Estimation of Output Measurement Variances for EVM Parameter Estimation

Kaveh Abdi and Kimberley B. McAuley*

Department of Chemical Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada

ABSTRACT

Error-in-variables model (EVM) methods require information about input and output measurement variances when estimating model parameters. In EVM, using replicate experiments for estimating output measurement variances is complicated because true values of inputs may be different when multiple attempts are made to repeat an experiment. To address this issue, we categorize attempted replicate experiments as: i) true replicates (TRs) when uncertain inputs are the same in replicated runs and ii) pseudo replicates (PRs) when *measured* inputs are the same, but unknown true values of inputs are different. We propose methodologies to obtain output measurement variance estimates and associated parameter estimates for both situations. We also propose bootstrap methods for obtaining joint-confidence information for the resulting parameter estimates. A copolymerization case study is used to illustrate the proposed techniques. We show that different assumptions noticeably affect the uncertainties in the resulting reactivity-ratio estimates.

Keywords: Error-in-variables model (EVM), Replicate experiments, Reactivity-ratio estimation, Copolymerization, Mayo-Lewis equation

INTRODUCTION

Fundamental mathematical models are widely used for chemical process development and improvement. These models usually contain unknown parameters that require estimation.[1,2] In conventional parameter-estimation methodologies such as weighted least-squares (WLS) estimation, the model inputs are assumed to be perfectly known (i.e., measured without error), while random measurement errors are considered for the model outputs.[3,4] Nevertheless, there are important situations where model inputs are not perfectly known, but contain significant uncertainties. The Error-in-Variables Model (EVM) technique was developed to account for measurement uncertainties in both inputs and outputs during parameter estimation.[5,6] In EVM, the true values of the measured inputs are estimated along with the model parameters. EVM[7-30] has been used in a variety of chemical engineering modeling studies (see **Table 1** and **Table 2**).

The following single-response model can be used to illustrate assumptions that are made during an EVM parameter estimation wherein n_i replicated experiments have been attempted at N different target operating conditions (for $i = 1, \dots, N$):

$$Y_{ij} = g\left(u_{ij}, \boldsymbol{x_i}, \boldsymbol{\theta}\right) + \varepsilon_{Y,ij} \qquad (1)$$

$$U_{ij} = u_{ij} + \varepsilon_{U,ij} \qquad (2)$$

In Equation (1), Y_{ij} is the measured value of the response for the j^{th} ($j = 1,\dots, n_i$) run conducted using the i^{th} target experimental condition, g is a solution of nonlinear model equation(s), u_{ij} is the true value of the uncertain input for the j^{th} run at the i^{th} target condition, $\boldsymbol{x_i} \in \boldsymbol{R^{N_x}}$ is a vector containing any perfectly-known model inputs for the i^{th} experimental condition, $\boldsymbol{\theta} \in \boldsymbol{R^{N_\theta}}$ is the vector of model parameters, and $\varepsilon_{Y,ij}$ is the random measurement noise for the response $g\left(u_{ij}, \boldsymbol{x_i}, \boldsymbol{\theta}\right)$. In Equation (2), U_{ij} is the measured value of the uncertain input and $\varepsilon_{U,ij}$ is the

corresponding random measurement noise. We assume that $\varepsilon_{Y,ij}$ and $\varepsilon_{U,ij}$ in Equations (1) and (2) are independent random variables with $\varepsilon_{Y,ij} \sim N(0, \sigma_Y^2)$ and $\varepsilon_{U,ij} \sim N(0, \sigma_U^2)$.

Based on the model in Equations (1) and (2), a maximum-likelihood approach can be used to obtain the following objective function for use in EVM parameter estimation:[31]

$$J_{EVM} = \sum_{i=1}^{N} \sum_{j=1}^{r_i} \left[\frac{\left(y_{m,ij} - \boldsymbol{g}(\boldsymbol{x_i}, u_{ij}, \boldsymbol{\theta})\right)^2}{\sigma_Y^2} + \frac{\left(u_{\boldsymbol{m,ij}} - u_{\boldsymbol{ij}}\right)^2}{\sigma_U^2} \right] \quad (3)$$

where $y_{m,ij}$is the measured value of output from the j^{th} run conducted using the i^{th} target run condition, and $u_{m,ij}$ is the measured value of the corresponding uncertain input. The objective function in Equation (3) is minimized to simultaneously determine the parameter estimates $\widehat{\boldsymbol{\theta}}$ and estimates $\hat{u}_{ij}$ for the true values of the uncertain inputs. As shown in Equation (3), variances of the measured model inputs σ_U^2 and outputs σ_Y^2 are weighting factors in the EVM objective function and therefore influence the parameter estimates. Also, confidence intervals for the parameters depend on these variances.[31]

In conventional WLS parameter-estimation studies, where inputs are perfectly known, measurement variances are often estimated from replicate experiments.[32-34] The influence of all uncertainties and disturbances associated with setup of the equipment, measurement and mixing of reactants, conduct of the experiments, collection of any samples for analysis, and measurement of the desired responses is lumped together and treated as part of the output uncertainty.[34] However, when some of the inputs are acknowledged to contain important uncertainties, so that EVM parameter estimation is required, the issue of replicate experiments and output variance estimation is more complex. In some situations, the experimentalist may believe that *true replicate* experiments have been conducted (i.e., all experiments conducted at the i^{th} target run condition have the same unknown value of u_{ij}). For example, the modeler will have true replicates when

repeated experiments are conducted using the same batch of input material that has uncertain properties. In other situations, it may not be appropriate to assume that u_{ij} is the same for all runs conducted at the i^{th} target condition (e.g., when attempted replicate experiments use different batches of input material with uncertain compositions). In the current article, we refer to these runs with the same target settings, but different values of u_{ij} as *pseudo-replicate* runs. When a modeler is confronted with *pseudo-replicate* experiments, the issue of estimating σ_Y^2 from data becomes complicated because the measured values $y_{m,ij}$ are influenced by random errors in the inputs as well as the corresponding output.

Table 1 and Table 2 summarize previous EVM studies in the chemical engineering literature where data from experiments with uncertain inputs were used for parameter estimation.[4,7-12,14,16,17,19,22,24,25,27,30,35-55] Studies in Table 1 involve reactivity-ratio estimation in copolymerization and terpolymerization studies. The authors used EVM for these studies because important model inputs (i.e., initial feed compositions or comonomer concentrations in the reactor) have significant uncertainties compared to the measurement uncertainty in the resulting copolymer or terpolymer composition. Other studies involving different types of chemical engineering models are summarized in Table 2. These studies are related to vapour-liquid-equilibrium, gas-phase catalytic hydrogenation, gas-solid adsorption, oxygen consumption in wastewater, water gas-shift reaction, ion-exchange equilibrium, gas purification, mass-transfer in liquid mixtures, a low-density polyethylene reactor, and a natural gas network. The uncertain inputs in these systems include measured inlet temperatures, inlet flowrates, inlet mole fractions, partial pressures, and liquid-phase concentrations.

In most EVM parameter-estimation studies, variances of the uncertain model inputs are estimated from data or are assumed to be known based on previous studies.[2,7,9,22,30,39,47-50,54] Users of EVM

can usually obtain reliable values of σ_U^2 for use in their objective functions, but it is much more difficult to obtain a reliable estimate of σ_Y^2. Repeating measurements of the same output sample only captures the portion of the variability that is attributable to noise in the analytical device. Other types of variability that should be included in σ_Y^2 (e.g., due to equipment set up and experimental procedures) can only be detected when full experiments are repeated. This difficult issue has not received adequate consideration in the EVM literature.

As shown in Table 1 and Table 2, values of σ_Y^2 were reported in 29 of the 39 studies, while the remaining authors were silent about the weighting factors used for parameter estimation. Authors from 23 of the 39 studies assumed that σ_Y^2 was known *a priori* (indicated using AK in the 4^{th} columns), while 14 studies were silent about how σ_Y^2 was obtained (indicated using ? in the 4th columns). In three of the 39 studies, the authors estimated σ_Y^2 from replicate data (indicated using R in the 4^{th} columns). Replicate experimental results were reported in 16 out of the 39 EVM studies. In one of these 16 studies, it is clear that true replicate experiments were conducted so that σ_Y^2 was relatively easy to estimate.[47] In this study, Keeler and Reilly purchased several standard ethylene/propylene copolymer samples and used them to fit a calibration curve between IR (the model input) and NMR (the model output) measurements. Repeated measurements of the NMR response were obtained for each standard sample, which has a constant composition and therefore a fixed true value of the unknown (noise-free) IR response. In another of the 16 studies, it is clear that Scott and Penlidis obtained pseudo-replicate data.[44] In this terpolymerization study, they made multiple solutions of monomers with the same target concentrations and used them to repeat some experiments. As such, their reported input measurements might have different true values, even though the corresponding reported input values are the same.[44] In the remaining 14 studies with replicates, the authors do not provide sufficient details about how the repeated runs

were conducted, so it is difficult to ascertain whether true replicates or pseudo replicates were performed (indicated using R? in the 5^{th} columns).[10,12,14,16,17,38] Based on the articles summarized in Table 1 and Table 2, it is apparent that users of EVM are not attempting to obtain reliable estimates of σ_Y^2 from data. Note that EVM parameter estimation has been used by engineers and scientists in a variety of fields beyond chemical engineering.[(e.g.,56-62)] To our knowledge, reliable methods for estimating σ_Y^2 from data have not been reported in these other EVM studies.

