Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 22, Iss. 1, pp. 4-16, 2024

Inference of the parameters of a bioreactor via approximate Bayesian evidence

Filipe P. de Farias, Davi de Lima Cruz, Michela Mulas
Department of Teleinformatics Engineering, Federal University of Ceará
{filipepfarias, davilcruz}@alu.ufc.br, michela.mulas@ufc.br

Resumo – No âmbito da recente crise hídrica, o desenvolvimento e melhoria de tecnologias para tratar eficientemente águas residuais se torna cada vez mais necessário. Alinhado a isto, o processo de lodo ativado é amplamente usado nas plantas biológicas de tratamento de águas residuais para remover dela carbono e nutrientes. Embora grandes avanços tenham ocorrido nos últimos anos na modelagem desses processos para melhorar sua performance e eficiência energética, os modelos obtidos ainda são complexos e difíceis de calibrar utilizando dados de estações reais. Neste trabalho utilizamos frameworks de inferência Bayesiana para estimar os parâmetros do processo de lodo ativado. Empregamos os processos Gaussianos para estimar a solução das equações diferenciais do modelo do processo de lodo ativado e ajustamos essa estimativa à dados fabricados de água residual. Obtemos que o uso do framework de inferência Bayesiana se sobressai à abordagem clássica em cenários em que medições são mais fortemente afetadas por ruído.

Palavras-chave –Lodo ativado, tratamento de águas residuais, modelagem de processos, Inferência Bayesiana, Fenrir, Probabilistic Numerics.

Abstract –In the context of the recent water crisis, the development and improvement of technologies for efficiently treating wastewater becomes increasingly necessary. Aligned with this, the activated sludge process is widely used in biological wastewater treatment plants to remove carbon and nutrients from wastewater. Although significant advancements have been achieved in recent years in modelling these processes to improve their performance and energy efficiency, the models are still complex and difficult to calibrate using data from real plants. In this work, we use Bayesian inference frameworks to estimate the parameters of the activated sludge process. We employ Gaussian processes to estimate the solution to the differential equations of the activated sludge process model and we fit this estimation to manufactured data of the observations of the wastewater. We find that the use of the Bayesian inference framework outperforms the classical approach in scenarios where measurements are more strongly affected by noise.

Keywords – Activated sludge process, Wastewater treatment plants, Process modeling, Bayesian inference, Fenrir, Probabilistic Numerics.

1 Introduction

Wastewater treatment plants play an essential role in preserving the environment and public health contributing significantly to the achievement of multiple Sustainable Development Goals (SDGs, [1]) by promoting environmental sustainability, public health, and socio-economic well-being. In a biological wastewater treatment plant (WWTP), activated sludge process (ASP) is a widely method for carbon and nutrients removal [2, 3]. The process aims to reduce the concentration of biodegradable matter and nutrients in the effluent, with the compromise of keeping low sludge production using microorganisms mixed with suspended solids in the mixture to treat domestic and industrial wastewater. Several efforts have been made in the last decades to improve these processes in terms of energy efficiency and operational cost optimisation [4]. To aid this, the modelling of such systems is fundamental to predict their operational states without dealing with the high costs of the real plant [5, 6]. The first Activated Sludge Model (ASM1, [7]) was developed to describe the removal of organic carbon substances and nitrogen with simultaneous consumption of oxygen and nitrate as electron acceptors in urban wastewater treatment plants. Later, a full family of ASP models (ASM2, ASM2d, ASM3, [8]) was developed to take into account the even more complex processes such as biological phosphorus removal and, more recently, also process-related greenhouse gas emissions (especially N_2O, see for instance [9]).

Despite these significant advances and the development of more complete mathematical models a common issue remains the overparametrisation of these models, that include many kinetic and stoichiometric parameters to be adjusted for taking into account the variations in microbial community structure and dominant species [10, 11]. In addition, data from full-scale plants or pilot- and lab-scale reactors may not yield reliable estimations for all parameters simultaneously. Model calibration is crucial to ensure that the model accurately represents the multifaceted interactions within the treatment system, allows the specific characteristics of the wastewater being treated and ensures the model's applicability to diverse scenarios. To tackle that, guidelines have been developed [12] to approach the calibration of the ASM models. Such guidelines consist of defining modelling objectives, collecting the data, setting up the model, and calibration. The later is in general the most challenging step once the parameters of the model should be optimised to match with the data of the WWTP such the model be a realistic representation of the process.

In general, the approach used to estimate the parameters of bioreactors consists of simulating the system and evaluating the difference between the states and observed data available from the WWTP [13–16]. This mostly involves four steps: i) characterisation of the influent wastewater; ii) definition of the dynamic influent loading data; iii) manual estimation of the model parameters; and iv) model validation. Few works adopt a probabilistic approach [17–19], whereas other procedures use information about the gradient of the dynamics of the system to help with the computational burden [20]. Systematic analysis approaches mainly consist of parameter identification [21], sensitivity analysis [22], and error propagation [23], which have been developed and tested. A numerical optimal approaching procedure has been used for the systematic calibration of ASMs [24]. By means of global factors sensitivity analysis for factors fixing and pseudo-global parameter correlation analysis for non-identifiable factors detection, the procedure effectively determines the optimal parameter subset and performs model calibration for two different ASP systems. Recently, a steady-state calibration methodology was proposed [25] to effectively tackle key challenges in WWTP modelling, showing that accurate calibration is still an open problem crucial for producing reliable results.

In this work, we propose: to use the framework of probabilistic numerics [26] to address the uncertainty over the the ASMs using Gaussian process (GP); and to address the estimation of the parameters via the optimization of the marginal likelihood, differently of the maximum likelihood approach [18,27]. Also the use of the language Julia [28] for the modelling and calibration of ASMs, which is not known by the authors to have been used before for the calibration of wastewater treatment models, as well as, although outside the scope of the article, its automatic differentiation environment [29] to speedup the calculations. Our study is presented as follows. Section 2 overviews the methodology and the assumptions made to consider the uncertainty over the dynamical model. In Section 2.1 we follow [30] to set the GP representation of the solution of the dynamical model. Section 2.2 details the steps of the Fenrir framework [31] used in this work to define the marginal likelihood to estimate the parameters of the ASM which fits the data better. An overview of the method adopted to perform the sensitivity analysis is presented in Section 2.3 for the selection of the model parameters. The ASM1 model is briefly presented in Section 3 and, eventually, the results obtained for the calibration of the model are provided in Section 4.

2 Theory and Methodology

We consider the state-space representation of the dynamical system of the bioreactor model in the equations (1).

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f(x(t), \theta, t), \tag{1a}$$

$$y(t) = g(x(t), \theta, t). \tag{1b}$$

The evolution in time of the state $x(t) \in \mathbb{R}^{D_x}$ depends on a given initial value $x_\dagger$ and the vector of parameters θ. The measurement of the states are described by $y(t) \in \mathbb{R}^{D_y}$. We follow [31] to propose a probability model in which θ are the parameters of the differential equation of the bioreactor model which are to be identified using a Bayesian approach. We consider the solution $x(t)$ of equation (1a) evaluated at each time step of $\{t_1, t_2, ..., t_n\} = \{t_n\}_{n=1}^N$ giving $\{x_n\}_{n=1}^N = \{x(t_n)\}_{n=1}^N$. The measurements occur at times $\{t_n\}_{n=1}^N$ giving $\{y_n\}_{n=1}^N = \{y(t_n)\}_{n=1}^N$.

The Bayesian formulation of this problem is summarised by the following probabilistic proposition: We treat x_n and y_n as being the random variables X_n and Y_n which are marginalizations of the respective random processes $X(t)$ and $Y(t)$. We assume the following probability density functions

$$p(X(t) \mid \theta) = \mathcal{GP}(m(t), k(t, t') \mid \theta), \tag{2a}$$

$$p(Y_n \mid X_n, \theta) = \mathcal{N}(g(X_n), \sigma_y^2). \tag{2b}$$

In equation (2a) the time evolution of the states is modelled by a Gaussian process as $X(t)$ with the same dimension of $x(t)$ and being described by its mean $m(t)$ and covariance $k(t, t')$ functions [32]. The measurements are assumed to be corrupted by additive Gaussian noise in equation (1b) resulting in the Gaussian density in equation (2b).

