ACS OMEGA

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

Article

Integration of Materials and Process Informatics: Metal Oxide and Process Design for CO_2 Reduction

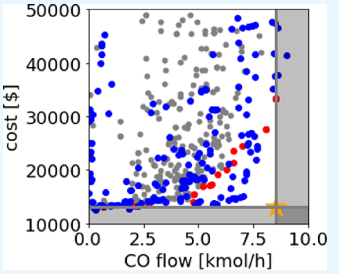
Ryo Iwama and Hiromasa Kaneko*

Cite This: *ACS Omega* 2022, 7, 46922−46934

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In materials informatics, a mathematical model constructed between the synthesis conditions of materials and their properties and activities is used to design synthesis conditions in which the properties and activities have the desired values. In process informatics, a mathematical model constructed between the process conditions for devices and industrial plants and product quality and cost is used to design process conditions that can produce the desired products. In this study, we propose a method to simultaneously design the synthesis conditions of materials and the process conditions of products by integrating materials and process informatics in the reverse water-gas shift chemical looping (RWGS-CL) reaction, which produces CO from CO_2 using metal oxides via the RWGS-CL process. Four methods: Gaussian process regression-Bayesian optimization (GPR-BO), Gaussian mixture regression−Bayesian optimization (GMR-BO), GMR-BO-multiple, and GPR-GMR-BO were investigated for the optimization. All four proposed methods outperformed the results of a random search. GPR-BO achieved the highest performance and proposed 27 promising candidates for the synthesis conditions and metal oxides. The selected metals did not include Cu and Ga, which tended to have high predicted CO_2 and H_2 conversion rates, but Fe and La, which had slightly lower predicted CO_2 and H_2 conversion rates. These results indicate that a combination of metal oxides with lower predicted CO_2 and H_2 conversion rates and optimized process conditions was important for the optimization of both materials and processes, which was achieved by integrating materials and process informatics via the proposed method. Thus, we confirmed that it is possible to simultaneously optimize the combination of metals, composition ratios, synthesis conditions of the material or the metal oxide, and the process conditions using experimental datasets, process simulations, and machine learning, such as GPR, GMR, BO, and multiobjective optimization with a genetic algorithm.



INTRODUCTION

Various chemicals are manufactured from fossil fuels, such as crude oil, coal, and natural gas, using chemicals and polymers such as ethylene and propylene. After use, some chemicals are recycled; however, most are incinerated or landfilled. A large amount of carbon dioxide (CO_2) is emitted during the manufacture and incineration of chemicals, which is known to cause global warming.

Many efforts have been made to capture and recycle CO_2 emitted from factories, power plants, and other sources. In this study, we focused on the reverse water-gas shift chemical looping (RWGS-CL) reaction,[1] which produces carbon monoxide (CO) from CO_2 using metal oxides (MO_x, MO_{x-1}) as oxygen carriers. The RWGS-CL reaction proceeds via two separate half-reactions, as follows:

$$H_2 + MO_x \rightleftarrows H_2O + MO_{x-1} \quad (1)$$

$$CO_2 + MO_{x-1} \rightleftarrows CO + MO_x \quad (2)$$

The RWGS-CL reaction can achieve a higher efficiency than the RWGS reaction because the reaction of H_2O and CO (water-gas reaction) is irreversible. In addition, a lack of contact between H_2 and CO prevents side reactions and methanation. Furthermore, because H_2O and CO need not be separated, a distillation column need not be installed. The RWGS-CL reaction can be used to investigate the effects of metals such as Cu, Sr, La, Ca, Co, and Mn on perovskite oxides[2−7] and supports such as alumina and silica[8−13] and process analysis based on thermodynamic data,[14] to find metal oxides with high CO_2 and H_2 conversion rates and their experimental and process conditions. However, because the redox reactions in eqs 1 and 2 are repeated, the RWGS-CL reaction is difficult to control, and the CO_2 and H_2 conversion rates vary significantly with metal oxides and their synthesis conditions. Owing to the trade-off between them, high CO_2 and H_2 conversion rates necessitate the development of metal oxides and synthesis and process conditions.

Manufacturing the desired product via the RWGS-CL reaction process requires metal oxide and process designs and testing in pilot and actual plants. In recent years, both

Received: September 16, 2022
Accepted: November 22, 2022
Published: December 6, 2022

metal oxide[15] and process designs[16] have been conducted using machine learning. A statistical model $y = f(x)$ is constructed using a dataset between x and y, which represent the synthesis conditions and the properties and activities, respectively, and the dataset is an experimental dataset in metal oxide design. In models wherein x and y represent the process condition and product quality, respectively, the dataset is a simulation dataset in process design. By substituting various values of x, such as synthesis and process conditions, into the statistical model, property, activity, and product quality values can be estimated without experiments and simulations. Experiments and simulations are conducted only under synthesis and process conditions that have desirable property, activity, and process quality values, allowing an efficient search for target synthesis and process conditions. However, in conventional material and product development, the synthesis conditions with the target property and activity values and process conditions for the target process are designed independently and separately. Thus, information on the process is not considered while designing metal oxides and that on the metal oxides is not considered during process design, preventing the determination of promising synthesis and process conditions from both perspectives and a global optimum solution incorporating both material and process.

The objective of this study is to simultaneously optimize metal oxides, their synthesis conditions for high CO_2 and H_2 conversion rates, and process conditions in an actual plant and to integrate metal oxide and process designs. The synthesis conditions of new metal oxides and process conditions were proposed via the Bayesian optimization using a mathematical model in which x represents the synthesis conditions and metal oxide descriptors and y is the experimental CO_2 and H_2 conversion rates and another mathematical model in which x represents the predicted results of the first mathematical model and process conditions in the RWGS-CL process and y is the target process requirement. Gaussian process regression (GPR)[17] and Gaussian mixture regression (GMR)[18] are regression model construction methods. The GPR model can predict not only the y values but also their variance, by the inputting x for which the experimental and simulation results are unknown. From the prediction results, the x values with a high probability of achieving the target process requirements are predicted. The GMR model can directly predict x values by inputting target y values into the model; this is called direct inverse analysis.[19,20] The process simulations of the RWGS-CL were then conducted for the designed x values to obtain the y values. By repeating the GPR and GMR model construction for designing the x values, and the process simulations using them, both the synthesis conditions of the optimal metal oxides and the optimal process conditions for the target process requirements were achieved via a few process simulations.

RESULTS AND DISCUSSION

Data and Simulation. Figure 1 shows the RWGS-CL process flow diagram.[21] The raw materials, CO_2 and H_2, are switched in the oxidation state of the metal oxide. The blue and red arrows indicate the flow for the reactions in eqs 1 and 2, respectively.

Experimental data and descriptors of metal oxides, such as Pymatgen[22] and Materials project descriptors,[23] were obtained from a previous report,[15] and regression models were constructed to predict CO_2 and H_2 conversion rates as

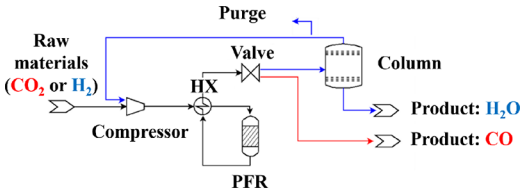


Figure 1. Process flow diagram of RWGS-CL. The blue, red, and black lines represent H_2O synthesis, CO synthesis, and common flows, respectively.

previously reported.[15] In the case of the CO_2 conversion rate, x represented the experimental conditions and Pymatgen and materials project descriptors. For the H_2 conversion rate, represented the experimental conditions and Pymatgen descriptors. The GPR models were constructed using transfer learning. Figure 2 shows the results of double-cross-

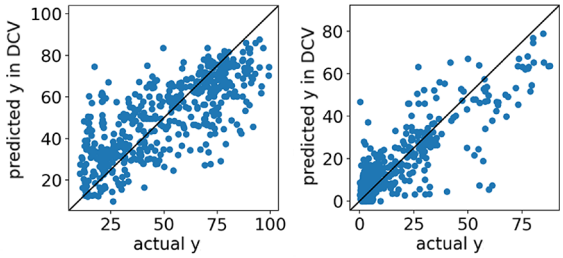


Figure 2. Plots of the actual and predicted y in DCV. (a, b) Mean CO_2 conversion rate and H_2 conversion rate, respectively.

validation[24] for each prediction model construction method. It was confirmed that GPR models can predict CO_2 and H_2 conversion rates with a certain degree of predictive ability, even for new metal oxides. Using the GPR models, we predicted the CO_2 and H_2 conversion rates at 500 and 650 °C and obtained the reaction-rate equations shown in eqs 3 and 4, respectively.

