Journal of Water Process Engineering 37 (2020) 101494

Contents lists available at ScienceDirect

Journal of Water Process Engineering

journal homepage: www.elsevier.com/locate/jwpe

Real-time monitoring of enhanced biological phosphorus removal in a multistage EBPR-MBBR using a soft-sensor for phosphates

Abhilash M. Nair[a,*], Blanca M. Gonzalez-Silva[b], Finn Aakre Haugen[c], Harsha Ratnaweera[a], Stein W. Østerhus[b]

[a] Faculty of Science and Technology, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway
[b] Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology (NTNU), S. P. Andersens vei 5, Trondheim 7491, Norway
[c] University of South-Eastern Norway, Kjølnes ring 56, Porsgrunn, Norway

ARTICLE INFO

Keywords:
Biological phosphorus removal
Extended Kalman Filter
MBBR process
Soft-sensor
Hybrid model

ABSTRACT

Enhanced biological phosphorus removal (EBPR) from municipal wastewater has been achieved in a multistage Moving Bed Biofilm Reactor (MBBR) configuration. The process operations can be further optimized by real-time monitoring of water quality parameters in the individual chambers of the EBPR-MBBR process. This work presents a hybrid, soft-sensor as a cost-effective monitoring option for real-time estimation of phosphates (PO_4^{3-}-P) and soluble COD (sCOD) concentrations in the anaerobic chambers of a multistage EBPR-MBBR pilot plant. The soft-sensor is developed by implementing an Extended Kalman filter on a reduced-order nutrient removal model. The hybrid model is constructed by combining mechanistic elements of phosphorus release kinetics in anaerobic conditions, and a statistical model correlating PO_4^{3-}-P and sCOD concentration with conductivity measurements. A systematic method for developing, calibrating a reduced-order model and tuning of the Kalman filter parameters have been discussed in this work. The drift in soft-sensor performance was studied and practical solutions were suggested for re-calibrating the model utilizing data from periodic lab measurements. The estimation results are successfully validated against standardized lab measurements to demonstrate the accuracy of the soft-sensing algorithm.

1. Introduction

Enhanced biological phosphorus removal (EBPR) process is a treatment configuration designed for phosphorus removal from wastewater. The EBPR process consists of an anaerobic stage before the aeration stage, where a specific species of biomass called polyphosphate-accumulating organisms (PAO) is enriched [1]. A novel configuration was reported by [13] for achieving EBPR in a continuous moving bed biofilm reactor (MBBR) process. The reactor performance is currently monitored using off-line laboratory analysis of PO_4^{3-}-P, soluble Chemical Oxygen Demand (sCOD) volatile fatty acids (VFA) ammonium (NH_4^+-N), nitrites (NO_2^--N) and nitrates (NO_3^--N). However, off-line monitoring implies a low sampling frequency, leading to potential non-optimal conditions between the sampling instances, and a subsequent delay between sampling time, hampering adjusting of the process conditions, and data availability. Manual sample collection, sample preparation, and lab analysis involve additional man-hours and costs associated with chemicals and kits required for lab analysis.

Real-time measurement of nutrient concentration can result in faster detection of process abnormalities. Online monitoring of nutrients in the plant can also enable the possibility of implementing various control strategies to ensure stable and optimal performance of the treatment process [2]. However, the unavailability of sensor probes and the high price of online nutrient analyzers often discourage their use in treatment plants [3]. Soft-sensors are a viable alternative for expensive composition analyzers or unreliable sensor probes [4]. A review on the use of data-driven soft-sensors for enhancing online monitoring in wastewater treatment operations is presented in [5]. Several case-studies on simulator-based evaluation as well as full-scale implementation of soft-sensors in wastewater treatment processes are reported in literature [6,7].

Parameters such as pH, oxidation-reduction-potential (ORP), dissolved oxygen (DO), and electrical conductivity (EC) can be measured using inexpensive and low-maintenance sensors. Conductivity measurements, which directly correlate to the ionic strength of the solution, can be used to predict the nutrient concentrations in biological

* Corresponding author.
E-mail addresses: muralabh@nmbu.no (A.M. Nair), blanca.g.silva@ntnu.no (B.M. Gonzalez-Silva), Finn.Haugen@usn.no (F.A. Haugen), harsha.ratnaweera@nmbu.no (H. Ratnaweera), stein.w.osterhus@ntnu.no (S.W. Østerhus).

https://doi.org/10.1016/j.jwpe.2020.101494
Received 11 May 2020; Received in revised form 22 June 2020; Accepted 3 July 2020
Available online 09 July 2020

Nomenclature

Variables

x_k	State variable at discrete time instance k
T_s	Time step (h)
z_k	Measurement vector
τ	Residence time in each chamber (h)
$X_{s,in}$	Particulate biodegradable in influent ($g\ COD.\ m^{-3}$)
$X_{s,i}$	Particulate biodegradable in chamber i ($g\ COD.\ m^{-3}$)
$S_{s,in}$	Soluble biodegradable in influent ($g\ COD.\ m^{-3}$)
$S_{s,i}$	Soluble biodegradable in chamber i ($g\ COD.\ m^{-3}$)
$S_{po,in}$	Soluble ortho-phosphates in influent ($g\ COD.\ m^{-3}$)
$S_{po,i}$	Soluble ortho-phosphates in chamber i ($g\ P.\ m^{-3}$)
C_{in}	Conductivity measurements in the influent ($\mu S/cm$)
C_i	Conductivity measurements in the chamber i ($\mu S/cm$)
f	Nonlinear state transition function
h	Nonlinear measurement function
K_H	Saturation coefficient of fermentation ($g\ COD.\ m^{-3}$)
K_S	Saturation coefficient of PHA storage ($g\ COD.\ m^{-3}$)
r_1	Rate constant for hydrolysis ($g\ P.\ m^{-3}h^{-1}$)
r_2	Rate constant for PHA storage ($g\ COD.\ m^{-3}h^{-1}$)
θ_1	Temperature coefficient for r_1
θ_2	Temperature coefficient for r_2
Y_{PO}	Yield coefficient for PO release ($g\ P/g\ COD$)
α_i	Regression parameters
f_{OBJ}	Objective Function
$\mathcal{N}$	Number of datapoints
φ_{exp}	Experimental Data
φ_{model}	Model Predicted Values
$\mathcal{N}$	Number of points in the dataset
F	Linearized state transition matrix
H	Linearized measurement vector
P_0^+	Initial estimate of Autocovariance matrix
x_0^+	Initial estimate of EKF
R	Covariance matrix of measurement noise
Q	Covariance matrix of process noise

wastewater treatment processes [8]. Both mechanistic [9,10], as well as data-driven models [11], have been used to demonstrate a positive correlation between conductivity measurements and the concentration of PO_4^{3-}-P as well as acetates (which can be one of the components of sCOD) concentrations in biological nutrient removal process. However, most work has been done either in a lab-scale using synthetic wastewater or in a sequential batch reactor [12]. We have not yet found reports of its implementation in a continuous treatment process with unmeasured variations in the influent wastewater quality.

The multistage MBBR is a novel process, designed for EBPR from municipal wastewater [13]. Several lab-scale studies have established a correlation between the anaerobic phase-length and the overall PO_4^{3-}-P removal during EBPR process. Control strategies aimed at manipulating the anaerobic phase-length in pilot plant or the dosage of external carbon source requires real-time monitoring of water quality parameters in the multistage MBBR process. The unfeasibility of installing nutrient sensors in each chamber establishes the importance of a cost-effective online monitoring system in the pilot plant. This work attempts to enhance the online monitoring system in a multistage EBPR-MBBR process [13] by implementing a conductivity based soft-sensor for real-time estimation of water quality parameters such as PO_4^{3-}-P and sCOD in their anaerobic stages.

