COLLABORATIVE STUDY

of

REFERENCE METHOD FOR DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)

Herbert C. McKee
Ralph E. Childers
Oscar Saenz, Jr.

Contract CPA 70-40
SwRI Project 21-2811

Prepared for
Office of Measurement Standardization
Division of Chemistry and Physics
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, N. C. 27709

September 1971

SOUTHWEST RESEARCH INSTITUTE
SAN ANTONIO CORPUS CHRISTI HOUSTON

COLLABORATIVE STUDY

of

REFERENCE METHOD FOR DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)

Herbert C. McKee
Ralph E. Childers
Oscar Saenz, Jr.

Contract CPA 70-40
SwRI Project 21-2811

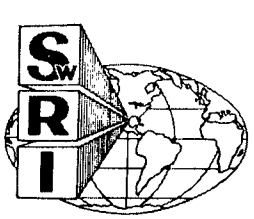
Prepared for
Office of Measurement Standardization
Division of Chemistry and Physics
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, N. C. 27709

September 1971

Approved:

Herbert C. McKee

Herbert C. McKee
Assistant Director
Department of Chemistry
and Chemical Engineering



SOUTHWEST RESEARCH INSTITUTE - HOUSTON
3600 SOUTH YOAKUM BOULEVARD, HOUSTON, TEXAS 77006

SUMMARY AND CONCLUSIONS

This report presents information obtained in the evaluation and collaborative testing of a reference method for measuring the sulfur dioxide content of the atmosphere. Different variations of this method have been used extensively by many laboratories since the original publication in 1956, and it has been found to be reliable and reasonably free of interferences.

This method was recommended as a tentative standard method by the Intersociety Committee, a cooperative group now consisting of representatives of nine scientific and engineering societies.* It was published as Tentative Method 42401-01-69T in **Health Laboratory Science**, January 1970, Part Two, pp 4-12. It was then tested, as a part of this program, by means of collaborative tests involving a total of eighteen laboratories.

A statistical analysis of the data of fourteen laboratories provided the following results, based on the analysis of pure synthetic atmospheres using the 30-min sampling procedure and the sulfite calibration method prescribed.

- The standard deviation for replication varies linearly with concentration from 7 $\mu g/m^3$ at zero to 17 $\mu g/m^3$ at 1000 $\mu g/m^3$
- The standard deviation for within-laboratory variation (repeatability) varies linearly with concentration from 15 $\mu g/m^3$ at zero to 36 $\mu g/m^3$ at 1000 $\mu g/m^3$
- The standard deviation for between-laboratory variation (reproducibility) varies linearly with concentration from 29 $\mu g/m^3$ at zero to 70 $\mu g/m^3$ at 1000 $\mu g/m^3$.
- No systematic error, bias, or inaccuracy was detected.
- The lower limit of detection is 25 $\mu g/m^3$ (95 percent confidence level).

In addition, this report presents other results with respect to the use of control samples and reagent blank samples, the minimum number of samples required to establish validity of results within stated limits, and the statistical evaluation of various steps included in the method.

These results show that the method can give satisfactory results only when followed rigorously by experienced laboratory personnel.

This method was published by the Environmental Protection Agency in the *Federal Register*, April 30, 1971, as the reference method to be used in connection with Federal ambient air quality standards for sulfur dioxide. That publication is reproduced as Appendix A of this report.

---

* Air Pollution Control Association
American Chemical Society
American Conference of Governmental Industrial Hygienists
American Industrial Hygiene Association
American Public Health Association
American Society for Testing and Materials
American Society of Civil Engineers
American Society of Mechanical Engineers
Association of Official Analytical Chemists

The Intersociety Committee receives partial financial support through EPA Contract 68-02-0004.

ACKNOWLEDGEMENT

The authors wish to express appreciation to the Project Officer, Mr. Thomas W. Stanley, and staff members of the Office of Measurement Standardization, for assistance in the planning and execution of the collaborative study. The assistance of Mr. John H. Margeson and others of the OMS staff in providing space, facilities, and training contributed significantly to the success of the Method Familiarization Session. Also acknowledged are the efforts of Mr. Clarence A. Boldt, Jr., of Southwest Research Institute, who conducted the bulk of the laboratory evaluation of the method and assisted in preparation for the Method Familiarization Session.

The assistance and cooperation of the participating laboratories is also acknowledged with sincere appreciation for the voluntary efforts of the staff members who represented each organization. The representatives and organizations participating in one or more phases of the collaborative test program were as follows:

Name	Organization
Robert M. Bethea	Texas Tech University Lubbock, Texas
James S. Caldwell	Environmental Protection Agency Cincinnati, Ohio
Gary Carlson	St. Louis County Health Department Clayton, Missouri
Emil R. deVera	Air & Industrial Hygiene Laboratory California Department of Public Health Berkeley, California
B. I. Ferber	Bureau of Mines United States Department of the Interior Pittsburgh, Pennsylvania
Harriet Klinger	Gary Air Pollution Control Division Gary, Indiana
W. D. Langley	Texas A&M University College Station, Texas
Harold E. Meyer	Galveston County Air Control Division Texas City, Texas
M. Rodney Midgett	Environmental Protection Agency Research Triangle Park, North Carolina

Name	Organization
Gordon D. Nifong	Bethlehem Steel Corporation Bethlehem, Pennsylvania
Walter Oyung	Bay Area Air Pollution Control District San Francisco, California
Robert E. Pattison	Air Pollution Control Laboratory Canton City Health Department Canton, Ohio
Rolf A. Paulson	Institute for Materials Research National Bureau of Standards Washington, D.C.
M. J. Rohlinger	Air Pollution Control Laboratory State of Illinois Springfield, Illinois
Karl Schoenemann	Air Pollution Chemistry Laboratories Los Angeles County Air Pollution Control District Los Angeles, California
F. Wopat, Jr.	Shell Oil Company Wood River Refinery Wood River, Illinois
Sandra Wroblewski	Department of Air Pollution Control City of Chicago Chicago, Illinois
Karl J. Zobel	Environmental Protection Agency Research Triangle Park, North Carolina

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

Sulfur dioxide is one of the more common atmospheric pollutants which result from the activities of man. Many urban areas throughout the world experience some degree of pollution from this contaminant due to the burning of sulfur-containing fuels, with less prevalent but occasionally severe problems due to emissions from industrial and other sources. In a few limited areas, volcanoes or sulfur springs add a natural source to the many man-made sources which exist. Sulfur dioxide in sufficiently high concentrations can be objectionable in many ways, including adverse effects on human health, damage to vegetation, corrosion of metals and other materials, and formation of haze which restricts visibility. During short-term episodes involving high concentrations, detrimental effects can occur in a few hours. At lower levels, long-range chronic effects can also occur over long periods of time, although the documentation of these is more difficult.

Because of the many different adverse effects which can occur, sulfur dioxide has traditionally received a major share of attention as an atmospheric pollutant. For this reason, methods to measure sulfur dioxide concentrations in the atmosphere have been known for many years, although some of the methods used in the past possessed serious disadvantages. Perhaps the chemical method most widely used in the past decade or more has been the pararosaniline method, also known as the West-Gaeke method from the original publication.[1]* The method is essentially a colorimetric technique, in which sulfur dioxide is removed from the air by absorption in a liquid solution, followed by reaction with pararosaniline dye to form a color proportionate to the amount of sulfur dioxide present, after which the color is then measured in a conventional laboratory spectrophotometer. As with most colorimetric methods, results are compared with a calibration curve developed from a chemically pure standard material in order to obtain quantitative results.

The pararosaniline method can be used by any laboratory equipped for conventional colorimetric analysis by merely adding the absorbers necessary for atmospheric sampling and a few other items of equipment. As with other colorimetric methods, careful, precise laboratory technique is required if accurate results are to be obtained, but this method is no more difficult to carry out than many other colorimetric methods that are widely used for a variety of purposes.

Since the original publication, many research investigations have been conducted to develop variations of this method aimed at minimizing interferences, increasing accuracy, and in other ways improving on the basic method. Because of these many investigations, a number of different variations of the original method have been published. Most of these vary in only minor details, such as the method used in purifying the reagent dye, the method of plotting calibration curves, and similar details. Except for various minor effects on precision, most of the differences do not exert a major effect on the results if the same procedure is used for preparing the calibration curve and for analyzing samples. However, to obtain the degree of precision which is possible, all details of the method as published in Appendix A of this report should be adhered to rigidly.

In order to obtain comparable data so that interlaboratory comparisons of results would be possible, the Office of Measurement Standardization (OMS) has been working for some time to develop standard methods which could be used by all persons making air quality measurements. A number of scientific and engineering societies have also been active in the development of standard methods, including several of those now participating in the Intersociety Committee whose members are listed in the Summary and Conclusions.

Following the development of a tentative standard method by the Intersociety Committee, the final

*Superscript numbers in parentheses refer to the List of References.

step in the standardization process is to conduct a collaborative test, or interlaboratory comparison, of the proposed standard method. This procedure, also called "round-robin testing," has been used to evaluate many different methods of measurement in such diverse fields as water chemistry, metallurgy, paint and surface coatings, food and related products, and many others. A test of this nature by a representative group of laboratories is the only way that the statistical limits of error inherent in any method can be determined with sufficient confidence. This report presents the results of a series of collaborative tests of the pararosaniline method conducted by Southwest Research Institute and the Office of Measurement Standardization, together with the statistical analysis of the data obtained. In planning for the collaborative test, it was also necessary to develop methods for generating test atmospheres so that each laboratory participating in the collaborative test could have an assured test atmosphere for experimental purposes. The information obtained in the development of these procedures is also presented as background information relating to the collaborative test program and as information helpful in understanding the capabilities and limitations of this standard method.

II. COLLABORATIVE TESTING OF THE METHOD

An important step in the standardization of any method of measurement is the collaborative testing of a proposed method to determine, on a statistical basis, the limits of error which can be expected when the method is used by a typical group of investigators. The collaborative, or interlaboratory, test of a method is an indispensable part[2] of the development and standardization of an analytical procedure to insure that (1) the procedure is clear and complete, and that (2) the procedure does give results with precision and accuracy in accord with those claimed for the method. Among other organizations, the Association of Official Analytical Chemists (AOAC) and the American Society for Testing and Materials (ASTM) have been active in the field of collaborative testing and have published guidelines of the proper procedure for conducting collaborative tests and evaluating the data obtained.[3-5] Publications of both of these organizations were used extensively in planning and conducting the collaborative tests of this method to measure sulfur dioxide.

After the development of techniques for generating test atmospheres, a detailed collaborative test was undertaken to obtain the necessary data to make a statistical evaluation of the method. This section of the report describes the various phases of the test plan that was developed.

A. Generation of Test Atmospheres

In order to facilitate interlaboratory comparison of results, a method must be evaluated by a collaborative test in which each of the various participants works in his own laboratory. Therefore, it was necessary to develop a procedure whereby each participant could generate a standard test atmosphere in his own laboratory for use in collaborative testing. Several methods are available for doing this, including dilution of cylinder gases into plastic bags, successive dilution stages using purified air, and others. Fortunately, the recent development of calibrated permeation tubes provided a more accurate and reproducible method of generating test atmospheres, and this method was chosen for the investigations reported here. A major advance in this field is the recent availability of certified permeation tubes for sulfur dioxide from the National Bureau of Standards.[6] By using these certified tubes, and controlling all experimental conditions which influence the rate of permeation, each laboratory could be assured of an accurate primary standard to use in evaluating the test method.

The permeation tubes used consist of a small cylindrical tube of Teflon containing liquid sulfur dioxide. The rate of diffusion of sulfur dioxide through the walls of the cylinder depends only on temperature and is reproducible within a reasonable temperature range. The certification available from the National Bureau of Standards covered the range

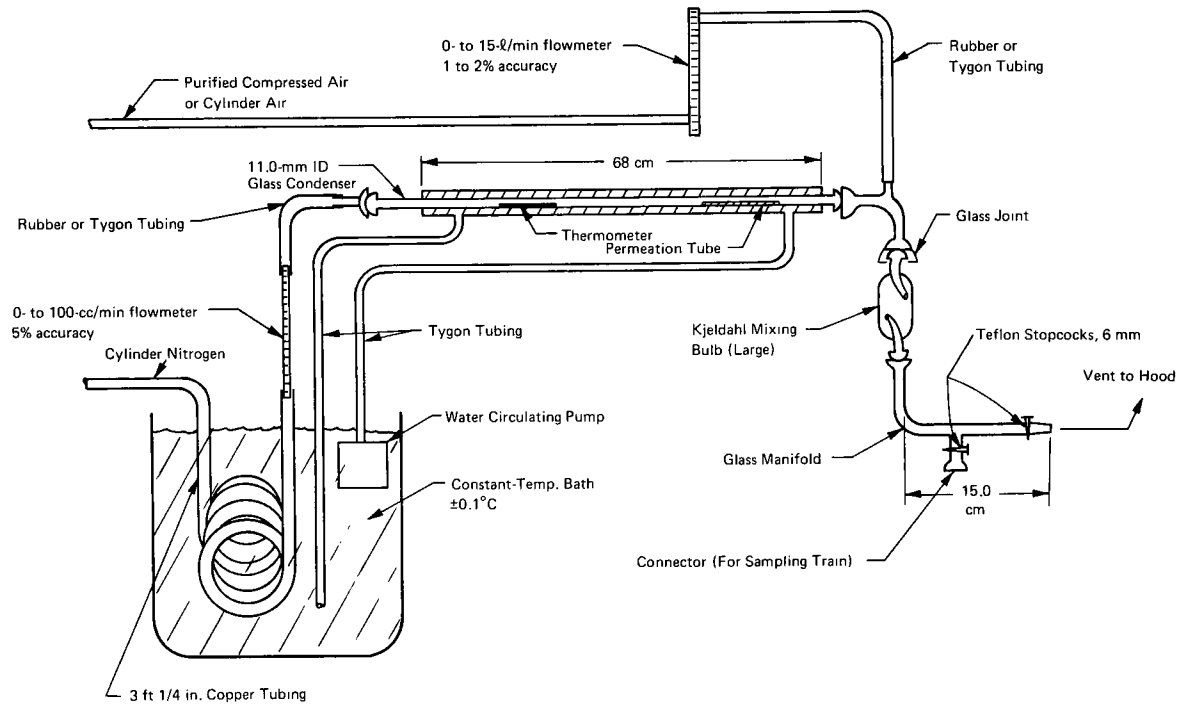


FIGURE 1. SPECIFICATIONS FOR PERMEATION TUBE SYSTEM USED IN COLLABORATIVE TESTS

of 20° to 30°C and provided sufficient accuracy if temperature control to within 0.1°C was maintained.

If the rate of permeation is controlled accurately through controlling the temperature, the only other variable controlling the concentration of the test atmosphere is flow rate. By passing air through the permeation tube apparatus at a controlled flow rate, and thus diluting the sulfur dioxide which passed through the walls of the tube by diffusion, the concentration of sulfur dioxide in the final air stream could be accurately controlled. A special apparatus was developed for this purpose which is illustrated in Figure 1. Major portions of this system were fabricated from pyrex glass, and temperature control was achieved by enclosing the permeation tube holder in a water jacket supplied by circulating water controlled to within 0.1°C. Purified air used for dilution was measured accurately with calibrated rotameters.

The apparatus consisted primarily of a condenser capable of accommodating a permeation tube and a 0.1°C thermometer, a large Kjeldahl trap to be used as a mixing bulb, and a manifold with Teflon stopcocks for sampling. The glassware is connected by ground-glass ball joints. Associated parts for the system include a calibrated flowmeter covering the range of 0 to 100 cc/min with an accuracy of 5 percent, a flowmeter covering the range of 0 to 15 ℓ/min with an accuracy of 1 to 2 percent, a 0.1°C thermometer, and a constant-temperature bath equipped with a circulating pump to continuously supply water to the condenser. The bath must be capable of maintaining the temperature within ±0.1°C. Cylinder air or compressed air, purified by carbon filters and driers (e.g., silica gel, molecular sieve), and cylinder nitrogen are required to complete the system.

A sulfur dioxide permeation tube obtained from the National Bureau of Standards was inserted into the condenser and the system assembled as shown in Figure 1. Nitrogen was passed continuously through the condenser housing the permeation tube

and the 0.1°C thermometer at a rate of 50 cc/min. It is advisable to maintain this flow through the system continuously in order to avoid sulfur dioxide accumulation in the condenser tube. The temperature in the system was adjusted to the desired temperature (usually 25.0°C). After the permeation tube had been equilibrated 24 hr, the dilution air was introduced into the system and the flow adjusted to produce the desired test atmosphere. Up to one-half of the total flow of the system may be sampled. The concentration of sulfur dioxide in the standard atmosphere generated was calculated according to the formula found in Section 8.2.2.2 of the method (see Appendix A). In order to conserve dilution air, it was shut off at the end of a sampling day; however, the constant-temperature bath and purge nitrogen gas were normally left on.

Following the development of this system, the permeation tube holder and other components were fabricated and the necessary equipment for a complete system was supplied to each participant in the collaborative test series for use in his own laboratory. Complete instructions were also supplied for using this system to generate test atmospheres.

B. Selection of Collaborators

If a collaborative test is to achieve the desired objectives, it is necessary that the participants in the test be representative of the large group that will ultimately make use of the method being tested. Since air pollution measurements are of interest to many different groups, it was desirable to include in the group of collaborators a variety of governmental agencies, universities, industrial laboratories, and others. The final selection of participants in all of the testing which was performed included five participants from federal laboratories, nine from state and local air pollution control agencies, two from industry, and two from universities. A complete list of the participants and their affiliation is given elsewhere in this report.

Even more important than the type of laboratory is the degree of skill of the persons who participated. Each laboratory was asked to assign a person to this test who had previous experience with the pararosaniline method and was competent in carrying out determinations by this method. This was done to avoid errors and greater variation in results which might be produced by a group of inexperienced workers. Each laboratory had previous experience in the use of the method and thus possessed the necessary equipment for collection of samples, preparation and standardization of solutions, and analysis by the colorimetric method used.

C. First Collaborative Test

For the first collaborative test, permeation tubes were furnished to seventeen collaborating laboratories, together with the special holders, calibrated flowmeters, and other equipment described previously to be used to generate test atmospheres. One permeation tube was furnished for use in familiarization runs plus two tubes to be used for test purposes. Nominal test concentrations used were 130 and 780 $\mu g/m^3$ (approximately 0.04 and 0.3 ppm).

Participants were instructed to set up the equipment for generation of test atmospheres, complete a number of familiarization runs, and then analyze samples obtained with the two tubes furnished for test purposes. When the results were received, it was found that many of the participants had deviated in one way or another from the desired procedure and it was necessary to normalize the data and recalculate results to produce a set of "adjusted" data. If very rigorous statistical and procedural requirements had been adopted, it would have been necessary to discard data obtained from fifteen of the sixteen laboratories completing the test and reporting results. In order to salvage as much useful information as possible from this test series, the adjustment of data was used and included such changes as the following:

(1) Elimination of calibration points higher than 24 μg of sulfur dioxide (see method, Appendix A, Section 2.3), eight laboratories.

(2) Insertion of a zero-zero point on the calibration curve for those laboratories (nine) which plotted net absorbance and failed to recognize (all nine) that a zero-zero point exists.

After adjusting the data, only about one-half of the participants (nine out of seventeen) submitted results that could be used for statistical analysis. The remainder included laboratories which did not sample at the specified rate, some who used the wrong temperature for the permeation system, or those which were rejected on statistical grounds as outliers.

