Journal of Process Control 148 (2025) 103404

Contents lists available at ScienceDirect

Journal of Process Control

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

Real-time chlorate by-product monitoring through hybrid estimation methods

E.A. Ross [a,b], R.M. Wagterveld [a], M.J.J. Mayer [c], J.D. Stigter [b], K.J. Keesman [a,b,*]

[a] Wetsus, European Centre of Excellence for Sustainable Water Technology, Oostergoweg 9, 8911 MA, Leeuwarden, The Netherlands
[b] Mathematical and Statistical Methods - Biometris, Wageningen University and Research, PO Box 16, 6700 AA, Wageningen, The Netherlands
[c] EasyMeasure B.V., Breestraat 22, 3811 BJ, Amersfoort, The Netherlands

ARTICLE INFO

Keywords:
Monitoring
Machine learning
Observer
Hybrid
Electrochlorination

ABSTRACT

Since the strict regulations regarding chlorate concentrations in drinking water and in food, there exists a need to monitor this by-product stemming from electrochlorination. Since, currently, there are no chlorate-specific sensors, Sensor Data Fusion is proposed as an alternative.

The objective of this paper is to investigate and design Sensor Data Fusion algorithms that are accurate over a broader set of circumstances.

Two different estimators are explored, both of which combine a first-principles model with a machine learning algorithm. The first-principles models are based on a nonlinear, reduced-order state-space model. The data-driven models investigated were multiple linear regression, K nearest neighbors, a gradient-boosting decision tree and support vector regression, with optimized hyperparameters and a two-stage validation process.

It was found that the addition of a first-principles model reduced the cross-validation mean squared error by 58%, and allows accurate scaling with the fluid flow rate, when used in combination with support vector regression. Furthermore, a relatively simple hybrid approach, with state-space and data-driven models in series, was sufficient in terms of accuracy, when compared to a more complex series–parallel hybrid version. The latter does provide information regarding the free chlorine concentration and current efficiencies in real-time, as well as an estimate of the uncertainties associated with the process states. The 1 σ confidence interval converged to 14% of the chlorate estimate.

The results indicate that a hybrid approach is viable in the design of a Sensor Data Fusion algorithm for chlorate monitoring, and preferable over a purely data-driven approach.

1. Introduction

Electrochlorination is a widely used means of disinfecting water, as in drinking water, food, and medical equipment, among others. Its primary advantage over the use of chlorine gas is that it is safer. Rather than chlorine gas canisters, only water and sodium chloride salt are needed in an electro-chemical system [1,2].

Recently, the European Union strictly regulated the permissible chlorate concentration in drinking water and food [3] as it was found to be potentially harmful to human health [4]. Chlorate is a by-product of chlorine-based disinfection methods. Thus, this calls for careful, preferably real-time monitoring of the chlorate formation in electrochlorination. Unfortunately, there are no chlorate sensors commercially available. Sensor Data Fusion (SDF; see Table 1 for a list of abbreviations used in this work) has been proposed as an alternative method of monitoring the process, as it is able to monitor in real-time, and can work with common, affordable, and robust sensors [5,6].

SDF has been applied extensively to the monitoring of electrochemical processes [7–13]. However, its application to the monitoring of chlorate formation during electrochlorination is relatively novel [14]. Recently, a first-principles approach has been developed [5], based on an Extended Kalman Filter (EKF) observer and a reduced-order model [15–17]. However, the resulting estimator was found to be sensitive to the current efficiency parameters. This means frequent calibration is necessary to ensure that the estimates remain accurate in terms of bias. Consequently, robustness of the EKF is expected to be limited.

* Corresponding author at: Mathematical and Statistical Methods - Biometris, Wageningen University and Research, PO Box 16, 6700 AA, Wageningen, The Netherlands.
E-mail address: karel.keesman@wur.nl (K.J. Keesman).

https://doi.org/10.1016/j.jprocont.2025.103404
Received 24 April 2023; Received in revised form 23 November 2024; Accepted 20 February 2025
Available online 6 March 2025

Table 1
Abbreviations.

CV	Cross-Validation
EC	Electrical Conductivity
EKF	Extended Kalman Filter
GBDT	Gradient Boosting Decision Tree
IC	Ion Chromatography
KNN	K Nearest Neighbors
MLR	Multiple Linear Regression
MSE	Mean Squared Error
SDF	Sensor Data Fusion
SVR	Support Vector Regression

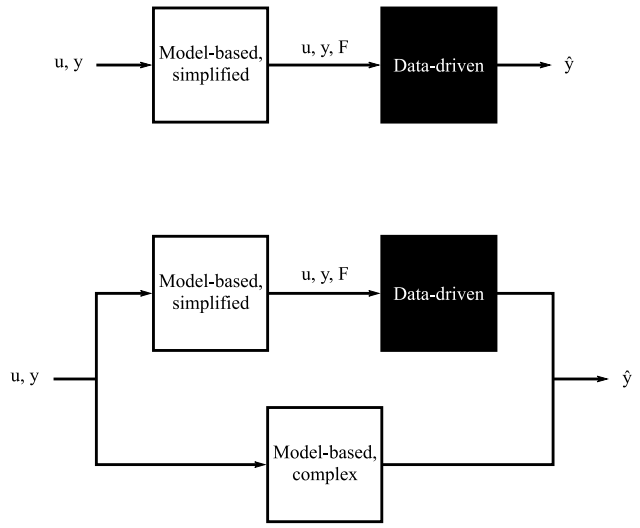

Fig. 1. The two hybrid structures investigated in this work, namely series hybrid (top) and series–parallel hybrid (bottom). In the schemes, u are the inputs, y the sensor signals, F the engineered features and $\hat{y}$ the estimates.

The bias of the estimates may be reduced by utilizing a data-driven model, in combination with first-principles knowledge of the electrochlorination process, using a hybrid estimator [18]. In such an approach, a first-principles submodel captures the aspects of the system that can be robustly modeled, such as the system's dynamics, while the data-driven submodel captures the difficult-to-model relations related to the current efficiency [19]. The main benefit over a purely data-driven estimator is that such an approach requires less data, since part of the knowledge comes from first principles instead. This is critical for the application in question, as each datapoint requires sampling and laboratory analysis to obtain. Thus, there will be relatively few datapoints available. There are many means of integrating first-principles and data-driven knowledge for process monitoring [20,21]. In this work we explored two approaches to estimate chlorate formation during electrochlorination: A so-called series hybrid estimator, as well as a series–parallel hybrid estimator (see Fig. 1). In the first approach, physics-informed features were provided to an otherwise data-driven model. These physics-informed features were derived from a first-principles model, which is the 'simplified white box' in Fig. 1. In the second approach, a series–parallel hybrid estimator used the output of the series hybrid estimator to correct a more complex first-principles model (the 'complex white box' in Fig. 1), via an Extended Kalman Filter (EKF). The goal of this paper was to investigate and design more robust and more accurate SDF algorithms for use in electrochlorination.

In Section 2, we will explain both the first-principles and data-driven submodels, their integration, the data gathering and the validation strategy. The results will be shown in Section 3, both for the first-principles method, the data-driven methods, and the combination of the two. The results will subsequently be discussed in Section 4.

2. Method

In this section we will describe the models used in the SDF algorithms. First, we will formulate the full first-principles model. Then, the data-driven models will be described, followed by a description of the dataset and validation procedure. Finally, the hybrid algorithms will be explained.

2.1. First-principles model

The intention of the first-principles model is to capture the known physics of the system, shown in Fig. 2, in particular its dynamics and the relation between the chemical matrix in the bulk of the cell and the chemical matrix in the sensor volume.

