

Article

Not peer-reviewed version

Model-Based Catalyst Screening and Optimal Experimental Design for the Oxidative Coupling of Methane

[Anjana Puliyaanda](#) *

Posted Date: 12 December 2023

doi: 10.20944/preprints202312.0761.v1

Keywords: catalyst screening; catalyst informatics; high-throughput experiments; optimal experimental design; random forests; genetic algorithms; Bayesian optimization



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Model-Based Catalyst Screening and Optimal Experimental Design for the Oxidative Coupling of Methane

Anjana Puliyaanda * ¹ University of Alberta; puliyand@ualberta.ca

* Correspondence: puliyand@ualberta.ca

Abstract: The oxidative coupling of methane (OCM) to produce ethane and ethylene (C₂ compounds) as platform chemicals involves complex chemistry with reactions both in the gas phase and on the catalyst surface, resulting in a distribution of products at the expense of C₂ selectivity. This work uses experimental data from a variety of mixed metal oxides on supports at different reaction conditions (temperature, contact time, and reactant flow rates) to train a random forest regressor that predicts methane conversion and C₂ selectivity (key performance indicators (KPIs)), and deploys it to locate optimal conditions that maximize C₂ yield for a catalyst. Investigating the regressor interpretability via feature importance reveals that the choice of metals and support are crucial to C₂ selectivity predictions, while the predictions of methane conversion are driven by the reaction conditions. The machine learning (ML) regressor is used as a surrogate to develop performance curves for each of the catalysts via a multi-objective optimization routine that seeks to maximize the KPIs in the decision space of reaction conditions, is seen to locate optimal conditions at which the maximum C₂ yields for catalysts are predicted to be 15%, higher on average. Analyzing the catalysts in the space of their performance curves with respect to a popular OCM catalyst, Mn-Na₂WO₄/SiO₂, reveals distinct patterns based on intrinsic properties of metals and supports. Further, the decision space with catalyst descriptors and reaction conditions is optimized for high C₂ yields using the ML surrogate, in a static multi-objective optimization routine, and an adaptive Bayesian routine, where the latter was found to have a wider field focus in proposing catalyst formulations and conditions. Transition metal oxides on a variety of supports were proposed but not their lanthanide oxide counterparts.

Keywords: catalyst screening; catalyst informatics; high-throughput experiments; optimal experimental design; random forests; multiobjective optimization; pareto curves; bayesian optimization; genetic algorithms

1. Introduction

Synthesis of platform chemicals via catalyzed reactions lead to a wide product distribution because every catalyst has different active sites, composition and response to operating conditions, making it complex to identify their role in reaction mechanisms. This challenges the selective and economical manufacturing of target products at scale, as evidenced by studies on the oxidative coupling of methane (OCM) where the selective formation of ethane/ethylene (C₂) for the polymer manufacturing chain is limited by the thermodynamically favored over-oxidation [1]. The Edisonian approach to material design has been replaced by information-driven platforms that seamlessly integrate digitized database with modeling and optimization for hypothesis-driven design decisions [2]. Central to these platforms are machine learning (ML) surrogates that map properties of candidate materials in the database to process performance outcomes, so that appropriate materials can be recommended for experimental synthesis [3]. This work seeks to use the high-throughput experimental database for a variety of mixed metal oxide catalysts ($M_1 - M_2 - M_3O_4/\text{Support}$) to train ML surrogates for catalyst screening and to devise future experimental strategies that meet the selectivity-conversion targets for the extensively studied OCM chemistry.

The digitized data to build catalyst informatics platforms curated exhaustively from literature [4] are associated with inconsistencies in data, methods and reproducibility because of which there has been a shift towards high-throughput experimentation (HTE) [5], and high-throughput theoretical calculations (HTCs) [6] to reliably record catalyst performance across scales from the level of reaction energetics to process operations. Web-based visualization tools to deploy exploratory data analysis on HTE data using co-ordinated multiple views (CMVs) to discover apparent trends in the reaction performance across a variety of catalysts and operating conditions can provide insights for future experimentation [7]. Sophisticated ML tools to uncover the not so apparent insights require quantitative descriptors of a catalyst from elemental properties (atomic numbers, electron affinity, ionization energy, density) of constituent metal atoms from the periodic table to characterize its activity [8], or HTC-based reaction energetics descriptors from density functional theory [9,10]. Once the catalyst design space has been quantified by descriptors, unsupervised clustering can be used to identify catalyst groupings based on how they impact reaction performance, across different experimental conditions [11]. ML has been used to develop supervised descriptor-based reaction performance prediction models, and to minimize the time and cost in strategizing recommendations for physical experiments or theoretical calculations to guide exploration of the design space for materials discovery [12]. Descriptor-based ML models have been used to screen electrocatalysts for carbon-dioxide reduction [13], and also for the adaptive electrocatalyst and photocatalyst discovery either by human-in-the-loop learning, where the ML model is updated once the outcome has been observed via experimental/theoretical runs at algorithmically sampled points of the design space [14–16]. Alternatively, descriptor-based ML models have also been used for goal-driven exploration via Bayesian optimization or evolutionary genetic algorithms [17], to create self-driving laboratories that integrate databases (literature, HTE, HTC), ML and automated experimentation [18,19].

Most of the aforementioned approaches are yet to reveal catalyst candidates for OCM chemistry with a C₂ yield > 30%, a threshold considered practical for industrial applications that are limited by the maximum achievable C₂ yields because the reactant methane, is much less reactive to oxygen than the target C₂ products leading to selectivity-conversion tradeoffs. Analysis of 1868 literature reported OCM catalysts, reveals that most of them barely meet 20% C₂ yields, with just ~ 12 of them surpassing the thresholds for feasible industrial production [4]. The inconsistencies of literature-reported data (missing data, mass balance errors), not only pose an obstacle to reproducibility but are shown to result in poorly trained regression models to predict reaction performance that register prediction outliers on literature data with C₂ yields greater than 30%. For instance, the support vector regression trained on literature-mined data for OCM chemistry to predict C₂ yields has $R^2 \sim 0.5 - 0.6$, which is not impressive, because of which catalyst candidates *discovered* by it when used as a surrogate in Bayesian optimization lacks diversity in predicted materials, with a narrow field around La₂O₃ derivatives, and a maximum C₂ yield of ~ 15-16% [20]. To ensure reliability of the database used to propose catalyst candidates for OCM chemistry, HTE data has been used with informatics tools for visualization, supervised ML and catalyst networks to uncover patterns among dynamically evolving factors like catalyst synthesis, composition and operating conditions on reaction performance [21]. However, going beyond the interpolation filling abilities of ML in multi-dimensional data to predict rare targets with C₂ yields > 30% when the HTE dataset used to train the ML surrogates covers yields capped as much lesser values ~ 20 %, is still an ongoing research effort. In that spirit, this manuscript represents an effort to create informed serendipity using ML surrogates to enhance discovery of catalyst candidates by avoiding a narrow field focus.

Most works outlined herein develop descriptor-based ML to predict C₂ yields and CH₄ conversion using random forest regression, or neural network formalisms with mass balance reconciliation for the same [22]. However, using these ML models to develop catalyst performance curves by tuning operating conditions that maximize both C₂ selectivity (S_{C_2}) and methane conversion (X_{CH_4}) for each of the catalysts, followed by using these performance curves to screen $M1 - (M2)_{1-2} - M3O_x$ /Support type catalysts with respect to the popularly used $Mn - Na_2 - WO_4/SiO_2$ for OCM chemistry, is yet

to be investigated. Most ML models have been rationalized in terms of feature importances of the descriptors in predicting reaction outcomes, however, ML model validation to deduce activation barriers via lumped reaction kinetics models for methane conversion, and subsequent overoxidation of C2, is also pending investigation. Additionally, the descriptor-based ML models have also been used to test an evolutionary framework of exploring combinations of catalyst descriptors and process conditions in tandem, that maximize the two-fold selectivity-conversion targets, in an attempt to propose new candidates for synthesis. A multi-objective optimization routine using the NSGA-2 genetic algorithm has been contrasted against a Bayesian optimization routine to propose candidates and operating conditions, in an attempt to analyse field focus in proposing candidates, and the number of generations/ sampling iterations that are required to arrive at proposed entities with the highest achievable C2 yields. Assessing the synthesis feasibility of the proposed candidates and their experimental validation is out of scope of this manuscript.

2. Methodology

2.1. Dataset

The HTE database for OCM chemistry using 40 types of $M1 - (M2)_{1-2} - M3O_x$ /Support catalyst, and 19 references across 216 experimental conditions leading to ~ 12700 data points [5], has been used in this work. The dataset is hosted on a web-based informatics platform called the *Catalyst Acquisition by Data Science* (CADS), and records reaction outcomes at the end of each run in sequentially programmed experimental campaigns across combinations of temperature (900, 850, 800, 775, 750 and 700 °C), total reagent flowrate (10, 15, 20 ml/min), CH_4/O_2 ratio (2,3,4 and 6 mol/mol), and contact times (0.75, 0.50, or 0.38 s) [23]. The conversion of methane and yields/selectivity of the target products, ethane and ethylene (C2 products), as well as the undesired products (carbon dioxide, carbon monoxide) as a result of over-oxidation have been recorded. The design space is defined by quantitative descriptors of the catalyst and the operating conditions. Catalyst descriptors encompass atomic numbers of its constituent metal atoms (M1, M2, M3), the composition of these metal elements in mol%, and finally the nature of the support indicated by its unique index identification, all of which have been indicated in the CADS data repository. Descriptors of the reaction conditions encompass temperature (T), contact time (t), total flow rate (\dot{V}°), methane flow rate ($\dot{V}_{CH_4}^\circ$) and $CH_4 : O_2$, and have also been furnished in the dataset. Hence, there are 12 descriptors in all, when it comes to defining the design space that has been investigated via HTE to find optimal combinations that maximize methane conversion and C2 selectivity for OCM chemistry.

2.2. Random forest regression

Data collection under controlled experimental conditions eliminates most inconsistencies due to variability across experimental platforms. Yet, one cannot avoid uncertainties in the recorded measurements either because of material balance violation or instrumental errors [24]. The use of an ensemble model that learns aggregate predictions in this work, runs a lower risk of overfitting such biases [25], placing lesser emphasis on the need to eliminate data points with higher mass balance uncertainties. Also, random forest regressors learn via decision thresholds on descriptors to segment the design space in which aggregate predictions are made, making it agnostic to scale of the features and eliminating the need for much data pre-processing. When highly parametrized set of ML models like neural networks that run the risk of overfitting to the artefacts in data is used, good pre-processing, efforts to embed the training with mass balances, and even truncation of datapoints with mass balance violation beyond a fixed threshold have been widely considered [22,26].

