Title: Predicting Radioactive Waste Glass Dissolution with Machine Learning.

Authors: Joseph N. P. Lillington[1,*], Thomas L. Goût[1], Mike T. Harrison[2], Ian Farnan[1].

[1]Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, UK.

[2]National Nuclear Laboratory, Central Laboratory, Sellafield, Seascale, Cumbria, CA20 1PG, UK.

*Corresponding author:

Email address: jnpl2@cam.ac.uk (J. Lillington).

Declarations of interest: none

Keywords

Machine learning; Leaching; Nuclear waste glass; Dissolution

Abstract

The vitrification of high-level nuclear waste within borosilicate glass and its disposition within a multi-barrier repository deep underground is accepted as the best form of disposal. Here, the ability of machine learning to predict both static and dynamic glass leaching behavior is analysed using large-scale unstructured multi-source data, covering a diverse range of experimental conditions and glass compositions. Machine learning can accurately predict leaching behavior, predict missing data, and time forecast. Accuracy depends upon the type of learning algorithm, model input variables, and diversity or size of the underlying dataset. For static leaching, the bagged random forest method predicts well, even when either pH or glass composition are neglected as input variables, additionally showing potential in predicting independent glass dissolution data. For dynamic leaching, accuracy improves if replacing final pH with a species dissolution rate as an input variable, although results show no preferred output species (Si, Na, or Al).

1. Introduction

Historically in the UK, high-level radioactive waste (HLW) from the reprocessing of spent nuclear fuel is vitrified into a borosilicate glass matrix [1]; a well-established method of waste-form immobilisation [2,3]. The glass is solidified within stainless steel containers and, as of 2016, there was 870 m^3 of vitrified HLW contained within 5,780 containers at Sellafield [4]. Current government policy is to store this glass within a multi-barrier geological disposal facility (GDF) deep underground [4]. For the safety case, this will require confidence that the initially contained radionuclides will not be released in any significant quantity into the environment. This represents a major challenge, given that glass dissolution is known to depend on many different factors, including temperature, pH, groundwater flow-rate, and both glass and groundwater compositions [5–8]. This issue is also an international one as many of the major nuclear waste generating countries have chosen vitrification as part of their radioactive waste strategy [9].

Such complexity ensures that robust techniques are needed to predict glass-leaching behaviour as a function of time, which is particularly difficult given the expected million-year design life of a GDF [10]. In the literature, these methods have primarily been mechanistic models, and arguably, the French *glass reactivity with allowance for the alteration layer* (GRAAL) model [11] is the current state of the art. Whilst it is widely accepted that glass dissolution evolves following distinct initial dissolution, rate-drop, residual-rate, and potentially rate resumption regimes [12], there are two competing theories of diffusion controlled corrosion versus interfacial dissolution and reprecipitation [13]; inevitably, differences do exist across computational models of these processes [14–17]. Therefore, it remains a challenge to have one model that can predict experimental leaching dissolution behaviour robustly, under a variety of different experimental conditions, for a range of different glass compositions.

As an alternative, predictive machine learning (ML) methods are potentially of value, particularly given that they reduce the need to make assumptions about underlying glass leaching mechanisms and that they could utilise the considerable amount of data that has been both collected and published in the field over the previous decades. Such techniques are becoming transformative across healthcare, manufacturing, consumer goods, financial services, the media, as well as other industries [18–23]. Nonetheless, their application to nuclear waste glass dissolution has been extremely limited. Krishnan *et al.* [24] demonstrated their value, accurately predicting logarithmic silicon initial dissolution rates from eight different aluminosilicate glasses. In addition, Jantzen *et al*. [25] applied an informatics approach to the ALTGLASS database, analysing the correlation between gel compositions and zeolite generation. Nonetheless, further research is required to examine predictive leaching performance of machine learning on both large-scale static and alternative dynamic datasets.

To further analyse the capability of machine learning to predict glass leaching behaviour, this study first explores the ability to predict dissolution behavior using large-scale static leaching glass dissolution data. This includes: comparing leaching predictive performance across 14 different learning methods, examining the effect of different experimental features on prediction, exploring the ability of machine learning to predict given missing experimental data, discerning the effect of dataset size on leaching prediction, and understanding the performance of trained networks on both group-independent data and in time-forecasting. Additionally, machine learning techniques are also further applied to predict glass initial dissolution rates using various dynamic-flow glass leaching data, building upon the work of Krishnan *et al.* [24] .

This study is novel in a number of respects. To our knowledge, this is the first study to explore the effectiveness of machine learning prediction in static leaching and for nuclear waste glass dissolution in general. It makes use of primarily unstructured data, taken from across the

literature, internal to the University of Cambridge, and from multiple industrial vitrification campaigns. Furthermore, in a substantial expansion of the work of Krishnan *et al.* [24], dynamic leaching prediction considers both 'geological' and nuclear waste glass dissolution data, uses additional machine learning methods, and considers the effect of varying dataset size. In addition, rate prediction is not solely limited to three component (sodium aluminosilicate) glasses but to complex multi-component glasses, the effect of different experimental features on prediction is considered, and prediction is not solely limited to silicon release, but includes the release of species with more varied solubility such as sodium and aluminium.

2. Methods

This paper separates glass dissolution prediction into two categories. Machine learning is firstly applied to static glass leaching data and then subsequently to dynamic flow data. The different machine learning methods are stated in Section 2.1 with the underlying experimental data being outlined in Section 2.2. The specific simulations performed are then detailed in Section 2.3. All code has been implemented using MATLAB [26] and is available upon request.

2.1. Machine Learning Methods

Machine learning [27,28] aims to predict one (or multiple) output variables as a function of different input variables. Each method learns a correlation using a training dataset, prior to determining its predictive ability using an independent test dataset. Neural networks use an additional validation dataset as part of training whilst tuning hyperparameters. The 14 supervised machine-learning methods considered in this study to predict glass leaching behaviour are: neural networks, multiple, lasso, ridge, and elastic-net regression, support vector machines (SVM), Gaussian Process Regression (GPR), individual regression trees, boosted ensembles, and bagged random forests. SVM regression used either Gaussian, linear, or

polynomial kernel functions. GPR used either MATLAB ‘exponential’, ‘squaredexponential’, and ‘ardsquaredexponential’ kernel functions. The techniques are more extensively described in Table S1, and several of the methods have also been discussed by Krishnan *et al.* [24].

2.2. Experimental Data

Dataset A

Dataset A consists of 53 static leaching experiments on simulant UK Magnox radioactive waste glasses obtained at 90.0 ± 0.2^{o}C. Tests primarily used an initial surface-area-to-volume ratio (SA/V) of 2000 (range 1726-2131) m^{-1}, a 75-150 µm powder of mass 4.00 (range 3.01-4.06) g, and initial deionised water leachant volume of 40.0 (30.1-40.2) mL. Leaching used perfluoroalkoxy alkane (PFA, Savillex) 60 mL “standard jars”. The dataset was provided courtesy of the Nuclear Decommissioning Authority (NDA), taken over many vitrification campaigns. Glass composition and density have consequently varied significantly between experiments. Experiments used variable Magnox waste loadings, different ratios of Magnox to Thermal Oxide Reprocessing Plant (THORP) waste blends, and newer Ca/Zn base glass frits with and without different loadings of standard HLW and Molybdenum-rich post operational clean out (POCO) waste. All experiments were run for varying leaching time periods for a minimum of 100 days.

Dataset B

Dataset B consists of 18 static leaching experiments, obtained internally at the University of Cambridge. A deionised water leachant was used with a method which followed the ASTM product consistency test (PCT) [29]. The dataset contains: two different international simple glass (ISG) [30] compositions in which lithium had been substituted for sodium as two different Li:Na ratios were each leached at 40 and 90^{o}C [31], a complex simulant Magnox waste glass of 25 wt.% waste loading (MW25, (Mixture Windscale)) glass at 40, 60, 70, 80, and 90^{o}C for

up to 28 days, and two simple lithium-sodium borosilicate base glass frits employed in the UK vitrification process leached at 40 and 90 ^{o}C. A SA/V of 2000 m^{-1} was used for all experiments.

Dataset C

Dataset C contains nine variable composition sodium borosilicate experiments leached at 90 ^{o}C using deionised water leachant. These results were previously published by Gin *et al*. [32].

Dataset D

Dataset D represents 12 long-term French complex simulant waste glass (SON68) experiments, taken under both static and dynamic conditions. These results were previously published by Frugier *et al.* [11].

Dataset E

Dataset E was data obtained using single-pass-flow-through (SPFT) experiments, extracted from the work of Vienna *et al*. [33] where boron initial dissolution rates were obtained for 19 different complex glasses established across many different countries, each repeated at different temperature and pH values.

Dataset F

Dataset F contains nine SPFT experiments leached at pH 3 and 9, reported by Guo *et al*. [34] Simplified glass compositions with Si, B, and (Na) at molar ratios similar to UK glass were leached using deionised water leachant at 90^{o}C.

Dataset G

Dataset G represents two MW25 initial dissolution rates, obtained at 40 and 90^{o}C by Iwalewa *et al*. [6] using SPFT techniques.

Dataset H

Dataset H gives initial dissolution rates computed by Ferrand *et al*. [35] using both SON68 and German designed, alkali-borosilicate (PAMELA) glasses, leached under alkaline conditions.

Dataset I

Dataset I provides initial dissolution rates found by Elia *et al*. [36] using ISG glass, leached under alkaline conditions.

Dataset J

Dataset J gives initial dissolution rates determined by Backhouse *et al*. [37] using ISG glass, under both acidic and hyper-alkaline (up to pH 11) conditions.

Dataset K

Dataset K is 299 initial dissolution rates obtained for nine sodium aluminosilicate glasses by Hamilton *et al*. [38], previously used in the Krishnan *et al*. machine learning study [24].

2.3. Description of Simulations

2.3.1. Static Leaching Simulations

Using the static leaching data (Datasets A-D), this study firstly analyses the ability of the different machine learning methods listed in Section 2.1 to predict normalised boron (B) release (gm^{-2}) as a function of different experimental input variables (see Table 1). This allows for both a comparison across different machine learning methods, as well as an analysis of the effect of different algorithms on predictive performance. Boron is considered because of its generally high release during leaching and inability to form secondary precipitates [39], as such it acts as a proxy for overall glass alteration. Concentrations measured in the static experiments were normalised to the mass fraction of the element within the pristine glass and SA/V after mass loss (of leachate) and blank corrections, known as the *normalised release,* NL_i of element, i. With the exception of neural networks, each simulation has been performed by dividing either

the complete static data (Datasets A-D) or only Dataset A into training and test datasets using ratios of 0.7 and 0.3 respectively. For neural networks, ratios of 0.55, 0.15, and 0.3 have been used for the training, validation, and test datasets respectively. These 'whole experiment' simulations partition the data on a whole experiment basis, rather than partitioning specific time measurements within each individual experiment.

Table 1: *The different input variable combinations used in this study's static simulations. Note that 'All variables' represent the combined experimental variables: elemental mass fractions, glass density, average powder mass (in time), average leachant volume (in time), average surface area to volume ratio (in time), time, pH (in time), elemental normalised release of Cr, Li, Mg, Mo, Na, and Si, in time.*

Simulation Number	**Input Variable Combination**	**Simulation Number**	**Input Variable Combination**
1	*All variables*	10	*All variables*, excluding Li elemental release
2	*All variables*, excluding pH	11	*All variables*, excluding Mg elemental release
3	*All variables*, excluding surface area to volume ratio	12	*All variables*, excluding time
4	*All variables*, excluding powder mass	13	*All variables*, excluding Cr, Li, Mg, and Mo elemental release
5	*All variables*, excluding leachant volume	14	*All variables*, excluding elemental mass fractions

6	*All variables*, excluding glass density	15	*All variables*, excluding all species elemental release
7	*All variables*, excluding pH and all species elemental release	16	*All variables*, excluding Cr, Li, Mg, Mo, and Na elemental release
8	*All variables*, excluding Si elemental release	17	*All variables*, excluding Cr, Li, Mg, Mo, and Na elemental release, adding flow rate to surface area
9	*All variables*, excluding Na elemental release	-	-

Other related static leaching simulations have also been performed. These include examining the ability of the different machine learning algorithms to predict normalised B releases when there is missing experimental data. Here, different individual time-point results have been randomly removed from each experiment, using the same training/validation/test ratios and input combinations stated previously. Both these 'missing data' simulations and the 'whole experiment' simulations described above have also been implemented by using different starting fraction ratios of the full data. Ratios of 0.2, 0.4, 0.6, and 0.8 have been applied, referring to the fraction of data initially removed prior to training/test set partition. This allows for the effect of dataset size to be determined. In addition, the ability of each Dataset A trained model to predict independent group data (Datasets B-D) has been examined. The ability of each model to time-forecast under each Table 1 input variable combination was also assessed. This was achieved by using the initial half of each experimental duration in either Dataset A or Datasets A-D to predict the behaviour during the second half of the leaching duration. Again,

the effect of dataset size was analysed in time-forecasting, by using the same fraction values (0.2, 0.4, 0.6, and 0.8) as stated above.