The nonlinear, reduced-order model is based on our earlier work [5], and constitutes the following state equations, with $[\mathrm{ClO^-}]_c$ a nonlinear function of the states (see Eq. (15)) [22–24].

$$\frac{dx_1}{dt} = I\theta_1 - \nu I\theta_2[\mathrm{ClO^-}]_c - \frac{Q}{V_c}x_1 \tag{1}$$

$$\frac{dx_2}{dt} = \frac{Q}{V_c}[\mathrm{OH^-}]_{in} - \frac{Q}{V_c}x_2 \tag{2}$$

$$\frac{dx_3}{dt} = I\theta_2[\mathrm{ClO^-}]_c - \frac{Q}{V_c}x_3 \tag{3}$$

$$\frac{dx_4}{dt} = \frac{Q}{V_s}x_1 - \frac{Q}{V_s}x_4 \tag{4}$$

$$\frac{dx_5}{dt} = \frac{Q}{V_s}x_2 - \frac{Q}{V_s}x_5 \tag{5}$$

$$\frac{dx_6}{dt} = \frac{Q}{V_s}x_3 - \frac{Q}{V_s}x_6 \tag{6}$$

With:

$$x_1 := [\mathrm{ClO^-}]_c + [\mathrm{HClO}]_c \tag{7}$$

$$x_2 := [\mathrm{OH^-}]_c - [\mathrm{HClO}]_c \tag{8}$$

$$x_3 := [\mathrm{ClO_3^-}]_c \tag{9}$$

$$x_4 := [\mathrm{ClO^-}]_s + [\mathrm{HClO}]_s \tag{10}$$

$$x_5 := [\mathrm{OH^-}]_s - [\mathrm{HClO}]_s \tag{11}$$

$$x_6 := [\mathrm{ClO_3^-}]_s \tag{12}$$

Here, Q is the volume flow rate (l/s), V the volume (l), I the current over the electrodes (A) and ν a stoichiometric parameter $(-)$. The subscript c means the variable regards the cell volume, and the subscript s means that it regards the sensor volume. Furthermore, the parameters are defined as follows:

$$\theta_1 := \frac{\eta_{\mathrm{ClO^-}}}{FV_c e_{\mathrm{ClO^-}}} \tag{13}$$

$$\theta_2 := \frac{\eta_{\mathrm{ClO_3^-}}}{FV_c e_{\mathrm{ClO_3^-}}} \tag{14}$$

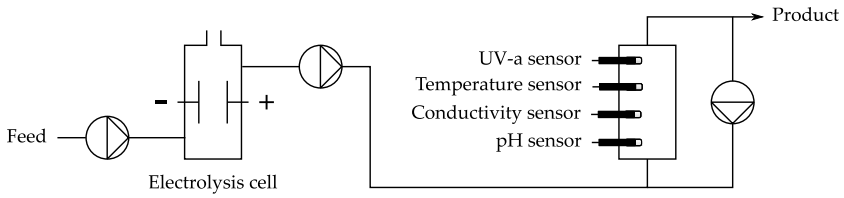

Fig. 2. Scheme depicting the laboratory electrochlorination setup.

e_{ClO^-} and $e_{ClO_3^-}$ the stoichiometric number of electrons transferred in the hypochlorite and chlorate half-reactions, respectively (−). The hypochlorite concentration in the cell $[ClO^-]_c$ (mol/l), see Eqs. (1) and (3), can be determined as follows:

$$[ClO^-]_c = \frac{k_{eq} + 2x_1 + x_2 - \sqrt{k_{eq}^2 + 4k_{eq}x_1 + 2k_{eq}x_2 + x_2^2}}{2} \tag{15}$$

In the original work, the parameters θ_1 and θ_2 were assumed to be constant [5]. In this work, the current efficiencies η_{ClO^-} and $\eta_{ClO_3^-}$ are assumed to vary over time. To accommodate the time-varying parameters, the state vector $x := [x_1\ x_2\ \cdots\ x_6]^T$ has been augmented with two states, which are defined as follows:

$$x_7 := \theta_1 \tag{16}$$

$$x_8 := \theta_2 \tag{17}$$

The derivatives of these states are set to zero. Thus, these states will only change due to the correction step in the Kalman filter. As will be explained in Section 2.6, a data-driven submodel will be used to this effect.

A UV-a absorbance sensor and a pH probe are used to guide the estimator. Prior to use in this estimator, these signals are corrected for temperature and converted to the hypochlorite and hydroxide concentration, respectively. Thus, the output vector $y := [y_1\ y_2]^T$ is defined as follows [5]:

$$y_1 = [ClO^-]_s = \frac{k_{eq} + 2x_4 + x_5 - \sqrt{k_{eq}^2 + 4k_{eq}x_4 + 2k_{eq}x_5 + x_5^2}}{2} \tag{18}$$

$$y_2 = [OH^-]_s = \frac{-k_{eq} + x_5 + \sqrt{k_{eq}^2 + 4k_{eq}x_4 + 2k_{eq}x_5 + x_5^2}}{2} \tag{19}$$

The system is controlled through the current I, flow rate Q and the feed hydroxide concentration $[OH^-]_{in}$ (mol/l). Hence the control input vector $u := [u_1\ u_2\ u_3]^T$ is defined with:

$$u_1 := I \tag{20}$$

$$u_2 := Q \tag{21}$$

$$u_3 := [OH^-]_{in} \tag{22}$$

The disturbances that this process is subject to are mainly conditional changes that affect the electrolysis at the electrodes, such as changes in temperature and impurities in the feed, which may also affect the model quality to some extent.

2.2. Data-driven models

For this work, several methods of regression have been implemented, namely multiple linear regression (MLR), K nearest neighbors (KNN), gradient boosting decision tree (GBDT) and support vector regression (SVR) [25–27]. An important distinction between these models is that KNN and GBDT are inherently non-linear. MLR and SVR are linear in principle, but can be extended to nonlinear curves via kernel functions.

Table 2
Range of conditional parameters used to generate the dataset.

Process parameter	Range
Salt concentration of the feed	24 – 32 g/l
Flow velocity	4 – 8.1 ml/min
Temperature of the feed	15 – 45 °C
Recirculation flow velocity	0 – 5.4 ml/min
pH buffer added to the feed	0 – 3 mmol/l $NaHCO_3$

2.3. Data collection

The data used for the regression models consists of 224 datapoints, from 16 electrochlorination experiments using a custom electrochlorination cell. Each experiment recorded the response to a step-change in current, from zero to 1500 A/m^2, under constant flow and constant feed chemistry. The conditions of the experiments were varied, as summarized in Table 2. The cell and sensor volumes were 0.105 and 0.288 l, respectively.

The responses were recorded using pH, EC, temperature and UV-a absorption sensors. The pH and EC readings have been temperature-corrected, prior to use as features for the machine learning algorithms.

In addition, the differences between the current sensor readings to that at $t = 0$ were provided as features, as well, in order to simulate a differential sensor setup. The readings at $t = 0$ can be considered to correspond to those of the feed, under the assumption that the properties of the feed remained constant throughout the experiment. This extension of the features has been done for the EC and temperature readings, but not for pH, as the pH at $t = 0$ is ill-defined due to the very low pH buffer capacity of the feed. This extension has not been done for the UV-a absorption data either, as the absorption at $t = 0$ is almost zero.

Consequently, the available features are: pH, EC, EC difference to that at $t = 0$, temperature in Kelvin, temperature difference to that at $t = 0$, UV-a absorption, $[e^-]_{s,1}$ and $[e^-]_{s,2}$. For the fitting of models, the data has been normalized.

