LA-10974-MS

UC-15
Issued: April 1987

LA--10974-MS

DE87 010833

Measurement Control Program for New Special Recovery

S. -T. Hsue
K. Campbell
G. Barlich

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

CONTENTS

CONTENTS (cont)

CONTENTS (cont)

MEASUREMENT CONTROL PROGRAM FOR NEW SPECIAL RECOVERY

by

S. -T. Hsue, K. Campbell, and G. Barlich

ABSTRACT

This report summarizes the design of the measurement control (MC) program for the New Special Recovery facility. The MC program is divided into two levels. Level 1 MC checks are performed at the individual instrument computer and will always be functional even when the instrument-control computer is down. The level 1 MCs are divided into statistical checks for both bias and precision, and diagnostic checks. All the instruments are connected on line to an instrument-control computer to which the measurement results can be communicated. Level 2 MC analyses are performed at this computer. The analyses consist of control charts for bias and precision and statistical tests used as analytic supplements to the control charts. They provide the desired detection sensitivity and yet can be interpreted quickly and easily. Recommendations are also made in terms of the frequency of the tests, the standards used, and other operational aspects of the MC program.

I. INTRODUCTION

The purpose of this report is to summarize the design of the measurement control (MC) program for the New Special Recovery (NSR) project.

NSR is a new plutonium scrap recovery facility being constructed at the Savannah River Plant (SRP). The Los Alamos National Laboratory is collaborating with SRP to develop an integrated system of modern, automated nondestructive assay (NDA) instrumentation that will provide nuclear materials accounting and process monitoring information to the operators of this facility. The goal is to provide an accountability system that is capable of drawing frequent material balances with minimum reliance on laboratory measurements of analytical samples.

This state-of-the-art instrumentation has been designed and fabricated by the Los Alamos National Laboratory, Lawrence Livermore National Laboratory, Mound Laboratories, and Savannah River Laboratory.[1] Los Alamos is also serving as system coordinator, combining the individual components into an integrated package with the nine NDA instruments reporting to a central instrument-control computer (ICC). The ICC, in turn, is integrated into a facility computer network that includes other computers dedicated to process control and nuclear materials accounting functions. The integrated system is illustrated schematically in Fig. 1.

Since the accountability system receives inputs from the various NDA instruments, it is crucial to have a well-developed quality assurance program

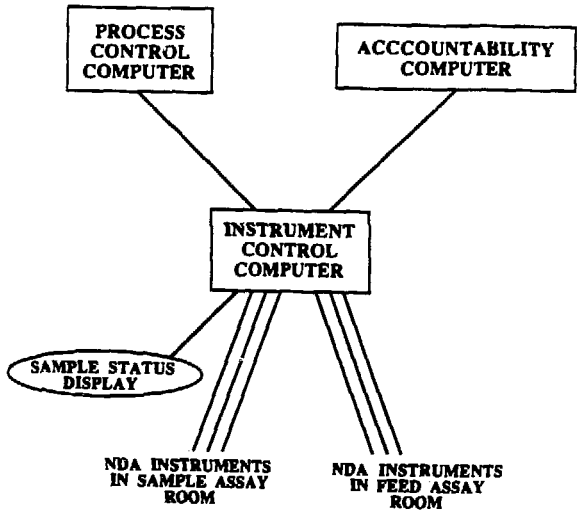


Fig. 1. Schematic diagram of the major computer network system for New Special Recovery.

for these various instruments. The purpose of the MC program is to provide this quality assurance on a routine basis. There are no generally accepted methods of providing the quality assurance for NDA, partly because most of the NDA instruments have been developed only recently. Destructive chemical analysis, however, has a well-developed quality assurance program. We follow the principles established by the quality assurance for chemical analysis, but there are intrinsic differences between NDA and chemical analysis. For example, the chemical analyses are very much subject to human factors; for NDA instruments, the operators in general have very little impact on the assay results. We also follow the guide outlined in "A Measurement Control Program for Nuclear Material Accounting" prepared by Battelle Northwest.[2]

Because of the presence of the central computer and because all NDA instruments have individual computers [Digital Equipment Corporation (DEC) Micro-11s], the MC checks are divided into two levels. Level 1 MC checks, which include several simple statistical checks and diagnostic checks, are performed at the individual NDA computers. Level 2 MC is performed at the ICC computer, where more extensive statistical checks and trend analyses are carried out. The division of the two levels is shown in Fig. 2.

The level 1 MC statistical checks in this report are based on those implemented at the Los Alamos Plutonium Facility (LAPF) during the DYMAC demonstration,[3,4] which have been in use since 1979. In developing subsequent more up-to-date NDA systems for the LAPF, we have added diagnostic MC checks. We have found these useful in detecting instrumental problems.[5] The level 1 MC checks will be discussed in Sec. III.

Since 1979, there have been several studies of the MC experience at LAPF.[6,7,8] Based on these studies, several recommendations were made. Most of these recommendations have been incorporated into level 2 of the MC. The level 2 MC checks will be discussed in Sec. IV.

In establishing the control limits, we assume that the measurement results from all NDA instruments are normally distributed at least approximately. The method to generate the control limits for different instruments will be discussed in Sec. V. Additional tests of statistical assumptions such as normality are also included in Sec. V.

Although the MC here is designed for the NSR facility, the methods described in this report are quite general and can be easily adapted to other

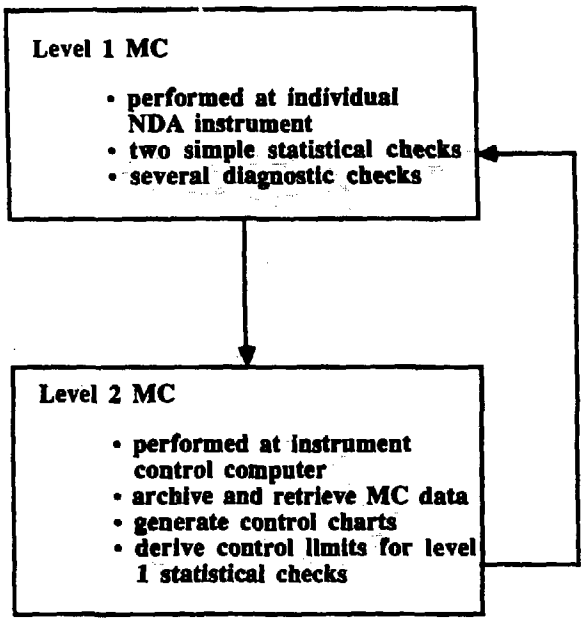


Fig. 2. Functions of level 1 and level 2 measurement control checks.

facilities. This report presents our most up-to-date design of the MC program for an NDA counting laboratory.

II. FACILITY AND MEASUREMENT SYSTEM

A. Facility Description

This new facility is designed to recycle and recover off-specification plutonium, excess process waste, and process scrap. The layout of the process is shown in Fig. 3 and a simplified description of the process is shown in Fig. 4. Input materials scheduled for recovery are brought from a storage vault to a dedicated feed assay room (FAR) where their total plutonium contents are determined for input accountability purposes using a series of NDA measurements. Following the input accountability measurements, the materials are transferred into a feed glove box train (feed) for preparation before their dissolution. Generally, the solids isotopic analyzer measurement will be used in conjunction with either the calorimeter measurement or the feed coincidence counter (FCC) measurement to determine the total plutonium mass of input materials in the FAR.

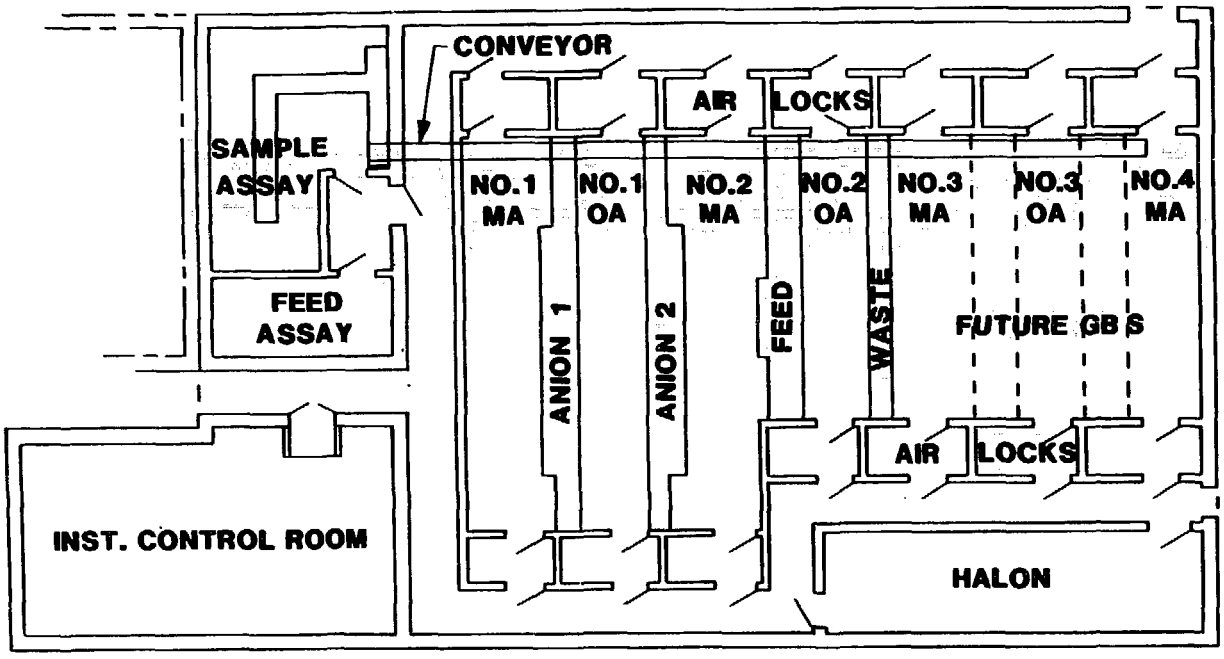


Fig. 3. Physical layout of the plutonium scrap recovery facility.

