

Machine Learning Enabled Models to Predict Sulfur Solubility in Nuclear Waste Glasses

Xinyi Xu, Taihao Han, Jie Huang, Albert A. Kruger, Aditya Kumar,* and Ashutosh Goel*

Cite This: <https://doi.org/10.1021/acsami.1c10359>

Read Online

ACCESS |



Metrics & More



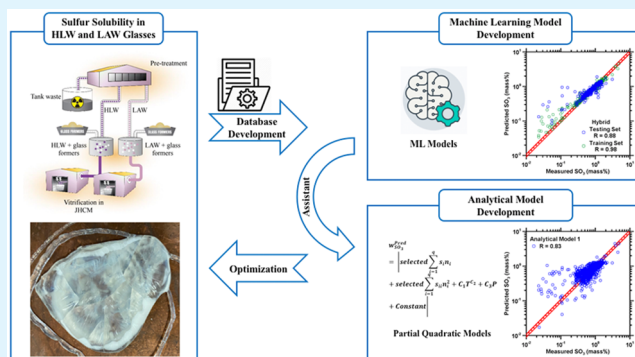
Article Recommendations



Supporting Information

ABSTRACT: The U.S. Department of Energy is considering implementing the direct feed approach for the vitrification of low-activity waste (LAW) and high-level waste (HLW) at the Hanford site in Washington state. If implemented, the nuclear waste with a higher concentration of alkali/alkaline-earth sulfates (than expected under the previously proposed vitrification scheme) will be sent to the vitrification facility. It will be difficult for the existing empirical models to predict sulfate solubility in these glasses or design glass formulations with enhanced sulfate loadings in such a scenario. Further, the existing models are unable to produce reliable predictions when applied to HLW glasses whose composition falls outside of the range encompassed by the database used to develop/calibrate the models. Accordingly, this study harnesses the power of artificial intelligence (machine learning, ML) with a goal to address the limitations of the existing models. Toward this, three ML models have been trained using a large database; comprising >1000 LAW and HLW glasses and encompassing a wide range of glass compositions and processing variables. Next, the ML model with the best prediction performance has been used to quantitatively assess and rank the influence (i.e., importance) of glasses' compositional/processing variables on the SO_3 solubility in the glasses. Finally, on the premise of such understanding of influential and inconsequential variables, two closed-form analytical models—with disparate degrees of complexity (one highly parametrized and one with fewer input variables)—have been developed. Results show that both analytical models produce predictions of SO_3 solubility in LAW and HLW glasses with an accuracy analogous to ML models and substantially higher than the analytical models that represent the current state-of-the-art. Overall, this study's outcomes present a roadmap—informed by data and channeled by artificial intelligence—that can be leveraged in the future to design nuclear waste glasses with unprecedented sulfur loadings.

KEYWORDS: sulfate solubility, nuclear waste glass, machine learning, analytical model, prediction



1.0. INTRODUCTION

Sulfur (in the form of sulfate, SO_4^{2-})—owing to its low solubility (<1 mass %) in borosilicate glasses—is one of two critical factors impeding the loading of Hanford low-activity waste (LAW) in vitreous waste forms¹ [N.B.: *long-term chemical durability is the other critical factor*]. Exceeding the solubility limit of SO_4^{2-} (incorporated and measured as SO_3) in borosilicate-based nuclear waste glass melt results in the formation of an immiscible sulfate-rich salt layer (known as “gall” or yellow phase, whose formation is facilitated by the presence of chromates in the waste) that floats on the top of the melt.^{2–5} This low-viscosity, electrically conductive salt layer, on account of being rich in water-soluble alkali sulfates along with specific radionuclides (e.g., Tc, Sr, and Cs), can cause several problems for the melter, for example, corrosion of sidewalls and metallic components, thus severely impacting its performance and reduce its longevity.^{1,6–8} In order to avoid the above-mentioned complications—and for safe, long-term operation of the melter—empirical (analytical) models have

been developed to predict the practical limit of salt (SO_3) solubility in the melt as a function of feed composition. For example, Vienna et al.,¹ on the basis of 253 simulated Hanford LAW glass compositions, established an empirical model (shown in eq 1) to predict SO_3 solubility in LAW glasses as a function of the concentrations of different components in the melter feed.

$$W_{\text{SO}_3}^{\text{Pred}} = \sum_{i=1}^q s_i n_i + \text{selected} \left(\sum_{i=1}^q s_i n_i^2 + \sum_{j=1}^{q-1} \sum_{k=j+1}^q s_{jk} n_j n_k \right) \quad (1)$$

Special Issue: Artificial Intelligence/Machine Learning for Design and Development of Applied Materials

Received: June 4, 2021

Accepted: August 25, 2021

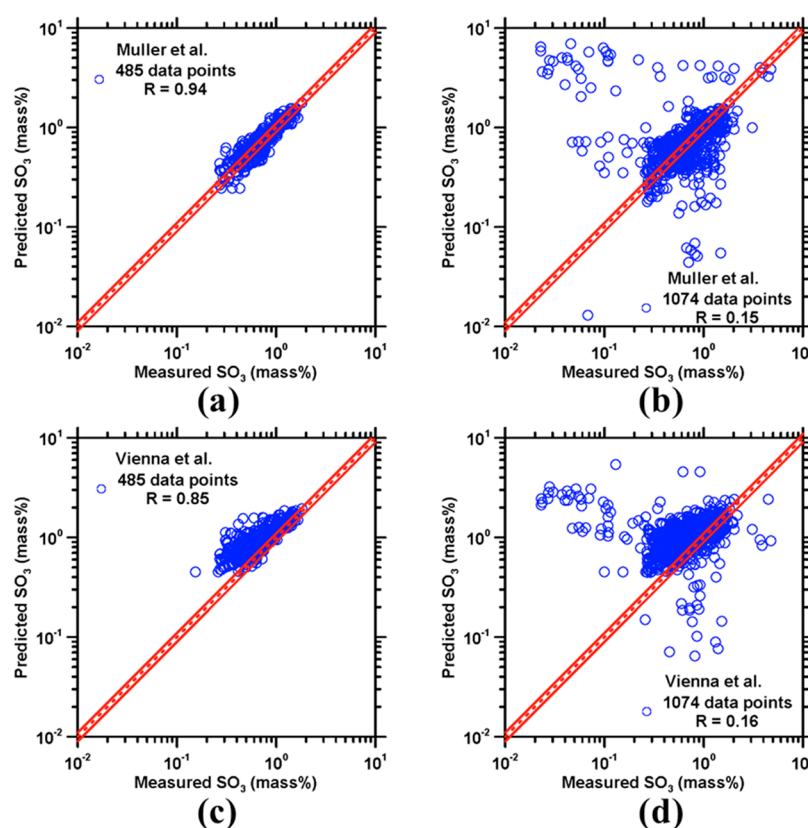


Figure 1. Predictions of SO_3 solubility—as produced by empirical models advanced by Vienna et al.¹ and Muller et al.⁹—in (a, c) 485 LAW and (b, d) 1074 LAW and HLW glasses compared against experimental measurements. As a measure of the models' prediction performances, the Pearson correlation coefficient (R) of SO_3 solubility predictions and the corresponding models used to make the predictions, are indicated in the legends. The dashed line represents the line of ideality, and the solid lines represent a $\pm 10\%$ bound.

Here, $W_{\text{SO}_3}^{\text{Pred}}$ is the predicted sulfur solubility (in the units of mass %); q represents the number of compositional components in the glass; and s_i , s_{jk} , n_j , and n_k are coefficients and normalized mass fractions, respectively, corresponding to each component (i , or j , or k , or jk). Muller et al.⁹ extended this model to a 31 term, partial quadratic mixture model (shown in eq 2) using the data of 485 LAW glasses, belonging to Waste Treatment and Immobilization Plant (WTP) and the U.S. Department of Energy, Office of River Protection (DOE-ORP) streams, with SO_3 content in the glasses ranging from 0.15 to 1.83 mass %.

$$W_{\text{SO}_3}^{\text{Pred}} = \sum_{i=1}^{18} s_i n_i + \text{selected} \left\{ \sum_{i < j}^{q-1} \sum_j^q s_{ij} n_i n_j \right\} \quad (2)$$

where $W_{\text{SO}_3}^{\text{Pred}}$ is the predicted sulfur solubility (in the units of mass %); q represents the number of components in the composition; and s_i , s_{ij} , n_i , and n_j are coefficients and normalized mass fractions, respectively, of each component (i or j or ij).¹⁰

Both empirical models—advanced by Vienna et al.¹ and Muller et al.⁹—produce reasonably accurate predictions of SO_3 solubility, provided the following caveat is met: the models are applied to borosilicate-based LAW glasses that are covered in the limited compositional space used to develop the respective model. As can be seen in Figure 1, when the aforesaid caveat is met, the Pearson correlation coefficient (R) of predictions produced by the models proposed by Vienna et al.¹ and Muller et al.⁹ are high, that is, 0.85 and 0.94, respectively.

Furthermore, the dependency of SO_3 solubility on the composition of LAW glasses, as predicted by these models, is in good agreement with those reported in the literature.^{2,3,11–14} Conversely, when applied to a larger data set—comprising 1074 sulfur-containing LAW as well as high-level waste (HLW) glasses (discussed further in Section 2.0)—the predictions produced by the two models feature substantial departures from experimental measurements. This is demonstrated in Figure 1, which shows that the R of predictions of SO_3 solubility in the aforementioned 1074 glasses, as produced by the models proposed by Vienna et al.¹ and Muller et al.⁹ are only 0.16 and 0.15, respectively. Such precipitous decline in the models' prediction performances—that is, a drop in R from a high value of ~ 0.90 to a very low value of ~ 0.15 —can be attributed to the following reasons.

