



HAL
open science

Assessing practical identifiability during calibration and cross-validation of a structured model for high-solids anaerobic digestion

Vicente Pastor-Poquet, Stefano Papirio, Jérôme Harmand, Jean-Philippe Steyer, Eric Trably, Renaud Escudié, Giovanni Esposito

► To cite this version:

Vicente Pastor-Poquet, Stefano Papirio, Jérôme Harmand, Jean-Philippe Steyer, Eric Trably, et al.. Assessing practical identifiability during calibration and cross-validation of a structured model for high-solids anaerobic digestion. *Water Research*, 2019, 164, 10.1016/j.watres.2019.114932 . hal-02623443

HAL Id: hal-02623443

<https://hal.inrae.fr/hal-02623443>

Submitted on 20 Dec 2021

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 - NonCommercial 4.0 International License

1 **Assessing Practical Identifiability during Calibration and Cross-**
2 **Validation of a Structured Model for High-Solids Anaerobic Digestion**

3

4 Vicente Pastor-Poquet ^{a,b,*}, Stefano Papirio ^c, Jérôme Harmand ^b, Jean-Philippe Steyer ^b,
5 Eric Trably ^b, Renaud Escudie ^b, and Giovanni Esposito ^c

6

7 ^a Department of Civil and Mechanical Engineering, University of Cassino and Southern
8 Lazio, via Di Biasio 43, 03043 Cassino (FR), Italy

9 * Corresponding author. E-mail: vicente.pastor.poquet@gmail.com

10 ^b LBE, Univ. Montpellier, INRA, 102 Avenue des Etangs, 11100, Narbonne, France

11 ^c Department of Civil, Architectural and Environmental Engineering, University of
12 Napoli Federico II, via Claudio 21, 80125 Napoli, Italy

13

14

15 **ABSTRACT**

16 High-solids anaerobic digestion (HS-AD) of the organic fraction of municipal solid
17 waste (OFMSW) is operated at a total solid (TS) content $\geq 10\%$ to enhance the waste
18 treatment economy, though it might be associated to free ammonia (NH₃) inhibition.
19 This study aimed to calibrate and cross-validate a HS-AD model for homogenized
20 reactors in order to assess the effects of high NH₃ levels in HS-AD of OFMSW, but also
21 to evaluate the suitability of the reversible non-competitive inhibition function to
22 reproduce the effect of NH₃ on the main acetogenic and methanogenic populations. The
23 practical identifiability of structural/biochemical parameters (i.e. 35) and initial
24 conditions (i.e. 32) was evaluated using batch experiments at different TS and/or

25 inoculum-to-substrate ratios. Variance-based global sensitivity analysis and
26 approximate Bayesian computation were used for parameter optimization. The
27 experimental data in this study permitted to estimate up to 8 biochemical parameters,
28 whereas the rest of parameters and biomass contents were poorly identifiable. The study
29 also showed the relatively high levels of NH_3 (i.e. up to 2.3 g N/L) and ionic strength
30 (i.e. up to 0.9 M) when increasing TS in HS-AD of OFMSW. However, the NH_3 non-
31 competitive function was unable to capture the acetogenic/methanogenic inhibition.
32 Therefore, the calibration emphasized the need for target-oriented experimental data to
33 enhance the practical identifiability and the predictive capabilities of structured HS-AD
34 models, but also the need for further testing the NH_3 inhibition function used in these
35 simulations.

36

37 **Keywords:** High-solids Anaerobic Digestion Model; Ammonia Inhibition; Ionic
38 Strength; Global Sensitivity Analysis; Approximate Bayesian Computation.

39

40 **1 INTRODUCTION**

41 High-solids anaerobic digestion (HS-AD) of the organic fraction of municipal solid
42 waste (OFMSW) is operated at total solids (TS) $\geq 10\%$ to minimize the reactor volume,
43 the water addition and the digestate dewatering (Pastor-Poquet et al., 2019a). HS-AD
44 can also lead up to 80 % TS removal, easing the digestate post-treatment. However, HS-
45 AD of OFMSW is usually associated to free ammonia (NH_3) inhibition, resulting in
46 volatile fatty acids (VFA) accumulation. NH_3 affects both acetoclastic and
47 hydrogenotrophic methanogens but also the rest of VFA degraders (acetogens) in
48 anaerobic digestion (AD), being the inhibition related to the operative parameters (i.e.
49 temperature and pH) and the biomass acclimation (Rajagopal et al., 2013).

50

51 Adding lignocellulosic materials to OFMSW permits to minimize the buildup of total
52 ammonia nitrogen (TAN), while their low hydrolysis rates permit to increase the TS
53 content and to counteract the VFA accumulation in HS-AD (Capson-Tojo et al., 2017;
54 Pastor-Poquet et al., 2019a). However, including lignocellulosic waste in OFMSW
55 depends on the season or the local waste management strategy. Whether or not a
56 lignocellulosic co-substrate is used, understanding the effects of NH_3 inhibition is
57 crucial to optimize HS-AD of OFMSW.

58

59 A HS-AD model was recently developed for homogenized reactors to evaluate the NH_3
60 inhibition in HS-AD of OFMSW (Pastor-Poquet et al., 2018, 2019b). This structured
61 model, based on the Anaerobic Digestion Model No. 1 (ADM1) (Batstone et al., 2002),
62 gathers the main bio-physical-chemical mechanisms in HS-AD. In the HS-AD model,
63 apparent (i.e. $\text{kmol/kg H}_2\text{O}$) instead of global (i.e. kmol/kg) concentrations determine

64 the effect of TS upon solutes, as a consequence of the low water content within HS-AD.
 65 Meanwhile, an extended set of mass balances allows the simulation of the organic mass
 66 removal from the biogas production. A liquid solution ‘non-ideality’ subroutine was
 67 subsequently included in the model as a function of the ionic strength (I), since ‘non-
 68 ideality’ determines the pH, CO₂ liquid-gas transfer and NH₃ inhibition in HS-AD
 69 (Pastor-Poquet et al., 2019b). As a preliminary assumption, the HS-AD model included
 70 a reversible non-competitive inhibition function by NH₃ (I_{nh3}) [Equation 1] in the
 71 biochemical rates of acetogenic and methanogenic populations, a mathematical resource
 72 commonly used in structured AD models (Astals et al., 2018).

$$I_{nh3} = \frac{K_{i,Snh3}}{K_{i,Snh3} + S_{nh3,App}} \quad (1)$$

74
 75 Simulating the NH₃ inhibition at high TS with the HS-AD model requires an adequate
 76 set of input parameters, θ , to be estimated by calibration (Pastor-Poquet et al., 2018,
 77 2019b). θ includes both the structural/biochemical parameters, θ_P , and the initial and
 78 influent conditions, θ_B : $\theta = (\theta_P, \theta_B)$. Nonetheless, calibration of structured AD models is
 79 not trivial due to the equation complexity and large number of θ involved (Dochain &
 80 Vanrolleghem, 2001; Donoso-Bravo et al., 2011).

81
 82 To calibrate a mathematical model, θ must be structurally and practically identifiable,
 83 instead of correlated. The θ structural identifiability is theoretically assessed, assuming
 84 noise-free experimental data and error-free model structure. Noteworthy, nearly all θ in
 85 ADM1 are (locally) structurally identifiable (Nimmegeers et al., 2017). This is a
 86 prerequisite to assess the θ practical identifiability and calibrate the HS-AD model using

87 ‘imperfect’ experimental data. Unfortunately, the reduced number of experimental data
88 often available and/or the presence of experimental errors yield non-identifiable
89 parameters; i.e. parameters that cannot be uniquely estimated. These are known as
90 practical identifiability issues.

91

92 Calibration usually consists of minimizing an objective function, $J(\theta)$, that condenses
93 the ‘goodness of fitting’ between the experimental data, y , and the model outputs, $y^{sim}(\theta)$,
94 being these a function of N input parameters, θ (Dochain & Vanrolleghem, 2001;
95 Flotats et al., 2010). Several $J(\theta)$ can be used to calibrate AD models as the weighted
96 sum of squares or any user-defined alternative (Donoso-Bravo et al., 2011; Ratto et al.,
97 2001). Assuming the existence of a global minimum (optimum) for an objective
98 function, $J(\theta_{opt})$, this value is reached using the optimal set of input parameters, θ_{opt} .

99

100 Practical identifiability issues commonly translate into $J(\theta)$ showing many local
101 optimums and/or flat valleys, where the precise value of θ cannot be easily determined
102 (Guisasola et al., 2006; Rodriguez-Fernandez et al., 2006). Thus, only practical
103 identification of a reduced θ subset (i.e. $N' < N$) is often possible for ADM1-based
104 models (Nimmegeers et al., 2017). This is called model over-parameterization, where
105 the modification of two individual θ , θ_i (with $i = 1, \dots, N$), can lead to a similar model
106 output. Particularly, when using batch experiments – highly dependent on the initial
107 conditions – to calibrate AD models, different sets of experimental conditions, including
108 different inoculum-to-substrate ratios (ISR), are required to reduce the θ correlation
109 (Donoso-Bravo et al., 2011; Flotats et al., 2010).

110

111 Two main approaches can be used to calibrate complex models: the Bayesian and the
112 frequentist. The frequentist approach searches for optimal θ values, θ_{opt} , whereas the
113 Bayesian approach considers θ_{opt} as probabilistic distributions conditioned on the
114 experimental data, $p(\theta_{opt}|y)$, instead of single values (Ratto et al., 2001; Rodriguez-
115 Fernandez et al., 2006; Saltelli et al., 2006). In both approaches, when facing over-
116 parameterization, it must be assessed which θ_i significantly influence $y^{sim}(\theta)$ (sensitivity
117 analysis) and need to be adequately calibrated.

118

119 ADM1-based models contain several θ_P (i.e. ≥ 35) and θ_B (i.e. ≥ 24). θ_P might be
120 obtained from literature, though a different model structure – from where these θ_P were
121 obtained – potentially influences the optimal θ_P values/distributions (Pastor-Poquet et
122 al., 2019b). On the other hand, θ_B might not be easily determined due to the lack of
123 experimental data or the difficulties to translate the data into adequate model units
124 (Donoso-Bravo et al., 2011; Flotats et al., 2003).

125

126 Parameter inference based on the Bayes' theorem [Equation 2] is particularly suited to
127 calibrate structured AD models since it can deal with complex $J(\theta)$ showing several
128 optima or flat geometries, where frequentist inference might not be well suited
129 (Kennedy & O'Hagan, 2001; Toni & Stumpf, 2010). In Bayesian inference, the prior
130 parameter distribution, $p(\theta)$, is sampled to obtain the posterior parameter distribution,
131 $p(\theta|y)$, conditioned on the experimental data, y , and the likelihood function, $p(y|\theta)$,
132 while $p(y)$ can be considered as a normalizing constant. Importantly, any user-defined
133 $J(\theta)$ arising from $p(y|\theta)$ can be used in 'informal' statistical approaches [Equation 3], as

134 variance-based global sensitivity analysis (GSA) and approximate Bayesian
 135 computation (ABC) (Donoso-Bravo et al., 2011; Nott et al., 2012).

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (2)$$

$$J(\theta) = f(p(y|\theta)) \quad (3)$$

136

137 Variance-based GSA provides an appropriate assessment about the potentiality of θ_i to
 138 influence the model outputs and the correlations existing with the rest of θ , θ_j (with $j =$
 139 $1, \dots, N$ and $i \neq j$) (Kennedy & Petropoulos, 2017; Oakley & O'Hagan, 2004). Similarly,
 140 ABC permits also to highlight practical identifiability issues yielding simultaneously the
 141 most likely $p(\theta|y)$ (Beaumont et al., 2009; Filippi et al., 2013; Toni & Stumpf, 2010).

142 As main disadvantage, Bayesian inference is often computationally intensive.

143

144 The mathematical performance of the HS-AD model was previously verified, though
 145 the model was only validated for a single HS-AD batch experiment due to the elevated
 146 number of θ requiring calibration (i.e. > 30) (Pastor-Poquet et al., 2018, 2019b). Instead,
 147 this study aimed to fully calibrate and cross-validate the HS-AD model to simulate the
 148 effect of high NH_3 levels in HS-AD of OFMSW, while testing the non-competitive NH_3
 149 inhibition function [Equation 1] on the main acetogenic/methanogenic populations. In
 150 particular, this study assessed the practical identifiability of 35 θ_P and 32 θ_B , by using
 151 nine HS-AD batch digesters at different TS and/or ISR as a source of experimental data.
 152 Identifiability was assessed by variance-based GSA and ABC permitting to approximate
 153 also $p(\theta|y)$. Importantly, the proposed methodology can be easily readapted to account
 154 for further HS-AD datasets (e.g. batch, continuous) and/or model configurations.

155

156

157 **2 METHODOLOGY**158 **2.1 Experimental Data**

159 To calibrate and cross-validate the HS-AD model, while further evaluating the effects of
160 increasing the initial TS content on HS-AD simulations, four different batch

161 experiments were used at thermophilic (55°C) conditions from 10 to 30 % TS [Table 1].

162 The laboratory-scale reactor design volume (V_{Reactor}) was either 160 or 280 mL for the
163 different experiments. In all experiments, centrifuged inoculum was used to increase

164 simultaneously the initial TS and ISR. These small-scale digesters were manually

165 shaken when the biogas production was measured. The batch experiments are described

166 next, whereas a thorough description of these experiments and the bio-physical-

167 chemical analyses performed was reported elsewhere (Pastor-Poquet et al., 2019a).

168

169 Experiment 1 consisted of a sacrifice test for mono-digestion of OFMSW using ISR =

170 1.00 g VS/g VS. In this experiment, the main physical-chemical dynamics (i.e. biogas

171 production and composition, TS, VS, VFA, TAN, and mono-valent ions) were

172 evaluated at different operational times. In Experiments 2 to 4, the biogas production

173 and composition were measured at different experimental times, whereas the rest of

174 physical-chemical analyses (i.e. TS, VS, VFA, TAN and ions) were only performed

175 before starting and after ending each experiment. Non-sacrifice experiments included

176 mono-digestion of OFMSW using ISR = 1.50 g VS/g VS (Experiment 2) and ISR =

177 0.50 g VS/g VS (Experiment 3), but also co-digestion of OFMSW and beech sawdust

178 using ISR = 0.16 g VS/g VS (Experiment 4).

