Univariate model-based deadband alarm design for nonlinear processes

Aditya Tulsyan* and R. Bhushan Gopaluni*

Department of Chemical and Biological Engineering, The University of British Columbia, Vancouver, BC, Canada.

E-mail: tulsyan@ualberta.ca; bhushan.gopaluni@ubc.ca

Abstract

Alarm design is an essential industrial problem with significant implications for safety and performance. Standard alarm design algorithms are based on the assumption that the data are uncorrelated and stationary. In this article, we relax these assumptions and develop a novel approach to design alarms for processes modeled as a stochastic nonlinear time-series model. In particular, we develop an algorithm to design deadband alarms by minimizing the false and missed alarm rates. The resulting algorithm is illustrated through extensive simulations on a reactor system.

Introduction

Modern day industrial plants are highly interconnected and complex due to the increasing demands on performance, safety, environment and profitability standards. As a result, they are also highly automated and fitted with hundreds of thousands of sensors that continuously monitor a similar number of process variables.[1,2] The performance and safety of industrial processes depend critically on processing the information from these sensors to raise various

safety and performance alarms. These alarms have to be adequately designed and commissioned. Given the volume of alarms that need to be designed and tuned, often alarms are used with their default settings without a proper systematic approach to design. This leads to the plant operators being inundated with far more alarms than what is deemed safe by various international industrial standards. It is therefore vital to efficiently design alarms, process alarm information and presents it to an operator to take appropriate and timely actions.

The standard industry practice is to design sensitive alarms that bring to the attention of operators every minor safety or performance violation. This approach to designing alarms may result in operators being flooded with a large number of alarms. In fact, operators sometimes completely ignore certain recurring, but uncritical alarms. This can lead to complacency on the part of operators and the management. There are numerous historical examples, where a large number of alarms, in conjunction with complacent behavior of operators have resulted in catastrophic events with significant human and material losses. For instance, the Deepwater Horizon oil spill in 2010,[3] the Buncefield fire in 2005,[4] the BP Texas City refinery explosion in 2005[5] have been attributed partly to poor design and management of process alarms and complacent behavior of operators. According to the Abnormal Situation Management Consortium (ASM), process plants are losing billions of dollars a year due to poor management of abnormal situations resulting from the inefficient design of alarms.[6] According to ISA[1] 18.2 and EEMUA[2] 191 v.2 industrial standards, an operator, should receive no more than six alarms per hour during normal operation of a plant; however in practice, in most industrial plants, operators constantly receive a large number of alarms that easily exceed these standards, and that is mostly false or nuisance in nature (chattering, redundant, etc.). According to EEMUA, the average number of alarms in the oil/gas industry is thirty-six alarms per hour per operator.

Despite the importance of designing alarms, there is minimal literature that provides

[1]International Society for Automation

[2]The Engineering Equipment and Materials Users Association

rigorous and systematic approaches to design alarms. It is only relatively recently that more rigorous mathematical and data-based approaches have been used for designing alarms.[7–10] Alarm design methods can be broadly classified into those that are based purely on historical data and those that are based on models.[11] Data-based methods rely on assumptions such as independence and stationarity of process data;[12–15] however, the corresponding design algorithms are often easy to implement on industrial control systems. Further, these methods do not require explicit process models. Although these methods perform reliably under certain operating conditions, the underlying assumptions are often difficult to satisfy in closed-loop nonlinear industrial systems. Model-based approaches can potentially provide superior performing algorithms for designing alarms;[7,16] however, they require high-fidelity models that are often difficult to build for large-scale processes. All rigorous alarm design algorithms require a probability density function of the data under normal and faulty conditions. The data and model-based algorithms differ in how these density functions are constructed. In data-based algorithms, the density functions are obtained from historical data, and in model-based algorithms, they are estimated using a process model.

There are several approaches to triggering an alarm. The most straightforward approach is the limit-checking method. This approach triggers an alarm when a process variable exceeds a pre-defined alarm threshold.[7,17] Although simple to implement, the process noise and process disturbances can lead to several unnecessary nuisance alarms. The nuisance alarms can be reduced by using other alarm-triggering approaches, such as delay-timers, generalized delay-timers, and deadbands. In an n-sample delay timer, an alarm is triggered if the past consecutive n samples of a process variable are above or below a pre-defined alarm threshold.[14,16,18,19] Similarly, for deadbands, an alarm is triggered once the process variable crosses a limit, but it is turned off only when it leaves a pre-defined deadband.[19–22] In industry, the delay-timer and deadband alarms are both widely used.

We develop a model-based deadband alarm design algorithm for a nonlinear stochastic process. The authors had previously considered this problem in[18] in the context of

delay-timer alarm configuration. This article extends the problem in[18] to deadband alarm configuration, which is far more commonly used in industry. The design parameters for these algorithms are the optimal upper and lower thresholds for turning "on" and turning "off" the deadband alarm. This is a challenging problem due to the non-stationary, non-Gaussian, and correlated nature of the process variables. Furthermore, the process can exhibit complex nonlinear behavior. The proposed approach tackles these challenges by proposing new definitions of false and missed alarm rates that are valid for nonlinear, non-stationary and non-Gaussian processes. A recursive approach to calculate the proposed false and missed alarms for dead-band alarm configuration is also provided. The effectiveness of the algorithm is illustrated on a continuous stirred tank reactor (CSTR) system. To the authors' best knowledge, this is the first method to design deadband alarms for nonlinear and non-stationary processes. None of those above challenges have been addressed previously in the alarm design community. It is noteworthy to highlight that in a recent study;[21] the authors proposed an approach for designing deadband alarms for non identically and independently distributed (non-IID) processes. While the article[21] consider a similar problem, there are several key differences with the proposed method: (a) first, the method[21] assesses whether it is suitable for deadband alarms to remove nuisance alarms from a system, and if the answer is affirmative, it designs the deadband alarm to reduce the nuisance alarms. In contrast, our method assumes that there are no nuisance alarms, such that an alarm state is always indicative of abnormal process operation; (b) the article[21] considers a data-based alarm design; whereas, our method considers a model-based design; (c) while the method[21] considers deadband alarm design that achieves the best-weighted balance between the nuisance-alarm duration ratio and the pseudo-detection delay, our approach, in contrast, considers the alarm design that delivers the best-weighted balance between the false alarm rate and the missed alarm rate; and (d) while the method[21] assumes that the alarm trip threshold is known a priori (or can be estimated using historical data), and proposes a method to design the width of the deadband optimally, our method considers a simultaneous optimal design of the

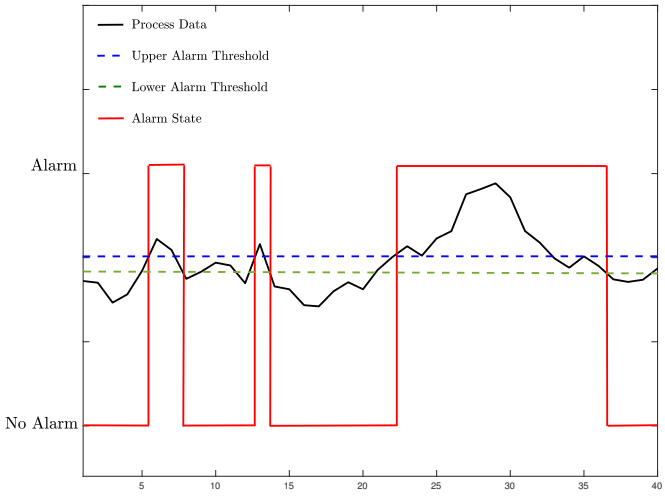

Figure 1: Process data (black curve) along with the deadband alarming strategy (upper threshold is indicated by broken-blue line and the lower threshold indicated by broken-green line). An alarm is raised if a process crosses the upper threshold and is cleared only if it crosses below the lower threshold. The alarm state is represented by the solid-red line.

alarm trip threshold and the width of the deadband. Finally, it is instructive to highlight that article[21] mentions that analyzing false and missed alarm rates for a deadband alarm configuration for non-IID signals is a challenging problem. We believe that the proposed method makes an effort in addressing that problem by introducing a recursive approach to compute the false and missed alarm rates for non-IID processes.

The remainder of the paper is organized as follows. First, a list of notation used in the article is provided. In the second section, the deadband alarm configuration and its alarming strategy are introduced. In the third section, the performance assessment of deadband alarms in terms of FAR and MAR are derived for a class of stochastic nonlinear processes. In the fourth section, the overall design of deadband alarms to reduce FAR and MAR are discussed in detail. In the fifth section, several case studies are presented to demonstrate the efficacy of the proposed deadband alarm design. Finally, some concluding remarks are provided.

