Supplementary Information

A statistical approach dealing with multicollinearity among predictors in microfluidic reactor operation to control liquid-phase oxidation selectivity

Muhammad N. Siddiquee [a,b], Kaushik Sivaramakrishnan [a*], Yucheng Wu [a], Arno de Klerk [a] and Neda Nazemifard [a]

[a] *Department of Chemical and Materials Engineering,*
12-203 Donadeo Innovation Centre for Engineering,
9211-116 Street NW, University of Alberta, Edmonton, Alberta T6G 1H9, Canada.
[b]*Department of Chemical Engineering and Polymer Science, Shahjalal University of Science & Technology, Sylhet, Bangladesh*
** Corresponding author – Email: sivarama@ualberta.ca; Tel.: +1-780-2002049*

1. Introduction

The manuscript involved experimental investigation using a microfluidic reactor, as well as detailed statistical analysis of the data. In order to provide sufficient detail so that this work could be repeated, the sections dealing with the description of the experimental work and statistical analysis were extensive. It was therefore decided to place most of the detailed descriptions in the Supplementary Information.

2. Experimental

2.1 Materials

In addition to the materials used for the experimental work, some authentic compounds were used to confirm key oxidation products, namely, 1,2,3,4-tetrahydro-1-naphthol (alcohol of tetralin) and α-tetralone (ketone of tetralin). These were used as described before. [1]

2.2 Equipment details: Microfluidic reactor

The microfluidic experimental setup consisted of a microfluidic reactor (Dolomite Microfluidics, Charlestown, MA, USA), syringe pump (KDS-210, KD Scientific, USA), oxygen and nitrogen gas cylinders (Praxair Inc., Edmonton, Canada), pressure transducer (Swagelok, Canada), gas flow meter (Swagelok, Canada), pressure bomb (Swagelok, Canada) and backpressure regulator (Swagelok, Canada), Heidolph MR Hei-Standard hot plate (Model: 505-20000-01-2, Heldolph Instruments, Germany), a surface mounted thermocouple (Model: CO 1, Cement-on Thermocouple, Omega Engineering, Inc., USA), a Flea3FL3-U3-13E4M camera (Point Grey Research Inc., Canada), a Fiber–Lite lamp (Model: 3100, Dolan-Jenner Industries, Inc., USA) and PTFE tubing, 1/16″ OD x 0.8 mm ID (Dolomite Microfluidics, Charlestown, MA, USA).

2.3 Equipment details: Gas Chromatograph

An Agilent 7890A GC-FID equipped with DB-5 MS column 30 m × 0.25 mm × 0.25 μm column was used for quantitative analysis. The injector temperature of the GC was 250 °C and the

split ratio was 10 : 1. Helium was used as a carrier gas, which was flowed through the column at a constant flowrate of 2 mL/min during the experiments. Oven temperature was varied throughout the experiments. Initially, the oven temperature was 75 °C, which was kept constant for 0.5 minutes and then temperature was raised from 75 °C to 325 °C at a rate of 20 °C/min, and finally, the temperature was kept constant at 325 °C for 5 minutes. HPLC grade chloroform was used for sample preparation and hexachlorobenzene was used as an internal standard.

The oxidation products of tetralin were identified by using an Agilent 7820A GC coupled with an Agilent 5977E mass spectrometer. The products were separated on an HP-5 30 m × 0.25 mm × 0.25 μm column which have identical separation characteristics of DB-5 MS column used in GC-FID. The temperature programs of both the GC-FID analyses and GC-MS analyses were the same. Oxidation products were classified as primary (alcohol and ketones of tetralin), secondary (products containing more than one ketone and/or alcohol functional groups) and addition products (products containing at least a dimer having different functional groups). GC-MS spectrum of commercially available alcohol and ketone of tetralin were used to identify the primary oxidation products whereas secondary and addition products were identified using GC-MS spectrum and NIST library. Details of product identification are available from a previous study. [1]

2.4 Calculations

Different hydrodynamic parameters and mass transfer coefficients were calculated from the images captured during experiments in microfluidic reactor as elaborated in our previous paper. [2]

(a) a (gas liquid interfacial area per unit liquid slug volume) was calculated from the dimension of the rectangular channel reactor ($h \times w$) and image analysis of gas bubbles and liquid slugs.
Surface area of gas bubble:

$$S_G = 2\left(wL_{G,\ actual} + hL_{G,actual}\right) + 4\pi((w+h)/4)^2 \quad (1)$$

$$L_{G,\ actual} = L_G - (w+h)/2 \quad (2)$$

Volume of liquid slug:

$$V_L = whL_S + wh[(w+h)/2] - (4/3)\pi[(w+h)/4]^3 \quad (3)$$

Gas liquid interfacial area per unit liquid slug volume:

$$a = S_G/V_L \quad (4)$$

Here, S_G is the surface of the gas bubble, L_G and L_S are the lengths of the gas bubble and liquid slug respectively, and w and h are the width and depth of the reactor channel, respectively.
Approximated radius of the cap of liquid slug:

$$r_{cap} = (w+h)/4 \quad (5)$$

Since geometry formed by the two liquid caps is not a complete sphere, the approximation was made.

(b) U_L (superficial liquid slug velocity) and U_G (gas bubble velocity) were calculated from the distance travelled by the slug and bubble in a particular time. Two phase superficial velocity (U_{TP}) was calculated as follows:

$$U_{TP} = \varepsilon_G U_G + (1 - \varepsilon_G) U_L \quad (6)$$

Here, the volume fraction of gas bubble (ε_G) was calculated as:

$$\varepsilon_G = \frac{V_G}{V_G + V_L} \quad (7)$$

V_L is the volume of liquid slug was calculated according to equation 3.
V_G is the volume of gas bubble, given as:

$$V_G = whL_{G,\ actual} + (4/3)\pi((w + h)/4)^2 \quad (8)$$

(c) Average residence time:
The two-phase superficial velocity (U_{TP}) was divided by the reactor length to calculate the average residence time.

(d) δ (liquid film thickness surrounding a gas bubble) by using the correlations provided by Yun et al. [3] for a rectangular microchannel reactor as follows:

$$\frac{\delta_{max}}{D_h} = 0.39\ We^{0.09} \quad (9)$$

$$\frac{\delta_{min}}{D_h} = 0.02\ We^{0.62} \quad (10)$$

Here, Weber number,

$$We = \frac{D_h U_T^2 \rho_l}{\sigma_l} \quad (11)$$

Hydraulic diameter of the channel (m),

$$D_h = 2[wh/(w + h)] \quad (12)$$

δ_{max} and δ_{min} are the maximum and minimum thicknesses of the liquid film (m), respectively.
U_{TP} (m/s) is the two-phase superficial gas velocity, ρ_l is the density of liquid and σ_l is the surface tension of liquid (N/m).
w and h are the width and depth of the reactor channel, respectively.

(e) $k_L a$ (volumetric mass transfer coefficient, s^{-1}) was calculated from k_L and a. Film theory was applied to calculate k_L [4] as follows:

$$k_L = \frac{D_A}{\delta} \tag{13}$$

Here, D_A is the diffusivity of oxygen in tetralin, δ is the thickness of liquid film surrounding the oxygen bubble.
$k_{L(max)}a$ and $k_{L(min)}a$ were based on the equation 9 and equation 10, respectively.

