Supporting Information to: An Active Learning Approach to Model Solid-Electrolyte Interphase Formation in Li – ion Batteries

*Mohammad Soleymanibrojeni Celso Ricardo Caldeira Rego Meysam Esmaeilpour Wolfgang Wenzel**

Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

*Corresponding author. Email Address: wolfgang.wenzel@kit.edu

Description

Supporting information to the above mentioned manuscript, which provides more insights to the context and clarify some concepts.

A Dirichlet sample, π_i as shown in Eq. (1), of 4 independent gamma distributions γ_i, of rate 1, and a shape parameter that is $\alpha_i = 1 + \alpha_\epsilon$ for the label, otherwise $\alpha_i = \alpha_\epsilon$ ($\alpha_\epsilon \ll 1$), such that [1]:

$$\pi_i = \frac{\gamma_i}{\sum_{c=1}^{C} \gamma_c}, \quad \text{where } \gamma_i \sim \text{Gamma}(\alpha_i, 1) \tag{1}$$

And, to have Gaussian likelihoods, the gamma-distributed γ_i is approximated with a log-normal distribution (Eq. 2) so that the transformed labels become:

$$\tilde{y}_i = \log \alpha_i - \tilde{\sigma}_i^2/2\,, \qquad \tilde{\sigma}_i^2 = \log(1/\alpha_i + 1) \tag{2}$$

Probability of observing actual organic SEI class, given predicted as organic SEI class:

$$P(AO|PO) = \frac{P(PO|AO) \times P(AO)}{P(PO|AO)\times P(AO)+P(PO|AE)\times P(AE)+P(PO|AI)\times P(AI)+P(PO|AU)\times P(AU)} = \frac{0.91\frac{944}{2988}}{0.91\frac{944}{2988} + 0.096\frac{197}{2988} + 0.007\frac{935}{2988} + 0.056\frac{912}{2988}} = 0.918 \tag{3}$$

In this equation, AO and PO are the actual class organic SEI and predicted class organic SEI, respectively, and similarly, AE, AI, and AU, are actual class empty, inorganic SEI and unfinished, respectively. The values of 859 and 944 correspond to the predicted and actual number of observations of class organic SEI, $2,988$ is the size of the test set, 197 is the size of class empty, and 935 is the number of observations of class inorganic SEI and 912 of class unfinished, respectively.

