

Accounting for erroneous model structures in biokinetic process models

Kris Villez^{a,b}, Dario Del Giudice^{a,c,d}, Marc B. Neumann^{e,f}, Jörg Rieckermann^a

^a*Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland*

^b*ORNL: Oak Ridge National Laboratory, Oak Ridge, TN, USA*

^c*ETHZ: Swiss Federal Institute of Technology, Stefano-Frascini-Platz 5, 8093 Zürich, Switzerland*

^d*Department of Civil, Construction & Environmental Engineering, NC State University, Mann Hall 311, 2501 Stinson Drive, Raleigh, NC, 27695, USA*

^e*Basque Centre for Climate Change (BC3), Scientific Campus of the University of the Basque Country, Sede Building 1, 1st floor, 48940 Leioa, Spain*

^f*IKERBASQUE, Basque Foundation for Science, Maria Diaz de Haro 3, 6 solairua, 48013 Bilbao, Spain*

Abstract

In engineering practice, model-based design requires not only a good process-based model, but also a good description of stochastic disturbances and measurement errors to learn credible parameter values from observations. However, typical methods use Gaussian error models, which often cannot describe the complex temporal patterns of residuals. Consequently, this results in overconfidence in the identified parameters and, in turn, optimistic reactor designs. In this work, we assess the strengths and weaknesses of a method to statistically describe these patterns with autocorrelated error models. This method produces increased widths of the credible prediction intervals following the inclusion of the bias term, in turn leading to more conservative design choices. However, we also show that the augmented error model is not a universal tool, as its application cannot guarantee the desired reliability of the resulting wastewater reactor design.

Keywords: bias description; kinetic model; process design; wastewater treatment; uncertainty

Email address: villezk@ornl.gov (Kris Villez)

Copyright notice

This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doepublic-access-plan>).

1 Introduction

2 In current environmental engineering practice, deterministic process-based
3 modeling is a common tool to better understand the functioning of complex
4 wastewater collection and treatment systems. The gold standard is to im-
5 prove prediction performance of our models by fitting them to observations.
6 Consequently, the advent of ubiquitous sensing leads to an unintended yet
7 commonly observed situation where sensors reveal more details than mech-
8 anistic models can capture. When this is the case, uncertainty estimates
9 obtained from statistical inference with mechanistic models are almost cer-
10 tainly too narrow as the applied model structure is too restrictive relative
11 to the observed reality. A long-standing question is therefore whether risk-
12 based design, based on uncertainty estimates from statistical inference with
13 mechanistic models, is actually feasible. In this work, we test one method
14 designed to address this issue to a case of WWTP design and discuss its
15 potential and limitations.

16 Accounting for model parameter uncertainty is crucial for risk-based decision-
17 making, including infrastructure design and operations (e.g., [Cagno et al., 2011](#);
18 [Scheidegger et al., 2013](#); [Kabir et al., 2015](#); [Scheidegger et al., 2015](#);
19 [Jensen and Jerez, 2018](#)). Conventional methods for uncertainty analysis are
20 based on a two-step approach, consisting of (a) quantification of input un-
21 certainty, measurement uncertainty, and subsequent uncertainty of model
22 parameters followed by (b) propagation of the quantified uncertainty to the
23 system performance measure of interest (e.g., [Van Griensven and Meixner, 2007](#);
24 [Sin et al., 2009](#); [Guo and Murphy, 2012](#); [Del Giudice et al., 2016](#)).

25 However, it has been demonstrated before how systematic deficiencies in
26 model structure, next to input and measurement uncertainty, also lead to bi-
27 ased model parameters and, consequently, incorrect design of infrastructural
28 elements, such as biological reactor systems (Neumann and Gujer, 2008).
29 Unfortunately, to our knowledge, no one has attempted to provide a method
30 to solve this particular problem, i.e. to identify systematic discrepancies in
31 process-based models so to account for them during model-based design.

32 Recently, statisticians have been suggesting a promising approach to solve
33 this dilemma. The underlying idea is to not assume identically and independ-
34 ently distributed (i.i.d.) errors for mismatches between models and obser-
35 vations (Liu and Zachara, 2001), but to explicitly account for mismatches by
36 adding a stochastic auto-correlated process to the i.i.d. measurement error
37 model (Craig et al., 2001; Kennedy and O’Hagan, 2001; Bayarri et al., 2007).
38 This is known as the bias description method and focuses on the modeling
39 of the symptoms of a mismatch between model structure and reality. While
40 this does not identify or tackle the root cause of these symptoms, it has
41 been proven to be a computationally efficient tool to increase the reliability
42 of model-based predictions compared to standard regression approaches in a
43 variety of systems from lakes to natural catchments to urban hydrology (Di-
44 etzel and Reichert, 2012; Reichert and Schuwirth, 2012; Del Giudice et al.,
45 2015). Therefore, we expect that the bias description method also improves
46 the reliability of predictions with structurally deficient wastewater treatment
47 models in view of risk-based design. Specifically, adding a stochastic mea-
48 surement error term to a model given the same amount of experimental
49 measurements is expected to reduce the relative information-richness of the
50 experimental data and lead to larger credibility intervals of model param-
51 eters, wider prediction intervals and, by avoiding overconfident predictions, a
52 more trustworthy design.

53 Note that the bias description method can be regarded as a grey box or
54 hybrid modelling strategy. Indeed, the resulting model consists of a mech-
55 anistic model for the studied process (white box) and a stochastic model
56 for auto-correlated measurement errors (black box). Other grey box ap-
57 proaches may be based on the inclusion of time-variant parameters (Reichert
58 and Mieleitner, 2009; Lin and Beck, 2012) or integration of non-parametric
59 elements into a model structure that is mechanistic otherwise (Mašić et al.,
60 2017).

61 In this contribution, we apply the bias description method to investigate
62 the impact of model structure deficits for process design. We use Neumann

63 and Gujer (2008) as a benchmark to evaluate the benefits and limitations
64 of the bias description method for model-based design and refer to it as *the*
65 *reference study*. While this reference study concerns a conceptually simple
66 case, using it in this study highlights (a) that the apparent simplicity of this
67 case is rather deceptive and (b) that challenges associated with model-reality
68 mismatch are to be expected for both simple and complex systems.

69 2. Material and methods

70 2.1. Applied error models

71 In a vast majority of environmental modeling studies, the measurement
72 error is assumed to be i.i.d. For example, Hauduc et al. (2015) compares an
73 extensive list of model performance criteria for wastewater treatment mod-
74 elling yet does not list any criterion which accounts for autocorrelated model
75 prediction errors. One approach considered in Cierkens et al. (2012), consists
76 of downsampling time series to avoid the appearance of autocorrelation. As
77 explained in the same study, this leads to an inefficient use of the available
78 data and, more importantly, cannot account at all for model structure deficits
79 as a potential root cause of autocorrelated residuals. Ignoring the presence
80 of autocorrelated residuals was shown to lead to overconfidence in the pro-
81 duced model and, subsequently, poor decision-making, as was shown also in
82 the reference study. Most often, a Gaussian distribution is assumed for the
83 measurement errors. Such a model of measurement error cannot account for
84 systematic deviations between the assumed model and the observed mea-
85 surements, i.e. bias. One way of accounting for bias is by adding terms, such
86 as a stochastic autocorrelated error bias term, to the measurement equation
87 (Craig et al., 2001; Kennedy and O’Hagan, 2001; Bayarri et al., 2007). In
88 this work, we describe the observable output time-series (i.e., measured con-
89 centration, y_o) as a sum of a deterministic dynamic model output (y , the
90 modeled concentration), a classical Gaussian measurement error ($e(\psi)$), and
91 an auto-correlated error term ($b(\psi)$):

$$y_o(\theta, \gamma, \psi) = y(\theta) + \gamma + b(\psi) + e(\psi) \quad (1)$$

92 where θ and γ are parameters of the deterministic parts of the model and
93 ψ are those of the stochastic parts (errors). The bias term $b(\psi)$ describes an
94 autocorrelated error ($b(\psi) \sim \mathcal{N}(0, \Sigma_b(\sigma_b, \tau))$) and can be included to account

95 for time-dependent deviations between model and observations (Reichert and
 96 Schuwirth, 2012). Note that this bias term represents a stochastic process,
 97 thus describing aleatory uncertainty, although the deviations between model
 98 and observations may actually be systematic, possibly even deterministic.
 99 These deviations are expected to be systematic when they are caused by a
 100 lack of knowledge about the true data-generating process. This lack of knowl-
 101 edge is typically characterized as a source of epistemic uncertainty rather than
 102 aleatory uncertainty.

