

ARTICLE

<https://doi.org/10.1038/s41467-019-08325-8>

OPEN

A meta-analysis of catalytic literature data reveals property-performance correlations for the OCM reaction

Roman Schmack¹, Alexandra Friedrich², Evgenii V. Kondratenko³, Jörg Polte⁴, Axel Werwatz² & Ralph Kraehnert¹

Decades of catalysis research have created vast amounts of experimental data. Within these data, new insights into property-performance correlations are hidden. However, the incomplete nature and undefined structure of the data has so far prevented comprehensive knowledge extraction. We propose a meta-analysis method that identifies correlations between a catalyst's physico-chemical properties and its performance in a particular reaction. The method unites literature data with textbook knowledge and statistical tools. Starting from a researcher's chemical intuition, a hypothesis is formulated and tested against the data for statistical significance. Iterative hypothesis refinement yields simple, robust and interpretable chemical models. The derived insights can guide new fundamental research and the discovery of improved catalysts. We demonstrate and validate the method for the oxidative coupling of methane (OCM). The final model indicates that only well-performing catalysts provide under reaction conditions two independent functionalities, i.e. a thermodynamically stable carbonate and a thermally stable oxide support.

¹Technische Universität Berlin, Institut für Chemie, Str. des 17. Juni 124, 10623 Berlin, Germany. ²Technische Universität Berlin, Institut für Volkswirtschaftslehre und Wirtschaftsrecht, FG Ökonometrie und Wirtschaftsstatistik, Straße des 17. Juni 135, 10623 Berlin, Germany. ³Leibniz Institute for Catalysis (LIKAT Rostock), Albert-Einstein-Str. 29 a, 18059 Rostock, Germany. ⁴Humboldt-Universität zu Berlin, Institut für Chemie, Brook-Taylor-Straße 2, 12489 Berlin, Germany. Correspondence and requests for materials should be addressed to R.K. (email: ralph.kraehnert@tu-berlin.de)

Meta-analysis¹ is a powerful tool to rigorously assess the findings of published research. Successful meta-analysis studies were reported in research fields where quantitative analysis of independently conducted experiments is prevalent, e.g., medical research², genetics^{3,4}, biology⁵, and economics⁶. Particularly in medical research, meta-analysis is used to aggregate individual studies aiming at the same treatment effect and employing the same research design (randomized control trials). In this case, the main goal and benefit of meta-analysis is to obtain a more precise and robust effect estimate than any individual study can deliver.

Despite the existence of a vast amount of well-documented experiments also in heterogeneous catalysis, hardly any corresponding meta-analyses have been reported. Efforts to collect and analyze sets of literature data have been reported primarily for the water gas shift reaction (WGS)⁷, CO-oxidation^{8–10}, transesterification in biodiesel production¹¹, and electro-catalytic oxidation of alcohols in direct alcohol fuel cells¹². However, neither did these studies identify new structure–activity relationships nor could they provide simple chemical explanations for the observed statistical effects.

The shortcomings of these reports can be attributed, e.g., to (1) rather small datasets being used (data from 85 publications or less^{7–12}), (2) statistical learning methods that search exploratively for relationships in the data, and (3) failure to incorporate existing chemical knowledge to inform and direct the statistical work. Another challenge faced in heterogeneous catalysis results (4) from the unsystematic heterogeneity of the available data. Each experimental report typically explores a narrow range of catalyst compositions and reaction conditions. However, the employed reaction conditions vary widely between different publications, which influences the catalytic performance via chemical kinetics, and makes a meta-analysis challenging.

A significant break-through in terms of the dataset quality was reported by Zavyalova et al. in 2011¹³ and in a recent follow up¹⁴ (see Supplementary Notes 1 and 2 for details). The dataset of Zavyalova et al.¹⁵ was made freely available and covered 1866 distinct catalyst compositions collected from 421 reports. The dataset provides for the oxidative coupling of methane (OCM) information on catalyst composition, reaction conditions and catalyst performance. The study employed formal mathematical approaches (multiway ANOVA, Pearson's and Spearman's correlation coefficients, regression trees) to search for correlations between catalyst composition and catalyst performance. The study concluded that some combinations of key elements contained in a catalyst can contribute to high C₂ selectivity or yield, but did not result in a general model that correlates catalyst composition to catalytic performance with proven statistical significance and robustness. Unfortunately, a simple correlation between elemental composition and performance is also not likely to exist, since materials with the same nominal composition can often adopt very different structures and surface terminations that should result in different OCM performances. Hence, descriptors other than just elemental composition are needed.

We report a meta-analysis method that can identify statistically significant correlations between the physico-chemical properties assigned to a catalyst material and its performance in a given reaction. The method employs three distinct sources of information to achieve this goal. It incorporates in addition to (i) the experimental data reported in literature also (ii) general textbook knowledge about fundamental material properties and (iii) the experienced intuition of a chemist or material scientist about possible property–performance correlations. The OCM is used here as an example to illustrate the approach and derive property–performance correlations.

Results

Meta-analysis approach. The method is schematically outlined in Fig. 1. It starts out from the chemist's intuition expressed as a hypothesis (iii) about a supposed relationship between the properties of a catalyst material and its catalytic performance. In a first step, data on the composition, reaction conditions and performance of different catalysts are assembled from the body of literature (i) into a dataset (1) for this reaction. Then, related textbook knowledge (ii) on all relevant elements is collected in the form of tables that provide, e.g., the ability of an element to form certain types of chemical compounds (e.g., oxides, carbonates), their thermal stability, formation enthalpies, and melting points. Moreover, descriptor rules (2) are derived that define how to calculate the so-called physico-chemical descriptors that are required to express the hypothesis (iii). Applying the descriptor rules to each catalyst entry in the dataset creates an extended dataset (3) that includes for each catalyst additional physico-chemical property descriptors. One particular strength of this procedure is the fact that the descriptors can be computed as a function of temperature and pressure, i.e., at conditions that closely reflect the individual conditions at which a catalyst was experimentally tested in the original experimental report. Hence, the computed properties closely reflect the state of the catalyst during the actual performance testing.

The chemical hypothesis (iii) can now be translated into a set of formal sorting rules (4). These sorting rules express in a formal way, which descriptors are assumed to correlate with high catalytic performance. More precisely, a sorting rule states that a catalyst will be assigned to one specific property group (5) if the catalyst's descriptors possess values within a defined range. Applying the rules to the extended dataset (3) divides the data into smaller subsets of so-called property groups (5). Then, the measured performance value of each catalyst is used to compute for each property group a density distribution of the performance indicator (6), its averaged value as well as a standard deviation. If the descriptors chosen in step 4 represent properties that are indeed relevant to the catalysts performance, then the performance distributions (6) should differ significantly between the corresponding property groups. However, these performance distributions do not account yet for the fact that each catalyst was measured at different reaction conditions.

In order to compensate for these differences in reaction conditions, a multivariate regression analysis (7) is invoked that approximates the influence of, e.g., temperature (T) and the ratio of feed gases (here: $p_{\text{CH}_4}/p_{\text{O}_2}$). Moreover, it considers to which of two compared property groups a given catalyst had been assigned. The regression procedure adjusts all regression coefficients β_i in order to minimize the deviation between experimental and computed yield values over all entries in the two groups. The computed regression results (8) quantify the influence of temperature (β_T), pressure ratio ($\beta_{\text{CH}_4/\text{O}_2}$), and most importantly, the effect of belonging to either one of the two compared catalyst groups (β_{group}) on the catalysts performance (Y_{C_2}). The results significance is judged via a t -test that calculates for each β -regression coefficient a corresponding p -value. Low p -values ($p < 0.05$) indicate a high statistical significance, with, e.g., $p < 0.05$ implying a confidence exceeding 95% that two compared catalyst groups are indeed distinctly different in their performance (All p -values reported in this paper were obtained via t -test, see method section and Supplementary Note 8 for further details.).

In step (9), the obtained models are compared to additional spectroscopic, computational and fundamental evidence from literature. An iterative refinement of hypotheses (iii), descriptors (3), sorting rules (4) and if possible also the studied data (2) results in robust and statistically significant property–performance models.

