Implementing an Extended Kalman Filter for estimating nutrient composition in a sequential batch MBBR pilot plant

Abhilash M. Nair **IWA**, Abaynesh Fanta, Finn Aakre Haugen **IWA** and Harsha Ratnaweera **IWA**

ABSTRACT

Online monitoring of water quality parameters can provide better control over various operations in wastewater treatment plants. However, a lack of physical online sensors, the high price of the available online water-quality analyzers, and the need for regular maintenance and calibration prevent frequent use of online monitoring. Soft-sensors are viable alternatives, with advantages in terms of price and flexibility in operation. As an example, this work presents the development, tuning, implementation, and validation of an Extended Kalman Filter (EKF) on a grey-box model to estimate the concentration of volatile fatty acids (VFA), soluble phosphates (PO_4-P), ammonia nitrogen (NH_4-N) and nitrate nitrogen (NO_3-N) using simple and inexpensive sensors such as pH and dissolved oxygen (DO). The EKF is implemented in a sequential batch moving bed biofilm reactor (MBBR) pilot scale unit used for biological phosphorus removal from municipal wastewater. The grey-box model, used for soft sensing, was constructed by fitting the kinetic data from the pilot plant to a reduced order version of ASM2d model. The EKF is successfully validated against the standard laboratory measurements, which confirms its ability to estimate various states during the continuous operation of the pilot plant.

Key words | biological phosphorus removal, Extended Kalman Filter, sequential batch biofilm process, soft-sensor

Abhilash M. Nair IWA (corresponding author)
Harsha Ratnaweera IWA
Faculty of Science and Technology,
Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Aas,
Norway
E-mail: *muralabh@nmbu.no*

Abaynesh Fanta
Department of Hydraulic and Environmental
  Engineering,
Norwegian University of Science and Technology
  (NTNU),
S. P. Andersens vei 5, Trondheim 7491,
Norway

Finn Aakre Haugen IWA
University of South-Eastern Norway,
Kjølnes ring 56, Porsgrunn,
Norway

INTRODUCTION

There has been a rapid increase in the implementation of advanced control strategies in process industries, including water resource recovery facilities (WRRF). These control strategies are essential for the optimal operation of process plants (O'Brien *et al.* 2011). Advanced control strategies such as model predictive control (MPC) would require continuous, real-time information of various wastewater compositions (Liukkonen *et al.* 2013).

A number of online instruments for measuring nutrient composition are available in the market today. However, their use is often limited to large-scale urban treatment facilities (Häck & Wiese 2006). Despite considerable developments in online instrumentation in the past decade,

physical sensors for real-time measurement of some of the wastewater quality parameters are either extremely expensive or do not exist. Moreover, sensors available in today's market are vulnerable to fouling, drift or other inaccuracies resulting in unreliable measurement (Olsson & Jeppsson 2006). A lot of effort is put into fault detecting systems (either manual or real-time) to remove errors and correct sensors readings. Most online sensors (both ion selective and optical) require a lot of attention in terms of data quality checks, through regular inspection and adjustment (Olsson 2012). On the other hand, composition analyzers provide an accurate measurement of parameters such as ammonia, nitrate, chemical oxygen demand (COD), and phosphates, but are extremely expensive. Due to these problems, most WRRF depend on standardized laboratory measurements rather than online monitoring. However, parameters such as pH, oxidation-reduction potential (ORP), dissolved

oxygen (DO) and conductivity can be measured with relative ease and higher accuracy. Therefore, these sensors, which are often inexpensive, are installed in almost all treatment plants including small to medium WRRF.

Soft sensors are an alternative to expensive composition analyzers and unreliable in-situ sensors (Thürlimann *et al.* 2018). Soft sensors, also denoted as state estimators, virtual sensors, etc., use system knowledge enclosed in the process model together with available measurements. These software sensors can be used not only as alternative to expensive analyzers but also as a support system for the existing composition analyzers by acting as a reliable backup in case of an instrument malfunction (Haimi *et al.* 2013). A detailed compilation of statistical models such as artificial neural networks (ANN), partial least-square regression (PLSR) or auto-regressive models (ARX) used for real-time estimation is provided in Haimi *et al.* (2013). Mechanistic models describing biological processes such as activated sludge model (ASM) can also be used with state estimators such as Extended Kalman Filter (EKF) or Moving Horizon Estimator (MHE) for developing soft sensors (Busch *et al.* 2013). However, most of these state estimators are tested and validated on the BSM 1 simulator. We have not found reports of implementation of state estimators with mechanistic models for soft sensing in a real system.

The aim of this work is to develop a dynamic state-estimator to provide real-time estimations of PO_4-P, volatile fatty acids (VFA), NH_4-N, and NO_3-N in a sequential batch reactor, by using data obtained from online sensors such as pH and DO.

MATERIALS AND METHODS

Pilot plant setup

The sequential batch moving bed Bio-P (SB-MBBR Bio-P) pilot plant, located in the wastewater laboratory at Norwegian University of Science and Technology (NTNU) was used to test and validate the soft-sensor algorithm. Figure 1 depicts the pilot plant and the data acquisition system. The sequential batch reactor (SBR) contains polyurethane carriers (with 60% filling) which are exposed to raw wastewater to ensure biofilm growth and attachment to the carriers (see Supplementary Figure 1, available with the online version of this paper). The reactor has a total working volume of 13 L. It operates with a fixed cycle time of 8 h and a temperature of $16 \pm 0.5\,°C$. The carrier media have a cylindrical shape (10 mm in diameter and length) with a cross
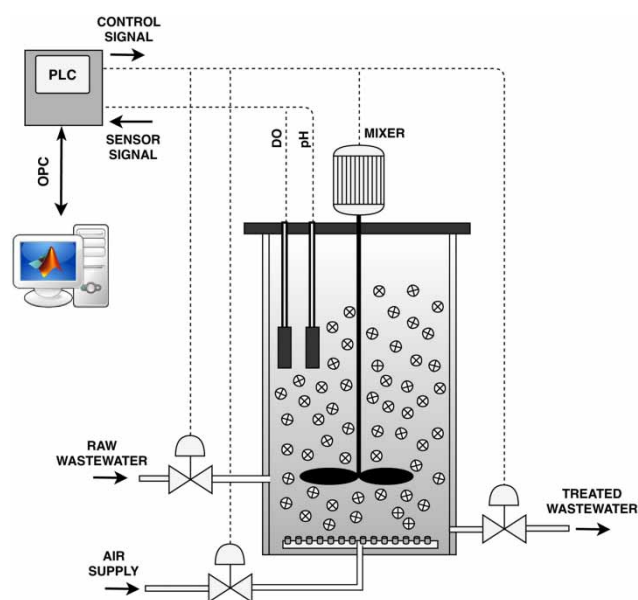

Figure 1 │ Pilot plant layout and the data acquisition system.

inside the cylinder and longitudinal fins outside, providing a large surface area for biofilm growth (Ødegaard *et al.* 1994). The SBR is fed with wastewater from a storage tank, which receives steady supply of raw wastewater from a nearby municipal sewer. The storage tank has a total volume of $3.5\ m^3$ and a residence time of 24 h, which ensures an uninterrupted supply of wastewater while maintaining sufficient variations in influent wastewater quality.