103 The bias term has two parameters, the standard deviation σ_b and the
 104 correlation length τ :

$$\Sigma_b(i, j) := \sigma_b^2 \cdot e^{-|t_i - t_j|/\tau}. \quad (2)$$

105 The random measurement error is temporally independent ($e \sim \mathcal{N}(0, \Sigma_e(\sigma_e))$)
 106 and is characterized by the parameter σ_e :

$$\Sigma_e(i, j) := \begin{cases} \sigma_e^2, & i = j \\ 0, & i \neq j \end{cases}. \quad (3)$$

107 Together, these error terms with parameters $\psi = \{\tau, \sigma_b, \sigma_e\}$ account for
 108 the fact that the deterministic model may not reproduce the modeled data
 109 set exactly. Note that the symbols σ_b and σ_e are chosen to convey the idea
 110 that they both describe the magnitude of variation of a stochastic term in the
 111 measurement equation. The symbol for the correlation length, τ , is chosen
 112 to highlight the fact that it describes a time-scale.

113 The statistical formulation in Eq. 1 naturally leads to the likelihood
 114 function $\mathcal{L}(y_o|\theta, \psi)$ which describes how likely the considered model with
 115 parameters (θ, ψ) generated the recorded data, y_o . The likelihood of the
 116 measurements conditional to the model parameters is:

$$\mathcal{L}(y_o|\theta, \psi) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2} [y_o - y]^T (\Sigma)^{-1} [y_o - y]\right) \quad (4)$$

117 where Σ is the variance-covariance matrix for the stochastic deviations
 118 between model and observations:

$$\Sigma := \Sigma_e + \Sigma_b, \tag{5}$$

119 with Σ_b and Σ_e defined in 2 and 3, and with n equal to the number of
 120 measurements. In order to interpret the results of parameter estimation, we
 121 define the parameters α and σ such that $\sigma_e^2 := (1 - \alpha) \sigma^2$ and $\sigma_b^2 := \alpha \sigma^2$.
 122 This means we can express the variance-covariance matrix above equivalently
 123 as:

$$\Sigma := \sigma^2 \cdot \left[(1 - \alpha) I_{n \times n} + \alpha K \right] \tag{6}$$

$$K(i, j) := e^{-|t_i - t_j|^2 / \tau} \tag{7}$$

124 In this form, σ is a measure for the overall spread of the deviations be-
 125 tween the model and the measurements and α is a parameter that defines
 126 the relative importance of the bias in the overall variance-covariance matrix.
 127 Meaningful values for α are between 0 and 1, with $\alpha = 1$ leading to the omis-
 128 sion of the independent measurement noise ($\sigma_e = 0$) and $\alpha = 0$ expressing
 129 that there is no bias ($\sigma_b = 0$). Note that setting $\alpha = 0$ reproduces the model
 130 without a bias term. Put otherwise, the model with bias term includes the
 131 model without bias term as special case.

132 The addition of a bias term accounts for underestimation of parameter
 133 uncertainty when a conventional yet unrealistic distribution for the model
 134 error is assumed (e.g., uncorrelated). This is expected to produce a wider
 135 predictive distribution, possibly leading to a better quantification of and a
 136 reduction of the risk of under-design or over-design. However, special atten-
 137 tion must be given to the parameter estimation method as increased model
 138 flexibility can lead to unidentifiability (see e.g., [Renard et al., 2010](#)).

139 2.1.1. Model parameter estimation

140 We apply a Bayesian approach for two reasons. First, we favour a Bayesian
 141 framework as a way to make prior beliefs explicit. Second, without any form
 142 of prior, some of the parameters of the variance-covariance matrix Σ can be
 143 structurally unidentifiable (for definitions, see [Dochain et al., 1995](#); [Dochain
 144 and Vanrolleghem, 2001](#); [Petersen et al., 2003](#)). More specifically, when $\tau = 0$
 145 the matrix K equals the identity matrix and likelihood $\mathcal{L}(y_o | \theta, \psi)$ becomes
 146 insensitive to the value of α . As a result, no unique value for α can be

147 identified under any circumstances as long as $\tau = 0$, i.e. α is structurally
148 unidentifiable. In the formulation with σ_e and σ_b , any increase of σ_e can
149 be compensated exactly by an equivalent decrease of σ_b when $\tau = 0$. For
150 small values of τ , e.g. close to the measurement interval or smaller, this is
151 expected to lead to a lack of practical identifiability, even if structural identi-
152 fiability could be guaranteed in principle. In early experiments with uniform
153 priors for τ , we observed that this can induce a lack of convergence and poor
154 mixing conditions for the applied sampling methods, similar to observations
155 described in [Renard et al. \(2010\)](#). Applying an informative prior solves this
156 identifiability problem and can therefore also be interpreted as a form of
157 regularization (e.g., [Scales and Tenorio, 2001](#); [Murphy, 2012](#); [Hastie et al.,](#)
158 [2015](#)).

159 Bayesian calibration aims at characterizing the distribution described by
160 the posterior likelihood $\mathcal{L}(\theta, \psi | y_o) \propto \mathcal{L}(y_o | \theta, \psi) \cdot \mathcal{L}(\theta, \psi)$, where the prior
161 likelihood $\mathcal{L}(\theta, \psi)$ expresses the prior beliefs about the parameters. In this
162 work, the posterior distribution is approximated with a sample of $\mathcal{L}(\theta, \psi | y_o)$
163 drawn with a Markov Chain Monte Carlo (MCMC) sampler (see “Numerical
164 implementation”).

165 *2.2. Biokinetic model parameter identification with batch experiments*

166 To study the effects of model structure error and the utility of the bias de-
167 scription method, we execute simulations with the dynamic biokinetic model
168 used in the reference study. Concretely, a series of batch experiments is sim-
169 ulated in which a substrate, with concentration $s(t)$, is consumed by a cell
170 culture with a fixed concentration. The conversion rate $r(t)$ depends on the
171 substrate by means of time-invariant Tessier kinetics so that one can write:

$$\frac{ds(t)}{dt} = -r(t) \tag{8}$$

$$s(t = 0) = s_0 \tag{9}$$

$$r(t) = r_{max}^{Tessier} \cdot \left(1 - \exp \left(- \frac{s(t)}{K_S^{Tessier}} \right) \right) \cdot x(t) \tag{10}$$

172 During the experiment, noisy measurements of the true substrate con-
173 centration, $y_o(t)$, are simulated by the following measurement error model,
174 which is a zero-mean Gaussian noise term:

$$y_o(t) = s(t) + e(t) \quad (11)$$

$$e(t) \sim N(0, \sigma_e) \quad (12)$$

175 Fixed parameters for each simulation are the same as in the reference
 176 study: s_0 (initial substrate concentration, 5 g/m^3), r_{max} (maximum con-
 177 version rate, $1 \text{ g/m}^3.h$). The simulated time is $T = 8$ hours. The affinity
 178 constant ($K_S^{Tessier}$) and measurement error standard deviation (σ_e) are varied
 179 yet constant in every simulated experiment. $K_S^{Tessier}$ is varied from 0.1 g/m^3
 180 to 1.5 g/m^3 in steps of 0.2 g/m^3 . This allows simulating a wide range of
 181 process conditions, including both low and high values for $K_S^{Tessier}$ relative
 182 to the initial substrate concentration. Two values for the simulated σ_e are
 183 considered, as in the reference study. In the low-noise case, σ_e takes the value
 184 0.01 g/m^3 . In the high-noise case, it takes the value 0.1 g/m^3 . The vector
 185 θ equals $[s_0, r_{max}, K_S^{Tessier}]^T$. The two simulated noisy time series obtained
 186 with $K_S^{Tessier} = 0.7 \text{ g/m}^3$ are shown in the supplementary information (Fig.
 187 S.1 and Fig. S.2).

