

# Einstein-Roscoe Regression for the Slag Viscosity Prediction Problem in Steelmaking

Hiroto Saigo (✉ [saigo@inf.kyushu-u.ac.jp](mailto:saigo@inf.kyushu-u.ac.jp))

Kyushu University

K.C. Dukka Bahadur

Michigan Technological University

Noritaka Saito

Kyushu University

---

## Research Article

**Keywords:** Einstein-Roscoe regression, slag viscosity prediction problem, steelmaking, classical machine learning, temperature

**Posted Date:** January 3rd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1209885/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Einstein-Roscoe regression for the slag viscosity prediction problem in steelmaking

Hiroto Saigo<sup>1,\*</sup>, K.C. Dukka Bahadur<sup>2</sup>, and Noritaka Saito<sup>3</sup>

<sup>1</sup>Kyushu University, Department of Electrical Engineering and Computer Science, 744, Motooka, Nishi-ku, 819-0395, Japan

<sup>2</sup>Michigan Technological University, Department of Computer Science, Houghton, MI, 49931, USA

<sup>3</sup>Kyushu University, Department of Materials Science and Engineering, 744, Motooka, Nishi-ku, 819-0395, Japan

\*saigo@inf.kyushu-u.ac.jp

## ABSTRACT

In classical machine learning, regressors are trained without attempting to gain insight into the mechanism connecting inputs and outputs. Natural sciences, however, are interested in finding a robust interpretable function for the target phenomenon, that can return predictions even outside of the training domains. This paper focuses on viscosity prediction problem in steelmaking, and proposes Einstein-Roscoe regression (ERR), which learns the coefficients of the Einstein-Roscoe equation, and is able to extrapolate to unseen domains. Besides, it is often the case in the natural sciences that some measurements are much more expensive than the others due to physical constraints. To this end, we employ a transfer learning framework based on Gaussian process, which allows us to estimate the regression parameters using the auxiliary measurements available in a reasonable cost.

In experiments using the viscosity measurements in high temperature slag system, ERR is compared favorably with various machine learning approaches in interpolation settings, while outperformed all of them in extrapolation settings. Furthermore, after estimating parameters using the auxiliary dataset obtained at room temperature, increase in accuracy is observed in the high temperature dataset, which corroborates the effectiveness of the proposed approach.

## Introduction

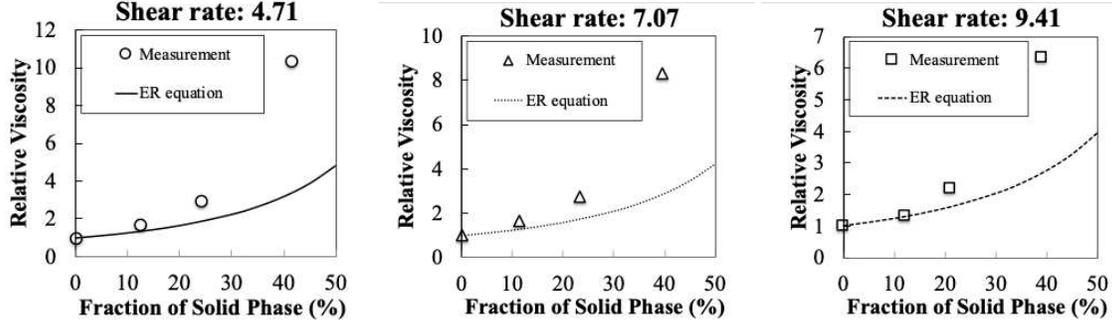
In classical machine learning, regressors are trained with a training dataset, and the generalization performances are measured using a test set. If the dataset at hand is dense enough, then the problem boils down to an *interpolation* problem. However, it is often the case that the measurements are available in a limited domain due to the constraints imposed by the physical environment. In natural sciences, still regressors are desired to return robust and accurate predictions even outside of the training domain, since they can help researchers formulating hypotheses about the observed phenomenon. We call this situation as *extrapolation*, and aim at building robust and accurate regressors both in the interpolation and extrapolation domains.

We are particularly interested in the slag viscosity prediction problem in steelmaking industry. The slag viscosity is known to be a key parameter in controlling and understanding industrial process, and is represented as a function of the fraction, shape and size of the solid phase particles of silicate melts. Numerous viscosity models including Einstein-Roscoe [1], Krieger-Dougherty [2] have been developed due to its industrial importance. However, there exists no universally valid model primarily due to an over simplification of the model despite the complicated target system [3]. For a specific example, we showcase the fitting of the Einstein-Roscoe equation to our in-house dataset measured in a high temperature (1773K) experimental system in Figure 1. It is observed that the fitting of the Einstein-Roscoe equation consistently underestimated the true measurements. In order to overcome this situation, we propose a novel regression algorithms specifically designed for the slag viscosity prediction problem in steelmaking.

