Signal Processing 82 (2002) 853 – 859

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Short communication

Detection and identification of potential faults via multi-level hypotheses testing ☆

Asok Ray[a,*], Shashi Phoha[b]

[a] Department of Mechanical and Nuclear Engineering, The Pennsylvania State University, 137 Reber Building, University Park, PA 16802-1412, USA

[b] Applied Research Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

Received 23 April 2001; received in revised form 9 January 2002

Abstract

This short communication 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 the Instrumentation & Control system of an operating power plant.

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 [1]. 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.

☆ The reported work has been supported in part by the U.S. Army Research Office under Grant No. DAAD19-01-1-0646.

* Corresponding author. Tel.: +1-814-865-6377; fax: +1-814-863-4848.

E-mail addresses: axr2@psu.edu (A. Ray), sxp26@psu.edu (Shashi Phoha).

2. Multi-level hypotheses testing

Let $\{\eta_k, k = 1, 2, 3, \ldots\}$ 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 0th 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 $i = 1, 2, \ldots, M$:

$$\begin{aligned} &H_k^0\text{: normal a priori pdf } f^0(\bullet) \equiv f(\bullet|H^0), \\ &H_k^i\text{: abnormal a priori pdf } f^i(\bullet) \equiv f(\bullet|H^i). \end{aligned} \tag{1}$$

We assume a one-to-one correspondence between the set of $(M + 1)$ events and the set of hypotheses, H_k^j,

0165-1684/02/$ - see front matter
PII: S0165-1684(02)00160-3

$j = 0, 1, 2, \ldots, M$, of their occurrence at the kth 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 jth event at the kth sample as

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

based on the cumulative observations $Z_k \equiv \{z_1, z_2, \ldots, z_k\}$ over k consecutive samples where the observations $z_i \equiv \{\eta_i \in B_i\}$, $i = 1, 2, \ldots,$ are mutually statistically independent and B_i is the region of interest at the ith sample. The sampling instants are not necessarily uniformly spaced in time.

The problem is to derive a recursive algorithm for a posteriori probabilities, π_k^j: $j = 1, 2, \ldots, M$, at the kth 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 kth sample

$$\Pi_k \equiv P\left[\bigcup_{j=1}^{M} H_k^j | Z_k\right]$$

$$= \sum_{j=1}^{M} P[H_k^j | Z_k] \Rightarrow \Pi_k = \sum_{j=1}^{M} \pi_k^j. \tag{3}$$

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

$$\text{Joint probability}: \quad \xi_k^j \equiv P[H_k^j, Z_k], \tag{4}$$

$$\text{A priori probability}: \quad \lambda_k^j \equiv P[z_k | H_k^j], \tag{5}$$

$$\text{Transition probability}: \quad a_k^{i,j} \equiv P[H_k^j | H_{k-1}^i]. \tag{6}$$

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

$$\xi_k^j = P[H_k^j, z_k, Z_{k-1}]$$

$$= P[z_k | H_k^j] P[H_k^j, Z_{k-1}]. \tag{7}$$

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

$$P[H_k^j, Z_{k-1}] = \sum_{i=0}^{M} P[H_k^j, H_{k-1}^i, Z_{k-1}]$$

$$= \sum_{i=0}^{M} P[Z_{k-1} | H_{k-1}^i] P[H_k^j | H_{k-1}^i] P[H_{k-1}^i]$$

$$= \sum_{i=0}^{M} P[H_k^j | H_{k-1}^i] P[H_{k-1}^i, Z_{k-1}]. \tag{8}$$

A combination of Eqs. (4)–(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 $\psi_k^j \equiv \xi_k^j / \xi_k^0$ 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

$$\pi_k^j = \frac{P[H_k^j, Z_k]}{P[Z_k]} = \frac{P[H_k^j, Z_k]}{\sum_{i=0}^{M} P[H_k^i, Z_k]}$$

$$= \frac{\xi_k^j}{\xi_k^0 + \sum_{i=1}^{M} \xi_k^i} = \frac{\psi_k^j}{1 + \sum_{i=1}^{M} \psi_k^i}. \tag{11}$$

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

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

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 $j = 1, 2, \ldots, M$. Therefore, in Eq. (4), $\xi_0^0 = 1$ and $\xi_0^j = 0$ for $j = 1, 2, \ldots, M$.

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 $i = 1, 2, \ldots, M$, 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 $a_k^{0,i} = p/M$.

Assumption 4(a) (*For Example a*). The transition from an abnormal mode to any abnormal mode including itself is equally likely, i.e., $a_k^{i,j} = 1/M \ \forall k$ and $i, j = 1, 2, \ldots, M$. The implication is a high noise-to-signal ratio or erratic behavior of instrumentation components.

