



**HAL**  
open science

# Kinetic modelling: Regression and validation stages, a compulsory tandem for kinetic model assessment

Sébastien Leveneur

► **To cite this version:**

Sébastien Leveneur. Kinetic modelling: Regression and validation stages, a compulsory tandem for kinetic model assessment. Canadian Journal of Chemical Engineering, In press, 10.1002/cjce.24956 . hal-04103190

**HAL Id: hal-04103190**

**<https://normandie-univ.hal.science/hal-04103190v1>**

Submitted on 23 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## RESEARCH ARTICLE

# Kinetic modelling: Regression and validation stages, a compulsory tandem for kinetic model assessment

Sébastien Leveueur 

INSA Rouen Normandie, Univ Rouen  
Normandie, Normandie Université,  
Rouen, France

**Correspondence**

Sébastien Leveueur, INSA Rouen  
Normandie, Univ Rouen Normandie,  
Normandie Université, LSPC, UR 4704,  
F-76000 Rouen, France.  
Email: [sebastien.leveueur@insa-rouen.fr](mailto:sebastien.leveueur@insa-rouen.fr)

**Funding information**

Métropole Rouen Normandie

**Abstract**

The development of robust and reliable kinetic models is vital to build safe, eco-friendly, and cost-competitive chemical processes. Establishing kinetic models for complex chemical systems such as biomass valorization is cumbersome because the kinetic modeller must test different models and fit several experimental observables (or concentrations). Usually, in chemical reaction engineering, kinetic model assessment is based solely on the regression stage outputs. The implementation of a validation stage can aid in choosing the most reliable kinetic models, essentially in the case of complex chemical systems. We studied the solvolysis of 5-hydroxymethylfurfural (5-HMF) to butyl levulinate (BL) as a model reaction constituting several consecutive and parallel reaction steps. From an existing kinetic model, we created 60 synthetic runs in batch conditions. In the first part, we tested four different models with 5 degrees of noise, and we carried out the modelling on the 60 synthetic runs. In the second part, two types of holdout methods were evaluated. In the last part, cross-validation, namely the k-fold method, was used. We found that the 10-fold method allowed more efficient selection results even when the noise level was high. Besides, k-fold allows for not scarifying experimental runs and selecting the most reliable model.

**KEYWORDS**

AIC, cross-validation, holdout, k-fold, kinetic model, validation

## 1 | INTRODUCTION

The kinetic model has a central role in the development of chemical processes. From this model, we can optimize the kinetics of production and, thus, the economic viability of a chemical process.

The more robust and reliable a kinetic model is, the better the optimization steps are.<sup>[1–3]</sup> The kinetic modelling stage is time-demanding because it requires experimental data, analytical method development, and model

testing. To develop a reliable kinetic model, we need reliable analytical methods to determine the evolution qualitatively and quantitatively of different species in permanent or transient regimes. The kinetic modeller keeps in mind that the complexity of his/her model depends on the accuracy of the analytical methods. We do not necessarily want a complex kinetic model describing all the reaction steps, but rather a model with the ability to describe the production kinetics of targeted chemicals in a reliable way.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Author. The *Canadian Journal of Chemical Engineering* published by Wiley Periodicals LLC on behalf of Canadian Society for Chemical Engineering.

In chemical reaction (or reactor) engineering,<sup>[4–6]</sup> there is a clear tendency in the current literature to develop complex kinetic models and microkinetic ones. Mechanical quantum calculations such as density functional theory (DFT) are increasingly used to unravel elementary reaction steps and some kinetic/thermodynamic constants.<sup>[7–11]</sup> This combination allows for decreasing the number of parameters to estimate. The drawback of this approach is that the quantum mechanical estimation could be time-demanding and uncertain. For instance, solvent effect and non-homogeneous surface catalysts are still challenging to include in the quantum mechanical calculation. To the best of our knowledge, there are no microkinetic models including a validation stage. Thus, their industrial application could be limited.

In research on kinetic models describing surface-catalyzed reactions, we can also notice a great effort to assess several surface reaction models and a big one for parameter estimation. Despite this significant effort, we regret that the validation stage is rare in such papers.<sup>[12–14]</sup> Usually, kinetic modellers discriminate models based on the output from the regression stage. For instance, the modeller could compare the residual sum of squares, plot the parity plot, or calculate the Akaike information criterion (AIC) to find the most reliable models.<sup>[15–18]</sup>

The fact that the validation stage is not commonly used could be due to the fact that the kinetic modeller only considers the holdout validation method, leading to the sacrifice of some experimental runs from the regression stage. The cross-validation (CV) approach, not common in kinetic modelling articles, can be used to overcome this issue. Some articles have shown the benefits of this approach compared to the holdout method, especially as a tool for model selection.<sup>[12–14]</sup>

To fill this gap, this article compares three assessment approaches: regression, regression-validation, and regression-CV applied to the production of butyl levulinate (BL) from the alcoholysis of 5-hydroxymethylfurfural (5-HMF).<sup>[14]</sup> This molecule comes from the valorization of lignocellulosic biomass that is not in competition with the alimentary sector. BL is considered a promising platform molecule and promising biofuel and/or fuel additive.<sup>[19–21]</sup>

Developing reliable and robust kinetic models for valorizing lignocellulosic biomass is vital for developing these sustainable processes.<sup>[22]</sup> The transformation of lignocellulosic biomass into platform molecules comprises several unidentified reaction steps leading to the production of several intermediates or side products. For instance, the acid-catalyzed hydrolysis of lignocellulosic biomass to levulinic acid (LA) produces more than 15 intermediates or final products.<sup>[23]</sup> Based on these multiple observables (experimental concentrations),

kinetic modellers could be tempted to develop complex kinetic models, requiring the estimation of several parameters and increasing the risk of overfitting.<sup>[24–27]</sup> A validation stage can determine the presence of overfitting.

In kinetic modelling, the implementation of a validation stage is more relevant for complex kinetic systems than simple kinetic systems, that is, chemical systems comprising less than two reaction steps. As mentioned earlier, one can develop several possible kinetic models for complex kinetic systems, and the validation stage is a supplementary tool in model selection. It was also important for the author to show the benefit of this approach in a real case study of biomass valorization.

This article proposes the evaluation of the efficiency of the k-fold method compared to the holdout method and the no-validation method. For that, synthetic data were generated by simulating a kinetic model for the solvolysis of 5-HMF over Amberlite IR-120.<sup>[14]</sup> Then, different noise levels were added to the synthetic data, that is, concentrations. We did not remove the outliers because this article aimed to assess the different validation methods even when the noise was high.

## 2 | NUMERICAL EXPERIMENTS

Di Meno Di Buchianico et al.<sup>[14]</sup> found that the reaction pathway displayed in Figure 1 was the most reliable. They assessed different models for the solvolysis of fructose to BL over Amberlite IR-120 in an autoclave. Their study used  $\gamma$ -valerolactone (GVL) as a co-solvent for solubility reasons. A previous study found that internal and external mass transfer resistance can be neglected.<sup>[20]</sup> The rate expressions are represented by Equations (1)–(8). In this study, it was assumed that this model is the true one, and the values of kinetic constants are displayed in Table S1.

From Figure 1, one can notice that 5-HMF can be degraded into humins, a hydrocarbon polymer. 5-HMF can also be transformed into formic acid (FA) and Int1, then Int1 can be transformed into LA. FA and LA can be esterified by butanol into butyl formate (BF) and BL. 5-HMF can also be etherified into 5-butoxymethylfurfural (5-BMF). This chemical (5-BMF) can be transformed into BF and BL.

The reaction rates are expressed as follows:

$$R_1 = k_1 \cdot [\text{HMF}] \cdot [\text{Prot.}] \quad (1)$$

$$R_2 = k_2 \cdot [\text{INT1}] \cdot [\text{Prot.}] \quad (2)$$

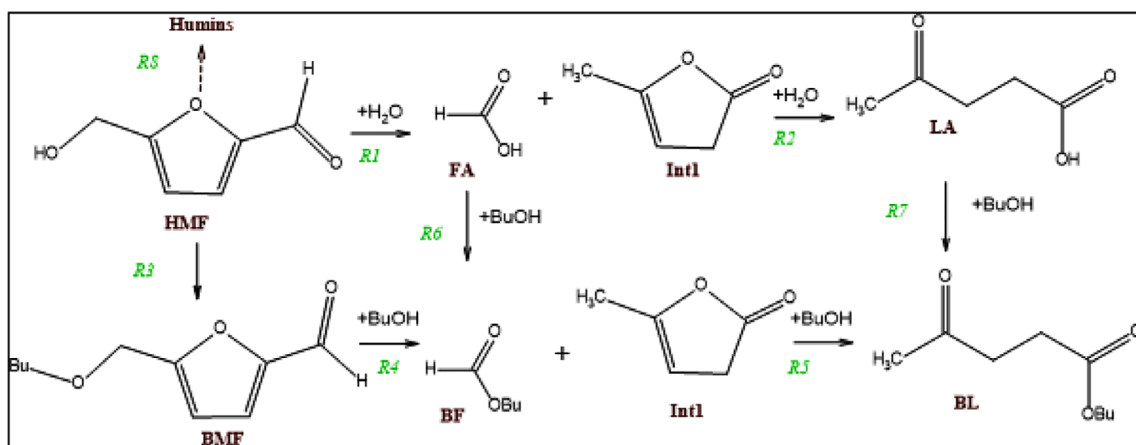


FIGURE 1 Reaction pathway for the solvolysis of 5-hydroxymethylfurfural (5-HMF) over Amberlite IR-120. BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; FA, formic acid; LA, levulinic acid.

$$R_3 = k_3 \cdot [\text{HMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}] \quad (3)$$

$$R_4 = k_4 \cdot [\text{BMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}] \quad (4)$$

$$R_5 = k_5 \cdot [\text{INT1}] \cdot [\text{BuOH}] \cdot [\text{Prot.}] \quad (5)$$

$$R_6 = k_6 \cdot [\text{FA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}] \quad (6)$$

$$R_7 = k_7 \cdot [\text{LA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}] \quad (7)$$

$$R_8 = k_8 \cdot [\text{HMF}] \cdot [\text{Prot.}] \quad (8)$$

$$\frac{d[\text{INT1}]}{dt} = R_1 - R_2 + R_4 - R_5 \quad (11)$$

$$\frac{d[\text{LA}]}{dt} = R_2 - R_7 \quad (12)$$

$$\frac{d[\text{BL}]}{dt} = R_5 + R_7 \quad (13)$$

$$\frac{d[\text{BuOH}]}{dt} = -R_3 - R_4 - R_5 - R_6 - R_7 \quad (14)$$

$$\frac{d[\text{FA}]}{dt} = R_1 - R_6 \quad (15)$$

$$\frac{d[\text{BF}]}{dt} = R_4 + R_6 \quad (16)$$

$$\frac{d[\text{Humins}]}{dt} = R_8 \quad (17)$$

The term  $[\text{Prot.}]$  is the concentration of protons. The proton capacity of Amberlite IR-120 is 4.4 meq of proton per dried gram of catalyst.<sup>[28-31]</sup> We have considered a free proton model, meaning that the degree of freedom is high. The reaction volume,  $V_{\text{Reaction}}$ , was calculated based on butanol density at the corresponding temperature.<sup>[32]</sup>

$$[\text{Prot.}] = \frac{m_{\text{dried catalyst}} \cdot \text{Capacity} \left( \frac{\text{mol of Proton}}{m_{\text{dried catalyst}}} \right)}{V_{\text{Reaction}}}$$

One can notice that protons catalyze all reaction steps.

Material balances by assuming an ideal batch reactor lead to the following ordinary differential equations (ODEs):

$$\frac{d[\text{HMF}]}{dt} = -R_1 - R_3 - R_8 \quad (9)$$

$$\frac{d[\text{BMF}]}{dt} = R_3 - R_4 \quad (10)$$

Table S1 shows the estimated kinetic constant obtained from Di Meno Di Buchianico et al.<sup>[14]</sup>

ODEs (9)–(17) were solved by the DDAPLUS solver, based on a modified Newton algorithm,<sup>[33]</sup> and implemented in Athena Visual studio.<sup>[34]</sup> We can generate synthetic data over operating conditions using these true kinetic constants (Table S1) and solve these ODEs (Table S2).

Table S2 shows the initial operating conditions for the 60 runs. This experimental matrix was designed by varying different initial operating conditions as a kinetic experimenter would have done. We varied initial concentrations of different species, catalyst loadings, and reaction temperatures. Some experiments were carried out

(numerically) with an initial amount of FA or LA to correctly estimate the kinetic constants of esterification (Figure 1).

### 3 | ERRORS

The noise was added to the synthetic data via white Gaussian noise. The Matlab function `awgn` was used to add this noise.

The previous article mentioned that the following species concentrations were used as observables: 5-HMF, BMF, LA, BL, and BF.<sup>[14]</sup> Thus, different levels of noise were added to these species concentrations.

Table 1 shows the  $SSE_i$  at different levels of noise. For LA concentration, the added noise was lower because in the work of Di Menno Di Bucchianico et al.,<sup>[14]</sup> the analytical error was lower than the other species. SSE was defined as follows:

$$SSE_i = \sum (C_{\text{true value of } i} - C_{\text{noised value of } i})^2 \quad (18)$$

where  $C_{\text{true value of } i}$  is the true concentration of species  $i$  and  $C_{\text{noised value of } i}$  is the noised concentration of species  $i$ .

