

Methods for Binning and Density Estimation of Load Parameters for Prognostics and Health Management

NIKHIL M. VICHARE, PETER RODGERS, and MICHAEL G. PECHT*

CALCE Electronic Products and Systems Center, University of Maryland, College Park, MD 20742, USA

Abstract: Environmental and usage loads experienced by a product in the field can be monitored in-situ and used with prognostic models to assess and predict the reliability of the product. This paper presents an approach for recording in-situ monitored loads in a condensed form without sacrificing the load information required for subsequent prognostic assessments. The approach involves optimally binning data in a manner that provides the best estimate of the underlying probability density function of the load parameter. The load distributions were developed using non-parametric histogram and kernel density estimation methods. The use of the proposed binning and density estimation techniques with a prognostic methodology were demonstrated on an electronic assembly.

Key Words: *electronic prognostics, health and usage monitoring system (HUMS), condition monitoring, reliability, optimal binning, density estimation*

1. Prognostics and Health Management

Prognostics and health management (PHM) permits the reliability of a system to be assessed in its actual application conditions [1] [2]. PHM is employed by integrating sensor data and prediction models that enable in-situ assessment of the extent of deviation or degradation of a product from an expected normal operating condition (i.e., the system's "health" or reliability). The objective is to provide: (1) advance warning of failures; (2) minimizing unscheduled maintenance, extending maintenance cycles, and maintaining effectiveness through timely repair actions; (3) reducing the life-cycle cost of equipment by decreasing inspection costs, downtime, and inventory; and (4) improving qualification and assisting in the design and logistical support of fielded and future systems [2].

One method for PHM implementation involves collecting the environmental and usage loads experienced by the product in its life cycle, and using models to assess the degradation of the product [2] [3] [4] [5]. Typical environmental loads include temperature, vibrations, shock,

* Corresponding author's email: pecht@calce.umd.edu

pressure, acoustic levels, strain, stress, inert environments, humidity levels, and contamination. Usage loads include usage frequency, usage severity, usage time, power, heat dissipation, current, and voltage. These life cycle loads, either individually or in various combinations, may lead to performance or physical degradation of the product and subsequently reduce its service life. The extent and rate of product degradation depends upon the nature, magnitude, and duration of exposure of the loads. The damage inflicted, and hence the life consumed can be assessed by using the monitored data in conjunction with physics-of-failure [6] based stress and damage models [7] [8].

In applications such as civil infrastructure, aircrafts, telecommunication systems and military equipment, life cycle loads required for PHM are often remotely monitored using sensor systems. Due to either limited accessibility, or logistics and operational constraints, one of the major challenges in the implementation of these sensor systems lies in efficient management of power and memory consumption [9] [10] [11] [12] [13]. Thus sensing and data collection strategies that enable reduction of on-board storage, power consumption, and uninterrupted data collection over longer durations, are generally desired.

One of the emerging methods to achieve in-situ monitoring involves integrating embedded processing capabilities with on-board processors of the sensor systems [5] [11] to enable immediate and localized processing of the “raw” sensor data. For wireless sensor systems, embedded processing enables transmitting fewer amounts of data (processed instead of raw data) to the base station, and hence results in lower power consumption. For large sensor systems working in a network, this allows decentralization of computational power and facilitates efficient parallel processing of data [11] [12] [14].

Data simplification is a way to obtain gains in computing speed and testing time, condense load histories without sacrificing important damage characteristics, preserve the interaction of load parameters, and provide an estimate of the error introduced by reducing and simplifying the data. Data simplification can be achieved using a variety of tools such as filters, Fourier transforms, wavelets, Hayes method, ordered overall range, etc.

It is essential to simplify or pre-process the “raw” sensor data to make it compatible with the damage models and algorithms needed to conduct prognostics. In particular, it may be necessary to extract the relevant load parameters including cyclic mean, amplitudes, ramp rates, hold periods, power spectral densities, etc. Commonly used load parameter extraction methods include: cycle counting algorithms for extracting cycles from time-load signal, Fast Fourier transforms (FFT) for extracting the frequency content of signals, etc. Depending on the application and type of signal, custom load extraction methods may be required.

