



HAL
open science

Subsampling approach and data-driven models to predict silicate glass melt viscosity

Damien Perret, Alexandra Garcin, Carole Soual, François Bergeret

► To cite this version:

Damien Perret, Alexandra Garcin, Carole Soual, François Bergeret. Subsampling approach and data-driven models to predict silicate glass melt viscosity. *Materials Letters*, 2024, 379, pp.137691. 10.1016/j.matlet.2024.137691 . cea-04815427

HAL Id: cea-04815427

<https://cea.hal.science/cea-04815427v1>

Submitted on 2 Dec 2024

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.



Distributed under a Creative Commons Attribution 4.0 International License



Subsampling approach and data-driven models to predict silicate glass melt viscosity

Damien Perret^{a,*}, Alexandra Garcin^a, Carole Soual^b, François Bergeret^b

^a CEA, DES, ISEC, DPME, Univ Montpellier, Marcoule, France

^b Ippon Innovation, Toulouse, France

ARTICLE INFO

Keywords:
Glass
Viscosity
Machine learning
Database

ABSTRACT

This study focused on developing a predictive tool for calculating glass melt viscosity between 900 °C and 1500 °C. A database containing approximately 16 000 silicate glass compositions was built using both literature data and a proprietary dataset. The approach integrates statistical techniques, including design of experiments, machine learning, and subsampling strategies for model training. Prediction accuracy was found to be highly promising for the various types of silicate glasses studied. The relative error in viscosity prediction at 1200 °C was approximately 20 % for simple SBN compositions, and less than one order of magnitude for more complex compositions.

1. Introduction

In situations where theoretical models cannot be efficiently applied to calculate glass properties, empirical statistical models are often required. This is typically the case when the glass contains a high number of components. Since the end of the 19th century, it has been known that under certain conditions, silicate glass properties can be expressed as a simple linear combination of oxide contents. This “Principle of Additivity” was initially introduced to calculate heat capacity of glass, and was later extended during the 20th century to a larger number of properties: optical, thermal and mechanical properties [1]. However, this calculation can only be applied to predict glass properties within a narrow range of compositions. Since the 2000s, the significant increase in computational power has enabled the use of highly efficient algorithms in data mining predictive methods.

However, glass melt viscosity prediction remains difficult. From strain point to melting temperature, glass viscosity extends on over 13 orders of magnitude, making prediction very challenging on wide range of temperatures. Even within a narrow range of temperature, glass melt viscosity variation may be significant. Viscosity-temperature dependence is highly sensitive to phase separation and crystallization. The melt transition from a Newtonian to a non-Newtonian behavior during crystallization is a phenomenon that has been observed and extensively discussed in the literature [2–4]. Therefore, glass viscosity prediction remains difficult even for compositions with few oxides. To illustrate,

three simple SBN (SiO₂-B₂O₃-Na₂O) glass compositions were selected to compare their viscosity at 1200 °C as predicted by several models available in the literature. Results are given in Fig. 1. Predictions were compared to the experimental value measured by using a device and protocol well documented in the literature [5]. Significant discrepancy was observed among the predicted values calculated from the different models.

In the case of glasses containing a higher number of oxides, it has been shown that viscosity predictions obtained from recent Machine Learning (ML) models based on Neural Nets can deviate by up to seven orders of magnitude far from the reported values [7]. For these reasons, a new approach was developed to better account for the broad variations in glass melt behavior across composition and temperature ranges.

2. Data collection and subsampling method for model training

Database for glass viscosity prediction was built using both data available in the literature [6] and a proprietary dataset. The global database includes 15 587 glasses, all silicate glasses (Table 1). Each of these glasses contains either one or more viscosity values within the temperature range of 900 °C to 1500 °C, or Vogel-Fulcher-Tammann (VFT) coefficients.

Several traditional ML techniques were initially applied to the full dataset to identify the most predictive viscosity model. The techniques included artificial neural network, boosted tree, random forest, SVM and

* Corresponding author.

E-mail address: damien.perret@cea.fr (D. Perret).

multilinear regression. A benchmark of these techniques was performed on the same dataset by comparing their performance in predicting glass viscosity on a test data sample. Classical approach widely used in data science consisting in separating the total sample into random subsamples was applied. At the end, no model was found to be sufficiently predictive due to the very wide range of compositions investigated.

Instead of partitioning the entire dataset into training, validation and test subsets, it was decided to implement a dynamic and automatic method of subsampling. This method was dynamic because the dataset selected to train the model depended on the glass composition of interest. Thus, only similar glasses were used to build the viscosity model ensuring more accurate predictions, as presented in the Results and discussion section. The method was automatic because subsampling was automatically performed by algorithms integrated in the tool.