(A) Limited volume/diversity of the database: In general, a given empirical model is only valid (and, thus, applicable) within the *compositional domain* encompassed by the database (i.e., collection of data records representing SO_3 solubility in glasses of different compositions) used to develop, calibrate/fine-tune, and validate the model. In the context of nuclear waste glasses, the *compositional domain* can be described in terms of the compositional range (minimum-to-maximum) of each elemental and elemental-oxide component (e.g., SiO_2 , Al_2O_3 , SO_3 , and B_2O_3 , etc.) of the glasses embodied in the database. The above-discussed empirical models were developed on the bases of only 253 (Vienna et al.¹) and 485 (Muller et al.⁹) unique data records (i.e., glass compositions). As databases of limited volume were used in the models'

development, the models' abilities to produce accurate predictions are—unsurprisingly—limited to glass compositions that are similar (or identical) to those of glasses included in the original database used to develop/calibrate/fine-tune the models in the first place. The prediction performance of each model is further compromised due to constraints that arise from using a database with limited compositional diversity. For example, the model advanced by Muller et al.⁹ was developed/calibrated on the basis of—and, thus, is only applicable for—LAW glasses in which the sum of alumina (Al_2O_3) and silica (SiO_2) contents is strictly between 36.6 and 59.5 mass %. When $\text{Al}_2\text{O}_3 + \text{SiO}_2$ of the given glass falls outside of the aforesaid range, as is the case with several LAW glass compositions that are a part of the database used in the present study, the prediction of SO_3 solubility in such glass, as produced by the model, is no longer reliable. To put it simply, a larger, more diverse, and more up-to-date experimental database, comprising a wide range of SO_3 solubility in glasses spanning a broad compositional space, is essential to develop models with high prediction accuracy.

(B) Differences in LAW and HLW glass chemistry: The implementation of the *direct feed approach* not only poses a challenge for the established models in terms of predicting SO_3 solubility in LAW glasses but it is also expected to exacerbate the complexity of HLW vitrification by considerably increasing the sulfate content in the waste stream (since the pretreatment of HLW for sulfate removal before vitrification shall be bypassed).¹⁵ Nevertheless, the existing SO_3 solubility models do not perform well with HLW glass compositions (Figure 1). This is because these models were developed and calibrated on the basis of SO_3 solubility in LAW glasses, in which the compositional ranges of the constituent oxides are significantly different from their HLW counterparts.¹⁶ Thus, a unified model needs to be developed that can be used to predict SO_3 solubility in both LAW and HLW glasses.

(C) Unaccountability of influential experimental conditions: The databases used to develop the two models^{1,9} were originally consolidated by collecting data pertaining to sulfur-containing glasses that were prepared by melting the mixture of sulfur-free glass frit and Na_2SO_4 at 1150 °C for 1 h. However, the literature on this subject discusses different approaches to saturate the glass melt with sulfur: for example, multiple melting cycles of the glass frit and the sulfate salt;¹⁷ or introducing sulfur in the melt in the form of a gas mixture comprising SO_2 , O_2 , and N_2 .¹ Depending on the technique used to saturate the glass melt with sulfur (i.e., one of the three techniques discussed above), the resultant sulfur loading in the LAW glasses has been shown to produce different results. However, as the existing models were developed on the basis of LAW glasses prepared using only one specific sulfur-saturation technique, they are essentially unable to account for the impact of experimental/processing conditions on the sulfur loading in the nuclear waste glasses. Therefore, to improve the models' prediction performance, it is critical they be extended to account for potential disparities in sulfur loading of the glasses that arise from different experimental/processing conditions (e.g., melting time and temperature).

The work presented in this article aims to overcome the above-discussed limitations of the current models by advancing closed-form analytical models—derived from machine learning (ML) techniques—that can produce prompt and accurate predictions of SO_3 solubility in not just LAW but also HLW glasses. In the field of glass science, ML models have been

applied to predict a wide range of glass properties,¹⁸ including Young's modulus,^{19,20} liquidus temperature,²¹ solubility,²² and dissolution kinetics.^{23,24} It is acknowledged that SO_3 solubility in borosilicate glasses is a complex function of their chemical and structural descriptors, as well as the processing techniques used to produce the glasses. Admittedly, due to this complexity, the dependency of SO_3 solubility on the glasses' compositional/processing variables is not well understood. Therefore, it cannot be readily expressed in the form of simple, closed-form equations. Thus, herein, three ML models, i.e., random forests (RFs), support vector machine (SVM), and hybrid model (metaheuristic hybridization of RF and SVM models), have been employed to establish and reveal such dependency of SO_3 solubility on compositional/processing variables of the glasses in an explicit, quantitative manner [N.B.: short descriptions of these ML models, methods used to optimize each model's hyperparameters, and relevant references to published articles (wherein the underlying algorithms/structures implemented within the models are described in further detail) are provided in the Supporting Information]. To account for both LAW and HLW glasses, the models are trained using a larger database that comprises the following: (1) 1074 distinct glass compositions, wherein the compositions encompass a blend of LAW and HLW glasses; and (2) a wide range of experimental/processing conditions (e.g., melting time and temperature, the method to introduce SO_3 in the melt, and so on). The ML model with the best performance (as evaluated, statistically, on the basis of each model's prediction performance) is used to quantitatively assess and rank the influence (i.e., importance) of glasses' compositional/processing variables on the SO_3 solubility in the glasses. Finally, on the premise of such understanding/identification of influential and inconsequential variables, two closed-form (partial quadratic) analytical models—a highly parametrized model and a simple model with fewer parameters—are developed to predict SO_3 solubility in both LAW and HLW glasses. To evaluate the analytical models' prediction performance, their predictions are benchmarked against experimental measurements of SO_3 solubility in 1074 distinct LAW and HLW glasses. As the models presented herein are trained and tested against a comprehensive database, they are applicable to a wide range of LAW and HLW glasses.

2.0. CONSOLIDATION AND EVALUATION OF DATABASE

The data pertaining to SO_3 solubility in 1074 LAW and HLW glass compositions have been collected from 23 publications.^{2,3,6,9,17,25–42} The majority of the data has been drawn from studies on sulfur tolerance in simulated LAW glasses;^{3,6,9,17,25–35,37,42} and the remaining data have been extracted from the experimental studies on simulated HLW glasses.^{2,31,38–41} Sulfur solubility (in nuclear waste glasses), as extracted from literature and used in this study, corresponds to the limiting sulfate concentration in the glasses measured just before the formation (first appearance) of the salt layer on the surface of the glass/melt.^{1,2,7,9,43} Here, it may be helpful to differentiate between two terms—"solubility" and "retention." The solubility is defined as the concentration of sulfur at an established equilibrium between dissolved and atmospheric sulfur (most likely in the form of oxides). However, the glassmaking process generally does not allow molten glass to reach equilibrium. While the portion of a component dissolved in the amorphous phase(s) may be unsaturated (i.e., below the

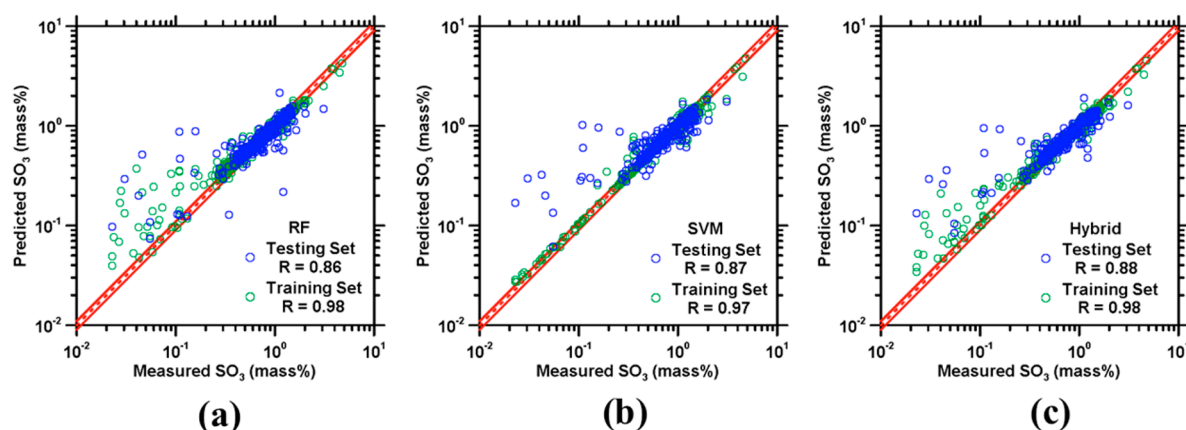


Figure 2. Predictions of SO_3 solubility in LAW and HLW glasses (in the training and testing sets, partitioned from the parent database consolidated from prior studies) as produced by (a) RF, (b) SVM, and (c) hybrid models compared against experimental measurements. 75% of the database was used for training, and the remaining 25% was used for testing. The dashed line represents the line of ideality, and the solid lines represent a $\pm 10\%$ bound.

solubility), a substantial portion may simultaneously exist in the form of salt (on the melt surface) or inclusions. Therefore, the retention is a dynamic quantity that depends on the glassmaking conditions.^{44–46} In the present study, glasses with SO_4^{2-} retention have not been considered for the model development.

In the database, each data record has 51 inputs and 1 output. The inputs include 51 pertinent experimental parameters employed in glass synthesis: normalized masses (mass %) of 47 distinct constituents (i.e., elemental and elemental oxides such as I, F, SnO_2 , and Na_2O) of the glasses, glass melting temperature, melting time (hour), SO_3 source (1 = NaSO_4 , and 2 = other sources), and method of addition of SO_3 to the melt (1 = salt saturation, and 2 = gas bubbling). The output is the SO_3 solubility in the glass (mass %). Statistical parameters pertaining to the database are shown in Supporting Information Table S1. Here, it should be noted that the glass compositions were characterized by different techniques: inductively coupled plasma–optical emission spectroscopy; inductively coupled plasma–atomic emission spectroscopy; X-ray fluorescence; electron probe microanalyzer; and energy dispersive X-ray spectroscopy. Typical standard deviations of these characterizations (e.g., of sulfate composition) range from 0.1 to 0.4 mass %.^{2,3,17} On the basis of this range, it is posited that 0.2 mass % is a suitable measure of standard deviation in experimental measurements; and, therefore, the same value will be used as an additional benchmark to evaluate the prediction performance of ML models (described below).

