ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

# Multiobjective Bayesian Optimization Framework for the Synthesis of Methanol from Syngas Using Interpretable Gaussian Process Models

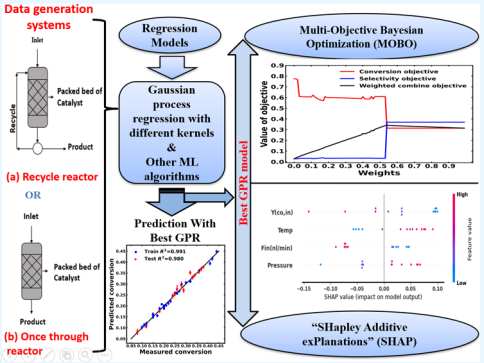
Avan Kumar, Kamal K. Pant, Sreedevi Upadhyayula,* and Hariprasad Kodamana*

Cite This: *ACS Omega* 2023, 8, 410−421

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Methanol production has gained considerable interest on the laboratory and industrial scale as it is a renewable fuel and an excellent hydrogen energy storehouse. The formation of synthesis gas ($CO/H_2$) and the conversion of synthesis gas to methanol are the two basic catalytic processes used in methanol production. Machine learning (ML) approaches have recently emerged as powerful tools in reaction informatics. Inspired by these, we employ Gaussian process regression (GPR) to the model conversion of carbon monoxide (CO) and selectivity of the methanol product using data sets obtained from experimental investigations to capture uncertainty in prediction values. The results indicate that the proposed GPR model can accurately predict CO conversion and methanol selectivity as compared to other ML models. Further, the factors that influence the predictions are identified from the best GPR model employing "Shapley Additive exPlanations" (SHAP). After interpretation, the essential input features are found to be the inlet mole fraction of CO (Y(CO, in)) and the net inlet flow rate (Fin(nL/min)) for our best prediction GPR models, irrespective of our data sets. These interpretable models are employed for Bayesian optimization in a weighted multiobjective framework to obtain the optimal operating points, namely, maximization of both selectivity and conversion.




## INTRODUCTION

Methanol is a main chemical that may be used as a fuel, a fuel substitute, or a chemical feedstock. It can be used in combustion turbines for rapid-starting and as a gasoline substitute or a blend with gasoline in engines.[1] It is a good hydrogen carrier as well.[2] Methanol is frequently utilized as a raw material feedstock for various industrial chemicals, including methyl methacrylate, acetic acid, formaldehyde, and methyl tertiary-butyl ether.[3] There are two main catalytic routes for methanol production from syngas ($CO/H_2$) by hydrogenation. In the first route, syngas is produced from fossil sources by reforming technologies, i.e., dry and steam.[4,5] However, in the second route, pyrolysis or gasification of coal and biomass-based feedstocks is employed for syngas production. Hydrogenation is the proceeding step for methanol synthesis.[6] In addition, syngas production from shale gas is also prevalent in North American countries due to the abundance of shale reserves.[7]

Catalytic reduction of carbon monoxide (CO) with hydrogen ($H_2$) from coal-derived syngas into methanol can be achieved using both gas-phase and liquid-phase processes (LPMeOHTM) as developed by air products and chemicals.[8] The following are the relevant reactions

$$2H_2 + CO \rightleftharpoons CH_3OH \quad (1)$$

$$CO_2 + 3H_2 \rightleftharpoons CH_3OH + H_2O \quad (2)$$

$$CO + H_2O \rightleftharpoons CO_2 + H_2 \quad (3)$$

All of the above presented three reactions eqs 1−3 are highly exothermic.[9−11] The conventional commercial gas-phase process is carried out in a fixed-bed reactor at high pressure. Depending on the catalyst, the methanol synthesis reaction is usually conducted at about 40−110 barg and 200−270 °C. Even though eqs 1−3 are highly exothermic, the temperature rise across an adiabatic reactor is controlled by concrete recycling of the $H_2$ affluent process gas. After dilution with recycled $H_2$, the CO content at the reactor input is usually controlled to around 10−15%. Methanol synthesis catalysts are often a ternary combination of copper, zinc oxide, and alumina with a promoter such as magnesia.[12,13] Recent advances have also resulted in the possibility of a novel catalyst made of carbon, nitrogen, and platinum.[14] Methanol has also been produced from syngas using membrane, electrochemical, photoelectrochemical, and slurry reactors.[15,16]

**Received:** August 3, 2022
**Accepted:** November 8, 2022
**Published:** December 16, 2022

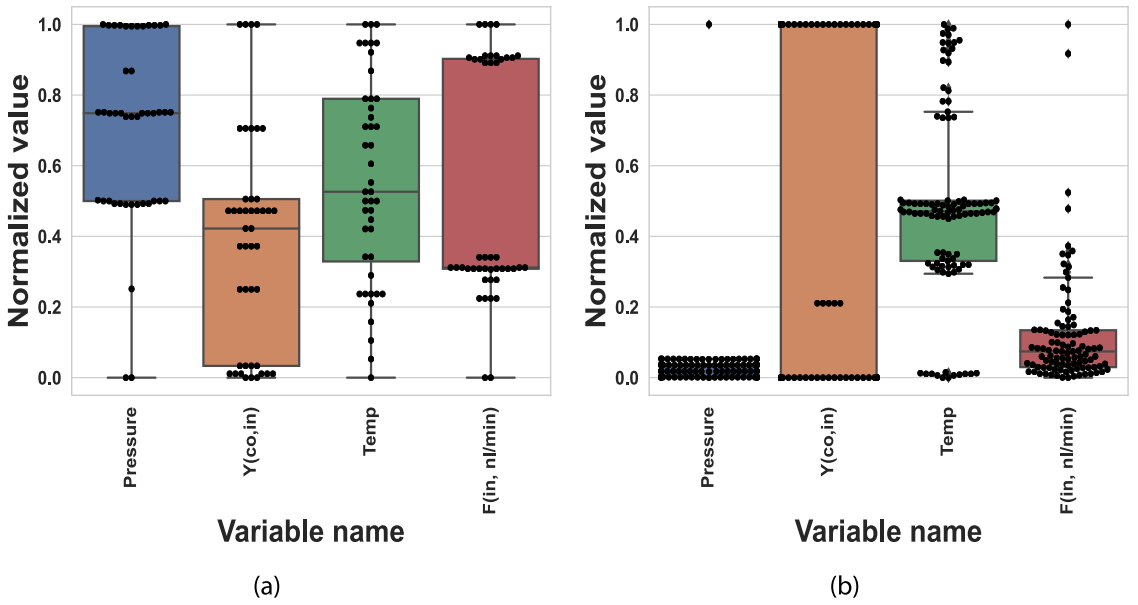



**Figure 1.** Distribution of all descriptors represented in the form of box plots, where (a) and (b) represent recycle and once-through reactor data sets, respectively.

To optimize the conditions for methanol production from syngas, it is imperative to model the CO conversion and methanol selectivity in the reaction (eqs 1−3) as functions of feed and process conditions. To this end, kinetic modeling and reactor design of methanol formation have been studied by many researchers in the last decade.[17−19] However, being developed from first principles, they are bound by various assumptions and are tedious to solve in real time. Alternatively, one can use data-driven machine learning (ML) approaches to model the underlying trend of reaction kinetics if sufficient and representative data are available from the experiments.[20,21] A critical impediment in this direction is the availability of flexible machine learning models that can capture the tricky hidden underlying trend in the experimental data sets with sufficient accuracy. Further, machine learning and applications in catalysis are witnessing a progressive trend in recent times.[22−25]

Drawing inspiration from the above, the following are proposed in this study: (1) a Gaussian process regression (GPR) modeling framework for predicting the conversion of CO and methanol selectivity as target/output variables using various descriptors from laboratory-scale experimental data; (2) SHAP-based interpretation of the developed GPR models; and (3) Multi-Objective Bayesian Optimization (MOBO) to identify the optimal operational trajectory using the developed interpretable models. To this end, we have used laboratory-scale experimental data sets of gas-phase methanol synthesis from a fixed-bed once-through reactor and a Berty recycle continuous reactor,[26] where the utilized catalyst contains CuO, ZnO, MgO, Ni-Cu/$Al_2O_3$, and K−ZnCr.[27−29] GPR has been frequently employed in the nonparametric Bayesian framework for the data-driven modeling of complex systems.[30−32] The advantage of the GPR model is that it can quantify the uncertainty in modeling. The GPR's basic premise is that a multivariate Gaussian distribution may be used to describe a collection of any arbitrary function value.[33] Also, GPR has an impeccable ability to capture the hidden trends in sparse data sets.[31,34−36] In this study, we extensively leverage the prediction capabilities of GPR by tuning hyperparameters and kernels. The results of GPR are compared to those of deep neural nets, support vector machines, and regression models, among other machine learning methods.[25]

Features are vital in terms of their relevance and contribution in a data-driven modeling framework as they critically affect the prediction of the target variables.[37] Shapley Additive exPlanation (SHAP) is a framework that helps interpret an ML model effectively. It provides precise information on the relevance of each feature instance and the order of importance of features for target variable prediction. These are calibrated using the "Shapley" value. Many researchers have implemented this technique to interpret the black-box nature of the ML-based models. Ekanayake et al. have implemented a decision tree, adaptive boost (AdaBoost), and extreme gradient boost (XGBoost) in the SHAP framework.[38] Zaki et al. utilized SHAP to study the optical properties of glass by implying the ML model.[39] Onsree et al. investigated the effect of features on the accuracy of the ML model using SHAP.[40] Park et al. presented a comparative study of eight machine learning algorithms by implementing the SHAP to identify the best combination of features.[41] Similarly, Liang et al. worked with ensemble machine learning models, namely, random forest (RF), extreme gradient boosting machine (XGBoost), and light gradient boosting machine (LGBM), and interpreted by means of SHAP.[42] However, to the best of the author's knowledge, the "SHAP"-based framework has not been used in interpreting the GPR models.