2.3.2. Dynamic Leaching Simulations

Using the dynamic leaching data (Datasets E-J), this study has aimed to predict initial (log/non-log) B glass dissolution rates as a function of temperature, pH, with and without mole percentage of oxides/halogens (Table 2). Note that due to the low solubility of halogens in glass, these form a very minor contribution to overall glass composition. Again, the performance of different algorithms was compared, as was the effect of dataset size, using the same fraction ratios (0.2, 0.4, 0.6, and 0.8) as in the static leaching simulations stated above. Training was implemented using either Dataset E or E-I. Trained models solely developed using Dataset E were subsequently applied on the remaining datasets F-J to analyse their ability to independently predict rates. Finally, following the approach of Krishnan *et al*. [24], the performance of different machine learning algorithms (considering variable dataset size) was assessed using Dataset K. This was to build upon the original work, going beyond predicting Si release, in order to: determine the relative accuracy of Si, Na, and Al initial dissolution rate prediction; determine the effect of the other input variables; analyse the influence of dataset size; and consider alternative learning algorithms (including SVM kernel variability, ridge regression, GPR (additionally considering kernel variability), and boosting). Nine different input-output variable combinations were trialled, as shown in Table 2.

231 ***Table 2:*** *The different input-output variable combinations trialled using Datasets E(-J) and K.*

Simulation Number Datasets E(-J)	**Input Variable Combination**	**Output Variable Combination**	**Simulation Number Datasets E(-J)**	**Input Variable Combination**	**Output Variable Combination**
1	Temperature, pH, mole percentage of oxides/halogens	B log-initial dissolution rate	3	Temperature, pH	B log-initial dissolution rate
2	Temperature, pH, mole percentage of oxides/halogens	B initial dissolution rate	4	Temperature, pH	B initial dissolution rate
Simulation Number Datasets K	**Input Variable Combination**	**Output Variable Combination**	**Simulation Number Datasets K**	**Input Variable Combination**	**Output Variable Combination**
1	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, final pH	Si log-initial dissolution rate	6	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, Si log-initial dissolution rate	Na log-initial dissolution rate
2	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages,	Na log-initial dissolution rate	7	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, Al	Na log-initial dissolution rate

	initial pH, final pH			log-initial dissolution rate	
3	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, final pH	Al log-initial dissolution rate	8	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, Si log-initial dissolution rate	Al log-initial dissolution rate
4	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, Na log-initial dissolution rate	Si log-initial dissolution rate	9	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, Na log-initial dissolution rate	Al log-initial dissolution rate
5	SiO_2, Na_2O_3, Al_2O_3 mass oxide percentages, initial pH, Al log-initial dissolution rate	Si log-initial dissolution rate			

For all static and dynamic leaching simulations, the performance of different algorithms was judged by computing R^2 and mean square errors (MSE) across both training and test datasets. For neural networks, validation set errors were also considered. Other fine-tuned parameters included: regularisation parameters (lasso, ridge, elastic net regression), leaf size (individual

regression trees), number of trees (boosting/bagging), and neural network hidden layer sizes. These are important for optimising the learnt algorithm performances. Due to the random nature of dataset partition, averages on both R^2 and MSE were performed across 100 iterations. In optimisation, regularisation parameters up to 0.01, leaf size/number of trees up to 150, and neuron numbers up to 52 were considered. Single hidden layer neural networks were considered with feed forward networks, Levenberg-Marquardt optimisation, with a maximum of 1000 epochs in the training. For bagged random forests, the minimum number of observations per leaf was 5. For boosted ensembles, the 'LSBoost' algorithm was used.

3. Results

For improved presentation, machine learning algorithms are given the following labels: multiple linear (1), SVM with Gaussian kernel (2), SVM with linear kernel (3), SVM with polynomial kernel (4), GPR with exponential kernel (5), GPR with square exponential kernel (6), GPR with 'ardsquaredexponential' kernel (7), lasso (8), ridge (9), elastic net (10), single regression tree (11), bagged random forest (12), boosted ensemble (13), and neural network (14). Input/Output (I/O) combinations are numbered consistently with the values given in Tables 1 and 2. Additional results are provided in Tables S1-9 and Figure S1.

3.1. Static Leaching Results

3.1.1. 'Whole Experiment' Simulations

Table 3 states the 'whole experiment' mean R^2/MSE test errors for Dataset A, which indicate the level of agreement between the simulated and experimental normalised B release curves over the test data. The term 'whole experiment' refers to simulations that partition the data into training and test sets on a whole experiment basis. Note that negative R^2 errors indicate a fit worse than just using a horizontal straight line [40]. Each trained model was established using

Dataset A, 17 I/O combinations and 14 machine learning algorithms. See Section 4.1 for a discussion of these and other static leaching results.

Table 3: *'Whole experiment' mean R^2/MSE test errors as a function of I/O combinations and machine learning algorithms. Training and testing were performed using Dataset A considering the full available data. I/O numbers are given in Table 1. Machine learning algorithm numbers correspond to the algorithms given at the beginning of Section 3. Three relatively good and bad performing algorithms are highlighted in green and red respectively for each I/O combination.*

I/O	Error	Machine Learning Algorithm													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	R^2	<-10000	-0.03	0.94	-0.05	0.91	0.83	0.96	0.96	0.95	0.95	0.95	0.99	0.97	0.65
	MSE	>10000	51.3	2.75	52.37	3.92	7.84	2.43	1.72	2.16	2.36	2.45	0.19	0.59	16.26
2	R^2	<-10000	-0.03	0.95	-0.07	0.92	0.82	0.96	0.96	0.95	0.95	0.95	0.96	0.98	0.58
	MSE	>10000	50.87	2.37	50.76	3.58	7.98	2.39	1.7	2.15	2.33	2.32	0.48	0.85	18.9
3	R^2	<-10000	0.35	0.94	-0.12	0.98	0.96	0.96	0.96	0.95	0.94	0.95	0.99	0.99	0.63
	MSE	>10000	34.01	2.69	60.25	1.11	2.31	1.93	1.73	2.1	2.29	2.64	0.48	0.6	17.15
4	R^2	<-10000	-0.05	0.95	-0.07	0.92	0.86	0.96	0.96	0.94	0.95	0.95	0.98	0.95	0.67
	MSE	>10000	55.67	2.22	53.88	3.89	7.02	2.04	1.77	2.69	2.44	2.42	0.52	1.16	14.52
5	R^2	<-10000	-0.03	0.96	-0.06	0.91	0.82	0.96	0.96	0.95	0.95	0.95	0.98	0.98	0.63
	MSE	>10000	49.96	1.81	52.61	4.21	7.6	2.34	1.65	2.28	2.31	2.5	0.3	0.85	17.35
6	R^2	<-10000	-0.03	0.94	-0.05	0.91	0.85	0.96	0.96	0.96	0.94	0.95	0.99	0.96	0.6
	MSE	>10000	51.24	3.09	55.55	4.43	7.33	2.05	1.66	1.97	2.3	2.6	0.27	0.73	18.1
7	R^2	<-10000	-0.02	0.27	-0.05	0.16	0.13	0.19	-7.81	-5.16	-3.71	0.32	0.8	0.67	-0.02
	MSE	>10000	49.12	34.53	48.68	37.13	39.72	35.84	314.32	217.5	224.34	31.22	7.01	11.96	49.81
8	R^2	<-10000	-0.04	0.94	-0.07	0.91	0.83	0.96	0.96	0.95	0.95	0.95	0.99	0.97	0.53
	MSE	>10000	53.03	2.55	52.35	4.11	7.62	2.3	1.64	2.09	2.36	2.51	0.43	0.71	18.01

9	**R^2**	<-10000	-0.03	0.95	-0.08	0.9	0.88	0.96	0.95	0.94	0.95	0.95	0.98	0.95	0.65
	MSE	>10000	53.02	2.33	52.64	4.45	6.06	2.59	2.19	2.69	2.44	2.71	0.52	1.16	15.58
10	**R^2**	<-10000	-0.03	0.94	-0.06	0.85	0.27	0.95	0.93	0.92	0.9	0.92	0.97	0.93	0.56
	MSE	>10000	51.15	2.63	55.38	7.1	35.85	2.27	3.5	4.27	4.9	3.8	0.68	1.14	20.18
11	**R^2**	<-10000	-0.03	0.95	-0.04	0.92	0.83	0.96	0.96	0.95	0.95	0.95	0.99	0.97	0.55
	MSE	>10000	54.15	2.7	57.43	3.85	7.83	1.83	1.73	2.17	2.39	2.44	0.45	0.76	17.93
12	**R^2**	<-10000	-0.02	0.95	-0.07	0.92	0.86	0.94	0.96	0.95	0.95	0.95	0.99	0.95	0.6
	MSE	>10000	56.58	2.64	53.76	3.64	6.99	3.09	1.69	2.09	2.31	2.54	0.35	0.7	18.14
13	**R^2**	<-10000	-0.03	0.94	-0.06	0.85	0.87	0.9	0.95	0.92	0.93	0.93	0.93	0.92	0.49
	MSE	>10000	53.93	2.71	54.57	6.83	5.7	5.6	2.46	3.62	3.1	3.23	1.44	1.49	22.94
14	**R^2**	0.94	-0.03	0.95	-0.05	0.93	0.83	0.89	0.97	0.97	0.97	0.95	0.99	0.98	0.93
	MSE	2.3	50.77	2.47	52.87	3.51	7.56	6.24	1.46	1.45	1.34	2.34	0.46	0.93	3.63
15	**R^2**	<-10000	-0.01	0.32	-0.06	0.14	0.12	0.27	-4.29	-1.63	-0.54	0.37	0.77	0.58	0.07
	MSE	>10000	50.93	36.53	54.82	42.8	41.09	33.92	250.1	106.37	81.07	33.05	7.67	12.78	43.59
16	**R^2**	<-10000	-0.04	0.35	-0.05	0.16	0.08	0.32	-2.55	-0.68	-0.28	0.36	0.78	0.58	0.09
	MSE	>10000	55.83	31.99	49.79	38.37	45.49	31.61	163.66	73.41	67.38	31.39	5.43	12.56	43.11
17	**R^2**	<-10000	-0.02	0.33	-0.06	0.25	0.15	0.33	-2.55	-0.68	-0.28	0.36	0.86	0.7	0.11
	MSE	>10000	49.89	33.66	51.99	39.62	38.69	31.03	163.66	73.41	67.38	31.22	4.31	12.83	45.15

Using Table 3, the effect of I/O combinations and machine learning algorithms on 'whole experiment' predictive performance can be examined. As a first example, Figure 1 shows Dataset A mean R^2/MSE test errors as a function of the 17 input/output combinations using both boosted ensemble and GPR ('ardsquaredexponential' kernel) methods. Maximised R^2 and minimised MSE errors indicate I/O combination 7, and 15-17 performed poorly for both algorithms, although the remaining I/O combinations performed well. I/O 7 excludes all

elemental concentrations and pH, and I/O 15-17 also exclude all or potentially important species concentrations. Therefore, a poor performance is expected based on our existing knowledge of the leaching process (see Section 4). As a second example, Figure 2 presents Dataset A R^2/MSE test errors across the machine learning algorithms for I/O combinations 1 and 7. Again, R^2/MSE test errors worsen for I/O combination 7 for which all elemental concentrations and pH are excluded from prediction. To finish, the predicted static leaching performance is illustrated (see Figure 3) for I/O combination 1 for both Dataset A test data using bagged random forest and GPR (square exponential kernel) methods and for independent group data (Datasets B-D) using a trained Dataset A bagged random forest method. Note that predicted normalised releases lie within experimental error of the dissolution data. In particular, Figure 3c demonstrates an important result because although algorithm training was achieved using dissolution data obtained with a complex glass, independent testing was performed with data that used a simplified four component glass.