Researchers who perform EVM parameter estimation are interested in quantifying the uncertainties in their parameter estimates. As shown in Table 1, all research groups who estimated reactivity ratios obtained joint confidence regions (JCRs) and/or confidence intervals (CIs) for their parameters. Only three of the 28 studies in Table 1 provide details about how their parameter uncertainty information was obtained.[7,9,39] All three studies relied on the assumption that output measurement variances were perfectly known.

The objectives of the current study are: i) to propose approaches for estimating σ_Y^2 from true-replicate and pseudo-replicate data, ii) to use a copolymerization case study to illustrate how σ_Y^2 can be reliably estimated in situations involving true replicates and pseudo replicates, and iii) to estimate joint confidence (JC) information for the estimated parameters in the copolymerization model. The remainder of the article is organized as follows. First, we describe three different situations involving replicate experiments in EVM and provide information about suitable EVM objective functions. Next, methods for calculating $\hat{\sigma}_Y^2$ in EVM situations involving different types of replicate data are explained. Finally, copolymerization of n-butyl methacrylate and n-butyl acrylate is used as a case study to show how σ_Y^2 and $\boldsymbol{\theta}$ should be estimated in EVM situations involving true replicates and pseudo replicates. We also propose a bootstrapping method to obtain JC information for reactivity-ratio estimates, based on the estimated value for $\hat{\sigma}_Y^2$.

Table 1- Summary of Studies where EVM used for Reactivity-ratio Estimation

Authors	Monomers	Variance estimates reported	How variances obtained	Type of replicate experiments	Parameter uncertainty estimates
Dube et al. [14,35]	Styrene/butyl acrylate	Yes	?	R?	JCR
Brar et al. [4]	Styrene/butyl acrylate	Yes	AK	None	JCR and CIs
Brar and Charan[36]	Vinyl acetate/methyl acrylate	No	?	None	CIs
Brar and Charan[37]	Vinyl acetate/ethyl methacrylate	Yes	AK	None	CIs
Dube and Penlidis[38]; Scott and Penlidis[39]	Butyl acrylate/methyl methacrylate, butyl acrylate/ vinyl acetate, and methyl methacrylate/vinyl acetate	Yes	AK	R?	JCR and CIs
Schoonbrood et al. [25]	Styrene/2-Hydroxyethyl methacrylate	No	?	None	JCR and CIs
Brar et al. [40]	Acrylonitrile/methacrylic acid	Yes	?	None	JCR and CIs
Brar and Malhorta[41]	Vinylidene chloride/methyl acrylate	Yes	?	None	CIs
McManus and Penlidis [42]; Scott and Penlidis[39]	Styrene/ethyl acrylate	Yes	AK	R?	JCR
Suddaby et al. [16]	Methyl methacrylate/n-butyl methacrylate	Yes	AK	R?	JCR
Brar et al. [27]	Acrylonitrile/glycidyl methacrylate	No	?	None	JCR and CIs
Hakim et al. [43]	Butyl acrylate/methyl methacrylate	No	?	R?	JCR
Baradie et al. [17]	Tetrafluoroethylene/ vinyl acetate, chlorotrifluoroethylene/ vinyl acetate, and vinylidene fluoride/ vinyl acetate	Yes	R	R?	JCR and CIs
Kazemi et al.[8]	Acrylonitile/methyl acrylate	Yes	AK	None	JCR
Zhang and Dube[11]	N-Butyl Methacrylate/d-limonene	No	?	R?	JCR
Kazemi et al. [12]	N9-(4-vinylbenzyl)-9Hcarbazole/methyl methacrylate	Yes	AK	R?	JCR
Mathew and Duever[7]	Di-n-butyl itaconate (DBI)/methyl methacrylate	Yes	AK	None	JCR
Ren et al.[2]	N-Butyl acrylate/n-Butyl methacrylate	Yes	AK	R?	JCR
Scott and Penlidis[39]	2-methylene-1,3-dioxepane/vinyl acetate	Yes	AK	None	JCR
Scott and Penlidis[44]	2-acrylamido-2-methylpropane sulfonic acid/acrylamide /acrylic acid	No	?	Pseudo replicate	JCR
Gabriel and Dube[45]	N-Butyl Acrylate/2-Ethylhexyl Acrylate/Methyl Methacrylate	No	?	R?	JCR
Yousefi et al.[10]	Triisopropylsilyl Acrylate/ Methyl Methacrylate/ Butyl Acrylate	No	?	R?	JCR
Krieger et al. [46]	Itaconic Acid/N-Vinyl-2-Pyrrolidone	No	?	None	JCR
Santos et al.[9]	Styrene/VeoVa-10 copolymerization	Yes	AK	None	JCR and CIs

Table 2- Summary of Studies Involving other Chemical Engineering Models where EVM used for Parameter Estimation

Authors	Model of study	Variance estimates are reported	How variances were determined	Type of replicate experiments reported
Keeler and Reilly[47]	Copolymer composition calibration between IR and NMR measurements	No	R	True replicate
Sutton and MacGregor[30]	Vapour-liquid equilibrium for ethanol-isooctane and benzene- heptane systems	Yes	AK	None
Duever et al.[24]	Vapor-liquid equilirium	Yes	AK	None
Rod and Hancil[48]	Gas-phase hydrogenation of phenol on a palladium catalyst	Yes	AK	None
High and Danner [22]	Gas-solid adsorption	Yes	AK	None
Valko and Vajda[49]	Biological oxygen consumption and chemical oxygen consumption relationship in a wastewater	Yes	AK	None
Kim et al.[50]	Vapour-liquid equilibrium for methanol and 1,2-dichloroethane	Yes	AK	None
Bardow and Marquardt [51]	Liquid-liquid diffusion	Yes	AK	None
Kim et al.[52]	Water-gas shift reaction	Yes	AK	None
Vamos and Haas[53]	Binary ion-exchange equilibrium	Yes	R	R?
Faber and Wozny[54]	Coke-oven-gas purification process	Yes	AK	None
Zavala and Biegler[19]	Low-density polyethylene tubular reactor	No	?	None
Leung et al.[55]	Gas lateral network of TransCanada Pipeline system	Yes	AK	None

PROPOSED CLASSIFICATION OF REPLICATE EXPERIMENTS IN EVM

Table 3 and **Table 4** provide information about model equations and objective functions corresponding to three different sets of assumptions about replicate (or nearly replicated) experiments. The first two classifications, TR1 and TR2, correspond to "true replicate" experiments. By true replicates, we mean that, when several experiments are conducted using the same target settings, the *true values of the uncertain inputs can be assumed the same* for the replicated runs. The third classification corresponds to "pseudo-replicate" experiments, denoted by PR. By pseudo-replicate runs, we mean that the *true values for the uncertain inputs are different* when several experiments are conducted using the same target conditions. Two different

situations are considered for true replicates because sometimes modelers will have only one measured value for an uncertain input used in the repeated runs (situation TR1) and sometimes modelers will have new measured values for uncertain inputs in each replicated run, even though the modeler assumes that the underlying true values have not changed.

Equation (3.1) in Table 3 is the model equation used to describe the measured outputs for situations TR1 and TR2. Here $\boldsymbol{Y_{ij}} \in R^{N_Y}$ is the vector of measured responses for the j^{th} replicate experiment conducted at the i^{th} target run condition, where N is the number of distinct run conditions in the data set (i.e., $i = 1, \ldots, N$) and n_i is the number of repeated runs performed at the i^{th} target run condition (i.e., $j = 1,\ldots, n_i$). In Equation (3.1), $\boldsymbol{g} \in R^{N_Y}$ is a vector of solutions of nonlinear equations, $\boldsymbol{u}_i \in R^{N_U}$ is the vector of unknown true values for the uncertain model inputs for the i^{th} target condition and $\boldsymbol{\varepsilon_{Yij}} \in R^{N_Y}$ is a vector of random noise for the corresponding vector of output measurements $\boldsymbol{Y}_{ij}$. Equation (3.2) in

is the model equation used to describe the measured values of the uncertain inputs in situation TR1. $\boldsymbol{U_i}$ is the vector of uncertain input measurements for the i^{th} target experimental condition and $\boldsymbol{\varepsilon_{Ui}} \in R^{N_U}$ is the corresponding vector of random measurement noise. In theTR1 situation, only one measurement is obtained for each uncertain input. As a result, the vector of input measurement noise $\boldsymbol{\varepsilon_{Ui}}$ in Equation (3.1) has a single index i. However, the symbol for the vector of output measurement random noise $\boldsymbol{\varepsilon_{Yij}}$ has two indices, because n_i measurements are obtained for the model outputs at each target condition.

