Chemometrics and Intelligent Laboratory Systems, 4 (1988) 33–45
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

A Bayesian Approach To Gross Error Detection in Chemical Process Data

Part I *: Model Development

AJIT C. TAMHANE *

Department of Industrial Engineering and Management Sciences, and Statistics, Northwestern University, Evanston, IL 60208 (U.S.A.)

CORNELIU IORDACHE ** and RICHARD S.H. MAH

Department of Chemical Engineering, Northwestern University, Evanston, IL 60208 (U.S.A.)

(Received 15 June 1987; accepted 16 February 1988)

ABSTRACT

Tamhane, A.C., Iordache, C. and Mah, R.S.H., 1988. A Bayesian approach to gross error detection in chemical process data. Part I: Model development. *Chemometrics and Intelligent Laboratory Systems*, 4: 33–45.

A new statistical test based on the Bayesian approach for detecting gross errors in chemical process data is presented in this paper. Part I gives the theoretical development of the underlying model and the proposed test, while Part II gives the results of a simulation study for assessing the performance of the test. In Part I we first develop a one-time application of the Bayes test, and then embed it in a sequential setting. For this setting a probabilistic model is proposed for updating the prior probabilities of gross error occurrences in the light of accumulating data. Modifications in the basic model are suggested to take account of unknown magnitudes of gross errors and the aging of measuring instruments. Some practical difficulties in the application of the Bayes test, e.g., adjustments for unknown delay times in detecting gross errors in instruments, adjustments of instrument lifetimes when updating their failure probabilities, and computational complexity of the scheme, are discussed and heuristic methods for their amelioration are suggested.

1 INTRODUCTION

Process data in a chemical plant are subject to omnipresent random errors as well as gross errors. The latter errors are caused by nonrandom events such as instrument biases, malfunctioning measurement devices, and leaks or depositions in process units. It is important to have efficient methods for detection and removal of gross errors because their presence invalidates the basis for further treatment of the data and confounds the methods for monitoring process performance. Therefore considerable research effort has been

* Part II: *Chemometrics and Intelligent Laboratory Systems*, 4, No. 2 (1988) in press.

** Present address: ChemShare, P.O. Box 1885, Houston, TX 77251, U.S.A.

0169-7439/88/$03.50

devoted to this problem. Tamhane and Mah [1] have given a detailed review of the various gross error detection methods studied in the literature. In this paper we shall restrict consideration to gross errors arising from measurement biases.

The central issue in gross error detection is how to enhance the frequency of correct detections (power) and to reduce the associated time delays without increasing the incidence of false detections. Thre are two ways to bring about this enhancement: First, to include as much relevant information as possible in the detection and identification procedure, and second, to utilize this information as efficiently as possible. Thus far most of the statistical methods for gross error detection have been based upon the following basic information inputs: conservation constraints, covariance matrix of measurement errors, bounds on variables, and measured process data. The widespread use of computerized data accural and management systems with capabilities for on-line access suggest the possibility of using historical information concerning the past failure data on measuring instruments for further enhancement of the test procedure performance. The present paper represents the first attempt to model the occurrences of gross errors so that such historical information can be efficiently utilized.

The Bayesian approach in statistics [2,3] provides a natural framework for this purpose. It has been employed for outlier detection by several authors including Box and Tiao [4], de Alba and Van Ryzin [5], and Kitagawa and Akaike [6]. Its use in reliability applications has been extensively studied in Martz and Waller [7]. However, it has not been applied to the problem of gross error detection in process data. The objective of the present two-part paper is to develop this application of the Bayesian approach. Part I develops the basic model and the theory of the Bayes test, while Part II reports the results of a simulation study for assessing its performance, particularly in relation to the measurement test proposed by Mah and Tamhane [8], which is based on the classical Neyman-Pearson approach to hypothesis testing.

Another point of departure in the present research is that it takes into account the fact that the gross error detection tests are applied sequentially over time to the process data. The conventional tests have not considered this aspect as they were not designed to use the past data. For the Bayes test, however, the extent of past data, and how they are utilized in updating the prior information are important aspects.

The summary of Part I is as follows. Section 2 introduces the basic model for process data and gross errors. Section 3 considers a one-time application of the Bayes test for a single snap-shot of the process. Section 4 considers the sequential setting. For this setting a model is proposed for proper updating of the information on the occurrences of gross errors. Some modifications of the basic model resulting from the necessity to relax certain restrictive assumptions, and other difficulties encountered with the basic model are discussed in Section 5. Finally some concluding remarks pertaining to Part I are given in Section 6. Sections in Part II are numbered consecutively following this paper.

2 BASIC MODEL, ASSUMPTIONS AND NOTATION FOR A ONE-TIME APPLICATION OF THE BAYES TEST

Throughout this paper we assume that the process is in a steady state. In this and the next section we restrict attention to a one-time application of the Bayes gross error detection test. Toward this end we consider a single measurement period consisting of $N \geqq 1$ successive observations on a vector of $n \geqq 1$ variables of interest (e.g., total mass flow rates in n streams of the process network). Typically the data vectors, which are assumed to be mutually independent, are automatically sampled and recorded at regular time intervals of 0.5 to 5 minutes. We assume that gross errors can occur in any of the n measurements only at the beginning of a measurement period consisting of N data vectors, and that period is relatively short, say 10 to 30 minutes, so that no further changes take place during it; the Bayes test is applied at the end of the measurement period followed by checking and appropriate corrective actions on those measuring instruments in which gross errors are indicated by the Bayes test.

The foregoing assumptions imply that we can base the Bayes test on the average of N data

vectors; we denote this average by $\mathbf{y}: n \times 1$ which is assumed to satisfy the following linear model:

$$\mathbf{y} = \mathbf{D}x + \epsilon + \delta \otimes \eta \tag{2.1}$$

Here $\mathbf{D}: n \times p$ is a known full column rank matrix, $x: p \times 1$ is a vector of unknown true values of the state variables, $\epsilon: n \times 1$ is a vector of random errors, $\delta: n \times 1$ is a vector of known nonzero magnitudes of gross errors, $\eta: n \times 1$ is a vector of 0's and 1's, and $\otimes$ denotes the componentwise product, i.e. the ith component of $\delta \otimes \eta$ equals $\delta_i \eta_i$ $(i = 1, 2, \ldots, n)$. If $\eta_i = 1$ (resp., $= 0$), then a gross error of known magnitude δ_i is present (resp., not present) in the ith measurement $(i = 1, 2, \ldots, n)$. (Throughout this paper we follow the notation that the ith component of vector a is a_i, and the (i, j)th component of matrix $\mathbf{A}$ is a_{ij}.)

We assume that ϵ has a multivariate normal distribution with a zero mean vector and a known covariance matrix $\mathbf{Q} = (1/N)\Sigma$ (denoted by $\epsilon \sim N(\mathbf{0}, \mathbf{Q})$) where Σ is the covariance matrix of the individual data vectors. (In practical applications Σ is unknown but can be accurately estimated by cumulatively pooling its estimates over time. We do not address the problem of the estimation of Σ here.)