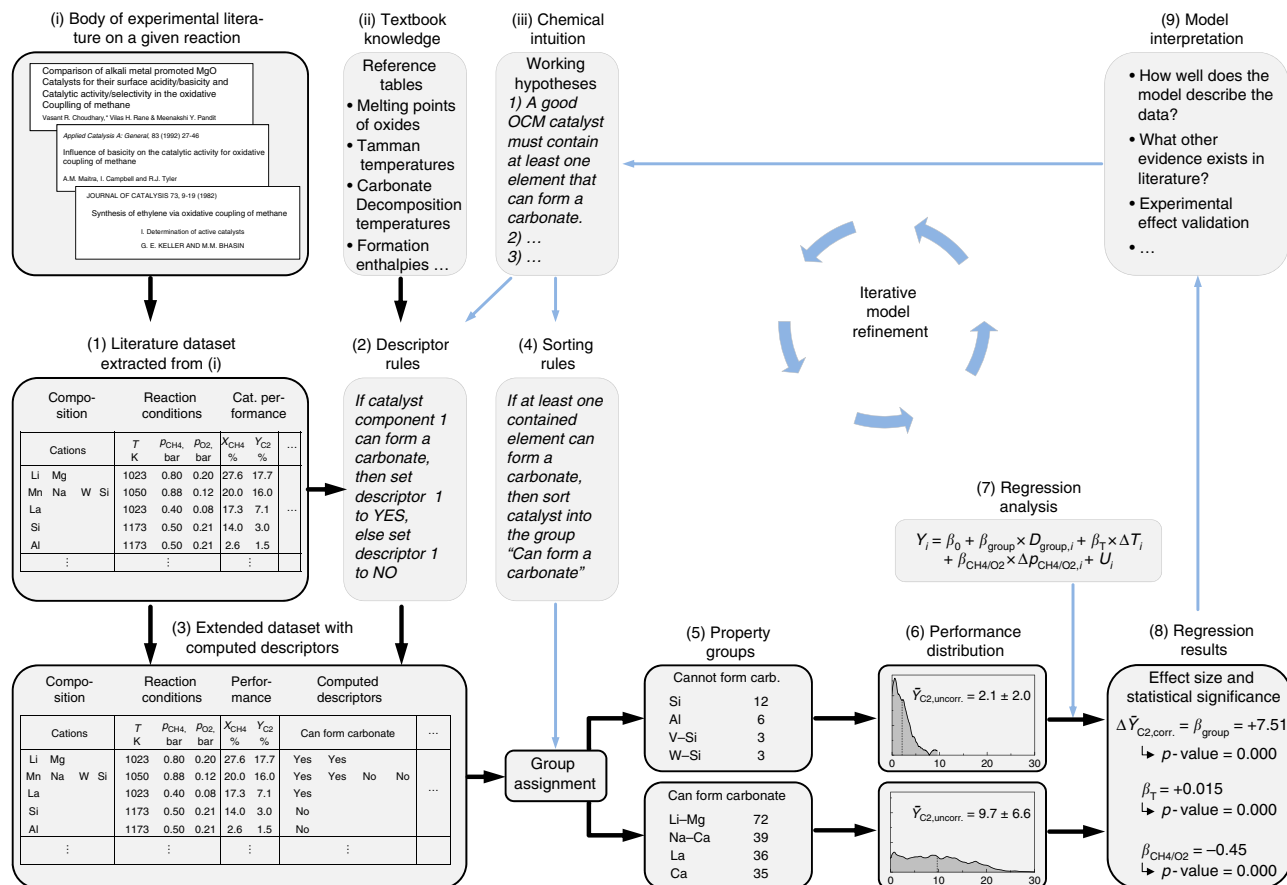


Fig. 1 Proposed meta-analysis approach. Starting out with (i) experimental catalytic data reported in literature, (ii) general textbook knowledge and (iii) a researchers chemical intuition a working hypothesis is formulated, processed, tested against the data for its statistical significance, and iteratively refined into a property-performance model (steps 1-9). Bold arrows mark the flow of the data. All specific examples relate to the studied OCM reaction and hypothesis (1). (See Method section and Supplementary Notes 5-9 for further details)

Literature OCM dataset. The oxidative coupling of methane aims to convert methane and oxygen selectively into C_2 coupling products (ethane, ethylene) while avoiding total oxidation¹⁶. The OCM reaction is catalyzed by numerous materials with very diverse elemental compositions¹⁷. Economic viability of the process would require C_2 yields well above the best values (ca. 30%) reported so far^{17,18}. Despite decades of research, the reaction mechanism is not fully understood even for extensively studied catalyst such as Li/MgO and Mn-Na₂WO₄/SiO₂^{19,20}. It is generally presumed that the reaction network involves gas-phase reactions coupled to reactions at the catalyst surface²¹⁻²³, where surface defects facilitate C-H bond activation. Unfortunately, the required high reaction temperatures (500 to 1000 °C)²⁰ obstruct many common in situ analytical tools. Hence, definite proof exists so far neither for the relevance of such defects under practical operation conditions nor the critical contribution of gas-phase reactions. Moreover, no generally valid correlation between a catalyst's composition, its structure and its OCM performance has been established yet.

We illustrate the developed meta-analysis approach in a complete re-analysis of the OCM data compiled by Zavyalova et al.¹³ after corrections and outlier removal from the data (see Supplementary Note 3). The data is represented by a table that contains 1802 rows (different catalyst compositions) and 37 columns (see Supplementary Note 4 and Supplementary Data 1). The columns encode for each catalyst (a) the elements contained in the catalyst and their molar fractions, (b) parameters indicating the reaction conditions (T , p_{CH_4} , p_{O_2} , p_{CH_4}/p_{O_2} , p_{total} , contact

time) and (c) numerical values that describe catalyst performance (X_{O_2} , X_{CH_4} , S_{CO_x} , S_{ethane} , S_{ethene} , S_{C_2} , Y_{C_2}). Out of these categories, elemental composition, reaction temperature (T), reactant partial pressure ratio (p_{CH_4}/p_{O_2}) and C_2 yield (Y_{C_2}) were evaluated. Figure 2 displays for the corrected dataset exemplarily Y_{C_2} as a function of (a) reaction temperature, (b) p_{CH_4}/p_{O_2} , and (c) contact time to illustrate the covered data range and the data heterogeneity. (See also Supplementary Note 20 for individual catalyst examples Li-Mg and Mn-Na-W-Si.)

Derived final model. Numerous hypotheses were tested. The derived best model employs four different hypotheses to divide the OCM dataset into a total of 18 different property groups as illustrated by the tree-like representation shown in Fig. 3. Each displayed box represents a group of N catalysts. Starting out from the complete dataset (group 0, $n = 1802$), the physico-chemical criteria related to hypothesis (1) divide the contained catalysts into two subgroups (groups 1a, 1b). The subsequently applied hypotheses (2), (3), and (4) divide these subgroups further, resulting in 10 terminal subgroups 1a, 4a...4h, and 2d. For each terminal subgroup, the respective five most frequent catalyst compositions are listed in Fig. 3 (see Supplementary Note 10 for a complete list of the catalysts contained in each group). Bold box frames indicate on each level the catalyst group that shows the best catalytic performance in terms of average C_2 yield.

Figure 4a displays in a similar representation the most important results of the regression analysis and statistical

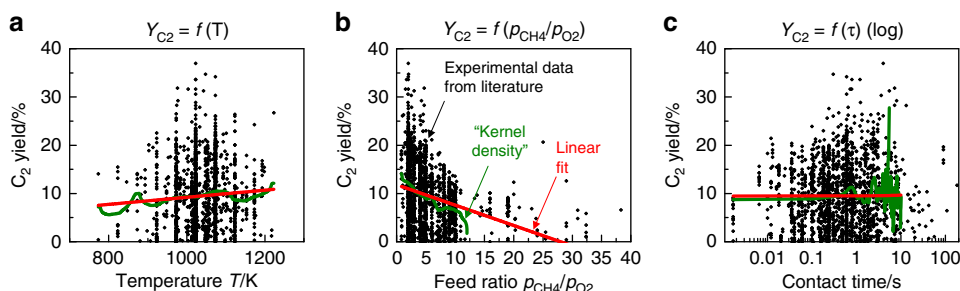


Fig. 2 Distribution of data in the corrected OCM dataset (1802 data points). The data are plotted in terms of C_2 yield **a** as a function of temperature (T), **b** pressure ratio methane/oxygen in the feed gas (p_{CH_4}/p_{O_2}) and **c** contact time (τ , log). Each point represents one individual catalyst. Red lines indicate a linear fit. Green lines represent a so-called Epanechnikov kernel density function

