Internship in Chemistry

# Machine learning driven parameter investigation of the galvanostatic Kolbe reaction of aliphatic organic acids

Peter Fichtelmann

University of Leipzig
Master in Chemistry

Supervisor: Julia Westermayr

# Abstract

AI and machine learning are on the rise, but they are still in their infancy when it comes to chemistry. However, these tools offer new improvements for green chemistry to become economically feasible. In this proof-of-concept work, supervised machine learning and analysing explainable AI (XAI) methods were demonstrated on data from galvanostatic flow cell and batch Kolbe electrolysis experiments by Falk Harnisch and co-workers. 12 different types of regression models were screened to predict coulombic efficiency (CE), selectivity and total yield as output. The influence of input parameters was investigated and intuitive conclusions from a chemist's perspective were confirmed. Trade-offs between the outputs for different input parameters were revealed, where a reinforcement learning based reaction optimization was demonstrated.

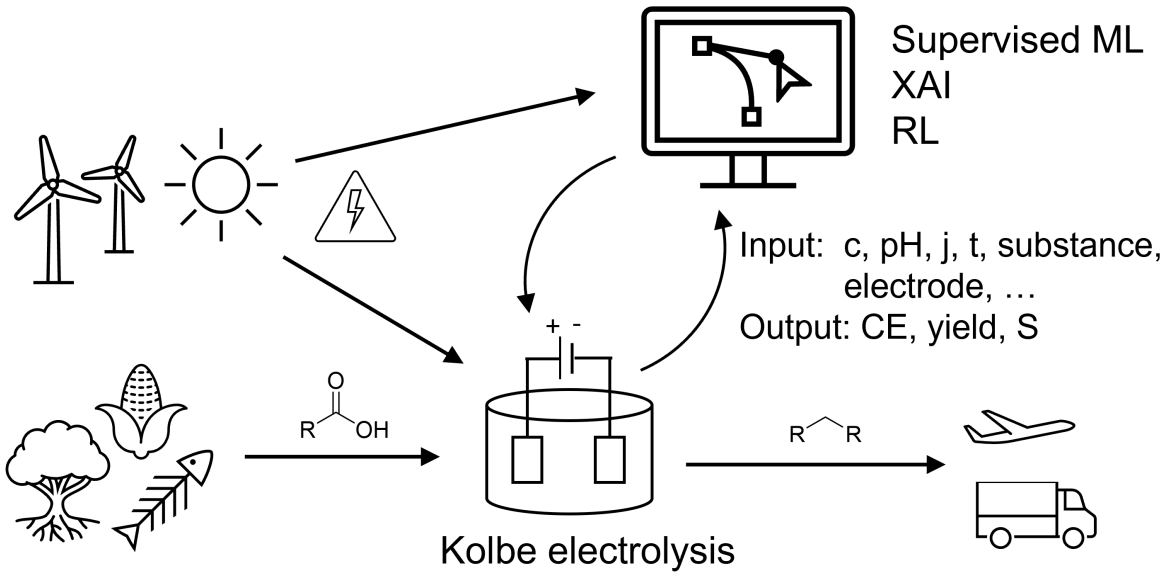



**Figure 1:** Graphical abstract.

# Abbreviations

| | |
|---|---|
| ALE | accumulated local effects |
| CE | coulombic efficiency |
| dielectric | dielectric constant |
| dH_ f | standard formation enthalpy |
| dH_ R | reaction enthalpy |
| ECFP | extended-connectivity fingerprint |
| FE | faradaic equivalents |
| ICE | individual conditional expectations |
| j | current density |
| knn | k-nearest-neighbour |
| log $K_{OW}$ | logarithm of the octanol-water partition coefficient |
| ML | machine learning |
| MW | molecular weight |
| MSE | mean squared error |
| PCA | principal component analysis |
| PDP | partial dependence plot |
| PPO | proximal policy optimization |
| RL | reinforcement learning |
| RMSE | root mean squared error |
| S | selectivity |
| SHAP | shapley additive explanations |
| sklearn | scikit-learn |
| Tb | boiling point |
| Tm | melting point |
| T_ flash | flash point |
| XAI | explainable artificial intelligence |
| Y | yield |

# Contents

# 1 Introduction

The burning of non-renewable fossil fuel resources is causing greenhouse gas emissions, air pollution, water and land contamination.[1] The transportation sector is the main consumer of fossil fuels,[2] as 90% of energy consumption was based on fossil oil in 2022.[3] Its replacement by more environmentally friendly substitutes, such as a) biomass or b) electrochemical power-to-x based hydrocarbons, is a high priority. In this context, electrochemistry can also fulfill important criteria of green chemistry, such as (i) mild reaction conditions at ambient temperature (ii) water as environmentally friendly solvent, (iii) electrons as immaterial agent instead of any oxidizing or reducing agents, and (iv) the minimization of waste products.[4] From an economic point of view, galvanostatic (constant current) operation is more interesting than potentiostatic (constant voltage) electrolysis.[5] Electrobiorefineries combine the concepts of biomass valorization and electrochemistry. Microbially produced medium-chain carboxylates have been discussed as promising renewable platform chemicals that can be electrochemically upgraded to fuels to power a bio-based economy.[6] A proof-of-concept of the combined microbial and electrochemical pathway to more climate-friendly hydrocarbons was demonstrated in 2017.[7] The applied electrochemical oxidation of organic acids is known as Kolbe electrolysis,[8] which was first reported 168 years ago in 1849.[9] Since then, many reaction parameters of galvanostatic batch experiments have been investigated, such as electrolyte,[8, 10] surfactants,[10] anode material,[11, 12], chain length of aliphatic acid starting material,[10, 13, 14] current density (j),[10, 13] Faradaic equivalents (FE, charge per concentration of starting material)[13], starting pH[8, 10] and temperature.[8] The number of reaction parameters and their complex interactions call for new emerging optimization and analysis tools like machine learning algorithms. To my knowledge, Kolbe electrolysis has not yet been studied with AI tools. Therefore, the aim of this project is to use machine learning methods to study the galvanostatic Kolbe electrolysis of aliphatic acids.

# 2 Theory

## 2.1 Kolbe reaction

The Kolbe reaction is one of the oldest and best known electroorganic syntheses.[8] Figure 2 illustrates the reaction pathway with n-hexanoic acid as the reactant.

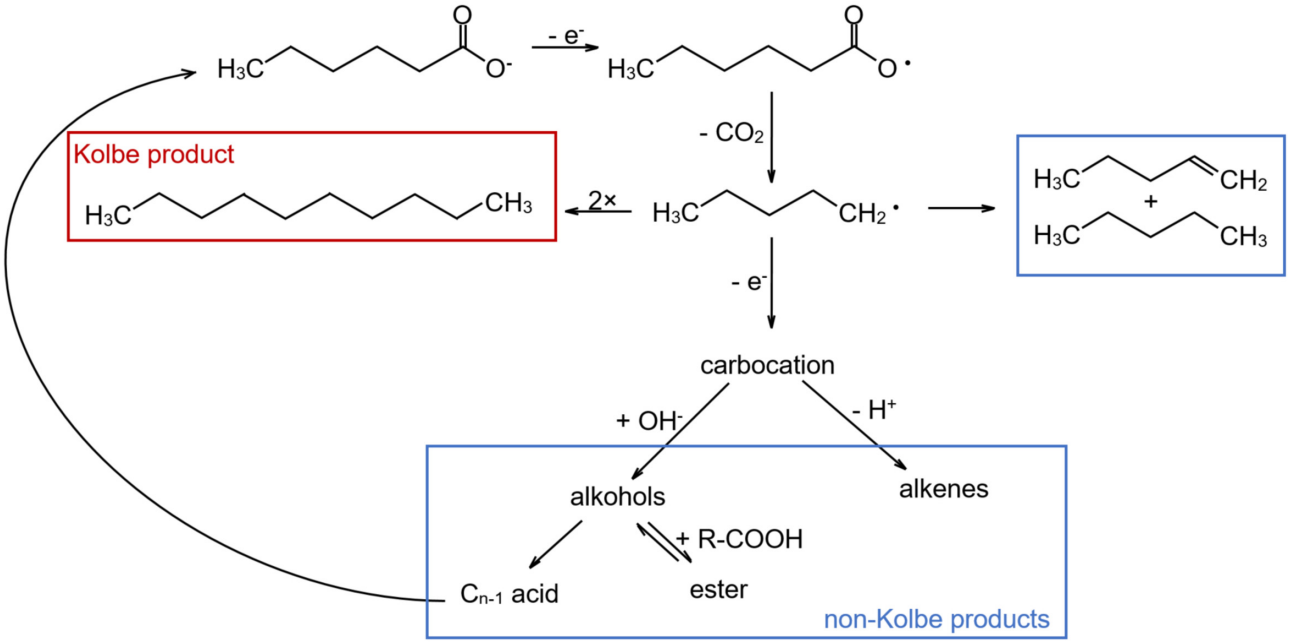



**Figure 2:** Schematic reaction mechanism of the Kolbe electrolysis of n-hexanoic acid to the Kolbe product (red) and side reactions to alkanes, alkenes, esters and alcohols (blue). Reprinted with permission from ref. [11]. Copyright CC BY-NC 4.0 DEED.

First, the organic acid is deprotonated and oxidized at the anode surface. After the one-electron oxidation, the formed radical is decarboxylated. Then, the alkyl radical can react to different products depending on the organic acid, reactant concentration, solvent, electrolyte type and concentration, electrode composition and surface. Dimerization of two alkyl radicals results in the long chained Kolbe product. If the radical disproportionate, short-chain alkanes and alkenes are formed. Further oxidation of the radical yields a carbocation, followed by deprotonation to alkenes or nucleophilic hydroxide addition to alcohols. Subsequently, alcohols can react with the starting acid to form esters. Alternatively, oxidation can occur to an one carbon shorter acid which can react again as a reactant.[8] The Kolbe product and the other liquid hydrocarbons formed are suitable as fuel.[7]

Two different general setups of reaction management have been studied.[2, 11, 13] In a batch system, the reaction mixture is stirred in the same vessel until the end of the reaction. On the other hand, the reaction solution is pumped with a flow rate through the reactor in a flow cell.

## 2.2 Basics of machine learning

In machine learning a model is trained on input data to produce a desired output. A schematic is shown in Figure 3.

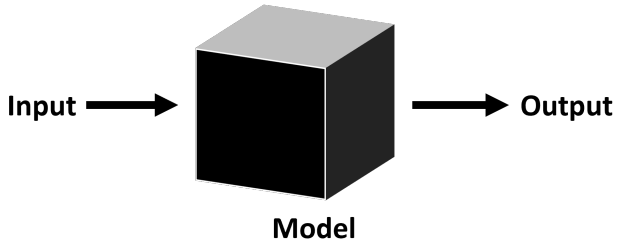



**Figure 3:** General schematic of the training of a machine learning model. The model is represented as black box as it is usual difficult to achieve model transparency.

In unsupervised learning the model's algorithms never get in touch with the output variables, but only the input parameters (called features).[15] In contrast, supervised learning uses labeled features with their corresponding output parameters. It is common practice to split the data into train and test sets to evaluate the trained model on unseen test data.[15] If the algorithm uses distance measurements, standardization or normalization of the input features is recommended so that large absolute values do not dominate the learning of the model. Most models have various hyperparameters (e.g. number of layers in a neural network or number of trees in a random forest) which can be tuned to improve performance. Another type of machine learning is reinforcement learning (RL). A schematic of the use of RL for reaction optimization is shown in Figure 4.

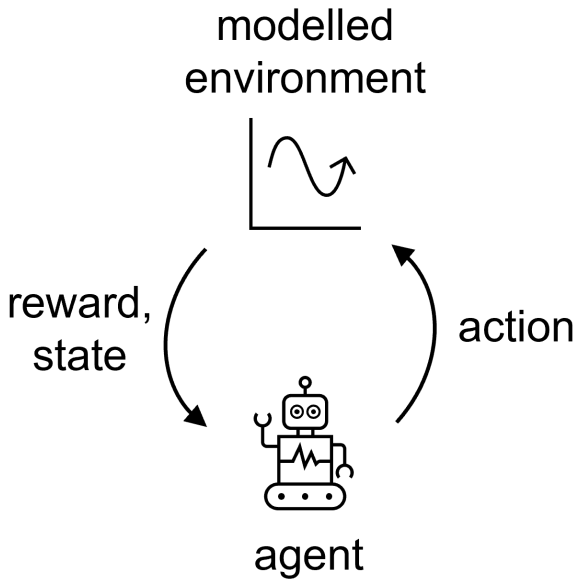



**Figure 4:** Schematic of a reinforcement learning (RL) reaction optimization.

Instead of data, an environment is required. The model's agent interacts with it by taking actions following certain strategies (called policies). After an action, the agent receives the

current state of the environment and a reward, followed by the next iteration of this cycle. The agent can be programmed with a value function to maximize the long-term reward. In combination with electrochemistry, the experimental data is useful to create a supervised machine learning model that acts as the environment in a RL reaction optimization.

## 2.3 Model-agnostic XAI methods

Data can be used to gain knowledge. In addition, a black-box machine learning model itself becomes a source of knowledge.[16] XAI methods are useful to extract relevant patterns captured by the model. Model-agnostic XAI methods are suitable for any model (as opposed to model-specific methods). These methods are divided into global and local methods. The latter provide insight into what led to individual predictions while global methods characterize the average of the model and its general mechanisms.[16]
The following methods have been investigated in this project and will be briefly introduced:

- Local
  - SHapley Additive exPlanations (SHAP)
  - Individual Conditional Explanations (ICE)
- Global
  - Permutation Feature Importance
  - SHapley Additive exPlanations (SHAP) summary plot
  - Partial Dependence Plot (PDP)
  - Accumulated Local Effects (ALE)

### 2.3.1 Permutation feature importance

Permutation feature importance is a global XAI method that measures the increase in predicted error after permuting the values of input features.[16] The steps of the calculation are shown in Figure 5.

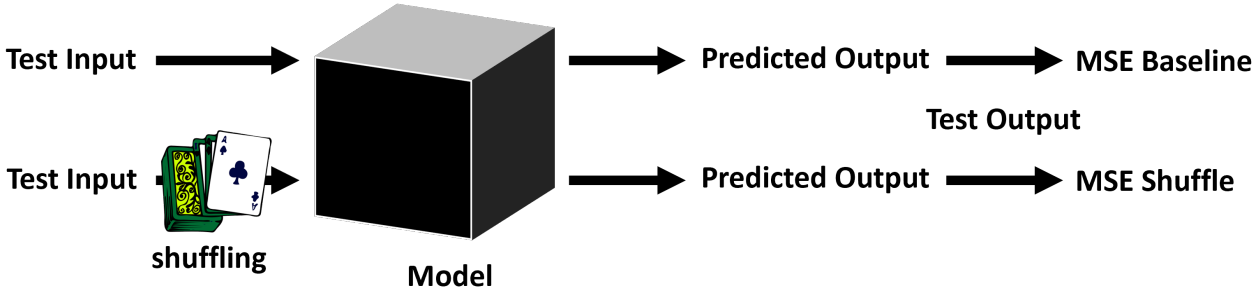



**Figure 5:** Steps of a permutation feature importance calculation using MSE as evaluation metric.

First, the MSE (or any other evaluation metric) of the test set is calculated from the deviation of the predicted output from the test output as the MSE Baseline. Test data is used instead of training data because training errors are meaningless.[16] In addition, the test input feature's values are shuffled (permuted) and the predicted output is used to calculate the MSE Shuffle. The difference between these two MSEs of a feature is the permutation feature importance of a feature. Normalization by dividing by the summed permutation feature importance of all features allows for comparability.

### 2.3.2 Shapley additive explanations (SHAP)

Shapley additive explanations (SHAP) are a local method where individual results can be summarized to yield global interpretations. SHAP is based on optimal Shapley values from coalitional game theory. Each feature of a data point is considered as a player in a game where the prediction is the payout and the Shapley value describes the contribution of a feature to the payout.[16] A brief description of the mathematical background can be found in Christoph Molnar's free book *Interpretable Machine Learning*[16]

In contrast to permutation feature importance, SHAP summary plots reveal more details, such as heterogeneous behavior or positive and negative impacts of features on the predicted outcome.

### 2.3.3 Partial dependence (PDP) and individual conditional explanation (ICE) plots

Partial dependence plots (PDPs) show the global effect of one or two features on the predicted outcome,[17] revealing linear, monotonic or more complex relationships between features and the output parameter. The PDP of a regression is defined as:[16]

$$\hat{f}_S(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_S, x_C^{(i)}) \tag{1}$$

where $x_S$ are the features of interest and $X_C$ the other features. For each data point, $X_C$ is held constant while $x_S$ is varied. The average of the changed output variable for each $x_S$ is the PDP.

PDPs hide heterogeneity caused by feature interactions. Accordingly, the local equivalent for individual data points, called the individual conditional expectation (ICE) plot, refines the interpretation.[18] ICE plots show (one line per data point) how the examined features of that data point affect the prediction. Thus, the average of all ICE plots of a feature is the PDP of that feature.

### 2.3.4 Accumulated local effects (ALE) plots

The previously introduced methods produce inaccurate results when features are correlated due to the consideration of unlikely combinations of feature values. Similar to PDP, but unbiased and faster to compute, accumulated local effect plots describe the global relationship of one to two features and the outcome. As shown in Figure 6, only small intervals close to each data point are used to compute the difference in prediction instead of averages.[16] The differences are then accumulated and centered to yield the ALE curve.