Similarly, Equation (3.3) is used to describe measured values of the uncertain inputs for situation TR2. Here $\boldsymbol{U_{ij}}$ is the j^{th} vector of uncertain input measurements for the i^{th} target experimental

condition, and $\boldsymbol{\varepsilon_{Uij}}$ is the corresponding vector of measurement noise. The extra subscript *j* compared with Equation (3.2) appears because, in situation TR2, each repeated run has its own set of measurements for the unknown inputs. Equations (3.4) and (3.5) are the model equations for the measured outputs and measured uncertain inputs for the situation where pseudo-replicate runs have been conducted. Notice that Equation (3.4) is the same as Equation (3.1) except that the vector of true values for the uncertain inputs $\boldsymbol{u_{ij}}$ has an extra subscript, *j*. Here, the modeler assumes that the true inputs are different for each attempted replicate experiment. As a result, Equation (3.5) also contains the vector $\boldsymbol{u_{ij}}$.

Objective functions for parameter estimation using the three different assumptions about replicates are provided in Table 4. The objective function in Equation (4.1) is minimized to simultaneously to obtain the vector of parameter estimates $\widehat{\boldsymbol{\theta}}$ and the estimates of uncertain inputs in vector $\widehat{\boldsymbol{u}}_{\boldsymbol{i}}$ for each target run condition. In objective function (4.1), $\boldsymbol{y_{m,ij}}$ is the vector of output measurements from the j^{th} run conducted using the i^{th} target condition, and $\boldsymbol{u_{m,i}}$ is the vector of uncertain input measurements for the i^{th} target run. Note that the lower-case letters y and u in $\boldsymbol{y_{m,ij}}$ and $\boldsymbol{u_{m,i}}$, respectively, are used to show that these quantities are samples obtained from random variables $\boldsymbol{Y_{ij}}$ and $\boldsymbol{U_i}$. The total number of data values used for parameter estimation in situation TR1 is $N_Y \sum_{i=1}^{N} n_i + NN_U$ as shown in the 3rd column of Table 3. Similarly, Equation (4.2) is the objective function for situation TR2, where $\boldsymbol{u_{m,ij}}$ is the vector of uncertain input measurements for the j^{th} run using the i^{th} experimental target. The total number of data values in situation TR2 is larger than in TR1 because more uncertain input measurements are available. Equation (4.3) is used when true inputs are assumed to be different whenever replicate experiments are attempted. As a result, more unknown input values (in vectors $\boldsymbol{u_{ij}}$) require estimation than in

situations TR1 and TR2. As shown in Table 3, the PR situation requires estimation of $N_U \sum_{i=1}^{N} r_i$ unknown input values, while situations TR1 and TR2 require estimation of NN_U unknown inputs.

Table 3- Error-in-Variables Model Equations for Different Assumptions about Replicate Runs

Case	EVM Equations	Number of Data Values	Number of parameters and inputs requiring estimation
TR1	$\boldsymbol{Y_{ij} = g(x_i, u_i, \theta) + \varepsilon_{Y,ij}}$ **(3.1)** $\boldsymbol{U_i = u_i + \varepsilon_{U,i}}$ **(3.2)**	$N_Y \sum_{i=1}^{N} n_i + NN_U$	$N_\theta + NN_U$
TR2	$\boldsymbol{Y_{ij} = g(x_i, u_i, \theta) + \varepsilon_{Y,i}}$ **(3.1)** $\boldsymbol{U_{ij} = u_i + \varepsilon_{U,ij}}$ **(3.3)**	$\sum_{i=1}^{N} n_i (N_Y + N_U)$	$N_\theta + NN_U$
PR	$\boldsymbol{Y_{ij} = g(x_i, u_{ij}, \theta) + \varepsilon_{Y,ij}}$ **(3.4)** $\boldsymbol{U_i = u_{ij} + \varepsilon_{U,ij}}$ **(3.5)**	$\sum_{i=1}^{N} n_i (N_Y + N_U)$	$N_\theta + N_U \sum_{i=1}^{N} r_i$

Table 4- Objective Functions for Parameter Estimation in Error-in-Variables Models with Different Assumptions about Replicate Runs

Case	EVM objective function
TR1	$J_{TR1} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \boldsymbol{\left(y_{m,ij} - g(x_i, u_i, \theta)\right)^T \Sigma_Y^{-1} \left(y_{m,ij} - g(x_i, u_i, \theta)\right)} + \sum_{i=1}^{N} \boldsymbol{(u_{mi} - u_i)^T \Sigma_U^{-1} (u_{mi} - u_i)}$ (4.1)
TR2	$J_{TR2} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \boldsymbol{\left[\left(y_{m,ij} - g(x_i, u_i, \theta)\right)^T \Sigma_Y^{-1} \left(y_{m,ij} - g(x_i, u_i, \theta)\right) + \left(u_{mij} - u_i\right)^T \Sigma_U^{-1} \left(u_{mij} - u_i\right)\right]}$ (4.2)
PR	$J_{PR} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \boldsymbol{\left[\left(y_{m,ij} - g(x_i, u_{ij}, \theta)\right)^T \Sigma_Y^{-1} \left(y_{m,ij} - g(x_i, u_{ij}, \theta)\right) + \left(u_{mi} - u_{ij}\right)^T \Sigma_U^{-1} \left(u_{mi} - u_{ij}\right)\right]}$ (4.3)

PROPOSED OUTPUT MEASUREMENT-VARIANCE ESTIMATION IN SITUATIONS INVOLVING REPLICATE EXPERIMENTS

In this section, methods for estimation of output measurement variances are proposed for situations involving data values from replicate or pseudo-replicate runs. In situations TR1 and TR2, where true replicate experiments have been performed, the methods for calculation of variance estimates are similar. The estimated variance in the measurements for the k^{th} model output obtained using the i^{th} run condition is:

$$\hat{\sigma}^2_{Y_{ik}} = \frac{\sum_{j=1}^{n_i}(y_{m,ijk}-\bar{y}_{m,ik})}{r_i-1} \qquad (i = 1, \dots, N \text{ and } k = 1, \dots, N_Y) \quad (4)$$

where $y_{m,ijk}$ is a measured value obtained from the j^{th} replicate run and $\bar{y}_{m,ik}$ is the corresponding average for the measurements obtained from the n_i repeated runs. For situations TR1 and TR2, measurement variances for each model output can be estimated by pooling variances obtained from different run conditions:

$$\hat{\sigma}^2_{Y_k} = \frac{\sum_{i=1}^{N}(n_i-1)\hat{\sigma}^2_{Y_{ik}}}{\sum_{i=1}^{N}(n_i-1)} \quad (k = 1, \dots, N_Y) \qquad (5)$$

When pseudo replicates are conducted, variance estimates obtained from Equations (4) and (5) could be too large, because uncertainties associated with the different true inputs may inflate the variability of the measured outputs at each target condition.

We propose a linearization-based approach for the PR situation so that more-accurate estimates of $\sigma^2_{Y_k}$ can be computed. Using a Taylor-series expansion and Equation (3.4) in Table 3, the k^{th} element of measurement vector $\boldsymbol{Y_{ij}}$ can be estimated:

$$Y_{ijk} \approx g_k(\boldsymbol{x_i}, \boldsymbol{u_{m,i}}, \boldsymbol{\theta}) + \frac{\partial g_k}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{m,i}}, \boldsymbol{x_i}, \boldsymbol{\theta}} (\boldsymbol{u_{ij}} - \boldsymbol{u_{m,i}}) + \varepsilon_{Y_{ijk}} \qquad (6)$$

by linearizing around measured values $\boldsymbol{u_{m,i}}$ from a PR run. In Equation (6), the Jacobian $\frac{\partial g_k}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{m,i}},\boldsymbol{x_i},\boldsymbol{\theta}}$ is a row vector with N_U columns, which contains derivatives of model predictions for the k^{th} response variable with respect to each of the uncertain inputs and $\varepsilon_{Y_{ijk}}$ is the random noise for the measurements of the k^{th} output obtained at the j^{th} replicate of the i^{th} target condition. Taking variances of both sides of Equation (6) gives:

$$\sigma^2_{PR,ik} \approx \left(\frac{\partial g_k}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{m,i}},\boldsymbol{x_i},\boldsymbol{\theta}}\right)\boldsymbol{\Sigma_U}\left(\frac{\partial \boldsymbol{g_k}}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{m,i}},\boldsymbol{x_i},\boldsymbol{\theta}}\right)^T + \sigma^2_{Y_{ik}} \qquad (7)$$

assuming that uncertainties in measured inputs and outputs are independent. In Equation (7), $\sigma^2_{PR,ik}$ is the variance of the k^{th} model output due to variability in both inputs and outputs, whereas $\sigma^2_{Y_{ik}}$ is the measurement variance for the k^{th} response variable.

