Heterogeneous & Homogeneous & Bio- & Nano-
CHEMCATCHEM
CATALYSIS

Accepted Article

Title: Data Driven Determination of Reaction Conditions in Oxidative Coupling of Methane via Machine Learning

Authors: Junya Ohyama, Shun Nishimura, and Keisuke Takahashi

To be cited as: *ChemCatChem* 10.1002/cctc.201900843

Link to VoR: http://dx.doi.org/10.1002/cctc.201900843

A Journal of
ChemPubSoc Europe

WILEY-VCH

www.chemcatchem.org

FULL PAPER WILEY-VCH

Data Driven Determination of Reaction Conditions in Oxidative Coupling of Methane via Machine Learning

Junya Ohyama,*[a, b] Shun Nishimura,[c] and Keisuke Takahashi*[d,e]

Abstract: The challenge in catalytic reactions lies within its complexity coming from high dimensional experimental factors. In order to solve such complexity, machine learning is implemented to treat experimental conditions in high dimensions. Oxidative coupling of methane, methane to C_2 compounds (ethylene and ethane), is chosen as the prototype reaction where 156 data consisting of various experimental conditions is prepared. Machine learning reveals that the relationship between experimental conditions and C_2 yield is non-linear matter. In particular, extreme tree regression is found to accurately reproduce the experimental data. In addition, machine learning predictions can be a good indicator for designing experiments. Thus, machine learning can be a powerful approach towards understanding and determining experimental conditions in high dimension.

Introduction

Understanding and controlling the catalytic reactions are considered to be challenging matters. Catalytic reactions involve multiple experimental factors and surface chemistries in a complex manner, making it difficult to create a complete model. However, the implementation of data science could potentially navigate and reveal such complex matter within catalysis as machine learning can treat multiple factors in high dimensions.[1, 2] In earlier years, neural network was implemented to treat heterogeneous catalysis where such data science techniques are proposed to be effective tools for simulating catalyst properties and the performance of solid materials.[3-8] Since then, rapid development of machine learning algorithms occurred due to the introduction of random forest and support vector machine, which greatly improve and expand the ways of understanding the data.[9-12] Thus, one can consider that such machine learning techniques could potentially solve the mystery of how reaction conditions determine the catalytic activities in heterogeneous catalysis.

Catalytic reactions depend on various parameters which can be classified into three categories: catalyst structure, reaction condition, and catalyst stability. The catalyst structure includes composition, oxidation state, particle size, shape, crystallinity, and strain; the reaction conditions include reactor shape, temperature, catalyst, reactant concentration, and flow rate; and the catalyst change during reactions includes aggregation/dispersion, redox, and surface reconstruction. Consequently, all of these parameters can be descriptor variables for representing catalytic reactions. In general, it is challenging to manipulate and handle high dimension descriptor variables by the human brain. Therefore, researchers design catalysts by considering few major descriptors such as particle size, d-band center, and reaction temperature.[13-15] For instance, in the case of gold catalysts, the particle size has been elucidated as the main descriptor of their catalytic performance for a variety of reactions including CO oxidation, aldehyde hydrogenation, and the water gas shift reaction.[16] Hence, gold catalysts under the guideline of particle size control can be well designed. Previous studies have also suggested, however, that factors other than particle size such as support effect, cluster size or single atom effect, and crystal structure also greatly contribute to the performance of gold catalysts.[17-19] In other words, the effect of the catalyst structure is a multi-dimensional problem. If such multi-dimensional problem can be treated within machine learning, further sophisticated catalyst design can be made achievable in principle. The same can be said about the effect of catalytic reaction conditions. One usually explores the influence and importance of each reaction parameter on the target products by conducting reactions under various conditions, and then either find the optimal experiment conditions empirically or instinctively when the data dimensions becomes high and complex. If the effects of the reaction conditions can be organized and if accurate models can be constructed using machine learning, then one could then survey the information surrounding the effects of each reaction parameter and select the optimum reaction conditions from predicted data, even if the range is not investigated through experiments.