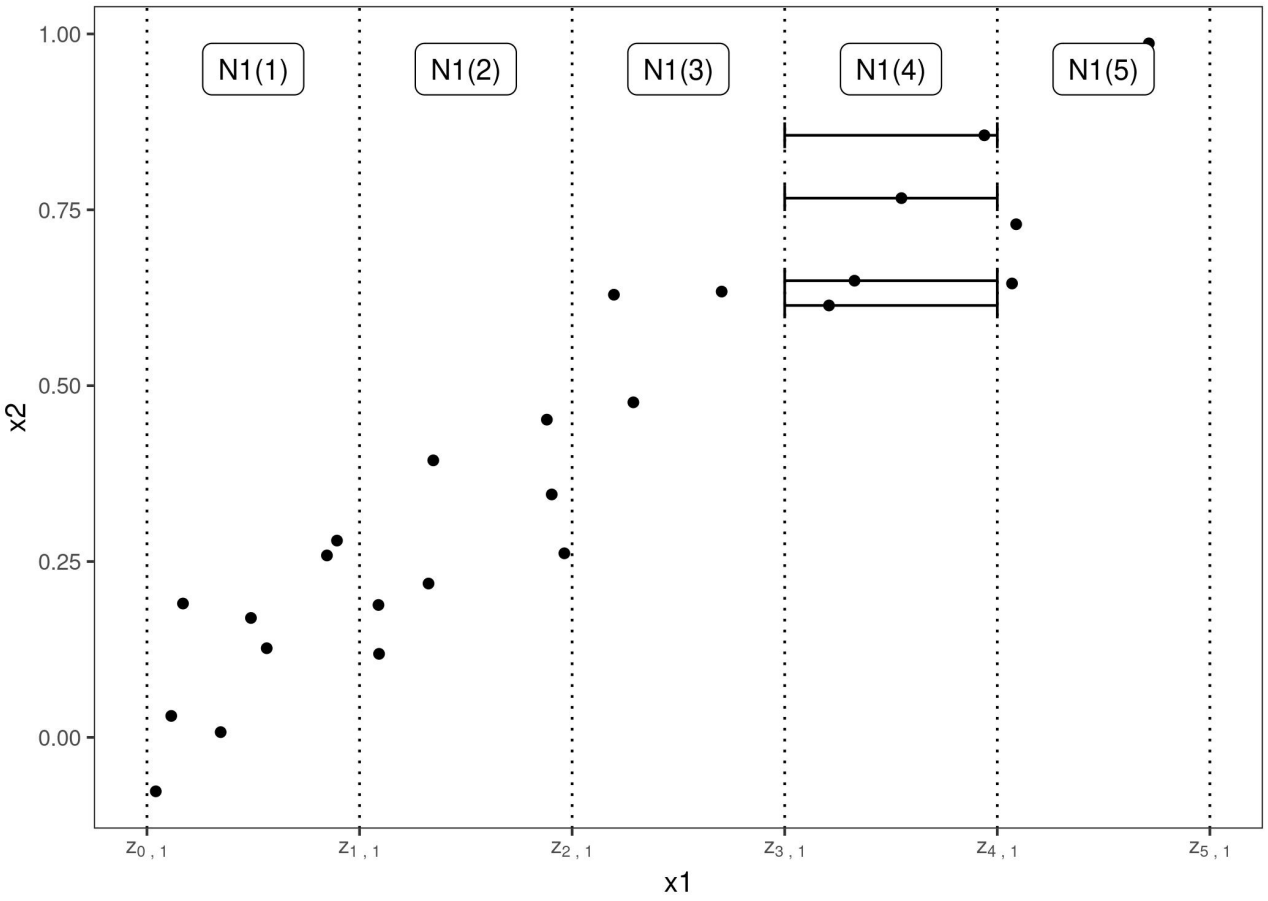



**Figure 6:** Calculation of ALE for feature x1 correlated to feature x2. Reprinted with permission from [16]. Copyright CC BY-NC-SA 4.0 DEED.

# 3 Method

## 3.1 Package versions

Important packages and the version used are listed below:

- Numpy Version 1.26.0
- Pandas Version 2.1.1
- Scikit-Learn Version 1.3.2

## 3.2 Input feature calculation and selection

Recently, the reaction parameters current density (j), flow rate, and Faraday equivalents (FE) of the Kolbe reaction of hexanoic acid in a flow cell have been studied.[2] The latter variable describes how much charge was supplied to theoretically (CE=100%) convert a certain partition of substrate. It is defined as:

$$FE = \frac{j \cdot A \cdot t}{F \cdot c \cdot V} \tag{2}$$

where $j$ is the current density, $A$ the geometric surface of the anode, $t$ the reaction time, $F$ the Faraday constant, $c$ the total concentration of the organic acids and $V$ the volume of the anolyte. FE is strongly correlated with the reaction time and the latter was chosen as feature due to its easier interpretability. In addition, pH data at the start of the experiments were available and added as the fourth input feature.

Considering the batch data set, most of the available experimental data used a two chamber setup. Data from one chamber experiments were added due to only small variations. In contrast to the flow cell data, more reaction parameters were available due to previous investigations.[11, 13] In addition to hexanoic acid, butanoic and octanoic acid have been tested on their own and in a mixture of all three acids. Therefore, three features were used to describe the concentrations of the highest(main), second highest (2nd) and lowest concentrated (3rd) acid. Furthermore, substance features were added. Molecular weight (MW) [g/mol], the logarithm of the octanol-water partition coefficient (log $K_{OW}$), boiling point (Tb) [°C], melting point (Tm) [°C], flash point (T_ flash) [°C], density (at 20°C) [g/cm$^3$], viscosity (at 20-25°C) [mPa·s], refraction index (at 20°C) and p$K_a$ were acquired from PubChem.[19] Dielectric constant and dipole moment [D][20] as well as the standard enthalpy of formation (dH_ f)[21] were researched. The latter was used to calculate the standard reaction enthalpy (dH_ R) of the following general Kolbe reaction.

$$2\,\mathrm{R{-}COOH\,(l)} \longrightarrow 2\,\mathrm{CO_2(g)} + \mathrm{R{-}R\,(l)} + \mathrm{H_2(g)} \tag{3}$$

In addition, the smiles notation of the main compound was used to calculate the extended-connectivity fingerprint (ECFP) which was decomposed into one dimension by a principal component analysis (PCA). Besides substance features, two electrode descriptors were used to include the effects of different electrode materials, which have been recently investigated.[11] The onset potential and tafel slope of the electrode at pH7 with 0.25 M $Na_2SO_4$ were determined as previously done by Chen *et al.*[22] The onset potential describes the required potential to oxidize the acid on the electrode surface and the tafel slope is an electro-kinetic indicator of the electron transfer.[22]

## 3.3 Output variable selection

Coulombic efficiency (CE), selectivity (S) and total yield (Y) were chosen as output variables due to data availability and representation of different desirable reaction characteristics. In the flow cell, CE is related to the fuel, whereby selectivity (without gases) and yield are based on the Kolbe product decane. The outcomes of the batch experimental data are related to the sum of alkanes formed.

## 3.4 Data preprocessing

The following data points which deviated significantly from their replicates were removed from the flow cell data set: 21065, 21090, 21091, 21097, 21092, 21080. When CE was selected as the output variable, additional data was removed due to missing CE data: 21025, 21026, 21027, 21030, 21036.
The entire data sets were randomly split into 80% training data (train set) and 20% testing data (test set) using sklearns train_ test_ split. Then, the input features were standardized with sklearns StandardScaler and normalized with sklearns MinMaxScaler to allow attribution of all features.

## 3.5 Code availability

The code is available on GitHub: `https://github.com/peter1999de/Machine-learning-driven-parameter-investigation-of-the-Kolbe-reaction/`

# 4 Results and Discussion

## 4.1 Model screening

12 different types of sklearn models were screened. Hyperparameters tuning was performed by bayesian optimization using the MSE of the train set. Due to the significant influence of the splitting, the preprocessing, hyperparameter optimization and training were conducted for 10 different random states (42-51) for each kind of model. The mean values of the evaluation metrics MSE, RMSE and $R^2$ and their corresponding standard deviations are listed in the appendix (flow-cell: Table A1, Table A2, Table A3; batch: Table A4, Table A5, Table A6).
The simple non-linear models k-nearest-neighbour (knn) and decision tree as well as the more complex ensembles of trees (gradient tree boosting and random forest) performed best on the flow cell data set (4 features). Multiple linear regression showed higher errors as only linear patterns could be captured in the data. More complex models, such as xDNN, gaussian progress and kernel ridge, are more prone to overfit the limited amount of data. The RMSE of the model screening on the flow data with S as the outcome is shown in Figure 7.
knn regression also performed best on the batch data (25 features) e.g. see in Figure 8. Replicates can decrease the errors of knn because it uses the closest neighbouring data points for prediction. However,of all the models trained with the means of the batch data instead of replicates knn models were still the best (see Table A7).
The champion model (depending on random state and hyperparameters) of the best type of models was selected for further analysis. Predicted and test outcomes (CE and selectivity) were plotted in Figure 9 (flow cell) and Figure 10 (batch) to spot anomalies.
With respect to the flow cell, the CE-knn-regression-model follows the linear trend, but is prone to under- and overestimate CEs below and above 40%, respectively. Test data is sparse for CEs above 45% and more corresponding data points would improve the evaluation there. Regarding the S-gradient-tree-boosting-model, the test data is better distributed and was predicted with only up to 1% deviation. The better performance can be attributed to less noise of the batch data of S in contrast to CE and Y.
The scatter plots of the batch data also show a clear linear trend and an uneven distribution of the data. Particularly for S, limited data below 60% existed.

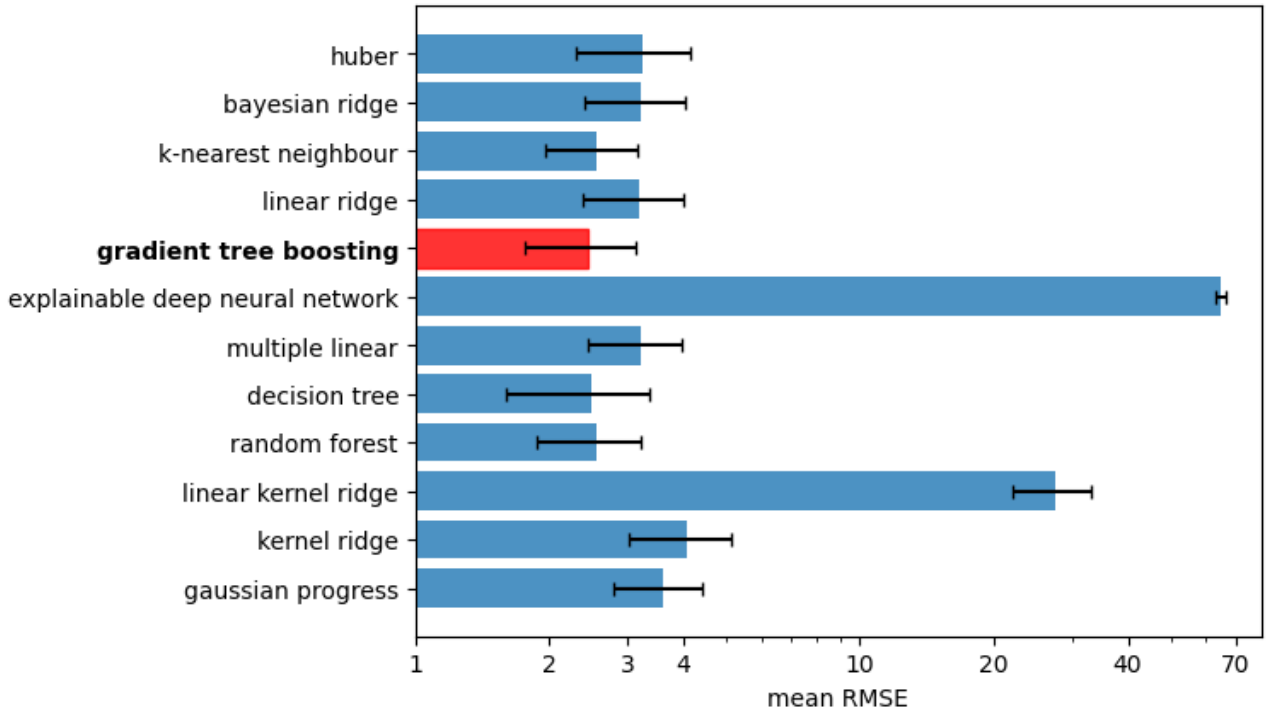



**Figure 7:** RMSE of the model screening (10 different random states) of the flow data with the selectivity of decane (without gases) as the output parameter. Error bars are the standard deviation.

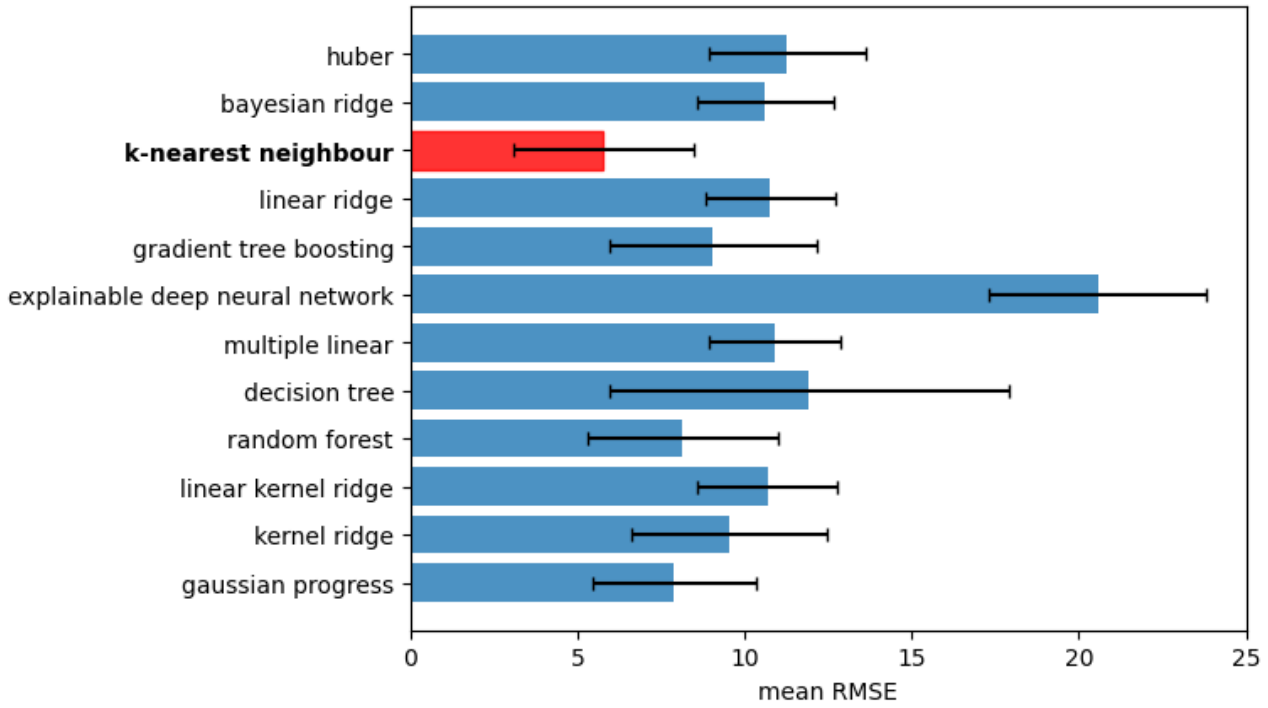
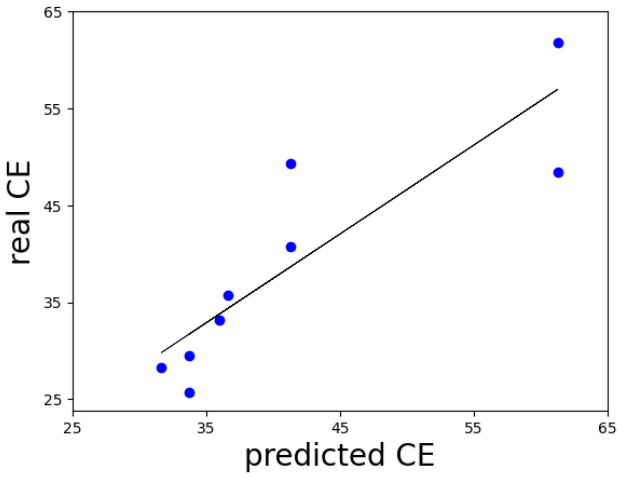
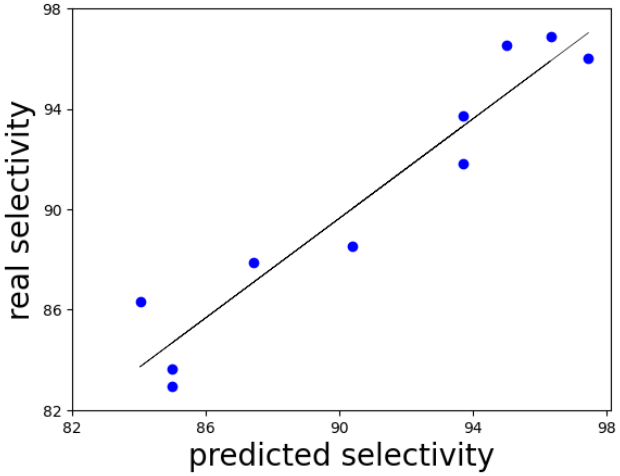



**Figure 8:** RMSE of the model screening (10 different random states) of the batch data with the selectivity of alkanes as the output parameter. Error bars are the standard deviation.

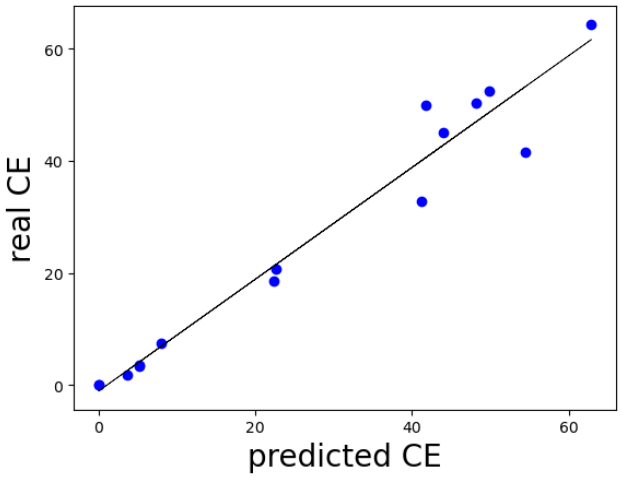



**(a)** CE

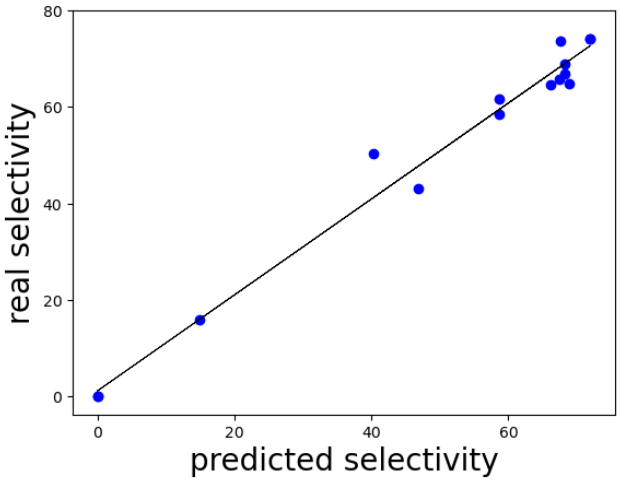



**(b)** Selectivity

**Figure 9:** Flow cell: predicted vs. test set outputs of the champion models for CE and S (blue dots). The black line is the corresponding linear fit.