All equipment and operating conditions were set as x for the process conditions, based on the PFD shown in Figure 1. Table 1 presents the process conditions and their ranges. The

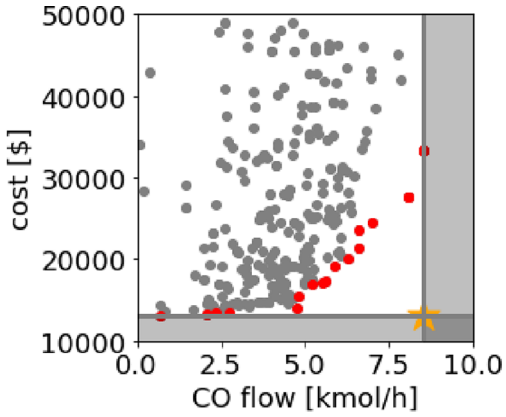
Table 1. Settings of the Process Conditions

	unit	place in Figure 1	lower limit	upper limit
pressure of CO_2	bar	compressor	5	10
temperature of CO_2	K	HX	400	1.00×10^3
pipe length	m	PFR	1	6
inner diameter of the pipes	m	PFR	0.01	0.1
number of pipes		PFR	10	3.00×10^3
pressure of H_2	bar	compressor	5	10
temperature of H_2	K	HX	500	1.00×10^3

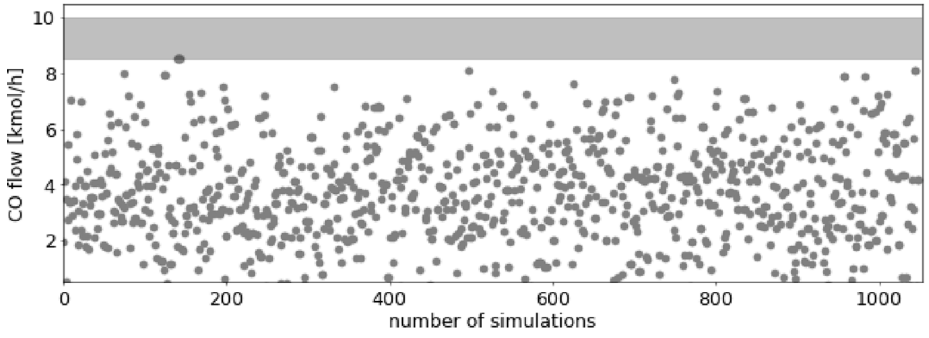
flow rate of the product, CO, and the cost of production were set to y. The cost was the sum of the construction, operation, metal oxide, and feedstock costs, and y was calculated via a process simulation based on the reaction-rate equations in eqs 3 and 4 and set as the process conditions.

Normally, target y values are set based on the performance and cost required in a plant. However, in this study, to verify the efficiency and effectiveness of the proposed method, 1000 samples of synthesis and process conditions were randomly selected, and the Pareto optimal solution obtained from the

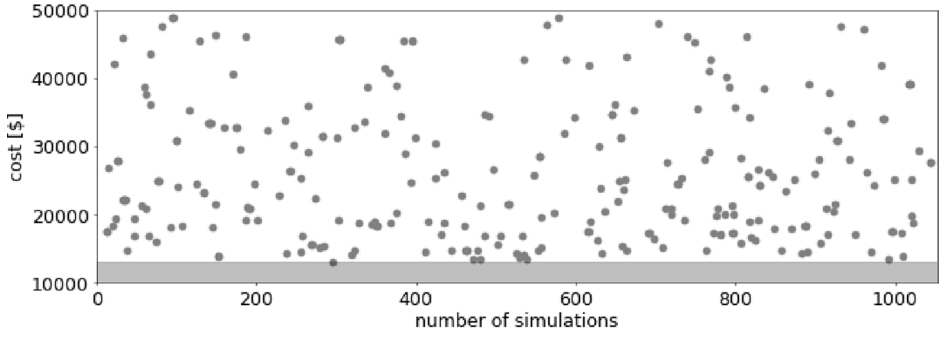
https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934



(a) CO flow versus cost



(b) CO flow per simulation



(c) Cost per simulation.

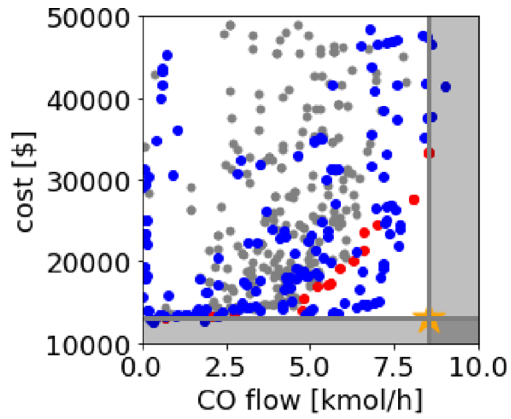
Figure 3. Results of running simulations with randomly selected synthesis and process conditions. (a−c) Mean CO flow versus cost, CO flow per simulation, and cost per simulation, respectively. The gray regions, gray points, red points, and yellow star represent the target ranges, randomly selected samples, Pareto optimal set of gray points, and target point, respectively.

1000 simulations was set as the target y value. Figure 3 shows the results of the 1000 simulations. The maximum and minimum values of the red points, corresponding to the Pareto optimal solutions in Figure 3a, were used as the target y values.

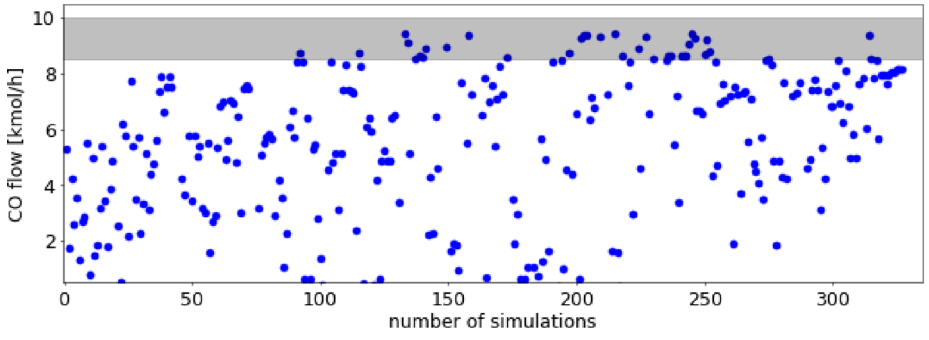
Thirty candidates for the synthesis and process conditions were randomly selected as the initial dataset of x to construct the GPR and GMR models. The simulations were run 30 times to obtain the y values and construct the first GPR model in GPR-BO or the first GMR model in GMR-BO between x and y.

Design of Materials and Processes. In GPR-BO, the GPR model was constructed using x as the experimental condition, metal oxide features, CO_2 and H_2 conversion rates, and process conditions and y as the CO flow rate and cost. The results of the cross-validation evaluation showed that the CO-flow and cost prediction models had r^2 = 0.642 and 0.584, respectively. After the model construction, 30 candidates for an x with a high PTR were investigated using the non-dominated sorting genetic algorithm (NSGA-II), and three candidates were selected via D-optimal programming. Figure 4 shows the

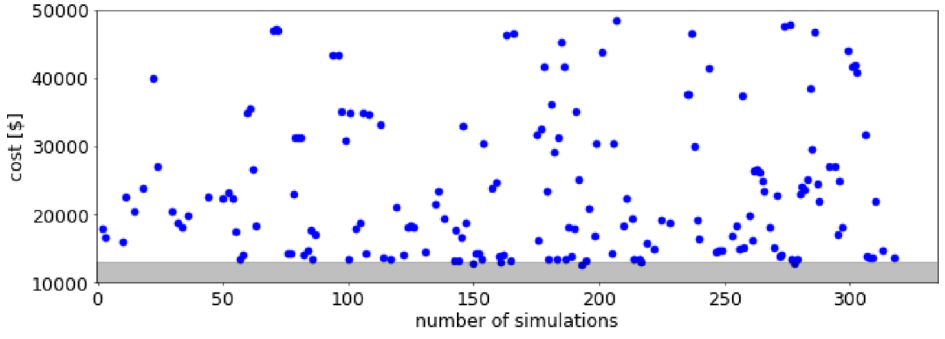
https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934



(a) CO flow versus cost



(b) CO flow per simulation



(c) Cost per simulation.