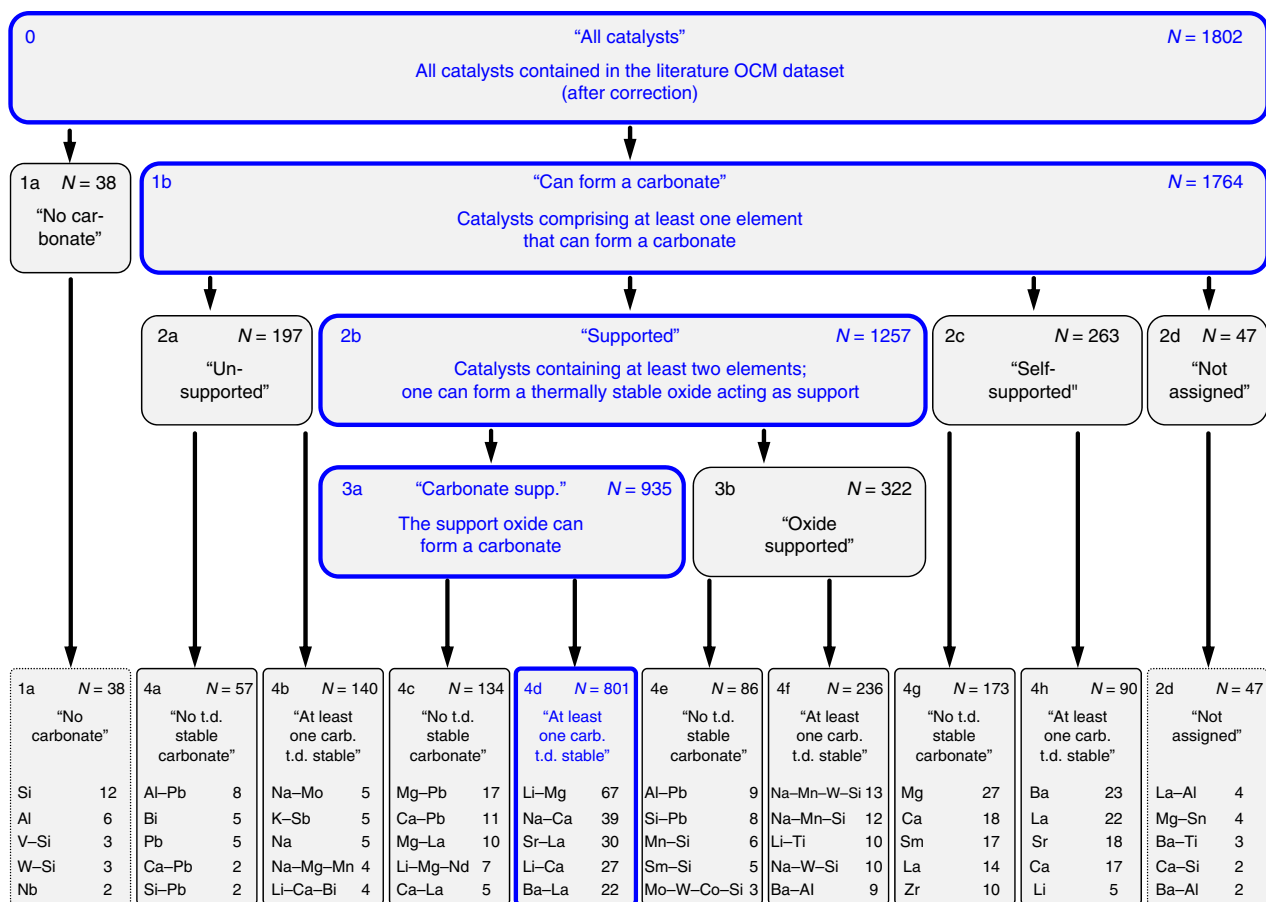


Fig. 3 Tree representation of the best derived OCM property-performance model. Each box represents a distinct property subgroup obtained from the applied sorting criteria. Starting out from the initial dataset (group 0), each catalyst is assigned to one out of ten terminal subgroups (1a; 4a-4h; 2d). For each terminal subgroup also the five most frequent catalyst compositions (i.e., cation combinations) and the corresponding number of observations are listed. N denotes the number of catalysts in each group. Blue colored bold frames mark the group with the highest averaged C_2 yield on each hypothesis level. "t.d." and "carb." refer to "thermodynamically" and "carbonate", respectively. A full list of catalyst compositions (i.e., cation combinations) is provided in Supplementary Data 5

evaluation that produced the final model displayed in Fig. 3. For each group, a diagram is provided that illustrates the C_2 yield distribution within the group along with the mean C_2 yield ($\bar{Y}_{C_2, \text{uncorr}}$) marked by a dashed line, i.e., Y_{C_2} prior to regression analysis) and its standard deviation. The horizontal arrows that connect two respective boxes represent the performed statistical tests. Figure 4b summarizes the final model outcome in terms of performance ($\bar{Y}_{C_2, \text{corr}}$) for each terminal group. All observed p -values are lower than 0.005, which underlines the extremely high statistical significance of the derived final model.

Hypothesis 1—ability to form at least one carbonate compound: An initial assessment of the dataset gave the impression that many OCM catalysts contain components that can form a carbonate. Hypothesis 1 therefore proposes that a good OCM catalyst contains at least one chemical element that is known to form a carbonate compound.

All catalysts that can form at least one carbonate were assigned to group 1b "Can form a carbonate", all other catalysts to group 1a "Cannot form a carbonate" (Fig. 3). Only a small fraction of catalysts ($n = 38$) is not able to form any carbonate, whereas the

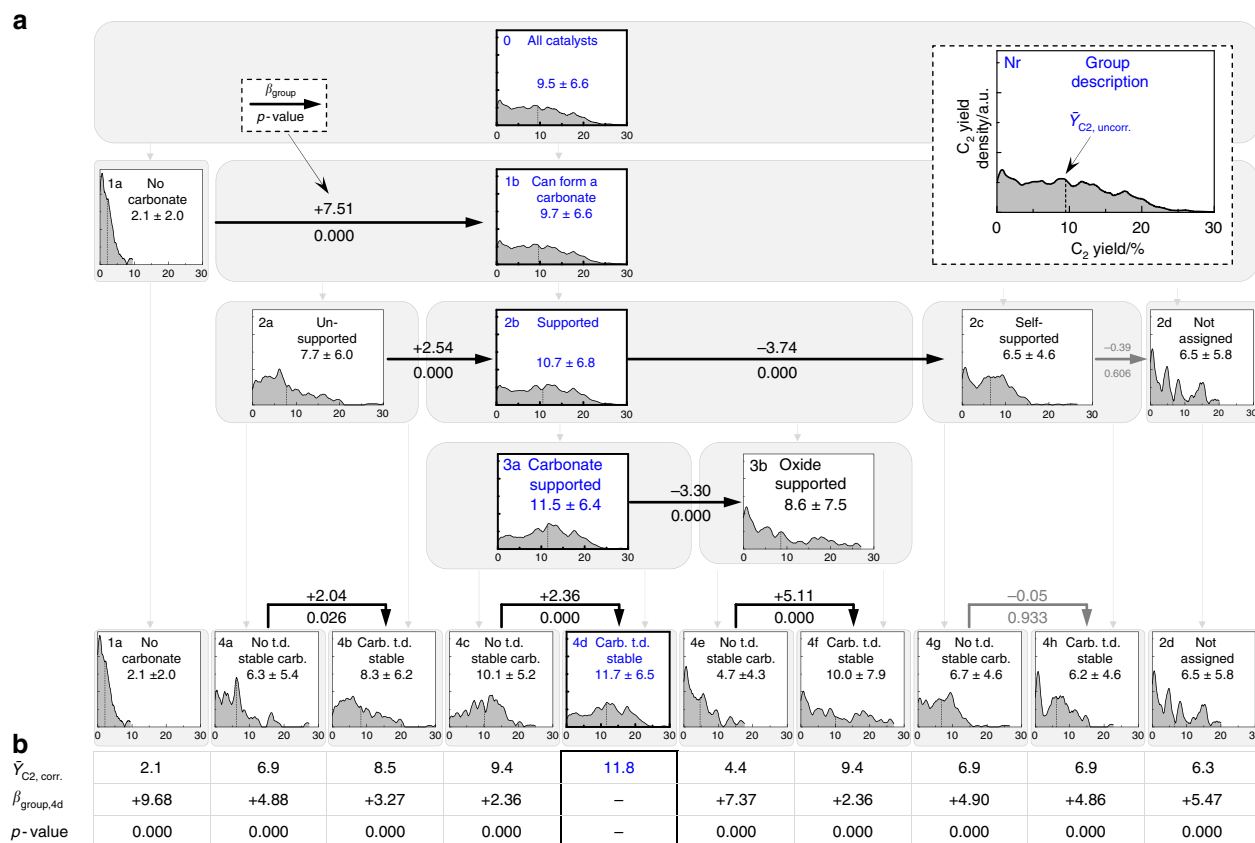


Fig. 4 Statistical evaluation for the best derived property–performance model. **a** Tree representation of the model with the same property groups as shown in Fig. 3. Diagrams display for each property group the C_2 yield–density distribution, the mean C_2 yield $\bar{Y}_{C_2, uncorr}$, and its standard deviation (marked by \pm). The arrow labels provide exemplarily the regression results, i.e., the value of β_{group} derived from the regression analysis, which corresponds to a difference in mean yield ($-\Delta\bar{Y}_{C_2, corr}$) between the two groups after correction for the influence of temperature T and pressure ratio p_{CH_4}/p_{O_2} . Moreover, the p -value corresponding to β_{group} is displayed, where smaller p -values represent a higher statistical significance. Blue color marks on each hypothesis level the group with the highest mean C_2 yield. “t.d.” and “carb.” refer to “thermodynamically” and “carbonate”, respectively. **b** Final model outcome in terms of performance ($\bar{Y}_{C_2, corr}$) for each terminal group. Regression coefficient β_{group} and p -values are reported for the comparison of the best-performing group 4d with each other terminal subgroup. Note that for reasons of clarity only a fraction of the performed tests is shown. The complete model contains each possible pairwise comparison among all formed subgroups. The complete list of all obtained regression coefficients (β_{group} , β_T , β_{CH_4/O_2}) and the respective p -values are provided in Supplementary Note 11a/b/c)

vast majority of reported catalysts can form at least one carbonate (1b: $n = 1764$) (Fig. 3).

