Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Monitoring a segmented fluid bed dryer by hybrid data-driven/knowledge-driven modeling

Francesco Destro* Andrew J. Salmon Pierantonio Facco* Constantinos C. Pantelides**,*** Fabrizio Bezzo* Massimiliano Barolo***

**CAPE-Lab, Department of Industrial Engineering, University of Padova, Padova, Italy*
*** Process Systems Enterprise Ltd, London W6 7HA, UK*
****Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK*

Abstract: Many data-driven and knowledge-driven methods for process monitoring have been developed in the last decade. In this study we show that the combined use of techniques from both categories can potentially outperform their standalone use. The proposed hybrid approach for fault detection and diagnosis is grounded in conventional multivariate statistical process monitoring. However, the datasets subject to analytics include not only field measurements, but also data obtained from a state estimator based on a mathematical model of the process. We apply the proposed methodology to a pharmaceutical case study, using the mechanistic model of a segmented fluid bed dryer from gPROMS FormulatedProducts. The hybrid framework demonstrates improved fault detection and diagnosis performances, when compared to data-driven monitoring or state estimation taken in isolation.

Keywords: process monitoring, process control, fault detection, fault diagnosis, hybrid model, gPROMS, Industry 4.0

1. INTRODUCTION

The economic loss associated with the occurrence of a fault in a plant increases quickly with the time and resources needed for its detection and diagnosis. Efficient online monitoring systems must be able to promptly detect deviations from normal operating conditions (NOC) and to provide effective tools to assess the cause of the abnormal behavior.

The techniques proposed in the last decade to fulfil these tasks can broadly be divided into two categories: data-driven (Jiang *et al.*, 2019) and knowledge-driven methods (Venkatasubramanian *et al.*, 2003). Latent-variables models (LVMs; Qin, 2012) are probably the most popular data-driven techniques used for monitoring purposes in the process industry. LVMs are multivariate correlative models, trained on field measurements representing the NOC of a plant. Faults are detected on multivariate charts when new field measurements are foreign to the NOC correlation structure (Nomikos and MacGregor, 1995). Contribution plots (Miller *et al.*, 1998) are then used to assess which of the available measurements are most correlated to the abnormal condition.

One drawback of this approach is that the fault can be detected only after it has manifested itself as a change in one or more measured variables. Therefore, if the measured variables do not relate to the fault directly, the fault cannot be detected promptly, because it must propagate into the process until it becomes visible through the available measurements. Fault diagnosis may then become difficult: since the process status has been altered significantly, several measurements might display high contributions to the fault (an issue known as the smearing-out effect; Qin, 2012), but none of them unequivocally pinpoints the root-cause of the fault. The monitoring task becomes even more difficult when only a limited number of measurement sensors are available. In such cases, fault detection may be delayed until those few measurements are affected by the fault, but the contribution plots will also unavoidably point always to the same measured variables, even for different fault scenarios, thus making fault diagnosis more challenging.

The fault detection and diagnosis problems may be addressed through knowledge-driven monitoring, based on a mechanistic model that provides comprehensive information about the internal variables characterizing the process, even if they are not measured (Mohd *et al.*, 2015). The main drawback is that the development of first-principles models is usually time-consuming, hence more expensive than data-driven modeling. In addition, knowledge-driven models are affected by process-model mismatch. Nonetheless, mechanistic models, describing at least the main physical phenomena occurring in process units, are being increasingly used with the advent of the Industry 4.0 digitalization paradigm, and the interest towards their real-time application is also increasing (Pantelides and Renfro, 2013). Currently, their application for industrial process monitoring is mainly in soft-sensing or univariate process monitoring.

In this study we propose a novel hybrid framework for fault detection and diagnosis, which combines multivariate data analytics techniques with information from a mechanistic model of the process under investigation. State-of-the-art approaches to hybrid monitoring focus on using data-driven models to compensate for process/model mismatch (Bonvin *et al.*, 2016). We propose instead a straightforward and effective fusion strategy between the data-driven and knowledge-driven modeling approaches. First, we implement

Copyright lies with the authors

a state estimator, based on a mechanistic model of the process, for soft sensing. Then, we use an LVM for multivariate monitoring. The data-driven model is built on a dataset consisting of *i*) field measurements, and *ii*) variables estimated by the state estimator. This ensures that the available first-principles knowledge on the process is embedded in the LVM effectively, and it is actively used for fault detection and diagnosis in addition to its soft-sensing features. Furthermore, partially uncompensated process/model mismatch can be tolerated as it becomes part of the correlation structure of the NOC dataset.

