Parameter estimation and prediction uncertainties for multi-response kinetic models with uncertain inputs

Kaveh Abdi[1], Benoit Celse[2], Kimberley B. McAuley[1*]

1. Department of Chemical Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada
2. IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France

Abstract

Error-in-variables model (EVM) methods are used for parameter estimation when independent variables are uncertain. During EVM parameter estimation, output measurement variances are required as weighting factors in the objective function. These variances can be estimated based on data from replicate experiments. However, conducting replicates is complicated when independent variables are uncertain. Instead, pseudo-replicate runs may be performed where the target values of inputs for repeated runs are the same, but the true input values may be different. Here, we propose a method to estimate output-measurement variances for use in multivariate EVM estimation problems, based on pseudo-replicate data. We also propose a bootstrap technique for quantifying uncertainties in resulting parameter estimates and model predictions. The methods are illustrated using a case study involving n-hexane hydroisomerization in a well-mixed reactor. Case-study results reveal that assumptions about input uncertainties can have important influences on parameter estimates, model predictions and their confidence intervals.

KEYWORDS: Error-in-variables model, Parameter estimation, Uncertainty quantification, Bootstrap, Kinetic model

1- Introduction

Fundamental kinetic models are important for chemical process development and improvement. These models usually contain kinetic parameters and activation energies, which are estimated using conventional parameter estimation techniques such as weighted least squares (WLS). In WLS parameter estimation, the independent variables (model inputs) are assumed to be perfectly known and measurement errors are considered for the dependent variables (model outputs). However, in some experimental situations, input errors are relatively large compared to output measurement errors, which can result in inaccurate parameter estimates.[1,2] Error-in-variables model (EVM) parameter estimation methods have been developed to address this issue by accounting for input uncertainties during parameter estimation. EVM objective functions for multi-response models are typically of the form:[3]

$$J_{EVM} = \sum_{i=1}^{N} \sum_{k=1}^{N_Y} \frac{\left(y_{m\,ik} - g_k(\boldsymbol{x_i}, \boldsymbol{u_i}, \boldsymbol{\theta})\right)^2}{\sigma_{Y_k}^2} + \sum_{i=1}^{N} \sum_{l=1}^{N_U} \frac{(u_{m\,il} - u_{il})^2}{\sigma_{U_l}^2} \qquad (1)$$

where $y_{m\,ik}$ is the k^{th} $(k = 1, \ldots, N_Y)$ measured output response obtained using the i^{th} target condition, $g_k(\boldsymbol{x_i}, \boldsymbol{u_i}, \boldsymbol{\theta})$ is the corresponding model solution obtained from the perfectly known model inputs $\boldsymbol{x_i} \in R^{\boldsymbol{N_x}}$, the imperfectly-known inputs $\boldsymbol{u_i} \in R^{\boldsymbol{N_U}}$ and the model parameters $\boldsymbol{\theta} \in R^{\boldsymbol{N_\theta}}$. In equation (1), $\sigma_{Y_k}^2$ is the measurement variance for the k^{th} model output and $\sigma_{U_l}^2$ is the measurement variance for the l^{th} $(l = 1, \ldots, N_U)$ uncertain input. In the second term, $u_{m\,il}$ is the measured value of the l^{th} uncertain input at the i^{th} experimental setting and u_{il} is the l^{th} element in the corresponding vector $\boldsymbol{u_i}$ of true values for the uncertain inputs.

During EVM parameter estimation, parameters $\boldsymbol{\theta}$ and uncertain inputs $\boldsymbol{u_i}$ for each experimental run are estimated by minimizing J_{EVM}. EVM parameter estimation is more complicated than WLS because EVM requires additional information about the uncertainties in inputs (i.e., values $\sigma_{U_l}^2$) and EVM involves additional decision variables in the optimization problem. Variances $\sigma_{U_l}^2$ in process inputs are usually assumed known, based on the information that experimentalists have regarding the reproducibility of inputs

(e.g., based on variations involved in analytical devices). In the chemical engineering literature, output measurement variances are also usually assumed to be known *a priori* in most EVM parameter estimation studies.[4] However, modelers can also use replicate data to estimate the output measurement variances $\sigma_{Y_k}^2$.[2,4-6] The resulting variance estimates should account for the measurement variability for the output responses as well as all other sources of experimental variations (e.g., ambient condition and sample collection) except variability associated with $\boldsymbol{u_i}$.

Conducting replicate experiments in EVM situations is not straight-forward because uncertain inputs should be kept constant during different replicate runs. In some situations, it is possible to perform true replicate (TR) experiments (e.g., when the same uncertain stock feed solution is used to perform multiple experiments). In other situations involving pseudo replicate (PR) experiments, the true values of uncertain inputs cannot be enforced to be the same during attempted replicate runs. In TR situations, quantifying output measurement variances $\sigma_{Y_k}^2$ is relatively straightforward (e.g., using pooled variance estimates from several different experimental conditions).[4] Estimation of output measurement variances from pseudo-replicate data is more difficult because the contribution of uncertain inputs to the overall variability in the outputs should be taken into account. In a previous study, we devised a method to estimate $\sigma_{Y_k}^2$ based on pseudo-replicates, but we only considered a simple single-response copolymerization model.[4] One of the objectives in the current paper is to extend this method for use in multi-response models. We use a n-hexane hydroisomerization model as a case study to illustrate the proposed methodology.[7] Further, we extend a bootstrap technique developed for parameter uncertainty quantification in single-response models so that it can be used for multi-response models and we demonstrate the technique using the case study. We also show how to obtain model prediction uncertainties based on the bootstrap parameter uncertainties, input uncertainties and estimated output uncertainties.

2- Background

2.1- Using Pseudo-Replicate Experiments to Estimate Parameters and Obtain Output Measurement Variances

Consider a situation where an experimenter attempts to perform n_i replicate experiments at the i^{th} target run condition where the desired flow rate of one species is $5.0\frac{mmol}{s}$ and the desired flow rate of a second species is $0.10\ \frac{mmol}{s}$. Assume that both of the flow rates contain significant uncertainty and that the experimenter makes adjustments to valves until the corresponding measured flow rates are $\boldsymbol{u_{m\,i}} = \left[5.0\frac{mmol}{s}, 0.10\frac{mmol}{s}\right]^T$. If attempted replicate experiments are conducted on different days, where the measured values are the same on each day, but the underlying true flow rates are different, then the experimenter performs pseudo replicates.

In general, for a multi-response system with several uncertain inputs, an error-in-variables model of the process can be expressed as:[4]

$$\boldsymbol{Y_{ij}} = \boldsymbol{g}\left(\boldsymbol{x}_i, \boldsymbol{u}_{ij}, \boldsymbol{\theta}\right) + \boldsymbol{\varepsilon_{Y\,ij}} \qquad (2)$$

$$\boldsymbol{u_{m\,i}} = \boldsymbol{u_{ij}} + \boldsymbol{\varepsilon_{U\,ij}} \qquad (3)$$

where $\boldsymbol{Y_{ij}} \in R^{N_Y}$ is a vector of uncertain measurements for N_Y dependent variables obtained from the j^{th} $(j = 1, \ldots, n_i)$ pseudo-replicate experiment conducted using the i^{th} target condition. A capital $\boldsymbol{Y}$ is used to indicate that $\boldsymbol{Y}_{ij}$ is a random variable due to the measurement noise $\boldsymbol{\varepsilon_{Y,ij}}$ on the right-hand side of equation (2). Model predictions in vector $\boldsymbol{g}\left(\boldsymbol{x}_i, \boldsymbol{u}_{ij}, \boldsymbol{\theta}\right)$ depend on the model parameters $\boldsymbol{\theta}$, the true values of the uncertain inputs $\boldsymbol{u}_{ij}$ and any perfectly-known inputs $\boldsymbol{x}_i$. In equations (2) and (3), $\boldsymbol{\varepsilon_{Y\,ij}}$ and $\boldsymbol{\varepsilon_{U\,ij}}$ are vectors of normally-distributed random noise for the output and input measurements, respectively. Note that $\boldsymbol{\varepsilon_{Y\,ij}}$ accounts mainly for measurement error in the j^{th} dependent variable, as well as additional variability in the

experiments which is not attributed to $\boldsymbol{\varepsilon_{U\,ij}}$. For simplicity, throughout the remainder of this article we will refer to $\boldsymbol{\varepsilon_{Y\,ij}}$ as output measurement uncertainty.

Assuming that input and output measurement noise vectors in equations (2) and (3) are independent, an EVM objective function can be obtained from maximum-likelihood arguments for situations involving pseudo-replicate experiments:[4]

$$J_{EVM} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\left(\boldsymbol{y_{m\,ij}} - \boldsymbol{g}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta})\right)^{\mathbf{T}} \boldsymbol{\Sigma_Y^{-1}}\left(\boldsymbol{y_{m\,ij}} - \boldsymbol{g}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta})\right) + \left(\boldsymbol{u_{m\,i}} - \boldsymbol{u_{ij}}\right)^{T}\boldsymbol{\Sigma_U^{-1}}\left(\boldsymbol{u_{m\,i}} - \boldsymbol{u_{ij}}\right)\right] \quad (4)$$

where $\boldsymbol{y_{m\,ij}}$ is the vector of measured values for the dependent variables from the j^{th} run conducted at the i^{th} experimental target condition. In the current study, we assume that $\boldsymbol{\Sigma_Y}$ and $\boldsymbol{\Sigma_U}$ the covariance matrices for random errors in the dependent and independent variables, respectively, are diagonal. Diagonal elements in $\boldsymbol{\Sigma_U}$ (i.e., $\sigma_{U_l}^2$ where $l = 1, \ldots, N_U$) are variances associated with the measurement or implementation of the uncertain inputs, which are assumed to be known based on the experimenter's experience with the experimental setup. The diagonal elements in $\boldsymbol{\Sigma_Y}$ (i.e., $\sigma_{Y_k}^2$ where $k = 1, \ldots, N_Y$) are unknown and need to be estimated from the pseudo-replicate data.[4] In a previous study, we developed a linearization-based approach to obtain estimates of output measurement variances for each of the N_Y dependent variables based on the pseudo-replicate data obtained at the i^{th} target condition. For the k^{th} dependent variable:

$$\hat{\sigma}_{Y_{ik}}^2 = \frac{\sum_{j=1}^{n_i}\left(y_{m\,ijk} - \bar{y}_{m\,ik}\right)^2}{n_i - 1} - \left(\left.\frac{\partial g_k(\boldsymbol{x_i}, \boldsymbol{u}, \boldsymbol{\theta})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}}, \boldsymbol{x_i}, \widehat{\boldsymbol{\theta}}}\right)\boldsymbol{\Sigma_U}\left(\left.\frac{\partial g_k(\boldsymbol{x_i}, \boldsymbol{u}, \boldsymbol{\theta})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}}, \boldsymbol{x_i}, \widehat{\boldsymbol{\theta}}}\right)^{T} \quad (5)$$

where $y_{m\,ijk}$ is the measured value of the k^{th} dependent variable obtained from the j^{th} pseudo-replicate run at the i^{th} experimental condition and $\bar{y}_{m\,ik}$ is the average of the n_i measured values for these runs. The second term on the right-hand side of equation (5) accounts for the contribution of uncertain inputs on the overall variability of the dependent variables. Sometimes, $\hat{\sigma}_{Y_{ik}}^2$ obtained from equation (5) can be negative, so a cut-off value δ_k is used to obtain reliable variance estimates, based on the modeler's prior knowledge:

$$\hat{\sigma}^2_{Y_{ik}} = \max\left(\frac{\sum_{j=1}^{n_i}\left(y_{m\,ijk}-\bar{y}_{m\,ik}\right)^2}{n_i-1} - \left(\left.\frac{\partial g_k(\boldsymbol{x_i},\boldsymbol{u},\boldsymbol{\theta})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}},\boldsymbol{x_i},\widehat{\boldsymbol{\theta}}}\right)\boldsymbol{\Sigma_U}\left(\left.\frac{\partial g_k(\boldsymbol{x_i},\boldsymbol{u},\boldsymbol{\theta})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}},\boldsymbol{x_i},\widehat{\boldsymbol{\theta}}}\right)^T, \delta_k\right) \quad (6)$$

The estimated variances obtained from equation (6) can be pooled together to obtain:

$$\hat{\sigma}^2_{Y_k} = \frac{\sum_{i=1}^{N}(n_i-1)\hat{\sigma}^2_{Y_{ik}}}{\sum_{i=1}^{N}(n_i-1)} \quad (k = 1, \ldots, N_Y) \qquad (7)$$

Note that, the parameter estimates $\widehat{\boldsymbol{\theta}}$ are required in equation (6) to calculate the Jacobian matrix $\left.\frac{\partial g_k(\boldsymbol{x_i},\boldsymbol{u},\boldsymbol{\theta})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}},\boldsymbol{x_i},\widehat{\boldsymbol{\theta}}}$. So, an iterative approach was developed to update the elements of $\left.\frac{\partial g_k(\boldsymbol{x_i},\boldsymbol{u},\boldsymbol{\theta})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}},\boldsymbol{x_i},\widehat{\boldsymbol{\theta}}}$ as improved parameter estimates become available (see algorithm in Table 1).[4] This method is used to obtain an estimate of $\boldsymbol{\Sigma_Y}$ in the hexane hydroisomerization case study in section 4.

Table 1- Procedure for EVM Parameter Estimation Using Pseudo-Replicate Data[4]

1- Assign appropriate cut-off values δ_k $(k = 1, \ldots, N_Y)$ for each of the measured outputs variances. Set the step counter to *s*=0.
2- Obtain an initial guess for the output measurement variances by treating the attempted replicates as true replicates (i.e., using $\hat{\sigma}^2_{Y_{ik}} = \frac{\sum_{j=1}^{n_i}\left(y_{m\,ijk}-\bar{y}_{m\,ik}\right)^2}{n_i-1}$ to calculate variance in each experimental setting) and using equation (7) to pool variance estimates.
3- Obtain initial parameter estimates $\widehat{\boldsymbol{\theta}}^{(0)}$ using weighted-least squares parameter estimation.
4- For each PR target condition and measured output, calculate $\hat{\sigma}^2_{Y_{ik}}$ from Equation (6) using the most-recent parameter estimates $\widehat{\boldsymbol{\theta}}^{(s)}$.
5- Use equation (7) to pool the variance estimates obtained in step 4.
6- Use J_{EVM} in equation (4) to perform an EVM parameter estimation, using the $\hat{\sigma}^2_{Y_k}$ as diagonal elements of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$, resulting in updated parameter estimates $\widehat{\boldsymbol{\theta}}^{(s+1)}$.
7- Calculate the relative change in the parameter values $e = \sqrt{\sum_{\boldsymbol{p=1}}^{\boldsymbol{N_\theta}}\left(\frac{\hat{\theta}_p^{(s+1)}-\hat{\theta}_p^{(s)}}{\hat{\boldsymbol{\theta}}_{\boldsymbol{p}}^{(s)}}\right)^{\mathbf{2}}}$ where subscript p denotes the p^{th} element in $\widehat{\boldsymbol{\theta}}$. If e is smaller than a tolerance set by the user, stop and report the parameter values. Otherwise, increase the value of *s* by one and return to step 4.

2.2- Parameter Uncertainty Quantification Using Bootstrapping for a Single-response Model with a Single Uncertain Input

In a previous study, we extended a parametric bootstrap technique for models with perfectly-known independent variables[8] so that it can be used for parameter uncertainty quantification in EVM.[4] Synthetic

data sets are generated multiple times, using typical input and output uncertainties, and parameters are estimated for each synthetic data set. The resulting cloud of bootstrap parameter estimates is then used for quantifying uncertainties in parameter estimates. Previously, we used this bootstrap method to quantify uncertainties in estimated reactivity-ratio parameters in a single-response Mayo-Lewis model with a single uncertain input.[4] In section 3.1, this bootstrapping methodology is extended for use in multi-response models with multiple uncertain inputs.

An alternative approach to bootstrapping would be to sample from a posterior distribution for parameter estimates obtained using an Bayesian EVM approach based on Markov-Chain Monte Carlo simulations.[1] In the current article, we decided to extend the bootstrapping approach instead, because our hexane hydroisomerization case study does not involve prior information about parameter values that should be imposed during estimation. Also, we believe that the proposed bootstrapping approach will be simpler for model developers to use than Bayesian methods.

2.3- Quantifying Prediction Uncertainties based on Uncertainties in Inputs and Parameter Estimates

In a previous study, we reviewed methods for quantifying prediction uncertainties when inputs are uncertain.[9] Linearization-based and MC-based techniques are typically used for this purpose.[9] MC-based methods are usually more accurate, but they require computationally intensive simulations when the model equations are complex. Nevertheless, they are relatively straight-forward and can readily handle situations where a cloud of parameter estimates is available from bootstrap computations. Past researchers have sampled from the distribution of the parameters and inputs simultaneously to propagate the corresponding uncertainties into the model predictions.[10-20] Table 2 provides an algorithm for quantifying uncertainties in model predictions using a cloud of b_{max} parameter estimates available from bootstrapping. This algorithm quantifies the uncertainties in a vector of model predictions $\hat{\boldsymbol{y}}_{\boldsymbol{c}}$ at an experimental condition of interest c. The results from Table 2 provide 95 % confidence intervals for noise-free values of the dependent variables

at the condition of interest. If the modeler is interested instead in a 95% confidence interval on noisy measured values that might be obtained from a future experiment, then additive measurement errors in the dependent variables must also be considered.[9] Both types of prediction uncertainties are computed in the case study in section 4.

Table 2. MC method for obtaining approximate 95% confidence intervals for mean model predictions at an experimental condition of interest [9]

1- Specify experimental settings for inputs of interest (i.e., values of $\boldsymbol{x_c}$ and $\boldsymbol{u_{c\,m}}$) and the elements of $\boldsymbol{\Sigma_U}$. Sample from the distribution of plausible true values $\boldsymbol{u_c}$ for the uncertain inputs to obtain b_{max} candidate values for the uncertain input vector (i.e., $\boldsymbol{u_c^{(1)}}$, $\boldsymbol{u_c^{(2)}}$ ... $\boldsymbol{u_c^{(b_{max})}}$), where b_{max} is a large number.

2- Use the cloud of parameters from a previous bootstrap (Monte Carlo) uncertainty quantification ($\boldsymbol{\hat{\theta}^{(1)}}, \boldsymbol{\hat{\theta}^{(2)}}, \dots, \boldsymbol{\hat{\theta}^{(b_{max})}}$).

3- For $b = 1$ to b_{max} , calculate the corresponding predicted values for the model outputs:

$$\hat{y}_{c\,k}^{(b)} = g_k(\boldsymbol{\hat{\theta}^{(b)}}, \boldsymbol{u_c}^{(b)}, \boldsymbol{x_c}) \quad (k = 1, \dots, N_Y) \quad (2.1)$$

4- Obtain the 2.5 and 97.5 percentile values of $\hat{y}_{c\,k}^{(b)}$. These values are approximate bounds for the 95% confidence intervals for the mean predicted responses.

3- PROPOSED METHODOLOGY

3-1 Bootstrap Method for Quantifying Parameter Uncertainties in Multi-response EVM

In the current study, we extend a bootstrap algorithm for quantifying parameter uncertainties in single-response models so that it can be used for multi-response models with several uncertain inputs. Steps required for implementing the extended algorithm are shown in Table 3. In the first step, parameters and elements of the output covariance matrix $\boldsymbol{\Sigma_Y}$ are estimated using the algorithm in Table 1. In steps 2 and 3, synthetic data are generated and their corresponding bootstrap parameters are estimated. In step 4, uncertainty information is extracted from the cloud of bootstrap parameter estimates.

Table 3- Bootstrap steps to obtain parameter uncertainties for multi-response error-in-variable models with several uncertain inputs using pseudo-replicate experiments

1- Estimate the model parameters $\widehat{\boldsymbol{\theta}}$ and the diagonal output measurement covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$ using the algorithm in Table 1. Assign bootstrap counter b=1 and set the desired number of bootstrap runs b_{max}.

