Control Engineering Practice 147 (2024) 105907

Contents lists available at ScienceDirect

Control Engineering Practice

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

Iterative real-time optimization of a membrane pilot plant

Afaq Ahmad [a,*], Radoslav Paulen [b], Richard Valo [b], Miroslav Fikar [b], Sebastian Engell [a,*]

[a] Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Strasse 70, Dortmund 44227, Germany

[b] Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Radlinského 9, 81237 Bratislava, Slovakia

ARTICLE INFO

Keywords:
Real-time optimization
Effective model adaptation
Plant-model mismatch
Modifier adaptation
Membrane processes
Model adequacy
Nanofiltration
Reverse osmosis
Surrogate model
Quadratic approximation

ABSTRACT

In this paper, we describe the implementation and experimental validation of iterative real-time optimization (RTO) at a continuously operated membrane separation pilot plant. The objective is to achieve an economically optimal operation using the transmembrane pressure and the feed flow rate as adjustable operational parameters (set-points for the lower-level plant controllers). In spite of the popularity of RTO based on rigorous nonlinear plant models in the research literature, its practical use so far remains limited to large scale plants in the petrochemical industries. The main issue behind the limited application of RTO schemes is the necessary effort to derive and maintain sufficiently accurate models of the plants. In case of a significant plant-model mismatch, the results are not reliable, the optimal performance is not reached and even systematic violations of constraints may occur. The applicability of RTO could be extended significantly if simple process models sufficed for its successful application. Then an adaptation of the model or of the optimization problem to the real behavior of the plant is necessary to achieve an optimal operation. For this example, we tackle this issue by applying an iterative scheme called modifier-adaptation with quadratic approximation in combination with model adaptation to ensure the convergence of the plant to the true process optimum despite the fact that the model has deficiencies. The experimental results show that, in spite of a considerable plant-model mismatch, the presented RTO scheme is capable of boosting the process economics.

1. Introduction

In the last decades, many traditional separation processes have been replaced by membrane separations. Membranes are e.g. industrially used for the concentration, separation and purification of solvents and of one or more solutes with a broad range of particle or molecule sizes (Cheryan, 1998; Scott & Hughes, 1996; Zeman & Zydney, 1996). A variety of applications have been found for membrane processes such as food (Yazdanshenas et al., 2005), chemicals (Ahmad et al., 2012), and biotechnology (Jaffrin & Charrier, 1994). Based on the pores size of the membrane, membrane separation processes are categorized into nanofiltration (NF), microfiltration (MF), reverse osmosis (RO) and ultrafiltration (UF).

NF represents membrane separations where molecules of molecular mass between 200–1000 $\mathrm{g\,mol^{-1}}$ are retained by a nanofilter membrane. The pressure range is 3–50 bar (Artug, 2007). This separation process is e.g. commonly used in the processing of vegetable oil, softening of water, wastewater treatment, beverage manufacturing and in the dairy (Chen et al., 2017), juice, and sugar industries (Conidi et al., 2017; Salehi, 2014). Several experimental studies were carried out to solve model based optimization problems to optimize the operation of continuous membrane processes (Paulen & Fikar, 2019; Sharma et al., 2019), assuming that an accurate plant model is available. If the available process model does not predict the behavior of the plant accurately, the optimal operating conditions found via numerical optimization may result in a sub-optimal economic performance or even in violating process constraints.

Real-time optimization (RTO) has the goal to drive an industrial process to optimal economic operating conditions, without jeopardizing the quality of the product and without violating process constraints. RTO, as a model-based upper-level optimization process, provides lower-level controllers with the set points to optimize the economics of the plant (Engell, 2007). The optimum operating points are usually calculated using stationary nonlinear first principles based models and nonlinear optimization solvers.

The performance of RTO relies on the accuracy of the model that is used for optimization. The effort needed for the development and maintenance of the model is a bottleneck in the implementation of RTO solutions. Moreover, any model, no matter how sophisticated, will

* Corresponding authors.
E-mail addresses: afaq.ahmad@tu-dortmund.de (A. Ahmad), radoslav.paulen@stuba.sk (R. Paulen), richard.valo@stuba.sk (R. Valo), miroslav.fikar@stuba.sk (M. Fikar), sebastian.engell@tu-dortmund.de (S. Engell).

https://doi.org/10.1016/j.conengprac.2024.105907
Received 14 December 2023; Received in revised form 11 February 2024; Accepted 6 March 2024
Available online 16 March 2024

be of limited accuracy and may sooner or later after its construction loose some of its predictive capabilities. To improve the accuracy of the model and to prevent model degradation, models can be re-identified using available measured data. However, when there are structural imperfections, a re-identified model will only be able to predict the plant behavior in a narrow range of operating conditions where the parameters were fitted. The approach where certain model parameters are adapted using plant measurements so that the model more accurately predicts the outputs of the plant and then new set-points are computed on the basis of the adapted model is called the two-step approach (Chen & Joseph, 1987; Darby et al., 2011; Jang et al., 1987). This is an intuitive method and is applied in industrial practice to large-scale chemical processes (Naysmith & Douglas, 1995). However, in the presence of structural plant-model mismatch, this approach may converge to sub-optimal solutions (Ahmad, Gao, & Engell, 2018; Forbes et al., 1994; Mandur & Budman, 2015). Moreover it is not guaranteed that sufficiently many parameters can be adapted based on the measured quantities of the plant.

Responding to this challenge, a modified two-step approach, called Integrated System Optimization and Parameter Estimation (ISOPE), was proposed by Roberts (1979). A parameter estimation problem is solved to make the model outputs match the measured plant outputs. Then, assuming that the plant gradients of the plant outputs to the inputs can be estimated experimentally, a gradient correction term is added to the cost of the optimization problem. Tatjewski (2002) pointed out that alternatively, one can use measurements to compute zeroth- and first-order correction terms to the cost function without model adaptation, and that this leads to convergence to the true plant optimum if the estimated gradients are correct. In Gao and Engell (2005), this approach was extended to problems with constraints, and a strategy to estimate the gradients from the measurements at the past operating points was proposed. Later this approach was termed Modifier Adaptation (MA) and its properties were investigated in detail (Marchetti et al., 2009). By adding terms to the cost and constraint functions of the optimization problem, the first-order necessary conditions of optimality of the true plant are satisfied. Therefore using the modifier adaptation approach, an iterative optimization scheme can converge to the optimum of the real plant despite the fact that the plant model is inaccurate. The iterative MA approach performs well provided that (i) the empirical gradients are accurately estimated from the usually noisy measurements and (ii) besides the KKT conditions for the models-based optimization problem, the optimality conditions of the second-order are fulfilled at the true plant optimum (Gao et al., 2015; Marchetti et al., 2009). The latter is called model adequacy (Ahmad et al., 2017; Gao et al., 2016; Marchetti et al., 2016). The critical element of modifier adaptation is the computation of the gradients of the cost function and of the constraints with respect to the manipulated inputs from measured data.

Various methods for obtaining the gradient information have been proposed. The plant gradients can be calculated by using finite differences (Roberts & Williams, 1981) which however is sensitive to noise. When using finite differences, usually one want to avoid additional probing moves of the inputs which perturb the operation of the plant, so it is preferable to compute the gradients based on the measurements at the past operating points. This however may lead to ill-conditioned problems in which case additional probing moves are necessary (Gao & Engell, 2005). Recently, Gao et al. (2016) proposed to estimate the plant gradients from local quadratic approximations of the cost and constraint functions to reduce the influence of the noise. This idea was borrowed from *derivative-free optimization* (Conn et al., 2009).

Addressing problem (ii), François and Bonvin (2013) proposed to only use convex models in order to satisfy the second-order optimality conditions. However, by this the rate of convergence can be reduced, since the accuracy of the model prediction is decreased by using local convex approximations of the objective function and of the constraints. Modifier Adaptation With Quadratic Approximation (MAWQA) monitors the model quality and switches between model-based iterative optimization with gradient correction and a purely data-based optimization using the fitted quadratic models (Gao et al., 2016) depending on the accuracy of the model predictions. MAWQA requires initial iterations, e.g. using the method from Gao and Engell (2005) until a sufficient number of data points has been collected to build the quadratic approximations of the cost function and of the constraint functions. The model adequacy of MA can be addressed by modifying the model parameters to ensure that the second order optimality operating conditions are met upon convergence (Ahmad et al., 2017). This led to the idea to combine MA and *Effective Model Adaptation* (EMA) so that the optimality conditions needed for convergence are fulfilled (Ahmad, Gao, & Engell, 2018). This approach has been further improved to enforce model adequacy by explicitly considering the adequacy criterion in the parameter estimation step (Ahmad, Gao, & Engell, 2019).

Ahmad, Gottu Mukkula, and Engell (2019) combined the advantages of EMA and MAWQA in response to both of the aforementioned difficulties in order to develop an efficient RTO scheme. The proposed approach is attractive for practical applications as it improves the robustness to measurement noise and guarantees convergence and speeds up the convergence rate to an optimum of the plant.

Although iterative RTO based upon Modifier Adaptation has been theoretically analyzed significantly improved and expanded in recent years, experimental validations of the method have only been reported in few studies so far (Bunin et al., 2012; Hernandez et al., 2018; Mukkula, Kern, et al., 2020; Navia et al., 2016). An experimental validation of RTO for a nanofiltration membrane separation process was presented in Mukkula et al. (2020b). In this study, the nominal model was a surrogate model that was fitted to experimental data instead of a first principles based model.

In this contribution, we discuss the implementation of the MAWQA with EMA iterative RTO scheme to a membrane separation process in a pilot-scale plant. The goal is to demonstrate the optimal operation of a continuous nanofiltration process in the presence of significant plant-model mismatch, where the optimization criterion is to reduce the operating costs and to achieve a specified recovery ratio.

The remainder of this paper is structured in the following manner. The experimental setup and the process model are explained and the optimization problem is formulated in Section 2. The iterative optimization problem of RTO in the presence of plant-model mismatch is summarized in Section 3. Section 4.1 discusses simulations based on a surrogate model that was fitted to the experimental behavior of the plant as the "real" plant in order to discuss the advantage of MAWQA with EMA over MAWQA without model adaptation. The experimental results which demonstrate the good performance of the proposed scheme are presented in Section 4.2. A summary and conclusions are provided in Section 5. In the appendix, the MAWQA algorithm extended by EMA (enforcing model adequacy) is presented in detail.