Here, machine learning is implemented in order to determine the reaction conditions in heterogeneous catalysis. In particular, oxidative coupling of methane (OCM) is chosen as a prototype reaction where the reaction involves conversion of

[a] Dr, J, Ohyama
Faculty of Advanced Science and Technology
Kumamoto University
2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan
E-mail: ohyama@kumamoto-u.ac.jp
[b] Dr, J, Ohyama
Elements Strategy Initiative for Catalysts and Batteries (ESICB)
Kyoto University
Katsura, Kyoto 615-8520, Japan
[c] Dr, S, Nishimura
Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
E-mail: s_nishim@jaist.ac.jp
[d] Dr, K, Takahashi
Center for Materials research by Information Integration (CMI^2)
National Institute for Materials Science (NIMS)
1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan
E-mail: Takahashi.Keisuke@nims.go.jp
[e] Dr, K, Takahashi
Institute for Catalysis
Hokkaido University
N21,W10,Kita-ku, Sapporo, 001-0021, Japan
E-mail: keisuke.takahashi@eng.hokudai.ac.jp

Supporting information for this article is given via a link at the end of the document

methane to C_2 compounds ethylene and ethane. Typically, the OCM is performed on basic catalysts containing alkaline or alkaline earth metals at a very high temperature ranging from 973-1173 K for abstraction of H from CH_4 to form CH_3 radical intermediates of the C_2 compounds.[20-22] To put the OCM towards an industrial application, the catalytic reaction systems should have high C_2 selectivity by suppressing undesirable reactions- in particular, oxidation of CH_3 radicals on catalyst surface as well as in gas phase- while maintaining CH_4 conversion rate.[20] In addition, high durability is required for the catalyst materials to stably produce the C_2 compounds at such high reaction temperatures.[23] In order to develop the OCM catalysts with these functions, hundreds of inorganic materials have been investigated as summarized in the previous review articles.[20, 21] Among the reported catalysts, Mn-Na_2WO_4/SiO_2 exhibits relatively high activity, selectivity, and stability for the OCM.[23-26] The catalytic performance of Mn-Na_2WO_4/SiO_2 has been further improved by meso-ordered structure of SiO_2 support as well as by a membrane reactor with separated feeding of CH_4 and O_2.[27, 28] However, no catalytic reaction systems affording sufficient C_2 yield and selectivity for practical large-scale application have been developed.[20, 23] One can consider that OCM involves complex reaction conditions; therefore, machine learning is implemented to reveal how reaction conditions act in high dimension during the OCM reaction. Mn-Na_2WO_4/SiO_2 catalysts are chosen due to high catalytic performance in OCM reaction.[21] Here, reaction conditions in the OCM reaction using Mn-Na_2WO_4/SiO_2 are investigated in high dimension via machine learning while determination of the reaction conditions is performed.

Results and Discussion

OCM reaction over Mn-Na_2WO_4/SiO_2 is performed under various conditions; more specifically, a 156 data set is prepared at 156 experimental conditions by changing five reaction parameters including reaction temperature, CH_4/O_2 ratio, concentration of CH_4 and O_2 gases (CH_4+O_2 conc.), total flow rate, and catalyst weight. The total yields of ethylene and ethane (C_2 yields) obtained under the various reaction conditions are listed in Table S1. C_2 yield varies with every parameter of the reaction conditions. The C_2 yields are consistent with previously reported yields obtained using Mn-Na_2WO_4 catalysts under similar reaction conditions including the artificial neural network results, where the C_2 yields of 22% reported in this work are consistent with C_2 yields of 23% previously reported using artificial neural network.[23, 29, 30, 31] Given the nature of the random forest regression model, the impact of the descriptors are also further evaluated. During data acquisition, a reaction parameter value is changed until C_2 yield reaches a local maximum or a plateau value as long as the variable parameter is within safety limits of the experiment (e.g. explosion limits); meanwhile, the other parameter values are kept constant. Several data at $CH_4/O_2 < 2$ are obtained under low CH_4+O_2 conc. to avoid explosion. It should also be noted that temperature < 973 K is not investigated since Mn-Na_2WO_4/SiO_2 hardly shows catalytic activity. In addition, the catalyst weight of > ca. 100 mg is not investigated due to concern of significant increase of back pressure by a thick catalyst bed. Each reaction parameter is changed to construct 156 data which is considered to form a parameter coordinate network capturing the reaction feature including the C_2 maximum yield.

A part of the reaction data is picked up to present the complex effect of the reaction conditions on C_2 yield as shown in Fig. 1. Fig. 1 (a) shows C_2 yield dependence on reaction temperature. After the C_2 compounds are generated from 973 K, C_2 yield increases and then decreases as the temperature increases. It should be noted that the peak temperature for C_2 yield is shifted by CH_4/O_2 ratio as shown in Fig. 1(a). The result indicates that the effect of reaction temperature on C_2 yield is influenced by the CH_4/O_2 ratio. C_2 yield also depends on reaction parameters other than reaction temperature as presented in Fig. 1(b)-(e). The figures also indicate that the effect of each parameter is affected by the reaction temperature. The effects of reaction parameters are changed by reaction parameters other than reaction temperature as represented by Fig. 1(f), which shows the effect of catalyst weight impacted by CH_4/O_2 ratio. Thus, the effects of the five parameters on C_2 yield cannot be described simply as they are intertwined. In the present case, the data set is six-dimensional including the five reaction parameters and C_2 yield. Thus, it is a difficult task to organize the effects of the reaction parameters and to specify the optimum reaction conditions using the human brain.

