

An adaptive real-time grey-box model for advanced control and operations in WRRFs

Cheng Yang ^a, Peter Seiler ^b, Evangelia Belia ^c and Glen T. Daigger ^{a,*}

^a Civil and Environmental Engineering, University of Michigan, 2350 Hayward St, G.G. Brown Building, Ann Arbor, MI 48109, USA

^b Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, EECS Building, Ann Arbor, MI 48109, USA

^c Primodal Inc., 145 Rue Aberdeen, Quebec City, Quebec, Canada

*Corresponding author. E-mail: gdaigger@umich.edu

 CY, 0000-0002-9352-0675; PS, 0000-0003-3423-1109; EB, 0000-0002-8741-4403; GTD, 0000-0003-3857-8345

ABSTRACT

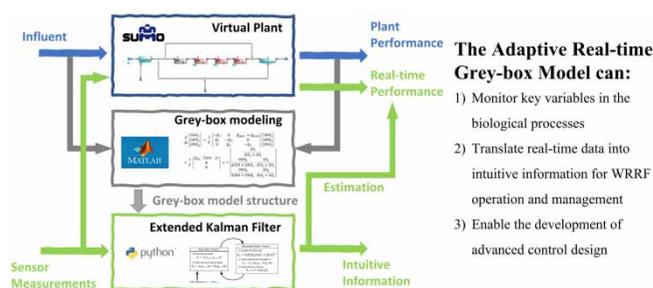
Grey-box models, which combine the explanatory power of first-principle models with the ability to detect subtle patterns from data, are gaining increasing attention in wastewater sectors. Intuitive, simple structured but fit-for-purpose grey-box models that capture time-varying dynamics by adaptively estimating parameters are desired for process optimization and control. As an example, this study presents the identification of such a grey-box model structure and its further use by an extended Kalman filter (EKF), for the estimation of the nitrification capacity and ammonia concentrations of a typical Modified Ludzack-Ettinger (MLE) process. The EKF was implemented and evaluated in real time by interfacing Python with SUMO (Dynamita™), a widely used commercial process simulator. The EKF was able to accurately estimate the ammonia concentrations in multiple tanks when given only the concentration in one of them. In addition, the nitrification capacity of the system could be tracked in real time by the EKF, which provides intuitive information for facility managers and operators to monitor and operate the system. Finally, the realization of EKF is critical to the development of future advance control, for instance, model predictive control.

Key words: activated sludge, extended Kalman filter, grey-box model, parameter estimation, SUMO

HIGHLIGHTS

- The development of an adaptive real-time grey-box model with intuitive information is presented.
- The need of model adaptivity was identified and fulfilled by the extended Kalman filter.
- The extracted real-time intuitive information will help WRRFs staff in operations and management.
- Model structure simplicity and development pathway encourages applications for other processes.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Wastewater treatment plants are being repurposed to water resource recovery facilities (WRRFs), addressing nutrient recovery and energy neutrality to deal with stricter emission regulation, increasing water scarcity and rapid urbanization (Therrien *et al.* 2020). The practice of WRRF design, operation, and control needs adequate models that encode

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

knowledge and information relevant to its designed objectives (Regmi *et al.* 2019; Therrien *et al.* 2020). In recent decades, the wastewater research community has established generally acknowledged mechanistic and phenomenological models (white-box models) such as the Activated Sludge Model (ASM) family (Rieger *et al.* 2012; Henze *et al.* 2015), Biofilm Models (Rittmann *et al.* 2018), and Anaerobic Digestion Model No. 1 (ADM1) (Batstone *et al.* 2002). Meanwhile, purely data-driven models (black-box models), mainly focusing on prediction power rather than interpretability, are becoming more widely developed and used thanks to advances in big data analytics and artificial intelligence. A comprehensive overview by Haimi *et al.* (2013) surveyed and reviewed applications of black-box models in biological processes. While black-box models have often been criticized for their lack of transparency, a mixture of fundamental white-box structure and empirical black-box components has emerged and gained increasing attention, namely the grey-box models. These models integrate components of first-principle models with the data-driven schemes of black-box models for improved ability to estimate unmeasurable variables and kinetics, capture unmodelled dynamics, and predict system performance (Haimi *et al.* 2013; Newhart *et al.* 2019; Therrien *et al.* 2020). Typical integrations include: (1) using black-box models to reconstruct needed but not available information for white-box models, for instance, influent data (Zahedi *et al.* 2005), kinetic parameters (Zahedi *et al.* 2005; Shiva Kumar & Venkateswarlu 2012); (2) using only partial white-box model structure, relying on data to complete the model (similar to black-box models) (Nair *et al.* 2019; Stentoft *et al.* 2019).

Models are developed for various purposes, including but not limited to design of new plants, upgrade and optimization of existing ones, prediction of future behaviour, education, and process control. The intended use determines the modelling approach and its associated model complexity. For instance, white-box models embedded with as many biochemical reactions as possible are promising teaching tools to educate operators and young water professionals, but not appropriate for real-time control design (e.g., model predictive control) because of their computational intensiveness and unmeasurable parameters. Similarly, black-box models yield good prediction power but are case specific and depend heavily on data availability and quality, in addition to significant effort required for model selection and training. A simply-structured but fit-for-purpose dynamic grey-box model, with data-driven techniques to complete the model, is a candidate solution for real-time prediction and control. One common downstream application is advanced control design based on the developed models. For instance, Stare *et al.* (2006) developed a reduced non-linear nitrification model by modifying expressions in ASM1 for an attached growth pilot plant, with model parameters estimated from real measurements. Different control strategies were then compared based on the identified model. Another common application is soft sensors (also known as state estimators and observers), a virtual asset acting like sensors for prediction and control. Stentoft *et al.* (2019) rewrote ASM1 into a simpler stochastic grey-box model and developed an online soft sensor to predict ammonium and nitrate removal in a small recirculating WRRF facility. Nair *et al.* (2019) developed a soft sensor based on a grey-box model to estimate volatile fatty acids, phosphate, ammonia and nitrate concentrations based on inputs from inexpensive sensors such as pH and dissolved oxygen. Richer real-time information extracted by soft sensors can improve the efficiency of control and operation.

One potential drawback of a simple grey-box model structure is the need to adaptively update the estimated parameters. Grey-box model identification (the procedure of estimating parameter values from data) often occurs offline, with a limited series of input and output data collected in advance. Estimated parameter values often vary substantially over time in WRRFs, due to slow changes that shift system equilibriums over weeks or months (e.g., biomass, temperature, and wastewater compositions). Adaptive approaches can be used to overcome this issue. One is the Moving Horizon Estimator (Robertson *et al.* 1996; Rao *et al.* 2003), where the states and parameters are re-estimated periodically after collecting sufficient new measurements. Another is the extended Kalman filter (EKF) (Welch & Bishop 2006), where parameter and state estimates are updated efficiently and recursively with new measurements in a continuous mode. Busch *et al.* (2013) compared and demonstrated the effectiveness of both approaches in estimating unmeasurable states in the Benchmark Simulation Model No. 1 (BSM1) (Alex *et al.* 2008). In this paper, the EKF method was selected. Once the grey-box model is equipped with an adaptive scheme, the simplicity of its model structure becomes an advantage for its wide compatibility for other processes, especially for control design of newly emerging biological wastewater treatment processes whose mechanisms are not fully understood.

