Concurrent estimation of time-to-failure and effective wear

Markus Loecher*
Siemens Corporate Research
755 College Road East
Princeton, New Jersey 08540

Christian Darken†
Naval Postgraduate School
Computer Science Department
Monterey, California 93943

Abstract

We propose a novel algorithm for the estimation of remaining lifetime of a generic device based on an intuitive and widely-applicable model of failure due to accumulation of effective wear. The simultaneous quantification of an abstract, multivariate wear function is achieved by a new learning algorithm which we term "cumulative effect regression". We develop the theory of the algorithm and compare its performance to traditional anomaly and pattern detection tools. Experimental results from X-ray tubes strongly validate the algorithm and demonstrate the utility of Operating Charactersitics (OC) curves as a powerful evaluation tool.

1 Introduction

When a machine or other device fails unexpectedly, the costs can be tremendous. The very worst case would be a device such as the gear box of a helicopter, where failure could easily cause the death of everyone aboard. In an industrial context, loss of life is more rare, but unexpected device failure can result in tremendous financial costs. These costs do not only include the obvious ones: the labor and material cost to repair or replace the device. Depending upon the use of the device,

*Markus.Loecher@scr.siemens.com
†cjdarken@nps.navy.mil

the costs can extend to collateral damage to product or devices, lost business because the device is unavailable, loss of the use of other devices whose use depend in some way on the broken one, as well as lateness penalties and future business lost because of customer dissatisfaction.

The traditional method of avoiding these costs is preventive maintenance, i.e. regular servicing or replacement. For some devices, however, elapsed calendar time, or even cumulative time of use, is a poor predictor of device failure. In such cases, it can be inordinately expensive to reduce the frequency of unexpected failures to an acceptable level. The costs include the labor to perform an unnecessary part replacement, the remaining life on the replaced part, and other costs from loss of availability while servicing is performed. For these reasons, it is very desirable to move from preventive maintenance to condition-based, or "predictive", maintenance, i.e. servicing or replacement only when called for by the actual condition of the device in question. Because condition-based maintenance is so obviously attractive to industry, providing measurement, data storage, and analysis tools for condition-based maintenance has become a small industry in its own right.

A key functionality to enable condition-based maintenance is the ability to predict when a device will fail given knowledge of its past use. Previous approaches to this problem have either been limited to a specific domain (the most general of which is perhaps the vibration analysis of rotating machines) or have provided information short of time-to-failure prediction, most commonly generating alarms when anomalous conditions are encountered.

In this paper, we propose an algorithm for failure prediction that achieves two goals concurrently. It estimates the remaining lifetime of a device, and, in parallel and at no extra computational cost, yields a quantitative estimate of the effective wear as a function of measurables. In the following section we take a closer look at the strenghts and shortcomings of a family of autoassociative anomaly detection (AAD) algorithms. We present two case studies that highlight a successful early warning as well as the failure of the AAD tool to predict failure. It is for the latter cases that our proposed wear estimator proves to be a valuable and effective alternative and/or augmentation. It is based on a simple, widely applicable, and intuitively appealing model of failure as a result of accumulated wear, as presented in section 3. We then apply a new learning algorithm, cumulative effect regression, to the task of wear estimation in section 4. Operating Characteristics (OC) curves are utilized in section 5 to assess the performance of the proposed algorithm on a fleet of X-ray tubes. Surface plots of the estimated wear function are presented in section 6, which is followed by a brief treatment of various possible further extensions of the technique (section 7), and discussion of related work (section 8).

2 Anomaly Detection: Success and Failure

A large fraction of the algorithms that support predictive scheduling of maintenance essentially are anomaly detection algorithms. A model is built of the normal range of the measurable signals. When the system departs from the normal range, an alarm is sounded. Various statistical technologies have been used to model the normal signals such as neural networks [1] [2], principal components analysis [3], or multivariate state estimation [4, 5, 6].