**(a)** CE



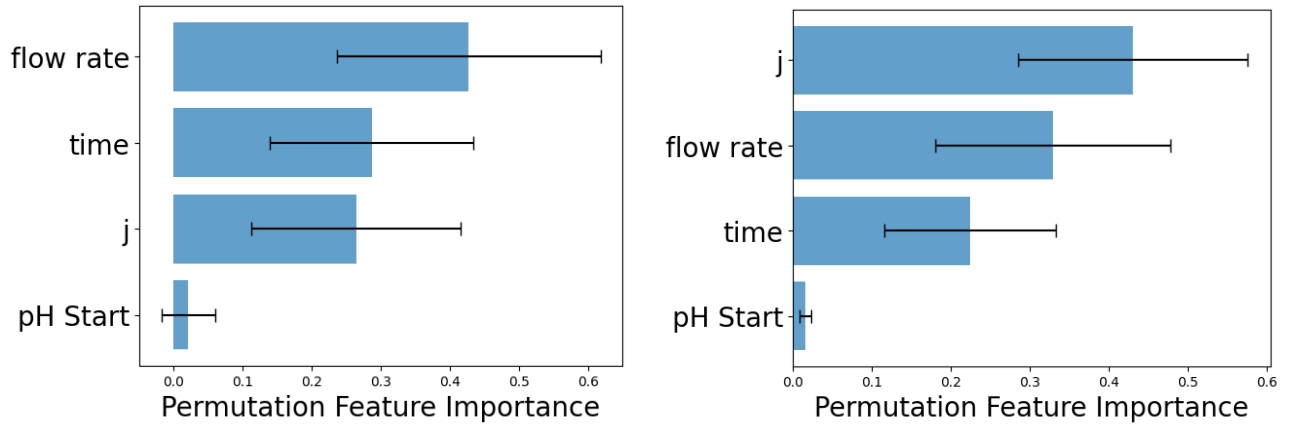

**(b)** S

**Figure 10:** Batch: predicted vs. test set outputs of the champion models for CE and S. The black line is the corresponding linear fit.

## 4.2 Model agnostic XAI explainability

### 4.2.1 XAI flow cell data set

In this project, feature oriented methods were chosen to gain insight into the influence of the adjustable input features. This is valuable for further reaction optimization. However, these methods lack human level explanations of how and why the model came to its decisions.[23] Permutation feature importance and SHAP were used to examine feature importance. Plots of the former are shown in Figure 11 (CE and S) and in Figure A2a (Y).

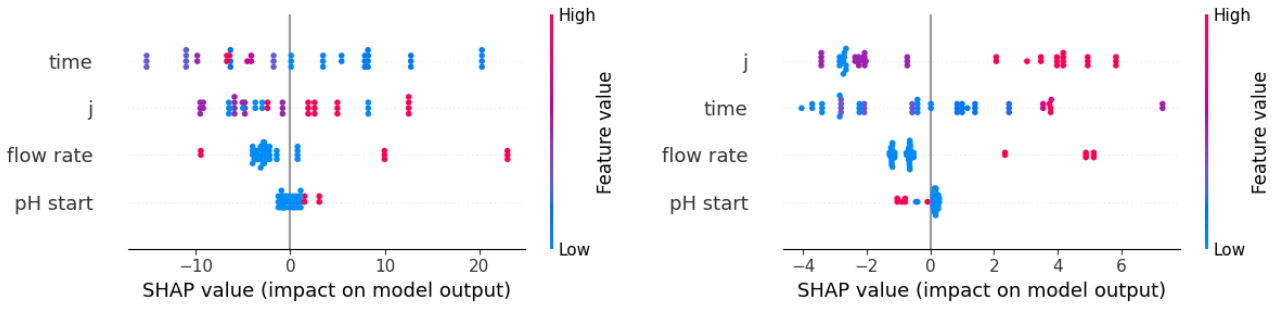



**(a)** CE **(b)** S

**Figure 11:** Flow cell: average permutation feature importance (500 permutations) on the test set of the champion models of CE and S. Error bars are the standard deviation of the permutations.

Permutation feature importance calculations were performed on the test set based on the MSE, which does not contain any meaningful information on the train set. Furthermore, it was repeated 500 times to achieve a stable level of reproducibility of the results. Flow rate, time and current density j were the most important features for the champion models of CE, S and Y. However, their order changed depending on the output variable. Thus, the most important feature was flow rate for CE, j for S and time for Y. This interpretation should be treated with caution due to the significant standard deviation of the different permutations. However, it is clear that the starting pH plays a minor role in all three models.

Permutation feature importance does not provide information about whether features have a positive or negative impact on the prediction. In addition, individual predictions are not explained due to the global nature of the explanation. A representation of both issues is provided by SHAP. Feature effects can be visualized as forces which contribute differently to each prediction. Feature importance is shown in SHAP summary plots in Figure 12 (CE and S) and Figure A2b (Y), where each point is the Shapley value of a feature with a heat map of its feature value. Since SHAP is based on output-related Shapley values, the entire data set, including train and test set, was examined.




**(a)** CE **(b)** S

**Figure 12:** Flow cell: SHAP summary plots of the champion models of CE and S.

The order of the three most important features flow rate, j and time changed for all models in contrast to the permutation feature importance. However, the starting pH was still the feature with the lowest impact. A low pH of 7 had hardly any effect on the models, but higher pH values of 11 showed an increase in CE, a decrease in S and mixed influence on Y predictions. Regarding the flow rate, 1 L/min reduced the predicted variable while 2 L/min generated only for S a clear positive behavior. Less reaction time resulted in higher CE and lower S as well as Y and more reaction time vice versa. The same trend was observed for j for all three champion models investigated. An increase in j was followed by a higher prediction, although it is more blurred for CE and Y. In addition, the less amount of data at high feature values of pH, flow rate and time must be mentioned. Therefore, the corresponding interpretations should be considered as assumptions rather than reliable facts.

Thus, partial dependence plots (PDPs) were used as another method to explain trends. Individual conditional expectation (ICE) plots of the train set are also shown in Figure 13 (CE), Figure 14 (S) and Figure A3 (Y) to reveal heterogeneity.

Regarding the CE-champion model, a marginal positive trend of the pH could be confirmed. However, a CE above 60% at pH 7 decreased to pH 11. A similar but more drastic was observed for the flow rate. Nevertheless,there were only two different values for these characteristics in the data set, which limits the reliability of the interpretation. Non-monotonic trends were observed for j and time, with a minimum at 300 mA/cm$^2$ and a maximum around 8 min as well as a minimum at 24 min, respectively. ICE curves with high CE at low j or time also followed divergent trends.

In comparison to the CE- and Y-champion models, the S-champion model showed a less divergent behavior of single instances. Most of the ICE lines followed the average PDP trend. Starting pH was negative and flow rate as well as j were positive monotonic, confirming the SHAP interpretation. In the detail of j, the slope increased between 300 and 450 mA/cm$^2$. The feature time exhibited a more complex pattern of general increase, but with several turning points.

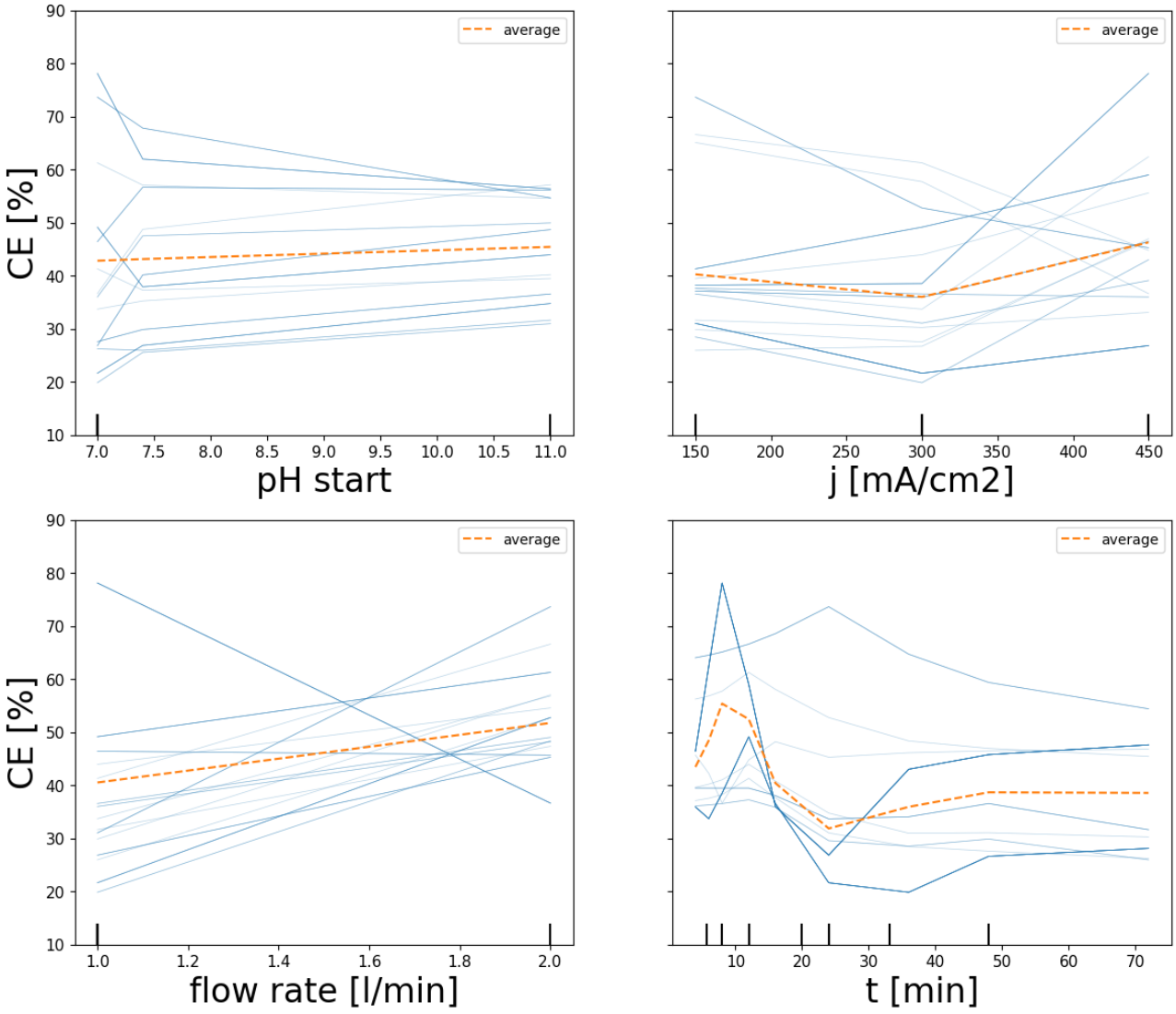



**Figure 13:** Flow cell - CE: PDP (orange dotted line) and ICE plots (blue solid lines) of the champion model of CE.

The Y champion model showed a clearer positive monotony for time, with the slope decreasing as reaction time increased. However, the other features had more heterogeneous ICE curves which hamper clear interpretations.

All hitherto investigated XAI methods have lower accuracy when features are correlated due to the integration of meaningless combinations of feature values.[16] Correlation of the features was not expected due to domain knowledge. However, Figure 15 shows the correlation matrix of all input features and output variables. Correlations cannot be excluded because of the biased selection of performing experiments which promised positive results.

Notably, time appears to be correlated with pH start and j. Accumulated local effect (ALE) plots deal better with correlated features because only small feature changes are calculated. ALE plots confirm the results of the PDPs and are shown in Figure A4, Figure A5 and Figure A6a. Two-dimensional ALE plots allow visualization of secondary interaction effects between features that add up to the main effects. The interaction between time and j was

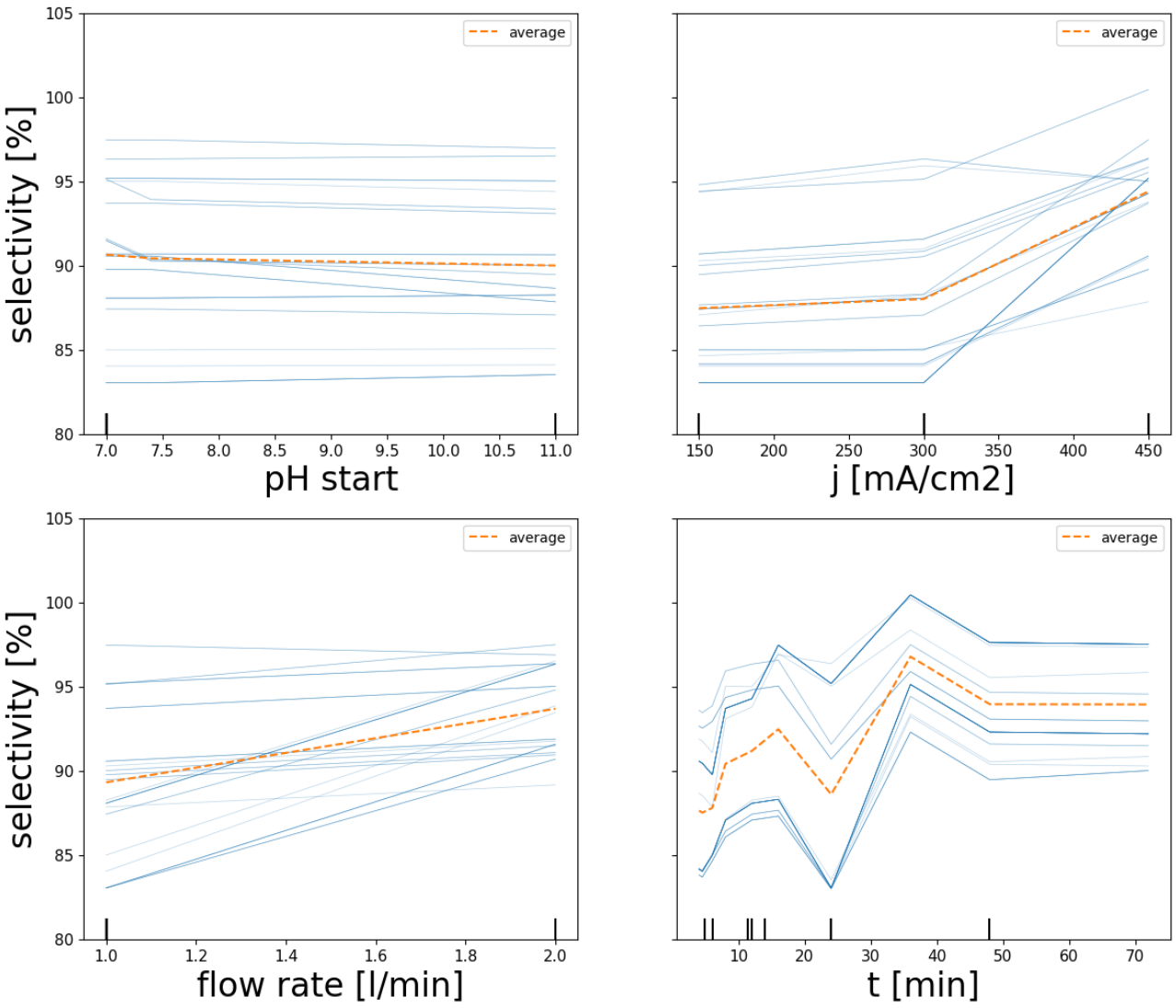



**Figure 14:** Flow cell - S: PDP (orange dotted line) and ICE (blue solid lines) plots of the champion model of S.

examined in Figure 16 (CE and S) and Figure A6b (Y).

All outcome variables showed marginal effects for j less than 350 mA/cm$^2$. Regarding j above 350 mA/cm$^2$, a negative and positive interaction was observed for less and more reaction time, respectively.