Besides the recording of sensor data, samples were taken every 15 min, and analyzed for free chlorine and chlorate concentrations. Due to the relatively slow flow rate three minutes were required to take one sample, while the sensor set provided data every 4.2 s. For each entry in the dataset, the sample data corresponding to the time span required to take a physical sample has been averaged.

2.4. Data-driven modeling: validation, hyperparameter tuning and feature selection

As the data consists of a collection of time series, not all data points are independent. A simple shuffling of the data, and splitting into training and validation sets, would overestimate the data driven model's performance, as the data in the training and validation sets will regard the same experiments. Instead, for each validation, data from one complete and separate experiment is used. In this way independence of the training and validation sets is guaranteed, as no information regarding that experiment is used for training. The features and chlorate concentrations were normalized when training the models.

The overall approach is two-tiered. First, the model's hyperparameters were tuned and features were selected using cross-validation,

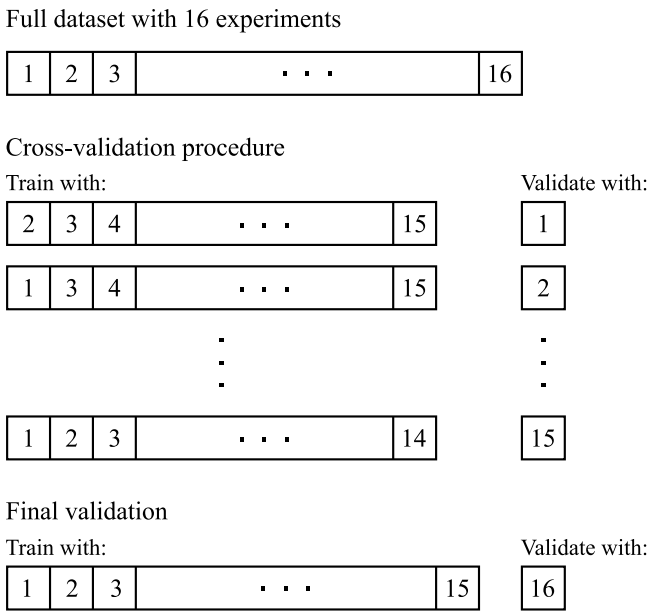

Fig. 3. Data usage throughout the validation strategy.

where the goal is to make the Mean Squared Error (MSE) of all the estimated data points across all cross-validation folds is as low as possible [28]. Data from one experiment is excluded from this. This data is used as validation data for the final validation, used for visualization and to validate the hybrid estimator. The data was excluded from the cross-validation to prevent specifically tuning the hyperparameters and feature selection to the final validation case shown in this work. The data usage throughout these procedures is shown graphically in Fig. 3.

The methods of tuning the data-driven model hyperparameters can be found in Appendix A. The features were selected by repeating the cross-validation procedure and hyperparameter optimization for each possible combination of features.

The extrapolation performance was also assessed, by repeating the tuning and validation, with the experiment with the highest chlorate concentration used exclusively for the final validation step. In that particular experiment, the chlorate concentration was relatively high due to the low flow velocity of 4 ml/min, mainly because there was simply less fluid to dissolve the reaction products into, per unit of time. This yielded a maximal chlorate concentration that was 27% higher than the next-highest in the dataset.

2.5. Simplified first-principles model and series hybrid estimator

To assist the regression models, a simplified first-principles model has been developed. The aim is to provide the data-driven models with real-time data that is linear with the hypochlorite and chlorate concentrations, but is otherwise as simple as possible and depends only on the input vector u.

The simplified model output $[e^-]_{s,1}$ describes the quantity of electrons that have participated in a reaction per supplied volume of fluid, subjected to the same mixing dynamics as the physical concentrations. This yields a "virtual electron concentration" in the sensor volume, which correlates linearly with the physical concentration of hypochlorite in the sensor volume, in case of a fixed current efficiency. The derivation of the next equations can be found in Appendix B. The virtual electron balances are given by:

$$\frac{d[e^-]_{c,1}}{dt} = \frac{I}{FV_c} - \frac{Q}{V_c}[e^-]_{c,1} \tag{23}$$

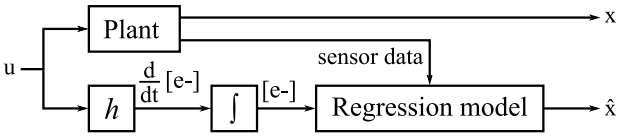

Fig. 4. Scheme of the series hybrid estimator.

$$\frac{d[e^-]_{s,1}}{dt} = \frac{Q}{V_c}[e^-]_{c,1} - \frac{Q}{V_s}[e^-]_{s,1} \tag{24}$$

Here, $[e^-]_{c,1}$ and $[e^-]_{s,1}$ are the first order virtual electron concentrations in the cell and sensor volumes, respectively (mol/l).

In contrast to the hypochlorite formation, the chlorate formation is expected to not only be linearly dependent on the current, but on the hypochlorite concentration as well, as is expressed in Eq. (3) [23]. Thus, the chlorate formation is linear with both the current and $[e^-]_{s,1}$, and is expected to follow a second order response to input u. Assuming the dynamics do still depend solely on (perfect) mixing, this second order $[e^-]$ responses for the fixed cell and sensor volumes may be calculated using the following differential equations:

$$\frac{d[e^-]_{c,2}}{dt} = \frac{I}{FV_c}[e^-]_{c,1} - \frac{Q}{V_c}[e^-]_{c,2} \tag{25}$$

$$\frac{d[e^-]_{s,2}}{dt} = \frac{Q}{V_c}[e^-]_{c,2} - \frac{Q}{V_s}[e^-]_{s,2} \tag{26}$$

With $[e^-]_{c,2}$ and $[e^-]_{s,2}$ the second order virtual electron concentrations in the cell and sensor volumes, respectively (mol^2/l^2), which are expected to be approximately linear with the chlorate concentrations in the cell and sensor volume. The derivation of these equations can be found in Appendix B, as well. Notice that in case I and Q are constant, an analytical solution can be found.

The calculated trajectories of $[e^-]_{s,1}$ and $[e^-]_{s,2}$ are used to fit the data-driven models, as those pertain to the volume where the sensors are located, and for which the chlorate concentration is estimated. The resulting series hybrid estimator is depicted in Fig. 4. There, $[e^-] := [[e^-]_{s,1}\ [e^-]_{s,2}]^T$, and the vector function h contains the right-hand side functions of the set of differential equations used to generate $[e^-]_{s,1}$ and $[e^-]_{s,2}$. Input $u_3 := [OH^-]_{in}$ is not used for this estimator.

2.6. Series–parallel hybrid estimator

To combine the full first-principles state-space model as described in Eqs. (1)–(19) with the series hybrid model described in Section 2.5, a series–parallel hybrid estimation scheme has been developed. The estimator is based on an Extended Kalman Filter (EKF). The output $y_1 := [ClO^-]_s$ is compared to real-time data from on the UV-a absorption sensor, which was related to $[ClO^-]_s$ using a linear calibration curve. The output y_2 is compared to $[OH^-]_s$ data as determined in real-time by the pH probe.

To connect the first-principles model to the series hybrid model, the output vector y has been augmented with the chlorate concentration in the sensor volume as determined by the data-driven model:

$$y_3 = [ClO_3^-]_s = x_6 \tag{27}$$

For $\bar{u}_1, \bar{u}_2 > 0$ and non-zero parameters, the linearized version of the augmented model is observable in any point on the state trajectory (see the Supplementary Material). The covariance matrices P_0, Q and R used in the EKF are given in Appendix A.