Random forests are an ensemble model comprising many decision trees. Each decision tree has binary nodes, where the number of samples n at each node splits into n_1 and n_2 samples based on the optimal segmentation of the j^{th} descriptor into 2 subnodes, based on a decision threshold θ , as outlined in Eqn 1

$$\begin{aligned} R_1(j, \theta) &= \{x_{ij} | x_{ij} \leq \theta\} \forall i = 1, 2, \dots, n_1 \\ R_2(j, \theta) &= \{x_{ij} | x_{ij} > \theta\} \forall i = 1, 2, \dots, n_2 \end{aligned} \quad (1)$$

The descriptor j and its value is chosen such that the residual sum of squares is minimized for the binary split at that tree node, achieved by minimizing the residual sum of squares as in Eqn 2, where \bar{y}_1 and \bar{y}_2 are the average target response of the samples in each of the subnodes $R_1(j, \theta)$ and $R_2(j, \theta)$.

$$\min_{j, \theta} \left[\min_{\bar{y}_1} \sum_{x_{ij} \in R_1(j, \theta)} (y_i - \bar{y}_1)^2 + \min_{\bar{y}_2} \sum_{x_{ij} \in R_2(j, \theta)} (y_i - \bar{y}_2)^2 \right] \quad (2)$$

Repeating the process, stratifies the d dimensional design space into W regions R_1, R_2, \dots, R_W to generate a decision tree given in Eqn 3, where $\mathbb{I}(\cdot)$ is the indicator function, i.e. $\mathbb{I} = 1$ if $x \in R_w$ else it is 0, w is the number of decision threshold splits, and \bar{y}_w is the average target response of the samples in the region R_w .

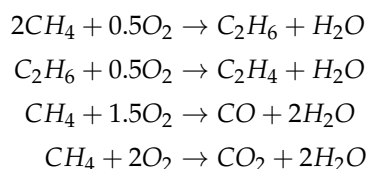
$$f(x) = \sum_{w=1}^W \bar{y}_w \mathbb{I}(x \in R_w) \quad (3)$$

This procedure is repeated on an ensemble of decision trees, $f_t(x)$, where $t = 1, 2, 3 \dots N_{trees}$ to generate a random forest model that aggregates the predictions across the learners in order to map the descriptors to the target response, as $\bar{f} : X \rightarrow Y$ given in Eqn 4

$$\bar{f}(x) = \frac{1}{N_{trees}} \sum_{t=1}^{N_{trees}} f_t(x) \quad (4)$$

2.3. Model validation by power-law reaction kinetics

This work demonstrates the ability of ML models to connect descriptors to parameters of lumped kinetic models to facilitate interpretability when used as a catalyst informatics tool. Evidence of ML being rationalized either by incorporating domain knowledge as constraints during its training or postfacto by the use of metrics like SHAP, feature importance and locally interpretable model explanations are quite popular. Here, in addition to just feature importance, we seek to interpret the relationship between descriptors and kinetic parameters using a lumped kinetic model for the otherwise complex reaction mechanism for OCM chemistry that involves both the gas and surface reactions [27]. The reactions considered in this work are as follows:



Once the random forest regression has been trained using the descriptors, the stoichiometry of the global reaction scheme by combining the above equations can be used to regress the lumped power-law kinetic model given in Eqn 5.

$$\begin{aligned} r_{CH_4} &= k_{10} \exp \left[\frac{-E_{aCH_4}}{RT} \right] \bar{P}_{CH_4}^a \bar{P}_{O_2}^b \\ r_{C_2} &= k_{20} \exp \left[\frac{-E_{aC_2}}{RT} \right] \bar{P}_{CH_4}^{a*} \bar{P}_{O_2}^{b*} \end{aligned} \quad (5)$$

Latin hypercube sampling (LHS) is used to randomly sample the process operation descriptors comprising temperature (T), contact time (t), total inlet volumetric flow rate (\dot{V}°), inlet volumetric flowrate of methane ($\dot{V}_{CH_4}^\circ$) and $CH_4 : O_2$ molar ratio, for a given set of catalyst descriptors comprising details of the elemental metals, their molar percentages and support ID. The reaction is performed in a continuous flow reactor, and is operated at a pressure P of 1 bar under isothermal conditions. Using this information the partial pressures of the reactants at the start of the reaction is calculated. The random forest regressor is then used to predict the conversion and selectivity for each set of descriptors across all samples, to obtain $\hat{X}_{CH_4}\%$ and $\hat{S}_{C_2}\%$, using which the reaction rates are expressed as given in Eqn 6, where $\dot{n}_{CH_4}^\circ$ is the molar flowrate of methane entering the reactor. The HTE data has been reported for 1g mass of catalyst support impregnated with the mixed metal oxides [5], and hence the kinetic models implicitly fit the specific reaction rates.

$$\begin{aligned} r_{CH_4} &= \dot{n}_{CH_4}^\circ \frac{\hat{X}_{CH_4}}{100} \\ r_{C_2} &= \dot{n}_{CH_4}^\circ \frac{\hat{X}_{CH_4} \hat{S}_{C_2}}{100} \end{aligned} \quad (6)$$

The stoichiometry of the global reaction scheme and the predicted methane conversion from the random forest regression is then used to calculate the partial pressure of the species at the end of the reaction. The average of the initial and the final partial pressures for oxygen and methane, along with the reaction rate expressions in Eqn 6 is substituted into the power-law kinetic expression in Eqn 5 for all the LHS sample points to estimate via regression the kinetic parameters *viz.* the Arrhenius pre-exponential factors for methane conversion (k_{10}) and C2 formation (k_{20}), their corresponding apparent activation energies (Ea_{CH_4} , Ea_{C_2}), and the orders of the species in each of the reactions (a,b,a*, b*). It must be noted that the HTE datasets report reaction performances only at the end of the contact time, owing to which fitting kinetic models to species concentration profiles is approximated by the average of the initial and final partial pressures. The estimated kinetic parameters are therefore treated as coarse estimates to characterize the impact of different catalysts on the reaction performance across varying operating conditions sampled by LHS.

2.4. Genetic algorithm for multi-objective optimization

The reaction performance for OCM chemistry in terms of methane conversion and C2 selectivity are known to exhibit a tradeoff, and depend on the operating conditions and the type of catalyst used. The combination of descriptor values $x \in R^{d \times 1}$, that maximizes both reaction performance indicators is posed as a multiobjective optimization problem (Eqn 7), where $f_{X_{CH_4}}(\cdot)$ and $f_{S_{C_2}}(\cdot)$ are the trained random forest regressors to predict the corresponding indicators.

$$\begin{cases} \max_x f_{X_{CH_4}}(x) \\ \max_x f_{S_{C_2}}(x) \end{cases} \quad (7)$$

$$\text{S.T.} \quad lb \leq x \leq ub \quad (8)$$

$$M1\% + M2\% + M3\% = 100 \quad (9)$$

$$\dot{V}^\circ - \dot{V}_{CH_4}^\circ - \dot{V}_{CH_4}^\circ \frac{1}{CH_4 : O_2} > 0 \quad (10)$$

The multiobjective optimization is constrained by limits of the descriptors (Eqn 8), given by the their range bounds in the HTE dataset, and Eqn 10 to ensure that the inlet volumetric flow of the inert Ar gas is non-negative. When solving the above optimization to develop performance curves for catalyst formulations outlined in the HTE datasets, the elemental metals (M1, M2, M3), their

corresponding molar percentages and their supports are fixed descriptor values, with the decision variables comprising just the operating conditions. However, when it comes to proposing different catalysts, their descriptors along with the operating conditions are treated as variables in the decision space in solving the multiobjective optimization, wherein the additional constraint in Eqn 9 enforces closure in the molar percentages of the elemental metals of the catalyst. The elemental metals and their supports are treated as categorical descriptor values but the rest of the descriptors are continuous. The multiobjective optimization is solved using an evolutionary approach via the NSGA-2 genetic algorithm [28], using 50 individuals, over 200 iterations with a mutation probability of 0.2 and crossover probability of 0.8. The constraints are implemented via the Delta penalty approach [29], where the fitness is penalized for invalid individuals by constant factor delta that is subtracted from the objectives we seek to maximize.

2.5. Bayesian optimization for adaptive experimentation

Bayesian optimization is a sequential global optimization approach that iteratively samples the design space of decision variables using a probabilistic surrogate model [30], like a Gaussian process regressor that captures the distribution of target predictions, $P(y|x)$ for a given $x \in R^{d \times 1}$, and an acquisition function like expected improvement (EI) to guide sampling as given in Eqns 11-13, subject to the constraints in Eqns 8-10.

$$f(x) = \frac{f_{X_{CH_4}}(x)f_{S_{C_2}}(x)}{100} \quad (11)$$

$$x^* \leftarrow \underset{x \in R^{d \times 1}}{\operatorname{argmax}} f(x) \quad (12)$$

$$EI(x) = \sum_y (y - f(x^*))P(y|x) \quad (13)$$

The idea is to start with an initial number of LHS samples, say 10, from the space of decision variables ($x \in R^{d \times 1}$), and use the above method to sample by *exploitation* to find the most likely optimal solutions based on the posterior distribution, while also resorting to *exploration* by sampling from points in areas with low probability density in order to be able to find the combination of descriptor values that maximize the yield of the desired C2 products. This approach has been widely used to encourage serendipity while navigating the combinatorial explosion of the decision space of design descriptors for the goal-driven enumeration of candidates in material science [31].

3. Results and Discussion

Section 3.1 presents the results from estimating kinetic parameters for different catalysts in the HTE dataset for OCM chemistry using power-law models based on the predictions of the descriptor-based random forest regressors. Section 3.2 discusses the performance curves of the given catalyst by tuning the decision space of operating conditions to maximize reaction performance indicators. The catalysts are screened with respect to a reference with the aid of the performance curves. Section 3.3 is an attempt to use two different techniques *viz.* multiobjective and Bayesian optimization to navigate both the catalyst and operating condition descriptors to propose new candidates.