After ensuring the attachment of active biofilm to the carrier, the reactor operates in an alternating anaerobic–aerobic cyclic process with a hydraulic retention time of 8 h (3 h anaerobic, 5 h aerobic, 3 min filling and 3 min emptying). A sieve at the outlet line retains the carriers in the SBR, allowing just the treated water and detached biomass to exit the reactor. The reactor operates as a biofilm process and all active biomass is attached to the carrier media. Therefore, the SBR can be operated with a 100% volume exchange ratio, resulting in a cycle time equal to the hydraulic retention time. The system is mechanically agitated using a custom-made mixer for the entire cycle period. During the aerobic phase, air is supplied with a constant flow rate of $4.5\ L\ min^{-1}$. The algorithm to execute the cycle time and the operational sequence is stored in the Supervisory Control and Data Acquisition (SCADA) system of the plant.

During the anaerobic-stage, the active biomass (primarily PAOs) utilizes the acetate from raw wastewater and converts it into polyhydroxy-alkanoates (X_{PHA}). As a result, PO_4-P is released due to Poly-P degradation, thus increasing its concentration in bulk solution. In anaerobic conditions,

the readily biodegradable organic matter (BOD) is also converted to VFA by fermentation, and the PAOs utilize it for further PO$_4$-P release. When aeration begins, the PAOs utilize the stored X_{PHA} as an energy source to grow and take up PO$_4$-P from the water phase and synthesize new storage products within the biomass (X_{PP}). The biomass also contains significant proportions of nitrifying organisms, which convert NH$_4$-N to NO$_3$-N. The biomass also contains heterotrophs, which consume BOD during the aerobic stage.

When the biofilm layer on the carriers reach a steady thickness, the system exhibits more than 80% PO$_4$-P removal. In this period, a constant amount of PO$_4$-P is added at the beginning of every cycle to avoid low influent PO$_4$-P concentration due to dilution of wastewater by storm-water (9 ± 3 mg P L^{-1}). However, the soluble COD (127 ± 30 mg COD L^{-1}), NH$_4$-N (28 ± 16 mg N L^{-1}) is not adjusted.

After the biomass activity shows steady conditions, kinetic tests are conducted in the reactor. Samples are collected for the entire cycle every 30 min and immediately filtered through 0.45 μm filters to determine the concentration of soluble COD (sCOD), orthophosphate (PO$_4$-P), ammonium nitrogen (NH$_4$-N) and nitrate nitrogen (NO$_3$-N). All these compositions are measured using Dr. Lange cuvette and HACH LANGE GmbH (DR 1900, China) spectrophotometer. All analyses are performed immediately during the cycle test. The DO and pH sensor installed in the pilot plant are connected to the plants SCADA system. The data from the online sensors are logged every minute. The data received from the laboratory analyses and the online sensors are used to calibrate the model.

Mathematical model

A number of mechanistic models explaining the process of removal of phosphorus, nitrogen, and carbon in sequential batch reactors are available in the literature. However, using a complex model such as ASM2d (Henze et al. 1999) for estimation and control would be challenging, especially when the number of measurements available in the reactors are limited (Jeppsson & Olsson 1994). The usual approach is to reduce the number of state variables by either lumping two or more states or assuming some of the states with relatively slower dynamics as constants (Steffens et al. 1997). We find similar model-reduction strategies implemented on ASM1 model, for use in state estimation and control (Julien et al. 1999; Gómez-Quintero et al. 2000; García-Usach et al. 2006).

Model reduction strategy

The original ASM2d model (Henze et al. 1999) consists of 19 states and 21 processes. It describes the growth and decay of three types of biomass; the phosphate accumulating organisms (X_{PAO}), autotrophs (X_{AUT}) and heterotrophs (X_H), followed by the subsequent uptake and release of various substrates and storage products in the biomass. The following assumptions, similar to the simplification strategies presented in Zhang et al. (2010), Cadet (2014) and Li et al. (2018) were made to reduce the model.

- The effect of particulate hydrolysis reactions appears to be insignificant in the process and has been ignored.
- The non-reacting states in the model: particulate inert (X_I), soluble inert (S_I) and alkalinity (S_{ALK}) are eliminated.
- Since there is no anoxic condition in the SBR's operational sequence, all kinetics pertaining to denitrifiers in the biomass are removed. However, the denitrification reactions by X_{PAO} has been retained.
- The effect of growth and decay kinetics of all three biomass types can be excluded. Hence, the model only considers the kinetics of substrate degradation processes.

This reduced order model explains the dynamics of storage and consumption of soluble substrates such as readily biodegradable substrate (S_F), volatile fatty acids (S_A), phosphates (S_{PO}), ammonia (S_{NH}) and nitrates (S_{NO}). The model also includes the kinetics of storage products such as polyhydroxy-alkanoates (X_{PHA}) and polyphosphates (X_{PP}). The description of substrates and storage products used in the model is presented in Table 1. The list of processes included in the reduced order ASM2d model is presented in Table 2.

Table 1 | State variables describing the model

States	Description	Unit
S_F	Readily biodegradable substrate	mg COD L^{-1}
S_A	Fermentation products (Volatile fatty acids/acetate)	mg COD L^{-1}
X_{PP}	Stored poly-phosphate of PAO	mg P L^{-1}
X_{PHA}	Organic storage products of PAO	mg COD L^{-1}
S_{NH}	Ammonium	mg N L^{-1}
S_{NO}	Nitrate	mg N L^{-1}
S_{PO}	Phosphate	mg P L^{-1}
S_O	Dissolved oxygen	mg O$_2$ L^{-1}

Table 2 │ Peterson matrix and rate kinetics for the reduced order model (see Supplementary Figure 2, available with the online version of this paper)

Process	S_F	S_A	X_{PP}	X_{PHA}	S_{NH}	S_{NO}	S_{PO}	S_o	Rate kinetics
1 COD uptake	−1							−1	$r_1 * \dfrac{S_F}{K_S + S_F} * \dfrac{S_o}{K_0 + S_o}$
2 Fermentation	−1	1			0.1		0.01		$r_2 * \dfrac{S_F}{K_{feH} + S_F} * \dfrac{K_O}{K_0 + S_o} * \dfrac{K_{NO}}{K_{NO} + S_{NO}}$
3 Storage of PHA Release of P		−1	−Y_{PO4}	1			Y_{PO4}		$r_3 * \dfrac{S_A}{K_A + S_A} * \dfrac{X_{PP}}{K'_{PP} + X_{PP}}$
4 Storage of PP Uptake of P			1	−Y_{PHA}			−1	−Y_{PHA}	$r_4 * \dfrac{S_o}{K_0 + S_o} * \dfrac{S_{po}}{K_{PS} + S_{po}} * \dfrac{X_{PHA}}{K'_{PHA} + X_{PHA}} * \dfrac{K'_{MAX} - X_{PP}}{K'_{PP} + (K'_{MAX} - X_{PP})}$
5 Nitrification					−1	1		−4.32	$r_5 * \dfrac{S_{NH}}{K_{0-AOB} + S_o} * \dfrac{S_o}{K_{NH-AOB} + S_{NH}}$
6 Denitrification						−1			$r_6 * \dfrac{K_O}{K_0 + S_o} * \dfrac{S_{NO}}{K_{NO} + S_{NO}} * \dfrac{X_{PHA}}{K_{PHA} + X_{PHA}} * \dfrac{K'_{MAX} - X_{PP}}{K'_{PP} + (K'_{MAX} - X_{PP})}$
7 Aeration (u)								$K_L a$	$(8.5 - S_O)$

The following modifications are implemented to reduce the number of parameters in the process rate kinetics.