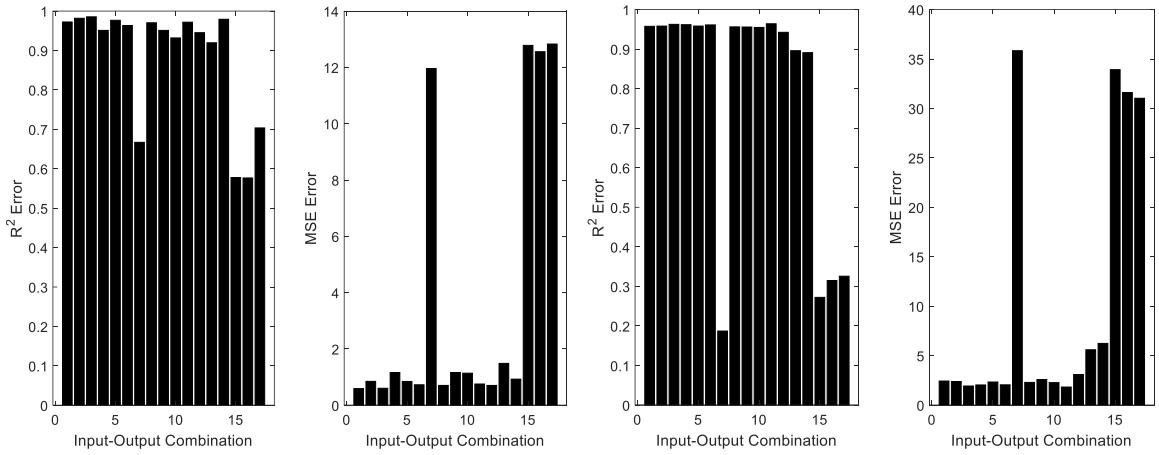


Figure 1: *Mean R^2/MSE test errors as a function of the 17 I/O combinations using boosted ensemble [Left] and GPR ('ardsquaredexponential' kernel) [Right] methods. Training and testing utilised the full Dataset A.*

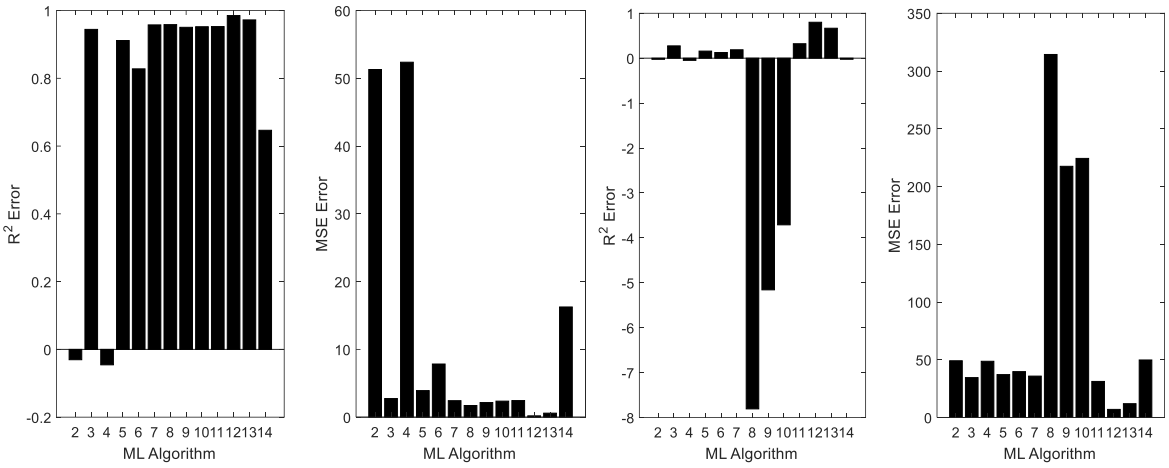


Figure 2: *Mean R^2/MSE test errors as a function of the 14 machine learning algorithms for I/O combinations 1 [Left] and 7 [Right]. Due to the large errors associated with multiple linear regression, this algorithm's results are excluded from the graph. Training and testing utilised the full Dataset A.*

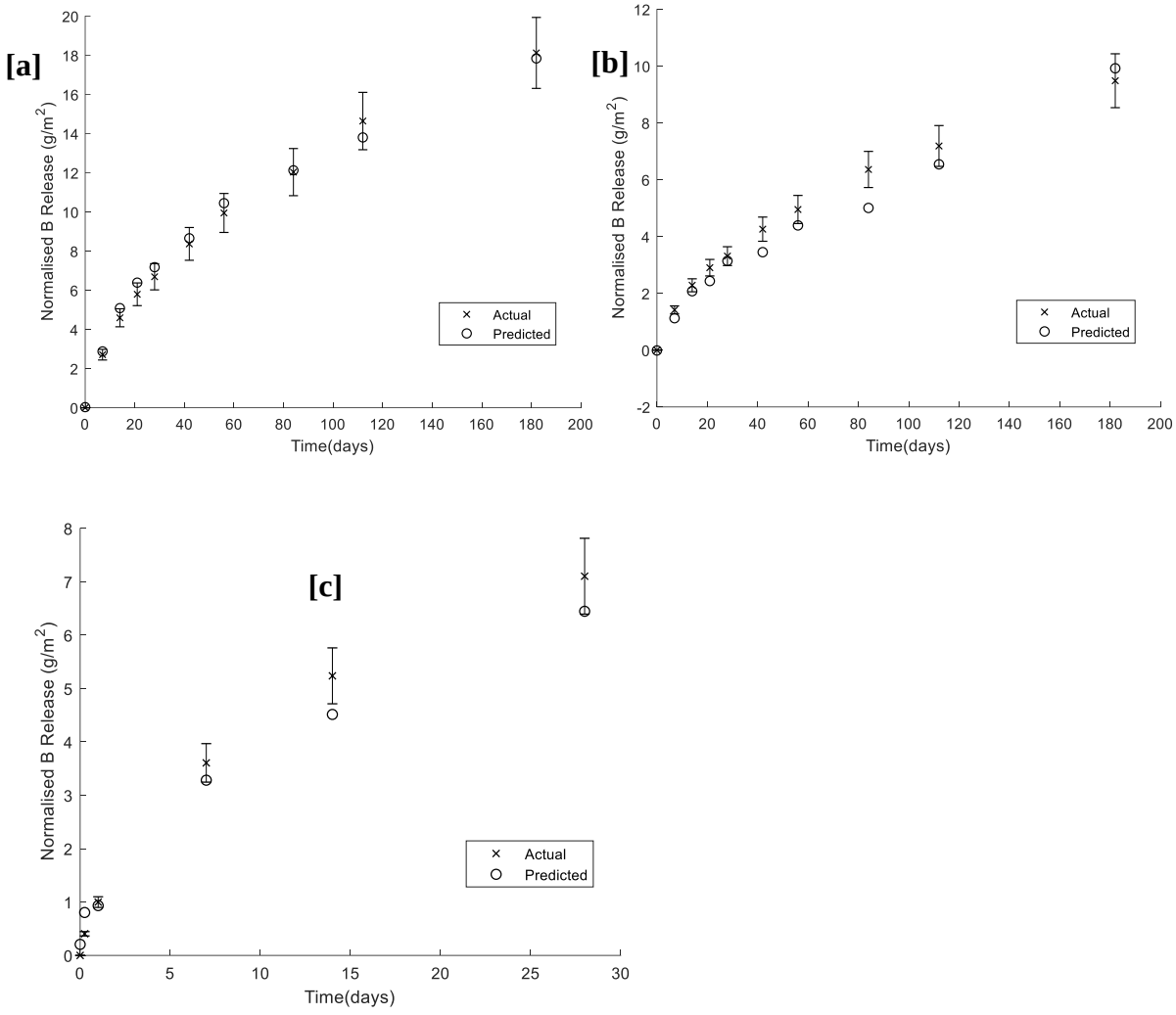


Figure 3: *Example predicted vs measured normalised B release versus time test data curves. Individual leaching experiments and their associated simulated predictions were selected. Training and testing used the full Dataset A with bagged random forest* ***[a]*** *and GPR (square exponential kernel) methods* ***[b]****. Additionally shown* ***[c]*** *is a full Dataset A bagged random forest trained model prediction made on data independent of Dataset A (a simplified four component glass). I/O combination 1 was used in all three cases. Whilst experimental triplicate errors were less than 10% on the mean, conservative 10% error bars have been added to experimental data in all of the plots.*

3.1.2. 'Missing Data' Simulations

Table 4 presents the 'missing data' mean R^2/MSE test errors for Dataset A, whereby 'missing data' refers to simulations that have partitioned training and test data on a specific time

measurement basis within each individual experiment. The R^2/MSE errors again indicate the level of agreement between the simulated and experimental B normalised release curves over the test data. Model training used Dataset A, 17 I/O combinations and 14 machine learning algorithms.

***Table 4**: 'Missing data' mean R^2/MSE test errors as a function of I/O combinations and machine learning algorithms. Training and testing were performed using Dataset A considering the fully available data. I/O numbers are given in Table 1. Machine learning algorithm numbers correspond to the algorithms given at the beginning of Section 3. Three relatively good and bad performing algorithms are highlighted in green and red respectively for each I/O combination.*

I/O	Error	Machine Learning Algorithm													
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1	**R^2**	0.99	-0.01	0.97	-0.01	0.96	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.61	52.06	1.37	50.76	2.16	0.76	0.54	0.58	0.57	0.59	2.12	0.38	1.04	0.48
2	**R^2**	0.99	-0.01	0.98	-0.01	0.96	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.58	50.32	1.29	51.07	2.07	0.71	0.28	0.56	0.56	0.58	1.93	0.34	1.06	0.52
3	**R^2**	0.99	0.33	0.97	-4.44	0.98	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.57	35.95	1.32	267.54	0.79	0.69	0.27	0.56	0.57	0.59	2.08	0.41	1.04	0.46
4	**R^2**	0.99	-0.01	0.97	-0.01	0.96	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.62	51.43	1.37	51.38	2.32	0.7	0.43	0.57	0.57	0.59	2.07	0.44	1.04	0.5
5	**R^2**	0.99	-0.01	0.97	-0.01	0.96	0.98	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.57	52.25	1.44	52.54	2.28	0.8	0.36	0.57	0.57	0.59	2.23	0.31	1.04	0.53
6	**R^2**	0.99	-0.01	0.97	-0.01	0.96	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.56	52.86	1.35	51.75	1.99	0.75	0.3	0.56	0.57	0.59	2.22	0.47	1.04	0.44
7	**R^2**	0.59	0	0.31	-0.01	0.39	0.37	0.87	0.6	0.6	0.6	0.76	0.85	0.72	0.83
	MSE	20.58	50.76	34.99	50.45	31.36	30.63	6.13	18.88	19.37	19.53	10.99	4.85	10.07	8.33

