Chemometrics and Intelligent Laboratory Systems, 4 (1988) 131–146
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

A Bayesian Approach to Gross Error Detection in Chemical Process Data

Part II *: Simulation Results

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 II: Simulation results. *Chemometrics and Intelligent Laboratory Systems*, 4: 131–146.

The performance of the gross error detection scheme based on the Bayesian test is evaluated using Monte Carlo simulation methods. Effects of selected control factors (implementation options) and noise factors (e.g., violation of assumptions and misspecification of priors) are studied. A comparison is made with the gross error detection scheme based on the non-Bayesian measurement test of Mah and Tamhane. The Bayesian scheme is found to be relatively robust. It performs better than the measurement test scheme when gross error occurrences are not infrequent. However, its performance characteristics converge rather slowly and hence accurate prior estimates of the various unknown parameters are necessary before the method can be put to practical use.

In conclusion, the Bayesian approach offers the promise of improving gross error detection and identification capabilities by using past failure data. Its technical feasibility is demonstrated by this investigation, but much remains to be done to make it a practical method.

7 SUMMARY OF PART II

In Part I we have given a theoretical development of the Bayesian scheme for detecting gross errors in process data. In the present Part II we study selected performance characteristics of this scheme. Analytical evaluation of these performance characteristics is difficult if not impossible because of the complex nature of the detection scheme. Therefore simulation methods must be employed. Using the same methods, performance characteristics of the gross error detection scheme based on the measurement test of Mah and

* Part I: *Chemometrics and Intelligent Laboratory Systems*, 4 (1988) 33–45.

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

0169-7439/88/$03.50

Tamhane [1] are also evaluated and compared with those of the detection scheme based on the Bayes test. Recall that the measurement test does not make use of the prior information while the Bayes test does. Therefore the objective of the aforementioned comparison is to assess the extent to which the prior information on instrument failures makes a difference in gross error detection and identification. We also study, via sensitivity analysis, the robustness of the Bayesian scheme to misspecification of the prior information. Both Bayesian scheme and the measurement test scheme are expected to be robust to non-normality of individual data vectors because both are based on the averages of N such vectors in each measurement period and for large N (we used $N = 30$ in the present simulations), by the central limit theorem, the averaged data vectors should be approximately normally distributed. For this reason we did not study robustness to non-normality in the present simulations.

The following is an outline of Part II. Section 8 gives a step-by-step description of the Bayesian gross error detection scheme as it is implemented in the simulation program. An analogous description of the measurement test scheme is also given. Details of the simulation study are given in Section 9. The measures used to assess and compare the performances of the Bayesian and the measurement test schemes are defined in Section 10. Simulation results are presented and discussed in Section 11. Conclusions following from these results and recommendations for the use of appropriate gross error detection scheme are given in Section 12. The same notation as in Part I is followed in this paper, and frequent references are made to the equations and discussion in Part I.

8 GROSS ERROR DETECTION SCHEMES BASED ON THE BAYESIAN AND MEASUREMENT TESTS

8.1 Gross error detection scheme based on the Bayesian test

In this section we outline the steps involved in the implementation of the Bayesian gross error detection scheme. Explanations for the modifications of the basic scheme are given in Section 5 of Part I.

Step 0. (i) Input the following information: For $i = 1, 2, \ldots, n$, initial estimates $\hat{\theta}_i^{(0)}$ and $\hat{\delta}_i^{(0)}$ of θ_i and δ_i, the number of confirmed detections $a_i(0) = 0$ and total detections $b_i(0) = 0$, and the value of s_i (cf. (4.11)). Compute the starting values of $l_i^{(0)}$ and $m_i^{(0)}$ from $\hat{\theta}_i^{(0)}$ and s_i using (4.12). In addition, input the following quantities: the design matrix $\mathbf{D}$, the constraint matrix $\mathbf{B}$, the error covariance matrix $\boldsymbol{\Sigma}$, and the number of data vectors observed per measuring period, N. The covariance matrix $\mathbf{Q}$ of the averaged data vector for each period is then $(1/N)\boldsymbol{\Sigma}$.

(ii) Initialize the ages $\tau_i(0)$ of all instruments and compute the prior probabilities $p_i(0)$ and $\pi_I(0)$ using (4.5) and (4.6), respectively. Also initialize sets $I_1(0)$ and $I_2(0)$ of flagged * and unflagged instruments, respectively **.

Step t ($t = 1, 2, \ldots$). (i) The following quantities are available from the previous step, i.e., the $(t-1)$th step (how these quantities are computed for period t is explained in subsequent steps of the algorithm):

$I_1(t-1)$ = Set of flagged instruments,
$I_2(t-1)$ = Set of unflagged instruments,
$\tau_i(t-1)$ = Age of instrument i,
$p_i(t-1)$ = Individual prior probabilities ***,
$\pi_I(t-1)$ = Group prior probabilities,
$a_i(t-1)$ = No. of confirmed occurrences of gross errors in instrument i,
$b_i(t-1)$ = No. of detections (confirmed or not) of gross errors in instrument i,

* A flagged instrument is one in which a gross error has been detected but not confirmed because it has not been inspected since the last detection. An unflagged instrument is one in which no gross error has been detected since the last inspection.

** In our simulation program we used $\tau_i(0) = 0$ for all i which makes $p_i(0) = 0$ for all i and therefore $\pi_I(0) = 0$ for all nonempty groups I. Hence no gross error detection test needs to be applied at $t = 1$. We also set $I_1(0) = \emptyset$, the empty set, and $I_2(0) = \{1, 2, \ldots, n\}$, the whole set.

*** From (8.8) we see that $p_i(t-1) = 0\ \forall i \in I_1(t-1)$ so that no gross errors will be detected in flagged instruments.

$l_i^{(a)}$, $m_i^{(a)}$, $\hat{\theta}_i^{(a)}$ = Parameters of the beta distribution associated with instrument i after $a = a_i(t-1)$ confirmed gross error occurrences,

$\hat{\delta}_i^{(b)}$ = Estimate of δ_i after $b = b_i(t-1)$ detections (confirmed or not).

(ii) Record N data vectors for period t and compute their average $y(t)$.

(iii) Compute the posterior probabilities $\tilde{\pi}_{\{i\}}(t)$ for configurations in which a gross error is present in only one of the instruments $i \in I_2(t-1)$. For calculating $\tilde{\pi}_{\{i\}}(t)$, use formula (4.3) with the corresponding $\boldsymbol{\delta}_{\{i\}}$ vector having the following components:

jth component of $\boldsymbol{\delta}_{\{i\}}$

$$= \begin{cases} \hat{\delta}_j^{(b)} & \text{for } j = i \text{ or } j \in I_1(t-1) \\ 0 & \text{for } j \neq i,\ j \in I_2(t-1) \end{cases} \tag{8.1}$$

where $b = b_j(t-1)$.

(iv) Choose the top quartile of the instruments in $I_2(t-1)$ based on their $\tilde{\pi}_{\{i\}}(t)$ values. Call this subset of $I_2(t-1)$ as $I(t)$.

(v) Find set $I^* = I^*(t) \subseteq I(t)$ using decision rule (3.7). In the search for the maximizing subset $I^*(t)$, restrict to subsets of $I(t)$ of cardinality $\leqq 3$.