2- Generate synthetic bootstrap data using $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}$:

$\boldsymbol{u}_{\boldsymbol{ij}}^{(\boldsymbol{b})} = \boldsymbol{u}_{\boldsymbol{m}\,\boldsymbol{i}} - \boldsymbol{\varepsilon}_{\boldsymbol{u}\,\boldsymbol{ij}}^{(\boldsymbol{b})}$ for $i = 1, \dots, N;\ j = 1, \dots, n_i$ where $\boldsymbol{\varepsilon}_{\boldsymbol{u}\,\boldsymbol{ij}}^{(\boldsymbol{b})} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{U}})$

$\boldsymbol{y}_{\boldsymbol{m}\,\boldsymbol{ij}}^{(\boldsymbol{b})} = g\left(\boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{u}_{\boldsymbol{ij}}^{(\boldsymbol{b})}, \widehat{\boldsymbol{\theta}}\right) + \boldsymbol{\varepsilon}_{\boldsymbol{y}\,\boldsymbol{ij}}^{(\boldsymbol{b})}$ for $i = 1, \dots, N;\ j = 1, \dots, n_i$ where $\boldsymbol{\varepsilon}_{\boldsymbol{y}\,\boldsymbol{ij}}^{(\boldsymbol{b})} \sim N(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}_{Y})$

3- Obtain bootstrap parameter estimates $\widehat{\boldsymbol{\theta}}^{(b)}$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}_{Y}^{(b)}$ using the b^{th} synthetic data set. If $b = b_{max}$ go to step 4; otherwise increase b by one and go to step 2.

4- Use the bootstrap parameter estimates $\widehat{\boldsymbol{\theta}}^{(\boldsymbol{b})}$ $(b = 1,2, \dots, b_{max})$ to construct confidence intervals or box plots for the parameters.

3-2- Bootstrap Method for Quantifying Prediction Uncertainties from Error-in-Variables Models

After parameter uncertainties are obtained using the multi-variate bootstrap technique in Table 3, a cloud of b_{max} parameter estimates is available. Using this cloud of parameter estimates, it is straight-forward to obtain the corresponding cloud of model predictions using the algorithm in Table 2. Note that the algorithm in Table 2 provides an estimate of uncertainty in the mean predicted responses at the target condition of interest. In section 4, we will demonstrate how to also obtain the uncertainty associated with measured responses for a single future experiment. Uncertainty associated with a possible measured value from a single new experiment is larger than the corresponding uncertainty in the mean response due to the associated measurement error.

4- Case Study: Catalytic n-Hexane Hydroisomerization in a Well-mixed Reactor

4-1- Model Equations and Experimental System

Here we present a multi-response model developed by Toch et al. to describe hydroisomerization of hexane (nC_6) in a catalytic well-mixed continuous reactor. The reaction scheme for this process is shown in Table 4. Products 2-methyl pentane (2MP) and 3-methyl pentane (3MP) are produced along with propane (C_3) as an undesirable side product.[7] The corresponding algebraic equations obtained from mole

balances are shown in Table 5. Derivations are provided in the supplementary information. In Table 5, F_q^0 and F_q are the molar inflow and outflow rates for species q ($q = H_2, nC_6$, *2MP*, *3MP* and C_3), respectively, W is the weight of catalyst (i.e., H-ZSM-5) and K_{phys} is the equilibrium constant for physisorption of n-hexane on the catalyst. The rate coefficients k_r ($r = 1,2,3$) and the equilibrium constant K_{phys} are temperature-dependant and are calculated using the equations in Table 6 where $k_{ref\,r}$ is the rate coefficient of the r^{th} reaction at the reference temperature $T_{ref} = 531.48\,K$ and K_{ref} is the corresponding equilibrium coefficient. In equations (6.1) and (6.2), $E_{a\,r}$ is the activation energy of the r^{th} reaction and ΔH^0 is the enthalpy of physisorption.[7]

As shown in the second column of Table 6, there are 8 unknown parameters $\boldsymbol{\theta} = \left[k_{ref\,1}, k_{ref\,2}, k_{ref\,3}, K_{ref}, E_{a\,1}, E_{a\,2}, E_{a\,3}, \Delta H^0\right]^T$ requiring estimation. When the values of the parameters and the experimental inputs (i.e., reactor pressure *P*, reactor temperature *T*, catalyst weight *W* and flow rates $F_{nC_6}^0$ and $F_{H_2}^0$) are given, the three equations in three unknowns in Table 5 can be solved numerically to predict F_{2MP}, F_{3MP} and F_{C3}. After predictions for the outputs are obtained, yield can be calculated:

$$Yield = \frac{(F_{2MP}+F_{3MP})}{F_{nC_6}^0} \qquad (8)$$

and selectivity for different species can also be calculated:

$$S_q = \frac{\alpha_q}{6}\frac{F_q}{F_{nC_6}^0 - F_{nC_6}} \qquad (q = 2MP, 3MP \text{ and } C_3) \qquad (9)$$

where α_q is the number of carbon atoms in the species molecules (i.e., $\alpha_{2MP} = \alpha_{3MP} = 6$ and $\alpha_{C_3} = 3$).

Table 4- Reaction Scheme Proposed for nC_6 Hydroisomerization[7]

$nC_6 \xrightarrow{k_1} 2MP$
$nC_6 \xrightarrow{k_2} 3MP$
$2MP + H_2 \xrightarrow{k_3} 2\ C_3$

Table 5- Algebraic Equations based on the Mole Balance Equations for Different Species

Species	Corresponding Mole Balance Equation
2MP	$F_{2MP} - \left(\frac{\frac{k_1 (F^0_{nC_6} - F_{2MP} - F_{3MP} - \frac{F_{C_3}}{2}) - k_3 F_{2MP}}{F^0_{H_2} - \frac{F_{C_3}}{2}}}{1 + K_{phys} P \left(\frac{F^0_{nC_6} - \frac{F_{C_3}}{2}}{F^0_{H_2} + F^0_{C_6}} \right)} \right) W = 0$ (5.1)
3MP	$F_{3MP} - \left(\frac{\frac{k_2 (F^0_{nC_6} - F_{2MP} - F_{3MP} - \frac{F_{C_3}}{2})}{F^0_{H_2} - \frac{F_{C_3}}{2}}}{1 + K_{phys} P \left(\frac{F^0_{nC_6} - \frac{F_{C_3}}{2}}{F^0_{H_2} + F^0_{C_6}} \right)} \right) W = 0$ (5.2)
C_3	$F_{C_3} - \left(\frac{\frac{2k_3 \frac{F_{2MP}}{F^0_{H_2} - \frac{F_{C_3}}{2}}}{1 + K_{phys} P \left(\frac{F^0_{nC_6} - \frac{F_{C_3}}{2}}{F^0_{H_2} + F^0_{C_6}} \right)} \right) W = 0$ (5.3)

Table 6- Kinetic rate and equilibrium constant equations for solving balance equations and the corresponding parameters requiring estimation

Kinetic rate and equilibrium constant equation	Parameters Requiring Estimation
$k_r = k_{ref\ r}\ e^{-\frac{E_{a\ r}}{R}(\frac{1}{T} - \frac{1}{T_{ref}})} \quad (r = 1,2,3)$ (6.1)	$k_{ref\ 1}, k_{ref\ 2}, k_{ref\ 3}, E_{a\ 1}, E_{a\ 2}, E_{a\ 3}$
$K_{phys} = K_{ref} e^{-\frac{\Delta H^0}{R}(\frac{1}{T} - \frac{1}{T_{ref}})}$ (6.2)	K_{ref} and ΔH^0

Toch et al. performed 36 experimental runs at a variety of conditions to estimate the model parameters, assuming that all inputs were perfectly known (see Table S.1 in the supplementary information). In the current case study, we consider two uncertain inputs $\boldsymbol{u} = \left[F_{H_2}^0 \text{ and } F_{nC_6}^0\right]^T$ and three perfectly-known inputs $\boldsymbol{x} = [P, T, W]^T$. We assume that the replicate experiments reported by Toch et al. are pseudo-replicates because slightly different true values of inputs $F_{H_2}^0$ and $F_{nC_6}^0$ would have been used when experiments were repeated. For the 36 experimental runs, there are 25 unique experimental conditions and 11 repeated runs conducted at a variety of different conditions. During each run, Toch et al. measured three outputs (i.e., measured flow rates for 2MP, 3MP and C_3). As shown in Table 7, three different levels of input uncertainties are considered, which are 1% and 5 % and 10% of the corresponding average measured input values in Table S.1. Corresponding measurement standard deviations are shown in the second and the third columns. In the current case study, we use the data in Table S1 multiple times, considering different levels of uncertainties in independent variables $F_{H_2}^0$ and $F_{nC_6}^0$ that could have been encountered by Toch et al. during their experiments.

Table 7- Different uncertainty level for inputs with measurement uncertainties

Levels of input uncertainties	Standard deviation for measurements of $F_{H_2}^0$ $(\sigma_{F_{H_2}^0})$	Standard deviation for measurements of $F_{nC_6}^0$ $(\sigma_{F_{nC_6}^0})$
1%	$1.63 \times 10^{-5} \frac{mol}{s}$	$2.56 \times 10^{-7} \frac{mol}{s}$
5%	$8.16 \times 10^{-5} \frac{mol}{s}$	$1.28 \times 10^{-6} \frac{mol}{s}$
10%	$1.63 \times 10^{-4} \frac{mol}{s}$	$2.56 \times 10^{-6} \frac{mol}{s}$

4-2 EVM Parameter and Variance Estimation

The EVM objective function and decision variables for estimating the case-study parameters are shown in the first row of Table 8. The vector of model predictions $\boldsymbol{g}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta})$ obtained from the j^{th} pseudo-replicate run at the i^{th} target experimental condition has three elements:

$$\boldsymbol{g}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta}) = \left[F_{2MP}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta}), F_{3MP}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta}), F_{C_3}(\boldsymbol{x_i}, \boldsymbol{u_{ij}}, \boldsymbol{\theta})\right]^T \quad (10)$$

where F_{2MP}, F_{3MP} and F_{C3} are predicted outflow rates obtained by solving the three equations in three unknowns in Table 5. As shown in Table 8, the vector $\boldsymbol{x_i}$ contains the perfectly-known inputs (i.e., temperature, pressure and catalyst weight) and $\boldsymbol{u}_{ij}$ contains the uncertain inputs (i.e., inflow rates of hydrogen and n-hexane). In equation (8.1), the vector $\boldsymbol{y_{m\,ij}} = \left[y_{2MP\,ij}, y_{3MP\,ij}, y_{C_3\,ij}\right]^T$ contains measured data for the dependent variables and the vector $\boldsymbol{u_{m\,i}} = \left[u_{mC_6\,i}, u_{mH_2\,i}\right]^T$ contains measured data for the uncertain independent variables.