The Bayesian inference framework consists of fitting a probability model to a set of data and representing the result of the fitting by another probability distribution of the non-observed (or latent) states and conditioned by the observations and the parameters of the probability model [33]. The assumption of the probability model allows to approach the uncertainties we may have about any aspect of the problem to be solved through the use of probability distributions and reason about these assumptions using the probability theory [34,35]. We define the following notation: the latent variable X with the probability density $p(X|\theta)$ called prior probability; the observed variable $Y(t)$ with the probability $p(Y|X, \theta)$ likelihood modelling the relation between the latent variables and the observations. The variable θ represents the parameters of the model. The Bayesian inference consists in evaluating the Bayes' rule such that

$$p(X|Y, \theta) = \frac{p(Y|X, \theta)p(X|\theta)}{\Omega(Y, \theta)},$$

where $p(X|Y, \theta)$ is the probability density, called posterior probability, of the probability model fitted to the data. To ensure $p(X|Y, \theta)$ to be a probability density, the numerator is normalised by $\Omega(Y, \theta) = \int p(Y|X, \theta)p(X|\theta)\mathrm{d}X$ known as marginal likelihood or evidence. $\Omega(Y, \theta)$ depends on both the observations Y and the parameters θ. For a fixed set of observations

$Y = \{Y_1, ..., Y_N\}$, the evidence ranks the different models represented by different values of θ [36]. Our objective is to find the best ASM model, i.e. the best set of parameters θ, based on the maximisation of $\Omega(\{Y_1, ..., Y_N\}, \theta)$. The problem may be stated equivalently due to the non-decreasing behaviour of the logarithm as finding $\theta^\star$ such that

$$\theta^\star = \arg\min_\theta \left\{ -\log \Omega(\{Y_n\}_{n=1}^N, \theta) \right\}. \tag{3}$$

where we express the marginal likelihood as

$$\Omega(\{Y_n\}_{n=1}^N, \theta) = \int p(\{Y_n\}_{n=1}^N | \{X_n\}_{n=1}^N, \theta) p(\{X_n\}_{n=1}^N | \theta) \mathrm{d}X_1 ... \mathrm{d}X_N, \tag{4}$$

quantifying the plausibility of the data y_n to have been generated by the solution X_n for $n = 1, ..., N$ and the parameters θ.

In Section 2.1 we construct the prior used to evaluate $p(\{X_n\}_{n=1}^N | \theta)$ and in Section 2.2 we compute the integration in equation (4). In Section 2.3 we define the theoretical background for the sensitivity analysis needed for the selection of the parameters θ to be calibrated in the dynamical model understudy.

2.1 Prior model

In this section, we build the solution to the system of differential equations as a GP prior to be used in the marginal likelihood. A previous work [37] treats $f(x(t), \theta, t)$ in equation (1a) as a GP on which a regression is made by fitting it to the measurements Y. In [38] $f(x(t), \theta, t)$ is provided as a constraint to the GP prior which together with the measurements Y constructs the *maximum a posteriori* estimate of the parameters and the states of the differential equation. The work of [31] which is followed in this paper is similar to the former, but assuming a GP with Markovian properties. This allows a state-space representation of the GP alleviating the known computational burden of the GP regression. We refer to [39, 40] for further discussions on state-space representation of GP with stationary covariance functions. The main idea is that the solution of special stochastic differential equations (SDEs, [41]) can be represented as a GP and vice-versa. As the ordinary differential equations, these special SDEs admit a state-space representation and the probability distribution of the solution can be obtained via Kalman filter.

To obtain the GP in state-space form we define the vector $\mathbf{X}(t)$, the collection of time derivatives of $X(t)$ in equation (2a), as

$$\mathbf{X}(t) = \left[\left(X^{(0)}(t) \right)^\top, \left(X^{(1)}(t) \right)^\top, ..., \left(X^{(q)}(t) \right)^\top \right]^\top. \tag{5}$$

being $X^{(q)}(t)$ the q-th time derivative of $X(t)$. Then, we define the following SDE as

$$\mathrm{d}\mathbf{X}(t) = \mathbf{F}\mathbf{X}(t)\mathrm{d}t + \mathbf{L}\mathrm{d}W(t). \tag{6}$$

To use the matrix form we define is the m-th canonical basis vector $\mathrm{e_m}$[1] in $\mathbb{R}^{q+1}$, I_{D_x} a identity matrix of size $D_x \times D_x$ and $\otimes$ the Kronecker product. $\mathbf{F} \in \mathbb{R}^{D_x(q+1) \times D_x(q+1)}$ is the state transition matrix of the SDE, $\mathbf{L} \in \mathbb{R}^{D_x(q+1)}$ is a diffusion matrix and $\mathrm{d}W$ is the increment of the Wiener process. The solution of equation (6) is

$$\mathbf{X}(t) = \exp(\mathbf{F}t)\mathbf{X}_\dagger + \int_0^t \exp(\mathbf{F}t)\mathbf{L}dW(\tau). \tag{7}$$

If the initial condition $\mathbf{X}_\dagger$ for $t = t_\dagger$ is Gaussian distributed, i.e. $\mathbf{X}_\dagger \sim \mathcal{N}(\boldsymbol{\mu}_\dagger, \boldsymbol{\Sigma}_\dagger)$, the solution $\mathbf{X}(t)$ is a GP such that [42]

$$\begin{aligned}
\mathbf{X}(t) &\sim \mathcal{GP}\left(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t)\right), \\
\boldsymbol{\mu}(t) &= \exp(\mathbf{F}t)\boldsymbol{\mu}_\dagger, \\
\boldsymbol{\Sigma}(t) &= \exp(\mathbf{F}t)\boldsymbol{\Sigma}_\dagger \left[\exp(\mathbf{F}t)\right]^\top + \mathbf{Q}, \\
\mathbf{A}(t - t_\dagger) &= \exp(\mathbf{F}h), \\
\mathbf{Q}(t - t_\dagger) &= \int_0^{t - t_\dagger} \exp(\mathbf{F}(t - t_\dagger - \tau))\mathbf{L}\mathbf{L}^\top \exp(\mathbf{F}^\top(t - t_\dagger - \tau))\mathrm{d}\tau,
\end{aligned}$$

We proceed by considering the q-times integrated Wiener process, $\mathrm{IWP}(q)$, defined by the assumption of $\mathbf{F} = \sum_{i=1}^q \mathrm{e}_i \mathrm{e}_{i+1}^\top \otimes I_{D_x}$ and $\mathbf{L} = \sqrt{\kappa}\mathrm{e}_{q+1} \otimes I_{D_x}$ where κ is a hyperparameter defining the variance of $\mathbf{X}(t)$. Observe that $\mathbf{F}$ results in a upper shift matrix that when substituted in equation (6) gives that $\mathrm{d}X^{(i)}(t) = X^{(i+1)}(t)$ for $i = 0, ..., q-1$. By the end $\mathbf{L}$ results in $\mathrm{d}X^{(q)}(t) = \sqrt{\kappa}dW(t)$.

We need to adjust the GP in order to incorporate the information of the differential equation in (1a). To do this we generate ideal measurements of the vector field f, i.e. evaluations of the right hand side of equation (1a) assuming no uncertainty over it. We do it so by defining the following random variable

$$Z(t) \triangleq X^{(1)}(t) - f(X^{(0)}(t), \theta, t) \equiv 0. \tag{9}$$

[1] Also called standard basis, defined as $\mathrm{e_m} = [\delta_{1,\mathrm{m}}, ..., \delta_{\mathrm{D_x},\mathrm{m}}]$.

The product $\mathrm{E}_m \mathbf{X}(t)$ with $\mathrm{E}_m = \mathrm{e}_m \otimes I_{D_x}$ selects the m-th row of equation (5) containing the vector with m-th time-derivative of all the d states. This enables us to recover solely the dimensions of $\mathbf{X}(t)$ we are interested in, in this case the solution of the differential equation $X^0(t)$. We rewrite $Z(t)$ as

$$Z(t) \triangleq \mathrm{E}_1 \mathbf{X}(t) - f(\mathrm{E}_0 \mathbf{X}(t), \theta, t) \equiv 0. \tag{10}$$

To implement algorithmically the Gauss-Markov process we need to discretized the solution $\mathbf{X}(t)$. We define the conditional probability density below for $h = t_n - t_{n-1}$

$$p(\mathbf{X}_{n+1}|\mathbf{X}_n) = \mathcal{N}(\mathbf{X}_{n+1}|\mathbf{A}(h)\boldsymbol{\mu}_n, \mathbf{A}(h)\boldsymbol{\Sigma}_n\mathbf{A}^\top(h) + \mathbf{Q}(h)), \tag{11}$$

where $\mathbf{X}_n$ is the marginalisation of $\mathbf{X}(t)$ at $t = t_n$ for $n = 1, ..., N$ of the GP such that $\mathbf{X}_n \sim \mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$, where $\boldsymbol{\mu}_n = \boldsymbol{\mu}(t_n)$ and $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}(t_n)$. We also define a discretized version of $Z(t)$ at $t = t_n$ as Z_n, which for this case is assumed to be Gaussian distributed with null variance (which can be understood as a Dirac delta centered at zero). Thus we have

$$p(Z_n|\mathbf{X}_n, \theta) = \mathcal{N}(\mathrm{E}_1 \mathbf{X}_n - f(\mathrm{E}_0 \mathbf{X}_t, \theta, t), \mathbf{R}). \tag{12}$$

with $\mathbf{R} = 0$. The method can be extended for values of $\mathbf{R} \neq 0$ when an uncertainty over the function f is assumed, but to compare the current method with the classical method to be defined we keep the zero variance assumption. Assuming $p(\mathbf{X}_0) = \mathcal{N}(\mathbf{X}_0|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ with $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ given, the equations (11) and (12) give the posterior $p(\mathbf{X}_n|\{Z_1, ..., Z_n\}, \theta)$ computed using the Kalman filter algorithm [43].

