SUPPLEMENTARY INFORMATION to Predicting the Dissolution Kinetics of Silicate Glasses by Topology-Informed Machine Learning

Han Liu[a], Tony Zhang[a], N M Anoop Krishnan[a,b,c], Morten M. Smedskjaer[d], Joseph V. Ryan[e], Stéphane Gin[f], Mathieu Bauchy[a]*

[a] *Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), Department of Civil and Environmental Engineering, University of California, Los Angeles, California, 90095, USA*

[b] *Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India*

[c] *Department of Material Science and Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India*

[d] *Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark*

[e] *Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, 99352, USA*

[f] *CEA Marcoule, DE2D SEVT, F-30207 Bagnols-sur-Ceze, France*

* Corresponding author: bauchy@ucla.edu

Model Parameters

(1) Model I: $\mathrm{DR} = f(\mathrm{pH}, \mathrm{Na_2O}, \mathrm{Al_2O_3})$

See model parameters in attached file: Model_parameter.xlsx

(2) Model II: $\log(\mathrm{DR}) = f(\mathrm{pH}, \mathrm{Na_2O}, \mathrm{Al_2O_3})$

See model parameters in attached file: Model_parameter.xlsx

(3) Model III: $\log(\mathrm{DR}) = f(\mathrm{pH_{acid}}, \mathrm{pH_{base}}, \mathrm{Na_2O}, \mathrm{Al_2O_3})$

a) Interpolation: $\log(\mathrm{DR}) = 0.18034663 \times \mathrm{pH_{acid}} + 0.40036747 \times \mathrm{pH_{base}} + 0.04457994 \times \mathrm{Na_2O} + 0.03066934 \times \mathrm{Al_2O_3} - 14.23734127$

b) Extrapolation: $\log(\mathrm{DR}) = 0.14388426 \times \mathrm{pH_{acid}} + 0.37579389 \times \mathrm{pH_{base}} - 6.9388939E^{-17} \times \mathrm{Na_2O} + 0.01969264 \times \mathrm{Al_2O_3} - 12.86885610$

(4) Model IV: $\log(\mathrm{DR}) = f(\mathrm{pH_{acid}}, \mathrm{pH_{base}}, n_c, \mathrm{Na_2O})$

a) Interpolation: $\log(\mathrm{DR}) = 0.17960919 \times \mathrm{pH_{acid}} + 0.40040157 \times \mathrm{pH_{base}} - 2.11986337 \times \mathrm{n_c} + 0.00254944 \times \mathrm{Na_2O} - 6.432819099$

b) Extrapolation: $\log(\mathrm{DR}) = 0.14344481 \times \mathrm{pH_{acid}} + 0.37623316 \times \mathrm{pH_{base}} - 1.33972787 \times \mathrm{n_c} + 2.44249065E^{-15} \times \mathrm{Na_2O} - 8.59776591$

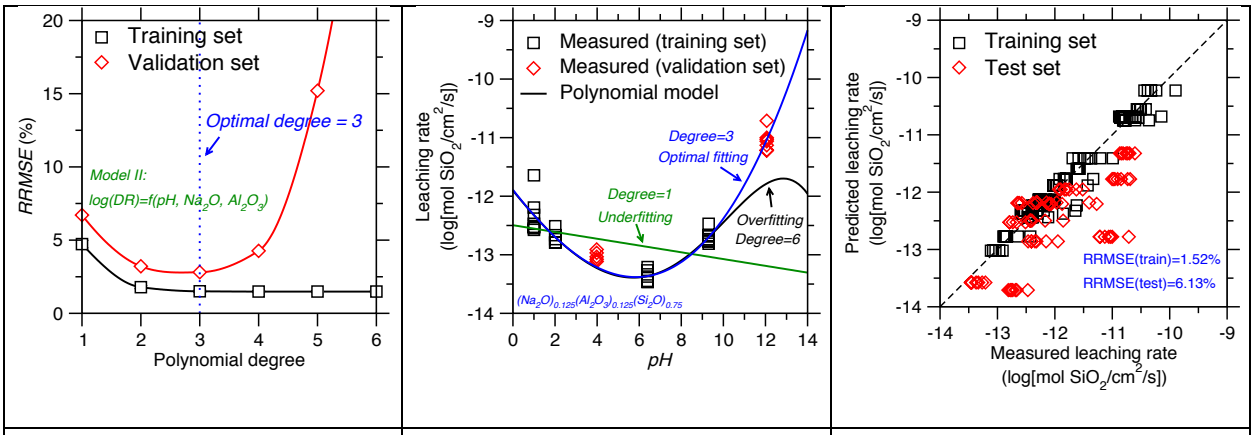
(5) Model IV-a (Interpolation): $\log(\mathrm{DR}) = f(\mathrm{pH_{acid}}, \mathrm{pH_{base}}, \mathrm{BS}, \mathrm{BB})$

$$\log(\mathrm{DR}) = 0.17960919 \times \mathrm{pH_{acid}} + 0.40040157 \times \mathrm{pH_{base}} - 0.24683845 \times \mathrm{BS} - 0.91514435 \times \mathrm{BB} - 12.12852431$$