Subsequently, a multiobjective Bayesian optimization is employed to optimize process variables. Bayesian optimization is an effective tool for finding the optimal global solution by building a data-driven surrogate model like GPR.[43,44] The surrogate model forms the probabilistic model of the objective function, which is then optimized with the help of acquisition functions. Some of the popularly used acquisition functions are expected improvement (EI), maximum probability improvement (MPI), lower confidence bound (LCB), etc. Bayesian optimization is becoming popular in process optimization as well, for instance, multiobjective optimization based on the optimal design of the reactor using CFD data,[45,46] optimization-based design and optimization of toluene di-isocyanate (TDI)


https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

reactors,[47] robust optimization with uncertainty quantification,[48] development of data-driven decision-making systems for chemical synthesis,[49,50] optimization of process parameters for the conversion of a mixture of waste terpenes to *p*-cymene,[51] process design,[52] and an optimized process model built by implying aspen plus.[53] The multiobjective optimization for maximizing methanol production and reducing carbon emission uses the genetic algorithm.[54] Using machine learning and Bayesian optimization, they created metal oxides while jointly optimizing experimental parameters to meet the target $CO_2$ and $H_2$ conversion predictions.[55] However, to the best of the authors' knowledge, hardly any work has been reported in the literature that employs interpretable GPR models in a multiobjective Bayesian optimization framework, specifically for syngas-to-methanol conversion.

The following are the details of how the rest of the article is organized. The data sets and preprocessing of data sets are briefly discussed in Section 2. The GPR modeling method and the multiobjective Bayesian optimization strategy are discussed in Section 3. The study's results and discussions are presented in Section 4, and concluding remarks are presented in Section 5.

## DATA SETS AND PREPROCESSING

We consider two distinct data sets obtained from experiments/literature for synthesizing methanol from syngas.[26,27] The first route employs a recycle reactor, and the second one utilizes a once-through reactor. After normalization of salient features, box plots are generated for the recycled reactor data and the once-through reactor in Figure 1(a),(b), respectively.

In this study, we have considered that the selectivity of methanol and the conversion of CO are target variables for prediction and optimization. The set of input descriptors includes the following variables: (a) reactor pressure (pressure) (40−110 barg), (b) reaction temperature (Temp) (200−270 °C), (c) inlet stream flow rate (Fin(nL/min)), and (d) inlet mole fractions of CO (Y(CO, in)), as mentioned in Figure 1. These feature sets are utilized for the prediction of conversion of CO and selectivity of methanol.

## METHODS: GPR MODELING, SHAP, AND MULTIOBJECTIVE BAYESIAN OPTIMIZATION

This section describes the GPR modeling, the SHAP framework, and the multiobjective Bayesian optimization, which are the three pillars of the proposed study.

**Gaussian Process Regression.** The Gaussian process models a set of random variables such that they follow a joint Gaussian distribution. For a set of inputs $x \in \mathbb{R}^n$, let us define the following distribution of functions

$$f(x) = \mathcal{GP}(\mu(x), k(x, x')) \tag{4}$$

where

$$\mu(x) = E[f(x)] \tag{5}$$

$$k(x, x') = E[(f(x)-\mu(x))^T (f(x') - \mu(x'))] \tag{6}$$

Here, $\mu(x)$ is the mean function and $k(x,x')$ is the covariance function. For a data set with $N$ sample points, $k(x,x')$ is defined as follows

$$k(x, x') = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ k(x_2, x_1) & \cdots & k(x_2, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix} \tag{7}$$

The kernel matrix $k(x,x')$ has a dimension of $N \times N$, symmetric, and positive definite. A suitable kernel is chosen based on assumptions such as data smoothness and patterns. According to a reasonable assumption, the correlation between two points should diminish as the distance between them increases. It indicates that nearby data points are likely to behave more similarly than farther away data points.

In the GPR modeling paradigm, the output $y$ and input $x$ are modeled as

$$y = f(x) + \epsilon \tag{8}$$

where $f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$ and the noise term $\epsilon \sim N(0,\sigma^2)$. In this study, we have evaluated the performance of various kernels whose functional forms are tabulated in Table 1.

**Table 1. Description of Different Kernels Used in the GPR Modeling**

| kernel | functional form $k(x_i, x_j)$ |
|---|---|
| radial basis functions (RBFs) | $\exp\left(-\frac{d(x_i,x_j)^2}{2l^2}\right)$ |
| RBF() + constant kernel | $\exp\left(-\frac{d(x_i,x_j)^2}{2l^2}\right)$ + constant value |
| rational quadratic | $\left(1 + \frac{d(x_i,x_j)^2}{2\alpha l^2}\right)^{-\alpha}$ |
| rational quadratic + constant kernel | $\left(1 + \frac{d(x_i,x_j)^2}{2\alpha l^2}\right)^{-\alpha}$ + constant value |
| matern | $\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)$ |
| matern + constant kernel () | $\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)$ + constant value |
| dot product | $\sigma_o^2 + x_i \cdot x_j$ |
| dot product + white kernel | $(\sigma_o^2 + x_i \cdot x_j)$ + noise level if $x_i = x_j$ else 0 |
| exponentiation (rational quadratic, p) | $\left(\left(1 + \frac{d(x_i,x_j)^2}{2\alpha l^2}\right)^{-\alpha}\right)^p$ |
| exponentiation (matern, p) | $\left(\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)\right)^p$ |

The radial basis function (RBF), commonly known as the "squared exponential kernel", is the most extensively utilized kernel. RBF is an infinitely differentiable function and has a length-scale parameter ($l > 0$), where the term $d(.,.)$ stands for Euclidean distance in the equation.[56,57] The rational quadratic kernel is an infinite summation or scale mixture of RBF kernels with different length scales with a scale mixture ($\alpha > 0$).[58,59] The Matern kernel is a modified version of the RBF kernel with an extra parameter $\nu$, which measures the smoothness of the function. As the value of $\nu$ increases, the nature of the Matern kernel approaches the simple RBF kernel, and for $\nu = \frac{1}{2}$, it becomes an exponential kernel. In the equation of the Matern kernel, $K_\nu(.)$ is the modified Bessel function and $\Gamma(.)$ is the $\gamma$ function.[60,61] The constant kernel is part of a sum kernel, where it modifies the mean of the Gaussian process and returns a specific value. It is generally added to other kernels to modify



https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

their mean, for instance, the RBF + constant kernel and the Matern + constant kernel. The dot product kernel has a parameter that defines the homogeneity ($\sigma_0^2 = 0$) or non-homogeneity ($\sigma_0^2 \neq 0$).[62,63] It is used separately and combined with a white kernel, which introduces the noise signal. The exponentiation kernel uses one base kernel and a parameter $p$ as the exponent power to the base kernel. The base kernel that is widely used in the exponentiation kernel is RBF or Matern or rational quadratic kernels with an exponent.[64] The deep learning kernel is another kernel that has the following general functional form

$$k(x_i, x_j|\Theta) \rightarrow k(g(x_i, w), g(x_j, w)|w) \tag{9}$$

The term $g(x,w)$ is specially designed for capturing complex signatures in the data set by deep learning, with the optimized weights $w$.[65] The deep kernel structure that we employed is given below as follows $k(x,x^* \mid \Theta)$

$$\sum_{p=1}^{Q} a_p \frac{|\sum_p|^{0.5}}{(2\pi)^{D/2}} \exp\left(-\frac{1}{2}\left|\left|\sum_p^{0.5}(x-x^*)\right|\right|^2\right) \cos(x - x^*, 2\pi\mu_p) \tag{10}$$

where the parameter ($\Theta$) of the above kernel is equivalent to a combination of $a_p$, $\sum_p$, and $\mu_p$, which is a mixture of weight, bandwidth, and frequencies, respectively.[66] For model predictions of GPR for an unseen new input data $x^*$, the following relation will hold

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu^* \end{bmatrix}\begin{bmatrix} k(x, x) + \sigma^2 I & k(x, x^*) \\ k(x^*, x) & k(x^*, x^*) \end{bmatrix}\right) \tag{11}$$

where $f^*$ would follow the distribution as follows

$$f^*|x, y, x^* \sim \mathcal{N}(\mu(f^*), cov(f^*)) \tag{12}$$

with

$$\mu(f^*) = \mu^* + k(x^*, x)[k(x, x) + \sigma^2 I]^{-1}(y - \mu) \tag{13}$$

$$cov(f^*) = k(x^*, x^*) - k(x^*, x)[k(x, x) + \sigma^2 I]^{-1}k(x, x^*) \tag{14}$$

Equations 12−14 are used to perform model predictions with the unseen input instance $x^*$.

For the GPR modeling, the data sets (with labeled inputs and outputs) are split randomly into two parts: (a) train set and (b) test set. The test set is used to benchmark the performance of the built models, while the train set is utilized for training the models. In our work, the data set is divided into 70% and 30% as a training set and a test set, respectively. Before the training phase, a fraction of the train set is randomly chosen as a validation set. The data set is divided into train and test sets numerous times at distinct random instances to eliminate bias in the results.