Notation: We denote a random process variable measured at time $t \in \mathbb{R}_+$ with Z_t and

its particular realization by z_t. The probability density function (PDF) of Z_t is defined by $p(z_t)$ and it is written as $Z_t \sim p(z_t)$. The probability of $Z_t \leq a$ for some constant $a \in \mathbb{R}$ is denoted by $P(Z_t \leq a)$. In addition, we use the following compact notation to denote a sequence of random process measurements from time $t = 1$ to $t = N$, $Z_{1:t} = \{Z_1, \ldots, Z_N\}$ or $\{Z_t\}_{t\in\mathbb{N}}$ to denote a sequence with arbitrary length. $\mathbb{R}_+$, $\mathbb{P}$, and $\mathbb{N}$ respectively denote the sets of non-negative real numbers, the set of real numbers in the interval $[0, 1]$, and the set of integers.

Deadband Alarms

Deadband alarms are widely used in industry to remove spurious alarms that may arise due to process disturbances. A deadband alarm is triggered, when the process variable reaches a value greater (or lower) than a threshold and is turned off only if it is lower (or greater) than a different threshold. The unique feature of this alarm is that the thresholds for turning on and turning off an alarm are different (see Figure 1 for illustration). Now, for a given process state, $X_t \in \mathbb{R}$, the deadband (i.e., the upper and lower thresholds) divides the state-space into three disjoint sets (see Figure 1 for illustration), which are generically denoted here as follows

$$\mathcal{C}_t \equiv \{X_t : X_t < \underline{S}_x\}, \qquad \text{(Lower)} \tag{1a}$$

$$\mathcal{B}_t \equiv \{X_t : \underline{S}_x \leq X_t < \overline{S}_x\}, \qquad \text{(Deadband)} \tag{1b}$$

$$\mathcal{U}_t \equiv \{X_t : X_t \geq \overline{S}_x\}, \qquad \text{(Upper)}, \tag{1c}$$

where $\overline{S}_x \in \mathbb{R}$ and $\underline{S}_x \in \mathbb{R}$ are the upper and lower thresholds, $\mathcal{C}_t \subseteq \mathbb{R}$ is the 'lower alarm' region, $\mathcal{B}_t \subseteq \mathbb{R}$ is the deadband region, and $\mathcal{U}_t \subseteq \mathbb{R}$ is the 'upper alarm' region. Note that the alarming regions $\mathcal{C}_t$, $\mathcal{B}_t$ and $\mathcal{U}_t$ are defined solely on the basis of the alarm parameters $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$. It is straightforward to check that $\mathbb{R} = \mathcal{C}_t \cup \mathcal{B}_t \cup \mathcal{U}_t$ for all $t \in \mathbb{N}$.

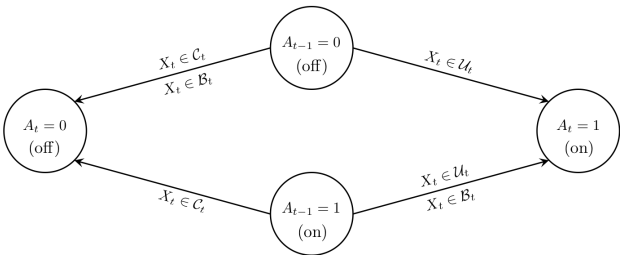

Figure 2: The alarming strategy for a deadband alarm.

Note that at any time $t \in \mathbb{N}$, the process X_t can only be in one of the three alarming regions. Now based on in which alarming region, $\{X_t\}_{t\in\mathbb{N}}$ is, the state of the alarm can be appropriately defined. If suppose $A_t \in \{0, 1\}$ denotes the state of the alarm at time $t \in \mathbb{N}$, such that $A_t = 0$ represents the "off" state and $A_t = 1$ represents the "on" state; then A_t can be defined in terms of A_{t-1}. For example, if $A_{t-1} = 0$, then $A_t = 1$ if and only if $X_t \in \mathcal{U}_t$ (i.e., $X_t \geq \overline{S}_x$). Similarly, if $A_{t-1} = 1$, then $A_t = 0$ if and only if $X_t \in \mathcal{C}_t$ (i.e., $X_t \leq \underline{S}_x$). The other possible scenarios that describe the complete dynamics of A_t are given in Figure 2 or Table 1. Observe that the "on" state of the alarm, or $A_t = 1$ can be mathematically

Table 1: The alarming strategy for a deadband alarm

A_{t-1}	A_t	Conditions
0	0	$X_t \in \mathcal{C}_t$ or $X_t \in \mathcal{B}_t$
0	1	$X_t \in \mathcal{U}_t$
1	0	$X_t \in \mathcal{C}_t$
1	1	$X_t \in \mathcal{U}_t$ or $X_t \in \mathcal{B}_t$

represented as follows

$$A_t = I_{\mathcal{U}_t}(X_t) + A_{t-1} I_{\mathcal{B}_t}(X_t), \tag{2}$$

where $I_\chi(X_t)$, for any arbitrary set, χ is an indicator function such that

$$I_\chi(X_t) \equiv \begin{cases} 1 & \text{if } X_t \in \chi \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

According to (2), if $A_{t-1} = 0$, then $A_t = 1$, if and only if $X_t \in \mathcal{U}_t$; and if $A_{t-1} = 1$, then $A_t = 1$ either when $X_t \in \mathcal{B}_t$ or $X_t \in \mathcal{U}_t$. Note that (2) only represents the "on" state of an alarm, i.e., for $A_t = 1$. Similarly, the "off" state of the alarm can be represented using the following equation

$$A_t = A_{t-1} I_{\mathcal{B}_t}(X_t) + A_{t-1}(1 - I_{\mathcal{C}_t}(X_t)). \tag{4}$$

Again, from (4), if $A_{t-1} = 0$ then $A_t = 0$ if $X_t \in \mathcal{C}_t$ or $X_t \in \mathcal{B}_t$, i.e., when $X_t < \overline{S}_x$. Similarly, if $A_{t-1} = 1$ then $A_t = 0$ if and only if $X_t \in \mathcal{C}_t$ (i.e., $X_t \leq \underline{S}_x$). Observe that the alarm triggering mechanisms in (2) and (4) are completely defined by the alarm parameters $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$. In the following sections, we assess the performance of a deadband alarm, and discuss how it can be used to optimally guide the design of $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$.

Performance Assessment

In this section, we discuss the performance assessment of a deadband alarm configuration based on the FAR and MAR. Let $\mathcal{R}_N$ and $\mathcal{R}_F$ represent the normal and faulty operating regions of a non-stationary process, $\{X_t\}_{t\in\mathbb{N}}$, respectively, such that $X_t \in \mathcal{R}_N$ for $t < t_F$ and $X_t \in \mathcal{R}_F$ for $t \geq t_F$, where t_F is the time of fault. Designing alarms for $\{X_t\}_{t\in\mathbb{N}}$ often results in two types of errors, namely the false and missed alarm error. A false alarm is an alarm that is triggered when the process variable $\{X_t\}_{t\in\mathbb{N}}$ is in the normal operating region. Similarly, a missed alarm is an alarm that occurs when $\{X_t\}_{t\in\mathbb{N}}$ is behaving abnormally, but no alarm is raised. While the false alarms reduce the trustworthiness of alarm systems, missed alarms severely degrade the designed functionality of alarm systems. In an alarm design problem, false alarm rate (FAR) and missed alarm rate (MAR) are two important measures of alarm performance. Given $\mathcal{R}_N$ and $\mathcal{R}_F$, FAR for a deadband alarm designed for

a non-stationary process, $\{X_t\}_{t\in\mathbb{N}}$, can be defined as follows

$$F_t(\underline{S}_x, \overline{S}_x) = \begin{cases} P(A_t = 1 | X_t \in \mathcal{R}_N), \text{ for } t = 0, \ldots, t_F - 1, \\ 0, \text{ for } t = t_F, \ldots, t_N, \end{cases} \tag{5}$$

where $F_t \in \mathbb{P}$ is the FAR at $t \in \mathbb{N}$, and $P(A_t = 1 | X_t \in \mathcal{R}_N)$ is the probability of observing an alarm given $X_t \in \mathcal{R}_N$. Note that $F_t = 0$ for $X_t \in \mathcal{R}_F$. For a given $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$, F_t is time-varying if $\{X_t\}_{t\in\mathbb{N}}$ is non-stationary, and assumes a constant value if $\{X_t\}_{t\in\mathbb{N}}$ is stationary. This is because for stationary processes, $P(A_t = 1 | X_t \in \mathcal{R}_N)$ is time-invariant. Similarly, MAR for deadband alarms can be defined as follows

$$M_t(\underline{S}_x, \overline{S}_x) = \begin{cases} 0, \text{ for } t = 0, \ldots, t_F - 1, \\ P(A_t = 0 | X_t \in \mathcal{R}_F), \text{ for } t = t_F, \ldots, t_N. \end{cases} \tag{6}$$

where $M_t \in \mathbb{P}$ is the MAR at $t \in \mathbb{N}$, and $P(A_t = 0 | X_t \in \mathcal{R}_F)$ is the probability of not observing an alarm when $X_t \in \mathcal{R}_F$. As (5), M_t in (6) is time-varying for non-stationary processes.