In this paper we also address another important problem in experimental science and industry; supplementing the number of high-cost experiments using dataset measured in a reasonable cost. In our specific case, the measurements of viscosity in the high temperature blast furnace system are quite expensive. However, a room temperature experimental system has been designed to mimic the behavior in the high temperature system such that the number of measurements can be compensated [4]. Then we desire to estimate important parameters for the prediction model in the high temperature system using the auxiliary measurements from the room temperature system. The corresponding problem is known as *transfer learning* in machine learning. We employ Gaussian process [5] for this task, and attempt to estimate its parameters in the direction of maximizing the marginal likelihood using the auxiliary dataset.

We make two main contributions in this paper: 1) proposal of the Einstein-Roscoe Regression (ERR), a model based on the Einstein-Roscoe equation, and 2) the development of the parameter selection strategy using the auxiliary dataset.

This paper is organized as follows. In the next section, we describe the main results including the main contributions. In Discussion section, we argue the limitation and the possible extension of the proposed approach. Methods section describes the procedures of the computational experiments and the physical experiments including the parameter settings.



**Figure 1.** Underestimation results of the viscosity by Einstein-Roscoe equations. The measurements are obtained in a matrix of CaO-Al<sub>2</sub>O<sub>3</sub>-SiO<sub>2</sub>-MgO slag at 1773K. Each plot corresponds to a different shear rate. The details regarding the experimental system is described in the Methods section.

## Results

In this section, we first illustrate the problem, introduce the proposed model, and evaluate it both in the simulated datasets and the high temperature blast furnace datasets. We then extend the proposed method such that it can make use of the auxiliary dataset measured at room temperature.

### Einstein-Roscoe Regression (ERR) in the simulated datasets

For simplicity, we begin by estimating the viscosity solely based on the fraction of the solid phase, and ignore the other parameters for the moment. The simulation dataset is generated based on the Einstein-Roscoe equation [1], a popular equation for modeling the viscosity of heterogeneous silicate melts;

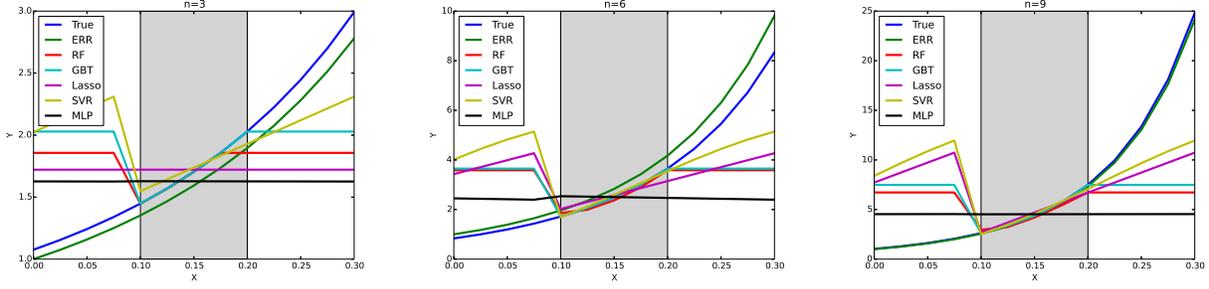
$$\eta_r = (1 - \phi)^{-n}, \quad (1)$$

where  $\eta_r$  stands for relative viscosity and  $\phi$  stands for the fraction of solid phase.  $n$  is typically determined by the particle size, shear rate, and kinetic viscosity of the liquid phase. In [4],  $n$  is determined using the Reynolds number such that  $n = 0.362Re^{-0.189}$ , where  $Re$  is determined by  $\frac{(d/2)^2\gamma\rho}{\eta_L}$ , and  $d$ ,  $\rho$  and  $\eta_L$  are the diameter of the particle, shear rate and the viscosity of the liquid phase, respectively. The coefficients 0.362 and  $-0.189$  are obtained by non-linear regression to the measurements [4]. It is based on a system of physical models, and allows us to understand the reason of the viscosity, but did not reproduce the viscosity in the high temperature system as shown in Figure 1. To this end, we attempt a data science approach; estimate the coefficients  $n$  of Einstein-Roscoe equation by least squares, that is, we solve

$$\log(1 - \phi) \tilde{n} = -\log(\eta_r), \quad (2)$$

for  $\tilde{n}$ . This approach is completely data-driven, and different from any physical model proposed in the literature [3]. We call this approach as Einstein-Roscoe Regression (ERR) below.

The results of fitting ERR and various machine learning algorithms to the simulated datasets are shown in Figure 2. In this simulation experiment, we first fixed  $n$ , generated data points according to Einstein-Roscoe equation, and added Gaussian noise. Then we trained ERR and the baseline methods using the data points whose domain are limited to  $\phi = \{0.1, 0.125, 0.15, 0.175, 0.2\}$  for simulating the extrapolation settings. The gray band in the middle of each figure corresponds to the training domain, and the neighboring left and right domains correspond to the test domains. The baseline methods we tested includes Random Forest (RF) [6], Gradient Boosted Tree (GBT) [7], Lasso [8], Support Vector Regression (SVR) [9] and Multi Layer Perceptron (MLP) [10]. It is clear from Figure 2 that baseline approaches perform well in *interpolation* settings, but are quite erroneous in *extrapolation* settings. Our proposed method ERR, on the other hand, correctly captures the *smooth* and *nondecreasing* properties of the viscosity with respect to the solid state rate. The resulting good performance in both the *interpolation* and *extrapolation* problems is due to the usage of the Einstein-Roscoe equation as a prior knowledge.