Figure 1 is a schematic of the in-situ monitoring, pre-processing, and storage of environmental and usage loads. A time-load signal is monitored in-situ using sensors, and further processed to extract (in this case) cyclic range (Δs), cyclic mean load (S_{mean}), and rate of change of load

(ds/dt) using embedded load extraction algorithms. The extracted load parameters are stored in appropriate bins to achieve further data reduction. The binned data is downloaded to estimate the distributions of the load parameters for use in damage assessment, remaining life estimation, and the accumulation of the products use history. Since the accuracy of the damage assessment and remaining life prognostics depend on the accurate estimate of the probability density functions (pdf) of the loads derived from the binned data, it is essential to select the appropriate bin-width and number of bins a-priori.

This paper provides a method for applying optimal binning techniques for enabling condensed storage of in-situ monitored load histories. Optimal bin-widths relations are used in conjunction with non-parametric density estimation techniques, namely the histograms and kernels. The approach is then demonstrated.

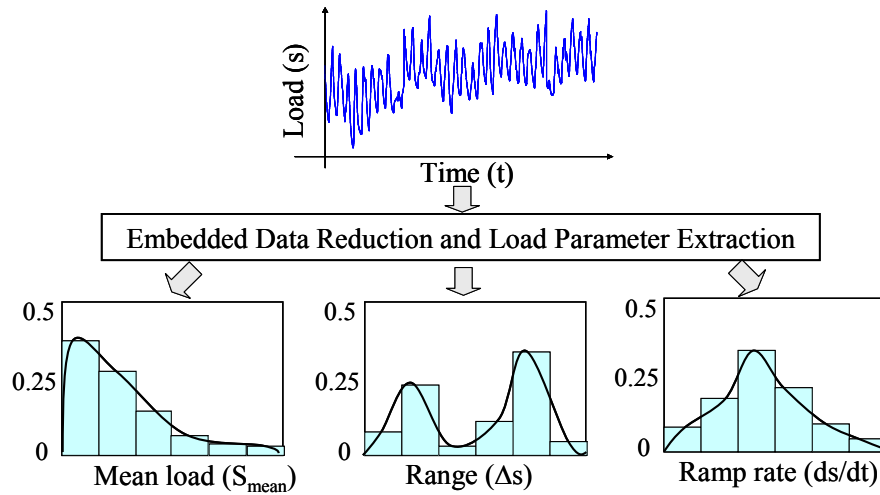


Figure 1. Load parameters extracted and optimally binned

2. Density Estimation and Binning Optimization

If the entire data set, without data binning, is retained, a parametric method can be used for density estimation. In that case the objective is to find a best fitting parametric function to the given data. Standard parametric functions include normal, lognormal, exponential, and Weibull, described by parameters such as mean and standard deviation. If the underlying distribution is non-standard (e.g., multi-modal), fitting a single standard parametric function to the data can result in errors, and fitting multiple parametric functions can make the problem solution non-unique. Non-parametric methods can be used to estimate the pdf without assuming any parametric structure [15] [16]. Examples of non-parametric density estimation methods include histograms, Kernels, orthogonal series estimation, and nearest neighbor method. In our study, histograms and kernel estimators were used.

2.1 Histograms

A histogram is a graphical representation of a frequency distribution in which the height of the bins represents the observed frequencies. The choice of bin-width primarily controls the representation of the actual data. Smaller bin-widths may present too much details (undersmoothing) and larger bin-widths may present or too less details (oversmoothing), of the true distribution [15]. Histograms are based on an equally spaced bin-width h_n where n denotes the sample size. If $\hat{f}(x)$ is the estimate of the true density $f(x)$, then the global accuracy of the estimate can be evaluated by the integrated mean square error that is defined by.