Next, we define the terms necessary to the implementation of the Kalman filter algorithm:

$$p(\mathbf{X}_n|Z_{1:n}, \theta) = \mathcal{N}(\boldsymbol{\mu}_n^F, \boldsymbol{\Sigma}_n^F), \tag{13a}$$
$$\mathbf{S}_n = \mathbb{V}ar\left[\mathrm{E}_1 \mathbf{X}_n - f(\mathrm{E}_0 \mathbf{X}_n, \theta, t_n)|\{Z_1, ..., Z_{n-1}\}\right] + \mathbf{R}, \tag{13b}$$
$$\mathbf{K}_n = \mathrm{Cov}\left[\mathbf{X}_n, \mathrm{E}_1 \mathbf{X}_n - f(\mathrm{E}_0 \mathbf{X}_n, \theta, t_n)|\{Z_1, ..., Z_{n-1}\}\right] \mathbf{S}_n^{-1}, \tag{13c}$$
$$\hat{Z}_n = \mathbb{E}\left[\mathrm{E}_1 \mathbf{Z}_n - f(\mathrm{E}_0 \mathbf{X}_n, \theta, t_n)|\{Z_1, ..., Z_{n-1}\}\right], \tag{13d}$$
$$\boldsymbol{\mu}_n^F = \boldsymbol{\mu}_n^P + \mathbf{K}_n(Z_n - \hat{Z}_n), \tag{13e}$$
$$\boldsymbol{\Sigma}_n^F = \boldsymbol{\Sigma}_n^P - \mathbf{K}_n\mathbf{S}_n\mathbf{K}_n^\top. \tag{13f}$$

The statistics in equations (13) are taken with respect to $\{Z_1, ..., Z_{n-1}\}$ in the case of the prediction step of the Kalman filter and $\{Z_1, ..., Z_n\}$ in the case of the update step. The superscript P denotes the prediction coming from the estimate mean and variance given incomplete information, i.e., the estimation of $\mathbf{X}_n$ given the information up to $n-1$ as $Z_{1:n-1}$, and F the update of that estimate by its filtering using the most recent information Z_n. The equations (13b) to (13d) are tractable if f is linear, otherwise an approximation required (see the details on Section 2.1.1). The prediction towards $n+1$ is done by using equation (11):

$$\boldsymbol{\mu}_{n+1}^P = \mathbf{A}(h)\boldsymbol{\mu}_n^F, \tag{14a}$$
$$\boldsymbol{\Sigma}_{n+1}^P = \mathbf{A}(h)\boldsymbol{\Sigma}_n^F\mathbf{A}^\top(h) + \mathbf{Q}(h). \tag{14b}$$

We repeat the steps in equations (13) and (14) until all the collection $\{t_n\}_{n=1}^N$ have been evaluated. We obtain the estimation $\boldsymbol{\mu}_n^F$ of $\mathbf{X}_n$ for $n = 1, ..., N$, as well as the variance $\boldsymbol{\Sigma}_n^F$ of this estimation. It is worth to mention that once computed the filtered estimates $\boldsymbol{\mu}_n^F$ and $\boldsymbol{\Sigma}_n^F$, one can use the Rauch–Tung–Striebel (RTS) smoother algorithm to compute other estimates of $\mathbf{X}(t)$ in other time steps not covered by $\{Z_n\}_{n=1}^N$, thus avoiding the evaluation of the function f in equation (1a). Basically, the filter and smooth equations for the Gauss-Markov process estimate and interpolate the solution of the differential equation. We write the terms of the RTS smoother for $0 \leq t_{n'} < t_n \leq t_N$

$$p(\mathbf{X}_{n'}|\{Z_n\}_{n=1}^N) = \mathcal{N}(\mathbf{X}_{n'}|\boldsymbol{\mu}_{n'}^S, \boldsymbol{\Sigma}_{n'}^S), \tag{15a}$$
$$\boldsymbol{\mu}_{n'+1}^P = \mathbf{A}(h')\boldsymbol{\mu}_n^F, \tag{15b}$$
$$\boldsymbol{\Sigma}_{n'+1}^P = \mathbf{A}(h')\boldsymbol{\Sigma}_n^F\mathbf{A}^\top(h') + \mathbf{Q}(h'), \tag{15c}$$
$$\mathbf{G}_{n'} = \boldsymbol{\Sigma}_n^F\mathbf{A}^\top(h)\left[\boldsymbol{\Sigma}_{n'+1}^P\right]^{-1}, \tag{15d}$$
$$\boldsymbol{\mu}_{n'}^S = \boldsymbol{\mu}_n^F + \mathbf{G}_{n'}\left[\boldsymbol{\mu}_{n'+1}^S - \boldsymbol{\mu}_{n'+1}^P\right], \tag{15e}$$
$$\boldsymbol{\Sigma}_{n'}^S = \boldsymbol{\Sigma}_n^F - \mathbf{G}_{n'}\left[\boldsymbol{\Sigma}_{n'+1}^S - \boldsymbol{\Sigma}_{n'+1}^P\right]\mathbf{G}_{n'}^\top. \tag{15f}$$

where we assume $h' = t_{n'} - t_n$, $\boldsymbol{\mu}_N^S = \boldsymbol{\mu}_N^F$ and $\boldsymbol{\Sigma}_N^S = \boldsymbol{\Sigma}_N^F$. The RTS smother algorithm runs backwards in time using the estimates obtained during the time-forward run of the Kalman filter [43]. For the case of estimating the parameters of the model, the smoothing step is not necessary, only the filtered estimates of the states, as it is shown in the next section.

2.1.1 Affine approximation of the vector field

The first order approximation of $f(x(t), \theta, t)$ in the context of the Kalman filter is known in the literature as the extended Kalman filter (EKF, [30]). The EKF consists of approximating the trajectory of the system by its tangent at each point in time, obtaining an affine vector field. This eases the evaluation of equations (13b) to (13d) due to the Gaussian assumption. The approximation consists in truncating the Taylor expansion of $f(x(t), \theta, t)$ around the predicted estimate $\boldsymbol{\mu}_n^P$ of $x(t)$

$$f(x(t), \theta, t) \approx f(\mathrm{E}_0\, \boldsymbol{\mu}_n^P, \theta, t) + \mathbf{J}_f|_{x(t)=\mathrm{E}_0\, \boldsymbol{\mu}_n^P}\, \mathrm{E}_0 \left[\mathbf{X}_n - \boldsymbol{\mu}_n^P\right],$$

where $\mathbf{J}_f$ is the Jacobian matrix whose elements are defined as $[\mathbf{J}]_{ij} = \partial f_i / \partial x_j$, where f_i is the i-th dimension of $f(x(t), \theta, t)$ and x_j the j-th dimension of $x(t)$. The partial derivatives are evaluated substituting $x(t) = \mathrm{E}_0\, \boldsymbol{\mu}_n^P$. It is also possible to consider the filtered estimate $\mathrm{E}_0\, \boldsymbol{\mu}_n^F$ of $x(t)$ [44]. The Kalman filter parameters in equations (13) is given by the equations in (16).

$$\mathbf{H}_n = \mathrm{E}_1 - \mathbf{J}_f|_{x(t)=\mathrm{E}_0\, \boldsymbol{\mu}_n^P}\, \mathrm{E}_0, \tag{16a}$$
$$\mathbf{S}_n = \mathbf{H}_n \boldsymbol{\Sigma}_n^P \mathbf{H}_n^\top + \mathbf{R}, \tag{16b}$$
$$\mathbf{K}_n = \boldsymbol{\Sigma}_n^P \mathbf{H}_n^\top \mathbf{S}_n^{-1}, \tag{16c}$$
$$\hat{Z}_n = \mathrm{E}_1\, \boldsymbol{\mu}_n^P - f(\mathrm{E}_0\, \boldsymbol{\mu}_n^P, \theta, t_n), \tag{16d}$$
$$\boldsymbol{\mu}_n^F = \boldsymbol{\mu}_n^P + \mathbf{K}_n (Z_n - \hat{Z}_n), \tag{16e}$$
$$\boldsymbol{\Sigma}_n^F = \Sigma_n^P - \mathbf{K}_n \mathbf{S}_n \mathbf{K}_n^\top. \tag{16f}$$