In PR situations, $\hat{\sigma}^2_{Y_{ik}}$ in Equation (4) provides an estimate of the overall output variance $\sigma^2_{PR,ik}$. Substituting this expression for $\sigma^2_{PR,ik}$ in Equation (7), replacing $\boldsymbol{\theta}$ with $\boldsymbol{\hat{\theta}}$, and rearranging gives:

$$\hat{\sigma}^2_{Y_{ik}} = \frac{\sum_{j=1}^{\boldsymbol{n_i}}(y_{m,ijk}-\bar{y}_{m,ik})}{n_i-1} - \left(\frac{\partial g_k}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{m,i}},\boldsymbol{x_i},\boldsymbol{\hat{\theta}}}\right)\boldsymbol{\Sigma_U}\left(\frac{\partial \boldsymbol{g_k}}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{m,i}},\boldsymbol{x_i},\boldsymbol{\hat{\theta}}}\right)^T \qquad (8)$$

In our experience, occasionally $\hat{\sigma}^2_{Y_{ik}}$ computed using Equation (8) can be a negative number, especially if it is computed using a small number of pseudo-replicate runs, and uncertainties in the model inputs are relatively large.´ As a result, we recommend a cut-off value δ_k be used, based on the modeler's knowledge about a reasonable size for the output measurement variance:

$$\hat{\sigma}^2_{Y_{ik}} = \max\left(\frac{\sum_{j=1}^{\boldsymbol{n_i}}(y_{m,ijk}-\bar{y}_{m,ik})}{n_i-1} - \left(\frac{\partial g_k}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{mi}},\boldsymbol{x_i},\boldsymbol{\hat{\theta}}}\right)\boldsymbol{\Sigma_U}\left(\frac{\partial \boldsymbol{g_k}}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u_{mi}},\boldsymbol{x_i},\boldsymbol{\hat{\theta}}}\right)^T, \delta_k\right) \quad (k = 1, \ldots, N_Y) \qquad (9)$$

Equation (5) can then be used to pool these output-measurement variance estimates, resulting in an appropriate value of $\hat{\sigma}^2_{Y_k}$ for use in EVM parameter estimation.

Because parameter estimates $\widehat{\boldsymbol{\theta}}$ are required to calculate the Jacobian matrix $\frac{\partial g_k}{\partial \boldsymbol{u}}$ in Equation (9), a procedure that updates $\left.\frac{\partial g_k}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u}_{m,i},\boldsymbol{x}_i,\widehat{\boldsymbol{\theta}}}$ based on the most recent estimate $\widehat{\boldsymbol{\theta}}$ is recommended. **Table 5** provides suggested steps for this iterative procedure. In Steps 2 and 3, initial estimates for the output variances and parameters are obtained by neglecting the influence of the uncertain inputs. Steps 4 to 6 involve iterations that improve these estimates by taking the input uncertainties into account.

Table 5- Procedure for EVM Parameter Estimation Using Pseudo-Replicate Data

1- Assign appropriate cut-off values δ_k $(k = 1, \dots, N_Y)$ for each of the measured outputs variances. Set the step counter to *s*=0.
2- Obtain an initial guess for the output measurement variances using Equations (4) and (5).
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 (9) using the most-recent parameter estimates $\widehat{\boldsymbol{\theta}}^{(s)}$.
5- Use Equation (5) to pool the variance estimates obtained in step 4.
6- Use J_{PR} in Table 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 $e = \sqrt{\sum_{\boldsymbol{p=1}}^{\boldsymbol{N_\theta}} \left(\frac{\hat{\theta}_p^{(s+1)} - \hat{\theta}_p^{(s)}}{\hat{\boldsymbol{\theta}}_{\boldsymbol{p}}^{(s)}} \right)^{\boldsymbol{2}}}$ 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.

CASE STUDY: EVM REACTIVITY-RATIO ESTIMATION USING MAYO-LEWIS EQUATION

The Mayo-Lewis equation describes the relationship between instantaneous comonomer composition in the reaction medium f_1 and the corresponding copolymer composition F_1:

$$F_1 = \frac{r_1 f_1^2 + f_1(1-f_1)}{r_1 f_1^2 + 2 f_1(1-f_1) + r_2(1-f_1)^2} \qquad (10)$$

Many researchers have used the Mayo-Lewis equation to estimate reactivity ratios from low-conversion batch-reactor data (i.e., at conversions below 10%), where the initial monomer composition and cumulative copolymer composition are assumed to be similar to the instantaneous compositions f_1 and F_1 in Equation (10), respectively.[35,63] Reactivity ratios r_1 and r_2 are used to account for the relative rates of the four propagation reactions shown in **Table 6**, where $R_1^\bullet$ and $R_2^\bullet$ are growing polymer chains with terminal monomers M_1 and M_2, respectively:

$$r_1 = \frac{k_{11}}{k_{12}} \qquad (11)$$

$$r_2 = \frac{k_{22}}{k_{21}} \qquad (12)$$

Table 6- Important reactions corresponding to a free-radical copolymerization terminal model

$R_1^\bullet + M_1 \overset{k_{11}}{\longrightarrow} R_1^\bullet$	(6.1)
$R_1^\bullet + M_2 \overset{k_{12}}{\longrightarrow} R_2^\bullet$	(6.2)
$R_2^\bullet + M_1 \overset{k_{21}}{\longrightarrow} R_1^\bullet$	(6.3)
$R_2^\bullet + M_2 \overset{k_{22}}{\longrightarrow} R_2^\bullet$	(6.4)

In the current case study, copolymerization data provided by Ren et al., with n-butyl methacrylate as component 1 and n-butyl acrylate as component 2, are used to estimate reactivity ratios r_1 and r_2 in situations involving different assumptions about the replicate runs.[2] Data values obtained by Ren et al. are provided in **Table 7**.[2] Their data set contains replicate data at two initial monomer feed compositions (i.e., $f_1 = 0.487$ and $f_1 = 0.196$, respectively). The information provided by Ren et al. about how they conducted their experiments is not sufficiently detailed so that we can ascertain whether their replicated experiments are true replicates or pseudo replicates. Nevertheless, because the reported measured inputs at each replicate condition are the same, either

a TR1 or PR situation is plausible. A TR2 situation would usually lead to different reported values of f_1 when replicate experiments are attempted.

Table 7- The experimental data collected for copolymerization of n-Butyl Methacrylate and n-Butyl Acrylate from Ren et al.[2]

f_1	F_1
0.100	0.187
0.200	0.335
0.300	0.459
0.410	0.620
0.501	0.668
0.601	0.762
0.700	0.820
0.801	0.882
0.897	0.968
0.487	0.656
0.487	0.654
0.487	0.651
0.487	0.655
0.196	0.334
0.196	0.348
0.196	0.344
0.196	0.353

Based on the results in Table 7, Ren et al. might have prepared a mixture of comonomers and used it to run four copolymerization experiments at the corresponding replicate condition where $f_1 = 0.487$, resulting in situation TR1. Alternatively, they might have run four copolymerization

experiments at that target replicate condition using four different monomer mixtures, after making some adjustments to ensure that the corresponding measured values of f_1 are the same, resulting in a PR situation. When estimating reactivity ratios from their data, Ren et al. assumed that the input and output measurement variances were both already known (AK). They did not explicitly estimate σ_y^2 using their replicate experiments.