For non-stationary processes, calculating F_t and M_t is challenging as it requires access to $P(A_t = 1 | x_t \in \mathcal{R}_N)$ and $P(A_t = 0 | X_t \in \mathcal{R}_F)$ at each time point. The next two theorems provide a recursive approach to calcualte F_t and M_t.

Theorem 1. *The FAR for a deadband alarm designed for a non-stationary process, $\{X_t\}_{t\in\mathbb{N}}$, can be calculated as follows*

$$F_t = P(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N) + P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) F_{t-1}, \tag{7}$$

for all $t = 0, 1, \ldots, t_F - 1$, *where:*

$$P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) = \int_{\underline{S}_x}^{\overline{S}_x} p_N(x_t) dx_t; \tag{8a}$$

$$P(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N) = \int_{\overline{S}_x}^{+\infty} p_N(x_t) dx_t; \tag{8b}$$

and $p_N(\cdot)$ *is the PDF for* $\{X_t\}_{t\in\mathbb{N}}$, *when* $X_t \in \mathcal{R}_N$.

Proof: See Appendix A for the detailed proof.

Theorem 2. *The MAR for a deadband alarm designed for a non-stationary process,* $\{X_t\}_{t\in\mathbb{N}}$, *can be calculated as follows*

$$M_t = P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F) + P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F) M_{t-1}, \tag{9}$$

for all $t = t_F, t_F + 1, \ldots, t_N$, *where:*

$$P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F) = \int_{\underline{S}_x}^{\overline{S}_x} p_F(x_t) dx_t; \tag{10a}$$

$$P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F) = \int_{-\infty}^{\underline{S}_x} p_F(x_t) dx_t; \tag{10b}$$

and $p_F(\cdot)$ *is the PDF for* $\{X_t\}_{t\in\mathbb{N}}$, *when* $X_t \in \mathcal{R}_F$.

Proof: See Appendix B for the detailed proof.

Given $0 \leq F_0 \leq 1$ and $0 \leq S_0 \leq 1$, equations (7) and (9) provide a recursive approach to calculate F_t and M_t for all $t \in \mathbb{N}$. Note that F_t and M_t calculated using n (7) and (9), respectively, satisfy the conditions $0 \leq F_t \leq 1$ and $0 \leq M_t \leq 1$ for all $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$ and for all $t \in \mathbb{N}$ (see Appendix C for the detailed proof). This ensures that (7) and (9) are bounded for all $0 \leq F_0 \leq 1$, $0 \leq M_0 \leq 1$ and $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$.

Despite the recursive relations in (7) and (9), calculating integrals in (8a)-(8b) and (10a)-(10b) require the complete distribution of $\{X_t\}_{t\in\mathbb{N}}$ under normal and faulty operations, i.e.,

p_N and p_F. A traditional data-based approach uses process data to estimate p_N and p_F. This is done as follows. For a given sequence $x_{1:T_N}$, change-point detection algorithms, such as moving-average charts and generalized likelihood-ratio test [25,26] are used to estimate the time of fault, t_F, in $x_{1:T_N}$. Given an estimate of t_F, the sequence, $x_{1:T_N}$, can be readily decomposed into normal and faulty operations. Once decomposed, empirical density estimation methods, such as histogram can be used to construct the required PDFs. [14,23,24] For example, using histograms, p_F and p_N can be estimated as

$$p_N \approx \texttt{hist}(x_1, \ldots, x_{t_F-1}), \quad p_F \approx \texttt{hist}(x_{t_F}, \ldots, x_{t_N}), \tag{11}$$

where $\texttt{hist}(\cdot)$ is a histogram function. While the approximations in (11) are reliable, they are only valid when $\{X_t\}_{t\in\mathbb{N}}$ is an i.i.d. sequence. In fact, for non-i.i.d. or non-stationary processes, histogram-based approximations are no longer valid.

To calculate p_N and p_F for non-stationary processes, we propose a model-based approach. Like with data-based methods, given a sequence of $x_{1:T_N}$, we first estimate the time of fault, t_F. Next, given an estimate of t_F, we decompose the PDF of $\{X_t\}_{t\in\mathbb{N}}$ as follows

$$p(x_t) = \begin{cases} p_N(x_t), & \text{for } t = 0, \ldots, t_F - 1, \\ p_F(x_t), & \text{for } t = t_F, \ldots, t_N, \end{cases} \tag{12}$$

Given (12), p_N and p_F are approximated using a process model. Note that since the distribution of $\{X_t\}_{t\in\mathbb{N}}$ may be influenced by other states in the system, we consider an extended state vector, $\{\bar{X}_t\}_{t\in\mathbb{N}}$, that includes the state, X_t. It is assumed that $\{\bar{X}_t\}_{t\in\mathbb{N}}$ is defined by the nonlinear time-series model

Model 1. Stochastic nonlinear time-series model

$$\bar{X}_0 \sim p(\bar{x}_0), \tag{13a}$$

$$\bar{X}_{t+1} = f_t(\bar{X}_t, V_t), \tag{13b}$$

where $\bar{X}_t \in \mathbb{R}^n$ is the process state; $p(\bar{x}_0)$ is the PDF for the initial state $\bar{X}_0$; $f_t : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is an n-dimensional state transition function; and $V_t \in \mathbb{R}^n$ is the state noise sequence.

For mathematical convenience, exogenous process inputs have been dispensed with in Model 1; however, the results presented in the paper hold with inputs included.

Assumption 1. *$\{V_t\}_{t\in\mathbb{N}} \in \mathbb{R}^n$ is a sequence of independent random variables distributed according to $V_t \sim p(v_t)$, and is independent of the initial state $\bar{X}_0 \sim p(\bar{x}_0)$. Further, $p(\bar{x}_0)$ and $p(v_t)$ have known parametric probability density functions.*

Finally, given Model 1 and the sequence, $x_{1:T_N}$, we use particle methods to compute p_N and p_F at each sampling time. A particle approach to approximate the densities in (12) is discussed in Appendix D. Note that the densities computed in Appendix D are valid since Model 1 used with the particle method is a nonlinear, non-stationary, first-order time-series model.

Design of Deadband Alarms

FAR and MAR are undesirable as it is detrimental to the performance of an alarm system for it makes the alarm system less trustworthy and the system perceptible to catastrophic failures. As shown in the section on performance assessment, FAR and MAR of a deadband alarm configuration depend on alarm parameters $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$. In fact, both FAR and MAR can be reduced by appropriately choosing the alarm parameters $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$.

