Bioprocess and Biosystems Engineering (2019) 42:1793–1808
https://doi.org/10.1007/s00449-019-02176-9

RESEARCH PAPER

# Sensitivity analysis and reduction of a dynamic model of a bioproduction of fructo-oligosaccharides

R. Fekih-Salem[1] · L. Dewasme[2] · Cristiana Cordeiro Castro[3] · C. Nobre[4] · A.-L. Hantson[3] · A. Vande Wouwer[2]

Received: 8 April 2019 / Accepted: 15 July 2019 / Published online: 2 August 2019


## Abstract

Starting from a relatively detailed model of a bioprocess producing fructo-oligosaccharides, a set of experimental data collected in batch and fed-batch experiments is exploited to estimate the unknown model parameters. The original model includes the growth of the fungus *Aureobasidium pullulans* which produces the enzymes responsible for the hydrolysis and transfructosylation reactions, and as such contains 25 kinetic parameters and 16 pseudo-stoichiometric coefficients, which are not uniquely identifiable with the data at hand. The aim of this study is, therefore, to show how sensitivity analysis and quantitative indicators based on the Fisher information matrix can be used to reduce the detailed model to a practically identifiable model. Parametric sensitivity analysis can indeed be used to progressively simplify the model to a representation involving 15 kinetic parameters and 8 pseudo-stoichiometric coefficients. The reduced model provides satisfactory prediction and can be convincingly cross validated.

**Keywords** Mathematical modeling · Fisher information matrix · Parameter identification · Biotechnology

## List of symbols

| | |
|---|---|
| GF | Sucrose concentration ($g\,L^{-1}$) |
| $GF_2$ | 1-Kestose concentration ($g\,L^{-1}$) |
| $GF_3$ | Nystose concentration ($g\,L^{-1}$) |
| $GF_4$ | Fructofuranosylnystose concentration ($g\,L^{-1}$) |
| F | Fructose concentration ($g\,L^{-1}$) |
| G | Glucose concentration ($g\,L^{-1}$) |
| $X$ | Biomass concentration ($g\,L^{-1}$) |
| $r_i, i = 1, \ldots, 4$ | Hydrolysis reaction rates ($g\,L^{-1}\,h^{-1}$) |
| $r_i, i = 5, 6, 7$ | Transfructosylation reaction rates ($g\,L^{-1}\,h^{-1}$) |
| $r_i, i = 8, 9$ | Biomass production reaction rates ($g\,L^{-1}\,h^{-1}$) |
| $k_i, i = 1, \ldots, 14$ | Pseudo-stoichiometric coefficients |
| $Y_F$ | Biomass yield coefficient from fructose |
| $Y_G$ | Biomass yield coefficient from glucose |
| $Vmh_{GF}$ | Maximum hydrolysis rate for sucrose ($g\,L^{-1}\,h^{-1}$) |
| $Kmh_{GF}$ | Michaelis–Menten constant for sucrose ($g\,L^{-1}$) |
| $Vmh_{GF_i}$ | Maximum hydrolysis rate for $GF_i$ ($g\,L^{-1}\,h^{-1}$) |
| $Kmh_{GF_i}$ | Michaelis–Menten constant for $GF_i$ ($g\,L^{-1}$) |
| $Kih_{GF_i}$ | Substrate inhibition constant for $GF_i$ ($g\,L^{-1}$) |
| $VmT_{GF}$ | Maximum transfructosylation rate for sucrose ($g\,L^{-1}\,h^{-1}$) |
| Kmst | Michaelis–Menten constant for sucrose ($g\,L^{-1}$) |
| Ksts | Substrate inhibition constant for sucrose ($g\,L^{-1}$) |
| Kgst | Competitive inhibition constant for glucose |
| $VmT_{GF_i}$ | Maximum transfructosylation rate for $GF_i$ ($g\,L^{-1}h^{-1}$) |
| $Kmt_{GF_i}$ | Michaelis–Menten constant for $GF_i$ ($g\,L^{-1}$) |
| $Kit_{GF_i}$ | Competitive inhibition constant for $GF_i$ |

✉ L. Dewasme
laurent.dewasme@umons.ac.be

[1] LAMSIN (LR-99-ES20), National Engineering School of Tunis, University of Tunis El Manar, 1002 Tunis, Tunisia

[2] Automatic Control Laboratory, Faculty of Engineering, University of Mons, 7000 Mons, Belgium

[3] Department of Chemical and Biochemical Engineering, Faculty of Engineering, University of Mons, 7000 Mons, Belgium

[4] Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal

| | |
|---|---|
| $\mu_{mF}$ | Maximum specific growth rate for fructose ($h^{-1}$) |
| $K_F$ | Monod constant of fructose ($g\,L^{-1}$) |
| $\mu_{mG}$ | Maximum specific growth rate for glucose ($h^{-1}$) |
| $K_G$ | Monod constant of glucose ($g\,L^{-1}$) |
| $V$ | Culture volume (L) |

## Introduction

Human health and well-being are dependent on the metabolic activity of the bacterial community present in the gastrointestinal tract. Pre-, pro-, and synbiotics can be used to control the intestinal function through modulation of microbiota composition and activity [23]. Epidemiological and clinical studies indicate that beneficial commensal bacteria, known as probiotic bacteria, may be useful to prevent cancer and to slow down the progression of tumors [2, 3]. Because of their bifidogenic nature and their health-promoting properties, fructo-oligosaccharides (FOS), are classified as prebiotics. Functional properties, as well as the technological potential of FOS, make them attractive in food and pharmaceutical applications [1, 15, 20, 21].

The main FOS include 1-kestose ($GF_2$), nystose ($GF_3$), and fructofuranosylnystose ($GF_4$). They can be found in trace amounts in fruits, vegetables, and honey [12]. Industrially, FOS can be produced from sucrose (GF) by $\beta$-fructofuranosidase enzymes with transfructosylating and hydrolytic activity, provided by fungi, such as *Aureobasidium pullulans* [6, 8, 25, 30]. To improve the production of FOS, screening and selection of carriers for immobilization of *A. pullulans* cells were studied in [4]. FOS production yields can be affected by sucrose concentration in the medium as well as small saccharides, such as glucose (G) and fructose (F) that can inhibit the fructosyltransferase enzymes and trigger FOS hydrolysis.

The maximization of the productivity of the fermentative production, as well as the minimization of the small monosaccharides in the medium, can be achieved by a tight process control. To this end, a dynamic model of the FOS production process is needed. However, only few dynamic models of FOS production are readily available. In [18], the synthesis reactions of FOS by fructosyltransferase are considered, while [9] introduces the nystose hydrolysis reaction. A fermentative process based on cultures of *A. Pullulans* is described in [24], which includes biomass growth, and 1-kestose and 1-fructosylfuranosyl nystose hydrolysis reactions. However, this model is delicate to identify and validate from experimental data, as discussed in [10].

The objective of the present study is to revisit the model of [24], and the preliminary identification results of [10], to eventually propose a reduced model that would be practically identifiable. and to develop a careful parameter identification procedure using local sensitivity analysis, the Fisher Information Matrix, and subset selection based on QR decomposition to reduce the number of parameters to a minimum. Parameter identifiability could also be assessed using global sensitivity analysis (see for instance, [11, 16, 19, 28] for overviews). However, as stated in [7], where local and global sensitivities are computed for a non-linear biological system, local sensitivity analysis is significantly easier to perform and provides a handful of information, if applied carefully.

The resulting dynamic model involves 8 stoichiometric coefficients and 15 kinetic parameters, and can be convincingly validated. The identification is made possible by the collection of informative data from several batch and fed-batch experiments.

This bottom–up approach has to be distinguished from the top–down approach presented in [27], where a dynamic model is developed from scratch, starting from the experimental database, and applying principal component analysis and parameter estimation to develop a data-driven model. Without deflowering the content of this article and the present one, it can nevertheless be concluded that the two approaches lead to models whose predictive capacity is similar. One of our main interests in the present study is to proceed along the more classical approach in biology, which consists in starting from the accumulated a priori knowledge, and to demonstrate the interest of known but probably underused techniques such as the QR decomposition of information matrices.

The paper is organized as follows. In Sect. "Materials and methods", the experimental setup and the experimental data collected in batch and fed-batch modes are presented. The model proposed in [24] is then introduced in Sect. "Modeling the fermentative FOS production". Section "Model parameter identification" presents the weighting least squares identification criterion and the quantitative indicators of parameter identifiability. Using the available experimental data, parameter identification is achieved in Sect. "Numerical results" as well as parametric sensitivity analysis based on Fisher information. An iterative parameter reduction procedure is applied and a FOS model is validated. Conclusions are drawn in Sect. "Conclusion".

## Materials and methods

### Fermentation for FOS production

An inoculum of *A. pullulans* was prepared by transferring 1 mL of spores suspension with $9 \times 10^7$ spores $mL^{-1}$ to a 500 mL Erlenmeyer flask with 100 mL of medium (100g $L^{-1}$ sucrose, 0.5g $L^{-1}$ KCl, 0.35g $L^{-1}$ $K_{2SO_4}$, 0.5g $L^{-1}$

$MgSO_4.7H_2O$, 0.01g $L^{-1}$ $FeSO_4.7H_2O$, 5g $L^{-1}$ $NaNO_3$ and 4g $L^{-1}$ $KH_2PO_4$). The inoculum was grown at 28 °C and 150 rpm and transferred after 3 days to a 5 L bioreactor—BIOSTAT® B module (Sartorius, Germany), with a working volume of 3 L of culture medium (200g $L^{-1}$ sucrose and the same salt concentrations as the ones used in the inoculum). Bioreactor fermentations were carried out at 32°C and 385 rpm with a fixed pH of 5.5.

### Sugar analysis

The sugar analysis was performed according to [5, 22]. A HPLC (Jasco) equipped with a refractive index detector working at 30 °C and a Prevail Carbohydrate ES 5u column (5 µm, 25 × 0.46 cm length × diameter) (Alltech) were used to analyze samples. The mobile phase consisted of a mixture of acetonitrile (HPLC Grade, Carlo Erba, France) in pure water (70 : 30 *v*/*v*) with 0.04% of ammonium hydroxide in water (HPLC Grade from Sigma, Germany). Samples were eluted at 1 mL $min^{-1}$ flow rate at room temperature. Chromatograms were further integrated using a Star Workstation software (Varian, USA). All the chemical standards used were of analytical grade.

### Operating conditions and data sampling

Batch cultures imply no bioreactor inlet and oulet flow rates and are, therefore, easily achieved in practice. However, the experimental data that can be collected in this operating mode usually suffers from a lack of information, particularly on the half saturation coefficients of the Michaelis–Menten laws. Fed-batch operation (where the outlet flow is still nill but the inlet flow can be manipulated) is a convenient way to overcome this lack of information issue since it allows the process to be driven to different substrate concentration levels. In our preliminary investigation [10], experimental design showed that the volumetric flow rate $Q_{in}(t)$ in the fed-batch experiments should vary as a combination of ramps and exponentials:

$$Q_{\text{in}(t)} = \begin{cases} 0 & \text{for} \quad t \in [t_0, t_1[\cup[t_2, t_3[\cup[t_4, t_f[, \\ Q_0 e^{\beta t} & \text{for} \quad t \in [t_1, t_2[, \\ Q_1 + \gamma t & \text{for} \quad t \in [t_3, t_4], \end{cases} \tag{1}$$

with, for instance, $\beta = 0.02$ and $\gamma = 2 \times 10^{-4}$. The transitions times could be defined as $t_0 = 0$ h, $t_1 = 50$ h, $t_2 = 100$ h, $t_3 = 130$ h, $t_4 = 160$ h and $t_f = 240$ h, so as to explore a relatively large operation range and generate informative data (see Fig. 1). The maximum volume is fixed at 3.5 L.

