United States Environmental Protection Agency
Acid Rain Division (6204J) Washington, D.C. 20460
EPA/430-R-99-009a
June 1999

EPA Flow Reference Method Testing and Analysis: Findings Report

EPA FLOW REFERENCE METHOD TESTING AND ANALYSIS

FINDINGS REPORT

Prepared for

U.S. Environmental Protection Agency
Acid Rain Division, 6204J
401 M Street, SW
Washington, DC 20460

Contract No. 68-D7-0061
Task Order No. 02

Prepared by

The Cadmus Group, Inc.
1920 Highway 54, Suite 100
Durham, NC 27713

June 1999

EXECUTIVE SUMMARY

In the summer of 1997, the U.S. Environmental Protection Agency (EPA) and the Cadmus Group, Inc. (Cadmus) conducted a series of week-long field tests at three electric utility sites to evaluate potential improvements to the Agency's reference method for measuring volumetric flow in stacks (Test Method 2, 40 CFR Part 60, Appendix B). Method 2 does not include procedures to account for yaw or pitch angles of flow when the flow in the stack is not axial and allows the use of only two probe types, the Type S and the Prandtl. The three test sites were chosen to provide a range of yaw- and pitch-angle flow conditions. After undergoing extensive wind tunnel evaluation, seven types of velocity sensing probes (Prandtl, manual Type S, automated Type S, French, modified Kiel, DAT, and spherical) were included in field tests. The manual and automated versions of the Type S probe were operated with and without yaw angle determination. The resulting nine probe/procedure modes were evaluated for their field worthiness and potential to improve the accuracy and precision of Method 2.

The experimental design for the field tests was structured to yield data that would allow drawing statistically and scientifically supportable conclusions about inter- and intraprobe performance, test team effects, and the impact of velocity decay near stack walls on volumetric flow rate measurements. Four test teams using multiple copies of each tested probe performed the field tests. The in-stack measurements were compared to estimates of volumetric flow obtained using five engineering methods: four based on boiler stoichiometry and one based on boiler efficiency. Extensive measures were taken before, during, and after the field tests to ensure data quality. Two rounds of data analyses were performed: one in preparation for peer review, another in response to peer review. A battery of analytical and statistical tools was used to assess the accuracy and variability of the probes and procedures.

The analysis resulted in the following key findings.

- Using various measures of accuracy, three probes yielded the most favorable results: the automated and manual Type S yaw-nulled probes, and the DAT probes. Excluding the low-load tests at one site (which presented difficulties for all probes except the automated Type S), across all three test sites, the analysis revealed that all three probe types produced velocity values within ±3% of the mean of all methods and the engineering baseline.

- Of all the tested probes, the automated Type S, in either straight-up or yaw-nulled mode, yielded the most consistent results. Across all test sites and all load conditions, its coefficient of variation when performing a 16-point traverse was always below 1%.

- The spherical, modified Kiel, and French probes yielded the most variable results with coefficients of variation as high as 19.7%, 5.2%, and 4.2%, respectively, and deviations from the first-round grand mean as large as 12.3%, 7.3%, and 3.3%. A number of peer reviewers, however, questioned the plausibility of the spherical probe data obtained in the second field test. If the spherical probe results from the second field test are not included, the probe's accuracy and variability results are comparable to the accuracy and variability of the automated and manual Type S yaw-nulled probes and the DAT probe. Velocity values

produced by the spherical probe were within ±2.3% of the mean of all methods and ±2.4% of the engineering baseline. Its coefficient of variation did not exceed 1.4%.

- A procedure to account for velocity decay near the stack wall was used during the three field tests and in follow-up testing at additional sites. Accounting for "wall effects" resulted in volumetric flow reductions ranging from 0.6% to 2.4% after excluding anomalous results. The average percent difference in velocity due to wall effects was -1.5% for steel stacks and -1.9% for brick and mortar stacks.

The above results suggest that disparities that have sometimes been reported between engineering estimates of flow and in-stack measurements using Method 2 can be eliminated by using (1) probes that can reliably account for the yaw and pitch angles of flow (e.g., manual and automated Type S yaw-nulled, DAT, and spherical probes); and (2) a procedure that can account for velocity decay near the stack wall.

OVERVIEW

PURPOSE OF THE STUDY

This report describes an experimental program sponsored by the U.S. Environmental Protection Agency (EPA) to evaluate potential improvements to the Agency's current reference method for measuring volumetric flow (Method 2, 40 CFR Part 60, Appendix B). Method 2 [Determination of Stack Gas Velocity and Volumetric Flow Rate (Type S Pitot Tube)] specifies measurements to determine volumetric flow, but does not prescribe specific procedures to account for yaw or pitch angles of flow when the flow in the stack is not axial. Method 2 also allows the use of only two probe types, the Type S and the Prandtl.

During the summer of 1997, EPA conducted week-long field tests at three electric utility sites to evaluate potential improvements to Method 2. The candidate improvements grew out of a technical review of Method 2 and draft Method 2F (an unpromulgated, mid-1993 proposal for using three-dimensional probes in situations where flow velocity has significant yaw and pitch components), extensive preliminary wind tunnel and field testing, and comments provided by industry on the performance of currently available technology.

Three test sites were chosen to provide a range of flow conditions under which the tested equipment and procedures could be rigorously evaluated. One site had near-axial flow, another had flow with moderate yaw and pitch angles, and the third had flow with a significant yaw component. Two sites were gas-fired and the third was coal-fired. This provided an opportunity to compare in-stack measurements of gas flow with calculations of volumetric flow using engineering methods for more than one type of fuel. "Wall effects" tests were conducted at the three primary sites and six additional stacks to measure the impact of velocity decay near the stack wall on volumetric flow.

The four primary goals of the field study were to:

- identify changes to Method 2 that could improve the accuracy of the method under various flue gas flow conditions,
- ensure that each contemplated change would result in measurements that have acceptably low variability,
- ensure that each change would be practicable for use in the field and able to withstand the conditions typically encountered in electric utility stacks, and
- ensure that any changes to the methods would not result in harm to the environment because of systematically low measurements of volumetric flow.

Two collateral goals were to (1) seek improvements to Method 2, which might reduce the disparity industry had reported to occur under certain circumstances between the heat rate derived from in-stack measurements of flow and heat rate calculated from combustion parameters; and (2) collect data that would quantify wall effects.

FIELD STUDY PROCEDURE

Tests compared flow results for seven different types of probes, some operated in different modes, for a total of nine probe/procedure combinations. The nine combinations were the manual Type S probe straight-up (the current Method 2 procedure), manual Type S probe yaw-nulled, automated Type S probe straight-up, automated Type S probe yaw-nulled, DAT, spherical, modified Kiel, French, and Prandtl. Probe types were initially qualified for inclusion in the study through testing performed during 1996 at the Merrill Subsonic Wind Tunnel at North Carolina State University (NCSU) in Raleigh. All of the probes used in the field tests were provided to the test teams by EPA, except for the DAT probes, which the test teams supplied.

Four stack-testing teams performed simultaneous measurements using Method 2 and pre-selected alternatives to Method 2. The multiple-team approach ensured that the data collected on each tested alternative would be representative of and support an assessment of the extent and sources of variability in the measurements. Concurrent in-stack measurements were made using the automated Type S system (the Autoprobe) and using a cross-correlation infrared monitor. The cross-correlation monitor was evaluated as an adjunct to this study. Simultaneously with the in-stack measurements, plant parameter data were collected so that volumetric flow could be estimated using the engineering methods.

A three-part experimental design was used in the field tests:

1. Matrix A—In this interprobe comparison, simultaneous measurements from different types of probes were compared to examine the relative magnitudes of volumetric flow measurements determined by different probe types.

2. Matrix B—In this intraprobe comparison, multiple copies of the same types of probes were compared. In conjunction with Matrix A, Matrix B was designed to provide more extensive data on the variability in each probe's measurements.

3. Matrix C—The wall-effects study prescribed the collection of data in 18 1-in. increments from the stack wall. This study was conducted to evaluate the extent of decay in flue gas velocity near the stack wall.

QUALITY ASSURANCE

Extensive measures were taken to ensure data quality, the most important of which were probe calibrations and pre-analysis screening of data.

Probe Calibration. Routine calibrations of all copies of all probes were conducted at NCSU before the start of field testing (designated as pre-test calibrations) and after all testing was completed (designated as post-test calibrations). Intermediate calibrations were performed for some probes between the first and second field tests because the probes were physically changed. Field test data were derived using the most recent calibration performed before each field test.

The NCSU post-test calibrations were used to determine if the probe calibrations changed over the course of testing and the extent to which any changes affected test results. Independent calibrations of the test probes were performed by the National Institute of Standards and Technology (NIST) at

its Dual-Test Section Wind Tunnel in Gaithersburg, Maryland, using the same written protocol used during the NCSU calibrations.

Data Screening. A three-step, systematic procedure was implemented to identify individual volumetric flow values or specific methods that should not be included in data analyses due to data quality flaws or to problems in the operation of the method.

In the first step of the data screening process, plots of volumetric flow versus run were examined to identify measurements that appeared to differ substantially from other values in the same or proximal runs. Raw data records and supporting information associated with all values thus identified were then reviewed. Data quality problems, equipment malfunctions, and procedural problems were noted. Correctable errors identified by reviewing data records and supporting information were corrected. Values that were flawed due to documented procedural or equipment problems were excluded from analysis.

After this initial screening was completed, the interquartile range (IQR) was used to identify additional suspicious values. The IQR is the difference between the 25th percentile value and the 75th percentile value of all measurements taken for a specific method. Values falling outside 1.5 times the IQR were flagged, and the raw data and supporting information were reviewed.

Finally, the stability of each unit's operation during the course of the field test was evaluated by plotting volumetric flow, O_2 concentration at the economizer outlet, and unit operating load for each run for each site. The resulting effect of unit instability on measured volumetric flow values was examined.

INITIAL DATA ANALYSES

Comparison of Volumetric Flow Measurements Across Probes and Methods. The Matrix A (interprobe) and Matrix B (intraprobe) test results were used to examine differences across probe types and methods. One round of field data analysis was conducted before the results were peer reviewed. A second round of analysis (discussed below) was conducted that incorporated the recommendations provided to EPA by the peer reviewers. Method comparisons were made using several analytical approaches, including descriptive statistics and graphs to describe the flow characteristics at each site and to compare the volumetric flow results obtained by the various methods; rank order analyses to investigate patterns among the methods; analysis of variance (ANOVA) on the Matrix A volumetric flow values to determine whether statistically significant differences can be detected among the individual methods and classes of methods; and central tendency analyses to determine whether certain methods are more likely to be good indicators of the average, long-term volumetric flow.

Each in-stack method was examined individually to determine whether the choice of particular probe copies and test teams affected the results obtained. The variability of each method was also examined. The expected uncertainty in engineering methods was compared with the expected uncertainty in the probe methods.

Analysis of Wall Effects. Wall effects tests were conducted at each of the three steel-stack field test sites during the day under steady load conditions (Matrix C runs) using manual probes and

Autoprobes, and at night under nonsteady-state load conditions using only Autoprobes. Additional tests using only Autoprobes were conducted at six sites with brick and mortar stacks to investigate the effect on volumetric flow of velocity decay for rough stack surfaces.

A standard 12-, 16-, or 20-point cross-stack traverse was first performed as prescribed in Method 1 and the average stack gas velocity was calculated. Velocity was then measured at 1-in. increments across the width of each of the four Method 1 equal-area sectors adjacent to the stack wall, usually starting at 1 in. from the stack wall and moving toward the center of the stack. The near-wall measurements were used to calculate a replacement velocity value for the equal-area sector adjacent to the stack wall for each of the four sampling ports. The average cross-stack velocity was then recalculated by replacing the measured velocity values for the four equal-area sectors closest to the stack wall with the calculated replacement velocity values. The percent difference between the original average cross-stack velocity and the average cross-stack velocity making use of replacement velocity values was then calculated. The calculated percent difference indicates the effect on volumetric flow due to velocity drop-off close to the stack wall. The variation of velocity drop-off as a function of stack material was examined.

Supplemental Data Analyses

The second round of data analyses included a refinement of the central tendency analysis, the addition of a comparison of probe measurements to an engineering baseline, and the performance of a sensitivity analysis of the effect of excluding certain data points that appeared implausible to peer reviewers, even though no documented problems could be found to justify their exclusion based on the criteria established for identifying outliers. A comparison of pressure data collected by manual and electronic devices at two wind tunnels and several field test sites was also conducted.

The concerns expressed by the peer reviewers regarding the Round 1 central tendency analysis centered on how the grand mean was derived. Reviewers indicated that calculating the grand mean of the volumetric flow values using the measurements obtained by every probe type used in the study may not be appropriate, for several reasons. Some probe types used in the study produced uniformly high or highly variable results. Some probe types and methods were over-represented in the grand mean calculations (e.g., there were four Type S probes, each of which was operated straight-up and yaw-nulled and five engineering methods). Because probes or methods of the same type tend to produce similar flow measurements, the Round 1 grand mean value reflects the over-represented probe types or methods, and therefore may not be a good estimate of the true flow in the stack. Probes that do not account for yaw angles generally give higher readings than those that do and should not be included in calculations of the grand mean. And finally, because velocity decay near the stack wall is not adequately captured by a standard Method 1 sampling traverse, the volumetric flow values used in the Round 1 grand mean calculations do not account for wall effects and are therefore high.

Based on these concerns and suggestions, a second central tendency analysis was performed, calculating the grand mean using only those methods reviewers considered to be most technically credible and using wall effects-adjusted values, based on data derived during the wall effects runs. Additionally, Run 3 for the spherical probe at Lake Hubbard was excluded because it seemed implausible to several peer reviewers.

The refined grand mean was calculated using flow measurements from the following four probe types and two engineering methods: manual Type S yaw-nulled, DAT, spherical, baseline Autoprobes 16-point yaw-nulled, MMBtu, and BTCE. For DeCordova and Lake Hubbard, the Autoprobes 16-point yaw-nulled, DAT probes, MMBtu, and BTCE each contributed eight flow values to the grand mean, while the Type S yaw-nulled and spherical probes each contributed four flow values. For Homer City, each probe type and method contributed four values to the grand mean, except for the spherical probes, which contributed eight.

KEY FINDINGS

DAT Probe. The DAT probe gave favorable results with respect to approaching the central tendency of the data (within ±1.0% of the grand mean at DeCordova, Lake Hubbard—high load, and Homer City and 2.3% below the grand mean at Lake Hubbard—low load). This finding is supported by the refined central tendency analysis in Section 4 (DAT probes were always within ±1.3% of the grand mean across all sites). The DAT probe's variability was typical of manual probes (the coefficient of variation ranged from 2.08% at Homer City to 3.14% for the low-load runs at Lake Hubbard), but was significantly higher than that typical of automated probes. The ANOVA analyses of volumetric flow indicate that variation in flow is attributable to different copies of the probe and, at one site, test team was a significant factor in flow variation.

Prandtl Probe. Although the coefficient of variation of the Prandtl probe at the two gas sites where it was tested was the lowest of any of the manual probes (0.63% and 1.27%), mean flow readings at both sites were higher than the central tendency. The Prandtl probe had the highest deviation from the central tendency (+4.3%) at the near-axial site (DeCordova); the deviation was less severe (+2.2%) at the moderate yaw/moderate pitch site (Lake Hubbard). This finding is consistent with the refined central tendency analysis in Section 4 (3.53% from the central tendency at DeCordova and 2.08% at Lake Hubbard).

Spherical Probe. The four original spherical probes (damaged at DeCordova and subsequently repaired and re-calibrated before Lake Hubbard) produced moderately high flow values at DeCordova and by far the lowest flow values for both the high- and low-load tests at Lake Hubbard (12.3% and 10.6% below the grand mean). If Run 3 is excluded, as suggested by peer-reviewers, the average flow measurement for spherical probes for Lake Hubbard high-load tests is lower than the grand mean by 3.45%. At Lake Hubbard low-load, the percent deviation from the grand mean was three times greater than that of any other method. On the other hand, at Homer City where both the new and old sets of spherical probes were tested, the volumetric flow values were very close to each other and much closer to the central tendency of the data (2.0% and 2.1% below the grand mean). The rank order of the spherical probes was relatively high at DeCordova, lowest at Lake Hubbard, and in the middle at Homer City. Consistent results were observed from the refined central tendency analysis in Section 4, except that the deviation from the central tendency was smaller across all sites.

Autoprobes Yaw-Nulled. Of all the tested probe types, the baseline Type S Autoprobe system operated in the yaw-nulled mode with 16-point traverses was closest to the central tendency across all three sites and load levels, as found in both the original and refined central tendency analyses (0.3%, -0.1%, -0.4%, and -1.6% for DeCordova, Lake Hubbard high- and low-load, and Homer City, respectively, from the central tendency in the original analysis, and -0.23%, -0.20%, 0.61%, and

-0.53% for DeCordova, Lake Hubbard high- and low-load, and Homer City, respectively, from the central tendency in the refined analysis). Results for 12-, 16-, and 48-point traverses for the baseline Autoprobe system and the 16-point traverses for the single manual Autoprobe were comparable. The 12-, 16-, and 48-point baseline Autoprobes yaw-nulled and the manual Autoprobe yaw-nulled displayed the lowest variability among all the methods tested, with coefficients of variation consistently below 1%.

Autoprobes Straight-Up. Straight-up operation of the Autoprobes consistently produced higher flow values than yaw-nulled operation. As expected, the difference between the straight-up and yaw-nulled modes increased as flow angularity increased from near-axial to moderate yaw/moderate pitch to high-yaw flow conditions. For 16-point traverses, the deviation from the grand mean for straight-up operation increased from 0.3% at DeCordova, to 1.2% and 0.9% for Lake Hubbard high- and low-load, to 2.7% for Homer City. At the near-axial site (DeCordova) the deviation from the grand mean for straight-up operation (for 16-point, 48-point, or manual 16-point) was no more than 0.2% larger than the deviation from the grand mean for yaw-nulled operation. At the other sites, differences between the straight-up and yaw-nulled values were substantially larger (up to 1.4% at Lake Hubbard and 4.4% at Homer City). The low variability observed for yaw-nulled operation was also found for straight-up operation.

Type S Probe Yaw-Nulled. The Type S probe operated in the yaw-nulled mode consistently produced flow results 2.2% to 2.9% higher than the grand mean. The results of the refined central tendency analysis in Section 4 are consistent with these results. The coefficients of variation for the Type S probe yaw-nulled across all three sites ranged from 1.76% to 3.23%, which was typical for the manual probes tested. No significant differences among probe copies were found, but at Lake Hubbard some test team effect on flow measurement was detected. As with the Autoprobes, yaw-nulled operation consistently produced lower flow values than straight-up operation, and the difference between the two modes increased as flow angularity increased from near-axial to moderate yaw/moderate pitch to high-yaw flow conditions.

Type S Probe Straight-Up. The Type S probe operated in the straight-up mode has been the standard method for measuring volumetric flow for over 20 years. The Type S probe straight-up had the highest positive difference (6.9%) from the central tendency at Homer City and in low-load operation at Lake Hubbard (3.4%), second highest for high-load at Lake Hubbard (6.0%), and third highest at DeCordova (3.5%). In the refined central tendency analysis in Section 4, Type S probe straight-up had the highest positive deviation from the central tendency at Lake Hubbard high- and low-load and Homer City (5.69%, 4.15%, and 7.62%, respectively), and the second highest at DeCordova (2.52%). At the same time, its measurements at the near-axial site (DeCordova) were slightly lower than those of the Prandtl, which is the accepted standard pitot used to calibrate other probes in wind tunnels with axial flow. The coefficient of variation of the Type S straight-up measurements was 2.31% or less at all three sites.

French Probe. At the moderate yaw/moderate pitch and high-yaw angle sites (Lake Hubbard and Homer City), the French probe produced volumetric flow values that were lower than the central tendency (-2.0% and -3.3%). At Homer City the French probe produced the lowest measurements of all the tested in-stack methods (3.3% below the grand mean). The French probe's coefficient of variation ranged from 1.62% at the gas-fired site with axial flow (DeCordova) to 4.18% at the coal-

fired site, where it was the most variable of all tested methods. Its rank order relative to the grand mean, which varied from +6 at DeCordova to -6 at Lake Hubbard and -8 at Homer City, suggests a strong dependence of flow measurement capability on yaw angle.

Modified Kiel Probe. The modified Kiel probe produced the highest volumetric flow values of all the tested methods at Lake Hubbard high-load (7.3% above the grand mean) and second highest at DeCordova (3.6% above the grand mean). However, at Homer City, the modified Kiel probe had the second smallest deviation from the grand mean (0.3%) of all tested methods. At Lake Hubbard low-load, it had the smallest deviation (2.0%) from the grand mean of all the tested in-stack manual methods. The modified Kiel probe's coefficient of variation at Homer City (1.24%) was lowest of any of the tested manual probes. At DeCordova, the coefficient of variation (1.39%) was in the middle of the range for manual probes. The probe had the second highest coefficient of variation at Lake Hubbard (5.19% at high-load and 4.04% at low-load).

Comparison to Engineering Baseline. At DeCordova, Lake Hubbard high- and low-load, and Homer City, respectively, the following percent differences from the MMBtu method flow values were found: 0.39%, -1.88%, -3.55%, and -1.44% for the Autoprobes 16-point yaw-nulled; 2.45%, 0.04%, -1.16%, and 1.88% for the Type S probe yaw-nulled; and 0.37%, -2.71%, -5.33%, and -0.86% for the DAT probe.

Comparison of Pressure-Measuring Devices. From a practical standpoint, differences in pressure readings between manual and electronic pressure devices were generally small. Where statistically significant differences were detected, the differences were not consistent for different types of probes and different velocity conditions. When used with Type S probes, however, manual devices tend to read higher pressures than electronic transducers in field tests, but read lower than electronic transducers in the wind tunnel tests. Generally, the differences between the manual and the electronic devices are smaller on average and less variable in the wind tunnel tests than in the field tests.

Wall Effects. Wall-effects adjustments to flow generally resulted in a decrease in calculated velocity. The average percent difference in average velocity across all probe types and field test sites for which complete wall effects data sets were obtained is -1.72% (s.d. = 0.23%, n = 5 sites). No significant difference was found among probe types. A small difference was observed between the percent difference in the average velocities measured in smooth stacks and rough stacks, but these differences were not statistically significant at the 95% confidence level. The percent decrease in average velocity between unadjusted and wall-effects adjusted traverses becomes smaller as the number of points in the Method 1 traverse increases.

CALIBRATIONS

One- and Two-Dimensional Probes. Comparison of the NCSU pre-and post-test results showed that all the post-test C_p values derived at 60 ft/sec were within ±1.5% of the pre-test values (the values used to calculate flow in the field tests). The largest differences between pre- and post-test C_p values occurred at 30 ft/sec. For all probes, the post-test 30 ft/sec C_p values were higher than the pre-test coefficients, ranging from 0.4% higher for one Prandtl probe to 5.1% higher for one Type S probe. Pre- and post-test C_p values obtained at 60 and 90 ft/sec were generally consistent. Even with the comparatively large changes observed at 30 ft/sec, C_p values averaged over all three velocities

changed by less than 1.5% between pre- and post-test calibrations. Although the C_p values measured at NIST were generally higher than the corresponding NCSU post-test values, the NIST post-test C_p values were within ±2.2% of the NCSU values.

DAT Probes. For three of the four DAT probes, differences in velocities calculated between pre- and post-test NCSU calibrations were less than 2% in the -20° to +20° pitch angle range. In the pitch angle range of -10° and +10°, which is comparable to the range of pitch angles measured at the three utility stacks during this field study, the differences in velocity were less than 1%. For reasons that are unclear, the velocity differences ranged from 3.35% to 4.18% within the -10° to +10° pitch angle range for the fourth DAT probe. A comparison of calculated velocities between the NIST and post-test NCSU calibrations shows that, within the -10° to +10° pitch angle range, the percent difference in calculated velocities over the three nominal wind tunnel velocity settings (i.e., 30, 60, and 90 ft/sec) were 1.7% or less for each probe. In the -30° to +30° pitch angle range, the average percent difference of calculated velocity between the NIST and post-test NCSU calibrations was less than 3.8% for each probe.

Spherical Probes. For the four original spherical probes, the results of the four sets of calibrations at NCSU showed that the calibration values varied by up to 5% over all pitch angles. For the second set of probes, the change in calibration was approximately 4% to 5% within the -10° to +10° pitch angle range. A comparison of the calculated velocities for the NIST and NCSU post-test spherical probe calibrations shows that, within the -30° to +30° pitch angle range, the calculated NCSU velocity for one probe was low relative to NIST (-1.2%). For all other probes, the NCSU calculated velocities were 1.9% to 6.4% higher than those obtained by NIST.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (continued)

SECTION 1
BACKGROUND AND GOALS OF STUDY

1.1 BACKGROUND

This report describes an experimental program sponsored by the U.S. Environmental Protection Agency (EPA) to evaluate potential improvements to the Agency's current reference method for measuring volumetric flow (Method 2, 40 CFR Part 60, Appendix B). Method 2 [Determination of Stack Gas Velocity and Volumetric Flow Rate (Type S Pitot Tube)] specifies measurements to determine volumetric flow, but does not prescribe specific methods to account for yaw or pitch angles of flow when the flow in the stack is not axial. Method 2 also allows the use of only two probe types, the Type S and the Prandtl.

During the summer of 1997, EPA conducted week-long field tests at three electric utility sites to evaluate potential improvements to Method 2. The candidate improvements grew out of a technical review of Method 2 and draft Method 2F (an unpromulgated, mid-1993 proposal for using three-dimensional probes in situations where flow velocity has significant yaw and pitch components), extensive preliminary wind tunnel and field testing, and comments provided by industry on the performance of currently available technology.

Three test sites were chosen to provide a range of flow conditions under which the tested equipment and procedures could be rigorously evaluated. One site had near-axial flow, another had flow with moderate yaw and pitch angles, and the third had flow with a significant yaw component. Two sites were gas-fired and the third was coal-fired, which provided an opportunity to compare in-stack measurements of gas flow with calculations of volumetric flow using engineering methods for more than one type of fuel. "Wall effects" tests were conducted at the three primary sites and six additional stacks to measure the impact of velocity decay near the stack wall on volumetric flow.

