Combining deep learning and thermodynamics for trans-theoretical prediction of material properties: application to aluminosilicate glasses and melts

[1,2,3,\*]Charles Le Losq, [2]Daniel R. Neuville, [3]Bjorn O. Mysen

[1] Research School of Earth Sciences, Australian National University, Canberra, 2601, Australia

[2] Institut de Physique du Globe de Paris, Université de Paris, France, 75005, France

[3] Geophysical Laboratory, Carnegie Institution for Science, Washington D.C., 20001, U.S.A.

\* Corresponding Author : [lelosq@ipgp.fr](mailto:lelosq@ipgp.fr)

Keywords: Glass, Liquid, Viscosity, Densisty, Structure, Raman, neural networks, deep learning

**Abstract**

Predicting the properties of materials like minerals, liquids or glasses represent a critical goal in many disciplines. Among them, geo- and material sciences share a common interest for the properties of aluminosilicate melts. Indeed, the latter constitute most of Earth's volcanic magmas and lavas, and can be quenched to produce technological glasses like hand-held devices' screens. While several avenues allow property predictions, an extended core model linking chemical composition to structure and property measurements, allowing systematic and extended exploration of those links, do not exist to date. Focusing on alkali aluminosilicate melts, we here show how thermodynamic theories can be combined with a core deep neural network to build a framework for prediction of liquid properties like viscosity as well as of glass properties like density, or even the glass Raman signals. Such approach is a new tool to unravel the links between chemistry, structure and properties of materials.

**1. Introduction**

The prediction of the properties of materials represents a cornerstone for many disciplines, which rely on such predictions to solve problems related to material engineering and manufacturing, or to address key questions in the area of physics, chemistry and geosciences. Particular efforts have been deployed to understand how the chemical composition of aluminosilicate melts and glasses affect their structure, and, in turn, their thermodynamic and dynamic properties (Mysen and Richet, 2005; Le Losq et al., 2019b). Indeed, those materials form the liquid part of magmas, and their viscosity (resistance to movement) directly influence the effusive or explosive character of volcanic eruptions (Dingwell, 1996; Gonnermann et al., 2013; Gonnermann, 2015), and, hence, influence risk mitigation during eruptive crisis. Furthermore, aluminosilicate glasses are used to manufacture technological glasses used in many areas. Knowledge and prediction of the glass and melt properties is thus key to solve many material and Earth sciences problems.

Most models of the viscosity (*η*) of aluminosilicate melts rely on parametric modeling of the link between melt chemical composition and viscosity (Bottinga and Weill, 1972; Shaw, 1972, p. 19; Persikov, 1991; Hui and Zhang, 2007; Giordano et al., 2008; Persikov and Bukhtiyarov, 2009; Mauro et al., 2013). For instance, it is possible to express the fitting parameters A, B and C of the empirical Tamman-Vogel-Fulcher equation:

, (1)

with T the temperature, as a function of polynomial functions that depend on melt composition (Giordano et al., 2008). Such model are either focused on geological melts (e.g. Hui and Zhang, 2007; Giordano et al., 2008) or on industrial compositions (Mauro et al., 2013). Contrary to this approach, recent works demonstrated the possibility of developing models that link melt composition, structure, and properties. Melt structure can be described as a disrupted network of interconnected tetrahedral silica units. Interestingly, at the glass transition, melt structure is frozen-in such that it can be observed in glasses via the analysis of 29Si Nuclear Magnetic Resonance or Raman spectra (e.g. H. Maekawa et al., 1991; Mysen, 1995). Such knowledge allows calculating melt thermodynamic properties at the glass transition like its configurational entropy at the glass transition temperature Tg, Sconf(Tg), or its configurational heat capacity Cpconf (Mysen, 1995; Le Losq and Neuville, 2017). Such knowledge can then be leverage using theories like that of Adam and Gibbs (1965) to calculate melt viscosity from

, (2)

where *Ae* is a pre-exponential term proportional to the viscosity at infinite temperature, and *Be* is a constant proportional to the energy barriers to the cooperative re-arrangement of melt structure. We must note that the Adam and Gibbs (AG) theory relies on viscous flow occuring through cooperative re-arrangement of molecular subunits, and such events have been identified via high-temperature 29Si NMR spectroscopy (Stebbins, 1991) or direct observations (Huang et al., 2013).