Assumption 4(b) (*For Example b*). No transition is allowed from an abnormal mode, i.e., $a_k^{i,j} = \delta_{ij} \ \forall k$ and $i, j = 1, 2, \ldots, M$. 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 Assumptions 1–4(a), and using Eq. (10) for $j = 1, 2, \ldots, M$ to yield

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

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

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

Similarly, another recursive relation for Ψ_k can be generated based on Assumptions 1–4(b), and using Eq. (10) for $j = 1, 2, \ldots, M$ to yield

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

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

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

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 the expression $z_k \equiv \{\eta_k \in B_k\}$ approaches zero measure [7]. This Radon–Nikodym derivative is simply the likelihood ratio $f^j(\eta_k)/f^0(\eta_k)$, $j = 1, 2, \ldots, M$, where $f^j(\bullet)$ is the a priori density function conditioned on the hypothesis H^j, $j = 0, 1, 2, \ldots, M$. Accordingly, given $\Psi_0 = 0$, the recursive relations in Eqs. (14a) and (14b) combined with Eq. (12) become

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

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

Eqs. (15a) and (15b) recursively compute the total a posteriori probability Π_k based on the observations $\{\eta_k, k = 1, 2, 3, \ldots\}$ 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°F (560°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. Since these sensors are not spatially collocated, they become asynchronous under severe transients due to the transport lag. The information on the estimated average temperature, derived from these sensors, is used for health monitoring and damage prediction in the main steam header as well as for coordinated feedforward–feedback control of the power plant under steady-state and transient operations [3].

The readings of all four temperature sensors were collected over a period of 10 h at the sampling frequency of once every 15 sec. The collected data, after bad data suppression (e.g., elimination of obvious outliers following built-in tests such as limit check and rate check), show that each sensor exhibits temperature fluctuations due to inherent thermal-hydraulic disturbances and process transients in addition to the instrumentation noise. For this specific application, the filter parameters of the hypotheses test algorithm are selected as described below.

4. Filter parameters

In this section, we evaluate the parameters and functions that are necessary for the hypotheses testing algorithm. In this application, 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 kth sample as

$$y_k = Hx_k + \varepsilon_k, \tag{16}$$

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: $H = [1\ 1\ 1\ 1]^{\mathrm{T}}$; 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 $E(\varepsilon_k \varepsilon_\ell^{\mathrm{T}}) = R_k \delta_{k\ell}$ 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 [4,2,6] as

$$\eta_k = V y_k \tag{17}$$

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

$$VH = 0_{3\times 1}; \quad VV^{\mathrm{T}} = I_{3\times 3} \tag{18}$$

and

$$V = \begin{bmatrix} \sqrt{\frac{3}{4}} & -\sqrt{\frac{1}{12}} & -\sqrt{\frac{1}{12}} & -\sqrt{\frac{1}{12}} \\ 0 & \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{6}} \\ 0 & 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix}. \tag{19}$$

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 jth sensor undergoes a positive (negative) fault, then the component of η_k along the jth failure direction (i.e., jth column of the V matrix in Eq. (19)) grows in the positive (negative) sense and thus identifies the faulty sensor and its failure mode [6]. Following (16)–(18), the mean and covariance of parity residual vector are given as

$$E(\eta_k) = 0_{3\times 1} \quad \text{and} \quad E(\eta_k \eta_k^{\mathrm{T}}) = VRV^{\mathrm{T}} = \sigma^2 I_{3\times 3}. \tag{20}$$

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:

$$f^0(\varphi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\varphi}{\sigma}\right)^2\right)$$

normal operation,

$$f^1(\varphi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\varphi - \theta}{\sigma}\right)^2\right)$$

high failures, (21)

$$f^2(\varphi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\varphi + \theta}{\sigma}\right)^2\right)$$

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 the allowable probability α of false alarms for each of the four sensors [5]. Numerical values of the parameters, σ, θ, p, and α that have been generated from the archived data of power plant operation are presented below:

- The standard deviation of the a priori Gaussian density functions of each sensor (measurement noise only) is: $\sigma = 2^\circ$F (1.11°C).
- The failure threshold parameter is: $\theta = 3^\circ$F (1.67°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 1 year

