Robust Measurement Selection Design for Experimental Systems with Input Uncertainty

Ke Wang and Hong Yue

Wind Energy & Control Centre, Department of Electronic and Electrical Engineering
University of Strathclyde
Glasgow G1 1XW, UK
k.wang@strath.ac.uk; hong.yue@strath.ac.uk

Abstract—**Input uncertainty in experimental implementation deteriorates data quality for parameter estimation. This work aims to examine the influence of input uncertainty, in particular the inaccurate setting of initial states, to parameter estimation and explore methods to mitigate the effects. First, a Monte-Carlo method is employed to generate input-output data. The input uncertainty is assumed to follow Gaussian distribution. Samples are taken from the uncertainty region and used to produce output through the dynamic system. Statistical characteristics are utilised to quantify uncertainty in outputs. Then a robust experimental design (RED) is proposed, in which the states that are less affected by input uncertainty are selected as measurement state variables. In addition, two different residual functions are used in parameter estimation to compare the estimation robustness against data uncertainty. Simulation studies are conducted using a benchmark enzyme reaction system. Compared to the non-designed experimental settings, improved parameter estimation is achieved via robust design.**

Index Terms—**robust experimental design (RED), input uncertainty, optimal experimental design (OED), parameter estimation, residual function**

I. INTRODUCTION

Building a reliable model for complex systems requires data that adequately describe system dynamics. Optimal experimental design (OED) can be used to find experimental settings that will produce the most informative data for modelling purposes such as model discrimination and parameter estimation. In a typical OED for parameter estimation, the Fisher information matrix (FIM) is used to evaluate the information content of data based on the Cramer-Rao inequality. The optimisation design problem is formulated based on selected metrics of FIM.

OED is a design methodology based on local (parameter) sensitivities, therefore applies mostly to systems with small model uncertainties. Robust experimental design (RED) is required for systems with large uncertainties. Typical RED methods include the pseudo-Bayesian method [1] and the maximin method [2], which can be implemented in one experimental design cycle. Other design strategies such as iterative design [3], [4] or online experimental redesign [5], [6] can also be used but would cost more time and resources.

Apart from model uncertainties, another uncertainty comes from the malfunction in experimental implementation, especially on the input settings. This issue is addressed in some recent studies, aiming to find the input for experiments so that the produced outputs can be constrained by certain thresholds. One example is from manufacturing industry where it is difficult to accurately control the input conditions during machine operation. The input uncertainty range relevant to machine operation performance is identified through a level set estimation method [7]. In a batch production process in chemical engineering, the impact of input uncertainty is assessed by formulating the input variations with polynomial chaos expansions, and the output of every batch is controlled within a given range to avoid possible substrate inhibition effect [8]. These works reveal input uncertainties in different applications. The influence of input uncertainty on OED and parameter estimation are yet to be investigated.

In this work, we aim to understand the impact of input uncertainty on experimental design and explore how to reduce this influence to parameter estimation through RED. When input uncertainty exists in an experiment, the measurement data will deviate from the expected values under OED. This will inevitably affect parameter estimation since the inverse problem is likely to be ill-conditioned [9]. In most cases, explicit forms of input uncertainty projected to outputs and further to parameter estimations are hardly obtainable. The Monte-Carlo method can be used instead. Samples collected from the input uncertainty range can be used to generate the pseudo experimental data. Then, each set of pseudo data are used to solve the inverse problem to obtain a set of parameter estimates, from which the statistical quality of parameter estimation can be determined. To reduce the influence of input uncertainty on parameter estimation, we propose a measurement set selection design which selects measurable state variables that are less affected by the input uncertainty. To further reduce the influence of input uncertainty to parameter estimation, we propose to replace the typical L-2 norm in the inverse problem with L-1 norm since the latter is less sensitive to outliers [9].

The remaining paper includes the following. Section II gives preliminaries on experimental design. The main methodology is presented in Section III including OED of input design,

mapping of input uncertainty to output and RED of measurement set selection. Numerical studies are undertaken on a benchmark enzyme reaction system and results of different methods are compared in Section IV. Conclusions are given in Section V.