For the purposes of training and validating the ML models, the database was split into two subsets: (1) Training and (2) Testing. 75% (i.e., 806 of 1074) of randomly selected data records from the parent database were used to rigorously train the ML models, while the remaining 268 data records (which were kept hidden from the ML models during their training) were used as the testing set to quantitatively evaluate the ability of each ML model to predict SO_3 solubility in nuclear waste glasses of unknown (to the ML model) compositions. Justification for the 3:1 split in the parent database for training and testing of the ML models is provided in our prior publications^{24,47–50} and Supporting Information. A 10-fold cross-validation (CV) method,^{47,51} in conjunction with the grid-search method,^{48,49} was employed during the models' training. These methods were employed to (1) fine-tune the

underlying hyperparameters of the ML models; (2) develop quantitative, logical correlations between the inputs and output (SO_3 solubility), while ensuring that outliers in the database are accounted for (as opposed to being excluded from the analyses of input–output correlations) with the same gravity as data records that fall into trends; and (3) simultaneously reduce the variance (underfitting) and bias (overfitting) to their global minima [N.B.: negating overfitting is especially important, because failing to do so would result in unreliable/inaccurate predictions for glasses whose compositions deviate significantly from those included in the training database, thus significantly diminishing the applicability of the ML model²³].

In order to ascertain the prediction performance of the ML models, five independent statistical parameters—Pearson correlation coefficient (R); coefficient of determination (R^2); mean absolute percentage error (MAPE); root mean squared error (RMSE); and mean absolute error (MAE)—were used to measure the prediction errors. Mathematical formulations for each of these parameters can be found in our previous studies.^{24,47,52}

3.0. RESULTS AND DISCUSSION

3.1. Comparison of Prediction Performance of ML Models. Figure 2 presents comparisons of predictions of SO_3 solubility in nuclear waste glasses produced by the ML models (RF, SVM, and hybrid model) against measurements. Statistical parameters (R , R^2 , MAE, MAPE, and RMSE), describing the prediction accuracy of models are tabulated in Table 1.

As can be seen in Figure 2 and Table 1, all ML models produce reasonably accurate predictions of SO_3 solubility in the glasses, with R spanning a narrow range from 0.86 to 0.88

Table 1. Five Statistical Parameters (i.e., R , R^2 , MAE, MAPE, and RMSE) Used to Quantify the Accuracy of ML Models' Predictions of SO_3 Solubility in Nuclear Waste Glasses Included in the Testing Set

ML model	R	R^2	MAE, mass %	RMSE, mass %	MAPE, %
RF	0.8551	0.7312	0.1105	0.2068	28.35
SVM	0.8678	0.7531	0.1165	0.2010	31.48
hybrid	0.8765	0.7683	0.1055	0.1949	28.86

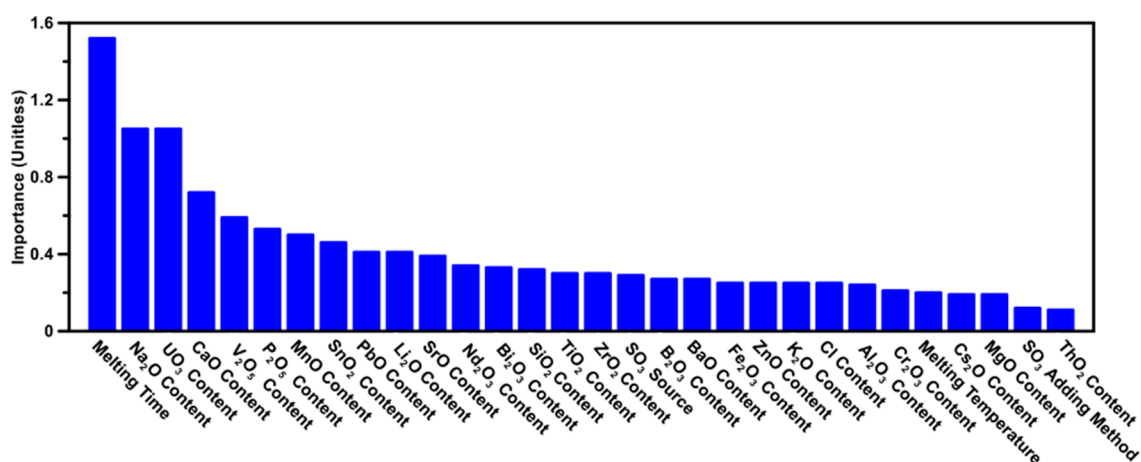


Figure 3. Quantitative evaluation of the influence of 25 (out of 30) input variables in terms of their contribution toward SO_3 solubility in nuclear waste glasses.

and RMSE ranging from 0.21 to 0.19 mass %. On the basis of the values of R , the ML models can be ranked as hybrid > SVM > RF. Other statistical parameters (i.e., R^2 , MAE, and RMSE), likewise, confirm that the hybrid model produces more accurate predictions than the other two ML models. If we consider the RMSE of the best-performing ML model (i.e., the hybrid model), the margin of error in predictions of SO_3 solubility (± 0.19 mass %) is commensurable to the typical standard deviation (± 0.20 mass %) in experimental measurements.⁶ This is significant as it entails that if the hybrid model were to be employed to predict SO_3 solubility in a LAW or HLW glass of a new composition, the prediction would—in essence—be as reliable as experimental measurement. In Figure 2, it is worth pointing out all three ML models appear to consistently overestimate the SO_3 solubility when its actual (measured) value is <0.1 mass %. The error distributions of these predictions are shown in Supporting Information. This observation can be attributed to (1) artifacts associated with plotting in logarithmic scale, which inherently results in an exaggeration of deviations between measurements and predictions at low magnitudes of the data points; and (2) bias introduced in the ML models since the majority of the data used to train the models has higher magnitude (i.e., >0.5 mass %), thereby predisposing the models to predict, more often than not, higher (than actual) magnitudes of SO_3 solubility. This issue can be readily resolved by (re)training the ML models with a larger database featuring a significant number of glasses with not just high but also low/very low SO_3 solubility.

The RF model predicts SO_3 solubility with reasonable accuracy because of features relevant to its internal architecture;^{53–55} important ones among them are recapped ahead. The large number of deep, unpruned *classification-and-regression* trees (CARTs)—all of which are grown in the ensemble (forest) by following the *three-stage randomization* principle—simultaneously minimize bias and variance while also guaranteeing that each CART generates outputs that are distinct and derived individually (i.e., with no correlation whatsoever with the other CARTs). Furthermore, cause–effect correlations established by the RF model are generally robust; this is primarily because the structure [i.e., number of secondary, tertiary, and terminal (leaf) nodes] of each CART in the forest is self-determining as it evolves—during the model’s training—strictly, and exclusively, in relation to the

unique bootstrap assigned to it. Additionally, utilization of the 10-fold CV method⁵¹ in tandem with the grid-search method⁴⁹ autonomously optimizes the hyperparameters of the RF model, thereby obviating the likelihood of biases that could have been introduced if the parameters were to be optimized manually (e.g., by fine-tuning each parameter in a stepwise manner, while keeping other parameters fixed, until R^2 is maximized or MAPE is minimized).

As can be seen in Table 1, the prediction performance of the SVM model is superior—although only slightly—compared to the RF model. Here, again, the SVM model’s prediction performance can be attributed to the underlying algorithms that drive its ability to establish correlations between the inputs and the output. The SVM model is premised on classification and optimization of the database’s assembly, wherein virtual hyperplanes (constructed, usually, using nonlinear kernel functions) are drawn within the multidimensional database to enhance the dimensionality of clusters of data records (commonly referred to as mapping) and classify them into nonoverlapping families. Because of the data being classified from lower- to higher-dimensional planes, the model is able to streamline the complex relationship between *compositional/processing variables pertaining to the glasses* and SO_3 solubility into a collection of simple ones.⁵⁶ Furthermore, in this study, the hyperparameters and kernel functions of SVM were rigorously optimized (using the 10-fold CV and grid-search methods), to ensure optimal dimensionality enhancement and classification of the data, which, in turn, are expected to improve the prediction performance of the model in new/unpopulated data domains.⁴⁷

While the standalone RF and SVM models produce reliable predictions of SO_3 solubility, they do have drawbacks. For example, the RF model produces predictions of SO_3 solubility with lower MAPE when the SO_3 content ranges between 0.5 to 1 mass%. Outside of this range, the RF model’s prediction performance declines. Past studies have demonstrated that the prediction performance of standalone ML models can be improved through genetic programming,⁵⁷ or gradient boosting techniques;⁵⁴ or by using theoretically guided constraints (e.g., by manually or autonomously removing data records that feature unrealistic values of SO_3 solubility or glass composition). In this study, to enhance the RF and SVM models’ prediction performance, the pair of models were metaheuristically hybridized into a singular, ensemble model

(i.e., the hybrid model). Metaheuristic hybridization involves the assimilation of the outputs produced independently by the two standalone models—using the *weighted voting method*⁵⁸—into a singular, ensemble output. By doing so, prediction errors and drawbacks from one model over a given subset of the data can be compensated for by the superior prediction accuracy of the other model. As shown in Figure 2 and Table 1, the hybrid model indeed yields predictions of SO₃ solubility in HLW and LAW glasses—as a function of their chemical composition, processing conditions, and sulfur incorporation methods—with superior fidelity as compared to the standalone RF and SVM models.

Application of the hybrid ML model—more specifically, the RF model included within the hybrid model—offers an important advantage: inputs–output relationships learned by the model during its training can be leveraged to quantitatively evaluate the influence of each input variable (commonly referred to as *variable importance* quantified on the basis of each variable's Gini score^{59,60}) in terms of affecting the output (i.e., SO₃ solubility). Figure 3 shows the magnitudes of *variable importance*, organized in descending order of influence of input variables on SO₃ solubility. Minor elemental and elemental-oxide components of the glass—whose contents are <5 mass % and have *variable importance* of <0.05—are not included in the list. These minor components include AgO, As₂O₃, CdO, Ce₂O₃, CoO, CuO, F, Gd₂O₃, I, La₂O₃, MoO₃, NiO, PdO, Re₂O₇, Rh₂O₃, Sb₂O₃, SeO₂, TeO₂, Ti₂O, and Y₂O₃. To the best of the authors' knowledge, none of the variables listed above have been reported to exert substantial impacts on SO₃ solubility in the prior literature.^{1,9} Exclusion of these variables is also essential from the standpoint of developing closed-form, analytical models discussed in Section 3.4.