Catalysts that can form at least one carbonate (1b) show on average a 7.6 % higher yield of C_2 hydrocarbons ($\bar{Y}_{C_2, uncorr} = 9.7$ %) than catalysts in group 1a ($\bar{Y}_{C_2, uncorr} = 2.1$ %). The difference in C_2 yield computed via multiple regression analysis ($\Delta\bar{Y}_{C_2, corr} = \beta_{group} = +7.51$) proves to be statistically highly significant on a >99.5 % confidence level ($p < 0.005$). Also the regression coefficients obtained for temperature ($\beta_T = +0.0015$; $p = 0.000$) and pressure ratio p_{CH_4}/p_{O_2} ($\beta_{CH_4/O_2} = -0.452$; $p = 0.000$) show high statistical significance with higher C_2 yields corresponding to higher temperatures and higher oxygen partial pressures (see Supplementary Note 9 for a discussion and interpretation).

Group 1a contains mostly Si, Al, V, W, Nb, Sn, Pt, Ti, and combinations thereof (Fig. 3), i.e., elements typically not associated with high OCM activity. In contrast, group 1b includes all the usual suspects of high OCM performance, i.e., the Li/MgO system (Li–Mg), catalysts based on lanthanum oxide and calcium oxide, as well as variations of the Mn– Na_2WO_4/SiO_2 system (see Supplementary Note 10 for full list). Hence, hypothesis 1 is able to sort catalysts that are not very active (silica, alumina) or combust methane unselectively (vanadium on silica, Pt) into

group 1a using an objective chemical criterion. Thus, statistical noise is removed from the remaining set of catalysts (1b).

Hypothesis 2—presence of a stable oxide acting as a support: Many reported OCM catalysts feature two components. One of the components often appears to be able to form a thermally stable metal oxide. Such metal oxides are frequently used in heterogeneous catalyst to support an active phase. Hypothesis 2 therefore defines two criteria that attempt to describe a property “catalyst support”. Criterion (i) states that among all the oxides that can possibly be formed by a catalyst’s cations, at least one oxide is present that possesses a Tammann temperature higher than the individual temperature at which the catalyst was tested in OCM. An additional criterion (ii) imposes that the mass content of this stable oxide amounts to at least 50 wt% of the catalyst’s constituents, assuming that they are present in the form of their respective most stable oxides. An additional criterion (iii) used for the group assignment is the presence of an element other than the designated support oxide, and its ability to form a carbonate.

A catalyst was assigned to group 2a “unsupported” ($n = 197$) if none of a catalyst’s components fulfilled both support criteria (i) and (ii) (Fig. 3). Group 2b “supported” ($n = 1257$) represents all catalysts where one component fulfils both support criteria (i, ii),

and where an additional cation is present that is able to form a carbonate. Catalysts were assigned to group 2c “self supported” ($n = 263$) when a catalyst contained only one component, and if this component satisfied both support criteria AND could form a carbonate. All remaining catalysts were assigned to group 2d “not assigned” ($n = 47$) for further noise reduction.

Catalysts in group 2b “supported” showed higher C_2 yields ($\bar{Y}_{C_2, \text{uncorr}} = 10.7\%$) than catalysts in the group 2a “unsupported” ($\bar{Y}_{C_2, \text{uncorr}} = 7.7\%$) and group 2c “self-supported” ($\bar{Y}_{C_2, \text{uncorr}} = 6.5\%$). The differences are statistically highly significant.

The best-performing group (2b “supported”) features the highest number of catalyst entries ($n = 1257$) and contains most of the classical OCM catalysts (e.g., Li/MgO, Mn–Na₂WO₄/SiO₂). Most group entries consist of a binary combination of La or alkali metals with an alkaline earth element (Li–Mg, Na–Ca, Li–Ca). Typically, the support criteria are satisfied by the thermally stable oxide of the alkaline earth element. Many of the cation combinations that include heavy metals (Pb, Bi) and alkali elements (Li, Na, K, Cs) are assigned to group 2a “unsupported” because their respective (super)oxides fail the thermal stability criterion (i), or because the stable oxide is not present in sufficiently large amounts (ii). The “self-supported” catalysts 2c contain only a single cation, which is either a lanthanide or an alkaline earth element, and forms a thermally stable oxide. The 47 catalysts placed in group 2d “not assigned” comprise mostly binary combinations that contain a large amount of a stable support oxide, but fail criterion (iii) because the second element (that is not the designated support) cannot form a carbonate.

Hypothesis 3—ability of the support oxide to form a carbonate: Hypotheses 1 and 2 suggest that the ability to form a carbonate plays a critical role in OCM catalysis. All catalysts retained in group 2b contain at least one element that can form a carbonate, and another element that can form a thermally stable oxide. Hypothesis 3 tests, if also for this support oxide the ability to form a carbonate impacts a catalyst’s OCM performance.

All catalysts for which the designated support oxide can form a carbonate were assigned to group 3a “carbonate-supported” ($n = 935$), all other catalysts to group 3b “oxide-supported” ($n = 322$). Most of the “carbonate supported” catalysts consist of La or an alkali element combined with an alkaline earth element where typically the alkaline earth elements provide the support functionality. The “oxide-supported” catalysts (the designated support oxide cannot form a carbonate) comprise mostly alkali and alkaline earth elements or Na₂WO₄ supported on either alumina, silica or titania (see element combinations listed in groups 4e and 4f in Fig. 3).

The carbonate-supported catalysts (group 3a, $\bar{Y}_{C_2, \text{uncorr}} = 11.5\%$) clearly outperform the oxide-supported group 3b (8.6%) by about 2.9% C_2 yield ($\beta_{\text{group}} = -3.30$) and with high statistical significance ($p = 0.000$), providing further evidence for the beneficial and important role of carbonates in OCM catalysis.

Hypothesis 4—thermodynamic stability of carbonates during OCM: A property often attributed to good OCM catalysts is their so-called basicity^{24–27}. One possible measure for the basicity of a metal oxide is its ability to form a carbonate and the carbonate’s thermal stability. If such a carbonate is thermodynamically stable also at the high temperatures of OCM, then the carbonates are likely to be present on the catalyst surface during the OCM reaction.

Hypothesis 4 therefore assesses the thermodynamic stability of all carbonates that can be formed by the elements present in a catalyst by comparing the decomposition temperature of the most stable carbonate to the temperature at which the OCM performance was measured. A catalyst is assigned to a subgroup “carbonate thermodynamically stable” if the OCM test was performed at a reaction temperature that is lower

than the temperature at which the most stable contained carbonate decomposes (corrected for a constant offset that accounts for experimental uncertainties and dependence of carbonate decomposition equilibria on CO₂ partial pressure, here: $T_{\text{OCM measurement}} < T_{\text{carbonate decomposition}} + 100 \text{ K}$; for other offsets see Supplementary Note 14h-k).

Hypothesis 4 was applied to all subgroups that had been derived via positive sorting criteria in the previous hypotheses (i.e., not to the “not-groups” 1a and 2d). It divides the groups 2a, 3a, 3b, and 2c into the respective subgroups 4a, c, e, g (“no thermodynamically stable carbonate”), and subgroups 4b, d, f, h (“at least one thermodynamically stable carbonate”) (Fig. 3). Three out of four corresponding comparisons, i.e., 4a/4b, 4c/4d, 4e/4f, show a strong statistical significance indicated by p -values being equal or lower than 0.026 (Fig. 4a). In all three statistically significant cases the catalysts that can form at least one thermodynamically (“t.d.”) stable carbonate outperform the corresponding catalysts that cannot form a stable carbonate, with C_2 yields $\bar{Y}_{C_2, \text{corr}}$ being higher by 2.04% (4a/4b), 2.36% (4c/4d), and 5.11% (4e/4f) (see β_{group} values in Fig. 4a).

