A PHYSICS-BASED DATA-DRIVEN MODEL FOR CO_2 GAS DIFFUSION ELECTRODES TO DRIVE AUTOMATED LABORATORIES

Ivan Grega[*], **Félix Therrien**[*], **Abhishek Soni**[○], **Karry Ocean**[○], **Kevan Dettelbach**[○],
Ribwar Ahmadi[○], **Mehrdad Mokhtari**[○], **Curtis P. Berlinguette**[○], **Yoshua Bengio**[*,‡]
[*]Mila - Quebec AI Institute [○]University of British Columbia [‡]Université de Montréal

ABSTRACT

The electrochemical reduction of atmospheric CO_2 into high-energy molecules with renewable energy is a promising avenue for energy storage that can take advantage of existing infrastructure especially in areas where sustainable alternatives to fossil fuels do not exist. Automated laboratories are currently being developed and used to optimize the composition and operating conditions of gas diffusion electrodes (GDEs), the device in which this reaction takes place. Improving the efficiency of GDEs is crucial for this technology to become viable. Here we present a modeling framework to efficiently explore the high-dimensional parameter space of GDE designs in an active learning context. At the core of the framework is an uncertainty-aware physics model calibrated with experimental data. The model has the flexibility to capture various input parameter spaces and any carbon products which can be modeled with Tafel kinetics. It is interpretable, and a Gaussian process layer can capture deviations of real data from the function space of the physical model itself. We deploy the model in a simulated active learning setup with real electrochemical data gathered by the AdaCarbon automated laboratory and show that it can be used to efficiently traverse the multi-dimensional parameter space.

1 INTRODUCTION

The electrolysis of atmospheric CO_2 using renewable electricity is an alternative way to store energy that could allow sectors with limited sustainable alternatives (e.g. aviation, marine, chemical industry) to reach carbon neutrality faster (Breyer et al., 2019; Zanatta, 2023). Electrolyzers take CO_2 as input and convert it into valuable chemicals such as carbon monoxide (CO), ethylene (C_2H_4), and ethanol (C_2H_5OH) (Chen et al., 2024). This reaction takes place in a part of the device called the gas diffusion electrode (GDE) whose performance depends on many variables from chemical composition and processing parameters to the choice of the operating conditions, and its optimization is challenging.

Here, we present a physics-based framework for CO_2 reduction that can be utilized for the Bayesian optimization of GDE designs in automated self-driving labs. The framework uses an analytical one-dimensional continuum model for the GDE cathode which is based on the work by Blake et al. (2021). We implemented the model in differentiable PyTorch code and extended it to multi-product reactions. [1] Certain latent parameters of the model which cannot be experimentally measured are inferred from the data. This data-driven analytical model is then used as the mean function of a Gaussian process (GP). The benefit of this hybrid model is twofold: the GP learns a correction to the physics model, and it provides an uncertainty measure to use in Bayesian optimization.

We demonstrate the use of the model on the experimental data collected from AdaCarbon platform, which consists of a team of seven robots that work together to accelerate the fabrication, characterization and testing of multiple GDEs of varying catalytic compositions. (Soni et al., 2025) As key contributions, we (1) present a physics-based framework for *multi-product* CO_2 reduction reactions, (2) which is interpretable and uncertainty-aware; (3) demonstrate the use of the model in simulated pool-based active learning with real experimental data.

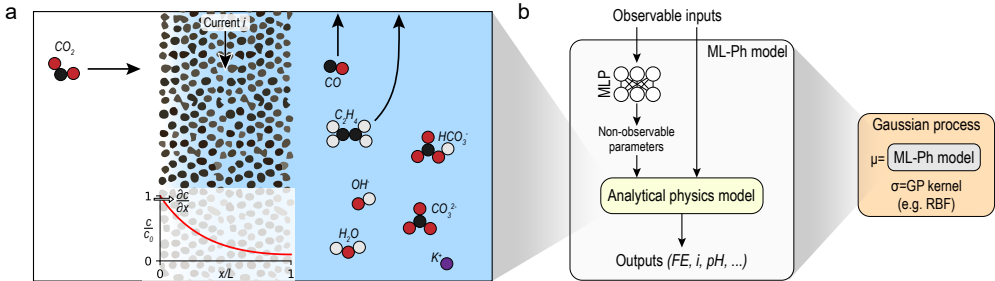[1]Code is available at github.com/igrega348/CO2-catalysis