We apply the hybrid framework to monitor the operation of a segmented fluid bed dryer, a cyclic batch process used in the pharmaceutical industry to emulate continuous processing. A first-principles model of this unit is available in the gPROMS FormulatedProducts software (Process Systems Enterprise, 2020). We employ the extended Kalman filter (EKF; Ray, 1981) solver available in gPROMS for state estimation, whereas we use principal component analysis (PCA; Jackson, 1991) as the LVM.

The remainder of this article is organized as follows. In Section 2 the mathematical methods employed are outlined. In Section 3, we present the proposed hybrid monitoring system. In Section 4, we apply the hybrid framework to the case study, and discuss the fault detection and diagnosis results. Final conclusions are addressed in Section 5.

2. MATHEMATICAL METHODS

2.1 Extended Kalman filter for state estimation

A nonlinear mechanistic model of a process, formulated as a set of ordinary differential equations, can be represented as:

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

where $\mathbf{x}$ is the state vector, $\mathbf{u}$ is the input vector, t denotes the time, $\mathbf{f}$ is a nonlinear function, and the process noise $\mathbf{w}$ is assumed to be normally distributed with $\mathbf{0}$ mean and covariance $\mathbf{Q}$. The vector $\mathbf{y}$ of measurements from the plant sensors, available at finite sampling times t_k, can be related to the system states through a measurement model:

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

where $\mathbf{h}$ is a nonlinear function, and the measurement noise $\mathbf{v}$ is assumed to follow a normal distribution with $\mathbf{0}$ mean and covariance $\mathbf{R}$.

Based on (1) and (2), the discrete-time data EKF algorithm (Ray, 1981) produces estimates $\hat{\mathbf{x}}$ of the state vector and $\hat{\mathbf{P}}$ of the state covariance through subsequent prediction and update steps. The EKF is first initialized with the initial estimation of the states $\mathbf{x}_0$ and the initial state covariance $\mathbf{P}_0$. Between sampling times t_{k-1} and t_k, the predictions of the states $\hat{\mathbf{x}}(t_k|t_{k-1})$ and of the state covariance $\hat{\mathbf{P}}(t_k|t_{k-1})$ are obtained upon integration of (3) and (4), with the updated estimations of states $\hat{\mathbf{x}}(t_{k-1}|t_{k-1})$ and state covariance $\hat{\mathbf{P}}(t_{k-1}|t_{k-1})$ at t_{k-1} as initial conditions:

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

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

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

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

The estimates are updated at the sampling times with:

$$\hat{\mathbf{x}}(t_k \mid t_k) = \hat{\mathbf{x}}(t_k \mid t_{k-1}) + \mathbf{K}\left[\mathbf{y}(t_k) - \mathbf{h}\left(\hat{\mathbf{x}}(t_k|t_{k-1}), \mathbf{u}(t_k), t_k\right)\right] \tag{6}$$

$$\hat{\mathbf{P}}(t_k \mid t_k) = (\mathbf{I} - \mathbf{KH})\, \hat{\mathbf{P}}(t_k \mid t_{k-1}) \quad , \tag{7}$$

where the Kalman gain $\mathbf{K}$ and the Jacobian matrix $\mathbf{H}$ are respectively calculated according to:

$$\mathbf{K} = \hat{\mathbf{P}}(t_k \mid t_{k-1})\mathbf{H}^{\mathbf{T}}\left[\mathbf{H}\hat{\mathbf{P}}(t_k|t_{k-1})\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right]^{-1} \tag{8}$$

$$\mathbf{H} = \left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right)_{\hat{\mathbf{x}}(t_k|t_{k-1}), \mathbf{u}(t_k), t_k} \tag{9}$$

In this study, we adopt a generalized version of the EKF (Cheng *et al.*, 1997) for models expressed as mixed systems of algebraic and ordinary differential equations, instead of differential equations only as in (1). This state estimation algorithm is implemented within the solvers available in gPROMS FormulatedProducts.

