

Supplementary Materials for

Deep learning prediction of material properties

Charles Le Losq, Andrew Valentine, Daniel R. Neuville, Bjorn O. Mysen

Correspondence to: lelosq@ipgp.fr

**This PDF file includes:**

Materials and Methods

Supplementary Text

Figs. S1 to S7

Tables S1 to S3

**Other Supplementary Materials for this manuscript include the following:**

Software repository at github.com/charlesll/neuravi

Materials and Methods

***Experimental Design***

Developing the model required the collection and compilation of viscosity, density, refractive index as well as Raman spectra data on glasses and melts in the K2O-Na2O-Al2O3-SiO2 quaternary diagram (Fig. S1). Viscosity of supercooled melts remains to be better defined for peralkaline compositions in this diagram, such that we realized new experiments to complete the existing dataset. We further compiled existing data as specified below, prior to developing the model. The model is developed via the PyTorch library in the Python programming language, and can be run using Jupyter Notebooks. All the code and data necessary to reproduce this study are provided as an archive file and can also be found on the software repository Github at the web address <https://github.com/charlesll/neuravi> .

***Data compilation***

Existing data regarding Raman spectra, optical refractive index, 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 studies is critical and was checked on compositions like sodium trisilicate, albite and jadeite. Publications presenting deviations higher than 0.1 log Pa · s compared to the general trend on such compositions were entirely discarded.

***Sample synthesis and new measurements***

New glass samples were synthesized at IPGP in Paris from reagent-grade K2CO3, Na2CO3, Al2O3 and SiO2 dried oxide powders to complete the dataset following the protocol described in (*1*). Viscosity and density measurements follow the protocol described in (*1*)(*1*), such that the reader is refereed to this publication for further information. Chemical compositions have been measured using a Cameca SX50 electron microprobe (Table S1), with a 30 nA current, U = 30 kV, and 5 seconds of counting. The values reported in Table S1 are the statistical mean of 10-20 individual measurements. Viscosity measurement values for glasses reported in Table S1 are provided in Table S2.

***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 100-150 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 were added to the database (see Supplementary Materials for data and references). 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 then corrected from temperature and excitation line effects (*1*), 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 the 400-1250 cm-1 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.

***Machine learning modelling***

*Datasets*

Data used in this study are available in a spreadsheet (see supplementary archive as well as github.com/charlesll/neuravi) that was used as a database. Four 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);

- *Doptical*, the database of optical refractive index, composed of *Xoptical* chemical composition entries (mole fractions) as well as their associated wavelength (µm) and *yrefractive index* observations.

The idea of using four different set of observations was to leverage the fact that neural networks learning multiple tasks show better predictions capacities compared to those trained to learn only one task (*2*). *Dviscosity*, *Ddensity* and *Doptical* cover an important part of the glass-forming domain of alkali aluminosilicates (Fig. S1); they were thus used to train the artificial neural network with a performance oriented mindset. *DRaman* covers a more limited set of compositions (Fig. S1). It was used as a way of improving multitask learning as well as an way of introducing structural information in the model.

*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, but 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 (Fig. S1). 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,the data were randomly separated by composition (*3*) to avoid the pitfall of having the same composition in different *training*, *validation* and *testing* subsets in *Dviscosity* or *Doptical*. *Dviscosity, Doptical* and *Ddensity* were separated in three splits following this rule (Fig. S1). *DRaman* was divided in only two *train* and *validation* subsets due to its small size. This is not a problem as we do not want precise predictions of Raman spectra but uses this dataset as a way to improve the predictive capacity of the trained neural network.

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 (*4*). 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 and remains coherent in regard of the datasets. 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, and scaling does not affect network convergence. However, one key trick was to set the bias of the output layer of the neural network to values in the expected range of the predictions to be made, as done for instance for Mixture Density Network (*5*).

After their pre-processing, the different subsets were kept in HDF5 files for their future use in the model.

*Implementation*

The model presented in Fig. 1 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, each one having *k* activation units (a.k.a neurons). Different tests led to choose the now popular rectifier function (*6*), which for an input *x* give the output *y = max(0,x)*, as the function of the activation units. The outputs of this core network were fed to two final linear layers (Fig. 1). The first output layer returns vectors of Raman spectra calculated from the linear sum of the output layer’ inputs. The second output layer returns 15 different values:

- the parameters *Ae*, *AAM*, *ACG* and *ATVF* (eqs. 1, 2, 3, 4), as well as the coefficients *B1* to *B3* and *C1* to *C3* of the Sellmeir equation (see eq. S1) for the calculation of the glass refractive index *n* are directly given by the linear outputs;

- *Sconf(Tg)*, *CCG*, *Tg*, *To*, *T1*, the melt fragility *m*, the glass density *d* are calculated from the exponential of network outputs.

The use of the exponential function in the latter case was inspired by the strategy proposed by (*5*) to avoid negative values in outputs of 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. One trick for this method to work is to properly set the initial values of the biases of the last output layer to realistic values when creating the network. We note that it further allowed avoiding any scaling of the network outputs, which thus directly provide values in realistic ranges that do not need any transformation after prediction. After their calculations, *Sconf(Tg)*, *Tg*, *m,* *Ae* and *Aa-m* are used in the equations 1-4 for final predictions of melt viscosity, and the coefficients *B1* to *B3* and *C1* to *C3* to predict the refractive index at given wavelength, , via the Sellmeier equation:

*. (S1)*

The architecture of the hidden layers was optimized by performing a global random search (*30*). This allowed us to observe how the number of layers and hidden units affect the generalization ability of the model (Fig. S2). After training 2000 models in the same conditions on the same datasets, we observed than moderately deep networks with 3 to 5 layers and 200-300 units per layer perform best (Fig. S2); accordingly, best performance are in general reached with more than 1000 neurons. The dropout method, which consists in turning off *p* percent of neurons per layer at each training iteration in order to prevent overfitting (*7*), slightly helps preventing overfitting but is not a critical feature (Fig. S2). From this random search, we were able to select the 10 best models with the lowest error on the validation data subset. Predictions are made from those 10 best models, again their combination allowing better estimations to be made, following the well-known bagging principle (*8*).

*Training*

The least-square deviations between measurements and the viscosity predictions from eqs. 1-4 as well as density, optical 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. 1 are multiple, given the extreme correlation between *Be* and *Sconf(Tg)* originating from the intervention of the intrinsic entropy *Sc\** in both *Be* and *Sconf(Tg)*(*9*). 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

*. (S2)*

A similar strategy was adopted to calculate BFV and BTVF from the other parameters.

Batch training was performed monitoring the global loss on the *training* and *validation* data subsets. Early stopping (*10*) was performed to avoid overfitting: when the loss function on the *validation* data subset stopped decreasing for more than 50 epochs, training was stopped and the network presenting the best validation loss was saved. This was combined with the other strategies (dropout, bagging, multitask learning) in order to avoid overfitting as much as possible. From the 2000 random runs, the value of *p* is not always as critical as originally thought to avoid overfitting (Fig. S2): indeed, in the best saved best models obtained from the random exploration of the network architecture (see above), *p* usually is lower than 0.2. We still note that in general, higher *p* values result in less overfitting.

