

Optimization of the Mizoroki-Heck reaction using Design of Experiment (DoE)

Varinder K. Aggarwal,^{†} Anne C. Staubitz[†] and Martin Owen[†],*

[†] University of Bristol, School of Chemistry, Cantock's Close, Bristol, BS8 1TS, UK

[†] GlaxoSmithKline, Stevenage, UK

Supplementary Material

Statistical Optimization Models

Factorial schemes resolve the parameter space better than “One factor at a time schemes” and can give good indications about the relative importance and interaction of factors. They are especially good in the beginning of an optimization study. The models derived from factorials have only linear and interaction terms. It is possible to detect curvature, when points in the centre of the factorial are included (centre points), but curvature cannot be included in the model. Therefore, to reach the optimum, other techniques such as simplex technique or response surface method techniques (RSM) are often used.

Simplex schemes employ a linear algorithm for the search of an optimum on the parameter surface. They are inherently sequential and require a very high precision. They can be effective in quickly reaching the optimum but don't facilitate modeling the parameter space around the optimum.

RSM techniques on the contrary enable non linear models to be developed to describe the parameter surface more accurately as well as providing accurate point prediction to find the optimum. They are usually, but not necessarily, done in a sequential manner, often following a factorial screening design (either a separate screening study or as the first part of the RSM).

Factorial study

In the factorial study, the factors volume (A), catalyst loading (B), equivalents of base (C), temperature (D) and time (E) were examined.

The study was performed with two parallel reactors using the following setup:

Day 1: Reactor A: 2 centre points. Reactor B: 8 reactions at low temperature.

Day 2: Reactor A: 8 reactions at high temperature. Reactor B 2 centre points.

The run order was followed as proscribed with the exception of experiments where it was more convenient due to the limitations of the equipment at hand, i.e. reactions of the same temperature were run on the same day.

Table 1 shows the conditions and the relative yields (calculated from GC-MS chromatograms, assuming full conversion from aldehyde to product, corrected with a calibration factor. This will not lead to a big error because the reaction is very chemoselective. It shows less than approximately 1 % of byproducts such as the homocoupling product or the hydrogenated product.)

Table 1. Parameters and results of the screening study.

Standard order	Run order	Volume mL	Catalyst / eq.	Base / eq.	T / °C	Time / h	Yield / %
1	8	1.6	0.005	1.50	50.0	9.0	16
2	11	5.0	0.005	1.50	50.0	4.0	5
3	6	1.6	0.050	1.50	50.0	4.0	29
4	3	5.0	0.050	1.50	50.0	9.0	49
5	12	1.6	0.005	4.00	50.0	4.0	14
6	2	5.0	0.005	4.00	50.0	9.0	15

7	10	1.6	0.050	4.00	50.0	9.0	30
8	7	5.0	0.050	4.00	50.0	4.0	27
9	20	1.6	0.005	1.50	90.0	4.0	32
10	17	5.0	0.005	1.50	90.0	9.0	69
11	15	1.6	0.050	1.50	90.0	9.0	44
12	9	5.0	0.050	1.50	90.0	4.0	73
13	5	1.6	0.005	4.00	90.0	9.0	34
14	16	5.0	0.005	4.00	90.0	4.0	54
15	19	1.6	0.050	4.00	90.0	4.0	46
16	18	5.0	0.050	4.00	90.0	9.0	75
17	13	3.3	0.028	2.75	70.0	6.5	46
18	4	3.3	0.028	2.75	70.0	6.5	44
19	14	3.3	0.028	2.75	70.0	6.5	46
20	19	3.3	0.028	2.75	70.0	6.5	50

The values were not transformed, which led to a model with a (just) not significant lack of fit, but significant curvature. Often, bounded values such as yield are better transformed using the logit transformation especially when values close to the bounds are achieved. In this case the values were not especially close to 0 or 100% (particularly 100% which is the region of interest) and a transformed model had significant lack of fit. Therefore, a non transformed model was used.

RSM study

In the RSM study, the factors volume (A), catalyst loading (B), temperature (C) and time (D) were examined, following the same experimental principles discussed earlier for the factorial study.

Table 2 shows the results obtained in this study.

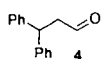
Table 2. Results of the optimization (RSM) study.