8	**R^2**	0.99	-0.01	0.97	-0.01	0.96	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.57	52.87	1.48	51.54	2.12	0.66	0.25	0.57	0.59	0.59	2.28	0.35	0.85	0.46
9	**R^2**	0.98	-0.01	0.97	-0.01	0.95	0.99	0.99	0.98	0.98	0.98	0.96	0.99	0.97	0.99
	MSE	1.09	49.83	1.85	52.16	2.45	0.74	0.73	1.01	1.01	0.96	2.28	0.45	1.3	0.69
10	**R^2**	0.98	-0.01	0.96	-0.01	0.92	0.96	0.99	0.98	0.98	0.98	0.94	0.97	0.96	0.99
	MSE	0.99	51.72	1.83	53.08	4.12	1.78	0.46	0.94	0.97	0.98	3.21	0.76	1.41	0.52
11	**R^2**	0.99	-0.01	0.97	-0.01	0.96	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.57	52.06	1.44	51.25	2.16	0.74	0.4	0.56	0.56	0.58	2.18	0.48	1	0.48
12	**R^2**	0.99	0.11	0.97	-0.01	0.99	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99
	MSE	0.58	45.55	1.48	50.88	0.54	0.35	0.38	0.57	0.57	0.59	2.27	0.45	1.04	0.5
13	**R^2**	0.97	-0.01	0.95	-0.01	0.91	0.98	0.98	0.97	0.97	0.97	0.94	0.98	0.97	0.97
	MSE	1.64	51.7	2.48	52.32	4.68	1.25	1.05	1.37	1.33	1.34	3.02	0.84	1.62	1.56
14	**R^2**	0.98	-0.01	0.97	-0.01	0.96	0.98	0.96	0.98	0.98	0.98	0.96	0.98	0.98	0.98
	MSE	0.95	50.72	1.51	53.2	2.28	0.93	2.45	0.91	0.92	0.89	2.05	0.59	0.99	1.03
15	**R^2**	-5.07	0	0.35	-0.01	0.38	0.37	0.89	0.63	0.61	0.61	0.74	0.85	0.74	0.71
	MSE	394.43	51.92	32.58	53.29	32.88	33.53	5.83	19.2	19.63	19.36	12.8	5.36	8.57	14.82
16	**R^2**	0.62	-0.01	0.37	-0.01	0.37	0.38	0.9	0.65	0.65	0.66	0.73	0.9	0.77	0.71
	MSE	20.04	53	33.67	52.8	32.48	32.51	5.05	17.39	17.73	17.56	13.25	3.5	9.5	14.78
17	**R^2**	-4.11	-0.01	0.38	-0.01	0.38	0.36	0.9	0.66	0.65	0.66	0.75	0.92	0.77	0.71
	MSE	233.26	53.81	31.91	51.63	30.62	33.49	5.04	17.47	17.73	17.56	13.08	3.43	9.5	14.96

Using Table 4 data, Figure 4 presents mean R^2/MSE test errors for Dataset A as a function of the 17 input/output combinations using both GPR ('ardsquaredexponential' kernel) and neural network methods. Figure 5 shows mean R^2/MSE test errors for Dataset A across the different machine learning algorithms for I/O combinations 2 and 8. Maximised R^2 and minimised MSE errors indicate GPR ('ardsqexponential' kernel) (learning algorithm 7) and bagged random

forest (learning algorithm 12) methods gave high and the most accurate predictions across both combinations. Figure 6 shows predicted (test data) versus measured normalised B release for the neural network method and I/O combination 1 when training/testing using only Dataset A versus when training/testing using Datasets A-D. All normalised B release test data across experiments are shown in the graphs. Predictive accuracy is shown to be high in both cases.

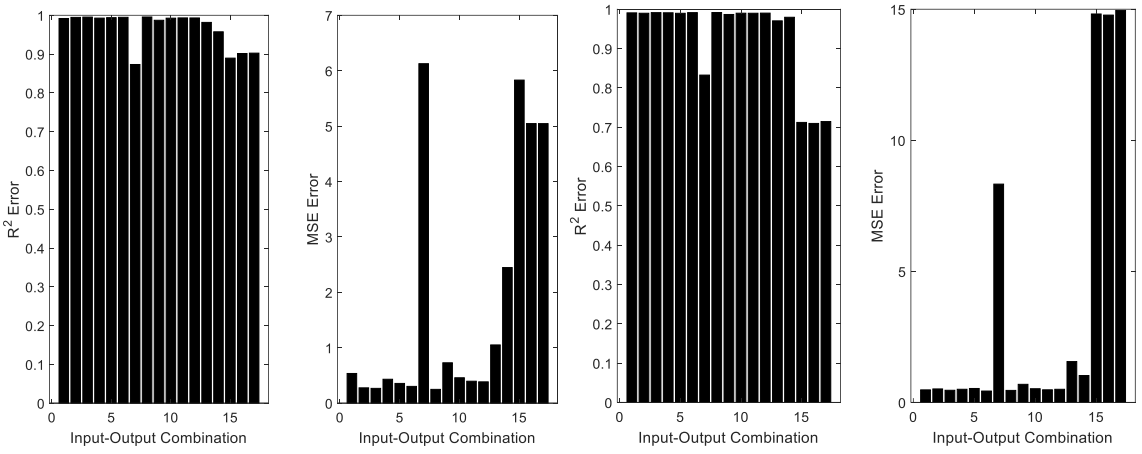


Figure 4: *'Missing data' mean R^2/MSE test errors as a function of the 17 I/O combinations using both GPR ('ardsquaredexponential' kernel) [Left] and neural network [Right] methods. Training and testing utilised the complete Dataset A.*

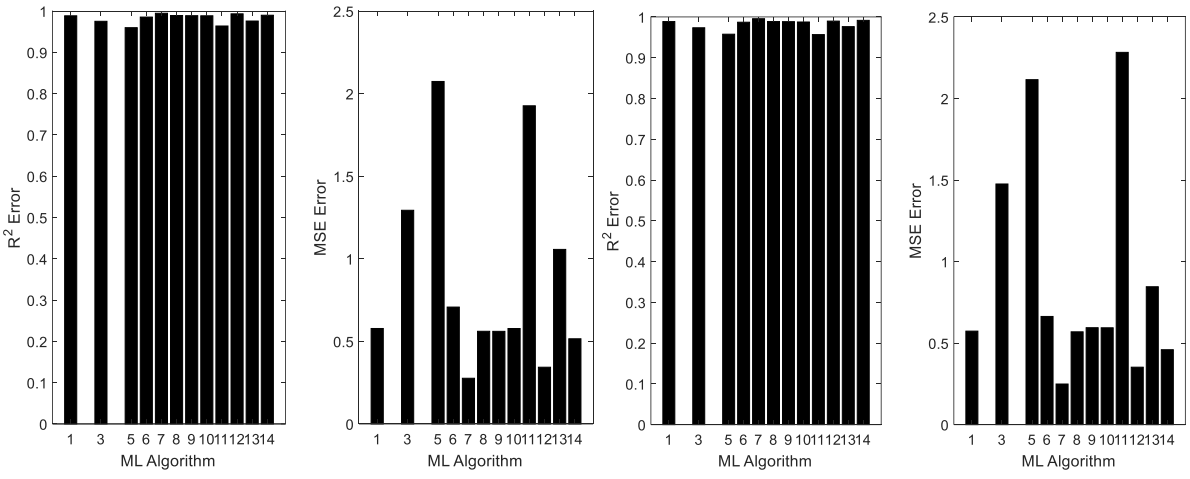


Figure 5: Mean R^2/MSE test errors as a function of the 14 machine learning algorithms for I/O combinations 2 [Left] and 8 [Right]. Due to the large errors associated with ML algorithms 2 and 4, these algorithms are excluded from the graph. The full Dataset A has been used in the model training.

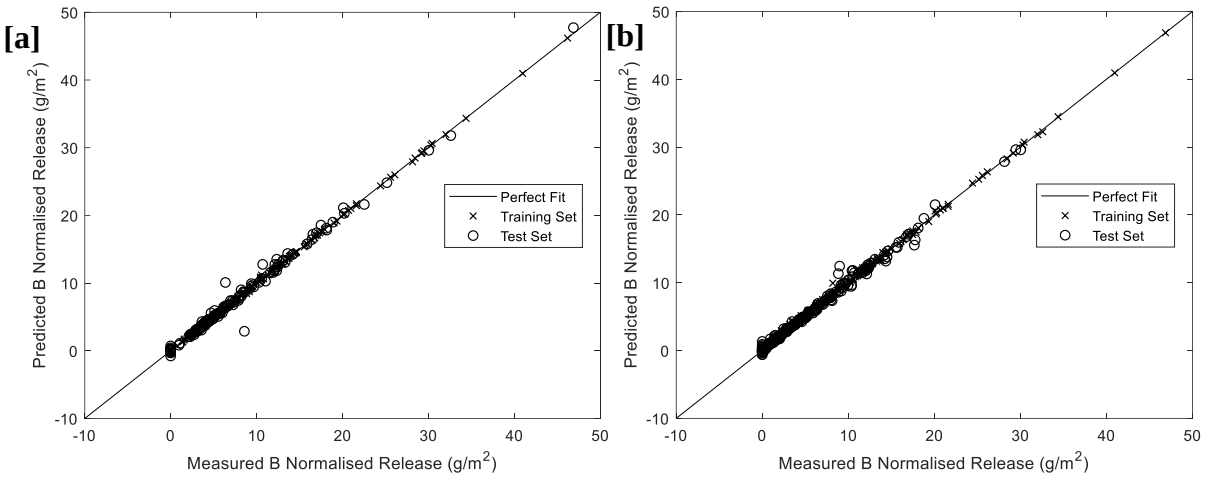
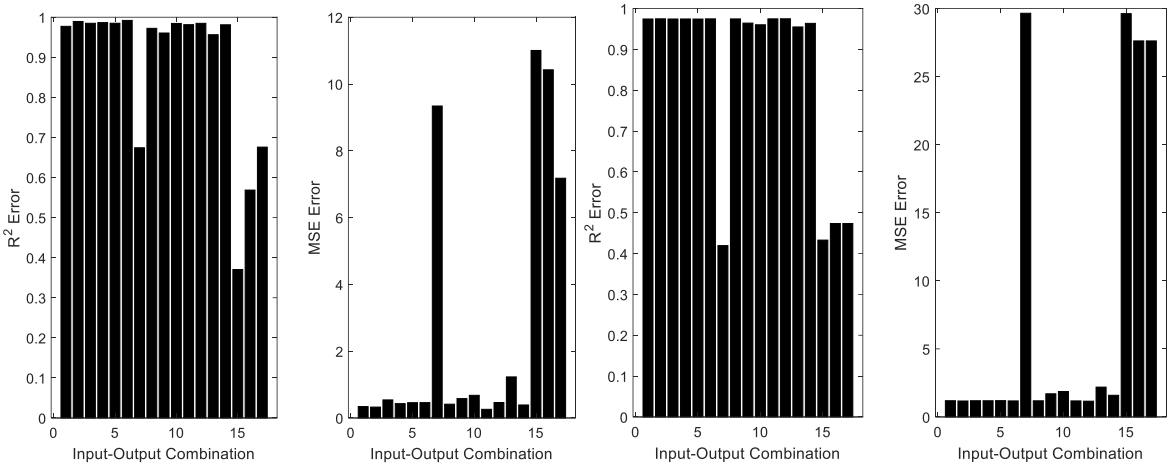


Figure 6: Predicted (test data) versus measured normalised B release curves. These used a neural network with I/O combination 1 with training being applied either on the full Dataset A **[a]** or on the full Dataset A-D **[b]**. Perfect performance would have training/test results following the black straight line. All normalised B release test data across experiments are shown in the graphs.

3.1.3. 'Forecasting' Simulations

Figures 7-8 present selected results taken from the 'forecasting' simulations. These were achieved by using the first half of each experimental duration to predict the behaviour in the second half. Figure 7 shows Dataset A's mean R^2/MSE test errors as a function of the 17 I/O combinations using both bagged random forest and elastic net methods. Maximised R^2 and minimised MSE errors indicate the bagged random forest predicts accurately for I/O combinations 1-6, and 8-14. Figure 8 shows Dataset A's mean R^2/MSE test errors as a function of the 14 machine learning algorithms and I/O combinations 1 and 14. Here, results are presented after training using the full Dataset A data and 80 % of the full Dataset A data. Errors indicate the level of agreement between simulated and test experimental data for normalised B release in the second half of the experiments. In general, considering the different machine learning algorithms, performance is shown here to decrease (although not substantially) as higher fractions of data are removed prior to training/test partition.



361

362 **Figure 7:** 'Forecasting' mean R^2/MSE test errors as a function of the 17 I/O combinations using both
363 bagged random forest [Left] and elastic net [Right] methods. Training and testing utilised the full
364 Dataset A.