Supplementary figures

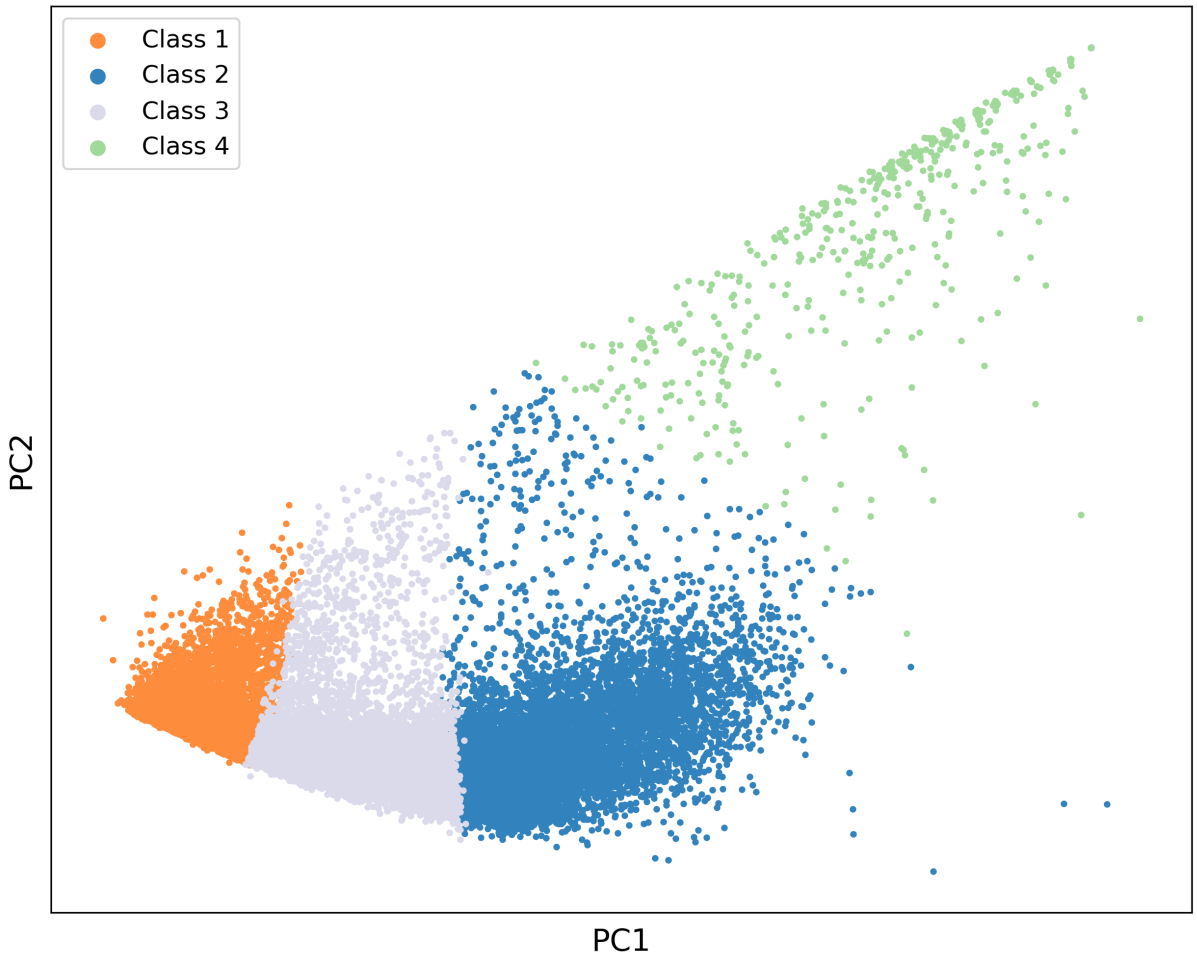
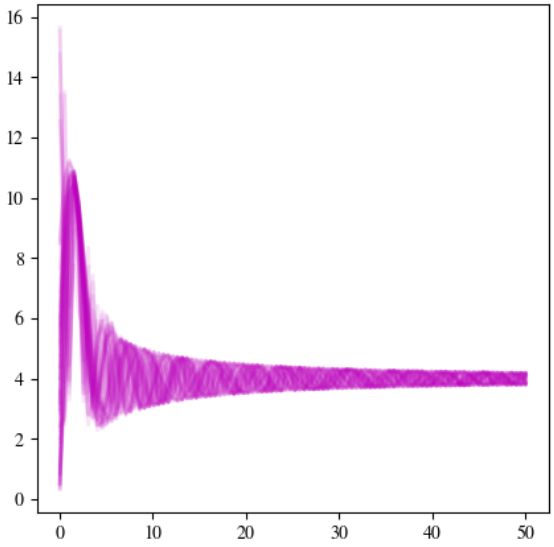
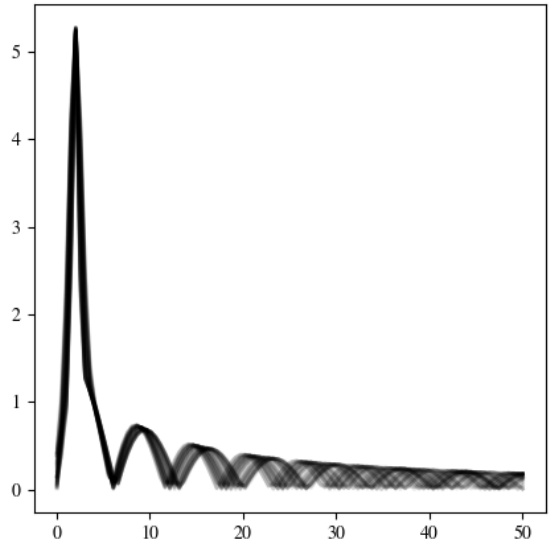


Figure 1: The effect of not using ICA on the descriptor. By performing K-mean directly on the descriptor, the labeling of the data is not deterministic, meaning there will be overlaps in properties between the labeled data points

(a)



(b)



(c)