2. Experimental setup, description of the process model and formulation of the optimization problem

This section presents the experimental setup of the plant and describes the process model and the optimization problem.

2.1. Plant description

Fig. 1 shows the flow diagram of the membrane pilot plant (see Table 1 for a explanation of the abbreviations), and a photograph of the membrane pilot plant can be seen in Fig. 2. Except for its size, the pilot plant is similar to standard industrial hardware and is influenced by the same kinds of disturbances. The membrane module is operated at a constant temperature. It is composed of four sections, the feed side,

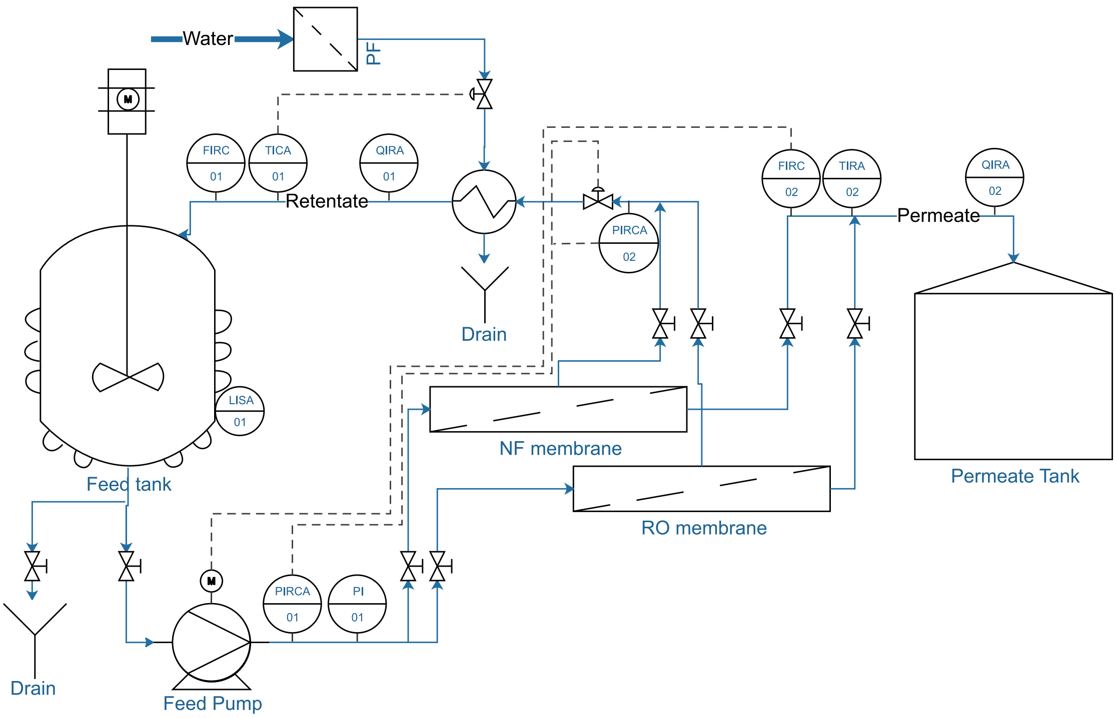


Fig. 1. Flow diagram of membrane pilot plant.

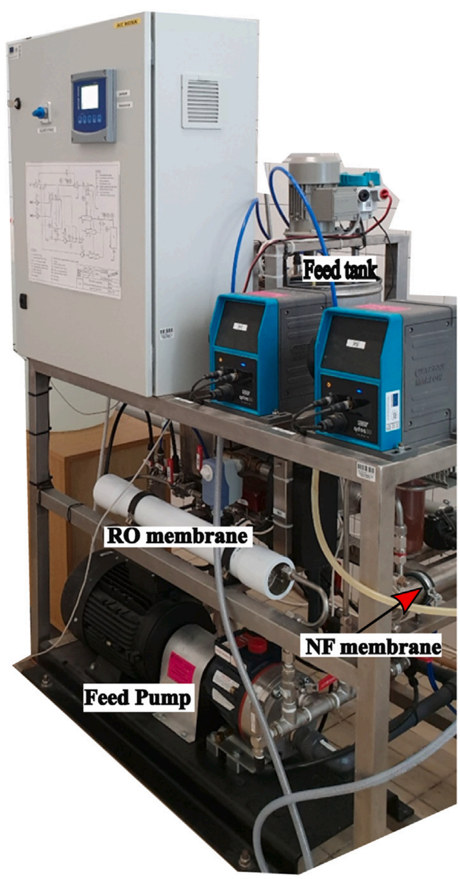


Fig. 2. Photograph of the membrane pilot plant.

the internal membrane area, the retentate side and the permeate side. The solution is added to the feed tank for separation.

Table 1
Abbreviations used in Fig. 1.

Abbreviation	Meaning
NF	Nanofiltration
RO	Reverse Osmosis
LISA	Level Indication Switch Alarm
PIRC	Pressure Indicator, Recorder, Control, Alarm
PI	Pressure Indicator
QIRA	Quantity Indication Recording Alarm (conductivity)
TICA	Temperature Indicator, Control, Alarm
TIRA	Temperature Indicator, Recorder, Alarm
FIRC	Flowrate Indicator, Recorder, Control
PF	Pre-filter

The centrifugal pump provides the feed to the membrane in cross-flow mode at a certain trans-membrane pressure. The filtered flow, i.e. the permeate stream, passes through the membrane to the permeate tank. The rejected membrane stream (retentate) returns to the feed tank as shown in Fig. 1. Initially, the feed tank is filled with the feed solution composed of the solvent (water), the macro-solute (lactose monohydrate) and the micro-solute (sodium chloride). The driving force of the separation is created by the centrifugal feed pump and by the manipulation of the valve at the retentate side of the membrane module. The flow rates of retentate and permeate are measured in L/h using FIRC01 and FIRC02. The concentrations of sodium chloride in the retentate and permeate are inferred from conductivity measurements (sensors QIRA01 and QIRA02) using the experimentally calibrated linear model

$$c\,[\mathrm{g/L}] = 0.0007 \times \mathrm{QIRA}\,[\mu\mathrm{S/cm}] - 0.6949. \tag{1}$$

The feed flowrate (q_f) is controlled by the rotational speed (v_{pump}) of the centrifugal pump. The valve at the retentate (V_r) side of the membrane module controls the trans-membrane pressure (ΔP) defined as:

$$\Delta P(P_f, P_r, P_p) = \frac{P_f + P_r}{2} - P_p, \tag{2}$$

where P_f, P_r and P_p represent the measured feed pressure, retentate pressure and permeate pressure.

Table 2
Parameters and inputs of the membrane process.

Parameters	Description	Value
A	Membrane area	$0.465\,\mathrm{m}^2$
d_h	Hydraulic diameter	$0.092\,\mathrm{m}$
a_1	Virial coefficient	$0.208\,\mathrm{kPa}\ \mathrm{m}^3/\mathrm{kg}$
a_2	Virial coefficient	$0.002\,\mathrm{kPa}\ (\mathrm{m}^3/\mathrm{kg})^2$
a_3	Virial coefficient	$0.00001\,\mathrm{kPa}\ (\mathrm{m}^3/\mathrm{kg})^3$
R_m	Membrane resistance to pure water (constant)	$85\,\mathrm{kPa\,h\,m^2/L}$
D_i	Diffusion coefficient	$3 \times 10^{-11}\ \mathrm{m}^2/\mathrm{s}$
ν	Kinematic viscosity of water	$1.004 \times 10^{-6}\ \mathrm{m}^2/\mathrm{s}$

The solution is kept at a constant temperature of 20 °C by a heat exchanger with pre-filtered coolant and a temperature controller (Sharma et al., 2018, 2019). The process is continuously operated in a closed-loop mode. The permeate was circulated back to the feed tank for this experiment.

2.2. Process model

A fundamental model of membrane processes was presented in Guadix et al. (2004). The behavior of the plant is described by the following equations:

$$q_f = q_r + q_p, \tag{3}$$

where the permeate flow rate is determined by the osmotic pressure model (Cheryan, 1998):

$$q_p(c_f, q_f, \Delta P) = A\frac{\Delta P - \Pi(c_w)}{R_m}. \tag{4}$$

The osmotic pressure depends on the lactose concentration at the membrane wall which can be expressed in terms of the virial coefficients as follows (Cheryan, 1998):

$$\Pi(c_w) = a_1 c_w + a_2 c_w^2 + a_3 c_w^3. \tag{5}$$

Therefore, the permeate flow rate is:

$$q_p(c_f, q_f, \Delta P) = A\frac{\Delta P - (a_1 c_w + a_2 c_w^2 + a_3 c_w^3)}{R_m}. \tag{6}$$

Following this, the concentration at the wall can be determined from Noronha et al. (2002):

$$q_p(c_f, q_f, \Delta P) = k(q_f) A \ln\frac{c_w}{c_f}. \tag{7}$$

For the process conditions at this plant, the mass transfer coefficient is given by Noronha et al. (2002), Schock and Miquel (1987)

$$k(q_f) = 0.065 D_i^{0.75} d_h^{-0.125} \nu^{-0.625}\left(\frac{q_f}{A}\right)^{0.875}, \tag{8}$$

where D_i is a diffusion coefficient in m^2/s, d_h is the hydraulic diameter in m (twice the channel thickness), ν is the kinematic viscosity of water in m^2/s, and A is the membrane area in m^2.

Eqs. (6) and (7) are used to obtain the permeate flowrate and the lactose concentration at the membrane wall. For the modeling of the separation efficiency, we assume $c_p = c_f = c_r$. The parameters which are used in the process model are given in Table 2. These model parameters were determined using historical experimental data. However, the preliminary experiments showed significant variations in the prediction accuracy of the model over longer periods of time (days). This may be due to fouling, but other experimental factors could also be the cause of the plant-model mismatch.

Table 3
Manipulated variables of the membrane process.

