Under [illegible] as amended, the U.S. Environmental Protection Agency (EPA), on December 23, 1971, promulgated its first group of new source performance standards (NSPS), which placed restrictions on the allowable emissions from new plants in five industrial categories. These were followed in 1974 by standards for seven additional industries, and standards covering several others have now either been promulgated or are in varying stages of development. Moreover, four substances (asbestos, beryllium, mercury, and vinyl chloride) have been designated as hazardous air pollutants, and emission standards for the first three were promulgated in 1973 under Section 112 of the Clean Air Act.

Of fundamental importance to enforcement of the above standards is the measurement process. At the time that the initial NSPS were established, many of the measurement methods used to determine compliance with these standards had not been fully evaluated, nor had their precision, accuracy, and general reliability in the hands of typical users been determined. It is for this reason that the Quality Assurance Branch, Environmental Monitoring and Support Laboratory, EPA, has for the past three years been engaged in a systematic program to standardize or validate those source test methods that will be used to determine compliance with these federal emission standards.

Collaborative tests have been, and are being developed to establish monitoring procedures for keeping tabs on Clean Air Act compliance

M. Rodney Midgett
U.S. Environmental Protection Agency
Research Triangle Park, N.C. 27711

Traditionally, the within-laboratory and between-laboratory precision of test methods is determined through collaborative testing (round-robin testing). The collaborative test is designed so that each participant makes one or more measurements on identical samples by use of the same test method. Then, from a statistical analysis of the results, an estimate is made of the within-laboratory and between-laboratory precision of the test method. This general technique has been used very widely for the validation of methods for the analysis of such materials as water, drugs, food and agricultural products, fertilizers, coal, and ores.

Experience has shown that before a stationary source test method can be successfully collaboratively tested, it must be described in sufficient detail to ensure that each collaborator uses exactly the same sampling and analysis procedures. Further, it must give repeatable results when one laboratory analyzes the same sample several times. This repeatability can be assured only through intensive method evaluation, which now constitutes a large portion of the total program. This evaluation, followed by collaborative testing, is resulting in more fully described methods of known precision and accuracy, and of proven reliability.

The standardization process

The validation of source test methodology is a complex, lengthy, and costly process, but years of experience have indicated the need for a complete and systematic examination even for those methods and measurement principles with fairly extensive histories of usage. Basically, this examination consists of the following steps:

- The method is examined for technical accuracy, clarity, and completeness of detail.

- The method is subjected to a thorough and rigorous laboratory evaluation, which may include investigations of sample collection efficiency, applicable concentration range, mode of calibration, and effects of interferences.
- The method receives field evaluation at an applicable test site; statistically designed experiments, with novel and original evaluation techniques to determine its performance under typical field conditions, are often performed.
- Finally, the method is submitted to an interlaboratory collaborative test at an appropriate test site. Qualified participants determine its precision, accuracy, and field reliability. Based on the test results and other information gained from the test, a final draft of the method is prepared and recommendation is made for its adoption by the agency.

Collaborative test design

For collaborative tests of a stationary source emission method, both the sampling and the analytical procedures must be evaluated, usually at a real source representative of those where the method will be used. All participants must have access to the same pollutant concentration in the stack, for, if they cannot obtain identical samples, they surely will not get reproducible results. For gaseous pollutants, this can frequently be accomplished by extracting a side stream from the stack and piping it to ground level, where it is delivered through a manifold to the collaborators who simultaneously sample the gaseous pollutant.

Collaborative testing of methods for pollutants that exist in particulate form is complicated by the requirement that all test teams sample the material isokinetically directly from the stack. Here the problem becomes one of the simultaneous extraction of representative samples from the stack by each of the collaborative test teams. Since spatial and temporal variations may constantly be occurring in both the velocity profile and the pollutant profile, an attempt must be made to compensate for this so that each participant has access to statistically identical or equivalent samples.