Despite the fact that the two most common OCM catalyst types (Li/MgO, Mn–Na₂WO₄/SiO₂) were assigned to different groups on level 3 (Li–Mg: 3a; Na–Mn–W–Si: 3b), both end up in the respective subgroups that can form at least one t.d. stable carbonate (Li–Mg: 4d; Na–Mn–W–Si: 4f). In general, the role of the thermodynamically most stable carbonate is typically fulfilled by alkaline, alkaline earth, or lanthanide elements in the subgroups 4b/d/f/h. In absence of these elements no thermodynamically stable carbonate can be formed and lower yields are observed (see Supplementary Note 13 for a more detailed discussion).

Hypothesis 4 thus illustrates another strength of the method: the same sorting criteria can be applied to different groups, in this case groups derived on hypothesis levels 2 and 3. Consistent results obtained across different groups point towards a general validity of the proposed effect.

Model robustness. The robustness of the final model (Figs. 3 and 4) was assessed using different approaches of parameter variation, changing the independent variable to $\ln(Y_{C_2})$ and with a so-called “robust regression” (see Supplementary Notes 14–18 for details). The model retained the same statistically significant correlations, except for few instances where some correlations on level 4 lost their statistical significance. The final model is therefore highly robust despite the data heterogeneity, potential publication bias within the data and the small size of some of the formed terminal groups.

Discussion

The derived model reveals clear property–performance correlations. It establishes for the first time a generalized set of physico-chemical properties that discriminate between high- and low-performing OCM catalysts. Four simple hypotheses suffice to divide 1802 catalysts into 10 groups of distinct properties and OCM performance (Fig. 4b). Following the path of the best-performing groups on each hypothesis level indicates a combination of properties that correlates with high C_2 yields. The model predicts the highest C_2 yield (4d, $\bar{Y}_{C_2, \text{corr}} = 11.8\%$) when (1b) at least one of the elements contained in the catalyst can form a carbonate, (2b) the catalyst contains an additional element that forms a thermally stable oxide, (3a) this oxide is also able to form a carbonate, and (4d) at least one of the possible carbonates is thermodynamically stable at the applied OCM reaction temperature. Also a comparison of the best-performing terminal group 4d with each other terminal group (1a, 4a–h, 2d) in terms

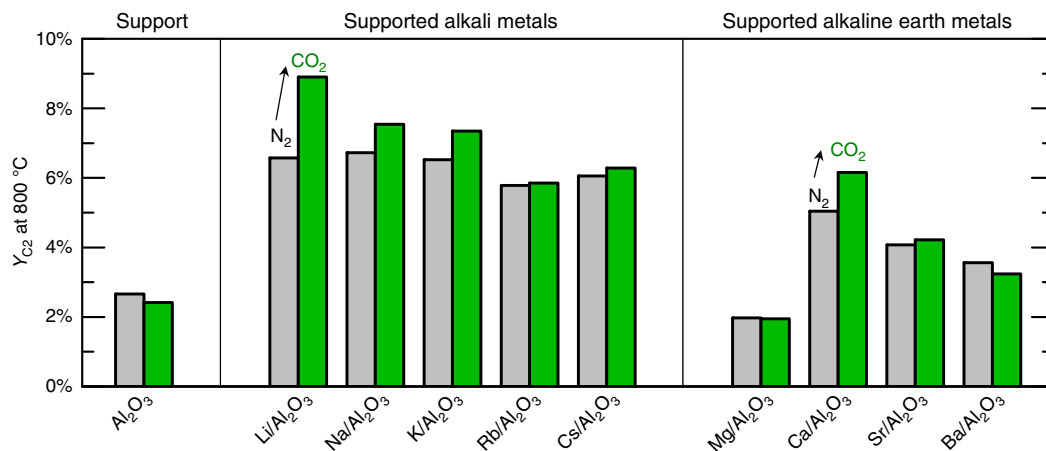


Fig. 5 Experimental verification of the influence of CO_2 co-feed in OCM. The effect of adding CO_2 to the OCM feed gas was measured for nine different alumina-supported catalysts as well as the bare alumina support. The yield of C_2 hydrocarbons (ethane plus ethylene) measured at 800°C is plotted for two different feed conditions, i.e. with CO_2 co-feed (green bars, $\text{CH}_4/\text{O}_2/\text{N}_2/\text{CO}_2 = 26.2:14.8:3.0:56.0$ and without CO_2 co-feed (gray bars, $\text{CH}_4/\text{O}_2/\text{N}_2 = 26.2:14.8:59.0$). Within the limits of experimental accuracy, CO_2 addition either increases Y_{C_2} , or shows a negligible effect for all studied catalyst. All catalysts were prepared and measured under identical conditions, changing only the precursors employed to support the different active metals

of yield difference (β_{group}) and p -value shows a very high statistical significance ($p < 0.005$).

The nature of the hypothesis formulation in terms of chemical conditions facilitates a direct physico-chemical interpretation. Good OCM catalysts comprise at least two elements, with one of the elements being able to form a carbonate at the temperatures of OCM reaction, and a second element that can form a thermally stable (non-sintering) oxide under OCM conditions (hypothesis 2). Good OCM catalysts thus provide two functionalities, i.e., a “support” and an “active phase”. The thermally stable oxide (alkaline earth and lanthanide oxides, alumina, silica, titania) acts as support and provides a high surface area during OCM (2b). The active phase appears to be related to the ability to form a carbonate. The ability to form a carbonate²⁴ is one out of several possible descriptors for basicity (Smith scale²⁸; optical basicity²⁹, an ionic-covalent parameter³⁰). Basicity has been proposed by many authors as an important feature of OCM catalysts^{13,24–26,31}. However, no general quantitative model for the impact of basicity on OCM performance has been established so far.

The ability to retain a stable carbonate at the studied OCM temperature results in higher C_2 yields (hypothesis 4). Hence, the actual presence of carbonates during OCM is likely. Carbonates could therefore contribute to the reaction, i.e., via ongoing transitions from carbonate into oxide and vice versa, which would continuously produce new defect sites³². However, the presence of carbonate phases can assist also the formation and stabilization of peroxide species^{33–36}. Such peroxide species could activate methane in a selective fashion to generate methyl radicals^{32,35,37}. Otsuka et al.^{38,39} reacted CH_4 with bulk peroxides (Na, Sr, Ba) and observed C_2 coupling products already at 400°C .

C_2 yields benefit also from the ability of the support oxide to form a carbonate (3a “carbonate-supported”). Possibly, carbonate formation suppresses unselective combustion of methane and C_2 hydrocarbons, which was reported to occur on oxide-supports (Al_2O_3 ⁴⁰, SiO_2 ⁴¹, and TiO_2 ⁴², group 3b “oxide-supported”).

The essential role of carbonates in the model provides a clear incentive to rethink the classical OCM reaction mechanism and in particular the role of CO_2 . Carbon dioxide was so far typically regarded only as an undesired byproduct. However, the only way that carbonates can form during OCM is a reaction with CO_2 . Hence, CO_2 could be essentially required to stabilize carbonate phases that contribute directly to C_2 formation or prevent further oxidation of C_2 products. This could provide also a new

explanation of the observation that a maximum of ca. 30% C_2 yield has not been overcome so far when feeding CH_4/O_2 mixtures to the reactor¹⁷. Methane combustion would be essentially required to form CO_2 and the carbonate phase, thus limiting the achievable C_2 yield. Supplying the required CO_2 to the catalyst via CO_2 co-feed instead of methane combustion could provide a simple path to improved C_2 yields.

The influence of CO_2 was experimentally verified. The stability of a carbonate depends on both temperature and CO_2 partial pressure. Increasing the CO_2 partial pressure should increase the carbonate’s stability, which, according to the model, should increase the C_2 yield. We tested this effect experimentally using nine different catalysts belonging to group 3b (see Supplementary Note 19). Figure 5 compares the C_2 yields measured at 800°C , both in the absence and in the presence of CO_2 co-feed. When carbon dioxide is added to the OCM feed the observed C_2 yields either increase (by up to 2.5%) or remain constant. For the alumina support alone, which cannot form a carbonate, CO_2 addition does not have a measurable impact. The experimental data further support the model interpretation that carbonates are likely to be formed during OCM and result in higher C_2 yields.

The following paragraphs discuss the power of the developed meta-analysis method as well as current limitations. The exemplarily performed evaluation of 1802 observations in OCM catalysis provides a clear picture of physico-chemical properties that have a statistically significant impact on the observed C_2 yield. What sets the presented meta-analysis apart from previous work is that chemical knowledge and chemical reasoning become an integral part of the analysis, i.e., a chemical instead of mathematical perspective.

The method yields a robust and meaningful model, with C_2 yield distributions becoming narrower with each model level. Yet, a considerable variety of C_2 yields remains within each property group. This could be improved by further refined hypotheses, additional property descriptors, or more complicated combinations/interactions of the catalyst constituents.

A substantial contribution to the broadness of the distribution results from the variance in the underlying data (see Supplementary Note 20 for yields reported, e.g., Li/Mg and Mn/Na/W/Si). This could result from the diversity of employed experimental conditions, but also subtle changes in the synthesis procedure that can strongly influence the structure and performance of a catalyst. In future work it would be desirable to incorporate also intricate

details such as catalyst structures, synthesis recipes or kinetic information.