In the objective of increasing the robustness of the prediction, two methods of subsampling were developed, evaluated, and implemented. The first method was based on the design of experiments (DOE) methodology, while the second method was based on a procedure of iterative enlargement around the composition of interest.

2.1. Method #1 – virtual design of experiment

DOE methodology has been extensively described in the literature [8] and has demonstrated its applicability for predicting glass properties on narrow composition domains [9,10]. For the present work, an optimal DOE was automatically generated and centered around the composition of interest. The optimality criterion was related to the average prediction variance. Then, algorithms were used to replace each run of the virtual DOE by the most similar experimental data present in the global database (Fig. 2a). Similarity criteria were based on mathematical distance calculations and physics-informed knowledge on the role played by glass oxides on melt viscosity. Factors such as the oxide type (e.g. former, modifier, intermediate, none) was taken into consideration, as well as oxide mass fraction in the glass composition. Final DOE was not rigorously optimal compared to initial virtual DOE, however, it was observed that predictions were much better than considering the whole database. Three types of models were applied on this training set, as described later.

2.2. Method #2 – iterative enlargement

The iterative enlargement procedure involved generating training sets with progressively increasing amounts of data around the composition of interest and predictive models were built on this training set. Similarly to the first method, three types of predictive models were built. Statistical criteria based on R^2 , adjusted R^2 , and prediction variance were computed and stored. Then, a second dataset was generated around the composition of interest by increasing distances, leading to a

Table 1

(a) Number of compositions and oxide content distribution in the viscosity database. Only oxides present in at least 10 % of the total number of glasses are indicated. (b) Distribution of viscosity values in the database.

Oxide	Compositions containing the oxide		Oxide content (weight %) in the dataset				
	(count)	(%)	25 % quartile	50 % quartile	75 % quartile	min	max
SiO ₂	15 587	100	42,0	55,5	64,5	19,4	100,0
Al ₂ O ₃	11 770	76	0,1	6,9	15,4	0	55,9
CaO	11 426	73	0	5,8	18,7	0	61,8
MgO	8 329	53	0	0,2	4,0	0	42,8
Na ₂ O	7 924	51	0	0,1	10,4	0	55,8
K ₂ O	4 999	32	0	0	0,8	0	61,1
B ₂ O ₃	4 720	30	0	0	2,7	0	74,8
BaO	3 638	23	0	0	0	0	75,0
SrO	3 010	19	0	0	0	0	63,3
Fe ₂ O ₃	2 777	18	0	0	0	0	65,6
ZeO ₂	2 462	16	0	0	0	0	24,9
TiO ₂	2 246	14	0	0	0	0	50,0
Li ₂ O	2 178	14	0	0	0	0	40,3
FeO	1 976	13	0	0	0	0	75,0
ZnO	1 635	10	0	0	0	0	62,3

Temperature (°C)	Number of viscosity values in the dataset (count)	Distribution of viscosity values in the dataset (dPa.s)				
		25% quartile	50% quartile	75% quartile	min	max
900	4 122	2.5E4	1.2E5	2.1E6	1.7	2.9E13
1000	4 613	3.7E3	1.4E4	1.1E5	1.0	3.0E13
1100	5 566	9.3E2	3.6E3	2.6E4	1.0	3.1E13
1200	7 154	3.0E3	1.3E3	8.5E3	1.0	3.2E13
1300	8 094	5.6E1	4.2E2	2.9E3	1.0	2.5E13
1400	8 463	1.4E1	1.5E2	1.0E3	1.0	5.6E12
1500	6 807	5.0E0	5.6E1	4.5E2	1.0	4.7E12

more populated training set. Additional larger and larger training sets were generated and statistical criteria were computed and stored (Fig. 2b). When the predictive quality decreased, algorithm stopped the process.