3.1. Assessment and validation of random forest regression via kinetic parameter estimation

A random forest (RF) regressor model is fitted to map the descriptors to the reaction performances by way of methane conversion and C2 selectivity using a 5-fold cross validation for model hyperparameter tuning to prevent it from overfitting to the training data. An 85% train-test split is used for the same and predictions are assessed on completely new test data. The parity plots shown in Figure A1 indicates that the RF model adequately captures trends in the training data, and generalizes well on the test data too. Error from instruments or intrinsic phenomena like sintering [24]

can potentially lead to mass balance inconsistencies in HTE datasets quantified in terms of the total carbon balance error based on the difference between the methane conversion and the yields of the products formed, as follows:

$$\text{Total carbon balance\%} = \frac{X_{CH_4} - Y_{C_2H_6} - Y_{C_2H_4} - Y_{CO} - Y_{CO_2}}{X_{CH_4}} \times 100$$

The ensemble nature of the RF regressors make them robust to fitting such biases as can be seen in Figure A2, where the total carbon balance is uncorrelated with the prediction errors of the RF models for both conversion and selectivity. RF regressors also have the inherent procedure of calculating feature/descriptor importance [32], based on the optimal choice of descriptor that most reduces the residual sum of squares (Eqn 2).

It can be seen from Figure 1 that the operational descriptors like temperature, inlet volumetric flowrate of methane and the $CH_4 : O_2$ ratio are more important than any of the catalyst descriptors to the prediction model for methane conversion. While for the C2 selectivity model, the catalyst descriptors encompassing atomic numbers of metal elements M1 and M2, and the support ID are important in addition to the operational descriptors of temperature and inlet volumetric flowrate of methane and $CH_4 : O_2$ ratio. Although the catalyst surface active sites at OCM conditions are still unknown, the rate determining step involves hydrogen abstraction by C-H bond cleavage of methane either via surface-active oxygen (Langmuir-Hinshelwood kinetics) or via oxygen from the lattice sites (Mars-Van Krevelen), and requires high temperatures [33]. Coupling of methyl radicals to form ethane that dehydrogenates to ethylene is thermodynamically less favored than its further oxidation to CO_x because of which, although high flowrates of reagents are known to increase methane conversion, lower proportions of oxygen is used [34]. Also, the use of M2 type promoters in mixed metal oxide catalysts of $M_1M_3O_4$ type are found to suppress further oxidation by hindering the exposure of the tetrahedral $M_3O_4^{2-}$ active site, thereby increasing C2 selectivity [35]. Hence, prediction of methane conversion is dominated by the identified operating condition descriptors, and that of C2 selectivity is governed by the said catalyst descriptors as seen in Figure 1.

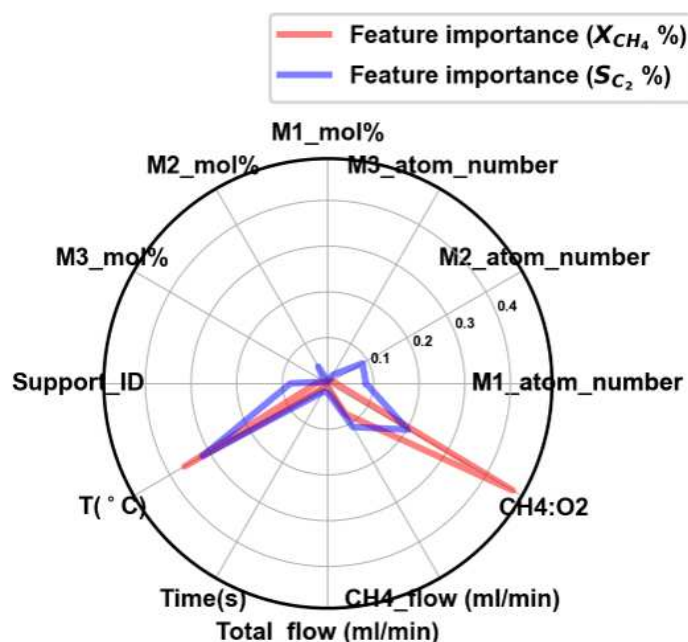


Figure 1. Feature importance of the RF regressors trained to predict methane conversion and C2 selectivity

Once the RF models have been duly fit and assessed, it is important to validate them via fitting power-law kinetic parameters for each of the catalysts in the HTE dataset. As specified in Section 2.3,

the points sampled via LHS in the space of the descriptors for the operating conditions are outlined in Table A1, and are used to regress the rate expressions to estimate the kinetic parameters for each catalyst. The $MnNa_2WO_4/SiO_2$ is a popular OCM catalyst that has registered high experimental C2 yields ~ 14 -27% and stability [36] and hence has been chosen as reference with respect to which the other catalysts are screened in this work. Figures 2b,c presents the regression to estimate the kinetic parameters for methane conversion and C2 formation, using the RF model predicted conversion and selectivity values at the LHS points as shown in Figure 2a.

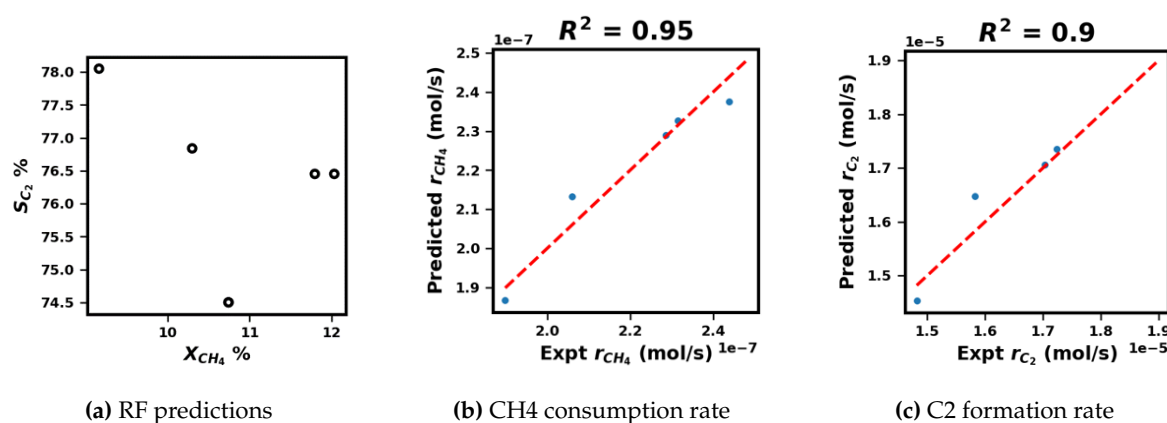


Figure 2. Regression fits for kinetic parameter estimation for $MnNa_2WO_4/SiO_2$ via the RF models.

The same procedure is followed for the other catalysts, where Figure A3 reports the regression fits, Figure A4 reports the species orders, and Figure 3c reports the apparent activation energies. The apparent activation energies for $MnNa_2WO_4/SiO_2$ in Figure 3a is within the literature ballpark ~ 200 -270 kJ/mol [27]. Clear groupings among the catalysts when the supports are varied for the $MnNa_2WO_4$ catalyst in Figure 3a, and the metal atoms are varied for a fixed SiO_2 support in Figure 3b have been observed. Transition metal oxide supports are reported to have better C2 yields than SiO_2 at similar process conditions [6], and even when porous aluminosilicates/zeolite-like supports or SiC are used as it leads to the formation of highly dispersed active sites after calcination [11], thereby having lower apparent activation energies for OCM reactions, as seen in Figure 3a. It must be noted that the regression calculated activation energies for zeolite-like supported $MnNa_2WO_4$ catalysts is very low, and is not the case in reality. This is primarily an artefact of *approximately* estimating kinetic parameters using RF predicted reaction KPIs at the end of contact time, and at the given LHS sample points by neglecting concentration gradients and potentially different chemistries. However, it is a reasonable approach to screen for trends and groupings among different catalysts. Similarly, in Figure 3b shows 3 groupings among alkali/alkaline earth metals, transition metals, and lanthanides/actinide metals on a fixed SiO_2 support with increasing apparent activation energies is seen. This can largely be attributed to the electronic properties of the mixed metal oxides by way of their electronegativity and ionization energies that impact the activation of methane and gas phase oxygen [37].

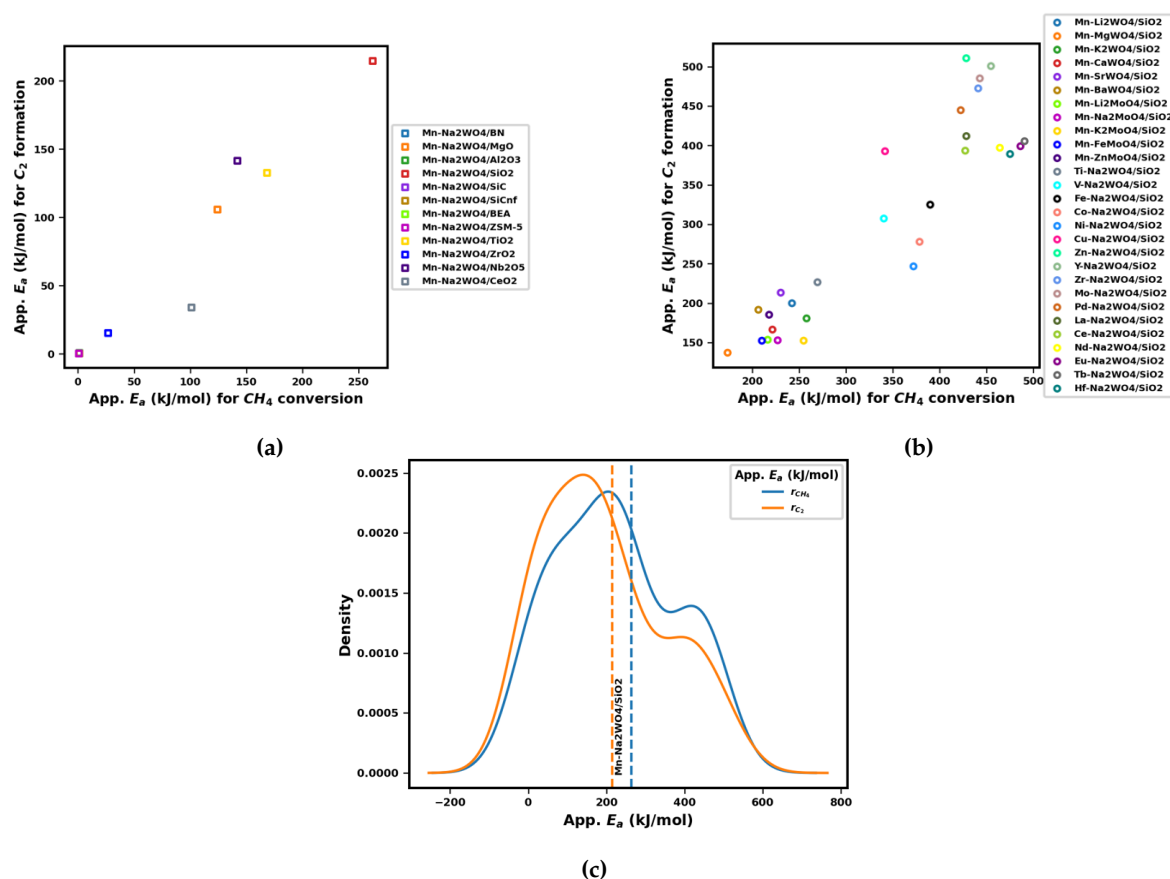


Figure 3. Apparent activation energy for methane consumption and C_2 formation estimated for (a) MnNa_2WO_4 across different supports, (b) different mixed metal oxide catalysts on SiO_2 support, and (c) distribution of the activation energies across the catalysts

3.2. Performance curves for catalyst screening

A good approach to compare how different catalysts i.e. ones with different supports or different mixed metal oxides impact the reaction KPIs of OCM chemistry is to use kinetic models to ascertain the best performance which can be achieved by a given catalyst over a range of operating conditions [38]. Using RF models as kinetic surrogates to maximize both methane conversion and C_2 selectivity for a given catalyst with fixed loading of metal atoms as outlined in the HTE dataset, in the decision space of just the 5 operating condition descriptors has resulted in the S-X performance curves shown in Figure 4.