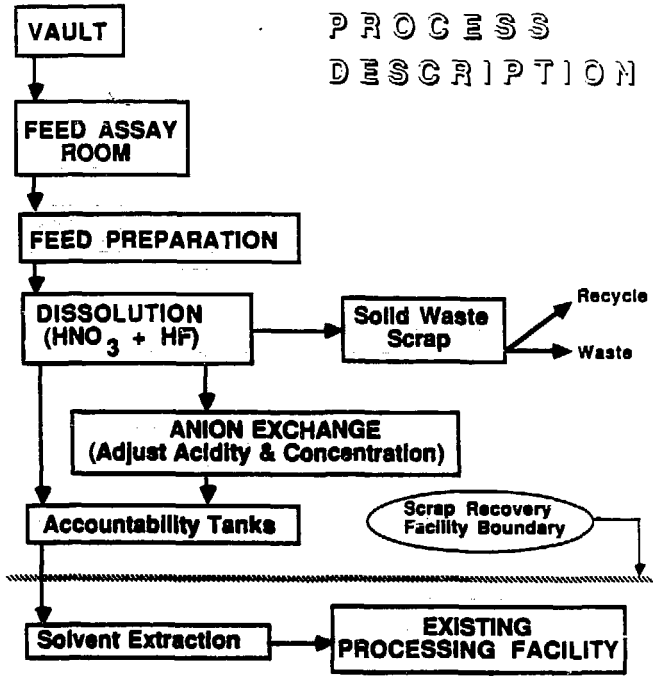


Fig. 4. Overall description of the processes in the recovery operation.

The plutonium scraps are dissolved in nitric and hydrofluoric acids. After dissolution, the plutonium-bearing solution can either be transferred directly to the accountability tank or to one of the two anion-exchange columns, depending on the fluoride tolerance of the existing processing facility. Anion-exchange columns will separate americium, uranium, and other impurities from the plutonium. Required NDA measurements of dissolver/anion-exchange cabinet solutions are performed on samples obtained from process vessels and transferred by conveyor to a dedicated glove box in a sample assay room (SAR). In the SAR, the integrity of samples is assured by first checking for suspended solids with the turbidimeter and a visual inspection and then comparing the sample density, measured with the densitometer, to the measured density, obtained with differential pressure transducers, of the tank from which the sample was obtained. If these tests are successful, the plutonium concentration of the sample is then measured in either the x-ray fluorescence analyzer, the gamma-ray pulse-height analyzer (PHA), or the low-solution assay instrument (LOSAI) analyzer. Duplicate samples will be drawn from critical accountability points and be assayed by two independent techniques. Anticipated sample throughput is 35 per day.

Solid waste from the process will be transferred by conveyor to the waste-handling cabinet, recovered by a wash/rinse process, and then assayed for plutonium content by NDA before removal from the process glove box. Liquid waste will be sampled and assayed for plutonium content by NDA in the SAR. In addition to the instrumentation discussed above, a sodium iodide gamma-ray detector array will be deployed at key points in the process. Measurements from this array will be used to determine holdup in process equipment, set criticality alarm limits, and monitor the process.

B. NDA Measurement System Description

The design of the integrated NDA measurement system is shown in detail in Fig. 5. Each NDA instrument that is part of the integrated system is microcomputer-based and thus is capable of stand-alone operation if the ICC is out of service. The instruments, their suppliers, the nature of the NDA measurement, and the location of the instrument in the facility are given in Table I. Extensive performance tests were made by each vendor with du Pont/SRP representatives using formal acceptance procedures at each vendor site. These tests involved measurements of plutonium standards and solutions over and above the

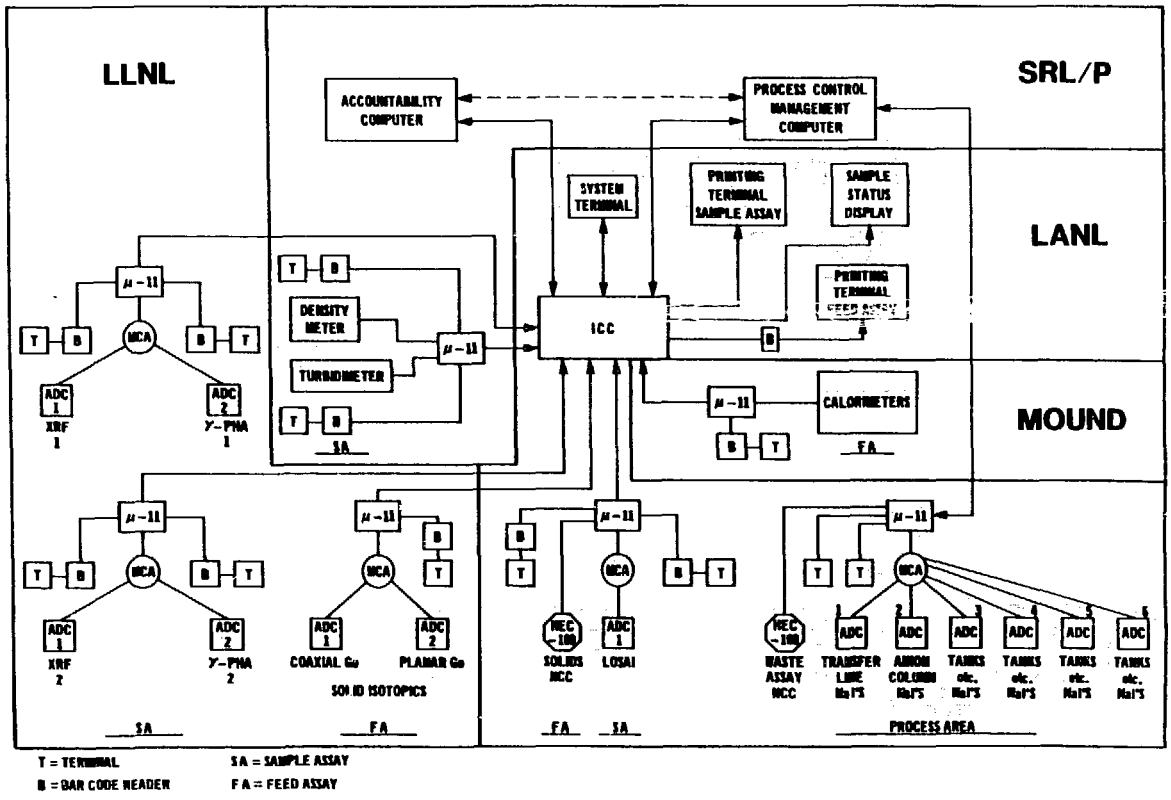


Fig. 5. Detailed design of integrated NDA system for New Special Recovery.

TABLE I

NDA INSTRUMENTS FOR INTEGRATED MEASUREMENT SYSTEM

Instrument	Supplier	Measurement Type	Location
FCC	LANL	Effective ^{240}Pu mass	FAR
Calorimeter (CAL) (4)[a]	Mound	Heat	FAR
Solids isotopic analyzer (ISO)	LLNL	Pu isotopic fractions	FAR
Turbidimeter (TUR)	SRL	Suspended solids	SAR
Densitometer (DEN)	SRL	Solution density	SAR
X-ray fluorescence (XF) (2)[a]	LLNL	Pu concentration	SAR
Gamma PHA (GP) (2)[a]	LLNL	Pu concentration	SAR
LOSAI	LANL	Pu concentration (Low)	SAR
Waste coincidence counter (WCC)	LANL	Effective ^{240}Pu mass	Process area
NaI monitor array (NaI)(13)[a]	LANL	^{239}Pu mass	Process area

[a]Number of individual measurement units.

anticipated measurement range for NSR samples. In addition, 100-h endurance testing of each system was performed successfully to ensure component reliability. The following is a brief description of each of the instruments in the measurement system:

1. Feed Coincidence Counter. The FCC was designed for the passive neutron coincidence counting of feed material up to 2.25 kg of plutonium and will be used in the feed assay room. This counter was designed for high count rates and uniform response. The coincidence count rate from a sample varies ±1% from the bottom to the top of the sample cavity; a sample moved 1 cm off center in a radial direction changes the coincidence rate by +0.2%. Counting rates up to 500 kHz can be processed. Sample cans up to 8 in. in diameter by 12 in. high can be placed into the sample carrier.

One special feature of the FCC is the electric elevator capable of lifting the upper end plug so samples can be inserted by raising the elevator and opening the cage door. The door is electrically and mechanically interlocked to prevent the elevator from moving when the cage door is open.

The FCC has been calibrated for pure PuO_2 powder; assays can be performed with or without neutron multiplication corrections. There is room in the parameter files for many more sets of calibration data to suit specialized measurement requirements. These files are easily created or modified by a user who knows the supervisory access code.

SRP has gained extensive experience in the measurement of plutonium receipts using similar instrumentation. Even though the system has been calibrated with PuO_2, it is anticipated that it will be useful for other reasonably homogeneous scrap materials. The FCC can be calibrated by selecting representative items from a shipment, determining the plutonium mass by combining gamma isotopics and calorimetry, and then using them as calibration standards for the FCC. SRP experience indicates that results in the 1-2% range can be obtained in less than 10 min.

For MC purposes, a ^{252}Cf or a plutonium oxide sample can be used.

2. Calorimeter. The calorimeters are of the isothermal (constant temperature) resistance bridge type designed to cover 2 to 12 W. The accuracy of the calorimeter is 0.2% over this wattage range. The calorimeter housing, incorporating the dual-stage heat exchanger with temperature-controlled

circulation bath, is designed to minimize space requirements. All isothermal calorimeters require a thermally stable environment. This heat exchange and bath maintain the calorimeter sensing element at 25°C ± 0.001°C as long as the room temperature does not vary beyond ±1.1°C. A ^{238}Pu heat source and PuO_2 standards will be used for MC purposes.

3. Solids Isotopics System. A gamma-ray spectrometry system utilizing two detectors has been designed to measure plutonium isotopic ratios in feed scrap materials in sealed containers. Both the FCC and the calorimeter require plutonium isotopic distribution to obtain the plutonium mass of the sample. A low-energy photon-type germanium detector (also called a low-energy photon spectrometer, or LEPS) is used to measure the low-energy plutonium gamma-ray spectrum. At the same time this spectrum is being acquired, a large coaxial germanium detector (COAX) is used to measure the higher-energy gamma rays. The LEPS detector is required, whereas the COAX detector is optional. Once both spectra are obtained, a sophisticated analysis code, the Multiple Group Analysis (MGA) code, is used to simultaneously solve 120 linear equations representing all of the major peak regions in both spectra and to determine the isotopic abundance with their attendant errors. For MC purposes, a plutonium standard with well-characterized isotopic distribution will be used.

