, Melting and premelting of silicates: Raman spectroscopy and X-ray diffraction of Li<sub>2</sub>SiO<sub>3</sub> and Na<sub>2</sub>SiO<sub>3</sub>, *Phys. Chem. Miner.* 23 (3) (1996) 157–172. doi:10.1007/BF00220727.
  - [12] P. Colomban, Polymerization degree and Raman identification of ancient glasses used for jewelry, ceramic enamels and mosaics, *J. Non-Cryst. Solids* 323 (1) (2003) 180–187.
  - [13] P. Colomban, A. Tournié, L. Bellot-Gurlet, Raman identification of glassy silicates used in ceramics, glass and jewellery: a tentative differentiation guide, *J. Raman Spectrosc.* 37 (8) (2006) 841–852.
  - [14] L. Robinet, A. Bouquillon, J. Hartwig, Correlations between Raman parameters and elemental composition in lead and lead alkali silicate glasses, *J. Raman Spectrosc.* 39 (5) (2008) 618–626.
  - [15] A. Quaranta, A. Rahman, G. Mariotto, C. Maurizio, E. Trave, F. Gonella, E. Cattaruzza, E. Gibaudo, J. Broquin, Spectroscopic investigation of structural rearrangements in silver ion-exchanged silicate glasses, *J. Phys. Chem. C* 116 (5) (2012) 3757–3764. doi:10.1021/jp2095399.
  - [16] T. Seuthe, M. Grehn, A. Mermillod-Blondin, H. Eichler, J. Bonse, M. Eberstein, Structural modifications of binary lithium silicate glasses upon femtosecond laser pulse irradiation probed by micro-Raman spectroscopy, *Opt. Mater. Express* 3 (6) (2013) 755–764.
  - [17] T. Deschamps, C. Martinet, J. L. Bruneel, B. Champagnon, Soda-lime silicate glass under hydrostatic pressure and indentation: a micro-Raman study, *J. Phys. Condens. Matter* 23 (3) (2011) 035402.
  - [18] A. Perriot, D. Vandembroucq, E. Barthel, V. Martinez, L. Grosvalet, C. Martinet, B. Champagnon, Raman microspectroscopic characterization of amorphous silica plastic behavior, *J. Am. Ceram. Soc.* 89 (2) (2006) 596–601.
  - [19] J. F. Stebbins, Identification of multiple structural species in silicate glasses by <sup>29</sup>Si NMR, *Nature* 330 (6147) (1987) 465–467.
  - [20] T. Furukawa, K. E. Fox, W. B. White, Raman spectroscopic investigation of the structure of silicate glasses. III. Raman intensities and structural units in sodium silicate glasses, *Chem. Phys.* 75 (7) (1981) 3226–3237. doi:10.1063/1.442472.
  - [21] P. F. McMillan, Structural studies of silicate glasses and melts: Applications and limitations of Raman spectroscopy, *Am. Mineral.* 69 (1984) 622–644.
  - [22] P. McMillan, B. Piriou, A. Navrotsky, A Raman spectroscopic study of glasses along the joins silica-calcium aluminate, silica-sodium aluminate, and silica-potassium aluminate, *Geochim. Cosmochim. Acta* 46 (11) (1982) 2021–2037. doi:10.1016/0016-7037(82)90182-X.
  - [23] D. R. Neuville, Viscosity, structure and mixing in (Ca, Na) silicate melts, *Chem. Geol.* 229 (1-3) (2006) 28–41. doi:10.1016/j.chemgeo.2006.01.008.
  - [24] T. Furukawa, W. B. White, Raman spectroscopic investigation of sodium borosilicate glass structure, *J. Mater. Sci.* 16 (10) (1981) 2689–2700.
  - [25] D. Manara, A. Grandjean, D. Neuville, Advances in understanding the structure of borosilicate glasses: a Raman spectroscopy study, *Am. Mineral.* 94 (5-6) (2009) 777–784.
  - [26] D. W. Matson, S. K. Sharma, J. A. Philpotts, The structure of high-silica alkali-silicate glasses. A Raman spectroscopic investigation, *J. Non-Cryst. Solids* 58 (2-3) (1983) 323–352. doi:10.1016/0022-3093(83)90032-7.
  - [27] N. Zotov, I. Ebbsjö, D. Timpel, H. Keppler, Calculation of Raman spectra and vibrational properties of silicate glasses: Comparison between Na<sub>2</sub>Si<sub>4</sub>O<sub>9</sub> and SiO<sub>2</sub> glasses, *Phys. Rev. B* 60 (9) (1999) 6383–6397. doi:10.1103/PhysRevB.60.6383.
  - [28] G. Spiekermann, M. Steele-MacInnis, C. Schmidt, S. Jahn, Vibrational mode frequencies of silica species in SiO<sub>2</sub>-H<sub>2</sub>O liquids and glasses from ab initio molecular dynamics, *J. Chem. Phys.* 136 (15) (2012) 154501. doi:10.1063/1.3703667.