One major advantage of our approach is the combination of experimental data with sound chemical textbook knowledge. Thus, hypotheses with a high level of abstraction can be formulated. Pioneering work of, e.g., Schueth^{43–46}, Rothenberg^{47–51}, and Baumes^{52,53} provides access to more advanced descriptors. Also descriptors based on DFT calculations^{54–58}, e.g., efficiently computed d-band centres⁵⁹, could be incorporated. Alternative basicity descriptors such as the Smith scale²⁸, optical basicity²⁹, or the ionic-covalent parameter³⁰ should be implemented to test whether OCM performance is primarily governed by basicity or indeed the presence of carbonates. In the future, hypotheses could be generated via rapidly improving algorithms and machine-learning approaches (see e.g., refs. 60–62 for excellent reviews), or following schemes suggested by Rothenberg and co-workers^{50,63,64}.

In general, statistical tools can explore only correlations between variables that are quantitatively accessible. Quantity, quality, and distribution of literature data could be improved by adopting standardized testing and benchmarking procedures for a particular reaction^{65,66}, depositing complete sets of original data in open repositories and using data formats designed for facile data sharing⁶⁷.

The developed meta-analysis method allows the identification of catalytic property–performance correlations hidden in the vast body of existing experimental research. Applying the method to the OCM, we identify general correlations between a material’s physico-chemical properties and its OCM performance. Good catalysts comprise at least two elements, with one element being able to form a thermodynamically stable carbonate at the temperatures of OCM reaction, and a second element forming a thermally stable (non-sintering) oxide under OCM conditions. Good catalysts apparently require a support that provides a high surface area at OCM temperatures, and carbonate(s) that either contribute directly to C₂ formation and/or prevent subsequent unselective oxidation of the C₂ products. Future work could assess the specific role of CO₂ and carbonates in OCM using, e.g., in-situ spectroscopic experiments or quantum-chemical calculations. In a broader context, the developed meta-analysis method provides a generic and open platform for data science in heterogeneous catalysis. The concept allows the extensions to other reactions, more complex and fundamental descriptors derived, e.g., from quantum chemistry, and ultimately also the incorporation of a catalyst’s structure. Consequently, it could empower all chemists to work as data scientists and explore the full potential of their valuable experimental data.

Methods

Dataset corrections. The dataset compiled by Zavyalova et al.¹³ was corrected for errors (wrong temperature or composition, duplicate entries) based on a comparison with the original literature. Some entries were removed that contained highly unusual feed compositions and contact times, lacked essential information (T , p_{CH_4} , p_{O_2}) or were not measured in continuous-flow with methane and oxygen co-feed. Moreover, we removed the predefined data categories “promotor” and “support”, to which Zavyalova et al.¹³ had assigned some elements, and reassigned the respective elements to the more general categories “cation” and “anion” (see Supplementary Note 3 and Supplementary Note 4). The corrected dataset is available as Supplementary Data 1 (20160720_corrected-dataset_xls).

Physico-chemical properties. The physico-chemical properties of individual chemical elements and chemical compounds formed by these elements were compiled from literature and own measurements. The properties include the molar mass of an element, its position in the periodic table, the ability to form a carbonate, and the carbonate’s decomposition temperature (thermal stability, indicator for basicity). Moreover, properties of potentially formed oxides were incorporated (oxidation number, the stoichiometry of the most stable oxide, and the oxide’s melting point). From this melting point, the Tammann temperature⁶⁸ was computed as a measure of the onset of sintering and loss of surface area.

Supplementary Note 5 describes the complete property table, the respective data are provided as Supplementary Data 2 (20160811_element-properties.xls).

Physico-chemical descriptors. The physico-chemical descriptors were computed by a set of instructions that combine values extracted from the OCM catalyst table as well as the elemental-property table. The descriptor values were written into new columns (so-called dummy variables) of an extended data file. Important descriptor categories employed in the present study are the ability of each catalyst component (e.g., “cation”) (i) to form a carbonate compound, (ii) the thermal stability of this carbonate, and (iii) the ability to form a sufficiently stable oxide that could act as catalyst support. A value “true” (“1”) is assigned to a carbonate descriptor (i) if the respective catalyst component is able to form a carbonate. Descriptor (ii) describes whether this carbonate is stable under OCM reaction conditions. A value “true” is assigned if the OCM reaction is measured at a temperature that is lower than the actual temperature at which the carbonate starts to decompose, corrected for an offset that accounts for experimental uncertainties ($T_{\text{OCM measurement}} < T_{\text{carbonate decomposition}} + 100 \text{ K}$). A “true” value is assigned to the respective support descriptors (iii) if the catalyst contains a cation that can form a thermally stable oxide, i.e., an oxide that possesses a Tammann temperature ($T_{\text{Tammann}} = T_{\text{melt,oxide}} \times 0.6$) higher than the actual temperature at which the OCM data were measured for the catalyst, and if the concentration of this oxide amounted to at least 50 wt% of the catalysts constituents, each in the form of its most stable oxide. The descriptor rules take the reaction temperature explicitly into account for each catalyst, hence the same descriptor can adopt different values depending on the actual conditions under which a catalyst was measured. Supplementary Note 6 provides further details on descriptor computation and resulting values. Respective data are provided in Supplementary Data 3 (20160729b_corrected_preprocessed_dataset_descriptors_xlsx).

Sorting rules. Sorting rules assigned each catalyst entry to one of at least two distinct “property groups” based on the values of the corresponding descriptor variables. In the studied OCM example, sorting rules were implemented to generate property groups that, e.g., (1) contain at least one element that can form a carbonate (“can form carbonate”), that (2) contain at least one additional element that fulfils the proposed support criterion of a stable and abundant oxide (“supported”), evaluate (3) whether the element that has been assigned the support function is also able to form a carbonate (“carbonate-supported”), and probe (4) whether the carbonate with the highest decomposition temperature is thermodynamically stable under the respective OCM conditions. Thus, very complex and hierarchically nested hypotheses can be formulated. Supplementary Note 7 discusses the employed sorting rules in detail. Respective data are provided in Supplementary Data 4 (20160729b_corrected_preprocessed_dataset_property_groups_xlsx).

Regression analysis. A multiple regression analysis was employed in order to quantify the performance difference between catalysts assigned to two different property groups while compensating also for the effect of differences in the OCM measurement conditions. OCM catalysts are often compared based on the obtained yield of C₂ products. However, it is obvious from chemical kinetics that the measured C₂ yield is critically influenced not only by a catalyst’s composition, but also the reaction conditions (T , p_{CH_4} , p_{O_2} , p_{total} , contact time) employed in a catalytic test. Due to the lack of standardized OCM test protocols, employed operation conditions vary significantly between different literature reports. Unfortunately, the limited amount of catalytic data available in the studied dataset did not allow a complete kinetic description for each catalyst. In order to mitigate this problem to some extent, the regression equation (1) implements a linear correction for the variables “ratio $p_{\text{CH}_4}/p_{\text{O}_2}$ ” and “temperature”:

$$Y_i = \beta_0 + \beta_{\text{group}} \times D_{\text{group},i} + \beta_T \times \Delta T_i + \beta_{\text{CH}_4/\text{O}_2} \times \Delta p_{\text{CH}_4/\text{O}_2,i} + U_i \quad (1)$$

whereas the dummy variable $D_{\text{group},i} \in \{0, 1\}$ describes to which of the two compared property groups each catalyst i had been assigned. The regression procedure then adjusts all β values in order to minimize the sum of squares of the residuals U_i between experimental (Y_i) and computed yield values over all entries in the two groups. The derived β values quantify the influence of temperature (β_T), pressure ratio ($\beta_{\text{CH}_4/\text{O}_2}$), and most importantly, the effect of belonging to either one of the two compared catalyst groups (β_{group}) on Y_{C_2} . Hence, β_{group} is the obtained quantitative measure (in terms of corrected C₂ yield) for the influence of a set of physico-chemical properties on the catalytic performance in OCM.

Significance testing. The statistical significance of information derived from each β regression coefficient was judged via a t -test. The t -test relates the value of a β regression coefficient to its estimated standard error. From this, a probability value p is calculated for each β regression coefficient. It gives the significance of the observed effect by the probability that an estimated β value is decided to be non-zero, whereas the true value is zero. Lower p -values correspond to a lower probability of erroneous assignment of an effect, and thereby a high statistical significance of the observed effect. A p -value < 0.05 corresponding to a confidence of 95% is typically used as indicator of statistical significance. The p -values $p = 0.000$ (equivalent to $p < 0.0005$) obtained for most of the hypotheses of our

OCM example corresponds to a confidence exceeding 99.95% that two compared catalyst groups are indeed distinctly different in their C_2 yield. Supplementary Note 8 provides details on the employed regression analysis and significance testing. Supplementary Note 9 illustrates the results of a typical regression analysis exemplarily for the tested OCM hypothesis 1 with respect to group assignment, pressure ratio, temperature, β values, p -values, and property distributions.