179

180 Within Experiments 2 and 4, different initial TS contents – dilutions – were evaluated,
181 though all the initial batch conditions contained exactly the same amount of substrate
182 and centrifuged inoculum. Briefly: 1) the inoculum was centrifuged; 2) the same
183 amount of substrate and inoculum was added to each bottle; 3) distilled water was
184 added to reach the different TS contents; and 4) each bottle was manually homogenized.
185 This strategy was aimed to use the mass balances among the different initial TS
186 conditions, as explained in Section 2.3.1, since soluble materials were partially removed
187 when centrifuging the inoculum. In total, nine different HS-AD batch conditions were
188 assessed at different TS, ISR and/or co-digestion ratios, subsequently named as “Batch
189 No. 1 to 9” [Table 1].

190

191 **2.2 HS-AD Model**

192 The HS-AD model included the main biochemical rates of Pastor-Poquet et al. (2018),
193 though some minor modifications were also implemented [Table 2]. Firstly, a reversible
194 non-competitive NH_3 inhibition function [Equation 1] was included in the valeric (S_{va}),
195 butyric (S_{bu}), propionic (S_{pro}), and hydrogen (S_{h2}) uptake rates, similarly to the NH_3
196 (S_{nh3}) inhibition on the acetate (S_{ac}) uptake, aiming to simulate the VFA accumulation
197 observed in HS-AD experiments likely consequence of the NH_3 buildup (Pastor-Poquet
198 et al., 2019a). Secondly, carbohydrates (X_{ch}) were split into readily-biodegradable
199 ($X_{ch,fast}$) and slowly-biodegradable ($X_{ch,slow}$) to simulate the relatively lower hydrolysis
200 rates of sawdust and the longer methane production observed in co-digestion
201 experiments (i.e. ≥ 200 d) [Table 1]. Importantly, the hydrolysis of both $X_{ch,fast}$ and
202 $X_{ch,slow}$ pooled into soluble sugars (S_{su}).

203

204 **2.3 Model Calibration and Validation**

205 A common set of biochemical parameters was used for all HS-AD simulations at
 206 different ISR and/or TS. All biochemical parameters for thermophilic (55°C) AD were
 207 extracted from Batstone et al. (2002), though some of those needed to be calibrated (i.e.
 208 θ_P) to improve the model fitting. The initial conditions of the batch experiments were
 209 predefined according to the experimental data available, as described by Pastor-Poquet
 210 et al. (2018) and also mentioned next. Moreover, different ranges of initial biomass
 211 concentrations were used (i.e. θ_B) to assess the potential interrelationship of θ_B with θ_P
 212 in batch experiments.

213

214 **2.3.1 Initial Conditions**

215 The initial conditions used for each batch simulation are shown in Table 3. Since the
 216 same amount of substrate and inoculum was used along different initial TS contents in
 217 Experiment 2, but also in Experiment 4, mass balances were used to reduce the number
 218 of ‘unknown’ initial conditions [Table 1]. To use mass balances, the anaerobic
 219 biodegradability (BD) of each substrate-inoculum mixture in Experiment 2 and
 220 Experiment 4 was assumed constant, whereas the soluble (S) and particulate (X)
 221 components were assumed proportional, among the different TS contents. With these
 222 assumptions, mass balances permitted to extrapolate the molar and chemical oxygen
 223 demand (COD) concentrations as a function of the initial reactor content mass (M_{Global}),
 224 reactor content specific weight (ρ_{Global}) and reactor content volume (V_{Global}). For
 225 example, the concentration of acetoclastic methanogens (X_{ac}) in Batch No. 3 was
 226 approximated from Batch No. 2 as: $X_{ac, Batch3} = X_{ac, Batch2} \cdot V_{Global, Batch2} / V_{Global, Batch3}$.

227

228 The nitrogen content of soluble (S_i) and particulate (X_i) inert in each substrate-inoculum
229 mixture ($N_{i,subs}$) determined the initial protein (X_{pr}) + amino-acid (S_{aa}) content,
230 according to the nitrogen balance, as shown elsewhere (Pastor-Poquet et al., 2018).
231 Thus, $N_{i,subs}$ was different for each batch experiment [Table 3]. On the other hand,
232 despite the initial conditions of batch experiments were maintained for all the
233 simulations, different biomass concentrations (i.e. degraders of sugars, X_{su} ; amino acids,
234 X_{aa} ; long-chain fatty acids, X_{fa} ; valerate, X_{c5} ; butyrate, X_{bu} ; propionate, X_{pro} ; acetate,
235 X_{ac} ; and hydrogen, X_{h2}) were also assessed, as mentioned before. Since mass balances
236 were used among different TS in Experiments 2 and 4, only the initial conditions of
237 Batch No. 1, 2, 6 and 7 [Table 1] were evaluated.

238

239 **2.3.2 Biochemical Parameters, Biomass Concentrations and Calibration Ranges**

240 The modified biochemical parameters (θ_P) and modified biomass concentrations (θ_B) in
241 this study, including their initial values and potential variability ranges, are shown in
242 Table 4. In total, 35 θ_P and 32 θ_B were evaluated. θ_P related to the hydrolysis, sugar
243 fractioning (f_{su}), maximum growth rate (k_m) and half-saturation constant (K_S), but also
244 the pH, NH_3 and H_2 inhibition constants (K_i), since all these θ_P are simultaneously
245 associated to the substrate under study, are correlated among themselves, and strongly
246 regulate the biogas production/composition from solid substrates (Batstone et al., 2002;
247 Garcia-Gen et al., 2015). On the other hand, θ_B included all the initial biomass
248 concentrations in the HS-AD simulations (i.e. X_{su} , X_{aa} , X_{fa} , X_{c5} , X_{bu} , X_{pro} , X_{ac} and X_{h2}),
249 as these concentrations might not only strongly influence the biogas production during
250 the initial days of batch experiments, but might also be potentially interrelated among
251 themselves and/or to the previous θ_P .

252
 253 The sugars uptake yields butyrate ($f_{bu,su}$), propionate ($f_{pro,su}$), acetate ($f_{ac,su}$) and hydrogen
 254 ($f_{h2,su}$) as COD fractions in ADM1. Therefore, a maximum of three fractions can be
 255 selected simultaneously to fulfill the COD balance: $f_{bu,su} + f_{pro,su} + f_{ac,su} + f_{h2,su} = 1$. In
 256 this study, $f_{bu,su}$, $f_{ac,su}$ and $f_{h2,su}$ were selected, while $f_{pro,su}$ was recalculated: $f_{pro,su} = 1 -$
 257 $f_{bu,su} - f_{ac,su} - f_{h2,su}$. Importantly, further structural parameters and initial conditions need
 258 to be induced in the HS-AD model as, for example, the amino-acid (AA, S_{aa})
 259 fractioning and the biomass yield coefficients (Y_b), though these were not assessed here
 260 aiming to reduce the problem under study. In either case, the proposed methodology for
 261 calibration/validation can easily include any further θ .

262
 263 Variability ranges for structural parameters are suggested in ADM1 (Batstone et al.,
 264 2002). However, considerably wider ranges were assessed in this study to emphasize
 265 the absence of prior knowledge about the optimal values. For simplicity, all θ_P were
 266 allowed to vary by 90 % from their initial values, $\theta_{P,0}$: $(1 - 0.9) \cdot \theta_{P,0} \leq \theta_P \leq (1 +$
 267 $0.9) \cdot \theta_{P,0}$; uniform $p(\theta_P)$ [Table 4]. As the only exception, the lower pH threshold for
 268 acetoclastic methanogens ($pH_{LL,ac}$) was bounded between a pH value where
 269 methanogenesis potentially collapses (i.e. ≤ 5.0) and the upper pH threshold for
 270 acetoclastic methanogens ($pH_{UL,ac}$, i.e. 7.0), to maintain the suitability of the Hill
 271 function to simulate the pH inhibition [Table 2].

272
 273 Different methods are available to approximate the initial conditions (i.e. biomass
 274 concentrations) in batch simulations as, for example, simulating a continuous reactor
 275 fed with exactly the same substrate until reaching steady state, and then use these steady

276 conditions to initialize the batch simulations (Dochain & Vanrolleghem, 2001; Donoso-
 277 Bravo et al., 2011). However, since the inoculum was centrifuged right before setting
 278 the batch experiments in this study, absence of prior knowledge about the initial
 279 biomass concentrations was preferred. To explore homogeneously different orders of
 280 magnitude in the biomass concentrations (often $0 < \theta_B < 1 \text{ kmol COD/m}^3$), the
 281 logarithm-transformed θ_B were allowed to vary by 50 % from their initial values, $\theta_{B,0}$: (I
 282 $+ 0.5) \cdot \log_{10}(\theta_{B,0}) \leq \log_{10}(\theta_B) \leq (I - 0.5) \cdot \log_{10}(\theta_{B,0})$; uniform $p(\log_{10}(\theta_B))$ [Table 4].

283

284 2.3.3 Objective Function

285 The weighted sum of squares between all the available experimental and the
 286 corresponding simulated values was used as objective function, $J(\theta)$ [Equation 4]. $J(\theta)$
 287 was adapted from Flotats et al. (2003) to assess the model ‘goodness of fitting’, being: θ
 288 the structural parameters and/or initial conditions implemented in the HS-AD model; R
 289 the number of batch simulations; D the number of experimental datasets; t_{exp} the
 290 experimental time of each batch experiment; $y_{i,j,t}$ the experimental measurements;
 291 $y_{i,j,t}^{sim}(\theta)$ the simulated values; and $w_{i,j}$ the weighting coefficients – calculated as a
 292 function of the average experimental data, $\bar{y}_{i,j}$ [Equation 5]. With this approach, $J(\theta)$
 293 was lower-bounded by the – preliminarily unknown – global minimum: $J(\theta) \geq J(\theta_{opt})$.

$$J(\theta) = \sum_{i=1}^R \sum_{j=1}^D \sum_{t=t_1}^{t_{exp}} w_{i,j} \left(y_{i,j,t} - y_{i,j,t}^{sim}(\theta) \right)^2 \quad (4)$$

$$w_{i,j} = \frac{1}{\sum_{t=t_1}^{t_{exp}} (y_{i,j,t} - \bar{y}_{i,j})^2} \quad (5)$$

294

295 Noteworthy, only 6 out of 9 experimental conditions were used to calculate $J(\theta)$ in this
 296 study: Batch No. 1, 3, 5, 6, 7 and 9 [Table 1]. Meanwhile, 3 out of 9 experimental

297 conditions were used for cross-validation: Batch No. 2, 4 and 8 [Table 1]. Other
298 experimental combinations could be used for calibration/cross-validation as, for
299 example, Batch No. 2, 4, 5, 7 and 8 for calibration and Batch No. 1, 3, 6 and 9 for cross-
300 validation. However, the configuration used in this study permitted to include the most
301 extreme HS-AD conditions for model calibration (i.e. $TS \leq 10 \%$, $TS \geq 30 \%$), while
302 intermediate conditions (i.e. $10 \leq TS \leq 30 \%$) were used for cross-validation. This
303 strategy also aimed to enhance the number of experimental data used for calibration and
304 to ensure the most diverse operative conditions where θ might be representative.
305 Additionally, the proposed configuration included the most informative dataset (i.e.
306 Batch No. 1) to increase the complexity of $J(\theta)$ and remove identifiability issues related
307 to the lack of bio-physical-chemical dynamics in the overall dataset, as further discussed
308 in Section 3.1.3.

309

310 **2.3.4 Global Sensitivity Analysis**

311 GSA was aimed to highlight the most influential θ to be calibrated with the available set
312 of experimental data. For GSA, multiple θ combinations were used to evaluate $J(\theta)$
313 [Equation 4]. Latin-hypercube sampling (LHS) served to explore the global θ space
314 (Solon et al., 2015). Subsequently, $J(\theta)$ arrays and their corresponding θ were assessed
315 by the GSA engine of Kennedy and O'Hagan (2001) and Oakley and O'Hagan (2004).
316 The GSA engine calculates the individual (IE) and global (GE) effects of θ_i upon the
317 global model output (e.g. $p(y|\theta)$, $J(\theta)$) variance. On the other hand, the GSA engine
318 provides also all the double correlations between θ_i and θ_j (i.e. $\theta_i \cdot \theta_j$, being $i \neq j$). Both
319 IE and $\theta_i \cdot \theta_j$ are expressed as a percentage of the global variance explained. Thus, θ_i
320 showing a relatively low IE are associated to poor practical identifiability, since these θ_i

321 influence minimally the model output. Meanwhile, GE comprises IE, the double
322 correlations $\theta_i \cdot \theta_j$ and all the rest of multiple correlations for a single θ_i . Therefore, GE
323 itself should not be used to assess the identifiability. Instead, the overall correlation of θ_i
324 with θ can be inferred by the relative difference between IE and GE – though IE and GE
325 are expressed in different units, as interactions can be repeated in the GE of two or more
326 θ_i (Oakley & O'Hagan, 2004; Saltelli et al., 2006). Summarizing, both a low IE and a
327 large IE to GE difference mean poor identifiability for a single θ_i , though *a priori* no
328 fixed threshold – neither for IE, nor for GE minus IE – permits to distinguish whether θ_i
329 should be considered as non-identifiable.

330

331 A maximum θ subset of $N' = 30$ and/or 400 simulations of $J(\theta)$ can be evaluated
332 simultaneously with the GSA engine (Kennedy & Petropoulos, 2017). Therefore, to
333 assess θ interactions when $N > 30$, a combination strategy was followed. Firstly, 30 θ_i
334 were randomly selected and evaluated by GSA (i.e. GSA No. 1). From these θ , only
335 those showing the smallest IE (e.g. $< 1\%$) were disregarded as non-identifiable,
336 removed from the initial θ subset, and not used for further GSA. Importantly, these non-
337 identifiable θ_i were fixed at their initial values [Table 4] for all subsequent model
338 simulations, since non-identifiability implies that these θ can be fixed at any value
339 within $p(\theta)$ (Dochain & Vanrolleghem, 2001; Guisasola et al., 2006). Then, θ_P and/or θ_B
340 non-previously-assessed by GSA were combined with the non-removed θ subset, and a
341 new GSA was performed (i.e. GSA No. 2). The GSA methodology was repeated until
342 the last remaining θ subset was considered as ‘potentially identifiable’, θ' .