188 For each simulation experiment, parameter identification is executed with
 189 four distinct model structures. The first model matches the above model
 190 structure (Eq. 8-Eq. 12) exactly. This represents an idealized situation where
 191 the structure of the calibrated model matches reality (ground truth) exactly.
 192 The identified parameters are S_0 , $\mu_{max}^{Tessier}$, $K_S^{Tessier}$, and σ_e . A second model
 193 is obtained by replacing the Tessier kinetics with the alternative and more
 194 commonly used Monod kinetics. Practically, Eq. 10, is replaced with the
 195 following equation:

$$r(t) = r_{max}^{Monod} \cdot \frac{s(t)}{K_s^{Monod} + s(t)} \quad (13)$$

196 The estimated parameters are now s_0 , μ_{max}^{Monod} , K_s^{Monod} , and σ_e with
 197 $\theta = [s_0, r_{max}, K_S^{Monod}]^T$. This case represents the likely situation that a
 198 modeling practitioner uses the common-place Monod model structure and
 199 does not observe the bias that results. This is very likely in the high-noise
 200 case (see reference study). Given this difficulty, the stochastic bias term de-
 201 scribed above is included to capture the systematic deviations between the
 202 model predictions and measurements. To achieve this, the previously applied
 203 measurement equation (Eq. 11) is replaced with the following equations:

$$y_o(t) = s(t) + b(t) + e(t) \quad (14)$$

$$e(t) \sim \mathcal{N}(0, \sigma_e) \quad (15)$$

$$b \sim \mathcal{N}(0, \Sigma_b(\sigma_b, \tau)) \quad (16)$$

204 with the parameters τ defined as above and σ_b and σ_e reparametrized with
 205 α and σ . This results in a third model, where a Tessier model is combined
 206 with the statistical bias description and which requires specification of the
 207 parameters s_0 , $\mu_{max}^{Tessier}$, $K_S^{Tessier}$, σ , α , and τ . As the Tessier model has
 208 the same structure as the data-generating model, one can expect a good
 209 model fit with α close to 0 and estimates of s_0 , $\mu_{max}^{Tessier}$, $K_S^{Tessier}$, and σ
 210 that are close to ground truth values. Finally, the fourth model combines
 211 the presumed Monod kinetics with the statistical bias description and the
 212 identified parameters are s_0 , μ_{max}^{Monod} , K_S^{Monod} , σ , α , and τ . In this case we
 213 can expect that the present model structure bias is accommodated by means
 214 of the statistical bias description. If so, this should increase the width of the
 215 prediction intervals and thereby improve the reliability of the model ([Reichert
 216 and Schuwirth, 2012](#)).

217 *2.3. Numerical implementation*

218 The biokinetic model, parameter estimation, and uncertainty propaga-
 219 tion were implemented in Matlab (R2019a). The prior probabilities for the
 220 parameters were set based on the authors' experience. They are all independ-
 221 ent of each other. All priors are uniform, except for σ and τ . The prior
 222 likelihood for σ is proportional to its inverse and is equivalent to the Jeffreys
 223 prior conditional to fixed values for all other parameters (see [Box and Tiao,
 224 1973](#)). The prior likelihood for τ is the sine function supported between 0 and
 225 $2 T$. This prior equals zero at $\tau = 0$ and $\tau = 2 T$ and one at $\tau = T$. This ex-
 226 presses the subjective belief that the autocorrelation length of the deviations
 227 due to model structure error is expected to be similar to the duration of the
 228 experiment. The priors are specified completely in [Table 1](#). We first run an
 229 adaptive MCMC algorithm ([Vihola, 2012](#)) to find a good guess for the maxi-
 230 mum a posteriori estimates and a good proposal variance-covariance matrix.
 231 With these results, we execute a (non-adaptive) MCMC algorithm to obtain
 232 20,000 samples from $\mathcal{L}(\theta, \psi | y_o)$. The first 10,000 samples are considered to
 233 correspond to the burn-in phase of the sampler, during which effects of the
 234 initial sample may still be apparent. These samples are therefore discarded,
 235 as is common in practice ([Gilks et al., 1996](#)).

Table 1: Prior probability distributions for the parameters.

Parameter	Distribution	Lower bound	Upper bound	Unit
s_0	Uniform	0	$+\infty$	g/m^3
r_{max}	Uniform	0	$+\infty$	$g/m^3 \cdot h$
K_S	Uniform	0	$+\infty$	g/m^3
σ	Inverse ($\propto \frac{1}{\sigma}$)	e^{-12}	e^{+12}	g/m^3
α	Uniform	0	1	—
τ	Sine ($\propto \sin(\frac{\pi}{2} \frac{\tau}{T})$)	0	$2 T$	h

236 *2.4. Using information gained in parameter estimation for design*

237 The obtained parameter estimates for each model are inspected by means
 238 of visual inspection. In addition, we test the reliability of the calibrated
 239 models for subsequent process design. To this end, the steady state substrate
 240 concentration is computed for a CSTR with cells growing according to the
 241 aforementioned ground truth Tessier kinetics. The dilution rate D is set at
 242 $0.5 h^{-1}$, corresponding to a hydraulic residence time of 2 h. The ground
 243 truth steady state substrate concentration can be computed as follows:

$$s(t \rightarrow \infty)_{CSTR} = -K_s^{Tessier} \cdot \log \left(1 - \frac{D}{r_{max}} \right) \quad (17)$$

244 The above equation is also used to simulate the model-based concentra-
 245 tion estimate by replacing the true value of $K_s^{Tessier}$ with its estimates. To
 246 do the same with the models exhibiting Monod kinetics, the steady state
 247 concentration estimate is computed as follows:

$$s(t \rightarrow \infty)_{CSTR} = \frac{D \cdot K_s^{Monod}}{r_{max} - D} \quad (18)$$

248 The steady state concentrations are computed assuming that a perfectly
 249 accurate value for r_{max} is available from a separate washout experiment (as in
 250 the reference study). The affinity constant, K_s^{Monod} or $K_s^{Tessier}$, is therefore
 251 the only parameter whose values are based on the batch experiment. In all
 252 cases, the reliability of the predicted steady-state concentration in the CSTR
 253 is assessed by visual inspection. Considering that the perfect information
 254 about r_{max} is not accounted for during estimation of the affinity constant, we

255 also provide results in the supplementary information for a modified version
256 of our method. In this case, we assume no washout experiment has been
257 executed and use the sampled values for r_{max} in Eq. 17 and Eq. 18 instead
258 of the assumed ground truth value.

259 2.5. Software availability

260 All data and numerical methods to reproduce our results are available as a
261 stand-alone package for the Matlab (R2019b) platform. The version used for
262 this article is added to the supplementary materials. The most recent version
263 can be found on Gitlab (<https://gitlab.com/krisvillez/biasdescription>).

264 3. Results

265 3.1. Parameter estimation

266 The effect of bias and the use of the statistical bias description are demon-
267 strated first by means of the marginal posterior parameter distributions ob-
268 tained with data simulated with the Tessier model with $K_s^{Tessier} = 0.7 \text{ g/m}^3$
269 and $\sigma_e = 0.01 \text{ g/m}^3$. In Fig. 1 the empirical cumulative density functions
270 are shown for the idealized case where the model structure is correct. The
271 results without bias description are shown with a dashed red line. One can
272 observe easily that the estimates of the deterministic part of the model are
273 fairly accurate. Indeed, the distributions are both narrow (all relative stan-
274 dard deviations are below 2% in magnitude) and close to the true values
275 (all median estimates are within $\pm 1.5\%$ of the ground truth). This is not
276 surprising. In contrast, the median estimate of σ is about 35% smaller than
277 the simulated value while its precision remains small also (relative standard
278 deviation: 1.1%).

