



Project Summary

Precision and Accuracy Assessments for State and Local Air Monitoring Networks - 1987

Jack C. Suggs

Precision and accuracy data obtained from state and local agencies during 1987 are analyzed. Pooled site variances and average biases which are relevant quantiles to both precision and accuracy determinations are statistically compared within and between states to assess the overall effectiveness and consistency in the application of various quality assurance programs. Individual site results are evaluated for consistent performance throughout the year. Reporting organizations, states and regions which demonstrate consistent precision and accuracy data as the result of effectively administered quality assurance programs are identified. This information is intended as a guide for identifying problem areas, for taking corrective action from the standpoint of improving the effectiveness of quality assurance programs, and for providing more knowledgeable decisions concerning attainment status with regards to ambient air quality standards. An approach to dealing with accuracy data for individual sites is presented, and an alternative sampling design for generating precision and accuracy data is discussed.

This Project Summary was developed by EPA's Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

In accordance with revisions to Appendix A, 40 CFR Part 58 promulgated March 19, 1986, site-specific precision and accuracy data were submitted as actual test results for the first full year beginning January 1987. The availability of individual site data and the opportunity to assess the performance of specific instruments was cited as a way to improve the usefulness of the data quality estimates associated with the NAMS/ SLAMS monitoring network. The regulations did not, however, specify how this would be accomplished except that EPA would now be responsible for calculating the pooled precision and accuracy probability limits formerly calculated and reported by the reporting organizations. The objectives of this report are to analyze and interpret individual site data as they pertain to:

1. Identifying extreme measurement errors.
2. Evaluating the effectiveness of SLAMS quality assurance programs.
3. Validating models used to describe precision and accuracy data.
4. Improving decisions concerning attainment of air quality standards as they relate to specific instruments.

The goal is to provide an overall assessment of various quality assurance programs at the reporting organization, state, and regional levels. Routine calculations are provided only to verify assumptions required to satisfy the objectives. Otherwise, routine information is available through various programs that have access to PARS data files.

Manual SO₂ and NO₂ data were not included in this report because so little data is available. Also, since site codes for 1987 were converted to FIPS codes for this report, site-to-site comparisons between the PARS data and the National Performance Audit Program data were not possible since a complete cross reference conversion file was not available.

Data Analysis

The primary aim of this report is to assess the overall effectiveness of quality assurance programs administered by the states and local agencies. Accomplishing this within the scope of current guidelines for calculating precision and accuracy estimates requires verification of certain basic underlying assumptions either implied or explicitly stated in the models presented in Section 5 of the amendments to 40 CFR Part 58. These models use weighted site averages and pooled within-site variances to calculate quarterly probability limits for percent differences of precision data from a given reporting organization. Accuracy calculations treat different instruments (analyzers) for a given quarter and reporting organization as having been drawn from a homogeneous group of instruments with similar statistical properties. The 95th percentiles used for the probability limits assume that the percent differences are normally distributed.

If it is assumed that individual percent differences were taken from the same normal population for a given pollutant (and level in the case of accuracy data), then annual probability limits could be calculated on a nationwide basis as shown in Tables 1 and 2. The pooled within-site standard deviations for precision data over all sites and quarters is equivalent to the root mean-square error obtained by applying an analysis-of-variance type linear model to each pollutant. Comparisons of site means within reporting organizations, within and between states and regions, and across quarters using F-tests require the same basic assumption: that all site variances are homogeneous. The limits given in Tables 1 and 2 would be valid in this case if assumptions of homogeneity of variance were correct and there were no significant analysis of variance F-tests. However, this is an extreme case and one that is highly unlikely to occur. There may, however, be some reporting organizations, states, or even regions whose sites have equal bias as compared to

within-site differences. In dealing with quality control data, samples should be in subgroups that are as homogeneous as possible so that if differences are present, they show up as differences between subgroups rather than as differences between numbers within a group. This latter assumption (samples should be in subgroups that are as homogeneous as possible) is the basic premise of this investigation which examines the effectiveness of various quality assurance programs.