343

344 In total, seven GSA with different θ combinations were progressively performed [Table
 345 5]. In this study, the only criterion for non-identifiability was assumed as $IE \leq 0.20 \%$,
 346 which also coincided with a relatively low GE (i.e. < 4.5), though this conservative
 347 criterion could be modified as mentioned in Section 3.1.1. To enhance the GSA
 348 representativeness in presence of a high number of θ (i.e. $20 \leq N' \leq 30$) and/or wide
 349 variability ranges (i.e. $\pm 50 \%$), each GSA was conducted in triplicate and the results
 350 averaged. Finally, all $J(\theta)$ arrays used for GSA were searched for the minimum
 351 observed value, $J_{min}(\theta)$ (i.e. $\geq J(\theta_{opt})$), to be subsequently used in ABC.

352

353 **2.3.5 Approximate Bayesian Computation**

354 The θ' posterior distribution, $p(\theta'|y)$, was assessed by ABC (Toni & Stumpf, 2010). In
 355 short, multiple simulations were carried at different θ' combinations sampled by LHS,
 356 whereas relatively high $J(\theta')$ values were discarded by a progressively stringent
 357 criterion based in a tolerance coefficient, ϵ (i.e. > 1.0). In other words, only $J(\theta') -$
 358 $J_{min}(\theta') \leq \epsilon$ were accepted for posterior evaluation: $p(\theta'|J(\theta') - J_{min}(\theta') \leq \epsilon)$. With this
 359 approach, θ' identifiability was further assessed by the convergence in the confidence
 360 range.

361

362 In this study, ϵ was successively reduced from 2.50 to 1.05 (i.e. 2.50, 1.80, 1.30, 1.10
 363 and 1.05). Within each explored $J(\theta') - J_{min}(\theta') \leq \epsilon$ range, 400 simulations were used. θ'
 364 were allowed to vary within the same range used for GSA [Table 4]. Meanwhile, the 5
 365 to 95 % interquartile range of each θ' was used as confidence range, but also as a
 366 criterion for identifiability/convergence. The posterior mean, median, mode, Kurtosis,

367 Skewness and correlation matrix were also evaluated, as described in Martin and Ayesa
368 (2010).

369

370 **2.3.6 Cross-Validation**

371 Cross-validation assesses the model ‘goodness of fitting’ in experiments not used for
372 calibration (Bennett et al., 2013). In this study, the θ' posterior mean was considered as
373 θ_{opt} . Thus, θ_{opt} were used to simulate all batch experiments, including the three
374 experimental conditions selected for cross-validation: Batch No. 2, 4 and 8 [Table 1].

375

376

377 **3 RESULTS AND DISCUSSION**

378 **3.1 GSA – Selecting the Most Influencing Input Parameters for Calibration**

379 **3.1.1 Preliminary Identifiability Assessment**

380 GSA results are summarized in Table 5. GSA was started with 30 θ_P and progressively
381 led to only 14 θ' : 13 θ_P (i.e. $K_{h,pr}$, $K_{h,ch,slow}$, $k_{m,fa}$, $k_{m,c5}$, $k_{m,c4}$, $k_{m,pro}$, $k_{m,ac}$, $k_{m,h2}$, k_d , $pH_{LL,ac}$,
382 $f_{bu,su}$, $f_{ac,su}$, $f_{h2,su}$) and 1 θ_B (i.e. $X_{ac,Batch7}$). In this study, only the θ_i showing $IE \leq 0.20$ %
383 were fixed at their initial values [Table 4] to enhance the capabilities of GSA and ABC
384 for calibrating structured AD models, as mentioned in Section 2.3.4. The overall $J(\theta)$
385 variance explained by the GSA engine was around 70 % in all cases, confirming the
386 validity of this methodology to assess the most influential θ in the HS-AD model
387 (Oakley & O'Hagan, 2004).

388

389 IE provides a relative measure of the θ_i practical identifiability, while a high θ_i

390 correlation – high difference between IE and GE – suggests that θ_i cannot be

391 independently calibrated with the available set of experimental data (Saltelli et al.,
 392 2006). As an example, GSA No. 1 showed that 40.7 % out of 72.4 % of the global $J(\theta)$
 393 variance was explained by adding up the IE of 30 θ_P [Table 5], while the remaining
 394 31.7 % variance was explained by correlations among these. For example, $k_{m,c4}$
 395 explained individually around 4.4 % of the global variance (i.e. IE), though showing GE
 396 up to 17.9. This meant that the correlation of $k_{m,c4}$ with other θ was high, and any
 397 improvement in $J(\theta)$ obtained by solely modifying $k_{m,c4}$ could be partially or totally
 398 compensated by the modification of a correlated θ_i . In this line, the $k_{m,c4} \cdot f_{bu,bu}$
 399 correlation in GSA No. 1 explained up to 2.8 % of the global variance (data not shown),
 400 due to the influence of $f_{bu,su}$ in $J(\theta)$ (i.e. IE = 2.0 % and GE = 24.4) [Table 5].

401

402 In this study, θ_i were disregarded by a single and low-demanding criterion (i.e. IE \leq
 403 0.20 %) [Table 5], since any chosen criterion for ‘potential identifiability’ would
 404 influence the GSA results when $N \geq 30$. Meanwhile, GSA also depends on N' (i.e. ≤ 30)
 405 and/or the particular combination of θ_P and θ_B used. Thus, using a more demanding
 406 identifiability criterion (e.g. IE ≤ 0.50 % instead of 0.20 %) might have led to discard θ_i
 407 during preliminary GSA runs, which would be subsequently characterized as
 408 ‘potentially identifiable’. For example, GSA No. 1 showed an IE = 0.22 % for $k_{m,c5}$,
 409 whereas GSA No. 7 eventually showed an IE = 0.67 % [Table 5]. To enhance
 410 identifiability, a strategy to reduce the gap between IE and GE for each θ_i is required
 411 (Kennedy & Petropoulos, 2017). However, reducing the IE to GE gap in this study
 412 would require to readapt the criterion used for ‘potential identifiability’. One strategy
 413 might consist on fixing those θ_i showing, for example, IE ≤ 0.30 % in GSA No. 7 (i.e.
 414 $f_{ac,su}$), and then perform further rounds of GSA (i.e. GSA No. 8) until IE \sim GE for all θ_i .

415
416 With all the above, a second assessment for identifiability can be useful when using
417 variance-based GSA for structured AD models. ABC is a well-suited tool in this regard,
418 yielding also $p(\theta|y)$, in contrast to GSA. Importantly, both methodologies should yield
419 equivalent results regarding the θ_i identifiability, though ABC is much computationally
420 intensive than GSA, as explained in Section 3.2.1.

421

422 **3.1.2. Importance of the $p(\theta)$ for Model Calibration**

423 It must be noted that any $p(\theta)$ could be used to calibrate AD models, provided that θ do
424 not contradict biochemical laws (e.g. $\theta_P \geq 0$ and $\theta_B \geq 0$) or disrupt mathematical
425 resources (e.g. $\text{pH}_{\text{UL,ac}} > \text{pH}_{\text{LL,ac}}$) and the overall $p(\theta)$ range is as densely and
426 homogeneously sampled as possible. On the other hand, the $p(\theta)$ distribution used (e.g.
427 uniform, log-transformed uniform, normal) can be also crucial to determine $p(\theta|y)$ and
428 the overall θ estimation.

429

430 In this study, using triplicates permitted to enhance the GSA representativeness (overall
431 sampling) in presence of large N and $p(\theta)$ ranges. Meanwhile, a uniform distribution
432 with mostly a 90 % modification was predefined for θ_P , whereas a log-transformed
433 uniform distribution with a 50 % modification was allowed for θ_B [Table 4], as
434 mentioned in Section 2.3.2. These specific $p(\theta)$ ranges were considered sufficiently
435 wide for the objectives of this study, and were based mainly on experience and visual
436 analysis of the overall model results. Particularly, the order of magnitude of many θ_P are
437 relatively well characterized in literature according to their corresponding bio-physical-
438 chemical meaning. Nonetheless, the order of magnitude of θ_B is highly unknown,

439 particularly after the inoculum centrifugation used in this study. Thus, using a uniform
 440 $p(\theta)$ permitted to emphasize the precision over punctual θ_P values within the $p(\theta_P)$ range,
 441 whereas the precision over the ‘unknown’ order of magnitude in the θ_B range was
 442 emphasized by using a log-transformed uniform distribution $p(\log_{10}(\theta_B))$.

443

444 With all the above, using a different $p(\theta)$ range or distribution might alter the results of
 445 variance-based GSA and the overall calibration of structured AD models. For example,
 446 using considerably narrower $p(\theta_P)$ ranges in GSA No. 1 (i.e. $(1 - 0.3) \cdot \theta_{P,0} \leq \theta_P \leq (1 +$
 447 $0.3) \cdot \theta_{P,0}$; *uniform* $p(\theta_P)$) resulted in some θ_P – which were highlighted as ‘potentially
 448 identifiable’ in this study – being disregarded as non-identifiable (e.g. $K_{h,pr}$ and $k_{m,c4}$),
 449 likely because part of their ‘optimal’ or ‘sub-optimal’ θ_P values were left outside the
 450 $p(\theta_P)$ range [Supplementary Information]. On the other hand, using wider ranges (i.e.
 451 $0.01 \cdot \theta_{P,0} \leq \theta_P \leq 3 \cdot \theta_{P,0}$; *uniform* $p(\theta_P)$), the results were not substantially different; i.e.
 452 only the $k_{m,c5}$ identifiability was additionally disregarded by GSA No. 1. However,
 453 using a log-transformed uniform distribution (i.e. $(1 + 0.5) \cdot \log_{10}(\theta_{P,0}) \leq \log_{10}(\theta_P) \leq (1$
 454 $- 0.5) \cdot \log_{10}(\theta_{P,0})$; *uniform* $p(\log_{10}(\theta_P))$) relatively altered the results of GSA No. 1; i.e.
 455 the order of magnitude of $k_{m,c5}$ and $k_{m,c4}$ was disregarded, while the order of magnitude
 456 of $K_{h,ch,slow}$ and $K_{S,Xh2}$ has was highlighted as highly important for the HS-AD model
 457 calibration. Therefore, for the correct calibration of structured AD models, both the $p(\theta)$
 458 range and distribution should be set accordingly to all the prior information available for
 459 these θ , minimizing the $p(\theta)$ range explored alongside the number of simulations
 460 required for the objectives of the study.

461

462 3.1.3 Importance of the Available Data for Model Calibration

463 Provided the θ are structurally identifiable, practical identifiability relates to the quantity
464 as well as the quality (i.e. experimental errors and/or the sampling frequency) of the
465 experimental data available (Donoso-Bravo et al., 2011; Guisasola et al., 2006;
466 Nimmegeers et al., 2017; Rodriguez-Fernandez et al., 2006). Particularly, a reduced
467 number of experimental data associated to some model dynamics usually prevents
468 practical identifiability of the θ involved in these specific dynamics. For example, Y_b
469 might not be identifiable in AD models provided that the biomass concentration
470 dynamics were measured (Bernard et al., 2001).

471

472 In this study, the hydrolysis constant of readily-biodegradable carbohydrates ($K_{h,ch,fast}$)
473 and lipids ($K_{h,li}$), but also the maximum growth rate of sugar ($k_{m,su}$) and amino acid
474 ($k_{m,aa}$) degraders, showed a reduced influence in $J(\theta)$ by GSA. These results suggest that
475 either insufficient experimental data was available to calibrate $K_{h,ch,fast}$, $K_{h,li}$, $k_{m,su}$ and
476 $k_{m,aa}$, or that the biogas production in the HS-AD batch experiments [Table 1] was
477 mostly influenced by the VFA uptake – as the main limiting step. In the same line, due
478 to the Monod properties, k_m and K_s might be correlated when using batch experiments
479 for calibration (Dochain & Vanrolleghem, 2001; Flotats et al., 2010; Guisasola et al.,
480 2006). Nevertheless, GSA showed negligible influence for all K_s in this study, likely
481 due to using different batch experiments (i.e. ISR and TS) and/or a reduced number of
482 experimental data to obtain $J(\theta)$ [Equation 4].

483

484 The liquid-gas transfer coefficient (kLa) was also disregarded during the initial steps of
485 GSA (i.e. $IE = 0.07\%$ and $GE = 0.39$) [Table 5]. The kLa coefficient is related to the
486 homogenization and mixing strategy, as well as other operational parameters in AD (e.g.

487 temperature and pH) (Batstone et al., 2002). In this study, all the HS-AD digesters were
488 manually shaken. However, increasing the TS content can hamper the liquid-gas
489 transfer mechanisms in HS-AD (Pastor-Poquet et al., 2019a). Therefore, a calibration
490 strategy should be specifically envisaged to correctly calibrate kLa (e.g. using different
491 stirring velocities/devices for mixing), since the presence of several θ in HS-AD models
492 can prevent identifiability of kLa with the reduced number of experimental data usually
493 available.

494

495 Importantly, both the NH_3 and H_2 inhibition parameters [Table 2] were shown as non-
496 identifiable in this study, despite the strong influence of these parameters to regulate the
497 biogas production in an ADM1-based model, as mentioned in Section 2.3.2. These
498 results were associated to the reduced TAN and VFA dynamics in the experimental data,
499 since only one single sacrifice experiment was available for calibration/validation.

500 Therefore, despite using different initial conditions (i.e. ISR and/or TS) for model
501 calibration, the NH_3 inhibition parameters in HS-AD of OFMSW cannot be assessed by
502 using traditional batch experiments, where only the biogas production and composition
503 are (usually) dynamically evaluated.