279 The red dashed lines in Fig. 2 show the empirical cumulative density
280 functions produced by using the Monod model structure while using the
281 conventional error model without bias description. This results in biased
282 estimates of the conversion rate (r_{max}^{Monod}). These estimates do not reflect
283 the values used in the ground truth simulation, as discussed in the reference
284 study. For instance, the relative deviation median estimates are +19% for
285 r_{max}^{Monod} . This is explained as a consequence of using a different rate function
286 during estimation. Indeed, the parameter values for r_{max} (and K_S) are used
287 to compensate for the model structure error. The relative standard devia-
288 tions for all parameters remain below 1.1% in magnitude, except for K_s^{Monod}
289 (2.7%).

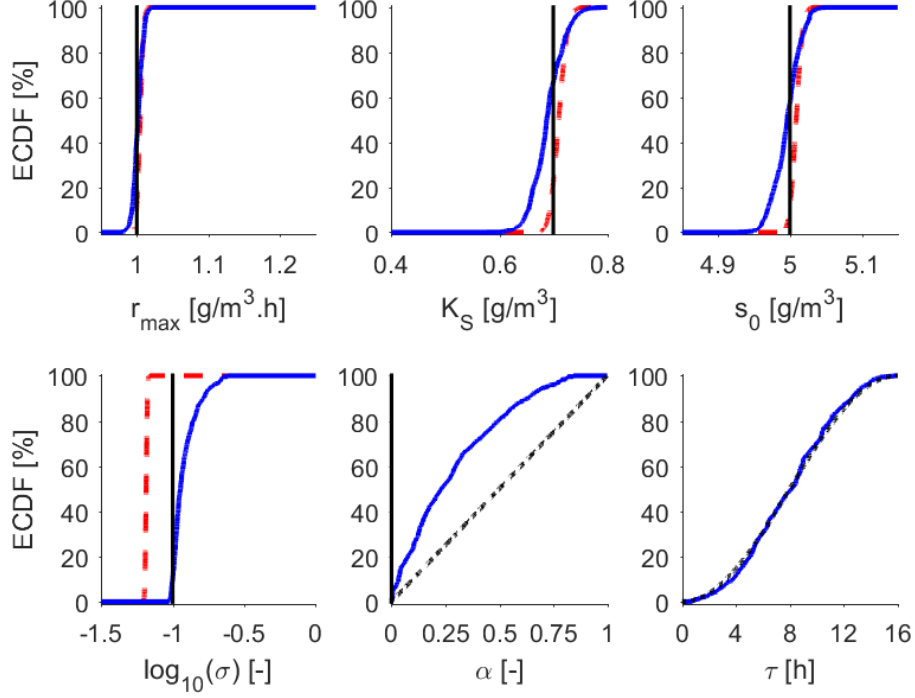


Figure 1: Batch experiment simulated with the Tessier model and $K_s = 0.7 \text{ g/m}^3$ and $\sigma_e = 0.01 \text{ g/m}^3$. Marginal posterior distributions for the calibrated Tessier model without bias description (dashed red line) and with bias description (full blue line). Vertical full lines indicate the true data-generating parameter values (r_{max} , K_S , s_0 , σ , α). Black dotted lines indicate the informative prior (α , τ).

290 This is different in the case where the Monod model structure is used
 291 for estimation (Fig. 2). As expected, the posterior distribution for α is now
 292 located to the right of its prior and indicates a large contribution of bias
 293 to the prediction error. At the same time, the posterior for τ is to the
 294 left of the prior, which means the autocorrelation length is shorter than
 295 the experimental time length (8h). The residuals obtained with maximum
 296 likelihood estimation of the Monod model now appear correlated as well (see
 297 Fig. S.2 in the supplementary information).

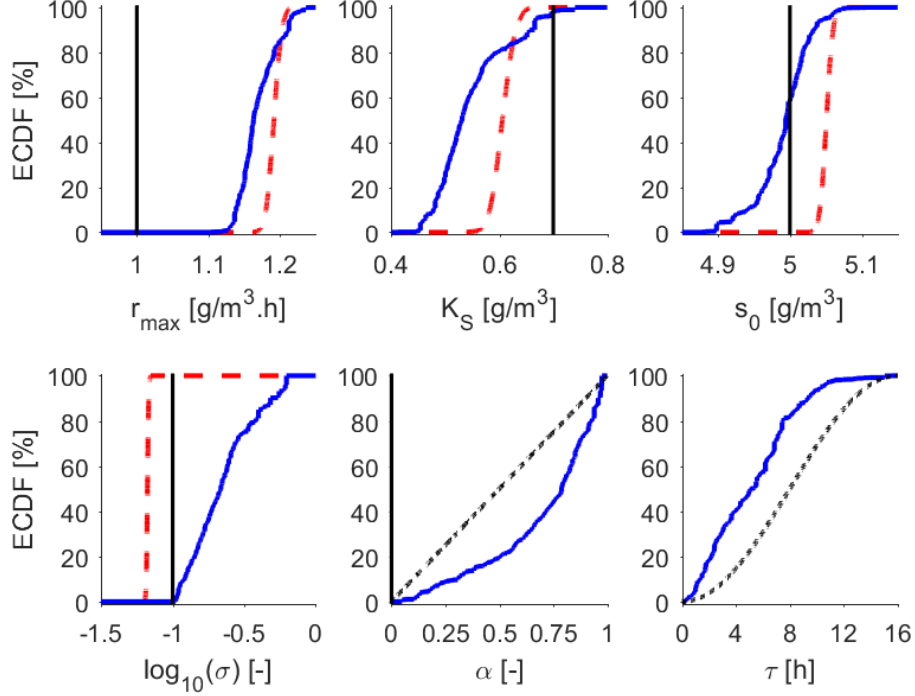


Figure 2: Batch experiment simulated with the Monod model and $K_s = 0.7 \text{ g/m}^3$ and $\sigma_e = 0.01 \text{ g/m}^3$. Marginal posterior distributions for the calibrated Monod model without bias description (dashed red line) and with bias description (full blue line). Vertical full lines indicate the data-generating parameter values in the Tessier model used for simulation (r_{\max} , K_S , s_0 , σ , α). Black dotted lines indicate the informative prior (α , τ).

298 3.2. Model-based prediction: use of parameter estimates for CSTR design

299 For the purpose of model-based design it is important to predict the
 300 steady state concentration properly for a given desired dilution rate. As
 301 indicated above, the steady state concentration is predicted at a dilution
 302 rate of 0.5 h^{-1} for both the ground truth as well as with the four constructed
 303 models.

304 The low-noise case is discussed first. The identified Monod model pa-
 305 rameter sets are used to predict the steady-state concentration by means of
 306 Eq. 18. This is repeated for every simulated value of K_S^{Tessier} . The ratios of
 307 the predicted steady-state concentrations to the true steady-state concentra-

308 tion are visualized in Fig. 3. It is clear that using the Monod model leads
 309 to a significant bias in these predictions. This situation is however easy to
 310 identify by inspection of the posterior of α (as explained above). In addi-
 311 tion, model structure error in the low-noise case is also identified easily by
 312 means of frequentist methods (cfr. reference study). We therefore assume
 313 that the modeler applies one of these tools and thereby successfully identifies
 314 the presence of bias.

