

Supporting Information

Autonomous kinetic model identification using optimal experimental design and retrospective data analysis: methane complete oxidation as a case study

Arun Pankajakshan, Solomon Gajere Bawa, Asterios Gavriilidis,* and Federico Galvanin*
Department of Chemical Engineering, University College London, London, WC1E 7JE, United Kingdom.
a.gavriilidis@ucl.ac.uk, f.galvanin@ucl.ac.uk

1 Experimental Data

The full set of experimental conditions and corresponding data obtained in this work is provided in Table 1. An experiment ID has been used for each experiment to indicate the experimental campaign to which it belongs to. Each experiment is defined by the set of inputs (reaction temperature, mass flow rate, oxygen/methane mole ratio and inlet methane mole fraction) and the set of outputs (measured steady state mole fractions of methane, oxygen and carbon dioxide).

Table 1 Full set of experimental data showing experiment ID, inputs and outputs. Temperature is measured in the reactor, while all the other variables are at the reactor inlet.

Exp. No.	Exp. ID	Inputs				Outputs		
		Temperature (°C)	Mass flow rate (Nml/min)	Oxygen/Methane mole ratio (mol/mol)	Methane mole fraction (mol/mol)	Methane mole fraction (mol/mol)	Oxygen mole fraction (mol/mol)	Carbon dioxide mole fraction (mol/mol)
1	# Factorial DoE 1	253.9	20.0	2.0	0.005	0.00382	0.00601	0.00111
2	# Factorial DoE 2	355.5	20.0	4.0	0.005	0	0.00808	0.00495
3	# Factorial DoE 3	253.9	20.0	2.0	0.015	0.01286	0.02607	0.00175
4	# Factorial DoE 4	355.5	20.0	4.0	0.015	0	0.02832	0.01488
5	# Factorial DoE 5	253.9	20.0	4.0	0.025	0.02156	0.09596	0.00373
6	# Factorial DoE 6	355.5	20.0	2.0	0.025	0.00212	0.00082	0.02261
7	# Factorial DoE 7	253.9	30.0	4.0	0.005	0.00402	0.01790	0.00079
8	# Factorial DoE 8	355.5	30.0	2.0	0.005	0	0	0.00420
9	# Factorial DoE 9	253.9	30.0	4.0	0.015	0.01320	0.05398	0.00146
10	# Factorial DoE 10	355.5	30.0	2.0	0.015	0.00114	0	0.01374
11	# Factorial DoE 11	253.9	30.0	2.0	0.025	0.02323	0.04893	0.00179
12	# Factorial DoE 12	355.5	30.0	4.0	0.025	0.00085	0.04791	0.02401
13	# MBDoE-MD 1	313.8	22.2	2.3	0.024	0.01262	0.03071	0.01140
14	# MBDoE-MD 2	325.9	27.7	3.9	0.022	0.00908	0.05990	0.01236
15	# MBDoE-PP 1	327.1	20.2	4.0	0.017	0.00399	0.04028	0.01269
16	# MBDoE-PP 2	308.4	22.2	2.1	0.013	0.00641	0.01097	0.00682
17	# MBDoE-PP 3	281.2	20.2	3.8	0.025	0.01989	0.08633	0.00445
18	# MBDoE-PP 4	337.4	28.2	2.0	0.025	0.00953	0.01439	0.01508
19	# MBDoE-PP 5	281.3	20.1	3.9	0.008	0.00519	0.02187	0.00250
20	# MBDoE-PP 6	308.4	22.2	2.1	0.013	0.00706	0.01066	0.00584

2 Parameter Estimation Results for Model 2

The parameter estimation results including the estimated values of model parameters, 95 % confidence intervals and the t -values for Model 2 after the MBDoE-MD experimental campaign, i.e., at the end of experiment 14, are provided in Table 2. The same results after the MBDoE-PP experimental campaign, i.e., at the end of experiment 20, are shown in Table 3.

Results of parameter estimation at the end of MBDoE-PP campaign suggest that parameter 4 and 6 are not estimated precisely as indicated by their large confidence intervals and small t -values. In case of parameter 6, the estimated value is negligibly small (in the order of $1E-07$) compared to the magnitude of experimental error, which is in the order of $1E-04$. This suggests that the estimated value of parameter 6 is affected strongly by insignificant variations in data and its estimation cannot be explained based on random variations in data samples. This fact leads to the conclusion that it is better to keep the value of parameter 6 to zero and remove it from the estimation method.