Anomaly detection techniques are very useful for device monitoring, and in fact are the only possibility in applications where identical devices similarly used nonetheless produce very different measurements. However, an intrinsic difficulty with these techniques is that anomalous measurements do not only occur when a device is near failure, but may also be due to changes in how the device is being used. In addition, the variables available for measurement may not contain any

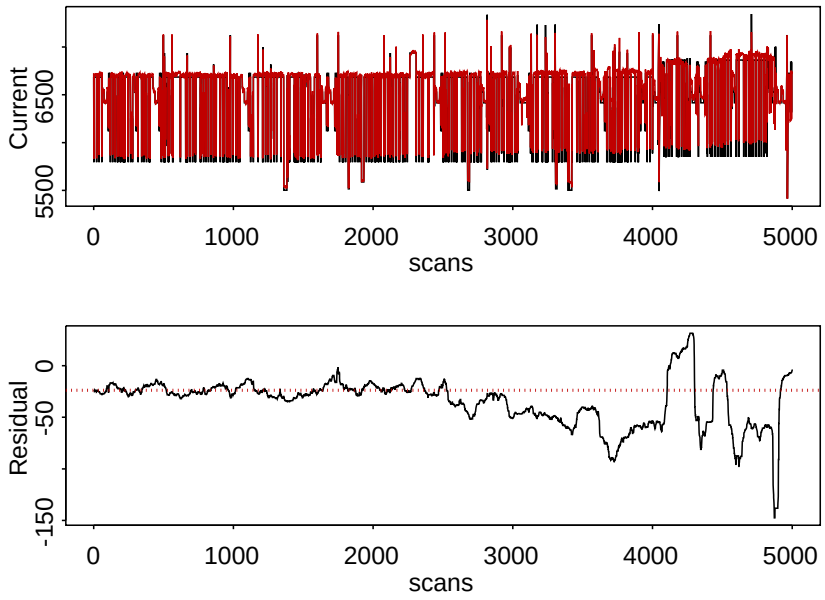

Figure 1: Actual (black) and estimated (red) current are shown in the upper figure. The smoothed residuals are displayed in the lower panel. The onset of drift in residuals could be detected up to 2000 scans before the tube's failure at scan number 5000.

information at all about the deteriorating state of the device. Even if they did, the

warning window might not be sufficient to be of any practical value whatsoever.

This section merely serves to give anecdotal evidence for the varying success of AAD algorithms; we do not attempt a thorough statistical or methodological classification of devices and circumstances in which AAD tools will either work or fail. The particular device we applied one particular AAD algorithm technique to are modern rotating-anode X-ray tubes used in a clinical setting. These devices are highly complex, having moving parts including bearings and high voltages applied at various points inside the device. A total of 19 proprietary variables are continuously measured and fed into the AAD tool. After an initial training period, the AAD algorithm returns the corresponding 19 *estimated* variables. The magnitude of the differences, the *residuals*, is a measure for the deviation from a normal, healthy state. Figure 1 displays one particular variable inside one of the X-ray tubes monitored along with its estimated value and the residual. The slowly developing drift indicating the onset of a malfunction is clearly discernible both in the original variable as well as in the residuals. A moderately smart diagnostic tool would have given a (correct) warning far in advance of the actual failure. This highly successful example provides credible evidence for the power and utility of the AAD appproach. Figure 2 displays the time series of a different variable - the number of high voltage arcs per scan - of a different X-ray tube.In this case, the residuals do not suggest any anomalous behavior until very shortly before the actual failure. The warning window of of a few scans is too short to be of any practical value. We consider this example as a case of missed detection.