365

366

367

368

369

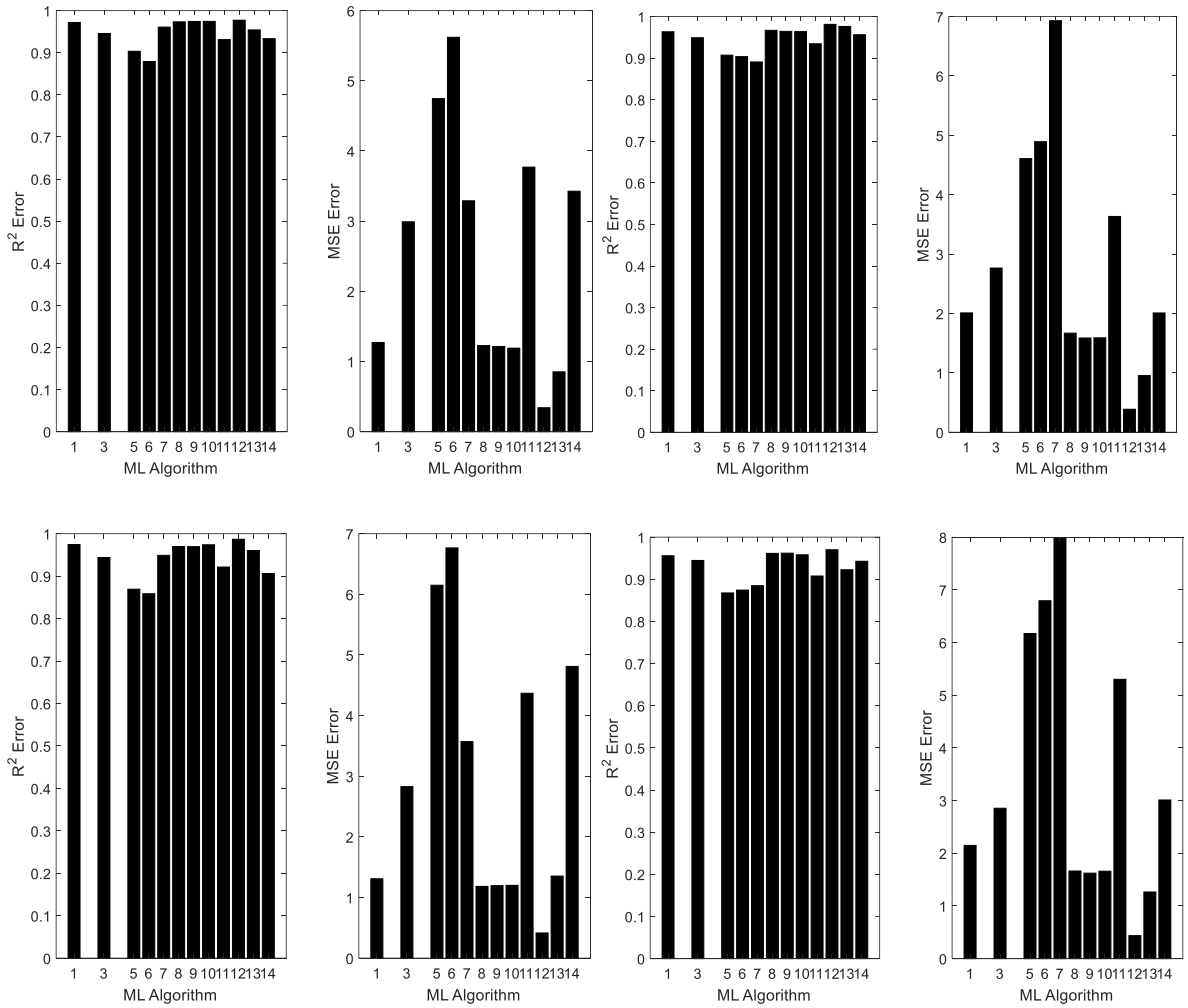


Figure 8: *Mean R^2/MSE test errors as a function of the 14 machine learning algorithms for I/O combination 1 (full Dataset A) [Top Left], I/O combination 14 (full Dataset A) [Top Right], I/O combination 1 (80% Dataset A) [Bottom Left], and I/O combination 14 (80% Dataset A). Due to the large errors associated with ML algorithms 2 and 4, these algorithms are excluded from the plots.*

3.2. Dynamic Leaching Results

3.2.1. Dataset E(-J) Simulations

Table 5 states Dataset E mean R^2/MSE test errors across the four I/O combinations and 14 machine learning algorithms. These explore the ability of machine learning to predict initial (log/non-log) B glass dissolution rates as a function of temperature, pH, with and without mole percentage of oxides/halogens. The effect of the four I/O combinations on Dataset E mean

R^2/MSE test errors is shown in Figure 9 for both neural network and bagged random forest methods. Performance is shown to improve for both learning algorithms when using log B dissolution rate (see for example, I/O combination 1) as an output variable than when using B dissolution rate (see for example, I/O combination 2). This is likely due to an algorithmic issue occurring from the wider range of fitable values if considering B versus log B rates. The effect of the machine learning algorithms is then illustrated in Figure 10 for the same data with I/O combinations 2 and 4. Results indicate that performance does not always diminish with the removal of mole percentage of oxides/halogens as an input variable. This is because for all other input variables the same, I/O 4 has the mole percentage removed whereas I/O 2 includes it. See Section 4.2 for a further discussion of these and other dynamic leaching results.

Table 5: *Mean R^2/MSE test errors as function of I/O combinations and machine learning algorithms. Training and testing used the full Dataset E. I/O numbers are given in Table 2. Machine learning algorithm numbers correspond to the algorithms given at the beginning of Section 3. Three relatively good and bad performing algorithms are highlighted in green and red respectively for each I/O combination.*

I/O	Error	Machine Learning Algorithm													
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1	**R^2**	0.93	0.38	0.93	-5.6	0.93	0.94	0.94	0.93	0.93	0.93	0.82	0.84	0.94	0.95
	MSE	0.07	0.6	0.07	6.4	0.07	0.06	0.05	0.06	0.06	0.06	0.18	0.16	0.06	0.05
2	**R^2**	0.22	0.01	0.19	<-1000	0.39	0.31	0.41	0.27	0.28	0.28	0.46	0.53	0.51	0.65
	MSE	1.64	2.36	2.07	>10000	1.38	1.73	1.36	1.38	1.47	1.43	1.12	1.02	1.05	0.7
3	**R^2**	0.86	0.8	0.86	<-10000	0.86	0.87	0.87	0.87	0.87	0.87	0.84	0.85	0.87	0.86
	MSE	0.14	0.19	0.13	>10000	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.14	0.13	0.13
4	**R^2**	0.31	0.36	0.13	<-10000	0.5	0.54	0.46	0.32	0.33	0.28	0.53	0.56	0.52	0.48

MSE	1.52	1.72	2.25	>10000	1.05	1.12	1.37	1.38	1.4	1.4	0.98	0.94	1.01	1.12

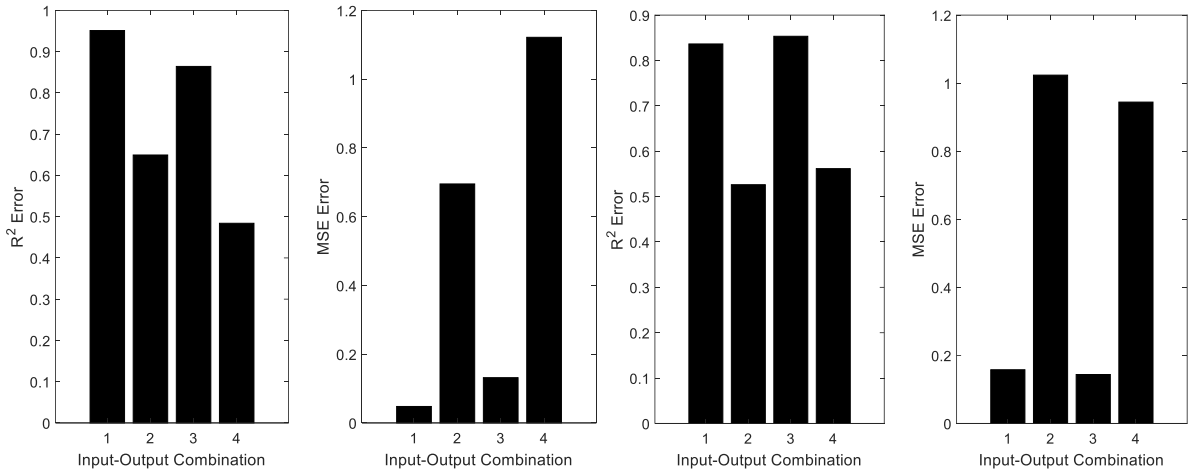


Figure 9: *Dataset E Mean R^2/MSE test errors as a function of the four I/O combinations for neural network [Left] and bagged random forest [Right] methods.*

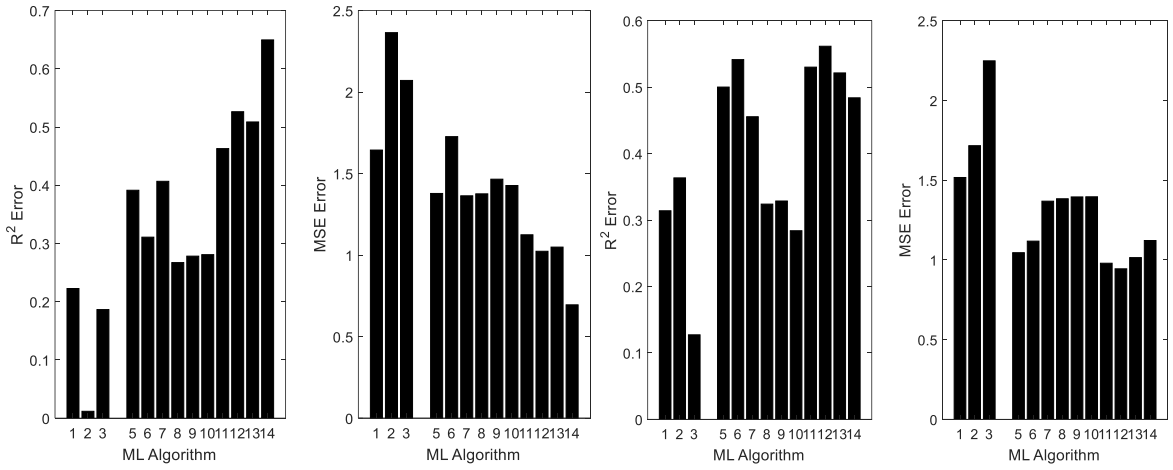


Figure 10: *Dataset E mean R^2/MSE test errors as a function of the 14 machine learning algorithms for I/O combinations 2 [Left] and 4 [Right]. Due to the large errors associated with ML 4, this algorithm is excluded from the graph for improved presentation.*

Model training was also applied on the collective Dataset E-J. Figure 11 presents predicted versus measured mean test boron log dissolution rates for I/O combination 1 using a neural

network. The two plots signify predictions determined after either training using only Dataset E or with the collective Dataset E-J.

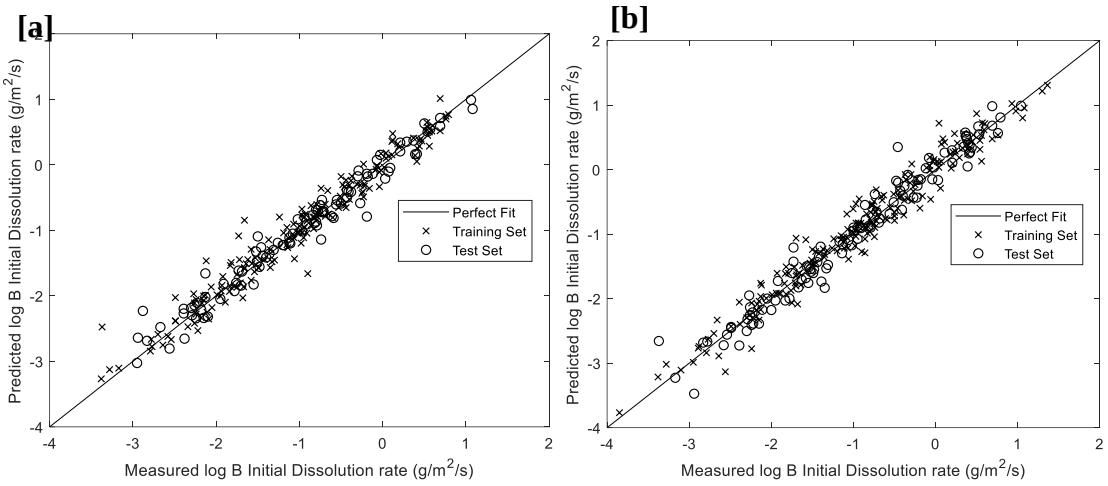


Figure 11: *Predicted versus measured mean training/test dissolution rates for I/O combination 1 with a neural net method. The two plots signify predictions determined after either training with only the complete Dataset E* ***[a]*** *or with the collective complete Dataset E-J* ***[b]****. Perfect performance would have training/test results following the black straight line.*

3.2.2. Dataset K Simulations

Table 6 states mean R^2/MSE test errors across the nine I/O combinations and 14 machine learning algorithms. These illustrate the agreement between predicted and experimental test Na, Si, or Al initial dissolution rates, which represent a species that is soluble, moderately soluble and insoluble, respectively. To explore these results, Figure 12 shows predicted dissolution rates versus measured dissolution rates for I/O combinations 1 and 3 with a random forest method. Both combinations considered final pH as one of the inputs as opposed to a species initial dissolution rate, and the graphs indicate high predictive performance for both Si (I/O 1) and Al (I/O 3) rate prediction. As a second example, Figure 13 presents R^2/MSE test errors as a function of the nine I/O combinations for both GPR (exponential kernel) and ridge methods. Predictive performance is shown to be worse for I/O combinations 1-3 (where final pH is considered as an input) than for I/O combinations 4-9 (whereby a species dissolution rate

is considered as input) across the different machine learning algorithms. Finally, Figure 14 presents full Dataset K R^2/MSE test errors as a function of machine learning algorithm with I/O combination 8. Using four of the high performing algorithms, the effect of dataset size is also illustrated. This I/O combination included Si initial dissolution rate as an input variable and aimed to predict Al initial dissolution rate. The graphs indicate high predictive accuracy in GPR (any kernel) and neural network methods. Moreover, performance appears to be approximately constant up to about a 0.2 ratio (fraction of data removed prior to training/test data partition) and then errors increase at an increasing rate as the fraction of data removed increases.