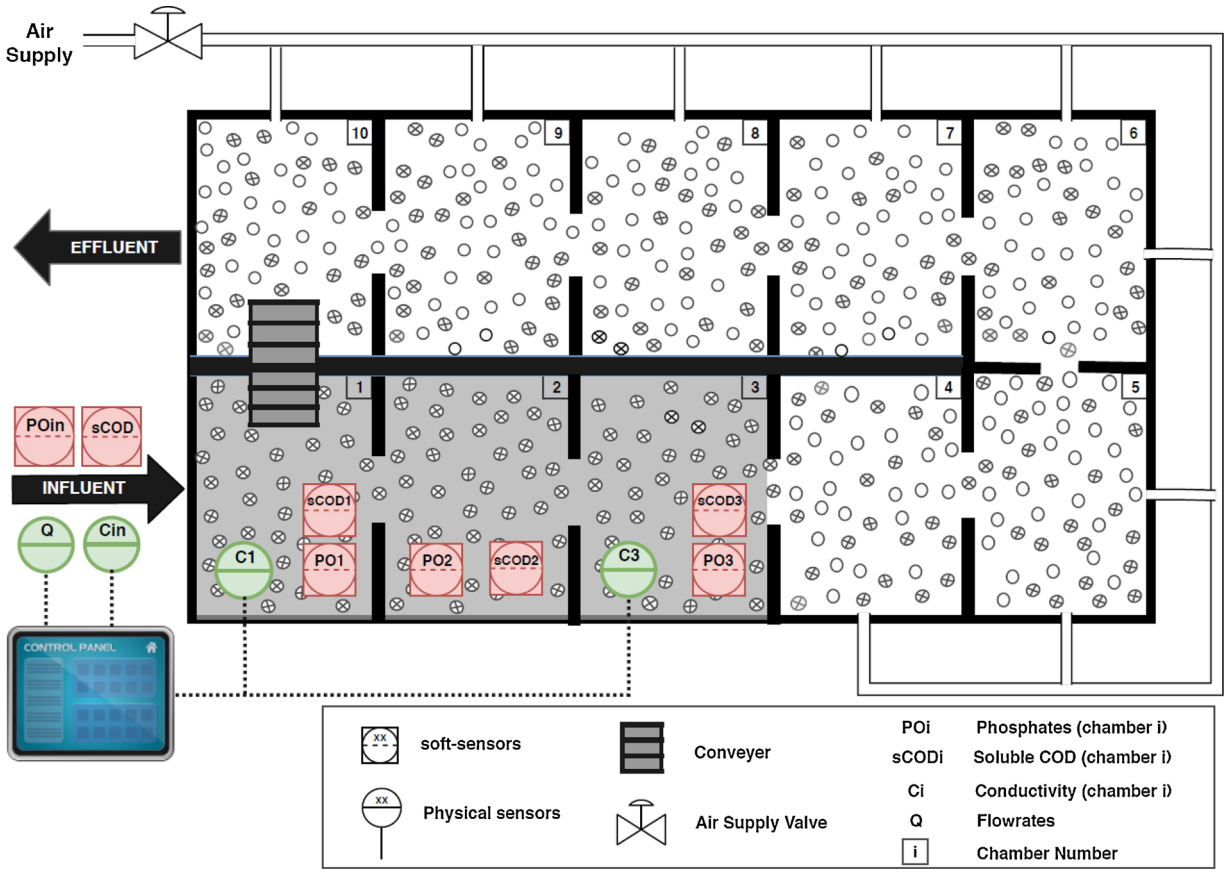

Fig. 1. Process flow diagram (top view) and the sensor network installed in the pilot plant. (Figure modified from [13]).

2. Materials and methods

2.1. Pilot plant setup and operational data

The pilot plant located in the wastewater lab at the Norwegian University of Science and Technology (NTNU) was used to test and validate the phosphate soft-sensor [13]. The reactor has a total working volume of 1 m^3, divided into 10 chambers of equal volume, operating with a hydraulic retention time (HRT) of 6.5 h (Fig. 1). The pilot plant contains the standard biofilm carriers (Kaldnes K1) with 60 % filling in each of the chambers. The opening in the separation walls of the chamber allows the flow of wastewater and carriers between different zones. Mechanical agitators ensure sufficient mixing of the carriers in the anaerobic zones. The first three chambers of the pilot plant operate as anaerobic reactors, and the subsequent chambers are aerated. The carriers from the last aerobic zone (Chamber 10) are transported to the first anaerobic zone (Chamber 1) by a conveyor belt. The pilot plant layout, online monitoring, and the data acquisition system are presented in Fig. 1. The pilot plant is fed continuously with wastewater (3.6 m^3/day) from a storage tank which receives fresh wastewater from a nearby sewer once every hour. The raw wastewater undergoes primary treatment, where it is passed through a fine sieve to remove larger particles, before sending it to the biological stage (MBBR).

The pilot plant is equipped with a state-of-the-art online monitoring system with remote data access capabilities. It has a conductivity sensor in chamber 1, chamber 3, and the influent (Fig. 1). All the online sensors are connected to a Supervisory Control And Data Acquisition System (SCADA) provided by DOSCON AS.

The mean values of sCOD and PO_4^{3-}-P in the influent and effluent streams and the average removal (%) in the pilot plant are presented in Table 1. The multistage MBBR can operate with a flexible anaerobic/aerobic phase length by controlling the aeration in Chamber 2 and 3 of the pilot plant. Monitoring PO_4^{3-}-P concentrations in chamber 1–3 provides real-time information on the phosphorus released in the anaerobic phase. Real-time control of anaerobic phase length based on feedback from the PO_4^{3-}-P soft-sensors can provide an indirect way of controlling the percentage of PO_4^{3-}-P removed by the pilot plant.

2.2. Mathematical model

Various models explaining biological nutrient removal exist in the literature [14]. These models vary from comprehensive mechanistic models such as ASM 2d [15] to data-driven statistical models [16]. Literature also mentions the use of a reduced-order version of the comprehensive ASM models for state estimation [17,18]. Mathematical models combining the mechanistic understanding of the nutrient removal process with a statistical correlation between nutrient composition and physical parameters such as pH, DO, or conductivity can also be used to develop soft-sensors for biological nutrient removal processes [19].

2.2.1. Reduced-order PO_4^{3-}-P release model

In this work, a simplified version of the phosphorus release model was implemented. The kinetic equations explaining PO_4^{3-}-P release during the anaerobic phase of the EBPR can be found in literature [15]. These equations were used as a basis to develop the mathematical model for the anaerobic chamber of the multistage MBBR pilot plant. The sub-model includes the following three processes occurring in parallel.

a) Hydrolysis of particulate biodegradable COD (X_s) to soluble biodegradable COD (S_s).
b) Uptake of soluble biodegradable COD (S_s) to store poly-hydroxy-alkanoates (PHA) in the biomass.
c) Consumption of stored polyphosphates (PP) in the biomass and simultaneous release of soluble ortho-phosphates (S_{PO}).

The original ASM2d model with 19 states and 21 processes poses a significant challenge in their utility as a model for soft-sensing. The usual strategy is to reduce the number of states in the original AMS2d model by either assuming the states with relatively slower dynamics as constants or by eliminating processes with a relatively insignificant effect on rate-kinetics. Several examples regarding the use of reduced-order sub-models for soft-sensing are reported in literature [20,21]. The model reduction strategies discussed in [22–24] were used to make the following assumptions.

a) The growth and decay kinetics of biomass are ignored. Hence, the model only considers the kinetics of substrate degradation processes.
b) The release of ammonium (S_{NH}) due to ammonification is considered insignificant in the anaerobic chambers, and hence it is ignored.
c) The release of soluble ortho-phosphates (S_{PO}) during hydrolysis of particulate biodegradable COD (X_s) is ignored since it is negligible compared to (S_{PO}) released due to the consumption of stored polyphosphates (PP).
d) The rate of storage of PHA would not be affected by PHA concentration in the biomass.
e) The temperature dependency of the rate constant for hydrolysis (r_1) and rate constant for PHA storage (r_2) follow the exponential term with coefficients (θ_1) and (θ_2).

The schematics of the process equations and the rate-kinetics occurring during the anaerobic stages of the biological phosphorus removal, and the bench-scale setup used to study the reaction kinetics are presented in Fig. 2.

2.2.2. State-space equations

Every chamber in the multistage MBBR system is modelled as a continuous-stirred tank reactor (CSTR). The mass balances and rate kinetics are adapted for all three anaerobic chambers of the CSTR. The influent concentrations of particulate COD ($X_{s,in}$), soluble COD ($S_{s,in}$), and phosphates ($S_{po,in}$) are modelled as 'random-walk' and included in the state-vector [29]. The augmented state variables, to be estimated by the soft-sensor, is presented in Eq. 1. The Monod's rate-kinetics in the continuous state-space form is presented in Eq. 2. The discrete state-space form of the model (Eq. 5) is obtained by discretizing the continuous-time model using an explicit Euler forward method with time-step T_S.

$$x = [X_{s,in}\ S_{s,in}\ S_{po,in}\ X_{s,1}\ S_{s,1}\ S_{po,1}\ X_{s,2}\ S_{s,2}\ S_{po,2}\ X_{s,3}\ S_{s,3}\ S_{po,3}\ ] \tag{1}$$

$$\frac{dx}{dt} = f_i = \begin{cases} 0 & i = 1,\ 2,\ 3 \\ \tau^{-1}\,(x(i-3) - x(i)) - r_1 \dfrac{x(i)}{K_H + x(i)}\,\theta_1^{T-20} & i = 4,\ 7 \\ \tau^{-1}\,(x(i-3) - x(i)) + r_1 \dfrac{x(i-1)}{K_H + x(i-1)}\theta_1^{T-20} - r_2 \dfrac{x_k(i)}{K_S + x(i)}\,\theta_2^{T-20} & i = 5,\ 8 \\ \tau^{-1}\,(x(i-3) - x(i)) + Y_{PO}\, r_2 \dfrac{x(i)}{K_S + x(i)}\,\theta_2^{T-20} & i = 6,\ 9 \end{cases} \tag{2}$$

Table 1
Influent and effluent wastewater quality to the pilot plant and average removal percentages.