$$IMSE = \int E[\hat{f}(x) - f(x)]^2 dx \quad (1)$$

The IMSE is the sum of the integrated square bias and integrated variance, the bias being the difference between the true density and the estimate. The bin-width that minimizes the IMSE is required for accurate estimation of the true pdf $f(x)$. Scott, [17], derived the asymptotically optimal choice of bin width, h_n^* as;

$$h_n^* = \left(6^{1/3} \left(\int_{-\infty}^{\infty} f'(x)^2 dx \right)^{-1/3} \right) n^{-1/3} \quad (2)$$

Since h_n^* depends on the unknown pdf $f(x)$, an estimate $\hat{f}(x)$ of $f(x)$ can be plugged into equation 2. For the normal distribution, the approximate optimal bin-width is;

$$h_n^* = 3.49sn^{-1/3} \quad (3)$$

where s is the sample standard deviation and n is the sample size. Freedman and Diaconis [18] suggested another bin width formula that can be applied to non-Gaussian data;

$$h_n^* = 2(IQR)n^{-1/3} \quad (4)$$

where IQR is the inter-quartile range of the data (Q3-Q1). Several other studies have focused on more quantitative and qualitative aspects of bin width selection [19] [20] [21] [22] [23]. Once the bin width is calculated, the histogram is defined as;

$$\hat{f}(x) = \frac{1}{nh} (\text{number of } X_i \text{ in the same bin as } x) \quad (5)$$

2.2 Kernel Estimators

Kernel density estimators [24] are non-parametric methods used to estimate $f(x)$ using the kernel estimator K (also called a window function) defined as;

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x-x_j}{h}\right) \quad (6)$$

where h is the bin width, also called as smoothing parameter or band width. Equation 6 can be considered as a pdf obtained by averaging the effect of a set of local functions K centered at each data point x .

A variety of kernel functions with different mathematical properties have been reported [24] [25] [26] [27] [28]. Commonly used kernel functions include uniform, Gaussian, triangle,

Epanechnikov, and biweight. Kernel functions are generally chosen based on the ease of computation and desired properties of the function.

For example, the Gaussian kernel is used for its continuity and differentiability. The optimal bin width for the Gaussian kernel is obtained by minimizing the integrated mean square error (equation 1).

$$h^{opt} = 1.06\hat{\sigma}n^{-1/5} \quad (7)$$

where $\hat{\sigma}$ is the estimate of standard deviation. The Gaussian kernel can be used for non-Gaussian data, since the kernel is used only as a local weighting function [16].

3. Approach

For in-situ monitoring over long durations, one can use the load measurements obtained in the past or the data monitored over initial few periods to obtain the estimate of sample size and standard deviation. The relations provided in equation (3) and (7) are then used for bin width calculations. Once the bin-widths are calculated, the data for the subsequent time periods is stored directly in the bins. At the end of the monitoring period, the bin-widths and frequency is used in conjunction with equations (5) and (6) to obtain the probability distributions. These distributions are then used for damage assessment and prognostics (see Figure 2).

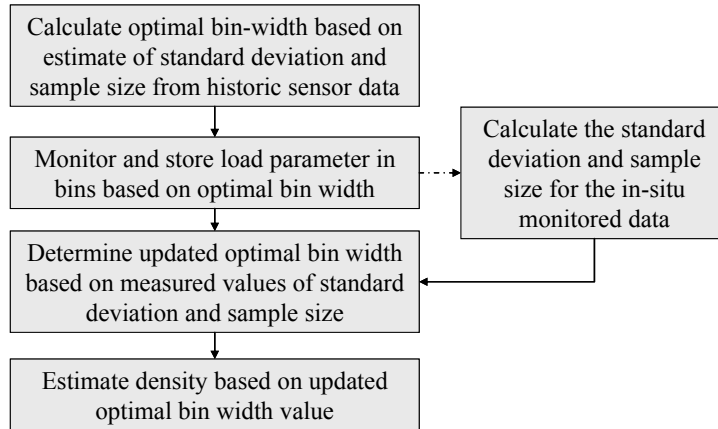


Figure 2. Method for estimating probability density using binned load data

The load parameters obtained during monitoring depend on the usage and environmental conditions during the monitoring period. Thus, depending on usage conditions, there could be differences between the estimated and actual values of sample size and standard deviation. These differences can lead to inaccuracies in the subsequent density estimation process. One approach to overcome this problem is by recording the actual values of standard deviation and sample size during monitoring and using these values for density estimation. One can update the value of standard deviation every time a data point is added to the bins and also keep a track on

the sample size. The advantage of continuously updating the standard deviation is that it does not require storing the complete data set. Optimal bin-widths can then be recalculated based on actual values of sample size and standard deviation to estimate the probability distribution (see Figure 2).