The vector x is assumed to satisfy the following set of linear balance constraints (e.g., mass and energy balances):

$$\mathbf{B}x = \mathbf{0} \tag{2.2}$$

where $\mathbf{B}$ is a $q \times p$ known full row rank matrix (referred to as the constraint matrix). More generally, the balance constraints may involve, apart from x, additional unmeasured state variables (which do not appear in equation (2.1)), and the constraints may be nonhomogeneous. It can be shown, however (see Crowe et al. [9] and Iordache et al. [10]), that such general constraints can be reduced to (2.2). See ref. 8 for various examples and special cases of the model specified by (2.1) and (2.2). A special case of particular interest is $\mathbf{D} = \mathbf{I}$ (where $\mathbf{I}$ denotes an identity matrix of appropriate order which equals n here); in this case the y_i's are direct observations on the unknown x_i's $(i = 1, 2, \ldots, n)$.

We next discuss the assumptions concerning the indicator vector of gross errors, η, which is the parameter vector of primary interest here. As stated before, η is assumed to remain fixed during a given measurement period. The prior probability that a gross error is present in the ith measurement at the beginning of the given measurement period will be denoted by p_i $(i = 1, 2, \ldots, n)$. Thus each η_i is a Bernoulli random variable taking values 1 and 0 with probabilities p_i and $1 - p_i$, respectively $(i = 1, 2, \ldots, n)$. Hence for the unknown vector η we have 2^n possible values which may be denoted by η_I for all subsets I of $\{1, 2, \ldots, n\}$ where the ith component of η_I equals 1 if $i \in I$ and 0 if $i \notin I$ $(i = 1, 2, \ldots, n)$. In other words, if $\eta = \eta_I$ then gross errors are present in group I of measurements. In particular, if I is an empty set (denoted by ϕ), we have $\eta_\phi = \mathbf{0}$ (the null vector) which means that all measurements are free of gross errors. Furthermore, as stated before, we will assume (at least initially) that if a gross error is present in the ith measurement, then it will have a known fixed magnitude $\delta_i \neq 0$ $(i = 1, 2, \ldots, n)$.

Let us further assume that gross errors occur independently in different measurements (or sometimes we will say that the corresponding measuring instruments 'fail' independently of each other). Then the prior probability that $\eta = \eta_I$ is given by

$$\pi_I = \prod_{i \in I} p_i \prod_{i \notin I} (1 - p_i) \tag{2.3}$$

We will refer to the π_I's as the group prior probabilities. In particular, the prior probability that no gross errors are present in any of the measurements is given by

$$\pi_\phi = \prod_{i=1}^{n} (1 - p_i) \tag{2.4}$$

Thus, given the p_i's we can compute the group prior probabilities for all 2^n possible states of nature (i.e., possible values of η).

3 ONE-TIME APPLICATION OF THE BAYES TEST

In the simplest form, the basic idea in the Bayes test is to use the current data to update the group prior probabilities π_I to obtain the corre-

sponding group posterior probabilities $\tilde{\pi}_I$ via the Bayes rule; if $\tilde{\pi}_{I^*} = \max_I \pi_I$ then η_{I^*} is declared as the true state of nature, i.e., the measurements in group I^* are declared to contain gross errors. If $I^* = \phi$ then all measurements are declared to be free of gross errors.

A general formula for the posterior probability of η_I is given by

$$\tilde{\pi}_I = \frac{\pi_I f(\text{data} \mid \eta_I)}{\sum_J \pi_J f(\text{data} \mid \eta_J)} \tag{3.1}$$

where $f(\text{data} \mid \eta_I)$ denotes the probability density function (p.d.f.) of the data given that the true state of nature is η_I, and the summation in the denominator is over all 2^n subsets J of $\{1, 2, \ldots, n\}$. Now, for the data we cannot use y directly because its p.d.f. involves the unknown true state variable vector x. Instead we shall use a transformed vector $\mathbf{C}y$ such that the p.d.f. of $\mathbf{C}y$ is free of x and depends only on η; here $\mathbf{C}$ is an $m \times n$ nonrandom matrix whose row dimension m will become clear from the following discussion.

The transformed vector $\mathbf{C}y$ will have a multivariate normal distribution with mean vector $\mathbf{C}\mathbf{D}x + \mathbf{C}(\delta \otimes \eta)$ and covariance matrix $\mathbf{C}\mathbf{Q}\mathbf{C}'$. Therefore $\mathbf{C}$ must be such that $\mathbf{C}\mathbf{D}x = \mathbf{0}$ (where $\mathbf{0}$ denotes a null vector of appropriate dimension). This condition, using (2.2), is equivalent to

$$\mathbf{C}\mathbf{D} = \mathbf{A}\mathbf{B} \tag{3.2}$$

for some $m \times q$ matrix $\mathbf{A}$. Moreover, $\mathbf{C}\mathbf{Q}\mathbf{C}'$ must be nonsingular (for $\mathbf{C}y$ to have a p.d.f.), i.e., $\mathbf{C}$ and hence $\mathbf{A}$ must be full row rank. Then using $\mathbf{C}y$ for data in (3.1) we have

$$\tilde{\pi}_I = \Big(\pi_I \exp\Big\{-\tfrac{1}{2}(y-\delta_I)'\mathbf{C}'(\mathbf{C}\mathbf{Q}\mathbf{C}')^{-1} \times \mathbf{C}(y-\delta_I)\Big\}\Big) \Big/ \Big(\sum_J \pi_J \exp\Big\{-\tfrac{1}{2}(y-\delta_J)' \times \mathbf{C}'(\mathbf{C}\mathbf{Q}\mathbf{C}')^{-1}\mathbf{C}(y-\delta_J)\Big\}\Big) \tag{3.3}$$

where $\delta_I = \delta \otimes \eta_I$, i.e., the ith component of δ_I equals δ_i if $i \in I$ and 0 if $i \notin I$. Note that any nonsingular transformation of $\mathbf{C}y$ yields the same posteriors $\tilde{\pi}_I$.

A possible choice for $\mathbf{C}$ is one that corresponds to $\mathbf{A} = \mathbf{I}$ in (3.2) and has $m = q$ rows, e.g.,

$$\mathbf{C} = \mathbf{B}(\mathbf{D}'\mathbf{Q}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{Q}^{-1} \tag{3.4}$$

However, such a choice is not unique, e.g., an alternate choice is

$$\mathbf{C} = \mathbf{B}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' \tag{3.5}$$

Note that (3.5) cannot be obtained from (3.4) by a nonsingular transformation, and thus yields different posteriors $\tilde{\pi}_I$.

The above discussion raises the following question: What is the maximum row dimension m for the transformation matrix $\mathbf{C}$? It is desirable to use the maximal dimensional transformation because such a transformation involves minimal loss of information about η, the parameter vector of interest. An algebraic lemma in the Appendix shows that the maximum value of m is $n - p + q$. The Appendix also gives a derivation of a transformation of this dimensionality, which results in the following formula for the posterior probability:

$$\tilde{\pi}_I = \frac{\pi_I \exp\{-\frac{1}{2}(y-\delta_I)'\mathbf{W}(y-\delta_I)\}}{\sum_J \pi_J \exp\{-\frac{1}{2}(y-\delta_J)'\mathbf{W}(y-\delta_J)\}} \tag{3.6}$$

where $\mathbf{W}$ is defined in equation (A.14) of the Appendix.

