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-Franscini-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)

1. Introduction

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

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

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

Recently, statisticians have been suggesting a promising approach to solve this dilemma. The underlying idea is to not assume identically and independently distributed (i.i.d.) errors for mismatches between models and observations (Liu and Zachara, 2001), but to explicitly account for mismatches by adding a stochastic auto-correlated process to the i.i.d. measurement error model (Craig et al., 2001; Kennedy and O'Hagan, 2001; Bayarri et al., 2007). This is known as the bias description method and focuses on the modeling of the symptoms of a mismatch between model structure and reality. While this does not identify or tackle the root cause of these symptoms, it has been proven to be a computationally efficient tool to increase the reliability of model-based predictions compared to standard regression approaches in a variety of systems from lakes to natural catchments to urban hydrology (Dietzel and Reichert, 2012; Reichert and Schuwirth, 2012; Del Giudice et al., 2015). Therefore, we expect that the bias description method also improves the reliability of predictions with structurally deficient wastewater treatment models in view of risk-based design. Specifically, adding a stochastic measurement error term to a model given the same amount of experimental measurements is expected to reduce the relative information-richness of the experimental data and lead to larger credibility intervals of model parameters, wider prediction intervals and, by avoiding overconfident predictions, a more trustworthy design.

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

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

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

2. Material and methods

2.1. Applied error models

In a vast majority of environmental modeling studies, the measurement error is assumed to be i.i.d. For example, Hauduc et al. (2015) compares an extensive list of model performance criteria for wastewater treatment modelling yet does not list any criterion which accounts for autocorrelated model prediction errors. One approach considered in Cierkens et al. (2012), consists of downsampling time series to avoid the appearance of autocorrelation. As explained in the same study, this leads to an inefficient use of the available data and, more importantly, cannot account at all for model structure deficits as a potential root cause of autocorrelated residuals. Ignoring the presence of autocorrelated residuals was shown to lead to overconfidence in the produced model and, subsequently, poor decision-making, as was shown also in the reference study. Most often, a Gaussian distribution is assumed for the measurement errors. Such a model of measurement error cannot account for systematic deviations between the assumed model and the observed measurements, i.e. bias. One way of accounting for bias is by adding terms, such as a stochastic autocorrelated error bias term, to the measurement equation (Craig et al., 2001; Kennedy and O'Hagan, 2001; Bayarri et al., 2007). In this work, we describe the observable output time-series (i.e., measured concentration, y_o) as a sum of a deterministic dynamic model output (y, the modeled concentration), a classical Gaussian measurement error ($\mathrm{e}(\psi)$), and an auto-correlated error term ($\mathrm{b}(\psi)$):

$$\mathrm{y}_o(\theta, \gamma, \psi) = \mathrm{y}(\theta) + \gamma + \mathrm{b}(\psi) + \mathrm{e}(\psi) \tag{1}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.1.1. Model parameter estimation

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

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

Bayesian calibration aims at characterizing the distribution described by the posterior likelihood $\mathcal{L}(\theta, \psi|\mathrm{y}_o) \propto \mathcal{L}(\mathrm{y}_o|\theta, \psi) \cdot \mathcal{L}(\theta, \psi)$, where the prior likelihood $\mathcal{L}(\theta, \psi)$ expresses the prior beliefs about the parameters. In this work, the posterior distribution is approximated with a sample of $\mathcal{L}(\theta, \psi|\mathrm{y}_o)$ drawn with a Markov Chain Monte Carlo (MCMC) sampler (see "Numerical implementation").