The characteristics of the three field test sites are summarized in Section 1.4. Stack diagrams showing port and traverse point locations are presented in Appendix A. A full description of each site and the specific tests performed appears in the separate site data reports for each site[1]. Tests were run at six additional stacks to measure the impact, if any, of the "wall effect" on flow measurement accuracy.

1.2 GOALS OF THE FIELD STUDY

The field tests had four primary goals. The first was to identify changes to Method 2 procedures that could improve the accuracy of the method under various flue gas flow conditions. The second was

---

[1] The Cadmus Group, Inc. 1998, " EPA Flow Reference Method Testing and Analysis: Data Report, Texas Utilities, DeCordova Steam Electric Station, Volume I: Test Description and Appendix A (Data Distribution Package)," EPA/430-R-98-015a.

The Cadmus Group, Inc. 1998, " EPA Flow Reference Method Testing and Analysis: Data Report, Texas Utilities, Lake Hubbard Steam Electric Station, Volume I: Test Description and Appendix A (Data Distribution Package)," EPA/430-R-98-017a.

The Cadmus Group, Inc. 1998, " EPA Flow Reference Method Testing and Analysis: Data Report, Pennsylvania Electric Co., G.P.U. Genco Homer City Station: Unit 1, Volume I: Test Description and Appendix A (Data Distribution Package)," EPA/430-R-98-018a.

to ensure that each contemplated change would result in measurements that have acceptably low variability. The third goal was to ensure that each change would be practicable for use in the field and able to withstand the conditions typically encountered in electric utility stacks. The fourth was to ensure that any changes to the methods would not result in harm to the environment because of systematically low measurements of volumetric flow. In addition to these four primary goals, two collateral goals were identified. One was to seek improvements to Method 2 procedures that might reduce the disparity that industry had reported to occur under certain circumstances between the heat rate derived from in-stack measurements of flow and heat rate calculated from combustion parameters. The second was to collect data that would quantify the "wall effects," that is, the decline in flow velocity near the stack wall.

The field study addressed the four primary goals by evaluating the performance of nine probe/procedure combinations considered for inclusion in a revision to Method 2. The nine tested probe/procedure combinations were the manual Type S probe straight-up (the current Method 2 procedure), manual Type S probe yaw-nulled, automated Type S probe straight-up, automated Type S probe yaw-nulled, DAT, spherical, modified Kiel, French, and Prandtl. Probe types similar to those used in the field tests were initially qualified for inclusion in the study through testing performed during 1996 at the Merrill Subsonic Wind Tunnel at North Carolina State University (NCSU) in Raleigh[2]. All of the probes used in the field tests were provided to the test teams by EPA, except for the DAT probes, which were supplied by the test teams. Diagrams of each probe are presented in Appendix B. The probe/procedure combinations are described below.

The manual Type S probe has been used extensively to measure flow in accordance with the testing requirements of 40 CFR Part 75. When used in accordance with the procedures currently prescribed in Method 2, the probe is referred to as the "straight-up" probe; that is, it is positioned so that an imaginary line connecting the impact port and the static port is parallel to the longitudinal axis of the stack. In this configuration, it measures total velocity. During the test program, the manual Type S probe was also operated in the "yaw-null" mode that included determining the yaw component of the flow.

An automated version of the Type S probe, manufactured by United Sciences Testing, Inc.—the USTI Autoprobe Type S probe—was also evaluated in both the straight-up and yaw-nulled modes. Two configurations of the Autoprobes were operated in this field study: a single Autoprobe was moved from port to port like the other tested pitots, and a four-probe configuration (referred to as the "baseline" Autoprobes) was mounted in four separate ports and collected data from these locations for the duration of the field tests.

3-D probes provide a measure of both yaw and pitch angles of gas flow using five ports on the sensing head of the probe. For this field study, two types were tested: the DAT and the spherical probes. The DAT probe is prescribed in draft Method 2F. The spherical probe was recommended for consideration by the Electric Power Research Institute (EPRI). It is not commonly available. EPA had the spherical probes used for this study specially fabricated in stainless steel. Two sets of

[2] The Cadmus Group, Inc. 1997, "Flow Reference Method Testing and Analysis: Wind Tunnel Experimental Results," EPA/430/R-97-013.

four spherical probes, a total of eight probes, were ultimately fabricated after the welds on the first set sustained thermal damage during the first field test.

The modified Kiel probe is a standard Kiel probe with a wake port and two Fechheimer ports added. The modified Kiel probe was suggested for use by EPA and is not available commercially. EPA therefore had copies of the probe specially fabricated for this study. A prototype modified Kiel probe was tested in the NCSU wind tunnel to determine whether this probe was sufficiently promising to warrant further consideration as an acceptable method for measuring flow. Results obtained in the wind tunnel experiments were used in finalizing the probe's design for use in the field testing.

Recommended for consideration by EPRI, the French probe is a cylindrical probe with a solid stainless steel probe head, an impact port with a 15° chamfer on the probe head, and a wake port located at a 180° offset from the impact port. This probe head design does not meet current specifications in Method 2. The probe, designed by Jennifer French of Southern Company Services, is not commercially available. EPA therefore had copies of the probe specially fabricated for this study.

The Prandtl probe is the accepted standard that is used both to calibrate flow rates in wind tunnels and to establish calibration coefficients for other types of pitot tubes. Also called the "standard pitot," the Prandtl probe used in this project was a 0.25-in. outside diameter stainless steel tube with a 90° bend located approximately 3.5 in. from the hemispherical probe tip. Four, 0.035-in. inside diameter static pressure ports are located 1 in. from the probe tip and positioned 90° apart around the tube.

In conjunction with testing the probes, new field calibration and operational procedures were performed and evaluated as potential method improvements. Digital inclinometers were also tested to evaluate these angle measuring devices for performing yaw-nulling, rotational position calibration, and horizontal straightness checks. A laser device was used to measure stack diameter, and the results were compared to values obtained by calculating the diameter from measurements of the circumference of the stack.

This study was designed to test in-stack methods (i.e., the probes currently allowed or being considered for inclusion in new draft methods). However, to address industry's concern over instances of disparity between heat rates based on conventional industry calculational methods and heat rates based on continuous emission monitoring system (CEMS) measurements, five engineering methods, not under consideration for inclusion in new methods, were included in the study. These five engineering methods were used to evaluate the extent of the disparity between the in-stack and engineering approaches to calculating heat rate. To reduce the sources of variability in these engineering methods, great effort was taken to ensure process stability, collect a vast array of combustion and process parameter data, perform frequent fuel sampling, and obtain multiple, independent fuel analyses. Several specific measures were taken to ensure that the engineering methods were implemented as rigorously as possible. First, non-routine plant operating data were required at each site. Examples of these non-routine data include continuous automated natural gas sampling with analysis by an on-line gas chromatograph, described in the site data reports for the two gas-fired sites. At the coal-fired site, separate coal samples were taken corresponding to each stack run, and laboratory analysis was performed on each sample. Second, unit load was held steady

during the entire week of testing at each site to accommodate the different response times between the engineering methods and the in-stack measurements. Finally, a specialist from Lehigh University's Energy Research Center (ERC) assembled the necessary data and performed the engineering volumetric flow calculations. Copies of the combustion and process parameter data were also provided to field test observers representing EPRI, so that they could offer an independent spot check of the engineering data and calculations.

1.3 Approach

A collaborative field testing protocol was used under which multiple stack-testing teams performed simultaneous measurements using Method 2 and pre-selected alternatives to Method 2. The collaborative (or multi-team) approach ensured that the data collected on each tested alternative are representative of and support an assessment of the extent and sources of variability in the measurements. Concurrent with the test team measurements, in-stack measurements were made using the Autoprobe and using a cross-correlation infrared monitor. The cross-correlation monitor was being evaluated as an adjunct to this study. Simultaneously with the in-stack measurements, plant parameter data were collected to enable volumetric flow to be estimated using the five engineering methods.

A three-part experimental design was used in the field tests:

1. Matrix A—In this interprobe comparison, simultaneous measurements from different types of probes were compared to examine the relative magnitudes of volumetric flow measurements determined by different probe types.

2. Matrix B—In this intraprobe comparison, multiple copies of the same types of probes were compared. In conjunction with Matrix A, Matrix B was designed to provide more extensive data on the variability in each probe's measurements.

3. Matrix C—The wall-effects study prescribed the collection of data in 18 1-in. increments from the stack wall. This study was conducted to evaluate the extent of decay in flue gas velocity near the stack wall.

Table 1-1 provides an overview of Matrices A, B, and C, showing the methods tested and the number of runs performed at each site. A run consisted of each of the four test teams concurrently performing a full 16-point stack traverse. Simultaneous traverses were performed by the four baseline Autoprobes that were installed in separate ports. A full description of the experimental design used at each test site is found in each site data report.

1.4 Key Characteristics of Field Test Sites

Characteristics of the three primary field sites are presented in Table 1-2. Test dates and average velocity, yaw angles, and pitch angles, measured during the tests are presented in Table 1-3.

DeCordova has the highest nameplate capacity, followed by Homer City, and then Lake Hubbard. Operating loads during the test were generally maintained near each unit's maximum capacity. Four low-load runs were also conducted at the Lake Hubbard site.

Table 1-1. Cross-Site[a] Test Matrix Summary: Number of Runs Performed by Each Probe or Method

Method	Matrix A				Matrix B			Matrix C		
		LH								
	DC	High Load	Low Load	HC	DC	LH	HC	DC	LH	HC
DAT	8	9	4	8	4	4	4	1	1	1
Type S Straight-up	4	5	4	4	4	4	4	1	—	1
Type S Yaw-nulled	4	5	4	4	4	4	4	1	—	1
Modified Kiel	4	5	4	4	4	—	4	1	—	1
Spherical A[b]	3	5	4	4	4	4	4	1	—	1
Spherical B[b]	—	—	—	4	—	—	4	—	—	1
Prandtl	4	4	—	—	4	—	—	1	—	—
French	4	4	—	4	4	—	4	1	—	1
Manual Autoprobe 16-point Straight-up	4	4	—	4	—	—	—	0	—	—
Manual Autoprobe 16-point Yaw-nulled	4	4	—	4	—	—	—	0	—	—
Baseline Autoprobes 12-point Straight-up	—	—	—	8	—	—	9	—	—	2
Baseline Autoprobes 12-point Yaw-nulled	—	—	—	8	—	—	9	—	—	2
Baseline Autoprobes 16-point Straight-up	8	9	4	8	23	12	24	6	1	6
Baseline Autoprobes 16-point Yaw-nulled	8	9	4	8	14	12	20	4	1	5
Baseline Autoprobes 20-point Straight-up	—	—	—	0	—	—	15	—	—	4
Baseline Autoprobes 20-point Yaw-nulled	—	—	—	0	—	—	11	—	—	3
Baseline Autoprobes 48-point Straight-up	8	8	4	0	23	12	—	5	1	—
Baseline Autoprobes 48-point Yaw-nulled	8	8	4	0	14	12	—	3	1	—
Infrared Monitor A[c]	8	1	4	8	8	12	24	—	—	—
Infrared Monitor B[c]	—	—	—	8	—	—	24	—	—	—
O_2 F-Factor	8	9	4	8	24	12	24	—	—	—
CO_2 F-Factor	8	9	4	8	24	12	24	—	—	—
MMBtu	8	9	4	8	24	12	24	—	—	—
ASME PTC 4.1	8	9	4	8	24	12	24	—	—	—
BTCE	8	9	4	8	24	12	24	—	—	—
Installed CEMS	—	—	—	8	—	—	24	—	—	—

[a] DC = DeCordova, LH = Lake Hubbard, HC = Homer City.

[b] Four additional spherical probes were tested at Homer City; the original set of spherical probes was designated Spherical "A," and the additional set was designated Spherical "B."

[c] Two (redundant) infrared monitors (designated as A and B) were tested at Homer City.

Table 1-2. Key Physical Characteristics of Field Test Sites

Characteristic	DeCordova	Lake Hubbard	Homer City
Location	40 mi. S.W. of Fort Worth, TX	20 mi. E. of Dallas, TX	Homer City, PA
Fuel	Natural gas	Natural gas	Bituminous coal
Boiler Type	Babcock & Wilcox 5,445,000 lb/hr opposed-fired	Babcock & Wilcox 3,825,000 lb/hr opposed-fired	Foster Wheeler 4,620,000 lb/hr opposed-fired
Nameplate Capacity	799 MWe	531 MWe	660 MWe
Load Range During Tests	784 MWe–786 MWe	508 MWe–510 MWe (high-load runs) 251 MWe (low-load runs)	594 MWe–655 MWe
Inside Stack Diameter	23.08 ft	22.5 ft	24.0 ft
Stack Height	209 ft	199 ft	800 ft
Test Port Distance from Downstream Disturbance	28.1 ft (1.2 diam.)	53.5 ft (2.4 diam.)	420 ft (17.5 diam.)
Test Port Distance from Upstream Disturbance	71.8 ft (3.1 diam.)	46.0 ft (2.0 diam.)	199 ft (8.3 diam.)

Table 1-3. Flow Characteristics of Field Test Sites

Characteristic[a]	DeCordova	Lake Hubbard	Homer City
Test Dates	6/3/97–6/8/97	7/8/97–7/12/97	8/19/97–8/24/97
Velocity (ft/sec) (mean ± s.d.)	94.05 ± 6.24	65.36 ± 17.78 (high load) 39.13 ± 10.65 (low load)	77.89 ± 4.35
Average Absolute Yaw Angle (°) (mean ± s.d.)	2.34 ± 2.00	6.84 ± 6.01 (high load) 7.28 ± 6.40 (low load)	13.93 ± 5.14
Average Absolute Pitch Angle (°) (mean ± s.d.)	3.42 ± 2.49	4.01 ± 3.60 (high load) 4.53 ± 4.47 (low load)	2.24 ± 1.87
Average Yaw Angle (°) (mean ± s.d.)	-0.54 ± 3.03	-0.21 ± 9.10 (high load) -0.82 ± 9.67 (low load)	13.93 ± 5.14
Average Pitch Angle (°) (mean ± s.d.)	2.71 ± 3.24	3.19 ± 4.34 (high load) 2.59 ± 5.83 (low load)	1.17 ± 2.67

[a] Equations for calculating these averages are as follows:

$$\text{Average Absolute Yaw Angle} = \frac{\sum_{i=1}^{n}|Y_i|}{n} \pm \text{s.d.}, \quad \text{Average Absolute Pitch Angle} = \frac{\sum_{i=1}^{n}|P_i|}{n} \pm \text{s.d.}$$

$$\text{Average Yaw Angle} = \frac{\sum_{i=1}^{n}Y_i}{n} \pm \text{s.d.}, \quad \text{Average Pitch Angle} = \frac{\sum_{i=1}^{n}P}{n} \pm \text{s.d.}$$

where:

Y_i = individual measured yaw angle,
P_i = individual measured pitch angle,
n = number of measurements, and
s.d. = standard deviation.

Stack diameter was approximately the same at all three sites (22-24 ft). Stack height was approximately 200 ft at the two gas-fired sites and 800 ft at the coal-fired site. Average flue gas velocity (as an average measured by all probe types) determined during the high-load runs was approximately 95 ft/sec at DeCordova, 65 ft/sec at Lake Hubbard, and 78 ft/sec at Homer City. During the low-load runs at Lake Hubbard, flow velocity averaged about 39 ft/sec.

At DeCordova, the average absolute yaw angle was 2.34° with a standard deviation of 2.00°. The average absolute pitch angle was 3.42° with a standard deviation of 2.49°. Flow measurements obtained from each of the ports were generally uniform. Because the test location was within 1.2 diameters of the stack exit, wind gusts over the top of the stack occasionally caused pulses in the stack pressures.

The flow profile at Lake Hubbard was quite different from the nearly axial flow at DeCordova. Gas flow at the Lake Hubbard stack was stratified, with significantly higher gas velocities at the back of the stack where Ports B and C were located than those measured at the front of the stack (see the relatively high standard deviations for the velocity values presented in Table 1-3). Note that the relatively high standard deviations for the yaw angles at Lake Hubbard are due to the large variation in yaw angles; the high standard deviation does not reflect measurement error. The flow may also have exhibited a double swirl pattern, as indicated by the measured yaw angles in each port; the yaw angles were generally positive in Ports A and B and generally negative in Ports C and D. The average absolute yaw angles were 6.84° with a standard deviation of 6.01°over the high-load runs and 7.28° with a standard deviation of 6.40° over the low-load runs. The average absolute pitch angles at Lake Hubbard were 4.01° with a standard deviation of 3.60° over the high-load runs and 4.53° with a standard deviation of 4.47° over the low-load runs.

The test ports on the Homer City stack were installed at a location that satisfies Method 1 criteria for 12-point traverses (8 diameters downstream and 2 diameters upstream from the nearest flow disturbance). The average absolute yaw angle was 13.93° with a standard deviation of 5.14°. The average absolute pitch angle was 2.24° with a standard deviation of 1.87°.

Box-and-whisker plots included in Appendix A illustrate the yaw and pitch angles at each site.

1.5 Document Organization

This section (Section 1) describes the background and goals of EPA's field flow test program. The section is supported by Appendix A, which illustrates the stack cross-sections and flow characteristics at the DeCordova, Lake Hubbard, and Homer City test sites, and Appendix B, which presents diagrams of the probes tested in this study.

Section 2 describes the procedures used to develop a final data set for analysis, to compare how changes in the calibration coefficients derived before and after the field tests affect volumetric flow results, and to evaluate the error in flow estimates derived from the five engineering methods. Appendix C provides the supporting information for the calibration coefficient comparison (Appendices C1 and C2) and the error analysis of the five engineering methods (Appendix C3).

Section 3 presents the results of the first round of the data analyses (performed before this document was peer reviewed). The section has several major subsections, including comparisons of results across methods (3.1); comparisons of results within methods (3.2); an analysis of method variability (3.3); a summary of central tendency and variability analysis (3.4); and a comparison of the extent of uncertainty inherent in in-stack versus engineering methods (3.5). Five appendices provide

supporting information and data for Section 3: Appendix D, Rank Order Tables; Appendix E: Analysis of Variance; Appendix F, Confidence Interval Plots; Appendix G, Dispersion Analysis; and Appendix H, Input Parameters and Results for Uncertainty Analysis.

Section 4 presents the results of the supplemental data analyses, which were conducted based on recommendations made to EPA by a panel of peer reviewers. The results of a refined central tendency analysis are presented in Section 4.1; Section 4.2 presents a comparison of probe measurements to an engineering baseline; and Section 4.3 presents the results of a sensitivity analysis to determine the effect on the conclusions of excluding certain data points that appeared implausible to the peer reviewers. Section 4.4 presents the results of a comparison of manual and electronic pressure measuring devices, using data from four sources: (1) wind tunnel tests conducted at the National Institute of Standards and Technology (NIST); (2) wind tunnel tests conducted by the Fossil Energy Research Corporation (FERCo) at their swirl tunnel; (3) field test data collected by Radian Corporation at an electric utility plant in Pennsylvania; and field test data collected by FERCo at the Coal Creek Power Plant, located in Washburn, North Dakota and the Columbia Power Plant, located in Portage, Wisconsin. Supporting data for the analyses in this section are provided in Appendix I.

Section 5 presents a detailed discussion of wall effects. Supporting information is provided in Appendix J.

Section 6 presents a discussion of the major findings of this study. Findings are presented by probe type. Additional findings are presented on probe calibration, wall effects, and other issues.

Section 7 presents recommended revisions to equipment and procedures used to determine volumetric flow. The recommendations are based on the major findings discussed in Section 6 and field test observations.

SECTION 2
DATA SCREENING PROCEDURES, PROBE CALIBRATIONS, AND ENGINEERING METHOD ERROR ANALYSIS

This section describes preparatory activities performed on the data before the first round of analysis. Section 2.1 presents the systematic data screening procedures used to identify outlier volumetric flow measurements. Section 2.2 discusses the procedures used to verify probe calibrations. Section 2.3 describes the error analysis performed on the engineering methods. As a result of peer review of the procedures described in this section and the analyses presented in Section 3, additional screening activities were conducted as part of supplemental sensitivity analyses of the field test data. These additional screening activities and the resulting analyses are presented in Section 4.

2.1 DATA SCREENING PROCEDURES

A three-step, systematic procedure was implemented to identify individual volumetric flow values or specific methods that should not be included in data analyses due to data quality flaws or to problems in the operation of the method. The first step was visual inspection of plots of volumetric flow data by run; the second was the application of a classical statistical tool to identify possible outliers; and the third was visual inspection of plots of volumetric flow, O_2 concentration at the economizer outlet, and unit operating load (MWe) for each run at the three sites.

In the first step of the data screening process, plots of volumetric flow versus run were examined visually to identify measurements that appeared to differ substantially from other values in the same or proximal runs. Raw data records and supporting information associated with all values thus identified were then reviewed. Any indication of errors or other data quality problems, including data recording errors, transcription errors, errors in calibration data, calculational errors, equipment malfunctions, or procedural problems was noted. Correctable errors identified through review of data records and supporting information were corrected. Values that were flawed due to documented procedural or equipment problems were excluded from analysis.

After this initial screening was completed, a statistical tool—the interquartile range, or IQR—was used to identify additional suspicious values. The IQR is the difference between the 25th percentile value and the 75th percentile value of all measurements taken for a specific method. The IQR is a standard metric for representing the range of values within which the majority of the measurements occur. Values falling outside of 1.5 times the IQR were flagged, and the raw data and supporting information were reviewed. The 1.5-times-the-IQR criterion was also used to examine individual methods that substantially differed from other methods.

Finally, the stability of each unit's operation during the course of the experiment was evaluated by plotting volumetric flow, O_2 concentration at the economizer outlet, and unit operating load for each run for each site. The resulting effect of unit instability on measured volumetric flow values was examined.

2.1.1 Data Excluded Due to Documented Equipment Problems

Review of raw data records and supporting information for the values identified through initial visual examination of data plots disclosed that two volumetric flow values were flawed and therefore

inappropriate for analysis. The two values were (1) the value obtained by the modified Kiel probe (MK-4) for Run 2 at Lake Hubbard and (2) the value obtained by one of the DAT probes for Run 11 at Homer City. The former was flawed because the probe was bent during the run. The bend was eliminated before the only other use of the MK-4 probe at Lake Hubbard in Run 23. The latter was flawed because moisture was discovered in some of the probe lines after the run. This problem was not observed for any other runs involving the DAT probe at Homer City. These two volumetric flow values, and all data used to generate these values, were excluded from any further analysis.

2.1.2 Statistical Evaluation of the Data

Figures 2-1 through 2-3 present box-and-whisker plots of the volumetric flow values (including Matrix A and Matrix B values, but excluding the two flawed values identified in Section 2.1.1) for each probe at the three field test sites. Values identified on the plots with an asterisk (*) fall outside the "whiskers" representing the 1.5-IQR value range for a specific probe.

When applied to the probe and engineering method data, the 1.5-IQR criterion resulted in the flagging of 4 values at DeCordova, 5 values at Lake Hubbard, and 18 values at Homer City. All 4 of the values at DeCordova and 11 of the 18 values at Homer City were ones that *had not* been identified by visually examining data plots; all 5 of the values at Lake Hubbard were ones that *had* been identified by visually examining data plots. The raw data and supporting information were reviewed for any indication of errors, underlying technical or engineering problems, or other data quality problems, as discussed above.

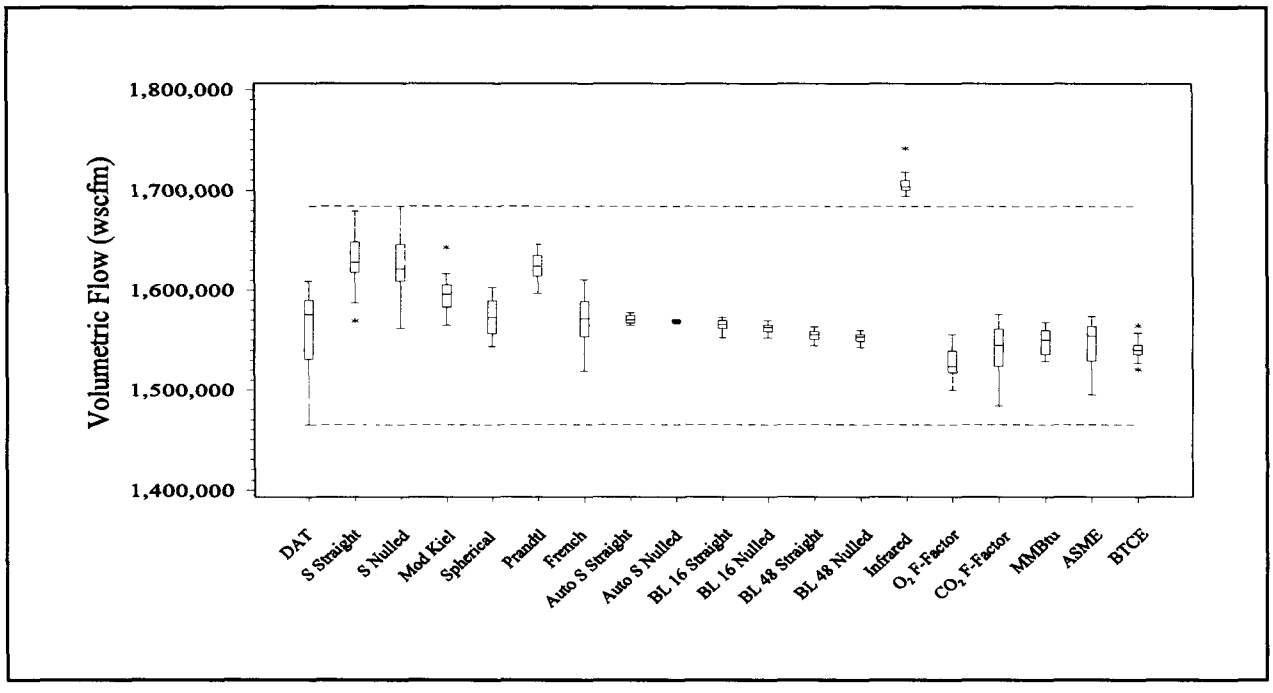

Figure 2-1. DeCordova: Volumetric flow by probe/method (horizontal lines represent the minimum and maximum 1.5-IQR values for all methods except infrared monitor).

