Original Article

Journal of Time Series Analysis

First version received November 2011

Published online in Wiley Online Library: 2012

(wileyonlinelibrary.com) DOI: 10.1111/j.1467-9892.2012.00791.x

Extreme value analysis of optimal level-crossing prediction for linear Gaussian processes

Rodney A. Martin[a,*,†]

A novel approach of combining the practical appeal of Kalman filtering with the design of an optimal alarm system for the prediction of level-crossing events was introduced in earlier work. Here, the aim is to perform a more detailed extreme value analysis using the critical threshold that enables definition of the level-crossing event. It will be rigorously proven that the approximations and baseline methods previously used yield important intuitive conclusions about the impact of low measurement noise and high levels on improved capability of level-crossing predictors. Where possible, elegant closed-form solutions for a well-known alarm system metric in face of those limiting considerations are also provided.

Keywords: Alarm systems; approximation methods; kalman filtering; level-crossing problems; prediction methods.

1. INTRODUCTION

In earlier work (Martin, 2010), a novel state-space approach to the optimal alarm systems literature was introduced, which contributed to the Kalman filter-based fault detection literature from a different theoretical angle. Originally, it was shown by Svensson (1998) and Svensson *et al.* (1996) that an optimal alarm system can be constructed by finding relevant alarm system metrics as a function of a design parameter by way of an optimal alarm condition. These alarm system metrics are often derived from those that are used in receiver operating characteristic (ROC) curve analysis, enabling the design of tradeoffs between false alarm and missed detection probabilities. The optimal alarm condition is fundamentally an alarm region or decision boundary based on a likelihood ratio criterion via the Neyman–Pearson lemma (de Maré, 1980; Lindgren, 1985). This allows for the design of an optimal alarm system that will elicit the fewest possible false alarms for a fixed detection probability. This becomes important when considering the numerous applications that might benefit from an intelligent tradeoff between false alarms and missed detections. An extreme value analysis using the critical threshold that enables definition of the level-crossing event will be investigated, to secure limiting conditions related to certain approximations to the optimal level-crossing predictor and baseline methods for comparison that were also previously introduced (Martin, 2010).

There are several examples of level-crossing events, varying from a simple one-sided case to a more complicated two-sided case. The former one-sided case involves exceedances and/or upcrossings of a single level by a discrete-time process. The difference between exceedances and upcrossings is related to the number of time points spanned by the event. An exceedance is a one-dimensional level-crossing event, where some critical threshold level is exceeded by a process involving only a single time point. An upcrossing is a two-dimensional level-crossing event, such that some critical threshold level is exceeded by a process involving two adjacent time points. The process is below the threshold at the first time point and above the threshold at the second time point. One-sided exceedances and upcrossings have traditionally been studied in previous work and invoke ARMA(X) prediction methods (de Maré, 1980; Lindgren, 1985; Svensson *et al.*, 1996; Svensson, 1998). The two-sided case involves a level-crossing event that spans many time points, exceeding upper levels and fading below lower levels symmetric about the mean of the process. A variant of the latter more complicated two-sided case has been previously investigated in the context of Kalman filter-based fault detection algorithm development (Kerr, 1982). This case is more practically relevant when monitoring residuals that may be derived from the output of other machine learning algorithms or transformed parameters that relate to critical performance of a system or process. The two-sided case was also investigated in the earlier work (Martin, 2010) and serves as a precursor to this article. As such, it will be investigated here as well.

The prediction of such a level-crossing event is also very similar to what has been established as the state of the art for newly minted spacecraft engines, as previously studied by Butas *et al.* (2007); however there were no optimality measures in place. This provides additional practical motivation for investigating a level-crossing event that spans many time steps, moving beyond what has previously been studied in this vein. In general, the design of optimal alarm systems demonstrates practical potential to enhance reliability and support health management for space propulsion, civil aerospace applications, and other related fields. Because of the

[a]*NASA Ames Research Center*
**Correspondence to: Rodney A. Martin, NASA Ames Research Center, Mail stop 269-1, P.O. Box 1, Moffett Field, CA 94035-0001, USA.*
[†]*E-mail: rodney.martin@nasa.gov*

great costs, not to mention potential dangers associated with a false alarm due to evasive or extreme action taken as a result of false indications, there are great opportunities for cost savings/cost avoidance and enhancement of overall safety. Nonetheless, the intent of this article is to demonstrate the utility of optimal level-crossing prediction from a more theoretical perspective.

Optimal alarm regions cannot be expressed in closed form, and thus ROC curve analysis must be performed either empirically or with a computationally prohibitive generic Monte Carlo numerical integration routine where the integration region is given by storing precomputed points along the boundary of the alarm region (Svensson, 1998). As such, one of the aims of the previous study (Martin, 2010) was to investigate approximations for the design of an optimal alarm system. The resulting false alarm and missed detection probabilities that can be derived from ROC curve analysis were also compared to competing baseline methods and may provide some level of predictive capability, but have no provision for minimizing false alarms for the prediction of level-crossing events. When using approximations to the theoretically optimal predictor for a stationary linear Gaussian process, it was demonstrated that there was a negligible loss in overall accuracy at the advantage of greatly reduced computational complexity. The negligibility of the loss in accuracy was demonstrated by comparing approximations to the optimal level-crossing predictor to the baseline methods. The approximations clearly outperformed the baseline methods; however, it was also observed that caution needed to be exercised when using these approximations in the design of alarm systems when they are defined with small threshold values. The character of these approximations and competing baseline methods when considering extreme values is thus one of the key objectives of this study.

As such, this article presents an extreme value analysis of all approximations to the optimal level-crossing predictor, as well as competing baseline methods. These alarm systems can be used to bootstrap an investigation into intuitive observations in face of the relevant limiting considerations that may also translate into providing some insights about performance for typical critical threshold values. All of the limiting considerations described herein will be evaluated using the area under the ROC curve (AUC), for which closed-form relationships can be derived under certain technical conditions. The AUC is a performance metric that characterizes the ability of an alarm system to accurately predict the level-crossing event. More precisely, it quantifies the Mann-Whitney–Wilcoxon *U*-test statistic, which is equivalent to the probability of correctly ranking two randomly selected data points, one belonging to the level-crossing event class and the other not. The AUC has been deemed as a theoretically legitimate metric for model selection and algorithmic comparison (Bradley, 1997). Closed-form expressions for the AUC will be derived explicitly as a function of relevant system parameters where possible.

2. METHODOLOGY

2.1. Preliminaries

The investigation of all alarm systems will be conducted in the context of predicting a level-crossing event defined with a critical level, L, that is assumed to have a fixed, static value. The level is exceeded by some critical parameter than can be represented by a dynamic process and is often modelled as a zero-mean stationary linear dynamic system driven by Gaussian noise. The theory that follows is based on this standard representation of the optimal level-crossing problem. As such, it is an underlying assumption that measured or transformed data can be fitted well to a model represented by a linear dynamic system driven by Gaussian noise. The state-space formulation is shown in eqns (1)–(3), demonstrating propagation of the state, $\mathbf{x}_k \in \mathbb{R}^n$ which is corrupted by process noise $\mathbf{w}_k \in \mathbb{R}^n$. For convenience of presentation, it will be assumed that propagation of all state covariance matrices can reasonably well be approximated by their steady-state counterparts. This approximation, while it introduces error with regards to the probability of a level-crossing event at a specific point in time, is ostensibly negligible and will provide for a great computational advantage in the design of an alarm system. Instead of designing an alarm system for each time step, a single alarm system can be designed for all time steps. The approximation is based on the limiting statistics that are reached at steady-state, which greatly reduces the computational burden, as previously identified (Martin, 2010). As such, the solution of the steady-state Lyapunov function, $\mathbf{P}_{ss}^{L}$, suffices for evolution of the unconditional state covariance matrix. The output $y_k \in \mathbb{R}$ is univariate and is corrupted by measurement noise $v_k \in \mathbb{R}$.

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{w}_k, \tag{1}$$

$$y_k = \mathbf{C}\mathbf{x}_k + v_k, \tag{2}$$

$$\mathbf{P}_{ss}^{L} = \mathbf{A}\mathbf{P}_{ss}^{L}\mathbf{A}^{\top} + \mathbf{Q}, \tag{3}$$

where

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad \mathbf{Q} \succeq \mathbf{0}$$
$$v_k \sim \mathcal{N}(0, R), \quad R > 0.$$

A summary of the basic mathematical notation not defined elsewhere is provided in Table 1. There is great flexibility in constructing a mathematical representation for the level-crossing event, C_k; however, in this article the event of interest is shown in eqn (4). This level-crossing event represents at least one exceedance outside of the threshold envelope specified by $[-L,L]$ of the process y_k within the specified look-ahead prediction window, d.

Table 1. Summary of mathematical notation

Mathematical representation	Nomenclature
$\triangleq$	is defined as
$\setminus$	Set subtraction
$\succeq$	Positive semi-definite
$(\bullet)'$	Not (Set complement)
$\mathcal{I}$	Universe of all events
$(\bullet)^\top$	Transpose
$P(\bullet)$	Probability
$E[\bullet]$ or $\mu_\bullet$	Expected value
VC[$\bullet$]	Variance–covariance matrix
$\hat{\bullet}_{k+j\|k}$	$E[\bullet \mid y_0, \ldots, y_k]$ (conditional expectation)
$\Phi(\bullet)$	Univariate Gaussian cumulative distribution function
$\phi(\bullet)$	Univariate Gaussian probability density function
$\mathcal{N}(\mu, \Sigma)$	Multi-variate Gaussian distribution with mean μ and covariance Σ
$\mathcal{N}(\mathbf{x}; \mu, \Sigma)$	Multi-variate Gaussian distribution evaluated at $\mathbf{x}$ with mean μ and covariance Σ

$$C_k \triangleq \bigcup_{j=1}^{d} S_{k+j} = \bigcup_{j=1}^{d} E'_{k+j} = \mathcal{I} \setminus \bigcap_{j=1}^{d} E_{k+j}, \tag{4}$$

where

$$E_{k+j} \triangleq \{|y_{k+j}| < L\}, \quad \forall j \geq 1,$$
$$S_{k+j} \triangleq \begin{cases} E'_{k+j} & j = 1 \\ \bigcap_{i=1}^{j-1} E_{k+i}, E'_{k+j} & \forall j > 1. \end{cases}$$

The level-crossing event C_k represents at least one exceedance over the span of d time steps. Mathematically, it can also be represented as the union of disjoint subevents, $\bigcup_{j=1}^{d} S_{k+j}$, or as the union of overlapping subevents, $\bigcup_{j=1}^{d} E'_{k+j}$. E'_{k+j} is based on the subevent E_{k+j}, and represents an exceedance at the jth time step. S_{k+j} is a sequence of such subevents, of which only the jth represents an exceedance. Equation (5) represents the unconditional probability of the level-crossing event in its most compact representational form.

$$P(C_k) = 1 - \int_{-L}^{L} \cdots \int_{-L}^{L} \mathcal{N}(\mathbf{y}_d; \mu_{\mathbf{y}_d}, \Sigma_{\mathbf{y}_d})\, \mathrm{d}\mathbf{y}_d, \tag{5}$$

where

$$\mathbf{y}_d \triangleq \begin{bmatrix} y_{k+1} \\ \vdots \\ y_{k+d} \end{bmatrix}, \quad \mu_{\mathbf{y}_d} = \mathbf{0}_d = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$
$$\lambda_c \triangleq \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^\top + R,$$
$$\Sigma_{\mathbf{y}_d} \triangleq \begin{cases} \lambda_c & \forall i = j \in [1, \ldots, d] \\ \mathbf{C}\mathbf{A}^{j-i}\mathbf{P}_{ss}^{L}\mathbf{C}^\top & \forall j > i \in [1, \ldots, d]. \end{cases}$$

The mathematical underpinnings for the optimal alarm condition corresponding to the level-crossing event, shown here as eqn (6), have previously been introduced in Martin (2010). Alternatively, the optimal alarm condition can be expressed in terms of the subevents E_{k+j}, as shown in eqn (7).

$$P(C_k \mid y_0, \ldots, y_k) \geq P_b, \tag{6}$$

$$P\left(\bigcap_{j=1}^{d} E_{k+j} \mid y_0, \ldots, y_k\right) \leq 1 - P_b. \tag{7}$$

The optimal alarm condition has therefore been derived from the use of the likelihood ratio resulting in the conditional inequality as given in eqn (6). This basically says 'give alarm when the conditional probability of the event, C_k, exceeds the level P_b'. However, prior to implementing this simple algorithmic decision rule, it is necessary to find the alarm regions defined by this inequality to design the alarm system according to specified missed detection and/or false alarm constraints (i.e. ROC curve analysis). The alarm region is parameterized by predicted future process output values, covariances, and P_b, the 'border' probability. P_b is an extremely important parameter, as it effectively defines the space spanned by the alarm region and therefore controls the tradeoff between false alarms and missed detections. It is thus chosen with respect to these metrics and is the key parameter used for design of an optimal alarm system.