In this section, we propose an optimization framework to optimize deadband alarm design

by minimizing FAR and MAR. This is done as follows. Let $J : \mathbb{R}\times\mathbb{R} \to \mathbb{R}_+$ be a cost function associated with $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$, then we define J as follows

$$J_t(\underline{S}_x, \overline{S}_x) \equiv w_1 F_t^p(\underline{S}_x, \overline{S}_x) + w_2 M_t^q(\underline{S}_x, \overline{S}_x), \tag{14}$$

where: J_t is the cost function at time $t \in \mathbb{N}$; $p, q \geq 0$ are positive integers; and $w_1, w_2 \geq 0$ are non-negative weights. The cost function in (14) is defined as a weighted sum of FAR and MAR; however, this need not be the case, in general. For the cost function in (14), alarm parameters can be optimally designed by solving the following constrained optimization problem

$$\begin{aligned} (\underline{S}_x^*, \overline{S}_x^*) \in \underset{(\underline{S}_x, \overline{S}_x)\in\mathbb{R}\times\mathbb{R}}{\arg\min} \quad & J(F_t(\underline{S}_x, \overline{S}_x), M_t(\underline{S}_x, \overline{S}_x)), \\ \text{s.t.} \quad & G \begin{bmatrix} \underline{S}_x, \overline{S}_x \end{bmatrix}^{\mathrm{T}} = c, \\ & F_t(\underline{S}_x, \overline{S}_x) \leq \eta_1, \\ & M_t(\underline{S}_x, \overline{S}_x) \leq \eta_2, \end{aligned} \tag{15}$$

where $(\underline{S}_x^*, \overline{S}_x^*) \in \mathbb{R} \times \mathbb{R}$ is an optimal estimate of $(\underline{S}_x, \overline{S}_x)$ at $t \in \mathbb{N}$, $G \in \mathbb{R}^{m\times 2}$, $c \in \mathbb{R}^m$ and $\eta_1, \eta_2 \in \mathbb{P}$ are defined as per the process safety standards. Note that the formulation in (15) considers optimization over two parameters, namely $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$. Physically, solving (15) yields the optimal width and the optimal location for placing the deadband. Note that, in applications, where only the deadband width is critical to optimize, it is possible to reformulate (15) in terms of $\Delta S_x = \underline{S}_x - \overline{S}_x$. Note that with ΔS_x, (15) reduces to a single-variable optimization problem.

Note that since F_t and M_t used in (15) are time-varying, (15) needs to be solved at each sampling time to calculate optimal alarm parameters. Note that designing an alarm system at each sampling time is not only impractical, it may also lead to serious process upsets and unsafe process operations. To bypass the use of time-varying FAR and MAR in the alarm

design, we consider the following time-invariant definitions

$$F^E(\underline{S}_x, \overline{S}_x) \equiv \frac{1}{t_F} \sum_{t=0}^{t_F-1} F_t(\underline{S}_x, \overline{S}_x), \tag{16a}$$

$$M^E(\underline{S}_x, \overline{S}_x) \equiv \frac{1}{(t_N - t_F + 1)} \sum_{t=t_F}^{t_N} M_t(\underline{S}_x, \overline{S}_x), \tag{16b}$$

where F^E and M^E are expected values. With new definitions, an expected-case optimization-based alarm design can be computed by replacing F_t and M_t in (15) with F^E and M^E. Alternatively, we can also consider the following definitions

$$F^W(\underline{S}_x, \overline{S}_x) \equiv \max_{t \in \{0,\ldots,t_F-1\}} F_t(\underline{S}_x, \overline{S}_x), \tag{17a}$$

$$M^W(\underline{S}_x, \overline{S}_x) \equiv \max_{t \in \{t_F,\ldots,t_N\}} M_t(\underline{S}_x, \overline{S}_x), \tag{17b}$$

where F^W and M^W are the worst-case values for FAR and MAR, respectively. A worst-case alarm design is obtained by replacing F_t and M_t in (15) with their worst-case values. Finally, the procedure for designing expected-case and worst-case deadband alarms for a nonlinear process is outlined in Algorithm 1.

Case Study

In this section, we illustrate the efficacy of the deadband alarm design in Algorithm 1 on a simulated example. The process considered here is a non-isothermal continuous stirred tank reactor (CSTR) system. Consider a CSTR reaction system of volume $\gamma \in \mathbb{R}_+$, and with the following three parallel, irreversible, exothermic reactions[16]

$$\mathrm{A} \xrightarrow{\mathrm{k}_1} \mathrm{B}, \quad \mathrm{A} \xrightarrow{\mathrm{k}_2} \mathrm{U}, \quad \mathrm{A} \xrightarrow{\mathrm{k}_3} \mathrm{R}, \tag{20}$$

Algorithm 1 Proposed deadband alarm design

Input: Model 1 in (13a) and (13b); input signal $\{U_t\}_{t\in\mathbb{N}}$; time of fault t_F; final time t_N; total number of particles; M; and optimization parameter set, $\{w_1, w_2, G, c, \eta_1, \eta_2\}$

Output: Optimal deadband alarm parameters, $(\underline{S}_x^\star, \overline{S}_x^\star)$

Set $F_0 \leftarrow 0$ and $M_0 \leftarrow 0$

for t=0 to t_N **do**

  if $t < t_F$ **then**

    Compute an M particle approximation of the PDF $p_N(dx_t)$ using (54), such that

$$\tilde{p}_N(x_t)dx_t \leftarrow \frac{1}{M}\sum_{i=1}^{M}\delta_{X_t^i}(dx_t).$$

    Compute an M particle approximation of $P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N)$ and $P(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N)$ using (8a) and (8b), respectively, such that

$$\tilde{P}(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) \leftarrow \frac{1}{M}\sum_{j=1}^{M}\mathbf{1}_{\mathcal{B}_t}(X_t^j), \quad \tilde{P}(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N) \leftarrow \frac{1}{M}\sum_{j=1}^{M}\mathbf{1}_{\mathcal{U}_t}(X_t^j).$$

    Compute F_t using (7), such that

$$F_t \leftarrow \tilde{P}(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N) + \tilde{P}(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N)F_{t-1},$$

  end if

  if $t \geq t_F$ **then**

    Compute an M particle approximation of the PDF $p_F(dx_t)$ using (55), such that

$$\tilde{p}_F(x_t)dx_t \leftarrow \frac{1}{M}\sum_{i=1}^{M}\delta_{X_t^i}(dx_t).$$

    Compute an M particle approximation of $P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F)$ and $P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F)$ using (10a) and (10b), respectively, such that

$$\tilde{P}(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F) \leftarrow \frac{1}{M}\sum_{j=1}^{M}\mathbf{1}_{\mathcal{B}_t}(X_t^j), \quad \tilde{P}(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F) \leftarrow \frac{1}{M}\sum_{j=1}^{M}\mathbf{1}_{\mathcal{C}_t}(X_t^j).$$

    Compute M_t using (9), such that

$$M_t \leftarrow \tilde{P}(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F) + \tilde{P}(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F)M_{t-1},$$

  end if

end for

Compute (F^E, M^E) in (16a) and (16b) for expected-case alarm design or (F^W, M^W) in (17a) and (17b) for worst-case alarm design

Optimize (15) with the objective function in (14) to obtain $(\underline{S}_x^\star, \overline{S}_x^\star)$

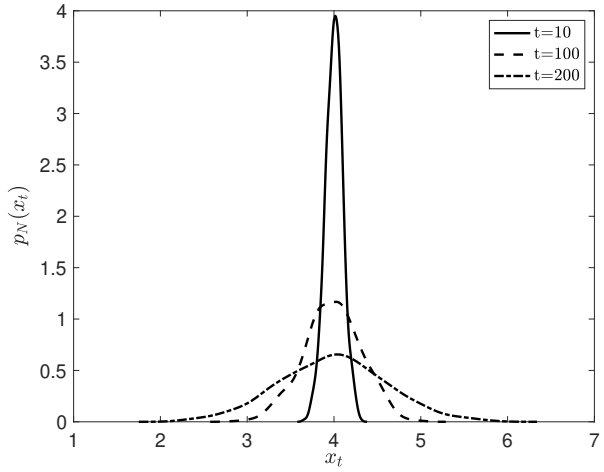

Figure 3: The PDFs for the process state under normal operations at time $t = 10$, $t = 100$ and $t = 200$ seconds.

Table 2: Nominal parameter values for the non-isothermal CSTR reaction system.