  - [29] A. Rahmani, M. Benoit, C. Benoit, Signature of small rings in the Raman spectra of normal and compressed amorphous silica: A combined classical and ab initio study, *Phys. Rev. B* 68 (18) (2003) 184202. doi:10.1103/PhysRevB.68.184202.
  - [30] B. O. Mysen, L. W. Finger, D. Virgo, F. A. Seifert, Curve-fitting of Raman spectra of silicate glasses, *Am. Mineral.* 67 (7-8) (1982) 686–695.
  - [31] D. R. Neuville, L. Cormier, D. Massiot, Al environment in tectosilicate and peraluminous glasses: a <sup>27</sup>Al MQ-MAS NMR, Raman, and XANES investigation, *Geochim. Cosmochim. Acta* 68 (24) (2004) 5071–5079. doi:10.1016/j.gca.2004.05.048.
  - [32] B. Parkinson, D. Holland, M. E. Smith, C. Larson, J. Doerr, M. Affatigato, S. A. Feller, A. Howes, C. Scales, Quantitative measurement of Q<sup>3</sup> species in silicate and borosilicate glasses using Raman spectroscopy, *J. Non-Cryst. Solids* 354 (17) (2008) 1936–1942. doi:10.1016/j.jnoncrysol.2007.06.105.
  - [33] B. O. Mysen, A. Lucier, G. D. Cody, The structural behavior of Al<sup>3+</sup> in peralkaline melts and glasses in the system Na<sub>2</sub>O-Al<sub>2</sub>O<sub>3</sub>-SiO<sub>2</sub>, *Am. Mineral.* 88 (11-12) (2003) 1668–1678.
  - [34] C. Le Losq, R. Moretti, D. R. Neuville, Speciation and amphoteric behaviour of water in aluminosilicate melts and glasses: high-temperature Raman spectroscopy and reaction equilibria, *Eur. J. Mineral.* 25 (5) (2013) 777–790.
  - [35] D. Macnaughtan, L. B. Rogers, G. Wernimont, Principal-component analysis applied to chromatographic data, *Anal. Chem.* 44 (8) (1972) 1421–1427. doi:10.1021/ac60316a016.
  - [36] A. Hodroj, P. Simon, P. Florian, M.-H. Chopinet, Y. Vaills, Phase separation and spatial morphology in sodium silicate glasses by AFM, light scattering and NMR, *J. Am. Ceram. Soc.* 96 (8) (2013) 2454–2460.

- [37] Y. B. Monakhova, S. A. Astakhov, A. Kraskov, S. P. Mush-takova, Independent components in spectroscopic analysis of complex mixtures, *Chemometr. Intell. Lab.* 103 (2) (2010) 108–115. doi:10.1016/j.chemolab.2010.05.023.
- [38] P. O. Hoyer, Non-negative matrix factorization with sparseness constraints, *J. Mach. Learn. Res.* 5 (2004) 1457–1469.
- [39] H. Li, T. Adal, W. Wang, D. Emge, A. Cichocki, A. Cichocki, Non-negative Matrix Factorization with Orthogonality Constraints and its Application to Raman Spectroscopy, *J VLSI Sign Process Syst Sign Im* 48 (1-2) (2007) 83–97. doi:10.1007/s11265-006-0039-0.
- [40] F. Guimet, R. Boqué, J. Ferré, Application of non-negative matrix factorization combined with Fisher’s linear discriminant analysis for classification of olive oil excitation–emission fluorescence spectra, *Chemometr. Intell. Lab.* 81 (1) (2006) 94–106.
- [41] M. Cuny, E. Vigneau, G. L. Gall, I. Colquhoun, M. Lees, D. N. Rutledge, Fruit juice authentication by <sup>1</sup>H NMR spectroscopy in combination with different chemometrics tools, *Anal. Bioanal. Chem.* 390 (1) (2008) 419–427. doi:10.1007/s00216-007-1708-y.
- [42] N. Gillis, R. J. Plemmons, Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis, *Linear Algebra Appl.* 438 (10) (2013) 3991–4007. doi:10.1016/j.laa.2012.04.033.
- [43] W. J. Malfait, V. P. Zakaznova-Herzog, W. E. Halter, Quantitative Raman spectroscopy: High-temperature speciation of potassium silicate melts, *J. Non-Cryst. Solids* 353 (44–46) (2007) 4029–4042. doi:10.1016/j.jnoncrysol.2007.06.031.
- [44] W. J. Malfait, V. P. Zakaznova-Herzog, W. E. Halter, Amorphous materials: Properties, structure, and durability Quantitative Raman spectroscopy: Speciation of Na-silicate glasses and melts, *Am. Mineral.* 93 (10) (2008) 1505–1518. doi:10.2138/am.2008.2783.
- [45] W. J. Malfait, Quantitative Raman spectroscopy: speciation of cesium silicate glasses, *J. Raman Spectrosc.* 40 (12) (2009) 1895–1901. doi:10.1002/jrs.2338.