Figure 4. Results of running simulations with synthesis and process conditions selected via GPR-BO. (a−c) Mean CO flow versus cost, CO flow per simulation, and cost per simulation, respectively. The gray regions, gray points, red points, yellow star, and blue points represent target ranges, randomly selected samples, Pareto optimal set of gray points, target point, and samples selected via GPR-BO, respectively.

results of 100 iterations of this process. The gray, red, and blue points represent the random results, Pareto optimal solutions for the gray samples, and GPR-BO results, respectively. Although no samples reached the target point, as shown in Figure 4a, we could propose candidates for x that exceeded the Pareto optimal solution. As shown in Figures 4b,c, y values closer to the target region were obtained with increasing number of simulations, because the GPR model appropriately evaluated the extrapolation regions of x and trained the relationship between x and y near the target region. The two y variables had a trade-off relationship, and the target y values were realistically unattainable, making it more difficult for the cost to converge than the CO flow rate. In an actual operation using the proposed method, setting realistic target y values would facilitate convergence.

Table 2 shows five examples of the 27 candidates that exceeded the Pareto optimal solution and had relatively high CO_2 conversion rates. Table 3 shows the results for the metal oxides that were found to have high CO_2 and H_2 conversion rates. Oxides of metals containing 0.3−0.5 mole fractions of Cu and Ga, the remaining being alkali metals or alkaline earth metals, tended to yield high CO_2 and H_2 conversion rates. By

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

Table 2. Five Candidates in GPR-BO That Exceeded the Pareto Optimal Set of Randomly Selected Samples

metal ratios in metal oxides	$Mn_{0.01}Fe_{0.02}La_{0.97}$	$Y_{0.31}Zn_{0.42}La_{0.27}$	$Mn_{0.23}Fe_{0.73}W_{0.04}$	$Ca_{0.03}Fe_{0.87}Bi_{0.10}$	$Mg_{0.43}Al_{0.44}K_{0.13}$
CO flow [kmol/h]	6.45	7.24	7.2	7.38	7.66
cost [$]	1.66×10^4	2.46×10^4	1.92×10^4	2.34×10^4	2.40×10^4
CO_2 conv. [-]	62.2	90.8	56	71.6	60.9
H_2 conv. [-]	1.76	23.1	29.7	44	15
reactor length [m]	1	1	1.73	6	6
reactor diameter [m]	0.1	0.1	0.0463	0.0492	0.0492
number of tubes [-]	1.92×10^3	1.97×10^3	2.98×10^3	1.37×10^3	2.13×10^3
temp. (CO_2/H_2) [K]	1.00×10^3/1.00×10^3	1.00×10^3/500	1.00×10^3/1.00×10^3	500/500	500/1.00×10^3
pressure (CO_2/H_2) [bar]	8.00/5.00	8.00/5.00	8.00/5.00	8.00/5.00	8.00/5.00

Table 3. Candidates Whose Predicted CO_2 and H_2 Conversion Rates Were Both High

metal ratios in metal oxides	CO_2 conv. [-]	H_2 conv. [-]
$K_{0.18}Cu_{0.51}Ga_{0.31}$	86.92	77.81
$Mg_{0.17}Cu_{0.45}Ga_{0.38}$	86.93	76.20
$Ca_{0.14}Cu_{0.43}Ga_{0.43}$	83.43	74.36
$Cu_{0.47}Ga_{0.34}Sr_{0.19}$	84.97	69.22
$Mn_{0.12}Cu_{0.46}Ga_{0.42}$	79.26	57.31

simultaneously optimizing the synthesis and process conditions, the metal oxides that did not achieve the highest CO_2 and H_2 conversion rates were selected because even metal oxides with somewhat lower predicted CO_2 and H_2 conversion rates could be optimal solutions if optimized simultaneously with the process conditions. The candidates shown in Table 2 were also selected owing to a difference in the temperature dependence. The difference in the CO_2 conversion rate between 500 and 650 °C for the candidates selected via GPR-BO was 15−20%, while that for the candidates with higher predicted CO_2 and H_2 conversion rates was 5−10%. The GPR-BO candidates could convert the raw material CO_2 without waste in the reaction kinetic equation. As indicated by the process conditions of the proposed candidates in Table 2, the process conditions often had minimum or maximum values, owing to the characteristics of BO in the extrapolated regions of x.

In GMR-BO, the GMR model was constructed using x as the experimental condition, metal oxide molar ratio, and process condition and y as the CO flow rate and cost. The results of the cross-validation evaluation showed that both the CO-flow and cost prediction models had $r^2 = 0.001$. Although highly predictive models could not be constructed, subsequent data analysis was performed. After the GMR model had been constructed, the target y values (CO flow = 8.3 kmol h^{-1}; cost = 13,500 $) were input, and a mixture of multivariate normal distributions of x was output. Thirty candidates of x, with high-probability-density constraints of the mixture of x were investigated using GA, and subsequently three candidates were selected using D-optimal programming. For the three candidates, a process simulation was performed to obtain the y values, and the GMR model was then updated. Figure 5 shows the results of 100 iterations of this process. The gray, red, and blue points represent the results of random selection, Pareto optimal solutions of the gray points, and results of GMR-BO, respectively. Although no samples reached the target point as shown in Figure 5a, we could propose candidates that exceeded the Pareto optimal solution; the reason for a smaller number of candidates exceeding the Pareto optimal solution than GPR-BO was that a GMR model with a high predictive ability could not be constructed. However, as shown in Figures 5b,c, y values closer to the target region were obtained with increasing number of simulations. Compared to Figures 4b,c, the y values became more consistent and convergent with each simulation, especially for the CO flow rate.

Table 4 shows that the results for the three candidates exceeded the Pareto optimal solution for the GMR-BO. In all three candidates, Fe, W, and Zr were chosen as the metals, which would have affected the convergence of the y value in Figure 5. From the candidates of the process conditions proposed in GMR-BO, no minimum or maximum values were obtained, whereas the candidates in GPR-BO reached minimum or maximum values for the process conditions. Realistic x values were designed instead of the maximum or minimum values because the GMR model could account for the correlation between the x variables of the dataset.

Fixed metals were proposed in GMR-BO because the target y values were too far away from the training data; therefore, the target y values were changed and set as multiple points. This method is called GMR-BO-multiple. Three target y points (CO flow [8.3 kmol h^{-1}]; cost [$]) = (8, 25,000), (7, 19,000), and (5, 15,000) were set near the Pareto optimal solution. The GA was run 10 times for each target y point, and the samples with the highest probability density value were selected, resulting in a total of three candidates. For the three candidates, a process simulation was performed to obtain the y values, and the GMR model was then updated. Figure 6 shows the results of 100 iterations of this process. Although Figure 6a indicates that no samples reached the target y point, we could propose candidates exceeding the Pareto optimal solution, with results similar to those shown in Figure 5. Additionally, the metals were fixed, as was the case with GMR-BO. Thus, the proposed metals were not fixed because of the effect of the target y values.