2.2 Principal component analysis for multivariate process monitoring

Given a NOC dataset $\mathbf{Z}$ [$N \times M$] with N observations and M variables, PCA extracts A principal components (PCs) expressing the directions of maximum variability in the data, and decomposes $\mathbf{Z}$ as:

$$\mathbf{Z} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{\mathbf{T}} + \mathbf{E} \quad , \tag{10}$$

where $\mathbf{t}_a$ [$N \times 1$] and $\mathbf{p}_a$ [$M \times 1$] are respectively the score and the loading vectors for the a-th PC, and $\mathbf{E}$ is the residuals matrix, consisting of random noise if A is selected appropriately (Ku *et al.*, 1995). Upon calibration of a PCA model, the Hotelling T^2 and the squared prediction error (SPE) statistics are calculated for observation n by:

$$T_n^2 = \sum_{a=1}^{A} t_{a,n} \lambda_a^{-1} t_{a,n} \tag{11}$$

$$\mathrm{SPE}_n = \mathbf{e}_n \mathbf{e}_n^{\mathbf{T}} \quad , \tag{12}$$

where $t_{a,n}$ and $\mathbf{e}_n$ are respectively the element in the a-th score vector and the row vector in $\mathbf{E}$ associated to observation n. λ_a is the eigenvalue of the ath PC, calculated via the PCA algorithm. In this study, we build monitoring charts as in Nomikos and MacGregor (1995), namely using the χ^2-distribution and the F-distribution 99% confidence limits for SPE and T^2 respectively.

When a new observation $\mathbf{z}_{new}(t_k)$ [$1 \times M$] from the process becomes available at time t_k, SPE_{new} and T_{new}^2 are calculated and monitored in the corresponding charts. If some (e.g.,

three) consecutive observations are out-of-control for a given statistic, a fault is alarmed, and contribution plots (Miller *et al.*, 1998) are looked at for diagnosis. Contributions to SPE or T^2 are M-dimensional vectors that, when calculated upon detection of a fault, provide meaningful indication of the variables in $\mathbf{Z}$ that are most related to the abnormal condition. The SPE and T^2 contribution vectors for observation n are obtained by:

$$\mathbf{c}_n^{\mathrm{SPE}} = \mathbf{e}_n \tag{13}$$

$$\mathbf{c}_n^{T^2} = \sum_{a=1}^{A} t_{a,n} \lambda_a^{-0.5} \mathbf{p}_a^{\mathrm{T}} \quad . \tag{14}$$

To further aid visual assessment of the variables mainly responsible for the fault, we calculate the Gaussian 99% confidence limits on the contributions for the NOC dataset (Westerhuis *et al.*, 2000).

To monitor a dynamic process, like the one considered in this study, the standard PCA procedure should be modified to consider also the auto-correlation of the data in $\mathbf{Z}$. To this purpose, we use multi-model moving-window PCA (Camacho *et al.*, 2008). Namely, at each time step t_k we build a PCA model on the data matrix $\mathbf{Z}_{dyn}(t_k)$, consisting of $\mathbf{z}(t_k)$ and l lags, i.e. the observations at the previous l time steps:

$$\mathbf{Z}_{dyn}(t_k) = \left[\mathbf{z}^{\mathrm{T}}(t_k) \quad \mathbf{z}^{\mathrm{T}}(t_k - 1) \quad \ldots \quad \mathbf{z}^{\mathrm{T}}(t_k - l)\right] \quad . \tag{15}$$

3. PROPOSED HYBRID MONITORING FRAMEWORK

The proposed hybrid monitoring framework is based on a serial-parallel approach (Fig. 1). Let us consider a process with a set of measured inputs $\mathbf{u}$ and measured outputs $\mathbf{y}$, subject to unmeasured disturbances $\mathbf{d}$. A knowledge-driven block receives the measurements $\mathbf{u}$ and $\mathbf{y}$ and performs state estimation, based on an available mechanistic model of the process. A subset $\tilde{\mathbf{x}}$ of the estimated variables is then used in the data-driven block, together with the measurements from the plant, for fault detection and diagnosis through an LVM. Hence, the dataset upon which we build the LVM is:

$$\mathbf{Z} = \left[\mathbf{u} \quad \mathbf{y} \quad \tilde{\mathbf{x}}\right] \quad . \tag{16}$$