The experimental data of batch experiments (B1, B2, B3 and B4 of Table 1) and fed-batch experiments (FB1 and FB2 of Table 1) are shown in Fig. 2 and the following ones. These data sets are divided in two partitions: B1, B2, FB1, and FB2 are dedicated to the parameter identification procedure and the model direct validation while B3 and B4 are used for model cross validation.

**Table 1** Initial concentration GF(0) (g $L^{-1}$) and feed concentration $GF_{in}$ (g $L^{-1}$)

| Batch fermentation | B1 | B2 | B3 | B4 |
|---|---|---|---|---|
| GF(0) | 206.59 | 205.94 | 172.11 | 212.79 |
| Fed-batch fermentation | FB1 | FB2 | | |
| $GF(0) = GF_{in}$ | 200.63 | 102.62 | | |

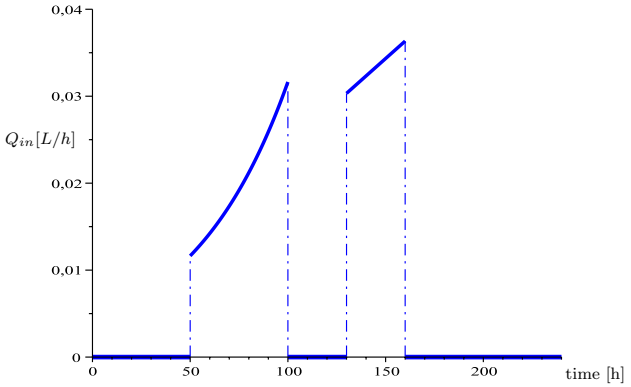


**Fig. 1** Evolution of the feed flow rate in a fed-batch experiment

The measured variables are the concentrations of sucrose GF, fructose F, glucose G, 1-kestose $GF_2$, nystose $GF_3$, and fructofuranosylnystose $GF_4$. Table 1 describes the experimental field, i.e., the several batch and fed-batch runs which differ in the initial condition of GF. Due to its heterogenous growth, fungal biomass is unfortunately not measured. This aspect is further discussed in Sect. 5. Data were collected with a varying measurement sampling time from 2 to 10 h during the first phase of sucrose transformation (up to almost 50 h). Then fewer samples were taken, e.g., every 6–16 h as the sucrose converges toward a stationary state. In fed-batch experiments, more frequent sampling was achieved (from 1 to 15 h sampling period) and a time-varying inlet flow rate $Q_{in}(t)$ (L $h^{-1}$) was used. The inlet and initial medium concentrations were the same, i.e., $GF(0) = GF_{in}$.

## Modeling of the fermentative FOS production

Based on the works of [1, 8, 24], a dynamic model of FOS bioproduction can be proposed. The model is based on a reaction scheme including enzymatic reactions and biological pseudo-reactions describing the yeast-like fungus *A. Pullulans* growth.

The enzymatic reactions are divided on the one hand, in hydrolysis reactions, representing FOS and sucrose degradation:

$$\begin{aligned}
&\mathrm{GF} \xrightarrow{r_1} k_1\mathrm{G} + k_2\mathrm{F}\\
&\mathrm{GF}_2 \xrightarrow{r_2} k_3\mathrm{GF} + k_4\mathrm{F}\\
&\mathrm{GF}_3 \xrightarrow{r_3} k_5\mathrm{GF}_2 + k_6\mathrm{F}\\
&\mathrm{GF}_4 \xrightarrow{r_4} k_7\mathrm{GF}_3 + k_8\mathrm{F}
\end{aligned}$$

where GF, G, F, and $\mathrm{GF}_i$, $i = 2, 3, 4$, denote sucrose, glucose, fructose, 1-kestose, nystose, and 1-fructofuranosyl nystose concentrations ($\mathrm{gL}^{-1}$), respectively, $r_i$, $i = 1, \ldots, 4$, represent the hydrolysis rate ($\mathrm{g\,L^{-1}\,h^{-1}}$), and $k_i$, $i = 1, \ldots, 8$ the pseudo-stoichiometric coefficients.

On the other hand, the transfructosylation reactions, which describe FOS synthesis, are given by:

$$\begin{aligned}
&2\mathrm{GF} \xrightarrow{r_5} k_9\mathrm{GF}_2 + k_{10}\mathrm{G}\\
&2\mathrm{GF}_2 \xrightarrow{r_6} k_{11}\mathrm{GF}_3 + k_{12}\mathrm{GF}\\
&2\mathrm{GF}_3 \xrightarrow{r_7} k_{13}\mathrm{GF}_4 + k_{14}\mathrm{GF}_2.
\end{aligned}$$

Biomass growth is described by:

$$\begin{aligned}
&Y_\mathrm{G}\mathrm{G} \xrightarrow{r_8} X\\
&Y_\mathrm{F}\mathrm{F} \xrightarrow{r_9} X,
\end{aligned}$$

where $X$ is the biomass concentration ($\mathrm{g\,L^{-1}}$), and $Y_\mathrm{G}$, $Y_\mathrm{F}$ represent the biomass yield coefficient from glucose and fructose, respectively.

Applying mass balancing, the following ordinary differential equation system is obtained:

$$\begin{cases}
\dot{\mathrm{GF}} &= D(\mathrm{GF_{in}} - \mathrm{GF}) - r_1 + k_3 r_2 - r_5 + \frac{k_{12}}{2} r_6\\
\dot{GF}_2 &= -D \times \mathrm{GF}_2 - r_2 + k_5 r_3 + \frac{k_9}{2} r_5 - r_6 + \frac{k_{14}}{2} r_7\\
\dot{\mathrm{GF}}_3 &= -D \times \mathrm{GF}_3 - r_3 + k_7 r_4 + \frac{k_{11}}{2} r_6 - r_7\\
\dot{\mathrm{GF}}_4 &= -D \times \mathrm{GF}_4 - r_4 + \frac{k_{13}}{2} r_7\\
\dot{\mathrm{F}} &= -D \times \mathrm{F} + k_2 r_1 + k_4 r_2 + k_6 r_3 + k_8 r_4 - Y_\mathrm{F} r_9\\
\dot{G} &= -D \times \mathrm{G} + k_1 r_1 + \frac{k_{10}}{2} r_5 - Y_\mathrm{G} r_8\\
\dot{X} &= -D.X + r_8 + r_9\\
\dot{V} &= Q_\mathrm{in} - Q_\mathrm{out},
\end{cases} \tag{2}$$

where $\mathrm{GF_{in}}$ represents the inlet sucrose concentration, $Q_{in}(t)$ and $Q_\mathrm{out}(t)$ the volumetric inlet and outlet flow rates, $V(t)$ the liquid volume inside the reactor, and $D(t) = \frac{Q_\mathrm{in}(t)}{V(t)}$ the dilution rate. Note that in fed-batch operation $Q_\mathrm{out}(t) = 0$, but a small outlet flow rate can sometimes be considered to take the sample collection into account (small volumes of liquid are sampled for off-line analysis and measurement purposes).

For convenience, the previous equations can be cast into a matrix–vector representation, introducing the pseudo-stoichiometric matrix $K$ defined by:

$$K = \begin{bmatrix}
-1 & k_3 & 0 & 0 & -1 & \frac{k_{12}}{2} & 0 & 0 & 0\\
0 & -1 & k_5 & 0 & \frac{k_9}{2} & -1 & \frac{k_{14}}{2} & 0 & 0\\
0 & 0 & -1 & k_7 & 0 & \frac{k_{11}}{2} & -1 & 0 & 0\\
0 & 0 & 0 & -1 & 0 & 0 & \frac{k_{13}}{2} & 0 & 0\\
k_2 & k_4 & k_6 & k_8 & 0 & 0 & 0 & -Y_F & 0\\
k_1 & 0 & 0 & 0 & \frac{k_{10}}{2} & 0 & 0 & 0 & -Y_G\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}. \tag{3}$$

as well as the state, rate, and transport vectors:

$$\underline{\xi}^T(t) = \left[\xi_1, \xi_2, \ldots, \xi_7\right] = \left[\mathrm{GF}, \mathrm{GF}_2, \mathrm{GF}_3, \mathrm{GF}_4, \mathrm{F}, \mathrm{G}, X\right], \tag{4}$$

$$\underline{r}^T\left(\underline{\xi}\right) = \left[r_1, r_2, \ldots, r_9\right], \quad \underline{\mathfrak{F}}^T(t) = \left[D(t)\mathrm{GF_{in}}, 0, \ldots, 0\right]. \tag{5}$$

With this notation, the mass balance model (2) can be rewritten as:

$$\begin{cases}
\dot{\underline{\xi}}(t) &= \mathrm{K}\underline{r}\left(\underline{\xi}\right) - D(t)\underline{\xi} + \underline{\mathfrak{F}}(t)\\
\dot{\underline{V}}(t) &= Q_\mathrm{in} - Q_\mathrm{out}.
\end{cases} \tag{6}$$

A candidate structure for the sucrose hydrolysis reaction rate is the following Michaelis–Menten law [8, 24]:

$$r_1 = \frac{\mathrm{Vmh_{GF}GF}}{\mathrm{Kmh_{GF}} + \mathrm{GF}}, \tag{7}$$

where $\mathrm{Vmh_{GF}}$ represents the maximum hydrolysis rate ($\mathrm{g\,L^{-1}\,h^{-1}}$) and $\mathrm{Kmh_{GF}}$ is the Michaelis–Menten half-saturation constant ($\mathrm{g\,L^{-1}}$).

The FOS hydrolysis rates are given by the modified Michaelis–Menten laws describing substrate inhibition

$$r_i = \frac{\mathrm{Vmh_{GF_i}GF}_i}{\mathrm{Kmh_{GF_i}} + \mathrm{GF}_i + \frac{\mathrm{GF}_i^2}{\mathrm{Kih_{GF_i}}}}, \quad i = 2, 3, 4, \tag{8}$$

where $\mathrm{Vmh_{GF_i}}$ represents the maximum hydrolysis rate ($\mathrm{g\,L^{-1}\,h^{-1}}$), $\mathrm{Kmh_{GF_i}}$ the Michaelis–Menten constant ($\mathrm{g\,L^{-1}}$), and $\mathrm{Kih_{GF_i}}$ is the substrate inhibition constant ($\mathrm{g\,L^{-1}}$).

The sucrose transfructosylation kinetic rate is given by a modified Michaelis

–Menten law describing the substrate inhibition and competitive glucose inhibition

$$r_5 = \frac{\mathrm{VmT_{GF}GF}}{\mathrm{Kmst} + \mathrm{GF} + \frac{\mathrm{GF}^2}{\mathrm{Ksts}} + \frac{\mathrm{G}}{\mathrm{Kgst}}}, \tag{9}$$

where $\mathrm{VmT_{GF}}$ is the maximum transfructosylation rate ($\mathrm{g\,L^{-1}\,h^{-1}}$), Kmst the Michaelis–Menten constant ($\mathrm{g\,L^{-1}}$), Ksts the substrate inhibition constant ($\mathrm{g\,L^{-1}}$), and Kgst is the competitive inhibition constant ($\mathrm{g\,L^{-1}}$) for glucose.