2.2. Biokinetic model parameter identification with batch experiments

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

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

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

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

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

$$y_o(t) = s(t) + e(t) \tag{11}$$
$$e(t) \sim N(0, \sigma_e) \tag{12}$$

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

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

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

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

$$y_o(t) = s(t) + b(t) + e(t) \tag{14}$$
$$e(t) \sim \mathcal{N}(0, \sigma_e) \tag{15}$$
$$\mathrm{b} \sim \mathcal{N}(0, \Sigma_b(\sigma_b, \tau)) \tag{16}$$

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

2.3. Numerical implementation

The biokinetic model, parameter estimation, and uncertainty propagation were implemented in Matlab (R2019a). The prior probabilities for the parameters were set based on the authors' experience. They are all independent of each other. All priors are uniform, except for σ and τ. The prior likelihood for σ is proportional to its inverse and is equivalent to the Jeffreys prior conditional to fixed values for all other parameters (see Box and Tiao, 1973). The prior likelihood for τ is the sine function supported between 0 and $2\,T$. This prior equals zero at $\tau = 0$ and $\tau = 2\,T$ and one at $\tau = T$. This expresses the subjective belief that the autocorrelation length of the deviations due to model structure error is expected to be similar to the duration of the experiment. The priors are specified completely in Table 1. We first run an adaptive MCMC algorithm (Vihola, 2012) to find a good guess for the maximum a posteriori estimates and a good proposal variance-covariance matrix. With these results, we execute a (non-adaptive) MCMC algorithm to obtain $20,000$ samples from $\mathcal{L}(\theta, \psi|\mathrm{y}_o)$. The first $10,000$ samples are considered to correspond to the burn-in phase of the sampler, during which effects of the initial sample may still be apparent. These samples are therefore discarded, 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	$+\infty$	$g/m^3.h$
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

2.4. Using information gained in parameter estimation for design

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

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

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

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

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

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

2.5. Software availability

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

3. Results

3.1. Parameter estimation

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

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

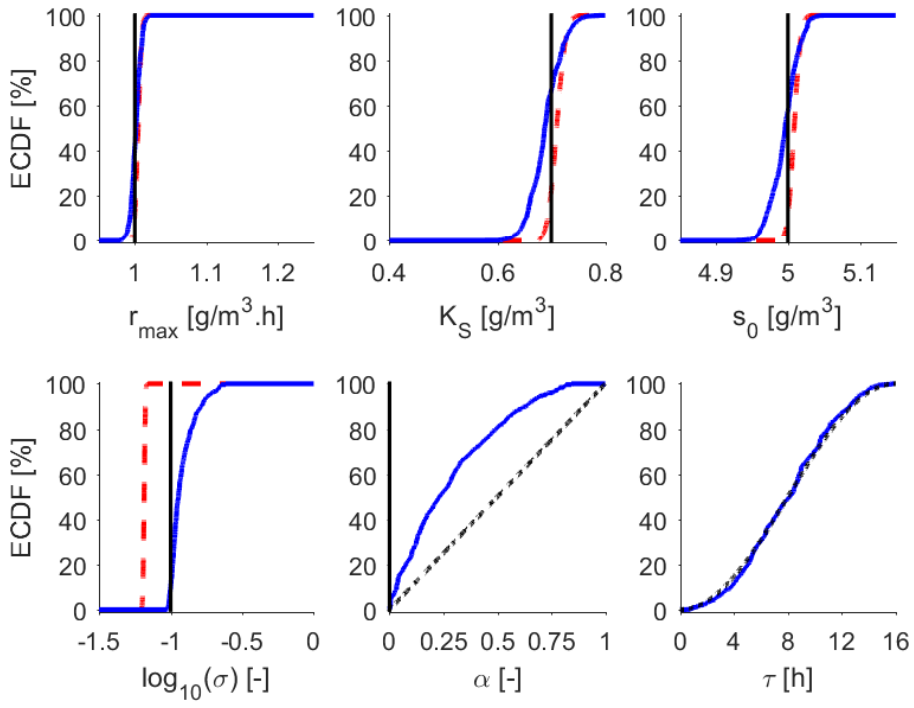

Figure 1: Batch experiment simulated with the Tessier model and $K_s = 0.7\ g/m^3$ and $\sigma_e = 0.01\ g/m^3$. Marginal posterior distributions for the calibrated Tessier model without bias description (dashed red line) and with bias description (full blue line). Vertical full lines indicate the true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