We finally note that since the Bayes rule is the maximum posterior decision rule here (declare η_{I^*} to be the true state of nature if $\tilde{\pi}_{I^*} = \max_I \tilde{\pi}_I$), and since all the $\tilde{\pi}_I$'s share a common denominator, it suffices to calculate only their numerators and base the decision on them. Furthermore, it is computationally simpler to calculate the logarithms of the numerators of (3.6). Thus the final form of the decision rule used in our work is:

Decide $\eta = \eta_{I^*}$ if I^* maximizes

$$\{\ln \pi_I - \tfrac{1}{2}(y-\delta_I)'\mathbf{W}(y-\delta_I)\} \text{ over all } I \tag{3.7}$$

4 SEQUENTIAL APPLICATION OF THE BAYES TEST

We assume that the data vectors are sampled at regular time intervals and the Bayes test derived in Section 3 is applied after every N samplings of the data vectors. Let us index the time periods when the gross error detection test is applied by t ($t = 1, 2, \ldots$); $t = 0$ will denote the beginning of the first time period.

Let $y(t): n \times 1$ denote the average of $N \geqq 1$ data vectors observed during the time period t;

$y(t)$ is assumed to follow (in analogy with (2.1)) the linear model

$$y(t) = \mathbf{D}x + \epsilon(t) + \delta \otimes \eta(t-1) \quad (t = 1, 2, \ldots) \tag{4.1}$$

In (4.1) x is subject to balance constraints (2.2), $\epsilon(t)$ is a vector of random errors which is assumed to have an $N(\mathbf{0}, \mathbf{Q})$ distribution, and $\eta(t-1) = (\eta_1(t-1), \ldots, \eta_n(t-1))'$ is the indicator vector of gross errors present at time $t-1$, i.e., at the beginning of time period t, where $\eta_i(t-1) = 1$ (resp., $\eta_i(t-1) = 0$) if a gross error is present (resp., not present) in the ith measurement at time $t-1$ ($i = 1, 2, \ldots, n$, $t = 1, 2, \ldots$).

Note that **B**, **D**, **Q**, and x are assumed to remain fixed over the time horizon of interest; this corresponds to the steady state assumption stated at the beginning of Section 2. Although mathematically it causes no difficulty if these quantities vary with t, the assumption of knowing **B**, **D** and **Q** for every t would be untenable in practice. Also, fluctuations in these quantities may give rise to time lags that the model is not equipped to handle. So it seems safer to restrict the present analysis to the steady state situation.

Let the prior probability that a gross error (of magnitude δ_i) is present in measurement i at the beginning of time period t, i.e., at time $t-1$, be $p_i(t-1)$ ($i = 1, 2, \ldots, n$). By the independence assumption stated in Section 2, the group prior probability at time $t-1$ is given by (in analogy with (2.3))

$$\pi_I(t-1) = \prod_{i \in I} p_i(t-1) \prod_{i \notin I} (1 - p_i(t-1)) \tag{4.2}$$

Given the set of group prior probabilities $\pi_I(t-1)$ and the averaged data vector $y(t)$, group posterior probabilities $\tilde{\pi}_I(t)$ are calculated using (cf. (3.6)):

$$\tilde{\pi}_I(t) = \left(\pi_I(t-1) \exp\left\{ -\tfrac{1}{2} (y(t) - \delta_I)' \times \mathbf{W}(y(t) - \delta_I) \right\} \right) \Big/ \left(\sum_J \pi_J(t-1) \times \exp\left\{ -\tfrac{1}{2} (y(t) - \delta_J)' \mathbf{W} (y(t) - \delta_J) \right\} \right) \tag{4.3}$$

The Bayes decision rule is applied using (3.7). Let $I^* = I^*(t)$ be the group of instruments that is declared to contain gross errors. If $I^*(t)$ is nonempty then instruments in that group are checked. If $\eta_i(t-1) = 0$ and $i \in I^*(t)$ then we have a type I error for measurement i at time t. We assume that upon checking the measuring instrument it is known whether a type I error is committed, and corrective actions (e.g., repair or replacement) are taken only on those instruments $i \in I^*(t)$ that are actually found to contain gross errors. Note that if $\eta_i(t-1) = 1$ and $i \notin I^*(t)$ then we have a type II error for that instrument. Type II errors are not detected because the corresponding instruments are not checked.

We assume that corrective actions are perfect in the sense that after a corrective action is taken, the measuring instrument is completely free of a gross error. Thus if $i \in I^*(t)$ then $\eta_i(t) = 0$. If no corrective action is taken because $i \notin I^*(t)$ (in which case the true state of instrument i is not known), and if $\eta_i(t-1) = 1$ then $\eta_i(t) = 1$.

The model described thus far is essentially the same as that for a one-time application of the Bayes test, except that we now have a time parameter t. The main new feature in the sequential setting is the incorporation of the previous experiences with 'failures' of measuring instruments (i.e., gross error occurrences in them) in updating the prior probabilities $p_i(t-1)$. We now proceed to propose a probabilistic model for this purpose.

Initially we will assume that instrument failures follow independent Bernoulli processes with a constant (with respect to time) failure rate θ_i for the ith instrument ($i = 1, 2, \ldots, n$). In other words, the probability that the ith instrument fails (if it is not already in a failed state) when it is τ time periods old ($\tau = 1, 2, \ldots$) is the same, namely θ_i ($i = 1, 2, \ldots, n$), and the failures of a given instrument in different time periods are independent. (By the perfect corrective action assumption, an instrument is not assumed to fail at $\tau = 0$, i.e., immediately after it is repaired or replaced, and a repaired instrument is as good as new.) Thus the probability that instrument i lasts exactly τ time periods is given by the geometric distribution:

$$\theta_i (1 - \theta_i)^{\tau - 1} \quad \tau = 1, 2, \ldots \ (i = 1, 2, \ldots, n) \tag{4.4}$$

Let $\tau_i(t)$ denote, at time t, the time since instrument i was last checked. If instrument i has never been checked since the beginning of the test ($t = 0$) then we put $\tau_i(t) = t$. We refer to $\tau_i(t)$ as the age of the ith instrument at time t. Note that $\tau_i(t)$ is not always the actual age because the instrument's last checking may have been the result of a type I error (false detection), and therefore it may not have been repaired or replaced at that time. However, in such a case the information that the instrument had not failed at that time together with the memoryless property of the geometric distribution imply that probabilistically the instrument can be regarded as starting a new life at the time it was last checked. If the instrument was repaired or replaced at the time of its last checking then, of course, the instrument can be regarded as physically starting a new life (since repair is assumed to be perfect). Also note that the instrument may have failed before time t but the Bayes test may not have detected it by time t.

The ideas introduced thus far will become clearer if one plots how $\eta_i(t)$ and $\tau_i(t)$ change with t. Sample realizations of these two quantities are plotted in Fig. 1. In this figure $t_i^{(j)} \geqq 1$ and $d_i^{(j)} \geqq 1$ denote, respectively, for instrument i the time in service (lifetime) before its jth failure, and the delay in detecting that failure ($j = 1, 2, \ldots$). Note that the $t_i^{(j)}$'s are not observable but the $(t_i^{(j)} + d_i^{(j)})$'s are observable. Also note that the $\eta_i(t)$-process is a simple level change process. A final observation is that $\tau_i(t)$ is simply the time since the last detection (correct or false).

