ACS
Applied Catalysis

Open Access
This article is licensed under CC-BY 4.0

pubs.acs.org/acsappliedcat
Article

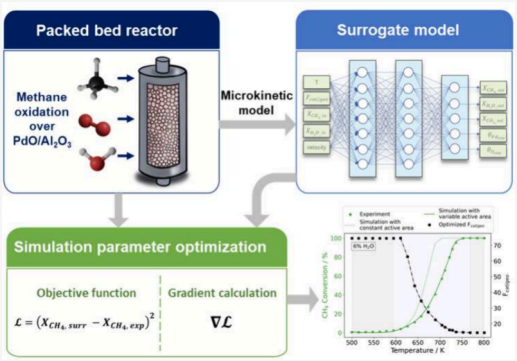
Surrogate Modeling of Catalytic Packed Bed Reactor Using Deep Learning Architectures and Its Application in Modeling Catalyst Performance

Rinu Chacko, Markus Götz, Sofia Angeli,* and Olaf Deutschmann

Cite This: https://doi.org/10.1021/acsappcat.6c00044
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The field of machine learning is evolving rapidly, and deep learning–based surrogate modeling has emerged as an efficient approach for accelerating numerical studies. In this work, multiple deep learning architectures were investigated as surrogate models for simulating methane oxidation in a packed-bed reactor. Among the tested architectures, a Long Short-Term Memory (LSTM)-based model demonstrated superior performance on unseen data and was able to capture the underlying relationships within the data set. The surrogate model was subsequently employed for a gradient-based optimization task to accurately simulate the deactivation behavior of palladium-based catalysts when exposed to oxidizing, humid conditions.



KEYWORDS: *Surrogate Modeling, Parameter Optimization, Gradient-based Optimization, Kinetic Modeling, Methane Oxidation*

INTRODUCTION

A core objective of catalysis research is to design high-performance catalysts while deepening the understanding of the underlying mechanisms that govern their behavior.[1,2] Multiscale modeling approaches are valuable for this purpose, as they aim to capture catalytic phenomena across relevant length and time scales. Data-driven methods utilizing machine learning and deep learning techniques are increasingly being used for supporting both catalyst design and gaining mechanistic insight within multiscale modeling frameworks.[1,3−6] Such models aim to extract knowledge from available data and construct models that learn intrinsic patterns, and in the context of supervised learning, determine the relationships between input features and target outputs.[6−8]

The complex and nonlinear nature of catalytic phenomena requires model architectures capable of capturing the intricate dependencies within the system. At the same time, the rapid expansion of machine learning as a field has greatly broadened the range of such advanced models available for predictive tasks. The fundamental criterion for selecting an appropriate architecture from the wide range of available options, is to consider both the complexity of the underlying problem and the characteristics of the data. Catalyst performance data are typically available as axially resolved, sequential data measured or simulated along the length of the catalytic channel, covering multiple species and reactor conditions. Such structured, sequential data sets naturally motivate the use of architectures that are designed to handle temporal dependencies. Deep learning architectures such as, Recurrent neural networks,[9] transformer-based architectures[10] and the more recent structured state-space models (SSMs)[11−13] represent suitable options for handling data of this nature.

Depending on their architecture, these models can function as low-cost predictive tools, and when interpretability is built in, as a means of elucidating the underlying structure of the learned relationships.[8,14−19] As predictive models, they are particularly valuable in investigations where computational cost is a limiting factor. Once trained, they can be evaluated at a fraction of the computational cost required for full physics-based models, making them attractive alternatives to expensive simulations; for this reason, they are commonly referred to as surrogate models.[20−24]

Several studies have demonstrated the use of such surrogate models to approximate the catalytic activity of catalytic materials for achieving acceleration.[18,24−31] Recently, Döppel et al.[32] developed a physics-informed neural network (PINN) model for a catalytic channel reactor, incorporating a kinetic surrogate model that approximated species source terms for given gas-phase compositions and temperatures. The authors

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906539/4/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

ACS Publications
© XXXX The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

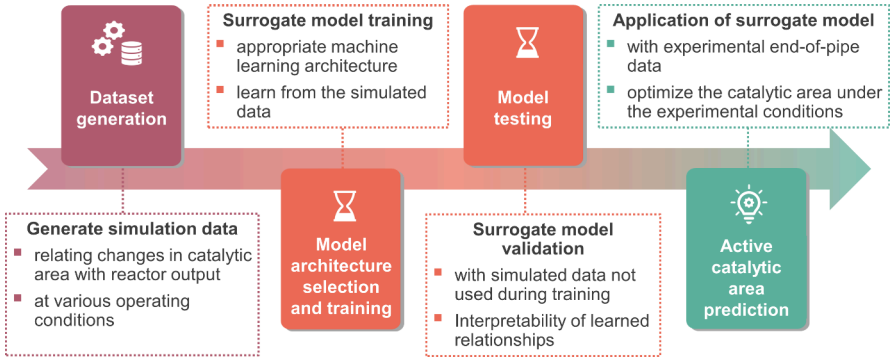


Figure 1. Schematic representation of the surrogate model development and application workflow.

reported that the resulting model achieved a computational cost roughly 1,000 times lower than that of a full numerical simulation of the catalytic channel performed with Cantera.[32,33]

Once the surrogate models are trained accurately, they provide not only a substantial reduction in computational cost during prediction but also convenient access to gradient information. Such gradients are often unavailable or prohibitively expensive to compute in traditional numerical simulations, depending on the complexity of the governing equations and the dimensionality of the system. Access to gradients enables interpretation of the learned relationships and can be leveraged to estimate or optimize unknown parameters.[34−37] Furthermore, surrogate models are particularly valuable in scale-up or optimization studies that require repeated evaluations of reactor behavior, where direct numerical simulations might be too slow or computationally intensive.[26,38,39]

In light of these advantages, in this work, a surrogate model is trained and its gradients are leveraged to capture the dynamic change in catalyst properties under harsh operating conditions, in order to simulate them more accurately and better understand the experimental behavior. We first evaluate three deep learning architectures suitable for sequential data sets namely, Long Short-Term Memory (LSTM),[40] Transformer,[10,41] and Mamba[13] and compare their performance as surrogate models for simulating the oxidation of methane in a catalytic packed-bed reactor operating under lean conditions. The most accurate surrogate model is then employed to study the dynamic deactivation behavior typically observed experimentally in palladium-based catalysts for methane oxidation under oxidizing and humid environments,[42−47] by exploiting the gradient information provided by the surrogate.

METHODS

Experimental Data for Optimization

The experimental data used in this work are taken from the study of Keller et al.,[48] who studied methane oxidation over PdO/Al_2O_3 in a packed bed reactor under lean, humid conditions. The experimental procedure is described here briefly. Catalytic tests employed varying H_2O concentrations (0−12 vol %) and a 3 K min^{-1} temperature ramp between 500 and 825 K. Before testing, catalyst samples were degreened at 825 K under reactive conditions (3200 ppm of CH_4, 10 vol % O_2 in N_2) to ensure comparable initial properties across samples used at different condtions.

Tests were carried out in a packed bed reactor with 300 mg PdO/Al_2O_3 (catalyst loading 2.3%) and diluted with 700 mg quartz to avoid hot-spots. The catalytic bed length was 1.5 cm, the channel diameter 8 mm, and the bed porosity about 56%. The average particle diameter was 200 μm. TEM measurements showed an average PdO nanoparticle size of 1.7 nm.

Accounting for Available Catalytic Surface Area in Simulations

In this work, the packed bed reactor code ($DETCHEM^{PBR}$) in the DETCHEM simulation package[49,50] is used for simulating the reactor. Within DETCHEM,[49] the active surface area in porous catalysts is accounted for via the $F_{cat/geo}$ factor,[51] which is applied as a multiplier to the surface reaction rate during the simulation. This factor is defined as the ratio of the catalytic surface area available for reaction (A_{cat}) to the geometric surface area (A_{geo}) as given in eq 1.

$$F_{cat/geo} = \frac{A_{cat}}{A_{geo}} \tag{1}$$

It should be noted that a change in $F_{cat/geo}$ value can reflect a range of phenomena affecting catalyst activity and should not be interpreted exclusively as changes in the physical surface area of the catalyst particles. The catalytic surface area in the packed bed is estimated from the average nanoparticle diameter of the active phase d_{np}, through eq 2, assuming spherical particles[52] with half of their surface exposed to the bulk.

$$A_{cat} = 0.5\frac{m_{activespecies}}{\rho_{activespecies}}\frac{6}{d_{np}} \tag{2}$$

Here, $m_{active\ species}$ is the mass of the active phase, determined from the catalyst loading, and $\rho_{active\ species}$ is its density. The geometric area for a cylindrical packed bed reactor is calculated through eqs 3 and 4 based on the surface area per unit volume of the solid fraction in the catalyst bed[53] such that

$$A_{geo} = \frac{6(1-\varepsilon)}{d_p}V_{bed} \tag{3}$$

$$A_{geo} = \frac{3\pi}{2}(1-\varepsilon)\frac{D^2L}{d_p} \tag{4}$$

where ε is the bed porosity, D is the diameter of the reactor, L is the length of the reactor and d_p is the average diameter of particles in the catalyst bed.

Surrogate Modeling and Application

A schematic representation of the methodology followed in this work is shown in Figure 1, which is used to obtain a reliable surrogate model and its subsequent application in studying the deactivation behavior of PdO catalysts under humid conditions. In particular, for the application of the surrogate model in studying the deactivation behavior, the $F_{cat/geo}$ parameter used as the simulation input is critical, as it relates to the available catalytic area at any given reaction condition, which changes as the catalyst undergoes deactivation during the experimental measurement. The following subsections outline each step in detail.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

Simulation Data Generation and Storage. A crucial requirement in the development of a neural network-based surrogate model is the availability of a high-quality data set for training. Since the surrogate model needs to simulate the reactor outputs on the basis of the inlet composition and operating conditions alone, the model must learn the inherent kinetics from the data itself, a data set was generated through simulations of the packed bed reactor model with variations in several input parameters such as temperature, water concentration in the inlet, and the $F_{cat/geo}$ parameter representing the catalytic surface area (Table 1).