Inputs	Description	Value
ΔP	Transmembrane pressure	1000–2500 kPa
q_f	Feed flowrate	300–500 L/h

2.3. Materials

Sodium chloride and lactose monohydrate manufactured by Centralchem (Slovakia) are used as solutes. Reverse osmosis water is used as a solvent to prepare the solutions. The concentrations of sodium chloride and lactose monohydrate in the feed solution are $1.0\,\mathrm{g/L}$ and $40\,\mathrm{g/L}$, respectively. A NFW-1812F nanofilter membrane is used, manufactured by Synder Filtration, USA. It has a cut-off range of 300–500 Daltons and a membrane area of $A = 0.465\,\mathrm{m}^2$.

2.4. Optimization problem for the membrane process

The optimization problem is formulated to minimize the operational costs while maintaining the recovery ratio (RR) and permeate flowrate above prescribed thresholds. The manipulated variables are the feed flowrate q_f and the transmembrane pressure ΔP.

The optimization problem for the membrane process is given below:

$$\begin{aligned} \min_{q_f, \Delta P} \quad & J(q_f, \Delta P) \quad := p_e q_f \Delta P \\ \text{s.t.} \quad & \mathrm{RR} \geq \phi_1, \\ & q_p \geq \phi_2, \end{aligned} \tag{9}$$

where $\mathrm{RR} := \frac{q_p c_p}{q_f c_f}$ is the recovery ratio. $\phi_1 = 0.01$ and $\phi_2 = 7$ are the minimum thresholds of the recovery ratio RR and permeate flowrate q_p, respectively. $p_e = 0.07$ €/kWh is the price of electricity.

The flow rates of the permeate q_p and of the retentate q_r are obtained at 15 s intervals during operation. The model parameters a_1 and a_2 are estimated from the flow rate measurements and the remaining model parameters are the same as those of the nominal model. The ranges of the manipulated variables are given in Table 3.

3. The MAWQA with EMA RTO scheme

3.1. Problem formulation

The steady-state optimization of the operation of a plant can be represented as:

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_p(\mathbf{u}) := \tilde{J}_p(\mathbf{u}, \mathbf{y_p}(\mathbf{u})) \\ \text{s.t.} \quad & \mathbf{C}_p(\mathbf{u}) := \tilde{\mathbf{C}}_p(\mathbf{u}, \mathbf{y_p}(\mathbf{u})) \leq \mathbf{0}, \\ & \mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max}, \end{aligned} \tag{10}$$

where $\mathbf{y_p} \in \mathbb{R}^{n_y}$ and $\mathbf{u} \in \mathbb{R}^{n_u}$ are the vectors of the plant measurements and of the control inputs. $J_p : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}$ represents the cost function to be minimized and $\mathbf{C}_p : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_C}$ is the vector of constraints. The upper and lower bounds of the input variables are $\mathbf{u}_{max}$ and $\mathbf{u}_{min}$. The solution to the optimization problem (10) is represented by $\mathbf{u}_p^*$.

The exact input–output mapping, $\mathbf{u} \to \mathbf{y_p} : \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$, is however not known and only an approximate plant model is available. The optimization problem based on the model can be described as follows:

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_m(\mathbf{u}, \theta) := \tilde{J}(\mathbf{u}, \mathbf{y}_m(\mathbf{u}, \theta)) \\ \text{s.t.} \quad & \mathbf{C}_m(\mathbf{u}, \theta) := \tilde{\mathbf{C}}(\mathbf{u}, \mathbf{y}_m(\mathbf{u}, \theta)) \leq \mathbf{0}, \\ & \mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max}, \end{aligned} \tag{11}$$

where $\mathbf{y}_m(\mathbf{u}, \theta)$ represents the output vector that is predicted by the approximate model. It is a function of the vector of model parameters

$\theta \in \mathbb{R}^{n_\theta}$ and of the input variables $\mathbf{u}$. J_m and $\mathbf{C}_m$ are the cost function and the vector of constraint functions of the model.

In the presence of a plant-model mismatch, the solution of the optimization problem (11) is generally different from the true plant optimum $\mathbf{u}_p^*$. The aim of iterative optimization is to find the optimum value of the actual plant $\mathbf{u}_p^*$ by iteratively adjusting and solving the optimization problem (11) on the basis of measured information.

In principle, the model parameters can be adapted to cope with the plant-model mismatch if the plant can be represented accurately by the model, i.e. there it is no *structural* mismatch, and the operating conditions lead to sufficient excitation for the estimation of the uncertain model parameters. This approach is known as the two-stage approach, where the model parameters are determined by fitting the plant measurements at the current operating point using least-square value fitting as given below:

$$\begin{aligned} \theta^{(k)} :=&\quad \arg\min_{\theta} \|\mathbf{y}_p(\mathbf{u}^{(k)}) - \mathbf{y}_m(\mathbf{u}^{(k)}, \theta)\|^2, \\ \text{s.t.}&\quad \theta_{min} \le \theta \le \theta_{max}, \end{aligned} \tag{12}$$

where $\mathbf{y}_p(\mathbf{u}^{(k)})$ is the vector of plant measurements at the current operating point $\mathbf{u}^{(k)}$. The parameters have lower and upper bounds, θ_{min} and θ_{max}. The updated model is then used in the optimization problem (11) to generate a new set-point. This approach may converge to a sub-optimal point because the plant optimality conditions are not satisfied in the presence of structural plant-model mismatch.

3.2. Modifier adaptation

In contrast, the parameters of the model are not modified in the Modifier Adaptation (MA) scheme. In MA, correction terms of the bias and of the gradients are added to the cost function and to the process constraints of the nominal optimization problem (11). The modified optimization problem is defined below (Gao & Engell, 2005).

$$\begin{aligned} \min_{\mathbf{u}}&\quad J_{ad}^{(k)}(\mathbf{u}, \theta) := J_m(\mathbf{u}, \theta) + \varepsilon_J^{(k)} + \lambda_J^{(k)}\left(\mathbf{u} - \mathbf{u}^{(k)}\right) \\ \text{s.t.}&\quad \mathbf{C}_{ad}^{(k)}(\mathbf{u}, \theta) := \mathbf{C}_m(\mathbf{u}, \theta) + \varepsilon_C^{(k)} + \Lambda_C^{(k)}\left(\mathbf{u} - \mathbf{u}^{(k)}\right) \le \mathbf{0}, \\ &\quad \mathbf{u}_{min} \le \mathbf{u} \le \mathbf{u}_{max}, \end{aligned} \tag{13}$$

where $\varepsilon_J^{(k)}$ represents a zero order (bias) correction of the cost function that is a scalar quantity and the vector of the bias modifiers for the constraint functions is represented by $\varepsilon_C^{(k)} \in \mathbb{R}^{n_C}$. $\lambda_J^{(k)} \in \mathbb{R}^{n_u}$ and $\Lambda_C^{(k)} \in \mathbb{R}^{n_u \times n_C}$ represent the first-order modifications in the cost and constraint functions. The bias correction of the cost function can be omitted as it does not influence the computed optimum.

The zero- and first-order modifiers in the kth RTO iteration are defined as:

$$\varepsilon_J^{(k)} = J_p(\mathbf{u}^{(k)}) - J_m(\mathbf{u}^{(k)}, \theta), \tag{14}$$

$$\lambda_J^{(k)} = \left(\nabla J_p(\mathbf{u}^{(k)}) - \nabla J_m(\mathbf{u}^{(k)}, \theta)\right)^T, \tag{15}$$

$$\varepsilon_C^{(k)} = \mathbf{C}_p(\mathbf{u}^{(k)}) - \mathbf{C}_m(\mathbf{u}^{(k)}, \theta), \tag{16}$$

$$\Lambda_C^{(k)} = \left(\nabla \mathbf{C}_p(\mathbf{u}^{(k)}) - \nabla \mathbf{C}_m(\mathbf{u}^{(k)}, \theta)\right)^T, \tag{17}$$

where $\nabla \mathbf{C}_p(\mathbf{u}^{(k)})$ and $\nabla \mathbf{C}_m(\mathbf{u}^{(k)}, \theta)$ represent of the plant and the model gradients of the constraint functions. The gradients are assumed to be available at the current operating condition $\mathbf{u}^{(k)}$. In order to avoid excessive corrective actions, the operating conditions are usually updated after filtering:

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{*(k+1)} + \mathbf{K}(\mathbf{u}^{(k)} - \mathbf{u}^{*(k+1)}). \tag{18}$$

The solution to the problem of modified optimization (13) is represented by $\mathbf{u}^{*(k+1)}$, $\mathbf{K}$ is the filter matrix.

MA has the potential to identify the true optimum of the plant upon convergence. Let assume that $\mathbf{u}^{(\infty)} := \lim_{k\to\infty} \mathbf{u}^{(k)}$ is the asymptotic solution of the problem of modified model-based optimization (13). Then $\mathbf{u}^{(\infty)}$ also fulfills the necessary optimality conditions of the plant optimization problem (10). MA converges to the optimal input $\mathbf{u}^{(\infty)}$ if the gradients of the plant and of the constraints can be estimated correctly and if the solution of the modified optimization problem (13) meets the second-order optimality conditions upon convergence (see Appendix A).

3.3. MAWQA with EMA

MAWQA (Gao et al., 2016) is based on the central idea of DFO (Conn et al., 2009) and uses local quadratic approximations (QA) around the current operating conditions to capture the curvature information of the function which is optimized in order to achieve a better local convergence rate. MAWQA exploits the inherent smoothness of the approximating function to improve the robustness of the estimation of the gradients to noise (details are given in Appendix B). The computation of the quadratic approximation requires at least $(n_u+1)(n_u+2)/2$ data points (i.e. responses of the plant at different set-points). Before this number of data points are available, MA based on finite differences or probing moves must be applied. In practice, it is important that an iterative scheme approaches the optimal operation fast and without oscillations. Therefore, we combine MAWQA with model adaptation such that the advantages of both approaches are combined.