Figure 1: *a)* Schematic of the analytical model of the cathode. CO_2 dissolves in the electrolyte as it enters the porous catalyst layer. The concentration profile of CO_2 (c/c_0) is indicated on the schematic. Gaseous reaction products leave the system while other molecules and ions are dissolved in the electrolyte and in equilibrium. *b)* The analytical model is embedded in a data-driven framework (ML-Ph) where non-observable parameters are inferred from the data. *c)* The ML-Ph model is used as the mean function of a Gaussian process to enable uncertainty-aware predictions.

2 METHODS

Data structure and model inputs   We operate with a GDE dataset collected from a self-driving laboratory, AdaCarbon (Soni et al., 2025). AdaCarbon is a platform comprising seven robotic modules for automated GDE fabrication and characterization, along with an Automated Test Cell capable of performing zero-gap CO_2 electrolysis. We used this platform to fabricate and test 90 GDEs (30 unique GDEs, tested in triplicate), with varying compositions of Cu-Ag metals and Nafion-Sustainion ionomer bilayers, aiming to enhance ethylene selectivity.

The available per-sample attributes in the dataset are AgCu ratio x_{Ag}, Nafion volume v_{Naf}, Sustanion volume v_{Sus} and catalyst mass loading m. As a pre-processing step, we calculate the thickness of the catalyst layer corresponding to zero porosity as $L_0 = m/(\bar{\rho}A)$ where $\bar{\rho}$ is the mass-averaged density given by

$$\bar{\rho} = (1 - x_{Ag})\rho_{Cu} + x_{Ag}\rho_{Ag}$$

This way we obtain the set of 5 input variables $x_{Ag}, v_{Naf}, v_{Sus}, m, L_0$. There are two output variables: the Faradaic efficiencies with respect to CO and ethylene, FE_{CO} and $FE_{C_2H_4}$, respectively.

In addition to the per-sample attributes, there are further parameters which characterize the operating conditions of the CO2 reduction system which is run at steady-state conditions. The values of these parameters and other assumed physical constants are listed in the Appendix. Notably, the predictions of the model are very sensitive to certain parameters which are not readily available from the experiments (e.g. porosity of the cathode ε, radius of particles r, diffusion coefficient K_{dl}). We employ an MLP with layer dimensionality $[5, 64, 64, 64, 6]$, ReLU nonlinearities and dropout with $p = 0.1$ between the layers to map from the five per-sample input variables to the six per-sample inputs to the physical model ($\varepsilon, r, K_{dl}, \theta_1, \theta_2, \theta_3$). In addition to the per-sample inputs, there are 6 trainable parameters in the model which do not depend on sample attributes. These are the Tafel reaction constants i_i^*, α_i which are defined in the Appendix.