Parameter	Influent			Effluent			Average Removal (%)
	mean	min	max	mean	min	max	
sCOD	155.4	32	357.1	30.9	17.1	62	80.1
PO_4^{3-}-P	4.3	0.5	7.6	1.01	0.096	2.41	76.1

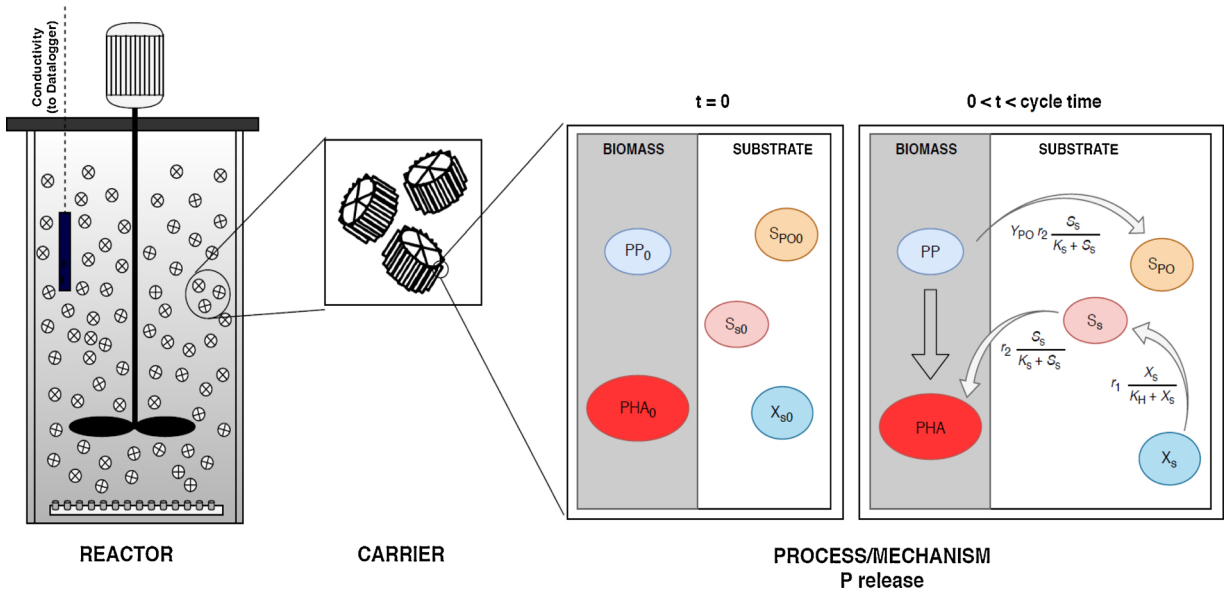

Fig. 2. Setup for batch kinetic experiment and associated mechanism for phosphorus release during the anaerobic phase.

Table 2
Stoichiometric and Kinetic parameters for the reduced order model.

Parameter	Description	Value	Min	Max	Units
K_H	Saturation coefficient of hydrolysis	0.10	0.01	6	g COD m^{-3}
K_S	Saturation coefficient of PHA storage	5.548	0.1	50	g COD m^{-3}
r_1	Rate constant for hydrolysis	205	1	2000	g P m^{-3} h^{-1}
r_2	Rate constant for PHA storage	421	1	2000	g COD m^{-3} h^{-1}
θ_1	Temperature coefficient for r_1	1.093	1	5	
θ_2	Temperature coefficient for r_2	1.047	1	5	
Y_{PO}	Yield coefficient for PO_4^{3-}-P	0.064	0	1	g P/ g COD

$$y_k = h(x(3i-1), x(3i)) \quad i = 1, 3, 5 \tag{3}$$

$$z_k = [C_{in}\ C_1\ C_3] \tag{4}$$

$$x_k = x_{k-1} + T_S f(x_{k-1}) \tag{5}$$

In Eqs. (1)–(5), k is the discrete-time index and τ is the residence time in each chamber. x_k is the state-vector at time instance k. The variables $X_{s,j}$, $S_{s,j}$, and $S_{po,j}$ correspond to the particulate biodegradable, soluble biodegradable and soluble ortho-phosphates in chamber j (j = 1, 2, 3). f and h are the nonlinear state transition and measurement function respectively. z_k is the measurement vector, and y_k is the model predicted value of the conductivity measurement at time instance k. The kinetic parameters K_H, K_S, r_1, r_2, θ_1, θ_2, and the stoichiometric parameters Y_{PO} (explained in Table 2) are determined by fitting the model to the data obtained from performing kinetic tests in a batch scale reactor.

2.2.3. Measurement equation

The data from the batch kinetic experiments, as well as results presented in literature [9,10], indicate a clear relationship between electrical conductivity and various nutrient compositions present in wastewater. The mathematical equations to quantitatively express conductivity as a function of ionic species in the wastewater had been previously reported [8,25]. However, the simplified model used in this work has just two species, represented as state variables S_s and S_{po} influencing the changes in electrical conductivity. Therefore, we use a

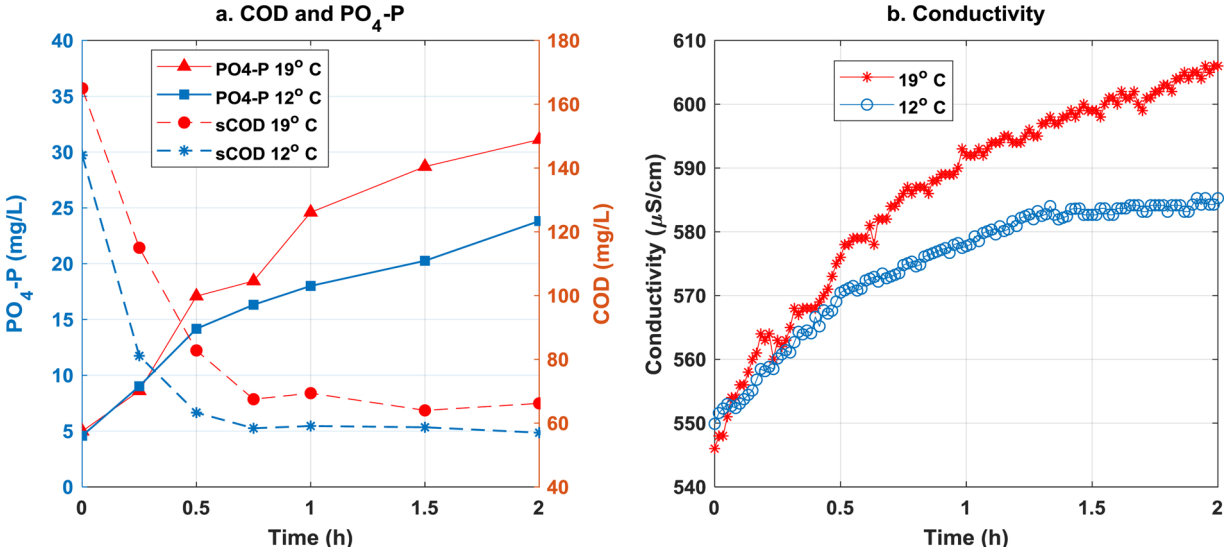

Fig. 3. Variations in (a) PO_4^{3-}-P and sCOD (b) Conductivity during the batch tests.