Implementation of the optimal alarm system hinges on use of standard Kalman filter and predictor equations which are omitted for the sake of brevity. However, it is important to introduce relevant predicted future process output values, covariances and cross-covariances, given in eqns (8)–(10) respectively. These equations rely on Kalman filter formalisms and will be used in subsequent formulae. $\mathbf{P}_{ss}^{R}$ is the solution to the discrete algebraic Riccati equation (eqn 11), and $\hat{\mathbf{P}}_{ss}^{R}$ is the steady-state *a posteriori* covariance matrix given in eqn (12), where both rely on the Kalman gain defined in eqn (13). The approximations shown in eqns (9) and (10) will provide for a great computational advantage in design of the optimal alarm system and its corresponding approximations for reasons stated previously. The assumption of stationarity is required for the design of an optimal alarm system, and holds here as well.

$$\hat{y}_{k+j|k} = \mathbf{C}\mathbf{A}^{j}\hat{\mathbf{x}}_{k+j|k}, \tag{8}$$

$$\mathbf{P}_{k+j|k} \approx \mathbf{A}^{j}(\hat{\mathbf{P}}_{ss}^{R} - \mathbf{P}_{ss}^{L})(\mathbf{A}^{\top})^{j} + \mathbf{P}_{ss}^{L}, \tag{9}$$

$$\mathbf{P}_{k+i,k+j|k} \approx \mathbf{A}^{j}(\hat{\mathbf{P}}_{ss}^{R} - \mathbf{P}_{ss}^{L})(\mathbf{A}^{\top})^{i} + \mathbf{A}^{j-i}\mathbf{P}_{ss}^{L}, \tag{10}$$

$$\mathbf{P}_{ss}^{R} = \mathbf{A}\left(\mathbf{P}_{ss}^{R} - \mathbf{F}_{ss}\mathbf{C}\mathbf{P}_{ss}^{R}\right)\mathbf{A}^{\top} + \mathbf{Q}, \tag{11}$$

$$\hat{\mathbf{P}}_{ss}^{R} \triangleq \mathbf{P}_{ss}^{R} - \mathbf{F}_{ss}\mathbf{C}\mathbf{P}_{ss}^{R}, \tag{12}$$

$$\mathbf{F}_{ss} \triangleq \mathbf{P}_{ss}^{R}\mathbf{C}^{\top}(\mathbf{C}\mathbf{P}_{ss}^{R}\mathbf{C}^{\top} + R)^{-1}. \tag{13}$$

A more formal representation of the optimal alarm region is shown in eqn (15), which essentially defines a sublevel set, $L_A^-(g)$, of the function shown in eqn (14), $g(\hat{\mathbf{y}}_d)$. Equation (15) plays a pivotal role in enabling enforcement of the alarm region for an optimal alarm system. Using this condition allows it to outperform other alarm systems with respect to the minimization of false alarms.

$$g(\hat{\mathbf{y}}_d) \triangleq P\left(\bigcap_{j=1}^{d} E_{k+j} \,|\, y_0, \ldots, y_k\right), \tag{14}$$

$$\begin{aligned} A_k &\triangleq \left\{\bigcap_{i=1}^{d} \hat{y}_{k+i|k} : P(C_k \,|\, y_0, \ldots, y_k) \geq P_b\right\} \\ &= \left\{\bigcap_{i=1}^{d} \hat{y}_{k+i|k} : P(\bigcap_{j=1}^{d} E_{k+j} \,|\, y_0, \ldots, y_k) \leq 1 - P_b\right\} \\ &= \{\hat{\mathbf{y}}_d : g(\hat{\mathbf{y}}_d) \leq 1 - P_b\} \\ &= L_A^-(g), \end{aligned} \tag{15}$$

where

$$\hat{\mathbf{y}}_d \triangleq E[\mathbf{y}_d \,|\, y_0, \ldots, y_k] = \begin{bmatrix} \hat{y}_{k+1|k} \\ \vdots \\ \hat{y}_{k+d|k} \end{bmatrix}.$$

Equations (17) and (18) give the integral form of eqn (14), a multi-variate normal probability to be computed numerically. It should be noted that an important boundary condition exists for $g(\hat{\mathbf{y}}_d)$, shown in eqn (16), given eqns (5), (17), and (18).

$$g(\mu_{\mathbf{y}_d}) = g(\mathbf{0}_d), \tag{16}$$

$$g(\hat{\mathbf{y}}_d) = P\left(\bigcap_{j=1}^{d} E_{k+j} \,|\, y_0, \ldots, y_k\right) = \int_{-L}^{L} \cdots \int_{-L}^{L} \mathcal{N}(\mathbf{y}_d; \hat{\mathbf{y}}_d, \Sigma_{\varepsilon_d})\, \mathrm{d}\mathbf{y}_d \tag{17}$$

$$= \int_{-L-\hat{y}_{k+1|k}}^{L-\hat{y}_{k+1|k}} \cdots \int_{-L-\hat{y}_{k+d|k}}^{L-\hat{y}_{k+d|k}} \mathcal{N}(\mathbf{y}_d; \mathbf{0}_d, \Sigma_{\varepsilon_d})\, \mathrm{d}\mathbf{y}_d, \tag{18}$$

where

$$\begin{aligned} \boldsymbol{\varepsilon}_d &\triangleq \mathbf{y}_d - \hat{\mathbf{y}}_d, \\ \Sigma_{\varepsilon_d} &\triangleq \begin{cases} V_{k+i|k} & \forall i = j \in [1, \ldots, d] \\ \mathbf{C}\mathbf{P}_{k+i,k+j|k}\mathbf{C}^{\top} & \forall i \neq j \in [1, \ldots, d], \end{cases} \\ V_{k+i|k} &\triangleq \mathbf{C}\mathbf{P}_{k+i|k}\mathbf{C}^{\top} + R. \end{aligned}$$

The feasible region for values of P_b can easily be determined by applying an intermediate value theorem from calculus which provides sufficient conditions for finding a level set solution. The sufficient conditions are shown in eqns (19) and (20), and the resulting level set $L_A(g)$ is shown in eqn (21).

$$g(\mu_{\mathbf{y}_d}) \geq 1 - P_b, \tag{19}$$

$$\lim_{|\hat{\mathbf{y}}_d| \setminus \hat{y}_{k+j|k} \to \infty} g(\hat{\mathbf{y}}_d) < 1 - P_b, \quad \forall j \in [1, \ldots, d], \tag{20}$$

$$L_A(g) \triangleq \{\hat{\mathbf{y}}_d : g(\hat{\mathbf{y}}_d) = 1 - P_b\}, \tag{21}$$

$$P_{b_{\text{crit}}} = 1 - g(\mu_{\mathbf{y}_d}) = 1 - g(\mathbf{0}_d). \tag{22}$$

The notation that represents the limiting condition shown in eqn (20) is $|\hat{\mathbf{y}}_d| \setminus \hat{y}_{k+j|k} \to \infty$, and it is meant to indicate that all elements of $\hat{\mathbf{y}}_d$ other than $\hat{y}_{k+j|k}$ approach $\pm\infty$. Application of this condition yields $P_b < 1$, which is true by definition, and application of the sufficient condition shown in eqn (19) yields $P_b \geq P_{b_{\text{crit}}}$. $P_{b_{\text{crit}}}$ is defined for the optimal alarm region in eqn (22) and can be derived from the previous observation about an existing boundary condition, $g(\mu_{\mathbf{y}_d}) = g(\mathbf{0}_d)$. Thus, the feasible region for P_b is $P_b \in [1 - g(\mathbf{0}_d), 1]$. It is important to note that any approximation to the optimal alarm region may change the value for $P_{b_{\text{crit}}}$.

It is not possible to obtain a closed-form representation of the parametrization for the optimal alarm region shown in eqn (15). As such, a Monte Carlo approach must be used to generate relevant alarm design (ROC curve) statistics. This allows the ROC curve statistics to be estimated empirically with observational truth data generated from the existing model and corresponding computations of the conditional probability of level-crossing events given by eqns (17) and (18). However, with the aid of two distinct approximations or competing alternatives, ROC curve statistics can be generated by numerically integrating expressions for the computation of relevant multi-variate normal probabilities. These multi-variate probability computations are performed by using an adaptation of a special purpose algorithm (Genz, 1992), which is based on a robust and computationally efficient technique designed to be used for integrations in multiple dimensions for multi-variate normal distributions. This provides a tool necessary for the design of approximations to an optimal alarm system. As such, computationally intensive simulation runs using MonteCarlo empirical estimation or computationally prohibitive generic Monte Carlo numerical integration routines where the integration region is given by storing precomputed points along the boundary of the alarm region (Svensson, 1998) can be avoided.

The two approximations and competing alternatives will allow the formulae for the true and false positive rates to be computed by appealing to eqns (23) and (24). These are the two primary equations that provide the basis of alarm system design, and represent the metrics used to construct the ROC curve. The extreme value analysis conducted in this article hinges upon the use of these formulae as a starting point.

True positive rate:

$$P_d = P(A_k \mid C_k) = \frac{P(C_k, A_k)}{P(C_k)}. \tag{23}$$

False positive rate:

$$\begin{aligned} P_{fa} = P(A_k | C_k') &= \frac{P(C_k', A_k)}{P(C_k')} \\ &= \frac{P(A_k) - P(C_k, A_k)}{1 - P(C_k)}. \end{aligned} \tag{24}$$

The formula for $P(C_k)$ has already been introduced in eqn (5) and holds regardless of the alarm system being used. Thus, only the additional expressions for $P(C_k, A_k)$ and $P(A_k)$ are necessary for computing P_d and P_{fa}, specific to the alarm system under consideration.

2.2. Alarm system formulation details

Two competing baseline alternatives, the 'redline' and 'predictive' alarm systems, will be compared with the optimal alarm system approximations. They will also attempt to predict the level-crossing event defined by eqn (4). The redline alarm system defines an envelope, $[-L_A, L_A]$, outside of which an alarm will be triggered to forewarn of the impending level-crossing event. Thus, it is a simple alarm level crossing used to predict a second more critical level-crossing. In this case two levels are used: L as the critical threshold and L_A as the design threshold. The probabilities necessary to compute P_d and P_{fa} based on eqns (23) and (24) for this alarm system are provided in eqns (26)–(29), where a redefinition of A_k in eqn (25) now holds, such that the alarm is based only on the current process value.

$$A_k = \{|y_k| > L_A\}, \tag{25}$$

$$P(A_k) = P(|y_k| > L_A), \tag{26}$$

$$= 2\Phi\left(\frac{-L_A}{\sqrt{\lambda_c}}\right), \tag{27}$$

$$P(C_k, A_k) = P(C_k) - P(A'_k) + P(C'_k, A'_k), \tag{28}$$

$$P(C'_k, A'_k) = \int_{-L_A}^{L_A}\int_{-L}^{L}\cdots\int_{-L}^{L}\mathcal{N}(\mathbf{z};\mu_{\mathbf{z}},\Sigma_{\mathbf{z}})\,d\mathbf{z}, \tag{29}$$

where

$$\mathbf{z} \triangleq \begin{bmatrix} y_k \\ \mathbf{y}_d \end{bmatrix},$$
$$\mu_{\mathbf{z}} \triangleq \begin{bmatrix} \mu_{y_k} \\ \mu_{\mathbf{y}_d} \end{bmatrix} = \mathbf{0}_{d+1},$$
$$\Sigma_{\mathbf{z}} \triangleq \begin{bmatrix} \lambda_c & \Lambda_c \\ \Lambda_c^\top & \Sigma_{\mathbf{y}_d} \end{bmatrix} \approx \begin{cases} \lambda_c & \forall i = j \in [0,\ldots,d] \\ \mathbf{C}\mathbf{A}^{j-i}\mathbf{P}_{ss}^{L}\mathbf{C}^\top & \forall j > i \in [0,\ldots,d], \end{cases}$$
$$\Lambda_c = \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{A}^\top\mathbf{O}^\top.$$
$$\mathbf{O} \triangleq \begin{bmatrix} \mathbf{C} \\ \vdots \\ \mathbf{C}\mathbf{A}^{d-1} \end{bmatrix}.$$

The 'predictive' alarm system incorporates the use of predicted future process values and defines the same envelope, $[-L_A, L_A]$, outside of which an alarm will be triggered to forewarn of the impending level-crossing event. However, the alarm definition differs from the redline method as shown in eqn (30). The predicted future process value $\hat{y}_{k+d|k}$ is found from standard Kalman filter eqn (8). The probabilities necessary to compute P_d and P_{fa} based on eqns (23) and (24) for this alarm system are provided in eqns (31)–(34).