Table 6: *Mean R^2/MSE Dataset K test errors as a function of I/O combinations and machine learning algorithms. The complete Dataset K was used for model training. I/O numbers are given in Table 2. Machine learning algorithm numbers correspond to the algorithms given at the beginning of Section 3. Three relatively good and bad performing algorithms are highlighted in green and red respectively for each I/O combination.*

I/O	Error	Machine Learning Algorithm													
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
1	**R^2**	0.1	0.94	-0.04	<-10000	0.98	0.98	0.98	0.1	0.1	0.1	0.96	0.94	0.77	0.97
	MSE	1.32	0.08	1.54	>10000	0.03	0.04	0.03	1.26	1.26	1.27	0.06	0.08	0.33	0.05
2	**R^2**	0.4	0.89	0.34	<-10000	0.94	0.94	0.94	0.4	0.4	0.41	0.93	0.93	0.81	0.93
	MSE	0.71	0.13	0.8	>10000	0.07	0.07	0.07	0.72	0.72	0.72	0.09	0.09	0.22	0.08
3	**R^2**	0.3	0.94	0.28	<-10000	0.96	0.96	0.96	0.32	0.31	0.26	0.94	0.94	0.84	0.95
	MSE	1.07	0.09	1.11	>10000	0.06	0.06	0.06	1.04	1.04	1.07	0.08	0.09	0.24	0.07
4	**R^2**	0.93	0.97	0.92	0.11	0.99	0.99	1	0.93	0.93	0.93	0.94	0.96	0.95	0.99
	MSE	0.11	0.04	0.11	1.3	0.01	0.01	0.01	0.1	0.1	0.1	0.09	0.05	0.07	0.01

5	**R^2**	0.96	0.96	0.95	<-10000	0.99	0.99	0.99	0.96	0.96	0.96	0.97	0.97	0.98	0.99
	MSE	0.06	0.06	0.06	2.0	0.01	0.01	0.01	0.06	0.06	0.06	0.03	0.04	0.02	0.01
6	**R^2**	0.95	0.95	0.95	<-10000	0.98	0.98	0.98	0.95	0.95	0.95	0.95	0.96	0.97	0.98
	MSE	0.06	0.07	0.06	>1000	0.02	0.02	0.02	0.06	0.06	0.06	0.06	0.05	0.04	0.02
7	**R^2**	0.92	0.93	0.92	<-10000	0.98	0.98	0.99	0.92	0.92	0.92	0.93	0.94	0.96	0.98
	MSE	0.07	0.06	0.08	>100	0.02	0.02	0.01	0.07	0.07	0.07	0.06	0.05	0.03	0.02
8	**R^2**	0.96	0.96	0.96	<-10000	0.99	0.99	0.99	0.96	0.96	0.96	0.97	0.97	0.98	0.99
	MSE	0.06	0.06	0.06	>100	0.01	0.01	0.01	0.06	0.06	0.06	0.04	0.04	0.02	0.02
9	**R^2**	0.91	0.96	0.91	0.14	0.99	0.99	1	0.91	0.91	0.91	0.94	0.95	0.97	0.99
	MSE	0.14	0.06	0.14	1.3	0.01	0.01	0.01	0.13	0.13	0.13	0.09	0.07	0.05	0.01

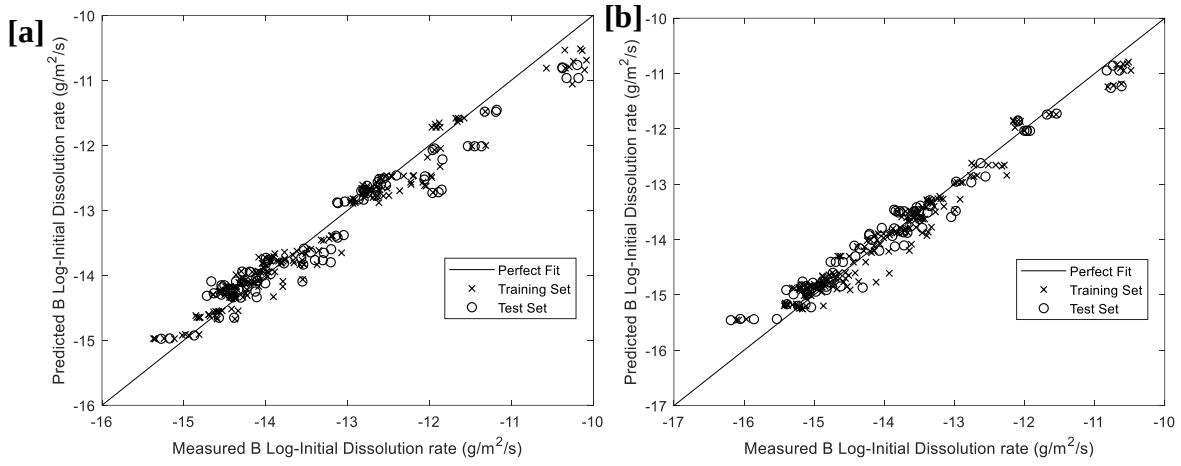
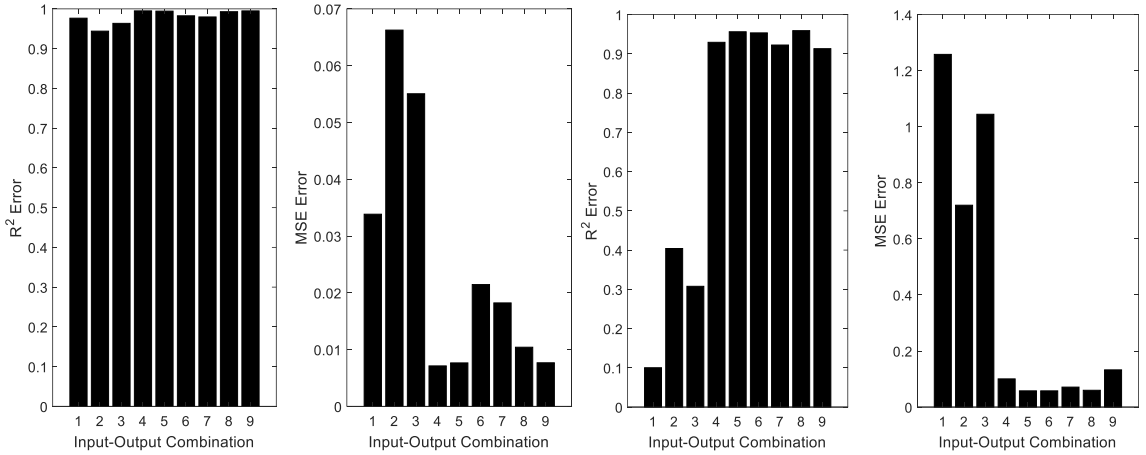
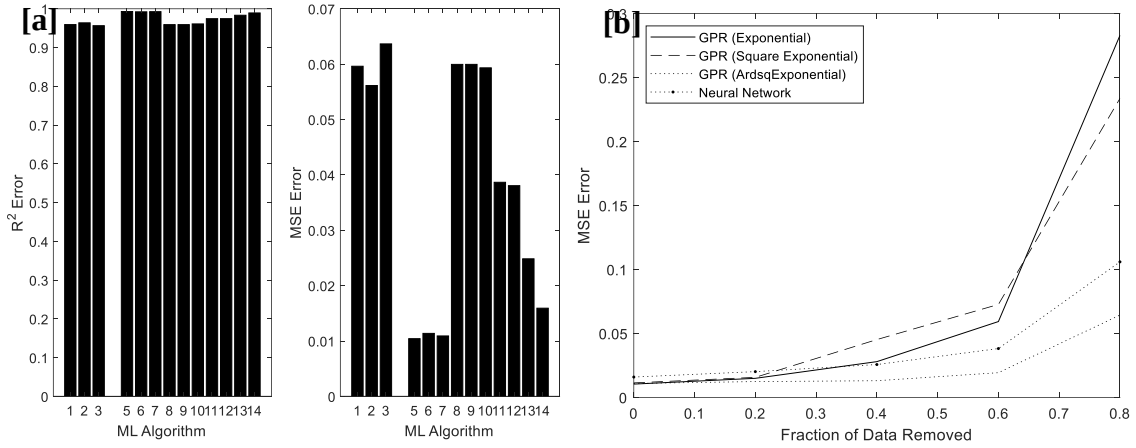


***Figure 12**: Predicted training/test dissolution rates vs measured dissolution rates for I/O combinations 1 **[a]** and 3 **[b]** using a bagged random forest method. The full Dataset K was considered. Perfect performance would have training/test results following the black straight line.*



446

447 ***Figure 13**: R^2/MSE test errors as a function of the nine I/O combinations with both GPR (exponential*
448 *kernel) [Left] and ridge [Right] methods. The full Dataset K was used.*



449

450 ***Figure 14:** Dataset K R^2/MSE test errors as a function of machine learning algorithm with I/O*
451 *combination 8 **[a]**. Here, the full Dataset K was considered in model training. Due to the large errors*
452 *associated with ML 4, this algorithm is excluded from the graph for better visualisation. With four of*
453 *the high performing algorithms, the effect of dataset size is demonstrated using I/O combination 8 **[b]***
454 *by plotting MSE test errors as a function of different starting ratios of Dataset K.*

455 ## 4. Discussion

456 ### 4.1. Static Leaching

457 Following 'whole experiment' model training using Dataset A, predictive performance has
458 been shown to be poor for multiple linear and SVM (Gaussian/polynomial) methods,

irrespective of the I/O combination (Table 3). This is likely due to these algorithms ineffectively treating the non-linear nature of the data. High performing algorithms were bagged random forest and boosted ensemble methods for which these could accurately predict normalised B release for I/O combinations 1-6 and 8-14 (Figures 1 and 3, Table 3). For these combinations, errors were close in magnitude. This suggests that several experimental condition variables including SA/V, powder mass, leachant volume, as well as glass density, pH, Si, Na, Li and Mg elemental release in isolation, dissolution time, and species mass fraction within the pristine glass all had a small individual influence on the ability of these algorithms to predict glass leaching behaviour. The greater importance of pH/elemental release relative to initial experimental conditions is consistent with Figure S1 which shows much lower feature importance for initial experimental conditions (1-30) relative to elemental releases (32-37) for the bagged random forest. The accurate predictions show the value in making use of unstructured data obtained across many campaigns, even though the data may appear somewhat separate (for example, c.f. Dataset A with Ca/Zn versus MW25 glasses). Note also that pH may have had a small effect in isolation on prediction because all tests used deionised water as leachant, and as a consequence, the range of established pH values was relatively small (~7-10).