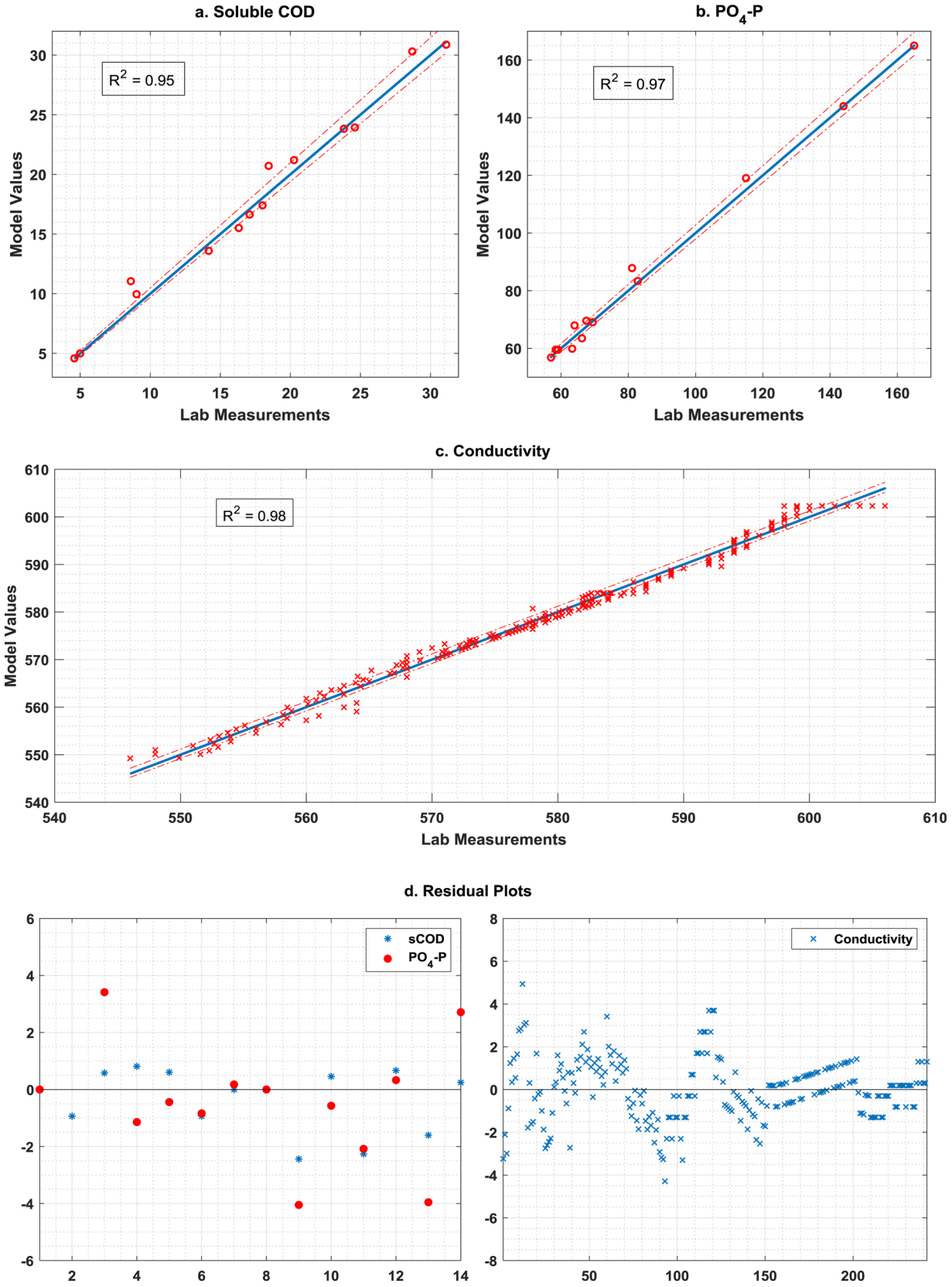

Fig. 4. Regression plot experimental results versus model predicted values (a) soluble COD, (b) soluble $PO_4^{3-}P$, (c) Conductivity, and (d) Residual plots.

statistical method to regress the data obtained from the kinetic tests to a generic equation form which includes higher powers and binary interaction terms of the predictors (S_s and S_{po}). The use of similar model forms to correlate water quality parameters have been previously reported in literature [33].

$$h(S_s,\ S_{po}) = C = \alpha_0 + \alpha_1 S_s + \alpha_2 S_{po} + \alpha_3 {S_s}^2 + \alpha_4 {S_{po}}^2 + \alpha_5 S_s S_{po} \quad (6)$$

The Regression Learner Toolbox in MATLAB was used to identify a

Table 3
Regression parameters for the measurement model.

Parameter	α_0	α_1	α_2	α_3	α_4	α_5
Values	539	−0.0031	1.91	0.000610	0.00141	−0.0220
Confidence Interval	± 14.8	± 0.0005	± 0.12	± 0.00007	± 0.0008	± 0.0014
p-Value	1.28e-11	3.63e-4	8.43e-8	4.31e-5	2.17e-16	1.81e-7

correlation between the species S_{po}, S_s and the data received from the conductivity sensor installed in the experimental unit used for the kinetic study. The final model form (Eq. 6) was derived by assessing the significance of each regression coefficients using t-statistic and selecting the parameters that significant at 5% significance level.

2.3. Batch experiments and model calibration

To evaluate the biomass kinetics and to estimate the kinetic and stoichiometric coefficients of the reduced-order model, a series of batch experiments were conducted. To perform the experiment, the carriers colonized with PAOs were taken from the MBBR pilot and filled in a 1 L beaker at 60 % filling degree and raw wastewater. Nitrogen gas was initially bubbled through the reactor to remove the oxygen from the wastewater and ensure perfect anaerobic conditions. The beaker was continuously mixed using a mechanical stirrer. The batch reactor was equipped with an online conductivity sensor for continuous monitoring of electrical conductivity during the cycle. A data logger was connected to the conductivity sensor and was configured to log the conductivity measurement every 1 min during the entire duration of the batch test. The breaker was sealed to maintain anaerobic conditions during the test. The schematic of the test beaker is presented in Fig. 2. Samples for analysis of sCOD and PO_4^{3-}-P were taken every 15 min for an entire test period of 2 h. The water quality parameters were measured using Dr. Lange cuvettes (LCK 350 and LCK 348 for PO_4^{3-}-P and LCI400 and LCK 314 for COD) and HACH LANGE GmbH (DR 1900, China) spectrophotometer.

The data obtained from the batch experiment was fit to the mathematical model explained in Eq. (1) to Eq. (6). The optimization procedure, using the minimization of quadratic-error function [26], was used to fit the model and obtain the kinetic parameters. The optimization problem is expressed in Eqs. 7 and 8.

$$\min_{p}\left\{\sum_{i=1}^{\mathcal{N}}[\varphi_{exp}-\varphi_{model}(p)]^2\right\} \tag{7}$$

$$lb \le p \le ub \tag{8}$$

'*p*' is a vector consisting of the model parameters that are to be estimated. The terms *ub* and *lb* are the upper and lower bounds of the parameter vector. $\mathcal{N}$ is the number of datapoints obtained from the kinetic experiment. The objective function is defined as the least-square error between the experimental values $\varphi_{exp} = [sCOD PO_4^{3-}\text{–P}]$, and their corresponding model predicted values $\varphi_{model}(p) = [S_s S_{po}]$. The *lb* and *ub* of the parameter vector *p*, along with the results of the optimization problem are presented in Table 2. Due to the lack of an appropriate initial guess for the parameters, a high range was provided to *lb* and *ub* to ensure a wider search area for the optimization algorithm. The optimization problem was solved with MATLAB. The fmincon function using the interior-point algorithm was combined with the grid search method to solve the optimization problem and obtain the model parameters.

2.4. Soft-sensor algorithms

The Kalman Filter is a commonly used algorithm for estimating the state variables in linear systems. Several versions of the original Kalman Filter algorithm can be used to estimate the states of a nonlinear system. The Extended Kalman Filter (EKF) computes the Kalman gain by using a linearized version of the nonlinear functions describing the state-space as well as measurement equations. Eqs. (9) – (15) describe the EKF algorithm used for estimating the states of the MBBR system. The equations are written in the order they have been programmed in Python.

$$F_{k-1} = I + T_s \frac{\partial f}{\partial x}\bigg|_{x_{k-1}^{+}, u_k} \tag{9}$$

$$x_k^{-} = x_{k-1}^{+} + T_s f(x_{k-1}^{+}, u_{k-1}) \tag{10}$$

$$P_k^{-} = F_{k-1} P_{k-1}^{+} F_{k-1}^{T} + Q_{k-1} \tag{11}$$

$$H_k = \frac{\partial h}{\partial x}\bigg|_{x_k^{-}, u_k} \tag{12}$$

$$K_k = P_k^{-} H_k^{T} (H_k P_k^{-} H_k^{T} + R_{k-1})^{-1} \tag{13}$$

$$x_k^{+} = x_k^{-} + K(z_k - h(x_k^{-})) \tag{14}$$

$$P_k^{+} = (I - K_k H_k) P_k^{-} \tag{15}$$

In the equations above, x_k is the state variable at time instance k, z_k is the measurement vector, Q and R are the covariance matrices of the process and measurement noise, respectively. f and h are the nonlinear state space and measurement functions. F is the linearized state transition matrix, and H is the linearized 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.

2.5. Implementation in pilot plant

The pilot plant SCADA system provides access to real-time data from the online sensors through a remote SQL server. The mathematical model (Eqs. (1)–(6)) and the EKF algorithm (Eqs. (9)–(15)) are written as a Python script and can be executed on a single board computer (Raspberry Pi 3B+). This open-source, non-intrusive, remote deployment strategy [27] is used to implement the soft-sensor algorithm in the pilot plant. The Python script was used for simulator-based testing, tuning of the EKF parameter, and the real-time deployment of the soft-sensing algorithm.

3. Results and discussion

3.1. Model calibration results

The results obtained from the batch experiments are presented in Fig. 3. The plots indicate an increase in conductivity measurements when the PO_4^{3-}-P ions are released and sCOD is consumed under anaerobic conditions, These observations are consistent with the results obtained from similar experiments [12,25,8].

The results from the parameter estimation exercise, showing a comparison between model-predicted and experimental values followed by their residual plots are presented in Fig. 4. The R^2 values of 0.96 and 0.97 indicate a good correlation between the experimental data and the model. The stoichiometric and kinetic parameters obtained by implementing the parameter estimation algorithm (Eqs.