3 Wear Model

Our technique is based upon the following model of failure. We define a “device” as any object that can be considered to fail, be it homogeneous and monolithic, such as a lining on a single component, or composed of many diverse parts, such as a motor. Each such device may be considered to have a state that describes its condition, from new and perfect to failed and broken. For complex devices, accurate modeling of the state for the purpose of predicting failure requires estimating many quantities. Our model, by contrast, models the state of a device by only a single quantity, called “wear”. (It is possible to extend our model to include one quantity per failure mode. See section 7.) As compared to more detailed models of a device that may be comprised of quantities that have a direct physical interpretation which is in principle directly measurable, the wear is a theoretical quantity, and may or may not correspond to a measurable quantity. A device in perfect condition is considered to have a wear of zero. The wear then generally increases over time as a device is used (although maintenance or certain modes of use may decrease

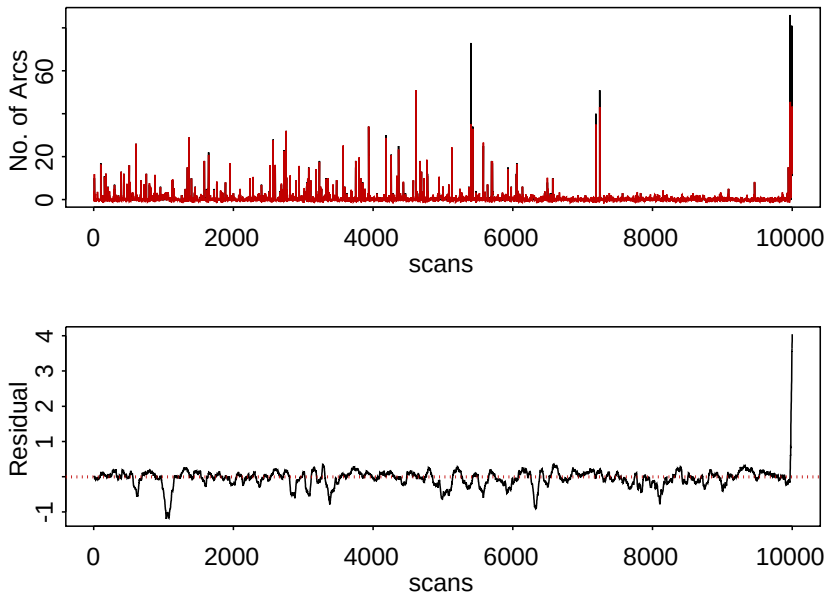

Figure 2: Actual (black) and estimated (red) current are shown in the upper figure. The smoothed residuals are displayed in the lower panel. No onset of drift in residuals could be detected until a few scans before the tube's failure at scan number 10000.

it). A wear of 1.0 corresponds to a device which has failed, i.e. has worn out.

In order to apply our technique, it is necessary to have access to one or more quantities that contain information about changes to the state of wear of the device. These quantities may include ones that are commonly used for maintenance scheduling purposes, such as calendar time since installation or minutes of use. They may also include control information sent to the device, parameters describing material being processed by the device, and of course, measurements relevant to the status or performance of the device, such as temperature, etc. To simplify our presentation, we assume that these quantities are available once per use for devices which perform discrete operations (e.g. punch presses) or are regularly sampled in the case of devices for which no such discrete operations can be defined (e.g. a continuously running induction motor). It is not necessary that the wear be a deterministic function of the available data in order for our technique to be useful. It is sufficient that the variance of the wear after taking the data into account be reduced to a significant degree. The accuracy of the wear prediction is judged by

its effectiveness in predicting device failure.

Given a data set of devices that have been measured until failure, our technique is able to compute an estimate of the current wear of a similar device and the variance of the estimate. The remaining wear of the device (i.e. 1.0 minus the current estimate of the wear) can then be used to forecast when the device will fail. The remaining wear can be used directly as a trigger for scheduling maintenance. Alternatively, the remaining wear can be transformed into a failure prediction in terms of calendar time either using average patterns of usage, or a proposed schedule of operation for the device.