**Figure 2.** Prediction results of the regressors in the simulated datasets. Each plot corresponds to a different coefficient, and they are  $n = 3$ (left),  $n = 6$ (center), and  $n = 9$ (right). All the regressors except for ERR show poor extrapolation abilities.

### Fitting ERR to the high temperature blast furnace data

We demonstrate the effectiveness of the proposed approach using the dataset measured in high temperature blast furnace experimental system. The procedure for the measurements are described in Method section. Figure 3 shows the results of fitting ERR to the measurements across different shear rates and bead (particle) sizes. The resulting  $r^2$  was between 0.878 and 0.982, indicating the good fitting performance. Moreover, we have two interesting observations from the estimated coefficients  $n$ ; 1)  $n$  increases with the decreasing bead size, and 2)  $n$  increases with the decreasing shear rate. Both facts are consistent with the previous observation [4]. We can see in Figure 3 that we have obtained 9 estimates of the coefficient  $n$  of Einstein-Roscoe equation. In our ERR framework below, we use these estimated coefficients  $n$  as true response label, and attempt to estimate it from the other experimental parameters;

$$n = f(d, \gamma), \quad (3)$$

where  $d$  denotes bead size, and  $\gamma$  denotes shear rate. Notice that once  $n$  is determined, then the viscosity  $\eta_r$  can be obtained by Equation 1.

### Extrapolation experiments using the high temperature blast furnace data

**Table 1.** Errors in terms of mean squares in both interpolation and extrapolation settings in the high temperature viscosity prediction problem. Statistically significant results are highlighted in a bold font.

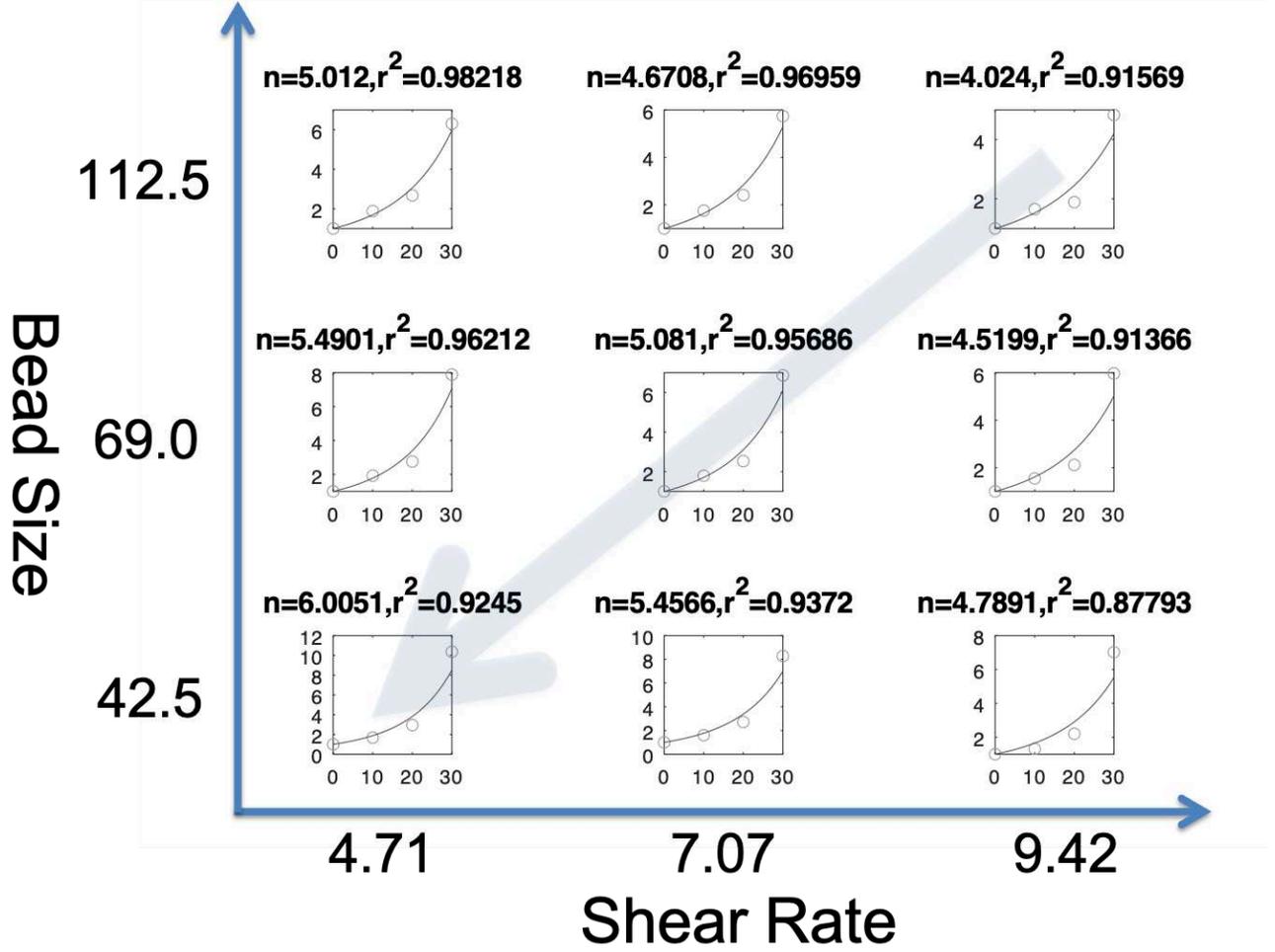
Domain	ERR	ER equation	RF	GBT	LASSO	SVR	MLP
Interpolation	0.075±0.065	0.61±0.29	0.054±0.11	0.14±0.096	0.037±0.0087	0.18±0.12	0.18±0.12
Extrapolation	<b>3.2±4.1</b>	13±9.2	12±7.0	14±8.4	10±5.5	14±8.6	14±8.6