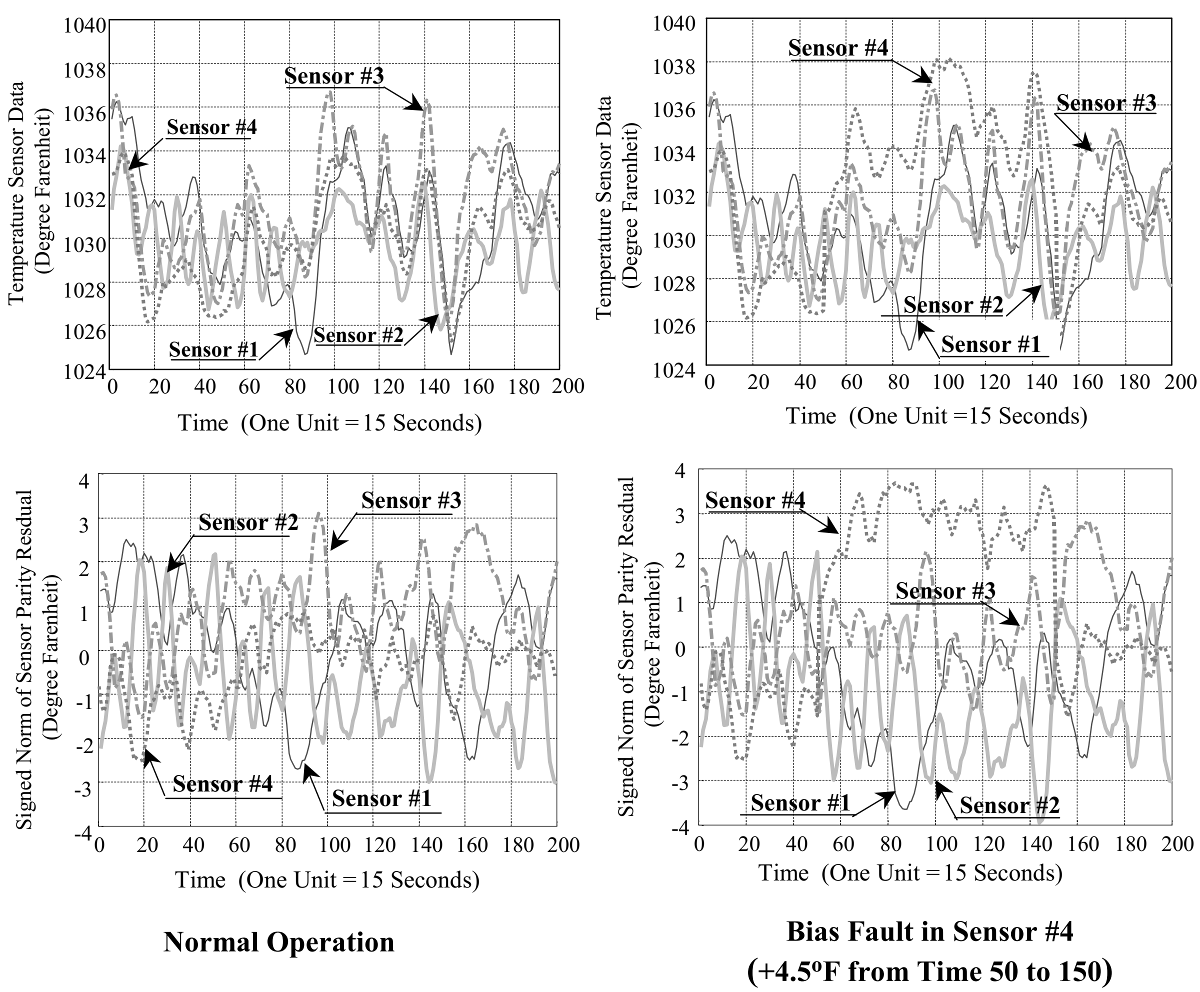

Fig. 1. Sensor data and parity residuals.

