Journal of Process Control 92 (2020) 333–351

Contents lists available at ScienceDirect

Journal of Process Control

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

A hybrid framework for process monitoring: Enhancing data-driven methodologies with state and parameter estimation

Francesco Destro [a], Pierantonio Facco [a], Salvador García Muñoz [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] Small Molecule Design & Development, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN, 46285, USA

ARTICLE INFO

Article history:
Received 26 September 2019
Received in revised form 21 April 2020
Accepted 10 June 2020
Available online 16 July 2020

Keywords:
Fault detection
Fault diagnosis
Process monitoring
Hybrid modeling
State estimation
Industry 4.0
Extended Kalman filter

ABSTRACT

In this study we bridge traditional standalone data-driven and knowledge-driven process monitoring approaches by proposing a novel hybrid framework that exploits the advantages of both simultaneously. Namely, we design a process monitoring system based on a data-driven model that includes two different data types: *i*) "actual" data coming from sensor measurements, and *ii*) "virtual" data coming from a state estimator, based on a first-principles model of the system under investigation. We test the proposed approach on two simulated case studies: a continuous polycondensation process for the synthesis of poly-ethylene terephthalate, and a fed-batch fermentation process for the manufacturing of penicillin. The hybrid monitoring model shows superior fault detection and diagnosis performances with respect to conventional monitoring techniques, even when the first-principles model is relatively simple and process/model mismatch exists.

1. Introduction

Process monitoring is a key task in the process industry, as detecting a fault and assessing its cause before the production is compromised can save valuable assets. Several data-driven (DD) methodologies for fault detection and diagnosis have been proposed in the last decades [1,2]. Among them, latent-variables models [3] (LVMs) are a powerful class of DD multivariate approaches that proved very effective for fault detection and diagnosis [4], and gained increased relevance with the Industry 4.0 "big data" era. LVMs detect a fault when new measurements coming from the plant sensors are unknown to the correlation structure of the training data, which define the normal operating conditions (NOC) for the process. Faults are typically detected using multivariate control charts [5]. Contribution plots [6] can then be exploited to pinpoint the measurements most related to the faulty conditions. However, when the number of available measurements is relatively small, detection of a fault through an LVM may be delayed, because the fault must propagate into the system until the few measured variables are affected. On the other side, diagnosing the root-cause of a fault may be challenging due to the smearing-out effect [1]. In fact, since LVMs are not cause–effect models, it may be difficult to identify causality patterns between measurements under abnormal process conditions. This issue is particularly relevant if the variables embodying the root-cause of the fault are not measured, and therefore cannot be included in the LVM.

To overcome this limitation, monitoring methodologies exploiting first-principles knowledge about the process under investigation may be considered. Process monitoring methodologies based on knowledge-driven (KD) models have been thoroughly reviewed elsewhere [7,8]. The most popular KD approaches are based on parity relations [9] or on state estimators [10–14], possibly implemented for simultaneous state and parameter estimation [15]. Generally speaking, KD models have the advantage of embedding the available understanding on the mechanisms driving the process under investigation. This piece of information can help fault detection and diagnosis, and is missing in DD monitoring approaches. However, KD models are generally more complex to develop than their DD counterparts and, when used for monitoring, the performances can be severely affected by process-model mismatch. In addition, the fault models have typically to be known a priori [13,14].

Hybrid models [16,17] combine DD methods with the information available from first-principles knowledge about the process, and are promising techniques for overcoming the limitations of DD and KD monitoring [18,19]. Hybrid models for process monitoring usually consist of a KD soft-sensing framework in which a DD component is added to make up for missing

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

https://doi.org/10.1016/j.jprocont.2020.06.002

deterministic information [20,21]. Other hybrid approaches [22] use DD techniques to monitor the residuals of a KD model, or they develop complex schemes with subsequent DD and KD steps, tailored to specific applications [23]. Recent contributions [24, 25] showed the benefits of building advanced control charts to monitor the states estimated by a state estimator. However, these approaches are typically univariate, and therefore suffer from the well-known limitations of univariate monitoring with respect to its multivariate counterpart [3,4,26]. Even though all these methodologies are gaining interest, state-of-the-art hybrid monitoring methods still lack a general framework to combine results coming from DD and KD modeling approaches [22].

In this study, we couple the easy-to-design features of DD process monitoring approaches to the descriptive capability of KD models, in order to develop a novel multivariate monitoring methodology with improved fault detection and diagnostic capabilities. Namely, we propose a latent-variable-based monitoring model that uses an augmented data matrix including two different data types: (*i*) "actual" data coming from sensor measurements, and (*ii*) "virtual" data coming from a dynamic KD model able to capture the main features of the system under investigation. Process-model mismatch for the KD model is (at least partially) compensated for by using a state estimator, which returns a set of virtual data consisting of estimated system states, adapted parameters, and reconstructed measurements to be included in the DD model. We test the proposed methodology on two simulated processes: a continuous poly-ethylene terephthalate (PET) polymerization process, and a fed-batch fermentation process for the production of penicillin.

The remainder of this article is organized as follows. In Section 2 the mathematical methodologies later applied are briefly summarized. The proposed hybrid monitoring framework is outlined in Section 3. The case studies are presented in Section 4, and the results are discussed in Sections 5 and 6. Some conclusions to the study are finally reported in Section 7.

2. Mathematical background

2.1. Process monitoring by extended Kalman filtering

Let the first-principles model (FPM) of a dynamic system be expressed as a set of ordinary differential equations:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) + \mathbf{w}(t) \tag{1}$$

where $\mathbf{f}$ is a nonlinear function, $\mathbf{x}(t)$ denotes the system state vector at time t, $\mathbf{u}(t)$ is the input vector, and $\mathbf{w}(t)$ is the process noise vector, which is assumed to be a white Gaussian process with mean $\mathbf{0}$ and covariance $\mathbf{Q}(t)$. Assuming that measurements are available at discrete time steps t_k from the plant, they can be related to the system states through an appropriate measurement model:

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), t) + \mathbf{v}(t) \tag{2}$$

where $\mathbf{y}(t)$ is the measurement vector, and $\mathbf{v}(t)$ is the measurement noise vector, which is assumed to be a white Gaussian process with mean $\mathbf{0}$ and covariance $\mathbf{R}(t)$.

The discrete time data extended Kalman filter (EKF) [27] provides the estimated state vector $\hat{\mathbf{x}}(t)$ and the state covariance $\mathbf{P}(t)$, given the initial estimation of the states $\hat{\mathbf{x}}_0$ and the initial state covariance $\mathbf{P}_0$. The algorithm includes two steps, prediction and update, which are alternatively performed at each time point $k = 1, 2, \ldots, K$. The predictions of the states and of the state covariance at time t_k *before* the measurements are available (t_k^-) are respectively referred to as $\hat{\mathbf{x}}(t_k^-)$ and $\mathbf{P}(t_k^-)$, whereas the corrected estimations *after* the sampling time (t_k^+) are denoted as $\hat{\mathbf{x}}(t_k^+)$ and $\mathbf{P}(t_k^+)$.

During the prediction step, $\hat{\mathbf{x}}(t_{k-1}^+)$ and $\mathbf{P}(t_{k-1}^+)$ are propagated, with the integration of Eqs. (3)–(4), to obtain (respectively) $\hat{\mathbf{x}}(t_k^-)$ and $\mathbf{P}(t_k^-)$:

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), t) \tag{3}$$

$$\dot{\mathbf{P}}(t) = \mathbf{F}\mathbf{P} + \mathbf{P}\mathbf{F}^{\mathrm{T}} + \mathbf{Q}(t) \tag{4}$$

where $\mathbf{F}$ is the Jacobian matrix:

$$\mathbf{F} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)_{\hat{\mathbf{x}}(t),u(t),t} \tag{5}$$

At each sampling point k, the predicted estimations are corrected with the update equations:

$$\hat{\mathbf{x}}(t_k^+) = \hat{\mathbf{x}}(t_k^-) + \mathbf{K}(t_k)\boldsymbol{\gamma}(t_k) \tag{6}$$

$$\mathbf{P}(t_k^+) = \mathbf{P}(t_k^-) - \mathbf{K}(t_k)\mathbf{H}_k\mathbf{P}(t_k^-) \tag{7}$$

where the Kalman gain $\mathbf{K}(t_k)$, the innovation $\boldsymbol{\gamma}(t_k)$, and the Jacobian matrix $\mathbf{H}_k$ are respectively calculated with:

$$\mathbf{K}(t_k) = \mathbf{P}(t_k^-)\mathbf{H}_k^{\mathrm{T}}\mathbf{V}(t_k)^{-1} \tag{8}$$

$$\boldsymbol{\gamma}(t_k) = \mathbf{y}(t_k) - \mathbf{h}(\hat{\mathbf{x}}(t_k^-), \mathbf{u}(t_k), t_k) \tag{9}$$

$$\mathbf{H}_k = \left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right)_{\hat{\mathbf{x}}(t_k^-),\mathbf{u}(t_k),t_k} \tag{10}$$

The matrix $\mathbf{V}(t_k)$ in Eq. (8) is calculated with:

$$\mathbf{V}(t_k) = \mathbf{H}_k\mathbf{P}(t_k^-)\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}(t_k) \tag{11}$$

Rigorous methods to design $\mathbf{P}_0$, $\mathbf{Q}$ and $\mathbf{R}$ exist [28]. In this study, we take a simpler approach that proved effective also in several other studies [29–31]. Namely, we design $\mathbf{P}_0$ as a diagonal matrix, based on the expected uncertainty on the initial estimation for each state. Additionally, we set $\mathbf{Q}$ as a time-invariant diagonal matrix, and we tune it by trial and error in such a way as to obtain robust convergence. Finally, we design $\mathbf{R}$ as:

$$\mathbf{R} = \mathrm{diag}(\sigma_i), \tag{12}$$

where σ_i is the variance of the noise of the i-th measurement sensor.

The EKF provides reconstructed values for $\mathbf{y}(t_k)$. Specific implementations of the EKF can be set up for tackling colored noise or measurement bias [32], if that emerges useful for monitoring.

The EKF can also perform online parameter adaptation, upon augmentation of the state vector $\hat{\mathbf{x}}$ with the subset $\hat{\mathbf{p}}$ of the FPM parameters that one seeks to adapt in real time. The nominal values $\mathbf{p}_0$ of the parameters are taken as initial conditions for $\hat{\mathbf{p}}$, and negligible dynamics with additive process noise of small variance can be assumed. State augmentation can be resorted to for improving the estimation of unmeasured states through bias estimation [33], too. Bias estimation consists in the insertion of an additive term (bias) at the right-hand side of Eq. (3) for selected states. Biases are assigned null initial values and a random walk model of small variance. To meet the observability conditions, the augmented states cannot exceed in number the available measurements.

Conventional KD fault detection approaches [11,12] are based on the assumption that the innovation sequence $\boldsymbol{\gamma}(t)$ of the EKF without augmented states follows a white $\mathbf{0}$ mean Gaussian distribution with covariance $\mathbf{V}(t)$. Deviations from this behavior

indicate a fault and can be detected by monitoring for each time instant the test statistic in Eq. (13), which follows the central χ^2 distribution with R degrees of freedom (where R is the number of measurements in $\mathbf{y}$):

$$\epsilon\,(t_k) = \boldsymbol{\gamma}\,(t_k)^{\mathrm{T}}\,\mathbf{V}\,(t_k)^{-1}\,\boldsymbol{\gamma}(t_k). \tag{13}$$

Upon rejection of the null hypothesis for $\epsilon\,(t_k)$ at time t_k, the innovation sequence over a time window of size S (tuning parameter) is used for confirming the fault condition, with the following test statistic:

$$\epsilon\,(t_k, S) = \sum_{i=k}^{k+S} \boldsymbol{\gamma}\,(t_i)^{\mathrm{T}}\,\mathbf{V}\,(t_i)^{-1}\,\boldsymbol{\gamma}(t_i). \tag{14}$$

If $\epsilon\,(t_k, S)$ violates the confidence limit for the central χ^2 distribution with $R\cdot S$ degrees of freedom, the fault is confirmed. Fault isolation is then typically carried out by online adaptation of all the parameters related to possible fault conditions, and by identifying which parameter is drifting from its nominal value [34,35].

2.2. Multivariate process monitoring by principal component analysis

Principal component analysis (PCA) is a dimensionality reduction technique aimed at extracting the few underlying factors (called principal components, PCs) that explain most of the variability from a NOC dataset $\mathbf{Z}$ $[N\times M]$ including N observations on M variables [3]. The data matrix $\mathbf{Z}$ is auto-scaled to zero mean and unit variance, and the PCs are extracted by decomposing $\mathbf{Z}$ as [36]:

$$\mathbf{Z} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{\mathrm{T}} + \mathbf{E}, \tag{15}$$

where $\mathbf{t}_a$ are $[N\times 1]$ score vector of the a-th PC, $\mathbf{p}_a$ $[M\times 1]$ is the loading vector for the same PC, and $\mathbf{E}$ is the matrix of residuals, which are random noise if an appropriate number A of PCs is selected. For each observation n, two monitoring statistics can be calculated from the PCA model, namely the Hotelling T^2 and the squared prediction error (SPE), according to:

$$T_n^2 = \sum_{a=1}^{A} t_{a,n}\,\lambda_a^{-1} t_{a,n} \tag{16}$$

$$\mathrm{SPE}_n = \boldsymbol{e}_n \boldsymbol{e}_n^{\mathrm{T}} \tag{17}$$

where $t_{a,n}$ is the element in $\mathbf{t}_a$ corresponding to observation n, λ_a is the eigenvalue associated to the a-th PC, and $\boldsymbol{e}_n$ is the residual vector. Confidence limits can be obtained for both statistics from the available set of NOC. In this study, we obtained the confidence limits $\mathrm{SPE}_{\mathrm{lim}}$ (on SPE) and T^2_{lim} (on T^2) by means (respectively) of the Jackson-Mudholkar equation [36], and of the F-distribution confidence limit equation [36]. We calculated both limits at 99% confidence.

When a new observation $\mathbf{z}_{\mathrm{new}}$ $[1\times M]$ becomes available from the process at time t_{new}, it is first normalized on the mean and variance of the NOC dataset, then it is projected onto the model space, and finally $\mathrm{SPE}_{\mathrm{new}}$ and T^2_{new} are calculated. If at least one of the two statistics exceeds its relevant confidence limit for some (e.g., three) consecutive observations, a fault is alarmed. The fault can be diagnosed using contribution plots [6], which point to the variables included in $\mathbf{Z}$ that most contribute to the confidence limit violation. The $[1\times M]$ contribution vectors for $\mathbf{z}_{\mathrm{new}}$ can be built by calculating the contributions for each variable m as:

$$\mathrm{c}^{\mathrm{SPE}}_{\mathrm{new},m} = e_{\mathrm{new},m}, \tag{18}$$

$$\mathrm{c}^{T^2}_{\mathrm{new},m} = \sum_{a=1}^{A} t_{a,new}\,\lambda_a^{-0.5} p_{a,m}, \tag{19}$$

where $e_{\mathrm{new},m}$ and $p_{a,m}$ are the elements corresponding to variable m respectively in $\boldsymbol{e}_{\mathrm{new}}$ and $\mathbf{p}_a$. We use the residuals as contributions to SPE in order to preserve the sign of the error, which proves useful for fault diagnosis. The vector of residuals $\mathbf{c}^{\mathrm{SPE}}_{\mathrm{new}}$ is known to follow a normal distribution if A is selected appropriately [36]. For the case studies under investigation, we found that also the $\mathbf{c}^{T^2}_{\mathrm{NEW}}$ contributions are normally distributed, an occurrence that has been noted also by other investigators [37,38]. Therefore, we calculated Gaussian control limits (at 99% confidence) for both contributions. Although these limits should not be considered to have statistical significance, they are nevertheless helpful for comparing contributions presenting significantly different magnitudes also with respect to the NOC.