$$A_k = \{|\hat{y}_{k+d|k}| > L_A\}, \tag{30}$$

$$P(A_k) = P(|\hat{y}_{k+d|k}| > L_A) \tag{31}$$

$$= 2\Phi\left(\frac{-L_A}{\sqrt{\lambda_c - V_{k+d|k}}}\right), \tag{32}$$

$$P(C_k, A_k) = P(C_k) - P(A'_k) + P(C'_k, A'_k), \tag{33}$$

$$P(C'_k, A'_k) = \int_{-L}^{L}\cdots\int_{-L}^{L}\int_{-L_A}^{L_A}\mathcal{N}(\mathbf{z};\mu_{\mathbf{z}},\Sigma_{\mathbf{z}})\,d\mathbf{z}, \tag{34}$$

where

$$\mathbf{z} \triangleq \begin{bmatrix} \mathbf{y}_d \\ \hat{y}_{k+d|k} \end{bmatrix},$$
$$\mu_{\mathbf{z}} \triangleq \begin{bmatrix} \mu_{\mathbf{y}_d} \\ \mu_{\hat{y}_{k+d|k}} \end{bmatrix} = \mathbf{0}_{d+1},$$
$$\Sigma_{\mathbf{z}} \triangleq \begin{bmatrix} \Sigma_{\mathbf{y}_d} & \Lambda_a^\top \\ \Lambda_a & \lambda_c - V_{k+d|k} \end{bmatrix},$$
$$\Lambda_a = \mathbf{C}\mathbf{A}^d(\mathbf{P}_{ss}^{L} - \mathbf{P}_{ss}^{R})\mathbf{O}^\top.$$

The first approximation to the optimal alarm system is termed the 'closed-form' approximation. Fundamentally, this approximation is constructed by solving for asymptotic bounds on the exact alarm region. By using asymptotes, a geometrical approximation is implicitly formed with a hyperbox around the alarm region. By using two successive approximations to find these asymptotes as previously described in Martin (2010), the alarm region $A_k = L_A^-(g)$ can be approximated in 'closed-form'. The first approximation is $A_k = L_A^-(g) \approx \bigcup_{j=1}^{d} A_k^j$, and the second approximation uses the definition of A_k^j given in eqns (36) and (37). The expression for the asymptotic bounds, L_{A_j}, is shown in eqn (35). The closed form approximation domain of feasibility is $P_b \in [P_{b_{\text{crit}}}, 1]$ where $P_{b_{\text{crit}}}$ is given in eqn (38), again as previously derived in Martin (2010).

$$L_{A_j} \triangleq L + \sqrt{V_{k+j|k}}\Phi^{-1}(P_b), \tag{35}$$

$$A_k^j = \{\hat{y}_{k+j|k} : P(E_{k+j}|y_0, \ldots, y_k) \leq 1 - P_b\} \tag{36}$$

$$\approx |\hat{y}_{k+j|k}| \geq L_{A_j}, \tag{37}$$

$$P_{b_{\text{crit}}} = \Phi\left(\frac{-L}{\sqrt{V_{k+d|k}}}\right). \tag{38}$$

The second approximation to be investigated is termed the 'root-finding' approximation, as previously described (Martin, 2010). As with the 'closed-form' approximation, fundamentally this approximation is constructed by solving for asymptotic bounds on the exact alarm region. Again, by using asymptotes, a geometrical approximation is implicitly formed with a hyperbox around the alarm region. However, instead of using two successive approximations to find the asymptotes, the alarm region $A_k = L_A^-(g)$ can be approximated with a single approximation $A_k = L_A^-(g) \approx \bigcup_{j=1}^d \Omega_{A_j}$ by solving a root-finding problem which yields bounding asymptotes. The root-finding problem, eqn (42), can be solved numerically. An alternate definition for L_{A_j} is also used, which overloads the same notation used for the 'closed-form' approximation asymptotes. Supporting expressions for $f(\hat{y}_{k+j|k})$ and Ω_{A_j} are given in eqns (39) and eqns (40) and (41) respectively.

$$f(\hat{y}_{k+j|k}) = \int_{-L}^{L} \cdots \int_{-L-\hat{y}_{k+j|k}}^{L-\hat{y}_{k+j|k}} \cdots \int_{-L}^{L} \mathcal{N}(\mathbf{y}_d; \mathbf{0}_d, \Sigma_{\varepsilon_d})\, \mathrm{d}\mathbf{y}_d, \tag{39}$$

$$\Omega_{A_j} = \{\hat{y}_{k+j|k} : f(\hat{y}_{k+j|k}) \leq 1 - P_b\} \tag{40}$$

$$= \{|\hat{y}_{k+j|k}| \geq L_{A_j}\}, \tag{41}$$

$$L_{A_j} \triangleq \{|\hat{y}_{k+j|k}| : f(\hat{y}_{k+j|k}) = 1 - P_b\}. \tag{42}$$

Furthermore, the feasible region for values of $P_b \in [P_{b_{\text{crit}}}, 1]$ is the same for the root-finding method as for the optimal alarm region, where $P_{b_{\text{crit}}} = P(C_k)1 - g(\mathbf{0}_d)$ as given in eqn (22). For both of these approximations to the optimal alarm system, eqns (43) and (44) can be used.

$$P(A_k) = \begin{cases} 1 - P(A_k') & P_b > P_{b_{\text{crit}}} \\ 1 & P_b = P_{b_{\text{crit}}} \end{cases}, \tag{43}$$

$$P(C_k, A_k) = \begin{cases} P(C_k) - P(A_k') + P(C_k', A_k') & P_b > P_{b_{\text{crit}}} \\ P(C_k) & P_b = P_{b_{\text{crit}}} \end{cases}, \tag{44}$$

where

$$P(A_k') \triangleq P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) = \int_{-L_{A_1}}^{L_{A_1}} \cdots \int_{-L_{A_d}}^{L_{A_d}} \mathcal{N}(\hat{\mathbf{y}}_d; \mu_{\mathbf{y}_d}, \Sigma_{\hat{\mathbf{y}}_d})\, \mathrm{d}\hat{\mathbf{y}}_d,$$

and

L_{A_j} :Asymptotic bound on alarm region formed by either approximation

$$\Sigma_{\hat{\mathbf{y}}_d} \triangleq \Sigma_{\mathbf{y}_d} - \Sigma_{\varepsilon_d} = \mathbf{O}(\mathbf{P}_{ss}^L - \mathbf{P}_{ss}^R)\mathbf{O}^\top.$$

Since the predictor $\hat{\mathbf{y}}_d$ (with covariance matrix $\Sigma_{\hat{\mathbf{y}}_d}$) and the residual $\boldsymbol{\varepsilon}_d$ (with covariance matrix Σ_{ε_d}) are uncorrelated, a more intuitive way to express the relationship between the corresponding covariance matrices is $\Sigma_{\mathbf{y}_d} = \Sigma_{\hat{\mathbf{y}}_d} + \Sigma_{\varepsilon_d}$.

Furthermore,

$$P(C_k', A_k') = P\left(\bigcap_{j=1}^{d} E_{k+j}, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) = \int_{-L}^{L} \cdots \int_{-L}^{L} \int_{-L_{A_1}}^{L_{A_1}} \cdots \int_{-L_{A_d}}^{L_{A_d}} \mathcal{N}(\mathbf{z}; \mu_{\mathbf{z}}, \Sigma_{\mathbf{z}})\, \mathrm{d}\mathbf{z},$$

where

$$\mathbf{z} \triangleq \begin{bmatrix} \mathbf{y}_d \\ \hat{\mathbf{y}}_d \end{bmatrix},$$
$$\mu_{\mathbf{z}} \triangleq \begin{bmatrix} \mu_{\mathbf{y}_d} \\ \mu_{\mathbf{y}_d} \end{bmatrix},$$
$$\Sigma_{\mathbf{z}} \triangleq \begin{bmatrix} \Sigma_{\mathbf{y}_d} & \Sigma_{\hat{\mathbf{y}}_d} \\ \Sigma_{\hat{\mathbf{y}}_d} & \Sigma_{\hat{\mathbf{y}}_d} \end{bmatrix}.$$

2.3. Small value level-crossing prediction

In general, solutions to problems given by eqns (45) and (46) must precede finding the solution to eqn (47). As such, eqn (48) follows naturally from the fact that $\lim_{L\to 0} P(C_k) = 1$, by using eqn (5) and applying eqns (23) and (24).

$$P_{fa}^0 \triangleq \lim_{L\to 0} P_{fa}, \tag{45}$$

$$P_d^0 \triangleq \lim_{L\to 0} P_d, \tag{46}$$

$$\mathrm{AUC}_0 \triangleq \lim_{L\to 0} \mathrm{AUC}, \tag{47}$$

$$P_d^0 = \lim_{L\to 0} P(C_k, A_k) = \lim_{L\to 0} P(A_k). \tag{48}$$

To find a well-defined limit for P_{fa}^0, l'Hôpital's rule must be invoked due to the fact that it is in an indeterminate form. For the redline and predictive methods, applying l'Hôpital's rule yields eqns (49) and (50) respectively, which are both derived using Theorems 3 and 4 provided in Appendix A.

$$P_{fa}^0 = 2\Phi\left(\frac{-L_A}{a_{\mathrm{red}}\sqrt{\lambda_c}}\right), \tag{49}$$

$$P_{fa}^0 = 2\Phi\left(\frac{-L_A}{a_{\mathrm{pred}}\sqrt{\lambda_c - V_{k+d|k}}}\right), \tag{50}$$

where

$$a_{\mathrm{red}} \triangleq \sqrt{1 - \frac{\Lambda_c \Sigma_{\mathbf{y}_d}^{-1} \Lambda_c^\top}{\lambda_c}},$$
$$a_{\mathrm{pred}} \triangleq \sqrt{1 - \frac{\Lambda_a \Sigma_{\mathbf{y}_d}^{-1} \Lambda_a^\top}{\lambda_c - V_{k+d|k}}}.$$

Combining eqns (48) and either (49) or (50), P_d^0 can be expressed as a function of P_{fa}^0 as shown in eqn (51). The solution to AUC_0 is given in eqn (52), and its generic derivation is provided in Theorem 5 in Appendix A. This formula can be used for either the redline or predictive methods in the case where $a{=}a_{\mathrm{red}}$ or $a{=}a_{\mathrm{pred}}$ respectively.

$$P_d^0 = 2\Phi\left(a\Phi^{-1}\left(\frac{P_{fa}^0}{2}\right)\right), \tag{51}$$

$$\begin{aligned} \mathrm{AUC}_0 &= \int_0^1 P_d^0(P_{fa}^0)\,\mathrm{d}P_{fa}^0 \\ &= 1 - \frac{2}{\pi}\arcsin\left(\frac{a}{\sqrt{1+a^2}}\right) \\ &= 1 - \frac{2}{\pi}\arctan\, a. \end{aligned} \tag{52}$$

As it turns out, eqns (48)–(52) are valid when $d \geq 1$ due to the fact that the alarm region is one-dimensional. As such, the closed form expression for AUC_0 shown in eqn (52) can be expressed explicitly as both a function of d, and other model parameters. These ideas will be explored further in Section 4.