Following this adjustment and recalculation of the data, a statistical analysis was performed on the remaining subset of data from nine laboratories. The results of this statistical analysis appeared to be straightforward and reasonable, in view of the expected results and the previously published information concerning the precision and reliability of the method. However, it was felt that any test in which almost one-half of the data could not be used was hardly valid to provide a rigorous statistical analysis concerning the capabilities of the method as it was intended to be used. Therefore, it was concluded from this test that the method is inherently satisfactory but is extremely vulnerable to misinterpretation, as shown by the large number of participants that failed to follow the method as specified and by the large number of outliers. Therefore, additional work was undertaken to provide a better test of the tentative method, and these steps are described subsequently.

D. Familiarization Session and Second Collaborative Test

In order to be sure that all participants followed all details precisely in generating test atmospheres and in carrying out the analysis by the tentative method, a familiarization session was held to review the entire procedure with the participants. The session was held at the EPA training laboratory in Durham, North Carolina, over a 3-day period. All details of the method and test procedure were reviewed, and each participant conducted all laboratory operations in the training laboratory. Data from these test runs were reviewed qualitatively, but no detailed statistical analysis of the data was undertaken. However, the results did appear to be consistent with previous test results.

Following the familiarization session, the participants were again given permeation tubes and instructions for a complete collaborative test series, with each again working in his own laboratory. The value of this familiarization session is shown by the fact that all participants in this test series returned usable data, in distinct contrast to the original series in which mistakes in procedure invalidated much of the data obtained. The results of this test series were then used for detailed statistical analysis, and this series is considered to provide the best evaluation available of the method being tested.

III. STATISTICAL DESIGN AND ANALYSIS

Several fundamental requirements must be met in order to provide the maximum reliability of the collaborative test. First, the conditions of the test must be representative of a specified population; each factor involved must be a representative sample of a population about which inferences are to be drawn. Second, the collaborative test must be unbiased; precautions must be taken to avoid the introduction of any bias in the collaborative test procedure. It is important that the collaborators assume a responsibility to try to eliminate any bias by carefully following the instructions of the collaborative procedure and the method. Every detail is important and even the slightest departure from the specified procedures may bias the results. Third, the results of the collaborative test must be reproducible; that is, the conditions for the test should be such that similar results would be obtained if the collaborative test were repeated. The fourth requirement involves the scope of the test; the materials and conditions for which the analytical method was designed must be included in the test. Finally, the collaborative test must be practical and economically feasible. Since funds and facilities are never available for an unlimited testing program, it is

necessary to accept less than the ideal testing procedures in order to accomplish the program. Thus, fundamental requirements may not be completely fulfilled, since any practical compromise introduces limitations on the inferences that can be drawn. If pursued too far, compromises from practical considerations may render the collaborative test useless.

The conditions for the test were very carefully prescribed. The method was tested for 30-min sampling using the calibration procedure with sulfite solution (see Sections 7 and 8, respectively, of the method in Appendix A). The collaborative test procedure required that a separate calibration curve be prepared for each separate day. In addition, certain reagents were specified to be made fresh each day. Stability was not the only criterion, since it was desired to include the variation associated with these operations in the between-days within-laboratories variation (repeatability). These specific reagents are identified below followed by the respective section number in the method (see Appendix A). The reagents prepared fresh were as follows: sulfamic acid (6.2.1), formaldehyde (6.2.2), standardized sulfite solution for preparation of working sulfite-TCM solution (6.2.8), and working sulfite-TCM solution (6.2.9). A sufficient quantity of pararosaniline solution conforming to the specifications of the method (see Section 6.2.10.2 of the method in Appendix A) was supplied to each collaborator in the test.

Appendix B contains the complete and detailed description of the design and analysis of the formal collaborative test which followed the Method Familiarization Session. The results of Appendix B are summarized in this section.

The primary purpose was to establish the reliability of the method in terms of its systematic variation, precision, and accuracy. More emphasis was placed on the quality of the method when properly used than upon the performance of the laboratories. At the same time, it was necessary to retrieve information which would allow the investigation of various steps within the method; therefore, intermediate data were obtained relating to calibration curves, control samples, and blanks.

The statistical planning of a program is limited in scope and depends upon what information is desired. The scope is limited by what a collaborating laboratory can conveniently and economically accomplish as well as by the number of collaborators that can be accommodated. Under these limitations, it was possible to examine the effects of laboratories, concentrations, and days upon the precision of the method in addition to estimating the replication error. The main experiment, as well as secondary experiments, was designed so that the analysis of variance technique could be used.

Fourteen laboratories took part in the final test program. An analyst representing each laboratory attended the Method Familiarization Session and subsequently conducted the formal collaborative testing. These individuals and their affiliations have been identified elsewhere in this report. These laboratories constitute a random sample of a rather large population of experienced laboratories. Three different concentrations were analyzed by each laboratory. The concentrations were nominally 150, 275, and 820 $\mu g/m^3$. These concentrations, in addition to having practical significance, were selected to approximate the low range, the optimum range, and the high range for the method. Each of the three concentrations was analyzed in triplicate on each of three separate days using independently prepared reagents, standards, and calibration curves. This resulted in a total of 378 individual determinations.

The collaborative test was designed to allow the analysis of the results using the most efficient statistical methods available. The form of the analysis depends upon the statistical model under consideration. The experiment was designed so that the data could be analyzed according to three different models or techniques which are listed and described in Appendix B. In addition to providing a comparison of the techniques, this approach also provided the opportunity to use the technique the results of which were most convenient to apply. It was shown that excellent agreement was obtained, and that an analysis of variance with the data for all concentrations analyzed together (with data transformation)

provided the most convenient and useful application of the results.

Supplementary or secondary experiments were incorporated to evaluate errors associated with various steps within the method. Each of these supplementary experiments was designed so that the data could be analyzed by the analysis of variance to determine whether variations between days and between laboratories were significant. The results of each of these experiments will be discussed below.

A. Outlying Observations

In accordance with the experiment design, the full plan was carried out satisfactorily and without any missing data. Two cases of atypical results were present. These were dealt with in accordance with the discussion and conclusions in Appendix B. A logical substitute was made for each case to allow the remainder of the data for the laboratory to be utilized.

Since the emphasis was upon the quality of the method and not upon the performance of the laboratories, all arithmetic errors were corrected, and the arithmetic error problem was evaluated qualitatively. Only three of the fourteen laboratories submitting results exhibited any errors. Four instances of inadvertent errors in arithmetic operations were noted and corrected. The method contains complex calculational procedures and consequently is vulnerable to arithmetic and procedural errors. However, the majority of the collaborators demonstrated the capability to handle this complexity. There is no reason to believe that a careful checking procedure could not eliminate this problem.

The results, after the disposition of the two outlying observations, are believed to be an excellent data base for the statistical analyses to follow. The data are believed to be representative and unbiased. The number of outliers was small, especially in comparison with the first test where nearly two-thirds of the laboratories produced one or more atypical results, many of which could not be corrected to obtain usable data.

B. Analysis of Variance

In Appendix B, three analyses are described and the results discussed. The first two are classic analysis of variance cases. The first handles each concentration separately, while the second combines all concentration data into a single analysis, and includes the evaluation and application of the necessary data transformation to allow this treatment. The third case, the linear model analysis, does not strictly constitute a classic analysis of variance case but does involve the technique.

The comparison of methods is best made by referring to the results in graphic form shown in Figure 2. This figure compares the replication error, the repeatability*, and the reproducibility* for each method of statistical analysis. First, the overall agreement between these methods is very good. The agreement for the replication error is exact because the methods have this much in common. The point estimates seem to imply a minimum for repeatability and reproducibility in the midrange of concentration. The other methods, because of their fundamental assumptions, do not recognize any minimum. In this respect, the results are inconclusive, and an experiment incorporating many more intermediate concentrations would be required to verify such a condition. In consideration of the optimum absorbance range of most spectrophotometers, such a minimum could be entirely possible.

The point estimates are therefore of limited use

*The terms "repeatability" and "reproducibility" have been in use for many years, and it has not always been clear from the context of each publication just what is the precise definition. A very recent publication [Mandel, John, "Repeatability and Reproducibility," *Materials Research and Standards, Am. Soc. Testing & Mats.*, Vol 11, No. 8, p 8, (August 1971)] clears the confusion by giving rigid definitions. While the use of these two terms in this report is not exactly consistent with the definitions of Mandel, they are, nevertheless, well defined and are easily relatable.

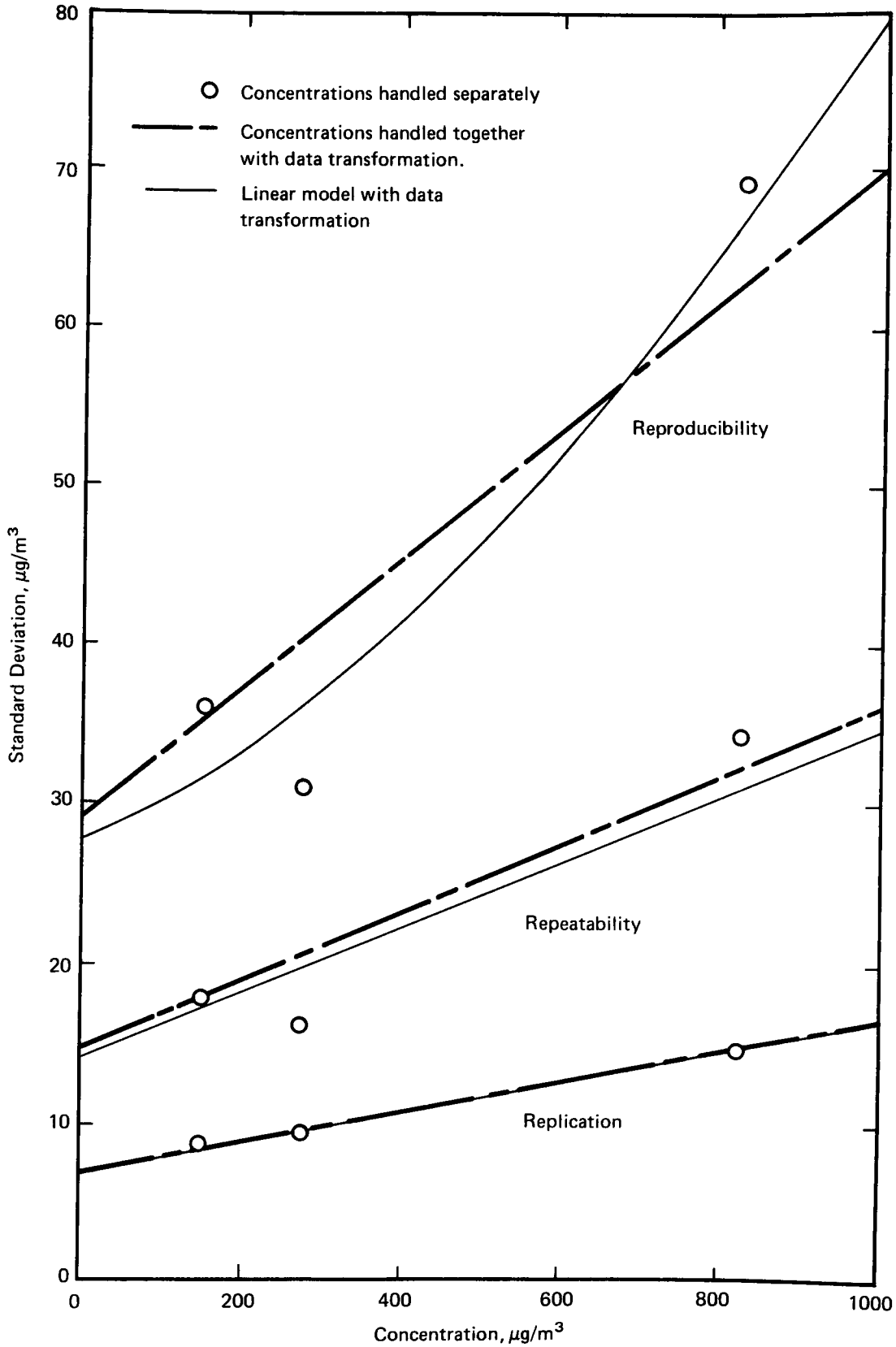


FIGURE 2. REPLICATION ERROR, REPEATABILITY, AND REPRODUCIBILITY VERSUS CONCENTRATION FOR THREE DIFFERENT METHODS OF DATA ANALYSIS.

since we must also make inferences between these points. The other two methods of analysis provide for objective estimates at other concentrations. The second method, the analysis of variance with concentrations handled together, has the advantage of a simple expression for the standard deviations as a function of concentration. The linear model expresses the standard deviations as a function of concentration; however, the function is more complex. The linear model has an advantage because of its ability to describe and compare the various sources of error more thoroughly. The linear model is less sensitive to outlying observations.

On the basis of the preceding discussion, it was concluded that the analysis of variance handling all concentrations together (with data transformation) offers the most convenient and practical method of expressing the replication error, the repeatability, and the reproducibility as a function of concentration. All subsequent treatment of these parameters will be made accordingly.

C. Various Sources of Error Within the Analytical Method

No study would be complete without at least a superficial examination of certain steps within the analytical procedure. This subsection deals with the determination of the existence and the estimation of the magnitude of some of these sources of error.

1. Calibration Curves

The experiment was designed so that complete data for each and every calibration curve were retrieved. Each individual absorbance-concentration point for each day for each laboratory was recorded. These voluminous data are not reproduced here. The purpose of this secondary experiment was to examine the deviations from linearity of the curves, to examine the distribution of the slopes of these curves, and to evaluate the variability in slope between days and between laboratories.

From these data, the slopes, the intercepts, and the standard errors of estimate were computed using the least squares technique. There were fourteen laboratories and 3 days, thus yielding a table of 42 entries for each parameter. These data are shown in Table C-III in Appendix C.

We will first examine the standard errors of estimate of these calibration curves. The overall mean standard error of estimate was found to be 0.010 absorbance unit (41 degrees of freedom). At an average slope of 0.030 absorbance unit per microgram sulfur dioxide, this corresponds to 0.33 μg of sulfur dioxide. Thus, the lower limit of detection for sulfur dioxide (see Section 7.1.1 of the method in Appendix A) at the 95 percent level of confidence is 2.02×0.33 or 0.67 μg, representing a concentration of 22 $\mu g/m^3$ for a 30-ℓ air sample. This is in excellent agreement with the 25-$\mu g/m^3$ limit claimed for the method (see Section 2.2 of the method in Appendix A). Other independent estimates of the lower detection limit will be made later.

The overall linearity of the calibration curves was not examined explicitly since the method proposes a straight line relationship (see Section 2.1 of the method in Appendix A). A qualitative examination was made resulting in the general conclusion that the upper end of the calibration curves contributed most heavily to the standard error of estimate. This does not necessarily imply nonlinearity but suggests increasing inaccuracy in absorbance readings at the higher end. These effects could be minimized by using a weighted least squares technique[7-8] in which the relative deviations are minimized rather than the absolute deviations. In view of the magnitude of the overall precision of the method, this refinement, although no more difficult or complex than the conventional least squares technique, is probably not justified.

The data for the slopes of the calibration curves were normally distributed with the possible exception of Laboratory 926, whose slopes appeared to be atypically low. These data were subjected to an analysis of variance involving two factors–labora-

tories and days. The variation between laboratories was significant (95 percent level of significance) with respect to the variation between days. The component of variance due to laboratories constituted about 80 percent of the total variance. The standard deviation of the slope for variation between days was 0.00082 absorbance unit per microgram (28 degrees of freedom) while the corresponding standard deviation for between-laboratories variation was 0.00195 absorbance unit per microgram (13 degrees of freedom). The overall mean slope was 0.0298 absorbance unit per microgram. The 95 percent confidence interval was therefore 0.0298 ± 0.0017 for within-laboratories and 0.0298 ± 0.0042 for between-laboratories. The overall mean is in excellent agreement with that claimed for the method, and the within-laboratory confidence interval corresponds almost exactly to the claim for the method (see Section 6.2.10.1 of the method in Appendix A). The important observation is that the variation between laboratories is approximately 2.5 times as much as the variation within laboratories, which partially accounts for the relatively poor interlaboratory precision demonstrated previously.

These calibration curves were further investigated with respect to the absorbance-axis intercept (see Table C-III in Appendix C). The absolute magnitude of the intercept as well as its respective deviation from the zero-standard absorbance was investigated. The variation between laboratories was significant (95 percent level of significance) with respect to the variation between days for the absolute magnitude of the intercept. The component of variance due to laboratories contributed two-thirds of the total variance. The standard deviation of the intercept, for variation between days, was 0.013 absorbance unit (28 degrees of freedom), while the corresponding standard deviation for variation between laboratories was 0.022 absorbance unit (13 degrees of freedom). The overall mean was 0.163 absorbance unit. The 95 percent confidence intervals were therefore 0.163 ± 0.026 for within-laboratories and 0.163 ± 0.048 for between-laboratories. Since the reagent blank is temperature sensitive, it is not surprising to find significant interlaboratory variation. The overall mean of 0.163 is comparable to the value suggested as a guide by the method (see Section 6.2.10.1 of the method in Appendix A); however, because of temperature effects, no rigid comparison is valid.

The deviations of the zero-absorbance standards from the intercepts of the calibration curves were analyzed, and the effects due to laboratories were found to be not significant (95 percent level of significance). The overall mean was not significantly different from zero. The standard deviation (pooled estimate with 41 degrees of freedom) was 0.006 absorbance unit corresponding to a 95 percent confidence interval of ±0.011 absorbance unit. This is well below a value of 0.03 specified in the method (see Section 8.2.1 of the method in Appendix A).

2. Control Samples

The method prescribes that a control sample, consisting of an aliquot of standard sulfite solution, is to be included with each set of determinations (see Section 7.2.2 of the method in Appendix A). The results of all of these control samples were analyzed.

Some laboratories ran more control samples than others and one laboratory (788) did not run any. A set of data was constructed consisting of three randomly selected control samples for each of twelve laboratories. Most laboratories had analyzed three control samples (one each day), except Laboratory 578 which analyzed only one. Consequently, Laboratories 578 and 788 were not included in this analysis.

The data consisted of a table of 36 values (twelve laboratories × 3 days), which were the result of subtracting the amount taken from the amount found, both in micrograms of sulfur dioxide. These data are shown in Table C-IV in Appendix C. All data were of the approximate same order of magnitude, allowing this approach. The differences were normally distributed.

The analysis of variance found the between-laboratory variability to be significant (95 percent level of significance) with respect to the within-laboratory variability. The between-laboratory variance accounted for 42 percent of the total variance while the within-laboratory variance accounted for the remaining 58 percent. The standard deviation for within-laboratory variation was 0.4 μg (24 degrees of freedom), and the standard deviation for between-laboratory variation was 0.5 μg (11 degrees of freedom). The overall mean was found to be insignificantly different from zero. The 95 percent confidence interval for within-laboratory variation was therefore ±0.77 μg, and the corresponding interval for between-laboratory variation was ±1.08 μg. In terms of concentration, for a 30-ℓ air sample, these values become 26 $\mu g/m^3$ and 36 $\mu g/m^3$, respectively. The first figure is another independent estimate of the lower limit of detection and also verifies the claim for the method (see Section 2.1 of the method in Appendix A).

The method (see Section 7.2.2 of the method in Appendix A) specifies that a control sample be run with each set of determinations, but does not set any specifications regarding the results obtained. If the difference (in micrograms of sulfur dioxide) between the amount taken and the amount found exceeds 0.8 μg, either the control sample or the calibration curve is suspect and should be checked accordingly.