***Statistical Analysis***

Following the bagging method that consists in averaging the results of several models to improve model generalisation abilities, predictions are made from the mean of the 10 best models obtained from the random exploration of the network architecture (see above). This later step allowed a statistical analysis of the influence of the network size on the predictive ability of the model (see previous section as well as Fig. S2). The influence of the dataset size was explored training 10 neural networks with the same architecture with different *training* subset of *Dviscosity*. Results of this experiment thus represent the mean of those 10 different models. Finally, the correlation between the different predicted parameters was explored using the Spearman correlation coefficient that allows observing non-linear correlations between different variables.

Supplementary Text

***The structure of aluminosilicate melts***

Aluminosilicate melts are formed by a disorganized network of SiO2 and AlO2 tetrahedral units forming bonds between them, with some of the bonds being disrupted by network modifier elements like alkali and alkaline-earth elements. The later also compensate the deficit of charge around AlO2 tetrahedral units. In such melt, viscous flow is ensured by cooperative movements of the tetrahedral units, facilitated by the presence of network modifiers. Rapid cooling of such melts allows to cross the glass transition, where the melt structure is frozen-in. Melt structure close to the glass transition can thus be observed in their glasses via, e.g., 29Si Nuclear Magnetic Resonance or Raman spectroscopy.

It is important to consider melt structure and the role of elements in it because this directly determines melt properties. For instance, in the present case, variations in *Sconf(Tg)* and *m* can be understood once we consider that aluminum and non-network former metal cations have important and complex roles in the melt  (see review of *11*). In Al-free silicate glasses, network modifier alkali cations break Si-O-Si bonds, forming alkali channels percolating in the disrupted SiO2 tetrahedral network as described by the Modified Random Network (MRN) (*12*–*16*). Adding aluminum, entering as network forming AlO4- tetrahedral units, changes such picture: alkalis switch their role from network modifiers to charge compensators of AlO4- tetrahedra to ensure charge balance (*17*). Alkali distribution still is non-random, localized in compensator channels as described by the Compensated Continuous Random Network (CCRN) model (*18*, *19*). In term of properties, changing the role of alkali metals from network modifiers to charge compensators of Al results in an average decrease in *Sconf(Tg)*, particularly marked in the case of potassium compositions (Fig. 3A,B). MRNs thus may generally present higher *Sconf(Tg)* than CCRNs, at least for alkali aluminosilicates. Furthermore, mixing alkalis result in different MAE effects as the alkalis reside in MRN (Fig. 3C) or CCRN (Fig. 3D). In the former case, mixing Na and K results from an excess entropy of mixing caused by the hindering of the diffusions of alkali cations in modifier percolation channels (*18*). In the latter case, variations in *Sconf(Tg)* with the Na/K ratio are close to a linear mechanical mixing of two sub-networks (Na-Al-Si-O and K-Al-Si-O subnetworks) because K and Na occupy different environments and do not really interact upon mixing (*1*, *19*).

***Viscosity equations***

Aside the free volume and Adam-Gibbs models described in the main text, many other theories have been proposed to describe the viscous flow of liquids. Among those, some are empirical like the Tamman-Vogel-Fulcher (TVF) equation, or semi-empirical like the Avramov and Milchev (AM) (*12*) model. We can also cite the MYEGA model (*13*) that directly derives from eq. (1). The TVF equation is

, (S3)

with ATVF, BTVF and T1 adjustable parameters. The AM model proposes the equation

, (S4)

with AAM a pre-exponential terms proportional to , *Tg(x)* and *m(x)* the melt glass transition temperature and fragility. The MYEGA equation can be written in a quite similar form:

**,** (S5)

with *Ae*  a pre-exponential term proportional to .Those equations remain empirical or semi-empirical as they are not relating measurable variables, like heat capacity, to viscosity as done by the Adam-Gibbs equation. However, they still model well the viscosity dependence to temperature (Fig. S3), and for the MYEGA and AM equations allow leveraging knowledge of melt Tg and fragility as predicted by the model.

***The difficulty of modeling aluminosilicate melts properties***

The AG theory assumes that melt viscous flow occurs through cooperative re-arrangement of molecular subunits, and such events have been identified for instance via high-temperature 29Si NMR spectroscopy (*20*) or even direct observations (*21*, *22*) in silicate melts. The structure of simple Al-free silicate melts has been related to *Sconf(Tg)* (*23*) and *Cpconf* (*24*), and this was leveraged to calculate melt viscosity in the ternary Na2O-K2O-SiO2 system with an unrivaled precision of 0.18 log Pa·s (*23*). However, it is very difficult to extend to more complex composition like aluminosilicates because of the many new degrees of complexity generated by the addition of one critical elements like Al.

Indeed, it is actually difficult to experimentally validate models of melt structure for aluminosilicate compositions, representative of most natural and industrial glasses. In such compositions, 29Si NMR spectroscopy, which usually bring the necessary information to quantify the connectivity of SiO2 tetrahedral units, becomes blind due to Si-Al interactions resulting in significant signal broadening; signal interpretation relies on various hypothesis and back-end models of melt structure (*25*). Raman spectroscopy, another method to explore glass structure, does not solve this problem as it cannot be calibrated against reliable NMR data to distinguish the molecular subunits in the glasses. Furthermore, the aluminum content of the melt also affects interactions between, and the environment of the metal cations, as well as Al-Si ordering, Al coordinance and the potential presence of three-fold coordinated oxygen (see *11* for a review). Such problems severely affect our abilities to construct models in presence of aluminum, and strongly question the theoretical viability of proposed models based on untested structural calculations. This complexity pushed existing models to simply link chemical composition of aluminosilicate melts to their viscosity using a set of polynomial equations (*26*–*29*). A more complex model was proposed by Starodub et al. (*30*) for the system Na2O-K2O-Al2O3-SiO2. They proposed an associate-solution model for the structure of melts in the quaternary. Their model is interesting but one should keep in mind that it was not validated by experimental data, and therefore there may be biases due to our inability to determine the structure of aluminosilicate compositions.

***Internal consistency of the model***

The ability to predict the melt fragility *m* further allows testing the internal consistency of the model predictions. Indeed, experimental data indicate that a direct correlation between *m* and the ratio between the configurational heat capacity at *Tg* and *Sconf*(*Tg), Cpconf(Tg)/Sconf(Tg)* shall be observed (*31*, *32*). The model predicts this linear correlation (Figure S5), albeit some scattering arising from the way melt and glass *Cpconf* values are determined in the model. Indeed, the present model calculates *Cpconf(Tg)* as *Cpliquid(Tg) – Cpglass(Tg),* with *Cpliquid(Tg)* predicted from the model of *(33*)and *Cpglass(Tg)* calculated from the Dulong-Petit limit of 3R, with R the perfect gas constant*.* Models of *Cpliquid(Tg)* and the theoretical calculation of *Cpglass(Tg)* do not yield exact values, and this most probably explains the moderately good correlation between predicted *m* and *Cpconf(Tg)/Sconf(Tg)* values in the present work in comparison of other studies that used experimental values of *Cp*. In turn, this result indicates a critical need for better heat capacity models of aluminosilicate melts*.*