Among all the variables calculated by the state estimator, only those considered to have a physical meaning useful for fault detection and diagnosis are selected for inclusion in $\tilde{\mathbf{x}}$. As a result, $\tilde{\mathbf{x}}$ includes a subset of the differential states and algebraic variables defined in the mechanistic model. Since a subset of the estimated variables might also be measured, $\mathbf{Z}$ will include those variables twice: once in $\mathbf{y}$ (measurements from sensors) and once in $\tilde{\mathbf{x}}$ (measurements reconstructed by the estimator). State estimators are often also employed for online adjustment of the model parameters. In this situation, it might be convenient to include the estimated parameters in $\tilde{\mathbf{x}}$, especially if they are related to specific fault conditions to be monitored.

The proposed hybrid framework has important advantages over its individual building blocks taken in isolation. Augmentation of the measurements dataset with the estimated variables allows the LVM to access information on the physical phenomena occurring in the process, thus resulting in improved fault detection and diagnosis over what is achievable via data-driven monitoring on its own. Besides, inclusion of estimated variables within the multivariate framework is more effective for fault detection and diagnosis than using the estimated variables by themselves in a univariate fashion. In fact, faults often manifest themselves earlier as a (possibly small) co-variation of states rather than as a co-variation of measurements, and this can be detected promptly by multivariate analysis using the proposed framework. On the other hand, monitoring by knowledge-driven modeling alone may be problematic in the presence of process-model mismatch, which can partially mask drifts in unmeasured variables, thus resulting in a state estimator that captures only small deviations spread across several variables. In such a situation, there are further benefits from using a multivariate data-driven block.

In the example under investigation, we use an EKF in the knowledge-driven block, and PCA in the data-driven block. However, other state estimators or multivariate modeling techniques could be used within the same framework.

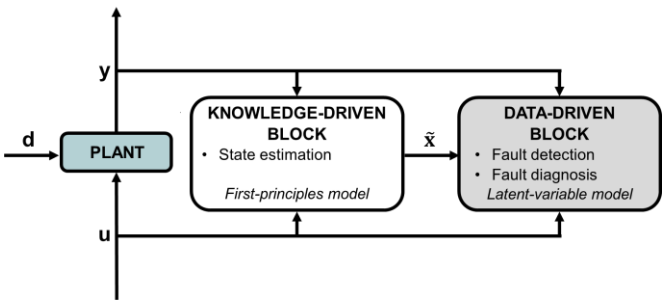

Fig. 1. Proposed hybrid monitoring framework.

4. HYBRID MONITORING OF A SEGMENTED FLUID BED DRYER

4.1 The process

To test the proposed hybrid monitoring framework, we generate synthetic data with the segmented fluid bed dryer model available in gPROMS FormulatedProducts. The model (223 differential equations and 6256 algebraic equations) represents the physical phenomena occurring in real fluid bed dryers (Burgschweiger and Tsotsas, 2002).

We use the dryer model to simulate a pharmaceutical process (Fig. 2) in which the moisture content of wet granules fed to the unit is reduced by flowing hot air. The dryer receives a continuous feed of wet granules, with each of the six segments behaving as a fluidized bed that cycles through four phases: loading, drying, discharging and waiting. We refer to the sequence of loading, drying and discharging phases in a given segment as a "batch"; the waiting phase is not considered here as the segment is empty in that period of time. During the dryer operation, a batch is processed in each of the segments. In Fig. 2, segment #2 is being loaded with wet granules, thus starting a new batch for that segment. When segment #2 is fully loaded, the loading of segment #3 (which was in the waiting phase until that moment) starts. Meanwhile, segment #4 is discharging, while segments #1, #5 and #6 are in the drying phase. Hot air is continuously fed

to the dryer, and its flow is distributed between the six segments through a distributor plate (not shown in Fig. 2). Details about the model assumptions and equations can be found in Burgschweiger and Tsotsas (2002).
We assume that, as in typical industrial settings, measurement sensors are available for five inputs (total flowrate, temperature and relative humidity of the inlet air; total flowrate and moisture of the wet granules) and six outputs (temperature in each segment). The air flows to individual segments are not measured. Note that, for a given batch, only one output measurement is available (namely, the temperature of the segment wherein the batch is being carried out).