Table 2 Parameter estimation results showing the estimated values, 95 % confidence interval (C.I) and t -values of parameters of Model 2 at the end of MBDoE-MD experimental campaign. Note that the value of t_{ref} is 1.68. Underlined values indicate a poor estimation of parameter, indicated by a confidence interval larger than the parameter estimate

Parameter	Estimate \pm 95% C.I	t -value
θ_1	8.66 ± 0.26	33.50
θ_2	8.21 ± 1.89	4.34
θ_3	2.48 ± 1.11	2.24
θ_4	4.45 ± 8.89	<u>0.50</u>
θ_5	4.69 ± 0.54	8.61
θ_6	$1.75E-06 \pm 0.20$	<u>8.5E-06</u>

Table 3 Parameter estimation results showing the estimated values, 95 % confidence interval (C.I) and t -values of parameters of Model 2 at the end of MBDoE-PP experimental campaign. Note that the value of t_{ref} is 1.67. Underlined values indicate a poor estimation of parameter, indicated by a confidence interval larger than the parameter estimate

Parameter	Estimate \pm 95% C.I	t -value
θ_1	8.64 ± 0.17	50.21
θ_2	8.73 ± 1.19	7.34
θ_3	2.48 ± 0.67	3.72
θ_4	2.19 ± 6.61	<u>0.33</u>
θ_5	3.91 ± 0.23	16.95
θ_6	$5.94E-07 \pm 0.06$	<u>1.02E-05</u>

3 Prediction Density Plots

The prediction density plots for models 2 and 3, which approximate the probability distributions of model predictions at each of the experimental conditions are shown in Figure 1. The method of evaluating the prediction density plots is explained in Section 2.6 of the main text. The prediction density plots are shown in multiple pages. Each row corresponds to the Exp. No. (see Table 1) and in each row from left to right, the prediction density plots for methane, oxygen and carbon dioxide are shown respectively. The observed experimental values within the error limits are also shown in the density plots.

The prediction density plots can be used to study the uncertainty of model predictions, which is essential to ensure the reliability of validated models. This is important because usually the models are validated at the best fitted parameter estimates and a good fit merely establishes the fact that there is no reason to reject the model based on the data at hand. However, this alone does not ensure that the model predictions are completely reliable with respect to the sampling distribution of the experimental data, which is better understood from the density plots. Prediction density plots along with the experimental data within their error limits provide insight about both the uncertainty as well as adequacy of the model in representing the data and their distribution. Another objective of studying the prediction density plots, particularly the joint prediction density plots of models 2 and 3 in Figure 1 is to understand the degree of discrimination between the two models at the performed experimental conditions.