***Deep learning for small dataset?***

Using the viscosity dataset, we tested the effects of dataset size and network architecture on the model predictive abilities. Indeed, experimental data are often scarse because difficult to obtain, each experiment requiring a large amount of work, time, and potentially funds. The common idea is that this usually prohibits the use of “datathristy” machine learning methods. The present datasets are small, even extremely small (e.g. that of Raman spectra), raising the question: are we not simply overfitting or even encoding our data? We tested this by training the model on different training data subsets with variable size and monitoring the RMSE loss on the same unseen test data subset (Fig. S2A). As expected, the model predictive ability directly depends on the dataset size. Results become interesting only after reaching a threshold of around ~80 different compositions in the training data subset. This yields testing RMSE lower than 0.6 log Pa·s on the *testing* *Dviscosity* subset (Fig. S2A), a reasonable achievement as existing parametric models have RMSE values higher than 0.6 log Pa·s on the full 100-1015 Pa·s range (*27*, *28*). Lower numbers of compositions in the *training* subset result in the model constantly over-fitting the data. With more data, the lowest achievable RMSE probably lies around 0.36 Pa·s as shown by the evolution of the training RMSE (Fig. S2A) and reflects errors affecting the dataset (see below).

Aside the number of training data, the network architecture has a direct effect on the model predictive ability. For small datasets, a common outdated advice to avoid overfitting is to use networks with a limited number of activation units and layers. In our case, this is not true. More than a thousand hidden activation units are necessary to achieve good validation and testing RMSE on the viscosity dataset (Fig. S2). More precisely, network with more than 3 hidden layers provide better predictive abilities (Figure S2), confirming that deep networks perform better than shallow ones even on small datasets.

***Can we detect a lowest achievable limit for the model predictive error?***

Interestingly, the training RMSE on viscosity is always of ~0.36 log Pa·s (Fig. S2A), regardless of the training data subset size. Artificial neural networks are extremely flexible and very prone to overfit the training data subset. Therefore, the fact that models trained on a small number of compositions clearly overfit the training data subset but do not provide viscosity predictions better than ~0.36 log Pa·s brings an important information: this places a lowest achievable RMSE at ~0.36 log Pa·s on the present data subset. We infer that this limit reflects contributions from experimental errors affecting viscosity measurements and chemical composition values, those errors varying between different laboratories, as well as from the accuracy of the glass and melt heat capacity determination (see above).

***Configurational entropy at the glass transition: chemical and topological contributions***

Model results show large, non-linear variations of *Sconf(Tg)* with glass composition. Such variations are better understood when considering that *Sconf(Tg)* should be considered as originating from various sources, of topological and chemical nature. As a result, it varies in a complex way with melt chemistry, and those variations are difficult to rationalize without making severe simplifications and/or assumptions. Even for the simple SiO2 glass, not all the configurational entropy can be assigned to a topological origin with certainty. Indeed, the Raman signal of the A1 vibrational doublet at high frequency arising from symmetric stretching of the SiO2 tetrahedral units is split in two components, which fractions represent 0.68 and 0.32 of the A1 signal (*34*); assuming an ideal mixing of the two vibrators, one finds a Raman-derived *Sconf(Tg)* = 5.2 ± 0.4 J mol-1 K-1. This value is surprisingly realistic, very close to the 5.1 ± 2 J mol-1 K-1 value returned from calorimetric measurements (*35*) and not very different from the 8.3 ± 2.8 J mol-1 K-1 value calculated from the SiO2 viscosity data (*23*). The same treatment for the nepheline glass yields a value of 4.6 ± 0.7 J mol-1 K-1, close to the calorimetric value of 4.85 J mol-1 K-1, this being explained by the high ordering of the Si-Al distribution in nepheline (*36*) and hence a limited chemical mixing of various units. However, as soon as chemical mixing becomes important, such calculation fails. This is shown by comparing the value of such calculation for albite, which yields a Raman-derived *Sconf(Tg)* = 3.8 ± 1.1 J mol-1 K-1, a value much lower than the calorimetric one equal to 9.2 ± 2 J mol-1 K-1 (*35*). Assuming that Na repartition in CCRN is not accompanied by any mixing effect (this seems reasonable as Na is the only cation that can occupy such environment), this difference can be assigned to Al-Si mixing. Assuming ideal mixing, one can calculate it and add it to the 3.8 ± 1.1 J mol-1 K-1 value to finally obtain a value of 8.46 ± 1.1 J mol-1 K-1, a value in much better agreement with the measured one. Such analysis may not be easily extrapolated to other compositions, but rather is an interesting exposition of the source of entropy affecting the melt configurational entropy at the glass transition, and, hence, the variations of its viscosity. In turn, the complexity of *Sconf(Tg)* indicates that values for different glasses do not necessary embed interesting information about their relative ordering. Indeed, while two glasses may present similar *Sconf(Tg)*, their decomposition may yield very different topological and mixing contributions, making any attempt in discussing glass structure using *Sconf(Tg)* pointless*.*

***Configurational Entropy in the quaternary Na2O-K2O-Al2O3-SiO2 diagram***

Variations in *Sconf(Tg)* with melt composition are non-monotonic, with maxima close to, or on the silicate joint (alkali-silica, Al-free), while in general a decrease in *Sconf(Tg)* is observed upon addition of aluminum (Fig. 3A,B). Increasing the fraction of MAlO2 (M=Na, K) decreases *Sconf(Tg)*, the effect being more linear and pronounced for potassic compositions (Fig. 3B). Mixing alkalis result in different effects as a function of the M/Al ratio of the melt (Fig. 3C,D). For silicate compositions, the Mixed Alkali Effect (MAE) (*15*) is observed: a maximum in *Sconf(Tg)* is generally observed as Na and K mix, except at very low (< 0.55) or very high (> 0.95) silicate contents. At high Al concentrations, the variations in entropy with the Na/K ratio are closer to a linear trend, indicating a different MAE (*2*).

***Links between the Adam-Gibbs and Free Volume theories?***

The trans-theoretical character of the model allows adopting a new vision about a problem like melt viscous flow and glass transition. For aluminosilicate melts, it allows observing the link between two theories: Adam-Gibbs (AG) and Free Volume (FV). In the FV theory, solid-like and liquid-like molecular cells are distinguished and separated by a critical volume *v\**, and viscous flow occurs via cooperative molecular movements between liquid-like cells. In the AG theory, viscous flow occurs via cooperative motions of molecular segments of a size *z\*(T),* characterized by an intrisic entropy *Sc\**. The two theories thus share some common philosophical background: viscous flow is assumed to occur via some sort of cooperative movements of molecular entities in the melt. This background can be retrieved when diving into the details of the parameters of eq. 1 and 2. Indeed, *BCG* embeds some structural information because it depends on *v\**:

*BCG = v\* zo, (5)*

where *zo* is an adjustable parameter. Similarly, the ratio *Be/Sconf(Tg)* embeds molecular subunits lengthscale information as