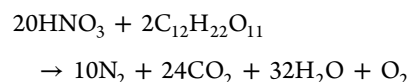
3.2. Impact of Melting Time/Temperature, SO₃ Source and Method of SO₃ Addition. On the basis of the quantification of *variable importance* (Figure 3), melting time appears to be the most influential input variable governing the solubility of SO₄^{2−} in nuclear waste glasses during the processing of the waste feed in the nuclear waste melter. This finding, derived from the hybrid ML model, is in good agreement with what has been reported in the literature. For example: in a prior study,⁶¹ it has been shown that the melting time longer than 10 h results in significant sulfate volatilization from the glass melt. In the database used in the present study, the melting time ranges between 1 and 26 h, with several data records on either side of the 10 h threshold, thus instituting melting time as a determining factor to optimize SO₄^{2−} volatility from the glass melts.

It is noteworthy that although the glass melting time has been found to be an influential variable that exerts a substantial effect on the SO₄^{2−} loading in the glass, it may be difficult to measure or control this parameter in industrial processes. In such cases, the analytical models presented in Section 4.0 of this study can be used to estimate the glass melting time.

Further, the hybrid model predicts the glass melting temperature to have minimal impact on SO₃ solubility (Figure 3). While this is an interesting finding, it is a justifiable one, considering that the melting temperatures in the database (used in the present study) range between 1100 and 1200 °C, i.e., the operating temperature range of Joule Heated Ceramic Melters (JHCMS) at the Hanford site. According to the literature,^{26,43,62} the volatilization of sulfate from alkali borosilicate and alkali aluminoborosilicate melts at temperatures below 1250 °C is relatively low. The majority of the

sulfate is removed (via volatilization) from the glass melt only when the melting temperature exceeds 1300 °C. On the basis of these findings, it can be postulated that inclusion of SO₃ solubility data at melting temperatures greater than 1250 °C would amplify its variable importance.

Figure 3 reveals another important point: the source of sulfur in the glass appears to exert a moderate impact on sulfate loadings, while the method employed to incorporate SO₃ in the glass is largely inconsequential. As has been reported before,^{1,3} sulfur can exist in glass melt in a wide range of oxidation states under various melting redox conditions, from sulfate (SO₄^{2−}) under oxidizing conditions to sulfide (S^{2−}) in more reducing atmosphere. Because the Hanford waste feeds are highly oxidizing in nature (due to the presence of nitrates),¹⁵ sulfur is expected to be present primarily as SO₄^{2−} in the Hanford LAW and DF-HLW.^{1,15,17} During vitrification, although, the reductants (for example, sucrose or oxalic acid) will be added to the waste feed to decompose nitrates, as described in the following reaction scheme, the amount added will be insufficient for the complete decomposition of nitrates or reduction of sulfur from +6 oxidation state in SO₄^{2−}.^{63,64}



Therefore, the method of addition of SO₃ to the melt should have minimal impact on its solubility in the glass structure (as observed in Figure 3).

3.3. Impact of Glass Composition on SO₃ Solubility in Nuclear Waste Glasses. According to the quantification of variable importance (Figure 3), the solubility of SO₃ in the glass melts is largely governed by the glasses' chemical descriptors (including the composition and redox conditions). This is anticipated, considering that chemical composition is at the origin of nearly every physiochemical property—including structure (i.e., degree of polymerization in the glass/melt structure and ionic field strength of the nonframework cations)^{27,62}—of the glasses (and all other materials, for that matter). From the perspective of predicting the impact of individual elemental and elemental-oxide components of the nuclear waste glasses on the SO₃ solubility, the concentrations (normalized mass, expressed as mass %) of Na₂O, CaO, V₂O₅, P₂O₅, SnO₂, and Li₂O, ranking from high to low, in the glasses are among the top 10 (out of 30) influential factors (Figure 3).

According to the literature,^{1–3,43} Na₂O content is one of the primary glass constituents driving the glasses' SO₃ solubility. This is because Na⁺ can preferentially charge-compensate SO₄^{2−} in the glasses over other higher ionic field strength cations such as Li⁺ and Ca²⁺.^{2,43} Therefore, it is not surprising that the hybrid model predicted Na₂O as the second most influential parameter (after melting time) governing SO₃ solubility in the nuclear waste glasses.

Li⁺ and Ca²⁺, being high ionic field strength cations, are known to reduce SO₃ solubility in borosilicate glasses when substituted for Na⁺.⁴³ However, when added to the glass composition (as opposed to being used as substitutes for another nonframework cation), these cations act as network modifiers, thus depolymerizing the glass structure by creating nonbridging oxygens (NBOs), resulting in higher SO₃ solubility. To better illustrate this, the case of CaO is consider, which when added to a glass system with the composition (100 − x) (25 Na₂O − 25 Al₂O₃ − 10 B₂O₃ − 40 SiO₂) − xCaO,

where $x = 0\text{--}15$ mol %, appears to depolymerize the silicate network instead of increasing the fraction of BO_3 units in the glasses (unpublished data), as is usually observed upon substitution of Na_2O by CaO .⁴⁶ The same scenario is expected in nuclear waste glasses, where Li_2O and CaO additions result in enhancement of the glasses' SO_3 solubility,^{2,65} as has been predicted in the present study (Figure 3).

On the impact of V_2O_5 on the sulfate solubility in nuclear waste glasses: the structural role of V_2O_5 in borosilicate glasses is intriguing owing to its co-existence in 4+ and 5+ oxidation states in the glass structure.^{1,2} According to McKeown et al.,⁶⁶ vanadium primarily exists as V^{5+}O_4 tetrahedra in the borosilicate glass structure, while a minor fraction (20–25%) exists as V^{4+}O_5 . Overall, V_2O_5 is known to increase the SO_3 solubility in borosilicate glasses,^{1,3,67} although the mechanism driving this increase is still unclear. According to the hypothesis proposed in the literature, vanadium and sulfate ions do not interact in the glass structure. However, V^{5+} loosens the voids in the borosilicate network, which favors the incorporation of SO_4^{2-} in the glass matrix.^{3,66,67} Per recent results (unpublished and obtained from experiments conducted in a parallel study), SO_3 solubility increases from 1.7 to 5.8 mass % with increasing V_2O_5 concentration from 0 to 7.5 mol % in the glass system $(100 - x)(30 \text{ Na}_2\text{O} - 5 \text{ Al}_2\text{O}_3 - 15 \text{ B}_2\text{O}_3 - 50 \text{ SiO}_2) - x \text{ V}_2\text{O}_5$. Thus, the high influence of V_2O_5 on sulfate solubility (as predicted by the hybrid ML model; see Figure 3) is in good agreement with the literature (and unpublished findings from related experiments). Further, P_2O_5 has been shown to improve SO_3 solubility when present in low concentration (≤ 3 mass %); its role/impact is reversed at higher concentrations.^{1,68} Also, SnO_2 has been shown to moderately reduce the SO_3 solubility in borosilicate glasses.^{1,65,69} Not surprisingly, in Figure 3, P_2O_5 and SnO_2 are ranked as influential variables.

Results shown in Figure 3 predict UO_3 to be one of the most influential factors governing the SO_3 solubility in nuclear waste glasses. This result is counterintuitive, and the authors were unable to find corroborating data in the literature. This disparity between the prediction (in the current study) and the literature can be attributed to two possible reasons: (1) majority of the literature on SO_3 solubility in borosilicate glasses is focused on LAW glass compositions, while UO_3 is primarily contained in HLW;¹⁵ and (2) although a few glass compositions (i.e., 11 samples) in the database used in the present study contain UO_3 , the SO_3 solubility in these samples exhibit much larger variation vis-à-vis the variations in glasses that were devoid of UO_3 . Similarly, MnO is a minor component (≤ 3 mass %) in LAW glasses, whereas its concentration can be as high as 10 mass % in HLW glasses. Nevertheless, due to insufficient literature elucidating the compositional dependence of SO_3 solubility in HLW glasses, the impact of MnO on SO_3 solubility is not clear from the literature. Therefore, though the results in Figure 3 show MnO to be the seventh most influential variable affecting SO_3 solubility in nuclear waste glasses, this may not be true and should be attributed to the lack of experimental data on this subject. Along the same lines, the concentration of PbO in the glass has been ranked as the ninth most important variable governing sulfate solubility. Greater variable importance of PbO may be attributed to the large variation in the PbO content of the glasses (i.e., with values as high 41 mass % in a few compositions and <1 mass % PbO in the majority of the glasses), which—purely because of such high variance in its

content—exerts significant influence in the glasses' SO_3 solubility. Finally, both chloride ions and Cr^{3+} are known to lower the sulfur solubility in borosilicate glasses.⁷⁰ However, owing to the considerable dearth of literature demonstrating the impact of these two ions on sulfur solubility—thereby resulting in limited variability in concentrations of these ions in the database used in this study—the results in Figure 3 predict them to be the least influential variables.

Other elemental-oxide components of the glasses have lower variable importance compared to the ones described above in this section. This is because these components, in spite of spanning a wide range of mass percentages in the LAW and HLW glasses, only result in imperceptible changes in SO_3 solubility.

3.4. Development of Closed-Form Analytical Models to Predict SO_3 Solubility. Results described in Sections 3.1–3.3 show that ML models—once they are adequately trained—can produce reliable predictions of SO_3 solubility in LAW and HLW glasses, solely on the basis of the glasses' compositional/processing variables. ML models, however, can be challenging to use, especially for end users (including researchers and professionals) who do not have a computer science background. Furthermore, the input–output correlations developed by ML models are difficult to interpret (e.g., the hybrid ML model expresses such correlations in the form of a highly intricate decision-tree structure); and, as such, it is difficult to gain insights into how certain physical/chemical variables influence the ultimate property of a given glass. Therefore, for widespread application and better interpretability, it is essential that ML models are distilled down to simple, closed-form analytical predictive models, without sacrificing much of the original ML model's prediction performance. This section describes two such analytical models that have been developed on the basis of results produced by the best-performing ML model (i.e., the hybrid model). In these analytical models, the input variables predicted to exert significant influence on the SO_3 solubility in nuclear waste glasses (shown in Figure 3) have been used to refine the analytical models proposed by Vienna et al.¹ (eq 1) and Muller et al.⁹ (eq 2). The objective here is to leverage the knowledge gained from the hybrid ML model to develop simple, closed-form (partial quadratic) analytical models that can produce accurate predictions of SO_3 solubility in both LAW and HLW glasses.