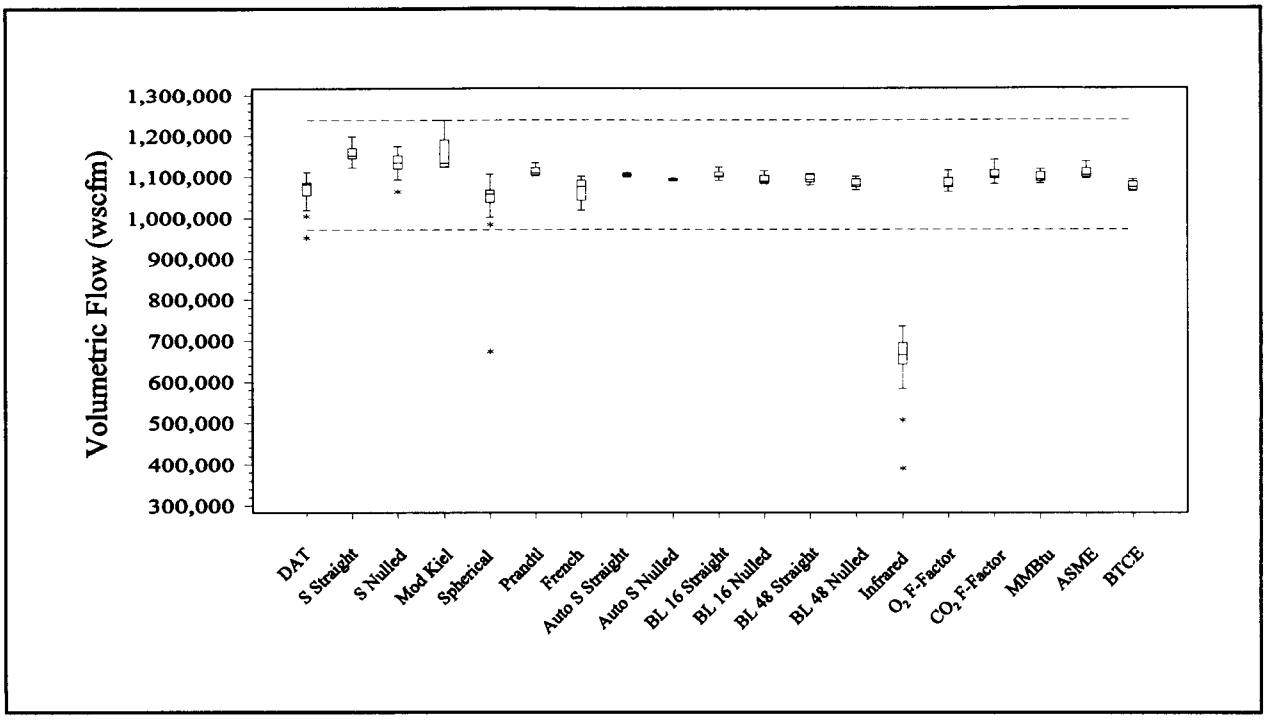

Figure 2-2. Lake Hubbard: Volumetric flow by probe/method (horizontal lines represent minimum and maximum 1.5-IQR values for all methods except infrared monitor).

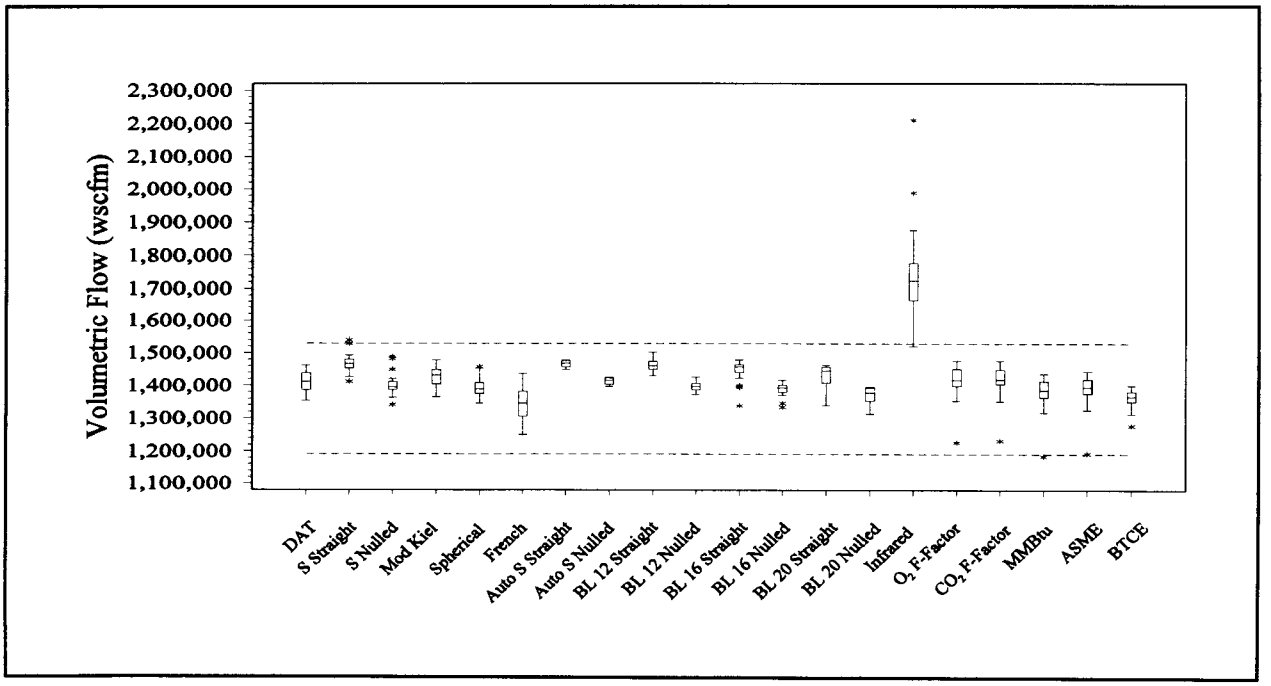

Figure 2-3. Homer City: Volumetric flow by probe/method (horizontal lines represent minimum and maximum 1.5-IQR values for all methods except infrared monitor).

Examination of the raw data and supporting information showed no engineering or data-based reasons for identifying the values as flawed. Therefore, because no plausible reason for excluding the values from further analysis was evident, the values were retained in the analytical data set.

2.1.3 Evaluation of Outlier Probes

Examination of the probe-specific box-and-whisker plots shows that the infrared monitor values are not consistent with the measurements resulting from the other probes and engineering methods (see Figures 2-1 through 2-3). As a test of the difference between the infrared monitor values and the other methods, the minimum and maximum 1.5-IQR values obtained across all other methods, excluding the infrared monitor, were identified. The minimum and maximum 1.5-IQR values are represented by the dashed horizontal lines in Figures 2-1 through 2-3. As seen in these figures, the infrared monitor consistently produced values that were below the minimum or above the maximum 1.5-IQR values.

The infrared monitor values were consistently and substantially higher than values obtained by the other methods at DeCordova (Figure 2-1). At Lake Hubbard, however, the infrared monitor produced volumetric flow values that were always substantially lower than those by the other methods (Figure 2-2). Investigation of the apparent problem indicated by the Lake Hubbard field test data led to the installation and operation of two infrared monitors for the third field test at Homer City. Nonetheless, as shown in Figure 2-3, the average values produced by the two infrared monitors at Homer City were consistently and substantially higher than values obtained by the other methods.

In addition to these data quality problems, the infrared monitors produced less than complete data for two of the field tests. The infrared monitor failed to produce values for 20 runs at the DeCordova field test and for eight runs at the Lake Hubbard field test.

Taken together, the large variability of the infrared monitor values, the large and erratic differences between these values and values produced by other methods, and the incomplete data sets suggest unresolved technical problems with the application of this technology in measuring volumetric flow under the conditions prevailing at the three test sites. To have included the infrared monitor values in the overall analysis of field test data would have produced skewed and unrepresentative results. For these reasons, the data from the infrared monitors have been presented in the site data reports, but are not included in the analyses for this report.

2.1.4 Process Stability Analysis

Maintaining process stability was an important consideration in attempting to limit the introduction of confounding factors in the comparison of flue gas flow rates from different methods. Process stability was particularly important for comparing in-stack and engineering methods because of the likelihood that the two approaches would display different response times in the presence of transient operating conditions. Thus, the field test data were screened to eliminate data collected during unstable conditions. Unit operating load, O_2 concentrations in the flue gas at the economizer outlet, and volumetric flow measurements taken by the baseline Autoprobes (operated in the straight-up mode) were used as indicators of process stability. The process was considered to be stable during a run if load was within ±2.5% of the corresponding values measured during Run 1. (Run times varied between approximately 45 and 75 minutes.)

Figures 2-4 through 2-6 show the relationship between load values and volumetric flow measurements taken by the baseline Autoprobe for each run at DeCordova, Lake Hubbard, and Homer City , respectively. Figures 2-7 through 2-9 show the relationship between O_2 concentration at the economizer outlet and volumetric flow (also measured by the Autoprobe) at the three sites. At DeCordova (Figures 2-4 and 2-7), no substantial change in the operation of the power plant during the time of the experiment is indicated. The maximum absolute deviation from the first run's flow and load values were 0.31% (Run 29) and 0.12% (Run 32), respectively. Plant operating conditions therefore have no substantial impact on the interpretation of the analytical results or on the data chosen for analysis.

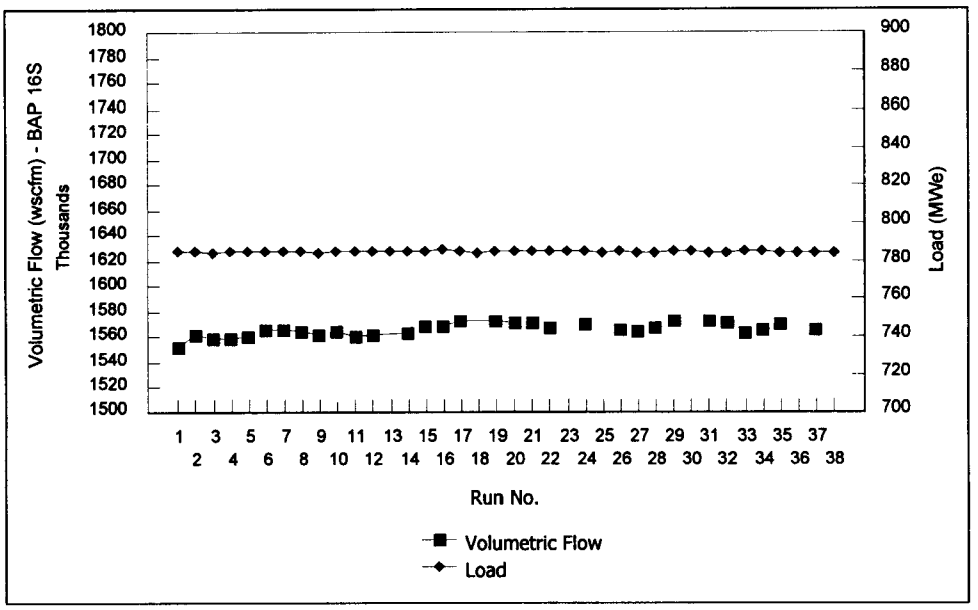

Figure 2-4. Process stability at DeCordova, illustrated by volumetric flow (determined by the baseline Autoprobe, 16-point, straight-up) and unit operating load.

The load and O_2 values at Lake Hubbard during the "high-load" operation (Runs 1-20 and Run 26) also show no substantial change in the operation of the power plant during the time of the experiment (Figures 2-5 and 2-8, respectively). [Run 21, a wall effects run conducted at high load, was not included in the process stability analysis because this run was analyzed separately as part of the wall effects analysis (see Section 5)]. With Runs 21-25 excluded, the maximum deviations from the first run's flow and load values were 2.0% (Run 9) and 0.47% (Run 14), respectively.

As indicated in the experimental design section, Runs 22 through 25 are low-load (~250 MWe) replicates of Runs 1 through 4, which were performed at a high-load condition. Due to this change in load conditions, Runs 22 through 25 are evaluated individually in all subsequent data analyses. Isolating these values from the measurements taken at all other runs minimizes the influence of load on the interpretation of the analytical results. In addition, Run 26 is a replicate of Run 1, taken after the plant regained full-load conditions. However, testing was suspended after Run 26, resulting in an unbalanced set of measurements relative to the experimental design. Therefore, for those analyses that require an equal number of samples per probe (see Section 3 for a discussion of balanced designs), Run 26 was dropped from consideration. For those analyses that are descriptive in nature (e.g., calculation of means and variances), Run 26 was included in the analytical data set.

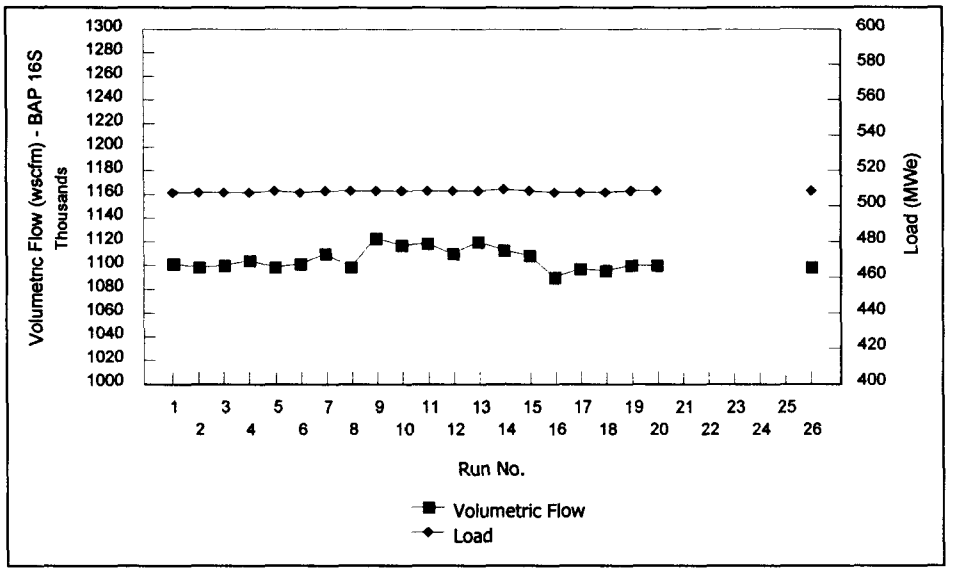

Figure 2-5. Process stability at Lake Hubbard (high load), illustrated by volumetric flow (determined by the baseline Autoprobe, 16-point, straight-up) and unit operating load.

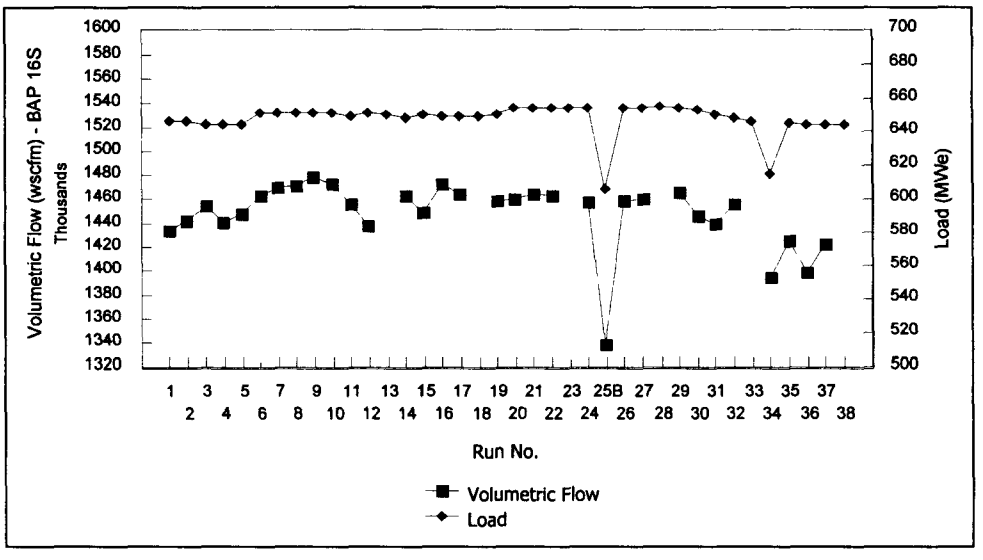

Figure 2-6. Process stability at Homer City, illustrated by volumetric flow (determined by the baseline Autoprobe, 16-point, straight-up) and unit operating load.

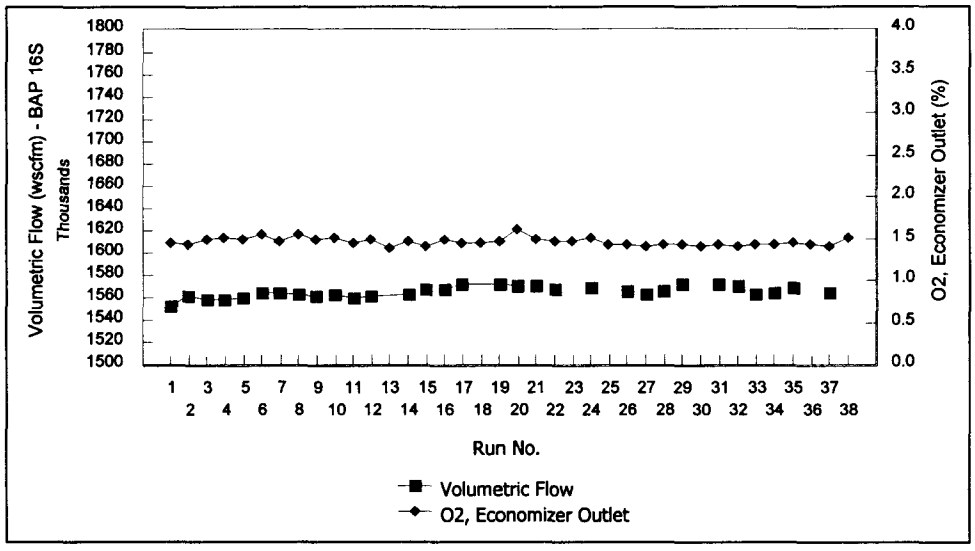

Figure 2-7. Process stability at DeCordova, illustrated by volumetric flow (determined by the baseline Autoprobe, 16-point, straight-up) and O_2 concentration at the economizer outlet.

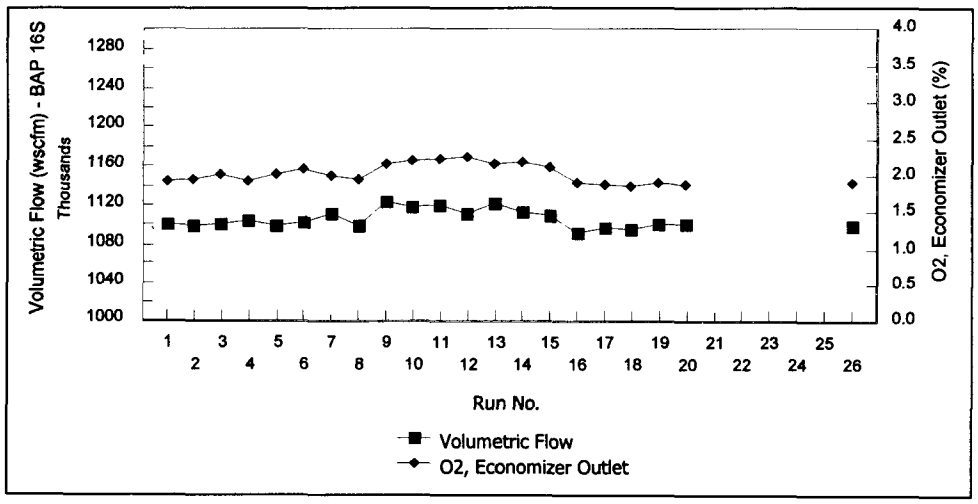

Figure 2-8. Process stability at Lake Hubbard (high load), illustrated by volumetric flow (determined by the baseline Autoprobe, 16-point, straight-up) and O_2 concentration at the economizer outlet.

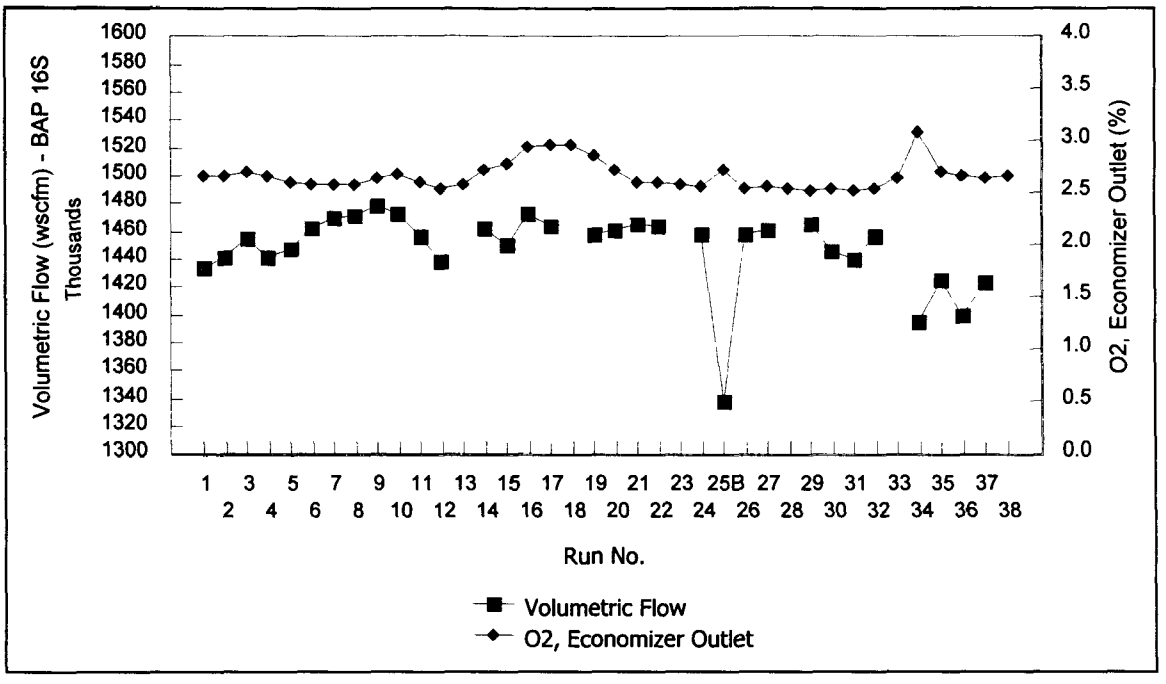

Figure 2-9. Process stability at Homer City, illustrated by volumetric flow (determined by the baseline Autoprobe, 16-point, straight-up) and O_2 concentration at the economizer outlet.

At Homer City, the load and flow values for Run 25B[3] (Matrix B testing of the French probe) were notably lower than the values for the other runs (Figures 2-6 and 2-9). The load and flow values for Run 25B were 6.29% and 6.62% lower than the corresponding values for Run 1. Across all other runs, the maximum deviations of load and flow values from the values in the first run were 1.30% (Run 28) and 3.15% (Run 9), respectively. For this reason, Run 25B was excluded from all subsequent analyses.

2.2 Verification of Probe Calibrations

All copies of all probes used in the field were calibrated before the start of field testing (designated as pre-test calibrations) and after all testing was completed (designated as post-test calibrations). In addition, several types of probes were recalibrated after the start of testing due to physical changes made to these probes either during or after the first field test. The Type S probes were recalibrated after their thermocouples were modified because the initial construction made temperature stabilization unacceptably slow. The original spherical probes were recalibrated after being repaired to eliminate thermal stress fractures sustained during the Matrix A tests (Runs 1-4) in the first field test at DeCordova (see Section 2.2.1). A replacement DAT probe was also initially calibrated between the second and third field test, when the original DAT probe developed an irreparable leak.

[3] Due to a transmission line problem, the plant had to reduce load by approximately 323 MWe during Run 25. The run was invalidated due to the within-run load reduction. The invalid run was designated as Run 25A. The run was then repeated and designated as Run 25B.

All of the field test data reported in this document were derived using the most recent calibration performed prior to each field test, that is, either the initial pre-test calibrations or the recalibrations. The NCSU post-test calibrations were used to determine if the probe calibrations changed over the course of testing, and, if so, the extent to which the changes affected test results. Independent calibrations of the test probes were performed by the National Institute of Standards and Technology (NIST) after the NCSU post-test calibrations were completed. The NIST calibrations were used to independently verify the NCSU calibrations and to ensure that no errors were introduced due to the design of the Merrill Wind Tunnel, the equipment used in the NCSU testing, or the procedures followed by testing personnel.

The NIST calibrations were conducted in the NIST Dual-Test Section Wind Tunnel in Gaithersburg, Maryland, using the same written protocol used during the NCSU calibrations[4]. During the NIST calibrations, the probe pressures were read by three types of differential pressure measuring devices that were connected in parallel: an inclined manometer, an electronic manometer, and a set of magnehelic gauges. All the NCSU measurements were obtained using magnehelic gauges.

The tests at NCSU and NIST did not include the calibration of a "control" probe, that is, one that had not been involved in the field tests. Therefore, it is not possible to distinguish the extent of the change in calibration coefficients attributable to field use as opposed to variability inherent in the performance of the calibrations.

Details of the probe calibration procedures and data analyses, including side-by-side comparisons of the NCSU pre-test, NCSU post-test, and NIST calibration factors, are presented in Appendices C1 and C2.

2.2.1 Calibration Comparison: Pre-Test NCSU to Post-Test NCSU

One- and Two-Dimensional Probes

Calibration coefficients for the Prandtl, French, Type S, Autoprobe Type S, and modified Kiel probes were derived at wind tunnel velocities of 30, 60, and 90 ft/sec. The field test results were computed for these probes using pre-test C_p values at 60 ft/sec. Review of the pre- and post-test probe calibration coefficients derived at 60 ft/sec, which are presented in Table 2-1, show that all the post-test C_p values were within ±1.5% of the pre-test values. Method 2 contains a precision specification limiting the deviation between the three individual C_p values (obtained from the three repetitions at a single velocity) and the average C_p for each probe to 0.01 or less. All probes satisfied this specification.