3. Reagent Blanks

The method prescribes that a reagent blank is to be included with each set of determinations (see Section 7.2.2 of the method in Appendix A). An analysis of all reagent blanks run along with samples was made, and the differences between the blank and the intercept of the calibration curve were analyzed. All data were normally distributed, except that from Laboratory 788 which contained an unusually high *and* an unusually low difference. These data are shown in Table C-V in Appendix C. In this case, a graphic analysis was made by plotting the results on normal probability graph paper. All results, including those of Laboratory 788, were within ±0.04 absorbance unit, and 95 percent of the results were within ±0.03 absorbance unit. The overall mean was zero. These results conform to the criteria set forth in the method (see Section 7.2.2 of the method in Appendix A) and also with the results from Section III-C-1 above which compared the intercept of the calibration curve with the zero-absorbance standard (reagent blank).

D. Application of the Results

A more detailed application of results is given in Appendix B. In this subsection, the various measures of precision are summarized. Unless otherwise stated below, a 95 percent level of significance is assumed. The results apply for 30-min sampling and the use of the calibration procedure with sulfite solution.

The expressions for the replication error (σ_ϵ), the within-laboratory (single-replicate, single-analyst) variation (repeatability) (σ_D), and the between-laboratory (single-replicate, single-day, single-analyst) variation (σ_L) from Figure 1 are restated as follows:

$$\sigma_\epsilon = (0.7 + 0.001y)\,(10)$$

$$\sigma_D = (0.7 + 0.001y)\,(21)$$

$$\sigma_L = (0.7 + 0.001y)\,(41)$$

where y is the concentration in $\mu g/m^3$. All statements regarding the precision of the method are derived from these expressions. With these equations, the precision for any desired case can be computed. Some of the simpler cases are shown below.

Replication will not materially assist in increasing the precision of the method, and will, in general, be a waste of time and effort. Nevertheless, a measure of acceptability of replicates should be provided. The expression for the checking limits for duplicates is

$$R_{\max} = (2.77)\,(0.7 + 0.001y)\,(10)$$

where $R_{\max}$ is the maximum permissible difference

between duplicates. Two such replicates should be considered suspect if they differ by more than R_{max}.

Agreement between duplicates better than 5 percent cannot be expected below 900 $\mu g/m^3$ Agreement better than 10 percent cannot be expected below 300 $\mu g/m^3$, and agreement better than 20 percent cannot be expected below 100 $\mu g/m^3$.

To compare two single-replicate observations made by the same analyst on the same sample on different days, the following expression is used:

$$R_{max} = (2.82)(0.7 + 0.001y)\ (21)$$

where R_{max} is the maximum permissible difference between the two results. Two such values may not be considered to belong to the same population if they differ by more than R_{max}. Conversely, the two values are not significantly different if they differ by less than R_{max}.

The method cannot detect a difference smaller than 10 percent between two observations by the same analyst in the range of 0 to 1000 $\mu g/m^3$ A difference of 20 percent or less may be detected above 300 $\mu g/m^3$, and a difference of less than 50 percent may be detected above 100 $\mu g/m^3$

As an example of the futility of replication, the factor 21 in the equation above would be reduced to 20 for duplicates, approximately the same for triplicates, and to 19 for an infinite number of replicates.

To compare two single-replicate observations made by different laboratories on the same sample, the following expression is used:

$$R_{max} = (3.06)(0.7 + 0.001y)\ (41)$$

where R_{max} is the maximum permissible difference between the observations. Two such values may not be considered to belong to the same population if they differ by more than R_{max}. Conversely, the two values are not significantly different if they differ by less than R_{max}.

The method cannot detect a difference of less than 20 percent between single-replicate observations of two laboratories in the range of 0 to 1000 $\mu g/m^3$. At a level of 100 $\mu g/m^3$, a difference of less than 100 percent is not detectable.

Various statistical methods are available for the comparison of means or the comparison of a mean and a fixed value.[9-11] These methods are straightforward and are applied independently of the results of this study. That is, whether or not a mean is significantly different from some fixed value is dependent upon the actual standard deviation of the sample population. The variance of the sample population includes both the variance of the true values and the variance due to the measurement method. A limiting case is discussed in Appendix B under the assumption that all variation is due to the measurement method. The case is an extremely unlikely, if not impossible, situation; however, a certain amount of guidance can be obtained in terms of the numbers of observations required to provide a specified degree of agreement. These numbers are sufficient only to compensate for the variation of the method. An additional quantity, dependent on the variation in the true values, will always be required. Interested readers may refer to Figures B-4 and B-5 and the respective discussions in Section B-V of Appendix B where two illustrative examples are given.

Three independent estimates of the lower limit of detection were made. These are in very good agreement, and it is not too important which one is used; therefore, a value of 25 $\mu g/m^3$ is proposed as a practical figure. A single determination less than this value is not significantly different from zero. Whether the mean of several observations, each a single determination, is significantly different from zero is dependent upon the number of observations and their distribution, regardless of the magnitude of the mean.

Recorded results of a determination using this method should carry no more than two significant digits. Originators or recorders of data should assume the responsibility of appending confidence limits (95 percent) to their data.

The overall average deviations from the expected values for each concentration tested were not significant, and therefore no systematic error, bias, or inaccuracy was detectable.

LIST OF REFERENCES

1. West, P.W., and Gaeke, G.C., "Fixation of Sulfur Dioxide as Sulfitomercurate III and Subsequent Colorimetric Determination," *Anal. Chem. 28,* pp 1816 (1956).

2. Youden, W.J., "The Collaborative Test," *Journal of the AOAC,* Vol 46, No. 1, pp 55-62 (1963).

3. **1968 Book of ASTM Standards,** Part 30, Recommended Practice for Developing Precision Data on ASTM Methods for Analysis and Testing of Industrial Chemicals, ASTM Designation: E180-67, pp 459-480.

4. **Handbook of the AOAC**, Second Edition, October 1, 1966.

5. ASTM Manual for Conducting an Interlaboratory Study of a Test Method, ASTM STP No. 335, *Am. Soc. Testing & Mats.* (1963).

6. National Bureau of Standards, "New Sulfur Dioxide Permeation Tube," *NBS Technical News Bulletin,* p 106 (April 1971).

7. Cook, Peter P., and Grady, Roger A., "Analysis of Flow Sensor Calibration Data," *Instruments and Control Systems,* Vol 44, No. 4, pp 101-102 (April 1971).

8. Southwest Research Institute, Houston, Texas, Computer Subroutine WTLSQ for Weighted Least Squares Regression Analysis, Unpublished (1971).

9. Dixon, Wilfred J., and Massey, Frank J., Jr., **Introduction to Statistical Analysis,** McGraw-Hill Book Company, Inc., New York, Chapter 9, pp 112-129 (1957).

10. Duncan, Acheson J., **Quality Control and Industrial Statistics,** Third Edition, Richard D. Irwin, Inc., Homewood, Illinois, Chapters XXV and XXVI, pp 473-521 (1965).

11. Bennett, Carl A., and Franklin, Norman L., **Statistical Analysis in Chemistry and the Chemical Industry,** John Wiley and Sons, New York, Chapter 5, pp 149-164 (1954).

APPENDIX A

REFERENCE METHOD FOR THE DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)

Reproduced from Appendix A, "National Primary and Secondary Ambient Air Quality Standards," *Federal Register*, Vol 36, No. 84, Part II, Friday, April 30, 1971.

APPENDIX A.—REFERENCE METHOD FOR THE DETERMINATION OF SULFUR DIOXIDE IN THE ATMOSPHERE (PARAROSANILINE METHOD)

1. *Principle and Applicability.* 1.1 Sulfur dioxide is absorbed from air in a solution of potassium tetrachloromercurate (TCM). A dichlorosulfitomercurate complex, which resists oxidation by the oxygen in the air, is formed (*1, 2*). Once formed, this complex is stable to strong oxidants (e.g., ozone, oxides of nitrogen). The complex is reacted with pararosaniline and formaldehyde to form intensely colored pararosaniline methyl sulfonic acid (*3*). The absorbance of the solution is measured spectrophotometrically.

1.2 The method is applicable to the measurement of sulfur dioxide in ambient air using sampling periods up to 24 hours.

2. *Range and Sensitivity.* 2.1 Concentrations of sulfur dioxide in the range of 25 to 1,050 $\mu g/m.^3$ (0.01 to 0.40 p.p.m.) can be measured under the conditions given. One can measure concentrations below 25 $\mu g./m.^3$ by sampling larger volumes of air, but only if the absorption efficiency of the particular system is first determined. Higher concentrations can be analyzed by using smaller gas samples, a larger collection volume, or a suitable aliquot of the collected sample. Beer's Law is followed through the working range from 0.03 to 1.0 absorbance units (0.8 to 27 $\mu g.$ of sulfite ion in 25 ml. final solution computed as SO_2).

2.2 The lower limit of detection of sulfur dioxide in 10 ml. TCM is 0.75 $\mu g._3$ (based on twice the standard deviation) representing a concentration of 25 $\mu g./m^3 SO_2$ (0.01 p.p.m.) in an air sample of 30 liters.

3. *Interferences.* 3.1 The effects of the principal known interferences have been minimized or eliminated. Interferences by oxides of nitrogen are eliminated by sulfamic acid (*4, 5*), ozone by time-delay (*6*), and heavy metals by EDTA (ethylenediaminetetroacetic acid, disodium salt) and phosphoric acid (*4, 6,*). At least 60 $\mu g.$ Fe (III), 10 $\mu g.$ Mn(II), and 10 $\mu g.$ Cr(III) in 10 ml. absorbing reagent can be tolerated in the procedure. No significant interference was found with 10 $\mu g.$ CU (II) and 22 $\mu g.$ V(V).

4. *Precision, Accuracy, and Stability.* 4.1 Relative standard deviation at the 95 percent confidence level is 4.6 percent for the analytical procedure using standard samples. (*5*)

4.2 After sample collection the solutions are relatively stable. At 22° C. losses of sulfur dioxide occur at the rate of 1 percent per day. When samples are stored at 5° C. for 30 days, no detectable losses of sulfur dioxide occur. The presence of EDTA enhances the stability of SO_2 in solution, and the rate of decay is independent of the concentration of SO_2. (*7*)

5. *Apparatus.*

5.1 *Sampling.*

5.1.1 *Absorber.* Absorbers normally used in air pollution sampling are acceptable for concentrations above 25 $\mu g./m.^3$ (0.01 p.p.m.). An all-glass midget impinger, as shown in Figure A1, is recommended for 30-minute and 1-hour samples.

For 24-hour sampling, assemble an absorber from the following parts:

Polypropylene 2-port tube closures, special manufacture (available from Bel-Art Products, Pequannock, N.J.).

Glass impingers, 6 mm. tubing, 6 inches long, one end drawn to small diameter such that No 79 jewelers will pass through, but No. 78 jewelers will not. (Other end fire polished.)

Polypropylene tubes, 164 by 32 mm. Nalgene or equal).

5.1.2 *Pump.* Capable of maintaining an air pressure differential greater than 0.7 atmosphere at the desired flow rate.

5.1.3 *Air Flowmeter or Critical Orifice.* A calibrated rotameter or critical orifice capable of measuring air flow within ±2 percent. For 30-minute sampling, a 22-gauge hypodermic needle 1 inch long may be used as a critical orifice to give a flow of about 1 liter/minute. For 1-hour sampling, a 23-gauge hypodermic needle five-eighths of an inch long may be used as a critical orifice to give a flow of about 0.5 liter/minute. For 24 hour sampling, a 27-gauge hypodermic needle three-eighths of an inch long may be used to give a flow of about 0.2 liter/minute. Use a membrane filter to protect the needle (Figure A1a).

6.2 *Analysis.*

5.2.1 *Spectrophotometer.* Suitable for measurement of absorbance at 548 nm. with an effective spectral band width of less than 15 nm. Reagent blank problems may occur with spectrophotometers having greater spectral band width. The wavelength calibration of the instrument should be verified. If transmittance is measured, this can be converted to absorbance:

$$A = \log_{10}(1/T)$$

6. *Reagents.*

6.1 *Sampling.*

6.1.1 *Distilled water.* Must be free from oxidants.

6.1.2 *Absorbing Reagent [0.04 M Potassium Tetrachloromercurate (TCM)].* Dissolve 10.86 g. mercuric chloride, 0.066 g. EDTA (thylenediaminetetraacetic acid, disodoum salt), and 6.0 g. potassium chloride in water and bring to mark in a 1,000-ml. volumetric flask. (Caution: highly poisonous. If spilled on skin, flush off with water immediately). The pH of this reagent should be approximately 4.0, but it has been shown that there is no appreciable difference in collection efficiency over the range of pH 5 to pH 3.(7) The absorbing reagent is normally stable for 6 months. If a precipitate forms, discard the reagent.

5.2 *Analysis.*

6.2.1 *Sulfamic Acid (0.6 percent).* Dissolve 0.6 g. sulfamic acid in 100 ml. distilled water. Prepare fresh daily.

6.2.2 *Formaldehyde (0.2 percent).* Dilute 5 ml. formaldehyde solution (36–38 percent) to 1,000 ml. with distilled water. Prepare daily.

6.2.3 *Stock Iodine Solution (0.1 N).* Place 12.7 g. iodine in a 250-ml. beaker; add 40 g. potassium iodide and 25 ml. water. Stir until all is dissolved, then dilute to 1,000 ml. with distilled water.

6.2.4 *Iodine Solution (0.01 N).* Prepare approximately 0.01 N iodine solution by diluting 50 ml. of stock solution to 500 ml. with distilled water.

6.2.5 *Starch Indicator Solution.* Triturate 0.4 g. soluble starch and 0.002 g. mercuric iodide (preservative) with a little water, and add the paste slowly to 200 ml. boiling water. Continue boiling until the solution is clear; cool, and transfer to a glass-stoppered bottle.

6.2.6 *Stock Sodium Thiosulfate Solution (0.1 N).* Prepare a stock solution by dissolving 25 g. sodium thiosulfate ($Na_2S_2O_3 \cdot 5H_2O$) in 1,000 ml. freshly boiled, cooled, distilled water and add 0.1 g. sodium carbonate to the solution. Allow the solution to stand 1 day before standardizing. To standardize, accurately weigh, to the nearest 0.1 mg., 1.5 g. primary standard potassium iodate dried at 180° C. and dilute to volume in a 500-ml. volumetric flask. To a 500-ml. iodine flask, pipet 50 ml. of iodate solution. Add 2 g. potassium iodide and 10 ml. of 1 N hydrochloric acid. Stopper the flask. After 5 minutes, titrate with stock thiosulfate solution to a pale yellow. Add 5 ml. starch indicator solution and continue the titration until the blue color disappears. Calculate the normality of the stock solution:

$$N = \frac{W}{M} \times 2.80$$

N = Normality of stock thiosulfate solution.
M = Volume of thiosulfate required, ml.
W = Weight of potassium iodate, grams.

$$2.80 = \frac{10^3 (\text{conversion of g. to mg.}) \times 0.1 (\text{fraction iodate used})}{35.67 (\text{equivalent weight of potassium iodate})}$$

6.2.7 *Sodium Thiosulfate Titrant (0.01 N).* Dilute 100 ml. of the stock thiosulfate solution to 1,000 ml. with freshly boiled distilled water.

Normality = Normality of stock solution × 0.100.

6.2.8 *Standardize Sulfite Solution for Preparation of Working Sulfite-TCM Solution.* Dissolve 0.3 g. sodium metabisulfite ($Na_2S_2O_5$) or 0.40 g. sodium sulfite (Na_2SO_3) in 500 ml. of recently boiled, cooled, distilled water. (Sulfite solution is unstable; it is therefore important to use water of the highest purity to minimize this instability.) This solution contains the equivalent of 320 to 400 $\mu g./ml.$ of SO_2. The actual concentration of the solution is determined by adding excess iodine and back-titrating with standard sodium thiosulfate solution. To back-titrate, pipet 50 ml. of the 0.01 N iodine into each of two 500-ml. iodine flasks (A and B). To flask A (blank) add 25 ml. distilled water, and to flask B (sample) pipet 25 ml. sulfite solution. Stopper the flasks and allow to react for 5 minutes. Prepare the working sulfite-TCM Solution (6.2.9) at the same time iodine solution is added to the flasks. By means of a buret containing standardized 0.01 N thiosulfate, titrate each flask in turn to a pale yellow. Then add 5 ml. starch solution and continue the titration until the blue color disappears.

6.2.9 *Working Sulfite-TCM Solution.* Pipet accurately 2 ml. of the standard solution into a 100 ml volumetric flask and bring to mark with 0.04 M TCM. Calculate the concentration of sulfur dioxide in the working solution:

$$\mu g\ SO_2/ml. = \frac{(A - B)(N)(32{,}000)}{25} 25 \times 0.02$$

A = Volume thiosulfate for blank, ml.
B = Volume thiosulfate for sample, ml.
N = Normality of thiosulfate titrant.
32,000 = Milliequivalent wt. of SO_2, $\mu g.$
25 = Volume standard sulfite solution, ml.
0.02 = Dilution factor.

This solution is stable for 30 days if kept at 5° C. (refrigerator). If not kept at 5° C., prepare daily.

6.2.10 *Purified Pararosaniline Stock Solution (0.2 percent nominal).*

6.2.10.1 *Dye Specifications.* The pararosaniline dye must meet the following performance specifications: (1) the dye must have a wavelength of maximum absorbance at 540 nm. when assayed in a buffered solution of 0.1 M sodium acetate-acetic acid; (2) the absorbance of the reagent blank, which is temperature-sensitive (0.015 absorbance unit/°C), should not exceed 0.170 absorbance unit at 22° C. with a 1-cm. optical path length, when the blank is prepared according to the prescribed analytical procedure and to the specified concentration of the dye; (3) the calibration curve (Section 8.2.1) should have a slope of 0.030±0.002 absorbance units/$\mu g.$ SO_2 at this path length when the dye is pure and the sulfite solution is

properly standardized.

6.2.10.2 *Preparation of Stock Solution.* A specially purified (99–100 percent pure) solution of pararosaniline, which meets the above specifications, is commercially available in the required 0.20 percent concentration (Harleco*). Alternatively, the dye may be purified, a stock solution prepared and then assayed according to the procedure of Scaringelli, et al. (*4*)

6.2.11 *Pararosaniline Reagent.* To a 250-ml. volumetric flask, add 20 ml. stock pararosaniline solution. Add an additional 0.2 ml. stock solution for each percent the stock assays below 100 percent. Then add 25 ml. 3 M phosphoric acid and dilute to volume with distilled water. This reagent is stable for at least 9 months.

7. *Procedure.*

7.1 *Sampling.* Procedures are described for short-term (30 minutes and 1 hour) and for long-term (24 hours) sampling. One can select different combinations of sampling rate and time to meet special needs. Sample volumes should be adjusted, so that linearity is maintained between absorbance and concentration over the dynamic range.

7.1.1 *30-Minute and 1-Hour Samplings.* Insert a midget impinger into the sampling system, Figure A1. Add 10 ml. TCM solution to the impinger. Collect sample at 1 liter/minute for 30 minutes, or at 0.5 liter/minute for 1 hour, using either a rotameter, as shown in Figure A1, or a critical orifice, as shown in Figure A1a, to control flow. Shield the absorbing reagent from direct sunlight during and after sampling by covering the impinger with aluminum foil, to prevent deterioration. Determine the volume of air sampled by multiplying the flow rate by the time in minutes and record the atmospheric pressure and temperature. Remove and stopper the impinger. If the sample must be stored for more than a day before analysis, keep it at 5° C. in a refrigerator (see 4.2).