Table 1. Details on the Variation in Input Parameters during the Data Generation

Input parameter	Range of values and units
Temperature	450−900 K
H_2O mole fraction in the feed	0−15 vol %
$F_{cat/geo}$	5−100% of reference value (78.48)

The data set comprised the mole fractions of the gas-phase species and the surface coverages of the surface species, along the length of the reactor, generated from running the simulations using the DETCHEM^{PBR} code[50] with the detailed kinetic model for PdO/Al_2O_3[54] under varying inlet and reactor operating conditions. Hence, the data needed to be stored as three-dimensional arrays of the following shape: [number of sample points (N), number of length steps (L), number of variables (D)]. The first dimension of the array corresponds to the total number of simulations with variations in any of the input parameters. The second dimension in the data set corresponds to the resolution of the axial coordinate. In this work, data were captured at every 0.5 mm of the reactor length. The third dimension corresponds to the number of variables (or features) in the data set, namely, the temperature, the mole fractions of the gas-phase species, the coverages of the surface species, pressure, velocity, residence time, and the $F_{cat/geo}$ value at each axial point within the reactor. In this work, N = 68614, L = 31, and D = 33.

In order to reduce the complexity in the data set and focus the surrogate model on learning the dependency of $F_{cat/geo}$ on the output, certain input parameters were kept constant. These included the input mole fraction of reactants other than H_2O, the porosity of the catalytic bed, the reactor diameter, the reactor length, the particle diameter, and the pressure. A Python-based[55] script was used to parallelize and automate the various steps in the generation of the data set including, the generation of the input files needed for DETCHEM^{PBR}, the execution of the simulations, the extraction of the relevant data from the output files, and the storage of the data in an appropriate data format to make them compatible for training the surrogate model and ensure adherence to FAIR principles.[56] Overall 68,614 simulations were performed to generate the complete data set containing 2,058,420 axially resolved data points. The simulations were performed using an AMD Ryzen Threadripper 3990X processor, leveraging 10 cores in parallel for computations, and required approximately 1.5 h to generate the data set. The data set was stored in Hierarchical Data Format version 5 (HDF5)[57] to allow data slicing during batch training.

Surrogate Model Training and Validation. To model catalytic performance in an isothermal packed-bed reactor, the species mass balance and continuity equations are solved numerically for specified inlet compositions, flow conditions, and reactor geometries. Reaction rates enter as source terms derived from kinetic models, which may range from simplified lumped models to detailed microkinetic mechanisms. Surrogate models trained to emulate such simulations learn to predict axial concentration profiles for all species on the basis of the given inlet feed and operating conditions. Because these profiles are sequential and only the concentrations at the reactor inlet are known, model training requires an autoregressive strategy in which the surrogate model uses its own prediction at the previous axial position to predict the composition at the next axial position. This makes training challenging as errors can accumulate along the reactor length, necessitating architectures with sufficient complexity to handle both the autoregressive nature of the task and the inherent complexity of the chemical kinetics and transport behavior. Considering this and the sequential nature of the axially resolved data, we used model architectures that are useful in handling multivariate time-series data, i.e., LSTM,[40] autoregressive self-attention Transformer,[10,41] and Mamba.[13] It should be emphasized that nonsequential model architectures may also be useful in learning the reactor behavior, but models designed for capturing sequential dependencies mirror the actual physics-based reactor simulations and enable them to capture the progression of the concentrations and coverages along the length of the reactor. Moreover, as the sequential models learn the history of the output variable at each step, they may be able to capture the underlying kinetic relationships or reactor geometry effects that are otherwise missing from raw data.

All three architectures were used for surrogate model training to find the best performing architecture. The following procedure was followed for all architectures involving the steps related to data preprocessing, model training, and validation. The model input consisted of the 33-dimensional array at z = 0, while the remaining sequence was used to compute the loss function during the autoregressive training. The models were built and trained using the Pytorch library[58] for the Python programming language.[55] In the data preprocessing step, all data were standardized using eq 5 to scale the data to have a mean value of 0 and a standard deviation of 1. This step helped ensure that all features are on the same scale during the model training.

$$X_{std} = \frac{X - \mu}{\sigma} \tag{5}$$

In eq 5, X_{std} is the standardized value, X is the original value of the sample point for a given feature, and μ and σ are the mean and standard deviation of all data points for that feature, respectively. Before training, the data set was split in an 80:20 ratio to obtain training data and validation data, respectively. The validation data was further split in a 50:50 ratio to obtain a test data set, reserved as unseen samples on which the final model performance was tested. The validation set excluding the test set was used to evaluate the model after each training loop. The test set is used to estimate the generalization error of the model.

The training procedure followed was as follows: All models were trained for at least 50 epochs where data were fed in batches. After each epoch the batches were shuffled to avoid overfitting. During each epoch, the prediction of the model during the forward pass were compared with the actual simulated values using the mean squared error (MSE) loss function as defined in eq 6, where $(\hat{y}_{n,l,d})$ is the predicted and $(y_{n,l,d})$ is the ground truth value for all N samples, L length-steps, and D features. The objective of the model training is to minimize this loss function so that the model predictions closely match the values obtained from direct simulations using DETCHEM^{PBR}.

$$\mathcal{L} = \frac{1}{N \cdot L \cdot D} \sum_{n,l,d} (\hat{y}_{n,l,d} - y_{n,l,d})^2 \tag{6}$$

Based on the gradients obtained through the loss function during the backward pass, the Adam optimizer[59] was used to update the model parameters. The model weights were saved as checkpoints after each epoch and as the best model whenever the loss improved, ensuring that progress was not lost in case the model training was interrupted. The complete model training was conducted autoregressively, that is, only the model predictions were used as inputs for making predictions after the inlet data was provided in the first length-step. After each training loop, the model performance was evaluated on the validation data set. Once the losses converged, model training was considered to be complete.

Each architecture has a set of hyperparameters that are fixed before training is started. However, the optimal values for these hyperparameters are not known beforehand, hence, multiple combinations of these hyperparameters need to be tested to find the optimal set, a

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

process known as hyperparameter tuning. In this work, the study of several distinct architectures with multiple hyperparameters made a complete grid search for hyperparameter tuning computationally impractical. Instead, a manual search was performed within a predefined range of values of hyperparameters by iteratively refining hyperparameter combinations based on model performance. During hyperparameter tuning, only the validation set was used to test performance. The model with the best performance on the validation set was used for inference on unseen data and compared to other architectures to obtain the final surrogate model.

The surrogate model training was performed on the high-performance computing cluster "Helmholtz AI computing resources (HAICORE)@KIT". The deep learning models were built using the libraries `Pytorch 2.3.1`,[58] `mambapy 1.2.0`, and `mamba-ssm 2.2.2` in Python.[55] Other additional Python packages such as `matplotlib 3.7.5`,[60] `NumPy 1.24.4`,[61] `pandas 2.0.3`,[62] and `h5py 3.11.0`[57] were used for processing of data and plotting. The Distributed Data Parallel (DDP) in `Pytorch` was used to parallelize model training on multiple Graphics Processing Units (GPUs) to decrease model training time.

Surrogate Model Analysis. The best surrogate model obtained among the investigated architectures was analyzed using Captum[63] to understand how the model made its predictions, i.e., which features had the most influence on determining the model outputs. The Integrated Gradients method[64] with a baseline of 0 representing the mean value was used to obtain the feature attribution vectors for various data samples with varying feed compositions and temperature. To visualize these attribution values, the mean attribution values were calculated for samples corresponding to each water concentration in the feed (158 total samples across 7 water concentrations). These averaged attribution profiles were then projected into a two-dimensional space using the Uniform Manifold Approximation and Projection (UMAP) method[65] (`n_neighbors` = 3, `min_dist` = 0.3, `metric` = `euclidean`). To assess robustness with respect to the random initialization of the UMAP embedding, the analysis was repeated using five different random seeds. The resulting embeddings were clustered based on their similarity using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm[66] (`min_cluster_size` = 3, `min_samples` = 1).

Surrogate Model Application: Optimization of Active Catalyst Area. Several studies[42−47] have shown that palladium-based catalysts suffer from deactivation when they are exposed to humid, oxidizing conditions during methane oxidation. This behavior was also experimentally observed in the light-off curves of PdO supported over alumina catalyst in a packed bed reactor under lean, humid operating conditions, in the study of Keller et al.[48] To accurately simulate the light-off curves in light of the deactivation behavior due to continuous exposure to water and varying temperature conditions, the change in the catalytically active area needs to be estimated. We used the trained surrogate model to optimize this unknown value of the catalytically active area which relates to the input $F_{cat/geo}$ parameter in the simulations, so that the experimental trends could be accurately reproduced, and investigated the impact of the reaction conditions on the catalytically active area.

The optimization process was similar to the training of a neural network architecture, where the optimizer learned to update the input parameters whose values are unknown, instead of the model parameters as done during a typical model training. In this scenario, the loss function acted as an objective function for minimization of error between the surrogate predictions and actual experimental outputs. Therefore, during the optimization process, the surrogate model weights were kept constant and the input $F_{cat/geo}$ parameter was defined as a trainable parameter, which would be updated using the optimizer. The experimental data measured at the end of the reactor, i.e., at z = 0.015 m for a given reactor temperature was used as the ground truth and to calculate the loss function as expressed in eq 7, where $X_{CH_4,surr}$ is the value of CH_4 mole fraction predicted autoregressively by the surrogate model based on the inlet data provided as inputs, and $X_{CH_4,exp}$ is the experimentally measured mole fraction of CH_4 at the end of the reactor.

$$\mathcal{L} = (X_{CH_4,\mathrm{surr}} - X_{CH_4,\mathrm{exp}})^2 \tag{7}$$

The Adam optimizer,[59] used during the surrogate training, was also used here for updating the $F_{cat/geo}$ value after each iteration. The total number of iterations for the optimization were kept to be 2000. The optimization process was performed for every temperature and inlet feed condition from 1−12 vol % H_2O. For each feed condition, at least 20 temperature points were considered for optimization between 500 and 800 K. The optimization process was stopped either when the set number of total iterations exceeded or when the loss value was lower than a given threshold value (in this work, threshold = 10^{-5} with the scaled data).

RESULTS AND DISCUSSION

Surrogate Model Performance across Various Architectures

In this section, the performance of the various deep learning architectures is compared on unseen samples of data that were not used in the model training and validation loops. Table 2 shows the range of hyperparameters tested for each architecture and their values in the best performing model. The purpose of each hyperparameter may be found in the Supporting Information. As a preliminary test, a feed-forward neural network with a single hidden layer (64 neurons) was also evaluated following the same data preprocessing, training and validation strategy used for the sequence-based architectures described in the Methods section. Although this architecture provided faster training due to substantially lower number of learnable parameters (4256) compared to the other architectures investigated in this work, its validation loss (1.301) was much higher than that of the sequence-based models (<0.3), indicating that it was unable to adequately capture the spatially correlated species profiles present in the data set. This is in part due to the limited learning capacity of the model with only a single hidden layer of 64 dimensions, and it arguably follows that a larger or deeper model could achieve better accuracy. However, the aim of this work is to investigate sequential architectures and their predictive performance as a surrogate to DETCHEM^{PBR} and hence, a complete hyperparameter tuning was not performed for the standard neural network. The following subsections showcase the results obtained with the sequence-based architectures and the optimal set of hyperparameters.