We also validated and benchmarked different ML models against the GPR models, including linear regression (LR), lasso regression (LOR), ridge regression (RR), elastic-net regression (ELR), principal component regression (PCR), partial least-square regression (PLSR), support vector regression (SVR), and multilayer perceptron (MLP). The performance of all ML models, including GPR, is estimated by two metrics: (1) coefficient of determination ($R^2$) and (2) root-mean-square error (RMSE).[67] The simulations are appropriately performed in the Python environment using Scikit-learn and Pytorch libraries.

**Interpreting the GPR Prediction Model Using SHAP.** ML models tend to dismiss the physical relevance of input features on the predictions. The model would become more generally explainable and easy to interpret if the effect of input qualities on outputs could be established. To this extent, recently, there have been various techniques proposed in the literature, such as LIME,[68] DeepLIFT,[69] layer-wise relevance propagation,[70] Shapley regression values,[71] Shapley sampling values, and quantitative input influence,[72] among others.

The Shapley Additive exPlanation (SHAP) is a technique that helps interpret the impact of a feature value on prediction using the "Shapley value". It is the average value of a feature's marginal contribution over all possible feature combinations. Suppose a regression model and their prediction function be $f'(x_1,\cdots,x_n)$, where $x_1,\cdots,x_n$ are the features. By means of SHAP, the contribution of the $j$th feature is computed as follows

$$\phi_j(f') = f'(x_j) - E(f'(X_j)) \tag{15}$$

where the upper-case letter stands for the feature random variable. $E(f'(X_j))$ is the mean effect estimate for the $j$th feature and their contribution, computed as the difference between the feature prediction and the average estimation. If we add up every feature contribution at once, we get the following

$$\Sigma_{j=1}^{N}\phi_j(f') = \Sigma_{j=1}^{N}(f'(x_j) - f'(E(X_j))) \tag{16}$$

Let $f_x(S)$ denote the estimate feature values in the set $S$ that are minimized over features not in the set $S$

$$f_x(S) = \int f'(x_1 \cdots x_N)\mathrm{d}P_{x\notin S} - E_X(f'(X))) \tag{17}$$

The Shapley value calculation for a feature value is its contribution to the prediction, weighted and summed across all feasible feature value combinations. It is calculated as a feature's average marginal contribution to overall feasible coalitions of features. Thus, the Shapley value $\phi_j(f_x)$ is estimated as follows

$$\phi_j(f_x) = \sum_{S\subseteq N/\{j\}} \frac{|S|!(N-|S|-1)!}{N!}[f_x(S\cup\{j\}) - f_x(S))] \tag{18}$$

where $S$ denotes the subset of features and $N$ denotes the total number of input features. It would be either negative or positive, and a negative value means the feature instance negatively impacts the target value and vice versa.

The explanatory technique implements the additive feature contribution, which relies on the linear combination of a feature set as follows.

$$g(z') = \phi_o + \sum_{j=1}^{N}\phi_j z_j' \tag{19}$$

where $z' \in\{0,1\}^N$ is the coalition vector, $N$ is the number of features, and $\phi_j$ is the attribute by the $j$th feature.[73] In this study, we have interpreted the best-performing GPR models by utilizing the "KernelSHAP" framework. It is appropriate for nonlinear models such as GPR and interprets the feature importance by evaluating the Shapley values.

The KernelSHAP calculates the contributions of each feature value to the estimate for an instance $x$. The KernelSHAP algorithm involves five major steps as follows.



https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

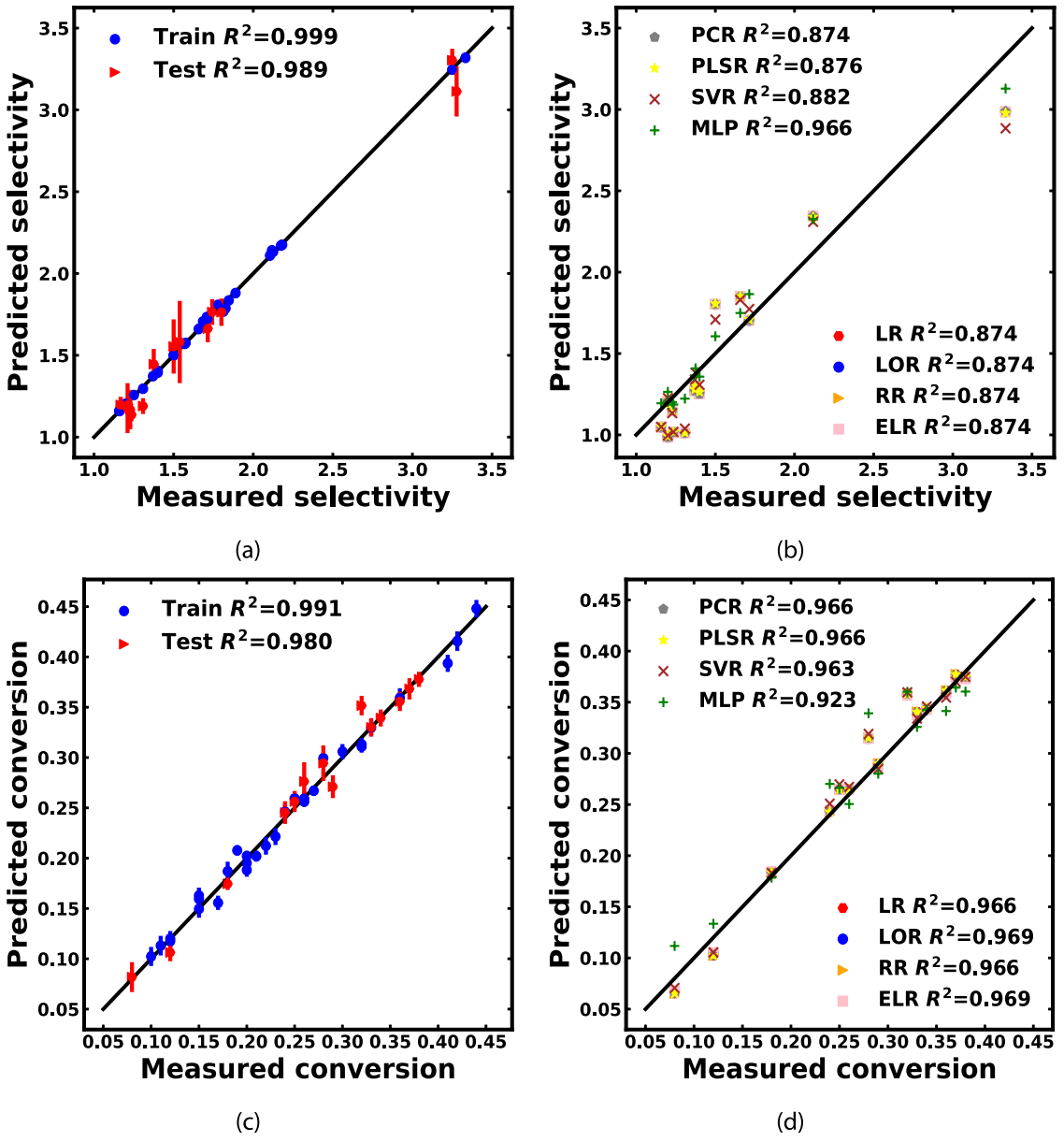



**Figure 2.** Prediction results of recycle reactor models presented as scatter plots. Measured and predicted selectivity along with $R^2$ values. (a) GPR with rational quadratic kernel shown with a standard deviation around mean. (b) Combined results of selectivity prediction of all ML models. Measured and predicted conversion. (c) GPR rational quadratic kernel with standard deviation around mean. (d) Combined results of all models used for conversion prediction. The other ML models used for comparison purposes are LR, LOR, RR, ELR, PCR, PLSR, SVR, and MLP.

1. Sample coalitions $z'_k \in \{0,1\}^M$, where $k = \{0,1,\cdots,K\}$, 1 means the coalition has feature and 0 means feature is absent.
2. Get prediction for each $z'_k$: $f'(h_x(z'_k))$, where $h_x$: $\{0, 1\}^M \rightarrow \mathbb{R}^P$.
3. Estimate the weight for each $z'_k$ using the SHAP kernel $\Pi_x(z')$[73]

$$\Pi_x(z') = \frac{(M-1)}{\binom{M}{|z'|}|z'|(M-|z'|)} \tag{20}$$

where $M$ indicates the maximum size of coalition and $|z'|$ stands for the number of features present for a given instance.

4. Fit the weighted linear model by minimizing the loss $L = \sum_{z'}[f'(h_x(z')) - g(z')]^2\Pi_x(z')$.
5. Return Shapley values for $\phi_k$, the linear model's coefficients.

The computed Shapley values are finally presented as the normalized average of absolute Shapley values per feature across the feature data, $I_k = 1/n \sum_i^n \|\phi_k^i\|$. The high value of $I_k$ means it is a crucial feature while predicting the target variables and vice versa. In this study, SHAP analysis is performed in the Python environment.