Finally, we studied a method combining GPR-BO and GMR-BO, called GPR-GMR-BO. In GPR-GMR-BO, two candidates of x were selected via the GPR model and one via the GMR model. The process simulation was run with a total of three candidates of x to obtain the y values. The GPR and GMR models were then updated based on the results of 100 iterations of this process, which are shown in Figure 7. The gray, red, and blue points represent the randomly selected results, Pareto optimal solutions, and GPR-GMR-BO results, respectively. Although no samples reached the target y point, as shown in Figure 7a, we could propose candidates that exceeded the Pareto optimal solution. All samples that exceeded the Pareto optimal solution were those proposed via GPR-BO, while those proposed via GMR-BO had a combination of metals fixed to Gd, Zn, and Zr. Although the metal combination changed from Fe, W, and Zr to Gd, Zn, and

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

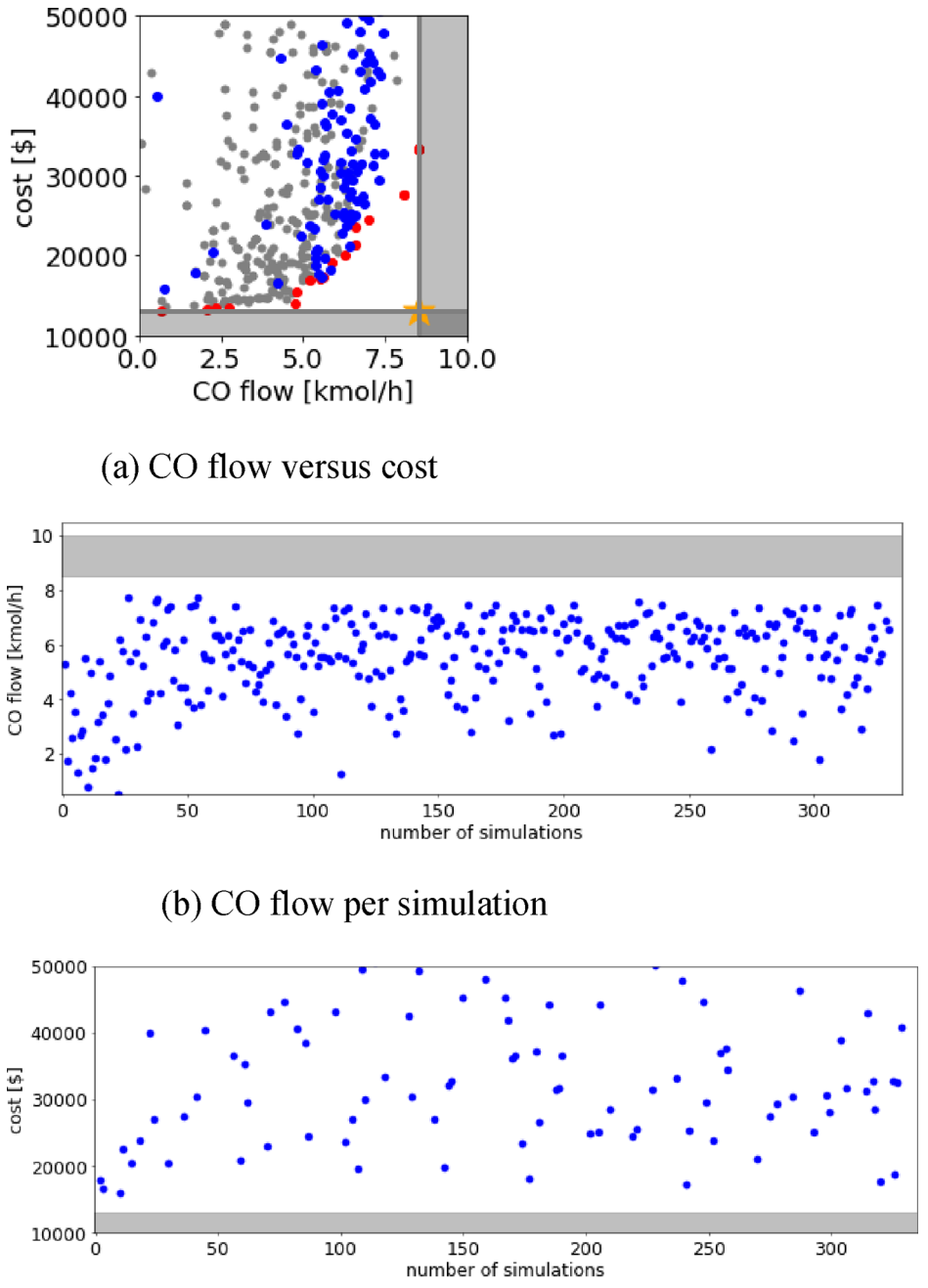


Figure 5. Results of running simulations with synthesis and process conditions selected via GMR-BO. (a−c) Mean CO flow versus cost, CO flow per simulation, and cost per simulation, respectively. The gray regions, gray points, red points, yellow star, and blue points represent target ranges, randomly selected samples, Pareto optimal set of gray points, target point, and samples selected via GMR-BO, respectively.

Zr, there were no effects on the optimization efficiencies of GPR-BO and GMR-BO. The metal combination was fixed to Gd, Zn, and Zr in GMR-BO because the characteristics of the metal oxides were not considered in the features of x, which is an issue to be addressed in the future.

Table 5 shows five of the ten candidates that exceeded the Pareto optimal solution in GPR-GMR-BO. Zr, which was not selected via GPR-BO, was one of them, which could have been owing to the influence of Gd, Zn, and Zr, which were selected via GMR-BO. For the process conditions, the results with the maximum or minimum values were fewer compared to the GPR-BO results. As with the metal oxides, the process conditions were also believed to have been affected by GMR-BO.

The initial 30 samples for the four proposed methods were changed five times for comparison. Figure 8 shows box plots of the number of candidates that exceeded the Pareto optimal solution. GPR-BO was the most stable among the four proposed methods in terms of affording a large number of promising candidates, because it properly evaluated the extrapolation region of x by considering the variance of the predicted y values, even in the initial dataset of 30 samples. In

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

Table 4. Three Candidates in GMR-BO That Exceeded the Pareto Optimal Set of Randomly Selected Samples

metal ratios in metal oxides	$Fe_{0.33}W_{0.33}Zr_{0.34}$	$Fe_{0.35}W_{0.33}Zr_{0.33}$	$Fe_{0.35}W_{0.33}Zr_{0.33}$
CO flow [kmol/h]	5.83	5.57	6.44
cost [$]	1.82×10^4	1.72×10^4	2.11×10^4
CO_2 conv. [-]	56.35	56.87	57.05
H_2 conv. [-]	33.64	1.76	23.07
reactor length [m]	2.26	3.04	4.29
reactor diameter [m]	0.0419	0.0234	0.0325
number of tubes [-]	694	881	941
temp. (CO_2/H_2) [K]	885/762	890/929	941/871
pressure (CO_2/H_2) [bar]	6.26/7.39	7.11/7.08	6.37/6.32

contrast, GMR-BO proposed more than 20 candidates only once; however, it did not propose a stable number of candidates, indicating that it was dependent on the initial dataset.

CONCLUSIONS

In this study, we developed methods to simultaneously optimize the components of metal oxides, their synthesis conditions, and RWGS-CL process conditions to achieve target values of CO flow and cost in the RWGS-CL process. The synthesis and process conditions were designed simultaneously using an adaptive design of experiments employing various models for predicting the CO_2 and H_2 conversion rates from compositions of metal oxides and their synthesis conditions and the CO flow rate and cost from process conditions including the metal oxides used. GPR and GMR were considered for model construction; GPR was used as the regression analysis method in GPR-BO, and NSGA-II was used to determine an x with a high probability of achieving the target y values. Simulations were run using three candidates of x with high probabilities, and if the targets were met, the design was terminated; otherwise, the GPR model was updated with the simulation results. In GMR-BO, the GMR model was used as the regression analysis method: the target y values were input into the GMR model, and a mixture of multivariate normal distributions of x was output. GA was used to determine constrained x values with high probability densities. Simulations were run using the three candidates of x with the highest probability densities, and if the targets were met, the simulation was terminated; otherwise, the GMR model was updated with its results.

In addition to GPR-BO and GMR-BO, GMR-BO-multiple and GPR-GMR-BO were also discussed. In GMR-BO-multiple, three target y values were selected near the Pareto front optimum of the random selection. In GPR-GMR-BO, two candidates for x were selected using GPR-BO and one using GMR-BO. All four proposed methods outperformed the results of the random search. Furthermore, GPR-BO achieved the highest performance and proposed 27 promising candidates for x. With an initial dataset of only 30 samples, GPR-BO properly evaluated the extrapolation regions of x. The selected metals did not include Cu and Ga, which tended to have high predicted CO_2 and H_2 conversion rates, but Fe and La, which had slightly lower predicted CO_2 and H_2 conversion rates. These results indicate that a combination of metal oxides with lower predicted CO_2 and H_2 conversion rates and optimized process conditions was important for the simultaneous optimization of both materials and processes, which was achieved by integrating materials and process informatics via the proposed method.

Future studies may include the development of features for accurately representing metal oxides and using features such as x for improving the predictive ability of the model, predicting new metal oxides, and achieving a high-performing design of synthesis and process conditions. The proposed method may efficiently determine superior synthesis and process conditions for both material and process design. In future, this study may be validated via experiments and manufacturing methods employing the proposed candidates as the synthesis and process conditions.