4 Wear Estimation through cumulative effect regression

We will model the wear as a sum of wear increments, one per use (or period of use) of the device. Maintenance to a device can also be modeled as a "use", albeit one with a negative wear increment. Let w_{ij} be the wear increment associated with the j'th use of device i. Then the wear after p uses is given by

$$W_{i,p} \equiv \sum_{j=1}^{p} w_{ij}. \tag{1}$$

Since we do not have access to the wear increments, they must be estimated based on usage data that is available. Let x_{ij} be a vector containing all available information on the j'th use (or period of use) of device i. It is assumed for now that for each device represented, the usage data is complete, i.e. the device had no uses before it failed that are not in the data set. Let n be the number of devices for which we have data, and let N_i be the number of uses before failure for the i'th device. Let $w(\theta, x_{ij})$ be an estimate of the incremental wear w_{ij} given usage data x_{ij}. We propose to construct semiparametric estimates, so the size of θ will be chosen based on the usage data.

How can the incremental wear function be estimated? For the i'th device, the usage data has the form $(x_{i1}, x_{i2}, \ldots, x_{iN_i})$. The device failed after N_i uses, so its wear is 1.0. Unlike in ordinary (supervised) regression, the only information provided is the sum of the dependent variable values over arbitrary sequences instead of the individual values. We refer to this particular type of learning with highly impoverished supervision as *cumulative effect regression*. Compared to traditional regression, cumulative effect regression performs essentially the same task of predicting the individual increments, but with less information. From this perspective, it is somewhat remarkable that this problem can be solved at all, but in fact, for the linear case, it has a relatively simple closed form solution. In general, the quantity

to be minimized is the expected prediction error

$$\mathcal{E}(\theta) \equiv \sum_{i=1}^{n} \left(\sum_{j=1}^{N_i} w(x_{ij}, \theta) - 1 \right)^2 . \tag{2}$$

It should be readily apparent that θ can be estimated easily if we choose a linear form for w, i.e.

$$w(x_{ij}, \theta) = \sum_{k=1}^{m} \theta_k g_k(x_{ij}). \tag{3}$$

In that case, given n failed devices, we can compute the estimate $\hat{\theta}$ by minimizing the expected prediction error

$$\hat{\theta} = \arg\min \mathcal{E}(\theta) = \left(\sum_{i=1}^{n} s_i s_i^T \right)^{-1} \sum_{i=1}^{n} s_i, \tag{4}$$

where

$$s_i \equiv \sum_{j=1}^{N_i} g(x_{ij}), \tag{5}$$

and we have taken $g(x_{ij})$ to be the vector

$$(g_1(x_{ij}), g_2(x_{ij}), \ldots, g_m(x_{ij}))^T. \tag{6}$$

The derivation and structure of the solution is reminiscent of that for regression [7]. The standard tools used in conventional regression to improve generalization, i.e. cross-validation, boot-strapping, bagging, and the like, all seem to be applicable to cumulative effect regression. For the radial basis function network we will use later, the g_k are Gaussians with means and variances chosen as described in [8].

5 Performance Assessment

A reasonable goodness-of-fit criterion would be the expected prediction error (2) computed for a test set of devices, which is independent of the set used for training. Alternatively, bootstrap or cross-validation estimates of the latter could be utilized. Unfortunately, as any one of these measures is defined in the abstract "wear" space, a direct quantitative interpretation of its value is rather difficult. A more comprehensive and easily interpretable diagnostic tool is provided by *Operating Characteristic* (OC) curves, which are parametric plots of 'false alarms" versus "missed detection" as an implicit function of the sensitivity threshold. For

simplicity we will illustrate the fundamentals of our proposed predictive maintenance scheme via the simple example of exchanging a device when its lifetime T exceeds a threshold t specified by the user. The inherent tradeoff of this approach, which is common to most statistical classifiers, is characterized most lucidly by its associated error rates:

- A *type I error*, generally denoted by α, is committed by exchanging a device too early. The most obvious cost resulting from this error is the mean lost lifetime.
- A *type II error*, generally denoted by β, corresponds to a device failing before its lifetime assumes t. The associated costs are highly business and context specific. We choose to simply compute the percentage of unexpected failures.

