Chemical Engineering Science 283 (2024) 119413

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Hybrid modeling of a biorefinery separation process to monitor short-term and long-term membrane fouling

Elia Arnese-Feffin [a], Pierantonio Facco [a], Daniele Turati [b], Fabrizio Bezzo [a], Massimiliano Barolo [a,*]

[a] CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova PD, Italy
[b] Novamont S.p.A., via G. Fauser 8, 28100 Novara NO, Italy

ARTICLE INFO

Keywords:
Biorefinery
Hybrid model
Membrane filtration
Reversible fouling
Irreversible fouling
Fouling

ABSTRACT

Membrane filtration is commonly used in biorefineries to separate cells from fermentation broths containing the desired products. However, membrane fouling can cause short-term process disruption and long-term membrane degradation. The evolution of membrane resistance over time can be monitored to track fouling, but this calls for adequate sensors in the plant. This requirement might not be fulfilled even in modern biorefineries, especially when multiple, tightly interconnected membrane modules are used. Therefore, characterization of fouling in industrial facilities remains a challenge. In this study, we propose a hybrid modeling strategy to characterize both reversible and irreversible fouling in multi-module biorefinery membrane separation systems. We couple a linear data-driven model, to provide high-frequency estimates of trans-membrane pressures from the available measurements, with a simple nonlinear knowledge-driven model, to compute the resistances of the individual membrane modules. We test the proposed strategy using real data from the world's first industrial biorefinery manufacturing 1,4-bio-butanediol via fermentation of renewable raw materials. We show how monitoring of individual resistances, even when done by simple visual inspection, offers valuable insight on the reversible and irreversible fouling state of the membranes. We also discuss the advantage of the proposed approach, over monitoring trans-membrane pressures and permeate fluxes, from the standpoints of data variability, effect of process changes, interaction between module in multi-module systems, and fouling dynamics.

1. Introduction

Biorefineries are facilities that integrate sustainable biomass conversion processes and equipment to output a range of products (Velidandi et al., 2023), among which fuels, such as ethanol (Delgenes, 1996) or biodiesel (McCurdy et al., 2014), and building-block chemicals, such as succinic acid (Mancini et al., 2020) or 1,4-bio-butanediol (Burgard et al., 2016). Concepts of sustainability, social impact, and circular economy are accounted for in biorefinery process design and life cycle assessment (Ioannidou et al., 2020; Julio et al., 2017; Sikdar, 2003), aiming to ensure a sustainable production of substances traditionally derived from petroleum (Martín and Grossmann, 2013). Process operations typically include: feedstock and media preparation, preliminary growth of microorganisms, and large scale fermentation in the upstream section; broth sterilization, cell separation, and product recovery and purification in the downstream section (Bähner et al., 2021).

Fermentation broths containing the desired products are usually diluted solutions, which entails high downstream processing cost (Cuellar and Straathof, 2020; Martín and Grossmann, 2013). Membrane filtration has been identified as an effective technology to remove cells (and compounds with high molecular weight) from the solutions containing the main product (Prochaska et al., 2018). The topic has been widely investigated (Abels et al., 2013; Carstensen et al., 2012; Ennaceri et al., 2022; Gerardo et al., 2014; Jiang and Zhu, 2013; Saha et al., 2017) and is in fact becoming increasingly relevant in biorefineries (Abels et al., 2013; Ennaceri et al., 2022; Saha et al., 2017), due to the lower operating costs compared to conventional thermal separation processes (Ennaceri et al., 2022; Gerardo et al., 2014; Jiang and Zhu, 2013; Saha et al., 2017).

Pressure-driven membrane separation processes (e.g., ultrafiltration and nanofiltration operated (semi-)continuously) are most used to separate biomass from the fermentation products (Rudolph et al., 2019). However, when such processes are applied to the filtration of fermentation broths, they can strongly suffer from membrane fouling (Arnese-

* Corresponding author.
E-mail address: max.barolo@unipd.it (M. Barolo).

https://doi.org/10.1016/j.ces.2023.119413
Received 22 August 2023; Received in revised form 27 September 2023; Accepted 17 October 2023
Available online 21 October 2023

Feffin et al., 2023; Mancini et al., 2020; Prochaska et al., 2018). Membrane fouling can be characterized as reversible or irreversible: the former is relatively fast, triggers short-term process disruption, and can be removed by hydraulic or chemical cleaning; the latter acts slowly and causes long-term membrane degradation (Huang et al., 2021; Shi et al., 2014). Both fouling types affect membranes at the same time and cause decrease of permeate flux in constant pressure filtration (Abels et al., 2013) or increase of trans-membrane pressure (TMP, pressure difference across the membrane) in constant flux separation (Klimkiewicz et al., 2016), also implying an increase in energy expenditure in the latter case. However, monitoring fouling by means of these process variables might be cumbersome: they have been reported to exhibit strong variability (Philippe et al., 2013), which is in fact determined not only by membrane fouling, but also by the variability of process conditions.

Microfiltration and ultrafiltration can be generally described by the integral form of Darcy's equation (Meindersma et al., 1997; Whitaker, 1986), which relates the permeate volume-flux and the TMP to the membrane resistance to flow (or to its reciprocal, the membrane permeability). In a way, membrane resistance represents the health state of a membrane and provides an indirect measure of its fouling, as proved by several studies of membrane fouling focusing on resistance/permeability modeling and prediction (Barello et al., 2014; Dologlu and Sildir, 2022; Geissler et al., 2005; Han et al., 2020, 2020; Huang et al., 2021; Kallioinen et al., 2006; Philippe et al., 2013; Ruiz-García and Nuez, 2016). In the context of plant operation, the assessment of fouling through online monitoring of membrane resistance is of paramount importance to guarantee prompt processing of upstream fermentation broth, smooth downstream operation, and economical optimality of production.

Estimation of membrane resistance (or permeability) by Darcy's equation is straightforward, provided that one has access to online measurements of permeate flux and TMP. However, even such simple demands might not be met in full-scale industrial processes, or even in pilot plants. Online measurements of permeate flux and TMP are available in plants employing a single membrane module (Chen et al., 2014; Han et al., 2020), but this is not usually the case when multiple modules are used, despite multi-module membrane separation being a common occurrence in industrial practice (Dologlu and Sildir, 2022; Geissler et al., 2005; Kallioinen et al., 2006; Klimkiewicz et al., 2016; Ruiz-García and Nuez, 2016). Given the limited availability of appropriate online data, only the average resistance/permeability of the membrane ensemble is estimated, thus neglecting the actual resistances/permeabilities of single membrane modules. This clearly offers limited insight on the actual fouling state of the modules and hinders the identification of severe fouling events acting on single modules.

On the other hand, online measurements are not the only data source available in industrial processes. In fact, offline measurements are collected during process operations to monitor critical variables not available through online sensors, or that cannot be acquired automatically by cheap and/or reliable sensors (Kadlec et al., 2009). Therefore, available data are typically multi-rate, featuring online variables automatically acquired by the data acquisition system at a high sampling rate (time scale is seconds or minutes), and offline variables manually measured by operators at a low sampling rate (time scale is variable, for example one point per day or per week (Lin et al., 2009)). Assuming, for instance, that available data for each membrane module operating in the plant consist of high sampling rate permeate flux measurements coming from the data acquisition system, and low sampling rate TMP measurements coming from operator-read manometers installed on modules, single-module resistances can still be estimated at the TMP sampling rate (the lowest one). This solution might nonetheless be unsatisfactory, because the resolution of estimates could be too low to properly characterize relatively fast reversible fouling, therefore hindering prompt detection.

Despite the aforementioned limitations, many literature studies aim at modeling the evolution of membrane resistance by exploiting the information hidden in process data. Recent literature reviews (Bagheri et al., 2019; Velidandi et al., 2023) highlight that the most common approach is to consider only the average resistance in multi-module systems, and to focus on either reversible or irreversible fouling, with limited attempts to resolve the two types (Chan et al., 2017; Huang et al., 2021; Klimkiewicz et al., 2016). Furthermore, strongly nonlinear models, such as neural networks, are used by default, which requires massive datasets (Rendall et al., 2019; Sun and Braatz, 2021) to ensure robustness and to discern relevant phenomena (i.e., fouling) from common-cause process variability (i.e., to avoid overfitting). However, such datasets are typically not available for large-scale processes (Rendall et al., 2019; Sun and Braatz, 2021). On the other hand, simpler, linear modeling approaches are less demanding in terms of data amount, but are seldom used as they may require sophisticated measurements (e.g., concentration of relevant compounds in the inlet stream) acquired in specifically designed experimental campaigns (Philippe et al., 2013) to achieve good performance. Linear models can still achieve good performance in the prediction of TMP rather than resistance, as proved by Kaneko and Funatsu (2015), but TMP may not be the most appropriate variable to monitor when the purpose is discriminating between reversible and irreversible fouling.

A different approach is to develop a soft-sensor (Kadlec et al., 2009) combining data and process knowledge in a hybrid modelling framework (Reis et al., 2019; Sansana et al., 2021; Solle et al., 2017; Tidriri et al., 2016). Whereas the potential of soft sensors has been recognized to enhance sustainable process operation (Perera et al., 2023), their application to biorefineries and membrane separation processes at the industrial scale is still limited. In fact, few studies (Chan et al., 2017; Chew et al., 2017; Hwang et al., 2009; Piron et al., 1997) considered the hybrid modelling approach, and most of them (Chew et al., 2017; Hwang et al., 2009; Piron et al., 1997) used neural networks to predict the parameters of a knowledge-driven model (the cake filtration model) aimed at describing the evolution of resistance due to reversible fouling (therefore neglecting irreversible fouling). Such studies considered a single-module pilot plant (Chew et al., 2017), or multi-module lab equipment (Piron et al., 1997) and pilot plants (Hwang et al., 2009), estimating only the average resistance in the latter cases. Chan et al. (2017) adopted a different approach, using the cake filtration model to estimate the energy requirement of single process runs and Gaussian process regression to model the prediction mismatch between runs, indirectly providing separate models for reversible and irreversible fouling. However, they achieved such results in a single laboratory-scale membrane module operated under controlled fouling conditions.

In this study, we address the problem of characterizing both reversible and irreversible fouling in multi-module industrial biorefinery membrane separation systems by a hybrid modeling strategy that enables high-frequency estimation of the resistances of individual membrane modules. High sampling-rate process data, together with low sampling-rate TMP data, are first used to calibrate (and then use) a PLS soft sensor (Geladi and Kowalski, 1986; Wold et al., 2001) that estimates at high frequency the TMPs of each membrane module. The Darcy's equation is then used to obtain high-frequency estimates of the resistances of each module. To test the proposed strategy, we use real data from two years of operation of an industrial-scale biorefinery that manufactures 1,4-bio-butanediol (bio-BDO) via fermentation of renewable raw materials (Novamont S.p.A., 2016; Satam et al., 2019; Silva et al., 2020). The microfiltration section separates cells from the fermentation broth and features seven interconnected membrane modules equipped with online sensors that measure the permeate fluxes; only offline manometers are available to measure the TMPs, thus requiring manual readings by operators. We show how individual resistance monitoring, even when done by simple visual inspection, offers valuable insight on the reversible and irreversible fouling state of membranes. We also discuss the advantage of monitoring individual resistances, rather than TMPs and permeate fluxes, from the standpoints of data variability, effect of process changes, interaction between

modules in multi-module systems, and dynamic evolution of fouling.