- All the biological processes involved in the aerobic uptake of (S_F) are written as a Monod equation with a rate constant r_1 and saturation coefficients K_S and K_O.
- The maximum fermentation rate q_{fe} is lumped with X_H and represented as a single term r_2.
- The rate constant for X_{PHA} storage q_{PHA} is lumped with X_{PAO} to form a term r_3.
- The rate constant for X_{PP} storage q_{PP} is lumped with X_{PAO} to represent a single term r_4.
- The maximum growth rate of autotrophs μ_{AUT} and Y_A is lumped with X_{AUT} to represent a single term r_5. This rate equation assigned for aerobic growth of autotroph is used to explain the kinetics of ammonia nitrification.
- All the reactions pertaining to denitrification by PAO are lumped with X_{PAO} to represent a single term r_6.
- The Monod saturation coefficients associated with X_{PHA} (K_{PHA}) and X_{PP} (K_{PP}) and (K_{MAX}) was lumped with X_{PAO} to obtain the terms K'_{PP}, K'_{PHA} and K'_{MAX}, respectively.

After considering the aforementioned modifications, the rate kinetics can be transformed to the equations mentioned in Table 2. All the kinetic parameters mentioned in this model will be estimated in the model calibration stage.

State-space equations

The state-space equation denotes the rate of changes in the state variables. Equation (1) presents the model equation in the continuous-time state space form; where $x = [S_F \; S_A \; X_{PP} \; X_{PHA} \; S_{NH} \; S_{NO} \; S_{PO} \; S_O]$ are the state variables and $u = K_L a$ is the manipulated variable. The process rate equations f_c are presented as a Petersen matrix in Table 2.

$$\dot{x} = f_c(x, u) \tag{1}$$

The reduced order ASM2d model has to be calibrated before use in a specific case (García-Usach et al. 2006). The data obtained from the kinetic studies conducted in the pilot plant are used to estimate the stoichiometric and kinetic parameters of the model. The optimization procedure involving minimization of a quadratic error function (Marsili-Libelli et al. 2001; Nair et al. 2017) is used to fit the experimental data to the model. The optimization problem statement is presented in Equations (2)–(4).

$$\min_p f_{OBJ}(p) \tag{2}$$

$$f_{obj}(p) = \sum_{i=1}^{N_i} (y_i - m_i(p))^2 \tag{3}$$

$$lb \le p \le ub \tag{4}$$

'p' is a vector consisting of the model parameters that are to be estimated; the upper and lower bounds of the parameter vector is represented as ub and lb, respectively. The objective function f_{OBJ} is defined as the least-square error between the experimental values y_i = [sCOD NH_4-N NO_3-N PO_4-P] and its corresponding model predicted values $m_i(p)$ = [$(S_A + S_F)$ S_{NH} S_{NO} S_{PO}]. The half-saturation constants and stoichiometric parameters were constrained within -100% and $+150\%$ of the original ASM2d model parameters, but a wider constraints range is provided to the rate constants. The initial concentrations of S_F and X_{PP} are also taken as parameters, which would be estimated by the optimizer. The non-linear optimization solver function in MATLAB (fmincon) is used to solve the optimization problem and obtain the parameters.

Observation equations

The observation equation expresses the online measurements as functions of the state variables (see Supplementary Table 1, available with the online version of this paper). The work presented in Serralta *et al.* (2004) and Aguado *et al.* (2006) demonstrates the possibility of expressing the pH as a function of various ionic species available in the wastewater treatment system (S_A, S_{PO}, S_{NH}, S_{NO}). A generic equation, which contains a weighted combination of the linear terms, squared terms, and binary interaction terms, is used as the preliminary model structure. The mathematical form of the equation is presented in Equations (5) and (6). A similar model structure can be found in dosing control predictors, and are found to fit a wide range of data (Ratnaweera & Fettig 2015). The regression learner app in MATLAB is used to establish a correlation between the ionic species ($\theta = [S_A \; S_{NH} \; S_{NO} \; S_{PO}]$) and the data received from pH sensors and obtain the coefficients α_i β_i and $\gamma_{i,j}$. The second measurement, dissolved oxygen is equivalent to the state S_O in the model. The measurement function $h(x)$ can be expressed as Equation (7).

$$pH_{MODEL} = \alpha_0 + \sum_{i=1}^{n} \alpha_i \theta_i + \sum_{i=1}^{n} \beta_i \theta_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \gamma_{i,j} \theta_i \theta_j \tag{5}$$

$$\theta = [S_A \; S_{NH} \; S_{NO} \; S_{PO}] \tag{6}$$

$$h(x) = \begin{bmatrix} pH_{MODEL}(S_A, S_{NH}, S_{NO}, S_{PO}) \\ S_O \end{bmatrix} \tag{7}$$

State estimator

The Kalman Filter is an optimal state estimator for linear systems with Gaussian noise. However, a number of versions of original Kalman Filter algorithm such as Unscented Kalman Filter (UKF), and particle filters (PF) can be used to estimate the states of a non-linear system. In this case, the Extended Kalman Filter is used as a state estimator for the SBR. The equations of the EKF are given in Equations (8)–(15). The equations were written in the same order in which they were implemented in the PLC.

$$x_k^- = x_{k-1}^+ + T_s f_c(x_{k-1}^+, u_{k-1}) \tag{8}$$

$$y_k = h(x_k^-) \tag{9}$$

$$P_k^- = F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1} \tag{10}$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_{k-1})^{-1} \tag{11}$$

$$x_k^+ = x_k^- + K_k(z_k - y_k) \tag{12}$$

$$P_k^+ = (I - K_k H_k) P_k^- \tag{13}$$

$$F_k = I + T_s \frac{\partial f_c}{\partial x}\bigg|_{x_k^+, u_k} \tag{14}$$

$$H_k = \frac{\partial h}{\partial x}\bigg|_{x_k^-, u_k} \tag{15}$$

The discrete form of state-space equation presented in Equation (8) is derived from the continuous-time system function f_c (in Equation (1)) by an explicit Euler discretization method with a constant sampling time of T_S. f_c and h are the nonlinear state transition and measurement functions. In the equations above, x_k is the state variable at time k, z_k is the measurement, Q and R are the covariance matrices of the process and measurement noise, respectively. F is the state transition matrix, and H is the measurement matrix. I is the identity matrix, x_k^- is the *a priori* estimate of the state, x_k^+ the *a posteriori* estimate of the state, K_k the Kalman gain, P_k^- the covariance of *a priori* estimation error, and P_k^+ the covariance matrix of the *a posteriori* estimation error.