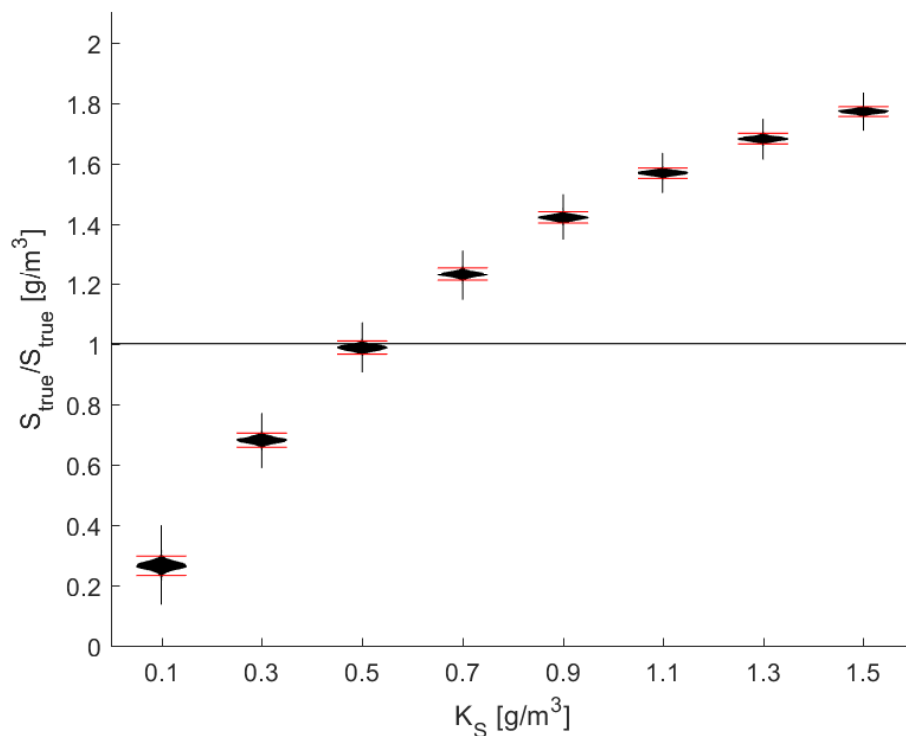


Figure 3: Distributions of the ratio of predicted steady state concentrations to the ground truth concentration as a function of the affinity constant (K_S) – Low noise case ($\sigma_e = 0.01 \text{ g/m}^3$). Red horizontal whiskers indicate the two-sided 99% credible intervals. Results are shown for the Monod model without bias description. As the ideal ratio of 1 falls outside the credibility intervals, we see that using a wrong model for reactor design will lead to a failure to achieve the desired pollutant removal capacity. This situation is easily detected in this low-noise case, also when using traditional methods.

315 As the high-noise case is more challenging for frequentist methods, we

316 discuss it in greater detail. Fig. 4 shows all results obtained for every assumed
 317 structure for the deterministic model (Monod/Tessier) and for the stochastic
 318 model parts (with and without bias). The top panel displays the results
 319 obtained with the Tessier model. It is easy to see that the predictions are
 320 unbiased as well as precise when no bias description is incorporated in the
 321 error model (left-side bean plots), except when $K_S = 0.1 \text{ g/m}^3$. We return
 322 to the latter case below. If the stochastic bias term is added (right-side
 323 bean plots), the distribution of the predictions becomes wider without any
 324 meaningful shift of the average predictions ($K_S = 0.3 \text{ g/m}^3$ and higher).
 325 Consider that the process design is based on the shown 99% credible limits.
 326 As adding bias increases the predicted upper limit, this will result in a larger
 327 design volume to account for the perceived increase in uncertainty. As such,
 328 accounting for uncertainty in the design would lead to a more conservative
 329 design compared to the case without bias description.

330 The predictions obtained with a Monod model are shown in the bot-
 331 tom panel of Fig. 4. Depending on the ground truth value for $K_s^{Tessier}$,
 332 rather severe under-prediction or over-prediction results when no bias term
 333 is added (left-side bean plots). The predictive uncertainty (spread) is similar
 334 to the case with the Tessier models without bias description. As a result,
 335 the two-sided 99% credible intervals include the ground truth in only one
 336 of the simulated cases ($K_s = 0.5 \text{ g/m}^3$), indicating the lack of reliability
 337 of the model-based prediction intervals. The addition of a bias description
 338 term abates this issue to some extent. Even though bias is still present, the
 339 addition of a phenomenological bias term leads to increased widths of the
 340 credible intervals and a more conservative design. However, the ground truth
 341 is included in the 99% credible intervals in two cases only ($K_s = 0.5 \text{ g/m}^3$,
 342 $K_s = 0.7 \text{ g/m}^3$). Thus, the bias description cannot ensure a reliable design
 343 without further modification. Inspecting the posterior of α is more useful.
 344 As one can see in Fig. 5, in all cases without model structure error (Tessier
 345 model), the posterior of α is shifted towards the left of its prior, thus sug-
 346 gesting that the model structure is defined well. For the case with model
 347 structure error (Monod model), the posterior is shifted to the right or re-
 348 mains relatively close to the prior, except for $K_S = 0.1 \text{ g/m}^3$. One reason
 349 is that this value is equal to the measurement error standard deviation. A
 350 second likely factor is that the length of time during which the produced
 351 experimental data are sensitive to the value of K_S is very short. Indeed,
 352 the substrate concentration in the simulated experiment is between half and
 353 twice the value for K_S for less than 3% of the duration of the experiment.

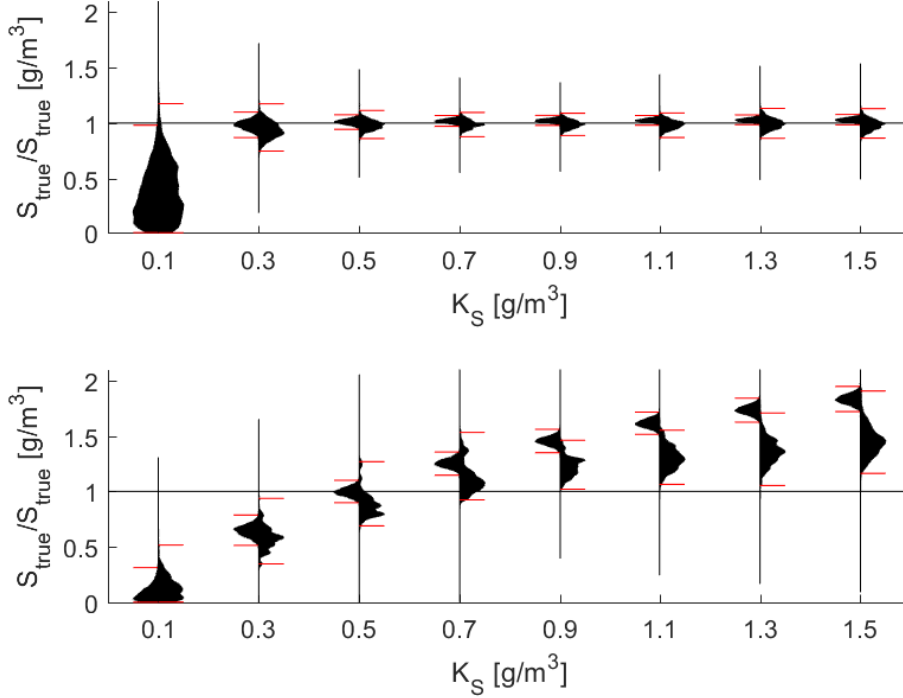


Figure 4: Distributions of the ratio of predicted steady state concentrations to the ground truth concentration as a function of the affinity constant (K_S) – High noise case ($\sigma_e = 0.1 \text{ g/m}^3$). Red horizontal whiskers indicate the two-sided 99% credible intervals. **Left side beans:** without bias description; **Right side beans:** with bias description. **Top:** Tessier model - All credible intervals include the ideal ratio (equal to 1), except for the simulation with $K_S = 0.1 \text{ g/m}^3$ without bias term. The uncertainty increases when a bias term is added to the model. **Bottom:** Monod model – Including the bias term in the model increases the reliability of the credible intervals. These intervals include the ideal ratio for two cases ($K_S = 0.5$ and 0.7 g/m^3).

354 For $K_S = 0.3 \text{ g/m}^3$ this already amounts to 8.6%. Since the model structure
 355 error primarily relates to the curvature of the conversion rate in this region,
 356 it follows that model structure error will always be difficult to detect when
 357 this time fraction is low.