Standard order	Run order	Volume^o / mL	Catalyst / eq.	T / °C	Time / h	Byproduct^{a)} / %	Yield^{a)} / %
1	18	5.0	0.0050	90.0	8.0	1.37	57.14
2	16	10.0	0.0050	90.0	8.0	0.67	73.47
3	6	5.0	0.0500	90.0	8.0	1.86	73.47
4	17	10.0	0.0500	90.0	8.0	0.63	79.78
5	5	5.0	0.0050	120.0	8.0	4.68	82.59
6	29	10.0	0.0050	120.0	8.0	1.36	98.64
7	28	5.0	0.0500	120.0	8.0	1.99	98.01
8	20	10.0	0.0500	120.0	8.0	0.57	99.43
9	15	5.0	0.0050	90.0	24.0	2.14	66.44
10	7	10.0	0.0050	90.0	24.0	0.84	86.01
11	3	5.0	0.0500	90.0	24.0	2.45	80.43
12	8	10.0	0.0500	90.0	24.0	0.71	89.79
13	12	5.0	0.0050	120.0	24.0	5.91	94.09
14	27	10.0	0.0050	120.0	24.0	0.87	99.13
15	22	5.0	0.0500	120.0	24.0	0.44	99.56
16	21	10.0	0.0500	120.0	24.0	0.10	99.90
17	1	5.0	0.0275	105.0	16.0	4.38	86.21
18	11	10.0	0.0275	105.0	16.0	0.96	99.04
19	10	7.5	0.0050	105.0	16.0	2.84	88.55
20	25	7.5	0.0500	105.0	16.0	1.18	97.72
21	2	7.5	0.0275	90.0	16.0	1.32	81.28
22	26	7.5	0.0275	120.0	16.0	0.75	99.25
23	13	7.5	0.0275	105.0	8.0	1.35	86.55
24	9	7.5	0.0275	105.0	24.0	1.77	98.23
25	23	7.5	0.0275	105.0	16.0	1.52	93.47
26	19	7.5	0.0275	105.0	16.0	1.54	95.80
27	4	7.5	0.0275	105.0	16.0	1.84	98.16

28	14	7.5	0.0275	105.0	16.0	1.56	97.28
29	30	7.5	0.0275	105.0	16.0	1.78	96.85
30	24	7.5	0.0275	105.0	16.0	1.60	95.47

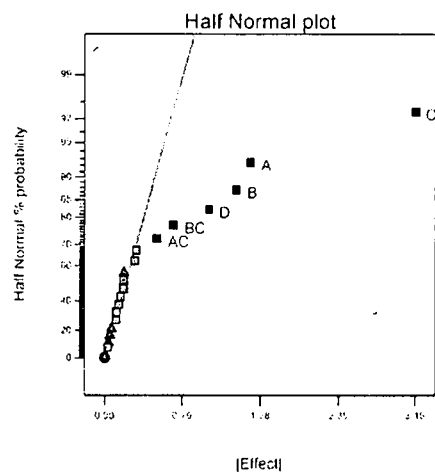
^{a)} For the accuracy of these values see ANOVA in the experimental part.

In this study, we could see that many of the yields obtained came very close to 100 %. Therefore, we chose to no longer ignore traces of the byproduct (the hydrogenated aldehyde **4**). This product was not isolated (identification by GC-MS). Because of its similarity to the product, the same response factor was used which will lead to a certain error; however, the trends shown in the statistical models should not be distorted very much.



For the yield, the values were transformed using the logit function (boundaries 0 to 100); then a quadratic model reduced by the not significant terms was found. For the byproduct, the yields were also transformed using a logit transformation (boundaries 0 to 100). A linear, 2FI or quadratic model would all have had the problem of a significant lack-of-fit. Ideally the model should fit though sometimes in response surface designs it can be found that there is statistically significant lack of fit (due to variation at the centre points being smaller than at surrounding design points and/or because a quadratic model doesn't quite model a flat plateau sufficiently) but it is not possible to identify which cubic terms should be included and the practical difference made by adding them in to the model is very small. Here, the cubic reduced model was chosen, even though it contained aliased terms, since these could be resolved unambiguously when using the assumption that a main factor that is significant would also have more impact in an aliased term than one which is non significant. The cubic model enabled a better fit to the flat plateau¹. As there is a 2-level factorial implicit in a CCD, the data can also be analyzed as a factorial in order to view a half-normal plot and determine the most important factors (Figure 1).

Figure 1. Factorial analysis of the optimization (RSM) study – significant effects.