Using this theory, (Le Losq and Neuville, 2017) proposed a model expressing Sconf(Tg) and *Be* via the knowledge of glass structure, which allows predicting η in Na2O-K2O-SiO2 melts with an unrivaled precision of 0.18 log Pa · s. However, extension to aluminosilicate compositions, representative of most natural and industrial glasses, encounters several major difficulties. 29Si NMR spectroscopy becomes blind in aluminosilicate glasses due to Si-Al interactions resulting in significant signal broadening; signal interpretation relies on various hypothesis and back-end model of melt structure (e.g. Mysen et al., 2003). Raman spectroscopy does not solve this problem as it cannot be calibrated against reliable NMR data to distinguish the molecular subunits in the glasses. Furthermore, the aluminium content of the melt also affect interactions between, and the environment of the metal cations, as well as Al-Si ordering or Al coordinance (Mysen et al., 2003; Neuville et al., 2006; Le Losq et al., 2014; Lee et al., 2016, p. 201; Le Losq and Neuville, 2017; Le Losq et al., 2017). Such problems severely affect our abilities to construct models in presence of aluminum, and question the theoretical viability of proposed models based on unfiltered data (Starodub et al., 2019).

Aside such problems, one must also notice that, while eq. (2) is proposed as a plausible model for silicate melt viscosity, other theories or modification of the Adam-Gibbs theory exist. For instance, eq. (2) has been modified by Mauro et al. (2009) to propose the MYEGA equation, which can be then used to predict viscosity only from the melt fragility *m* (see this public for fragility) and *Tg*:

**Eq MYEGA mod** . (3)

This equation is in some way similar to that proposed by Avramov and Milchev (1988), which also predicts viscosity from melt fragility:

**Eq A-M** (4)

Eq. (4) has different theoretical grounds compared to eq. (2), XXX. Arguments exist to support different equations and theories for different liquids (REF) or even different temperature range (Bottinga et al., 1995). However, and despite the Adam-Gibbs approach is usually the one recommended for silicate melts, we must consider that there is no totally accepted consensus behind one or the other theories. With this in mind, any modelling using a particular theory makes already a strong leap of faith toward a particular view of silicate melts, and more generally liquids. Trans-theoretical modelling may be a way to solve such problem, but traditionally has been impossible to implement as a single, unique model.

We address such issues in this study, presenting a model where a core neural network is used to predict (i) the melt viscosity via different theories, as well as the glass (ii) density, (iii) glass transition temperature, (iv) fragility, and (v) Raman spectra. The network was trained via multitask learning, employing early-stopping as well dropout to promote generalisation while training on a relatively restrained dataset (1753 samples, XXX compositions). Results demonstrate the viability of this approach, which offers a new tool for systematic and quantitative exploration of the links between material chemistry, structure and properties.

**2. Material and methods**

**2.1 Data compilation**

Existing data regarding the density and viscosity of alkali aluminosilicate glasses were selected by hand via a review of the existing literature. Cross-verification of the accuracy of viscosity data from different authors was checked on compositions as sodium trisilicate, albite and jadeite. Publications presenting deviations higher than 0.1 log Pa · s on such compositions were entirely discarded. Silica viscosity data were selected following the recommendation of Doremus (2002).

**2.2 Sample synthesis and new measurements**

New glass samples were synthesized at IPGP in Paris from reagent-grade K2CO3, Na2CO3 and SiO2 dried oxide powders to complete the dataset following the protocol described in (Le Losq et al., 2014; Le Losq and Neuville, 2017; Le Losq et al., 2017). Viscosity and density measurements follow the protocol described in those articles, such that the reader is refereed to those publications for further information. Chemical compositions have been measured using a Cameca SX50 electron microprobe (Table 1), with a 30 nA current, U=30kV, and 5s of counting. The values reported in Table 1 are the statistical mean of 10-20 individual measurements.

**2.3 Raman spectroscopy**

Raman spectra of silicate and aluminosilicate glasses acquired at IPGP in Paris were recorded using a T64000 Jobin-Yvon® Raman spectrometer equipped with a confocal system, a 1024 charge-couple detector (CCD) cooled by liquid nitrogen and an Olympus® microscope. The optimal spatial resolution allowed by the confocal system is 1-2 μm2 with a x100 Olympus® objective, and the spectral resolution is 0.7 cm-1. A Coherent® laser 70-C5 Ar+, having a wavelength of 488.1 or 514.532 nm, has been used as the excitation line. Unpolarized Raman spectra were acquired between 20 and 1500 cm-1 on pieces of glass from the starting materials that were excited with a laser power of 100 to 150 mW on the sample.