358 In the supplementary information, we provide results obtained with the
 359 modified method. We omit the information obtained during the washout ex-
 360 periment during prediction. In this case, the uncertainty in the predictions

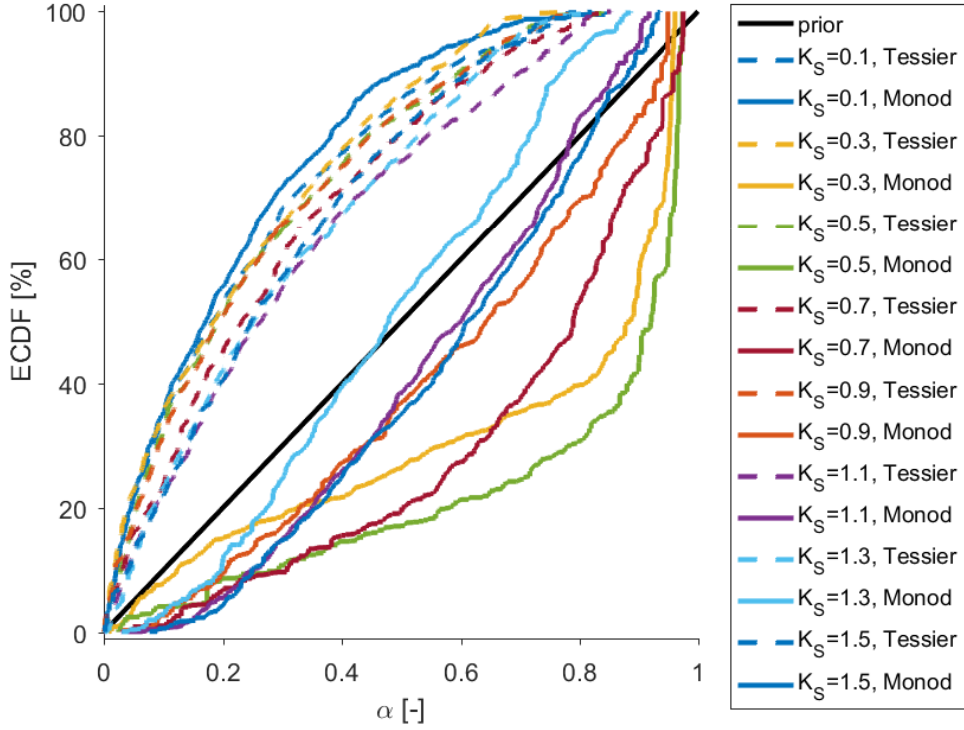


Figure 5: Distributions of the parameter α for both models with a bias term in all high-noise cases ($\sigma_e = 0.1 \text{ g/m}^3$). When the Tessier model is selected (no model structure error), the posterior distribution of α is shifted to the left of the prior, thus suggesting the kinetic model structure is adequate. In contrast, the posterior of α is similar to or located at the right of the prior when the Monod model is used in all but one case ($K_S = 0.1 \text{ g/m}^3$), thus providing a useful indication of model structure error.

361 is reduced significantly to the point that none of the 99% credible intervals
 362 include the ground truth (see Fig.S.3). This is explained by the fact that
 363 the estimates for r_{max} and K_S exhibit strong correlation (see supplement-
 364 ary information for details). However, since the modification relates to the
 365 prediction step only, one can still use the posteriors for α as a detection
 366 mechanism for bias.

367 4. Discussion

368 4.1. Summary and limitations of the experimental simulation study

369 *Summary.* The numerical results described above suggest that the inclusion
370 of an additive auto-correlated error process into a measurement error model
371 can improve the reliability of model-based designs. This is true even when
372 only a subset of the identified parameters are used during prediction (here we
373 only used the estimates for K_S) and even when the experimental setting for
374 prediction (steady state) is different from the experimental conditions used
375 for model identification (batch experiment). In our case, the bias description
376 method improves the reliability in all cases. Despite this improvement, the
377 computed credible intervals include the ground truth value only in a lim-
378 ited number of cases with model structure error, meaning that guaranteed
379 reliability cannot be obtained with the studied method. Thus, the inclusion
380 of a bias description term for the purpose of prediction can be advised as
381 a relatively fast and easy way to account for errors in the proposed model
382 structure, however only when one is unable to modify the model structure
383 itself. This is especially relevant in engineering applications where one is re-
384 stricted to specific process representations (e.g., Monod kinetics) or software
385 with limited flexibility. While the bias description method improves the re-
386 liability of the model predictions only in a limited way, it is very useful as
387 a tool to detect the presence of bias during model identification, especially
388 when reformulated with the α parameter.

389 *Limitations.* In this study, a simple case was chosen deliberately for two
390 reasons. First, this enabled an objective comparison of the bias descrip-
391 tive method with the historical results in the reference study ([Neumann and](#)
392 [Gujer, 2008](#)). Second, the apparent simplicity of the case also highlights the
393 challenge of generating reliable predictions with mechanistic models, induced
394 by the typical lack of flexibility of such models. The chosen scope also means
395 that our study comes with some limitations, which are:

- 396 • The general applicability of the bias description method is not demon-
397 strated. However, the bias description method could easily be adapted
398 to more complex systems. One could incorporate a bias term to ex-
399 press correlation between multiple measurements, of the same or dis-
400 tinct variables measured in the same location or different locations. In
401 this case, the covariance between two measurements, as expressed by Σ ,
402 would not only be a function of (*a*) the time difference ($t_i - t_j$, see (6)),

403 as in our study, but also of *(b)* spatial distance in one or more dimen-
404 sions and *(c)* effects of measurement error correlation between distinct
405 sensors measuring the same or distinct variables. This generalization
406 of the present model is likely most convenient when the bias error term
407 is modelled as spatio-temporal Gaussian process (e.g., [De Cesare et al.,](#)
408 [2001](#); [Gneiting, 2002](#); [Stein, 2005](#)).

- 409 • The methods applied in both the reference study and ours are based on
410 methods that account for aleatory uncertainty only. However, the lack
411 of knowledge about the model structure is typically epistemic in nature
412 and may therefore be difficult to account for in this way. Epistemic
413 uncertainty may however be reduced by using more flexible models
414 ([Mašić et al., 2017](#)) while increasing parametric uncertainty, which can
415 be handled as an aleatory source of uncertainty with currently available
416 methods. Still, the adoption of alternative frameworks for uncertainty
417 analysis ([Parsons, 2001](#); [Rao et al., 2008](#)) may be suited to handle
418 epistemic uncertainty directly. In summary, the handling of epistemic
419 uncertainty deserves more attention.

420 *4.2. General consequences for practical uncertainty and reliability analysis*

421 *Utility of the bias description method.* In our opinion, the detection of sys-
422 tematic deviations between the assumed model structure and the data-generating
423 process is the most useful feature of the bias description method. For this
424 reason, we recommend that a model is inspected for bias by *(a)* adding a bias
425 term in the assumed model, specifying a prior for α concentrated around
426 a strictly positive value, as suggested here, and *(b)* inspecting the posterior
427 of α whenever an inappropriate model structure is suspected. Reformulation
428 of the error model (bias + measurement error) with α , σ , and τ proved very
429 helpful as it enables interpreting α as an indicator for the relative impor-
430 tance of model structure error. In cases where the posterior probability mass
431 is not shifted towards zero, relative to the prior, the modeler should suspect
432 the presence of bias. When this is detected, potential model improvements
433 may include the use of time-dependent parameters ([Reichert and Mieleitner,](#)
434 [2009](#); [Lin and Beck, 2012](#)) and/or input errors ([Del Giudice et al., 2016](#)) or
435 a change in model structure ([Del Giudice et al., 2015](#); [Mašić et al., 2017](#)).
436 While the method increases the reliability of the obtained steady-state pol-
437 lutant concentration predictions, it is important to note that the observation
438 of this benefit depends strongly on the root cause of the observed bias. For

439 this reason, detection of bias should be followed by exploratory analysis of
440 the residuals and development of a better model structure (e.g., Reichert
441 and Mieleitner, 2009; Del Giudice et al., 2013). We do not recommend ex-
442 ploiting the bias term for prediction without search for the underlying causes
443 for model structure deficits, especially considering that the ground truth is
444 rarely included in the produced credible intervals. Ultimately, the utility
445 of any approach depends on whether it can successfully describe the rele-
446 vant sources of the deviations between model predictions and the measured
447 variables (Brynjarsdóttir and O’Hagan, 2014; Wani et al., 2019).