4. PHM Implementation Case Study

The use of the proposed binning and density estimation techniques with a PHM methodology were demonstrated on an electronic assembly. The remaining life of the electronic assembly is assessed by modeling the damage accumulation due to cumulative exposures to temperature variations (see Figure 3).

4.1 Experimental Setup

An electronic assembly (Figure 3) was exposed to a completely irregular temperature profile for 120 days. During the testing, the temperatures on the assembly (board and components) were measured. An example of the temperature profile measured in-situ for the first 20 days is shown in Figure 4. The raw sensor data was processed using the Ordered Overall Range (OOR) and 3-parameter rainflow algorithm [29]. The OOR converted the irregular time-temperature data into a sequence of peaks and valleys. This information was fed to rainflow cycle counting algorithm to count the number of cycles and extract the load parameters, namely the cyclic mean, the range, and the ramp time.

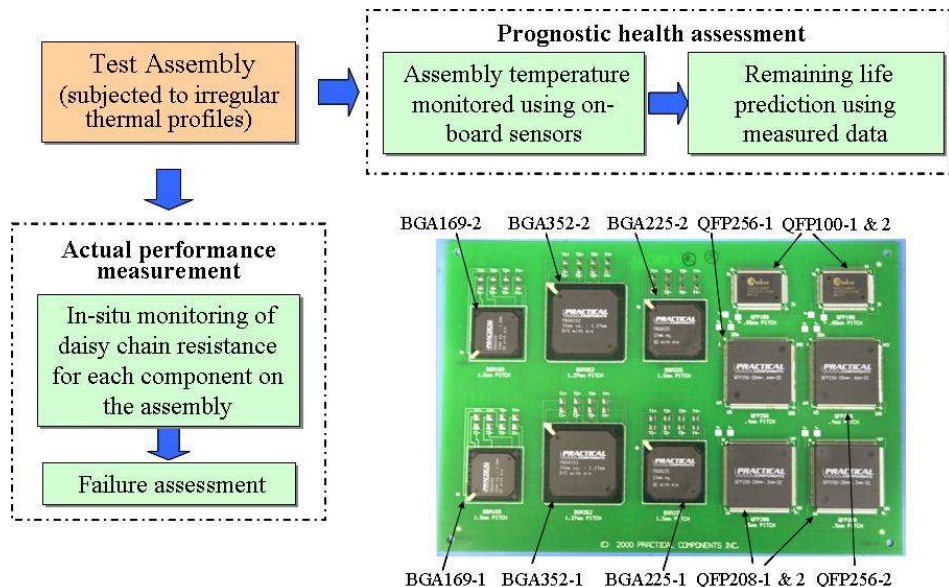


Figure 3. Experimental setup for PHM implementation on electronic assembly

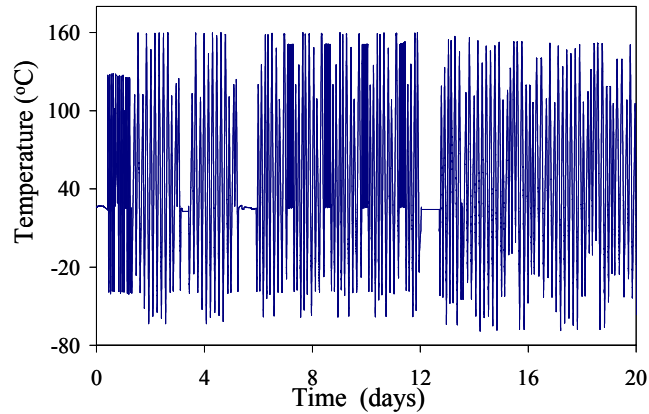


Figure 4. Illustration of temperatures measured on the board

Based on the output from the rainflow cycle counting, the distributions of load parameters were obtained. These distributions were then sampled in a Monte Carlo simulation of a damage model for the solder joint interconnect of the electronic components on the assembly. The Monte Carlo simulation provided the distribution of damage fractions. The damage fractions due to each exposure were then accumulated linearly using Miner's hypothesis [6]. A damage fraction equal to 1 indicates failure or end of life of the product being monitored.

Along with the temperature, the daisy chain resistance (actual electrical performance) of the solder interconnects was measured in-situ. The failure criteria for the solder joints was based on IPC standard IPC-SM 785, that states that failure manifests itself in the form of resistance spikes of short durations $>300 \Omega$ and failure is defined as the first interruption of electrical continuity that is confirmed by 9 additional interruptions with an additional 10% of cycle life. The actual (measured) failure of the solder joint was then compared with the prediction made using the PHM methodology.

4.2 Assessment of Binning Techniques

The rest of the discussion will focus only on cyclic T_{mean} , since the procedure used for binning and the result obtained are similar for all load parameters. For the first five days of the experiment, the cyclic T_{mean} data was extracted and stored without any binning. The average samples size (n), standard deviation (s), and inter-quartile range (IQR) for each load parameter was then calculated (see Table 1). These values were used as estimates of s , n , and IQR in equations (3), (4), and (7) to calculate the optimal bin widths for both histograms and kernels. For the histograms, the optimal bin widths obtained from equations (3) and (4) were within $\pm 5\%$. This is in agreement with results reported in the literature [16].