II. PRELIMINARIES

A. Optimal Experimental Design (OED)

A dynamic nonlinear system can be represented by the following state-space model:

$$\dot{\mathbf{X}}(t) = \boldsymbol{f}(\mathbf{X}(t), \boldsymbol{\theta}), \mathbf{X}(t_0) = \mathbf{X}_0$$
$$\mathbf{Y}(t) = \mathbf{h}(\mathbf{X}(t)) + \boldsymbol{\xi}(t) \tag{1}$$

where $\boldsymbol{f}(\cdot)$ are a series of dynamics functions which are continuous and first-order derivable; $\mathbf{X} = [x_1, x_2, \ldots, x_{N_X}]^T \in \mathbb{R}^{N_X}$ is the vector of state variables. $\mathbf{X}_0$ is the initial state vector, among which the N_I nonzero entries are grouped in $\bar{\mathbf{X}}_0 \in \mathbb{R}^{N_I}$. $\boldsymbol{\theta} \in \mathbb{R}^{N_P}$ is the parameter vector. The output $\mathbf{Y} = [y_1, y_2, \ldots, y_{N_Y}]^T \in \mathbb{R}^{N_Y}$ is the vector of measurable state variables, which is chosen from the N_X state variables; $\mathbf{h}(\cdot)$ is a selection function for measurement; $\boldsymbol{\xi}$ is the vector of measurement errors, which is assumed to be independent at different time t and have Gaussian distribution.

For a system in (1), the local parametric sensitivity matrix is defined by $\boldsymbol{S} = \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}}$, in which only the state variables selected by $\mathbf{h}(\cdot)$ are included in the matrix so that $\boldsymbol{S} \in \mathbb{R}^{N_Y \times N_P}$. The time derivative of $\boldsymbol{S}$ is

$$\dot{\boldsymbol{S}} = \frac{d}{dt}\frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{f}}{\partial \mathbf{X}} \cdot \frac{\partial \mathbf{X}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}} \tag{2}$$

Sensitivity analysis quantifies the effect of parameter variations on state variables, which is useful for OED in at least two ways. One is to select the most sensitive parameters for parameter estimation, keeping those less influential ones at their nominal values, the dimension of OED is thus reduced. Another use of the local sensitivity matrix is to form FIM, based on which the OED objective function is established.

In model-based OED, the FIM is used as an indicator of information content of the data because the inverse of FIM provides a lower bound for parameter estimation error covariance, following the Cramer-Rao inequality. The FIM for (1) can be written as

$$\mathbf{FIM}(\boldsymbol{\psi}, \boldsymbol{\theta}) = \boldsymbol{S}^T(\boldsymbol{\psi}, \boldsymbol{\theta}) \cdot \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} \cdot \boldsymbol{S}(\boldsymbol{\psi}, \boldsymbol{\theta}) \tag{3}$$

where $\boldsymbol{\psi}$ is the experimental design vector, which can be dynamic stimuli, initial state, sampling time, measurement selection weightings and so on; $\boldsymbol{\Sigma}_{\mathbf{Y}}$ is the measurement error covariance matrix which is a diagonal matrix $\boldsymbol{\Sigma}_{\mathbf{Y}} = diag[\sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_Y}^2] \in \mathbb{R}^{N_Y \times N_Y}$. The OED is formulated as an optimisation problem, i.e.

$$\boldsymbol{\psi}^* = \arg\min_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \phi\left((\mathbf{FIM}(\boldsymbol{\psi}, \boldsymbol{\theta}))^{-1}\right) \tag{4}$$

where $\phi(\cdot)$ is a scaler function to extract selected characteristics of FIM. The most commonly used scalarisation criteria are the determinant (D-optimal), L2 norm (E-optimal), and trace (A-optimal).

B. Parameter Estimation

Parameter estimation can be obtained by minimising the residual between the measured output $\mathbf{Y}$ and the estimated output $\hat{\mathbf{Y}}(\boldsymbol{\theta})$.

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\mathbf{Y} - \hat{\mathbf{Y}}(\boldsymbol{\theta})\| \tag{5}$$

The L-2 norm is often used in (5). When weighted by the inverse of the measurement error covariance matrix, it makes the well-known least square estimation (LSE). LSE gives the maximum likelihood estimation under independent Gaussian measurement noise. However, the Gaussian assumption does not always apply, especially when there exists input uncertainty. An alternative criterion, the L-1 norm, can be used to make the estimation more tolerant to outliers [9].