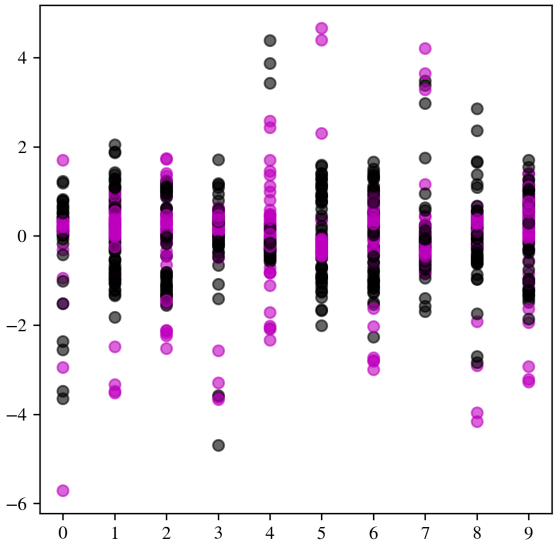
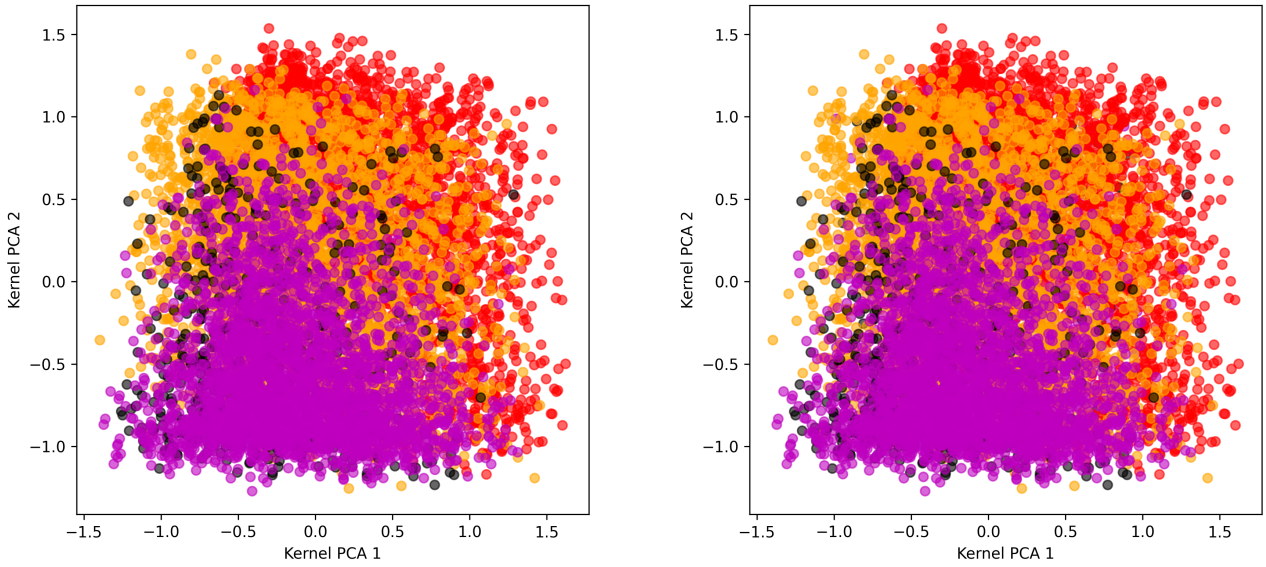


Figure 2: Insights about applying ICA to the descriptor. The two plots (a) and (b) are synthetic data that represent two types of model output classes, meaning organic SEI and empty, respectively. Each plot is made of 50 synthetic plots with 100 data points. After applying the ICA and reducing the components to 10, the output will be (c). The K-mean classifies the data points that have a size of 50×10

(a)

(b)



(c)

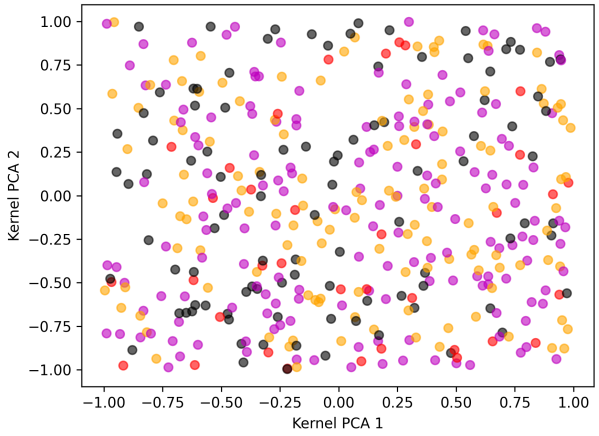


Figure 3: The sampling of the reaction barrier space (a) is the presentation of the first and second components of the 15 components in the transformed space from the initial dataset. (b) is a similar plot after 15 cycles of active learning. The difference is not perceptible with the eyes, but it is separately plotted in (c), which shows the sampling between (-1, 1). These data points were transformed back into the reaction barrier space using the inverse transformation using the kernel matrix. Each color represents a data point. red: inorganic SEI, magenta: organic SEI, orange: unfinished, black: empty