The hybrid estimation process is as depicted in Fig. 5. In that figure, f is the set of differential equations of the state space model, g is the set of algebraic equations of the state space model and h is the set of differential equations used to determine the virtual electron concentrations (Eqs.(23)–(26)). The block K represents the algorithm that determines the Kalman gain. For this estimation process all three inputs in vector u are used.

Table 3
Feature selection, optimization results and model MSE scores.

Model	Features	Hyperparameters	CV MSE ($mmol^2$)	Final MSE ($mmol^2$)	Extrap. MSE ($mmol^2$)
MLR	pH, EC, EC difference, T, $[e^-]_{s,2}$		0.1273	0.0129	0.0395
KNN	T difference, $[e^-]_{s,2}$	Uniform weights $K = 4$	0.1642	0.1438	1.0882
GBDT	T, T difference, $[e^-]_{s,1}$, $[e^-]_{s,2}$	depth = 4 $n_e = 40$ $\eta = 0.10$ Row subsample = 0.89 Column subsample = 0.64	0.1225	0.0192	0.9250
SVR	pH, EC, EC difference, T, $[e^-]_{s,2}$	Linear kernel $C = 76$ $\epsilon = 1.5 \times 10^{-5}$	0.0895	0.0055	0.0085

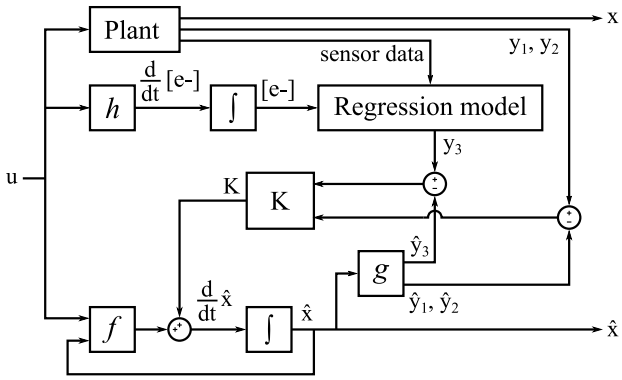

Fig. 5. Scheme of the series–parallel hybrid estimator.

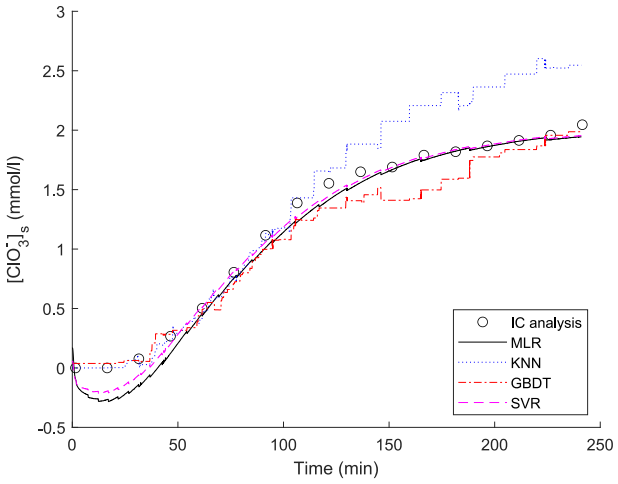

Fig. 6. Estimates of the chlorate concentration in the final validation experiment, using the series hybrid models.

3. Results

In this section, we will first present the results of the series hybrid models, and then of the series–parallel hybrid models.

3.1. Series hybrid chlorate estimation

The feature selection, hyperparameter tuning and model scores are summarized in Table 3, and the final validation estimation results are shown in Fig. 6. As can be seen in Fig. 6, the chlorate concentration does not rise sharply from the start, as initially the hypochlorite concentration is too low. This corresponds with Eq. (3).

Especially the algorithms that use the MLR and SVR data-driven submodels are able to estimate the chlorate concentration well in the validation test. As can be seen in Table 3, the cross-validation (CV) scores are worse than the validation score, which is as expected, as the validation test regards average process conditions. This indicates that the accuracy of the estimators in each case depends on the process conditions. Thus, it is not uniform.

The $[e^-]_{s,2}$ was part of the optimal feature set regardless of ML model, indicating that the series hybrid approach leads to an improved CV accuracy. It is also notable that the UV-a absorption feature was not optimal for any ML model in the series hybrid case.

For the GBDT a rather large number of estimators n_e was found, in combination with a low learning rate (η), which suggests an extended, but conservative, boosting process is best. An optimal depth of 4 suggests a moderate model complexity. Note that for the SVR algorithm, the linear kernel provided better results than the radial basis function kernel. This is because the algorithm relies heavily on the second order virtual electron concentration $[e^-]_{s,2}$ feature, which is designed to be linear with the chlorate concentration in the sensor volume. The optimal feature set is the same as for MLR.

As shown in Fig. 6, the estimators based on MLR and SVR provide good estimates especially at high concentrations. In practice, negative estimates would be set to zero.

As both MLR and SVR use linear relations, a compromise in accuracy at different chlorate concentrations has to be made. In contrast, in KNN and GBDT no linear relationship is assumed. These models are both very accurate at low chlorate concentrations but less so at higher concentrations. This may be due to a skewed distribution of chlorate levels in the dataset, as discussed in more detail in the discussion and the Supplementary Material.

MLR and, to a lesser extent, SVR show a slightly jagged result; this is likely due to the low resolution of the temperature sensor. It should be noted that this temperature data was used to temperature-correct the EC and pH data, as well.

3.2. Extrapolation performance

The series hybrid estimators based on MLR and SVR are able to estimate high chlorate concentrations due to a low flow rate with good accuracy, as is shown in Fig. 7 and Table 3. When the process conditions exceed any entry in the dataset, and such conditions can be modeled accurately through the implemented first-principles model, these data-driven models are still able to provide an accurate estimate.

In contrast, the estimators based on KNN and GBDT perform poorly when extrapolating, even when the process conditions can be modeled accurately through first principles. The KNN and GBDT submodels cannot estimate chlorate concentrations that are not covered by the training data. Thus, the robustness that a first-principles model offers cannot be utilized with these models.

Table 4
Feature selection, optimization results and model MSE scores without $[e^-]_{s,1}$ and $[e^-]_{s,2}$ features.

Model	Features	Hyperparameters	CV MSE ($mmol^2$)	Final MSE ($mmol^2$)	Extrap. MSE ($mmol^2$)
MLR	pH, EC, T, UV-a absorption		0.3832	0.1105	1.4666
KNN	EC, UV-a absorption	Uniform weights K = 15	0.3824	0.0409	1.0372
GBDT	EC, T, T difference, UV-a absorption	depth = 6 n_e = 44 η = 0.07 Row subsample = 0.85 Column subsample = 0.74	0.3135	0.0742	1.4034
SVR	pH, EC, UV-a absorption	Radial basis function kernel C = 8.5 $\epsilon = 4.7 \times 10^{-5}$	0.2123	0.1410	0.8331

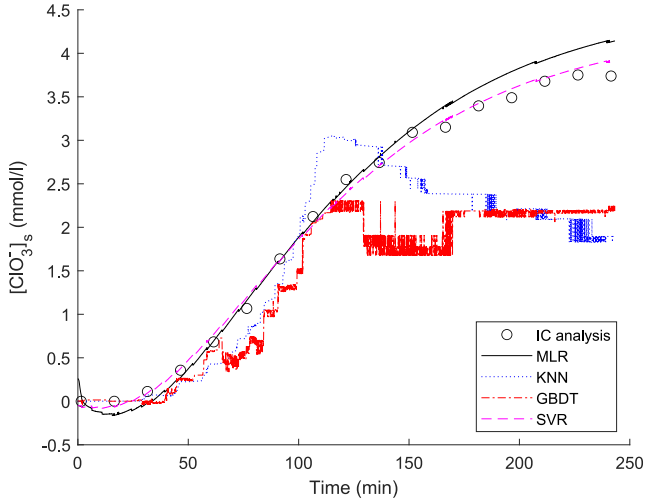

Fig. 7. Estimates of chlorate concentrations for an experiment with higher chlorate concentrations than those in the dataset used for training.