In effective model adaptation (EMA) as proposed in Ahmad et al. (2017), Ahmad, Gao, and Engell (2018, 2019), Ahmad, Singhal, et al. (2018) the model parameters are adapted such that the adapted plant model meets the model adequacy conditions which helps to prevent oscillations of the iterative optimization. A vector of updated model parameters $\theta^{(k+1)}$ at the set-point $\mathbf{u}^{(k+1)}$ is accepted in EMA if and only if the following two conditions are fulfilled:

$$\nabla^2 J_m(\mathbf{u}^{(k+1)}, \theta^{(k+1)}) \text{ is positive definite,} \tag{19}$$

$$\left|J_{ad}^{(k)}(\mathbf{u}^{(k+1)}, \theta^{(k+1)}) - J_p(\mathbf{u}^{(k+1)})\right| < \left|J_{ad}^{(k)}(\mathbf{u}^{(k+1)}, \theta^{(k)}) - J_p(\mathbf{u}^{(k+1)})\right|. \tag{20}$$

The first condition of EMA ensures that for the model with adapted parameters the Hessian of the Lagrangian of problem (13) is positive definite at the kth RTO iteration. The second condition results from the application of the Taylor series expansion of $J_{ad}^{(k)}(\mathbf{u}, \theta) - J_p(\mathbf{u})$ at $\mathbf{u}^{(k)}$ while omitting the terms of higher than 2$^{\text{nd}}$ order:

$$\begin{aligned} &J_{ad}^{(k)}(\mathbf{u}, \theta) - J_p(\mathbf{u}) \approx \\ &\qquad \frac{1}{2}\left(\mathbf{u} - \mathbf{u}^{(k)}\right)^T \left(\nabla^2 J_m^{(k)} - \nabla^2 J_p^{(k)}\right)\left(\mathbf{u} - \mathbf{u}^{(k)}\right). \end{aligned} \tag{21}$$

The prediction error of the adapted objective function at $\mathbf{u}^{(k+1)}$, i.e. $J_{ad}^{(k)}(\mathbf{u}^{(k+1)}, \theta) - J_p(\mathbf{u}^{(k+1)})$, can be used to infer the difference of the Hessian matrices at $\mathbf{u}^{(k)}$. Hence, reducing the difference between $\nabla^2 J_m(\mathbf{u}, \theta)$ and $\nabla^2 J_p(\mathbf{u})$ ensures that the adapted model parameters accelerate the rate of convergence to the process optimum.

Solving a least-squares estimation problem to improve the accuracy of the model prediction does not ensure that the adapted parameters will meet the model adequacy criterion. In Ahmad, Gao, and Engell (2019), slack variables are used in a soft constraint to enforce model adequacy by solving

$$\begin{aligned} \theta^{(k)} :=&\quad \arg\min_{\theta, \omega} \|\mathbf{y}_p(\mathbf{u}^{(k)}) - \mathbf{y}_m(\mathbf{u}^{(k)}, \theta)\|^2 + \omega^T\ \Gamma\ \omega, \\ \text{s.t.}&\quad -\mathbf{eig}\left(\nabla^2 \mathcal{L}_{ad}(\mathbf{u}^{(k)}, \theta, \mu^{(k)})\right) - \omega < \mathbf{0}, \\ &\quad \theta_{min} \le \theta \le \theta_{max}, \\ &\quad \omega \ge \mathbf{0}, \end{aligned} \tag{22}$$

where Γ is the $n_u \times n_u$ diagonal weighting matrix with positive elements and $\mathbf{eig}$ represents the n_u- dimensional vector of the eigenvalues of the Hessian of the Lagrangian function. $\omega \in \mathbb{R}^{n_u}$ is the vector of slack variables. Each element of ω describes the amount of constraint violation of the corresponding eigenvalue. If all elements of ω are equal to 0, all eigenvalues of the Hessian of the Lagrangian function are non-negative for a given value of the parameters θ.

Fig. 3 shows the flowchart of the MAWQA with EMA algorithm. The iterative RTO scheme starts by generating the probing moves and

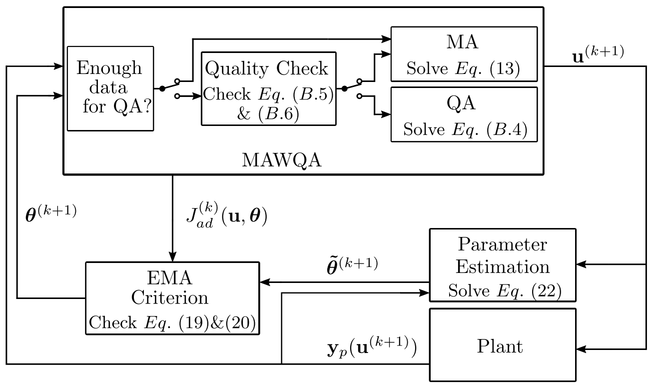


Fig. 3. MAWQA with EMA algorithm.

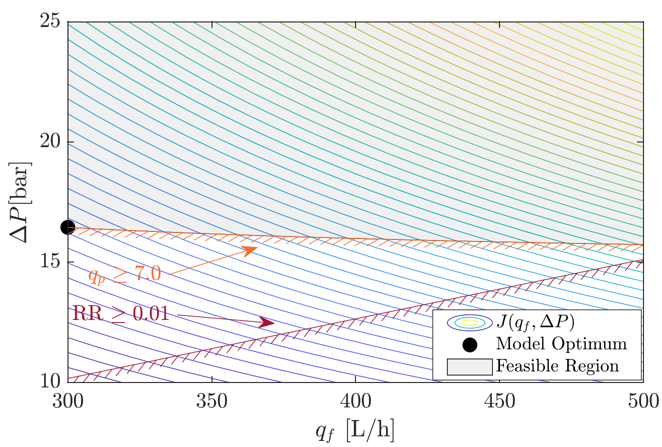


Fig. 4. Contour plot of the cost function with the constraint functions of the model of the membrane plant. The lines represent the constraints RR $\geq \phi_1$ and $q_p \geq \phi_2$. The shaded region at the top is feasible. The black dot represents the model optimum.

after the gradients have been computed from the results of the probing moves solves the modified optimization problem (13) until a sufficient number of data points (input–output pairs) is gathered for building quadratic approximations. New model parameters are estimated at each RTO iteration. The flowchart denotes the solution to the parameter estimation problem (22) as $\tilde{\theta}^{(k+1)}$ to illustrate that the EMA criteria have to be fulfilled for the parameter vector update ($\tilde{\theta}^{(k+1)} \rightarrow \theta^{(k+1)}$) to be accepted. The quadratic approximations are built when a sufficient number of data points are available. Depending on the quality assessment, the subsequent inputs are calculated by solving either the modified optimization problem (13) or the optimization problem based solely on the quadratic approximation (B.4).

4. Results

4.1. Simulations based on a surrogate model of the membrane plant

In this section, simulations based upon a surrogate model of the real plant are used to demonstrate the performance of the proposed MAWQA with EMA RTO scheme and compare it with two-step approach, MAWQA without parameter adaptation and MAWQA with model adaptation (MA). In each iteration of MAWQA with MA, the model parameters are estimated via problem (12). It is important to highlight that neither the checking of the EMA criterion nor the enforcing of model adequacy are considered in MAWQA with MA. The manipulated variables are the feed flowrate q_f and the transmembrane pressure ΔP. The permeate flow rate q_p and the retentate flow rate

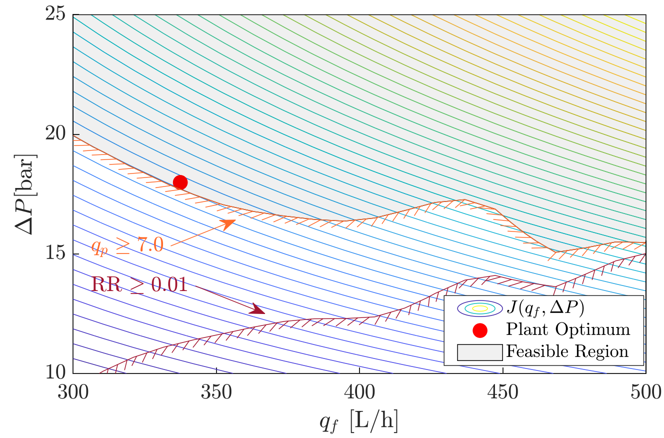


Fig. 5. Contour plot of the cost function and the constraint functions of the surrogate model of the real membrane plant. The lines are the constraints RR $\geq \phi_1$ and $q_p \geq \phi_2$. The shaded region at the top is feasible. The red dot represents the plant optimum for the surrogate model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

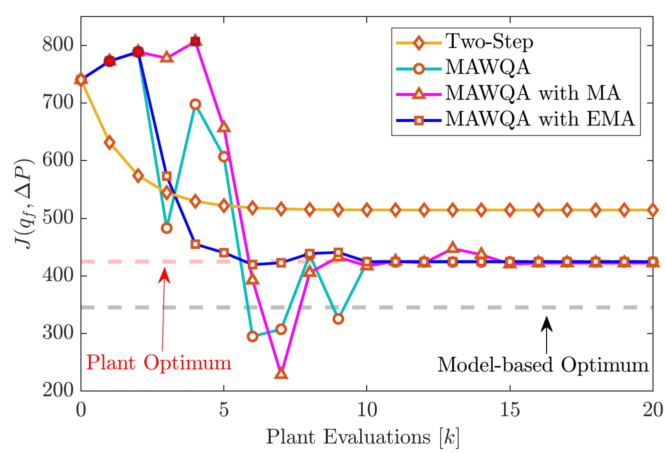


Fig. 6. Comparison of the iterative optimization schemes with respect to the evolution of the plant cost using the two-step approach, MAWQA, MAWQA with MA, and the proposed approach i.e. MAWQA with EMA. The dashed red line marks the real plant optimum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

q_r are measured in 15-second intervals during operation. The model parameters a_1 and a_2 are estimated based on these flow rate measurements, while the other model parameters remain at their nominal values (see Table 2).

Fig. 4[1] shows the contour plot of the cost function of the first principles based model (see Section 2.2) and the constraints. As the goal is to minimize the operational cost while satisfying the constraints, the solution of the model-based optimization is [300 L/h, 16.4489 bar]. It can be observed that the model optimum is defined by the constraint on the minimum permeate flow rate, $q_p \geq 7.0$.

The surrogate plant model that represents the real plant in the simulations was developed using plant measurements. The model structure is an artificial neural network with one hidden layer consisting of eight neurons. There are 1108 measurement points in the available data set. From the available data sets, 70% were used for training, while 15% each are used for validation and testing.

[1] For the graphical representation of the results, we use a consistent color coding throughout the paper. The meaning of the graphical elements used is introduced at the first occurrence via a legend entry.