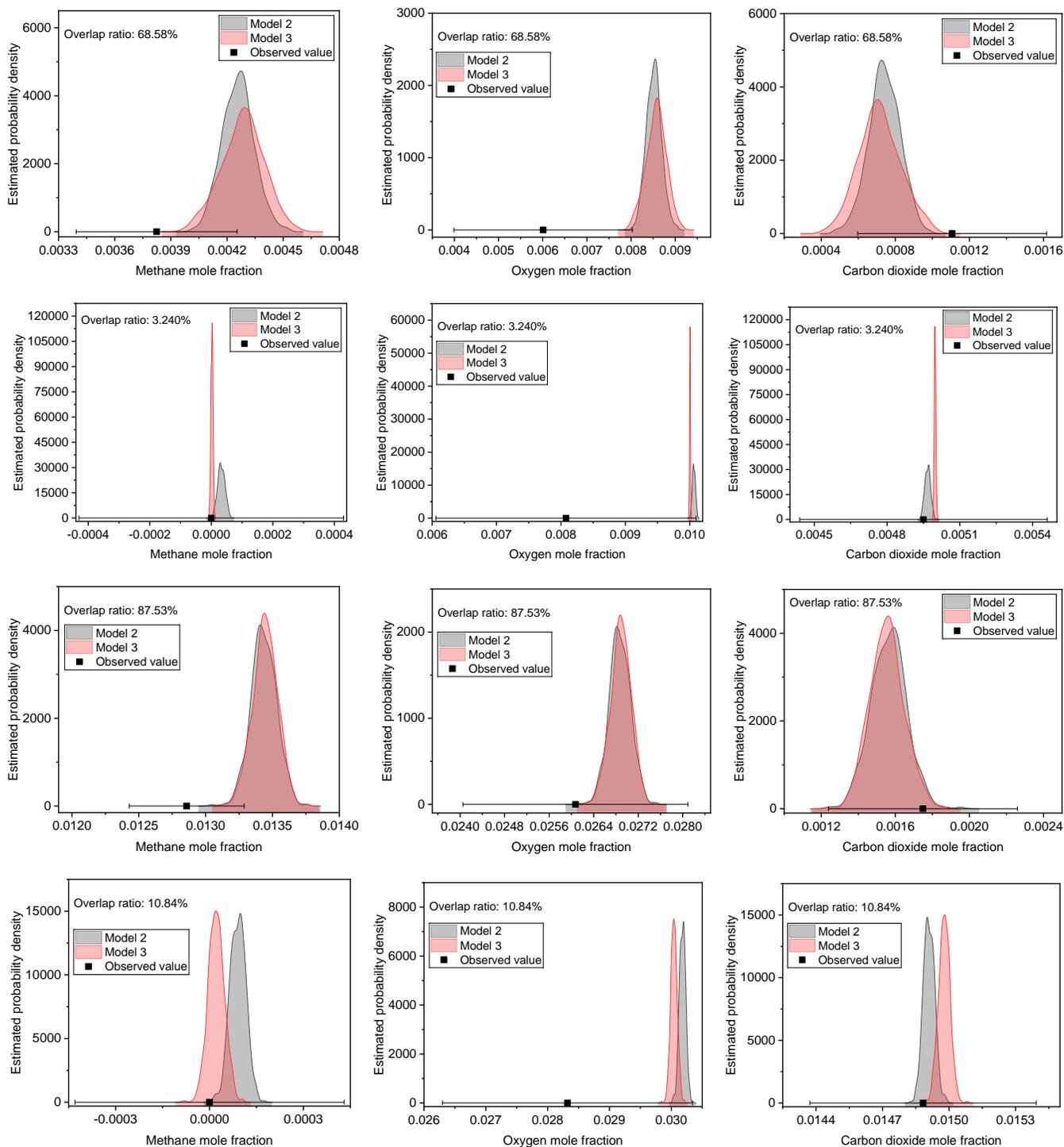


Fig. 1 Prediction density plots showing uncertainty in predictions of models 2 and 3 at the performed experiments (experiments 1 - 4 along the rows). The observed value is shown as a point with the error bar (\pm standard deviation).

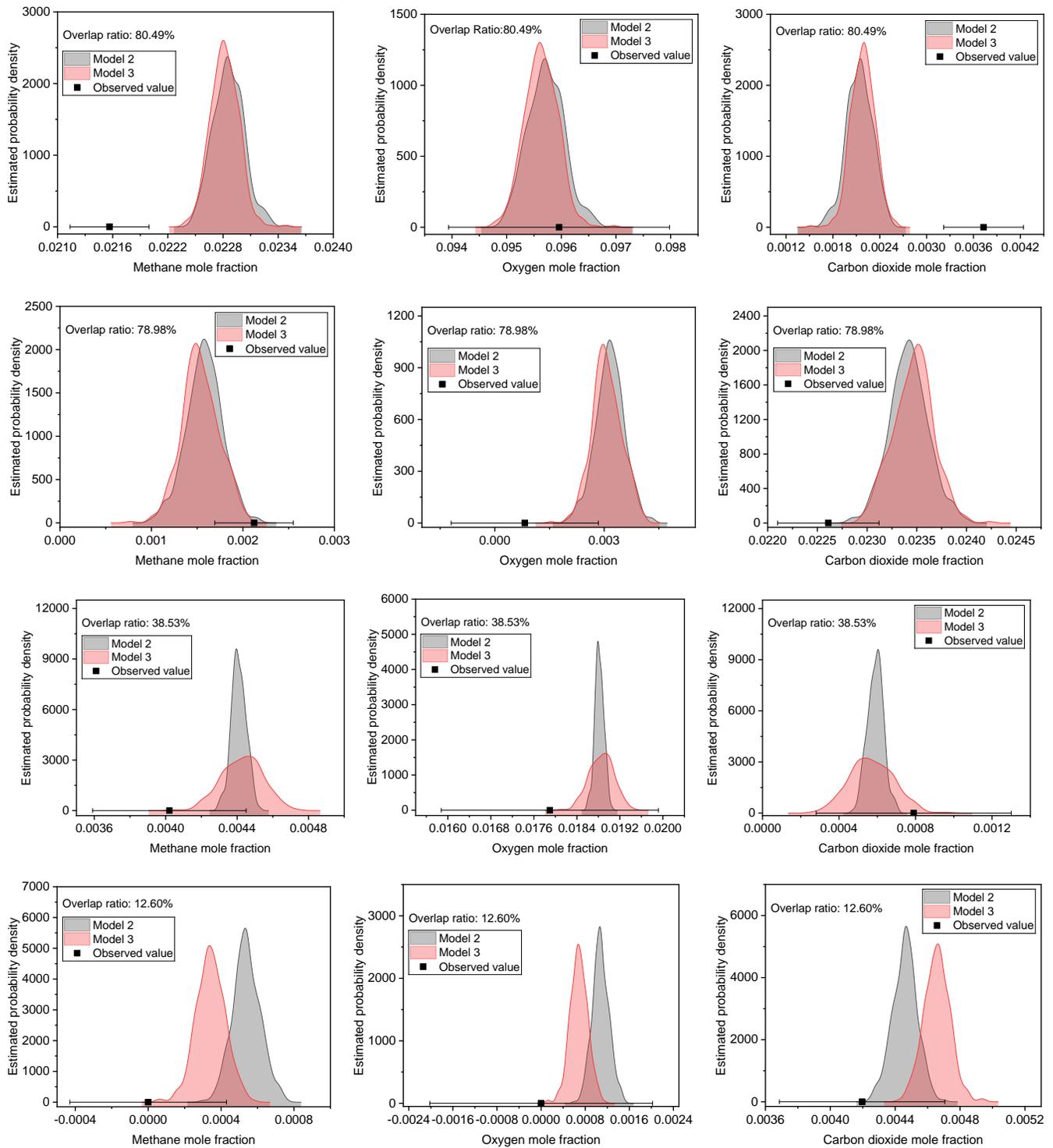


Fig. 1 Prediction density plots showing uncertainty in predictions of models 2 and 3 at the performed experiments contd. (experiments 5 - 8 along the rows). The observed value is shown as a point with the error bar (\pm standard deviation).