Differentiable inverse solver   In the basic operation, the model takes the cathode voltage along with the inputs as mentioned above and solves for concentrations of products, current densities and the associated Faradaic efficiencies. In our experiments, however, it is the cathode current rather than cathode voltage which is imposed. Moreover, in the current experimental setup the cathode voltage is not measured (only the voltage across the entire cell could be measured). Therefore, given the current state of the model parameters and inputs, the outputs are calculated for a set of 1000 voltages ranging between $-1.25\,\text{V}$ and $0\,\text{V}$. We then find the operating point as a linear interpolation of the 2 voltages which provide the overall current closest to the target value of $I = 233\,\text{mAcm}^{-2}$. This procedure is efficient, enables backpropagation of gradients to the MLP, and proves empirically stable.
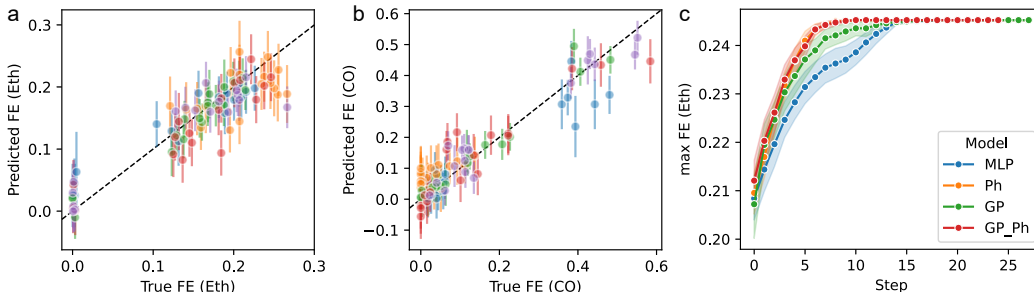
Figure 2: *a,b)* Parity plots over test set for Faradaic efficiency of ethylene and CO, respectively for the GP+Ph model. Error bars are $\pm\sigma$. Colors represent independent trials in 5-fold cross-validation. *c)* Maximum FE in the training pool as a function of step number in simulated active learning.

Uncertainty-aware model In order to use the model in Bayesian optimization of experiments, every prediction needs to come with an associated uncertainty measure. We experiment with two frameworks: (i) model ensembling and (ii) Gaussian process (GP) with embedded physics mean function. The ensembles are trained with a bagging approach whereby each of the $N = 50$ models has access to a fraction $m = 1/3$ of the data during training. At inference time, predictions are obtained from each model independently and variance of the predictions is used as the resulting uncertainty. In the GP with embedded physics mean function, a single physics model is used as the mean function and GP covariance kernel is trained on top of it to provide an uncertainty measure at inference time.

3 RESULTS

3.1 UNCERTAINTY-AWARE PREDICTIONS

As outlined in Section 2, we experiment with two approaches: (i) model ensembling (Ph ensemble) and (ii) Gaussian process with physics mean function (GP + Ph). In addition, we consider two other models: (i) ensemble of MLPs (MLP ensemble) and (ii) constant-mean Gaussian process (GP). Both these models are trained end-to-end with 5 inputs $(x_{\text{Ag}}, v_{\text{Naf}}, v_{\text{Sus}}, m, L_0)$ and 2 outputs (Faradaic efficiency with respect to carbon monoxide and ethylene, FE_{CO} and $\text{FE}_{\text{C}_2\text{H}_4}$, respectively).

Table 1 shows the negative log likelihood and mean average error for the two-target predictions of both CO_2 and C_2H_4 Faradaic efficiencies. Of the four evaluated approaches, the Gaussian process with embedded physics mean function performs best on both metrics. Figure 2a,b shows the parity plots over withheld test set for the GP+Ph model. In all five runs of the 5-fold validation, the predicted values are close to the true values and the predicted uncertainty is well calibrated.
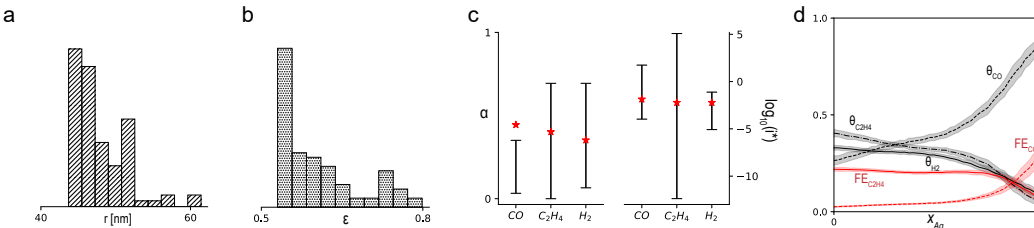

Figure 3: *a,b)* Histograms of inferred particle radius r and porosity ε for our dataset. *c)* Electrode reaction kinetics parameters α and i^*. Vertical bars show ranges of values reported in the literature (Bui et al., 2022) on *Cu* substrates. Red stars show values inferred for our dataset (mixture of Cu *and* Ag substrates). *d)* Faradaic efficiencies and surface coverage parameters θ_i for a virtual x_{Ag} sweep. Shaded bands indicate dropout-enabled variance (Gal & Ghahramani, 2016).