The modified Michaelis–Menten laws with competitive glucose inhibition for 1-kestose and nystose are given by:

$$r_j = \frac{\mathrm{VmT_{GF_i}}\,\mathrm{GF}_i}{\mathrm{Kmt_{GF_i}} + \mathrm{GF}_i + \frac{\mathrm{G}}{\mathrm{Kit_{GF_i}}}} \quad j = 6,7;\ \ i = 2,3, \tag{10}$$

where $\mathrm{VmT_{GF_i}}$ is the maximum transfructosylation rate ($\mathrm{g\,L^{-1}\,h^{-1}}$), $\mathrm{Kmt_{GF_i}}$ the Michaelis–Menten constant ($\mathrm{g\,L^{-1}}$), and $\mathrm{Kit_{GF_i}}$ is the competitive inhibition constant ($\mathrm{g\,L^{-1}}$) for glucose.

The Monod laws are given by:

$$r_j = \frac{\mu_{mj} S_j X}{S_j + K_{S_j}} \quad j = 8,9 \tag{11}$$

where $S_j$ ($j = 8,9$) are the fructose and glucose concentrations ($\mathrm{g\,L^{-1}}$), respectively, $\mu_{mj}$ the maximum specific growth rate, and $\mathrm{K}_{S_j}$ is the half-saturation constant ($\mathrm{g\,L^{-1}}$).

The resulting model contains 16 pseudo-stoichiometric coefficients and 25 kinetic parameters. These parameters have to be calibrated to fit actual experimental observations. Since fungi grow in a heterogeneous form, it is difficult to measure biomass and in the following, it is, therefore, assumed that this information is not available.

## Model parameter identification

### Weighted least squares

The parameter set $\underline{\theta}$, i.e., 25 kinetic parameters, 16 pseudo-stoichiometric coefficients and possibly the initial concentrations (initial conditions of the mass balance equations) are identified based on the available data sets (batch and fed-batch experiments) using the parameter values of [22] as initial guesses (see Table 3). The identification problem is formulated as the minimization of a weighted least squares (WLS) criterion measuring the distance between experimental measurements, represented by the vector $\underline{y}_j^{\exp}$, and model prediction, represented by the vector $\underline{y}_j^{\mathrm{sim}}$:

$$J\left(\underline{\theta}\right) = \sum_{j=1}^{N_{\exp}} \sum_{i=1}^{Nt_j} \left[\underline{y}_j^{\exp}\left(t_{ij}\right) - \underline{y}_j^{\mathrm{sim}}\left(t_{ij};\underline{\theta}\right)\right]^T W_j^{-1} \left[\underline{y}_j^{\exp}\left(t_{ij}\right) - \underline{y}_j^{\mathrm{sim}}\left(t_{ij};\underline{\theta}\right)\right] \tag{12}$$

where

$$\underline{\theta} = \left[\underline{\theta}_r^T \quad \underline{\theta}_K^T \quad \underline{\xi}_1^T(0) \cdots \underline{\xi}_{N_{\exp}}^T(0)\right];$$

$\underline{\theta}_r$ is the vector of kinetic parameters $\left(\dim\left(\underline{\theta}_r\right) = 25\right)$, $\underline{\theta}_K$ is the vector of stoichiometric parameters $\left(\dim\left(\underline{\theta}_K\right) = 16\right)$, $\underline{\xi}_j(0)$, $j = 1,\ldots,N_{\exp}$, are the initial conditions of the $j$th experiment, $t_{ij}$ denotes the $i$th sampling time of the $j$th experiment, $N_{exp}$ denotes the number of experiments, $N_{y_j}$ the number of measured states during the $j$th experiment and $Nt_j$ represents the number of samples taken during the $j$th experiment. A scaling matrix

$$W_j = \mathrm{diag}\left(\max_{1\le i\le Nt_j}\left(y_{j1}^{\exp}\left(t_{ij}\right)\right)^2, \ldots, \max_{1\le i\le Nt_j}\left(y_{jN_{y_j}}^{\exp}\left(t_{ij}\right)\right)^2\right)$$

is used to normalize the distances calculated in (12) with respect to the magnitude of each signal.

The Nelder–Mead method as implemented in the MATLAB function *fminsearchbnd* is used to minimize the WLS criterion. This version of the algorithm allows setting bounds restricting the search space to positive parameters and initial concentrations. Typically, the algorithm is called several times in a row, with re-initialization with the previously found optimum. These successive calls improve the performance of the Nelder–Mead method and allow a better convergence toward the sought minimum (see for instance [32]).

The optimization runs imply a large number of simulations of the model ordinary differential equations and the MATLAB solver *ode23tb* is selected to enhance the simulation speed and accuracy.

A posterior estimate of the measurement error variance of the $j$th experiment $\hat{\sigma}_j^2$ can be obtained by:

$$\hat{\sigma}_j^2 = \frac{J_j\left(\hat{\underline{\theta}}\right)}{N_{y_j} Nt_j - N_p}, \quad j = 1,\ldots,N_{\exp},$$

where $N_p = \dim\left(\underline{\theta}_r\right) + \dim\left(\underline{\theta}_K\right)$ and the WLS estimator is

$$\hat{\underline{\theta}} = \arg\min_{\underline{\theta}} J\left(\underline{\theta}\right).$$

### Parameter sensitivity analysis

Let $S_\theta\left(t;\underline{\theta}\right)$ denote the local sensitivity matrix of the state vector $\underline{\xi}$ to the parameter $\underline{\theta}$ defined by

$$S_\theta\left(t;\underline{\theta}\right) = \frac{\partial \underline{\xi}\left(t;\underline{\theta}\right)}{\partial \underline{\theta}}$$

Under the assumption of Gaussian white noise, the Fisher Information Matrix is determined by the following equation [29]

$$\mathrm{FIM}\left(\underline{\theta}, \hat{\Sigma}\right) = \sum_{j=1}^{N_{\exp}} \sum_{i=1}^{Nt_j} S_\theta^T\left(t_{ij};\underline{\theta}\right) \hat{\Sigma}_j^{-1} S_\theta\left(t_{ij};\underline{\theta}\right)$$

where $\hat{\Sigma}_j$ is the a posteriori covariance matrix of the measurement errors of experiment $j$, calculated as:

$$\hat{\Sigma}_j = \hat{\sigma}_j^2 W_j \tag{13}$$

The inverse of the FIM provides an optimistic estimate of the parameter error covariance matrix (Cramer–Rao bound) [29]

$$C_{\underline{\theta}} \succ \text{FIM}^{-1}\left(\hat{\underline{\theta}}, \hat{\Sigma}\right). \tag{14}$$

The main limitation of the FIM is that it is based on a model linearization. The propagation of so-called sigma points through the non-linear model usually allows a better estimation of the covariance matrix [26]. Alternatively, Monte Carlo simulation and the bootstrap method [17] allow the consideration of more general error distributions (non-Gaussian, non-symmetric, etc). FIM is, however, a simple and practical tool that we will exploit in the sequel of this study.

The standard deviation $\sigma_k$ of the parameter estimate $\hat{\theta}_k$ can be obtained from the square root of the $k$th diagonal element of $C_{\underline{\theta}}$

$$\sigma_k = \sqrt{C_{\underline{\theta}_{kk}}}, \quad k = 1, \ldots, N_\text{p}$$

It is possible to estimate the confidence intervals for a given confidence level and, in this work, a 95% confidence level (under the assumption of a Gaussian distribution) is considered:

$$\left[\hat{\theta}_k - 2\sigma_k, \hat{\theta}_k + 2\sigma_k\right], \quad k = 1, \ldots, N_\text{p}. \tag{15}$$

Alternatively, the variation coefficients of the estimated parameters can also be considered (i.e., the standard deviation $\sigma_k$ normalized by the respective estimated parameter value $\hat{\theta}_k$). The correlation matrix, which is generically defined by the term as:

$$\rho_{ij} = \sqrt{\frac{C_{\underline{\theta}_{ij}}}{C_{\underline{\theta}_{ii}} C_{\underline{\theta}_{jj}}}}, \quad i,j = 1, \ldots, N_\text{p} \tag{16}$$

can help to explain parameter inter-dependency. However, high parameter correlation is not a rejection criterion since it does not necessarily induce over-parametrization.

A frequent problem in identification studies is that the FIM can be ill-conditioned due to a lack of informative experimental data or an over-parametrization of the model. To assess the non-singularity of the FIM, the LAPACK reciprocal condition estimator (which is given by the method *rcond* in MATLAB) should at least satisfy the following [31]:

$$\text{rcond(FIM)} > 10\epsilon, \tag{17}$$

where $\epsilon$ is the floating point relative accuracy.

To investigate singularity issues, a deeper study of the sensitivities is required to detect the parameters which are possibly linearly dependent.

This study should be based on the normalized sensitivities with respect to the $L$ outputs and $P$ parameters:

$$S_{i,k}(\underline{\theta}, t) = \frac{\theta_k}{y_i} \frac{\partial y_i}{\partial \theta_k}(t) \tag{18}$$

and a matrix summarizing the sensitivity information can be constructed

$$S(\underline{\theta}, t) = \begin{bmatrix} S_{1,1}(\underline{\theta}, t_{1,1}) & \cdots & S_{1,P}(\underline{\theta}, t_{1,1}) \\ S_{1,1}(\underline{\theta}, t_{2,1}) & \cdots & S_{1,P}(\underline{\theta}, t_{2,1}) \\ \vdots & & \\ S_{1,1}(\underline{\theta}, t_{Nt_1,1}) & \cdots & S_{1,P}(\underline{\theta}, t_{Nt_1,1}) \\ S_{1,1}(\underline{\theta}, t_{1,2}) & \cdots & S_{1,P}(\underline{\theta}, t_{1,2}) \\ \vdots & & \\ S_{1,1}(\underline{\theta}, t_{Nt_{N_\text{exp}},N_\text{exp}}) & \cdots & S_{1,P}(\underline{\theta}, t_{Nt_{N_\text{exp}},N_\text{exp}}) \\ S_{2,1}(\underline{\theta}, t_{1,1}) & \cdots & S_{2,P}(\underline{\theta}, t_{1,1}) \\ \vdots & \vdots & \\ S_{L,1}(\underline{\theta}, t_{Nt_{N_\text{exp}},N_\text{exp}}) & \cdots & S_{L,P}(\underline{\theta}, t_{Nt_{N_\text{exp}},N_\text{exp}}) \end{bmatrix} \tag{19}$$

A first simple quantitative indicator consists of calculating the L2 norm of each column of $S(\underline{\theta}, t)$ and classifying the parameters from the most sensitive to the less sensitive (according to the norm criterion). This indicator is useful, but not sufficient as it does not give any information on the possible dependencies between the parameters.