3.3. Influence of physics-informed features

To assess the impact of providing the data-driven models with physics-informed features, the models have been fit without such information, as well. The results are shown in Table 4. The lowest MSE score among the results that do use physics-informed features (shown in Table 3) was 87% lower than the lowest MSE score among the results that do not use physics-informed features. Thus, the series-hybrid approach offers a large improvement over one that only uses a data-driven model. The difference in extrapolation performance is much larger still.

In the absence of $[e^-]_{s,1}$ and $[e^-]_{s,2}$, the radial basis function kernel performed best for SVR. This means that, in the cross-validation procedure, the relationships between the features and the chlorate concentration were not as close to linear as before. The MSE score of the final validation of the SVR-based estimator is worse than that of the estimator based on MLR, but it still performed relatively well.

A notable difference between the purely data-driven estimators and the series hybrid estimators is that, without $[e^-]_{s,1}$ and $[e^-]_{s,2}$, the UV-a absorption data was used instead.

3.4. Series–parallel hybrid estimator results

The results of the series–parallel hybrid estimator are shown in Table 5. The settings for the data-driven models are as in Table 3, and the settings for the EKF can be found in Appendix A. The initial current efficiencies were set at the final values from a previous run.

Table 5
MSE scores for the series–parallel hybrid estimators.

Model	Final MSE ($mmol^2$)	Extrapolation MSE ($mmol^2$)
MLR	0.0170	0.0373
KNN	0.0565	0.3101
GBDT	0.0336	0.6696
SVR	0.0076	0.0076

The series–parallel hybrid approach worsens the result for the GBDT-based series estimator, and improves that of the estimator using the KNN submodel. This is mainly because the dynamics of the current efficiency variables are fairly slow. As can be seen in Fig. 8 for the series–parallel hybrid estimator using the GBDT submodel, the chlorate concentration estimation follows the series hybrid estimation well initially, but is slow to correct after the series hybrid estimator jumps towards a more accurate prediction at around 190 min. For the KNN-based estimators the opposite holds: The series–parallel estimation is slow to correct towards the series hybrid overestimation of the chlorate concentration, which improves the MSE score. The extent to which this holds depends on the covariance matrix of the measurement noise R: A lower value for the covariance of the measurement noise associated with y_3 would mean that the estimator follows the estimations made by the series-hybrid more closely. This effect is smaller when using the MLR and SVR submodels, as these are already rather accurate.

The EKF's state covariance matrix gives an estimate of the uncertainties of the algorithm's estimates. For the estimator with the SVR algorithm, with the P_0, Q and R matrices given in Appendix A, the 1 σ confidence interval converges to 14% of the estimate.

To demonstrate the functionality of the series–parallel hybrid structure, x_7 and x_8 have been given a value that is 25% higher than it would be when correctly calibrated. The results with SVR is shown in Fig. 9. The results of a first-principles estimator using the full model described in Section 2.1, but without data-driven correction, is also shown. These results are labeled "First-principles".

In spite of the erroneous initial current efficiency estimates, the series–parallel estimator produces accurate results, especially at the end of the experiment. Both hybrid approaches are a clear improvement over the estimator that does not apply a machine learning submodel, as the latter is unable to correct the current efficiency parameters. In terms of accuracy, the series–parallel estimator does not improve upon the series hybrid, as the latter does not use the incorrect parameters.

The corrections to the chlorate current efficiency is shown in Fig. 10, based on the estimator with SVR. Note that the chlorate current efficiency depends on the hypochlorite concentration as well (see Eq. (14)) [23].

4. Discussion

The results of this work show that a limited inclusion of the process' physics, in the form of the series hybrid architecture, strongly improves the estimator's performance, in terms of Mean Squared Error.

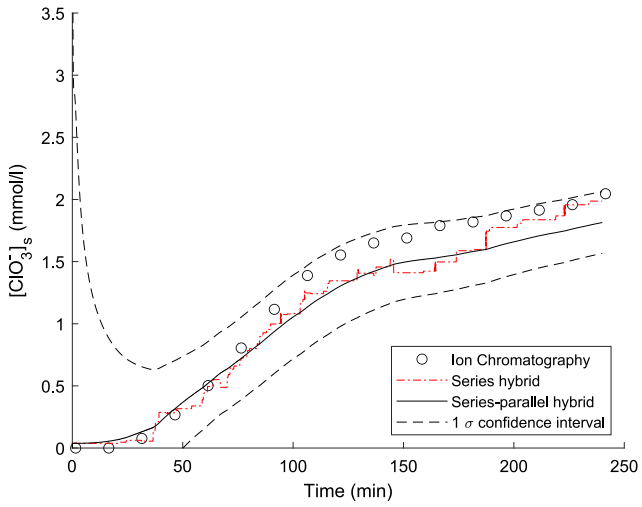

Fig. 8. Estimates of the chlorate concentration in the final validation experiment, using the GBDT-based series–parallel hybrid model.

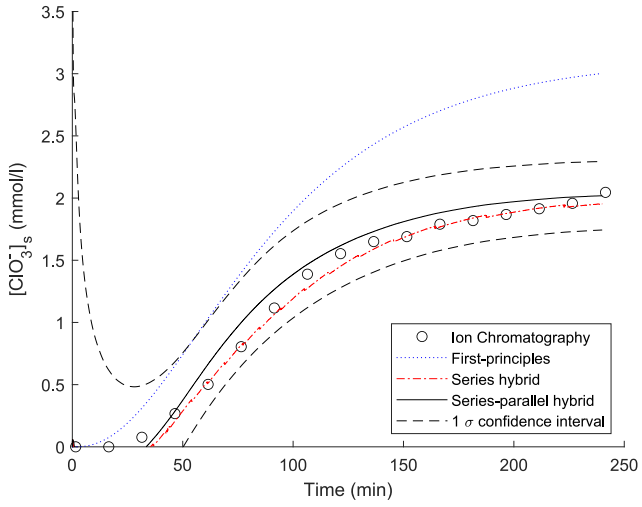

Fig. 9. Estimates of the chlorate concentration in the final validation experiment, with a 25% error in the initial current efficiencies, using the SVR model.

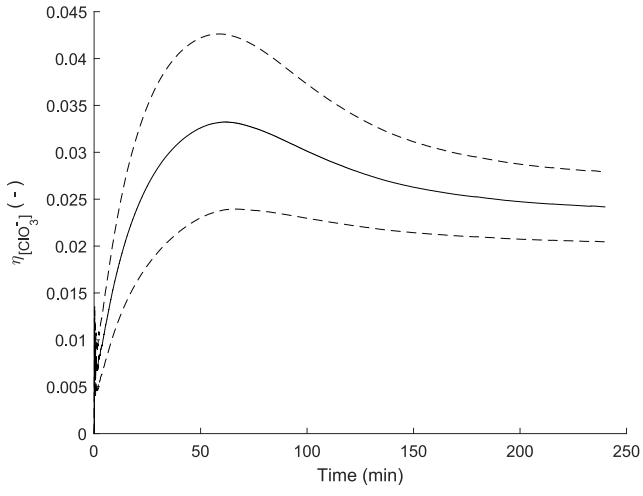

Fig. 10. Estimates of the chlorate current efficiency, with 1 σ confidence interval.