For the 'closed-form' approximation, if $d = 1$, eqn (53) holds, following naturally from a special case of eqn (57) shown later.

$$P_d^0 = \lim_{L\to 0} P(C_k, A_k^1) = \lim_{L\to 0} P(A_k^1) = \begin{cases} 2\Phi(-a_{cf}\Phi^{-1}(P_b)) & P_b > 0.5 \\ 1 & P_b = 0.5 \end{cases}, \tag{53}$$

where

$$a_{cf} \triangleq \sqrt{\frac{V_{k+1|k}}{\lambda_c - V_{k+1|k}}}.$$

As before, to find a well-defined limit for P_{fa}^0, l'Hôpital's rule must be invoked. Thus, eqn (54) holds by applying l'Hôpital's rule, and the derivation is provided in Corollary 1 in Appendix A.

$$P_{fa}^0 = \begin{cases} 2(1 - \Phi(b_{cf}\Phi^{-1}(P_b))) & P_b > 0.5 \\ 1 & P_b = 0.5, \end{cases} \tag{54}$$

where

$$b_{cf} \triangleq \sqrt{1 + a_{cf}^2} = \frac{\lambda_c}{\lambda_c - V_{k+1|k}}.$$

Combining eqns (53) and (54), P_d^0 can be expressed as a function of P_{fa}^0 as shown in eqn (55). The solution to AUC_0 is given in eqn (56), and its derivation is provided in Theorem 2 in Appendix A.

$$P_d^0 = 2\Phi\left(-a_{cf}b_{cf}\Phi^{-1}\left(1 - \frac{P_{fa}}{2}\right)\right). \tag{55}$$

$$\begin{aligned} \mathrm{AUC}_0 &= \int_0^1 P_d^0(P_{fa}^0)\, \mathrm{d}P_{fa}^0 \\ &= 1 - \frac{2}{\pi}\arcsin\left(\frac{c}{\sqrt{1+c^2}}\right) \\ &= 1 - \frac{2}{\pi}\arctan c \\ &= 1 - \frac{2}{\pi}\arctan\left(\frac{\sqrt{\lambda_c V_{k+1|k}}}{\lambda_c - V_{k+1|k}}\right), \end{aligned} \tag{56}$$

where

$$c \triangleq a_{cf}b_{cf} = \frac{\sqrt{\lambda_c V_{k+1|k}}}{\lambda_c - V_{k+1|k}}.$$

When $d > 1$, it is not possible to derive a closed form expression for AUC_0 as a function of model parameters. However, it is possible to find simplified expressions for P_d^0 and P_{fa}^0, which are given in eqns (57) and (58) respectively. Here AUC_0 can be computed numerically for various values of P_b. Equations (57) and (58) are thus generalizations of eqns (53) and (54) that were respectively expressed in closed-form when $d = 1$. Detailed proofs for eqns (57) and (58) (for $d > 1$) are provided in Theorem 1, which can be seen in Appendix A. Similarly, detailed proofs for eqns (53) and (54), which naturally follow from Theorem 1, are also provided in Corollary 1 in Appendix A.

$$P_d^0 = \lim_{L\to 0} P\left(C_k, \bigcup_{j=1}^d A_k^j\right) = \lim_{L\to 0} P\left(\bigcup_{j=1}^d A_k^j\right) = \begin{cases} 1 - P(\bigcap_{j=1}^d |\hat{y}_{k+j|k}| \le L_{A_j}^0) & P_b > 0.5 \\ 1 & P_b = 0.5, \end{cases} \tag{57}$$

$$P_{fa}^0 = \begin{cases} 1 - P\left(\bigcap_{j=1}^d |\hat{y}_{k+j|k}| \le L_{A_j}^0\right) & P_b > 0.5 \\ 1 & P_b = 0.5, \end{cases} \tag{58}$$

where

$$P\left(\bigcap_{j=1}^{d}|\hat{y}_{k+j|k}| \le L_{A_j}^0\right) = \int_{-L_{A_1}^0}^{L_{A_1}^0} \cdots \int_{-L_{A_d}^0}^{L_{A_d}^0} \mathcal{N}\left(\hat{\mathbf{y}}_d; \mathbf{0}_d, \Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d}\Sigma_{\mathbf{y}_d}^{-1}\Sigma_{\hat{\mathbf{y}}_d}^{\top}\right) \mathrm{d}\hat{\mathbf{y}}_d$$
$$L_{A_j}^0 \triangleq \sqrt{V_{k+j|k}}\Phi^{-1}(P_b).$$

For the root-finding approximation method, all of limiting results as $L \to 0$ change dramatically due in part to the much different domain of feasibility for P_b as compared to the 'closed-form' approximation. As pointed out in the previous section, the domain of feasibility or feasible region for values of $P_b \in [P_{b_{\text{crit}}}, 1]$ is the same for the root-finding method as for the optimal alarm region, unlike the closed-form approximation. Solutions to the small value level-crossing prediction problem $\forall d \ge 1$ can be combined, because it will be found that there is no distinction between the solutions for the cases $d = 1$ and $d > 1$ as there was with the closed-form approximation. In fact, it can be proven that the solution is theoretically prohibitive by recognizing the limiting value for $P_{b_{\text{crit}}} = 1 - g(\mathbf{0}_d)$ shown below:

$$\begin{aligned}\lim_{L\to 0} P_{b_{\text{crit}}} &= \lim_{L\to 0} 1 - g(\mathbf{0}_d)\\ &= 1.\end{aligned}$$

Since the resulting revised domain is $P_b \in [1,1]$ when $L \to 0$, the only value in the domain is the maximum value of $P_b = 1$ for the root-finding approximation. As shown below, it then follows that any value of $\hat{y}_{k+j|k}$ suffices for $L_{A_j}^0$, given eqns (39) and (42).

$$\begin{aligned}&\lim_{L\to 0} f(\hat{y}_{k+j|k}) = 0, \quad \forall \hat{y}_{k+j|k}\\ &\text{and } L_{A_j}^0 = \{|\hat{y}_{k+j|k}| : P_b = 1\}.\end{aligned}$$

This implies that the resulting alarm system will never yield positive indications of the level-crossing event. As such, intuitively it makes sense that $P_d = 0$ and $P_{fa} = 0$, which are proven mathematically in eqns (59) and (60).

$$\lim_{L\to 0} P_d = \lim_{L\to 0} P\left(C_k, \bigcup_{j=1}^{d}\Omega_{A_j}\right) = \lim_{L\to 0} P\left(\bigcup_{j=1}^{d}\Omega_{A_j}\right) = 1 - P\left(\bigcap_{j=1}^{d}|\hat{y}_{k+j|k}| \le \infty\right) = 0, \tag{59}$$

$$\lim_{L\to 0} P_{fa} = 1 - P\left(\bigcap_{j=1}^{d}|\hat{y}_{k+j|k}| \le \infty\right) = 0. \tag{60}$$

As there is only a single alarm point, a fully formed ROC curve cannot be formed and it is thus impossible to compute or find an expression for the AUC as was performed for the closed form approximation.

2.4. Large value level-crossing prediction

In general, we will cover solutions to the problems given by the definitions shown in eqns (61)–(63) in this section.

$$P_{fa}^{\infty} \triangleq \lim_{L\to\infty} P_{fa}, \tag{61}$$

$$P_{d}^{\infty} \triangleq \lim_{L\to\infty} P_{d}, \tag{62}$$

$$\text{AUC}_{\infty} \triangleq \lim_{L\to\infty} \text{AUC}. \tag{63}$$

For both the redline and predictive methods, solutions to eqns (61) and (62), $\forall d \ge 1$, are provided in eqns (64) and (65) respectively.

$$\begin{aligned}P_{fa}^{\infty} &= \lim_{L\to\infty} \frac{P(A_k) - P(C_k, A_k)}{1 - P(C_k)}\\ &= \lim_{L\to\infty} \frac{P(C_k') - P(C_k', A_k')}{P(C_k')}\\ &= \lim_{L\to\infty} \frac{P\left(\bigcap_{j=1}^{d}|y_{k+j}| < L\right) - P\left(\bigcap_{j=1}^{d}|y_{k+j}| < L, A_k'\right)}{P\left(\bigcap_{j=1}^{d}|y_{k+j}| < L\right)}\\ &= P(A_k),\end{aligned} \tag{64}$$

$$\begin{aligned} P_d^{\infty} &= \lim_{L\to\infty} \frac{P(C_k, A_k)}{P(C_k)} \\ &= \frac{0}{0} \\ &= \text{Undefined.} \end{aligned} \tag{65}$$

Because P_d^{∞} results in an indeterminate form $\left(\frac{0}{0}\right)$, l'Hôpital's rule must be applied. However, even after infinite differentiations, as the operation $\lim_{L\to\infty} \lim_{m\to\infty} \frac{\partial^{md}}{\partial L^{md}}$, $m \in \mathbb{Z}^+$ is applied to both the numerator and denominator of eqn (65), the quotient P_d^{∞} is undefined. As such, it is evident that $\text{AUC}_{\infty} = \int_0^1 P_d^{\infty}(P_{fa}^{\infty})\, \mathrm{d}P_{fa}^{\infty} = 1$.

For the closed-form approximation, solutions to the problems given by $\lim_{L\to\infty} P_{fa}$ and $\lim_{L\to\infty} P_d$, $\forall d \geq 1$ are provided in eqns (66) and (67) respectively. The last two steps prior to the result shown in eqn (66) are evident after application of eqn (38) for $P_{b_{\text{crit}}}$.

$$\begin{aligned} P_{fa}^{\infty} &= \lim_{L\to\infty} \frac{P(A_k) - P(C_k, A_k)}{1 - P(C_k)} \\ &= \lim_{L\to\infty} \frac{P(C_k') - P(C_k', A_k')}{P(C_k')} \\ &= \lim_{L\to\infty} \frac{P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right) - P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right)}{P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right)} \\ &= 1 - \lim_{L\to\infty} P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) \\ &= \begin{cases} 1 - \lim_{L\to\infty} P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L + \sqrt{V_{k+j|k}}\Phi^{-1}(P_b)\right) & P_b > 0 \\ 1 - \lim_{L\to\infty} P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L\left(1 - \sqrt{\frac{V_{k+j|k}}{V_{k+d|k}}}\right)\right) & P_b = 0 \end{cases} \\ &= \begin{cases} 1 - P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < \infty\right) & P_b > 0 \\ 1 - P\left(\bigcap_{j=1}^{d-1} |\hat{y}_{k+j|k}| < \infty, |\hat{y}_{k+d|k}| < 0\right) & P_b = 0 \end{cases} \\ &= \begin{cases} 0 & P_b > 0 \\ 1 & P_b = 0 \end{cases}, \end{aligned} \tag{66}$$

$$\begin{aligned} P_d^{\infty} &\triangleq \lim_{L\to\infty} \frac{P(C_k, A_k)}{P(C_k)} \\ &= \lim_{L\to\infty} \frac{1 - P(C_k') - P(A_k') + P(C_k', A_k')}{1 - P(C_k')} \\ &= \lim_{L\to\infty} \frac{0 - P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) - P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right)}{0} \\ &= \frac{0}{0} \\ &= \text{Undefined.} \end{aligned} \tag{67}$$

As before, P_d^{∞} results in an indeterminate form $\left(\frac{0}{0}\right)$ even after infinite differentiations using l'Hôpital's rule. As such, it is also evident that $\text{AUC}_{\infty} = \int_0^1 P_d^{\infty}(P_{fa}^{\infty})\, \mathrm{d}P_{fa}^{\infty} = 1$. For the root-finding approximation, solutions to the problems posed above are identical to those presented for the closed form approximation. The main difference is in the definition of the domain of P_b, however $\lim_{L\to\infty} P_{b_{\text{crit}}} = 0$ applies for both approximations, using eqns (22) and (38).

3. EXAMPLE

The example to be used for the presentation of the results has no specific application, but is generic and based on the same example used in previous work (Svensson *et al.*, 1996; Martin, 2010). The model parameters are provided in eqns (68)–(71).

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -0.9 & 1.8 \end{bmatrix}, \tag{68}$$

$$\mathbf{C} = \begin{bmatrix} 0.5 & 1 \end{bmatrix}, \tag{69}$$

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{70}$$

$$R = 0.08. \tag{71}$$

Unless otherwise stated, all cases to be compared will use a prediction window of $d = 5$ while varying L.

4. RESULTS AND DISCUSSION

A comparison of the AUC for all alarm systems for a prediction window of $d = 5$ while varying $L \in [0,40]$ is shown in Figure 1. It is very clear that the optimal alarm system and its approximations outperform the redline and predictive methods, over the entire range of values shown for L, as expected and shown previously (Martin, 2010). Another important point to note is that the closed-form and root-finding methods shown as dashed and dotted blue lines respectively approximate the exact optimal performance (shown as a solid blue line) quite well over most of the range of values shown for L. The exact optimal results were obtained by using a Monte Carlo simulation over only a portion of the entire range shown. This is due to the fact that it would be infeasible to find the AUC with this method as $L \to 0$, and it would take prohibitively long to find the AUC for higher values of L, unlike with the approximations.

However, the approximations break down as evidenced by the notable divergence of AUC values. The closed-form approximation breaks down with higher values of L, and the root-finding approximation breaks down with lower values of L. These observations are purely empirical and unsubstantiated by theoretical backing, the latter of which is due to the finding that it is not possible to compute AUC_0. However, for the closed-form approximation, it is empirically evident in Figure 1 that a small divergence from the optimal AUC as $L \to 0$ also exists. There is theoretical backing for this divergence, given the expressions for AUC_0 derived in the previous section. The technical conditions that may lead to more egregious divergences than this one will be explored later in this section. Furthermore, as proven in the previous section $\text{AUC}_\infty = 1$ for *both* approximations to the optimal predictor and the redline and predictive methods. This is evident in Figure 1 and is also well substantiated by the same apparent behaviour of the results for the AUC when using the optimal level-crossing predictor itself.

Due to the closed-form representation of AUC for the baseline redline and predictive methods when $L = 0$, AUC_0 can easily be expressed explicitly and thus plotted as a function of d. The results are shown in Figure 2. However, the same is not true for the closed-form approximation, and so for this case alone the AUC values shown as a function of d in Figure 2 were all computed numerically with the exception of the point corresponding to $d = 1$. A more detailed theoretical analysis of the limiting considerations as $d \to \infty$, and for $L \geq 0$ for the closed-form approximation, will require functional and infinite dimensional analysis, and as such will be studied in earnest in subsequent work. However, it is clear from Figure 2, as it was previously found in Martin (2010), that the closed-form approximation outperforms the baseline redline and predictive methods, $\forall d$.