2.3. Predictive modeling

For both subsampling methods, three types of models were implemented (Fig. 2c). The first was a Partial Quadratic Model (PQM), which included main effects and selected cross-effects. PQM has previously demonstrated strong predictive performance for glass melt viscosity [11]. Cross effects included in PQM were specific to each composition. They were automatically selected at stepwise regression step according

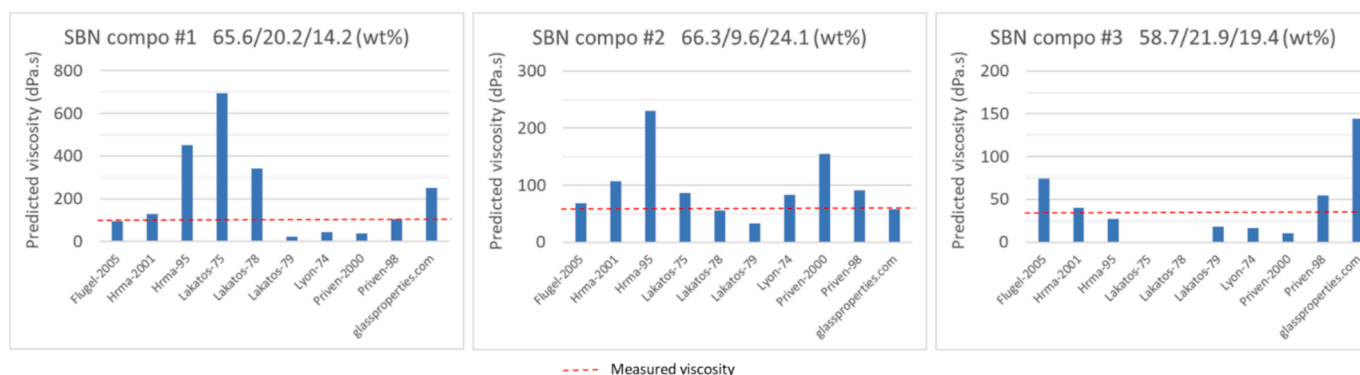


Fig. 1. Comparison of SBN glass melt viscosity at 1200 °C predicted from models available in the literature [6].