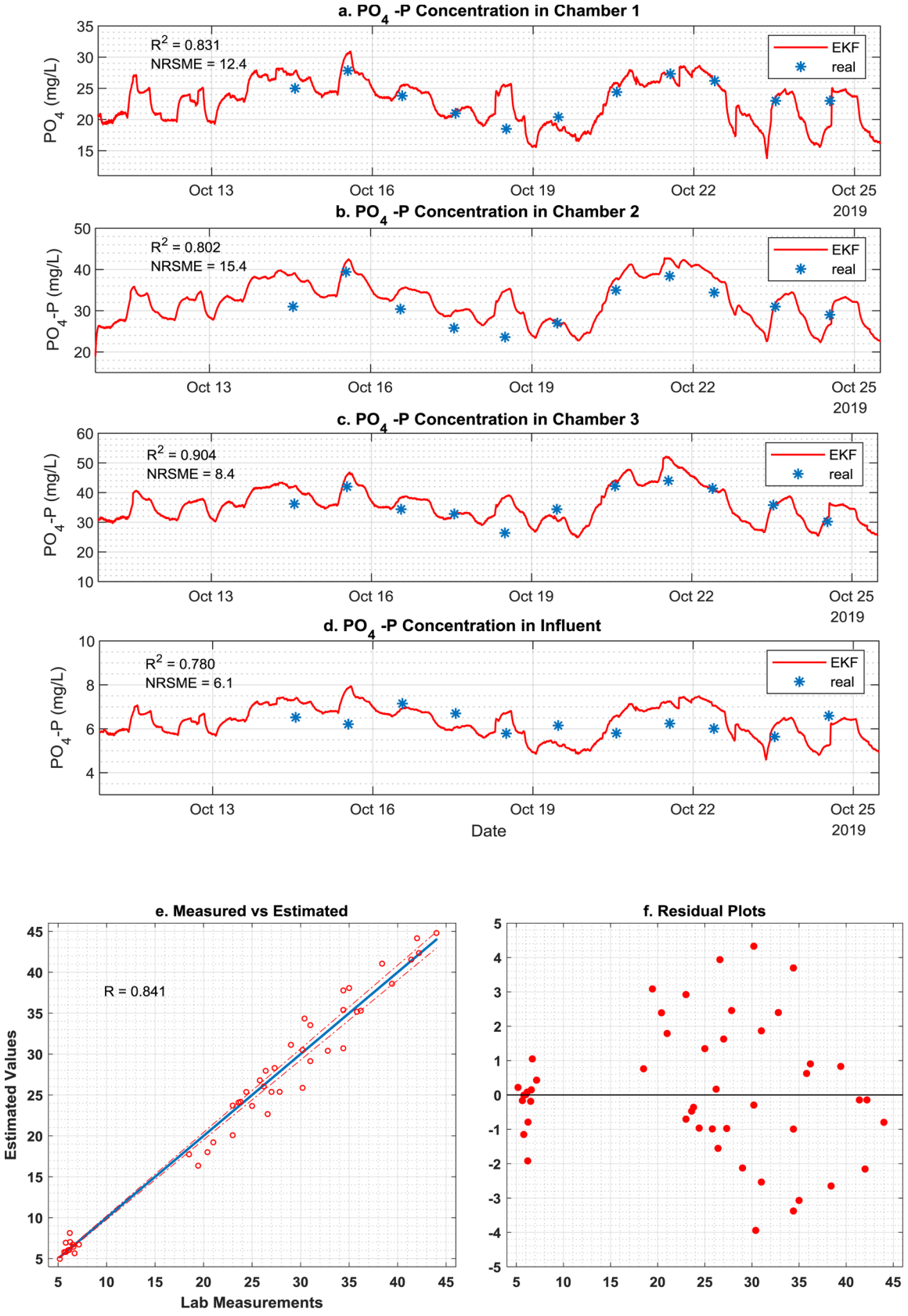

Fig. 5. Soft-sensor validation for phosphates (a) Chamber 1 (b) Chamber 2 (c) Chamber 3 (d) Influent (e) Measured versus estimated values (f) Residual plots.

(7)–(8)) on the data obtained from the batch experiments are presented in Table 2. A good match between experimental versus model predicted values and the corresponding R values close to 1 and a random distribution of the errors along the zero-error line (presented in the residual plot) validates the accuracy of the model.

The coefficients of the measurement model (Eq. 6) obtained by

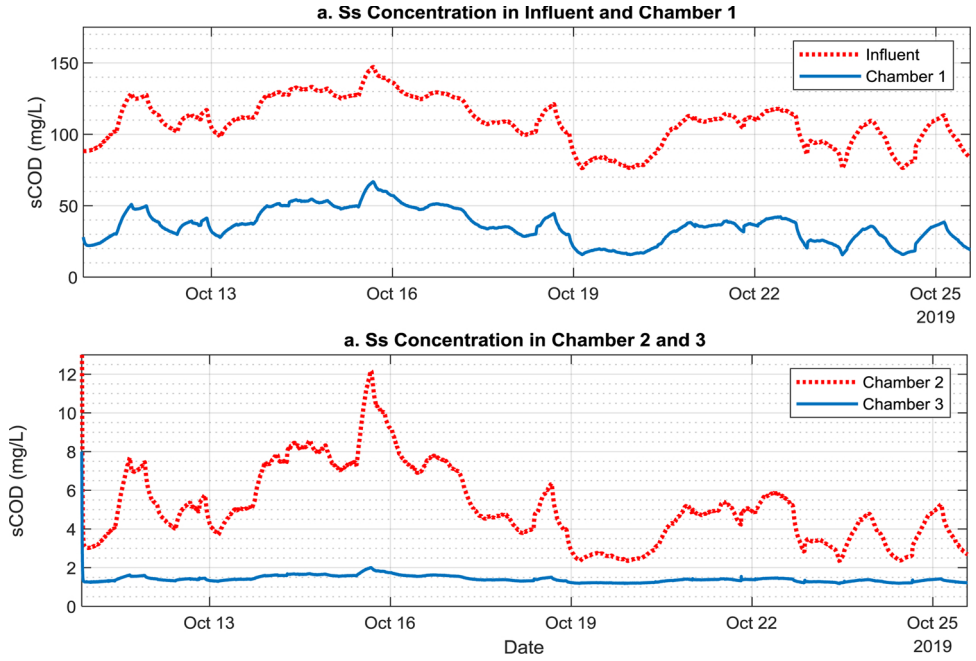

Fig. 6. Soft sensing results for sCOD concentration in the anaerobic chambers.

correlating the ionic concentration to the conductivity measurement, using the Regression Learner toolbox is presented in Table 3. Table 3. also presents the results of significance tests for all the estimated parameters obtained by conducting the t-statistics on the model. The p-value for the t-statistic of the hypothesis test indicates that all the parameters are significant within the 5% significance interval.

3.2. Tuning the EKF parameters

The tuning parameters of Kalman filter x_0^+, P_0^+, Q, and R described in Eq. (9) to Eq. (15) have to be obtained before using it to estimate the concentration of phosphates (PO_4^{3-}-P) and sCOD in the pilot plant. Some guidelines for tuning EKF parameters have been suggested in literature [28]. However, in practical applications, a little trial-and-error combined with some operational knowledge of the process would be used to tune these parameters. The guidelines mentioned in [29] have been used as a base for the simulator-based testing of the EKF and to obtain the tuning parameters.

Tuning of x_0^+ : A reasonable initial guess of x_0^+ is its steady-state values. In our case, this value of x_0^+ is set equal to the designed steady-state values of the state variables in each chamber.

$$x_0^+ = [230 \quad 150 \quad 4.4 \quad 169 \quad 54.9 \quad 14.0 \quad 109 \quad 9.1 \quad 20.6 \quad 48.9 \quad 3.4 \quad 30.1] \tag{16}$$

Tuning of P_0^+: The initial guess for the apriori estimation error is an $n\,x\,n$ diagonal matrix. The values are decided based on the confidence levels in the initial estimation of state variables.

$$P_0^+ = \text{diag}\,[k_i\, x_i\,] \tag{17}$$

By trial and error, it was found that the following values provide a good estimation, where the estimated states approach the simulated state within a reasonable short time-span (30 min in our case).

$$k_i = [15 \quad 5.1 \quad 6 \quad 2 \quad 1 \quad 0.5 \quad 3 \quad 2 \quad 1.1 \quad 2 \quad 1 \quad 0.5] * 10^{-3} \tag{18}$$

Tuning of R: A gaussian noise with a standard deviation of 0.01 was observed in the measurements provided by conductivity sensors. Therefore, the measurement noise covariance matrix R was calculated as the square of the standard deviation of the noise observed in conductivity sensors installed in chamber 1, 3, and the influent.

$$R = [0.01^2\ 0.01^2\ 0.01^2] = [1 \quad 1 \quad 1] * 10^{-4} \tag{19}$$

Tuning of Q: The process noise covariance matrix, Q, is assumed to be a diagonal matrix, with diagonal elements related to the initial guess of the pertinent state variable scaled through the factors l_i:

$$Q = \text{diag}([l_i\, x_{o,i}^+]^2) \tag{20}$$

In the simulator-based testing, values of l_i are adjusted so that the estimated values converge faster (within 30 min) to the assumed true value and at the same time do not exhibit too much noise in the estimated values. By trial and error, the values were found to be as follows.

$$\{l_i\} = [1.2 \quad 0.21 \quad 1.4 \quad 2 \quad 1 \quad 3.7 \quad 1 \quad 1 \quad 1.6 \quad 2 \quad 2.1 \quad 3.3] * 10^{-3} \tag{21}$$

3.3. Validation in pilot plant

The soft sensor was implemented in the pilot plant, and the estimation results for phosphate concentration in the anaerobic chamber and the influent are observed for a period of two weeks. During the evaluation period, one grab sample was collected every day and analyzed for PO_4^{3-}-P concentration. Fig. 5 presents a comparison between the phosphate concentration estimated by the soft-sensor and the values measured using standardized lab tests. The comparisons are provided for the influent and all three anaerobic chambers. The normalized root-mean-square error (NRSME) and R^2 values are also presented in Fig. 5. The R^2 value above 0.8 even with a limited data series demonstrates the EKF's potential to estimate the phosphate concentration in the pilot plant.