A second quantitative indicator is obtained by calculating the numerical rank of the matrix $S^T S$. If the singular values are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, the $\epsilon$–numerical rank could be defined as

$$r_\epsilon = \min\left\{r : \sigma_r \leq \epsilon\right\}. \tag{20}$$

Usually (when deciding whether or not a given matrix is numerically rank-deficient or not), $\epsilon \approx u\sigma_1$, where $u$ is the square root of the machine precision epsilon. If the numerical rank $r$ is smaller than the number of parameters to estimate, this indicates that parameter dependencies do exist. The full matrix cannot be used, and a reduced set of $r$ parameters should be considered.

A third tool, called subset selection can be used at this stage. There exist several powerful algorithms, but we restrict ourselves to QR decomposition [13] to give a ranking of the parameter dependencies. Indeed, the QR decomposition algorithm allows, following the introduction of a specific permutation matrix $P$, to order the parameters from least to most linearly independent following:

$$S^T S P = QR \tag{21}$$

where:

$$P^T = \left[1 \cdots N_p\right]^T, \tag{22}$$

In practice, this operation can be easily implemented in MATLAB using the function $[Q, R, P] = qr(S^T S)$.

## Numerical results

In [10], a preliminary study of the identification of model (2) is proposed, mostly based on simulation data, as well as some experimental data in batch cultures. The motivation behind the current study is to revisit the identification problem using the systematic tools presented in the previous section and additional experimental data sets. Indeed, the simulation study reveals that a few batch experiments are not sufficient to estimate all the model parameters. The additional experiments, which are carried out in fed-batch mode, will therefore provide part of the missing information, while the sensitivity analysis and subset selection will help in reducing the model parametrization.

Note the following facts:

– In practice, not only the 25 kinetic parameters have to be considered as in [10, 22], but also the 16 pseudo-stoichiometric coefficients and possibly the initial concentrations which are measured and are, therefore, uncertain. This leads to an identification problem with 41 (48 with the initial conditions) parameters for one batch culture, and up to 69 parameters for a combination of 2 experiments in batch and 2 in fed-batch.

To illustrate the procedure in a systematic way, the data from only one batch culture (fermentation B1) is considered first for parameter identification of model (2) where the initial condition of sucrose concentration is provided in Table 1. The direct validation results are shown in Figs. 2 and 3, and are quite satisfactory. However, the FIM is ill-conditioned with rcond(FIM) $= 3 \times 10^{-53}$, that is, condition (17) is not satisfied and rank(FIM) $= 35$ while dim ( FIM ) $= 41$ (see Table 2). As such, the model is, therefore, not identifiable with the data of only one batch experiment.

To provide more information, several combinations of the experiments B1, B2, FB1, FB2 are selected, including one batch (1B), one fed-batch (1FB),and two experiments in batch and two in fed-batch (2B–2FB). The several experiments convey different information as they were carried out with different initial conditions of sucrose concentration (see Table 1). The remaining experiments (fermentations B3 and B4) are not used for identification, but for cross validation (i.e., test of the model prediction against unused data).

Figures 2 and 3 show that model (2) reproduces quite well the experimental data in direct validation. For the experiment combinations 1FB and 2B–2FB, rcond(FIM) has been significantly increased (see Table 2). Nevertheless, the values of some parameters and their associated confidence intervals can vary widely depending on the identification data base. Logically, the richest database 2B–2FB leads to more reasonable values, but still does not allow to capture all the parameters reliably (see Table 3). This is logical since the FIM is still singular (since rank(FIM) = 38 while dim ( FIM ) = 41 when not considering the initial conditions at first in the unknown paramter sets—see Table 2). So, the additional data are not enough to ensure the identifiability of the full model.

It can, therefore, be useful to turn to the sensitivity analysis and subset selection tools introduced in the previous section. Considering $u$ about $10^{-8}$ in (20), it follows that $\epsilon$−numerical rank(S) = 39 and $\epsilon$−numerical rank($S^T S$) = 35 for the identification of model (2) with the experiment combination 2B–2FB. Using the subset selection based on the QR decomposition (21), the parameters with the lowest ordering in Table 4 could be questioned.

Looking at the complete identification results in Table 3 and the correlation matrix in Fig. 4 can also be a good guide in selecting the parameters that could be eliminated from the problem. In particular, high-variance parameters can be removed when they have a high degree of correlation with other parameters [14]. Using these indicators, it seems legitimate to perform the following simplifications:

**Table 2** Root-mean-square error (RMSE) and singularity according different models and culture modes

| Model | RMSE | dim $(\underline{\theta}_r)$ | dim $(\underline{\theta}_K)$ | Rank (FIM) | rcond (FIM) | FIM |
|---|---|---|---|---|---|---|
| (2)-1B | 5.20 | 25 | 16 | 35 | $3\times10^{-53}$ | Singular |
| (2)-1FB | 3.94 | 25 | 16 | 36 | $7\times10^{-25}$ | Singular |
| (2)-2B-2FB | 6.23 | 25 | 16 | 38 | $6\times10^{-21}$ | Singular |
| (23)-2B-2FB | 6.70 | 20 | 11 | 31 | $2\times10^{-10}$ | Regular |
| (25)-2B-2FB | 7.68 | 20 | 10 | 30 | $10^{-12}$ | Regular |
| (25-26)-2B-2FB | 6.72 | 17 | 10 | 27 | $6\times10^{-13}$ | Regular |
| (27–29)-2B-2FB | 7 | 15 | 8 | 23 | $2\times10^{-12}$ | Regular |

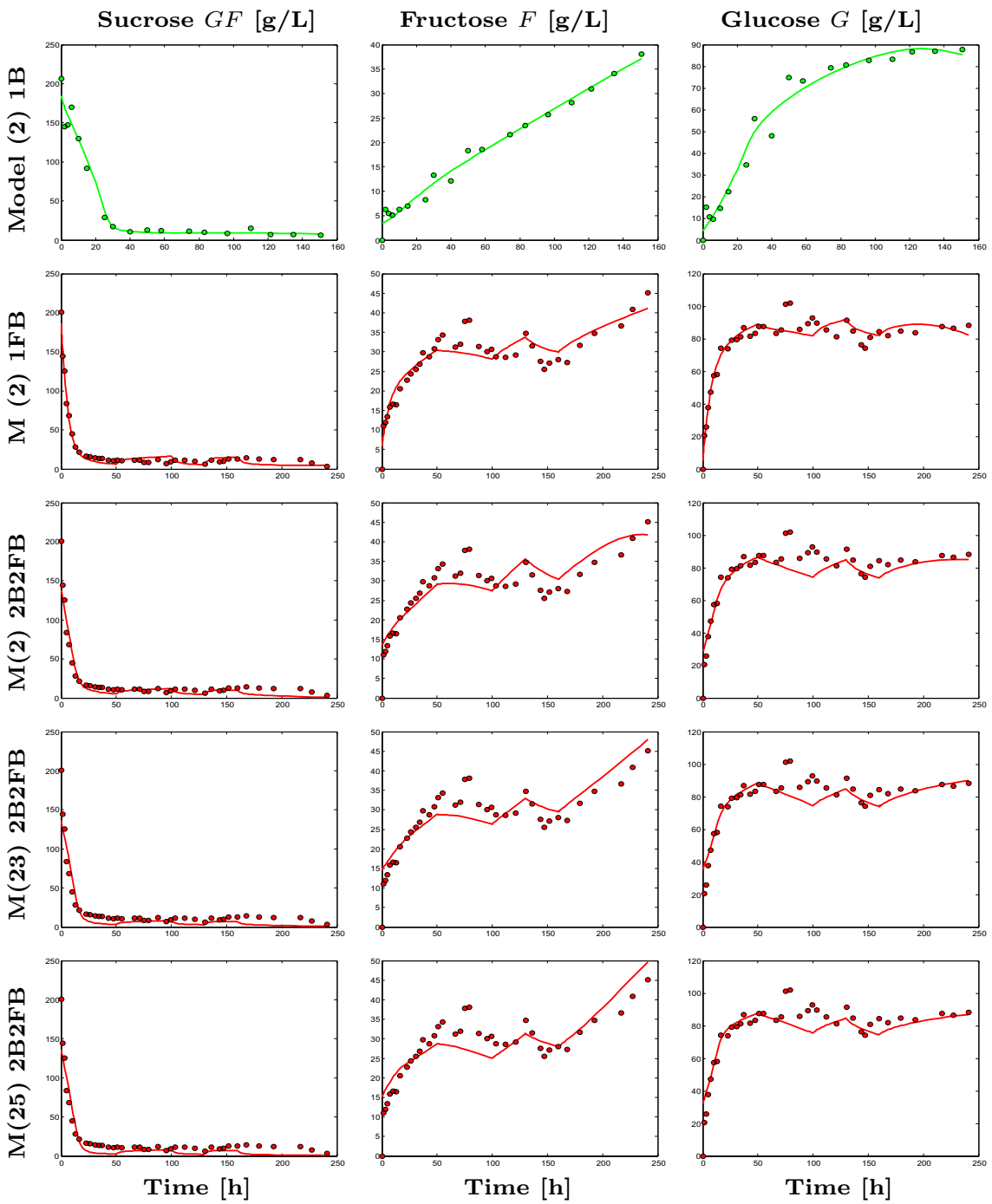


**Fig. 2** Sucrose, fructose, and glucose concentrations: measurements of fermentations B1 (green dots) and FB1 (red dots) and prediction of different models. Model simulations are depicted with continuous lines while discrete time experimental data are represented with dots

– do not consider the fungus growth pseudo-reactions $r_8$ and $r_9$, thus eliminating $Y_F$, $Y_G$, KmG, KmF and $\mu_{mG}$;
– neglect $k_1$ and $k_8$, since they assume very small values;
– neglect the term $G/Kit_{GF_3}$ in reaction rate (10) as the competitive inhibition constant $Kit_{GF_3}$ is quite large.