Summary tables that present the calibration results for each probe at each tested velocity are presented in Appendix C1. The largest differences between pre-test and post-test C_p values occurred at 30 ft/sec. For all probes, the post-test 30 ft/sec C_p values were higher than the pre-test coefficients, ranging from 0.4% for Prandtl probe PR-2 to 5.1% for Type S probe S-11. Pre- and post-test C_p values obtained at 60 and 90 ft/sec were generally consistent. Even with the comparatively large changes observed at 30 ft/sec, C_p values averaged over all three velocities changed by less than 1.5% between pre- and post-test calibrations.

---

[4] The Cadmus Group, Inc. 1997. "Flow Reference Method Testing and Analysis, Test Plan: Velocity Probe Calibrations at the NIST Wind Tunnel," prepared for the U.S. Environmental Protection Agency under Contract No. 68-W6-0050, Work Assignment No. 0007AA-23, 16 pgs. plus appendices.

Table 2-1. One- and Two-Dimensional Probe Calibration Coefficients Derived at 60 ft/sec Nominal Wind Tunnel Velocity

Probe Type	Probe ID	NCSU Pre-Test Calibrations (March-May 1997) C_p	NCSU Post-Test Calibrations (October-November 1997) C_p	% Difference[a]
Prandtl	PR-1	0.976	0.982	0.6
	PR-2	0.970	0.970	0.0
	PR-3	0.976	0.972	-0.4
	PR-4	0.976	0.970	-0.6
French	FR-1	0.711	0.705	-0.8
	FR-2	0.704	0.699	-0.7
	FR-3	0.710	0.711	0.1
	FR-4	0.708	0.699	-1.3
Type S	S-10	0.802	0.790	-1.5
	S-11	0.803	0.797	-0.7
	S-12	0.801	0.796	-0.6
	S-13	0.801	0.797	-0.5
Autoprobe Type S	026	0.785	0.784	-0.1
	027	0.788	0.785	-0.4
	028	0.783	0.779	-0.5
	033	0.789	0.780	-1.1
	036	0.770	0.775	0.6
Modified Kiel	MK-1	0.736	0.734	-0.3
	MK-2	0.736	0.747	1.5
	MK-3	0.736	0.746	1.4
	MK-4	0.736	0.744	1.1

[a] % Difference = $[(C_p^{NCSU\ post} - C_p^{NCSU\ pre}) / C_p^{NCSU\ pre}] \times 100$

The magnitudes of the changes observed at 30 ft/sec are within the error expected for magnehelic gauges such as those used in the NCSU probe calibrations. According to the NCSU principal investigator, a 2% to 3% error could be expected in gauge readings, given the lower pressures generated at the low tunnel speeds (as low as 0.2 in. H_2O measured by the Prandtl probe).

Three-Dimensional Probes

Comparisons of calibration data for the 3-D probes are made on a velocity basis. To facilitate these comparisons, a pre-test velocity data set was generated for each probe using the pre-test pressure data and the resulting set of calibration values (F_1 and F_2 at each pitch angle). A velocity was computed for each probe pitch angle at each wind tunnel velocity setting. Post-test velocities were then computed using the post-test F_1 and F_2 calibration records that were applied to the same pre-test point-by-point pressure data. DAT and spherical probe calibration data comparisons are presented in Appendix C1.

DAT Probes. For three of the four DAT probes, differences in velocities calculated using pre- and post-test NCSU calibrations were less than 2% in the -20° to +20° pitch angle range. In the pitch angle range of -10° and +10°, which is comparable to the range of pitch angles measured at the three

utility stacks during this field study, the differences in velocity were less than 1%. For reasons that are unclear, the velocity differences ranged from 3.35% to 4.18% within the ±10° pitch angle range for the Cadmus DAT probe (3D-3).

Spherical Probes. The first set of spherical probes (copies SPH-1 through SPH-4) was calibrated on four occasions in 1997: in April, before the first field test; in June, after repairs were made to the probes and before performing the second field test; in July after the second field test; and in October, following the completion of the third test. After the first set of probes (referred to collectively as Spherical A) was repaired, four new spherical probes (designated as copies SPH-5 through SPH-8 and referred to collectively as Spherical B) were purchased for use during the third field test. These Spherical B probes were calibrated in August before the third field test and again in October, following the third field test.

Because the April calibrations on the Spherical A probes were negated by the repairs, the June results were used as the pre-test calibrations for the Spherical A probes; the July calibrations for the Spherical A probes are referred to as "intermediate" calibrations. The August results for the four Spherical B probes were used as the pre-test calibrations for those probes. For both sets of probes, the October results were used as the post-test calibrations.

A review of all four sets of NCSU calibrations of the Spherical A probes showed two distinct groups. First, the final (October) calibration records compared well with the initial calibration, even though it was believed that the probe repairs would have changed the initial probe calibration; second, the June and July calibrations compared well. The differences between these two pairs of calibration records are approximately 2% to 3% within the -10° to +10° pitch angle range; the differences increase at pitch angles beyond ±10°. The results of the four calibrations show that the calibration values may vary by up to 5% over all pitch angles.

The Spherical B probe copies were calibrated only twice, once before and once after the Homer City field test. The differences in these pre- and post-test calibrations were greater than those determined for the Spherical A probe copies. On average, the change in calibration was approximately 4% to 5% within the -10° to +10° pitch angle range.

2.2.2 Calibration Comparison: NIST to NCSU

In the probe calibrations performed at NIST, excellent agreement was achieved between the resulting calibration factors generated using the electronic manometer, inclined manometer, and magnehelic gauges; the differences, in most cases, were less than 1%. Because of this agreement, the NIST calibrations used in all comparisons presented below are based exclusively on the inclined manometer readings. The complete presentation of calibration factors generated using the three pressure devices is included in Appendix C2.

One and Two-Dimensional Probes

The calibration procedures performed at NIST were identical to those performed at NCSU. Calibration coefficients for the Prandtl, French, Type S, Autoprobe Type S, and modified Kiel probes were derived at nominal wind tunnel velocities of 30, 60, and 90 ft/sec, and the resulting C_p values were calculated. Review of the NIST and NCSU post-test probe calibration coefficients derived at 60 ft/sec, which are presented in Table 2-2, show that the NCSU post-test C_p values were

within ±2.2% of the NIST values. NIST C_p values generally were higher than the corresponding NCSU post-test values; percent differences between the NIST and NCSU values varied slightly from probe to probe.

Summary tables of the calibration results for each probe at each tested velocity are presented in Appendix C2.

Table 2-2. One- and Two-Dimensional Probe Calibration Coefficients Derived at 60 ft/sec Nominal Wind Tunnel Velocity

Probe Type	Probe ID	NCSU Post-Test Calibrations (October-November 1997) C_p	NIST Calibrations C_p	% Difference[a]
Prandtl	PR-1	0.982	0.9970	-1.5
	PR-2	0.970	0.9885	-1.9
	PR-3	0.972	0.9890	-1.7
	PR-4	0.970	0.9887	-1.9
French	FR-1	0.705	0.7093	-0.6
	FR-2	0.699	0.7093	-1.5
	FR-3	0.711	0.7100	0.1
	FR-4	0.699	0.7094	-1.5
Type S	S-10	0.790	0.8022	-1.5
	S-11	0.797	0.8051	-1.0
	S-12	0.796	0.7909	0.6
	S-13	0.797	0.8088	-1.5
Autoprobe Type S	026	0.784	0.7941	-1.3
	027	0.785	0.7931	-1.0
	028	0.779	0.7963	-2.2
	033	0.780	0.7921	-1.5
	036	0.775	0.7852	-1.3
Modified Kiel	MK-1	0.734	0.7483	-1.9
	MK-2	0.747	0.7485	-0.2
	MK-3	0.746	0.7508	-0.6
	MK-4	0.744	0.7471	-0.4

[a] % Difference = $[(C_p^{NCSU\ post} - C_p^{NIST}) / C_p^{NIST}] \times 100$

Three-Dimensional Probes

DAT Probes. Comparisons of the NIST and NCSU post-test calibration results for 3-D probes were made on a velocity basis, as described in the earlier subsection on comparisons of NCSU pre- and post-test calibration results for 3-D probes. For the NIST vs. NCSU post-test comparisons, a baseline velocity data set was generated for each probe using the NCSU pre-test pressure data and calibration values (F_1 and F_2 at each pitch angle). The NIST and NCSU post-test calibration values for each probe were applied to the NCSU pre-test pressure data for that probe and the resulting velocities were compared.

A comparison of calculated velocities within the -10° to +10° pitch angle range shows good agreement between the NIST and post-test NCSU calibrations. Within this pitch range, the average

percent difference between the NCSU and NIST calculated velocities over the three nominal wind tunnel velocity settings (i.e., 30, 60, and 90 ft/sec) was 1.7% or less for each of the tested DAT probes (Table 2-3).

Table 2-3. Comparisons of Calculated Velocities Using the NIST and NCSU Post-test Calibrations of DAT Probes

DAT Probe ID	Average % Difference[a]: NIST vs. NCSU Post-Test Calibrations		
	0°, ±5°, ±10° Pitch	±15° and ±20° Pitch	±25° and ±30° Pitch
E-DAT	-0.3	-2.0	-3.0
K-DAT	-0.6	-1.0	-1.6
T-DAT	-1.7	-3.0	-2.4
3D-3	-1.5	-3.3	-3.8

[a] % Difference = [($NCSU^{post}$ - NIST) / NIST] × 100.

Beyond the ±10° pitch range, the differences between the velocities derived using the NCSU and NIST calibration curves are greater. Table 2-3 shows the average differences between NIST and the NCSU calibrations for tests conducted at the following groups of pitch settings: (1) 0°, ±5°, and ±10°; (2) ±15° and ±20°; and (3) ±25° and ±30°. For all DAT probes, the average percent difference between the NIST and NCSU post-test velocities was greater for pitch angles outside the range of ±10° than for pitch angles within the range of ±10°. Because the pitch angles found at all three sites were generally in the range of ±10°, only the results in the ±10° pitch range are relevant to this study. Considering the research-grade quality of both the NCSU and NIST wind tunnels, the high level of quality control exercised during both calibrations, and the professional experience of the test personnel, the results shown in Table 2-3 suggest that aerodynamic factors may make obtaining stable calibration curves at pitch angle settings beyond ±10° increasingly difficult.

Spherical Probes. A comparison of the calculated velocities for the NIST and NCSU post-test spherical probe calibrations (Table 2-4) shows that, within the -10° and +10° pitch angle range, the calculated NCSU velocity for one probe (SPH-1) was low relative to NIST (-1.2%). For all other probes, the NCSU calculated velocities were 1.9% to 6.4% higher than those obtained by NIST.

Unlike the DAT probes, the differences between the velocities derived using the NCSU and NIST calibration curves did not change significantly at settings outside the ±10° pitch range (Table 2-4). Table 2-4 also shows the average differences between NIST and the NCSU calibrations for each probe for tests conducted at the following groups of pitch angles: (1) 0°, ±5°, and ±10°; (2) ±15° and ±20°; and (3) ±25° and ±30°.

Summary tables of the calibration results for each DAT and spherical probe at each tested velocity are presented in Appendix C2.

Table 2-4. Comparisons of Calculated Velocities Using the NIST and NCSU Post-test Calibrations of Spherical Probes

Spherical Probe ID	Average % Difference[a]: NIST vs. NCSU Post-Test Calibrations		
	0°, ±5°, ±10° Pitch	±15° and ±20° Pitch	±25° and ±30° Pitch
SPH-1	-1.2	-1.7	-0.2
SPH-2	4.4	3.6	3.7
SPH-3	3.4	2.5	3.0
SPH-4	1.9	2.0	2.3
SPH-5	6.1	5.3	4.3
SPH-6	4.6	3.3	2.8
SPH-7	5.7	4.1	2.9
SPH-8	6.4	5.8	5.1

[a] % Difference = [($NCSU^{post}$ - NIST) / NIST] × 100.

2.3 ENGINEERING METHOD ERROR ANALYSIS

For this study, two types of engineering methods were used to estimate volumetric flow: (1) four methods based on combustion stoichiometry (O_2 F-factor, CO_2 F-factor, MMBtu, and ASME PTC 4.1) and (2) one method based on conservation of mass and energy (BTCE). Two of the four stoichiometric methods use F-factors. F-factor methods of estimating flue gas velocity are based on estimating the ratio of flue gas generated from combustion of fuel to the gross calorific value of the fuel burned. The ratio, which is specific to the type of fuel burned, is expressed in terms of standard cubic feet (scf) per million Btu (MMBtu). The dry F-factor (F_d), used in the O_2 F-factor method, includes all components of combustion less water, while the carbon F-factor (F_c), used in the CO_2 F-factor method, includes only CO_2. In the O_2 F-factor method, F_d is used in combination with measured values of flue gas O_2 concentration and total heat input to the boiler to estimate the volumetric flow rate of the flue gas. In the CO_2 F-factor method, F_c is used with measurements of CO_2 concentration in the flue gas and total heat input to the boiler to estimate volumetric flow.

The MMBtu method is the third stoichiometric method. This method is based on the principle that the weight of air required to burn a unit weight of fuel is proportional to the amount of heat input. Several steps are involved in the calculation procedure, including deriving the weights of wet and dry combustion air and wet gaseous products of combustion, which are expressed in units of MMBtu fired. The weight of gaseous combustion products per MMBtu is multiplied by the total heat input to the boiler in MMBtu to determine the flue gas mass flow rate, which is then divided by the density of the flue gas to derive the flue gas volumetric flow rate.

In the fourth stoichiometric method, which is based on procedures for determining plant heat rate described in the American Society for Mechanical Engineers Power Test Code (ASME PTC) 4.1, the flue gas mass flow rate is calculated using the as-fired (AF) fuel mass flow rate and the calculated weight of wet flue gas per pound of AF fuel. The weight of wet flue gas per pound of AF fuel is calculated from the weight of dry flue gas per pound of AF fuel, which is based on measured concentrations of CO_2, O_2, and CO in the flue gas and stoichiometric combustion of fuel carbon and sulfur, along with the weight of flue gas moisture per pound of AF fuel, which takes into account moisture in the AF fuel and combustion air. The weight of ash per pound of AF fuel and ash heating

value are also taken into account in determining the weight of dry flue gas per pound of AF fuel. The flue gas volumetric flow rate is determined by multiplying the flue gas mass flow rate by the flue gas density, which is based on measured concentrations of H_2O, CO_2, O_2, and CO in the flue gas.

The Boiler Turbine Cycle Efficiency (BTCE) method can be used to estimate the flue gas volumetric flow rate generated from fossil fuel combustion based on conservation of mass and energy. Unlike the four methods described above, which require quantitative data on fuel feed rate, the BTCE method does not require measurement of fuel feed rate. It does, however, require approximately 20 measured plant parameters, such as economizer excess O_2 concentration, fuel composition, and steam, air, and fuel temperatures, as well as calculated performance data, such as turbine cycle heat rate and air preheater leakage rate. These data are then used as inputs to solve a series of complex equations based on conservation of mass and energy for the combustion side of the power plant.

A formal quantitative error analysis, called first-order error analysis, was implemented for each engineering method evaluated at the three test sites. First-order error analysis answers the question: "What is the total uncertainty in any single estimate of volumetric flow based on a specific engineering calculation?" The analysis was performed by the Lehigh University ERC before the field studies began, using data from historical studies and literature values, and again after the field tests were completed, using data generated during the course of the field testing. The results of the analyses are presented in this section.

It is important to note that error analysis (also referred to as uncertainty analysis) provides a measure of the variability (i.e., uncertainty) inherent in a method, not its accuracy. Error analysis reveals the variation that can be expected in a method's results based on the variation in its input parameters, irrespective of whether the method is a good or poor indicator of the true flow value. Accepting the assumptions of the method, error analysis quantifies the variability that can be expected in the method's results.

The first-order error analysis consists of several steps. First, an equation providing an approximation (to first order) of the error in a single calculation of volumetric flow is derived. The first-order equation propagates the error in each term of the engineering method equation into an error in the calculation of volumetric flow. Assuming independence among the input parameters for each of the engineering method calculations, the overall (resulting) uncertainty in the calculated flow rate can be determined from the uncertainties in the primary measurement parameters (such as temperature and pressure) using the following approach:

$$U(R) = \{[\theta_1 U(x_1)]^2 + [\theta_2 U(x_2)]^2 + ... + [\theta_n U(x_n)]^2\}^{0.5} \qquad \textbf{Eq. 2-1}$$

where:

- $U(R)$ = resulting coefficient of variation (CV) in calculated volumetric or mass flow rate of flue gas;
- θ_i = sensitivity coefficient for the i-th parameter, i = 1 to n;
- $U(x_i)$ = measurement uncertainty in the i-th parameter as represented by the sampling variance of the parameter, i = 1 to n;
- x_i = the i-th parameter, i = 1 to n; and
- n = the total number of parameters.

Sensitivity coefficients for the stoichiometry-based engineering methods were determined by differentiation, while sensitivity testing of the engineering equation was used to determine the sensitivity coefficients for the BTCE method. This analysis was performed using the HEATRT computer code developed by the ERC at Lehigh University. Tables C3-1 and C3-2 in Appendix C3 present the input data used in the first-order error equation for each engineering method at each field test site.

After solving the first-order error equation, the uncertainty in the calculation of a single volumetric flow value for each of the engineering equations was generated. The results are presented in Table 2-5 for both the pre- and post-test uncertainty analysis. The pre-test uncertainties were based on data from historical studies and literature values. The post-test uncertainties shown in this table were derived from the data generated during the field tests. The post-test uncertainties shown in Table 2-5 are expressed as coefficients of variation, a statistical term obtained by dividing the standard deviation of a parameter by its mean value and multiplying the result by 100 to obtain a percent value. Plots of the volumetric flow values and the respective uncertainties associated with each engineering method are presented in Figures C-3-1 through C-3-3 in Appendix C-3.

Examination of Table 2-5 provides the following information:

- The MMBtu method had the largest uncertainty at Homer City, while the CO_2 F-factor method had the largest uncertainty at DeCordova and Lake Hubbard.
- The coefficients of variation at the coal-fired site were two to three times larger than those at the gas-fired sites.
- The uncertainties at the gas-fired sites were, except for the CO_2 F-factor method, less than half of those at Homer City.

Table 2-5. Pre- and Post-test Error Estimates of Engineering Methods

Flow Calculation Method	Pre-Test Uncertainty Analysis		Post-Test Uncertainty Analysis		
	Resulting Uncertainty (%)		Coefficient of Variation (%)		
	Natural Gas-fired[a]	Coal-fired[b]	DeCordova	Lake Hubbard	Homer City
O_2 F-Factor	2.5 to 3.5	2.5 to 3.5	1.20	0.52	2.88
CO_2 F-Factor	5 to 6	5 to 6	1.62	0.68	2.97
MMBtu	2 to 3.5	2 to 3.5	0.65	0.40	3.16
ASME PTC 4.1	5 to 6	5 to 6	1.40	0.49	2.89
BTCE	3 to 3.5	3 to 3.5	0.77	0.35	2.39

[a] In calculating the uncertainties for each engineering method at the gas-fired sites, the following uncertainties were associated with key input parameters: 1-2% in HHV, 1-2% in M_{fuel}, 5% in O_2, 5% in CO_2, 2-3% in HR_{cycle}, 0.1% in $O_{2,stack}$, 5% in excess air coefficient, 1% in fuel composition, 1% in F_d, 1% in F_c, and 1% in A_{dry}. (See Appendix H for definitions of terms.)

[b] At the coal-fired site, the same uncertainties were associated with the input parameters as at the gas-fired sites. In addition, an uncertainty of 2% Abs. was associated with loss on ignition (LOI).

SECTION 3
ANALYSIS OF MATRIX A AND MATRIX B

This section analyzes the Matrix A (interprobe) and Matrix B (intraprobe) test results. Section 3.1 presents between-method comparisons using a series of analytical approaches. Section 3.2 examines each in-stack method individually to determine whether the choice of particular probe copies and test teams affected the results obtained. In Section 3.3, the variability of each method is examined. Section 3.4 summarizes the results presented in Sections 3.1 and 3.2 in terms of two key analytical factors: proximity to the central tendency of the data and the comparative variability of the methods. Section 3.5 compares the expected uncertainty in engineering methods with the expected uncertainty in the probe methods.

The analysis presented in this section represents the first of two rounds of analyses of the field test data. A supplemental round of analyses was conducted in response to recommendations provided to EPA by a panel of peer reviewers based on their review of the analyses presented in this section. These follow-up analyses are presented in Section 4.

3.1 COMPARISON ACROSS PROBES AND METHODS

Four analytical approaches are used in the sections below. Section 3.1.1 uses descriptive statistics and graphs to describe the flow characteristics at each site and to compare the volumetric flow results obtained by the various methods. Section 3.1.2 presents the results of rank order analyses to reveal patterns among the methods more clearly. Section 3.1.3 applies an analysis of variance (ANOVA) to the Matrix A volumetric flow values to determine whether statistically significant differences can be detected among the individual methods and classes of methods. Section 3.1.4 analyzes the proximity of each method's measurements to the central tendency of the data to discover whether certain methods are more likely to be good indicators of the average, long-term volumetric flow.

3.1.1 Relationships Among Probes and Methods

The following analyses use descriptive statistics and graphs to evaluate the relationships among probes and methods in the Matrix A and Matrix B data sets. Specific analyses are presented that provide information on the range of the volumetric flow measurements among all methods, characterization of the range of volumetric flow measurements among in-stack methods only, and the range of volumetric flow measurements among engineering methods only. In addition, volumetric flow values derived from in-stack measurements are compared with those derived from the engineering methods (see Table 3-4). The manual Type S probe straight-up measurements are also compared with the engineering methods (see Table 3-5) to identify patterns among the methods across the DeCordova, Lake Hubbard, and Homer City test sites.

Overall Range of Measurements

This analysis evaluates the extent of the differences in volumetric flow measurements among the methods tested in this study. Only those volumetric flow values obtained during stable load conditions were used in the analysis (e.g., Run 25B at Homer City was dropped from the analysis, as described in Section 2.1.4). The average volumetric flow value for each method tested in the experimental design was calculated. The methods with the largest and smallest average values are shown in Table 3-1. In addition to the average values for these methods, the difference between the maximum and minimum average value is presented to show the range of volumetric flow

measurements at a specific site. Because the flow is stable across the experiment (see Section 2.1.4), differences in volumetric flows can be attributed to inherent differences among methods. To simplify an examination of the range across different sites, the ratio of the range (maximum minus minimum) to the average of all methods was computed and is displayed in Table 3-1.

Table 3-1. Overall Range in Flow Measurements

	DeCordova		Lake Hubbard[a]		Homer City	
	Avg. Value	**Method**	**Avg. Value**	**Method**	**Avg. Value**	**Method**
Max (wscfm)	1,631,513	S Straight-up	1,157,629	Modified Kiel[b]	1,466,549	S Straight-up
Min (wscfm)	1,528,137	O_2 F-factor	1,036,472	Spherical	1,359,861	French
Difference	103,176 (6.58%)[c]		121,157 (11.03%)[c]		106,688 (7.56%)[c]	

[a] High-load runs only.
[b] The modified Kiel probe produced the highest measurement of flow when measurements taken only under high-load operation were included. When low-load measurements were also included, the Type S probe straight-up gave the highest average measured flow; see Tables 3-6 through 3-8.
[c] Ratio of difference relative to average of all in-stack and engineering methods, expressed as a percent.

Table 3-1 shows that the range of average flow measurements was greatest at Lake Hubbard and least at DeCordova. At two of the three sites, the maximum mean difference was produced by the manual Type S probe operated in straight-up mode. At Lake Hubbard (under high-load operating conditions only), the maximum average value was produced by the modified Kiel probe. At two sites, the lowest measurements were produced by two new probes being evaluated in this study, the French and spherical probes.

In-stack methods characterization

In Table 3-2, the range of volumetric flow values is characterized for in-stack methods only. The objective was to identify which particular methods defined the extremes of the range. As before, average volumetric flow values were calculated for each in-stack method with Run 25B at Homer City dropped from the calculation.

Table 3-2. Range in Flow Measurements for In-stack Methods

	DeCordova		Lake Hubbard[a]		Homer City	
	Avg. Value	**Method**	**Avg. Value**	**Method**	**Avg. Value**	**Method**
Max (wscfm)	1,631,513	S Straight-up	1,157,629	Modified Kiel[b]	1,466,549	S Straight-up
Min (wscfm)	1,552,637	BLAP 48-point yaw-nulled	1,036,472	Spherical	1,359,861	French
Difference	78,876 (4.99%)[c]		121,157 (11.01%)[c]		106,688 (7.53%)[c]	

[a] High-load runs only.
[b] The modified Kiel probe produced the highest measurement of flow when measurements taken only under high-load operation were included. When low-load measurements were included, the Type S probe straight-up gave the highest average measured flow; see Tables 3-6 through 3-8.
[c] Ratio of difference relative to average of all in-stack methods, expressed as a percent.

Table 3-2 shows that, at Lake Hubbard and Homer City, the in-stack measurements (by the spherical and French probes, respectively) defined the lower extremes in the average volumetric flow. At DeCordova, the lowest mean average volumetric flow was obtained from the baseline Autoprobe 48-point, yaw-nulled. The highest average in-stack methods were the manual Type S straight-up at DeCordova and Homer City and the modified Kiel at Lake Hubbard.

Engineering methods characterization

The results of performing the same calculations only on the engineering method data are presented in Table 3-3. Not surprisingly, the difference among engineering methods was greatest at the coal-fired facility, Homer City, where tracking combustion factors is most difficult.

Table 3-3. Range in Flow Measurements for Engineering Methods

	DeCordova		Lake Hubbard[a]		Homer City	
	Avg. Value	Method	Avg. Value	Method	Avg. Value	Method
Max (wscfm)	1,547,958	MMBtu	1,109,560	ASME	1,420,873	CO_2 F-factor
Min (wscfm)	1,528,137	O_2 F-factor	1,077,270	BTCE	1,366,783	BTCE
Difference	19,821 (1.29%)[b]		32,290 (2.95%)[b]		54,089 (3.87%)[b]	

[a] High-load runs only.

