

An Interpretable Linear Model Bridging Data-Driven Analysis and Chemical Intuition for Eu^{2+} -Phosphor Emissions

Yukinori Koyama^{a*}, Ryusei Hayasaka^b, Yuta Matsushima^b, Takayuki Nakanishi^c, Takashi Takeda^c and Naoto Hirosaki^c

^aCenter for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Japan; ^bDepartment of Applied Chemistry, Chemical Engineering, and Biochemical Engineering, Yamagata University, Yonezawa, Japan; ^cResearch Center for Electronic and Optical Materials, National Institute for Materials Science, Tsukuba, Japan

Supplemental material

S1. Hyperparameter Optimization

The hyperparameter α in ridge regression, which controls the regularization strength, was determined using 10-fold cross-validation to minimize the average root mean squared error (RMSE) on the validation sets. Figure S1 shows the average RMSE for the training and validation sets as a function of α . The validation RMSE is minimized at $\alpha = 0.01$. A tendency for overfitting is observed when α is smaller than this value, whereas a tendency for underfitting is observed for larger values. Therefore, we adopted $\alpha = 0.01$ as the optimal hyperparameter.

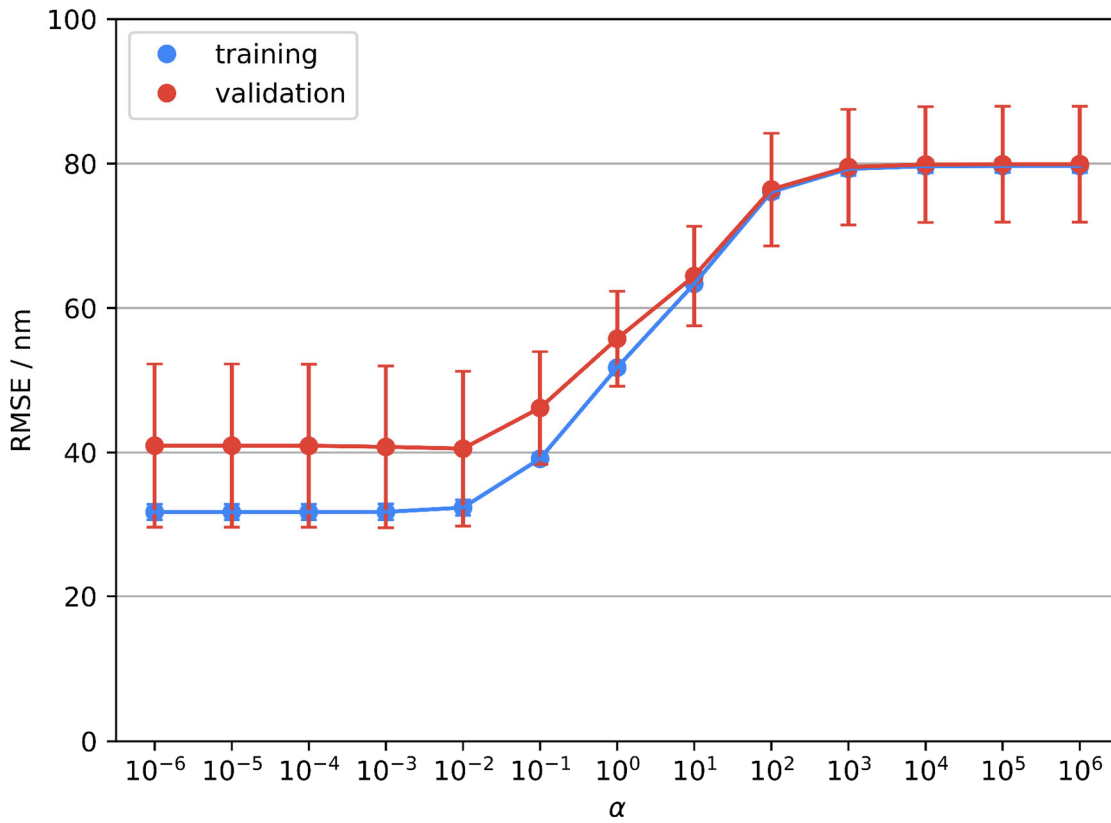


Figure S1. Hyperparameter (α) optimization for ridge regression. The plot shows the average root mean squared error (RMSE) from 10-fold cross-validation for the training (blue) and validation (red) sets as a function of α . The error bars represent the standard deviations of the RMSE across the 10 folds.