It must be stressed that eliminating the selected parameters does not induce that the corresponding physical mechanisms do not exist. Following an optimal experiment design, based on the proposed model structure (2), and therefore generating adapted new data sets would certainly alleviate, at least, part

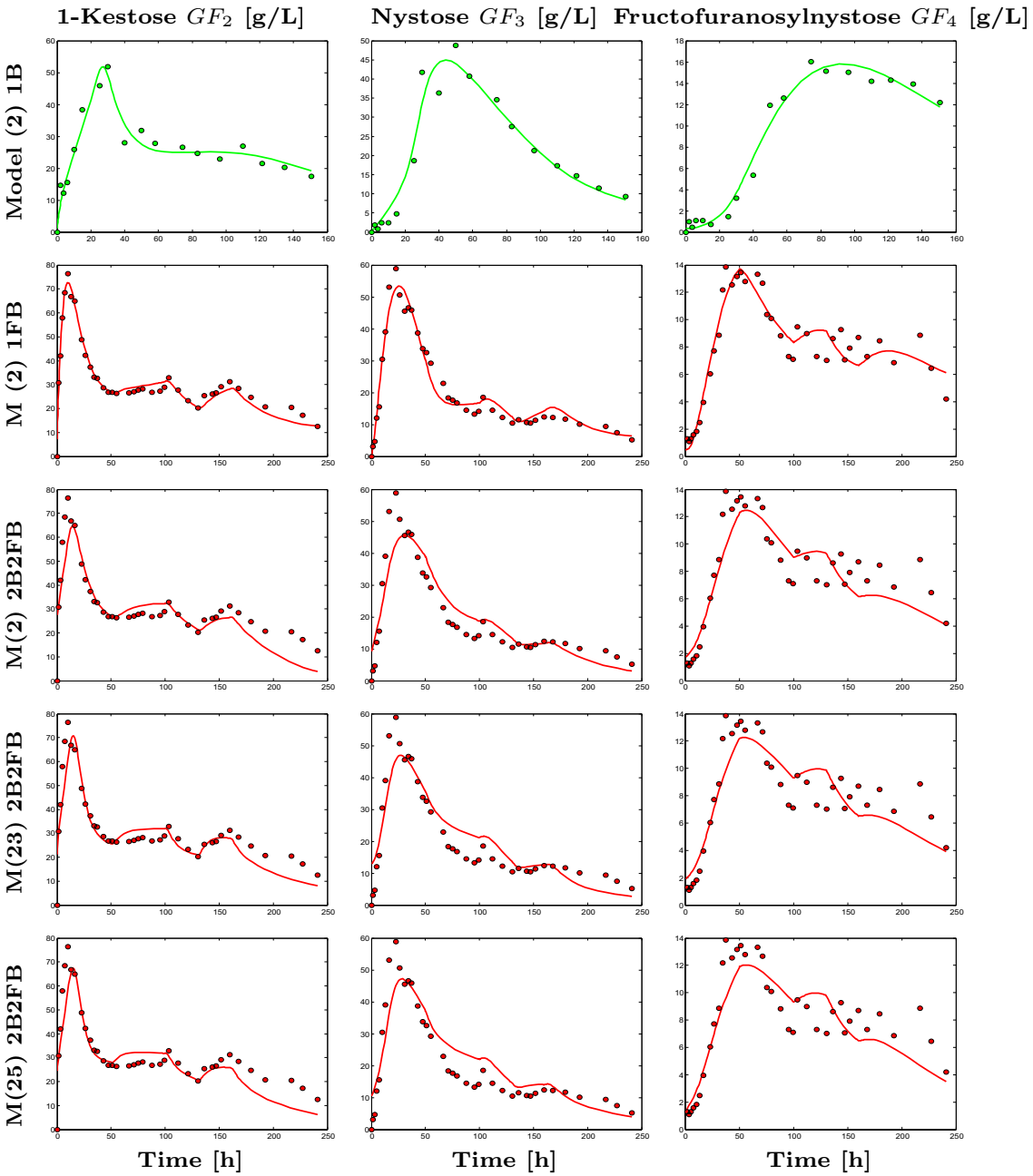


**Fig. 3** FOS concentrations: measurements of fermentations B1 (green dots) and FB1 (red dots) and prediction of different models. Model simulations are depicted with continuous lines while discrete time experimental data are represented with dots

of this parameter identifiability issue. The above reasoning is, therefore, limited to the available information since no other data set may currently be obtained. Fixing the non-identifiable parameter values using, for instance, knowledge from the literature may also introduce a strong degree of uncertainty and severe bias on the identified parameters. This last solution is, therefore, rejected.

The differential equations of the reduced model can be rewritten as follows:

$$\begin{cases} \dot{\mathrm{GF}} &= D(\mathrm{GF_{in}} - \mathrm{GF}) - r_1 - r_5 + \frac{k_{12}}{2} r_6 \\ \mathrm{GF_2} &= -D \times \mathrm{GF_2} - r_2 + k_5 r_3 + \frac{k_9}{2} r_5 - r_6 + \frac{k_{14}}{2} r_7 \\ \mathrm{GF_3} &= -D \times \mathrm{GF_3} - r_3 + k_7 r_4 + \frac{k_{11}}{2} r_6 - r_7 \\ \mathrm{GF_4} &= -D \times \mathrm{GF_4} - r_4 + \frac{k_{13}}{2} r_7 \\ \mathrm{F} &= -D \times \mathrm{F} + k_2 r_1 + k_4 r_2 + k_6 r_3 \\ \mathrm{G} &= -D \times \mathrm{G} + \frac{k_{10}}{2} r_5 \\ \dot{V} &= Q_{\mathrm{in}} - Q_{\mathrm{out}}. \end{cases} \tag{23}$$

**Table 3** Parameter identification results (with 95% confidence intervals) of model (2) considering different experiment combinations for parameter identification

| No. | Parameter | Literature[a] | Model (2) 1 B | Model (2) 1 FB | Model (2) 2B-2FB |
|---|---|---|---|---|---|
| 1 | VmhGF | 1.43 | 0.68 ±6.6 | 9.2±55 | 10± 3.3 |
| 2 | $VmhGF_2$ | 7.58 | 6.2 ± 5.4 | 0.32 ±319 | 28.9± 4.7 |
| 3 | $VmhGF_3$ | 7.97 | 12 ± 9 | 13.8 ±8.4 | 22.8 ± 0.6 |
| 4 | $VmhGF_4$ | 7.35 | 7.6 ± 36 | 8.9 ± 2.9 | 8.5 ± 2.2 |
| 5 | VmtGF | 49.99 | 85 ± 45 | 67 ± 45 | 440 ± 118 |
| 6 | $VmtGF_2$ | 41.63 | 205 ± 86 | 57.2 ± 28.7 | 217 ± 14 |
| 7 | $VmtGF_3$ | 11.53 | 11 ± 71 | 6.1 ± 3.9 | 8.8 ± 4.6 |
| 8 | $\mu_{mF}$ | 0.0097 | 2 ± 89 | 0.07 ± 1.3 | 0.1 ± 0.3 |
| 9 | $\mu_{mG}$ | $2.89\times10^{-5}$ | $5\times10^{-3}\pm17$ | $9\times10^{-3}\pm1$ | 0.09 ± 2.7 |
| 10 | KmhGF | 111.57 | $136\pm3\times10^{3}$ | 611± 697 | $10^{3}\pm456$ |
| 11 | $KmhGF_2$ | 0.61 | $136\pm3\times10^{3}$ | $8\times10^{-12}\pm705$ | 250 ± 244 |
| 12 | $KmhGF_3$ | 177.41 | 20.5 ± 52 | 79 ± 68 | 85 ± 8 |
| 13 | $KmhGF_4$ | 724.07 | $468\pm10^{3}$ | 409 ± 391 | 484 ± 69 |
| 14 | Kmst | 70.22 | 90 ±254 | 269 ± 77 | 210 ± 101 |
| 15 | $KmtGF_2$ | 239.88 | $10^{3}\pm10^{3}$ | 405 ± 148 | 627 ± 84 |
| 16 | $KmtGF_3$ | 333.07 | $384\pm10^{3}$ | 87 ± 146 | 313 ± 39 |
| 17 | KmF | 11.45 | $8\times10^{3}\pm3\times10^{16}$ | $238\pm3\times10^{8}$ | 496 ± 357 |
| 18 | KmG | 397.98 | $7\times10^{3}\pm5\times10^{5}$ | 461 ± 3239 | $10^{4}\pm10^{4}$ |
| 19 | $KihGF_2$ | 2.72 | $10^{3}\pm10^{5}$ | 0.3± 133 | 0.24± 0.07 |
| 20 | $KihGF_3$ | 10.52 | 22 ± 43 | 15 ± 14.8 | 8.3 ± 0.8 |
| 21 | $KihGF_4$ | 6.21 | $10^{11}\pm4\times10^{22}$ | $403.8\pm4\times10^{5}$ | 2.9 ± 5.9 |
| 22 | Ksts | 911.16 | 23 ± 9.5 | $406\pm2\times10^{3}$ | 3.1 ± 0.9 |
| 23 | Kgst | 24.57 | 15 ± 182 | $10^{4}\pm10^{8}$ | 0.06 ± 0.02 |
| 24 | $KitGF_2$ | 49.96 | $10^{6}\pm2\times10^{13}$ | 4.85 ± 91 | $0.08\pm5\times10^{-3}$ |
| 25 | $KitGF_3$ | 49.95 | 2.7 ± 168 | 32 ± 851 | $2\times10^{3}\pm3\times10^{6}$ |
| 26 | $Y_F$ | 79.34 | $4\times10^{-8}\pm10^{5}$ | $6\times10^{-4}\pm5\times10^{3}$ | 129 ± 272 |
| 27 | $Y_G$ | 29.23 | $7\times10^{3}\pm2\times10^{7}$ | $3\times10^{3}\pm3\times10^{5}$ | $416\pm10^{4}$ |
| 28 | $k_1$ | 0.53 | $3\times10^{-10}\pm6$ | 1 ± 9 | $3\times10^{-7}\pm0.2$ |
| 29 | $k_2$ | 0.53 | 0.26 ± 2.7 | 1 ± 6 | 0.2 ± 0.15 |
| 30 | $k_3$ | 0.68 | 1.4 ± 1.7 | $71.7\pm5\times10^{4}$ | 0.14 ± 0.3 |
| 31 | $k_4$ | 0.36 | 0.05 ± 0.3 | $10^{-13}\pm155$ | 0.49 ± 0.3 |
| 32 | $k_5$ | 0.76 | 0.7 ± 1 | 0.8 ± 0.88 | 0.6 ± 0.1 |
| 33 | $k_6$ | 0.27 | $2\times10^{-10}\pm0.7$ | 0.03 ± 0.1 | 0.09 ± 0.04 |
| 34 | $k_7$ | 0.8 | 3.4 ± 13 | $3\times10^{-18}\pm5.8$ | 7.7 ± 1.7 |
| 35 | $k_8$ | 0.22 | $10^{-3}\pm7.5$ | $7.7\times10^{-15}\pm1.4$ | $10^{-6}\pm1.3$ |
| 36 | $k_9$ | 1.47 | 1.27 ± 0.4 | 1.4 ± 0.8 | 1.78 ± 0.1 |
| 37 | $k_{10}$ | 0.53 | 0.25 ± 0.2 | 0.5 ± 1.1 | 0.47 ± 0.05 |
| 38 | $k_{11}$ | 1.32 | 1.64 ± 0.5 | 1.6 ± 0.5 | 1 ± 0.1 |
| 39 | $k_{12}$ | 0.68 | 0.85 ± 2 | 0.8 ± 0.8 | 0.85 ± 0.1 |
| 40 | $k_{13}$ | 1.24 | 0.84 ± 5 | 0.5 ± 0.7 | 0.7 ± 0.4 |
| 41 | $k_{14}$ | 0.76 | 1.33 ± 3.4 | 0.03 ± 1.3 | 0.46 ± 1 |

[a]Parameter values (dim $(\underline{\theta})= 41$) of model of Rocha et al. [24]

with the reaction rate $r_7$, initially defined by (10), simplified to:

$$r_7 = \frac{\mathrm{VmT}_{\mathrm{GF}_3}\mathrm{GF}_3}{\mathrm{Kmt}_{\mathrm{GF}_3} + \mathrm{GF}_3} \tag{24}$$

Parameter identification of model (23) with the experiment combination 2B–2FB still shows a good predictive capability in direct validation (see Figs. 2 and 3). This a posteriori confirms the legitimacy of the model reduction. The identification results of Table 5 show that most of the uncertainties,