Some other interesting and intuitive observations alluded to earlier hinge upon the elegant closed-form representations for AUC_0 for all alarm systems. Equations (52) and (56) demonstrate that AUC_0 can be expressed explicitly as a function of model parameters. However, again these representations are specific to the alarm system under consideration. For the closed-form approximation, AUC_0 can only be expressed explicitly as a function of model parameters when $d = 1$. Nonetheless, the corresponding expression for AUC_0 allows for some intuitive observations to be made in face of the limiting consideration $L \to 0$, as shown in eqn (72), which can easily be derived from eqn (56).

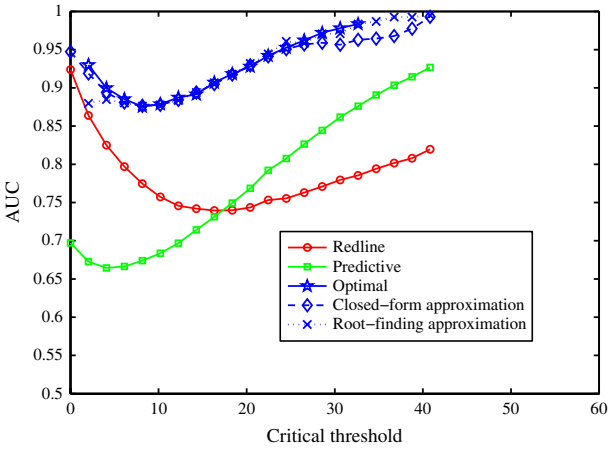

Figure 1. AUC for all alarm systems a function of critical threshold, L

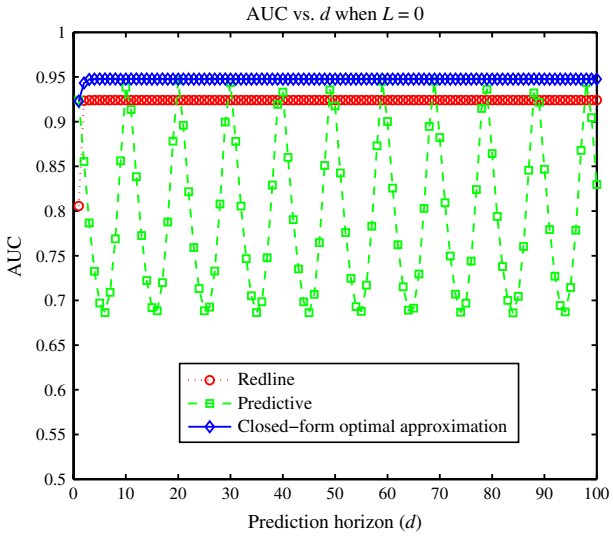

Figure 2. AUC for all alarm systems a function of prediction horizon, d, when $L = 0$

$$\lim_{V_{k+1|k}\to\lambda_c} \mathrm{AUC}_0 = 0. \tag{72}$$

The natural and intuitive interpretation of eqn (72) is that a relationship exists between the level-crossing predictive capability (*event* prediction) as quantified by AUC_0 and the Kalman prediction error (*value* prediction) as quantified by $V_{k+1|k}$. *Value* prediction was previously defined in Lindgren (1985) as the ability to predict future process values of a stationary linear Gaussian random process, in the sense of least squares. *Event* prediction is the ability to accurately predict a level-crossing event constructed over a defined prediction horizon, involving exceedances of a given critical threshold by the random process. It is clear that AUC_0 is directly influenced by $V_{k+1|k}$ as it approaches the upper bound λ_c. In this case, the level-crossing performance quantified by $\mathrm{AUC}_0 \to 0$ yields a completely imperfect prediction. In short, intuitively when $d = 1$, as the uncertainty in the value prediction increases, event prediction capability of level-crossings worsens. Although intuitive, the novelty of this result lies in the fact that it is now supported by rigorous theoretical proof. In subsequent work, the more general case when $d \geq 1$ will be investigated, as well a more thorough exploration of the relationship between *value* prediction and *event* prediction using a different contextual basis.

It is also of interest to examine what happens as $V_{k+1|k}$ approaches a lower bound. It is well known that the Kalman prediction error ($V_{k+1|k}$) and thus value prediction uncertainty is directly influenced by the measurement noise (R), among other factors. The measurement noise has the most straightforward relationship to AUC_0 in the context of eqn (72), and as such it is best candidate to study the effect of varying the Kalman prediction error ($V_{k+1|k}$). As the measurement noise approaches zero ($R \to 0$), a lower bound for the prediction error is given by $V_{k+1|k} \to \mathbf{CQC}^\top$. The full relationship is given in eqn (73), where it is clear that as $R \to 0$, the limiting value for $\mathrm{AUC}_0 < 1$ will not yield perfect prediction. However, under certain very specific circumstances related to system stability governed by $\mathbf{A}$, the limiting value for $\mathrm{AUC}_0 = 1$ will result in perfect prediction when $R \to 0$. This observation will be studied in depth in a sequel paper.

$$\lim_{V_{k+1|k}\to\mathbf{CQC}^\top} \mathrm{AUC}_0 = \lim_{R\to 0} \mathrm{AUC}_0 = 1 - \frac{2}{\pi}\arctan\left(\frac{\sqrt{(\mathbf{CP}_{ss}^{L}\mathbf{C}^\top)(\mathbf{CQC}^\top)}}{\mathbf{CAP}_{ss}^{L}\mathbf{A}^\top\mathbf{C}^\top}\right). \tag{73}$$

All of the closed-form approximation results that have been presented so far have been for the case in which $L = 0$ and $d = 1$. However, this extreme case can also be used to gain some useful insight into the AUC performance for non-zero L values. The Signal-to-noise ratio (SNR) can be used to provide this insight by adjusting the value of R appropriately, recognizing that $\lim_{R\to 0}\mathrm{SNR} = \infty$, and $\lim_{R\to\infty}\mathrm{SNR} = 1$. The top two panels of Figure 3 illustrate that a limiting value of $\mathrm{AUC}_0 < 1$ exists as $R \to 0$ (SNR $\to \infty$) when $L > 0$, just as when $L = 0$ for both the closed-form and the root-finding approximations. The case $L = 0$ is not shown for the root-finding approximation, based on the previous finding that it is impossible to compute AUC_0 in this instance. When using the closed-form approximation, the bottommost line shown in blue on the top-left panel corresponds to the case when $L = 0$, directly illustrating the result from eqn (72). Observing the remaining curves in the top two panels when $L > 0$, it is clear that as the measurement noise increases (the SNR decreases) $\lim_{R\to\infty}$ AUC $= 0.5$, and that as measurement noise decreases (the SNR increases), $\lim_{R\to 0} AUC = 1$ as $L \to \infty$. The latter observation provides evidence of the large value level-crossing prediction derivation provided earlier that $\mathrm{AUC}_\infty = 1$. These observations in general serve only as preliminary empirical evidence to support characterization of the limiting values for AUC when addressing measurement noise, R, as well as the critical level, L. A more detailed theoretical analysis will be presented in a sequel paper.

When $L = 0$ and $d = 1$, important observations can also be made for the baseline redline and predictive methods in face of the limiting consideration $R \to \infty$ as shown in eqns (74) and (75). These observations are empirically evident in the bottom two panels of

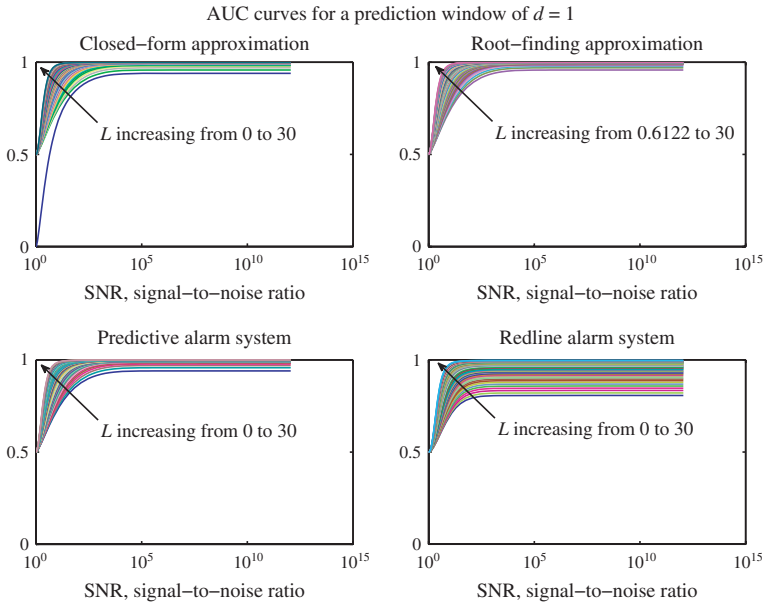

Figure 3. AUC for all alarm systems a function SNR

Figure 3, again with the bottommost line shown in blue. These equations can also be derived by a straightforward application of the specific conditions given by $L = 0$ and $d = 1$, and using the definitions of a_{red} and a_{pred} provided in eqns (49) and (50), as well as eqn (52).

$$\lim_{R\to\infty} \text{AUC}_0 = \frac{1}{2}, \tag{74}$$

$$\lim_{R\to\infty} \text{AUC}_0 = \frac{1}{2}. \tag{75}$$

As the measurement noise decreases, limiting values for AUC ($\text{AUC}_0 < 1$) are given by eqns (76) and (77) for the redline and predictive methods respectively and are analogous to eqn (73) for the closed-form approximation.

$$\lim_{V_{k+1|k}\to\mathbf{C}\mathbf{Q}\mathbf{C}^\top} \text{AUC}_0 = \lim_{R\to 0} \text{AUC}_0 = 1 - \frac{2}{\pi}\arctan\left(\sqrt{1 - \left(\frac{\mathbf{C}\mathbf{P}_{ss}^L\mathbf{A}^\top\mathbf{C}^\top}{\mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^\top}\right)^2}\right), \tag{76}$$

$$\lim_{V_{k+1|k}\to\mathbf{C}\mathbf{Q}\mathbf{C}^\top} \text{AUC}_0 = \lim_{R\to 0} \text{AUC}_0 = 1 - \frac{2}{\pi}\arctan\left(\sqrt{1 - \frac{\left(\mathbf{C}\mathbf{P}_{ss}^L\mathbf{A}^\top\mathbf{C}^\top\right)^2}{\left(\mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^\top\right)\left(\mathbf{C}\mathbf{A}\mathbf{P}_{ss}^L\mathbf{A}^\top\mathbf{C}^\top\right)}}\right). \tag{77}$$

As in the analysis of the closed-form approximation, the extreme case of $L = 0$ can be used to gain some useful insight into the AUC performance for values of $L > 0$, as illustrated in the bottom two panels of Figure 3. In the more general case when $L > 0$ as measurement noise increases, empirically eqns (74) and (75) appear to hold true, indicating that the event prediction capability of level-crossings worsens to random guessing when $L > 0$ as well. Furthermore, the AUC appears to approach 1 as $L \to \infty$ and $R \to 0$, again providing evidence of the derivation provided earlier that $\text{AUC}_\infty = 1$.

It is also evident that neither the baseline redline nor predictive methods worsen to the extent that the closed-form approximation does, and they only degenerate to random guessing ($\text{AUC}_0 = 0.5$) as opposed to completely imperfect prediction ($\text{AUC}_0 = 0$). This is a particularly interesting result due to the fact that the performance of the predictive method has previously been proven as identical to the optimal level-crossing predictor when $d = 1$ (Svensson, 1998). This notable discrepancy manifests only when $L = 0$ by using the closed-form approximation.

However, the same conclusion that was drawn for the case of the closed-form approximation concerning the influence of value prediction uncertainty on event prediction capability also applies to the baseline redline and predictive methods. The natural and intuitive interpretation of eqns (74) and (75) is that as the measurement noise (R) increases, event prediction capability of level-crossings worsens, even when $L > 0$ in addition to when $L = 0$. As such, the closed-form representations of AUC_0 for all alarm systems which are relevant for specific conditions on L and d can be used to gain insight for how the AUC varies more generally in the presence of measurement noise.