504

505 The above results condense the importance of an adequate sampling to enhance
506 identifiability in AD models, but also to test hypotheses regarding the effects of
507 inhibitory substances in HS-AD. Particularly, an extensive sampling for VFA, pH and
508 TAN at different operational times during batch experiments is required to identify
509 crucial parameters regarding the NH_3 inhibition in HS-AD of OFMSW. Therefore,
510 sacrifice experiments and/or any sampling technique for batch setups – allowing the

511 thorough characterization of the reactor content variables in dynamic mode – should be
 512 recommended to calibrate structured HS-AD models using batch experiments. On the
 513 other hand, including (semi-)continuous datasets when available might also alleviate
 514 these identifiability issues during calibration (Bennett et al., 2013; Nimmegeers et al.,
 515 2017).

516

517 **3.1.4 The Importance of Initial Conditions for Model Calibration**

518 Interestingly, all θ_B except $X_{ac, Batch7}$ were shown as non-identifiable in this study [Table
 519 5]. The reason presumably lies on the high k_m of all microorganisms ‘shading’ the effect
 520 of their initial concentration. For example, X_{pro} was associated to a maximum growth
 521 rate ($k_{m, pro}$) around 10 d^{-1} [Table 4]. Thus, X_{pro} doubles within 1 h (i.e. $\frac{\log(2)}{k_m} =$
 522 $\frac{\log(2) \cdot 24}{10} = 0.6 \text{ h}$), whereas the HS-AD batch experiments lasted considerably longer
 523 than 20 days [Table 1].

524

525 X_{ac} and X_{h2} are important variables to avoid batch acidification during the initial 0 - 10
 526 days of HS-AD simulations, due to the rapid changes occurring in the bio-physical-
 527 chemistry during these days and the influence of these two microbial populations to
 528 define the buffering capacity (Batstone et al., 2002; Capson-Tojo et al., 2017).

529 Nonetheless, these biomass concentrations were also rapidly disregarded by GSA in this
 530 study [Tables 3 and 4], except the ‘potential identifiability’ of $X_{ac, Batch7}$ suggested by
 531 GSA No. 7 which was likely explained by the influence of the order of magnitude of
 532 this particular biomass content to regulate the risk of acidification of the most extreme
 533 HS-AD condition in Experiment 7 (i.e. Batch No. 9, TS = 30 %) [Table 1].

534

535 With all the above, it is likely that only the initial biomass magnitude – not a precise
 536 value – was needed to calibrate the HS-AD model based on batch experiments. In other
 537 words, approximate biomass concentrations serve mainly to avoid acidification in HS-
 538 AD batch simulations, since these might not influence significantly the model
 539 calibration. The θ influence on $J(\theta)$ in this study was further assessed by ABC.

540

541 **3.2 Parameter Optimization**

542 **3.2.1 Second Identifiability Assessment**

543 Figure 1 shows $p(\theta|y)$ using an $\epsilon = 1.05$. The main statistics for these $p(\theta|y)$ are
 544 summarized in Table 6, including the confidence ranges (i.e. 5 - 95 % interquartile
 545 range), since a reliable assessment of the θ confidence range is as important as the value
 546 of θ_{opt} themselves (Guisasola et al., 2006). The correlation matrix is included as
 547 Supplementary Information. Figure 2 shows the 5 - 95 % interquartile range vs. ϵ , since
 548 reducing progressively ϵ permitted to assess the convergence of the θ' posterior as
 549 second identifiability assessment.

550

551 Parameter identifiability is roughly associated to the ‘sharpness’ of the posterior
 552 distribution, $p(\theta|y)$ (Martin & Ayesa, 2010; Toni & Stumpf, 2010). In this line, $K_{h,pr}$,
 553 $K_{h,ch,slow}$, $k_{m,fa}$, $k_{m,c5}$, $k_{m,c4}$, $k_{m,ac}$, $pH_{LL,ac}$ and $f_{bu,su}$ showed relatively well-defined bell-
 554 shaped distributions by ABC, suggesting an adequate identifiability [Figure 1].
 555 Meanwhile, $k_{m,pro}$, $k_{m,h2}$, k_d , $X_{ac,Batch7}$, $f_{ac,su}$ and $f_{h2,su}$ showed a more uniform-like
 556 distribution, suggesting a poorer identifiability. The substantial reduction observed in
 557 the interquartile range for $K_{h,pr}$, $K_{h,ch,slow}$, $k_{m,fa}$, $k_{m,c5}$, $k_{m,c4}$, $k_{m,ac}$, $pH_{LL,ac}$ and $f_{bu,su}$ (i.e. 60
 558 - 80 %) corroborated their adequate identifiability in this study, in contrast to $k_{m,pro}$,

559 $k_{m,h2}$, k_d , $X_{ac,Batch7}$, $f_{ac,su}$ and $f_{h2,su}$ that showed a much constant interquartile range (i.e. \leq
 560 50 % reduction) [Figure 2]. The poor practical identifiability of these last θ' is explained
 561 by their high correlation with the rest of θ' . For example, the $f_{ac,su} \cdot f_{bu,su}$ correlation was -
 562 0.82, while $k_{m,pro} \cdot k_d$ was 0.72 – data not shown. As suggested in Section 3.1.4, $X_{ac,Batch7}$
 563 served mainly to counteract the potential acidification in Batch No. 9, since the poor
 564 reduction in the interquartile range (i.e. 23 %) alongside the high correlation with other
 565 θ' (i.e. $pH_{LL,ac} \cdot X_{ac,Batch7} = 0.24$) indicated that only an approximate biomass content is
 566 needed to calibrate structured HS-AD models based on batch experiments.

567

568 As expected, ABC supported the identifiability assessment previously performed by
 569 GSA. In particular, θ' showing IE < 1.5 % in GSA No. 7 (i.e. $k_{m,pro}$, $k_{m,h2}$, and $f_{ac,su}$)
 570 [Table 5] were associated to a poor identifiability. However, some parameters showing
 571 an IE ≥ 1.5 % in GSA No. 7 (i.e. k_d , $X_{ac,Batch7}$ and $f_{h2,su}$) were also indicated as non-
 572 identifiable by ABC in contrast to GSA, suggesting that ABC was a more sensitive
 573 methodology for parameter identifiability in this study. With all the above, a more
 574 restrictive IE threshold (i.e. 0.50 % instead of 0.20 %) could have been used in further
 575 GSA rounds, once fixing poorly-identifiable parameters to any value within the prior, as
 576 mentioned in Section 3.1.1.

577

578 ABC is computationally intensive due to the high level of $J(\theta') - J_{min}(\theta') \leq \epsilon$ rejection,
 579 particularly when using highly-demanding ϵ (Filippi et al., 2013; Toni & Stumpf, 2010).
 580 For example, the acceptance ratio was 0.129 when using $\epsilon = 1.80$, meaning that only 1
 581 out of 8 simulations was accepted for posterior evaluation, whereas the acceptance ratio
 582 was 0.004 when using $\epsilon = 1.10$ – data not shown. Thus, ABC is not recommended to

583 assess identifiability in complex models with a large number of θ (i.e. $N \geq 30$). Different
584 upgrades have been proposed to increase the ABC efficiency (Beaumont et al., 2009;
585 Filippi et al., 2013; Toni & Stumpf, 2010), though the evaluation of these upgrades for
586 calibrating structured AD models was out of the scope of this study. Conversely, the
587 GSA engine relies upon a Bayesian emulator to speed up the analysis of model outputs
588 (Kennedy & O'Hagan, 2001; Oakley & O'Hagan, 2004). Therefore, GSA can be an
589 adequate tool to reduce the global computation required for parameter optimization, by
590 preliminarily reducing the number of θ' to be further assessed by ABC as shown in this
591 study.

592

593 **3.2.2 Batch Simulations**

594 Using the θ' mean as θ_{opt} [Table 6] led to a good approximation of both the methane
595 production [Figure 3] and the rest of variables at the end of all batch experiments
596 [Figure 4] used either for calibration (i.e. Batch No. 1, 3, 5, 6, 7 and 9) or cross-
597 validation (i.e. Batch No. 2, 4 and 8) [Table 1]. Therefore, the θ' mean might be a good
598 approximation of θ_{opt} , particularly for those θ' where practical identifiability was likely
599 (i.e. $K_{h,pr}$, $K_{h,ch,slow}$, $k_{m,fa}$, $k_{m,c5}$, $k_{m,c4}$, $k_{m,ac}$, $pH_{LL,ac}$ and $f_{bu,su}$). Importantly, the HS-AD
600 model was able to capture particularly well the TS and TAN contents, but also VS (data
601 not shown), in all mono- and co-digestion experiments, confirming the suitability of the
602 hypotheses used for the model development (Pastor-Poquet et al., 2018).

603

604 Some disagreements were also observed between the simulations and the experimental
605 results. Particularly, the implementation of a reversible NH_3 inhibition function
606 [Equation 1] in all the VFA and H_2 degrading populations [Table 2] was unable to

607 capture the VFA accumulation at the end of HS-AD experiments [Figure 4]. As the
608 most noticeable example, calibration failed to represent the S_{bu} accumulation in Batch
609 No. 6 [Figure 4e], yielding also a slight miss-adjustment in the methane production
610 [Figure 3c]. More in general, S_{ac} and S_{bu} at the end of all experiments were poorly
611 represented [Figure 4], despite the butyrate ($k_{m,c4}$) and acetate ($k_{m,ac}$) growth rates were
612 adequately identified and the NH_3 inhibition upon the acetate uptake is a relatively well-
613 established strategy in structured AD models (Batstone et al., 2002).

614

615 Two main reasons might explain the VFA disagreement between the model simulations
616 and the experimental data available. The first reason relates to the relatively low amount
617 of experimental data hampering calibration, as mentioned in Section 3.1.3. In this line,
618 the NH_3 half-inhibition parameters in the VFA and/or H_2 uptakes ($K_{i,nh3}$) were
619 disregarded as unimportant by GSA to represent the experimental data [Table 5], mainly
620 because only Batch No. 1 contained the VFA, pH and TAN dynamics. The second
621 reason relates to the poor suitability of the reversible non-competitive NH_3 inhibition
622 function [Equation 1] to explain the VFA accumulation in HS-AD simulations, as
623 discussed in next section.

624

625 **3.3 Main Effects of Increasing the TS Content in HS-AD of OFMSW**

626 In this study, calibration/cross-validation served to further test the hypotheses used for
627 model construction (e.g. mass balances), particularly regarding the TS and VS
628 simulation. Noteworthy, the correct simulation of TS is crucial in HS-AD, as TS
629 determines the apparent concentration of soluble compounds subsequently affecting all
630 the HS-AD bio-physical-chemical dynamics (Pastor-Poquet et al., 2019b). For example,

631 TS = 20 % supposes approximately 20 % higher apparent concentrations (i.e. kmol/kg
632 H₂O) regarding the corresponding global concentrations (i.e. kmol/kg).

633

634 The HS-AD model was also calibrated/validated to assess the effects of increasing TS
635 upon the NH₃ inhibition in HS-AD. Specifically, a high solute content – potentially
636 associated to a high TS – exacerbates the solution ‘non-ideality’, affecting all the HS-
637 AD dynamics (e.g. pH, NH₃ concentration, CO₂ transfer) (Pastor-Poquet et al., 2019b).
638 More in detail, ‘non-ideality’ can lower S_{nh3} , serving as a potential source of NH₃
639 inhibition abatement in HS-AD of OFMSW. In this study, I ranged from 0.22 to 0.93 M
640 [Figure 4c], emphasizing the need for an adequate ‘non-ideal’ bio-physical-chemical
641 approach (Hafner & Bisogni, 2009; Solon et al., 2015). Importantly, despite the high I
642 observed, S_{nh3} reached up to 0.13 mol N/kg in this study [Figure 4d] – equivalent to
643 0.16 mol N/kg H₂O (i.e. 2.3 g N/L). These S_{nh3} were considerably high, since reactors
644 operated at $S_{nh3} \geq 1.0$ g N/L usually show an inefficient VFA conversion (Rajagopal et
645 al., 2013).

646

647 The inefficient VFA conversion in HS-AD experiments was not well simulated by the
648 reversible NH₃ inhibition function, as mentioned in Section 3.2.2. To understand the
649 poor VFA simulation, it is necessary to consider the relatively flat inhibition described
650 by Equation 1 but also the COD fluxes in the HS-AD model. Briefly, I_{nh3} is 0.50 when
651 $S_{nh3,App} = K_{i,Snh3}$, whereas I_{nh3} is 0.33 when $S_{nh3,App} = 2 \cdot K_{i,Snh3}$ [Equation 1]. In other
652 words, a non-competitive reversible inhibition by NH₃ might be far too ‘blunt’ to
653 describe the actual effect of NH₃ upon the anaerobic biomass (Astals et al., 2018). On
654 the other hand, due to the COD fractioning used in this study, approximately 54 % of

655 the COD from S_{su} [Table 6] and 26 % of the COD from S_{aa} flowed through the butyrate
656 pathway, whereas up to 80 % of the COD either from S_{su} , S_{aa} and/or long-chain fatty
657 acids (S_{fa}) flowed through the acetate pathway (Batstone et al., 2002). Meanwhile, a
658 considerable proportion of the initially biodegradable COD (i.e. 75 - 85 %) was
659 assigned to $X_{ch} + S_{su} + X_{pr} + S_{aa}$ [Table 3].

660

661 With all the above, the ‘blunt’ definition of the NH_3 inhibition function, alongside the
662 high COD flowing through the butyrate and acetate pathways, presumably favored the
663 X_{bu} and X_{ac} growth even at considerably high S_{nh3} (i.e. up to 2.3 g N/L) during
664 simulations. Summarizing, the high substrate content counterbalanced the effect of the
665 NH_3 inhibition, preventing the correct simulation of S_{bu} and S_{ac} accumulation at the end
666 of the HS-AD experiments. Therefore, the reversible non-competitive NH_3 inhibition
667 function [Equation 1] in the VFA and/or H_2 uptakes [Table 2] requires further testing to
668 represent the VFA accumulation observed in HS-AD of OFMSW. To this particular aim,
669 using extensive data regarding the VFA, pH and TAN dynamics in HS-AD simulations
670 is strongly recommended.