**Multiobjective Bayesian Optimization (MOBO).** Bayesian optimization (BO) is a framework to maximize or minimize the objective functions that are nonparametric. In the BO framework, surrogate black-box functions are used to approximate the actual objective function, and subsequently, acquisition functions are utilized to sample the objective


https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

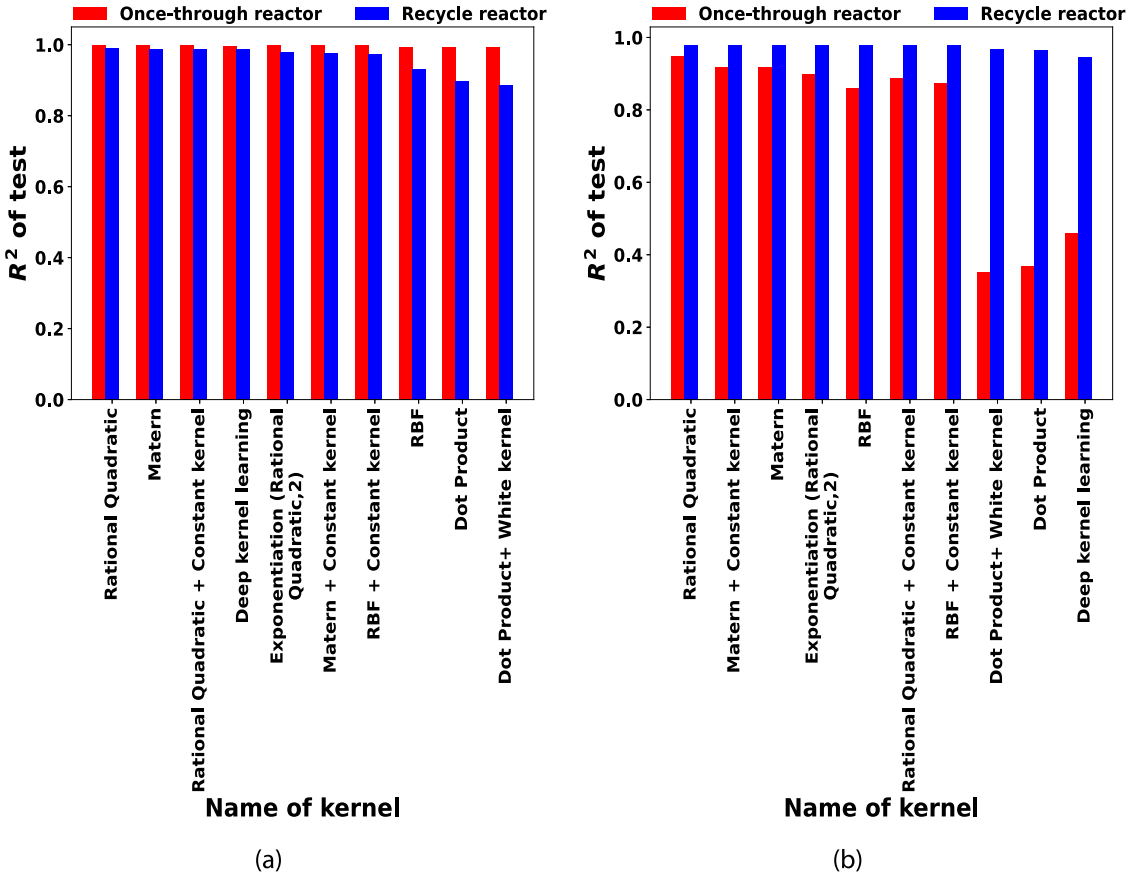



**Figure 3.** $R^2$ metric of the test set for GPR performance with different kernels is shown in bar plots for selectivity and conversion parameters, as displayed in (a) and (b), respectively.

function probabilistically. The most widely used surrogate function in BO is GPR because of its ability to capture complex nonlinear characteristics and the uncertainty in prediction.[74,75]

A key component of BO is the acquisition function. Expected improvement is one of the most extensively used acquisition functions in the BO framework.

$$\mathrm{EI}(x) = (\mu(x) - f(x^+))\Psi\left(\frac{\mu(x) - f(x^+)}{\sigma(x)}\right) + \sigma(x)\phi\left(\frac{\mu(x) - f(x^+)}{\sigma(x)}\right) \tag{21}$$

The mean and variance of the GPR model output at a given point x are $\mu(x)$ and $\sigma(x)$, respectively. $f$ is the objective function, optimized with an estimated maximum at $x^+$, and $\psi(z)$ and $\phi(z)$ denote a standard Gaussian distribution's cumulative and density functions, respectively. The lower confidence bound (LCB) of the acquisition function is defined as follows

$$\mathrm{LCB}(x) = \mu(x) - \kappa\sigma(x) \tag{22}$$

where $\kappa$ is a parameter used for balancing exploitation and exploration.[76] Maximum probability of improvement (MPI) is another option for an acquisition function.

$$\mathrm{MPI}(x) = \phi\left(\frac{E(f(x) - f(x^+) - \epsilon)}{\sigma(f(x))}\right) \tag{23}$$

where $\epsilon$ is used to indicate the percentage of exploitation. The Bayesian optimization is carried out in this work with the help of the GPyOpt toolbox[77] in the Python environment.

This study employs multiobjective Bayesian optimization (MOBO) in a weighted objective framework. A mathematical description of the weighted objective function is given as follows

$$obj_{cov}(x) = (Tar_{conv} - GP_1(x))^2 \tag{24x}$$

$$obj_{sel}(x) = (Tar_{sel} - GP_2(x))^2 \tag{24y}$$

$$obj_{weight}(x) = (w)obj_{cov}(x) + (1 - w)obj_{sel}(x) \tag{24z}$$

where $obj_{cov}(x)$, $obj_{sel}(x)$, and $obj_{weight}(x)$ are the objective function for conversion, the objective function for selectivity, and the multiobjective weighted function, respectively. The proposed interpretable GPR models with an explainable feature set are used as surrogate models for conversion and selectivity, $GP_1(x)$ and $GP_2(x)$, respectively, and $w \in \{0,1\}$. The $Tar_{conv}$ and $Tar_{sel}$ are the chosen valued target variables for conversion and selectivity, respectively, and have been normalized between 0 and 1, by dividing the maximum value of the target variable. eqs 24x and 24y refer to the objective function for conversion and selectivity irrespective of reactor type. Both objectives are combined in the weighted objective function as given in eq 24z.

## RESULTS AND DISCUSSION

First, we consider the prediction of selectivity of methanol and conversion of CO for the recycling reactor as target variables using GPR models. Figure 2a shows the GPR prediction of



https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

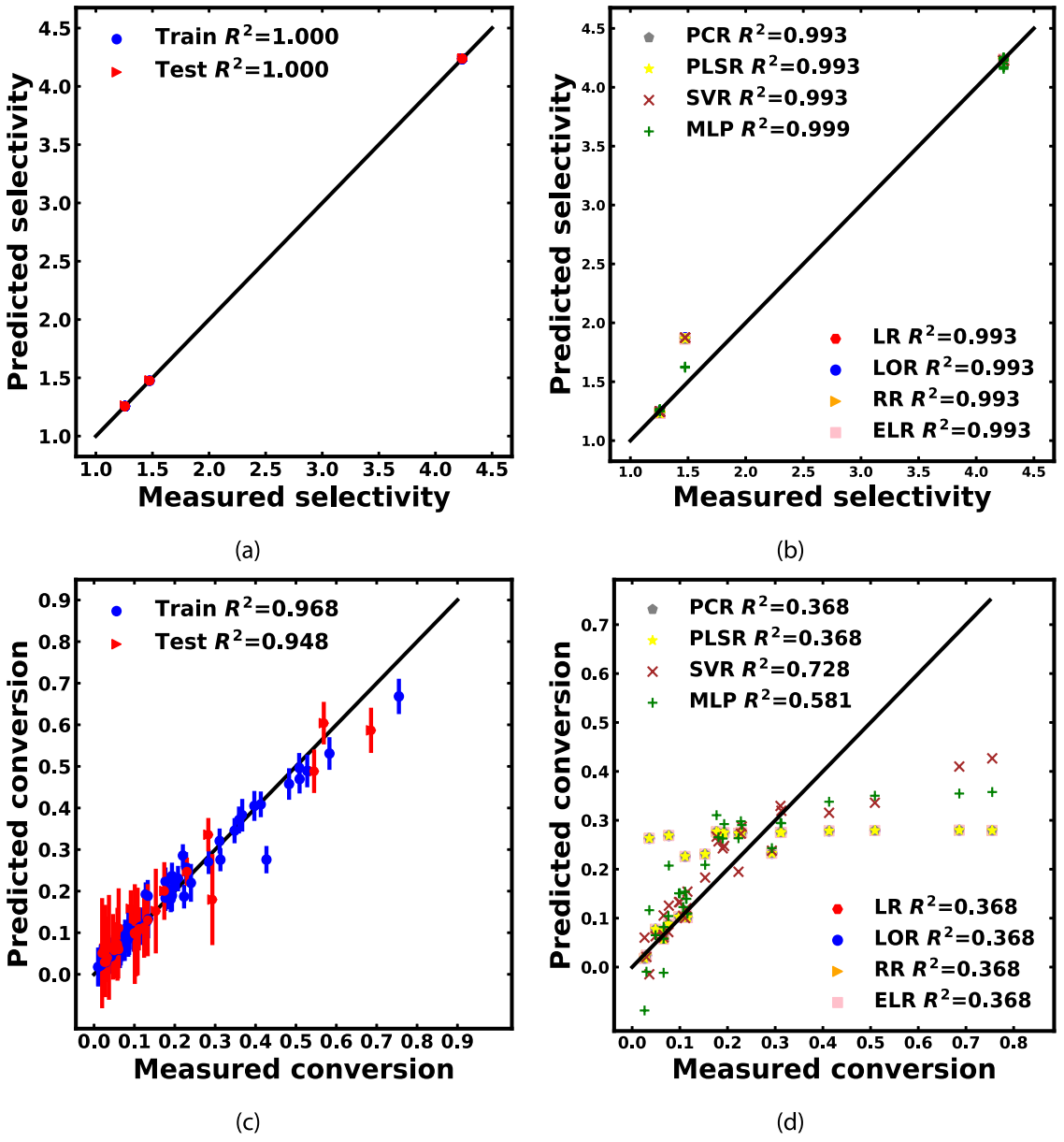



**Figure 4.** Prediction results of once-through reactor models presented as scatter plots. Measured and predicted selectivity along with $R^2$ values. (a) GPR with a rational quadratic kernel with a standard deviation around mean. (b) Combined results of selectivity prediction of all ML models. For conversion, (c) GPR rational quadratic kernel with a standard deviation around mean. (d) Combined results of all models used for conversion prediction. The other ML models used for comparison purposes are LR, LOR, RR, ELR, PCR, PLSR, SVR, and MLP.