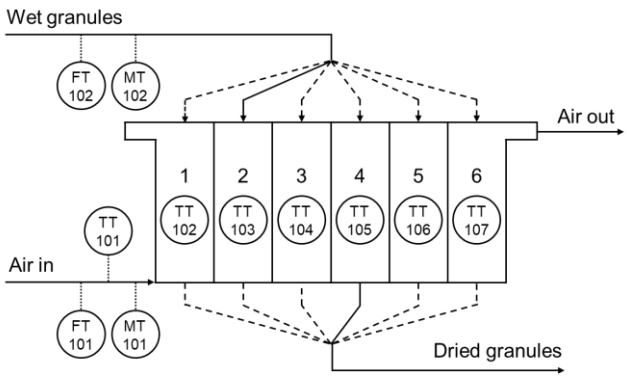

Fig. 2. Sketch of a six-segmented fluid bed dryer. The full lines represent the active streams in the current phase of the process, while the dashed lines are streams not currently under operation. FT, MT and TT represent flowrate, moisture/humidity and temperature transmitters. The air distributor plate is not shown.

We simulate 100 NOC batches, corresponding to a total of about 6 h of dryer operation. The duration of each batch is fixed and equal to 800 s. To mimic inter-batch variability under NOC, all the five measured inputs are varied according to smoothed pseudo-random binary sequences (PRBSs) of maximum amplitude ± 0.5% around the set-point. Also the split ratios of the air among the six segments follow a smoothed PRBS pattern of maximum amplitude ±1% around their nominal values. Thus, the air is not distributed evenly between the segments (not even under NOC), a situation that may arise in practice due to the distributor plate design.
We generate data for 3 faulty batches involving disturbances in the air flow to the segments, something which could be caused by partial blockage of the distributor plate. All abnormal batches start from NOC; at t = 300 s, for a given faulty batch we introduce a step (–5% for Fault #1, –10% for Fault #2) or a ramp (–0.025%/s for Fault #3) decrease in the air flow to the relevant segment. The segment air flow changes are simulated by changing the air split ratios. For each fault, Fig. 3 compares the inlet air flow to the segment to the flow variability (expressed as 99% confidence limits) induced by the application of the smoothed PRBSs on both the total inlet air inflow and the split ratios.
The input and output measurements are affected by white noise with standard deviations of 1% of the set-point for the inputs, and 0.05 °C for the outputs.

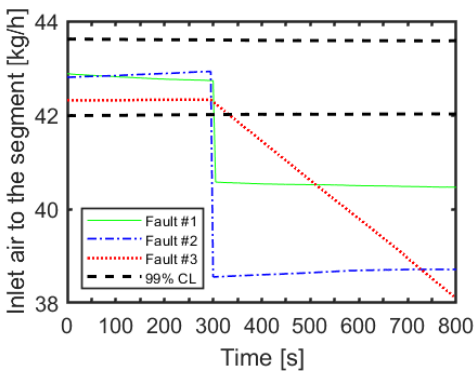

Fig. 3. Profiles of segment inlet air flow (unmeasured) for the three faulty batches.