Because they could not ensure for statistically identical or equivalent samples, previous test designs were considered imperfect, and often included source variability in the precision estimates. Thus, a different approach was sought for collaborative testing methods for pollutants that exist in particulate form. The result was a new test design that uses paired sampling trains in which two probe-pitot tube assemblies could simultaneously sample at very nearly the same point in the stack. Since the paired probe tips sample in rather close proximity, the effects of spatial and temporal stack variation on the samples collected by the adjacent probes are greatly minimized. This test design employed six independent test teams operating separate trains in three of the paired train systems for the entire duration of the test. Both trains in the remaining pair were operated by a single team, with one operator running both meter boxes. Since all equipment in each train in this pair was virtually identical, had been carefully calibrated, and was operated by the same individual, the sample pair collected during any given run could be considered replicates.

Estimates of the variability within a laboratory were based upon the differences in concentration reported by the paired-train laboratory for the replicate samples on each determination or run. Differences among laboratories were estimated by contrasts between paired trains that were operated by the six single-train laboratories.

Data analysis

Prior to evaluation of the precision of a method, the determinations are tested for equality of variance by means of Bartlett's test for homogeneity of variances. In addition, the determinations are passed through either of two common-variance stabilizing transformations; the logarithmic or the square root, and Bartlett's test is again applied. The use of transformations serves two purposes. First of all, it can put the data into an acceptable form for an analysis of variance; and secondly, it can provide information concerning the true nature of the distribution of sample points. The transformation that provides the highest degree of run equality of variance is accepted and used for deriving the precision estimates.

When the distributional nature of the data is such that its original or linear form provides the highest degree of equality of variance, a constant variance that is independent of the mean level is implied. In this case, the variances are estimated by a pooled analysis of variance on the original data.

Probe. *A technician takes samples from stack gases; another records the data from the instrumentation*

In order to provide the maximum useful information, the test must be designed, and the data analyzed in such a fashion that the precision estimates for a determination can be partitioned into their respective variance components. The variance components of interest are:

- the within-laboratory standard deviation, σ, which measures the dispersion in replicate single determinations made by one laboratory team sampling the same concentration level
- the between-laboratory standard deviation, σ_b, which measures the total variability in a determination caused by simultaneous determinations by different laboratories sampling the same true-stack concentration
- the laboratory bias standard deviation, $\sigma_L = \sqrt{\sigma_b^2 - \sigma^2}$, which is that portion of the total variability that can be ascribed to differences in the field operators, analysts, and the instrumentation, and to different manners of performance of procedural details left unspecified in the method.

With respect to the accuracy of a method, an attempt is made to define its absolute accuracy; this is, how well the measurement value agrees with the actual or true value. Estimates of method accuracy must frequently be based on the analysis of standard cylinder gases. One approach is to have each collaborator measure the concentration of the cylinder gas (or other material), after which a mean and a standard deviation are calculated for the group of collaborators. A 95% confidence interval is then calculated around this mean. If the true concentration of the cylinder gas lies within this 95% confidence interval, then the method is said to be unbiased and accurate within the limits of its precision.

A more common means of stating method accuracy consists of averaging the respective biases of all collaborators and expressing this average as a percentage (either positive or negative) of the overall mean, or of the true value, when known. Both approaches to stating method accuracy will be found in the various collaborative test reports.

Since the initiation of the program in August 1972, evaluations and collaborative studies have been done on a number of methods. Table 1 lists those methods for which some collaborative testing has already been completed, and a discussion of the results of these investigations will now follow. Those readers interested in a more detailed discussion and interpretation of these results are referred to the review by Midgett, and to the individual collaborative test reports.

Velocity

For example, collaborative tests of the Type S Pitot Tube Method for stack gas velocity and volumetric flow rate (EPA Method 2) were conducted at a Portland cement plant, a coal-fired power plant, and two different municipal incinerators. The precision components were shown to be proportional to the mean of the determinations and are expressed as percentages of the true mean in Table 2 for both the velocity and the volumetric flow rate determinations.