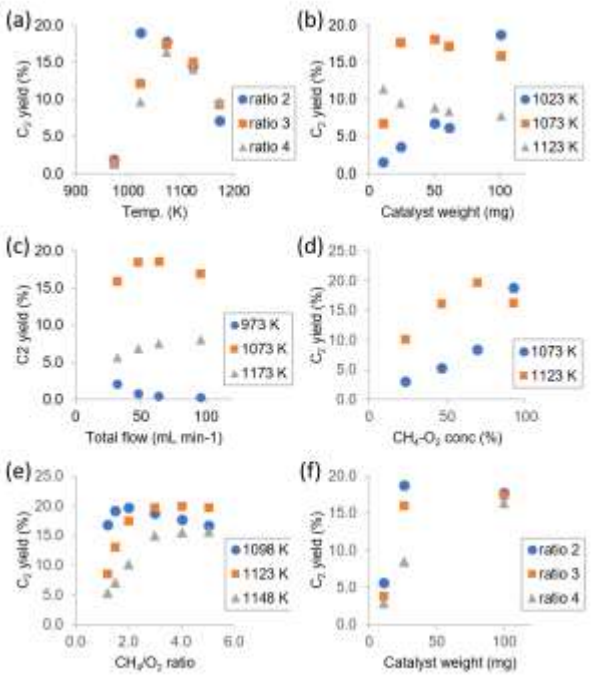


Figure 1. C_2 yield as (a) a function of *reaction temperature* for the OCM reaction over Mn-Na_2WO_4/SiO_2 (catalyst weight: 50 mg, total flow: 32 mL min^{-1}, CH_4-O_2 conc.: 94%, *CH_4/O_2 = 2 (blue), 3 (orange), 4 (gray)*); (b) a function of *catalyst weight* (total flow: 32 mL min^{-1}, CH_4-O_2 conc.: 94%, CH_4/O_2 = 3, *temp.: 1023 (blue), 1073 (orange), 1123 K (gray)*); (c) a function of *total flow* (catalyst

weight: 100 mg, CH_4-O_2 conc.: 94%, CH_4/O_2 = 3, temp.: *973 (blue),1073 (orange), and 1173 K (gray)*); (d) a function of CH_4-O_2 conc. (catalyst weight: 26 mg, total flow: 26 mL min^{-1}, CH_4/O_2 ratio: 2, *temp.: 1073 (blue) and 1123 K (orange)*); (e) a function of *CH_4/O_2 ratio* (catalyst weight: 40 mg, total flow: 30 mL min^{-1}, CH_4-O_2 conc.: 40%, *temp.: 1098 (blue), 1123 (orange), 1148 K (gray)*); (f) a function of *catalyst weight* (total flow: 26 mL min^{-1}, CH_4-O_2 conc.: 92%, *CH_4/O_2 = 2 (blue), 3 (orange), 4 (gray)*, temp: 1073 K).

Machine learning is implemented to order to understand how experimental conditions and C_2 yields interact with each other in high dimensions. If machine learning can capture experimental conditions, one can interpolate or determine the catalytic activity such as in OCM reaction. In order to introduce machine learning, it is necessary to find appropriate machine learning methods for predicting C_2 yield as there are numerous machine learning algorithms. Here, five linear and non-linear machine learning algorithms are applied to the prepared 156 data set where the average scores in cross validation and corresponding information are collected in Table 1 and Fig. S1. The following five descriptor variables are chosen: temperature, catalysts weight, total weight, CH_4-O_2 concentration, and, CH_4/O_2 ratio. Meanwhile, C_2 yield is set to the objective variable. Note that basic experimental conditions are chosen as descriptors for C_2 yield because the Mn-Na_2WO_4/SiO_2 catalyst is known to have high stability for the OCM under harsh conditions; however, catalyst information including surface poisoning, particle size, and oxidation state could be potential descriptors for catalytic reactions involving catalyst degradation. The linear regression algorithms chosen include LSLR and SVR (L), and the non-linear regression algorithms chosen include SVR with SVR (RBF), RFR, and extreme tree regression ETR.[9, 10, 32] . Table 1 indicates that two liner regression algorithms give low accuracy while three non-linear regression algorithms give relatively high accuracy. Thus, one can conclude that the relationship between experimental conditions and C_2 yield is a non-linear matter. Furthermore, the ETR has the highest score among the non-linear machine learning with a score of 86%.