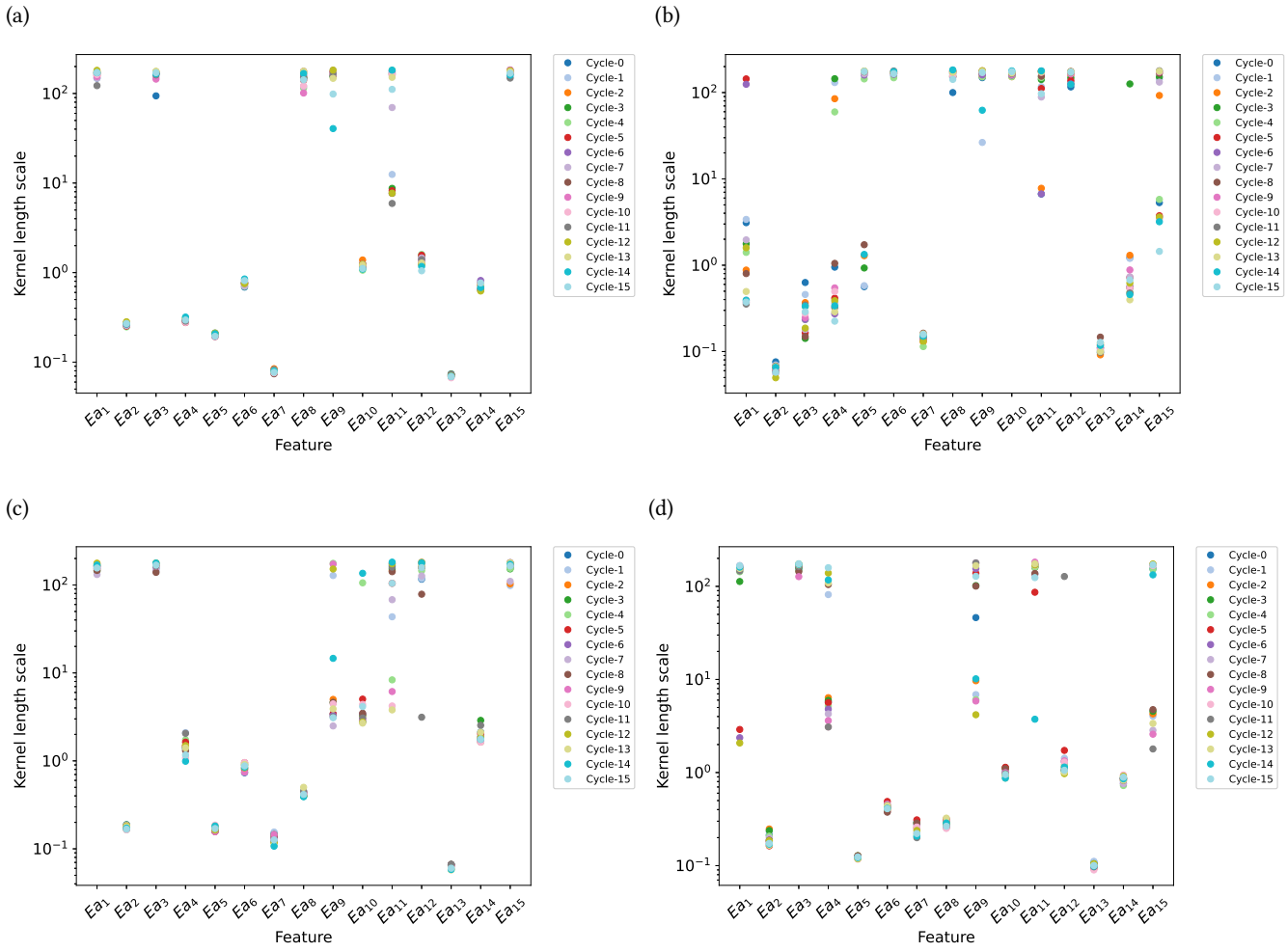
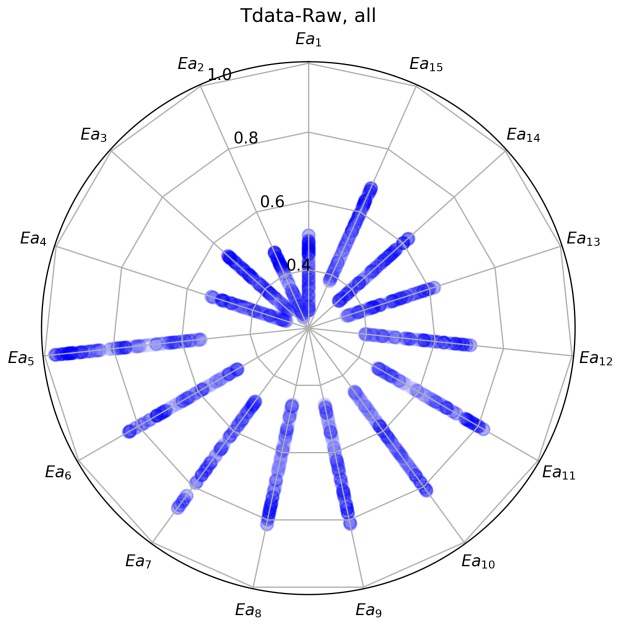