7.1.2 *24-Hour Sampling.* Place 50 ml. TCM solution in a large absorber and collect the sample at 0.2 liter/minute for 24 hours from midnight to midnight. Make sure no entrainment of solution results with the impinger. During collection and storage protect from direct sunlight. Determine the total air volume by multiplying the air flow rate by the time in minutes. The correction of 24-hour measurements for temperature and pressure is extremely difficult and is not ordinarily done. However, the accuracy of the measurement will be improved if meaningful corrections can be applied. If storage is necessary, refrigerate at 5° C. (see 4.2).

7.2 *Analysis.*

7.2.1 *Sample Preparation.* After collection, if a precipitate is observed in the sample, remove it by centrifugation.

7.2.1.1 *30-Minute and 1-Hour Samples.* Transfer the sample quantitatively to a 25-ml. volumetric flask; use about 5 ml. distilled water for rinsing. Delay analyses for 20 minutes to allow any ozone to decompose.

7.2.1.2 *24-Hour Sample.* Dilute the entire sample to 50 ml. with absorbing solution. Pipet 5 ml. of the sample into a 25-ml. volumetric flask for chemical analyses. Bring volume to 10 ml. with absorbing reagent. Delay analyses for 20 minutes to allow any ozone to decompose.

7.2.2 *Determination.* For each set of determinations prepare a reagent blank by adding 10 ml. unexposed TCM solution to a 25-mL volumetric flask. Prepare a control solution by adding 2 ml. of working sulfite-TCM solution and 8 ml. TCM solution to a 25-ml. volumetric flask. To each flask containing either sample, control solution, or reagent blank, add 1 ml. 0.6 percent sulfamic acid and allow to react 10 minutes to destroy the nitrite from oxides of nitrogen. Accurately pipet in 2 ml. 0.2 percent formaldehyde solution, then 5 ml. pararosaniline solution. Start a laboratory timer that has been set for 30 minutes. Bring all flasks to volume with freshly boiled and cooled distilled water and mix thoroughly. After 30 minutes and before 60 minutes, determine the absorbances of the sample (denote as A), reagent blank (denote as A_o) and the control solution at 548 nm. using 1-cm. optical path length cells. Use distilled water, not the reagent blank, as the reference. (NOTE: This is important because of the color sensitivity of the reagent blank to temperature changes which can be induced in the cell compartment of a spectrophotometer.) Do not allow the colored solution to stand in the absorbance cells, because a film of dye may be deposited. Clean cells with alcohol after use. If the temperature of the determinations does not differ by more than 2° C. from the calibration temperature (8.2), the reagent blank should be within 0.03 absorbance unit of the y-intercept of the calibration curve (8.2). If the reagent blank differs by more than 0.03 absorbance unit from that found in the calibration curve, prepare a new curve.

7.2.3 *Absorbance Range.* If the absorbance of the sample solution ranges between 1.0 and 2.0, the sample can be diluted 1:1 with a portion of the reagent blank and read within a few minutes. Solutions with higher absorbance can be diluted up to sixfold with the reagent blank in order to obtain onscale readings within 10 percent of the true absorbance value.

8. *Calibration and Efficiencies.*

8.1 *Flowmeters and Hypodermic Needle.* Calibrate flowmeters and hypodermic needle (*8*) against a calibrated wet test meter.

8.2 *Calibration Curves.*

8.2.1 *Procedure with Sulfite Solution.* Accurately pipet graduated amounts of the working sulfite-TCM solution (6.2.9) (such as 0, 0.5, 1, 2, 3, and 4 ml.) into a series of 25-ml. volumetric flasks. Add sufficient TCM solution to each flask to bring the volume to approximately 10 ml. Then add the remaining reagents as described in 7.2.2. For maximum precision use a constant-temperature bath. The temperature of calibration must be maintained within ±1° C. and in the range of 20° to 30° C. The temperature of calibration and the temperature of analysis must be within 2 degrees. Plot the absorbance against the total concentration in μg. SO_2 for the corresponding solution. The total μg. SO_2 in solution equals the concentration of the standard (Section 6.2.9) in μg. SO_2/ml. times the ml. sulfite solution added (μg. SO_2= μg./ml. SO_2×ml. added). A linear relationship should be obtained, and the y-intercept should be within 0.03 absorbance unit of the zero standard absorbance. For maximum precision determine the line of best fit using regression analysis by the method of least squares. Determine the slope of the line of best fit, calculate its reciprocal and denote as B_s. B_s is the calibration factor. (See Section 6.2.10.1 for specifications on the slope of the calibration curve). This calibration factor can be used for calculating results provided there are no radical changes in temperature or pH. At least one control sample containing a known concentration of SO_2 for each series of determinations, is recommended to insure the reliability of this factor.

8.2.2 *Procedure with SO_2 Permeation Tubes.*

8.2.2.1 *General Considerations.* Atmospheres containing accurately known amounts of sulfur dioxide at levels of interest can be prepared using permeation tubes. In the systems for generating these atmospheres, the permeation tube emits SO_2 gas at a known, low, constant rate, provided the temperature of the tube is held constant (±0.1° C.) and provided the tube has been accurately calibrated at the temperature of use. The SO_2 gas permeating from the tube is carried by a low flow of inert gas to a mixing chamber where it is accurately diluted with SO_2-free air to the level of interest and the sample taken. These systems are shown schematically in Figures A2 and A3 and have been described in detail by O'Keeffe and Ortman (*9*), Scaringelli, Frey, and Saltzman (*10*), and Scaringelli, O'Keeffe, Rosenberg, and Bell (*11*).

8.2.2.2 *Preparation of Standard Atmospheres.* Permeation tubes may be prepared or purchased. Scaringelli, O'Keeffe, Rosenberg, and Bell (*11*) give detailed, explicit directions for permeation tube calibration. Tubes with a certified permeation rate are available from the National Bureau of Standards. Tube permeation rates from 0.2 to 0.4 μg./minute inert gas flows of about 50 ml./minute and dilution air flow rates from 1.1 to 15 liters/minutes conveniently give standard atmospheres containing desired levels of SO_2 (25 to 390 μg./m.³; 0.01 to 0.15 p.p.m. SO_2). The concentration of SO_2 in any standard atmosphere can be calculated as follows:

$$C = \frac{P \times 10^3}{R_d + R_i}$$

Where:

C = Concentration of SO_2, μg./m.³ at reference conditions.
P = Tube permeation rate, μg./minute.
R_d = Flow rate of dilution air, liter/minute at reference conditions.
R_i = Flow rate of inert gas, liter/minute at reference conditions.

8.2.2.3 *Sampling and Preparation of Calibration Curve.* Prepare a series (usually six) of standard atmospheres containing SO_2 levels from 25 to 390 μg. SO_2/m.³. Sample each atmosphere using similar apparatus and taking exactly the same air volume as will be done in atmospheric sampling. Determine absorbances as directed in 7.2. Plot the concentration of SO_2 in μg./m.³ (x-axis), against $A-A_o$ values (y-axis), draw the straight line of best fit and determine the slope. Alternatively, regression analysis by the method of least squares may be used to calculate the slope. Calculate the reciprocal of the slope and denote as B_s.

8.3 *Sampling Efficiency.* Collection efficiency is above 98 percent; efficiency may fall off, however, at concentrations below 25 μg./m.³. (*12, 13*)

9. *Calculations.*

9.1 *Conversion of Volume.* Convert the volume of air sampled to the volume at reference conditions of 25° C. and 760 mm. Hg. (On 24-hour samples, this may not be possible.)

$$V_R = V \times \frac{P}{760} \times \frac{298}{t+273}$$

V_R = Volume of air at 25° C. and 760 mm. Hg, liters.
V = Volume of air sampled, liters.
P = Barometric pressure, mm. Hg.
t = Temperature of air sample, °C.

9.2 *Sulfur Dioxide Concentration.*

9.2.1 When sulfite solutions are used to prepare calibration curves, compute the concentration of sulfur dioxide in the sample:

$$\mu g.\ SO_2/m.^3 = \frac{(A-A_o)\ (10^3)\ (B_s)}{V_R} \times D$$

A = Sample absorbance.
A_o = Reagent blank absorbance.
10^3 = Conversion of liters to cubic meters.
V_R = The sample corrected to 25° C. and 760 mm. Hg, liters.

*Hartmen-Leddon, 60th and Woodland Avenue, Philadelphia, PA 19143.

B_s = Calibration factor, μg./absorbance unit.

D = Dilution factor.

For 30-minute and 1-hour samples, D=1.

For 24-hour samples, D=10.

9.2.2 When SO_2 gas standard atmospheres are used to prepare calibration curves, compute the sulfur dioxide in the sample by the following formula:

$$SO_2, \mu g./m.^3 = (A - A_o) \times B_s$$

A = Sample absorbance.

A_o = Reagent blank absorbance.

B_s = (See 8.2.2.3).

9.2.3 *Conversion of μg./m.³ to p.p.m.* = If desired, the concentration of sulfur dioxide may be calculated as p.p.m. SO_2 at reference conditions as follows:

$$\text{p.p.m. } SO_2 = \mu g.\ SO_2/m.^3 \times 3.82 \times 10^{-4}$$

10. *References.*

(1) West, P. W., and Gaeke, G. C., "Fixation of Sulfur Dioxide as Sulfitomercurate III and Subsequent Colorimetric Determination", *Anal. Chem. 28*, 1816 (1956).

(2) Ephraims, F., "Inorganic Chemistry," p. 562, Edited by P.C.L. Thorne and E. R. Roberts, 5th Edition, Interscience. (1948).

(3) Lyles, G. R., Dowling, F. B., and Blanchard, V. J., "Quantitative Determination of Formaldehyde in Parts Per Hundred Million Concentration Level", *J. Air Poll. Cont. Assoc. 15*, 106 (1965).

(4) Scaringelli, F. P., Saltzman, B. E., and Frey, S. A., "Spectrophotometric Determination of Atmospheric Sulfur Dioxide", *Anal. Chem. 39*, 1709 (1967).

(5) Pate, J. B., Ammons, B. E., Swanson, G. A., Lodge, J. P., Jr., "Nitrite Interference in Spectrophotometric Determination of Atmospheric Sulfur Dioxide", *Anal. Chem. 37*, 942 (1965).

(6) Zurlo, N. and Griffini, A. M., "Measurement of the SO_2 Content of Air in the Presence of Oxides of Nitrogen and Heavy Metals", *Med. Lavoro, 53*, 330 (1962).

(7) Scaringelli, F. P., Elfers, L., Norris, D., and Hochheiser, S., "Enhanced Stability of Sulfur Dioxide in Solution", *Anal. Chem. 42*, 1818 (1970).

(8) Lodge, J. P. Jr., Pate, J. B., Ammons, B. E. and Swanson, G. A., "Use of Hypodermic Needles as Critical Orifices in Air Sampling," *J. Air Poll. Cont. Assoc. 16*, 197 (1966).

(9) O'Keeffe, A. E., and Ortman, G. C., "Primary Standards for Trace Gas Analysis", *Anal. Chem. 38*, 760 (1966).

(10) Scaringelli, F. P., Frey, S. A., and Saltzman, B. E., "Evaluation of Teflon Permeation Tubes for Use with Sulfur Dioxide", *Amer. Ind. Hygiene Assoc. J. 28*, 260 (1967).

(11) Scaringelli, F. P., O'Keeffe, A. E., Rosenberg, E., and Bell, J. P., "Preparation of Known Concentrations of Gases and Vapors with Permeation Devices Calibrated Gravimetrically", *Anal. Chem. 42*, 871 (1970).

(12) Urone, P., Evans, J. B., and Noyes, C. M., "Tracer Techniques in Sulfur Dioxide Colorimetric and Conductiometric Methods", *Anal Chem. 37*, 1104 (1965).

(13) Bostrom, C. E., "The Absorption of Sulfur Dioxide at Low Concentrations (p.p.m.) Studied by an Isotopic Tracer Method", *Intern. J. Air Water Poll. 9*, 33 (1965).

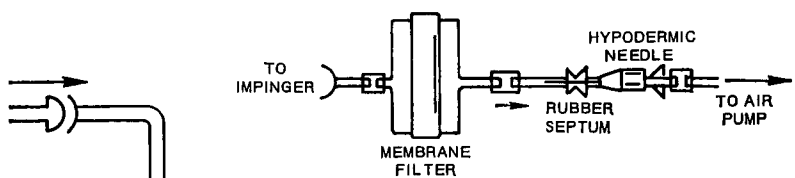


Figure A1a. Critical orifice flow control.

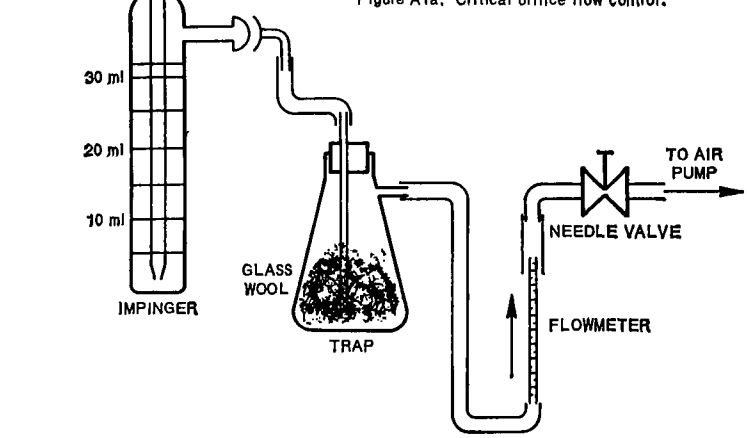


Figure A1. Sampling train.

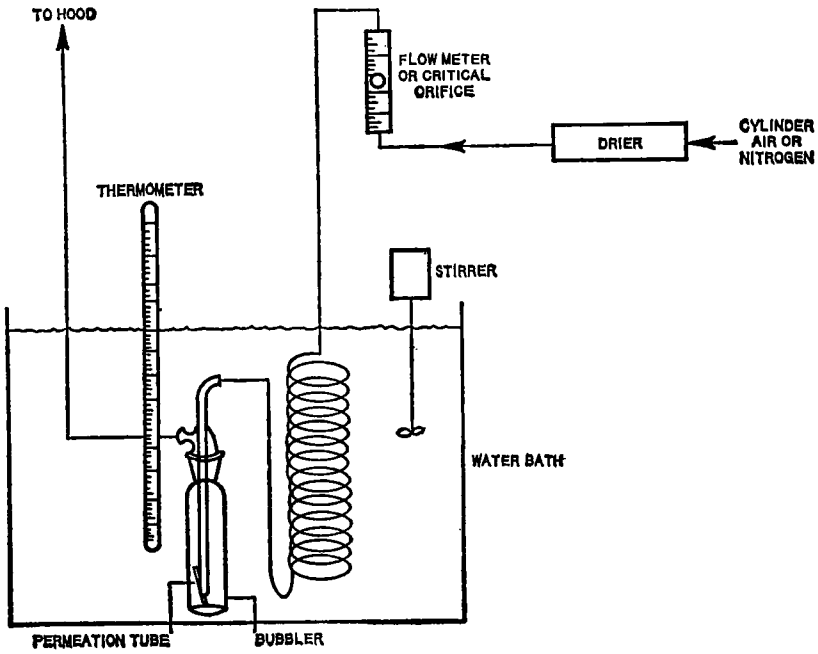


Figure A2. Apparatus for gravimetric calibration and field use.

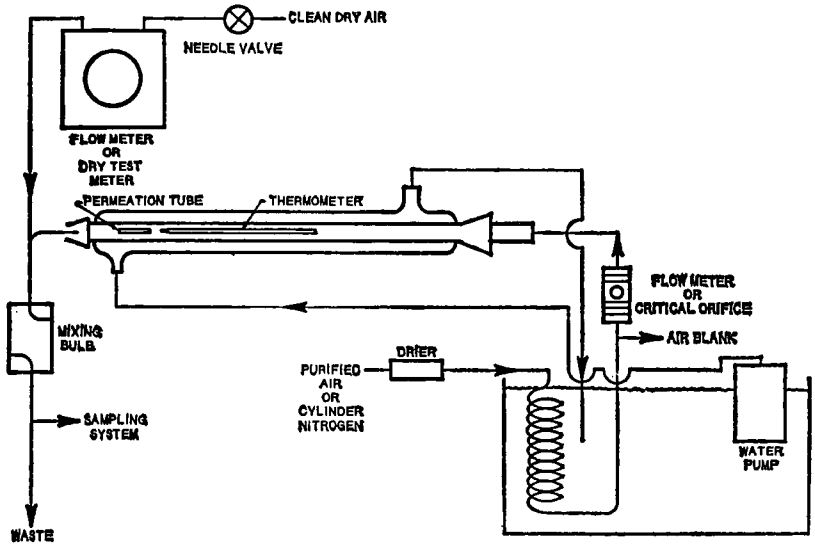


Figure A3. Permeation tube schematic for laboratory use.

APPENDIX B

STATISTICAL DESIGN AND ANALYSIS

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

APPENDIX B

STATISTICAL DESIGN AND ANALYSIS

I. INTRODUCTION

In the application of interlaboratory testing techniques, the first step is to determine the exact purpose of the program. There are many, and the particular one must be established. All subsequent details of the program must be planned keeping the prime objective in mind. This appendix describes the design and analysis of the formal collaborative test which followed the Method Familiarization Session. The Method Familiarization Session has been described in the main report.

A. Purpose and Scope of the Experiment

The primary purpose was to establish the reliability of the method in terms of systematic variation, precision, and accuracy. More emphasis was placed on the inherent quality of the method when properly used than upon the performance of the laboratories–from the standpoint of the selection of collaborators as well as from the standpoint of the disposition of outlying results.

At the same time, it was desirable to retrieve information which would allow the investigation of various steps within the method; therefore, emphasis was placed upon obtaining intermediate data relating to calibration curves, control samples, and blanks. As a result, a substantial amount of data was obtained in addition to the end result of the analytical procedure.

The statistical planning of the program, which necessarily must be limited in scope, depends upon what information is desired. The scope is limited by what a collaborating laboratory can conveniently and economically accomplish, as well as by the number of collaborators that can be accommodated. Under these limitations, it was possible to examine the effects of laboratories, concentrations, and days upon the precision of the method in addition to estimating the replication error. The main experiment and all secondary experiments were designed so that the analysis of variance technique could be used.

Fourteen laboratories took part in the program. An analyst representing each laboratory attended the Method Familiarization Session and subsequently conducted the formal collaborative testing. These individuals and their affiliations have been identified elsewhere in the main report. These laboratories constitute a random sample from a rather large population of experienced laboratories.

Three different concentrations were analyzed by each laboratory. The concentrations were nominally 150, 275, and 820 $\mu g/m^3$. These concentrations were selected to approximate the low range, the optimum range, and the high range for the method. Due to variations among permeation tubes and to variation in atmospheric pressure and temperature, it was not possible for each laboratory to generate test atmospheres having the exact values above; however, the expected concentration can be determined accurately as a function of permeation tube temperature, dilution air temperature and pressure, and volumetric flow rate of the dilution air. The permeation tube system has been described in the main report. In most instances, it was the deviations of the observed values from the expected values that were subjected to statistical analysis.