METHODS

Process Simulation. For the H_2O synthesis process, the raw material H_2 was assumed to have been produced via the electrolysis of water and to contain no impurities; H_2 was introduced under high temperature and pressure through a compressor and heat exchanger and reacted with the metal oxides in a piston-flow reactor (PFR) containing catalysts to produce H_2O. After the reaction, flash distillation was performed under atmospheric pressure to separate the product H_2O from unreacted H_2, which was purged and recycled as a raw material.

For the CO synthesis process, the flow of the raw material, CO_2, which was assumed to have been separated from the exhaust gas from factories and not contain impurities, was set to 10 kmol h^{-1}. CO_2 reacts with metal oxides under high temperature and pressure to produce CO, similar to H_2. The produced CO was shipped as a product after lowering the temperature and pressure. The similarity of CO_2 and CO components increases the cost of separation; therefore, separation is not conducted, and the reactants of CO and CO_2 are separated in the next process.

In the PFR, simulations were performed using the reaction rate equation, the Arrhenius equation, as follows:

$$k = Ae^{-E/RT} \tag{3}$$

where A represents the frequency factor, E [J mol^{-1}] the activation energy of the reaction, R [J mol^{-1} K^{-1}], the gas constant, is set as 8.314, and T [K] is the temperature. By taking the logarithm of both sides of eq 3, the following equation is obtained:

$$\ln k = \ln A - \frac{E}{RT} \tag{4}$$

As indicated by eq 4, the relationship between ln k calculated from the CO_2 and H_2 conversion rates and the reciprocal of the reaction temperature $1/T$ is linear, E is obtained from the slope $-E/R$, and the frequency factor A is obtained from the intercept ln A.

Robustness is important when results are obtained by inputting random data into the process simulator; therefore, we used SimCentral and AVEVA process simulation software.[25] With AVEVA process simulation, we received an error flag and discarded the results when the solution was divergent and could not be obtained. In an iterative process, a sequential simulator requires a long time to converge; however, AVEVA process simulation affords a convergent solution rapidly. In addition, x can be set as needed in the process designs because the specifications can be freely changed. Furthermore, AVEVA process simulation can be performed

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

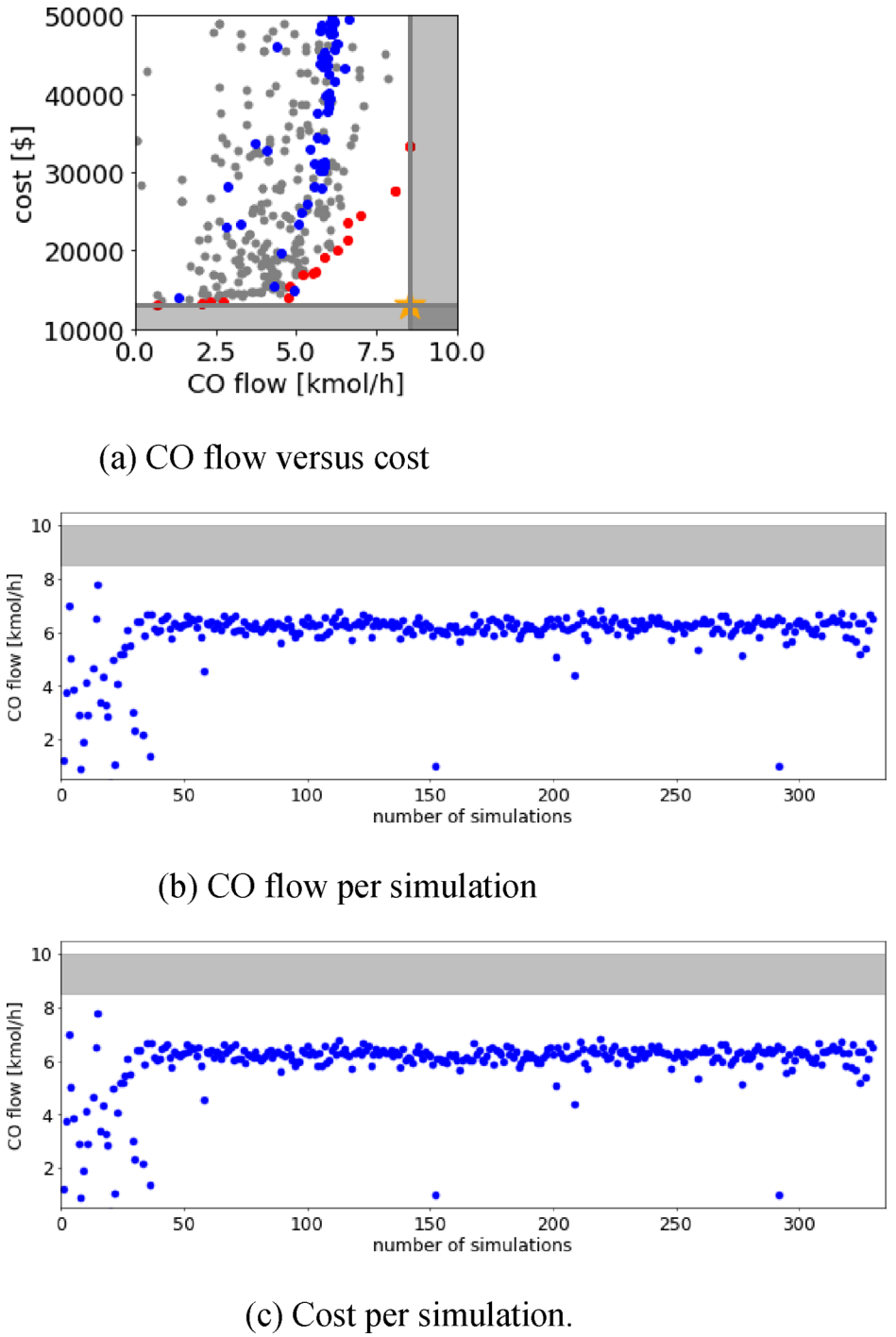


Figure 6. Results of running simulations with synthesis and process conditions selected via GMR-BO-multiple. (a−c) Mean CO flow versus cost, CO flow per simulation, and cost per simulation, respectively. The gray regions, gray points, red points, yellow star, and blue points represent target ranges, randomly selected samples, Pareto optimal set of gray points, target point, and samples selected via GMR-BO-multiple, respectively.

using JavaScript and automated in association with Python programs, meaning that the proposed method can be easily combined with AVEVA process simulation.

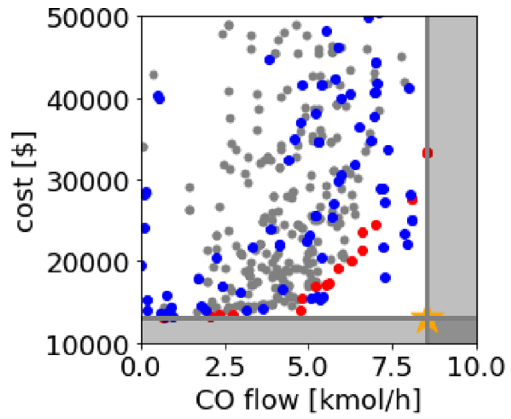
Gaussian Process Regression. GPR, a linear regression analysis method, can be extended to the construction of nonlinear regression models using kernel functions. GPR models can predict y values and calculate their standard deviation variance, allowing a discussion on the reliability of the predicted y values.