Further Raman spectra acquired at the Geophysical Laboratory on glasses along the K2Si4O9-K2(KAl)4O9 and K2Si4O9-K2(KAl)4O9 joins (Mysen, 1996, 1999) were added to tha database. Those spectra were acquired with a Dilor XY confocal microRaman spectrometer equipped with a cryogenic Thompson Model 4ooO CCD . The 488 nm line of a SpectraPhysics model 2025 Ar+ laser operating at several hundred mW at the sample was used for sample excitation.

Preprocessing of the spectra was kept to minimum: (i) a linear baseline was fitted to the minima in the 700-800 and 1200-1300 cm-1 portions of the spectra and then subtracted to obtained baseline-corrected spectra, (ii) the spectra were corrected from temperature and excitation line effects (see eq. 1 in Le Losq et al., 2014), and (iii) the spectra were normalised to their maximum intensity such that the intensity in each spectrum varies between 0 and 1 (min-max scaling in data science terminology). Only signals in this range were saved as different spectra had different starting and ending Raman shift values. After pre-processing, spectra were saved in a HDF5 file for their future use.

**2.4 Machine learning modelling**

*2.4.1 Datasets*

Data used in this study are available in a spreadsheet (see supplementary materials) that was used as a database. Three different streams of data are present:

- *Dviscosity*, the database of viscosity measurements, composed of *Xviscosity* chemical composition entries (mole fractions) as well as their associated temperatures (Kelvin) and *yviscosity* observations (Pa·s);

- *Ddensity*, the database of density measurements, composed of *Xdensity* chemical composition entries (mole fractions) and *ydensity* observations (g cm-3);

- *DRaman*, the database of Raman spectra, composed of *XRaman* chemical composition entries (mole fractions) and *yRaman* spectra observations (min-max scaled Raman intensities).

The idea of using three different set of obsetvations (viscosity, density and Raman spectra) was to leverage the fact that neural networks learning multiple tasks show better predictions capacities compared to those trained to learn only one task (Caruana, 1997). *Dviscosity* and *Ddensity* cover an important part of the glass-forming domain of alkali aluminosilicates (Figure 1); they were thus used to train the artificial neural network to be able to make precise predictions of density and chemistry given composition. *DRaman* covers a more limited set of compositions (Figure 1). It was used as a way of improving multitask learning as well as a philosophical way to introduce structural information in the model.

*2.4.2 Train-test splitting and standardisation*

As machine learning algorithms are powerful interpolators, one key goal when training them is to avoid overfitting. The latter corresponds to the case where the algorithm predictions on the dataset used for training are excellent, but predictions on new, unseen data are poor. This indicates that the algorithm basically encoded the inputs-outputs relationships from the training data subset, and did not capture the mathematical relationships between inputs and outputs. For artificial neural networks, overfitting is usually monitored by splitting the available dataset in three different, randomly chosen *training*, *validation* and *testing* data subsets (Figure 1). The *training* subset is used for training the neural network while the *validation* subset is used for monitoring overfit during the training. This allows adjusting hyperparameters to avoid this overfit (i.e. parameters of the algorithm that control the size of the network, the learning rate, etc.), and to select potential candidates for a final training step. The final predictive capacity of the trained neural networks are then evaluated using the unseen *testing* data subset. In the present study, we followed the approached developed in Le Losq et al. (2019a), which consist in separating randomly the data by composition. As for each chemical composition several viscosity measurements were done at different temperatures, separating them completely randomly will result in having viscosity datapoints from the same composition in the different data subsets, and hence introducing a bias that results in over-estimating the predictive ability of the network. Random separation by composition avoids this pitfall. The viscosity and density datasets were separated in three splits following this rule (Figure 1). The Raman dataset was divided in only two *train* and *validation* subsets due to its small size. This is not a problem as we do not expect precise predictions of Raman spectra but uses this dataset as a way to improve the predictive capacity of the trained neural network. After their separation, the different subsets were kept in HDF5 files for their future use in the model.

After train-validation-test splitting, the important step in any machine learning data preprocessing is standardization of the data. Indeed, convergence of the machine learning algorithms strongly depends on the scaling of the data, i.e. in having features in the data that follow a normal distribution or are scaled between 0 and 1 (e.g., see Murphy, 2012). Such scaling promotes feature variations close to unity and prevents having different features covering very different numerical ranges; as such, it helps gradient back-propagation that is key for training neural networks.

