Chemical Science

rsc.li/chemical-science

algorithmic
workflow

conventional
screening

ROYAL SOCIETY
OF CHEMISTRY | Celebrating
IYPT 2019

EDGE ARTICLE
Oliver Trapp *et al.*
Identifying high-performance catalytic conditions for carbon
dioxide reduction to dimethoxymethane by multivariate
modelling

Check for updates

Identifying high-performance catalytic conditions for carbon dioxide reduction to dimethoxymethane by multivariate modelling†

Max Siebert, [ID] Gerhard Krennrich, Max Seibicke, Alexander F. Siegle and Oliver Trapp [ID] *

In times of a warming climate due to excessive carbon dioxide production, catalytic conversion of carbon dioxide to formaldehyde is not only a process of great industrial interest, but it could also serve as a means for meeting our climate goals. Currently, formaldehyde is produced in an energetically unfavourable and atom-inefficient process. A much needed solution remains academically challenging. Here we present an algorithmic workflow to improve the ruthenium-catalysed transformation of carbon dioxide to the formaldehyde derivative dimethoxymethane. Catalytic processes are typically optimised by comprehensive screening of catalysts, substrates, reaction parameters and additives to enhance activity and selectivity. The common problem of the multidimensionality of the parameter space, leading to only incremental improvement in laborious physical investigations, was overcome by combining elements from machine learning, optimisation and experimental design, tripling the turnover number of 786 to 2761. The optimised conditions were then used in a new reaction setup tailored to the process parameters leading to a turnover number of 3874, exceeding by far those of known processes.
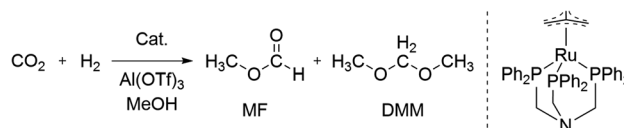
Introduction

In this century, our ecosystem faces severe problems such as global warming, environmental pollution and resource depletion. The negative impact of humankind as well as its responsibility in solving these problems can no longer be ignored.[1,2] Besides the necessity for future-oriented global politics,[3,4] both the scientific community and the chemical industry must provide answers to crucial questions regarding sustainable process development, alternative energies as well as recycling of waste and pollutants.[5–9] In this context, more efficient techniques must be developed and applied in research to reduce time, cost and resources.[10–12]

In catalytic investigations, system optimisation is typically approached by one-factor-at-a-time (OFAT) methods, successively screening along one parameter axis. Once optimised, a parameter is kept constant for the subsequent experiments. In this univariate analysis, variables are treated as being independent of each other. Beside the vast number of experiments that must be performed, local maxima with higher performances might be missed. Consequently, algorithm-based screening and optimisation techniques have been among the fastest growing research areas in recent years.[13–24] Considering the interactions of parameters, optimised results can be achieved with minimal experimental effort.[25,26]

Recently, we applied a univariate optimisation approach including several hundred catalytic reactions to improve the selective ruthenium-catalysed transformation of carbon dioxide to dimethoxymethane (DMM) reaching a turnover number (TON) of 786 (Scheme 1).[27,28] The product DMM itself is a high value feedstock for biofuels, but can also be hydrolysed yielding formaldehyde and methanol or directly employed as a formaldehyde synthon.[29,30] Beside the desired product, only methyl formate (MF) was formed with TONs of up to 1290 (Scheme 1).[27] Previously, two studies on the selective hydrogenation of CO_2 by the group of Klankermayer showed the formation of DMM and MF by using a homogeneous ruthenium catalyst[31] with TONs of 214 and 104 or a cobalt catalyst[32] with TONs of 157 and 37, respectively. Further selective reductions toward the formaldehyde oxidation state were reported utilising hydroboration,[33–37] hydrosilylation,[38,39] and frustrated Lewis pairs,[40,41] however, being mainly of academic interest due to the stoichiometric use of reducing reagents.

Department Chemie, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany. E-mail: oliver.trapp@cup.uni-muenchen.de

† Electronic supplementary information (ESI) available: Synthetic procedures, multivariate modelling procedure, additional tables, catalysis data, R-code. See DOI: 10.1039/c9sc04591k

Scheme 1 Reductive transformation of CO_2 towards the formaldehyde oxidation level yielding methyl formate (MF) and dimethoxymethane (DMM).

Inspired by previous reports, which underline the importance of a holistic view on the results obtained,[13,19] we now utilised multivariate analysis[42,43] to further optimise the investigated catalysis. Based on the large amount of experimental data and the dependency of the catalytic system on seven different parameters, we envisioned to increase its performance by modelling and predicting conditions optimally accounting for parameter interaction. For this purpose, we devised a multistep process combining elements from machine learning, optimisation and experimental design (design of experiments, DoE). Fig. 1 gives a schematic overview of the easy-to-use algorithmic workflow, which was developed in this work and is applicable for any catalytic screening. Experimental data (Fig. 1(I)) was modelled using the random forest (RF) algorithm (Fig. 1(II)) to identify promising subspaces with high catalytic performance (Fig. 1(III)). To better understand the origin of the exceptional activity and to extend the amount of data for further modelling, the subspace was augmented (Fig. 1(IV)) by additional experiments (Fig. 1(V)) that were based on experimental design. Starting from these first experiments, an iterative workflow – consisting of DoE (Fig. 1(V) and (VI)), optimisation (relaxation; Fig. 1(VII) and (VIII)) and evaluation (Fig. 1(IX)) – was applied until the final optimum was reached (Fig. 1(X)).