Based upon the results of these tests, the precision of the volumetric flow-rate determination seems adequate for use with other test methods in determining pollutant emission rates. A previous single-laboratory study indicated that for nonturbulent streams, Method 2 provides an accurate estimate of the true stack gas velocity at velocities of 55–60 ft/s.

TABLE 1

Methods collaboratively tested

Parameter	Method of determination	EPA Method no.
Stack gas velocity and volumetric flow rate	S-type pitot tube	2
Stack gas molecular weight, CO_2, O_2	Orsat	3
Stack gas moisture content	Condensation and volumetric measurement	5
Particulates	Dry filtration and gravimetric determination	5
Sulfur dioxide	Selective absorption and barium thorin titration	6
Nitrogen oxides	Phenol disulfonic acid	7
Sulfuric acid mist/sulfur dioxide	Selective absorption and barium thorin titration	8
Opacity of stack	Visual estimation of percent opacity	9
Carbon monoxide	Nondispersive infrared absorption	10

TABLE 2

Estimates of precision (standard deviation proportional to mean value)

Method no.	Parameter, units	Standard deviations, percent of mean value (δ)[a]		
		σ	σ_b	σ_L
2	Velocity, ft/s	3.9	5.0	3.2
2	Volumetric flow rate, ft^3/h	5.5	5.6	1.1
5	Particulate matter, mg/m^3	10.4	12.1	6.1
6	SO_2, mg/m^3	4.0	5.8	4.2
7[b]	NO_x, mg/m^3	6.6	9.5	6.9
7[c]	NO_x, mg/m^3	14.9	18.5	10.5
8	H_2SO_4 mist (including SO_3), mg/m^3	58.5	66.1	30.8

[a] σ = within-laboratory deviation; σ_b = between-laboratory deviation; σ_L = laboratory bias. [b] Pooled power plant/pilot combustion plant data. [c] Nitric acid plant data.

Gas analysis

Several collaborative tests of the Orsat methodology for the determination of carbon dioxide (CO_2), excess air, and stack gas molecular weight (EPA Method 3) have been conducted to investigate various aspects of the method's performance. The most recent collaborative field test, by use of a revised version of Method 3, was conducted at a municipal incinerator. Estimates of precision for the various parameters are summarized in terms of standard deviation in Table 3.

Based upon the results of all studies completed, it is concluded that:

- The Orsat method is tedious and requires great attention to detail and technique.
- The original EPA Method 3 operator performance criterion (three consecutive analyses that vary no more than 0.2% by volume) is not easily met in the field, and even meeting this criterion does not ensure that highly reproducible and accurate results will be obtained.
- The use of Orsat data to convert particulate catches to such

TABLE 3

Precision estimates (standard deviation independent of mean value)

Method no.	Parameter, units	Standard deviations, parameter units[a]		
		σ	σ_b	σ_L
3	CO_2, %	0.20	0.40	0.35
3	O_2, %	0.32	0.61	0.52
3	Dry mol wt, g/g-mol	0.035	0.048	0.033
5	Moisture fraction	0.009	0.012	0.008
8	SO_2, mg/m^3	123	115	99
9	Opacity, %	2.05	2.42	1.29
10	CO, mg/m^3	14.3	32.3	29.0

[a] σ = within-laboratory deviation; σ_b = between-laboratory deviation; σ_L = laboratory bias.

reference conditions as 12% CO_2 and 50% excess air can introduce significant errors into the corrected particulate loading.

- The Orsat method is, however, quite satisfactory for use in determining stack gas molecular weights.

Particulates

Collaborative tests of EPA Method 5 for determination of particulate matter emissions were done at a coal-fired power plant, a Portland cement plant, and a municipal incinerator, with the early test design that has subsequently been abandoned. Although attempts were made to find the best test sites available at the time, the power plant testing was conducted at a far less-than-desirable sampling location. Also, the cement plant was subject to extreme load variation over the duration of the test. For the purpose of statistical treatment, the determinations were grouped into blocks by means of the most appropriate blocking criteria that could be devised for each test. A coefficient of variation approach was then used to calculate a within-laboratory, between-laboratory, and laboratory bias component for each test. These ranged from 25.3–31.1%, 36.7–58.4%, 19.6–51.0%, respectively, for the three tests.