Calculation of conversion and product selectivity from GC analysis:

(f) Product selectivity was obtained from the relative peak area of the products as follows:

$$\text{Product selectivity (\%)} = \frac{relative\ peak\ area\ of\ specific\ product}{sum\ of\ relative\ peak\ area\ of\ all\ the\ products}\ x\ 100 \tag{14}$$

Ketone-to-alcohol selectivity in primary oxidation products was calculated by dividing ketone selectivity and alcohol selectivity.

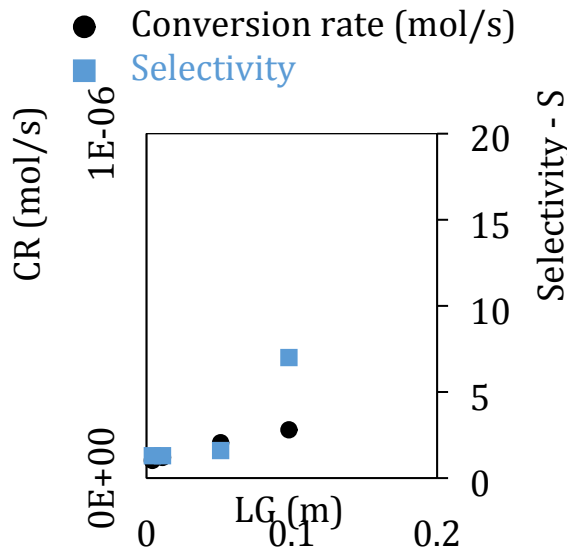
(g) Tetralin conversion was calculated by using GC-FID response factor (detailed calculation of conversion and response factor is given in section 8 of this document). In the selectivity calculations, response factors for products were not used due to the diversity of oxidation products.

3. Regression Modelling Methods

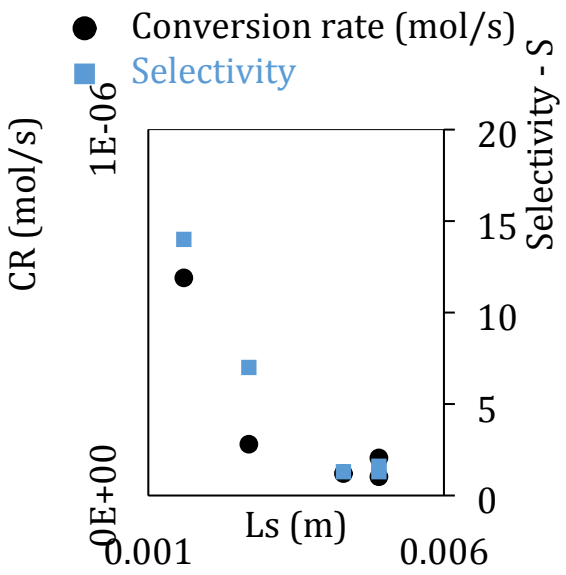
3.1 Assumptions in regression analysis

3.1.1 Linearity –

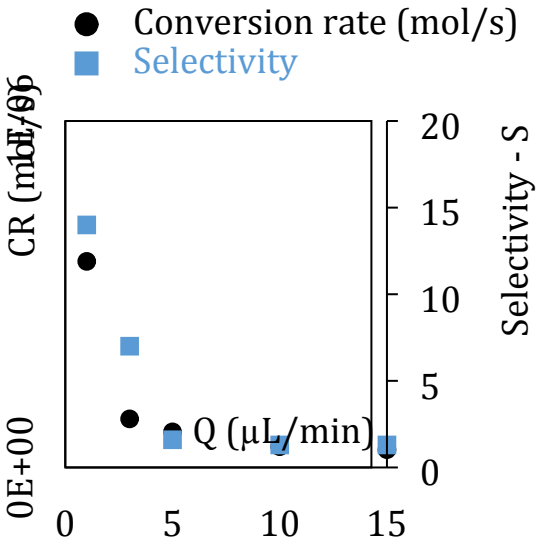
Figure S1 shows the scatter plots for both the outputs varying with the explanatory variables. By looking at this data that was obtained from our previous work, [2] it was realized that not all variations were entirely linear in nature, which was expected. This did not hinder the development of linear regression models as different powers of the variables were also checked for comparing the fits in both the calibration ('Data for model building/calibration' section under "Regression modeling methods" in the manuscript) and validation data sets ('Independent data set for prediction' section under "Results and Discussion" in the manuscript) before arriving at the best fits for further analyses and confirmation.


Conversion rate (mol/s)
Selectivity
CR (mol/s)
1E-06
0E+00
Selectivity - S
20
15
10
5
0
LG (m)
0
0.1
0.2

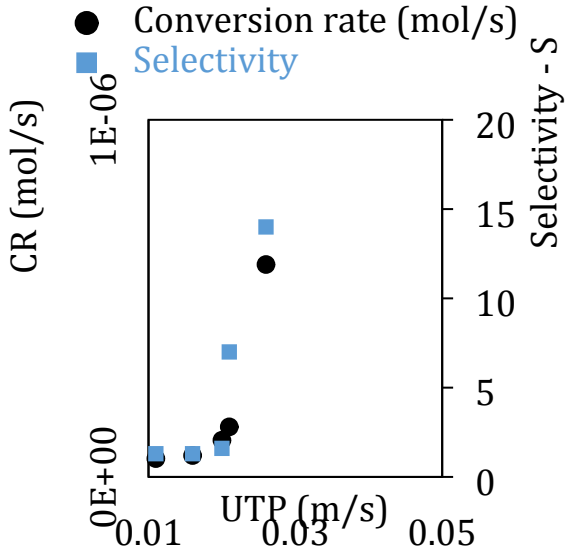
(a)


Conversion rate (mol/s)
Selectivity
CR (mol/s)
1E-06
0E+00
Selectivity - S
20
15
10
5
0
Ls (m)
0.001
0.006

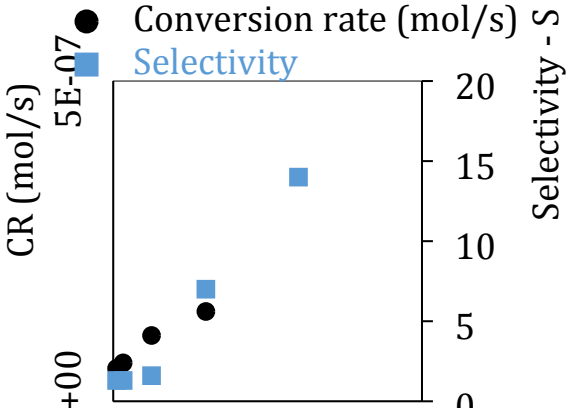
(b)


Conversion rate (mol/s)
Selectivity
CR (mol/s)
0E+00
Selectivity - S
20
15
10
5
0
Q (µL/min)
0
5
10
15

(c)


Conversion rate (mol/s)
Selectivity
CR (mol/s)
1E-06
0E+00
Selectivity - S
20
15
10
5
0
UTP (m/s)
0.01
0.03
0.05

(d)


Conversion rate (mol/s)
Selectivity
CR (mol/s)
5E-07
Selectivity - S
20
15
10
5
0

(e)

Figure S1. Scatter plots for tetralin conversion rate and oxidation product selectivity versus: (a) length of gas bubble (L_G); (b) length of liquid slug (L_S), (c) tetralin flow rate (Q), (d) two-phase superficial velocity (U_{TP}), (e) gas-liquid interfacial area (a).

3.2 Software employed

The information regarding the version of the software package as well as the operating system that it was run on is provided in the manuscript.