Results and discussion

Catalytic system

The selective transformation of CO_2 to DMM was performed using the ruthenium catalyst $[Ru(N\text{-triphos}^{Ph})(tmm)]$ (tmm =

trimethylenemethane dianion) for hydride transfer and $Al(OTf)_3$ to facilitate the acetalisation with methanol (Scheme 1). Beneficially, the robust catalyst can be obtained on a large scale in a simple three-step procedure, which underlines its suitability for industrial and large scale applications.[27] The catalysis is mainly affected by the reaction temperature (T), partial pressure of H_2 (p_{H_2}) and CO_2 (p_{CO_2}), reaction time (t), the amount of catalyst (n_{cat}) and Lewis acid (n_{add}) as well as the reaction volume (V).

Experimental data analysis

Following the rationale outlined in the previous section, the optimisation started with a corpus of 49 unique settings of the seven process factors (X: $X_1, X_2, ..., X_7 = T, p_{H_2}, p_{CO_2}, t, n_{cat}, n_{add}, V$), two responses ($Y$: $Y_1, Y_2 = TON_{DMM}, TON_{MF}$) and 144 observations, which are measured values of the responses (Table 1, Fig. 1(I), ESI Table 1†).[27] Each setting was run as a triple replicate, except for three settings that were realised as simple replicates, thus making up the 144 cases of the dataset. Goal of the optimisation was finding reaction conditions X^*: $X_1^*, X_2^*, ..., X_7^*$, that maximise TON_{DMM} (for a theoretical introduction on this approach, see the Methods section). As process factors, six of the previously investigated parameters, namely temperature, partial pressure of H_2 and CO_2, reaction time as well as the amount of catalyst and Lewis acid, proved suitable for the optimisation. The influence of varying the nature of the Lewis acid was neglected here, because statistical analysis of the dataset revealed $Al(OTf)_3$ as the only promising candidate. Instead, the volume of the catalysis solution was considered for the multivariate analysis to account for mass transfer phenomena and gas solubility effects (concentration effects) potentially influencing the performance of the catalyst.

Modelling and analysis of datasets based on empirical optimisation (e.g. OFAT) is complex and more difficult than estimating parametric surrogate models from well-designed data. The difficulties arise primarily from non-linearities, which simple parametric models cannot appropriately describe. Furthermore, OFAT variations often and inadvertently lead to co-varying and thereby confounding the effects of process parameters. For instance, in the present dataset the process parameters n_{cat} and n_{add} were found to be highly correlated with a Pearson correlation coefficient $r(n_{cat} - n_{add}) = 0.895$, rendering the data non-informative in terms of the independent effects of n_{cat} and n_{add}.
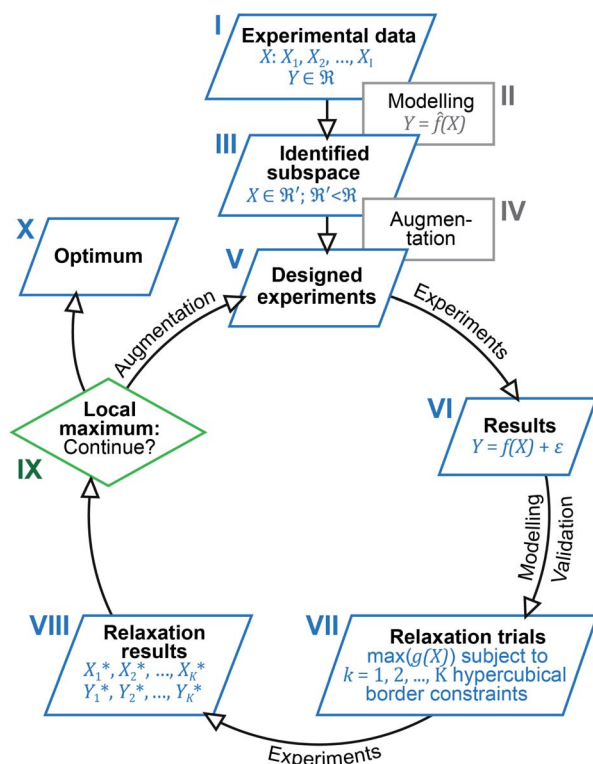

Fig. 1 Outline of the algorithm-based workflow combining experimental data analysis, DoE and optimisation.