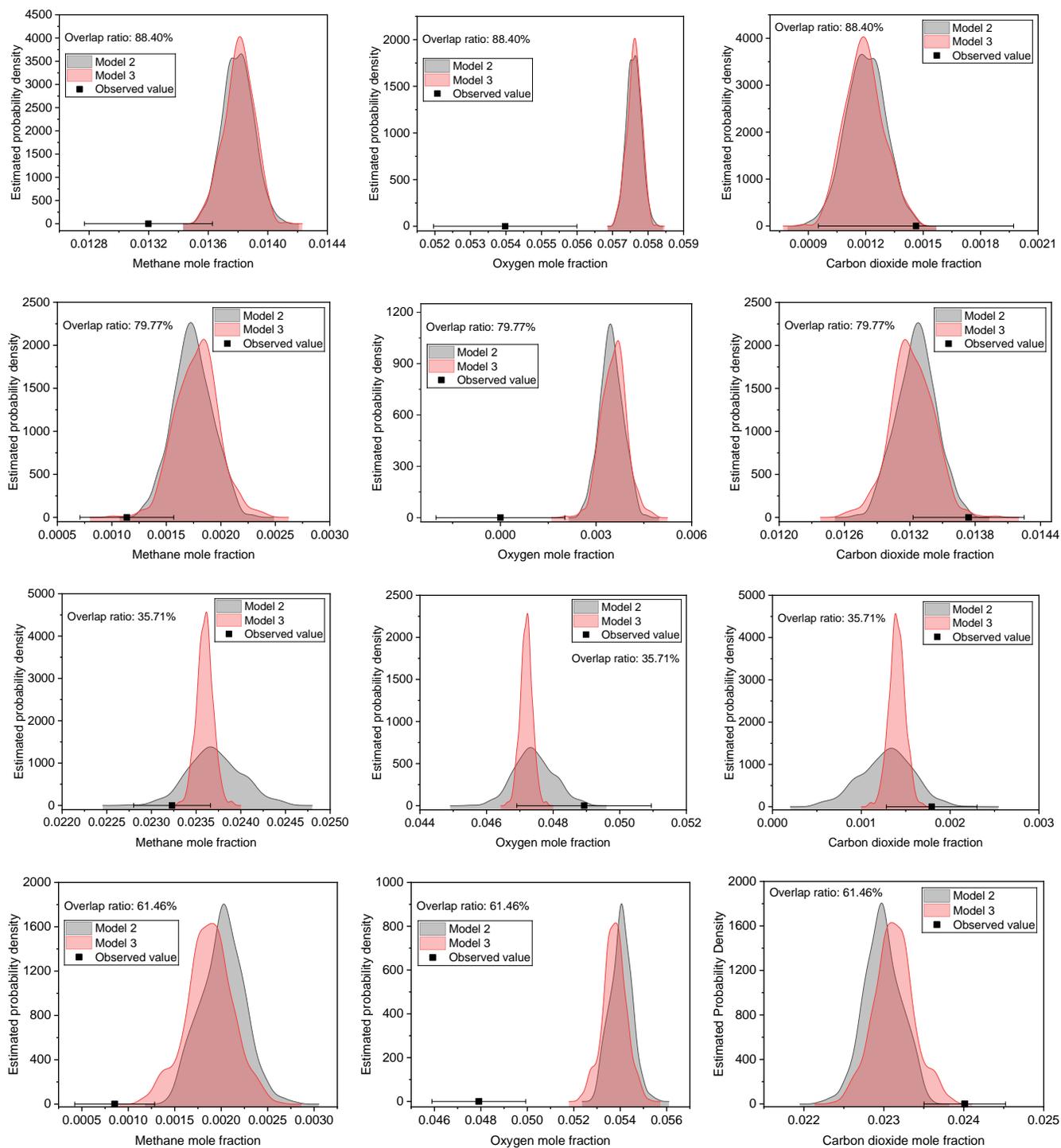


Fig. 1 Prediction density plots showing uncertainty in predictions of models 2 and 3 at the performed experiments contd. (experiments 9 - 12 along the rows). The observed value is shown as a point with the error bar (\pm standard deviation).