*Be/Sconf(Tg) = [ Δµ z\*(Tg) ] / R, (6)*

with *Δµ* the energy barriers opposed to the rearrangement of molecular subunits of size *z\*(Tg)*, and R the perfect gas constant. Now let’s consider *v\** and *z\** as structural parameters embedding information about the volume or lengthscale of cooperative molecular regions. In this case, they should directly depend on melt or glass structure. This is confirmed by the fact that both *BCG* and *Be/Sconf(Tg)* correlate very well with *RRaman* (Fig. 5E,F, *rs* = 0.987 and 0.985 respectively), which provides information about the network connectivity; the higher *RRaman* is, the higher the interconnection between polyhedral units, thus the higher the 3D network topology *(19*)*.* In detail, this implies that it actually should be possible to develop a free-volume version of the AG theory, as it has been proposed (*20*, *21*). More generally, the links between *BCG,* *Be/Sconf(Tg)* and *RRaman* support the hypothesis that melt viscous flow occurs when a critical molecular lengthscale is reached. This lengthscale can be determined from Raman spectra (Fig. 5E,F) and strongly influences the glass transition temperature *Tg* (Fig. 5B). In detail, entropic effects (like the excess of entropy resulting from the MAE) also affect *Tg* (*18* and references cited therein) but their influence remains limited compared to that of the polyhedral network topology (Fig. 5B). On the other hand, such entropic effects strongly control the rate at which supercooled melt viscosity changes as a function of T, or in other terms, the melt fragility *m* because (Fig. S5)*.* Actually, we may consider that such mixing phenomena are allowed by the non-ergotic nature of melts, their inhomogeneities being frozen into glasses below *Tg*. Following from this idea, it is possible to relate *m* to ongoing density fluctuations inherited from dynamic heterogeneities formed at supercooled T, i.e. Tliquidus < T < Tg (*22*). The nature of such heterogeneities is expected to change largely with melt composition, as we can recognize different cases for network organization as described by the Random Network for simple AX2 glasses like SiO2 (*23*), the Modified Random Network for silicate liquids (*24*) and the Continuous Compensated Random Network (*25*, *2*) for aluminum-rich aluminosilicates, each one presenting distinct and different extent of heterogeneities at medium and long range order with various associated cationic mixing within such heterogeneities that drive large and complex changes in glass configurational entropy (Supplementary Text) and hence viscosity (Fig. 2A).

1. C. Le Losq, D. R. Neuville, Effect of the Na/K mixing on the structure and the rheology of tectosilicate silica-rich melts. *Chemical Geology*. **346**, 57–71 (2013).

2. R. Caruana, Multitask Learning. *Machine Learning*. **28**, 41–75 (1997).

3. C. Le Losq, A. J. Berry, M. A. Kendrick, D. R. Neuville, H. S. C. O’Neill, Determination of the oxidation state of iron in Mid-Ocean Ridge basalt glasses by Raman spectroscopy. *American Mineralogist*. **104**, 1032–1049 (2019).

4. K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (The MIT Press, Cambridge, Massachusetts, 2012).

5. C. M. Bishop, *Pattern recognition and machine learning* (Springer, New York, 2006), *Information science and statistics*.

6. X. Glorot, A. Bordes, Y. Bengio, in *International Conference on Artificial Intelligence and Statistics* (2011; http://machinelearning.wustl.edu/mlpapers/paper\_files/AISTATS2011\_GlorotBB11.pdf), pp. 315–323.

7. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*. **15**, 1929–1958 (2014).

8. L. Breiman, L. Breiman, in *Machine Learning* (1996), pp. 123–140.

9. G. Adam, J. H. Gibbs, On the temperature dependence of cooperative relaxation properties in glass-forming liquids. *The journal of chemical physics*. **43**, 139–146 (1965).

10. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016; http://www.deeplearningbook.org).

11. C. Le Losq, M. R. Cicconi, G. N. Greaves, D. R. Neuville, in *Handbook of Glass* (Springer, 2019; https://www.springer.com/us/book/9783319937267).

12. G. N. Greaves, A. Fontaine, P. Lagarde, D. Raoux, S. J. Gurman, Local structure of silicate glasses. *Nature*. **293**, 611–616 (1981).

13. G. N. Greaves, Exafs and the structure of glass. *Journal of Non-Crystalline Solids*. **71**, 203–217 (1985).

14. G. N. Greaves, EXAFS, glass structure and diffusion. *Philosophical Magazine Part B*. **60**, 793–800 (1989).

15. A. Meyer, J. Horbach, W. Kob, F. Kargl, H. Schober, Channel formation and intermediate range order in sodium silicate melts and glasses. *Physical Review Letters*. **93**, 1–4 (2004).

16. F. Kargl, A. Meyer, M. M. Koza, H. Schober, Formation of channels for fast-ion diffusion in alkali silicate melts: A quasielastic neutron scattering study. *Physical Review B*. **74**, 014304 (2006).

17. C. Le Losq, D. R. Neuville, P. Florian, G. S. Henderson, D. Massiot, The role of Al3+ on rheology and structural changes of sodium silicate and aluminosilicate glasses and melts. *Geochimica et Cosmochimica Acta*. **126**, 495–517 (2014).

18. G. N. Greaves, K. L. Ngai, Reconciling ionic-transport properties with atomic structure in oxide glasses. *Phys. Rev. B*. **52**, 6358–6380 (1995).

19. C. Le Losq, D. R. Neuville, W. Chen, P. Florian, D. Massiot, Z. Zhou, G. N. Greaves, Percolation channels: a universal idea to describe the atomic structure and dynamics of glasses and melts. *Scientific Reports*. **7**, 16490 (2017).

20. J. Stebbins, NMR evidence for five-coordinated silicon in a silicate glass at atmospheric pressure. *Nature*. **351**, 638–639 (1991).

21. P. Y. Huang, S. Kurasch, A. Srivastava, V. Skakalova, J. Kotakoski, A. V. Krasheninnikov, R. Hovden, Q. Mao, J. C. Meyer, J. Smet, D. A. Muller, U. Kaiser, Direct Imaging of a Two-Dimensional Silica Glass on Graphene. *Nano Letters*. **12**, 1081–1086 (2012).

22. P. Y. Huang, S. Kurasch, J. S. Alden, A. Shekhawat, A. A. Alemi, P. L. McEuen, J. P. Sethna, U. Kaiser, D. A. Muller, Imaging Atomic Rearrangements in Two-Dimensional Silica Glass: Watching Silica’s Dance. *Science*. **342**, 224–227 (2013).

23. C. Le Losq, D. R. Neuville, Molecular structure, configurational entropy and viscosity of silicate melts: Link through the Adam and Gibbs theory of viscous flow. *Journal of Non-Crystalline Solids*. **463**, 175–188 (2017).

24. B. O. Mysen, Experimental, in situ, high-temperature studies of properties and structure of silicate melts relevant to magmatic processes. *European Journal of Mineralogy*. **7**, 745–766 (1995).

25. B. O. Mysen, A. Lucier, G. D. Cody, The structural behavior of Al3+ in peralkaline melts and glasses in the system Na₂O-Al₂O₃-SiO₂. *American Mineralogist*. **88**, 1668–1678 (2003).

26. J. K. Russell, D. Giordano, A model for silicate melt viscosity in the system CaMgSi2O6-CaAl2Si2O8-NaAlSi3O8. *Geochimica et Cosmochimica Acta*. **69**, 5333–5349 (2005).