671

672 To end up, the HS-AD model (Pastor-Poquet et al., 2018) is a suitable platform to
673 understand the inner mechanisms of HS-AD, whereas further model developments
674 and/or model configurations (i.e. inhibition functions) should be also tested to enhance
675 our understanding about the VFA accumulation within HS-AD of OFMSW. Similarly,
676 additional experimental data is needed to thoroughly understand the role of NH_3
677 inhibition in HS-AD simulations. Specifically, extensive data regarding the main
678 species driving ‘non-ideality’ (i.e. VFA, pH, TAN) and/or further bio-physical-chemical

679 mechanisms (i.e. precipitation) in HS-AD seem to be crucial, due to the outstanding
680 importance of ‘non-ideality’ for the biochemical parameter optimization (Pastor-Poquet
681 et al., 2019b). In this scheme, the relatively simple calibration/validation methodology
682 presented in this study can serve to test the θ practical identifiability and confidence
683 ranges in presence of any set of experimental data and/or HS-AD model structure.

684

685

686 **4 CONCLUSIONS**

687 Nine different batch conditions were used to calibrate and cross-validate the HS-AD
688 model. For parameter optimization, variance-based GSA in tandem with ABC served to
689 evaluate the practical identifiability of 35 θ_P and 32 θ_B . Among all these, mostly 8 θ_P
690 were correctly identified with the available data, as corroborated by the convergence of
691 $p(\theta|y)$. The study also showed that HS-AD may be operated at $S_{\text{nh}_3} \geq 2.3$ g N/L and $I \geq$
692 0.9 M, whereas a reversible non-competitive NH_3 inhibition function was not able to
693 explain the VFA accumulation in HS-AD of OFMSW. Therefore, further datasets about
694 the VFA, pH and TAN dynamics are required to enhance the θ practical identifiability,
695 whereas further model configurations should be tested to enhance the simulation of NH_3
696 inhibition in HS-AD.

697

698

699 **Acknowledgements**

700 This project has received funding from the European Union’s Horizon 2020 research
701 and innovation programme under the Marie Skłodowska-Curie grant agreement No.

702 643071. Angela Andreella is gratefully acknowledged for revising all the statistical
703 concepts.

704

705

706

707

708 REFERENCES

709

- 710 Astals, S., Peces, M., Batstone, D.J., Jensen, P.D., Tait, S. 2018. Characterising and
711 modelling free ammonia and ammonium inhibition in anaerobic systems. *Water*
712 *Res.*, **143**, 127-135.
- 713 Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A.,
714 . . . Vavilin, V.A. 2002. The IWA Anaerobic Digestion Model No. 1 (ADM1).
715 *Water Sci. Technol.*, **45**(10), 65-73.
- 716 Beaumont, M.A., Cornuet, J.-M., Marin, J.-M., Robert, C.P. 2009. Adaptive
717 approximate Bayesian computation. *Biometrika*, **96**(4), 983-990.
- 718 Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H.,
719 Jakeman, A.J., . . . Andreassian, V. 2013. Characterising performance of
720 environmental models. *Environ. Model. Software*, **40**, 1-20.
- 721 Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., Steyer, J.P. 2001. Dynamical
722 model development and parameter identification for an anaerobic wastewater
723 treatment process. *Biotechnol. Bioeng.*, **75**, 424-438.
- 724 Capson-Tojo, G., Trably, E., Rouez, M., Crest, M., Steyer, J.P., Delgenès, J.P., Escudié,
725 R. 2017. Dry anaerobic digestion of food waste and cardboard at different
726 substrate loads, solid contents and co-digestion proportions. *Bioresour. Technol.*,
727 **233**, 166-175.
- 728 Dochain, D., Vanrolleghem, P. 2001. *Dynamical Modelling and Estimation in*
729 *Wastewater Treatment Processes*. IWA Publishing, London, UK.
- 730 Donoso-Bravo, A., Mailier, J., Martin, C., Rodríguez, J., Aceves-Lara, C.A., Vande
731 Wouwer, A. 2011. Model selection, identification and validation in anaerobic
732 digestion: A review. *Water Res.*, **45**(17), 5347-5364.
- 733 Filippi, S., Barnes, C.P., Cornebise, J., Stumpf, M.P. 2013. On optimality of kernels for
734 approximate Bayesian computation using sequential Monte Carlo. *Stat. Appl.*
735 *Genet. Mol. Biol.*, **12**(1), 87-107.
- 736 Flotats, X., Ahring, B.K., Angelidaki, I. 2003. Parameter identification of thermophilic
737 anaerobic degradation of valerate. *Appl. Biochem. Biotechnol.*, **109**(1-3), 47-62.
- 738 Flotats, X., Palatsi, J., Fernandez, B., Colomer, M.A., Illa, J. 2010. Identifying
739 anaerobic digestion models using simultaneous batch experiments. *Environ.*
740 *Engineer. Manag. J.*, **9**(3), 313-318.
- 741 Garcia-Gen, S., Sousbie, P., Rangaraj, G., Lema, J.M., Rodriguez, J., Steyer, J.P.,
742 Torrijos, M. 2015. Kinetic modelling of anaerobic hydrolysis of solid wastes,
743 including disintegration processes. *Waste Manage.*, **35**, 96-104.

- 744 Guisasola, A., Baeza, J.A., Carrera, J., Sin, G., Vanrolleghem, P.A., Lafuente, J. 2006.
 745 The influence of experimental data quality and quantity on parameter estimation
 746 accuracy: Andrews inhibition model as a case study. *Educ. Chem. Eng.*, **1**(1),
 747 139-145.
- 748 Hafner, S.D., Bisogni, J.J. 2009. Modeling of ammonia speciation in anaerobic
 749 digesters. *Water Res.*, **43**(17), 4105-4114.
- 750 Kennedy, M.C., O'Hagan, A. 2001. Bayesian calibration of computer models. *J. Roy.*
 751 *Stat. Soc. Ser. B. (Stat. Method.)*, **63**(3), 425-464.
- 752 Kennedy, M.C., Petropoulos, G.P. 2017. GEM-SA: The gaussian emulation machine for
 753 sensitivity analysis. in: *Sensitivity Analysis in Earth Observation Modelling*,
 754 (Ed.) G.P. Petropoulos, Elsevier. Amsterdam, Netherlands.
- 755 Martin, C., Ayesa, E. 2010. An Integrated Monte Carlo Methodology for the calibration
 756 of water quality models. *Ecol. Model.*, **221**(22), 2656-2667.
- 757 Nimmegeers, P., Lauwers, J., Telen, D., Logist, F., Impe, J.V. 2017. Identifiability of
 758 large-scale non-linear dynamic network models applied to the ADM1-case
 759 study. *Math. Biosci.*, **288**, 21-34.
- 760 Nott, D.J., Marshall, L., Brown, J. 2012. Generalized likelihood uncertainty estimation
 761 (GLUE) and approximate Bayesian computation: What's the connection? *Water*
 762 *Resour. Res.*, **48**(12).
- 763 Oakley, J.E., O'Hagan, A. 2004. Probabilistic sensitivity of complex models: A
 764 Bayesian approach. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)*, **66**(3), 751-769.
- 765 Pastor-Poquet, V., Papirio, S., Steyer, J.-P., Trably, E., Escudié, R., Esposito, G. 2018.
 766 High-solids anaerobic digestion model for homogenized reactors. *Water Res.*,
 767 **142**, 501-511.
- 768 Pastor-Poquet, V., Papirio, S., Trably, E., Rintala, J., Escudié, R., Esposito, G. 2019a.
 769 High-Solids Anaerobic Digestion requires a tradeoff between Total Solids,
 770 Inoculum-to-Substrate Ratio and Ammonia Inhibition. *Int. J. Environ. Sci.*
 771 *Technol.*, 1-15.
- 772 Pastor-Poquet, V., Papirio, S., Steyer, J.-P., Trably, E., Escudié, R., Esposito, G. 2019b.
 773 Modelling Non-Ideal Bio-Physical-Chemical Effects on High-Solids Anaerobic
 774 Digestion of the Organic Fraction of Municipal Solid Waste. *J. Environ.*
 775 *Manage.*, **238**, 408-419.
- 776 Rajagopal, R., Masse, D.I., Singh, G. 2013. A critical review on inhibition of anaerobic
 777 digestion process by excess ammonia. *Bioresour. Technol.*, **143**, 632-641.
- 778 Ratto, M., Tarantola, S., Saltelli, A. 2001. Sensitivity analysis in model calibration:
 779 GSA-GLUE approach. *Comput. Phys. Commun.*, **136**, 212-224.
- 780 Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R. 2006. Novel metaheuristic for
 781 parameter estimation in nonlinear dynamic biological systems. *BMC*
 782 *Bioinformatics*, **7**, 483.
- 783 Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F. 2006. Sensitivity analysis
 784 practices: Strategies for model-based inference. *Reliab. Eng. Syst. Saf.*, **91**(10-
 785 11), 1109-1125.
- 786 Solon, K., Flores-Alsina, X., Mbamba, C.K., Volcke, E.I., Tait, S., Batstone, D., . . .
 787 Jeppsson, U. 2015. Effects of ionic strength and ion pairing on (plant-wide)
 788 modelling of anaerobic digestion. *Water Res.*, **70**, 235-245.
- 789 Toni, T., Stumpf, M.P. 2010. Simulation-based model selection for dynamical systems
 790 in systems and population biology. *Bioinformatics*, **26**(1), 104-110.
 791

792 **TABLE CAPTIONS**

793 **Table 1.** Batch experiments and initial conditions used for HS-AD model calibration
794 and cross-validation

795 **Table 2.** Biochemical rates used for the HS-AD model in this study

796 **Table 3.** Initial conditions used for all batch simulations in this study

797 **Table 4.** Main input parameters (θ) used for global sensitivity analysis (GSA), including
798 the initial values, lower and upper thresholds

799 **Table 5.** Global sensitivity analysis (GSA) of input parameters (θ): Individual (IE) and
800 global (GE) effects upon the objective function, $J(\theta)$, variance

801 **Table 6.** Calibration of potentially-identifiable input parameters (θ'): Prior and posterior
802 distributions

803

804

805 **FIGURE CAPTIONS**

806

807 **Figure 1.** Posterior parameter distribution using 400 simulations and $\varepsilon = 1.05$ 808 **Figure 2.** Interquartile range (percentiles 5 to 95 %) of the posterior parameter809 distribution using $\varepsilon \geq 1.05$ and $\varepsilon \leq 2.50$ 810 **Figure 3.** Methane production with mono-digestion of dried OFMSW at a) $ISR = 1.00$;811 b) $ISR = 1.50$; and c) $ISR = 1.00$; and co-digestion of dried OFMSW and sawdust at d)812 $ISR = 0.16$. Dots represent experimental data, while lines represent simulated values.813 **Figure 4.** Main variables at the end of the four batch experiments: a) Total solids (TS);814 b) total ammonia nitrogen (TAN, S_{in}); c) ionic strength (I); d) free ammonia nitrogen815 (NH_3 , S_{nh3}); e) acetic acid (S_{ac}); f) propionic acid (S_{pro}); g) butyric acid (S_{bu}); and h)816 valeric acid (S_{va}). Crosses represent experimental data, while geometries represent

817 simulated values

818

819

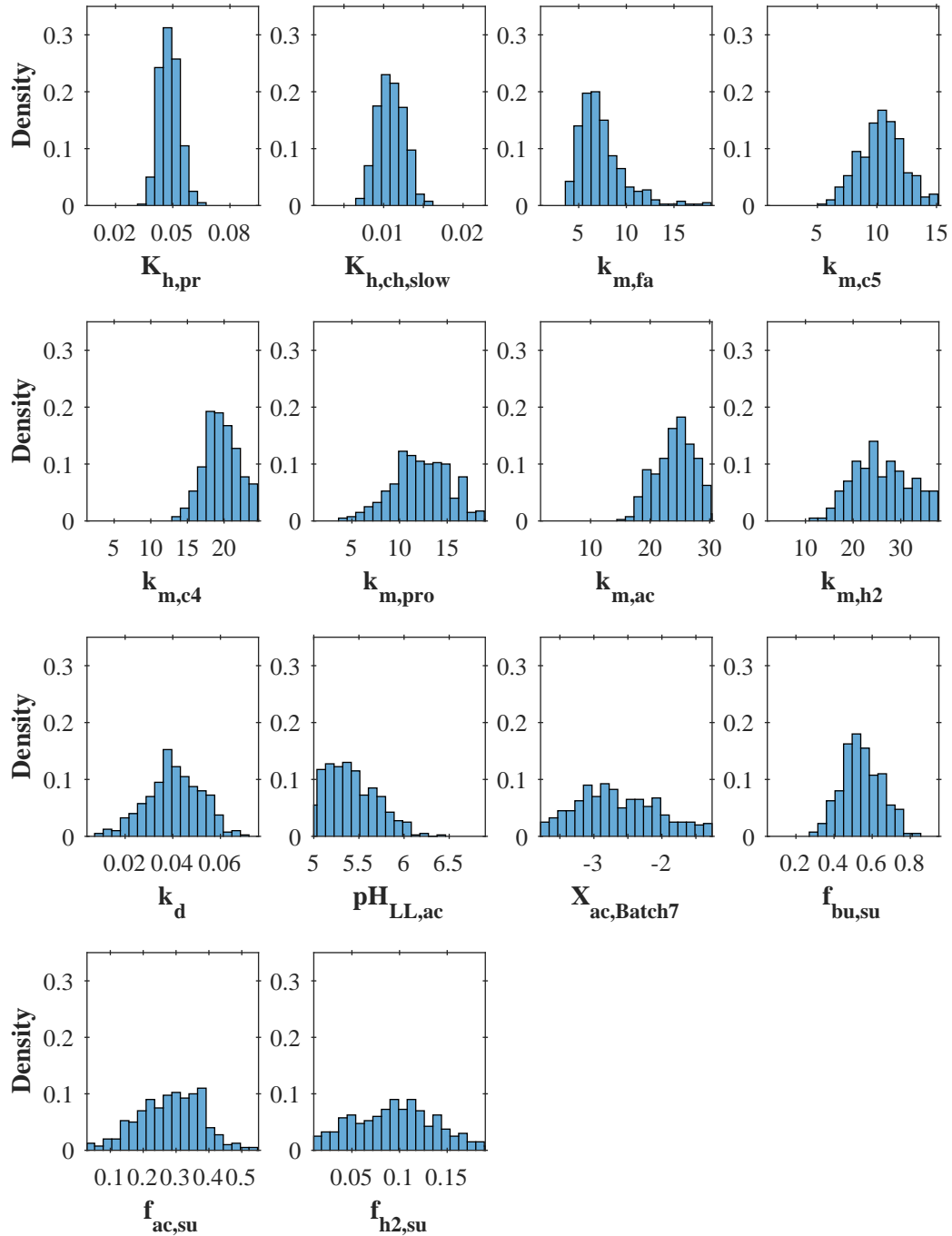


Figure 1. Posterior parameter distributions using 400 simulations and $\epsilon = 1.05$

