15th Triennial World Congress, Barcelona, Spain

FAULT DETECTION AND IDENTIFICATION VIA MULTI-LEVEL HYPOTHESES TESTING†

Asok Ray
Email: axr2@psu.edu

Shashi Phoha
Email: sxp26@psu.edu

The Pennsylvania State University
University Park, PA 16802, USA

Abstract: This paper formulates a recursive algorithm of multi-level hypotheses testing for real-time detection and identification of potential faults from continuous sensor signals. The usage of the recursive algorithm is illustrated on a data set of temperature sensors, collected from an operating power plant. *Copyright © 2002 IFAC*

Keywords: Fault detection; Fault diagnosis; Hypothesis; Statistical analysis; Stochastic systems

1. INTRODUCTION

Multi-level hypotheses testing provides a more precise characterization of potential faults than the bi-level fail/no-fail hypothesis testing, and is often essential for early warning and timely detection and identification of soft failures in degrading devices [Basseville '93]. The contribution of this short communication is analytical formulation of a recursive algorithm that is built upon the statistical decision-theoretic principles of multi-level hypotheses testing. The algorithm is potentially applicable to real-time condition monitoring, early warning, and fault identification in complex dynamical systems like undersea vehicles, advanced aircraft, spacecraft, and power plants.

2. MULTI-LEVEL HYPOTHESES TESTING

Let $\{\eta_k,\ k=1,2,3,\cdots\}$ be statistically independent observations of a continuous random process at consecutive sampling instants. For example, these observations could be (zero-mean) residuals obtained from noisy sensor data and/or analytical measurements.

We assume M distinct possible modes of abnormal operation (i.e., faults) in addition to the normal (i.e., no-fault) condition that is denoted as the 0^{th} mode such that exactly one of the $(M+1)$ modes is occupied at each instant. Occupancy of each of these $(M+1)$ modes is designated as an event. These $(M+1)$ events constitute a set of mutually exclusive and exhaustive Markov states. Correspondingly, the following hypotheses are defined for $i=1,2,\cdots,M$:

$$\begin{aligned} H_k^0 &: \text{Normal } a \text{ priori pdf } f^0(\bullet) \equiv f\left(\bullet \,\middle|\, H^0\right) \\ H_k^i &: \text{Abnormal } a \text{ priori pdf } f^i(\bullet) \equiv f\left(\bullet \,\middle|\, H^i\right) \end{aligned} \tag{1}$$

We assume a one-to-one correspondence between the set of $(M+1)$ events and the set of hypotheses, $H_k^j,\ j=0,1,2,\cdots,M$, of their occurrence at the k^{th} sample. The terms, event, mode, and hypothesis, are therefore synonymously used in the sequel.

We define the *a posteriori* probability π_k^j of the j^{th} event at the k^{th} sample as:

$$\pi_k^j \equiv P[H_k^j \,|\, Z_k],\ j=0,1,2,\cdots,M \tag{2}$$

based on the cumulative observations $Z_k \equiv \{z_1, z_2, \cdots, z_k\}$ over k consecutive samples where the observations, $z_i \equiv \{\eta_i \in B_i\}$, are mutually

† The reported work has been supported in part by U.S. Army Research Office under Grant No. DAAD19-01-1-0646 and National Science Foundation under Grant No. ECS-9912495.

statistically independent and B_i is the region of interest at the i^{th} sample. The sampling instants are not necessarily uniformly spaced in time.

The problem is to derive a recursive algorithm for *a posteriori* probabilities, $\pi_k^j : j = 1, 2, \cdots, M$, at the k^{th} sample in real time. This information also leads to evaluation of the total *a posteriori* probability Π_k of occurrence of any one of the M abnormal events at the k^{th} sample:

$$\begin{aligned} \Pi_k &\equiv P\left[\bigcup_{j=1}^{M} H_k^j \,|\, Z_k\right] \\ &= \sum_{j=1}^{M} P\left[H_k^j \,|\, Z_k\right] \Rightarrow \ \Pi_k = \sum_{j=1}^{M} \pi_k^j \end{aligned} \tag{3}$$

Equation (3) holds because of the exhaustive and mutually exclusive properties of the Markov states, H_k^j, $j = 1, 2, \cdots, M$. To construct a recursive relation for Π_k, we define the following:

Joint probability: $\xi_k^j \equiv P[H_k^j, Z_k]$ (4)

A priori probability: $\lambda_k^j \equiv P\left[z_k \,|\, H_k^j\right]$ (5)

Transition probability: $a_k^{i,j} \equiv P\left[H_k^j \,|\, H_{k-1}^i\right]$ (6)

Then, because of independence of the events z_k and Z_{k-1}, Eq. (4) takes the following form:

$$\begin{aligned} \xi_k^j &= P[H_k^j, z_k, Z_{k-1}] \\ &= P[z_k \,|\, H_k^j] P[H_k^j, Z_{k-1}] \end{aligned} \tag{7}$$

Furthermore, the exhaustive and mutually exclusive properties of the Markov states H_k^j, $j = 0, 1, 2, \cdots, M$, and independence of Z_{k-1} and H_k^j lead to:

$$\begin{aligned} P\left[H_k^j, Z_{k-1}\right] &= \sum_{i=0}^{M} P\left[H_k^j, H_{k-1}^i, Z_{k-1}\right] \\ &= \sum_{i=0}^{M} P\left[Z_{k-1} \,|\, H_{k-1}^i\right] P\left[H_k^j \,|\, H_{k-1}^i\right] P\left[H_{k-1}^i\right] \\ &= \sum_{i=0}^{M} P\left[H_k^j \,|\, H_{k-1}^i\right] P\left[H_{k-1}^i, Z_{k-1}\right] \end{aligned} \tag{8}$$

A combination of Eqs. (4) to (8) yields the following relation:

$$\xi_k^j = \lambda_k^j \sum_{i=0}^{M} a_k^{i,j} \xi_{k-1}^i \tag{9}$$

We introduce a new term $\psi_k^j \equiv \dfrac{\xi_k^j}{\xi_k^0}$ that reduces to the following form by use of Eq. (9):

$$\psi_k^j = \left(\frac{\lambda_k^j}{\lambda_k^0}\right)\left(\frac{a_k^{0,j} + \sum_{i=1}^{M} a_k^{i,j} \psi_{k-1}^i}{a_k^{0,0} + \sum_{i=1}^{M} a_k^{i,0} \psi_{k-1}^i}\right) \tag{10}$$

and we obtain the *a posteriori* probability π_k^j in Eq. (2) in terms of ξ_k^j and ψ_k^j as:

$$\begin{aligned} \pi_k^j &= \frac{P[H_k^j, Z_k]}{P[Z_k]} = \frac{P[H_k^j, Z_k]}{\sum_{i=0}^{M} P\left[H_k^i, Z_k\right]} \\ &= \frac{\xi_k^j}{\xi_k^0 + \sum_{i=1}^{M} \xi_k^i} = \frac{\psi_k^j}{1 + \sum_{i=1}^{M} \psi_k^i} \end{aligned} \tag{11}$$

A combination of Eqs. (3) and (11), leads to the total *a posteriori* probability Π_k as:

$$\Pi_k = \frac{\Psi_k}{1 + \Psi_k} \ \text{ with } \ \Psi_k \equiv \sum_{j=1}^{M} \psi_k^j \tag{12}$$

Two examples show how the above expressions can be realized by simple recursive relations under the following assumptions:

- ***Assumption 1*** (for Examples a and b): At the starting point (i.e., k=0), the device operates in the normal mode, i.e., $P[H_0^0] = 1$ and $P[H_0^j] = 0$ for $j = 1, 2, \cdots, M$. Therefore, in Eq. (4), $\xi_0^0 = 1$ and $\xi_0^j = 0$ for $j = 1, 2, \cdots, M$.
- ***Assumption 2*** (for Examples a and b): No transition takes place from an abnormal mode to the normal mode, i.e., $a_k^{i,0} = 0$ for $i = 1, 2, \cdots, M$, and all k. The implication is zero probability of an abnormally operating device returning to the normal operation (unless replaced or repaired).
- ***Assumption 3*** (for Examples a and b): The transition from the normal mode to any abnormal mode is equally likely. That is, if p is the *a priori* probability of failure during one sampling interval, then $a_k^{0,0} = 1 - p$ and $a_k^{0,i} = \frac{p}{M}$.
- ***Assumption 4a*** (for Example a): The transition from an abnormal mode to any abnormal mode including itself is equally likely, i.e., $a_k^{i,j} = \frac{1}{M}$ $\forall k$ and $i, j = 1, 2, \cdots, M$. The implication is a high noise-to-signal ratio or erratic behavior of instrumentation components.
- ***Assumption 4b*** (for Example b): No transition is allowed from an abnormal mode, i.e., $a_k^{i,j} = \delta_{ij}$ $\forall k$ and $i, j = 1, 2, \cdots, M$. The implication is that a device remains at any one of the abnormal modes for a long period (e.g., slow drift or bias error of a sensor).