Table 2. Range of Values of Various Hyperparameters Used during the Hyperparameter Tuning

Architecture	Hyperparameter name	Range	Best value
Transformer	e_layers	2, 3	2
	factor	5, 10	10
	d_model	64−256	64
	n_heads	4, 6, 8	8
	d_ff	64−512	64
Mamba	n_layers	2	2
	d_state	64−256	64
	d_model	64−512	256
	d_head	8−64	64
LSTM	n_layers	1−4	2
	d_model	64−512	256

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906539/4/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

Autoregressive Self-Attention Transformer. This model architecture aims to capture complex dependencies in multivariate time-series data by using cross-dimension self-attention to model interactions across the features for each point in the sequence.[10,41] The architecture contains five hyperparameters that need to be tuned to obtain the best-performing model. The best model had overall 418,883 trainable parameters and took approximately 47 h to train on 2 GPUs. Figure 2 demonstrates the variation in the MSE loss across all training and validation samples with each epoch.

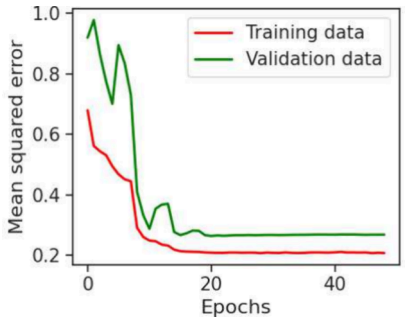


Figure 2. Training history of the transformer model with best set of hyperparameters.

The model losses converged after 20 epochs, and the MSE loss on the validation data was calculated to be 0.263 corresponding to a root-mean-square error of approximately 0.51. Since the target variable is standardized (i.e., has unit variance), the error indicates that the model predictions deviate from the true values by about half of a standard deviation on average. Compared to a naive baseline predictor, this represents a substantial reduction in error, indicating that the model captures meaningful structures in the data. However, the significant difference between the training loss and the validation loss suggests that the model could not generalize well for unseen data.

This is more clearly seen in Figure 3, which shows the transformer model predictions compared to the ground truth, i.e., the simulated axially resolved values obtained with DETCHEM^{PBR} simulations with the chemical model[54] for a subset of features. Figure 3 shows this comparison for a randomly selected set of samples not used during training or validation. In the nomenclature of the surface species, "a" represents the Pd_{cus} sites and "b" represents the O_{cus} sites of the supported PdO catalyst as described in the Supporting Information. The comparison between the numerically simulated and predicted axial curves show that the transformer model struggled to learn the complex reaction kinetics in the packed bed reactor. Although the reactor temperature and the $F_{cat/geo}$ parameter are fixed input parameters in the simulation model, here in the transformer these are still predicted along the length with the rest of the outputs. This is because this model architecture does not support using input and output sequences of different sizes. As can be observed in the Figure 3,

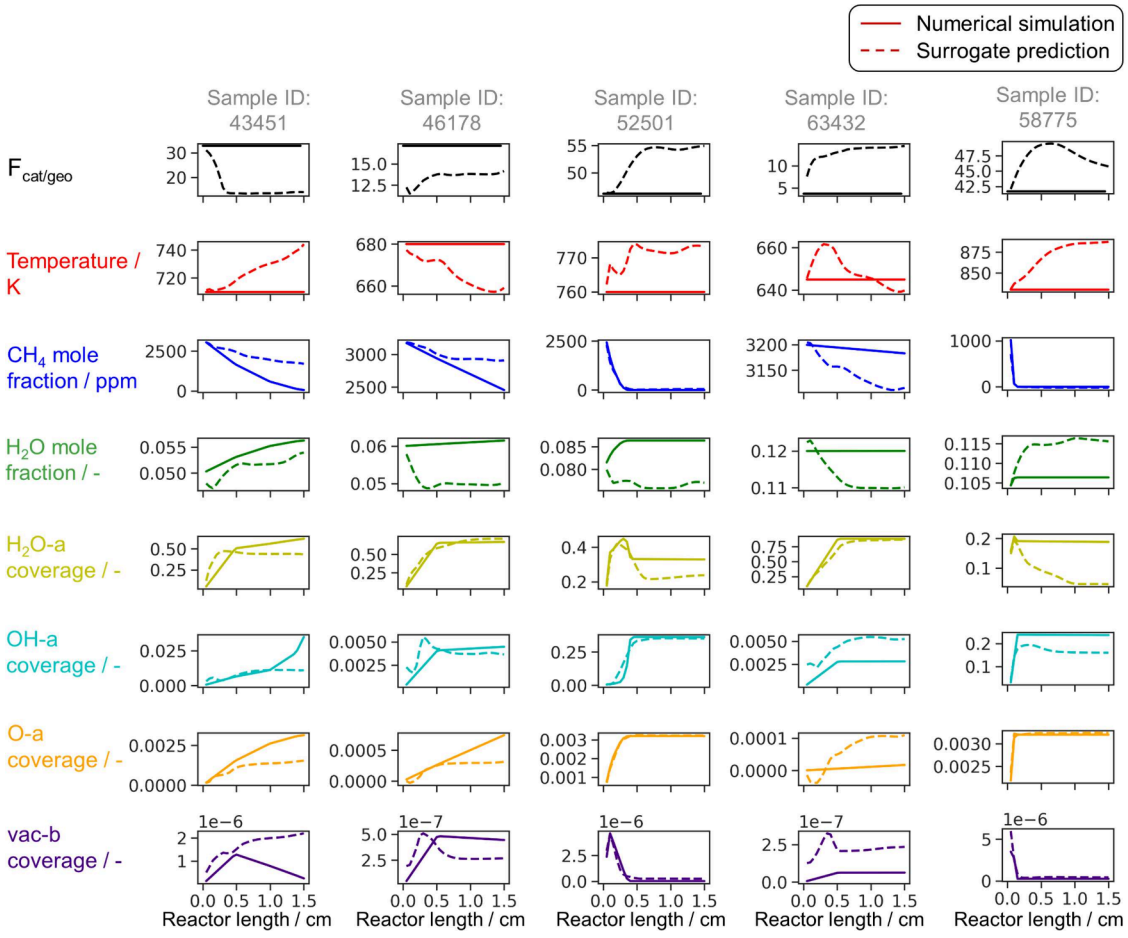


Figure 3. Comparison of the ground truth with the predictions from the autoregressive self-attention transformer model on unseen data.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

the model does not accurately predict the isothermal behavior and the constant $F_{cat/geo}$ values. The model was able to predict the methane mole fractions under the conditions of full conversion quite well but performed poorly at temperatures where conversion is low. Additionally, it was unable to capture the water production trends, leading to the conclusion that the autoregressive self-attention transformer architecture was unsuitable as a surrogate to the numerical simulation model.

Such transformer-based model has been shown to work well on multivariate data,[41] but it relies on seeing a larger window of the sequential data (24 and higher) for accurate performance as compared to this work where the input window was kept as 1 due to the autoregressive nature of the simulation task. It is possible that the architecture was not well-suited for regression tasks of this nature when the window was equal to 1 or the model lacked sufficient complexity to capture the underlying trends in the data. The significant training time, despite parallel training on multiple GPUs, further constrained the exploration of deeper network configurations.

Mamba. The Mamba architecture builds upon structured state space models (SSMs) and introduces a selective state space mechanism for efficient sequence modeling.[11,12] It was proposed as an alternative to the attention-based transformer architecture,[10] offering linear-time sequence processing while retaining strong performance on large sequential data sets.[13] Figure 4 shows the progression of the loss values with the training and validation data sets during the model training with best set of hyperparameters (Table 2).

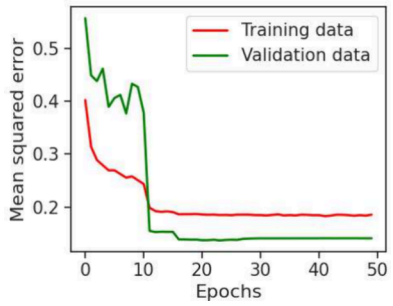


Figure 4. Training history of Mamba model with the best set of hyperparameters.

The validation loss at the beginning of the training was noisy, likely due to the autoregressive training approach. However, both the training and validation losses converged after around 20 epochs. The minimum validation loss associated with the best model was 0.1366. The total number of trainable parameters in the best model was 880,719 and the model took approximately 17 h to train on a single GPU. The time taken for training the Mamba model was significantly less compared to the transformer architecture, as its hardware-aware implementation makes its training on GPUs significantly faster than transformer-based architectures. Additionally, the Mamba model showed a lower mean-squared error value on the validation data set compared to the transformer architecture.

The performance of the best model on the testing set samples are shown in Figure 5. Since Mamba could be configured to have different sizes for the input and the output sequences, the features, namely, the reactor temperature and the $F_{cat/geo}$ parameter, were used only as inputs and not predicted with the outputs. The results show that similar to the transformer model, the Mamba model struggled to capture the water production trends, particularly at temperatures where conversion is low. However, the coverages of certain surface species, such as O-a and H_2O-a, were predicted to be reasonably well. While the model correctly captured the overall trend of methane conversion, it underpredicted the outlet concentrations for some of the samples. Therefore, the Mamba model was also deemed unsuitable for use as a surrogate for DETCHEM^{PBR}. The reasons for the inadequate performance remain unclear, given the black box nature of the model and its limited interpretability.

LSTM. An LSTM network models sequential data by using gated memory cells to capture and retain long-term temporal dependencies. In comparison to the other models, the LSTM architecture has the lowest number of hyperparameters, namely `n_layers` relating to the number of LSTM layers and `d_model` relating to the dimensionality of the hidden state. During LSTM training, reactor temperature and $F_{cat/geo}$ were only considered as inputs to the model and not predicted. In order to improve the performance of the model, a penalty was also imposed during training on the error in prediction of the mole fractions of methane and water, hence the loss function for LSTM model training had an additional penalty term as expressed in eq 8, where $\lambda = 10$, is the penalty weight,

$$\mathcal{L} = \frac{1}{N \cdot L \cdot D} \sum_{n,l,d} (\hat{y}_{n,l,d} - y_{n,l,d})^2 + \lambda \cdot \frac{1}{N \cdot L} \sum_{n,l} [(\hat{y}_{n,l,CH_4} - y_{n,l,CH_4})^2 + (\hat{y}_{n,l,H_2O} - y_{n,l,H_2O})^2] \quad (8)$$

Figure 6 compares the penalized training and validation loss values after each epoch during the model training with the best set of hyperparameter values, as tabulated in Table 2. The best model had 832,287 trainable parameters and took approximately 6.5 h to train on a single GPU. The model converged after 25 epochs with a minimum penalized validation loss of 0.155. The train and validation losses showed similar trend and converged values suggesting that the model was trained well. The model training with the LSTM architecture was easier than the other architectures investigated in this work, due to short sequence lengths in the data set and simpler model architecture.