Implementing the EKF

Before implementing the estimator to the pilot plant a series of simulations are conducted to identify the appropriate tuning parameters that would result in a reasonably good correction of the states by the EKF. For this purpose, the

mathematical model and the EKF equation described in the previous section were implemented in MATLAB. The model equations are written symbolically and the Jacobian function in MATLAB's symbolic toolbox is used to linearize the process model and obtain the F and H matrix at every time step. The implementation schematic is presented in Figure 2. The block mentioned as 'simulator' consists of the mathematical model and 'plant' represents the real pilot plant. The EKF equations are written as a separate section in MATLAB, which can accept the measurement z_k and the control variable u_k from either a simulator or from the plant's SCADA system. This shift from simulator based testing to a pilot plant implementation is done with the switch S1 and S2, as indicated in Figure 2. During the tuning of the EKF parameters, the simulated state vector, x_k generated

by the model is compared to the estimated state x_k^+ predicted by the EKF to test the estimator's ability to provide a reliable estimation.

The SBR pilot plant SCADA provides an Open Platform Communication (OPC) server, which enables communication with other devices through the standard OPC protocol. A communication layer is written in MATLAB using its OPC toolbox to pull real-time data from the OPC server into the MATLAB workspace. This method enables scripts written in MATLAB to be used directly in any industrial PLC without the hassle of re-writing the code in a different programming language. While implementing the EKF in the pilot plant, a few overriding constraints are included in the estimator algorithm to prevent EKF from overshooting during the emptying and filling stages of the
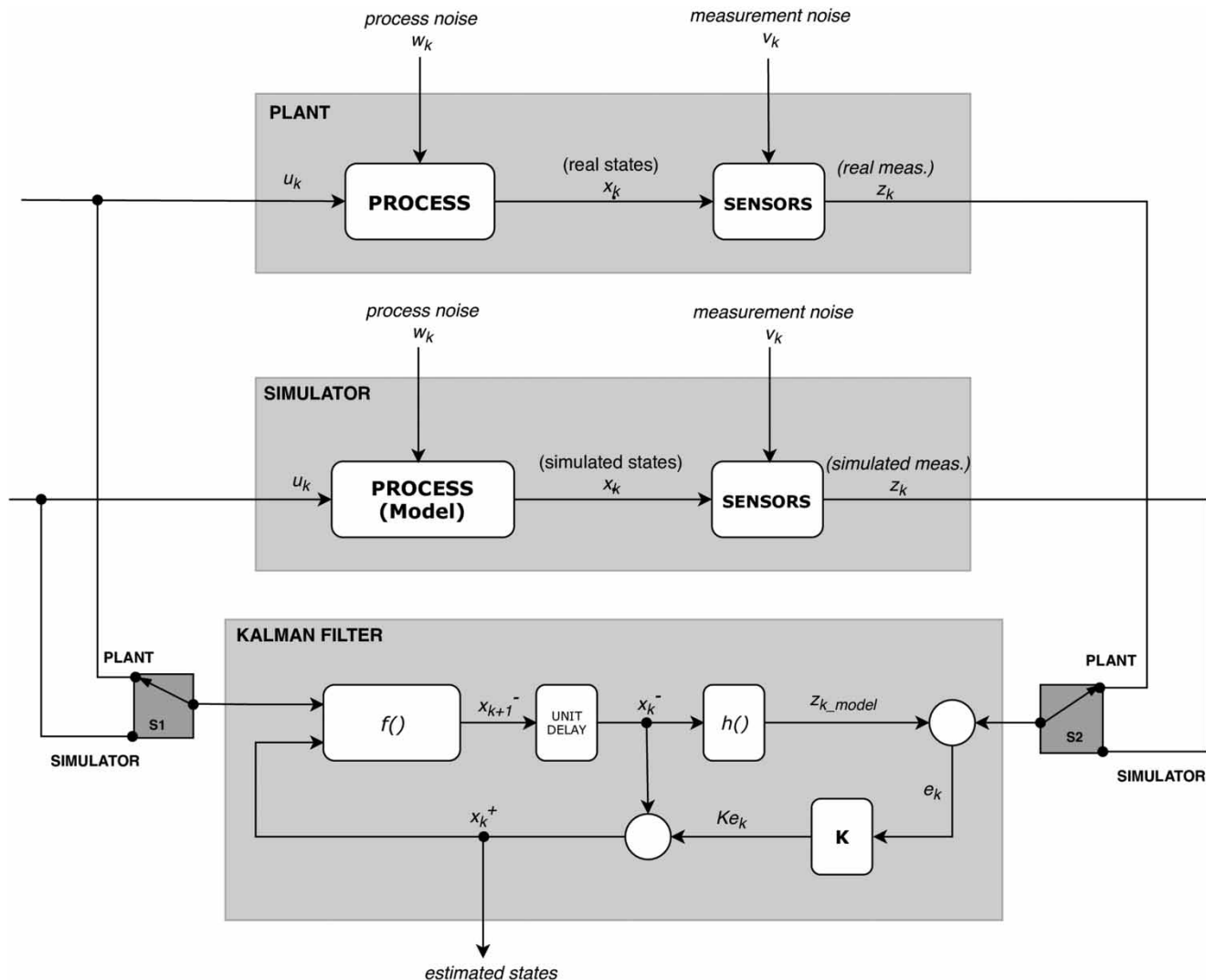

Figure 2 | Implementation schematic of EKF code in simulator and real plant.

SBR. The states and the auto-covariance matrix are reinitialized at the beginning of every cycle and the estimator is kept on hold until the anaerobic conditions begin ($DO = 0$) and the pH sensor value stabilizes. Once the EKF is started, the algorithm continues to run until the end of the aerobic stage. When the emptying stage begins, the estimated states are frozen until the next filling stage begins.

RESULTS AND DISCUSSION

Model calibration results

The results of the model calibration step, obtained by the proposed optimization procedure, are presented in Table 3. The graphs presented in Figure 3 show the comparison between the model predicted values and the data obtained from the kinetic studies conducted in the pilot plant. The figure also provides the normalized root-mean-square error (NRSME) and R^2 values obtained during the model calibration. The R^2 values presented in the graph show that the values predicted by the model are close to

the measured data collected from the pilot plant. Hence, the model can be considered to provide a good representation of the SBR process.

The optimization algorithm used for estimating the model parameters (Equations (2)–(4)) is also included in the MATLAB code. The model parameters presented in Table 3 can be updated by running the optimizer code with a new data set obtained from laboratory analysis. This strategy provides and effective platform for model calibration and effortless update of parameters whenever a drift is observed between the laboratory results and estimated values. It also generates a possibility of adapting model parameters such that the soft-sensor algorithm can be implemented in any SBR plant with similar operational sequence.