For precision data, biweekly samples for a quarter determine the basic subgroup. Variation within this subgroup is used to detect differences between sites, between quarters for a given site, or between groups of sites. With regards to the structure of higher level groupings, the basic experimental design for characterizing precision data could be considered as a completely nested design with respect to location parameters (i.e., sites within reporting organizations within states, within regions) with repeated sampling (i.e., biweekly measurements) and quarters providing replication. Under the present model of 40 CFR Part 58, accuracy data does not provide a way of comparing individual sites. Rather, site differences within a reporting organization provide the within-subgroup variation to derive quarterly probability limits for each audit level. In the context of experimental design, accuracy data also follow a nested-like hierarchal structure with respect to location parameters, and quarters provide replication, but there is no repeat biweekly sampling. If analysis of variance techniques had been applicable to the PARS data, between-site variances pooled across reporting organizations and quarters would provide the experimental error for making within and between-state and region comparisons at each separate level.

The assessment of precision and accuracy data involves examining the variance of biweekly samples (between-site variance for accuracy data) and bias (weighted averages) as relevant quantities to both precision and accuracy probability limit calculations. Bias is a systematic error that can often be adjusted through the use of calibration procedures. The variance is a measure of precision (or rather imprecision) that is indicative of random uncontrolled errors that are more serious. Due to extreme imbalance in the data and the need to verify basic assumptions of homogeneity of variance, analysis of variance techniques were not

used to compare biases between groups based on location and time of year.

The approach used to analyze the precision and accuracy data in this report began with the basic subgroup of sites (accuracy data) and biweekly samples (precision data) and worked up the hierarchical structure and across quarters. This provided an independent assessment of the validity of the assumptions of homogeneity of variance and equality of means as they apply to calculating probability limits for each group. The assumption of normality was accepted since all statistical tests used in this report rely to some extent on this assumption, and since it is difficult to verify this on the basis of 6-7 biweekly samples per site per quarter for precision data. For accuracy data there was usually only one site per quarter available per reporting organization.

The basic statistical tools used in this analysis are Bartlett's chi-square test for homogeneity of variance and Welch's test for equality of means. Bartlett's test is well suited to precision data with unequal numbers of biweekly measurements from site to site and accuracy data with unequal numbers of sites across reporting organizations. It is, however, sensitive to departures from normality as well as to unequal variances. Welch's F-test is equivalent to an analysis-of-variance test when within-subgroup variances are homogeneous, but it does not require that variances be homogeneous. The effectiveness of individual quality assurance programs should not be judged on the width of the probability limits for various groups, but also by the validity of assumptions used in calculating probability limits as determined by statistical tests. As a rule, both multiple comparison tests must not be rejected at the $\alpha = 0.01$ significance level to show justification for claiming a real difference between biases or between variances. This level was chosen as an extra measure of precaution to prevent erroneously rejecting comparisons as being significant due to the possible lack of normality in the data. Since Welch's F-test does not require that variances be homogeneous, the two tests are independent. A valid probability limit is an indication that the measurement system(s) is in control and producing uniform results with respect to the group of sites to which the probability limit applies. The width of the probability limits can be judged against the PAF goals for 95% probability target limits of $\pm 15\%$ (all precision checks and flow rate audits) and $\pm 20\%$ (accuracy audits).

Table 1. National Probability Limits for 1987 PARS Precision Data

Pollutant	N	Weighted Average	Pooled std. dev.	Lower 95% prob. limit	Upper 95% prob. limit
CO	13347	0.09	3.56	-6.9	7.0
NO ₂	6410	-0.31	5.15	-10.4	9.7
O ₃	17018	-0.71	4.65	-9.8	8.4
Pb	1247	-1.63	10.82	-22.8	19.5
PM-10	1936	1.28	6.85	-12.1	14.7
SO ₂	24908	-0.87	4.31	-9.3	7.5
TSP	14909	0.07	5.26	-10.2	10.3