It has become a usual practice to develop a white-box model of treatment facilities, often based for instance on the IWA ASMs and ADM1, either as part of the initial design or major upgrade of a facility. If available, the white-box model could be used to simulate a wide range of facility loading and operating conditions, from start-up to design influent flows and constituent loadings, plant operating conditions, and seasonal factors such as wastewater temperature variations. In this

paper, such a white-box model was used as a digital twin to develop and evaluate grey-box models. The aim of this paper was to develop an adaptive real-time dynamic model (also known as soft sensors and observers) for advanced control and operations in WRRFs, and evaluate it comprehensively with the various scenarios simulated with its corresponding white-box model. This paper presents the development of the model based on grey-box modelling and EKF. SUMO (Dynamita™), an extensively used commercial simulator, was used to simulate a typical and well understood bioprocess, acting like a virtual WRRF to generate data. A grey-box model structure was identified and validated under different scenarios. This model structure was then converted into an EKF to overcome the adaptivity issue. Finally, performance of the EKF was evaluated by comparing the outputs of the EKF-based model to SUMO simulation results.

2. MATERIALS AND METHODS

The realization pathway of this paper is shown in Figure 1. SUMO was used as a virtual plant for data generation. Plant performance simulated with SUMO, at design influent flow and loads and different operational scenarios, was used as a reference to evaluate the performance of the grey-box model. The study was divided into two phases: (I) Grey-box modelling, and (II) Implementation of EKF. In Phase I, input and output data under different scenarios were collected from SUMO simulations, and a grey-box model structure was identified and validated in MATLAB offline. In Phase II, the grey-box model structure was converted into an EKF, the adaptive dynamic model, in Python. Critical steps included discretizing the grey model structure in Phase I and setting parameters to estimate as new states. Influent flow, loads and operations data streams were then generated in Python with noise and fed into SUMO, and performance data streams were retrieved from SUMO, also with added noise. Noise addition was intended to further simulate real sensor signals. The same noisy data streams were used as input to the EKF, and outputs from the EKF were compared with SUMO simulation results. Intuitive information, which requires less professional knowledge to understand, interpret and take actions, was transformed and updated from data produced by the EKF-based model for operations and control. Phase II can be viewed as an upgrade of Phase I in the following aspects: (1) Phase II was a real-time implementation while Phase I was offline; (2) Phase II used EKF to adaptively estimate the parameters in the grey-box model by setting them as new states; (3) Phase II further reduced the number of required sensor signals for estimation.

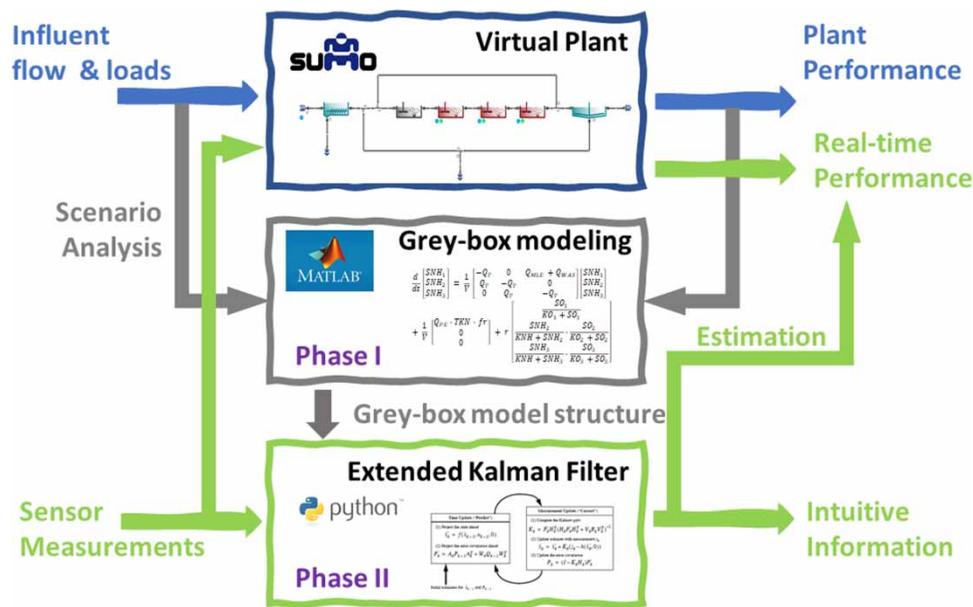


Figure 1 | The realization pathway of the adaptive real-time grey-box model in this paper. The blue arrows represent data in and out of the virtual plant. The grey arrows represent inputs and outputs of the virtual plant under different scenarios that were used in Phase I for identifying a grey-box model structure. After validation, the grey-box model structure was used to develop an EKF. Simulated sensor data streams were fed into both the virtual plant and the EKF in real-time and outputs from both were compared. Additionally, the EKF generated intuitive information for plant operation and management.

2.1. Virtual plant

2.1.1. Model configuration

The virtual plant process simulated was a typical Modified Ludzack-Ettinger (MLE) activated sludge process, which is widely used for biological wastewater treatment (Grady *et al.* 2011). It consisted of a bioreactor with an anoxic zone, an aerobic zone with three sequential stages, and a mixed liquor recirculation from the end of the bioreactor to the anoxic zone. The layout as represented in SUMO is depicted in Figure 2. The anoxic zone and aerobic zones were represented as continuously stirred tank reactors (CSTRs) in series, each with a volume of $2,000 \text{ m}^3$. The primary and secondary clarifiers were modelled as ideal separators. The primary clarifier had a suspended solid removal efficiency of 60%, and the secondary clarifier had a fixed effluent solids concentration of 10 mg total suspended solids per litre. The recirculated activated sludge flow rate (QRAS) and the internal recirculation flow rate (QMLE) were set to $23,400 \text{ m}^3/\text{d}$ (roughly 100% of QPE) and $36,000 \text{ m}^3/\text{d}$ (roughly 150% of QPE), respectively. Diurnal patterns are considered in the flow rate, which was observed in primary effluent (QPE), whose average value was around $23,400 \text{ m}^3/\text{d}$. Influent total Kjeldahl nitrogen (TKN) was varied on a dynamic basis, with an average around 56.6 mg-N/L . Dissolved oxygen levels ($\text{SO}_{2,i}$) varied from $0.5\text{--}3 \text{ mg/L}$. Detailed statistics are provided in Table S1 in the supplementary material. Biological kinetics were left at their default values (full plant model-SUMO1, Version 20~nb201104), except for the half saturation of ammonia for nitrifiers, which was set to 0.5 mg-N/L , and the half saturations of oxygen for nitrifiers in the three aerobic tanks, which were $0.6, 0.4, 0.2 \text{ mg-O}_2/\text{L}$, respectively. Adjustments to half saturations of oxygen for nitrifiers are often found useful in practice and reflect decreasing competition of nitrifiers for dissolved oxygen with heterotrophs because of decreasing heterotrophic activity through the bioreactor (B. Johnson, G. Daigger, personal communication, March 4, 2019). All other settings not mentioned, including the default COD-based wastewater characteristics, were left at the default values in SUMO. A Excel file including all needed information to reproduce the virtual plant has been provided in the supplementary material.