The noise decreases in the following order: Error 1 < Error 4 < Error 3 < Error 2 < Error 5.

Figures S1–S5 show the true concentration of  $i$  versus the noise-corrupted concentration of  $i$  at different noise levels. True concentrations with error 5 are the most scattered data.

### 4 | MODELLING

As shown in a previous article of our research group,<sup>[14]</sup> the elementary reaction steps for this system are still under debate and are unclear. We have proposed four probable reaction models, which are briefly explained below. Figure 2 shows a simplified pathway for 5-HMF alcoholysis, where there is no intermediate. Table 2

shows the rate expressions for the different reactions in Model 1, and Table 3 for Model 2.

Material balances for Models 1 and 2 lead to the following ODEs:

$$\frac{d[\text{HMF}]}{dt} = -R_1 - R_2 - R_6 \quad (19)$$

$$\frac{d[\text{BMF}]}{dt} = R_2 - R_3 \quad (20)$$

$$\frac{d[\text{LA}]}{dt} = R_1 - R_4 \quad (21)$$

$$\frac{d[\text{BL}]}{dt} = R_3 + R_4 \quad (22)$$

$$\frac{d[\text{BuOH}]}{dt} = -R_2 - 2 \cdot R_3 - R_4 - R_5 \quad (23)$$

$$\frac{d[\text{FA}]}{dt} = R_1 - R_5 \quad (24)$$

$$\frac{d[\text{BF}]}{dt} = R_3 + R_5 \quad (25)$$

$$\frac{d[\text{Humins}]}{dt} = R_6 \quad (26)$$

Figure 1 shows the mechanism for Models 3 and 4. Table 4 shows the rate expression for the different reactions in Model 3, and Table 5 for Model 4. Model 3 was the true one used to produce the synthetic data using the estimated kinetic constants from Di Menno Di Bucchianico et al.<sup>[14]</sup> (Table S1).

Regression and validation stages were carried out by the commercial software Athena Visual Studio.<sup>[34,35]</sup>

HMF, LA, BMF, BF, and BL concentrations were used as observables. These concentrations, named experimental concentrations, are obtained from the synthetic data plus the noise. For such a multi-response system, the Bayesian framework, implemented in Athena Visual

TABLE 1  $SSE_i$  at different levels of noise.

	SSE_HMF (mol <sup>2</sup> /L <sup>2</sup> )	SSE_BMF (mol <sup>2</sup> /L <sup>2</sup> )	SSE_LA (mol <sup>2</sup> /L <sup>2</sup> )	SSE_BL (mol <sup>2</sup> /L <sup>2</sup> )	SSE_BF (mol <sup>2</sup> /L <sup>2</sup> )
Error 1	3.05E – 03	3.01E – 03	3.00E – 05	3.02E – 03	3.05E – 03
Error 4	8.16E – 03	8.79E – 03	8.34E – 05	8.54E – 03	8.49E – 03
Error 3	2.64E – 02	2.64E – 02	2.40E – 04	2.50E – 02	2.51E – 02
Error 2	2.41E – 01	2.33E – 01	2.13E – 03	2.52E – 01	2.48E – 01
Error 5	2.40E + 00	2.32E + 00	2.13E – 03	2.52E – 01	9.92E + 00

Abbreviations: BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSE, sum of squares for error.

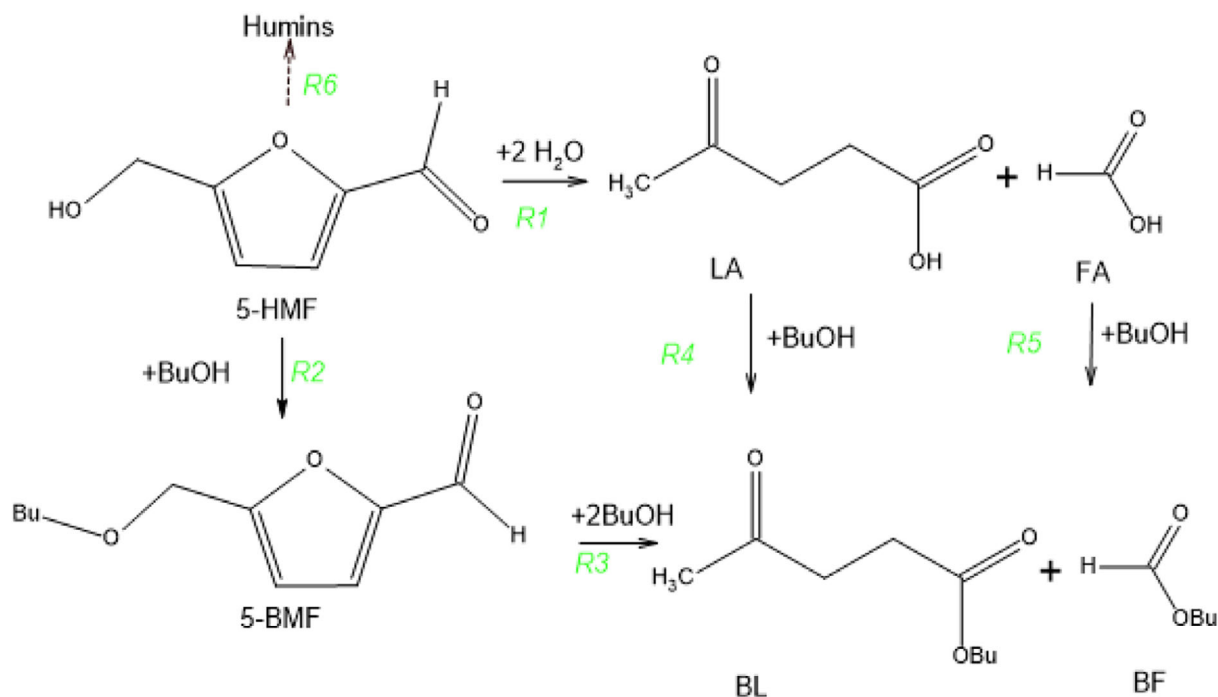


FIGURE 2 Mechanism for the alcoholysis of 5-hydroxymethylfurfural (5-HMF): Models 1 and 2. BF, butyl formiate; BL, butyl levulinate; 5-BMF, butoxymethylfurfural; FA, formic acid; LA, levulinic acid.

TABLE 2 Model 1 and kinetic constants to estimate.

Reaction	Rate expression	Constants to estimate
$5\text{-HMF} + 2\text{H}_2\text{O} \xrightarrow{\text{Prot.}} \text{LA} + \text{FA}$	$R_1 = k_1 \cdot [\text{HMF}] \cdot [\text{Prot.}]$	$\ln(k_1(T_{\text{ref}}))$ and $\frac{E_{a1}}{R \cdot T_{\text{ref}}}$
$5\text{-HMF} + \text{BuOH} \xrightarrow{\text{Prot.}} \text{BMF}$	$R_2 = k_2 \cdot [\text{HMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_2(T_{\text{ref}}))$ and $\frac{E_{a2}}{R \cdot T_{\text{ref}}}$
$\text{BMF} + 2\text{BuOH} \xrightarrow{\text{Prot.}} \text{BL} + \text{BF}$	$R_3 = k_3 \cdot [\text{BMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_3(T_{\text{ref}}))$ and $\frac{E_{a3}}{R \cdot T_{\text{ref}}}$
$\text{LA} + \text{BuOH} \xrightarrow{\text{Prot.}} \text{BL} + \text{H}_2\text{O}$	$R_4 = k_4 \cdot [\text{LA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_4(T_{\text{ref}}))$ and $\frac{E_{a4}}{R \cdot T_{\text{ref}}}$
$\text{FA} + \text{BuOH} \xrightarrow{\text{Prot.}} \text{BF} + \text{H}_2\text{O}$	$R_5 = k_5 \cdot [\text{FA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_5(T_{\text{ref}}))$ and $\frac{E_{a5}}{R \cdot T_{\text{ref}}}$
$5\text{-HMF} \xrightarrow{\text{Prot.}} \text{Humins}$	$R_6 = k_6 \cdot [\text{HMF}] \cdot [\text{Prot.}]$	$\ln(k_6(T_{\text{ref}}))$ and $\frac{E_{a6}}{R \cdot T_{\text{ref}}}$

Abbreviations: BF, butyl formiate; BMF, butoxymethylfurfural; FA, formic acid; HMF, hydroxymethylfurfural; LA, levulinic acid.

TABLE 3 Model 2 and kinetic constants to estimate.

Reaction	Rate expression	Constants to estimate
$5\text{-HMF} + 2\text{H}_2\text{O} \xrightarrow{\text{Prot.}} \text{LA} + \text{FA}$	$R_1 = k_1 \cdot [\text{HMF}] \cdot [\text{Prot.}]$	$\ln(k_1(T_{\text{ref}}))$ and $\frac{E_{a1}}{R \cdot T_{\text{ref}}}$
$5\text{-HMF} + \text{BuOH} \xrightarrow{\text{Prot.}} \text{BMF}$	$R_2 = k_2 \cdot [\text{HMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_2(T_{\text{ref}}))$ and $\frac{E_{a2}}{R \cdot T_{\text{ref}}}$
$\text{BMF} + 2\text{BuOH} \xrightarrow{\text{Prot.}} \text{BL} + \text{BF}$	$R_3 = k_3 \cdot [\text{BMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_3(T_{\text{ref}}))$ and $\frac{E_{a3}}{R \cdot T_{\text{ref}}}$
$\text{LA} + \text{BuOH} \xrightarrow{\text{Prot.}} \text{BL} + 2\text{H}_2\text{O}$	$R_4 = k_4 \cdot [\text{LA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_4(T_{\text{ref}}))$ and $\frac{E_{a4}}{R \cdot T_{\text{ref}}}$
$\text{FA} + \text{BuOH} \xrightarrow{\text{Prot.}} \text{BF} + \text{H}_2\text{O}$	$R_5 = k_5 \cdot [\text{FA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_5(T_{\text{ref}}))$ and $\frac{E_{a5}}{R \cdot T_{\text{ref}}}$
$5\text{-HMF} \xrightarrow{\text{Prot.}} \text{Humins}$	$R_6 = k_6 \cdot [\text{HMF}]^2 \cdot [\text{Prot.}]$	$\ln(k_6(T_{\text{ref}}))$ and $\frac{E_{a6}}{R \cdot T_{\text{ref}}}$

Abbreviations: BF, butyl formiate; BMF, butoxymethylfurfural; FA, formic acid; HMF, hydroxymethylfurfural; LA, levulinic acid.

Reaction	Rate expression	Constants to estimate
$5 - \text{HMF} + \text{H}_2\text{O} \xrightarrow{\text{Prot}} \text{FA} + \text{Int1}$	$R_1 = k_1 \cdot [\text{HMF}] \cdot [\text{Prot.}]$	$\ln(k_1(T_{\text{ref}}))$ and $\frac{E_{a1}}{R \cdot T_{\text{ref}}}$
$\text{Int1} + \text{H}_2\text{O} \xrightarrow{\text{Prot}} \text{LA}$	$R_2 = k_2 \cdot [\text{INT1}] \cdot [\text{Prot.}]$	$\ln(k_2(T_{\text{ref}}))$ and $\frac{E_{a2}}{R \cdot T_{\text{ref}}}$
$5 - \text{HMF} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BMF}$	$R_3 = k_3 \cdot [\text{HMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_3(T_{\text{ref}}))$ and $\frac{E_{a3}}{R \cdot T_{\text{ref}}}$
$\text{BMF} + \text{BuOH} \xrightarrow{\text{Prot}} \text{Int1} + \text{BF}$	$R_4 = k_4 \cdot [\text{BMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_4(T_{\text{ref}}))$ and $\frac{E_{a4}}{R \cdot T_{\text{ref}}}$
$\text{Int1} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BL}$	$R_5 = k_5 \cdot [\text{INT1}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_5(T_{\text{ref}}))$ and $\frac{E_{a5}}{R \cdot T_{\text{ref}}}$
$\text{FA} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BF} + \text{H}_2\text{O}$	$R_6 = k_6 \cdot [\text{FA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_6(T_{\text{ref}}))$ and $\frac{E_{a6}}{R \cdot T_{\text{ref}}}$
$\text{LA} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BF} + \text{H}_2\text{O}$	$R_7 = k_7 \cdot [\text{LA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_7(T_{\text{ref}}))$ and $\frac{E_{a7}}{R \cdot T_{\text{ref}}}$
$5 - \text{HMF} \xrightarrow{\text{Prot}} \text{Humins}$	$R_8 = k_8 \cdot [\text{HMF}] \cdot [\text{Prot.}]$	$\ln(k_8(T_{\text{ref}}))$ and $\frac{E_{a8}}{R \cdot T_{\text{ref}}}$

Abbreviations: BF, butyl formate; BMF, butoxymethylfurfural; FA, formic acid; HMF, hydroxymethylfurfural; LA, levulinic acid.

TABLE 4 Model 3 and kinetic constants to estimate.