Table 2. National Probability Limits for 1987 PARS Accuracy Data

Pollutant	Level	N	Weighted average	Pooled std. dev.	Lower 95% prob. limit	Upper 95% prob. limit
CO	1	1004	0.14	6.64	-12.8	13.1
	2	888	0.34	3.40	-6.3	7.0
	3	877	-0.01	3.17	-6.2	6.2
	4	6	0.00	2.00	-3.9	3.9
NO ₂	1	440	0.76	9.53	-17.9	19.4
	2	387	0.09	5.42	-10.5	10.7
	3	380	-0.21	4.67	-9.3	8.9
O ₃	1	1336	-0.81	6.04	-12.6	11.0
	2	1238	-0.73	4.96	-10.4	9.0
	3	1226	-0.75	3.88	-8.3	6.8
	4	82	-2.31	2.22	-6.6	2.0
Pb	1	600	-0.42	5.29	-10.8	9.9
	2	572	-1.03	3.79	-8.4	6.3
Pb flow	2	55	4.58	4.85	-4.9	14.1
PM-10	2	445	0.10	4.74	-9.1	9.4
SO ₂	1	1353	-0.22	5.65	-11.3	10.8
	2	1183	-0.24	5.15	-10.3	9.8
	3	1188	-0.35	4.82	-9.8	9.0
	4	105	1.39	4.18	-6.8	9.6
TSP	2	3963	0.11	3.34	-6.4	6.6

Precision Results

The analysis of precision data was begun at the reporting organization level since according to Section 3 of 40 CFR Part 58, probability limits at this level are derived from pooled estimates of variances and weighted means of percent differences from stations (sites, instruments, etc.) that are expected to be reasonably homogeneous, as a result of common factors. Reporting organizations having homogeneous within-site variances and equal site means were identified on a quarterly basis. From this

group, states with homogeneous variances and equal means across reporting organizations were identified. Table 3 lists those states by pollutant that demonstrated effective quality control practices because they produced uniform results on a quarterly basis.

As shown in Table 1, probability limits for precision data do not appear valid on a nationwide scale under the assumptions of homogeneity of variance and equal means. However, probability limits at this level do apply to the examination of trends of ambient air quality data

where acceptable probability limits for reporting organizations may be as wide as $\pm 100\%$, and homogeneity of variance is not of primary consideration. Use of the PARS data for purposes other than trends assessment may require that data be examined in smaller groups especially when the validity of assumptions concerning the uniformity of data is important.

Although the requirements for precision and accuracy data were not established for the specific purposes of setting standards or determining attainment

Table 3. States with Homogeneous Variance and Equal Means Across Reporting Organizations