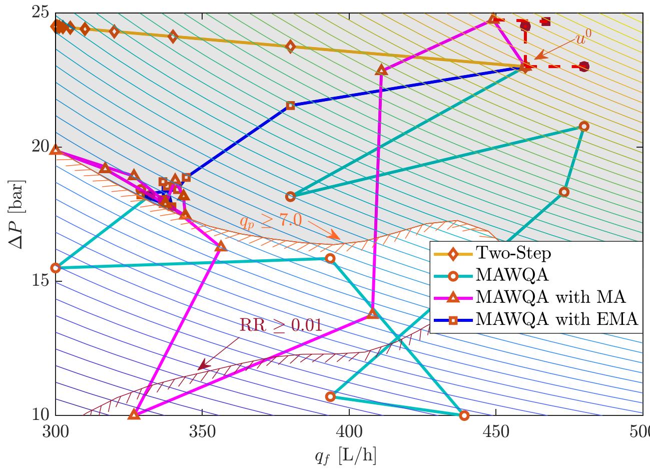


Fig. 7. Simulation result of the iterative optimization using the two-step approach, MAWQA, MAWQA with MA, and MAWQA with EMA: Sequence of set-points in the input space with additional perturbations (■). $J(q_f, \Delta P)$ is depicted by iso cost lines. The shaded area in the figure marks the feasible region.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5 presents the plant optimum and the constraints of the plant as predicted by the surrogate model. It can be seen that the optimal operating point predicted by the first principles based model is quite different from the actual plant optimum. In fact, the optimum identified by the optimization based on first principles based plant model computes an infeasible operating point as the constraint on the permeate flowrate would be violated.

The comparison of the two-step approach, MAWQA, MAWQA with MA, and the proposed approach, MAWQA with EMA, is shown in Figs. 6–8.evaluation. Fig. 6 shows the operational cost of the membrane process for each plant Fig. 7 shows the input moves on cost contours. Initial perturbations that are needed to estimate the plant gradients are illustrated by the red dotted lines. Figs. 8(a) and 8(b) show the trajectories of the constrained variables RR and q_p.

The screening method for the points in $\mathbb{U}^{(k)}$ in order to identify $\mathcal{U}^{(k)}$ is used same as in Gao et al. (2016). The screening parameter $\Delta\mathbf{u}$ and the initial step-size h are set to 0.05 and 0.1, respectively. The scaling parameter γ is set to 3. After the initial probing moves, the plant gradients are computed as in Gao and Engell (2005), i.e. by a finite-difference approximation with additional perturbations around $\mathbf{u}^{(k)}$ if the data matrix becomes close to singular in order to reduce the number of plant evaluations. Afterwards, when at least $(n_u+1)(n_u+2)/2 \equiv 6$ data points are available, quadratic approximation of the cost and constraint functions is used. The following observations can be made:

(1) The convergence of the two-step approach leads to a suboptimal point that is significantly different from the true plant optimum. The reason for this is the presence of structural plant-model mismatch.

(2) Without model adaptation, MAWQA moves slowly towards the optimal operational cost of the membrane process as an inadequate model is used until a minimal number of $(n_u+1)(n_u+2)/2$ data points have been collected. Therefore, the 4th and 5th operating points diverge from the plant optimum. After enough data points ($(n_u+1)(n_u+2)/2 = 6$) have been collected, quadratic approximations of the cost and of the constraints functions can be constructed and the rate of convergence is improved. It can be observed that, at the 6th and 7th steady states of the plant, the operational cost decreases significantly. However, the recovery ratio RR threshold is violated at these operating points. Also the threshold for the permeate flow rate q_p is violated at the 6th, 7th, 8th, and 9th plant evaluations. At the optimum, both constraints are satisfied while the operational cost is minimal.

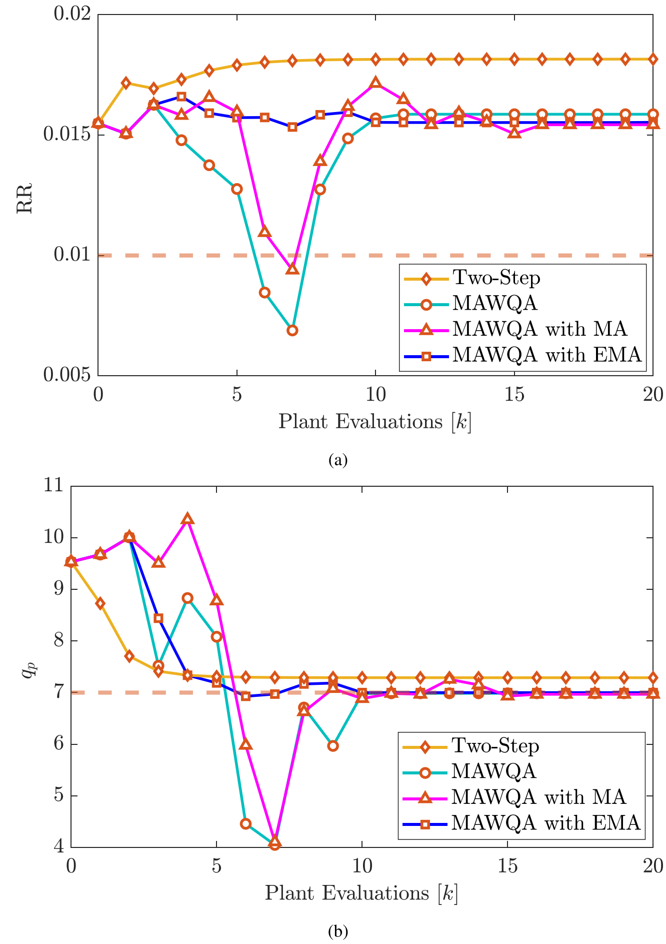


Fig. 8. Simulation results of the two-step approach, MAWQA, MAWQA with MA, and MAWQA with EMA: (a) Evolution of the recovery ratio (RR). (b) Evolution of the permeate flowrate (q_p). The orange dashed lines (– –) indicate the minimum values of the constraints. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

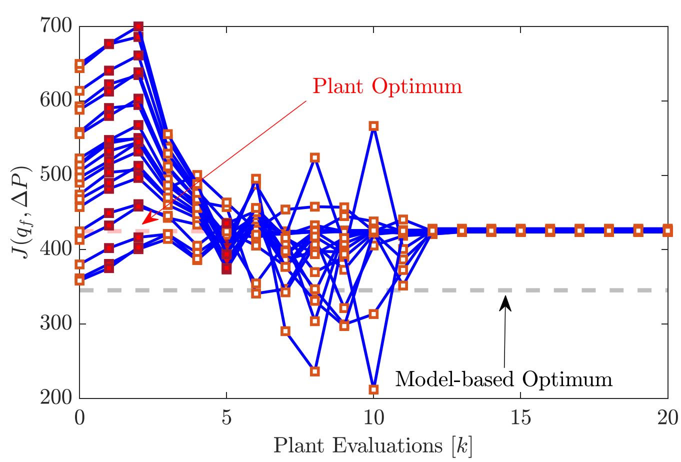


Fig. 9. Simulation result of the iterative optimization using MAWQA with EMA for different initial operating conditions: The black dashed line represents the model optimum. The red dashed line represents the plant optimum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

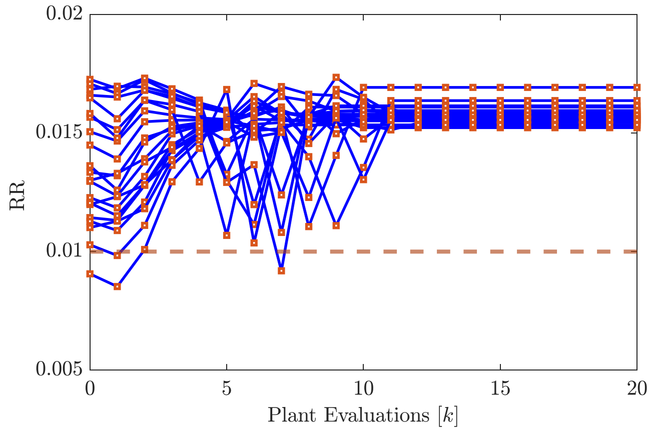


Fig. 10. Simulation results for different initial operating conditions: Evolution of the recovery ratio (RR). The orange dashed line represents the minimum threshold of the RR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

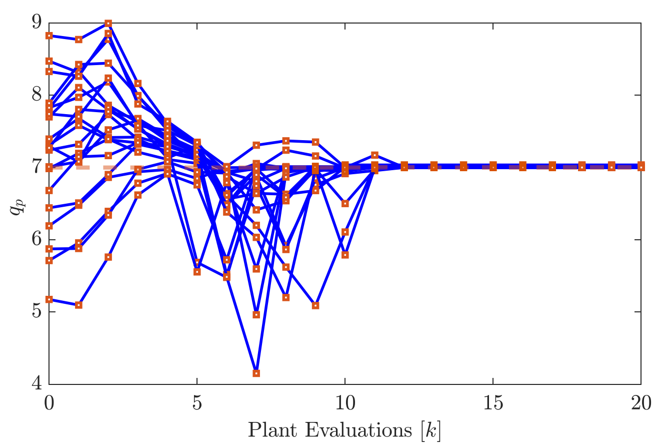


Fig. 11. Simulation results using MAWQA with EMA for different initial operating conditions: Evolution of the permeate flowrate (q_p). The minimum threshold of the permeate flowrate q_p is equal to the value to which the optimization converges.

(3) With model adaptation via problem (12), MAWQA with MA initially moves far from the optimum. This indicates that it is not ensured that model adaptation always provides an adequate model. Model adaptation in MAWQA without EMA may therefore not improve the performance of the iterative optimization. It can be observed that the 3rd and the 4th operating points diverge from the plant optimum. At the 5th plant evaluation, MAWQA with MA determines model parameters that reduce the operating cost. Afterwards it starts oscillating again, until a sufficient number of data points ($(n_u + 1)(n_u + 2)/2 = 6$) are collected. At this point, quadratic approximations of the cost and constraints functions are constructed, leading to an enhanced rate of convergence to the process optimum.