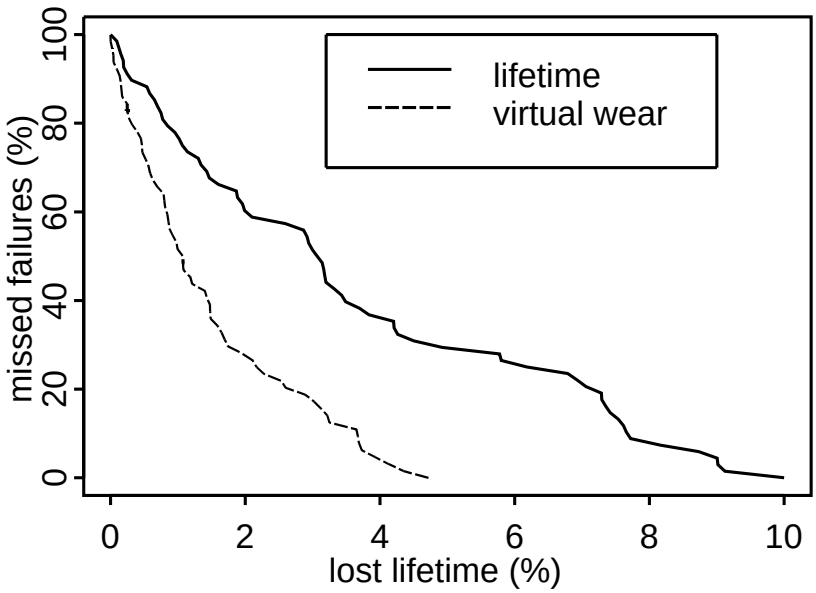

Figure 3: Experimentally obtained Operating Characteristics for X-ray tubes. The OC curve corresponding to the virtual wear model is significantly better than the one obtained from naively thresholding lifetime (scan seconds).

An OC curve, such as Figure 3, parametrically plots pairs of (α, β) while varying the threshold t through a reasonable range. We trained the wear model on a fleet

of the same X-ray tubes described above. An automated search process resulted in the selection of four of the available measurements to be used as inputs. The actual physical variables used are proprietary. The data set contained 69 failed tubes. Regular lifetime was measured in "scan seconds", which is the total number of seconds of use. We estimated the incremental wear function (equation 3) via a radial basis function network with twelve basis functions. An OC curve was generated by leave-one-out cross validation, i.e. training the model on the remaining 68 tubes when studying its performance on the remaining tube. Figure 3 compares the OC curve for this virtual wear model to the OC curve obtained when straightforward scan seconds thresholding is performed. The improvement is rather striking. For any acceptable fraction of missed failures, the amount of lost lifetime under the new algorithm is about half what it would be under simple time thresholding. This means that a significant maintenance cost, i.e. the cost of discarding parts that still have some life in them, has been halved.

6 Quantifying wear

The benefits of the cumulative wear model extend beyond providing an estimate of the remaining lifetime of a device, the quality which can be assessed by an OC curve. By estimating the parameters $\hat{\theta}$, the algorithm effectively approximates the incremental wear function (3). Being able to quantify the actual wear and therefore directly compare the effective device "damage" across different operating regimes can be invaluable not only for the user but especially for device manufacturers. Understanding how the wear of a particular device depends upon its operating parameters can lead to design improvements that further reduce maintenance cost. Graphical renderings of $w(\vec{x}, \theta)$ as a function of its arguments $\vec{x}$ provides a comprehensive summary of its dependence on the joint values of the input variables. Unfortunately, such visualization is limited to low-dimensional views. One way to visualize the wear curve or surface if the dimension of $\vec{x}$ is greater than two is a partial dependence plot, which are marginal averages over a subset of the input variables. It is possible in this way to concentrate on the effect of a selected subset of variables. Figure 4 is such a partial dependence plot, displaying the wear surface as a function of only two of the four variables. While the overall qualitative trend of this plot might not be a surprise to X-ray tube manufacturers, the ability to **quantify** the wear could be extremely useful.