Now a recursive relation for Ψ_k can be generated based on the assumptions 1, 2, 3 and 4a, and using Eq. (10) for $j = 1, 2, \cdots, M$ to yield:

$$\psi_k^j = \frac{p + \sum_{i=1}^{M} \psi_{k-1}^i}{(1-p)M} \left(\frac{\lambda_k^j}{\lambda_k^0} \right) \text{ given } \psi_0^j = 0 \qquad (13a)$$

which is simplified via the relation $\Psi_k \equiv \sum_{i=1}^{M} \psi_k^i$ in Eq. (12):

$$\Psi_k = \left(\frac{p + \Psi_{k-1}}{(1-p)M} \right) \sum_{j=1}^{M} \frac{\lambda_k^j}{\lambda_k^0} \text{ given } \Psi_0 = 0 \qquad (14a)$$

Similarly, another recursive relation for Ψ_k can be generated based on the assumptions 1, 2, 3 and 4b, and using Eq. (10) for $j = 1, 2, \cdots, M$ to yield:

$$\psi_k^j = \frac{\frac{p}{M} + \psi_{k-1}^j}{1-p} \left(\frac{\lambda_k^j}{\lambda_k^0} \right) \text{ given } \psi_0^j = 0 \qquad (13b)$$

which is simplified via the relation $\Psi_k \equiv \sum_{i=1}^{M} \psi_k^i$ in Eq. (12):

$$\Psi_k = \left(\frac{1}{1-p} \right) \sum_{j=1}^{M} \left(\frac{p}{M} + \psi_{k-1}^j \frac{\lambda_k^j}{\lambda_k^0} \right) \text{given } \Psi_0 = 0 \qquad (14b)$$

If the probability measure in each abnormal mode is absolutely continuous relative to that in the normal mode, then the ratio $\lambda_k^j / \lambda_k^0$ of *a priori* probabilities converges to a Radon-Nikodym derivative as the region B_k in $z_k \equiv \{\eta_k \in B_k\}$ approaches zero measure [Wong '85]. This Radon-Nikodym derivative is simply the likelihood ratio $f^j(\eta_k) / f^0(\eta_k)$, $j = 1, 2, \cdots, M$, where $f^j(\bullet)$ is the *a priori* density function conditioned on the hypothesis H^j, $j = 0, 1, 2, \cdots, M$. Accordingly, given $\Psi_0 = 0$, the recursive relations in Eqs. (14a) and (14b) combined with Eq. (12) become:

$$\Psi_k = \left(\frac{p + \Psi_{k-1}}{(1-p)M} \right) \sum_{j=1}^{M} \frac{f^j(\eta_k)}{f^0(\eta_k)}; \quad \Pi_k = \frac{\Psi_k}{1+\Psi_k} \qquad (15a)$$

$$\psi_k^j = \left(\frac{1}{1-p} \right) \left(\frac{p}{M} + \psi_{k-1}^j \left(\frac{f^j(\eta_k)}{f^0(\eta_k)} \right) \right); \Pi_k = \frac{\sum_{j=1}^{M} \psi_k^j}{1 + \sum_{j=1}^{M} \psi_k^j} \qquad (15b)$$

Equations (15a) and (15b) recursively compute the total *a posteriori* probability Π_k based on the observations $\{\eta_k,\ k = 1, 2, 3, \cdots\}$ for different operating conditions as delineated under Assumptions 4(a) and 4(b), respectively.

