Recent Results on “Approximations to Optimal Alarm Systems for Anomaly Detection”

Rodney A. Martin
NASA Ames Research Center
Mail Stop 269-1
Moffett Field, CA 94035-1000, USA
(650) 604-1334
Rodney.Martin@nasa.gov

***Abstract*— An optimal alarm system and its approximations may use Kalman filtering for univariate linear dynamic systems driven by Gaussian noise to provide a layer of predictive capability. Predicted Kalman filter future process values and a fixed critical threshold can be used to construct a candidate level-crossing event over a predetermined prediction window. An optimal alarm system can be designed to elicit the fewest false alarms for a fixed detection probability in this particular scenario.**

I. Introduction

Recent studies [11], [9] have served as a foundation for the application of a novel idea for anomaly detection that is derived from the collusion of decades-old theory [15],[2] with more recent techniques [16],[17]. It was shown by Svensson [16], [17] 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. The optimal alarm condition is fundamentally an alarm region or decision boundary based upon a likelihood ratio criterion via the Neyman-Pearson lemma, as shown in [3], [8]. This allows us to design an optimal alarm system that will elicit the fewest possible false alarms for a fixed detection probability.

Due to the fact that the alarm regions cannot be expressed in closed form, one of the aims of previous studies has been to investigate approximations for the design of an optimal alarm system. Such an alarm system uses Kalman filtering along with temporally varying auxiliary thresholds to provide a layer of predictive capability. The resulting metrics can easily be compared to methods that incorporate auxiliary fixed thresholds or redlines that may also provide a similar layer of predictive capability, but have no provision for minimizing false alarms.

The design of optimal alarm systems demonstrates potential to enhance reliability and support health management for space propulsion, civil aerospace applications, and more fundamentally to aeronautics research. Due to the great costs, not to mention potential dangers associated with a false alarm due to evasive or extreme action taken as a result of such a false indication, there are great opportunities for cost savings/cost avoidance, enhancement of overall safety, and reduction of technical risks of NASA programs and projects. Furthermore, within NASA’s space program, a missed detection can yield a catastrophic result of the loss of mission, crew, and/or vehicle that may be encountered when failing to abort in the presence of valid indicators.

Even though recent studies have been limited to application-specific datasets, our intent is to demonstrate the utility of the technique from a much broader perspective. In [11] level-crossing events of the type most amenable to monitoring of control system error were used to derive the design framework for an optimal alarm system via the ROC curve. From the applications perspective, we assume that the control system has already been designed, is robust to environmental disturbances, and rejects them expediently. Therefore, when unexpected large transients in the control system error occur, this may be indicative of an impending fault or change in system that may be cause for further diagnostic investigation. This error can be compared against a threshold whose selection is based upon the physics of the system and the margin of safety required. The threshold may also be determined from domain experts, experimentally, in flight tests, or by using statistical models.

Alternatively, a serial architecture can be used to preprocess a full feature space, implicitly reducing the entire feature space into a univariate signal while retaining salient operational signatures [10]. This is performed by using the composite score generated by any algorithm with favorable properties as training data for a linear dynamic system. This is potentially a far more effective approach than using only a small fraction of the feature space by using the control system error alone. As such we may potentially allow for many more anomalies to be detected by using this paradigm. Furthermore, allowing for this sort of preprocessing lifts the restriction of this algorithm to the control systems domain, and addresses our objective of demonstrating the utility of the technique from a much broader perspective.

II. Background

Coincidentally, the techniques investigated as part this research have their origins in application to legacy NASA platforms. Rudolf E. Kalman found a unique application of his now very well-known Kalman filter for the Apollo program and more broadly to aerospace applications in general, due in part to finding support at NASA Ames Research Center in the mid 1960’s [15].

Although tremendously popular and ubiquitous in today's aerospace systems, practical applications of Kalman filtering for aerospace have largely been relegated to state estimation for guidance, navigation, and control purposes. The study of auxiliary failure detection and bad data rejection algorithms have been developed in concert with Kalman filters [15], [18], [5], however the main purpose of those Kalman filters were for state estimation in guidance, navigation, and control systems.