Table 1. Results of cross validation using various machine learning methods. Score is R^2 and SD is standard deviation in cross validation.

Type	Machine	Score	SD	Hyper Parameter
Linear	Least Squares Linear Regression (LSLR)	0.32	0.16	N/A
Linear	Support Vector Regression with linear kernel (SVR (L))	0.27	0.19	N/A
Non-Linear	SVR (RBF)	0.77	0.10	C = 10 gamma = 0.001
Non-Linear	Random Forest Regression (RFR)	0.78	0.12	Number of tree = 1000
Non-Linear	Extreme Tree Regression (ETR)	0.86	0.07	Number of tree = 1000

For further validation of the models created by the three nonlinear methods, prediction of C_2 yield using the three non-linear methods is performed in order to evaluate how each model represents the C_2 prediction. The three-dimensional surface plots in Fig. 2 illustrate the predicted C_2 yield using the non-linear machine learning models as a function of reaction temperature and CH_4/O_2 ratio. The experimental data is superimposed onto the three-dimensional plots as indicated by red circles. The predictions by ETR (Fig. 2(a)-(c)) and RFR (Fig. 2(d)-(f)) present a shift of the peak reaction temperature for C_2 yield where the peak temperature decreases with a decrease of CH_4/O_2 ratio. This peak shift is the same as peak shift of the experimental data as clearly seen in Fig. 1(a). On the other hand, the prediction by SVR (RBF) (Fig. 2(g)-(i)) does not show such peak shift, and exhibits an unusual notched change around 1100 K. Thus, the ETR and RFR models are revealed to accurately reproduce the experimental data.

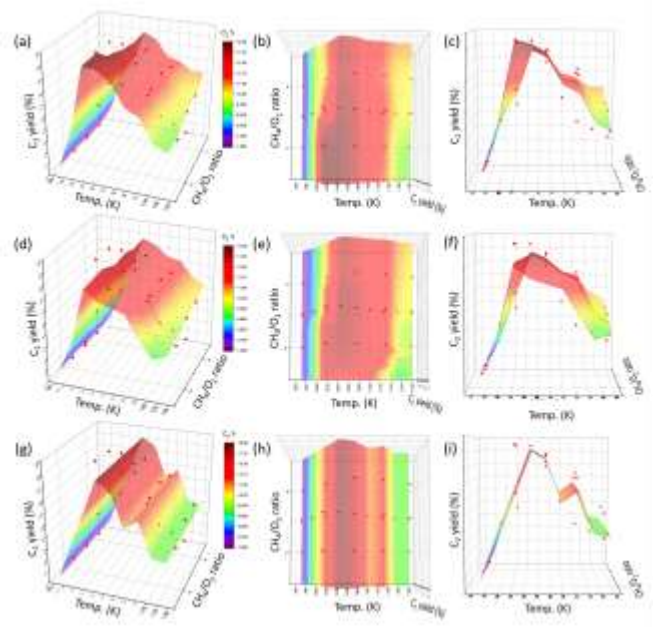


Figure 2. 3D surface plots of the predicted C_2 yield against reaction temperature and CH_4/O_2 ratio: (a-c) ETR; (d-f) RFR; (g-i) non-linear SVR prediction. (b), (e), (h): top view of (a), (d), (g). (c), (f) (i): side view of (a), (d), (g). The experimental data are over-layed on the figures as red circles. The other reaction parameters: catalyst weight 100 mg, CH_4+O_2 conc 90%, total flow 30 mL min^{-1}.

Fig. 3 presents another three-dimensional surface plot of the predicted C_2 yield against catalyst weight and total flow rate. In the ETR prediction (Fig. 3(a)), the surface plot exhibits a top and slopes on both sides of catalyst weight and total flow rate, capturing the feature of the experimental data. However, such feature is not observed in the RFR prediction (Fig. 3(b)), where a simple cliff-like change is seen along with the catalyst weight. In the case of the RFR prediction, a bimodal shape along with the catalyst weight, an unusual variation, is observed. Therefore, the ETR gives the most accurate model for OCM reactions for the

present set of catalysts and models tested, which is consistent with the cross validation score. In the ETR algorithm, the importance of descriptors is not weighted for making decision trees; however, it is weighted in the RFR algorithm. This difference can be viewed as the reason for the higher accuracy of the ETR model.