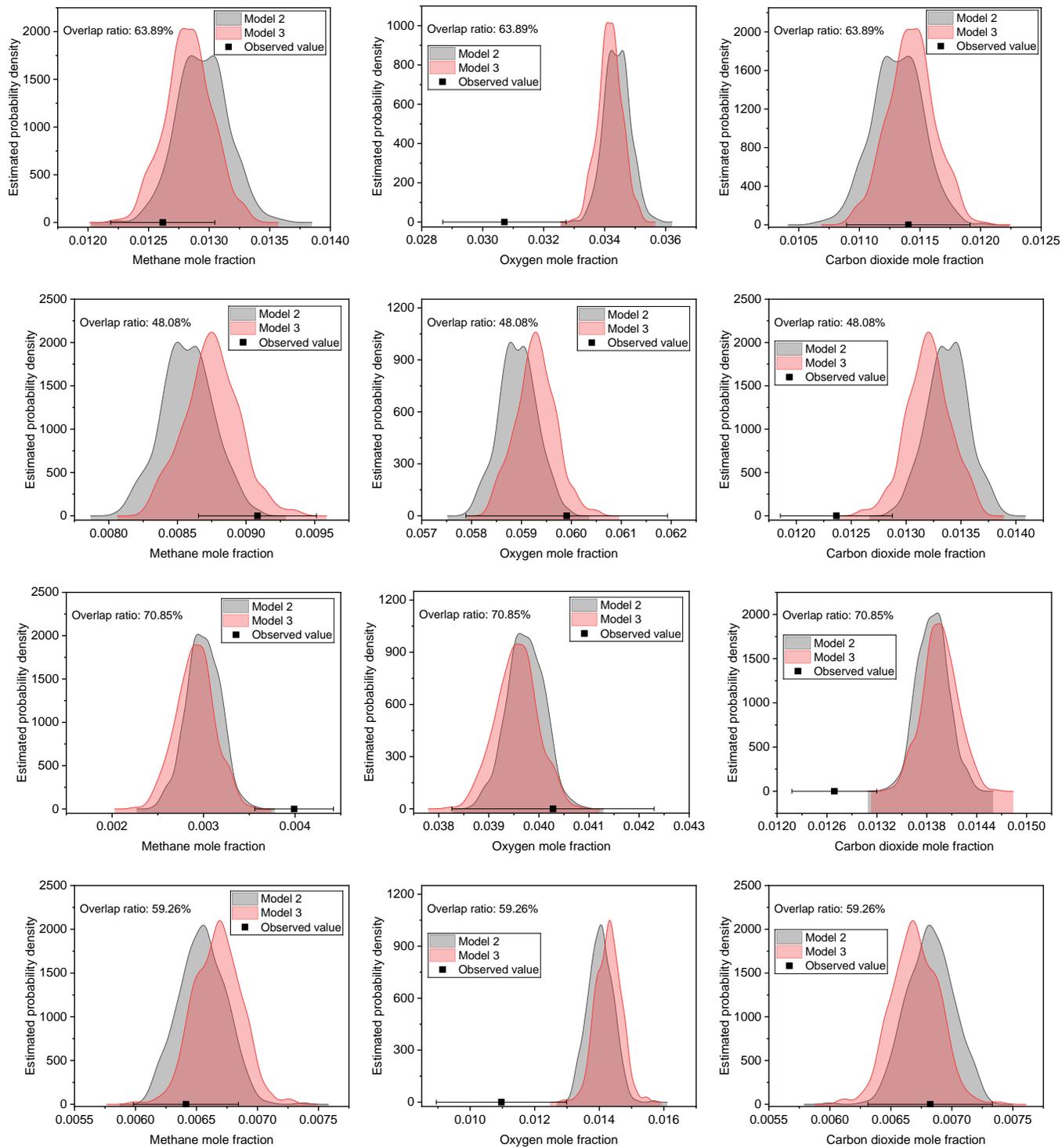


Fig. 1 Prediction density plots showing uncertainty in predictions of models 2 and 3 at the performed experiments contd. (experiments 13 - 16 along the rows). The observed value is shown as a point with the error bar (\pm standard deviation).