Estimation strategy. The estimation strategy evaluates four different property–performance hypotheses labeled (1, 2, 3, 4) that divide the initial OCM data successively into smaller subgroups (1a, 1b,...). All individual hypotheses are then assembled into one overall property–performance model (Figs. 3 and 4). The best-performing property group is identified and compared to each other subgroup in terms of Y_{C_2} .

Robustness. The robustness tests for the final model varied systematically the threshold values of important descriptor values ($4\times$ wt% of support oxide (Supplementary Note 14a–d), $3\times$ Tammann factors (Supplementary Note 14e–g), $4\times$ offsets for carbonate stability (Supplementary Note 14h–k)). Moreover, the regression analysis was performed also without reaction-condition compensation (Supplementary Note 15), and with a frequency-weighted multiple regression that accounts for the different frequency of elemental catalyst composition reported in the database (Supplementary Note 16). Moreover, the regression was performed with $\ln(Y_{C_2})$ as independent variable instead of Y_{C_2} , (Supplementary Note 17) and using a so-called robust regression approach that reduces the influence of outliers in the data (Supplementary Note 18).

Result presentation. The regression results are presented in terms of the number of catalysts assigned to each group (N), the effective (corrected) difference in C_2 yield (regression coefficient β_{group} , in mol% C_2) between two compared groups along with the p -value corresponding to the regression coefficients (three significant digits). Moreover, the arithmetic mean of uncorrected C_2 yield (\bar{Y} , in mol% C_2) and corrected (\bar{Y}_{corr}) are reported along with the corresponding standard deviation. The presented diagrams display for each subgroup the density of yield observations vs. the uncorrected C_2 yields (see Supplementary Note 12 for details on the calculation of density plots via Epanechnikov kernel density). Supplementary Data 5 (20160713_list_cat-combination_property-groups_xlsx) provides for the final model a full list of catalyst compositions (i.e., cation combinations) contained in each property group.

Catalytic testing and experimental validation. Catalytic tests were performed to validate the main findings. Alumina-supported catalysts containing similar loadings of the precursor of either one alkali element (Li, Na, K, Rb, Cs) or one alkaline earth element (Mg, Ca, Sr, Ba) were prepared via impregnation of the corresponding carbonate or acetate onto the same pre-calcined alumina support. The employed $\gamma\text{-Al}_2\text{O}_3$ support was supplied by Südchemie/Clariant as pellets ($\text{Al}_2\text{O}_3\text{-100}$), which were ground, sieved to a size fraction of 200–500 μm , and calcined for 12 h at 800 °C in air. The support was then loaded via wet impregnation with aqueous solutions (alkali carbonates, alkaline earth acetates) and dried at 60 and 150 °C under vacuum, i.e., following a procedure adapted from Kusche et al.⁶⁹ Catalyst loadings amounted to 23.1 wt% of the employed salt in anhydrous state (alkali elements: carbonates; alkaline earth elements: acetates).

OCM catalytic tests were performed in a parallel fixed bed reactor featuring 48 quartz tube reactors with 4 mm inner diameter. Each tube carried 50 mg catalyst powder 300 mg of pre-calcined SiC granules placed upstream to ensure efficient feed preheating. Catalytic performance was measured in 25 K intervals (450 °C \rightarrow 5 K/min ramp \rightarrow 475 °C,...) up to 850 °C, holding each temperature to analyze product composition via GC¹⁴. OCM testing was performed with two different feed compositions, i.e., $\text{CH}_4/\text{O}_2/\text{N}_2$ (26.2:14.8:59.0) and $\text{CH}_4/\text{O}_2/\text{N}_2/\text{CO}_2$ (26.2:14.8:3.0:56.0), using fresh catalysts for each run and a flow rate of 14.7 Nml/min per reactor channel.

Data availability

All data employed in this work are available in the Supplementary Files.

Received: 4 September 2018 Accepted: 4 January 2019

Published online: 25 January 2019

References

- Glass, G. V. Primary, secondary, and meta-analysis of research. *Educ. Res.* **5**, 3–8 (1976).
- Baigent, C., Sudlow, C., Collins, R. & Peto, R. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* **324**, 71–86 (2002).
- Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197 (2015).
- Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
- Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024–1026 (2011).
- Orlitzky, M., Schmidt, F. L. & Rynes, S. L. Corporate social and financial performance: a meta-analysis. *Organ. Stud.* **24**, 403–441 (2003).
- Odabaşı, Ç., Günay, M. E. & Yildirim, R. Knowledge extraction for water gas shift reaction over noble metal catalysts from publications in the literature between 2002 and 2012. *Int. J. Hydrog. Energy* **39**, 5733–5746 (2014).
- Günay, M. E. & Yildirim, R. Neural network analysis of selective CO oxidation over copper-based catalysts for knowledge extraction from published data in the literature. *Ind. Eng. Chem. Res.* **50**, 12488–12500 (2011).
- Günay, M. E. & Yildirim, R. Developing global reaction rate model for CO oxidation over Au catalysts from past data in literature using artificial neural networks. *Appl. Catal. A* **468**, 395–402 (2013).
- Günay, M. E. & Yildirim, R. Knowledge extraction from catalysis of the past: a case of selective CO oxidation over noble metal catalysts between 2000 and 2012. *ChemCatChem* **5**, 1395–1406 (2013).
- Baroi, C. & Dalai, A. K. Review on biodiesel production from various feedstocks using 12-tungstophosphoric acid (TPA) as a solid acid catalyst precursor. *Ind. Eng. Chem. Res.* **53**, 18611–18624 (2014).
- Tapan, N. A., Günay, M. E. & Yildirim, R. Constructing global models from past publications to improve design and operating conditions for direct alcohol fuel cells. *Chem. Eng. Res. Des.* **105**, 162–170 (2016).
- 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* **3**, 1935–1947 (2011).
- Kondratenko, E. V., Schluter, M., Baerns, M., Linke, D. & Holena, M. Developing catalytic materials for the oxidative coupling of methane through statistical analysis of literature data. *Catal. Sci. Technol.* **5**, 1668–1677 (2015).
- Zavyalova, U., Holena, M., Schlögl, R., & Baerns, M. Oxidative coupling of methane. <http://www.fhi-berlin.mpg.de/acnew/departments/pages/ocmdata.html> (2011).
- Keller, G. E. & Bhasin, M. M. Synthesis of ethylene via oxidative coupling of methane. *J. Catal.* **73**, 9–19 (1982).
- Kondratenko, E. V. et al. Methane conversion into different hydrocarbons or oxygenates: current status and future perspectives in catalyst development and reactor operation. *Catal. Sci. Technol.* **7**, 366–381 (2017).
- Su, Y. S., Ying, J. Y. & Green, W. H. Jr Upper bound on the yield for oxidative coupling of methane. *J. Catal.* **218**, 321–333 (2003).
- Arndt, S. et al. A critical assessment of Li/MgO-based catalysts for the oxidative coupling of methane. *Catal. Rev.* **53**, 424–514 (2011).
- Horn, R. & Schlögl, R. Methane activation by heterogeneous catalysis. *Catal. Lett.* **145**, 23–39 (2015).
- Holmen, A. Direct conversion of methane to fuels and chemicals. *Catal. Today* **142**, 2–8 (2009).
- Schlögl, R. Heterogeneous catalysis. *Angew. Chem. Int. Ed.* **54**, 3465–3520 (2015).
- Simon, Y. et al. Detailed mechanism of the oxidative coupling of methane. *Stud. Surf. Sci. Catal.* **147**, 571–576 (2004).
- Maitra, A. M., Campbell, I. & Tyler, R. J. Influence of basicity on the catalytic activity for oxidative coupling of methane. *Appl. Catal., A* **85**, 27–46 (1992).
- JASP, Carreiro & Baerns, M. Oxidative coupling of methane. *J. Catal.* **117**, 258–265 (1989).
- Lunsford, J. H. The catalytic oxidative coupling of methane. *Angew. Chem. Int. Ed. Engl.* **34**, 970–980 (1995).
- Choudhary, V. R., Mulla, S. A. R. & Rane, V. H. Surface basicity and acidity of alkaline earth-promoted La₂O₃ catalysts and their performance in oxidative coupling of methane. *J. Chem. Technol. Biotechnol.* **72**, 125–130 (1998).
- Smith, D. W. An acidity scale for binary oxides. *J. Chem. Educ.* **64**, 480 (1987).
- Duffy, J. A. A review of optical basicity and its applications to oxidic systems. *Geochim. Cosmochim. Acta* **57**, 3961–3970 (1993).
- Lebouteiller, A. & Courtine, P. Improvement of a bulk optical basicity table for oxidic systems. *J. Solid State Chem.* **137**, 94–103 (1998).
- Choudhary, V. R., Mulla, S. A. R. & Uphade, B. S. Oxidative coupling of methane over alkaline earth oxides deposited on commercial support pre-coated with rare earth oxides. *Fuel* **78**, 427–437 (1999).
- Dubois, J.-L. & Cameron, C. J. Synergy between stable carbonates and yttria in selective catalytic oxidation of methane. *Chem. Lett.* **20**, 1089–1092 (1991).
- Appleby, A. J. & Nicholson, S. The reduction of oxygen in molten lithium carbonate. *J. Electroanal. Chem. Interfacial Electrochem.* **53**, 105–119 (1974).
- Chen, L.-J., Lin, C.-J., Zuo, J., Song, L.-C. & Huang, C.-M. First spectroscopic observation of peroxocarbonate/peroxodicarbonate in molten carbonate. *J. Phys. Chem. B* **108**, 7553–7556 (2004).