Assuming the value of y in the ith sample to be $y^{(i)}$ and the vector of x in the ith sample to be $\mathbf{x}^{(i)} \cap R^m$ (where m is the number of descriptors or features), the relationship between $y^{(i)}$ and $\mathbf{x}^{(i)}$ can be given as follows:

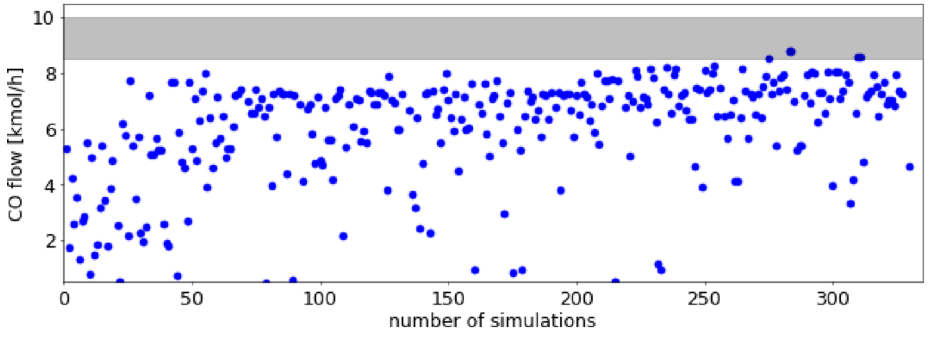
$$y^{(i)} = \mathbf{x}^{(i)}\mathbf{b} + e^{(i)} \tag{5}$$

where $\mathbf{b} \in R^m$ is a vector of regression coefficients and $e^{(i)}$ represents the noise in y. Assuming the mean of each regression coefficient to be zero, the regression coefficients

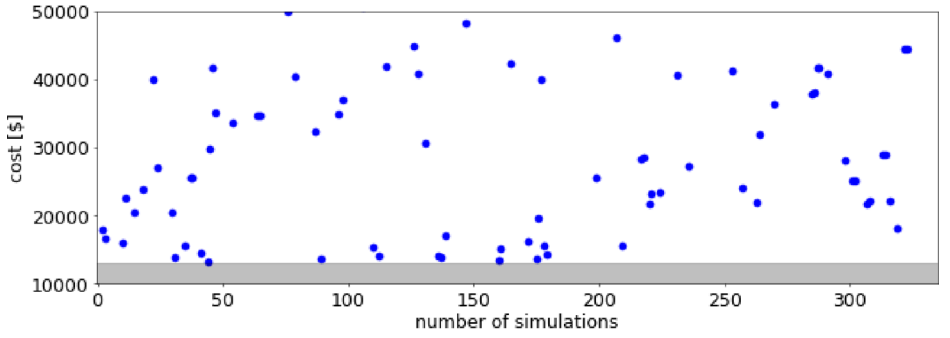
https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934



(a) CO flow versus cost



(b) CO flow per simulation



(c) Cost per simulation

Figure 7. Results of running simulations with synthesis and process conditions selected via GPR-GMR-BO. (a−c) Mean CO flow versus cost, CO flow per simulation, and cost per simulation, respectively. The gray regions, gray points, red points, yellow star, and blue points represent target ranges, randomly selected samples, Pareto optimal set of gray points, target point, and samples selected via GPR-GMR-BO, respectively.

to be independent, and the mean of $y^{(i)}$ to be zero, the covariance ${\sigma_{i,j}}^2$ between $y^{(i)}$ and $y^{(j)}$ can be given as follows:

$$ {\sigma_{i,j}}^2 = \sigma_b^2 \mathbf{x}^{(i)} \mathbf{x}^{(j)\mathrm{T}} + \delta_{i,j} {\sigma_e}^2 \tag{6} $$

where ${\sigma_b}^2$, ${\sigma_e}^2$, and $\delta_{i,j}$ represent the variance of the regression coefficients, variance of the noise in y, and delta of the Kronecker, respectively.

When x is transformed by the non-linear function g, $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ become $g(\mathbf{x}^{(i)})$ and $g(\mathbf{x}^{(j)})$, respectively. Then, eq 2 can be transformed as follows:

$$ {\sigma_{i,j}}^2 = \sigma_b^2 g(\mathbf{x}^{(i)}) g(\mathbf{x}^{(j)})^{\mathrm{T}} + \delta_{i,j} {\sigma_e}^2 \tag{7} $$

GPR uses the kernel function, K, as follows:

$$ K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sigma_b^2 g(\mathbf{x}^{(i)}) g(\mathbf{x}^{(j)})^{\mathrm{T}} + \delta_{i,j} {\sigma_e}^2 \tag{8} $$

This study uses the following kernel function:

$$ K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \theta_0 \exp\left\{ -\frac{\theta_1}{2} \left\| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \right\|^2 \right\} + \theta_2 \tag{9} $$

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

Table 5. Five Candidates in GPR-BO That Exceeded the Pareto Optimal Set of Randomly Selected Samples

metal ratios in metal oxides	$Zr_{0.01}Mo_{0.36}Zn_{0.63}$	$K_{0.01}Fe_{0.01}Zr_{0.98}$	$Co_{0.26}Zr_{0.62}Nb_{0.12}$	$Si_{0.01}Ca_{0.01}Fe_{0.98}$	$Mg_{0.01}Si_{0.02}Fe_{0.97}$
CO flow [kmol/h]	5.08	5.34	7.21	7.86	8.06
cost [$]	1.56×10^4	1.53×10^4	2.18×10^4	2.34×10^4	2.50×10^4
CO_2 conv. [-]	48.12	62.71	67.76	63.17	66.09
H_2 conv. [-]	21.2	5.83	9.6	35.52	19.17
reactor length [m]	1.00	1.00	1.00	2.79	5.13
reactor diameter [m]	0.100	0.0966	0.100	0.0898	0.0998
number of tubes [-]	3000	3000	3000	1035	565
temp. (CO_2/H_2) [K]	1.00×10^3/1.00×10^3	1.00×10^3/977	1.00×10^3/1.00×10^3	991/1.00×10^3	1.00×10^3/1.00×10^3
pressure (CO_2/H_2) [bar]	5.00/8.00	5.00/5.92	8.00/5.00	8.00/8.00	5.00/8.00

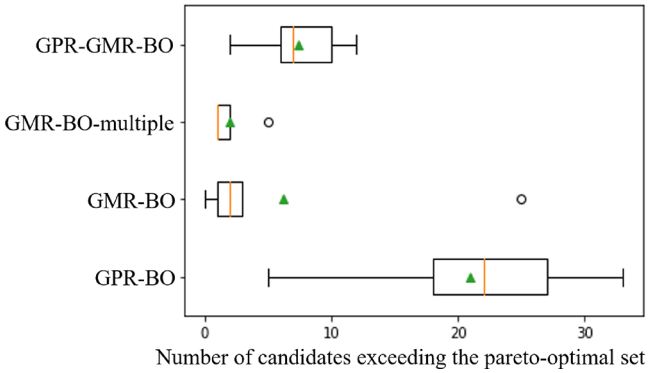


Figure 8. Box plots of the number of candidates that exceeded the Pareto optimal set of randomly selected samples. The green triangles represent averages.

where θ_0, θ_1, and θ_2 are the hyperparameters optimized using the maximum likelihood estimation method. Equation 7 can then be expressed as follows:

$$\sigma_{i,j}{}^2 = \theta_0 \exp\left\{-\frac{\theta_1}{2}\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|^2\right\} + \theta_2 \tag{10}$$

For an n number of samples in the training data, the y value of the $(n + 1)$th sample can be predicted. The posterior probability of the $(n + 1)$th value of y given n training samples can be calculated using the multiplicative theorem of probability. When the mean of $y^{(n+1)}$ is μ_{n+1} and its variance is $\sigma_{n+1}{}^2$, they can be calculated as follows:

$$\mu_{n+1} = \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{y} \tag{11}$$

$$\sigma_{n+1}{}^2 = K(\mathbf{x}^{(n+1)}, \mathbf{x}^{(n+1)}) - \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k} \tag{12}$$

where, $\mathbf{y} \in R^n$, $\mathbf{k} \in R^n$, and $\mathbf{K} \in R^{n\times n}$ represent the vector of y in the training data, vector of a kernel function between all samples of the training data and the $(n + 1)$th sample, and gram matrix of a kernel function between all the samples of the training data, respectively. The GPR calculation used the GaussianProcessRegressor[26] from the scikit-learn library, and the GPR models were constructed after autoscaling for both x and y.

Gaussian Mixture Regression. The GMR is a regression analysis approach that relies on the use of a Gaussian mixture model (GMM) to construct a model that expresses a dataset by superimposing multiple Gaussian distributions.

For a sample $\mathbf{x} \in \mathrm{R}^m$ (where m is the number of variables in x), the probability distribution function $p(\mathbf{x})$ of the GMM is given by

$$p(\mathbf{x}) = \sum_{i=1}^{n} \pi_i N(\mathbf{x}|\mu_i, \Sigma_i) \tag{13}$$

where n is the number of Gaussians and μ_i and Σ_i are the mean vector and variance−covariance matrix of the ith Gaussian, respectively. π_i is the weight of the ith Gaussian that satisfies the following conditions:

$$0 \le \pi_i \le 1, \quad \sum_{i=1}^{n} \pi_i = 1 \tag{14}$$

μ_i, Σ_i, and π_i can be determined using the EM algorithm, based on the log-likelihood function, as follows:

$$\log\{p(\mathbf{x}|\mu, \Sigma, \pi)\} = \sum_{k=1}^{N} \log\left\{\sum_{i=1}^{n} \pi_i N(\mathbf{x}_k|\mu_i, \Sigma_i)\right\} \tag{15}$$

where N is the number of samples.