In the present study, we have implemented a custom approach. All chemical compositions inputs are in mole fractions, which corresponds to a modification of min-max (0-1) scaling. After testing their scaling or leaving them as mole fraction, the later approach was adopted as it does not influence training. Raman spectra were normalised to be comprised between 0 and 1. Viscosity and density were not scaled, as they are not directly calculated from the neural network outputs (section 2.4.3), and scaling does not affect network convergence.

*2.4.3 Implementation*

The model presented in figure 2 was implemented in Python using the Pytorch library. It consists in inputting four inputs, i.e. the mole fractions of SiO2, Al2O3, Na2O and K2O, in a neural network composed of *n* hidden layers with *k* activation units (a.k.a neurons) each (square architecture; Figure 2). Different tests led to choose the now popular rectifier function (Glorot et al., 2011), which for an input *x* give the output *y = max(0,x)*, for the functions of the activation units. The outputs of this core network were fed to two final linear layers (Figure 2). The first output layer returns vectors of Raman spectra calculated from the linear sum of the output layer’ inputs. The second output layer returns 6 different values:

- *Sconf(Tg)*, *Tg*, the melt fragility *m* and the glass density *d* are calculated from the exponential of the four first linear outputs;

- the parameters *Ae* as well as *Aa-m* (**eqs.** **……)** are calculated from the last two linear outputs.

The use of the exponential function there was inspired by the strategy proposed by (Bishop, 2006) to avoid negative values in outputs of neural networks for Mixture Density Networks. In the present case, it prevents thermodynamic and static physical quantities to take impossible negative values. It further helps for rapid convergence during training. After their calculations, *Sconf(Tg)*, *Tg*, *m,* *Ae* and *Aa-m* are used in the equations 3-5 for final predictions of melt viscosity.

The architecture of the hidden layers was optimized by performing a global random search (Bergstra and Bengio, 2012). This allowed us to observe how the number of layers and hidden units affect the generalization ability of the model (Figure 3). After training more than XXXX models, we observed than moderately deep networks with ~5 layers and 200-300 units per layer perform best (Fig. 3A,B). Best performance are in general reach with more than 1000 neurons (Fig. 3C). Using elevated dropout, a technique that consists in turning off *p* percent of neurons per layer at each training iteration in order to prevent overfitting, also helps preventing overfitting (Fig. 3D, see section 2.4.4 for explanation). From this random search, we were able to select the model with the lowest error on the validation data subset. This model is composed of XXXXXXX. It is used in all predictions reported in this publication.

*2.4.4 Training*

The least-square deviations between the viscosity predictions from eq. 3-5 as well as density and Raman spectra predictions were used as a metric of goodness of fit. We further added loss functions for known viscous *Tg* and *Sconf(Tg)* values in the dataset *Dviscosity*. This allowed better constraining calculations of *Sconf(Tg)*. This parameter is usually hard to calculate as solutions of eq. 2 are multiple, given the extreme correlation between *Be* and *Sconf(Tg)* originating from the intervention of the intrinsic entropy *S\** in both *Be* and *Sconf(Tg)* (Adam and Gibbs, 1965). The present approach was used as a way to make the network less sensitive to this correlation. This also is why the network does not predict directly *Be* ; this term is calculated in the model from

**Eq BE** , (5)

Batch training was performed monitoring the global loss on the *training* and *validation* data subsets. Early stopping was performed to avoid overfitting: when the loss function on the *validation* data subset for more than 50 epochs, training was stopped and the network presenting the best validation loss was saved. Overfitting was further prevented by using dropout (Srivastava et al., 2014), i.e. at each training iteration a fraction *p* of the activation units was not used for predictions. From the XXXX random runs, it is best to set *p* at high values for preventing overfitting in general (Fig. 3D). In the saved best model, *p* was of XXX.

**3. Results**

**3.1 Viscosity predictions**

Viscosity is predicted by eqs. 2 to 4 within 0.6 log Pas (Table 1). Errors are slightly higher in the supercooled domain (viscosity > 107 Pas) for all theories (Fig. 4). We note that the Adam-Gibbs and Avramov-Milchev equations perform similarly whereas the MYEGA equation returns predictions affected by higher errors (Table 1).