[b] Ratio of difference relative to average of all engineering methods, expressed as a percent.

Comparison of in-stack and engineering methods

In-stack and engineering volumetric flow methods are compared in Table 3-4. The numbers of runs analyzed are shown in parentheses. First, the average volumetric flow measurements of all in-stack methods and of all engineering methods were computed separately. The difference in these values was computed, and the ratio of the difference to the average engineering value was calculated. The resulting value is the percent difference in the two classes relative to the engineering method average volumetric flow. Second, the same calculation was performed using only the manual Type S straight-up method in place of all in-stack methods, and only the engineering values for the runs when the manual Type S probe was tested were used to compute the average. The Type S straight-up method is the current Method 2 approach.

Table 3-4. In-stack and Engineering Methods Comparison[a]

Comparison	DeCordova	Lake Hubbard	Homer City
In-stack Avg. vs. Engineering Avg.	2.50% (32)	0.26% (20)	1.41% (31)
S-straight Avg. vs. Engineering Avg.	7.10% (8)	5.44% (8)	5.43% (8)

[a] Number of runs analyzed is shown in parentheses.

The difference between the engineering methods and the manual Type S probe straight-up is further shown in Table 3-5. Table 3-5 presents the average and maximum percent difference of the manual Type S probe's volumetric flow measurements from the volumetric flow values derived by each of the engineering methods. These statistics were calculated using all Matrix A and Matrix B runs in which a Type S probe was tested. The number of tests used in the calculations is also shown.

Table 3-5. Average Difference of Manual Type S Probe Straight-up from Engineering Methods for Matrix A and Matrix B Runs

	DeCordova		Lake Hubbard (High Load)		Homer City	
	20 Tests		21 Tests		20 Tests	
Method	Average	Maximum	Average	Maximum	Average	Maximum
O_2 F-Factor	7.57%	10.54%	5.56%	11.95%	3.28%	9.61%
CO_2 F-Factor	7.40%	10.14%	3.54%	9.51%	3.35%	9.92%
MMBtu	6.28%	9.26%	4.33%	10.11%	6.02%	12.65%
ASME PTC 4.1	7.02%	9.82%	3.15%	8.80%	5.83%	12.49%
BTCE	6.17%	8.94%	6.40%	11.26%	9.42%	13.79%

Table 3-5 shows that, depending on the engineering method used, the average difference ranged from 6.17% to 7.57% at DeCordova, 3.15% to 6.40% at Lake Hubbard, and 3.28% to 9.42% at Homer City. If the BTCE results at Homer City are excluded, the data indicate that the difference was greatest at the site where the flow was closest to axial (DeCordova) and lower at the other two sites. This observation is inconsistent with the general perception that non-axial flow heightens the disparity between manual Type S straight-up measurements and engineering method calculations.

Consistent patterns across sites

Using the data in Tables 3-6 through 3-8, patterns among the methods were identified across DeCordova, Lake Hubbard, and Homer City. To facilitate comparisons across sites with significantly different volumetric flow values, the flow values produced by each tested probe/ procedure combination and engineering method were normalized relative to measurements produced by a baseline method.

The baseline chosen for the current analysis was the set of measurements taken by the 16-point baseline Type S Autoprobes operated in straight-up mode, that is, without yaw-nulling. (Comparisons using two other methods as baselines, i.e., the BTCE and MMBtu methods, appear in the series 13 and 14 tables included in the site data reports for the DeCordova, Lake Hubbard, and Homer City field tests.) As was noted in the original test plans for the field studies, comparison of measurements to a particular "baseline" is simply a tool for analyzing and displaying the relative magnitudes and variability of the volumetric flow obtained by each of the tested methods[5]. *Choosing the 16-point straight-up Autoprobe measurements as a common denominator for comparing the various methods does not imply that its measurements represent or approximate the true flow rate any more accurately than the other tested methods.* In part, this set of data was chosen because the baseline Autoprobes were operated during all runs at all three sites and are allowed to be used in current Method 2.

[5] See, for example, The Cadmus Group, Inc., 1997, "Flow Reference Method Testing and Analysis: Field Test Plan, DeCordova Steam Electric Station," EPA/430/R-97-024.

Table 3-6. Summary of Volumetric Flow Results—DeCordova

Probe/Method	No. Tests	Percent Difference From Baseline[a]			
		Mean	Minimum	Maximum	s.d.
DAT	30	-0.18%	-4.43%	2.86%	2.03%
Type S Straight-up	20	4.50%	1.08%	7.38%	1.62%
Type S Yaw-nulled	20	4.00%	0.59%	7.98%	1.80%
Modified Kiel	20	1.76%	-0.42%	5.41%	1.37%
Spherical	9	0.75%	-1.09%	2.81%	1.35%
Prandtl	20	3.70%	1.89%	5.28%	0.92%
French	20	0.09%	-3.37%	2.94%	1.68%
Manual Autoprobe 16-point Straight-up	4	0.48%	-0.03%	1.15%	0.50%
Manual Autoprobe 16-point Yaw-nulled	4	0.35%	0.10%	0.60%	0.21%
Baseline Autoprobes 16-point Straight-up[a]	*31*	*0.00%*	*0.00%*	*0.00%*	*0.00%*
Baseline Autoprobes 16-point Yaw-nulled	22	-0.14%	-0.45%	0.33%	0.19%
Baseline Autoprobes 48-point Straight-up	31	-0.64%	-0.91%	-0.49%	0.11%
Baseline Autoprobes 48-point Yaw-nulled	22	-0.75%	-0.97%	-0.39%	0.16%
O_2 F-Factor	31	-2.38%	-4.17%	-1.04%	0.76%
CO_2 F-Factor	31	-1.51%	-4.64%	0.40%	1.32%
MMBtu	31	-1.14%	-2.31%	-0.15%	0.63%
ASME PTC 4.1	31	-1.21%	-4.20%	0.36%	1.18%
BTCE	31	-1.58%	-2.49%	-0.46%	0.41%

[a] Baseline is baseline Autoprobes 16-point straight-up.

The box-and-whisker plots shown in Figures 3-1 through 3-6 provide additional visual representations of the data presented in Tables 3-6 through 3-8[6]. (Note that Figures 3-4 through 3-6 illustrate the same results as presented in Tables 3-1 through 3-3, but the scales have been normalized.) The following observations can be made:

- At each site, the performance characteristics of the single manual Autoprobe were very similar to those of the baseline four-Autoprobe system.
- At all three sites, the manual Type S straight-up measurements exhibited a higher average positive percent difference from the baseline than all other methods.
- The Prandtl probe exhibited the lowest standard deviation of percent difference in volumetric flow from the baseline of any of the in-stack methods tested at DeCordova and Lake Hubbard, except the Autoprobes. (The Prandtl probe was not tested at the coal-fired Homer City site, because of the possibility of plugging.)

6 The box-and-whisker plot is a useful tool for displaying the distribution of a data set. The plots show the median (middle horizontal line), the 25th and 75th percentiles (lower and upper horizontal lines of the boxes), and the minimum and maximum values (lower and upper ends of the whiskers, or lowest and highest asterisks [*]). Potential outliers are indicated by asterisks (*) and the criterion for identifying the potential outliers is any value beyond the distance of 1.5 times the interquartile range (75th percentile value minus the 25th percentile value) from the box.

Table 3-7. Summary of Volumetric Flow Results— Lake Hubbard

Probe/Method	No. Tests	Percent Difference From Baseline[a]			
		Mean	Minimum	Maximum	s.d.
DAT	29	-2.98%	-13.44%	0.54%	3.10%
Type S Straight-up	25	3.65%	0.98%	7.92%	1.83%
Type S Yaw-nulled	25	1.72%	-3.70%	4.93%	2.16%
Modified Kiel	8	3.11%	-4.24%	12.04%	4.71%
Spherical	25	-7.17%	-38.73%	-1.22%	7.84%
Prandtl	4	0.93%	-0.30%	3.12%	1.52%
French	4	-3.24%	-7.37%	-0.16%	3.16%
Manual Autoprobe 16-point Straight-up	4	0.09%	-0.71%	0.97%	0.75%
Manual Autoprobe 16-point Yaw-nulled	4	-0.93%	-1.58%	-0.29%	0.52%
Baseline Autoprobes 16-point Straight-up[a]	*25*	*0.00%*	*0.00%*	*0.00%*	*0.00%*
Baseline Autoprobes 16-point Yaw-nulled	25	-1.08%	-2.93%	0.47%	0.66%
Baseline Autoprobes 48-point Straight-up	24	-0.99%	-1.48%	-0.18%	0.36%
Baseline Autoprobes 48-point Yaw-nulled	24	-2.06%	-3.13%	-0.88%	0.51%
O_2 F-Factor	25	-1.75%	-3.60%	0.78%	1.12%
CO_2 F-Factor	25	-0.02%	-1.78%	2.69%	0.96%
MMBtu	25	-0.50%	-1.99%	1.21%	0.78%
ASME PTC 4.1	25	0.49%	-0.81%	2.37%	0.76%
BTCE	25	-2.31%	-3.31%	1.60%	0.99%

[a] Baseline is baseline Autoprobes 16-point straight-up.

Table 3-8. Summary of Volumetric Flow Results—Homer City

Probe/Method	No. Tests	Percent Difference From Baseline[a]			
		Mean	Minimum	Maximum	s.d.
DAT	23	-3.20%	-6.15%	0.59%	1.83%
Type S Straight-up	20	3.48%	1.20%	5.98%	1.23%
Type S Yaw-nulled	20	-1.12%	-3.87%	2.78%	1.67%
Modified Kiel	20	-2.39%	-6.70%	1.11%	1.99%
Spherical A	20	-3.88%	-6.36%	-0.15%	1.70%
Spherical B	20	-4.36%	-8.65%	0.17%	2.06%
French	20	-6.29%	-11.78%	-2.30%	2.85%
Manual Autoprobe 16-point Straight-up	4	0.18%	0.04%	0.45%	0.19%
Manual Autoprobe 16-point Yaw-nulled	4	-3.41%	-3.57%	-3.22%	0.18%
Baseline Autoprobes 12-point Straight-up	17	0.27%	-1.73%	2.56%	1.05%
Baseline Autoprobes 12-point Yaw-nulled	17	-4.14%	-5.64%	-2.59%	0.71%
Baseline Autoprobes 16-point Straight-up[a]	*32*	*0.00%*	*0.00%*	*0.00%*	*0.00%*
Baseline Autoprobes 16-point Yaw-nulled	28	-4.33%	-4.77%	-3.68%	0.33%
Baseline Autoprobes 20-point Straight-up	15	-0.55%	-2.27%	0.63%	0.74%
Baseline Autoprobes 20-point Yaw-nulled	11	-4.65%	-5.81%	-2.55%	0.88%
O_2 F-Factor	32	-2.23%	-14.78%	2.18%	2.96%
CO_2 F-Factor	32	-2.03%	-14.39%	2.58%	2.90%
MMBtu	32	-4.60%	-17.71%	-0.59%	3.01%
ASME PTC 4.1	32	-4.06%	-17.12%	-0.12%	2.94%
BTCE	32	-5.81%	-8.56%	-4.30%	0.98%

[a] Baseline is baseline Autoprobes 16-point straight-up.

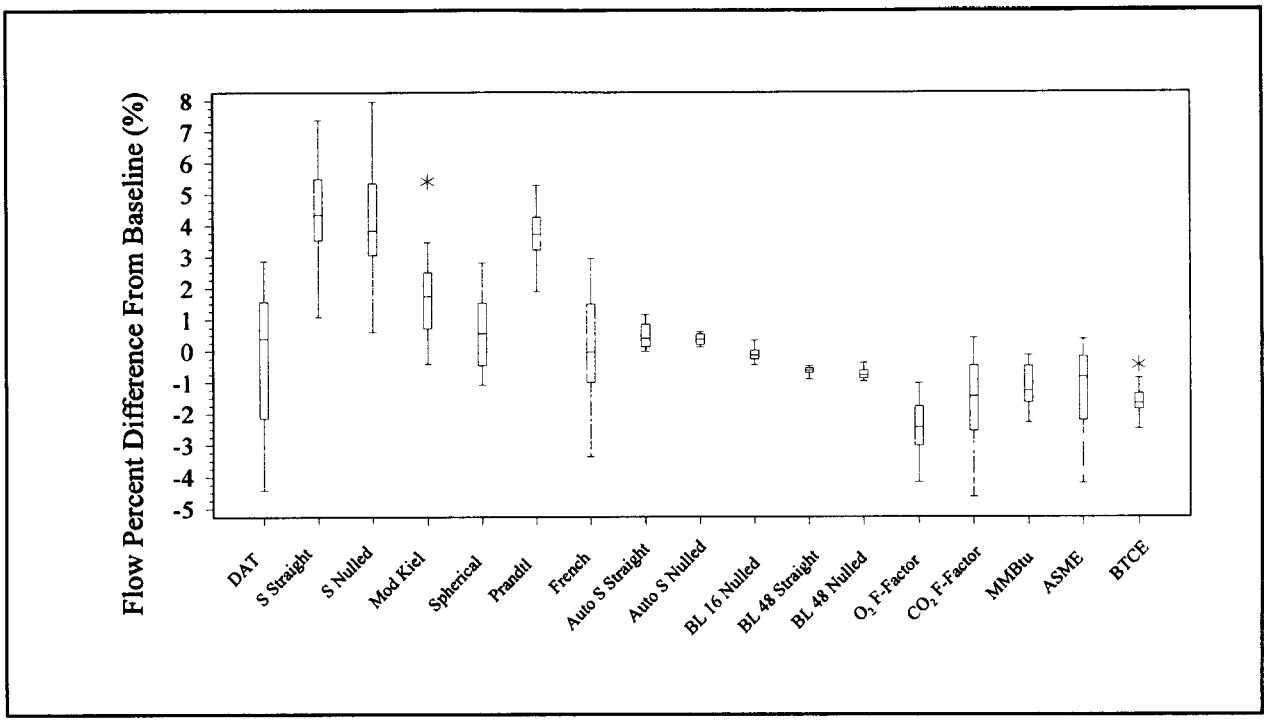

Figure 3-1. Flow percent difference from baseline Autoprobes 16-point, straight-up at DeCordova (scale 1).

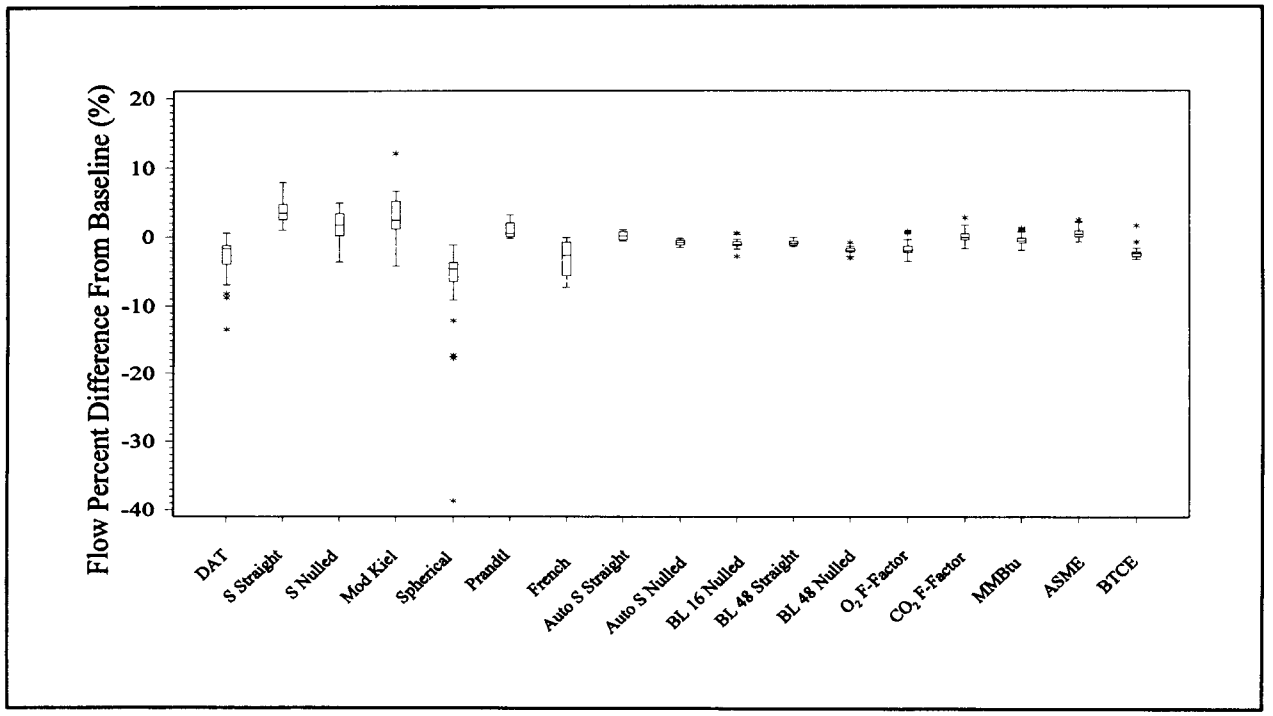

Figure 3-2. Flow percent difference from baseline Autoprobes 16-point, straight-up at Lake Hubbard (scale 1).

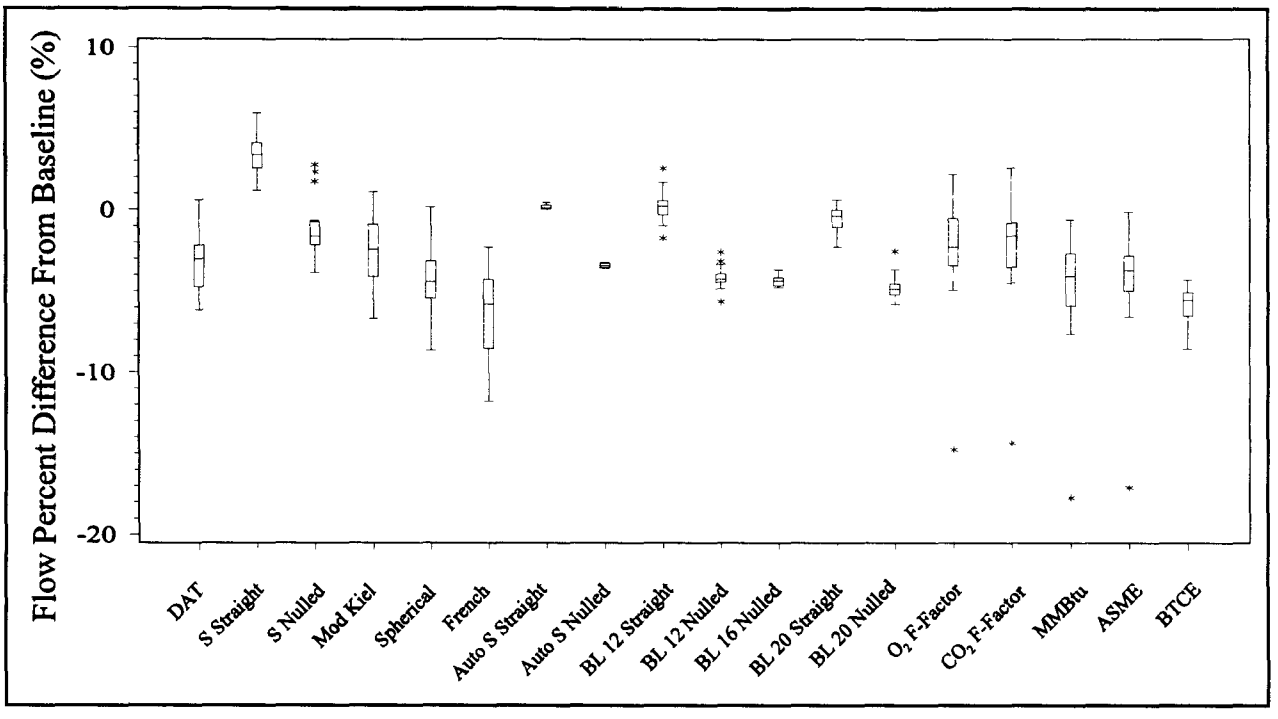

Figure 3-3. Flow percent difference from baseline Autoprobes 16-point, straight-up at Homer City (scale 1).

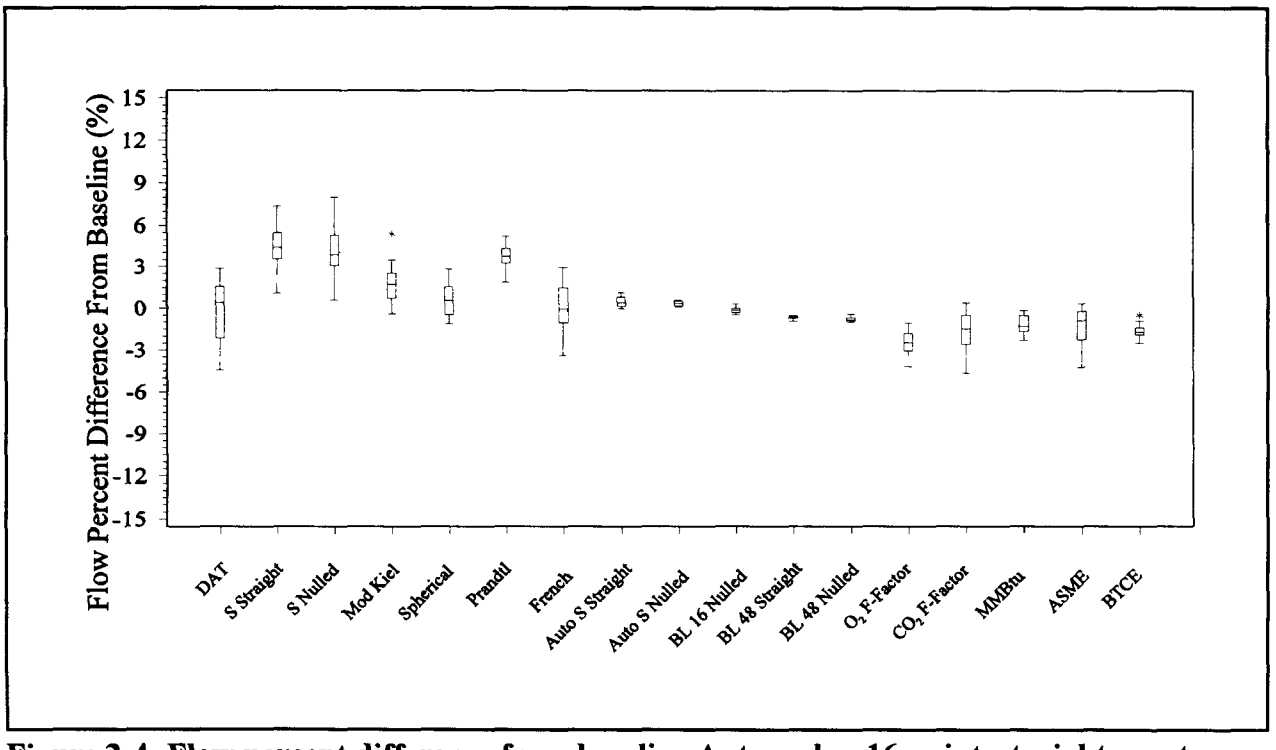

Figure 3-4. Flow percent difference from baseline Autoprobes 16-point, straight-up at DeCordova (scale 2).

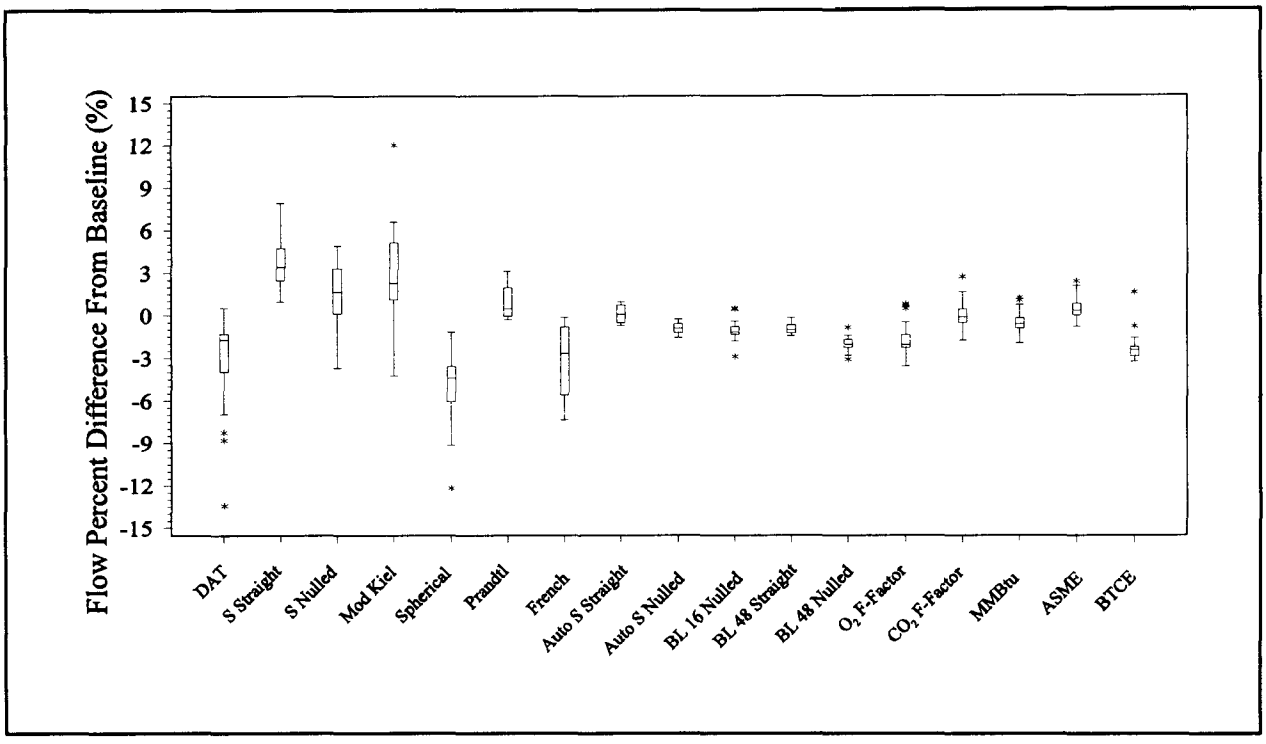

Figure 3-5. Flow percent difference from baseline Autoprobes 16-point, straight-up at Lake Hubbard (scale 2).