(vi) If an instrument is included in set I^* in 2 out of 3 consecutive time periods, then include it in the subset of instruments in which gross errors have been detected but not confirmed yet. Symbolically,

$$I_1(t) = I_1(t-1) \cup I^{**}(t) \tag{8.2}$$

where

$$I^{**}(t) = I^*(t) \cap [I^*(t-1) \cup I^*(t-2)] \tag{8.3}$$

Let

$$I_2(t) = \{1, 2, \ldots, n\} - I_1(t) \tag{8.4}$$

(vii) Set

$$b_i(t) = \begin{cases} b_i(t-1) + 1 & \forall i \in I^{**}(t) \\ b_i(t-1) & \forall i \notin I^{**}(t) \end{cases} \tag{8.5}$$

(viii) Estimate δ_i for each $i \in I^{**}(t)$ using the procedure given in Section 5.1. Let $\hat{\delta}_i(t)$ denote this estimated value. (Note that this procedure yields estimates $\hat{\delta}_i(t)$ for all $i \in I_1(t)$, but we only use the estimates obtained for the newly detected gross errors, i.e., for $i \in I^{**}(t)$.) Compute weighted estimates for $i \in I^{**}(t)$:

$$\hat{\delta}_i^{(b)} = \frac{1}{b+1}\left\{\hat{\delta}_i(t) + b\hat{\delta}_i^{(b-1)}\right\} \tag{8.6}$$

where $b = b_i(t)$. Note that the estimates $\hat{\delta}_i^{(b)}$ remain unchanged for $i \notin I^{**}(t)$.

(ix) If t is not a scheduled inspection time, then let $a_i(t) = a_i(t-1)$ for all i,

$$\tau_i(t) = \min\left(\tau_i(t-1) + 1,\ 1/\hat{\theta}_i^{(a)}\right) \text{ for } i \in I_2(t) \tag{8.7}$$

where $\hat{\theta}_i^{(a)} = l_i^{(a)}/(l_i^{(a)} + m_i^{(a)})$ and $a = a_i(t)$. Let

$$p_i(t) = 0 \quad \forall i \in I_1(t) \tag{8.8}$$

and compute $p_i(t)$ for $i \in I_2(t)$ using formula (4.7). Let $t \to t+1$ and return to the beginning of step t.

(x) If t is a scheduled inspection time, then check all instruments $i \in I_1(t)$. Take appropriate corrective actions on instruments that are found to contain gross errors. For these instruments set $\tau_i(t) = 0$, $a_i(t) = a_i(t-1) + 1$, and update their l_i and m_i values using the formula (cf. (4.9)):

$$\begin{aligned} l_i^{(a)} &= l_i^{(a-1)} + 1, \\ m_i^{(a)} &= m_i^{(a-1)} + \left(t_i^{(a)} + d_i^{(a)}\right) - 1 \end{aligned} \tag{8.9}$$

Here $a = a_i(t)$, $t_i^{(a)}$ is the actual age of the ith instrument at which its presently corrected gross error occurred and $d_i^{(a)}$ is the delay in detecting that gross error.

(xi) For all other instruments set $a_i(t) = a_i(t-1)$. For instruments $i \in I_1(t)$ that are checked but found to be not in error, set $\tau_i(t) = \text{APC}$ (age post-checking), which is a small fixed number provided as input to the program.

(xii) Reset $I_1(t) = \phi$ and $I_2(t) = \{1, 2, \ldots, n\}$. Let $t \to t+1$ and return to the beginning of step t.

8.2 Gross error detection scheme based on the measurement test

Since the measurement test does not make use of any prior information, there is no need in this

scheme to keep track of such quantities as instrument ages and estimates of the θ_i's and δ_i's. Other aspects of the scheme, e.g., the deferred decision rule (2/3 rule), checking of the instruments only at scheduled inspection times, restriction to at most 3 undetected gross errors at any time, etc. are the same as before. The following are the steps in this scheme.

Step 0. Input the quantities **D**, **B**, $\boldsymbol{\Sigma}$, N and $\mathbf{Q} = (1/N)\boldsymbol{\Sigma}$. Also fix a level of significance α. Let $\beta = \frac{1}{2}\{1 - (1-\alpha)^{1/n}\}$ and let z_β be the upper β point of the standard normal distribution. Let $I_1(0) = \phi$ and $I_2(0) = \{1, 2, \ldots, n\}$.

Step t ($t = 1, 2, \ldots$). (i) Let $I_1(t-1)$ and $I_2(t-1)$ be as defined for the Bayesian scheme.

(ii) Record N data vectors for the time period t, and calculate the averaged data vector $\boldsymbol{y}(t)$.

(iii) Compute the vector of transformed residuals

$$\boldsymbol{d}(t) = \mathbf{Q}^{-1}\boldsymbol{r}(t) = \mathbf{Q}^{-1}(\mathbf{I} - \mathbf{DM})\boldsymbol{y}(t) \qquad (8.10)$$

where **M** is defined in (A.3). Next calculate the standardized residuals *

$$z_i(t) = \frac{d_i(t)}{\sqrt{w_{ii}}} \quad \text{for } i \in I_2(t-1) \qquad (8.11)$$

where w_{ii} is the ith diagonal element of matrix **W** given by (A.14), which can be shown to be equal to the covariance matrix of $\boldsymbol{d}(t)$.

(iv) Pick the three largest (in absolute value) $z_i(t)$-statistics, $i \in I_2(t-1)$. If any of these three statistics satisfy the condition

$$|z_i(t)| > z_\beta \qquad (8.12)$$

then a gross error is suspected in the corresponding instrument and it is included in set $I^*(t)$.

(v) Let $I^{**}(t)$ be as defined in (8.3) and let $I_1(t)$ be as defined in (8.2).

(vi) If t is not a scheduled inspection time, then let $t \to t+1$ and return to the beginning of step t.

(vii) If t is a scheduled inspection time, then check all instruments $i \in I_1(t)$. Take appropriate corrective actions on instruments that are found to contain gross errors. Reset $I_1(t) = \phi$ and $I_2(t) = \{1, 2, \ldots, n\}$. Let $t \mapsto t+1$ and return to the beginning of step t.

We remark that in step (iv) above an alternative test for multiple gross errors described in Cook and Weisberg (ref. 3, pp. 28–31) could have been employed. However, that test involves many more computations (computations of statistics for all subsets of cardinality ≤ 3), and it is not clear that it is significantly more powerful.

9 DETAILS OF THE SIMULATION STUDY

In this section we give details of the simulation study conducted to assess the performance of the gross error detection scheme based on the Bayesian test relative to that based on the measurement test. The simulation program follows closely the algorithms for the two schemes described in the

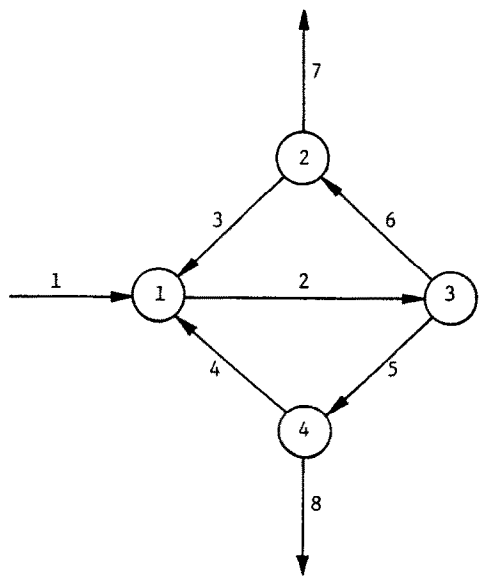


Fig. 1. Network 1.

* As noted in ref. 1, the test statistics (8.11) for detecting gross errors are based on transformed residuals because they have certain optimal power properties shown in ref. 2. If **Q** is diagonal, then there is no need to use transformed residuals. In that case, instead of (8.11), we simply use

$$z_i(t) = \frac{r_i(t)}{\sqrt{v_{ii}}} \quad \text{for } i \in I_2(t-1)$$

where v_{ii} is the ith diagonal entry of matrix **V** defined in (A.6).