When the observations in $\mathbf{Z}$ are auto-correlated, the standard PCA approach should be modified to account for the effect of time. For example, dynamic PCA (DPCA, [39]) can be used, where L lagged measurements are included in $\mathbf{Z}$ to obtain the NOC matrix $\mathbf{Z}_{dyn}$:

$$\mathbf{Z}_{dyn} = \begin{bmatrix} \mathbf{z}^{\mathrm{T}}(t_1) & \mathbf{z}^{\mathrm{T}}(t_0) & \dots & \mathbf{z}^{\mathrm{T}}(t_1 - L) \\ \mathbf{z}^{\mathrm{T}}(t_2) & \mathbf{z}^{\mathrm{T}}(t_1) & \dots & \mathbf{z}^{\mathrm{T}}(t_2 - L) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}^{\mathrm{T}}(t_M) & \mathbf{z}^{\mathrm{T}}(t_M - 1) & \dots & \mathbf{z}^{\mathrm{T}}(t_M - L) \end{bmatrix}. \tag{20}$$

Alternatively, a multi-model moving-window PCA (MW-PCA, [40]) approach can be resorted to, with a PCA model calibrated at each time point t_k on the data for the previous W time points, where W is the time window width. The NOC matrix $\mathbf{Z}_{MW}\,(t_{\mathrm{k}})$ for the PCA model at time t_k is defined as:

$$\mathbf{Z}_{MW}\,(t_{\mathrm{k}}) = \begin{bmatrix} \mathbf{z}^{\mathrm{T}}(t_k) & \mathbf{z}^{\mathrm{T}}(t_k - 1) & \dots & \mathbf{z}^{\mathrm{T}}(t_k - W) \end{bmatrix}. \tag{21}$$

3. Proposed hybrid monitoring framework

The proposed hybrid monitoring framework is based on the approach sketched in Figure 1. A process with measured inputs $\mathbf{u}$ and unknown disturbances $\mathbf{d}$ produces a set of measured outputs $\mathbf{y}$. The KD block exploits an FPM and the online measurements $\mathbf{u}$ and $\mathbf{y}$ to perform state estimation, online adaptation of the FPM parameters, and measurement reconstruction. In this study, an EKF [27] has been used to carry out these tasks, though other estimators might be used [10]. The DD block is based on an LVM that extracts operation-relevant information from the available set of field measurements ($\mathbf{u}$ and $\mathbf{y}$) as well as from the virtual data (estimated states $\hat{\mathbf{x}}$, adapted model parameters $\hat{\mathbf{p}}$, and reconstructed measurements $\hat{\mathbf{y}}$) returned by the KD block. The augmented overall data matrix becomes:

$$\mathbf{Z} = [\hat{\mathbf{x}}\ \hat{\mathbf{p}}\ \hat{\mathbf{y}}\ \mathbf{u}\ \mathbf{y}] \tag{22}$$

The advantage of this architecture is that, through the state estimator, the DD multivariate monitoring model can receive information related also to the inner mechanisms driving the system (in the form of states and FPM parameters), which can be very useful to monitor the process, but would otherwise not be accessible in the absence of a KD block. Therefore, the proposed hybrid approach is multivariate in nature, a feature that is known to offer significant advantages over single-variable methods [3,4,26]. The main features of the hybrid monitoring framework are the following:

- the estimated states (and possibly adapted parameters) provide meaningful indications about the phenomena involved in faults, which can facilitate fault detection and diagnosis.

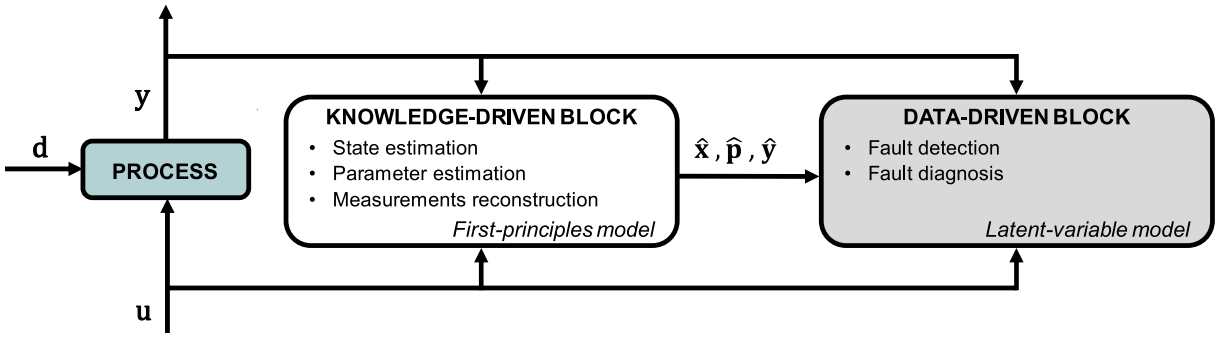

Fig. 1. Proposed hybrid monitoring framework.

Large contributions from an estimated state or an adapted parameter indicate that the fault might be related to a physical phenomenon linked to that state or parameter. For this reason, the parameters to be selected for online adaptation (if any) are those mostly related to specific faults to be monitored. Although parameter estimation for fault detection and diagnosis has already been discussed in the literature [34,35], the distinctive advantage of the proposed framework is that the overall *co-variation* of states, parameters, and measurements is assessed by the hybrid framework, which can improve the monitoring performance;

- a subset of the measured inputs and outputs (**y** and **u**) may be not modeled by the FPM. Yet, due to their inclusion in the **Z** matrix, the LVM can exploit the deterministic information they embed, by assessing how they correlate not only with the other measurements, but also with the estimated variables;
- measurements are considered twice by the LVM: once in terms of **y** and once in terms of $\hat{\mathbf{y}}$. Hence, the filter innovations $\boldsymbol{\gamma}(t)$, which are sometimes monitored for fault detection in KD monitoring approaches [7,11,12], are (indirectly) fed to the monitoring system and are analyzed in a multivariate fashion.

4. Case studies

In this section, the two simulated case studies used to test the proposed methodology are presented. For each case study, two models are employed, namely:

- a detailed model is used to represent the true plant behavior; this model will be referred to as "the process";
- a simplified model is used to design the state estimator; this model will be referred to as "the FPM" or, more simply, "the model".

The main features of the two case studies are summarized in Table 1. Parametric and structural process-model mismatch exist in both case studies, although in Case study 2 the structural mismatch is more significant. In Case study 1, only state estimation and measurement reconstruction are carried out by the EKF in the KD block, whereas in Case study 2 the EKF is also exploited for online parameter adaptation. DPCA is used in the DD block for Case study 1, whereas MW-PCA is used in Case study 2.

4.1. Case study 1: PET manufacturing

PET synthesis occurs through three main steps: transesterification/esterification, pre-polymerization, and polycondensation. In this study, we refer to the polycondensation step. The process is

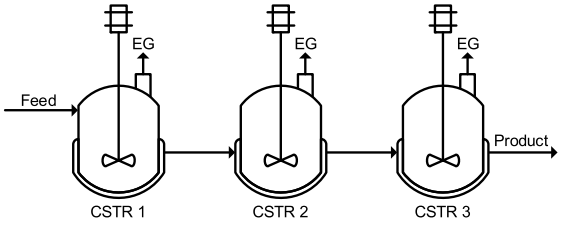

Fig. 2. Case study 1: scheme of the PET polycondensation process with three CSTRs in series.
Source: Adapted from [41].

constituted by a series of three CSTRs (Figure 2) described by the following set of equations [41]:

$$\frac{dc_{\mathrm{EG},i}}{dt} = \frac{1}{\tau}\left(c_{\mathrm{EG},i-1} - c_{\mathrm{EG},i}\right) - h_i\left(c_{\mathrm{EG},i} - c^*_{\mathrm{EG},i}\right) + 0.5R_{\mathrm{pol},i} \tag{23}$$

$$\frac{dc_{\mathrm{OH},i}}{dt} = \frac{1}{\tau}\left(c_{\mathrm{OH},i-1} - c_{\mathrm{OH},i}\right) - R_{\mathrm{pol},i} \tag{24}$$

$$\frac{dc_{\mathrm{COOH},i}}{dt} = \frac{1}{\tau}\left(c_{\mathrm{COOH},i-1} - c_{\mathrm{COOH},i}\right) + R_{\mathrm{degr},i} \tag{25}$$

$$\frac{dc_{\mathrm{ESTER},i}}{dt} = \frac{1}{\tau}\left(c_{\mathrm{ESTER},i-1} - c_{\mathrm{ESTER},i}\right) + 0.5R_{\mathrm{pol},i} - R_{\mathrm{degr},i} \tag{26}$$

$$R_{\mathrm{pol},i} = k_{\mathrm{pol},i}(c^2_{\mathrm{OH},i} - 8c_{\mathrm{EG},i}c_{\mathrm{ESTER},i}) \tag{27}$$

$$R_{\mathrm{degr},i} = k_{\mathrm{degr},i}c_{\mathrm{ESTER},i} \tag{28}$$

$$c^*_{\mathrm{EG},i} = \frac{P_i}{P^{sat}_{\mathrm{EG},i}(T)\hat{v}_{\mathrm{EG}}\exp(1+\chi)} \tag{29}$$

where i identifies a given reactor ($i = 1, 2, 3$).

In the above model equations, EG, OH, COOH and ESTER respectively denote ethylene glycol, hydroxyl end-groups, acid end-groups, and ester groups; $c_{j,i}$ is the concentration of species j in reactor i ($c_{j,0}$ being the inlet concentration of species j to the first reactor); τ is the residence time in a reactor, and is the same for all reactors; $R_{\mathrm{pol},i}$ and $R_{\mathrm{degr},i}$ respectively refer to the rates for the polycondensation and degradation in reactor i. In the Flory–Huggins equation (29), $c^*_{\mathrm{EG},i}$ is the equilibrium concentration of ethylene glycol (the only species existing in the vapor phase) in reactor i, P_i is the pressure in reactor i, $P^{sat}_{\mathrm{EG},i}$ is the EG vapor pressure, $\hat{v}_{\mathrm{EG}}$ is the EG molar volume in the liquid phase, and χ is the polymer-solvent interaction parameter. The

Table 1
Comparison of the main features of the two case studies investigated.

Feature	Case study 1	Case study 2
Process name	PET manufacturing	Penicillin manufacturing
Process type	Continuous	Fed-batch
Process-model mismatch	parametric and (mild) structural	Parametric and structural
Unmodeled measurements	Pressures	Temperature
Knowledge-driven block	State estimation; measurement reconstruction	State estimation; measurement reconstruction; parameter adaptation
Data-driven block	Dynamic PCA	Moving-window PCA

values of the parameters of the Flory–Huggins equation are set as in [41]. The process is assumed to be isothermal, so the energy balance is neglected. The meaning of the rest of the symbols is reported in Table 2, together with the values of all parameters and feed conditions. Their nominal values are taken from [41], but fluctuations are added as smoothed pseudo-random binary signals in order to more closely mimic a real situation where process noise increases normal process variability.

The key performance indicator is the degree of polymerization in the third reactor (DP_3), which cannot be measured online and is calculated as:

$$DP_3 = 1 + \frac{2c_{\mathrm{ESTER},3}}{c_{\mathrm{OH},3} + c_{\mathrm{COOH},3} + c_{\mathrm{E_g},3}} \tag{30}$$

where $c_{\mathrm{E_g},3}$ is the concentration of a byproduct, which is assumed to be equal to $c_{\mathrm{COOH},3}$ [41].

The 9 measurements available online from the process are:

- $y_{c_{\mathrm{OH},i}}$[mol L^{-1}], concentration of hydroxyl end-groups in reactor i;
- $y_{c_{\mathrm{COOH},i}}$[mol L^{-1}], concentration of acid end-groups in reactor i;
- y_{P_i} [Pa], pressure in reactor i.

White noise with standard deviation of typical industrial sensors is added to the detailed model outputs; namely, the standards deviations are 7E-3 mol L^{-1} for $y_{c_{\mathrm{OH},i}}$, 3E-4 mol L^{-1} for $y_{c_{\mathrm{COOH},i}}$, and 1 Pa for y_{P_i} (notice that the types of sensors for OH and COOH concentration measurements are different [41]). The measurement intervals for $y_{c_{\mathrm{OH},i}}$ and $y_{c_{\mathrm{COOH},i}}$ are set to 1 min and 10 min, respectively [41]. Measurements for pressure are recorded every 10 min, because they are not needed at greater frequency for process monitoring.

The NOC dataset includes data from 5300 min of steady-state operation. We generate faulty datasets by running the process as under NOC, but applying the fault after 300 min from the start of the NOC sequence. Simulations are run for 1300 min. We consider four faulty sequences, all of which eventually result in an impactful decrease of DP_3:

- Fault #1: slowly decreasing ester concentration in the feed ($c_{\mathrm{ESTER},0}$ decreases by 0.01% per min). As a consequence, the concentration of ester in the first reactor starts decreasing, eventually reducing DP_3 (Eq. (30)).
- Fault #2: minor fault in the agitation system of the second reactor (h_2 decreases by 0.05% per min). The smaller mixer speed hinders the mass transfer of ethylene glycol from the liquid to the vapor phase. Ethylene glycol starts accumulating in the second reactor, thus reducing the rate of polycondensation Eq. (27), and eventually affecting DP_3 (Eq. (30)).
- Fault #3: significant fault in the agitation system of the second reactor (h_2 decreases by 0.1% per min). The consequences are as for Fault #2, but with greater magnitude.
- Fault #4: increasing pressure in the second reactor (P_2 increases by 3% per min). The pressure increase inhibits ethylene glycol mass transfer (Eqs. (23) and (29)), with effects on DP_3 similar to Faults #2 and #3.

To assess reproducibility of the results for different patterns of measurement and process noise, we consider 10 different realizations of each fault scenario.

The simplified FPM model employs Eqs. (23)–(28), but with constant parameters and feed conditions. In addition, the effect of pressure on $c^*_{\mathrm{EG},i}$ in each reactor i (Eq. (29)) is neglected and $c^*_{\mathrm{EG},i}$ is assumed constant. As a result, the FPM presents parametric and (mild) structural mismatch.