Tuning parameters of the EKF

The values of x_0^+, P_0^+, Q and R described in Equations (8)–(15) can be used as the tuning parameters of the Kalman Filter. In practical applications, these parameters are not known precisely, and therefore, trial-and-error

Table 3 | Stoichiometric and Kinetic parameters for the reduced order model

| Parameter	Parameter	Model Value	Min	Max	Units
K_{feH}	Saturation coefficient of fermentation	3.58	2	6	g COD m^{-3}
K_A	Saturation coefficient of acetate	2.20	2	6	g COD m^{-3}
K'_{PP}	Saturation coefficient of polyphosphate	10.19	1	50	g P m^{-3}
K_O	Saturation coefficient of oxygen	2.96	0.1	4	g O$_2$ m^{-3}
K'_{PHA}	Saturation coefficient of polyphosphate	119	1	200	g COD m^{-3}
K_{PS}	Saturation coefficient of phosphorus	0.2	0.1	1	g P m^{-3}
K'_{MAX}	Maximum values of X_{PP}	132	1	200	g P m^{-3}
K_{NO}	Saturation coefficient of nitrate	1.02	0.1	1.5	g N m^{-3}
K_{O_AOB}	Saturation coefficient of oxygen (nitrifiers)	1.57	0.1	4	g O$_2$ m^{-3}
K_{NH_AOB}	Saturation coefficient of ammonia (nitrifiers)	1	0.1	4	g Nm^{-3}
r_1	Rate constant for aerobic uptake of BOD	500	50	5000	g COD m^{-3}d^{-1}
r_2	Rate constant for fermentation	135	50	5000	g COD m^{-3}d^{-1}
r_3	Rate constant for PHA storage	2690	50	5000	g COD m^{-3}d^{-1}
r_4	Rate constant for PP storage	4110	50	5000	g P m^{-3}d^{-1}
r_5	Rate constant for nitrification	430	50	5000	g N m^{-3}d^{-1}
r_6	Rate constant for denitrification	310	50	5000	g N m^{-3}d^{-1}
Y_{PHA}	PHA requirement for PP storage	1.496	0.1	4.0	g COD g^{-1} P
Y_{PO}	PP requirement per PHA storage	0.577	0.1	0.6	g P g^{-1} COD
S_{F0}	Initial concentration of biodegradables	45	20	120	mg COD L^{-1}
X_{PP0}	Initial concentration of PP	65	20	120	mg P L^{-1}
k_{La}	Mass transfer coefficient for dissolved oxygen	912	100	1500	d^{-1}
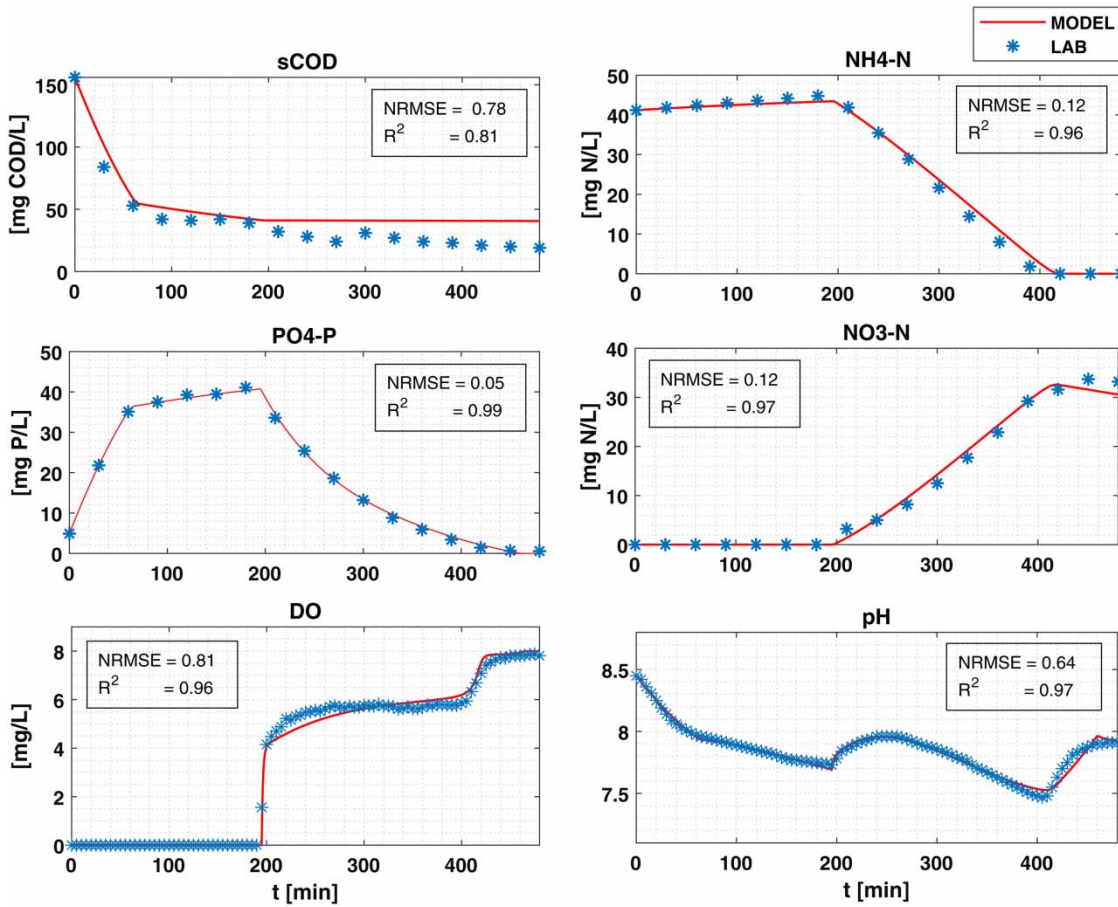
Figure 3 | Model calibration results – reduced order model versus laboratory measured values.

tuning must be expected. We have tuned the EKF as explained below, mainly following the guidelines in Haugen *et al.* (2014).

Tuning of x_0^+: A reasonable value of x_0^+ is a representative steady state, here set equal to the designed operating point of the process. For an SBR operation, we chose the design values of influent water quality as x_0^+.

$$x_0^+ = [S_{F0}\ S_{A0}\ X_{PP0}\ X_{PHA0}\ S_{NH0}\ S_{NO0}\ S_{PO0}\ S_{O0}]$$
$$= [45\ 20\ 65\ 0\ 45\ 0\ 5\ 0] \tag{16}$$

Tuning of P_0^+: It is commonly set as a diagonal matrix. The values are adjusted based on how close our initial estimates are to the real value of states. The knowledge on raw wastewater composition can help provide a good guess for the initial composition of some of the states in the process.

- The raw wastewater does not contain any nitrates, therefore the value of S_{NO} is 0 g-N m^{-3}.

- The stored PHA in the biomass usually depletes before the cycle ends, therefore we can assume that the initial concentration of $X_{PHA0} = 0$ mg COD L^{-1}.
- $S_O = 0$ mg L^{-1} since there is no dissolved oxygen in the anaerobic stage.
- The states S_F and X_{PP} do not affect the measurements (pH) in the anaerobic zone. These states exist in the model merely for the purpose of maintaining the model dynamics. A fixed initial value of $X_{PP} = 65$ mg P L^{-1} and $S_F = 45$ mg COD L^{-1} is provided to make the model fit to the kinetic data. These values should not be confused with the actual P content in the biomass or the available soluble BOD.