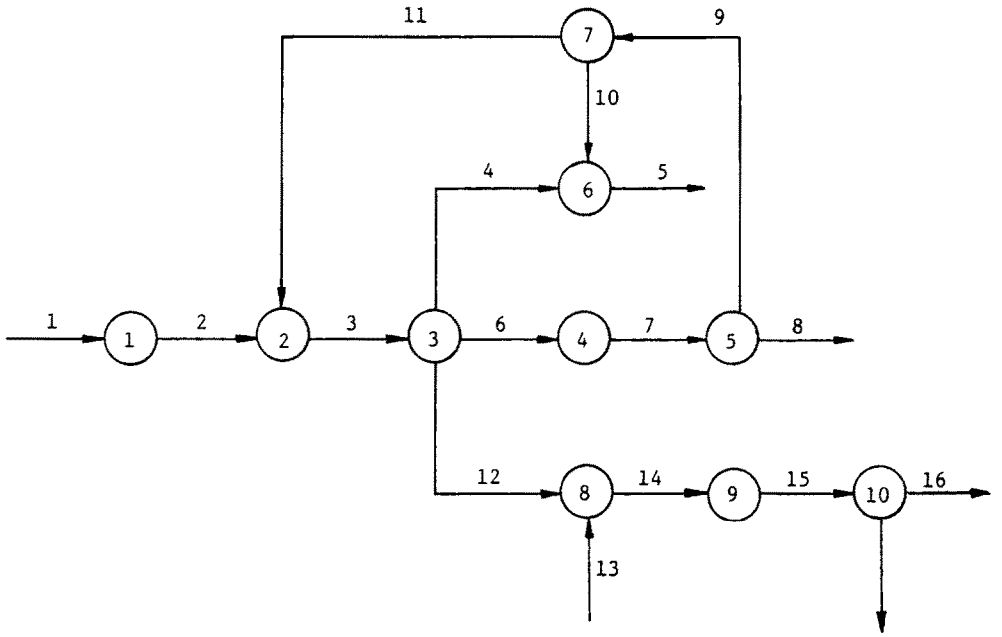


Fig. 2. Network 2.

previous section. Selected performance measures (defined in Section 10) are evaluated periodically in the program.

Simulation runs were performed for five different networks (shown in Figs. 1–5), which differ in size and structure. For the first four networks the

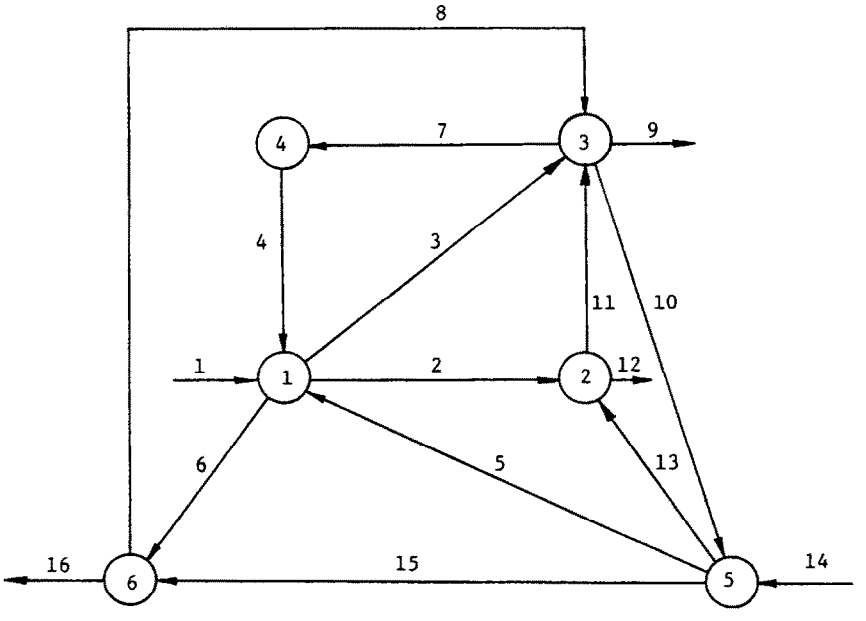


Fig. 3. Network 3.

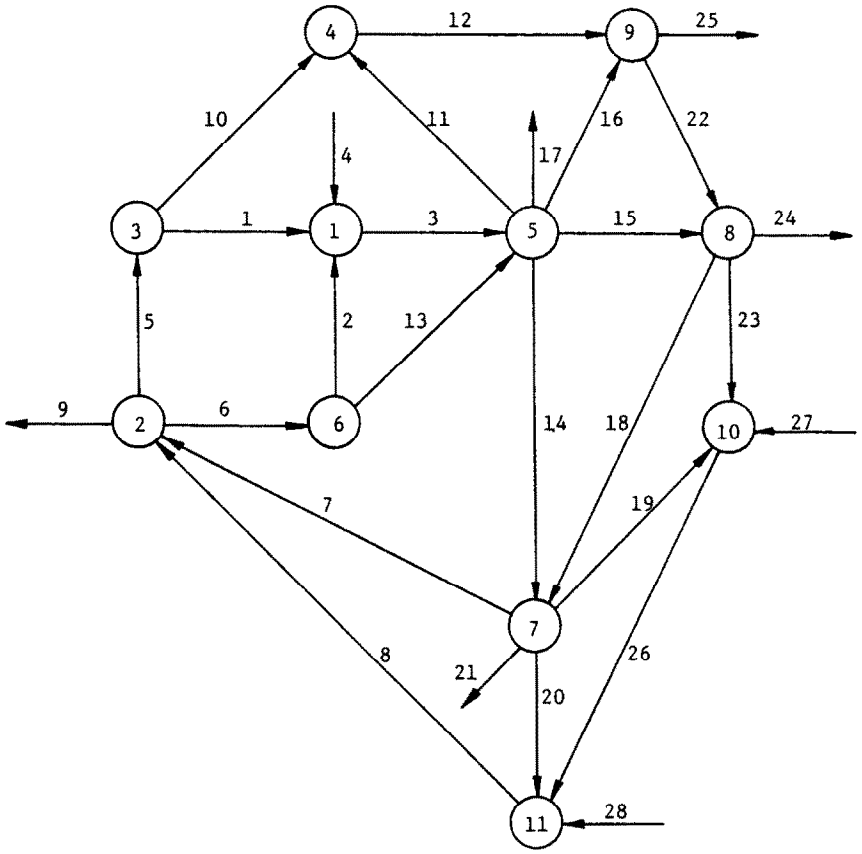


Fig. 4. Network 4 (steam system for a methanol synthesis unit, by Serth and Heenan [3]).

constraints are total mass flow balance constraints, so **B** is the incidence matrix of the corresponding digraph in each case. Network 1 shown in Fig. 1 with 4 nodes and 8 streams is the smallest in size. Networks 2 and 3 shown in Figs. 2 and 3, respectively, have the same number of streams but different structures (different numbers of nodes). Network 4 in Fig. 4 is the largest (11 nodes and 28 streams). It represents the steam-metering system that was used by Serth and Heenan [4] to test their gross error detection algorithm. Networks 1–3 do not have a direct physical interpretation. Network 5 shown in Fig. 5 is for the ammonia synthesis loop studied by Crowe et al. [5]. The **B** matrix for this network is not simply the incidence matrix. It is the matrix denoted by **C** in Tamhane and Mah (ref. 6, p. 419) with ζ, the unknown fraction of N_2, H_2 and Ar purged in the splitter, estimated to be 0.02. This matrix is given in Table 7. For details of calculation of this matrix, see ref. 6.