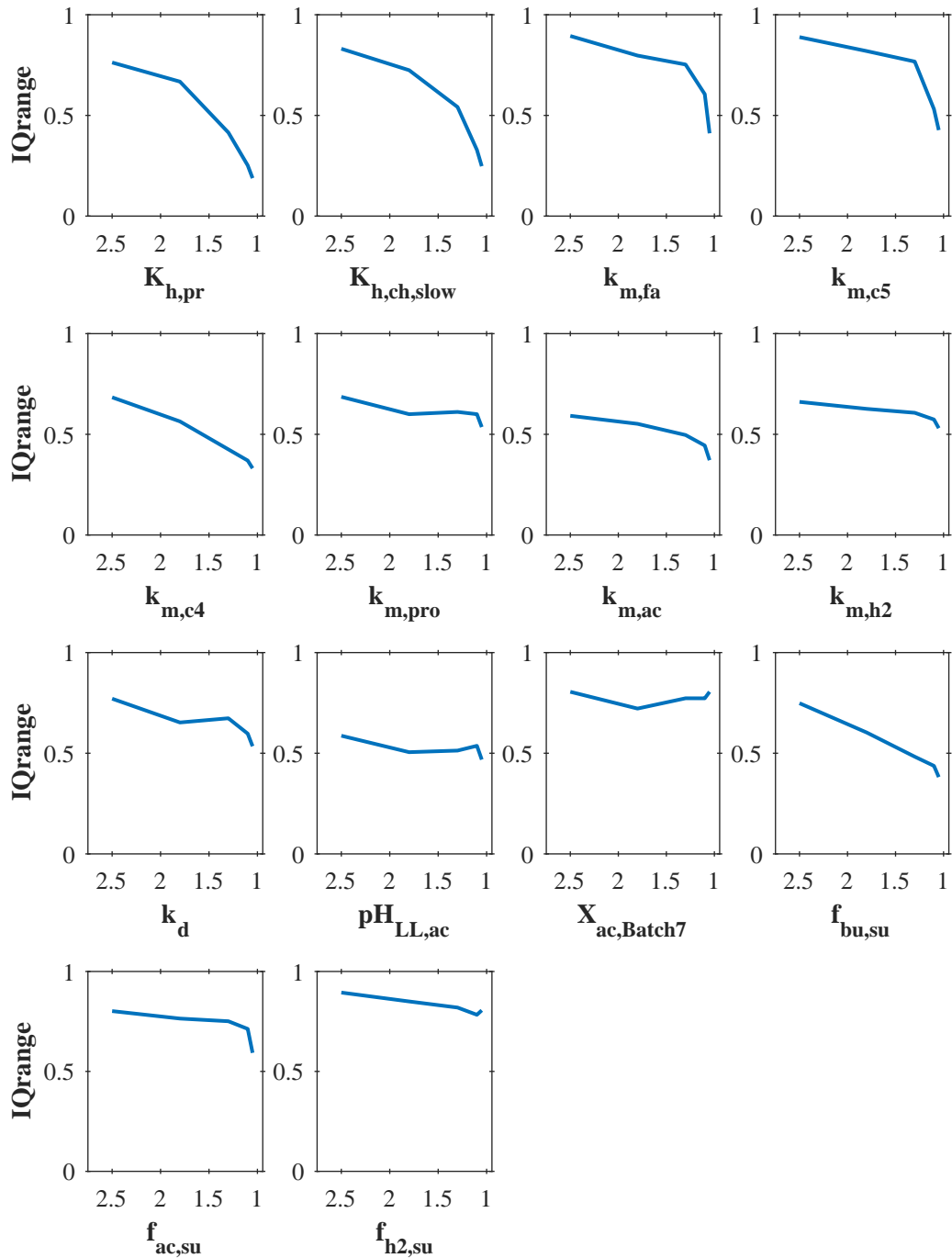


Figure 2. Interquartile range (percentiles 5 to 95 %) of the posterior parameter distributions using $\epsilon \geq 1.05$ and $\epsilon \leq 2.50$

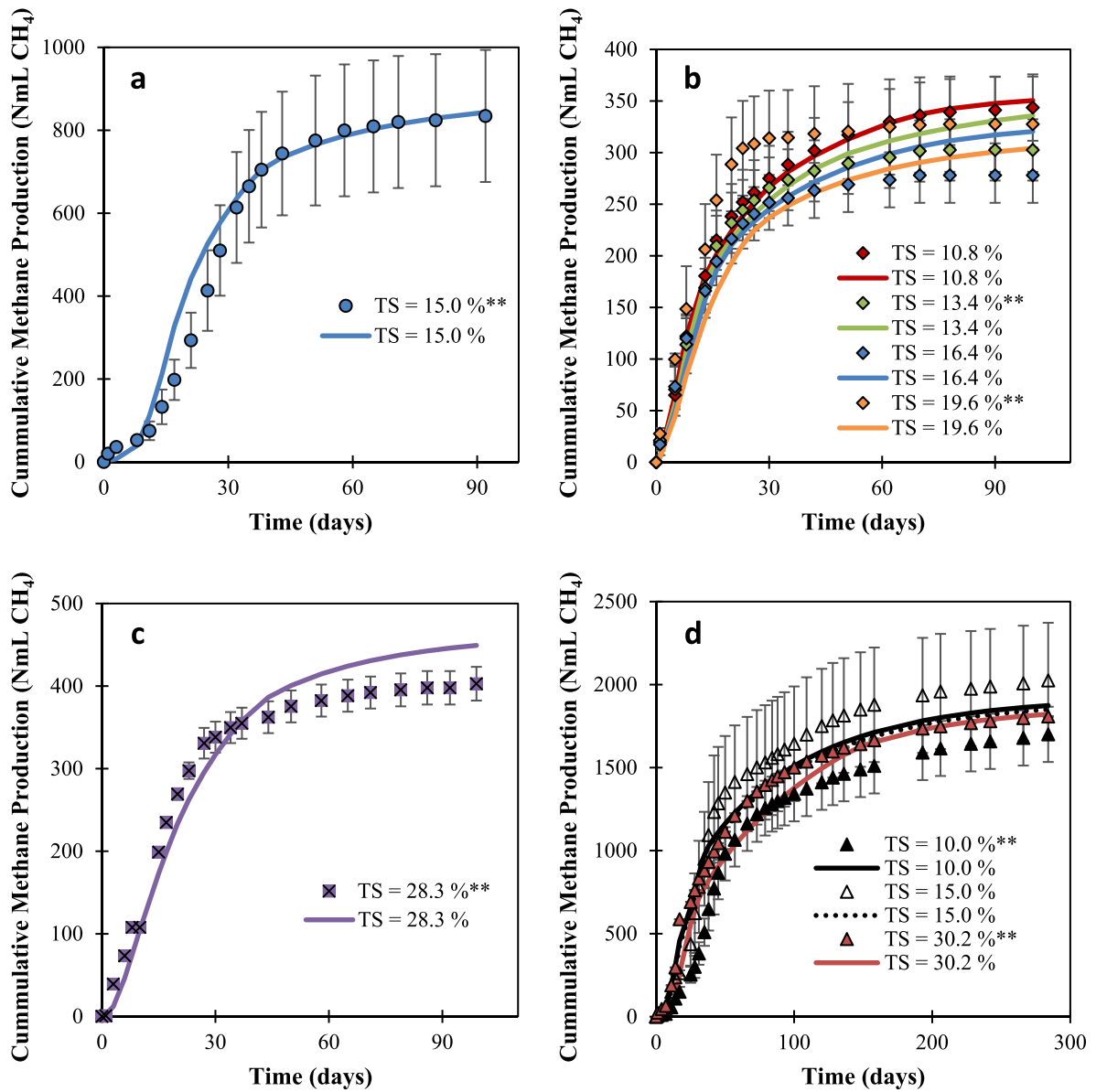


Figure 3. Methane production with mono-digestion of dried OFMSW at a) ISR = 1.00; b) ISR = 1.50; and c) ISR = 1.00; and co-digestion of dried OFMSW and sawdust at d) ISR = 0.16. Dots represent experimental data, while lines represent simulated values.

**Experimental data used for calibration

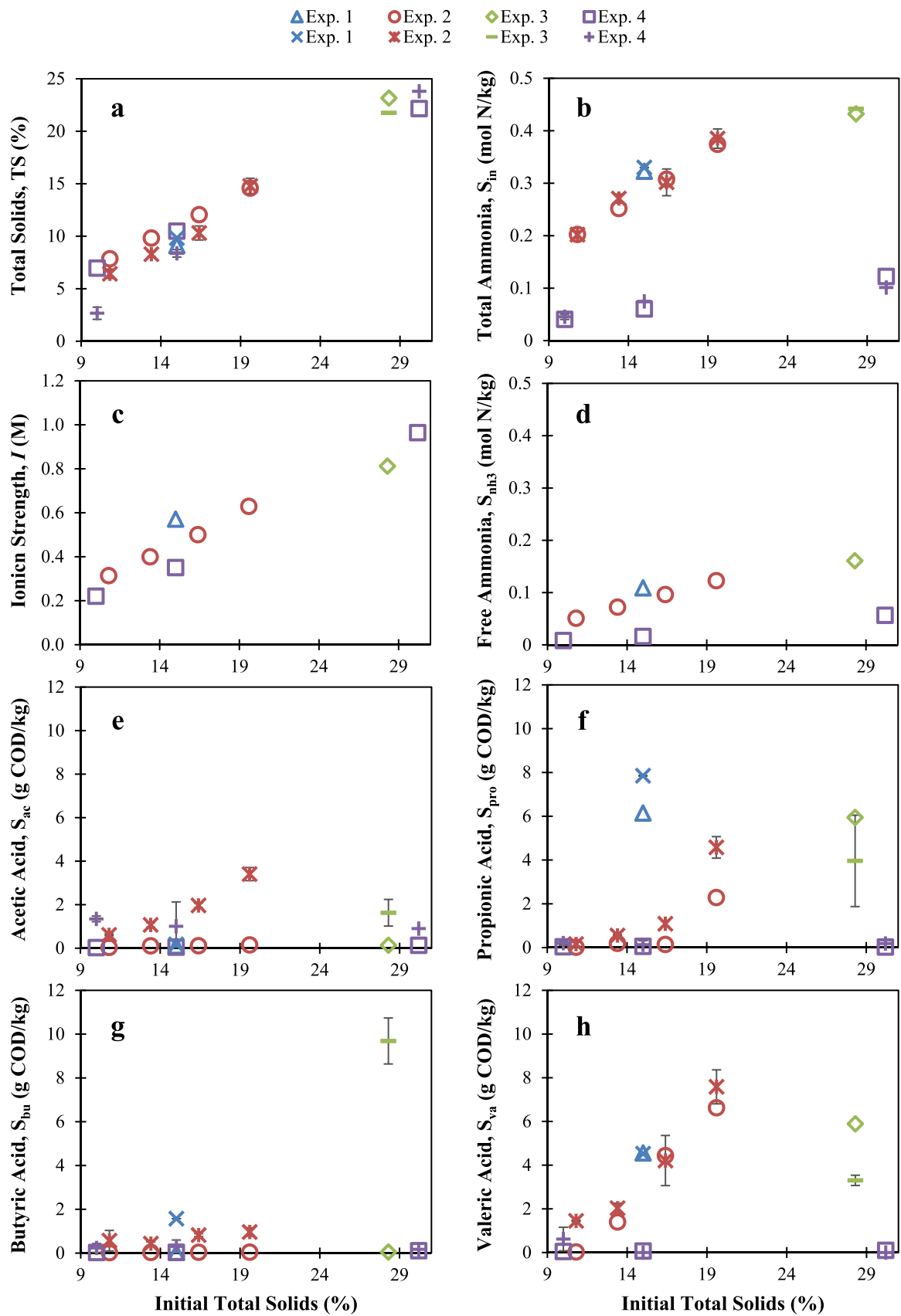


Figure 4. Main variables at the end of the four batch experiments: a) Total solids (TS); b) total ammonia nitrogen (TAN, S_{in}); c) ionic strength (I); d) free ammonia nitrogen (NH_3 , S_{nh3}); e) acetic acid (S_{ac}); f) propionic acid (S_{pro}); g) butyric acid (S_{bu}); and h) valeric acid (S_{va}). Crosses represent experimental data, while geometries represent simulated values

Table 1. Batch experiments and initial conditions used for HS-AD model calibration and cross-validation

Substrate	Exp. No.	Batch No.	ISR (g VS/g VS)	Initial TS (%)	No. Replicates	Design Volume (mL)	Exp. Time (days)	Calibration or Validation
OFMSW	1	1	1.00	15.0	15	280	92	C
		2		10.8	3			V
	2	3	1.50	13.4	3	160	100	C
		4		16.4	3			V
		5		19.6	3			C
OFMSW + Sawdust	3	6	0.50	28.3	2	280	99	C
	4	7	0.16	10.0	3	280	284	C
		8		15.0	2			V
9		30.2		1	C			