Below we demonstrate how to obtain output measurement variance estimates from the replicates provided by Ren et al. in two different ways: *i*) assuming that replicate data in Table 7. are true replicates of type 1, and *ii*) assuming that pseudo replicates were performed. In our analysis, we make the common assumption (used by Ren et al.) that random errors in the measured comonomer feed composition and copolymer composition are multiplicative rather than additive (i.e., the anticipated percentage error is constant, so that larger measured values tend to have larger errors).[12,63] Writing the model equation in terms of natural logarithms of the measured inputs and responses results in additive random-error terms with constant variances. Model Equations (8.1) and (8.2) in **Table 8** are used for situations where the output measurement variance is assumed known (indicated by AK in the first column of Table 8). In this situation, we do not need to pay attention to whether replicates have been conducted or not, and each pair of (f_1, F_1) values can be treated as if it were obtained at a distinct run condition. In Equations (8.1) and (8.2) the index i, where $i = 1 \ldots 17$, is the index for the 17 experimental runs listed in Table 7. On the left-hand side of Equation (8.1), we use a capital *L* in the symbol $LnF_{1,i}$ to indicate that the natural logarithm of the i^{th} measured output is a random variable due to the random measurement error $\varepsilon_{LnF_1,i}$. On the right-hand side of Equation (8.1), lower-case symbols r_1, r_2 and $f_{1,i}$ are used to indicate unknown true values of the reactivity ratios and the input comonomer compositions that require estimation.

Similarly, in Equation (8.2), $Lnf_{1,i}$ is a random variable, $\ln f_{1,i}$ on the right-hand side is the corresponding true value, and $\varepsilon_{Lnf_1,i}$ is the random measurement error.

The third row of Table 8 shows the EVM equations for the situation where TR1 replicates are conducted. In this situation, there are 11 unique experimental conditions (i.e., $i = 1, \dots, 11$). Data from replicated runs are shown for the last two target conditions (i.e., $n_{10} = n_{11} = 4$). In equation (8.3), $LnF_{1,ij}$ is the output for the j^{th} replicate at the i^{th} target condition. Equation (8.2) shows the relationship between input measurements and their corresponding true values in the TR1 situation. The difference between Equation (8.2) when it is used for the TR1 and AK situations is that, in the TR1 situation, the index i has the values from 1 to 11, whereas it has the values from 1 to 17 in the AK situation. Equations (8.4) and (8.5) are used for the PR situation.

Table 8- EVM Mayo-Lewis for the AK, TR1 and PR situations

Case	EVM equation	Counter values
AK	$\mathrm{Ln}F_{1,i} = \ln\left(\frac{r_1 f_{1,i}^2 + f_{1,i}(1-f_{1,i})}{r_1 f_{1,i}^2 + 2f_{1,i}(1-f_{1,i}) + r_2(1-f_{1,i})^2}\right) + \varepsilon_{\mathrm{Ln}F_1,i}$ (8.1) $Lnf_{1,i} = \ln f_{1,i} + \varepsilon_{Lnf_1,i}$ (8.2)	$i = 1, \dots, 17$
TR1	$\mathrm{Ln}F_{1,ij} = \ln\left(\frac{r_1 f_{1,i}^2 + f_{1,i}(1-f_{1,i})}{r_1 f_{1,i}^2 + 2f_{1,i}(1-f_{1,i}) + r_2(1-f_{1,i})^2}\right) + \varepsilon_{\mathrm{Ln}F_1,ij}$ (8.3) $Lnf_{1,i} = \ln f_{1,i} + \varepsilon_{Lnf_1,i}$ (8.2)	$i = 1, \dots, 11$, $j = 1, \dots, n_i$ $n_i = \begin{cases} 1 & if\ 1 \le i \le 9 \\ 4 & if\ 10 \le i \le 11 \end{cases}$
PR	$\mathrm{Ln}F_{1,ij} = \ln\left(\frac{r_1 f_{1,ij}^2 + f_{1,ij}(1-f_{1,ij})}{r_1 f_{1,ij}^2 + 2f_{1,ij}(1-f_{1,ij}) + r_2(1-f_{1,ij})^2}\right) + \varepsilon_{\mathrm{Ln}F_1,ij}$ (8.4) $Lnf_{1,i} = \ln f_{1,ij} + \varepsilon_{Lnf_1,ij}$ (8.5)	$i = 1, \dots, 11$, $j = 1, \dots, n_i$ $n_i = \begin{cases} 1 & if\ 1 \le i \le 9 \\ 4 & if\ 10 \le i \le 11 \end{cases}$

The corresponding objective functions for the models in Table 8 are provided in **Table 9**. Equation (9.1) is the objective function when the output variance is assumed known, where $F_{1m,i}$ and $f_{1m,i}$ are measurements for copolymer composition and comonomer composition, respectively. A lower-case letter l is used in $lnF_{1m,i}$ and $lnf_{1m,i}$, because they are numerical values obtained from an experiment rather than random variables. Equations (9.2) and (9.3) are the corresponding

objective functions for the TR1 and PR situations. In all of the objective functions, we assume a known variance for the inputs $\sigma^2_{Lnf_1} = 1 \times 10^{-4}$, which was reported by Ren et al.[2] In the AK situation, we also use their assumed output measurement variance $\sigma^2_{lnF_1} = 2.5 \times 10^{-3}$. For the TR1 and PR situations, we estimate the output measurement variances using the proposed methodology, obtaining the results shown in **Table 10**. The estimate of $\hat{\sigma}^2_{LnF_1}$ for the PR situation is 47% smaller than the estimate for the TR1 situation. This result makes sense because, in the PR situation, some of the variation in the measured outputs is attributed to uncertainties in $lnf_{1,m}$. However, in the TR1 situation, all of the variation in the measured outputs for the true replicate runs is assigned to output measurement noise (and to common-cause variation not associated with f_1).

Table 9- EVM-based Objective Functions for Reactivity-ratio Estimation Corresponding to the AK, TR1 and PR Situations

Case	Objective function
AK	$J_{AK} = \sum_{i=1}^{17}\left[\frac{\left(ln\,F_{1m,i} - ln\,F_{1,i}\right)^2}{\sigma^2_{LnF_1}} + \frac{\left(ln\,f_{1m,i} - ln\,f_{1,i}\right)^2}{\sigma^2_{\mathrm{Ln}\,f_1}}\right]$ (9.1)
TR1	$J_{TR1} = \sum_{i=1}^{11}\sum_{j=1}^{n_i}\frac{\left(ln\,F_{1m,ij} - ln\,F_{1,i}\right)^2}{\hat{\sigma}^2_{LnF_1}} + \sum_{i=1}^{11}\frac{\left(ln\,f_{1m,i} - ln\,f_{1,i}\right)^2}{\sigma^2_{\mathrm{Ln}\,f_1}}$ (9.2) where $n_i = \begin{cases} 1 & if\ 1 \le i \le 9 \\ 4 & if\ 10 \le i \le 11 \end{cases}$
PR	$J_{PR} = \sum_{i=1}^{11}\sum_{j=1}^{n_i}\frac{\left(lnF_{1m,ij} - ln\,F_{1,ij}\right)^2}{\hat{\sigma}^2_{lnF_1}} + \sum_{i=1}^{11}\sum_{j=1}^{n_i}\frac{\left(ln\,f_{1m,i} - ln\,f_{1,ij}\right)^2}{\sigma^2_{lnf_1}}$ (9.3) where $n_i = \begin{cases} 1 & if\ 1 \le i \le 9 \\ 4 & if\ 10 \le i \le 11 \end{cases}$

Estimated values of the parameters obtained using three different assumptions about uncertainties are shown in columns 3 and 4 of Table 10. The estimated values of the reactivity ratios for the different situations are similar. However, uncertainties in these parameter estimates are noticeably

different. Figure 1a shows a linearization-based 95% JCR (the red ellipse) corresponding to the AK situation. Detailed information about how this JCR was computed is provided in Appendix A. We obtained the red ellipse by extending a methodology developed by Hamilton[64] who obtained JCRs for important subsets of model parameters in nonlinear least-squares regression. While constructing these JCRs, Hamilton accounted for uncertainties in nuisance parameters that were not part of the important subsets. In our proposed method, described in Appendix A, the uncertain inputs f_1 are treated as nuisance parameters and r_1 and r_2 are important parameters. The JCR shown in Figure 1a is constructed using a chi-squared distribution with 2 degrees of freedom (corresponding to the 2 important parameters being estimated).

When output measurement variances are not assumed known, it is appropriate to use an F distribution for computing the corresponding JCR. However, it is not obvious how a modeler should determine the denominator degrees of freedom for this F distribution (corresponding to the degrees of freedom for the estimated parameter covariance matrix). As a result, we opted to use an empirical bootstrapping technique to obtain information about parameter uncertainties. Information about the proposed bootstrapping algorithm is provided in Appendix B. In this method, we generate bootstrap composition measurements and use them to determine plausible reactivity-ratio estimates, which are shown by green dots in Figures 1a), 1b), and 1c). We validated the proposed bootstrapping method by comparing the results obtained for the AK situation with the red ellipse. Bootstrapping results are in good agreement with this analytical result, because only 4 out of 100 bootstrap estimates falls outside of the linearization-based 95% JCR. Using 10000 random seeds for bootstrapping resulted in 533 plausible parameter estimates falling outside of the red ellipse. We attribute the small deviation between 533/10000 and the theoretical value of 5.0% to nonlinearity of the model in Equation (8.1).