Pollutant	State	Qtr.	N	Weighted average	Pooled std. dev.	Lower 95% prob. limit	Upper 95% prob. limit
CO	WV	1	31	-1.39	7.12	-15.3	12.5
	IN	1	28	-0.76	3.63	-7.9	6.3
	NM	3	15	-5.13	3.54	-12.0	1.8
	MO	4	33	4.03	5.14	-6.0	14.1
	NE	1	20	0.85	1.53	-2.1	3.8
	NE	2	20	1.41	1.96	-2.4	5.2
	AZ	2	42	-0.23	2.33	-4.8	4.3
NO ₂	PA	1	43	0.46	5.06	-9.4	10.3
	PA	2	38	-0.44	6.20	-12.6	11.7
	PA	3	40	-2.48	7.16	-16.5	11.5
	WV	4	31	0.00	5.97	-11.6	11.7
	FL	4	13	7.65	6.78	-5.6	20.9
	OH	2	33	-2.40	4.83	-11.8	7.0
	OH	3	38	2.90	5.24	-7.3	13.1
	OK	1	29	0.38	3.63	-6.7	7.5
	MO	4	26	-0.53	4.36	-9.0	8.0
	O ₃	PA	4	31	-0.54	4.49	-9.3
AL		1	10	0.06	4.33	-8.4	8.5
AL		4	24	-2.71	6.47	-15.4	9.9
NC		4	20	-1.82	3.76	-9.1	5.5
TN		1	22	1.03	2.69	-4.2	6.3
TN		2	27	1.30	3.20	-4.9	7.5
TN		3	28	1.26	2.15	-2.9	5.4
TN		4	28	-0.36	3.13	-6.5	5.7
IN		3	71	-0.50	3.38	-7.1	6.1
OH		4	26	-0.49	2.50	-5.4	4.4
IA		1	27	0.37	2.21	-3.9	4.7
IA		4	28	0.31	3.35	-6.2	6.8
AZ		4	60	-1.12	3.10	-7.2	4.9
PM-10	PA	3	45	-1.30	3.63	-8.4	5.8
SO ₂	ME	3	44	-2.45	3.59	-9.4	4.5
	ME	4	44	-1.24	2.94	-7.0	4.5
	VA	3	69	-1.66	4.07	-9.6	6.3
	KY	2	49	-4.42	4.76	-13.7	4.9
	OK	2	39	-0.38	2.76	-5.8	5.0
	IA	2	54	-2.55	4.85	-12.0	6.9
TSP	AL	1	90	-1.23	6.99	-14.9	12.4
	NC	2	57	-0.44	5.48	-11.1	10.3
	AR	1	34	-3.86	5.01	-13.6	5.9
	MO	3	21	-1.51	1.92	-5.2	2.2
	AZ	1	39	1.11	6.08	-10.8	13.0

status, the extent to which an ambient concentration is nonattainment due to measurement error cannot be ignored. When attention is focused on individual site data, more emphasis is placed on the attainment of short-term standards relative to the performance or adequacy of specific methods or types of monitoring instruments. Although precision (and accuracy) data may not be directly applicable to the determination of attainment status to long-term standards due

to averaging times and spatial differences, the overriding requirements are in conjunction with the interpretation of air quality data. In this regard, site-specific bias and imprecision (variance) provide basic information upon which to draw inference or to simply calculate probability limits. Due to space limitations, it is impossible to list each site that demonstrated uniform results across the year. It is worth noting that approximately 56 percent of the 2528 sites reporting data

maintained an effective quality control program, i.e., one that produced uniform results for the entire year.

Accuracy Results

The basic calculations for computing probability limits for accuracy data are the arithmetic mean and between-site standard deviation for each reporting organization at each audit level on a quarterly basis. Minimum requirements for these calculations are given in

CFR Part 58. Pooling of variances across quarters is not required. However, in the interest of assessing the overall effectiveness of various quality assurance programs, it was necessary to expand the basic calculations to derive annual results and to make comparisons within and between states. In these cases, the homogeneity of between-site variances and equal quarterly averages within reporting organizations had to hold for the probability limits to be statistically valid. Table 4 provides the probability limits for regions that have uniform results for 1987 at a given audit level.

As with precision data, satisfying the assumptions of homogeneous variance and equal means is not necessary to the study of trends so Table 2 may be adequate for this purpose. The validity of these assumptions is necessary for assessing the effectiveness of quality assurance programs. However, a program that is effective at one audit level and not at another is not completely effective.

Under the condition of homogeneity of variance, the average of the percent difference across levels can be considered an estimate of the bias of the slope of a linear regression line through the origin as compared to an ideal slope of 1.00. This model would apply to data with two or more levels such as gases and analytical lead results. The basic statistical assumption of homogeneity of variance required for the validity of using this model on accuracy data can be tested using Bartlett's chi-square statistic. Table 5 lists probability limits and slope estimates for regions that demonstrated an effective quality control program with respect to homogeneity of variance across levels when Bartlett's statistic was used.