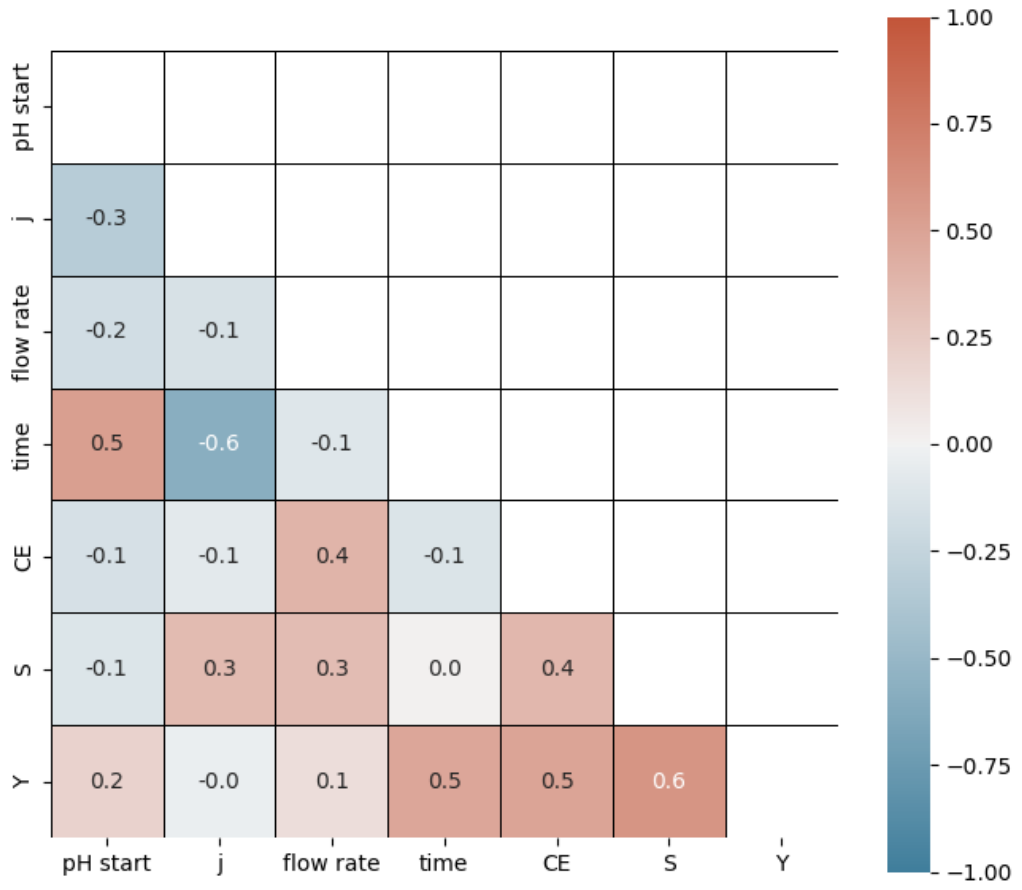
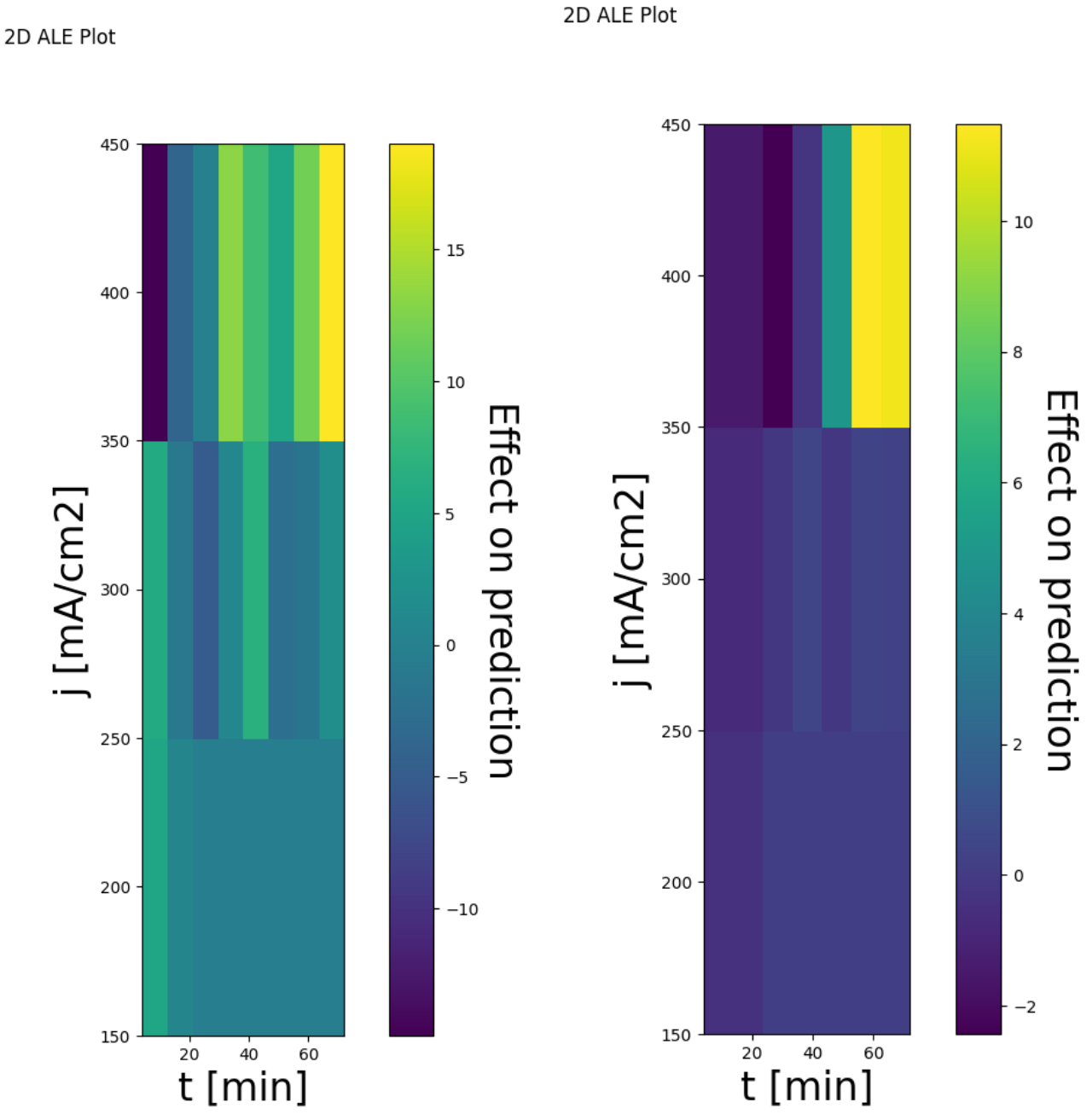



**Figure 15:** Correlation matrix of the input features and output variables. Positive correlations are colored red and negative ones blue.



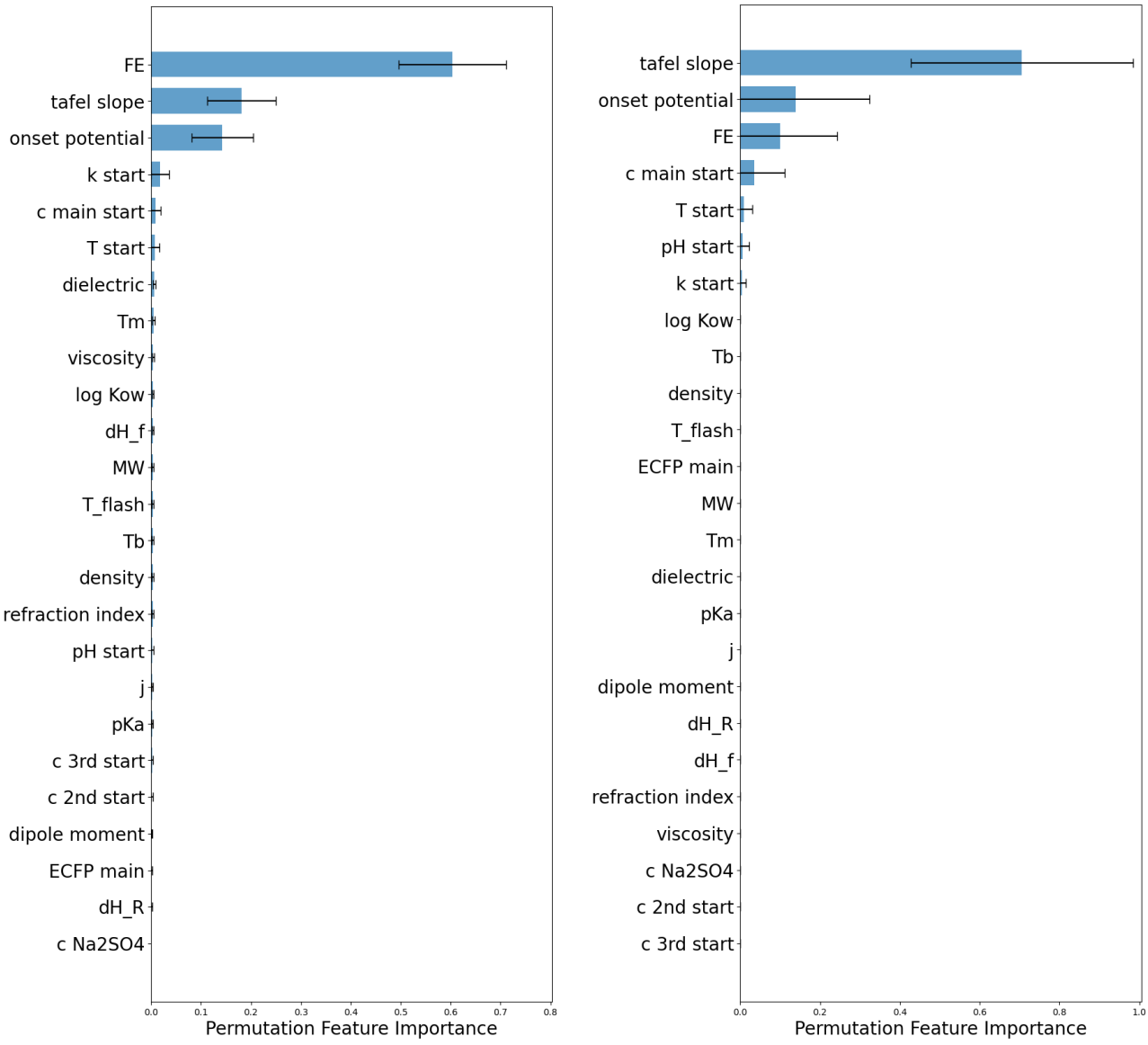

**(a)** CE

**(b)** S

**Figure 16:** Flow cell: 2D ALE plots of the champion models of CE and S. Only interaction effects are shown, not the main effects of the features.

### 4.2.2 XAI batch data set

The batch data set was refined with 21 more features than the flow cell data set. Therefore, discussing each effect of each feature on the model is considered too tedious for this project. Instead, the focus is on the main effects.
Permutation feature importance plots are shown in Figure 17 (CE and S) and Figure A7a (Y).




**(a)** CE **(b)** S

**Figure 17:** Batch: average permutation feature importance (500 permutations) on the test set of the champion models of CE and S. The error bars are the standard deviation of the permutations.

The theoretical conversion rate based on the applied electron charge (FE) and the electrode characteristics tafel slope and onset potential had a significant influence on all three champion models. In particular, the tafel slope had a greater effect than the onset potential. In addition, the starting conductivity k was found to have some influence on the models

investigated.

Further examination of the feature importance was carried out using SHAP summary plots in Figure 18 (CE), Figure 19 (S) and Figure A7b (Y). This confirmed the importance of FE, tafel slope and onset potential. The investigated electrodes without platinum on the surface had a higher tafel slope and lower onset potential which contributed to lower CE, S and Y. In addition, a trade-off with respect to FE was observed. Increased FE correlated with significantly lower CE and S, but also with increased Y and vice versa. Slight importance could be ascribed to k start. However, only for S a trend was observed where S decreased with increasing k start. Furthermore, the concentration of the most concentrated acid c main played a role, but no clear trend was observed. Substance describing features showed less importance in the plots. Nevertheless, a correlation matrix in the appendix in Figure A11 revealed their strong correlation. This is attributed to the lower importance of each individual feature. Selection of these feature could be beneficial for further machine learning investigations. Due to correlation, ALE plots are shown instead of PDP plots in Figure 20 (CE), Figure 21 (S) and Figure A12 (Y).

A general increasing (onset potential) and decreasing (FE and tafel slope) trend was observed for the three most important features. However, a shift from the trend occurred between pure platinum and platinized titanium, where a slight change in CE was caused. In addition, the maximum of FE was reached at 0.5. The substance features showed two different behaviors: 1) a negative slope, starting from hexanoic acid to butanoic and octanoic acid for dielectric constant, dipole moment and reaction enthalpy dH_R; 2) a slight increase from butanoic to hexanoic acid, followed by a drastic decrease to octanoic acid for all other substance features (log Kow, MW, Tb, viscosity, density, T_flash, Tm, refraction index, dHf and pKa). Notably, the pH-ALE plot had a significant maximum at 7.4, but there was little data for pH above 7.3.

Regarding the S-model, S was significantly reduced at higher pH values. Furthermore, the negative trend of FE in SHAP was confirmed. Notably, no influence of the substance features was observed. Instead, S of the starting conductivity k decreased until 41 mS/cm, followed by a slight increase to a stable level. A similar trend, but opposite behavior in the first interval, was shown by the ALE plot of the CE- and Y-champion models. The latter model had generally similar trends to the CE model.

So far, the influence of different acidic aubstrates has been neglected because the majority of the data focused on hexanoic acid alone (72%). However, the batch data set also included butanoic (4%) and octanoic (4%) acid as well as mixtures of the three acids (20%). SHAP summary plots of the compounds and features without hexanoic acid alone are shown in the appendix in Figure A8, Figure A9 and Figure A10. Most of the features showed a negative effect on all three output parameters for butanoic and octanoic acid. The effect was more significant for butanoic acid. However, FE, tafel slope and onset potential had a positive

impact on CE for octanoic acid. In this case only platinized titanium electrodes were used. Therefore, tafel slope and onset potential had only one feature value each. Nevertheless, they influenced the decision of the models especially on S. Notably, the dielectric constant and melting point (Tm) had a more significant influence on CE and Y. Regarding the mixture of acids, the SHAP values of less important features were between butanoic and octanoic acid. In contrast, more important features showed an ambiguous effect.