Each of the three concentrations was analyzed in triplicate on each of three separate days using independently prepared reagents, standards, and calibration curves. The reagents which were to be prepared fresh each day are identified in the main report. In the designation of the reagents to be prepared fresh each day, stability was not the only criterion since these operations represent a portion of the variation between days.

B. Design of the Experiment

A properly planned collaborative test should

allow the analysis of the results by the analysis of variance technique or by a procedure which incorporates this technique.[1-3]* In general, analysis of variance techniques are more efficient than the simpler control chart techniques. Since the cost of statistical analysis is small compared to the total cost involved in a collaborative test, it is desirable to use the most efficient statistical methods available in analyzing the results. High efficiency in data utilization becomes more important when the amount of data is limited.

The form of the analysis depends upon the statistical model under consideration. The experiment was designed so that the data could be analyzed according to three different models or techniques as follows:

- Analysis of variance with the data for each concentration analyzed separately
- Analysis of variance with the data for all concentrations analyzed together, including data transformation if required
- Linear model analysis with data transformation if required.

In addition to providing a comparison of the techniques, this approach also provided the opportunity to use the technique for which results were most convenient to apply, provided that all techniques gave comparable results. It will be shown subsequently that excellent agreement was obtained, and that the second technique with data transformation provided the most convenient and useful application of the results. Each of the techniques will be described in more detail in later subsections.

The overall design of the experiment can best be shown by the diagram in Figure B-1. It can be seen that one analyst in each of fourteen laboratories analyzed each of three concentrations in triplicate on each of three separate days resulting in a total of 378 individual determinations. The data are presented appropriately in the next subsection. The data in this form may readily be analyzed by each of the techniques listed above in accordance with the respective statistical model.

In collaborative testing, two general sources of variability can be readily detected. First, the variability between laboratories (reproducibility) can be

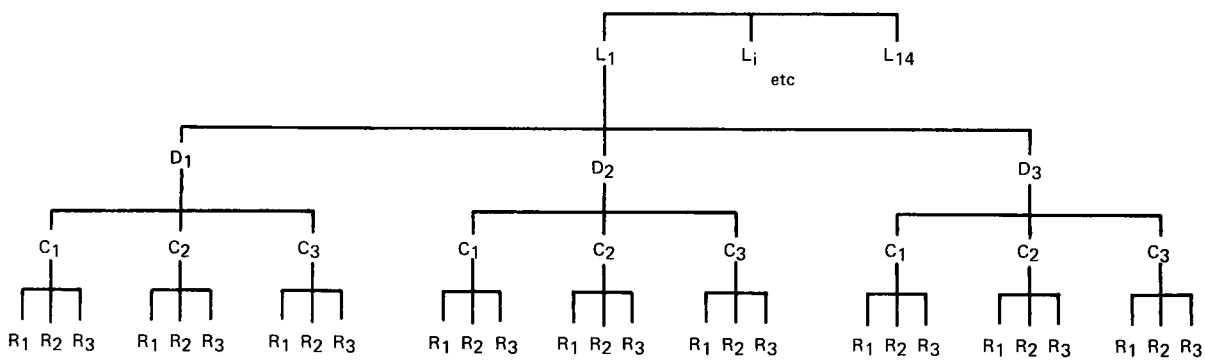


FIGURE B-1. DESIGN OF SULFUR DIOXIDE METHOD EXPERIMENT. L, LABORATORIES; D, DAYS; C, CONCENTRATIONS; R, REPLICATES.

*Superscript numbers in parentheses refer to the List of References at the end of this appendix.

estimated. This is frequently the largest source of variability and is not under the control of the investigator. Second, the within-laboratory variability (repeatability) can be estimated. This source is under the control of the investigator to the extent that the separate components which make up this source may be identified separately. These separate components, of varying magnitude and importance, may be measured if the proper design has been employed. Alternatively, the separate sources may be confounded or lumped into a single variable by altering the design. By employing the design above, separate estimates can be made of the variability between days and of the variability between replicates. These two components, appropriately combined, constitute the within-laboratory source of variability.

A useful purpose could have been served by employing more than one analyst in each laboratory or by including additional different concentrations. Also, sensitivity could have been improved by increasing the number of days or the number of replicates. These innovations would have required considerably more effort on the part of the volunteer collaborators. A conservative estimate for the effect of analysts can be made by assuming that a different analyst is analogous to a different laboratory. It was believed that sufficient sensitivity for the purpose was obtainable with the numbers of laboratories, days, and replicates noted above.

Additional assumptions and rationale for each of the techniques listed above will be stated later as the technique is described and applied. If appropriate, the statistical model will be stated in the respective discussions.

Supplementary or secondary experiments were incorporated to evaluate errors associated with various steps within the method. To accomplish this, a relatively large amount of intermediate data was retrieved. These data consisted of (1) the individual points for each and every calibration curve, (2) the concentration and absorbance of all control samples analyzed, and (3) the absorbances of all blanks. Carefully prepared instructions and data forms were used to retrieve these data uniformly from every collaborator. Each of these supplementary experiments was designed so that the data could be analyzed by the analysis of variance to determine whether variations between days and between laboratories were significant. The results of each of these supplementary experiments have been discussed in the main report.

II. PRELIMINARY DATA ANALYSIS

In accordance with the experiment design described above and with the collaborative test procedure described in the main report, the full plan was carried out satisfactorily and without any missing data.

In a carefully planned program, dishonesty, carelessness, or incompetence can readily be detected and the data eliminated. There were no data of this category present in the results of this test. However, extreme or atypical results must be dealt with. There is no problem in detecting these results. Whether a result is out-of-line or not may be decided by obvious explanation (either by the investigator or the collaborator), visual observation, or by statistical methods. What disposition is to be made of the results of a laboratory that are responsible for outlying points? Three obvious alternatives are available; they are (1) retain all data in the analysis, (2) delete the entire data, or (3) make some logical substitution for the atypical data so that the other data of the laboratory can be utilized. No amount of discussion or statistical testing can substitute for a straightforward facing of the problem. Clearly, the use of the first alternative, retention of all data, puts more emphasis upon the performance of the laboratories rather than the quality of the procedure. This would be in direct opposition to the objectives stated previously.

Use of the second alternative, deletion of data, is in accord with the objective; however, it carries with it two distinct disadvantages. First, there is the tendency to make the method appear more reliable than perhaps it is. Second, good data must be sacrificed to eliminate suspicious data. The proportion of good data to bad data is usually high, making the sacrifice a costly one, and at the same time reducing the sensitivity of the experiment by reducing the cor-

responding degrees of freedom. For example, an individual laboratory in this collaborative test produced 27 individual results consisting of nine sets of triplicate analyses. The statistical techniques require that there be no missing data. Therefore, if one individual result or one set of replicates were bad, a considerable amount of good data would be lost if the laboratory were omitted from the analysis.

The foregoing discussion makes the third alternative an attractive and logical compromise. Several methods, both simple and complex, are available to replace missing data. Only the more simple methods appear to be justified in this case. The one selected was to replace the outlying observation by its closest neighbor. For example, if an individual replicate is an outlier, it is replaced by the nearest value in its respective set of three replicates. If the mean of a set of three replicates is an outlier, the set of replicates is replaced by the set of replicates from the same concentration whose mean is nearest the mean of the outlier set. Replacements beyond these two cases will rarely be required and are probably not justified. This outlier replacement technique accomplishes two objectives simultaneously. It salvages good data without a high risk, and it avoids the tendency to make the method appear more reliable than it might be.

Finally, what is to be done about the arithmetic errors that sometimes appear? A check for these errors must be made to avoid the significant ones. This procedure, using computer technology, is relatively inexpensive and does not require a great amount of effort. At the same time, a set of error-free data is generated which is consistent from laboratory to laboratory in calculation sequence and round-off. Since the emphasis is upon the quality of the method and not upon the performance of the laboratories, it seems justifiable to use these error-free data for all subsequent analysis, and to evaluate the arithmetic error problem from a qualitative standpoint. Certainly, if a procedure is so complex or cumbersome in its calculations that it is overly vulnerable to errors, that fact should be pointed out. However, it makes no sense to express arithmetic erors quantitatively in terms of means or variances or of components thereof.

A. Presentation of Data

The data resulting from the experiment are rather voluminous; however, it is essential that these data be tabulated for future reference. In addition to their necessity as supporting information for the problem at hand, the data are also valuable academically as a source of data for the development, evaluation, and comparison of new statistical techniques. Therefore, the more voluminous raw data will be found in Appendix C.

The volume of data can be considerably reduced and the presentation simplified by tabulating the deviations of the observed values from the expected values. That is, the expected result was subtracted from the observed result and the algebraic difference was tabulated. These data are presented in Table B-I and were derived from the data of Tables C-I and C-II. Each section of Table B-I represents a different concentration level and shows each individual replicate for each laboratory for each day. All errors have been corrected. It can be seen by referring to Table C-II in Appendix C that the ranges of the expected values are sufficiently narrow so that comparison of deviations of observed values from expected values within these ranges can be done without risk of error.

Additional description of these data will be presented as appropriate in the discussion of the analysis which follows.

B. Tests for Outlying Observations

The data were first visually inspected for unusually large departures from the expected values. Only one such set of replicates was noted, and these had been pointed out by the collaborator submitting them. These data for Laboratory 926 on the second day can be seen in Table B-I. The absolute values of these observations were so near zero that some quite unusual error was implied. Several explanations come to mind; however, it is hardly worth the conjecture to list them. Following the previously stated philosophy,

TABLE B-I. DEVIATION FROM EXPECTED VALUES FOR EACH REPLICATE FOR EACH CONCENTRATION FOR EACH DAY FOR EACH LABORATORY, MICROGRAMS PER CUBIC METER.

Laboratory Code Number	Day 1			Day 2			Day 3		
Low Concentration									
271	2	−8	−13	4	−2	−4	−9	−8	−7
274	7	7	11	2	7	12	3	9	19
305	−4	0	−6	−5	−2	−8	−2	−4	−10
345	19	40	30	31	43	48	−1	10	−13
500	22	−3	−3	−19	−19	−19	−8	4	−11
509	−54	−54	−56	−3	−12	−5	−35	−30	−35
526	−27	−26	−28	−24	−17	−18	−17	−29	−25
571	24	12	0	60	26	14	2	−3	−3
578	−8	−15	−15	−18	−26	−18	0	−4	−9
655	109	133	129	88	97	99	87	108	90
788	16	−14	−1	90	57	30	−19	−20	−23
920	4	5	−6	−20	−19	−21	0	−10	−10
926	24	19	24	−142	−137	−137	35	30	30
927	−5	−5	−5	−13	−5	−15	2	−8	−8
Intermediate Concentration									
271	−16	−14	−15	−24	−21	−24	−20	−15	−19
274	22	29	29	14	20	28	45	45	45
305	−3	−3	−3	−8	−5	−7	−18	−10	−18
345	−13	4	−18	−26	21	32	−11	0	−6
500	10	7	14	−11	−2	−11	−3	0	42
509	−45	−45	−48	−26	−17	−17	−30	−43	−36
526	−38	−38	−35	−25	−30	−30	−25	−19	−16
571	24	14	12	−15	−7	5	7	−6	−10
578	10	−6	−7	−28	−25	−21	−10	−7	−7
655	87	76	87	53	44	63	48	55	54
788	91	45	38	36	22	29	12	−14	−20
920	−12	−7	−16	−20	−22	−22	−15	−18	−7
926	30	26	26	27	25	18	62	62	55
927	−10	−12	−10	−29	−19	−19	−12	−14	−14
High Concentration									
271	−58	−67	−60	−40	−37	−43	−62	−69	−57
274	37	37	86	60	60	60	94	94	118
305	−44	−33	−44	−74	−64	−79	−73	−71	−64
345	−1	−23	−50	25	154	165	−56	−45	−45
500	0	−39	−27	−25	−37	−67	−70	−58	−70
509	−95	−93	−95	−34	−34	−88	−72	−74	−61
526	−103	−109	−103	−64	−51	−26	−58	−74	−52
571	72	35	23	−9	−24	−52	−17	−61	−39
578	−14	−17	−11	−144	−134	−132	−111	−118	−99
655	103	92	103	31	54	31	56	67	67
788	16	−30	−56	−40	−54	−67	−101	−101	−127
920	−34	−29	−39	−46	−49	−41	−29	−16	−36
926	91	91	111	74	70	84	164	164	170
927	−69	−69	−59	−65	−65	−65	−53	−43	−49

these replicates were replaced by the set of replicates for the first day for this concentration.

A quite thorough statistical examination for outlying observations was then performed. It is believed to be beyond the scope of this report to describe this examination in detail; therefore, the description will be superficial and the methods used will be cited appropriately. Needless to say, computer techniques were used to facilitate these analyses.

Two methods were used to test for outliers among laboratory means, among day means within laboratories, and among replicates. The respective means or observations were examined by Dixon's test[4] and also by a method attributed to David and described by ASTM.[5] The second method was accomplished by making the analysis of variance, which will be described in the next subsection. Although this might seem to be premature, it is fair to say that these computer-assisted analyses are quick and inexpensive and may be repeated, if necessary, after the disposition of outliers or after later transformation of data. The output of the analysis of variance computer program[6] was used as the input for a special program[7] to test for outliers among the various means. Testing was done at a high level of confidence (99 percent), and any outliers detected by this method were tested using Dixon's test. Borderline cases were not rejected. Relative outliers between days were ignored unless the magnitude was also significant. It should be noted that these outlier tests were applied to the data as they appear in Table B-I, and, when a data transformation was later found to be appropriate, the outlier tests were repeated on the transformed data for verification.

The only outlying observations under these criteria were the data for Laboratory 345 for the high concentration on the second day (see Table B-I). The disposition of these data will be described following the description of the tests for homogeneity of variances below.

Bartlett's test[8] and Cochran's test[9] were used to test for homogeneity of variances. The variances of the laboratory means for each of the three concentrations were found to be nonhomogeneous by both tests. This indicated the necessity to handle each concentration separately or to find an appropriate data transformation in order to stabilize the variance. This was not unexpected, considering the difference in magnitude of the deviations in Table B-I which appear to vary with concentration.

These tests of variances were pursued further to find any laboratory with data substantially more scattered than the main cluster. This examination showed the data for Laboratory 345 for the high concentration for the second day to be inconsistent. The variance of this set of replicates was extremely high and in a class all by itself. The mean for this set was shown to be an outlier above.

Since the remainder of the data for this laboratory did not contain any other atypical results, it was logical to replace this set of replicates by the set of replicates for the same concentration for the first day (see Table B-I).

C. Discussion of Results of Preliminary Data Analysis

In the presentation of the results in this report, all arithmetic errors, no matter how small or insignificant, have been corrected. No offense to the collaborators is intended since this was simply a built-in part of the design of the experiment. Collaborators may note differences of one or two units in the least significant digit of their reported values and the values shown in Table C-I. Deviations beyond this magnitude reflect more significant arithmetic errors or other errors in the calculation procedure.

A relatively small number of these types of errors of varying degrees of magnitude was noted. Only three of the fourteen laboratories submitting results exhibited any errors of this type. Four instances of inadvertent errors in arithmetic operations were noted and corrected. More important, however, was the fact that two laboratories had difficulty with the least squares technique which was a part of the method. There are many separate mathematical operations in this procedure and, conse-

quently, many opportunities for errors, especially in transcription of intermediate results. The method is one containing complex calculational procedures and is consequently vulnerable to arithmetic and procedural errors. However, the majority of collaborators have demonstrated the capability to handle this complexity. There is no reason to believe that a careful checking procedure would not eliminate this problem.

The data produced as a result of the disposition of the two outlying observations described above are believed to be an excellent basis for the statistical analyses to follow. The data are believed to be representative and unbiased. The number of outliers is small, especially in comparison with the first test where nearly two-thirds of the laboratories produced one or more atypical results.

In all cases of outlier analysis, the statistical tests were interpreted with a good deal of judgment. The quantitative results of a given test were viewed as a guide and a part of the picture, not as an obligation to delete or retain the observation. This combination of objectivity and judgment is considered to be extremely important and to be the backbone of all inferences following.

III. ANALYSIS OF VARIANCE

In this subsection, three analyses will be described and the results discussed. The first two are classic analysis of variance cases. The first handles each concentration separately, while the second combines all concentration data into a single analysis, and includes the evaluation and application of the necessary data transform to allow this treatment. The third case, the linear model analysis, does not strictly constitute a classic analysis of variance case but does involve the technique. It is therefore logically included here. Each of these cases will be discussed under its respective heading below, and, finally, a comparison of the results of each will be presented along with conclusions derived from the most applicable method. This multiple analysis is another approach to insure the reliability of the final conclusions.

A. Analysis of Variance of Concentrations Separately

This analysis is made in accordance with recommended practices for conducting an interlaboratory study using one material.[1,3] Discussions and recommendations have also been presented in other sources.[10,11] An extremely valuable and flexible computer program[6] facilitated the accomplishment of the mathematical treatment. At the same time, a valuable but very specific computer program[12] was developed to materially ease the calculation of the components of variance.

The purpose of this analysis was to compare, for each concentration, the magnitudes of three of the four sources of variation which this study was designed to examine. These are the relative variations among laboratories, among days within laboratories, and among replicates run together on the same day. The mathematical model is as follows:

$$y_{ikm} = A + L_i + D_{k(i)} + e_{m[k(i)]} \qquad \text{(B-1)}$$

where

i = 1, 2, 3 . . . p designates a laboratory

k = 1, 2, 3 w designates a day

m = 1, 2, 3 . . . n designates an individual replication

The term y_{ikm} represents an individual measurement, A represents the overall average, L_i represents the effect of the ith laboratory, $D_{k(i)}$ represents the effect of the kth day nested in the ith laboratory, and $e_{m[k(i)]}$ represents the random deviation associated with an individual measurement.

In this study, there were p = 14 laboratories, w = 3 days, and n = 3 replicates.

The analysis was applied to the deviations in Table B-I after the substitutions for the two outlier cells. The results are shown in the analysis of variance tables in Table B-II. All effects in each of the tables are significant at the 95 percent level of significance.

The components of variance and the repeatability and reproducibility were calculated and are shown in Table B-III. The percent of the total variance accounted for by each component is shown along with the degrees of freedom and the 95 percent confidence intervals for each. The confidence intervals for the components were computed by a method from Scheffé.[13] The degrees of freedom for the repeatability and reproducibility were estimated

TABLE B-II. ANALYSIS OF VARIANCE FOR EACH CONCENTRATION. DATA IN ORIGINAL SCALE. THREE FACTORS: L, LABORATORIES; D, DAYS; R, REPLICATES.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square
Low Concentration				
L	124796.0000	13	9599.6923	$\sigma_R + 3\sigma_D + 9\sigma_L$
D(L)	22459.7778	28	802.1349	$\sigma_R + 3\sigma_D$
R(LD)	6565.3333	84	78.1587	σ_R
Intermediate Concentration				
L	89408.6349	13	6877.5873	Same
D(L)	16944.2222	28	605.1508	
R(LD)	7528.0000	84	89.6190	
High Concentration				
L	455465.8810	13	35035.8370	Same
D(L)	87629.1111	28	3129.6111	
R(LD)	18054.6667	84	214.9365	

TABLE B-III. COMPONENTS OF VARIANCE FOR EACH CONCENTRATION. DATA IN ORIGINAL SCALE. THREE FACTORS: L, LABORATORIES; D, DAYS; R, REPLICATES.