Reaction	Rate expression	Constants to estimate
$5 - \text{HMF} + \text{H}_2\text{O} \xrightarrow{\text{Prot}} \text{FA} + \text{Int1}$	$R_1 = k_1 \cdot [\text{HMF}] \cdot [\text{Prot.}]$	$\ln(k_1(T_{\text{ref}}))$ and $\frac{E_{a1}}{R \cdot T_{\text{ref}}}$
$\text{Int1} + \text{H}_2\text{O} \xrightarrow{\text{Prot}} \text{LA}$	$R_2 = k_2 \cdot [\text{INT1}] \cdot [\text{Prot.}]$	$\ln(k_2(T_{\text{ref}}))$ and $\frac{E_{a2}}{R \cdot T_{\text{ref}}}$
$5 - \text{HMF} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BMF}$	$R_3 = k_3 \cdot [\text{HMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_3(T_{\text{ref}}))$ and $\frac{E_{a3}}{R \cdot T_{\text{ref}}}$
$\text{BMF} + \text{BuOH} \xrightarrow{\text{Prot}} \text{Int1} + \text{BF}$	$R_4 = k_4 \cdot [\text{BMF}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_4(T_{\text{ref}}))$ and $\frac{E_{a4}}{R \cdot T_{\text{ref}}}$
$\text{Int1} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BL}$	$R_5 = k_5 \cdot [\text{INT1}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_5(T_{\text{ref}}))$ and $\frac{E_{a5}}{R \cdot T_{\text{ref}}}$
$\text{FA} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BF} + \text{H}_2\text{O}$	$R_6 = k_6 \cdot [\text{FA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_6(T_{\text{ref}}))$ and $\frac{E_{a6}}{R \cdot T_{\text{ref}}}$
$\text{LA} + \text{BuOH} \xrightarrow{\text{Prot}} \text{BF} + \text{H}_2\text{O}$	$R_7 = k_7 \cdot [\text{LA}] \cdot [\text{BuOH}] \cdot [\text{Prot.}]$	$\ln(k_7(T_{\text{ref}}))$ and $\frac{E_{a7}}{R \cdot T_{\text{ref}}}$
$5 - \text{HMF} \xrightarrow{\text{Prot}} \text{Humins}$	$R_8 = k_8 \cdot [\text{HMF}]^2 \cdot [\text{Prot.}]$	$\ln(k_8(T_{\text{ref}}))$ and $\frac{E_{a8}}{R \cdot T_{\text{ref}}}$

Abbreviations: BF, butyl formate; BMF, butoxymethylfurfural; FA, formic acid; HMF, hydroxymethylfurfural; LA, levulinic acid.

TABLE 5 Model 4 and kinetic constants to estimate.

Studio, is more suitable than the non-linear least squares approach.<sup>[36,37]</sup> In the Bayesian framework, the minimization of the objective function (OF) requires the determination of the determinant criterion.<sup>[38]</sup>

The GREGPLUS subroutine minimizes the OF, determines the credible intervals for each estimated parameter, and calculates the normalized parameter covariance.

The minimization of the OF is done via successive quadratic programming.<sup>[34,37]</sup>

$$\text{OF} = (a + b + 1) \cdot \ln|v| \quad (27)$$

where  $|v|$  is the determinant of the covariance matrix of the responses,  $b$  is the number of responses, and  $a$  is the number of events in response. Each element of this matrix is as follows:

$$v_{ij} = \sum_{u=1}^n [C_{iu} - \hat{C}_{iu}] \cdot [C_{ju} - \hat{C}_{ju}] \quad (28)$$

where  $C_{iu}$  is the experimental concentration and  $\hat{C}_{iu}$  is the estimated value for response  $i$  and event  $u$ ; and  $C_{ju}$  the experimental concentration and  $\hat{C}_{ju}$  the estimated value for response  $j$  and event  $u$ .

The credible intervals of the estimated parameters were evaluated by the marginal highest posterior density (HPD).

A modified Arrhenius equation expresses the rate constants to consider the temperature effect. To decrease the strong correlation between the pre-exponential factor and the activation energy, Buzzi-Ferraris recommended linearizing the Arrhenius equation<sup>[39]</sup> as follows:

$$k(T) = \exp \left[ \ln(k(T_{\text{ref}})) + \frac{E_a}{R \cdot T_{\text{ref}}} \cdot \left( 1 - \frac{T_{\text{ref}}}{T} \right) \right] \quad (29)$$

where  $T_{\text{ref}}$  is the reference temperature.

Three different approaches were used for the kinetic modelling:



- Use of all experiments for the regression and no validation stage.
- Use of the holdout approach<sup>[40,41]</sup>: 80% of experiments for regression and 20% for validation; 90% of experiments for regression and 10% for validation.
- Use of CV<sup>[12,40–46]</sup>: 5-fold and 10-fold methods.

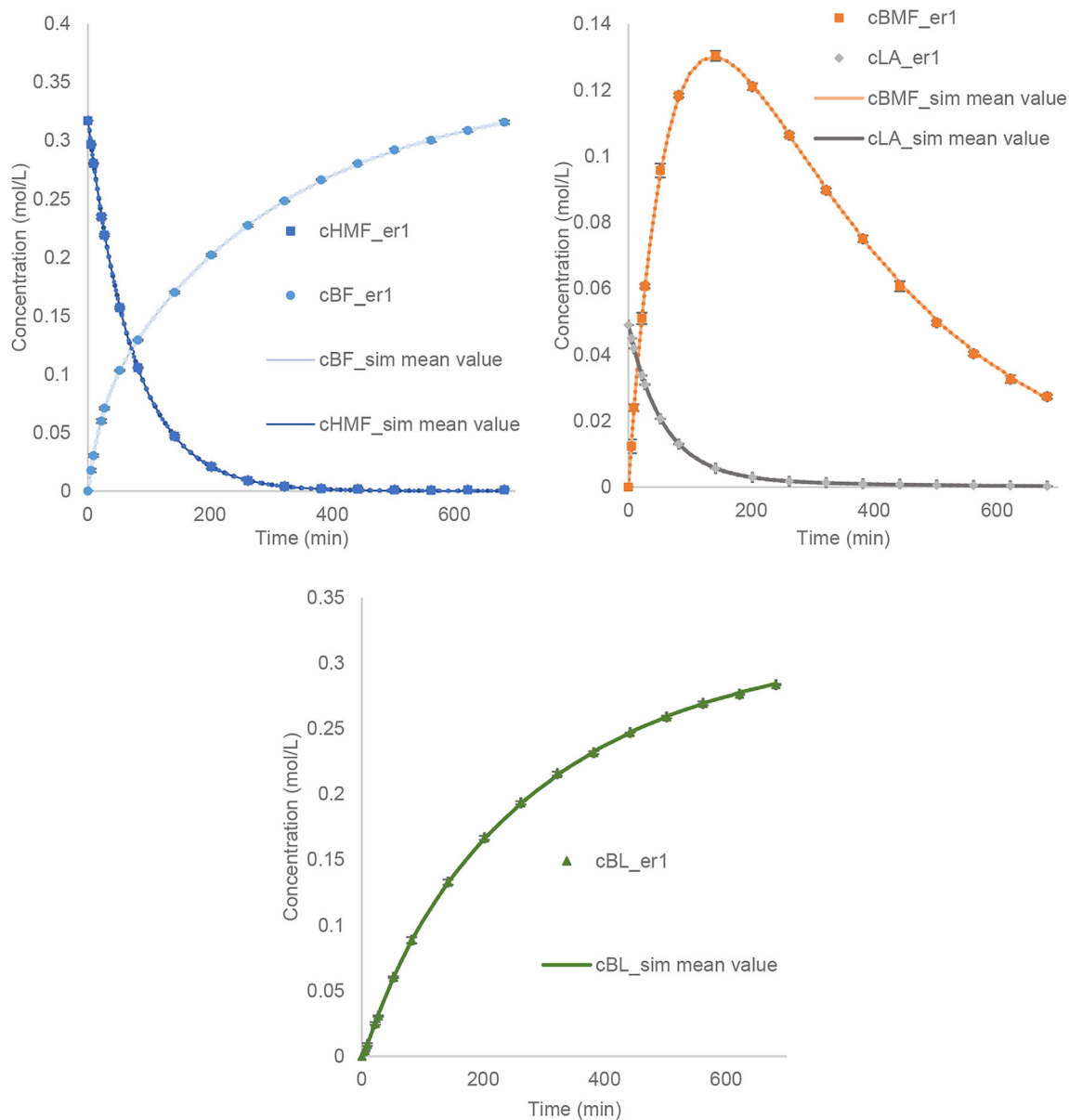
Visual fitting, parity plots, SSR, or CV evaluate regression or validation stages. In the multi-response system, one should calculate these properties for each observable.

#### 4.1 | All experiments

The 60 experiments (Table S2) with different errors were used for the regression to estimate the kinetic constants.

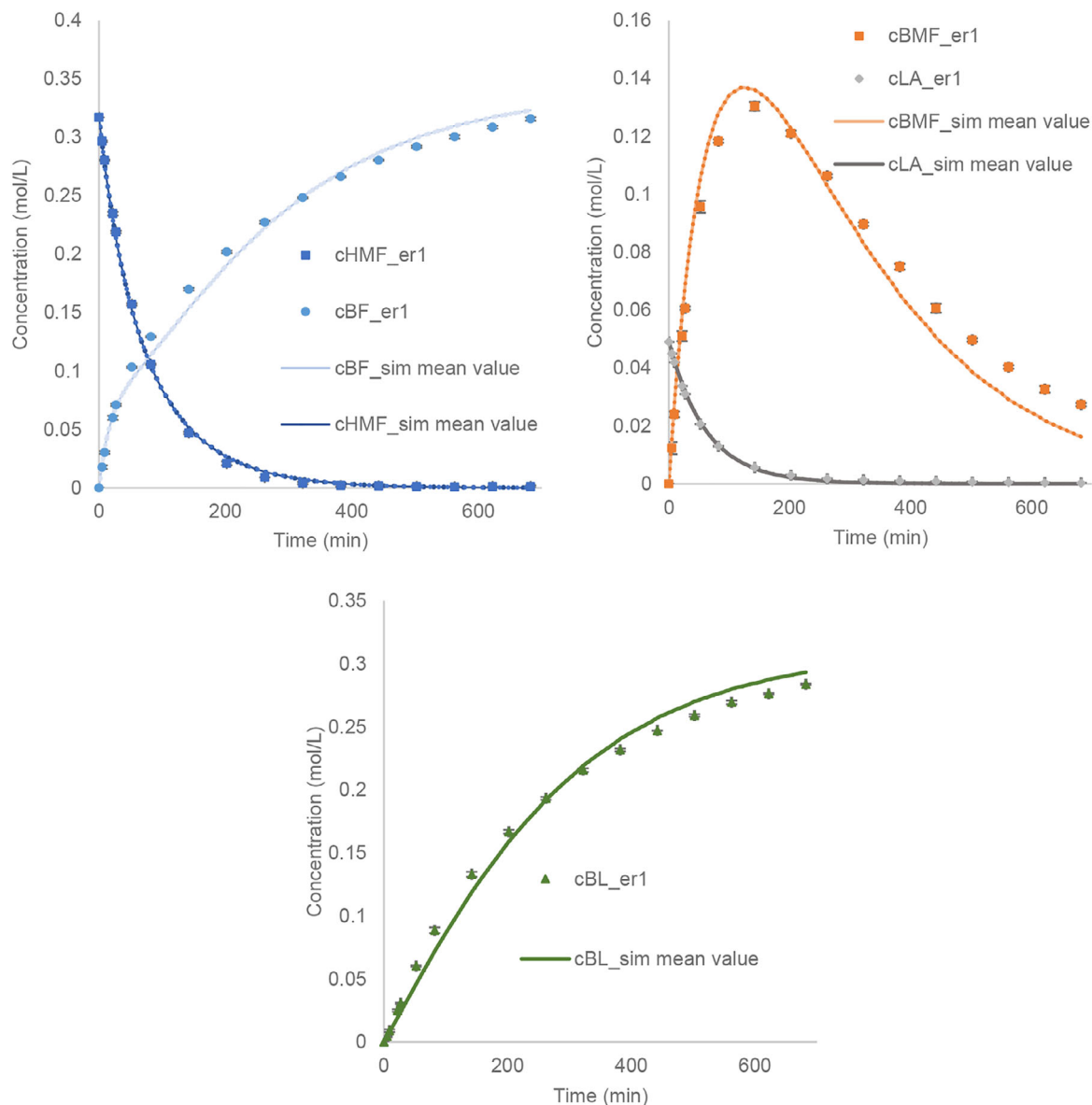
The fit of the model to the experimental data can give a first overview regarding the model reliability. Nevertheless, it could be challenging to discriminate different models just based on the fitting.

Figures 3–6 show the fit to Run 55 for Models 1–4 with the lowest error, that is, error 1. In Run 55, a significant



**FIGURE 3** Fit of Model 1 to the experimental concentrations (error 1) with error standard deviations for Run 55. cBF, concentration of butyl formate; cBL, concentration of butyl levulinate; cBMF, concentration of 5-butoxymethylfurfural; cHMF, concentration of 5-hydroxymethylfurfural; cLA, concentration of levulinic acid.