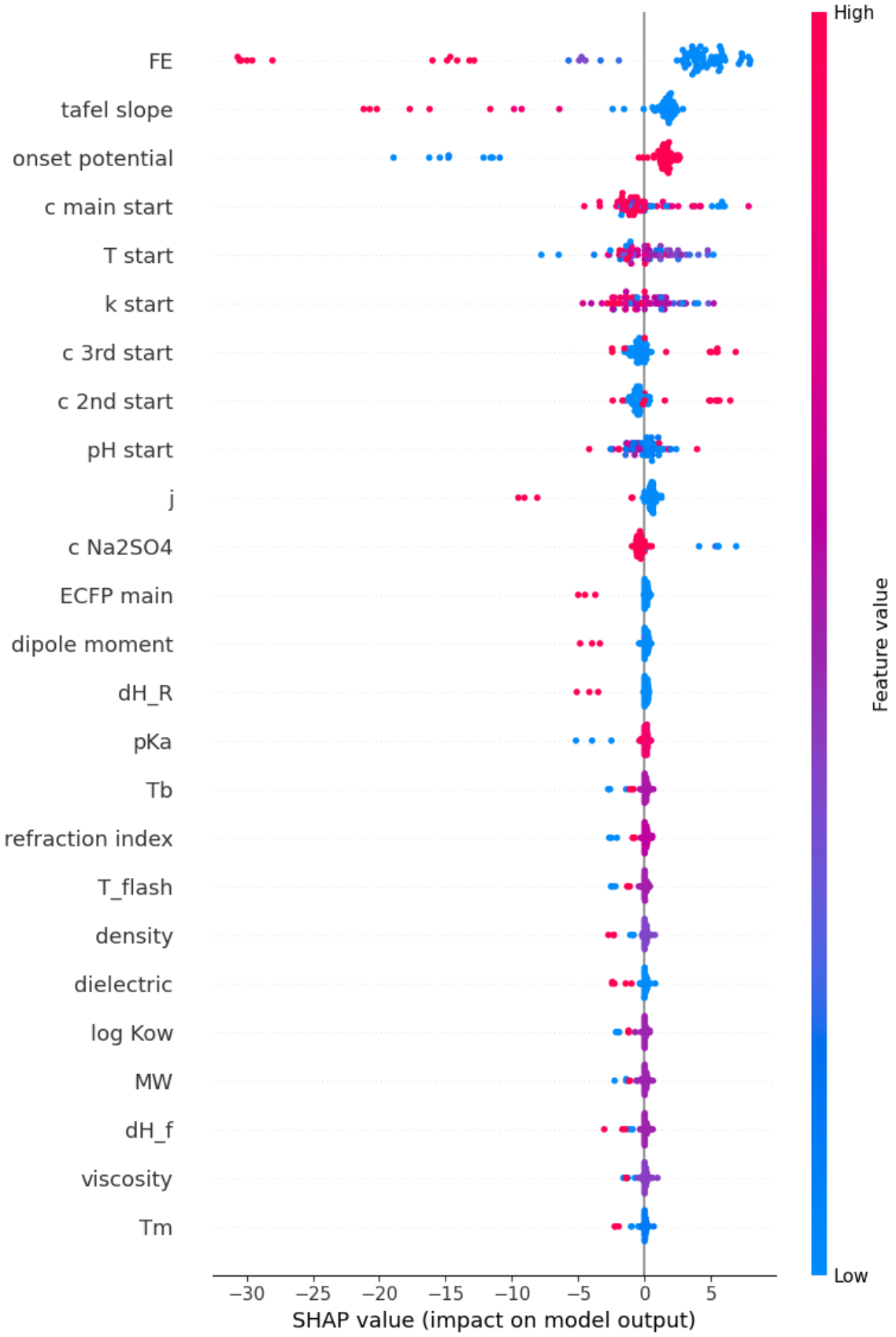



**Figure 18:** Batch: SHAP summary plot of the champion model of CE.

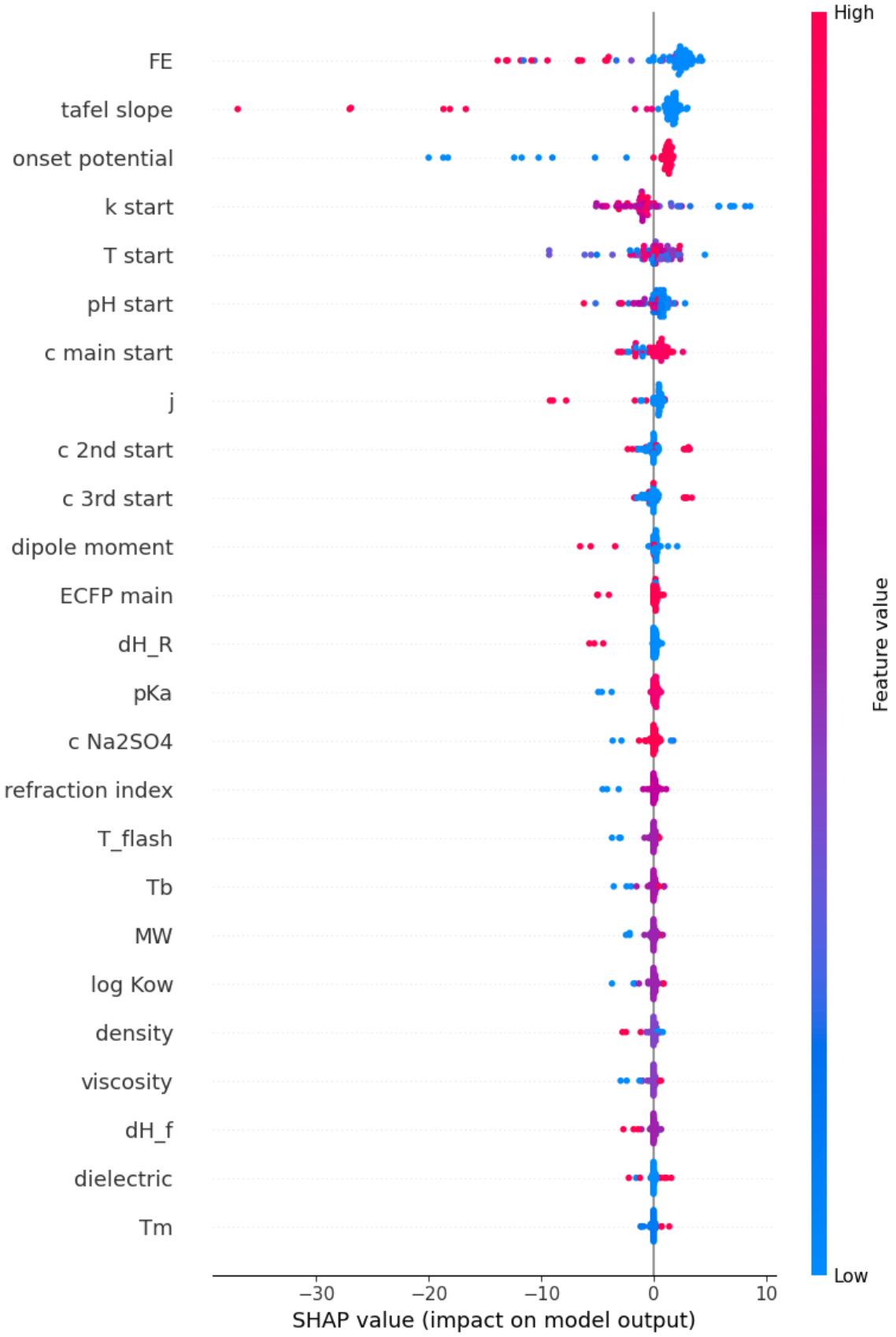



**Figure 19:** Batch: SHAP summary plot of the champion model of S.

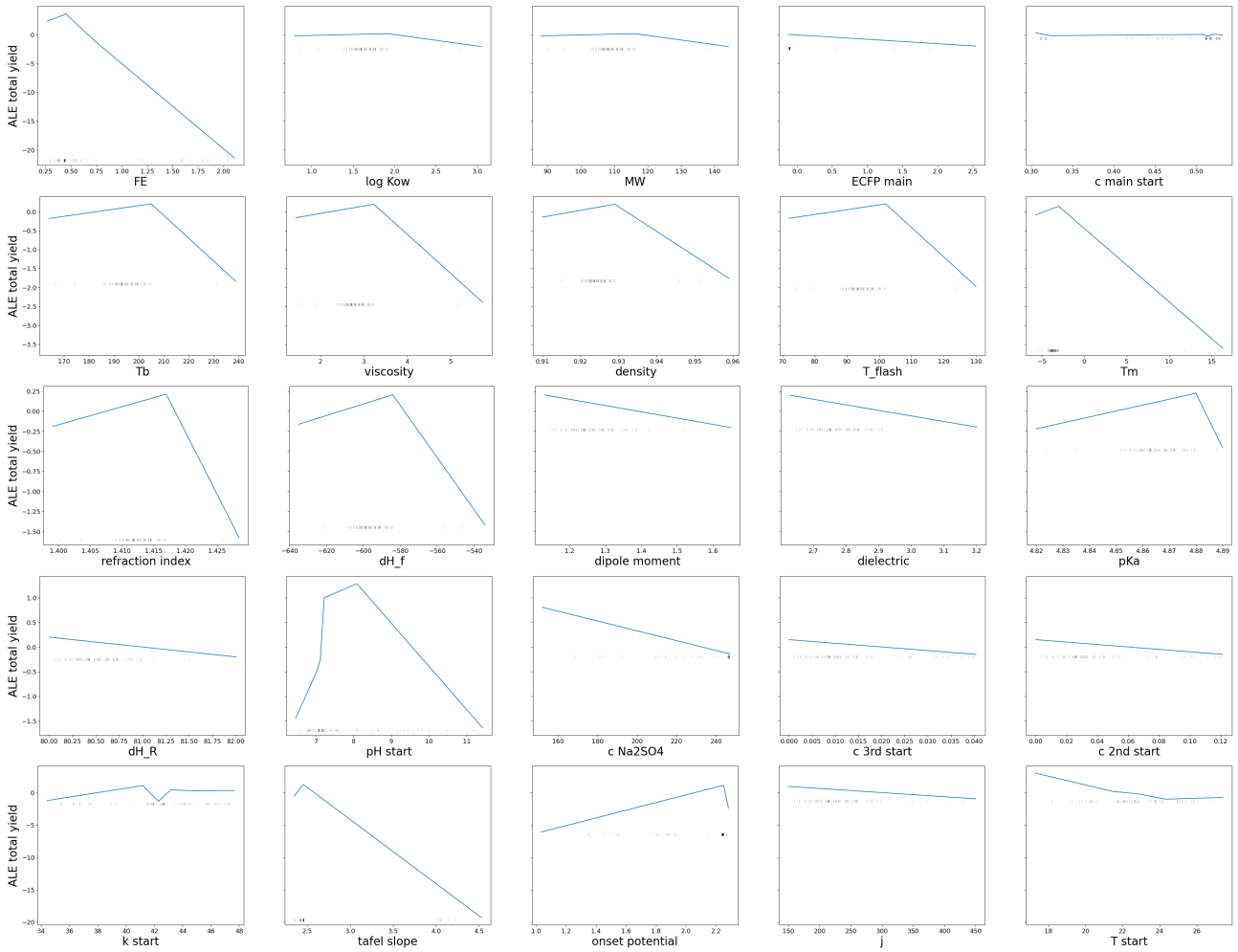



**Figure 20:** Batch - CE: ALE of the champion model of CE.

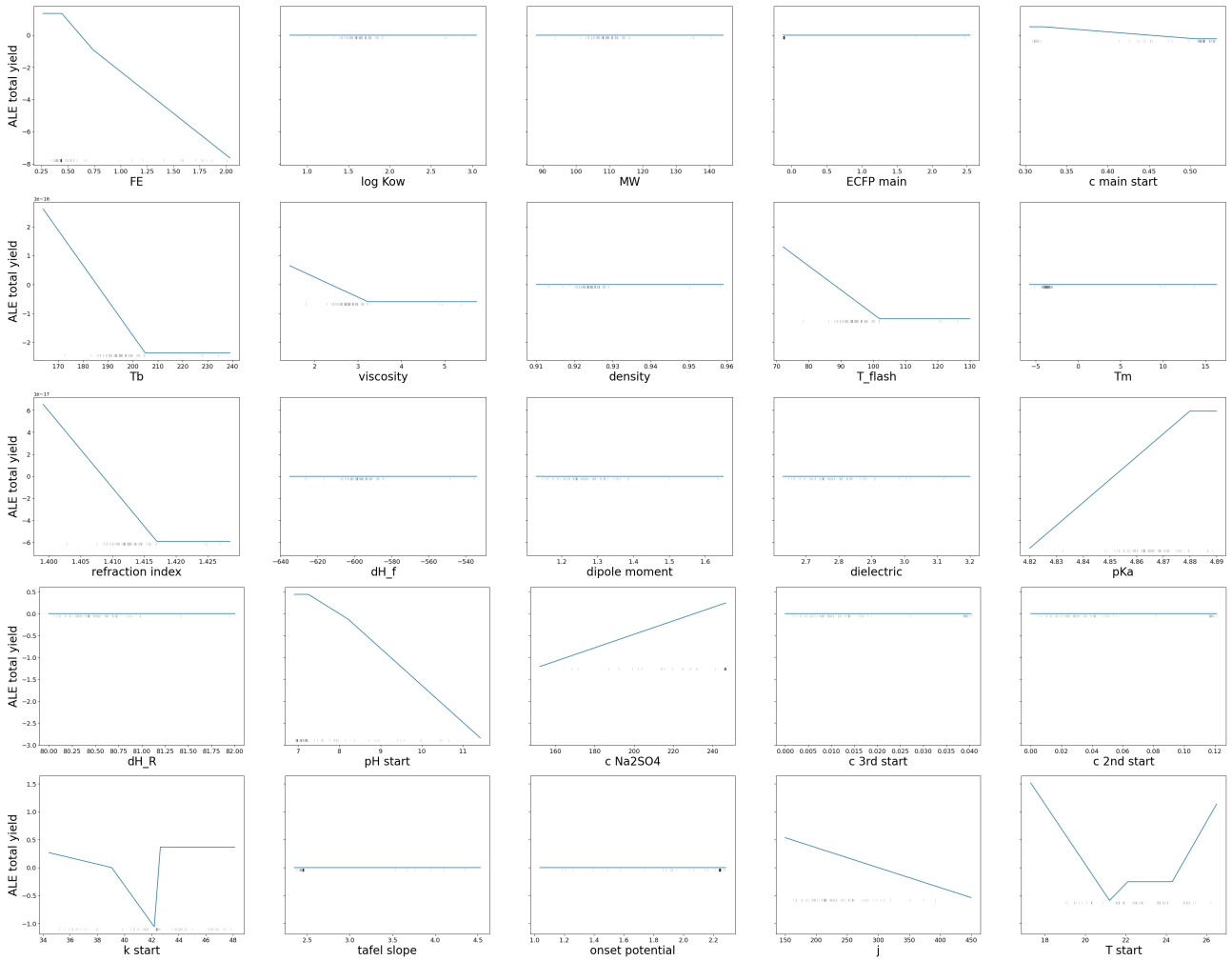



**Figure 21:** Batch - S: ALE of the champion model of S.

## 4.3 Reinforcement learning for reaction optimization

Reinforcement learning was tested using the Proximal Policy Optimization (PPO) algorithm[24] to optimize the Kolbe reaction in a flow cell. The flow cell data set was chosen due to smaller number of features to optimize. The model was previously developed by the working group members Oesterreich and Gerstenberg[25] and was adapted to this reaction. The reward function R was defined as:

$$R = CE + S + Y \tag{4}$$

Other reward functions could be engineered in the future to give more weight to certain output variables. In addition, other factors could be introduced such as costs as the cost of electricity consumed. The environment was a knn regression model which was the champion of the best performing type of model with the three outputs CE, S and Y. Six runs with one million timesteps resulted in the same optimized outputs of 73.6% CE, 91.2% S and 27.6% Y. However, different input feature values were computed, as shown in Figure 22. Only the flow rate was determined to be the same optimal value of 2 L/min in each run. The other features varied significantly between the runs. Notably, the optimized j was below 300 mA/cm$^2$ in 5 out of 6 runs, challenging the XAI result of the favorable influence of high j on CE, S and Y.

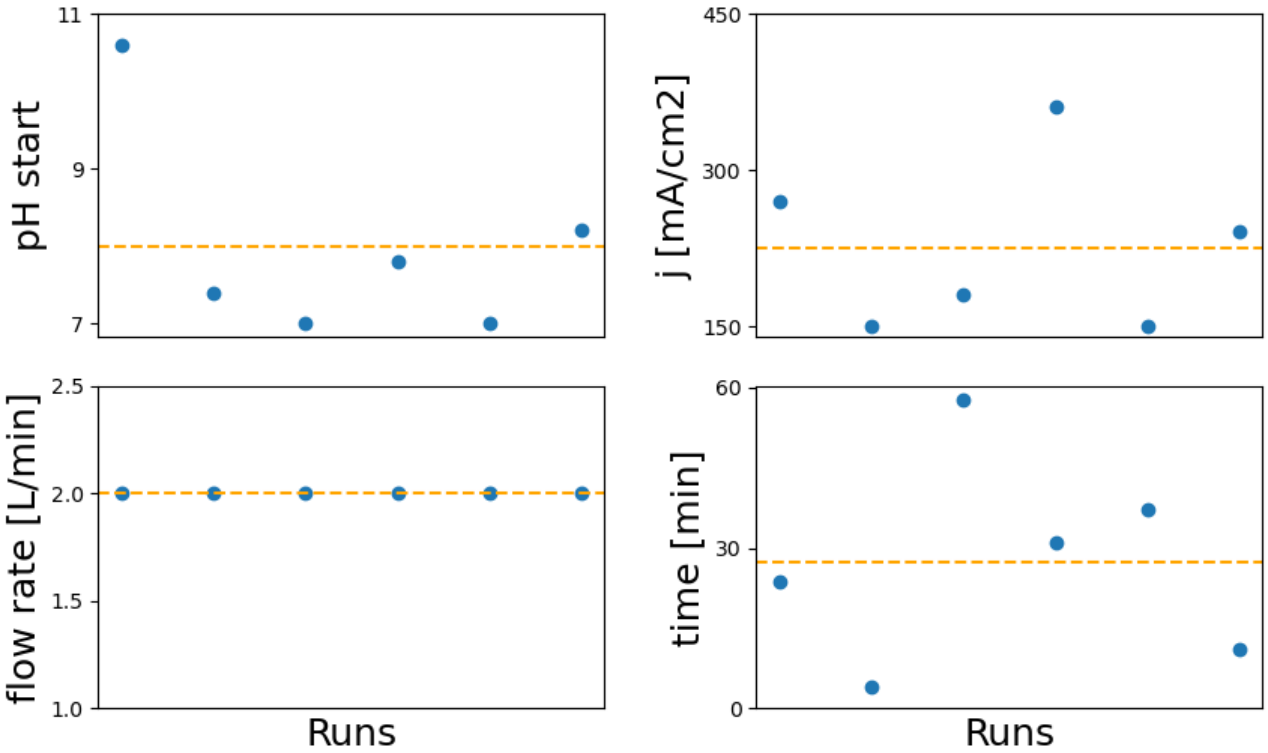



**Figure 22:** PPO reinforcement learning optimised input features for an optimal sum of CE, S and Y of 6 runs (blue dots) and their average (orange dotted line). Reward R = CE + S + Y.

# 5 Conclusion and Outlook

Supervised machine learning and XAI methods were demonstrated on data from galvanostatic flow cell and batch Kolbe electrolysis. After screening of 12 different types of regression models, the champion models were investigated. The simple non-linear models knn and decision tree as well as the more complex tree-based ensemble models (random forest and gradient tree boosting) showed the best performance.
With respect to the batch data set, FE and the electrode descriptors (onset potential and tafel slope) dominated the decisions of the models. A lower importance of substance related features was observed, but this can also be attributed to the larger amounts of these features. Therefore, feature selection is recommended for further model trainings on this data set.
Regarding the flow cell data, current density j, flow rate and time were most important to the model. The starting pH had less influence. Trade-offs between the outputs for the input parameters were visualized. Thus, reinforcement learning driven reaction optimisation was demonstrated. High reproducibility of the best outputs and the flow rate was achieved, while the other input features varied significantly between six runs. The reward engineering of RL offers further optimization for application to include non-chemical aspects, such as electricity costs.
The most crucial aspect for further analysis is the need for more data to validate and enrich the results of this work. Literature data could be helpful, but the lack of documentation of the reaction parameters significantly limits its use. Therefore, this work calls for the provision of more detailed and accessible research data.

# Acknowledgements

I want to thank all the people, who have supported me in my work: Julia Westermayr for her supervision and feedback, Toni Oesterreich for his technical introductions to the cluster- and office-infrastructure and his help with code, Falk Harnisch, Micjel Chávez Morejón and Katharina Röhring for providing the data, the university data center for the cluster-infrastructure, OpenAI's ChatGTP 3.5 for generating code and answering my corresponding questions and all members of the Westermayr and Tonner-Zech groups for their moral support.

# Conflict of Interest

I have written this work myself. The passages, which are taken in the meaning of other works, are marked in each individual case under indication of the reference. This assurance

also refers to the figures. I declare no conflict of interest.

## Data Availability Statement

I can confirm that all relevant data are included in this report, its appendix and the following GitHub page: https://github.com/peter1999de/Machine-learning-driven-parameter-investigation-of-the-Kolbe-reaction/

# References

[1] Cid Gomes, L., 2023. *Light-Induced Routes to Sustainable Biocrudes for Fuels and Lubricant Oils*, Band 2227 von *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology*. Acta Universitatis Upsaliensis, Uppsala. ISBN 978-91-513-1682-6.

[2] Rosa, L. F.; Röhring, K.; Harnisch, F., 2024. Electrolysis of medium chain carboxylic acids to aviation fuel at technical scale. *Fuel* 356, 129590.

[3] International Energy Agency. World Energy Outlook 2022.

[4] Nilges, P.; dos Santos, T. R.; Harnisch, F.; Schröder, U., 2012. Electrochemistry for biofuel generation: Electrochemical conversion of levulinic acid to octane. *Energy & Environmental Science* 5 (1), 5231–5235.

[5] Dörr, M.; Hielscher, M. M.; Proppe, J.; Waldvogel, S. R., 2021. Electrosynthetic Screening and Modern Optimization Strategies for Electrosynthesis of Highly Value–added Products. *ChemElectroChem* 8 (14), 2621–2629.

[6] Harnisch, F.; Urban, C., 2018. Electrobiorefineries: Unlocking the Synergy of Electrochemical and Microbial Conversions. *Angewandte Chemie (International ed. in English)* 57 (32), 10016–10023.

[7] Urban, C.; Xu, J.; Sträuber, H.; dos Santos Dantas, T. R.; Mühlenberg, J.; Härtig, C.; Angenent, L. T.; Harnisch, F., 2017. Production of drop-in fuels from biomass at high selectivity by combined microbial and electrochemical conversion. *Energy & Environmental Science* 10 (10), 2231–2244.

[8] Stang, C.; Harnisch, F., 2016. The Dilemma of Supporting Electrolytes for Electroorganic Synthesis: A Case Study on Kolbe Electrolysis. *ChemSusChem* 9 (1), 50–60.

[9] Kolbe, H., 1849. Untersuchungen über die Elektrolyse organischer Verbindungen. *Justus Liebigs Annalen der Chemie* 69 (3), 257–294.

[10] Lang, L.; Li, Y.; Lam, J. C.-H.; Ding, Y.; Yin, X.; Wu, C., 2022. Enhancing the selectivity of hydrocarbons during the Kolbe electrolysis of biomass-derived short-chain carboxylic acids by anionic surfactants. *Sustainable Energy & Fuels* 6 (11), 2797–2804.

[11] Neubert, K.; Schmidt, M.; Harnisch, F., 2021. Platinized Titanium as Alternative Cost-Effective Anode for Efficient Kolbe Electrolysis in Aqueous Electrolyte Solutions. *ChemSusChem* 14 (15), 3097–3109.

[12] Qiu, Y.; Lopez-Ruiz, J. A.; Sanyal, U.; Andrews, E.; Gutiérrez, O. Y.; Holladay, J. D., 2020. Anodic electrocatalytic conversion of carboxylic acids on thin films of RuO2, IrO2, and Pt. *Applied Catalysis B: Environmental* 277, 119277.

[13] Neubert, K.; Hell, M.; Chávez Morejón, M.; Harnisch, F., 2022. Hetero-Coupling of Bio-Based Medium-Chain Carboxylic Acids by Kolbe Electrolysis Enables High Fuel Yield and Efficiency. *ChemSusChem* 15 (21), e202201426.