Table 2. Biochemical rates used for the HS-AD model in this study

Process	Rate (r_j , kgCOD $m^{-3} d^{-1}$)
Hydrolysis of Fast Biodegradable Carbohydrates	$k_{h,ch,fast} \cdot X_{ch,fast}$
Hydrolysis of Slow Biodegradable Carbohydrates	$k_{h,ch,slow} \cdot X_{ch,slow}$
Hydrolysis of Proteins	$k_{h,pr} \cdot X_{pr}$
Hydrolysis of Lipids	$k_{h,li} \cdot X_{li}$
Sugars Uptake	$k_{m,su} \cdot S_{su,App} / (K_{S,Xsu} + S_{su,App}) \cdot X_{su} \cdot I_{pH} \cdot I_{in}$
Aminoacids Uptake	$k_{m,aa} \cdot S_{aa,App} / (K_{S,Xaa} + S_{aa,App}) \cdot X_{aa} \cdot I_{pH} \cdot I_{in}$
LCFA Uptake	$k_{m,fa} \cdot S_{fa} / (K_{S,Xfa} + S_{fa}) \cdot X_{fa} \cdot I_{pH} \cdot I_{in} \cdot I_{h2}$
Valerate Uptake	$k_{m,c5} \cdot S_{va,App} / (K_{S,Xc5} + S_{va,App}) \cdot X_{c5} \cdot I_{pH} \cdot I_{in} \cdot I_{h2} \cdot I_{nh3}$
Butyrate Uptake	$k_{m,c4} \cdot S_{bu,App} / (K_{S,Xc4} + S_{bu,App}) \cdot X_{c4} \cdot I_{pH} \cdot I_{in} \cdot I_{h2} \cdot I_{nh3}$
Propionate Uptake	$k_{m,pro} \cdot S_{pro,App} / (K_{S,Xpro} + S_{pro,App}) \cdot X_{pro} \cdot I_{pH} \cdot I_{in} \cdot I_{h2} \cdot I_{nh3}$
Acetate Uptake	$k_{m,ac} \cdot S_{ac,App} / (K_{S,Xac} + S_{ac,App}) \cdot X_{ac} \cdot I_{pH} \cdot I_{in} \cdot I_{nh3}$
Hydrogen Uptake	$k_{m,h2} \cdot S_{h2,App} / (K_{S,Xh2} + S_{h2,App}) \cdot X_{h2} \cdot I_{pH} \cdot I_{in} \cdot I_{nh3}$
Sugar Degraders Decay	$k_d \cdot X_{su}$
Aminoacids Degraders Decay	$k_d \cdot X_{aa}$
LCFA Degraders Decay	$k_d \cdot X_{fa}$
Valerate Degraders Decay	$k_d \cdot X_{c5}$
Butyrate Degraders Decay	$k_d \cdot X_{c4}$
Propionate Degraders Decay	$k_d \cdot X_{pro}$
Acetate Degraders Decay	$k_d \cdot X_{ac}$
Hydrogen Degraders Decay	$k_d \cdot X_{h2}$

with

$$I_{in} = S_{in,App} / (K_{i,Sin} + S_{in,App})$$

$$I_{h2} = K_{i,Sh2} / (K_{i,Sh2} + S_{h2,App})$$

$$I_{pH} = K_{pH}^{N_{pH}} / (K_{pH}^{N_{pH}} + S_{h+}^{N_{pH}})$$

$$I_{nh3} = K_{i,Snh3} / (K_{i,Snh3} + S_{nh3,App})$$

Table 3. Initial conditions used for all batch simulations in this study

Name	Mono-digestion						Co-digestion			Units
	ISR = 1.00	ISR = 1.50				ISR = 0.50	ISR = 0.16			
	TS=15.0%	TS=9.5%	TS=13.5%	TS=16.5%	TS=19.4%	TS=28.3%	TS=10.0%	TS=15.0%	TS=30.2%	
S _{su}	9.761	6.920	8.776	10.861	13.245	6.201	1.800	2.779	6.496	kg COD m ⁻³
S _{aa}	3.187	5.856	7.346	9.099	11.201	7.679	0.972	1.503	3.571	kg COD m ⁻³
S _{fa}	2.610	1.656	2.186	2.702	3.184	1.467	0.377	0.579	1.300	kg COD m ⁻³
S _{va}	0.791	1.015	1.282	1.582	1.936	1.061	1.467	2.269	5.314	kg COD m ⁻³
S _{bu}	0.500	0.195	0.244	0.302	0.370	1.518	0.230	0.355	0.831	kg COD m ⁻³
S _{pro}	2.059	0.877	1.109	1.368	1.674	2.565	1.367	2.115	4.952	kg COD m ⁻³
S _{ac}	0.103	0.035	0.044	0.054	0.066	0.871	0.058	0.090	0.210	kg COD m ⁻³
S _{h2}	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	kg COD m ⁻³
S _{ch4}	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	kg COD m ⁻³
S _{ic}	0.029	0.014	0.017	0.021	0.026	0.037	0.008	0.013	0.030	kmol C m ⁻³
S _{in}	0.186	0.125	0.157	0.194	0.238	0.229	0.033	0.051	0.120	kmol N m ⁻³
S _i	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	kg COD m ⁻³
S _{i,subs}	34.706	29.233	37.076	45.883	55.954	90.351	27.565	42.559	99.486	kgCOD m ⁻³
X _{ch,fast}	29.283	13.840	17.553	21.722	26.490	31.003	5.400	8.337	19.489	kg COD m ⁻³
X _{ch,slow}	-	-	-	-	-	-	27.360	42.242	98.743	kg COD m ⁻³
X _{pr}	28.680	11.713	14.692	18.197	22.402	38.393	5.834	9.016	21.425	kg COD m ⁻³
X _g	18.271	3.312	4.373	5.405	6.367	5.866	2.637	4.053	9.102	kg COD m ⁻³
X _{su} (*)	0.050	0.050	0.063	0.078	0.095	0.150	0.005	0.008	0.018	kg COD m ⁻³
X _{aa} (*)	0.050	0.050	0.063	0.078	0.095	0.060	0.005	0.008	0.018	kg COD m ⁻³
X _{fa} (*)	0.010	0.020	0.025	0.031	0.038	0.030	0.001	0.001	0.002	kg COD m ⁻³
X _{c5} (*)	0.005	0.010	0.013	0.016	0.019	0.030	0.001	0.001	0.002	kgCOD m ⁻³
X _{c4} (*)	0.001	0.050	0.063	0.078	0.095	0.030	0.001	0.001	0.002	kg COD m ⁻³
X _{pro} (*)	0.005	0.020	0.025	0.031	0.038	0.030	0.001	0.001	0.003	kg COD m ⁻³
X _{ac} (*)	0.024	0.150	0.190	0.234	0.286	0.100	0.003**	0.005	0.011	kg COD m ⁻³
X _{h2} (*)	0.050	0.200	0.253	0.312	0.382	0.090	0.003	0.005	0.011	kg COD m ⁻³
X _i	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	1.00E-08	kg COD m ⁻³
X _{i,subs}	86.765	73.083	92.689	114.706	139.885	225.877	68.914	106.398	248.714	kg COD m ⁻³
S _{cat}	0.100	0.060	0.075	0.091	0.109	0.166	0.040	0.059	0.120	kmoleq m ⁻³
S _{an}	0.051	0.040	0.050	0.060	0.073	0.069	0.020	0.030	0.060	kmoleq m ⁻³
M _{Global}	37.12	29.92	24.02	19.80	16.47	23.45	138.23	93.13	46.13	g
ρ _{Global}	1078	1059	1075	1093	1113	1128	1088	1134	1316	kg m ⁻³
TS	15.0	10.8	13.4	16.4	19.6	28.3	10.0	15.0	30.2	%
VS	12.4	9.1	11.4	13.9	16.6	24.0	9.6	14.2	28.6	%
N _{i,subs}	0.0010	0.0012	0.0012	0.0012	0.0012	0.0010	0.0004	0.0004	0.0004	kmol N kg COD ⁻¹
V _{Reactor}	280	160	160	160	160	280	160	160	160	mL

* These values were assessed by Global Sensitivity Analysis (GSA). ** This value was also assessed by Approximate Bayesian Computation (ABC). The inoculum-to-substrate ratio (ISR) is expressed in g VS/g VS

Table 4. Main input parameters (θ) used for global sensitivity analysis (GSA), including the initial values, lower and upper thresholds

Model Parameter	Units	ADM1	Initial Value	Lower Threshold	Upper Threshold	Initial Concentration	Units	Initial Value*	Lower Threshold*	Upper Threshold*
$K_{h,ch,fast}$	d^{-1}	10	0.120	0.012	0.228	$X_{su,Batch1}$	kg COD m^{-3}	-1.301	-1.952	-0.651
$K_{h,pr}$	d^{-1}	10	0.050	0.005	0.095	$X_{su,Batch2}$	kg COD m^{-3}	-1.301	-1.952	-0.651
$K_{h,li}$	d^{-1}	10	0.080	0.008	0.152	$X_{su,Batch6}$	kg COD m^{-3}	-0.824	-1.236	-0.412
$K_{h,ch,slow}$	d^{-1}	-	0.012	0.001	0.023	$X_{su,Batch7}$	kg COD m^{-3}	-2.301	-3.452	-1.151
$k_{m,su}$	d^{-1}	70	70	7	133	$X_{aa,Batch1}$	kg COD m^{-3}	-1.301	-1.952	-0.651
$k_{m,aa}$	d^{-1}	70	70	7	133	$X_{aa,Batch2}$	kg COD m^{-3}	-1.301	-1.952	-0.651
$k_{m,fa}$	d^{-1}	10	10	1	19	$X_{aa,Batch6}$	kg COD m^{-3}	-1.222	-1.833	-0.611
$k_{m,c5}$	d^{-1}	30	8	1	15	$X_{aa,Batch7}$	kg COD m^{-3}	-2.301	-3.452	-1.151
$k_{m,c4}$	d^{-1}	30	13	1	25	$X_{fa,Batch1}$	kg COD m^{-3}	-2.000	-3.000	-1.000
$k_{m,pro}$	d^{-1}	20	10	1	19	$X_{fa,Batch2}$	kg COD m^{-3}	-1.699	-2.548	-0.849
$k_{m,ac}$	d^{-1}	16	16	2	30	$X_{fa,Batch6}$	kg COD m^{-3}	-1.523	-2.284	-0.761
$k_{m,h2}$	d^{-1}	35	20	2	38	$X_{fa,Batch7}$	kg COD m^{-3}	-3.301	-4.952	-1.651
k_d	d^{-1}	0.040	0.040	0.004	0.076	$X_{c5,Batch1}$	kg COD m^{-3}	-2.301	-3.452	-1.151
$K_{S,Xsu}$	kg COD m^{-3}	1.00	1.00	0.1	1.9	$X_{c5,Batch2}$	kg COD m^{-3}	-2.000	-3.000	-1.000
$K_{S,Xaa}$	kg COD m^{-3}	0.30	0.30	0.03	0.57	$X_{c5,Batch6}$	kg COD m^{-3}	-1.523	-2.284	-0.761
$K_{S,Xfa}$	kg COD m^{-3}	0.40	0.40	0.04	0.76	$X_{c5,Batch7}$	kg COD m^{-3}	-3.222	-4.833	-1.611
$K_{S,Xc5}$	kg COD m^{-3}	0.40	0.40	0.04	0.76	$X_{c4,Batch1}$	kg COD m^{-3}	-3.000	-4.500	-1.500
$K_{S,Xc4}$	kg COD m^{-3}	0.40	0.40	0.04	0.76	$X_{c4,Batch2}$	kg COD m^{-3}	-1.301	-1.952	-0.651
$K_{S,Xpro}$	kg COD m^{-3}	0.30	0.30	0.03	0.57	$X_{c4,Batch6}$	kg COD m^{-3}	-1.523	-2.284	-0.761
$K_{S,Xac}$	kg COD m^{-3}	0.30	0.30	0.03	0.57	$X_{c4,Batch7}$	kg COD m^{-3}	-3.222	-4.833	-1.611
$K_{S,Xh2}$	kg COD m^{-3}	5.00E-05	5.00E-05	0.000005	0.000095	$X_{pro,Batch1}$	kg COD m^{-3}	-2.301	-3.452	-1.151
$pH_{LL,ac}$	-	6.00	5.80	5.00	6.90	$X_{pro,Batch2}$	kg COD m^{-3}	-1.699	-2.548	-0.849
$K_{i,Sh3,Xc5}$	kmol N m^{-3}	-	0.0070	0.0007	0.0133	$X_{pro,Batch6}$	kg COD m^{-3}	-1.523	-2.284	-0.761
$K_{i,Sh3,Xc4}$	kmol N m^{-3}	-	0.0100	0.0010	0.0190	$X_{pro,Batch7}$	kg COD m^{-3}	-3.097	-4.645	-1.548
$K_{i,Sh3,Xpro}$	kmol N m^{-3}	-	0.0100	0.0010	0.0190	$X_{ac,Batch1}$	kg COD m^{-3}	-1.620	-2.430	-0.810
$K_{i,Sh3,Xac}$	kmol N m^{-3}	0.0110	0.0040	0.0004	0.0076	$X_{ac,Batch2}$	kg COD m^{-3}	-0.824	-1.236	-0.412
$K_{i,Sh3,Xh2}$	kmol N m^{-3}	-	0.0150	0.0015	0.0285	$X_{ac,Batch6}$	kg COD m^{-3}	-1.000	-1.500	-0.500
$f_{bu,su}$	kg COD kg COD $^{-1}$	0.130	0.500	0.050	0.950	$X_{ac,Batch7}$	kg COD m^{-3}	-2.523	-3.784	-1.261
$f_{ac,su}$	kg COD kg COD $^{-1}$	0.270	0.290	0.029	0.551	$X_{h2,Batch1}$	kg COD m^{-3}	-1.301	-1.952	-0.651
$f_{h2,su}$	kg COD kg COD $^{-1}$	0.190	0.100	0.010	0.190	$X_{h2,Batch2}$	kg COD m^{-3}	-0.699	-1.048	-0.349
$K_{i,Sh2,Xfa}$	kg COD m^{-3}	3.00E-05	1.00E-05	1.00E-06	1.90E-05	$X_{h2,Batch6}$	kg COD m^{-3}	-1.046	-1.569	-0.523
$K_{i,Sh2,Xc5}$	kg COD m^{-3}	3.00E-05	3.00E-05	3.00E-06	5.70E-05	$X_{h2,Batch7}$	kg COD m^{-3}	-2.523	-3.784	-1.261
$K_{i,Sh2,Xc4}$	kg COD m^{-3}	3.00E-05	3.00E-05	3.00E-06	5.70E-05	-	-	-	-	-
$K_{i,Sh2,Xpro}$	kg COD m^{-3}	1.00E-05	1.00E-05	1.00E-06	1.90E-05	-	-	-	-	-
k_{La}	d^{-1}	200	200	20	380	-	-	-	-	-