3

Table 1: Negative log likelihood (NLL), mean average error (MAE) and acceleration factor (AF) for the benchmark of 4 models.

	Model	NLL ↓	MAE ↓	AF ↑
Benchmark	MLP ensemble	-1.4 ± 0.3	0.045 ± 0.02	2.0
	GP	-1.5 ± 0.4	0.048 ± 0.01	2.9
Ours	Ph ensemble	-1.5 ± 0.3	0.048 ± 0.01	**3.2**
	GP + Ph	$\mathbf{-1.9 \pm 0.3}$	$\mathbf{0.025 \pm 0.01}$	3.1

3.2 INTERPRETABILITY OF THE MODEL

Certain non-observable parameters of the model (e.g. porosity ε, particle radius r) are per-sample while others (e.g. electrode reaction kinetics) are shared across samples. Characteristics of the inferred parameters are displayed in Fig. 3a,b. Values of particle radius r and porosity ε are in sensible ranges ($40\,\mathrm{nm} < r < 65\,\mathrm{nm}$, $0.5 < \varepsilon < 0.8$). In Fig. 3c we plot the inferred values of electrode reaction kinetics α_i and i_i^* and we compare them to the values reported in the literature (Bui et al., 2022). Note that unlike reaction potentials η_i which are physical constants depending on the molecule formation energies, there is a large variability in values of α_i and i_i^* reported in the literature. Nonetheless, our inferred values fall within the acceptable ranges.[2]

Finally, the analytical model allows for physical interpretability of internal parameters. In a "virtual experiment", we vary the AgCu ratio while keeping all other variables at their mean values. The Faradaic efficiencies change accordingly and we observe that internally the inferred per-sample values of surface coverages θ_i indicate that reactions switch from preferential ethylene production on copper-rich surfaces ($\theta_{C_2H_4} > \theta_{CO}$ at $x_{Ag} \sim 0$) to carbon monoxide production on silver-rich surfaces ($\theta_{C_2H_4} < \theta_{CO}$ at $x_{Ag} \sim 1$) (Figure 3d).

3.3 SIMULATED POOL-BASED ACTIVE LEARNING

We carry out simulated active learning for maximization of ethylene Faradaic efficiency. All available data points are split into 2 pools: training pool and candidate pool. The models are calibrated using the data from the training pool and the next candidate is proposed from the candidate pool based on the acquisition function (here *expected improvement*). Once a data point is selected, it is included in the training pool for the next round of model calibration. See Appendix for more details.

In Figure 2c we plot the maximum Faradaic efficiency over the training pool as a function of acquisition step. The shaded bands correspond to variability over 100 independent trials. It is clear that the models with embedded physics perform better than the pure GP while an MLP ensemble has the worst performance. In Table 1 we include the *acceleration factor* (AF) with respect to random sampling. The ensemble of MLPs is on average $2\times$ more efficient than random sampling while the other three models are approximately $3\times$ more efficient.

4 CURRENT LIMITATIONS AND OUTLOOK

In this work, we developed a physics-based data-driven model for gas diffusion electrodes where non-observable parameters are inferred from experimental data. We have shown that the model can capture multi-product reactions with CO and ethylene products, but it is easily extensible to other carbon products. We carried out simulated pool-based active learning to evaluate the capacity of the framework to guide automated experiments. The Gaussian process (GP) augmented with a physics model achieves best performance. While the performance gain over standard GP is small, more experiments are needed to verify how the models perform in real scenarios with continuous optimization domain. Moreover, the physics-based model provides physical understanding and explainability not found in black-box ML models. In this work we focused on optimization deployed on a specific robot, but we are working on expanding the capabilities for other systems.