In this study, we compare the results of the EVM parameter estimation, assuming several different levels of input uncertainties, with results obtained from WLS parameter estimation using objective function (8.2). The WLS parameter estimation problem is considerably simpler than the EVM problem because it involves fewer decision variables (i.e., 8 compared to 44).

Further, the estimation of the output measurement variances is simpler when using WLS. In WLS situations, the attempted replicates are assumed to be true replicates, from which the modeler can readily obtain pooled estimates of output measurement variances. In EVM situations considered in this case study, output measurement variances and parameters should be estimated simultaneously using the algorithm in Table 1. Note the that the vector of perfectly-known inputs $\boldsymbol{x_i}$ in the EVM case have three elements for each experimental setting, but x_i in the WLS situation has 5 elements because all the inputs are assumed to be perfectly known.

The results of the parameter estimation and the output measurement variance estimation for WLS and EVM situations are shown in Table 9. The variance estimates are provided in the 2^{nd} to the 5^{th} rows and the parameter estimates in the 6^{th} to 14^{th} rows. As shown in the Table, estimated values of the variances and parameters obtained using WLS are similar to corresponding values from EVM with 1 % input uncertainties, as expected. As such, in situations where input uncertainties are small relative to output uncertainties, we recommend that modelers use WLS rather than EVM because: i) WLS objective functions are simpler than EVM objective functions; ii) WLS estimation involves fewer decision variables than EVM; and iii) output measurement variances are simpler to estimate using the WLS assumption of true replicates, compared to the EVM assumption of pseudo replicates. Table 9 shows that, when larger input uncertainties are assumed (i.e., 5 % and 10 % of typical input values) the estimated parameters and output measurement variances are noticeably different using EVM compared to WLS. As expected, the estimated output measurement variances become smaller (see the 2^{nd} to the 5^{th} rows in Table 9) for increasing levels of input uncertainties, because a higher amount of variability in measured outputs in pseudo-replicate runs is attributed to the input uncertainties. Note that, the output measurement variances reported in Table 9 were computed using equations (6) and (7). We used cut-off values of $\delta_{F_{2MP}} = 1.44 \times 10^{-17}$, $\delta_{F_{3MP}}$=6.84×10^{-18} and $\delta_{F_{C_3}} = 1.94 \times 10^{-19}$ in the first step of the algorithm in Table 1. These cut-off values are 0.01 % of the pooled variance estimates for the dependent variables obtained from equation (7) when input uncertainties are neglected and all variability in the measurement values of the outputs in attempted replicate run is attributed to $\boldsymbol{\varepsilon}_{Yij}$.

Figure 1 shows the contribution of input uncertainties $\boldsymbol{\varepsilon}_{U\,ij}$ and output uncertainties $\boldsymbol{\varepsilon}_{Y\,ij}$ to the overall uncertainty in F_{2MP}. The heights of the red bars in Figure 1 were obtained from $\hat{\sigma}^2_{F_{2MP\,U}}/\hat{\sigma}^2_{F_{2MP\,o}}$ where the overall variance in the F_{2MP} responses was estimated from the pseudo-replicate data:

$$\hat{\sigma}^2_{F_{2MP\,o}} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(y_{2MP\,ij}-\bar{y}_{2MP\,i}\right)^2}{\sum_{i=1}^{N}(n_i-1)} \qquad (11)$$

and the input contribution to this overall variance was estimated from:

$$\hat{\sigma}^2_{F_{2MP}\,U} = \frac{\sum_{i=1}^{N}(\mathrm{n_i}-1)\left(\left.\frac{\partial F_{2MP}\left(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta}\right)}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}},\boldsymbol{x_i},\hat{\boldsymbol{\theta}}}\right)\boldsymbol{\Sigma_U}\left(\left.\frac{\partial F_{2MP}\left(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta}\right)}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u_{m\,i}},\boldsymbol{x_i},\hat{\boldsymbol{\theta}}}\right)^T}{\sum_{i=1}^{N}(n_i-1)} \quad (12)$$

In equation (11), $\bar{y}_{2MP\,i}$ is the average measured value for F_{2MP} obtained from data collected at the i^{th} target condition. As shown in Figure 1, assigning a typical 1% error in the uncertain inputs results in pseudo-replicate experiments where nearly all the variability is due to $\boldsymbol{\varepsilon_{Y_{ij}}}$ rather than $\boldsymbol{\varepsilon_{U_{ij}}}$. Conversely, when 10 % typical input uncertainty is assumed, most of the variability in the responses arises from the input uncertainties.

Table 8- The objective function and their corresponding decision variables for EVM and WLS objective functions

$$J_{EVM} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[\left(\boldsymbol{y_{m\,ij}} - \boldsymbol{g}(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta})\right)^{\mathbf{T}}\boldsymbol{\Sigma_Y^{-1}}\left(\boldsymbol{y_{m\,ij}} - \boldsymbol{g}(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta})\right) + \left(\boldsymbol{u_{m\,i}} - \boldsymbol{u_{ij}}\right)^T\boldsymbol{\Sigma_U^{-1}}\left(\boldsymbol{u_{m\,i}} - \boldsymbol{u_{ij}}\right)\right]$$

$$= \sum_{i=1}^{25}\sum_{j=1}^{n_i}\left[\frac{\left(y_{2MP\,ij}-F_{2MP}\left(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta}\right)\right)^2}{\hat{\sigma}^2_{F_{2MP}}} + \frac{\left(y_{2MP\,ij}-F_{3MP}\left(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta}\right)\right)^2}{\hat{\sigma}^2_{F_{3MP}}} + \frac{\left(y_{C_3\,ij}-F_{C_3}\left(\boldsymbol{x_i},\boldsymbol{u_{ij}},\boldsymbol{\theta}\right)\right)^2}{\hat{\sigma}^2_{F_{C_3}}} + \frac{\left(u_{mC_6\,i}-F^0_{C_6\,ij}\right)^2}{\sigma^2_{F^0_{nC_6}}} + \frac{\left(u_{mH_2\,i}-F^0_{H_2\,ij}\right)^2}{\sigma^2_{F^0_{H_2}}}\right] \quad (8.1)$$

where the decision variables are:

$\boldsymbol{u_{ij}} = \left[F^0_{C_6\,ij}, F^0_{H_2\,ij}\right]^T$ and $\boldsymbol{\theta} = \left[k_{ref\,1}, k_{ref\,2}, k_{ref\,3}, K_{ref}, E_{a\,1}, E_{a\,2}, E_{a\,3}, \Delta H^0\right]^T$

and the perfectly-known inputs are $\boldsymbol{x_i}$=$[P_i, T_i, W]^T$.

$$J_{WLS} = \sum_{i=1}^{25}\sum_{j=1}^{n_i}\left[\frac{\left(y_{2MP\,ij}-F_{2MP}(\boldsymbol{x_i},\boldsymbol{\theta})\right)^2}{\hat{\sigma}^2_{F_{2MP}}} + \frac{\left(y_{2MP\,ij}-F_{3MP}(\boldsymbol{x_i},\boldsymbol{\theta})\right)^2}{\hat{\sigma}^2_{F_{3MP}}} + \frac{\left(y_{C_3\,ij}-F_{C_3}(\boldsymbol{x_i},\boldsymbol{\theta})\right)^2}{\hat{\sigma}^2_{F_{C_3}}}\right] \quad (8.2)$$

where the decision variables are $\boldsymbol{\theta} = \left[k_{ref1}, k_{ref2}, k_{ref3}, K_{ref}, E_{a1}, E_{a2}, E_{a3}, \Delta H^0\right]^T$,

and the perfectly-known inputs are $\boldsymbol{x_i} = \left[P_i, T_i, W, F^0_{H_2\,i}, F^0_{nC_6\,i}\right]^T$.

Table 9- Values of parameter estimates obtained by minimizing objective function in Equation (34) when different levels of uncertainties in model inputs

	WLS	EVM with 1% uncertainties in model inputs	EVM with 5% uncertainties in model inputs	EVM with 10% uncertainties in model inputs
$\hat{\sigma}^2_{F_{2MP}}(10^{-13}mol^2/s^2)$	1.44	1.43	1.26	0.96
$\hat{\sigma}^2_{F_{3MP}}(10^{-14}mol^2/s^2)$	6.84	6.78	5.46	1.85
$\hat{\sigma}^2_{F_{C_3}}(10^{-14}mol^2/s^2)$	1.94	1.86	0.8	0.16
$k_{ref1}(10^{-6}mol\ s^{-1}\ g_{cat}^{-1})$	202.3	200.7	195.7	183.3
$k_{ref2}(10^{-6}mol\ s^{-1}\ g_{cat}^{-1})$	115.6	114.7	112.0	105.1
$k_{ref3}(10^{-6}mol\ s^{-1}\ g_{cat}^{-1})$	14.8	14.6	13.8	11.8
$K_{phys}(10^{-6}Pa^{-1})$	10.6	10.0	7.3	3.5
$E_{a1}(kJ\ mol^{-1})$	54.1	54.3	54.0	58.2
$E_{a2}\ (kJ\ mol^{-1})$	62.2	62.3	62.0	66.2
$E_{a3}\ (kJ\ mol^{-1})$	68.3	68.7	70.7	77.5
$\Delta H^0(kJ\ mol^{-1})$	−80.9	−83.5	−98.9	−119.8