Source of Variation	Component	Percent of Total	Degrees of Freedom	Standard Deviation	95 Percent Confidence Interval
Low Concentration					
L	977.5064	75.4	13	31.27	22 to 52
D(L)	241.3254	18.6	28	15.53	12 to 22
R(LD)	78.1587	6.0	84	8.84	8 to 10
Repeatability	319.4841		28	17.87	
Reproducibility	1296.9905		13	36.01	
Intermediate Concentration					
L	696.9374	72.7	13	26.40	18 to 44
D(L)	171.8439	17.9	28	13.11	10 to 18
R(LD)	89.6190	9.4	84	9.47	8 to 11
Repeatability	261.4629		28	16.17	
Reproducibility	958.4003		13	30.96	
High Concentration					
L	3545.1362	74.9	13	59.54	41 to 99
D(L)	971.5582	20.5	28	31.17	24 to 43
R(LD)	214.9365	4.6	84	14.66	13 to 17
Repeatability	1186.4947		28	34.45	
Reproducibility	4731.6309		13	68.79	

according to ASTM recommended practice.[14]

These point estimates and their corresponding confidence intervals show the effects at the highest concentration to be significantly different from those at the two lower concentrations, especially the replication error and the variation between days.

Some very important conclusions can be drawn at this point. First, it is now clear that the replication error varies with the magnitude of the concentration. This means that a data transformation is in order and the basis for that transform has been established. The transform will be evaluated and applied in the next subsection.

Second, the relative contribution of each of the factors to the total variance is now evident. As a result, it can be seen that the component of variance due to laboratories accounts for three-fourths of the total variance–far larger than any other component. The component due to days accounts for about one-fifth of the total variance, and the component due to the replication is small compared to either of the other components. It accounts for roughly one-twentieth of the total variance, which suggests that replication is probably a waste of time and effort.

These point estimates will be compared graphically with the other methods later.

B. Analysis of Variance for All Concentrations Analyzed Together

The results of the analysis of variance of the concentrations analyzed separately in the preceding subsection have set the stage for the analysis in this discussion. It was evident that it was not worthwhile to make an analysis of variance with all concentrations handled together without first making a data transformation because of the lack of homogeneity of variances between the different concentrations tested. The use of a data transformation stabilizes the variance and allows the analysis.

The work of Mandel[15,16] illustrates a technique for determining an appropriate transform when the variance (replication) varies with the magnitude of the concentration. The relationship of the standard deviation for replicates to concentration was obtained from the analysis in the preceding subsection. The line formed by these points can be expressed in terms of its slope and intercept as

$$\text{Standard deviation} = 7 + 0.01 \times \text{concentration}$$

Accordingly, the appropriate transform for each observation is given by

$$z = K \log_e (A + By) - G \qquad \text{(B-2)}$$

where z is the transformed variable, K and G are arbitrary constants chosen for convenience, $A = 7$ from above, and $B = 0.01$ from above. The value of G was chosen as zero and the value of K as 1000.

This transformation was applied to each of the observations in Table C-I (with substitutions described previously for outlying observations) and to each of the expected values in Table C-II. A table of differences analogous to Table B-I was generated and subsequently subjected to an analysis of variance. This table has not been reproduced here.

The purpose of this analysis was the same as the preceding analysis except that the laboratory effects, day effects, and replication error could be expressed as functions of the concentration. This has a distinct advantage over the point estimates of the preceding subsection.

Bartlett's test[8] and Cochran's test[9], in addition to the outlier tests[4,5], were applied to the transformed differences for verification and assurance. These transformed data were analyzed handling the concentrations separately according to the technique of the preceding subsection to demonstrate that the variances were indeed homogeneous. The resulting assurances from this additional work were well worth the small effort, and the relative contributions of the different components for each concentration were found to be nearly identical.

The mathematical model for this analysis is:

$$y_{ikm} = A + L_i + M_j + D_{k(i)} + (LM)_{ij} + (DM)_{k(i)j} + e_{m[k(i)j]} \quad \text{(B-3)}$$

where

$j = 1, 2, 3 \ldots q$ designates a material (concentration).

All other subscripts have been defined previously. The term M_j represents the main effect of materials, $(LM)_{ij}$ represents the laboratory-material interaction, and $(DM)_{k(i)j}$ represents the day-material interaction. All terms involving materials (concentrations) are viewed as fixed effects while all other effects, except the overall mean, are viewed as random effects. All other terms have been previously defined.

For this analysis, there were $p = 14$ laboratories, $q = 3$ materials (concentrations), $w = 3$ days nested within a laboratory, and $n = 3$ replicates.

The resulting analysis of variance table is shown in table B-IV and the corresponding components of variance in Table B-V. All effects can be seen to be significant from the 95 percent confidence intervals in Table B-V. Remember that these estimates are now based on transformed data. The various standard deviations may be converted to the original scale by the linear approximation.(15)

$$\sigma_y = \frac{A + By}{KB}\sigma_z = (0.7 + 0.001y)\sigma_z \quad \text{(B-4)}$$

where σ_z is a standard deviation in the transformed scale, σ_y is the corresponding standard deviation in the original scale, and y is the concentration. The accuracy is sufficient for values of σ_z/K of less than 0.05. These relationships result in straight lines with non-zero intercepts. They are simple and easy to use and may be seen graphically for the replication error, repeatability, and reproducibility by looking ahead to Figure B-3 in a following subsection where they are compared with the results from the point estimates of the preceding analysis.

TABLE B-IV. ANALYSIS OF VARIANCE FOR ALL CONCENTRATIONS TOGETHER. DATA IN TRANSFORMED SCALE. FOUR FACTORS: L, LABORATORIES; M, MATERIALS OR CONCENTRATIONS; D, DAYS; R, REPLICATES.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square
L	373851.5296	13	28757.8100	$\sigma_R + 9\sigma_D + 27\sigma_L$
M	39722.1032	2	19861.0516	$\sigma_R + 3\sigma_{DM} + 9\sigma_{LM} + 126\sigma_M$
D(L)	57762.8631	28	2062.9594	$\sigma_R + 9\sigma_D$
LM	76098.7512	26	2926.8750	$\sigma_R + 3\sigma_{DM} + 9\sigma_{LM}$
DM(L)	28917.0515	56	516.3759	$\sigma_R + 3\sigma_{DM}$
R(LDM)	24063.7298	252	95.4910	σ_R

TABLE B-V. COMPONENTS OF VARIANCE FOR ALL CONCENTRATIONS TOGETHER. DATA IN TRANSFORMED SCALE. FOUR FACTORS: L, LABORATORIES; M, MATERIALS OR CONCENTRATIONS; D, DAYS; R, REPLICATES.

Source of Variation	Component	Percent of Total	Degrees of Freedom	Standard Deviation	95 Percent Confidence Interval
L	988.6982	53.6	13	31.44	22 to 52
M	134.3982	7.3	2	11.59	2 to 100
D(L)	218.6076	11.8	28	14.79	12 to 20
LM	267.8332	14.5	26	16.37	12 to 24
DM(L)	140.2950	7.5	56	11.84	10 to 15
R(LDM)	95.4910	5.2	252	9.77	9 to 11
Repeatability	454.3936		84	21.32	
Reproducibility	1710.9250		13	41.36	

Referring to Table B-V, it can be seen that the repeatability includes the day-material interaction component in addition to the replication error and the day component. The reproducibility includes all components included in the repeatability plus the laboratory component and the laboratory-material interaction component.

C. Linear Model Analysis

The approach in this analysis is different from that of the previous analysis. The assumption is made that systematic differences exist between sets of measurements made by the same observer at different times or by different observers in different laboratories, and that these systematic differences are linear functions of the magnitude of the measurements. Hence, the technique is called "the linear model."[1,15,16] The linear model leads to a simple design, but requires a special method of statistical analysis, geared to the practical objectives of a collaborative test.

The general design is: to each of p laboratories, q materials have been sent for test, and each laboratory has analyzed each material n times. We still have p = 14 laboratories running n = 3 replicates; however, we now view each of the three concentrations on each of the 3 days as a separate material and thus have q = 9 materials. These 9 materials cover the concentration range of interest for the method under study. Now, the n determinations made by the ith laboratory on the jth material constitute what will be denoted as the "i,j cell." The n replicates of any particular cell are viewed as a random sample from a theoretically infinite population of measurements within that cell. The laboratories, however, are now considered as a random sample from a larger population of laboratories, but are considered as fixed variables. Therefore, the inferences involving the variability among laboratories are limited, at least theoretically, to those laboratories participating in the test. The set of values which correspond to the q materials is viewed as a fixed variable, but each material is considered to be a random selection from a population of materials with the same "value."

This model allows for nonconstant, nonrandom differences between laboratories where the previous method does not. The method is not as sensitive to outliers as is the conventional analysis of variance where even a single outlier may result in an unusually large interaction term.

The first step is to examine the relation between the replication error and the magnitude of the measurement. If the standard deviation varies with the concentration, then an appropriate data transformation must be made. This step is exactly the same as previously, except that now we are dealing only with the observations in Table C-I which have been assembled into the i,j cells defined above. From the assumption of linear relationships among the p laboratories, it follows that the values obtained by each laboratory are linearly related to the corresponding average values of all laboratories.

Next, we may plot the transformed measured values versus their respective means. This should be a linear function, and the points corresponding to each line may be represented by three parameters: a mean; a slope; and a quantity related to the deviation from linearity, the standard error of estimate. These parameters are determined by a least squares regression analysis, and the results are shown in Table B-VI. They are more easily compared from the graphic presentation in Figure B-2 where they have been sorted into an ascending order relative to the means. This sorting often reveals effects not readily visible otherwise. Control limits, based upon the deviation from linearity, are shown for each parameter in Figure B-2. These 95 percent control limits indicate several points to be "out of control." This indicates that there are other important sources of error significantly larger than the replication error.

Examination of Figure B-2 reveals which laboratories were responsible for the greatest deviations from linearity, which laboratories showed the greatest departure from unit slope, and which laboratories showed the greatest departures from the overall mean. When viewing this figure, it is important to

TABLE B-VI. MEANS, SLOPES, AND STANDARD ERRORS OF ESTIMATE FOR LINEAR MODEL ANALYSIS. DATA IN TRANSFORMED SCALE.

Laboratory Code Number	Mean	Slope	Standard Error of Estimate
271	2359	0.9906	8.2
274	2409	1.0995	15.3
305	2361	0.9762	6.2
345	2380	0.9661	17.1
500	2369	0.9933	15.6
509	2339	1.0207	21.4
526	2344	1.0040	16.6
571	2384	1.0083	16.9
578	2351	0.9557	24.6
655	2453	0.9334	24.4
788	2379	0.9282	35.3
920	2363	1.0197	10.3
926	2426	1.1191	21.1
927	2358	0.9852	8.0
Mean	2377	1.0000	19.5*

*Pooled estimate.

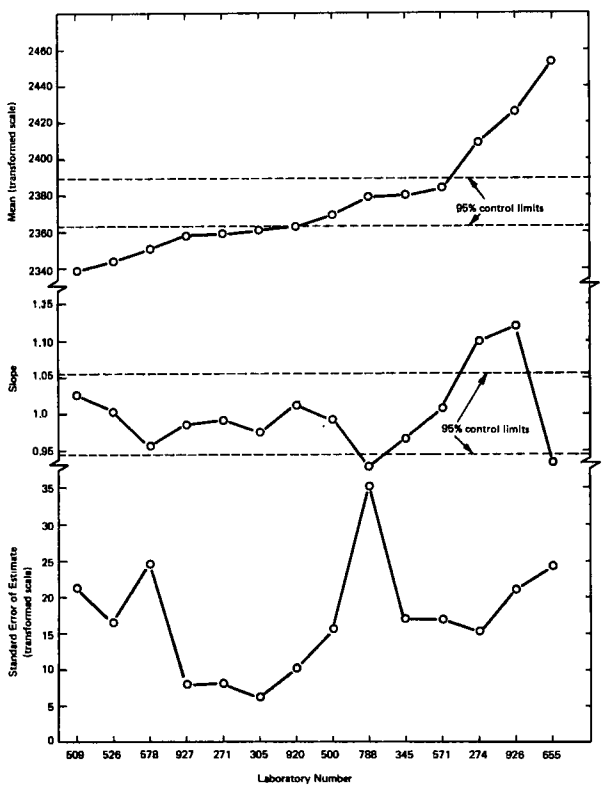


FIGURE B-2. CONTROL CHARTS FOR MEANS, SLOPES, AND STANDARD ERRORS OF ESTIMATE FOR LINEAR MODEL ANALYSIS. DATA IN TRANSFORMED SCALE.

watch for relationships between the parameters. The linear model dictates that the correlation between the means and the slopes be investigated, and this investigation revealed practically no correlation between these two parameters. It can be noticed that there are two laboratories (578 and 788) which produced results with a low slope and a high deviation from linearity. These particular laboratories obtained disproportionately lower results on the highest concentration in comparison with the two lower concentrations, thus accounting for this situation. No other unique combinations or relationships are evident.

The general model for the analysis of the results, classified according to two criteria, laboratories and materials, is:

$$y_{ij} = A + L_i + M_j + (LM)_{ij} \qquad \text{(B-5)}$$

where

$i = 1, 2, 3 \ldots p$ designates a laboratory

$j = 1, 2, 3 \ldots q$ designates a material

The term y_{ij} represents an individual measurement, A represents the overall average, L_i represents the effect of laboratory i, M_j represents the effect of material j, and $(LM)_{ij}$ represents the interaction effect between laboratory i and material j and includes the replication error.

The interaction term is partitioned further as follows:

$$(LM)_{ij} = (b_i - \bar{b})(c_j - \bar{c}) + d_{ij} \qquad \text{(B-6)}$$

where the first term on the right is the linear term in which b_i is the slope determined by the ith laboratory; $\bar{b}$ is the slope of the average response line, which in this case is equal to one; c_j represents the true value for the jth material; and $\bar{c}$ represents the true mean value for all materials. The second term, d_{ij}, is the deviation from linear term. The linear term indicates the difference in slope of the line for a particular laboratory and the average slope for all laboratories, and the nonlinear term expresses the departures from linearity for this individual line.

Starting with the ordinary two-factor analysis of variance, the deviation from linear component of the interaction sum of squares was computed. The

sum of squares for the linear component was obtained by difference. Finally, a single degree of freedom was extracted from the linear component of interaction by multiplying the linear component sum of squares by the square of the correlation coefficient of the means versus the slopes. This is denoted the concurrence term. The nonconcurrence term is computed by difference. These terms are computed for the sake of completeness, although it was apparent that no appreciable correlation existed between the means and the slopes. The final analysis is shown in Table B-VII from which variance components can be computed.

The components of variance were computed using the technique of Mandel[15] with the data from Table B-VII. A computer program was prepared to expedite these computations. The components are defined as follows:

$V(\epsilon)$ = the component of variance due to variability among replicates,

$V(\lambda)$ = the component of variance characterizing the differential response of different laboratories to interfering properties; it represents the irreducible experimental error of the method,

$V(\mu)$ = the component of variance due to that part of the between-laboratory variability involving the variability of the means of the response lines,

$V(\delta)$ = the component of variance due to that part of the between-laboratory variability involving the portion of the variability of the slopes of the response lines which is unrelated to the means.

In the transformed scale, $V(\epsilon)$ and $V(\lambda)$ are, of course, constant and have the values 95.4 and 322.7, respectively. $V(\mu)$ and $V(\delta)$, however, are dependent upon the magnitudes of the measurement. The relative contributions to the total variance are independent of which scale is used, and, since the transformed scale values are difficult to visualize, it is advantageous to reconvert to the original scale. This can be done according to Equation (B-4), and the results are shown in Table B-VIII for several values of concentration. Also shown is the fraction of the total variance accounted for by each component.

Since $V(\epsilon)$, the replication component, is small compared to $V(\lambda)$, it is again evident that replication is probably a waste of time since $V(\lambda)$ constitutes a lower limit. The total between-laboratory variability

TABLE B-VII. ANALYSIS OF VARIANCE FOR LINEAR MODEL. DATA IN TRANSFORMED SCALE.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Laboratories	122799.1556	13	9446.0889
Materials	7134311.6753	8	891788.9594
Laboratory X Material	51929.1766	104	499.3190
Linear	17192.1	13	1322.47
Concurrence	693.196	1	693.196
Nonconcurrence	16498.9	12	1374.91
Deviation from Linear	34737.1	91	381.726
Replication Within Cells	24034.5137	252	95.3751

TABLE B-VIII. COMPONENTS OF VARIANCE AND THEIR RELATIVE IMPORTANCE FOR THE LINEAR MODEL ANALYSIS. COMPONENTS ARE EXPRESSED AS STANDARD DEVIATIONS IN THE ORIGINAL SCALE.

Concentration, $\mu g/m^3$	Source of Variation								Total† Standard Deviation
	Within-Laboratory				Between-Laboratory				
	Replication		λ-Variation		μ-Variation		δ-Variation		
	Standard Deviation	Percent*	Standard Deviation	Percent*	Standard Deviation	Percent*	Standard Deviation	Percent*	
10	6.9	6	12.8	21	19.5	49	13.8	24	27.9
20	7.0	6	12.9	21	19.9	50	13.5	23	28.2
50	7.3	6	13.5	22	21.0	53	12.6	19	28.9
100	7.8	7	14.4	23	22.9	57	11.1	13	30.3
150	8.3	7	15.3	23	24.9	61	9.4	9	31.8
200	8.8	7	16.2	23	16.9	65	7.5	5	33.5
275	9.5	7	17.5	23	30.0	68	4.5	2	36.3
500	11.7	6	21.6	21	39.5	71	6.0	2	46.9
700	13.7	5	25.1	18	48.3	68	17.1	8	58.7
820	14.8	5	27.3	17	53.8	65	24.4	13	66.7
1000	16.6	4	30.5	15	62.1	61	36.1	20	79.8

*Percent of total variance.
†Based on a single determination per laboratory including all sources of variation.

is large compared to the total within-laboratory variability throughout the table. $V(\delta)$ is generally small compared to $V(\mu)$ throughout the intermediate part of the range but becomes appreciable at both the high and the low ends. This is most likely related to the poor readability of absorbance values above approximately 0.8 absorbance unit on most spectrophotometers. No speculation can be made regarding the lower end since results below 150 $\mu g/m^3$ are extrapolated. It is also possible that the calibration curves, absorbance versus concentration, may deviate from linearity at each end. This will be examined superficially in a later section.

The repeatability and reproducibility must now be defined and computed. The repeatability is defined as the square root of the total within-laboratory variance as

$$\text{Repeatability} = \sqrt{V(\epsilon) + V(\lambda)} \qquad \text{(B-7)}$$

The reproducibility is defined as the square root of the total variance as

$$\text{Reproducibility} = \sqrt{V(\epsilon) + V(\lambda) + V(\mu) + V(\delta)} \qquad \text{(B-8)}$$

These parameters, in the original scale, are functions of the magnitude of the measurement just as in the preceding analysis. However, in this case, the relationships are nonlinear and are not easily expressed in terms of the original variables. For comparison with the other estimates, see Figure B-3 in the following subsection.

D. Comparison of Methods and Discussion of Results

In this subsection, the three methods described above will be compared and certain inferences will be made. However, it is not within the scope of this report to make a detailed comparison in more complex statistical terms. It is more appropriate to review the results from a practical viewpoint.