---

[2]The plotted range of reported values of α corresponds to copper cathodes while our cathodes are a mixture of copper and silver. Reported values of α_{CO} on silver cathodes are higher.

REFERENCES

J.W. Blake, J.T. Padding, and J.W. Haverkort. Analytical modelling of co2 reduction in gas-diffusion electrode catalyst layers. *Electrochimica Acta*, 393:138987, 2021. ISSN 0013-4686. doi: https://doi.org/10.1016/j.electacta.2021.138987. URL https://www.sciencedirect.com/science/article/pii/S0013468621012779.

Christian Breyer, Mahdi Fasihi, Cyril Bajamundi, and Felix Creutzig. Direct air capture of co2: a key technology for ambitious climate change mitigation. *Joule*, 3(9):2053–2057, 2019.

Justin C. Bui, Eric W. Lees, Lalit M. Pant, Iryna V. Zenyuk, Alexis T. Bell, and Adam Z. Weber. Continuum modeling of porous electrodes for electrochemical synthesis. *Chemical Reviews*, 122 (12):11022–11084, 2022. doi: 10.1021/acs.chemrev.1c00901. URL https://doi.org/10.1021/acs.chemrev.1c00901. PMID: 35507321.

Yuanjun Chen, Rui Kai Miao, Christine Yu, David Sinton, Ke Xie, and Edward H Sargent. Catalyst design for electrochemical co2 reduction to ethylene. *Matter*, 7(1):25–37, 2024.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016.

Irving Langmuir. The adsorption of gases on plane surfaces of glass, mica and platinum. *Journal of the American Chemical Society*, 40(9):1361–1403, 1918. doi: 10.1021/ja02242a004. URL https://doi.org/10.1021/ja02242a004.

Abishek Soni, Siwei Ma, Karry Ocean, Kevan Dettelbach, Daniel Lin, Connor Rupnow, Mehrdad Mokhtari, Christopher Waizenegger, Guiseppe Crescenzo, and Curtis Berlinguette. Accelerated development of gas diffusion electrodes for co2 electrolyzers. *ChemRxiv*, 2025. doi: 10.26434/chemrxiv-2024-znh0f.

Marcileia Zanatta. Materials for direct air capture and integrated co2 conversion: advancement, challenges, and prospects. *ACS Materials Au*, 3(6):576–583, 2023.

A APPENDIX

A.1 SENSITIVITY ANALYSIS

The analytical model provides a detailed view of the micro-environment at the cathode. In Figure 4a we plot the dependence of various parameters on the half-cell voltage with respect to the reversible hydrogen electrode V_{RHE} for a representative set of parameters. As cathode voltage becomes increasingly more negative, the overall cathode current i increases exponentially (as governed by Tafel kinetics). Faradaic efficiencies $\text{FE}_{\text{C}_2\text{H}_4}$ and FE_{CO} first increase but then begin to drop again below $V_{\text{RHE}} \sim -0.5$. This behavior is well known in the literature (Blake et al., 2021) and is due to the excessive production of OH ions from hydrogen evolution, the environment becoming highly alkaline (see the associated increase in pH) and the associated reduction in the solubility of CO_2 in the electrolyte, $CO_{2,sol}$.