We investigate the effectiveness of ERR compared with the baseline ER equation and various machine learning algorithms in Figure 4. The experimentally measured data points are denoted as True (black circles), and the prediction result of each method is displayed in a line. In this dataset, we have only four data points along the  $x$  axis;  $x = \{0, 0.1, 0.2, 0.3\}$ , but we have reserved 0 and 0.3 for the extrapolation settings, and used only 0.1 and 0.2 for training. The resulting problem is quite hard, since we have only two points for estimating a non-linear curve.

As expected, all the existing regressors performed well inside of the training domain, but poorly outside (See Figure 4). The ER equation underestimated the viscosity for most of the situations. In contrast to that, our proposed method ERR performed well in both interpolation and extrapolation settings. In order to make a detailed comparison, we summarized the errors in terms of mean squares of each regressor in Table 1. In the interpolation settings, there was no statistically significant difference (except for the inferior result of ER equation) by two sample t-test with 5% significance level. In the extrapolation settings, the differences between ERR and all the others were statistically significant by two sample t-test (highlighted in a bold font).

### Extension of the basis model and the estimation of model parameters using the auxiliary data

Thus far, we had only 9 data points for the prediction of the coefficient  $n$  in the high temperature blast furnace dataset due to the high cost required for the measurements. On the other hand, we have established a room temperature experimental system using polyethylene beads and silicon oil [4]. With this system, we have collected 80 auxiliary data points, and attempt to improve the prediction performance in the high temperature dataset by correctly estimating the common underlying parameters.



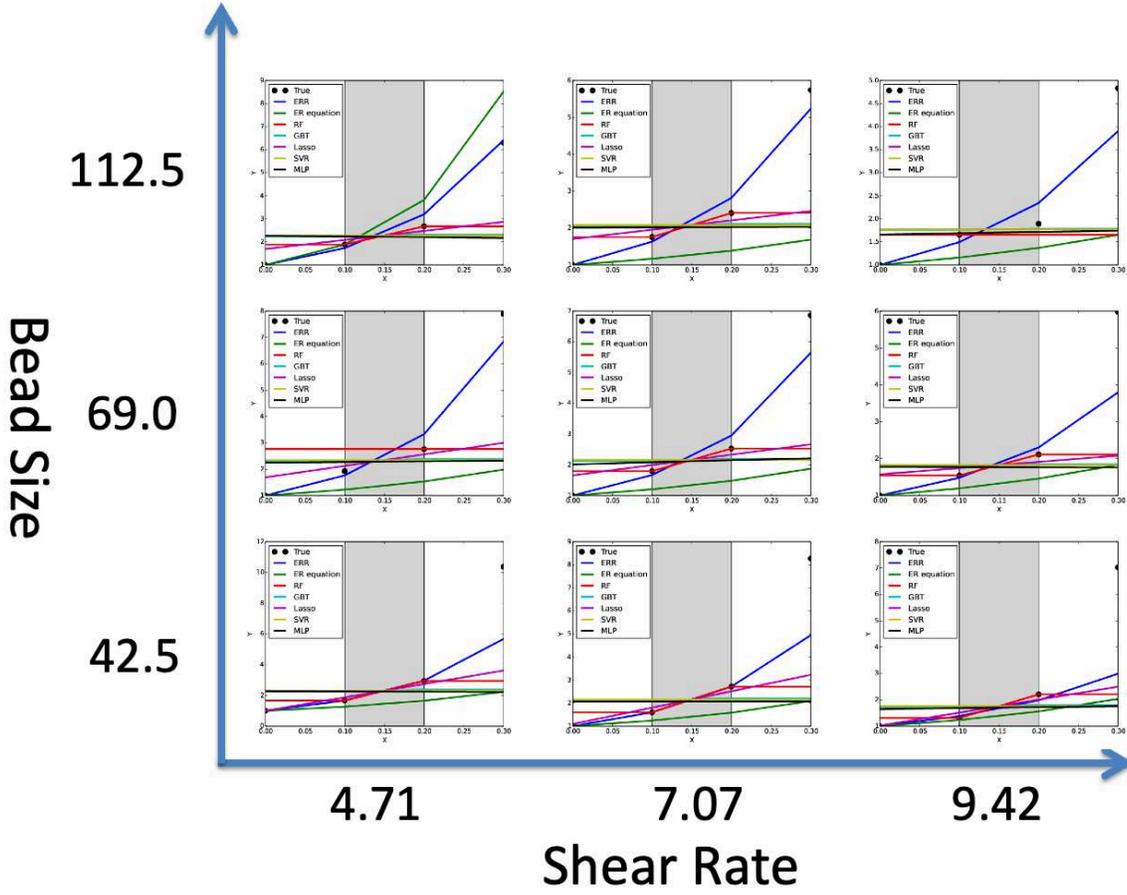
**Figure 3.** Least squares fitting by ERR to the high temperature blast furnace data. The fitting performance in terms of  $r^2$  was as high as between 0.88 and 0.98 across different bead sizes and shear rates. Notice that there are linear relationships (shown as an arrow) between  $n$  and the bead size, and between  $n$  and the shear rate.