2.1.2. Data management

Sensors and meters assumed to be present and used in the analysis are denoted by the circles in Figure 2. In total, three flow rates (Q) were measured: (1) the primary effluent, QPE; (2) the internal recirculation from the last aerobic tank to the head of the anoxic tank, QMLE; and (3) the recirculated activated sludge, QRAS. The soluble ammonia into the system was needed for the grey-box modelling. However, in this study, the TKN was used as input because the chosen SUMO influent unit did not include soluble ammonia as input. The soluble ammonia signal was indirectly simulated as 70% of the TKN. In practice, ammonia sensors are used as TKN sensors are not available. Dissolved oxygen ($\text{SO}_{2,i}$) and soluble ammonia concentrations

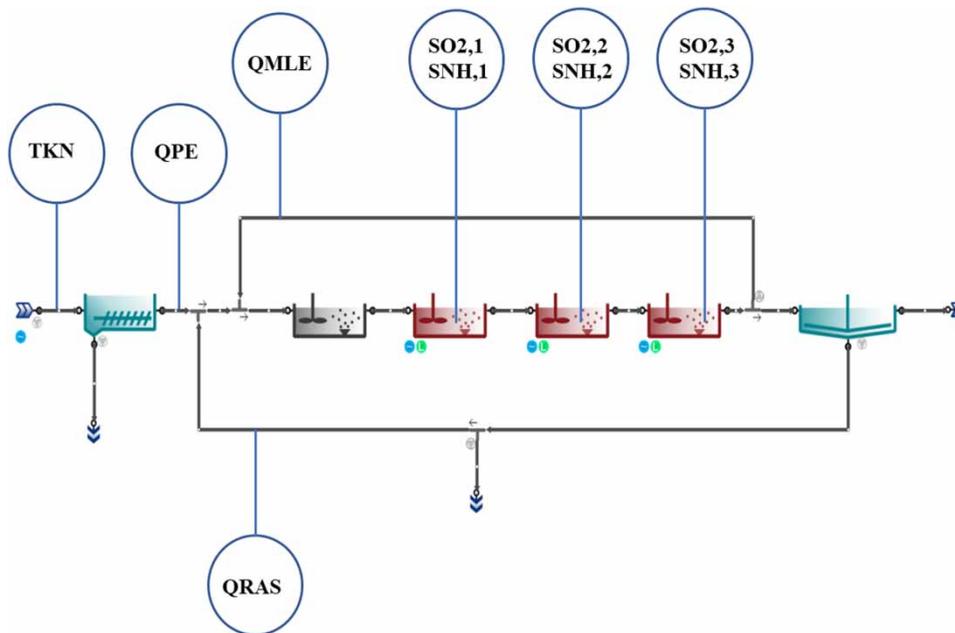


Figure 2 | The process layout and locations of assumed sensors and meters.

(SNH_i) in the aerobic tanks were extracted from SUMO for model fitting and evaluation. The sampling interval for all measurements was 10 minutes, and the simulation step size in SUMO was around 1 minute.

2.2. Grey-box model structure, identification and validation

2.2.1. Grey-box model

The development of a grey-box model typically begins with the general mathematical model for biochemical reactions in one reactor as described by the mass balance for substrates and biological reaction rate, as in Equation (1):

$$\frac{dC}{dt} = \frac{1}{V} (Q_{in} \cdot C_{in} - Q_{out} \cdot C_{out}) + \rho \quad (1)$$

where:

- $\frac{dC}{dt}$ is the net concentration change rate of the substrate.
- V is the reactor volume.
- Q_{in} and Q_{out} are flows in and out of the reactor.
- C_{in} and C_{out} are substrates concentrations in and out of the reactor.
- ρ is the biochemical reaction rate.

For a white-box model ρ would be a complex function of many other model state variables (such as the nitrifier biomass concentration) and model parameters (such as kinetic and stoichiometric factors). The authors considered a wide number of simplifications of the ρ function as the basis for the grey-box model presented here. Ultimately, the authors selected a highly simplified form consisting of a maximum reaction rate (r) multiplied by relevant Monod functions modifying the maximum reaction rate for DO ($SO_{2,i}$) and soluble ammonia (SNH_i) concentration in each aerobic reactor. The complete model for the three aerobic reactors is depicted in matrix form in Equation (2) where one can observe the biochemical reactor rate described above:

$$\frac{d}{dt} \begin{bmatrix} SNH_1 \\ SNH_2 \\ SNH_3 \end{bmatrix} = \frac{1}{V} \begin{bmatrix} -Q_T & 0 & Q_{MLE} + Q_{WAS} \\ Q_T & -Q_T & 0 \\ 0 & Q_T & -Q_T \end{bmatrix} \begin{bmatrix} SNH_1 \\ SNH_2 \\ SNH_3 \end{bmatrix} + \frac{1}{V} \begin{bmatrix} Q_{PE} \cdot TKN \cdot fr \\ 0 \\ 0 \end{bmatrix} + r \begin{bmatrix} \frac{SO_{2,1}}{KO_1 + SO_{2,1}} \\ \frac{SNH_2}{KNH + SNH_2} \cdot \frac{SO_{2,2}}{KO_2 + SO_{2,2}} \\ \frac{SNH_3}{KNH + SNH_3} \cdot \frac{SO_{2,3}}{KO_3 + SO_{2,3}} \end{bmatrix} \quad (2)$$

where:

- SNH_i , $SO_{2,i}$ are the soluble ammonia and dissolved oxygen concentrations in the i^{th} aerobic tank, respectively, which are measured in these simulations by sensors.
- Q_{PE} , Q_{MLE} , Q_{RAS} are the measured volumetric flow rates as depicted in Figure 1, and Q_T is the sum of Q_{PE} , Q_{MLE} , Q_{RAS} .
- TKN is the measured total Kjeldahl nitrogen.
- KO_i , KNH and r are parameters to be estimated, which represent half saturation concentrations for oxygen and ammonia and the maximum ammonia change rate, respectively.
- V and fr are fixed parameters that stand for the reactor volume and fraction of ammonia in TKN.