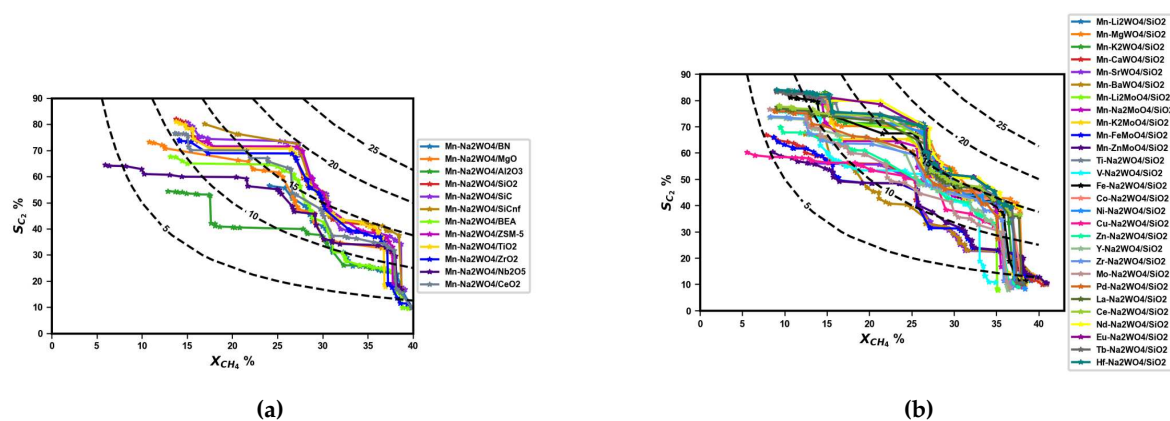
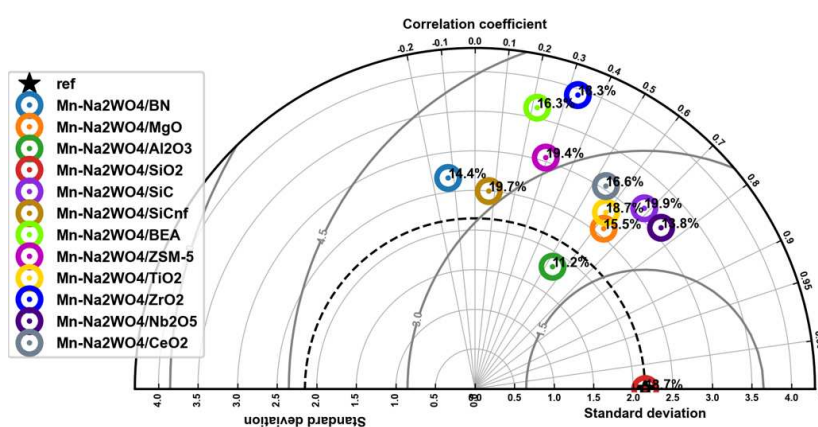


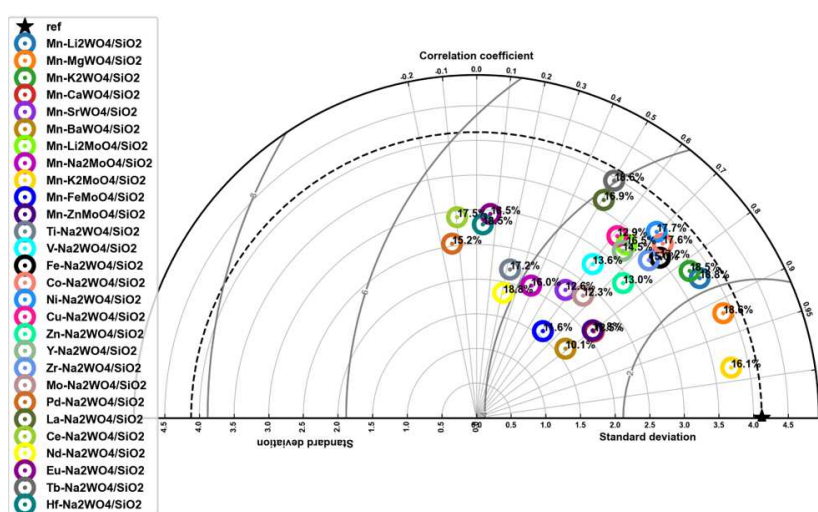
Figure 4. S-X performance curves for the OCM catalysts outlined in the HTE dataset.

The operating conditions corresponding to the best C2 yields obtained from these S-X performance curves have been tabulated in Tables A2 and A3, and shows on average across all catalysts $\sim 15\%$ improvement over the experimental values of the HTE dataset within the limits of the total carbon balance (TCB%). The C2 yield is defined as the product of the methane conversion and C2 selectivity that the RF models have been trained to predict, and combines the tradeoff between 2 of the reaction KPIs for the OCM reactions. Comparatively screening the catalysts based on the C2 yields in the locus of their S-X performance curves from multiobjective optimization, with respect to the reference catalyst reveals distinct groupings as seen in Figure 5. The figures indicate the RMS distance of the points on the performance curves of the catalysts from that of the reference, also the standard deviation of the performance curve C2 yields of the reference within which those of the other catalysts lie, and also the correlation of the same with the reference. The points are annotated with the maximum C2 yields of the performance curves that have been tabulated with their associated operating conditions in Appendix B.

These RF model-based screening plots reveal insights into interactions between the mixed metal oxides and the supports that contribute to surface properties linked to catalyst activity and C2 selectivity. Na-Mn-W oxides supported on SiO_2 is subject to severe restructuring owing to Na-induced phase change crystallization, but its superior thermal stability at high temperatures required for OCM reactions has made it a popular baseline [39]. The use of supports with different acid, basic or amphoteric nature to explore its synergies with Mn and W for redox cycles for O_2 activation and CH_4 activation, respectively can easily be deduced from Figure 5a. The use of acidic metal oxide supports like TiO_2 acts as a sink for alkali metal dopants, and stabilizes active species (MnO_x , WO_x) to facilitate lower temperatures to activate gas phase oxygen, thereby suppressing undesirable further oxidation to form CO_x [40]. The non-selective oxidation during high temperature exothermic reactions when acidic metal oxide supports are used is mitigated by using less acidic supports like SiO_2 instead of Al_2O_3 , alkali metal promoters (M_2) to neutralize acidic sites in the support, altogether using basic metal oxides for supports (MgO, BaO, CaO) despite their inability to stabilize WO_x required for CH_4 activation, and even by using an inert gas stream like Ar to dissipate hotspots [37]. SiO_2 and aluminosilicate-based materials like zeolite supports are known to phase transform by crystallization due to the alkali metal dopant (Na), which is conducive for the dispersion and stabilization of the WO_x active species [41]. The phase change of the support is seen to cause a drastic decrease in the surface area for Al_2O_3 , ZrO_2 and SiO_2 , but a sharp increase for SiC supports that have an added benefit of thermal stability [38]. The surface area and porosity of the supports correlates positively with C2 yield owing to better dispersion of active sites, as with zeolite supports [42]. However, if the supports are highly porous and also acidic in nature, alkali metal dopants are used to poison surface sites to limit excessive unselective oxidation [37]. It must be noted that neither the role of active sites nor individual supports is clear in OCM chemistry because even blank tests have shown good activity and C2 selectivity leading to questions about the actual contribution of the support to the chemistry [36]. However, it must be pointed that surface area, porosity, acidity/basicity, thermal stability and phase change amenability of the different supports for Na-Mn-W oxides justify the insights deduced from the patterns of the screening plot in Figure 5a.



(a)



(b)

Figure 5. Screening catalysts based on correlation, standard deviation and root mean squared distances of their C2 yields from the S-X curves, with respect to that of the reference $MnNa_2WO_4/SiO_2$.

The screening plot in Figure 5b presents a comparative assessment of the C2 yields from the S-X performance loci for different mixed metal oxide catalysts on SiO_2 supports with respect to the reference catalyst. The patterns reveals a coupling between the electronic properties of the metal elements and the OCM reaction performance. Host oxides doped with alkali/alkaline earth metals are reported to increase C2 selectivity due to their low electronegativity and ionization energy [11], with mainly Na known to induce crystallization of SiO_2 supports, while most of these dopants are known to distort the active site (WO_x^{2-}) for methane activation [37]. But the stability of these dopants at the harsh reaction conditions is challenged, as evidenced by catalyst degradation when the highly volatile (low melting point metal dopant) Li is used because of which either promoted lanthanide oxides, or

alkali dopants with higher melting points are generally preferred for OCM reactions. The positive correlation between the conductivity of the M1 metals with catalytic performance has been discussed, with Manganese (Mn) having the high electrical conductivity and hence preferred [43]. The type of M1 oxide and its oxidation state is known to be impacted by the nature of the oxo anion, and the choice of tungsten (W) and Molybdenum (Mo) as the M3 elements are discussed to have reasonable catalytic performance [43]. In line with this discussion, the three major groups identified in Figure 5b are seen to comprise alkali/alkaline earths, transition metals, and lanthanides.

3.3. Proposed candidates across combinations of catalysts and operating conditions

Different metal elements in the mixed metal oxides, their molar proportions, interactions with the support they are impregnated on, and even the operating conditions are known to exhibit synergies impacting reaction performance, from the discussions thus far. Most literature, until recently had deemed to achieve the best performance only if Mn, Na or K, and W were present [43]. The combinatorial explosion in navigating the design space of these catalyst and operating condition descriptors to create serendipity in proposing new candidates, is demonstrated by goal-driven sampling discovery via Bayesian optimization for C2 yield maximization as shown in Figure 6a. Alternatively, using the NSGA-2 genetic algorithm to solve multi-objective optimization in the decision space of catalyst and operating condition descriptors to maximize methane conversion and C2 selectivity, was seen to take longer to converge to values of maximum C2 yields, as seen in Figure A5.