Figure 7 shows the predictions of the LSTM model on samples from the test set in comparison with the ground truth. The results show that the LSTM model performed the best on unseen data samples compared with the transformer and Mamba models. The additional penalty term contributed to more accurate predictions of the methane and water mole fractions. Overall, the LSTM model was able to reproduce the general spatial trends under most conditions. Although the model exhibited slightly reduced accuracy at low conversion temperatures, the associated errors between the prediction and ground truth were minimal. It should be noted that the comparison plots are scaled to emphasize discrepancies, making even small differences appear to be more pronounced.

Because the final layer in the architecture is a "linear" layer, the surrogate model can theoretically predict negative concentrations or coverages. In practice, such values were observed only for species with very low coverages at the reactor inlet region (for example, in Figure 7, vac-b species coverage for Sample ID: 63432) and did not occur for the reactor outlet values relevant to the present application.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

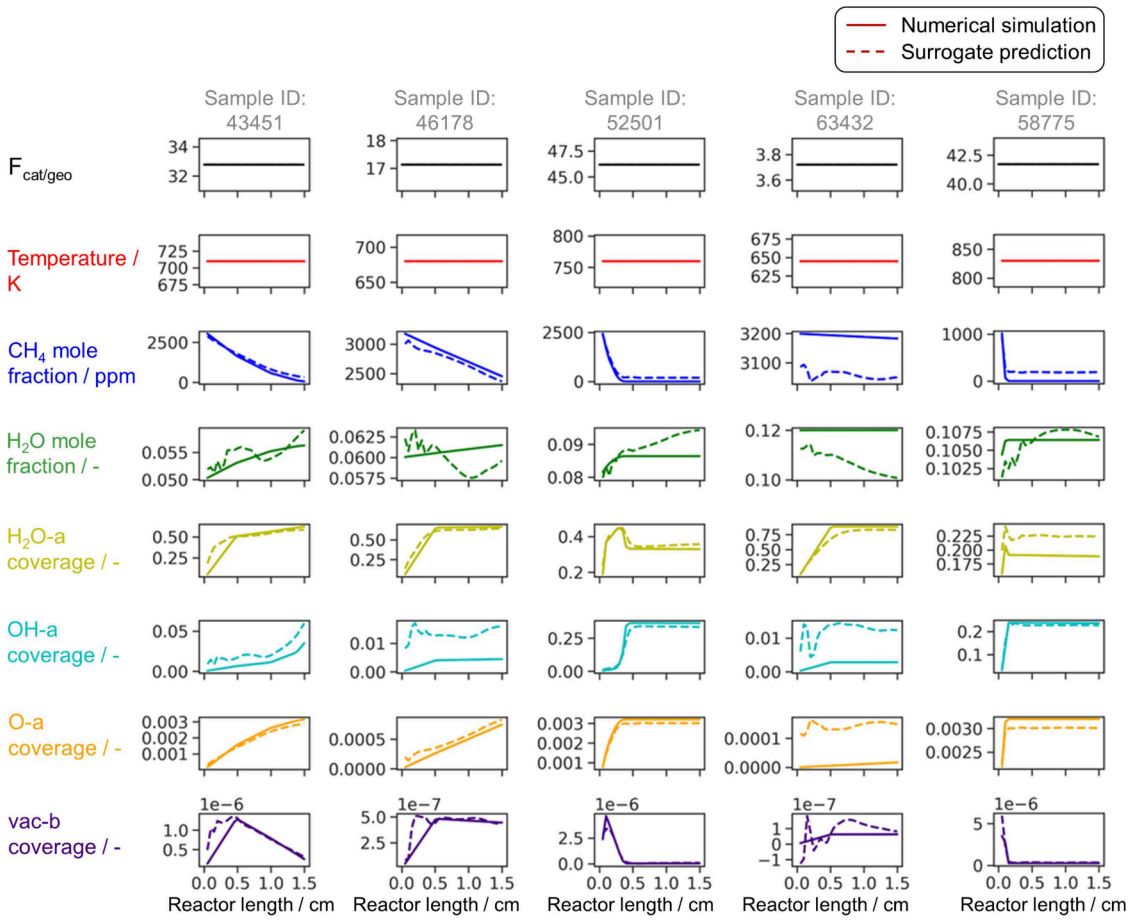


Figure 5. Comparison of the ground truth with the predictions from the Mamba model on unseen data.

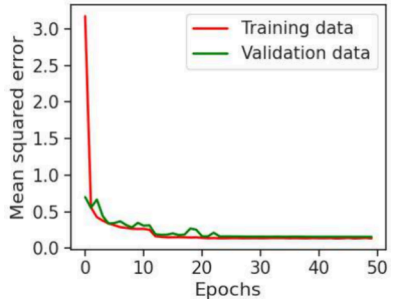


Figure 6. Training history of LSTM model with the best set of hyperparameters.

Furthermore, while mass balance constraints were not explicitly enforced during training, the LSTM model reproduced the corresponding mass balance relationships with mean absolute errors below 10^{-8}, indicating that the model learned the conservation behavior implicitly from the training data. To guarantee stricter adherence to the conservation relationships and prevent nonphysical outputs, the corresponding constraints could be incorporated explicitly into the training loss function. Based on the performance on the unseen samples, the LSTM architecture with 2 layers and hidden state of size 256 was shown to be the best candidate for a surrogate model, among the studied architectures. The LSTM-based surrogate reduced the average computation time from 42.9 ± 24.26 s for the DETCHEM^{PBR} model to 0.037 ± 0.01 s per sample, resulting in an approximate 1150-fold acceleration while maintaining predictive accuracy.

It should be noted that under the plug-flow assumptions employed in the reactor model simulations, the $F_{cat/geo}$ parameter effectively scales the reactor coordinate and thus reduces the complexity of the machine learning task. However, it was intentionally used as an explicit surrogate model input to be consistent with the inputs used in the DETCHEM^{PBR} model. Furthermore, $F_{cat/geo}$ is a physically meaningful reactor design parameter, and its explicit inclusion allowed the surrogate model to provide gradients, enabling the subsequent gradient-based optimization of $F_{cat/geo}$ under the experimental conditions.

Comparison of Surrogate Predictions at Experimental Conditions

In order to thoroughly compare the performance of the LSTM-based surrogate model to the DETCHEM^{PBR} simulation model at relevant experimental conditions, predictions were made at several temperatures between 500 and 800 K for various water concentrations in the feed. Figure 8 compares the predicted end-of-pipe conversion of methane by the surrogate model, the numerically simulated conversion using DETCHEM^{PBR}, and the measured experimental data under varying conditions. For the simulations and the surrogate predictions, a reference value of $F_{cat/geo}$ = 74.48 was used, which was calculated based on the nanoparticle diameter of the as-prepared catalyst samples.

The comparison shows that the surrogate model closely follows the simulated conversions using the DETCHEM^{PBR}

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906539/4/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

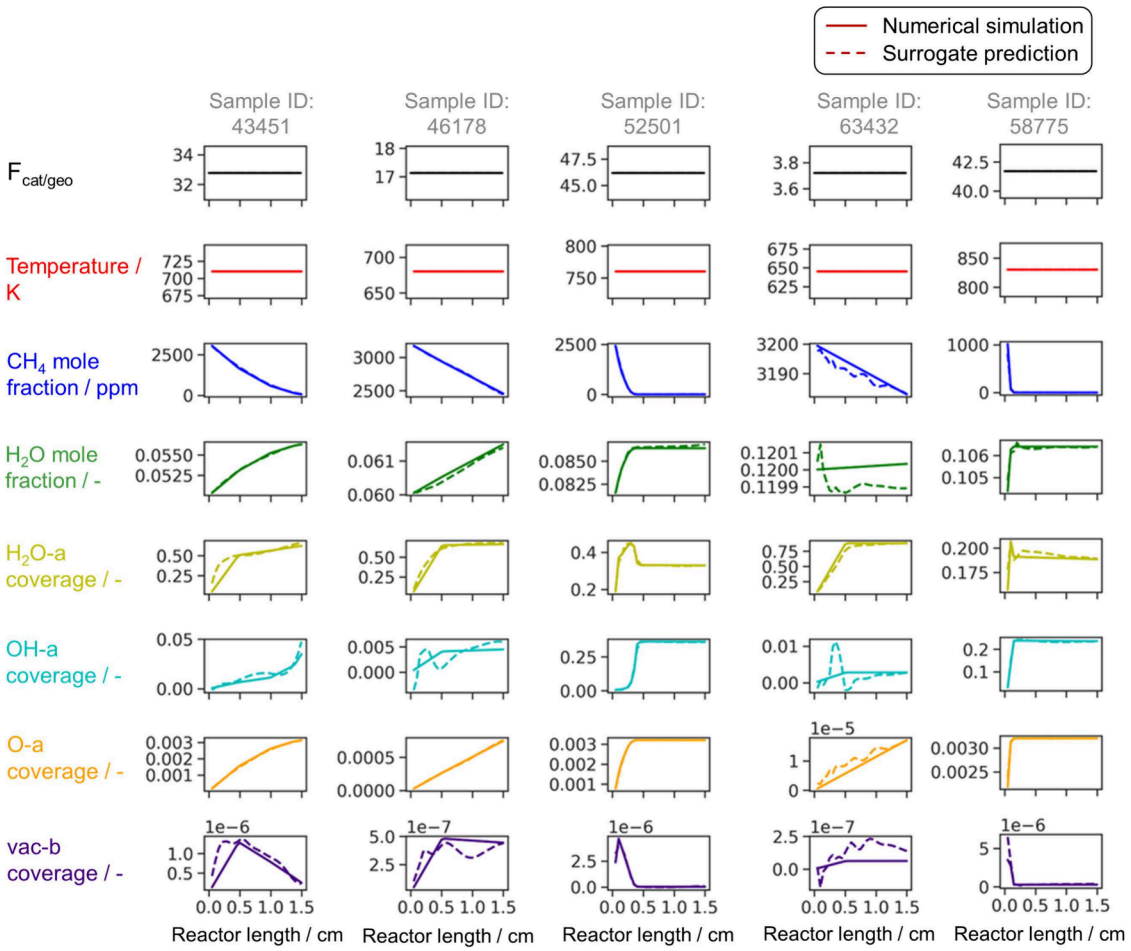


Figure 7. Comparison of the ground truth with the predictions from the LSTM model on unseen data.

code under all of the conditions. The deviations between the simulated (both the output of DETCHEM^{PBR} and the surrogate model) and the experimental data clearly demonstrate that a constant value of catalytically active area (i.e., constant $F_{cat/geo}$ value) leads to overprediction of the conversions at all operating conditions. The simulations correctly capture the onset of light-off, as the kinetic model accounts for the blocking of active sites by adsorbed water molecules resulting from the increasing water concentration in the feed, thereby predicting the shift in activity toward higher temperatures. However, the deactivation of the active sites happening due to continuous exposure of water and changing reactor temperature is not considered in the kinetic model from Stotz et al.[54] and requires consideration of a variable value of active area. This shows that in order to reproduce the experimental data accurately, the effect of the reaction conditions on the active area needs to be correctly estimated.