The extensive inclusion of process physics in the form of the series–parallel hybrid estimator yielded no benefit in accuracy over the series hybrid estimator, though it was able to bring the dynamics of the KNN and GBDT-based estimators more in line with the expected process. It is expected that a low-pass filter achieves a similar result. However, the series–parallel structure does bring several advantages: It estimates the hypochlorite and hypochlorous acid concentrations, both in the cell and at the sensor location, as well as the current efficiencies. Crucially, it also provides estimates of the uncertainty associated with the estimates of the species concentrations and current efficiencies.

In addition, this approach may be expanded upon. For instance, an adaptive algorithm might decide between the two parallel paths (simplified model-based and data-driven estimators in series, and a complex model-based estimator, as shown in the bottom part of Fig. 1). Traditionally, when using an EKF, model and measurement covariance matrices have to be defined. These are constant, and irrespective of the current situation. In reality, the uncertainty of a machine learning model depends strongly on whether the dataset used for training is relevant to the current situation, and even relates to the amount of data in the dataset regarding the current situation, as well as the accuracy of that data. In the present case, the uncertainty associated with y_3 may be underestimated in situations not covered well by the dataset, or over-estimated in case it is covered well by the dataset. Ideally, a machine learning algorithm should provide uncertainties with its estimates. This could then be used in the EKF of the series–parallel hybrid estimator, to optimally weigh the state-space and data-driven model estimates in all cases. Research regarding the uncertainty quantification of machine learning algorithms is ongoing [29–31]. In all likelihood, this will mean it is necessary to gather more data, as part of the data will be used to quantify the uncertainty.

In the hybrid architectures of this work the dynamics are described in the first principles models that are embedded in the algorithms. This approach forms an alternative to fully data driven algorithms. In such a case the dynamics are described by a machine learning model that explicitly considers a sequence of time steps, such as with one based on LSTM NN or TCN. Such timeseries ML models typically contain a large number of parameters to fit. Thus, more data is required, relative to the methods proposed in this work. In a heavily data-constrained problem such as electrochemical process monitoring, the more data-efficient methods are expected to be superior. Additional research is required to prove this theory.

Another avenue of research could be modeling the electrochemical process in greater detail, through the Butler–Volmer equation, or relations derived from it, or through an equivalent circuit model. An advantage in the SDF context is that this allows the usage of voltage potentials as supplementary data. Such data closely relates to the formation of species during electrochlorination, and as such is expected to be valuable in further improving the estimator's accuracy.

While some experiments resulted in a higher chlorate concentration than others, all experiments went through a phase with little chlorate. As such, while the full range of chlorate concentrations is in the dataset, the dataset is skewed towards low chlorate concentrations. In practice, this is unfavorable. In an electrochlorination system, there typically exists a buffer tank that temporarily stores the product. The relative accuracy of the high concentrations of chlorate will impact the accuracy of an estimate of the chlorate concentration more than the relative accuracy of the low chlorate concentrations would. In addition, electrochlorinators will spend most of their time producing a solution with a high chlorate concentration, which compounds the problem. So, in case the goal is to estimate high chlorate concentrations, the current dataset is more limited than the number of samples suggests. The skewed composition of the dataset may contribute to the difference in performance between MLR and linear SVR on the one hand, and KNN and GBDT on the other, since with MLR and SVR the lower concentration data is used to fit model parameters that are used to estimate high concentrations as well. In the KNN and GBDT methods,

the higher concentration estimates are independent from lower concentration data. A histogram of the chlorate concentrations is provided in the Supplementary Material.

Interestingly, the UV-a absorption data was mainly used in the absence of physics-informed features. This means that, by providing first-principles-based information, the use of a hardware sensor can be avoided. Of course, the data streams required to calculate the physics-based features need to be available, instead (being the current I and fluid flow rate Q, see Fig. 4 and Eqs. (23)–(26)). The series–parallel hybrid still requires the UV-a absorption sensor for the EKF algorithm.

5. Conclusion

The concentration of chlorate, a by-product formed during electrochlorination, was estimated through two different hybrid approaches, namely a series and a series–parallel approach. It was found that the series approach was sufficient in terms of accuracy. The series–parallel approach does provide more information regarding the process, and is able to provide estimates of the uncertainties, as well. Of the four machine learning methods, Support Vector Regression performed best. The use of first-principles models contributed significantly to the accuracy of the estimators, reducing the MSE of the cross validation by 58%, and the MSE of the final validation by 87%. It allowed accurate scaling beyond the dataset in case of a lower flow rate, when used with SVR.

In practice, the hybrid approach constitutes a more data-efficient method, compared to an approach that is only data-driven. This means that more accurate results may be obtained with the same amount of data, as is shown here, or that less data is required for the same performance. The extrapolation results demonstrate that a hybrid approach is more robust, depending on the chosen data-driven submodel, though only for a change in a condition that is described by the simplified first-principles submodel. Finally, the results show that the hybrid estimators work best with a different set of sensors compared to the purely data-driven approach.

CRediT authorship contribution statement

E.A. Ross: Writing – original draft. **R.M. Wagterveld:** Supervision. **M.J.J. Mayer:** Writing – review & editing. **J.D. Stigter:** Supervision. **K.J. Keesman:** Supervision.

Funding

Wetsus is co-funded by the Dutch Ministry of Economic Affairs and Climate Policy, the European Union Regional Development Fund, the Province of Fryslân and the Northern Netherlands Provinces, with additional support from the participants of the research theme "Sensoring" and Wageningen University and Research.

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.

Acknowledgments

This work was performed in the cooperation framework of Wetsus, European Centre of Excellence for Sustainable Water Technology (https://www.wetsus.eu). Wetsus is co-funded by the Dutch Ministry of Economic Affairs and Climate Policy, the European Union Regional Development Fund, the Province of Fryslân and the Northern Netherlands Provinces. The authors like to thank the participants of the research theme "Sensoring" for the fruitful discussions and financial support.

Appendix A. Detailed model parameters

The methods by which the hyperparameters of the machine learning algorithms are found are shown in Table A.6. The optimization based on the tree-structured Parzen estimator was performed using the Optuna framework [28], with 400 trials per optimization study.

The initial covariance matrix P_0 of the EKF used for the series–parallel hybrid estimator was a diagonal matrix with 10^5 on the diagonal, except for the elements relating to the current efficiency states. These were set to 10^{-1}. The covariance matrix of the system noise Q was also diagonal, with the diagonal elements set to 5×10^{-5} times the normalized steady state values of the states. The covariance matrix of the measurement noise R was diagonal with 10^2, 10^3 and 5×10^4 on the diagonal. With these values, the estimates did not strongly follow the measurements while still being corrected by the measurements over time.

Appendix B. Derivation of the simplified model

The premise of the simplified model is to provide features to a regression model that are based on the current and flow rate, and are linear with the chlorate concentration in the sensor set. This is expected to be possible, as the amount of product and by-product formed is highly dependent on the amount of current per volume flow, which may be described as the concentration of available virtual electrons per liter.

The quantity of electrons transferred from the electrode into the bulk solution of the cell in moles per second is simply:

$$\frac{d\mathrm{e}^-}{dt} = \frac{I}{F} \tag{B.1}$$

These are assumed to be equally distributed over the electrolysis cell volume, V_c, leading to what can be considered a virtual concentration in the absence of flow.

$$\frac{d[\mathrm{e}^-]_{c,1}}{dt} = \frac{I}{FV_c} \tag{B.2}$$

To account for the flow through the cell volume, the virtual electron concentration can be subjected to the same mixing dynamics as the physical species as described in Section 2.1.

$$\frac{d[\mathrm{e}^-]_{c,1}}{dt} = \frac{I}{FV_c} - \frac{Q}{V_c}[\mathrm{e}^-]_{c,1} \tag{B.3}$$

The concentration in the cell is equal to the concentration of the feed to the sensor volume. When accounting for flow through the sensor volume in a similar fashion as the cell, the differential equation for the virtual electron concentration in the sensor volume becomes as follows.

$$\frac{d[\mathrm{e}^-]_{s,1}}{dt} = \frac{Q}{V_c}[\mathrm{e}^-]_{c,1} - \frac{Q}{V_s}[\mathrm{e}^-]_{s,1} \tag{B.4}$$

These virtual electron concentrations are expected to be approximately linear with the hypochlorite concentrations, as is shown in what follows.