4.2. Case study 2: penicillin manufacturing

The manufacturing of penicillin by biomass fermentation is modeled by Birol et al. [42]. The process is carried out in a reactor operating batchwise for the first 50 h (growth phase). During this period, the concentration of biomass grows, and no penicillin is produced. Then, the substrate feed is turned on, and in this fed-batch phase the biomass concentration grows slowly and the penicillin concentration increases.

For ease of reading, the set of equations defining the detailed model [42] is reported in Appendix A, together with the values of all inputs, parameters, and process noise characteristics. The simplified FPM model includes significant parametric and structural mismatch, and is also reported in Appendix A.

The 6 measurements available online from the process are pH (y_{pH}), temperature (y_T), oxygen concentration ($y_{\mathrm{O_2}}$), volume (y_V), CO_2 concentration ($y_{\mathrm{CO_2}}$), and feed flow rate (y_F). A measurement interval of 3 min is considered for process monitoring. Note that y_T and y_{pH} are not accounted for by the FPM, y_F is the only input of the FPM, and the other available measurements correspond to states of the FPM. White noise is added to all measurements, consistently with the typical precision of industrial instrumentation. The standard deviations of the noise signals are 0.05 for y_{pH}, 0.05 K for y_T, 0.0025 $\mathrm{g_{O_2}}$ L^{-1} for $y_{\mathrm{O_2}}$, 0.01 L for y_V, 0.06 $\mathrm{mol_{CO_2}}$ L^{-1} for $y_{\mathrm{CO_2}}$, and 0.0002 L h^{-1} for y_F.

The NOC dataset includes 35 batches, each one lasting 300 h. Four faulty batches are considered, with the same length as the normal ones. The faulty batch characteristics are as follows:

- Fault #1: slow ramp decrease in the aeration rate (unmeasured variable), which causes a drop in the oxygen mass transfer coefficient and in the oxygen concentration in the reactor;
- Fault #2: slow ramp decrease of the substrate concentration in the feed (unmeasured variable), which inhibits the biomass growth;
- Fault #3: slow ramp decrease of the maximum growth rate kinetic parameter, which reduces the biomass concentration in the reactor;

Table 2
Case study 1: parameters and feed conditions in the detailed and simplified models. In the detailed model, fluctuations are added to the nominal values as smoothed-pseudo random binary signals with the indicated maximum/minimum amplitudes. In the simplified model, constant (nominal) values are used.

Parameter or feed condition	Symbol	Units	Nominal value	Max/min amplitude
Residence time (all reactors)	τ	min	60	0
Mass transfer coefficient reactor 1	h_1	min^{-1}	2.70	0.80
Mass transfer coefficient reactor 2	h_2	min^{-1}	2.03	0.61
Mass transfer coefficient reactor 3	h_3	min^{-1}	1.35	0.41
Kinetic constant, reaction 1	$k_{pol,i}$	L mol^{-1} min^{-1}	6.66E−02	0.017
Kinetic constant, reaction 2	$k_{degr,i}$	min^{-1}	8.34E−06	1.25E−06
Pressure in reactor i	P_i	Pa	130	6.5
Feed concentration of ethylene glycol	$c_{EG,0}$	mol L^{-1}	6.50E−03	9.75E−04
Feed concentration of OH end-groups	$c_{OH,0}$	mol L^{-1}	0.40	0.080
Feed concentration of COOH end-groups	$c_{COOH,0}$	mol L^{-1}	2.57E−03	2.57E−04
Feed concentration of ester groups	$c_{ESTER,0}$	mol L^{-1}	11.20	0.022

- Fault #4: high cooling water temperature (unmeasured variable), which causes the reactor temperature to rise.

Numerical details on the faulty sequences are reported in Appendix A. We implement ten different realizations of each fault scenario.

5. Results and discussion for case study 1

5.1. Design of the hybrid monitoring model

The 21 variables selected for inclusion in the augmented data matrix of the hybrid monitoring model are listed in Table 3. They comprise the 9 available field measurements and the 12 states estimated by the EKF using the FPM and the available measurements. No model parameters are included in the augmented matrix.

The difference in the concentration measurement intervals (1 min vs. 10 min, see Section 4.1) is dealt with using a two time-scale EKF (Figure 3), which we borrow from [43]. EKF-1 receives the frequent measurements and performs the prediction and correction steps for all states except $c_{COOH,i}$, which is predicted at open loop due to the observability conditions. EKF-2 receives the infrequent measurements and provides corrections for $c_{COOH,i}$. The accuracy of estimation of the unmeasured states $c_{EG,i}$ and $c_{ESTER,i}$ is improved through bias estimation [33]. Following the approach adopted in [43], to satisfy the observability conditions three biases are updated in EKF-1 ($b_{EG,1}$, $b_{EG,2}$ and $b_{EG,3}$), whereas the other three biases are updated in EKF-2 ($b_{ESTER,1}$, $b_{ESTER,2}$ and $b_{ESTER,3}$).

We design the initial state covariance matrices ($\mathbf{P}_{0,EKF-1}$ and $\mathbf{P}_{0,EKF-2}$) as diagonal matrices with zero variance for the biases, and the same variance (equal to 1E-6) for all the other states. We tune $\mathbf{Q}_{EKF-1}$ and $\mathbf{Q}_{EKF-2}$ heuristically to achieve quick and robust convergence, resulting in:

$$\mathbf{Q}_{EKF-1} = \text{diag}(4E-6 \;\; 0 \;\; 0 \;\; 0 \;\; 1E-6 \;\; 1E-8 \;\; 0 \;\; 0 \;\; 1E-6 \;\; 2.5E-9 \;\; 0 \;\; 0 \;\; 4E-10 \;\; 0 \;\; 4E-10 \;\; 0 \;\; 4E-10 \;\; 0) \quad (31)$$

$$\mathbf{Q}_{EKF-2} = \text{diag}(0 \;\; 0 \;\; 2.5E-9 \;\; 0 \;\; 0 \;\; 0 \;\; 2.5E-9 \;\; 0 \;\; 0 \;\; 0 \;\; 2.5E-9 \;\; 0 \;\; 0 \;\; 3.6E-7 \;\; 0 \;\; 3.6E-7 \;\; 0 \;\; 3.6E-7) \quad (32)$$

In Eqs. (31) and (32), the first twelve elements along the diagonal correspond to the estimated states (variable nos. 10–21 in Table 3), whereas the last six refer to the bias vector [$b_{EG,1}$ $b_{ESTER,1}$ $b_{EG,2}$ $b_{ESTER,2}$ $b_{EG,3}$ $b_{ESTER,3}$]. Finally, $\mathbf{R}_{EKF-1}$ and $\mathbf{R}_{EKF-2}$ are built according to Eq. (12), with $\sigma_{y_{OH,i}}$ = 4.9E-5 and $\sigma_{y_{COOH,i}}$ = 9E-8.

The typical estimation performance during a transient is shown in Figure 4 for one unmeasured state in the second reactor. The estimation accuracy is satisfactory.

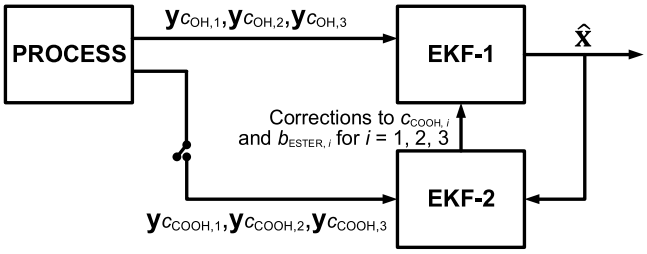

Fig. 3. Architecture of the knowledge-driven block (two time-scale extended Kalman filter) implemented for Case Study 1.
Source: Adapted from [43].

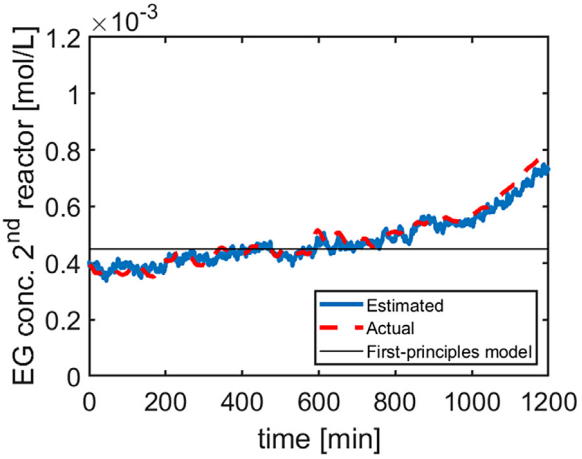

Fig. 4. Case study 1: EKF estimation performance for one unmeasurable state in the second reactor during a transient.

We design a DPCA model on the 21 variables listed in Table 3 under NOC, sampling all signals every 10 min. The number L of lagged measurements to be included in $\mathbf{Z}_{dyn}$ (Eq. (20)) is heuristically derived from the overall residence time in the three reactors (3 × 60 = 180 min), resulting in L = 18. The resulting size of $\mathbf{Z}_{dyn}$ is [512×399]. Eight PCs are used (this number being found by cross-validation), explaining 54% of data variability on T^2, with the remaining 46% being explained by SPE. The resulting control charts under NOC are shown in Figure 5: no false alarms are issued.

5.2. Fault detection and diagnosis

We compare the monitoring performances of the hybrid model to those of a standard DD monitoring model and of a standard KD monitoring approach. The standalone DD model uses only the 9

Table 3
Case study 1: list of variables included in the augmented data matrix of the hybrid monitoring model.

Var. no.	Variable name	Reactor no.	Symbol	Units	Variable type
1	Pressure	1	y_{P_1}	Pa	Measurement
2	Hydroxyl end-groups concentration	1	$y_{c_{\mathrm{OH},1}}$	mol L^{-1}	Measurement
3	Acid end-group concentration	1	$y_{c_{\mathrm{COOH},1}}$	mol L^{-1}	Measurement
4	Pressure	2	y_{P_2}	Pa	Measurement
5	Hydroxyl end-group concentration	2	$y_{c_{\mathrm{OH},2}}$	mol L^{-1}	Measurement
6	Acid end-group concentration	2	$y_{c_{\mathrm{COOH},2}}$	mol L^{-1}	Measurement
7	Pressure	3	y_{P_3}	Pa	Measurement
8	Hydroxyl end-group concentration	3	$y_{c_{\mathrm{OH},3}}$	mol L^{-1}	Measurement
9	Acid end-group concentration	3	$y_{c_{\mathrm{COOH},3}}$	mol L^{-1}	Measurement
10	Ethylene glycol concentration	1	$c_{\mathrm{EG},1}$	mol L^{-1}	Estimated state
11	Hydroxyl end-group concentration	1	$c_{\mathrm{OH},1}$	mol L^{-1}	Estimated state/Reconstructed measurement
12	Acid end-group concentration	1	$c_{\mathrm{COOH},1}$	mol L^{-1}	Estimated state/Reconstructed measurement
13	Ester end-group concentration	1	$c_{\mathrm{ESTER},1}$	mol L^{-1}	Estimated state
14	Ethylene glycol concentration	2	$c_{\mathrm{EG},2}$	mol L^{-1}	Estimated state
15	Hydroxyl end-group concentration	2	$c_{\mathrm{OH},2}$	mol L^{-1}	Estimated state/Reconstructed measurement
16	Acid end-group concentration	2	$c_{\mathrm{COOH},2}$	mol L^{-1}	Estimated state/Reconstructed measurement
17	Ester end-group concentration	2	$c_{\mathrm{ESTER},2}$	mol L^{-1}	Estimated state
18	Ethylene glycol concentration	3	$c_{\mathrm{EG},3}$	mol L^{-1}	Estimated state
19	Hydroxyl end-group concentration	3	$c_{\mathrm{OH},3}$	mol L^{-1}	Estimated state/Reconstructed measurement
20	Acid end-group concentration	3	$c_{\mathrm{COOH},3}$	mol L^{-1}	Estimated state/Reconstructed measurement
21	Ester end-group concentration	3	$c_{\mathrm{ESTER},3}$	mol L^{-1}	Estimated state

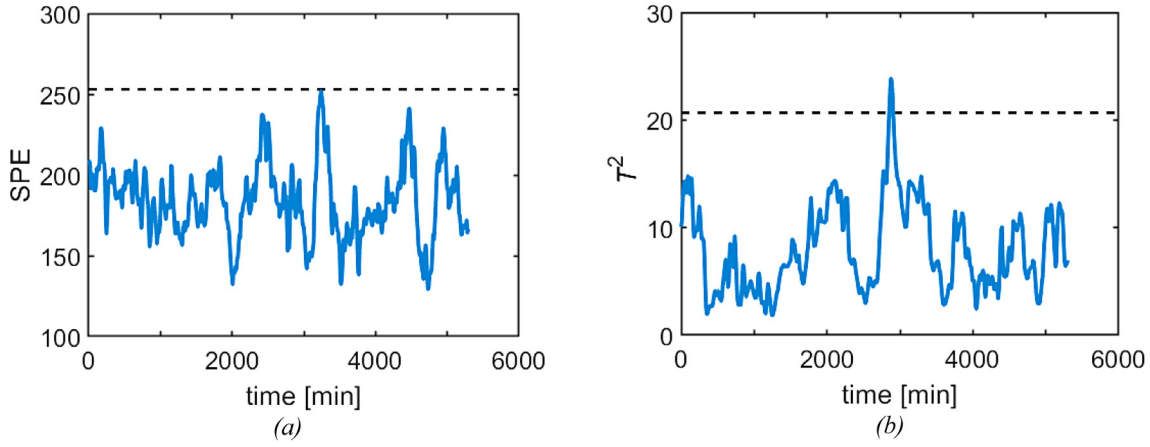

Fig. 5. Case study 1: control charts under normal operating conditions: (a) SPE chart, (b) Hotelling T^2 chart, (c) contributions to SPE. The dashed lines represent the 99% confidence limits.

measurements available from the process and is based on a DPCA model designed with the same characteristics of the hybrid model (18 lagged measurements spaced by 10 min). Cross-validation suggests using 8 PCs for this model, too. Monitoring through a standalone KD model is carried out using an EKF and χ^2 tests on $\epsilon\,(t_k)$ and $\epsilon\,(t_k, S)$ [12]. The confirmation test is implemented with the same moving window of the hybrid and DD models; no state augmentation is used.

For each fault scenario, the fault detection time is reported in Table 4 as an average across the relevant fault realizations. Whereas the detection performances for Fault #4 result the same for the hybrid and the DD models, the hybrid model detects Fault #1 and Fault #3 much more promptly than the DD one (170 and 320 min earlier, respectively). Additionally, Fault #2 (a subtle one) goes undetected by the DD model, whereas it is correctly detected by the hybrid model. The KD method leads to the worst detection performance, as it is severely compromised by the process-model mismatch. Fault #2 is not detected, and the average detection times for the other fault scenarios are greater than those of the hybrid and DD models (changing the confirmation test window size does not lead to any substantial improvement). For this reason, the KD model will not be investigated further for this case study. Incidentally, if the process is monitored by univariate control charts on measurements or estimated states, unsatisfactory detection performances for all fault scenarios are obtained (results are expected and are not reported for conciseness).