In GMR, the joint probability distribution of x and y is determined by calculating the GMM using a dataset that combines sample $\mathbf{y} \in \mathrm{R}^p$ of y (p is the number of y variables) and sample $\mathbf{x}$ of x. Based on the probability-multiplication and Bayes' theorems, the y and x values can be predicted by calculating the posterior probability distributions $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x}|\mathbf{y})$, respectively, for given $\mathbf{x}$ and y, respectively. Essentially, a direct inverse analysis of the model is possible.

Explicitly separating x and y into eq 13 generates the following expression for the joint probability distribution of the variables:

$$p(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \pi_i N\left([\mathbf{x}\ \mathbf{y}]\ \middle|\ [\mu_{x,i}\ \ \mu_{y,i}], \begin{bmatrix}\Sigma_{xx,i} & \Sigma_{yx,i}\\ \Sigma_{xy,i} & \Sigma_{yy,i}\end{bmatrix}\right) \tag{16}$$

where $\mu_{x,i}$ and $\mu_{y,i}$ are the mean vectors of x and y of the ith Gaussian, respectively; $\Sigma_{xx,i}$ and $\Sigma_{yy,i}$ are the variance−covariance matrices of x and y for the ith Gaussian, respectively; and $\Sigma_{xy,i}$ and $\Sigma_{yx,i}$ are the covariance matrices of x and y for the ith normal distribution, respectively.

The estimation of x from y corresponds to the determination of the posterior probability distribution $p(\mathbf{x}|\mathbf{y})$ of x for a given $\mathbf{y}$, which can be transformed using the probability-multiplication and Bayes' theorems, as follows:

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

$$p(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^{n} p(\mathbf{x}|\mathbf{y}, \mu_{y,i}, \Sigma_{yy,i}) p(\mu_{y,i}, \Sigma_{yy,i}|\mathbf{x})$$
$$= \sum_{i=1}^{n} p(\mathbf{x}|\mathbf{y}, \mu_{y,i}, \Sigma_{yy,i}) \frac{p(\mathbf{y}|\mu_{y,i}, \Sigma_{yy,i}) p(\mu_{y,i}, \Sigma_{yy,i})}{\sum_{j=1}^{n} p(\mathbf{y}|\mu_{y,j}, \Sigma_{yy,j}) p(\mu_{y,j}, \Sigma_{yy,j})}$$
$$= \sum_{i=1}^{n} p(\mathbf{x}|\mathbf{y}, \mu_{y,i}, \Sigma_{yy,i}) \frac{\pi_i p(\mathbf{y}|\mu_{y,i}, \Sigma_{yy,i})}{\sum_{j=1}^{n} \pi_j p(\mathbf{y}|\mu_{y,j}, \Sigma_{yy,j})}$$
$$= \sum_{i=1}^{n} w_{y,i} p(\mathbf{x}|\mathbf{y}, \mu_{y,i}, \Sigma_{yy,i}) \tag{17}$$

$$w_{y,i} = \frac{\pi_i p(\mathbf{y}|\mu_{y,i}, \Sigma_{yy,i})}{\sum_{j=1}^{n} \pi_j p(\mathbf{y}|\mu_{y,j}, \Sigma_{yy,j})} \tag{18}$$

where $p(\mathbf{x}|\mathbf{y}, \mu_{y,i}, \Sigma_{yy,i})$ and $w_{y,i}$ represent the multivariate Gaussian distribution of the estimated x value for the ith Gaussian distribution and weight of the multivariate Gaussian distribution, respectively. For $p(\mathbf{x}|\mathbf{y}, \mu_{y,i}, \Sigma_{yy,i})$, the mean vector $\mathbf{m}_i(\mathbf{y})$ and variance−covariance matrix $\mathbf{S}_i(\mathbf{y})$ are given as

$$\mathbf{m}_i(\mathbf{y}) = \mu_{x,i} + (\mathbf{y} - \mu_{y,i})\Sigma_{yy,i}^{-1}\Sigma_{yx,i} \tag{19}$$

$$\mathbf{S}_i(\mathbf{y}) = \Sigma_{xx,i} - \Sigma_{xy,i}\Sigma_{yy,i}^{-1}\Sigma_{yx,i} \tag{20}$$

Here, y can be estimated from x by swapping x for y.

The GMM calculation employed the GaussianMixture[27] from the scikit-learn library, whereas the GMR calculation relied on DCEKit.[28]

Genetic Algorithm (GA) and Non-dominated Sorting GA (NSGA-II). The genetic algorithm (GA)[29] mimics biological evolution and searches for the optimal solution from a large solution space. The GA performs optimization in the following flow:

1. Generation of an initial population containing N individuals
2. Calculation of the fitness of each individual
3. Selecting parent individuals according to their fitness
4. Conducting genetic manipulations, such as crossover and mutation
5. Running steps 3 and 4 until N new individuals are obtained to generate a child population
6. Running steps 2−5 a specified number of times (number of generations)

Because multiple objective variables can be the fitness in the GA, the non-dominated sorting genetic algorithm (NSGA-II),[30] which is an extension of GA to multi-objective optimization problems, was used in this study. The NSGA-II performs optimization in the following flow:

1. Generation of an initial population containing N individuals
2. Calculation of the fitness of each individual
3. Ranking each individual by non-dominant and congestion sorting
4. Selection of parental individuals via congestion tournament selection
5. Conducting genetic manipulations, such as crossover and mutation
6. Running steps 4 and 5 until N new individuals are obtained to generate a child population
7. Running steps 2−6 a specified number of times (number of generations)

Herein, the GA and NSGA-II calculations employed the DEAP library.[31]

Gaussian Process Regression-Bayesian Optimization (GPR-BO). Using the dataset of the experimental and simulation results, a GPR model was constructed between x, which represented the synthesis conditions and metal oxide characteristics, CO_2 and H_2 conversion rates, and process conditions, and y, which represented the target process requirements. By inputting the x values for unknown simulation results to the GPR model, the predicted y values and their variances were output. Based on the prediction results, the next candidate for the simulation, x, was selected.

In Bayesian optimization (BO), an acquisition function was calculated after the GPR models had predicted the y values, whose variances were also calculated. Because y has target ranges in this study, we used probability in the target range (PTR)[16] expressed in terms of probability to unify them into an acquisition function. PTR is the probability of the predicted y values falling within a target range of y. Considering a normal distribution, in which the predicted y values and their variances from the GPR models correspond to the mean and variance, respectively, we integrated the normal distribution from the lower limit of y, y_{LOWER}, to the upper limit, y_{UPPER}, and calculated PTR. $\mathrm{PTR}(\mathbf{x}^{(n+1)})$, which is the value of PTR for a new sample $\mathbf{x}^{(n+1)}$, is given as follows:

$$\mathrm{PTR}(\mathbf{x}^{(n+1)}) = \int_{y_{LOWER}}^{y_{UPPER}} \frac{1}{\sqrt{2\pi\sigma_{n+1}^2}} \exp\left\{-\frac{1}{2\sigma_{n+1}^2}(x - \mu_{n+1})^2\right\} dx \tag{21}$$

We first optimized x using NSGA-II and generated 30 samples of x to yield a high probability of eq 19 and subsequently performed D-optimal programming[32] to determine the next three samples to be simulated from the 30. The three simulations were performed using the selected candidates of x, and the GPR model was then updated using the three samples of x and y—the simulation results. This was repeated 100 times, and the results were evaluated. The proposed method is called "Gaussian process regression-Bayesian optimization" (GPR-BO).

Gaussian Mixture Regression-Bayesian Optimization (GMR-BO). Using the dataset of experimental and simulation results, a GMR model was constructed between x, representing the experimental conditions and metal oxide characteristics, CO_2 and H_2 conversion rates, and process conditions, and y, representing the target process requirements. The probability distributions of x were estimated by inputting the target y values into the GMR model. Considering the restrictions of x, such as molar ratios, GA was used to determine the constrained x values for which the probability density was high.[18] Thirty iterations of optimization with GA were conducted to obtain 30 candidates of x, and D-optimal programming was then performed to determine the next three samples to be simulated from the 30. The three simulations

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

were performed using the selected candidates of x, and the GMR model was then updated using the three samples of x and y, which represented the simulation results. This was repeated 100 times, and the results were evaluated. This method is called "Gaussian mixture regression-Bayesian optimization" (GMR-BO).