Table 1 Factors and responses of the dataset[27]

Entry	Variable	Mean	Median	Min	Max
1	T (°C)	89.2	90	20	120
2	p_{H_2} (bar)	82.1	90	40	100
3	p_{CO_2} (bar)	17.9	20	5	40
4	t (h)	26.5	18	1	168
5	n_{cat} (μmol)	1.3	1.5	0	3
6	n_{add} (μmol)	5.4	6.25	0	12.5
7	V (mL)	0.5	0.5	0.25	0.5
8	TON_{DMM}	275	263	0	906
9	TON_{MF}	145	71	0	1377

Next, the problem for these datasets is finding an appropriate functional representation, $\hat{f}()$, of the target response as a function of the process factors X_i, that is $TON_{DMM} = f(T, p_{H_2}, p_{CO_2}, t, n_{cat}, n_{add}, V) + \varepsilon$, without making any prior assumptions about the analytical form of the 'true' function $f()$. Here, we used RF as a powerful non-linear and easy-to-use method for the empirical model building of complex datasets (Fig. 1(II)).[44,45]

The RF models describe 79% of the responses' variance on average ($R^2(TON_{DMM}) = 0.83$; $R^2(TON_{MF}) = 0.74$), which is a good result given the heterogeneous nature of the dataset. Further, tuning the RF hyperparameters using 10-fold cross validation with consecutive blocks (mtry* = 2 with $R^2 = 0.83$) revealed that the default RF hyperparameters describe the data appropriately. The effect structure of the models can be conveniently explored by plotting the RF model predictions $\hat{f}(X)$ against the process factors $X_1, X_2, \ldots, X_7$. Fig. 2 shows the effects of the process parameters T, p_{CO_2}, n_{cat} and n_{add}, given the median values $p_{H_2} = 90$ bar, $t = 18$ h and $V = 0.5$ mL, as a conditional trellis plot.

As a tree-based ensemble method, RF splits and mean-aggregates the experimental space into hyperrectangles, the consequence being that the RF model surface becomes non-smooth (Fig. 2). This property of RF is a particular strength when it comes to identifying promising domains of the experimental space and was another motivation for choosing RF as modelling technique.

Evidently, n_{cat} and n_{add} exert strong non-linear, step like negative effects on the process performance (TON_{DMM}), dividing the $n_{cat} \times n_{add}$ space into domains of different performance. The pressure, p_{CO_2}, reveals a positive and temperature, T, a small convex effect, both, however, negligible compared to the dominant effects of n_{cat} and n_{add}. At this point, the erroneous impression may arise that the heuristic interpretation of the experimental data could lead to a further improvement by a trivial reduction of the amount of catalyst, which is often reduced to minimise catalyst deactivation. However, a closer look at the experimental data shows that a reduction in the amount of catalyst without changing other parameters leads to a decrease in catalytic activity with respect to DMM. This case shows exemplarily how important multivariate analytical methods can be, because they can easily identify complex correlations of process parameters that are not detectable by a univariate analysis of the scientist.

Fig. 2 Trellis plot of the random forest predictions $\hat{f}(T, p_{H_2}, p_{CO_2}, t, n_{cat}, n_{add}, V)$ with $p_{H_2} = 90$ bar, $t = 18$ h and $V = 0.5$ mL. Note the difference of max(TON_{DMM}) in the experimental data (Table 1, entry 8) and the model. The RF values are predictions causing the range of the z-axis to be smaller than the range of the empirical data.

With the condition $TON_{DMM} > 400$ deduced from the landscape of TON_{DMM} over $n_{cat} \times n_{add}$ (Fig. 2), a subset of 26 observations with 9 unique settings was selected from the data (Fig. 1(III)). In this subset, the parameters T, p_{H_2}, p_{CO_2} and t were found constant at $T = 90\ °C$, $p_{H_2} = 90$ bar, $p_{CO_2} = 20$ bar and $t = 18$ h, thereby suggesting to conditionally optimise n_{cat}, n_{add} and V first, while keeping T, p_{H_2}, p_{CO_2} and t constant for the time being.

First DoE

These 9 candidate points (identified unique settings mentioned above) were subsequently augmented by an additional number of 6 points to fully support all second order effects of n_{cat}, n_{add} and V (Fig. 1(IV)). The intention behind this augmentation was to render the formerly confounded effects of n_{cat} and n_{add} estimable as well as, assuming high complexity, to allow for interactions and non-linearities of the process parameters as a solid basis for further optimisation (Fig. 1(V)).

The 6 augmentation trials were realised as triple replicates in the lab, added to the 9 settings already available and the responses (TON_{DMM}, TON_{MF}) were modelled as a function of the least square parameters with stepwise ordinary least squares (OLS; Fig. 1(VI)). Both models accurately describe the data within the replication error and explain 91% ($R^2(TON_{DMM}) = 0.91$; DF = 37) and 97% ($R^2(TON_{MF}) = 0.97$; DF = 35) of the responses' variance with DF denoting the degrees of freedom (number of data points minus number of estimated model parameters).

The local effects of n_{cat}, n_{add} and V are depicted in Fig. 3A as response surface trellis plot. Again, there is a strong negative effect for n_{cat} along with positive effects for n_{add} and V. Together with the positive, synergistic effect between n_{add} and V, the effect structure suggests to decrease n_{cat} and to increase n_{add} and V to further maximise the catalytic performance beyond the best result of the first DoE (Fig. 4A, entry 4).