The reason why the hybrid model performs better than the DD model in fault detection can be explained as follows. Thanks to the presence of the KD block (Figure 1), the hybrid model embeds more information about the inner working of the process, namely on how the measured variables *and* the states are expected to co-vary under NOC. Hence, mutual deviations of measured variables and states from the relevant reference trajectories (such as those occurring after the onset of a fault) can be detected effectively. For example, Fault #3 starts impacting on the (estimated) states earlier than it does on the measurements, as can be seen from Figure 6 for the representative profiles of the ethylene glycol

Table 4
Case study 1: fault detection time using the hybrid monitoring model, the data-driven monitoring model and the knowledge-driven monitoring model. The detection time runs from the fault onset, and the reported values are averaged across the 10 realizations of the relevant fault. For the hybrid and data-driven models, the monitoring statistic alarming the fault is also indicated..

Fault number and type	Primarily affected variable or parameter	Hybrid model detection time (min)	Data-driven model detection time (min)	Knowledge-driven model detection time (min)
#1: Ester feed concentration decrease	$c_{ESTER,0}$	550 (SPE)	720 (T^2)	880
#2: Minor agitation fault	h_2	590 (SPE)	(undetected)	(undetected)
#3: Major agitation fault	h_2	380 (SPE)	700 (SPE)	930
#4: Pressure increase	P_2	40 (SPE)	40 (SPE)	890

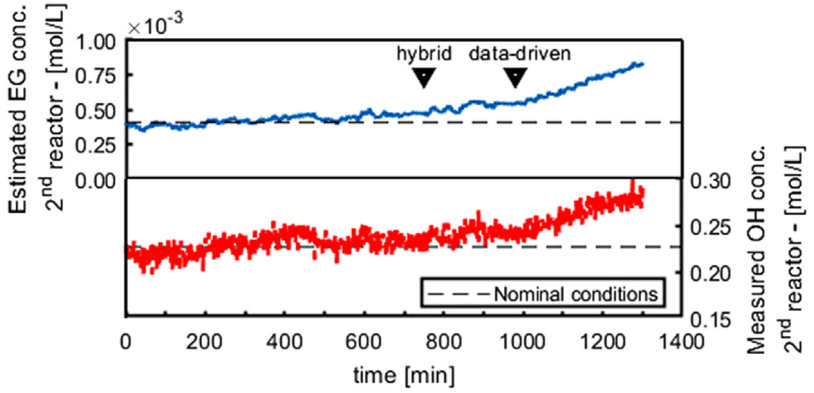

Fig. 6. Case study 1: time profiles of the ethylene glycol concentration (estimated state) and of the ethylene glycol concentration (measured variable) before and after the onset of Fault #3 (time = 300 min). The triangles indicate the detection instant of the hybrid model and of the data-driven model.

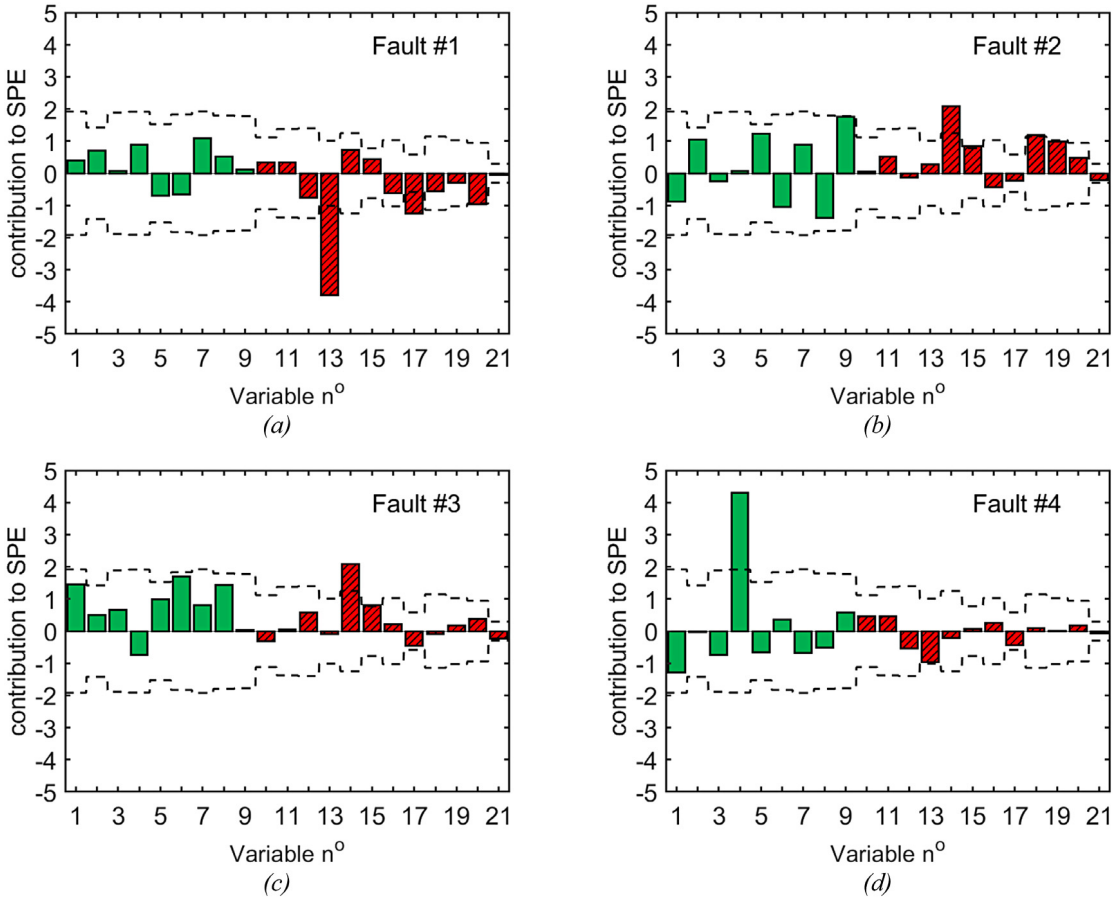

Fig. 7. Case study 1: representative contribution plots for the hybrid monitoring model at the first out-of-control observation for (a) Fault #1, (b) Fault #2, (c) Fault #3, and (d) Fault #4. Variables are numbered as in Table 3. In all plots, the contributions of field measurements are in green, while those of estimated/reconstructed variables are in red with diagonal lines. Control limits at 99% confidence are shown as dashed lines.

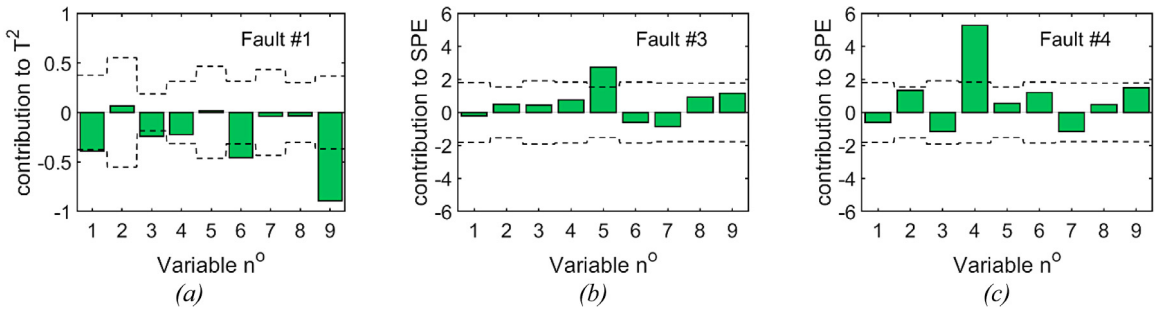

Fig. 8. Case study 1: representative contribution plots for the purely data-driven model at the first out-of-control observation for (a) Fault #1, (b)Fault #3, and (c) Fault #4. Variables are numbered as is in Table 3. Control limits at 99% confidence are shown as dashed lines.

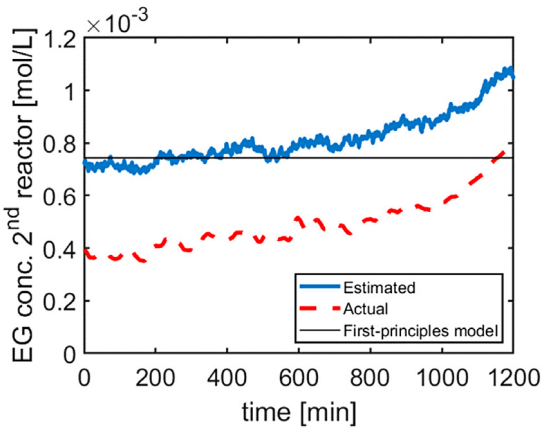

Fig. 9. Case study 1: EKF estimation performance on the unmeasurable state and under the same dynamic conditions of Figure 4, in the presence of significant process-model mismatch and bad filter tuning.

concentration (an estimated state in the second reactor) and of the hydroxyl end-group concentration (a measured output in the same reactor). This piece of information is captured by the DD block, and this allows anticipating the fault detection. On the other hand, if the fault impacts on a measured variable directly, rather than through (or after) the change in one or more states (as occurs for reactor 2 pressure in Fault #4), it is unlikely that the information provided by the KD block to the hybrid model can lead to a significant improvement in the detection performance.

Contribution plots of the hybrid and of the DD models at the first out-of-control signal are used for fault diagnosis. For Fault #1, Figure 7a shows that the hybrid monitoring model points to the concentration of ester in the first reactor (variable no. 13 in Table 3, an estimated state) as most related to the deviation from the NOC, which would straightforwardly suggest an abnormal feed concentration change as a possible root-cause of the fault. On the other hand, the information provided by the DD monitoring model for fault diagnosis is more ambiguous: Figure 8a misleadingly draws the attention to the concentration of acid end-groups in the third reactor (variable no. 9), a measurement that is not related to the root-cause of the fault directly. In ∼50% of the realizations of this fault, the DD model provides T^2 contributions outside the control limits for acid end-groups also in the first and in the second reactors. Indeed, due to the reduced ester concentration in the reactors caused by the fault, also the acid end-group concentrations are expected to decrease because of the ester degradation kinetics Eq. (28), but this is only a secondary (slower) effect.

With respect to Faults #2 and #3, the hybrid model (Figure 7b and 7c) clearly identifies an abnormal ethylene glycol concentration in the second reactor (variable no. 14, an estimated state) as the variable most directly related to the fault, thus correctly pointing the attention to a possibly abnormal mass transfer of EG in that reactor. On the other hand, as already mentioned, Fault #2 is not detected by the DD monitoring model because it is too small in magnitude, whereas for Fault #3 the DD model points to measured variable no. 5, i.e. to the hydroxyl end-group concentration in reactor 2 (Figure 8b). In fact, accumulation of ethylene glycol due to the fault hinders the polycondensation reaction (Eq. (27)), leading to an increase in $y_{c_{OH,2}}$. However, attributing a reduction in the polycondensation rate of reaction to a high ethylene glycol concentration may not be straightforward. Finally, Fault #4 can be diagnosed very easily both by the hybrid model (Figure 7d) and by the DD model (Figure 8c), because this fault impacts on reactor 2 pressure (measured variable no. 4) directly.

Note that very accurate state estimation is not required for the hybrid monitoring system to perform well. In fact, the main task of the KD block in Figure 1 is only to provide information on how the estimated states (together with reconstructed measurements and possibly adapted parameters) co-vary during the process operation, regardless of the fact that the actual values of the states may be somewhat different from the actual (and unknown) ones. To clarify this point, we consider a stronger process/model mismatch by altering the FPM dynamics through summation of a constant term (equal to −2E-3) to the right-hand side of Eq. (24), even though the process dynamics remains the same. Additionally, we degrade the filter performance by tuning $\mathbf{Q}_{EKF-2}$ as:

$$\mathbf{Q}_{EKF-2} = \text{diag}(0\ \ 0\ \ 2.5E-9\ \ 0\ \ 0\ \ 0\ \ 2.5E-9\ \ 0\ \ 0\ \ 0\ \ 2.5E-9\ \ 0\ \ 0\ \ 4E-8\ \ 0\ \ 4E-8\ \ 0\ \ 4E-8). \quad (33)$$

This results in unsatisfactory state estimation, as shown in Figure 9 for the same state and transient considered in Figure 4.

Nevertheless, the fault detection and fault diagnosis performance of the hybrid monitoring model are almost the same as those reported in Table 4 and shown in Figure 7 (results are not reported for conciseness).

6. Results and discussion for case study 2

6.1. Design of the hybrid monitoring model

The augmented data matrix of the hybrid monitoring model includes the 15 variables listed in Table 5. To improve the monitoring performance, in addition to state estimation the EKF performs online adaptation of three FPM parameters: $\mu_{X,\max}$ (maximum growth parameter), K_{la} (mass transfer coefficient) and s_F (substrate concentration in the feed). We select these parameters because they can provide useful insights for typical potential faults that may affect the reactor, namely changes in kinetics, mass transfer, or feed composition.

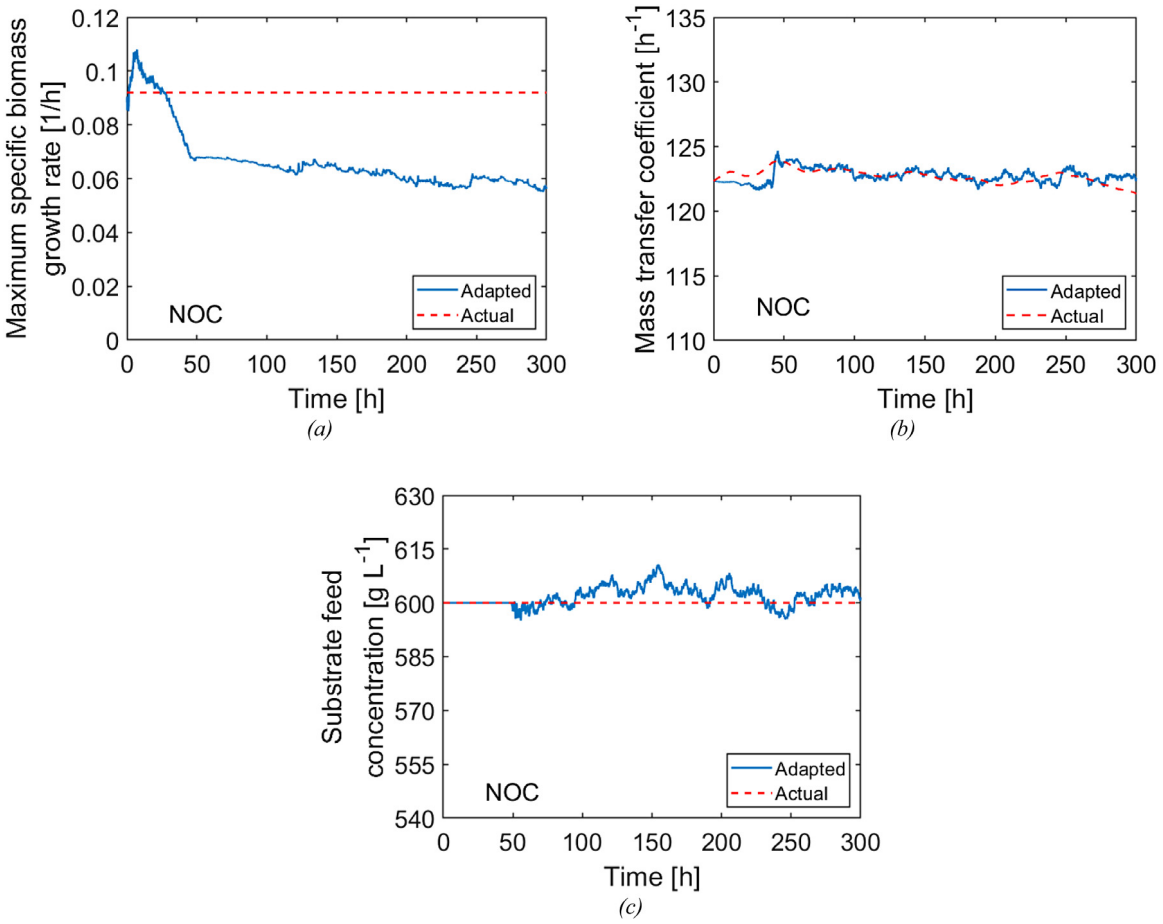

Fig. 10. Case study 2: EKF parameter adaptation performance during a representative NOC batch of (a) the maximum specific biomass growth rate $\mu_{X,\max}$, *(b) the mass transfer coefficient* K_{la}, and (c) the substrate feed concentration s_F.