5. CONCLUSIONS AND FUTURE WORK

In an earlier article, a novel state-space approach to the optimal alarm systems literature was introduced, which was enabled by use of the Kalman filter. In doing so, it was demonstrated that there is a negligible loss in overall accuracy when using approximations to the theoretically optimal predictor for a stationary linear Gaussian process, at the advantage of greatly reduced computational complexity. In this study, the limiting effects of using those approximations were investigated with an extreme value analysis by varying the critical threshold, L. Depending on the magnitude of the measurement noise, it was found when $L \to 0$, using the closed-form approximation will provide a worse result than the redline or predictive methods. As such, the original hypothesis postulated that care needs to be exercised when applying approximations to the optimal level-crossing prediction has held true. However, this extreme case was also used to gain useful insight into the AUC performance when $L > 0$. Specifically, when L increases, the predictive capability as quantified by the AUC improves, for which theoretical proof was also provided to support the finding of improvement to perfect prediction ($\text{AUC}_\infty = 1$). It was also found that closed-form expressions for the AUC can be derived as $L \to 0$. In particular, expressions were derived for the AUC when $d = 1$ using the closed-form approximation and when $d \geq 1$ using the redline and predictive methods. In doing so, an intuitive conclusion was recognized that increased measurement noise negatively impacts event prediction capability.

A closed-form expression for the AUC does not exist for the more general case of $d > 1$, nor can a limiting value for the AUC be derived for the root-finding approximation. Despite these minor drawbacks, it is clear that for all methods the AUC approaches 1 as $L \to \infty$. The implications of this and other findings can be summarized by stating that given the assumed technical conditions, level-crossings of a linear Gaussian process can be predicted with the greatest accuracy for extremely high levels or very low measurement noise.

In future work, limiting values for relevant statistics as R and d approach the extremities of their feasible domains will be examined in more depth. Doing so will help to facilitate a new and broader context for the design of an optimal alarm system related to important engineering design parameters. Furthermore, control theoretic implications and ramifications of using the Kalman filter in tandem with optimal alarm theory that naturally follow will be investigated. One such example is the implication of the effect of stability on predictive capability which was briefly alluded to in the results section. Following this investigation, it will also be possible to gain further insight into important engineering design considerations for both the analysis and synthesis of algorithms used for mitigation of potential adverse events from a practical standpoint. Relaxing some of the inherent assumptions made in this article to the point where non-parametric methods such as Gaussian process regression and particle filtering are accessible may also provide a natural vehicle for the extension of optimal alarm theory to other machine learning research domains. Finally, extension of this work to systems containing both multi-variate inputs and outputs is important and has practical appeal to the field of data mining. As such, scalability and complexity will remain important considerations in current and future work.

APPENDIX A: THEOREMS AND COROLLARIES

THEOREM 1. *False alarm probability for small value level-crossing prediction using closed-form approximation*

$$P_{fa}^0 = \begin{cases} 1 - P\left(\bigcap_{j=1}^d |\hat{y}_{k+j|k}| < L_{A_j}^0\right) & P_b > 0.5 \\ 1 & P_b = 0.5, \end{cases}$$

where

$$P\left(\bigcap_{j=1}^d |\hat{y}_{k+j|k}| < L_{A_j}^0\right) = \int_{-L_{A_1}^0}^{L_{A_1}^0} \cdots \int_{-L_{A_d}^0}^{L_{A_d}^0} \mathcal{N}\left(\hat{\mathbf{y}}_d; \mathbf{0}_d, \Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d}\Sigma_{\mathbf{y}_d}^{-1}\Sigma_{\hat{\mathbf{y}}_d}^\top\right) \mathrm{d}\hat{\mathbf{y}}_d,$$

$$L_{A_j}^0 \triangleq \sqrt{V_{k+j|k}}\Phi^{-1}(P_b).$$

PROOF. Equations (24), (38), (43) and (44) can be used to prove eqn (58), by using the series of equations shown below. It is clear from eqn (38) that when using the closed-form approximation, $P_{b_{\text{crit}}} = 0.5$.

$$\begin{aligned} P_{fa}^0 &= \lim_{L\to 0} \frac{P(A_k) - P(C_k, A_k)}{1 - P(C_k)} \\ &= \begin{cases} \lim_{L\to 0} \frac{P(C_k') - P(C_k', A_k')}{P(C_k')} & P_b > 0.5 \\ 1 & P_b = 0.5 \end{cases} \\ &= \begin{cases} \lim_{L\to 0} \frac{P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right) - P\left(\bigcap_{j=1}^d |y_{k+j}| < L, \bigcap_{j=1}^d |\hat{y}_{k+j|k}| < L_{A_j}\right)}{P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right)} & P_b > 0.5 \\ 1 & P_b = 0.5. \end{cases} \end{aligned}$$

Because P_{fa}^{0} results in an indeterminate form $\left(\frac{0}{0}\right)$ when $P_b > 0.5$, l'Hôpital's rule must be applied:

$$\begin{cases} \lim_{L\to 0} \dfrac{\frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right) - \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right)}{\frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right)} & P_b > 0.5 \\ 1 & P_b = 0.5 \end{cases}$$

To find $\frac{\partial^d}{\partial L^d} P(\bigcap_{j=1}^{d} |y_{k+j}| < L)$ it must be recognized that the form of the multi-variate cumulative probability function shown in eqn (78) must be employed. Similarly, to find the following:

$$\frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right),$$

the multi-variate cumulative probability function shown in eqn (79) must be employed.

$$P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right) = \sum_{j=1}^{2^d} \prod_{i=1}^{d} (-1)^{s_d(i,j)} P\left(\bigcap_{i=1}^{d} y_{k+i} < L(-1)^{s_d(i,j)}\right), \tag{78}$$

$$\begin{aligned} &P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) \\ &= \sum_{j=1}^{4^d} \prod_{i=1}^{2d} (-1)^{s_{2d}(i,j)} P\left(\bigcap_{i=1}^{d} y_{k+i} < L(-1)^{s_{2d}(i,j)}, \bigcap_{i=1}^{d} \hat{y}_{k+i|k} < L_{A_i}(-1)^{s_d(i,j)}\right), \end{aligned} \tag{79}$$

where

$$s_d(i,j) \triangleq \left\lceil \frac{j + 2^{d-i}}{2^{d-i}} \right\rceil,$$
$$s_{2d}(i,j) \triangleq \left\lceil \frac{j + 2^{2d-i}}{2^{2d-i}} \right\rceil.$$

and $\lceil \cdot \rceil$ is the 'ceiling' operator that denotes rounding to the next highest integer value. The next sequence of equalities follows from the standard definition of a multi-variate cumulative probability function as it relates to the multi-variate Gaussian density function through the fundamental theorem of calculus, with a need to use the chain rule as a consequence of the alternating signs of the integration limits.

$$\begin{aligned} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right) &= \sum_{j=1}^{2^d} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{i=1}^{d} y_{k+i} < L(-1)^{s_d(i,j)}\right) \\ &= \sum_{j=1}^{2^d} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \underbrace{\frac{\partial^d P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L(-1)^{s_d(i,j)}\right)}{\partial \mathbf{x}_L}}_{\mathcal{N}(\mathbf{x}_L; \mathbf{0}_d, \Sigma_{\mathbf{y}_d})} \cdot \underbrace{\frac{\partial \mathbf{x}_L}{\partial L^d}}_{\prod_{i=1}^{d} (-1)^{s_d(i,j)}} \\ &= \sum_{j=1}^{2^d} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \mathcal{N}(\mathbf{x}_L; \mathbf{0}_d, \Sigma_{\mathbf{y}_d}), \end{aligned}$$

where

$$\mathbf{x}_L = \begin{bmatrix} L(-1)^{s_d(1,j)} \\ L(-1)^{s_d(2,j)} \\ \vdots \\ L(-1)^{s_d(d,j)} \end{bmatrix}.$$

Thus,

$$\begin{aligned} \lim_{L\to 0} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L\right) &= \sum_{j=1}^{2^d} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \mathcal{N}(\mathbf{0}_d; \mathbf{0}_d, \Sigma_{\mathbf{y}_d}), \\ &= \sum_{j=1}^{2^d} \prod_{i=1}^{d} (-1)^{s_d(i,j)} \frac{\prod_{i=1}^{d} (-1)^{s_d(i,j)}}{(2\pi)^{\frac{d}{2}} |\Sigma_{\mathbf{y}_d}|^{\frac{1}{2}}}. \end{aligned} \tag{80}$$

Similarly,

$$\begin{aligned}
&\frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) \\
&= \sum_{j=1}^{4^d} \prod_{i=1}^{2d} (-1)^{s_{2d}(i,j)} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{i=1}^{d} y_{k+i} < L(-1)^{s_{2d}(i,j)}, \bigcap_{i=1}^{d} \hat{y}_{k+i|k} < L_{A_i}(-1)^{s_d(i,j)}\right) \\
&= \sum_{j=1}^{4^d} \prod_{i=1}^{2d} (-1)^{s_{2d}(i,j)} \underbrace{\frac{\partial^d P\left(\bigcap_{i=1}^{d} y_{k+i} < L(-1)^{s_{2d}(i,j)}, \bigcap_{i=1}^{d} \hat{y}_{k+i|k} < L_{A_i}(-1)^{s_d(i,j)}\right)}{\partial \mathbf{x}_{2L}}}_{P\left(\bigcap_{i=1}^{d} \hat{y}_{k+i|k} < L_{A_i}(-1)^{s_d(i,j)} \,\middle|\, \bigcap_{i=1}^{d} y_{k+i} = L(-1)^{s_d(i,j)}\right) \mathcal{N}(\mathbf{x}_L; \mathbf{0}_d, \Sigma_{\hat{\mathbf{y}}_d})} \cdot \underbrace{\frac{\partial \mathbf{x}_{2L}}{\partial L^d}}_{\prod_{i=1}^{d} (-1)^{s_{2d}(i,j)}} \\
&= \sum_{j=1}^{4^d} \prod_{i=1}^{2d} (-1)^{s_{2d}(i,j)} \prod_{i=1}^{d} (-1)^{s_{2d}(i,j)} \mathcal{N}(\mathbf{x}_L; \mathbf{0}_d, \Sigma_{\hat{\mathbf{y}}_d}) p_j,
\end{aligned}$$

where the mean, $\mathbf{x}_L$, and covariance, $\Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} \Sigma_{\hat{\mathbf{y}}_d}^{\top}$, necessary for computation of the integral for p_j below were obtained by using the Schur complement, and $\mathbf{x}_{2L}$ is defined below.

$$\begin{aligned}
\mathbf{x}_{2L} &\triangleq \begin{bmatrix} L(-1)^{s_{2d}(1,j)} \\ L(-1)^{s_{2d}(2,j)} \\ \vdots \\ L(-1)^{s_{2d}(d,j)} \end{bmatrix} \\
p_j &\triangleq P\left(\bigcap_{i=1}^{d} \hat{y}_{k+i|k} < L_{A_i}(-1)^{s_d(i,j)} \,\middle|\, \bigcap_{i=1}^{d} y_{k+i} = L(-1)^{s_d(i,j)}\right) \\
&= \int_{-\infty}^{L_{A_1}(-1)^{s_d(1,j)}} \cdots \int_{-\infty}^{L_{A_d}(-1)^{s_d(d,j)}} \mathcal{N}(\hat{\mathbf{y}}_d; \mathbf{x}_L, \Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} \Sigma_{\hat{\mathbf{y}}_d}^{\top}) \, \mathrm{d}\hat{\mathbf{y}}_d.
\end{aligned}$$

Use of the Schur complement to obtain the conditional mean and covariance of $\hat{\mathbf{y}}_d$, given $\mathbf{y}_d$, is shown below.

$$\begin{aligned}
E[\hat{\mathbf{y}}_d \,|\, \mathbf{y}_d] &= E[\mathbf{y}_d] + \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} (\mathbf{x}_L - E[\mathbf{y}_d]) \\
&= \mu_{\mathbf{y}_d} + \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} (\mathbf{x}_L - \mu_{\mathbf{y}_d}) \\
&= \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} \mathbf{x}_L \\
VC[\hat{\mathbf{y}}_d \,|\, \mathbf{y}_d] &= \Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} \Sigma_{\hat{\mathbf{y}}_d}^{\top}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\lim_{L \to 0} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^{d} |y_{k+j}| < L, \bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L_{A_j}\right) &= \sum_{j=1}^{4^d} \prod_{i=1}^{2d} (-1)^{s_{2d}(i,j)} \prod_{i=1}^{d} (-1)^{s_{2d}(i,j)} \mathcal{N}(\mathbf{0}_d; \mathbf{0}_d, \Sigma_{\hat{\mathbf{y}}_d}) p_j^0 \\
&= \sum_{j=1}^{4^d} \prod_{i=1}^{2d} (-1)^{s_{2d}(i,j)} \prod_{i=1}^{d} (-1)^{s_{2d}(i,j)} \frac{p_j^0}{(2\pi)^{\frac{d}{2}} |\Sigma_{\mathbf{y}_d}|^{\frac{1}{2}}},
\end{aligned} \tag{81}$$

where

$$\begin{aligned}
p_j^0 &\triangleq \int_{-\infty}^{L_{A_1}^0(-1)^{s_d(1,j)}} \cdots \int_{-\infty}^{L_{A_d}^0(-1)^{s_d(d,j)}} \mathcal{N}(\hat{\mathbf{y}}_d; \mathbf{0}_d, \Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d} \Sigma_{\mathbf{y}_d}^{-1} \Sigma_{\hat{\mathbf{y}}_d}^{\top}) \, \mathrm{d}\hat{\mathbf{y}}_d, \\
L_{A_j}^0 &\triangleq \sqrt{V_{k+j|k}} \Phi^{-1}(P_b).
\end{aligned}$$