3.3 Relevant formulae and statistical calculations

3.3.1 Pearson's correlation coefficient (r) –

For a dependent variable defined as a vector $y = \{y_i \in E, i = 1, 2, .., n)\}$ consisting of n observations (E is a one-dimensional vector space), the correlation with an explanatory variable $X = \{x_i \in E, i = 1, 2, .., n)\}$ also comprising n observations is given in equation 15 as:

$$r_{X-y} = \frac{\sum_{i=1}^{n}(x_i - x_{mean})(y_i - y_{mean})}{\sqrt{\sum_{i=1}^{n}(x_i - x_{mean})^2}\sqrt{\sum_{i=1}^{n}(y_i - y_{mean})^2}} \quad (15)$$

where x_{mean} and y_{mean} are the average of all X and y observations, respectively.
Since the definition describes the correlation as the average of the product of mean-subtracted random variables, 'product-moment' is appended to the name. The square of the bivariate correlation coefficient gives the proportion of shared variance between the two variables (also same as r^2 for simple linear regression).

3.3.2 Significance of a correlation –

Determination of the significance of a correlation involves formulation of a null ($H_0: \rho = 0$) and an alternative hypothesis ($H_A: \rho \neq 0$), where ρ is the population correlation coefficient between the two variables. It is assumed that the sampling distribution of the test statistic (t-statistic is used here) of the Pearson's correlation coefficient follows a t-distribution with 3 degrees of freedom (n - 2) under the null hypothesis in our case. The t-statistic in this case takes the form:

$$t = \frac{r - \rho}{\sqrt{\left(\frac{1 - r^2}{n - 2}\right)}} \tag{16}$$

However, since we do not have the population data, we can only estimate the population correlation in terms of the sample coefficients. The null hypothesis states that the population correlation coefficient is near to 0 while the alternative hypothesis is that it is significantly different from 0, implying a linear relationship between the variables in question. Finally, the p-value is the probability (simply, area under the t-distribution curve) of obtaining future values of the standardized form (t-statistic) of r at least as extreme as that observed value for the existing correlation in the direction of the H_A, still assuming that the null hypothesis was true. In other words, if it is quite low, there is lesser possibility that the correlation coefficient is near to zero, i.e. null can be rejected.

3.3.3 <u>SLR and MLR models</u> –

The simple linear regression (SLR) models in Table 7 (S1 – S5) and Table 9 (SS1 – SS5) of the manuscript can be viewed as given by equation 17:

$$y = b_0 + b_1 X + \varepsilon \tag{17}$$

where b_0 is the intercept (constant) term and b_1 is the regression coefficient estimate (unstandardized) corresponding to the single explanatory variable X; ε is the normally distributed error with mean as 0 and variance σ^2 ($\varepsilon \sim N(0,\sigma^2)$). In case of a quadratic fit, b_1 would be the coefficient estimate of the quadratic term and the linear term is excluded. Again, this is because of the limited number of observations available in the calibration set. When X and y are standardized, the intercept term disappears and the standardized coefficient equals the simple correlation coefficient between X and y.

The expression for multiple linear regression (MLR) with two variables, such models M1 – M4 (Table 7 in manuscript) and SM1 – SM4 (Table 9 in manuscript) is given in equation 18 as:

$$y = b_0 + b_1 X_1 + b_2 X_2 + \varepsilon \tag{18}$$

where b_0 is the intercept, b_1 and b_2 are the unstandardized slope estimates for the two explanatory variables X_1 and X_2 each consisting of n observations, respectively. Here again, ε is normally distributed as given by $N(0,\sigma^2)$. The standardized coefficients in MLR do not equal the respective simple correlation coefficients of the predictors with y as in SLR, but depend on both the simple correlations between y and X as well as on the correlation between X_1 and X_2.

The standard errors for b_1 and b_2 are given as:

$$SE_{b_1} = \sqrt{\frac{SSE}{(n-k-1)\left(\sum_{i=1}^{n}(X_{1i}-X_{1mean})^2\right)\left(1-r_{X_1{}^2-X_2}\right)}} \tag{19}$$

$$SE_{b_2} = \sqrt{\frac{SSE}{(n-k-1)\left(\sum_{i=1}^{n}(X_{2i}-X_{2mean})^2\right)\left(1-r_{X_1{}^2-X_2}\right)}} \tag{20}$$

where SSE is the sum of squared errors of the regression model and is directly proportional to the standard deviation of the residuals; k is the number of explanatory variables. SSE is defined as:

$$SSE = \sum_{i=1}^{n}\left(y_i - y_i^*\right)^2 \tag{21}$$

and $r_{X_1-X_2}$ is the Pearson correlation coefficient between X_1 and X_2.

Like with the Pearson's correlation, the regression coefficient estimates for the MLR and SLR are tested for significance with the help of a t-statistic defined as:

$$t_{b_i} = \frac{b_i}{SE_{b_i}} \tag{22}$$

where t_{b_i} is the t-statistic corresponding to the i^{th} coefficient estimate (b_i).

3.3.4 <u>F-Statistic</u> –

This statistic is given by equation 23 as:

$$F = \frac{MSM}{MSE} \tag{23}$$

where MSM and MSE are the mean squared regression model and mean squared error defined as:

$$MSM = \frac{\sum_{i=1}^{n}\left(y_i^* - y_{imean}\right)^2}{k} \tag{24}$$

$$MSE = \frac{\sum_{i=1}^{n} \left(y_i - y^*_i\right)^2}{n - k - 1} \tag{25}$$

where y^*_i is the predicted value of each observation in y through regression.

It follows an F-distribution with k and $n - k - 1$ degrees of freedom (DF). The computed F-statistic for a regression model is compared to the critical value obtained from the F-table for 5% level of significance corresponding to (1, 3) DF for SLR and (2, 2) DF for MLR in our study.

The second type of F-statistic used for testing the significance of the addition of a second variable to an existing SLR model in improving the r^2 of the existing SLR model. Let R^2_{MLR} be the coefficient of determination of the MLR model including the added second variable, R^2_{SLR} be the coefficient of determination of the already existing SLR model, k_{MLR} and k_{SLR} be the number of predictors in the MLR and SLR respectively. Then, the F-statistic distributed with ($k_{MLR} - k_{SLR}$) and ($n - k_{MLR} - 1$) DF is given by:

$$F_{inc} = \frac{(R^2_{MLR} - R^2_{SLR})/(k_{MLR} - k_{SLR})}{(1 - R^2_{MLR})/(n - k_{MLR} - 1)} \tag{26}$$

3.4 RMSE and R^2

It is important to note that r^2 for SLR is also equal to the squared Pearson's correlation between input and output variable but R^2 in MLR is the proportion of variance explained in y due to both explanatory variables in our case. However, it can be calculated as the correlation between the predicted and experimental output values also. The RMSE and R^2 are given as:

$$RMSE = (MSE)^{\frac{1}{2}} = \left(\frac{1}{n - k - 1}\sum_{i=1}^{n} |y_i - y^*_i|^2\right)^{\frac{1}{2}} \tag{27}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y^*_i\right)^2}{\sum_{i=1}^{n}\left(y_i - y_{imean}\right)^2} \tag{28}$$

3.5 Data used for regression

The manuscript provides information about the calibration dataset used for constructing the regression models, i.e. the calibration set.

4. Results and Discussion

4.1 Bivariate correlations among input and output variables

The manuscript provides sufficient details about the analysis and interpretations of the correlations between the different variables investigated in this study.