The predictive ability of the model directly depends on the size of the dataset (Fig. 5), as can be expected given the well-known sensitivity of neural networks’ generalization abilities to dataset size. Figure 5 shows that, for the present problem composed of four input chemical variables, more than 100 different compositions in the training data are necessary to obtain RMSE values lower than 0.6 log Pas on the testing data subset; we consider in this publication that a predictive RMSE value of 0.6 log Pas on the full viscosity range 10-2-1016 Pas as a good achievement given that it will be lower than predictive RMSE claimed by existing models like those of (Giordano et al., 2008) or (Hui and Zhang, 2007). Lower numbers of compositions in the *training* data subset result in the model constantly overfitting the data. Interestingly, in figure 5 the training RMSE is always of ~0.36 ±0.04 log Pas, regardless of the training dataset size. Artificial neural networks are extremely flexible and very prone to overfit their training dataset. Therefore, the fact that models trained on a small number of compositions clearly overfit the training dataset but do not provide viscosity predictions better than ~0.36 log Pas brings an important information: this places the lowest achievable RMSE at ~0.36 log Pas on the present dataset. This limit probably arises from the combination of experimental errors affecting viscosity and chemical composition values.

**3.2 Thermodynamic and static property predictions**

The interest of the present model is not pure viscosity calculations, but the possibility to calculate their thermodynamic properties of aluminosilicate melts, or even the density of their glasses. Multitask learning clearly improves the predictive abilities of the artificial neural network and prevents overfitting (Figure 6). Besides, the model can now predict *Tg*, *Sconf(Tg)* and the density of alkali aluminosilicate glasses within XX K, XX J mol-1 K-1 and XX g cm-3, respectively (RMSE on *testing* dataset; see Figure 7A,B,C). Its ability to predict the melt fragility *m* further allows comparing it to the ratio between the configurational heat capacity at *Tg* and *Sconf*(*Tg), Cpconf(Tg)/Sconf(Tg)* (Fig. 7D). We detect a linear trend between those quantities, as expected from the works of Webb (2008) and Russell and Giordano (2017). However, contrary to the latter studies, the present predictions show a larger scattering between *m* and *Cpconf(Tg)/Sconf(Tg)*. Webb (2008) and Russell and Giordano (2017) used *Cpconf* data in their calculations, whereas, due to the more general nature of the present model, we calculated *Cpconf (Tg)* as *Cpliquid(Tg) – Cpglass(Tg),* with *Cpliquid(Tg)* predicted from the model of Richet and Bottinga*(1985)* and *Cpglass(Tg)* calculated from the Dulong-Petit limit of 3R*.* This most probably explains the moderately good correlation between predicted *m* and *Cpconf (Tg)/Sconf(Tg)* values in the present work. In turn, this result indicates a critical need for better heat capacity models of silicate melts, particularly for aluminosilicate compositions that where not the target of the Richet and Bottinga*(1985)* model*.*

**3.3 Raman spectra**

While prediction of the Raman spectra was not a goal in this study (they are used as a way to perform multitask learning and introduce structural information in the model), the model performs reasonable predictions of the Raman spectra of glasses (Figure 8). The average error between predicted and measured Raman spectra is of 3% of the total integrated intensity of the spectra. The model thus definitely captured the general link between glass composition and features observed in the Raman spectra. Most prediction errors are associated with signals in the 300-500 cm-1 and 800-1100 cm-1, indicating that the Raman dataset was probably too small for the model to adequately capture the links between glass composition, polymerisation, viscosity and Raman spectra. In the future, better results will be obtained by improving the coverage of the Raman spectral database.