35. Moneuse, C., Cassir, M., Martin, G. & Devynck, J. Oxidative dimerization of methane in molten $\text{Na}_2\text{CO}_3\text{-K}_2\text{CO}_3$ eutectic at 800 °C: Comparison with other molten salts. *Appl. Catal. A* **85**, 147–155 (1992).
36. Moutiers, G., Cassir, M. & Devynck, J. Electrochemical characterization of the $\text{O}_2^-/\text{O}_2^{2-}$ system in molten $\text{Na}_2\text{CO}_3+\text{K}_2\text{CO}_3$ eutectic at 750 °C. *J. Electroanal. Chem. Interfacial Electrochem.* **315**, 103–112 (1991).
37. Conway, S. J., Szanyi, J. & Lunsford, J. H. Catalytic properties of lithium carbonate melts and related slurries for the oxidative dimerization of methane. *Appl. Catal.* **56**, 149–161 (1989).
38. Otsuka, K., Murakami, Y., Wada, Y., Said, A. A. & Morikawa, A. Oxidative couplings of methane, ethane, and propane with sodium peroxide at low temperatures. *J. Catal.* **121**, 122–130 (1990).
39. Otsuka, K., Said, A. A., Jinno, K. & Komatsu, T. Peroxide anions as possible active species in oxidative coupling of methane. *Chem. Lett.* **16**, 77–80 (1987).
40. Mariscal, R., Soria, J., Peña, M. A. & Fierro, J. L. G. Structure and reactivity of undoped and sodium-doped $\text{PbO}/\alpha\text{-Al}_2\text{O}_3$ catalysts for oxidative coupling of methane. *Appl. Catal. A* **111**, 79–97 (1994).
41. Carreiro, J.A.S.P. & Baerns, M., Oxidative coupling of methane: I. Alkaline earth compound catalysts. *Journal of Catalysis* **117**, 258–265 (1989).
42. Gong, M. C., Zhou, J. L., Xu, Z. H., Chen, Y. Q. & Chen, Y. Proceedings of the second workshop on C1-C3 hydrocarbon conversion study on the oxidative coupling of methane: XRD and XPS study of TiO_2 -based catalysts promoted by different additives. *Catal. Today* **24**, 259–261 (1995).
43. Klanner, C. *Evaluation of Descriptors for Solids* (Ruhr-Universität Bochum, Bochum, 2004).
44. Klanner, C. et al. The development of descriptors for solids: teaching “catalytic intuition” to a computer. *Angew. Chem. Int. Ed.* **43**, 5347–5349 (2004).
45. Klanner, C., Farrusseng, D., Baumes, L., Mirodatos, C. & Schüth, F. How to design diverse libraries of solid catalysts? *Qsar. Comb. Sci.* **22**, 729–736 (2003).
46. Schüth, F. et al. High throughput experimentation in oxidation catalysis: Higher integration and “intelligent” software. *Catal. Today* **117**, 284–290 (2006).
47. Madaan, N., Shiju, N. R. & Rothenberg, G. Predicting the performance of oxidation catalysts using descriptor models. *Catal. Sci. Technol.* **6**, 125–133 (2016).
48. Morra, G., Farrusseng, D., Bouchy, C. & Morin, S. Knowledge based catalyst design by high throughput screening of model reactions and statistical modelling. *Oil Gas. Sci. Technol.* **68**, 487–504 (2013).
49. Ras, E.-J., Louwse, M. J., Mittelmeijer-Hazeleger, M. C. & Rothenberg, G. Predicting adsorption on metals: simple yet effective descriptors for surface catalysis. *Phys. Chem. Chem. Phys.* **15**, 4436–4443 (2013).
50. Ras, E.-J. & Rothenberg, G. Heterogeneous catalyst discovery using 21st century tools: a tutorial. *RSC Adv.* **4**, 5963–5974 (2014).
51. Ras, E.-J. *Descriptors for Solid Catalysts: 21st Century Discovery Tools* (University of Amsterdam, Amsterdam, 2014).
52. Baumes, L. A., Serna, P. & Corma, A. Merging traditional and high-throughput approaches results in efficient design, synthesis and screening of catalysts for an industrial process. *Appl. Catal. A* **381**, 197–208 (2010).
53. Serna, P., Baumes, L. A., Moliner, M. & Corma, A. Combining high-throughput experimentation, advanced data modeling and fundamental knowledge to develop catalysts for the epoxidation of large olefins and fatty esters. *J. Catal.* **258**, 25–34 (2008).
54. Ferrin, P. et al. Modeling ethanol decomposition on transition metals: a combined application of scaling and Bronsted-Evans-Polanyi Relations. *J. Am. Chem. Soc.* **131**, 5809–5815 (2009).
55. Latimer, A. A. et al. Mechanistic insights into heterogeneous methane activation. *Phys. Chem. Chem. Phys.* **19**, 3575–3581 (2017).
56. Latimer, A. A. et al. Understanding trends in C-H bond activation in heterogeneous catalysis. *Nat. Mater.* **16**, 225 (2017).
57. Medford, A. J. et al. Activity and selectivity trends in synthesis gas conversion to higher alcohols. *Top. Catal.* **57**, 135–142 (2014).
58. Vojvodic, A. et al. On the behavior of Bronsted-Evans-Polanyi relations for transition metal oxides. *J. Chem. Phys.* **134**, 244509 (2011).
59. Takigawa, I., Shimizu K-i, Tsuda, K. & Takakusagi, S. Machine-learning prediction of the d-band center for metals and bimetallics. *RSC Adv.* **6**, 52587–52595 (2016).
60. Zunger, A. Inverse design in search of materials with target functionalities. *Nat. Rev. Chem.* **2**, 0121 (2018).
61. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
62. Goldsmith, B. R., Esterhuizen, J., Liu, J.-X., Bartel, C. J. & Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **64**, 2311–2323 (2018).
63. Landman, I. R., Paulson, E. R., Rheingold, A. L., Grotjahn, D. B. & Rothenberg, G. Designing bifunctional alkene isomerization catalysts using predictive modelling. *Catal. Sci. Technol.* **7**, 4842–4851 (2017).
64. Maldonado, A. G. & Rothenberg, G. Predictive modeling in homogeneous catalysis: a tutorial. *Chem. Soc. Rev.* **39**, 1891–1902 (2010).
65. Bligaard, T. et al. Toward benchmarking in catalysis science: best practices, challenges, and opportunities. *ACS Catal.* **6**, 2590–2602 (2016).
66. McCrory, C. C. L., Jung, S., Peters, J. C. & Jaramillo, T. F. Benchmarking heterogeneous electrocatalysts for the oxygen evolution reaction. *J. Am. Chem. Soc.* **135**, 16977–16987 (2013).
67. Kitchin, J. R. Examples of effective data sharing in scientific publishing. *ACS Catal.* **5**, 3894–3899 (2015).
68. Tammann, G. & Sworykin, A. Zur Bestimmung der Temperatur des Zusammenbackens. *Z. für Anorg. und Allg. Chem.* **176**, 46–48 (1928).
69. Kusche, M. et al. Enhanced activity and selectivity in catalytic methanol steam reforming by basic alkali metal salt coatings. *Angew. Chem. Int. Ed.* **52**, 5028–5032 (2013).

Acknowledgements

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC-2008/1 (UniSysCat) - 390540038 as well as EXC 314 (UniCat) are gratefully acknowledged. R.K. thanks Einstein Foundation Berlin for generous support provided by an Einstein-Junior-Fellowship (EJF-2011-95). The support of Denis Bernsmeier during manuscript revisions is highly appreciated.

Author contributions

R.K. and A.W. with support of J.P. designed the study R.S. and A.F. performed the study and developed the statistical approach. E.V.K. contributed to the experimental validation. All authors contributed to the manuscript writing.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-08325-8>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019