Surrogate Model Interpretability

Before using the surrogate model to optimize the $F_{cat/geo}$ parameter, it was essential to verify that the model had indeed learned the underlying kinetics and inherent relationships between the reactor operating conditions and the resulting reactor outputs. As described in the Methods section, importance scores were computed for each input feature based on the gradients of the model outputs with respect to the inputs. These high-dimensional feature importance values (shape = (158, 30, 33)) were projected into a two-dimensional embedding space to enable clustering and visualization of the feature importance profiles. Figure 9 shows the resulting clusters obtained from attribution profile embeddings averaged for each water concentration and reactor position. The clustering results were found to be stable across multiple dimensionality reduction runs with different random seeds. The attributions were averaged based on their inlet water concentration to facilitate visualization and to examine whether the resulting attribution profiles exhibited patterns across the water concentration. Based on the feature importance scores obtained for predicting the methane output mole fraction, the low water concentrations (0−3 vol % H_2O) were grouped into one cluster, whereas the samples with higher water concentrations (6−12 vol %) formed a second cluster. This suggests that the surrogate model relies on different attribution patterns for samples with low and high water concentration. While the clustering analysis is purely associative and was performed primarily to visualize the high-dimensional attribution values, the observed separation is consistent with the understanding that the deactivation effect induced by water tends to saturate at higher concentrations.[48]

Figure 10 shows the averaged absolute feature importance values for the two clusters for predicting the outlet methane mole fractions. The importance values clearly demonstrate that the surrogate model identified the temperature as the most important feature to determine the outlet methane mole fraction for both clusters. Among other input features, the $F_{cat/geo}$ parameter relating to the catalytically active surface area

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

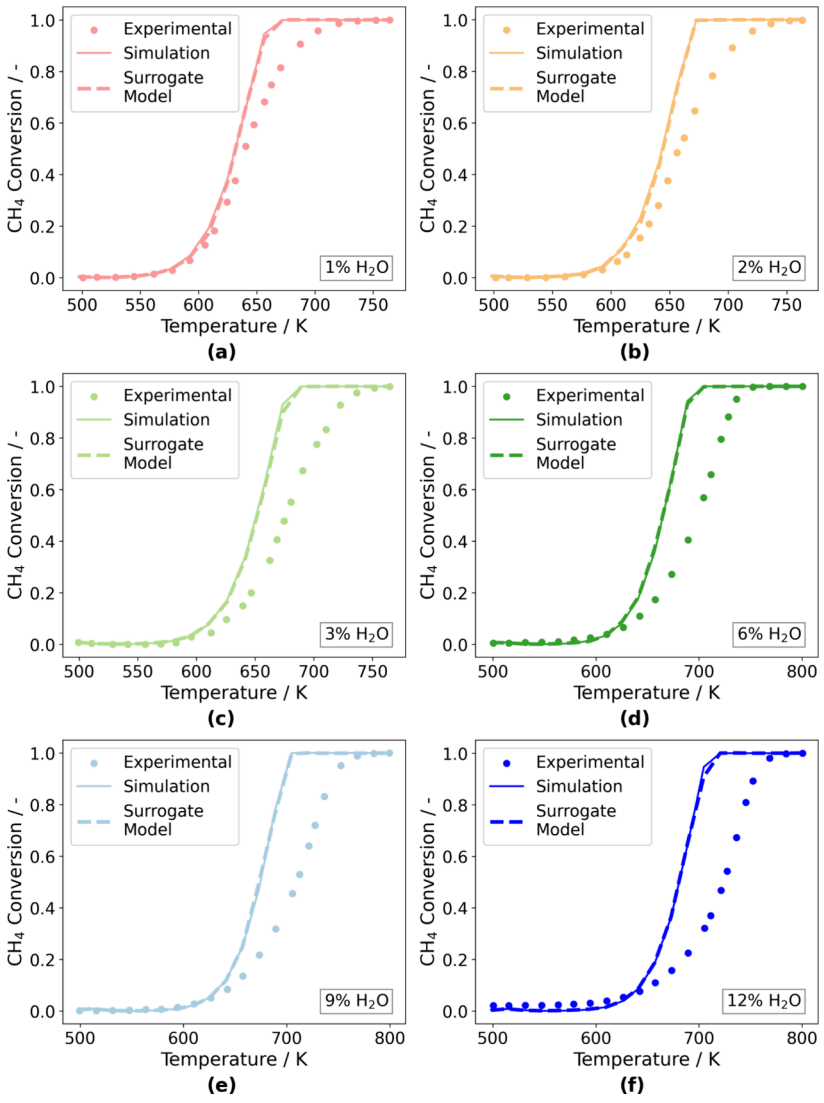


Figure 8. End-of-pipe performance of the surrogate under experimental conditions, with varying inlet water concentrations: (a) 1 vol % H_2O, (b) 2 vol % H_2O, (c) 3 vol % H_2O, (d) 6 vol % H_2O, (e) 9 vol % H_2O, and (f) 12 vol % H_2O.

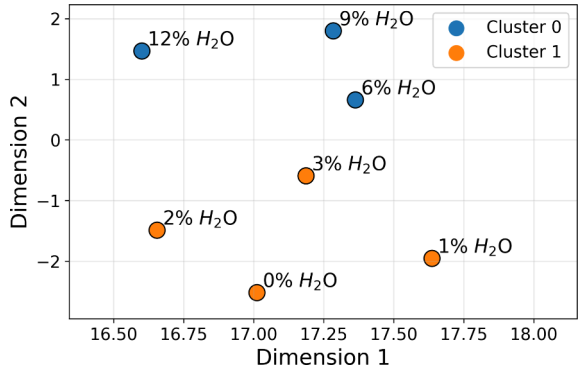


Figure 9. UMAP clustering with the two-dimensional embeddings of the feature importance scores.

is deemed almost equally important for both clusters. The inlet concentration of water (and the correlated inert N_2 concentration) had higher importance for determining the reactor outputs for the cluster with low water concentrations than the high-water-concentration cluster. Interestingly, the model also identified the surface species, namely, the free a and b sites, H_2O-a, H-a, and H-b species as the main contributors for its predictions, which is in agreement with the dominant surface intermediates in the kinetic model[54] used for generating the data. Similar patterns are observed for all other outputs as well (Figures S2 and S3 in the Supporting Information).

These patterns demonstrate that the surrogate model has reliably captured the underlying relationships among the various inputs and the reactor outputs. It is important to note that the data were scaled before training and computing feature importance, preventing bias from differences in the feature magnitudes. Consequently, the reliability of the surrogate model is established not only in its ability to accurately predict the simulation outputs but also in its capacity to correctly identify inherent relationships within the data. Therefore, using the gradients from this surrogate model for the subsequent optimization of the $F_{cat/geo}$ parameter is justified, serving as an effective replacement for the more computationally intensive DETCHEM^{PBR} simulation model that does not provide gradient information.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906539/4/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

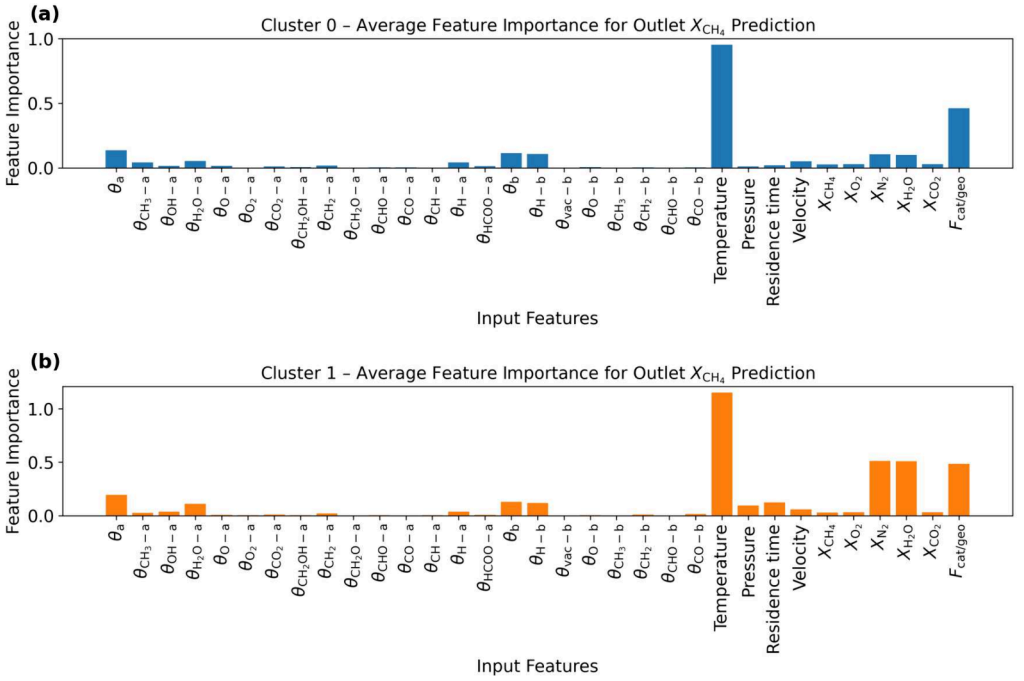


Figure 10. Averaged absolute feature importance values obtained from the gradients of the surrogate model for (a) cluster 0 with high water concentration samples and (b) cluster 1 with low water concentration samples.

Simulated Results with Optimized Active Area

As shown in Figure 8, the simulated data assuming a constant catalytic area (i.e., constant $F_{cat/geo}$ parameter) deviate significantly from the experimental results. This discrepancy arises because continuous exposure to humid and oxidizing conditions leads to catalyst deactivation, making the active surface area variable. To estimate the effective value of the active area with changing operating conditions, a gradient-based optimization was performed using the surrogate model. The $F_{cat/geo}$ values were optimized by minimizing the difference between the experimental end-of-pipe methane mole fractions and the predictions from the surrogate model. The optimization was performed for at least 20 temperature points for each inlet water concentration in the feed.