Now the conditional (for given θ_i) prior probability that instrument i is in a failed state at time $t-1$ (i.e., $\eta_i(t-1) = 1$) is given by

$$\begin{aligned} p_i(t-1 \mid \theta_i) &= P\{\text{the } i\text{th instrument has lifetime} \\ &\qquad \leqq \tau_i(t-1) \mid \theta_i\} \\ &= \sum_{\tau=1}^{\tau_i(t-1)} \theta_i(1-\theta_i)^{\tau-1} \quad \text{(using (4.4))} \\ &= 1 - (1-\theta_i)^{\tau_i(t-1)} \quad (i = 1, 2, \ldots, n) \end{aligned} \tag{4.5}$$

The instrument failure probabilities θ_i are unknown parameters, so the $p_i(t-1 \mid \theta_i)$'s defined above cannot be computed. We will put prior distributions on the θ_i's and uncondition $p_i(t-1 \mid \theta_i)$'s with respect to these prior distributions to obtain unconditional prior probabilities $p_i(t-1)$. This approach has another advantage; namely, additional failure experiences with instruments can be used to update the prior distributions on the θ_i's that reflect our additional knowledge about them.

Convenient prior distributions for this purpose are independent beta distributions since they are conjugate priors with respect to the geometric distributions (4.4) (ref. 11, p. 395) that are followed by the instrument lifetimes. Let the prior distribution on θ_i be a beta distribution with parameters l_i and m_i:

$$f_i(\theta_i) = \frac{\Gamma(l_i + m_i)}{\Gamma(l_i)\Gamma(m_i)} \theta_i^{l_i - 1}(1-\theta_i)^{m_i - 1} \quad (i = 1, 2, \ldots, n) \tag{4.6}$$

where $\Gamma(\cdot)$ denotes the gamma function. Using this prior distribution we obtain the following expression for the unconditional prior probability that instrument i is in a failed state at time $t-1$ when its age is $\tau_i(t-1)$:

$$\begin{aligned} p_i(t-1) &= \int_0^1 p_i(t-1 \mid \theta_i) f_i(\theta_i)\, \mathrm{d}\theta_i \\ &= \int_0^1 \left\{1 - (1-\theta_i)^{\tau_i(t-1)}\right\} \\ &\qquad \times \frac{\Gamma(l_i + m_i)}{\Gamma(l_i)\Gamma(m_i)} \theta_i^{l_i - 1}(1-\theta_i)^{m_i - 1}\, \mathrm{d}\theta_i \\ &= 1 - \frac{\Gamma(l_i + m_i)\Gamma(m_i + \tau_i(t-1))}{\Gamma(l_i)\Gamma(l_i + m_i + \tau_i(t-1))} \\ &\qquad (i = 1, 2, \ldots, n) \end{aligned} \tag{4.7}$$

These $p_i(t-1)$-values will be substituted in (4.2) to obtain the group prior probabilities $\pi_I(t)$.

In (4.6) the parameters l_i and m_i have the following interpretation: l_i is the number of previous failures for instrument i and m_i is the sum of previous lifetimes for instrument i ($i = 1, 2, \ldots, n$). The mean of the prior distribution (4.6) is $\hat{\theta}_i = l_i/(l_i + m_i)$ ($i = 1, 2, \ldots, n$).

Given the starting values $l_i^{(0)}$ and $m_i^{(0)}$ of the parameters l_i and m_i for the beta prior distribu-

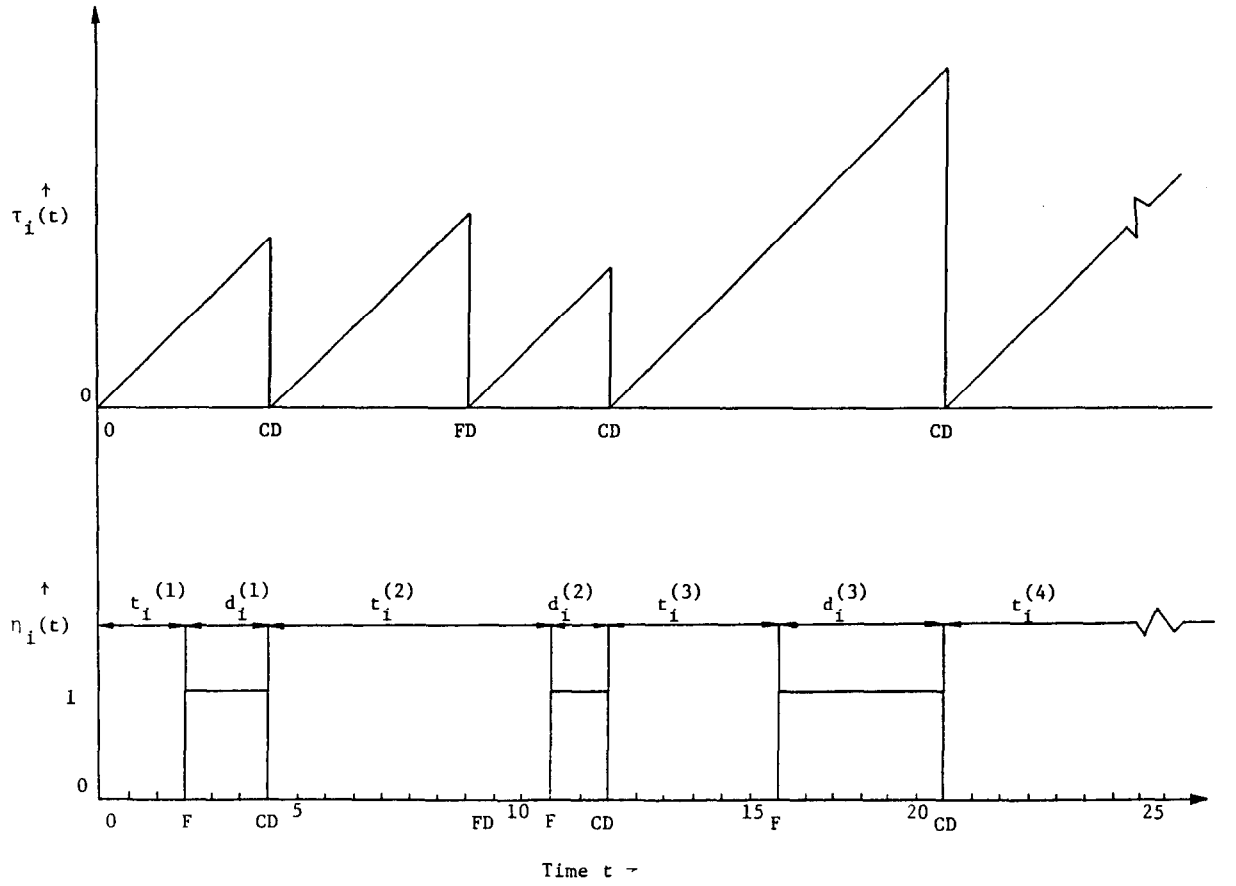

Fig. 1. Sample realizations of $\eta_i(t)$ and $\tau_i(t)$. F = Failure; CD = correct detection; FD = false detection.