Because of problems and uncertainties in the original test designs, as well as difficulties with the test sites, a fourth collaborative test of Method 5, using the paired sampling train test design previously discussed, was undertaken. The test, conducted at a municipal incinerator in September 1975, used a revised and more detailed version of Method 5, since the original method write-up was considered deficient.

Data analysis produced the precision estimates shown in Table 2. (See also Table 3 for the moisture fraction results.) These test results show the precision capabilities of Method 5 to be considerably greater than had been previously thought, and this improvement may be due in part to better test design. There was no feasible way that the accuracy of the method could be estimated under field conditions. Those readers seeking a complete discussion of the probable reasons for the poor precision obtained on the earlier tests are referred to the review by Midgett.

SO_2/NO_x

EPA Method 6 for SO_2 was evaluated, and then was collaboratively tested at two different sites—a 140-megawatt coal-fired electric generating plant, and an oil-fired pilot combustion plant. Analysis of the data by means of a coefficient-of-variation approach provided estimates of the precision components listed in Table 2. From these values, it is evident that Method 6 is capable of good precision when used by competent personnel. Analysis of standard sulfate solutions indicated that most of the precision variation is found in the field-sampling phase of Method 6, as opposed to the analytical phase.

A gas cylinder accuracy test first showed Method 6 to be accurate at SO_2 concentrations of up to about 480 mg/m^3, but indicated that it acquires a significant negative bias above the range of about 480–800 mg/m^3. However, more recent work within EPA has indicated that this conclusion was in error, and the method is unbiased up to SO_2 concentrations of at least 5000 mg/m^3.

EPA Method 7 for NO_x was evaluated for interference effects in the laboratory, and then subjected to collaborative testing at the same two sites used for the Method 6 tests described above. A third test was conducted at a nitric acid plant. The data from the first two tests were pooled to provide a larger data base, and then analyzed with a coefficient-of-variation approach. A similar analysis was performed on the nitric acid plant data. The resulting precision estimates are presented in Table 2, first for the pooled power plant/pilot combustion plant data, and then for the nitric acid plant data.

Because of the larger data base resulting from the pooling of the data from the first two tests, and because of the frequently unstable conditions encountered at the nitric acid plant, it is felt that more reliance may be placed on the precision estimates obtained from the former tests. Analysis of a standard test atmosphere established that Method 7 is unbiased and accurate within the limits of its precision.

Sulfuric acid mist

EPA Method 8 for the measurement of sulfuric acid (H_2SO_4) mist (including any free SO_3) and SO_2 was collaboratively tested at a dual-absorption contact process sulfuric acid plant. Simultaneous samples were collected by four collaborative test teams in a manner analogous to that previously described for the earlier Method 5 tests.