27. H. Hui, Y. Zhang, Toward a general viscosity equation for natural anhydrous and hydrous silicate melts. *Geochimica et Cosmochimica Acta*. **71**, 403–416 (2007).

28. D. Giordano, J. K. Russell, D. B. Dingwell, Viscosity of magmatic liquids: A model. *Earth and Planetary Science Letters*. **271**, 123–134 (2008).

29. X. Duan, A model for calculating the viscosity of natural iron-bearing silicate melts over a wide range of temperatures, pressures, oxygen fugacites, and compositions. *American Mineralogist*. **99**, 2378–2388 (2014).

30. K. Starodub, G. Wu, E. Yazhenskikh, M. Müller, A. Khvan, A. Kondratiev, An Avramov-based viscosity model for the SiO2-Al2O3-Na2O-K2O system in a wide temperature range. *Ceramics International*. **45**, 12169–12181 (2019).

31. S. L. L. Webb, Configurational heat capacity of Na₂O–CaO–Al₂O₃–SiO₂ melts. *Chemical Geology*. **256**, 92–101 (2008).

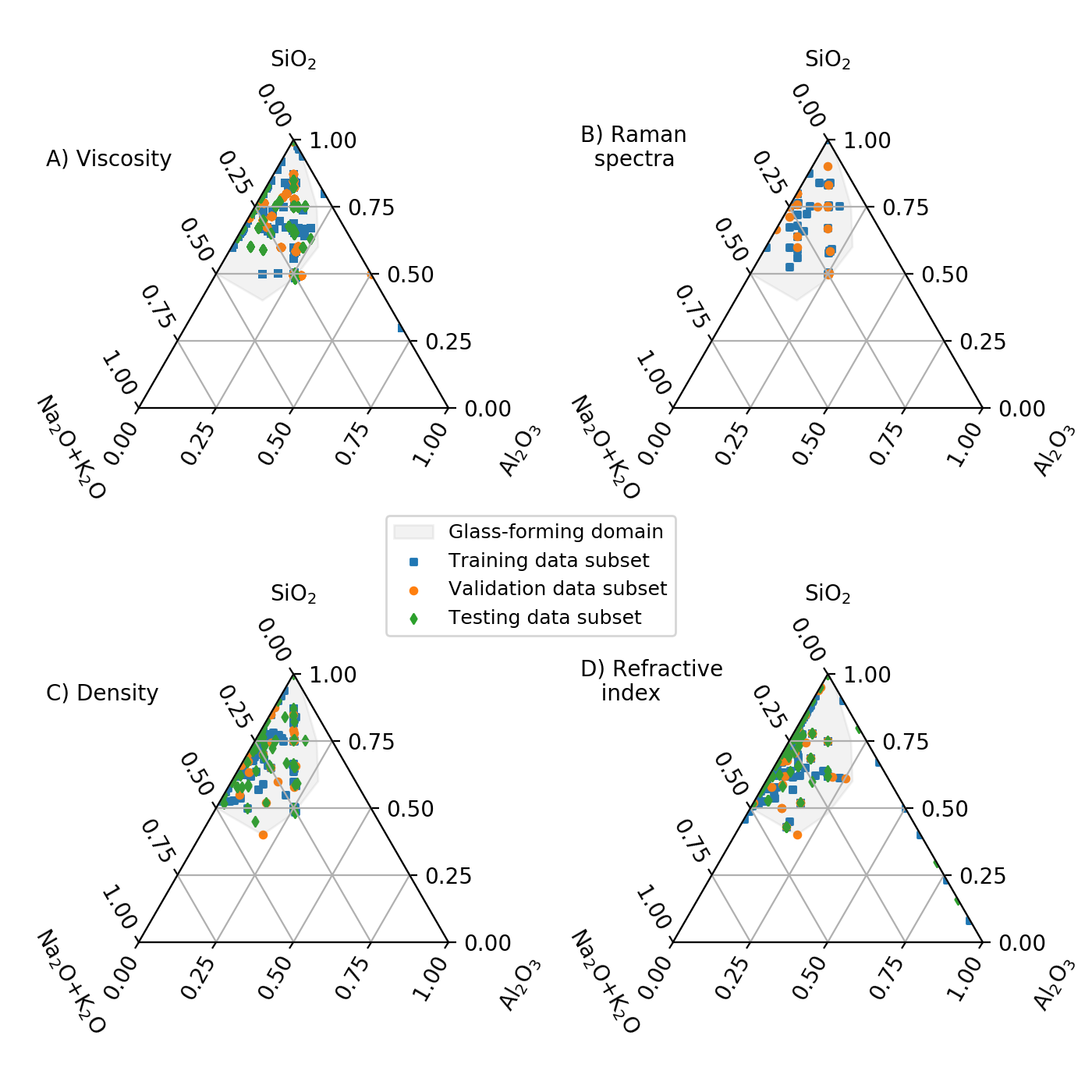
32. J. K. Russell, D. Giordano, Modelling configurational entropy of silicate melts. *Chemical Geology*. **461**, 140–151 (2017).

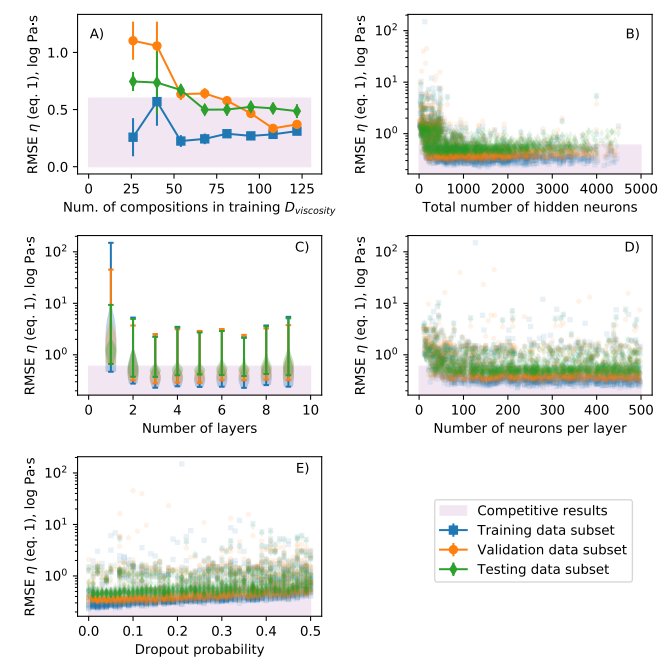
33. P. Richet, Y. Bottinga, Heat capacity of aluminum-free liquid silicates. *Geochimica et Cosmochimica Acta*. **49**, 471–486 (1985).