448 *Parameter interpretation and transferability.* The mechanistic interpretation
449 of identified values for the parameters in the deterministic part of the model
450 is nearly impossible when bias is present. Adding an auto-correlated additive
451 error term contributes to a better reliability of the model predictions but can-
452 not provide a clearer interpretation of the parameter values or a direction to
453 a more appropriate model structure. Indeed, the parameter estimates remain
454 biased. Importantly, this is a likely scenario in wastewater engineering due to
455 the extremely simplified representation of biological processes during model
456 construction. Furthermore, obtaining proofs of a lack of bias is extremely
457 difficult to achieve so that a straightforward interpretation of parameter val-
458 ues is unlikely, even when the model structure may be appropriate. However,
459 grey-box or hybrid models may offer interpretability and transparency at the
460 cost of computational efforts (see introduction above).

461 *Data quality.* The quality of the simulated measurements in the studied case
462 is fairly high relative to current experience in the wastewater sector. However,
463 sensor hardware has become increasingly robust in the last three decades
464 (Olsson, 2012) and there is no obvious reason why this trend should stop
465 now. It is therefore reasonable to expect that the presence of bias can be
466 detected easily in the future, either by statistical tests for auto-correlation
467 of the residuals, as in the reference study, or with descriptive methods, as in
468 this study. This will also facilitate the modification of the model structure
469 in accordance to the envisioned high-quality data.

470 4.3. Future work

471 Through this work, we identified several avenues of further research.
472 These include:

- 473 • Develop and study methods for parameter estimation and parameter
474 interpretation under presence of model structure error.
- 475 • Develop a systematic approach to the formulation of prior distributions,
476 especially when flexibility is at odds with model structure or parameter
477 identifiability.
- 478 • Evaluation of experimental design methods to improve the chances of
479 detection of model structure errors.
- 480 • Adopt and evaluate methods to handle epistemic uncertainty in model-
481 based process design and operation.

482 5. Conclusions

483 In this paper, we investigated the challenge of structural model deficits in
484 risk-based reactor design. This is a relevant problem, because digitalization
485 will improve sensor resolution and spatial coverage of reactors, which will
486 reveal mismatches (i.e, bias) in our common engineering models (which have
487 been developed in the data-scarce past, often by grab sampling). Auto-
488 correlated mathematical formulations have been suggested to improve the
489 description of such biases.

490 In summary, our study shows that

- 491 • Adding auto-correlation terms in the measurement error model as a
492 way to account for model structure deficits significantly improves the
493 reliability of biokinetic models.
- 494 • Bias description enables accounting for predictive uncertainty during
495 process design to a large degree. This does not produce a guaranteed
496 reliability of the resulting design however. It is therefore not a bullet-
497 proof solution to the presence of model-reality mismatch.
- 498 • The studied bias description method is an adequate tool to identify the
499 presence of model structure deficits in presence of noisy experimental
500 data.

501 **Acknowledgments**

502 We thank Peter Reichert and Sanda Dejanic for their helpful insight into
503 the studied problem. Marc B. Neumann acknowledges financial support pro-
504 vided by the Spanish Government through the BC3 María de Maeztu ex-
505 cellence accreditation 2018-2022 (MDM-2017-0714) and the Ramón y Cajal
506 grant (RYC-2013-13628); and by the Basque Government through the BERC
507 2018-2021 program.

508 **6. Author contributions**

509 Dario del Giudice: Methodology, Software, Data analysis, Writing - Re-
510 view & Editing. Marc B. Neumann: Methodology, Software, Data analysis,
511 Writing - Review & Editing. Jörg Rieckermann: Conceptualization, Method-
512 ology, Writing - Review & Editing. Kris Villez: Methodology, Software,
513 Formal analysis, Data analysis, Visualization, Writing - Original Draft.

514 **References**

- 515 Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J.,
516 Lin, C.H., Tu, J., 2007. A framework for validation of computer models.
517 *Technometrics* 49, 138–154. doi:[10.1198/004017007000000092](https://doi.org/10.1198/004017007000000092).
- 518 Box, G.E.P., Tiao, G.C., 1973. *Bayesian Inference in Statistical Analysis*.
519 Addison-Wesley, Reading, MA, USA.
- 520 Brynjarsdóttir, J., O’Hagan, A., 2014. Learning about physical parame-
521 ters: The importance of model discrepancy. *Inverse Problems* 30, 114007.
522 doi:[10.1088/0266-5611/30/11/114007](https://doi.org/10.1088/0266-5611/30/11/114007).
- 523 Cagno, E., De Ambroggi, M., Grande, O., Trucco, P., 2011. Risk analysis
524 of underground infrastructures in urban areas. *Reliability Engineering &*
525 *System Safety* 96, 139–148. doi:[10.1016/j.ress.2010.07.011](https://doi.org/10.1016/j.ress.2010.07.011).
- 526 Cierkens, K., Plano, S., Benedetti, L., Weijers, S., de Jonge, J., Nopens, I.,
527 2012. Impact of influent data frequency and model structure on the quality
528 of WWTP model calibration and uncertainty. *Water Science & Technology*
529 65, 233–242. doi:[10.2166/wst.2012.081](https://doi.org/10.2166/wst.2012.081).

- 530 Craig, P.S., Goldstein, M., Rougier, J.C., Seheult, A.H., 2001. Bayesian
531 forecasting for complex systems using computer simulators. *Journal*
532 *of the American Statistical Association* 96, 717–729. doi:[10.1198/
533 016214501753168370](https://doi.org/10.1198/016214501753168370).
- 534 De Cesare, L., Myers, D.E., Posa, D., 2001. Estimating and modeling space-
535 time correlation structures. *Statistics & Probability Letters* 51, 9–14.
536 doi:[10.1016/S0167-7152\(00\)00131-0](https://doi.org/10.1016/S0167-7152(00)00131-0).
- 537 Del Giudice, D., Albert, C., Rieckermann, J., Reichert, P., 2016. Describing
538 the catchment-averaged precipitation as a stochastic process improves pa-
539 rameter and input estimation. *Water Resources Research* 52, 3162–3186.
540 doi:[10.1002/2015WR017871](https://doi.org/10.1002/2015WR017871).
- 541 Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., Rieck-
542 ermann, J., 2013. Improving uncertainty estimation in urban hydrological
543 modeling by statistically describing bias. *Hydrology and Earth System*
544 *Sciences* 17, 4209–4225. doi:[10.5194/hess-17-4209-2013](https://doi.org/10.5194/hess-17-4209-2013).
- 545 Del Giudice, D., Reichert, P., Bareš, V., Albert, C., Rieckermann, J., 2015.
546 Model bias and complexity - Understanding the effects of structural deficits
547 and input errors on runoff predictions. *Environmental Modelling & Soft-*
548 *ware* 64, 205–214. doi:[10.1016/j.envsoft.2014.11.006](https://doi.org/10.1016/j.envsoft.2014.11.006).
- 549 Dietzel, A., Reichert, P., 2012. Calibration of computationally demand-
550 ing and structurally uncertain models with an application to a lake wa-
551 ter quality model. *Environmental Modelling & Software* 38, 129–146.
552 doi:<http://dx.doi.org/10.1016/j.envsoft.2012.05.007>.
- 553 Dochain, D., Vanrolleghem, P.A., 2001. *Dynamical modelling & estimation*
554 *in wastewater treatment processes*. IWA Publishing, London, UK.
- 555 Dochain, D., Vanrolleghem, P.A., Van Daele, M., 1995. Structural identifica-
556 bility of biokinetic models of activated sludge respiration. *Water Research*
557 29, 2571–2578. doi:[10.1016/0043-1354\(95\)00106-U](https://doi.org/10.1016/0043-1354(95)00106-U).
- 558 Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. *Markov Chain Monte*
559 *Carlo in practice*. Chapman and Hall, London, UK.