First relaxation

To optimise towards the direction of maximal improvement, following the procedure outlined in the Methods section [eqn (5)], the experimental space was relaxed with 10% step size. The triple obtained from relaxing the design space together with the achieved experimental results are listed in Fig. 4B (Fig. 1(VII) and (VIII); see the Methods section).

The joint condition max(TON_{DMM}), max(TON_{MF}) was best met by the 20% relaxation trial (Fig. 4B, entry 2), and the factor setting $n_{cat}^* = 0.075$ μmol, $n_{add}^* = 3.576$ μmol and $V^* = 0.550$ mL thus became the reference point for further optimisation. The 30% relaxation trial was very poor, indicating that a local maximum had been exceeded (Fig. 4B, entry 3). The pronounced drop in activity, resulting most likely from the reduced catalyst loading, indicates a molecular deactivation pathway of the catalytically active species due to potential inhibitors, such as carbon monoxide, moisture and oxygen, which are probably present in low concentrations.

At this point, there were two alternatives to proceed: (1) create an experimental design around n_{cat}^*, n_{add}^* and V^* to fully identify the topology around the relaxation point at the desired resolution (complexity). (2) Consider $n_{cat}^* = 0.075$ μmol, $n_{add}^* = 3.576$ μmol and $V^* = 0.550$ mL as locally optimal and switch to optimising the candidates T, p_{H_2}, p_{CO_2} and t, which had so far been kept constant.

The poor outcome of the third relaxation trial (Fig. 4B, entry 3) showed that not much was to be expected from exploring the three-dimensional environment of the 20% relaxation trial any further. Therefore, option 2 was chosen and the 20% relaxation trial became the reference point for optimising the candidates T, p_{H_2}, p_{CO_2} and t.

Second DoE

A small linear design [eqn (2)] with 5 runs in the ranges listed in Fig. 4C was created with the reference point $n_{cat}^* = 0.075$ μmol, $n_{add}^* = 3.576$ μmol and $V^* = 0.550$ mL, $T^* = 90\ °C$, $p_{H_2}^* = 90$ bar, $p_{CO_2}^* = 20$ bar bar, $t^* = 18$ h at the design centre as replicate (Fig. 1(V); see the Methods section). The experiments were realised in the lab as triple replicate to provide a measure of accuracy (Fig. 1(VI)).

The measured responses (TON_{DMM}, TON_{MF}) were linearly modelled as a function of the process parameters with stepwise OLS. The models explain 92% ($R^2(TON_{DMM}) = 0.92$; DF = 16) and 78% ($R^2(TON_{MF}) = 0.78$; DF = 18) variance of the responses thus indicating a large signal-to-noise ratio for TON_{DMM} and to a lesser extent for TON_{MF}.

Fig. 3B shows the linear effects of the process parameters on TON_{DMM} as trellis response surface plot. The factors T and p_{H_2} both have strong positive effects on TON_{DMM}, whereas t reveals only a small positive and p_{CO_2} a moderate negative effect on TON_{DMM}. Optimal conditions were found in the upper left panel and these are the conditions of the top candidate found in the design list with T, p_{H_2}, t at the upper and p_{CO_2} at the lower bound, yielding respective TONs for DMM and MF of 2610 and 2356 (Fig. 4C, entry 4).

Second relaxation

Following eqn (5), the experimental space was relaxed in 25% and 50% steps and the relaxation trials were experimentally realised in the lab (Fig. 4D and 1(VII); see the Methods section).

Again, we saw a small improvement of the 25% relaxation trial (Fig. 4D, entry 1) compared with the best candidate from the second DoE (Fig. 4C, entry 4), whereas the 50% relaxation candidate performed comparatively poorly (Fig. 4D, entry 2; Fig. 1(VIII)). With these relaxation trials, we reached the technical limits of our setup regarding hydrogen gas pressure, and therefore, the conditions of the 25% relaxation experiment can be considered locally optimal given the constraints of technical feasibility (Fig. 1(IX) and (X)). We would like to point out that the catalytic conditions optimised by the here presented strategy may still represent a local maximum. Modification of the catalyst and the additive might also result in further improvement, but as demonstrated in this work the identification of high-performance catalytic conditions should be a prerequisite for a design strategy of new catalysts.

A complete overview of the results obtained at each step of the optimisation project is given in Fig. 4E, overall tripling the