Table 1. Best candidates for C2 yield maximization in the space of catalyst and operating condition descriptors across sampling iterations of Bayesian optimization.

Catalyst	M1 atom	M2 atom	M3 atom	M1 mol%	M2 mol%	M3 mol%	Support ID	T(°C)	time(s)	V(ml/min)	V _{CH₄} (ml/min)	CH ₄ : O ₂	Y _{C₂} %
Mn-Na ₂ WO ₄ /CeO ₂	25	11	74	40.00	40.00	20.00	5	700.00	0.75	10.00	2.00	2.00	8.71
Mn-Li ₂ MoO ₄ /SiO ₂	25	3	42	40.00	40.00	20.00	11	775.00	0.75	10.00	7.30	6.00	9.07
Ti-Na ₂ WO ₄ /SiO ₂	22	11	74	40.00	40.00	20.00	11	700.00	0.75	10.00	2.40	4.00	9.78
Mn-FeMoO ₄ /SiO ₂	25	26	42	40.00	40.00	20.00	11	850.00	0.38	20.00	4.50	3.00	10.92
Mn-CaWO ₄ /SiO ₂	25	20	74	40.00	40.00	20.00	11	700.00	0.38	20.00	4.50	3.00	11.63
Fe-Li ₂ MoO ₄ /Nb ₂ O ₅	26	3	42	44.81	27.25	26.48	8	815.49	0.65	15.76	11.26	2.69	12.32
Mo-Li ₂ MoO ₄ /ZrO ₂	42	3	42	44.24	27.44	26.84	13	821.44	0.49	16.37	6.30	5.73	12.80
Mo-Na ₂ MoO ₄ /ZrO ₂	42	11	42	44.02	27.77	26.37	13	799.38	0.70	10.16	3.62	5.59	13.73
Mn-CaWO ₄ /TiO ₂	25	20	74	44.17	27.01	26.96	12	726.27	0.38	15.31	12.00	4.60	15.07
Cu-K ₂ WO ₄ /SiO ₂	29	19	74	44.49	27.94	26.83	11	823.97	0.74	17.49	3.30	4.81	17.11
Ti-K ₂ MoO ₄ /SiCnf	22	19	42	44.44	27.54	26.10	10	799.36	0.49	15.55	2.11	4.26	17.42
V-K ₂ WO ₄ /CeO ₂	23	19	74	44.98	27.74	26.68	5	784.46	0.50	12.53	2.12	5.73	17.53
Mn-K ₂ MoO ₄ /SiCnf	25	19	42	44.93	27.70	26.99	10	818.22	0.73	15.25	2.10	5.94	19.04
Ti-MgMoO ₄ /ZSM-5	22	12	42	44.11	27.03	26.21	14	790.11	0.50	17.69	2.20	2.25	19.25
Mn-Li ₂ WO ₄ /SiO ₂	25	3	74	44.87	27.80	26.99	11	804.92	0.51	18.37	2.01	5.48	19.36

The descriptor values of the 50 individuals (annotated with black numbers in increasing order of the C2 yields) at the end of 3000 solution generations have been visualized in a 2d t-SNE plot of Figure 6b. Additionally, the descriptors corresponding to the best C2 yield among the samples picked by the acquisition function in each iteration of the Bayesian optimization have also been visualized in the 2d t-SNE plot, and are annotated in pink by their sampling indices in increasing order of the C2 yields. The descriptors corresponding to the experimental dataset are also visualized on the t-SNE plot, and all the points are shaded by their associated C2 yields predicted by the RF models. The catalysts proposed by the multiobjective optimization was found to have a narrow field focus in proposing candidates, as most of the catalysts proposed were *MnWO₄* doped with alkali/alkaline earth metals on either SiC, SiCnf or *SiO₂* supports at their respective optimal operating conditions. Hence, only the descriptor candidates corresponding to the best C2 yields among the sampling iterations from Bayesian optimization have been presented in Table 1. It can be seen that a number of new candidates like mixed metal oxides of transition metals on a wide variety of supports have been proposed to have comparable performance as the best performing Mn-Na-W family of baseline catalysts. However, no new mixed metal lanthanide oxides have been proposed. Lanthanide group elements have been reported to hinder the exposure of *WO_x²⁻*, and thereby lower catalytic activity [11]. The feasibility of synthesizing the tabulated catalysts, characterizing them and designing experiments around the

specified optimal operating conditions, encourages goal-driven approaches to experimentation in the future.

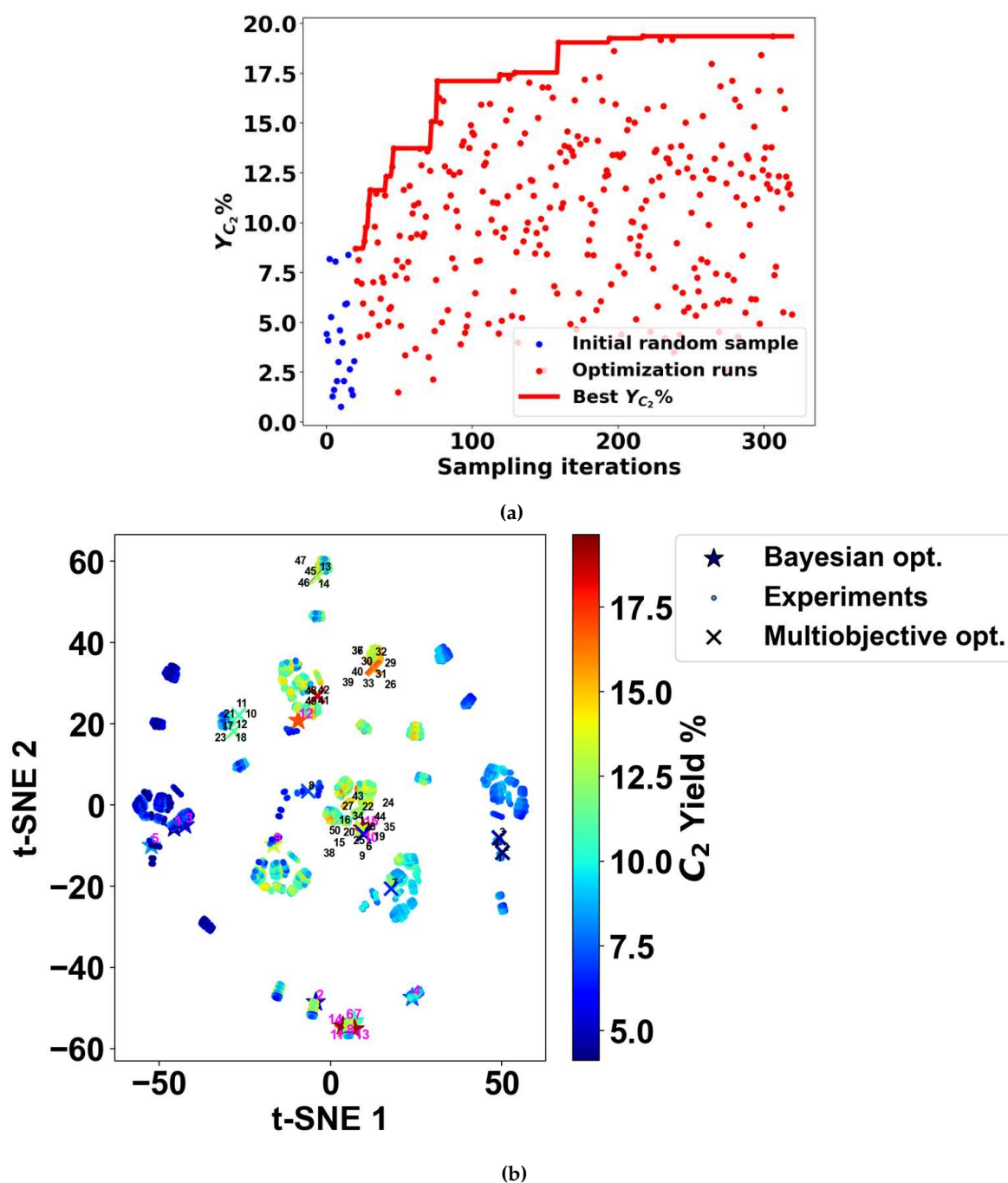


Figure 6. Goal-driven design of experiments to maximize C₂ yields by evaluating combinations of catalyst constituents and operating conditions.

4. Conclusions

This work develops a descriptor-based random forest regression model that maps to the reaction KPIs of methane conversion and C₂ selectivity furnished in the HTE dataset for OCM chemistry, spanning a wide variety of catalysts and operating conditions. The synergies among the catalyst constituents and operating conditions impact the conversion of reactant methane to selectively form ethane and ethylene by suppressing their undesirable total oxidation to form CO_x. Ethane and ethylene are viable platform chemicals in polymer processing. The feasibility of their industrial production

necessitates the OCM reactions to be designed with catalysts and operating conditions resulting in C2 yields greater than 30%. Most literature and trial-and-error experimental efforts have fallen short. Also, the maximum reported C2 yield in the HTE dataset in this work is 20%, within the limits of the total carbon balance. This limits the models trained herein to make higher predictions. However, the focus is to deploy the models in (i) screening catalysts to identify relationships between catalyst properties and reaction KPIs, (ii) optimizing the operating conditions for catalyst formulations in the HTE datasets to maximize both methane conversion and C2 selectivity, and (iii) proposing new catalysts and their optimal operating conditions by a goal-driven Bayesian optimization for C2 yield maximization to guide future experimentation. The ability of the RF models to capture lumped kinetics has been validated and is shown to reveal patterns among SiO_2 supported catalysts, and among MnWO_4 catalysts across different supports, in the space of the estimated kinetic parameters. The RF models were found to improve the C2 yield by $\sim 15\%$ on average when used to optimize operating conditions for catalysts in the HTE dataset to meet both methane conversion and C2 selectivity targets. Screening the catalysts in the space of the best performance achieved across a range of operating conditions along the S-X curves are found to reveal similar patterns, as in the space of the kinetic parameters estimated by the RF model. A number of transition metal oxides on different supports have been proposed by the Bayesian optimization routine, but lanthanide metal oxides were not sampled.

Author Contributions: A.P. was solely responsible for all aspects of the manuscript preparation.