The soft-sensor can also potentially estimate the sCOD in the system. The estimation results for sCOD in each chamber are presented in Fig. 6. Although no rigorous validation tests (similar to the PO_4^{3-}-P measurement) were conducted to assess the accuracy of sCOD estimations in each anaerobic chamber, a comparison between the influent sCOD estimation and the corresponding sCOD lab measurements presented in Fig. 7b demonstrates their potential in estimating influent sCOD concentrations. However, more rigorous validations have to be conducted to confirm their ability to estimate sCOD concentrations in all the anaerobic chambers.

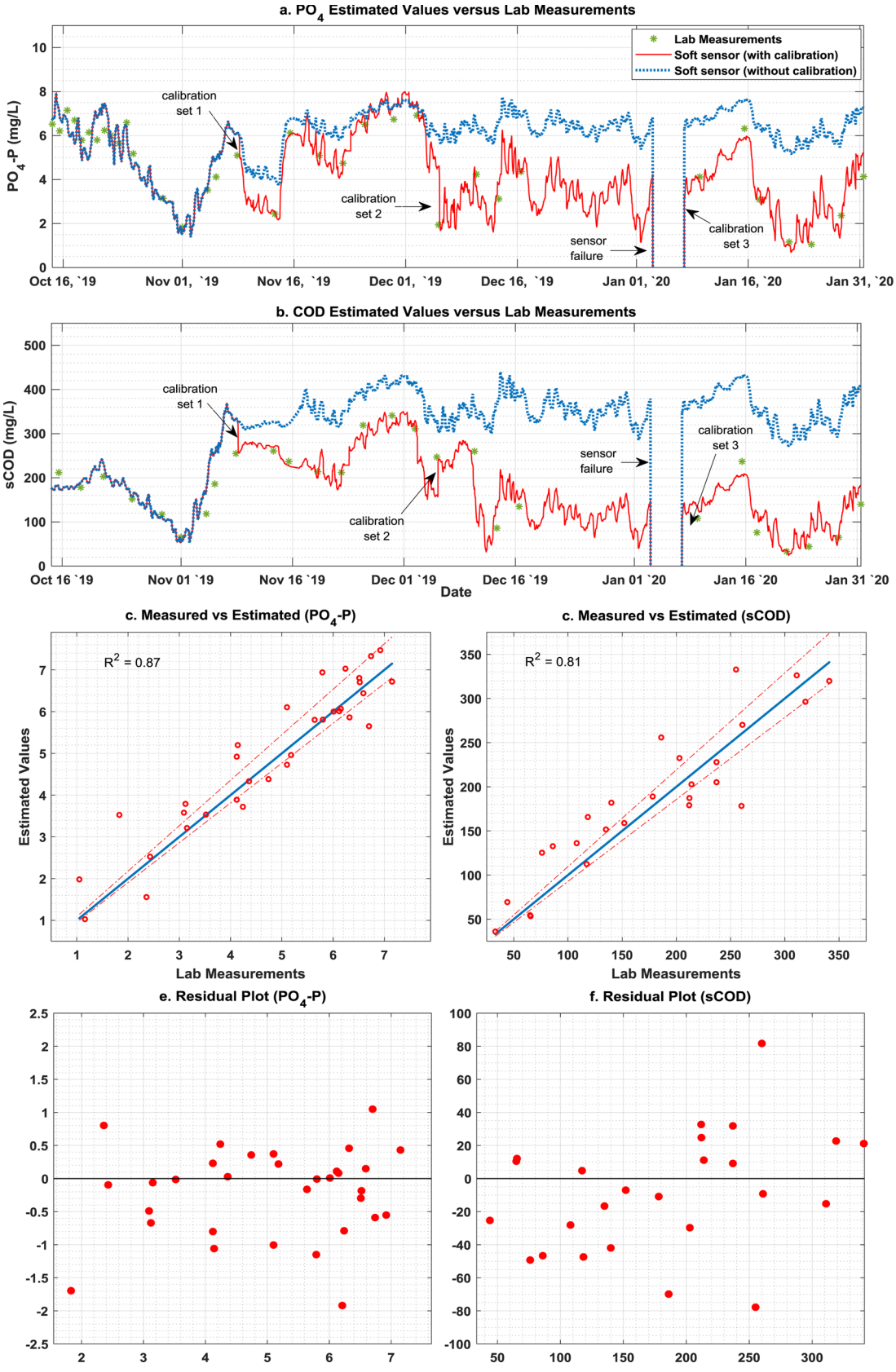

Fig. 7. Long-term evaluation of soft-sensor performance and influence of recalibration strategy on (a) influent PO_4^{3-}-P (b) influent sCOD. Measured vs estimated values (c) PO_4^{3-}-P (d) sCOD. Residual Plots (e) PO_4^{3-}-P (f) sCOD.

Table 4
Regression parameters of the measurement model obtained during recalibration.

Parameter	Calibration Date	R^2	α_0	α_1	α_2	α_3	α_4	α_5
Set 1	08−11-2019	0.93	527 (± 14.8)	−0.0036 (± 0.00050)	2.21 (± 0.12)	0.000600 (± 0.00007)	0.00321 (± 0.0008)	−0.0220 (± 0.0035)
Set 2	05−12-2019	0.96	519 (± 11.8)	−0.0016 (± 0.00015)	2.91 (± 0.16)	0.000510 (± 0.00006)	0.00241 (± 0.0004)	−0.0202 (± 0.0077)
Set 3	07−01-2020	0.87	511 (± 13.8)	−0.0036 (± 0.00025)	2.01 (± 0.11)	0.000110 (± 0.00002)	0.00671 (± 0.0001)	−0.0402 (± 0.0060)

3.4. Influence of recalibration strategy

To assess the robustness of the conductivity based soft-sensor and to study their long-term performance, the estimated values $S_{s,in}$ and $S_{po,in}$ were monitored for an additional period of 3 months (14th October 2019 to 06th February 2020). The assessment was performed by comparing the lab results from the biweekly measurements of sCOD and PO_4^{3-}-P concentrations in raw wastewater to the corresponding values of $S_{s,in}$ and $S_{po,in}$, respectively, estimated by the soft-sensor. Preliminary assessment of soft-sensor showed that, although the estimation results were very close to the lab measurements for the first 3 weeks, from the beginning of week 4 (on 5th November 2019), the first sign of drift from the lab measurements started to appear.

The accuracy of the mathematical model is a vital element in the performance of the soft sensor. A possible reason could be attributed to the drift in the measurement model adopted in this work. It should be noted that the parameters (α_i) mentioned in Eq. (6) is obtained by correlating the conductivity measurements to the concentration of sCOD and PO_4^{3-}-P obtained from the kinetic tests on a batch scale. However, in reality, the wastewater used in the batch tests contains many other ionic species affecting the conductivity measurements [8]. A few prior works, such as [30], report a stable correlation between electrical conductivity and PO_4^{3-}-P release during anaerobic conditions of an EBPR process. However, a substantial change in the ratio between various ionic species of the influent wastewater over an extended period may affect the accuracy of the initial measurement model obtained from the batch experiments. Including the additional ionic species as extra state variables in the model could offer a possible solution, but it would affect the observability of the model and undermine the idea of using a simple reduced-order model, for estimating the states using an EKF.

An alternative solution for this problem would be to frequently calibrate the measurement model and update the model parameters presented in Table 3. The significance of regular calibration of mathematical models has been emphasized in [31]. The measurement model was updated using the recent values of sCOD, PO_4^{3-}-P, and the corresponding conductivity measurements. The calibration algorithm used for obtaining initial regression parameters was used to obtain the new set. The model calibration followed by the update of model parameters was carried out once every month or whenever the percentage error between the estimated value and the lab measurement was more than 50 %. For the purpose of comparison, the soft-sensor using the initial measurement model (presented in Table 3) was also executed in parallel with the model that was regularly calibrated. Both these data were logged in the remote SQL server.

The soft sensor was calibrated at three different instances, as indicated in Fig. 7a and b. A comparison between the estimated values of sCOD and PO_4^{3-}-P along with the experimental data for chamber 1, 2 and 3 are provided in Fig. 7c and d. The updated parameters of the measurement model obtained from each calibration instance along with their 95 % confidence intervals are mentioned in Table 4. The first calibration of the soft-sensor was carried out on 8th November 2019 followed by another two on 05th December 2019, and 07th January 2020 subsequently. The first update of the measurement model (Set 1) demonstrated a minor difference between the calibrated and uncalibrated estimation. However, the difference increased significantly over time, especially after the second (Set 2) and third (Set 3) calibration instances. The confidence interval presented in Table 4 indicates that the parameters (α_i where $i = 1, 2,..5$) estimated during all three calibration instances are well within the 95 % significance intervals.

A comparison plot, along with the corresponding R^2 values, is presented in Fig. 7. The plots presented in Fig. 7a and b demonstrate that the estimations of sCOD and PO_4^{3-}-P with a regularly calibrated model was closer to the actual values (obtained from standardized lab measurements) when compared to the uncalibrated model. This could infer that periodic calibration of the soft-sensor model by updating the measurement function would solve the estimation error caused by the drift in the measurement model. Additional plots regarding the estimated PO_4^{3-}-P and sCOD in chamber 1, 2, and 3 are provided in Appendix as Fig. A1 and Fig. A2.