S2. Linear Model Coefficients

For reference, the raw coefficients β_j of the final ridge regression model ($\alpha = 0.01$) are shown in Figure S2. The elements are sorted in ascending order of their coefficient values. Although certain trends can be observed at first glance, these coefficients possess gauge freedom, as detailed in Section 2.4 of the main text, and are thus not suitable for interpretation. In particular, direct comparison of the magnitudes of the coefficients between elements with different oxidation numbers is not justifiable because of this gauge freedom. To resolve this interpretation issue, we have introduced the gauge-invariant elemental contribution coefficient (ECC), γ_j , in the main text.

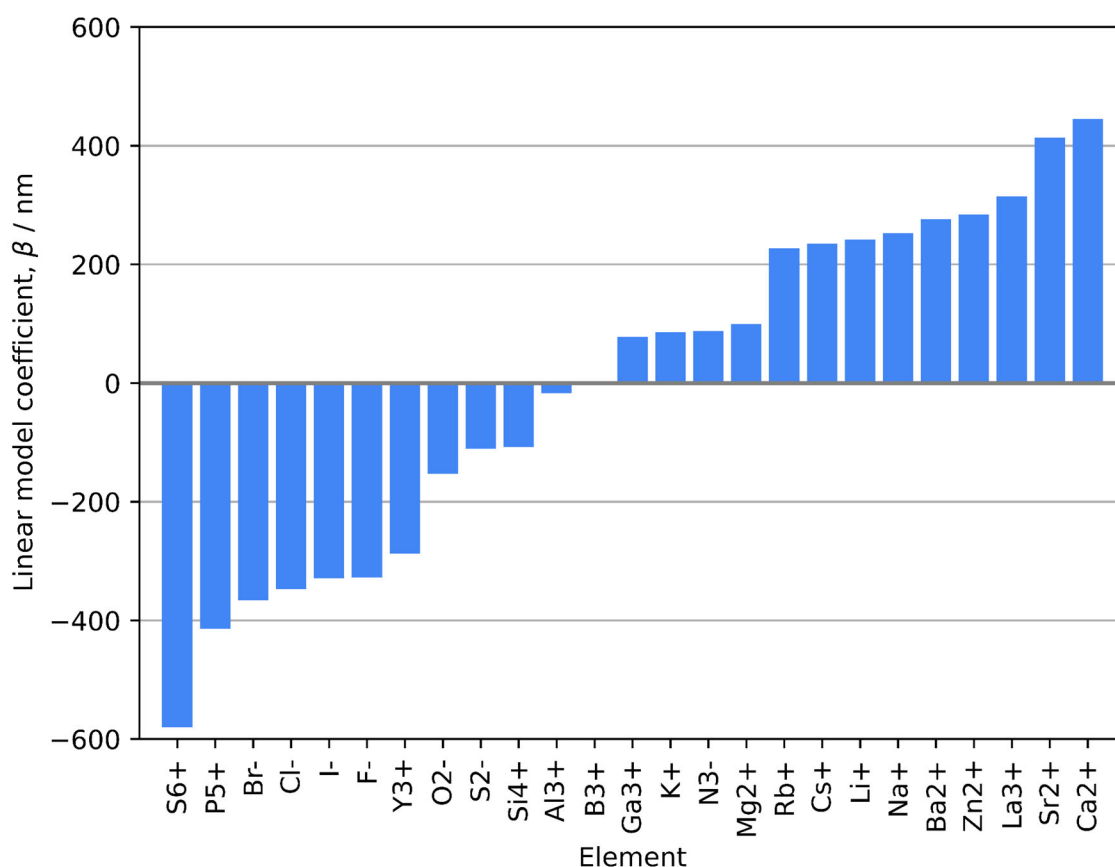


Figure S2. Linear model coefficients, β_j , for each element, as obtained from the ridge regression model. The elements on the x-axis are sorted in ascending order of their coefficient values.