A **Response surface trellis plot for $g_3(n_{cat}, n_{add}, V)$**

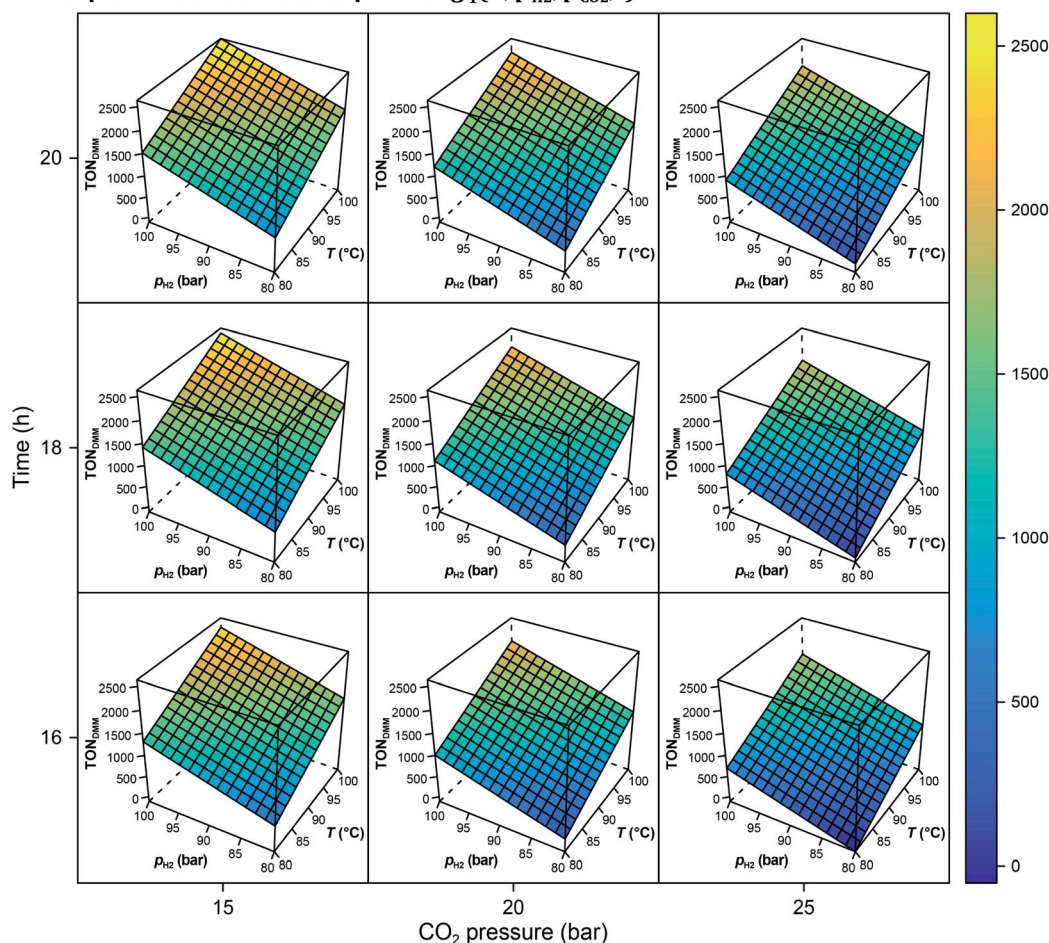**B** **Response surface trellis plot for $g_1(T, p_{H2}, p_{CO2}, t)$**

Fig. 3 Calculated response surfaces. (A) Trellis plot of the response surface for $g_3(n_{cat}, n_{add}, V)$ from OLS modelling with $T = 90$ °C, $p_{H_2} = 90$ bar, $p_{CO_2} = 20$ bar and $t = 18$ h. (B) Response surface trellis plot of the linear OLS model for $g_1(T, p_{H_2}, p_{CO_2}, t)$ with $n_{cat}^* = 0.075$ µmol, $n_{add}^* = 3.576$ µmol and $V^* = 0.550$ mL. For a mathematical definition of the polynomial parametric surrogates, $g_{1,3}()$, see the Methods section.

initial TON_{DMM} value of 786 to a final value of 2761. As illustrated, the standard deviation (SD) of TON_{DMM} tends to increase with increasing mean value of TON_{DMM} (Fig. 4E), which might be a joint effect from decreasing n_{cat} and increasing both n_{add} and V over the course of the optimisation project. Decreasing n_{cat} and increasing V is equivalent to reducing the concentration of the catalyst, which presumably renders the system more

susceptible to random disturbances by catalyst deactivation, thereby providing an explanation for the observed increase of the standard deviation. Simple Spearman rank correlation analysis of the relationship between SD(TON_{DMM}), SD(TON_{MF}) and the process parameters X_i supports this hypothesis by revealing a negative and a positive association of SD(TON_{DMM}) with n_{cat} and V, respectively (ESI Table 26†).

A First DoE

entry	n_{cat} (μmol)	n_{add} (μmol)	V (mL)	TON_{DMM}	TON_{MF}
1	0.750	0.781	0.250	204	90
2	0.750	0.781	0.500	119	205
3	0.750	1.953	0.375	342	116
4	0.188	3.125	0.250	1116	437
5	0.750	3.125	0.250	314	79
6	0.469	3.125	0.375	617	201

B First relaxation

entry	n_{cat} (μmol)	n_{add} (μmol)	V (mL)	TON_{DMM}	TON_{MF}
1	0.131	3.359	0.525	1374	1560
2	0.075	3.576	0.550	1375	2762
3	0.019	3.659	0.575	180	1040

C Second DoE

entry	T (°C)	p_{H2} (bar)	p_{CO2} (bar)	t (h)	TON_{DMM}	TON_{MF}
RS1	90	90	20	18	1310	2858
1	80	100	25	20	889	3281
2	100	80	25	20	1181	2936
3	80	80	15	16	466	1996
4	100	100	15	20	2610	2356
5	100	100	25	16	1788	3189