Kalman filtering has seen limited practical application *dedicated* to system reliability and health management as related to exceedance of predetermined failure thresholds in aerospace systems. The difference in the approach that we take with this investigation is that the Kalman filter machinery will be implemented for the express purpose of system reliability and health management, invoking more recently available data mining and machine learning techniques [4], [12], [13], [10] to develop suitable models.

Almost in parallel with Kalman's breakthrough, a perhaps lesser known study, [6], was conducted by Ross Leadbetter and Harald Cramér who are pioneers in the field of the statistics of level crossings and extremes. This study was also funded by NASA, and yielded interesting results on the more theoretical aspects of level-crossing behavior of random processes. The motivation behind the work was as a result of Gertrude Cox's charter to Ross Leadbetter and Harald Cramér at the time to "make comprehensive statistical models for manned space-flight systems." They ended up supporting a small corner of that effort, having to do with the reliability of guidance systems, approaching the problem by modeling the error in a guidance system and declaring failure if it went out of prescribed limits in a mission period - leading to their work on crossings and extremes [7].

All three researchers are legendary, celebrated mathematicians/statisticians in their own right; however, the work was never truly developed to its fullest potential for its intended purpose. Over the years Leadbetter's younger Swedish colleagues developed theories which ultimately yielded the idea of optimal alarm systems [17], which is used in this study. There are still parts of Leadbetter's original theoretical constructs which have gone unused for its originally intended target application.

As such, a future research objective is to marry the largely uncultivated portions of Leadbetter's theory for its intended purpose and the results generated by his younger Swedish colleagues, enabled by none other than the Kalman filter, coming full circle. Therefore, with further development and implementation across a broad spectrum of NASA aerospace platforms, this activity also has the potential to generate new knowledge that has evolved from the results of NASA-based legacy programs.

III. Methodology

Our underlying assumption is that we can fit measured or transformed data 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 both the state and the covariance matrix with time-invariant parameters.

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{w}_k \qquad (1)\\ y_k &= \mathbf{C}\mathbf{x}_k + v_k \qquad (2)\\ \mathbf{P}_{k+1} &= \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{Q} \qquad (3)\end{aligned}$$

where

$$\begin{aligned}\mathbf{w}_k &\sim \mathcal{N}(0,\mathbf{Q})\\ v_k &\sim \mathcal{N}(0,R)\\ \mathbf{x}_0 &\sim \mathcal{N}(\mu_{\mathbf{x}},\mathbf{P}_0)\\ \mu_{\mathbf{x}} &= E[\mathbf{x}_k]\\ \mathbf{P}_k &= E[(\mathbf{x}_k-\mu_{\mathbf{x}})(\mathbf{x}_k-\mu_{\mathbf{x}})^T]\end{aligned}$$

The parameters to be learned are specified below, as the parameter θ. These parameters are also shown in Fig. 1, which specify them in relation to the probabilistic graphical modeling paradigm which may be used for machine learning purposes.

$$\theta = (\mu_{\mathbf{x}}, \mathbf{P}_0, \mathbf{A}, \mathbf{C}, \mathbf{Q}, R) \qquad (4)$$

The essence of the optimal alarm system is derived from the use of the likelihood ratio resulting in the conditional inequality: $P(C_k|y_0,\ldots,y_k) \geq P_b$. This basically says "give alarm when the conditional probability of the event, C_k, exceeds the level P_b." Here, P_b represents some optimally chosen border or threshold probability with respect to a relevant alarm system metric. It is necessary to find the alarm regions in order to design the alarm system. The event, C_k, can be chosen arbitrarily, and is usually defined with respect to a pre-specified critical threshold, L, as well as a prediction window, d. In this paper, the event of interest is shown in Eqn. 5, and 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.