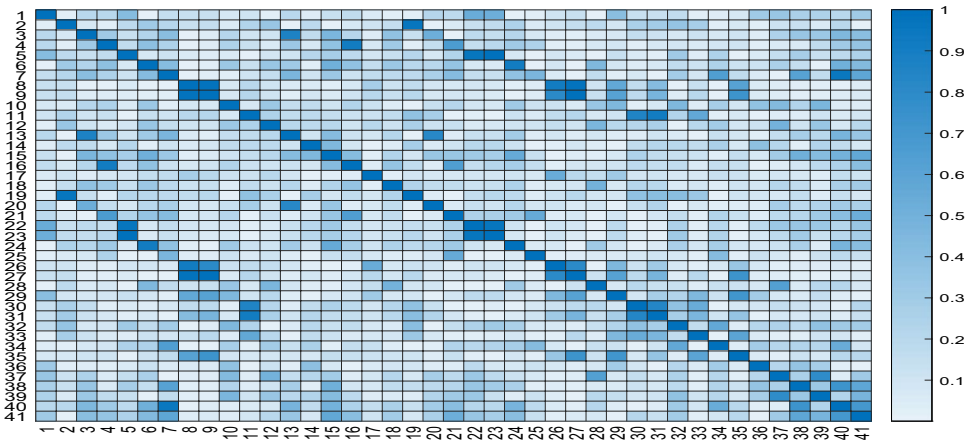

**Fig. 4** Correlation matrix of model (23) represented as a heatmap. Parameters are numbered from 1 to 41 following the sequence of the first column of Table 3

but not all of them, are reduced. Whereas the FIM is regular with condition (17) satisfied and rank (FIM) = dim ( FIM ) = 31 (see Table 2), $\epsilon$−numerical rank (S) = 30 and $\epsilon$−numerical rank$(S^TS) = 30$ instead of $N_p = 31$. It is, therefore, advisable to eliminate one more parameter, and the next in the QR subset selection would be $k_6$. The L2 norm method is also pointing to this parameter and, moreover, the current identification results indicate that this pseudo-stoichiometric coefficient is very small (see Table 5), and could probably be neglected.

In the light of this last reduction, the 30-parameter model is given by:

$$\begin{cases} \dot{\mathrm{GF}} &= D(\mathrm{GF_{in}} - \mathrm{GF}) - r_1 - r_5 + \frac{k_{12}}{2} r_6 \\ \dot{\mathrm{GF}}_2 &= -D \times \mathrm{GF}_2 - r_2 + k_5 r_3 + \frac{k_9}{2} r_5 - r_6 + \frac{k_{14}}{2} r_7 \\ \dot{\mathrm{GF}}_3 &= -D \times \mathrm{GF}_3 - r_3 + k_7 r_4 + \frac{k_{11}}{2} r_6 - r_7 \\ \dot{\mathrm{GF}}_4 &= -D \times \mathrm{GF}_4 - r_4 + \frac{k_{13}}{2} r_7 \\ \dot{\mathrm{F}} &= -D \times \mathrm{F} + k_2 r_1 + k_4 r_2 \\ \dot{\mathrm{G}} &= -D \times \mathrm{G} + \frac{k_{10}}{2} r_5 \\ \dot{V} &= Q_{\mathrm{in}} - Q_{\mathrm{out}}. \end{cases} \tag{25}$$

The identification results for this latter model show that the FIM is regular since condition (17) is satisfied and rank (FIM) = dim ( FIM ) = $\epsilon$−numerical rank $(S) = 30$ while $N_p = 30$ (see Table 2).

However, some parameters remain associated with fairly large uncertainties, and it is proposed to eliminate the parameters Kgst, $\mathrm{Kih}_{\mathrm{GF}_2}$ and $\mathrm{Kih}_{\mathrm{GF}_4}$ for which the coefficients of variation are more than 50% (not explicitly shown in Table 5, but the second column indicates the confidence intervals). Therefore, the simplified reaction rates $r_i$, $i = 2, 4, 5$ become

$$r_i = \frac{\mathrm{Vmh}_{\mathrm{GF}_i}\ \mathrm{GF}_i}{\mathrm{Kmh}_{\mathrm{GF}_i} + \mathrm{GF}_i}, \quad r_5 = \frac{\mathrm{VmT}_{\mathrm{GF}}\ \mathrm{GF}}{\mathrm{K}mst + GF + \frac{\mathrm{GF}^2}{\mathrm{Ksts}}}, \quad i = 2, 4. \tag{26}$$

The coefficients of variation are also large for $\mathrm{K}mh_{\mathrm{GF}}$, $\mathrm{Kmh}_{\mathrm{GF}_2}$ and $\mathrm{Kmh}_{\mathrm{GF}_4}$ but they are a priori kept in the model in view of their important role. Note also that their precision is also likely to improve after the model simplification and re-identification.

The identification results for model (25, 26) with the data combination 2B–2FB show again that the FIM is regular with condition (17) satisfied and rank(FIM)=dim (FIM ) while $N_p = 27$ (see Table 2).

Note however that rank$(S)$ = $\epsilon$−numerical rank$(S) = 26$. According to the L2-norm of the sensitivities, one could then try to eliminate parameter $k_2$, whose estimated value is very small and is associated to a large uncertainty (see Table 5). Following this indication, reaction $r_1$ would be eliminated with the corresponding parameters $\mathrm{Vmh}_{\mathrm{GF}}$ and $\mathrm{Kmh}_{\mathrm{GF}}$.

Moreover, $\epsilon$−numerical rank$(S^TS) = 25$ and, according to the QR subset selection, one could eliminate the two pseudo-stoichiometric coefficients $k_2$ (already pointed out by the L2-norm indicator) and $k_{14}$ which is also very small and largely uncertain (see Table 5).

Therefore, the model parametrization is further reduced to a set of 23 parameters, and the mass balance ODEs simplify to

$$\begin{cases} \dot{\mathrm{GF}} &= D(\mathrm{GF_{in}} - \mathrm{GF}) - r_5 + \frac{k_{12}}{2} r_6 \\ \dot{\mathrm{GF}}_2 &= -D \times \mathrm{GF}_2 - r_2 + k_5 r_3 + \frac{k_9}{2} r_5 - r_6 \\ \dot{\mathrm{GF}}_3 &= -D \times \mathrm{GF}_3 - r_3 + k_7 r_4 + \frac{k_{11}}{2} r_6 - r_7 \\ \dot{\mathrm{GF}}_4 &= -D \times \mathrm{GF}_4 - r_4 + \frac{k_{13}}{2} r_7 \\ \dot{\mathrm{F}} &= -D \times \mathrm{F} + k_4 r_2 \\ \dot{\mathrm{G}} &= -D \times \mathrm{G} + \frac{k_{10}}{2} r_5 \\ \dot{V} &= Q_{\mathrm{in}} - Q_{\mathrm{out}}, \end{cases} \tag{27}$$

with the reaction rates

$$r_i = \frac{\mathrm{Vmh}_{\mathrm{GF}_i}\ \mathrm{GF}_i}{\mathrm{Kmh}_{\mathrm{GF}_i} + \mathrm{GF}_i}, \quad i = 2, 4, \quad r_3 = \frac{\mathrm{Vmh}_{\mathrm{GF}_3}\ \mathrm{GF}_3}{\mathrm{Kmh}_{\mathrm{GF}_3} + \mathrm{GF}_3 + \frac{\mathrm{GF}_3^2}{\mathrm{Kih}_{\mathrm{GF}_3}}}, \tag{28}$$

**Table 4** Classification of the model parameters based on the data of 2B–2FB according to the L2 norm and the QR decomposition

| | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 |
|---|---|---|---|---|---|---|---|---|
| Ranking L2 | $k_1$ | $k_8$ | $KitGF_3$ | $Y_G$ | $\mu_{mG}$ | KmG | $KihGF_4$ | $k_2$ |
| Ranking QR | $k_8$ | $k_1$ | $KitGF_3$ | $Y_G$ | $Y_F$ | $k_3$ | $KihGF_4$ | KmG |
| | 33 | 32 | 31 | | | | | |
| Ranking L2 | KmF | $k_3$ | $Y_F$ | | | | | |
| Ranking QR | $k_{14}$ | KmF | $\mu_{mG}$ | | | | | |

$$r_5 = \frac{\mathrm{VmT_{GF}\ GF}}{\mathrm{Kmst} + GF + \frac{\mathrm{GF}^2}{Ksts}}, \quad r_i = \frac{\mathrm{VmT_{GF_j}\ GF}_j}{\mathrm{Kmt_{GF_j}} + \mathrm{GF}_j + \frac{\mathrm{G}}{Kit_{\mathrm{GF}_j}}}, \quad i = 6, 7; \quad j = 2, 3. \tag{29}$$

The parameter identification of model (27–29) with the 2B–2FB data configuration (see Table 5) show that all parameter uncertainties are very significantly reduced. The FIM is regular (see Table 2) with rank(FIM) = rank($S$) = $\epsilon$−numerical rank($S$) = dim ( FIM ) = 23. To complete the identifiability analysis, the correlation matrix obtained from