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

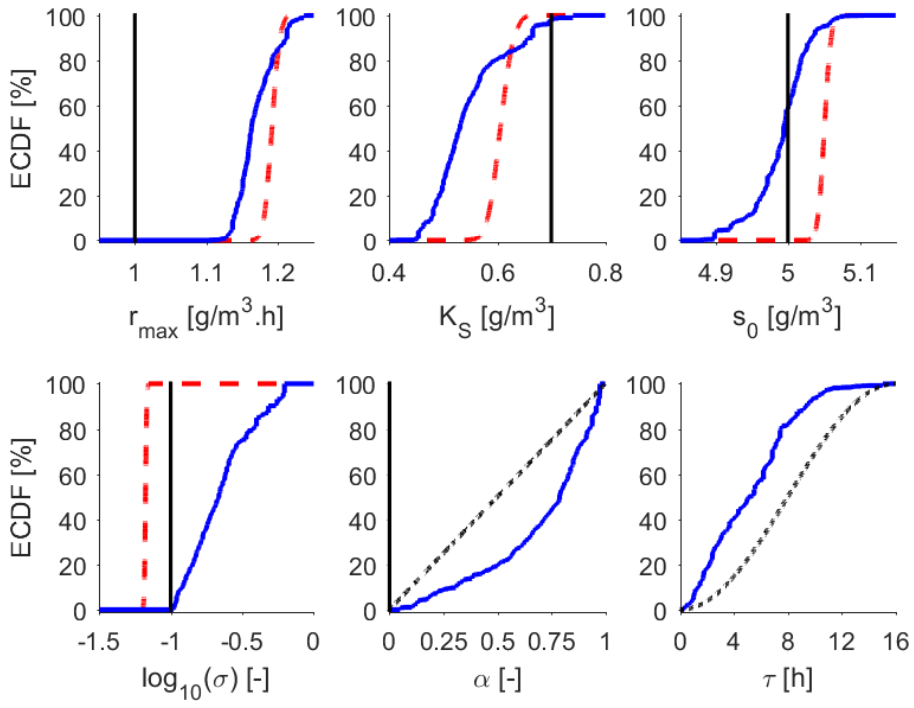

Figure 2: Batch experiment simulated with the Monod model and $K_s = 0.7\ g/m^3$ and $\sigma_e = 0.01\ g/m^3$. Marginal posterior distributions for the calibrated Monod model without bias description (dashed red line) and with bias description (full blue line). Vertical full lines indicate the data-generating 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).