2.2 Marginal likelihood via Kalman filtering

The next step is to evaluate the marginal likelihood in equation (4). Following [31], the proposition is to take the posteriors $p(\mathbf{X}_n | Z_{1:n}, \theta)$ computed via the Kalman filter to compute iteratively also via Kalman filter the marginal likelihood. Thus we need to store all the $\boldsymbol{\mu}_n^P$s and $\boldsymbol{\Sigma}_n^P$s for $n = 1, ..., N$. The idea is, once the Kalman filter finishes obtaining the posteriors $p(\mathbf{X}_n | Z_{1:n}, \theta)$, we run another Kalman filter backwards in time to compute the marginal likelihood using the observations Y_n[2]. We assume the following probability model

$$p(\mathbf{X}_n | \mathbf{X}_{n+1}, Z_{1:n}, \theta) = \mathcal{N}(\mathbf{X}_n | \boldsymbol{\xi}_n, \mathbf{P}_n), \tag{17a}$$
$$p(Y_n | \mathbf{X}_n) = \mathcal{N}(Y_n | \mathbf{B}\, \mathrm{E}_0\, \mathbf{X}_n, \sigma_y^2 \mathbf{I}). \tag{17b}$$

The conditional probability backward evolution of the states $p(\mathbf{X}_n | \mathbf{X}_{n+1}, Z_{1:n}, \theta)$ is computed via Gaussian algebra by the inversion of equation (11), giving the parameters $\boldsymbol{\xi}_n = \boldsymbol{\mu}_{n+1}^F + \mathbf{G}_n \left[\mathbf{X}_{n+1} - \boldsymbol{\mu}_{n+1}^P\right]$ and $\mathbf{P}_n$) with $\mathbf{G}_n = \boldsymbol{\Sigma}_n^F \mathbf{A}^\top(h) \left[\boldsymbol{\Sigma}_{n+1}^P\right]^{-1}$ and $\mathbf{P}_n = \boldsymbol{\Sigma}_n^F - \mathbf{G}_n \left[\boldsymbol{\Sigma}_{n+1}^P\right] \mathbf{G}_n^\top$. We define matrix $\mathbf{B}$ whose rows are e_m if the m-th state is observed. The variance of the noise is σ_y^2 and $\mathbf{I}$ the identity matrix. Then we use the prediction error decomposition [45] to compute

$$\Omega(\{Y_n\}_{n=1}^N, \theta) = \prod_{n=N-1}^{1} \mathcal{N}(Y_n | \boldsymbol{\xi}_n^F, \boldsymbol{\Lambda}_n^F), \tag{18a}$$
$$\boldsymbol{\xi}_n^F = \boldsymbol{\xi}_n^P + \mathbf{M}_n \left[Y_n - \mathbf{C} \boldsymbol{\xi}_n^P\right], \tag{18b}$$
$$\boldsymbol{\Lambda}_n^F = \boldsymbol{\Lambda}_n^P - \mathbf{M}_n \mathbf{L}_n \mathbf{M}_n^\top, \tag{18c}$$
$$\mathbf{C} = \mathbf{B}\, \mathrm{E}_0, \tag{18d}$$
$$\mathbf{L}_n = \mathbf{C} \boldsymbol{\Sigma}_n^P \mathbf{C}^\top + \mathbf{R}_y, \tag{18e}$$
$$\mathbf{M}_n = \boldsymbol{\Lambda}_n^P \mathbf{C}_n^\top \mathbf{L}_n^{-1} \tag{18f}$$
$$\boldsymbol{\xi}_n^P = \boldsymbol{\mu}_n^F + \mathbf{G}_n \left[\boldsymbol{\xi}_{n+1}^F - \boldsymbol{\mu}_{n+1}^P\right], \tag{18g}$$
$$\boldsymbol{\Lambda}_n^P = \mathbf{G}_n \boldsymbol{\Lambda}_{n+1} \mathbf{G}_n^\top + \mathbf{P}_n. \tag{18h}$$

The package in Julia language (`Fenrir.jl`) implements equation (18) considering the prior in equation (16) for nonlinear differential equations. The optimisation problem in equation (3) is solved here using the Limited memory Broyden-Fletcher-Goldfarb-Shanno, LBFGS, algorithm [46], which implements a Newton method. The LBFGS takes the gradient of the marginal likelihood with respect to the variables to be optimized, in this case θ, but estimates the Hessian between the iterations in order to improve performance. We summarise the steps of the whole framework in Algorithm 1.

[2] We also refer to [44] for a time-forward implementation.

Algorithm 1 Inference of the parameters of the ASM1 model

Input: $f(x(t), \theta, t)$ (Eq. 1), $\{Y_1, ..., Y_n\}$, ϵ (convergence criterion)
  $\boldsymbol{\mu}_0^F \leftarrow 0$, $\boldsymbol{\Sigma}_0^F \leftarrow \mathbf{I}$, $\theta \leftarrow \theta_\dagger$ (initial condition).
  while $||\partial_\theta \Omega|| \leq \epsilon$ **do** (LBFGS)
    while $n \leq N$ **do**
      Evaluate $p(\mathbf{X}_n | Z_{1:n}, \theta)$ in Eq. 13
      Store $\boldsymbol{\mu}_n^F, \boldsymbol{\Sigma}_n^F$
      $n \leftarrow n + 1$
    end while
    $\boldsymbol{\xi}_N^F \leftarrow \boldsymbol{\mu}_N^F$, $\boldsymbol{\Lambda}_N^F \leftarrow \boldsymbol{\Sigma}_N^F$
    while $n \geq 1$ **do**
      Evaluate equations (18a) to (18h)
      Store $\boldsymbol{\xi}_n^F, \boldsymbol{\Lambda}_n^F$
      $n \leftarrow n - 1$
    end while
    To LBFGS $\Omega(\{Y_n\}_{n=1}^N, \theta)$ and $\partial_\theta \Omega(\{Y_n\}_{n=1}^N, \theta)$.
  end while

We compare the estimates of the parameters using the marginal likelihood of `Fenrir.jl` with a classic approach as the least squares [47] defined as

$$l(\theta, \{y_1, ..., y_N\}) = \sum_{n=1}^{N} (y_n - g(\hat{x}(t_n), \theta, t_n))^2 \tag{19}$$

where y_n is the observation at time t_n and $\hat{x}$ the solution of the differential equation using classic numerical methods. The numerical integration depends on the parameters θ and the initial conditions $x_\dagger$. Thus at each step of the optimisation, we must integrate the differential equation to obtain the value of $l(\theta, \{y_1, ..., y_N\})$.

2.3 Sensitivity analysis

We perform a sensitivity analysis to check how the parameters are prune to variations of the observations [48]. The method chosen is the Morris method [49] which consists in varying one parameter which may cause a variation in the output, which is also measured. We define the elementary effect EE_i as

$$EE_i(\theta_r) = \frac{\hat{x}(t_N, \theta_r + \Delta \mathrm{e}_i, x_\dagger) - \hat{x}(t_N, \theta_r, x_\dagger)}{\Delta},$$

where $\hat{x}$ is the solution of the dynamical system obtained with numerical methods, here evaluated at the last time step t_N using the parameters θ_r. The term e_i is a vector of the canonical base and Δ is a predetermined multiple of $1/(m-1)$, being m the number of elements of the discretisation of each dimension of the parameter space. In the r-th step, a random point θ_r in the parameter space is chosen. Then, for i-th parameter the elementary effect is calculated. By the end of a suitably chosen of iterations R, the following metrics are obtained

$$\mu_M = \frac{1}{R} \sum_{r=1}^{R} |EE_i(\theta_r)|,$$

$$\sigma_M = \sqrt{\frac{1}{R} \sum_{r=1}^{R} \left(EE_i(\theta_r) - \frac{1}{R} \sum_{r=1}^{R} |EE_i(\theta_r)| \right)^2}.$$

μ_M indicates the influence of the i-th parameter on the output, and σ_M how non-linear are the effects of the i-th parameter on the output. If μ_M is small it suggests that the effects of the parameters are weak with the output.

3 Application to a biological reactor model

To demonstrate the methodology exposed in Section 2, we explore the widely used Activated Sludge Model No. 1 (ASM1, [7]). In this work, the ASM1 models a bioreactor consisting of the aerated continuous stirred tank reactor with perfect mixing shown in Figure 1, where the influent load, x^{IN}, is given in terms of the thirteen ASM1 state variables and the flow-rate, Q, that equals the effluent one, the set of sensors in the effluent stream gives the the measured variables: dissolved oxygen, nitrate and nitrite nitrogen, ammonia nitrogen and alkalinity, that is S_O, S_{NO}, S_{NH} and S_{ALK}, respectively, and $K_L a$ gives an adjustable variable considered representative of the diffused aeration in the bioreactor.

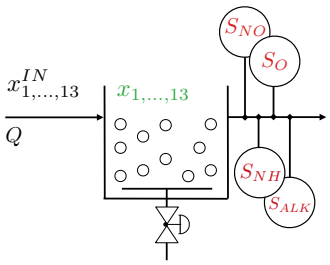


Figure 1: Bioreactor schematic.