After the fifth day, the load parameters were extracted and stored in the bins calculated from the five day estimate. The bin frequencies were downloaded and distributions of load parameters were derived using histograms (see equation 5) and kernel functions (see equation 6). The

percentage of data reduction due to use of the bins was measured. The ability of histograms and kernel functions (using binned data) to represent the complete data set was assessed by comparing the histograms and kernel plots with the parametric distributions that were best fitted to the complete data set.

Table 1. Estimate of parameters used for the bin width calculation

| Day | Sample size (n) | Inter-quartile range (IQR) | Standard deviation (s) |
|---------|-----------------|----------------------------|------------------------|
| 1 | 97 | 110.77 | 59.10 |
| 2 | 80 | 131.20 | 66.70 |
| 3 | 72 | 101.81 | 60.94 |
| 4 | 80 | 36.63 | 48.89 |
| 5 | 90 | 135.96 | 68.68 |
| Average | 84 | 103.27 | 60.68 |

5. Results and Discussion

The best fitting distribution for the complete data set of T_{mean} values obtained on day 6 was found to be a normal distribution (see Figure 5). The distribution obtained using histogram (equation 5), based on data binned as per equation 3, is shown in Figure 5. The solid dark line in the figure shows the distribution of T_{mean} values obtained from binned data and using the Gaussian kernel. It is observed that the true shape of the distribution obtained from the data is bi-modal, which is accurately represented by the kernel and histogram. This bi-modal nature of T_{mean} values was hidden by the smooth normal distribution.

When the distributions were used in the Monte Carlo simulation model for damage assessment, the samples of T_{mean} values drawn from the kernel resulted in more accurate estimates of damage, since the samples were drawn from a distribution that accurately represented the measured data. In comparison to the histogram, the kernel density estimate is smoother. Hence during random sampling, more number of distinct samples were drawn from the kernel distribution as compared to the histogram. This further improved the accuracy of the damage distribution resulting from the Monte Carlo.

In terms of data reduction, the use of kernels and histograms resulted in 78% and 85% less data per day compared to using the complete monitored data set. Assuming the same amount of data reduction, the sensor system with binning can now be used uninterrupted for 6.6 more days with histogram and 4.7 more day with kernels to consume the same storage space without binning. Similar results were obtained when the analysis was conducted from day seven onwards. This clearly demonstrates the importance of these methods to enable uninterrupted monitoring at low power and memory consumption.