4.2 Implementation of the hybrid monitoring framework

The hybrid system of Fig. 1 is implemented to monitor the batches operated in the dryer segments. The measured inputs **u** and outputs **y** for a batch are summarized in Table 1 (Variables #1-6).
The EKF available within the gPROMS platform is employed as the state estimator in the knowledge-driven block. Process-model mismatch arises from the fact that the EKF is not aware of the disturbances in the split ratios. The update step of the EKF is performed every 5 s. To improve the EKF robustness, we filter its inputs with a moving average approach. The state estimator is initialized with the true initial states $\mathbf{x}_0$ under NOC and with a null $\mathbf{P}_0$ matrix, because at the beginning of the process the dryer is empty and there is no uncertainty on this condition. The only non-null entries of the **Q** matrix are the diagonal elements corresponding to the vapor phase enthalpy of each segment, which are set equal to 1. **R** is a diagonal matrix, whose i^{th} element is the variance of the sensor noise for measurement y_i.
We apply the EKF to each of the 100 NOC batches to reconstruct the differential states and the algebraic variables. We arrange dataset **Z** as in (16), thus augmenting the set of 5 input measurements **u** and 1 output measurement **y** (Table 1, Variables #1-6) with 9 estimated variables $\tilde{\mathbf{x}}$ (Table 1, Variables #7-15) selected to provide additional information on the unmeasured phenomena occurring in the dryer.
We divide each batch of 800 s into time intervals of 10 s, and construct a separate dynamic PCA model at each of the corresponding 81 time points. By trial and error, we select a number of lags l = 4 (cf. (15)), resulting in a [100×75] $\mathbf{Z}_{dyn}(t_k)$ matrix for each time point. We retain 10 PCs in each of the 81 PCA models, with the explained variance ranging from 75% to 85%. Confidence limits for SPE, T^2 and the contributions are derived as discussed in Section 2.
The hybrid monitoring system can be implemented for real-time applications as the EKF, i.e., the most demanding component of the framework, requires a computational time smaller than the sampling time.

4.3 Fault detection and diagnosis

The hybrid model proves capable of detecting all faults (Fig. 4), with the first out-of-control signal always coming from the T^2 chart. The larger step decrease (–10%, Fault #2) in the air flow to a segment is detected earlier than the ramp fault

(Fault #3). The hardest fault to detect is the smaller step (–5%, Fault #1). The SPE does increase sharply at the beginning of the discharging phase (t = 750 s), but the faults are already detected well before that on the basis of the T^2 criterion.

Table 1. List of variables included in the augmented data matrix of the hybrid monitoring model

#	Variable	Unit	Variable type
1	Total flowrate of inlet air to the dryer	kg/h	Input ($\in \mathbf{u}$)
2	Relative humidity of inlet air to the dryer	%	Input ($\in \mathbf{u}$)
3	Temperature of inlet air to the dryer	°C	Input ($\in \mathbf{u}$)
4	Total flowrate of inlet granules	kg/h	Input ($\in \mathbf{u}$)
5	Moisture of inlet granules	kg/kg	Input ($\in \mathbf{u}$)
6	Segment temperature	°C	Output ($\in \mathbf{y}$)
7	Heat loss rate	J/s	Estimated ($\in \tilde{\mathbf{x}}$)
8	Mass of air in the segment	kg	Estimated ($\in \tilde{\mathbf{x}}$)
9	Mass of granules in the segment	kg	Estimated ($\in \tilde{\mathbf{x}}$)
10	Temperature of air in the segment	°C	Estimated ($\in \tilde{\mathbf{x}}$)
11	Temperature of granules in the segment	°C	Estimated ($\in \tilde{\mathbf{x}}$)
12	Drying rate	kg/s	Estimated ($\in \tilde{\mathbf{x}}$)
13	Moisture of granules in the segment	kg/kg	Estimated ($\in \tilde{\mathbf{x}}$)
14	Absolute humidity of air in the segment	g/kg	Estimated ($\in \tilde{\mathbf{x}}$)
15	Relative humidity of air in the segment	%	Estimated ($\in \tilde{\mathbf{x}}$)

Contributions to T^2 shortly after fault detection are similar for all faults, and Fig. 5 shows an example for Fault #2. Note that most of the out-of-limit contributions to T^2 result not from measurements (green bars), but from estimated values (red bars). Fig. 5 shows that, for the batch under investigation, the segment temperature is smaller than normal, both for the raw measurement (Variable #6) and for its value as reconstructed by the EKF (Variable #10). In addition, the drying rate (Variable #12) is smaller than normal. Taken together, these two results suggest than the fault is probably due to a reduced energy exchange (low temperature) in the segment, which is causing a reduction of the drying rate. Since the source of energy for the process is the total hot air feed, one may diagnose the fault as a problem in the air feed received by the segment. This diagnosis is corroborated by the fact that the relative air humidity in the segment (Variable #15) is abnormally high despite the lower drying rate (Variable #12). This indicates that the flux of water being vaporized is picked up by a lower air flow.