4. Conclusions

A reduced-order model explaining the dynamics of phosphorus release during the anaerobic stage of biological wastewater treatment process was developed. The hybrid model combining statistical methods with the mechanistic elements of biological nutrient removal process, adequately explains the reactions while still retaining its ability to be used in conjunction with EKF for estimating the states of the model. The simulator-based testing of the mathematical model provides theoretical validations for the ability of the soft-sensor to provide reliable estimations of states in the anaerobic stages of the EBPR-MBBR pilot plant. The implementation of the EKF based soft sensor in the pilot plant and the corresponding validation studies demonstrate the ability of the soft-sensor to provide reliable estimations of PO_4^{3-}-P and sCOD concentration in the anaerobic chamber of a multistage EBPR-MBBR pilot plant. The work also emphasizes the importance of regular model calibration. A systematic calibration strategy, discussed in this work, addresses the issue of drift in soft-sensor caused by the measurement model mismatch. The approach involving the regular model calibration provides a practical solution to the drift caused by model error. A comparison of the soft-sensor performance using the uncalibrated model and the frequently calibrated model with the lab measurements demonstrates the importance of calibration in enhancing the performance of the soft-sensor. The cost-effective method for estimation of influent PO_4^{3-}-P and sCOD concentration as well as the PO_4^{3-}-P concentrations in the anaerobic chambers of a continuous multistage MBBR pilot plant may be beneficial in the implementation of advanced control strategies for optimal operation of wastewater treatment plants designed for enhanced biological phosphorus removal.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the Norwegian Research Council (NFR) Project RECOVER. The authors acknowledge the technical and financial support from DOSCON AS.

Appendix A

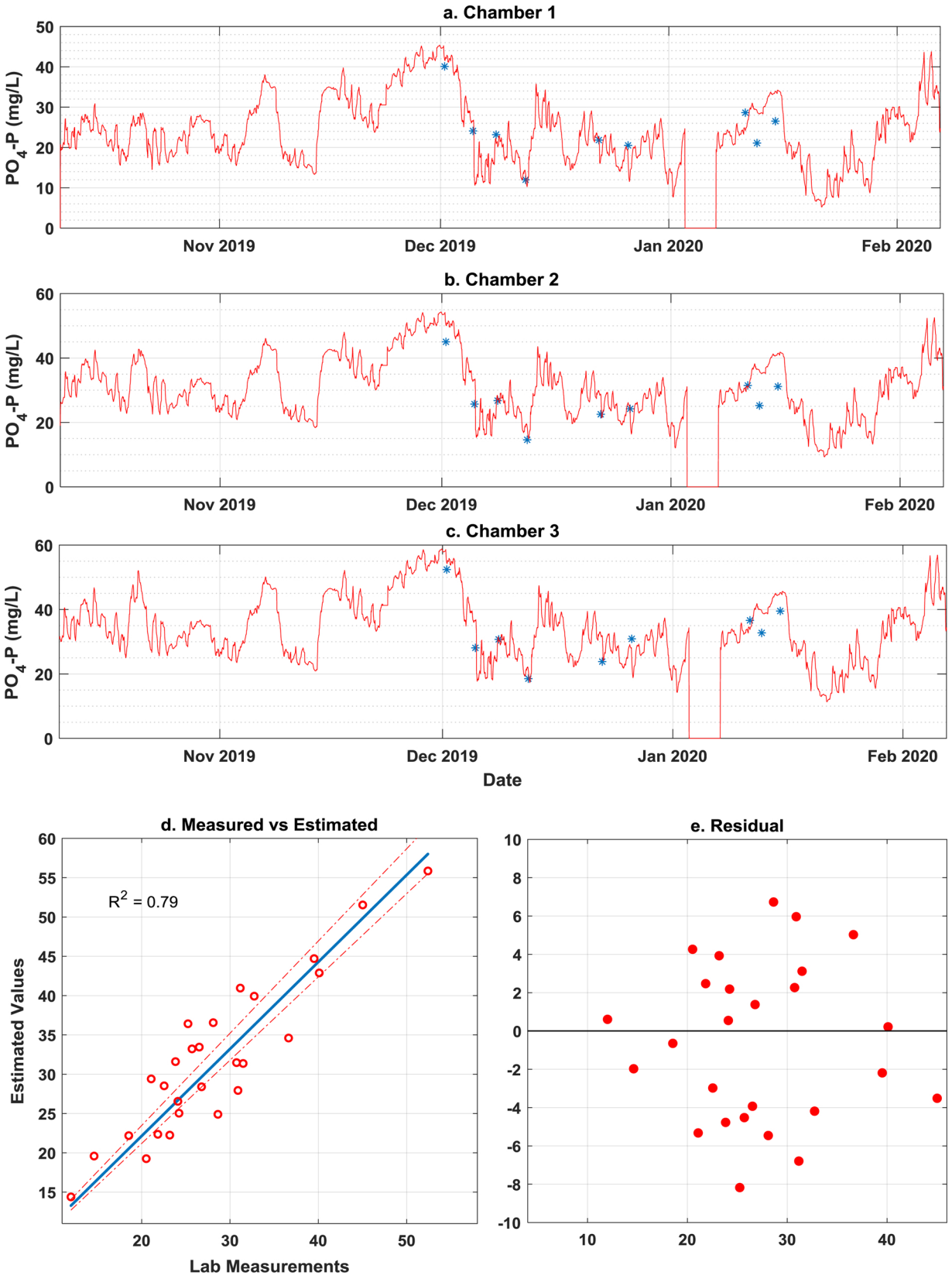

Fig. A1. Long-term measured versus estimated plots for $PO_4^{3-}P$ in (a) Chamber 1, (b) Chamber 2, (c) Chamber 3. (d) Measured versus estimated (e) Residual Plots.

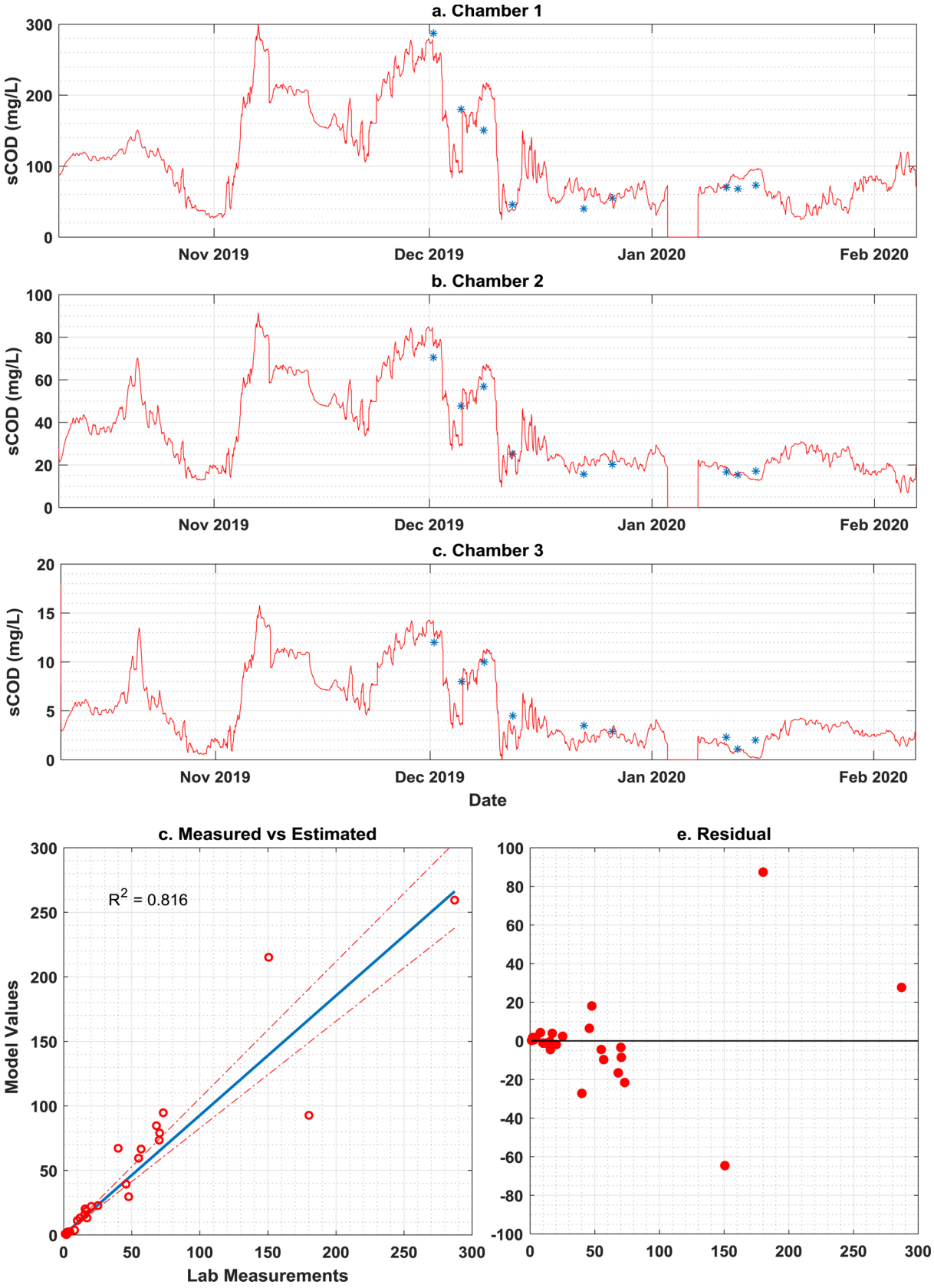

Fig. A2. Long-term measured versus estimated plots for sCOD in (a) Chamber 1, (b) Chamber 2, (c) Chamber 3. (d) Measured versus estimated (e) Residual Plots.