The resulting trends in the estimated $F_{cat/geo}$ values during light-off for various wet feed conditions, based on the accurate simulation of the experimental data, are illustrated in Figure 11. In the figure, the light blue background highlights the temperature range where a variation in $F_{cat/geo}$ has an influence on the reactor output. The areas shaded in gray correspond to temperature regions where either the rates are too low to have any conversion or so high that the inhibition effects become negligible, therefore, the $F_{cat/geo}$ values under these conditions are irrelevant. As shown in Figure 11, the use of optimized $F_{cat/geo}$ values leads to a much closer agreement between the experimental light-off data and the simulations using DETCHEM^{PBR}, compared to simulations assuming a constant $F_{cat/geo}$ throughout the light-off.

The estimated $F_{cat/geo}$ values generally decrease with an increase in temperature, and the rate of this decrease depends on the inlet water concentration. Notably, at higher temperatures, the $F_{cat/geo}$ values converge to approximately 10 across all water concentrations, indicating that water inhibition effects are minimal in this regime, consistent with the findings of Mortensen et al.[46] It should be emphasized that since the $F_{cat/geo}$ parameter can reflect a range of phenomena affecting the available catalytic area, the estimated values obtained here cannot be used to identify the underlying cause of the reduction in the available catalytic area. Moreover, the present approach is intended to estimate the value of $F_{cat/geo}$ specific to the investigated experiments and should not be interpreted as a predictive model for catalyst activity under new operating conditions.

Additionally, the discrepancies observed in simulations using a constant $F_{cat/geo}$ at high temperatures may also stem from internal mass transfer limitations not accounted for in the model, leading to overestimated reaction rates. Parameter fitting that directly affects reaction rates may inadvertently mask the influence of diffusion limitations. If present, these effects should be explicitly incorporated into the data set and, consequently, into the surrogate modeling framework as well. The reduction in $F_{cat/geo}$ in the optimized case effectively lowers the predicted rates, aligning them more closely with experimental observations. The Weisz-Prater number[67] was calculated for temperatures of 10% conversion (T_{10}), 50% conversion (T_{50}) and 90% conversion (T_{90}) to estimate if internal mass transfer limitations existed. The calculated values were lower than 0.2 for both dry and humid conditions at all conditions except at T_{90}. At T_{90}, the values were estimated to be 0.29 and 0.32 for 0 vol % and 12 vol % H_2O in feed, respectively. These values imply that internal mass transfer is not limiting under the studied conditions but may have some effect at higher temperatures ($\geq T_{90}$). Therefore, given that water inhibition is negligible at elevated temperatures,[68] the analysis of $F_{cat/geo}$ trends should primarily focus on low temperatures, where its impact is more pronounced.

A comparison of the $F_{cat/geo}$ trends across different water concentrations in Figure 11 reveals that the rate at which $F_{cat/geo}$ changes with reactor temperature is strongly influenced by the water content in the feed. With 1 vol % water in the feed, $F_{cat/geo}$ initially decreases gradually up to the temperature corresponding to approximately 60% methane conversion.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

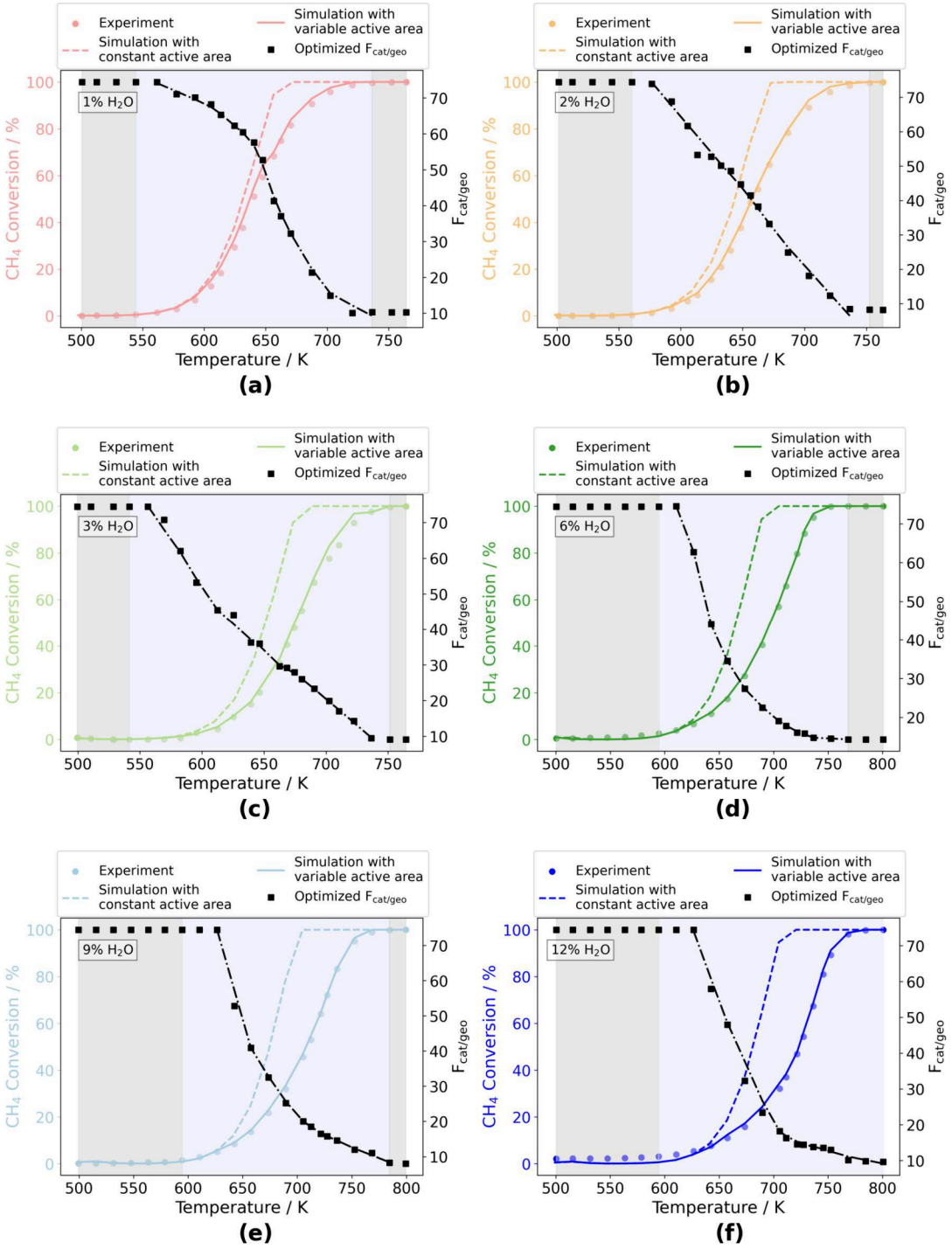


Figure 11. End-of-pipe simulated data with DETCHEM^{PBR} using the optimized $F_{cat/geo}$ values through the surrogate model, under experimental conditions with varying inlet water concentrations: (a) 1 vol % H_2O, (b) 2 vol % H_2O, (c) 3 vol % H_2O, (d) 6 vol % H_2O, (e) 9 vol % H_2O, and (f) 12 vol % H_2O.

Beyond this point, the decline becomes steeper until full conversion is achieved. For feeds containing 2 vol % and 3 vol % H_2O, the decrease in $F_{cat/geo}$ follows an almost linear trend throughout the temperature range. However, at higher water concentrations, the behavior becomes distinctly nonlinear. In these cases, $F_{cat/geo}$ exhibits a sharp initial drop, followed by a more gradual decline as conversion increases. This effect is particularly pronounced at 12 vol % water, where $F_{cat/geo}$ drops dramatically from an initial value of 74.48 to around 15 at temperatures corresponding to only 25% conversion. This suggests that a higher water content in the feed intensifies the inhibition effect at lower temperatures, leading to a more significant early loss of active surface area available for the reaction.

The loss in the effective catalytically active area may be ascertained from the decrease in the $F_{cat/geo}$ parameter values compared to the active area before the catalyst was exposed to reactive conditions (Figure 12). The results are grouped similar to Figure 10 as low and high water concentration in feed. For 1 vol % water in feed, the optimized data indicate

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906539/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

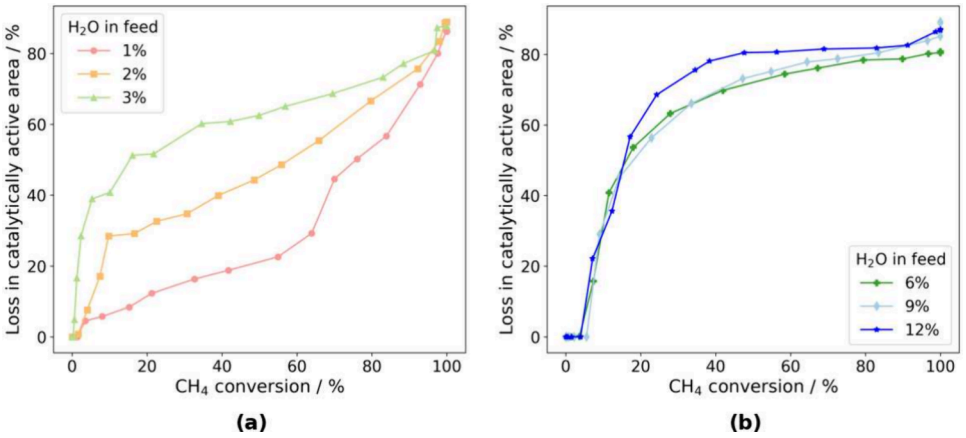


Figure 12. Loss in catalytically active area with respect to conversion for various feed conditions: (a) low water concentration at the inlet and (b) high water concentration at the inlet.

that 50% of the active area was lost after more than 70% conversion is achieved (cf. Figure 12(a)). This much loss in active area is encountered earlier for water concentrations above 3 vol % at temperatures of approximately 20% conversion, due to the intensified inhibition effect. At lower concentrations of water, the loss in active area progresses gradually as more conversion takes place with an increase in temperature. At higher water concentrations, there is a loss of active area overlap (Figure 12(b)). This indicates a saturation point in the inhibition with respect to water content, similar to the findings of Keller et al.[48] In Figure 12(b), at conversions less than 5%, the loss in active area is negligible. This may be attributed to the fact that the loss is calculated relative to the as-prepared catalyst, whereas the catalyst properties likely changed after the degreening step. As a result, the initial $F_{cat/geo}$ values at the start of the light-off may have been underestimated, underscoring the need for accurate postdegreening characterization of catalyst samples.