Test errors did increase for I/O combinations 7 (pH and all elemental normalised release excluded) and 15-17 (either no species normalised release, or all species except Si normalised release excluded, or all species except Si normalised release excluded additionally including the flow rate to glass surface to volume ratio respectively). The error on I/O combination 17 is redundant because flow rate is included as an input variable, while Dataset A was obtained solely under static leaching conditions. High I/O combination 15 and 16 errors suggest although a single species elemental release can be neglected as an input variable, multiple species elemental release cannot. Moreover, whilst the errors for I/O combination 7 were lower

than 15-17, they still indicate an inability to robustly predict static leaching behaviour when solely using experimental initial conditions as input variables and that predictive performance is strongly influenced by the combined effect of pH and the normalised release of elemental species.

Test errors mostly increased when training using the collective Dataset A-D (Supplementary Table S2). This is expected given that many of the added experiments were significantly different from Dataset A, either in regard to composition or experimental methodology such as temperature, SA/V, or long-term dynamic flow compared with static conditions. Consequently, it would be expected that machine learning is less capable of making accurate predictions since there is a larger diversity in the methodologies of the combined data.

For Dataset A-D simulation, bagged random forest followed by boosted ensemble methods again exhibited highest predictive accuracy across the I/O combinations. Both algorithms likely performed well due to their ability to handle non-linear data and having used multiple models to reduce the effect of weak learners. The type of kernel affected SVM and GPR methods as it also did for Dataset A simulation and neural network accuracy improved when adding the Datasets B-D, likely because Dataset A was too small individually for sufficient training/validation/test data partition. The accuracy division between both the I/O combinations 7, 15-17 and the remaining combinations also remained. Results show higher variation across I/O combinations 1-6 and 8-14 than in the case of a Dataset A trained model. Again, this is expected given the larger variation in Dataset A-D experimental conditions. Moreover, for these combinations, it can be seen that in the case of the highest performing algorithm (bagged random forest), either neglecting lithium (Li) elemental release or element mass fractions caused the largest relative increase in predictive error; highlighting the importance of considering the releases of all of the most mobile glass species.

Overall, this study finds that predictions made after training using Dataset A-D can accurately predict static leaching behaviour using a bagged random forest and I/O combination in the range 1-6 or 8-14. This is despite errors being worse than after training solely with Dataset A. Results additionally suggest (Supplementary Table S3) that for specific I/O combinations, the bagged random forest method can accurate predict Dataset B-D behaviour after training using Dataset A. Indeed, it appears that the bagged random forest does have the ability to predict leaching behaviour when static conditions, glass composition, and temperature are not substantially different from the underlying training data. This is illustrated in Figure 3, whereby a trained Dataset A model, could accurately predict the leaching behaviour of a substantially simpler composition glass. If significant differences exist between the training and test data, then predictive inaccuracies occur. However, this is expected given that the underlying training dataset should be sufficiently diverse to formulate appropriate model behaviour for the required experimental conditions. Note that performance generally decreased as higher fractions of data were removed prior to training/test partition.

For the 'missing data' simulations and Dataset A trained models, both bagged random forest and GPR ('ardsqexponential' kernel) methods gave high and the most accurate predictions across the I/O combinations (Table 4, Figure 5). Neural network predictions were also often high (Figure 6), although SVM (Gaussian and polynomial kernel) performance was consistently poor. For Dataset A-D trained models, lasso, ridge, elastic net, and neural networks were also arguably high performing for I/O combinations 1-6 and 8-14 (Supplementary Table S4). Again, there was a division in predictive accuracy between I/O combinations 7, 15-17 and the remaining I/O combinations for both Dataset A and A-D trained models. However, results indicate that in the presence of missing data, machine learning can predict the data accurately, and this appears also largely true both for independent group data (Supplementary Data S5) and for I/O combination 7 whereby only experimental input conditions are given as inputs for

model training. Additionally, kernel variability again influenced predictive accuracy, and performance decreased at an increasing rate as a higher fraction of data was removed prior to training/test data partition.

For the 'forecasting' simulations, additional data is given in Supplementary Tables S6 and S7. Small errors for I/O combinations 1-6 or 8-14 (Table S6, Figures 7 and 8) show the ability of the bagged random forest to forecast well considering Dataset A for model training. Again, this suggests that several experimental condition variables including SA/V, powder mass, leachant volume, as well as glass density, pH, Si, Na, Li and Mg elemental release in isolation, dissolution time, and species mass fraction within the pristine glass all individually had a small influence on the ability of these algorithms to predict glass leaching behaviour. As in the case of 'whole experiment' simulations, results suggest that it is still not possible to forecast when just considering experimental initial conditions as input variables (I/O 7). Moreover, with large errors associated with I/O combinations 15-17, it again does not appear possible to predict when considering the effect of flow rate or when neglecting multiple species elemental releases as inputs. After adding Datasets B-D, test errors increased (Supplementary Table S7) and it generally does not seem possible to accurately forecast due to the increased diversity of the collective dataset. Again, note that kernel type had an effect on predictions, and that performance generally decreased as higher fractions of data were removed prior to training/test partition (Figure 8).

4.2. Dynamic Leaching

After performing model training on Dataset E and the collective Dataset E-J, results indicate that predictive performance was higher for a given machine learning algorithms using log B dissolution rate as an output variable than using B dissolution rate (Table 5, Figure 9). Performance did not always diminish with the removal of mole percentage of oxides/halogens

as an input variable (Table 5, Figure 10). This is not surprising given that for many algorithms increasing the number of features may lead to overfitting. Whether or not performance increases or decreases varies between the machine learning algorithms. For example, if considering log B initial dissolution rate as an output, removing the mole percentage of oxides/halogens as an input variable reduces the ability of the elastic net method but this is not the case for the SVM (Gaussian kernel) method.

Following Dataset E training, results did not indicate one unique algorithm that outperformed all other algorithms across all I/O combinations (Table 5). For I/O combination 1, where logarithmic B rates were predicted using temperature, pH, and mole percentage of oxides/halogens as inputs, predictive performance was high, with neural networks producing smallest test errors, although many algorithms including lasso, ridge, elastic net, GPR (any kernel), and boosted ensemble methods also performed well. For I/O combination 2, which predicted B rates using temperature, pH, and mole percentage of oxides/halogens, neural networks performed best, although errors were significantly higher than I/O combination 1 (Figure 10). For I/O combination 3, which predicted logarithmic B rates using temperature and pH, algorithm performance was similar to I/O combination 1 with predictive performance being high. Finally, for I/O combination 4, which predicted B rates using temperature and pH, the bagged random forest method performed best (Figure 10), although overall performance was still worse than for I/O combinations 1 and 3. It appears that SVM (both Gaussian and polynomial kernels) performed poorly across I/O combinations, as did multiple linear and SVM (linear kernel) methods for I/O combinations 2 and 4.

Following Dataset E-J model training (results given in Supplementary Table S8), GPR ('ardsqexponential') demonstrated smallest test errors for I/O combination 1, although neural network method errors were similar. The performance of both algorithms was again similar for I/O combination 2, although neural network performed best for I/O combination 3. Moreover,

582 for I/O combination 4, it was GPR (exponential kernel) and neural networks that gave the
583 smallest errors. Note that in general, SVM (both Gaussian and polynomial kernels) performed
584 poorly across I/O combinations, and for I/O combinations 2-4, multiple linear, SVM (linear
585 kernel), lasso, ridge, and elastic net methods were unable to accurately predict B initial
586 dissolution rates. Furthermore, similarly to Dataset E, results have showed that the type of
587 kernel does significantly influence predictive performance in both SVM and GPR methods.

588 The addition of Datasets F-J to Dataset E reduced the performance of machine learning
589 algorithms for a given I/O combination (Figure 11, Supplementary Table S8). Nonetheless, in
590 the cases where predictions were accurate for Dataset E training, they remained high in the
591 case of the collective Dataset E-J. The small reduction in performance is likely to be because
592 many of the additional data added were obtained under highly alkaline conditions, and
593 therefore models may have been unable to learn the effective correlations from the bulk Dataset
594 E. Test F-J errors support this view (Supplementary Table S9) because the poor performance
595 computed across the machine learning algorithms and I/O combinations indicate that model
596 correlations learnt using Dataset E were unable to accurately predict the additional data. Note
597 that as expected, as a higher fraction of data was removed from either Dataset E or Dataset E-
598 J, predictive performance decreased. Results were approximately constant up to 20 percent of
599 the data being removed indicating some robustness in the machine learning algorithms.

600 Considering the Dataset K simulations that extend the work of Krishnan *et al*. [24]., predictive
601 performance has been shown to be worse for I/O combinations 1-3 than for I/O combinations
602 4-9 (Table 6, Figure 13) across the different machine learning algorithms. This indicates that it
603 may be better to consider species (Si, Na, or Al) dissolution rate as an input variable than final
604 pH. For the different machine learning algorithms, there was no output species (Si, Na, Al) for
605 which rate prediction consistently showed better performance considering the I/O

606 combinations 1-9 (Table 6, Figure 13). This is despite there being significant differences in the
607 solubility across the species.

608 For I/O combinations 1-3, where final pH was considered as one of the inputs as opposed to a
609 species initial dissolution rate, GPR (any kernel), single regression tree, bagged random forest,
610 and neural networks demonstrated high predictive performance (Table 6, Figures 12-13).
611 Multiple linear, SVM (linear and polynomial kernels), lasso, ridge, and elastic net methods
612 performed poorly (Table 6, Figure 13). For I/O combinations 4-9, whereby a species initial
613 dissolution rate was considered as one of the inputs as opposed to the final pH, GPR (any
614 kernel) and neural network methods consistently predicted initial dissolution rates accurately
615 (see Table 6, Figure 14). The remaining algorithms performed relatively worse, although errors
616 were still small. Note that unlike Krishnan *et al*. [24], this study finds that SVM (Gaussian
617 kernel) produces small test errors for I/O combination 1. The remaining results on the
618 suitability of the other learning algorithms are consistent with those obtained by Krishnan *et*
619 *al*. [24].

620 Again, Dataset K simulation results have demonstrated that the type of kernel appears to
621 influence predictive performance. It is unsurprising that kernel type had an effect because of
622 their differing functional ability to map features to outputs. For example, the study found larger
623 variation in errors across kernel type for SVM than GPR methods. As expected, as a higher
624 fraction of data was removed from Dataset K predictive performance decreased (Figure 14).
625 Performance appears to be approximately constant up to about a 0.2 ratio (fraction of data
626 removed prior to training/test data partition) and then errors increase at an increasing rate as
627 the fraction of data removed increases. This suggests that machine learning algorithm initial
628 dissolution rate prediction may be reasonably robust to dataset size.

Both the static leaching and dynamic leaching simulation results described above are important because they demonstrate that the machine learning methods previously applied to simplistic, three component, non-nuclear glasses by Krishnan *et al*. [24] can be used to accurately predict the dissolution behaviour of more compositionally complex nuclear glasses. The work of Krishnan *et al*. [24] demonstrated that machine learning can be used to predict initial dissolution rates within the envelope of well-structured experimental data. However, this study has used highly unstructured data, and shows the value in using machine learning to predict both static and dynamic leaching behaviour, making use of data that is not well designed for machine learning analysis.

5. Conclusion

Machine learning techniques can predict both the static and dynamic leaching behaviour of radioactive waste glasses. The use of large datasets obtained from a variety of different sources, covering a diverse range of experimental conditions and glass compositions shows an accurate performance that is comparable with similar methods applied to simplistic non-nuclear glasses from more limited datasets. Machine learning can accurately predict leaching behavior, predict missing data, and time forecast. This is provided that the type of machine learning algorithm, model input variables, and diversity or size of the underlying dataset are carefully chosen.

For static leaching, the bagged random forest method can yield highly accurate predictive performance, even when either pH or individual species normalised release or glass composition or several experimental initial condition variables (glass density, powder mass, etc.) are neglected as input variables. It also shows potential in predicting independent group dissolution data, except when using data with increased diversity in the experimental methodology, including where substantial variations in leaching temperature, glass composition, and dynamic compared with static conditions exist.