C. Monte-Carlo Method and Latin Hypercube Sampling

The Monte-Carlo method is widely used in sampling, optimisation and uncertainty propagation. The main idea is to repeat the experiment (or simulation) many times to obtain sufficient data of interested objectives [10]. According to the law of large numbers, the statistical characteristics can be approximated by those data. The Monte-Carlo method is often used for problems where analytic forms are not available.

By repeating the experiment or model simulation of (1) N_{sp} times, a set of data pairs $[\tilde{\mathbf{X}}^{(k)}, \hat{\mathbf{Y}}^{(k)}]$ are obtained, where $k = 1, 2, \ldots, N_{sp}$. The probability, mean and variance of the ith output y_i can be obtained as follows [11]:

$$prob(y_i \leq Y) \simeq W \sum_{k=1}^{N_{sp}} \delta_Y(\tilde{y}_i^{(k)}) \tag{6}$$

$$E(y_i) \simeq W \sum_{k=1}^{N_{sp}} \tilde{y}_i^{(k)} \tag{7}$$

$$V(y_i) \simeq W \sum_{k=1}^{N_{sp}} [\tilde{y}_i^{(k)} - E(y_i)]^2 \tag{8}$$

where $\delta_Y(\tilde{y}_i^{(k)})$ equals to 1 if $\tilde{y}_i^{(k)} \leq Y$ and 0 otherwise; W is usually $1/N_{sp}$ or $1/(N_{sp} - 1)$ according to the sampling method.

Various sampling strategies have been proposed for the Monte-Carlo method, among them the Latin hypercube sampling (LHS) is most commonly used since it requires less sampling size and is easy to implement [11]. In LHS, the range of each variable is divided into N_{sp} intervals of equal probability, and one sample is chosen randomly from every interval. Therefore, every element in $\mathbf{X}$ will have N_{sp} samplings. Those samplings are grouped in a random manner to form the set of initial states $[\tilde{\mathbf{X}}^{(1)}, \ldots, \tilde{\mathbf{X}}^{(N_{sp})}]$.

III. METHODOLOGY

Input uncertainty occurs in experiments when the setting of initial state deviates from the designed values. This uncertainty propagates to outputs through nonlinear system dynamics, will affect the data quality for parameter estimation. In this section, the OED for input design is firstly presented. Then

the input uncertainty is characterised by stochastic distribution and projected to output by LHS. Finally a robust measurement set selection design is proposed to reduce impact of input uncertainty to parameter estimation.

A. Optimal Input Design

The task of input design is to determine the values for the nonzero initial states $\bar{\mathbf{X}}_0$ so as to generate the most informative input-output data for parameter estimation. The OED on initial states can be formulated as

$$\bar{\mathbf{X}}_0^* = \arg\min \phi\left(\left(\mathbf{FIM}\left(\bar{\mathbf{X}}_0, \boldsymbol{\theta}\right)\right)^{-1}\right)$$
$$s.t. \quad \dot{\mathbf{X}}\left(t\right) = \boldsymbol{f}\left(\mathbf{X}\left(t\right), \boldsymbol{\theta}\right) \qquad (9)$$
$$\bar{\mathbf{X}}_L \leq \bar{\mathbf{X}}\left(t_0\right) \leq \bar{\mathbf{X}}_U$$

where $\bar{\mathbf{X}}_L$ and $\bar{\mathbf{X}}_U$ are the lower and upper bounds for the nonzero initial states.

The optimisation problem in (9) is non-convex in general for nonlinear systems. To obtain the global solution, multiple starting points of $\bar{\mathbf{X}}_0$ will be implemented. The initial states in $\bar{\mathbf{X}}_0$ are assumed to be independent and uniformly distributed within $[\bar{\mathbf{X}}_L, \bar{\mathbf{X}}_U]$. The LHS method will be used as the sampling strategy to construct the initial data for $\bar{\mathbf{X}}_0$.

B. Mapping of Uncertainty from Input to Output

A Gaussian distribution is assumed for the input uncertainty, that is, $\bar{\mathbf{X}}_0 \sim N(\boldsymbol{\mu}, \bar{\boldsymbol{\Sigma}})$, where $\boldsymbol{\mu}$ is the mean that can be taken as the designed initial states from OED, i.e., $\bar{\mathbf{X}}_0^*$. $\bar{\boldsymbol{\Sigma}}$ is an $N_I \times N_I$ diagonal variance matrix.