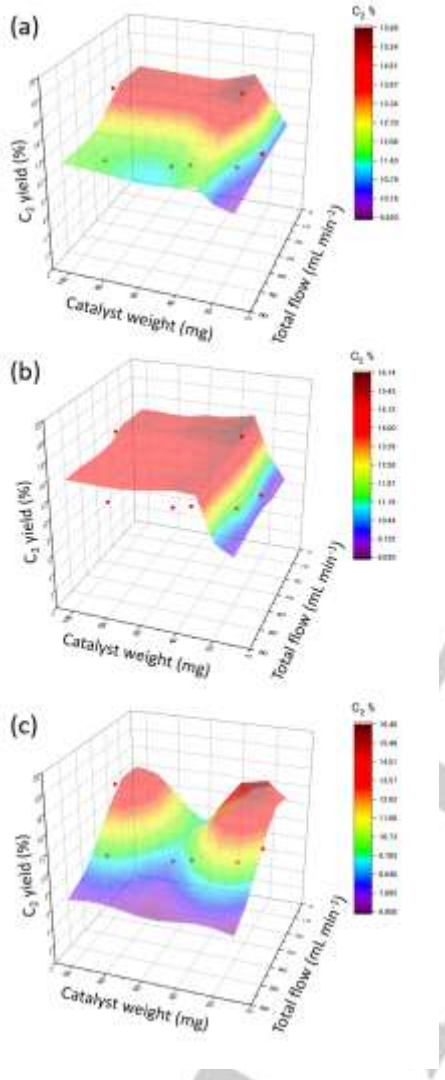


Figure 3. Surface 3D plots of the predicted C_2 yield against the catalyst weight and the total flow rate: (a) ETR and (b) SVR prediction. The experimental data are over-layed on the figures as red circles. The other reaction parameters: temperature 1123K, CH_4+O_2 conc 90%, CH_4/O_2 ratio 3.

Construction of accurate predictive models using a small sampling of experimental data is very attractive for boosting the development of catalyst reaction systems. Here, model construction using a small number of data is investigated using ETR, where the data is selected from 156 data while keeping the diversity of the data as indicated on Table S1. Although the average cross validation score of the ETR model constructed using 45 data is 0.45 , which is lower than that reported for the 156 data, the predicted results are still relatively accurate enough to act as a good guideline for designing experiments as shown in Fig. S1(f). Fig. 4 illustrates C_2 yield predicted by the ETR model based on the 45 data as a function of reaction temperature and CH_4/O_2 ratio (Fig. 4(a)) and that as a function of catalyst weight and total flow rate (Fig. 4(b)). Although the figures do not capture the detailed variation of the experimental C_2 yield by reaction conditions, they roughly agree with Fig. 2(a) and Fig. 3(a). Moreover, ETR with a small data set can be a good guideline for determining the next experimental conditions. In particular, Fig. 4(b) suggests that there might be hot spots for high C_2 yield at 20-40 mg of catalysts weight with low total flow. Hence, machine learning with a small data set can be used to design experiments for achieving high C_2 yield. Although generalizations of non-linear machine learning can be limited due to the nature of random forest regression, machine learning can be a powerful approach to accelerate the understanding of the effects of reaction conditions and for finding the optimal reaction conditions.

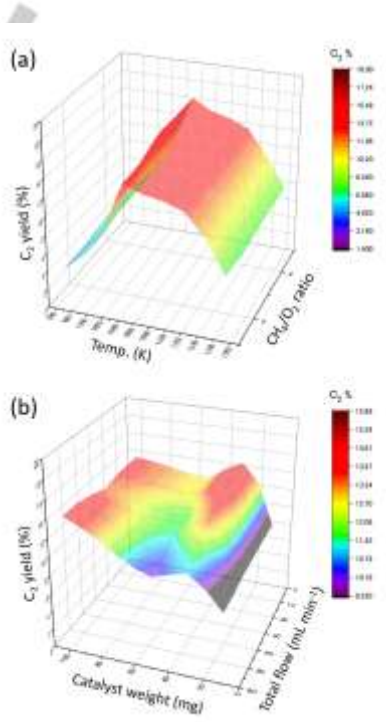


Figure 4. Surface 3D plots of the predicted C_2 yield by the ETR based on 45 data (a) as a function of reaction temperature and CH_4/O_2 ratio (catalyst weight 100 mg, CH_4+O_2 conc 90%, total flow 30 mL min^{-1}) and (b) the catalyst weight and the total flow rate (temperature 1123K, CH_4+O_2 conc 90%, CH_4/O_2 ratio 3) .