4.2 Multicollinearity Diagnostics

Two types of diagnostics are used to confirm the presence of multicollinearity in the data and the degree of impact on regression. These are the following:

1. *Variance Inflation Factor (VIF)* [5]

$$VIF = \frac{1}{1 - R_{x_1 x_k}^{2}} \tag{29}$$

where $R_{x_1 x_k}^{2}$ is the coefficient of determination when the variable x_1 is regressed on x_k, which represents the set of all other explanatory variables except x_1. A higher value of VIF indicates higher correlated variables. VIF would be 1 in a simple regression and higher in the case of multiple regression with collinear variables.

As with Pearson's correlation coefficient, there is a caution for false diagnosis of multicollinearity with VIF as well since there is no consensus on the threshold value. [6] Kutner et al. [7] suggest a minimum value of 10 while Vatcheva et al. [8] demonstrated that even a value of < 5 could be problematic. More than the absolute value, a change in VIF magnitude towards the higher side could provide evidence towards multicollinearity, which is what is pursued in this study by comparing multiple regression models with the simple regression counterparts as detailed in further sections in the manuscript. In addition, VIF can also be compared with $\frac{1}{1 - R_{model}^{2}}$ to know whether the correlation between the regressors is stronger than the overall regression model. [9]

2. *Eigen values (EV) and Condition Index (CI)*

The sum of the eigenvalues of the correlation matrix (obtained through eigenvalue decomposition) will equal the number of explanatory variables in the system but the distribution of the eigenvalues across the dimensions of the matrix would point towards the presence or absence of linear dependencies. [10] If the variables are linearly independent, all eigenvalues will equal unity and in the case of correlated variables, certain dimensions would show eigenvalues close to 0. The latter situation indicates that the regression parameter estimates when regressed using these input variables would be very sensitive to changes in the data. Condition index (CI) helps in amplifying the unequal distribution of the eigenvalues and is given in equation 30 as:

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_i}} \tag{30}$$

where λ_{max} and λ_i are the maximum and the i^{th} eigen value respectively.

According to Midi et al. [11], if the CI falls below 15, then multicollinearity is not a serious concern. Johnston [12] proclaimed of inconsequential collinearity till $CI < 20$. Further, the detection process is also assisted by the variance decomposition proportions for each predictor, i.e. the proportion of variance for the regression coefficient estimates of each input variable that belongs to every dimension. Significantly correlated variables would have higher variance proportions concentrated on the same eigenvalue dimension. We have considered this aspect in our study also. Another diagnostic that has been reported in the literature but used on lesser number of occasions is the determinant of the correlation matrix, where a lower value indicates multicollinearity. It was not used in this study.

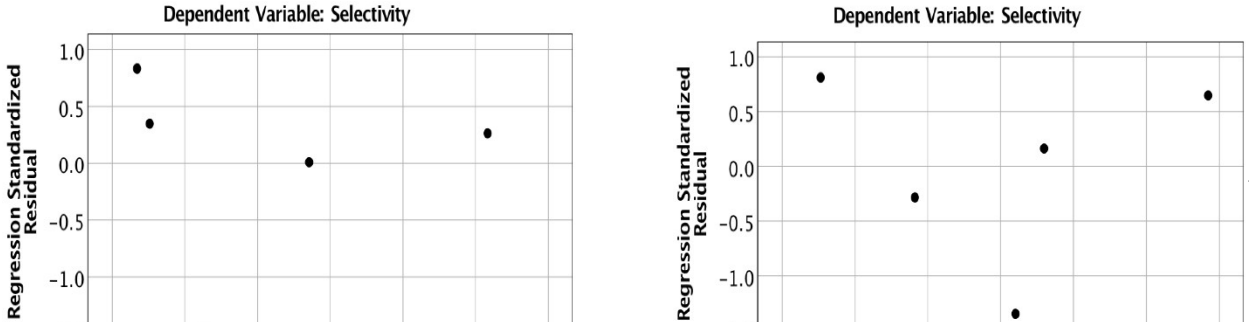
4.3 Validation dataset

This corresponds to new experimental data collected by oxidizing tetralin at 150 ˚C and 90 kPa (gauge) injected at flowrates different to the calibration set. This was solely intended towards this study and the purpose of this data was to find the best fits for the explanatory variables with the outputs through the SLR models by validating the output predicted values at the new flowrate conditions. RMSEP and r_P^2 values were used as performance evaluation measures for deciding the best fits.

4.4 SLR models with S as outcome and residual plots

When the output of the SLR was selectivity, the prediction errors for the best fit regression models with L_G, U_{TP} and Q were larger than that for a by 107%, 229% and 243%, respectively. Plus, the averaged RMSEP for all the fits combined was the smallest for a (1.35) with the errors for U_{TP}, L_G, Q and L_S being 55%, 168%, 214% and 431% higher than a. Similar to the model with CR as the outcome, the significance of a and L_G terms were maximum with almost equivalent p-values of 0.002 and 0.003, respectively. In addition, the amount of variance in the outcome explained with respect to the calibration set by the best fits of the explanatory variables were in the order: a (0.98) > L_G (0.96) > U_{TP}^2 (0.79) > Q (0.58). Interestingly, r_C^2 for a linear fit with L_S as the explanatory variable was 0.93 but it was not considered for analysis because of the poor performance in predicting the validation set selectivity data. Lowest value of calibration error averaged for all the fits (including exponential) was obtained for a (RMSEC average = 1.63) while the corresponding values for L_G, L_S, U_{TP} and Q were higher than that for a by 12%, 15%, 83% and 188%, respectively.

Figure S2 shows the residual vs. predicted value plots for the SLR models with selectivity and conversion rate as the output. It mainly helps in verification of the assumption of equal variance for validity of the regression model. Equal variance of the residuals essentially means that the observed experimental data points are spread out above and below the regression curve that represents the predicted values in a symmetric fashion and within the standard deviation of the residuals (which is also used as the estimate of the standard error for the population). Further interpretation of the plots are summarized in the 'Residual Plots' section of the manuscript.



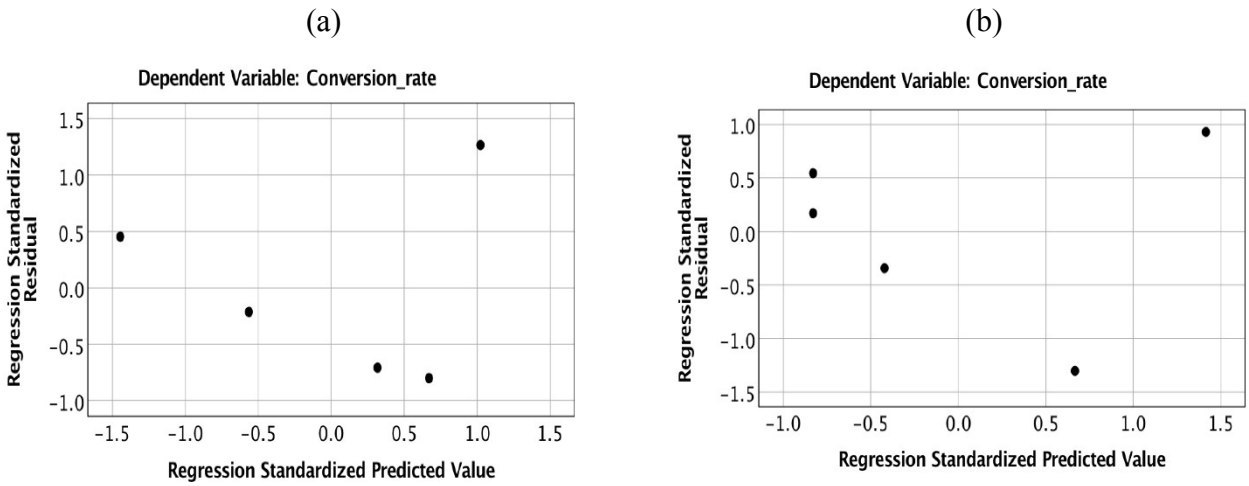


Figure S2. Residual vs. predicted value plots in standardized forms for SLR of: (a) S on a; (b) S on U_{TP}^2; (c) CR on Q and (d) CR on L_S.