3. AN APPLICATION EXAMPLE

The recursive algorithm of the multi-level hypotheses test algorithm, derived above, has been validated on a data set of temperature sensors in a 320 MWe coal-fired supercritical power plant. The set of redundant measurements of throttle steam temperature at $\sim 1040^o F\,(560^o C)$ is generated by four temperature sensors installed at different spatial locations of the main steam header that carries superheated steam from the steam generator into the high-pressure turbine via the throttle valves and governor valves. The readings of all four temperature sensors were collected over a period of 10 hours at the sampling frequency of once every 15 seconds. For this specific application, the filter parameters of the hypotheses test algorithm are selected as described below.

3.1 Filter Parameters

In this section, we evaluate the parameters and functions that are necessary for the hypotheses testing algorithm. The noise associated with each sensor output is assumed to be additive Gaussian that assures existence of the likelihood ratios in Eqs. (13a).and (13b).

The set of four temperature sensing instrumentation that are appropriately calibrated for zero bias error is modeled at the k^{th} sample as:

$$y_k = H x_k + \varepsilon_k \qquad (16)$$

where

y_k is the (4×1) sensor data vector;

H is the (4×1) *a priori* determined matrix of scale factor having rank 1. After conversion of the sensor data into engineering units, the scale factor matrix becomes: $H = [1 \quad 1 \quad 1 \quad 1]^T$;

x_k is the (1×1) vector of true (unknown) value of the measured temperature;

ε_k is the (4×1) vector of additive noise, such that $E[\varepsilon_k] = 0$ and $E(\varepsilon_k \varepsilon_\ell^T) = R_k \delta_{k\ell}$ with $R_k > 0$.

The noise associated with each of the four similar sensors was found to be stationary Gaussian and independent and identically distributed so that $R_k = R = \sigma^2 I_{4\times 4}$.

We now construct the (3×1) parity residual vector η_k from the sensor vector y_k, which is defined [Potter '77; Chow '80; Ray 91] as:

$$\eta_k = V y_k \qquad (17)$$

where the rows of the parity matrix $V \in \Re^{3\times 4}$ form an orthonormal basis of the left null space of the scale factor matrix $H \in \Re^{4\times 1}$ in Eq. (1)., i.e.,

$$VH = 0_{3\times 1}; \quad VV^T = I_{3\times 3} \qquad (18)$$

and $$V = \begin{bmatrix} \sqrt{\frac{3}{4}} & -\sqrt{\frac{1}{12}} & -\sqrt{\frac{1}{12}} & -\sqrt{\frac{1}{12}} \\ 0 & \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{6}} \\ 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} \qquad (19)$$

Note that the columns of V, often called *failure directions*, span the parity space. Under the normal condition when all sensor readings are clustered together, the magnitude of the parity residual vector η_k is small. Under an abnormal condition, if the j^{th} sensor undergoes a posittive (negative) fault, then the component of η_k along the j^{th} failure direction (i.e., j^{th} column of the V matrix in Eq. (19)) grows in the positive (negative) sense and thus identifies the faulty sensor and its failure mode [Ray '89]. Following (16), (17) and (18), the mean and covariance of parity residual vector are:

$$E(\eta_k) = 0_{3\times 1} \text{ and}$$
$$E\left(\eta_k \eta_k^T\right) = VRV^T = \sigma^2 I_{3\times 3} \qquad (20)$$

The structures of the *a priori* conditional density functions for a three-level ($M = 2$) hypotheses test based on the time series of the parity residuals, are chosen as follows:

$$f^0(\varphi) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2}\left(\frac{\varphi}{\sigma}\right)^2\right) \quad \textit{Normal Operation}$$
$$f^1(\varphi) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2}\left(\frac{\varphi-\theta}{\sigma}\right)^2\right) \quad \textit{High Failures} \qquad (21)$$
$$f^2(\varphi) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2}\left(\frac{\varphi+\theta}{\sigma}\right)^2\right) \quad \textit{Low Failures}$$

where σ is the standard deviation, and θ and $-\theta$ are the thresholds for high and low failures, respectively, for each component of the parity residual vector under the density functions $f^1(\bullet)$ and $f^2(\bullet)$, respectively.