For each network we performed a number of runs under different conditions, e.g., different prior estimates of the θ_i's and δ_i's, different implementation options (e.g., the 3/5 deferred decision rule versus the 2/3 rule mentioned in the descriptions of the schemes, different schedules of inspection times), different values of the true δ_i's and **Q**, etc. Each simulation run consists of N_R replications of the realization of the process over N_T time periods. All the operations carried out during each time period including generation of gross errors, generation of the averaged data vector, application of the Bayes test for detection of gross errors, etc. are together referred to as a simulation. Thus each simulation run consists of N_R times N_T simu-

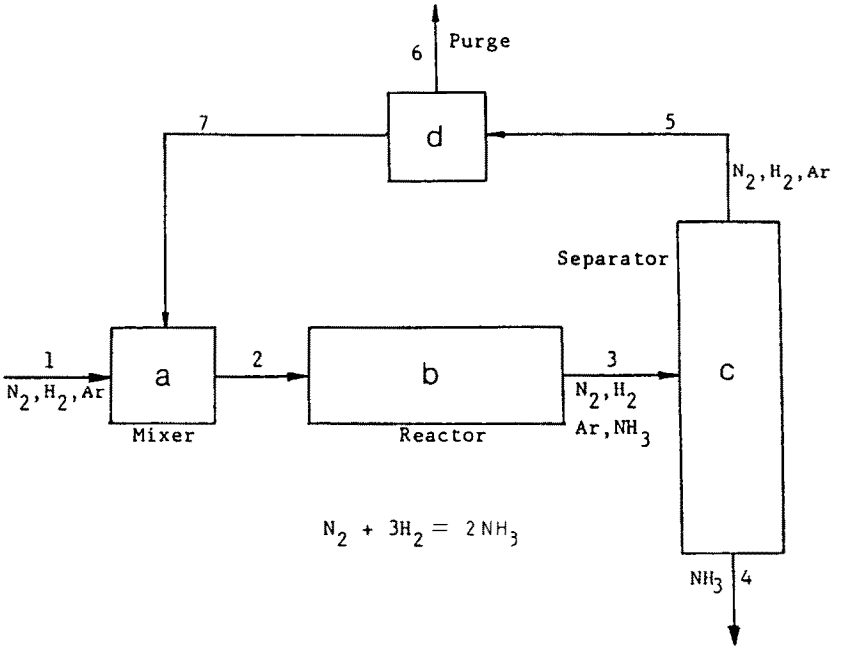
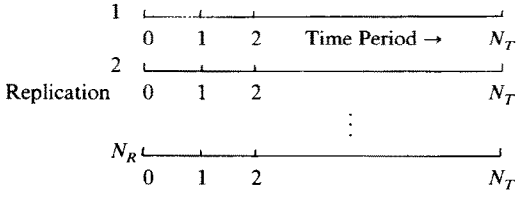


Fig. 5. Network 5 (ammonia synthesis loop). Measured flow rates: $y = (N_2^{(1)}, H_2^{(1)}, Ar^{(1)}, N_2^{(2)}, Ar^{(2)}, N_2^{(3)}, NH_3^{(4)}, H_2^{(5)})$.

lations. A schematic representation of a simulation run is shown below.



We assume that N measurements of the data vector are made in each time period. Notice that only the average of these N individual measurements needs to be generated for each period. For time period t, this average, $y(t)$, has an n-variate normal distribution with mean vector $\mathbf{D}x + \delta \otimes \eta(t-1)$ and covariance matrix $\mathbf{Q} = (1/N)\Sigma$. Without loss of generality x is taken to be a null vector. The vector $\eta(t-1)$ is obtained by generating Bernoulli random variables $\eta_i(t-1)$ for all instruments i in which gross errors are not already present at time $t-1$. Here $\eta_i(t-1) = 1$ with probability $\theta_i(T_i(t-1))$ and $= 0$ with probability $1 - \theta_i(T_i(t-1))$ where $\theta_i(T_i(t-1))$ is obtained from (5.6) and $T_i(t-1)$ is the actual age of instrument i at time $t-1$ ($i = 1, 2, \ldots, n$, $t = 1, 2, \ldots$). The values of $\theta_i(1)$ and β_i used in (5.6) are inputs to the simulation program; the program keeps track of the actual ages of all the instruments. Recall that $\beta_i = 0$ corresponds to the constant failure rate model, $\theta_i(T_i) = \theta_i$ for all $T_i \geq 1$ ($i = 1, 2, \ldots, n$).

As noted before, the output from each simulation run consists of averaged values of the various performance measures. Certain simulation runs were carried out especially for studying the convergence behavior of the Bayesian scheme. For these runs, the performance measures were evaluated on a cumulative basis at selected measurement periods N_M in each realization. The reported values are averages of these performance measures over N_R realizations. Although the program computed performance measures for each stream separately, to save space we only report, for networks 1–4, their averages over all the streams; only for network 5 have we reported performance measures for each stream separately in Table 7. The reasons for doing this will become clear in the discussion of Table 7 given in Section

11. In addition to the performance measures, the program also reports the final estimates of the θ_i's and the δ_i's (averaged over N_R realizations).

All runs were made using $N_R = 50$ and $N = 30$. For most runs we used $N_T = 10080$ except those runs that were conducted to study the convergence properties of the Bayesian scheme, which used $N_T = 50400$. For these latter runs, performance measures were computed periodically for $N_M = 1008$, 5040, 10080, 30240 and 50400. Assuming that each measurement period is 10 minutes long (i.e., the data vectors are recorded once every 20 seconds), we see that 48 periods correspond to one shift of 8 hours, 144 periods correspond to one day, 1008 periods correspond to one week, 10080 periods correspond to ten weeks, and 50400 periods correspond to 50 weeks or nearly a year.

Simulations were performed on Northwestern University's CDC Cyber 180/845 computer and on the University of Illinois's Cray X-MP/48 supercomputer. The supercomputer was used to perform runs with $N_T = 50400$ (runs for studying the convergence properties of the Bayesian scheme) and runs for network 4, which is the largest network studied; these runs would not have been feasible without access to a supercomputer. Fortran 77 was the language used in both the programs. The multiplicative congruential pseudo-random number generator of Downham and Roberts [7] was used to generate uniform [0, 1] random variates on the Cyber while the library function RANF was used on the Cray. The uniform random variates were transformed into standard normal variates using the polar method (see ref. 8, p. 105). The covariance matrix $\mathbf{Q} = (1/N)\boldsymbol{\Sigma}$ was assumed to be diagonal (with diagonal entries * equal to σ_i^2/N) in all of the simulation runs. Thus independently distributed measurements $y_i(t)$ for $i = 1, 2, \ldots, n$ were obtained by scaling the standard normal variates by $\sigma_i/\sqrt{N}$ and adding the δ_i's to those measurements in which gross errors have been generated but not removed yet (recall that we have taken $\boldsymbol{x} = \mathbf{0}$ without loss of generality). Generation of gross errors at time $t-1$ for those measurements in which gross errors were not already present was determined by simulation of Bernoulli random variates $\eta_i(t-1)$. The latter were generated from uniform [0, 1] variates by the usual method.

* Note that here we have denoted the diagonal entries of $\boldsymbol{\Sigma}$ by the σ_i^2 instead of the σ_{ii}, which is thus an exception to the notational convention followed elsewhere in this paper.

10 PERFORMANCE MEASURES

For a given replication of a simulation run consisting of N_T time periods we used the following quantities as measures of performance of a gross error detection scheme. Note that these quantities are sample estimates of the corresponding population parameters. Further note that the values reported in the Results and Discussion section of this paper for these performance measures are averages over $N_R = 50$ replications. Also, as noted before, some of the performance measures are evaluated periodically during each replication rather than only at the end of each replication.