- 560 Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-
561 time data. *Journal of the American Statistical Association* 97, 590–600.
562 doi:[10.1198/016214502760047113](https://doi.org/10.1198/016214502760047113).
- 563 Guo, M., Murphy, R.J., 2012. LCA data quality: sensitivity and uncertainty
564 analysis. *Science of the Total Environment* 435, 230–243. doi:[10.1016/j.
565 scitotenv.2012.07.006](https://doi.org/10.1016/j.scitotenv.2012.07.006).
- 566 Hastie, T., Tibshirani, R., Wainwright, M., 2015. *Statistical learning with
567 sparsity: the Lasso and generalizations*. Chapman and Hall/CRC, Boca
568 Raton, FL, USA.
- 569 Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., Van-
570 rollegheem, P.A., 2015. Efficiency criteria for environmental model quality
571 assessment: A review and its application to wastewater treatment. *Envi-
572 ronmental Modelling & Software* 68, 196–204. doi:[10.1016/j.envsoft.
573 2015.02.004](https://doi.org/10.1016/j.envsoft.2015.02.004).
- 574 Jensen, H.A., Jerez, D.J., 2018. A stochastic framework for reliability and
575 sensitivity analysis of large scale water distribution networks. *Reliability
576 Engineering & System Safety* 176, 80–92. doi:[10.1016/j.res.2018.04.
577 001](https://doi.org/10.1016/j.res.2018.04.001).
- 578 Kabir, G., Tesfamariam, S., Sadiq, R., 2015. Predicting water main failures
579 using Bayesian model averaging and survival modelling approach. *Relia-
580 bility Engineering & System Safety* 142, 498–514. doi:[10.1016/j.res.
581 2015.06.011](https://doi.org/10.1016/j.res.2015.06.011).
- 582 Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models.
583 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*
584 63, 425–464. doi:[10.1111/1467-9868.00294](https://doi.org/10.1111/1467-9868.00294).
- 585 Lin, Z.L., Beck, M.B., 2012. Accounting for structural error and uncertainty
586 in a model: An approach based on model parameters as stochastic pro-
587 cesses. *Environmental Modelling & Software* 27-28, 97–111. doi:[10.1016/
588 j.envsoft.2011.08.015](https://doi.org/10.1016/j.envsoft.2011.08.015).
- 589 Liu, C., Zachara, J.M., 2001. Uncertainties of Monod kinetic parameters
590 nonlinearly estimated from batch experiments. *Environmental Science &
591 Technology* 35, 133–141. doi:[10.1021/es001261b](https://doi.org/10.1021/es001261b).

- 592 Mašić, A., Srinivasan, S., Billeter, J., Bonvin, D., Villez, K., 2017. Shape
593 constrained splines as transparent black-box models for bioprocess mod-
594 eling. *Computers & Chemical Engineering* 99, 96–105. doi:[10.1016/j.
595 compchemeng.2016.12.017](https://doi.org/10.1016/j.compchemeng.2016.12.017).
- 596 Murphy, K.P., 2012. *Machine learning: A probabilistic perspective*. MIT
597 Press, Cambridge, MA, USA.
- 598 Neumann, M.B., Gujer, W., 2008. Underestimation of uncertainty in sta-
599 tistical regression of environmental models: Influence of model struc-
600 ture uncertainty. *Environmental Science & Technology* 42, 4037–4043.
601 doi:[10.1021/es702397q](https://doi.org/10.1021/es702397q).
- 602 Olsson, G., 2012. ICA and me - A subjective review. *Water Research* 46,
603 1585–1624.
- 604 Parsons, S., 2001. *Qualitative methods for reasoning under uncertainty*. MIT
605 Press, Cambridge, MA, USA.
- 606 Petersen, B., Gernaey, K., Devisscher, M., Dochain, D., Vanrolleghem, P.A.,
607 2003. A simplified method to assess structurally identifiable parameters
608 in Monod-based activated sludge models. *Water Research* 37, 2893–2904.
609 doi:[10.1016/S0043-1354\(03\)00114-3](https://doi.org/10.1016/S0043-1354(03)00114-3).
- 610 Rao, K.D., Kushwaha, H.S., Verma, A.K., Srividya, A., 2008. Epistemic
611 uncertainty propagation in reliability assessment of complex systems. *In-
612 ternational Journal of Performability Engineering* 4, 71–84.
- 613 Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty
614 of nonlinear dynamic models with stochastic, time-dependent parameters.
615 *Water Resources Research* 45, W10402. doi:[10.1029/2009WR007814](https://doi.org/10.1029/2009WR007814).
- 616 Reichert, P., Schuwirth, N., 2012. Linking statistical bias description to
617 multiobjective model calibration. *Water Resources Research* 48, W09543.
618 doi:[10.1029/2011WR011391](https://doi.org/10.1029/2011WR011391).
- 619 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Un-
620 derstanding predictive uncertainty in hydrologic modeling: The challenge
621 of identifying input and structural errors. *Water Resources Research* 46.
622 doi:[10.1029/2009WR008328](https://doi.org/10.1029/2009WR008328).

- 623 Scales, J.A., Tenorio, L., 2001. Prior information and uncertainty in inverse
624 problems. *Geophysics* 66, 389–397. doi:[10.1190/1.1444930](https://doi.org/10.1190/1.1444930).
- 625 Scheidegger, A., Leitao, J.P., Scholten, L., 2015. Statistical failure models
626 for water distribution pipes - A review from a unified perspective. *Water*
627 *Research* 83, 237–247. doi:[10.1016/j.watres.2015.06.027](https://doi.org/10.1016/j.watres.2015.06.027).
- 628 Scheidegger, A., Scholten, L., Maurer, M., Reichert, P., 2013. Extension of
629 pipe failure models to consider the absence of data from replaced pipes.
630 *Water Research* 47, 3696–3705. doi:[10.1016/j.watres.2013.04.017](https://doi.org/10.1016/j.watres.2013.04.017).
- 631 Sin, G., Gernaey, K.V., Neumann, M.B., van Loosdrecht, M.C., Gujer, W.,
632 2009. Uncertainty analysis in WWTP model applications: a critical dis-
633 cussion using an example from design. *Water Research* 43, 2894–2906.
634 doi:[10.1016/j.watres.2009.03.048](https://doi.org/10.1016/j.watres.2009.03.048).
- 635 Stein, M.L., 2005. Space-time covariance functions. *Journal of the American*
636 *Statistical Association* 100, 310–321. doi:[10.1198/016214504000000854](https://doi.org/10.1198/016214504000000854).
- 637 Van Griensven, A., Meixner, T., 2007. A global and efficient multi-
638 objective auto-calibration and uncertainty estimation method for wa-
639 ter quality catchment models. *Journal of Hydroinformatics* 9, 277–291.
640 doi:[10.2166/hydro.2007.104](https://doi.org/10.2166/hydro.2007.104).
- 641 Vihola, M., 2012. Robust adaptive metropolis algorithm with coerced ac-
642 ceptance rate. *Statistics and Computing* 22, 997–1008. doi:[10.1007/
643 s11222-011-9269-5](https://doi.org/10.1007/s11222-011-9269-5).
- 644 Wani, O., Scheidegger, A., Cecinati, F., Espadas, G., Rieckermann, J., 2019.
645 Exploring a copula-based alternative to additive error models for non-
646 negative and autocorrelated time series in hydrology. *Journal of Hydrology*
647 575, 1031–1040. doi:[10.1016/j.jhydrol.2019.06.006](https://doi.org/10.1016/j.jhydrol.2019.06.006).