The remainder of the paper is organized as follows. The process investigated in this study, with the available dataset, are described in Section 2. Section 3 focuses on the mathematical models. Results are discussed in Section 4, and conclusions are drawn in Section 5.

2. Ultrafiltration process and data

2.1. Ultrafiltration process

This study focuses on the world's first industrial biorefinery process operating a single-step bioconversion of sugars to bio-BDO by genetically engineered microorganisms (Burgard et al., 2016; Novamont S.p.A., 2016). Three main blocks can be identified in the process: materials preparation, upstream processing, and downstream processing. Sugars, formulated culture medium, and a microorganism inoculum are fed to a series of fed-batch bioreactors for bio-BDO production; mature fermentation broth is sent to continuous downstream processing with separation operations such as membrane filtration, ion-exchange chromatography, evaporation, and distillation.

The ultrafiltration unit processes the fermentation broth and is a critical operation in the downstream train due to the high fouling potential of the feed. A simplified process flow diagram of this operation is illustrated in Fig. 1. Seven membrane modules filter the fermentation broth to remove cells and high molecular weight compounds, yielding a clarified permeate stream containing the bio-BDO. Five of the modules employ a diafiltration strategy to maximize the bio-BDO recovery (Mulder, 1996). The feed flow rate is fixed, while flow rates of retentate and diafiltration solvent are adjusted based on preset ratios to the feed flow rate. The overall permeate flow rate (controlled variable) is kept constant by changing the overall TMP through feed pressure (by acting on the feed pump speed, manipulated variable) adjustment. Furthermore, the modules adopt the cross-flow configuration; therefore, the cross-flow velocities are adjusted by manipulating the speeds of the pumps incorporated in each of the modules (each speed can be manipulated independently), in such a way as to counteract the effects of fouling. The permeate flow rates of single modules are not controlled; therefore, they vary according to both membrane age (i.e., resistance) and applied TMP. The process is interrupted and cleaning is triggered when a preset volume of feed has been filtered or when the feed pressure exceeds a given threshold. Ultrafiltration is therefore run in semi-continuous mode, alternating operating and cleaning phases (Klimkiewicz et al., 2016; Philippe et al., 2013). The overall feed pressure and cross-flow velocity of one module on a selected timespan are reported as an example in Fig. 2, where shaded intervals identify cleaning operations. For confidentiality reasons, all variables will be reported as normalized values within the $[0, 1]$ interval in all figures throughout this paper.

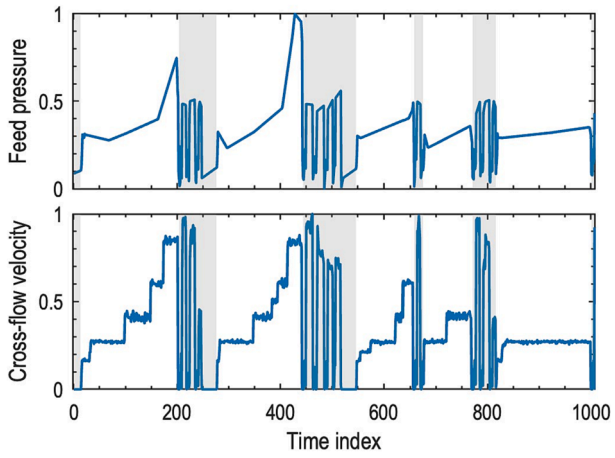

Fig. 2. Example of run/cleaning sequences in the ultrafiltration plant. Shaded intervals identify cleaning operations.

According to the current plant operation, the state of membrane modules is monitored using offline measurements. Readings of manometers installed on each module allow computing the TMPs (see Eq. (4)). Matching online measurements of permeate fluxes are then used to compute the flux-to-TMP ratios (one for each module), which are proportional to membrane permeabilities. An example is shown in Fig. 3 for a period where the membrane of the relevant module was replaced (vertical solid black line), as can be clearly inferred by the trend of flux-to-TMP ratio. Fig. 3 also shows the effects of both reversible fouling (permeability increases after cleaning) and irreversible fouling (average

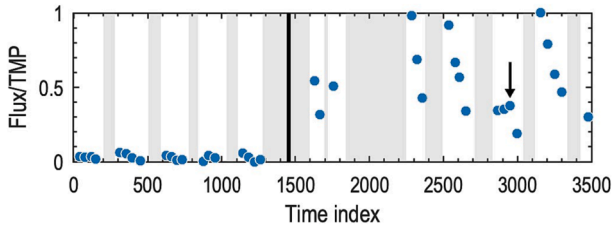

Fig. 3. Example of the effect of membrane replacement (vertical solid black line) onto the flux-to-TMP ratio. Shaded intervals identify cleaning operations. The arrow indicates the last observation before a significant fouling event, which was not detected until the following observation due to the low frequency at which measurements are taken.

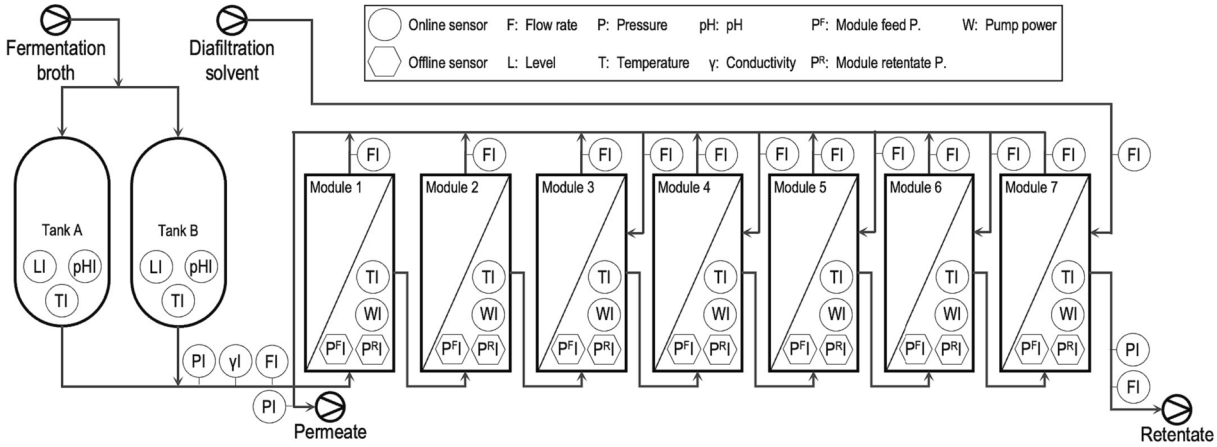

Fig. 1. Scheme of the ultrafiltration operation in the downstream processing section.

permeability decreases across runs), though the former is harder to characterize properly due to the low frequency of readings. For example, a fouling event was detected in the third-to-last run shown in Fig. 3 only after membrane permeability was largely degraded (see arrow in the figure), while an earlier detection would have helped operators to take action immediately.

2.2. Ultrafiltration data

Fig. 1 illustrates all the online variables available through the data acquisition system. Level, temperature, and pH are measured for each one of the feed tanks. The feed manifold features flow rate, pressure, and conductivity sensors. Flow rate measurements are available for the retentate and diafiltration solvent manifolds, while pressure sensors are installed on the retentate and permeate manifolds. Each one of the cross-flow membrane modules is operated with controlled cross-flow velocity (inferred from the measured pump powers), and mounts sensors for temperature, permeate flow rate, and diafiltration solvent flow rate (when relevant). Online pressure sensors are not available on membrane modules, but manometers installed on the feed and retentate pipes allow for manual readings of pressure, which is made available as offline data.

Besides thirty-eight online variables and fourteen offline variables, some additional ones are computed due to the valuable information they provide. The overall permeate flow rate, $\dot{V}^{\mathrm{P}}$, is computed as:

$$\dot{V}^{\mathrm{P}} = \sum_{l=1}^{7} \dot{V}_l^{\mathrm{P}} - \sum_{l=3}^{6} \dot{V}_l^{\mathrm{D}} \tag{1}$$

where $\dot{V}_l^{\mathrm{P}}$ and $\dot{V}_l^{\mathrm{D}}$ are permeate and diafiltration flow rates of module l, respectively. The volume conversion ratio (VCR) of the multi-module system is computed as:

$$\mathrm{VCR} = \min\left\{\dot{V}^{\mathrm{F}}, \frac{\dot{V}^{\mathrm{F}}}{\dot{V}^{\mathrm{R}}}\right\} \tag{2}$$

where the minimum operator is introduced so as to limit the maximum value of the VCR for cases where the retentate flow rate is very low, typically during the startup and shutdown phases of each run. The multi-module system TMP is defined as:

$$\Delta P = \frac{P^{\mathrm{F}} + P^{\mathrm{R}}}{2} - P^{\mathrm{P}} \tag{3}$$

namely, as the difference between the average feed-side pressure and the permeate pressure.

The variables available through relevant offline sensors are also illustrated in Fig. 1. These variables are used to compute the TMPs ΔP_l of single membrane modules:

$$\Delta P_l = \frac{P_l^{\mathrm{F}} + P_l^{\mathrm{R}}}{2} - P^{\mathrm{P}} \tag{4}$$

where P_l^{F} and P_l^{R} are offline readings of the feed and permeate pressures of module l, while P^{P} is the corresponding online measurement of the pressure of the permeate manifold, assumed to be equal for all modules.

Data covering almost two years of operation are available for modeling. Data for some periods are missing due to the change in operational production. A total of 496 runs is found in the datasets. The number of observations per batch for each online variables spans between 21 and 260, with an average of 172, thus entailing a strong variability in the batch duration. On the other hand, offline variables are recorded between 1 and 6 times per batch, with an average of 4; most offline measurements are unevenly spaced and not aligned across runs. However, measurements of both online and offline variables are time-stamped, in such a way that observations can be matched. The available data are used to build a PLS-based soft sensor for online prediction of TMPs of all membrane modules using online variables only. This will be discussed in the next section.

3. Background

3.1. Partial least-squares regression

PLS regression (Geladi and Kowalski, 1986; Wold et al., 2001) models the linear relationship between a set of input variables and output variables. N observations of the V_X inputs are collected in matrix $\mathbf{X} \in \mathbb{R}^N \times \mathbb{R}^{V_X}$, while observations of the V_Y outputs are gathered in matrix $\mathbf{Y} \in \mathbb{R}^N \times \mathbb{R}^{V_Y}$. These matrices are assumed to be autoscaled (mean-centered and scaled to unit variance) prior to PLS model calibration.