$$C_k \triangleq \{|y_k| > L\} \bigcup \left[\bigcup_{j=1}^{d}\left[\bigcap_{i=0}^{j-1} |y_{k+i}| < L, |y_{k+j}| > L\right]\right] \qquad (5)$$

There are three different alarm systems to compare which will all attempt to predict the level-crossing event defined by Eqn. 5, whose probability, $P(C_k)$, can be computed according to formulae presented in [11]. The first alarm system attempts to define an envelope, $[-L_A, L_A]$, outside of which an alarm will activate. In order to provide for a layer of predictive capability, L_A should be chosen such that $L_A < L$. An alarm probability can likewise be computed, $P(A_k) = P(|y_k| > L_A)$ and the details of this formula are also provided in [11]. This "redline" alarm system is termed as such in order to indicate that a simple level is used, and often the same terminology is used in practice. Even without the benefit of using any predicted future process values, this alarm system would be superior to a true redline system that uses only a

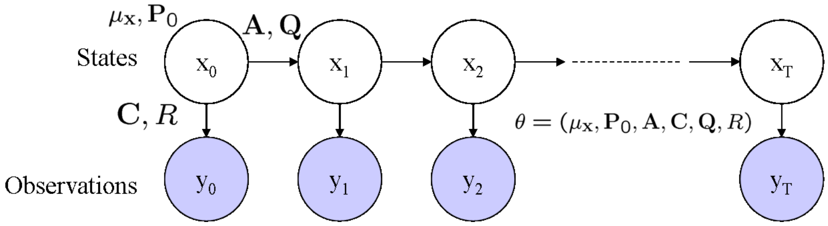


Fig. 1. Linear Dynamic System

single level L. However, in this case two levels are used, L as the failure threshold, and L_A as the design threshold.

The second alarm system incorporates the use of predicted future process values, and is called the "predictive" alarm system. This alarm system also defines an envelope, $[-L_A, L_A]$, outside of which an alarm will sound. Similarly, L_A should be chosen such that $L_A < L$ in order to provide for a layer of predictive capability. However, the alarm probability is defined in a different fashion than for the redline method, as $P(A_k) = P(|\hat{y}_{k+d|k}| > L_A)$, where the predicted future process value $\hat{y}_{k+d|k}$ is found from standard Kalman filter equations shown in Eqns. 6 - 11 by using the definitions below.

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &\triangleq E[\mathbf{x}_k | y_0, \ldots, y_k] \\ \mathbf{P}_{k|k} &\triangleq E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T | y_0, \ldots, y_k] \end{aligned}$$

$$\begin{aligned} \hat{y}_{k|k} &= \mathbf{C}\hat{\mathbf{x}}_{k|k} && (6) \\ \hat{\mathbf{x}}_{k+1|k} &= \mathbf{A}\hat{\mathbf{x}}_{k|k} && (7) \\ \mathbf{F}_{k+1|k} &\triangleq \mathbf{P}_{k+1|k}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k+1|k}\mathbf{C}^T + R)^{-1} && (8) \\ \mathbf{P}_{k+1|k} &= \mathbf{A}\mathbf{P}_{k|k}\mathbf{A}^T + \mathbf{Q} && (9) \\ \mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k} - \mathbf{F}_{k+1|k}\mathbf{C}\mathbf{P}_{k+1|k} && (10) \end{aligned}$$

Eqn. 8 represents the dynamically updated Kalman gain, and combining the two equations 9 and 10, we may obtain the Riccati equation (Eqn. 11).

$$\mathbf{P}_{k+1|k} = \mathbf{A}\mathbf{P}_{k|k-1}\mathbf{A}^T - \mathbf{A}\mathbf{F}_{k|k-1}\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{A}^T + \mathbf{Q} \quad (11)$$

The final alarm system to be compared to the previous two is the optimal alarm system, and has two approximations, but only the one presented as Eqn. 12 will be used for comparison in this paper. The alarm condition, $P(C_k | y_0, \ldots, y_k) \geq P_b$, can be approximated to form the alarm region specified in Eqn. 12.

$$A_k = \bigcup_{i=0}^{d} |\hat{y}_{k+i|k}| \geq L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b) \quad (12)$$

where $\Phi^{-1}(\cdot)$ represents the inverse cumulative normal standard distribution function, and $V_{k+i|k} = \mathrm{Var}(y_{k+i} | y_0, \ldots, y_k)$.