Overall, it has been demonstrated that the experimental observations during the light-offs are well-reproduced using the estimated $F_{cat/geo}$ values obtained from the optimization procedure. The reactor simulations with varying $F_{cat/geo}$ values show that the experimental data could be reproduced more accurately when a loss of active area is considered, consistent with the continuous exposure of the catalyst samples to humid conditions and increasing temperatures during the light-off measurements. Figure 12 shows that this loss in available catalytic area is a function of both water concentration and how much methane is converted at any point in the light-off. The dependency of the loss in catalytically active area on conversion values is consistent with the findings from Mortensen et al.,[46] who suggested that the inactive state of the catalyst is a result of the migration and redispersion of the highly mobile Pd^0 species formed when methane gets activated. The use of the surrogate model for gradient-based optimization enabled capturing the complex interactions of water concentration and the reaction temperature on the surface coverages and the observed end-of-pipe methane concentrations without relying on making prior assumptions about the conditions driving the deactivation. The surrogate model allowed computation of the gradients of methane outlet mole fraction with respect to the input $F_{cat/geo}$ values that are not easily available with direct simulation of the reactor model, in addition to fast computation of the reactor outputs with changing input conditions during the optimization process.

CONCLUSIONS

In this work, several deep learning architectures were explored to function as surrogate models for the $DETCHEM^{PBR}$ simulation model. Three architectures were trained on simulated data to capture the methane oxidation kinetics in a packed-bed reactor. Among them, the long short-term memory (LSTM) network delivered the best performance on test data and was selected as the surrogate model to replace the $DETCHEM^{PBR}$ simulation model. The analysis of the LSTM model reasoning patterns for its predictions showed a reliable capture of the underlying kinetics and relationships within the data. Therefore, the LSTM model could be successfully applied in the subsequent gradient-based optimization task. Depending on the task at hand, other architectures, such as PINNs that explicitly incorporate thermodynamic consistency or atom conservation, may be necessary. However, black-box predictive models such as LSTM with some degree of interpretability can be equally useful for simpler applications, as showcased in this work.

The surrogate model predictions, in agreement with the $DETCHEM^{PBR}$ model, showed deviations in simulated data in comparison with the experimental data using the microkinetic model by Stotz et al.[54] for PdO/Al_2O_3. These deviations were traced to the assumption of a constant catalytically active area during light off. While the model correctly captured the shift in activity to higher temperatures in the presence of external water, experimental data showed additional inhibition not explained by site blocking alone, consistent with the irreversible deactivation demonstrated in recent studies.[46,47] Instead, the findings suggest that an irreversible loss of active sites due to prolonged exposure to reaction conditions should be considered to reproduce experimental observations. Through the optimization of the $F_{cat/geo}$ parameter, the variation in the catalytically active area with the water concentration could be investigated. At low water concentrations (<3 vol %), a gradual loss of active sites was observed as temperature and conversion increased. Higher water concentrations led to a more pronounced loss, up to 50% of the active area at 20% conversion.

The results also showed that the inhibition effect saturates with increasing water concentration and becomes negligible at

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX

higher temperatures where full conversion is achieved irrespective of the available catalytic area, aligning with the literature.[68] Overall, the findings highlight the significant role of external water in irreversible catalyst deactivation, in addition to reversible blocking of active sites. While the simulations with optimized $F_{cat/geo}$ values cannot pinpoint whether this deactivation is due to $Pd(OH)_2$ formation, or a change in the Pd^{2+} state, or redispersion of the Pd particles, they underscore the need to account for the change in catalyst state for accurate simulation of catalyst performance, especially under wet conditions.

In conclusion, this work highlights the effectiveness and practicality of surrogate modeling for accelerating optimization in heterogeneous catalysis. The LSTM-based surrogate model enabled accurate, efficient predictions and facilitated gradient-based optimization, offering a practical alternative to costly numerical simulations. This demonstrated the value of using data-driven tools to bridge the gap between simulation and experimental studies. Beyond optimization, such surrogate models are especially advantageous in scenarios requiring numerous costly simulations, such as scale-up studies or design of systems consisting of multiple unit operations.

■ ASSOCIATED CONTENT

Data Availability Statement

The data set used for model training, the experimental data and the trained LSTM model parameters are made available at https://hdl.handle.net/21.11165/4cat/444x-kpks. The code used for model training is also publicly available at https://gitlab.com/sofia.angeli/dynamic_ch4_ox.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsappcat.6c00044.

Description of the modeling equations, conversion profiles from a sample of the data set, the surrogate model architecture hyperparameters, and the input feature importance values for all model outputs for the chosen surrogate model (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Sofia Angeli − *Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen 76344, Germany;* orcid.org/0000-0001-8623-047X; Email: sofia.angeli@kit.edu

Authors

Rinu Chacko − *Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology, Karlsruhe 76131, Germany*

Markus Götz − *Scientific Computing Center and Helmholtz AI, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen 76344, Germany*

Olaf Deutschmann − *Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology, Karlsruhe 76131, Germany; Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen 76344, Germany;* orcid.org/0000-0001-9211-7529

Complete contact information is available at: https://pubs.acs.org/10.1021/acsappcat.6c00044

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by the Helmholtz Association Initiative and Networking Fund under the Helmholtz AI platform and the HAICORE@KIT partition grants. The authors also acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under CRC-1441 (Project-ID 426888090) and NFDI4Cat (Project-ID 670389-NFDI 2/1). The authors gratefully acknowledge Kevin Keller for providing the experimental and characterization data for methane oxidation over a PdO catalyst. In addition, they thank omegadot software & consulting GmbH (Limburgerhof, Germany) for providing a cost-free academic license of DETCHEM for simulations.

■ REFERENCES

(1) Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311−2323.

(2) Schlexer Lamoureux, P.; Winther, K. T.; Garrido Torres, J. A.; Streibel, V.; Zhao, M.; Bajdich, M.; Abild-Pedersen, F.; Bligaard, T. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem.* **2019**, *11*, 3581−3601.

(3) Bozal-Ginesta, C.; Pablo-García, S.; Choi, C.; Tarancón, A.; Aspuru-Guzik, A. Developing Machine Learning for Heterogeneous Catalysis with Experimental and Computational Data. *Nature Reviews Chemistry* **2025**, *9*, 601−616.

(4) Mou, T.; Pillai, H. S.; Wang, S.; Wan, M.; Han, X.; Schweitzer, N. M.; Che, F.; Xin, H. Bridging the Complexity Gap in Computational Heterogeneous Catalysis with Machine Learning. *Nature Catalysis* **2023**, *6*, 122−136.

(5) Rangarajan, S.; Tian, H. Improving the Predictive Power of Microkinetic Models via Machine Learning. *Current Opinion in Chemical Engineering* **2022**, *38*, 100858.

(6) Abraham, B. M.; Jyothirmai, M. V.; Sinha, P.; Viñes, F.; Singh, J. K.; Illas, F. Catalysis in the Digital Age: Unlocking the Power of Data with Machine Learning. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2024**, *14*, e1730.

(7) Casillo, E.; Scattolin, T.; Nolan, S. P. Catalysis Meets Machine Learning: A Guide to Data-Driven Discovery and Design. *Chem. Commun.* **2025**, *61*, 18247−18272.

(8) Suvarna, M.; Pérez-Ramírez, J. Embracing Data Science in Catalysis Research. *Nature Catalysis* **2024**, *7*, 624−635.

(9) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533−536.

(10) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: 2017. DOI: 10.48550/arXiv.1706.03762.

(11) Gu, A.; Goel, K.; Ré, C. Efficiently Modeling Long Sequences with Structured State Spaces. *arXiv* (Computer Science.Machine Learning), 2022-08−05. https://arxiv.org/abs/2111.00396 (accessed 11/01/2025).

(12) Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv* (Computer Science.Machine Learning), 2024-05−31. https://arxiv.org/abs/2312.00752 (accessed 11/01/2025).

(13) Dao, T.; Gu, A. Transformers are SSMs: Generalized Models and Efficient Algorithms through Structured State Space Duality. In *Proceedings of the 41st International Conference on Machine Learning*; ICML'24; JMLR.org: 2024. DOI: 10.5555/3692070.3692469.

(14) Khaldi, M. K.; Al-Dhaifallah, M.; Taha, O.; Mahmood, T.; Alharbi, A. Computational Modeling of a Fluid Catalytic Cracking Unit. *Ain Shams Engineering Journal* **2025**, *16*, 103661.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