(i) Overall probability of a type I error

This is estimated by

$$\text{PTI} = 1 - \frac{N_{00}}{N_0} \tag{10.1}$$

where N_0 is the number of simulations in which no gross error is present in any of the instruments, and N_{00} is the number of those N_0 simulations where no type I error is made.

(ii) Expected number of type I errors

This is estimated by

$$\text{ETI} = \frac{\text{Total number of type I errors}}{N_T} \tag{10.2}$$

Note that in Table 7 for network 5 this ETI is referred to as the overall ETI which is the sum of the ETI values for all the streams.

(iii) Overall expected delay in detection

Let $d_i^{(j)}$ be the delay in detecting the jth generated gross error in the ith instrument. Then

the expected delay for that instrument is estimated by

$$\mathrm{EDL}_i = \frac{\sum_{j=1}^{N_{gi}} d_i^{(j)}}{N_{gi}} \tag{10.3}$$

where N_{gi} = number of gross errors generated in instrument i ($i = 1, 2, \ldots, n$). The overall expected delay is estimated by

$$\mathrm{EDL} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{N_{gi}} d_i^{(j)}}{\sum_{i=1}^{n} N_{gi}} \tag{10.4}$$

(iv) Estimates of the θ_i and δ_i

For the Bayesian scheme it is of interest to know how large a and b (the number of confirmed gross errors and the number of detected gross errors in instrument i) must be so that the estimates $\hat{\theta}_i^{(a)}$ and $\hat{\delta}_i^{(b)}$ will converge to the corresponding true values θ_i (assuming constant failure rate) and δ_i. For the increasing failure rate model (5.4), $\hat{\theta}_i^{(a)}$ should converge to

$$\bar{\theta}_i = \{\text{expected lifetime of instrument } i\}^{-1} = \left\{ \sum_{t=1}^{\infty} t\theta_i(t) \right\}^{-1} \quad (i = 1, 2, \ldots, n) \tag{10.5}$$

Of course, the rapidity with which these estimates converge to the true values will depend on, among other things, how close the initial estimates $\hat{\theta}_i^{(0)}$ and $\hat{\delta}_i^{(0)}$ are to the true values. We have reported the estimates $\hat{\theta}_i$ and $\hat{\delta}_i$ (averaged over $N_R = 50$ replications) at the end of N_T time periods, and in some cases at selected intermediate time periods.

In addition to the above performance measures we also evaluated the overall powers of the Bayesian and the measurement test schemes for detecting gross errors. Since any generated gross error is eventually detected and since the 2/3 deferred decision rule was adopted in our work, we used the proportion of all gross errors that were detected within 3 time periods of their generation as our estimate of the overall power. The corresponding proportion for each instrument was used to estimate the instrument specific power. We have not reported these power estimates here because essentially the same information concerning the sensitivity of the scheme in detecting gross errors is more directly obtained from EDL.

11 RESULTS AND DISCUSSION

The simulation results are presented in Tables 1–7. The first five tables are for the Bayesian scheme applied to network 1; the results are given for different runs corresponding to different implementation options and other conditions. The objective of these runs was to make a comprehensive study of the performance of the Bayesian scheme for the simplest network. Table 6 gives comparative results for the Bayesian and the measurement test schemes for the first four networks. Finally Table 7 gives similar comparative results for the ammonia synthesis loop (network 5).

Each simulation run for the Bayesian scheme is labelled by B and that for the measurement test scheme by M. Each label is followed by a number indicating the network, followed by another number indicating the particular run.

All of the runs used $\mathbf{D} = \mathbf{I}$ and diagonal $\mathbf{Q} = (1/N)\boldsymbol{\Sigma}$ with diagonal entries $\sigma_i^2/N = \mathrm{var}(y_i(t))$, $i = 1, \ldots, n$, $t = 1, 2, \ldots$. All of the runs were performed for equal values of the parameters σ_i^2, θ_i (or $\theta_i(1)$ and β_i for the increasing failure rate model (5.4)), δ_i, s_i, and equal values of the initial estimates $\hat{\theta}_i^{(0)}$ and $\hat{\delta}_i^{(0)}$. Therefore we have dropped subscript i from these quantities for notational convenience. Also note that the final estimates $\hat{\theta}_i$ and $\hat{\delta}_i$ (averaged over N_R realizations) are further averaged over $i = 1, 2, \ldots, n$ (except for network 5), and reported as $\hat{\theta}$ and $\hat{\delta}$.

Table 1 studies the performance of the Bayesian scheme as a function of the following implementation options (which are under a process engineer's control): deferred decision rule, deferred checking of flagged instruments, the value of APC (age post-checking), and the value of ADJ (adjustment for delays in detection). The following observations may be made on this table.

TABLE 1

Performance of the Bayesian scheme for different implementation options ($N_T = 10080$)

$\theta(1) = 0.0001$, $\beta = 7.7 \times 10^{-5}$, $\bar{\theta} = 0.007$, $\hat{\theta}^{(0)} = 0.007$, $\sigma = 0.2857$, $s = 1.2$, $\delta = 1.0$, $\hat{\delta}^{(0)} = 0.85$.

Run	Deferred decision rule	Inspection frequency *	APC	ADJ	Simulation results				
					$\hat{\theta}$	$\hat{\delta}$	PTI	ETI	EDL
B1.1	1/1	Immed.	7	3	0.0071	0.9970	0.1974	0.2285	1.624
B1.2	2/3	Immed.	7	3	0.0071	0.9250	0.0662	0.0740	2.985
B1.3	3/5	Immed.	7	5	0.0069	0.9700	0.0332	0.0385	4.468
B1.4	2/3	Shift	7	5	0.0060	0.8700	0.0502	0.0580	4.932
B1.5	2/3	Day	7	5	0.0047	0.8600	0.0321	0.0232	11.592
B1.6	2/3	Shift	0	5	0.0061	0.8828	0.0458	0.0553	5.394
B1.7	2/3	Shift	**	5	0.0062	0.7113	0.1374	0.0956	3.567
B1.8	2/3	Shift	7	3	0.0061	0.8650	0.0496	0.0576	4.933
B1.9	2/3	Shift	7	10	0.0063	0.8607	0.0518	0.0588	4.872

* Shift = 48 periods, Day = 144 periods.

** APC is set equal to the actual age.

(i) From runs B1.1–B1.3 we see that a deferred decision rule reduces the incidence of type I errors but increases the detection delay. The 2/3 rule was thought to give a reasonable trade-off between these two performance measures, and hence was used in all of the later work. There is no significant effect of different deferred decision rules on the final estimates $\hat{\theta}$ and $\hat{\delta}$.

(ii) Runs B1.2, B1.4 and B1.5 differ only in the frequency of inspection of flagged instruments. We see that long postponements in inspection reduce PTI and ETI but only at the expense of a substantial increase in EDL. Also delays in inspection have deleterious effects on $\hat{\theta}$ in particular, and also on $\hat{\delta}$, because these estimates do not converge to the true values θ and δ; in other words, there is an asymptotic bias. For example, for the 2/3 deferred decision rule with immediate inspection, $\hat{\theta} = 0.0071$ and $\hat{\delta} = 0.9250$, which are quite close to the true values, while with daily inspection, $\hat{\theta} = 0.0047$ and $\hat{\delta} = 0.8600$, which are well below the true values. Considering that immediate (or very frequent) checking is generally infeasible and type II errors (longer EDL's) are generally costlier than type I errors, we adopted the once a shift (once every 48 periods) inspection schedule in subsequent simulation runs.