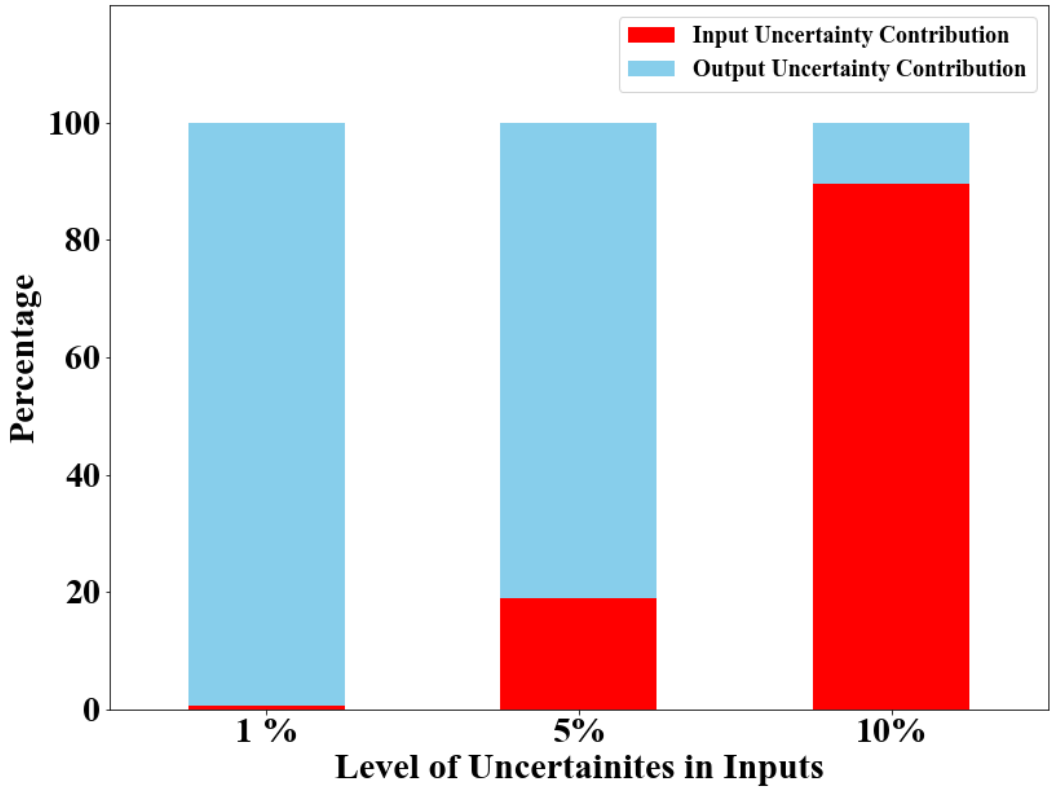

Figure 1- Contribution of input uncertainty and output uncertainty to the overall pooled estimates of variance $\hat{\sigma}^2_{F_{2MP}\ o}$ when different levels of input uncertainties are assumed during EVM parameter estimation.

4-3 Parameter and Prediction Uncertainty Quantification

Figure 2 shows boxplots for parameter uncertainties for two parameters (i.e., k_{ref1} and ΔH^0) based on results from 200 bootstrap parameter estimations using the algorithm in Table 3. Similar boxplots for the other 6 parameters are provided in the supplementary information. As shown in Figure 2, when uncertainties in the inputs are small (i.e., 1% for both $F^0_{H_2\,m}$ and $F^0_{nC_6\,m}$) the boxplots of the parameter estimates are very similar to those obtained using WLS parameter estimation, as expected. However, for higher input uncertainties (i.e., 5 % and 10 %), the confidence intervals for the parameter estimates (i.e., the distance between the ends of the whiskers) are noticeably larger than those obtained using WLS. It is interesting that the confidence interval for k_{ref1} when input uncertainties are 10 % is slightly smaller than the corresponding confidence interval when the input uncertainties are 5 %. This unanticipated trend is attributed to the decrease in output measurement variance estimates $\hat{\sigma}^2_{F_{2MP}}$, $\hat{\sigma}^2_{F_{3MP}}$ and $\hat{\sigma}^2_{F_{C_3}}$ when input uncertainties are assumed to be larger (see Table 9). Notice that the median of the bootstrap parameter estimates changes as the assumed input uncertainties increase, revealing that input uncertainties also have important effects on point estimates for the parameters. Based on the results in Figure 2, we recommend that modelers should use WLS in situations where input uncertainties are small. However, when the influence of input uncertainties becomes noticeable compared to output measurement uncertainties, EVM parameter estimation is recommended.

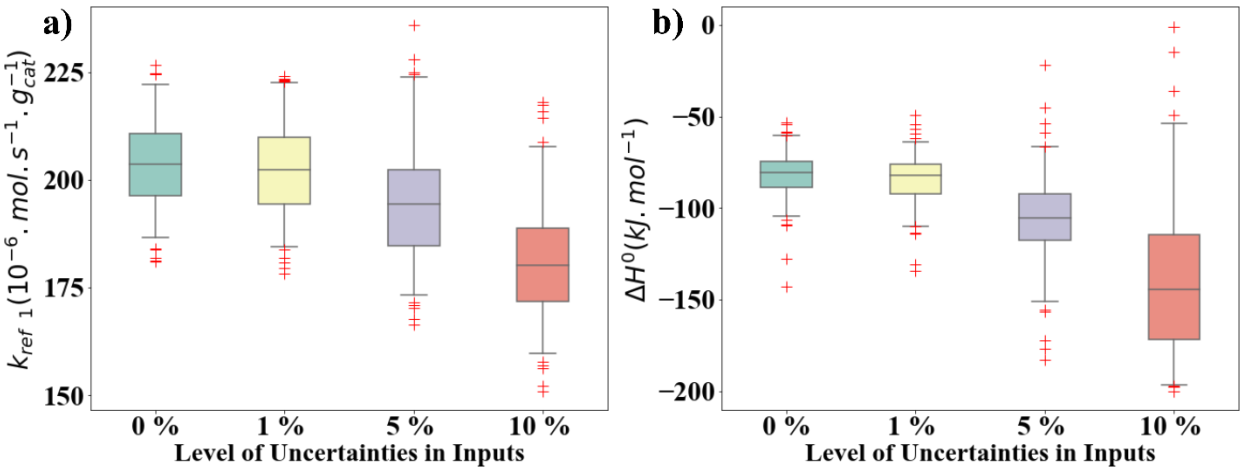

Figure 2- Parameter uncertainty boxplots for a) k_{ref1} and b) ΔH^0 constructed using 200 bootstrap parameter estimates when different levels of input uncertainties are considered for both $F^0_{H_2\,m}$ and $F^0_{nC_6\,m}$. The ends of the whiskers correspond to approximate 95 % confidence intervals.

Figure 3 shows model-prediction uncertainty boxplots for F_{2MP} and F_{C_3} based on 200 MC simulations of the 15^{th} experimental run (i.e., conducted at $T = 513\,K, P = 1\,MPa, F^0_{H_2\,m} = 1.3 \times 10^{-3}\,\frac{mol}{s}$, and $F^0_{nC_6\,m} = 2.56 \times 10^{-5}\,\frac{mol}{s}$) for different assumed levels of input uncertainties. The algorithm in Table 2 was used to obtain these results. Similar boxplots for the third output F_{3MP} are provided in the supplementary information. Figure 3 shows that, as expected, uncertainties in the mean predictions (see Figures 3a and 3c) when input uncertainties are small are similar to the corresponding uncertainties when WLS is used. When larger input uncertainties are assumed, the sizes of the confidence intervals increase. Figures 3b and 3d show boxplots for single future measurements of F_{2MP} and F_{C3} that might be obtained using the experimental condition for the 15^{th} experimental run. The cloud of 200 MC predictions for plausible single future measurements was obtained using the results of the 200 MC predictions for the mean response, with additional random measurement error added. As expected, the uncertainties in the single future measurements are larger than the uncertainties in the corresponding mean predictions. It is interesting that Figure 3d shows that the single-prediction uncertainties of F_{C3} are larger when the assumed input uncertainties are smaller. This trend occurs because the output measurement variances for F_{C_3} are larger when small input uncertainties are assumed than when larger input uncertainties are assumed (see Table 9).

It is also interesting that the trend in Figure 3b is the opposite of the trend in Figure 3d. For F_{2MP}, assuming larger input uncertainties results in increasing uncertainty about what might happen in a single new experiment. It seems that large increase in uncertainty for the mean prediction of F_{2MP} (see Figure 3a) for this output outweighs the decrease in $\hat{\sigma}^2_{F_{2MP}}$ shown in Table 9. In summary, the results in Figures 2 and 3 reveal that different assumptions about the size of input uncertainties can have important (and sometimes counter-intuitive) consequences for parameter and prediction uncertainties.