* *Logarithm-transformed: $\log_{10}(\theta_B)$*

Table 5. Global sensitivity analysis (GSA) of input parameters (θ): Individual (IE) and global (GE) effects upon the objective function, $J(\theta)$, variance

GSA No.1			GSA No.2			GSA No.3			GSA No.4			GSA No.5			GSA No.6			GSA No.7		
θ	IE (%)	GE	θ	IE (%)	GE	θ	IE (%)	GE	θ	IE (%)	GE	θ	IE (%)	GE	θ	IE (%)	GE	θ	IE (%)	GE
$K_{h, ch, fast}$	0.08	0.63	$K_{h, pr}$	1.75	5.23	$K_{h, pr}$	2.75	8.62	$K_{h, pr}$	2.21	7.33	$K_{h, pr}$	1.01	5.42	$K_{h, pr}$	2.00	10.44	$K_{h, pr}$	1.97	7.72
$K_{h, pr}$	2.01	7.56	$K_{h, li}$	0.25	0.69	$K_{h, li}$	0.26	2.54	$K_{h, li}$	0.11	2.85	$K_{h, ch, slow}$	4.28	14.61	$K_{h, ch, slow}$	4.07	14.47	$K_{h, ch, slow}$	4.43	17.86
$K_{h, li}$	0.29	2.07	$K_{h, ch, slow}$	5.87	16.21	$K_{h, ch, slow}$	5.83	14.32	$K_{h, ch, slow}$	5.22	16.59	$k_{m, fa}$	0.46	4.14	$k_{m, fa}$	1.15	5.66	$k_{m, fa}$	1.11	8.95
$K_{h, ch, slow}$	6.45	17.27	$k_{m, aa}$	0.07	0.43	$k_{m, fa}$	1.21	7.25	$k_{m, fa}$	0.56	4.81	$k_{m, c5}$	0.22	1.63	$k_{m, c5}$	1.59	8.34	$k_{m, c5}$	0.67	5.72
$k_{m, su}$	0.16	3.82	$k_{m, fa}$	0.81	5.97	$k_{m, c5}$	0.80	3.12	$k_{m, c5}$	0.73	3.31	$k_{m, c4}$	1.38	11.18	$k_{m, c4}$	1.50	13.79	$k_{m, c4}$	2.39	15.94
$k_{m, aa}$	0.53	2.43	$k_{m, c5}$	0.33	1.49	$k_{m, c4}$	2.23	17.17	$k_{m, c4}$	2.47	19.12	$k_{m, pro}$	1.77	8.66	$k_{m, pro}$	1.23	11.37	$k_{m, pro}$	0.72	9.24
$k_{m, fa}$	0.91	4.34	$k_{m, c4}$	2.12	12.96	$k_{m, pro}$	0.51	8.59	$k_{m, pro}$	0.75	10.21	$k_{m, ac}$	3.56	12.26	$k_{m, ac}$	5.11	17.49	$k_{m, ac}$	4.13	17.22
$k_{m, c5}$	0.22	3.64	$k_{m, pro}$	0.44	10.49	$k_{m, ac}$	7.18	17.17	$k_{m, ac}$	4.67	11.62	$k_{m, h2}$	3.72	13.46	$k_{m, h2}$	4.13	15.65	$k_{m, h2}$	1.46	12.35
$k_{m, c4}$	4.39	17.87	$k_{m, ac}$	3.16	10.71	$k_{m, h2}$	4.07	10.68	$k_{m, h2}$	3.57	11.84	k_d	5.16	15.78	k_d	3.37	9.47	k_d	4.43	13.37
$k_{m, pro}$	0.62	8.62	$k_{m, h2}$	2.60	8.32	k_d	3.67	9.00	k_d	3.22	8.23	$K_{S, Xh2}$	0.13	1.09	$pH_{LL, ac}$	10.24	24.61	$pH_{LL, ac}$	11.69	29.21
$k_{m, ac}$	4.88	16.35	k_d	2.86	7.83	$K_{S, Xh2}$	0.40	1.76	$K_{S, Xh2}$	0.22	0.91	$pH_{LL, ac}$	11.04	25.72	$f_{bu, su}$	3.34	26.74	$f_{bu, su}$	1.89	24.59
$k_{m, h2}$	1.60	11.78	$K_{S, Xh2}$	0.71	2.03	$pH_{LL, ac}$	11.09	22.10	$pH_{LL, ac}$	9.69	18.95	$K_{i, Snh3, Xh2}$	0.21	3.16	$f_{ac, su}$	0.96	13.49	$f_{ac, su}$	0.22	9.58
k_d	1.65	5.26	$pH_{LL, ac}$	11.48	25.57	$K_{i, Snh3, Xh2}$	0.32	2.61	$K_{i, Snh3, Xh2}$	0.39	1.91	$f_{bu, su}$	4.03	30.77	$f_{h2, su}$	5.00	14.87	$f_{h2, su}$	6.37	19.92
$K_{S, Xsu}$	0.05	0.34	$K_{i, Snh3, Xpro}$	0.20	1.79	$f_{bu, su}$	2.72	20.48	$f_{bu, su}$	2.93	24.05	$f_{ac, su}$	0.81	15.33	$X_{ac, Batch7}$	1.89	11.52	$X_{ac, Batch7}$	2.73	13.98
$K_{S, Xaa}$	0.15	1.12	$K_{i, Snh3, Xac}$	0.16	1.55	$f_{ac, su}$	0.41	10.59	$f_{ac, su}$	0.48	12.46	$f_{h2, su}$	7.05	17.03	$X_{h2, Batch7}$	0.00	0.00	-	-	-
$K_{S, Xfa}$	0.15	3.83	$K_{i, Snh3, Xh2}$	0.24	1.70	$f_{h2, su}$	6.02	13.39	$f_{h2, su}$	5.83	13.35	$K_{i, Sh2, Xc4}$	0.03	0.40	-	-	-	-	-	-
$K_{S, Xc5}$	0.07	0.63	$f_{bu, su}$	3.56	30.53	$K_{i, Sh2, Xc4}$	0.33	4.00	$K_{i, Sh2, Xc4}$	0.16	4.36	$X_{ac, Batch7}$	1.63	7.72	-	-	-	-	-	-
$K_{S, Xc4}$	0.00	0.00	$f_{ac, su}$	0.67	17.63	$K_{i, Sh2, Xpro}$	0.17	1.55	$K_{i, Sh2, Xpro}$	0.13	2.88	$X_{h2, Batch2}$	0.07	1.00	-	-	-	-	-	-
$K_{S, Xpro}$	0.00	0.00	$f_{h2, su}$	7.20	16.29	$X_{su, Batch6}$	0.06	0.69	$X_{c4, Batch2}$	0.06	0.28	$X_{h2, Batch6}$	0.00	0.00	-	-	-	-	-	-
$K_{S, Xac}$	0.13	1.13	$K_{i, Sh2, Xfa}$	0.00	0.00	$X_{aa, Batch6}$	0.02	0.38	$X_{c4, Batch6}$	0.06	0.33	$X_{h2, Batch7}$	0.40	3.15	-	-	-	-	-	-
$K_{S, Xh2}$	0.79	4.53	$K_{i, Sh2, Xc5}$	0.03	0.88	$X_{aa, Batch7}$	0.00	0.00	$X_{c4, Batch7}$	0.00	0.00	-	-	-	-	-	-	-	-	-
$pH_{LL, ac}$	6.73	16.94	$K_{i, Sh2, Xc4}$	0.22	2.74	$X_{fa, Batch1}$	0.03	0.10	$X_{pro, Batch1}$	0.05	0.44	-	-	-	-	-	-	-	-	-
$K_{i, Snh3, Xc5}$	0.02	0.29	$K_{i, Sh2, Xpro}$	0.25	2.26	$X_{fa, Batch2}$	0.00	0.00	$X_{pro, Batch2}$	0.00	0.00	-	-	-	-	-	-	-	-	-
$K_{i, Snh3, Xc4}$	0.16	2.64	k_{La}	0.07	0.39	$X_{fa, Batch6}$	0.00	0.00	$X_{pro, Batch6}$	0.05	0.64	-	-	-	-	-	-	-	-	-
$K_{i, Snh3, Xpro}$	0.33	4.71	$X_{su, Batch1}$	0.05	0.47	$X_{fa, Batch7}$	0.07	0.69	$X_{pro, Batch7}$	0.20	1.12	-	-	-	-	-	-	-	-	-
$K_{i, Snh3, Xac}$	0.99	6.99	$X_{su, Batch2}$	0.03	0.28	$X_{c5, Batch1}$	0.00	0.01	$X_{ac, Batch1}$	0.15	1.09	-	-	-	-	-	-	-	-	-
$K_{i, Snh3, Xh2}$	0.23	3.55	$X_{su, Batch6}$	0.22	3.38	$X_{c5, Batch2}$	0.16	0.72	$X_{ac, Batch2}$	0.14	0.38	-	-	-	-	-	-	-	-	-
$f_{bu, su}$	1.97	24.39	$X_{su, Batch7}$	0.04	0.29	$X_{c5, Batch6}$	0.03	0.14	$X_{ac, Batch6}$	0.16	1.89	-	-	-	-	-	-	-	-	-
$f_{ac, su}$	0.56	13.69	$X_{aa, Batch1}$	0.05	0.60	$X_{c5, Batch7}$	0.07	0.26	$X_{ac, Batch7}$	1.99	7.48	-	-	-	-	-	-	-	-	-
$f_{h2, su}$	4.53	12.54	$X_{aa, Batch2}$	0.00	0.00	$X_{c4, Batch1}$	0.06	0.76	$X_{h2, Batch1}$	0.18	0.74	-	-	-	-	-	-	-	-	-
Σ IE (%) =	40.67	-	-	45.46	-	-	50.46	-	-	46.38	-	-	46.94	-	-	45.57	-	-	44.21	-
Total (%) =	72.37	-	-	75.17	-	-	78.67	-	-	74.74	-	-	72.81	-	-	70.49	-	-	68.79	-

Table 6. Calibration of potentially-identifiable input parameters (θ'): Prior and posterior distributions

Parameter	Units	Initial Value	Initial Evaluation Range		Posterior Parameter Distribution						
			Minimum	Maximum	Mean	Median	Mode	5% Percentile	95% Percentile	Skewness	Kurtosis
$K_{h,pr}$	d^{-1}	0.050	0.005	0.095	0.048	0.048	0.044	0.040	0.057	0.217	2.83
$K_{h,ch,slow}$	d^{-1}	0.0120	0.0012	0.0228	0.0109	0.0108	0.0102	0.0084	0.0137	0.171	2.74
$k_{m,fa}$	d^{-1}	10.0	1.0	19.0	7.3	6.8	6.6	4.6	12.0	1.663	6.93
$k_{m,e5}$	d^{-1}	8.0	0.8	15.2	10.4	10.3	8.6	7.2	13.4	0.084	2.75
$k_{m,e4}$	d^{-1}	13.0	1.3	24.7	19.6	19.5	20.2	15.8	23.6	0.001	2.52
$k_{m,pro}$	d^{-1}	10.0	1.0	19.0	12.2	12.2	13.0	7.1	16.7	-0.148	2.66
$k_{m,ac}$	d^{-1}	16.0	1.6	30.4	24.2	24.5	23.2	18.7	29.4	-0.242	2.32
$k_{m,h2}$	d^{-1}	20.0	2.0	38.0	26.0	25.5	25.1	17.0	36.1	0.077	2.19
k_d	d^{-1}	0.040	0.004	0.076	0.040	0.040	0.045	0.020	0.059	-0.155	2.83
$pH_{L,Lac}$		5.8	5.0	6.9	5.4	5.4	5.2	5.0	5.9	0.589	2.73
$X_{ac,Batch7}$	$kg\ COD\ m^{-3}$	-2.523	-3.784	-1.261	-2.648	-2.713	-2.853	-3.574	-1.543	0.303	2.39
$f_{bu,su}$	$kg\ COD\ kg\ COD^{-1}$	0.50	0.05	0.95	0.54	0.54	0.49	0.38	0.72	0.161	2.67
$f_{ac,su}$	$kg\ COD\ kg\ COD^{-1}$	0.29	0.03	0.55	0.28	0.29	0.23	0.12	0.43	-0.180	2.73
$f_{h2,su}$	$kg\ COD\ kg\ COD^{-1}$	0.10	0.01	0.19	0.09	0.09	0.09	0.02	0.17	0.060	2.31

EXPERIMENTAL DATA

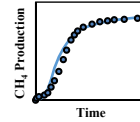


HS-AD batch experiments
fed with OFMSW
TS € [10-30] %

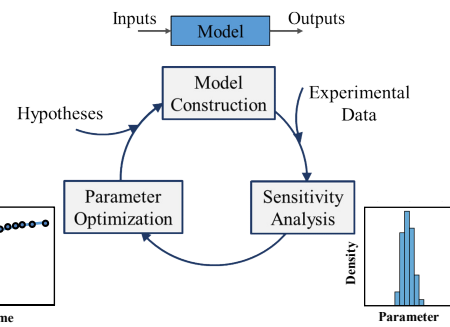
HYPOTHESIS

Reversible NH_3 inhibition in
all VFA and H_2 degraders

$$I_{nh3} = \frac{K_{L,Snh3}}{K_{L,Snh3} + S_{nh3,App}}$$



MODEL DEVELOPMENT



CONCLUSIONS

- Ionic Strength ≥ 0.9 M;
- $\text{NH}_3 \geq 2.3$ g N/L;
- Highly correlated parameters;
- Extensive experimental data needed to avoid non-identifiability;
- The reversible function poorly represents NH_3 inhibition.