tion for θ_i, if the first failure of the ith measuring instrument occurs after it has been in service for $t_i^{(1)}$ time periods, then it can be shown (see ref. 11, p. 396) that θ_i will have a beta posterior distribution with parameters

$$l_i^{(1)} = l_i^{(0)} + 1 \quad \text{and} \quad m_i^{(1)} = m_i^{(0)} + t_i^{(1)} - 1$$
$$(i = 1, 2, \ldots, n) \tag{4.8}$$

Note that $l_i^{(1)}$ is the new number of failures observed over a new total of $l_i^{(1)} + m_i^{(1)} = l_i^{(0)} + m_i^{(0)} + t_i^{(1)}$ time periods. Applying (4.8) recursively we see that after $a \geqq 1$ failures of the ith instrument have occurred, the lifetime corresponding to the jth failure being $t_i^{(j)}$ $(j = 1, 2, \ldots, a)$, the failure probability θ_i will have a beta posterior distribution with parameters

$$l_i^{(a)} = l_i^{(0)} + a \quad \text{and} \quad m_i^{(a)} = m_i^{(0)} + \sum_{j=1}^{a} t_i^{(j)} - a$$
$$(i = 1, 2, \ldots, n) \tag{4.9}$$

The Bayes estimate of θ_i will be the mean of this posterior distribution, which is

$$\hat{\theta}_i^{(a)} = \frac{l_i^{(a)}}{l_i^{(a)} + m_i^{(a)}} = \frac{l_i^{(0)} + a}{l_i^{(0)} + m_i^{(0)} + \sum_{j=1}^{a} t_i^{(j)}}$$
$$(i = 1, 2, \ldots, n) \tag{4.10}$$

From (4.9) we see that the larger the values of $l_i^{(0)}$ and $m_i^{(0)}$ (for a given prior mean $\hat{\theta}_i^{(0)} = l_i^{(0)}/(l_i^{(0)} + m_i^{(0)})$) the greater the influence the prior distribution will have, and the less the influence the observed failure data will have on the posterior distribution of θ_i $(i = 1, 2, \ldots, n)$. Therefore by a suitable choice of the values of $l_i^{(0)}$ and $m_i^{(0)}$, one can adjust the relative influences of the prior information and the data on the posterior of θ_i $(i = 1, 2, \ldots, n)$.

The following method of choosing $l_i^{(0)}$ and $m_i^{(0)}$ was proposed by Colombo and Constantini [12]: Define

$$s_i = \frac{(l_i^{(0)} + m_i^{(0)})(l_i^{(0)} + 1)}{l_i^{(0)}(l_i^{(0)} + m_i^{(0)} + 1)} \quad (i = 1, 2, \ldots, n) \tag{4.11}$$

which is the ratio of $\hat{\theta}_i^{(1)}$ to $\hat{\theta}_i^{(0)}$ if $t_i^{(1)} = 1$, i.e., if the failure occurs immediately after the first time period. Note that $1 < s_i < 2$,

$$l_i^{(0)} = \frac{1 - s_i\hat{\theta}_i^{(0)}}{s_i - 1} \quad \text{and} \quad m_i^{(0)} = \frac{(1 - \hat{\theta}_i^{(0)})l_i^{(0)}}{\hat{\theta}_i^{(0)}}$$

$$(i = 1, 2, \ldots, n) \tag{4.12}$$

Colombo and Constantini [12] suggest that one should choose $\hat{\theta}_i^{(0)}$ which is a prior estimate of θ_i without any current data, and s_i which tells by what factor one is willing to change the prior estimate $\hat{\theta}_i^{(0)}$ if the failure is observed immediately after the first time period; then $l_i^{(0)}$ and $m_i^{(0)}$ can be computed from (4.12). Note that choosing a large value of s_i (close to 2) yields small values of $l_i^{(0)}$ and $m_i^{(0)}$ which means that one is giving less weight to the prior information and more weight to the current data, and vice versa.

5 SOME MODIFICATIONS IN THE BASIC MODEL

Three assumptions that have been made in the basic model but that are not generally satisfied in practice are:

(i) The δ_i's are fixed known constants.
(ii) The instrument failure probabilities θ_i are independent of the instrument ages.
(iii) Checking and corrective actions are immediate and perfect.

In this section we relax these assumptions and discuss how the basic model needs to be modified accordingly. We also discuss some other difficulties inherent in the basic model, and how we deal with them.

5.1 Estimation of the gross error magnitudes

First we consider assumption (i) above. In practice the δ_i's are unknown and typically random quantities. Not only are the δ_i's random, they are also time-varying. We will not model all of these complexities here. Rather we will only consider the case that the δ_i's are unknown but fixed constants. In the following we give a method of estimating the δ_i's. It would have been logical in the present Bayesian framework to put prior distributions on the δ_i's and update them in light of the accumulating data. However, this fully Bayesian approach appears analytically intractable. Also it seems practically difficult to specify prior distributions for all the unknown parameters. Hence the following ad-hoc method is used instead.

From (A.5) we have

$$E(\boldsymbol{r}) = (\mathbf{I} - \mathbf{DM})(\boldsymbol{\delta} \otimes \boldsymbol{\eta}) \tag{5.1}$$

where $\boldsymbol{r}$ is the vector of residuals defined in (A.2). If we set $\boldsymbol{\eta}$ equal to the vector of all 1's, then $\hat{\boldsymbol{\delta}}$ can be obtained by solving the equation

$$(\mathbf{I} - \mathbf{DM})\boldsymbol{\delta} = \boldsymbol{r} \tag{5.2}$$

However, generally not all elements of $\boldsymbol{\eta}$ are 1, i.e., gross errors are not present simultaneously in all instruments. Permitting all elements of $\boldsymbol{\eta}$ to be 1 can lead to misleading estimates of the δ_i's. Hence we allow only those elements of $\boldsymbol{\eta}$ to be 1 that correspond to the measurements that are indicated to contain gross errors by the Bayes test. (In Section 5.3 we see that immediate checking is generally not feasible, and therefore η_i's for some measurements with type I errors are also allowed to be 1.) Let I be this group of measurements and let $\boldsymbol{\delta}_I = \boldsymbol{\delta} \otimes \boldsymbol{\eta}_I$. Then instead of (5.2) we solve the equation

$$(\mathbf{I} - \mathbf{DM})\boldsymbol{\delta}_I = \boldsymbol{r} \tag{5.3}$$

using the least squares method of Golub and

Reinsch [13]. The nonzero elements of the solution vector $\hat{\boldsymbol{\delta}}_I$ provide estimates of the δ_i's for $i \in I$. An alternative method of estimating δ_i's was given by Romagnoli [14].

To obtain more stable estimates of the δ_i's, weighted averages of the last estimates and the current estimates were used in our work. Note that estimates of the δ_i's are not updated for those measurements for which no gross errors are indicated by the Bayes test. To start the process, initial estimates of the δ_i's are provided for all measurements. The method of computing weighted averages is discussed in Part II. The estimates of the δ_i's obtained in a given period are used as 'known' fixed values when applying the Bayes test in the following period.