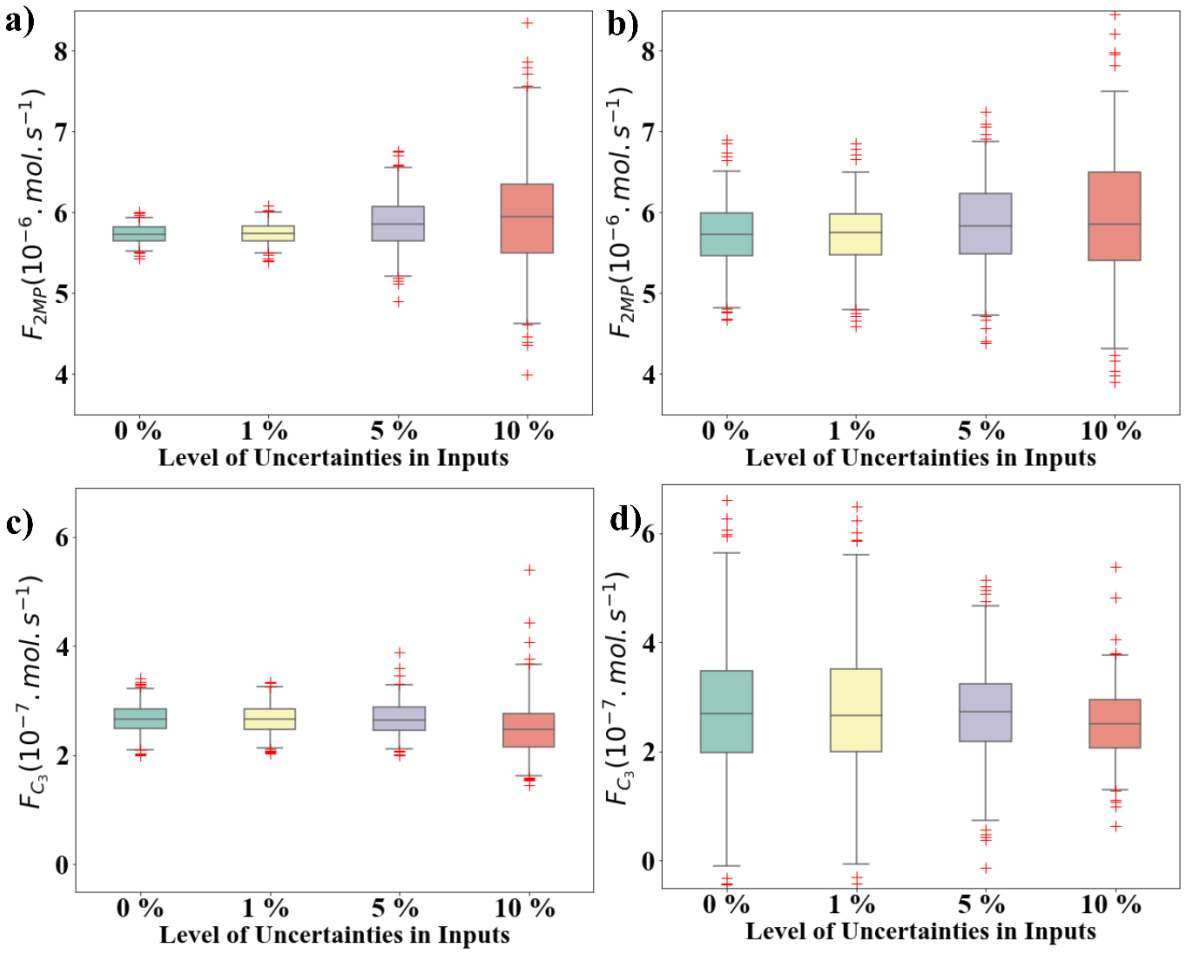

Figure 3- Prediction uncertainty boxplots for a) mean prediction of F_{2MP},b) a single-future measurement of F_{2MP}, c) mean prediction of F_{C_3} and d) a single future measurement of F_{C_3} constructed based on 200 MC simulations. Predictions correspond to the experimental condition at the 15th run in Table S.1 ($T = 513\ K, P = 1\ MPa,\ and\ , F^0_{H_2}{}_m = 1.3 \times 10^{-3}\frac{mol}{s}, and\ F^0_{nC_6}{}_m = 2.56 \times 10^{-5}\frac{mol}{s}$). The whisker ends correspond to approximate 95 % confidence intervals.

Figure 4 shows the boxplots for 200 MC simulations for the predictions of F_{2MP} and F_{C3} for a new experimental condition ($T = 580\ K, P = 1.0\ MPa,\quad F^0_{H_2\ m} = 1.3 \times 10^{-3}\frac{mol}{s}$, and $F^0_{nC_6\ m} = \quad 2.56 \times 10^{-5}\frac{mol}{s}$). Additional results for the other output F_{3MP} are provided in the supplementary information. As shown, the confidence intervals for the mean predictions and single future predictions increase when higher input uncertainties are assumed. Also, the confidence intervals for the single future measurements of F_{2MP} and F_{C_3} are noticeably larger than the corresponding confidence intervals for the mean predictions when input uncertainties are neglected or small (i.e., when it is zero or 1 %). This can be attributed to the larger estimates of the output measurement variances when the input uncertainties are small. The confidence intervals for the mean and a single future measurement prediction are similar for situations with higher input uncertainties. This can be attributed to higher values of measurement variance estimates in smaller input uncertainties (e.g., $\hat{\sigma}^2_{F_{C_3}} = 1.86\ \times 10^{-14}\ mol^2/s^2$ when there are 1% input uncertainties and $\hat{\sigma}^2_{F_{C_3}} = 0.16\ \times 10^{-14}\ mol^2/s^2$ when there are 5% input uncertainties).

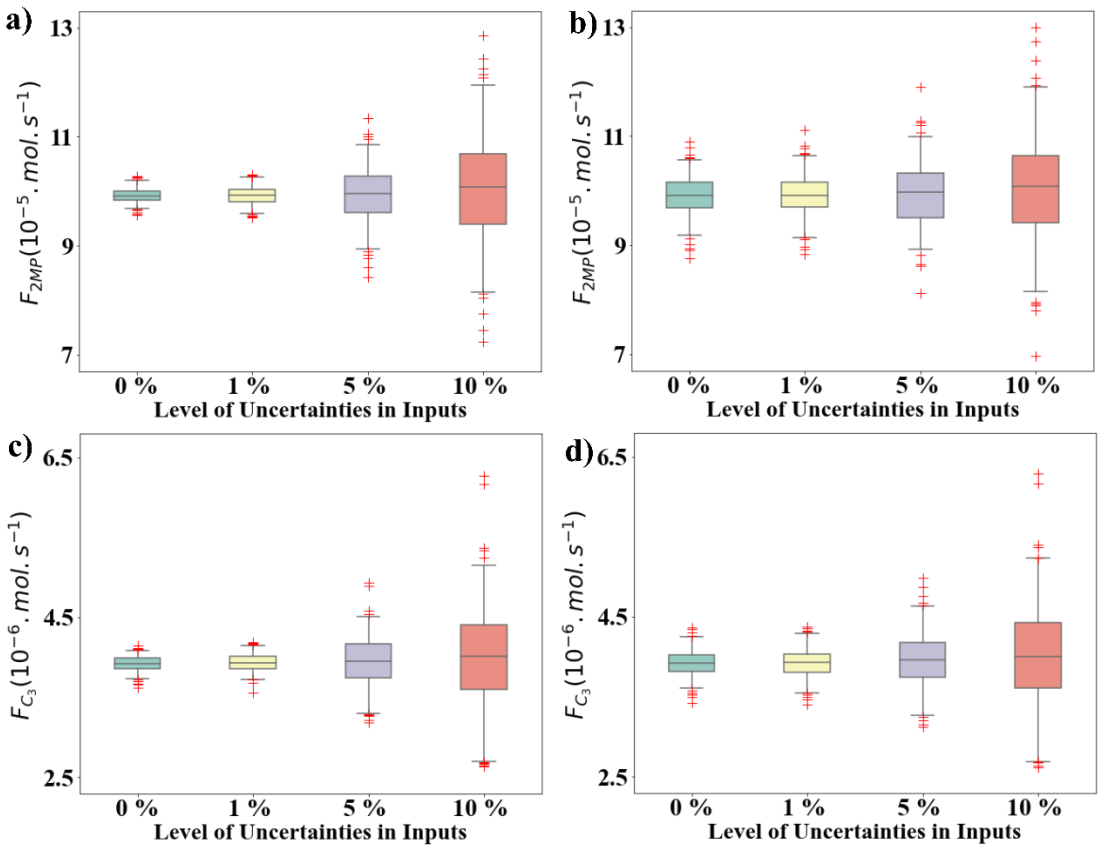

Figure 4- Prediction uncertainty boxplots for a) mean prediction of F_{2MP},b) a single future measurement of F_{2MP}, c) mean prediction of F_{C_3} and d) a single future measurement of F_{C_3} constructed based on 200 MC simulations. The model predictions are obtained for a new experimental condition ($T = 580\ K, P = 1.0\ MPa,\ F^0_{H_2\,m} = 1.3 \times 10^{-3}\,\frac{mol}{s}$, $and\ F^0_{nC_6\,m} = 2.56 \times 10^{-5}\,\frac{mol}{s}$). The whisker ends correspond to approximate 95 % confidence intervals.

4-4 Uncertainties in Predicted Yield and Selectivities

Figure 5 shows boxplots for yield predictions when different levels of input uncertainties are assumed. These results were obtained by substitution of points from the cloud of MC-based output predictions (i.e., for F_{2MP}, F_{3MP} and F_{C_3}) into equation (8). Figure 5a shows yield prediction uncertainty boxplots for the 15^{th} experimental run and Figure 5b shows yield prediction uncertainty boxplots for a new experimental condition that was not considered by Toch et al. during their experiments (i.e., $T = 580\ K, P = 1.0\ MPa$, $F^0_{H_2}{}_m = 1.3 \times 10^{-3}\ \frac{mol}{s}$, and $F^0_{nC_6}{}_m = 2.56 \times 10^{-5}\ \frac{mol}{s}$). As shown in Figure 5, the uncertainties in the yield predictions increase when a higher input uncertainty is assumed, as expected.

Figure 6a shows boxplots for the selectivity of 3MP. Additional boxplots for selectivities of other species and at different experimental conditions are provided in the supplementary information. It is interesting that in Figure 6a, when input uncertainties increase the confidence intervals for the selectivity predictions become smaller. This unanticipated result is attributed to increasing correlation between the numerator (F_{3MP}) and the denominator ($F^0_{nC_6} - F_{nC_6}$) for S_{3MP} in equation (9) when input uncertainties become larger. As shown in Figure 6b, when input uncertainties are large, the uncertainties in F_{3MP} and $F^0_{nC_6} - F_{nC_6}$ both become larger, but high correlation between them makes their ratio less uncertain than when input uncertainties are assumed to be small. These results reveal that assumptions about the size of the input uncertainties can have noticeable influence on predictions of selectivities, yields and other quantities that are calculated from model predictions and parameter values. Also, the size of prediction uncertainties can be influenced in important ways by the input uncertainty assumptions that are made by the modeler.