Analysis of Individual Site Data

Using the models customarily employed for precision and accuracy data, there is currently no provision for calculating probability limits for accuracy data at individual sites. However, if the conditions of homogeneity of variance can be assumed to hold across levels, then the average and variance of the percent difference (across levels) provide estimates for calculating probability limits for an individual site. In this case, the average is an estimate of the bias in the slope of a regression line through the origin as compared to an ideal slope of 1.00. For calculating variance, there are only one, two, or three degrees of freedom at most

depending on the number of levels audited. An alternative model that would be a more objective way of estimating the slope and consequently the bias is to regress the indicated value onto the designated value. Until more research can be conducted in this area, the risk must be accepted that substantial departures from the necessary assumptions will invalidate the estimates derived from this latter approach.

Conclusions

The availability of individual site data is invaluable in providing more detailed information concerning the performance of site-specific methods and in providing more informed attainment decisions pertaining to a specific site. In addition, having individual site data affords an opportunity to use statistical models to assess the overall effectiveness of specific quality assurance programs. Evaluations based solely on probability limits are inadequate for these purposes. It is evident that some statistical measure of internal comparability must be used in order to detect uniformly reliable precision and accuracy data. This is important for identifying states or reporting organizations that may have difficulty in consistently administering an effective quality assurance program.

The results presented in this report are by no means conclusive. The reporting organizations, states and regions listed as demonstrating consistently effective quality assurance programs were judged on the basis of the uniformity of results where there was no justification in claiming a real difference between biases or between the variances of the data. Precautions were taken to reduce risks of committing errors in this assessment. In fact, it was in the verification of basic assumptions required to validly calculate probability limits that results relating to the effectiveness of various quality assurance programs were derived. The outcome of some comparisons may be due to lack of normality in the data. Although normality is not a direct indicator of the effectiveness of a quality assurance program, it is important for statistically testing the homogeneity of variance and equality of biases as a way of assessing the uniformity of the data. It is hoped that this evaluation will provide Regional QA Coordinators with information to assist in their review of operations and quality control practices across the states in their region.

Recommendations

In the interest of providing precision and accuracy data that is more easily analyzed, interpreted, and costwise and timewise more economical, some consideration should be given to improving the efficiency of the PARS precision and accuracy sampling design. Recommendations are presented in this section that should prove beneficial in these regards to the PARS and similar data bases.

With the exception of flow rate audits, audits should be performed on all instruments at two nominal levels on a biweekly basis. In effect, there would be only one set of data rather than two: one for precision checks and one for accuracy audits. This would allow a regression approach to be used in analyzing and interpreting the data. To make use of the percent difference calculation for automated analyzers, the basic requirements would be that the regression line be a straight line through the origin, and that the error about the line be proportional to audit levels (i.e., homogeneous variance in percent difference across levels). In this case, levels could be placed near the extremes providing a minimum variance estimate of the slope using the percent difference calculation. An optional third level could be audited at a midpoint once per quarter as a test for curvature if needed. However, the use of two levels is best if the estimate of a slope is of primary importance. Probability limits would still be used as an indicator of the distribution of the bias between the average of ratios and the ideal slope of 1.00. Precision estimates would be calculated from the variance of biweekly samples at each level.

Collocated data from manual instruments would still be gathered on a biweekly schedule. This is basically a regression situation as it exists since ambient levels are paired to calculate the percent difference. However, the formula currently used for manual methods is approximately equal to the difference between the logs of the designated and indicated values. Therefore, if a lognormal distribution is assumed for the distribution of ambient particulate data (which is a common assumption), the percent difference is distributed as a normal variable around a mean of zero when there is no bias. Even in the case of accuracy data, averaging percent differences as an estimate of the slope of a regression line through zero is equivalent to taking logs of the indicated measurements in order to stabilize the variance