ASM1 is a state-of-the-art model commonly used for representing the carbon and nitrogen removal, oxygen consumption and sludge production. The growth of the microbiological population responsible for the processes occurring inside the reactor is described by Monod kinetics [50]. The mass balances for the state variables lead to a system of ordinary differential equations of the thirteen state variables in Table 1 that comprises stoichiometric and kinetic parameters in Table 2. As an example of the nonlinearity in the ASM1, the differential equation for the readily biodegradable organic matter, S_S in equation (20), is provided.

$$\frac{\mathrm{d}S_S}{\mathrm{d}t} = \frac{Q}{V}\left[S_S^{IN} - S_S\right] + R_{S_S} \tag{20a}$$

$$R_{S_S} = -\frac{\mu_H}{Y_H}\frac{S_S}{K_S + S_S}\left[\frac{S_O}{K_{OH} + S_O} + \eta_g \frac{K_{OH}}{K_{OH} + S_O}\frac{S_{NO}}{K_{NO} + S_{NO}}\right] X_{BH} + k_h \frac{X_S}{K_X X_{BH} + X_S}\left[\frac{S_O}{K_{OH} + S_O} + \eta_h \frac{K_{OH}}{K_{OH} + S_O}\frac{S_{NO}}{K_{NO} + S_{NO}}\right] X_{BH}. \tag{20b}$$

Symbol	Description	Units
S_I	Soluble inert organic matter	$g\,\mathrm{COD} \cdot m^{-3}$
S_S	Readily biodegradable substrate	$g\,\mathrm{COD} \cdot m^{-3}$
X_I	Particulate inert organic matter	$g\,\mathrm{COD} \cdot m^{-3}$
X_S	Slowly biodegradable substrate	$g\,\mathrm{COD} \cdot m^{-3}$
X_{BH}	Active heterotrophic biomass	$g\,\mathrm{COD} \cdot m^{-3}$
X_{BA}	Active autotrophic biomass	$g\,\mathrm{COD} \cdot m^{-3}$
X_P	Particulate products arising from biomass decay	$g\,\mathrm{COD} \cdot m^{-3}$
S_O	Dissolved oxygen	$g\,O_2 \cdot m^{-3}$
S_{NO}	Nitrate and nitrite nitrogen	$g\,N \cdot m^{-3}$
S_{NH}	$NH_4^+ + NH_3$ nitrogen	$g\,N \cdot m^{-3}$
S_{ND}	Soluble biodegradable organic nitrogen	$g\,N \cdot m^{-3}$
X_{ND}	Particulate biodegradable organic nitrogen	$g\,N \cdot m^{-3}$
S_{ALK}	Alkalinity	mol HCO_3^- $\cdot m^{-3}$

Table 1: ASM1 - State variables

The right-hand side of equation (20a) represents the mass balance and the R_{S_S} term, fully given in equation (20b), is the kinetic of the reaction producing (or consuming) S_S. S_S^{IN} is the inflow concentration of biodegradable substrate. We refer to the work in [51] for the full description of the system of equations. The biological reactor is simulated considering the settings established by the Benchmark Simulation Model No. 1 (BSM1, [52]), a simulation environment defining a plant layout, the ASM1 as simulation model, the influent load concentrations and flow-rate, as well as the test procedures and evaluation criteria. We consider the steady-states of the BSM1 computed in [53] as initial conditions to solve the differential equations of the ASM1.

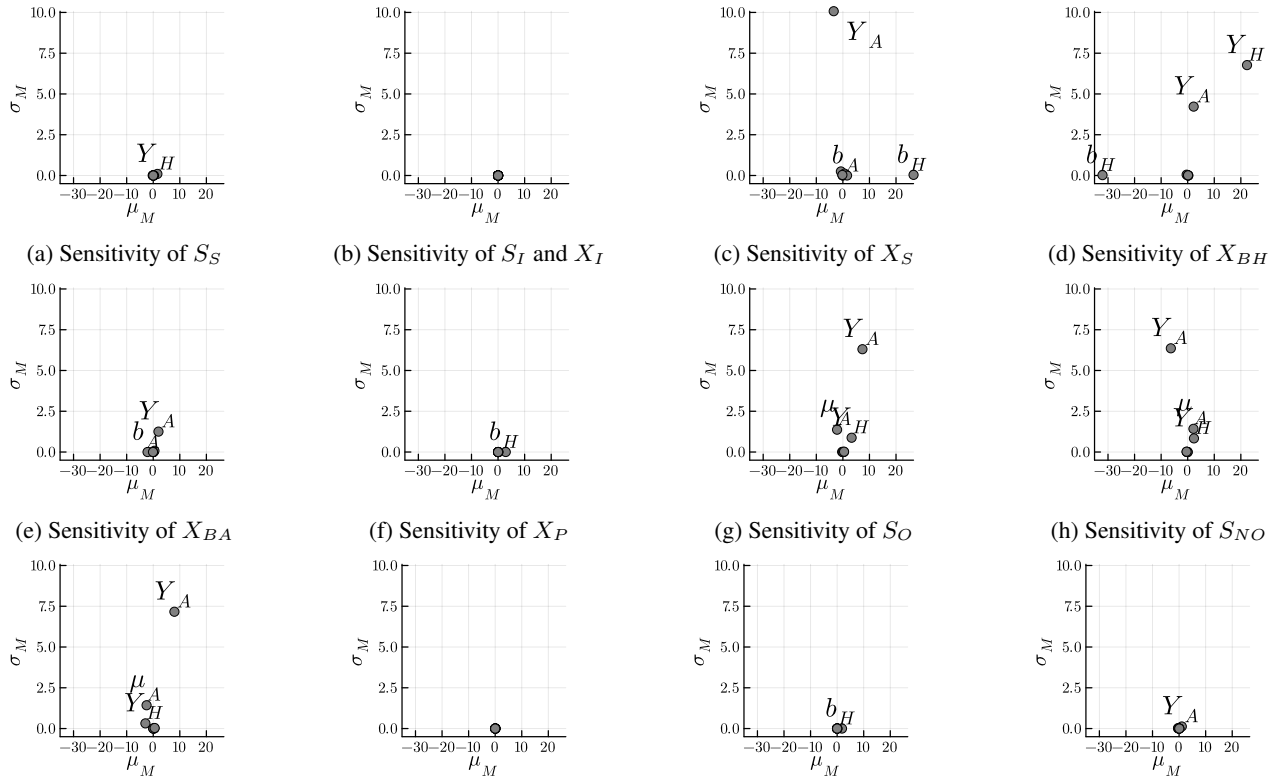
We consider that, beyond the measurements of nitrate and nitrite nitrogen, S_{NO}, and dissolved oxygen, S_O, given by the sensors setting in the simulation platform [53] and commonly used in modern WWTPs, it is also possible to measure ammonia nitrogen, S_{NH}, [54] and alkalinity, S_{ALK}, [55] in real-time. The number of sensors is realistic and low avoiding to increase operational costs in a real plat setting. The measurements occur ten times a day and are evenly spaced.

Symbol	Description	Value	Units
Y_A	Autotrophic yield	0.24	g X_{BA} COD formed $\cdot$ (g N oxidised)$^{-1}$
Y_H	Heterotrophic yield	0.67	g X_{BH} COD formed $\cdot$ (g COD utilized)$^{-1}$
f_P	Fraction of biomass to particulate products	0.08	g X_P COD formed $\cdot$ (g N decayed)$^{-1}$
i_{XB}	Fraction nitrogen in biomass	0.08	g $N \cdot$ (g COD)$^{-1}$ in biomass
i_{XP}	Fraction nitrogen in particulate products	0.06	g $N \cdot$ (g COD)$^{-1}$ in X_P
μ_H	Maximum heterotrophic growth rate	4.00	d^{-1}
K_S	Half-saturation (heterotrophic growth)	10.0	g COD m^{-3}
K_{OH}	Half-saturation (heterotrophic oxygen)	0.20	g O_2 m^{-3}
K_{NO}	Half-saturation (nitrate)	0.50	g NO_3-N m^{-3}
b_H	Heterotrophic decay rate	0.30	d^{-1}
ν_g	Anoxic growth rate correction factor	0.80	-
ν_h	Anoxic hydrolysis rate correction factor	0.80	-
k_h	Maximum specific hydrolysis rate	3.00	g X_S (g X_{BH} COD d)$^{-1}$
K_X	Half-saturation (hydrolysis)	0.10	g X_S (g X_{BH} COD)$^{-1}$
μ_A	Maximum autotrophic growth rate	0.50	d^{-1}
K_{NH}	Half-saturation (autotrophic growth)	1.00	g NH_4-N m^{-3}
b_A	Autotrophic decay rate	0.05	d^{-1}
K_{OA}	Half-saturation (autotrophic oxygen)	0.40	g O_2 m^{-3}
k_a	Ammonification rate	0.05	g X_S (g COD d)$^{-1}$

Table 2: ASM1 - Parameters

4 Results and discussion

The computation of the solution of the system of differential equations for both the sensitivity analysis and the least squares method is done using the package `DifferentialEquations.jl` implemented in Julia language [28, 56].