(iii) The effect of APC is studied in runs B1.4, B1.6 and B1.7. In run B1.6 the age of a wrongly flagged instrument is set equal to zero (as per theory) after it is inspected while in run B1.7 it is set equal to its actual age; run B1.4 is for an intermediate value of 7, which is approximately 5% of the expected lifetime ($1/\bar{\theta}$). Higher APC is associated with higher PTI and ETI but lower EDL. The intermediate value of roughly 5% to 10% of the expected lifetime is used in most simulation runs.

(iv) Runs B1.4, B1.8 and B1.9 use ADJ = 5, 3 and 10, respectively, with everything else the same. We see that the simulation results for the three runs are not significantly different from each other. In subsequent runs ADJ = 5 was adopted.

Table 2 studies the robustness of the Bayesian scheme to the violations of certain assumptions made and to the misspecification of the prior information (these violations and misspecifications being not under a process engineer's control). Robustness to non-normality was not studied for the reason mentioned in Section 7. We note the following conclusions.

(i) Run B1.4 in Table 1 and run B1.10 in Table 2 differ only in terms of the failure model used to generate gross errors; the former uses the increasing failure rate model (5.4) while the latter uses the constant failure rate model with $\theta =$ the average $\bar{\theta}$ for the former. Surprisingly we notice that the results for the two runs are very similar on all

TABLE 2

Sensitivity analysis of the Bayesian scheme ($N_T = 10080$)

$\theta(1) = 0.0001$, $\beta = 7.7 \times 10^{-5}$, $\bar{\theta} = 0.007$, $\delta = 1.0$, $\sigma = 0.2857$, APC = 7, ADJ = 5. Inspection frequency: once a shift. Decision rule: 2/3.

Run	$\hat{\theta}^{(0)}$	s	$\hat{\delta}^{(0)}$	Simulation results				
				$\hat{\theta}$	$\hat{\delta}$	PTI	ETI	EDL
B1.10 *	0.007	1.20	0.85	0.0062	0.8797	0.0525	0.0576	5.805
B1.11	0.0023	1.01	0.85	0.0030	0.9973	0.0139	0.0350	6.625
B1.12	0.0023	1.20	0.85	0.0054	0.9127	0.0337	0.0481	5.619
B1.13	0.0023	1.80	0.85	0.0059	0.8844	0.0428	0.0539	5.134
B1.14	0.021	1.01	0.85	0.0110	0.7584	0.1052	0.0865	3.849
B1.15	0.021	1.20	0.85	0.0065	0.8419	0.0605	0.0630	4.736
B1.16	0.021	1.80	0.85	0.0062	0.8618	0.0511	0.0585	4.892
B1.17	0.007	1.20	0.50	0.0062	0.8585	0.0514	0.0585	4.893
B1.18	0.007	1.20	2.00	0.0061	0.8915	0.0443	0.0550	5.180

* Run B1.10 used $\theta(1) = 0.007$ and $\beta = 0$, i.e., constant failure rate with $\theta = \bar{\theta} = 0.007$.

performance measures recept EDL, which is actually larger for the latter run. One would expect that under the constant failure rate model the Bayesian scheme should perform better because it is based on that assumption, but this is not the case at least in this particular run.

TABLE 3

Convergence study of the Bayesian scheme relative to prior estimate of $\bar{\theta}$ ($N_T = 50400$)

$\theta(1) = 0.0001$, $\beta = 7.7 \times 10^{-5}$, $\bar{\theta} = 0.007$, $\delta = 1.0$, $\hat{\delta}^{(0)} = 0.85$, $\sigma = 0.2857$, APC = 7, ADJ = 5. Inspection frequency: once a shift. Decision rule: 2/3.

Run	$\hat{\theta}^{(0)}$	s	N_M	Simulation results					
				a	$\hat{\theta}$	$\hat{\delta}$	PTI	ETI	EDL
B1.19	0.0023	1.20	1008	6	0.0034	0.941	0.0244	0.0410	6.563
			5040	29	0.0049	0.929	0.0296	0.0454	5.800
			10080	59	0.0054	0.913	0.0337	0.0481	5.619
			30240	177	0.0058	0.895	0.0399	0.0519	5.363
			50400	296	0.0059	0.889	0.0420	0.0532	5.269
B1.20	0.0023	1.80	1008	6	0.0050	0.900	0.0323	0.0477	5.679
			5040	29	0.0058	0.892	0.0402	0.0525	5.189
			10080	59	0.0059	0.884	0.0428	0.0539	5.134
			30240	178	0.0060	0.879	0.0455	0.0550	5.112
			50400	296	0.0060	0.877	0.0463	0.0558	5.092
B1.21	0.021	1.20	1008	6	0.0097	0.809	0.0725	0.0714	4.387
			5040	30	0.0069	0.831	0.0662	0.0660	4.599
			10080	59	0.0065	0.842	0.0605	0.0630	4.736
			30240	178	0.0062	0.858	0.0539	0.0597	4.910
			50400	297	0.0061	0.863	0.0523	0.0588	4.949
B1.22	0.021	1.80	1008	6	0.0072	0.851	0.0505	0.0604	4.816
			5040	30	0.0063	0.859	0.0526	0.0585	4.820
			10080	59	0.0062	0.862	0.0511	0.0578	4.892
			30240	178	0.0061	0.868	0.0495	0.0575	4.974
			50400	297	0.0060	0.870	0.0492	0.0573	5.008

(ii) The prior information regarding θ (or $\bar{\theta}$ in the case of increasing failure rates) as expressed via $\hat{\theta}^{(0)}$ has a significant influence on the performance of the Bayesian scheme. This can be seen from runs B1.11–B1.16. In the first three of these runs $\hat{\theta}^{(0)} = \bar{\theta}/3$ while in the latter three runs $\hat{\theta}^{(0)} = 3\bar{\theta}$. We see that initial underestimation of $\bar{\theta}$ results in a lower incidence of type I errors but longer detection delays, and vice versa. Note also that these effects are moderated if less weight is placed on the prior information and more weight on the data by specifying a larger value of s. Therefore if there is uncertainty about the accuracy of the initial estimate $\hat{\theta}^{(0)}$, then a larger value of s (≥ 1.5) is recommended.

(v) Inaccuracy in the initial estimate $\hat{\delta}^{(0)}$ of δ has a less significant effect. This is seen by comparing the results for runs B1.17 and B1.18 with those for run B1.4, the corresponding values of $\hat{\delta}^{(0)}$ being 0.5δ, 2δ and 0.85δ, respectively. The results are very similar in all three cases with higher $\hat{\delta}^{(0)}/\delta$ values yielding smaller PTI and larger EDL values.

Table 3 studies the convergence behavior of the Bayesian scheme over a long time horizon ($N_T = 50400$) as a function of $\hat{\theta}^{(0)}$.

(i) First we note that the final estimate $\hat{\theta}$ converges to a fixed value of about 0.0060 ($\neq$ true $\bar{\theta} = 0.0070$) for a wide range of $\hat{\theta}^{(0)}$ and s values. (In fact, from Table 1 we see that for the increasing failure rate model, $\hat{\theta}$ converges to the true $\bar{\theta}$ only under immediate inspection of flagged instruments. Also from run B1.10 in Table 2 we see that with deferred inspection, $\hat{\theta}$ does not converge to the true $\bar{\theta}$ even for the constant failure rate model.) The speed of convergence of $\hat{\theta}$ depends on the value of s. Both for $\hat{\theta}^{(0)} = \bar{\theta}/3$ or $3\bar{\theta}$, $\hat{\theta}$ converges in about 60 failures for $s = 1.8$, while for $s = 1.2$, about 300 failures are required. Notice that it would take a very long time in practice to observe so many failures.