(4) The proposed scheme, MAWQA with EMA, improves the rate of convergence to the process optimum. As the model parameters are adapted using (22) in the framework of EMA, the adapted model parameters are always adequate. The MAWQA with EMA scheme converges to the optimum after the 10th iteration. It achieves 96.48% of the possible improvement of the plant optimum in six iterations, whereas, for the MAWQA scheme, this value is only 69%.

Fig. 9 shows the performance of MAWQA with EMA for different initial setpoints of the plant. The operational cost of the simulated process is plotted for each plant evaluation all trajectories converge smoothly to the plant optimum after the $14th$ plant evaluation.

Figs. 10 and 11 show the trajectories of the constrained variables RR and q_p that are obtained during iterative optimization using MAWQA with EMA. After the 5th iteration, all simulations meet the constraint for the recovery ratio, i.e., RR ≥ 0.01. After the 14th iteration both constraints are satisfied while the operational cost is at the minimum.

4.2. Experimental results

In this section, the performance of the proposed RTO scheme is discussed for the real plant that was described in Section 2. The MAWQA with EMA algorithm is implemented in MATLAB and the communication with the plant automation system is realized via a dedicated web-server application.

The performance of MAWQA with EMA for the pilot plant can be seen in Fig. 12, where the operational cost of the membrane process is plotted for each plant evaluation. Fig. 13 shows the input moves on the contour plot of the cost function. The red dotted lines illustrate the additional plant perturbations that are needed to estimate the plant gradients at the beginning of the optimization process. The objective is to minimize the operational costs while maintaining a recovery ratio RR ≥ 0.01 and a flow rate $q_p \geq 7$. The iterative RTO scheme started from the feasible point at $u^0 = [460\,\mathrm{L/h}, 23\,\mathrm{bar}]$. From Fig. 12, it can be observed that the initial operating point of the experiment leads to a high operating cost that is far from the cost at the model-based optimum and at the plant optimum.

The screening method for the points in $\mathbb{U}^{(k)}$ in order to identify $\mathcal{U}^{(k)}$ is the same as in Gao et al. (2016). The screening parameter $\boldsymbol{\Delta u}$ and the initial step-size h are set to 0.1 and 0.1, respectively. The scaling parameter γ is set to 3.

In the initial iterations, the plant gradients are computed using finite-difference approximation, as described in Section 4.1, until enough data points are available for the quadratic approximation. From Figs. 12 and 13, it can be observed that MAWQA with EMA converges to the feasible region after the 8th iteration. From Fig. 13, it can be observed that some operating points that are computed via the modified model-based optimization violate the permeate flow rate q_p constraint because the process model does not accurately predict the behavior of the plant.

The experimental results that are obtained during the iterative RTO of the membrane process are presented over time in Fig. 14. Fig. 14(a) shows the evolution of the operational cost of the membrane process. The trajectories of the constrained variables RR and q_p are shown in Figs. 14(b) and 14(c). At the optimum, both constraints are satisfied while the operational cost is minimal. MA ensures the feasibility of the operating point upon convergence, however, as it has also been observed in other papers as well, the RTO iteration before convergence may lead to violations of the plant constraints (Marchetti et al., 2016). The Sufficient Conditions for Feasibility and Optimality (SCFO), as proposed by Bunin et al. (2013a, 2013b), can be combined with any RTO method to enforce the feasibility of the plant throughout the optimization process. However, SCFO may lead to slow convergence (see the examples in Bunin et al. (2013a)). This limitation arises due to the necessity to bound the uncertain plant constraints using Lipschitz constants.

5. Conclusions

In this contribution, an iterative real-time optimization approach based on the adaptation of some model parameters and modifier adaptation with quadratic approximation was applied to a continuous nanofiltration membrane process. Despite a significant plant-model mismatch, the iterative real-time optimization scheme is able to converge to the true plant optimum fast and without oscillations.

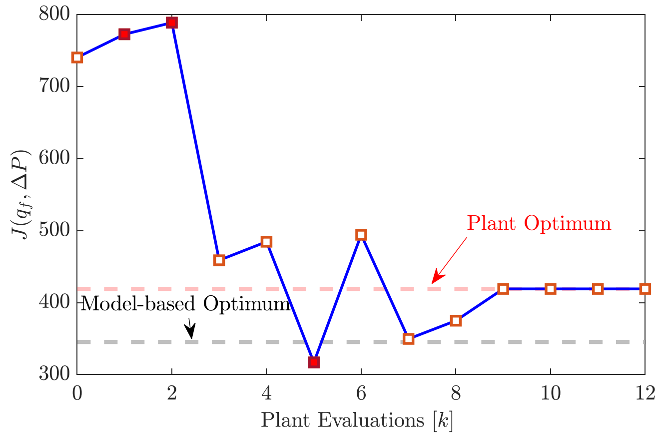


Fig. 12. Performance of the iterative optimization scheme using MAWQA with EMA at the pilot plant. The black dashed line represents the model optimum. The red dashed line represents the plant optimum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

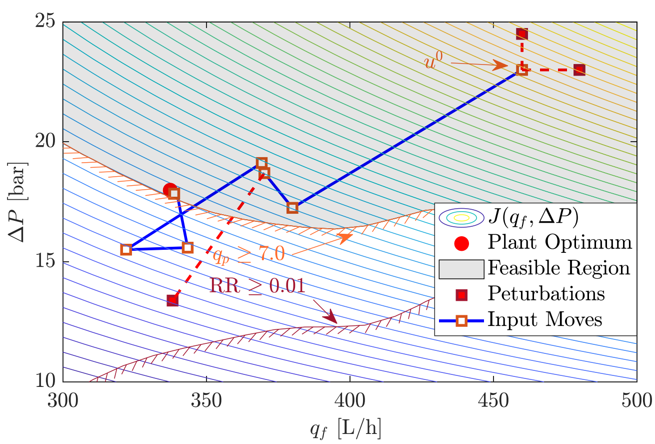


Fig. 13. Experimental results of the iterative optimization using MAWQA with EMA: Sequence of set-points in the input space with additional perturbations. The mapping of $J(q_f, \Delta P)$ is illustrated by the contour lines. The shaded area in the figure marks the feasible region as computed by the surrogate model.

A comparison of the two-step, MAWQA, MAWQA with MA, and MAWQA with EMA schemes was performed in simulations with a surrogate model that was built based on experimental data. The MAWQA, MAWQA with MA and MAWQA with EMA schemes converge to the true plant optimum whereas the two-step approach converges to a significantly sub-optimal point. MAWQA with EMA achieved significantly faster convergence to the optimum of the true plant compared to MAWQA and MAWQA with MA. This results from the model adaptation are not adequate during the initial iterations in MAQWA, therefore the first-order correction in the modified optimization problem leads to moves that do not approach the optimal point.

The obtained results show that the production cost can be reduced significantly by applying the MAWQA with EMA iterative RTO approach despite the fact that the original plant model is inaccurate and the optimum which is computed based on the original plant model violates the constraints.

In the future work, we will put into practice the idea that iterations should be stopped at a certain point, since repeated moves of control inputs are often undesired in industrial plants, and resumed if model errors are detected. First proposals in this direction can be found in Ye et al. (2018) and Mukkula et al. (2019).

CRediT authorship contribution statement

Afaq Ahmad: Writing – original draft, Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Radoslav Paulen:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Funding acquisition, Formal analysis, Data curation. **Richard Valo:** Writing – review & editing. **Miroslav Fikar:** Writing – review & editing. **Sebastian Engell:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

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.

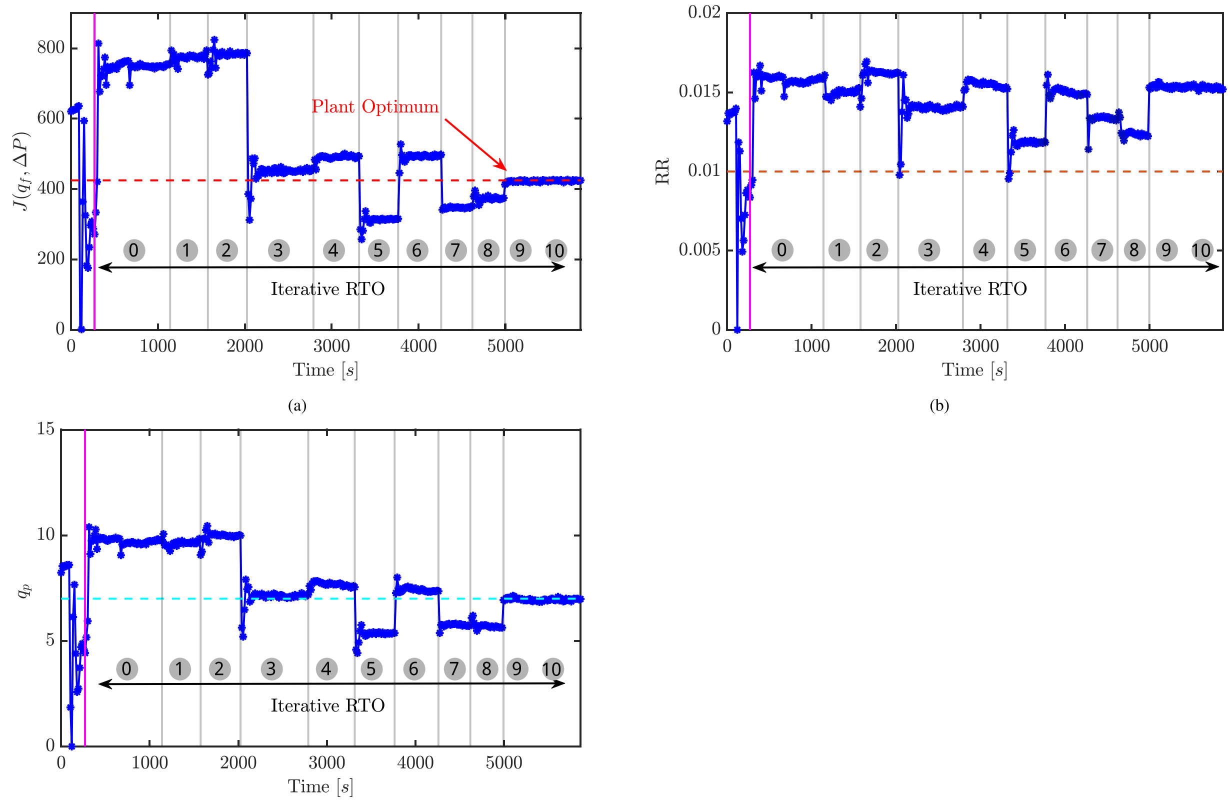
Acknowledgments

The research leading to these results has received funding from Deutscher Akademischer Austauschdienst (DAAD) and The Ministry of Education, Science, Research and Sport of the Slovak Republic under the grant "Reliable and Real-time Feasible Estimation and Control of Chemical Plants". RP, RV, and MF acknowledge the contribution of the Scientific Grant Agency of the Slovak Republic under the grant 1/0691/21, by the Slovak Research and Development Agency under the project APVV-21-0019, and by the European Commission under the grant no. 101079342 (Fostering Opportunities Towards Slovak Excellence in Advanced Control for Smart Industries). The authors appreciate fruitful discussions with Dr. Zoltán Kovács from Szent Istvan University, Budapest, Hungary and his comments on the work.