Essentially, the first term in Equation (2) represents internal ammonia transportation between tanks, and the second term is the external ammonia loading into the system. The last term is the ammonia changes due to biological reactions.

Several assumptions were made for this model structure:

- The change of ammonia due to biochemical reactions, including hydrolysis, growth and decay of both heterotrophs and nitrifiers, is combined in the last term, in the form of the maximum rate times the Monod Saturation terms. Non-dominant kinetics are reflected in the estimated KO_i , KNH , and r .

Table 1 | Summary of different scenarios

Scenario index	Flows	TCOD	TKN	SO _{2,i} 's	Scenario feature
1	Constant	Constant	Constant	Varied	Constant loadings
2	Varied, Periodic	Varied	Varied	Same as 1	Varying loadings
3		Same as 2 + noise			Measurement noise
4		Same as 2			Varying SRT
5		Same as 2			Varying temperature

- The influence of factors such as wastewater composition, active biomass and temperature is implicitly embedded in the estimated maximum nitrification rate (r). Section 3.1 of this paper discusses how this assumption caused model mismatches due to variations in r . The adaptive nature of the EKF proved to be effective in accounting for these variations.
- The ammonia concentration in the first aeration tank is often much greater than the ammonia half saturation coefficient. Therefore, the half saturation expression for ammonia is discarded in the first aeration tank to reduce non-linearity in the grey-box model.

Note that, since the biological reaction term is comprised of a generalized rate, the maximum reaction rate (r), modified by relevant half saturation terms, in its general form it can represent a variety of specific biological transformations. Consequently, Equation (2) may be viewed as a general biological reaction model with the constituents and half saturation functions adjusted to the particular biological reaction being considered.

2.2.2. Scenario analysis

Several scenarios were used to evaluate the ability of the model structure presented in Equation (2) to estimate ammonia concentrations in the virtual plant model and to identify when model parameters needed to be adjusted. A summary of the scenarios evaluated is provided in Table 1, and detailed statistics are provided in Table S1 in the supplementary material. Scenarios 1 & 2 provided a baseline evaluation. Scenario 3 evaluated the sensitivity of model parameter estimations to measurement noise. Scenarios 4 & 5 investigated how estimated parameters changed when factors that are expected to affect r varied, like sludge retention time (SRT) and temperature.

For each scenario, a 7-day period was simulated in SUMO and data was collected and divided equally into two sets: training and testing. The training set was used for parameter estimation and the testing set was used for model validation. The estimated parameters were accepted only when (1) no substantial deviations (within 2 standard deviations for most of the time) in the grey-box modelling prediction were observed, and (2) The Normalized Root Mean Square Error (NRMSE) and R^2 of the testing set equal or close to that of the training set.

2.2.3. Performance evaluation metrics and implementation

Performance of the grey-box model structure was evaluated by comparison of the virtual plant performance and the grey-box modelled values for ammonia concentrations in the three aerated tanks. The metrics used were the goodness of fit (R^2) and the NRMSE, as defined in Equations (3) and (4), respectively. Greater R^2 and smaller NRMSE generally indicate better performance:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$NRMSE = \frac{RMSE}{Max(y) - Min(y)} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{Max(y) - Min(y)} \quad (4)$$

where:

- y is the true value.
- \bar{y} is the mean of the true values.
- \hat{y} is the estimated value.
- RSS and TSS are residual sum of squares and total sum of squares, respectively.

Estimation of the parameters in the grey-box model was accomplished offline in MATLAB (R2020a) with the System Identification Toolbox Estimate Nonlinear Grey-Box Models (<https://www.mathworks.com/help/ident/ref/idnlngrey.html>). The search method for estimation was 'lsqnonlin' (Optimization toolbox).

2.3. Implementation of the extended Kalman filter

2.3.1. Discrete non-linear grey-box model

A general discrete non-linear grey-box model with noise can be written in state-space form as shown in Equation (5), with a subset of variables that could be measured as in Equation (6):

$$x_{k+1} = f(x_k, u_k) + v_k \quad (5)$$

$$z_k = h(x_k, u_k) + w_k \quad (6)$$

where:

- $(\cdot)_k$ denotes variables at the time step k .
- x_k, z_k, u_k denote the vectors of state, observed output, and input at time step k , respectively.
- $f(x_k, u_k)$ is the process model function, which is discretized from Equation (2), and $h(x_k, u_k)$ is the measurement model function.
- $v_k \sim N(0, V)$ is the gaussian process noise with zero mean and covariance Q , and the same for $w_k \sim N(0, W)$, the measurement noise.

2.3.2. Extended Kalman filter (EKF)

The EKF is a recursive method for state estimation of non-linear process models and measurement models (Equations (5) and (6)). It is an optimal state estimator in the sense that it minimizes the error variance. More details about its mathematical origin and practical implementation can be found in Welch & Bishop (2006). The general steps for the EKF update equations are listed in the sequence of its implementation, as follows:

$$\hat{x}_k^- = f(\hat{x}_{k-1}^-, u_{k-1}) \quad (7)$$

$$P_k^- = F_k P_{k-1}^- F_k^T + W \quad (8)$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + V)^{-1} \quad (9)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - h(\hat{x}_k^-, u_k)) \quad (10)$$

$$P_k = (I - K_k H_k) P_k^- \quad (11)$$

where:

- \hat{x}_k^- and \hat{x}_k denotes a *prior* and a *posterior* state.
- P_k^- and P_k are the covariance matrices of a *prior* and a *posterior* estimation error.
- K_k is the Kalman filter gain and I is the identity matrix.
- F_k and H_k are the Jacobian matrix of partial derivatives of f and h with respect to x , which are evaluated with estimates (x) and input (u) at time step k , that is, $F = \frac{\partial}{\partial x} f(x, u)|_{x=\hat{x}_k, u=\hat{u}_k}$ and $H = \frac{\partial}{\partial x} h(x, u)|_{x=\hat{x}_k, u=\hat{u}_k}$.
- W and V are covariance matrices of the process and measurement noises respectively. They can be seen as the 'tuning knobs' that determine how much one can trust in the measurements and the process.

2.3.3. Online implementation

Determination of variation in the maximum nitrification rate, r , was accomplished by including it as a new state in x , as $x = [SNH1 \quad SNH2 \quad SNH3 \quad r]^T$, and Equation (1), Equations (5)–(11) were adjusted correspondingly. In addition, unlike Phase I in which all three ammonia measurements are used ('measured'), the EKF can estimate all states by only measuring SNH_3 , therefore, $h(x, u) = [0 \quad 0 \quad 1 \quad 0]x$ and $H = [0 \quad 0 \quad 1 \quad 0]$. In Phase II, noise was added to all measurements before being fed into the EKF.

Communication between SUMO and Python was achieved by the SUMO-Python interface module developed by Dynami. The Jacobian matrix calculation was accomplished with the Python toolbox SymPy (<https://www.sympy.org/en/index.html>). The EKF was modified from the toolbox FilterPy (<https://github.com/rlabbe/filterpy>). The Python scripts used in this paper are shared in the supplementary material.