PLS modeling identifies a sequence of A pairs of mutually orthogonal latent variables (LVs) defined as linear combinations of input and output variables; one LV in the pair refers to input variables, the other to output variables. LVs are defined so as to maximize the input-output cross-covariance. In fact, models for $\mathbf{X}$ and $\mathbf{Y}$ matrices are provided as well:

$$\mathbf{X} = \mathbf{T}\cdot\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{5}$$

$$\mathbf{Y} = \mathbf{U}\cdot\mathbf{Q}^{\mathrm{T}} + \mathbf{F} \tag{6}$$

where $\mathbf{T} \in \mathbb{R}^N \times \mathbb{R}^A$ and $\mathbf{U} \in \mathbb{R}^N \times \mathbb{R}^A$ are called scores matrices, whose rows represent the projection of input and output observations (respectively) onto the space of LVs, $\mathbf{P} \in \mathbb{R}^{V_X} \times \mathbb{R}^A$ and $\mathbf{Q} \in \mathbb{R}^{V_Y} \times \mathbb{R}^A$ are called loading matrices, and $\mathbf{E} \in \mathbb{R}^N \times \mathbb{R}^{V_X}$ and $\mathbf{F} \in \mathbb{R}^N \times \mathbb{R}^{V_Y}$ are residual matrices of the models of $\mathbf{X}$ and $\mathbf{Y}$ (respectively); the symbol $\cdot$ represents the row-by-column matrix product.

LVs are identified in such a way as to optimize the linear relationship between corresponding columns of the score matrices. To this end, a weight matrix for inputs, $\mathbf{W} \in \mathbb{R}^{V_X} \times \mathbb{R}^A$, is computed column by column; for example, the first weight is the eigenvector of the cross-covariance $\mathbf{X}^{\mathrm{T}}\cdot\mathbf{Y}\cdot\mathbf{Y}^{\mathrm{T}}\cdot\mathbf{X}$ corresponding to its largest eigenvalue (Wold et al., 2001). A matrix of corrected weights, $\mathbf{W}^* \in \mathbb{R}^{V_X} \times \mathbb{R}^A$, is then defined as:

$$\mathbf{W}^* = \mathbf{W}\cdot\left(\mathbf{P}^{\mathrm{T}}\cdot\mathbf{W}\right)^{-1} \tag{7}$$

and used to compute the input scores:

$$\mathbf{T} = \mathbf{X}\cdot\mathbf{W}^* \tag{8}$$

Finally, an optimal linear relationship between corresponding columns of $\mathbf{T}$ and $\mathbf{U}$ is obtained solving the linear regression problem:

$$\mathbf{U} = \mathbf{T}\cdot\mathrm{diag}(\mathbf{b}) + \mathbf{H} \tag{9}$$

where $\mathrm{diag}(\mathbf{b})$ is a diagonal matrix containing vector $\mathbf{b} \in \mathbb{R}^A$, called inner regression coefficients, on the main diagonal, while $\mathbf{H} \in \mathbb{R}^N \times \mathbb{R}^A$ is the residual matrix of the regression model between scores. Once A has been set by the user, all entities involved in the model (i.e., $\mathbf{T}$, $\mathbf{U}$, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{W}$, and $\mathbf{b}$) are computed by any PLS calibration algorithm to respect all model equations. The reader may consult the papers by Geladi and Kowalski (1986) and by Wold et al. (2001) for more information.

After calibration, the PLS model can be used to compute $\hat{\mathbf{y}}_{\mathrm{new}}$, which is an estimate of the unknown output observation $\mathbf{y}_{\mathrm{new}} \in \mathbb{R}^{V_Y}$, given the corresponding input observation $\mathbf{x}_{\mathrm{new}} \in \mathbb{R}^{V_X}$. The input observation is first projected onto the space of input LVs by Eq. (8):

$$\mathbf{t}_{\mathrm{new}}^{\mathrm{T}} = \mathbf{x}_{\mathrm{new}}^{\mathrm{T}}\cdot\mathbf{W}^* \tag{10}$$

The best estimate of the output scores in a least-squares sense, $\hat{\mathbf{u}}_{\mathrm{new}}^{\mathrm{T}}$, is then obtained using the inner regression coefficients as in Eq. (9):

$$\hat{\mathbf{u}}_{\mathrm{new}}^{\mathrm{T}} = \mathbf{t}_{\mathrm{new}}^{\mathrm{T}}\cdot\mathrm{diag}(\mathbf{b}) \tag{11}$$

Finally, the estimate of the output observation is computed as for Eq. (6):

$$\hat{\mathbf{y}}_{\text{new}}^{\text{T}} = \hat{\mathbf{u}}_{\text{new}}^{\text{T}} \cdot \mathbf{Q}^{\text{T}} \tag{12}$$

Eqs. (10) to (12) can be summarized in a single equation directly relating $\mathbf{x}_{\text{new}}$ and $\hat{\mathbf{y}}_{\text{new}}$:

$$\hat{\mathbf{y}}_{\text{new}}^{\text{T}} = \mathbf{x}_{\text{new}}^{\text{T}} \cdot \left(\mathbf{W}^{*} \cdot \text{diag}(\mathbf{b}) \cdot \mathbf{Q}^{\text{T}}\right) = \mathbf{x}_{\text{new}}^{\text{T}} \cdot \mathbf{B} \tag{13}$$

where $\mathbf{B} \in \mathbb{R}^{V_X} \times \mathbb{R}^{V_Y}$ is the matrix of PLS (outer) regression coefficients.

3.2. Estimation of membrane resistances

Permeation and filtration in membranes are generally modelled through the transport theory. A thorough treatment of such a theory is out of the scope of this study; interested readers are referred to notable literature resources (Baker, 2004; Mulder, 1996; Spiegler and Kedem, 1966; Vilker et al., 1984).

Filtration through porous membranes can be described by the pore-flow model, i.e., Darcy's equation (Darcy, 1856) in its integral form (Meindersma et al., 1997):

$$v = \frac{1}{\mu R} \Delta P \tag{14}$$

where v is the permeate volume-flux (permeate flow rate divided by the flow surface), μ is the dynamic viscosity of the permeate, R is the membrane resistance to flow, ΔP is the TMP (average pressure difference between the feed side and the permeate side of the membrane). Osmotic pressure is usually neglected for membrane filtration of fermentation broths because it is negligible for solutes with high molecular weight (Vilker et al., 1984; Wankat, 2009). Given the available online measurements, Eq. (14) allows to compute online the average resistance of the sequence of membrane modules in Fig. 1. The resistances of each module can also be computed using offline measurements and applying Eq. (14) to each membrane module l:

$$R_l = \frac{1}{\mu_l v_l} \Delta P_l, l \in \{1, \cdots, 7\} \tag{15}$$

The resistance computed in Eq. (15) represents the overall resistance of the membrane, including contributions from both reversible and irreversible fouling. However, since the frequency of offline measurements is low, the trend of the overall resistance does not allow to properly monitor reversible fouling. Therefore, we seek to obtain high-frequency estimates of the resistances of each membrane module in Fig. 1. A hybrid estimation approach is proposed, combining data-driven and knowledge-driven model components. Namely, offline observations of TMPs for each module, together with the corresponding observations of a subset of downsampled online and computed variables, are first used to calibrate a PLS model that estimates the TMP; subsequently, the model is used to obtain high-frequency estimates of TMPs thanks to online variables at their original sampling rate (details on the PLS model calibration and assessment are provided in Section 4.1). This corresponds to the data-driven component of the hybrid estimation approach. The knowledge-driven component is made by Eq. (15), which is used to estimate the module resistances once the TMP estimates are available. The logic of the soft sensor is graphically represented in Fig. 4.

Viscosity measurements are not available online. Expert knowledge from plant operators is leveraged to assume a reasonable value for viscosity. Water-like behavior is postulated for permeate viscosities on all modules, and the effect of pressure and composition variation is deemed negligible. The temperature dependence of viscosity is modeled empirically (Perry and Green, 2008):

$$\mu_l = \exp\left(-52.843 + \frac{3703.6}{T_l} + 5.866\ln(T_l) - \left(5.879 \cdot 10^{-29}\right) T_l^{10}\right) \tag{16}$$

where temperature is in [K] and viscosity is obtained in [Pa s]. Note that, even if the aforementioned assumptions on viscosity might seem strong, the purpose of the soft sensor is not to yield extremely accurate estimate of the membrane resistance, but to accurately represent its trend so as to allow to properly monitor the process.

4. Results and discussion

Results of the study are presented in this section. Section 4.1 presents the workflow for PLS model development and assessment, which are entirely achieved on MATLAB R2022a (The Mathworks, 2022) with in-house-developed code. Sections 4.2 and 4.3 discuss the advantages of using resistances rather than fluxes and TMPs for process monitoring.

4.1. PLS model calibration and assessment

Direct prediction of membrane resistance by data-driven modeling

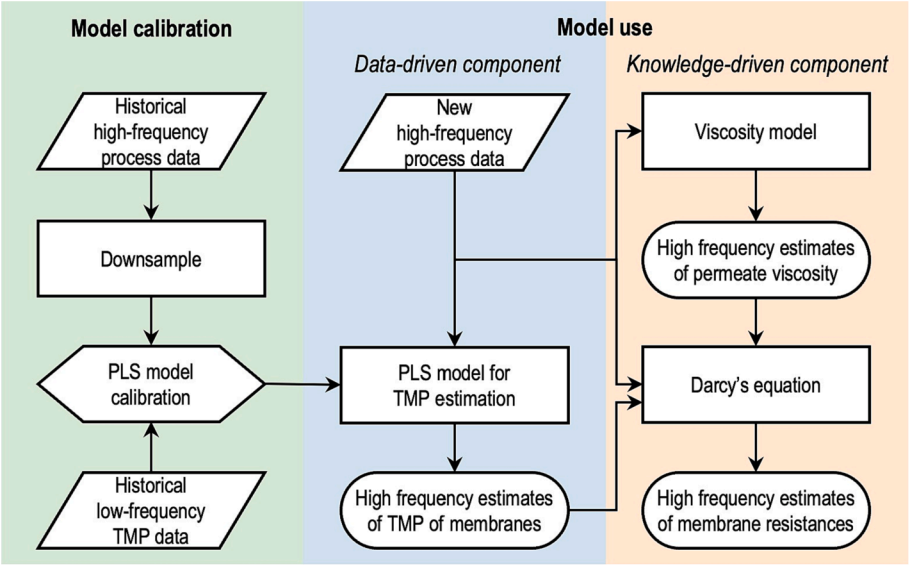

Fig. 4. Logic of the proposed soft sensor. The middle column represents the data-driven component, while the column on the right represents the knowledge-driven component.

usually requires strongly nonlinear models (Bagheri et al., 2019; Velidandi et al., 2023), while linear modeling may require to carry out ad-hoc experiments to acquire sophisticated measurements, e.g., concentration of foulants (Philippe et al., 2013). In fact, preliminary tests with a linear data-driven model (PLS) to directly estimate resistances yielded very unsatisfactory results: residuals featured a clear trend not captured by the model, thus denoting the need for a nonlinear model (results are not shown for brevity); furthermore, significant autocorrelation was detected in the residuals, indicating that the process dynamics was not modelled. Therefore, as discussed in the previous section, we tackle the problem by using a hybrid modeling approach: namely, we use a linear data-driven model (PLS) in the form of a soft sensor to provide estimations of TMPs from the available measurements; then, we couple this model to a simple nonlinear knowledge-driven model (Darcy's equation) to estimate the individual resistances from TMPs. A somewhat similar approach, yet in a different context, was used by Kaneko and Funatsu (2013). The advantages of using a linear data-driven model is that, compared to a nonlinear data-driven model, model calibration is simplified and model robustness is improved.