Eqn. 12 plays a pivotal role in enabling the enforcement of the approximation to the alarm region for an optimal alarm system. Using this approximation allows it to outperform the other alarm systems with respect to the minimization of false alarms. All of the three alarm systems described will be compared using the area under the ROC curve (AUC). This provides a performance metric with which to assess and compare the performance of each alarm system. The ROC curve parametrically displays the true positive rate against the false positive rate. The AUC has been deemed as a theoretically valid metric for model selection and algorithmic comparison [14].

The parameters of interest are L_A for the redline and predictive methods, and P_b for the approximation to the optimal alarm system. It is possible to generate formulae for the true and false positive rates as a function of these parameters (L_A, P_b) as well as the model parameters (θ) by appealing to Eqns. 13-14. The details for constructing these formulae are provided in [11].

True positive rate:

$$P(C_k | A_k) = \frac{P(C_k, A_k)}{P(A_k)} \quad (13)$$

False positive rate:

$$P(A_k | C_k') = \frac{P(C_k', A_k)}{P(C_k')} \quad (14)$$

IV. Results

The example to be used for the presentation of our results has no specific application, but is generic, and the model parameters are provided in Eqns. 15-18.

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 0 & 1 \\ -0.9 & 1.8 \end{bmatrix} && (15) \\ \mathbf{C} &= \begin{bmatrix} 0.5 & 1 \end{bmatrix} && (16) \\ \mathbf{Q} &\triangleq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} && (17) \\ R &\triangleq 0.08 && (18) \end{aligned}$$

Unless otherwise stated, for all three cases to compare: redline, predictive, and optimal, the threshold is $L = 16$, and the prediction window is $d = 5$. Fig. 2 represents the optimal alarm region decision boundary for a sample system and two level-crossing events that span a prediction window of three time steps. The figure shown on the right is of the same form that we are investigating in Eqn. 12. Approximations to this sort of alarm region are required for the most computationally efficient generation of a ROC curve or other similar alarm system design metrics.

Some recent results of computing the AUC as a function of the prediction window, d, are shown on the left of Fig. 3. We show the AUC for the three methods described thus far to be compared. Clearly, the approximations to the optimal alarm system outperform the redline and predictive methods, for the entire prediction horizon. This figure can also be used as a preliminary design step for choice of maximal prediction window corresponding to a minimum allowable AUC as the criterion for selection.

For example, if AUC$_{min} = 0.95$ is set as the minimum allowable AUC, the maximal prediction window is obtained by using the optimal alarm system, and corresponds to $d = 5$. The final design step will involve choosing the ROC curve corresponding to this maximal prediction window. Using this ROC curve, a value of P_b can be selected based upon the desired tradeoff between true and false positive rates.

For contrast, shown on the right on Fig. 3 is a plot of the prediction variance, $V_{k+d|k}$, and the bounded uncertainty. $V_{k+d|k}$ is a function of model parameters as shown in Eqn. 19. Taking $\lim_{d\rightarrow\infty} V_{k+d|k} = \mathbf{C}\mathbf{P}_k\mathbf{C}^T + R$ provides the finite bound on uncertainty for an infinite prediction horizon, and as such represents the maximum uncertainty for predicted future process values. The finiteness of the bound is guaranteed only if $\rho(\mathbf{A}) < 1$, where $\rho(\cdot)$ is the spectral radius operator.

$$V_{k+d|k} = \mathbf{C}\left[\mathbf{A}^d(\mathbf{P}_{k|k} - \mathbf{P}_k)(\mathbf{A}^d)^T + \mathbf{P}_k\right]\mathbf{C}^T + R \quad (19)$$