**FIGURE 4** Fit of Model 2 to the experimental concentrations (error 1) with error standard deviations for Run 55. cBF, concentration of butyl formate; cBL, concentration of butyl levulinate; cBMF, concentration of 5-butoxymethylfurfural; cHMF, concentration of 5-hydroxymethylfurfural; cLA, concentration of levulinic acid.

amount of LA was added. One can notice that Models 1 and 2 (Figures 3 and 4) show a slightly lower fitting quality than Models 3 and 4. It is difficult to discriminate between models based on fitting visuals. In SI (Figures S6–S9), we notice that parity plots for error 3 show that Model 3 is slightly better than Model 4.

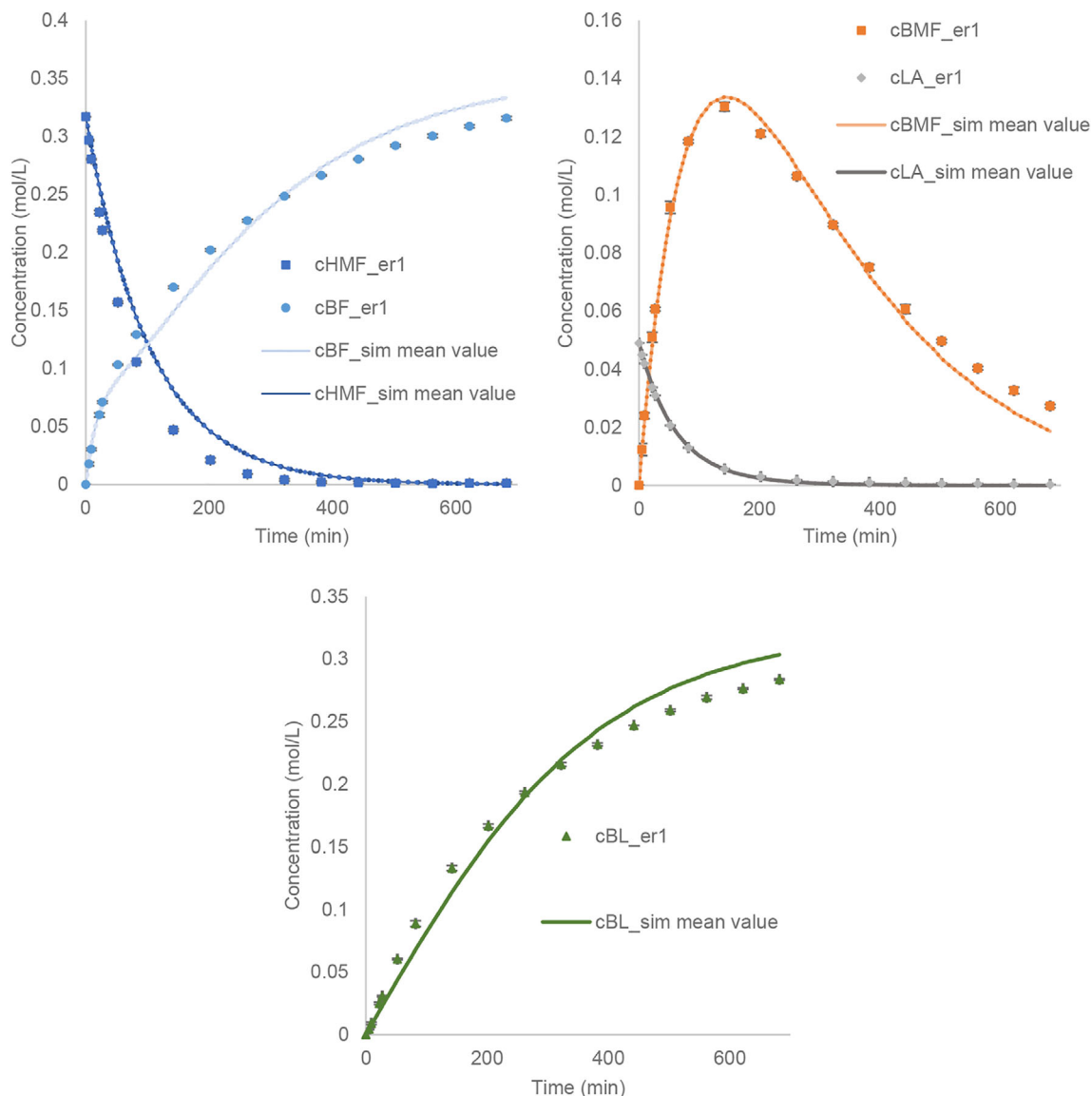
Figure 7 displays the sum of squared residuals from regression ( $SSR_{\text{Reg}}$ ) for the different models and errors.

$$SSR_{\text{Reg}_i} = \sum (C_{\text{simulated value of } i} - C_{\text{experimental value of } i})^2 \quad (30)$$

Figure 7 shows that for errors 1, 4, and 3, we can observe that Model 3 gives the lowest values of  $SSR_{\text{Reg}_i}$ . Nevertheless, we cannot discriminate between Models 3 and 4 for errors 2 and 5, that is, the ones with the higher degree of noise.

The AIC is another model selection tool that considers the trade-off between the number of estimated parameters and  $SSR_{\text{Reg}}$  values.<sup>[14,43,47–50]</sup>

$$\begin{aligned} \text{AIC} = & \text{number of independent events} \\ & \cdot \ln \left( \frac{SSR_{\text{Reg\_All}}}{\text{number of independent events}} \right) \\ & + 2 \text{ number of estimated parameters} \end{aligned} \quad (31)$$



**FIGURE 5** Fit of Model 3 to the experimental concentrations (error 1) with error standard deviations for Run 55. cBF, concentration of butyl formiate; cBL, concentration of butyl levulinate; cBMF, concentration of 5-butoxymethylfurfural; cHMF, concentration of 5-hydroxymethylfurfural; cLA, concentration of levulinic acid.

where  $SSR_{\text{Reg\_All}}$  is the sum of squared residuals from regression for BF, BL, LA, BMF, and HMF. Table 6 shows the AIC values for different models,  $SSR_{\text{Reg\_All}}$ , and the number of estimated parameters. For errors 1, 4, 3, and 2, Table 6 shows that Model 3 is the most reliable because AIC is the lowest for this model. For error 5, that is, the one with the highest noise level, Model 2 is predicted to be the most reliable based on AIC value.

Tables S3–S6 show the estimated values for all Models with credible intervals, represented by HPD. For errors with low levels of noise (errors 1, 4, and 3), the credible intervals are low due to the wide range of operating conditions. By increasing the noise level, Models have difficulty estimating some kinetic constants. For example,

Model 3 cannot estimate some kinetic constants for Reactions (2) or (5) when errors 2 and 5 are applied. Indeed, these two reactions involve the intermediate Int1 that we do not track.

The correlation matrix (Table S7) shows no correlation between estimated parameters, meaning that they are well-identified.

Figure S10 shows the estimated constants with their credible intervals for Model 3 compared to true constants.

We were able to estimate all parameters quite well (Figure S10). When the level of noise increases, it is more challenging to estimate kinetic constants for Reactions (2) and (5), because we do not track the intermediates.

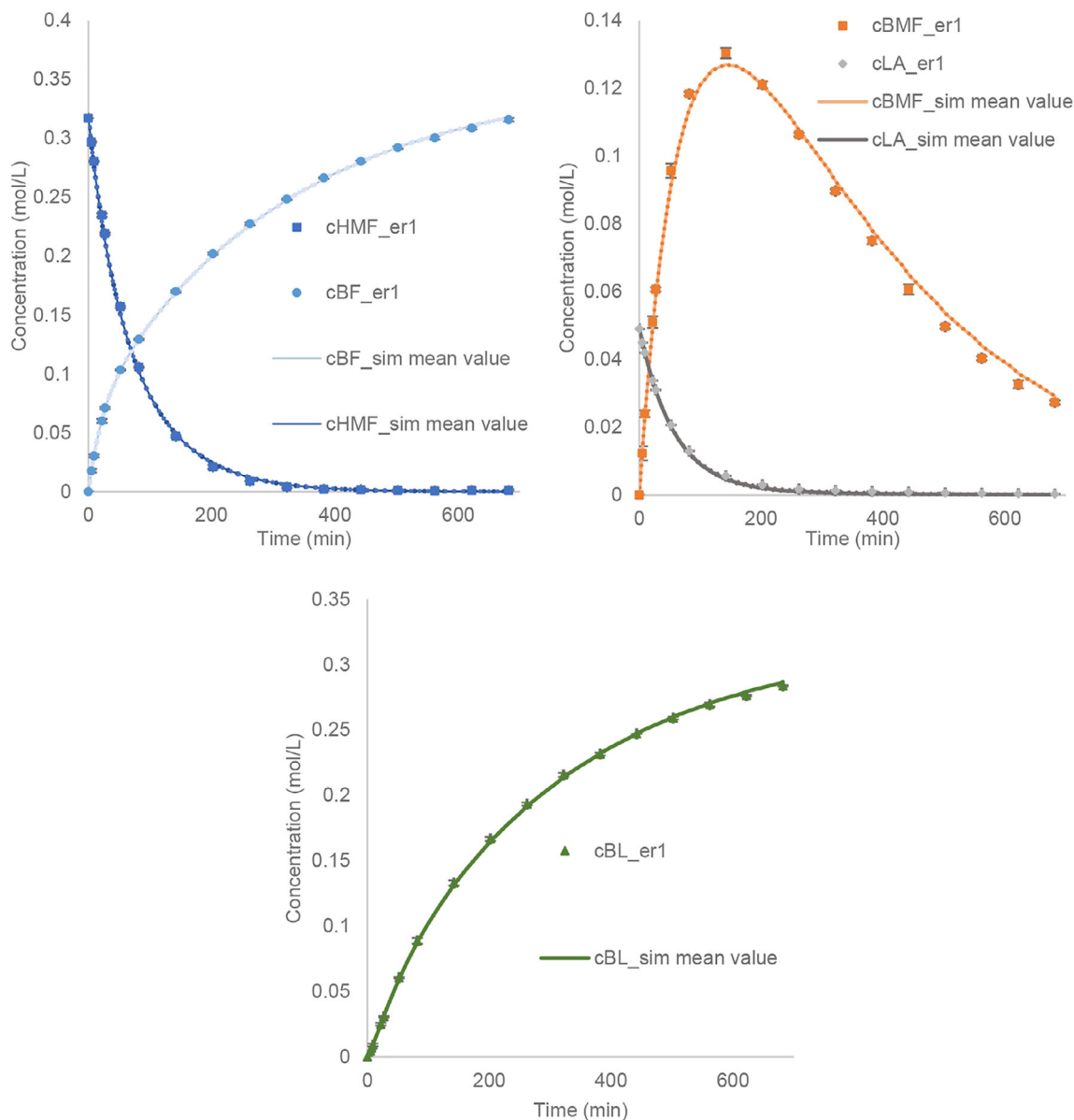


FIGURE 6 Fit of Model 4 to the experimental concentrations (error 1) with error standard deviations for Run 55. cBF, concentration of butyl formate; cBL, concentration of butyl levulinate; cBMF, concentration of 5-butoxymethylfurfural; cHMF, concentration of 5-hydroxymethylfurfural; cLA, concentration of levulinic acid.

## 4.2 | Holdout method

The holdout method uses a part of the experimental data for the regression and the other part for the validation.

Besides  $SSR_{Reg\_i}$ , we define the sum of squared residuals from the validation stage  $SSR_{Val\_i}$  as follows:

$$SSR_{Val\_i} = \sum (C_{\text{simulated value of } i} - C_{\text{experimental value from regression set of } i})^2 \quad (32)$$

From this method, we can analyze the quality of regression and validation.

We considered two options: 80% of the 60 runs are used for regression, and 90% of the 60 runs are used for regression.

### 4.2.1 | 80% of experiments for regression and 20% for validation

Experiments used for validation were 3, 24, 28, 40, 41, 45, 46, 48, 51, 55, 57, and 58. These experiments were randomly selected from all experiments.