Funding: This research received no external funding.

Data Availability Statement: The dataset in this study can be accessed using the following link: <https://cads.eng.hokudai.ac.jp/datamanagement/datasources/21010bbe-0a5c-4d12-a5fa-84eea540e4be/>.

Acknowledgments: Anjana Puliya acknowledges discussions with Vinay Prasad.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Assessing the random forest regression model

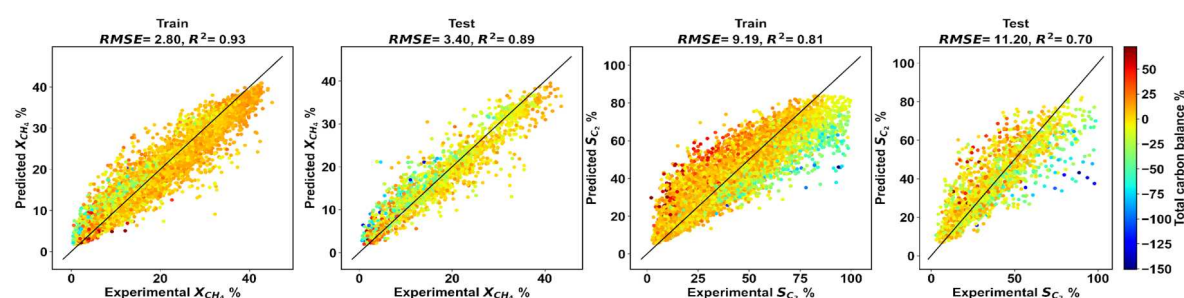


Figure A1. Parity plots between the model predictions for methane conversion and C2 selectivity against the experimental data shaded by the total carbon balance.

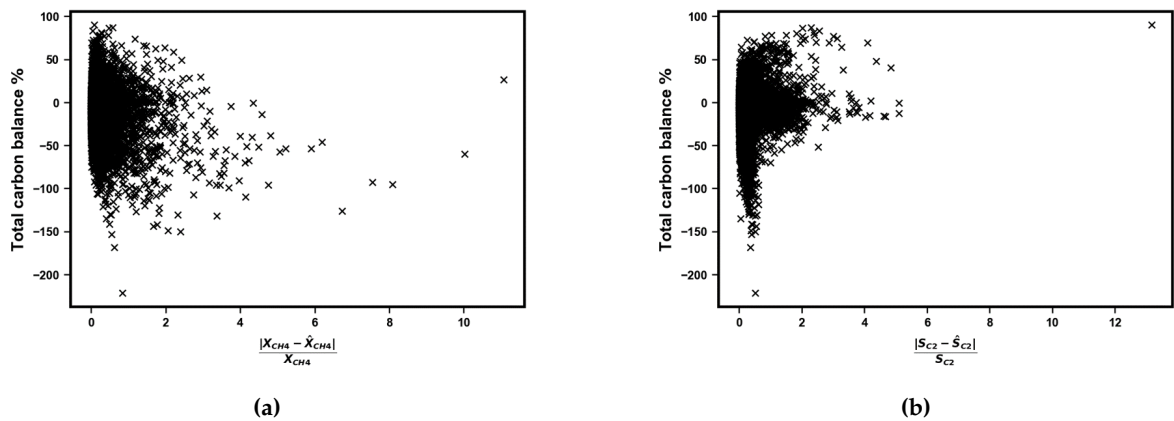


Figure A2. Random forest regressor predictions are shown to be uncorrelated to the total carbon balance

Appendix A.2. Tabulation of LHS sample points

Table A1. Points sampled by LHS from the space of operating condition descriptors

T(°C)	time(s)	\dot{V} (ml/min)	\dot{V}_{CH_4} (ml/min)	CH ₄ :O ₂ (mol:mol)
756.17	0.60	21.62	10.92	15.41
747.83	0.41	21.04	10.42	16.67
751.17	0.53	20.88	10.08	12.74
752.83	0.47	21.63	10.75	12.29
749.50	0.72	21.13	10.58	19.54
754.50	0.66	21.21	10.25	10.70

Appendix A.3. Fits and orders of power-law kinetic parameter estimation

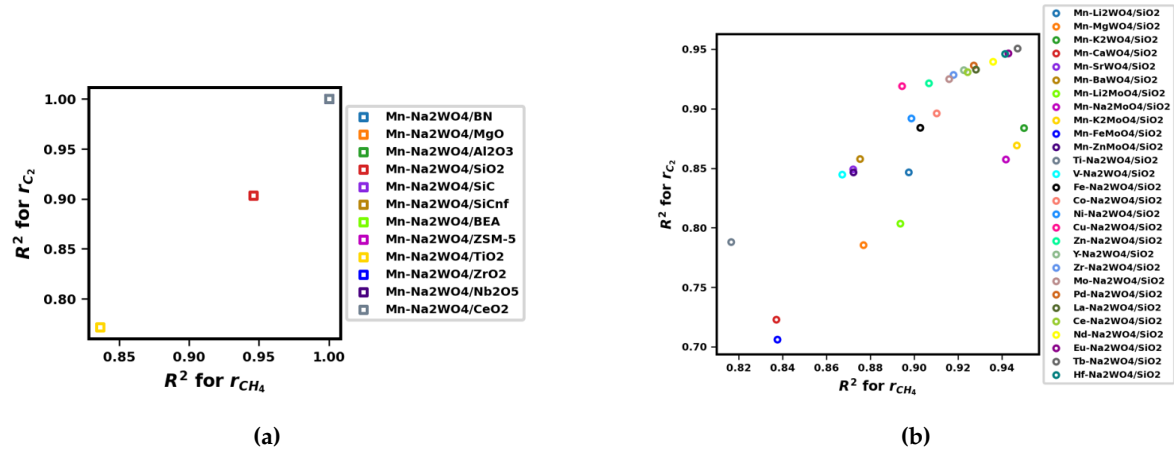


Figure A3. Regression fits for the estimation of power-law kinetic parameters

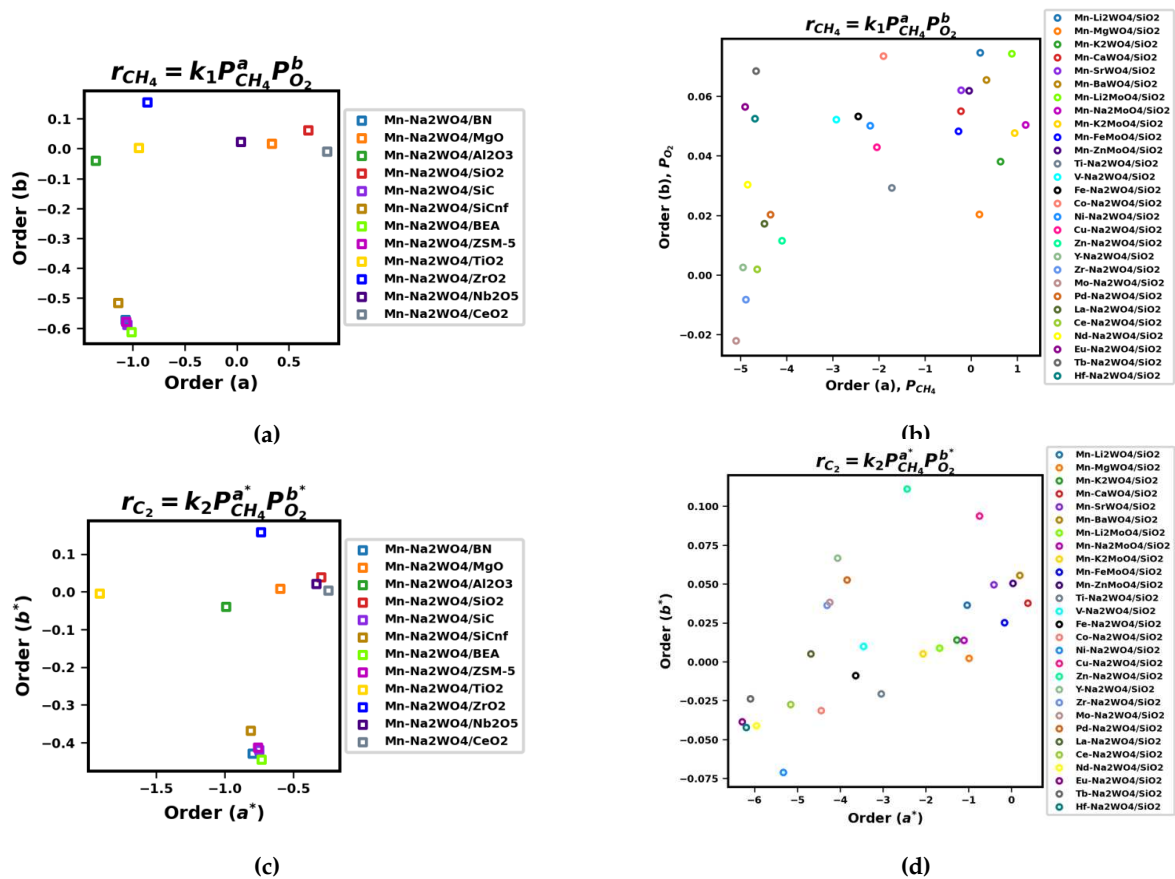


Figure A4. Estimated orders for the consumption of methane (a),(b), and those for the C2 formation (c),(d) for different catalysts.