The Monte-Carlo method is used to map the uncertainty from the inputs to the outputs. N_{sp} samples $[\bar{\mathbf{X}}_0^{(1)}, \bar{\mathbf{X}}_0^{(2)}, \ldots, \bar{\mathbf{X}}_0^{(N_{sp})}]$ are drawn from the Gaussian distribution using LHS. By solving the dynamic system model in (1), N_{sp} pseudo-data $\tilde{\mathbf{Y}}$ can be produced. Due to the nonlinear dynamics, the uncertainty in outputs won't follow Gaussian distribution. At time t_j, each output variable y_i is a random variable $y_i(t_j)$ with statistics calculated by (6) for cumulative probability, (7) for mean and (8) for variance.

C. Robust Measurement Selection Design

To select the measurement variables that are less affected by the input uncertainty, a criterion is prosposed to keep the squared deviation between the sampled pseudo data $\tilde{\mathbf{Y}}$ and $\mathbf{Y}\left(\bar{\mathbf{X}}_0^*\right)$ as small as possible. Assume N_{SY} measurable variables are to be selected from the N_Y outputs. A weighting factor w_i is assigned to each output, $w_i = 1$ means y_i is selected and 0 otherwise. Denote $\boldsymbol{w} = [w_1, \cdots, w_{N_Y}]^T$, the OED can be formed as an linear integer programming problem.

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}^T \mathbf{D}$$
$$s.t. \quad \mathbf{1}^T \boldsymbol{w} = N_{SY} \qquad (10)$$

where $\mathbf{D} = [D_1, D_2, \ldots, D_{N_Y}]^T \in \mathbb{R}^{N_Y}$ is a vector of squared deviation between $\tilde{\mathbf{Y}}$ and $\mathbf{Y}\left(\bar{\mathbf{X}}_0^*\right)$ in which $D_i = \sum_{j=1}^{N}\sum_{k=1}^{N_{sp}}\left(\tilde{y}_i^{(k)}(t_j) - y_i(\bar{\mathbf{X}}_0^*, t_j)\right)^2$, N is the total number of sampling points in time.

Since this optimisation problem is linear in both objective function and constraint, it can be solved by ranking the outputs according to deviation D_i. Here the minimum number for N_{SY} is set to be one. It should be noted that this robust design is focused on reducing the impact to outputs from input uncertainty, it is not the typical RED on measurement set selection based on FIM. The latter is meant to include more information in measurement variables for parameter estimation.

IV. CASE STUDY AND RESULTS

In this section, a benchmark model — enzyme reaction model, is used for simulation study. This model represents typical kinetically controlled synthesis. The system has 10 state variables, only five of them can be measured. The model is represented by 10 ODEs established following the mass action principles (see appendix). The model details can be found in [12].

The time duration of model response is set to be 6000s to allow output responses reach steady-state under all simulation conditions. For each measurable state, 21 samples are take uniformly in the time horizon, the sampling time points are $[0 : 300 : 6000]s$. Three most sensitive parameters, $[k_2, k_{-3}, k_5W]$ (with nominal value of $[100, 200, 5000]$) [12], need to be precisely estimated and included in the experimental design.

A. Optimal Input Design

The nonzero initial state vector is the experimental factor to be designed to maximise the information content in experimental data. In this enzyme reaction system, there are three nonzero initial states: the donor substrate S, the nucleophile N and the enzyme E. The upper and lower bounds for the three states and their assumed initial values are listed in TABLE I.

TABLE I
DESIGN SPACE AND INITIAL ASSUMPTION OF $\bar{\mathbf{X}}_0$

States (mol/L)	Lower	Upper	Inital Assumption
S_0	0	1	0.8
N_0	0	1	0.9
E_0	1.5e-6	1.5e-4	1.5e-5

The input design is formed as a constrained optimisation problem in (9). Here the E-optimal criterion is chosen as the scaler function ψ in OED. The local solution can be obtained using the active-set sequential programming available in the MATLAB optimisation toolbox. Multiple starting points are implemented to get a set of local optimum solutions, from which the global one is taken. The design results for the initial states are listed in TABLE II.