The development of analytical models requires a careful selection of parameters. Inclusion of influential parameters is critical for maintaining prediction accuracy. Simultaneously, exclusion of inconsequential parameters is necessary to ensure that the model's structure/format is simple (and, thus, not unnecessarily complex or highly parametrized). In the models proposed by Vienna et al.¹ (eq 1) and Muller et al.⁹ (eq 2), input parameters such as glass melting time, melting temperature, SO_3 source, and SO_3 adding method were ignored. However, as Figure 3 clearly shows, some (e.g., melting time)—though not all—of these variables exert a finite influence on SO_3 solubility in the glasses; disregarding them would unavoidably undermine the models' prediction performance. Thus, in the analytical models developed in the present study, while the general forms of the models have deliberately been kept similar to those proposed by Vienna et al.¹ and Muller et al.,⁹ all of the influential variables have been incorporated, whereas the variables that are broadly inconsequential to SO_3 solubility have been excluded. To be more

Table 2. Optimal Values of 41 Coefficients and 1 Constant (for 25 Input Variables Corresponding to the Glasses' Chemical Descriptors and Processing Parameters) for Analytical Model 1

CaO	0.057907	Na ₂ O	−0.041117	UO ₃	0.18774
MnO	−0.020066	V ₂ O ₅	0.030485	P ₂ O ₅	0.003074
Li ₂ O	0.106227	SnO ₂	−0.044722	PbO	0.02916
SiO ₂	0.052658	SrO	0.039883	Nd ₂ O ₃	0.03032
B ₂ O ₃	0.055648	TiO ₂	0.00705	ZrO ₂	−0.0376
ZnO	0.048613	BaO	0.043697	Fe ₂ O ₃	−0.01251
Cl	−0.23531	Cr ₂ O ₃	−0.05642	Bi ₂ O ₃	0.083248
V ₂ O ₅ × V ₂ O ₅	−0.00026	Na ₂ O × Na ₂ O	0.00235	MnO × MnO	0.009639
SnO ₂ × SnO ₂	0.001441	P ₂ O ₅ × P ₂ O ₅	0.000148	ZnO × ZnO	−0.002857
SiO ₂ × SiO ₂	−0.00053	Li ₂ O × Li ₂ O	−0.00362	Fe ₂ O ₃ × Fe ₂ O ₃	0.00097
B ₂ O ₃ × B ₂ O ₃	−0.00115	ZrO ₂ × ZrO ₂	0.002191	Al ₂ O ₃ × Al ₂ O ₃	0.000976
K ₂ O × K ₂ O	0.001603	BaO × BaO	−0.0003	Cr ₂ O ₃ × Cr ₂ O ₃	0.005447
Cl × Cl	0.012622	UO ₃ × UO ₃	−0.0196	C ₁	0.853466
C ₂	0.0684	C ₃	−0.13081	constant	−2.169832

specific, all input variables represented in Figure 3, with the exception of the five least significant variables, were included in the development of the analytical models. For example, the melting temperature and SO₃ adding method were excluded, but melting time and source of SO₃ were included on account of their higher ranking.

During the development of *form* (e.g., a combination of weighted linear function and partial quadratic function, as used in the models advanced by Vienna et al.¹ and Muller et al.⁹) of an analytical model, it is important to assign greater or smaller weights (i.e., by assigning a multiplication factor represented by a coefficient and/or by raising the variable's power from 1 to a different value) to variables that exert great or little-to-no influence, respectively, on the output (SO₃ solubility, in the present case). As shown in Figure 3, melting time is the most influential processing variable; as such, in the analytical models developed herein, this variable has been assigned greater weight to enhance its influence (i.e., the weight is assigned by raising the power of the melting time). Other variables were also treated in the same manner; that is, they were assigned weights which were commensurate with their influence on the output (SO₃ solubility).

The general form of the first analytical model (subsequently referred to as Analytical Model 1)—developed on the basis of the above-discussed concepts/notions—is shown in eq 3. As can be seen, the form of Analytical Model 1, with 25 input variables, is similar to the ones depicted in eqs 1 and 2; the dissimilarities arise primarily from the inclusion of parameters that were ignored in the original models advanced by Vienna et al.¹ and Muller et al.⁹

$$w_{\text{SO}_3}^{\text{Pred}} = \left| \begin{aligned} &\text{selected} \sum_{i=1}^q s_i n_i \\ &+ \text{selected} \sum_{i=1}^q s_{ii} n_i^2 + C_1 T^{C_2} + C_3 P + \text{Constant} \end{aligned} \right| \quad (3)$$

Here $w_{\text{SO}_3}^{\text{Pred}}$ = predicted SO₃ solubility (in mass %); q = number of components in the waste glass (except for SO₃); n_i = normalized mass % of the i th component; s_i = coefficient of the i th component; s_{ii} = coefficient of the i th component squared; T = melting time (in hour); P = SO₃ source (1 = Na₂SO₄; and 2 = other sources); C_j = coefficients associated with input variables P and T ; and Constant = a coefficient representing a

baseline SO₃ solubility value. It should be noted that the units of all coefficients are assigned (empirically) on the basis of the variable associated with them, such that the final output has the units of mass %. For example, s_i is unitless; s_{ii} has the units of 1/(mass %); and C_3 has the units of mass %.

As discussed above, in eq 3, there are several coefficients (e.g., s_i , s_{ii} , and C_j)—and a constant—that ought to be optimized. Toward this, a nonlinear, gradient-descent scheme^{24,48,49}—based on the Nelder–Mead multidimensional simplex algorithm,⁷¹ which is also built into the RF, SVM, and hybrid models for hyperparameter optimization—has been used to determine the optimum values of each coefficient and the constant so as to minimize the deviations between the model's predictions of SO₃ solubility with respect to actual measurements in the 1074 glasses included in the original database (described in Section 2.0). During the optimization of these coefficients/constant, constraints (i.e., minimum or maximum value that is allowed) were not employed. Such a constraint-free optimization scheme has been deliberately employed since it allows automatic suppression, or even removal, of inconsequential variables (e.g., if the coefficient of a given variable acquires an optimum value of 0 or a finite number close to 0, the implication is that the variable is not important). Table 2 shows the optimal values of coefficients and the constant for Analytical Model 1.

The predictions of SO₃ solubility in both 1074 HLW and LAW glasses, as produced by Analytical Model 1, are compared against experimental measurements in Figure 4, while the statistical parameters describing the prediction performance in a quantitative manner are itemized in Table 3. As can be seen, Analytical Model 1 produces reasonably accurate predictions of SO₃ solubility, as evidenced by the value of $R = 0.83$. Compared to the models proposed by Muller et al. ($R = 0.15$)⁹ and Vienna et al. ($R = 0.16$), the predictive ability of Analytical Model 1 is noticeably and substantially superior. Likewise, compared to the hybrid ML model ($R = 0.88$), predictions of Analytical Model 1 are inferior but not by a significant margin. In Figure 4, it is worth noting that the predictions show large deviations from measurements when the SO₃ solubility is <0.2 mass %. This is because the database—used to develop and calibrate Analytical Model 1—includes a small number of glasses with such low SO₃. It is posited that the inclusion of more data records, spanning a wide range of SO₃ solubility, would alleviate this problem.

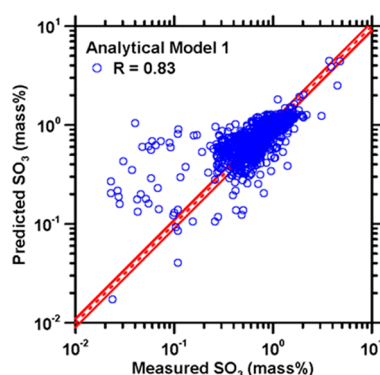


Figure 4. Predictions of SO_3 solubility in 1074 LAW and HLW glasses as produced by Analytical Model 1 (represented by eq 3) compared against experimental measurements. The dashed line represents the line of ideality and the solid lines represent a $\pm 10\%$ bound.

Table 3. Statistical Parameters Used to Quantify Accuracy of Predictions of SO_3 Solubility Produced by Analytical Model 1 and Analytical Model 2

analytical model	R	R^2	MAE, mass %	RMSE, mass %	MAPE, %
1	0.8301	0.6891	0.1771	0.2412	42.99
2	0.7785	0.6061	0.2030	0.2714	45.47

While Analytical Model 1 produces accurate predictions of SO_3 solubility in nuclear waste glasses—especially compared to the two widely used models that, essentially, represent the current state-of-the-art of analytical predictive models^{1,9}—the model requires 67 different inputs from end users. To be clear, the aforesaid 67 inputs include 25 compositional/processing variables pertaining to the glasses, in addition to the 42 coefficients/constant tabulated in Table 2. It is acknowledged that a large number of inputs could be a deterrent, impeding practical use of Analytical Model 1, in spite of its high prediction accuracy, ability to account for both LAW and HLW glasses, and simple/partial quadratic form. The authors believe that simpler analytical models, with fewer inputs, would be more appealing to end-users, even if the simplicity comes at the expense of prediction accuracy. On the basis of this belief, a much simpler model—named Analytical Model 2—was developed. The development of, and parameter optimization for, this model was carried out using the same principles as those used for the development of Analytical Model 1. To specifically reduce the model's complexity and the number of inputs, only the 10 most influential variables represented in Figure 3 were selected, while the remaining 30 variables were excluded. The general form of Analytical Model 2, with 10 input variables, is shown in eq 4. Optimal values of the 14 coefficients (and 1 constant) for the model—as determined from the nonlinear, gradient-descent scheme used in the

previous analytical model—are listed in Table 4. Predictions of SO_3 solubility in 1074 HLW and LAW nuclear waste glasses, as produced by Analytical Model 2, are compared against experimental measurements in Figure 5. Statistical parameters describing prediction errors are itemized in Table 3.

$$w_{\text{SO}_3}^{\text{Pred}} = \left| \text{selected} \sum_{i=1}^q s_i n_i + \text{selected} \sum_{i=1}^q s_{ii} n_i^2 + C_1 T^{C_2} + \text{Constant} \right| \quad (4)$$

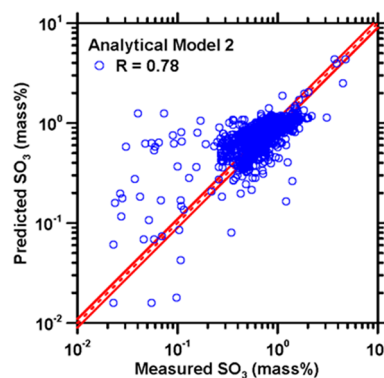


Figure 5. Predictions of SO_3 solubility in 1074 LAW and HLW glasses as produced by Analytical Model 2 (represented by eq 4) compared against experimental measurements. The dashed line represents the line of ideality, and the solid lines represent a $\pm 10\%$ bound.