Figure 8 shows the  $SSR_{Reg\_i}$  for all models. We can observe that the regression quality decreased compared

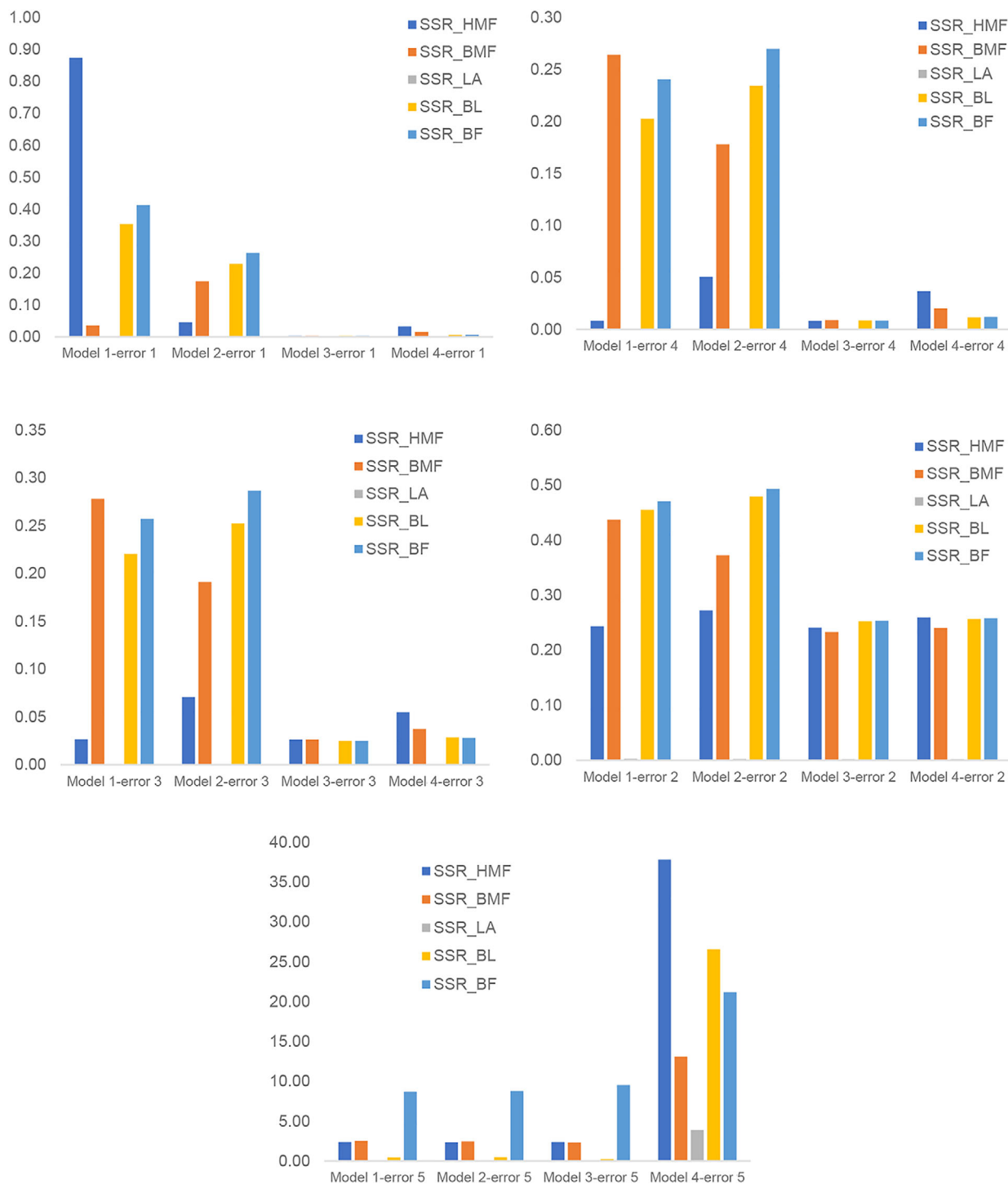


FIGURE 7  $SSR_{Reg}$  ( $\text{mol}^2/\text{L}^2$ ) for all experiments and models. BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid;  $SSR_{Reg}$ , sum of squared residuals from regression.

to the regression quality when all data were used (Figure 7). Surprisingly, with error 2, it is Model 4 that is the most reliable, that is, with lower  $SSR_{Reg_i}$ . With error 5, it is not possible to discriminate between the 4 models.

In the Supporting Information, we can find the estimated values and statistical data (Table S8).

Figure 9 shows the  $SSR_{Val_i}$  for all models. A similar observation can be done for the regression stage, that is, in error 2, we cannot discriminate between Models 3 and 4; in error 5, we cannot discriminate between all models.

Figure S11 shows that when the noise level increases, it is challenging to estimate the kinetic

	Number of estimated parameters	$SSR_{\text{Reg\_All}}$	AIC
Model 1–error 1	12	1.67	–122,760
Model 2–error 1	12	0.71	–134,422
Model 3–error 1	16	0.01	–189,770
Model 4–error 1	16	0.06	–168,226
Model 1–error 4	12	0.72	–134,344
Model 2–error 4	12	0.73	–134,016
Model 3–error 4	16	0.03	–175,822
Model 4–error 4	16	0.08	–164,112
Model 1–error 3	12	0.78	–133,131
Model 2–error 3	12	0.80	–132,813
Model 3–error 3	16	0.10	–160,748
Model 4–error 3	16	0.15	–155,680
Model 1–error 2	12	1.61	–123,309
Model 2–error 2	12	1.62	–123,216
Model 3–error 2	16	0.98	–130,035
Model 4–error 2	16	1.02	–129,560
Model 1–error 5	12	14.07	–93,731
Model 2–error 5	12	14.06	–93,749
Model 3–error 5	16	14.46	–93,353
Model 4–error 5	16	102.54	–66,645

TABLE 6 Akaike information criterion (AIC), number of estimated parameters, and  $SSR_{\text{Reg\_All}}$  for different numbers.

constants correctly from Reactions (2) and (5). Compared to the regression with all data, estimation quality is slightly lower for high errors, that is, errors 2 and 5.

#### 4.2.2 | 90% of experiments for regression and 10% for validation

Experiments used for validation were 9, 26, 30, 49, 55, and 58. These experiments were randomly selected from all experiments.

From Figure 10, we can observe the same trend as for the holdout approach 80/20. With error 2, we cannot discriminate between Models 3 and 4. With error 4, we cannot discriminate between any models.

From Figure 11, we can draw the same conclusions as from Figure 9 regarding the validation stage. This means that both approaches are good for this system.

In the Supporting Information, we can find the estimated values and statistical data (Table S9).

Figure S12 shows that kinetic constants for Reactions (2) and (5) are challenging to identify when the noise degree increases. This is because both reactions involve the intermediate that we do not track.

#### 4.2.3 | Comparison $\frac{SSR_{\text{Reg}_i}}{n_i \text{ from regression}}$ and $\frac{SSR_{\text{Val}_i}}{n_i \text{ from Validation}}$ between both holdout method

To compare both holdout methods, the following terms were calculated:  $\frac{SSR_{\text{Reg}_i}}{n_i \text{ from regression}}$  and  $\frac{SSR_{\text{Val}_i}}{n_i \text{ from Validation}}$ . The terms  $n_i \text{ from regression}$  and  $n_i \text{ from validation}$  are the number of experiments used for the regression and validation stages, respectively.

Figure S13 shows the value of  $\frac{SSR_{\text{Reg}_i}}{n_i \text{ from regression}}$  and  $\frac{SSR_{\text{Val}_i}}{n_i \text{ from Validation}}$  for error 3. For both methods, we can draw the same conclusions concerning the model selection. Thus, it could be more interesting to use 90/10 than 80/20 to improve the quality of estimated parameters.

### 4.3 | CV method

The 60 experiments were divided randomly into 5-fold (Table 7) and 10-fold (Table 8) methods. From the literature,  $K = 10$  is considered by several authors as the optimal value.<sup>[51]</sup> The regression stage was made on  $K$  minus 1 folds and the validation stage was on the remaining fold (Tables 9 and 10).

Kinetic constants were estimated from each regression, and these estimated constants were used for the

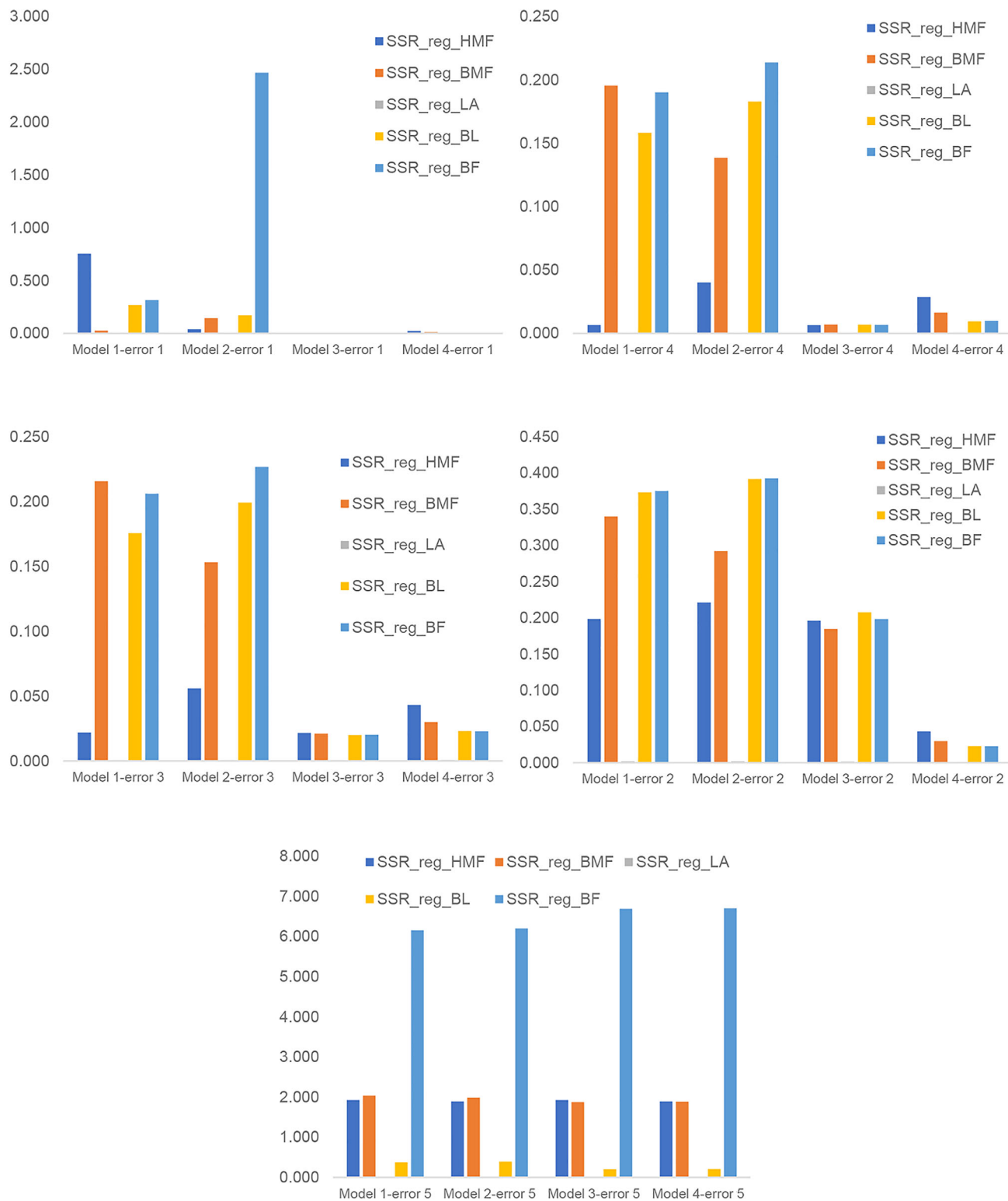


FIGURE 8 SSR<sub>Reg\_i</sub> for all models. BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSR, sum of squared residuals.