4.5 Comparison of MLR with SLR models

In this section, the regression coefficient estimates of the predictors, their standard errors and the significance of each term for the developed SLR models are compared with four MLR models involving gas-liquid interfacial area combined with the other 4 input variables one at a time, for both the outputs.

4.5.1 Effect of explanatory variables on tetralin conversion rate

Here, the analysis of the MLR models M3 and M4 (Table 7 in manuscript) involving a^2 & Q and a^2 & L_S in comparison with the respective SLR models are provided (Table S1).

Table S1. Parameter estimates, significances and output variances of SLR models with CR as the output and individual explanatory variables as the input. Compare the results with Table 7 in the manuscript that shows the parameter estimates for the MLR models.

Model [1]	Var. involved [2]	b_n [3]	SE [4]	Pr > $\lvert t_{crit} \rvert$ [5]	Std. coeff. *	R_C^2	F [6]	VIF	EV [7]	CI	Variance Proportions

											b_0	**T_1**
S1	b_0	5.2E-8	1.7E-8	.056			210.7 (.001)	1	1.56	1.00	.22	.22
	a^2 (T_1)	5.9E-18	4.1E-19	.001	.993	.986			.438	1.89	.78	.78
S2	b_0	-8.6E-8	7.9E-8	.356			17.87 (.024)	1	1.82	1.00	.09	.09
	U_{TP}^3 (T_1)	.0343	.0081	.024	.925	.856			.176	3.22	.91	.91
S3	b_0	-4.0E-9	5.0E-8	.941			30.25 (.012)	1	1.70	1.00	.15	.15
	L_G (T_1)	2.6E-6	4.7E-7	.012	.954	.910			.296	2.40	.85	.85
S4	b_0	3.8E-7	1.4E-7	.079			2.63 (.203)	1	1.80	1.00	.10	.10
	Q (T_1)	-2.8E-8	1.7E-8	.203	-.683	.467			.199	3.01	.90	.90
S5	b_0	6.8E-7	1.8E-7	.034			8.00 (.066)	1	1.94	1.00	.03	.03
	L_S (T_1)	-1.3E-4	4.7E-5	.066	-.853	.727			.058	5.77	.97	.97

[1] Prefix 'S' corresponds to simple linear regression models.

[2] b_0 represents the constant or intercept term in the regression equation. T_1 indicates the respective input variables used to specify the variance proportions associated with that variable.

[3] Regression coefficient or slope estimates.

[4] Standard errors associated with the respective coefficient estimate.

[5] p-value indicating the probability that the t-statistic for the term is greater than the critical value which depends on the degrees of freedom (DF) of the model. For SLR, DF = 3 & $t_{crit} = 3.182$.

[6] F-statistic, distributed as F with 1 & 3 DF for SLR ($F_{critical} = 10.12$). Statistic for the test of overall R^2 of the model. Value in bracket is the p-value for the significance of the overall model.

[7] Eigenvalue for the respective dimension.

* Standardized coefficient for SLR is the same as the zero-order Pearson's correlation coefficient between the variable and the output.

4.5.1.1 Effect of gas-liquid interfacial area and length of gas bubble

Sufficient details are provided in the manuscript for this section.

4.5.1.2 Effect of gas-liquid interfacial area and two-phase velocity

Here again, the reader is referred to the manuscript for the analysis of this section.

4.5.1.3 Effect of gas-liquid interfacial area and injection flowrate of tetralin

All the tables referred to in this section and the next, belong to the main manuscript so that it is not repeated at every point of referral. From Table 8, it can be recognized that a^2 and Q share the least correlation (-0.704) with each other compared to all other combinations of explanatory

variables with a^2. This is supported by the VIF values from model M3, which was the lowest among the MLR models M1-M4 (Table 7). Though the VIF value of 2 was larger than that for the SLR models S1 and S4, the effect of collinearity on the MLR model was not as much a problem in this case as was with the previous 2 combinations. This was because model S4 itself could explain only 46.7% of the variance in CR since the Pearson's correlation between Q and CR was on the lower side (-0.683).

So, it was quite lucid that a^2 was more adept at influencing the output variation than Q and that the combined model involving both these predictors would not inflate the standard errors of the respective regression coefficients by a huge amount. This premise was evident from the results in Table 7 as the standard error for the regression coefficient estimate of a^2 increased only by 69% (compare with 377% and 231% in the previous cases). The slope estimate was still significant in the MLR model with a p-value of 0.013. The presence of some extent of collinearity, between the predictors caused a sign reversal in the regression coefficient estimate of Q but surprisingly, the standard error of the coefficient estimate decreased by 73%. The change in the sign of the slope estimate for Q rendered it insignificant (p-value = .818), though the overall model was still significant with F (2, 2) = 72.65 and p-value = 0.014.

There was a marginal increase of 0.011 and 0.203 units in the standardized coefficient of Q in M3 as compared to U_{TP}^3 in M2 and L_G in M1, respectively. It is more a consequence of a lesser correlation between a^2 and Q than the individual contribution of Q itself because the standardized coefficient of a^2 was still much higher (1.014). A CI value of 5.49 indicated that collinearity was not a serious concern as the increase from the individual models (S1 and S3) was not as large as in the previous cases. Variance proportions of 71% and 90% provided further evidence of the diminished presence of collinearity in this case as compared to the previous combinations of predictors.

Examination of Table 8 for CR also reveals that the overall R^2 of model S1 is not significantly improved by adding Q (F (1, 2) = 0.00). On the other hand, the incremental effect of adding a^2 to model S4 (involving Q as the explanatory variable) was magnified with an F-statistic value of 74.14, which was much higher than $F_{critical}$ = 18.51. Moreover, the partial correlation between a^2 & CR while controlling for Q was 0.987, which implied that Q did not have a serious impact on the relationship between a^2 and CR, unlike the previous two situations. The correlation between Q and CR decreased by 73% (zero order correlation = 0.683; partial correlation = 0.182), while controlling for a^2. The effect of a^2 on the relationship between U_{TP}^3 & CR and L_G & CR was much greater as it lowered the correlations by 94% and 137%, respectively.

4.5.1.4 Effect of gas-liquid interfacial area and length of liquid slug

The first observation from Table 7 with a^2 and L_S as the predictors in model M4 was that the model produced the highest R_C^2 (0.997) with an overall significance of 0.003, which was the highest among other MLR models for CR explored in this study. It must also be noted that the pairwise correlation coefficient between a^2 and L_S (-0.905) was stronger than between a^2 & Q but weaker than the combinations of L_G and U_{TP}^3 with a^2. VIF showed a value of 6.00 for the variables involved

and reflected this observation. Moreover, a major proportion of the variances in the regression coefficient estimates for a^2 (87%) and L_S (100%) was centered on the third eigenvalue dimension, thus indicating sufficient linear dependency between the regressors. The % variance proportion for the regression coefficient of a^2 in the third dimension was still lesser than in the case of M1 and M2.