Figure 4: The kernel length scale for each class at the end of each active learning cycle

(a)



(b)

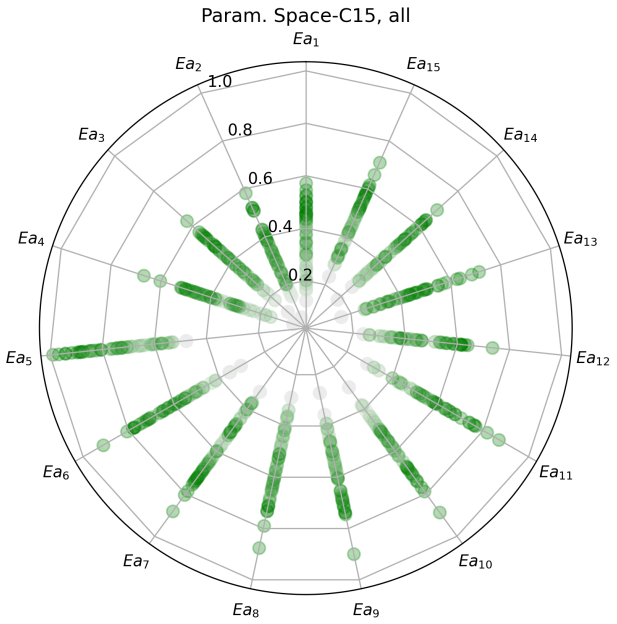


Figure 5: (a) Initial model parameters (reaction barriers) from the initial dataset without any processing (b) the parameter space sampled with the model after 15 cycles of active learning. The similarities and differences are clear. The model parameters have found their respective regions. This plots the parameters for all classes of model output, based on Table 1, main manuscript