**4. Discussion**

**4.1 Exploring the properties of alkali aluminosilicate melts**

The model allows a systematic exploration of the links between chemical composition and thermodynamic properties of melts. In figure 9, we show how *Sconf(Tg)* and *m* vary along different joints on the ternary diagrams Na2O-Al2O3-SiO2 and K2O-Al2O3-SiO2: the M2O-SiO2 silicate join, the join 3 (3 M2O for 1 Al2O3) and the join 1 (1 M2O for 1 Al2O3). Decreasing the silica concentration along the silicate join results in, first, an increase in *Sconf(Tg)* rapidly followed by a decrease to a minimum at the disilicate composition, and then an increase again at lower silica concentrations (Fig. 9A,B). Such variations originate from the competition between the topological and chemical entropy of the different *Qn* tetrahedral units in the melts (Le Losq and Neuville, 2017), which proportions change systematically with melt composition (e.g., Hideki Maekawa et al., 1991). With addition of aluminium, variations become interestingly smoother. In the Na2O-Al2O3-SiO2 diagram, a small maximum near mole[SiO2] = 0.75 is still observed on the joints 1 and 3. On the contrary, this is not visible anymore in the K2O-Al2O3-SiO2 diagram (Fig. 9A,B). Furthermore, while entropy values range between 6 and 12 J mol-1 K-1 in the Na2O-Al2O3-SiO2 diagram, they display larger variations in the K2O-Al2O3-SiO2 diagram; the entropy of the K-silicate melts are the highest, while the entropy of the K-tectosilicate melts (joint 1) are the lowest and actually barely vary for mole [SiO2]  [0.5-0.75].

Those variations bring important information regarding how the change of the role of alkali metal cations affect the thermodynamic properties of the melt. It has been documented that, as Al3+ enters the network of Na silicate melts, Na changes its role from network modifier to charge compensator of the AlO4- tetrahedral units (e.g., Le Losq et al., 2014 and references cited therein). From the present model results, changing the role of alkali metals from network modifier to charge compensator of Al results in an average decrease of the configurational entropy, particularly marked in the case of potassium. In detail, while *Sconf(Tg)* decreases monotonically from the silicate join to the join 1 at high silica contents in sodium compositions (Fig. 9A), it first slightly increases from the silicate join to the join 3, to then decrease when reaching the J1 at silica contents lowen than 0.8 in sodic melts (Fig. 9A), in agreement with Le Losq et al. (2014). In potassic melts, only a general, large and monotonic decrease of *Sconf(Tg)* is observe (Fig. 9B). The general decrease in *Sconf(Tg)* can be understood as a competition between alkali distributed in modifier percolation channels in silicate glasses, following the Random Network Model proposed by Greaves (1985), and alkali charge compensators distributed according to the Charge-Compensated Random Network (Greaves and Ngai, 1995; Le Losq et al., 2017). While sodium as a network modifier or charge compensator may not affect critically the general entropy of the network, Fig. 9B suggests that *Sconf(Tg)* change strongly as Al is added to potassic silicate melts and as potassium turns to a charge compensator of AlO4- units.

The large entropic variations observed in K-bearing systems are rationalized when considering that potassium behave in extreme ways in both silicate and aluminosilicate melts, as shown by a T-dependent liquid heat capacity for the former and significant K-clustering detected in the later (Richet and Bottinga, 1985; Le Losq et al., 2017). Such behavior may be assigned to its ionic radius, ranging between 146 pm and 168 pm in CN 6 to 12, respectively (Whittaker and Muntus, 1970). Potassium thus presents a ionic radius larger than oxygens (132 pm), and this may result in a particular structural behavior, like elevated cationic segregation and non-random distribution, which results in strong variations in thermodynamic properties of K-silicate and aluminosilicates melts and glasses.

Turning to fragility, we note that the variation from the silicate composition to the tectosilicate ones mirror those of *Sconf(Tg)*, but are much more restrained*.* The fragility *m* tends to monotonically decrease with the silica content of the glasses, and the ratio *M/Al* only results in small variations of *m*. This is explained by much larger variations in *Tg* that finally allow eq. *(4)* to be flexible enough to fit the viscosity data.

The general observations of the variations of *Sconf(Tg)* and *m* further reveal that, in all compositions, a significant change occurs close to the disilicate cpomposition at 0.67 mole SiO2. While in silicate glasses this is assigned to a maximum in the fractions of *Q3* units (3 BO per tetrahedra – 1 NBO) *(Hideki Maekawa et al., 1991)*, in tectosilicate glasses composed of (nearly) only *Q4* units, another explanation is required. The observation of predicted Raman spectra along the SiO2-NaAlO4 and SiO2-KAlO4 joins reveal an interesting change at such silica concentration (Figure 10). Indeed,

|  |  |  |  |
| --- | --- | --- | --- |
| Error using: | Train | Test | Valid |
| Adam-Gibbs (eq. 2) |  |  |  |
| MYEGA (eq. 3) |  |  |  |
| Avramov-Milchev (eq. 4) |  |  |  |