We now vary the input parameters of the model and observe the associated change in the output Faradaic efficiency for ethylene, $\text{FE}_{\text{C}_2\text{H}_4}$ (Fig. 4b). Sensitivity of FE to parameter x is defined as

$$\frac{\frac{\partial \text{FE}}{\text{FE}}}{\frac{\partial x}{x}} = \frac{\partial \text{FE}}{\partial x} \frac{x}{\text{FE}}$$

where the derivative $\frac{\partial \text{FE}}{\partial x}$ is estimated numerically by taking small deviations around x. The parameters with largest sensitivity are electrode kinetics parameters, α_i. For instance, a 1% change in
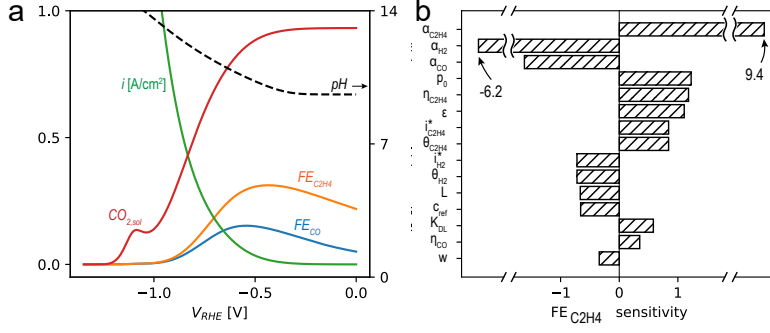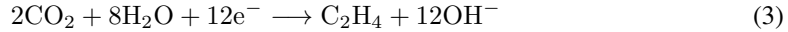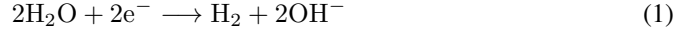
Figure 4: *a)* Dependence of selected quantities on half-cell voltage V_{RHE}. *b)* Sensitivity of ethylene Faradaic efficiency to model parameters.

α_{C2H4} will result in 9.4% change in $\text{FE}_{\text{C}_2\text{H}_4}$. An important observation is that many parameters which are not readily observable in experiments (e.g. $\alpha_i, i_i^*, \theta_i, \varepsilon, K_{DL}$) have a deciding impact on the output Faradaic efficiency. For this reason, we implement a data-driven approach in which we infer these latent parameters from the data.

A.2 DETAILS OF THE PHYSICS MODEL

The following three reactions are assumed to take place at the cathode:

$$2\text{H}_2\text{O} + 2\text{e}^- \longrightarrow \text{H}_2 + 2\text{OH}^- \tag{1}$$

$$\text{CO}_2 + \text{H}_2\text{O} + 2\text{e}^- \longrightarrow \text{CO} + 2\text{OH}^- \tag{2}$$

$$2\text{CO}_2 + 8\text{H}_2\text{O} + 12\text{e}^- \longrightarrow \text{C}_2\text{H}_4 + 12\text{OH}^- \tag{3}$$

The corresponding current densities are expressed using Tafel kinetics with first order dependence on the local concentration of reactants and exponential dependence on overpotentials η_i:

$$i_{H_2} = -\theta_1 i_{H_2}^* \exp\left(-\frac{\alpha_{H_2} F}{RT}\eta_{H_2}\right)$$

$$i_{CO} = -\theta_{CO} i_{CO}^* \frac{c_{\text{CO}_2}}{c_{\text{ref}}} \exp\left(-\frac{\alpha_{CO} F}{RT}\eta_{CO}\right)$$

$$i_{C_2H_4} = -\theta_{C_2H_4} i_{C_2H_4}^* \frac{c_{\text{CO}_2}}{c_{\text{ref}}} \exp\left(-\frac{\alpha_{C_2H_4} F}{RT}\eta_{C_2H_4}\right)$$

The factors $i_{H_2}^*, i_{CO}^*, i_{C_2H_4}^*, \alpha_{H_2}, \alpha_{CO}, \alpha_{C_2H_4}$ are trainable constants. As shown in review by Bui et al. (2022), the range of reported constants i_i^* and α_i varies by orders of magnitudes between different works in the literature. Therefore, we argue that it is reasonable to enable these constants to be trained from data. The competition between the three reactions is captured by *coverage parameters* $\theta_{H_2}, \theta_{CO}, \theta_{C_2H_4}$. These are similar to the surface coverage parameters outlined in (Bui et al., 2022) and introduced by Langmuir (1918). However, we introduce a different framework in which we do not calculate θ_i from temperature and voltage, but rather θ_i are functions of the input parameters (e.g. AgCu ratio). They sum to 1 which we enforce by softmax. For instance, as shown in Figure 3, the model trains to shift the balance of the reactions to ethylene on Cu-rich surfaces and to CO on cathodes without copper.