■ ASSOCIATED CONTENT

Data Availability Statement

Research data have not been shared because the experimental data are considered proprietary by the organization, and the algorithms of the proposed method have been described in the manuscript.

■ AUTHOR INFORMATION

Corresponding Author

Hiromasa Kaneko − *Department of Applied Chemistry, School of Science and Technology, Meiji University, Kawasaki-shi, Kanagawa-ken 214-8571, Japan;* orcid.org/0000-0001-8367-6476; Email: hkaneko@meiji.ac.jp

Author

Ryo Iwama − *Department of Applied Chemistry, School of Science and Technology, Meiji University, Kawasaki-shi, Kanagawa-ken 214-8571, Japan;* orcid.org/0000-0002-7718-4544

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c06008

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support of Sekisui Chemical Co. Ltd. This study was supported by a Grant-in-Aid for Scientific Research (KAKENHI [grant number 19K15352] from the Japan Society for the Promotion of Science.

■ REFERENCES

(1) Chen, X.; Chen, Y.; Song, C.; Ji, P.; Wang, N.; Wang, W.; Cui, L. Recent Advances in Supported Metal Catalysts and Oxide Catalysts for the Reverse Water-Gas Shift Reaction. *Front. Chem.* **2020**, *8*, 709.

(2) Daza, Y. A.; Kent, R. A.; Yung, M. M.; Kuhn, J. N. Carbon Dioxide Conversion by Reverse Water-Gas Shift Chemical Looping on Perovskite-Type Oxides. *Ind. Eng. Chem. Res.* **2014**, *53*, 5828−5837.

(3) Daza, Y. A.; Maiti, D.; Kent, R. A.; Bhethanabotla, V. R.; Kuhn, J. N. Isothermal reverse water gas shift chemical looping on $La_{0.75}Sr_{0.25}Co_{(1-Y)}Fe_YO_3$ perovskite-type oxides. *Catal. Today* **2015**, *258*, 691−698.

(4) Daza, Y. A.; Maiti, D.; Hare, B. J.; Bhethanabotla, V. R.; Kuhn, J. N. More Cu, more problems: Decreased CO_2 conversion ability by Cu-doped $La_{0.75}Sr_{0.25}FeO_3$ perovskite oxides. *Surf. Sci.* **2016**, *648*, 92−99.

(5) Maiti, D.; Hare, B. J.; Daza, Y. A.; Ramos, A. E.; Kuhn, J. N.; Bhethanabotla, V. R. Earth abundant perovskite oxides for low temperature CO_2 conversion. *Energy Environ. Sci.* **2018**, *11*, 648−659.

(6) Ramos, A. E.; Maiti, D.; Daza, Y. A.; Kuhn, J. N.; Bhethanabotla, V. R. Co, Fe, and Mn in La-perovskite oxides for low temperature thermochemical CO_2 conversion. *Catal. Today* **2019**, *338*, 52−59.

(7) Humayun, M.; Ullah, H.; Shu, L.; Ao, X.; Tahir, A. A.; Wang, C.; Luo, W. Plasmon Assisted Highly Efficient Visible Light Catalytic CO_2 Reduction Over the Noble Metal Decorated Sr-Incorporated g-C_3N_4. *Nano-Micro Lett.* **2021**, 1.

(8) Siriwardane, R.; Benincosa, W.; Riley, J.; Tian, H.; Richards, G. Investigation of reactions in a fluidized bed reactor during chemical looping combustion of coal/steam with copper oxide-iron oxide-alumina oxygen carrier. *Appl. Energy* **2016**, *183*, 1550−1564.

(9) Hare, B. J.; Maiti, D.; Daza, Y. A.; Bhethanabotla, V. R.; Kuhn, J. N. Enhanced CO_2 Conversion to CO by Silica-Supported Perovskite Oxides at Low Temperatures. *ACS Catal.* **2018**, *8*, 3021−3029.

(10) Hare, B. J.; Maiti, D.; Ramani, S.; Ramos, A. E.; Bhethanabotla, V. R.; Kuhn, J. N. Thermochemical conversion of carbon dioxide by reverse water-gas shift chemical looping using supported perovskite oxides. *Catal. Today* **2019**, *323*, 225−232.

(11) Hare, B. J.; Maiti, D.; Meier, A. J.; Bhethanabotla, V. R.; Kuhn, J. N. CO_2 Conversion Performance of Perovskite Oxides Designed with Abundant Metals. *Ind. Eng. Chem. Res.* **2019**, *58*, 12551−12560.

(12) Sun, H.; Wang, J.; Zhao, J.; Shen, B.; Shi, J.; Huang, J.; Wu, C. Dual functional catalytic materials of Ni over Ce-modified CaO sorbents for integrated CO_2 capture and conversion. *Appl. Catal., B.* **2019**, *244*, 63−75.

(13) Jo, A.; Kim, Y.; Lim, H. S.; Lee, M.; Kang, D.; Lee, J. W. Controlled template removal from nanocast $La_{0.8}Sr_{0.2}FeO_3$ for enhanced CO_2 conversion by reverse water gas shift chemical looping. *J. CO2 Util.* **2022**, *56*, No. 101845.

(14) Ma, L.; Qiu, Y.; Li, M.; Cui, D.; Zhang, S.; Zeng, D.; Xiao, R. Spinel-Structured Ternary Ferrites as Effective Agents for Chemical Looping CO_2 Splitting. *Ind. Eng. Chem. Res.* **2020**, *59*, 6924−6930.

(15) Iwama, R.; Takizawa, K.; Shinmei, K.; Baba, E.; Yagihashi, N.; Kaneko, H. Design and analysis of metal oxides for CO_2 reduction using machine learning, transfer learning, and Bayesian optimization. *ACS Omega* **2022**, *7*, 10709−10717.

(16) 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.

(17) Rasmussen, C. E.; Nickisch, H. Gaussian Processes for Machine Learning (GPML) Toolbox. *J. Mach. Learn. Res.* **2010**, *11*, 3011−3015.

(18) Kaneko, H. Extended Gaussian Mixture Regression for Forward and Inverse Analysis. *Chemom. Intell. Lab. Syst.* **2021**, *213*, No. 104325.

(19) Shimizu, N.; Kaneko, H. Direct Inverse Analysis Based on Gaussian Mixture Regression for Multiple Objective Variables in Material Design. *Mater. Des.* **2020**, *196*, No. 109168.

(20) Kaneko, H. True Gaussian Mixture Regression and Genetic Algorithm-based Optimization with Constraints for Direct Inverse Analysis. *Sci. Technol. Adv. Mater.: Methods* **2022**, *2*, 14−22.

(21) Wenzel, M.; Rihko-Struckmann, L.; Sundmacher, K. Continuous production of CO from CO_2 by RWGS chemical looping in fixed and fluidized bed reactors. *Chem. Eng. J.* **2018**, *336*, 278−296.

(22) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.

(23) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, No. 011002.

(24) Filzmoser, P.; Liebmann, B.; Varmuza, K. Repeated double cross validation. *J. Chemom.* **2009**, *23*, 160−171.

(25) *AVEVA Process Simulation* (formerly known as SimCentral); AVEVA.

(26) https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html (accessed 15 August 2022)

(27) https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html (accessed 15 August 2022).

(28) *DCEKit (Data Chemical Engineering toolKit)*; https://datachemeng.com/dcekit/ (accessed 15 August 2022).

(29) Whitley, D. A Genetic Algorithm Tutorial. *Stat. Comput.* **1994**, *4*, 65−68.

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934

(30) Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182−197.
(31) *DEAP/deap*; https://github.com/deap/deap (accessed 15 August 2022)
(32) Brandmaier, S.; Sahlin, U.; Tetko, I. V.; Öberg, T. PLS-Optimal: A Stepwise D-Optimal Design Based on Latent Variables. *J. Chem. Inf. Model.* **2012**, *52*, 975−983.

https://doi.org/10.1021/acsomega.2c06008
ACS Omega 2022, 7, 46922−46934