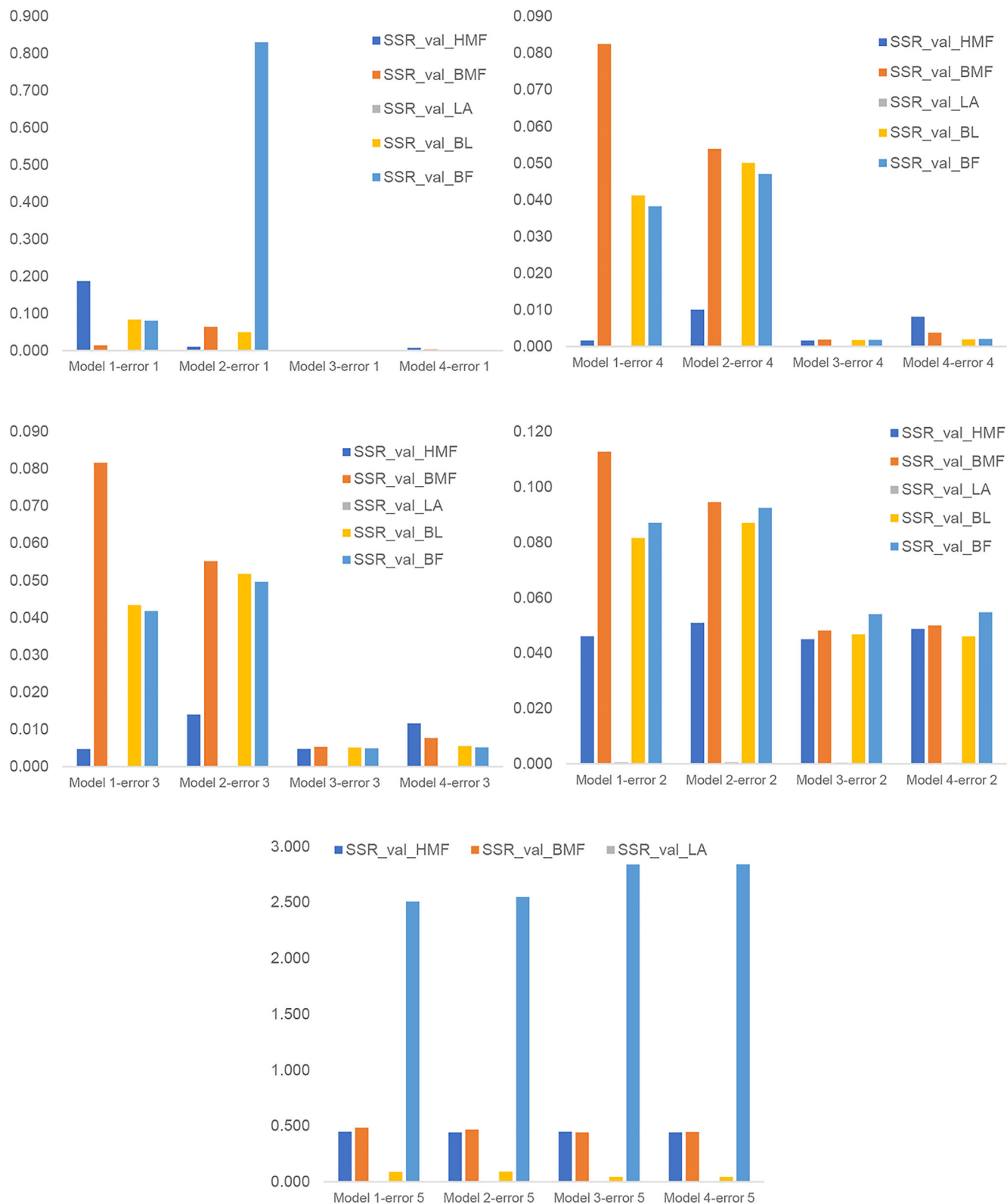
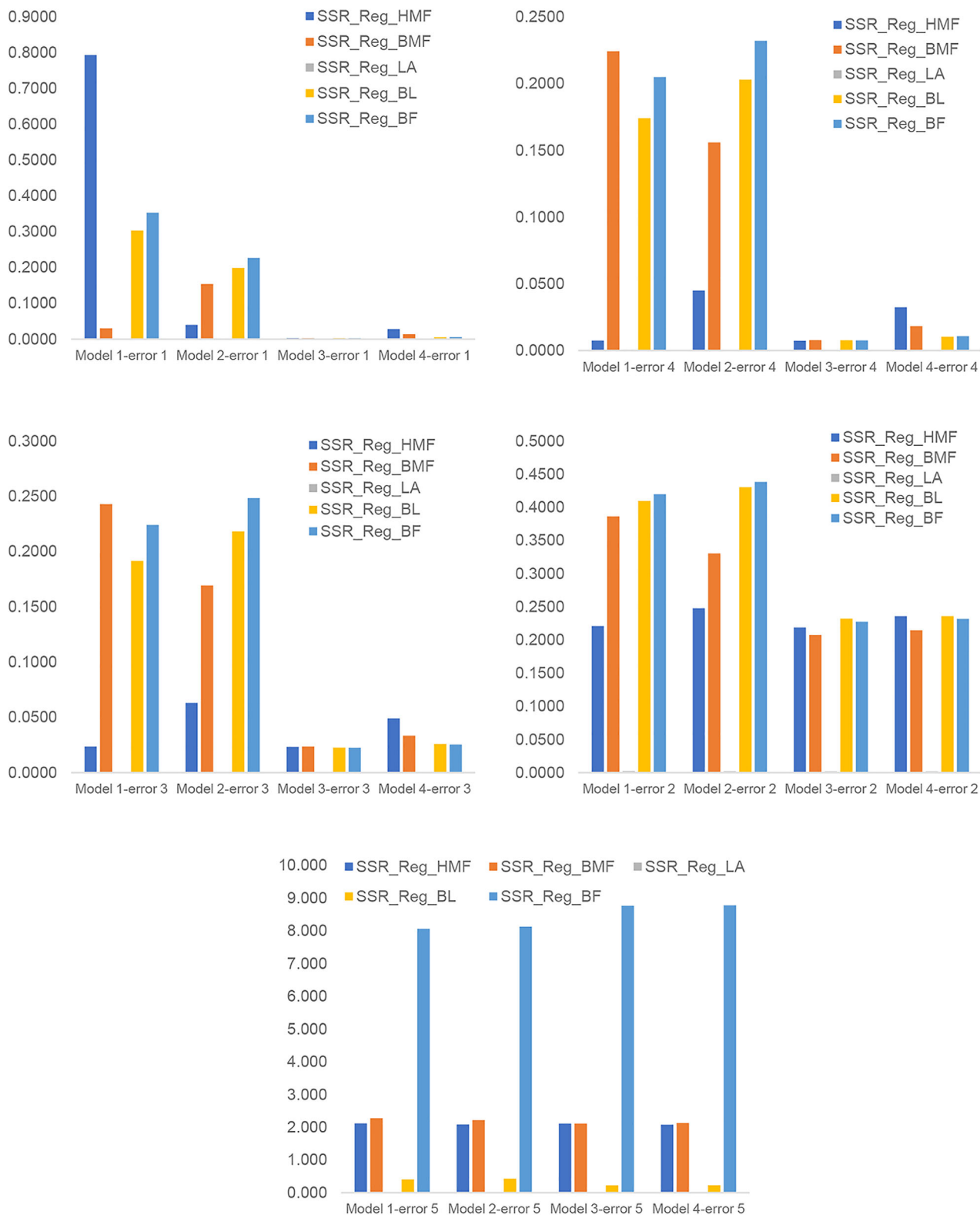


FIGURE 9  $SSR_{val_i}$  for all models. BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSR, sum of squared residuals.





**FIGURE 10**  $SSR_{Reg_i}$  for all models. BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSR, sum of squared residuals.

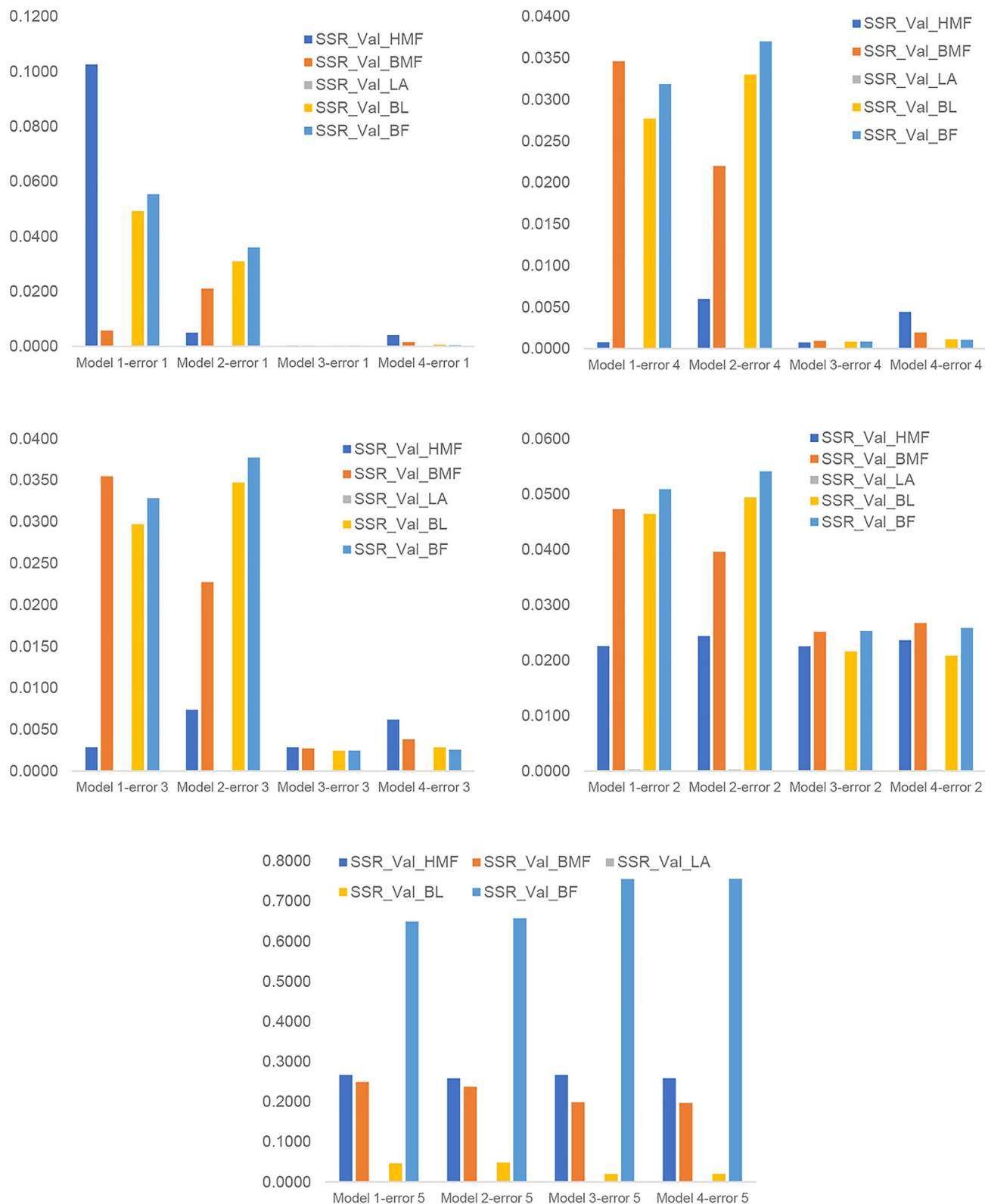


FIGURE 11  $SSR_{Val_i}$  for all models. BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSR, sum of squared residuals.

TABLE 7 Distribution of the 60 runs in the 5 folds.

	Runs		Runs		Runs		Runs		Runs
FOLD 1	5	FOLD 2	2	FOLD 3	11	FOLD 4	1	FOLD 5	6
	10		4		12		3		7
	15		9		23		8		17
	20		21		24		13		19
	22		26		29		14		28
	25		27		33		16		31
	30		32		38		18		34
	35		36		39		37		41
	44		46		40		42		43
	45		48		49		47		55
	50		52		53		51		56
	59		60		57		54		58

TABLE 8 Distribution of the 60 runs in the 10 folds.

Fold	Runs	Fold	Runs	Fold	Runs	Fold	Runs	Fold	Runs
FOLD 1	42	FOLD 2	2	FOLD 3	49	FOLD 4	8	FOLD 5	6
	35		1		46		23		21
	7		12		3		5		48
	44		17		38		47		41
	28		57		9		50		56
	53		36		18		58		24
FOLD 6	43	FOLD 7	34	FOLD 8	51	FOLD 9	52	FOLD 10	27
	40		60		59		14		55
	26		31		32		4		37
	11		16		39		29		19
	10		20		25		54		13
	33		22		45		15		30

TABLE 9 Different sets for regression and validation.

Set	Regression	Validation
Set 1	Folds 1-2-3-4	Fold 5
Set 2	Folds 5-1-2-3	Fold 4
Set 3	Folds 4-5-1-2	Fold 3
Set 4	Folds 3-4-5-1	Fold 2
Set 5	Folds 2-3-4-5	Fold 1

validation. To evaluate the capacity of prediction for a model, the  $CV_{(K)}$  number was calculated.

$$CV_{(K)} = \frac{1}{K} \cdot \sum (C_{i,\text{experimental from validation set}} - C_{i,\text{simulated}})^2_K \quad (33)$$

TABLE 10 Different sets for regression and validation for 10-fold method.

Set	Regression	Validation
Set 1	Folds 1-2-3-4-5-6-7-8-9	Fold 10
Set 2	Folds 10-1-2-3-4-5-6-7-8	Fold 9
Set 3	Folds 9-10-1-2-3-4-5-6-7	Fold 8
Set 4	Folds 8-9-10-1-2-3-4-5-6	Fold 7
Set 5	Folds 7-8-9-10-1-2-3-4-5	Fold 6
Set 6	Folds 6-7-8-9-10-1-2-3-4	Fold 5
Set 7	Folds 5-6-7-8-9-10-1-2-3	Fold 4
Set 8	Folds 4-5-6-7-8-9-10-1-2	Fold 3
Set 9	Folds 3-4-5-6-7-8-9-10-1	Fold 2
Set 10	Folds 2-3-4-5-6-7-8-9-10	Fold 1