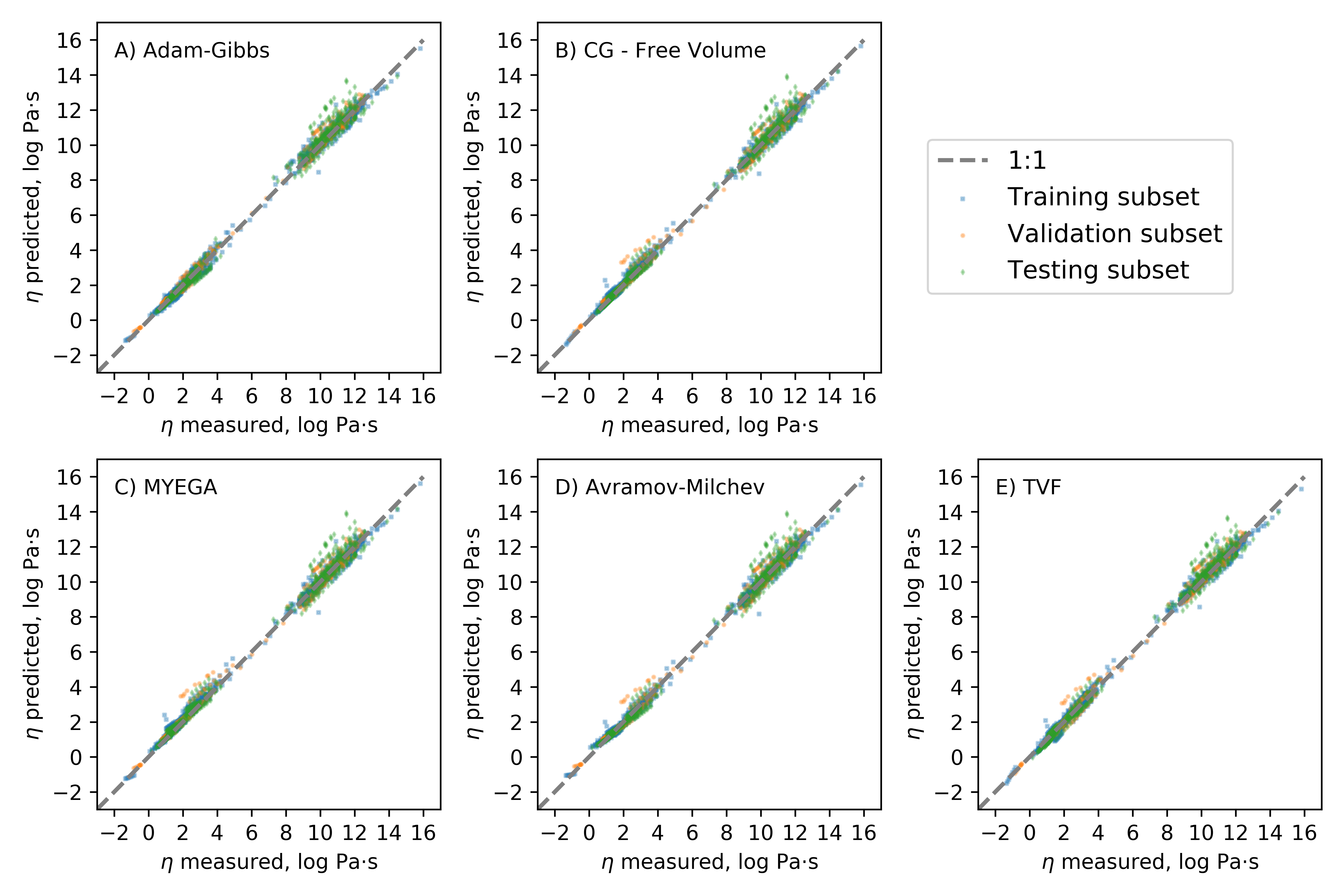
34. D. R. Neuville, B. O. Mysen, Role of aluminium in the silicate network: In situ, high-temperature study of glasses and melts on the join SiO2-NaAlO2. *Geochimica et Cosmochimica Acta*. **60**, 1727–1737 (1996).

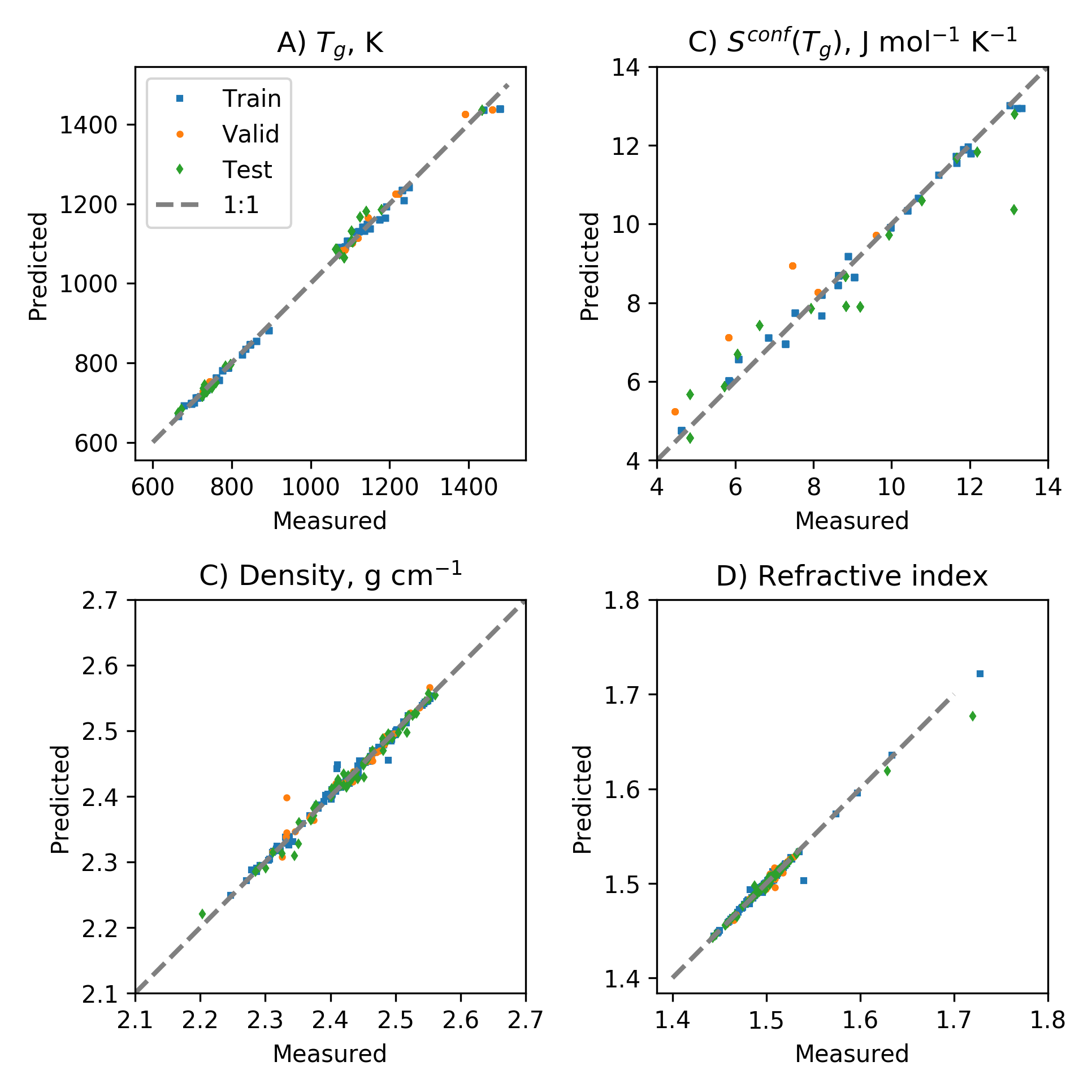
35. P. Richet, Y. Bottinga, Glass transitions and thermodynamic properties of amorphous SiO2, NaAlSinO2n+2 and KAlSi3O8. *Geochimica et Cosmochimica Acta*. **48**, 453–470 (1984).