The free chlorine concentration in the cell volume is modeled as follows (see Section 2.1 Eq. (1) of the main text):

$$\frac{d([\mathrm{ClO}^-]_c + [\mathrm{HClO}]_c)}{dt} = \frac{I\eta_{\mathrm{ClO}^-}}{FV_c e_{\mathrm{ClO}^-}} - \nu\frac{I\eta_{\mathrm{ClO}_3^-}}{FV_c e_{\mathrm{ClO}_3^-}}[\mathrm{ClO}^-]_c - \frac{Q}{V_c}([\mathrm{ClO}^-]_c + [\mathrm{HClO}]_c) \tag{B.5}$$

At a pH of 10, we can assume all free chlorine will be in the form of hypochlorite [24]. Additionally, we may assume the amount of free chlorine lost to chlorate formation is small. This simplifies Eq. (B.5) to:

$$\frac{d[\mathrm{ClO}^-]_c}{dt} = \frac{I\eta_{\mathrm{ClO}^-}}{FV_c e_{\mathrm{ClO}^-}} - \frac{Q}{V_c}[\mathrm{ClO}^-]_c \tag{B.6}$$

Table A.6
Hyperparameter values and optimization methods.

Model	Hyperparameter	Method of optimization
MLR	Feature subset	Exhaustive search
KNN	Feature subset	Exhaustive search
	Number of neighbors	Iterative grid search
	Weights	Exhaustive search
GBDT	Feature subset	Exhaustive search
	Number of estimators	Tree-structured Parzen estimator
	Learning rate	Tree-structured Parzen estimator
	Tree depth	Tree-structured Parzen estimator
	Subsampling of rows	Tree-structured Parzen estimator
	Subsampling of columns	Tree-structured Parzen estimator
SVR	Feature subset	Exhaustive search
	Kernel	Exhaustive search
	C	Tree-structured Parzen estimator
	ϵ	Tree-structured Parzen estimator

In the following it is shown that $[ClO^-]_c$ is proportional to $[e^-]_{c,1}$, that is:

$$[ClO^-]_c = C_1[e^-]_{c,1} \tag{B.7}$$

With C_1 assumed to be a constant. Substituting Eq. (B.7) into Eq. (B.6) gives:

$$\frac{d\left(C_1[e^-]_{c,1}\right)}{dt} = \frac{I\eta_{ClO^-}}{FV_c e_{ClO^-}} - \frac{Q}{V_c}C_1[e^-]_{c,1} \tag{B.8}$$

Since C_1 was assumed to be a constant, we find:

$$\frac{d[e^-]_{c,1}}{dt} = \frac{I\eta_{ClO^-}}{C_1 FV_c e_{ClO^-}} - \frac{Q}{V_c}[e^-]_{c,1} \tag{B.9}$$

Thus, through Eq. (B.3), Eq. (B.7) holds if:

$$C_1 = \frac{\eta_{ClO^-}}{e_{ClO^-}} \tag{B.10}$$

Assuming C_1 is constant. Given that e_{ClO^-} is a stoichiometric number, this implies η_{ClO^-} is assumed to be constant. In this case, and under the other mentioned assumptions, $[ClO^-]_c$ is linear with $[e^-]_{c,1}$.

The rate of formation of chlorate in the cell was found to be linear with both the current and hypochlorite concentration [23]. In Section 2.1, the chlorate mass balance reads as:

$$\frac{d[ClO_3^-]_c}{dt} = \frac{I\eta_{ClO_3^-}}{FV_c e_{ClO_3^-}}[ClO^-]_c - \frac{Q}{V_c}[ClO_3^-]_c \tag{B.11}$$

A corresponding, "second-order virtual electron concentration" is formulated, as follows.

$$\frac{d[e^-]_{c,2}}{dt} = \frac{I}{FV_c}[e^-]_{c,1} - \frac{Q}{V_c}[e^-]_{c,2} \tag{B.12}$$

The premise is that $[e^-]_{c,2}$ is linear with $[ClO_3^-]_c$, under the assumption of a constant $\eta_{ClO_3^-}$. So, in a similar vain as before, we aim to show:

$$[ClO_3^-]_c = C_2[e^-]_{c,2} \tag{B.13}$$

With C_2 again assumed to be a constant. We can use Eq. (B.13) to substitute $[ClO_3^-]_c$ in Eqs. (B.11), and (B.7) to substitute $[ClO^-]_c$. Consequently,

$$\frac{d\left(C_2[e^-]_{c,2}\right)}{dt} = \frac{I\eta_{ClO_3^-}}{FV_c e_{ClO_3^-}}C_1[e^-]_{c,1} - \frac{Q}{V_c}C_2[e^-]_{c,2} \tag{B.14}$$

Given that C_2 is a constant and using Eq. (B.10), we find:

$$\frac{d[e^-]_{c,2}}{dt} = \frac{I\eta_{ClO_3^-}\eta_{ClO^-}}{C_2 FV_c e_{ClO_3^-} e_{ClO^-}}[e^-]_{c,1} - \frac{Q}{V_c}[e^-]_{c,2} \tag{B.15}$$

Eq. (B.15) corresponds with Eq. (B.12) under the following condition:

$$C_2 = \frac{\eta_{ClO_3^-}\eta_{ClO^-}}{e_{ClO_3^-} e_{ClO^-}} \tag{B.16}$$

Thus, both $\eta_{ClO_3^-}$ and η_{ClO^-} are assumed to be constant.

We will show that under the same assumptions, $[ClO^-]_s$ is linear with $[e^-]_{s,1}$. At a pH of roughly 10, the free chlorine can be assumed to be only hypochlorite, as assumed before. This means Eq. (4) simplifies to:

$$\frac{d[ClO^-]_s}{dt} = \frac{Q}{V_c}[ClO^-]_c - \frac{Q}{V_s}[ClO^-]_s \tag{B.17}$$

Again, a linear relation between $[ClO^-]_s$ and $[e^-]_{s,1}$ implies:

$$[ClO^-]_s = C_3[e^-]_{s,1} \tag{B.18}$$

Using Eq. (B.18) to substitute $[ClO^-]_s$ in Eq. (B.17), and Eq. (B.7) to substitute $[ClO^-]_c$, we find:

$$\frac{d\left(C_3[e^-]_{s,1}\right)}{dt} = \frac{C_1 Q}{V_c}[e^-]_{c,1} - \frac{Q}{V_s}C_3[e^-]_{s,1} \tag{B.19}$$

Given that C_3 is assumed to be constant, this gives:

$$\frac{d[e^-]_{s,1}}{dt} = \frac{C_1 Q}{C_3 V_c}[e^-]_{c,1} - \frac{Q}{V_s}[e^-]_{s,1} \tag{B.20}$$

This is equivalent to Eq. (B.4), in case C_1 is equal to C_3.