Therefore, the only uncertain initial estimates are S_{A0}, S_{NH0}, and S_{PO0}. The P_0^+ corresponding to the states S_{F0}, X_{PP0}, X_{PHA0}, S_{NO0}, and S_{O0} can be set to a very low value or to zero. Therefore, the P matrix at the beginning of the cycle can be expressed as the following.

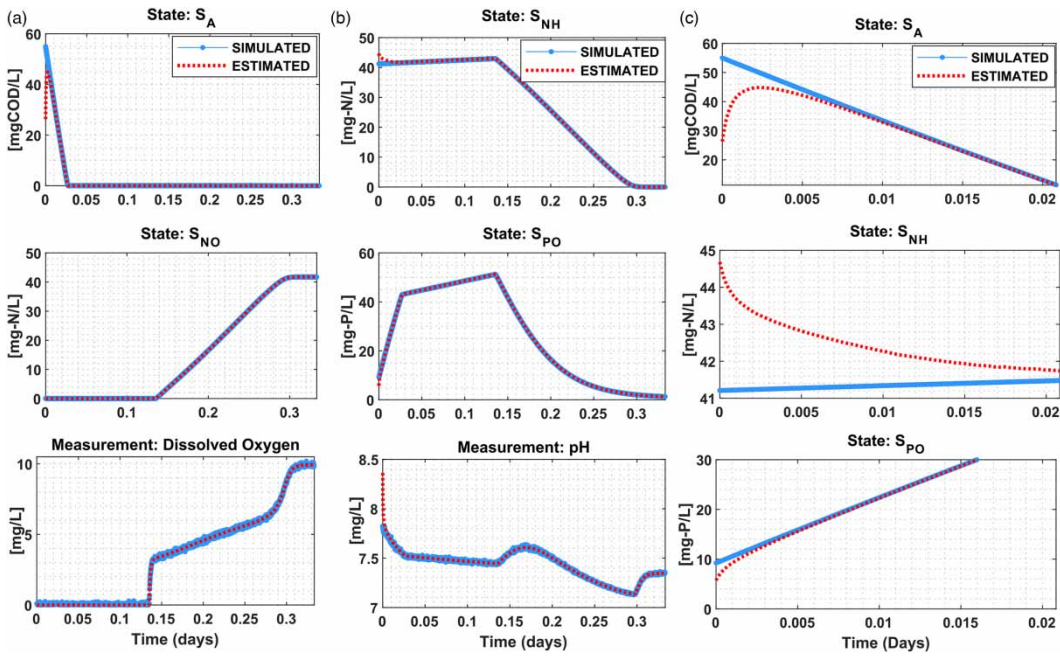$$P_0^+ = \text{diag}\ [0\ \ P_{S_A}\ \ 0\ \ 0\ \ P_{S_{NH}}\ \ 0\ \ P_{S_{PO}}\ \ 0] \tag{17}$$

Figure 4 │ (a)–(b) Simulated vs estimated values of states for a complete cycle of 8 h. (c) State-correction by EKF during the first 30 min of the cycle.
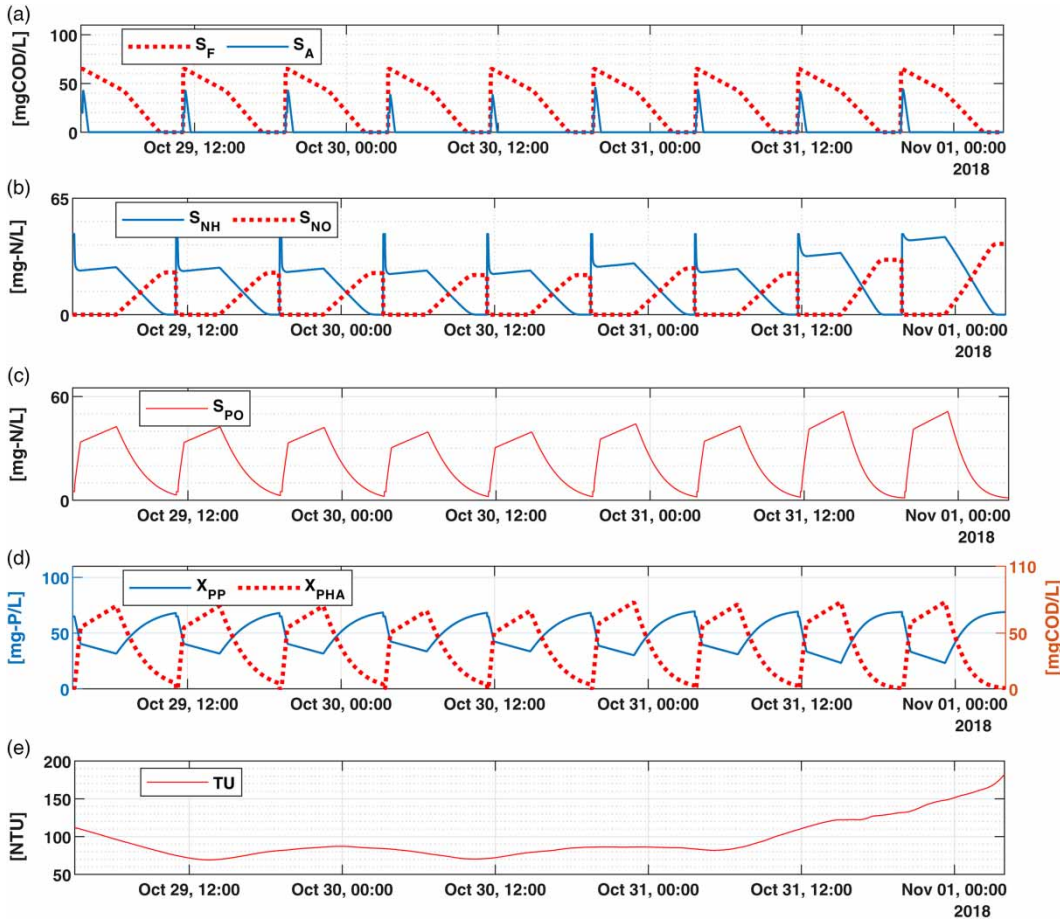

Figure 5 │ (a)–(d) Estimated states in the pilot plant for a 3-day period. (e) Turbidity measurement at influent raw water.

By trial and error, it was found that the following values provide a good estimation.

$$P_{S_A} = 316 \; ; \quad P_{S_{NH}} = 0.12; \quad P_{S_{PO}} = 0.10 \tag{18}$$

Tuning of R: It is estimated as the variance of the measurement time series collected from the SCADA system.

Tuning of Q: Typically, Q is used as the main, or final, tuning parameter. It is assumed a diagonal matrix with diagonal elements related to the initial guess of the pertinent state variable through factors l_i:

$$Q_c = \text{diag}([l_i \, x_{o,i}^+]^2) \tag{19}$$

In the simulator based testing, proper values of trial-and-error were found as

$$\{l_i\} = [1 \quad 1 \quad 1 \quad 1 \quad 2 \quad 1 \quad 1 \quad 1.5]^* 10^{-3} \tag{20}$$

Simulator test results

The EKF is updated with the tuning parameters presented in Equations (16)–(20). The results presented in Figure 4(a) and 4(b) show a comparison between the state variable obtained from the simulator and the states estimated by the tuned EKF during an 8 h cycle. The state correction by EKF is not clearly visible in Figure 4(a) and 4(b) since the estimated states usually approach the simulated states within the first 30 min of the SBR cycle. Therefore, the simulation result for the first 30 min is presented in Figure 4(c), where the state correction by the EKF is noticeable. The convergence of the estimated value to a randomly selected initial composition confirms that the EKF parameters are correctly tuned.