4. Turbidimeter and Densitometer. The turbidimeter and densitometer are used to check the integrity of the liquid samples.

MC checks of the turbidimeter are made before each sample analysis using a standard.

The densitometer system goes through an automatic MC self-check cycle during each sample measurement. This MC procedure involves measurement of the density of air and water as part of each measurement cycle to ensure proper performance.

5. X-Ray Fluorescence Analysis. The x-ray fluorescence system consists of a stainless-steel cell mounted inside the glove box into which ~15 mℓ of plutonium solution is drawn. Two collimated 20-mCi ^{57}Co sources are used to ionize K-shell electrons from plutonium atoms; ionization is immediately followed by K x-ray emission. The emitted x rays are measured by a LEPS detector located outside the glove box. A third, highly collimated ^{57}Co source

transmits a 122-keV beam through the solution into the detector. A computer-controlled tungsten shutter is used to eclipse the exciter sources so that the passive radioactivity spectrum can be measured. The data analysis software subtracts this spectrum from the fluorescence spectrum to obtain a net spectrum. The ratio of net fluoresced plutonium K x-ray intensity to 122-keV transmission intensity is measured to determine the plutonium concentration. This method measures concentrations independent of plutonium isotopics and minor variations in the composition and density of the solution. When solutions with high levels of ^{241}Am are measured, a rhodium absorber foil mounted on a slide is moved into place under computer control during the assay. There is also a thorium foil mounted in front of the ^{57}Co transmission source; the thorium x ray/122-keV transmission intensity ratio is monitored for MC purposes.

6. Gamma-Ray Pulse-Height Analyzer. The second type of NDA system assays plutonium solution concentrations by gamma-ray spectrometry. A stainless-steel cell mounted inside a glove box will hold 17 mℓ of nitric acid solution containing plutonium concentrations from 8 to 100 g/ℓ. The cell is closely coupled to a LEPS detector located outside the glove box. The plutonium concentration and isotopics are determined from the measured intensity of the emitted gamma rays. Two collimator sizes (1-1/8 and 7/8 in. in diameter) are available under computer control to adjust the input counting rate. The smaller collimator includes a 10-mil cadmium absorber and will be used when solutions contain high concentrations of ^{241}Am. A ^{109}Cd monitor source is used to correct for small changes in detector efficiency or inaccuracies in measuring the true counting duration. A sealed plutonium metal source can be moved into position to periodically check the system calibration.

7. Low-Solution Assay Instrument. The LOSAI is designed to assay solutions from the effluent stream of the anion-exchange column and other plutonium-bearing samples with low concentrations (5 to 1000 mg/ℓ). The 414-keV gamma ray is used as the assay peak for these effluents. Pulse pileups resulting from the summing of the 208-keV gamma rays are corrected. The system consists of a coaxial detector with 25% efficiency viewing a stainless-steel cell with a sample volume of 35 mℓ located inside the glove box. A ^{133}Ba

radioactive source is used to compensate for counting-rate-related losses. A sealed ^{239}Pu foil has been incorporated as part of the system for MC use.

8. Waste Coincidence Counter. The waste coincidence counter (WCC) is built around a well in the waste-handling glove box. This counter is designed to measure large cans of plutonium waste and scrap up to 14-in. in diameter by 16 in. high. The assay precision specification (1σ) is 1 g plutonium or 25%, whichever is larger. Samples are placed into an elevator in a glove box and are lowered into a glove-box well, which is surrounded by the coincidence counter. The counter is installed in the waste-handling glove box. It reports the assay to the process-control computer. A small plutonium oxide sample is used for MC.

9. NaI Monitor Array. The NaI system is used mainly for process control and to ensure nuclear criticality safety; it reports to the process-control computer. The 13 NaI detectors are installed on the anion/dissolver cabinets and the waste-handling cabinet. All of the monitors have a built-in ^{241}Am seed for self-checking. Some monitor locations use a ^{137}Cs check source for MC purposes; the transfer-line monitors may be checked with a sealed plutonium source.

III. LEVEL 1 MEASUREMENT CONTROL CHECKS

As mentioned in the introduction, there are two levels of MC checks. In this chapter we will discuss the level 1 MC. The level 1 MC checks are performed at the individual NDA instrument computer and are operational even when the ICC is down. Level 1 MC checks are the foundations upon which the level 2 checks are based.

The checks in level 1 can be roughly divided into two categories: statistical (external) checks and diagnostic (internal) checks. The statistical checks require the operator to perform assays on standards or foils on a regular basis. The standard nominally is one with known values established independently by chemical analysis. Sometimes a foil standard may be used with solution systems because solution standards are known to change with time. In these cases, the solution-equivalent-concentration values of these foil

standards cannot be determined independently. These are considered secondary standards. Because of the time required to perform these checks and for ease of operation, only one standard (primary or secondary) is used per system for MC purposes. If experience shows that one standard is not adequate to check the bias over the dynamic range of the instruments, then more standards should be used. Also, more standards will be used during the calibration. These statistical checks are also called external checks.

The diagnostic checks are internal checks that do not require any operator action. These checks monitor the performance of the system to ascertain that it is operating within the limits that the developers deem appropriate. Because each system is different, the diagnostic checks are highly system dependent. The great advantage of the diagnostic checks is the fact that they can be performed on every sample measured. Thus, they can guard against unknown samples not appropriate for the system. In this respect, they are different from the quality-assurance checks for chemical analysis.

A. Statistical Checks

Two components of the measurement uncertainties will be monitored: the bias and the precision.

1. Bias Check. The purpose of this check is to test the validity of the calibration and to monitor the system bias. This is accomplished by measuring a primary standard or a secondary standard on a regular basis.

- Frequency

  We recommend that this check be performed daily on all systems with the exception of calorimetry, which, because of the assay time, can be done on a weekly basis. Systems designed for process-control purposes, such as some of the NaI monitors, may have to be checked less frequently because of lower accuracy requirements and because of the difficulty of introducing a standard in front of the detector.

- Standard

  This check can guard against bias at one assay value and one sample type, namely that of the standard. For this reason, the standard

should be selected at the middle of the dynamic range of the instrument and the most usual sample type. This check does not guard against drifts at the extremes of the dynamic range or for other sample types. However, in most NDA systems the changes in "relative" detection efficiency of the system are nominally less likely than changes in "absolute" detection efficiency. If there is evidence that the detection efficiency changes differently at the extremes of the dynamic range, then two standards near the two extremes of the dynamic range should be measured routinely. At the present time only one standard is recommended for each system.

For the solution assay systems, a plutonium foil has been incorporated into each system with the exception of the x-ray fluorescence. The plutonium foil will be used for the statistical checks. However, the plutonium equivalent concentration of these foils cannot be determined reliably by chemical means. The effective values of these foils should be determined immediately after a calibration and probably verified after each calibration. The x-ray fluorescence system uses the measured ratio of the thorium foil x rays and 122-keV peak from the ^{57}Co transmission source for its bias checks because no plutonium foil is built in. The standards used in each of the systems are summarized in Table II.

Check

The bias check consists of the following:

W_0 is the value of the standard determined by an independent method,

W_i is the measured value of the standard, and

Std Dev is the historical standard deviation of W_i when the system is "in control." Methods to calculate this value will be discussed in Sec. V.

For each measurement, the measured W_i and W_0 are compared. If the difference is greater or equal to the warning limit but less than the action limit, a warning message is printed at the NDA terminal. If the difference is greater than or equal to the action limit, a warning message will be printed and the bias check flag will be set to false (nominally set to true). The control limits are listed in Table III.

- Comments

  This check has been modified from what has been in the MC program at LAPF. In the MC program at LAPF, a T test is performed in which

TABLE II

STANDARD USED FOR MEASUREMENT CONTROL FOR NSR

System	Standard	Comment
FCC	^{252}Cf	
CAL	^{238}Pu heat source	
ISO	Pu oxide	
XF	^{57}Co, Th foil	Built in
GP	Pu foil	Built in
LOSAI	Pu foil	Built in
WCC	Pu oxide	
NaI	Pu source, ^{137}Cs	

TABLE III

CONTROL LIMITS FOR BIAS CHECK

Type	Limit (Std Dev)	Probability (%)
Warning	1.96	5
Action	3.00	0.3

$$T = \frac{W_i - W_0}{S_a} \quad , \qquad (1)$$

where S_a is the estimated precision (from counting statistics). Essentially we have replaced the estimated precision by the standard deviation from historical data. The problem of using the estimated precision in the control limits is the fact that for most systems, the day to day variability of the system is usually greater than from the counting statistics alone.

The bias check as outlined above does not guard against bias in solution samples from other origins. For example, the volume of the solution sample (and therefore the assay results) may be off because of air bubbles. It is, therefore, desirable to assay the same sample occasionally by means of an independent analysis. The independent analysis can be a destructive chemical assay or a completely different NDA. We should mention that in the NSR facility, duplicate accountability samples will be assayed by two completely different NDA techniques.

2. Precision Check. The purpose of this check is to verify that the random error of the instrument is within control. The check is accomplished on repeated assays by comparing the calculated error with the standard deviation on repeated assays without disturbing the standard. Because it is impossible to account for all the sources of the random error that would give rise to the imprecision of a system, we will try to specify what we are and are not monitoring.

- Frequency
  We recommend that this check be done on a monthly basis for all systems with the exception of the calorimeter and process-control monitors, for which this check is not practical.

- Standard
  The same standard for the bias check can be used for this check.

- Check

  The precision check consists of 5 or 15 replicate measurements of the standard. The reduced chi-square for a series of measurement is

$$x^2/\nu = S_n^2/\sigma_n^2 \quad , \tag{2}$$

where S_n^2 is the variance based on observed variance from repeated measurements, and

σ_n^2 is the variance based on counting statistics propagated in the analysis program for the specific instrument.

Control limits used for the reduced chi-square parameter are shown in Table IV. Assuming normal distribution, the warning and action limits represent 5% and 1% failure probability, respectively.