Appendix A. Model adequacy

Forbes and Marlin (1996) first addressed the issue of model adequacy. In an RTO scheme, a process model is called adequate if a model-based optimization is able to converge to a point that is a local minimum for the RTO problem at the optimum of the plant $\mathbf{u}_p^*$. Specifically, a model is deemed to be adequate for use in RTO when an optimum of the plant $\mathbf{u}_p^*$ meets both the first and second order criteria of optimality for Eq. (10).

MA converges to a set-point $\mathbf{u}^{(\infty)}$, which satisfies the first-order conditions of the optimality. No guarantee is provided that $\mathbf{u}^{(\infty)}$ is a local minimum for the optimization problem of the plant Eq. (10).



(a) (b) (c)

Fig. 14. Experimental results: (a) Evolution of the operational cost. The red dashed line illustrates the plant optimum. (b) Evolution of the experimental recovery ratio (RR). The orange dashed line represents the minimum threshold of the RR. (c) Evolution of the permeate flowrate (q_p). The cyan dashed line illustrates the minimum threshold of the permeate flowrate q_p. The experiment started at the magenta line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Adequacy of the model requires a positive definite reduced Hessian of the adapted optimization problem Eq. (13) at the real optimum of the plant $\mathbf{u}_p^*$ (Marchetti et al., 2016).

$$\nabla^2 \mathcal{L}_{ad,r}\left(\mathbf{u}_p^*, \boldsymbol{\theta}, \boldsymbol{\mu}^*\right) > 0, \quad \text{(positive definite)} \tag{A.1}$$

where $\mathcal{L}_{ad}$ is the Lagrangian function of the problem of the adapted model-based optimization Eq. (13).

$$\mathcal{L}_{ad}(\mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\mu}) := \quad J_{ad}(\mathbf{u}, \boldsymbol{\theta}) + \boldsymbol{\mu}^T \mathbf{C}_{ad}(\mathbf{u}, \boldsymbol{\theta}), \tag{A.2}$$

$\boldsymbol{\mu}$ represents a Lagrange multiplier that is a n_C-dimensional vector.

Appendix B. Modifier adaptation with quadratic approximation in combination with EMA

The accuracy of the estimation of the gradients for the real plant and the satisfaction of the adequacy of the model are two major challenges for MA to converge to the true optimum of the plant. The MAWQA method combines MA with quadratic approximation (QA) that is used in derivative free optimization (DFO) to address these issues (Gao et al., 2016).

The fundamental concept is to estimate the cost of the plant $J_p(\mathbf{u})$ and the constraint functions $\mathbf{C}_p(\mathbf{u})$ by fitting general quadratic functions to the measurements as follows:

$$\mathcal{Q}(\mathbf{u}, \mathcal{P}) = \sum_{i=1}^{n_u} \sum_{j=1}^{i} a_{i,j} u_i u_j + \sum_{i=1}^{n_u} b_i u_i + c, \tag{B.1}$$

with the set of parameters $\mathcal{P} = \left\{a_{1,1}, \dots, a_{n_u,n_u}, b_1, \dots, b_{n_u}, c\right\}$. Assume that $\mathbb{U}^{(k)}$ is the collection of all available steady-state measurements of the plant up to the kth iteration. A subset $\mathcal{U}^{(k)}$ is selected from $\mathbb{U}^{(k)}$ in the kth iteration for QA such that the points in $\mathcal{U}^{(k)}$ are well distributed around $\mathbf{u}^{(k)}$ in the input space. The Regression set $\mathcal{U}^{(k)}$ is defined as $\mathcal{U}^{(k)} = \mathcal{U}_{nb} \cup \mathcal{U}_{dist}$,

$$\mathcal{U}_{nb} = \{\mathbf{u} : ||\mathbf{u} - \mathbf{u}^{(k)}|| \leq \Delta\mathbf{u}; \mathbf{u} \in \mathbb{U}\},$$

and the distant set $\mathcal{U}_{dist}$ are determined by

$$\min_{\mathcal{U}_{dist}} \quad \frac{\sum_{\mathbf{u} \in \mathcal{U}_{dist}} ||\mathbf{u} - \mathbf{u}^{(k)}||}{\varphi(\mathcal{U}_{dist})} \tag{B.2}$$
$$\text{s.t.} \quad \text{size}(\mathcal{U}_{dist}) = (n_u + 1)(n_u + 2)/2 - 1 \quad \mathcal{U}_{dist} \subset \mathbb{U} \setminus \mathcal{U}_{nb},$$

where $\Delta\mathbf{u}$ is a screening parameter and $\varphi(\mathcal{U}_{dist})$ is the minimal angle between all possible vectors that are defined by $\mathbf{u} - \mathbf{u}^{(k)}$. The set $\mathbb{U}^{(k)}$ must have a sufficient number of steady-state measurements of the plant to construct the quadratic cost and constraint functions, at least $(n_u+1)(n_u+2)/2$ points. The accuracy of the QA relies on how the points are distributed in $\mathcal{U}^{(k)}$. The quality of the distribution can be described by the condition number ($\mathbf{S}^{(k)}$) of the subset $\mathcal{U}^{(k)}$ (Brdyś & Tatjewski, 1995; Gao & Engell, 2005). Gao et al. (2016) and Wenzel et al. (2017) proposed two different screening algorithms to identify $\mathcal{U}^{(k)}$ from $\mathbb{U}^{(k)}$ in the kth iteration of QA.

The quadratic approximations of the plant cost $J_Q(\mathbf{u})$ and of the constraint functions $\mathbf{C}_Q(\mathbf{u})$ are obtained by fitting $\mathcal{Q}(\mathbf{u}, \mathcal{P})$ to the observed values for the points in $\mathcal{U}^{(k)}$. Then the estimated gradients of the cost $\nabla J_p(\mathbf{u}^{(k)})$ and of the constraints $\nabla \mathbf{C}_p(\mathbf{u}^{(k)})$ can be calculated analytically as the gradients of the function (B.1). However, there is no guarantee that the quadratic functions perfectly approximate the available plant measurements. Therefore, the next input movement is restricted to lie within a trust region around $\mathbf{u}^{(k)}$:

$$\mathcal{B}^{(k)} : (\mathbf{u} - \mathbf{u}^{(k)})^T M^{-1} (\mathbf{u} - \mathbf{u}^{(k)}) \leq \gamma^2, \tag{B.3}$$

where $M = \mathrm{cov}(\mathcal{U}^{(k)})$ is the covariance matrix and γ is a scaling parameter.

MAWQA ensures the convergence to the process optimum by monitoring the quality of the model and switching between the modified optimization problem (13) and a purely data-based optimization based on QA:

$$\begin{aligned} \min_{\mathbf{u}\in\mathcal{B}^{(k)}} \quad & J_Q(\mathbf{u}) := \tilde{J}(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \\ \text{s.t.} \quad & \mathbf{C}_Q(\mathbf{u}) := \tilde{\mathbf{C}}(\mathbf{u}, \mathbf{y}_p(\mathbf{u})) \leq \mathbf{0}, \\ & \mathbf{u}_{min} \leq \mathbf{u} \leq \mathbf{u}_{max}. \end{aligned} \tag{B.4}$$

The choice is made by comparing the quality of the modified cost and constraint functions in (13) and the approximated quadratic functions in (B.4) using the following equations:

$$\rho_m^{(k)} := \max\left\{ \begin{aligned} &\left|1 - \frac{J_{ad}^{(k)} - J_{ad}^{(k-1)}}{J_p^{(k)} - J_p^{(k-1)}}\right|, \left|1 - \frac{C_{ad,1}^{(k)} - C_{ad,1}^{(k-1)}}{C_{p,1}^{(k)} - C_{p,1}^{(k-1)}}\right|, \\ &\ldots, \left|1 - \frac{C_{ad,n_c}^{(k)} - C_{ad,n_c}^{(k-1)}}{C_{p,n_c}^{(k)} - C_{p,n_c}^{(k-1)}}\right| \end{aligned} \right\}, \tag{B.5}$$

$$\rho_Q^{(k)} := \max\left\{ \begin{aligned} &\left|1 - \frac{J_{Q}^{(k)} - J_{Q}^{(k-1)}}{J_p^{(k)} - J_p^{(k-1)}}\right|, \left|1 - \frac{C_{Q,1}^{(k)} - C_{Q,1}^{(k-1)}}{C_{p,1}^{(k)} - C_{p,1}^{(k-1)}}\right|, \\ &\ldots, \left|1 - \frac{C_{Q,n_c}^{k} - C_{Q,n_c}^{k-1}}{C_{p,n_c}^{k} - C_{p,n_c}^{k-1}}\right| \end{aligned} \right\}. \tag{B.6}$$

For the kth iteration, if $\rho_Q^{(k)} < \rho_m^{(k)}$, (B.4) is solved to compute the next input move $\mathbf{u}^{(k+1)}$ and if $\rho_Q^{(k)} \geq \rho_m^{(k)}$, (13) is solved.