Comparing M4 and S5, the sign of the slope estimate for L_S turned positive from being negative in S5, possibly due to the effect of collinearity. Interestingly, the standard error for this coefficient decreased by 72% from S5, but still the lower bound value at the 95% confidence interval was negative (-1.6E-5). Although the L_S term was insignificant, its p-value was lower than that of L_G, U_{TP}^3 and Q in models M1, M2 and M3, respectively. On the other hand, the standard error for the coefficient estimate of a^2 hiked by a paltry 24%, while the regression coefficient estimate increased by 23%, thus rendering the term still highly significant (p-value = 0.005). The relative of importance of a^2 in influencing the outcome was further evident with the larger standardized coefficient (1.220) in comparison with that of L_S (0.251). It was intriguing to note that the standardized coefficient of L_S was greater than Q, U_{TP}^3 and L_G in M3, M2 and M1 models by 0.221, 0.232 and 0.424 units, respectively.

In addition, augmenting L_S to model S1 resulted in maximum value of F-statistic (7.33 –Table 8) compared to the other three MLR models, though it was still insignificant ($F_{critical}$ (1, 2) = 18.51). The addition of a^2 to L_S caused the F-statistic to swell up to 180.00, portending that the incremental effect of a^2 over L_S in improving the model R^2 was the maximum out of all MLR models. Notably, the partial correlation of L_S with CR, controlling for a^2 increased to 0.903 from the absolute zero order correlation of 0.853. Also, there was a petty increase of 0.002 units from the zero-order correlation between a^2 and CR to the partial correlation between them, while controlling for L_S.

4.5.2 Effect of input variables on oxidation product selectivity

Here, the analysis of the MLR models SM3 and SM4 (Table 9 in manuscript) involving a & Q and a & L_S in comparison with the respective SLR models are provided (Table S2).

Table S2. Parameter estimates, significances and output variances of SLR models with S as the output and individual explanatory variables as the input. Compare the results with Table 9 in the manuscript that shows the parameter estimates for the corresponding MLR models.

Model [1]	**Var. involved** [2]	$\mathbf{b_n}$ [3]	**SE** [4]	**Pr > |t_{crit}|** [5]	**Std. coeff.** *	R_C^2	**F** [6]	**VIF**	**EV** [7]	**CI**	**Variance Proportions**	
											b_0	$\mathbf{T_1}$
SS1	b_0	.238	.614	.725					1.70	1.0	.15	.15

	a (T_1)	4.5E-5	4.0E-6	.002	.988	.977	125.4 (.002)	1	.302	2.4	.85	.85
SS2	b_0	.199	.790	.817					1.70	1.0	.15	.15
	L_G (T_1)	64.71	7.44	.003	.981	.856	70.71 (.003)	1	.296	2.4	.85	.85
SS3	b_0	-3.953	2.96	.275					1.90	1.0	.05	.05
	U_{TP}^2 (T_1)	23740.8	7020.5	.043	.890	.792	11.44 (.043)	1	.103	4.3	.95	.95
SS4	b_0	10.10	3.13	.048					1.80	1.0	.10	.10
	Q (T_1)	-.744	.369	.137	-.759	.576	4.07 (.137)	1	.199	3.0	.90	.90
SS5	b_0	18.50	2.20	.004					1.94	1.0	.03	.03
	L_S (T_1)	-3658.2	562.3	.007	-.966	.934	42.33 (.007)	1	.058	5.8	.97	.97

[1] 'SS' corresponds to simple linear regression models with S as the output.

[2] b_0 represents the constant or intercept term in the regression equation. T_1 indicates the respective input variable used to specify the variance proportions associated with that variable.

[3] Regression coefficient or slope estimates.

[4] Standard errors associated with the respective coefficient estimate.

[5] p-value indicating the probability that the t-statistic for the term is greater than the critical value of t-stat, that depends on the degrees of freedom (DF) of the model. For SLR, DF = 3 & t_{crit} = 3.182.

[6] F-statistic, distributed as F with 1 and 3 DF for SLR ($F_{critical}$ = 10.12). Statistic for the test of overall R^2 of the model. Value in bracket is the p-value for the significance of the overall model.

[7] Eigenvalue for the respective dimension

* Standardized coefficient for SLR is the same as the zero-order Pearson's correlation coefficient between the variable and the output.

4.5.2.1 Effect of gas-liquid interfacial area and length of gas bubble

Sufficient details are provided in the manuscript for this section.

4.5.2.2 Effect of gas-liquid interfacial area and two-phase velocity

Here again, the reader is referred to the manuscript for the analysis of this section.

4.5.2.3 Effect of gas-liquid interfacial area and injection flowrate of tetralin

Similar to the previous discussion, all references to tables in this section and the next belong to the main manuscript. Like in model M3 where the correlation between a^2 and Q was minimum among all other combinations, the simple correlation between a and Q was the least (-0.839) as compared to other correlations with a (Table 8). This was also reflected in the low value for VIF

(3.37) and CI (8.32). Although the absolute value of CI was lower than the thumb rule cut off (11), it was ~3.5 times and ~2.5 times higher than in SS1 and SS4 models (Table 9 and Table S2). The variance proportions for the regression coefficients of a (87%) and Q (90%) were more for the 3rd dimension of the eigenvalues but were lesser than that for SM1 and SM2, indicating moderate collinearity. The regression coefficient estimate of a did not experience any dramatic change in SM3 as compared to SS1, as a consequence of which the coefficient was significant in the MLR model (p-value = 0.008).

As with model M3, the slope estimate for Q became positive from being negative (in SS4) with a decrease in standard error by 72%. The coefficient was still insignificant (compare with SS4), but the overall model SM3 was significant with a F (2, 2) = 145.9 and p-value of 0.007. Moreover, the due to the low variance explained by Q in the output (r^2 = 0.576 for SS4), the addition of a to Q had a significant contribution in improving the r^2 (F (1, 2) = 119.14 – Table 8). On the other hand, Q did not produce a significant increase to the overall R^2 of SS1 when added to a as apparent from a much lower F-statistic value of 4.57 (F_{crit} = 18.51). Additionally, the partial correlation of a and S, while controlling for Q increased to 0.992 from the zero-order correlation of 0.988. Although the presence of a did not impact the relationship between Q and S in a negative way (value of 0.842), the partial correlation between Q and S was still lesser than that between a and S.

4.5.2.4 Effect of gas-liquid interfacial area and length of liquid slug

There was evidence for a strong linear dependency between a and L_S from the Pearson's simple correlation coefficient value of -0.945 (Table 8). This value was very similar to that between a and U_{TP}^2 but the direction of the correlation was opposite. It was interesting to note that though the VIF (9.35 –Table 9) in model SM4 was almost equal to that of model SM2, the value of CI for SM4 was higher than that of SM2, which only indicated that collinearity could impact regression negatively. 93% and 100% of the variance proportions of a and L_S were focused on the third eigenvalue dimension, like in previous instances.

The presence of collinearity between the predictors did not affect the sign of regression coefficient estimate of L_S but increased its standard error by 67% in SM4 as compared to SS5. It can be seen from Table 9 that both regression coefficients (for a and L_S) were insignificant in model SM4 but the overall model was still significant (p-value = 0.013, R^2 = 0.987). Thus, it made sense to interpret the slope estimates relative to each other because the overall regression was significant in all MLR models with S or CR as the output variable. It can be inferred that L_S affected the relationship between a and S more than other predictors because of the decreased standardized coefficient of a in SM4 (0.700) as compared with other MLR models. But this was much higher than that for L_S (-0.305), which signaled that S was more sensitive to changes in a than in L_S.