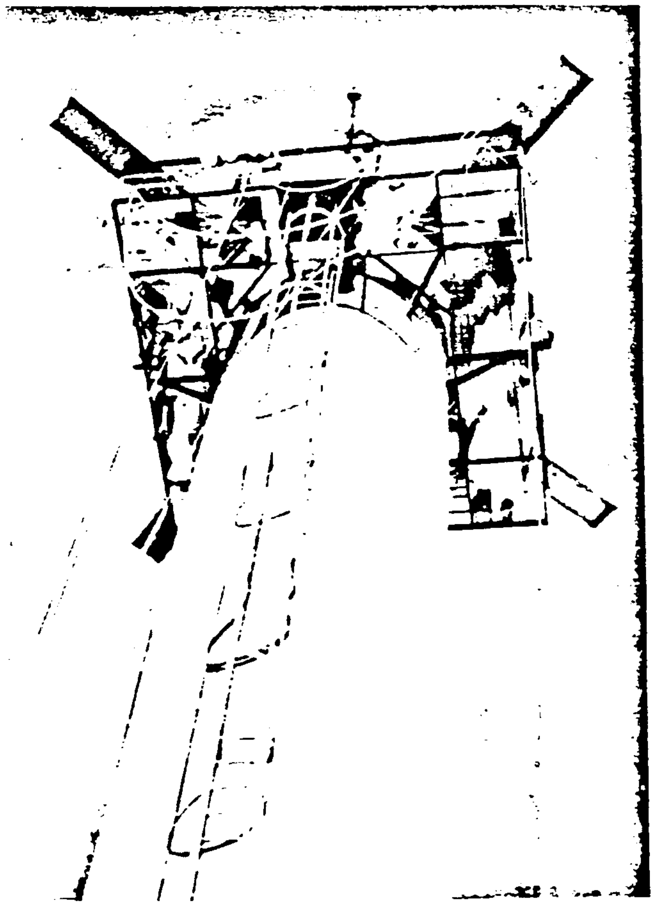
Inspection of the data revealed that H_2SO_4 mist concentrations varied by as much as an order of magnitude between collaborators within single runs, with several very high values occurring. SO_2 determinations varied by as much as a factor of two, and showed a significant negative correlation with the variation in the H_2SO_4 mist values.

The precision components shown in Table 2 for H_2SO_4 mist, and in Table 3 for SO_2, were developed after six extraordinarily high acid mist values were excluded from the data set. Thus, the precision of the acid mist determination was extremely poor in this test and the SO_2 determination, while better than that of the acid mist, was not highly precise either. However, the analytical phase was found to be precise and accurate.

Because of the significant negative correlation between the H_2SO_4 mist determinations and the SO_2 determinations, one is immediately led to suspect some intrinsic problem in the method. But at this time, it is impossible to say whether the imprecision observed is due to a real deficiency in the method, some phenomenon peculiar to the test site, or other unknown factors.

Opacity

Collaborative testing of EPA Method 9 for visual determination of opacity of emissions from stationary sources was conducted by certified observers at three collaborative test sites: a training smoke generator, a sulfuric acid plant, and a fossil fuel-fired



Stack. *Particulate matter test is being validated here* steam generator. The initial test on the training smoke generator was conducted to provide background information on the use of the method, while the test at the sulfuric acid plant and the fossil fuel-fired steam generator was conducted to obtain information on the use of the method on applicable sources under field conditions.

Composite precision estimates based upon the results of all the tests were derived, and these are shown in Table 3. With data from the training generator and from Test 3 at the steam station, a composite estimate of the accuracy of Method 9 was derived for ideal (clear sky) conditions. This estimate compares the expected deviation of the observer from the average metered opacity, and is given by the equation, deviation = 3.13 − 0.31 (meter average), for the range from 5–35%, average opacity.

These Methods are defined

Methods 1–9: *Federal Register*, **36** (247), 24876–24895 (1971)

Methods 10–11: *Federal Register*, **39** (47), 9308–9323 (1974)

Method 9 (revised): *Federal Register*, **39** (219), 39872–39876 (1974)

Methods 1–8 (proposed revisions): *Federal Register*, **41** (111), 23060–23090 (1976)

With respect to the other experimental factors and variables studied, it was concluded from the clear-sky data of Test 3 that:

- The angle of observation does affect the observer's determinations, and, in this study, the most accurate readings were made when the group was at an approximately 45-degree angle to the sun.
- The experienced observers were able to read average opacity more accurately than the inexperienced observers, but the difference occurred mainly in the high opacity range (≥25%).
- Attempts at reading opacity in increments of 1% produced greater within-observer variability and was less accurate than reading in 5% increments.
- Averaging the results of two observers yielded increased accuracy over the result of a single observer.

Based partly on the results of these studies, Method 9 was revised and improved and has now been repromulgated to replace the original method of 1971.