References

[1] T. Mino, M.C.M. van Loosdrecht, J.J. Heijnen, Microbiology and biochemistry of the enhanced biological phosphate removal process, Water Res. 32 (11) (1998) 3193–3207, https://doi.org/10.1016/S0043-1354(98)00129-8.

[2] P.A. Vanrolleghem, D.S. Lee, On-line monitoring equipment for wastewater treatment processes: state of the art, Water Sci. Technol. 47 (2) (2003) 1–34, https://doi.org/10.2166/wst.2003.0074.

[3] G. Olsson, ICA and me - A subjective review, Water Res. 46 (6) (2012) 1585–1624, https://doi.org/10.1016/j.watres.2011.12.054.

[4] D. Dochain, P. Vanrolleghem, Dynamical Modelling and Estimation in Wastewater Treatment Processes, IWA Publishing, London, U.K, 2001, pp. 250–274, https://doi.org/10.2166/9781780403045.

[5] H. Haimi, F. Corona, M. Mulas, L. Sundell, M. Heinonen, R. Vahala, Shall we use hardware sensor measurements or soft-sensor estimates? Case study in a full-scale WWTP, Environ. Model. Softw. 72 (2015) 215–229, https://doi.org/10.1016/j.envsoft.2015.07.013.

[6] J. Busch, D. Elixmann, P. Kühl, C. Gerkens, J.P. Schlöder, H.G. Bock, W. Marquardt, State estimation for large-scale wastewater treatment plants, Water Res. 47 (13) (2013) 4774–4787, https://doi.org/10.1016/j.watres.2013.04.007.

[7] N. Patel, J. Ruparelia, J. Barve, Prediction of total suspended solids present in effluent of primary clarifier of industrial common effluent treatment plant: mechanistic and fuzzy approach, J. Water Process. Eng. 34 (2020) 101146, , https://doi.org/10.1016/j.jwpe.2020.101146.

[8] K.S. Kim, J.S. Yoo, S. Kim, H.J. Lee, K.H. Ahn, I.S. Kim, Relationship between the electric conductivity and phosphorus concentration variations in an enhanced biological nutrient removal process, Water Sci. Technol. 55 (1–2) (2007) 203–208, https://doi.org/10.2166/wst.2007.053.

[9] D. Aguado, A. Ferrer, A. Seco, J. Ferrer, Comparison of different predictive models for nutrient estimation in a sequencing batch reactor for wastewater treatment, Chemom. Intell. Lab. Syst. 84 (1–2) (2006) 75–81, https://doi.org/10.1016/j.chemolab.2006.03.009.

[10] D. Aguado, T. Montoya, J. Ferrer, A. Seco, Relating ions concentration variations to conductivity variations in a sequencing batch reactor operated for enhanced biological phosphorus removal, Environ. Model. Softw. 21 (6) (2006) 845–851, https://doi.org/10.1016/j.envsoft.2005.03.004.

[11] D. Aguado, M. Zarzo, A. Seco, A. Ferrer, Process understanding of a wastewater batch reactor with block-wise PLS, Environmetrics 18 (2007) 551–560, https://doi.org/10.1002/env.828.

[12] J. Serralta, L. Borrás, C. Blanco, R. Barat, A. Seco, Monitoring pH and electric conductivity in an EBPR sequencing batch reactor, Water Sci. Technol. 50 (10) (2004) 145–152, https://doi.org/10.2166/wst.2004.0630.

[13] T. Saltnes, G. Sørensen, S. Eikås, Biological nutrient removal in a continuous biofilm process, Water Pract. Technol. 12 (4) (2017) 797–805, https://doi.org/10.2166/wpt.2017.083.

[14] M.F.R. Zuthi, W.S. Guo, H.H. Ngo, L.D. Nghiem, F.I. Hai, Enhanced biological phosphorus removal and its modeling for the activated sludge and membrane bioreactor processes, Bioresour. Technol. 139 (2013) 363–374, https://doi.org/10.1016/j.biortech.2013.04.038.

[15] M. Henze, W. Gujer, T. Mino, T. Matsou, M.C. Wentzel, Gv.R. Marais, M.C.M. Van Loosdrecht, Activated sludge model No. 2D, ASM2D, Water Sci. Technol. Technol. 39 (1) (1999) 165–182, https://doi.org/10.1016/S0273-1223(98)00829-4.

[16] K.B. Newhart, R.W. Holloway, A.S. Hering, T.Y. Cath, Data-driven performance analyses of wastewater treatment plants: a review, Water Res. 157 (2019) 498–513, https://doi.org/10.1016/j.watres.2019.03.030.

[17] U. Jeppsson, G. Olsson, Reduced order models for on-line paameter identification of the activated sludge, Water Sci. Technol. Technol. 28 (11) (1994) 173–183, https://doi.org/10.2166/wst.1993.0657.

[18] X. Yin, J. Liu, State estimation of wastewater treatment plants based on model approximation, Comput. Chem. Eng. 111 (2018) 79–91, https://doi.org/10.1016/j.compchemeng.2018.01.003.

[19] J. Carstensen, P. Harremoes, R. Strubee, Software sensors based on the grey-box modelling approach, Water Sci. Technol. 33 (1) (1996) 117–126, https://doi.org/10.1016/0273-223(96)00164-3.

[20] C. Gómez-Quintero, I. Queinnec, M. Spérandio, A reduced linear model of an activated sludge process, Ifac Proc. Vol. 37 (3) (2004) 219–224, https://doi.org/10.1016/S1474-6670(17)32586-7.

[21] V.K. Madyastha, V. Prasad, V. Mahendraker, Reduced order model monitoring and control of a membrane bioreactor system via delayed measurements, Water Sci. Technol. 64 (8) (2011) 1675–1684, https://doi.org/10.2166/wst.2011.437.

[22] M.A. Steffens, P.A. Lant, R.B. Newell, A systematic approach for reducing complex biological wastewater treatment models, Water Res. 31 (3) (1997) 590–606, https://doi.org/10.1016/S0043-1354(96)00273-4 (1997).

[23] T. Zhang, D. Zhang, Z. Li, Q. Cai, Evaluating the structural identifiability of the parameters of the EBPR sub-model in ASM2d by the differential algebra method, Water Res. 44 (9) (2010) 2815–2822, https://doi.org/10.1016/j.watres.2010.02.027.

[24] Z. Li, P. Lu, D. Zhang, T. Zhang, Practical identifiability analysis and optimal experimental design for the parameter estimation of the ASM2d-Based EBPR anaerobic submodel, Math. Probl. Eng. 2018 (2018) 1–9, https://doi.org/10.1155/2018/9201085.

[25] M. Maurer, W. Gujer, Monitoring of microbial phosphorous release in batch experiments using electric conductivity, Water Res. 29 (11) (1995) 2613–2617, https://doi.org/10.1016/0043-1354(95)00146-C.

[26] A. Nair, V.M. Cristea, P.S. Agachi, M. Brehar, Model calibration and feed-forward control of the wastewater treatment plant – case study for CLUJ-Napoca WWTP, Water Environ. J. 32 (2) (2018) 164–172, https://doi.org/10.1111/wej.12310.

[27] A.M. Nair, A. Hykkerud, H. Ratnaweera, A cost-effective IoT strategy for remote deployment of soft sensors – a case study on implementing a soft sensor in a multistage MBBR plant, Water Sci. Technol. (2020), https://doi.org/10.2166/wst.2020.067.

[28] D. Simon, Optimal State Estimation: Kalman, H, and Nonlinear Approaches, John Wiley & Sons, USA, 2006, https://doi.org/10.1002/0470045345.

[29] F. Haugen, R. Bakke, L. Bernt, State estimation and model-based control of a pilot anaerobic digestion reactor, J. Control. Sci. Eng. 2014 (2014) 572–621, https://doi.org/10.1155/2014/572621.

[30] A. Colin Wylie, Investigation of Electrical Conductivity As a Control Parameter for Enhanced Biological Phosphorus Removal in a Pilot Scale Sequencing Batch Reactor. Thesis, University of British, Columbia, 2001, https://doi.org/10.14288/1.0067103 2009. Print.

[31] F. Tscheikner-Gratl, P. Zeisl, C. Kinzel, J. Leimgruber, T. Ertl, W. Rauch, M. Kleidorfer, Lost in calibration: why people still do not calibrate their models, and why they still should – a case study from urban drainage modelling, Water Sci. Technol. 74 (10) (2016) 2337–2348, https://doi.org/10.2166/wst.2016.395.

[33] S. Daneshgar, P.A. Vanrolleghem, C. Vaneeckhaute, A. Buttafava, A.G. Capodaglio, Optimization of P compounds recovery from aerobic sludge by chemical modeling and response surface methodology combination, Sci. Total Environ. 668 (2019) 668–677, https://doi.org/10.1016/j.scitotenv.2019.03.055.