3. RESULTS AND DISCUSSION

3.1. Scenario analysis results

The estimated parameters and model performance metrics are presented in Tables 2 and 3, respectively. Scenario 1 represents constant ammonia loadings and operations, providing reference information about the nitrification capacity of the system. The estimated parameters were close to the values set in SUMO. Small deviations were expected given that unmodelled dynamics might act as corrections for the kinetic parameters. In Table 3, consistently large R^2 and small NRMSE in both the training and testing sets indicated little overfitting. It is important to note that, in general, the NRMSE of ammonia in the last tank is expected to be larger than the previous two. Because the ammonia concentration in the last tank has a relatively smaller range, the same absolute error results in larger NRMSE after normalization.

Scenario 2 investigated model performance when ammonia loadings varied. The loading patterns are shown in Figure 3(a). The influent flow (QPE) was designed to have a daily pattern with small shifts, and the TKN concentration fluctuated. Other inputs were kept the same as Scenario 1. Grey box model performance is shown in Figure 3(b) for both the training data and the testing data. By visual inspection, the model captured system dynamics, as further supported by the metrics in Tables 2 and 3 (large R^2 and small NRMSE). Scenario 3 added noise to all measurements to investigate the parameter estimation sensitivity to measurement noise. Although differences were found in estimated parameters (Table 2), model performance was acceptably good by visual inspection of Figure S1 and similar R^2 and NRMSE as Scenario 2 in Table 3.

Scenarios 4 and 5 investigated the effect of parameters known to change system performance, specifically temperature and SRT, on grey-box model performance. These scenarios investigate situations where the wastewater temperature (Scenario 4) or SRT (Scenario 5) in the actual treatment plant (simulated here using the virtual plant in SUMO) change. The grey-box model is calibrated to system performance for the previous operating condition (higher temperature or higher SRT) during the training period and then the performance of the previously calibrated grey-box model is evaluated compared to the performance of the virtual plant when the temperature or SRT decrease. The results indicated that model performance deteriorated when the system temperature and SRT changed. This is illustrated for Scenario 4 in Figure 4. When the temperature decreased, the parameters estimated from the training set (20 °C) no longer remained valid for the testing set (15 °C),

Table 2 | Estimated parameters under different scenarios

Scenario Index	r	KNH	KO1	KO2	KO3
TRUE	–	0.5	0.6	0.4	0.2
1	–325	0.49	0.56	0.32	0.18
2	–334	0.48	0.58	0.30	0.17
3	–321	0.35	0.50	0.30	0.10
4	–322	0.46	0.58	0.30	0.10
4 ^a	–244	0.45	0.52	0.34	0.20
5	–323	0.46	0.57	0.30	0.10
5 ^a	–303	0.50	0.63	0.37	0.27

^aIndicates parameters are re-estimated after the conditions (temperature and SRT) has changed.

Table 3 | Grey-box model performance under different scenarios

Scenario index	SNH1				SNH2				SNH3			
	Training set		Testing set		Training set		Testing set		Training set		Testing set	
	R ²	NRMSE										
1	0.92	0.10	0.93	0.10	0.94	0.08	0.94	0.09	0.91	0.13	0.90	0.23
2	0.97	0.06	0.97	0.06	0.97	0.07	0.97	0.05	0.95	0.07	0.96	0.33
3	0.97	0.06	0.97	0.06	0.96	0.07	0.97	0.06	0.95	0.13	0.95	0.53
4	0.97	0.06	0.88	0.24	0.97	0.06	0.86	0.32	0.95	0.08	0.84	1.08
4 ^a	0.95	0.23	0.93	0.08	0.91	0.48	0.92	0.09	0.86	1.37	0.93	0.29
5	0.97	0.06	0.94	0.11	0.97	0.06	0.93	0.16	0.96	0.08	0.90	0.82
5 ^a	0.97	0.10	0.95	0.07	0.96	0.17	0.94	0.08	0.94	0.35	0.92	0.35

^aIndicates parameters are re-estimated based on the testing set and validated based on the training set, when the conditions (temperature and SRT) have changed.

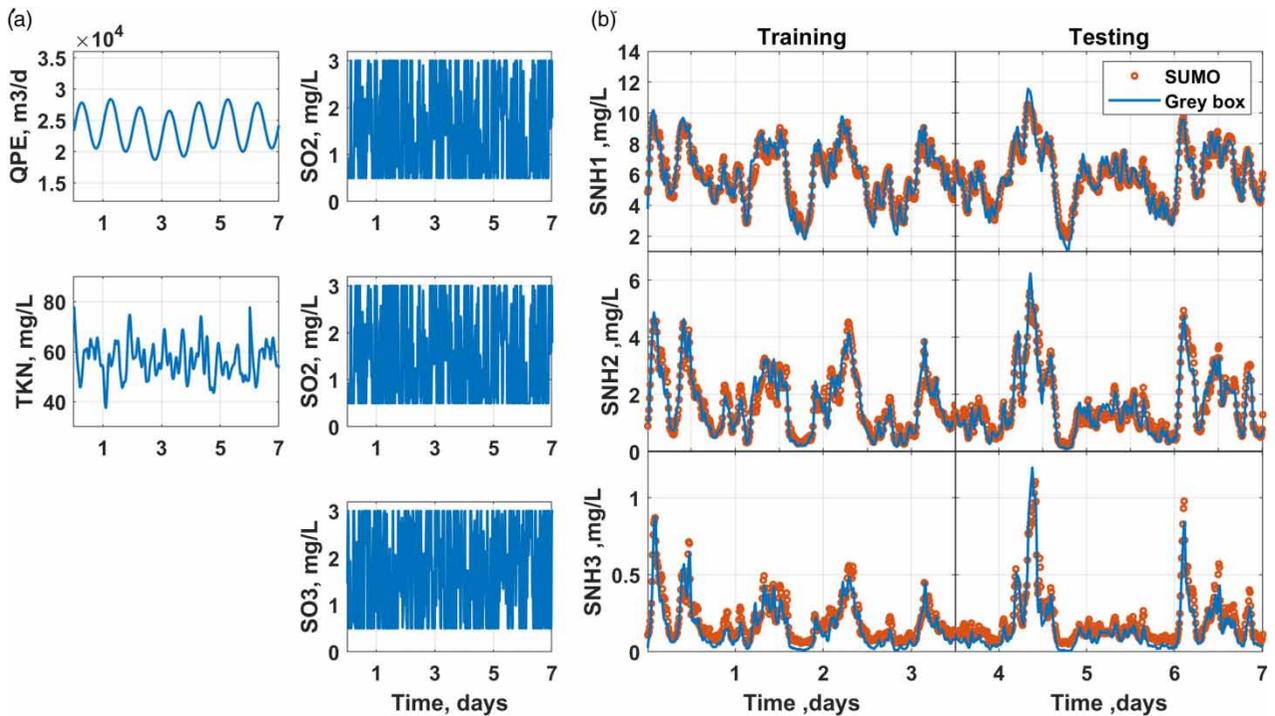


Figure 3 | An example of scenarios when the grey box model performs well (Scenario 2): (a) model inputs. QMLE and QRAS are flat lines therefore are not shown. (b) Predictions. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/wst.2021.408>.