For those who are involved in the instrument development, it is important to make sure that the variance based on the counting statistics propagated in the assay system is in agreement with the observed variance from repeated measurements; that is, that the average value of the reduced chi-square is 1. One should be careful about what is variable during the repeated assay and what remains constant. For example, background is taken once a day, and the same background is used in all subsequent assays. The error from the background, therefore, should not be included in the propagated error during the precision check. It is also desirable to verify that the assay results are indeed normally distributed because the warning and action limits are based on this assumption.

- Comments

  This precision check has been in use since 1979 at LAPF and has been found useful. This check guards against some of the random errors but not all. For example, the effect of random positioning of the sample in the instrument is not measured by the reduced chi-square statistics.

TABLE IV

REDUCED CHI-SQUARE CONTROL LIMITS

No. of Observations	Warning Limit		Action Limit	
	Lower	Upper	Lower	Upper
5	0.12	2.79	0.05	3.72
15	0.40	1.87	0.29	2.24

3. Background Check. In addition to the bias and the precision checks, we also require that the background radiation be measured regularly. The background measurement is particularly important for low-level systems or for systems in a high background. If the background limits are exceeded, then a cleanout is recommended.

- Frequency
  We recommend that the background be measured on a daily basis, except where the background may be rapidly changing. For certain NDA systems, such as the calorimeter and plutonium solid isotopic, the background assay is not necessary.

We have arbitrarily grouped the background check with the bias and precision checks because its implementation (and therefore its software structure) is similar to the other two statistical checks. A summary of the statistical checks and the recommended frequency is shown in Table V.

The operator will be notified by the instruments if the bias, precision, or the background measurement is out of date and should be performed or is outside of acceptable range. In addition, information will be sent to the ICC to indicate the MC status of each instrument. The statistical checks discussed here are not mandatory; assays still can be performed but with warning messages in the print-out.

B. Diagnostic Checks

Diagnostic checks are designed to verify that the system is performing within the bounds of the system; they can be used in a diagnostic manner to

TABLE V

SUMMARY OF THE STATISTICAL CHECKS FOR NSR

System	Bias (Daily)	Precision (Monthly)	Background (Daily)
FCC	X	X	X
CAL	X[a]		X[a]
ISO	X	X	
XF	X	X	X
GP	X	X	X
LOSAI	X	X	X
WCC	X	X	X

[a]Calorimeter systems require a baseline run, which is essentially a background assay. Because of the assay time the bias run for each calorimeter should be performed approximately once a week.

pinpoint problems with the system. In contrast with the statistical checks, these checks do not require special sample or special action by the operators and can be performed with every assay and on every sample. These checks have no statistical basis and in general are different for every system.

For the NSR project, each of the instruments has a maximum of 10 diagnostic checks described as follows.

1. Feed Coincidence Counter. In the FCC, there is an accidental-to-totals check, testing the consistency of the coincidence rate. If anything goes wrong with the system, this check would alarm promptly.

2. Calorimeter. In the calorimeter, there are seven diagnostic checks:

- lid opened on calorimeter during run,
- bath out of temperature limit during run,
- room out of temperature limit during run,
- bias run time limit exceeded,
- baseline run time limit exceeded, and
- bias value limit exceeded.

3. Solids Isotopics System. In the solid isotopic system, there are two detectors in the system: LEPS and COAX detectors. There are five diagnostic checks, which are listed below:

- full-width-half-maximum (FWHM) of americium 59.5-keV peak (LEPS),
- centroid of the 59.5-keV peak (LEPS),
- centroid of the 208-keV peak (LEPS),
- FWHM of the 208-keV peak (COAX), and
- centroid of the 413.7-keV peak (COAX).

4. Turbidimeter and Densitometer. Both these systems have internal checks that are somewhat different from the rest of the systems. For details of the checks, see the manuals provided by SRL.

5. X-Ray Fluorescence Assay. In the x-ray fluorescence assay system, there are two diagnostic checks:

- FWHM of the 122.1-keV peak, and
- centroid of the 122.1-keV peak.

6. Gamma-Ray Pulse-Height Analyzer. In the gamma PHA system, there are three diagnostic checks:

- FWHM of the 88.0-keV peak,
- centroid of the 88.0-keV peak, and
- centroid of the 208.0-keV peak.

Note: The 88.0-keV peak is due to the ^{109}Cd monitor source.

7. Low-Solution Assay Instrument. In the LOSAI system, there are four diagnostic checks against the amplifier gain drifts and detector resolution degradations.

- centroid of lower-energy stabilizing peak,
- FWHM of the same peak,
- centroid of higher-energy stabilizing peak, and
- FWHM of the same peak.

8. Waste Coincidence Counter. In the WCC, there is one diagnostic check, which is the reals-to-totals check. If the reals-to-totals ratio is too small,

it is an indication that the sample may have relatively high (α,n) induced neutron emission.

9. NaI Monitor Array. For each of the 13 NaI detectors, there are 3 diagnostics checks that mainly monitor the americium alpha seed pulses, which are used for both gain stabilization and deadtime correction:

- centroid of americium alpha peak,
- FWHM of the americium alpha peak, and
- integrated count rate of the americium alpha peak.

IV. LEVEL 2 MEASUREMENT CONTROL CHECKS

The daily (or, in the case of calorimeters, perhaps weekly) bias measurement for each instrument, discussed in Sec. III.A.1, is reported to and archived by the ICC, together with the nominal value of the standard in use and other information such as the propagated error associated with the measurement. These archived data are the bases for the level 2 MC checks described in this section, including

(1) control charts for bias,
(2) control charts for precision, and
(3) sequential tests for shifts in the mean.

Typically, the individual bias measurements made using a given standard on a given instrument will cluster around an average value with a spread reflecting uncontrollable factors affecting individual measurements differently. These factors might include counting error in the measurement of radioactive material, small variations in atmospheric conditions and temperature, or slight differences in the procedures used by different operators. As observed in Sec. III, not all of these sources of variation are accounted for by level 1 MC checks, but all will be reflected by a typical sequence of bias measurements taken over a period of several weeks. Level 2 MC checks are based on "standardized" observations, the differences between the actual measurements and their nominal value divided by the historical standard deviation, which describes this variability. In this section this historical standard deviation

is considered a known constant, or else a known function of stored parameters and the data. Derivation of these parameters from historical data is discussed in the following section.

Generally, the distribution of observations about their mean value is adequately described by a Gaussian, or normal, probability distribution, which is completely characterized by its mean value and standard deviation. This assumption underlies all tests described in this section. A test for its verification is included in Sec. V.

It is also assumed that the result of one measurement does not in any way affect the result of another (that is, that the observations are "independent"). A test of this assumption will also be found in Sec. V.

The objective of examining the archived sequence of control measurements is to determine whether the instrument, its associated procedures, and its operators continue to function at an acceptable level of precision and accuracy. If so, we say that the measurement process is "in control." Specifically, given a sequence of control measurements, we will assume that the process continues in control as long as the average value and standard deviation of the observations do not deviate significantly (in a statistical sense) from their historical values. Thus, the algorithms described below are designed (1) to produce graphical displays that will provide visual indications of deviation from the historical norms, and (2) to provide automatic indicators of such deviations in a timely manner, based on statistical tests of the null hypothesis that the measurement process is in fact in control.

A. Algorithms for the Standard Case: Constant Mean and Standard Deviation

In the simplest case, the measurement process is considered to be in control if the MC observations are independently normally distributed with constant mean and standard deviation. Of the instruments discussed in Sec. II.B, this is a reasonable model for measurements using coincidence counters (FCC and WCC), calorimeters, the solids isotopic system, and the LOSAI. (The LOSAI does contain a decaying source that can be expected to affect measurement precision over time, but its half-life of 13 years greatly exceeds the anticipated calibration period.) Thus, in this case we are dealing with the problems of detecting deviations in the measurement process from the known mean value (that is, detecting "bias"), and also with changes in the historically

determined standard deviation that provides a measure of the significance of differences between the actual observations and the known mean value.

1. Shewhart Control Charts. The use of control charts for bias and precision provide for visual inspection of data and enable an alert reviewer to spot problems possibly before they are detected by statistical tests.

Let μ be the known mean and σ^2 the known variance of the measurement process when it is in control. Let x_t denote the control measurement of the standard made at time t. We will consider the standardized measurement

$$z_t = \frac{(x_t - \mu)}{\sigma} \quad . \tag{3}$$

Under the standard assumptions discussed above, the z_t is an independent realization of a normal random variable with mean zero and unit variance. In particular, about 1 observation in 400 is expected to exceed 3 in absolute value.

A Shewhart control chart is a plot of the standardized measurement z_t vs time t, and includes horizontal lines at zero (the mean value of the standardized measurement) and, generally, at plus and minus three units (the "control limits"). (See Fig. 6.) If the standardized observation falls outside these control limits, this is considered to be an indication of bias, an alarm requiring some action such as recalibration of the instrument or at least a review of recent MC data.