With respect to the PLS model, TMPs of membrane modules, computed by Eq. (4) from observations of offline (i.e., low frequency) variables, are regarded as output variables, while input variables are the corresponding observations of a subset of online and computed variables, as reported in Table 1.

To build the PLS model, the online observations are downsampled to match the timestamps of online and offline observations; more sophisticated approaches to multi-rate modeling (Lin et al., 2009) were not needed for this study. Observations from all runs are stacked vertically to obtain the data matrices to be processed by the modeling algorithm. Stating the same in a multivariate batch data analysis parlance, the structures of data matrices is the variable-wise unfolding one (Lee et al., 2004). Cleaning periods are not included in the data matrices, as well as the startup and shutdown phases of each run. These preprocessing operations results in a $\mathbf{X}$ matrix with dimension $[1621 \times 30]$ and a $\mathbf{Y}$ matrix with dimension $[1621 \times 7]$.

The output variables feature a remarkable correlation (the minimum value of correlation is 0.9057), as can be inferred from the correlation matrix reported in Table 2. The correlation among TMPs reflects the action of the control system, which adjusts the TMPs to adapt to the fouling state of each individual membrane. This guarantees that the overall permeate flow rate matches the assigned set-point, while the permeate flow rates of individual modules are free to change according to the fouling state of each membrane. Since it is desirable to model such valuable information on the interaction among the modules, PLS is a natural choice due to its well-known ability to exploit correlation in inputs and outputs formulating LVs (Burnham et al., 1996).

Table 1
Input and output variables of the PLS soft sensor.

ID	Name	Symbol	Type
Input variables			
1	Feed flow rate	$\dot{V}^{\mathrm{F}}$	Measured
2	Retentate flow rate	$\dot{V}^{\mathrm{R}}$	Measured
3	Permeate flow rate	$\dot{V}^{\mathrm{P}}$	Computed by Eq. (1)
4	Diafiltration solvent flow rate	$\dot{V}^{\mathrm{D}}$	Measured
5	Overall TMP	ΔP	Computed by Eq. (3)
6	Feed pressure	P^{F}	Measured
7	Retentate pressure	P^{R}	Measured
8	Permeate pressure	P^{P}	Measured
9	VCR	VCR	Computed by Eq. (2)
10–16	Permeate flow rate of modules 1 to 7	$\dot{V}_l^{\mathrm{P}}$, with $l \in \{1, \cdots, 7\}$	Measured
17–23	Temperature of modules 1 to 7	T_l, with $l \in \{1, \cdots, 7\}$	Measured
24–30	Pump power of modules 1 to 7	W_l, with $l \in \{1, \cdots, 7\}$	Measured
Output variables			
1–7	TMP of modules 1 to 7	ΔP_l, with $l \in \{1, \cdots, 7\}$	Computed by Eq. (4)

Table 2
Correlation matrix of the TMPs (output variables of the PLS soft sensor) for the seven membrane modules.

	ΔP_1	ΔP_2	ΔP_3	ΔP_4	ΔP_5	ΔP_6	ΔP_7
ΔP_1	1.0000	0.9485	0.9323	0. 9196	0.9141	0.9183	0.9057
ΔP_2	0.9485	1.0000	0.9570	0.9434	0.9424	0.9390	0.9417
ΔP_3	0.9323	0.9570	1.0000	0.9664	0.9692	0.9567	0.9417
ΔP_4	0.9196	0.9434	0.9664	1.0000	0.9678	0.9560	0.9432
ΔP_5	0.9141	0.9424	0.9692	0.9678	1.0000	0.9600	0.9561
ΔP_6	0.9183	0.9390	0.9567	0.9560	0.9600	1.0000	0.9561
ΔP_7	0.9057	0.9216	0.9417	0.9432	0.9561	0.9561	1.0000

The available data feature a dynamic component (due to the effects of fouling and the control system), which naturally calls for the use of dynamic PLS (Baffi et al., 2000; Dong and Qin, 2015; Ricker, 1988; Zhu, 2021). However, lagged matrices (on either data or scores) would then be required, which would set a constraint on the time step of observations in the prediction phase. Furthermore, lagged matrices require an even time step, while available data are sampled irregularly. Therefore, a static PLS model is developed. This allows one to calibrate the soft sensor with low sampling rate data, and to use it with high sampling rate data in the prediction phase. The choice of a static model for dynamic data is also backed up by two additional considerations. Firstly, LV models can capture dynamic information in data using additional LVs (Vanhatalo and Kulahci, 2016). Secondly, dynamic modelling is not necessarily needed to capture the input/output relation, as the dynamics on outputs could be solely induced by dynamics on inputs (Sun and Braatz, 2021).

A preliminary PLS model is fitted on autoscaled data matrices by the SIMPLS algorithm (de Jong, 1993) selecting the number of LVs by repeated k-fold cross validation (Burman, 1989; Geisser, 1975) with one-standard-error-rule (Filzmoser et al., 2009; Hastie et al., 2009). The performance index is defined as the average of root-mean squared errors on all seven output variables, which results in nine LVs. Data from nine runs are removed due to high values of the squared prediction error (SPE) for reconstruction of input data (Nomikos and MacGregor, 1995), while data from three more runs are deemed as outliers and discarded due to high leverages (Berber and Akcay, 2005; Rousseeuw and Van Zomeren, 1990). The preliminary PLS model is re-calibrated after removal of observations, which reduces the number or rows in the $\mathbf{X}$ and $\mathbf{Y}$ matrices to 1579; cross-validation results in twelve LVs.

Residuals of the new model feature remarkable dynamics, tested by the significance of coefficients of the autocorrelation function (ACF; Box et al., 2016) on 95 % confidence limits computed by Bartlett's formula (Bartlett, 1946), as can be seen in Fig. 5(a). Therefore, additional LVs are included in the model to remove as much residual autocorrelation as possible (i.e., to improve the dynamics captured by the PLS model), while preserving smoothness of prediction. Such tuning is aided by visual assessment of ACFs of the seven outputs and of predictions, and results in a final PLS model with twenty latent variables as best compromise. Fig. 5(b) proves that most of the residual autocorrelation is removed. This approach was attempted also on the preliminary tests with the purely data-driven model for direct estimation of resistances, but significant dynamics was still left in the residuals even using all the available LVs. This further proves the value of the proposed hybrid modelling approach.

The generalization performance of the model are investigated by means of nested cross-validation (Varma and Simon, 2006) employing repeated k-fold splitting in both the inner and outer loops (Filzmoser

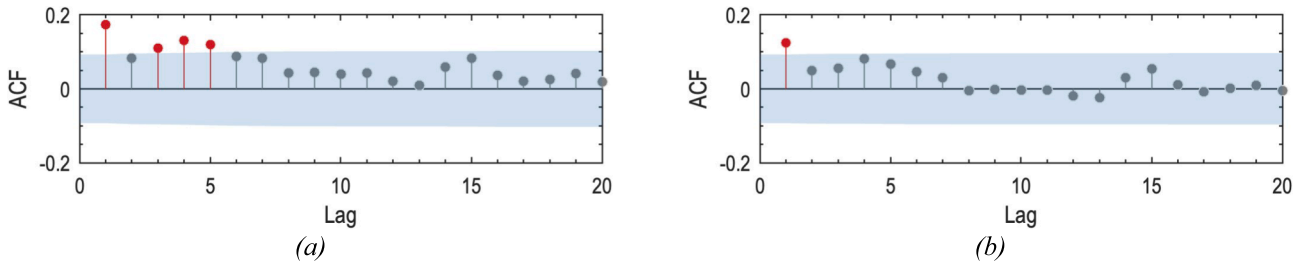

Fig. 5. Autocorrelation function of residuals for an example output variable of the (a) preliminary PLS model with twelve LVs and (b) final PLS model with twenty LVs. Significant coefficients are represented as red dots outside of the shaded envelope of the 95% confidence interval.

et al., 2009). Such a tool was also used to make sure that the manual tuning does not deteriorate generalization performance. Average determination coefficients of the preliminary and final PLS model in calibration, validation, and testing are reported in Table 3. The final model shows a satisfactory generalization performance. Furthermore, determination coefficients in calibration and testing are similar, denoting that the model does not overfit even after manual tuning of the number of LVs.

An example of prediction of the final PLS model on high frequency online data is shown in Fig. 6. The model successfully captures dynamic effects of both reversible and irreversible fouling. The effect of membrane replacements are excellently reconstructed, as can be seen in Fig. 6: for instance, considering module 2, predictions jump down after the membrane replacement occurred at observation no. 1650, and this follows the trend of experimental data. Predictions tend to be unreliable at the very beginning of a run, which is especially clear considering the seventh module; however, they tend to realign to the actual TMPs in a relatively short time (which typically corresponds to ~2 h). Prediction reliability can be assessed by means of the PLS model monitoring statistics, namely Hotelling's T^2 and SPE (Nomikos and MacGregor, 1995). Predictions are deemed as unreliable when SPE is beyond its 95 % control limit, while a T^2 beyond the limit denote that the process is drifting far from the average conditions. Fig. 7 reports such statistics in logarithmic form and allows one to clearly identify anomalous tails at the beginning of runs after observation no. 1950 in Fig. 6, which are associated to statistics well beyond their control limits.

4.2. Membrane resistances to monitor short-term fouling trends

The TMPs estimated by the PLS model are plugged in Eq. (15) with online measurements of permeate fluxes and temperatures to estimate resistances of all membrane modules online. Resistances allow for a better monitoring and understanding of the filtration process as opposed to fluxes and TMPs. Some examples concerning the monitoring of short-term fouling are discussed in this Section.