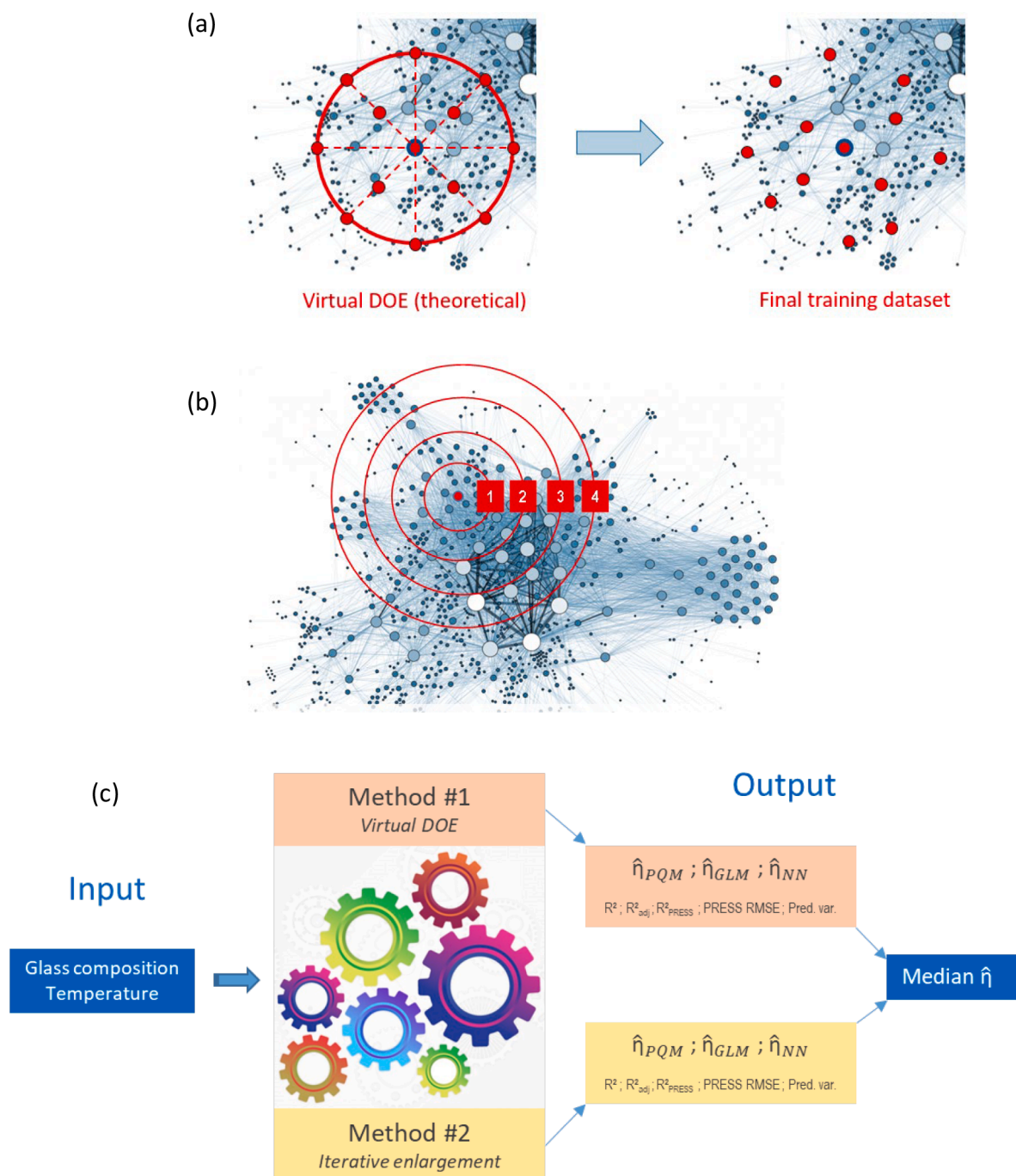


Fig. 2. Schematic representation of the subsampling methods using (a) virtual design of experiments and (b) iterative enlargement methodologies, and global description of the viscosity prediction tool (c).

to Bayes Information Criterion (BIC). Second model was a Generalized Linear Model (GLM). GLM is a relevant approach for non-normally distributed data [12]. Third model was Neural Nets (NN). NN are widely used in ML for several reasons, including their prediction capability in the presence of non-linear phenomena and without probabilistic assumptions on the response distribution. A simple architecture containing one hidden layer of three neurons with tanH activation functions was used.

3. Results and discussion

Predicted values of viscosity at 1200 °C for simple SBN compositions are given in Fig. 3a. These predictions were compared to values from

literature models and experimental measurements. Results show that the relative error of prediction, calculated as the mean of the relative errors obtained for the three SBN compositions, was approximately 20 %. This error was smaller than that obtained from literature models in their validity region, indicating that this subsampling approach was relevant for simple glass compositions. Tool capability was also evaluated by predicting 1200 °C viscosity on a dataset of 230 glasses, containing sodo-alumino silicate and borosilicate glass compositions (Fig. 3b). Number of oxides was up to 40 in the case of nuclear waste borosilicate glasses. Relative errors of prediction are given in Fig. 3c. Overall, highest relative errors were found to be less than one order of magnitude. For 90 % of the compositions, relative error was smaller than 77 %. Most accurate predictions were obtained on borosilicate nuclear