D Second relaxation

entry	T (°C)	p_{H2} (bar)	p_{CO2} (bar)	t (h)	TON_{DMM}	TON_{MF}
RS2	100	100	15	20	2563	2401
1	105	105	12.5	21	2761	1769
2	110	110	10	22	2209	1276

E Scatter plot TON_{MF} versus TON_{DMM}

Fig. 4 Optimisation summary. Process parameters and TONs for DMM and MF at different optimisation steps: (A) first DoE. (B) First relaxation (10, 20 and 30%). (C) Second DoE. (D) Second relaxation (25 and 50%). (E) Scatter plot TON_{MF} *versus* TON_{DMM} summarising the outcome of the optimisation project. The results of each optimisation step are marked according to the colour code. Replication studies (RS): the circled data points label the best hit of the first relaxation and the respective replication trials in the second DoE (RS1) as well as the best hit of the second DoE and the respective replication trials in the second relaxation (RS2).

Another interesting result in Fig. 4E refers to the independent replication error. The linear design of the second DoE includes two independent settings of the 20% reference as centre points with each point measured as triple replicate (ESI Tables 11 and 17†). This sextet from the second DoE (Fig. 4C, entry RS1) excellently matches the 20% relaxation triple (Fig. 4B, entry 2) indicating good repeatability and reliability of the system (Fig. 4E, ESI Tables 27 and 28†). In a similar way, the best candidate from the second DoE (Fig. 4C, entry 4) has been independently replicated when running the second relaxation sequence (Fig. 4D, entry RS2) and both replicates turned out

identical within the experimental error (Fig. 4E, ESI Tables 29 and 30†).

Technical adaption

The results of the non-biased mathematical modelling approach presented in this study revealed that a better catalytic performance is *inter alia* strongly correlated to a combination of lower catalyst loadings and higher reaction volumes. The heuristic interpretation of this finding indicates that the poor solubility of hydrogen gas in methanol and mass transfer might be limiting factors within our experimental setup. We anticipated that a larger reaction volume, a higher surface to volume

ratio and improved mixing would lead to better mass transfer and thus designed an experimental setup accordingly to enhance the catalytic performance. With this upscale autoclave, the TON for DMM increased to 3874, while the TON for MF reached a value of 1445, resulting in a higher selectivity toward DMM (ESI Table 21†). The result of this technical adaption shows that while the reaction is already optimised to a high degree with respect to reaction parameters, further improvements can be expected by focusing on engineering aspects of the reaction setup.

Conclusion

We demonstrated the power of multivariate optimisation for catalytic processes over the usually applied cumbersome one-factor-at-a-time method. In the homogeneously catalysed transformation of CO_2 to DMM (dimethoxymethane) and MF (methyl formate) using the ruthenium-triphos complex [Ru(N-triphos^{Ph})(tmm)], the TON (turnover number) for DMM was drastically increased to 2761 (with it: TON$_{MF}$ 1769) by an easy-to-use algorithmic workflow combined with only a small number of catalytic experiments. Given the complexity of the transformation, which depends on seven parameters, conventional OFAT screening techniques would have been very costly and time-consuming, with uncertain outcome.

Starting from catalytic data using RF (random forest) for empirical model building, an experimental subspace was identified and subsequently augmented to render the effects of a first set of three process factors estimable. Modelling and optimisation, followed by relaxation led to a sequence of relaxation trials with one candidate assumed to be locally optimal. With this candidate as reference, a linear design of the remaining four variables yielded another substantial improvement. Relaxation of the second design further enhanced the catalytic performance, thereby reaching the technical limits of the setup.

The optimised conditions were used in a specifically designed experimental setup and the highest TON for DMM of 3874 (with it: TON$_{MF}$ 1445) was obtained, which is, to the best of our knowledge, the by far highest value reported in the investigated catalysis.

Methods

Theoretical introduction

Experimental design (design of experiments, DoE) methods can be used to study the joint effects of several parameters X: X_1, X_2, …, X_I on response Y.[26,42,43] This can formally be written as:

$$Y = f(X) + \varepsilon \tag{1}$$

$f()$ denotes the true, however unknown function, linking the responses Y with the process conditions X: X_1, X_2, …, X_I, whereas ε is a random element taken from a normal distribution with variance σ^2, $\varepsilon \sim N(0,\sigma^2)$ to account for experimental uncertainties. Conceptually, nature evaluates in an experiment the function $f()$ known to her only at reaction conditions X, then

adds some random noise ε and returns the experimental results Y [eqn (1)].