For dynamic leaching, predictive performance is higher if replacing final pH with a species (Si, Na, or Al) dissolution rate as an input variable, although there is no preferred output species (Si, Na, or Al), despite the difference in solubility between these species. If predicting B rates, the bagged random forest method gives smallest errors using temperature and pH, although neural networks perform best if additionally using the mole percentage of oxides/halogens as an input.

Acknowledgements

The authors would like to thank the Nuclear Decommissioning Authority (NDA) for kindly providing Dataset A. They also wish to thank Stéphane Gin for providing Dataset F.

Funding

This project was funded and supervised as part of an EPSRC funded Imperial-Cambridge-Open University (ICO) Centre for Doctoral Training (CDT) PhD project (EPSRC Grant Number: EP/L015900/1).

References

[1] M.T. Harrison, Vitrification of High Level Waste in the UK, Procedia Mater. Sci. 7 (2014) 10–15. doi:10.1016/J.MSPRO.2014.10.003.

[2] L. Werme, I.K. Björner, G. Bart, H.U. Zwicky, B. Grambow, W. Lutze, R.C. Ewing, C. Magrabi, Chemical corrosion of highly radioactive borosilicate nuclear waste glass under simulated repository conditions, J. Mater. Res. 5 (1990) 1130–1146. doi:10.1557/JMR.1990.1130.

[3] A.J. Connelly, R.J. Hand, P.A. Bingham, N.C. Hyatt, Mechanical properties of nuclear waste glasses, J. Nucl. Mater. 408 (2011) 188–193. doi:10.1016/J.JNUCMAT.2010.11.034.

676 [4] NDA, Radioactive Wastes in the UK: A summary of the 2016 Inventory, 2017.
677 https://ukinventory.nda.gov.uk/wp-content/uploads/sites/18/2017/03/High-Level-
678 Summary-UK-Radwaste-Inventory-2016.pdf Accessed: 18/4/19.

679 [5] P. Jollivet, P. Frugier, G. Parisot, J.P. Mestre, E. Brackx, S. Gin, S. Schumacher,
680 Effect of clayey groundwater on the dissolution rate of the simulated nuclear waste
681 glass SON68, J. Nucl. Mater. 420 (2012) 508–518.
682 doi:10.1016/J.JNUCMAT.2011.10.026.

683 [6] T.M. Iwalewa, T. Qu, I. Farnan, Investigation of the maximum dissolution rates and
684 temperature dependence of a simulated UK nuclear waste glass in circum-neutral
685 media at 40 and 90°C in a dynamic system, Appl. Geochemistry. 82 (2017) 177–190.
686 doi:10.1016/J.APGEOCHEM.2017.05.018.

687 [7] K. Lemmens, The effect of clay on the dissolution of nuclear waste glass, J. Nucl.
688 Mater. 298 (2001) 11–18. doi:10.1016/S0022-3115(01)00590-6.

689 [8] S. Liu, K. Ferrand, K. Lemmens, Transport- and surface reaction-controlled SON68
690 glass dissolution at 30 °C and 70 °C and pH = 13.7, Appl. Geochemistry. 61 (2015)
691 302–311. doi:10.1016/J.APGEOCHEM.2015.06.014.

692 [9] S. Gin, A. Abdelouas, L.J. Criscenti, W.L. Ebert, K. Ferrand, T. Geisler, M.T.
693 Harrison, Y. Inagaki, S. Mitsui, K.T. Mueller, J.C. Marra, C.G. Pantano, E.M. Pierce,
694 J.V. Ryan, J.M. Schofield, C.I. Steefel, J.D. Vienna, An international initiative on
695 long-term behavior of high-level nuclear waste glass, Mater. Today. 16 (2013) 243–
696 248. doi:10.1016/J.MATTOD.2013.06.008.

697 [10] J.P. Busby, J.R. Lee, S. Kender, J.P. Williamson, S. Norris, Modelling the potential for
698 permafrost development on a radioactive waste geological disposal facility in Great
699 Britain, Proc. Geol. Assoc. 126 (2015) 664–674. doi:10.1016/J.PGEOLA.2015.06.001.

[11] P. Frugier, T. Chave, S. Gin, J.-E. Lartigue, Application of the GRAAL model to leaching experiments with SON68 nuclear glass in initially pure water, J. Nucl. Mater. 392 (2009) 552–567. doi:10.1016/J.JNUCMAT.2009.04.024.

[12] S. Gin, Open Scientific Questions about Nuclear Glass Corrosion, Procedia Mater. Sci. 7 (2014) 163–171. doi:10.1016/J.MSPRO.2014.10.022.

[13] Y. Gong, J. Xu, R.C. Buchanan, The aqueous corrosion of nuclear waste glasses revisited: Probing the surface and interfacial phenomena, Corros. Sci. 143 (2018) 65–75. doi:10.1016/J.CORSCI.2018.08.028.

[14] T. Ma, A.P. Jivkov, W. Li, W. Liang, Y. Wang, H. Xu, X. Han, A mechanistic model for long-term nuclear waste glass dissolution integrating chemical affinity and interfacial diffusion barrier, J. Nucl. Mater. 486 (2017) 70–85. doi:10.1016/J.JNUCMAT.2017.01.001.

[15] Y. Minet, B. Bonin, S. Gin, P. Frugier, Analytic implementation of the GRAAL model: Application to a R7T7-type glass package in a geological disposal environment, J. Nucl. Mater. 404 (2010) 178–202. doi:10.1016/J.JNUCMAT.2010.07.015.

[16] P.C. Rieke, S. Kerisit, J. V. Ryan, J.J. Neeway, Adaptation of the GRAAL model of Glass Reactivity to accommodate non-linear diffusivity, J. Nucl. Mater. 512 (2018) 79–93. doi:10.1016/J.JNUCMAT.2018.09.058.

[17] M. Fournier, P. Frugier, S. Gin, Application of GRAAL model to the resumption of International Simple Glass alteration, Npj Mater. Degrad. 2 (2018) 21. doi:10.1038/s41529-018-0043-4.

[18] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, Disease Prediction by Machine

723 Learning Over Big Data From Healthcare Communities, IEEE Access. 5 (2017) 8869–
724 8879. doi:10.1109/ACCESS.2017.2694446.

725 [19] L. Monostori, A. Markus, H. Van Brussel, E. Westkämpfer, Machine Learning
726 Approaches to Manufacturing, CIRP Ann. 45 (1996) 675–712. doi:10.1016/S0007-
727 8506(18)30216-6.

728 [20] T. Terano, Y. Ishino, K. Yoshinaga, Integrating Machine Learning and Simulated
729 Breeding Techniques to Analyze the Characteristics of Consumer Goods, in: Evol.
730 Algorithms Manag. Appl., Springer Berlin Heidelberg, Berlin, Heidelberg, 1995: pp.
731 211–224. doi:10.1007/978-3-642-61217-6_11.

732 [21] van Liebergen, Bart, Machine learning: A revolution in risk management and
733 compliance?, J. Financ. Transform. 45 (2017) 60–67.
734 https://ideas.repec.org/a/ris/jofitr/1592.html (accessed April 18, 2019).

735 [22] F. Marturana, S. Tacconi, A Machine Learning-based Triage methodology for
736 automated categorization of digital media, Digit. Investig. 10 (2013) 193–204.
737 doi:10.1016/J.DIIN.2013.01.001.

738 [23] J. Mei, J. Zhao, Prediction of HIV-1 and HIV-2 proteins by using Chou's pseudo
739 amino acid compositions and different classifiers, Sci. Rep. 8 (2018) 2359.
740 doi:10.1038/s41598-018-20819-x.

741 [24] N.M. Anoop Krishnan, S. Mangalathu, M.M. Smedskjaer, A. Tandia, H. Burton, M.
742 Bauchy, Predicting the dissolution kinetics of silicate glasses using machine learning,
743 J. Non. Cryst. Solids. 487 (2018) 37–45. doi:10.1016/J.JNONCRYSOL.2018.02.023.

744 [25] C.M. Jantzen, C.L. Trivelpiece, C.L. Crawford, J.M. Pareizs, J.B. Pickett, Accelerated
745 Leach Testing of GLASS (ALTGLASS): I. Informatics approach to high level waste

746 glass gel formation and aging, Int. J. Appl. Glas. Sci. 8 (2017) 69–83.
747 doi:10.1111/ijag.12262.

748 [26] MathWorks, MATLAB and Statistics Toolbox Release 2018b, (2018).

749 [27] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, Cambridge, 2006.
750 doi:10.1017/CBO9781107298019.

751 [28] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory to
752 Algorithms, Cambridge University Press, Cambridge, 2014.

753 [29] ASTM International, ASTM C1285-14 Standard Test Methods for Determining
754 Chemical Durability of Nuclear, Hazardous, and Mixed Waste Glasses and Multiphase
755 Glass Ceramics: The Product Consistency Test (PCT), (2014).
756 https://compass.astm.org/EDIT/html_annot.cgi?C1285+14#fn00001 (accessed April
757 18, 2019).

758 [30] M. Collin, M. Fournier, P. Frugier, T. Charpentier, M. Moskura, L. Deng, M. Ren, J.
759 Du, S. Gin, Structure of International Simple Glass and properties of passivating layer
760 formed in circumneutral pH conditions, Npj Mater. Degrad. 2 (2018) 4.
761 doi:10.1038/s41529-017-0025-y.

762 [31] T.L. Goût, M.T. Harrison, I. Farnan, Impacts of lithium on Magnox waste glass
763 dissolution, J. Non. Cryst. Solids. 517 (2019) 96–105.
764 doi:10.1016/J.JNONCRYSOL.2019.04.040.

765 [32] S. Gin, X. Beaudoux, F. Angéli, C. Jégou, N. Godon, Effect of composition on the
766 short-term and long-term dissolution rates of ten borosilicate glasses of increasing
767 complexity from 3 to 30 oxides, J. Non. Cryst. Solids. 358 (2012) 2559–2570.
768 doi:10.1016/J.JNONCRYSOL.2012.05.024.

769 [33] J.D. Vienna, J.J. Neeway, J. V. Ryan, S.N. Kerisit, Impacts of glass composition, pH,
770 and temperature on glass forward dissolution rate, Npj Mater. Degrad. 2 (2018) 22.
771 doi:10.1038/s41529-018-0042-5.

772 [34] R. Guo, C.T. Brigden, S. Gin, S.W. Swanton, I. Farnan, The effect of magnesium on
773 the local structure and initial dissolution rate of simplified UK Magnox waste glasses,
774 J. Non. Cryst. Solids. 497 (2018) 82–92. doi:10.1016/J.JNONCRYSOL.2018.03.002.

775 [35] K. Ferrand, K. Lemmens, Determination of the Forward Rate of Dissolution for
776 SON68 and PAMELA Glasses in Contact With Alkaline Solutions, MRS Proc. 1107
777 (2008) 287. doi:10.1557/PROC-1107-287.

778 [36] A. Elia, K. Ferrand, K. Lemmens, Determination of the Forward Dissolution Rate for
779 International Simple Glass in Alkaline Solutions, MRS Adv. 2 (2017) 661–667.
780 doi:10.1557/adv.2016.672.

781 [37] D.J. Backhouse, A.J. Fisher, J.J. Neeway, C.L. Corkhill, N.C. Hyatt, R.J. Hand,
782 Corrosion of the International Simple Glass under acidic to hyperalkaline conditions,
783 Npj Mater. Degrad. 2 (2018) 29. doi:10.1038/s41529-018-0050-5.

784 [38] J.P. Hamilton, J. Patrick, Corrosion behavior of sodium aluminosilicate glasses and
785 crystals, ProQuest Diss. Theses; Thesis (Ph.D.)--The Pennsylvania State Univ. 1999.;
786 Publ. Number AAI9937980; ISBN 9780599392762; Source Diss. Abstr. Int. Vol. 60-
787 07, Sect. B, Page 3498.; 234 P. (1999).
788 http://adsabs.harvard.edu/abs/1999PhDT.....194H (accessed March 26, 2019).

789 [39] B. Fleury, N. Godon, A. Ayral, S. Gin, SON68 glass dissolution driven by magnesium
790 silicate precipitation, J. Nucl. Mater. 442 (2013) 17–28.
791 doi:10.1016/J.JNUCMAT.2013.08.029.

[40] Mathworks, Evaluating Goodness of Fit, (2019). https://ww2.mathworks.cn/help/curvefit/evaluating-goodness-of-fit.html (accessed May 24, 2019).