selectivity with the standard deviation of prediction using the best-performing kernel compared to the actual experimental data points with values of $R^2$ for the train set and test set as 0.999 and 0.989, respectively. To achieve this, the hyperparameters of the GPR model are tuned with the help of a random search method. We have tried different types of kernels, as indicated in the bar plot in Figure 3; the order of kernels is descending based on their performance. The best-performing kernel is the rational quadratic kernel (length scale = 10.0, $\alpha = 0.5$), with the $R^2$ value of 0.989 for the test, as clearly visible in Figure 3a with the blue-color bar. All kernels' performance ($R^2$ of train and test) and corresponding hyperparameter values are collated in Table S6 in Section S2 of the Supporting Information.

Along similar lines, we proceed to predict the conversion of CO, a key parameter in the reaction. Figure 2c depicts GPR prediction and the standard deviation of prediction for the best kernel, the rational quadratic kernel, against the actual experimental data points. We obtain an $R^2$ of 0.991 for the training data set, and the test data set has an $R^2$ of 0.980. In this kernel, we have used the length scale ($l$) = 1.0 and $\alpha = 0.75$ as the tuned hyperparameter values. We also evaluated the performance of other kernels, and their results in the form of metric $R^2$ with tuned hyperparameters are collated in Table S5 of Section S2 of the Supporting Information and shown in Figure 3b, as indicated in red color. GPR model predictions match the experimental data in both cases as the prediction and the measured data coincide with the 45-degree line. For comparison purposes, the GPR prediction performance is evaluated using the $R^2$ metric on the test data and is reported against other standard ML models LR, LOR, RR, ELR, PCR, PLSR, SVR, and MLP, as shown in Figure 2b,d for selectivity and conversion, respectively. The hyperparameters of all ML models are


https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

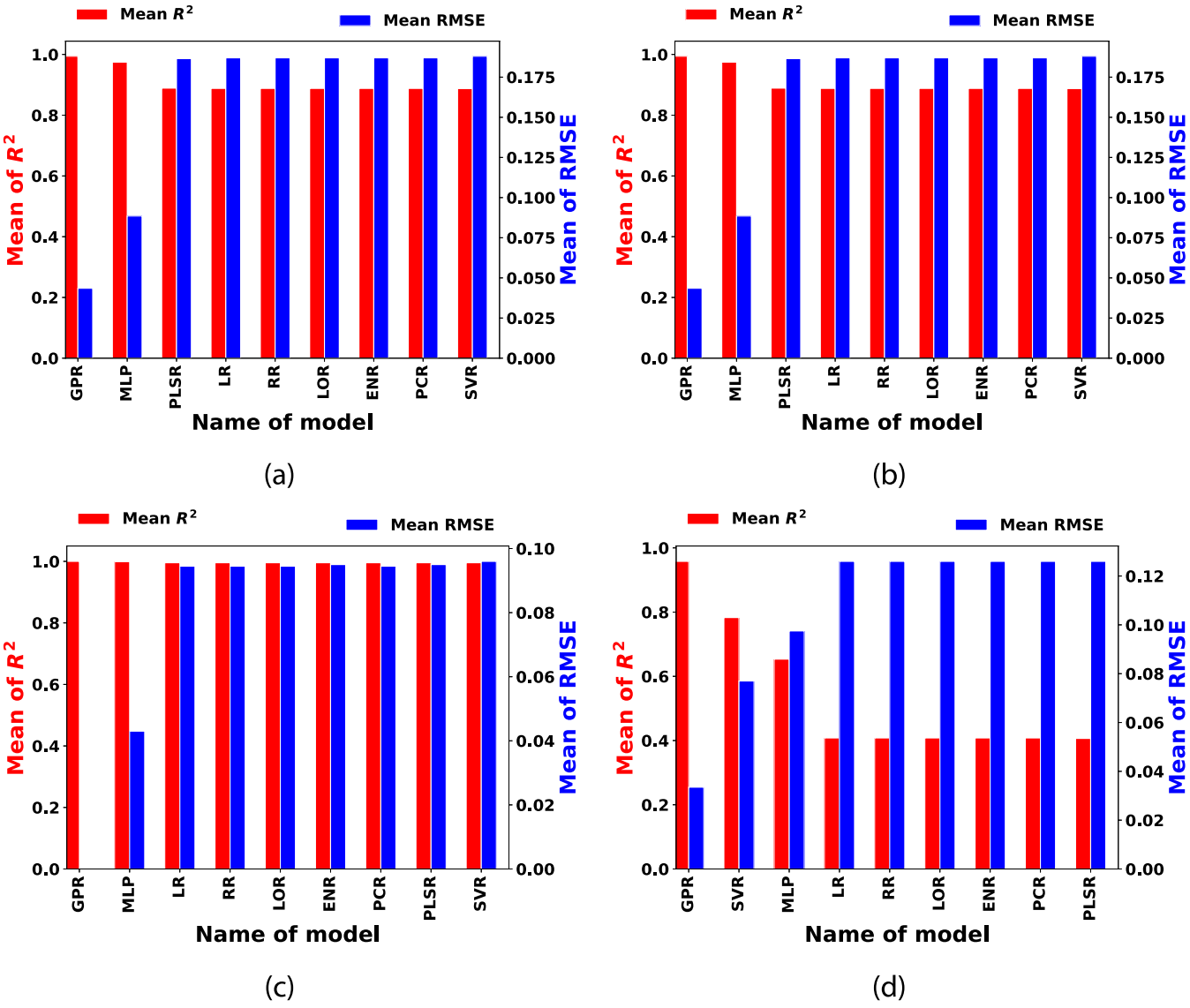



**Figure 5.** Evaluated mean $R^2$ of train and test and mean RMSE of train and test are shown in (a), (b), (c), and (d), where (a) and (b) represent selectivity and conversion in the recycle reactor, respectively. The bar plots (c) and (d) represent selectivity and conversion in the once-through reactor, respectively.

included in the Supporting Information. Tables S1 and S2 represent selectivity and conversion results, respectively. The mean RMSE and mean $R^2$ for all ML models are represented in the form of bar plots in Figure 5a,b for selectivity and conversion, respectively.

Second, we developed GPR models for methanol selectivity and CO conversion in once-through reactors. Interestingly, we do see that various kernels are giving comparable performances, as is clearly shown in Figure 3a with the blue-color bar and in Figure 3b with the red-color bar. In particular, the rational quadratic kernel performs exceptionally well, with $R^2$ values of 1.0 and 0.968 for the train data set and 1.0 and 0.948 for the test data set, as shown in Figure 4a,c for selectivity and conversion, respectively. This agrees well with recycle reactor results, wherein the rational quadratic kernel proved the best for the given cases. In this case, also, we have optimized the GPR model hyperparameters using random search techniques. For selectivity and conversion, respectively, all kernel results are tabulated in Tables S7 and S8 in the Supporting Information. The length-scale parameter of the optimized rational quadratic kernel is 1.0, and $\alpha$ is 0.1 for both case conversion and selectivity, respectively. Similar to the previous case, we have also compared the performance of other ML models against the GPR models. The scattered plots with the test data set's $R^2$ metric of various ML models are showcased in Figure 4b,d for selectivity and conversion, respectively. The detailed analysis regarding the prediction performance of all ML models is collated in Tables S3 and S4 for selectivity and conversion, respectively, and presented in the Supporting Information. Further, the models' mean RMSE and mean $R^2$ are shown in bar plots for selectivity and conversion in Figure 5c,d, respectively.