[14] ONO, Y.; KIM, S.-H.; YASUDA, M.; NONAKA, T., 1999. Kolbe Electrolysis of Carboxylates on a Hydrophobic Platinum Electrode Composite-plated with PTFE Particles. *Electrochemistry* 67 (11), 1042–1045.

[15] Palkovits, S., 2020. A Primer about Machine Learning in Catalysis – A Tutorial with Code. *ChemCatChem* 12 (16), 3995–4008.

[16] Christoph Molnar, 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2 Auflage.

[17] Jerome H. Friedman, 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29 (5), 1189–1232.

[18] Goldstein, A.; Kapelner, A.; Bleich, J.; Pitkin, E., 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics* 24 (1), 44–65.

[19] National Center for Biotechnology Information. National Library of Medicine - PubChem. `https://pubchem.ncbi.nlm.nih.gov/`.

[20] R, S. `http://www.stenutz.eu/`.

[21] National Institute of Standards and Technoloty. NIST Chemistry WebBook. `https://webbook.nist.gov/`.

[22] Chen, Y.; Tian, B.; Cheng, Z.; Li, X.; Huang, M.; Sun, Y.; Liu, S.; Cheng, X.; Li, S.; Ding, M., 2021. Electro-Descriptors for the Performance Prediction of Electro-Organic Synthesis. *Angewandte Chemie (International ed. in English)* 60 (8), 4199–4207.

[23] Angelov, P. P.; Soares, E. A.; Jiang, R.; Arnold, N. I.; Atkinson, P. M., 2021. Explainable artificial intelligence: an analytical review. *WIREs Data Mining and Knowledge Discovery* 11 (5).

[24] Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms.

[25] Toni Oesterreich, A. G., 2023. Towards Automated Reaction Optimization by Machine Learning: Investigation of Experimental Setups and ML Approaches: Report of a Project in Chemistry at the University Leipzig.

# Appendix

## A1 Model screening

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 356 | 159 | 18.4 | 4.19 | 0.0619 | 0.246 |
| kernel ridge | 423 | 565 | 18.1 | 9.75 | 0.0825 | 0.694 |
| linear kernel ridge | 808 | 358 | 27.8 | 5.99 | -1.43 | 1.49 |
| random forest | 240 | 209 | 14.6 | 5.18 | 0.322 | 0.456 |
| decision tree | 205 | 199 | **13.1** | 5.75 | 0.478 | 0.379 |
| multiple linear | 418 | 169 | 20.1 | 3.97 | -0.157 | 0.412 |
| explainable deep neural network | 826 | 300 | 28.2 | 5.66 | -1.29 | 0.855 |
| gradient tree boosting | 221 | 210 | 13.8 | 5.59 | 0.45 | 0.349 |
| linear ridge | 398 | 171 | 19.5 | 4.11 | -0.0698 | 0.307 |
| **k-nearest neighbour** | **192** | 118 | 13.1 | 4.45 | **0.494** | 0.297 |
| bayesian ridge | 396 | 163 | 19.5 | 3.99 | -0.0499 | 0.21 |
| huber | 379 | 169 | 19.0 | 4.19 | -0.0361 | 0.353 |

**Table A1:** Flow data, output parameter: CE of the fuel. Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 13.7 | 6.40 | 3.61 | 0.819 | 0.165 | 0.358 |
| kernel ridge | 17.8 | 8.45 | 4.09 | 1.06 | -0.0292 | 0.382 |
| linear kernel ridge | 787 | 291 | 27.5 | 5.44 | -54.8 | 33.2 |
| random forest | 6.98 | 3.80 | 2.55 | 0.674 | 0.559 | 0.278 |
| decision tree | 6.92 | 5.85 | 2.48 | 0.877 | **0.607** | 0.17 |
| multiple linear | 10.9 | 5.3 | 3.21 | 0.767 | 0.321 | 0.335 |
| explainable deep neural network | 4.17e+03 | 222 | 64.5 | 1.73 | -281 | 115 |
| **gradient tree boosting** | **6.50** | 3.98 | **2.46** | 0.687 | 0.586 | 0.216 |
| linear ridge | 10.9 | 5.79 | 3.20 | 0.826 | 0.336 | 0.327 |
| k-nearest neighbour | 6.90 | 3.33 | 2.56 | 0.596 | 0.509 | 0.332 |
| bayesian ridge | 11.1 | 5.75 | 3.23 | 0.820 | 0.330 | 0.317 |
| huber | 11.3 | 6.51 | 3.24 | 0.929 | 0.304 | 0.418 |

**Table A2:** Flow data, output parameter: S of decane (without gases). Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 62.4 | 31.3 | 7.70 | 1.77 | 0.353 | 0.162 |
| kernel ridge | 60.6 | 35.4 | 7.46 | 2.2 | 0.38 | 0.28 |
| linear kernel ridge | 62 | 30.4 | 7.65 | 1.86 | 0.372 | 0.147 |
| random forest | 57 | 50.6 | 6.95 | 2.96 | 0.412 | 0.475 |
| decision tree | 51.5 | 35.4 | 6.84 | 2.17 | 0.462 | 0.272 |
| multiple linear | 61.8 | 34.9 | 7.59 | 2.04 | 0.383 | 0.158 |
| explainable deep neural network | 72 | 39.4 | 8.21 | 2.17 | 0.28 | 0.191 |
| **gradient tree boosting** | **41.1** | 33.4 | **6.04** | 2.17 | **0.597** | 0.186 |
| linear ridge | 61.3 | 37.1 | 7.53 | 2.17 | 0.395 | 0.17 |
| k-nearest neighbour | 47.8 | 35.8 | 6.46 | 2.47 | 0.488 | 0.332 |
| bayesian ridge | 61.8 | 38.8 | 7.54 | 2.22 | 0.393 | 0.174 |
| huber | 61.7 | 42.1 | 7.46 | 2.46 | 0.397 | 0.239 |

**Table A3:** Flow data, output parameter: total yield of decane. Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 108 | 43.4 | 10.2 | 2.22 | 0.766 | 0.101 |
| kernel ridge | 103 | 40.8 | 9.93 | 2.15 | 0.772 | 0.103 |
| linear kernel ridge | 125 | 30.8 | 11.1 | 1.46 | 0.726 | 0.0998 |
| random forest | 128 | 41.8 | 11.2 | 1.86 | 0.719 | 0.121 |
| decision tree | 225 | 182 | 13.9 | 5.64 | 0.508 | 0.378 |
| multiple linear | 201 | 103 | 13.7 | 3.58 | 0.543 | 0.307 |
| explainable deep neural network | 382 | 85.0 | 19.4 | 2.13 | 0.201 | 0.069 |
| gradient tree boosting | 103 | 44.6 | 9.93 | 2.18 | 0.770 | 0.121 |
| linear ridge | 164 | 86.3 | 12.4 | 3.23 | 0.626 | 0.252 |
| **k-nearest neighbour** | **83.8** | 28.6 | **8.98** | 1.78 | **0.813** | 0.0807 |
| bayesian ridge | 127 | 35.8 | 11.2 | 1.64 | 0.718 | 0.118 |
| huber | 170 | 84.7 | 12.7 | 3.03 | 0.606 | 0.265 |

**Table A4:** Batch data, output parameter: CE of alkanes. Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 68.4 | 43.0 | 7.90 | 2.46 | 0.815 | 0.157 |
| kernel ridge | 99.4 | 58.4 | 9.53 | 2.92 | 0.714 | 0.286 |
| linear kernel ridge | 119 | 47.7 | 10.7 | 2.09 | 0.674 | 0.270 |
| random forest | 74.5 | 52.6 | 8.14 | 2.86 | 0.818 | 0.124 |
| decision tree | 178 | 129 | 11.9 | 5.99 | 0.524 | 0.54 |
| multiple linear | 123 | 44.7 | 10.9 | 1.98 | 0.693 | 0.176 |
| explainable deep neural network | 433 | 140 | 20.6 | 3.25 | 0.0423 | 0.237 |
| gradient tree boosting | 91.7 | 67.5 | 9.06 | 3.10 | 0.762 | 0.230 |
| linear ridge | 120 | 44.7 | 10.8 | 1.93 | 0.664 | 0.263 |
| **k-nearest neighbour** | **40.7** | 42.7 | **5.78** | 2.71 | **0.875** | 0.160 |
| bayesian ridge | 117 | 45.9 | 10.6 | 2.02 | 0.680 | 0.265 |
| huber | 133 | 56.3 | 11.3 | 2.36 | 0.618 | 0.336 |

**Table A5:** Batch data, output parameter: S of alkanes. Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 48.0 | 14.4 | 6.85 | 1.04 | 0.624 | 0.134 |
| kernel ridge | 45.4 | 27.2 | 6.48 | 1.85 | 0.646 | 0.184 |
| linear kernel ridge | 72.2 | 28.1 | 8.31 | 1.76 | 0.430 | 0.282 |
| random forest | 52.7 | 23.3 | 7.09 | 1.53 | 0.602 | 0.152 |
| decision tree | 109 | 76.2 | 9.85 | 3.38 | 0.167 | 0.485 |
| multiple linear | 123 | 80.6 | 10.4 | 3.88 | -0.0668 | 1.01 |
| explainable deep neural network | 93.3 | 36.9 | 9.45 | 1.97 | 0.346 | 0.0622 |
| gradient tree boosting | 58.3 | 24.9 | 7.45 | 1.69 | 0.545 | 0.213 |
| linear ridge | 97.9 | 59.0 | 9.44 | 2.98 | 0.155 | 0.731 |
| k-nearest neighbour | **35.1** | 11.3 | **5.86** | 0.911 | **0.726** | 0.0972 |
| bayesian ridge | 78.1 | 29.9 | 8.65 | 1.79 | 0.344 | 0.443 |
| huber | 110 | 60.2 | 10.1 | 2.94 | 0.0256 | 0.957 |

**Table A6:** Batch data, output parameter: total yield of alkanes. Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

| | MSE | STD | RMSE | STD | R2 | STD |
|---|---|---|---|---|---|---|
| gaussian progress | 206 | 180 | 12.9 | 6.38 | **-0.853** | 1.64 |
| kernel ridge | 1.44e+04 | 3.08e+04 | 71.0 | 96.5 | -197 | 411 |
| linear kernel ridge | 1.75e+04 | 3.31e+04 | 87.3 | 99.2 | -218 | 444 |
| random forest | 200 | 93.9 | 13.7 | 3.55 | -3.63 | 9.89 |
| decision tree | 238 | 134 | 14.7 | 4.67 | -5.92 | 16.2 |
| multiple linear | 1.16e+32 | 2.07e+32 | 5.66e+15 | 9.19e+15 | -1.04e+30 | 1.96e+30 |
| xDNN | 6.43e+03 | 8.5e+03 | 57.7 | 55.7 | -82.3 | 115 |
| gradient tree boosting | 210 | 108 | 14.0 | 3.85 | -5.06 | 14.2 |
| linear ridge | 1.35e+04 | 3.64e+04 | 63.0 | 97.4 | -183 | 487 |
| **k-nearest neighbour** | **132** | 82.5 | **10.8** | 3.83 | -2.38 | 7.68 |
| bayesian ridge | 1.17e+03 | 1.93e+03 | 26.4 | 21.7 | -18.2 | 33.6 |
| huber | 2.01e+04 | 4.92e+04 | 80.6 | 117 | -254 | 659 |

**Table A7:** Batch data with means instead of single replicates, output parameter: total yield of alkanes. Mean evaluation metrics of the model screening (10 different random states). STD is the standard deviation of the metrics to the left. Bold: the lowest values of each metrics and the method with the most lowest values.

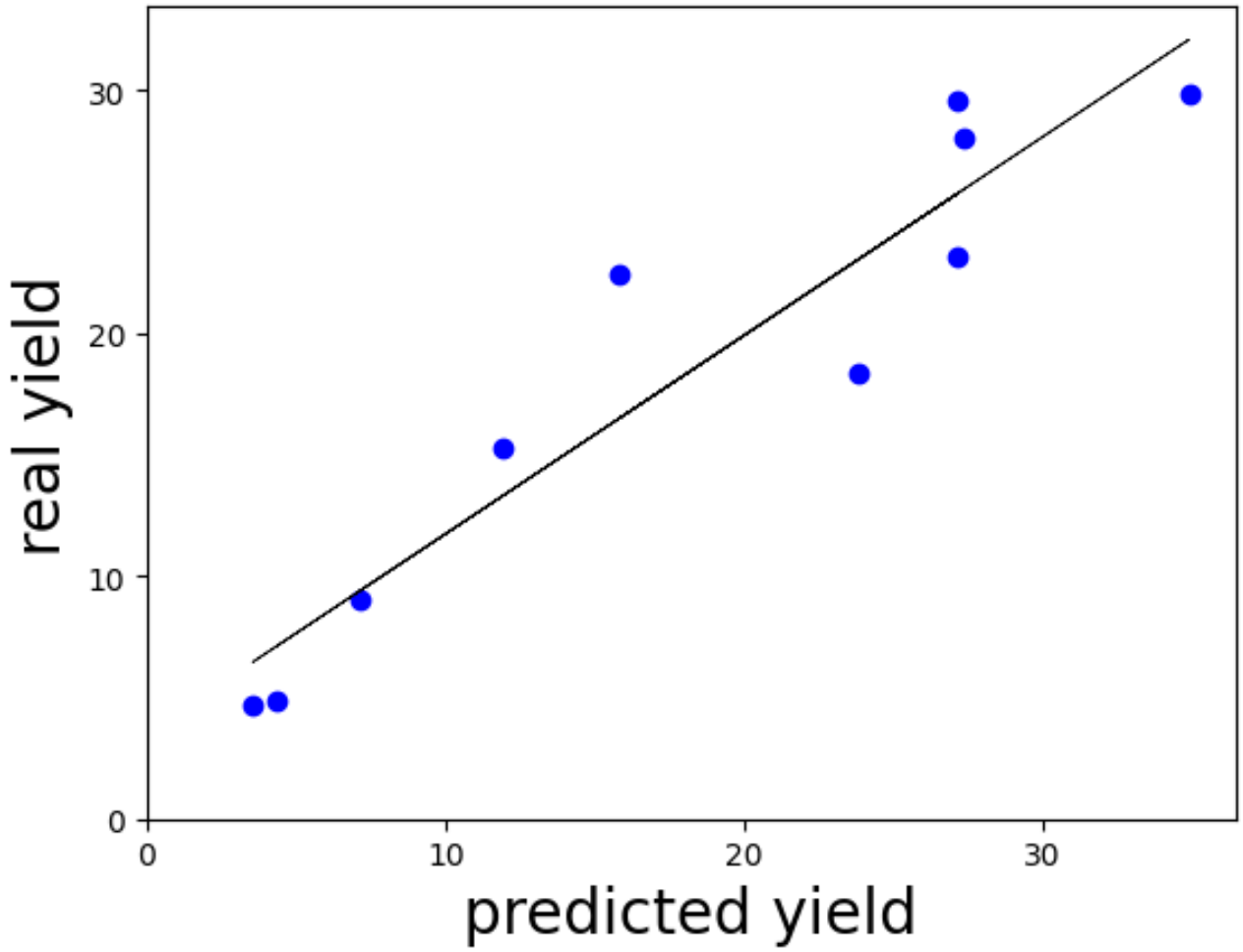



**Figure A1:** Flow cell: predicted vs. test set outputs of the champion models of Y (blue dots). The black line is the corresponding linear fit.

# A2 XAI - model explainability

## A2.1 XAI flow cell

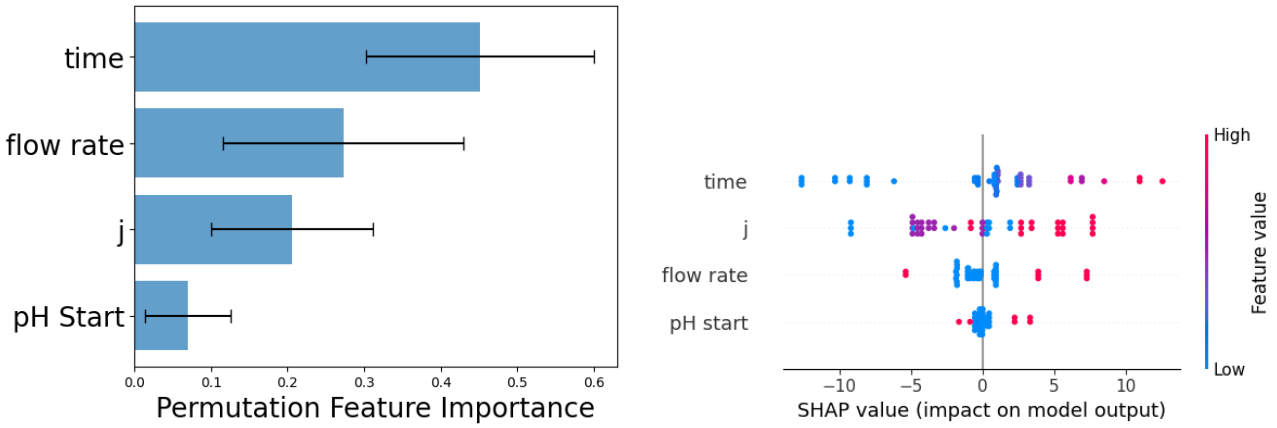



**Figure A2:** Flow cell - Y: (a) average permutation feature importance (500 permutations) on the test set and (b) SHAP on the whole dataset of the champion model of Y. The error bars are the standard deviation of the permutations.