Pilot plant implementation results

The tuned EKF algorithm is implemented in the pilot plant and the estimator results are recorded for a period of two weeks. Figure 5 presents a subset consisting of estimated states for a duration of 3 days (nine cycles). The variations in nutrient composition during the cycle follow similar trends to those observed in a typical SBR operation, as presented in literature such as Marsili-Libelli *et al.* (2001) and Sin & Vanrolleghem (2006). Due to the high holdup in the buffer tank, the diurnal variations in influent concentration is barely noticeable. However, a significant change in influent concentration can be observed during rain or snowmelt events, causing storm-water infiltration at the raw wastewater source. These variations can be

observed in Figure 5(e), which shows the data from the online turbidity sensor installed in the raw wastewater tank. The estimation results presented in Figure 5(a)–5(d) show the increase in the estimated value of S_{NH}, S_{PO}, and S_A on 31 October 2018 around 12.00 h. This event corresponds to the influx of wastewater with a rather higher nutrient concentration at 9:00 AM on the same day. This provides a qualitative validation for the soft-sensor's ability to detect variations in influent wastewater quality (see Supplementary Table 2, available with the online version of this paper).

Further validation tests were conducted on 4 November 2018. Samples were collected every 30 min during the complete 8-h cycle and the composition of NH$_4$-N, PO$_4$-P, and NO$_3$-N are measured using standardized laboratory tests. These values are compared to the states S_{NH}, S_{PO}, and S_{NO} estimated by the EKF algorithm during the cycle. The comparison plots along with the NRSME and R^2 values are presented in Figure 6. This demonstrates the EKF's potential to estimate the states in the pilot plant.

The estimated value of the effluent water quality is also recorded for each treatment cycle. This is executed by logging the estimated value of S_{NH}, S_{PO} and S_{NO} at the end of every SBR cycle. The pilot plant achieves complete
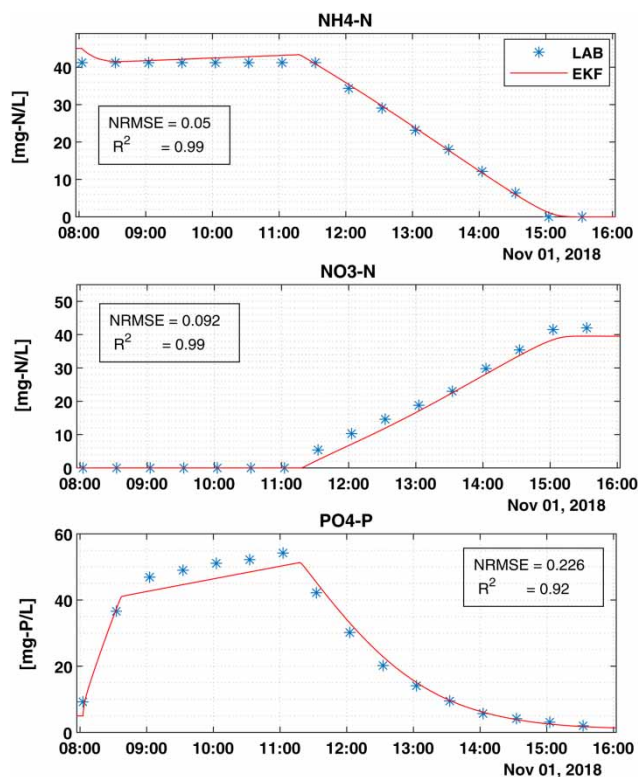

Figure 6 | Soft-sensor validation. Estimated values (continuous red line) versus laboratory measurements (blue discrete points). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wst.2019.272.
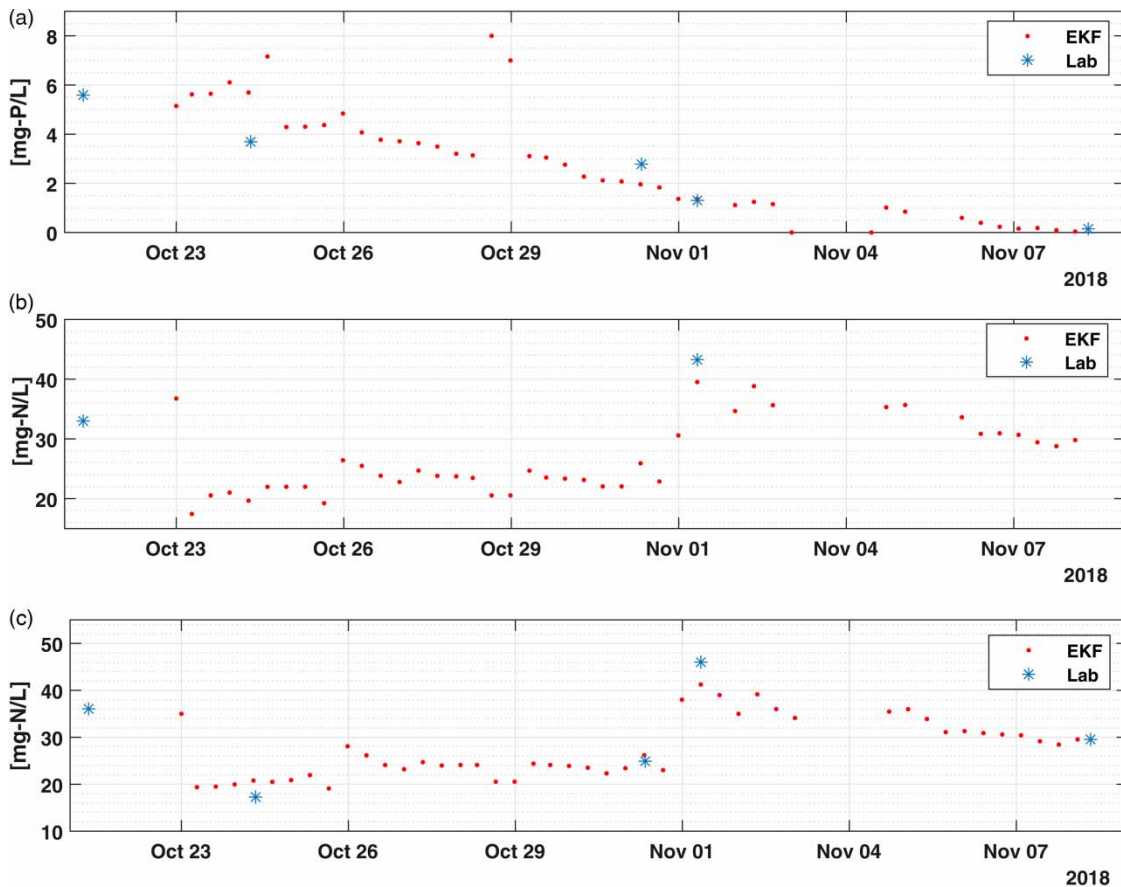
Figure 7 │ Comparison of estimated values of (a) effluent PO_4-P, (b) effluent NO_3-N and (c) influent NH_4-N composition with laboratory measurements.

nitrification, resulting in a very low effluent ammonia concentration close to zero. However, the effluent concentrations of S_{NO} and S_{PO} are rather susceptible to variations in raw wastewater quality. These variations can be observed in Figure 7(a) and 7(b).