(a) Sensitivity of S_S (b) Sensitivity of S_I and X_I (c) Sensitivity of X_S (d) Sensitivity of X_{BH}

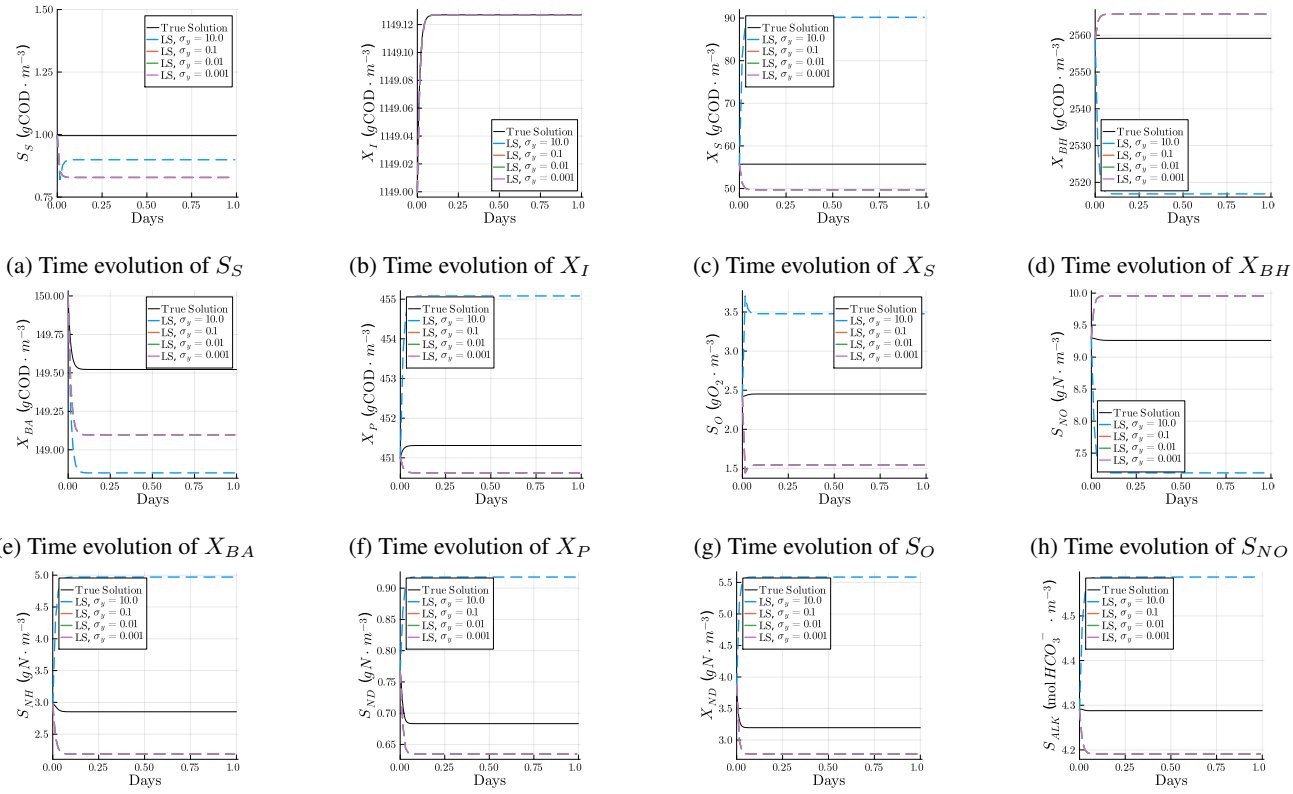
(e) Sensitivity of X_{BA} (f) Sensitivity of X_P (g) Sensitivity of S_O (h) Sensitivity of S_{NO}

(i) Sensitivity of S_{NH} (j) Sensitivity of S_{ND} (k) Sensitivity of X_{ND} (l) Sensitivity of S_{ALK}

Figure 2: Results of the Morris analysis, being μ_M and σ_M the metrics calculated for each output of the ASM1 model. All the parameters of the system are pictured in gray dots but only the ones obeying $\sqrt{\mu_M^2 + \sigma_M^2} \geq 1$ where labeled.

Figure 2 displays the results of the sensitivity analysis conducted using the Morris method. The findings align with those reported in [57] for Y_H, Y_A, μ_A and b_A. In the same work, the sensitivity of these parameters was explored concerning various

outputs derived as functions of state variables, specifically the chemical oxygen demand, total suspended solids, ammonia and pH. In contrast, our approach directly observes the system states. The parameters K_{NH} and K_{OA} previously identified as significant in [57] have their sensitivity reduced when considering the sensitivity of b_H and μ_A on the states directly measured. Consequently, we decided to comprehensively identify the set of parameters, $\theta = Y_H$, Y_A, μ_A, b_H, μ_A, K_{NH} and K_{OA}, simultaneously and evaluate the performance of parameter estimation across them. The plots in Figure 2 indicate that b_A and b_H exert a robust linear influence on model outputs due to the values of μ_M greater that 1 while the other parameters exhibit strong non-linear effects owing to the values of σ_M greater than 1. Additionally, we anticipate that the estimation, overall, will perform more effectively for parameters situated away from the origin in the sensitivity plots in Figure 2. Specifically, for the measured states S_{NO}, S_O,S_{NH} and S_{ALK} we anticipate suboptimal performance for parameters μ_H, b_h, K_{NH}, b_A and K_{OA} given that their sensitivity results satisfy $\sqrt{\mu_M^2 + \sigma_M^2} \leq 1$ suggesting weak sensitivity of the outputs on these parameters. This is attributed to the reliance on gradient methods, as employed in this work, where weak dependencies of outputs on parameters result in small gradients concerning the parameters during the evaluation of the optimisation problem's objective function.



(a) Time evolution of S_S (b) Time evolution of X_I (c) Time evolution of X_S (d) Time evolution of X_{BH}

(e) Time evolution of X_{BA} (f) Time evolution of X_P (g) Time evolution of S_O (h) Time evolution of S_{NO}

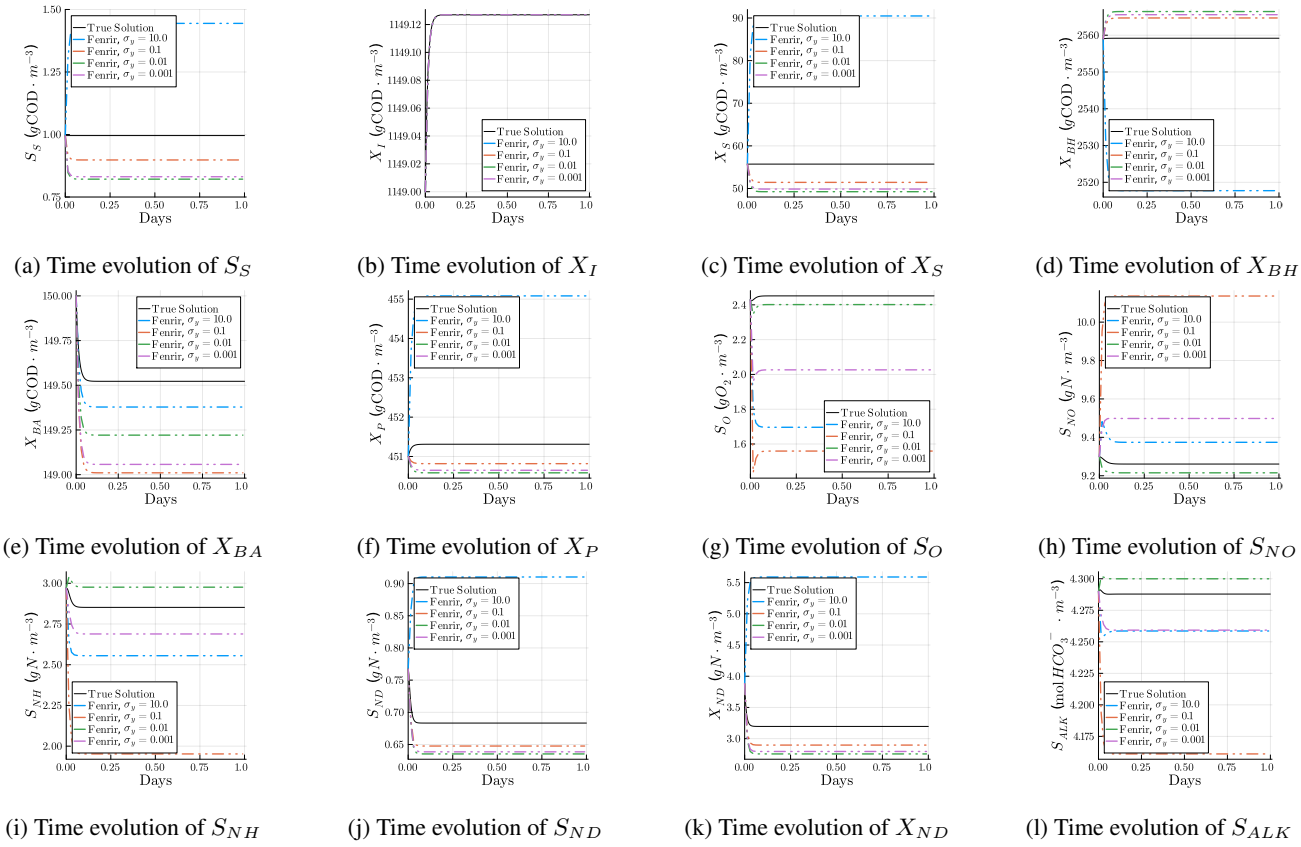
(i) Time evolution of S_{NH} (j) Time evolution of S_{ND} (k) Time evolution of X_{ND} (l) Time evolution of S_{ALK}

Figure 3: Simulation of the system with the set of parameters obtained from the optimisation of the least squares (LS) for different noise standard deviations σ_y.