Appendix B

Table A2. Performance curve conditions to improve C2 yields as compared to the best reported values in the HTE dataset for $MnNa_2WO_4$ catalysts with different supports

Catalyst	Experimental conditions							S-X performance curve conditions						
	T(°C)	time(s)	V(ml/min)	V_{CH_4} (ml/min)	CH ₄ O ₂ (mol/mol)	max Y_{C_2}	TCB%	T(°C)	time(s)	V(ml/min)	V_{CH_4} (ml/min)	CH ₄ O ₂ (mol/mol)	max Y_{C_2}	Y_{C_2} improvement%
Mn-Na ₂ WO ₄ /BN	800.00	0.50	15.00	9.60	3.00	7.75	-42.85	787.33	0.71	12.37	2.05	5.26	14.36	85.30
Mn-Na ₂ WO ₄ /MgO	800.00	0.50	15.00	3.40	3.00	9.32	6.05	812.79	0.46	15.67	2.10	5.61	15.47	66.02
Mn-Na ₂ WO ₄ /Al ₂ O ₃	750.00	0.38	20.00	12.80	3.00	8.08	-15.04	822.76	0.40	18.44	2.15	3.19	11.21	38.77
Mn-Na ₂ WO ₄ /SiO ₂	800.00	0.50	15.00	3.00	2.00	21.03	-0.71	788.05	0.53	13.95	2.10	5.39	18.72	-10.99
Mn-Na ₂ WO ₄ /SiC	800.00	0.50	15.00	3.40	3.00	19.59	2.06	808.30	0.59	16.55	2.06	5.66	19.90	1.59
Mn-Na ₂ WO ₄ /SiCnf	800.00	0.38	20.00	4.00	2.00	19.15	-1.83	812.97	0.59	14.74	2.04	5.79	19.69	2.80
Mn-Na ₂ WO ₄ /BEA	800.00	0.38	20.00	4.50	3.00	15.56	-0.77	792.61	0.50	12.74	2.04	5.23	16.33	4.93
Mn-Na ₂ WO ₄ /ZSM-5	800.00	0.38	20.00	4.50	3.00	19.90	-1.94	817.63	0.67	12.58	2.09	5.81	19.36	-2.71
Mn-Na ₂ WO ₄ /TiO ₂	750.00	0.38	20.00	4.00	2.00	18.29	5.69	821.57	0.57	14.80	2.11	5.31	18.71	2.29
Mn-Na ₂ WO ₄ /ZrO ₂	800.00	0.38	20.00	4.80	4.00	11.21	-3.64	793.97	0.60	14.02	2.05	5.26	18.28	63.11
Mn-Na ₂ WO ₄ /Nb ₂ O ₅	800.00	0.38	20.00	12.80	3.00	8.25	-11.21	813.40	0.62	17.65	2.11	5.86	13.81	67.44
Mn-Na ₂ WO ₄ /CeO ₂	775.00	0.75	10.00	2.00	2.00	18.04	0.23	819.77	0.55	18.39	2.15	5.94	16.62	-7.86

Table A3. Performance curve conditions to improve C2 yields as compared to the best reported values in the HTE dataset for mixed metal oxides on SiO₂ support.

Catalyst	Experimental conditions							S-X performance curve conditions						
	T(°C)	time(s)	V(ml/min)	V _{CH₄} (ml/min)	CH ₄ :O ₂ (mol:mol)	max Y _{C₂}	TCB%	T(°C)	time(s)	V(ml/min)	V _{CH₄} (ml/min)	CH ₄ :O ₂ (mol:mol)	max Y _{C₂}	Y _{C₂} improvement%
Mn-Li ₂ WO ₄ /SiO ₂	800.00	0.50	15.00	3.00	2.00	18.81	9.29	793.70	0.47	13.26	2.00	5.94	18.77	-0.21
Mn-MgWO ₄ /SiO ₂	775.00	0.50	15.00	3.00	2.00	16.08	5.92	805.43	0.45	13.66	2.09	5.87	18.59	15.59
Mn-K ₂ WO ₄ /SiO ₂	775.00	0.75	10.00	2.00	2.00	18.55	3.12	820.03	0.61	17.14	2.12	5.28	18.47	-0.45
Mn-CaWO ₄ /SiO ₂	850.00	0.38	20.00	4.80	4.00	8.51	10.87	870.22	0.39	17.95	2.02	5.08	12.55	47.46
Mn-SrWO ₄ /SiO ₂	850.00	0.38	20.00	4.80	4.00	10.65	12.74	833.07	0.39	18.57	2.06	5.77	12.61	18.40
Mn-BaWO ₄ /SiO ₂	850.00	0.38	20.00	5.10	6.00	10.17	13.48	788.44	0.52	19.85	12.02	4.84	10.05	-1.16
Mn-Li ₂ MoO ₄ /SiO ₂	800.00	0.38	20.00	4.00	2.00	14.00	7.74	769.54	0.63	11.26	2.13	5.98	16.45	17.53
Mn-Na ₂ MoO ₄ /SiO ₂	775.00	0.50	15.00	3.00	2.00	15.43	-0.58	798.53	0.54	17.36	2.14	5.02	16.01	3.74
Mn-K ₂ MoO ₄ /SiO ₂	800.00	0.38	20.00	4.50	3.00	16.60	-6.61	814.59	0.47	12.99	2.03	5.06	16.13	-2.84
Mn-FeMoO ₄ /SiO ₂	850.00	0.38	20.00	5.10	6.00	12.57	7.69	840.37	0.44	17.54	2.02	5.05	11.63	-7.45
Mn-ZnMoO ₄ /SiO ₂	850.00	0.50	15.00	3.90	6.00	12.96	15.70	856.03	0.41	19.40	2.06	5.38	11.78	-9.13
Ti-Na ₂ WO ₄ /SiO ₂	800.00	0.75	10.00	2.00	2.00	20.23	9.12	800.11	0.71	12.22	2.10	5.11	17.21	-14.95
V-Na ₂ WO ₄ /SiO ₂	775.00	0.50	15.00	6.00	2.00	8.58	-4.08	812.24	0.40	19.94	2.09	3.24	13.59	58.34
Fe-Na ₂ WO ₄ /SiO ₂	800.00	0.75	10.00	2.00	2.00	15.24	5.16	812.21	0.49	16.08	2.05	5.31	17.16	12.59
Co-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.50	3.00	16.14	7.41	823.83	0.51	14.33	2.14	5.90	17.64	9.32
Ni-Na ₂ WO ₄ /SiO ₂	800.00	0.50	15.00	3.00	2.00	17.66	8.01	806.45	0.47	12.85	2.06	5.64	17.74	0.47
Cu-Na ₂ WO ₄ /SiO ₂	800.00	0.38	20.00	8.00	2.00	9.11	-5.59	796.12	0.40	17.58	2.02	2.57	12.91	41.73
Zn-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.00	2.00	12.62	7.19	788.46	0.40	17.83	2.15	2.00	13.01	3.10
Y-Na ₂ WO ₄ /SiO ₂	850.00	0.50	15.00	3.40	3.00	12.56	-3.45	801.48	0.68	11.28	2.04	5.18	14.50	15.41
Zr-Na ₂ WO ₄ /SiO ₂	800.00	0.75	10.00	2.00	2.00	13.86	2.69	811.17	0.66	12.01	2.09	5.32	14.99	8.14
Mo-Na ₂ WO ₄ /SiO ₂	800.00	0.50	15.00	3.00	2.00	11.01	13.88	756.00	0.46	14.00	8.12	2.07	12.25	11.27
Pd-Na ₂ WO ₄ /SiO ₂	800.00	0.75	10.00	2.00	2.00	15.45	-2.62	794.41	0.73	10.61	2.15	5.16	15.20	-1.64
La-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.50	3.00	15.43	9.34	790.40	0.65	10.89	2.01	5.92	16.90	9.50
Ce-Na ₂ WO ₄ /SiO ₂	800.00	0.75	10.00	2.00	2.00	16.75	2.48	815.08	0.69	11.59	2.06	5.55	17.49	4.39
Nd-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.50	3.00	15.88	9.43	797.14	0.65	10.71	2.02	5.83	18.77	18.17
Eu-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.00	2.00	16.09	8.48	788.82	0.75	11.02	2.15	5.51	18.46	14.71
Tb-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.50	3.00	15.84	4.96	789.10	0.64	12.31	2.13	5.92	18.62	17.57
Hf-Na ₂ WO ₄ /SiO ₂	850.00	0.38	20.00	4.00	2.00	16.01	4.52	824.64	0.70	10.26	2.10	5.57	18.54	15.79

Appendix C

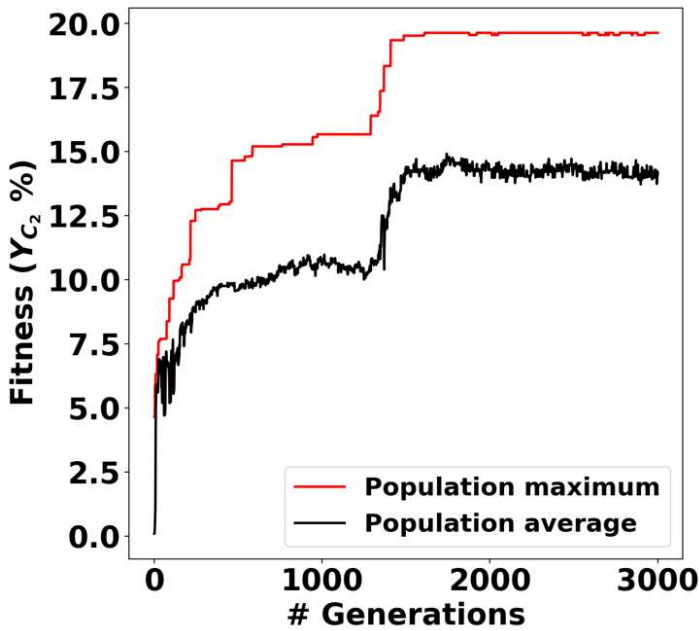


Figure A5. Genetic algorithm for multi-objective optimization of methane conversion and C2 selectivity across combinations of catalyst and operating condition descriptors.

References

1. Zhu, Z.; Guo, W.; Zhang, Y.; Pan, C.; Xu, J.; Zhu, Y.; Lou, Y. Research progress on methane conversion coupling photocatalysis and thermocatalysis. *Carbon Energy* **2021**, *3*, 519–540. doi:10.1002/cey2.127.
2. Weber, J.M.; Guo, Z.; Zhang, C.; Schweidtmann, A.M.; Lapkin, A.A. Chemical data intelligence for sustainable chemistry, 2021. doi:10.1039/d1cs00477h.
3. Takahashi, K.; Tanaka, Y. Materials informatics: A journey towards material design and synthesis, 2016. doi:10.1039/c6dt01501h.
4. Takahashi, K.; Miyazato, I.; Nishimura, S.; Ohyama, J. Unveiling Hidden Catalysts for the Oxidative Coupling of Methane based on Combining Machine Learning with Literature Data. *ChemCatChem* **2018**, *10*, 3223–3228. doi:10.1002/cctc.201800310.