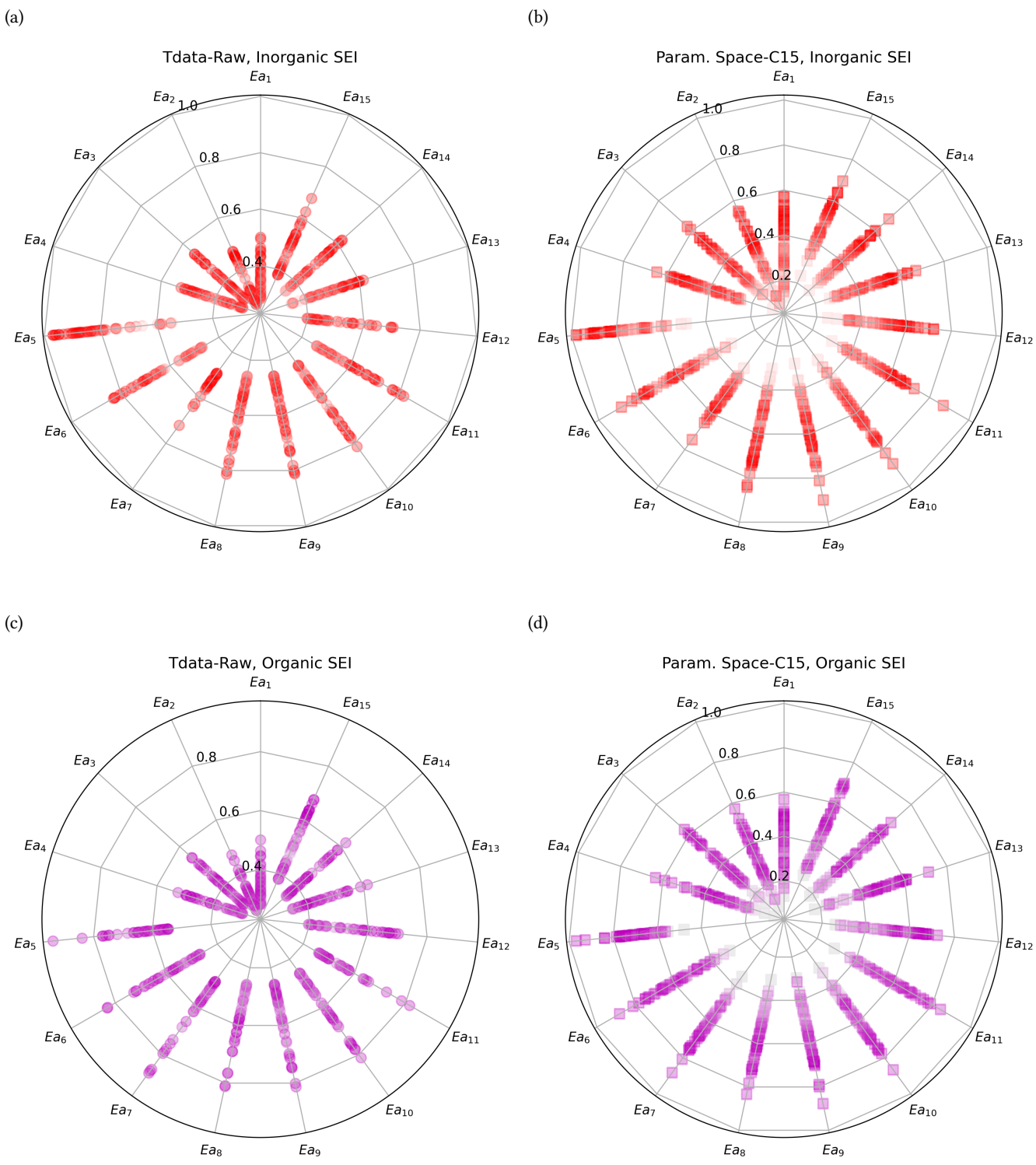


Figure 6: (a), (c) initial model parameters from the initial dataset without any processing, and (b), (d) the parameter space sampled with the model after 15 cycles of active learning, individually presented for inorganic SEI (a), (b), and organic SEI (c), (d) classes. The differences can be seen in how the samples from the trained model have a more clear distinction for the ranges of the parameters, based on Table 1, main manuscript