To compare parameter estimation results, we employ the least squares (LS) method outlined in equation (19). Subsequently, we optimize the objective function with respect to the ASM1 parameters θ using the LBFGS algorithm, setting boundaries of the optimisation problem based on the range of values established by the literature review in [57]. The measurement of the states S_O, S_{NO}, S_{NH} and S_{ALK} are subject to additive Gaussian noise with standard deviation σ_y. The assumed range of σ_y in the simulated data encompasses values wider than the precision of the available sensors [53]. Four scenarios were tested by varying the standard deviation of the noise in the observations, as detailed in the first column of Table 3. The absolute errors $|\theta^\star - \theta|$ are presented in Table 3, where $\theta^\star$ represents the result of the optimisation and θ denotes their true values shown in Table 2. Noticeably, LS was unable to reduce the error for σ_y values below 10 leading to the same parameter estimates on those scenarios.

σ_y	Y_A	Y_H	μ_H	b_H	μ_A	K_{NH}	b_A	K_{OA}
10	0.0400	0.0500	2.0000	1.3000	0.3000	**0.0000**	0.1000	1.6000
0.1	0.1679	0.0495	**0.0200**	0.2345	0.2920	0.2475	**0.0010**	0.0160
0.01	0.1679	0.0495	**0.0200**	**0.2345**	0.2920	0.2475	**0.0010**	0.0160
0.001	0.1679	0.0495	**0.0200**	0.2345	**0.2920**	0.2475	**0.0010**	0.0160

Table 3: Parameter absolute error for the LS objective function. The parameters for which the method performed better than the marginal likelihood method were highlighted.



(a) Time evolution of S_S (b) Time evolution of X_I (c) Time evolution of X_S (d) Time evolution of X_{BH}

(e) Time evolution of X_{BA} (f) Time evolution of X_P (g) Time evolution of S_O (h) Time evolution of S_{NO}

(i) Time evolution of S_{NH} (j) Time evolution of S_{ND} (k) Time evolution of X_{ND} (l) Time evolution of S_{ALK}

Figure 4: Simulation of the system with the set of parameters obtained from the optimisation of the marginal likelihood (Fenrir) for different noise standard deviations σ_y.

The marginal likelihood evaluation in this study is conducted using the Fenrir package developed in the Julia programming language (`Fenrir.jl`, [31]). It generates the negative log marginal likelihood as the objective function utilised in the optimisation problem. The optimisation is carried out using the same LBFGS settings as employed in the least squares method. Similarly to Table 3, Table 4 presents the absolute error $|\theta^\star - \theta|$ where $\theta^\star$ represents the result of the optimisation using the negative log marginal likelihood as the objective function and θ denotes their true values shown in Table 2.

Figures 3 and Figure 4 showcase the time evolution of each state obtained by simulating the ASM1 with the original and the optimised set of parameters for each scenario of noisy measurements. One of the states, the soluble inert organic matter S_I, is excluded as it exhibits no dynamic changes over time. This is supported by the null values in the sensitivity analysis of S_I and X_I in Figure 2b. Overall, the marginal likelihood objective function leads to parameter estimations that are closer to their true values in scenarios with higher noise. In contrast, the least squares method was unable to achieve parameter estimates closer to their true values for the values of σ_y below 10. Consequently, the solutions plotted in Figure 3 display only two lines: One for σ_y equals 10 in blue , and all the others superposed.

σ_y	Y_A	Y_H	μ_H	b_H	μ_A	K_{NH}	b_A	K_{OA}
10	**0.0039**	**0.0000**	**0.1659**	**1.3000**	**0.0975**	0.1674	**0.0999**	**0.0784**
0.1	**0.1404**	**0.0000**	0.1307	**0.1736**	**0.1930**	**0.2400**	0.0954	**0.0002**
0.01	**0.0636**	**0.0446**	0.0535	0.2500	**0.1590**	**0.2234**	0.0527	**0.0066**
0.001	**0.1505**	**0.0486**	0.0248	**0.2266**	0.3000	**0.2462**	0.0280	**0.0015**

Table 4: Parameter absolute error for the marginal likelihood method. The parameters for which the method performed better than the LS method were highlighted.

5 Conclusion and future works

We conclude noticing that the main disadvantage of the marginal likelihood is the computational cost. Both implementations, `Fenrir.jl` and the LS, make use of the automatic differentiation available in Julia, but only the LS experienced sped ups in terms of wall-clock time to convergence of the LBFGS algorithm at the cost of ignoring the uncertainty over the model. Hence, it may be advisable to consider classical methods, particularly in scenarios characterised by low noise. This recommendation

is grounded in the observation that classical methods can yield comparable errors with significantly less computational burden. The optimisations were conducted in a user-end 11th Gen Intel(R) Core(TM) i7-11800H 2.30GHz with 16.0 GB of RAM. The wall-clock time was 8600 seconds for the Fenrir loss while 200 seconds for the least squares. There is space for optimizations in the computation time and explorations of scenarios not considered in this work as the uncertainty of the solution for $\mathbf{R} \neq 0$. We judged the latter to need a deeper discussion with the convergence behavior in sight once this analysis is not known by the authors.

References

[1] United Nations, Department of Economic and Social Affairs - Sustainable Development. "Transforming our world: the 2030 Agenda for Sustainable Development", 2015.

[2] H. T. Tran, G. Lesage, C. Lin, T. B. Nguyen, X.-T. Bui, M. K. Nguyen, D. H. Nguyen, H. G. Hoang and D. D. Nguyen. *Activated sludge processes and recent advances*, pp. 49–79. Elsevier, 2022.

[3] D. Jenkins and J. Wanner. *Activated Sludge - 100 Years and Counting*. IWA Publishing, 2014.

[4] M. Faisal, K. M. Muttaqi, D. Sutanto, A. Q. Al-Shetwi, P. J. Ker and M. Hannan. "Control technologies of wastewater treatment plants: The state-of-the-art, current challenges, and future directions". *Renewable and Sustainable Energy Reviews*, vol. 181, pp. 113324, 2023.

[5] R. Alvarado. "Computer Simulations as Scientific Instruments". *Foundations of Science*, vol. 27, pp. 1183–1205, 2022.

[6] L. M. Ruiz, J. I. Pérez and M. A. Gómez. "Practical review of modelling and simulation applications at full-scale wastewater treatment plants". *Journal of Water Process Engineering*, vol. 56, pp. 104477, 2023.

[7] M. Henze, C. P. L. J. Grady, V. Gujer, G. v. R. Marais and T. Matsuo. "Activated Sludge Model No. 1. IAWPRC Scientific and Technical Reports No. 1." Technical report, IWA Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment, 1987.

[8] M. Henze, W. Gujer, T. Mino and M. Van Loosedrecht. "Activated sludge models ASM1, ASM2, ASM2d and ASM3". Technical report, IWA Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment, 2006.

[9] M. Maktabifard, K. Blomberg, E. Zaborowska, A. Mikola and J. Mąkinia. "Model-based identification of the dominant N2O emission pathway in a full-scale activated sludge system". *Journal of Cleaner Production*, vol. 336, pp. 130347, 2022.

[10] H. Hauduc, L. Rieger, A. Oehmen, M. van Loosdrecht, Y. Comeau, A. Heduit, P. Vanrolleghem and S. Gillot. "Critical review of activated sludge modeling: State of process knowledge, modeling concepts, and limitations". *Biotechnology and Bioengineering*, vol. 110, no. 1, pp. 24–46, 2013.

[11] H. Hauduc, S. Gillot, L. Rieger, T. Ohtsuki, A. Shaw, I. Takács and S. Winkler. "Activated sludge modelling in practice: an international survey". *Water Science and Technology*, vol. 60, no. 8, pp. 1943–1951, 2009.