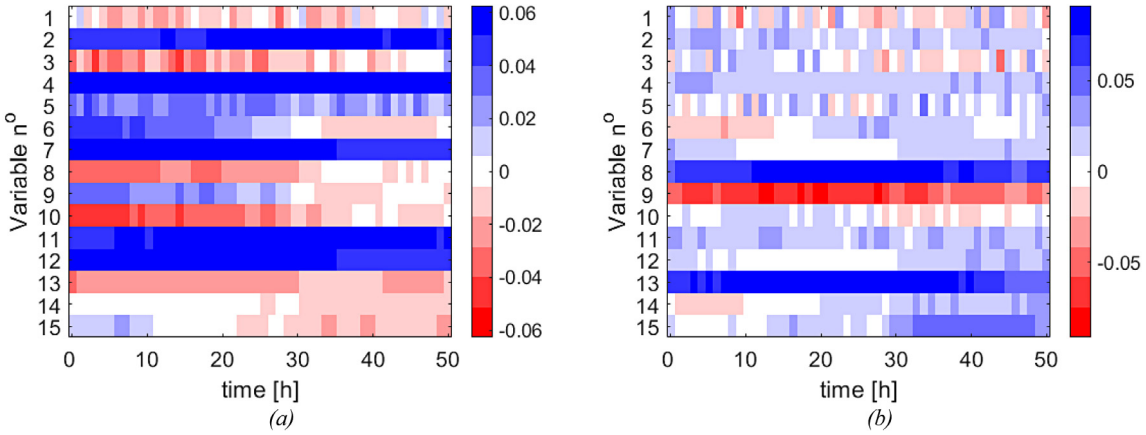

Fig. 11. Case study 2: loadings for (a) the first PC and (b) the second PC of the MW-PCA model over one window width during the production phase. The variables are numbered as in Table 5.

The modeled measurements (y_{O_2}, y_V, y_{CO_2}) and measured input (y_F) are supplied to the EKF every 3 min. The $\mathbf{P}_0$, $\mathbf{Q}$ and $\mathbf{R}$ matrices used are designed using the same criteria as in Case study #1, resulting in (states and measurements are ordered as is in Table 5):

$$\mathbf{P}_0 = \text{diag}(1.00\text{E}-4 \;\; 1.00\text{E}-4 \;\; 1.00\text{E}-4 \;\; 1.00\text{E}-4 \;\; 1.00\text{E}-4 \;\; 1.00\text{E}-4 \;\; 2.25\text{E}-4 \;\; 1.00 \;\; 0.00) \tag{34}$$

$$\mathbf{Q} = \text{diag}(2.50\text{E}-08 \;\; 1.00\text{E}-09 \;\; 9.00\text{E}-07 \;\; 2.50\text{E}-08 \;\; 9.00\text{E}-05 \;\; 2.50\text{E}-08 \;\; 6.50\text{E}-08 \;\; 2.50\text{E}-3 \;\; 8.00\text{E}-2) \tag{35}$$

$$\mathbf{R} = \text{diag}(6.25\text{E}-06 \;\; 1.00\text{E}-04 \;\; 3.60\text{E}-03) \tag{36}$$

The resulting state estimation performance is satisfactory (results are reported in Appendix B, Figure B.1). Nevertheless, as illustrated in Figure 10, the adapted parameters drift away from their "true" values even under NOC, because the EKF adjusts the model parameters in the attempt to compensate for the detected process-model mismatch. This does not represent an

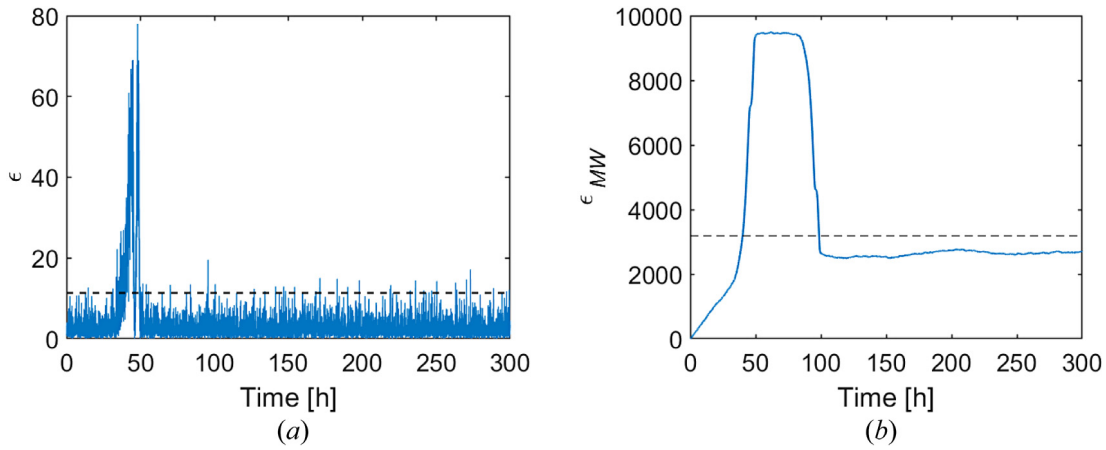

Fig. 12. Case study 2: control charts under normal operating conditions for the knowledge-driven approach. (a) χ^2 test on the single-point innovation (fault detection), (b) χ^2 test on the innovations on a 50 h moving window (fault confirmation). The dashed lines represent the 99% confidence limits.

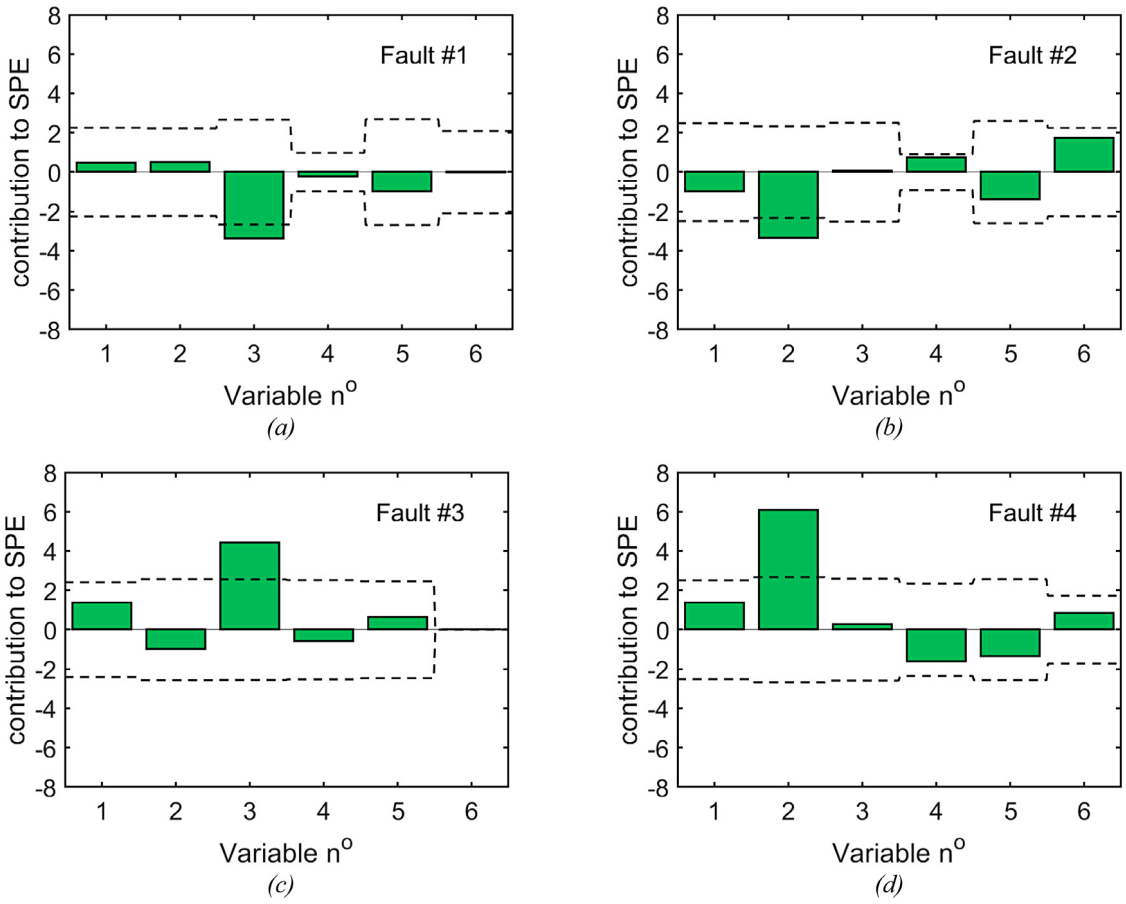

Fig. 13. Case study 2: representative contribution plots for the purely data-driven model at the first out-of-control observation for (a) Fault #1, (b) Fault #2, (c) Fault #3, and (d) Fault #4. Variables are numbered as is in Table 5. Control limits at 99% confidence are shown as dashed lines.

issue for the hybrid system, because it is the pattern of change of the adjusted parameter profiles together with the profiles of all other measured and estimated variables that matters for process monitoring, regardless of the fact that each single parameter is estimated accurately or not.

Incidentally, note that, in the presence of significant structural mismatch, a comparison between the "process" and the "model" parameters might even not be entirely appropriate, because the meaning of a parameter within the "process" may be different from the one the same parameter has in the "model". This is especially true when a model parameter (e.g., the maximum growth rate kinetic parameter; see Appendix A) is used to compactly represent a set of physical mechanisms that are expected to occur in the process, but are not described by the model equations accurately.

We design an MW-PCA model for the DD block of the hybrid monitoring system. Considering that the total batch duration is 300 h, measurements are retained every 1 h. After exploring several windows widths, we select $W = 50$ h as the width

Table 5
Case study 2: list of variables included in the augmented data matrix of the hybrid monitoring model.

#	Symbol	Variable	Unit	Type of variable
1	y_{pH}	pH	–	Measurement
2	y_T	temperature	K	Measurement
3	y_{O_2}	oxygen concentration	$g_{O_2}\ L^{-1}$	Measurement
4	y_V	volume	L	Measurement
5	y_{CO_2}	CO_2 concentration	$mol_{CO_2} L^{-1}$	Measurement
6	y_F	feed flow rate	$L\ h^{-1}$	Measurement
7	X	biomass concentration	$g_X\ L^{-1}$	Estimated state
8	P	penicillin concentration	$g_P\ L^{-1}$	Estimated state
9	S	substrate concentration	$g_S\ L^{-1}$	Estimated state
10	c_{O_2}	oxygen concentration	$g_{O_2}\ L^{-1}$	Estimated state/Reconstructed measurement
11	V	volume	L	Estimated state/Reconstructed measurement
12	c_{CO_2}	CO_2 concentration	$mol_{CO_2} L^{-1}$	Estimated state/Reconstructed measurement
13	$\mu_{X,max}$	maximum growth parameter	h^{-1}	Adapted parameter
14	K_{la}	mass transfer coefficient	h^{-1}	Adapted parameter
15	s_F	feed concentration	$g_S\ L^{-1}$	Adapted parameter

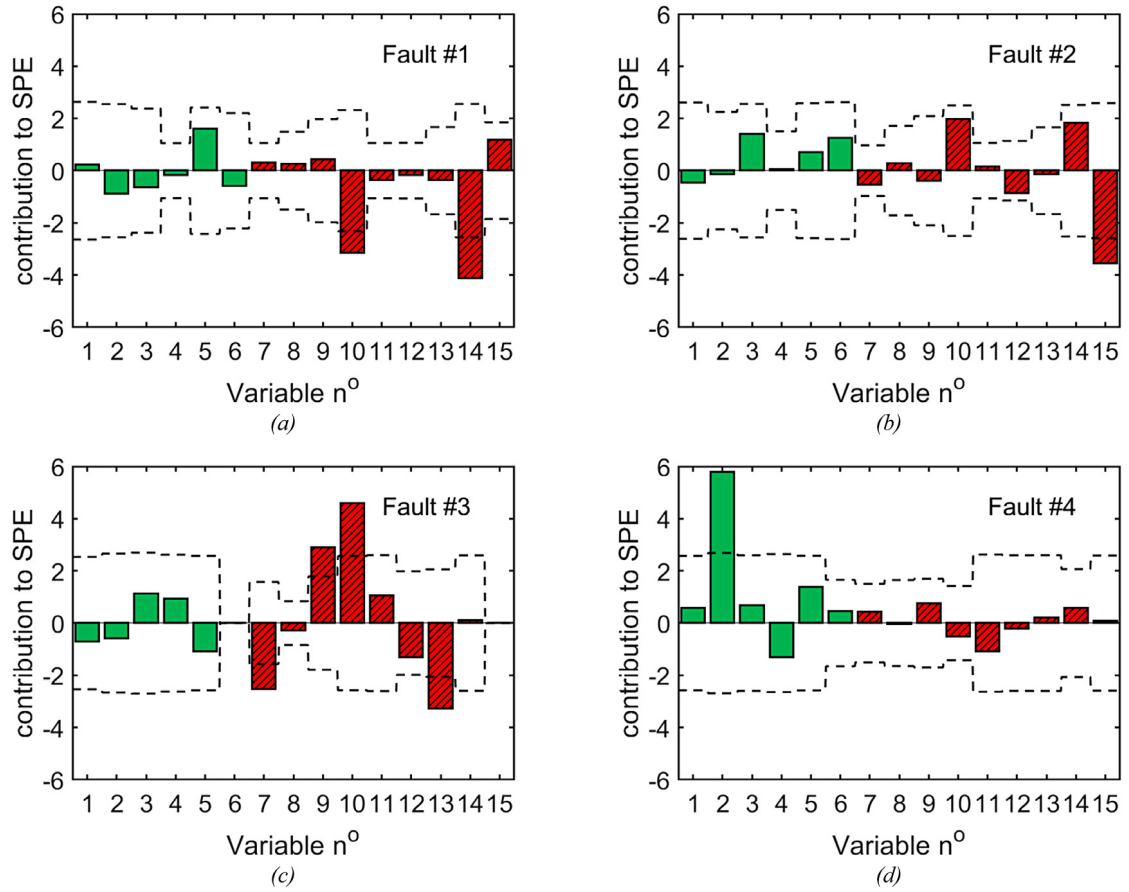

Fig. 14. Case study 1: representative contribution plots for the hybrid monitoring model at the first out-of-control observation for (a) Fault #1, (b) Fault #2, (c) Fault #3, and (d) Fault #4. Variables are numbered as in Table 5. In all plots, the contributions of field measurements are in green, while those of estimated/reconstructed variables are in red with diagonal lines. Control limits at 99% confidence are shown as dashed lines.

leading to satisfactory monitoring performances. The size of $\mathbf{Z}_{MW}$ is [35×15] after 1 h, and grows of 15 columns per hour until 50 h. Then, a fully developed $\mathbf{Z}_{MW}$ of size [35×750] is used until the end of the batch. Three PCs are used (found by cross-validation), with an explained variance profile of ∼45% in the production phase, after reaching a minimum of ∼25% at the switch between the growth and the production phases, due to the high variability of the switching instant across the NOC batches.