The importance of the 5 descriptor variables in the ETR model with 156 data is evaluated and shown in Fig. 5. Reaction temperature shows the highest importance among the five descriptors, which is consistent with the experimental results in Fig. 1(a)-(e) showing C_2 yield dependence on the temperature

as well as with the empirical rule of chemical reactions being very sensitive to temperature as usually represented by Arrhenius equation. After reaction temperature, catalyst weight exhibits high importance, but total flow rate shows a low degree of importance. These results suggest that a highly accurate model of the OCM reaction cannot be constructed only by using contact time calculated by catalyst weight and total flow. The higher importance of catalyst weight might reflect higher importance of the amount of active surface sites affected by the desorption rate of the fragments such as proton and hydroxyl groups on the catalyst surface simultaneously generated in the catalytic OCM reaction.[22] In fact, the previous studies have suggested that the desorption of the fragments needs higher energy than the activation of methane.[22] It is also interesting that CH_4+O_2 conc showed higher importance than CH_4/O_2 ratio. This result might reflect that the adsorption of CH_4 and O_2 promotes the desorption of the fragments rather than N_2. It should be noted that this matter must be carefully investigated using theoretical calculations or experimental spectroscopies to confirm whether the importance reflect the reaction mechanism, since the importance of ETR are determined from contribution during the tree development on ETR.

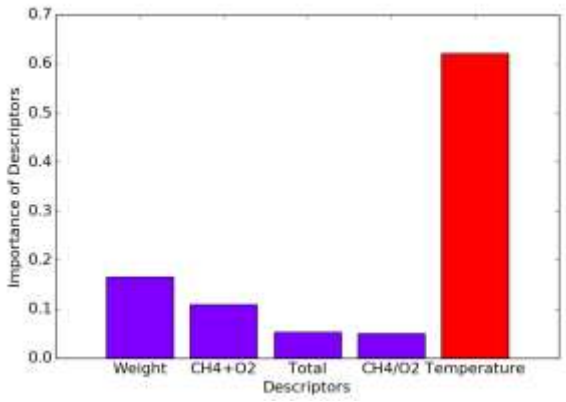


Figure 5. Importance of five descriptors in the ETR prediction.

Conclusions

Accurate models describing the effect of the OCM reaction conditions on the C_2 yield can be constructed by machine learning. Two linear regression methods (simple linear and linear SVR) and three non-linear regression methods (non-linear SVR, RFR, and ETR) of machine learning have been examined for constructing models using 156 data with six-dimensions composed of five reaction parameters as the descriptors and C_2 yield as the objective variable. It is revealed that the relationship between experimental conditions and C_2 yield lies upon non-linear matter where ETR is found to accurately predict the experimental data. Furthermore, ETR with a small data set of 45 data demonstrates that prediction from machine learning can be a good indicator for designing experiments for achieving high C_2 yield. This observation suggests that machine learning can be used earlier in the research stage, allowing researchers to apply machine learning on a smaller data set. The machine can do interpolation which allows for gaps in the data to be filled, thereby suggesting potential areas for further investigation. Researchers can then use those preliminary findings to help inform how the next set of experiments should be designed and conducted. Such approach could potentially reduce the time spent on conducting experiments as well as make the experiment design process more efficient. Thus, one can consider that machine learning is an effective tool for understanding and determining experimental conditions in high dimensions, thereby acting as a great aid for designing experiments.

Method

Experiment

Catalyst precursors, silica gel (60N), $Mn(NO_3)_2{\cdot}6H_2O$, and $Na_2WO_4{\cdot}2H_2O$ are supplied from Kanto Chemical Co., Fujifilm Wako Pure Chemical Co., and Sigma-Aldrich Co., respectively. Mn-Na_2WO_4/SiO_2 is prepared by a co-impregnation method. A 300 mL of aqueous suspension containing 5 g of SiO_2, 0.50 g of $Mn(NO_3)_2{\cdot}6H_2O$, and 0.45 g of $Na_2WO_4{\cdot}2H_2O$ is stirred at 323 K under stirring for 24 h. After evaporation of water at 338 K, the resulting solid is dried at 383 K overnight, and then calcined at 1273K for 3 h to obtain Mn-Na_2WO_4/SiO_2 containing 1.7wt% of Mn and 7.2wt% of Na_2WO_4.