To compare the monitoring performance of the proposed hybrid system to the one of a standard PCA approach, we also performed a PCA on a reduced dataset including sensor measurements only (Variables #1-6). Results are not shown for lack of space, but they nevertheless deserve discussion. Though the faults are still detected (with minor delay), the contributions can only point to the measured temperature in the segment (Variable #6) as responsible for the fault. Without the additional information generated by the EKF, diagnosing the fault is much harder and further investigation would therefore be required. Note that the multivariate analysis introduced by the PCA is essential for fault detection. Fig. 6 shows univariate charts for three of the five variables exhibiting out-of-control contributions in Fig. 5; the corresponding charts for the other two variables are very similar. We note that Fault #1 does not result in significant deviations from the confidence limits established under NOC. Faults #2 and #3 do result in some deviations, but their magnitude is very small: the strongest deviations, those in the measured temperature (Variable #6), are only ~0.2°C for Fault #2 and ~0.5°C for Fault #3.

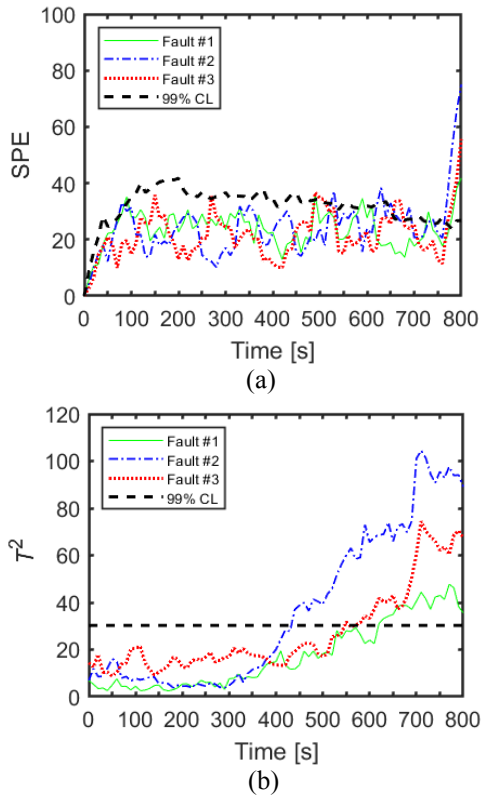

Fig. 4. Fault detection: (*a*) SPE monitoring chart and (*b*) T^2 monitoring chart for the three fault scenarios.

5. CONCLUSIONS

We presented a novel hybrid monitoring framework for process monitoring. Fault detection and diagnosis were performed through a latent-variable model built on a dataset comprising both process measurements and variables obtained in real-time through a state estimator. We tested the methodology on a detailed mechanistic model of a segmented fluid bed dryer. Even though only a single output measurement was available, the faults were detected promptly, and contribution plots demonstrated powerful diagnostic capabilities. The hybrid monitoring model

performed better than standalone data-driven and knowledge-driven monitoring approaches.

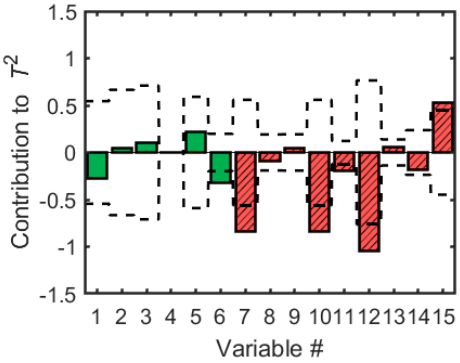

Fig. 5. Contributions to T^2 few instants after fault detection for Fault #2. Variables are numbered as in Table 1. Green bars refer to measured variables, red bars to estimated variables. Confidence limits at 99% for NOC are shown as black dashed lines.