First, we consider employing a flexible basis model to ERR. We consider  $n$  as a function of  $\gamma$  and  $d$ , and employ Gaussian process [5]. As a result, the prediction model can be represented as a weighted sum of kernels.

$$n(x) = \sum_i \alpha_i k(x_i, x), \quad (4)$$

where  $x$  corresponds to the measurements with different  $\gamma$ ,  $d$  and,  $i$  runs through the number of measurements, and  $\alpha$  represents the weights learned by Gaussian process. An advantage of Gaussian process is its ability to tune parameters using the training dataset only. Since we already know that a linear model was successful in the previous experiment (Figure 3), we consider mixing a linear kernel with a Gaussian kernel [11] such that an overall trend is captured by a linear model, and small fluctuations are captured by a Gaussian model. The procedure for this experiment is as follows. First we have split the room temperature data into training sets and test sets. Then we have estimated parameters  $\theta$  in the direction of maximizing marginal likelihood using the conjugate gradient descent. Finally we have run leave-one-out cross-validation in the high temperature dataset using the learned parameters. In order to highlight the effectiveness of the parameter selection, we compare the kernel before and after optimizing parameters. A kernel before the parameter optimization is given in the following form;

$$k_{Customized}(x, x' | \theta) = \langle x, x' \rangle + \exp(-|x - x'|^2) + \delta(x, x'). \quad (5)$$



**Figure 4.** The results of fitting in the extrapolation settings. The true measurements in the gray band are used for training, and the true measurements in the neighboring regions are used only for the testing. The ER equation returned the underestimated results for most of the cases. All the other baseline machine learning methods returned flat prediction that goes through the training data points. The proposed method ERR, on the other hand, correctly reproduced the smooth and nondecreasing characteristic due to the usage of Einstein-Roscoe model as a prior knowledge.

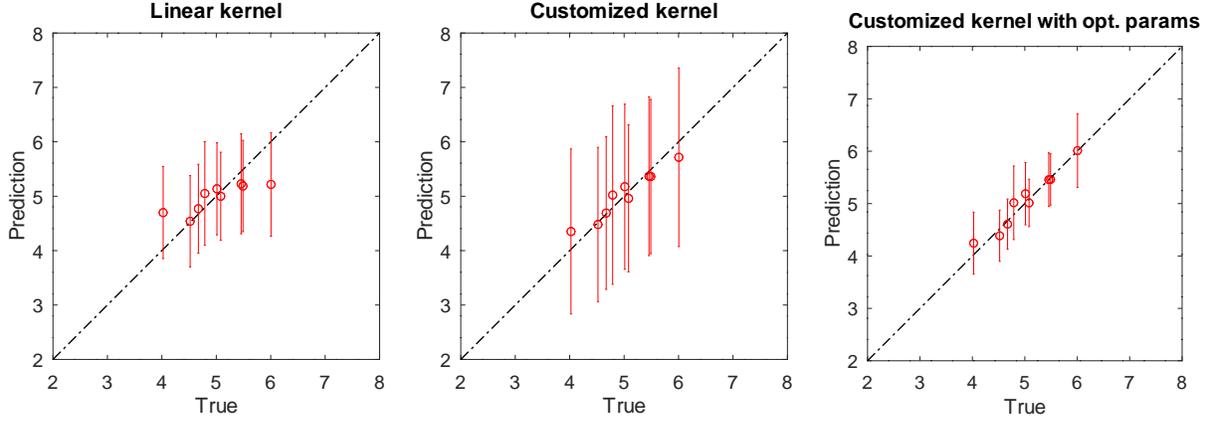
We also compare a kernel with only a linear term;