Parameter	**Value**	**Unit**
V	1	m^3
R	8.314	$\mathrm{kJ} \cdot \mathrm{kmol}^{-1} \cdot \mathrm{K}^{-1}$
ΔH_1	-5.0×10^4	$\mathrm{kJ} \cdot \mathrm{kmol}$
ΔH_2	-5.2×10^4	$\mathrm{kJ} \cdot \mathrm{kmol}$
ΔH_3	-5.4×10^4	kJkmol
k_{10}	3.0×10^6	h^{-1}
k_{20}	3.0×10^5	h^{-1}
k_{30}	3.0×10^5	h^{-1}
E_1	5.0×10^4	$\mathrm{kJ} \cdot \mathrm{kmol}^{-1}$
E_2	7.53×10^4	$\mathrm{kJ} \cdot \mathrm{kmol}^{-1}$
E_3	7.53×10^4	$\mathrm{kJ} \cdot \mathrm{kmol}^{-1}$
ρ	1000	$\mathrm{kg} \cdot \mathrm{m}^{-3}$
c_p	0.231	$\mathrm{kJ} \cdot \mathrm{kg}^{-1} \cdot \mathrm{K}^{-1}$

where A is the reactant, B is the desired product, and U and R are the undesired byproducts. The concentrations of A, B, U, and R are denoted by C_A, C_B, C_U, and C_R, respectively. The reactor is assembled with a jacket system to remove heat from the reactor. Given (20), the

concentrations of species and the reactor temperature can be modeled using the following

$$\dot{T}(t) = \frac{1}{\gamma}F(t)(T_{A0} - T(t)) + \sum_{i=1}^{3}\frac{(-\Delta H_i)}{\rho c_p}R_i(C_A(t), T(t)) + \frac{Q(t)}{\rho c_p \gamma}, \tag{21a}$$

$$\dot{C}_A(t) = \frac{1}{\gamma}F(t)(C_{A0} - C_A(t)) - \sum_{i=1}^{3}R_i(C_A(t), T(t)), \tag{21b}$$

$$\dot{C}_j(t) = -\gamma^{-1}F(t)C_j(t) + R_i(C_A(t), T(t)), \tag{21c}$$

where: $j = B, U, R$ and R_i for $i = 1, 2, 3$ are rate functions

$$R_i(C_A(t), T(t)) = k_{i0} \cdot exp(-E_i/RT(t))C_A(t); \tag{22}$$

ΔH_i, k_{i0}, and E_i for $i = 1, 2, 3$ denote the enthalpy, pre-exponential rate constant, and the activation energy for the three reactions in (20); T is the reactor temperature; c_p, ρ and R denote the heat capacity, fluid density, and the gas constant, respectively; and Q denote the rate of heat removal. The feed flow rate, denoted by F, is pure A of molar concentration C_{A0} and at temperature T_{A0}. The initial conditions in the CSTR are $T(0) = 300$ K, $C_A(0) = 4$ kmol·m^{-3}, and $C_B(0) = C_U(0) = C_R(0) = 0$ kmol·m^{-3}. The nominal values for all other model parameters are given in Table 2. Equations (21a) through (21c) are first discretized and represented in terms of Model 1 using the Euler's discretization method with a time-step 0.01 hr. For the sake of brevity, the discrete-time nonlinear time-series model representation of the network in (21a) through (21c) is not shown here, but is straightforward to derive. For the remainder of this section, (21a) through (21c) is represented by Model 1 with $\bar{X}_t \equiv [T(t)\ C_A(t)\ C_B(t)\ C_U(t)\ C_R(t)]^{\mathrm{T}}$ denoting the states and $U_t \equiv [F(t)\ Q(t)]^{\mathrm{T}}$ the inputs. To account for uncertainties in the parameters, we assume that the state noise in Model 1,

denoted by $V_t \sim \mathcal{N}(m_t, Q_t)$, is an additive multivariate Gaussian noise, where

$$m_t = \begin{cases} [0\ 0\ 0\ 0\ 0]^{\mathrm{T}}, \text{ for } t = 0, \ldots, t_F - 1, \\ [0.2\ 0.1\ 0.1\ 0.1\ 0.1]^{\mathrm{T}}, \text{ for } t = t_F, \ldots, t_N, \end{cases} \tag{23a}$$

$$Q_t = \begin{bmatrix} 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix}. \tag{23b}$$

The mean of the state noise in (23a) is assumed to be different in the normal and faulty operating conditions. It is further assumed that the initial state, $X_0 \in \mathbb{R}^5$ is imprecisely known, such that $X_0 \sim \mathcal{N}(m_{x_0}, Q_{x_0})$, where

$$m_{x_0} = \begin{bmatrix} 300 \\ 4 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ Q_{x_0} = \begin{bmatrix} 0.01 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{24}$$

Observe that the initial state distribution is defined independent of the state noise distribution (see Assumption 1). Note that even if X_0 and $\{V_t\}_{t\in\mathbb{N}}$ are Gaussian random variables, the distribution for $\{X_t\}_{t\in\mathbb{N}}$ is non-Gaussian and non-stationary. In fact, the non-stationary and nonlinear behavior of $\{X_t\}_{t\in\mathbb{N}}$ under normal process operations are illustrated in Figure 3.

In this section, we seek a univariate deadband alarm for the concentration of species A. Further, the alarm parameters $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$ are constrained to the space defined by the

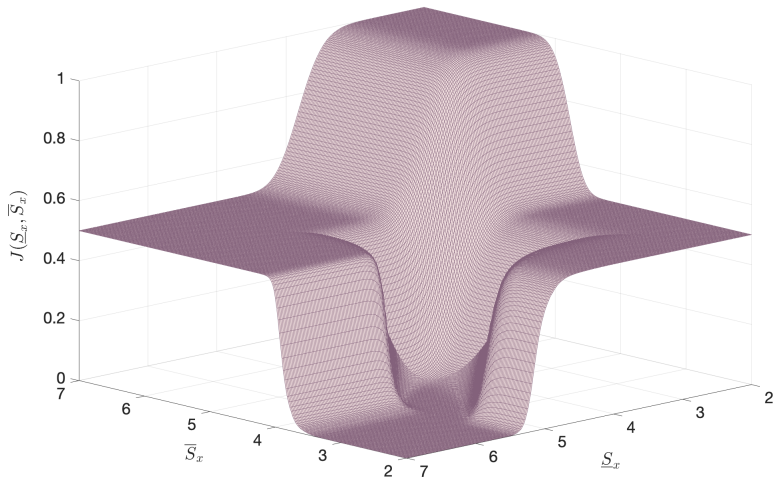

Figure 4: The objective function for the expected-case deadband alarm design as a function of $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$. The objective function is plotted for the unconstrained case.

following inequalities

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} \leq \begin{bmatrix} \underline{S}_x \\ \overline{S}_x \end{bmatrix} \leq \begin{bmatrix} 7 \\ 7 \end{bmatrix}, \tag{25a}$$

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \underline{S}_x \\ \overline{S}_x \end{bmatrix} \leq \begin{bmatrix} 1.5 \\ -0.3 \end{bmatrix}, \tag{25b}$$

where (25a) is the upper and lower bounds on $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$ and (25b) is the minimum and maximum width of the deadband. The inequalities in (25a)–(25b) are decided based on the process understanding and safety standards set for the process.

First, we consider an expected-case alarm design by assuming the following cost function

$$J(\underline{S}_x, \overline{S}_x) = 0.5[F^E(\underline{S}_x, \overline{S}_x)]^2 + 0.5[M^E(\underline{S}_x, \overline{S}_x)]^2. \tag{26}$$

Next, we minimize (26) given (25a) and (25b). Figure 4 gives a schematic of the cost function as alarm parameters are varied. From Figure 4 it is clear that the expected-case cost function is convex. The optimal alarm parameters, cost function, and the corresponding

Table 3: Results for the proposed expected-case and worst-case deadband alarm designs.

Design type	Optimal parameters	Optimal cost	Training		Cross-validation	
	$(\underline{S}_x^\star, \overline{S}_x^\star)$	$J(\underline{S}_x^\star, \overline{S}_x^\star)$	FAR (in %)	MAR (in %)	FAR (in %)	MAR (in %)
Expected-case	(3.70, 4.80)	5.98×10^{-5}	0.57	1.03	0.65	1.11
Worst-case	(3.53 4.02)	0.0061	13.01	18.4	13.37	18.76

FAR and MAR for the expected alarm design is shown in Table 3. It is easy to check that the optimal parameters $(\underline{S}_x^\star, \overline{S}_x^\star) = (3.70, 4.80)$ is a feasible solution. The expected FAR and MAR corresponding to the optimal parameters are 0.57% and 1.03%, respectively. Finally, Figure 5 (and Table 3) provides the expected FAR and MAR based on cross-validation. The results are based on 1000 Monte-Carlo simulations and are in close agreement with the results from the training set.

Next, for the worst-case alarm design, we replace the expected values for the FAR and MAR in (26) with their worst-case values. Again, solving (26) under (25a) and (25b) yields the worst-case alarm design. The optimal alarm parameters, cost function, and the worst-case FAR and MAR are tabulated in Table 3. Observe that the worst-case alarm design picks a different set of alarm parameters, compared to the expected-case design. Further, for the worst-case design, the worst-case FAR and MAR are 13.01% and 18.4%, which are much higher than their expected values calculated for the expected alarm design. Again, cross-validation results for the worst-case design are in close agreement with the training results, demonstrating the efficacy of the proposed deadband alarm design.