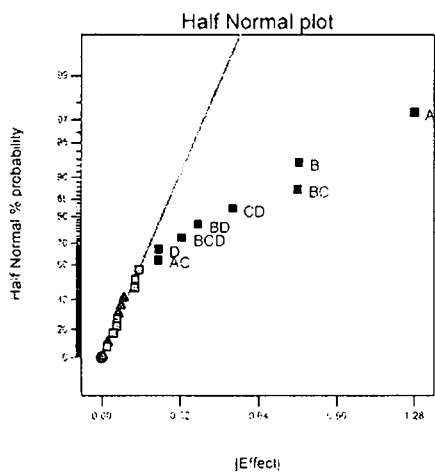
Product - analysis

A = volume

B = catalyst

C = temperature

D = time



Byproduct - analysis

A = volume

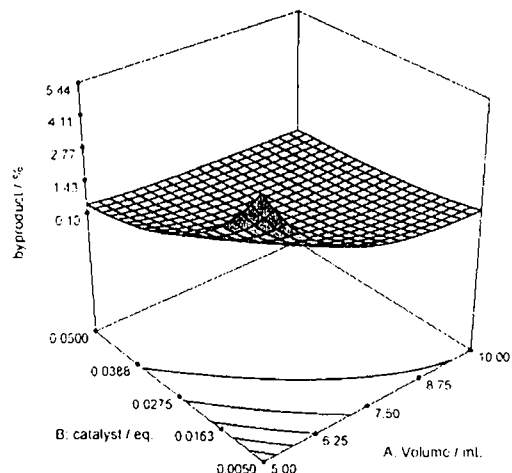
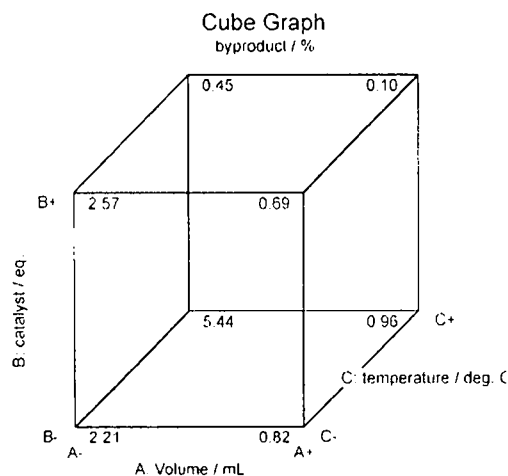
B = catalyst

C = temperature

D = time

The byproduct formation was analysed in a cubic model (Figure 2). There is no positive correlation between amount of byproduct and reaction time or temperature. Lower catalyst loading at high temperature and low volume seem to promote byproduct formation.

Figure 2. Graphical analysis for the byproduct.



$$A^- = 5 \text{ mL}; A^+ = 10 \text{ mL}$$

$$B^- = 0.005 \text{ eq.}; B^+ = 0.05 \text{ eq.}$$

$$C^- = 90 \text{ }^\circ\text{C}; C^+ = 120 \text{ }^\circ\text{C}$$

$$\text{time} = 24 \text{ h}$$

$$\text{temperature} = 120 \text{ }^\circ\text{C}$$

$$\text{time} = 24 \text{ h}$$

The isolated yields are always lower than the predicted yields by GC-MS. The fact that they are substantially lower when the product had to be separated from the starting material (original conditions) probably reflects the significantly longer column which had to be used.

The RSM derived model gives a very accurate prediction of the yields.

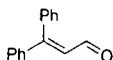
Even with low catalyst loading, excellent yields can be obtained. Yet, if the byproduct is a problem for subsequent reactions, a higher catalyst loading should be chosen.

Experimental

Flash chromatography was performed on silica gel (Merck silica gel 60 F₂₅₄, particle size 0.040 – 0.063 mm, 230 – 400 mesh). NMR spectra were measured on a JEOL delta 400. Chemical shifts (δ_H , δ_C and δ_P) are quoted in parts per million (ppm), referenced to the appropriate residual solvent peak. GC-

MS spectra were recorded on a Agilent GC-MS system (column: HP 190915-433, 5% phenyl methyl siloxane, 30 m * 250 μm * 0.25 μm).

3,3-Diphenylacrylaldehyde 3



All test reactions for the DoE study were carried out in a parallel reactor under inert gas conditions. For the factorial the scale was 1.0 mmol, for the RSM study 0.5 mmol respective of the cinnamaldehyde. The verification studies were done on 1.0 mmol scale respective of the cinnamaldehyde in round-bottom flasks.