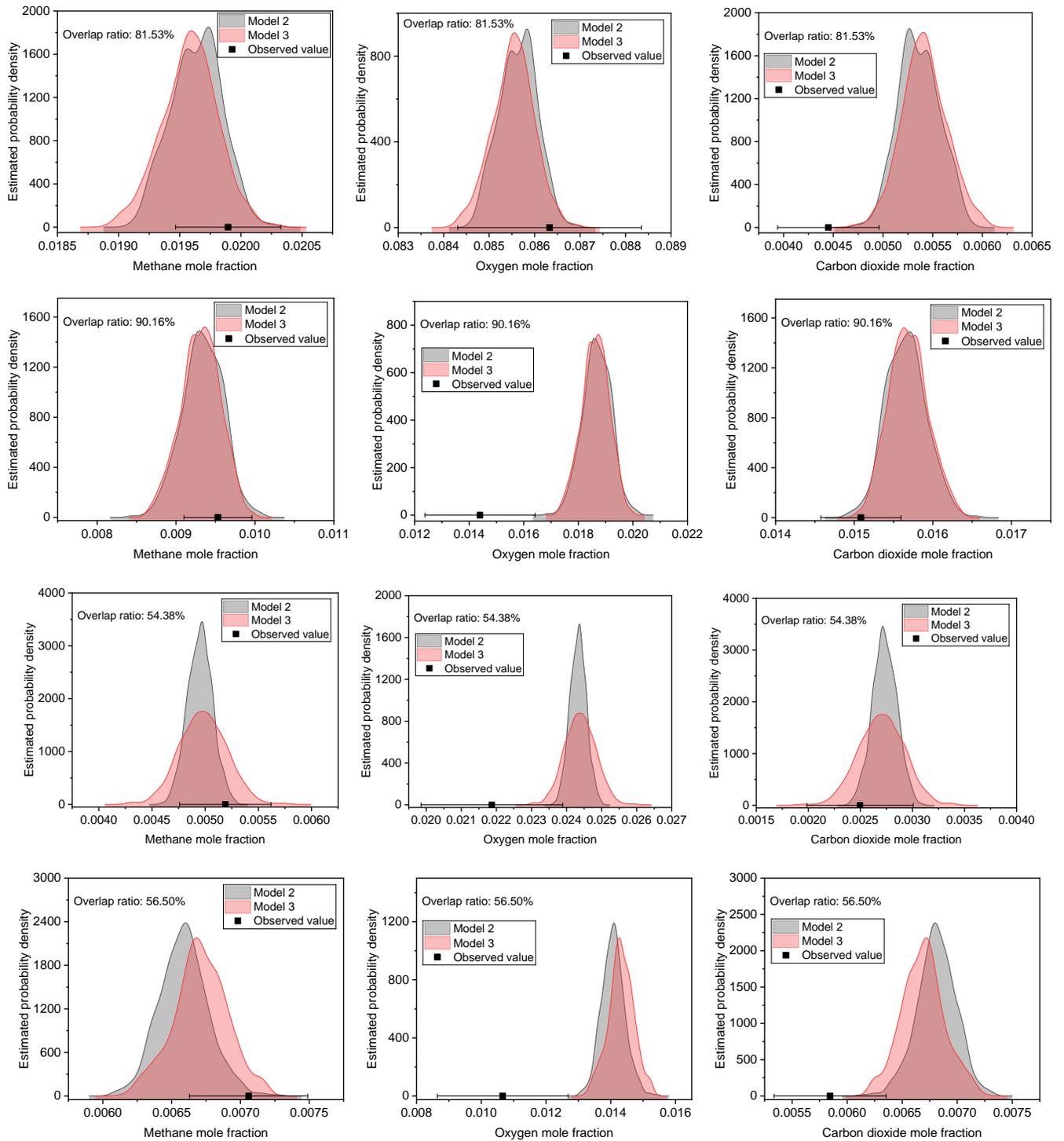


Fig. 1 Prediction density plots showing uncertainty in predictions of models 2 and 3 at the performed experiments contd. (experiments 17 - 20 along the rows). The observed value is shown as a point with the error bar (\pm standard deviation).

4 Simulation over an extended design space

The ranges of operating conditions used in this work were selected based on the feasible experimental limits achievable with the setup. However, to check if widening the design space can speed up the convergence of the procedure to identify an appropriate kinetic model, a simulation study was carried out over an extended design space. The extended ranges of inputs used in the simulation study are provided in Table 4 along with the original bounds. The results of simulation study are shown in Figure 2 and Figure 3. In Figure 2, the contours of MBDoE-MD objective function for discrimination between models 2 and 3, which is to be maximised are plotted and in Figure 3, the contours of MBDoE-PP objective function for improving precision of parameters of model 3, which is to be minimised are plotted. In both figures, the original design space is shown in a box.

The simulation study was conducted using the parameter estimates obtained in this work. The results suggest that at high inlet methane mole fraction of $0.05 \text{ mol mol}^{-1}$ (panels (b) and (d) of Figures 2 and 3), lowering the mass flow rate (less than 20 Nml/min) can result in better discrimination between the models and can also generate informative data to precisely estimate the parameters of model 3. Hence, extending the ranges of operating conditions could have helped to speed up the procedure to identify an appropriate model. However, the use of extended ranges of operating conditions was experimentally unfeasible.

5 Parity plot of methane conversion

The error bounds of methane conversions in panel (d) of Figure 11 in the main paper were computed as follows. Considering methane conversion X_{CH_4} as a function of initial $y_{\text{CH}_4}^{\text{in}}$ and final $y_{\text{CH}_4}^{\text{out}}$ mole fractions of methane, i.e., $X_{\text{CH}_4} = f(y_{\text{CH}_4}^{\text{in}}, y_{\text{CH}_4}^{\text{out}})$, and according to the principle of propagation of error, the error in methane conversion δX_{CH_4} can be computed as given in Equation 1.

$$\delta X_{\text{CH}_4} = \sqrt{\left(\frac{\partial f}{\partial y_{\text{CH}_4}^{\text{in}}} \delta y_{\text{CH}_4}^{\text{in}}\right)^2 + \left(\frac{\partial f}{\partial y_{\text{CH}_4}^{\text{out}}} \delta y_{\text{CH}_4}^{\text{out}}\right)^2} \quad (1)$$

In Equation 1, $\delta y_{\text{CH}_4}^{\text{in}}$ and $\delta y_{\text{CH}_4}^{\text{out}}$ are the random errors in measuring inlet and outlet methane mole fractions. According to the constant variance assumption used in this work, these errors are assumed to be the same and computed as $0.00043 \text{ mol mol}^{-1}$. Although the errors in measurement of methane mole fractions are assumed constant, Equation 1 leads to an expression in which the error in methane conversion δX_{CH_4} becomes a function of the measured mole fractions $y_{\text{CH}_4}^{\text{in}}$ and $y_{\text{CH}_4}^{\text{out}}$ and hence they are different for different values of $y_{\text{CH}_4}^{\text{in}}$ and $y_{\text{CH}_4}^{\text{out}}$. The error bounds shown in Figure 11 (panel (d)) represent \pm twice the standard deviation δX_{CH_4} of methane conversions.