Equation (82) is the result of combining eqns (80) and (81), which can be simplified by replacing the coefficient for p_j^0 in the numerator of the expression shown in eqn (82) as follows:

$$\begin{aligned}\prod_{i=1}^{2d}(-1)^{s_{2d}(i,j)}\prod_{i=1}^{d}(-1)^{s_{2d}(i,j)} &= \prod_{i=1}^{d}(-1)^{2s_{2d}(i,j)}\prod_{i=1}^{d}(-1)^{s_{2d}(i+d,j)}\\ &= \prod_{i=1}^{d}(-1)^{2s_{2d}(i,j)}(-1)^{s_{2d}(i+d,j)}\\ &= \prod_{i=1}^{d}(-1)^{2s_{2d}(i,j)}(-1)^{s_{d}(i,j)}\\ &= \prod_{i=1}^{d}(-1)^{\mathbb{Z}^{+}_{\text{even}}}(-1)^{s_{d}(i,j)}\\ &= \prod_{i=1}^{d}(-1)^{s_{d}(i,j)}.\end{aligned}$$

The simplification requires using one or more of the following identities (recognizing that all are true by definition):

$$\begin{aligned}s_{2d}(i+d,j) &= s_d(i,j),\\ s_d(i,j) &\in \mathbb{Z}^{+},\\ 2s_d(i,j) &\in \mathbb{Z}^{+}_{\text{even}},\\ 2s_{2d}(i,j) &\in \mathbb{Z}^{+}_{\text{even}}.\end{aligned}$$

The coefficient in the denominator of the expression shown in eqn (82) can easily be simplified as follows:

$$\begin{aligned}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)} &= \prod_{i=1}^{d}(-1)^{2s_{d}(i,j)}\\ &= \prod_{i=1}^{d}(-1)^{\mathbb{Z}^{+}_{\text{even}}}\\ &= 1.\end{aligned}$$

The overall simplification is therefore as follows:

$$\begin{aligned}&\lim_{L\to 0}\frac{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)-\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L,\bigcap_{j=1}^{d}|\hat{y}_{k+j|k}|<L_{A_j}\right)}{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)}\\ &=1-\frac{\sum_{j=1}^{4^d}\prod_{i=1}^{2d}(-1)^{s_{2d}(i,j)}\prod_{i=1}^{d}(-1)^{s_{2d}(i,j)}p_j^0}{\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}}\end{aligned}\tag{82}$$

$$\begin{aligned}&=1-\frac{\sum_{j=1}^{4^d}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}p_j^0}{\sum_{j=1}^{2^d}1}\\ &=1-2^{-d}\sum_{j=1}^{4^d}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}p_j^0\\ &=1-2^{-d}\left[2^d\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}p_j^0\right]\\ &=1-\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_{d}(i,j)}p_j^0\\ &=1-P\left(\bigcap_{j=1}^{d}|\hat{y}_{k+j|k}|<L_{A_j}^0\right).\end{aligned}$$

The last step of the final simplification is true by definition, and the preceding step requires recognition of the fact that there are only 2^d possible ways that p_j^0 can be expressed, based on the total number of possible permutations of alternating signs of the integration limits. Thus,

$$\lim_{L\to 0}P_{fa}=\begin{cases}1-P\left(\bigcap_{j=1}^{d}|\hat{y}_{k+j|k}|<L_{A_j}^0\right) & P_b>0.5\\ 1 & P_b=0.5.\end{cases}$$

COROLLARY 1. *False alarm probability for small value level-crossing prediction using closed-form approximation when* $d = 1$

$$\lim_{L\to 0} P_{fa} = \begin{cases} 2(1-\Phi(b_{cf}\Phi^{-1}(P_b))) & P_b > 0.5 \\ 1 & P_b = 0.5. \end{cases}$$

PROOF. From Theorem 1, eqn (58) holds

$$\lim_{L\to 0} P_{fa} = \begin{cases} 1 - P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L^0_{A_j}\right) & P_b > 0.5 \\ 1 & P_b = 0.5 \end{cases}$$

When $d = 1$, the following simplification can be made:

$$\begin{aligned} P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L^0_{A_j}\right) &= P\left(|\hat{y}_{k+1|k}| < L^0_{A_1}\right) \\ &= P\left(-L^0_{A_1} < \hat{y}_{k+1|k} < L^0_{A_1}\right) \\ &= \Phi\left(\frac{L^0_{A_1}}{\sqrt{\Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d}\Sigma^{-1}_{\mathbf{y}_d}\Sigma^{\top}_{\hat{\mathbf{y}}_d}}}\right) - \Phi\left(\frac{-L^0_{A_1}}{\sqrt{\Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d}\Sigma^{-1}_{\mathbf{y}_d}\Sigma^{\top}_{\hat{\mathbf{y}}_d}}}\right) \\ &= 2\Phi\left(\frac{L^0_{A_1}}{\sqrt{\Sigma_{\hat{\mathbf{y}}_d} - \Sigma_{\hat{\mathbf{y}}_d}\Sigma^{-1}_{\mathbf{y}_d}\Sigma^{\top}_{\hat{\mathbf{y}}_d}}}\right) - 1 \\ &= 2\Phi\left(\frac{\sqrt{V_{k+1|k}}\Phi^{-1}(P_b)}{\sqrt{(\lambda_c - V_{k+1|k}) - \frac{(\lambda_c - V_{k+1|k})^2}{\lambda_c}}}\right) - 1 \\ &= 2\Phi\left(\sqrt{\frac{\lambda_c}{\lambda_c - V_{k+1|k}}}\Phi^{-1}(P_b)\right) - 1 \\ &= 2\Phi(b_{cf}\Phi^{-1}(P_b)) - 1, \end{aligned}$$

where

$$\begin{aligned} b_{cf} &\triangleq \sqrt{1 + a^2_{cf}}, \\ a_{cf} &\triangleq \sqrt{\frac{V_{k+1|k}}{\lambda_c - V_{k+1|k}}}. \end{aligned}$$

Thus,

$$\begin{aligned} \lim_{L\to 0} P_{fa} &= \begin{cases} 1 - P\left(\bigcap_{j=1}^{d} |\hat{y}_{k+j|k}| < L^0_{A_j}\right) & P_b > 0.5 \\ 1 & P_b = 0.5 \end{cases} \\ &= \begin{cases} 1 - (2\Phi(b_{cf}\Phi^{-1}(P_b)) - 1) & P_b > 0.5 \\ 1 & P_b = 0.5 \end{cases} \\ &= \begin{cases} 2(1 - \Phi(b_{cf}\Phi^{-1}(P_b))) & P_b > 0.5 \\ 1 & P_b = 0.5. \end{cases} \end{aligned}$$

THEOREM 2. *AUC for small value level-crossing prediction using closed-form approximation when* $d = 1$ □

$$\begin{aligned} \mathrm{AUC}_0 &= \int_0^1 P^0_d(P^0_{fa})\, \mathrm{d}P^0_{fa} \\ &= 1 - \frac{2}{\pi}\arctan\left(\frac{\sqrt{\lambda_c V_{k+1|k}}}{\lambda_c - V_{k+1|k}}\right). \end{aligned}$$

PROOF. Using eqn (51),

$$\begin{aligned}\mathrm{AUC}_0 &= \int_0^1 P_d^0(P_{fa}^0)\,\mathrm{d}P_{fa}^0 \\ &= \int_0^1 2\Phi\left(-a_{cf}b_{cf}\Phi^{-1}\left(1-\frac{P_{fa}^0}{2}\right)\right)\mathrm{d}P_{fa}^0.\end{aligned}$$

Now making the following variable substitutions:

$$\begin{aligned}u &= 1-\frac{P_{fa}^0}{2},\\ q &= a_{cf}b_{cf}\Phi^{-1}(u),\\ m &= \frac{q}{a_{cf}b_{cf}}.\end{aligned}$$

Thus, $u = \Phi(m)$. Now, taking derivatives the following relationships can be obtained:

$$\begin{aligned}\frac{\mathrm{d}u}{dP_{fa}^0} &= -\frac{1}{2},\\ \frac{\mathrm{d}u}{\mathrm{d}m} &= \phi(m),\\ \frac{\mathrm{d}m}{\mathrm{d}q} &= \frac{1}{a_{cf}b_{cf}},\\ \frac{\mathrm{d}P_{fa}^0}{\mathrm{d}q} &= \frac{\mathrm{d}P_{fa}^0}{\mathrm{d}u}\frac{\mathrm{d}u}{\mathrm{d}m}\frac{\mathrm{d}m}{\mathrm{d}q}\\ &= -\frac{2\phi(m)}{a_{cf}b_{cf}}.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{AUC}_0 &= 4\int_0^\infty \Phi(-q)\frac{1}{a_{cf}b_{cf}}\phi\left(\frac{q}{a_{cf}b_{cf}}\right)\mathrm{d}q\\ &= 4\int_0^\infty\int_{-\infty}^{-q}\phi(s)\frac{1}{a_{cf}b_{cf}}\phi\left(\frac{q}{a_{cf}b_{cf}}\right)\mathrm{d}s\,\mathrm{d}q\\ &= 4P(s+q<0, q>0)\\ &= 4P(\mathbf{b}^\top\mathbf{z}<0, q>0),\end{aligned}$$

where

$$\mathbf{b} = \begin{bmatrix}1\\1\end{bmatrix},\quad \mathbf{z} = \begin{bmatrix}s\\q\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & (a_{cf}b_{cf})^2\end{bmatrix}\right),$$

$$\ell \triangleq \mathbf{b}^\top\mathbf{z} \sim \mathcal{N}(0, 1+(a_{cf}b_{cf})^2).$$

Let

$$\tilde{\mathbf{z}} = \begin{bmatrix}\ell\\q\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1+(a_{cf}b_{cf})^2 & b_{cf}^2\\ b_{cf}^2 & (a_{cf}b_{cf})^2\end{bmatrix}\right)$$

Finally,

$$\begin{aligned}\mathrm{AUC} &= 4P(\ell<0, q>0)\\ &= 4(P(q>0)-P(\ell>0, q>0))\\ &= 4\Phi(0)-4\left(\frac{1}{4}+\frac{1}{2\pi}\arcsin\left(\frac{a_{cf}b_{cf}}{\sqrt{1+(a_{cf}b_{cf})^2}}\right)\right)\\ &= 1-\frac{2}{\pi}\arcsin\left(\frac{a_{cf}b_{cf}}{\sqrt{1+(a_{cf}b_{cf})^2}}\right)\\ &= 1-\frac{2}{\pi}\arcsin\left(\frac{c}{\sqrt{1+c^2}}\right)\\ &= 1-\frac{2}{\pi}\arctan c\\ &= 1-\frac{2}{\pi}\arctan\left(\frac{\sqrt{\lambda_c V_{k+1|k}}}{\lambda_c - V_{k+1|k}}\right),\end{aligned}$$

where

$$c \triangleq a_{cf} b_{cf} = \frac{\sqrt{\lambda_c V_{k+1|k}}}{\lambda_c - V_{k+1|k}}.$$

THEOREM 3. *False alarm probability for small value level-crossing prediction using redline method*

$$\lim_{L\to 0} P_{fa} = 2\Phi\left(\frac{-L_A}{a_{\text{red}}\sqrt{\lambda_c}}\right),$$

where

$$a_{\text{red}} \triangleq \sqrt{1 - \frac{\Lambda_c \Sigma_{\mathbf{y}_d}^{-1} \Lambda_c^\top}{\lambda_c}}.$$

PROOF.

$$\begin{aligned} P_{fa}^0 &= \lim_{L\to 0} \frac{P(A_k) - P(C_k, A_k)}{1 - P(C_k)} \\ &= \lim_{L\to 0} \frac{P(C_k') - P(C_k', A_k')}{P(C_k')} \\ &= \lim_{L\to 0} \frac{P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right) - P\left(\bigcap_{j=1}^d |y_{k+j}| < L, |y_k| < L_A\right)}{P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right)}. \end{aligned}$$

Because P_{fa}^0 results in an indeterminate form $\left(\frac{0}{0}\right)$, l'Hôpital's rule must be applied:

$$\lim_{L\to 0} \frac{\frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right) - \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^d |y_{k+j}| < L, |y_k| < L_A\right)}{\frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right)}.$$