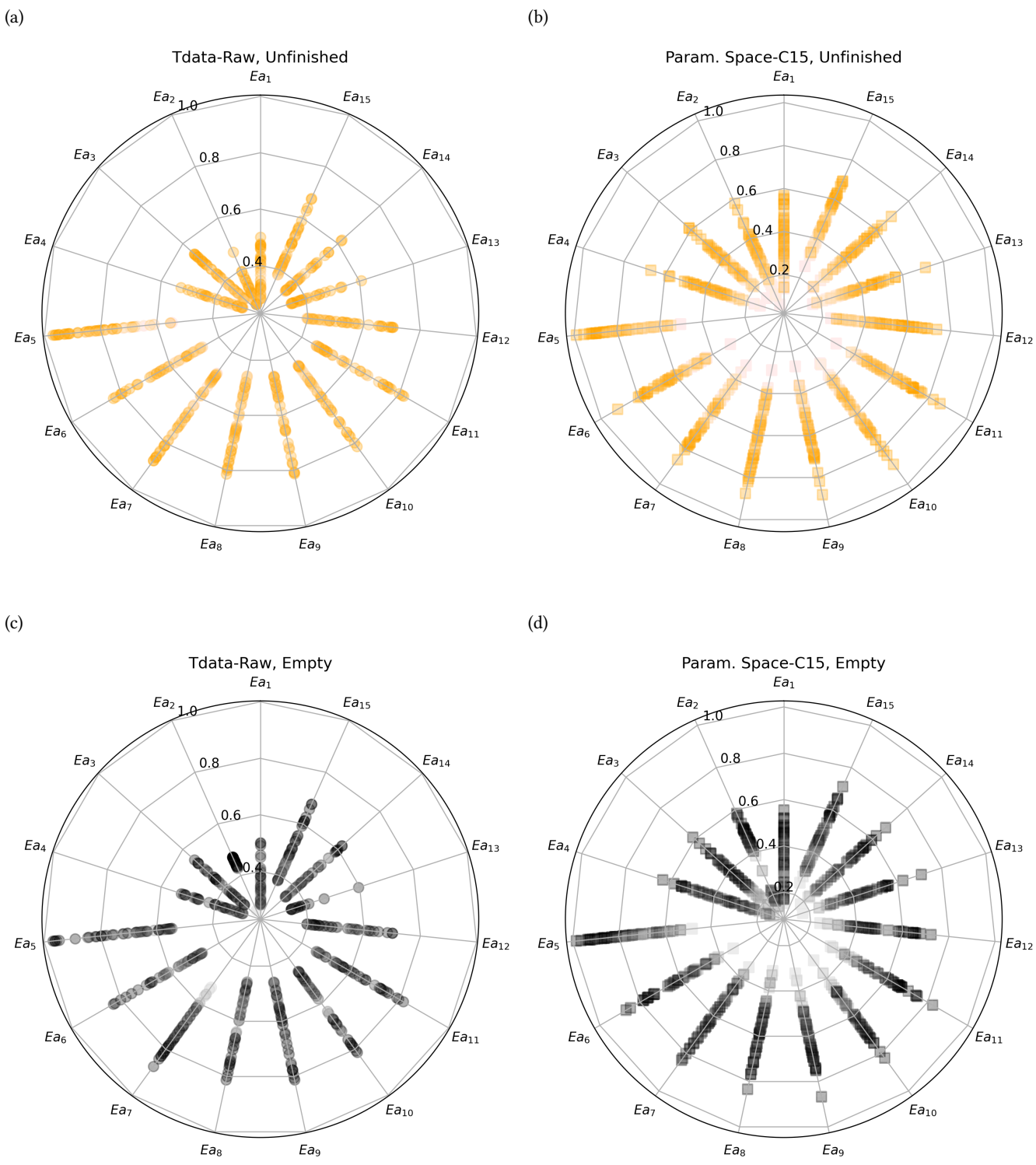


Figure 7: (a), (c) initial model parameters from the initial dataset without any processing, and (b), (d) the parameter space sampled with the model after 15 cycles of active learning, individually presented for unfinished (a), (b), and empty (c), (d) classes. The differences can be seen in how the samples from the trained model have a more clear distinction for the ranges of the parameters, based on Table 1, main manuscript

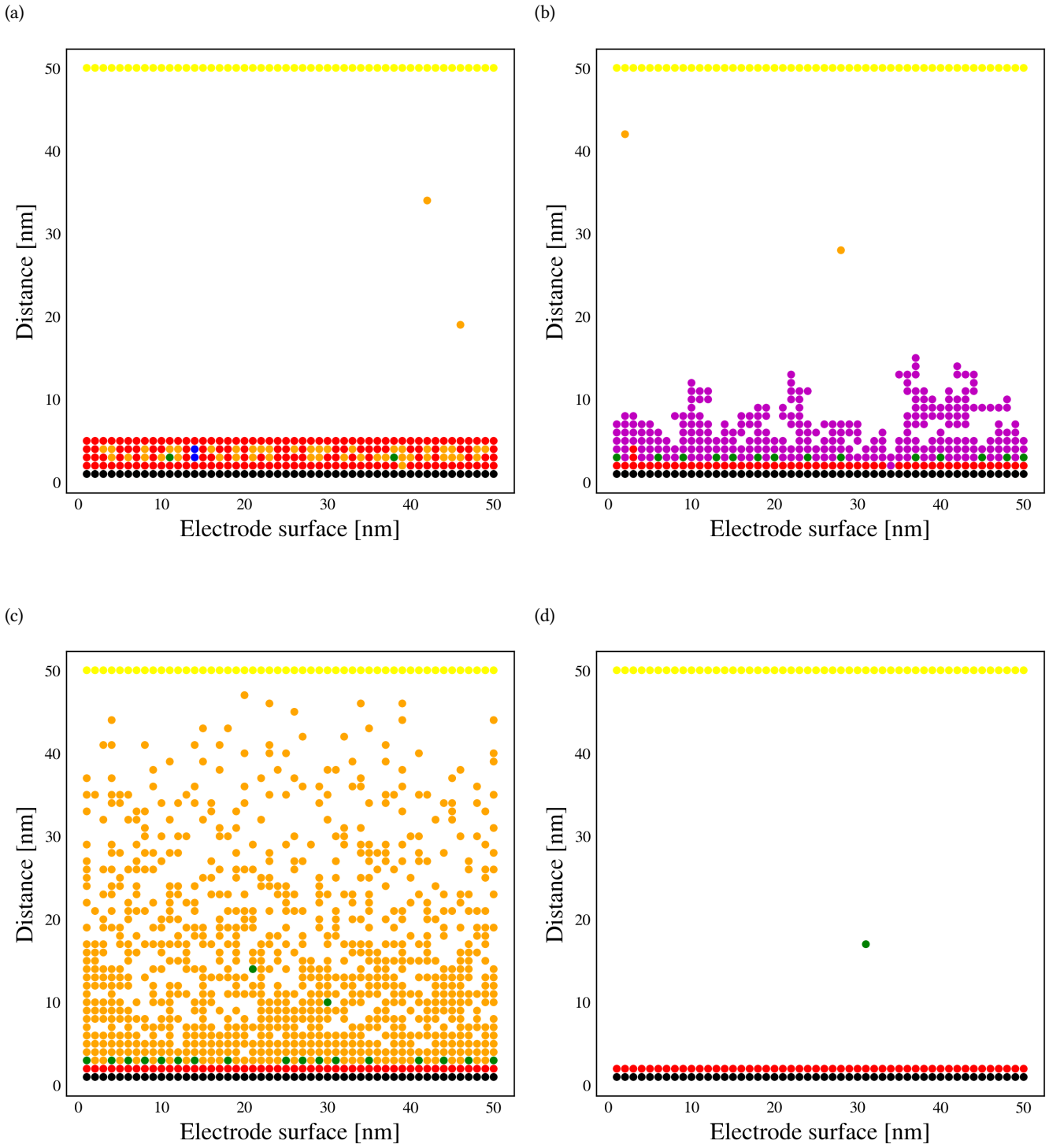


Figure 8: Using the trained model, the model parameters for each class of model output were obtained. By performing the kMC calculations using the generated model parameters, (a) shows the organic SEI, (b) the organic SEI, (c) the unfinished, and (d) the empty class, which matches the prediction of the trained model. Black: electrode, White: Li^+/EC^-, Red: Li_2CO_3, Magenta: organic SEI, Blue: $(Li_2EDC)_2$, Orange: Li_2EDC, Green: $Li^+.oEC^-$, Yellow: cell boundary