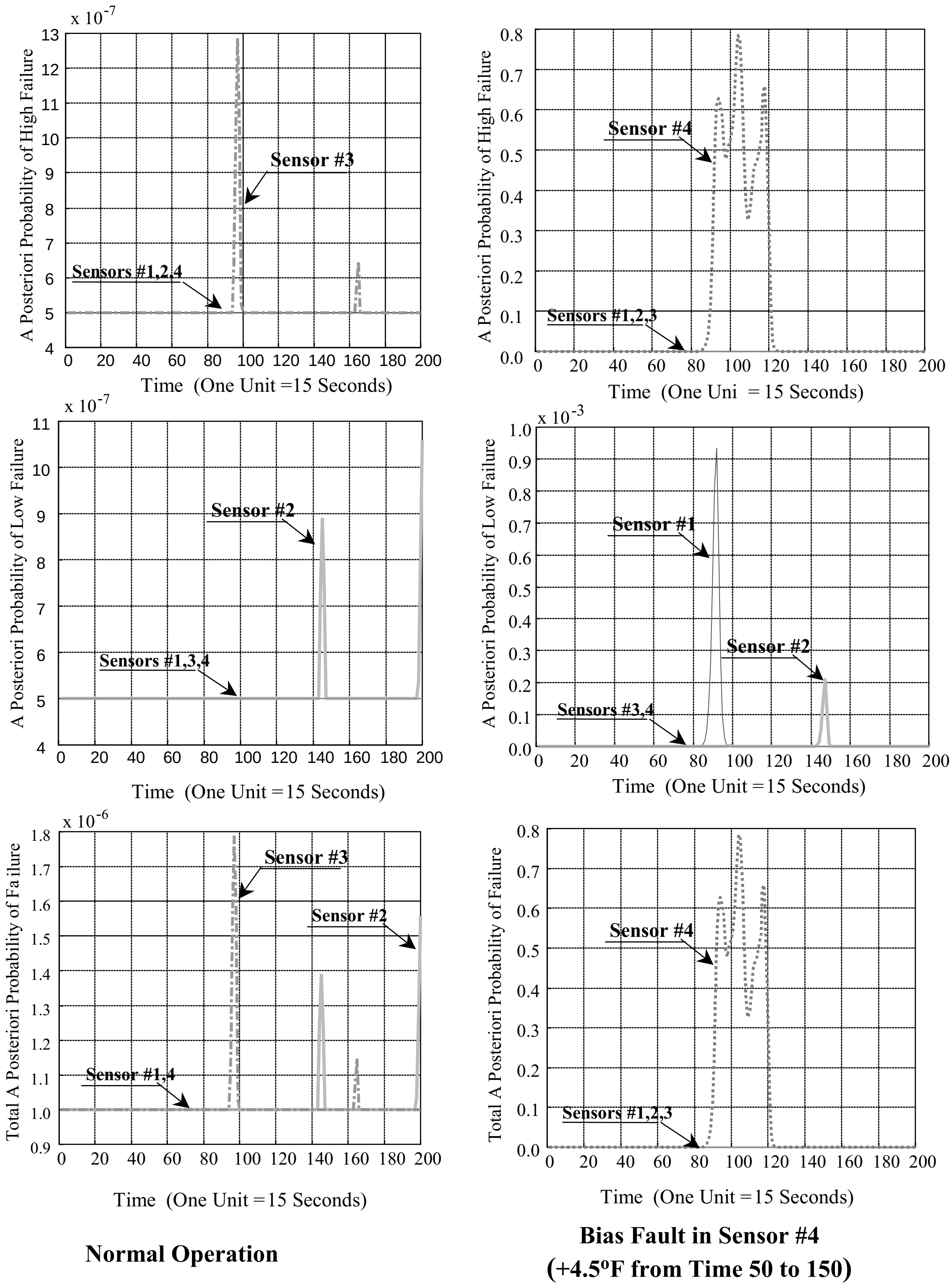

Fig. 2. A posteriori probabilities of failure.

of continuous operation. For a sampling interval of 15 sec, 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 1 year of continuous operation. For a sampling interval of 15 sec, this information leads to: $\alpha \approx 0.5 \times 10^{-6}$.

5. Filter performance based on experimental 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 4(b) is more valid than Assumption 4(a) in this application and the algorithms in Eqs. (13b), (14b) and (15b) have been used.

The plates on the left-hand side in Figs. 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 Figs. 1 and 2 represent an abnormal scenario in which a bias fault of $+4.5^{\circ}$F (2.5°C) is injected in one of the sensors, Sensor #4, over the period of 50–150 units of time. This is seen by comparison of the two plates in the top row of Fig. 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 Fig. 1. The norm of each of the remaining three components of η_k remains small.

The top row and middle row in Fig. 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 Fig. 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 Fig. 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.

6. 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 like undersea vehicles, advanced aircraft, spacecraft, and power plants. The algorithm is also suitable for identification of discrete events from continuous sensor signals and analytical measurements in hybrid control systems.

References

[1] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes, Prentice-Hall, Englewood Cliffs, NJ, 1993, pp. 185–225 (Chapter 5).

[2] E.Y. Chow, A failure detection system design methodology, Ph.D. Dissertation, MIT, L.I.D.S., 1980.

[3] M. Holmes, A. Ray, Fuzzy damage mitigating control of a fossil power plant, IEEE Trans. on Control Systems Technology 9 (1) (January 2001) 140–147.

[4] J.E. Potter, M.C. Suman, Thresholdless redundancy management with arrays of skewed instruments, NATO AGARDOGRAPH-224 (1977) 15-1–15-15.

[5] A. Ray, Sequential testing for fault detection in multiply-redundant systems, J. Dynamic Systems, Measurement and Control 111 (2) (June 1989) 329–332.

[6] A. Ray, R. Luck, Signal validation in multiply-redundant systems, IEEE Control Systems Magazine 11 (2) (February 1991) 44–49.

[7] E. Wong, B. Hajek, Stochastic Processes in Engineering Systems, Springer, New York, 1985, p. 28 (Chapter 1).