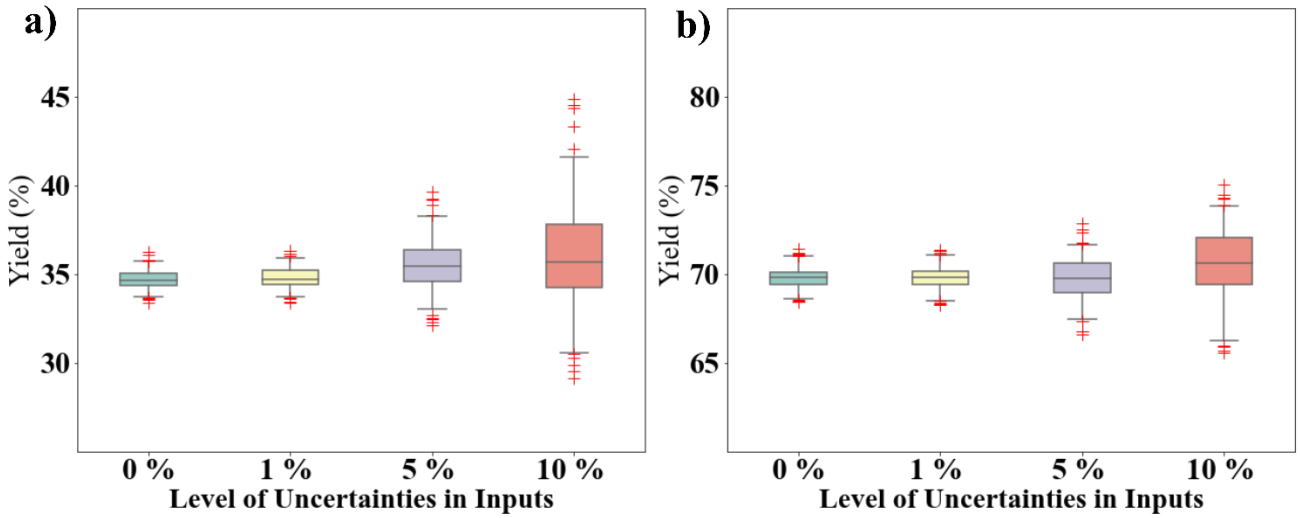

Figure 5- Boxplots for yield predictions constructed based on 200 MC simulations: a) for mean yield predictions at the target experimental condition with $T = 513\ K, P = 1\ MPa, F^0_{H_2}{}_m = 1.3 \times 10^{-3}\frac{mol}{s}$ and $F^0_{nC_6}{}_m = 2.56 \times 10^{-5}\frac{mol}{s}$; b) for mean yield predictions at a new experimental condition with higher temperature where $T = 580\ K, P = 1.0\ MPa,\ F^0_{H_2}{}_m = 1.3 \times 10^{-3}\frac{mol}{s}$ and $F^0_{nC_6}{}_m = 2.56 \times 10^{-5}\frac{mol}{s}$. The ends of the whiskers correspond to the 2.5 and 97.5 percentiles (i.e., approximate boundaries of 95% confidence intervals).

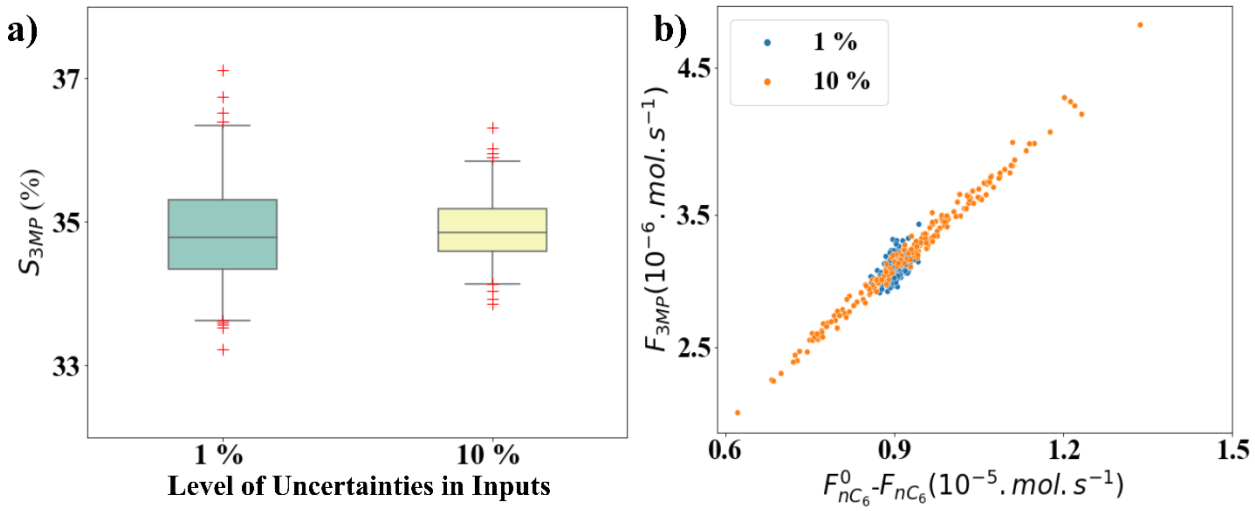

Figure 6- Selectivity predictions for 3MP constructed based on 200 MC simulations: a) boxplot for mean selectivity predictions at the experimental condition with $T = 513\ K, P = 1\ MPa, F^0_{H_2\,m} = 1.3 \times 10^{-3}\,\frac{mol}{s}$ and $F^0_{nC_6\,m} = 2.56 \times 10^{-5}\,\frac{mol}{s}$ for two levels of input uncertainties; b) the corresponding scatter plot for computed values of F_{3MP} vs. $F^0_{nC_6} - F_{nC_6}$ to investigate the correlation between the numerator and denominator in the selectivity expression in equation 9. Ends of the whiskers correspond to the approximate 95 % confidence intervals.

5- Conclusions

Conducting replicate experiments is not straightforward when inputs are uncertain, because it might not be possible to obtain the same true input values during repeated runs. Instead, experimenters can operate their equipment so that measured or target values of the inputs are the same, while acknowledging that the true values may be different. These types of attempted replicate experiments are called pseudo replicates.[21] Modelers can use pseudo-replicate data to estimate the output measurement variances required in EVM objective functions for parameter estimation, if they are careful to properly account for influence of uncertain inputs. Otherwise, their output-measurement-variance weighting factors will be too large.

In this study, we extend a methodology for estimating output measurement variances so that it can be used for multi-response models with multiple uncertain inputs. A case study involving n-hexane hydroisomerization, which has two uncertain inputs (i.e., n-hexane feed rate and hydrogen feed rate) is used to illustrate the proposed method. Kinetic and thermodynamic parameters in this hydroisomerization model are estimated using literature data.[7] Parameter estimates and model predictions are obtained using three

different levels of input uncertainty so that results can be compared with weighted-least-squares estimates, which neglect input uncertainties.

Further, we extend a bootstrap method for assessing parameter uncertainties so that it can be applied for multi-response models with several uncertain inputs. The results of the parameter estimation and parameter uncertainty calculations confirm that, when small input uncertainties are assumed (i.e., 1% of the typical flow rates), results from the EVM parameter estimation are similar to those obtained using WLS. However, when larger input uncertainties are assumed (i.e., 5 % or 10 % of the typical feed rates), the parameter estimates and their uncertainties are noticeably different compared to WLS results.

Increasing the input uncertainties resulted in several different trends for parameter confidence intervals. For example, for two of the model parameters (i.e., E_{a1} and ΔH^0) increasing input uncertainties resulted in increased width of the confidence intervals. However, for two of the parameters (i.e., k_{ref1}, k_{ref2}, k_{ref3} and K_{ref}), increasing the input uncertainties led to an initial increase in the width of the parameter confidence intervals followed by a decrease when input uncertainties increased from 5 % to 10 %. We attribute this complicated trend to the tradeoff between the influences of the larger input uncertainties and the corresponding smaller estimates for the output measurement variances.

In this study, we obtained prediction uncertainties for the *mean* response that would be obtained at several different target settings (e.g., $T = 513\ K, P = 1MPa, F^0_{H_2\ m} = 1.3 \times 10^{-3}\ \frac{mol}{s}$ and $F^0_{nC_6\ m} = 2.56 \times 10^{-5}\ \frac{mol}{s}$). We also obtained larger prediction uncertainties for a *single-future* experiment conducted at the same target. As input uncertainties increased, the confidence intervals for the predicted *mean* product outflow rates became wider for all three responses (i.e., F_{2MP}, F_{3MP} and F_{C_3}). However, confidence intervals for the corresponding *single-future* measured responses increased for two responses ($i.e., F_{2MP}$ and F_{3MP}) but decreased for the third. This result is complicated but not surprising due to the tradeoff between larger assumed input uncertainties and smaller estimated output measurement variances.

We also obtained uncertainties in the predicted yield and selectivities for 2MP and 3MP which are calculated from the predicted product outflow rates. These results showed that increasing the input uncertainties led to increasing width of the confidence interval for the yield. However, more complicated results were obtained for selectivity confidence intervals. For example, the selectivity confidence interval for 3MP became narrower when increasing input uncertainty was assumed. These results reveal that assumptions about input uncertainties can have important and sometimes complicated implications for parameter estimates, model predictions and other predicted quantities of interest to modelers and model users. We recommend that modelers should use weighted-least-squares estimation when they believe that uncertainties in experimental inputs have minimal influence on the overall reproducibility of their experiments. However, when input uncertainties are an important issue of concern, EVM parameter estimation should be used. The methods proposed in the current article should be helpful for assessing the resulting uncertainties in model parameters and model predictions.

Acknowledgements

Authors acknowledge the financial support from Natural Sciences and Engineering Research Council of Canada (NSERC) and EUROKIN. We also acknowledge professor Joris W. Thybaut for his technical recommendations for this study.

Nomenclature

Abbreviations

2MP	2-methyl pentane
3MP	3-methyl pentane
C_3	Propane
EVM	Error-in-variables model
H_2	Hydrogen
nC_6	n-Hexane
PR	Pseudo replicate
TR	True replicate
WLS	Weighted least squares

Symbols

b	Counter for the iterations of the bootstrapping algorithm
b_{max}	Maximum number of the bootstrap and Monte Carlo iterations
c	A condition of interest
e	The relative change of parameter estimates in two subsequent iterations of the algorithm used in PR situations to estimate output measurement variances
$E_{a\,r}$	Activation energy for the r^{th} reaction
F_{2MP}	Molar outflow rate of 2MP
F_{3MP}	Molar outflow rate of 3MP
F_{C_3}	Molar outflow rate of C_3
$F^0_{nC_6}$	Feed flow rate of nC_6
$F^0_{H_2}$	Feed flow rate of H_2
$F^0_{H_2\,m}$	Measured value of feed flow rate of hydrogen
$F^0_{nC_6\,m}$	Measured value of feed flow rate of n-hexane
$\boldsymbol{g}(.)$	The function returning a vector of solutions for the model predictions of a multi-output model
$g_k(.)$	The function returning solution for the prediction of the k^{th} output
J	Objective function
k_r	Kinetic rate for the r^{th} reaction
K_{ref}	Equilibrium constant at the reference temperature T_{ref}
$k_{ref\,r}$	Kinetic rate of the r^{th} reaction at the reference temperature T_{ref}
N	Number of target experimental condition
N_U	Number of uncertain model input
N_Y	Number of output predictions at each experimental condition
N_θ	Number of parameters
n_i	Number of replicate experiments at the i^{th} run condition
P	Reactor pressure
P_i	Reactor pressure at the i^{th} condition
p	Counter for the elements of parameter vector θ
s	Iteration counter in calculating the output measurement variance using pseudo-replicate data
S_q	Selectivity of the q^{th} species ($q = 2MP, 3MP$ and C_3)
T	Reactor temperature
T_i	Reactor temperature at the i^{th} experimental condition
T_{ref}	Reference temperature
u_{ij}	True value of the input for j^{th} replicate run of the i^{th} target condition
u_{il}	True value for the l^{th} uncertain input at the i^{th} experimental condition when there are no replicate data
$\boldsymbol{u_{ij}}$	Vector of true values of uncertain inputs for the j^{th} pseudo-replicate run at the i^{th} target condition
$\boldsymbol{u_{ij}^{(b)}}$	Bootstrap-generated value of the uncertain input for the j^{th} pseudo-replicate run at the i^{th} target condition
$u_{mC_6\,i}$	Measured value of F^0_{nC6} at the i^{th} experimental condition