(ii) A prior underestimate of θ ($\hat{\theta}^{(0)} = \bar{\theta}/3$) results in low initial PTI and ETI (resp., high EDL) values, which slowly increase (resp., decrease) to their true values. Opposite behavior is observed for prior overestimation of θ ($\hat{\theta}^{(0)} = 3\bar{\theta}$). It may be noted that the convergence of these performance measures is rather slow even for $s = 1.8$, but there is not more than 5% difference between their values at about 60 failures and at about 300 failures.

One may argue that it is unlikely that the process will remain in a steady state so long (1 year in the present example). It should be noted, however, that the failures observed over disjoint steady states can be pooled and the application of the Bayesian scheme can be carried over from one steady state period to the next as long as the type

TABLE 4

Convergence study of the Bayesian scheme relative to prior estimate of δ ($N_T = 50400$)

$\theta(1) = 0.0001$, $\beta = 7.7 \times 10^{-5}$, $\bar{\theta} = 0.007$, $\delta = 1.0$, $s = 1.8$, $\hat{\theta}^{(0)} = 0.007$, APC = 7, ADJ = 5, $\sigma = 0.2857$. Inspection frequency: once a shift. Decision rule: 2/3.

Run	$\hat{\delta}^{(0)}$	N_M	Simulation results					
			a	$\hat{\theta}$	$\hat{\delta}$	PTI	ETI	EDL
B1.23	0.50	1008	6	0.0067	0.815	0.0523	0.0601	4.828
		5040	30	0.0063	0.852	0.0529	0.0592	4.818
		10080	59	0.0062	0.858	0.0514	0.0585	4.893
		30240	198	0.0061	0.867	0.0496	0.0579	4.990
		50400	297	0.0060	0.869	0.0493	0.0572	5.010
B1.24	2.00	1008	6	0.0067	1.053	0.0235	0.0437	6.724
		5040	29	0.0062	0.912	0.0411	0.0536	5.287
		10080	59	0.0061	0.892	0.0443	0.0550	5.180
		30240	178	0.0061	0.882	0.0466	0.0561	5.105
		50400	296	0.0060	0.877	0.0472	0.0563	5.087

TABLE 5

Performance comparison of the Bayesian scheme with the measurement test scheme: effect of different failure rates ($N_T = 10080$) $\sigma = 0.2857$, $s = 1.20$, $\hat{\theta}^{(0)} = \bar{\theta}$, ADJ = 5, $\delta = 1.0$, $\hat{\delta}^{(0)} = 0.85$. Different values of $\theta(1)$, β and APC are used for different runs. Inspection: immediate. Decision rule: 2/3.

Run	$\bar{\theta}$	α	Simulation results				
			PMG *	PNG **	PTI	ETI	EDL
B1.25	0.00025	–	0.54	99.27	0.0091	0.0093	3.773
M1.25	0.00025	0.200	0.29	99.31	0.0105	0.0116	3.540
B1.26	0.001	–	2.33	97.02	0.0206	0.0218	3.758
M1.26	0.001	0.260	3.36	97.34	0.0188	0.0228	3.404
B1.2	0.007	–	9.36	85.17	0.0662	0.0740	2.985
M1.2	0.007	0.360	15.08	84.08	0.0420	0.0742	3.599
B1.27	0.021	–	20.92	65.38	0.1291	0.1461	2.667
M1.27	0.021	0.325	50.32	48.05	0.0408	0.1489	6.478

* PMG = Percentage of simulations with multiple gross errors.
** PNG = Percentage of simulations with no gross errors.

of instrumentation used and the basic process do not change.

Table 4 makes an analogous convergence study relative to the prior estimation of δ. We see that convergence of $\hat{\delta}$ is unaffected by the initial estimate $\hat{\delta}^{(0)}$. For both $\hat{\delta}^{(0)} = \frac{1}{2}\delta$ and $\hat{\delta}^{(0)} = 2\delta$, the final value of $\hat{\delta}$ is about the same but less than the true $\delta = 1$. Low $\hat{\delta}^{(0)}$ results in high (resp., low) initial PTI and ETI (resp., EDL) values which slowly decrease (resp., increase) to their true long run averages. Opposite behavior is noted for large $\hat{\delta}^{(0)}$.

Tables 5 and 6 give a comparative study of the two schemes based on the Bayesian test and the measurement test for gross error detection. Comparisons were made under identical implementation options (e.g., deferred decision, deferred inspection, etc.). To make a fair comparison between the powers (as measured via their EDL's) of the two schemes, the level of significance α used

TABLE 6

Performance comparison of the Bayesian scheme with the measurement test scheme: effect of different process networks ($N_T = 10080$) $\theta(1) = 0.0001$, $\beta = 7.7 \times 10^{-5}$, $\bar{\theta} = 0.007$, $\delta = 1.0$, $\hat{\delta}^{(0)} = 0.85$, $\hat{\theta}^{(0)} = 0.007$, $s = 1.2$, APC = 7, ADJ = 5, $\sigma = 0.2857$. Inspection: immediate. Decision rule: 2/3.

Run	α	Simulation results				
		PMG *	PNG **	PTI	ETI	EDL
B1.2	–	9.36	85.17	0.0662	0.0740	2.985
M1.2	0.360	15.08	84.08	0.0420	0.0742	3.599
B2.1	–	13.33	76.35	0.1262	0.1322	2.462
M2.1	0.650	15.86	75.16	0.1033	0.1333	2.666
B3.1	–	28.74	62.16	0.1188	0.1351	4.828
M3.1	0.235	94.26	6.82	0.0106	0.1398	50.050
B4.1	–	42.16	41.01	0.1766	0.2090	5.021
M4.1	0.250	99.50	0.66	0.0065	0.2078	72.449

* PMG = Proportion of simulations with multiple gross errors.
** PNG = Proportion of simulations with no gross errors.

for the measurement test was adjusted so that the ETI values for the two schemes were roughly equal.

Table 5 compares the two schemes for increasing frequencies (increasing $\bar{\theta}$) of gross error occurrences. The performance of the two schemes is very similar for low $\bar{\theta}$ (low incidence of multiple gross errors and high incidence of no gross errors). But for high $\bar{\theta}$ the Bayesian scheme is much more powerful (has low EDL values), although it has higher PTI too. (The ETI values for the two schemes are adjusted to be approximately equal.)

Table 6 compares the two schemes for networks 1–4. Here the average failure rate $\bar{\theta}$ is kept the same throughout. In that case, the larger the network the higher the incidence of multiple gross errors.

(i) First note that, as in Table 5, the Bayesian scheme dominates the measurement test scheme in terms of EDL when there is a greater likelihood of multiple gross errors being present. In fact, PMG and PNG (defined in footnotes to Table 6) can themselves be viewed as performance measures, and it may be noted that for the measurement test scheme multiple gross errors are present much more frequently (and very rarely are there no gross errors present) than is the case for the Bayesian scheme. Thus the poorer performance of the measurement test scheme in the presence of multiple gross errors has an adverse effect that builds upon itself.

An explanation for the improved performance of the Bayesian scheme under higher frequencies of occurrences of gross errors is that the effect of

TABLE 7

Performance comparison of the Bayesian scheme with the measurement test scheme for network 5 (NH_3 synthesis loop) ($N_T = 10080$)

$\theta(1) = 0.0001$, $\beta = 7.7 \times 10^{-5}$, $\bar{\theta} = 0.007$, $\delta = 1.0$, $\hat{\delta}^{(0)} = 0.85$, $\hat{\theta}^{(0)} = 0.007$, $s = 1.5$, APC = 7, ADJ = 5, $\sigma = 0.2857$. Inspection frequency: once a shift. Decision rule: 2/3.