The increase in interpretability is clear from Fig. 8, reporting profiles of permeate flux, TMP, and resistance for one of the separation modules during two example batches ((a) and (b) in the figure). In Fig. 8(a) the estimated TMP features an increasing trend, whereas the permeate flux exhibits an erratic behavior with both fast and slow variations, which somewhat casts doubt on the interpretation of the TMP behavior. However, the estimated resistance features a well-defined, mostly monotonic trend, which allows one to unambiguously monitor the evolution of reversible fouling along the batch for the membrane under investigation. A significant fouling event is visible around batch time equal to 0.8, where the resistance steps up and then steadily increases thereafter; in fact, this specific batch had to be interrupted soon after that event due to significant pressure build-up. Fig. 8(b) highlights how membrane resistance can capture the occurrence of hydraulic cleaning within a batch. In this case, the material being filtered sedimented on the membrane surface in the initial part of the batch. Such deposition was removed due an increase in cross-flow velocity dictated by the control system shortly after batch time 0.3, which caused a decrease of the resistance. The increased flux however eventually enhanced membrane fouling, as can be argued from the rapid increase of the resistance after batch time 0.6.

For this particular example, the trends of membrane resistance and TMP are not too different, and one might think that both variables are equally effective to monitor short-term membrane fouling. However, that is not generally the case, and in fact monitoring resistances, rather than TMPs, offers significant advantages from the process understanding point of view. To appreciate this, recall that the plant layout (Fig. 1) consists in several separation modules, but only the overall permeate flow rate is controlled, and this is achieved by manipulating the overall feed pressure. Therefore, a linear constraint acts on permeate fluxes from single modules, which are therefore not independent and must compensate for each other. Such a strong correlation between fluxes makes it difficult to trace flux variations back to the fouling state of each single membrane, and this can impact also TMP trends. Furthermore, promptly identifying fouling events that act on single modules is cumbersome or impossible if done by visual inspection of the recorded data, due to the low frequency of offline readings. However, both issues are fixed if membrane resistance is monitored by the proposed approach, as elucidated in Fig. 9 for two different batches ((a) and (b)). The effect of interdependence of permeates fluxes is clear from Fig. 9(a), where fluxes of modules 2 to 7 increase to make up for the decrease in flux of module 1. However, TMPs of modules 2 to 7 also increase, which makes it difficult to conjecture anything about the fouling state of membranes. On the other hand, resistances allow one to clearly understand that modules 1 and 2 are the ones mostly suffering from reversible fouling in this batch. Similar considerations can be done for the batch illustrated in Fig. 9(b). Additionally, this figure enables one to appreciate that the onset of significant fouling events affecting single modules becomes clear if resistances (rather than TPMs or fluxes) are monitored. These events occurred in module 5 and 7 shortly after the batch start, and in module 6 around batch time equal 0.5, where the slope of the resistance changes.

Table 3
Average determination coefficients R^2 of the preliminary and final PLS models in calibration, validation, and testing estimated by nested cross-validation.

Model	LVs	Calibration	Validation	Testing
Preliminary	12	0.9012	0.8979	0.8977
Final	20	0.9047	0.9005	0.9002

4.3. Membrane resistance to monitor long-term fouling trends

The proposed model is helpful also in the analysis of long-term trends in fouling, which can be obtained by computing averages of estimated resistances from the profiles over batches. While this is possible also using low frequency offline measurements, the high frequency estimations of resistances offer stronger reliability and increased robustness to outliers, thus allowing one to properly visualize and monitor long-term fouling trends for each single membrane module. We denote such batch-

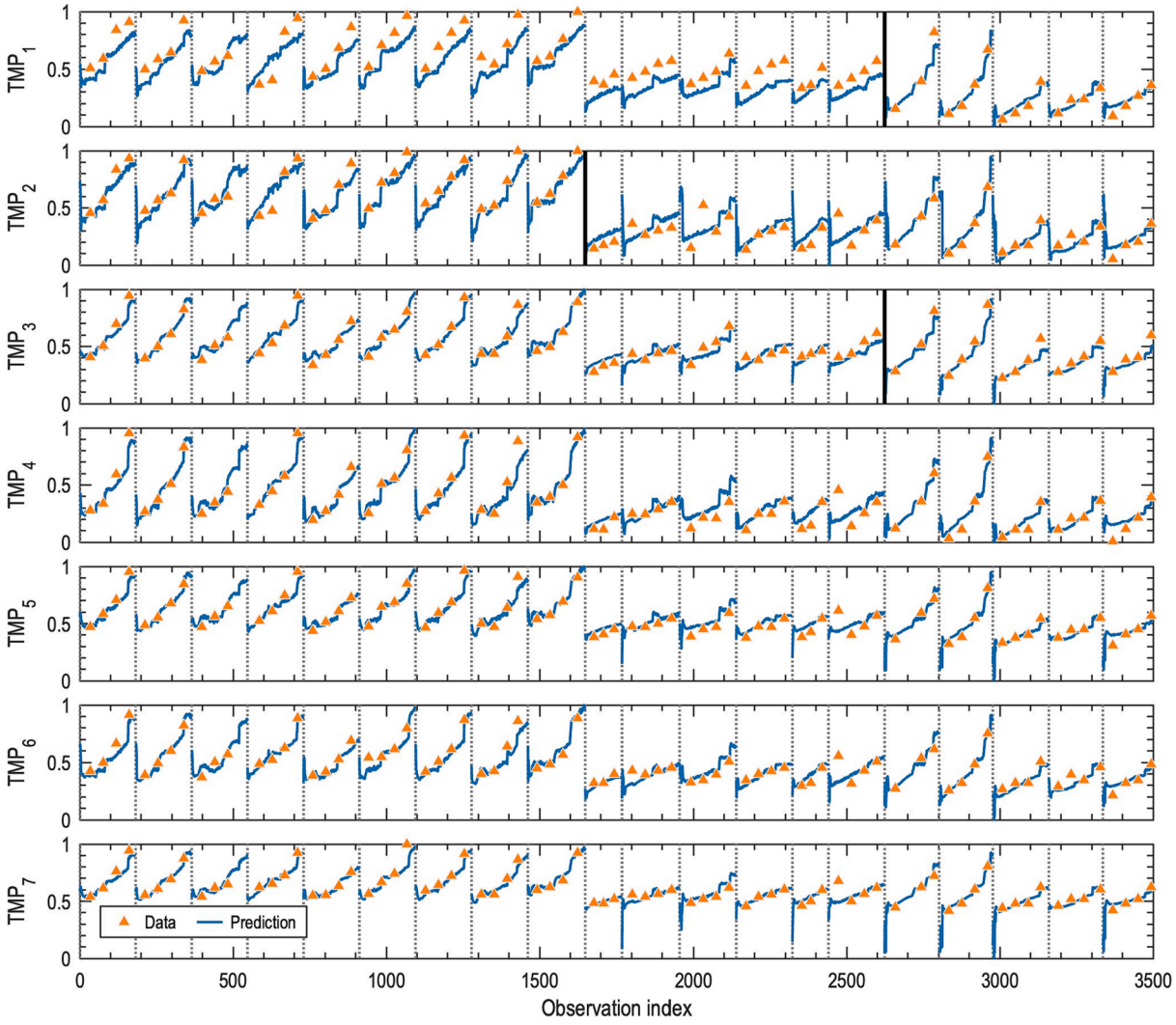

Fig. 6. Example of predictions of TMPs by the final PLS model. Low frequency, offline measurements are represented as orange triangles, while high frequency, online estimates are blue solid lines. Dotted lines delimit single process runs, while vertical solid black lines separate runs with a replacement of the membrane of the relevant module in between.

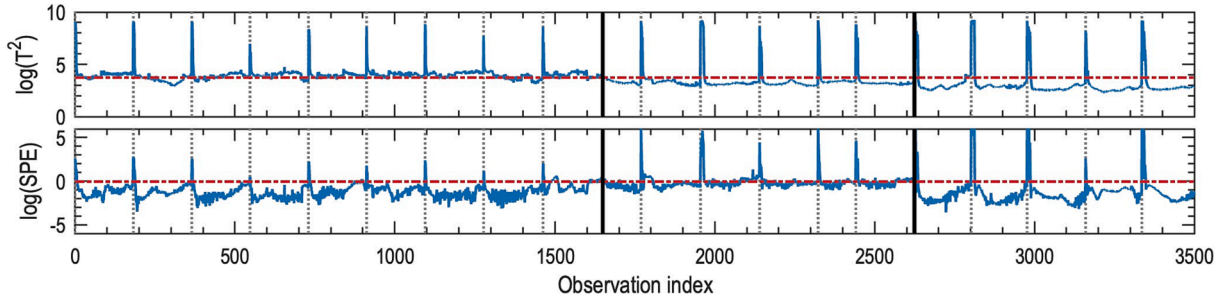

Fig. 7. Example of monitoring statistics to diagnose the reliability of predictions of the final PLS model; results refer to the predictions of Fig. 6. 95% control limits are represented as red dash-dotted lines, vertical dotted lines delimit single process runs, and vertical solid black lines separate runs with a replacement of the membrane of any module in between.

averaged variables as features in the following.

The most prominent advantage of using estimated resistances in the context of long-term fouling investigation is the possibility to decouple membrane aging effects from flux interdependencies across modules. As an example, consider Fig. 10, which illustrates the trends of the permeate flux, TMP, and resistance features for one separation module across several consecutive batches. The end-of-life replacement of the module membrane occurs at batch no. 112 and is indicated by a solid black line in the figure. This causes the flux to step up in that module. However, due to the interdependence of fluxes across modules, the other fluxes adjust accordingly (often stepping down). On the other hand, TMP values are so strongly affected by variability across batches that membrane replacement passes almost unnoticed. We conclude that analyzing fluxes or TMPs confounds the diagnosis of membrane health.