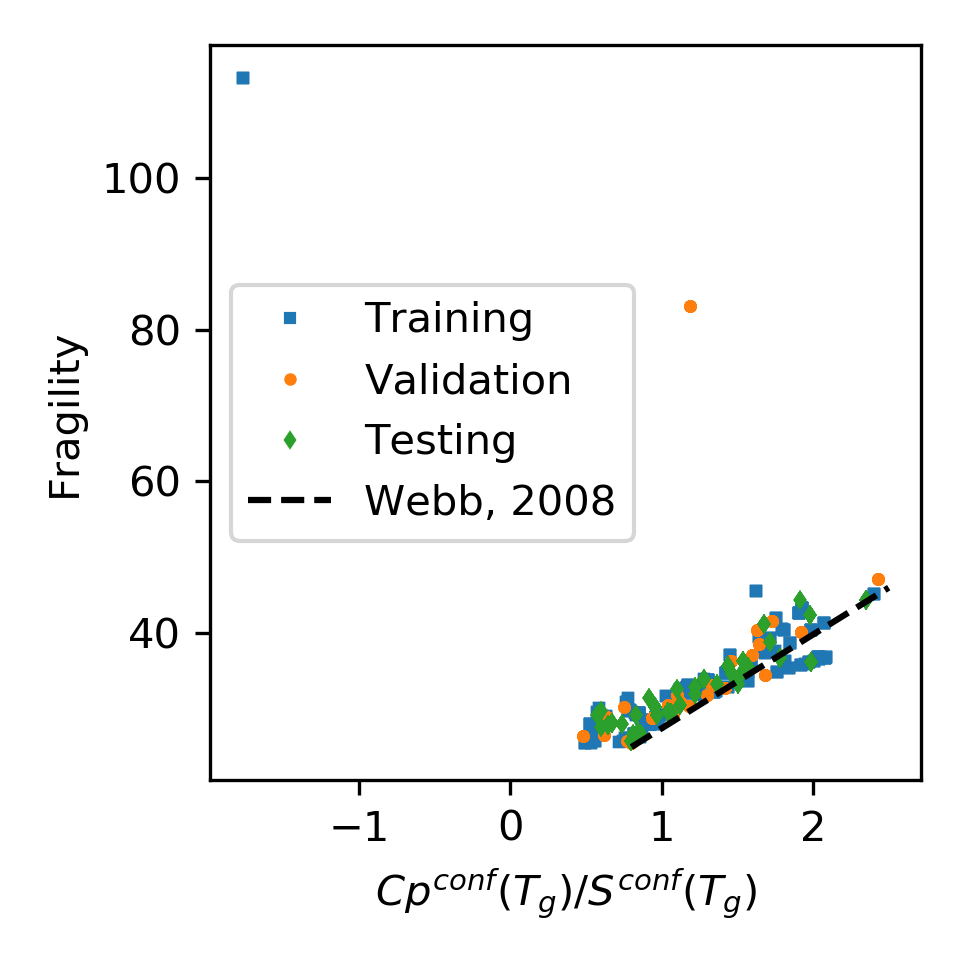
36. P. Richet, R. A. Robie, J. Rogez, B. S. Hemingway, P. Courtial, C. Téqui, Thermodynamics of open networks: ordering and entropy in NaAlSiO₄ glass, liquid and polymorphs. *Physics and Chemistry of Minerals*. **17**, 385–394 (1990).

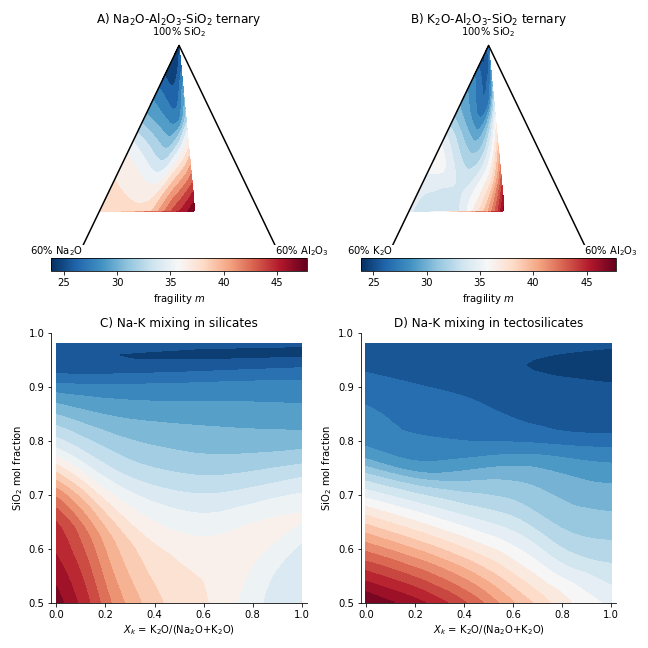
Fig. S1. Viscosity (A), Raman (B), density (C) and refractive index (D) datasets used in this publication. Each symbol corresponds to a composition. The glass-forming domain at usual laboratory cooling rates is indicated in grey.

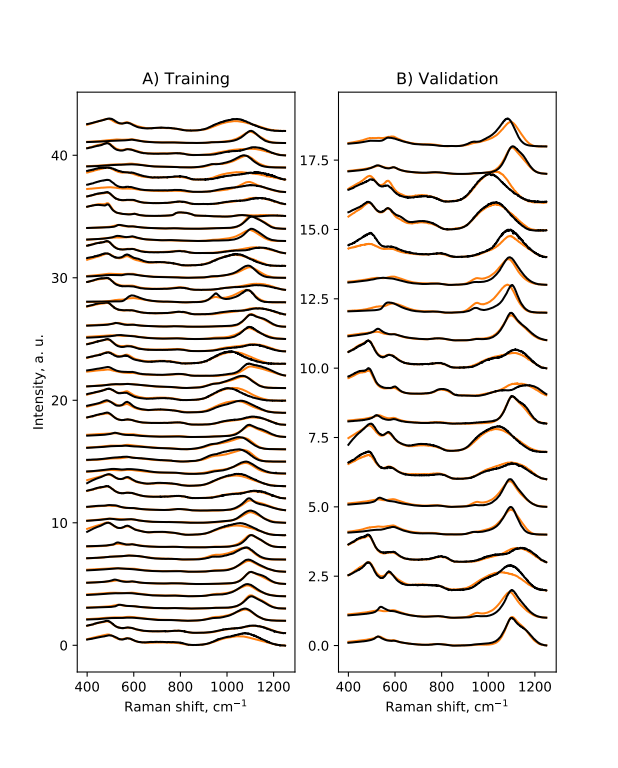
Fig. S2. Network architecture and dropout influence on model performance. RMSE between viscosity predictions made using the Adam-Gibbs model (eq. 1) and measurements in training, validation and testing data subsets as a function of the number of compositions in the training data subset (A), of the total number of hidden activation units (a.k.a. neurons, B), of the number of hidden layers (C), of the number of neurons per layer (D) and of the dropout probability (E). 2,000 models were randomly selected and trained to obtain those results. Subplot C is a violin plot with extreme values showed. Subplots B, D and E are scatter plots in which each slightly transparent symbol corresponds to a given model; less transparence is directly indicative of a higher number of models for a given X-Y value. Deep network still generalizes better than shallow one on this small problem. Overfitting is limited but present, as shown by systematically lower errors on the training data subset.