Shewhart charts are familiar in most facilities and are easy to interpret. A series of measurements plotted in this way can reveal abnormalities that may have crept into the measurement process, which are not immediately detectable out of their historical context. However, the statistical tests implicit in the drawing of control and warning limits are less than optimal, especially for the detection of small shifts in the mean value (although numerous ad hoc solutions to this problem have been developed[9]). In Sec. IV.A.2 a two-sided Page's test is described that, as a modification of a sequential probability ratio test, is known to have some optimal properties in terms of the length of

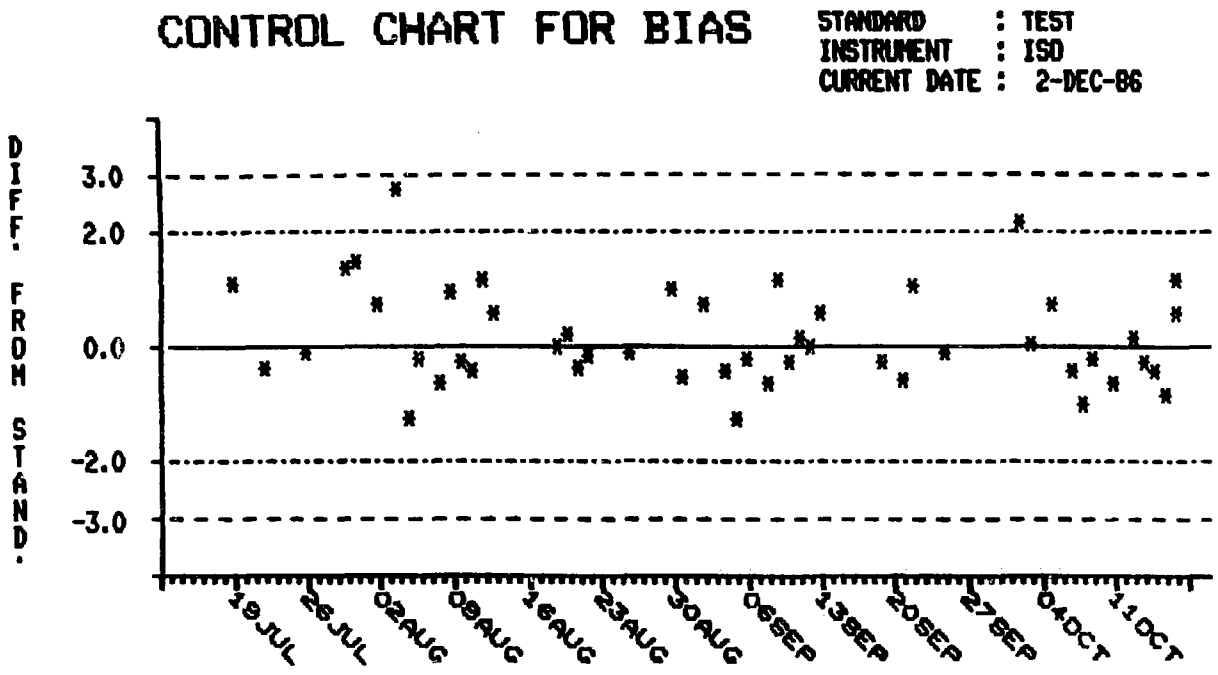


Fig. 6. Sample control chart for bias.

time (number of observations) required to detect shifts in the mean of the process.

A Shewhart control chart can also be developed for control of the standard deviation of the measurement process (Fig. 7). This chart is based on a statistic, which is the sample standard deviation of a sequence of individual observations of length r = 5. The statistic

$$(r - 1)\, s_t^2 = \sum_{i=t-r+1}^{t} (z_i - \bar{z})^2 = \frac{\sum_{i=t-r+1}^{t} (x_i - \bar{x})^2}{\sigma^2} \qquad (4)$$

(where z is the average of the r standardized observations z_i, i = t-r+1, ..., t, and similarly x is the average of the r observations x_i) has a chi-square distribution with r-1 degrees of freedom, and upper and lower control limits

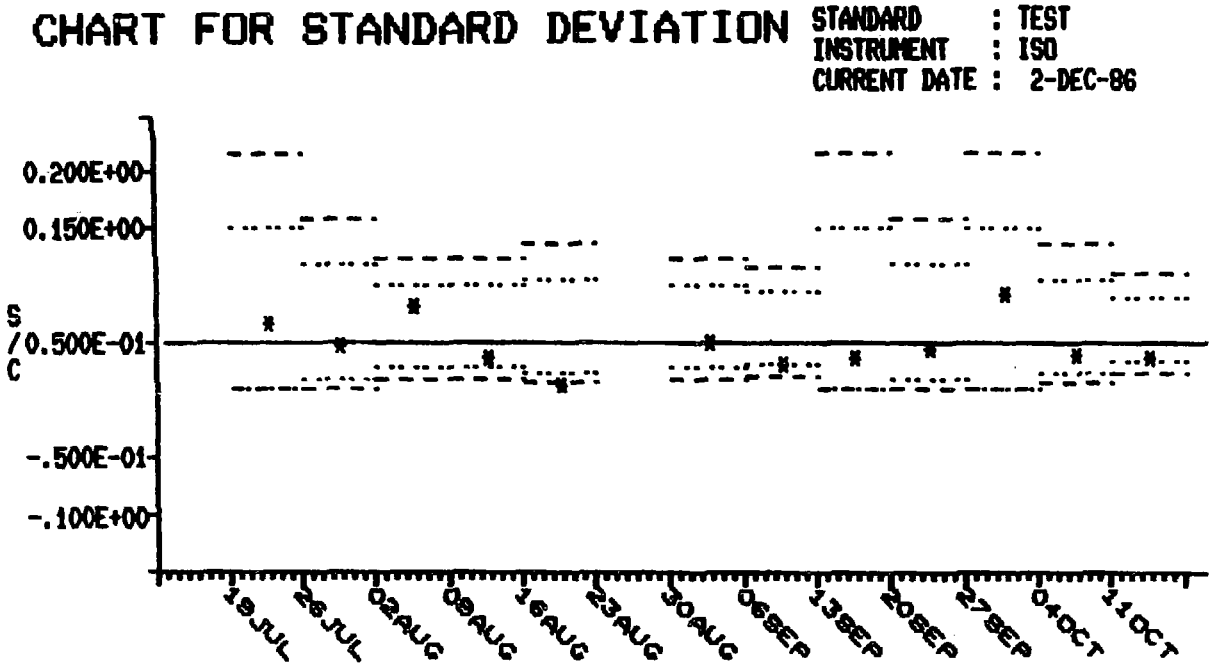


Fig. 7. Sample control chart for standard deviation.

for the sample standard deviations s_t, computed after every r observations, are constructed using the tabled percentiles of this distribution. Specifically, the mean value of s_t, defined by Eq. (4), is c_r, given in Table VI for r between 2 and 12; upper and lower action limits may be set at j_r^+ and j_r^-; and optional warning limits may be set at v_r^+ and v_r^-.

2. Sequential Test for Bias. Page's test is based on a statistic closely related to the cumulative sum of the past observations.[10] It has been described as a cusum test with a restart mechanism; this prevents recent problems from being obscured by a long history of satisfactory behavior. It can be alarmed by a single observation deviating from the mean by a large amount [$(h + k)\sigma$ in terms of the notation given below], by a short sequence of observations deviating from the mean by a smaller amount, or by a longer sequence of observations deviating by an even smaller amount. Thus many of the modifications proposed by Roberts[9] as alternatives to the simple test implied by a Shewhart chart with control limits at ±3 are captured in a single test.

TABLE VI

MEAN VALUE WITH ACTION AND WARNING LIMITS FOR THE SAMPLE STANDARD DEVIATION OF r STANDARDIZED OBSERVATIONS[a]

r	c_r	j_r^-	j_r^+	v_r^-	v_r^+
2	0.798	0.	3.29	0.03	2.24
3	0.886	0.03	2.63	0.16	1.92
4	0.921	0.09	2.34	0.27	1.77
5	0.940	0.15	2.15	0.35	1.67
6	0.952	0.21	2.03	0.41	1.60
7	0.959	0.25	1.93	0.45	1.55
8	0.965	0.29	1.86	0.49	1.51
9	0.969	0.35	1.81	0.52	1.48
10	0.973	0.36	1.76	0.55	1.45
11	0.975	0.38	1.72	0.57	1.43
12	0.978	0.41	1.69	0.59	1.41

[a]Based on chi-square distribution with r - 1 degrees of freedom.

A two-sided Page's test for deviations in the mean of the measurement process, based on the standardized observations z_t, takes the following form:

(1) Select a reference value k (typically about 0.5), and a decision value h (typically in the range 4 to 5). Set m(0) = 0 and M(0) = 0.

(2) For t = 1, 2, 3, ..., compute

$$m(t) = \max \{0, m(t-1) + z_t - k\}$$

and

$$M(t) = \max \{0, M(t-1) - z_t - k\}$$

(max {x, y} denotes the larger of the two numbers x and y).

(3) Take action (that is, declare the measurement process out of control) at time t if either m(t) or M(t) exceeds h.

Typically, h and k are chosen so that it will take about 10 or 12 observations to detect a shift in the mean of the standardized measurements z_t of magnitude 1 (that is, a shift of magnitude σ from the mean value μ of the raw observations x_t) while allowing an average run length of the process in control (that is, the average time until a false alarm) to be on the order of 300 to 400 observations.[11]

The interesting region of the (h,k) parameter space is indicated in Fig. 8. The two heavy curves bound a region in which the average run length of the process in control exceeds 300 observations and the average time until detection of a shift of 1 standard deviation is less than 12 observations. Values of (h,k) chosen from the right-hand side of this region will lead to tests that have longer run lengths when the process is in control, but also take longer, on the average, to detect a shift in the mean. Shaded regions indicate where (a) the average run length of the controlled process exceeds 400 observations, and (b) the average time until detection of a shift of 0.9 times the standard deviation is less than 12 observations. From the latter we see that it is not possible to improve on the time to detection of even marginally smaller shifts without driving the average run length of the controlled process to unacceptably small values. It is worth noting, also, that even with an average run length of 400 observations for the process in control, a significant number of runs will be shorter, because the distribution of run lengths is approximately exponential; almost one run in five will be less than 100 observations.

The probability of at least one alarm occurring as a function of run length is shown in Fig. 9 for (h,k) = (5,0.5) and Fig. 10 for (h,k) = (7.5,0.5). Plots of the Page statistics m(t) and M(t) are shown on the same chart with control limit at h and -h (Fig. 11). Small changes in bias are reportedly easier to spot on such charts.

B. Modifications for Nonstandard Cases

The modifications required for nonstandard cases consist in general of replacing the known constant mean μ and/or known constant standard deviation

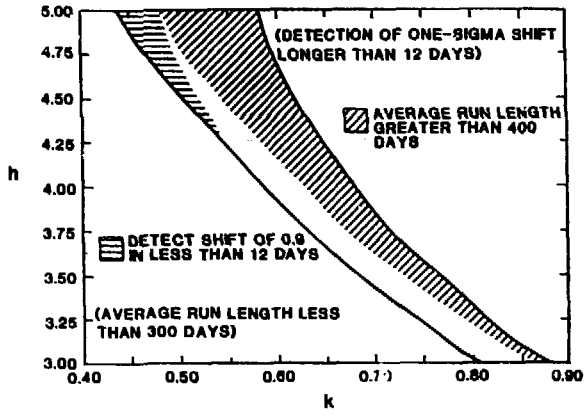
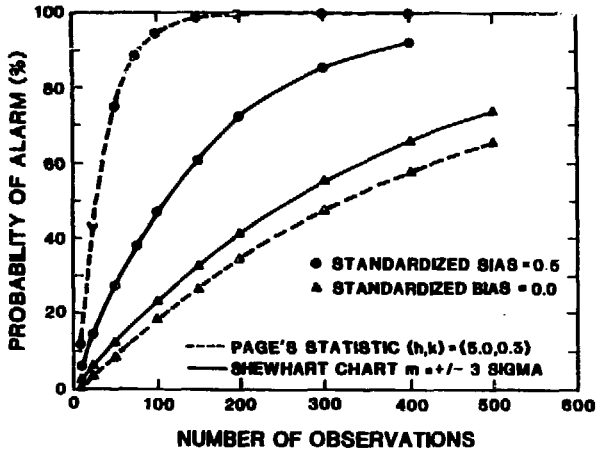


Fig. 8. Usable region of parameter space for two-sided Page test for shift in mean. Curves bound a region in which the average run length of the process in control exceeds 300 observations and the average time until detection of a shift of 1 standard deviation is less than 12 observations.