Conclusions

We have developed a deadband alarm design algorithm for nonlinear stochastic systems. The algorithm is valid for process variables that are correlated and non-stationary. We have derived mathematical expressions for recursive computation of false and missed alarm rates. These expressions depend on probability density functions, which are approximated using particle methods. The resulting approximate false and missed alarm rates are used to develop

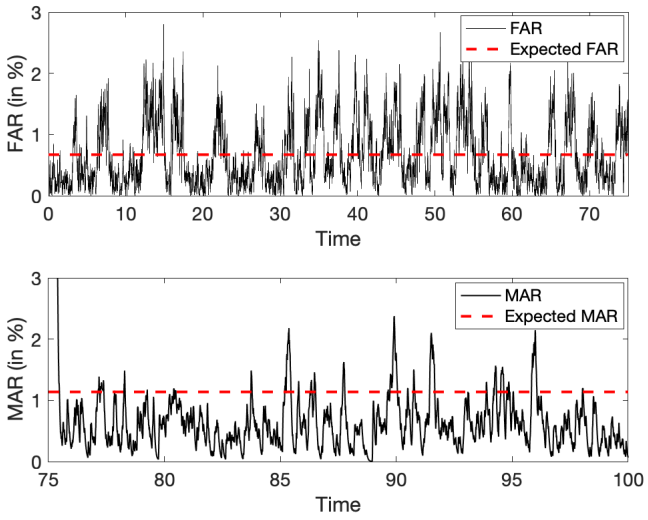

Figure 5: The FAR and MAR for the optimal expected-case alarm design in cross-validation (represented by black curves). The expected FAR and MAR are denoted by broken-red lines.

expected-case and worst-case algorithms. In particular, an alarm design algorithm based on a new optimization method is developed. These algorithms are effectively illustrated in a CSTR example.

Appendix A: Proof for Theorem 1

Using the Law of Marginalization, (5) can be written as

$$F_t = \sum_{A_{t-1}\in\{0,1\}} P(A_t = 1, A_{t-1}|X_t \in \mathcal{R}_N), \tag{27}$$

where the marginalization is over the alarm state A_{t-1}. Now using the law of conditional probability, (27) can be decomposed and written as follows

$$\begin{aligned} F_t &= P(A_t = 1|A_{t-1} = 0, X_t \in \mathcal{R}_N)P(A_{t-1} = 0) \\ &+ P(A_t = 1|A_{t-1} = 1, X_t \in \mathcal{R}_N)P(A_{t-1} = 1), \end{aligned} \tag{28}$$

where $P(A_t = 1|A_{t-1} = 0, X_t \in \mathcal{R}_N)$ is the probability that the alarm at t is "on", given that the alarm at $t-1$ is "off" and $X_t \in \mathcal{R}_N$, and $P(A_t = 1|A_{t-1} = 1, X_t \in \mathcal{R}_N)$ is the probability that the alarm at t is "on", given that the alarm at $t-1$ is "on" and $X_t \in \mathcal{R}_N$. Now to calculate F_t in (28), we need to first calculate the following density functions:

1. $P(A_t = 1|A_{t-1} = 0, X_t \in \mathcal{R}_N)$;
2. $P(A_t = 1|A_{t-1} = 1, X_t \in \mathcal{R}_N)$;
3. $P(A_{t-1} = 0)$;
4. $P(A_{t-1} = 1)$.

Next, we discuss the procedures to calculate the aforementioned density functions. From (2), if $A_{t-1} = 0$, then $A_t = 1$ if and only if $X_t \in \mathcal{U}_t$, such that

$$P(A_t = 1|A_{t-1} = 0, X_t \in \mathcal{R}_N) = P(X_t \in \mathcal{U}_t|X_t \in \mathcal{R}_N). \tag{29}$$

Similarly, if $A_{t-1} = 1$, then $A_t = 1$ if $X_t \in \mathcal{U}_t$ or $X_t \in \mathcal{B}_t$, such that the following relation holds

$$\begin{aligned} &P(A_t = 1|A_{t-1} = 1, X_t \in \mathcal{R}_N) = P(X_t \in \mathcal{U}_t|X_t \in \mathcal{R}_N) \\ &+ P(X_t \in \mathcal{B}_t|X_t \in \mathcal{R}_N), \end{aligned} \tag{30}$$

Now using the Law of Marginalization, $P(A_{t-1})$ is written as

$$P(A_{t-1}) = \sum_{i \in \{N,F\}} P(A_{t-1}, X_{t-1} \in \mathcal{R}_i), \tag{31a}$$

$$\begin{aligned} &= P(A_{t-1}|X_{t-1} \in \mathcal{R}_N)P(X_{t-1} \in \mathcal{R}_N) \\ &+ P(A_{t-1}|X_{t-1} \in \mathcal{R}_F)P(X_{t-1} \in \mathcal{R}_F). \end{aligned} \tag{31b}$$

Substituting $A_{t-1} = 0$ into (31b) yields

$$P(A_{t-1} = 0) = P(A_{t-1} = 0|X_{t-1} \in \mathcal{R}_N)P(X_{t-1} \in \mathcal{R}_N)$$
$$+ P(A_{t-1} = 0|X_{t-1} \in \mathcal{R}_F)P(X_{t-1} \in \mathcal{R}_F), \tag{32}$$

Now from (5), we have $P(A_{t-1} = 0|X_{t-1} \in \mathcal{R}_N) = 1 - F_{t-1}$, and from (6), we have $P(A_{t-1} = 0|X_{t-1} \in \mathcal{R}_F) = M_{t-1}$. Substituting these expressions into (32), we get

$$P(A_{t-1} = 0) = (1 - F_{t-1})P(X_{t-1} \in \mathcal{R}_N)$$
$$+ M_{t-1}P(X_{t-1} \in \mathcal{R}_F). \tag{33}$$

Similarly, substituting $A_{t-1} = 1$ into (31b) yields

$$P(A_{t-1} = 1) = P(A_{t-1} = 1|X_{t-1} \in \mathcal{R}_N)P(X_{t-1} \in \mathcal{R}_N)$$
$$+ P(A_{t-1} = 1|X_{t-1} \in \mathcal{R}_F)P(X_{t-1} \in \mathcal{R}_F). \tag{34}$$

Now from (5), we have $P(A_{t-1} = 1|X_{t-1} \in \mathcal{R}_N) = F_{t-1}$, and from (6), we have $P(A_{t-1} = 1|X_{t-1} \in \mathcal{R}_F) = 1 - M_{t-1}$. Substituting these expressions into (34), we get

$$P(A_{t-1} = 1) = F_{t-1}P(X_{t-1} \in \mathcal{R}_N)$$
$$+ (1 - M_{t-1})P(X_{t-1} \in \mathcal{R}_F). \tag{35}$$

Finally, substituting (29), (30), (33) and (35) into (31b), we get the following

$$\begin{aligned} F_t = {} & P(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N) \left[(1 - F_{t-1}) P(X_{t-1} \in \mathcal{R}_N) \right. \\ & \left. + M_{t-1} P(X_{t-1} \in \mathcal{R}_F) \right] + \left[P(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N) \right. \\ & \left. + P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) \right] \left[F_{t-1} P(X_{t-1} \in \mathcal{R}_N) \right. \\ & \left. + (1 - M_{t-1}) P(X_{t-1} \in \mathcal{R}_F) \right]. \end{aligned} \tag{36}$$

Now for $t < t_F$, where t_F is the time of fault, the process is in the normal operation region, such that $P(X_t \in \mathcal{R}_N) = 1$ and $P(X_t \in \mathcal{R}_F) = 0$ for all $t < t_F$. Substituting these expressions into (36) gives

$$\begin{aligned} F_t = {} & P(X_t \in \mathcal{U}_t | X_{t-1} \in \mathcal{R}_N) \\ & + P(X_t \in \mathcal{B}_t | X_{t-1} \in \mathcal{R}_N) F_{t-1}, \end{aligned} \tag{37}$$

which completes the proof.