Table 4 Range of control variables. Temperature is measured in the reactor, while all the other variables are at the reactor inlet.

Control variable	Temperature [°C]	Mass flow rate [Nm ³ min ⁻¹]	Oxygen to methane mole ratio [mol.mol ⁻¹]	Methane mole fraction [mol.mol ⁻¹]
Original range	250-350	20-30	2-4	0.005-0.025
Extended range	100-400	10-40	2-6	0.005-0.05

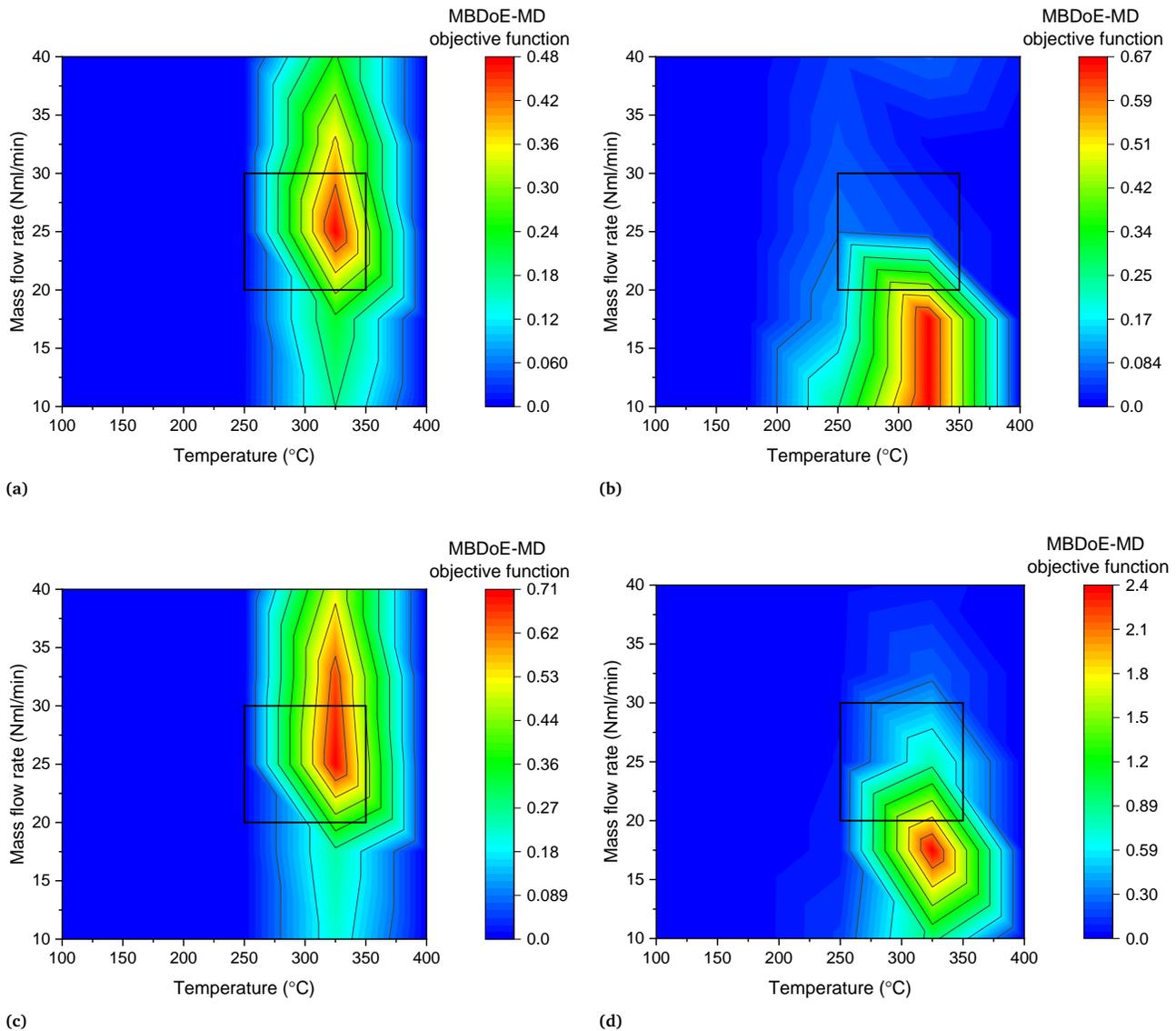


Fig. 2 Contour plots showing maps of objective function for model discrimination (discrimination between model 2 and model 3) as a function of reactor temperature and inlet mass flow rate, at inlet oxygen to methane mole ratio and inlet methane mole fraction of: (a) 5 mol mol⁻¹ and 0.0275 mol mol⁻¹, (b) 5 mol mol⁻¹ and 0.05 mol mol⁻¹, (c) 6 mol mol⁻¹ and 0.0275 mol mol⁻¹, (d) 6 mol mol⁻¹ and 0.05 mol mol⁻¹. The original design space is shown in the box.