**Table 5** Parameter identification results of models (23), (25), (25–26) and (27–29) using the data combination 2B–2FB

| No. | Parameter | (23) | (25) | (25-26) | (27–29) | Var[a] (%) |
|---|---|---|---|---|---|---|
| | VmhGF | 0.08 ± 0.27 | 1.78 ± 0.07 | 463 ± 29 | | |
| 1 | $VmhGF_2$ | 5.84 ± 3.91 | 15.15 ± 15.6 | 6.5 ± 0.3 | 0.1 ± 0.02 | 10 |
| 2 | $VmhGF_3$ | 7.37 ± 0.25 | 1.64 ± 0.44 | 450 ± 40 | 17.2 ± 0.57 | 1.66 |
| 3 | $VmhGF_4$ | 2.8 ± 1 | 4.98 ± 1 | 42.8±12.3 | 29.5 ± 10 | 16.95 |
| 4 | VmtGF | 125.7 ± 17 | 76.4 ± 10.9 | 151.6 ± 13 | 167 ± 4.8 | 1.44 |
| 5 | $VmtGF_2$ | 103.9 ± 6.4 | 401.4 ± 42.2 | 195.5 ± 19.9 | 191 ± 9.5 | 2.49 |
| 6 | $VmtGF_3$ | 9.58 ± 1.9 | 9.98 ± 0.75 | 8.9 ± 26.8 | 11.8 ± 1.7 | 7.20 |
| | KmhGF | 8.18 ± 4.8 | 1.43 ± 0.8 | 8747 ± 1234 | | |
| 7 | $KmhGF_2$ | 72.6 ± 154 | 0.08 ± 0.35 | 57.5 ± 40.5 | 0.36 ± 0.9 | 125 |
| 8 | $KmhGF_3$ | 18.9 ± 4.8 | 0.8 ± 0.45 | 1153 ± 227 | 47.8 ± 4.7 | 4.92 |
| 9 | $KmhGF_4$ | 138.4 ± 39.8 | 206.7 ± 21.8 | 2100 ± 303 | 1730 ± 348 | 10.06 |
| 10 | Kmst | 264 ± 62 | 50.3 ± 12.7 | 1508 ± 203 | 447 ± 81.6 | 9.13 |
| 11 | $KmtGF_2$ | 197.4 ± 71.5 | 483.9 ± 193.7 | 1327 ± 223 | 492 ± 92 | 9.35 |
| 12 | $KmtGF_3$ | 134.6 ± 27 | 121 ± 24 | 509 ± 52 | 252 ± 48 | 9.52 |
| | $KihGF_2$ | 0.46 ± 0.47 | 1 ± 1.2 | | | |
| 13 | $KihGF_3$ | 16.2 ± 5.16 | 11.25 ± 5.5 | 0.89 ± 0.09 | 7.97 ± 1.5 | 9.41 |
| | $KihGF_4$ | 4.48 ± 5.5 | 2.4 ± 2.1 | | | |
| 14 | Ksts | 11.5 ± 1.9 | 17.7 ± 4.4 | 3.1 ± 0.6 | 7.2 ± 0.6 | 4.17 |
| | Kgst | 6.6 ± 18 | 57.4 ± 357.5 | | | |
| 15 | $KitGF_2$ | 0.27 ± 0.03 | 0.05 ± 0.01 | 0.24 ± 0.1 | 0.11 ± 0.01 | 4.54 |
| | $k_2$ | 4.8 ± 16 | 0.15 ± 0.04 | $2 \times 10^{-13}$ ±0.02 | | |
| 16 | $k_4$ | 0.99 ± 1 | 0.16 ± 0.06 | 0.14 ± 0.07 | 2.46 ± 0.6 | 12.2 |
| 17 | $k_5$ | 0.9 ± 0.3 | 2.8 ± 0.4 | 0.64 ± 0.3 | 0.88 ± 0.12 | 6.82 |
| | $k_6$ | $9 \times 10^{-16}$ ±0.07 | | | | |
| 18 | $k_7$ | 7.8 ± 2 | 3.7 ± 1.4 | 11.4 ± 4.5 | 8.4 ± 1.5 | 8.93 |
| 19 | $k_9$ | 2.3 ± 0.4 | 2.27 ± 0.22 | 4.94 ± 0.9 | 2.23 ± 0.17 | 3.81 |
| 20 | $k_{10}$ | 0.4 ± 0.05 | 0.35 ± 0.06 | 1.56 ± 0.29 | 0.52 ± 0.05 | 4.81 |
| 21 | $k_{11}$ | 0.77 ± 0.19 | 0.55 ± 0.07 | 2.2 ± 0.7 | 0.99 ± 0.05 | 2.52 |
| 22 | $k_{12}$ | 0.56 ± 0.15 | 1.1 ± 0.14 | 0.17 ± 0.14 | 0.47 ± 0.09 | 9.57 |
| 23 | $k_{13}$ | 0.32 ± 0.11 | 0.3 ± 0.07 | 1.26 ± 3.9 | 0.45 ± 0.11 | 12.22 |
| | $k_{14}$ | 1.49 ± 0.78 | 1.46 ± 0.4 | 0.05 ± 4.3 | | |

[a]Variation coefficients of the estimated parameters of model (27–29)

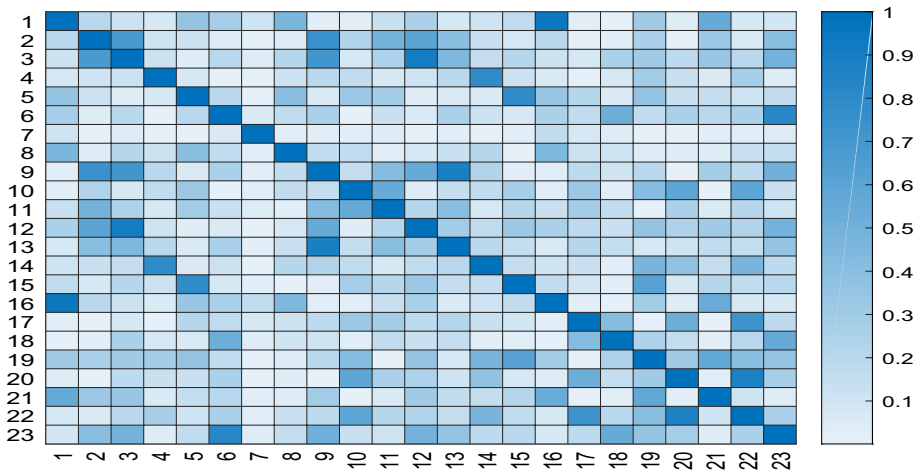


**Fig. 5** Correlation matrix of model (27–29) represented as a heatmap. Parameters are numbered from 1 to 23 following the sequence of the first column of Table 5

**Table 6** Remaining high parameter correlations

| First parameter | Second parameter | Correlation coefficient |
|---|---|---|
| $Kmh_{GF4}$ | $Vmh_{GF3}$ | − 0.739 |
| $Kmh_{GF4}$ | $Vmh_{GF4}$ | 0.715 |
| $Kmt_{GF3}$ | $Vmh_{GF4}$ | 0.898 |
| $Kih_{GF3}$ | $Kmh_{GF4}$ | − 0.885 |
| Ksts | $Vmt_{GF}$ | − 0.791 |
| $Kit_{GF2}$ | $Vmt_{GF2}$ | − 0.789 |
| $k_4$ | $Vmh_{GF2}$ | − 0.95 |
| $k_{12}$ | $k_{10}$ | − 0.725 |
| $k_{12}$ | $k_5$ | − 0.868 |
| $k_{13}$ | $Vmt_{GF3}$ | − 0.814 |

(16) is represented in Fig 5 and shows that, even if all parameters appear as identifiable, several strong parameter inter-dependencies remain ($\rho_{ij} \geq 0.7$). The corresponding correlation coefficients are shown in Table 6. Obviously, four kinds of inter-dependencies appear: (i) inner-kinetic dependency between the maximum rate coefficients ($Vmh_{GF4}$, $Vmt_{GF,GF2}$) and their respective half-saturation and/or inhibition constants ($Kmh_{GF4}$, Ksts,$Kit_{GF2}$), (ii) cross-kinetic dependencies between parameters as, for instance, $Kmt_{GF3}$ and $Vmh_{GF4}$ or $Kih_{GF3}$ and $Kmh_{GF4}$, (iii) stoichiometric ($k_4$, $k_{13}$)/kinetic ($Vmh_{GF2}$,$Vmt_{GF3}$) parameter inter-dependency, and (iv) stoichiometric inter-dependency between $k_{12}$ and $k_{10}/k_5$. (i) and (iii) can be relatively well understood since the presence of multiplicative terms containing parameters may reinforce parameter dependency while (ii) and (iv) can be interpreted as a consequence of the cascade form of model (27–29) considering rates 3 and 4, as well as 5–7.

The interpretation of model (27–29) would therefore be based on the following chemical/biological assumptions:

- The fractions of GF involved as reactants or products in hydrolysis reactions are negligible compared to the fractions consumed or produced by transfructosylations;
- The fraction of $GF_2$ produced by transfructosylation of $GF_3$ is negligible compared to all the fractions produced from the other considered reactions;
- The production of fructose is mainly explained by hydrolysis of $GF_2$ as well as the production of glucose from transfructosylation of GF;
- Under the considered operating conditions, biomass dynamics may be neglected.

To double check the predictive capability of the reduced model, cross validation is achieved with the batch data B3–B4. Figure 6 shows the fit with the data of batch B4. Note that cross validation implies the identification of the most likely initial conditions listed in Table 7.

To investigate the impact of the parameter uncertainties on the model prediction, additional computations can be achieved, including the estimation of the covariance of the model prediction error, or alternatively, Monte Carlo analysis which considers all scenarios generated from the

**Table 7** Identification of initial conditions of the fermentation B3 and B4 for cross validation

| Batch fermentation | GF(0) | $GF_2(0)$ | $GF_3(0)$ | $GF_4(0)$ | F(0) | G(0) |
|---|---|---|---|---|---|---|
| B3 | 160 | 4.4 | 10.1 | $10^{-3}$ | 6.3 | 10.1 |
| B4 | 177.4 | $5 \times 10^{-3}$ | 0.04 | 1.2 | 7.30 | 30.2 |