The sensitivity of the kernel densities to the difference between estimated and actual standard deviation and sample size were investigated by analyzing the experimental data for different time periods. For example, Figure 6 compares three density estimates estimated using

Gaussian kernel; distribution (1) is obtained using the complete data set and actual values of standard deviation and sample size for calculating is h_{opt} . Distribution (2) is based on binned data and estimated values standard deviation and sample size for calculating is h_{opt} . Distribution (3) is obtained from binned data and updated values standard deviation and sample size using the approach in Figure 2. In this example, the difference between the actual and estimated s was 38%. It is observed that the distribution (3) using updated value is a better estimate of the actual distribution (1). Similar results were obtained for different sets of data collected in the study.

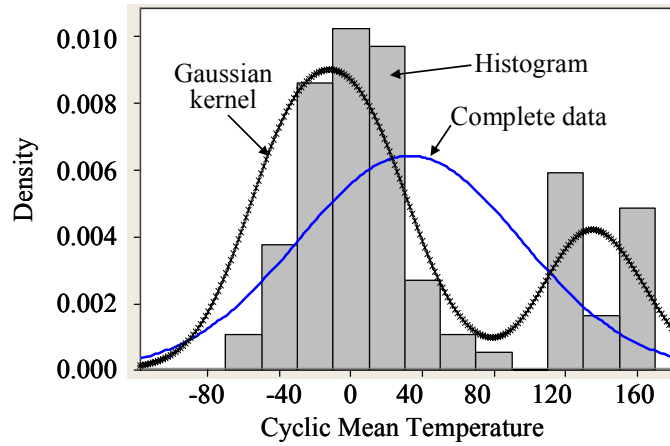


Figure 5. Comparison of density estimates for T_{mean} values of day 6

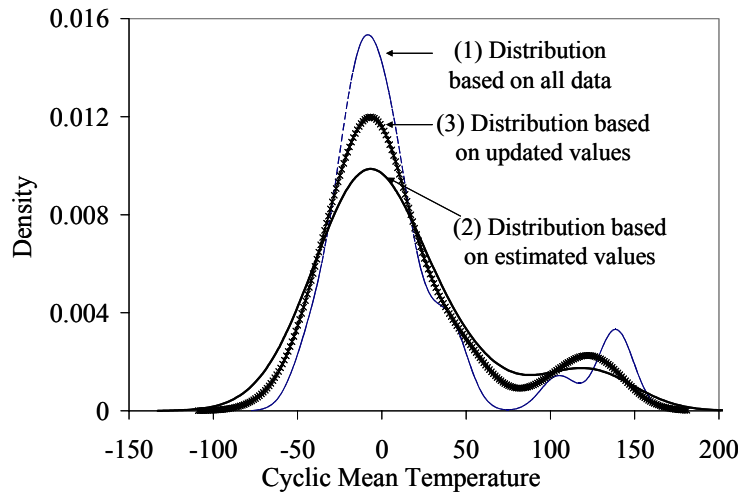


Figure 6. Comparison of distributions obtained from estimated versus actual parameters

6. Conclusions

For prognostics and health management, the application of optimal binning and density estimation using histogram and kernel estimators for condensed storage of load histories, can significantly improve the ability of the sensor nodes to monitor for extended durations without interruptions and conserve memory and power. It was found that optimal bin widths can be calculated a-priori based on estimates of sample size and standard deviation of representative data sets. In using the binned data for representing the entire data set, kernel methods provided a better estimate of the probability density compared to histograms. It was shown that the difference between the estimated and actual bin-width and the resulting error in density estimation can be minimized by recording the sample size and standard deviation during in-situ monitoring, and using the recorded (true) values for density estimation.