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

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

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

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

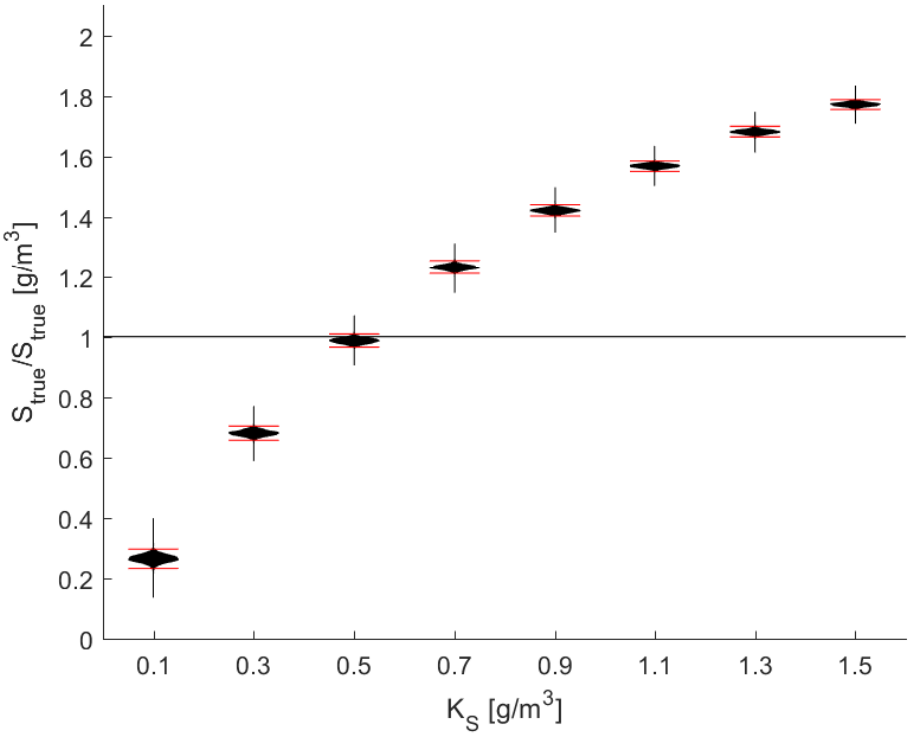

Figure 3: Distributions of the ratio of predicted steady state concentrations to the ground truth concentration as a function of the affinity constant (K_S) – Low noise case ($\sigma_e = 0.01\ g/m^3$). Red horizontal whiskers indicate the two-sided 99% credible intervals. Results are shown for the Monod 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.

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

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

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

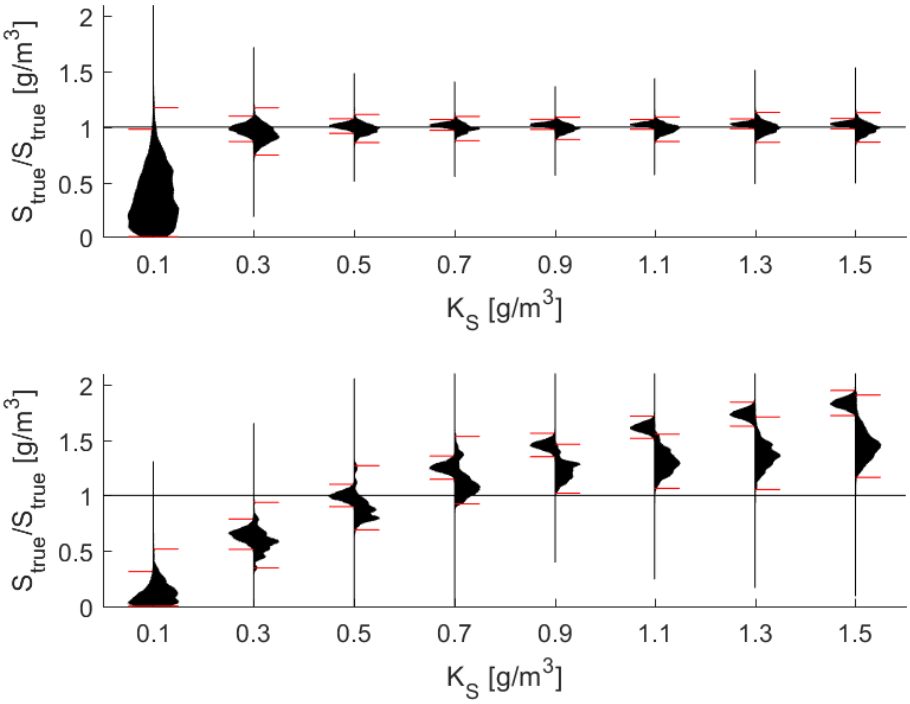

Figure 4: Distributions of the ratio of predicted steady state concentrations to the ground truth concentration as a function of the affinity constant (K_S) – High noise case (σ_e = 0.1 g/m^3). Red horizontal whiskers indicate the two-sided 99% credible intervals. **Left side beans:** without bias description; **Right side beans:** with bias description. **Top:** Tessier model - All credible intervals include the ideal ratio (equal to 1), except for the simulation with $K_S = 0.1$ 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).