(6) Model IV-b: $\log(\mathrm{DR}) = f(\mathrm{pH_{acid}}, \mathrm{pH_{base}}, n_c^{Si}, n_c^{Al})$

$$\log(\mathrm{DR}) = 0.17960919 \times \mathrm{pH_{acid}} + 0.40040157 \times \mathrm{pH_{base}} - 3.24001683 \times n_c^{\mathrm{Si}} - 2.64210665 \times n_c^{\mathrm{Al}} - 9.73325611$$

Model II: Interpolated Prediction



(a) **(b)** **(c)**

Fig. S1: Predictions from machine learning while explicitly accounting for the exponential dependence of the dissolution rate on the inputs ("Model II"). **(a)** Evolution of the relative root square mean square error (RRMSE) of the training and validation sets with respect to the polynomial degree p. The minimum in the RRMSE of the validation set indicates $p = 3$ as an optimal polynomial degree. **(b)** Predicted dissolution rate as a function of pH for $p = 1$ (underfitted model), $p = 3$ (optimal model), and $p = 6$ (overfitted model). The glass composition is here fixed as $(Na_2O)_{0.125}(Al_2O_3)_{0.125}(SiO_2)_{0.75}$. **(c)** Predicted dissolution rate for $p = 3$ as a function of the measured dissolution rate.

Fig. S1(a) shows the RRMSE of the training and validation sets as a function of the maximum polynomial degree p for Model II. Overall, (i) the RRMSE of the training set monotonically decreases with increasing model complexity (i.e., increasing polynomial degree) and (ii) the RRMSE of the validation set initially decreases with increasing model complexity in underfitted regions, shows a minimum (here found at $p = 3$), and eventually increases with increasing model complexity in overfitted regions. For illustration purpose, Fig. S1(b) shows the dissolution rate predicted by different models as a function of pH at fixed glass composition $(Na_2O)_{12.5}(Al_2O_3)_{12.5}(SiO_2)_{75}$. As expected, we observe that the underfitted model ($p = 1$)

is too simple to properly describe the non-linear pH-dependence of the dissolution rate. On the other hand, the overfitted model (p = 6) fits the noise of the data and does not succeed at properly predicting the validation set (e.g., at pH 12). In turn, the model presenting the optimal degree of complexity properly captures the overall trend.

Notably, we note that Model II is significantly less complex than Model I, since the optimal polynomial degree (i.e., for which the RRMSE of the validation set is minimum) is here found to be 3 (as compared to 5 for Model I). In addition, we find that the accuracy of the model is drastically improved as the RRMSE of the training set becomes equal to 1.52% (as opposed to 98% for Model I), which indicates that the model can properly interpolate the training set. More importantly, we find that Model II offers a very good prediction of the test set (RRMSE = 6.13%), which shows that the model can offer realistic predictions for unknown data. This demonstrates that predicting the logarithm of the dissolution rate (rather than the dissolution rate itself) yields a less complex model with a significantly enhanced accuracy.

Determination of $n_c^{guessed}$

The number of topological constraints per atom $n_c^{guessed}$ "guessed" by Model III can be determined by setting equal the exponent terms in Eqs. (5) and (6):

$$a[\mathrm{Na_2O}] + b[\mathrm{Al_2O_3}] = \left[\frac{-n_c E_0}{RT}\right]$$

so that $n_c^{guessed}$ can be estimated using the following format:

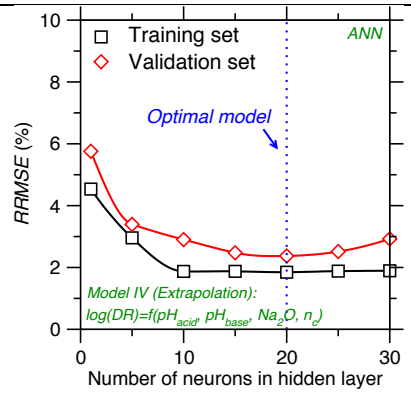
$$a[\mathrm{Na_2O}] + b[\mathrm{Al_2O_3}] = c\left[n_c^{\mathrm{guessed}}\right]$$

where c is a constant coefficient estimated by taking as the fitting coefficient of n_c used in Model IV. Finally, the guessed n_c can be expressed as:

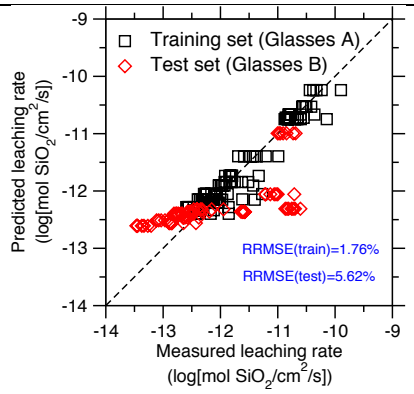
$$n_c^{\text{guessed}} = \frac{a[\text{Na}_2\text{O}] + b[\text{Al}_2\text{O}_3]}{c} + d$$

where d is a constant term added on purpose to make the calculated value comparable with true n_c values, which as $d = 11/3$ (i.e., the theoretical number of constraints of pure quartz [1]).

Model IV Extrapolation based on Artificial Neural Network (ANN)



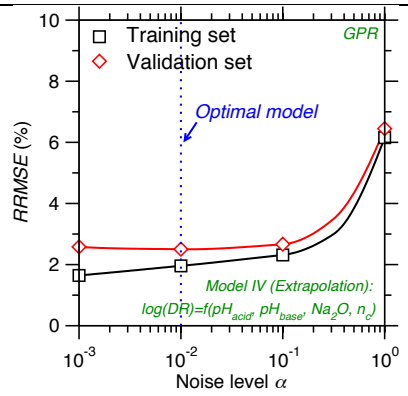
(a)



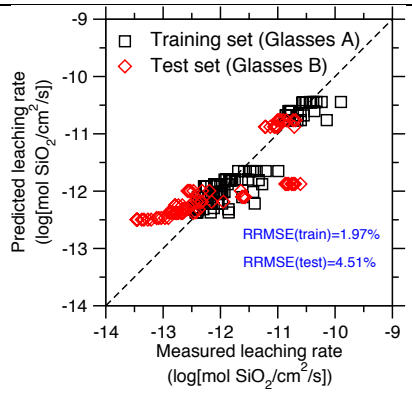
(b)

Fig. S2: Predictions from "topology-informed" machine learning ("Model IV") using artificial neural network (ANN). **(a)** Evolution of the relative root square mean square error (RRMSE) of the training and validation sets with respect to model complexity quantified as the number of neurons in hidden layer. Only one hidden layer is used in this case. The minimum in the RRMSE of the validation set indicates 20 as an optimal number of neurons. **(b)** Predicted dissolution rate for a number of neurons of 20 as a function of the measured dissolution rate—wherein the dissolution data of Glasses A ($(Na_2O)_{0.25}(Al_2O_3)_x(SiO_2)_{0.75-x}$, training set) are used as a training set to predict the dissolution kinetics of Glasses B ($(Na_2O)_x(Al_2O_3)_x(SiO_2)_{1-2x}$, test set).

Model IV Extrapolation based on Gaussian Process Regression (GPR)



(a)



(b)

Fig. S3: Predictions from "topology-informed" machine learning ("Model IV") using Gaussian process regression (GPR). **(a)** Evolution of the relative root square mean square error (RRMSE) of the training and validation sets with respect to model complexity quantified as the dataset's noise level α. The GPR kernel used herein is Matern kernel. The minimum in the RRMSE of the validation set indicates 10^{-2} as an optimal noise level. **(b)** Predicted dissolution rate for $\alpha = 10^{-2}$ as a function of the measured dissolution rate—wherein the dissolution data of Glasses A ($(Na_2O)_{0.25}(Al_2O_3)_x(SiO_2)_{0.75-x}$, training set) are used as a training set to predict the dissolution kinetics of Glasses B ($(Na_2O)_x(Al_2O_3)_x(SiO_2)_{1-2x}$, test set).

References

1. Bauchy, M., Qomi, M. J. A., Bichara, C., Ulm, F.-J. & Pellenq, R. J.-M. Rigidity Transition in Materials: Hardness is Driven by Weak Atomic Constraints. *Phys. Rev. Lett.* **114**, 125502 (2015).