5.2 Instruments with increasing failure rates

We next turn to assumption (ii). Generally speaking, the θ_i's will not be constant but will be increasing with the ages of the instruments, i.e., the instruments will have increasing failure rates. Let $\theta_i(T_i)$ denote the failure probability for the ith instrument when its actual age is T_i ($i = 1, 2, \ldots, n$). Note that here we use the actual age rather than the time since the last checking of the instrument.

In Part II, in which we study the performance of the Bayes test by Monte Carlo simulation techniques, we use the following model for $\theta_i(T_i)$ for generation of gross errors in instrument i:

$$\theta_i(T_i) = \begin{cases} 0 & \text{for } T_i = 0 \\ 1 - (1 - \theta_i(1)) \exp\{-\beta_i(T_i - 1)\} & \\ & \text{for } T_i = 1, 2, \ldots \end{cases} \tag{5.4}$$

where $0 < \theta_i(1) < 1$ and $\beta_i \geqq 0$ are given constants ($i = 1, 2, \ldots, n$). For $\beta_i = 0$ we obtain the constant failure rate model with $\theta_i(T_i) = \theta_i(1) = \theta_i$ (say). Furthermore, for $\beta_i > 0$, $\theta_i(T_i)$ increases with T_i (increasing failure rate) and as $T_i \to \infty$, $\theta_i(T_i) \to 1$ ($i = 1, 2, \ldots, n$). It can also be checked that this model satisfies Salvia and Bollinger's conditions for a proper distribution [15].

It should be emphasized that this model is used only for simulating the occurrence of gross errors. The Bayes test for their detection is still based on the constant θ_i assumption as described before. A Bayes test based on this model will be prohibitively complicated, and has not been worked out yet. This part of the simulation exercise will help us study the robustness of the Bayes detection scheme to the violation of the constant failure rate assumption.

5.3 Delays and imperfections in checking and corrective actions

We now turn to assumption (iii) which refers to immediate checking of instruments (followed by corrective actions if necessary) if they are indicated to contain gross errors. Usually this is not feasible in practice. First, one may want to see some sustained evidence of the presence of a gross error in a measurement before deciding to take a corrective action. Second, even after seeing sustained evidence, one may wish to postpone checking and corrective actions to a more convenient time (e.g., the end of the shift or the scheduled inspection/maintenance time).

In response to the first consideration (and also to reduce the excessive occurrence of type I errors) we adopted a rule which declares that a gross error is present in a given measurement if in at least 2 out of 3 successive time periods the Bayes test indicates a gross error in that measurement (referred to as the 2 out of 3 or 2/3 deferred decision rule). The second consideration implies that when a gross error is indicated by the Bayes test in a given measurement, it is usually not known until some time later (when the instrument is checked) whether the gross error is actually present (no type II error) or not (a type I error). Until then, for the purposes of estimating the δ_i's, a gross error is assumed to have been detected (but not confirmed) in that instrument. However, for the purposes of updating the l_i's and m_i's (using (4.9)), and hence the θ_i's (using (4.10)), we only use the confirmed correct detections (i.e., a = the number of actual gross error occurrences). Thus l_i, m_i and θ_i are updated only when the ith

flagged instrument is checked and found to contain a gross error.

5.4 Some other modifications and difficulties

One difficulty in implementing the basic model as stated in Section 4 has to do with the updating of the beta prior distribution parameters $l_i^{(0)}$ and $m_i^{(0)}$ according to equation (4.9). Notice that we need to know the lifetimes $t_i^{(j)}$ in order to apply this equation. However, as noted in Section 4, we do not observe the $t_i^{(j)}$'s but only the $(t_i^{(j)} + d_i^{(j)})$'s. Exactly how to adjust these observable upper bounds on the $t_i^{(j)}$'s downward is not entirely clear, since the $d_i^{(j)}$'s are unknown ($d_i^{(j)} \geqq 2$ because of the 2/3 deferred decision rule). If no adjustment is made then the instruments will be perceived to be more reliable than what they actually are; as a result, the prior probabilities of their being in failed states will be underestimated. An opposite effect would take place if excessive adjustment is applied.

In the simulation work we found that our detection scheme is quite robust with respect to the amount of adjustment applied (referred to as 'adjustment for delays in detection' and abbreviated as ADJ in the tables of simulation results given in Part II), particularly in the latter part of the time horizon (after a modest number of failures have been observed), as long as the adjustment is not excessive. This is because the failures are generally rare and the test is relatively powerful, which means that the $t_i^{(j)}$'s are generally much bigger than the $d_i^{(j)}$'s. Therefore a small downward adjustment to the observed $(t_i^{(j)} + d_i^{(j)})$'s suffices. Writing (4.9) as

$$m_i^{(a)} = m_i^{(0)} + \sum_{j=1}^{a} t_i^{(j)} - a$$

$$= m_i^{(0)} + \sum_{j=1}^{a} \left(t_i^{(j)} + d_i^{(j)}\right) - \sum_{j=1}^{a} d_i^{(j)} - a \tag{5.5}$$

we see that even for modest values of a, $\Sigma_{j=1}^{a} d_i^{(j)}$ becomes negligible relative to $\Sigma_{j=1}^{a}(t_i^{(j)} + d_i^{(j)})$, and hence may be ignored without incurring much error.

The next modification of the basic scheme has to do with the updating of the instrument ages $\tau_i(t)$. In Section 4 we defined, for current calendar time t, $\tau_i(t)$ as the time since instrument i was last checked. This means that $\tau_i(t)$ should be set equal to zero for all the instruments that are checked at time t. But for instruments that are falsely detected to be in error, empirically we found that the performance (in particular, the power) of the test procedure improves if $\tau_i(t)$ is set equal to some small fixed number (referred to as 'age post-checking' and abbreviated as APC in the tables of simulation results given in Part II).

Another modification in updating the value of $\tau_i(t)$ in practice is required because $\tau_i(t)$ can increase indefinitely if no gross error is detected in the ith instrument. Once $\tau_i(t)$ becomes sufficiently large, the prior probability given by (4.6) also becomes large, which eventually leads to a detection (correct or false) of a gross error in that instrument. Empirically we found that letting $\tau_i(t)$ increase in this fashion results in far too many false detections. Therefore, on an ad-hoc basis, we used the average age of the instrument as an upper bound on $\tau_i(t)$. For the constant failure rate model, the average age is given by $1/\theta_i$, which is used to calculate the upper bound on $\tau_i(t)$. Since θ_i is unknown, we used its most current estimate given by (4.10). For the increasing failure rate model, as noted before, the Bayes test is still based on the constant θ_i assumption, and therefore the same method is used to find the average age.

Finally we discuss a computational difficulty associated with the implementation of the Bayes test. Each application of the Bayes test requires an evaluation of 2^n posterior probabilities. This is a rather large number of computations even for a relatively modest size process network. Furthermore, most of the posterior probabilities are negligibly small and some can even lead to underflow errors (e.g., the posterior probability that all measurements have gross errors). To avoid these complications we evaluated only a small fraction of the 2^n posterior probabilities that are likely to be close to the maximum.