Fig. 9. The probability of at least one alarm's occurring based on the Page statistic with (h,k) = (5,0.5) for standardized measurement biases of 0 and 0.5. Also included is the probability of at least one alarm's occurring based on the Shewhart chart with control limits at m ± 3σ.



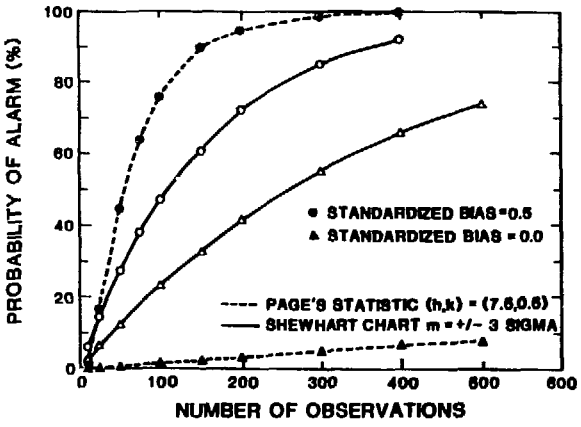


Fig. 10. The probability of at least one alarm's occurring based on the Page statistic with (h,k) = (7.5,0.5) for standardized measurement biases of 0 and 0.5. Also included is the probability of at least one alarm's occurring based on the Shewhart chart with control limits at m ± 3σ.

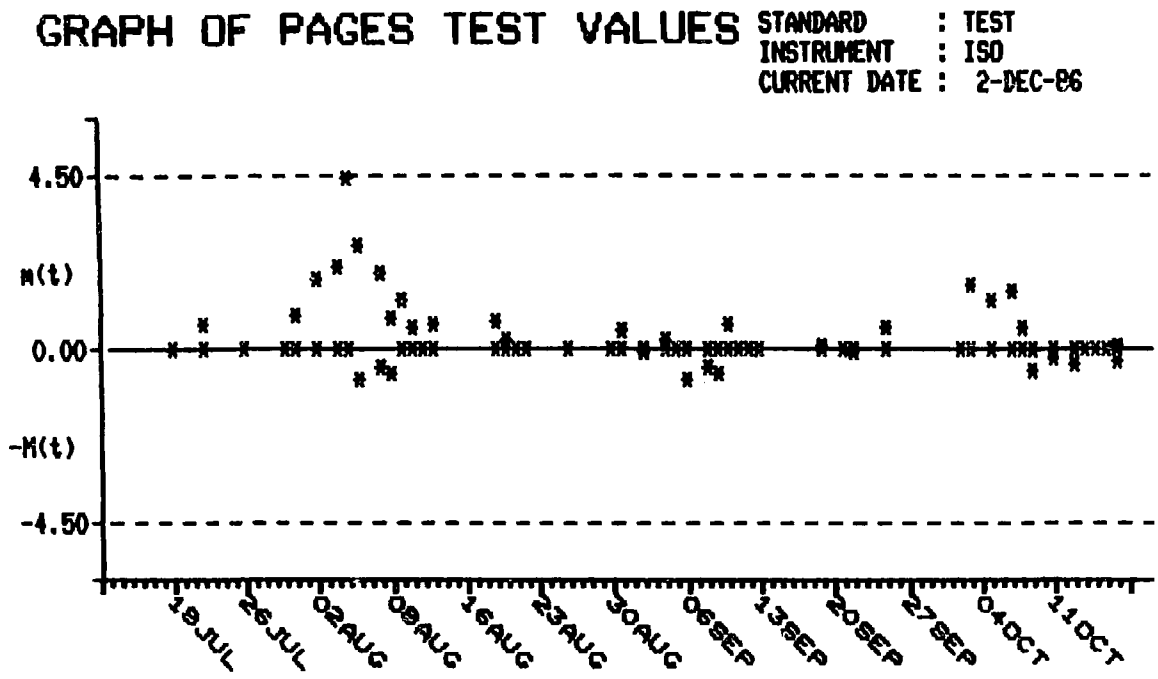


Fig. 11. Page statistics m(t) and M(t) as a function of time, with (H,k) - (4.5, 0.5).

σ used to standardize the observations in Eq. (3) by functions of t. Specifically, we consider the cases where the mean at time t is a known function of t, μ_t; where the standard deviation at time t is a known or estimated function of t, σ_t; and where there is the possibility of repeated measurements (more than one measurement of the same standard each time a control measurement is scheduled.)

1. Changing (Decaying) Standard Value. When the standard contains a radioactive element with a relatively short half-life, noticeable changes in the mean value μ of the measurement process may occur over the usable life of the standard. This type of change is easily accommodated if the change is not too significant, because the half-life of the element, as well as the initial quantity of material present, can be assumed known to high precision, and hence the mean value μ_t of the measurement x_t at time t can be computed and used to standardize the observation as above. Significant decay, however, may also affect the standard deviation of the observation, one component of which is counting error (proportional to the square root of the mean

in a simple Poisson model of the measurement process, or a more complicated function in a less direct measurement taking into account background, other radioactive sources, etc.). One way to compensate for this effect is to increase the counting time used in making the control measurement. Otherwise a model for the standard deviation $\sigma = \sigma_t$ in Eq. (3) must be constructed and used as discussed in the following section.

2. Modification for Deteriorating Measurement Precision. In the case of NDA instruments with short-lived sources and possibly when the standard itself is decaying, as mentioned above, one component of the variance may change over time, while the other component (the part due to other factors such as varying atmospheric conditions, sample placement, etc.) remains constant if the measurement process is in control. Both the x-ray fluorescence system and gamma-ray spectrometry use sources with a half-life of 453 days (^{109}Cd), which will affect the precision of measurements made with these systems. For this case we need a model for the changing component, or some measurement of the changing component, while the constant part is estimated using historical data in a manner similar to that used for the standard case (see Sec. IV.B). The time-dependent standard deviation $\sigma = \sigma_t$ needed for standardization of the observation [Eq. (3)] can be modeled as the square root of the sum of the computed variance τ_t^2 and the constant component η^2:

$$\sigma_t^2 = \eta^2 + \tau_t^2 \ . \tag{5}$$

3. Modifications When Repeated Measurements Are Made. Multiple daily measurements of the same standard (such as two control measurements always made at the same time) could be treated as independent observations. It is probably more instructive, however, to redefine the control measurement x_t as the average of these measurements, with the same mean μ as the original measurement process but with standard deviation that is smaller by the square root of n, the number of replicates. Shifts that can be detected by Page's test using the resulting standardized observations, for example, are similarly smaller by a factor of $\sqrt{n}$, with the same parameters as before. (For example, with three

daily measurements, we can expect to detect a shift on the order of 0.6σ in 12 days while maintaining an average run length of 400 days for the controlled process.)

V. COMPUTATIONS WITH HISTORICAL DATA

The discussion in the previous section made a number of assumptions about the sequence of bias measurements used as the basis for the tests described therein: that their standard deviation was known (at least as a known function of the data that might include an internally generated component of variance), that they were random, and that they were normally distributed. In this section we take up the problem of estimating the historical standard deviation, and discuss tests for verifying the other assumptions.

A. Estimation of the Historical Standard Deviation

We have assumed that the mean value μ is known to high precision, whether as the result of independent assay or of careful measurement following calibration (in the case of a secondary standard). The standard deviation of the process, on the other hand, can only be determined by analysis of historical data or of data from a designed experiment. These data must be collected under circumstances that accurately reflect all the sources of random error that will affect routine accountability measurements--for example, using several operators and following usual procedures over a reasonably extended period of time.

This historical standard deviation must be carefully distinguished from specification limits claimed by the manufacturer of the instrument or required by the user. If these are significantly smaller than the historical limits, this may be a problem that needs to be investigated, but is separate from the MC program. The limits that are relevant for MC are the historical, empirical ones. As long as the standard deviation of the observed measurement process is not significantly different from its historical value (and no bias, or shift in the mean value, is apparent), the measurement process is considered to be in (statistical) control, even if the precision or accuracy of the measurement fails to meet administrative requirements.

Therefore, in order to estimate σ we require an initial set of observations. These may be obtained during a trial period or, if necessary, during

the early phase of actual operation. To the extent possible, these data should be collected using procedures, operators, and environments that are typical of actual accountability measurements. (Re-estimation of σ may be indicated later, in which case the recent history of the measurement process can be used. However, once computed, σ, or η in the case of the modification described by Eq. (5), is assumed known and constant unless it becomes obvious, from one of the standard tests or control charts, that it has been incorrectly estimated or has changed since originally computed.)

1. Standard Case: Constant Mean and Standard Deviation. Suppose we have a history of N measurements of the standard. Because these historical data may have some long-term problems (for example, a slow trend or recalibration in the middle of the series), the estimate of σ is obtained by segmenting the data into small, relatively homogeneous subgroups, for example, by weeks; σ^2 is estimated by the within-group variance of the historical data, rather than by the overall variance. If the historical data contain some extremely divergent observations (outliers) for which an assignable cause can be determined, these observations should be deleted before performing the computations below, as they can strongly influence the estimate of σ^2. It may also be worthwhile testing the historical data for normality (see Sec. IV.C).