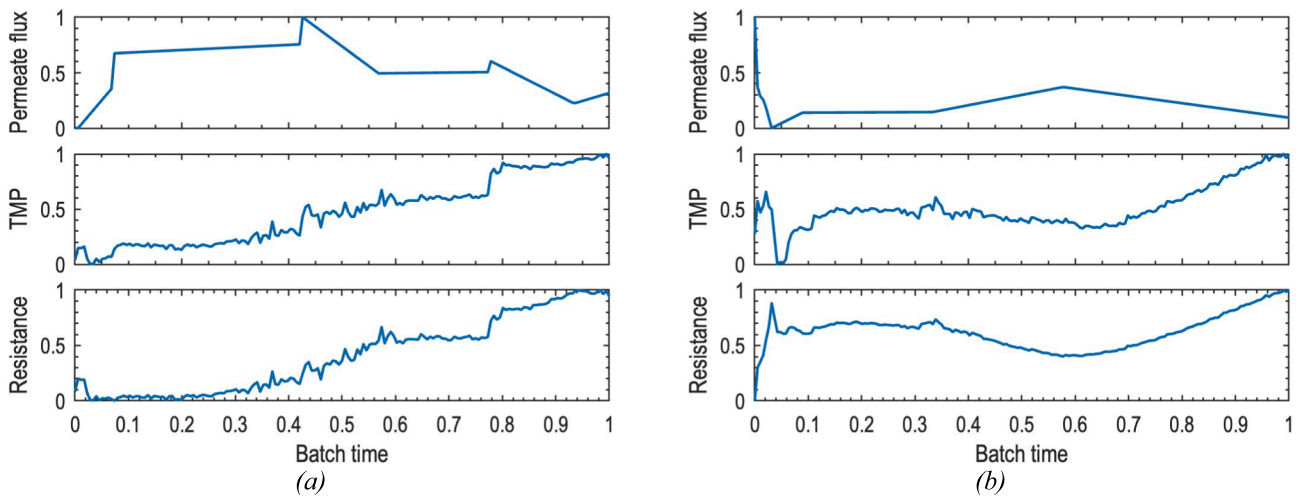

Fig. 8. Time profiles of fouling-related variables for one membrane module to highlight the increased interpretability of (a) reversible fouling trend in one batch, and (b) within-batch hydraulic cleaning phenomena due to the control system in a different batch.

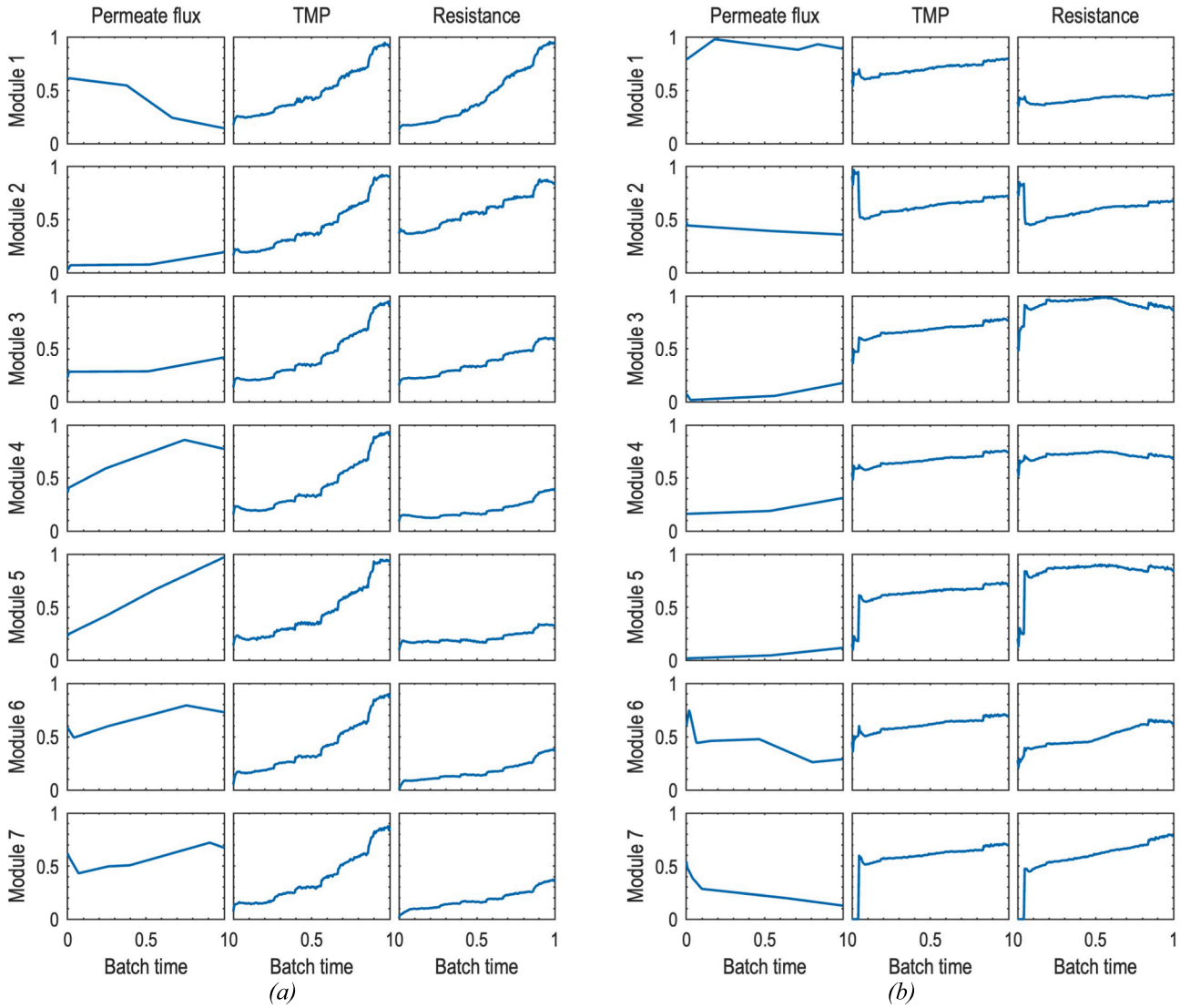

Fig. 9. Time profiles of fouling-related variables for all membrane modules to highlight (a) the increased interpretability due to the monitoring of membrane resistances in one batch, and (b) the identification of significant fouling events acting on a single module in a different batch.

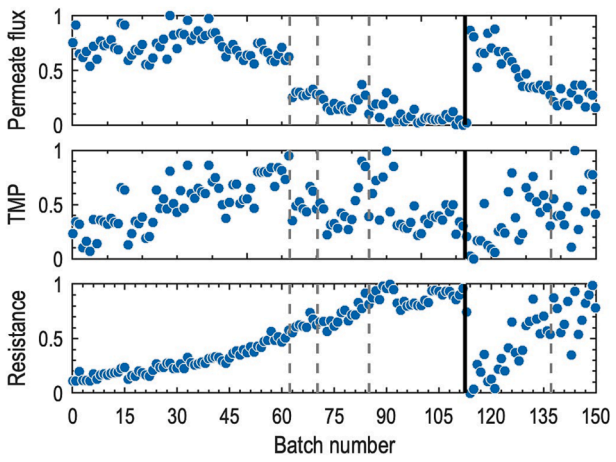

Fig. 10. Long-term trends of permeate flux, TMP, and resistance features of one membrane module across several consecutive batches. Each dot represents the average of profiles of the relevant variable on a batch. The vertical solid black lines indicate replacement of the membrane of the same module, while vertical dashed lines are membrane replacements of other modules.

Indeed, the true health state of the membrane is represented by its resistance, which clearly decreases right after replacement, thus facilitating monitoring. One additional advantage of monitoring the membrane resistance is that it does not change when the membranes in other modules are replaced (dashed lines in Fig. 10); conversely, both TMP and flux are affected by the replacement.

The last, less apparent advantage of the proposed approach regards the decoupling of reversible and irreversible fouling, which can be achieved by combining the proposed model with knowledge of the process operation rationale. Features computed by averaging resistances over an entire batch mix up the effects of reversible and irreversible fouling into one single indicator. Intuitively, irreversible fouling is represented by the "baseline resistance" of a membrane at the beginning of a batch, and the within-batch increase of resistance above the baseline is due to reversible fouling. This within-batch variation is compensated for by the process control systems, which adjusts the cross-flow velocity on discrete levels in response to reversible fouling, as shown in Fig. 11: while the first two levels are always reached under normal operation, the following levels are enforced only when the feed pressure rises above preset thresholds. Reversible fouling is believed to be under control (in the plant being considered) when the cross-flow velocity is set to the two/three lowest levels.

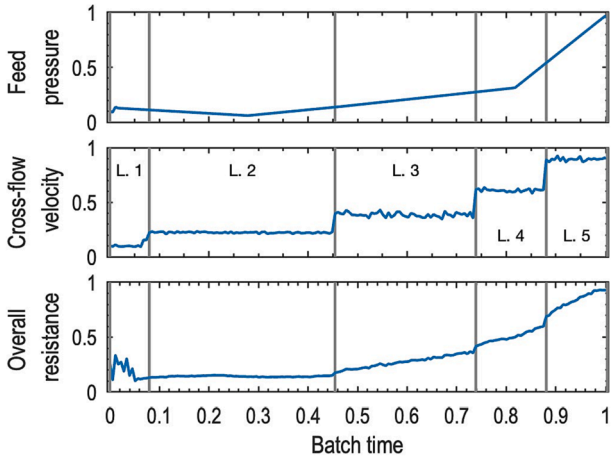

Fig. 11. Time profiles of fouling-related variables for one membrane module during one batch to illustrate the rationale of the variation of the cross-flow velocity on discrete levels (L.) to counteract the effects of reversible fouling. Time periods with constant cross-flow velocity are delimited by vertical lines.

The outlined rationale can be exploited to decompose the overall resistance feature in a batch into its contributions from irreversible and reversible fouling. The contribution due to irreversible fouling is computed as the average of the resistance profile where the cross-flow velocity is set to the three lowest levels; the contribution due to reversible fouling is computed as the difference between the average of the resistance profile along the entire batch and the contribution due to irreversible fouling calculated as above. This decomposition is exemplified in Fig. 12 for the membrane of one module. The overall resistance shows a clear long-term trend, which is associated to irreversible fouling and provides indirect indication of how the membrane state changes across a production campaign. However, batch-to-batch variability, due to the effect of within-batch reversible fouling, somewhat confounds the across-batch trend, particularly near membrane replacement (shortly before batch no. 70). The proposed resistance feature decomposition allows one to decouple the two effects: the resistance feature for irreversible fouling is affected by a much smaller variability, while the resistance feature for reversible fouling allow one to clearly identify batches that suffered from intense fouling, whose profiles can therefore undergo additional, in-depth investigation.

5. Conclusions

In this study, we developed a hybrid modeling strategy to estimate individual resistances of a multi-module membrane separation system in an industrial biorefinery that manufactures 1,4-bio-butanediol from fermentation of sustainable biomass. We combined a partial least-squares regression model, for online estimation of trans-membrane pressure (TMP) of each membrane module, and Darcy's equation for modeling of membrane resistances. The proposed modelling strategy achieved excellent generalization performance using a linear data-driven modeling component, as opposed to the dominant literature approaches using nonlinear data-driven models that require massive datasets for calibration. To the best of the author's knowledge, this is the first time that such results were achieved on a complex industrial biorefinery process with limited data.