This was done by first evaluating the posterior probabilities for all single gross error configurations $\eta_{(i)}$ $(i = 1, 2, \ldots, n)$. The top 25% of the

measurements with the highest such posterior probabilities were chosen as possible candidates for gross errors. Group posterior probabilities $\tilde{\pi}_I(t)$ were computed for only those groups I formed from this top quartile. Thus, e.g., for $n = 28$ the number of group posterior probabilities that need to be evaluated is drastically reduced from 2^{28} to $2^7 = 128$. Empirically we found that the inclusion of only this top quartile of measurements rarely gave a result different from the one obtained if all measurements were included. A further reduction in computational effort was obtained by assuming that at most three undetected gross errors can be present at any given time. In the foregoing example this assumption implies that out of a total of 128 possible group posteriors, only

$$\binom{7}{0} + \binom{7}{1} + \binom{7}{2} + \binom{7}{3} = 64$$

need be evaluated.

6 CONCLUDING REMARKS

In this paper we have given a theoretical development of a Bayesian model for occurrence of gross errors in process data, and a Bayesian test for their detection. As described in Section 5, numerous modifications need to be made in the basic scheme in order to implement it. Even with these modifications, the scheme is far from ready for application in practice because many of the practical realities are still not incorporated in the model, the present work being the first to attempt this modeling. It is probably worth mentioning some of the features that should be incorporated in the future developments of this model to bring it closer to reality and practical implementation.

(i) We have assumed that the data vectors are statistically independent. But, in fact, they may be serially correlated. This time-series feature should be considered in future modeling.

(ii) The magnitudes of the gross errors are assumed to be fixed in the present work; their estimates at any given time are regarded as their known values for computing the posterior probabilities. In fact, the magnitudes are likely to be random and time-varying. A unified Bayesian model should be developed to account for the randomness of the δ_i's.

(iii) Only gross errors due to measurement biases are considered in the present work. It is necessary to take into account gross errors arising from leaks and depositions, and misspecification of the model.

(iv) The covariance matrix Σ of the data vectors is assumed to be known and fixed. Both of these assumptions may not be true in practice. Even if Σ is assumed to be fixed, it is unclear how to estimate it in the presence of gross errors that cannot be always correctly detected.

(v) It is necessary to develop a Bayesian test for the increasing failure rate model. In the present work, although we have used such a model to generate gross errors, their detection is carried out using a Bayesian test based on the constant failure rate assumption.

APPENDIX

Lemma *The maximum row dimension of a full row rank matrix* $\mathbf{C}$ *satisfying* (3.2) *is* $n - p + q$.

Proof. The row space of $\mathbf{C}$ satisfying (3.2) is given by

$$\{\mathbf{c} \in \mathbb{R}^n : \mathbf{c}'\mathbf{D} = \mathbf{a}'\mathbf{B} \quad \text{for some } \mathbf{a} \in \mathbb{R}^q\}$$

$$= \left\{\mathbf{c} \in \mathbb{R}^n : (\mathbf{c}', \mathbf{a}')\begin{pmatrix}\mathbf{D} \\ -\mathbf{B}\end{pmatrix} = \mathbf{0} \right.$$

$$\left. \text{for some } \mathbf{a} \in \mathbb{R}^q \right\}$$

$$= \left\{\mathbf{c} \in \mathbb{R}^n : \begin{pmatrix}\mathbf{c} \\ \mathbf{a}\end{pmatrix} \in \text{Nullspace of} \begin{pmatrix}\mathbf{D} \\ -\mathbf{B}\end{pmatrix}\right\} \qquad \text{(A.1)}$$

Now $\binom{\mathbf{D}}{-\mathbf{B}}$ is an $(n+q) \times p$ matrix with full column rank. Therefore its nullspace has dimension $n - p + q$. Clearly, then, the dimension of the subspace (A.1) can be at most $n - p + q$. To see that it is in fact equal to $n - p + q$, choose a basis $(\mathbf{c}_1', \mathbf{a}_1'), (\mathbf{c}_2', \mathbf{a}_2'), \ldots, (\mathbf{c}_{n-p+q}', \mathbf{a}_{n-p+q}')$ for the nullspace of $\binom{\mathbf{D}}{-\mathbf{B}}$, and let $\mathbf{C} : (n-p+q) \times n$ and $\mathbf{A} : (n-p+q) \times q$ be matrices formed by taking the $\mathbf{c}_i$'s and $\mathbf{a}_i$'s as their rows, respectively. If the

dimension of the subspace (A.1) is less than $n-p+q$ then there exists a nonnull vector $\boldsymbol{v}:(n-p+q)\times 1$ such that $\boldsymbol{v}'\mathbf{C}=\mathbf{0}'$. Now $\boldsymbol{v}'\mathbf{AB}=\boldsymbol{v}'\mathbf{CD}=\mathbf{0}'$, and since $\mathbf{B}$ is full row rank $\boldsymbol{v}'\mathbf{A}=\mathbf{0}'$. Therefore

$$\boldsymbol{v}'(\mathbf{C},\mathbf{A})=\mathbf{0}'$$

But this is impossible since the rows of $(\mathbf{C},\mathbf{A})$ are linearly independent, being the basis vectors of the nullspace of $\binom{\mathbf{D}}{-\mathbf{B}}$. Hence the lemma follows.

DERIVATION OF THE FORMULA (3.6) FOR THE POSTERIOR PROBABILITY

To find a maximal dimensional transformation it is logical to start with the least squares residual vector $\boldsymbol{r}$ because it is known that the distribution of $\boldsymbol{r}$ is of rank $n-p+q$ and $E(\mathbf{r})$ depends on $\boldsymbol{\delta}\otimes\boldsymbol{\eta}$ but not on $\boldsymbol{x}$. More specifically, from Mah and Tamhane [8] we have

$$\boldsymbol{r}=(\mathbf{I}-\mathbf{DM})\boldsymbol{y} \tag{A.2}$$

where

$$\mathbf{M}=(\mathbf{I}-\mathbf{NB})(\mathbf{D}'\mathbf{Q}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{Q}^{-1} \tag{A.3}$$

and

$$\mathbf{N}=(\mathbf{D}'\mathbf{Q}^{-1}\mathbf{D})^{-1}\mathbf{B}'\left[\mathbf{B}(\mathbf{D}'\mathbf{Q}^{-1}\mathbf{D})^{-1}\mathbf{B}'\right]^{-1} \tag{A.4}$$

Thus

$$\begin{aligned}E(\boldsymbol{r})&=(\mathbf{I}-\mathbf{DM})E(\boldsymbol{y})\\&=(\mathbf{I}-\mathbf{DM})(\mathbf{D}\boldsymbol{x}+\boldsymbol{\delta}\otimes\boldsymbol{\eta})\\&=(\mathbf{D}-\mathbf{DMD})\boldsymbol{x}+(\mathbf{I}-\mathbf{DM})(\boldsymbol{\delta}\otimes\boldsymbol{\eta})\\&=\mathbf{DNB}\boldsymbol{x}+(\mathbf{I}-\mathbf{DM})(\boldsymbol{\delta}\otimes\boldsymbol{\eta})\\&\qquad\text{(using (A.3))}\\&=(\mathbf{I}-\mathbf{DM})(\boldsymbol{\delta}\otimes\boldsymbol{\eta})\ \text{(using (2.2))}\end{aligned} \tag{A.5}$$

which is free of $\boldsymbol{x}$, and

$$\operatorname{cov}(\boldsymbol{r})=\mathbf{V}=(\mathbf{I}-\mathbf{DM})\mathbf{Q}(\mathbf{I}-\mathbf{DM})' \tag{A.6}$$

which has rank $n-p+q$ [10]. Since $\boldsymbol{r}$ has a less than full rank distribution, i.e., it does not have a p.d.f., we need to further transform $\boldsymbol{r}$ to a $(n-p+q)$-dimensional random vector $\boldsymbol{s}$,

$$\boldsymbol{s}=\mathbf{H}\boldsymbol{r}=\mathbf{H}(\mathbf{I}-\mathbf{DM})\boldsymbol{y} \tag{A.7}$$