The partial correlations also tell a congruent story, giving a value of 0.894 for the relationship between a and S, while controlling for L_S and decreased from the zero-order value by 10%. On the contrary, a affected the link between L_S and S on a greater scale as the reduction from zero-order to partial correlation was 32% (-0.966 to -0.655). The improvement in r^2 of both models SS1 (0.977) and SS5 (0.934) to SM4 (0.987) caused by addition of L_S to a (F-stat = 1.54) and a to L_S

(F-stat = 8.15), respectively were both insignificant (F_{crit} = 18.51). But on a relative measure, a contributed more significantly to improve the model r^2 due to the higher value for F-statistic (Table 8).

5. Physicochemical Properties of Tetralin and Oxygen

Physicochemical properties of tetralin and oxygen used for the calculation shown in the study are listed in Table S3. Most of the properties are interpolated from the values obtained from the reported literatures.

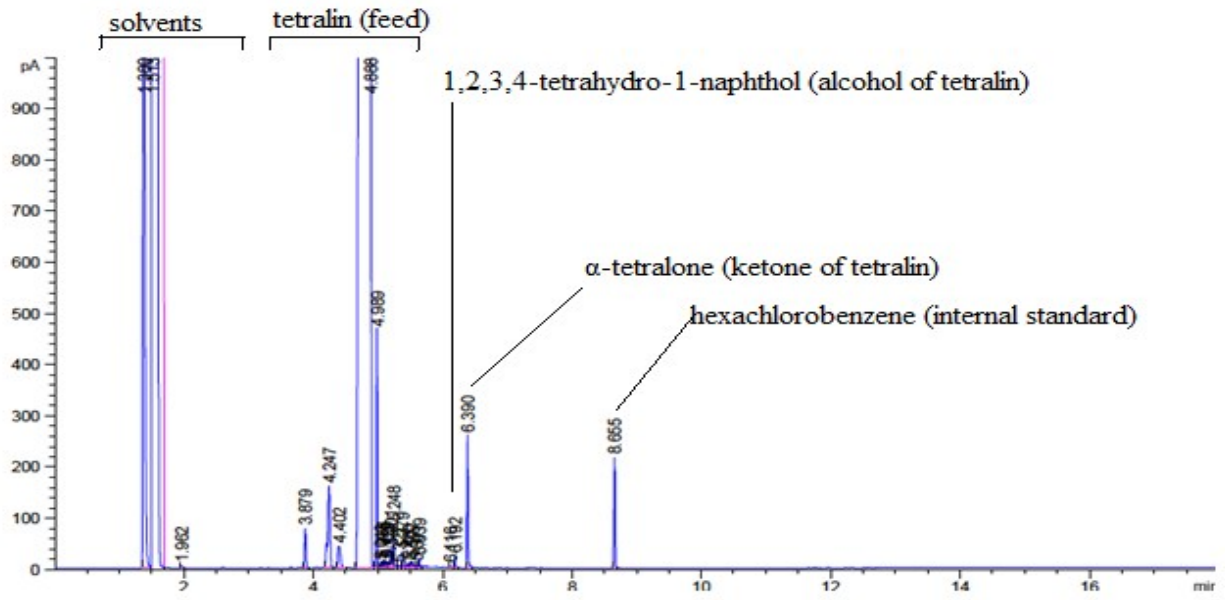
Table S3. Physicochemical properties of tetralin and oxygen at different experimental conditions

	Tetralin				Oxygen			
T (^{0}C)	**Density (kg/m^3)** [1]	**Surface tension [a] (N/m)** [4]	**Dynamic viscosity (Pa.s)** [1]	**Kinematic viscosity (m^2/s)** [1]	**Dynamic viscosity (Pa.s)** [3]	**Density [b] (kg/m^3)**	**Kinematic viscosity (m^2/s)** [3]	**D_A (m^2/s)** [2]
25	966	0.0351	1.17E-03	1.21E-06	2.15E-05	2.36E+00	9.08E-06	2.73E-09
120	887	0.0257	6.36E-04	7.17E-07	2.59E-05	1.79E+00	1.44E-05	2.14E-08
130	879	0.0248	5.84E-04	6.65E-07	2.64E-05	1.75E+00	1.51E-05	2.52E-08
140	871	0.0238	5.33E-04	6.13E-07	2.68E-05	1.71E+00	1.57E-05	2.93E-08
150	862	0.0228	4.83E-04	5.61E-07	2.73E-05	1.67E+00	1.64E-05	3.39E-08
160	854	0.0218	4.34E-04	5.08E-07	2.78E-05	1.63E+00	1.71E-05	3.90E-08

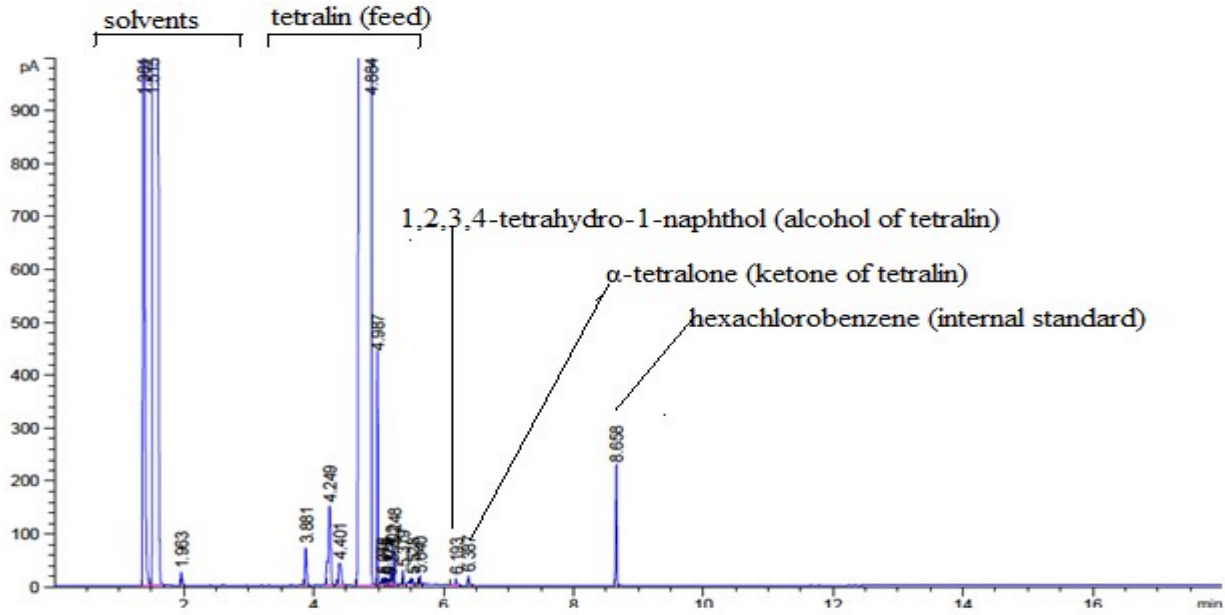
[a] with respect to air.
[b] density of oxygen was calculated at experimental pressure using ideal gas law.