As can be seen in Figure 5 and Table 3, Analytical Model 2 also produces reasonably accurate predictions of SO_3 solubility ($R = 0.78$). Compared to the Analytical Model 1 ($R = 0.83$), the predictive ability of Analytical Model 2 is—expectedly—inferior; but this loss in prediction performance is envisioned to be outweighed by the gains in model simplicity and substantial reduction in the number of input parameters [N.B., Analytical Model 1 requires 67 inputs, whereas Analytical Model 2 only requires 25]. Furthermore, compared to the models proposed by Muller et al. ($R = 0.15$)⁹ and Vienna et al. ($R = 0.16$)¹, the predictive ability of Analytical Model 2 is still substantially superior.

Finally, as discussed in Section 3.2, we acknowledge that in industrial processes, measuring—or controlling—glass melting time may be difficult or even impossible in some cases. In such cases, the analytical models presented in our study can be used to estimate the glass melting time. Specifically, this can be done by using eq 3 or 4, and plugging-in all compositional/processing variables, coefficients, and constant except for the glass melting time. Solving the equation will result in an estimate of the glass melting time. Suppose this procedure is

Table 4. Optimal Values of 14 Coefficients and 1 Constant (for 10 Input Variables Corresponding to the Glasses' Chemical Descriptors and Processing Parameters) for Analytical Model 2

CaO	0.058659	Na_2O	−0.049841	UO_3	0.072388
SnO_2	−0.0707	V_2O_5	0.020389	P_2O_5	−0.011248
Li_2O	0.068117	$\text{V}_2\text{O}_5 \times \text{V}_2\text{O}_5$	−0.000358	$\text{Na}_2\text{O} \times \text{Na}_2\text{O}$	0.002059
$\text{MnO} \times \text{MnO}$	0.003962	$\text{SnO}_2 \times \text{SnO}_2$	0.001901	$\text{Li}_2\text{O} \times \text{Li}_2\text{O}$	−0.00331
C_1	0.844376	C_2	0.0663	constant	−0.29307

repeated for a number (say ~ 10 or ~ 20) of glasses. In that case, the average of all estimates of glass melting time can then be uniformly used for subsequent predictions and optimizations.

4.0. CONCLUSION

Borosilicate glass—the most used glass system for nuclear waste vitrification—has been studied for decades, with the objective of determining the optimal formulation that increases waste loading capacity without compromising with the melter operation and its long-term performance. Sulfur (in the form of sulfate, SO_4^{2-}) loading capacity—generally defined as the limiting value of SO_3 solubility in the glass—is a decisive factor that dictates cost, performance, safety, and longevity of melter operations. To circumvent the above-mentioned complications—and to ensure safe, long-term operation of the melter—several empirical models have been developed to predict SO_3 solubility in the glasses in relation to the glasses' chemical composition and processing parameters. However, these empirical models were developed and calibrated on the basis of databases of limited volume, comprising $\ll 1000$ distinct Hanford LAW glass compositions. Furthermore, these models do not account for various processing parameters (e.g., time and temperature of glass melting, source of SO_3 , etc.), some of which exert substantial influence on SO_3 solubility. Owing to these *blind spots*, these models are unable to produce reliable predictions on SO_3 solubility in HLW and LAW glasses, whose composition or processing parameters fall outside of the domain encompassed by the databases used to develop/calibrate the empirical model.

This study attempts to address the limitations of the current empirical models by advancing simple yet comprehensive, analytical models—using a comprehensive data set of >1000 LAW and HLW glasses and a wide range of experimental/processing conditions—derived on the basis of three machine learning models (RF, SVM, and hybrid model) that can produce prompt, accurate predictions of sulfate solubility in not just LAW but also HLW glasses. The ML models were able to establish robust correlations between the inputs (glasses' compositional/processing variables) and the output (SO_3 solubility). Further, the ML model with the best prediction performance was used to quantitatively assess and rank the influence (i.e., importance) of glasses' compositional/processing variables on the SO_3 solubility. Finally, on the premise of such understanding of influential and inconsequential variables, two closed-form, partial quadratic, analytical models—a highly parametrized model (Analytical Model 1, comprising 25 input variables and 42 coefficients/constant) and a simple model (Analytical Model 2, comprising only 10 input variables and 15 coefficients/constant)—were developed to predict SO_3 solubility in both LAW and HLW glasses.

Prediction performances of both analytical models were then rigorously evaluated against the entire database, comprising 1074 distinct LAW and HLW glasses. It was shown that Analytical Model 1 produces predictions of SO_3 solubility in glasses with an accuracy ($R = 0.83$) that is comparable to that produced by the best-performing ML model ($R = 0.83$) and appreciably higher than the empirical models ($R \approx 0.16$) that represent the current state-of-the-art. Analytical Model 2—in spite of its simple (partial quadratic) form and inclusion of only 10 input variables—is able to produce predictions with R of 0.78, which is only slightly inferior to the prediction performances of Analytical Model 1 and the best-performing

ML model. The authors believe that Analytical Model 2—which is simple and easy to use and requires only a few inputs—would be more appealing to end users, even if its simplicity comes at the (minimal) expense of prediction accuracy (compared to Analytical Model 1 or ML models).

On a closing note, the authors acknowledge that there is still room for improvement in the prediction performance of the analytical models. Such improvement can be attained by utilizing a database of more extensive volume (e.g., inclusion of data records featuring very low SO_3 solubility in glasses) and even greater diversity. In particular, future work should focus on collecting sulfur solubility data in glass melts using large-scale melters. Notwithstanding, this study represents an essential milestone by demonstrating that by harnessing the power of artificial intelligence and data (past knowledge/experience), in the future it would indeed be possible to develop simple, closed-form analytical models that can reliably predict the behavior of materials. Such analytical models can be leveraged by end users—regardless of their level of expertise in computer science—to design new nuclear waste glasses with enhanced sulfur loadings.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsami.1c10359>.

Overview of machine learning models and summary of the database (PDF)

SO_3 solubility database (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Ashtosh Goel – Department of Materials Science and Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States; orcid.org/0000-0003-0139-9503; Phone: +1-848-445-4512; Email: ag1179@soe.rutgers.edu

Aditya Kumar – Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, Missouri 65409, United States; orcid.org/0000-0001-7550-8034; Phone: +1-573-341-6994; Email: kumarad@mst.edu

Authors

Xinyi Xu – Department of Materials Science and Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States

Taihao Han – Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, Missouri 65409, United States

Jie Huang – Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, Missouri 65409, United States; orcid.org/0000-0002-8659-2910

Albert A. Kruger – Office of River Protection, U.S. Department of Energy, Richland, Washington 99354, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsami.1c10359>

Notes

The authors declare no competing financial interest.

The database and the models presented are available from the corresponding authors (A.G., ag1179@soe.rutgers.edu; A.K., kumarad@mst.edu) upon request. The database—and associated details—are also provided in the [Supporting Information](#).

ACKNOWLEDGMENTS

This work was supported by funding provided by the Department of Energy (DOE), Office of River Protection, Waste Treatment & Immobilization Plant (WTP), through Contract No. 89304018CEM000006; UM system; the Leonard Wood Institute (LWI; Grant No. W911NF-07-2-0062); and the National Science Foundation (NSF-CMMI; Grant Nos. 1661609, 1932690, 2034871, and 2034856).