The comparison is best made by referring to the results in graphic form shown in Figure B-3. This figure compares the replication error, the repeatability, and the reproducibility for each method of statistical analysis. First, the overall agreement between these methods is very good. The agreement for the replication error is exact because the methods have this much in common. The point estimates seem to imply a minimum for repeatability and reproducibility in the midrange of concentration. The other methods, because of their fundamental assumptions, do not recognize any minimum. In this respect, the results are inconclusive, and an experiment incorporating many more intermediate concentrations would be required to verify such a condition. In consideration of the optimum absorbance range of most spectrophotometers, such a minimum could be entirely possible.

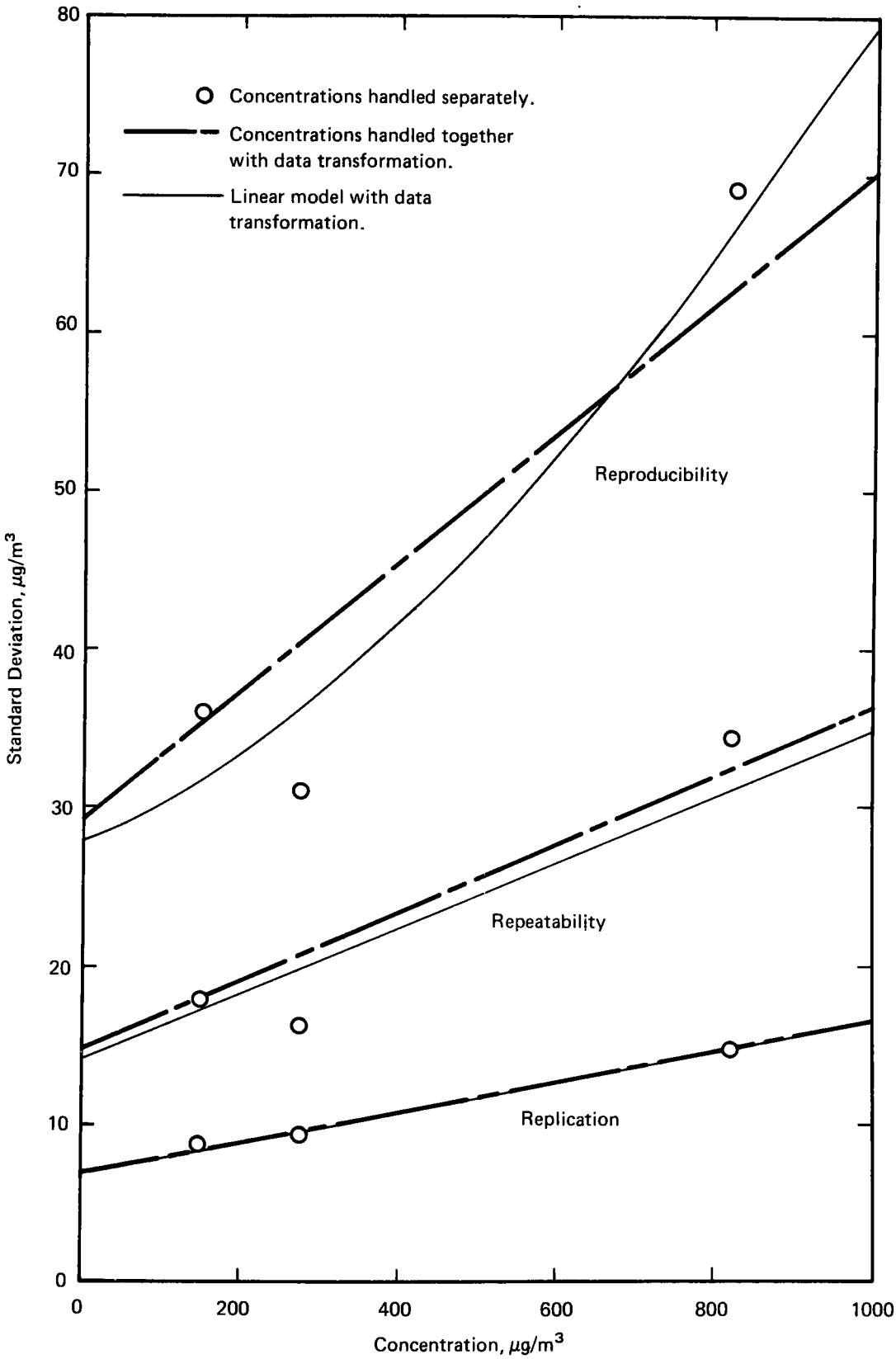


FIGURE B-3. REPLICATION ERROR, REPEATABILITY, AND REPRODUCIBILITY VERSUS CONCENTRATION FOR THREE DIFFERENT METHODS OF DATA ANALYSIS.

The point estimates are therefore of limited use since we must also make inferences between these points. The other two methods of analysis provide for objective estimates at other concentrations. The second method, the analysis of variance with concentrations handled together, enjoys the advantage of a simple expression for the standard deviations as a function of concentration. The linear model expresses the standard deviations as a function of concentration; however, there are also additional parameters derived from the collaborative test data and the function is more complex. In view of the relatively large repeatability and reproducibility, this complexity is not considered to be justified.

The linear model has an advantage because of its ability to describe and compare the various sources of error more thoroughly. As mentioned previously, the linear model is less sensitive to outlying observations than the other methods.

On the basis of the preceding discussion, we arrive at the conclusion that the analysis of variance handling all concentrations together with data transformation offers the most convenient and practical method of expressing the replication error, the repeatability, and the reproducibility as a function of concentration. All subsequent treatment of these parameters will be made accordingly.

IV. APPLICATION OF THE RESULTS

We are now in a position to apply the results of the previous section and answer some fundamental questions–thus fulfilling the objectives of this collaborative test. Unless otherwise stated below, a 95 percent level of significance is assumed. Let us also reclarify that the results apply for 30-min sampling and the use of the calibration procedure with sulfite solution.

A. Precision of the Method

We may use the expressions for the replication error (σ_ϵ), the within-laboratory (single-replicate, multiple-day) variation (repeatability) (σ_D), and the between-laboratory (single-replicate, single-day, single-analyst) variation (σ_L) from Figure B-3 which are restated as follows:

$$\sigma_\epsilon = (0.7 + 0.001y)(10) \tag{B-9}$$

$$\sigma_D = (0.7 + 0.001y)(21) \tag{B-10}$$

$$\sigma_L = (0.7 + 0.001y)(41) \tag{B-11}$$

where y is the concentration in $\mu g/m^3$. All statements regarding the precision of the method are derived from these expressions. With these equations, the precision for any desired case can be computed. Some of the more useful cases are shown below.

In the application of the results to test class means, we will resort to the studentized range.[17-20] If an estimate of the standard deviation (σ) is based on ν degrees of freedom and is independent of the class means to be compared, and if these class means are computed from N cases and selected from a group of g means, then the 0.05 allowance (95 percent confidence level) for any comparison is

$$\left|\bar{x}_1 - \bar{x}_2\right|_{max} = q_{0.05(g,\nu)}\, \sigma/\sqrt{N} \tag{B-12}$$

where $\bar{x}_1$ is the highest mean and $\bar{x}_2$ is the lowest class mean. Our interest will center around $g = 2$ because we will, in general, be interested in comparing two class means. The degrees of freedom ν will be taken from those corresponding to the independent estimates of the standard deviation. The value of N represents the number of observations that make up the means. In computing checking limits for duplicates, N is of course equal to one and the test is identical to ASTM recommended practice.[21]

An obvious limitation is that the means must all contain the same number of observations. When this is not the case, the standard normal deviate is appropriate and we shall use[22]

$$\left|\bar{x}_1 - \bar{x}_2\right|_{max} = 1.96\, \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \tag{B-13}$$

where $|\bar{x}_1 - \bar{x}_2|$ is the absolute value of the difference in the two class means $\bar{x}_1$ and $\bar{x}_2$, σ is an independent estimate, and N_1 and N_2 are the numbers of observations in $\bar{x}_1$ and $\bar{x}_2$, respectively. The results from this equation are the same as Equation (B-12) for $N = N_1 = N_2$ and ν is large. The results are adequate if N_1 and N_2 are relatively large (20 or more).

To test whether the true value of a mean is lower than a specified fixed value, the maximum permissible difference is[23]

$$(\bar{x} - \mu_0)_{max} = -1.645\sigma/\sqrt{N} \qquad \text{(B-14)}$$

which is a one-sided test (at the 95 percent level of confidence) where $\bar{x}$ is the mean, μ_0 is the fixed value, σ is again an independent estimate, and N is the number of observations in $\bar{x}$.

These techniques will be applied as appropriate to the three sources of variation below. The treatment will be in more depth for the precision between laboratories, which is of more practical interest.

1. Precision Between Replicates

We have already concluded that replication will not materially assist in increasing the precision of the method. Replication will, in general, be a waste of time and effort. Nevertheless, we would not be thorough if we did not consider the measure of acceptability of replicates. The expression from Equations (B-9) and (B-12) for the checking limits for duplicates is

$$R_{max} = (2.77)(0.7 + 0.001y)(10) \qquad \text{(B-15)}$$

where R_{max} is the maximum permissible range between duplicates. Two such replicates should be considered suspect if they differ by more than R_{max}.

It should be noted that agreement between duplicates better than 5 percent cannot be expected below 900 $\mu g/m^3$. Agreement better than 10 percent cannot be expected below 300 $\mu g/m^3$, and agreement better than 20 percent cannot be expected below 100 $\mu g/m^3$

2. Precision Between Days

In some instances, it will be necessary to compare observations made by the same analyst on different days. The following expression from Equations (B-10) and (B-12) allows this comparison (single-replicate):

$$R_{max} = (2.82)(0.7 + 0.001y)(21) \qquad \text{(B-16)}$$

where R_{max} is the maximum permissible range between two observations. Two such values may not be considered to belong to the same population if they differ by more than R_{max}. Conversely, the two values are not significantly different if they differ by less than R_{max}.

It can now be noted that the method cannot detect a difference of 10 percent between two such values in the range of 0 to 1000 $\mu g/m^3$. A difference of 20 percent may be detected above 300 $\mu g/m^3$.

As an example of the futility of replication, the factor 21 in Equation (B-16) would be reduced to 20 for duplicates, approximately the same for triplicates, and to 19 for an infinite number of replicates.

There may also be some occasions where it will be necessary to compare the means for each of two given sampling stations, where each mean was obtained by the same analyst, and consisted of a known number of single-replicate observations. The number of observations in each mean will not usually be equal. Their standard deviations will not usually be equal, and one or both may not be normally distributed. Where they are normally distributed, standard tests such as the t-test[24] may be applied.

A limiting case may be investigated if we assume that two means $\bar{x}_1$ and $\bar{x}_2$ are normally distributed with $\sigma_1 = \sigma_2 = \sigma_D$, where σ_D is given by

Equation (B-10). This is an unlikely, if not impossible, situation which could only result from absolutely constant concentrations at each of the sampling stations. Under these assumptions, we may apply Equations (B-10) and (B-13) and obtain

$$R_{max} = 1.96(0.7 + 0.001\bar{x}_1)(21)\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \quad \text{(B-17)}$$

where R_{max} is the maximum permissible range between means $\bar{x}_1$ and $\bar{x}_2$ containing N_1 and N_2 observations, respectively. If the range exceeds R_{max}, the means are significantly different and do not belong to the same population.

Under the same limiting assumptions, we may compare a mean $\bar{x}$ containing N observations with some fixed value μ_0 and be able to state whether the true value of $\bar{x}$ is less than μ_0. Equations (B-10) and (B-14) may be applied to this case resulting in

$$R_{max} = -1.645(0.7 + 0.001\mu_0)(21)/\sqrt{N} \quad \text{(B-18)}$$

where R_{max} is the maximum permissible range between $\bar{x}$ and μ_0. If $\bar{x} - \mu_0$ is less than R_{max}, then the true value of $\bar{x}$ is less than μ_0.

The variance of the values making up a mean can be compared to the variance above by an approximate F-test[25-27] of the variance of the sample data. The data must be transformed according to Equation (B-2). The denominator for the test may be obtained from Table B-V. An F-ratio below the critical value would indicate that all variation could be accounted for by the random variation of the analytical method. In other words, the set of data could have resulted from the repeated analysis of a sample whose true value was equal to the mean of the sample distribution.

3. Precision Between Laboratories

Probably the most frequent comparison to be made will be that involving observations of two different laboratories. The following expression from Equations (B-11) and (B-12) allows this comparison for a single-replicate single-analyst single-day:

$$R_{max} = (3.06)(0.7 + 0.001y)(41) \quad \text{(B-19)}$$

where R_{max} is the maximum permissible difference between the observations of two different laboratories. Two such values may not be considered to belong to the same population if they differ by more than R_{max}. Conversely, the two values are not significantly different if they differ by less than R_{max}.

It can be seen that the method cannot detect a difference of less than 20 percent between single observations of two laboratories. At a level of 100 $\mu g/m^3$, a difference of less than 100 percent is not detectable. It will, in general, be of limited usefulness to compare single observations of two laboratories.

Frequently, it will be necessary to compare the means for each of two given sampling stations. Each mean may be the result of observations by one or more different laboratories. Each mean may contain a different number of observations, each a single-determination. Their standard deviations will not usually be equal, and one or both may not be normally distributed. Where they are normally distributed, standard tests such as the t-test[24] may be applied.

Similar to the preceding subsection, a limiting case may be investigated if we assume that the two means $\bar{x}_1$ and $\bar{x}_2$ containing $N_1 = N_2 = N$ observations are normally distributed with $\sigma_1 = \sigma_2 = \sigma_L$, where σ_L is given by Equation (B-11). Here again, this is an unlikely, if not impossible, situation which could only result from absolutely constant concentrations at each sampling station. Nevertheless, a certain amount of guidance can be derived. If we apply Equations (B-11) and (B-12) to this case, we obtain

$$R_{max} = (3.06)(0.7 + 0.001\bar{x}_1)(41)/\sqrt{N} \quad \text{(B-20)}$$

where R_{max} is the maximum permissible range between the means $\bar{x}_1$ and $\bar{x}_2$. If the range exceeds

R_{max}, the means are significantly different and do not belong to the same population.

It is interesting to pursue this line of reasoning further in terms of the number of samples required to detect a specified difference under the limiting assumptions. Rearranging Equation (B-20) and solving for N, we obtain:

$$N = \left[\frac{(3.06)(0.7 + 0.001y)(41)}{R}\right]^2 \qquad \text{(B-21)}$$

This expression now gives the minimum number of observations (N) for any desired agreement (R) between two means at any level of concentration (y). These results are best illustrated in Figure B-4. This figure shows the agreement versus the concentration level for a family of sample sizes. This presentation is most convenient because of its linearity. Superimposed on the curve are percentage agreement lines for comparison purposes. For example, if we desired agreement better than 5 percent at a concentration of 300 $\mu g/m^3$, a minimum of 70 observations would be required. This figure may not be used for agreement of a mean with a fixed value.

Under the same assumptions as above, with the exception that N_1 may not equal N_2 but both are relatively large, Equations (B-11) and (B-13) are used, yielding

$$R_{max} = 1.96(0.7 + 0.001\bar{x}_1)(41)\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \qquad \text{(B-22)}$$

where R_{max} is the maximum permissible range between $\bar{x}_1$ and $\bar{x}_2$. If the range exceeds R_{max}, the means are significantly different and do not belong to the same population.

Under the same limiting assumptions, we may compare a mean $\bar{x}$ containing N observations with some fixed value μ_0 and be able to state whether the true value of $\bar{x}$ is less than μ_0. We again utilize Equations (B-11) and (B-14) for this type case yielding

$$R_{max} = -1.645(0.7 + 0.001\mu_0)(41)/\sqrt{N} \qquad \text{(B-23)}$$

where R_{max} is the maximum permissible range between $\bar{x}$ and μ_0. If $\bar{x} - \mu_0$ is less than R_{max}, then the true value of $\bar{x}$ is less than μ_0.

We may rearrange Equation (B-23) and solve for N obtaining

$$N = \left[\frac{1.645(0.7 + 0.001\mu_0)(41)}{R}\right]^2 \qquad \text{(B-24)}$$

This equation is exactly analogous to Equation (B-21). N is the minimum number of observations required to attain the agreement R under the limiting assumptions. Figure B-5, which is analogous to Figure B-4, best illustrates the resulting relationships. For example, a minimum of 20 observations would be required to establish that the true value of $\bar{x}$ is less than 300 $\mu g/m^3$, while the actual value is 285 $\mu g/m^3$ (a 5 percent difference). Stated differently, we may say that, given a set of 20 observations with a mean of 285 $\mu g/m^3$, we may be 95 percent confident that the true mean is less than 300 $\mu g/m^3$.

Analogous to the preceding subsection, the variance of the values making up a mean can be compared to the variance above by the approximate F-test. The denominator for the test is obtained from Table B-V and the numerator is the variance of the sample data, for the data transformed according to Equation (B-2). Just as in the preceding subsection, an F-ratio below the critical value would indicate that all variation could be accounted for by the random variation of the analytical method.

B. Lower Limit of Detection

We have previously made two independent estimates of the lower limit of detection. The first was based on calibration curve data (see Section III-C-1 of the main report) and is equal to 22 $\mu g/m^3$. The second was based on control sample data (see Section III-C-2 of the main report) and is equal to 26 $\mu g/m^3$ A third is possible, based on two standard deviations (replication), which we shall estimate from the data at the lowest concentration tested to be 18 $\mu g/m^3$ These are in very good

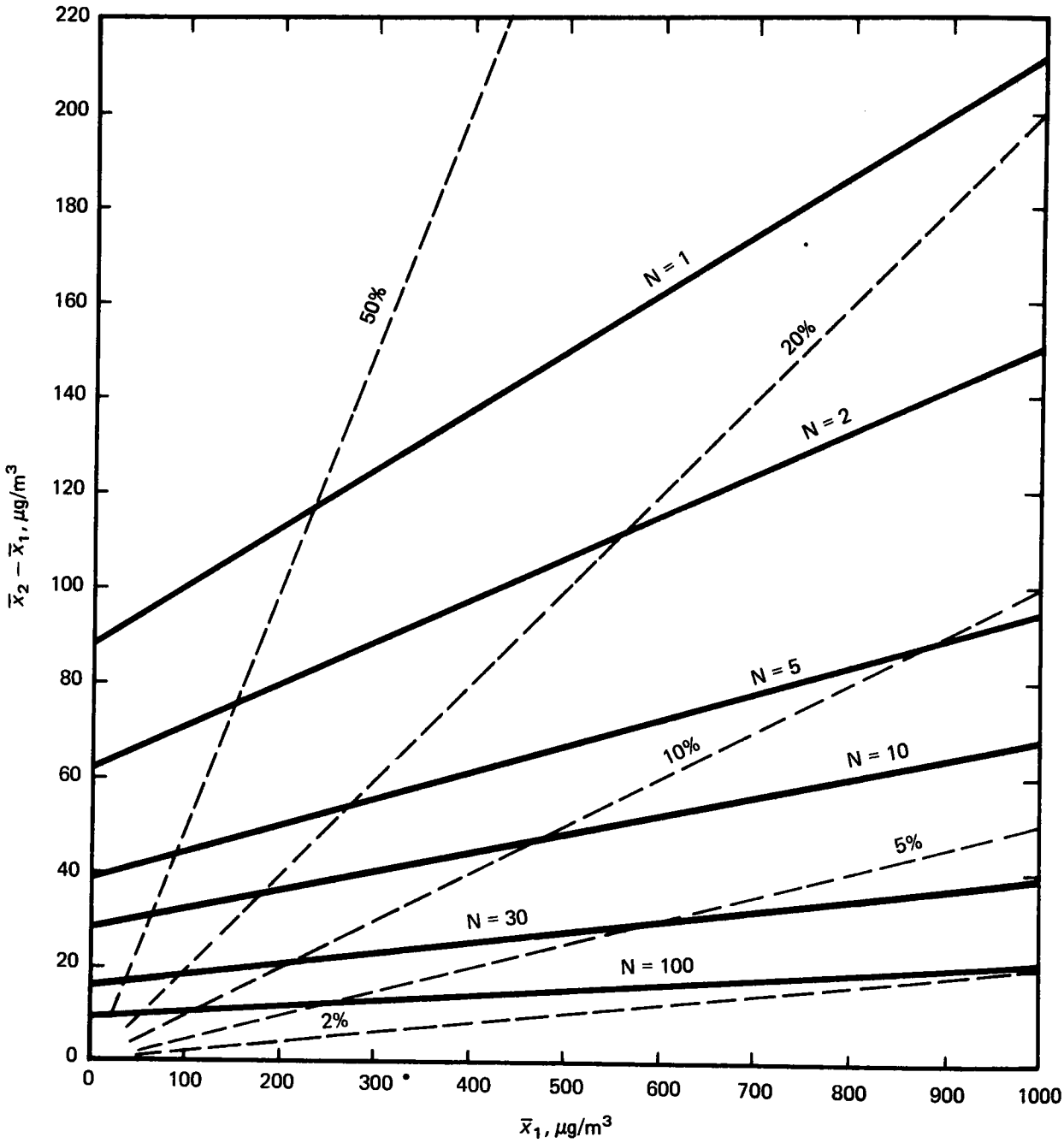


FIGURE B-4. EXPECTED AGREEMENT BETWEEN TWO MEANS VERSUS CONCENTRATION FOR VARIOUS NUMBERS OF OBSERVATIONS (95 Percent Level of Significance). EACH MEAN HAS N OBSERVATIONS WITH A STANDARD DEVIATION EQUAL TO $(0.7 + 0.001\bar{x}_1)$ (41).