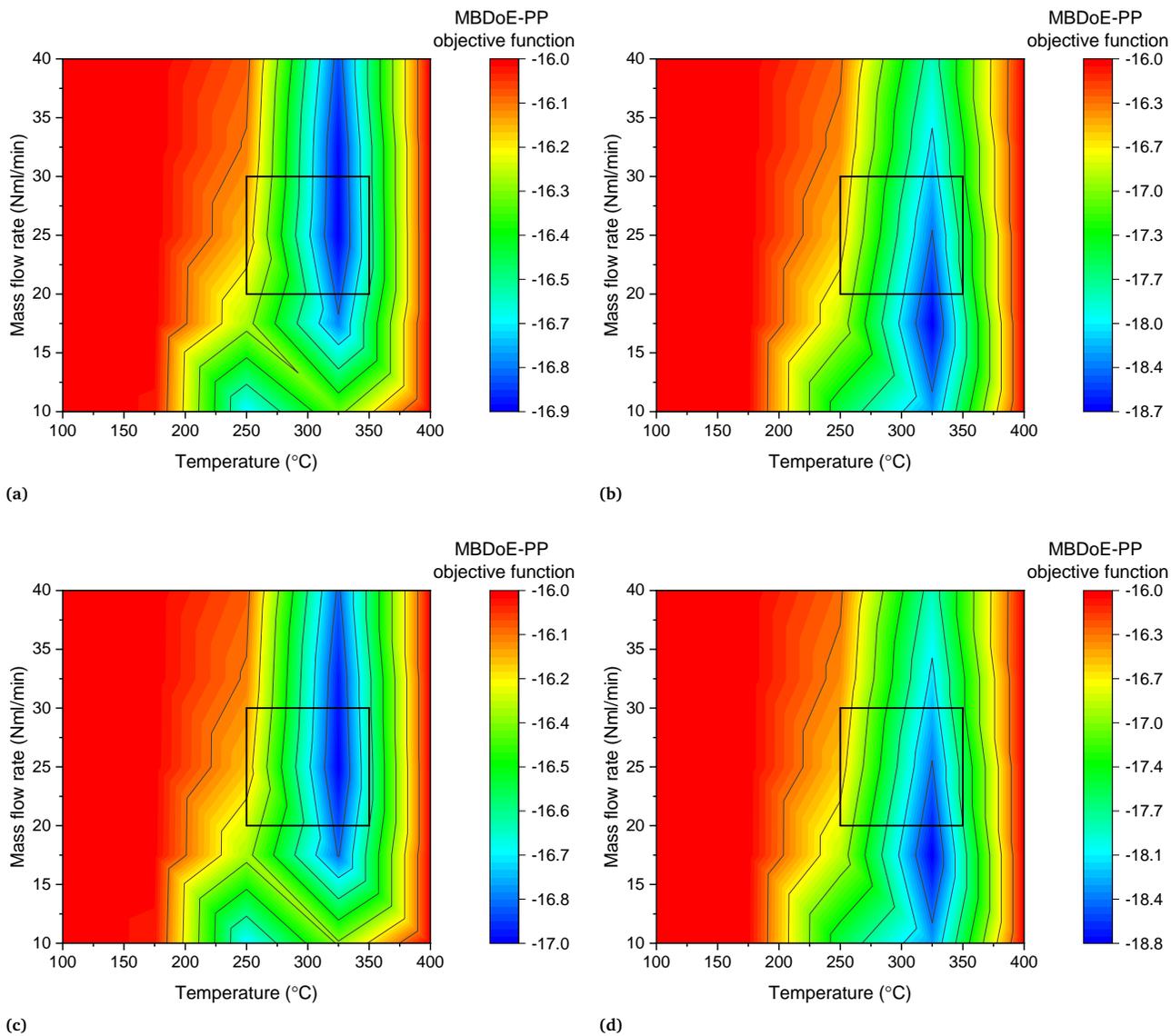


Fig. 3 Contour plots showing maps of objective function for improving parameter precision of model 3 as a function of reactor temperature and inlet mass flow rate, at inlet oxygen to methane mole ratio and inlet methane mole fraction of: (a) 5 mol mol^{-1} and $0.0275 \text{ mol mol}^{-1}$, (b) 5 mol mol^{-1} and $0.05 \text{ mol mol}^{-1}$, (c) 6 mol mol^{-1} and $0.0275 \text{ mol mol}^{-1}$, (d) 6 mol mol^{-1} and $0.05 \text{ mol mol}^{-1}$. The original design space is shown in the box.