The loadings of the first two PCs over one window width during the production phase (Figure 11) allow assessing the auto- and cross-correlation of the variables included in the augmented data matrix. MW-PCA provides a correlation model in which estimated states, adapted parameters and measurements (including those not accounted for by the FPM) are all intimately linked. For example, from the analysis of the loadings on the first PC (Figure 11a), it emerges that the reactor temperature (variable no.2, which is measured but not modeled by the FPM) is not only strongly auto-correlated, but also cross-correlated to several variables modeled by the FPM, such as for example the reactor volume (variable no.11). As will be shown in the next section, exploitation of the cross-correlation between states, measurements and parameters enhances the monitoring capability of the hybrid model.

Table 6
Case study 2: fault detection time using the hybrid monitoring model, the data- driven monitoring model and the knowledge-driven monitoring model. The detection time runs from the fault onset, and the reported values are averaged across the 10 realizations of the relevant fault. For the hybrid and data-driven models, the monitoring statistic alarming the fault is also indicated.

Fault number type	Primarily affected variable or parameter	Hybrid model detection time (h)	Data-driven model detection time (h)	Knowledge-driven model detection time (h)
#1: aeration rate decrease	K_{la}	45 (SPE)	110 (SPE)	70
#2: substrate feed concentration decrease	s_F	35 (SPE)	70 (SPE)	62
#3: growth rate decrease	$\mu_{X,\max}$	14 (SPE)	28 (SPE)	(undetected)
#4: cooling water temperature rise	T	35 (SPE)	35 (SPE)	(undetected)

Table A.1
Case study 2: states of the detailed model, with indication on whether they are measured, and whether or not they are included also in the simplified model.

State variable	Symbol	Initial value	Units	Measured?	In the FPM?
Biomass concentration	X	0.1	$g_X\ L^{-1}$	No	Yes
Penicillin concentration	P	0	$g_P\ L^{-1}$	No	Yes
Substrate concentration	S	15	$g_S\ L^{-1}$	No	Yes
Dissolved oxygen concentration	c_{O_2}	1.16	$g_{O_2}\ L^{-1}$	Yes	Yes
Volume	V	100	L	Yes	Yes
CO_2concentration	c_{CO_2}	0.5	$mmol_{CO_2}\ L^{-1}$	Yes	Yes
Hydrogen ion concentration	$[H^+]$	$10^{-5.1}$	mol L^{-1}	Yes	No
Temperature	T	297	K	Yes	No
Heat released	Q_r	0	cal	No	No

Table A.2
Case study 2: parameters the detailed model, with indication on whether or not they are included also in the simplified model.

Parameter	Symbol	Units	Value	In the FPM?
Penicillin hydrolysis rate constant	K	h^{-1}	0.04	Yes
Yield of biomass on substrate	$Y_{X/S}$	$g_X g_S^{-1}$	0.45	Yes
Yield of penicillin on substrate	$Y_{P/S}$	$g_P g_S^{-1}$	0.90	Yes
Maintenance coefficient on substrate	m_x	h^{-1}	0.014	Yes
Feed substrate concentration	s_F	$g_S\ L^{-1}$	600	Yes
Yield of biomass on oxygen	$Y_{X/O}$	$g_X\ g_{O_2}^{-1}$	0.04	Yes
Yield of product on oxygen	$Y_{P/O}$	$g_P\ g_{O_2}^{-1}$	0.20	Yes
Maintenance requirement of oxygen	m_o	h^{-1}	0.467	Yes
Solubility of oxygen in broth	$c^*_{O_2}$	$g_{O_2}L^{-1}$	1.16	Yes
Volume loss parameter	λ	h^{-1}	2.5E−4	Yes
Constant relating CO_2 to growth	α	$mmol_{CO_2}\ g_X^{-1}\ h^{-1}$	0.143	Yes
Constant relating CO_2 to maintenance energy	β	$mmol_{CO_2}\ g_X^{-1}\ h^{-1}$	4E7	Yes
Constant relating CO_2 to penicillin production	γ	$mmol_{CO_2}\ L^{-1}\ h^{-1}$	10E−4	Yes
Maximum specific biomass growth rate	$\mu_{X,\max}$	h^{-1}	0.092	Yes
Contois saturation constant	k_x	$g_X L^{-1}$	0.15	Yes
Oxygen limitations constants (no limitation)	K_{ox}, K_{op}	g_X^{-1} L	0	Yes
Oxygen limitations constants (with limitation)	K_{ox}, K_{op}	g_X^{-1} L	2E−2, 5E−4	Yes
Maximum specific rate of product formation	$\mu_{P,\max}$	h^{-1}	0.005	Yes
Inhibition constant	k_p	$g_S L^{-1}$	0.0002	Yes
Inhibition constant for product formation	K_I	$g_S L^{-1}$	0.10	Yes
Constant	P	–	3	Yes
Feed temperature of substrate	T_f	K	298	No
Constant for μ	K_1	[mol/L]	1E−10	No
Constant for μ #2	K_2	[mol/L]	7E−5	No
Arrhenius constant for growth	k_g	–	7E3	No
Activation energy for growth	E_g	cal mol^{-1}	5100	No
Arrhenius constant for cell death	k_d	–	1E33	No
Activation energy for cell death	E_d	cal mol^{-1}	50000	No
Density × heat capacity of the medium	ρc_p	cal $°C^{-1}\ L^{-1}$	1/1500	No
Density × heat capacity of the cooling liquid	ρc_{pc}	cal $°C^{-1}\ L^{-1}$	1/2000	No
Yield of heat generation	r_{q1}	cal g_X^{-1}	60	No
Constant in heat generation	r_{q2}	cal $g_X^{-1}h^{-1}$	1.6783E-4	No
Heat transfer coefficient	a	cal $h^{-1}\ °C^{-1}$	1000	No
Constant	b	–	0.60	No
Constant for K_{la}	α_1	–	70	No
Constant for K_{la}	α_2	–	0.4	No
Proportionality constant	γ	mol $[H^+]g_X^{-1}$	1E−5	No
Cooling water temperature	T_c	K	290	No

6.2. Fault detection and diagnosis

We compare the fault detection and diagnosis performances of the hybrid monitoring model to those of a standalone DD monitoring model that uses only the 6 measurements available from the process (Section 4.2 and first six entries in Table 5). We use a window width of 50 h also for the DD model, and cross-validation suggests retaining 3 PCs.

We also implement the KD fault detection and confirmation tests on the innovation sequence of the EKF without parameter

Table A.3
Case study 2: input variables of the detailed and simplified models, with indication on whether or not they are measured. In the detailed model, fluctuations are added around the nominal values as smoothed pseudo-random binary signals with the indicated maximum/minimum amplitudes. In the simplified model, constant (nominal) values are used.

Input variable	Symbol	Units	Nominal value	Max/min amplitude	Measured?	In the FPM?
Feed substrate	F	L h^{-1}	0.045*	8E−4	Yes	Yes
Aeration rate	f_g	L h^{-1}	8	0.08	No	No
Agitator power	P_w	W	30	0.45	No	No

*$F = 0$ L h^{-1} in the first 50 h of the batch.

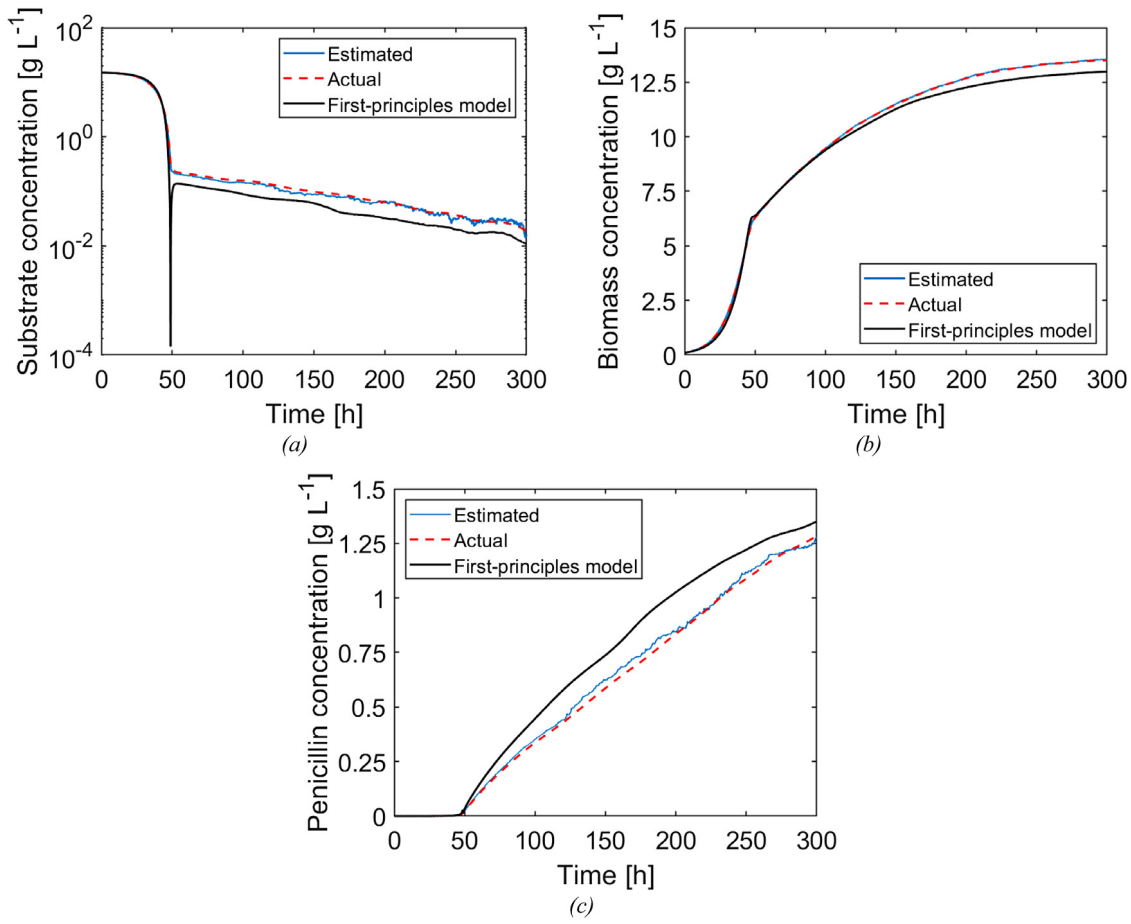

Fig. B.1. Case study 2: EKF state estimation performance for three unmeasured states during a representative NOC batch: (a) substrate concentration, (b) biomass concentration, and (c) penicillin concentration.

estimation for comparison. However, as shown in Figure 12, this approach performs poorly: due to process-model mismatch, false alarms are issued under NOC. Consequently, we build a benchmark KD approach by monitoring drifts of the adjusted parameters with respect to their nominal values [34]. Typical profiles of the adjusted parameters under faulty conditions are shown in Figures B.2–B.4 in Appendix B.

For each fault scenario, the fault detection time is reported in Table 6 as an average across the relevant fault realizations for the three monitoring approaches. The KD monitoring approach struggles with the high variability of the adjusted parameter profiles under NOC, which masks faulty parametric drifts. Only Fault #1 and Fault #2 are correctly detected and diagnosed (see also Figure B.3a and Figure B.4b in Appendix B), although much later than with the hybrid model. On the other hand, neither Fault #3 nor Fault #4 can be detected. This approach also suffers from an additional limitation. Due to the fact that EKF adjusts the model parameters to compensate for the process-model mismatch, the effect of a fault gets smeared into simultaneous variations of all parameters. This limits the possibility to monitor each single parameter in order to detect and isolate the faults. For example, looking at the very fluctuating substrate feed concentration estimation during certain realizations of Fault #3 (Figure B.4c), one might wrongly ascribe the faulty condition to this parameter.

On the other hand, all fault scenarios are detected with hybrid and DD monitoring models for all fault realizations. Still, the hybrid monitoring model alarms the faults much earlier than the DD one for all faults except Fault #4. For this fault, the reactor temperature (a measured variable) is affected by the fault at the same time as the states, and therefore the information the KD block passes to the DD block does not contribute to improve the detection performance.