The a posteriori probabilities π_k^j could ideally achieve the lower and upper bounds of 0 and 1, respectively. However, the lower bound of each π_k^j is set to p to accommodate the (non-zero) probability p of intra-sampling failure. This modification assists in achieving finite response time for fault detection from the normal operating condition. The upper bound of each π_k^j is set to $(1-\alpha)$ to account for thealloowable probability α of false alarms for each of the four sensors [Ray '89]. Numerical values of the parameters, σ, θ, p, and α that have been generated from the archived data of pwer plant operation are presented below:

- The standard deviation of the *a priori* Gaussian density functions of each sensor (measurement noise only) is: $\sigma = 2^o F\,(1.11^o C)$;
- The failure threshold parameter is: $\theta = 3^o F\,(1.67^o C)$,
- Operating experience at the power plant shows that the mean life of a resistance thermometer sensor, installed on the mean steam header, is about one year of continuous operation. For a sampling interval of 15 seconds, this information leads to: $p \approx 0.5\times 10^{-6}$
- The probability of false alarms is selected in consultation with the plant operating personnel. On the average, each sensor is allowed to generate a false alarm after approximately one year of continuous operation. For a sampling interval of 15 seconds, this information leads to: $\alpha \approx 0.5\times 10^{-6}$.

3.1 Filter Performance based on Test Data

Based on the sensor data collected from the power plant, we investigate efficacy of the proposed algorithm for early warning in the event of intermittent sensor degradation. The temperature sensors are more likely to be subjected to slow drift and bias errors than erratic behavior exhibiting a high noise-to-signal ratio. Therefore, Assumption 4b is more valid than Assumption 4a in this application and the algorithms in Eqs. (13b), (14b) and (15b) have been used.

The plates on the left hand side in Figures 1 and 2 exhibit the data and derived results under normal operation that include small deviations among the four sensor data as they are subjected to measurement noise and effects of thermo-fluid transients. The corresponding plates on the right hand side in Figures 1 and 2 represent an abnormal scenario in which a bias fault of +4.5°F (2.5°C) is injected in one of the sensors, Sensor #4, over the period of 50 to 150 units of time. This is seen by comparison of the two plates in the top row of Figure 1. Consequently, the (signed) norm of the component of η_k along the failure direction of Sensor #4, having the same sign as that of η_k, undergoes a change as seen in the right hand bottom-row plate of Figure 1. The norm of each of the remaining three components of η_k remains small.

The top row and middle row in Figure 2 show the probabilities of high failure and low failure of the four sensors, respectively. The high failure, injected in Sensor#4, induces positive growth of the parity residual norm along the respective failure direction (i.e., the fourth column of the fourth column of the matrix V). Consequently, the right hand plate in Figure 2 exhibits a significant growth in *a posteriori* probability π_k^1 of high failure for Sensor#4. Therefore, the right hand plate in the bottom row of Figure 2 shows a significant increase in the total a posteriori probability Π_k of failure in Sensor#4 within the time interval when the fault is prevalent. The probability of failure in the remaining sensors is significantly small.

4. SUMMARY AND CONCLUSIONS

A recursive algorithm is formulated and a filter software is coded for multi-level hypotheses testing of potential faults in real time. This algorithm is capable of small change detection, identification of incipient faults, and generation of early warnings for potentially pervasive failures. The usage of the recursive algorithm is illustrated on a data set of temperature sensors, collected from a power plant. The algorithm detects and identifies the faulty sensor and its failure mode. As such the algorithm could enhance the Instrumentation & Control System Software in tactical and transport aircraft, and nuclear and fossil power plants.

The algorithm is potentially applicable to real-time condition monitoring, early warning, and fault identification in complex dynamical systems. The algorithm is also suitable for identification of discrete events from continuous sensor signals in hybrid control systems.

REFERENCES

M. Basseville and I.V. Nikiforov (1993). *Detection of Abrupt Changes*, Prentice Hall, Englewood Cliffs, NJ, Ch. 5,pp. 185-225.

E.Y. Chow 1980). *A Failure Detection System Design Methodology,* (Ph.D. Dissertation, MIT, L.I.D.S.

J.E. Potter, and M.C. Suman (1977). "Thresholdless Redundancy Management with Arrays of Skewed Instruments," NATO AGARDOGRAPH-224, pp. 15-1 to 15-15.

A. Ray (1989). "Sequential Testing for Fault Detection in Multiply-Redundant Systems," *Journal Dynamic Systems, Measurement and Control*, **111(2)**, 329-332.