Fig. S3. Comparison between predicted and measured viscosity in the Na2O-K2O-Al2O3-SiO2 system. Predictions can be made using theories like Adam-Gibbs (A) and free volume (B), or empirical equations like MYEGA (C), Avramov-Milchev (D), and Tamman-Vogel-Fulcher (E). See table S3 for RMSE and text for the equations.

Fig. S4. Comparison between predictions and viscous glass transition temperature (A), *Sconf(Tg)* (values from *19*, *14*, *1*) (B), density (C) and refractive index (D) data. See table S3 for RMSE.

**Fig. S5 Glass fragility versus melt *Cpconf(Tg)/Sconf(Tg)* ratio.** Symbols are predictions of the model on the different subsets of the *Dviscosity* dataset. The back dotted line is the relationship observed by (*31*) using experimental heat capacity data. Except two extreme outliers that corresponds to Al2O3-SiO2 melts with more than 30 mol% Al2O3, a general good agreement is observed. This validates the internal consistency of the model. The scatter indicates that a better agreement could be obtained with using a better heat capacity model for aluminosilicate melts. Such *Cp* calculation could be integrated in the present global model in the future albeit the acquisition of additional *Cp* data for ternary and quaternary aluminosilicate melts.

Figure S6: Melt fragility variations with compositon. Fragility is represented in the glass forming domains of the ternary sodium (A) and potassium (B) aluminosilicate systems, as well as as a function of the silica fraction and the potassium to total alkali ratio of silicate (C) and tectosilicate (D) melts. No MAE is observed on melt fragility, which depends largely on melt silica content

**Fig. S7 Predicted (black lines) and measured (orange lines) Raman spectra of the glasses.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Glass name** |  | **%SiO2** | **%Al2O3** | **%K2O** | **%Na2O** | **Density, g cm-1** |
| KA80.05 | nom. mol% | 80.00 | 5.00 | 15.00 | 0.00 |  |
|  | nom. wt% | 71.40 | 7.60 | 21.00 | 0.00 |  |
|  | an. wt% | 74.85(44) | 7.56(14) | 15.12(22) | 0.00(4) | 2.320(1) |
| KA72.07 | nom. mol% | 72.00 | 7.00 | 21.00 | 0.00 |  |
|  | nom. wt% | 61.60 | 10.20 | 28.20 | 0.00 |  |
|  | an. wt% | 61.41(28) | 10.17(24) | 27.42(33) | 0.00(2) | 2.408(1) |
| KA65.09 | nom. mol% | 65.00 | 8.75 | 26.25 | 0.00 |  |
|  | nom. wt% | 53.70 | 12.30 | 34.00 | 0.00 |  |
|  | an. wt% | 53.32(55) | 12.49(43) | 31.68(28) | 0.00(3) | 2.451(10) |
| NA65.09 | nom. mol% | 65.00 | 8.75 | 0.00 | 26.25 |  |
|  | nom. wt% | 60.79 | 13.89 | 0.00 | 25.32 |  |
|  | an. wt% | 61.75(45) | 13.67(24) | 0.03(2) | 24.56(73) | 2.472(4) |
| NA58.10 | nom. mol% | 58.00 | 10.50 | 0.00 | 31.50 |  |
|  | nom. wt% | 53.55 | 16.45 | 0.00 | 30.00 |  |
|  | an. wt% | 54.61(28) | 16.42(18) | 0.05(2) | 28.92(37) | 2.502(5) |

**Table S1. Composition of the synthesized glasses.** Nominal (nom.) and analyzed (an.) compositions are reported. Standard deviations on measured values on 10 different spots (for EPMA measurements) or glass chips (for density measurements) are given in parenthesis (1 sigma confidence interval).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **T, K** | **KA80.05** | **T, K** | **KA72.07** | **T, K** | **KA65.09** | **T, K** | **NA65.09** | **T, K** | **NA58.10** |
| 1013.1 | 9.10 | 921.5 | 9.37 | 941.3 | 9.55 | 834.0 | 9.01 | 827.3 | 10.10 |
| 1001.8 | 9.32 | 891.0 | 10.17 | 935.1 | 9.71 | 829.0 | 9.18 | 836.9 | 9.73 |
| 989.6 | 9.51 | 872.0 | 10.75 | 919.4 | 10.13 | 813.8 | 9.61 | 817.6 | 10.51 |
| 981.6 | 9.78 | 852.0 | 11.40 | 913.6 | 10.32 | 803.2 | 9.94 | 796.2 | 11.42 |
| 967.6 | 10.05 |  |  | 898.7 | 10.78 | 798.6 | 10.09 | 805.7 | 10.97 |
| 949.7 | 10.50 |  |  | 892.0 | 10.96 | 787.5 | 10.50 | 847.1 | 9.36 |
| 940.3 | 10.83 |  |  | 882.1 | 11.28 | 779.9 | 10.81 | 855.9 | 9.09 |
| 928.2 | 11.05 |  |  | 867.3 | 11.86 | 773.7 | 11.07 | 828.3 | 10.01 |
| 918.4 | 11.32 |  |  | 855.4 | 12.29 | 772.9 | 11.07 | 834.4 | 9.81 |
| 905.1 | 11.63 |  |  |  |  | 762.6 | 11.52 | 787.4 | 11.80 |
| 896.6 | 11.92 |  |  |  |  | 756.3 | 11.74 | 777.4 | 12.36 |
|  |  |  |  |  |  | 752.1 | 11.95 |  |  |

**Table S2. Viscosity measurements made on the synthesized glasses.** Viscosity is in log10 Pas and was measured using a creep apparatus following the protocol described in (*19*). Errors on temperature are lower than 0.3 K, and errors on viscosity lower, or equal to 0.03 log10 Pas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data subset:** | **Training** | **Validation** | **Testing** |
| **Adam-Gibbs (eq. 1, log Pas)** | 0.25 | 0.30 | 0.46 |
| **Free Volume (eq. 2****, log Pas)** | 0.22 | 0.36 | 0.45 |
| **TVF (eq.S3, log Pas)** | 0.24 | 0.35 | 0.45 |
| **MYEGA (eq. S5, log Pas)** | 0.25 | 0.39 | 0.46 |
| **Avramov-Milchev (eq. S4, log Pas)** | 0.24 | 0.36 | 0.46 |
| **Density (g cm-1)** | 0.007 | 0.014 | 0.009 |
| **Raman spectra (%, LAD)** | 18 | 22 | - |
| **Refractive index** | 0.003 | 0.003 | 0.005 |

**Table S3: Root-mean-square errors of the model.** RMSE calculated between measured and predicted melt viscosity, density, refractive index, except for Raman spectra where a different metric is used (median least absolute deviation LAD).