The DD monitoring model does not allow diagnosing the faults unambiguously, because different faults generate qualitatively similar contribution plots. For example, both Fault #1 (Figure 13a) and Fault #3 (Figure 13c) point to the oxygen concentration (variable no.3), which is indeed a measurement strongly affected by both faults, but is not useful to clearly discriminate the root-causes of the faults. Reactor temperature (measured variable no.2) is pinpointed as the suspected variable for both Fault #2

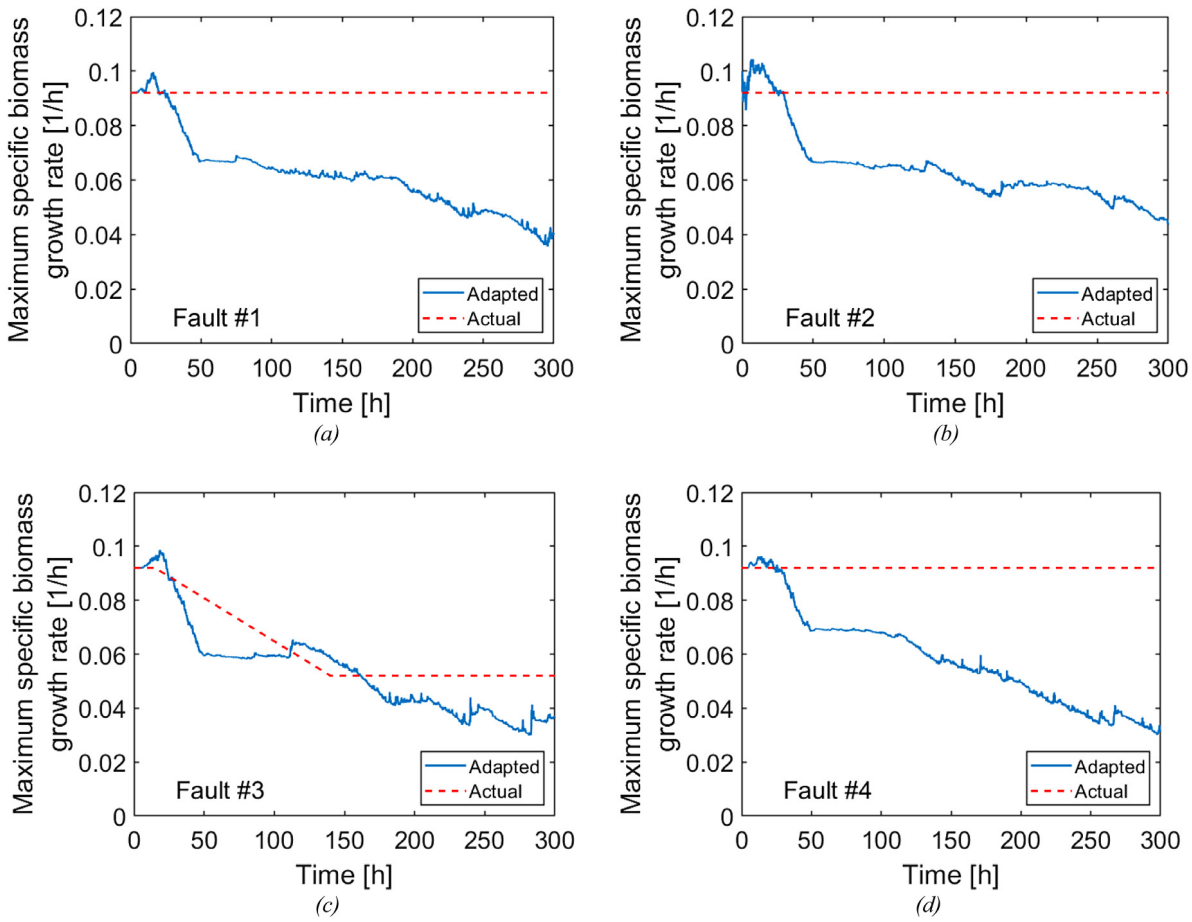

Fig. B.2. Case study 2: EKF parameter adaptation performance for the maximum specific biomass growth rate $\mu_{X,max}$ during a representative batch for (a) Fault #1, (b) Fault #2, (c) Fault #3, and (d) Fault #4.

(Figure 13b) and Fault #4 (Figure 13d), but this leaves the diagnosis problem open. The main difficulty with the DD monitoring model is that in this system there are too few measurements on which a fault can manifest. This implies that different faults become visible only through the same measurements, which makes fault diagnosis harder. On the other hand, in the hybrid monitoring system the estimated states and adapted parameters provide a set of additional "virtual measurements" that can capture a qualitative signature of the fault. Not only does this allow to anticipate fault detection, but it can also point to the root-cause of the fault in a more straightforward way, because — by design — the virtual measurements represent the underlying mechanisms through which the fault propagates into the system.

In fact, the contribution plots derived from the hybrid model (Figure 14) provide information that is very helpful for fault diagnosis. The aeration problem (Fault #1, Figure 14a) is marked by anomalously small contributions for the mass transfer coefficient (variable no. 14) and the oxygen concentration (variable no. 10), whereas in Fault #2 the abnormal feed concentration is clearly spotted (Figure 14b, variable no.15). For the biomass growth rate decrease problem (Fault #3; Figure 14c), the most significant contributions refer to variables related to the biomass reaction, including the online adapted kinetic parameter (variable no. 13); from this piece of information, a biomass growth reaction problem can be diagnosed in a relatively easy way. The only fault in which the main contribution is the reactor temperature is Fault #4 (variable no. 2; Figure 14d), but temperature is a measurement where an abnormal reactor cooling can leave a footprint directly. Hence, the hybrid model contribution plot does no better than the DD model one.

7. Conclusions

In this study, we proposed a novel framework for multivariate process monitoring based on a hybrid modeling approach. Real-time deterministic information about the process is first obtained in a knowledge-driven block from a state estimator in the form of estimated states, reconstructed measurements, and possibly adapted parameters. The information is then passed to a data-driven block, where it is exploited, in conjunction with the available field measurements, by a latent-variable model that accomplishes multivariate fault detection and diagnosis. The design of the two blocks is largely independent, which makes implementation of the proposed methodology easier.

We tested the hybrid methodology on two simulated case studies, namely a continuous process and a fed-batch one. It typically allowed for earlier fault detection than standard data-driven and knowledge-driven approaches taken in isolation, even when the state estimator did not perform entirely satisfactorily. In addition, using the hybrid approach significantly facilitated fault diagnosis.

The very satisfactory fault detection performance of the hybrid approach derives from the fact that the estimated states (and possibly the adapted parameters) provide a set of additional variables a fault can leave a footprint on. In most cases, these variables respond to the fault earlier than the measurements, causing an anticipated shift or break of the normal correlation structure of the data, which can be promptly captured and alarmed by the latent-variable model. With respect to fault diagnosis, if a fault manifests itself as an abnormal change in one or more states (or

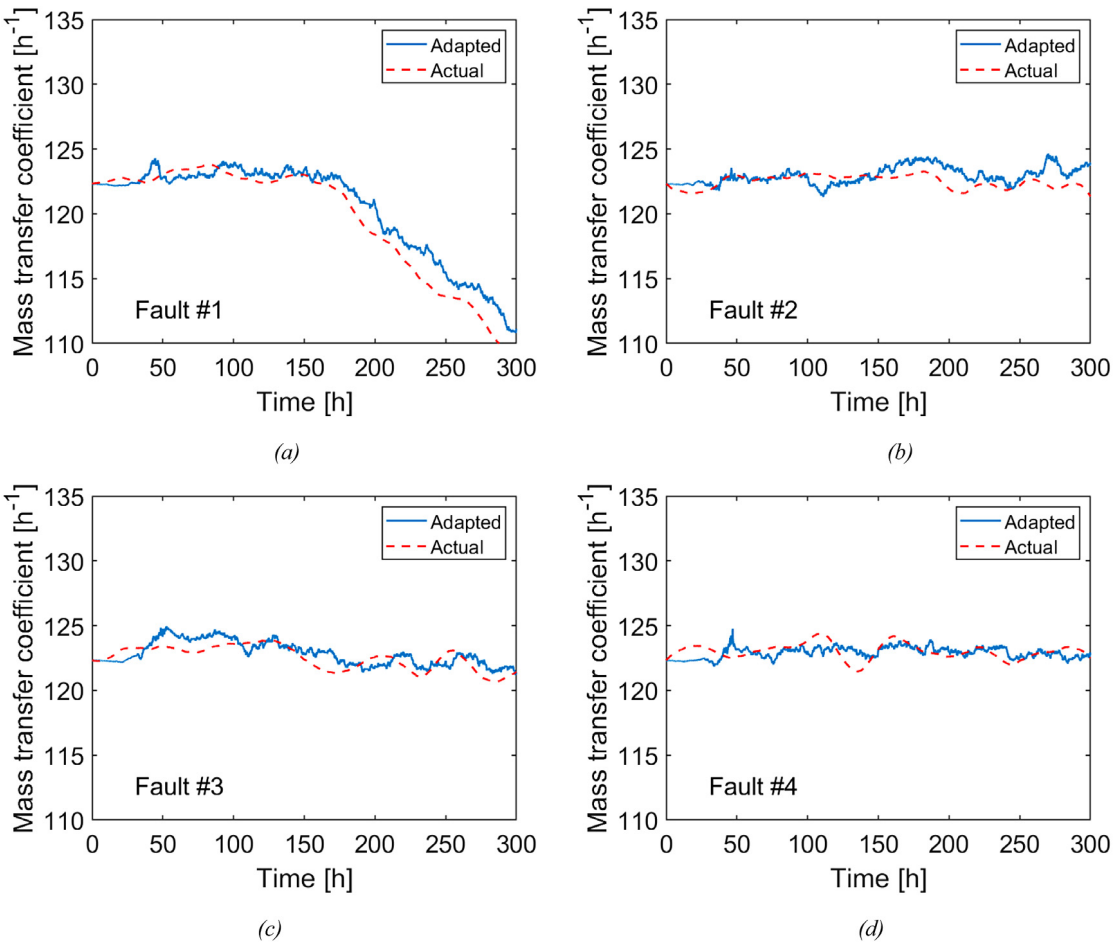

Fig. B.3. Case study 2: EKF parameter adaptation performance for the mass transfer coefficient K_{la} during a representative batch for (a) Fault #1, (b) Fault #2, (c) Fault #3, and (d) Fault #4.

parameters), diagnosing that fault with the hybrid model is generally easier, because the states and parameters straightforwardly point to the inner mechanism that is being impacted by the fault. This enables one to disclose the root-cause of the fault with less ambiguity than can be done using field measurements alone.

The successful performance of the hybrid monitoring system is due to the inclusion of the estimated states (and possibly of the adapted parameters) within a multivariate framework together with the measurements, rather than to the mere implementation of a knowledge-driven component. In fact, traditional knowledge-driven approaches (like innovation sequence monitoring or parametric drift detection) lack the well-known advantages of process monitoring by latent-variable modeling, and they were found not to be able to cope with even mild process-model mismatch.

As for all data-driven methodologies, a word of caution must be mentioned for the hybrid approach in relation to the region wherein the process is run. In fact, the monitoring performance may be compromised if the process is operated away from the region over which the data-driven component was calibrated.

CRediT authorship contribution statement

Francesco Destro: Conceptualization, Methodology, Software, Validation, Investigation, Writing - original draft, Visualization, Formal analysis. **Pierantonio Facco:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Salvador García Muñoz:** Conceptualization, Formal analysis, Writing - review & editing. **Fabrizio Bezzo:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing. **Massimiliano Barolo:** Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Supervision, Funding acquisition.

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

We gratefully acknowledge financial support from the University of Padova under project BIRD194889-SID 2019 “Augmenting data-driven models with knowledge-driven information to enhance process monitoring in the Industry 4.0 era (AUGH)”. F.D. gratefully acknowledges the CARIPARO Foundation for his PhD scholarship.

Appendix A. Penicillin manufacturing: detailed and simplified models

The fed-batch process for the manufacturing of penicillin by biomass fermentation is simulated using the detailed model by Birol et al. [42].

$$\frac{dX}{dt} = \mu_x X - \frac{X}{V}\frac{dV}{dt} \tag{A.1}$$

$$\frac{dP}{dt} = \mu_P X - KP - \frac{P}{V}\frac{dV}{dt} \tag{A.2}$$

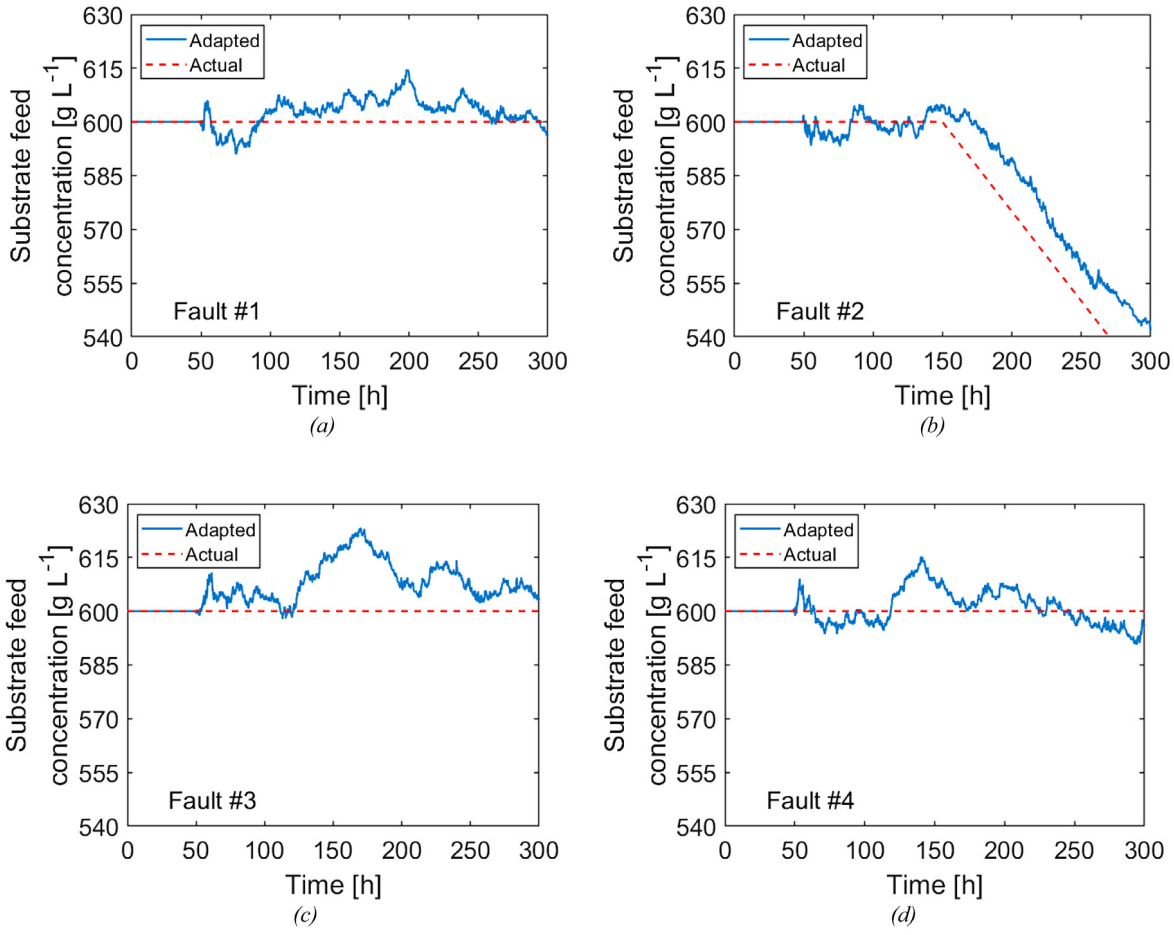

Fig. B.4. Case study 2: EKF parameter adaptation performance for the substrate feed concentration s_F during a representative batch for (a) Fault #1, (b) Fault #2, (c) Fault #3, and (d) Fault #4.