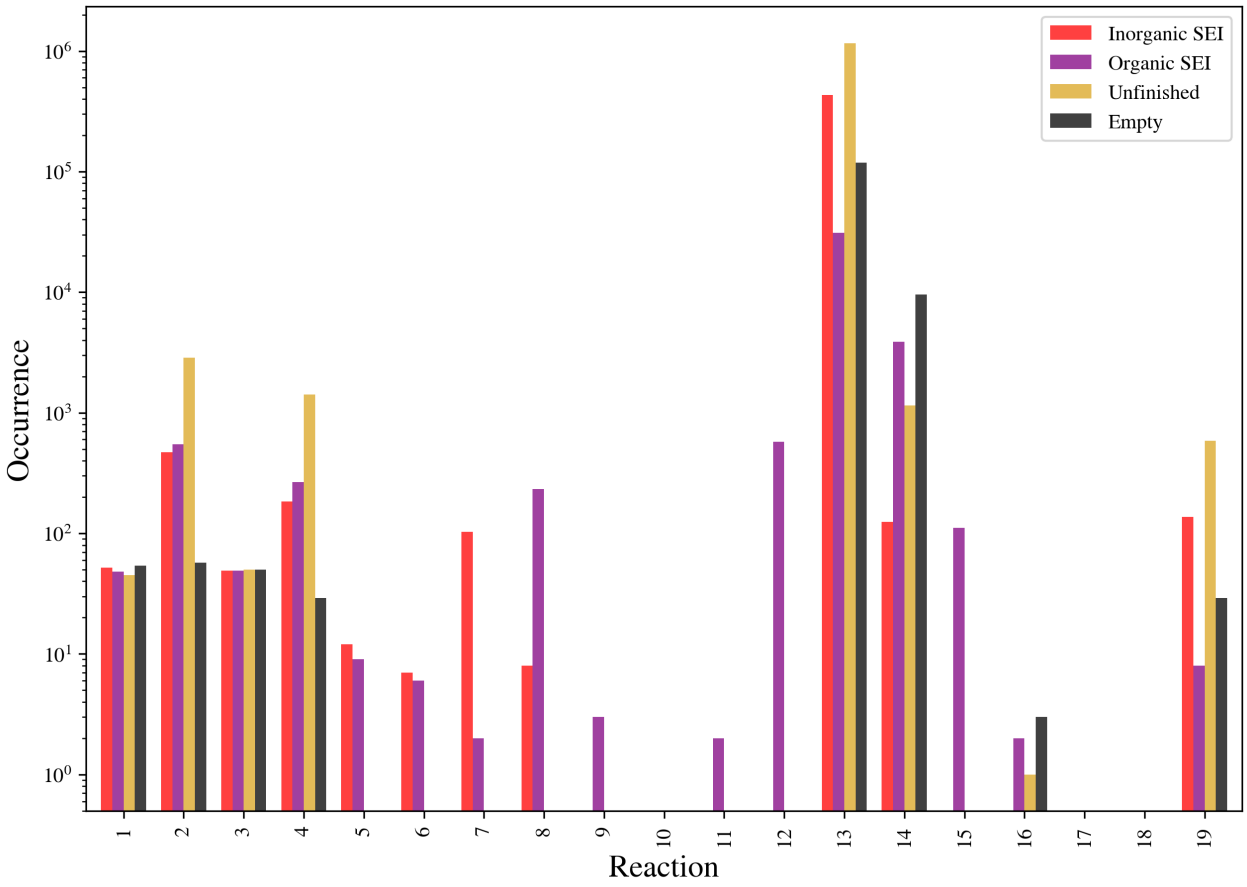


Figure 9: The occurrence of each reaction or event in the kMC calculations based on the generated model parameters of different SEI classes. Presented based on Table 1, main manuscript

References

[1] Dimitrios Milios, Raffaello Camoriano, Pietro Michiardi, Lorenzo Rosasco, and Maurizio Filippone. Dirichlet-based gaussian processes for large-scale calibrated classification. *Advances in Neural Information Processing Systems*, 31, 2018.