6. Product identification

GC-FID chromatograms of tetralin oxidized at 150 ˚C in a microfluidic reactor are shown in Figure S3 to illustrate the ketone-to-alcohol selectivity in primary oxidation product.



(a)



(b)

Figure S3. GC-FID chromatogram of tetralin oxidized at 150 °C in a microfluidic reactor at gas-liquid interfacial area: (a) 3×10^5 m^2/m^3 (Series A: Table 2 in manuscript) and (b) 5×10^3 m^2/m^3 (Series E: Table 2 in manuscript).

7. Flame Ionization Detector (FID) Response Factors

The flame ionization detector (FID) has different responses to various organic compounds. So, it is required to calculate response factors for accurate quantification of oxidative conversion by GC-FID. The Dietz-method [13] was used to calculate the response factors:

$$Response\ factor\ (RF)_{Dietz} = \frac{(Area\ of\ compound) \times (Mass\ of\ standard)}{(Mass\ of\ compound) \times (Area\ of\ standard)}$$

(31)

Heptane was used as the standard and its response factor was 1.00. The calculated relative response factors are tabulated in Table S4. The calculated relative response factors are very close the response factors reported in literature. [13–15] The FID response factors previously reported in literature are also listed in Table S4 for comparison.

Table S4. FID response factors of various compounds.

Compound Name	Retention Time (min)	Response Factor (RF)	Reported RF value
Heptane	1.72	1.00 ± 0.00	1.00 [13]
$CHCl_3$	1.52	0.09 ± 0.01	
Hexachlorobenzene	8.67	0.32 ± 0.01	0.31 [14]
Tetralin	4.90	1.08 ± 0.01	1.02 [15]
1, 2, 3, 4 – tetrahydro-1-naphthol	6.35	0.82 ± 0.02	
Alpha-tetralone	6.51	0.84 ± 0.01	0.80 [15]

8. Conversion Calculations

Conversion was calculated based on the tetralin disappearance and did not reflect the extent of oxidation. The percentage conversion was calculated as follows: [1]

$$W_i = \frac{A_i * W_{HCB}}{A_{HCB} * RRF_{i,HCB}}$$

(32)

where,

$$RRF_{i,\,HCB} = \frac{RF_i}{RF_{HCB}}$$

(33)

This is the relative response factor of model compounds with respect to hexachlorobenzene (internal standard).

In equations 32 and 33,

RF_{HCB} = Response factor of hexachlorobenzene with respect to heptane

RF_i = Response factor of model compound with respect to heptane

W_i = Weight % of model compounds

W_{HCB} = Weight % of hexachlorobenzene

A_i = Peak area of model compounds

A_{HCB} = Peak area of hexachlorobenzene

For the conversion less than 1 (% wt/wt), the tetralin conversion was calculated based on the formation of products. A conversion factor was calculated using the data obtained from oxidation of tetralin with air conducted in a semi-batch reactor (Table S5). [15] Conversion factor was multiplied by sum of relative peak areas of product area to get the conversion. Conversion factor was selected based on the sum of product area.

Table S5. Conversion data for oxidation of tetralin with air at 150 ^{0}C conducted in a semi-batch reactor. [15]

Time (h)	Conversion (% wt/wt)	Sum of oxidized products	Conversion factor
0.5	0.8	214.8	0.0035
1	1.1	643.3	0.0017
2	2.1	1128.1	0.0019
4	4.5	2922.5	0.0015
6	6.9	4628.7	0.0015

9. Diffusion Coefficient Calculation

Different correlations are available in literature to calculate the diffusivities in liquid. The correlation provided by Díaz et al. [16] can be used to calculate diffusivity of gases in liquid over wide temperature range. This correlation is used to calculate the diffusion coefficient of oxygen in tetralin (D_A) at 150 °C. The correlation goes as follows:

$$(D_A)_T = 4.996 \times 10^3 (D_{AB})_{T=25^0C} e^{-\frac{2539}{T}} \quad (34)$$

where,

$$(D_A)_{T=25^0C} = 6.02 \times 10^{-5} \frac{V_B^{0.36}}{\mu_B^{0.61} V_A^{0.64}} \quad (35)$$

$(D_A)_T$ is the diffusion coefficient of oxygen in tetralin at given temperature in cm^2/s.

$(D_A)_{T=25\,°C}$ is the diffusion coefficient of oxygen in tetralin at 25 °C in cm^2/s.

T is the absolute temperature (K) = 423 K.

μ_B is the viscosity of tetralin = 2 cp.

V_A is the molar volume of oxygen at the normal boiling point temperature (cm^3/gmol) = 27.9 cm^3/gmol.
V_B is the molar volume of tetralin at the normal boiling point temperature (cm^3/gmol) = 135.7 cm^3/gmol.
The values obtained at the different relevant temperatures were:

$$(D_A)_{T = 25\,^{\circ}C} = 2.7 \text{ x } 10^{-09} \text{ m}^2/\text{s}$$

$$(D_A)_{T = 150\,^{\circ}C} = 3.4 \text{ x } 10^{-08} \text{ m}^2/\text{s}$$

References

1. Siddiquee MN, De Klerk A. Hydrocarbon addition reactions during low-temperature autoxidation of oilsands bitumen. Energy and Fuels. 2014;28(11):6848–59.
2. Siddiquee MN, de Klerk A, Nazemifard N. Application of microfluidics to control product selectivity during non-catalytic oxidation of naphthenic-aromatic hydrocarbons. React Chem Eng. Royal Society of Chemistry; 2016;1(4):418–35.
3. Yun J, Lei Q, Zhang S, Shen S, Yao K. Slug flow characteristics of gas-miscible liquids in a rectangular microchannel with cross and T-shaped junctions. Chem Eng Sci. Elsevier; 2010;65(18):5256–63.
4. Doraiswamy LK. Chemical Reaction Engineering : Beyond the Fundamentals. Boca Raton: Taylor & Francis; 2014.
5. Daoud JI. Multicollinearity and Regression Analysis. J Phys Conf Ser. 2017 Dec;949:012009.
6. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant. 2007;41(5):673–90.
7. Kutner M, Nachtsheim C, Neter J. Applied Linear Statistical Models. Irwin: McGraw-Hill; 2004.
8. Vatcheva KP, MinJae L, McCormick JB, Rahbar MH. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiol Open Access. 2016;06(02).
9. Freund RJ, Wilson WJ. Regression Analysis: Statistical Modeling of a Response Variable. Amsterdam: Academic Press; 2006.
10. Belsley DA, Kuh E, Welsch RE. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley & Sons, Inc.; 1980.
11. Midi H, Sarkar SK, Rana S. Collinearity diagnostics of binary logistic regression model. J Interdiscip Math. 2010;13(3):253–67.
12. Johnston J. Econometric Methods. 3rd ed. New York: McGraw-Hill; 1984.
13. Dietz WA. Response Factors for Gas Chromatographic Analyses. J Chromatogr Sci. 1967 Feb 1;5(2):68–71.

14. Yieru H, Qingyu O, Weile Y. Characteristics of Flame Ionization Detection for the Quantitative Analysis of Complex Organic Mixtures. Anal Chem. 1990;62(18):2063–4.
15. Katritzky AR, Ignatchenko ES, Barcock RA, Lobanov VS, Karelson M. Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Qualitative Structure-Property Relationships Treatment. Anal Chem. 1994;66(11):1799–807.
16. Diaz M, Vega A, Coca J. Correlation for the Estimation of Gas-Liquid Diffusivity. Chem Eng Commun. 1987;52(4–6):271–81.