$$\frac{dS}{dt} = -\frac{\mu_x}{Y_{X/S}}X - \frac{\mu_P}{Y_{P/S}}X - m_x X + F\frac{s_F}{V} - \frac{S}{V}\frac{dV}{dt} \tag{A.3}$$

$$\frac{dC_{O_2}}{dt} = \frac{\mu_x}{Y_{X/O}}X - \frac{\mu_P}{Y_{P/O}}X - m_o X + K_{la}\left(c^*_{O_2} - c_{O_2}\right) - \frac{C_{O_2}}{V}\frac{dV}{dt} \tag{A.4}$$

with:

$$K_{la} = \alpha_1\sqrt{f_g}\left(\frac{P_w}{V}\right)^{\alpha_2}. \tag{A.5}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{U}{s_F} + F_{a/b} - V\lambda(e^{\frac{5(T-T_0)}{T_v-T_0}} - 1) \tag{A.6}$$

$$\frac{dc_{CO_2}}{dt} = \alpha\frac{dX}{dt} + \beta X + \gamma \tag{A.7}$$

$$\mu_P = \mu_{P,max}\frac{S}{k_p + S(1+S/K_I)}\frac{C^p_{O_2}}{k_{op}X + C^p_{O_2}} \tag{A.8}$$

$$\mu_X = \left[\frac{\mu_{x,max}}{1 + \frac{K_1}{[\mathrm{H}^+]} + \frac{[\mathrm{H}^+]}{K_2}}\right]\frac{S}{k_x X + S}\frac{C_{O_2}}{k_{ox}X + C_{O_2}} \times \left\{\left[k_g \exp\left(-\frac{E_g}{RT}\right)\right] - \left[k_d \exp\left(-\frac{E_d}{RT}\right)\right]\right\} \tag{A.9}$$

$$\frac{\mathrm{d}\left[\mathrm{H}^+\right]}{\mathrm{d}t} = \gamma\left(\mu X - \frac{FX}{V}\right) + \frac{\left[\frac{-\delta+\sqrt{(\delta^2+4\cdot 10^{-14})}}{2} - \left[H^+\right]\right]}{\Delta t} \tag{A.10}$$

with:

$$\delta = \left[\frac{10^{-14}}{[\mathrm{H}^+]} - \left[\mathrm{H}^+\right]\right]V - \frac{C_{a/b}\,(F_a + F_b)\,\Delta t}{V + (F_a + F_b)\,\Delta t}. \tag{A.11}$$

$$\frac{\mathrm{d}Q_r}{\mathrm{d}t} = r_{q1}\frac{\mathrm{d}X}{\mathrm{d}t}V + r_{q2}XV. \tag{A.12}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{F}{s_F}\left(T_f - T\right) + \frac{1}{V\rho c_p}\left[Q_r - \frac{aF_c^{b+1}(T - T_c)}{F_c + \left(\frac{aF_c^b}{2\rho c_{pc}}\right)}\right]. \tag{A.13}$$

We report in Table A.1 the symbols used for the states and outputs, and in Tables A.2 and A.3 those used for the parameters and the inputs, respectively.

To allow for more variability under NOC, thus making the monitoring problem more challenging, we detune the pH and temperature loops with respect to the original tuning in [42]. We also include process noise as fluctuations in the inputs, as reported in Table A.3.

The faulty batches discussed in Section 4.2 are initialized as the NOC ones, and then the following changes are considered in the detailed model:

- Fault #1: aeration rate decreases by 0.01 L h^{-1} from 150 h to 300 h;
- Fault #2: substrate feed concentration decreases by 0.50 g$_S$ L^{-1} h^{-1} from 150 h to 300 h;
- Fault #3: $\mu_{X,\max}$ decreases by 3.2E−4 h^{-2} from 15 h to 140 h;
- Fault #4: $T_c = 298$ K from the beginning of the batch.

The simplified model by Bajpai and Reuss [44], complemented with the CO_2 balance and a simplified version of the volume loss equation from Birol et al. [42], are used as the FPM. Namely, the FPM is composed by Eqs. (A.1)–(A.4), Eqs. (A.7)–(A.8) and the following equations as the volume balance and the biomass growth kinetics:

$$\frac{dV}{dt} = F - \lambda V \tag{A.14}$$

$$\mu_X = \mu_{X,\max} \frac{S}{k_x X + S} \frac{C_{O_2}}{k_{ox} X + C_{O_2}} \tag{A.15}$$

States, parameters and inputs are set as in Tables A.1–A.3. Note that the FPM retains only 6 of the 9 states of the process. The main sources of parametric and structural mismatch are as follows:

- protons, heat and energy balances Eqs. (A.10)–(A.13) are not included in the FPM;
- Eqs. (A.14) and (A.15) do not consider the effects of pH and temperature on the kinetic parameters and on the volume loss by evaporation, which are significant in the process and are accounted for in the detailed model;
- in the process, the mass transfer coefficient K_{la} depends on two inputs (f_g and P_w, Eq. (A.5)), which are subject to small fluctuations. The FPM neglects this dependency, assuming $K_{la} = 120.3\ \mathrm{h}^{-1}$.

Appendix B. Penicillin manufacturing: state estimation and parameter adaptation results

In this appendix we report representative results for state estimation and parameter adaptation in Case Study 2. Figure B.1 shows the estimations of three unmeasured states (substrate, biomass and penicillin concentrations) during a representative NOC batch. The adapted values of the maximum specific biomass growth rate, mass transfer coefficient, and substrate feed concentration during representative batches for each fault scenario are reported in Figs. B.2–B.4 (respectively).

References

[1] S.J. Qin, Survey on data-driven industrial process monitoring and diagnosis, Annu. Rev. Control. 36 (2012) 220–234, http://dx.doi.org/10.1016/j.arcontrol.2012.09.004.

[2] Q. Jiang, X. Yan, B. Huang, Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes, Ind. Eng. Chem. Res. 58 (2019) 12899–12912, http://dx.doi.org/10.1021/acs.iecr.9b02391.

[3] J.E. Jackson, A user's guide to principal components, John Wiley & Sons, New York (U.S.A.), 1991.

[4] J.V. Kresta, J.F. Macgregor, T.E. Marlin, Multivariate statistical monitoring of process operating performance, Can. J. Chem. Eng. 69 (1991) 35–47, http://dx.doi.org/10.1002/cjce.5450690105.

[5] P. Nomikos, J.F. Macgregor, Multivariate processes SPC charts for monitoring, Technometrics 37 (1995) 41–59, http://dx.doi.org/10.2307/1269152.

[6] P. Miller, R.E. Swanson, C.E. Heckler, Contribution plots: a missing link in multivariate quality control, Appl. Math. Comput. Sci. 8 (1998) 775–792.

[7] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault detection and diagnosis part I: quantitative model-based methods, Comput. Chem. Eng. 27 (2003) 293–311, http://dx.doi.org/10.1016/S0098-1354(02)00160-6.

[8] Z. Gao, C. Cecati, S.X. Ding, A survey of fault diagnosis and fault-tolerant techniques-part I: fault diagnosis with model-based and signal-based approaches, IEEE Trans. Ind. Electron. 62 (2015) 3757–3767, http://dx.doi.org/10.1016/j.arcontrol.2012.09.004.

[9] J. Gertler, Fault Detection and Diagnosis in Engineering Systems, Routledge, 1998.

[10] J. Mohd, N.H. Hoang, M.A. Hussain, D. Dochain, Review and classification of recent observers applied in chemical process systems, Comput. Chem. Eng. 76 (2015) 27–41, http://dx.doi.org/10.1016/j.compchemeng.2015.01.019.

[11] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, J. Schröder, Diagnosis and Fault-Tolerant Control, Springer, 2006.

[12] A.P. Deshpande, S.C. Patwardhan, S.S. Narasimhan, Intelligent state estimation for fault tolerant nonlinear predictive control, J. Process Control. 19 (2009) 187–204, http://dx.doi.org/10.1016/j.jprocont.2008.04.006.

[13] F. Caccavale, F. Pierri, M. Iamarino, V. Tufano, An integrated approach to fault diagnosis for a class of chemical batch processes, J. Process Control. 19 (2009) 827–841, http://dx.doi.org/10.1016/j.jprocont.2008.11.003.

[14] L.A. Rusinov, N.V. Vorobiev, V.V. Kurkina, Fault diagnosis in chemical processes and equipment with feedbacks, Chemometr. Intell. Lab. Syst. 126 (2013) 123–128, http://dx.doi.org/10.1016/j.chemolab.2013.03.015.

[15] D. Varshney, M. Bhushan, S.C. Patwardhan, State and parameter estimation using extended Kitanidis Kalman filter, J. Process Control. 76 (2019) 98–111, http://dx.doi.org/10.1016/j.jprocont.2018.11.007.

[16] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: past, present and future, Comput. Chem. Eng. 60 (2013) 86–101, http://dx.doi.org/10.1016/j.compchemeng.2013.08.008.

[17] S. Zendehboudi, N. Rezaei, A. Lohi, Applications of hybrid models in chemical, petroleum, and energy systems: a systematic review, Appl. Energy 228 (2018) 2539–2566, http://dx.doi.org/10.1016/j.apenergy.2018.06.051.

[18] Q.P. He, J. Wang, Statistical process monitoring as a big data analytics tool for smart manufacturing, J. Process Control. 67 (2018) 35–43, http://dx.doi.org/10.1016/j.jprocont.2017.06.012.

[19] M.S. Reis, G. Gins, T.J. Rato, Incorporation of process-specific structure in statistical process monitoring: a review, J. Qual. Technol. (2019) 1–15, http://dx.doi.org/10.1080/00224065.2019.1569954.

[20] R. Jia, Z. Mao, Y. Chang, L. Zhao, Soft-sensor for copper extraction process in cobalt hydrometallurgy based on adaptive hybrid model, Chem. Eng. Res. Des. 89 (2011) 722–728.

[21] D. Bonvin, C. Georgakis, C.C. Pantelides, M. Barolo, M.A. Grover, D. Rodrigues, R. Schneider, D. Dochain, Linking models and experiments, Ind. Eng. Chem. Res. 55 (2016) 6891–6903, http://dx.doi.org/10.1021/acs.iecr.5b04801.

[22] K. Tidriri, N. Chatti, S. Verron, T. Tiplica, Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: a review of researches and future challenges, Annu. Rev. Control. 42 (2016) 63–81, http://dx.doi.org/10.1016/j.arcontrol.2016.09.008.

[23] K. Ghosh, Y.S. Ng, R. Srinivasan, Evaluation of decision fusion strategies for effective collaboration among heterogeneous fault diagnostic methods, Comput. Chem. Eng. 35 (2011) 342–355, http://dx.doi.org/10.1016/j.compchemeng.2010.05.004.

[24] I. Baklouti, M. Mansouri, A. Ben Hamida, H.N. Nounou, M. Nounou, Enhanced operation of wastewater treatment plant using state estimation-based fault detection strategies, Int. J. Control. (2019) 1–12, http://dx.doi.org/10.1080/00207179.2019.1590735.

[25] I. Baklouti, M. Mansouri, A. Ben Hamida, H. Nounou, M. Nounou, Monitoring of wastewater treatment plants using improved univariate statistical technique, Process Saf. Environ. Prot. 116 (2018) 287–300, http://dx.doi.org/10.1016/j.psep.2018.02.006.

[26] D.E. Seborg, T.F. Edgar, D.A. Mellichamp, F.J. Doyle, Process dynamics and control, fourth ed., John Wiley & Sons, New York (U.S.A.), 2017.

[27] W.H. Ray, Advanced process control, McGraw-Hill, New York, 1981.

[28] R. Schneider, C. Georgakis, How to NOT make the extended kalman filter fail, Ind. Eng. Chem. Res. 52 (2013) 3354–3362, http://dx.doi.org/10.1021/ie300415d.

[29] N.L. Ricker, J.H. Lee, Nonlinear modeling and state estimation for the Tennessee Eastman challenge process, Comput. Chem. Eng. 19 (1995) 983–1005, http://dx.doi.org/10.1016/0098-1354(94)00113-3.

[30] G. Pérez, M. Garmendia, J.F. Reynaud, J. Crego, U. Viscarret, Enhanced closed loop state of charge estimator for lithium-ion batteries based on extended Kalman filter, Appl. Energy. 155 (2015) 834–845, http://dx.doi.org/10.1016/j.apenergy.2015.06.063.

[31] J.A. Delgado-Aguiñaga, G. Besançon, O. Begovich, J.E. Carvajal, Multi-leak diagnosis in pipelines based on extended Kalman filter, Control Eng. Pract. 49 (2016) 139–148, http://dx.doi.org/10.1016/j.conengprac.2015.10.008.

[32] D. Simon, Optimal state estimation: Kalman, H infinity, and nonlinear approaches, John Wiley & Sons, New York (U.S.A.), 2006.

[33] V. Liotta, C. Georgakis, M.S. El-Aasser, Real-time estimation and control of particle size in semi-batch emulsion polymerization, in: Proc. 1997 Am. Control Conf. (Cat. No. 97CH36041), IEEE, 1997, pp. 1172–1176, http://dx.doi.org/10.1109/ACC.1997.609717.

[34] W. Ku, R.H. Storer, C. Georgakis, Uses of state estimation for statistical process control, Comput. Chem. Eng. 18 (1994) S571–S575, http://dx.doi.org/10.1016/0098-1354(94)80093-6.

[35] E. Che Mid, V. Dua, Model-based parameter estimation for fault detection using multiparametric programming, Ind. Eng. Chem. Res. 56 (2017) 8000–8015.

[36] B.M. Wise, N.B. Gallagher, The process chemometrics approach to process monitoring and fault detection, J. Process Control. 6 (1996) 329–348, http://dx.doi.org/10.1016/0959-1524(96)00009-1.

[37] J.A. Westerhuis, S.P. Gurden, A.K. Smilde, Generalized contribution plots in multivariate statistical process monitoring, Chemometr. Intell. Lab. Syst. 51 (2000) 95–114, http://dx.doi.org/10.1016/S0169-7439(00)00062-9.
[38] C. Ündey, S. Ertunç, A. Çinar, Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis, Ind. Eng. Chem. Res. 42 (2003) 4645–4658, http://dx.doi.org/10.1021/ie0208218.
[39] W. Ku, R.H. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, Chemometr. Intell. Lab. Syst. 30 (1995) 179–196, http://dx.doi.org/10.1016/0169-7439(95)00076-3.
[40] J. Camacho, J. Picó, A. Ferrer, Bilinear modelling of batch processes. part I: theoretical discussion, J. Chemom. 22 (2008) 299–308, http://dx.doi.org/10.1002/cem.1113.
[41] C. Ling, C. Kravaris, State observer design for monitoring the degree of polymerization in a series of melt polycondensation reactors, Processes 4 (2016) 4, http://dx.doi.org/10.3390/pr4010004.
[42] G. Birol, C. Ündey, A. Çinar, A modular simulation package for fed-batch fermentation: Penicillin production, Comput. Chem. Eng. 26 (2002) 1553–1565, http://dx.doi.org/10.1016/S0098-1354(02)00127-8.
[43] C. Scali, M. Morretta, D. Semino, Control of the quality of polymer products in continuous reactors: Comparison of performance of state estimators with and without updating of parameters, J. Process Control. 7 (1997) 357–369, http://dx.doi.org/10.1016/S0959-1524(97)00013-9.
[44] R.K. Bajpai, M. Reuß, A mechanistic model for penicillin production, J. Chem. Technol. Biotechnol. 30 (1980) 332–344, http://dx.doi.org/10.1002/jctb.503300140.