$$k_{Linear}(x, x' | \theta) = \langle x, x' \rangle. \quad (6)$$

In Table 2, we compare different kernels in terms of root mean squared errors. The errors are measured by leave-one-out cross-validation in the high temperature dataset. We can observe that the amount of errors in the optimized kernels have decreased to almost half after parameter selection using the auxiliary dataset. The results of predicted  $n$  are plotted in Figure 5, where we can confirm that the predicted points after the parameter estimation lie closer to the diagonal, indicating improved correlation with the true label. We can also observe the smaller prediction intervals after parameter selection, which corresponds to the increase of the confidence in the prediction. Our kernel with optimized parameters are as follows;

$$k_{optimized}(x, x' | \theta) = 4.90 \langle x, x' \rangle + 0.105 \exp\left(-\frac{|x - x'|^2}{0.99}\right) + 2.92 \delta(x, x'). \quad (7)$$

A large coefficient to the linear model (4.90) relative to the non-linear model (0.105) indicates inherent linearity over non-linearity. Also a large coefficient to the noise term (2.92) suggests the existence of a relatively large noise in the measurements.



**Figure 5.** Prediction results of the coefficient  $n$  of Einstein-Roscoe equation. A plot with circles lying closer to the diagonal represents better prediction performance, and the predictions with smaller prediction intervals (vertical bar) represents higher confidence in the prediction. Customized kernel with optimized parameters (Right) not only achieved the best prediction performance, but displayed the higher confidence in the prediction than the other kernels.

**Table 2.** Errors in terms of mean squares by different kernels in the high temperature dataset.

kernel type	Linear	Customized	Customized with opt. parameters
MSE	0.147	0.0347	0.0176

## Discussion

Strong linearity in viscosity is observed in the two experiments; 1) fitting of ERR to the high temperature dataset and 2) optimization of the kernel parameters. However, an improved prediction performance is obtained by consideration of nonlinear effect into the model, which confirm the validness in the choice of our kernels. In order to achieve robust regression model, we estimated the coefficient  $n$  of Einstein-Roscoe regression through Equation (3). It worked successfully in our case at the cost of decreasing the number of measurements of viscosity in high temperature experiments. In practice, we obtained 9 coefficients  $n$ , at the cost of the 36 measurements. Generally in machine learning and statistics, a better prediction model comes with more measurements, so one of the future direction would be building a robust and accurate prediction model without decreasing the measurements towards improved prediction performance.

We employed Gaussian process as a regressor to predict the coefficient  $n$  of the Einstein-Roscoe model. In practice, we can employ any nonlinear regressor such as Multiple Layer Perceptron or Random Forest instead of Gaussian process. However, the properties of the Gaussian process such as the parameter selection ability using the training dataset and the availability of the prediction interval, are unique, and we have successfully made full advantage of them in the experiments.

## Methods

### Einstein-Roscoe Regression(ERR)

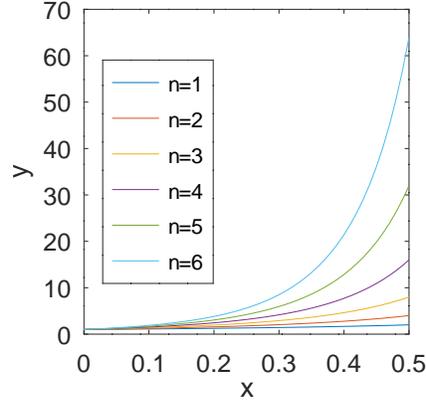
Einstein-Roscoe [1], Krieger-Dougherty [2] and many other models can be described in the common form as;

$$\eta_r = (1 - a\phi)^{-n}, \quad (8)$$

where  $\phi$  is the fraction of the solid phase particle, and  $a$  and  $n$  have been calculated from the shape and size of the solid phase particle in various ways [3]. In this study, we fix  $a = 1$ , and aim at estimating  $n$  from the training dataset, since  $n$  is the most important factor that determines the shape of the curve. Figure 6 illustrates that ER equation with various  $n$  can represent various curves with *smooth* and *nondecreasing* property. Thus we fix our basis model to Equation (8), and aim at predicting the coefficient  $n$  from the available measurements.