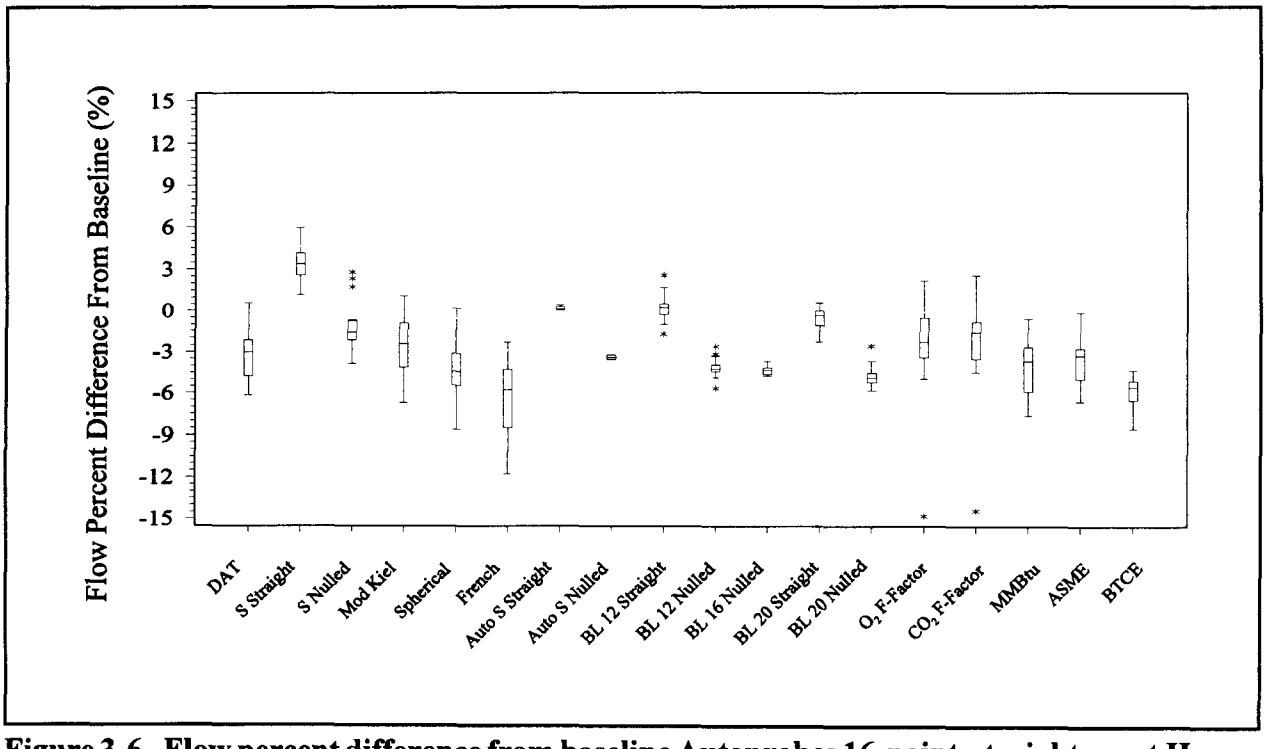

Figure 3-6. Flow percent difference from baseline Autoprobes 16-point, straight-up at Homer City (scale 2).

- The yaw-nulled Type S probes (both manual and automated) always resulted in volumetric flow values that were on average less than the Type S straight-up values.

- At all sites, the range (maximum minus minimum) in the percent differences from the baseline Autoprobe 16-point straight-up reported for the DAT probes, the yaw-nulled 16-point Autoprobes (both manual and baseline), and the MMBtu method was less than 2.5% (Table 3-9).

Table 3-9. DAT, Autoprobe Yaw-Nulled, and MMBtu Performance: Percent Difference From Baseline

Probe/Method	DeCordova	Lake Hubbard[a]	Homer City
DAT	-0.18%	-2.98%	-3.20%
Manual Autoprobe 16-point Yaw-nulled	0.35%	-0.93%	-3.41%
Baseline Autoprobes 16-point Yaw-nulled	-0.14%	-1.08%	-4.33%
MMBtu	-1.14%	-0.50%	-4.60%
Maximum Value	0.35%	-0.50%	-3.20%
Minimum Value	-1.14%	-2.98%	-4.60%
Range (maximum - minimum)	1.49%	2.48%	1.40%

[a] All loads included.

3.1.2 Rank Order Analysis

Rank order tables assign a discrete integer "rank" (in this case, 1 for the lowest measurement, 2 for the next lowest measurement, etc.) to the volumetric flow value determined by each method for each run. Assigning discrete integer values is useful for revealing among-methods patterns that are often difficult to detect when data are expressed in units of volumetric flow or percent difference from a baseline.

A complete set of rank order tables for each field test is found in the three individual site data reports. An aggregate rank order summary for the Matrix A tests across the three test sites is presented in Table 3-10. The table was created using the following steps: (1) within each of the eight Matrix A runs at each site, the methods were ranked from lowest to highest based on the volumetric flow value obtained; and (2) for each method, the median rank and the range in ranks across all runs were calculated. These values are displayed in Table 3-10; details are presented in Appendix D.

The rank order analysis enables three questions to be examined:

1. Are there discernible patterns in the magnitudes of the volumetric flow measurements among methods, for example, do some methods consistently display high rankings, while others display mid-range or low rankings?

2. How are the rankings affected by changes in flow conditions from site to site?

3. Are the rankings of some methods more variable than other methods?

Table 3-10. Aggregate Rank Order of Volumetric Flow at all Sites Across all Methods for Matrix A (Runs 1-8)

Probe/Method	DeCordova		Lake Hubbard		Homer City	
	Rank Median	Rank Range[a]	Rank Median	Rank Range[a]	Rank Median	Rank Range[a]
DAT	10.5	12	6.0	11	6.5	9
Type S Straight-up	13.0	2	13.0	1	14.0	0
Type S Yaw-nulled	12.5	4	12.0	11	11.0	6
Modified Kiel	13.0	2	14.0	2	9.0	8
Spherical A[b]	11.5	2	1.0	1	5.5	4
Spherical B[b]	–	–	–	–	6.0	7
Prandtl[c]	14.0	0	13.0	2	–	–
French	13.0	7	1.5	10	2.5	7
Baseline Autoprobes 12-point Straight-up[d]	–	–	–	–	13.0	3
Baseline Autoprobes 12-point Yaw-nulled[d]	–	–	–	–	4.5	8
Manual Autoprobe 16-point Straight-up	11.5	4	11.5	5	13.0	1
Manual Autoprobe 16-point Yaw-nulled	11.0	2	7.0	3	6.0	3
Baseline Autoprobes 16-point Straight-up	9.0	2	11.0	4	12.0	2
Baseline Autoprobes 16-point Yaw-nulled	9.0	3	6.5	4	4.0	3
Baseline Autoprobes 48-point Straight-up[e]	7.0	2	7.0	4	–	–
Baseline Autoprobes 48-point Yaw-nulled[e]	6.5	2	4.0	2	–	–
O_2 F-Factor	1.5	2	3.0	3	10.0	2
CO_2 F-Factor	2.0	2	9.5	5	10.0	3
MMBtu	4.5	3	7.5	3	5.0	7
ASME PTC 4.1	3.0	2	11.5	3	6.0	6
BTCE	5.0	2	2.0	3	1.5	2

[a] Rank range is the difference between the largest and smallest rank.
[b] A = spherical probe copies SPH-1, SPH-2, SPH-3, and SPH-4; see Section 2.2.1.
B = spherical probe copies SPH-5, SPH-6, SPH-7, and SPH-8; see Section 2.2.1.
[c] Prandtl probes were not tested at Homer City because the 0.035-in. I.D. static pressure ports could become clogged with particulate matter at this site.
[d] Twelve-point traverses were performed by the baseline Autoprobes at Homer City only, for comparison with results for 16- and 20-point traverses.
[e] At Homer City, 20-point traverses (results not shown in table) instead of 48-point traverses were performed by the baseline Autoprobes for comparison with results for 12- and 16-point traverses.

Examination of Table 3-10 provides the following observations:

- The DAT probe showed the largest range of ranks at all three sites, indicating a large variance in the run-by-run observations.

- The engineering methods had low ranks at DeCordova and low-to-moderate ranks at Lake Hubbard and Homer City. This observation indicates that the engineering methods provided the lowest volumetric flow values of all methods at the site with axial flow (DeCordova).

- The French probe showed the greatest change in ranking from site to site, providing a relatively high rank at DeCordova, but very low ranks at Lake Hubbard and Homer City.

- The baseline Autoprobe had generally moderate ranks at all sites, indicating that these methods provided measurements close to the overall median flow value.

- The Prandtl probe provided high readings relative to the other probes at the two sites where it was tested.

- The manual Type S (straight-up and yaw-nulled) probes displayed high ranks across all three sites, indicating a consistently high reading relative to the other probes.

The relatively high rank orders (10.5 and 11.5) for the 3-D probes at the axial-flow site are not surprising, because at DeCordova these probes function essentially as "straight-up," manual methods in axial flow conditions.

The low-to-moderate rank orders (6.0 and 6.5) for the DAT probes at Lake Hubbard and Homer City are as expected. The 3-D probes would be expected to provide lower results at these sites than the 1-D and 2-D probes, because significant yaw and/or pitch angles were present at both sites. The rank orders of 5.5 and 6.0 for the two sets of spherical probes tested at Homer City are also expected. The rank order of 1.0 for the spherical probe at Lake Hubbard, however, is not expected and implies that the spherical probe data from that site are suspect.

At Lake Hubbard and Homer City, where significant yaw angles were present, the rank orders of the12- and 16-point straight-up manual and baseline Autoprobes were relatively high, while the rank orders of the yaw-nulled Autoprobes were low to moderate. This observation is also as expected.

The rank order of 9.0 for the yaw-compensated modified Kiel probe at Homer City is as expected. However, the rank order of 14.0 for the probe at Lake Hubbard is surprising. Although the average yaw angles at Lake Hubbard were not extreme (6° to 7°), one would not expect yaw-compensated readings with the Kiel probe to be the *highest* readings among all of the methods.

3.1.3 Analysis of Variance of Matrix A Data

Discussion of Analysis of Variance Techniques

Analysis of variance (ANOVA) is a statistical method for indicating whether the mean values produced by different groupings (i.e., methods, probe copies, test teams) in this field test are

important predictors of volumetric flow. Multiple comparison methods (such as Duncan's Multiple Range Test), performed in conjunction with the ANOVA, can be used to determine which specific methods (or probe copies or test teams) produce results that statistically differ from one another.

In the first step of performing an ANOVA, a statistical model is created that is consistent with the experimental design. (See Appendix E for a more detailed description of the model.) The model, which consists of several terms, is fit to the data. Each term represents a different grouping. Measures of the relative amount of variation (called mean squares) associated with each grouping are calculated and compared to the remaining or unexplained variation in the data (the error term). If a mean square is large, that grouping is considered to be important in determining volumetric flow (given a pre-specified level of confidence). For example, a large mean square associated with the test teams term in the ANOVA model indicates that the teams themselves are important in influencing the volumetric flow measurements.

The second step of an ANOVA is performing a multiple comparison means test. The means test indicates which group member can be distinguished from the others (e.g., which probe copy can be distinguished from the three other probe copies). The mathematical representations of the models are presented in Appendix E.

Matrix A Analysis of Variance Description and Results

Several ANOVA models were applied to the Matrix A data and evaluated. For example, a model with a single grouping ("method") was evaluated as an approach for distinguishing the relative importance of the 16 methods as significant predictors of volumetric flow. The results of this model were inconclusive and showed that the relationships among methods were complex, with no single method or group of methods clearly distinguishable from the others.

To simplify the model and further the evaluations, another grouping called "class" was developed. Methods with similar characteristics were grouped into classes. The first variation of the approach (Variant 1) created four separate classes, one each for the 1-D, 2-D, and 3-D probes, and one for the engineering methods. The French probe, alone, was included in a fifth class because earlier analyses showed that the probe tended to produce lower results relative to other 1-D methods. A second set of classes (Variant 2) consisting of only the in-stack methods with the automated probes separated from their manual counterparts was also created. In Variant 2, the classes were 1-D manual, 2-D manual, 3-D, French, 1-D Autoprobes, and 2-D Autoprobes.

Within each class, each method had the same number of observations. The number of observations *across* classes is not consistent, but a sensitivity check of the model (using simulated data where the sample size was kept the same across classes) indicated that this issue had no effect on the results. Therefore, the ANOVA results are not dependent on the relative sample size for any one probe.

Both models (Variant 1 and Variant 2) were shown to be significant (α=0.05). Therefore, multiple comparison testing was performed on the class means. In multiple comparison tests, the means and variances of the data for each class are compared in succession to determine whether a statistically significant difference can be detected between each pair of class means. Table 3-11 presents the results of the means tests, using Duncan's Multiple Range Test. Other multiple comparison tests (e.g., Tukey's Multiple Range Test) produced similar results.

Table 3-11. Duncan's Multiple Range Test on Probe Classes (Matrix A, $\alpha = 0.05$)

Var.	DeCordova	Lake Hubbard		Homer City
		High Load	Low Load	
1[a]	(1D, 2D, F, 3D) ≠ E	(1D, 2D, E) ≠ 3D	(1D, 2D, E) ≠ 3D	1D ≠ 2D, E, 3D, F 2D ≠ F
2[b]	$1D_m \neq (F, 3D_m, 1D_a, 2D_a)$ $2D_m \neq (3D_m, 1D_a, 2D_a)$	$(2D_m, 1D_m, 1D_a) \neq 3D_m$	$(1D_m, 2D_m, 1D_a, 2D_a) \neq 3D_m$	$1D_m \neq 1D_a, 2D_m, 2D_a, 3D_m, F$ $1D_a \neq 2D_a, 3D_m, F$ $2D_m \neq 3D_m, F$

[a] Variant 1 Classes
- 1D = Type S straight-up, baseline Autoprobe straight-up, manual Autoprobe straight-up, Prandtl
- 2D = Type S yaw-nulled, baseline Autoprobe yaw-nulled, manual Autoprobe yaw-nulled, modified Kiel
- 3D = DAT, spherical
- F = French
- E = Engineering methods (O_2 F-factor, CO_2 F-factor, MMBtu, ASME PTC 4.1, BTCE)

[b] Variant 2 Classes
- $1D_m$ = manual 1-D methods (Type S straight-up and Prandtl)
- $2D_m$ = manual 2-D methods (Type S yaw-nulled and modified Kiel)
- $3D_m$ = 3-D Autoprobes methods (DAT and spherical)
- $1D_a$ = 1-D Autoprobes methods (baseline Autoprobe straight-up)
- $2D_a$ = 2-D methods (baseline Autoprobe yaw-nulled)
- F = French

In Variant 1 at DeCordova, the engineering methods are separable from the other methods. This relationship, however, does not hold at the other sites. For Variant 1, the relationships among the classes are not consistent from site to site. At the moderate-yaw/pitch (Lake Hubbard) and high-yaw (Homer City) sites, however, the mean flow values determined by the 1-D and 3-D probes are statistically distinguishable. Similar results are obtained for the Variant 2 classes. In Variant 2, a statistically significant difference between the means for the manual 3-D probes and manual 1-D probes is observed at all three sites.

3.1.4 Analysis of Central Tendency

Difference from Grand Mean

An important feature of any measurement method is the degree to which it provides values that are close to the long-term average or central tendency of measurements made by all viable methods. Although central tendency does not explicitly reflect the sampling variance of the method (for a direct analysis of sampling variance, see Section 3.3), methods with high central tendency do reflect the long-term average of all viable methods for measuring volumetric flow. In situations where the true or accurate value of volumetric flow at any point in time is unknown, central tendency can be used as an indicator of the closeness of a method's measurements to the true, but unknown, long-term mean.

All methods evaluated in this study were assumed to be credible candidate indicators of volumetric flow. The degree to which any one method differs from the long-term sample mean can be assessed by comparison with a baseline. A reasonable baseline for this assessment is the mean of all methods

(the grand mean). Those methods that consistently measure values that are much higher or lower than the grand mean are less likely to be good indicators of the true, long-term volumetric flow, as represented by the long-term sampling mean. In addition, those methods that consistently read much lower than the grand mean have the greatest chance of representing a risk to the environment by underestimating the long-term volumetric flow in the stack.

Central tendency analysis relies on the premise that the "majority rules." If on average, most methods offer a good estimate of true flow, methods that are outside the central tendency may be considered less accurate. Alternatively, if a particular method demonstrates superior detection capabilities over most of the other methods, its distance from the central tendency may in fact represent an improvement in accuracy relative to the majority of probes.

Table 3-12 presents the differences from the grand mean for all methods at each of the three sites. For each method, the average difference across all appropriate Matrix A runs is calculated as

$$D_k = \frac{\sum_{i=1}^{r} M_{ik}}{r} - G \qquad \textbf{(Eq. 3-1)}$$

where:

D_k = the difference from the grand mean of method k,
M_{ik} = volumetric flow measured at run i for method k,
G = grand mean of all methods (k) across all runs (i), and
r = number of runs in which method measurements were taken.

The grand mean used to derive the values in Table 3-12 was created by first calculating the mean of all probes on each run, and then calculating the means of those values. This procedure reduced the influence of sample size on the grand mean for individual probes.

The magnitude of a method's difference from the grand mean is an important consideration because the larger the difference, the more questionable the measurements. Additionally, methods that consistently measure low present a potential risk to the environment because they lead ultimately to underestimates of emissions (see discussion of the field study goals in Section 1.2). For these reasons, a bi-directional ranking of methods is presented in Table 3-12. That is, in the columns labeled "rank," the method with the greatest positive difference from the grand mean is given the highest positive ranking, and the method with the greatest negative difference from the grand mean is given the lowest negative rank. For example, at DeCordova, the method with the greatest positive difference and highest positive ranking is the Prandtl probe with a rank of +11, and the method with the greatest negative difference and lowest negative rank is the O_2 F-factor method with a rank of -7. The table shows differences from the grand mean in units of volumetric flow (wscfm) and on a percent basis. Different font attributes are used to indicate various degrees or categories of percent difference.

Table 3-12. Central Tendency Analysis on Matrix A Runs at Each Site

DeCordova				Lake Hubbard (High Load)				Lake Hubbard (Low Load)				Homer City			
Grand Mean = 1,556,062.95 wscfm				Grand Mean = 1,089,755.58 wscfm				Grand Mean = 712,314.81 wscfm				Grand Mean = 1,414,649.87 wscfm			
Rank	Method	Difference		Rank	Method	Difference		Rank	Method	Difference		Rank	Method	Difference	
		wscfm	%			wscfm	%			wscfm	%			wscfm	%
11	**Prandtl**	**67095.8**	**4.3%**	**11**	**Modified Kiel**	**79070.4**	**7.3%**								
10	**Modified Kiel**	**56057.3**	**3.6%**	**10**	**Type S Str.**	**65906.9**	**6.0%**								
9	**Type S Str.**	**55158.5**	**3.5%**	*9*	*Type S Nulled*	*23968.2*	*2.2%*					**9**	**Type S Str.**	**98056.1**	**6.9%**
8	*Type S Nulled*	*45875.8*	*2.9%*	*8*	*Prandtl*	*23726.2*	*2.2%*	**8**	**Type S Str.**	**23901.4**	**3.4%**	**8**	**Man AP S**	**51299.7**	**3.6%**
7	*Spherical A*	*35686.4*	*2.3%*	7	Man AP S	14430.3	1.3%	*7*	*Type S Nulled*	*17214.7*	*2.4%*	*7*	*BAP 12S*	*41645.1*	*2.9%*
6	French	28411.3	1.8%	6	ASME PTC 4.1	13568.9	1.2%	*6*	*ASME PTC 4.1*	*15306.7*	*2.1%*	*6*	*BAP 16S*	*38594.5*	*2.7%*
5	Man AP S	14,902.5	1.0%	5	BAP 16S	12650.9	1.2%	*5*	*Modified Kiel*	*14584.4*	*2.0%*	*5*	*Type S Nulled*	*31963.1*	*2.3%*
4	Man AP N	12,957.6	0.8%	4	CO_2 F-factor	6434.3	0.6%	4	MMBtu	11528.7	1.6%	4	O_2 F-factor	13545.9	1.0%
3	DAT	7554	0.5%	3	Man AP N	3227.7	0.3%	3	CO_2 F-factor	10225.2	1.4%	3	CO_2 F-factor	10073.3	0.7%
2	BAP 16S	4676.2	0.3%	2	MMBtu	2086.3	0.2%	2	BAP 16S	6451.2	0.9%	2	Modified Kiel	4481.9	0.3%
1	BAP 16N	4040.6	0.3%	1	BAP 48S	1988.5	0.2%	1	O_2 F-factor	5448.7	0.8%	1	Man AP N	-1096.7	-0.1%
-1	BAP 48S	-4355.5	-0.3%	-1	BAP 16N	-1226.4	-0.1%	-1	BTCE	-92.3	0.0%	-1	DAT	-12111.6	-0.9%
-2	BAP 48N	-5584.2	-0.4%	-2	DAT	-10633.5	-1.0%	-2	BAP 48S	-137.0	0.0%	-2	ASME PTC 4.1	-18789.5	-1.3%
-3	BTCE	-20682	-1.3%	-3	BAP 48N	-10777.2	-1.0%	-3	BAP 16N	-2964.6	-0.4%	-3	BAP 12N	-21491.2	-1.5%
-4	MMBtu	-22658	-1.5%	-4	O_2 F-factor	-15322.1	-1.4%	-4	BAP 48N	-10024.2	-1.4%	-4	BAP 16N	-21960.8	-1.6%
-5	*ASME PTC 4.1*	*-34888*	*-2.2%*	-5	BTCE	-16809.5	-1.5%	*-5*	*DAT*	*-16210.8*	*-2.3%*	-5	MMBtu	-22164.1	-1.6%
-6	*CO_2 F-factor*	*-40931*	*-2.6%*	*-6*	*French*	*-22242.1*	*-2.0%*	**-6**	**Spherical A**	**-75232.1**	**-10.6%**	-6	Spherical B	*-28033.9*	*-2.0%*
-7	*O_2 F-factor*	*-42508*	*-2.7%*	-7	**Spherical A**	**-133505.6**	**-12.3%**					-7	Spherical A	*-29580.9*	*-2.1%*
												-8	**French**	**-46088.4**	**-3.3%**
												-9	**BTCE**	**-47841.9**	**-3.4%**

[a] **bold = ±3.0% and beyond**, *italics =from ±2.0% up to ±3.0% of grand mean*; normal = within ±2.0% of grand mean

Values within ±2.0% of the grand mean appear in normal font; values from ±2.0% up to ±3.0% of the grand mean are shown in italics; and values that differ from the grand mean by ±3.0% or more are shown in bold.

Subdividing the results into such categories allows the following observations to be made on the probes tested:

- *Within ±2.0% of grand mean*: Both the manual and baseline Autoprobes, when operated in yaw-nulled mode, were always within ±2.0% of the grand mean.

- *Within ±3.0% of grand mean*: The DAT probe, the manual Type S probe yaw-nulled, and baseline Autoprobes straight-up were always within ±3.0% of the grand mean.

- *±3.0% and beyond of grand mean*: All other probes (Type S straight-up, Prandtl, manual Autoprobe straight-up, modified Kiel, French, spherical) had excursions beyond ±3.0% of the grand mean at one or more sites. Of these probes, several showed consistent patterns relative to the grand mean. That is, the manual Type S straight-up, Prandtl, manual Autoprobes straight-up, and modified Kiel were always higher than the grand mean. At the two sites where the flow was not near-axial (Lake Hubbard and Homer City), the spherical and French probes were lower than the grand mean.

Although the variability of the Prandtl probe at the two gas sites where it was tested was the lowest of any of the manual probes (coefficients of variation of 0.63% and 1.27% as shown in Section 3.3, below), mean flow readings at both sites were higher than the central tendency. The Prandtl had the highest difference from the central tendency (4.3%) at the near-axial site (DeCordova) and a less severe excursion (2.2%) at the moderate-yaw/moderate-pitch site (Lake Hubbard). Because the Prandtl probe is used as the standard to calibrate other probes in wind tunnels with axial flow, the high difference at the near-axial site was unexpected. On the other hand, its more central behavior at Lake Hubbard suggests that the flow reading taken with the probe may not be adversely affected by moderate yaw and pitch. Note that the relative flow values determined by the Prandtl probe are similar to those found for the manual Type S probe operated in the yaw-null mode.

The central tendency analysis using the grand mean provides an indication of how close the mean volumetric flow obtained by a specific type of probe is to the grand mean for all methods. In addition, line plots with various confidence intervals around the mean flow values for each run corroborate the observations made in the grand mean analysis. Appendix F presents the plots.

3.2 Within-Method Analysis: Analysis of Variance of Matrix B Data

In addition to examining the relationships among the various methods, the importance of specific variance components in the measurement of volumetric flow can be evaluated. The experimental design enables the relative variability contributed to the measurement of volumetric flow due to selection of test teams and the selection of different copies of a specific probe to be evaluated. Assessment of these variance components is an important consideration in determining whether a particular new method or procedure is appropriate for inclusion in a revised test method for volumetric flow. Data generated from Matrix B provide the information for this analysis.

3.2.1 ANOVA on Probe Copies

An ANOVA model with the groupings "probe copy" and "test team" was fit to the Matrix B data (see Appendix E for a description of the model). The model was fit independently to data from each type of probe (e.g., DAT, modified Kiel). In general, four copies of each probe type were analyzed.