These results unmistakably demonstrate the GPR prediction's advantage in terms of $R^2$ and RMSE. Since the GPR provides distribution for a prediction value rather than a single point as a prediction, it can quantify the prediction variability. Other ML models cannot account for this prediction variability. By choosing different kernel functions, it has the extraordinary ability to add prior information and specifications about the underlying data's signature. This supports the GPR model's overall ability to predict the reaction parameters.



https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

Next, we interpret our best GPR models by implementing the SHAP technique. It gives us some critical insights related to the feature set. Figures 6 and 7 indicate the SHAP summary plots,

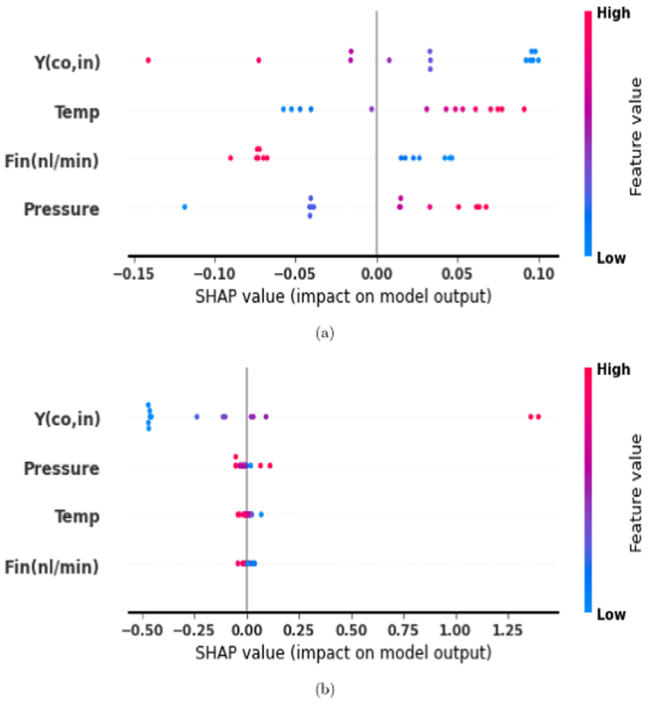



**Figure 6.** Summary plots of SHAP result of (a) conversion and (b) selectivity for the recycle reactor.

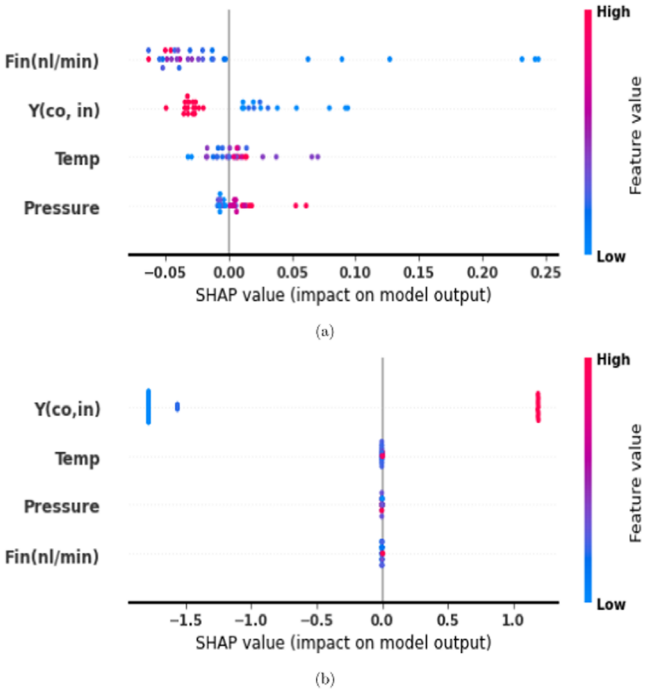



**Figure 7.** Summary plots of SHAP result of (a) conversion and (b) selectivity for the once-through reactor.

which depict the relevance of features and Shapley values. The importance of features is sorted in a descending order (top to bottom), and the color gradient represents the feature values as they progress from low to high. The "rational kernel" performs well with GPR for predicting the recycle reactor's target variables (conversion and selectivity). According to Figure 6a, Y(co, in) is the essential feature. The low values are more dominating (having a positive effect) than the high values of Y(co, in), and pressure seems to be the least contributing. The high pressure and high temperature are preferable for prediction of CO conversion. The Y(co, in) is also a key feature with preferable high values when predicting the selectivity. The least important feature is Fin(nL/min), which has minimal impact on prediction, as shown in Figure 6b.

Similarly, SHAP is used to investigate conversion and selectivity prediction models (best GPR models) for the once-through reactor. Here, GPR with a rational kernel also performed well for both conversion and selectivity as target variables. From the SHAP summary plots, it is evident that for predicting the conversion, Fin(nL/min) is the most relevant feature, and it is preferred to have a slow flow rate that results in a positive Shapley value (i.e., having a positive impact on prediction). Moreover, the least desired feature is pressure, but high pressure is preferred. Moreover, lower values of the Y(co, in) feature have a favorable effect, as shown in Figure 7a. While predicting the selectivity, Y(co, in) seems an essential feature for selectivity. Y(co, in) values are directly proportional to the Shapley value; thus, high values are preferable. The other three features seem to be neutral as their Shapley values are very close to zero. Fin(nL/min) is the least important among the three, as clearly shown in Figure 7b.

Now we proceed to present the results of the MOBO for maximizing both selectivity and conversion together. The three acquisition functions used in this study are EI, LCB, and MPI. The functional form of the weighted objective function is given in eq 24z. Here, we have attempted a range of weights ($w$), i.e., $w \in \{0,1\}$ with the interval of 0.01. The weighted objective function is minimized for each weight ($w$). Figure 8a,b plots show the comparative study of all objectives for the recycling reactor and the once-through reactor data sets, respectively. The black line indicates the weighted objective function, and the red and blue lines depict the conversion and selectivity objectives, respectively. The weighted objective for the recycle reactor in Figure 8a indicates that the conversion objective decreases as the weight $w$ increases from 0, while the selectivity objective stays close to zero. The weighted objective function grows until $w$ = 0.5 and then stays close to the individual objectives, possibly owing to the recycling effect. It is observed that when the weight ($w$) equals to 0.5, all three profiles coincide with each other. In this case, the optimal values of process parameters are 535, 70.3, 0.241, and 0.98 K for temperature, pressure (0.1 MPa), Y(co, in), and Fin(nL/min), respectively. However, for the once-through reactor in Figure 8b, we see that the conversion objective decreases as the weight $w$ increases from 0 and vice versa for the selectivity objective. In this case, it can be seen that the overall weighted objective is to increase initially, reaching a maximum when the weights are equal, and decrease afterward. When the overall weighted objective value is maximum, the optimal values for the process parameters of temperature, pressure (0.1 MPa), Y(co, in), and Fin(nL/min), are 515 K, 5.10, 0.382, and 0.00335, respectively.

## CONCLUSIONS

This study has employed GPR as a machine learning tool to predict methanol selectivity and CO conversion using sparse experimental data sets. We employed experimental data sets of syngas conversion to methanol using a once-through and a


https://doi.org/10.1021/acsomega.2c04919
ACS Omega 2023, 8, 410−421

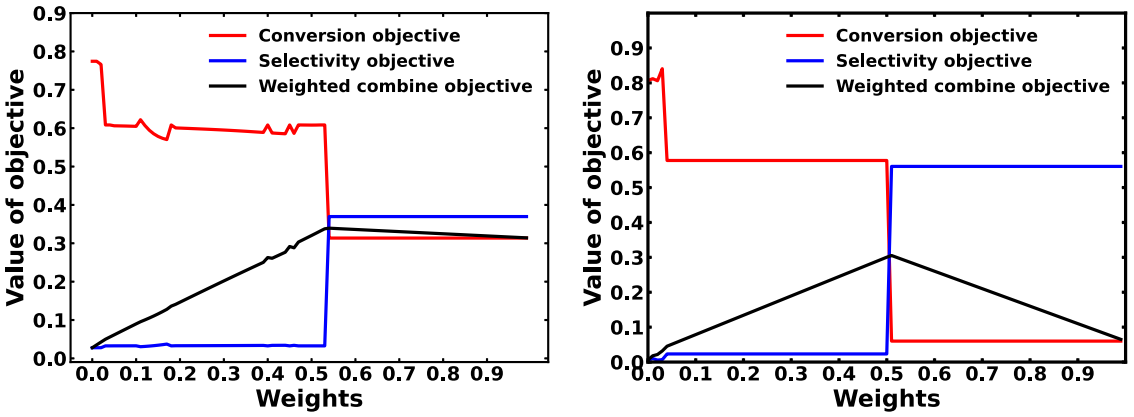



**Figure 8.** Value of objective functions with respect to weight, where (a) and (b) indicate the results of the recycle reactor and the once-through reactor, respectively.

recycle reactor. We have trained the GPR model with various kernels and found that the rational quadratic kernel is the best-performing kernel for conversion and selectivity target variables for either the recycle reactor or the once-through reactor. The $R^2$ metric values are 0.989 and 0.980 for the test data set for predicting selectivity and conversion, respectively, for the recycle reactor. Similarly, the predictions of selectivity and conversion for the once-through reactor are also good, with $R^2$ metric values of 1.0 and 0.948, respectively, for the test data set. We have benchmarked the performance of GPR models with other standard ML models such as LR, LOR, RR, ELR, PCR, PLSR, SVR, and MLP. The results indicate the superiority of the GPR models in predicting reaction progress parameters for methanol formation from syngas by quantifying the uncertainty of the prediction. Final selected GPR models are further interpreted by SHAP, which indicated that the inlet mole fraction of CO (Y(co, in)) and the net inlet flow rate (Fin(nL/min)) are salient features in the data sets that affect the target variables, positively. Further, multiobjective optimization in a weighted objective framework was also carried out using the Bayesian optimization technique for both target variables considering the developed GPR models as surrogate models to obtain the preferred operational trajectory. The MOBO worked well in optimizing the process input parameters with a normalized weight range ($w \in \{0,1\}$) and found that at weight equal to 0.5, the performance is optimal. Overall, it is expected that the adoption of the proposed framework will help accelerate the production of methanol as fuel and, thereby, contribute to lessening environmental footprints.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c04919.