A. Ray and R. Luck (1991). "Signal Validation in Multiply-Redundant Systems," *IEEE Control Systems Magazine*, **11 (2)**, 44-49..

E. Wong and B. Hajek (1985). *Stochastic Processes in Engineering Systems*, Springer-Verlag, New York, Ch.1, p. 28.

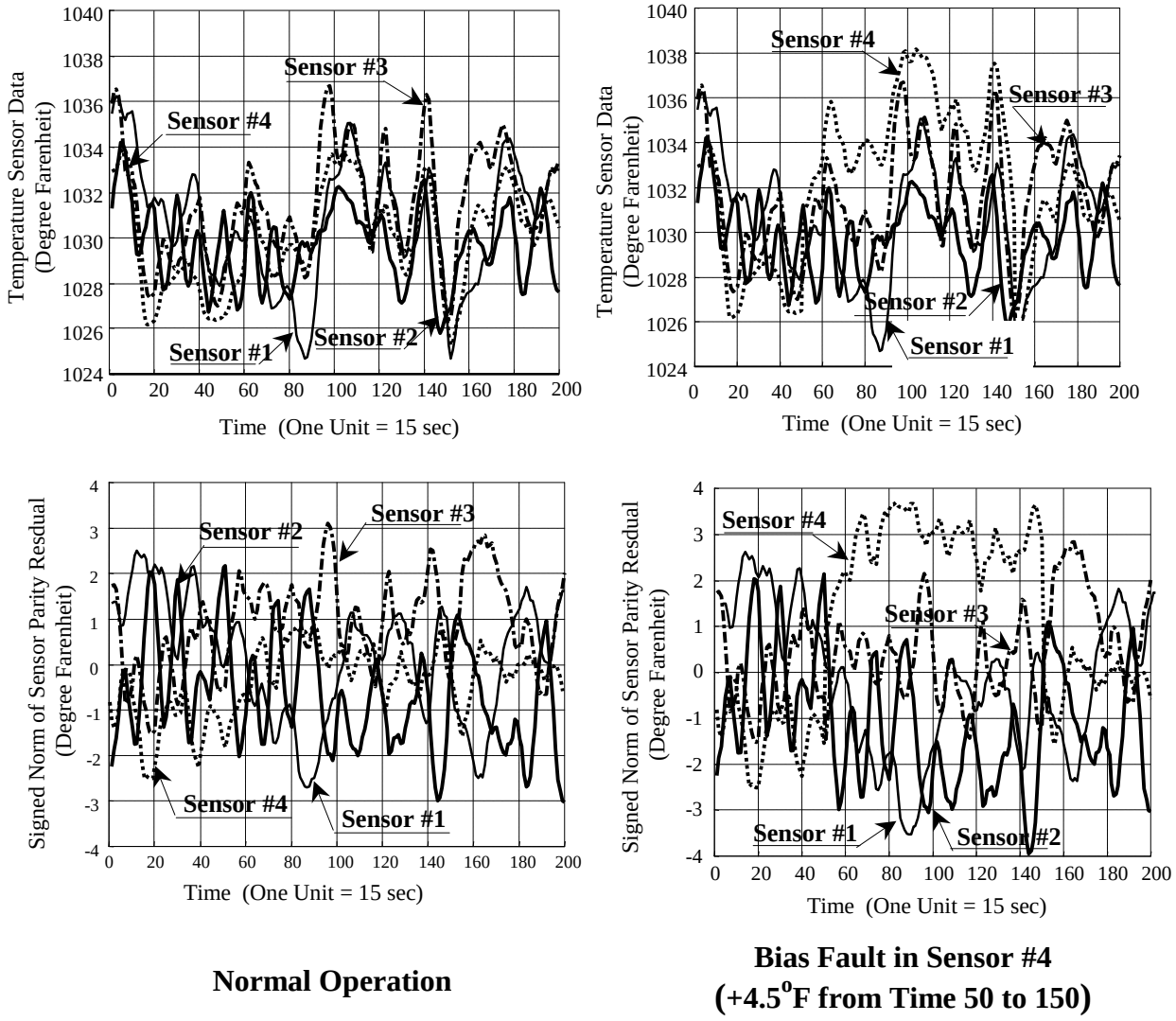
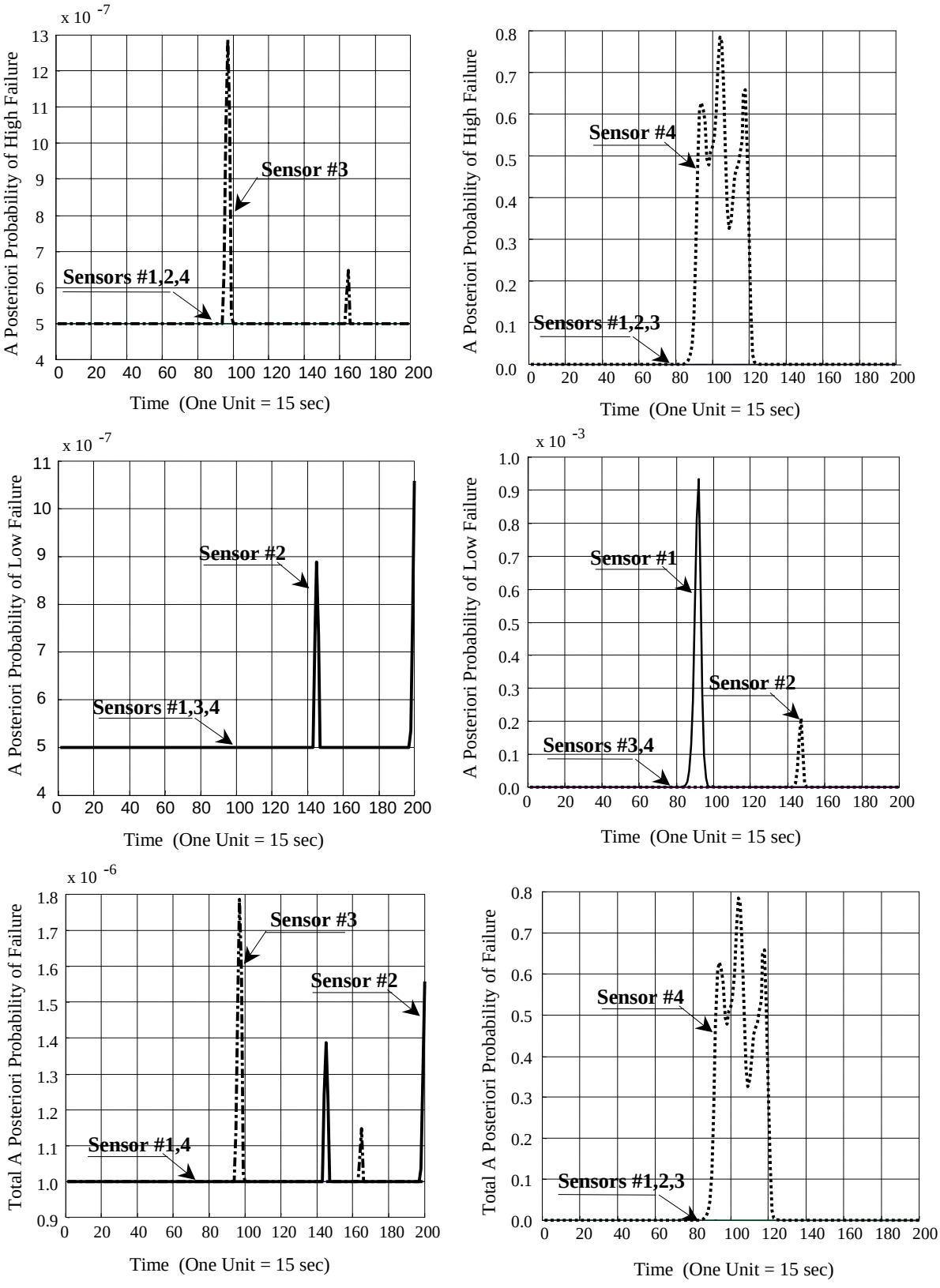


Figure 1 Sensor Data and Parity Residuals



Normal Operation

Bias Fault in Sensor #4 (+4.5°F from Time 50 to 150)

Figure 2 A Posteriori Probabilities of Failure