Due to the assumption of time-invariance for our model parameters, we require the necessary and sufficient conditions of controllability or stabilizability of $(\mathbf{A}, \sqrt{\mathbf{Q}})$ and observability or detectability of $(\mathbf{A}, \mathbf{C})$ in order to obtain a well-defined steady-state Kalman filter. The observability condition can easily be proven by taking $\lim_{k\rightarrow\infty} E[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T]$, where $\hat{\mathbf{x}}_k$ is the estimate of a generic observer. Our time-invariance assumption also allows for the optimal alarm system to designed off-line, rather than computing $\mathbf{P}_{k|k}$ and $\mathbf{P}_k$ and re-designing the alarm system at each time step.

As such, we can the use solution to the discrete algebraic Riccati equation, $\hat{\mathbf{P}}_{ss}^R$, in place of $\mathbf{P}_{k|k}$ for Eqn. 19. $\hat{\mathbf{P}}_{ss}^R$ is the aposteriori steady state covariance, and is a quadratic function of the apriori steady state covariance matrix, $\mathbf{P}_{ss}^R$. $\mathbf{P}_{ss}^R$ is the algebraic counterpart of Eqn. 11. Similarly, $\mathbf{P}_{ss}$, the solution to the discrete algebraic Lyapunov equation, can be used in place of its counterpart $\mathbf{P}_k$ from Eqn. 3.

$V_{k+d|k}$ therefore requires the solution to the both steady-state Riccati and Lyapunov equations, and its bound is dependent only on the Lyapunov equation, as indicated on the legend on the right of Fig. 3. The Riccati solution is inherently a conditional covariance matrix by definition of $\mathbf{P}_{k|k}$, and the Lyapunov solution is inherently an unconditional covariance matrix by definition of $\mathbf{P}_k$.

The graph on the right of Fig. 3 allows for us to obtain an estimate of the margin to maximum uncertainty for $\hat{y}_{k+d|k}$ when using the chosen maximal prediction window, $d = 5$. This estimate serves only as a relative indicator of uncertainty for $\hat{y}_{k+d|k}$. It also serves to contrast the optimal alarm to the predictive alarm system, the latter of which does not use uncertainty as part of its construction. This is apparent in the qualitative oscillations that evolve with increased prediction window for both the predictive alarm system on the left of Fig. 3, and the prediction variance on the right of Fig. 3.

V. Future Work

Because algorithms based upon the optimal alarm system concept appeal to data mining and machine learning techniques, they are clearly viable candidates for extension to techniques such as particle filtering. Performing this extension will enable event distributions and model parameters to be adaptively updated as in [1] rather than making convenient Gaussian and stationary assumptions. However with particle filtering, the formulation of the problem can involve non-Gaussian noise, as well as non-linearities which were not covered in [1].

Furthermore, we want to investigate improved approximations that would provide a tighter bound on the alarm regions shown in Fig. 2. We will also investigate and compare the discrepancy between the error accumulated due to techniques studied here, and those due to improved approximations. Future development will involve more rigorous testing and validation of the alarm systems discussed by using standard machine learning techniques and consideration of more complex, yet practically meaningful critical level-crossing events.

Finally, a more detailed investigation of model fidelity with respect to available data and metrics has been conducted [10]. As such, future work on modeling will involve the investigation of necessary improvements in initialization techniques and data transformations for a more feasible fit to the assumed model structure. Additionally, we will explore the integration of physics-based and data-driven methods in a Bayesian context, by using a more informative prior.

Acknowledgment

The author would like to thank various sponsors for supporting this research, and to the reviewers: John Wallerius and Nikunj Oza for their help.

References

[1] M. Antunes, A. Amaral Turkman, and K. F. Turkman. A Bayesian approach to event prediction. *Journal of Time Series Analysis*, 24(6):631–646, November 2003.