Table 4. *Regions with Homogeneous Variance Across States and Quarters by Level*

Pollutant	Region	Level	N	Weighted average	Pooled std. dev.	Lower 95% prob. limit	Upper 95% prob. limit
CO	3	2	29	-3.33	4.14	-11.4	4.7
	4	2	98	0.12	2.23	-4.2	4.5
	5	1	50	-3.09	7.15	-17.1	10.9
	5	2	32	0.01	3.12	-6.1	6.1
	6	1	45	1.61	5.88	-9.9	13.1
	6	2	56	0.15	4.74	-9.1	9.4
	6	3	26	-0.98	5.07	-10.9	8.9
NO ₂	3	1	31	4.76	7.28	-9.5	19.0
	3	2	50	2.88	4.82	-6.5	12.3
	3	3	61	1.90	4.23	-6.3	10.2
	5	2	43	-1.03	5.16	-11.1	9.0
	5	3	42	-1.14	5.49	-11.9	9.6
	6	1	31	0.05	11.46	-9.5	22.5
	6	2	31	-1.03	4.34	-7.6	7.4
	6	3	18	-0.51	3.66		6.6
O ₃	3	2	95	-0.65	3.20	-6.9	5.6
	3	3	147	-0.55	3.03	-6.5	5.4
	5	1	66	-2.81	4.93	-12.4	6.8
Pb	4	1	31	-2.13	4.31	-10.5	6.3
PM-10	5	2	48	-0.09	3.73	-7.4	7.2
	7	2	20	-3.03	2.74	-8.4	2.3
SO ₂	4	1	154	-0.43	5.17	-10.5	9.7
	5	1	80	-4.36	6.98	-18.0	9.3
	5	2	211	2.01	4.97	-7.7	11.7
	5	3	206	0.52	4.91	-9.1	10.1
	6	1	43	-1.63	7.87	-17.0	13.8
	6	2	32	-1.9	6.28	-14.2	10.3
	6	3	22	-2.18	6.74	-15.4	11.0
	7	2	30	-0.59	6.28	-12.9	11.7

Table 5. *Regions with Homogeneous Variance Across States, Quarters, and Levels*

Pollutant	Region	N	Weighted Average	Slope	Pooled std. dev.	Lower 95% prob. limit	Upper 95% prob. limit
CO	5	82	-1.87	0.98	5.7	-13.1	9.3
	6	127	0.43	1.00	5.2	-9.8	10.6
NO ₂	3	142	2.87	1.02	5.2	-7.4	13.1
	5	85	-1.09	0.98	5.3	-11.5	9.3
	6	80	-0.49	0.99	7.9	-16.1	15.1
O ₃	3	242	-0.59	0.99	3.1	-6.6	5.4
SO ₂	5	497	0.36	1.00	5.1	-9.8	10.5
	6	97	-1.85	0.98	7.1	-15.9	12.2

across levels when using standard regression techniques. Using this approach, the terms "accuracy" and "precision" are quantities that refer to the bias and variance, respectively, of all quality assurance data and not to two separate

sets of data. This is the usual statistical interpretation of these terms. Flow rate audits should no longer be classified as accuracy audits for the manual methods, but simply as a check on the quantity flow rate since accuracy (bias) and

precision (variance) are both relevant quantities to flow rate audits.

The regression approach would provide a single model for assessing all precision and accuracy data on a site-specific basis. Currently, there is no model for

xamining accuracy data on a site-specific basis. As the sampling design now exists, sites are treated as random for calculating probability limits. For precision data, biweekly measurements provide the random variation for calculating within-site variances, but at only one audit level. In quality control work, it is the behavior of the specific instruments (sites) at different concentration levels that is of interest and not the behavior of a random sample of instruments from some larger population of possible instruments. This is a weakness in the SLAMS PARS system's design that should be resolved.

The EPA author, **Jack C. Suggs**, is with the Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC 27711.

The complete report, entitled "Precision and Accuracy Assessments for State and Local Air Monitoring Networks 1987," (Order No. PB 89-155 246/AS; Cost: \$21.95, subject to change) will be available only from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650

The EPA author can be contacted at:

Atmospheric Research and Exposure Assessment Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Official Business
Penalty for Private Use \$300

EPA/600/S3-89/015

RECEIVED
FEDERAL MAIL
RECORDS MANAGEMENT
SECTION
MAY 15 1989

000085833 PS
U S ENVIR PROTECTION AGENCY
REGION 5 LIBRARY
230 S DEARBORN STREET
CHICAGO IL 60604