References

- [1]. Vichare N., P. Rodgers, V. Evely; M. Pecht, *In-Situ Temperature Measurement of a Notebook Computer - A Case Study in Health and Usage Monitoring of Electronics*, IEEE Transactions on Device and Materials Reliability, Vol. 4., No. 4, pp. 658-663, December 2004.
- [2]. Vichare, N., M. Pecht, *Prognostics and Health Management of Electronics*, IEEE Transactions on Components and Packaging Technologies, Vol. 29, No. 1, March 2006.
- [3]. Tumer, I., and A. Bajwa, *A Survey of Aircraft Engine Health Monitoring Systems*, Proceedings of AIAA Joint Propulsion Conference, Los Angeles, CA, June 1999.
- [4]. Chang, P., A. Flatau, S. Liu, *Review Paper: Health Monitoring of Civil Infrastructure*, Journal of Structural Health Monitoring, Vol. 3, No. 3, pp. 257-267, 2003.
- [5]. Vichare, N.; P. Rodgers, V. Evely, M. Pecht, *Monitoring Environment and Usage of Electronic Products for Health Assessment and Product Design*, IEEE Workshop on Accelerated Stress Testing and Reliability (ASTR), Austin, TX, USA, October 2-5, 2005.
- [6]. Dasgupta, A., *The Physics-of-Failure Approach at the University of Maryland for the Development of Reliable Electronics*, Proc. Third International Conf. on Thermal and Mechanical Simulation in (Micro) Electronics (EuroSimE), pp. 10-17, 2002.
- [7]. Ramakrishnan, A., M. Pecht, *A Life Consumption Monitoring Methodology For Electronic Systems*, IEEE Transactions on Components and Packaging Technologies, Vol. 26, No. 3, pp. 625-634, September 2003.
- [8]. Mathew, S.; D. Das, M. Osterman, M. Pecht, R. Ferebee, *Prognostic Assessment of Aluminum Support Structure on a Printed Circuit Board*, Accepted for publication in ASME Journal of Electronic Packaging, EP-05-1031
- [9]. Carroll, J., M. Fennell, *An Autonomous Data Recorder for Field Testing*, International Symposium on Transportation Recorders, Arlington, Virginia, May 3 - 5, 1999.
- [10]. Woodward, S., *Development and Flight Testing of an Adaptable Vehicle Health Monitoring Architecture*, Journal of Aircraft, Vol. 41, No. 3, pp. 531-539, May-June 2004.
- [11]. Lynch, J., A. Sundararajan, K. Law, A. Kiremidjian, *Embedding Damage Detection Algorithms in a Wireless Sensing Unit for Operational Power Efficiency*, Journal of