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

In the supplementary information, we provide results obtained with the modified method. We omit the information obtained during the washout experiment during prediction. In this case, the uncertainty in the predictions

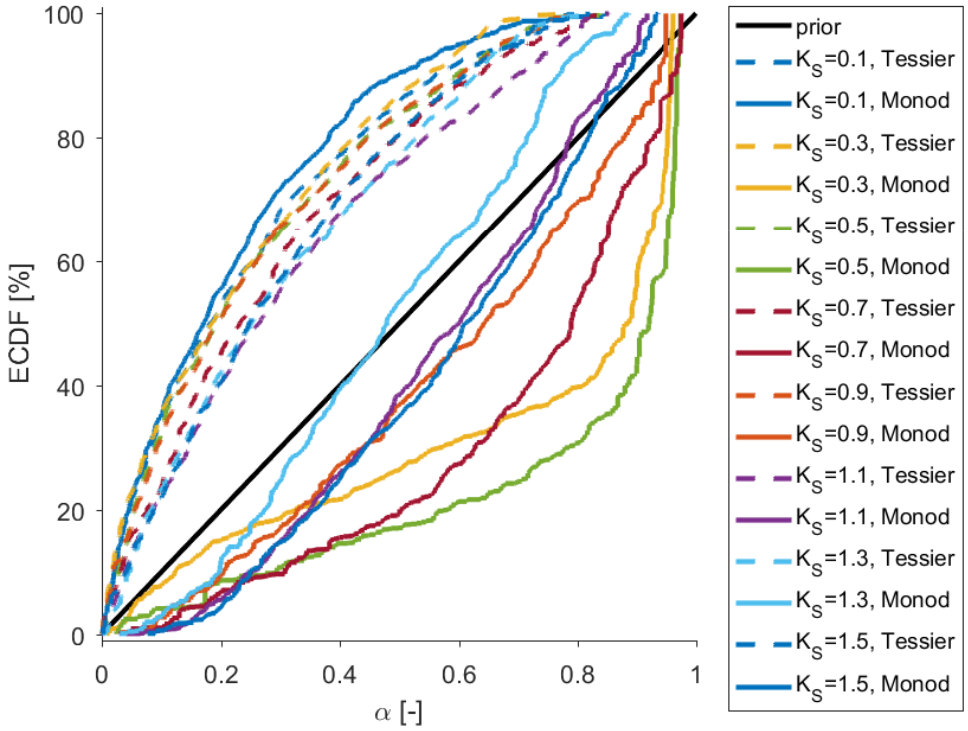

Figure 5: Distributions of the parameter α for both models with a bias term in all high-noise cases ($\sigma_e = 0.1\ g/m^3$). When the Tessier model is selected (no model structure error), the posterior distribution of α is shifted to the left of the prior, thus suggesting the kinetic model structure 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 ($K_S = 0.1\ g/m^3$), thus providing a useful indication of model structure error.

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

4. Discussion

4.1. Summary and limitations of the experimental simulation study

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

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

- The general applicability of the bias description method is not demonstrated. However, the bias description method could easily be adapted to more complex systems. One could incorporate a bias term to express correlation between multiple measurements, of the same or distinct variables measured in the same location or different locations. In this case, the covariance between two measurements, as expressed by Σ, would not only be a function of *(a)* the time difference ($t_i - t_j$, see (6)),

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

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

4.2. General consequences for practical uncertainty and reliability analysis

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

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

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

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

4.3. Future work

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

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

5. Conclusions

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

In summary, our study shows that

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

Acknowledgments

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

6. Author contributions

Dario del Giudice: Methodology, Software, Data analysis, Writing - Review & Editing. Marc B. Neumann: Methodology, Software, Data analysis, Writing - Review & Editing. Jörg Rieckermann: Conceptualization, Methodology, Writing - Review & Editing. Kris Villez: Methodology, Software, Formal analysis, Data analysis, Visualization, Writing - Original Draft.