The soft-sensor is also used to estimate the ammonia concentration in the raw wastewater. It is observed that the concentration of NH_4-N stays nearly constant during the anaerobic stage. However, it can also be observed from Figure 4(b) that the estimator converges to the real state variables within the first 30 min of the cycle time. Therefore, it would be safe to assume the value of S_{NH} at t = 30 min to be very close to the concentration of ammonia in raw wastewater. The estimated value of S_{NH} at t = 30 min for the entire testing period is presented in the third plot in Figure 7(c). These estimated values of the influent and the effluent wastewater composition is compared to the periodic laboratory measurements. Except for the erroneous estimates received on 3–5 November 2018 due to an unexpected power failure, the plots presented in Figure 7(c) demonstrate an acceptable

match between the experimental data and the values estimated by the EKF algorithm.

CONCLUSIONS

The grey box model discussed in this article explains the SBR with sufficient level of accuracy to constitute the basis of a state estimator for the reactor. The model calibration strategy discussed in this work provides the flexibility in adapting the model to any plant with similar process configuration. The validation tests conducted in the pilot plant demonstrates the possibility to estimate the concentration of VFA, NO_3-N, NH_4-N and PO_4-P in a SBR operating in alternative anaerobic–aerobic mode for the biological phosphorus removal process, by using physical sensors such as pH and DO. The EKF based soft-sensor can also provide accurate estimations of the ammonia concentration in the influent wastewater. Since the EKF algorithm is written in MATLAB's OPC compatible framework this soft-sensor code can easily be implemented in

any treatment plant which has an OPC enabled PLC, SCADA or a distributed control system (DCS), without additional hardware requirement or modifications to the MATLAB code.

The change in biomass concentration over a longer operational period could lead to minor drift in state estimation. However, periodic model calibration and parameter re-tuning can resolve this problem. The optimization code estimating the model parameters is also included as a part of the overall estimator package. This allows the model to be recalibrated with relative ease. A cost-effective method of state-estimation using soft sensors can be a prelude to implementing advanced model based control strategies in wastewater treatment and recovery processes.

REFERENCES

Aguado, D., Ferrer, A., Seco, A. & Ferrer, J. 2006 Comparison of different predictive models for nutrient estimation in a sequencing batch reactor for wastewater treatment. *Chemometrics and Intelligent Laboratory Systems* **84** (1–2), 75–81.

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.

Cadet, C. 2014 Simplifications of activated sludge model with preservation of its dynamic accuracy. *IFAC Proceedings Volumes* **19** (3), 7134–7139.

García-Usach, F., Ferrer, J., Bouzas, A. & Seco, A. 2006 Calibration and simulation of ASM2d at different temperatures in a phosphorus removal pilot plant. *Water Science and Technology* **53** (12), 199–206.

Gómez-Quintero, C., Queinnec, I. & Babary, J. P. 2000 A reduced nonlinear model of an activated sludge process. *IFAC Proceedings Volumes* **33** (10), 1001–1006.

Häck, M. & Wiese, J. 2006 Trends in instrumentation, control and automation and the consequences on urban water systems. *Water Science and Technology* **54** (11–12), 265–272.

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.

Haugen, F., Bakke, R. & Bernt, L. 2014 State estimation and model-based control of a pilot anaerobic digestion reactor. *Journal of Control Science and Engineering* **2014**, 1977.

Henze, M., Gujer, W., Mino, T., Matsou, T., Wentzel, M. C., Marais, G. V. R. & Van Loosdrecht, M. C. M. 1999 Activated sludge model NO. 2D, ASM2D. *Water Science and Technology* **39** (1), 165–182.

Jeppsson, U. & Olsson, G. 1994 Reduced order models for on-line paameter identification of the activated sludge. *Water Science and Technology* **28** (11), 173–183.

Julien, S., Barbary, P. & Lessard, P. 1999 A reduced order model for control of a single reactor activated sludge process. *Mathematical and Computer Modelling of Dynamical Systems* **5** (4), 337–350.

Li, Z., Lu, P., Zhang, D. & Zhang, T. 2018 Practical identifiability analysis and optimal experimental design for the parameter estimation of the ASM2d-Based EBPR anaerobic submodel. *Mathematical Problems in Engineering* **18**, 1–9.

Liukkonen, M., Laakso, I. & Hiltunen, Y. 2013 Advanced monitoring platform for industrial wastewater treatment: multivariable approach using the self-organizing map. *Environmental Modelling and Software* **48**, 193–201.

Marsili-Libelli, S., Ratini, P., Spagni, A. & Bortone, G. 2001 Implementation, study and calibration of a modified ASM2d for the simulation of SBR processes. *Water Science and Technology* **43** (3), 69–76.

Nair, A., Cristea, V. M., Agachi, P. Ş. & Brehar, M. 2017 Model calibration and feed-forward control of the wastewater treatment plant – case study for Cluj-Napoca WWTP. *Water and Environment Journal* **32** (2), 164–172.

O'Brien, M., Mack, J., Lennox, B., Lovett, D. & Wall, A. 2011 Model predictive control of an activated sludge process: a case study. *Control Engineering Practice* **19** (1), 54–61.

Ødegaard, H., Rusten, B. & Westrum, T. 1994 A new moving bed biofilm reactor – applications and results. *Water Science and Technology* **29** (10–11), 157–165.

Olsson, G. 2012 ICA and me – A subjective review. *Water Research* **46** (6), 1585–1624.

Olsson, G. & Jeppsson, U. 2006 Plant-wide control: dream, necessity or reality? *Water Science and Technology* **53** (3), 121–129.

Ratnaweera, H. & Fettig, J. 2015 State of the art of online monitoring and control of the coagulation process. *Water* **7** (11), 6574–6597.

Serralta, J., Ferrer, J., Borrás, L. & Seco, A. 2004 An extension of ASM2d including pH calculation. *Water Research* **38** (19), 4029–4038.

Sin, G. & Vanrolleghem, P. A. 2006 Evolution of an ASM2s-like model structure due to operational changes of an SBR process. *Water Science and Technology* **53** (12), 237–245.

Steffens, M. A., Lant, P. A. & Newell, R. B. 1997 A systematic approach for reducing complex biological wastewater treatment models. *Water Research* **31** (3), 590–606.

Thürlimann, C. M., Dürrenmatt, D. J. & Villez, K. 2018 Soft-sensing with qualitative trend analysis for wastewater treatment plant control. *Control Engineering Practice* **70**, 121–133.

Zhang, T., Zhang, D., Zhenliang, L. & Qing, C. 2010 Evaluating the structural identifiability of the parameters of the EBPR sub-model in ASM2d by the differential algebra method. *Water Research* **44** (9), 2815–2822.