Under the weak assumption that $f()$ is smooth and continuous, $f()$ can be locally approximated as polynomial parametric surrogates, $g_{1,2,3}()$, of increasing complexity, formally:

$$f(X) \cong g_1(X) = a_0 + \sum_i a_i X_i \tag{2}$$

$$f(X) \cong g_2(X) = a_0 + \sum_i a_i X_i + \sum_{j>i} a_{ij} X_i X_j \tag{3}$$

$$f(X) \cong g_3(X) = a_0 + \sum_i a_i X_i + \sum_i a_{ii} X_i^2 + \sum_{j>i} a_{ij} X_i X_j \tag{4}$$

These are linear [eqn (2)], bilinear [eqn (3)] or quadratic [eqn (4)] parametric surrogates of the true function $f()$. With the experimental values Y, X_1, X_2, …, X_I available, the unknown parameters a_i, a_{ii}, a_{ij} can be estimated from the data using ordinary least squares (OLS).[46]

Given process factors X_1, X_2, …, X_I, their ranges $X_i \in \{LB,UB\}$ with LB, UB denoting the lower and upper bounds of the process factors X_i and, depending on the expected complexity, the parametric form of the surrogate model, the design points X (experimental design) optimally supporting the chosen model can be calculated. However, in an early project phase it is often unclear which factors X_i and ranges should be chosen and what levels of complexity must be assumed for the domain under investigation. Therefore, DoE can benefit from experimental data analysis, with the latter helping to answer the questions arising in the former.

After first optimisation by an experimental design, ascending in the direction of maximal improvement can be easily achieved by increasing (relaxing) the experimental space in discrete steps and by solving a maximisation problem subject to a sequence of hypercubical constraints, that is

$$\max(g(X)) \text{ subject to } LB - k\Delta X < X < UB + k\Delta X \tag{5}$$

with ΔX being the step size of the relaxation, here taken to be 10% of the initial factor ranges, that is $\Delta X = 0.1[UB - LB]$, and LB, UB denoting the lower and upper bounds of the process factors X_i. Varying $k = 1, 2, …, K$ leads to a sequence of relaxation trials X_1^*, X_2^*, …, X_K^* to be realised in the lab.

Additional information

The detailed description of all experiments, the performed multivariate analysis, the spectroscopic data of compounds as well as the NMR spectra of compounds and catalysis samples can be found in the ESI.† In order to improve comprehensibility, simplified names were used in some cases rather than using exact IUPAC names.

All calculations were done using the statistical software R.[47] Random forest modelling was performed with the R-package 'randomForest'.[48] Experimental designs were calculated with the D-optimal criterion of the function optFederov() in the R-package 'AlgDesign'.[49] Optimisation was achieved with the

augmented Lagrange method from the R-package 'Rsolnp'.[50] Graphics were produced with the R-package 'lattice'.[51]

All catalyses and the corresponding analyses were performed following a procedure previously reported by our group.[27] The catalysis was also performed in the absence of a catalyst, a co-catalyst or both to demonstrate the need of the catalytic system for the formation of DMM and MF. In all cases, no significant conversions for both of these compounds were observed (ESI†).

Reproduction of modelling results

The R-code used as well as the catalytic data analysed (Excel sheet) are available online. The multivariate analysis can be reproduced by following the instructions in the ESI.†

Author contributions

M. Siebert and G. K. designed the experiments. M. Siebert performed the experiments. G. K. performed computer modelling and analysis of the data. All authors contributed to the interpretation of the data. The manuscript was co-written by M. Siebert and G. K. and all authors contributed to the manuscript. O. T. supervised the project.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 P. J. Crutzen, *Nature*, 2002, **415**, 23.

2 J.-L. Martin, V. Maris and D. S. Simberloff, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 6105–6112.

3 T. Sterner, E. B. Barbier, I. Bateman, I. van den Bijgaart, A.-S. Crépin, O. Edenhofer, C. Fischer, W. Habla, J. Hassler, O. Johansson-Stenman, A. Lange, S. Polasky, J. Rockström, H. G. Smith, W. Steffen, G. Wagner, J. E. Wilen, F. Alpízar, C. Azar, D. Carless, C. Chávez, J. Coria, G. Engström, S. C. Jagers, G. Köhlin, Å. Löfgren, H. Pleijel and A. Robinson, *Nature Sustainability*, 2019, **2**, 14–21.

4 Y. Geng, J. Sarkis and R. Bleischwitz, *Nature*, 2019, **565**, 153–155.

5 W. Keim, in *Carbon Dioxide as a Source of Carbon: Biochemical and Chemical Uses*, ed. M. Aresta and G. Forti, Springer, Dordrecht, 1st edn, 1987, DOI: 10.1007/978-94-009-3923-3.

6 M. Aresta, in *Carbon Dioxide as Chemical Feedstock*, ed. M. Aresta, Wiley-VCH, Weinheim, 2010.

7 E. J. Beckman, *Nature*, 2016, **531**, 180–181.

8 R. A. Sheldon, *Green Chem.*, 2016, **18**, 3180–3183.

9 J. A. Martens, A. Bogaerts, N. De Kimpe, P. A. Jacobs, G. B. Marin, K. Rabaey, M. Saeys and S. Verhelst, *ChemSusChem*, 2017, **10**, 1039–1055.