OCM reaction is performed on conventional fixed bed reactors at atmospheric pressure. Before the reaction, the catalyst inside quartz tubes with 4 mm inside diameter is pretreated under O_2 flow at 1173 K. OCM reaction is conducted under various conditions using 10-100 mg of the catalyst at 973-1173 K under CH_4/O_2/N_2 with 1-10 of CH_4/O_2 ratio and 8-94% of CH_4+O_2 conc. at total flow rate of 8-96 mL min^{-1} (See Table S1). The effluent gas is analyzed by gas chromatograph. The C_2 yield (%) is determined by ((effluent C_2 conc / effluent N_2 conc)/(initial CH_4 conc / initial N_2 conc)) ×2×100.

Machine Learning

Scikit–learn (version 0.17) is implemented for machine learning. The following five machine learning algorithms are implemented: least squares linear regression (LSLR), support vector regression with linear kernel regression (SVR (L)), SVR with RBF kernel (SVR (RBF)), random forest regression (RFR), and extreme tree regression (ETR).[33] Hyperparameters of support vector regression with RBF kernel and RFR, and ETR are also optimized. C and gamma are set to 10 and 0.001 in SVR with RBF kernel, respectively. The number of trees in RFR and ETR are set to 1000 where the max depth of trees is set to none, meaning that the tree expands until the minimum numbers of samples to split reach 2. Cross validation is used to evaluate the accuracy of each machine learning algorithm where data is split into 20% test data and 80% trained data. The average R^2 scores of ten random test and trained data are taken and evaluated. Descriptor importance is estimated during the random forest regression model. Random forest regression generates a bunch of decision trees. Within the decision tree, importance is defined according to the number of times each descriptor shows up throughout the decision tree as well as its place within the decision tree.