REFERENCES

- (1) Vienna, J. D.; Kim, D.-S.; Muller, I. S.; Piepel, G. F.; Kruger, A. A. Toward Understanding the Effect of Low-Activity Waste Glass Composition on Sulfur Solubility. *J. Am. Ceram. Soc.* **2014**, *97* (10), 3135–3142.
- (2) Bingham, P. A.; Vaishnav, S.; Forder, S. D.; Scrimshire, A.; Jaganathan, B.; Rohini, J.; Marra, J. C.; Fox, K. M.; Pierce, E. M.; Workman, P.; Vienna, J. D. Modelling the Sulfate Capacity of Simulated Radioactive Waste Borosilicate Glasses. *J. Alloys Compd.* **2017**, *695*, 656–667.
- (3) Manara, D.; Grandjean, A.; Pinet, O.; Dussossoy, J. L.; Neuville, D. R. Sulfur Behavior in Silicate Glasses and Melts: Implications for Sulfate Incorporation in Nuclear Waste Glasses as a Function of Alkali Cation and V_2O_5 Content. *J. Non-Cryst. Solids* **2007**, *353* (1), 12–23.
- (4) Karlina, O. K.; Ozhovan, M. I.; Sobolev, I. A.; Ovchinnikov, A. V.; Varlakova, G. A.; Polyakov, A. S. Solidifying Sulfate-Bearing Wastes to Make Glass Composite Materials. *At Energy* **1992**, *72* (2), 159–162.
- (5) Karlina, O. K.; Ovchinnikov, A. V.; Ozhovan, M. I. Immobilization of Sulfate-Containing Wastes in Borosilicate Glass Using Fluoride Additives. *At. Energy* **1994**, *76* (3), 236–239.
- (6) Skidmore, C. H.; Vienna, J. D.; Jin, T.; Kim, D.; Stanfill, B. A.; Fox, K. M.; Kruger, A. A. Sulfur Solubility in Low Activity Waste Glass and Its Correlation to Melter Tolerance. *Int. J. Appl. Glass Sci.* **2019**, *10* (4), 558–568.
- (7) Vienna, J.; Piepel, G.; Kim, D.-S.; Crum, J.; Lonergan, C.; Stanfill, B.; Riley, B.; Cooley, S.; Jin, T. *2016 Update of Hanford Glass Property Models and Constraints for Use in Estimating the Glass Mass to Be Produced at Hanford by Implementing Current Enhanced Glass Formulation Efforts*, PNNL-25835; Pacific Northwest National Laboratory: Richland, WA, USA, 2016.
- (8) Sobolev, I. A.; Dmitriev, S. A.; Lifanov, F. A.; Kobelev, A. P.; Stefanovsky, S. V.; Ojovan, M. I. Vitrification Processes for Low, Intermediate Radioactive and Mixed Wastes. *Glass Technol.* **2005**, *46* (1), 28–35.
- (9) Muller, I. S.; Gilbo, K.; Chaudhuri, M.; Pegg, I. L.; Joseph, I. K-3 *Refractory Corrosion and Sulfate Solubility Model Enhancement (Final Report)*, Technical Report ORP-63490_Rev_0; Hanford Site: Richland, WA, USA, 2018. DOI: [10.2172/1513833](https://doi.org/10.2172/1513833).
- (10) Du, L.-S.; Stebbins, J. F. Solid-State NMR Study of Metastable Immiscibility in Alkali Borosilicate Glasses. *J. Non-Cryst. Solids* **2003**, *315* (3), 239–255.
- (11) McKeown, D. A.; Muller, I. S.; Gan, H.; Pegg, I. L.; Stolte, W. C. Determination of Sulfur Environments in Borosilicate Waste Glasses Using X-Ray Absorption near-Edge Spectroscopy. *J. Non-Cryst. Solids* **2004**, *333* (1), 74–84.
- (12) Mishra, R. K.; Sudarsan, K. V.; Sengupta, P.; Vatsa, R. K.; Tyagi, A. K.; Kaushik, C. P.; Das, D.; Raj, K. Role of Sulfate in Structural Modifications of Sodium Barium Borosilicate Glasses Developed for Nuclear Waste Immobilization. *J. Am. Ceram. Soc.* **2008**, *91* (12), 3903–3907.
- (13) Vaishnav, S.; Hannon, A. C.; Barney, E. R.; Bingham, P. A. Neutron Diffraction and Raman Studies of the Incorporation of Sulfate in Silicate Glasses. *J. Phys. Chem. C* **2020**, *124* (9), 5409–5424.
- (14) Langowski, M. H. *The Incorporation of P, S, Cr, F, Cl, I, Mn, Ti, U, and Bi into Simulated Nuclear Waste Glasses: Literature Study*, Technical Report PNNL-10980; Pacific Northwest National Laboratory: Richland, WA, USA, 1996. DOI: [10.2172/208366](https://doi.org/10.2172/208366).
- (15) Goel, A.; McCloy, J. S.; Pokorny, R.; Kruger, A. A. Challenges with Vitrification of Hanford High-Level Waste (HLW) to Borosilicate Glass - An Overview. *J. Non-Cryst. Solids: X* **2019**, *4*, 100033.
- (16) Muller, I. S.; Gan, H.; Gilbo, K.; Pegg, I. L.; Joseph, I. *LAW Glass Property - Composition Models for K-3 Refractory Corrosion and Sulfate Solubility (Final Report)*, VSL-15R3270-1; UD Department of Energy, Office of River Protection, Richland, WA, 2015.
- (17) Jin, T.; Kim, D.; Darnell, L. P.; Weese, B. L.; Canfield, N. L.; Bliss, M.; Schweiger, M. J.; Vienna, J. D.; Kruger, A. A. Crucible Salt Saturation Method for Determining Sulfur Solubility in Glass Melt. *Int. J. Appl. Glass Sci.* **2019**, *10* (1), 92–102.
- (18) Ravinder; Venugopal, V.; Bishnoi, S.; Singh, S.; Zaki, M.; Grover, H. S.; Bauchy, M.; Agarwal, M.; Krishnan, N. M. A. Artificial Intelligence and Machine Learning in Glass Science and Technology: 21 Challenges for the 21st Century. *Int. J. Appl. Glass Sci.* **2021**, *12* (3), 277–292.
- (19) Bishnoi, S.; Singh, S.; Ravinder, R.; Bauchy, M.; Gosvami, N. N.; Kodamana, H.; Krishnan, N. M. A. Predicting Young's Modulus of Oxide Glasses with Sparse Datasets Using Machine Learning. *J. Non-Cryst. Solids* **2019**, *524*, 119643.
- (20) Onbaşı, M. C.; Tandia, A.; Mauro, J. C. Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass. In *Handbook of Materials Modeling: Applications: Current and Emerging Materials*; Andreoni, W., Yip, S., Eds.; Springer International: Cham, Switzerland, 2020; pp 1997–2019. DOI: [10.1007/978-3-319-44680-6_100](https://doi.org/10.1007/978-3-319-44680-6_100).
- (21) Mauro, J. C.; Tandia, A.; Vargheese, K. D.; Mauro, Y. Z.; Smedskjaer, M. M. Accelerating the Design of Functional Glasses through Modeling. *Chem. Mater.* **2016**, *28* (12), 4267–4277.
- (22) Brauer, D. S.; Rüssel, C.; Kraft, J. Solubility of Glasses in the System P_2O_5 -CaO-MgO-Na $_2$ O-TiO $_2$: Experimental and Modeling Using Artificial Neural Networks. *J. Non-Cryst. Solids* **2007**, *353* (3), 263–270.
- (23) Anoop Krishnan, N. M.; Mangalathu, S.; Smedskjaer, M. M.; Tandia, A.; Burton, H.; Bauchy, M. Predicting the Dissolution Kinetics of Silicate Glasses Using Machine Learning. *J. Non-Cryst. Solids* **2018**, *487*, 37–45.
- (24) Han, T.; Stone-Weiss, N.; Huang, J.; Goel, A.; Kumar, A. Machine Learning as a Tool to Design Glasses with Controlled Dissolution for Application in Healthcare Industry. *Acta Biomater.* **2020**, *107* (1), 286–298.
- (25) Aloy, A.; Soshnikov, R.; Trofimenko, A. V.; Vienna, J. D.; Elliott, M. L.; Holtzscheiter, E. W. Improved Loading of Sulfate—Limited Waste in Glass. *Waste Management 2006 Symposium*, INIS-US-09-WM-06331, Tucson, AZ, USA, Feb. 26–Mar. 2, 2006; WM Symposia: Tucson, AZ, USA, 2006.
- (26) Vienna, J. D.; Hrma, P. R.; Buchmiller, W. C.; Ricklefs, J. S. *Preliminary Investigation of Sulfur Loading in Hanford LAW Glass*, Technical Report PNNL-14649; Pacific Northwest National Laboratory: Redland, WA, USA, 2004.
- (27) Lenoir, M.; Grandjean, A.; Poissonnet, S.; Neuville, D. R. Quantitation of Sulfate Solubility in Borosilicate Glasses Using Raman Spectroscopy. *J. Non-Cryst. Solids* **2009**, *355* (28), 1468–1473.
- (28) Peeler, D. K. *An Assessment of the Sulfate Solubility Limit for the FRIT 418—Sludge Batch 2/3 System*, Technical Report WSR-TR-2004-00081; Savannah River Site: Aiken, SC, USA, 2004. DOI: [10.2172/824040](https://doi.org/10.2172/824040).
- (29) Bingham, P. A.; Hand, R. J. Sulphate Incorporation and Glass Formation in Phosphate Systems for Nuclear and Toxic Waste Immobilization. *Mater. Res. Bull.* **2008**, *43* (7), 1679–1693.
- (30) Muller, I. S.; Pegg, I. L.; Gan, H.; Buechele, A.; Rielley, E.; Bazemore, G.; Cecil, R.; Hight, K.; Mooers, C.; Lai, S.-T. T.; Kruger, A.