TABLE II
DESIGN RESULTS OF INITIAL STATES $\bar{\mathbf{X}}_0^*$

	$[E_0^*, S_0^*, N_0^*]$ (mol/L)
Non-design	[1.5e-05, 0.8, 0.9]
Local E-optimal	[3.18e-06, 0.57, 0.49]
Global E-optimal	[4.27e-06, 0.94, 0.07]

Using the three sets of initial states in TABLE II, the information contents for parameter estimation can be evaluated by the estimation error covariance matrix, which is approximately the inverse of FIM. The confidence interval (CI) ellipsoids of parameter pairs (k_2, k_{-3}) and $(k_2, k_5 W)$ are plotted in Fig. 1. It can be observed from Fig. 1 that, through OED of the initial states, the parameter estimation accuracy is improved. OED provides smaller CI for parameter estimation uncertainty compared to the non-designed settings. The global OED shows better performance than the local OED. The CIs for each parameter and the percentage uncertainty (ratio of CI range over the nominal value) under different design are compared in TABLE III.
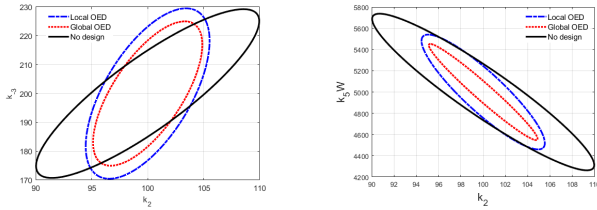

Fig. 1. Confidence interval of parameters with different initial state designs

TABLE III
CIs and Percentage Uncertainty under different input design

Parameter	Confidence Interval and Uncertainty		
	$k_2(100)$	$k_{-3}(200)$	$k_5 W(5000)$
No-design	[92.02 107.98]	[176.57 223.43]	[4409.6 5590.4]
	15.96%	23.43%	23.62%
Local OED	[95.60 104.40]	[177.49 222.51]	[4579.5 5420.5]
	8.8%	22.51%	16.82%
Global OED	[96.09 103.91]	[179.96 220.04]	[4637.6 5362.4]
	7.82%	20.04%	14.4%

B. Assessment of Input Uncertainty Impact

Once $\bar{\mathbf{X}}_0^*$ is designed, it is expected that the initial states can be accurately set in experiments following the designed values. This is hard to achieve in real operations, the deviation is taken as input uncertainty. As presented in Section III-B, this input uncertainty is assumed to be independent zero-mean Gaussian distribution. The standard deviation for the nonzero initial states is set to be $\bar{\Sigma} = diag[1e - 7, 0.01, 0.01]$, which ensures around 95% of CI within $[-1.96\sigma, 1.96\sigma]$ for input uncertainty.

The Monte-Carlo method is used to get a numerical description of the impact of uncertainty on outputs and further on parameter estimation. Using the LHS method, 1000 samples are drawn from the input uncertainty space, and the outputs can be obtained by solving the system dynamic model using those initial state samples.

The time responses of states R and N are shown in Fig. 2. Box charts are used for the outputs with mapped uncertainty, which shows the median (the line in the blue box), the lower and upper quartiles (the limits of the blue box), outliers (circles), and the minimum and maximum values that are not

outliers (the limits of whiskers). The projection of uncertainty from input to output state variables can be clearly seen in Fig. 2, and the uncertainty propagation is different from one state to another. As expected, the median of the outputs with input uncertainty is close to the output considering no input uncertainty.
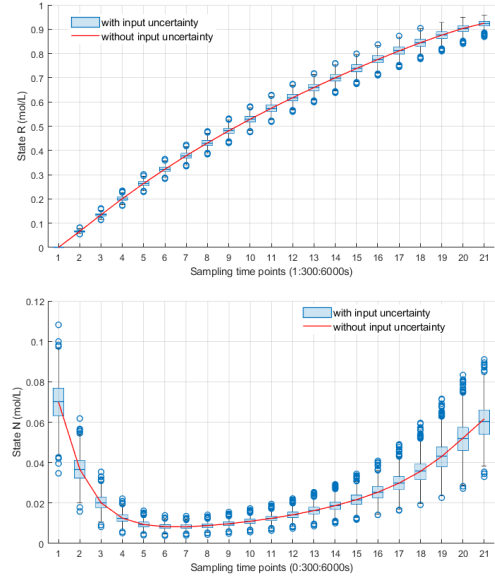