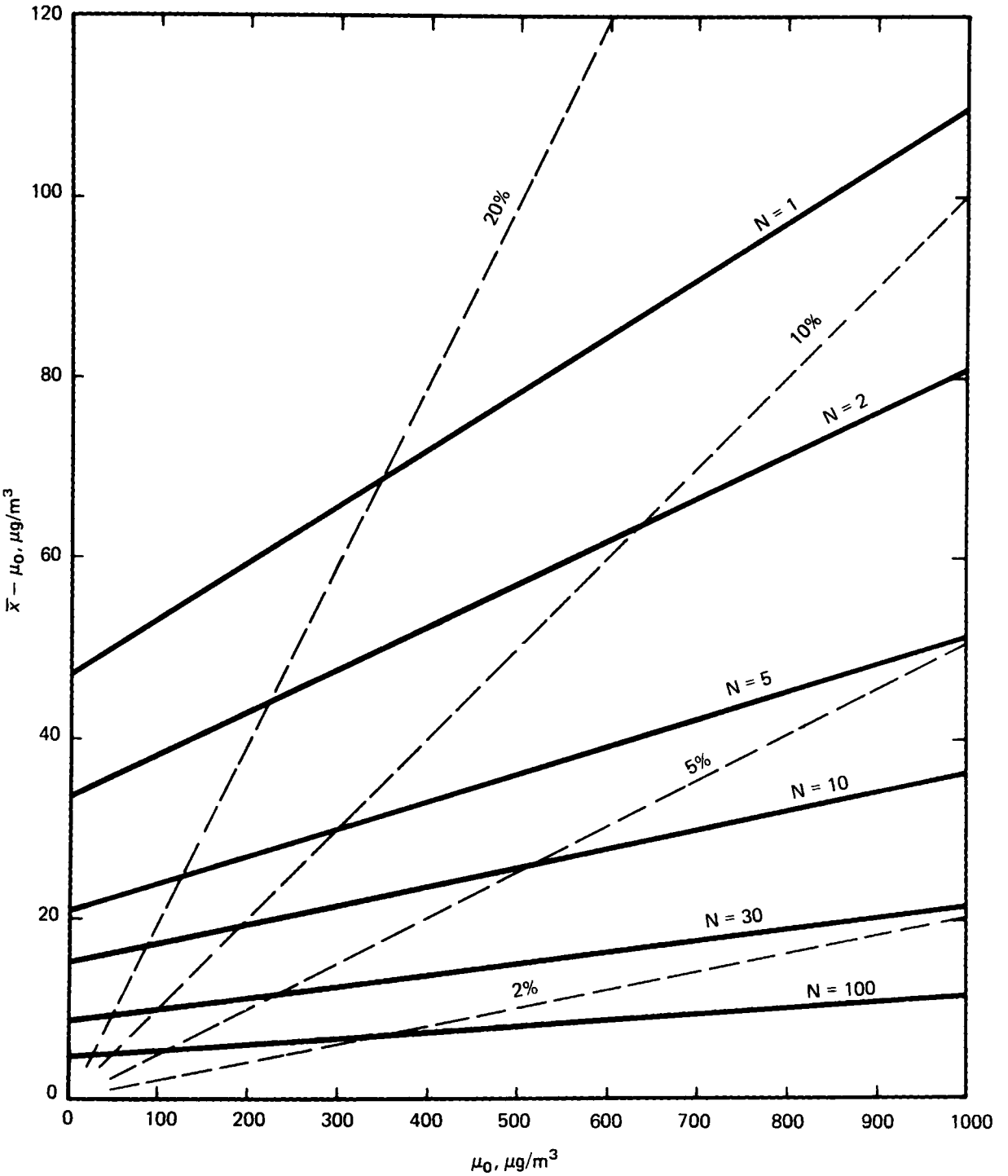


FIGURE B-5. EXPECTED AGREEMENT BETWEEN A MEAN AND A FIXED VALUE VERSUS CONCENTRATION FOR VARIOUS NUMBERS OF OBSERVATIONS (95 Percent Level of Significance). THE MEAN HAS N OBSERVATIONS WITH A STANDARD DEVIATION EQUAL TO $(0.7 + 0.001\mu_0)$ (41).

agreement, and it is not too important which one is used; therefore, a value of 25 $\mu g/m^3$ is proposed as a practical figure. A single observation less than this value is not distinguishable from zero. Whether the mean of several observations, each a single determination, is significantly different from zero is dependent upon the number of observations and their distribution, regardless of the magnitude of the mean.

Recorded results using this method should carry no more than two significant digits. Originators or recorders of data should assume the responsibility of appending confidence limits (95 percent) to their data.

C. Accuracy and Bias

The overall average deviations from the expected values for each concentration tested were 6.4, 2.2, and -23.9 $\mu g/m^3$ for concentrations of 150, 275, and 820 $\mu g/m^3$, respectively. These differences are not significant, and therefore no systematic error, bias, or inaccuracy was detectable.

LIST OF REFERENCES

1. ASTM Manual for Conducting an Interlaboratory Study of a Test Method, ASTM STP No. 335, *Am. Soc. Testing & Mats.* (1963).

2. **Handbook of the AOAC**, Second Edition, October 1, 1966.

3. **1968 Book of ASTM Standards**, Part 30, Recommended Practice for Developing Precision Data on ASTM Methods for Analysis and Testing of Industrial Chemicals, ASTM Designation: E180-67, pp 459-480.

4. **1968 Book of ASTM Standards**, op cit, Recommended Practice for Dealing with Outlying Observations, ASTM Designation: E178-68, pp 437-439.

5. Ibid, pp 444-447.

6. Dixon, W. J., (Ed.), BMD Biomedical Computer Programs, Second Edition, University of California Press, Berkeley and Los Angeles, pp 586-600 (1968).

7. Southwest Research Institute, Houston, Texas, Computer Program OUTLY, for test for Outlying Observations, Unpublished (1971).

8. Dixon, Wilfred J., and Massey, Frank J., Jr., **Introduction to Statistical Analysis**, McGraw-Hill Book Company, Inc., New York, Chapter 10, pp 179-180 (1957).

9. Ibid, p 180.

10. Nelson, Benjamin N., "Survey and Application of Interlaboratory Testing Techniques," *Industrial Quality Control*, Vol. 23, pp 554-559 (May 1967).

11. McArthur, D. S., Baldeschwieler, E. L., White, W. H., and Anderson, J. S., "Evaluation of Test Procedures," *Analytical Chemistry*, **26**, pp 1012-1018 (1954).

12. Southwest Research Institute, Houston, Texas, Computer Program COMPO for Computing Variance Components, Unpublished (1971).

13. Scheffé, Henry, **The Analysis of Variance**, John Wiley and Sons, Inc., New York, Chapter 7, pp 231-235 (1959).

14. **1968 Book of ASTM Standards**, op cit, p 476.

15. Mandel, J., "The Measuring Process," *Technometrics*, **1**, pp 251-267 (1959).

16. Mandel, J., and Lashof, T. W., "The Interlaboratory Evaluation of Testing Methods," *ASTM Bulletin 239*, pp 53-61 (1959).

17. Duncan, Acheson J., **Quality Control and Industrial Statistics**, Third Edition, Richard D. Irwin, Inc., Homewood, Illinois, Chapter XXXI, pp 632-636 (1965).

18. Ibid, p 909.

19. Bennett, Carl A., and Franklin, Norman L., **Statistical Analysis in Chemistry and the Chemical Industry**, John Wiley and Sons, Inc., New York, 1954, Chapter 4, p 111.

20. Ibid, p 185-189.

21. **1968 Book of ASTM Standards**, op cit, p 476.

22. Dixon and Massey, op cit, p 120.

23. Ibid, pp 114-115.

24. Ibid, pp 123-124.

25. Ibid, pp 106-107.

26. Bennett and Franklin, op cit, pp 192-196.

27. Duncan, op cit, pp 511-517.

APPENDIX C

TABULATION OF ORIGINAL DATA

TABLE C-I. OBSERVED VALUES FOR EACH REPLICATE FOR EACH CONCENTRATION FOR EACH DAY FOR EACH LABORATORY, MICROGRAMS PER CUBIC METER.

Laboratory Code Number	Day 1			Day 2			Day 3		
Low Concentration									
271	145	135	130	147	141	139	134	135	136
274	161	161	165	155	160	165	156	162	172
305	144	148	142	143	146	140	147	145	139
345	169	190	180	181	193	198	149	160	137
500	176	151	151	135	135	135	146	158	143
509	92	92	90	143	134	141	111	116	111
526	114	114	112	115	121	120	122	110	114
571	164	152	140	202	168	156	143	137	137
578	138	131	131	127	119	127	145	141	136
655	253	277	273	233	244	244	232	253	235
788	171	141	154	245	212	185	135	134	131
920	150	151	140	127	128	126	147	137	137
926	173	168	173	7	12	12	184	180	180
927	144	144	144	135	145	135	150	140	140
Intermediate Concentration									
271	247	249	248	238	241	238	242	247	243
274	306	313	313	296	302	310	325	325	325
305	270	270	270	264	267	265	256	264	256
345	264	281	259	251	298	309	266	277	271
500	294	291	298	272	281	272	280	283	325
509	223	223	220	242	251	251	238	225	232
526	220	220	225	229	226	224	231	238	240
571	282	272	270	247	253	265	268	255	249
578	278	262	261	238	241	245	256	259	259
655	352	341	352	322	313	332	314	321	320
788	375	329	322	322	308	315	296	270	264
920	258	263	254	250	248	248	255	252	263
926	303	299	299	301	299	292	338	338	331
927	264	262	264	246	256	256	261	261	261
High Concentration									
271	731	722	729	744	747	741	723	716	728
274	884	884	933	903	903	903	931	931	955
305	772	783	772	738	748	733	749	751	758
345	826	804	777	852	981	992	771	782	782
500	848	809	821	821	809	779	776	788	776
509	707	709	707	764	764	710	730	728	741
526	668	668	668	696	709	734	706	690	712
571	843	806	794	760	751	717	757	713	729
578	786	783	789	651	661	663	684	677	696
655	895	884	895	836	859	836	851	862	862
788	865	819	793	814	800	787	747	747	721
920	770	775	765	761	758	766	778	791	771
926	908	908	928	901	903	917	987	987	987
927	749	749	759	757	757	757	762	772	772

TABLE C-II. EXPECTED VALUES FOR EACH REPLICATE FOR EACH CONCENTRATION FOR EACH DAY FOR EACH LABORATORY, MICROGRAMS PER CUBIC METER.

Laboratory Code Number	Day 1			Day 2			Day 3		
Low Concentration									
271	143	143	143	143	143	143	143	143	143
274	154	154	154	153	153	153	153	153	153
305	148	148	148	148	148	148	149	149	149
345	150	150	150	150	150	150	150	150	150
500	154	154	154	154	154	154	154	154	154
509	146	146	146	146	146	146	146	146	146
526	141	140	140	139	138	138	139	139	139
571	140	140	140	142	142	142	141	140	140
578	146	146	146	145	145	145	145	145	145
655	144	144	144	145	147	145	145	145	145
788	155	155	155	155	155	155	154	154	154
920	146	146	146	147	147	147	147	147	147
926	149	149	149	149	149	149	149	150	150
927	149	149	149	148	150	150	148	148	148
Intermediate Concentration									
271	263	263	263	262	262	262	262	262	262
274	284	284	284	282	282	282	280	280	280
305	273	273	273	272	272	272	274	274	274
345	277	277	277	277	277	277	277	277	277
500	284	284	284	283	283	283	283	283	283
509	268	268	268	268	268	268	268	268	268
526	258	258	260	254	256	254	256	257	256
571	258	258	258	262	260	260	261	261	259
578	268	268	268	266	266	266	266	266	266
655	265	265	265	269	269	269	266	266	266
788	284	284	284	286	286	286	284	284	284
920	270	270	270	270	270	270	270	270	270
926	273	273	273	274	274	274	276	276	276
927	274	274	274	275	275	275	273	275	275
High Concentration									
271	789	789	789	784	784	784	785	785	785
274	847	847	847	843	843	843	837	837	837
305	816	816	816	812	812	812	822	822	822
345	827	827	827	827	827	827	827	827	827
500	848	848	848	846	846	846	846	846	846
509	802	802	802	798	798	798	802	802	802
526	771	777	771	760	760	760	764	764	764
571	771	771	771	769	775	769	774	774	768
578	800	800	800	795	795	795	795	795	795
655	792	792	792	805	805	805	795	795	795
788	849	849	849	854	854	854	848	848	848
920	804	804	804	807	807	807	807	807	807
926	817	817	817	827	833	833	823	823	817
927	818	818	818	822	822	822	815	815	821

TABLE C-III. CALIBRATION CURVE DATA FOR EACH DAY FOR EACH LABORATORY. SLOPE IN ABSORBANCE PER MICROGRAM OF SULFUR DIOXIDE DETERMINED BY LEAST SQUARES. OTHER PARAMETERS IN ABSORBANCE UNITS.

Laboratory Code Number	Day	Slope	Intercept	Intercept – Zero Standard	Standard Error of Estimate
271	1	0.0298	0.178	−0.002	0.0125
	2	0.0288	0.182	0.019	0.0151
	3	0.0293	0.159	0.002	0.0053
274	1	0.0327	0.169	0.005	0.0044
	2	0.0335	0.172	0.004	0.0037
	3	0.0305	0.185	0.008	0.0060
305	1	0.0293	0.181	0.001	0.0037
	2	0.0300	0.170	−0.001	0.0034
	3	0.0299	0.169	0.002	0.0035
345	1	0.0322	0.149	−0.001	0.0099
	2	0.0298	0.160	0.000	0.0023
	3	0.0303	0.163	0.006	0.0042
500	1	0.0282	0.141	0.009	0.0094
	2	0.0282	0.134	0.009	0.0090
	3	0.0279	0.117	0.002	0.0048
509	1	0.0300	0.175	−0.002	0.0021
	2	0.0303	0.160	0.007	0.0083
	3	0.0300	0.195	−0.001	0.0047
526	1	0.0312	0.149	−0.004	0.0028
	2	0.0306	0.158	−0.006	0.0038
	3	0.0302	0.150	−0.013	0.0094
571	1	0.0285	0.214	0.004	0.0121
	2	0.0289	0.224	0.004	0.0048
	3	0.0291	0.164	−0.001	0.0037
578	1	0.0271	0.142	−0.003	0.0057
	2	0.0304	0.159	−0.001	0.0041
	3	0.0305	0.161	0.001	0.0017
655	1	0.0299	0.188	0.006	0.0208
	2	0.0302	0.184	0.002	0.0038
	3	0.0303	0.188	0.007	0.0054
788	1	0.0316	0.148	0.006	0.0165
	2	0.0311	0.165	0.010	0.0116
	3	0.0326	0.140	−0.005	0.0062
920	1	0.0299	0.146	0.002	0.0071
	2	0.0293	0.143	0.002	0.0018
	3	0.0293	0.140	0.005	0.0038
926	1	0.0248	0.133	−0.007	0.0106
	2	0.0254	0.141	0.002	0.0018
	3	0.0249	0.146	0.002	0.0024
927	1	0.0319	0.174	−0.001	0.0062
	2	0.0319	0.157	−0.008	0.0151
	3	0.0318	0.158	0.003	0.0034

TABLE C-IV. CONTROL SAMPLE DATA, MICROGRAMS OF SULFUR DIOXIDE.

Laboratory Code Number	Day	Taken	Found	Difference
271	1	15.14	16.19	1.05
	2	15.14	15.85	0.71
	3	15.14	15.59	0.45
274	1	15.64	15.89	0.25
	2	14.66	14.34	−0.32
	3	15.06	14.34	−0.72
305	1	14.04	14.50	0.46
	2	13.90	14.10	0.20
	3	13.88	14.13	0.25
345	1	14.18	14.45	0.27
	2	15.46	14.96	−0.50
	3	13.80	13.19	−0.61
500	1	15.56	15.80	0.24
	2	15.60	16.59	0.99
	3	15.92	16.66	0.74
509	1	15.44	15.29	−0.15
	2	15.54	15.66	1.12
	3	15.48	14.83	−0.65
526	1	15.30	15.27	−0.03
	2	15.50	15.25	−0.25
	3	13.60	12.53	−1.07
571	1	15.30	16.14	0.84
	2	15.30	15.77	0.47
	3	15.40	15.48	0.08
655	1	13.85	13.63	−0.22
	2	14.16	13.83	−0.33
	3	12.56	13.19	0.63
920	1	16.00	16.53	0.53
	2	15.68	15.78	0.10
	3	15.70	15.90	0.20
926	1	6.72	6.94	0.22
	2	6.88	6.98	0.10
	3	8.24	8.39	0.15
927	1	14.60	14.87	0.27
	2	14.20	14.27	0.07
	3	14.80	14.77	−0.03

TABLE C-V. REAGENT BLANK DATA.
ABSORBANCE UNITS.

Laboratory Code Number	Day 1	Day 2	Day 3
271	0.198	0.164	0.158
274	0.168	0.168	0.165
305	0.185	0.170	0.167
345	0.153	0.147	0.158
500	0.138	0.131	0.116
509	0.164	0.170	0.220
526	0.153	0.163	0.160
571	0.210	0.210	0.180
578	0.145	0.160	0.160
655	0.182	0.182	0.181
788	0.145	0.125	0.180
920	0.144	0.141	0.133
926	0.140	0.139	0.144
927	0.175	0.165	0.160