(15) Wang, S.; Liu, Z.; Li, Z.; Liu, B.; Cui, Z.; Tian, W. Integrating Physical Modeling and Machine Learning for Reaction Activity Dynamic Prediction in Chemical Reactors. *Chemical Engineering Journal* **2025**, *522*, 167895.
(16) He, W.; Li, J.; Tang, Z.; Wu, B.; Luan, H.; Chen, C.; Liang, H. A Novel Hybrid CNN-LSTM Scheme for Nitrogen Oxide Emission Prediction in FCC Unit. *Mathematical Problems in Engineering* **2020**, *2020*, 8071810.
(17) Wang, Z.; Wu, Z. Towards Foundation Model for Chemical Reactor Modeling: Meta-Learning with Physics-Informed Adaptation. *Chem. Eng. Res. Des.* **2025**, *218*, 839−853.
(18) Suzuki, K.; Toyao, T.; Maeno, Z.; Takakusagi, S.; Shimizu, K.-i.; Takigawa, I. Statistical Analysis and Discovery of Heterogeneous Catalysts based on Machine Learning from Diverse Published Data. *ChemCatChem.* **2019**, *11*, 4537−4547.
(19) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable Machine Learning for Knowledge Generation in Heterogeneous Catalysis. *Nature Catalysis* **2022**, *5*, 175−184.
(20) McBride, K.; Sundmacher, K. Overview of Surrogate Modeling in Chemical Process Engineering. *Chemie Ingenieur Technik* **2019**, *91*, 228−239.
(21) Mora-Mariano, D.; Flores-Tlacuahuac, A. A Machine Learning Approach for the Surrogate Modeling of Uncertain Distributed Process Engineering Models. *Chem. Eng. Res. Des.* **2022**, *186*, 433−450.
(22) Wang, C.; Wang, B.; Wang, C.; Li, A.; Chang, Z.; Wang, R. A Machine Learning Model with Minimize Feature Parameters for Multi-Type Hydrogen Evolution Catalyst Prediction. *npj Computational Materials* **2025**, *11*, 111.
(23) Malpica Galassi, R.; Ciottoli, P. P.; Valorani, M.; Im, H. G. An Adaptive Time-Integration Scheme for Stiff Chemistry based on Computational Singular Perturbation and Artificial Neural Networks. *J. Comput. Phys.* **2022**, *451*, 110875.
(24) Döppel, F. A.; Votsmeier, M. Efficient Machine Learning Based Surrogate Models for Surface Kinetics by Approximating the Rates of the Rate-Determining Steps. *Chem. Eng. Sci.* **2022**, *262*, 117964.
(25) Odabaşı, Ç.; Günay, M. E.; Yıldırım, R. Knowledge Extraction for Water Gas Shift Reaction over Noble Metal Catalysts from Publications in the Literature between 2002 and 2012. *International journal of hydrogen energy* **2014**, *39*, 5733−5746.
(26) Maroor, S.; Spatenka, S.; Gupta, U. Application of Surrogate Modeling To Accelerate Design Space Exploration for Catalytic Reactor Systems. *Ind. Eng. Chem. Res.* **2026**, *65*, 3119−3131.
(27) Pétremand, H.; Witte, J.; Krocher, O.; Moioli, E. Machine Learning-Based Kinetic Modeling of the CO_2 Methanation Reaction over an Industrial Catalyst. *Ind. Eng. Chem. Res.* **2025**, *64*, 19864−19875.
(28) Doeppel, F. A.; Votsmeier, M. Efficient Neural Network Models of Chemical Kinetics using a Latent asinh Rate Transformation. *Reaction Chemistry & Engineering* **2023**, *8*, 2620−2631.
(29) Fedorov, A.; Perechodjuk, A.; Linke, D. Kinetics-Constrained Neural Ordinary Differential Equations: Artificial Neural Network Models Tailored for Small Data to Boost Kinetic Model Development. *Chemical Engineering Journal* **2023**, *477*, 146869.
(30) Lacerda de Oliveira Campos, B.; Oliveira Souza da Costa, A.; Herrera Delgado, K.; Pitter, S.; Sauer, J.; Ferreira da Costa Junior, E. Development of a Surrogate Artificial Neural Network for Microkinetic Modeling: Case Study with Methanol Synthesis. *Reaction Chemistry & Engineering* **2024**, *9*, 1047−1060.
(31) Lastrucci, G.; Theisen, M. F.; Schweidtmann, A. M. Physics-Informed Neural Networks and Time-Series Transformer for Modeling of Chemical Reactors. In *34th European Symposium on Computer Aided Process Engineering/15th International Symposium on Process Systems Engineering*, Manenti, F., Reklaitis, G. V., Eds.; Computer Aided Chemical Engineering, Vol. 53; Elsevier: 2024; pp 571−576. DOI: 10.1016/B978-0-443-28824-1.50096-X.
(32) Döppel, F. A.; Bracconi, M.; Maestri, M. Physically Consistent Neural Network Surrogates of Catalytic Reactors with Detailed Surface Kinetics. *ACS Engineering Au* **2026**, *6*, 359−371.
(33) Goodwin, D. G.; Speth, R. L.; Moffat, H. K.; Weber, B. W. Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes. *Zenodo* 2018. DOI: 10.5281/zenodo.4527812.
(34) Gusmão, G. S.; Retnanto, A. P.; Da Cunha, S. C.; Medford, A. J. Kinetics-Informed Neural Networks. *Catal. Today* **2023**, *417*, 113701.
(35) Fernandes, F. A. Optimization of Fischer−Tropsch Synthesis using Neural Networks. *Chem. Eng. Technol.* **2006**, *29*, 449−453.
(36) Kircher, T.; Doeppel, F. A.; Votsmeier, M. Global Reaction Neural Networks with Embedded Stoichiometry and Thermodynamics for Learning Kinetics from Reactor Data. *Chemical Engineering Journal* **2024**, *485*, 149863.
(37) Stagge, H.; Güttel, R. The Findability of Microkinetic Parameters by Heterogeneous Chemical Reaction Neural Networks (hCRNNs). *Chemical Engineering Journal* **2025**, *510*, 161460.
(38) Nentwich, C.; Engell, S. Application of Surrogate Models for the Optimization and Design of Chemical Processes. *2016 International Joint Conference on Neural Networks (IJCNN)* **2016**, 1291−1296.
(39) Fahmi, I.; Cremaschi, S. Process Synthesis of Biodiesel Production Plant using Artificial Neural Networks as the Surrogate Models. *Comput. Chem. Eng.* **2012**, *46*, 105−123.
(40) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735−1780.
(41) Zhang, Y., Yan, J. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In *The Eleventh International Conference on Learning Representations*; ICLR: 2023.
(42) Alyani, M.; Smith, K. J. Kinetic Analysis of the Inhibition of CH_4 Oxidation by H_2O on PdO/Al_2O_3 and $CeO_2/PdO/Al_2O_3$ Catalysts. *Ind. Eng. Chem. Res.* **2016**, *55*, 8309−8318.
(43) Gholami, R.; Alyani, M.; Smith, K. J. Deactivation of Pd Catalysts by Water during Low Temperature Methane Oxidation Relevant to Natural Gas Vehicle Converters. *Catalysts* **2015**, *5*, 561−594.
(44) Ciuparu, D.; Perkins, E.; Pfefferle, L. In situ DR-FTIR Investigation of Surface Hydroxyls on γ-Al2O3 Supported PdO Catalysts during Methane Combustion. *Applied Catalysis A: General* **2004**, *263*, 145−153.
(45) Velin, P.; Florén, C.-R.; Skoglundh, M.; Raj, A.; Thompsett, D.; Smedler, G.; Carlsson, P.-A. Palladium Dispersion Effects on Wet Methane Oxidation Kinetics. *Catalysis Science & Technology* **2020**, *10*, 5460−5469.
(46) Mortensen, R. L.; Noack, H.-D.; Pedersen, K.; Dunstan, M. A.; Wilhelm, F.; Rogalev, A.; Pedersen, K. S.; Mielby, J.; Mossin, S. Understanding the Reversible and Irreversible Deactivation of Methane Oxidation Catalysts. *Applied Catalysis B: Environment and Energy* **2024**, *344*, 123646.
(47) Chen, H.-Y.; Lugo-Jose, Y.; Fedeyko, J. M.; Toops, T. J.; Fidler, J. Irreversible Catalyst Deactivation Mechanisms of $PdO/\gamma\text{-}Al_2O_3$ Catalysts for Lean Methane Oxidation. *ACS Catal.* **2024**, *14*, 15751−15763.
(48) Keller, K.; Lott, P.; Stotz, H.; Maier, L.; Deutschmann, O. Microkinetic Modeling of the Oxidation of Methane over PdO Catalysts—Towards a Better Understanding of the Water Inhibition Effect. *Catalysts* **2020**, *10*, 922.
(49) Deutschmann, O.; Tischer, S.; Correa, C.; Chatterjee, D.; Kleditzsch, S.; Janardhanan, V. M.; Mladenov, N.; Minh, H. D.; Karadeniz, H.; Hettel, M. et al. *DETCHEM Software Package*; Karlsruhe, Germany, 2020. www.detchem.com.
(50) Mokashi, M.; Shirsath, A. B.; Çelik, A.; Lott, P.; Müller, H.; Tischer, S.; Maier, L.; Bode, J.; Schlereth, D.; Scheiff, F.; Flick, D.; Bender, M.; Ehrhardt, K.; Deutschmann, O. Methane Pyrolysis in Packed Ped Reactors: Kinetic Modeling, Numerical Simulations, and Experimental Insights. *Chemical Engineering Journal* **2024**, *485*, 149684.
(51) Hettel, M.; Daymo, E. A.; Sawant, N.; Deutschmann, O. 2D Lattice-Boltzmann Method with Detailed Surface Chemistry:

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906394/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

Conservation of Geometrical Surface. *Chem. Eng. Sci.* **2026**, *321*, 122797.
(52) Bergeret, G.; Gallezot, P. Particle Size and Dispersion Measurements In *Handbook of Heterogeneous Catalysis*, Ertl, G., Knozinger, H., Weitkamp, J., Eds.; Wiley-VCH: Weinheim, 1997; pp 439−491. DOI: 10.1002/9783527610044.hetcat0038.
(53) Carman, P. Fluid Flow through Granular Beds. *Chem. Eng. Res. Des.* **1997**, *75*, S32−S48.
(54) Stotz, H.; Maier, L.; Boubnov, A.; Gremminger, A.; Grunwaldt, J.-D.; Deutschmann, O. Surface Reaction Kinetics of Methane Oxidation over PdO. *J. Catal.* **2019**, *370*, 152−175.
(55) *Python Programming Language*. https://www.python.org/.
(56) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* **2016**, *3*, 160018.
(57) T*he HDF5 Library & File Format: High-Performance Data Management and Storage Suite*. https://www.hdfgroup.org/solutions/hdf5/, accessed 10/12/2024.
(58) Paszke, A. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2019. DOI: 10.5555/3454287.3455008.
(59) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* (Computer Science.Machine Learning), 2014-12−22. https://arxiv.org/abs/1412.6980 (accessed 11/01/2025).
(60) *Matplotlib: Visualization with Python*. https://matplotlib.org/.
(61) *NumPy: The Fundamental Package for Scientific Computing with Python*. https://numpy.org/.
(62) *pandas: Python Data Analysis Library*. https://pandas.pydata.org/.
(63) Kokhlikyan, N.; Miglani, V.; Martin, M.; Wang, E.; Alsallakh, B.; Reynolds, J.; Melnikov, A.; Kliushkina, N.; Araya, C.; Yan, S. et al. Captum: A Unified and Generic model interpretability library for pytorch. *arXiv* (Computer Science.Machine Learning), 2020-09−16. https://arxiv.org/abs/2009.07896 (accessed 12/01/2025).
(64) Sundararajan, M.; Taly, A.; Yan, Q. *Axiomatic Attribution for Deep Networks In Proceedings of the 34th International Conference on Machine Learning - Vol. 70*; JMLR.org: Sydney NSW, Australia, 2017; pp 3319−3328. DOI: 10.48550/arXiv.1703.01365.
(65) McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* (Statistics.Machine Learning), 2018−02−09. https://arxiv.org/abs/1802.03426 (accessed 12/01/2025).
(66) Campello, R. J. G. B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates In *Advances in Knowledge Discovery and Data Mining*; Pei, J., Tseng, V. S., Cao, L., Motoda, H., Xu, G., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 160−172. DOI: 10.1007/978-3-642-37456-2_14.
(67) Weisz, P.; Prater, C. Interpretation of Measurements in Experimental Catalysis. *Adv. Catal.* **1954**, *6*, 143−196.
(68) Burch, R.; Urbano, F.; Loader, P. Methane Combustion over Palladium Catalysts: The Effect of Carbon Dioxide and Water on Activity. *Applied Catalysis A: General* **1995**, *123*, 173−184.

Downloaded from pubs.acs.org/aacabd/article-pdf/doi/10.1021/acsappcat.6c00044/6906539/4/acsappcat.6c00044.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acsappcat.6c00044
ACS Appl. Catal. XXXX, XXX, XXX−XXX