The results are presented in Table 3-13. The columns labeled "ANOVA" indicate whether probe copy is a significant factor. A "yes" in this column indicates that the choice among the four available copies of the same probe type made a detectable difference in the flow value obtained. A "no" in this column indicates that the choice of a particular copy had no detectable impact on the measured volumetric flow.

Table 3-13. Results of ANOVA and Multiple Comparison Tests to Detect Differences Among Copies of the Same Probe Type in the Matrix B Data Set

Probe	DeCordova		Lake Hubbard		Homer City	
	ANOVA	Means Test[a]	ANOVA	Means Test[a]	ANOVA	Means Test[a]
DAT	Yes	M	No	-	Yes	M
Spherical A	-[b]	-[b]	No	-	No	-
Spherical B	-	-	-	-	Yes	O (SPH-8)
Modified Kiel	No	-	-	-	Yes	M
Type S Yaw-nulled	No	-	No	-	No	-
Type S Straight-up	Yes	M	No	-	No	-
Prandtl	No	-	-	-	-	-
French	No	-	-	-	No	-

[a] M=Multiple groups of probes identified, no individual outlier probes. O (Probe ID) = An individual outlier probe identified, whose ID is shown in parentheses.

[b] Due to damage of two spherical probes, the Matrix B model could not be run; see Section 2.2.1.

In those cases where a "yes" appears in the ANOVA column, a multiple comparison means test (Duncan's) was performed. The results are displayed in the column labeled "Means Test." If one probe copy was found to yield significantly different results from all other copies, the letter "O" (indicating "one copy") and the identifier code of the distinctive copy appears in this column. If the multiple comparison test indicates that more than one copy yielded significantly different results, the letter "M" appears in the "Means Test" column.

Table 3-13 shows that significant differences among probe copies most often occurred for the DAT probe. The DAT probe copies were shown to have a significant effect at two of the three test sites. However, as indicated by the letter "M" in the "Means Test" columns for these two sites, no individual probe gave significantly different results from the other three DAT probes. It should be recalled that whereas each test team supplied its own DAT probe, all other probes, including the spherical probes, were manufactured at the same time and supplied by EPA. Multiple probe copy effects (indicated the letter "M" in the "Means Test" columns) were also detected at one site for the Type S and Kiel probes. In addition, at Homer City, one of the spherical probes (SPH-8) was significantly different from the other spherical probe copies. No engineering or mechanical reasons

were identified to explain the differences between measurements by this copy of the spherical probe relative to measurements by the other copies of this type probe.

3.2.2 ANOVA on Test Teams

Table 3-14 shows that the test team effect was most often significant for the spherical probes. In all three sets of Matrix B runs performed on the spherical probes at Lake Hubbard and Homer City, the choice of test team significantly influenced the volumetric flow obtained. For the Spherical A probes, one test team's volumetric flow values were significantly different from the others at both Lake Hubbard and Homer City. For the Spherical B probes, the difference among teams was still significant, but no single test team was significantly different from the rest (see Section 2.2.1 for clarification of the definitions of the Spherical A and B probes).

Table 3-14. Results of ANOVA and Multiple Comparison Tests to Detect Differences Among Test Teams in the Matrix B Data Set

Probe	DeCordova		Lake Hubbard		Homer City	
	ANOVA	MeansTest[a]	ANOVA	MeansTest[a]	ANOVA	MeansTest[a]
DAT	No	-	No	-	Yes	M
Spherical A	-[b]	-[b]	Yes	O (Team 1)	Yes	O (Team 1)
Spherical B	-	-	-	-	Yes	M
Modified Kiel	No	-	-	-	No	-
Type S Yaw-nulled	No	-	Yes	M	No	-
S Straight-up	No	-	Yes	M	No	-
Prandtl	Yes	M	-	-	-	-
French	No	-	-	-	No	-

[a] M=Multiple groups of teams identified, no individual outlier teams. O (Team ID) = An individual team outlier identified, whose ID is shown in parentheses.

[b] Due to damage of two spherical probes, the Matrix B model could not be run.

Results obtained using the French and modified Kiel probes were not significantly influenced by the test team. The test team effect for the Prandtl probe was significant at the gas-fired site with near-axial flow (DeCordova), for the Type S probe straight-up and yaw-nulled at the gas-fired site with moderate yaw and pitch (Lake Hubbard), and for the DAT probe at the coal-fired site (Homer City) with high yaw components of flow.

These results indicate that the choice of test team has no consistent effect across sites with the possible exception of the spherical probes. Lack of a significant test-team effect for these methods indicates that the equipment was operated in a consistent manner, which resulted in data with small between test-team variance. In addition, except for the Spherical A probes at Lake Hubbard and Homer City, no one test team produced results that were significantly different from the other three teams. This finding indicates no problem, in general, with any specific probe, piece of equipment, or test team procedure.

The test-team effect, however, is likely to be more of a factor in routine test team performance of the methods than was evident in these field tests. Despite efforts to recruit a representative set of testers for this study, the test teams participating in an EPA-sponsored study of this magnitude, inevitably, represent some of the most qualified and experienced testers in the field. In addition, the field tests were deliberately performed under very controlled conditions with fewer time constraints than are typically faced by testers when performing routine emissions testing. All these factors would tend to minimize the differences among test team results. The differences in results among testers at large could be expected to be considerably greater than those experienced in this study.

3.3 Analysis of Method Variability

Table 3-15 presents a statistical analysis of the variability in each method using the Matrix A data. For each method, the coefficient of variation (CV, defined as the standard deviation divided by the mean times 100) of the Matrix A data is shown. Because the data were obtained over four to eight runs, the calculated CVs represent the temporal sampling variation associated with each probe. (This variation should not be confused with the variance components analysis presented in Section 3.5, where multiple sources of uncertainty are explicitly differentiated in the analysis.) A small CV indicates that the probe provided consistent measurements over the course of Matrix A. For example, a CV of 2% is equivalent to stating that the 95% confidence interval of the mean spans a range of approximately 8%, i.e., 4% on either side of the mean.

A graphical display of the variability of the Matrix B data can be found in Appendix G.

Using both Table 3-15 and Appendix G, the following conclusions can be drawn:

- The variances associated with the Autoprobes and the engineering methods are consistently small (CV < 2.3%) at all sites.
- At the two sites where they were tested, the Prandtl probes displayed the lowest variability of all the manual methods (CVs of 0.63% at DeCordova and 1.27% at Lake Hubbard).
- The original spherical probes (Spherical A) at Lake Hubbard (high and low load) have the highest CVs (19.70% and 7.39%, respectively).
- Relatively high CVs are associated with the modified Kiel probe (Lake Hubbard: 5.19% and 4.04%); French probe (Homer City: 4.18%, Lake Hubbard high load: 3.28%); DAT probe (Lake Hubbard: 3.02% and 3.14%); and Type S yaw-nulled probe (Lake Hubbard high load: 3.23% and Homer City: 3.18%).

3.4 Summary of Central Tendency and Variability Analyses

Figure 3-7 assembles the central tendency and variability analyses for the probe/mode combinations tested in the field study. The figure allows a visual examination in two dimensions of the statistical properties of the tested methods. The following conclusions can be drawn from this figure:

- The manual Autoprobe 16-point yaw-nulled and baseline Autoprobe 48-point straight-up achieved the best (i.e., lowest) central tendency (always within ±1% of the grand mean) and variability results (CVs always below 1%).

Table 3-15. Method Variability Analysis for Matrix A Runs

Probe/Method	CV (%)[a]			
		Lake Hubbard		
	DeCordova	High Load	Low Load	Homer City
DAT	2.66	3.02	3.14	2.08
Type S Straight-up	1.88	2.31	1.04	1.58
Type S Yaw-nulled	2.61	3.23	1.76	3.18
Modified Kiel	1.39	5.19	4.04	1.24
Spherical A	0.90	19.70	7.39	0.93
Spherical B	NA	NA	NA	1.42
Prandtl	0.63	1.27	NA	NA
French	1.62	3.28	NA	4.18
Manual Autoprobe 16-point Straight-up	0.34	0.43	NA	0.90
Manual Autoprobe 16-point Yaw-nulled	0.10	0.25	NA	0.93
Baseline Autoprobes 16-point Straight-up	0.27	0.37	0.91	0.98
Baseline Autoprobes 16-point Yaw-nulled	0.32	0.42	0.63	0.99
Baseline Autoprobes 48-point Straight-up	0.24	0.21	0.31	NA
Baseline Autoprobes 48-point Yaw-nulled	0.32	0.26	0.28	NA
Baseline Autoprobes 12-point Straight-up	NA	NA	NA	1.23
Baseline Autoprobes 12-point Yaw-nulled	NA	NA	NA	0.95
O_2 F-Factor	0.48	0.67	0.59	1.76
CO_2 F-Factor	1.25	0.69	0.76	1.90
MMBtu	0.26	0.38	0.64	1.95
ASME PTC 4.1	1.12	0.41	0.43	2.02
BTCE	0.57	0.30	2.22	0.98

[a] CV = coefficient of variation, (standard deviation / mean) × 100.

- Among the remaining methods, the baseline Autoprobes yaw-nulled (12-, 16-, and 48-point) always achieved the best (i.e., lowest) central tendency results (within ±2% of the grand mean) and variability results (CVs below 1%).

- The DAT, Type S yaw-nulled, and baseline Autoprobe 12-point and 16-point straight-up had comparable central tendency results (always within ±3% of the grand mean), but the variability of the baseline Autoprobes 12-point and 16-point straight-up was lower (CVs consistently below 2%) than the DAT and Type S yaw-nulled (for which neither CV was consistently below 3%).

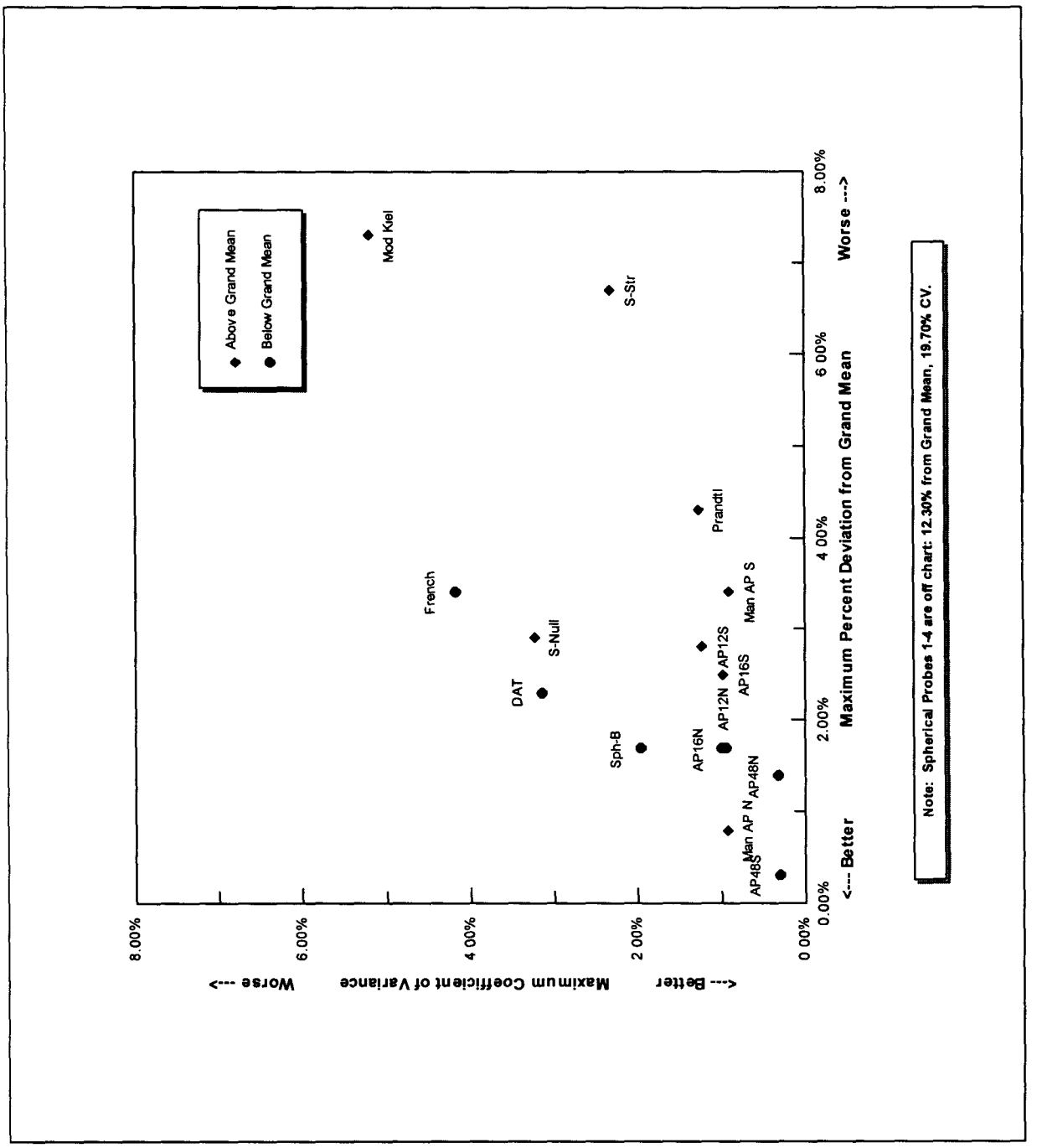

Figure 3-7. Summary of central tendency and variability results: All Matrix A runs, all sites.

- The Type S straight-up and the Prandtl probes had high central tendency results (always beyond ±3%), but mid-range variability results (CVs always below 2% for the Prandtl and always below 3% for the Type S Straight-up).

- The French, modified Kiel, and Spherical A probes had both central tendency and variability results that were high, i.e., not consistently below 3%.

- The Spherical B probes had better central tendency (always within ±2% of the grand mean) and variability results (CVs always below 2%) than the Spherical A (SPH 1-4) probes. (Note that the Spherical A probe values are off the scales and so are not shown in the figure.) Because both Spherical A and B probes, however, were available for testing only at the last field test site (Homer City), these findings should not be considered conclusive.

Table 3-16 summarizes the results of the central tendency and variability analyses.

Table 3-16. Summary of Central Tendency and Variability Analyses: Probe Achievement Levels[a]

Always Within	Central Tendency (% of Grand Mean)	Variability (CV in %)
±1%	BAP 48S Man AP N	BAP 48S BAP 16S Man AP S BAP 48N BAP 16N BAP 12N Man AP N
±2%	BAP 48N BAP 16N BAP 12N Spherical B	BAP 12S Prandtl Spherical B
±3%	BAP 16S BAP 12S Type S Yaw-nulled DAT	Type S Straight-up
Beyond ±3%	French Man AP S Prandtl Type S Straight-up Modified Kiel Spherical A	Type S Yaw-nulled French Modified Kiel DAT Spherical A

[a] Results are based on four sets of Matrix A runs (at DeCordova, Lake Hubbard high load, Lake Hubbard low load, and Homer City) except as follows: The Spherical B and Autoprobe 12-point results are based on one set of Matrix A runs (at Homer City), Prandtl on two (at DeCordova and Lake Hubbard high load), and Autoprobe 48-point on three (at all except Homer City).

3.5 Uncertainty in Engineering Methods and Probe Measurements

A comparison of the expected uncertainty in the engineering and probe methods is presented in this section. This analysis is an important consideration for addressing the field study's collateral goal of seeking improvements that might reduce the reported disparities between the heat input values derived from in-stack and engineering methods. Knowing the uncertainties associated with in-stack and engineering methods reveals when the flow values from each method are sufficiently close that they fall within each other's overlapping error bands.

The analysis in this section is similar in intent to the uncertainty analysis presented in Section 2.3 and the variability analysis presented in Section 3.3. In this section, however, the uncertainty analysis was performed in a manner (i.e., establishing sampling distributions for input parameters and performing repeated trials) that enables the expected uncertainty in the volumetric flow values generated by an engineering method to be directly compared with the expected uncertainty in a probe's measurements on matched runs. This direct comparison enabled the relative uncertainty in the methods at both gas-fired and coal-fired plants to be evaluated. The evaluation was limited to the MMBtu engineering method and the Type S probe straight-up measurement. The results of the error analysis using these approaches are considered to be representative of the expected uncertainty in most of the engineering and probe-specific techniques. In addition, the MMBtu and Type S probe straight-up are used extensively in many field applications. The analysis was performed at the DeCordova and Homer City test sites. Monte Carlo techniques are used as a basis for the error analysis.

As noted earlier in Section 2.3, error analysis, like that presented here, provides a measure of the variability (i.e., uncertainty) inherent in a method, not its accuracy. Error analysis reveals the variation that can be expected in a method's results based on the variation in its input parameters, irrespective of whether the method is a good or poor indicator of the true flow value.

Monte Carlo simulation is a widely used technique for estimating the prediction error in values generated by complicated models or equations. For each uncertain parameter in an equation, a sampling distribution is generated. The sampling distribution represents the uncertainty in the measurement of each parameter. For this analysis, the sampling distributions were assumed to be normal. A distribution, including the sufficient statistics of the distribution (i.e., the sample mean and variance), is assigned to each uncertain parameter in the model. The Monte Carlo procedure first draws a random sample from each of the parameter distributions. The model is run for these values, and the resulting volumetric flow value is recorded. This procedure is repeated many times (in this analysis, 10,000 times) resulting in a distribution of the predicted values. From this predictive distribution, the error in volumetric flow can be calculated (e.g., the variance of the predictive distribution can be determined).

Because of the complexity of the MMBtu method, the first-order error analysis for this method presented in Section 2.3 used only first-level equations. In contrast, the Monte Carlo technique uses all equations in performing its analyses, thereby providing a full characterization of the measurement error. Because the Monte Carlo technique is much easier to implement for complicated models than first-order error analysis, it is particularly well suited for analysis of approaches like the MMBtu method.

The equations for the MMBtu and Type S straight-up method are presented in Appendix H.

Table H-1 in Appendix H presents the inputs to the MMBtu method for each test site. In addition, the source of information used to calculate the sampling mean and variance are presented, along with the CV for each parameter. Information used to generate the sampling distributions include the raw data collected during the field studies, information concerning the allowable error as specified in 40 CFR Part 60 for specific methods, and expert judgement. For methods specified in 40 CFR Part 60, the total error allowable was calculated as specified in the method.

Table H-2 in Appendix H includes the same information for the parameters of the Type S probe straight-up measurements. The sources of information used to derive the sampling variances include allowable error specified in 40 CFR Part 60, expert judgement, and sensitivity testing of the equation used to calculate volumetric flow.

Table 3-17 presents the results of the Monte Carlo analysis. The mean and standard deviation of the predictive distribution generated by the Monte Carlo method is presented for each method at the DeCordova and Homer City test sites. The CV is also presented.

Table 3-17. Monte Carlo Error Analysis Results on Volumetric Flow (wscfm)

Method	DeCordova			Homer City		
	Mean	s.d.	CV (%)	Mean	s.d.	CV (%)
MMBtu	1,539,302	17,875	1.2	1,430,653	74,724	3.1
Type S Straight-up	1,612,854	28,847	1.8	1,452,558	27,770	1.9

Examination of the table shows that the expected error in volumetric flow for the Type S probe ranges between a CV of 1.8% and 1.9% across the test sites. The expected error in the MMBtu method ranges, however, from a CV of 1.2% to 3.1%.

These results can be used to address the question of whether differences in flow measured by the Type S probe straight-up and flow calculated by the MMBtu method are statistically significant. Using the larger CV for each site, differences of less than 3.6% (1.8% times 2) between the two measured values were within two standard deviations of each other and so were not statistically distinguishable at the gas-fired site. Differences of less than 6.2% (3.1% times 2) were not statistically distinguishable at the coal-fired site.

SECTION 4
SUPPLEMENTAL DATA ANALYSIS

The analyses presented in this section were conducted in response to recommendations provided to EPA by a panel of peer reviewers. This section also presents results of comparisons among various pressure measurement devices. Section 4.1 presents a refinement of the central tendency analysis. Section 4.2 presents a comparison of probe measurements to an engineering baseline. Section 4.3 presents a sensitivity analysis of the effect of excluding certain data points that appeared implausible to peer reviewers, even though no documented problems could be found to justify their exclusion based on the criteria established for identifying outliers. Section 4.4 presents a comparison of pressure data collected by manual and electronic devices at two wind tunnels and several field test sites.

4.1 REFINEMENT OF THE CENTRAL TENDENCY ANALYSIS

The concerns expressed by the peer reviewers regarding the Round 1 central tendency analysis presented in Section 3 are summarized below.

Calculating the central tendency, or the grand mean, of the volumetric flow values using the measurements obtained by every probe type used in the study may not be appropriate for the following reasons:

- As shown in Section 3, some probe types used in the study (the modified Kiel and French, for example) produced uniformly high or highly variable results. Values obtained by such probes should not be included in the calculation of the grand mean.
- Some probe types and methods were over-represented in the Round 1 grand mean calculations. For example, there were four Type S probes (manual Type S, manual Autoprobe Type S, baseline Autoprobes 16-point, and baseline Autoprobes 48-point), each of which was operated in both straight-up and yaw-nulled modes, and five engineering methods (O_2 F-factor, CO_2 F-factor, ASME PTC 4.1, MMBtu, and BTCE). Because probes or methods of the same type tend to produce similar flow measurements, a grand mean value that makes use of all probe types and methods will be overly influenced by the results of the over-represented probe types or methods, and therefore may not be a good estimate of the true flow in the stack.
- Accounting for the yaw component of flow generally results in more accurate measurements. Probes that do not account for yaw angles generally give higher readings than those that do and should not be included in calculations of the grand mean.

The decay of flue gas velocity near the stack wall is a widely accepted phenomenon and is observed for the wall-effects runs in this study. Because this decay is not adequately captured by a standard Method 1 sampling traverse, the volumetric flow values used in the Round 1 grand mean calculations do not account for wall effects and are therefore high. Adjusting the flow values obtained by all probe types for wall effects should result in a more accurate grand mean.

Based on these concerns and suggestions and the results from the Round 1 analysis presented in Section 3, a second central tendency analysis was performed, as described below. The intent of the Round 2 analysis was to implement peer reviewers' recommendations that the grand mean be

calculated only from the methods they considered to be most technically credible and that the results be evaluated to see how close these methods come to both the grand mean and each other.

The refined grand mean was calculated using flow measurements from the following four probe types and two engineering methods: manual Type S yaw-nulled, DAT, spherical (Run 3 at Lake Hubbard was excluded because it seemed implausible to several peer reviewers), baseline Autoprobes 16-point yaw-nulled, MMBtu, and BTCE. For DeCordova and Lake Hubbard, the Autoprobes 16-point yaw-nulled, DAT probes, MMBtu, and BTCE each contributed eight flow values to the grand mean, while the Type S yaw-nulled and spherical probes each contributed four flow values. For Homer City, each probe type and method contributed four values to the grand mean, except for the spherical probes, which contributed eight.

Flow measurements obtained by the four probe types were adjusted to account for wall effects prior to calculation of the grand mean. The wall effects adjusted volumetric flow values were calculated as follows:

$$F_{adj} = F_{unadj} \times (1 + A/100) \qquad \textbf{Eq. 4-1}$$

where:

F_{adj} = volumetric flow adjusted for wall effects;
F_{unadj} = volumetric flow unadjusted for wall effects (from Method 1 points); and
A = wall effects adjustment factor.

The wall effects adjustment factors, A, were derived using data collected during wall-effects tests at each of the three test sites. They are expressed in percent and bear a negative sign, for example, -1.57% for the DAT probe at DeCordova. The factors for each probe type are presented in Table I1-1 in Appendix I1-1.

Although not included in the Round 2 grand mean calculations, the Type S straight-up and Prandtl, which are the current allowable Method 2 probe types, are included in the tables and graphs presented in this section for comparison purposes.

Of the five engineering methods performed during the field tests (see Section 3), the MMBtu method was included in the Round 2 analysis, because the peer review panel indicated that it was the most widely accepted and commonly used of the four stoichiometry-based methods (O_2 F-factor, CO_2 F-factor, MMBtu, and ASME PTC 4.1). The BTCE method was included because it is based on conservation of energy rather than stoichiometry.

The results of the Round 2 central tendency analysis are summarized in Table 4-1. Following the convention used earlier in Table 3-12 and explained in Section 3.1.4 in this report, Table 4-1 uses a bi-directional ranking of methods, where the column labeled "rank" gives the highest positive rank to the method with the greatest positive deviation from the grand mean and the lowest negative rank to the method with the greatest negative deviation from the grand mean.

Table 4-1. Summary Statistics for Round 2 Central Tendency Analysis (Matrix A)

Probe/Method	Vol. Flow Grand Mean (wscfm)	Comparison to Grand Mean			
		Mean Diff.	s.d.	Percent Diff.	Rank
DeCordova					
Type S Straight-up[a]	1,542,922.4	38,813.9	29,722.4	2.52%	3
Prandtl[a]		54,428.1	9,986.7	3.53%	4
Type S Yaw-nulled		28,099.1	41,003.3	1.82%	2
DAT		-3,854.3	40,959.9	-0.25%	-2
Spherical		27,815.6	14,073.4	1.80%	1
BL AP 16-point Yaw-nulled		-3,568.3	4,984.1	-0.23%	-1
MMBtu		-9,517.0	3,956.0	-0.62%	-4
BTCE		-7,540.9	8,787.0	-0.49%	-3
Lake Hubbard (High Load, Without Spherical Run 3)					
Type S Straight-up[a]	1,073,469.6	61,044.1	26,192.1	5.69%	4
Prandtl[a]		22,307.9	13,863.3	2.08%	3
Type S Yaw-nulled		18,759.4	35,249.5	1.75%	2
DAT		-11,181.6	32,108.5	-1.04%	-3
Spherical		-37,086.3	13,984.5	-3.45%	-4
BL AP 16-point Yaw-nulled		-2,139.4	4,534.3	-0.20%	-2
MMBtu		18,372.2	4,106.0	1.71%	1
BTCE		-523.5	3,182.6	-0.05%	-1
Lake Hubbard (Low Load)					
Type S Straight-up[a]	693,929.4	28,814.1	7,549.9	4.15%	4
Prandtl[a]					
Type S Yaw-nulled		21,520.4	12,596.3	3.10%	3
DAT		-8,648.6	21,546.9	-1.29%	-1
Spherical		-65,256.1	46,432.3	-9.40%	-2
BL AP 16-point Yaw-nulled		4,213.1	4,364.4	0.61%	1
MMBtu		29,914.1	4,623.7	4.31%	5
BTCE		18,293.1	15,811.3	2.64%	2
Homer City					
Type S Straight-up[a]	1,379,855.1	105,168.4	23,523.2	7.62%	4
Type S Yaw-nulled		38,838.4	45,170.5	2.81%	3
DAT		663.3	28,651.4	0.05%	1
Spherical A		-13,069.1	12,662.5	-0.95%	-4
Spherical B		-11,542.6	19,446.4	-0.84%	-2
BL AP 16-point Yaw-nulled		-7,360.1	13,583.7	-0.53%	-1
MMBtu		12,630.6	27,188.9	0.92%	2
BTCE		-13,047.1	13,396.1	-0.95%	-3

[a] Not included in the grand mean calculation.