Such an $\mathbf{H}:(n-p+q)\times n$ is not unique, but a convenient choice is one that makes

$$\operatorname{cov}(\boldsymbol{s})=\mathbf{HVH}'=\mathbf{I} \tag{A.8}$$

By Theorem A3-4 of Seber [16] this $\mathbf{H}$ is given by

$$\mathbf{H}=\boldsymbol{\Psi}^{-1/2}\mathbf{G}' \tag{A.9}$$

Here $\boldsymbol{\Psi}^{-1/2}$ is an $(n-p+q)\times(n-p+q)$ diagonal matrix whose diagonal elements are $\psi_{ii}^{-1/2}$ $(i=1,2,\ldots,n-p+q)$ where the ψ_{ii}'s are the nonzero eigenvalues of $\mathbf{V}$, and the columns of $\mathbf{G}$ are the corresponding orthonormal eigenvectors of $\mathbf{V}$.

Now the p.d.f. of $\boldsymbol{s}$ is (apart from the normalizing constant $(2\pi)^{-(n-p+q)/2}$)

$$\begin{aligned}&\exp\{-\tfrac{1}{2}(\boldsymbol{s}-\mathbf{H}(\mathbf{I}-\mathbf{DM})(\boldsymbol{\delta}\otimes\boldsymbol{\eta}))'\\&\quad\times(\boldsymbol{s}-\mathbf{H}(\mathbf{I}-\mathbf{DM})(\boldsymbol{\delta}\otimes\boldsymbol{\eta}))\}\\&\qquad\text{(using (A.7) and (A.8))}\\&=\exp\{-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{\delta}\otimes\boldsymbol{\eta})'\\&\quad\times(\mathbf{I}-\mathbf{DM})'\mathbf{H}'\mathbf{H}(\mathbf{I}-\mathbf{DM})(\boldsymbol{y}-\boldsymbol{\delta}\otimes\boldsymbol{\eta})\}\end{aligned} \tag{A.10}$$

This expression can be simplified by making use of the following relations:

$$\begin{aligned}\mathbf{I}-\mathbf{DM}&=(\mathbf{I}-\mathbf{DM})(\mathbf{I}-\mathbf{DM})\\&\quad\text{(since }\mathbf{I}-\mathbf{DM}\text{ is idempotent)}\\&=(\mathbf{I}-\mathbf{DM})\mathbf{QQ}^{-1}(\mathbf{I}-\mathbf{DM})\\&=(\mathbf{I}-\mathbf{DM})\mathbf{Q}(\mathbf{I}-\mathbf{DM})'\mathbf{Q}^{-1}\\&\quad\text{(since }\mathbf{Q}^{-1}(\mathbf{I}-\mathbf{DM})\text{ is symmetric)}\\&=\mathbf{VQ}^{-1}\quad\text{(using (A.6))}\end{aligned} \tag{A.11}$$

and

$$\begin{aligned}\mathbf{H}'\mathbf{H}&=\mathbf{G}\boldsymbol{\Psi}^{-1}\mathbf{G}'\quad\text{(using (A.9))}\\&=\mathbf{V}^{-}\end{aligned} \tag{A.12}$$

where $\mathbf{V}^{-}$ denotes a generalized inverse of $\mathbf{V}$ satisfying

$$\mathbf{VV}^{-}\mathbf{V}=\mathbf{V} \tag{A.13}$$

We thus have

$$\begin{aligned}&(\mathbf{I}-\mathbf{DM})'\mathbf{H}'\mathbf{H}(\mathbf{I}-\mathbf{DM})\\&\quad=\mathbf{Q}^{-1}\mathbf{VV}^{-}\mathbf{VQ}^{-1}\quad\text{(using (A.11) and (A.12))}\\&\quad=\mathbf{Q}^{-1}\mathbf{VQ}^{-1}\quad\text{(using (A.13))}\\&\quad=\mathbf{Q}^{-1}(\mathbf{I}-\mathbf{DM})\quad\text{(using (A.11))}\\&\quad=\mathbf{W}\quad\text{(say)}\end{aligned} \tag{A.14}$$

Substituting this in (A.10) and the resulting expression for the p.d.f. in (3.1) we obtain the final formula for the posterior probability (3.6).

ACKNOWLEDGEMENT

This research was supported by NSF Grant CBT-8519182.

REFERENCES

1 A.C. Tamhane and R.S.H. Mah, Data reconciliation and gross error detection in chemical process networks, *Technometrics*, 27 (1985) 409–422.
2 G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1973.
3 J.O. Berger, *Statistical Decision Theory – Foundations, Concepts and Methods*, Springer Verlag, New York, 1980.
4 G.E.P. Box and G.C. Tiao, A Bayesian approach to some outlier problems, *Biometrika*, 55 (1968) 119–129.
5 E. de Alba and J. Van Ryzin, An empirical Bayes approach to outliers, *Journal of Statistical Planning and Inference*, 4 (1980) 217–236.
6 G. Kitagawa and H. Akaike, A quasi Bayesian approach to outlier detection, *Annals of the Institute of Statistical Mathematics*, 34 (1982) 389–398.
7 H.F. Martz and R.A. Waller, *Bayesian Reliability Analysis*, Wiley, New York, 1985.
8 R.S.H. Mah and A.C. Tamhane, Detection of gross errors in process data, *American Institute of Chemical Engineers Journal*, 28 (1982) 828–830.
9 C.M. Crowe, Y. Garcia Campos and A. Hrymak, Reconciliation of process flow rates by matrix projection, I. The linear case, *American Institute of Chemical Engineers Journal*, 29 (1983) 881–888.
10 C. Iordache, R.S.H. Mah and A.C. Tamhane, Performance studies of the measurement test for detection of gross errors in process data, *American Institute of Chemical Engineers Journal*, 31 (1985) 1187–1201.
11 N.R. Mann, R.E. Schafer and N.D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data*, Wiley, New York, 1974.
12 A.G. Colombo and D. Constantini, Ground-hypotheses for beta distribution as Bayesian prior, *Institute of Electrical and Electronic Engineers Transactions on Reliability*, R-29(1) (1980) 17–20.
13 G.H. Golub and C. Reinsch, Singular value decomposition and least-squares solutions, *Numerical Mathematics*, 14 (1970) 403–420.
14 J.A. Romagnoli, On data reconciliation: constraints processing and treatment of bias, *Chemical Engineering Science*, 38 (1983) 1107–1117.
15 A.A. Salvia and R.C. Bollinger, On discrete hazard functions, *Institute of Electrical and Electronic Engineers Transactions on Reliability*, R-31 (1982) 458–459.
16 G.A.F. Seber, *Linear Regression Analysis*, Wiley, New York, 1977.