Carbon monoxide

A collaborative test of EPA Method 10 for carbon monoxide (CO) was done at a petroleum refinery, where seven collaborators sampled the emissions from the CO boiler downstream of the fluid catalytic cracking unit catalyst regenerator. An indirect approach based upon the pairing of runs of similar concentration was used to estimate the precision components of the method shown in Table 3. From an analysis of cylinder gases supplied by the National Bureau of Standards (NBS), a somewhat similar between-laboratory term was calculated (26 mg/m^3 as compared to the 32 mg/m^3 shown in Table 3 for the field data). However, the standards data showed about a threefold improvement in the within-laboratory standard deviation over the field data (5.2 mg/m^3 vs. 14 mg/m^3), and this is probably due to the presence of some source variability in the field estimates.

As done in this study, Method 10 produced results with only moderate accuracy of ±101 mg/m^3 (2σ level), on the average, over the concentration range of 277–1048 mg/m^3 of CO. One factor that adversely affected the accuracy of Method 10 was the failure of many of the collaborators to correct adequately for the nonlinearity response characteristics of their instruments.

Another factor was the calibration gases themselves, some of which were in error by as much as 30%, when compared with the NBS standard gases.

Future plans

Results obtained thus far from the methods standardization program indicate that the program has been successful for the most part, although questions concerning the performance of some of the methods tested still remain. It has been shown that the methods for stack gas velocity and volumetric flow rate, particulates, sulfur dioxide, nitrogen oxides, and plume opacity (Methods 2, 5, 6, 7, 9) are indeed reliable if used properly under the conditions for which they were designed. The Orsat procedure (Method 3) is generally satisfactory, provided that its limitations are recognized, and that its limits of precision can be accepted. The method for carbon monoxide (Method 10) is thought to be capable of good accuracy and precision, but it appears that the suppliers of standard gases need to improve the state of their technology. The collaborative test also indicates that some users of nondispersive infrared (NDIR) instrumentation need further training in correcting for the nonlinear response characteristics of their instruments.

Results of the tests of the sulfuric acid mist/sulfur dioxide method (Method 8) indicate that either this method suffers from extremely poor precision, or that the test design was incapable of compensating for the normal range of concentration and velocity variation in time and space at the selected test site. The cause of this poor precision has been investigated, and another collaborative test of the method, using paired sampling trains, has recently been completed. The results of this test are not yet available.

It should be pointed out that EPA Methods 1 through 8 have recently been revised (1976), and that many of these revisions reflect refinements and improvements brought about through the methods standardization program. While the basic chemistry and procedures of these methods remain unchanged, the revisions supply much needed detail, and correct other deficiencies of the original versions. They are due for promulgation in a forthcoming issue of the *Federal Register*.

Several other test methods are now in various stages of the standardization process. These include methods for mercury, vinyl chloride, hydrogen sulfide, fluorides, polychlorinated biphenyls, lead, and arsenic.

The hydrogen sulfide method (Method 11) was evaluated in the laboratory and found to suffer a major interference from thiols, which are common constituents of such gas streams. A modified method, which is designed to eliminate this interference problem, was therefore developed, and this method has now been collaboratively tested. A report on the work is in progress. Field investigations of the fluoride method (Method 13) are complete and a collaborative test has recently been conducted. A report on this work will also be forthcoming.

Additional reading

"National Emission Standards for Hazardous Air Pollutants," U.S. Environmental Protection Agency, *Federal Register*, **38** (66), 8820–8850 (1973).

Midgett, M. R., "The EPA Program for the Standardization of Stationary Source Emission Test Methodology—A Review," U.S. Environmental Protection Agency, Research Triangle Park, N.C. Report No. EPA 600/4-76-044 (1976).

M. Rodney Midgett *is chief of the Source Section, Quality Assurance Branch, Environmental Monitoring and Support Laboratory, EPA. For the past 15 years, Midgett has been employed in various Federal programs concerned with both water and air pollution, and has numerous technical publications to his credit.*

Coordinated by JJ