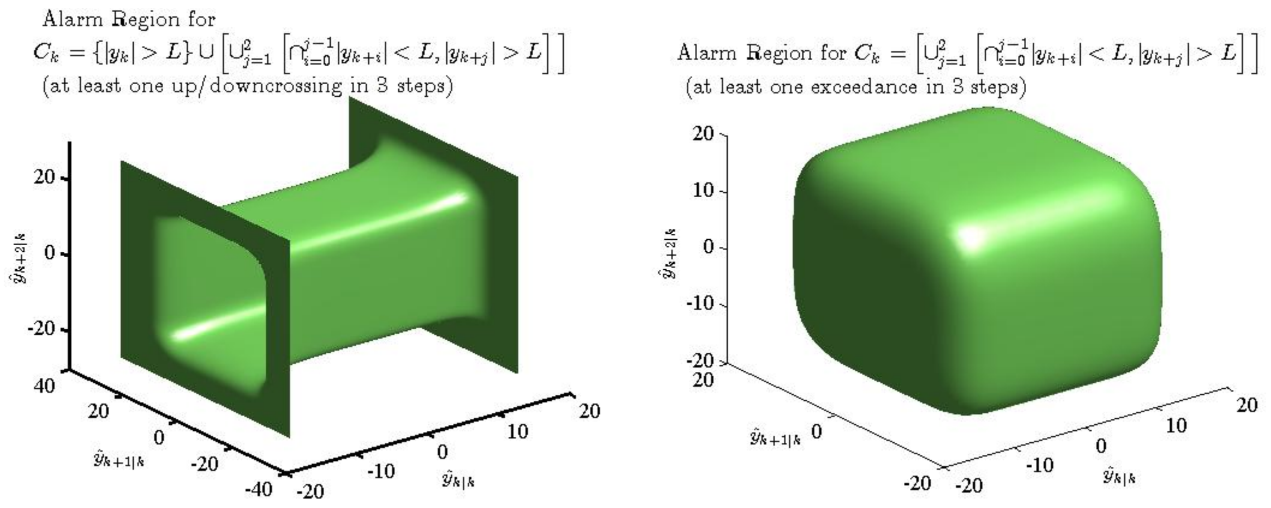
[2] Harald Cramér and M.R. Leadbetter. *Stationary and Related Stochastic Processes*. John Wiley and Sons, 1967.

[3] Jacques DeMaré. Optimal prediction of catastrophes with application to Gaussian processes. *Annals of Probability*, 8(4):840–850, August 1980.

[4] Michael I. Jordan. An introduction to probabilistic graphical models. Manuscript used for Class Notes of CS281A at UC Berkeley, Fall 2002.

[5] Thomas H. Kerr. 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):619–631, July 1982.

[6] M.R. Leadbetter. Development of reliability methodology for systems engineering. Technical report, NASA, April 1966.



Alarm Region for
$C_k = \{|y_k| > L\} \cup \left[\cup_{j=1}^{2} \left[\cap_{i=0}^{j-1} |y_{k+i}| < L, |y_{k+j}| > L\right]\right]$
(at least one up/downcrossing in 3 steps)
20
0
-20
40
20
0
-20
-40
-20
-10
0
10
20
$\hat{y}_{k+2|k}$
$\hat{y}_{k+1|k}$
$\hat{y}_{k|k}$
Alarm Region for $C_k = \left[\cup_{j=1}^{2} \left[\cap_{i=0}^{j-1} |y_{k+i}| < L, |y_{k+j}| > L\right]\right]$
(at least one exceedance in 3 steps)
20
10
0
-10
-20
20
0
-20
-20
-10
0
10
20
$\hat{y}_{k+2|k}$
$\hat{y}_{k+1|k}$
$\hat{y}_{k|k}$