[1] B. R. Goldsmith, J. Esterhuizen, J.-X. Liu, C. J. Bartel, C. Sutton, *AIChE J.* **2018**, *64*, 2311-2323.
[2] K. Takahashi, L. Takahashi, I. Miyazato, J. Fujima, Y. Tanaka, T. Uno, H. Satoh, K. Ohno, M. Nishida, K. Hirai, J. Ohyama, T. N. Nguyen, S. Nishimura, T. Taniike, *Chemcatchem* **2019**, *11*, 1146-1152.
[3] T. Hattori, S. Kito, *Catal. Today* **1995**, *23*, 347-355.
[4] S. Kito, T. Hattori, Y. Murakami, *Applied Catalysis a-General* **1994**, *114*, L173-L178.
[5] M. Sasaki, H. Hamada, Y. Kintaichi, T. Ito, *Applied Catalysis a-General* **1995**, *132*, 261-270.
[6] K. Huang, F.-Q. Chen, D.-W. Lü, *Applied Catalysis A: General* **2001**, *219*, 61-68.
[7] K. Huang, X.-L. Zhan, F.-Q. Chen, D.-W. Lü, *Chem. Eng. Sci.* **2003**, *58*, 81-87.
[8] Z.-Y. Hou, Q. Dai, X.-Q. Wu, G.-T. Chen, *Applied Catalysis A: General* **1997**, *161*, 183-190.
[9] L. Breiman, *Machine Learning* **2001**, *45*, 5-32.
[10] C. Cortes, V. Vapnik, *Machine Learning* **1995**, *20*, 273-297.
[11] C. C. Chang, C. J. Lin, *Acm Transactions on Intelligent Systems and Technology* **2011**, *2*.
[12] K. Takahashi, I. Miyazato, S. Nishimura, J. Ohyama, *ChemCatChem* **2018**, *10*, 3223-3228.
[13] J. K. Norskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, *Nature Chemistry* **2009**, *1*, 37-46.
[14] I. Takigawa, K. I. Shimizu, K. Tsuda, S. Takakusagi, *Rsc Advances* **2016**, *6*, 52587-52595.
[15] T. Fujitani, I. Nakamura, T. Akita, M. Okumura, M. Haruta, *Angew. Chem. Int. Ed.* **2009**, *48*, 9515-9518.
[16] T. Takei, T. Akita, I. Nakamura, T. Fujitani, M. Okumura, K. Okazaki, J. Huang, T. Ishida, M. Haruta, C. G. Bruce, C. J. Friederike, *Adv. Catal.* **2012**, *55*, 1-126.
[17] J. Ohyama, A. Esaki, T. Koketsu, Y. Yamamoto, S. Arai, A. Satsuma, *J. Catal.* **2016**, *335*, 24-35.
[18] J. Ohyama, T. Koketsu, Y. Yamamoto, S. Arai, A. Satsuma, *Chem. Commun.* **2015**, *51*, 15823-15826.
[19] M. M. Schubert, S. Hackenberg, A. C. van Veen, M. Muhler, V. Plzak, R. J. Behm, *J. Catal.* **2001**, *197*, 113-122.
[20] E. V. Kondratenko, T. Peppel, D. Seeburg, V. A. Kondratenko, N. Kalevaru, A. Martin, S. Wohlrab, *Catal. Sci. Tech.* **2017**, *7*, 366-381.
[21] U. Zavyalova, M. Holena, R. Schlögl, M. Baerns, *ChemCatChem* **2011**, *3*, 1935-1947.
[22] P. Schwach, X. Pan, X. Bao, *Chem. Rev.* **2017**, *117*, 8497-8520.
[23] S. Arndt, T. Otremba, U. Simon, M. Yildiz, H. Schubert, R. Schomäcker, *Applied Catalysis A: General* **2012**, *425-426*, 53-61.
[24] S. Pak, P. Qiu, J. H. Lunsford, *J. Catal.* **1998**, *179*, 222-230.
[25] H. Liu, X. Wang, D. Yang, R. Gao, Z. Wang, J. Yang, *Journal of Natural Gas Chemistry* **2008**, *17*, 59-63.
[26] U. Simon, O. Görke, A. Berthold, S. Arndt, R. Schomäcker, H. Schubert, *Chem. Eng. J.* **2011**, *168*, 1352-1359.
[27] M. Yildiz, Y. Aksu, U. Simon, K. Kailasam, O. Goerke, F. Rosowski, R. Schomäcker, A. Thomas, S. Arndt, *Chem. Commun.* **2014**, *50*, 14440-14442.
[28] H. R. Godini, A. Gili, O. Görke, S. Arndt, U. Simon, A. Thomas, R. Schomäcker, G. Wozny, *Catal. Today* **2014**, *236*, 12-22.
[29] S. Sadjadi, S. Jašo, H. R. Godini, S. Arndt, M. Wollgarten, R. Blume, O. Görke, R. Schomäcker, G. Wozny, U. Simon, *Catal. Sci. Tech.* **2015**, *5*, 942-952.
[30] S. Jašo, S. Sadjadi, H. R. Godini, U. Simon, S. Arndt, O. Görke, A. Berthold, H. Arellano-Garcia, H. Schubert, R. Schomäcker, G. Wozny, *Journal of Natural Gas Chemistry* **2012**, *21*, 534-543.
[31] Ehsani, Mohammad Reza, Hamed Bateni, and Ghazal Razi Parchikolaei, *KOREAN J CHEM ENG,* **2012**, 29, 855-861.
[32] P. Geurts, D. Ernst, L. Wehenkel, *Machine Learning* **2006**, *63*, 3-42.
[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *Journal of Machine Learning Research* **2011**, *12*, 2825-2830.

Acknowledgements

This work is funded by Japan Science and Technology Agency (JST) CREST Grant NumberJPMJCR17P2, JSPS KAKENHI Grant-in-Aid for Young Scientists (B) Grant Number JP17K14803, and Materials research by Information Integration (MI^2I) Initiative project of the Support Program for Starting Up Innovation Hub from JST.

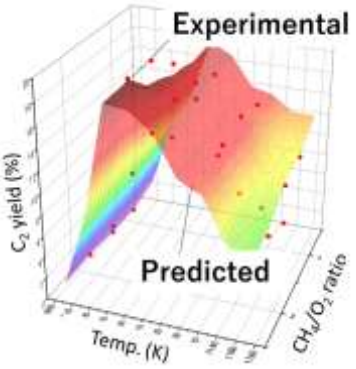
Keywords: Heterogeneous catalysis • machine learning • oxidative coupling of methane

10.1002/cctc.201900843

WILEY-VCH

FULL PAPER

The challenge in catalytic reactions lies within its complexity coming from high dimensional experimental factors. In order to solve such complexity, machine learning is implemented to treat experimental conditions in high dimensions. Oxidative coupling of methane, methane to C_2 compounds (ethylene and ethane), is chosen as the prototype reaction.



Junya Ohyama, Shun Nishimura, and Keisuke Takahashi**

Page No. – Page No.

Data Driven Determination of Reaction Conditions in Oxidative Coupling of Methane via Machine Learning