### Gaussian Process and the kernel parameter estimation

Suppose that we are given  $d$  dimensional feature  $x \in \mathcal{R}^d$  and the corresponding response  $y \in \mathcal{R}$ , then the training dataset with  $n$  examples is represented as  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . Suppose also that our test set is given as  $x_*$ , and we aim at predicting the function outputs  $f(x_*)$ . Let our *kernel function* be  $k(\cdot, \cdot)$ , and the element of the kernel matrix computed from the training



**Figure 6.** Einstein-Roscoe equation for  $a = 1$ , and various  $n$

dataset be  $K_{i,j} = k(x_i, x_j)$ , and those computed using the training dataset and test dataset be  $(k_*)_i = k(x_*, x_i)$ . The predictive distribution of Gaussian Process is given as

$$p(f(x_*) | X, y, X_*) = \mathcal{N}(f(x_*) | \mu_*, \Sigma_*),$$

where the mean and the covariance of the Gaussian distribution is given as  $\mu_* = k_*^\top K^{-1}y$  and  $\Sigma_* = k(x_*, x_*) - k_*^\top K^{-1}k_*$ , respectively. As a kernel function, we employ a mixture of linear and Gaussian kernels in the following form;

$$k(x, x' | \theta) = \theta_1 \langle x, x' \rangle + \theta_2 \exp\left(-\frac{|x - x'|^2}{\theta_3}\right) + \theta_4 \delta(x, x'). \quad (9)$$

Our high temperature dataset is as small as 9 data points, so we consider using the room temperature dataset for the estimation of the parameters. Since the objective function of the Gaussian process is given as log marginal likelihood

$$\log p(y | \theta) = -\frac{1}{2} \log |K| - \frac{1}{2} y^\top K^{-1} y - \frac{n}{2} \log(2\pi),$$

we consider taking gradient with respect to its parameters;

$$\frac{\partial \log p(y | \theta)}{\partial \theta} = -\frac{1}{2} \text{Tr}(K^{-1} \frac{\partial K}{\partial \theta}) - \frac{1}{2} y^\top K^{-1} \frac{\partial K}{\partial \theta} K^{-1} y$$

With our kernel represented by Equation (9), we can assume that all the parameters are *nonnegative*. Such constraints can be incorporated by rewriting the parameters as  $\theta'_i = e^{\theta_i}$ . Then our kernel can be rewritten as

$$k(x, x' | \theta) = e^{\theta'_1} \langle x, x' \rangle + e^{\theta'_2} \exp\left(-\frac{|x - x'|^2}{e^{\theta'_3}}\right) + e^{\theta'_4} \delta(x, x'). \quad (10)$$

Then the gradient with respect to each parameter can be obtained as follows.

$$\begin{cases} \frac{\partial k(x_n, x'_n)}{\partial \theta'_1} &= e^{\theta'_1} \langle x, x' \rangle \\ \frac{\partial k(x_n, x'_n)}{\partial \theta'_2} &= e^{\theta'_2} \exp\left(-\frac{|x_n - x'_n|^2}{e^{\theta'_3}}\right) = k(x_n, x'_n) - e^{\theta'_4} \delta(n, n') \\ \frac{\partial k(x_n, x'_n)}{\partial \theta'_3} &= e^{\theta'_2} \exp\left(-\frac{|x_n - x'_n|^2}{e^{\theta'_3}}\right) \cdot \frac{\partial}{\partial \theta'_3} \left(-\frac{|x_n - x'_n|^2}{e^{\theta'_3}}\right) = (k(x_n, x'_n) - e^{\theta'_4} \delta(n, n')) \cdot e^{-\theta'_3} |x_n - x'_n|^2 \\ \frac{\partial k(x_n, x'_n)}{\partial \theta'_4} &= e^{\theta'_4} \delta(n, n') \end{cases}$$

Given the gradient, we can employ any gradient based optimization method to update the parameters such that;

$$\theta \leftarrow \theta + \varepsilon \frac{\partial \log p(y | \theta)}{\partial \theta}.$$

In this work, we employed Conjugate Gradient descent optimizer [12].

## Experimental settings of baseline methods

In this subsection, we describe the parameters used for the baseline methods. The coefficients of Einstein-Roscoe (ER) equation is obtained based on the experimental condition as described in the Results section. The model is fully described, and there is no parameter to tune.

The baseline machine learning methods has more than or equal to 1 parameter to tune, which is found by cross validation with grid search in the training dataset. In both Random Forest (RF) and Gradient Boosting Trees (GBT), we chose the number of trees from {1,10,100,1000}. In Lasso, regularization parameter is chose from {0.01,0.1,1,10,100}. In SVR, regularization parameter is chose from {0.01,0.1,1,10,100}, and the kernel is chosen from either 'linear' or 'RBF', and the width parameter of RBF kernel is chosen from {0.01,0.1,1,10,100}. In MLP, the number of hidden layers is chosen from {1,2,3}, and the number of nodes are chosen from {1,3,5,10,20,30}, and the regularization parameter is chosen from {0.0001,0.05}. The max iteration is set to 1000, and the SGD optimizer is employed.