Results obtained using the proposed bootstrap methods for the TR1 and PR situations are shown by the green dots in Figures 1b) and Figure 1c), respectively, based on the corresponding estimates of $\hat{\sigma}^2_{LnF_1}$ shown in Table 10. The resulting reactivity-ratio uncertainties are noticeably smaller than for the AK situation (the red ellipse for the AK situation is repeated in Figures 1b) and 1c) for comparison). It makes sense that cloud of green dots obtained using the PR assumption is smaller than the corresponding cloud for the TR1 assumption, because the variance estimate obtained using the PR assumption is smaller. These results reveal that experimentalists should be careful when they make assumptions about uncertainties and repeated experiments in EVM parameter estimation. These assumptions can have an important influence on the resulting parameter uncertainties.

Table 10- Estimated (assumed) Values of Measurement Variance $\sigma^2_{LnF_1}$ and Estimated Reactivity Ratios Corresponding to EVM Situations with Different Underlying Assumptions

Case	Estimated (Assumed) output variance	r_1	r_2
AK	2.5×10^{-3}	2.0121	0.4615
TR1	8.1×10^{-5}	2.0556	0.4749
PR	4.3×10^{-5}	2.0081	0.4612

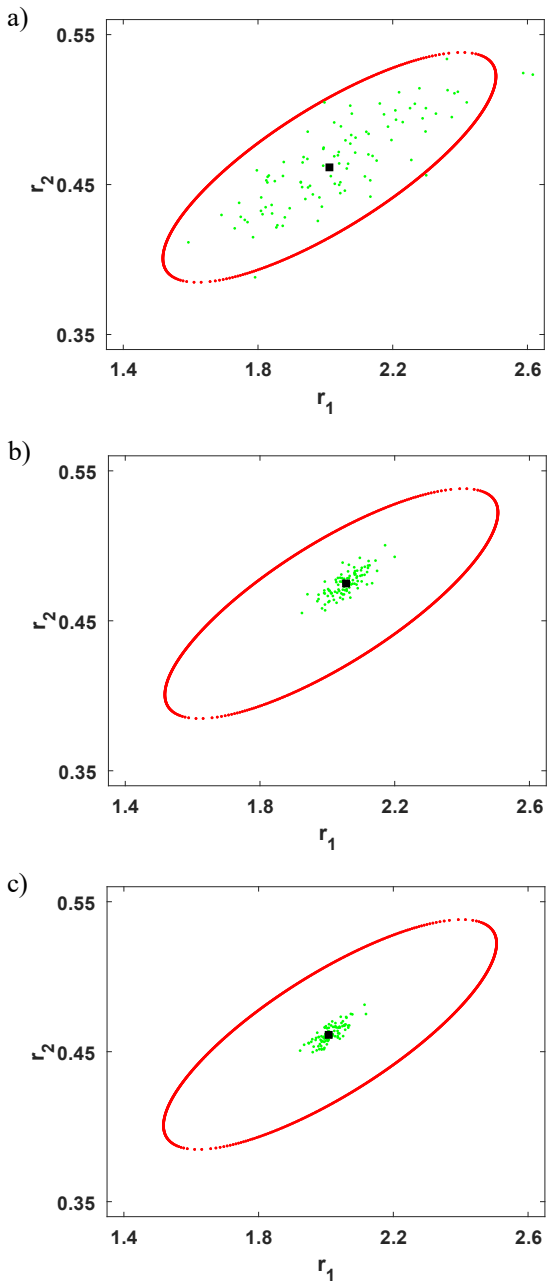

Figure 1. JC information for reactivity ratios obtained using data from Table 7 and different uncertainty assumptions. Different reactivity-ratio estimate, indicated by ▪ are obtained for situations where a) the output measurement variance $\sigma^2_{LnF_1} = 2.5 \times 10^{-3}$ is assumed known, b) the output measurement variance $\hat{\sigma}^2_{LnF_1}$ is estimated from true replicate runs and c) the output measurement variance $\hat{\sigma}^2_{LnF_1}$ is estimated assuming pseudo replicate runs. The red ellipse shown in all three plots is the linearization-based JCR obtained by assuming that $\sigma^2_{LnF_1}$ is perfectly known.

This ellipse is repeated in b) and c) as a helpful reference for comparison purposes. Green dots in a), b) and c) are plausible alternative reactivity-ratio estimates obtained using bootstrapping.

CONCLUSIONS

Conducting replicate experiments is complicated in error-in-variables (EVM) situations. Depending on how the experiments are performed, the true values of the uncertain inputs may or may not be the same for multiple runs conducted at the same target conditions. We call replicate experiments "true replicates", when true values of uncertain inputs are assumed to be the same for all runs conducted at a target condition. When measured values of uncertain inputs are the same for repeated runs, but the underlying true inputs may be different, we call these experiments pseudo replicates.

Obtaining accurate estimates of output measurement variances is important in EVM parameter estimation because these variance estimates appear in the EVM objective function, thereby influencing the parameter estimates. It is relatively straightforward to obtain output measurement variance estimates from true replicates, but situations involving pseudo replicates are more complex. In the current study, a linearization-based approach is proposed to estimate output measurement variances in pseudo-replicate (PR) situations.

We use a Mayo-Lewis case study involving literature data[2] for n-butyl methacrylate and n-butyl acrylate copolymerization to illustrate the proposed methodology. Reactivity-ratio parameters in the Mayo-Lewis equation are estimated using a data set containing replicate runs. We estimate the reactivity ratios making three different assumptions: i) output measurement variance is assumed known, ii) output measurement variance is unknown and replicate data are true replicates, and iii) output measurement variance is unknown and replicate data are pseudo replicates. We show similar point estimates for the reactivity ratios are obtained in all three situations.

To obtain information about parameter uncertainties, we extend a linearization-based method to obtain JCRs in situations where the output measurement variance is assumed known and we show that the parameter results are consistent with those from a proposed bootstrapping method. We develop variations for this bootstrapping technique to provide parameter uncertainty information for true-replicate and pseudo-replicate situations. The results of these bootstrapping calculations reveal that parameter uncertainty estimates are influenced to a great extent by the assumed type of replicates. In the current case study, uncertainties in the parameters are larger when true replicates are assumed than when pseudo replicates are assumed, because of the noticeable difference (by a factor of ∼2) in the corresponding estimates for the output measurement variance.

ACKNOWLEDGEMENT

Authors acknowledge the financial support from Natural Sciences and Engineering Research Council of Canada (NSERC).

CONFLICT OF INTEREST

The authors declare no conflict of interest

NOTATION USED IN MANUSCRIPT AND SUPPLEMENTARY INFORMATION

Abbreviations

AK = Assumed known

CI = Confidence interval

EVM = Error-in-variables model

JC= Joint confidence

JCR = Joint confidence region

PR = Pseudo replicate

TR1 = True replicate of type 1

TR2 = True replicate of type 2

WLS = Weighted least squares

Symbols

b= Counter for the iterations of the bootstrapping algorithm

b_{max}= Maximum number of the bootstrapping algorithm

e = The relative change of parameter estimates in two subsequent iterations of the algorithm used in PR situations to estimate output measurement variances

f_1 = Composition for comonomer 1

$f_{1,i}$ = True value for the composition of the comonomer of type 1 at the i^{th} target condition

$f_{1,ij}$ =The j^{th} true value for the composition of comonomer type 1 at the i^{th} target condition

$f_{1m,i}$= The measured value of for the composition of the comonomer type 1 obtained at the i^{th} target experimental condition

F_1 = The copolymer composition for component 1

$F_{1,ij}$= The j^{th} repeated value of the copolymer composition of type 1 obtained at the i^{th} target condition

$F_{1m,ij}$= The j^{th} measurement for the copolymer composition of type 1 at the i^{th} target experimental condition

$\boldsymbol{g}$= A vector of solutions for the model predictions of a multi-output model

g= A solution for the model prediction of a single-output model

$\boldsymbol{g}_{\boldsymbol{aug,s}}$ = Scaled augmented model prediction vector containing scaled output and scaled input predictions

g_k = Solution for the prediction of the k^{th} output

I_N= An identity matrix of size N

J_{TR1}= EVM-based objective function corresponding to the experiments involving true replicates of type 1