From Theorem 1, eqn (80) yields:

$$\lim_{L\to 0} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^d |y_{k+j}| < L\right) = \sum_{j=1}^{2^d} \prod_{i=1}^d (-1)^{s_d(i,j)} \frac{\prod_{i=1}^d (-1)^{s_d(i,j)}}{(2\pi)^{\frac{d}{2}} |\Sigma_{\mathbf{y}_d}|^{\frac{1}{2}}},$$

and the following holds:

$$\begin{aligned} \lim_{L\to 0} \frac{\partial^d}{\partial L^d} P\left(\bigcap_{j=1}^d |y_{k+j}| < L, |y_k| < L_A\right) &= \lim_{L\to 0} \sum_{j=1}^{2^d} \prod_{i=1}^d (-1)^{s_d(i,j)} \prod_{i=1}^d (-1)^{s_d(i,j)} \mathcal{N}(\mathbf{x}_L; \mathbf{0}_d, \Sigma_{\mathbf{y}_d}) p_j \\ &= \sum_{j=1}^{2^d} \prod_{i=1}^d (-1)^{s_d(i,j)} \frac{\prod_{i=1}^d (-1)^{s_d(i,j)}}{(2\pi)^{\frac{d}{2}} |\Sigma_{\mathbf{y}_d}|^{\frac{1}{2}}} p_j^0, \end{aligned}$$

where

$$p_j \triangleq P\left(|y_k| < L_A \,\middle|\, \bigcap_{i=1}^d y_{k+i} = L(-1)^{s_d(i,j)}\right),$$

$$p_j^0 \triangleq 2\Phi\left(\frac{L_A}{\sqrt{\lambda_c - \Lambda_c \Sigma_{\mathbf{y}_d}^{-1} \Lambda_c^\top}}\right) - 1,$$

and use of the Schur complement to obtain the mean and variance of y_k conditioned on $\mathbf{y}_d$ is shown below.

$$\begin{aligned} E[y_k \,|\, \mathbf{y}_d] &= E[y_k] + \Lambda_c \Sigma_{\mathbf{y}_d}^{-1} (\mathbf{x}_L - E[\mathbf{y}_d]) \\ &= \mu_{y_k} + \Lambda_c \Sigma_{\mathbf{y}_d}^{-1} (\mathbf{x}_L - \mu_{\mathbf{y}_d}) \\ &= \Lambda_c \Sigma_{\mathbf{y}_d}^{-1} \mathbf{x}_L, \\ VC[y_k \,|\, \mathbf{y}_d] &= \lambda_c - \Lambda_c \Sigma_{\mathbf{y}_d}^{-1} \Lambda_c^\top. \end{aligned}$$

Thus, the overall simplification is as follows:

$$\lim_{L\to 0}\frac{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)-\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L,|y_k|<L_A\right)}{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)}$$
$$=1-\frac{\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\prod_{i=1}^{d}(-1)^{s_d(i,j)}p_j^0}{\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\prod_{i=1}^{d}(-1)^{s_d(i,j)}}$$

$$\begin{aligned}&=1-2^{-d}\sum_{j=1}^{2^d}p_j^0\\&=1-2^{-d}2^dp_j^0\\&=1-p_j^0\\&=1-\left(2\Phi\left(\frac{L_A}{\sqrt{\lambda_c-\Lambda_c\Sigma_{\mathbf{y}_d}^{-1}\Lambda_c^\top}}\right)-1\right)\\&=2\Phi\left(\frac{-L_A}{a_{\text{red}}\sqrt{\lambda_c}}\right),\end{aligned}$$

where

$$a_{\text{red}}\triangleq\sqrt{1-\frac{\Lambda_c\Sigma_{\mathbf{y}_d}^{-1}\Lambda_c^\top}{\lambda_c}}.$$

THEOREM 4. *False alarm probability for small value level-crossing prediction using predictive method*

$$\lim_{L\to 0}P_{fa}=2\Phi\left(\frac{-L_A}{a_{\text{pred}}\sqrt{\lambda_c-V_{k+d|k}}}\right),$$

where

$$a_{\text{pred}}\triangleq\sqrt{1-\frac{\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}\Lambda_a^\top}{\lambda_c-V_{k+d|k}}}$$

PROOF.

$$\begin{aligned}P_{fa}^0&=\lim_{L\to 0}\frac{P(A_k)-P(C_k,A_k)}{1-P(C_k)}\\&=\lim_{L\to 0}\frac{P(C_k')-P(C_k',A_k')}{P(C_k')}\\&=\lim_{L\to 0}\frac{P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)-P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L,|\hat{y}_{k+d|k}|<L_A\right)}{P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)}.\end{aligned}$$

Because P_{fa}^0 results in an indeterminate form $\left(\frac{0}{0}\right)$, l'Hôpital's rule must be applied:

$$\lim_{L\to 0}\frac{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)-\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L,|\hat{y}_{k+d|k}|<L_A\right)}{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)}$$

From Theorem 1, eqn (80) yields:

$$\lim_{L\to 0}\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)=\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\frac{\prod_{i=1}^{d}(-1)^{s_d(i,j)}}{(2\pi)^{\frac{d}{2}}|\Sigma_{\mathbf{y}_d}|^{\frac{1}{2}}}$$

and the following holds:

$$\lim_{L\to 0}\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L,|\hat{y}_{k+d|k}|<L_A\right)=\lim_{L\to 0}\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\mathcal{N}(\mathbf{x}_L;\mathbf{0}_d,\Sigma_{\mathbf{y}_d})p_j$$
$$=\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\frac{\prod_{i=1}^{d}(-1)^{s_d(i,j)}}{(2\pi)^{\frac{d}{2}}|\Sigma_{\mathbf{y}_d}|^{\frac{1}{2}}}p_j^0,$$

where

$$p_j\triangleq P\left(|\hat{y}_{k+d|k}|<L_A\left|\bigcap_{i=1}^{d}y_{k+i}=L(-1)^{s_d(i,j)}\right.\right)$$
$$p_j^0\triangleq 2\Phi\left(\frac{L_A}{\sqrt{\lambda_c-V_{k+d|k}-\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}\Lambda_a^\top}}\right)-1,$$

and use of the Schur complement to obtain the mean and variance of $\hat{y}_{k+d|k}$ conditioned on $\mathbf{y}_d$ is shown below.

$$\begin{aligned}E[\hat{y}_{k+d|k}|\mathbf{y}_d]&=E[\hat{y}_{k+d|k}]+\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}(\mathbf{x}_L-E[\mathbf{y}_d])\\&=\mu_{y_k}+\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}(\mathbf{x}_L-\mu_{\mathbf{y}_d})\\&=\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}\mathbf{x}_L\\VC[\hat{y}_{k+d|k}|\mathbf{y}_d]&=\lambda_c-V_{k+d|k}-\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}\Lambda_a^\top\end{aligned}$$

Thus, the overall simplification is as follows:

$$\lim_{L\to 0}\frac{\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L\right)-\frac{\partial^d}{\partial L^d}P\left(\bigcap_{j=1}^{d}|y_{k+j}|<L,|\hat{y}_{k+d|k}|<L_A\right)}{\frac{\partial^d}{\partial L^d}P(\bigcap_{j=1}^{d}|y_{k+j}|<L)}$$
$$=1-\frac{\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\prod_{i=1}^{d}(-1)^{s_d(i,j)}p_j^0}{\sum_{j=1}^{2^d}\prod_{i=1}^{d}(-1)^{s_d(i,j)}\prod_{i=1}^{d}(-1)^{s_d(i,j)}}$$

$$\begin{aligned}&=1-2^{-d}\sum_{j=1}^{2^d}p_j^0\\&=1-2^{-d}2^dp_j^0\\&=1-p_j^0\\&=1-\left(2\Phi\left(\frac{L_A}{\sqrt{\lambda_c-V_{k+d|k}-\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}\Lambda_a^\top}}\right)-1\right)\\&=2\Phi\left(\frac{-L_A}{a_{\text{pred}}\sqrt{\lambda_c-V_{k+d|k}}}\right),\end{aligned}$$

where

$$a_{\text{pred}}\triangleq\sqrt{1-\frac{\Lambda_a\Sigma_{\mathbf{y}_d}^{-1}\Lambda_a^\top}{\lambda_c-V_{k+d|k}}}$$

THEOREM 5. *AUC for small value level-crossing prediction using redline or predictive method*

$$\text{AUC}_0=1-\frac{2}{\pi}\arctan\,a.$$

PROOF.

$$\begin{aligned}\text{AUC}_0&=\int_0^1 P_d^0(P_{fa}^0)\,\mathrm{d}P_{fa}^0\\&=\int_0^1 2\Phi\left(a\Phi^{-1}\left(\frac{P_{fa}^0}{2}\right)\right)\mathrm{d}P_{fa}^0.\end{aligned}$$

Now making the following variable substitutions:

$$q = a\Phi^{-1}\left(\frac{P_{fa}^0}{2}\right),$$
$$z = \frac{q}{a}.$$

Thus, $P_{fa}^0 = 2\Phi(z)$. Now, taking derivatives the following relationships can be obtained:

$$\frac{dP_{fa}^0}{dz} = 2\Phi(z),$$
$$\frac{dz}{dq} = \frac{1}{a},$$
$$\frac{dP_{fa}^0}{dq} = \frac{dP_{fa}^0}{dz}\frac{dz}{dq} = \frac{2\phi(z)}{a}.$$

Thus,

$$\begin{aligned}\text{AUC}_0 &= 4\int_{-\infty}^{0}\Phi(q)\frac{1}{a}\phi\left(\frac{q}{a}\right)dq \\ &= 4\int_{-\infty}^{0}\int_{-\infty}^{q}\phi(s)\frac{1}{a}\phi\left(\frac{q}{a}\right)dsdq \\ &= 4P(s-q<0, q<0) \\ &= 4P(\mathbf{b}^\top\mathbf{z}<0, q<0),\end{aligned}$$

where

$$\mathbf{b} = \begin{bmatrix}1\\-1\end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix}s\\q\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & a^2\end{bmatrix}\right)$$
$$\ell \triangleq \mathbf{b}^\top\mathbf{z} \sim \mathcal{N}(0, 1+a^2).$$

Let

$$\tilde{\mathbf{z}} = \begin{bmatrix}\ell\\q\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1+a^2 & -a^2\\-a^2 & a^2\end{bmatrix}\right).$$

Finally,

$$\begin{aligned}\text{AUC}_0 &= 4P(\ell<0, q<0) \\ &= 4(P(q<0) - P(\ell>0) + P(\ell>0, q>0)) \\ &= 4\left(\frac{1}{4} + \frac{1}{2\pi}\arcsin\left(\frac{-a^2}{\sqrt{a^2(1+a^2)}}\right)\right) \\ &= 1 - \frac{2}{\pi}\arcsin\left(\frac{a}{\sqrt{1+a^2}}\right) \\ &= 1 - \frac{2}{\pi}\arctan a.\end{aligned}$$

Acknowledgements

The author acknowledges the support of the Integrated Vehicle Health Management (IVHM) project and the Systemwide Safety Assurance Technologies (SSAT) project. Both the IVHM and SSAT projects were funded by the Aviation Safety Program of NASA's Aeronautics Research Mission Directorate. The author also thanks Dr. Nikunj Oza and Dr. Santanu Das for reviewing this article. Finally, the author thanks the reviewers selected by the editorial board of the *Journal of Time Series Analysis* for important points they made and for their valuable comments and criticisms that greatly helped to allow for the development of a more technically sound article.

REFERENCES

Bradley, A. P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30**(7), 1145–59.
Butas, J. P., Santi, L. M. and Aguilar, R. B. (2007) A tiered approach to J-2X health and status monitoring. In *Proceedings of the 54th Joint Army-Navy-NASA-Air Force Propulsion Meeting*. Johns Hopkins University, Baltimore, MD: Chemical Propulsion Information Analysis Center.
de Maré, J. (1980) Optimal prediction of catastrophes with application to Gaussian processes. *Annals of Probability* **8**(4), 840–50.

Genz, A. (1992) Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1**, 141–9.
Kerr, T. H. (1982) False alarm and correct detection probabilities over a time interval for restricted classes of failure detection algorithms. *IEEE Transactions on Information Theory* **IT-28**(4), 619–31.
Lindgren, G. (1985) Optimal prediction of level crossings in Gaussian processes and sequences. *Annals of Probability* **13**(3), 804–24.
Martin, R. A. (2010) A state-space approach to optimal level-crossing prediction for linear gaussian processes. *IEEE Transactions on Information Theory* **56**(10), 5083–96.
Svensson, A. (1998) Event Prediction and Bootstrap in Time Series. PhD thesis, Lund Institute of Technology.
Svensson, A., Holst, J., Lindquist, R. and Lindgren, G. (1996) Optimal prediction of catastrophes in autoregressive moving-average processes. *Journal of Time Series Analysis* **17**(5), 511–31.