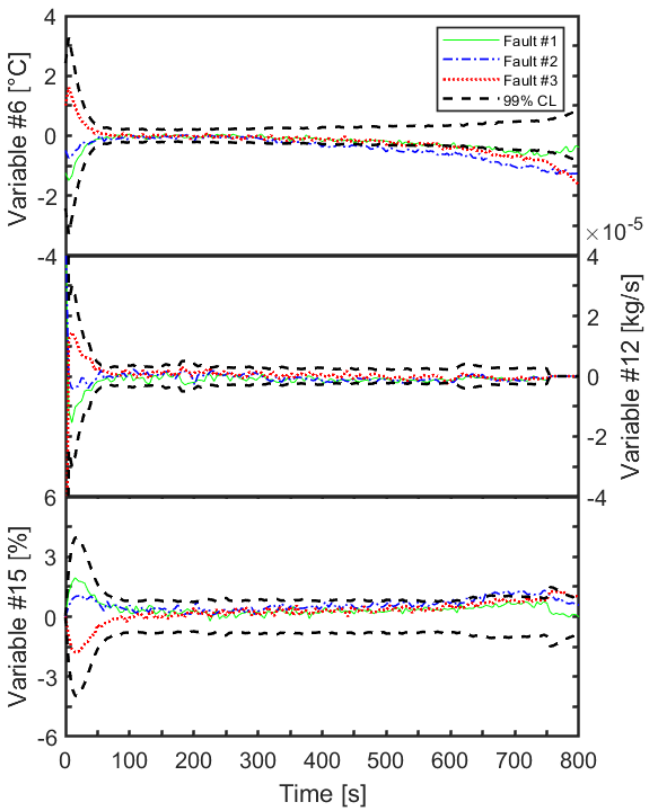

Fig. 6. Mean-centered univariate monitoring charts of selected variables displaying high contributions in Fig. 5. Variables are numbered as in Table 1.

ACKNOWLEDGEMENTS

We gratefully acknowledge the 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.

REFERENCES

Bonvin, D., Georgakis, C., Pantelides, C. C., Barolo, M., Grover, M. A., Rodrigues, D., Schneider, R. and Dochain, D. (2016). Linking Models and Experiments. *Industrial and Engineering Chemistry Research*, 55(25), pp. 6891–6903.

Burgschweiger, J. and Tsotsas, E. (2002). Experimental investigation and modelling of continuous fluidized bed drying under steady-state and dynamic conditions. *Chemical Engineering Science*, 57(24), pp. 5021–5038.

Camacho, J., Picó, J. and Ferrer, A. (2008). Bilinear modelling of batch processes. Part I: Theoretical discussion. *Journal of Chemometrics*, 22, pp. 299–308.

Cheng, Y.S., Mongkhonsi, T. and Kershenbaum, L.S. (1997). Sequential estimation for nonlinear differential and algebraic systems – Theoretical development and application. Computers and chemical engineering, 21(9), pp.1051-1067.Jackson, J. E. (1991). *A user's guide to principal components*. John Wiley & Sons, New York (U.S.A.).

Jiang, Q., Yan, X. and Huang, B. (2019). Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes. *Industrial & Engineering Chemistry Research*, 58(29), pp. 12899–12912.

Ku, W., Storer, R. H. and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30(1), pp. 179–196.

Miller, P., Swanson, R. E. and Heckler, C. E. (1998). Contribution plots: a missing link in multivariate quality control. *Applied Mathematics and Computer Science*, 8(4), pp. 775–792.

Mohd, J., Hoang, N. H., Hussain, M. A. and Dochain, D. (2015). Review and classification of recent observers applied in chemical process systems. *Computers and Chemical Engineering*, 76, pp. 27–41.

Nomikos, P. and MacGregor, J. F. (1995). Multivariate Processes SPC Charts for Monitoring. *Technometrics*, 37(1), pp. 41–59.

Pantelides, C. C. and Renfro, J. G. (2013). The online use of first-principles models in process operations: Review, current status and future needs. *Computers and Chemical Engineering*, 51, pp. 136–148.

Process Systems Enterprise Ltd. (2020). gPROMS FormulatedProducts. https://www.psenterprise.com/products/gproms/formulatedproducts.

Qin, S. J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2), pp. 220–234.

Ray, W. H. (1981). *Advanced process control*. McGraw-Hill, New York (U.S.A.).

Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. N. (2003). A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27, pp. 293–311.

Westerhuis, J. A., Gurden, S. P. and Smilde, A. K. (2000). Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 51(1), pp. 95–114.