Specifically, then, let the (edited or transformed, if necessary) historical observations be denoted by x_{ijk}, where $k = 1,\ldots,K$ for K calibration periods, $j = 1,\ldots,J_k$ for J_k subgroups within the k^{th} calibration period, and $i = 1,\ldots,I_{jk}$ for the individual measurements in the $(j,k)^{th}$ subgroup. Form the within-group means,

$$\bar{x}_{\bullet jk} = \frac{1}{I_{jk}} \sum_{i=1}^{I_{jk}} x_{ijk} \tag{6}$$

and compute

$$s^2 = \sum_{k=1}^{K} \sum_{j=1}^{J_k} \sum_{i=1}^{I_{jk}} (x_{ijk} - \bar{x}_{\bullet jk})^2 . \tag{7}$$

Then σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{s^2}{N - M} \quad , \tag{8}$$

where M is the total number of subgroups; that is,

$$M = \sum_{k=1}^{K} J_k \quad , \quad N = \sum_{k=1}^{K} \sum_{j=1}^{J_k} I_{jk} \quad . \tag{9}$$

With the data grouped as above into short intervals, the procedure described above is also suitable for the first nonstandard case considered in Sec. IV, namely a slowly changing mean value μ_t, provided that the standard deviation σ_t is not seriously affected as a result.

2. Modification for Time-Dependent Standard Deviation. Substantial complications are presented by NDA instruments with short-lived sources resulting in measurements of significantly decreasing precision during the periods between replacement. Similar considerations apply if a rapidly decaying standard implies changes in the counting statistics associated with its measurement. We assume that the data available for computation of the historical standard deviation cover a period sufficiently long to reflect such changes, and possibly at least one replacement of the source (or standard). The standard deviation σ_t to be used in Eq. (3) will be modeled as the square root of the sum of two variances: first, a component τ_t^2 associated with the observation at time t and estimated by the software package of the instrument, and second, a constant component η^2 to be estimated from the historical data (given the τ_t's associated with these data).

It is assumed that the mean value of the measurement process is approximately constant (or only slowly changing) during the period covered by the historical data. (For an instrument with a secondary foil standard, the "known" mean value may in fact change when the instrument is recalibrated using primary

standards. Measurements from two calibration periods should, of course, not be grouped together in a single subgroup below.)

Specifically, let the historical measurements be denoted by x_{ijk} as before, over K calibration periods, with J_k groups in the k^{th} calibration period (grouped observations obviously should not include replacement of the transmission source), and I_{jk} measurements in the $(j,k)^{th}$ group. Assume that the computed error (that is, the variable component of the measurement variance) is approximately constant in the $(j,k)^{th}$ group, and denote it by τ^2_{jk}. (In practice, τ^2_{jk} is computed as the average of the computed variances of the measurements in the $(j,k)^{th}$ group.) Compute the within-group means as in Eq. (6). Then η^2 is estimated by the solving the equation

$$\sum_{k=1}^{K} \sum_{j=1}^{J_k} \sum_{i=1}^{I_{jk}} \frac{(x_{ijk} - \bar{x}_{\bullet jk})^2}{\eta^2 + \tau^2_{jk}} = N - M \qquad (10)$$

numerically. (If τ^2_{min} and τ^2_{max} are the smallest and largest values of τ^2_{jk} in the historical data, and s^2 is defined as in Eq. (7), then the solution to Eq. (10) lies in the range

$$\left[\frac{s^2}{N - M} - \tau^2_{max}, \frac{s^2}{N - M} - \tau^2_{min}\right] ,$$

and a reasonable starting value for solving Eq. (10) numerically is

$$\frac{s^2}{N - M} - \frac{1}{M} \sum_{k=1}^{K} \sum_{j=1}^{J_k} \tau^2_{jk}$$

For a future measurement at time t, the expected value of the measurement is μ_t, the measured value of the standard obtained immediately after the

last calibration (or computed in the case of a decaying radioactive standard), and the standard deviation will be σ_t,

$$\sigma_t^2 = \eta^2 + \tau_t^2 \quad , \tag{11}$$

where τ_t^2 is the computed variance returned with that measurement.

B. Tests for Normality and Outliers

The control charts of Sec. II and other tests all assume that measurements obtained from the MC program are approximately normally distributed. Failure of normality can affect both the false-alarm rate of the tests and their "power"--that is, their ability to detect real problems in the process. Therefore we recommend occasionally running a test of normality. In particular, verification of normality of the historical data before estimating the standard deviation that will be used in future control charts and tests is important. If serious deviations from normality are discovered (and if these deviations cannot be ascribed to the presence of a few outliers), we may wish to consider the application of some normalizing transformation to our measurements. For example, distributions with long tails on the right may be made more symmetric by a logarithmic or square root transformation.

1. Shapiro-Wilks Test of Normality. The most powerful test of normality found in the literature is the Shapiro-Wilks test.[12] Application of this test requires a table of coefficients that are different for each sample size as well as a table of critical values. For practical purposes we set standard sample sizes and store the required coefficients $a_{n,i}$, $i = 1, 2, \ldots n/2$, for these sample sizes, along with the critical values c_n. Coefficients and critical values for significance levels (false-alarm rates) of 0.05 and 0.01 are tabulated in Tables VII and VIII for n = 20, 30, 40, and 50. The statistic W for a sample of n observations from the MC process is computed as follows:

(a) Order the n observations from smallest to largest. Label the ordered observations $x_1, \ldots, x_n$. (We assume n is even.)

TABLE VII

COEFFICIENTS $a_{n,i}$ FOR SHAPIRO-WILKS STATISTIC W[a]

Coefficient i	Number of Observations (n)			
	20	30	40	50
1	0.4734	0.4254	0.3964	0.3751
2	0.3211	0.2944	0.2737	0.2574
3	0.2565	0.2487	0.2368	0.2260
4	0.2085	0.2148	0.2098	0.2032
5	0.1686	0.1870	0.1878	0.1847
6	0.1334	0.1630	0.1691	0.1691
7	0.1013	0.1415	0.1526	0.1554
8	0.0711	0.1219	0.1376	0.1430
9	0.0422	0.1036	0.1237	0.1317
10	0.0140	0.0862	0.1108	0.1212
11		0.0697	0.0986	0.1113
12		0.0537	0.0870	0.1020
13		0.0381	0.0759	0.0932
14		0.0227	0.0651	0.0846
15		0.0076	0.0546	0.0764
16			0.0444	0.0685
17			0.0343	0.0608
18			0.0244	0.0532
19			0.0146	0.0459
20			0.0049	0.0386
21				0.0314
22				0.0244
23				0.0174
24				0.0104
25				0.0035

[a]From Ref. 12, pp. 603-604.

TABLE VIII

CRITICAL VALUES FOR SHAPIRO-WILKS TEST[a]

Level	Number of Observations (n)			
	20	30	40	50
0.01	0.868	0.900	0.919	0.930
0.05	0.905	0.927	0.940	0.947

[a]From Ref. 12, p. 605.

(b) Compute

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad ,$$

where $\bar{x}$ is the usual sample mean,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad .$$

(c) Compute

$$b = \sum_{i=1}^{n/2} a_{n,i} \, (x_{n-i+1} - x_i) \quad ,$$

where the coefficients $a_{n,i}$ are obtained from Table VII.

(d) Compute $W = b^2/S^2$, and compare the result to the critical value with the desired significance level in Table VIII. If W is smaller than the given critical value, the data appear to depart significantly from normality, and should be examined further to determine if the

problem is very light tails, heavy tails, skewness, or perhaps just a small number of outliers for which an assignable cause can be determined.

The storage requirements of this test preclude its routine use for samples of arbitrary size. If, say, 38 historical observations are available, we use the most recent 30.

2. Elimination of Outliers. A common form of nonnormality is a "contaminated normal" distribution; that is, data that are fundamentally normally distributed but that include occasional outliers caused by some change in the process generating the observations. When such observations are detected, and when the reason for the change can be determined, these observations should be deleted from the data used to test normality of the underlying distributions.

The most effective way to detect such outliers, given an adequate historical record, is by means of the control chart; points lying outside the action limits are considered outliers. In general, outlier tests look for observations that are substantially removed from the bulk of the data; for example, the tests described by Grubbs[13] are based on the statistic

$$T_{extreme} = \frac{x_{extreme} - \bar{x}}{s} ,$$

where $x_{extreme}$ is the observation in the set farthest removed from the sample mean $\bar{x}$ and s is the sample standard deviation. ($T_{extreme}$ is called the "extreme studentized residual.") Occasionally it may be useful to have a test for outliers that uses only the current week's observations. For detecting a single outlier in samples of sizes three to eight, the absolute value of $T_{extreme}$ above can be compared to the critical values in Table IX, with $x_{extreme}$ being considered an outlier if the critical value is exceeded. Generalizations for the detection of two or more outliers are considered by Rosner.[14]

TABLE IX

CRITICAL VALUES FOR ABSOLUTE VALUE OF $T_{extreme}$ [a]

n	Level 0.01	0.05
3	1.15	1.15
4	1.49	1.46
5	1.75	1.67
6	1.94	1.82
7	2.10	1.94
8	2.22	2.03

[a] From Ref. 13, p. 4.

C. Other Tests of Randomness

The statement that there is no "assignable cause" contributing to the variability in the MC data is equivalent to asserting that the remaining variability is "random," the result of uncontrollable factors, which vary from day to day or from measurement to measurement. While randomness is not precisely defined, we have some intuitive opinions of features that we would not consider random, such as extreme observations (four or more standard deviations removed from the mean), or a slow upward drift in the observations. Visible cycles, possibly with weekly or other natural periods, would alert us to look for potentially controllable factors affecting the measurements, as would correlations between measurements and other factors, such as the operator, if such information were available.

Von Neumann Test. A test that is quite sensitive to many types of nonrandom fluctuations is based on the Von Neumann ratio (the ratio of the mean square successive difference to the variance, or equivalently, the serial correlation coefficient).[15,16] This test statistic is computed as follows:

(a) Let $x_1, \ldots, x_n$ be the n observations in the order in which they were made.

(b) Compute

$$S^2 = \sum_{i=1}^{n-1} (x_i - \bar{x})^2 \quad ,$$

where $\bar{x}$ is the usual sample mean.