References

Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., Tu, J., 2007. A framework for validation of computer models. Technometrics 49, 138–154. doi:10.1198/004017007000000092.

Box, G.E.P., Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA, USA.

Brynjarsdóttir, J., O'Hagan, A., 2014. Learning about physical parameters: The importance of model discrepancy. Inverse Problems 30, 114007. doi:10.1088/0266-5611/30/11/114007.

Cagno, E., De Ambroggi, M., Grande, O., Trucco, P., 2011. Risk analysis of underground infrastructures in urban areas. Reliability Engineering & System Safety 96, 139–148. doi:10.1016/j.ress.2010.07.011.

Cierkens, K., Plano, S., Benedetti, L., Weijers, S., de Jonge, J., Nopens, I., 2012. Impact of influent data frequency and model structure on the quality of WWTP model calibration and uncertainty. Water Science & Technology 65, 233–242. doi:10.2166/wst.2012.081.

Craig, P.S., Goldstein, M., Rougier, J.C., Seheult, A.H., 2001. Bayesian forecasting for complex systems using computer simulators. Journal of the American Statistical Association 96, 717–729. doi:10.1198/016214501753168370.

De Cesare, L., Myers, D.E., Posa, D., 2001. Estimating and modeling space-time correlation structures. Statistics & Probability Letters 51, 9–14. doi:10.1016/S0167-7152(00)00131-0.

Del Giudice, D., Albert, C., Rieckermann, J., Reichert, P., 2016. Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation. Water Resources Research 52, 3162–3186. doi:10.1002/2015WR017871.

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., Rieckermann, J., 2013. Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. Hydrology and Earth System Sciences 17, 4209–4225. doi:10.5194/hess-17-4209-2013.

Del Giudice, D., Reichert, P., Bareš, V., Albert, C., Rieckermann, J., 2015. Model bias and complexity - Understanding the effects of structural deficits and input errors on runoff predictions. Environmental Modelling & Software 64, 205–214. doi:10.1016/j.envsoft.2014.11.006.

Dietzel, A., Reichert, P., 2012. Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model. Environmental Modelling & Software 38, 129–146. doi:http://dx.doi.org/10.1016/j.envsoft.2012.05.007.

Dochain, D., Vanrolleghem, P.A., 2001. Dynamical modelling & estimation in wastewater treatment processes. IWA Publishing, London, UK.

Dochain, D., Vanrolleghem, P.A., Van Daele, M., 1995. Structural identifiability of biokinetic models of activated sludge respiration. Water Research 29, 2571–2578. doi:10.1016/0043-1354(95)00106-U.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in practice. Chapman and Hall, London, UK.

Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. Journal of the American Statistical Association 97, 590–600. doi:10.1198/016214502760047113.

Guo, M., Murphy, R.J., 2012. LCA data quality: sensitivity and uncertainty analysis. Science of the Total Environment 435, 230–243. doi:10.1016/j.scitotenv.2012.07.006.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical learning with sparsity: the Lasso and generalizations. Chapman and Hall/CRC, Boca Raton, FL, USA.

Hauduc, H., Neumann, M.B., Muschalla, D., Gamerith, V., Gillot, S., Vanrolleghem, P.A., 2015. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. Environmental Modelling & Software 68, 196–204. doi:10.1016/j.envsoft.2015.02.004.

Jensen, H.A., Jerez, D.J., 2018. A stochastic framework for reliability and sensitivity analysis of large scale water distribution networks. Reliability Engineering & System Safety 176, 80–92. doi:10.1016/j.ress.2018.04.001.

Kabir, G., Tesfamariam, S., Sadiq, R., 2015. Predicting water main failures using Bayesian model averaging and survival modelling approach. Reliability Engineering & System Safety 142, 498–514. doi:10.1016/j.ress.2015.06.011.

Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, 425–464. doi:10.1111/1467-9868.00294.

Lin, Z.L., Beck, M.B., 2012. Accounting for structural error and uncertainty in a model: An approach based on model parameters as stochastic processes. Environmental Modelling & Software 27-28, 97–111. doi:10.1016/j.envsoft.2011.08.015.

Liu, C., Zachara, J.M., 2001. Uncertainties of Monod kinetic parameters nonlinearly estimated from batch experiments. Environmental Science & Technology 35, 133–141. doi:10.1021/es001261b.

Mašić, A., Srinivasan, S., Billeter, J., Bonvin, D., Villez, K., 2017. Shape constrained splines as transparent black-box models for bioprocess modeling. Computers & Chemical Engineering 99, 96–105. doi:10.1016/j.compchemeng.2016.12.017.

Murphy, K.P., 2012. Machine learning: A probabilistic perspective. MIT Press, Cambridge, MA, USA.

Neumann, M.B., Gujer, W., 2008. Underestimation of uncertainty in statistical regression of environmental models: Influence of model structure uncertainty. Environmental Science & Technology 42, 4037–4043. doi:10.1021/es702397q.

Olsson, G., 2012. ICA and me - A subjective review. Water Research 46, 1585–1624.

Parsons, S., 2001. Qualitative methods for reasoning under uncertainty. MIT Press, Cambridge, MA, USA.

Petersen, B., Gernaey, K., Devisscher, M., Dochain, D., Vanrolleghem, P.A., 2003. A simplified method to assess structurally identifiable parameters in Monod-based activated sludge models. Water Research 37, 2893–2904. doi:10.1016/S0043-1354(03)00114-3.

Rao, K.D., Kushwaha, H.S., Verma, A.K., Srividya, A., 2008. Epistemic uncertainty propagation in reliability assessment of complex systems. International Journal of Performability Engineering 4, 71–84.

Reichert, P., Mieleitner, J., 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resources Research 45, W10402. doi:10.1029/2009WR007814.

Reichert, P., Schuwirth, N., 2012. Linking statistical bias description to multiobjective model calibration. Water Resources Research 48, W09543. doi:10.1029/2011WR011391.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research 46. doi:10.1029/2009WR008328.

Scales, J.A., Tenorio, L., 2001. Prior information and uncertainty in inverse problems. Geophysics 66, 389–397. doi:10.1190/1.1444930.

Scheidegger, A., Leitao, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes - A review from a unified perspective. Water Research 83, 237–247. doi:10.1016/j.watres.2015.06.027.

Scheidegger, A., Scholten, L., Maurer, M., Reichert, P., 2013. Extension of pipe failure models to consider the absence of data from replaced pipes. Water Research 47, 3696–3705. doi:10.1016/j.watres.2013.04.017.

Sin, G., Gernaey, K.V., Neumann, M.B., van Loosdrecht, M.C., Gujer, W., 2009. Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Research 43, 2894–2906. doi:10.1016/j.watres.2009.03.048.

Stein, M.L., 2005. Space-time covariance functions. Journal of the American Statistical Association 100, 310–321. doi:10.1198/016214504000000854.

Van Griensven, A., Meixner, T., 2007. A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. Journal of Hydroinformatics 9, 277–291. doi:10.2166/hydro.2007.104.

Vihola, M., 2012. Robust adaptive metropolis algorithm with coerced acceptance rate. Statistics and Computing 22, 997–1008. doi:10.1007/s11222-011-9269-5.

Wani, O., Scheidegger, A., Cecinati, F., Espadas, G., Rieckermann, J., 2019. Exploring a copula-based alternative to additive error models for non-negative and autocorrelated time series in hydrology. Journal of Hydrology 575, 1031–1040. doi:10.1016/j.jhydrol.2019.06.006.