Table 1: XXXXX



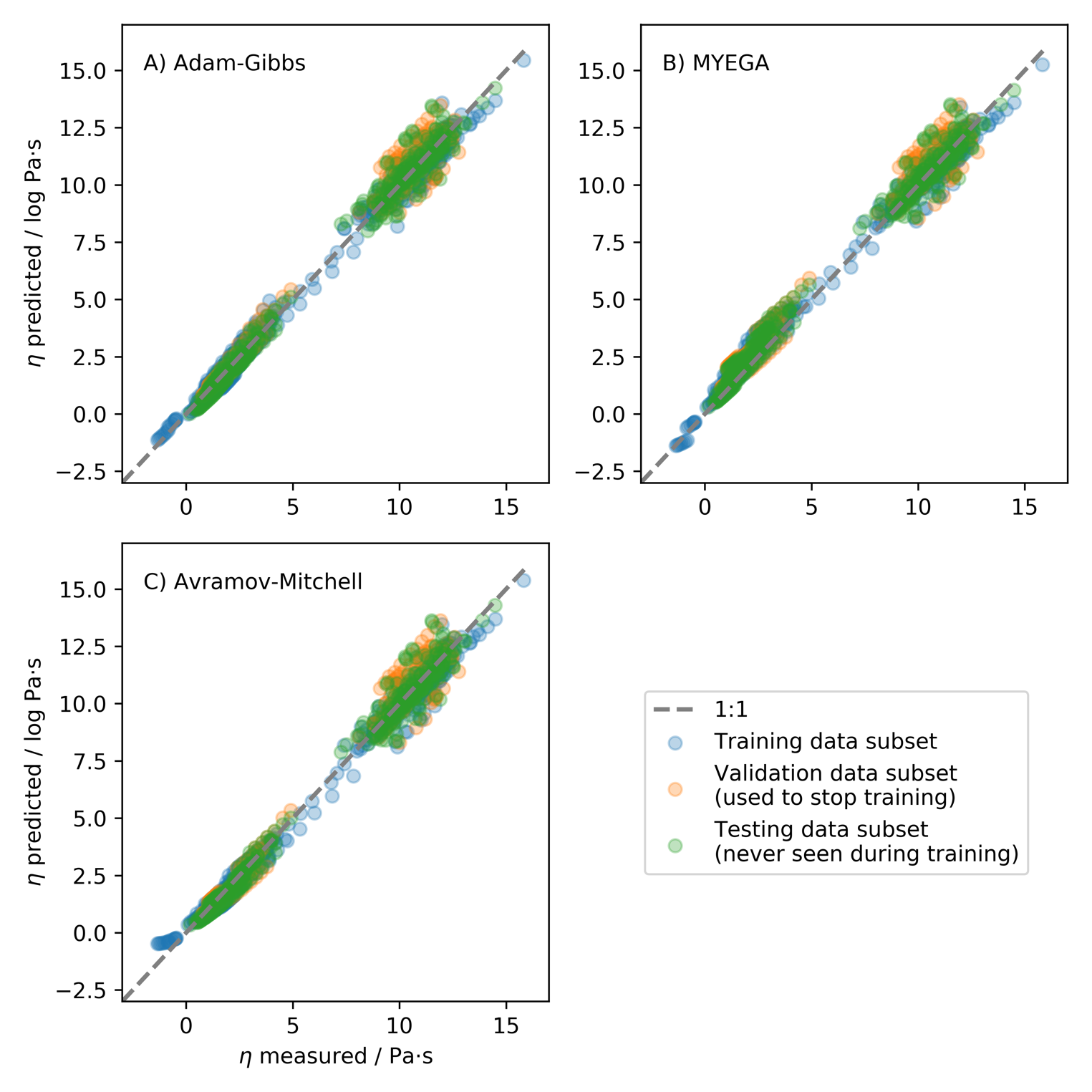
**Figure 1**



**Figure 2: Schematic drawing of the model.** The melt and glass chemical composition are inputted in the model in mole fractions, passed to the core neural network which provides thermodynamic and static parameters as well as the Raman spectra for the given composition.

****

**Figure 3: Random exploration of network parameters.** Mean square error (MSE) between the viscosity calculated with eq. 2 (Adam-Gibbs theory) and viscosity measurements reported as a function of the number of layers (A), neurons per layer (B), total number of hidden neurons (C) and the dropout probability (D) of the networks.