Fig. 2. Time response of states R and N with and without input uncertainty

Using the simulated pseudo-measurement data, 1000 estimated parameter sets are obtained by solving the optimisation problem in (5). Two norms, the L-1 norm and the L-2 norm, are used in the objective function, and the results are shown in Fig. 3. The three figures in the left column show CI ellipsoids approximated by scatter plots from the 1000 sets of estimated parameters and the CIs without input uncertainty. The three figures in the right column are the box charts illustrating the uncertainty in parameter estimation (median, quartiles, outliers and limits of non-outliers).

Fig. 3 shows that the uncertainty in parameter estimation becomes larger than that when no input uncertainty is introduced. k_{-3} is the most affected parameter with more outliers and wider range of non-outliers. Besides, the use of L-1 norm seems to be more robust than the L-2 norm for parameter estimation of k_2 and $k_5 W$, for it produces fewer outliers. However, this is not the case for k_{-3}, for which the L-2 norm result has less outliers.

C. Robust Measurement Set Selection Design

As discussed above, input uncertainty (even a small one) affects parameter estimation quality, leading to more outliers and larger CIs. To improve experiments against input uncertainty, a robust measurement set selection design is proposed in (10). This optimisation problem can be easily solved by ranking the squared deviation between outputs with and without input uncertainty. This deviation $\mathbf{D}$ is presented in Fig. 4 in time response profiles and the integration bar chart.
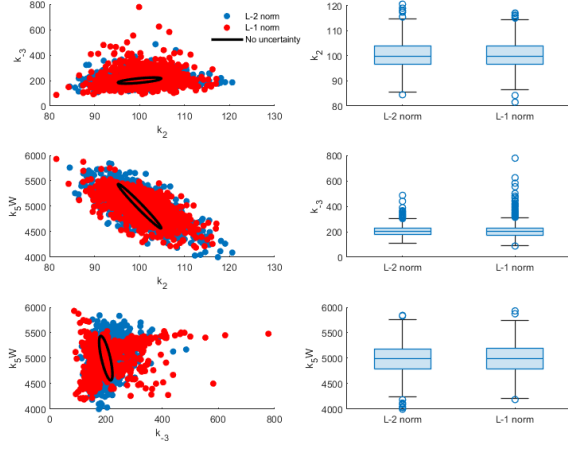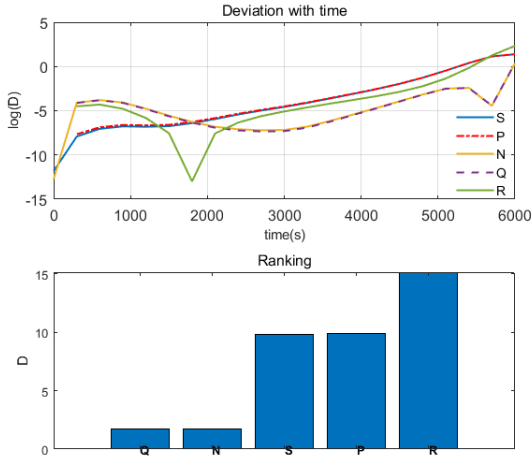
Fig. 3. Parameter estimation sets using L-1 & L-2 norm

Fig. 5. Parameter estimation sets (L-2 norm) using pseudo-measurements with robust measurement selection design and no RED

Fig. 4. The squared deviation (D) of outputs

Fig. 6. Parameter estimation sets (L-1 norm) using pseudo-measurements with robust measurement selection design and no RED

Following the results in Fig. 4, Q and N are selected as outputs since they are less affected by the input uncertainty among the five measurable states. Using the pseudo-data of Q and N, parameter estimation is simulated with L-1 and L-2 norms in (5). The parameter estimation results with L-2 norm are shown in Fig. 5, from which we it can be seen that when selecting Q and N only, the estimation error range for k_{-3} is smaller than using five states (no RED), but the estimation error ranges for k_2 and k_5W are larger compared to using all five states. The improved estimation in k_{-3} (the parameter mostly influenced by input uncertainty) shows that the robust measurement selection design has mitigated the effect of input uncertainty. The increased estimation errors in k_2 and k_5W may be caused by the loss of information content since three measurement (S, P, R) are excluded from the data.