In Section 3.3, it is mentioned that prior to the availability of $(n_u + 1)(n_u + 2)/2$ data points, it is necessary to compute the gradients based on the past moves or probing moves. The efficiency of an iterative RTO scheme in achieving the optimal operation quickly holds significant importance. Consequently, we merge MAWQA with EMA such that the advantages of both methodologies are combined. Fig. 3 presents the integration of EMA in MAWQA algorithm. The input move $\mathbf{u}^{(k+1)}$ is applied to the plant, and model parameters are estimated via problem (22). In EMA, the model parameters are adapted only if the adapted plant model meets the model adequacy conditions, so that oscillations of the iterative optimization are avoided. EMA adopts the new parameters if the conditions (19) and (20) are satisfied. The first condition (19) of EMA ensures convergence to the true plant optimum, while condition (20) ensures faster convergence to the process optimum.

References

Ahmad, A., Gao, W., & Engell, S. (2017). Effective Model Adaptation in Iterative RTO. *Computer Aided Chemical Engineering*, *40*, 1717–1722.

Ahmad, A., Gao, W., & Engell, S. (2018). Modifier adaptation with model adaptation in iterative real-time optimization. *Computer Aided Chemical Engineering*, *44*, 691–696.

Ahmad, A., Gao, W., & Engell, S. (2019). A study of model adaptation in iterative real-time optimization of processes with uncertainties. *Computers & Chemical Engineering*, *122*, 218–227.

Ahmad, A., Gottu Mukkula, A. R., & Engell, S. (2019). Model adaptation with quadratic approximation in iterative real-time optimization. In *2019 22nd international conference on process control* (pp. 250–255). IEEE.

Ahmad, F., Lau, K. K., Shariff, A. M., & Murshid, G. (2012). Process simulation and optimal design of membrane separation system for CO2 capture from natural gas. *Computers & Chemical Engineering*, *36*, 119–128.

Ahmad, A., Singhal, M., Gao, W., Bonvin, D., & Engell, S. (2018). Enforcing model adequacy in real-time optimization via dedicated parameter adaptation. *IFAC-PapersOnLine*, *51*(18), 49–54.

Artug, G. (2007). *Modelling and simulation of nanofiltration membranes*. Cuvillier Verlag.

Brdyś, M., & Tatjewski, P. (1995). An algorithm for steady-state optimizing dual control of uncertain plants. In *IFAC conference on new trends in design of control systems 1994* (pp. 215–220). Elsevier.

Bunin, G. A., François, G., & Bonvin, D. (2013a). Sufficient conditions for feasibility and optimality of real-time optimization schemes-I. Theoretical foundations. arXiv preprint arXiv:1308.2620.

Bunin, G. A., François, G., & Bonvin, D. (2013b). Sufficient conditions for feasibility and optimality of real-time optimization schemes-II. Implementation issues. arXiv preprint arXiv:1308.2625.

Bunin, G. A., Wuillemin, Z., François, G., Nakajo, A., Tsikonis, L., & Bonvin, D. (2012). Experimental real-time optimization of a solid oxide fuel cell stack via constraint adaptation. *Energy*, *39*(1), 54–62.

Chen, C. Y., & Joseph, B. (1987). On-line optimization using a two-phase approach: An application study. *Industrial and Engineering Chemistry Research*, *26*(9), 1924–1930.

Chen, Z., Luo, J., Wang, Y., Cao, W., Qi, B., & Wan, Y. (2017). A novel membrane-based integrated process for fractionation and reclamation of dairy wastewater. *Chemical Engineering Journal*, *313*, 1061–1070.

Cheryan, M. (1998). *Ultrafiltration and microfiltration handbook*. CRC Press.

Conidi, C., Cassano, A., Caiazzo, F., & Drioli, E. (2017). Separation and purification of phenolic compounds from pomegranate juice by ultrafiltration and nanofiltration membranes. *Journal of Food Engineering*, *195*, 1–13.

Conn, A. R., Scheinberg, K., & Vicente, L. N. (2009). *Introduction to derivative-free optimization*: *vol. 8*, Siam.

Darby, M. L., Nikolaou, M., Jones, J., & Nicholson, D. (2011). RTO: An overview and assessment of current practice. *Journal of Process Control*, *21*(6), 874–884.

Engell, S. (2007). Feedback control for optimal process operation. *Journal of Process Control*, *17*(3), 203–219.

Forbes, J. F., & Marlin, T. E. (1996). Design cost: A systematic approach to technology selection for model-based real-time optimization systems. *Computers & Chemical Engineering*, *20*(6–7), 717–734.

Forbes, J., Marlin, T., & MacGregor, J. (1994). Model adequacy requirements for optimizing plant operations. *Computers & Chemical Engineering*, *18*(6), 497–510.

François, G., & Bonvin, D. (2013). Use of convex model approximations for real-time optimization via modifier adaptation. *Industrial and Engineering Chemistry Research*, *52*(33), 11614–11625.

Gao, W., & Engell, S. (2005). Iterative set-point optimization of batch chromatography. *Computers & Chemical Engineering*, *29*(6), 1401–1409.

Gao, W., Wenzel, S., & Engell, S. (2015). Integration of gradient adaptation and quadratic approximation in real-time optimization. In *34th Chinese control conference* (pp. 2780–2785).

Gao, W., Wenzel, S., & Engell, S. (2016). A reliable modifier-adaptation strategy for real-time optimization. *Computers & Chemical Engineering*, *91*, 318–328.

Guadix, A., Sørensen, E., Papageorgiou, L. G., & Guadix, E. M. (2004). Optimal design and operation of continuous ultrafiltration plants. *Journal of Membrane Science*, *235*(1–2), 131–138.

Hernandez, R., Dreimann, J., Vorholt, A., Behr, A., & Engell, S. (2018). Iterative real-time optimization scheme for optimal operation of chemical processes under uncertainty: Proof of concept in a miniplant. *Industrial and Engineering Chemistry Research*, *57*(26), 8750–8770.

Jaffrin, M., & Charrier, J. P. (1994). Optimization of ultrafiltration and diafiltration processes for albumin production. *Journal of Membrane Science*, *97*, 71–81.

Jang, S.-S., Joseph, B., & Mukai, H. (1987). On-line optimization of constrained multivariable chemical processes. *AIChE Journal*, *33*(1), 26–35.

Mandur, J. S., & Budman, H. M. (2015). Simultaneous model identification and optimization in presence of model-plant mismatch. *Chemical Engineering Science*, *129*, 106–115.

Marchetti, A. G., Chachuat, B., & Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *Industrial and Engineering Chemistry Research*, *48*(13), 6022–6033.

Marchetti, A. G., François, G., Faulwasser, T., & Bonvin, D. (2016). Modifier adaptation for real-time optimization – [methods] and applications. *Processes*, *4*(4), 55.

Mukkula, A. R. G., Ahmad, A., & Engell, S. (2019). Start-up and shut-down conditions for iterative real-time optimization methods. In *2019 sixth Indian control conference* (pp. 158–163). IEEE.

Mukkula, A. G., Kern, S., Salge, M., Holtkamp, M., Guhl, S., Fleicher, C., Meyer, K., Remelhe, M., Maiwald, M., & Engell, S. (2020). An application of modifier adaptation with quadratic approximation on a pilot scale plant in industrial environment. *IFAC-PapersOnLine*, *53*(2), 11773–11779.

Mukkula, A. R. G., Valiauga, P., Fikar, M., Paulen, R., & Engell, S. (2020). Experimental real time optimization of a continuous membrane separation plant. In *21st IFAC world congress, vol. 21* (pp. 11967–11974). Preprints.

Navia, D., Villegas, D., Cornejo, I., & de Prada, C. (2016). Real-time optimization for a laboratory-scale flotation column. *Computers & Chemical Engineering*, *86*, 62–74.

Naysmith, M., & Douglas, P. (1995). Review of real time optimization in the chemical process industries. *Developments in Chemical Engineering and Mineral Processing*, *3*(2), 67–87.

Noronha, M., Mavrov, V., & Chmiel, H. (2002). Efficient design and optimisation of two-stage NF processes by simplified process simulation. *Desalination*, *145*(1–3), 207–215.

Paulen, R., & Fikar, M. (2019). Dynamic real-time optimization of batch processes using Pontryagin's minimum principle and set-membership adaptation. *Computers & Chemical Engineering*, *128*, 488–495.

Roberts, P. D. (1979). An algorithm for steady-state system optimization and parameter estimation. *International Journal of Systems Science*, *10*(7), 719–734.

Roberts, P. D., & Williams, T. (1981). On an algorithm for combined system optimisation and parameter estimation. *Automatica*, *17*(1), 199–209.

Salehi, F. (2014). Current and future applications for nanofiltration technology in the food processing. *Food and Bioproducts Processing*, *92*(2), 161–177.

Schock, G., & Miquel, A. (1987). Mass transfer and pressure loss in spiral wound modules. *Desalination*, *64*, 339–352.

Scott, K., & Hughes, R. (1996). *Industrial membrane separation technology*. Springer Science & Business Media.

Sharma, A., Valo, R., Kalúz, M., Paulen, R., & Fikar, M. (2018). Experimental validation and comparison of time-optimal and industrial strategy for membrane separation process. *IFAC-PapersOnLine*, *51*(2), 741–746.

Sharma, A., Valo, R., Kalúz, M., Paulen, R., & Fikar, M. (2019). Implementation of optimal strategy to economically improve batch membrane separation. *Journal of Process Control*, *76*, 155–164.

Tatjewski, P. (2002). Iterative optimizing set-point control – The basic principle redesigned. *IFAC Proceedings Volumes*, *35*(1), 49–54. http://dx.doi.org/10.3182/20020721-6-ES-1901.00994, URL http://www.sciencedirect.com/science/article/pii/S1474667015394155, 15th IFAC World Congress.

Wenzel, S., Yfantis, V., & Gao, W. (2017). Comparison of regression data selection strategies for quadratic approximation in RTO. *Computer Aided Chemical Engineering*, *40*, 1711–1716.

Yazdanshenas, M., Tabatabaeenezhad, A., Roostaazad, R., & Khoshfetrat, A. (2005). Full scale analysis of apple juice ultrafiltration and optimization of diafiltration. *Separation and Purification Technology*, *47*(1–2), 52–57.

Ye, L., Shen, F., Ge, Z., & Song, Z. (2018). On an aspect of implementing real-time optimization: Establishing the suspending and activating conditions incorporating process monitoring. *IFAC-PapersOnLine*, *51*(18), 79–84.

Zeman, L. J., & Zydney, A. (1996). *Microfiltration and ultrafiltration: principles and applications*. CRC Press.