The main advantage of the proposed resistance-based monitoring approach for membrane fouling characterization are the reduced variability and increased interpretability of resistances with respect to permeate fluxes and TMPs. We illustrated examples on how resistances feature clear and defined dynamics, allowing one to properly infer the reversible fouling state of membranes and to promptly identify the onset

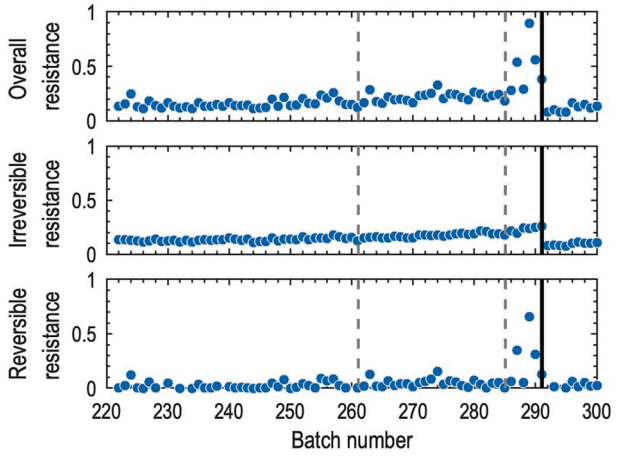

Fig. 12. Decomposition of overall resistance features (averages over batch profiles) of a module in their contributions form irreversible and reversible fouling. Vertical solid black lines indicate replacement of the membrane of the relevant module, while vertical dashed lines are membrane replacements of other modules.

of irreversible fouling. We showed how the resistance of a module is independent of the resistances of other modules (therefore of membrane replacements), a feature that fluxes lack when subject to linear constraints such as a control system controlling the overall permeate flow rate. Finally, we discussed how to aggregate resistances as batch-averages to monitor the long-term evolution of fouling, proposing a method to decompose the overall membrane resistance in contributions from reversible and irreversible fouling by leveraging process knowledge. Results show that the contribution due to irreversible fouling features a monotonic trend, while the contribution due to reversible fouling allows one to clearly identify batches that suffered from significant fouling issues, whose profiles can be analyzed more in-depth for diagnostic purposes.

CRediT authorship contribution statement

Elia Arnese-Feffin: Conceptualization, Methodology, Formal analysis, Data curation, Investigation, Software, Validation, Visualization, Writing – original draft. **Pierantonio Facco:** Conceptualization, Methodology, Writing – review & editing. **Daniele Turati:** Resources, Investigation, Writing – review & editing. **Fabrizio Bezzo:** Conceptualization, Methodology, Writing – review & editing. **Massimiliano Barolo:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Elia Arnese-Feffin reports financial support was provided by Novamont SpA, CARIPARO Foundation, UniSmart Foundation, and Intesa-SanPaolo. Daniele Turati reports a relationship with Novamont SpA that includes: employment.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors gratefully acknowledge financial support by Novamont S.p.A., CARIPARO Foundation, UniSmart Foundation, and Intesa-SanPaolo through the PhD Scholarship of E.A.F. The authors also wish to thank Ing. Diego Girotto and Ing. Giovanni Lembo (Novamont S.p.A., Bottrighe production plant) for valuable insight on the plant operation. Helpful discussion with Ing. Marco Cotti Comettini is also acknowledged.

References

Abels, C., Carstensen, F., Wessling, M., 2013. Membrane processes in biorefinery applications. Journal of Membrane Science 444, 285–317. https://doi.org/10.1016/j.memsci.2013.05.030.

Arnese-Feffin, E., Facco, P., Turati, D., Bezzo, F., Barolo, M., 2023. Troubleshooting high-pressure issues in an industrial biorefinery process by feature-oriented modeling, in: Proceedings of the 33rd European Symposium on Computer Aided Process Engineering (ESCAPE33). Presented at the ESCAPE33, Elsevier B.V., Athens (GR), pp. 163–168. https://doi.org/10.1016/B978-0-443-15274-0.50027-5.

Baffi, G., Martin, E.B., Morris, A.J., 2000. Non-linear dynamic projection to latent structures modelling. Chemometrics and Intelligent Laboratory Systems 52, 5–22. https://doi.org/10.1016/S0169-7439(00)00083-6.

Bagheri, M., Akbari, A., Mirbagheri, S.A., 2019. Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. Process Safety and Environmental Protection 123, 229–252. https://doi.org/10.1016/j.psep.2019.01.013.

Bähner, F.D., Prado-Rubio, O.A., Huusom, J.K., 2021. Challenges in Optimization and Control of Biobased Process Systems: An Industrial-Academic Perspective. Industrial and Engineering Chemistry Research 60, 14985–15003. https://doi.org/10.1021/acs.iecr.1c01792.

Baker, R.W., 2004. Membrane Technology and Applications, 2nd ed. Wiley, Chichester (UK).

Barello, M., Manca, D., Patel, R., Mujtaba, I.M., 2014. Neural network based correlation for estimating water permeability constant in RO desalination process under fouling. Desalination 345, 101–111. https://doi.org/10.1016/j.desal.2014.04.016.

Bartlett, M.S., 1946. On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. J. R. Stat. Soc. Ser. B Stat. Methodol. 8, 27–41. https://doi.org/10.2307/2983611.

Berber, R., Akcay, L., 2005. Monitoring and fault diagnosis by multivariate statistical methods in chemical processes, in: AIChE Annual Meeting, Conference Proceedings. p. 6855.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2016. Time Series Analysis. Wiley, Hoboken (NJ).

Burgard, A., Burk, M.J., Osterhout, R., Van Dien, S., Yim, H., 2016. Development of a commercial scale process for production of 1,4-butanediol from sugar. Current Opinion in Biotechnology 42, 118–125. https://doi.org/10.1016/j.copbio.2016.04.016.

Burman, P., 1989. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika 76, 503–514. https://doi.org/10.1093/biomet/76.3.503.

Burnham, A.J., Viveros, R., Macgregor, J.F., 1996. Frameworks for latent variable multivariate regression. Journal of Chemometrics 10, 31–45. https://doi.org/10.1002/(SICI)1099-128X(199601)10:1<31::AID-CEM398>3.0.CO;2-1.

Carstensen, F., Apel, A., Wessling, M., 2012. In situ product recovery: Submerged membranes vs. external loop membranes. Journal of Membrane Science 394–395, 1–36. https://doi.org/10.1016/j.memsci.2011.11.029.

Chan, L.L.T., Chou, C.-P., Chen, J., 2017. Hybrid model based expected improvement control for cyclical operation of membrane microfiltration processes. Chemical Engineering Science 166, 77–90. https://doi.org/10.1016/j.ces.2017.02.048.

Chen, F., Peldszus, S., Peiris, R.H., Ruhl, A.S., Mehrez, R., Jekel, M., Legge, R.L., Huck, P.M., 2014. Pilot-scale investigation of drinking water ultrafiltration membrane fouling rates using advanced data analysis techniques. Water Research 48, 508–518. https://doi.org/10.1016/j.watres.2013.10.007.

Chew, C.M., Aroua, M.K., Hussain, M.A., 2017. A practical hybrid modelling approach for the prediction of potential fouling parameters in ultrafiltration membrane water treatment plant. Journal of Industrial and Engineering Chemistry 45, 145–155. https://doi.org/10.1016/j.jiec.2016.09.017.

Cuellar, M.C., Straathof, A.J., 2020. Downstream of the bioreactor: advancements in recovering fuels and commodity chemicals. Current Opinion in Biotechnology 62, 189–195. https://doi.org/10.1016/j.copbio.2019.11.012.

Darcy, H., 1856. Les Fontaines publiques de la ville de Dijon. V. Dalmont, Paris (FR).

de Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems 18, 251–263. https://doi.org/10.1016/0169-7439(93)85002-X.

Delgenes, J., 1996. Comparative study of separated fermentations and cofermentation processes to produce ethanol from hardwood derived hydrolysates. Biomass and Bioenergy 11, 353–360. https://doi.org/10.1016/0961-9534(96)00019-0.

Dologlu, P., Sildir, H., 2022. Data driven identification of industrial reverse osmosis membrane process. Computers and Chemical Engineering 161, 107782. https://doi.org/10.1016/j.compchemeng.2022.107782.

Dong, Y., Qin, S.J., 2015. Dynamic-inner partial least squares for dynamic data modeling. IFAC-Pap. 28, 117–122. https://doi.org/10.1016/j.ifacol.2015.08.167.

Ennaceri, H., Fischer, K., Schulze, A., Moheimani, N.R., 2022. Membrane fouling control for sustainable microalgal biodiesel production: A review. Renewable and Sustainable Energy Reviews 161, 112335. https://doi.org/10.1016/j.rser.2022.112335.

Filzmoser, P., Liebmann, B., Varmuza, K., 2009. Repeated double cross validation. Journal of Chemometrics 23, 160–171. https://doi.org/10.1002/cem.1225.

Geisser, S., 1975. The predictive sample reuse method with applications. Journal of the American Statistical Association 70, 320–328. https://doi.org/10.1080/01621459.1975.10479865.

Geissler, S., Wintgens, T., Melin, T., Vossenkaul, K., Kullmann, C., 2005. Modelling approaches for filtration processes with novel submerged capillary modules in membrane bioreactors for wastewater treatment. Desalination 178, 125–134. https://doi.org/10.1016/j.desal.2004.11.032.

Geladi, P., Kowalski, B.R., 1986. Partial Least-Squares Regression: A Tutorial. Analytica Chimica Acta 185, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9.

Gerardo, M.L., Oatley-Radcliffe, D.L., Lovitt, R.W., 2014. Integration of membrane technology in microalgae biorefineries. Journal of Membrane Science 464, 86–99. https://doi.org/10.1016/j.memsci.2014.04.010.

Han, H.-G., Zhang, H.-J., Liu, Z., Qiao, J.-F., 2020. Data-driven decision-making for wastewater treatment process. Control Engineering Practice 96, 104305. https://doi.org/10.1016/j.conengprac.2020.104305.

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The Elements of Statistical Learning, 2nd ed. Springer, London (UK).

Huang, W., Zhu, Y., Wang, L., Lv, W., Dong, B., Zhou, W., 2021. Reversible and irreversible membrane fouling in hollow-fiber UF membranes filtering surface water: effects of ozone/powdered activated carbon treatment. RSC Advances 11, 10323–10335. https://doi.org/10.1039/D0RA09820E.

Hwang, T.-M., Oh, H., Choi, Y.-J., Nam, S.-H., Lee, S., Choung, Y.-K., 2009. Development of a statistical and mathematical hybrid model to predict membrane fouling and performance. Desalination 247, 210–221. https://doi.org/10.1016/j.desal.2008.12.025.

Ioannidou, S.M., Pateraki, C., Ladakis, D., Papapostolou, H., Tsakona, M., Vlysidis, A., Kookos, I.K., Koutinas, A., 2020. Sustainable production of bio-based chemicals and polymers via integrated biomass refining and bioprocessing in a circular bioeconomy context. Bioresource Technology 307, 123093. https://doi.org/10.1016/j.biortech.2020.123093.