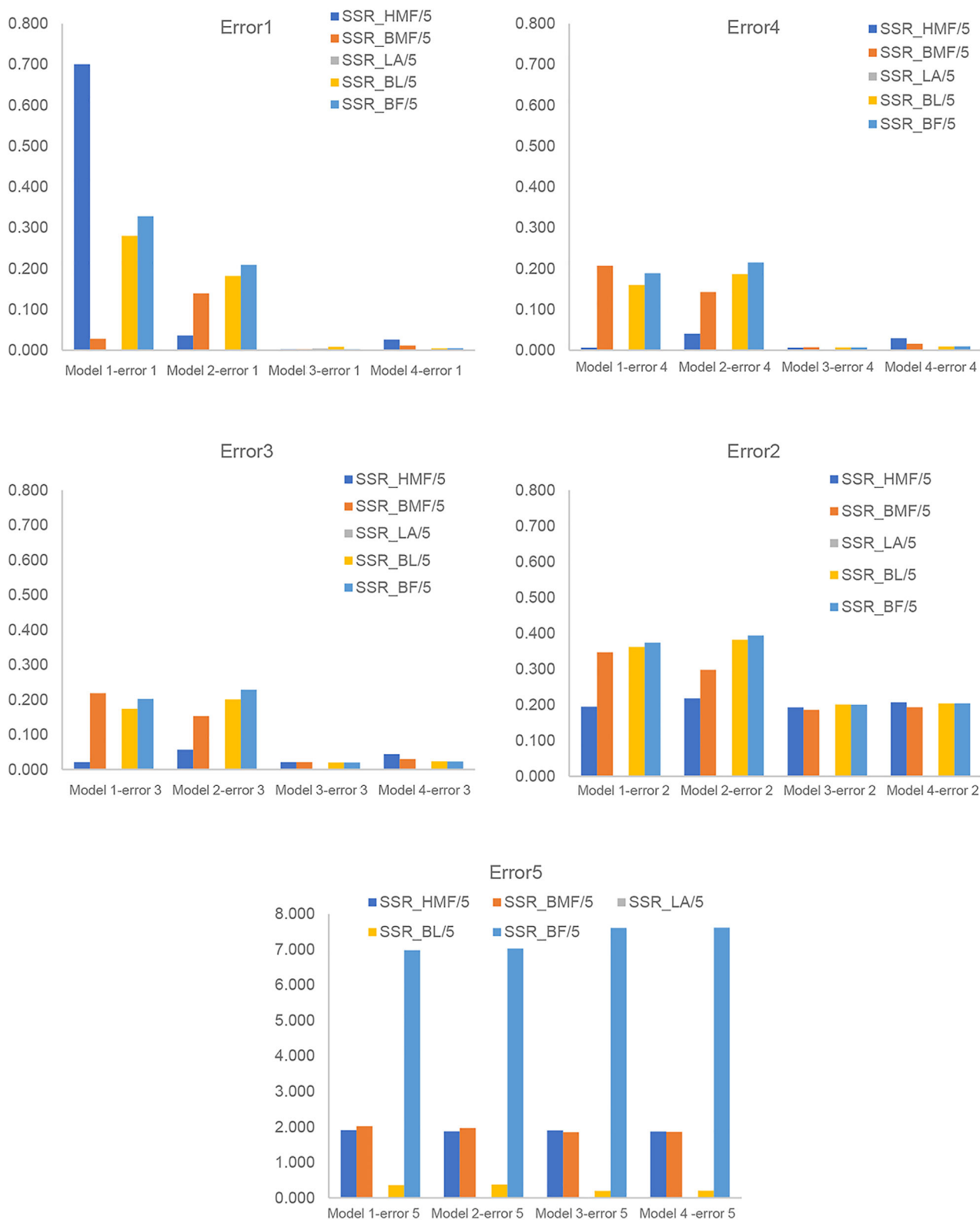


FIGURE 12  $\frac{SSR_i}{5}$  for all models using the 5-fold method (Table S10). BF, butyl formate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSR, sum of squared residuals.

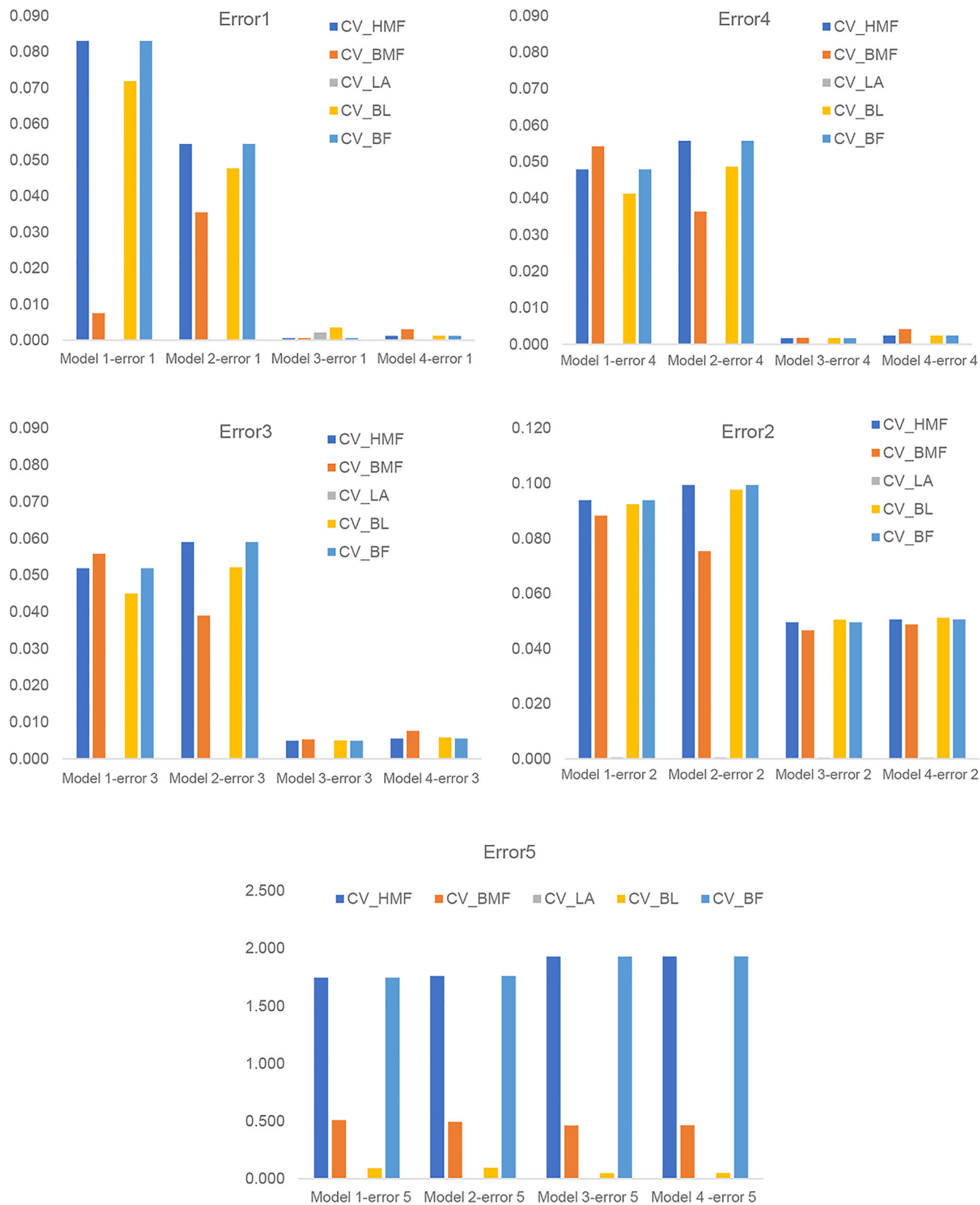


FIGURE 13 Cross-validation (CV) for 5-fold method (Table S12). BF, butyl formate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid.

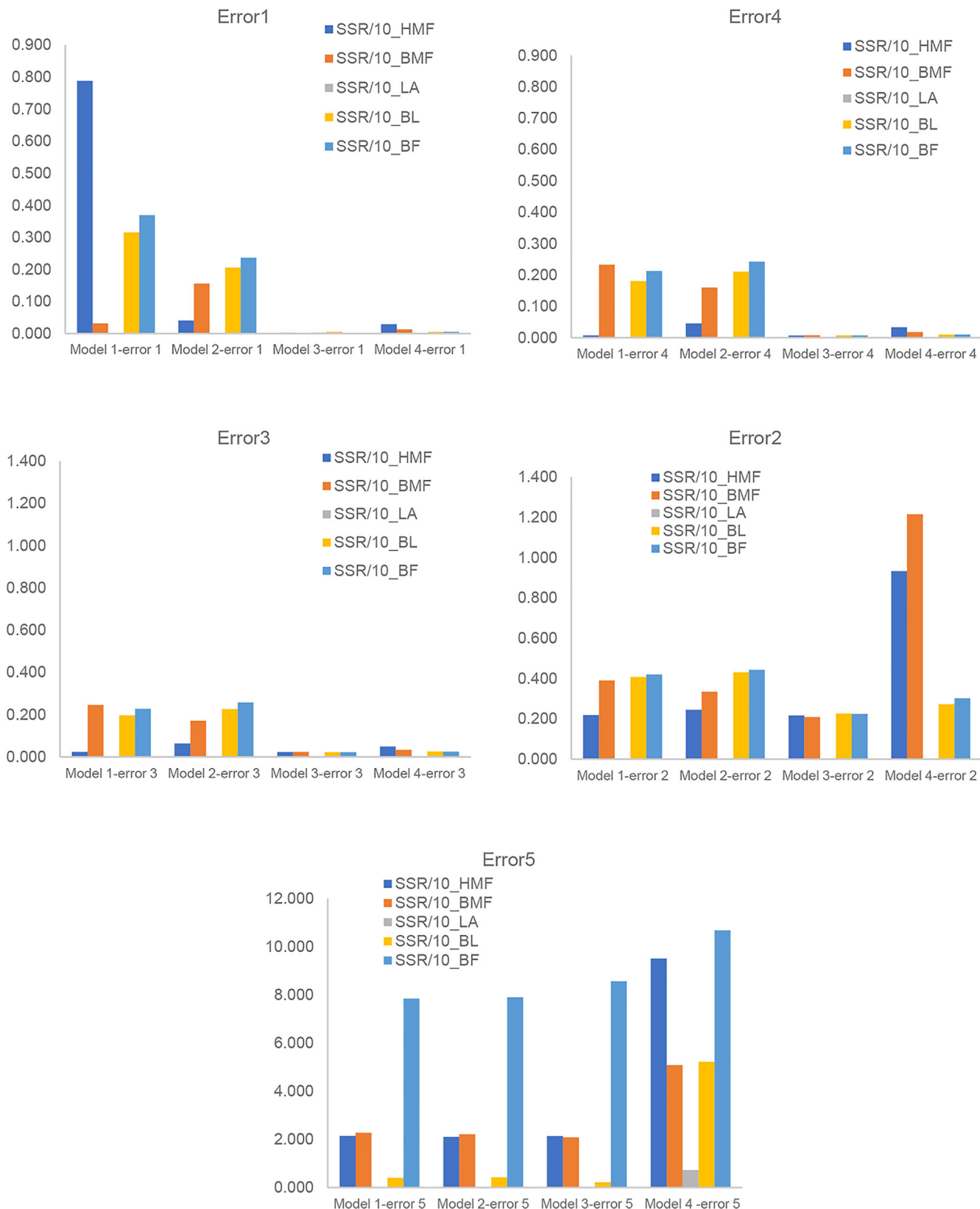
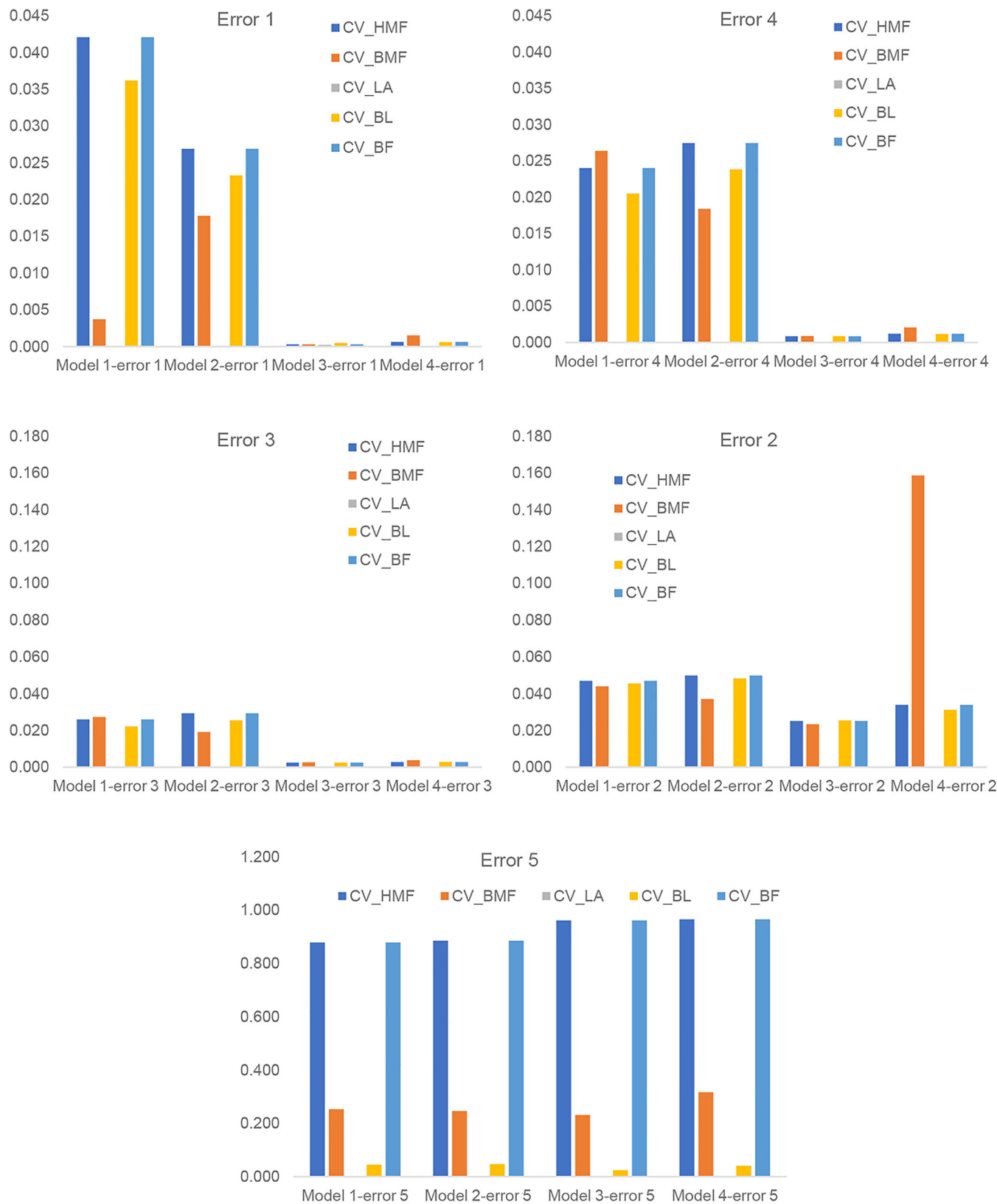


FIGURE 14  $\frac{SSR_i}{10}$  for all models for 10-fold method (Table S13). BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid; SSR, sum of squared residuals.