as large deviations were observed between the SUMO simulation and the grey-box model prediction in the testing set. The R² dropped and NRMSE increased dramatically for the testing set, as indicated by the results presented in Table 3. Grey-box model performance returned to previous levels when it was calibrated to the virtual plant performance for the altered operating conditions (lower temperature or SRT), as illustrated by Scenario 4* (results for the test data set are shown in Figure S3). However, in the training set, the grey-box model performance deteriorated for the 3.5 days prior to the temperature change. The major difference in estimated parameters for the grey-box model prior and after the change of temperature was the maximum ammonia reaction rate, r , whose absolute value dropped from 311 to 244 mg-N/L/day, implying that the system had smaller nitrification capacity. This is a logical outcome as it is commonly acknowledged that the correlation between biomass activity and temperature follows $r_T = r_{20} \times \theta^{(T-20)}$, where θ is between 1.03 to 1.1 in the literature (Grady *et al.* 2011; Rieger

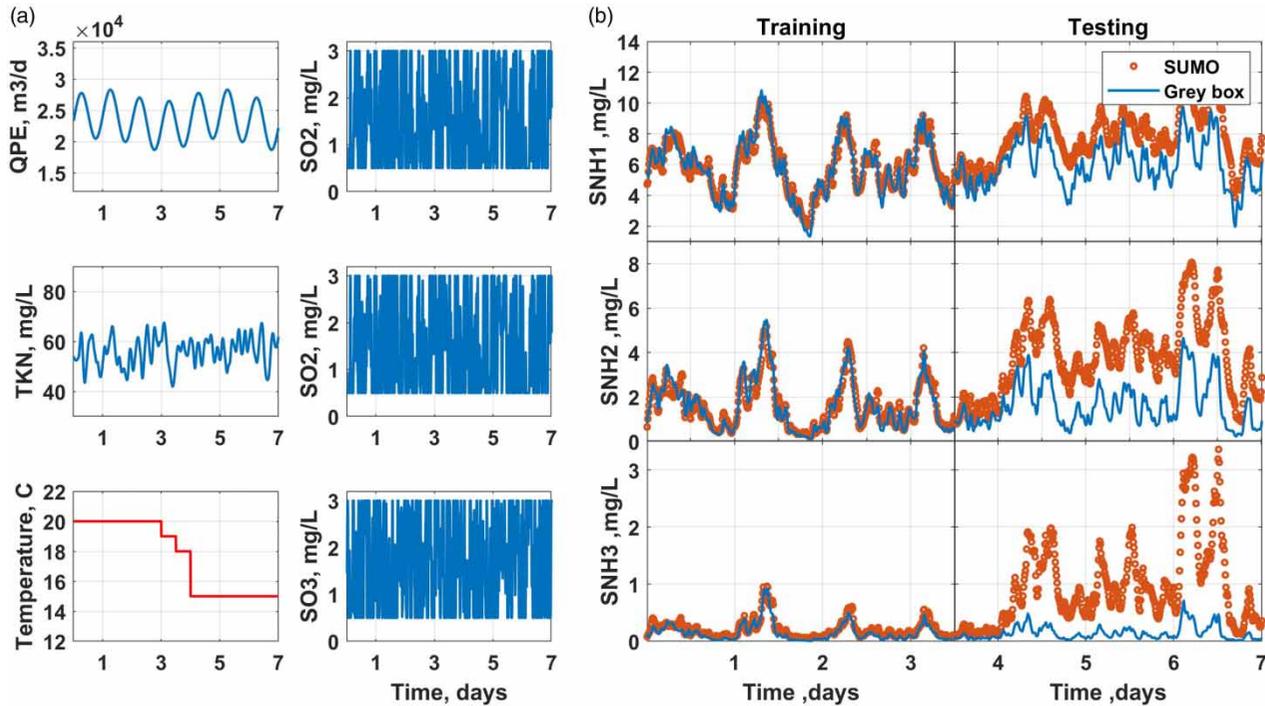


Figure 4 | An example of scenarios when the grey-box model fails (Scenario 5): (a) model inputs. QMLE and QRAS are flat lines therefore are not shown. (b) Predictions of three ammonia concentrations. Training set (early 3.5 days) is for parameter estimation. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/wst.2021.408>.

et al. 2012; Henze *et al.* 2015). In this study, θ equalled 1.05, obeying the temperature dependence. Similar results were observed when the SRT decreased (Figure S2 and S4). These results suggested that the parameters re-estimation of the grey-box model to plant data can detect changes in system dynamics (input-output mapping), as expressed in r . With different values of r , the same inputs (DOs) result in different outputs (ammonia concentrations). In this case, the change in r is expected and interpretable with known process knowledge. What was needed was recursive estimation of the time-varying model parameter, which motivated development of the EKF.

3.2. Performance of the extended Kalman filter

The EKF was implemented in real time in the SUMO–Python interface for different loading and operating conditions. An example, demonstrating the ability of the EKF to track the maximal nitrification rate, is presented in Figure 5. The temperature decreased from 20 to 15 °C on day 5, without changes in other conditions. Other examples are provided in Figure S5 and Figure S6. Noise was added for every other measurement. Another difference from Phase I was that only one ammonia measurement (last tank, the green line in Figure 5(b) bottom left panel) was fed into the EKF.

In Figure 5(b), reasonable estimations of the ammonia concentrations in the previous two tanks were observed as the estimation curves converged to simulated ‘true’ values in SUMO. The noisy fluctuation is due to measurement noise propagating through the EKF. Fine-tuning or extra filtering may further reduce such noise. Moreover, the maximal nitrification rate, r , followed the trend of the temperature drop as the curves started dropping at around day 5, indicating a good tracking of the r value.

Performance of the EKF was satisfactory in the sense that: (1) it was able to provide reasonable estimates of the ammonia concentrations and parameters in all three tanks with fewer signals. This can help reduce cost in practice because the number of sensors employed directly relates to installation and maintenance cost. (2) It recursively re-estimated r in real time and, therefore, remedied the deficit shown in Section 3.1 (the estimated parameters were no longer valid when system equilibriums changed). (3) It enabled downstream applications, for instance advanced control design like model predictive control and full-state feedback.