[12] L. Rieger, S. Gillot, G. Langergraber, T. Ohtsuki, A. Shaw, I. Takács and S. Winkler. *Guidelines for Using Activated Sludge Models*. IWA Publishing, 2012.

[13] A. Holmberg. "On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities". *Mathematical Biosciences*, vol. 62, no. 1, pp. 23–43, nov 1982.

[14] P. Reichert. "User manual of AQUASIM 2.0 for the identification and simulation of aquatic systems". *Swiss Federal Institute for Environmental Science and Technology, Dubendorf, Switzerland*, 1998.

[15] K. Gadkar, F. Doyle, T. Crowley and J. Varner. "Cybernetic Model Predictive Control of a Continuous Bioreactor with Cell Recycle". *Biotechnology Progress*, vol. 19, no. 5, pp. 1487–1497, oct 2003.

[16] G. Antonopoulou, M. Alexandropoulou, C. Lytras and G. Lyberatos. "Modeling of Anaerobic Digestion of Food Industry Wastes in Different Bioreactor Types". *Waste and Biomass Valorization*, vol. 6, no. 3, pp. 335–341, mar 2015.

[17] S. Sharifi, S. Murthy, I. Takács and A. Massoudieh. "Probabilistic parameter estimation of activated sludge processes using Markov Chain Monte Carlo". *Water Research*, vol. 50, no. 1 March 2014, pp. 254–266, 2014.

[18] Z. J. Zonta, X. Flotats and A. Magrí. "Estimation of parameter uncertainty for an activated sludge model using Bayesian inference: a comparison with the frequentist method". *Environmental Technology*, vol. 35, no. 1, pp. 1618–1629, 2014.

[19] P. A. Stentoft, T. Munk-Nielsen, L. Vezzaro, H. Madsen, J. K. Møller and P. S. Mikkelsen. "Towards model predictive control: Online predictions of ammonium and nitrate removal by using a stochastic ASM". *Water Science and Technology*, vol. 79, no. 1, pp. 51 – 62, 2019.

[20] A. Munack. "Optimal Feeding Strategy for Identification of Monod-Type Models by Fed-Batch Experiments". In *Computer Applications in Fermentation Technology: Modelling and Control of Biotechnological Processes*, pp. 195–204. Springer Netherlands, 1989.

[21] D. Dochain and P. Vanrolleghem. *Dynamical Modelling and Estimation in Wastewater Treatment Processes*. IWA Publishing, 2005.

[22] A. Cosenza, G. Mannina, P. A. Vanrolleghem and M. B. Neumann. "Global sensitivity analysis in wastewater applications: A comprehensive comparison of different methods". *Environmental Modelling & Software*, vol. 49, pp. 40–52, 2013.

[23] W. Gujer. *Systems Analysis for Water Technology*. Springer Berlin, Heidelberg, 2008.

[24] A. Zhu, J. Guo, B.-J. Ni, S. Wang, Q. Yang and Y. Peng. "A Novel Protocol for Model Calibration in Biological Wastewater Treatment". *Scientific Reports*, vol. 5, pp. 8493, 2015.

[25] A. Khajavian, A. Pourmohamadi, Y. Khatibi and S. Nazif. "Static calibration of wastewater treatment plant models: Investigating calibration processes and objective functions". *Journal of Water Process Engineering*, vol. 54, pp. 104016, 2023.

[26] P. Hennig, M. Osborne and H. Kersting. *Probabilistic Numerics*. Cambridge University Press, 2022.

[27] G. Mannina, A. Cosenza, G. Viviani and G. A. Ekama. "Sensitivity and uncertainty analysis of an integrated ASM2d MBR model for wastewater treatment". *Chemical Engineering Journal*, vol. 351, pp. 579–588, 2018.

[28] J. Bezanson, A. Edelman, S. Karpinski and V. B. Shah. "Julia: A fresh approach to numerical computing". *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.

[29] J. Revels, M. Lubin and T. Papamarkou. "Forward-mode automatic differentiation in Julia". *arXiv preprint arXiv:1607.07892*, 2016.

[30] F. Tronarp, H. Kersting, S. Särkkä and P. Hennig. "Probabilistic Solutions To Ordinary Differential Equations As Non-Linear Bayesian Filtering: A New Perspective", 2019.

[31] F. Tronarp, N. Bosch and P. Hennig. "Fenrir: Physics-Enhanced Regression for Initial Value Problems", 2023.

[32] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2005.

[33] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[34] P. de Laplace. *Essai philosophique sur les probabilités*. Courcier, 1814.

[35] E. Jaynes and G. Bretthorst. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[36] D. J. C. MacKay. "Bayesian Interpolation". *Neural Computation*, vol. 4, no. 3, pp. 415–447, May 1992.

[37] M. Heinonen, C. Yildiz, H. Mannerström, J. Intosalmi and H. Lähdesmäki. "Learning unknown ODE models with Gaussian processes". In *International conference on machine learning*, pp. 1959–1968. PMLR, 2018.

[38] P. Wenk, G. Abbati, M. A. Osborne, B. Schölkopf, A. Krause and S. Bauer. "ODIN: ODE-Informed Regression for Parameter and State Inference in Time-Continuous Dynamical Systems". 2019.

[39] A. Solin. "Hilbert space methods in infinite-dimensional Kalman filtering". Master's thesis, Aalto University, School of Science, 2012.

[40] S. Sarkka and J. Hartikainen. "Infinite-dimensional Kalman filtering approach to spatio-temporal Gaussian process regression". In *Artificial Intelligence and Statistics*, pp. 993–1001. PMLR, 2012.

[41] B. Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[42] S. Särkkä *et al.*. *Recursive Bayesian inference on stochastic differential equations*. Helsinki University of Technology, 2006.

[43] S. Särkkä and L. Svensson. *Bayesian filtering and smoothing*, volume 17. Cambridge university press, 2023.

[44] M. Wu and M. Lysy. "Data-Adaptive Probabilistic Likelihood Approximation for Ordinary Differential Equations", 2023.

[45] F. Schweppe. "Evaluation of likelihood functions for Gaussian signals". *IEEE transactions on Information Theory*, vol. 11, no. 1, pp. 61–70, 1965.

[46] D. C. Liu and J. Nocedal. "On the limited memory BFGS method for large scale optimization". *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989.

[47] Z. Li, M. R. Osborne and T. Prvan. "Parameter estimation of ordinary differential equations". *IMA Journal of Numerical Analysis*, vol. 25, no. 2, pp. 264–285, 2005.

[48] W. Coudron, A. Gobin, C. Boeckaert, T. De Cuypere, P. Lootens, S. Pollet, K. Verheyen, P. De Frenne and T. De Swaef. "Data collection design for calibration of crop models using practical identifiability analysis". *Computers and Electronics in Agriculture*, vol. 190, pp. 106457, 2021.

[49] B. Iooss and P. Lemaître. "A review on global sensitivity analysis methods". *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, pp. 101–122, 2015.

[50] J. Monod. "The growth of bacterial cultures". *Annual review of microbiology*, vol. 3, no. 1, pp. 371–394, 1949.

[51] O. B. Neto, M. Mulas and F. Corona. "About the classical and structural controllability and observability of a common class of activated sludge plants". *Journal of Process Control*, vol. 111, pp. 8–26, 2022.

[52] J. Alex, L. Benedetti, J. Copp, K. Gernaey, U. Jeppsson, I. Nopens, M. Pons, L. Rieger, C. Rosen, J. Steyer *et al.*. "Benchmark simulation model no. 1 (BSM1)". *Report by the IWA Taskgroup on benchmarking of control strategies for WWTPs*, vol. 1, 2008.

[53] K. V. Gernaey, U. Jeppsson, P. A. Vanrolleghem and J. B. Copp. *Benchmarking of Control Strategies for Wastewater Treatment Plants*. IWA Publishing, 2014.

[54] X. Wang, Y. Fan, Y. Huang, J. Ling, A. Klimowicz, G. Pagano and B. Li. "Solving Sensor Reading Drifting Using Denoising Data Processing Algorithm (DDPA) for Long-Term Continuous and Accurate Monitoring of Ammonium in Wastewater". *ACS ES&T Water*, vol. 1, no. 3, pp. 530–541, 2020.

[55] P. A. Vanrolleghem and D. S. Lee. "On-line monitoring equipment for wastewater treatment processes: state of the art". *Water Science and Technology*, vol. 47, no. 2, pp. 1–34, 2003.

[56] C. Rackauckas and Q. Nie. "Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia". *Journal of Open Research Software*, vol. 5, no. 1, pp. 15, 2017.

[57] A. Martins Jr, M. Silva and A. Benetti. "Evaluation and optimization of ASM1 parameters using large-scale WWTP monitoring data from a subtropical climate region in Brazil". *Water Practice & Technology*, vol. 17, no. 1, pp. 268–284, 2022.