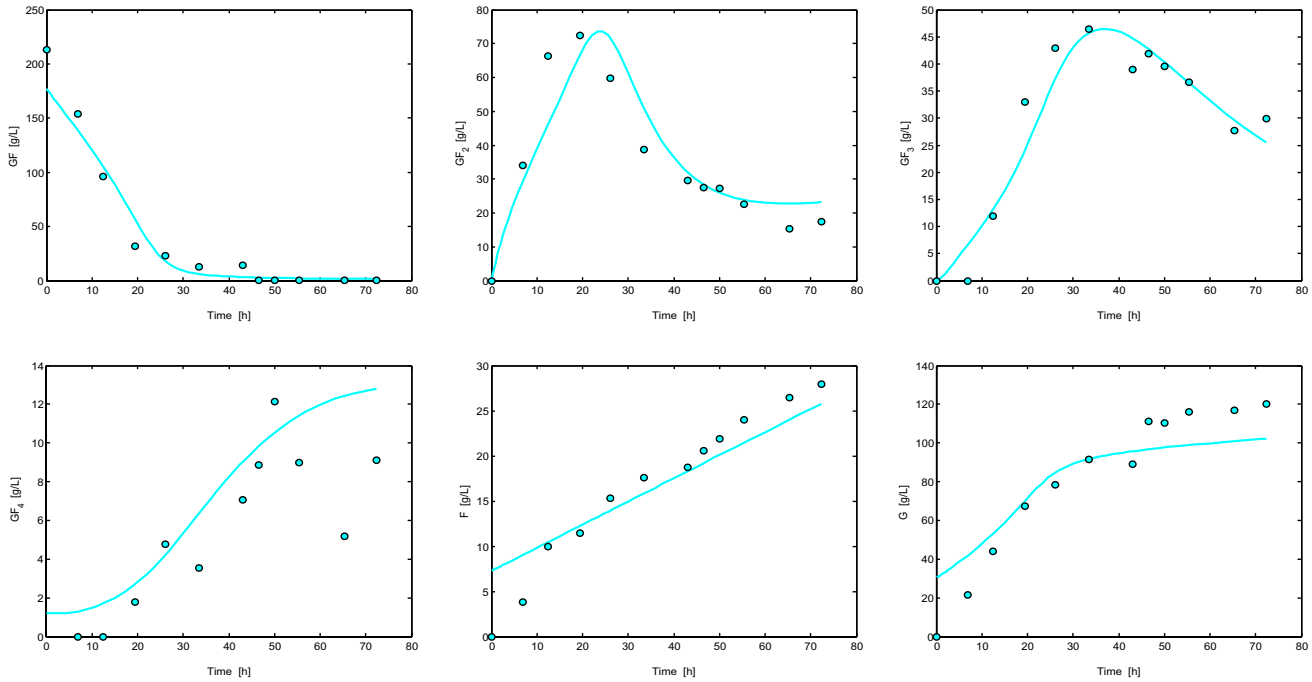


**Fig. 6** Cross validation of reduced model (27–29) using fermentation B4 (RMSE = 9.43)

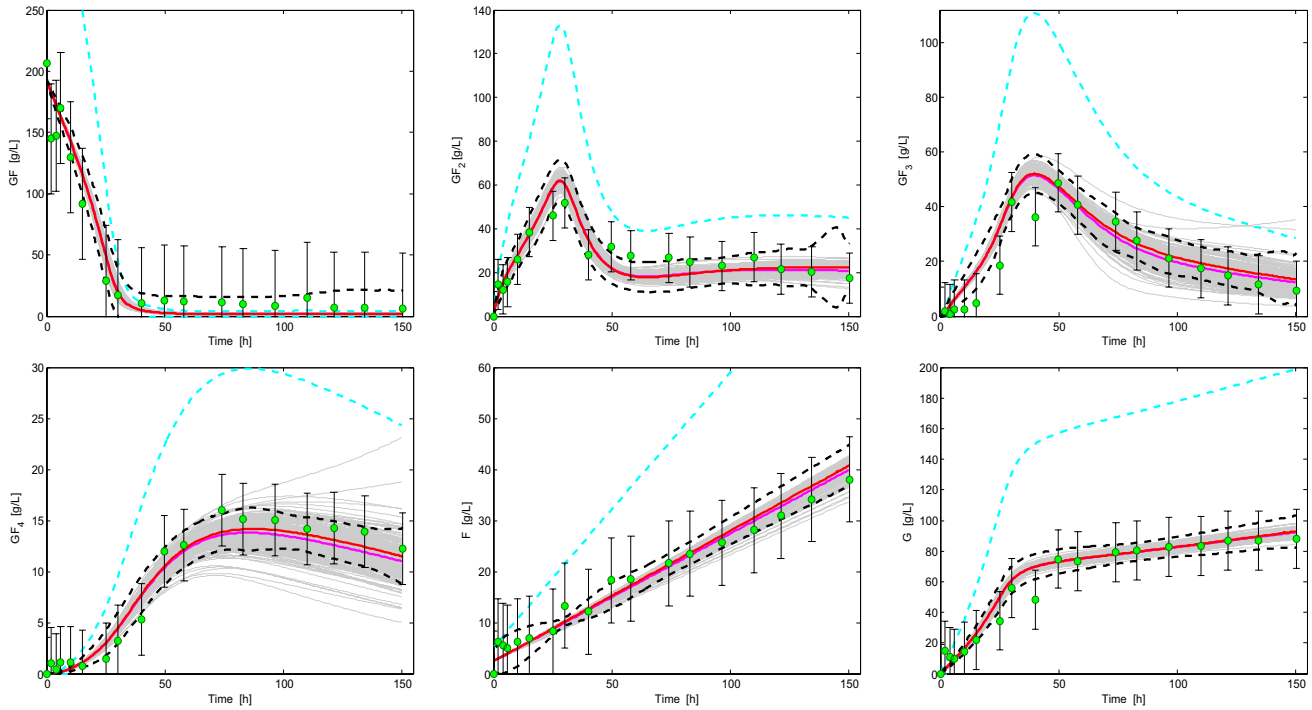


**Fig. 7** Time evolution of the concentrations in experiment B1. All the experimental samples (green dots with 95% black error bars) have a non-empty intersection with the 95% confidence intervals (blach dashed lines) of the simulated concentrations (red solid lines)—Monte Carlo simulation (200 trajectories in light grey) with a mean trajectory in purple, and 95%-confidence corridors in dashed cyan lines

estimation error covariance matrix $C_{\underline{\theta}}$ obtained by inverting the FIM.

In the case of batch B1, Fig. 7 shows:

- the trajectory predicted by the dynamic model (red continuous line);
- the 95% confidence intervals (black dashed lines) of the model prediction, which are calculated based on an estimation of the covariance of the model prediction error through $S^T \mathrm{FIM}^{-1} S$;
- Monte Carlo simulation (200 trajectories in light grey) corresponding to 200 realizations of the parameter probability distribution (generated in MATLAB by lhsnorm) with mean vector $\hat{\underline{\theta}}$ and covariance matrix $\mathrm{FIM}^{-1}\left(\hat{\underline{\theta}}, \hat{\Sigma}\right)$. The average trajectory is represented in magenta (and almost coincides with the model prediction in red), while the 95% confidence intervals are represented in cyan.

## Conclusion

The main objective of this paper is to show that sensitivity analysis, the Fisher Information matrix, and parameter subset selection based on the QR decomposition can be used as systematic tools to reduce detailed biological models to representations which are practically identifiable with the data at hand. Indeed, a priori process knowledge often leads to the derivation of models which are over-parametrized, at least when considering the limited information content of usual batch and fed-batch experiments. Of course, experiment design is of paramount importance and informative experiments allows improving parameter accuracy and precision, but at the expense of time consuming and delicate experimentation. As a case study, the reduction of a dynamic model of the bioproduction of FOS is discussed. The procedure allows reducing drastically the number of model parameters from 41 to 23, in several successive steps where intermediate models are estimated and analyzed.

**Acknowledgements** The authors thank the financial support from the F.R.S.-FNRS, the Belgium National Fund for the Scientific Research (Research Project 24643.08). C. Nobre thanks the Fundação para a Ciência e Tecnologia for the strategic funding of UID/BIO/04469 /2013 unit.

### Compliance with ethical standards

**Conflict of interest** The authors declare no conflict of interest.

## References

1. Alvarado-Huallanco M, Maugeri-Filho F (2010) Kinetics and modeling of fructooligosaccharide synthesis by immobilized fructosyltransferase from *Rhodotorula* sp. J Chem Technol Biotechnol 85:1654–1662
2. Aso Y, Akaza H, Kotake T, Tsukamoto T, Imai K, Naito S (1995) The BLP Study Group: preventive effect of a *Lactobacillus casei* preparation on the recurrence of superficial bladder cancer in a double-blind trial. Eur Urol 27:104–109
3. Aso Y, Akazan H, Group BS (1992) Prophylactic effect of a *Lactobacillus casei* preparation on the recurrence of superficial bladder cancer. Urol Int 49:125–129
4. Castro C, Nobre C, Duprez M, Weireld GD, Hantson A (2017) Screening and selection of potential carriers to immobilize *Aureobasidium pullulans* cells for fructo-oligosaccharides production. Biochem Eng J 118:82–90
5. Dias L, Velosoand A, Correia D, Rocha O, Torres D, Rocha I, Rodrigues L, Peres A (2009) UV spectrophotometry method for the monitoring of galacto-oligosaccharides production. Food Chem 113:246–252
6. Dominguez A, Nobre C, Rodrigues L, Peres A, Torres D, Rocha I (2012) New improved method for fructooligosaccharides production by *Aureobasidium pullulans*. Carbohydrate Polym 89:1174–1179
7. Donoso-Bravo A, Mailier J, Ruiz-Filippi G, Vande Wouwer A (2013) Identification in an anaerobic batch system: global sensitivity analysis, multi-start strategy and optimization criterion selection. Bioprocess Biosyst Eng 36:35–43
8. Duan K, Chen J, Sheu D (1994) Kinetic studies and mathematical model for enzymatic production of fructooligosaccharides from sucrose. Enzyme Microb Technol 16:334–339
9. Duan KJ, Chen JS, Sheu DC (1994) Kinetic studies and mathematical model for enzymatic production of fructooligosaccharides from sucrose. Enzyme Microbial Technol 16:334–339
10. Fekih-Salem R, Vande Wouwer A, Castro CD, Nobre C, Hantson A (2015) Parameter identification of the fermentative production of fructo-oligosaccharides by *Aureobasidium pullulans*. In: Proceedings of the 19th international conference on system theory, control and computing, pp 43–48
11. Fiordalis A, Georgakis C (2013) Data-driven, using design of dynamic experiments, versus model-driven optimization of batch crystallization processes. J Process Control 23(2):179–188
12. Gibson G (1998) Kinetic studies and mathematical model for enzymatic production of fructooligosaccharides from sucrose. Br J Nutr 80:209–212
13. Golub G, Loan CV (2013) Matrix computations. Studies in mathematical sciences, 4th edn. Johns Hopkins University Press, Baltimore
14. Grosfils A, Vande Wouwer A, Bogaerts P (2007) On a general model structure for macroscopic biological reaction rates. J Biotechnol 130:253–264
15. Guio F, Rugeles L, Rojas S, Palomino M, Camargo M, Sanchez O (2012) Kinetic modeling of fructooligosaccharide production using *Aspergillus oryzae* N74. Appl Biochem Biotechnol 167:142–163
16. Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. In: Dellino G, Meloni C (eds) Uncertainty management in simulation-optimization of complex systems. Operations research/computer science interfaces, vol 59. Springer, Boston
17. Joshi M, Seidel-Morgenstern A, Kremling A (2006) Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. Metab Eng 8:447–455
18. Jung K, Run JW, Kang KR, Lim JY, Lee JH (1989) Mathematical model for enzymatic production of fructo-oligosaccharides from sucrose. Enzyme Microb Technol 11:491–494
19. Kiparissides A, Georgakis C, Mantalaris A, Pistikopoulos EN (2014) Design of in silico experiments as a tool for nonlinear sensitivity analysis of knowledge-driven models. Ind Eng Chem Res 53(1):7517–7525

20. Mutanda T, Mokoena M, Olaniran A, Wilhelmi B, Whiteley C (2014) Microbial enzymatic production and applications of short-chain fructooligosaccharides and inulooligosaccharides: recent advances and current perspectives. J Ind Microbiol Biotechnol 41(6):893–906
21. Nishizawa K, Nakajima M, Nabetani H (2001) Kinetic study on transfructosylation by b-fructofuranosidase from *Aspergillus niger* ATCC 20611 and availability of a membrane reactor for fructooligosaccharide production. Food Sci Technol Res 7(1):39–44
22. Nobre C, Santos M, Dominguez A, Torres D, Rocha O, Peres A, Rocha I, Ferreira E, Teixeira J, Rodrigues L (2009) Comparison of adsorption equilibrium of fructose, glucose and sucrose on potassium gel-type and macroporous sodium ion-exchange resins. Anal Chim Acta 654(1):71–76
23. Preter VD, Hamer HM, Windey K, Verbeke K (2011) The impact of pre- and/or probiotics on human colonic metabolism: does it affect human health? Mol Nutr Food Res 55:46–57
24. Rocha O, Nobre C, Dominguez A, Torres D, Faria N, Rodrigues L, Teixeira J, Ferreira E, Rocha I (2009) A dynamical model for the fermentative production of fructooligosaccharides. In: 10th International Symposium on process systems engineering, pp 1–7
25. Sangeetha P, Ramesh M, Prapulla S (2005) Recent trends in the microbial production, analysis and application of fructooligosaccharides. Trends Food Sci Technol 16:442–457
26. Schenkendorf R, Kremling A, Mangold M (2009) Optimal experimental design with the sigma point method. IET Syst Biol 3:10–23
27. Schorsch J, Kinnaert M, Fekih-Salem R, Dewasme L, Castro C, Vande Wouwer A (2018) Identification and optimal control of fructo-oligosaccharide production. IFAC-PapersOnLine 51(18):678–683
28. Simeone M, Hogue I, Ray C, Kirschner D (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol 254(1):178–196
29. Walter E, Pronzato L (1997) Identification of parametric models from experimental data. Communications and control engineering. Springer, Berlin
30. Yun JW, Song SK (1993) The production of high-content fructooligosaccharides from sucrose by the mixed-enzyme system of fructosyltransferase and glucose-oxidase. Biotechnol Lett 15:573–576
31. Zak D, Gonye G, Schwaber J, Doyle F (2003) Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an in silico network. Genome Res 13:2396–2405
32. Zhao QH, Urosević D, Mladenović N, Hansen P (2009) A restarted and modified simplex search for unconstrained optimization. Comput Oper Res 36(12):3263–3271