- [46] F. Herzog, V. P. Zakaznova-Herzog, Quantitative Raman spectroscopy: Challenges, shortfalls, and solutions—Application to calcium silicate glasses, *Am. Mineral.* 96 (5-6) (2011) 914–927. doi:10.2138/am.2011.3508.
- [47] A. Fluegel, D. A. Earl, A. K. Varshneya, T. P. Seward, Density and thermal expansion calculation of silicate glass melts from 1000°C to 1400°C, *Phys. Chem. Glasses B* 49 (5) (2008) 245–257.
- [48] D. A. Long, *Raman spectroscopy*, McGraw-Hill, New York, 1977.
- [49] B. O. Mysen, D. Virgo, I. Kushiro, The structural role of aluminum in silicate melts; a Raman spectroscopic study at 1 atmosphere, *Am. Mineral.* 66 (7-8) (1981) 678–701.
- [50] F. A. Seifert, B. O. Mysen, D. Virgo, Three-dimensional network structure of quenched melts (glass) in the systems SiO<sub>2</sub>-NaAlO<sub>2</sub>, SiO<sub>2</sub>-CaAl<sub>2</sub>O<sub>4</sub> and SiO<sub>2</sub>-MgAl<sub>2</sub>O<sub>4</sub>, *Am. Mineral.* 67 (7-8) (1982) 696–717.
- [51] J. D. Hunter, Matplotlib: a 2d graphics environment, *Comput. Sci. Eng.* 9 (3) (2007) 90–95. doi:10.1109/MCSE.2007.55.
- [52] T. E. Oliphant, Python for Scientific Computing, *Comput. Sci. Eng.* 9 (3) (2007) 10–20. doi:10.1109/MCSE.2007.58.
- [53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [54] E. R. Malinowski, Theory of error in factor analysis, *Anal. Chem.* 49 (4) (1977) 606–612. doi:10.1021/ac50012a026.
- [55] E. R. Malinowski, Determination of the number of factors and the experimental error in a data matrix, *Anal. Chem.* 49 (4) (1977) 612–617. doi:10.1021/ac50012a027.
- [56] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, R. J. Plemmons, Algorithms and applications for approximate non-negative matrix factorization, *Comput. Stat. Data An.* 52 (1) (2007) 155–173. doi:10.1016/j.csda.2006.11.006.
- [57] D. Donoho, V. Stodden, When does non-negative matrix factorization give a correct decomposition into parts?, in: *Advances in neural information processing systems*, 2003, p. None.
- [58] C. Boutsidis, E. Gallopoulos, SVD based initialization: A head start for nonnegative matrix factorization, *Pattern Recogn.* 41 (4) (2008) 1350–1362. doi:10.1016/j.patcog.2007.09.010.
- [59] E. J. Candès, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.* 59 (8) (2006) 1207–1223. doi:10.1002/cpa.20124.
- [60] E. Segnit, Further data on the system Na<sub>2</sub>O-CaO-SiO<sub>2</sub>, *Am. J. Sci.* 251 (8) (1953) 586–601.
- [61] Y. Zhang, A modified effective binary diffusion model, *J. Geophys. Res.* 98 (B7) (1993) 11901–11920. doi:10.1029/93JB00422.
- [62] B. O. Mysen, D. Virgo, C. M. Scarfe, Relations between the anionic structure and viscosity of silicate melts—a Raman spectroscopic study, *Am. Mineral.* 65 (1980) 690–710.
- [63] A. R. Jones, R. Winter, G. N. Greaves, I. H. Smith, MAS NMR study of soda-lime-silicate glasses with variable degree of polymerisation, *J. Non-Cryst. Solids* 293–295 (2001) 87 – 92. doi:10.1016/S0022-3093(01)00656-1.
- [64] N. Trcera, S. Rossano, M. Tarrida, Structural study of Mg-bearing sodosilicate glasses by Raman spectroscopy, *J. Raman Spectrosc.* 42 (4) (2011) 765–772. doi:10.1002/jrs.2763.
- [65] B. Hehlen, G. Simon, The vibrations of vitreous silica observed in hyper-Raman scattering, *J. Raman Spectrosc.* 43 (12) (2012) 1941–1950. doi:10.1002/jrs.4110.
- [66] B. Hehlen, Inter-tetrahedra bond angle of permanently densified silicas extracted from their Raman spectra, *J. Phys.: Condens. Matter* 22 (2) (2010) 025401. doi:10.1088/0953-8984/22/2/025401.
- [67] B. Hehlen, D. R. Neuville, Raman Response of Network Modifier Cations in Alumino-Silicate Glasses, *J. Phys. Chem. B* 119 (10) (2015) 4093–4098. doi:10.1021/jp5116299.
- [68] X. Xue, J. F. Stebbins, M. Kanzaki, P. McMillan, B. T. Poe, Pressure-induced silicon coordination and tetrahedral structural changes in alkali oxide silica melts up to 12 GPa: NMR, Raman, and infrared spectroscopy, *Am. Mineral.* 76 (1-2) (1991) 8–26.