$u_{mH_2\,i}$	Measured value of $F_{H_2}^0$ at the i^{th} experimental condition
$\boldsymbol{u_{m\,i}}$	Vector of measured values for uncertain inputs obtained from the i^{th} experimental condition
$u_{m\,il}$	Measured value for the l^{th} uncertain input at the i^{th} experimental condition
W	Catalyst weight
$\boldsymbol{x_i}$	Vector of perfectly-known model inputs at the i^{th} target condition
$y_{2MP\,ij}$	Measured value of F_{2MP} for the j^{th} pseudo-replicate run at the i^{th} target condition
$y_{3MP\,ij}$	Measured value of F_{3MP} for the j^{th} pseudo-replicate run at the i^{th} target condition
$y_{C_3\,ij}$	Measured value of F_{C_3} for the j^{th} pseudo-replicate run at the i^{th} target condition
Y_{ij}	Output measurement random variable corresponding to a single-output model for the j^{th} replicate run at the i^{th} experimental target condition
Y_{ijk}	Output measurement corresponding for the k^{th}output of a multi-output model obtained for the j^{th}replicate run at the i^{th}experimental target condition
$\boldsymbol{y_{m\,ij}}$	Vector of measured values of outputs for the j^{th} pseudo-replicate run at the i^{th} condition
$\boldsymbol{y_{m\,ij}^{(b)}}$	Vector of bootstrap-generated output measurements corresponding to $\boldsymbol{y_{mij}}$
$y_{m\,ik}$	Measured value for the k^{th} model response at the i^{th} experimental condition when there not replicates
$y_{m\,ijk}$	Output measurement value for the k^{th} model output obtained from the j^{th} pseudo-replicate run at the i^{th} condition
$\bar{y}_{m\,ik}$	The average of the k^{th} model output measurements for replicate runs conducted at the i^{th} target condition
$\bar{y}_{2MP\,i}$	The average of the measured values of F_{2MP} for replicate runs at the i^{th} target condition

Subscripts

b	Counter for the iterations of the bootstrapping algorithm
b_{max}	Maximum number of the bootstrap and Monte Carlo iterations
c	A condition of interest
i	Counter for the unique target experimental conditions
j	Counter for the repeated replicate experiments at each condition
k	Counter for the elements of a model-output vector in a multi-output model
l	Counter for the elements of a model-input vector
r	Counter for the reactions
n	Nuisance parameters
p	Primary parameters
q	Letter for denoting different species (q =2MP, 3MP, and C_3)

Greek Symbols

α_q	The number of carbon atoms in the molecules of species q (q=2MP, 3MP and C_3).
δ_k	Cut-off value for the estimates of variance for the k^{th} model response
$\delta_{F_{2MP}}$	Cut-off value for the estimates of variance for the output measurements of F_{2MP}
$\delta_{F_{3MP}}$	Cut-off value for the estimates of variance for the output measurements of F_{3MP}
$\delta_{F_{C_3}}$	Cut-off value for the estimates of variance for the output measurements of F_{C_3}
ΔH^0	Enthalpy of the physisorption
$\boldsymbol{\varepsilon_{U\,ij}}$	Vector of random measurement noises for the input measurements
$\boldsymbol{\varepsilon_{Y\,ij}}$	Vector of output random measurement noise for the j^{th} replicated run at the i^{th} experimental condition
$\boldsymbol{\theta}$	Vector of parameters
$\widehat{\boldsymbol{\theta}}$	Parameter estimates
$\widehat{\boldsymbol{\theta}}^{(b)}$	Vector of estimated values of parameters from the b^{th} bootstrap-generated data
$\hat{\sigma}^2_{F_{2MP}}$	Output measurement variance for F_{2MP}
$\hat{\sigma}^2_{F_{3MP}}$	Output measurement variance for F_{3MP}
$\hat{\sigma}^2_{F_{C_3}}$	Output measurement variance for F_{C_3}
$\sigma^2_{F^0_{nC_6}}$	Measurement variance for uncertain input $F^0_{nC_6}$
$\sigma^2_{F^0_{H_2}}$	Measurement variance for uncertain input $F^0_{H_2}$
$\hat{\sigma}^2_{F_{2MP}\,o}$	Overall estimate of measurement variance for F_{2MP} accounting for input measurement noises $\boldsymbol{\varepsilon_{U\,ij}}$ and output measurement noises $\boldsymbol{\varepsilon_{Y\,ij}}$
$\sigma^2_{U_l}$	Measurement variance for the l^{th} uncertain input
$\sigma^2_{Y_k}$	Output measurement variance for the k^{th} model output
$\hat{\sigma}^2_{Y_k}$	Estimate of variance for the k^{th} model output measurements
$\hat{\sigma}^2_{Y_{ik}}$	Variance estimate for the k^{th} model output obtained from measurements of replicate data corresponding to the i^{th} experimental data
$\boldsymbol{\Sigma_U}$	Covariance matrix for the uncertain inputs
$\boldsymbol{\Sigma_Y}$	Covariance matrix for the model outputs
$\widehat{\boldsymbol{\Sigma}}_Y^{(b)}$	Estimated value of output measurements covariance matrix estimated from the b^{th} bootstrap-generated data

References

1. Mathew M, Duever T. Reactivity Ratio Estimation in Non-Linear Polymerization Models using Markov Chain Monte Carlo Techniques and an Error-In-Variables Framework. *Macromolecular Theory and Simulations.* 2015;24(6):566-579.

2. Keeler SE, Reilly PM. The error-in-variables model applied to parameter estimation when the error covariance matrix is unknown. *The canadian journal of chemical engineering.* 1991;69(1):27-34.

3. Britt H, Luecke R. The estimation of parameters in nonlinear, implicit models. *Technometrics.* 1973;15(2):233-247.

4. Abdi K, McAuley KB. Estimation of output measurement variances for error-in-variables model parameter estimation. *AIChE Journal.* 2022:e17735.

5. Vamos RJ, Haas CN. Reduction of ion-exchange equilibria data using an error in variables approach. *AIChE journal.* 1994;40(3):556-569.

6. Baradie B, Shoichet MS. Synthesis of fluorocarbon− vinyl acetate copolymers in supercritical carbon dioxide: insight into bulk properties. *Macromolecules.* 2002;35(9):3569-3575.

7. Toch K, Thybaut JW, Marin GB. A systematic methodology for kinetic modeling of chemical reactions applied to n-hexane hydroisomerization. *AIChE Journal.* 2015;61(3):880-892.

8. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine.* 2000;19(9):1141-1164.

9. Abdi K, Celse B, McAuley KB. Propagating Input Uncertainties into Parameter Uncertainties and Model Prediction Uncertainties- A Review. Submitted to *Industrial and Engineering Chemistry Research.* 2022.

10. Tørvi H, Hertzberg T. Methods for evaluating uncertainties in dynamic simulation—A comparison of performance. *Computers & chemical engineering.* 1998;22:S985-S988.

11. Fischer U, Hungerbühler K. Multiobjective batch process design aiming at robust performances. *Chemical engineering science.* 2007;62(21):6015-6031.

12. Chen Y, Hoo KA. Uncertainty propagation for effective reduced-order model generation. *Computers & Chemical Engineering.* 2010;34(10):1597-1605.

13. Morales-Rodriguez R, Meyer AS, Gernaey KV, Sin G. A framework for model-based optimization of bioprocesses under uncertainty: Lignocellulosic ethanol production case. *Computers & Chemical Engineering.* 2012;42:115-129.

14. Xu Z, Lai C, Marcy PW, et al. Predicting the performance uncertainty of a 1-MW pilot-scale carbon capture system after hierarchical laboratory-scale calibration and validation. *Powder technology.* 2017;312:58-66.

15. Behrooz HA, Hoseini M. Application of the unscented transform in the uncertainty propagation of thermodynamic model parameters. *Fluid Phase Equilibria.* 2018;475:64-76.

16. Morgan JC, Soares Chinen A, Omell B, et al. Development of a rigorous modeling framework for solvent-based CO2 capture. Part 2: Steady-state validation and uncertainty quantification with pilot plant data. *Industrial & Engineering Chemistry Research.* 2018;57(31):10464-10481.

17. Neba FA, Tornyeviadzi M, Asiedu NY, et al. Can the operating limits of biogas plants operated under non-isothermal conditions be defined with certainty? Modeling self-optimizing attainable regions. *Computers & Chemical Engineering.* 2020;141:107001.

18. Fenila F, Shastri Y. Stochastic optimization of enzymatic hydrolysis of lignocellulosic biomass. *Computers & Chemical Engineering.* 2020;135:106776.

19. Kimaev G, Chaffart D, Ricardez-Sandoval LA. Multilevel Monte Carlo applied for uncertainty quantification in stochastic multiscale systems. *AIChE Journal.* 2020;66(8):e16262.

20. Son J, Du Y. Comparison of intrusive and nonintrusive polynomial chaos expansion-based approaches for high dimensional parametric uncertainty quantification and propagation. *Computers & Chemical Engineering.* 2020;134:106685.
21. Abdi K, McAuley KB. Estimation of Output Measurement Variances for EVM Parameter Estimation. *AIChE Journal.* 2022:e17735.