- Smart Materials and Structures, Vol. 13, No. 2, pp. 800-810, 2004.
- [12]. Akyildiz, I.F., W. Su, Y. Sankarasubramaniam, E. Cayirci, *Wireless Sensor Networks a Survey*, Journal of Computer Networks, Vol. 38, No. 4, pp. 393-422, 2002.
- [13]. Porret, A., T. Melly, C. Enz, E. Vittoz, *A Low Power Low Voltage Transceiver Architecture Suitable for Wireless Distributed Sensor Networks*, IEEE International Symposium on Circuits and Systems, Geneva, Vol. 1, pp. 56-59, 2000.
- [14]. Sinha, A., A., Chandrakashan, *Dynamic Power Management in Wireless Sensor Networks*, IEEE Design and Test of Computers, Vol. 18, No. 2, pp. 62-74, March-April 2001.
- [15]. Wand, P., *Data Based Choice of Histogram Bin Width*, The American Statistician, Vol. 51, No. 1, pp.59-64, March 1997.
- [16]. Izenman, A. J., *Recent Developments in Non Parametric Density Estimation*, Journal of American Statistical Association, Vol. 86, No. 413, pp.205-13, March 1991.
- [17]. Scott, D., *On Optimal and Data Based Histograms*, Biometrika, Vol. 66, No. 3, pp. 605-10, 1979.
- [18]. Freedman, D., P. Diaconis, *On Maximum Deviation Between the Histogram and Underlying Density*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 58, pp. 139-167, 1981.
- [19]. He, K., G. Meeden, *Selecting the Number of Bins in a Histogram a Decision Theoretic Approach*, Journal of Statistical Planning and Inference, Vol. 61, No. 1, pp. 49-59, 1997.
- [20]. Jones, M., J. Marron, S. Sheather, *A Brief Survey of Bandwidth Selection for Density Estimation*, Journal of American Statistical Association, Vol. 91, No. 433, pp.401-407, 1996.
- [21]. Koyama, S., S. Shinomoto, *Histogram Bin Width Selection for Time-Dependant Poisson Process*, Journal of Physics A: Mathematics and General, Vol. 37, No. 29, pp. 7255-65, 2004.
- [22]. Song, M., R., Haralick, *Optimally Quantized and Smoothed Histograms*, Proceedings of Joint Conference on Information Sciences, Computer Vision, Pattern Recognition and Image Processing, pp. 894, 2002.
- [23]. Paulter, N.G., *The Effect of Histogram Size on Histogram- Derived Pulse Parameters*, IEEE Transactions on Instruments and Measurements, Vol. 47, No. 3, pp.609-12, 1998.
- [24]. Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.
- [25]. Fadda, D., E. Slezak, A. Bijaoui, *Density Estimation with Non-parametric Methods*, Astronomy and Astrophysics Supplement Series, Vol. 127, pp. 335-52, January 1998.
- [26]. Sheather, S., M. Jones, *A Reliable Data Based Bandwidth Selection Method for Kernel Density Estimation*, Journal of Royal Statistical Society, Vol. 53, No.3, pp.683-90, 1991.
- [27]. Chiu, S., *Bandwidth Selection for Kernel Density Estimation*, Annals of Statistics, Vol. 19, No. 4, pp. 1883-1905, 1991.
- [28]. Glavinovic, M., *Comparison of Parzen Density and Frequency Histograms as Estimators of Probability Density Function*, European Journal of Physiology, Vol. 433, No. 1-2, pp. 174-179, 1996.

[29]. Anzai, H., *Algorithm of Rainflow Method*, The Rainflow Method in Fatigue, Butterworth-Heinmann Oxford, pp 11-20, 1991.



Nikhil M. Vichare received the B.S. degree in production engineering from the University of Mumbai, India, and the M.S. degree in industrial engineering from the State University of New York at Binghamton. He is currently working towards the Ph.D. degree in mechanical engineering at the University of Maryland, College Park, in the area of electronic prognostics.

Peter Rodgers received the Ph.D. degree in mechanical engineering from the University of Limerick, Limerick, Ireland. He is currently an Associate Professor in the Mechanical Engineering Department at The Petroleum Institute, Abu Dhabi, United Arab Emirates. Formally he was with the CALCE Electronic Products and Systems Center, University of Maryland; Nokia Research Center, Finland; and Electronics Thermal Management, Ltd., Ireland. His research interests include conventional and high-performance cooling technologies, electronics reliability including health monitoring of electronics. He has authored or co-authored over 50 journal and conference publications on a broad range of topics in these areas, and has been an invited lecturer, keynote speaker, panelist, and session chair at international electronics thermal management conferences. Dr. Rodgers received the 1999 Harvey Rosten Memorial Award for his publications on the application of CFD analysis to electronics thermal design. He is a member of several international conference program committees, and is program co-chair for EuroSimE 2006.



Michael G. Pecht is a Chair Professor and the Director of the CALCE Electronic Products and Systems Center at the University of Maryland. Dr. Pecht has an MS in Electrical Engineering and an MS and PhD in Engineering Mechanics from the University of Wisconsin at Madison. He is a Professional Engineer, an IEEE Fellow, an ASME Fellow, and a Westinghouse Fellow. He has written eleven books on electronics products development. He has written six books on the electronics industry in S.E. Asia. He served as chief editor of the IEEE Transactions on Reliability for eight years and on the advisory board of IEEE Spectrum. He is currently the chief editor for Microelectronics Reliability. He serves as a consultant for various companies, providing expertise in strategic planning, design, test, and risk assessment of electronic products and systems.