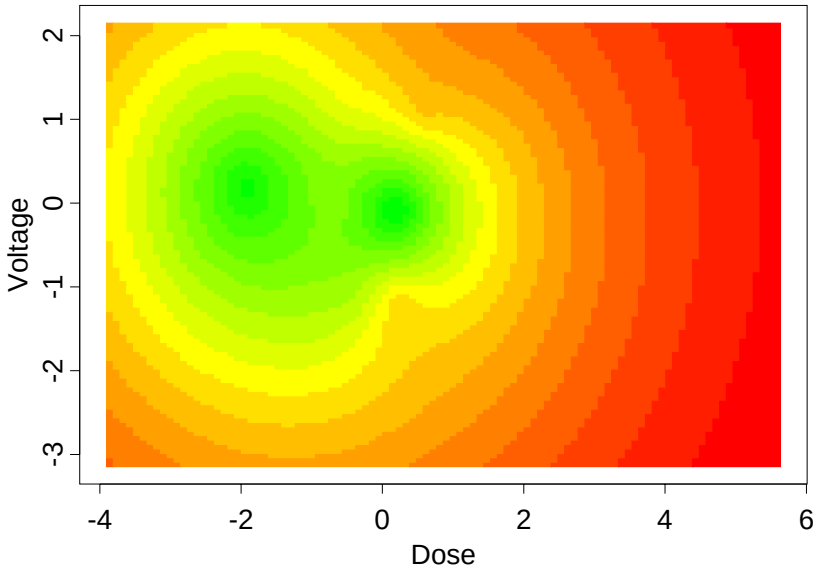

Figure 4: The wear surface as a function of two variables dose and voltage. Note that this is a partial dependence plot as there are a total of four variables. Shown is the effect averaged over the remaining two variables. The color coding extends from green (low) to yellow (intermediate) to red (high). We choose to omit the actual values of the wear function. Note that the variables have been rescaled to mean zero and unit variance.

7 Possible Extensions

Multiple modes of failure can be handled by multiple wear indices, one per mode, in the case that the failure modes are labeled, or by adaptations of "mixture of experts" techniques [9] in the case that they are unlabeled.

It is sometimes the case that some of the quantities comprising the usage data are not always available. For example, while some quantities might be measured for each use, others might be measured only occasionally. If, as is sometimes the case, the more rarely measured quantities change little between successive uses, they can simply be interpolated to provide estimates which are synchronous with the rest. Techniques adapted to more general settings are also available [10].

It might be desirable to have estimates of the incremental wear depend upon the instantaneous value of the total wear, e.g. for a particular type of use, wear

might accelerate as the device gets older. The algorithms required to optimize the parameters of the incremental wear function in this case are considerably more complicated, and a closed form computation is usually not sufficient [11]. This recursive wear estimate has not been explored by us.

8 Related Work

The subfield of statistics known as survival analysis encompasses multiple techniques for estimating remaining life, the most well-known of which is perhaps the proportional hazards method [12] [13]. These techniques are not generally applicable to the failure prediction problem described in this paper, as they rely upon the knowledge of a fixed number of boolean conditions, i.e. "risk factors" and "treatments". To apply these techniques to the context described in this paper, it is necessary to reduce the measurement sequences to a fixed number of conditions. This reduction requires a deep knowledge of the device that is often lacking. Additional points in favor of the algorithm described in this paper are the fact that periods in which the device is not in use are treated naturally in wear estimation (it would presumably have to be modeled as some sort of life-extending treatment in proportional hazards), and the fact that the proposed algorithm's adaptability to semi-parametric techniques may give it broader coverage as compared to the ad hoc parametric form of, e.g., proportional hazards.

An alternative approach is to apply system identification to determine the unknown parameters in a model of the state and observables of the device [14] [15] [16]. This approach requires a sufficiently accurate parametric model of the device, which is often not available. If the parametric model is very detailed, a common difficulty is in gathering enough data to determine all of the unknown parameters.

9 Conclusions and Discussion

We have introduced a new technique for predicting failure that involves the estimation of the amount of wear on the device modeled as a single number. Despite the simplicity of the wear model, the experimental result gives credible evidence that it can lead to useful failure predictions even on complex devices.