(c) Compute

$$\delta^2 = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \quad .$$

(d) Compare the value of the ratio $T = \delta^2/S^2$ with the critical values in Table X. In general, in MC programs, the alternative to no correlation between measurements that are of interest is the possibility of positive correlation between successive measurements, and so we perform a one-sided test. Positive correlation will tend to reduce the value of T, and thus we would reject the hypothesis of no correlation if T were smaller than the value indicated in Table X.

(e) For samples of size $n > 25$, compute

$$Z = \left(\frac{n^2 - 1}{n - 2}\right)^{1/2} \left(\frac{T}{2} - 1\right)$$

and compare Z with the percentiles of a standard normal. Z will be small for the positively correlated case; thus the hypothesis of no correlation would be rejected at the 5% level if $Z < -1.645$, or at the 1% level is $Z < -2.326$. A two-sided test for either positive or negative correlation between successive measurements would reject the null hypothesis at the 5% level if the absolute value of Z exceeded 1.96, or at the 1% level if the absolute value exceeded 3.09.

TABLE X

LOWER PROBABILITY POINTS FOR VON NEUMANN RATIO[a]

n	Level 0.01	0.05
5	0.538	0.820
6	0.561	0.890
7	0.614	0.936
8	0.665	0.982
9	0.709	1.025
10	0.752	1.062
11	0.791	1.096
12	0.828	1.128
13	0.862	1.156
14	0.893	1.182
15	0.922	1.205
16	0.949	1.227
17	0.974	1.247
18	0.998	1.266
19	1.020	1.283
20	1.041	1.300
21	1.060	1.315
22	1.078	1.329
23	1.096	1.342
24	1.112	1.355
25	1.128	1.367

[a]From Ref. 15, p. 287 and Ref. 16, p. 446.

VI. MC SYSTEM OPERATION AND DEVELOPMENT

The software tools provided in the integrated system are intended as a basis for a comprehensive MC program. In this section, we make recommendations for their use in an on-going MC program and review a few areas in which modifications may be found desirable based on experience.

A. Operational Recommendations

1. Establishment of Historical Standard Deviation. As emphasized in the descriptions of the various tests for lack of control of a given instrument, a key parameter is the standard deviation of the measurement process, which provides a scale factor with which deviations of individual measurements from the known value of the standard are to be compared. Statistical procedures for estimating this parameter are described in Sec. V; we reiterate that this is an empirical quantity, not some administratively dictated limit or a manufacturer's specification limit, although once established it (or at least one component of it) is treated as a constant.

Initially, it will be necessary to estimate this parameter using whatever test data are available--perhaps data from initial acceptance testing of the instruments and data from the first few weeks of actual operations. Accurate estimation requires a substantial amount of data, and ultimately several months of experience may be required before a satisfactory estimate is finally arrived at. During this period, the Shewhart control chart for the standard deviation will be a useful tool that can reveal whether the correct estimate is overly optimistic (when data points consistently lie above the mean, that is, above 1 on the standardized chart of Fig. 7) or pessimistic (when data points are below 1).

Once the estimate of the standard deviation appears to be correct and stable, it should not be altered without good reason. It should not, in particular, be necessary to re-estimate σ (or τ) when the instrument is recalibrated or when an internal source is replaced. An apparent change in σ should be examined as carefully as an apparent shift in the mean to see whether it could be a result of changes in operating procedures or of instrument malfunction.

2. Tests for Normality and Outliers. These are applicable primarily in conjunction with the estimation of the historical standard deviation and will probably not be applied regularly, although they may prove useful for trouble-shooting when one of the regularly scheduled tests indicates problems.

For the estimation of the historical standard deviation, a reasonably "clean" and approximately normally distributed set of observations is required. Extreme observations for which some explanation can be found should be deleted. For example, discrepant observations obtained under startup conditions are unlikely to be repeated. On the other hand, outliers for which there is no apparent explanation may in fact reflect the normal range of operating conditions and probably should be retained.

Significant deviations from normality will result in false-alarm rates and detection probabilities for statistical tests for bias, etc., which are larger or smaller than stated. (In this case, a transformation of the raw observations that brings the transformed observations closer to normality should be sought; commonly, for example, a logarithmic transformation is indicated for skewed observations.) The historical standard deviation should then be computed using the transformed observations, and all subsequent tests should be carried out on transformed data relative to the transformed mean with this standard deviation.

3. Shewhart Control Charts. Control charts for bias and for the standard deviation display 12 weeks of data, which is enough to reveal visual trends in the data even when they have not yet progressed far enough to generate an alarm. The most recent set of data should be examined on approximately a monthly basis.

4. Page's Test. A more timely indication of instrumental or procedural problems is provided by Page's test, but only if it is run frequently. Weekly application of this test to each sequence of MC observations is recommended. The output is simply pass/no pass, which does not require interpretation. A failure, of course, should be followed up by generation of corresponding control charts and further investigation.

5. Test for Randomness. Finally, the test for randomness should be applied periodically, perhaps once per quarter, as it may reveal some problem

that has escaped the inspector of the more frequent control charts. It should also be used on data for the estimation of the historical standard deviation.

B. Areas for Future Development

All of the suggestions below have been mentioned or implied elsewhere in this report, but they are gathered together here for convenience.

1. Multiple Standard. As mentioned in the introduction to Sec. III, it may well be appropriate to use more than one standard for some instruments to verify their accuracy in different parts of their range. Associated with each standard is its own historical standard deviation, as there is no reason to suppose that this will be constant over the range of the instrument. Thus, each standard used generates a completely independent sequence of MC observations. The option of using two standards is currently incorporated in the software.

2. Multiple Measurements. This refers to the possibility of making two or more measurements of one standard every time a MC observation is to be taken, not to the occasional duplication of a measurement. As suggested in Sec. III.B.3, this would enable faster detection of small shifts in the mean in cases in which the standard deviation of the process is so large that the shift detectable by a single measurement is deemed unacceptably large. The option of treating measurements in pairs or triples is currently programmed.

3. Automation of Page's Test. In theory, it would be possible to update the two Page statistics m(t) and M(t) for a sequence every time a new observation is received, without its being specifically requested. This would make optimal use of Page's test but is not currently programmed.

4. Automation of Transformations. If nonnormality of the observations indicates the desirability of applying a transformation to each observation, this transformation will have to be programmed, probably in the computer for the specific instrument.

5. Comparison with Results of Destructive Analysis. Although it lies outside the framework provided here, another check on the accuracy of an NDA

instrument is provided by comparing its measurement with the result of destructive analysis. For this, some estimate of the measurement error associated with each measurement is required. The historical standard deviation associated with the MC process provides such an estimate for the NDA measurement. (If this is known to vary over the range of the instrument, an appropriate interpolation between the standard deviations associated with two standards should be used.) For chemical analysis, Los Alamos experience indicates a standard deviation of about 0.2% can be obtained. (Locally relevant experience should be used to estimate this measurement error if at all possible, taking into account standard local procedures. For example, when the substance being measured is a powder that tends to absorb moisture, the error may be inflated if the moisture is not driven off before analysis.) Thus, the error of destructive assay is almost negligible by comparison with that of NDA measurement, and it may suffice to consider the destructive assay result as "truth" when the two measurement results are compared.

ACKNOWLEDGMENTS

We wish to thank L. Bruckner for his suggestions on the statistical portion of this work. We also wish to thank M. Krick, K. MacMurdo (SRP), W. Ruhter (Lawrence Livermore National Laboratory), and J. Wetzel (Mound Laboratories) for their contributions and comments.

REFERENCES

1. M. Baker, "An Integrated Nondestructive Assay System for a New Plutonium Scrap Recovery Facility," ESARDA 7th Symposium on Safeguards and Nuclear Material Management, Liege, Belgium (1985).

2. R. J. Brouns, F. P. Roberts, J. A. Merrill, and W. B. Brown, "A Measurement Control Program for Nuclear Material Accounting," Nuclear Regulatory Commission report NUREG/CR-0829, PNL-302 (June 1980).

3. W. R. Severe and C. C. Thomas, Jr., "Measurement Control Program for In-Line NDA Instruments," 20th Annual Meeting of the Institute of Nuclear Materials Management, Albuquerque, New Mexico, July 16-18, 1979; Nuclear Materials VIII, 620-633 (1979).

4. J. J. Malanify and D. C. Amsden, "Implementation of the DYMAC System at the New Los Alamos Plutonium Processing Facility - Phase II Report," Los Alamos National Laboratory report LA-9418 (August 1982).

5. S. -T. Hsue and T. Marks, "Measurement Control Program for NDA Instruments," Proc. ANS/INMM Conf. Safeguards Technol.: Process Safeguards Interface, Hilton Head Island, South Carolina, November 28-December 2, 1983 (US DOE/New Brunswick Laboratory, August 1984), CONF-831106, pp. 422-425.

6. A. S. Goldman and R. S. Marshall, "An Assessment of the Measurement Control Program for Balances at the LANL Plutonium Facility," Los Alamos National Laboratory report LA-9975-MS (December 1983).

7. A. Goldman, "Measurement Control Progress for TNCCs," Los Alamos National Laboratory document Q-4/85-304 (June 1985).

8. L. Bruckner, Los Alamos National Laboratory memoranda S-1/84-123 (February 27, 1984), S-1/84-228 (April 18, 1984), and S-1/84-190 (April 2, 1984).

9. S. W. Roberts, "Properties of Control Chart Zone Tests," Bell Syst. Tech. J. 37, 83-114 (1958).

10. E. S. Page, "Continuous Inspection Schemes," Biometrika 41, 100-115 (1954).

11. W. D. Ewan and K. W. Kemp, "Sampling Inspection of Continuous Processes With No Autocorrelation Between Successive Results," Biometrika 47, 363-380 (1960).

12. S. S. Shapiro and M. B. Wilks, "An Analysis of Variance Test for Normality (Complete Samples)," Biometrika 52, 591-611 (1965).

13. F. E. Grubbs, "Procedures for Detecting Outlying Observations in Samples," Technometrics 11, 1-21 (1969).

14. B. Rosner, "On the Detection of Many Outliers," Technometrics 17, 221-227 (1975).

15. C. Bingham and L. S. Nelson, "An Approximation for the Distribution of the von Neumann Ratio," Technometrics 23, 285-288 (1981).

16. B. I. Hart, "Significance Levels for the Ratio of the Mean Square Successive Difference to the Variance," Ann. Math. Stat. 12, 445-447 (1942).