J_{TR2} = EVM-based objective function corresponding to the experiments involving true replicates of type 2

J_{PR}= EVM-based objective function corresponding to the data set obtained from a pseudo-replicate situation

$\hat{\boldsymbol{J}}_{\boldsymbol{\theta_p}}$= The Jacobian matrix of the model predictions with respect to the primary parameters

k_{11}= Rate coefficient for the reaction between the growing polymer of type 1 and the comonomer of type 1

k_{22}= Rate coefficient for the reaction between the growing polymer of type 2 and the comonomer of type 2

k_{12}= Rate coefficient for the reaction between the growing polymer of type 1 and the comonomer of type 2

k_{21}= Rate coefficient for the reaction between the growing polymer of type 2 and the comonomer of type 1

M_1 = Monomer of type 1

M_2= Monomer of type 2

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 = Counter for the elements of parameter vector θ

$\widehat{\boldsymbol{Q}}$ = The weighting matrix for computing the covariance matrix of the primary parameters

$R_1^{\bullet}$= Macroradical with terminal monomer of type 1

$R_2^{\bullet}$= Macroradical with terminal monomer of type 2

U_{ij}= Input measurements random variable for a model with a single uncertain input for the j^{th} replicate run at the i^{th} experimental condition

$\boldsymbol{U_{ij}}$= Vector of uncertain input measurements random variable for the j^{th} replicate at the i^{th} experimental condition

r_1= Reactivity ratio indicating the relative tendency of macroradical type 1 to react with monomers type 1 and type 2

r_2= Reactivity ratio indicating the relative tendency of macroradical type 2 to react with monomers type 1 and type 2

s= Iteration counter in calculating the output measurement variance using pseudo-replicate data

u_{ij} = True value of the input for j^{th} replicate run of the i^{th} target condition

$\boldsymbol{u_{ij}}$= Vector of true values of uncertain inputs for the j^{th} replicate run of i^{th} target condition

$\boldsymbol{u_{mi}}$= Vector of measured values for uncertain inputs obtained from the i^{th} experimental condition

$\boldsymbol{x_i}$= Vector of perfectly-known model inputs for 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} replicate run at the i^{th} condition

$y_{m,ijk}$= Output measurement value for the k^{th} model output obtained from the j^{th} 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

Subscripts

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

n= Nuisance parameters

p = Primary parameters

Greek symbols

δ_k= Cut-off value for the estimates of variance for the k^{th} output measurement

ε = Random measurement noise for a single-output model with perfectly known inputs

$\varepsilon_{lnf_1,i}$= Random measurement noise at the i^{th} experimental condition for the natural logarithm of the measurements for the comonomer type 1 composition

$\varepsilon_{lnf_1,ij}$= Random measurement noise for the j^{th} pseudo-replicated run at the i^{th} experimental condition for the natural logarithm of the measurements for the comonomer type 1 composition

$\varepsilon_{lnF_1,i}$= Random measurement noise for the natural logarithm of the measurements of the copolymer composition at the i^{th} experimental condition

$\varepsilon_{lnF_1,ij}$= Random measurement noise for the natural logarithm of the j^{th} replicated measurement of the copolymer composition at the i^{th} experimental condition

$\boldsymbol{\varepsilon_{U,ij}}$= Vector of random measurement noises for the input measurements

$\varepsilon_{U,ij}$= Random noise for the input measurements of a model with a single uncertain input

$\boldsymbol{\varepsilon_{Y,ij}}$ = Vector of output random measurement noise for the j^{th} replicated run at the i^{th} experimental condition

$\varepsilon_{Y,ij}$= Random measurement noise for the j^{th} output measurement of a single-output model at the i^{th} experimental condition

$\varepsilon_{Y_{ijk}}$= Random noise for the k^{th} model output measurement obtained for the j^{th} replicated run at the i^{th} target condition

$\boldsymbol{\theta}$ = Vector of true values of parameters

σ_Y^2= Output measurement random noise for a single-output model

$\hat{\sigma}_{Y_k}^2$= Estimate of variance for the k^{th} model output

$\hat{\sigma}_{Y_{ik}}^2$= Variance estimate for the k^{th} model output obtained from measurements of replicate data corresponding to the i^{th} experimental data

$\sigma_{lnf_1}^2$= Variance for the natural logarithm of the comonomer type 1 composition measurements

$\sigma_{lnF_1}^2$= Variance for the natural logarithm of the copolymer composition measurements

$\sigma_{PR,ik}^2$= Variance of the k^{th} output measurements at the i^{th} target condition arising from output measurement noises and input uncertainties

$\widehat{\boldsymbol{\Sigma}}_{(\hat{r}_1,\hat{r}_2)}$= Covariance matrix for the reactivity ratio value estimates

$\boldsymbol{\Sigma_U}$= Covariance matrix for the uncertain input

$\boldsymbol{\Sigma_Y}$=Covariance matrix for the model outputs

$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta_p}}$= Covariance matrix estimate for the primary parameters

REFRENCES

1. Maria G. A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chemical and Biochemical Engineering Quarterly.* 2004;18(3):195-222.

2. Ren S, Hinojosa-Castellanos L, Zhang L, Dubé MA. Bulk Free-Radical Copolymerization of n-Butyl Acrylate and n-Butyl Methacrylate: Reactivity Ratio Estimation. *Macromolecular Reaction Engineering.* 2017;11(3):1600050.

3. Montgomery DC, Runger GC, Hubele NF. *Engineering statistics*: John Wiley & Sons; 2009.

4. Brar A, Satyanarayana C. Microstructure determination of styrene-butyl acrylate copolymers by NMR spectroscopy. *Polymer journal.* 1992;24(9):879-887.

5. Madansky A. The fitting of straight lines when both variables are subject to error. *Journal of the american statistical association.* 1959;54(285):173-205.

6. York D. Least-squares fitting of a straight line. *Canadian Journal of Physics.* 1966;44(5):1079-1086.

7. 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.

8. Kazemi N, Duever TA, Penlidis A. A powerful estimation scheme with the error-in-variables-model for nonlinear cases: Reactivity ratio estimation examples. *Computers & Chemical Engineering.* 2013/01/10/ 2013;48:200-208.

9. Santos MP, Torraga MGF, Barbosa NB, Tavares DT, Giudici R. Estimation of Reactivity Ratios in the Copolymerization of Styrene and VeoVa-10. *Macromol React Eng.* 2020;14(3):2000001.

10. Yousefi FK, Jannesari A, Pazokifard S, Saeb MR, Scott AJ, Penlidis A. Terpolymerization of triisopropylsilyl acrylate, methyl methacrylate, and butyl acrylate: Reactivity ratio estimation. *Macromol React Eng.* 2019;13(4):1900014.

11. Zhang Y, Dube MA. Copolymerization of n-Butyl Methacrylate and d-Limonene. *Macromolecular Reaction Engineering.* 2014;8(12):805-812.

12. Kazemi N, Lessard BH, Marić M, Duever TA, Penlidis A. Reactivity ratio estimation in radical copolymerization: from preliminary estimates to optimal design of experiments. *Ind Eng Chem Res.* 2014;53(18):7305-7312.

13. Kazemi N, Duever TA, Penlidis A. Demystifying the estimation of reactivity ratios for terpolymerization systems. *AIChE Journal.* 2014;60(5):1752-1766.

14. Dubé MA, Penlidis A, O'driscoll KF. A kinetic investigation of styrene/butyl acrylate copolymerization. *The Canadian Journal of Chemical Engineering.* 1990;68(6):974-987.

15. Chee K, Ng S. Estimation of monomer reactivity ratios by the error-in-variable method. *Macromolecules.* 1986;19(11):2779-2787.

16. Suddaby K, Hunt K, Haddleton D. MALDI-TOF mass spectrometry in the study of statistical copolymerizations and its application in examining the free radical copolymerization of methyl methacrylate and n-butyl methacrylate. *Macromolecules.* 1996;29(27):8642-8649.

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

18. McManus N, Dubé M, Penlidis A. High temperature bulk copolymerization of butyl acrylate/methyl methacrylate: reactivity ratio estimation. *Polymer Reaction Engineering.* 1999;7(1):131-145.

19. Zavala VM, Biegler LT. Large-scale parameter estimation in low-density polyethylene tubular reactors. *Ind Eng Chem Res.* 2006;45(23):7867-7881.

20. Arora N, Biegler LT. Parameter estimation for a polymerization reactor model with a composite-step trust-region NLP algorithm. *Ind Eng Chem Res.* 2004;43(14):3616-3631.

21. Yan F, Chu Y, Zhang K, et al. Determination of adsorption isotherm parameters with correlated errors by measurement error models. *Chemical Engineering Journal.* 2015;281:921-930.