The ability to quantify wear as a function of operating parameters should be extremely useful both for device manufacturers as well as the end user. Besides the potential for improvements in manufacturing or designing modified usage protocols, warranty and service contracts could be refined considerably.

The data typically available for use in wear estimation is censored, i.e. we do not have and can not get complete usage-to-failure data for all devices because

some have not yet failed. When devices which have not yet failed are simply ignored, this results in a pessimistic bias in the data (since those devices which are the best performers are most likely not to be included). The best method of correcting this bias in the context of the presented technique is currently unclear.

Acknowledgments

The authors would like to acknowledge useful discussion with Ekkehard Blanz, Matthias Berger, Elke Jennewein-Wolters, and Roland Schmidt.

References

[1] G. A. P. Fontaine, E. E. E. Frietman, and R. P. W. Duin, "Preventive and Predictive Maintenance Using Neural Networks", *Journal of Microelectronic Systems Integration*, 4(2):87–93, 1996.

[2] T. Petsche, A. Marcantonio, C. Darken, S. Hanson, G. Kuhn, and I. Santoso, "A Neural Network Autoassociator for Induction Motor Failure Prediction", *Proceedings of Neural Information Processing Systems (NIPS 95)*, 1995.

[3] D. R. Lewin, "Predictive Maintenance Using PCA", *Control Engineering Practice*, 3(3):415–421, 1995.

[4] R. M. Singer, K. C. Gross, J. P. Herzog, R. W. King, S. Wegerich, "Model-Based Nuclear Power Plant Monitoring and Fault Detection: Theoretical Foundations", Proc. 9th Intnl. Conf. On Intelligent Systems Applications to Power Systems, pp. 60-65, Seoul, Korea, 1997.

[5] N. Zavaljevski, K. C. Gross and S. Wegerich, "Regularization Methods for the Multivariate State Estimation Technique (MSET)", Proc. MC'99 Conference on Mathematics and Computation, Reactor Physics and Environmental Analysis in Nuclear Applications, Madrid, Spain (1999).

[6] N. Zavaljevski and K. C. Gross, "Support vector machines for nuclear reactor state estimation", American Nuclear Society International Topical Mtg. On "Advances in Reactor Physics, Mathematics, and Computation into the Next Millennium," Pittsburgh, PA, (May 7-11, 2000).

[7] Norman R. Draper and Harry Smith *Applied Regression Analysis*, Third edition. John Wiley and Sons, Inc., 1998.

[8] J. Moody and C. Darken, "Fast Learning in Networks of Locally-Tuned Processing Units", *Neural Computation*, 1:289–303, 1989.

[9] M. I. Jordan and R. A. Jacobs, "Hierarchical mixture of experts and the EM algorithm", *Neural Computation*, 1993.

[10] Roderick J. A. Little and Donald B. Rubin, *Statistical Analysis with Missing Data* John Wiley and Sons, 1987.

[11] Barak A. Pearlmutter, "Gradient calculation for dynamic recurrent neural networks: a survey.", *IEEE Transactions on Neural Networks*, 6(5):1212-1228, 1995.

[12] D. Cox, *Analysis of Survival Data*, CRC Press, 1984.

[13] David W. Hosmer, Jr. and Stanley Lemeshow, *Applied Survival Analysis: Regression Modeling of Time to Event Data* John Wiley and Sons, 1999.

[14] Q. Zhang, M. Basseville, and A. Benveniste, "Early Warning of Slight Changes in Systems", *Automatica*, 30(1):95–113, 1994.

[15] M. Basseville, A. Benveniste, B. Gach-Devauchelle, M. Goursat, D. Bonnecase, P. Dorey, M. Prevosto, and M. Olagnon, "*In Situ* Damage Monitoring in Vibration Mechanics: Diagnostics and Predictive Maintenance", *Mechanical Systems and Signal Processing*, 7(5):401–423, 1993.

[16] S. K. Yang and T. S. Liu, "State estimation for predictive maintenance using Kalman filter", *Reliability Engineering and System Safety*, 66:29–39, 1999.