The experimental procedure was the same for all reactions² apart from the factors which were intentionally varied: A mixture of NaOAc (123.0 mg to 328.1 mg, 1.5 to 4 mmol, (*n*Bu)₄NBr (322.4 mg, 1.0 mmol), iodobenzene (168 μL , 1.5 mmol) and *trans*-cinnamaldehyde (126 μL , 1.0 mmol) in DMF (1.2 mL to 7.5 mL, equal to $\frac{3}{4}$ of the total volume of added solvent) was heated to the given temperature (50°C to 120 °C). Pd(OAc)₂ (1.1 mg to 11.2 mg, 5 μmol to 50 μmol) was weighed directly in a syringe with which DMF (0.4 mL to 2.5 mL, equal to $\frac{1}{4}$ of the total volume of added solvent) was taken up directly and all air removed. After the catalyst had dissolved completely, this solution was added to the respective reaction mixture in the randomized run order given by the Experimental Design. After the preset reaction time (4 h to 24 h), a sample was taken and analyzed by GC-MS.

For the verification reactions, those samples were added back to the reaction mixture before the reactions were worked up. In these cases, the reaction mixture was poured on NaHCO_3 half-sat. (80 mL) and extracted with EtOAc (3 x 40 mL). The combined organic phases were washed with brine (40 mL), dried over Na_2SO_4 and concentrated in the rotary evaporator. In order to remove further residual DMF, the crude mixture was dried at 60 °C under high vacuum. Then the product was purified by flash chromatography (petrol ether / EtOAc = 20 / 1). For the reactions where there was still starting material present, a column (25 cm, \varnothing 2.5 cm) with an 9 times higher silica loading than the column (17 cm, \varnothing 0.5 cm) for the reactions that had gone to completion had to be used. As the crude reaction mixture had to be taken up in a little EtOAc, into the top of the conditioned column petrol ether had to be pressed in to ensure good separation. The product was isolated as a yellow oil.

$^1\text{H-NMR}$ (400 MHz, CDCl_3): 9.53 (d, $J = 8.3$ Hz, 1H, =CH-CHO), 7.54 – 7.19 (2 m, 10H, Ar-H), 6.60 (d, $J = 8.3$ Hz, 1H, =CH-CHO).

$^{13}\text{C-NMR}$ (100 MHz, CDCl_3): 193.2, 162.0, 139.7, 136.7, 130.6, 130.4, 129.6, 128.6, 128.3, 127.3.

$^1\text{H-NMR}$ (400 MHz, CD_2Cl_2): 9.49 (d, $J = 8.1$ Hz, 1H, =CH-CHO), 7.50 – 7.25 (2 m, 10H, Ar-H), 6.56 (d, $J = 8.1$ Hz 1H, =CH-CHO).

$^{13}\text{C-NMR}$ (100 MHz, CD_2Cl_2): 193.0, 161.9, 149.9, 136.9, 130.7, 130.4, 129.3, 128.6, 128.6, 128.3, 127.4.

MS (EI): 208 (M⁺, 72), 207 (100), 179 (35), 178 (53), 177 (10), 176 (13), 165 (16), 152 (13), 102 (32), 77 (10).

Statistical evaluation of the Screening study (half-factorial): yield

The following paragraph shows a design summary (Table 3), the ANOVA for the selected model (Table 4) for the screening study.

Table 3. Design summary for the half-factorial: yield.

Design: 2 level factorial

Experiments: 20 (of which 4 centre points)

Number of factors: 5

response	units	observations	minimum	maximum	transformation	model
yield	%	20	5.00	75.00	none	2FI

factor	name	units	low actual	high actual
A	volume	mL	1.60	5.00
B	catalyst	eq	0.005	0.050
C	base	eq	1.50	4.00
D	temperature	°C	50.0	90.0
E	time	h	4.00	9.00

Table 4. ANOVA for selected factorial model.

source	sum of square	DF	mean squares	F value	prob > F	Significance*
model	6441.75	4	1610.44	37.31	< 0.0001	significant
A	930.25	1	930.25	21.55	0.0004	significant