Projected constraint matrix **B**

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & -1 & -0.5 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0.98 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1.5 & -0.02 \\ 0 & 0 & 1 & 0 & -0.02 & 0 & 0 & 0 \end{bmatrix}$$

$y = (N_2^{(1)}, H_2^{(1)}, Ar^{(1)}, N_2^{(2)}, Ar^{(2)}, N_2^{(3)}, NH_3^{(4)}, H_2^{(5)})$

Run	α	Stream	Simulation results				
			PTI	ETI	EDL	$\hat{\delta}$	$\hat{\theta}$
B5.1	–	1		0.00100	2.38	1.060	0.0062
		2		0.00565	8.00	1.180	0.0048
		3		0.00178	2.05	0.930	0.0060
		4		0.00129	34.57	11.660	0.0050
		5		0.00436	49.97	20.080	0.0042
		6		0.00298	104.55	11.110	0.0037
		7		0.00238	4.30	1.200	0.0057
		8		0.00575	49.01	21.700	0.0040
		Overall	0.0226	0.02518	25.21	8.620	0.0049
M5.1	0.10	1		0.00119	3.02		
		2		0.00575	15.67		
		3		0.00020	2.45		
		4		0.00347	96.65		
		5		0.00258	90.39		
		6		0.00357	92.08		
		7		0.00456	6.85		
		8		0.00665	56.37		
		Overall	0.00130	0.02797	33.71		

misspecification of prior estimates of $\bar{\theta}$ and δ are wiped out faster by the accumulating data on instrument failures.

(ii) Networks 2 and 3 have the same number of streams but network 2 has 10 nodes while network 3 only has 6. From the results of the runs for these two networks we see that the performance of each scheme is much worse for network 3 compared to that for network 2 (PMG and EDL are higher and PNG is lower). This illustrates the point first noted by Iordache et al. [9] that the power of a gross error detection test is adversely affected by too many interconnections in the network. A convenient measure to characterize this property of a network is the average of the 'degrees' of the nodes in the network, where the degree of a node is defined as the total number of streams entering or leaving the node.

Table 7 gives results for the ammonia synthesis loop shown in Fig. 5. The constraint matrix **B** for this example, obtained by premultiplying the original constraint matrix by a projection matrix (see ref. 6), contains two pairs of proportional columns, namely, for measurements 2 and 8, and for 3 and 5; measurements 4 and 6 also have nearly proportional columns. (Note that in this network the measurements correspond to component flow rates, and not total stream flow rates. This correspondence is given in Fig. 5 and Table 7 via the y vector. Thus, e.g., measurement 1 corresponds to N_2 flow rate in stream 1, measurement 5 corresponds to Ar flow rate in stream 2, etc.) The adverse effects of these linear dependencies in the columns of the **B** matrix are reflected in the extraordinarily long expected delays (low powers) in detecting gross errors in streams 4, 5, 6 and 8. The final estimates of the δ's for these measurements are substantially upward biased. The $\hat{\theta}$ values for the same measurements are substantially downward biased; for others they are somewhat less so. Even longer expected delays are obtained for the measurement test in most cases. See ref. 9 for an explanation of why proportional columns result in low powers and hence long expected delays for the measurement test. To summarize, both the tests give their worst performances for this network (with the measurement tests's performance being worse than that of the Bayesian test) because of the proportional columns in the transformed balance matrix. To improve the performance of any detection scheme it would be necessary to augment the original balance matrix with additional constraints based on additional stoichiometric information if any.

In addition to the above we also made some runs using unequal parameter values σ_i, $\bar{\theta}_i$, δ_i, etc. The results of these runs are reported by Iordache [10], and they show that the Bayesian scheme offers better performance than the measurement test scheme under a variety of complex scenarios. However, highly unequal σ_i's result in highly biased estimates of the δ_i's using the Bayesian scheme.

12 CONCLUSIONS AND RECOMMENDATIONS

We have made a detailed study of the Bayesian scheme in this two-part paper. Despite the limitations of the basic model, which are listed in Section 6, the following conclusions can be drawn from this study, and consequent recommendations can be made.

(i) The Bayesian approach offers the promise of improving gross error detection and identification capabilities by using past failure data. Its technical feasibility is demonstrated by this investigation. However, much remains to be done to make it a practical method. In the mean time it may serve as a useful vehicle for accumulating potentially valuable data on the instrument failure rates.

(ii) The Bayesian scheme is relatively robust against misspecification of the prior information (particularly initial estimates of the δ_i's) and nonnormality of the data. However, its convergence properties are not very satisfactory. Starting with an initial guess of θ_i between 33% and 300% of the true value, a large number of observed failures (about 60 to 300) is needed before θ_i converges. A somewhat smaller number is needed for the key performance criteria to converge within $\pm 5\%$ of their steady state values. So accurate initial estimates of the θ_i's are needed before the method can be put to practical use. If there is uncertainty about the prior estimate of θ_i, then a greater weight should be placed on the current data by

choosing a larger value of s_i, which improves the performance of the scheme to some extent.

(iii) In general, the Bayesian scheme outperforms the measurement test scheme in the following situations: high frequencies of gross errors (multiple gross errors), large spreads in the magnitudes and frequencies of gross errors, and long delays in confirmation and repairs.

(iv) If instrument failures are infrequent, then the measurement test scheme performs slightly better than the Bayesian scheme at least for incorrect prior information. A suitable strategy in this case might be to initially apply the measurement test scheme until sufficient failures are observed to enable formation of accurate prior estimates of the θ_i's and the δ_i's. At that point the Bayesian scheme can take over. A point worth repeating in this context is that the failure data from disjoint steady state periods can be pooled.

(v) The computational effort involved in the application of the Bayesian scheme is substantially greater than that involved in the application of the measurement test scheme. A comparison on the Cray machine shows a ratio of 4:1 in the two computational times for network 4 (after implementing the computational shortcuts for the Bayesian scheme described in Section 5.4). This ratio will be larger for larger networks because the computations for the Bayesian scheme increase exponentially with n, while those for the measurement test scheme increase linearly with n. Further research is needed to alleviate this problem.

To conclude, we note some areas for future research. In Section 6 of Part I we have noted a number of practical features that are not incorporated in the present model. It would be desirable to analyze some real data to see which assumptions in the present model are most seriously violated. Then an extensive simulation study should be conducted to examine the robustness of the Bayesian scheme to the violations of those assumptions. This would help us in identifying the most critical practical features that should be incorporated in a future Bayesian model so that the resulting gross error detection scheme can then be applied in practice with reasonable confidence.

ACKNOWLEDGEMENT

This research was supported by NSF Grant CBT-8519182.

REFERENCES

1 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.
2 A.C. Tamhane, A note on the use of residuals for detecting an outlier in linear regression, *Biometrika*, 69 (1982) 488–489.
3 R.D. Cook and S. Weisberg, *Residuals and Influence in Regression*, Chapman & Hall, New York, 1982.
4 R.W. Serth and W.A. Heenan, Gross error detection and data reconciliation in steam-metering systems, *American Institute of Chemical Engineers Journal*, 32 (1986) 733–742.
5 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.
6 A.C. Tamhane and R.S.H. Mah, Data reconciliation and gross error detection in chemical process networks, *Technometrics*, 27 (1985) 409–422.
7 D.M. Downham and F.D.K. Roberts, Multiplicative congruential pseudo-random number generators, *Computer Journal*, 10 (1967) 74–77.
8 D.E. Knuth, *The Art of Computer Programming*, Vol. 2, Addison-Wesley, Reading, MA, 1969.
9 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.
10 C. Iordache, *A Bayesian Approach to Gross Error Detection in Process Data*, Ph.D. Dissertation, Northwestern University, Evanston, IL, 1987.