10 Y. Gil, M. Greaves, J. Hendler and H. Hirsh, *Science*, 2014, **346**, 171–172.

11 M. I. Jordan and T. M. Mitchell, *Science*, 2015, **349**, 255–260.

12 G. Rothenberg, *Catalysis: Concepts and Green Applications*, Wiley, Weinheim, 2017.

13 M. S. Sigman, K. C. Harper, E. N. Bess and A. Milo, *Acc. Chem. Res.*, 2016, **49**, 1292–1301.

14 M. H. S. Segler, M. Preuss and M. P. Waller, *Nature*, 2018, **555**, 604–610.

15 C. W. Coley, W. H. Green and K. F. Jensen, *Acc. Chem. Res.*, 2018, **51**, 1281–1289.

16 H. Gao, T. J. Struble, C. W. Coley, Y. Wang, W. H. Green and K. F. Jensen, *ACS Cent. Sci.*, 2018, **4**, 1465–1476.

17 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, *Science*, 2018, **360**, 186–190.

18 J. M. Granda, L. Donina, V. Dragone, D.-L. Long and L. Cronin, *Nature*, 2018, **559**, 377–381.

19 C. B. Santiago, J.-Y. Guo and M. S. Sigman, *Chem. Sci.*, 2018, **9**, 2398–2412.

20 S. Zhao, T. Gensch, B. Murray, Z. L. Niemeyer, M. S. Sigman and M. R. Biscoe, *Science*, 2018, **362**, 670–674.

21 A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science*, 2019, **363**, eaau5631.

22 J. P. Reid and M. S. Sigman, *Nature*, 2019, **571**, 343–348.

23 A. V. Brethomé, R. S. Paton and S. P. Fletcher, *ACS Catal.*, 2019, **9**, 7179–7187.

24 P. S. Gromski, A. B. Henson, J. M. Granda and L. Cronin, *Nat. Rev. Chem.*, 2019, **3**, 119–128.

25 G. E. P. Box, W. G. Hunter and J. S. Hunter, *Statistics for Experimenters*, John Wiley & Sons, New York, 1st edn, 1978.

26 G. E. P. Box, W. G. Hunter and J. S. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, John Wiley & Sons, Hoboken, 2st edn, 2005.

27 M. Siebert, M. Seibicke, A. F. Siegle, S. Kräh and O. Trapp, *J. Am. Chem. Soc.*, 2019, **141**, 334–341.

28 M. Seibicke, M. Siebert, A. F. Siegle, S. M. Gutenthaler and O. Trapp, *Organometallics*, 2019, **38**, 1809–1814.

29 S. Dabral and T. Schaub, *Adv. Synth. Catal.*, 2019, **361**, 223–246.

30 R. Sun, I. Delidovich and R. Palkovits, *ACS Catal.*, 2019, **9**, 1298–1318.

31 K. Thenert, K. Beydoun, J. Wiesenthal, W. Leitner and J. Klankermayer, *Angew. Chem., Int. Ed.*, 2016, **55**, 12266–12269.

32 B. G. Schieweck and J. Klankermayer, *Angew. Chem., Int. Ed.*, 2017, **56**, 10854–10857.

33 S. Bontemps, L. Vendier and S. Sabo-Etienne, *Angew. Chem., Int. Ed.*, 2012, **51**, 1671–1674.

34 S. Bontemps, L. Vendier and S. Sabo-Etienne, *J. Am. Chem. Soc.*, 2014, **136**, 4419–4425.

35 G. Jin, C. G. Werncke, Y. Escudié, S. Sabo-Etienne and S. Bontemps, *J. Am. Chem. Soc.*, 2015, **137**, 9563–9566.

36 A. Aloisi, J.-C. Berthet, C. Genre, P. Thuéry and T. Cantat, *Dalton Trans.*, 2016, **45**, 14774–14788.

37 M. R. Espinosa, D. J. Charboneau, A. Garcia de Oliveira and N. Hazari, *ACS Catal.*, 2019, **9**, 301–314.

38 T. T. Metsänen and M. Oestreich, *Organometallics*, 2015, **34**, 543–546.

39 P. Ríos, N. Curado, J. López-Serrano and A. Rodríguez, *Chem. Commun.*, 2016, **52**, 2114–2117.

40 Y. Jiang, O. Blacque, T. Fox and H. Berke, *J. Am. Chem. Soc.*, 2013, **135**, 7751–7760.

41 F. A. LeBlanc, W. E. Piers and M. Parvez, *Angew. Chem., Int. Ed.*, 2014, **53**, 789–792.

42 G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, New York, 1987.

43 D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, New York, 3rd edn, 1991.

44 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.

45 T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, New York, 2nd edn, 2009.

46 A. Sen and M. Srivastava, *Regression Analysis*, Springer-Verlag, New York, 1990.

47 R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017, https://www.R-project.org/.

48 A. Liaw and M. Wiener, *R News*, 2002, **2**, 18–22.

49 B. Wheeler, *AlgDesign: Algorithmic Experimental Design*, R package version 1.1-7.3, 2014, https://CRAN.R-project.org/package=AlgDesign.

50 A. Ghalanos and S. Theussl, *Rsolnp: General Non-Linear Optimization*, R package version 1.16, 2015, https://CRAN.R-project.org/package=Rsolnp.

51 D. Sarkar, *Lattice*, Springer-Verlag, New York, 2008.