Tables S1−S2, details regarding ML models' hyperparameters values; Tables S5−S8, details regarding hyperparameter values of kernels of GPR; Section S3, information regarding data sets (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

**Sreedevi Upadhyayula** − *Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India;* orcid.org/0000-0001-6042-4401; Email: sreedevi@chemical.iitd.ac.in

**Hariprasad Kodamana** − *Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India; Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi 110016, India;* orcid.org/0000-0003-3166-2712; Email: kodamana@iitd.ac.in

### Authors

**Avan Kumar** − *Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India*

**Kamal K. Pant** − *Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India;* orcid.org/0000-0002-0722-8871

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c04919


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors acknowledge the funding of Grant Nos. DST/TMD/CERI/MDME/2017/001(G) and DST/TM/EWO/MI/CCUS/28(G1) dtd. Sept 28th, 2019 from the Department of Science and Technology, Ministry of Human Resource Development, New Delhi, India, and 51/14/11/2019-BRNS and CRG/2018/001555, with BRNS and SERB India, respectively.

## ■ REFERENCES

(1) Meshkini, F.; Taghizadeh, M.; Bahmani, M. Investigating the effect of metal oxide additives on the properties of Cu/ZnO/Al2O3 catalysts in methanol synthesis from syngas using factorial experimental design. *Fuel* **2010**, *89*, 170−175.

(2) Leonzio, G. Methanol synthesis: optimal solution for a better efficiency of the process. *Processes* **2018**, *6*, 20.

(3) Cheng, W.-H.*Methanol Production and Use*; CRC Press, 1994.

(4) Schiaroli, N.; Volanti, M.; Crimaldi, A.; Passarini, F.; Vaccari, A.; Fornasari, G.; Copelli, S.; Florit, F.; Lucarelli, C. Biogas to syngas through the combined steam/dry reforming process: an environmental impact assessment. *Energy Fuels* **2021**, *35*, 4224−4236.

(5) Bermúdez, J.; Fidalgo, B.; Arenillas, A.; Menéndez, J. Dry reforming of coke oven gases over activated carbon to produce syngas for methanol synthesis. *Fuel* **2010**, *89*, 2897−2902.

(6) Wang, M.; Cai, Z.; Zhang, B.; Yang, K.; Shou, T.; Bernards, M. T.; Xie, P.; He, Y.; Shi, Y. Electrochemical Reduction of CO2 on Copper-Based Electrocatalyst Supported on MWCNTs with Different Functional Groups. *Energy Fuels* **2022**, *36*, 5833−5842.
(7) Galadima, A.; Muraza, O. From synthesis gas production to methanol synthesis and potential upgrade to gasoline range hydrocarbons: A review. *J. Nat. Gas Sci. Eng.* **2015**, *25*, 303−316.
(8) Heydorn, E.; Stein, V.; Tijm, P.; Street, B.; Kornosky, R. In *Liquid Phase Methanol (LPMEOHTM Project Operational Experience*, Presented at the Gasification Technology Council Meeting in San Francisco on October, 1998.
(9) Gutiérrez Ortiz, F.; Serrera, A.; Galera, S.; Ollero, P. Methanol synthesis from syngas obtained by supercritical water reforming of glycerol. *Fuel* **2013**, *105*, 739−751.
(10) Previtali, D.; Longhi, M.; Galli, F.; Di Michele, A.; Manenti, F.; Signoretto, M.; Menegazzo, F.; Pirola, C. Low pressure conversion of CO2 to methanol over Cu/Zn/Al catalysts. The effect of Mg, Ca and Sr as basic promoters. *Fuel* **2020**, *274*, No. 117804.
(11) Singh, J.; Kumar, A.; Jaiswal, A.; Suman, S.; Jaiswal, R. P. Luminescent down-shifting natural dyes to enhance photovoltaic efficiency of multicrystalline silicon solar module. *Solar Energy* **2020**, *206*, 353−364.
(12) Nakamura, J.; Uchijima, T.; Kanai, Y.; Fujitani, T. The role of ZnO in Cu/ZnO methanol synthesis catalysts. *Catal. Today* **1996**, *28*, 223−230.
(13) Damma, D.; Smirniotis, P. G. Recent advances in the direct conversion of syngas to oxygenates. *Catal. Sci. Technol.* **2021**, *11*, 5412−5431.
(14) Wong, W.; Daud, W.; Mohamad, A.; Kadhum, A.; Loh, K.; Majlan, E. Recent progress in nitrogen-doped carbon and its composites as electrocatalysts for fuel cell applications. *Int. J. Hydrogen Energy* **2013**, *38*, 9370−9386.
(15) Cui, X.; Kær, S. K. A comparative study on three reactor types for methanol synthesis from syngas and CO2. *Chem. Eng. J.* **2020**, *393*, No. 124632.
(16) Wang, Z.; Wang, Y.; Ning, S.; Kang, Q. Zinc-Based Materials for Photoelectrochemical Reduction of Carbon Dioxide. *Energy Fuels* **2022**, *36*, 11380−11393.
(17) Slotboom, Y.; Bos, M.; Pieper, J.; Vrieswijk, V.; Likozar, B.; Kersten, S.; Brilman, D. Critical assessment of steady-state kinetic models for the synthesis of methanol over an industrial Cu/ZnO/Al2O3 catalyst. *Chem. Eng. J.* **2020**, *389*, No. 124181.
(18) Chiavassa, D. L.; Collins, S. E.; Bonivardi, A. L.; Baltanás, M. A. Methanol synthesis from CO2/H2 using Ga2O3-Pd/silica catalysts: Kinetic modeling. *Chem. Eng. J.* **2009**, *150*, 204−212.
(19) Xu, D.; Wu, P.; Yang, B. Origin of CO 2 as the main carbon source in syngas-to-methanol process over Cu: theoretical evidence from a combined DFT and microkinetic modeling study. *Catal. Sci. Technol.* **2020**, *10*, 3346−3352.
(20) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.-i. Machine learning for catalysis informatics: recent applications and prospects. *ACS Catal.* **2020**, *10*, 2260−2297.
(21) Ortuño, M. A.; López, N. Reaction mechanisms at the homogeneous−heterogeneous frontier: insights from first-principles studies on ligand-decorated metal nanoparticles. *Catal. Sci. Technol.* **2019**, *9*, 5173−5185.
(22) Yan, Y.; Mattisson, T.; Moldenhauer, P.; Anthony, E. J.; Clough, P. T. Applying machine learning algorithms in estimating the performance of heterogeneous, multi-component materials as oxygen carriers for chemical-looping processes. *Chem. Eng. J.* **2020**, *387*, No. 124072.
(23) Shi, Z.; Liang, H.; Yang, W.; Liu, J.; Liu, Z.; Qiao, Z. Machine learning and in silico discovery of metal-organic frameworks: Methanol as a working fluid in adsorption-driven heat pumps and chillers. *Chem. Eng. Sci.* **2020**, *214*, No. 115430.
(24) Zhu, X.; Wan, Z.; Tsang, D. C.; He, M.; Hou, D.; Su, Z.; Shang, J. Machine learning for the selection of carbon-based materials for tetracycline and sulfamethoxazole adsorption. *Chem. Eng. J.* **2020**, *406*, No. 126782.
(25) Kumar, A.; Ganesh, S.; Gupta, D.; Kodamana, H. A text mining framework for screening catalysts and critical process parameters from scientific literature-a study on Hydrogen production from alcohol. *Chem. Eng. Res. Des.* **2022**, *184*, 90−102.
(26) Bos, A.; Borman, P.; Kuczynski, M.; Westerterp, K. The kinetics of the methanol synthesis on a copper catalyst: an experimental study. *Chem. Eng. Sci.* **1989**, *44*, 2435−2449.
(27) Kuczynski, M.; Browne, W.; Fontein, H.; Westerterp, K. Reaction kinetics for the synthesis of methanol from CO and H2 on a copper catalyst. *Chem. Eng. Process.* **1987**, *21*, 179−191.
(28) Waugh, K. Methanol synthesis. *Catal. Today* **1992**, *15*, 51−75.
(29) Khzouz, M.; Gkanas, E. I.; Du, S.; Wood, J. Catalytic performance of Ni-Cu/Al2O3 for effective syngas production by methanol steam reforming. *Fuel* **2018**, *232*, 672−683.
(30) Daemi, A.; Kodamana, H.; Huang, B. Gaussian process modelling with Gaussian mixture likelihood. *J. Process Control* **2019**, *81*, 209−220.
(31) Bishnoi, S.; Singh, S.; Ravinder, R.; Bauchy, M.; Gosvami, N. N.; Kodamana, H.; Krishnan, N. A. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning. *J. Non-Cryst. Solids* **2019**, *524*, No. 119643.
(32) Li, Z.; Hong, X.; Hao, K.; Chen, L.; Huang, B. Gaussian process regression with heteroscedastic noises—machine-learning predictive variance approach. *Chem. Eng. Res. Des.* **2020**, *157*, 162−173.
(33) Seeger, M. Gaussian processes for machine learning. *Int. J. Neural Syst.* **2004**, *14*, 69−106.
(34) Quiñonero, J.; Quiñonero-Candela, Q.; Rasmussen, C. E.; De, C. M. A Unifying View of Sparse Approximate Gaussian Process Regression. *J. Mach. Learn. Res.* **2005**, 1939−1959.
(35) Smola, A. J.; Bartlett, P. L. Sparse greedy Gaussian process regression. *Adv. Neural Inf. Process. Syst.* **2001**, 619−625.
(36) Bernardo, J.; Berger, J.; Dawid, A.; Smith, A.et al.Regression and Classification Using Gaussian Process Priors. In *Bayesian Statistics*; Oxford University Press, 1998; Vol. *6*, p 475.
(37) Wu, Y.; Zhou, Y. Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. *Constr. Build. Mater.* **2022**, *330*, No. 127298.
(38) Ekanayake, I.; Meddage, D.; Rathnayake, U.A Novel Approach to Explain the Black-Box Nature of Machine Learning in Compressive Strength Predictions of Concrete using Shapley Additive Explanations (SHAP). In *Case Studies in Construction Materials*, Elsevier, 2022; Vol. *16*, p e01059.
(39) Zaki, M.; Venugopal, V.; Bhattoo, R.; Bishnoi, S.; Singh, S. K.; Allu, A. R.; Krishnan, N. A. Interpreting the optical properties of oxide glasses with machine learning and shapely additive explanations. *J. Am. Ceram. Soc.* **2022**, *105*, 4046−4057.
(40) Onsree, T.; Tippayawong, N.; Phithakkitnukoon, S.; Lauterbach, J. Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass. *Energy* **2022**, *249*, No. 123676.
(41) Park, H.; Park, D. Y. Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms. *Build. Sci.* **2021**, *195*, No. 107744.
(42) Liang, M.; Chang, Z.; Wan, Z.; Gan, Y.; Schlangen, E.; Šavija, B. Interpretable Ensemble-Machine-Learning models for predicting creep behavior of concrete. *Cem. Concr. Compos.* **2022**, *125*, No. 104295.
(43) Kaddani, S.; Vanderpooten, D.; Vanpeperstraete, J.-M.; Aissi, H. Weighted sum model with partial preference information: Application to multi-objective optimization. *Eur. J. Oper. Res.* **2017**, *260*, 665−679.
(44) Loka, N.; Couckuyt, I.; Garbuglia, F.; Spina, D.; Van Nieuwenhuyse, I.; Dhaene, T.Bi-objective Bayesian Optimization of Engineering Problems with Cheap and Expensive Cost Functions. In *Engineering with Computers*; Springer, 2022; pp 1−11.
(45) Park, S.; Na, J.; Kim, M.; Lee, J. M. Multi-objective Bayesian optimization of chemical reactor design using computational fluid dynamics. *Comput. Chem. Eng.* **2018**, *119*, 25−37.
(46) Yamamoto, Y.; Yajima, T.; Kawajiri, Y. Uncertainty quantification for chromatography model parameters by Bayesian inference using sequential Monte Carlo method. *Chem. Eng. Res. Des.* **2021**, *175*, 223−237.