Appendix B: Proof for Theorem 2

The proof for Theorem 2 is similar to Appendix A. Nevertheless, a separate proof is provided here for the sake of completeness. First, using the Law of Marginalization, (6) can be written as

$$M_t = \sum_{A_{t-1} \in \{0,1\}} P(A_t = 0, A_{t-1} | X_t \in \mathcal{R}_F), \tag{38}$$

where the marginalization is over the alarm state A_{t-1}. Now using the law of conditional probability, (38) can be decomposed and written as follows

$$M_t = \sum_{A_{t-1}\in\{0,1\}} P(A_t = 0|A_{t-1}, X_t \in \mathcal{R}_F)$$
$$\times P(A_{t-1}|X_t \in \mathcal{R}_F), \tag{39a}$$
$$= \sum_{A_{t-1}\in\{0,1\}} P(A_t = 0|A_{t-1}, X_t \in \mathcal{R}_F)P(A_{t-1}), \tag{39b}$$
$$M_t = P(A_t = 0|A_{t-1} = 0, X_t \in \mathcal{R}_F)P(A_{t-1} = 0)$$
$$+ P(A_t = 0|A_{t-1} = 1, X_t \in \mathcal{R}_F)P(A_{t-1} = 1), \tag{39c}$$

where $P(A_t = 0|A_{t-1} = 0, X_t \in \mathcal{R}_F)$ is the probability that the alarm at t is "off", given that the alarm at $t-1$ is "off" and $X_t \in \mathcal{R}_F$, and $P(A_t = 0|A_{t-1} = 1, X_t \in \mathcal{R}_F)$ is the probability that the alarm at t is "off", given that the alarm at $t-1$ is "on" and $X_t \in \mathcal{R}_F$. Now to calculate M_t in (39c), we need to first calculate the following density functions:

1. $P(A_t = 0|A_{t-1} = 0, X_t \in \mathcal{R}_F)$;
2. $P(A_t = 0|A_{t-1} = 1, X_t \in \mathcal{R}_F)$;
3. $P(A_{t-1} = 0)$;
4. $P(A_{t-1} = 1)$.

Next, we discuss the procedures to calculate the aforementioned density functions. From (4), if $A_{t-1} = 0$, then $A_t = 0$ if $X_t \in \mathcal{B}_t$ or $X_t \in \mathcal{C}_t$, such that

$$P(A_t = 0|A_{t-1} = 0, X_t \in \mathcal{R}_F) = P(X_t \in \mathcal{B}_t|X_{t-1} \in \mathcal{R}_F)$$
$$+ P(X_t \in \mathcal{C}_t|X_{t-1} \in \mathcal{R}_F), \tag{40}$$

where

$$P(X_t \in \mathcal{B}_t | X_{t-1} \in \mathcal{R}_F) = \int_{\underline{S}_x}^{\overline{S}_x} p_F(x_t) dx_t, \tag{41a}$$

$$P(X_t \in \mathcal{C}_t | X_{t-1} \in \mathcal{R}_F) = \int_{-\infty}^{\underline{S}_x} p_F(x_t) dx_t. \tag{41b}$$

Similarly, if $A_{t-1} = 1$ then $A_t = 0$ if and only if $X_t \in \mathcal{C}_t$, such that the following relation holds

$$P(A_t = 0 | A_{t-1} = 1, X_t \in \mathcal{R}_F) = P(X_t \in \mathcal{C}_t | X_{t-1} \in \mathcal{R}_F). \tag{42}$$

Finally, the densities $P(A_{t-1} = 0)$ and $P(A_{t-1} = 1)$ in (39c), are given by (33) and (35), respectively. Therefore, substituting (40), (42), (33) and (35) into (39c), we get

$$\begin{aligned} M_t = & [P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F) + P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F)] \\ & \times [(1 - F_{t-1}) P(X_{t-1} \in \mathcal{R}_N) + M_{t-1} P(X_{t-1} \in \mathcal{R}_F)] \\ & + P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F)[F_{t-1} P(X_{t-1} \in \mathcal{R}_N) \\ & + (1 - M_{t-1}) P(X_{t-1} \in \mathcal{R}_F)]. \end{aligned} \tag{43}$$

Now for $t \geq t_F$, where t_F is the time of fault, the process is in the faulty operation region, such that $P(X_t \in \mathcal{R}_N) = 0$ and $P(X_t \in \mathcal{R}_F) = 1$ for all $t \geq t_F$. Substituting these expressions into (43) gives

$$M_t = P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_F) + P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_F) M_{t-1}, \tag{44}$$

which completes the proof.

Appendix C: Bounds on FAR and MAR

Theorem 3. *The F_t and M_t in Theorems 1 and 2, respectively, are such that they satisfy the following inequalities*

$$0 \leq F_t \leq 1, \tag{45a}$$

$$0 \leq M_t \leq 1, \tag{45b}$$

for all $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$ and for all $t \in \{0, 1, \ldots, t_N\}$.

Proof: First, we provide a proof for (45a). First observe that (45a) is trivially satisfied for $t = t_F, \ldots, t_N$ from the definition of FAR in (5). Now for a process $X_t \in \mathbb{R}$, for each $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$, the following equality holds

$$\begin{aligned} 1 = P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_N) + P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) \\ + P(X_t \in \mathcal{U}_t | X_t \in \mathcal{R}_N). \end{aligned} \tag{46}$$

This is because $X_t \in \mathbb{R} = \mathcal{C}_t \cup \mathcal{B}_t \cup \mathcal{U}_t$ for all $(\underline{S}_x, \overline{S}_x) \in \mathbb{R} \times \mathbb{R}$, and for all $t = 0, \ldots, t_F - 1$. Now using (46), F_t expression in (7) can be alternatively written as follows

$$\begin{aligned} F_t &= 1 - P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_N) - P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) \\ &+ P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) F_{t-1}, \qquad (47a) \\ &= 1 - P(X_t \in \mathcal{C}_t | X_t \in \mathcal{R}_N) - P(X_t \in \mathcal{B}_t | X_t \in \mathcal{R}_N) \\ &\times (1 - F_{t-1}). \qquad (47b) \end{aligned}$$

For $t < t_F$, where t_F is the time of fault, the process is in the normal operation region, such that $P(X_t \in \mathcal{R}_N) = 1$ and $P(X_t \in \mathcal{R}_F) = 0$ for all $t < t_F$. Substituting these expressions

into (34) yields

$$P(A_{t-1}=1)=P(A_{t-1}=1|X_{t-1}\in\mathcal{R}_N)=F_{t-1}. \tag{48}$$

Substituting (48) into (47b) yields

$$\begin{aligned} F_t &= 1-P(X_t\in\mathcal{C}_t|X_t\in\mathcal{R}_N)-P(X_t\in\mathcal{B}_t|X_t\in\mathcal{R}_N) \\ &\times(1-P(A_{t-1}=1)), \qquad (49a) \\ &= 1-P(X_t\in\mathcal{C}_t|X_t\in\mathcal{R}_N)-P(X_t\in\mathcal{B}_t|X_t\in\mathcal{R}_N) \\ &\times P(A_{t-1}=0), \qquad (49b) \end{aligned}$$

where the last equality is from $P(A_t=1)+P(A_t=0)=1$. Clearly, by definition $F_t\geq 0$ (see (5)); and in (49b), $0\leq P(X_t\in\mathcal{C}_t|X_t\in\mathcal{R}_N)\leq 1$ and $0\leq P(X_t\in\mathcal{B}_t|X_t\in\mathcal{R}_N)P(A_{t-1}=0)\leq 1$. Substituting these relations into (49b)

$$\begin{aligned} F_t = 1 &- \underbrace{P(X_t\in\mathcal{C}_t|X_t\in\mathcal{R}_N)}_{\geq 0 \text{ and } \leq 1} \\ &- \underbrace{P(X_t\in\mathcal{B}_t|X_t\in\mathcal{R}_N)P(A_{t-1}=0)}_{\geq 0 \text{ and } \leq 1} \quad \leq 1, \qquad (50a) \end{aligned}$$

which completes the proof. The proof for (45b) is similar, and for the sake of brevity is not included here.

Appendix D: Particle Methods

In this section, we present an approximation of the PDFs $p_N(\cdot)$ and $p_F(\cdot)$. First note that $p_N(x_t)$ is a marginal density of the joint density function $p_N(x_t, x_{t-1})$, which is given by

$$p_N(x_t) = \int p_N(x_t, x_{t-1}) dx_{t-1}, \tag{51a}$$

$$= \int p_N(x_t|x_{t-1}) p_N(x_{t-1}) dx_{t-1}, \tag{51b}$$

where $p_N(x_t|x_{t-1})$ is the state transition density in Model 1 under normal operating conditions. Given $p_N(x_t|x_{t-1})$ and $p(x_0)$, it is possible to use (51b) to recursively estimate $p_N(x_t)$ at any time t; however, often the integral in (51b) is too complex to evaluate directly.[27–29] A particle method provides an approximation of the integral in (51b). The basic idea of a particle method is to generate a number of 'particles' of the state through simulations and propagate them through the state transition model.[30] We refer the reader to[31] for a tutorial on particle methods.