5. Nguyen, T.N.; Nhat, T.T.P.; Takimoto, K.; Thakur, A.; Nishimura, S.; Ohyama, J.; Miyazato, I.; Takahashi, L.; Fujima, J.; Takahashi, K.; Taniike, T. High-Throughput Experimentation and Catalyst Informatics for Oxidative Coupling of Methane. *ACS Catalysis* **2020**, *10*, 921–932. doi:10.1021/acscatal.9b04293.
6. Takahashi, K.; Takahashi, L.; Le, S.D.; Kinoshita, T.; Nishimura, S.; Ohyama, J. Synthesis of Heterogeneous Catalysts in Catalyst Informatics to Bridge Experiment and High-Throughput Calculation. *Journal of the American Chemical Society* **2022**, *144*, 15735–15744. doi:10.1021/jacs.2c06143.
7. Fujima, J.; Tanaka, Y.; Miyazato, I.; Takahashi, L.; Takahashi, K. Catalyst Acquisition by Data Science (CADS): A web-based catalyst informatics platform for discovering catalysts. *Reaction Chemistry and Engineering* **2020**, *5*, 903–911. doi:10.1039/d0re00098a.
8. Ishioka, S.; Fujiwara, A.; Nakanowatari, S.; Takahashi, L.; Taniike, T.; Takahashi, K. Designing Catalyst Descriptors for Machine Learning in Oxidative Coupling of Methane. *ACS Catalysis* **2022**, *12*, 11541–11546. doi:10.1021/acscatal.2c03142.
9. Goldsmith, B.R.; Esterhuizen, J.; Liu, J.X.; Bartel, C.J.; Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE Journal* **2018**, *64*, 2311–2323, [\[https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.16198\]](https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.16198). doi:https://doi.org/10.1002/aic.16198.
10. Tamtaji, M.; Gao, H.; Hossain, M.D.; Galligan, P.R.; Wong, H.; Liu, Z.; Liu, H.; Cai, Y.; Goddard, W.A.; Luo, Z. Machine learning for design principles for single atom catalysts towards electrochemical reactions. *J. Mater. Chem. A* **2022**, *10*, 15309–15331. doi:10.1039/D2TA02039D.
11. Takahashi, K.; Takahashi, L.; Nguyen, T.N.; Thakur, A.; Taniike, T. Multidimensional Classification of Catalysts in Oxidative Coupling of Methane through Machine Learning and High-Throughput Data. *Journal of Physical Chemistry Letters* **2020**, *11*, 6819–6826. doi:10.1021/acs.jpcclett.0c01926.
12. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: Recent applications and prospects, 2017, [\[1707.07294\]](https://arxiv.org/abs/1707.07294). doi:10.1038/s41524-017-0056-5.
13. Zhang, N.; Yang, B.; Liu, K.; Li, H.; Chen, G.; Qiu, X.; Li, W.; Hu, J.; Fu, J.; Jiang, Y.; Liu, M.; Ye, J. Machine Learning in Screening High Performance Electrocatalysts for CO₂ Reduction, 2021. doi:10.1002/smt.202100987.
14. Mai, H.; Le, T.C.; Chen, D.; Winkler, D.A.; Caruso, R.A. Machine Learning for Electrocatalyst and Photocatalyst Design and Discovery, 2022. doi:10.1021/acs.chemrev.2c00061.
15. Masood, H.; Toe, C.Y.; Teoh, W.Y.; Sethu, V.; Amal, R. Machine Learning for Accelerated Discovery of Solar Photocatalysts, 2019. doi:10.1021/acscatal.9b02531.
16. Li, Z.; Achenie, L.E.; Xin, H. An Adaptive Machine Learning Strategy for Accelerating Discovery of Perovskite Electrocatalysts. *ACS Catalysis* **2020**, *10*, 4377–4384. doi:10.1021/acscatal.9b05248.
17. Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K.I. Machine Learning for Catalysis Informatics: Recent Applications and Prospects, 2020. doi:10.1021/acscatal.9b04186.
18. Chen, Y.Y.; Ross Kunz, M.; He, X.; Fushimi, R. Recent progress toward catalyst properties, performance, and prediction with data-driven methods, 2022. doi:10.1016/j.coche.2022.100843.
19. Moses, O.A.; Chen, W.; Adam, M.L.; Wang, Z.; Liu, K.; Shao, J.; Li, Z.; Li, W.; Wang, C.; Zhao, H.; Pang, C.H.; Yin, Z.; Yu, X. Integration of data-intensive, machine learning and robotic experimental approaches for accelerated discovery of catalysts in renewable energy-related reactions, 2021. doi:10.1016/j.matre.2021.100049.
20. Nishimura, S.; Li, X.; Ohyama, J.; Takahashi, K. Leveraging machine learning engineering to uncover insights into heterogeneous catalyst design for oxidative coupling of methane. *Catalysis Science & Technology* **2023**. doi:10.1039/d3cy00596h.
21. Takahashi, K.; Ohyama, J.; Nishimura, S.; Fujima, J.; Takahashi, L.; Uno, T.; Taniike, T. Catalysts informatics: paradigm shift towards data-driven catalyst design. *Chemical Communications* **2023**, *59*, 2222–2238. doi:10.1039/d2cc05938j.
22. Chen, K.; Tian, H.; Li, B.; Rangarajan, S. A chemistry-inspired neural network kinetic model for oxidative coupling of methane from high-throughput data. *AIChE Journal* **2022**, pp. 1–11. doi:10.1002/aic.17584.
23. Catalyst Acquisition by Data Science (CADS) homepage.
24. Nguyen, T.N.; Nakanowatari, S.; Tran, T.P.N.; Thakur, A.; Takahashi, L.; Takahashi, K.; Taniike, T. Learning Catalyst Design Based on Bias-Free Data Set for Oxidative Coupling of Methane. *ACS Catalysis* **2021**, *11*, 1797–1809. doi:10.1021/acscatal.0c04629.

25. Segal, M.R. Machine learning benchmarks and random forest regression **2004**.
26. Ziu, K.; Solozabal, R.; Rangarajan, S.; Takáč, M. A deep neural network for oxidative coupling of methane trained on high-throughput experimental data. *Journal of Physics: Energy* **2022**, *5*, 014009. doi:10.1088/2515-7655/aca797.
27. Daneshpayeh, M.; Khodadadi, A.; Mostoufi, N.; Mortazavi, Y.; Sotudeh-Gharebagh, R.; Talebizadeh, A. Kinetic modeling of oxidative coupling of methane over Mn/Na₂WO₄/SiO₂ catalyst. *Fuel Processing Technology* **2009**, *90*, 403–410. doi:10.1016/j.fuproc.2008.11.001.
28. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **2002**, *6*, 182–197. doi:10.1109/4235.996017.
29. Fortin, F.A.; De Rainville, F.M.; Gardner, M.A.; Parizeau, M.; Gagné, C. DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research* **2012**, *13*, 2171–2175.
30. Dewancker, I.; McCourt, M.; Clark, S. Bayesian Optimization for Machine Learning : A Practical Guidebook, 2016, [arXiv:cs.LG/1612.04858].
31. Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. *npj Computational Materials* **2023**, *9*, 42. doi:10.1038/s41524-023-01000-z.
32. Grömping, U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician* **2009**, *63*, 308–319, [https://doi.org/10.1198/tast.2009.08199]. doi:10.1198/tast.2009.08199.
33. Kiani, D.; Sourav, S.; Baltrusaitis, J.; Wachs, I.E. Oxidative Coupling of Methane (OCM) by SiO₂-Supported Tungsten Oxide Catalysts Promoted with Mn and Na. *ACS Catalysis* **2019**, *9*, 5912–5928. doi:10.1021/acscatal.9b01585.
34. Hu, L.; Pinto, D.; Urakawa, A. Catalytic Oxidative Coupling of Methane: Heterogeneous or Homogeneous Reaction? *ACS Sustainable Chemistry & Engineering* **2023**, *11*, 10835–10844. doi:10.1021/acssuschemeng.3c02088.
35. Zavyalova, U.; Holena, M.; Schlögl, R.; Baerns, M. Statistical Analysis of Past Catalytic Data on Oxidative Methane Coupling for New Insights into the Composition of High-Performance Catalysts. *ChemCatChem* **2011**, *3*, 1935–1947, [https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/cctc.201100186]. doi:https://doi.org/10.1002/cctc.201100186.
36. Ortiz-Bravo, C.A.; Chagas, C.A.; Toniolo, F.S. Oxidative coupling of methane (OCM): An overview of the challenges and opportunities for developing new technologies. *Journal of Natural Gas Science and Engineering* **2021**, *96*, 104254.
37. Amenomiya, Y.; Birss, V.I.; Golezdzinowski, M.; Galuszka, J.; Sanger, A.R. Conversion of methane by oxidative coupling. *Catalysis Reviews—Science and Engineering* **1990**, *32*, 163–227.
38. Yildiz, M.; Simon, U.; Otremba, T.; Aksu, Y.; Kailasam, K.; Thomas, A.; Schomäcker, R.; Arndt, S. Support material variation for the Mn_xO_y-Na₂WO₄/SiO₂ catalyst. *Catalysis Today* **2014**, *228*, 5–14. Natural Gas Conversion the Status and Potentials in the Light of NGCS-10, doi:https://doi.org/10.1016/j.cattod.2013.12.024.
39. fu Ji, S.; cun Xiao, T.; ben Li, S.; zhi Xu, C.; ling Hou, R.; Coleman, K.S.; Green, M.L. The relationship between the structure and the performance of Na-W-Mn/SiO₂ catalysts for the oxidative coupling of methane. *Applied Catalysis A: General* **2002**, *225*, 271–284. doi:https://doi.org/10.1016/S0926-860X(01)00864-X.
40. Airedy, D.R.; Roy, A.; Cullen, D.A.; Ding, K. TiO_x-supported Na-Mn-W oxides for the oxidative coupling of methane. *Catalysis Today* **2023**, *416*, 113977. SI:Natural gas catalysis, doi:https://doi.org/10.1016/j.cattod.2022.12.008.
41. Gu, S.; Kang, J.; Lee, T.; Shim, J.; Choi, J.W.; Suh, D.J.; Lee, H.; Yoo, C.; Baik, H.; Choi, J.; Ha, J.M. Na₂WO₄/Mn supported on all-silica delaminated zeolite for the optimal oxidative coupling of methane via the effective stabilization of tetrahedral WO₄: Elucidating effects of support precursors with different crystal structures, Al-addition, and morphologies. *Chemical Engineering Journal* **2023**, *457*, 141057. doi:https://doi.org/10.1016/j.cej.2022.141057.
42. Hayek, N.S.; Lucas, N.S.; Warwar Damouny, C.; Gazit, O.M. Critical Surface Parameters for the Oxidative Coupling of Methane over the Mn–Na–W/SiO₂ Catalyst. *ACS Applied Materials & Interfaces* **2017**, *9*, 40404–40411. doi:10.1021/acsaami.7b14941.

43. Arndt, S.; Otremba, T.; Simon, U.; Yildiz, M.; Schubert, H.; Schomäcker, R. Mn–Na₂WO₄/SiO₂ as catalyst for the oxidative coupling of methane. What is really known? *Applied Catalysis A: General* **2012**, *425–426*, 53–61. doi:<https://doi.org/10.1016/j.apcata.2012.02.046>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.