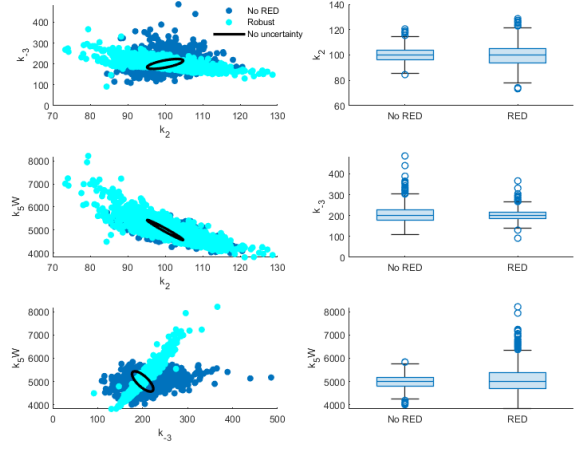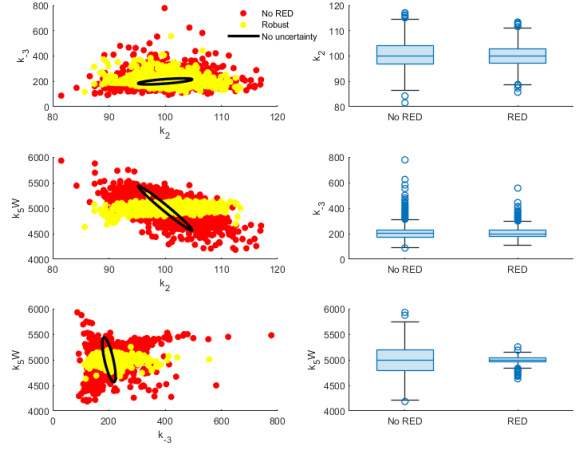The parameter estimation results with L-1 norm are shown in Fig. 6. Using the two measurement states selected by RED,

the estimation quality for all three parameters are improved compared to using the five states without RED.

The parameter estimation 'confidence intervals' and percentage uncertainty values are listed in TABLE IV. Four scenarios are compared: RED with L-1 and L-2 norm in parameter estimation, five states (no RED) with L-1 and L-2 norm in estimation. Here the most extreme 5% parameter samples are discarded, the range of the rest 95% samples is defined as 'confidence interval' for each parameter.

Comparing TABLE III and TABLE IV, it can be observed that for a nonlinear dynamic system, even with a small input uncertainty (standard deviation being about 1% of nominal value), the impacts on outputs and parameter estimation can be large when no RED is applied. For parameters that are heavily impacted by the input uncertainty (k_{-3} in the case study), the parameter estimation credibility can be quite low.

	$k_2(100)$	$k_{-3}(200)$	$k_5W(5000)$
no design (L-2)	[88.72 110.89] 22.17%	[118.21 284.57] 83.18%	[4394.9 5581.6] 23.73%
RED (L-2)	[83.18 116.28] 33.1%	[147.11 250.65] 51.77%	[3991.2 6176.5] 43.71%
no design (L-1)	[89.73 110.33] 20.6%	[87.53 330.21] 121.34%	[4487.3 5488.0] 20.02%
RED (L-1)	[90.38 109.26] 18.88%	[111.80 293.12] 90.66%	[4862.4 5131.0] 5.3%

With robust measurement set selection design and L-1 norm in parameter estimation, the impact of input uncertainty to parameter estimation can be reduced, although the improvement may not be ideal for parameters that are heavily impacted by input uncertainty.

V. CONCLUSIONS

Inaccurate setting of initial states is a source of input uncertainty in experiments that propagates to output through the dynamic system, and as a consequence, deteriorates data quality for parameter estimation. In this work, the impact of input uncertainty on parameter estimation is discussed. To reduce such impact, solutions of RED and use of more robust objective function in parameter estimation are suggested.

To examine how the input uncertainty impacts outputs and parameter estimation, the Monte-Carlo method is employed. Samples are drawn from input uncertainty space, which is assumed to obey Gaussian distribution with the mean being the designed input from OED and the variance being approximately 1% of the designed input. By simulation of the system dynamic model, output samples are obtained, and the population statistical characteristics of samples are utilised to approximate the output uncertainty. From the case study of an enzyme reaction system, it can be clearly seen that the outputs are influenced by input uncertainty. With those output samples, parameter estimations are implemented using the L-2 norm and the L-1 norm in the residual function. Results show that the uncertainty of parameter estimates is heavily impacted by input uncertainty. The L-1 norm seems to be more robust to data uncertainties compared to the L-2 norm.