<i>B</i>	1122.25	1	1122.25	26.00	0.0002	significant
<i>D</i>	3660.25	1	3660.25	84.81	< 0.0001	significant
<i>AD</i>	729.00	1	729.00	16.89	0.0011	significant
curvature	217.80	1	217.80	5.05	0.0413	significant
residual	604.25	14	43.16			
<i>lack of fit</i>	585.25	11	53.20	8.40	0.0530	not significant
<i>pure error</i>	19.00	3	6.33			
cor total	7263.80	19				

* Significance of factors, curvature and lack of fit tested at 0.05 level.

std. dev.	6.57	R-squared	0.9142
mean	39.90	adj R-squared	0.8897
C.V.	16.47	pred R-squared	0.8249
PRESS	1271.99	adeq precision	17.300

Final equation in terms of actual factors:

$$\text{yield} = +6.13521 - 9.41176 * \text{volume} + 372.22222 * \text{catalyst} + 0.10110 * \text{temperature} + 0.19853 * \text{volume} * \text{temperature}$$

Statistical evaluation of the Optimization (RSM) study

The following paragraph shows a design summary (Table 5), the ANOVA for the selected model (Table 6) for the product yield of the RSM study.

Table 5. Design summary for the RSM study: yield.

Design: Central Composite Design (RSM)

Experiments: 30 (of which 6 centre points)

Number of factors: 4

factor	name	units	low actual	high actual
A	volume	mL	5.0	10.0
B	catalyst	eq	0.005	0.050
D	temperature	°C	90.0	120.0
E	time	h	8.0	24.0

response	units	observations	minimum	maximum
yield	%	30 (20 of which were independent; 10 were measurements taken at a later time from the same sample.)	57.1	99.9

transformation model

Logit (0 to 100) quadratic, backward elimination regression with α to exit = 0.100

Removed terms: C^2 , AD, A^2 , BD, D^2 , CD, AB

Hierarchical terms added after backward elimination regression: none

Table 6. ANOVA for response surface of reduced quadratic model for the product yield.

source	sum square	of DF	mean squares	F value	prob > F	Significance*
model	71.95	7	10.28	49.21	< 0.0001	significant
A	9.72	1	9.72	46.54	< 0.0001	significant
B	7.91	1	7.91	37.85	< 0.0001	significant
C	44.70	1	44.70	214.00	< 0.0001	significant

<i>D</i>	4.99	1	4.99	23.89	< 0.0001	significant
<i>B</i> ²	1.24	1	1.24	5.92	0.0236	significant
<i>AC</i>	1.25	1	1.25	5.97	0.0231	significant
<i>BC</i>	2.15	1	2.15	10.31	0.0040	significant
residual	4.59	22	0.21			
lack of fit	3.54	17	0.21	0.99	0.5578	not significant
pure error	1.05	5	0.21			
cor total	76.54	29				

std. dev.	0.46 ^{a)}	R-squared	0.9400
mean	2.96 ^{a)}	adj R-squared	0.9209
C.V.	15.43 ^{a)}	pred R-squared	0.8889
PRESS	8.51 ^{a)}	adeq precision	29.659

Values are logit transformed.

* Significance of factors tested at 0.1 level (this is a less conservative level than the screening design and part of a generic RSM analysis typically used in GSK - since we expect most factors to have an effect given the design follows a screening study and often ranges are reduced resulting in a reduced signal/noise ratio). Lack of fit tested at 0.05 level.

Final equation in terms of actual factors:

$$\text{Logit}(\text{yield}) = \text{Ln}[(\text{yield} + 0.00)/(100.00 - \text{yield})] = -3.50695 - 0.48761 * \text{volume} - 39.66721 * \text{catalyst} \\ + 0.019337 * \text{temperature} + 0.065807 * \text{time} - 818.30807 * \text{catalyst}^2 + 7.44330\text{E-}003 * \text{volume} * \\ \text{temperature} + 1.08693 * \text{catalyst} * \text{temperature}$$

Statistical evaluation of the RSM-CCD: byproduct

The following paragraph shows a design summary (Table 7), the ANOVA for the selected model (Table 8) for the byproduct yield of the RSM study.

Table 7. Design summary for the byproduct.

response	units	observations	minimum	maximum
byprod.	%	30 (20 of which were independent; 10 were measurements taken at a later time from the same sample.)	0.1	5.9
transformation model				
Logit (0 to 100)	cubic, backward elimination regression with α to exit = 0.100			

Removed terms: AB, B², D², BA², ABD, ACD

Hierarchical terms added after backward elimination regression: AB

Alias list: AB² ~ AC² ~ AD²; BA² ~ BC² ~ BD²; CA² ~ CB² ~ CD², DA² ~ DB² ~ DC². For the model, after reduction, the decision which alias term was to be included relied on the significance of the quadratic part of the term regarding one variable as the linear term is the same for each alias of a term.