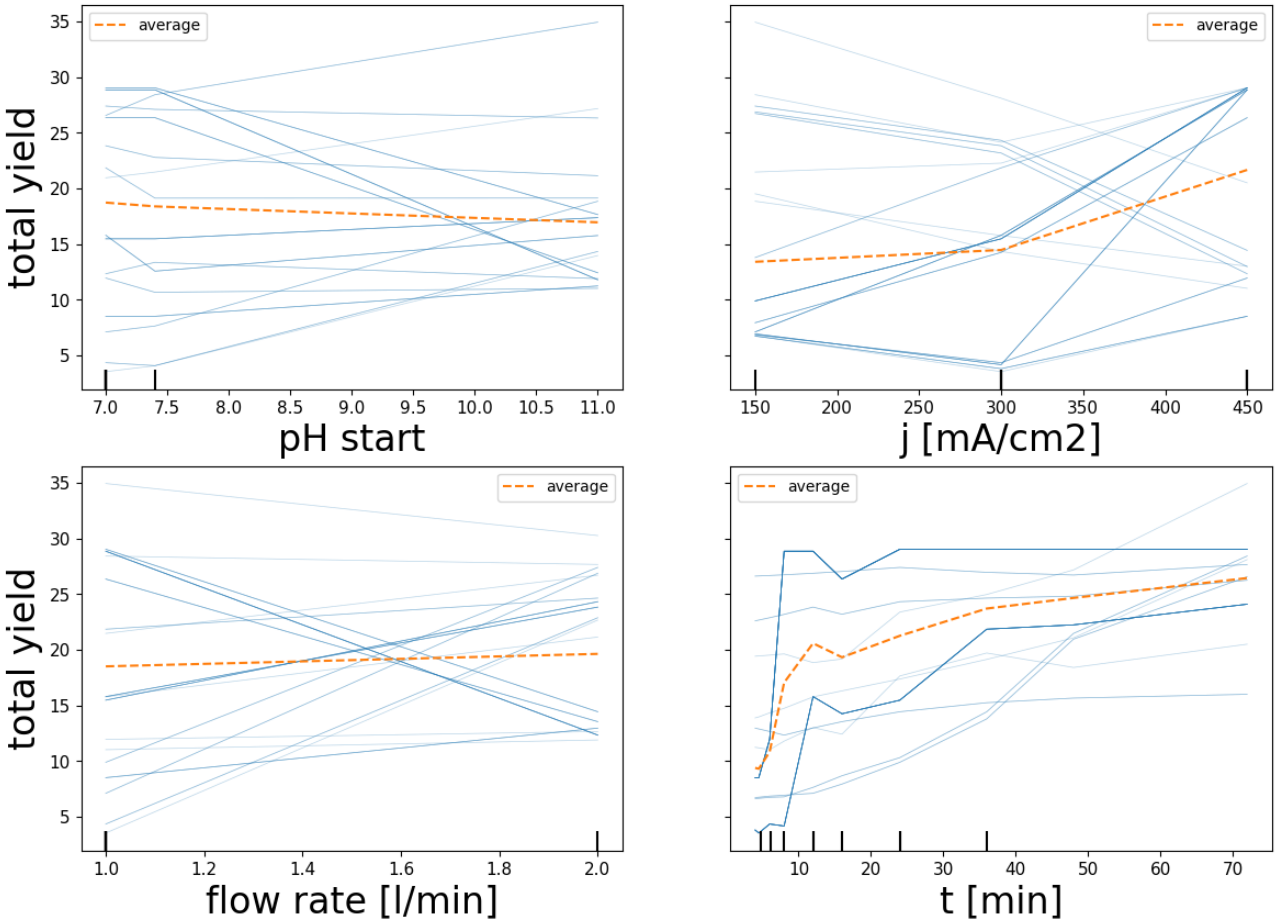



**Figure A3:** Flow cell - Y: PDP (orange dotted line) and ICE (blue solid lines) plots of the champion model of Y.

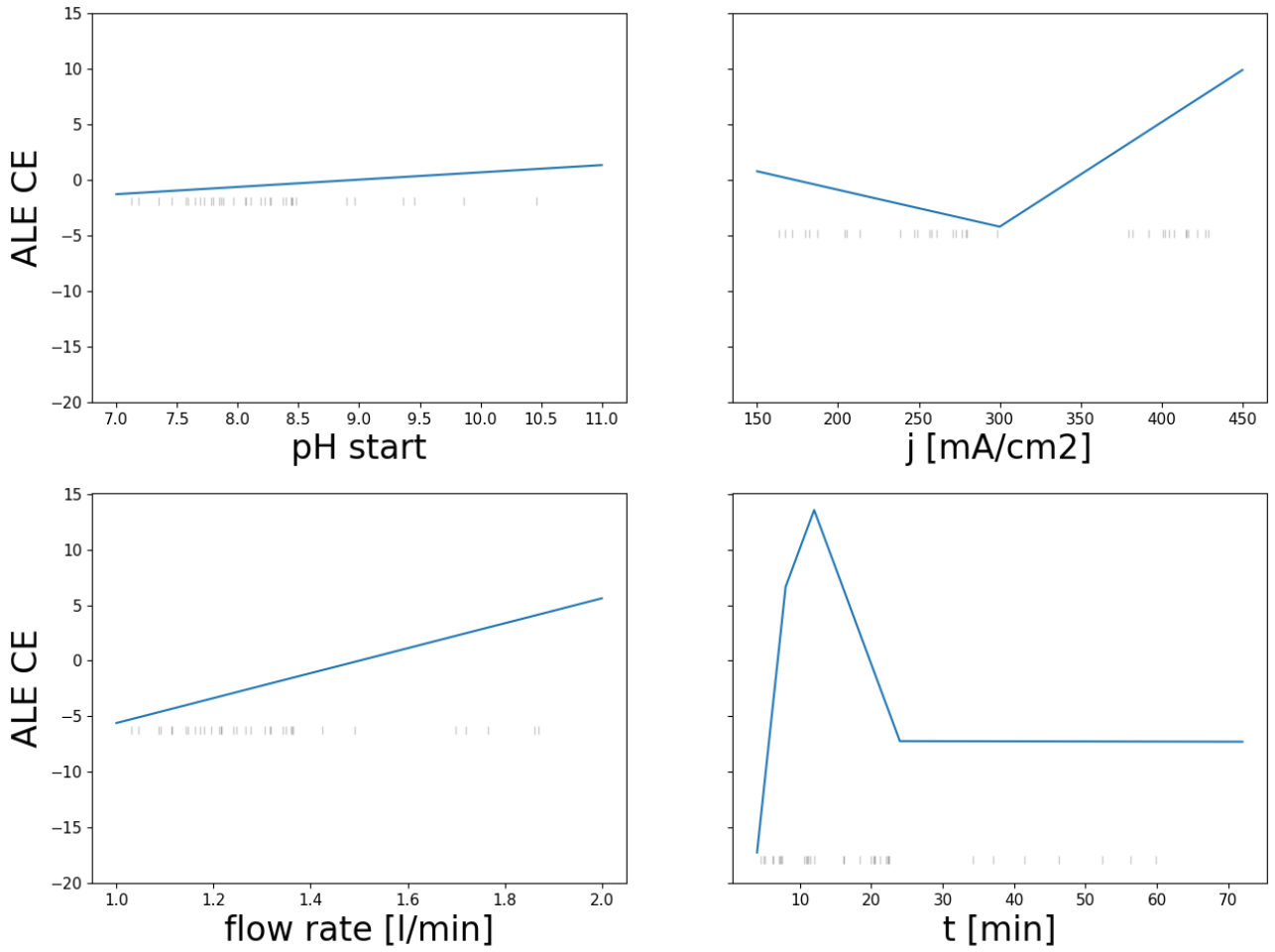



**Figure A4:** Flow cell - CE: ALE of the champion model of CE.

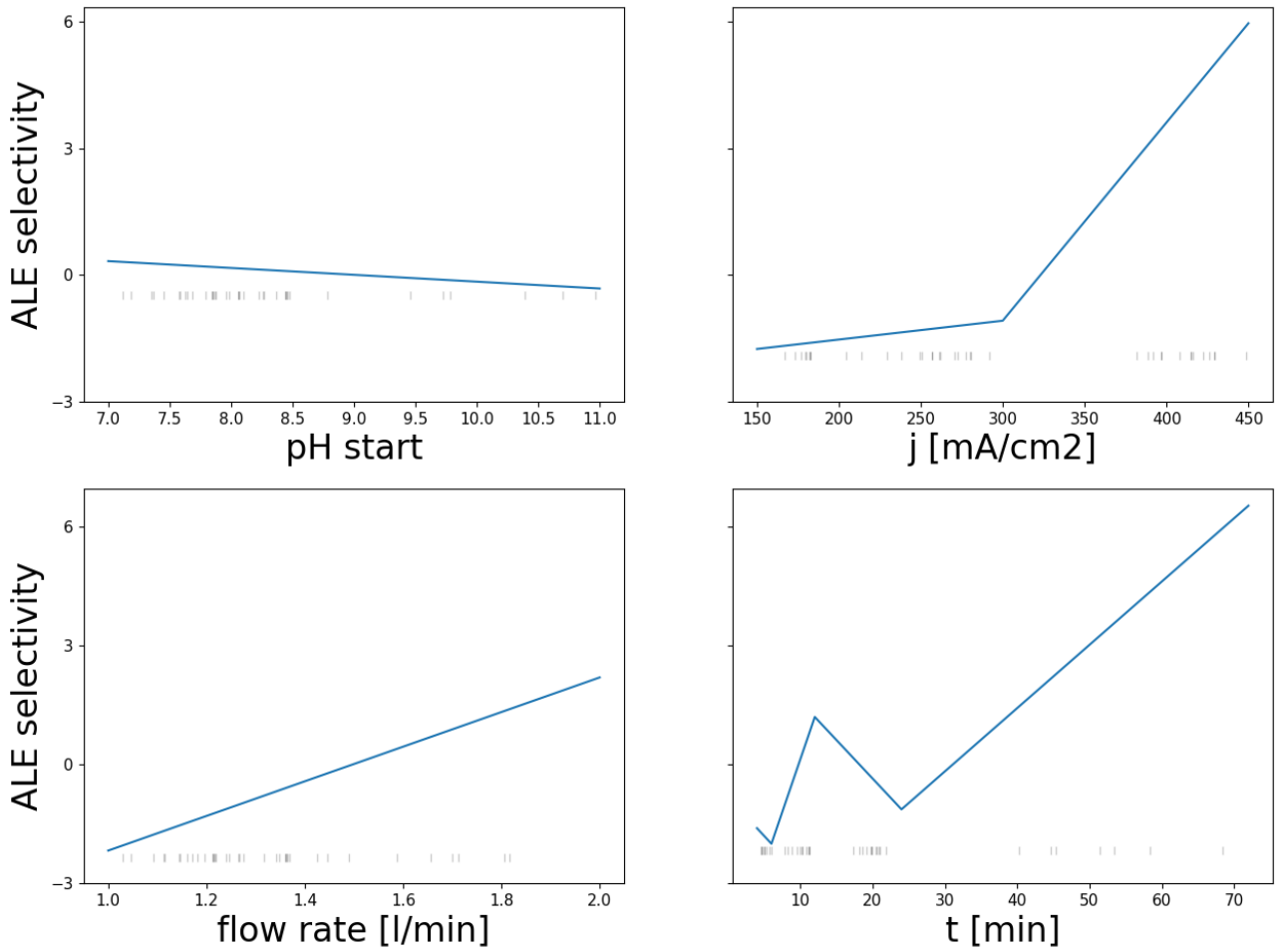
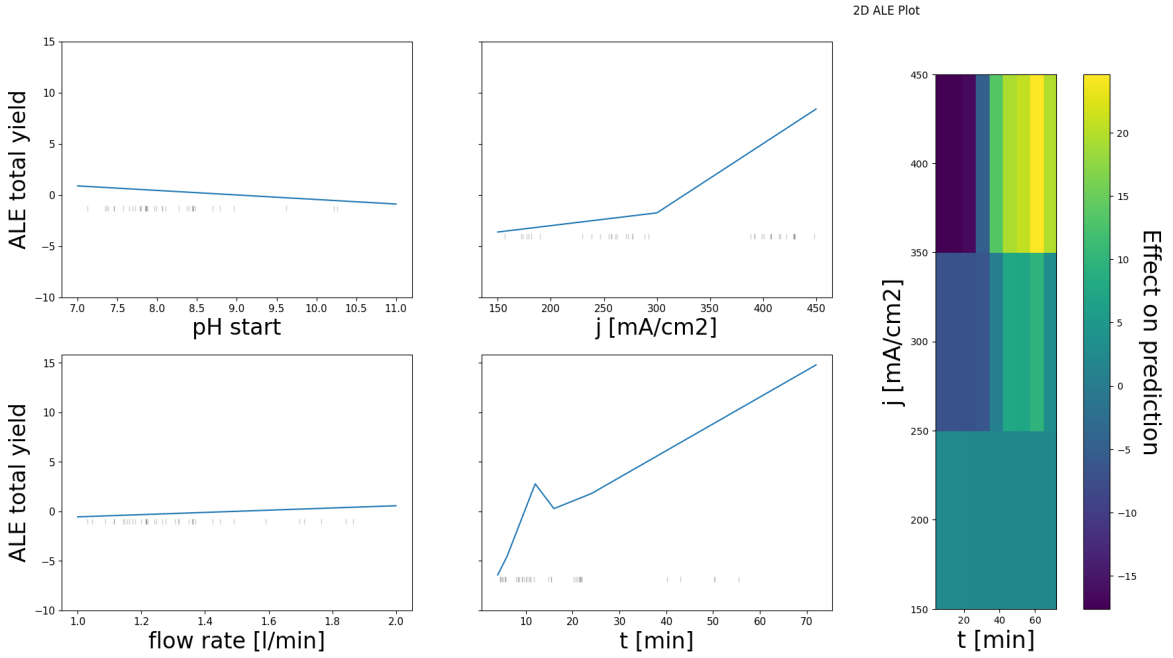



**Figure A5:** Flow cell - S: ALE of the champion model of S.




**(a)** ALE main effect **(b)** 2D ALE

**Figure A6:** Flow cell - Y: ALE of the champion model of Y with (a) the main effects and (b) only interaction effects between j and t.

## A2.2 XAI batch

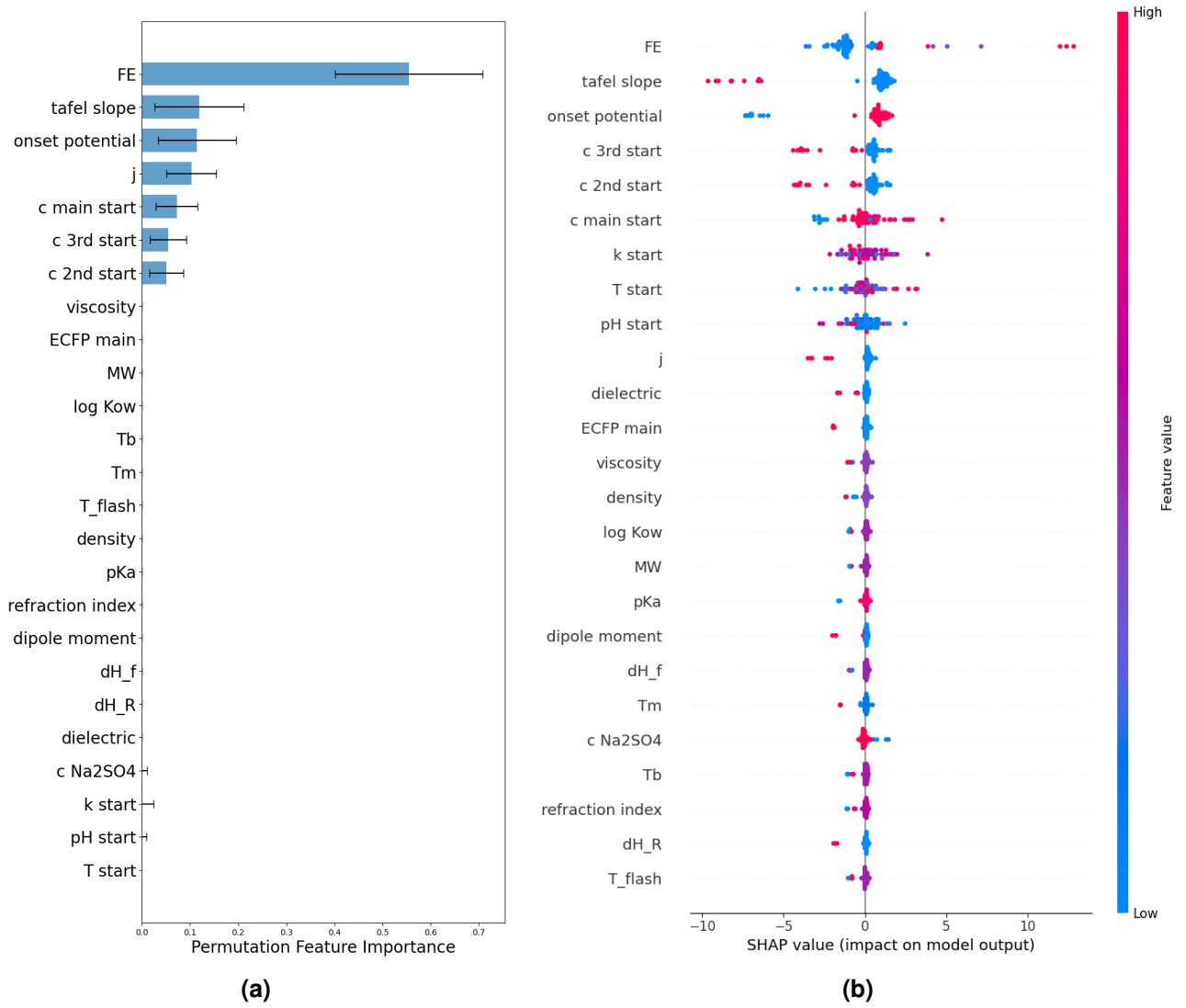



**Figure A7:** Batch - Y: (a) average permutation feature importance (500 permutations) on the test set and (b) SHAP on the whole dataset of the champion model of Y. The error bars are the standard deviation of the permutations.