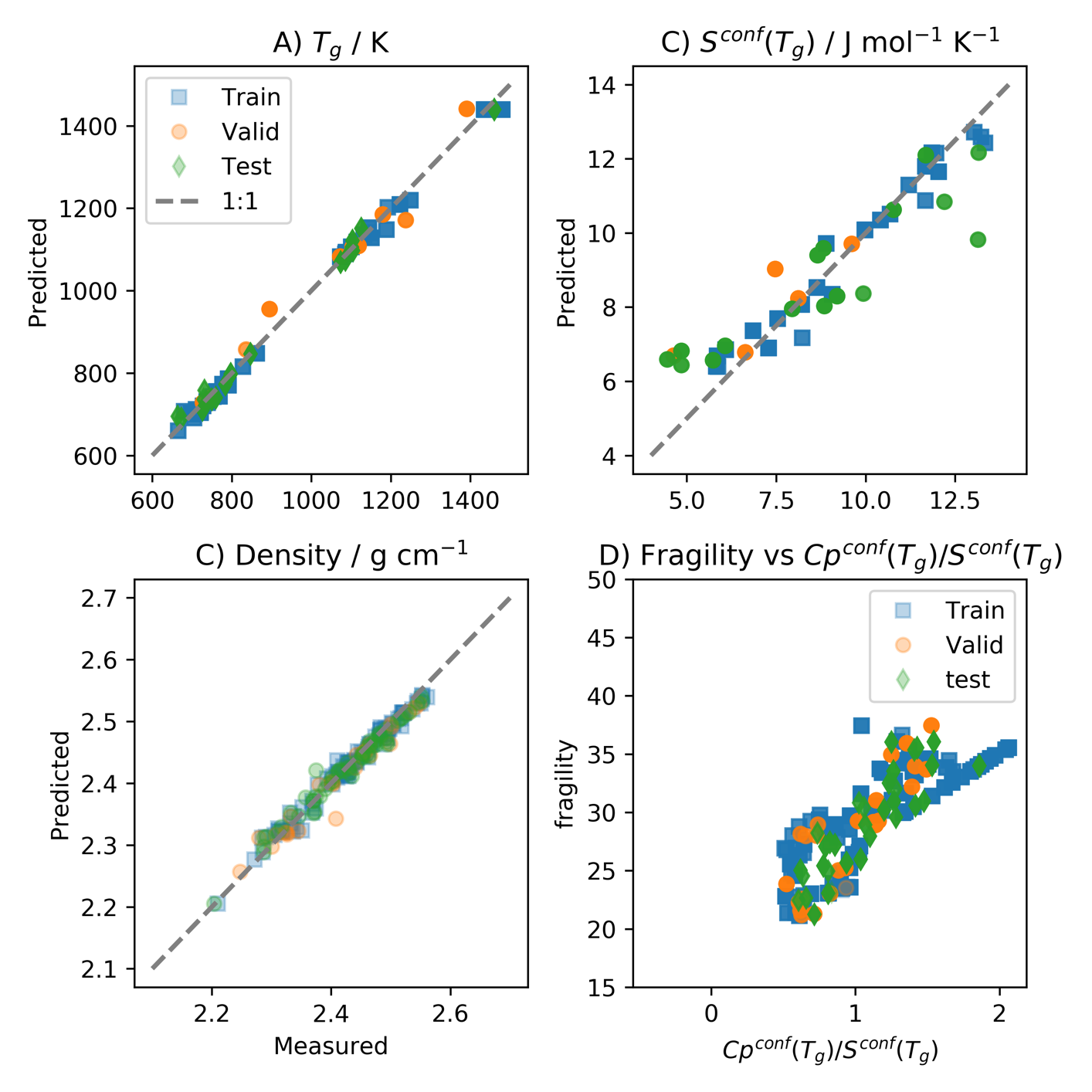


**Figure 4:** Comparison between viscosities predicted by the Adam and Gibbs theory (eq. 2, A), the derived MYEGA equation (eq. 3, B) or the Avramov-Milchev theory (eq. 4, C), and measurements.



**Figure 5: Effect of the size of the dataset on the predictive ability of the network.** This experiment was performed using a network with 4 hidden layers containing each 200 neurons, setting the dropout probability to 0.2.

**Figure 6: Multitask learning**



**Figure 7: Entropy vs compo**

**Figure 8 : Raman spectra**

**Figure 9: Entro – j1 j3**

**Figure 10 : Raman spectra along the tecto of SiO2-KAlO4 et SiO2-NaAlO4**