## Viscosity Measurement of Suspensions at Room Temperature

The viscosity measurement system consists of a rotational viscometer (DVII+ or DV2T, AMETEK Brookfield) and a suspension in a 300 ml beaker. Apparent viscosities were systematically measured for suspensions with different bead volume fractions, average diameters, shear rates, and liquid matrix viscosities. The shear rate was calculated from the rotational speed and the dimensions of the beaker (inner diameter 73 mm) and spindle (outer diameter 3.2 mm) using the following equation

$$\gamma = \frac{2\omega}{1 - (r_i/r_o)^2}$$

where  $\gamma$ ,  $\omega$ ,  $r_o$ ,  $r_i$  are the shear rate, angular velocity, inner radius of the outer cylinder (beaker), and radius of the inner cylinder (spindle), respectively. The apparent viscosity of the suspension was calculated from the torque generated in the spindle by rotation. The variation of the apparent viscosity was about 10%. Silicone oil (KF-96, Shin-Etsu Chemical) with viscosities of 0.5, 1.0, 2.0, and 3.0 Pa·s at 24 °C was used as the low polarity liquid matrix, and the relative permittivity ranged from 2.7 to 2.8, depending on the viscosity. Polyethylene beads (LE-1080, Sumitomo Seika) with average diameters of 9.35, 162.5, 340.0, and 602.5  $\mu\text{m}$  were selected as dispersed solid particles, whose sphericity and particle distribution were confirmed to be relevant by using scanning electron microscopy (SEM) imaging. Please refer to [4] for more details.

## Viscosity Measurement of Slag Suspensions at High Temperature

Calcined CaO and MgO powders were added to CaO and MgO saturated slag, respectively, to prepare a suspension of CaO and MgO particles dispersed in the slag. According to the phase diagram, the chemical composition of a quasi-ternary system 53CaO-35Al<sub>2</sub>O<sub>3</sub>-3SiO<sub>2</sub>-8MgO (mass %), has the eutectic temperature of lime (CaO) and periclase (MgO) at 1,773 K, suggesting that CaO and MgO do not chemically dissolve at the selected composition and at 1,773 K. As reagents, powders of CaCO<sub>3</sub>, Al<sub>2</sub>O<sub>3</sub>, SiO<sub>2</sub>, and MgO (Sigma-Aldrich Japan) were carefully weighed to achieve the specified composition and thoroughly mixed in an alumina mortar. The powder batches were pre-melted in a resistance furnace using a platinum crucible for 1 hour at 1873 K in air and quenched on a water-cooled copper plate. The respective reagent powders were calcined at 1,473 K for 30 min to prepare CaO and MgO particles for dispersion. The calcined and sieved CaO and MgO particles were mixed with a pre-melted 53CaO-35Al<sub>2</sub>O<sub>3</sub>-3SiO<sub>2</sub>-8MgO (mass%) slag to a predetermined composition before the viscosity measurement. Viscosity measurements were performed using the rotating crucible viscometer apparatus as described in the previous section. A Pt-20mass%Rh crucible filled with the mixed slag and CaO or MgO particles was placed in the crucible supporter in the furnace, heated to 1,773 K, and then the viscosity of the CaO or MgO slag dispersion was measured at the same temperature. Please refer to [4] for more details.

## References

1. Roscoe, R. The viscosity of suspensions of rigid spheres. *Br.J.Appl.Phys.* **3(8)**, 267–269 (1952).
2. Krieger, I. M. & Dougherty, T. M. A mechanism for non-newtonian flow in suspensions of rigid spheres. *The Soc. Rheol.* **3(137)**, 137–152 (1959).
3. Liu, Z. A., Pandelaers, L., Blanpain, B. & Guo, M. Viscosity of heterogeneous silicate melts: A review. *Metall. Mater. Transactions B* **49(B)**, 2469–2486 (2018).
4. Saito, N., Hara, D., Teruya, S. & Nakashima, K. Viscosity of slag suspensions with a polar liquid matrix. *ISIJ Int.* **60(12)**, 2807–2818 (2020).
5. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, 2005).

6. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
7. Friedman, J. Greedy function approximation: a gradient boosting machine. *Annals Stat.* **29**, 1189–1232 (2001).
8. Tibshirani, R. Regression shrinkage and selection via the LASSO. *J. Royal. Stat. Soc B.* **58(1)**, 267–288 (1996).
9. Smola, A. J. & Vapnik, V. N. Support vector regression machines. In *Proceedings of the advances in Neural Information Processing Systems* (1997).
10. Rosenblatt, F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan Books, Washington D.C., 1962).
11. Schölkopf, B. & Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, 2002).
12. Golub, G. H. & Loan, C. F. V. *Matrix computations* (Johns Hopkins University Press, 1996).

### **Acknowledgements (not compulsory)**

The work was supported by KAKENHI, Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science (JSPS), under Project 19H04176 and Project 21H01684.

### **Author contributions statement**

H.S., N.S. and K.C.D.B. conceived the project. N.S. coordinated the physical experiments in the room temperature and the high temperature experimental system. H.S. performed the computational experiments, and drafted the manuscript.

### **Additional information**