https://doi.org/10.1021/acsomega.2c04919
*ACS Omega* 2023, 8, 410−421

(47) Park, S.; Atwair, M.; Kim, K.; Lee, U.; Na, J.; Zahid, U.; Lee, C.-J. Bayesian optimization of industrial-scale toluene diisocyanate liquid-phase jet reactor with 3-D computational fluid dynamics model. *J. Ind. Eng. Chem.* **2021**, *98*, 327−339.
(48) Wauters, J. ERGO: a new Robust Design Optimization Technique combining Multi-Objective Bayesian Optimization with Analytical Uncertainty Quantification. *J. Mech. Des.* **2022**, *144*, 1−22.
(49) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021**, *590*, 89−96.
(50) Agrawal, D.; Sharma, R.; Ramteke, M.; Kodamana, H. Hierarchical two-tier optimization framework for the optimal operation of a network of hybrid renewable energy systems. *Chem. Eng. Res. Des.* **2021**, *175*, 37−50.
(51) Jorayev, P.; Russo, D.; Tibbetts, J. D.; Schweidtmann, A. M.; Deutsch, P.; Bull, S. D.; Lapkin, A. A. Multi-objective Bayesian optimisation of a two-step synthesis of p-cymene from crude sulphate turpentine. *Chem. Eng. Sci.* **2022**, *247*, No. 116938.
(52) Iwama, R.; Kaneko, H. Design of ethylene oxide production process based on adaptive design of experiments and Bayesian optimization. *J. Adv. Manuf. Process.* **2021**, *3*, No. e10085.
(53) Acquarola, C.; Ao, M.; Bhatelia, T.; Prakash, B.; Faka, S.; Pareek, V.; Shah, M. T. Simulations and Optimization of a Reduced CO2 Emission Process for Methanol Production Using Syngas from Bi-reforming. *Energy Fuels* **2021**, *35*, 8844−8856.
(54) Alarifi, A.; Alsobhi, S.; Elkamel, A.; Croiset, E. Multiobjective optimization of methanol synthesis loop from synthesis gas via a multibed adiabatic reactor with additional interstage CO2 quenching. *Energy Fuels* **2015**, *29*, 530−537.
(55) Iwama, R.; Takizawa, K.; Shinmei, K.; Baba, E.; Yagihashi, N.; Kaneko, H. Design and Analysis of Metal Oxides for CO2 Reduction Using Machine Learning, Transfer Learning, and Bayesian Optimization. *ACS Omega* **2022**, *7*, 10709−10717.
(56) Williams, C.; Seeger, M. In *The Effect of the Input Density Distribution on Kernel-Based Classifiers*, Proceedings of the 17th International Conference on Machine Learning, IEEE, 2000; pp 1159−1166.
(57) Chang, Q.; Chen, Q.; Wang, X. In *Scaling Gaussian RBF Kernel Width to Improve SVM Classification*, 2005 International Conference on Neural Networks and Brain, IEEE, 2005; pp 19−22.
(58) Wilson, A.; Adams, R. In *Gaussian Process Kernels for Pattern Discovery and Extrapolation*, International Conference on Machine Learning, PMLR, 2013; pp 1067−1075.
(59) Hong, S.; Zhou, Z.; Lu, C.; Wang, B.; Zhao, T. Bearing remaining life prediction using Gaussian process regression with composite kernel functions. *J. Vibroengineering* **2015**, *17*, 695−704.
(60) Melkumyan, A.; Ramos, F. In *Multi-Kernel Gaussian processes*, Twenty-Second International Joint Conference On Artificial Intelligence, 2011; pp 1408−1413.
(61) Borovitskiy, V.; Azangulov, I.; Terenin, A.; Mostowsky, P.; Deisenroth, M.; Durrande, N.Matérn Gaussian Processes on Graphs. In *Artificial Intelligence and Statistics*; PMLR, 2021; pp 2593−2601.
(62) Kar, P.; Karnick, H.Random Feature Maps for Dot Product Kernels. In *Artificial Intelligence and Statistics*; PMLR, 2012; pp 583−591.
(63) Smola, A.; Ovári, Z.; Williamson, R. C. Regularization with dot-product kernels. *Adv. Neural Inf. Process. Syst.* **2000**, *13*, 308−314.
(64) Phillips, P. C. B.; Sun, Y.; Jin, S. Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *Int. Econ. Rev.* **2006**, *47*, 837−894.
(65) Wilson, A. G.; Hu, Z.; Salakhutdinov, R.; Xing, E. P.Deep Kernel Learning. In *Artificial Intelligence and Statistics*; PMLR, 2016; pp 370−378.
(66) Rasmussen, C. E.Gaussian Processes in Machine Learning. In *Summer School on Machine Learning*; Springer, 2003; pp 63−71.
(67) Deng, B. Machine learning on density and elastic property of oxide glasses driven by large dataset. *J. Non-Cryst. Solids* **2020**, *529*, No. 119768.
(68) Ribeiro, M. T.; Singh, S.; Guestrin, C.et al. In *Why Should I Trust You? Explaining the Predictions of any Classifier*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2016; pp 1135−1144.
(69) Shrikumar, A.; Greenside, P.; Kundaje, A. In *Learning Important Features through Propagating Activation Differences*, International Conference on Machine Learning, PMLR, 2017; pp 3145−3153.
(70) Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* **2015**, *10*, No. e0130140.
(71) Lipovetsky, S.; Conklin, M.Analysis of Regression in Game Theory Approach. In *Applied Stochastic Models in Business and Industry*; John Wiley and Sons Ltd, 2001; Vol. *17*, pp 319−330.
(72) Štrumbelj, E.; Kononenko, I.Explaining Prediction Models and Individual Predictions with Feature Contributions. In *Knowledge and Information Systems*; Springer, 2014; Vol. *41*, pp 647−665.
(73) Lundberg, S.; Lee, S. A unified approach to interpreting model predictions, Adv. Neural Inf. *Process. Syst.* **2017**, 4766−4775.
(74) Snoek, J.; Rippel, O.; Swersky, K.; Kiros, R.; Satish, N.; Sundaram, N.; Patwary, M.; Prabhat, M.; Adams, R. In *Scalable Bayesian Optimization Using Deep Neural Networks*, International Conference on Machine Learning, PMLR, 2015; pp 2171−2180.
(75) Esche, E.; Weigert, J.; Rihm, G. B.; Göbel, J.; Repke, J.-U. Architectures for neural networks as surrogates for dynamic systems in chemical engineering. *Chem. Eng. Res. Des.* **2022**, *177*, 184−199.
(76) Lyu, W.; Yang, F.; Yan, C.; Zhou, D.; Zeng, X. In *Batch Bayesian Optimization via Multi-Objective Acquisition Ensemble for Automated Analog Circuit Design*, International Conference on Machine Learning, PMLR, 2018; pp 3306−3314.
(77) GPyOpt: A Bayesian Optimization Framework in Python, 2016. http://github.com/SheffieldML/GPyOpt.


https://doi.org/10.1021/acsomega.2c04919
*ACS Omega* 2023, 8, 410−421