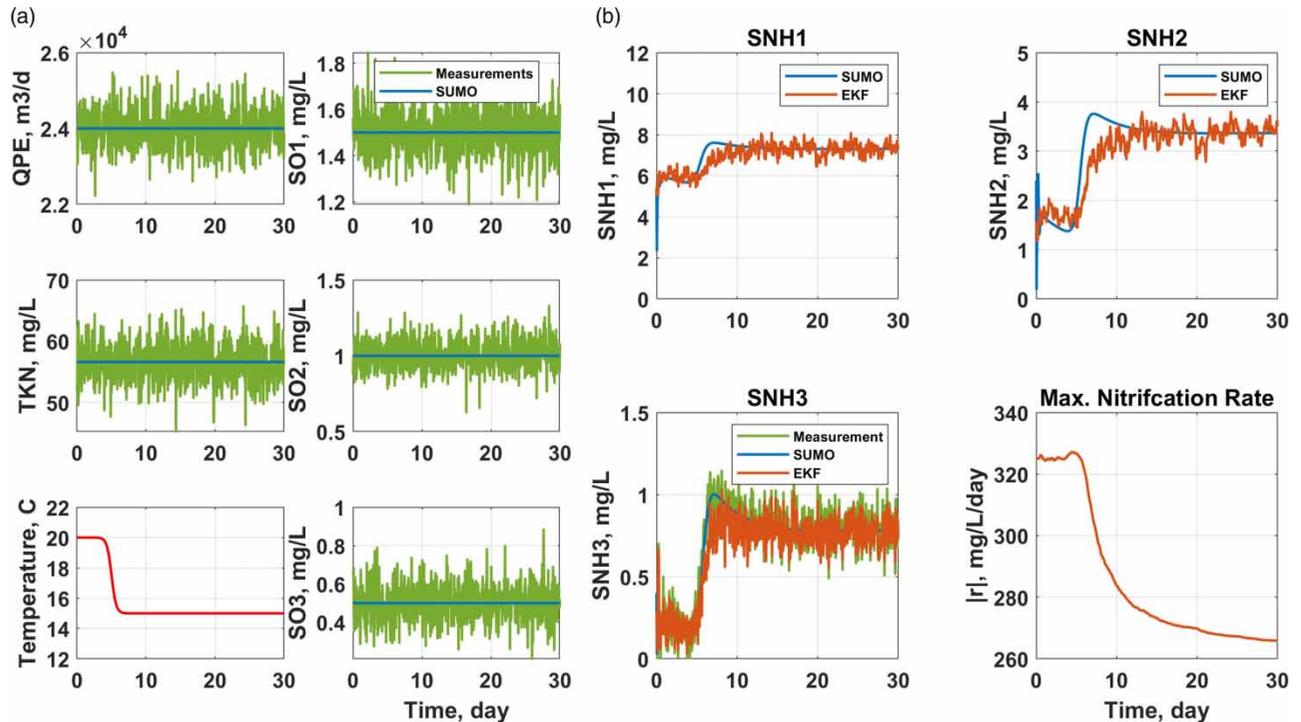


Figure 5 | The performance of EKF when temperature drops. (a) Model inputs. (b) Estimation of three ammonia concentrations and maximum nitrification rate, r . Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/wst.2021.408>.

3.3. Considerations for the implementation of the extended Kalman filter

3.3.1. Valid grey-box model structure

The structure of the grey-box model needs to be validated before implementation of the EKF, as was done in this study. A suggested approach would be to start with the full white-box model and remove less relevant components step by step based on appropriate assumptions (Nair *et al.* 2019; Nair *et al.* 2020). Another approach is to start from the major biochemical reactions and add extra components if models exhibit lack of fit with plant data (Stare *et al.* 2006). Grey-box model structure should be fit for purpose. In this study, the aim was to model nitrification alone, therefore, minimal but sufficient model components were used. For other processes such as simultaneous nitrification/denitrification and biological phosphate removal, more reaction equations and state variables may be needed.

3.3.2. Tuning the extended Kalman filter

The convergence rate and trade-offs between process and sensor noise depend on the two tuning parameters, V and W , which are the error covariance of the state and measurements. In practical applications, these are not known precisely and, therefore, trial-and-error tuning must be expected. In this paper, noise in flows, TKN, and DO propagated through the process model as process noise. Therefore, V was written as $J_k M J_k^T$, where M is the covariance matrix of inputs (i.e. flows, TKN, and $SO_{2,i}$'s) and J is the Jacobian matrix of partial derivatives of f with respect to u evaluated at time step k , $J = \frac{\partial}{\partial u} f(x, u)|_{x=\hat{x}_k, u=\hat{u}_k}$. Therefore, the 'tuning knob' in this paper was M and W , which were chosen from measurement covariance without further fine tuning.

3.3.3. Observability test

Observability is a system property which guarantees that, given inputs (u in Equation (5)) and measurements (z in Equation (6)), it is possible to estimate the state values. When a system is observable, the EKF can be developed and incorporated into control design as an observer. However, since not all systems are observable, observability tests should be performed in advance. For non-linear systems, observability tests are usually performed on the linear approximation of the system via conventional methods, such as the Popov–Belevich–Hautus (PHB) rank test (Kailath *et al.* 2000). In this study, investigations on

the observability revealed one ammonia signal was sufficient for full-state estimation. It is beyond the scope of this paper to discuss observability, but Busch *et al.* (2013) provide an integrated approach for observability testing for large-scale wastewater treatment plants.

3.4. Significance of intuitive information in real-time grey-box model

The EKF is not new in the wastewater sector, but most uses focus on estimating unmeasurable states instead of yielding intuitive information such as the maximum nitrification capacity as in this study (Ayasa *et al.* 1991; Stare *et al.* 2006; Busch *et al.* 2013; Nair *et al.* 2019). A monitoring system yielding intuitive information is valuable in that:

- (1) It supports decision making. A well-educated process engineer may quickly translate data into information based on knowledge and experience. Plant operation staff, especially those in their early career, might not have the same level of knowledge. Intuitive information is easier to understand and therefore prompt decision could be made. For instance, redundant capacity in r allows shorter SRT. It should be noted that the proper actions taken in response to a change in r still require basic process knowledge.
- (2) It assists system monitoring. Even if not used for automatic control, the intuitive information itself is still valuable in the sense of monitoring the system. In this case, the trend of r could be viewed as a soft sensor monitoring the nitrification capacity of the system. A dramatic change in r without explainable causes could be an alert for anomalies, warning operators to diagnose issues in operations and instrumentation.
- (3) It improves the control design by adaptive tuning of controller gains. The sensitivity of ammonia to dissolved oxygen (DO) relies on the r values. In other words, the same change in DO results in different changes in ammonia depending on system maximum nitrification capacity as reflected here in the numerical value of r . It was demonstrated in Section 3.2 that r is a time-varying parameter due to changes in temperature and operational settings such as SRT, among other potential factors. A controller tuned for one r value may be inappropriate as the value of r varies. With the adaptively estimated grey-box model, controller gains could be retuned once significant variation in r is observed. Alternatively, plant operating conditions can be adjusted to return to an r value that allows for better control.