- A. A. Final Report. Baseline LAW Glass Formulation Testing, VSL-03R3460-1, Rev. 0, Technical Report ORP-59016 Rev. 0; Hanford Site: Richland, WA, USA, 2015. DOI: 10.2172/1186012.
- (31) Jantzen, C. M.; Smith, M. E.; Peeler, D. K. Dependency of Sulfate Solubility on Melt Composition and Melt Polymerization. *Environmental Issues and Waste Management Technologies in the Ceramic and Nuclear Industries X*, Vol. 168; John Wiley & Sons, 2012; pp 141–152. DOI: 10.1002/9781118408438.ch13.
- (32) Kruger, A. A.; Pegg, I. L.; Matlack, K. S.; Joseph, I.; Muller, I. S.; Gong, W. Final Report—LAW Envelope C Glass Formulation Testing to Increase Waste Loading, VSL-05R5900-1, Technical Report ORP-56323 Rev 0; Hanford Site: Richland, WA, USA, 2013; DOI: 10.2172/1109496.
- (33) Kruger, A. A.; Matlack, K. S. Glass Formulation Testing to Increase Sulfate Incorporation—Final Report VSL-04R4960-1, Technical Report ORP-51808 Rev 0; Hanford Site: Richland, WA, USA, 2012. DOI: 10.2172/1035193.
- (34) Muller, I. S.; Joseph, I.; Pegg, I. L. Comparison of LAW Simulant, Actual Waste, and Melter Glasses (Final Report, Rev. 0), Technical Report ORP-63498-Rev.0; Hanford Site: Richland, WA USA, 2005. DOI: 10.2172/1515837.
- (35) Kruger, A. A.; Muller, I.; Gilbo, K.; Joseph, I.; Pegg, I. L. Final Report—Enhanced LAW Glass Property—Composition Models—Phase 1 VSL-13R2940-1, Technical Report ORP-56286 Rev 0; Hanford Site: Richland, WA USA, 2013. DOI: 10.2172/1105963.
- (36) Muller, I. S.; Joseph, I.; Pegg, I. L. Preparation and Testing of LAW High Phosphorus and High-Chromium Glasses (Final Report, Rev. 0), Technical Report ORP-63509 Rev. 0; Hanford Site: Richland, WA USA, 2006. DOI: 10.2172/1528775.
- (37) Kruger, A. A.; Pegg, I. L.; Matlack, K. S.; Joseph, I.; Muller, I. S.; Gong, W. Final Report—Enhanced LAW Glass Formulation Testing, VSL-07R1130-1, Rev. 0, Dated 10/05/07, Technical Report ORP-56293 Rev 0; Hanford Site: Richland, WA USA, 2013. DOI: 10.2172/1105974.
- (38) Kruger, A. A.; Pegg, I. L.; Feng, A.; Gan, H.; Kot, W. K. Final Report—Sulfate Solubility in RPP-WTP HLW Glasses, VSL-06R6780-1, Rev. 0, Technical Report ORP-56325 Rev 0; Hanford Site: Richland, WA USA, 2013. DOI: 10.2172/1109491.
- (39) Kruger, A. A.; Pegg, I. L.; Kot, W.; Gan, H.; Matlack, K. S. Enhanced Sulfate Management in HLW Glass Formulations VSL12R2540-1 REV 0, Technical Report ORP-53932-REV-0; Hanford Site: Richland, WA USA, 2012. DOI: 10.2172/1059479.
- (40) Billings, A. L.; Fox, K. M. Retention of Sulfate in Savannah River Site High-Level Radioactive Waste Glass. *Int. J. Appl. Glass Sci.* **2010**, *1* (4), 388–400.
- (41) Brendebach, B.; Denecke, M. A.; Roth, G.; Weisenburger, S. Sulfur Incorporation in High Level Nuclear Waste Glass: A S K-Edge XAFS Investigation. *J. Phys.: Conf. Ser.* **2009**, *190*, 012186.
- (42) Rielley, E.; Muller, I. S.; Pegg, I. L.; Gan, H.; Cecil, R.; Bazemore, G.; Moncrief, T.; Derr, K. D.; Watts, S.; Mooers, C.; Grant, S.; Feng, C.-F.; Lai, S. T.; Boice, C.; D'Angelo, N.; Su, L.; Liu, Y.; Cheron, J.; Kashnikov, E.; Goloski, L.; Singh, J. Final Report: Preparation and Testing of LAW Matrix Glasses, VSL-04R4480-1, Rev. 0, Technical Report ORP-63497, Rev. 0; Hanford Site: Richland, WA USA, 2004. DOI: 10.2172/1512798.
- (43) Xu, X.; Youngman, R. E.; Kapoor, S.; Goel, A. Structural Drivers Controlling Sulfur Solubility in Alkali Aluminoborosilicate Glasses. *J. Am. Ceram. Soc.* **2021**, *104* (10), 5030–5049.
- (44) Hrma, P. R. Retention of Halogens in Waste Glass, Technical Report PNNL-19361; Pacific Northwest National Laboratory: Redland, WA, USA, 2010.
- (45) McCloy, J. S.; Riley, B. J.; Goel, A.; Liezers, M.; Schweiger, M. J.; Rodriguez, C. P.; Hrma, P.; Kim, D.-S.; Lukens, W. W.; Kruger, A. A. Rhenium Solubility in Borosilicate Nuclear Waste Glass: Implications for the Processing and Immobilization of Technetium-99. *Environ. Sci. Technol.* **2012**, *46* (22), 12616–12622.
- (46) Brehault, A.; Patil, D.; Kamat, H.; Youngman, R. E.; Thirion, L. M.; Mauro, J. C.; Corkhill, C. L.; McCloy, J. S.; Goel, A. Compositional Dependence of Solubility/Retention of Molybdenum Oxides in Aluminoborosilicate-Based Model Nuclear Waste Glasses. *J. Phys. Chem. B* **2018**, *122* (5), 1714–1729.
- (47) Cook, R.; Lapeyre, J.; Ma, H.; Kumar, A. Prediction of Compressive Strength of Concrete: A Critical Comparison of Performance of a Hybrid Machine Learning Model with Standalone Models. *J. Mater. Civ. Eng.* **2019**, *31* (11), 04019255.
- (48) Lapeyre, J.; Han, T.; Wiles, B.; Ma, H.; Huang, J.; Sant, G.; Kumar, A. Machine Learning Enables Prompt Prediction of Hydration Kinetics of Multicomponent Cementitious Systems. *Sci. Rep.* **2021**, *11* (3), 3922.
- (49) Gomaa, E.; Han, T.; ElGawady, M.; Huang, J.; Kumar, A. Machine Learning to Predict Properties of Fresh and Hardened Alkali-Activated Concrete. *Cem. Concr. Compos.* **2021**, *115* (1), 103863.
- (50) Cook, R.; Han, T.; Childers, A.; Ryckman, C.; Khayat, K.; Ma, H.; Huang, J.; Kumar, A. Machine Learning for High-Fidelity Prediction of Cement Hydration Kinetics in Blended Systems. *Mater. Des.* **2021**, *208*, 109920.
- (51) Schaffer, C. Selecting a Classification Method by Cross-Validation. *Mach. Learn.* **1993**, *13* (1), 135–143.
- (52) Cook, R.; Monyake, K. C.; Hayat, M. B.; Kumar, A.; Alagha, L. Prediction of Flotation Performance of Metal Sulfides Using an Original Hybrid Machine Learning Model. *Eng. Rep.* **2020**, *2* (6), No. e12167.
- (53) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32.
- (54) Chen, X.; Ishwaran, H. Random Forests for Genomic Data Analysis. *Genomics* **2012**, *99* (6), 323–329.
- (55) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947–1958.
- (56) Smola, A. J.; Schölkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14* (3), 199–222.
- (57) Chopra, P.; Sharma, R. K.; Kumar, M. Prediction of Compressive Strength of Concrete Using Artificial Neural Network and Genetic Programming. *Adv. Mater. Sci. Eng.* **2016**, *2016*, 7648467.
- (58) Chou, J.-S.; Tsai, C.-F.; Pham, A.-D.; Lu, Y.-H. Machine Learning in Concrete Strength Simulations: Multi-Nation Data Analytics. *Constr. Build. Mater.* **2014**, *73*, 771–780.
- (59) Menze, B. H.; Kelm, B. M.; Masuch, R.; Himmelreich, U.; Bachert, P.; Petrich, W.; Hamprecht, F. A. A Comparison of Random Forest and Its Gini Importance with Standard Chemometric Methods for the Feature Selection and Classification of Spectral Data. *BMC Bioinf.* **2009**, *10* (1), 213.
- (60) Katuwal, R.; Suganthan, P. N.; Zhang, L. Heterogeneous Oblique Random Forest. *Pattern Recognit.* **2020**, *99*, 107078.
- (61) Lenoir, M.; Grandjean, A.; Neuville, D. R. A Kinetic Approach of Sulphur Behaviour in Borosilicate Glasses and Melts: Implications for Sulphate Incorporation in Nuclear Waste Glasses. *Atalante: Scientific Report 2008. Nuclear Fuel Cycle for a Sustainable Future*, Technical Report CEA-R-6181; CEA Marcoule: Montpellier, France, 2008.
- (62) Lenoir, M.; Neuville, D. R.; Malki, M.; Grandjean, A. Volatilization Kinetics of Sulphur from Borosilicate Melts: A Correlation between Sulphur Diffusion and Melt Viscosity. *J. Non-Cryst. Solids* **2010**, *356* (50), 2722–2727.
- (63) Kaushik, C. P.; Mishra, R. K.; Sengupta, P.; Kumar, A.; Das, D.; Kale, G. B.; Raj, K. Barium Borosilicate Glass - a Potential Matrix for Immobilization of Sulfate Bearing High-Level Radioactive Liquid Waste. *J. Nucl. Mater.* **2006**, *358* (2), 129–138.
- (64) Schreiber, H. D.; Kozak, S. J.; Leonhard, P. G.; McManus, K. K. Sulfur Chemistry in a Borosilicate Melt. Pt. 1. *Glastech. Ber.* **1987**, *60* (12), 389–398.
- (65) Vienna, J. D.; Buchmiller, W. C.; Crum, J. V.; Graham, D. D.; Kim, D.-S.; Macisaac, B. D.; Schweiger, M. J.; Peeler, D. K.; Edwards, T. B.; Reamer, I. A.; Workman, R. J. Glass Formulation Development for INEEL Sodium-Bearing Waste, Technical Report PNNL-14050; Pacific Northwest National Laboratory: Redland, WA, USA, 2002. DOI: 10.2172/1123261.

(66) McKeown, D. A.; Muller, I. S.; Matlack, K. S.; Pegg, I. L. X-Ray Absorption Studies of Vanadium Valence and Local Environment in Borosilicate Waste Glasses Using Vanadium Sulfide, Silicate, and Oxide Standards. *J. Non-Cryst. Solids* **2002**, 298 (2), 160–175.

(67) Stefanovskij, S. V.; Lifanov, F. A. Glasses for immobilization of sulphate-containing radioactive waste. *Radiokhimiya* **1989**, 31 (6), 129–134.

(68) Li, H.; Darab, J. G.; Matson, D. W.; Smith, P. A.; Hrma, P.; Chen, Y.; Liu, J. Phosphate-Sulfate Interaction in Simulated Low-Level Radioactive Waste Glasses. *MRS Online Proceedings Library (OPL)* **1995**, 412, 412.

(69) Plodinec, M. J.; Wiley, J. R. *Evaluation of Glass as a Matrix for Solidifying Savannah River Plant Waste: Properties of Glasses Containing Li_2O* , Technical Report DP-1498; Savannah River Laboratory, E. I. Du Pont de Nemours: Aiken, SC, USA, 1979. .

(70) Jin, T.; Riversa, E. L.; Kim, D. S.; Droubaya, T. C.; Jahrman, E. P.; Mara, D.; Weese, B. L.; Seidler, G. T.; Kruger, A. A. Effect of Chlorine and Chromium on Sulfur Solubility in Low Activity Waste Glasses. 2021, submitted for publication.

(71) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, 7 (4), 308–313.