Jiang, L.Y., Zhu, J.M., 2013. Separation technologies for current and future biorefineries—status and potential of membrane-based separation. WIREs Energy and Environment 2, 673–690. https://doi.org/10.1002/wene.73.
Julio, R., Albet, J., Vialle, C., Vaca-Garcia, C., Sablayrolles, C., 2017. Sustainable design of biorefinery processes: existing practices and new methodology. Biofuels Bioprod. Biorefining 11, 373–395. https://doi.org/10.1002/bbb.1749.
Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven Soft Sensors in the process industry. Computers and Chemical Engineering 33, 795–814. https://doi.org/10.1016/j.compchemeng.2008.12.012.
Kallioinen, M., Huuhilo, T., Reinikainen, S.P., Nuortila-Jokinen, J., Mänttäri, M., 2006. Examination of membrane performance with multivariate methods: A case study within a pulp and paper mill filtration application. Chemometrics and Intelligent Laborary Systems 84, 98–105. https://doi.org/10.1016/j.chemolab.2006.04.015.
Kaneko, H., Funatsu, K., 2013. A chemometric approach to prediction of transmembrane pressure in membrane bioreactors. Chemometrics and Intelligent Laborary Systems 126, 30–37. https://doi.org/10.1016/j.chemolab.2013.04.016.
Kaneko, H., Funatsu, K., 2015. Data density-based fault detection and diagnosis with nonlinearities between variables and multimodal data distributions. Chemometrics and Intelligent Laborary Systems 147, 58–65. https://doi.org/10.1016/j.chemolab.2015.07.016.
Klimkiewicz, A., Cervera-Padrell, A.E., van den Berg, F.W.J., 2016. Multilevel Modeling for Data Mining of Downstream Bio-Industrial Processes. Chemometrics and Intelligent Laborary Systems 154, 62–71. https://doi.org/10.1016/j.chemolab.2016.03.020.
Lee, J.M., Yoo, C.K., Lee, I.B., 2004. Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis. Journal of Biotechnology 110, 119–136. https://doi.org/10.1016/j.jbiotec.2004.01.016.
Lin, B., Recke, B., Schmidt, T.M., Knudsen, J.K.H., Jorgensen, S.B., 2009. Data-driven soft sensor design with multiple-rate sampled data: A comparative study. Industrial and Engineering Chemistry Research 48, 5379–5387. https://doi.org/10.1021/ie801084e.
Mancini, E., Mansouri, S.S., Gernaey, K.V., Luo, J., Pinelo, M., 2020. From second generation feed-stocks to innovative fermentation and downstream techniques for succinic acid production. Critical Reviews in Environmental Science and Technology 50, 1829–1873. https://doi.org/10.1080/10643389.2019.1670530.
Martín, M., Grossmann, I.E., 2013. On the systematic synthesis of sustainable biorefineries. Industrial and Engineering Chemistry Research 52, 3044–3064. https://doi.org/10.1021/ie2030213.
McCurdy, A.T., Higham, A.J., Morgan, M.R., Quinn, J.C., Seefeldt, L.C., 2014. Two-step process for production of biodiesel blends from oleaginous yeast and microalgae. Fuel 137, 269–276. https://doi.org/10.1016/j.fuel.2014.07.099.
Meindersma, G.W., Augeraud, J., Vergossen, F.H.P., 1997. Separation of a biocatalyst with ultrafiltration or filtration after bioconversion. Journal of Membrane Science 125, 333–349. https://doi.org/10.1016/S0376-7388(95)00081-X.
Mulder, M., 1996. Basic Principles of Membrane Technology, 2nd ed. Kluwer Academic Publisher, Dordrecht (NL).
Nomikos, P., MacGregor, J.F., 1995. Multi-way partial least squares in monitoring batch processes. Chemometrics and Intelligent Laborary Systems 30, 97–108. https://doi.org/10.1016/0169-7439(95)00043-7.
The Mathworks, Inc., 2022. MATLAB R2022a. Natick (MA).
Novamont S.p.A., 2016. Opening of the world's first industrial scale plant for the production of butanediol via fermentation of renewable raw materials [WWW Document]. URL https://novamont.it/eng/read-press-release/mater-biotech/ (accessed 4.1.23).
Perera, Y.S., Ratnaweera, D.A.A.C., Dasanayaka, C.H., Abeykoon, C., 2023. The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review. Engineering Applications of Artificial Intelligence 121, 105988. https://doi.org/10.1016/j.engappai.2023.105988.
Perry, R., Green, D.W., 2008. Perry's Chemical Engineers' Handbook, 8th ed. McGraw Hill, New York (NY).
Philippe, N., Stricker, A.-E., Racault, Y., Husson, A., Sperandio, M., Vanrolleghem, P., 2013. Modelling the long-term evolution of permeability in a full-scale MBR: Statistical approaches. Desalination 325, 7–15. https://doi.org/10.1016/j.desal.2013.04.027.
Piron, E., Latrille, E., René, F., 1997. Application of artificial neural networks for crossflow microfiltration modelling: "black-box" and semi-physical approaches. Computers and Chemical Engineering 21, 1021–1030. https://doi.org/10.1016/S0098-1354(96)00332-8.
Prochaska, K., Antczak, J., Regel-Rosocka, M., Szczygiełda, M., 2018. Removal of succinic acid from fermentation broth by multistage process (membrane separation and reactive extraction). Separation and Purification Technology 192, 360–368. https://doi.org/10.1016/j.seppur.2017.10.043.
Reis, M.S., Gins, G., Rato, T.J., 2019. Incorporation of process-specific structure in statistical process monitoring: A review. Journal of Quality Technology 51, 407–421. https://doi.org/10.1080/00224065.2019.1569954.
Rendall, R., Chiang, L.H., Reis, M.S., 2019. Data-driven methods for batch data analysis – A critical overview and mapping on the complexity scale. Computers and Chemical Engineering 124, 1–13. https://doi.org/10.1016/j.compchemeng.2019.01.014.
Ricker, N.L., 1988. The use of biased least-squares estimators for parameters in discrete-time pulse-response models. Industrial and Engineering Chemistry Research 27, 343–350. https://doi.org/10.1021/ie00074a023.
Rousseeuw, P.J., Van Zomeren, B.C., 1990. Unmasking Multivariate Outliers and Leverage Points. Journal of the American Statistical Association 85, 633–639. https://doi.org/10.1080/01621459.1990.10474920.
Rudolph, G., Virtanen, T., Ferrando, M., Güell, C., Lipnizki, F., Kallioinen, M., 2019. A review of in situ real-time monitoring techniques for membrane fouling in the biotechnology, biorefinery and food sectors. Journal of Membrane Science 588, 117221. https://doi.org/10.1016/j.memsci.2019.117221.
Ruiz-García, A., Nuez, I., 2016. Long-term performance decline in a brackish water reverse osmosis desalination plant. Predictive model for the water permeability coefficient. Desalination 397, 101–107. https://doi.org/10.1016/j.desal.2016.06.027.
Saha, K.R., Sikder, U.M.J., Chakraborty, S., Da Silva, S.S., Dos Santos, J.C., 2017. Membranes as a tool to support biorefineries: Applications in enzymatic hydrolysis, fermentation and dehydration for bioethanol production. Renewable and Sustainable Energy Reviews 74, 873–890. https://doi.org/10.1016/j.rser.2017.03.015.
Sansana, J., Joswiak, M.N., Castillo, I., Wang, Z., Rendall, R., Chiang, L.H., Reis, M.S., 2021. Recent trends on hybrid modeling for Industry 4.0. Computers and Chemical Engineering 151, 107365. https://doi.org/10.1016/j.compchemeng.2021.107365.
Satam, C.C., Daub, M., Realff, M.J., 2019. Techno-economic analysis of 1,4-butanediol production by a single-step bioconversion process. Biofuels Bioprod. Biorefining 13, 1261–1273. https://doi.org/10.1002/bbb.2016.
Shi, X., Tal, G., Hankins, N.P., Gitis, V., 2014. Fouling and cleaning of ultrafiltration membranes: A review. Journal of Water Process Engineering 1, 121–138. https://doi.org/10.1016/j.jwpe.2014.04.003.
Sikdar, S.K., 2003. Sustainable development and sustainability metrics. AIChE Journal 49, 1928–1932. https://doi.org/10.1002/aic.690490802.
Silva, R.G.C., Ferreira, T.F., Borges, É.R., 2020. Identification of potential technologies for 1,4-Butanediol production using prospecting methodology. Journal of Chemical Technology and Biotechnology 95, 3057–3070. https://doi.org/10.1002/jctb.6518.
Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., Prata, A., Steckenreiter, T., 2017. Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. Chemie-Ingenieur-Technik 89, 542–561. https://doi.org/10.1002/cite.201600175.
Spiegler, K.S., Kedem, O., 1966. Thermodynamics of hyperfiltration (reverse osmosis): criteria for efficient membranes. Desalination 1, 311–326. https://doi.org/10.1016/S0011-9164(00)80018-1.
Sun, W., Braatz, R.D., 2021. Smart process analytics for predictive modeling. Computers and Chemical Engineering 144, 107134. https://doi.org/10.1016/j.compchemeng.2020.107134.
Tidriri, K., Chatti, N., Verron, S., Tiplica, T., 2016. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. Annual Review in Control 42, 63–81. https://doi.org/10.1016/j.arcontrol.2016.09.008.
Vanhatalo, E., Kulahci, M., 2016. Impact of autocorrelation on principal components and their use in statistical process control. Quality and Reliability Engineering International 32, 1483–1500. https://doi.org/10.1002/qre.1858.
Varma, S., Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7, 1–8. https://doi.org/10.1186/1471-2105-7-91.
Velidandi, A., Kumar Gandam, P., Latha Chinta, M., Konakanchi, S., reddy Bhavanam, A., Raju Baadhe, R., Sharma, M., Gaffey, J., Nguyen, Q.D., Gupta, V.K., 2023. State-of-the-art and future directions of machine learning for biomass characterization and for sustainable biorefinery. J. Energy Chem. 81, 42–63. https://doi.org/10.1016/j.jechem.2023.02.020.
Vilker, V.L., Colton, C.K., Smith, K.A., Green, D.L., 1984. The osmotic pressure of concentrated protein and lipoprotein solutions and its significance to ultrafiltration. Journal of Membrane Science 20, 63–77. https://doi.org/10.1016/S0376-7388(00)80723-1.
Wankat, P.C., 2009. Separation Process Engineering, 4th ed. Prentice Hall, Boston (MA).
Whitaker, S., 1986. Flow in porous media I: A theoretical derivation of Darcy's law. Transport in Porous Media 1, 3–25. https://doi.org/10.1007/BF01036523.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laborary Systems 58, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.
Zhu, Q., 2021. Dynamic autoregressive partial least squares for supervised modeling. IFAC-Pap. 54, 234–239. https://doi.org/10.1016/j.ifacol.2021.08.364.