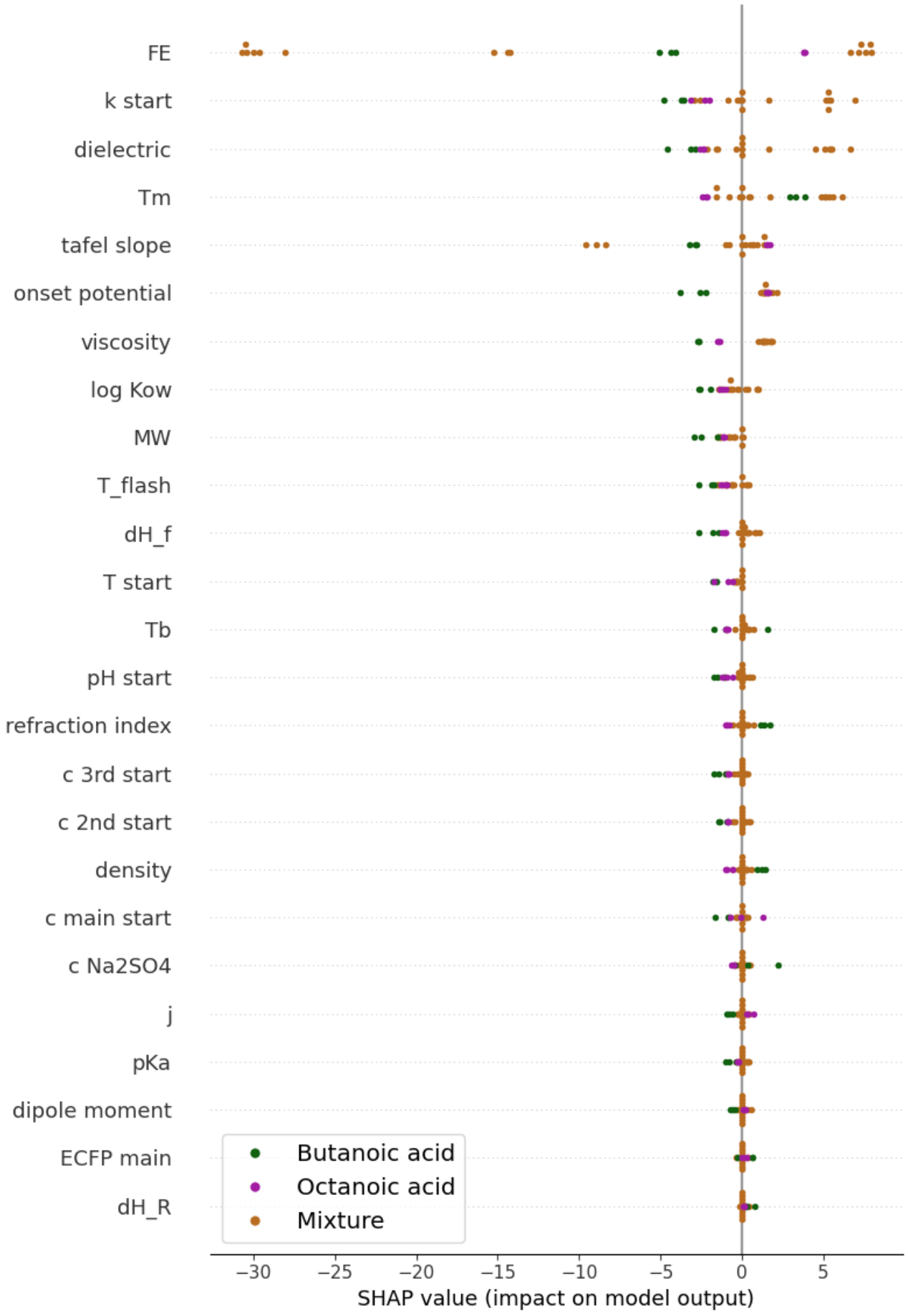



**Figure A8:** Batch - CE: SHAP on data of butanoic acid (green), octanoic acid (purple) and the mixture (brown) of butanoic, hexanoic and octanoic acid in the ratio 3:8:1.

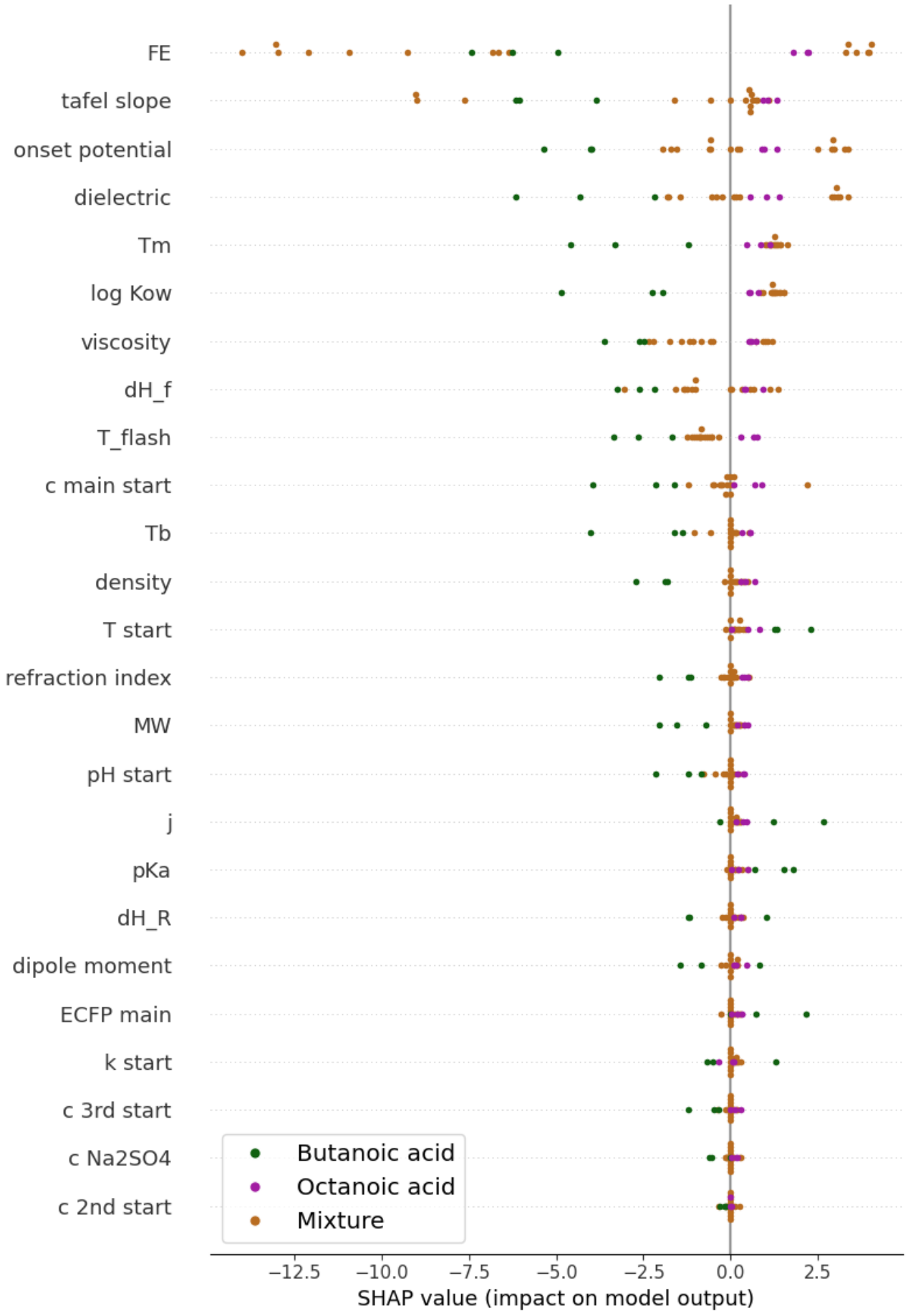



**Figure A9:** Batch - S: SHAP on data of butanoic acid (green), octanoic acid (purple) and the mixture (brown) of butanoic, hexanoic and octanoic acid in the ratio 3:8:1.

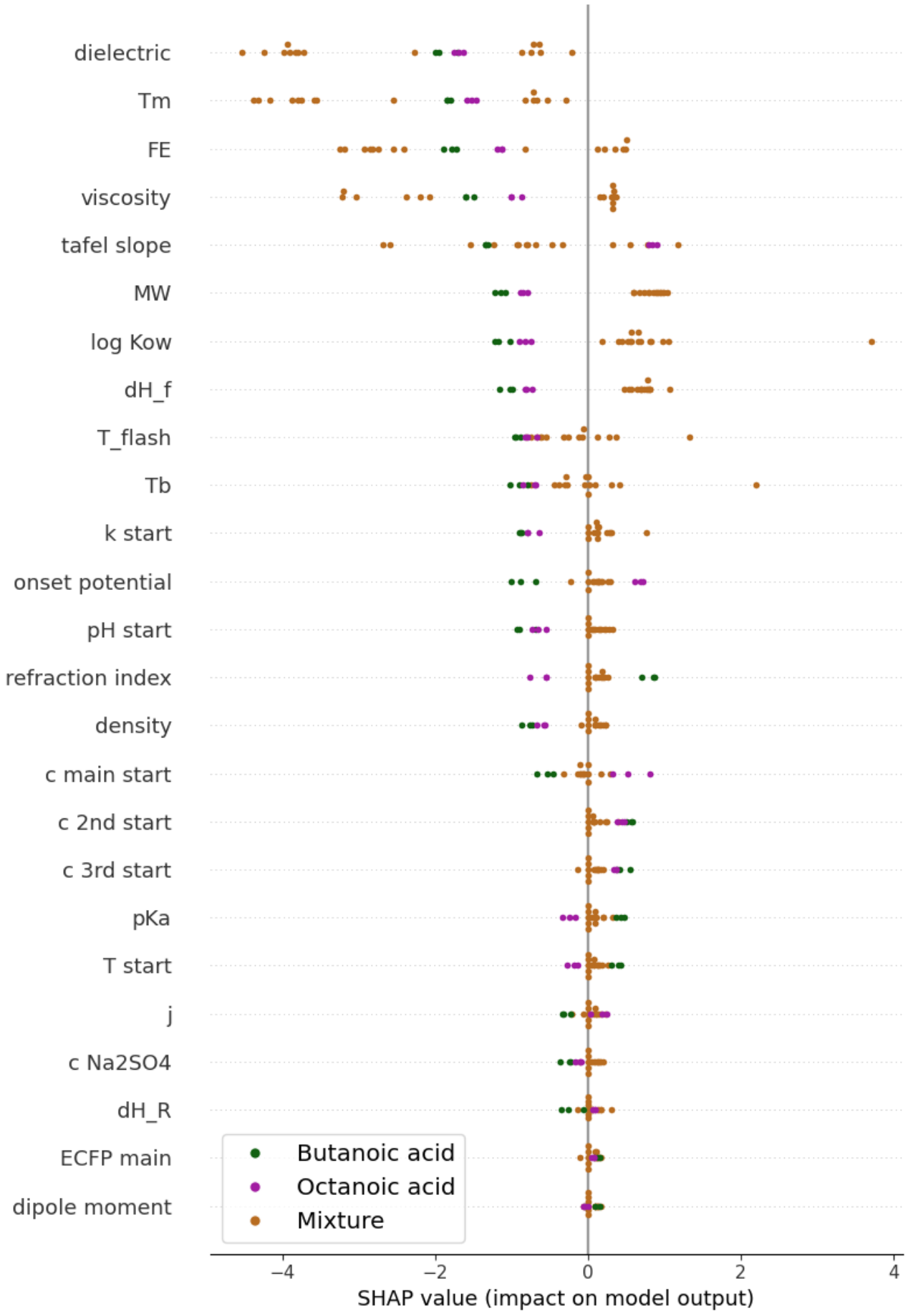



**Figure A10:** Batch - Y: SHAP on data of butanoic acid (green), octanoic acid (purple) and the mixture (brown) of butanoic, hexanoic and octanoic acid in the ratio 3:8:1.

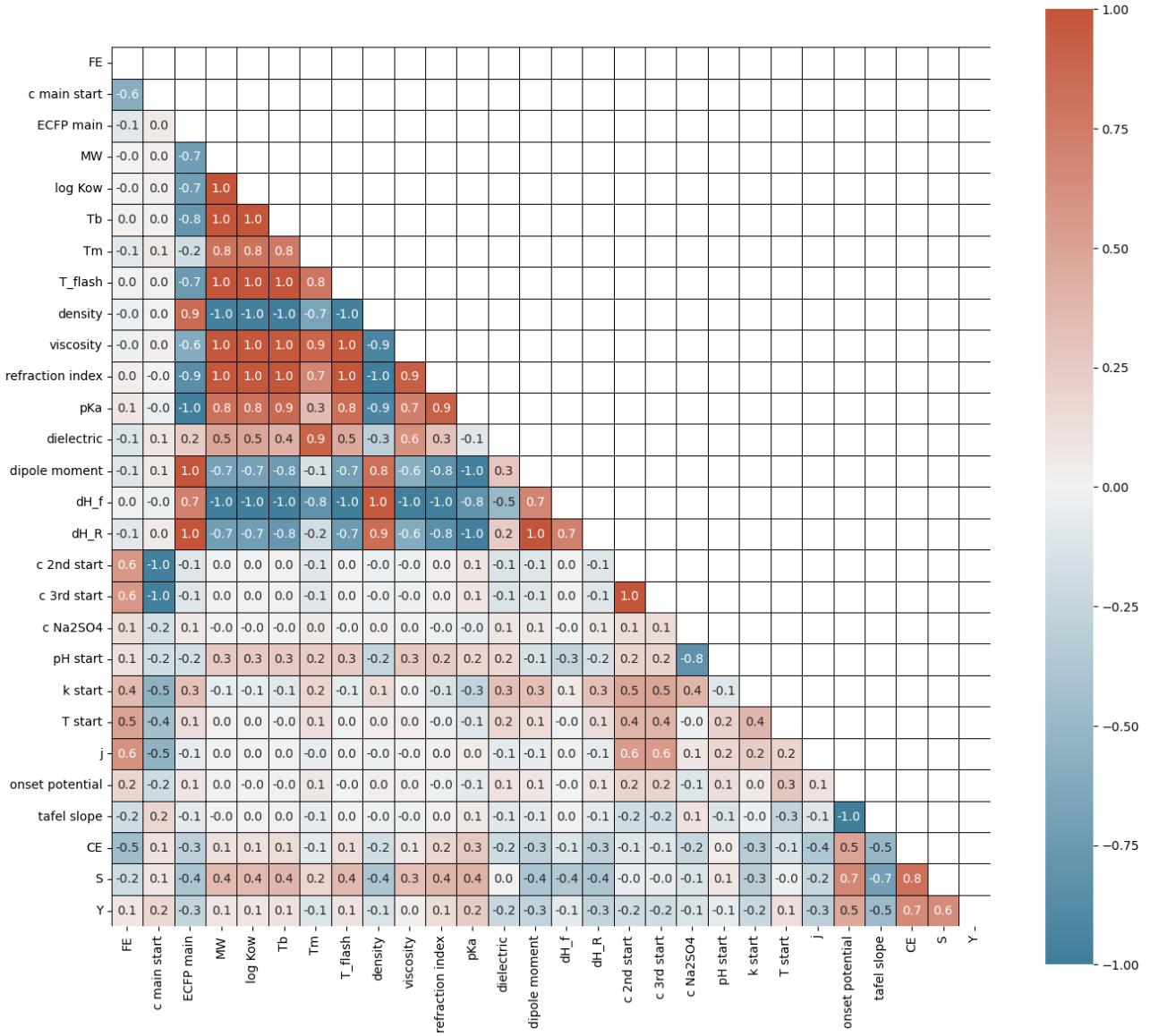



**Figure A11:** Correlation matrix of the input features and output variables of the batch data set. Positive correlations are marked red and negative ones blue.

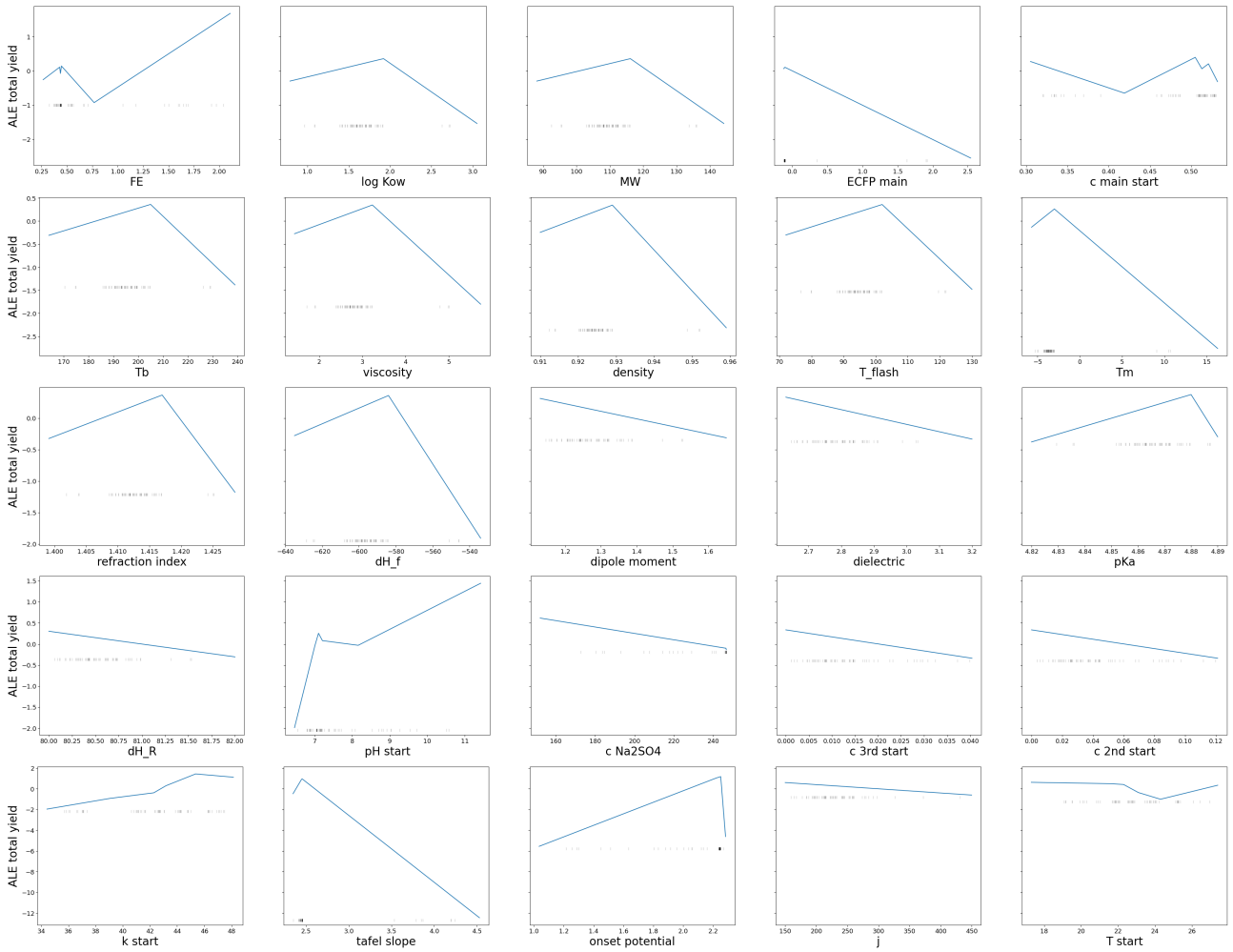



**Figure A12:** Batch - Y: ALE of the champion model of Y.