In order to reduce the impact of input uncertainty on parameter estimation, we propose a new RED method to select measurement state variables, in which those states with the smallest deviation from the nominal states (no input uncertainty introduced) are selected as measurement states in output. With the combined use of RED on measurement selection and the L-1 norm residual function in parameter estimation, the quality of parameter estimation has been improved. However, for parameters significantly influenced by input uncertainty, like k_{-3}, the improvement is still limited. Further experimental design strategies need to be explored to 'decouple' the influence from input uncertainty to parameter estimation.

REFERENCES

[1] É. Walter and L. Pronzato, "Optimal experiment design for nonlinear models subject to large prior uncertainties," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 253, no. 3, pp. R530–R534, 1987.

[2] S. Asprey and S. Macchietto, "Designing robust optimal dynamic experiments," *Journal of Process Control*, vol. 12, no. 4, pp. 545–556, 2002.

[3] G. Franceschini and S. Macchietto, "Model-based design of experiments for parameter precision: State of the art," *Chemical Engineering Science*, vol. 63, no. 19, pp. 4846–4872, 2008.

[4] F. Galvanin, S. Macchietto, and F. Bezzo, "Model-based design of parallel experiments," *Industrial & engineering chemistry research*, vol. 46, no. 3, pp. 871–882, 2007.

[5] J. Stigter, D. Vries, and K. Keesman, "On adaptive optimal input design: a bioreactor case study," *AIChE journal*, vol. 52, no. 9, pp. 3290–3296, 2006.

[6] F. Galvanin, M. Barolo, and F. Bezzo, "Online model-based redesign of experiments for parameter estimation in dynamic systems," *Industrial & Engineering Chemistry Research*, vol. 48, no. 9, pp. 4415–4427, 2009.

[7] S. Iwazaki, Y. Inatsu, and I. Takeuchi, "Bayesian experimental design for finding reliable level set under input uncertainty," *IEEE Access*, vol. 8, pp. 203982–203993, 2020.

[8] R. Hille, J. Mandur, and H. M. Budman, "Robust batch-to-batch optimization in the presence of model-plant mismatch and input uncertainty," *AIChE Journal*, vol. 63, no. 7, pp. 2660–2670, 2017.

[9] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems*. Elsevier, 2018.

[10] D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev, "Why the monte carlo method is so important today," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 6, no. 6, pp. 386–392, 2014.

[11] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003.

[12] H. Yue, P. Halling, and H. Yu, "Model development and optimal experimental design of a kinetically controlled synthesis system," *IFAC Proceedings Volumes*, vol. 46, no. 31, pp. 327–332, 2013.

APPENDIX

The ODEs of enzyme reaction system are:

$$\frac{dE}{dt} = -k_1 E \cdot S + k_{-1}ES + k_4 EQ - k_{-4}E \cdot Q + k_6 ER$$

$$\frac{dES}{dt} = k_1 E \cdot S - k_{-1}ES - k_2 ES + k_{-2}E^* \cdot P$$

$$\frac{dE^*}{dt} = k_2 ES - k_{-2}E^* \cdot P - k_3 E^* \cdot N + k_{-3}EQ - k_5 W \cdot E^* + k_{-5}ER$$

$$\frac{dEQ}{dt} = k_3 E^* \cdot N - k_{-3}EQ - k_4 EQ + k_{-4}E \cdot Q$$

$$\frac{dER}{dt} = k_5 W \cdot E^* - k_{-5}ER - k_6 ER$$

$$\frac{dS}{dt} = -k_1 E \cdot S + k_{-1}ES,$$

$$\frac{dP}{dt} = k_2 ES - k_{-2}E^* \cdot P$$

$$\frac{dN}{dt} = -k_3 E^* \cdot N + k_{-3}EQ$$

$$\frac{dQ}{dt} = k_4 EQ - k_{-4}E \cdot Q$$

$$\frac{dR}{dt} = k_6 ER$$