Finally, we can define a differential equation for the "second order virtual electron concentration" in the volume of the sensor set, $[e^-]_{s,2}$.

$$\frac{d[e^-]_{s,2}}{dt} = \frac{Q}{V_c}[e^-]_{c,2} - \frac{Q}{V_s}[e^-]_{s,2} \tag{B.21}$$

Using the same procedure as before, we find that $[e^-]_{s,2}$ is proportional to the chlorate concentration in the volume of the sensor set ($[ClO_3^-]_s$), through the following relation:

$$[ClO_3^-]_s = \frac{\eta_{ClO_3^-}\eta_{ClO^-}}{e_{ClO_3^-} e_{ClO^-}}[e^-]_{s,2} \tag{B.22}$$

This again assumes both $\eta_{ClO_3^-}$ and η_{ClO^-} are constant.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jprocont.2025.103404.

Data availability

The data presented in this study are available in the files Sample_data.csv, Sensor_data_validation.csv and Sensor_data_extrapolation.csv, submitted as supplementary material for this article.

References

[1] Black & Veatch Corporation, White's Handbook of Chlorination and Alternative Disinfectants, fifth ed., John Wiley and Sons, Inc., Hoboken, NJ, USA, 2009.
[2] J. Saha, S. Gupta, A novel electro-chlorinator using low cost graphite electrode for drinking water disinfection, Ionics 23 (2017) 1903–1913, http://dx.doi.org/10.1007/s11581-017-2022-0.

[3] The European Parliament and the Council of the European Union, Directive (EU) 2020/2184 of the European parliament and of the council of 16 December 2020 on the quality of water intended for human consumption (recast), 2020.
[4] J. Cotruvo, J.K. Fawell, Chlorine Dioxide, Chlorite and Chlorate in Drinking-Water, Background Document WHO/FWC/WSH/16.49, World Health Organization, Geneva, Switzerland, 2016.
[5] E. Ross, M. Wagterveld, M. Mayer, H. Stigter, B. Højris, Y. Li, K. Keesman, Sensor data fusion as an alternative for monitoring chlorate in electrochlorination applications, Sustain. 14 (10) (2022) http://dx.doi.org/10.3390/su14106119.
[6] L. Fortuna, S. Graziani, A. Rizzo, M. Xibilia, Soft sensors for monitoring and control of industrial processes, first ed., Advances in Industrial Control, Springer-Verlag London, 2007, http://dx.doi.org/10.1007/978-1-84628-480-9.
[7] P. Shrivastava, T.K. Soon, M.Y.I.B. Idris, S. Mekhilef, Overview of model-based online state-of-charge estimation using Kalman filter family for lithium-ion batteries, Renew. Sustain. Energy Rev. 113 (2019) 109233, http://dx.doi.org/10.1016/j.rser.2019.06.040.
[8] M.-F. Ng, J. Zhao, Q. Yan, G.J. Conduit, Z.W. Seh, Predicting the state of charge and health of batteries using data-driven machine learning, Nat. Mach. Intell. 2 (2020) 161–170, http://dx.doi.org/10.1038/s42256-020-0156-7.
[9] H. Yuan, H. Dai, X. Wei, P. Ming, Model-based observers for internal states estimation and control of proton exchange membrane fuel cell system: A review, J. Power Sources 468 (2020) 228376, http://dx.doi.org/10.1016/j.jpowsour.2020.228376.
[10] H. Liu, J. Chen, D. Hissel, J. Lu, M. Hou, Z. Shao, Prognostics methods and degradation indexes of proton exchange membrane fuel cells: A review, Renew. Sustain. Energy Rev. 123 (2020) 109721, http://dx.doi.org/10.1016/j.rser.2020.109721.
[11] O. Nolte, I.A. Volodin, C. Stolze, M.D. Hager, U.S. Schubert, Trust is good, control is better: A review on monitoring and characterization techniques for flow battery electrolytes, Mater. Horizons 8 (2021) 1866–1925, http://dx.doi.org/10.1039/D0MH01632B.
[12] J. Shi, Y. Yao, J. Bao, M. Skyllas-Kazacos, B.J. Welch, A. Jassim, M. Mahmoud, Advanced model-based estimation and control of alumina concentration in an aluminum reduction cell, JOM 74 (2) (2022) 706–717, http://dx.doi.org/10.1007/s11837-021-05073-3.
[13] D. Roman, S. Saxena, J. Bruns, R. Valentin, M. Pecht, D. Flynn, A machine learning degradation model for electrochemical capacitors operated at high temperature, IEEE Access 9 (2021) 25544–25553, http://dx.doi.org/10.1109/ACCESS.2021.3057959.
[14] E. Ross, R. Wagterveld, J. Stigter, M. Mayer, K. Keesman, Sensor data fusion in electrochemical applications: An overview and its application to electrochlorination monitoring, Comput. Chem. Eng. (2023) 108128, http://dx.doi.org/10.1016/j.compchemeng.2022.108128.
[15] A. Gelb, Applied Optimal Estimation, MIT Press, Cambridge, U.S., 1974.
[16] A. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, U.S. and London, U.K., 1970.
[17] H. Kwakernaak, R. Sivan, Linear Optimal Control Systems, John Wiley and Sons, New York, U.S., 1972.
[18] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput. Chem. Eng. 60 (2014) 86–101, http://dx.doi.org/10.1016/j.compchemeng.2013.08.008.
[19] M. Adaikkappan, N. Sathiyamoorthy, Modeling, state of charge estimation, and charging of lithium-ion battery in electric vehicle: A review, Int. J. Energy Res. (2021) 2141–2165, http://dx.doi.org/10.1002/er.7339.
[20] J. Sansana, M.N. Joswiak, I. Castillo, Z. Wang, R. Rendall, L.H. Chiang, M.S. Reis, Recent trends on hybrid modeling for Industry 4.0, Comput. Chem. Eng. 151 (2021) 107365, http://dx.doi.org/10.1016/j.compchemeng.2021.107365.
[21] G. Karniadakis, Y. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (2021) 422–440, http://dx.doi.org/10.1038/s42254-021-00314-5.
[22] L. Czarnetzki, Aspects of Electrochemical Production of Hypochlorite and Chlorate (Ph.D. thesis), Technische Universiteit Eindhoven, 1989, http://dx.doi.org/10.6100/IR316966.
[23] L.R. Czarnetzki, L.J.J. Janssen, Formation of hypochlorite, chlorate and oxygen during NaCl electrolysis from alkaline solutions at an RuO2/TiO2 anode, J. Appl. Electrochem. 22 (1992) 315–324, http://dx.doi.org/10.1007/BF01092683.
[24] Agency for Toxic Substances and Disease Registry (ATSDR), Toxicological Profile for Chlorine, Tech. Rep., U.S. Department of Health and Human Services, 2010.
[25] V. Vapnik, The Nature of Statistical Learning Theory, second ed., Springer, New York, 1995, http://dx.doi.org/10.1007/978-1-4757-3264-1.
[26] N. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Amer. Statist. 46 (3) (1992) 175–185, http://dx.doi.org/10.1080/00031305.1992.10475879.
[27] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, 2016, pp. 785–794, http://dx.doi.org/10.1145/2939672.2939785.
[28] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019, pp. 2623–2631, http://dx.doi.org/10.1145/3292500.3330701.
[29] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U.R. Acharya, V. Makarenkov, S. Nahavandi, A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Inf. Fusion 76 (2021) 243–297, http://dx.doi.org/10.1016/j.inffus.2021.05.008.
[30] R. Ghanem, D. Higdon, H. Owhadi, Handbook of Uncertainty Quantification, Springer International Publishing, 2017, http://dx.doi.org/10.1007/978-3-319-12385-1.
[31] T. Sullivan, Introduction to Uncertainty Quantification, Springer Cham, 2015, http://dx.doi.org/10.1007/978-3-319-23395-6.