**FIGURE 15** Cross-validation (CV) for 10-fold method (Table S15). BF, butyl formiate; BL, butyl levulinate; BMF, butoxymethylfurfural; HMF, hydroxymethylfurfural; LA, levulinic acid.



A model with a lower  $CV_{(K)}$  has a higher prediction capacity.

From the regression, we defined a similar number as follows:

$$\frac{SSR_i}{K} = \frac{1}{K} \cdot \sum (C_{i,\text{experimental from regression set}} - C_{i,\text{simulated}})^2_K \quad (34)$$

#### 4.3.1 | 5-fold method

Figure 12 shows  $\frac{SSR_i}{5}$  for all models using 5-fold method. We can notice that this approach allows for discriminating between all Models correctly for errors 1, 4, and 3, based on  $\frac{SSR_i}{5}$ . Increasing the noise level makes this selection more challenging for errors 2 and 5.

In the Supporting Information, we included Table S11 showing the estimated kinetic constants for the different sets and all experiments from Model 3 and error 3. The column AVERAGE shows the average values from the estimated constants from the different sets, and the column SD is its standard deviation. The column All\_exp is the estimated parameter obtained from the regression using all experiments, and the column HPD is the calculated credible intervals. Even if we can notice that the values of the columns SD and HPD are similar, their meanings are different. The column HPD represents the credible intervals when all runs are considered during the regression, whereas the column SD is the standard deviation of the average value of the estimated kinetic constants obtained from the different sets.

We want to stress that using the k-fold method requires estimating all the runs in the regression stage to get the estimated kinetic constants with their credible intervals. In other words, the kinetic modeller should present the estimated constants and credible intervals displayed in Tables S3–S6. Hence, the regression stage uses all the data, and its accuracy is better than the holdout method. The k-fold aids in discriminating (via CV number) and verifying if the estimated constants are similar from the different regressions using all runs. For the latter, the standard deviation of the different estimated constants helps the modeller. Table S11 shows that the estimated constants from the different sets and all runs are similar for Model 3 with error 3.

Figure 13 shows that Model 3 is the most reliable for errors 1, 4, and 3. Nevertheless, when the degree of noise increases, that is, errors 2 and 5, the selection is less evident.

#### 4.3.2 | CV 10-fold method

Figure 14 shows the  $\frac{SSR_i}{10}$  values for the different models for the 10-fold method. Compared to the 5-fold method, we can notice that it is possible to discriminate even with error 2. Figure 15 shows that it is possible to discriminate even error 2 for CV values. The 10-fold approach is more robust than the 5-fold one. In the Supporting Information, we included Table S14 showing the estimated kinetic constants for the different sets and all experiments from Model 3 and error 3 for the 10-fold method.

## 5 | CONCLUSIONS

This manuscript proposed testing different kinetic model assessments for the solvolysis of 5-HMF into BL over Amberlite IR-120. The most reliable kinetic model, obtained from a previous study, was used to create 60 synthetic runs with different operating conditions (reaction temperature, concentrations, and catalyst loading). These 60 synthetic runs were considered to be true. On these synthetic runs, we added different levels of noise.

Four models can describe the kinetics of 5-HMF solvolysis. In the first part, the model selection was based on the regression of all 60 synthetic runs by using the AIC. In the second part, we used a holdout method by comparing 80% of synthetic runs for regression and the rest for validation and 90% of synthetic runs for regression and the rest for validation. Both holdout methods gave similar conclusions regarding model selection. The benefit of using 90% of runs for the regression is that the estimated kinetic constants are more accurate.

In the last part, we studied the k-fold CV method by comparing the 5-fold and 10-fold methods. The 10-fold method was the most robust because even when there was a high degree of noise in some data, this method found the true kinetic model.

This article shows the benefit of using the validation stage in kinetic modelling, and highlights that model assessment should be done on the regression and validation stage outputs. Kinetic modellers might not want to use the holdout method because it sacrifices a part of the run for validation. CV such as the 10-fold method is a powerful tool because all data are used for the regression. This approach suits complex kinetic systems, even with significant analytical errors. Additionally, the k-fold approach could be extended to different chemical reactors in the same way as the classical regression analysis.

## AUTHOR CONTRIBUTIONS

**Sébastien Leveueur:** Conceptualization; investigation; project administration; software; validation; writing – original draft; writing – review and editing.

## ACKNOWLEDGEMENTS

This work was done in the framework of the PROMETEE project, standing for Processes to valoRize nOrman bio-Mass from renEwable energies: ciTizen scienceE and process safety. The author thanks Métropole Rouen Normandie for the funding.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


## PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/cjce.24956>.

## DATA AVAILABILITY STATEMENT

Data available upon request from the authors.

## ORCID

Sébastien Leveueur  <https://orcid.org/0000-0001-9528-6440>

## REFERENCES

- [1] L. P. De Oliveira, D. Hudebine, D. Guillaume, J. J. Verstraete, *Oil Gas Sci. Technol.* **2016**, 71, 45.
- [2] G. M. Ostrovsky, N. N. Ziyatdinov, T. V. Lapteva, I. V. Zaitsev, *Chem. Eng. Sci.* **2012**, 83, 119.
- [3] K. Li, P. Mahapatra, K. S. Bhat, D. C. Miller, D. S. Mebane, *React. Chem. Eng.* **2017**, 2, 550.
- [4] H. S. Fogler, *Elements of Chemical Reaction Engineering*, 5th ed., Pearson, London **2016**.
- [5] J. Villiermaux, *Génie de la réaction chimique*, Tec & Doc Lavoisier, Paris **1993**.
- [6] O. Levenspiel, *Chemical Reaction Engineering Essentials, Exercises and Examples*, John Wiley & Sons, London **1998**.
- [7] T. P. De Carvalho, R. C. Catapan, A. A. M. Oliveira, D. G. Vlachos, *Ind. Eng. Chem. Res.* **2018**, 57, 10269.
- [8] J. E. Sutton, D. G. Vlachos, *Chem. Eng. Sci.* **2015**, 121, 190.
- [9] M. Saeys, M. F. Reyniers, J. W. Thybaut, M. Neurock, G. B. Marin, *J. Catal.* **2005**, 236, 129.
- [10] K. Alexopoulos, M. John, K. Van Der Borgh, V. Galvita, M. F. Reyniers, G. B. Marin, *J. Catal.* **2016**, 339, 173.
- [11] N. Morlanés, G. Lezcano, A. Yerrayya, J. Mazumder, P. Castaño, *Chem. Eng. J.* **2022**, 433, 133201.
- [12] Y. Slotboom, M. J. Bos, J. Pieper, V. Vrieswijk, B. Likozar, S. R. A. Kersten, D. W. F. Brilman, *Chem. Eng. J.* **2020**, 389, 124181.
- [13] J. Delgado, W. N. Vasquez Salcedo, G. Bronzetti, V. Casson Moreno, M. Mignot, J. Legros, C. Held, H. Grénman, S. Leveueur, *Chem. Eng. J.* **2022**, 430, 133053.
- [14] D. Di Menno Di Bucchianico, A. Cipolla, J. Buvat, M. Mignot, V. Casson Moreno, S. Leveueur, *Ind. Eng. Chem. Res.* **2022**, 61, 10818.
- [15] Z. N. Abudi, Z. Hu, A. R. Abood, *Biomass Convers. Biorefin.* **2022**, 12, 275.
- [16] F. O. de Cintra, M. Takagi, *Chem. Eng.* **2018**, 35, 1305.
- [17] R. Chauhan, V. C. Srivastava, *Chem. Eng. Sci.* **2022**, 247, 117025.
- [18] A. Brink, T. Westerlund, *Chemom. Intell. Lab. Syst.* **1995**, 29, 29.
- [19] C. J. Chuck, J. Donnelly, *Appl. Energy* **2014**, 118, 83.
- [20] D. Di Menno Di Bucchianico, J. C. Buvat, M. Mignot, V. Casson Moreno, S. Leveueur, *Fuel* **2022**, 318, 123703.
- [21] K. Alamgir Ahmad, M. Haider Siddiqui, K. K. Pant, K. D. P. Nigam, N. P. Shetti, T. M. Aminabhavi, E. Ahmad, *Chem. Eng. J.* **2022**, 447, 137550.
- [22] D. Sengupta, R. W. Pike, *Chemicals from Biomass: Integrating Bioprocesses into Chemical Production Complexes for Sustainable Development*, CRC Press, Boca Raton, FL **2012**.
- [23] B. Girisuta, K. Dussan, D. Haverty, J. J. Leahy, M. H. B. Hayes, *Chem. Eng. J.* **2013**, 217, 61.
- [24] M. F. Gustavo, E. Székely, J. Tóth, *ACS Omega* **2020**, 5, 26795.
- [25] D. C. Manheim, R. L. Detwiler, *MethodsX* **2019**, 6, 1398.
- [26] G. Guillén-Gosálbez, A. Miró, R. Alves, A. Sorribas, L. Jiménez, B. M. C. Syst, *Biologia* **2013**, 7, 1.
- [27] C. Ruckebusch, S. Aloïse, L. Blanchet, J. P. Huvenne, G. Buntinx, *Chemom. Intell. Lab. Syst.* **2008**, 91, 17.
- [28] S. Leveueur, D. Y. Murzin, T. Salmi, J.-P. Mikkola, N. Kumar, K. Eränen, L. Estel, *Chem. Eng. J.* **2009**, 147, 323.
- [29] R. L. Musante, R. J. Grau, M. A. Baltanás, *Appl. Catal., A* **2000**, 197, 165.
- [30] S. Leveueur, J. Wärnä, K. Eränen, T. Salmi, *Chem. Eng. Sci.* **2011**, 66, 1038.
- [31] S. Leveueur, C. A. De Araujo Filho, L. Estel, T. Salmi, *Ind. Eng. Chem. Res.* **2012**, 51, 189.
- [32] H. Ariba, Y. Wang, C. Devouge-Boyer, R. P. Stateva, S. Leveueur, *J. Chem. Eng. Data* **2020**, 65, 3008.
- [33] M. Caracotsios, W. E. Stewart, *Comput. Chem. Eng.* **1985**, 9, 359.
- [34] W. E. Stewart, M. Caracotsios, *Computer-Aided Modeling of Reactive Systems*, Wiley & Sons, NJ **2008**.
- [35] W. E. Stewart, M. Caracotsios. Athena Visual Studio. **2010**, [www.athenavisual.com](http://www.athenavisual.com) (accessed: June 2020).
- [36] J. Kopyscinski, J. Choi, J. M. Hill, *Appl. Catal., A* **2012**, 445–446, 50.
- [37] W. E. Stewart, M. Caracotsios, J. P. Sørensen, *AIChE J.* **1992**, 38, 641.
- [38] M. A. J. S. Van Boekel, *J. Food Sci.* **1996**, 61, 477.
- [39] G. Buzzi-Ferraris, *Catal. Today* **1999**, 52, 125.
- [40] G. Ozbuyukkaya, R. S. Parker, G. Veser, *AIChE J.* **2022**, 68, e17538.
- [41] D. Hasdemir, H. C. J. Hoefsloot, A. K. Smilde, B. M. C. Syst, *Biologia* **2015**, 9, 1.
- [42] K. A. P. McLean, S. Wu, K. B. McAuley, *Ind. Eng. Chem. Res.* **2012**, 51, 6105.
- [43] S. Guo, B. Li, W. Yu, D. I. Wilson, B. R. Young, *Can. J. Chem. Eng.* **2021**, 99, 2405.
- [44] D. M. Hawkins, S. C. Basak, D. Mills, *J. Chem. Inf. Comput. Sci.* **2003**, 43, 579.
- [45] R. Kohavi, *International Joint Conference on Artificial Intelligence* **1995**, 2, 1137.
- [46] S. C. Chen, D. M. Hayden, S. S. Young, *J. Math. Chem.* **2015**, 53, 551.

- [47] A. K. Nayak, A. Pal, *J. Mol. Liq.* **2019**, 276, 67.
- [48] S. Capecchi, Y. Wang, J. Delgado, V. Casson Moreno, M. Mignot, H. Grénman, D. Y. Murzin, S. Leveueur, *Ind. Eng. Chem. Res.* **2021**, 60, 11725.
- [49] A. Cordier, M. Klinksiek, C. Held, J. Legros, S. Leveueur, *Chem. Eng. J.* **2023**, 451, 138541.
- [50] J. H. Badía, C. Fite, R. Bringue, E. Ramirez, M. Iborra, *React. Chem. Eng.* **2021**, 6, 165.
- [51] Z. Mahmood, S. Khan, *International Journal of Biostatistics* **2009**, 5. <https://doi.org/10.2202/1557-4679.1105>.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** S. Leveueur, *Can. J. Chem. Eng.* **2023**, 1. <https://doi.org/10.1002/cjce.24956>