Table 8. ANOVA for response surface of reduced cubic model for the byproduct.

source	sum square	of DF	mean squares	F value	prob > F	Significance
model	18.34	17	1.08	101.86	< 0.0001	significant
A	1.21	1	1.21	113.87	< 0.0001	significant
B	2.93	1	2.93	276.74	< 0.0001	significant
C	0.16	1	0.16	15.40	0.0020	significant
D	0.038	1	0.038	3.57	0.0831	significant
A ²	0.099	1	0.099	9.37	0.0099	significant
C ²	1.12	1	1.12	105.74	< 0.0001	significant
AB	0.002307	1	0.002307	0.22	0.6491	not

						<i>significant</i>
<i>AC</i>	0.24	1	0.24	22.81	0.0005	<i>significant</i>
<i>AD</i>	0.11	1	0.11	10.06	0.0080	<i>significant</i>
<i>BC</i>	2.89	1	2.89	273.18	< 0.0001	<i>significant</i>
<i>BD</i>	0.70	1	0.70	66.01	< 0.0001	<i>significant</i>
<i>CD</i>	1.30	1	1.30	122.66	< 0.0001	<i>significant</i>
<i>A²C</i>	0.11	1	0.11	9.98	0.0082	<i>significant</i>
<i>AC²</i>	0.040	1	0.040	3.82	0.0743	<i>significant</i>
<i>C²D</i>	0.15	1	0.15	13.83	0.0029	<i>significant</i>
<i>ABC</i>	0.085	1	0.085	8.07	0.0149	<i>significant</i>
<i>BCD</i>	0.48	1	0.48	45.77	< 0.0001	<i>significant</i>
residual	0.13	12	0.011			
<i>lack of fit</i>	0.093	7	0.013	1.97	0.2372	<i>not significant</i>
<i>pure error</i>	0.034	5	0.006775			
cor total	18.47	29				

* Significance of factors tested at 0.1 level (this is a less conservative level than the screening design and part of a generic RSM analysis typically used in GSK - since we expect most factors to have an effect given the design follows a screening study and often ranges are reduced resulting in a reduced signal/noise ratio). Lack of fit tested at 0.05 level

std. dev.	0.10 ^{a)}	R-squared	0.9931
mean	-4.32 ^{a)}	adj R-squared	0.9834
C.V.	-2.38 ^{a)}	pred R-squared	0.9160
PRESS	1.55 ^{a)}	adeq precision	51.231

^{a)} Values are logit transformed.

Final Equation in Terms of Actual Factors:

$$\text{Logit}(\text{byprod.}) = \text{Ln}[(\text{byprod.} + 0.00)/(100.00 - \text{byprod.})] = -45.49994 + 7.00090 * \text{volume} + 94.43645 * \text{catalyst} + 0.63280 * \text{temperature} - 1.61532 * \text{time} - 0.24592 * \text{volume}^2 -$$

$$\begin{aligned} & 1.99444\text{E-}003 * \text{temperature}^2 - 9.30644 * \text{volume} * \text{catalyst} - 0.10100 * \text{volume} * \text{temperature} - \\ & 4.08000\text{E-}003 * \text{volume} * \text{time} - 0.87799 * \text{catalyst} * \text{temperature} + 5.60747 * \text{catalyst} * \text{time} + \\ & 0.032888 * \text{temperature} * \text{time} + 2.60078\text{E-}003 * \text{volume}^2 * \text{temperature} + 2.68252\text{E-}004 * \\ & \text{volume} * \text{temperature}^2 - 1.59477\text{E-}004 * \text{temperature}^2 * \text{time} + 0.086600 * \text{volume} * \text{catalyst} * \\ & \text{temperature} - 0.064464 * \text{catalyst} * \text{temperature} * \text{time} \end{aligned}$$

¹ There are alternative ways to analyse this which produce slightly different models, but very similar predictions in the region of interest. There may be various views as to alternative strategies, designs or analysis in this particular case. However, those used are good, have been demonstrated to be effective and gave results which highly exceeded the chemists original hopes.

² The amounts of reactants and solvents given refer to the factorial study and were halved accordingly for the RSM study, see main paper.