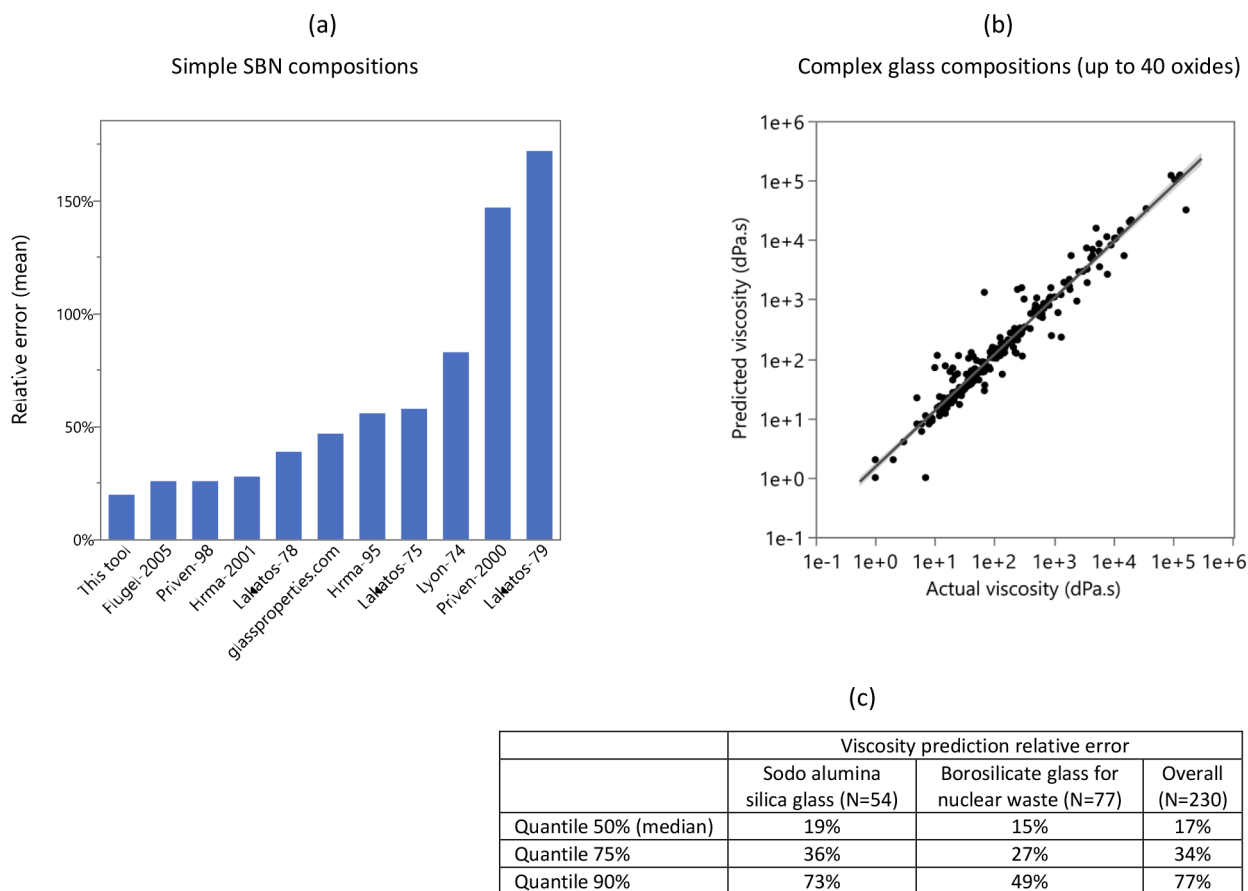


Fig. 3. (a) Relative error of prediction for glass melt viscosity at 1200 °C (simple glass compositions). (b) Predicted versus actual viscosity at 1200 °C for complex compositions and (c) quantile distribution of the relative errors.

glass compositions. Main reason was attributed to the reliability of the experimental data, in term of both device accuracy and glass melt homogeneity at the viscosity measurement temperature (1200 °C). Evaluating the tool predictive capability on a bigger test dataset and over the whole range of temperature will be an important perspective of this work.

4. Conclusion

This study focused on developing a predictive tool for glass melt viscosity calculation. A database containing about 16 000 compositions was built from both data available in the literature and proprietary dataset. An innovative approach for subsampling was developed for model training. PQM, GLM and NN models were implemented in the tool. The results demonstrated promising prediction accuracy for the viscosity at 1200 °C across various types of silicate glasses. This dynamic subsampling methodology could be applied to any type of materials and properties.

CRedit authorship contribution statement

Damien Perret: Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Alexandra Garcin:** Methodology, Investigation, Data curation. **Carole Soual:** Investigation. **François Bergeret:** Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was done in collaboration with Orano and EDF.

Data availability

The authors do not have permission to share data.

References

- [1] M.B. Volf, *Mathematical approach to glass*, Elsevier Science Publishers, Amsterdam, 1988.
- [2] A.-M. Lejeune, P. Richet, Rheology of crystal-bearing silicate melts: An experimental study at high viscosities, *J. Geophys. Res. Solid Earth* 100 (B3) (1995) 4215–4229.
- [3] D. Perret, et al., Crystallization and rheology of Na₂O–CaO–Al₂O₃–SiO₂ melt in the vitrification of technological waste, *Int. J. Appl. Glas. Sci.* 15 (3) (2024) 227–242.
- [4] J. Jiusti, et al., Precipitation of cerianite crystals and its effect on the rheology of a simplified nuclear glass melt, *Int. J. Appl. Glas. Sci.* 14 (4) (2023) 502–521.
- [5] J. Puig, et al., Rheological properties of nuclear glass melt containing platinum group metals, *Procedia Mater. Sci.* 7 (2014) 156–162.

- [6] SciGlass, *SciGlass database, publicly available under the Open Database License* (<https://github.com/epam/SciGlass>) 2019.
- [7] D.R. Cassar, ViscNet: Neural network for predicting the fragility index and the temperature-dependency of viscosity, *Acta Mater.* 206 (2021) 116602.
- [8] G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition*, ed, Wiley, 2005.
- [9] V. Piovesan, et al., Effect of composition on peraluminous glass properties: An application to HLW containment, *J. Nucl. Mater.* 483 (2017) 90–101.
- [10] B. Fleury, et al., *Development of an Experimental Design to Investigate the Effects of R7T7 Glass Composition on the Residual Rate of Alteration*, Elsevier Science Bv, Amsterdam, 2014.
- [11] J.D. Vienna, et al., *Glass Property Data and Models for Estimating High-Level Waste Glass Volume*, Battelle Rapport PNNL-18501 (2009).
- [12] P. McCullagh, J.A. Nelder, *Generalized Linear Models*, 2nd ed., Chapman & Hall, London, 1989.