A.2.1 PHYSICAL CONSTANTS

The nominal surface area of the cathode is $A = (1.85\,\text{cm})^2$, the overall current density is $i = 233\,\text{mAcm}^{-2}$, CO_2 is fed from a fixed pressure reservoir with $p_0 = 2.38\,\text{bar}$ at a flow rate of $Q = 30\,\text{cm}^3/\text{min}$. Flow channel dimensions are $1.5\,\text{mm} \times 0.5\,\text{mm} \times 20\,\text{mm}$.

A.3 CALIBRATION OF MODEL ENSEMBLES

When training ensemble models, we need to choose 2 hyperparameters: the number of models in the ensemble and the fraction of the data that is seen by each model. We carry out a hyperparameter
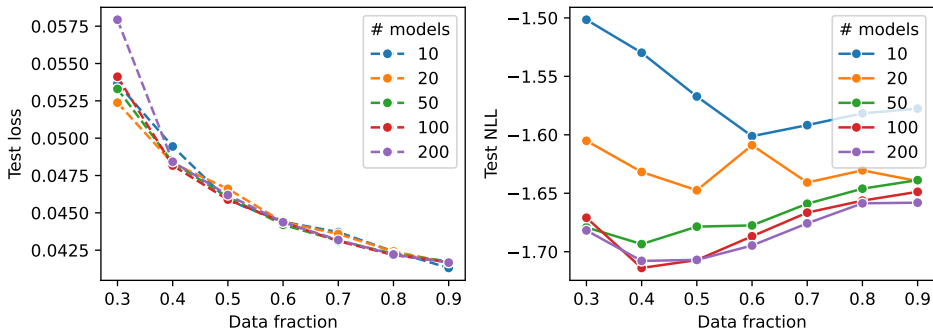
Figure 5: Hyperparameter exploration of model ensembles. Test loss and test negative log likelihood (NLL) for various data fractions and numbers of models.

sweep for MLP ensemble in which we include 10, 20, 50, 100, 200 models and the data fractions 0.3, 0.4, ..., 0.9. In Figure 5 we plot the test loss and test negative log likelihood for this parameter sweep. A few interesting observations emerge:

1. Based on test NLL, it is beneficial to include more models, but the gains are negligible beyond 50 models. Therefore, we select 50 models for the ensembles.

2. As the data fraction fed to each model increases, test loss continues to reduce. However, variance between model prediction reduces (models become increasingly confident) which is detrimental to NLL predictions. We select a data fraction of 0.33.

A.4 SIMULATED POOL-BASED ACTIVE LEARNING

All available data points are split into 2 pools: training pool and candidate pool. The models are calibrated using the data from the training pool and the next candidate is proposed from the candidate pool based on the *expected improvement* acquisition function. Since the data were acquired based on triplets, we follow the same principle in this method. At the start, the training pool contains 3 triplets. The candidate pool is always evaluated based on triplets. Note that 3 out of 5 input variables are identical for all points in the triplet. The remaining 2 variables (catalyst mass loading, zero porosity thickness) are averaged over the three data points in the triplet. Once a triplet is selected, all three points with their respective individual values of input variables are included in the training pool for the next round of model calibration.

Acceleration factor is defined as the ratio between the expected number of optimization steps with random candidate proposal and the corresponding expected number with the given strategy. Note that for a candidate pool size of n, the expectation for random proposal is $(n-1)/2$. We operate with an initial candidate pool size of $n = 27$. Therefore, $AF = 13/\mathbb{E}[N]$ where $\mathbb{E}[N]$ is the mean number of steps needed to get to the maximum FE and for the given strategy.