22. High M, Danner R. Treatment of gas-solid adsorption data by the error-in-variables method. *AIChE journal.* 1986;32(7):1138-1145.

23. Sörner M, Ström KH. Estimation of equation of state parameters for 1, 1, 1, 2-tetrafluoroethane in a modified benedict-webb-rubin equation. *The Canadian Journal of Chemical Engineering.* 1995;73(6):854-861.

24. Duever TA, Keeler SE, Reilly PM, Vera JH, Williams PA. An application of the Error-in-Variables Model—parameter estimation from Van Ness-type vapour-liquid equilibrium experiments. *Chemical Engineering Science.* 1987/01/01/ 1987;42(3):403-412.

25. Schoonbrood HA, Aerdts AM, German AL, van der Velden GP. Determination of the intra- and intermolecular microstructure of bulk and emulsion copolymers of styrene and 2-hydroxyethyl methacrylate by means of proton NMR and gradient polymer elution chromatography. *Macromolecules.* 1995;28(16):5518-5525.

26. Sünbül D, Giz AT. An Error in Variables Method for Determining the Implicit Penultimate Copolymerization Ratios by On-Line Monitoring of Reactions. *Macromolecular theory and simulations.* 2006;15(4):311-320.

27. Brar A, Dutta K. Acrylonitrile and glycidyl methacrylate copolymers: Nuclear magnetic resonance characterization. *Macromolecules.* 1998;31(15):4695-4702.

28. Giz A. An error-in-variables method for use when the reactivity ratios in copolymerization are close to one. *Macromolecular theory and simulations.* 1998;7(4):391-397.

29. Choliński J, Palczewska-Tulińska M, Szafrańska A, Wyrzykowskastankiewicz D. A new method of parameter adjustment and diagnostic checks on gamma models used in vapor—liquid equilibrium calculations. *Chemical Engineering Science.* 1981;36(1):173-181.

30. Sutton TL, Macgregor JF. The analysis and design of binary vapour-liquid equilibrium experiments. Part I: Parameter estimation and consistency tests. *The Canadian Journal of Chemical Engineering.* 1977;55(5):602-608.

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

32. Kou B, McAuley KB, Hsu C, Bacon DW, Yao KZ. Gas-phase ethylene/hexene copolymerization with metallocene catalyst in a laboratory-scale reactor. *Ind Eng Chem Res.* 2005;44(8):2443-2450.

33. Dubé MA, Penlidis A. Hierarchical data analysis of a replicate experiment in emulsion terpolymerization. *AIChE journal.* 1996;42(7):1985-1994.

34. Box GE, Hunter WH, Hunter S. *Statistics for experimenters.* Vol 664: John Wiley and sons New York; 1978.

35. Dube M, Sanayei RA, Penlidis A, O'driscoll K, Reilly P. A microcomputer program for estimation of copolymerization reactivity ratios. *Journal of Polymer Science Part A: Polymer Chemistry.* 1991;29(5):703-708.

36. Brar A, Charan S. Sequence determination of vinyl acetate–methyl acrylate copolymers by NMR spectroscopy. *Journal of applied polymer science.* 1994;53(13):1813-1822.

37. Brar A, Charan S. Reactivity ratios and microstructure determination of vinyl acetate–alkyl acrylate copolymers by NMR spectroscopy. *Journal of Polymer Science Part A: Polymer Chemistry.* 1995;33(1):109-116.

38. Dubé MA, Penlidis A. A systematic approach to the study of multicomponent polymerization kinetics—the butyl acrylate/methyl methacrylate/vinyl acetate example: 1. Bulk copolymerization. *Polymer.* 1995;36(3):587-598.

39. Scott AJ, Penlidis A. Computational package for copolymerization reactivity ratio estimation: Improved access to the error-in-variables-model. *Processes.* 2018;6(1):8.

40. Brar A, Dutta K, Kapur G. Complete spectral assignments and microstructures of photopolymerized acrylonitrile/methacrylic acid copolymers by NMR spectroscopy. *Macromolecules.* 1995;28(26):8735-8741.

41. Brar A, Malhotra M. Compositional assignments and sequence distribution of vinylidene chloride− methyl acrylate copolymers using one-and two-dimensional NMR spectroscopy. *Macromolecules.* 1996;29(23):7470-7476.

42. McManus N, Penlidis A. A kinetic investigation of styrene/ethyl acrylate copolymerization. *Journal of Polymer Science Part A: Polymer Chemistry.* 1996;34(2):237-248.

43. Hakim M, Verhoeven V, McManus N, Dubé M, Penlidis A. High-temperature solution polymerization of butyl acrylate/methyl methacrylate: Reactivity ratio estimation. *Journal of applied polymer science.* 2000;77(3):602-609.

44. Scott AJ, Penlidis A. Binary vs. ternary reactivity ratios: Appropriate estimation procedures with terpolymerization data. *European Polymer Journal.* 2018;105:442-450.

45. Gabriel VA, Dubé MA. Bulk Free-Radical Co-and Terpolymerization of n-Butyl Acrylate/2-Ethylhexyl Acrylate/Methyl Methacrylate. *Macromolecular Reaction Engineering.* 2019;13(2):1800057.

46. Krieger BV, de Araujo PHH, Sayer C. Reactivity Ratios Estimation of the Free-Radical Polymerization of Itaconic Acid and N-Vinyl-2-Pyrrolidone by the Error-in-Variables Methodology. *Macromol React Eng.* 2020:2000026.

47. 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.

48. Rod V, HanČl V. Iterative estimation of model parameters when measurements of all variables are subject to error. *Computers & Chemical Engineering.* 1980;4(2):33-38.

49. Valko P, Vajda S. An extended Marquardt-type procedure for fitting error-in-variables models. *Computers & chemical engineering.* 1987;11(1):37-43.

50. Kim IW, Liebman MJ, Edgar TF. Robust error-in-variables estimation using nonlinear programming techniques. *AIChE Journal.* 1990;36(7):985-993.

51. Bardow A, Marquardt W. Identification of diffusive transport by means of an incremental approach. *Computers & chemical engineering.* 2004;28(5):585-595.

52. Kim I-W, Edgar T, Bell N. Parameter estimation for a laboratory water—gas-shift reactor using a nonlinear error-in-variables method. *Computers & chemical engineering.* 1991;15(5):361-367.
53. Vamos RJ, Haas CN. Reduction of ion-exchange equilibria data using an error in variables approach. *AIChE journal.* 1994;40(3):556-569.
54. Faber R, Li P, Wozny G. Sequential parameter estimation for large-scale systems with multiple data sets. 2. Application to an industrial coke-oven-gas purification process. *Industrial & engineering chemistry research.* 2004;43(15):4350-4362.
55. Leung T, Hartloper C, Smith J, Botros K, Golshan H, Glen T. Application of EVM to determine pipe internal surface conditions in complex natural gas lateral networks. *Journal of Natural Gas Science and Engineering.* 2015;24:217-227.
56. Soukissian TH, Papadopoulos A. Effects of different wind data sources in offshore wind power assessment. *Renewable Energy.* 2015;77:101-114.
57. Lira I, Grientschnig D. Error-in-variables models in calibration. *Metrologia.* 2017;54(6):S133.
58. Andersen KK, Reddy TA. The error in variables (EIV) regression approach as a means of identifying unbiased physical parameter estimates: Application to chiller performance data. *HVAC&R Research.* 2002;8(3):295-309.
59. Sanchez B, Schoukens J, Bragos R, Vandersteen G. Novel estimation of the electrical bioimpedance using the local polynomial method. Application to in vivo real-time myocardium tissue impedance characterization during the cardiac cycle. *IEEE Transactions on Biomedical Engineering.* 2011;58(12):3376-3385.
60. Das SK, Basudhar PK. Parameter optimization of rock failure criterion using error-in-variables approach. *International Journal of Geomechanics.* 2011;11(1):36-43.
61. Doganaksoy N, van Meer H. An application of the linear errors-in-variables model in semiconductor device performance assessment. *Quality Engineering.* 2015;27(4):500-511.
62. Matei BC, Meer P. Estimation of nonlinear errors-in-variables models for computer vision applications. *IEEE transactions on pattern analysis and machine intelligence.* 2006;28(10):1537-1552.
63. Polic A, Duever T, Penlidis A. Case studies and literature review on the estimation of copolymerization reactivity ratios. *Journal of Polymer Science Part A: Polymer Chemistry.* 1998;36(5):813-822.
64. Hamilton D. Confidence regions for parameter subsets in nonlinear regression. *Biometrika.* 1986;73(1):57-64.