The development of the EKF relies on an adequate grey-box model structure. In this study, as shown in Equation (2), the grey-box model is relatively simple – mass balance plus reaction rates governed by Monod kinetics. Although the grey-box model lacks many traditional model components (e.g., biomass concentration, COD) when compared with state-of-art ASM models, the adaptive scheme can incorporate impacts from those neglected variables into the r values, ensuring the accuracy of the model. One benefit of this simplicity is the increased applicability to other processes, especially those that are not fully understood yet.

4. CONCLUSIONS

This paper presents the development of a grey-box model able to adaptively estimate the nitrification capacity of an MLE process in real time. Although simple, the grey-box model structure was designed with target information embedded. The model was completed by estimating its parameters from data and was validated under different scenarios. Results of scenario analysis revealed the need to update parameters adaptively to address the changing system dynamics.

An EKF was therefore developed with the grey-box model structure. With Python interfacing SUMO, a widely used process simulator, EKF performance was evaluated in real time. Results showed that the EKF was able to observe and track the nitrification capacity accurately with fewer sensor signals. Such an adaptive real-time model is valuable in that: (1) it provides intuitive information for decision making on operations; and (2) it enables advanced control design (e. g. model predictive control and state feedback control) and adaptive tuning for controller gains.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Alex, J., Benedetti, L., Copp, J., Germaey, K. V., Jeppsson, U., Nopens, I., Pons, M.-N., Rieger, L., Rosen, C., Steyer, J. P., Vanrolleghem, P. & Winkler, S. 2008 'Benchmark Simulation Model no. 1 (BSM1).' Report by the IWA Taskgroup on benchmarking of control strategies for WWTPs, pp. 19-20. Available online: <https://www.iea.lth.se/publications/Reports/LTH-IEA-7229.pdf>

- Ayesa, E., Florez, J., Garcia-Heras, J. L. & Larrea, L. 1991 State and coefficients estimation for the activated sludge process using a modified Kalman filter algorithm. *Water Science and Technology* 24 (6), pp. 235–247.
- Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., Sanders, W. T., Siegrist, H. & Vavilin, V. A. 2002 *The IWA anaerobic digestion model No 1 (ADM1)*. *Water Science and Technology* 45 (10), 65–73. (2021).
- Busch, J., Elixmann, D., Kühl, P., Gerkens, C., Schlöder, J. P., Bock, H. G. & Marquardt, W. 2013 *State estimation for large-scale wastewater treatment plants*. *Water Research* 47 (13), 4774–4787.
- Grady Jr, C. L., Daigger, G. T., Love, N. G. & Filipe, C. D. 2011 *Biological Wastewater Treatment*. CRC Press.
- Haimi, H., Mulas, M., Corona, F. & Vahala, R. 2013 *Data-derived soft-sensors for biological wastewater treatment plants: an overview*. *Environmental Modelling and Software* 47, 88–107.
- Henze, M., Gujer, W., Mino, T. & van Loosdrecht, M. 2015 *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*. IWA Publishing, London, 121 pp.
- Kailath, T., Sayed, A. H. & Hassibi, B. 2000 *Linear Estimation*. Prentice Hall, New Jersey.
- Nair, A. M., Fanta, A., Haugen, F. A. & Ratnaweera, H. 2019 *Implementing an extended Kalman filter for estimating nutrient composition in a sequential batch MBBR pilot plant*. *Water Science and Technology* 80 (2), 317–328.
- Nair, A. M., Gonzalez-Silva, B. M., Haugen, F. A., Ratnaweera, H. & Østerhus, S. W. 2020 *Real-time monitoring of enhanced biological phosphorus removal in a multistage EBPR-MBBR using a soft-sensor for phosphates*. *Journal of Water Process Engineering* 37, 101494.
- Newhart, K. B., Holloway, R. W., Hering, A. S. & Cath, T. Y. 2019 *Data-driven performance analyses of wastewater treatment plants: a review*. *Water Research* 157, 498–513.
- Rao, C. V., Rawlings, J. B. & Mayne, D. Q. 2003 *Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations*. *IEEE Transactions on Automatic Control* 48 (2), 246–258.
- Regmi, P., Stewart, H., Amerlinck, Y., Arnell, M., García, P. J., Johnson, B., Maere, T., Miletic, I., Miller, M., Rieger, L., Samstag, R., Santoro, D., Schraa, O., Snowling, S., Takács, I., Torfs, E., van Loosdrecht, M. C. M., Vanrolleghem, P. A., Villez, K., Volcke, E. I. P., Weijers, S., Grau, P., Jimenez, J. & Rosso, D. 2019 *The future of WRRF modelling – outlook and challenges*. *Water Science and Technology* 79 (1), 3–14.
- Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takacs, I. & Winkler, S. 2012 *Guidelines for Using Activated Sludge Models*. IWA Publishing, London and New York.
- Rittmann, B. E., Boltz, J. P., Brockmann, D., Daigger, G. T., Morgenroth, E., Sørensen, K. H., Takács, I., Van Loosdrecht, M. & Vanrolleghem, P. A. 2018 *A framework for good biofilm reactor modeling practice (GBRMP)*. *Water Science and Technology* 77 (5), 1149–1164.
- Robertson, D. G., Lee, J. H. & Rawlings, J. B. 1996 *A moving horizon-based approach for least-squares estimation*. *AIChE Journal* 42 (8), 2209–2224.
- Shiva Kumar, B. & Venkateswarlu, C. 2012 *Estimating biofilm reaction kinetics using hybrid mechanistic-neural network rate function model*. *Bioresource Technology* 103 (1), 300–308.
- Stare, A., Hvala, N. & Vrečko, D. 2006 *Modeling, identification, and validation of models for predictive ammonia control in a wastewater treatment plant – A case study*. *ISA Transactions* 45 (2), 159–174.
- Stentoft, P. A., Munk-Nielsen, T., Vezzaro, L., Madsen, H., Mikkelsen, P. S. & Møller, J. K. 2019 *Towards model predictive control: online predictions of ammonium and nitrate removal by using a stochastic ASM*. *Water Science and Technology* 79 (1), 51–62.
- Therrien, J. D., Nicolai, N. & Vanrolleghem, P. A. 2020 *A critical review of the data pipeline: how wastewater system operation flows from data to intelligence*. *Water Science and Technology* 82 (12), 2613–2634.
- Welch, G. & Bishop, G. 2006 *An introduction to the Kalman filter*. Available online: https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf
- Zahedi, G., Elkamel, A., Lohi, A., Jahanmiri, A. & Rahimpor, M. R. 2005 *Hybrid artificial neural network – first principle model formulation for the unsteady state simulation and analysis of a packed bed reactor for CO₂ hydrogenation to methanol*. *Chemical Engineering Journal* 115 (1–2), 113–120.

First received 8 July 2021; accepted in revised form 3 September 2021. Available online 20 September 2021