Fig. 2. Alarm Region for Sample Level-Crossing Events

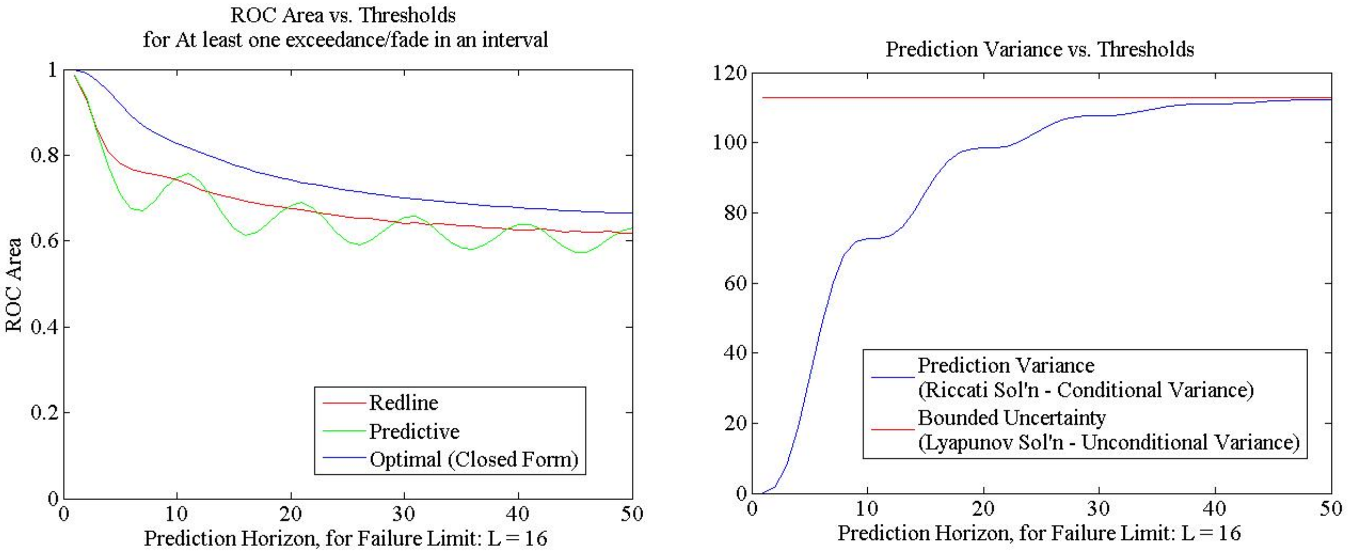


Fig. 3. Area Under the ROC Curve

[7] Ross Leadbetter. Re: NASA-development of reliability methodology for systems,engineering. E-mail correspondance, June 2007.

[8] Georg Lindgren. Optimal prediction of level crossings in Gaussian processes and sequences. *Annals of Probability*, 13(3):804–824, August 1985.

[9] Rodney Martin. Investigation of optimal alarm system performance for anomaly detection. In *National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*, Baltimore, MD, October 2007.

[10] Rodney Martin. An Investigation of State-Space Model Fidelity for SSME Data. In *Proceedings of the International Conference on Prognostics and Health Management*. IEEE (pending), 2008.

[11] Rodney A. Martin. Approximations of optimal alarm systems for anomaly detection. *IEEE Transactions on Information Theory (preprint)*, 2007.

[12] Kevin Murphy. Switching Kalman Filters. Technical report, Department of Computer Science, University of California, Berkeley, 1998.

[13] Kevin P. Murphy. The Bayes' Net Toolbox for MATLAB. *Computing Science and Statistics*, 33, 2001.

[14] Saharon Rosset. Model selection via the AUC. In *Proceedings of the Twenty-First International Conference on Machine Learning (ICML'04)*, Banff, Alberta, Canada, July 2004.

[15] Stanley F. Schmidt. The Kalman Filter: Its recognition and development for aerospace applications. *Journal of Guidance, Control, and Dynamics*, 4(1):4–7, 1981.

[16] Anders Svensson. *Event Prediction and Bootstrap in Time Series*. PhD thesis, Lund Institute of Technology, September 1998.

[17] Anders Svensson, Jan Holst, R. Lindquist, and Georg Lindgren. Optimal prediction of catastrophes in autoregressive moving-average processes. *Journal of Time Series Analysis*, 17(5):511–531, 1996.

[18] Alan S. Willsky and Harold L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21(1):108–112, 1976.