Let us assume that M random samples (also called particles) of $\{X_{t-1}^i\}_{i=1}^M$ distributed according to $p_N(x_{t-1})$ are available from a previous iteration. Then the distribution $p_N(dx_{t-1}) \equiv p_N(x_{t-1})dx_{t-1}$ can be approximated as

$$\tilde{p}_N(dx_{t-1}) = \frac{1}{M} \sum_{i=1}^{M} \delta_{X_{t-1}^i}(dx_{t-1}), \tag{52}$$

where $\tilde{p}_N(dx_{t-1})$ is an M-particle approximation of the corresponding distribution function, and $\delta_X(dx)$ is a Dirac delta measure centered at particle X. An approximation of the density function $p_N(x_t)$ can be obtained by substituting (52) in (51b),

$$\tilde{p}_N(x_t) = \int p_N(x_t|x_{t-1}) \frac{1}{M} \sum_{i=1}^{M} \delta_{X_{t-1}^i}(dx_{t-1}), \tag{53a}$$

$$= \frac{1}{M} \sum_{i=1}^{M} \int p(x_t|x_{t-1}) \delta_{X_{t-1}^i}(dx_{t-1}) = \frac{1}{M} \sum_{i=1}^{M} p(x_t|X_{t-1}^i). \tag{53b}$$

A set of M particles of $\{X_t^i\}_{i=1}^M$ can be generated by simply passing $\{X_{t-1}^i\}_{i=1}^M$ through $p_N(x_t|X_{t-1}^i)$ and using (53b). The new set of particles $\{X_t^i\}_{i=1}^M$ are approximately distributed according to $p_N(x_t)$, and can be represented as

$$\tilde{p}_N(x_t)dx_t = \frac{1}{M}\sum_{i=1}^{M}\delta_{X_t^i}(dx_t). \tag{54}$$

The PDF $p_F(x_t)$ can also be similarly approximated, such that

$$\tilde{p}_F(x_t)dx_t = \frac{1}{M}\sum_{i=1}^{M}\delta_{X_t^i}(dx_t), \tag{55}$$

represents an M article approximation of p_F.

References

(1) Tulsyan, A.; Garvin, C.; Ündey, C. Advances in industrial biopharmaceutical batch process monitoring: machine-learning methods for small data problems. *Biotechnology and Bioengineering* **2018**, *115*, 1915–1924.

(2) Tulsyan, A.; Garvin, C.; Ündey, C. Industrial batch process monitoring with limited data. *Journal of Process Control* **2019**, *In Press*.

(3) Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling. National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling. 2011.

(4) The Buncefield Incident 11 December 2005: The final report of the Major Incident Investigation Board. Buncefield Major Incident Investigation Board. 2008.

(5) Investigation report: refinery explosion and fire. US Chemical Safety and Hazard Investigation Board. 2007.

(6) Cochran, E.; Miller, C.; Bullemer, P. Abnormal situation management in petrochemical

plants: can a pilot's associate crack crude? In Proceedings of the IEEE 1996 National Aerospace and Electronics Conference. Dayton, USA, 1996.

(7) Ahnlund, J.; Bergquist, T.; Spaanenburg, L. Rule-based reduction of alarm signals in industrial control. *Journal of Intelligent and Fuzzy Systems* **2003**, *14*, 73–84.

(8) Brooks, R.; Thorpe, R.; Wilson, J. A new method for defining and managing process alarms and for correcting process operation when an alarm occurs. *Journal of Hazardous Materials* **2004**, *115*, 169–174.

(9) Rothenberg, D. *Alarm Management for Process Control: A Best-Practice Guide for Design, Implementation, and Use of Industrial Alarm Systems*; Momentum Press: New York, USA, 2009.

(10) Tulsyan, A.; Barton, P. I. Reachability-based fault detection method for uncertain chemical flow reactors. *IFAC-PapersOnLine* **2016**, *49*, 1–6.

(11) Bingyong, Y.; Zuohua, T.; Songjiao, S. A novel distributed approach to robust fault detection and identification. *International Journal of Electrical Power and Energy Systems* **2008**, *30*, 343–360.

(12) Izadi, I.; Shah, S.; Kondaveeti, S.; Chen, T. A framework for optimal design of alarm systems. In Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain, 2009.

(13) Izadi, I.; Shah, S.; David, S.; Chen, T. An introduction to alarm analysis and design. In Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain, 2009.

(14) Xu, J.; Wang, J.; Izadi, I.; Chen, T. Performance Assessment and Design for Univariate Alarm Systems Based on FAR, MAR, and AAD. *IEEE Transactions on Automation Science and Engineering* **2012**, *9*, 296–307.

(15) Kondaveeti, S. R.; Izadi, I.; Shah, S. L.; Shook, D. S.; Kadali, R.; Chen, T. Quantification of alarm chatter based on run length distributions. *Chemical Engineering Research and Design* **2013**, *91*, 2550–2558.

(16) Tulsyan, A.; Gopaluni, R. B. Robust model-based delay timer alarm for non-linear processes. In Proccedings of the American Control Conference. Boston, USA, 2016.

(17) Isermann, R. *Fault-diagnosis Systems: an Introduction From Fault Detection to Fault Tolerance*; Springer-Verlag: Berlin, Germany, 2006.

(18) Tulsyan, A.; Alrowaie, F.; Gopaluni, R. B. Design and Assessment of Delay Timer Alarm Systems for Nonlinear Chemical Processes. *AIChE Journal* **2017**, *1*, 77–90.

(19) Adnan, N. A.; Izadi, I.; Chen, T. On expected detection delays for alarm systems with deadbands and delay-timers. *Journal of Process Control* **2011**, *21*, 1318–1331.

(20) Hugo, A. Estimation of alarm deadbands. In Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain, 2009.

(21) Wang, Z.; Bai, X.; Wang, J.; Yang, Z. Indexing and designing deadbands for industrial alarm signals. *IEEE Transactions on Industrial Electronics* **2018**, *In Press*.

(22) Afzal, M. S.; Chen, T.; Bandehkhoda, A.; Izadi, I. Analysis and design of time-deadbands for univariate alarm systems. *Control Engineering Practice* **2018**, *71*, 96–107.

(23) Adnan, N. Performance Assessment and Systematic Design of Industrial Alarm Systems. Ph.D. thesis, Department of Electrical and Computer Engineering, University of Alberta, Canada, 2013.

(24) Adnan, N.; Izadi, I.; Chen, T. Computing detection delays in industrial alarm systems. In Proceedings of the American Control Conference. San Francisco, USA, 2011.

(25) Chen, J.; Gupta, A. K. *Parametric Statistical Change Point Analysis: with Applications to Genetics, Medicine, and Finance*; Springer Science & Business Media: Belin, Germany, 2011.

(26) Basseville, M.; Nikiforov, I. V. *Detection of Abrupt Changes: Theory and Application*; Prentice Hall: Englewood Cliffs, USA, 1993.

(27) Tulsyan, A.; Huang, B.; Gopaluni, R. B.; Forbes, J. F. Performance assessment, diagnosis, and optimal selection of non-linear state filters. *Journal of Process Control* **2014**, *24*, 460–478.

(28) Tulsyan, A.; Huang, B.; Gopaluni, R. B.; Forbes, J. F. Bayesian identification of non-linear state-space models: Part II-Error Analysis. *IFAC Proceedings Volumes* **2013**, *46*, 631–636.

(29) Tulsyan, A.; Khare, S.; Huang, B.; Gopaluni, B.; Forbes, F. A switching strategy for adaptive state estimation. *Signal Processing* **2018**, *143*, 371–380.

(30) Tulsyan, A.; Huang, B.; Gopaluni, R. B.; Forbes, J. F. On simultaneous on-line state and parameter estimation in non-linear state-space models. *Journal of Process Control* **2013**, *23*, 516–526.

(31) Tulsyan, A.; Gopaluni, R. B.; Khare, S. R. Particle filtering without tears: A primer for beginners. *Computers & Chemical Engineering* **2016**, *95*, 130–145.