The following observations can be made from examination of Table 4-1:

- The Round 2 analysis grand mean for each site was smaller than the Round 1 analysis grand mean [0.84%, 1.49%, 2.58%, and 2.46% smaller for DeCordova, Lake Hubbard (high load), Lake Hubbard (low load), and Homer City, respectively].

- At all sites, except Lake Hubbard at low load, all the methods included in the grand mean produced flow values that were close to the grand mean. At DeCordova, the flow values from all methods included in the grand mean were themselves within 2% of the grand mean. At Lake Hubbard (high load) and Homer City, only one method that was included in the grand mean produced values that were more than 2% from the grand mean—the spherical probes (3.45%) at Lake Hubbard (high load) and the Type S yaw-nulled probes (2.81%) at Homer City. In contrast, under the low load conditions at Lake Hubbard, four of the six methods included in the grand mean were themselves more than 2% from the grand mean.

- Of the in-stack methods, the Autoprobes 16-point yaw-nulled produced flow values that were always within 0.61% of the grand mean, and the DAT probes produced values that were always within 1.29% of the grand mean.

- The flow values for the Type S yaw-nulled probe were higher than the grand mean at all three sites, ranging from 1.75% at Lake Hubbard (high load) to 3.10% at Lake Hubbard (low load).

4.2 Comparison to Engineering Baseline

Some peer reviewers suggested that, at the gas-fired facilities, a mass balance calculation would yield the most accurate representation of flue-gas flow rates. These reviewers therefore recommended that the data analysis include comparisons of the in-stack probe measurements to an engineering baseline, rather than solely to the baseline Autoprobes 16-point straight-up, as presented in Section 3. Such a comparison had already appeared in Appendix A of the individual site data reports for the DeCordova, Lake Hubbard, and Homer City field tests, where Series 13 and 14 tables and figures used the BTCE and MMBtu methods, respectively, as baselines. Further comparisons are presented here in response to the reviewers' recommendations.

Table 4-2 compares flow values for the methods analyzed in Section 4.1 to the values obtained by the MMBtu method using Matrix A data only. The MMBtu method was selected as a baseline for this comparison because the peer review panel thought it to be generally accepted and commonly used by industry.

As in Table 4-1, Table 4-2 compares flow values for the methods selected for Round 2 analyses after the velocity decline near the stack wall has been taken into account. It should be noted that the Lake Hubbard high-load data exclude spherical Run 3, as requested by several peer reviewers. Table 4-2 presents the mean difference, standard deviation, and percent difference between each method and the average MMBtu value at each site.

Table 4-2. Summary Statistics for Round 2 Analysis Using MMBtu Method as the Baseline (Matrix A)

Probe/Method	Average MMBtu Vol. Flow (wscfm)	Comparison to MMBtu		
		Mean Diff.	s.d.	Percent Diff.
DeCordova				
Type S Straight-Up	1,533,405.4	48,330.9	29,722.4	3.15%
Prandtl		63,945.1	9,986.7	4.17%
Type S Yaw-Nulled		37,616.1	41,003.3	2.45%
DAT		5,662.8	40,959.9	0.37%
Spherical		37,332.6	14,073.4	2.43%
BL AP 16-point Yaw-Nulled		5,948.8	4,984.1	0.39%
MMBtu		*0.0*	*3,956.0*	*0.00%*
BTCE		1,976.1	8,787.0	0.13%
Lake Hubbard (High Load, Without Spherical Run 3)				
Type S Straight-Up	1,091,841.9	42,671.9	26,192.1	3.91%
Prandtl		3,935.6	13,863.3	0.36%
Type S Yaw-Nulled		387.1	35,249.5	0.04%
DAT		-29,553.9	32,108.5	-2.71%
Spherical		-55,458.5	13,984.5	-5.08%
BL AP 16-point Yaw-Nulled		-20,511.6	4,534.3	-1.88%
MMBtu		*0.0*	*4,106.0*	*0.00%*
BTCE		-18,895.8	3,182.6	-1.73%
Lake Hubbard (Low Load)				
Type S Straight-Up	723,843.5	-1,100.0	7,549.9	-0.15%
Prandtl		NA	NA	NA
Type S Yaw-Nulled		-8,393.8	12,596.3	-1.16%
DAT		-38,598.8	21,546.9	-5.33%
Spherical		-95,170.3	46,432.3	-13.15%
BL AP 16-point Yaw-Nulled		-25,701.0	4,364.4	-3.55%
MMBtu		0.0	4,623.7	0.00%
BTCE		-11,621.0	15,811.3	-1.61%
Homer City				
Type S Straight-Up	1,392,485.8	92,537.8	23,523.2	6.65%
Type S Yaw-Nulled		26,207.8	45,170.5	1.88%
DAT		-11,967.4	28,651.4	-0.86%
Spherical A		-25,699.8	12,662.5	-1.85%
Spherical B		-24,173.3	19,446.4	-1.74%
BL AP 16-point Yaw-Nulled		-19,990.8	13,583.7	-1.44%
MMBtu		*0.0*	*27,188.9*	*0.00%*
BTCE		-25,677.8	13,396.1	-1.84%

With respect to the in-stack methods, the primary observation that can be made from Table 4-2 is the closeness to the MMBtu values of the measurements by the Autoprobes 16-point yaw-nulled (0.39%, -1.88%, -3.55%, and -1.44%), the Type S yaw-nulled (2.45%, 0.04%, -1.16%, and 1.88%), and the DAT probes (0.37%, -2.71%, -5.33%, and -0.86%). This observation suggests that provision for the use of these two types of probes might address one of the collateral goals of the field study, that is, to reduce the disparity that industry had reported between in-stack and combustion-based calculations of heat rate and flow.

4.3 Sensitivity Analysis on Lake Hubbard Spherical Probe Data

As described in Section 2, the criterion used in this field study for identifying values as outliers and excluding them from the analytical data set was the existence of documented procedural or equipment problems. Any value that differed substantially from other values in the same or proximal runs and any value that exceeded the 1.5-IQR statistical screening criterion were flagged and investigated, but only those for which procedural or equipment problems could be documented were excluded from analysis as outliers.

A number of peer reviewers questioned the plausibility of several Lake Hubbard spherical probe values that remained in the analytical data set after application of the criterion described above. In particular, they questioned the values obtained by probe SPH-4 on Run 3, by probe SPH-2 on Run 22, and by probe SPH-4 on Run 24. (Individual probe values are presented by run in Tables 3A through 3E in Appendix A of the Lake Hubbard Site Data Report.)

Based on these comments, the three identified values were re-examined. No new evidence of procedural or equipment problems was found. However, whereas the values obtained on Runs 22 and 24 were consistent with proximal values obtained by other probes, the value obtained on Run 3 clearly was not. Of the three questioned values, the Run 3 value was also the only one that was implausible in view of the unit load at the time of the test.

A sensitivity analysis was therefore performed to evaluate the impact on the analysis presented in Section 3 of removing the Run 3 value for probe SPH-4 from the analytical data set. The results are presented in Tables 4-3 through 4-8. The most noticeable impact of removal of the SPH-4, Run 3 value was observed in the following results:

- In the central tendency analysis, removal of the SPH-4, Run 3 value brought the spherical probe closer to the grand mean, from -12.3% to -4.0%. By changing the grand mean itself, the removal also brought the modified Kiel, Type S straight-up and yaw-nulled, and Prandtl probes closer to the central tendency of the data and caused the DAT and French to depart somewhat further from the central tendency (Table 4-3).

- The CV for the spherical probes in the Lake Hubbard high-load runs dropped from 19.70% to 1.35%. The CV in the low-load runs at Lake Hubbard (7.39%) was unaffected by the removal of the Run 3 data (Table 4-4).

Table 4-3. Sensitivity Analysis on Round 1 Central Tendency Analysis for Matrix A Runs—Lake Hubbard (High Load) (Based on Table 3-12)[a]

With SPH-4, Run 3				Without SPH-4, Run 3			
Grand Mean = 1,089,755.58 wscfm				Grand Mean = 1,093,563.40 wscfm			
		Difference				Difference	
Rank	Method	wscfm	%	Rank	Method	wscfm	%
11	**Modified Kiel**	**79,070.4**	**7.3%**	**8**	**Modified Kiel**	**75,262.6**	**6.9%**
10	**Type S Straight-up**	**65,906.9**	**6.0%**	7	**Type S Straight-up**	**62,099.1**	**5.7%**
9	*Type S Yaw-nulled*	*23,968.2*	*2.2%*	*6*	Type S Yaw-nulled	20,160.4	1.8%
8	*Prandtl*	*23,726.2*	*2.2%*	*5*	Prandtl	19,918.4	1.8%
7	Manual AP Straight-up	14,430.3	1.3%	4	Manual AP Straight-up	10,622.5	1.0%
6	ASME PTC 4.1	13,568.9	1.2%	3	ASME PTC 4.1	9,761.1	0.9%
5	BL AP 16-point Straight-up	12,650.9	1.2%	2	BL AP 16-point Straight-up	8,843.1	0.8%
4	CO_2 F-factor	6,434.3	0.6%	1	CO_2 F-factor	2,626.5	0.2%
3	Manual AP Yaw-nulled	3,227.7	0.3%	-1	Manual AP Yaw-nulled	-580.1	-0.1%
2	MMBtu	2,086.3	0.2%	-2	MMBtu	-1,721.5	-0.2%
1	BL AP 48-point Straight-up	1,988.5	0.2%	-3	BL AP 48-point Straight-up	-1,819.3	-0.2%
-1	BL AP 16-point Yaw-nulled	-1,226.4	-0.1%	-4	BL AP 16-point Yaw-nulled	-5,034.2	-0.5%
-2	DAT	-10,633.5	-1.0%	-5	DAT	-14,441.3	-1.3%
-3	BL AP 48-point Yaw-nulled	-10,777.2	-1.0%	-6	BL AP 48-point Yaw-nulled	-14,585.0	-1.3%
-4	O_2 F-factor	-15,322.1	-1.4%	-7	O_2 F-factor	-19,129.9	-1.7%
-5	BTCE	-16,809.5	-1.5%	-8	BTCE	-20,617.3	-1.9%
-6	*French*	*-22,242.1*	*-2.0%*	*-9*	*French*	*-26,049.9*	*-2.4%*
-7	**Spherical A**	**-133,505.6**	**-12.3 %**	**-10**	**Spherical A**	**-43,316.4**	**-4.0%**

[a] **bold = ±3.0% of grand mean and beyond,** *italics =from ±2.0% up to ±3.0% of grand mean*; normal = within ±2.0% of grand mean.

Table 4-4. Sensitivity Analysis on Method Variability Analysis For Matrix A Runs (Based on Table 3-15)

Method	CV[a] (%)			
	DeCordova	Lake Hubbard		Homer City
		High Load	Low Load	
DAT	2.66	3.02	3.14	2.08
Type S Straight-up	1.88	2.31	1.04	1.58
Type S Yaw-nulled	2.61	3.23	1.76	3.18
Modified Kiel	1.39	5.19	4.04	1.24
Spherical A	**0.90**	**19.70**	**7.39**	**0.93**
Spherical A w/o SPH-4, Run 3 at LH		**1.35**		
Spherical B	NA	NA	NA	1.42
Prandtl	0.63	1.27	NA	NA
French	1.62	3.28	NA	4.18
Manual Autoprobe 16-point Straight-up	0.34	0.43	NA	0.90
Manual Autoprobe 16-point Yaw-nulled	0.10	0.25	NA	0.93
Baseline Autoprobes 16-point Straight-up	0.27	0.37	0.91	0.98
Baseline Autoprobes 16-point Yaw-nulled	0.32	0.42	0.63	0.99
Baseline Autoprobes 48-point Straight-up	0.24	0.21	0.31	NA
Baseline Autoprobes 48-point Yaw-nulled	0.32	0.26	0.28	NA
Baseline Autoprobes 12-point Straight-up	NA	NA	NA	1.23
Baseline Autoprobes 12-point Yaw-nulled	NA	NA	NA	0.95
O_2 F-Factor	0.48	0.67	0.59	1.76
CO_2 F-Factor	1.25	0.69	0.76	1.90
MMBtu	0.26	0.38	0.64	1.95
ASME PTC 4.1	1.12	0.41	0.43	2.02
BTCE	0.57	0.30	2.22	0.98

[a] CV = coefficient of variation, (standard deviation / mean) × 100.

The removal had limited, but discernible, impact on the following results:

- The mean percent difference of the spherical probes from the baseline Autoprobes 16-point straight-up in Matrices A and B changed from -7.17% to -5.86% (Table 4-5).

Table 4-5. Sensitivity Analysis on Volumetric Flow Summary—Lake Hubbard (Based on Table 3-7)

Probe/Method	Number of Tests	% Difference from Baseline[a]			s.d.
		Mean	Minimum	Maximum	
DAT	29	-2.98%	-13.44%	0.54%	3.10%
Type S Straight-up	25	3.65%	0.98%	7.92%	1.83%
Type S Yaw-nulled	25	1.72%	-3.70%	4.93%	2.16%
Modified Kiel	8	3.11%	-4.24%	12.04%	4.71%
Spherical	**25**	**-7.17%**	**-38.73%**	**-1.22%**	**7.84%**
Spherical w/o SPH-4 for Run 3	**24**	**-5.86%**	**-17.81%**	**-1.22%**	**4.37%**
Prandtl	4	0.93%	-0.30%	3.12%	1.52%
French	4	-3.24%	-7.37%	-0.16%	3.16%
Manual Autoprobe 16-point Straight-up	4	0.09%	-0.71%	0.97%	0.75%
Manual Autoprobe 16-point Yaw-nulled	4	-0.93%	-1.58%	-0.29%	0.52%
Baseline Autoprobes 16-point Straight-up[a]	*25*	*0.00%*	*0.00%*	*0.00%*	*0.00%*
Baseline Autoprobes 16-point Yaw-nulled	25	-1.08%	-2.93%	0.47%	0.66%
Baseline Autoprobes 48-point Straight-up	24	-0.99%	-1.48%	-0.18%	0.36%
Baseline Autoprobes 48-point Yaw-nulled	24	-2.06%	-3.13%	-0.88%	0.51%
O_2 F-Factor	25	-1.75%	-3.60%	0.78%	1.12%
CO_2 F-Factor	25	-0.02%	-1.78%	2.69%	0.96%
MMBtu	25	-0.50%	-1.99%	1.21%	0.78%
ASME PTC 4.1	25	0.49%	-0.81%	2.37%	0.76%
BTCE	25	-2.31%	-3.31%	1.60%	0.99%

[a] Baseline is baseline Autoprobes 16-point straight-up.

- The analysis of variance changed slightly. For the original ANOVA Variant 1, the engineering methods were not distinguishable from the 1-D and 2-D probes, but all were different from the 3-D probes. Without Run 3, the engineering methods and 3-D probes could not be distinguished, but both differed from the 1-D and 2-D probes (Table 4-6).

Table 4-6. Sensitivity Analysis on Duncan's Multiple Range Test—Lake Hubbard (High Load) (Based on Table 3-11)

Variant	With SPH-4, Run 3	Without SPH-4, Run 3
Variant 1[a]	(1D, 2D, Eng) ≠ 3D	(1D, 2D) ≠ (Eng, F, 3D)
Variant 2[b]	$(2D_m, 1D_m, 1D_a) \neq 3D_m$	$2D_m, 1D_m, 1D_a \neq 3D_m$ $(2D_m, 1D_m) \neq (2D_a, F, 3D_m)$ $2D_m \neq 1D_a, 2D_a, F, 3D_m$

[a] Variant 1 Classes
1D = Type S straight-up, baseline Autoprobes straight-up, manual Autoprobe straight-up, Prandtl
2D = Type S yaw-nulled, baseline Autoprobes yaw-nulled, manual Autoprobe yaw-nulled, modified Kiel
3D = DAT, spherical
F = French
Eng = Engineering methods (O_2 F-factor, CO_2 F-factor, MMBtu, ASME PTC 4.1, BTCE)

[b] Variant 2 Classes
$1D_m$ = manual 1-D methods (Type S straight-up and Prandtl)
$2D_m$ = manual 2-D methods (Type S yaw-nulled and modified Kiel)
$3D_m$ = 3-D methods (DAT and spherical)
$1D_a$ = Autoprobes 1-D methods (baseline Autoprobes straight-up)
$2D_a$ = Autoprobes 2-D methods (baseline Autoprobes yaw-nulled)
F = French

The following results were largely unchanged by the removal of the SPH-4, Run 3 value:

- overall range in all flow values (Table 4-7), and
- in-stack average vs. engineering average flow values (Table 4-8).

Table 4-7. Sensitivity Analysis on Overall Range of Flow Measurements—Lake Hubbard (High Load) (Based on Table 3-1)

	With SPH-4, Run 3		Without SPH-4, Run 3	
	Avg. Value	Method	Avg. Value	Method
Max (wscfm)	1,157,629	Modified Kiel[a]	1,157,629	Modified Kiel[a]
Min (wscfm)	1,036,472	Spherical	1,054,582	Spherical
Difference	121,157 (11.02%)[b]		103,047 (9.77%)[b]	

[a] The modified Kiel probe produced the highest measurement of flow when measurements taken only under high-load operation were included. When low-load measurements were also included, the Type S probe straight-up gave the highest average measured flow. See Tables 3-6 through 3-8.

[b] Ratio of difference relative to average of all in-stack and engineering methods.

Table 4-8. Sensitivity Analysis on In-stack and Engineering Methods Comparison—Lake Hubbard (High Load) (Based on Table 3-4)[a]

Comparison	With SPH-4 for Run 3	Without SPH-4 for Run 3
In-stack Avg. vs. Engineering Avg.	0.26% (20)	0.46% (20)
S Straight-up Avg. vs. Engineering Avg.	5.44% (8)	5.44% (8)

[a] Number of runs analyzed is shown in parentheses.

4.4 Comparison of Manual and Electronic Pressure Measuring Devices

Some peer reviewers expressed the opinion that manual pressure devices (i.e., the manometer and magnehelic gauges) provide consistently higher pressure readings than electronic transducers. To investigate this assertion, a series of comparisons was made of pressure readings taken with manual and electronic devices during test programs conducted as part of this study and by others. Data from the following tests were used: an EPA-sponsored swirl tunnel test conducted by Fossil Energy Research Corp. (FERCo); an EPRI-sponsored field test at the Coal Creek and Columbia power plants conducted by FERCo; an EPA-sponsored field test conducted by Radian (now known as the Energy Research Group); and a NIST wind tunnel test during which the probes used in this study were calibrated. In both the wind tunnel tests and the field tests, manual and electronic pressure devices were used in parallel to measure pressures.

Analysis of the five data sets produced the following findings:

- For different types of probes and different velocity conditions, no one type of pressure device consistently produced higher or lower measurements.

- Even where statistically significant differences were found between the manual and electronic devices, the differences in pressure readings were, from a practical standpoint, generally small.

- With Type S probes, manual devices tended to read higher than electronic transducers in the field tests. In the wind tunnel tests, the manual devices tended to provide lower pressure readings.

- The percent differences between the manual and the electronic devices were generally smaller on average and less variable in the wind tunnel tests than in the field tests.

- With electronic transducers, very little difference was found between taking a straight average of the individual differential pressure measurements (Approach 1) and squaring the average of the square roots of the individual pressure measurements (Approach 2). For the data examined, the average differential pressure using Approach 1 was always higher than the average differential pressure using Approach 2, but never by more than 0.44%.

A detailed description of the devices used, the type of data collected, and the analyses performed, and a fuller discussion of the findings appear in Appendix I2.

SECTION 5
WALL EFFECTS STUDY

A key element of EPA's evaluation of potential improvements to the Agency's current test methods for volumetric flow was the study of velocity decay near the stack wall, referred to as the "wall effect." The primary objective of the wall effects study was to investigate the impact of velocity decay close to the stack wall on volumetric flow. Collateral objectives were to (1) evaluate a procedure for adjusting for velocity decay close to the stack wall using actual field data and (2) obtain information on practical aspects of implementing the procedure under conditions typically encountered at electric utility stacks.

Wall effects tests were conducted at each of the three steel-stack field test sites during the day under steady load conditions (Matrix C runs) using manual probes and Autoprobes, and at night under nonsteady-state load conditions using only Autoprobes. Additional wall effects tests using only Autoprobes were conducted at six sites with brick and mortar stacks in order to investigate the effect on volumetric flow of velocity decay for rough stack surfaces. All of the over-night tests and some of the day-time tests at these additional sites were conducted under nonsteady-state load conditions. Because of the importance of flow stability in evaluating changes in velocity before and after adjustment for velocity decay close to the stack wall, all wall effects data collected under nonsteady-state conditions were examined for velocity stability by the procedure described in Exhibit J-1 in Appendix J before being used in the study. Data sets where the velocity at the Method 1 equal-area traverse points closest to the wall in the original (baseline) traverse differed by more than 8% from the velocity at the corresponding traverse point in the wall effects traverse were excluded from the analyses reported in this section. In addition, due to design features of some probes and/or physical constraints at some testing locations, measurements could not be made beginning at 1 in. from the stack wall for some probes at some field test sites. Data sets that did not contain a complete set of measurements beginning at 1 in. from the stack wall also were excluded from the analyses reported in this section. Table J-1 in Appendix J provides a complete summary of all wall effects data sets collected in the study that met the velocity stability criterion.

5.1 WALL EFFECTS DATA COLLECTION PROCEDURE

In the wall effects tests, a standard 12-, 16-, or 20-point cross-stack traverse was first performed as prescribed in Method 1. The velocity was measured at the centroid of each equal-area sector and the average cross-stack velocity was calculated from the resulting data for these baseline traverses.

Next, a wall effects traverse was performed in which velocity measurements were taken at 1-in. increments across the entire width of each of the four Method 1 equal-area sectors adjacent to the stack wall, usually starting at 1 in. from the wall and moving toward the center of the stack. As noted above, for some types of probes and for some sampling ports, obtaining measurements within a few inches of the stack wall was not possible due to the size and shape of the probe, because the sampling port nipple extended beyond the interior stack wall into the gas stream, or because the annular space in the stack did not allow moving the probe to within 1 in. of the stack wall. In such instances, the wall effects sampling procedure allows the first 1-in. incremented velocity measurement to be made at a distance of greater than 1 in. from the stack wall. In calculating the replacement velocity value (see following paragraph) in these cases, the procedure stipulates that the first actual measurement made in a sampling port should be used for all sampling points closer to

the stack wall where measurements cannot be made. The data in Table J-1 in Appendix J include some tests where measurements could not be made beginning at 1 in. from the stack wall in all ports. For these tests, the replacement velocity values were calculated in accordance with this provision of the procedure. However, only data sets where measurements could be made beginning at 1 in. from the stack wall were used in the analyses reported in this section.

The wall effects traverse measurements were used to calculate a replacement velocity value for the Method 1 traverse point closest to the stack wall for each of the four sampling ports. The replacement velocity values were calculated as described in Section 5.2 below. It should be noted that although measurements were taken across the entire width of each of the four Method 1 equal-area sectors adjacent to the stack wall in the field study in order to gather information about the decay in velocity close to the stack wall, the procedure followed in the study ("Method 2 Draft Revisions: Wall Effects Velocity Procedure"[7]) could easily be adapted to provide for sampling as few as two points. The calculational procedure described in Section 5.2 is much simpler when implemented for only a few points.

The average cross-stack velocity was then recalculated after replacing the measured velocity values for the four Method 1 traverse points closest to the stack wall with the calculated replacement velocity values. The percent difference between the original average cross-stack velocity and the average cross-stack velocity making use of replacement velocity values was then calculated. The calculated percent difference indicates the effect on volumetric flow due to velocity decay close to the stack wall.

5.2 Calculation of Replacement Velocity Values Using Wall Effects Data

Replacement velocity values for 12- and 16-point traverses were calculated in accordance with the procedure presented in Form 5-1A (or 5-1B), "Method 2 Draft Revisions: Wall Effects Velocity Procedure." This procedure was adapted for calculating replacement velocity values for the 20-point traverses that are reported in Table J-1 in Appendix J. Exhibit J-2 in Appendix J provides an example of a replacement velocity calculation for a 16-point traverse.

5.3 Summary of Wall Effects Data Obtained for the Study

Wall effects tests were conducted at each of the three field test sites (DeCordova, Lake Hubbard, and Homer City) using manual probes and Autoprobes. Because the three field test sites all have smooth steel stacks, additional wall effects tests were conducted at the six sites with brick and mortar stacks listed below in order to investigate the effect on volumetric flow of velocity decay for rough stack surfaces. The additional wall effects tests at these sites were conducted using only Autoprobes and were not observed by EPA or Cadmus personnel. While not subject to the same degree of on-site scrutiny as the tests at DeCordova, Lake Hubbard, and Homer City, the tests at these sites were conducted by the same test personnel, using the same or identical test equipment, following the same test protocol as was used at the three primary field tests sites. For these reasons, it was decided to include the data in the study.

---

[7] See, for example, The Cadmus Group, Inc., 1997, "Flow Reference Method Testing and Analysis: Field Test Plan, DeCordova Electric Station," EPA/430/R-97-024.

Sites for Additional Wall Effects Tests

- American Electric Power Co., Inc.
  Columbus Southern Power
  Conesville Units 1/2, 3, and 5
- American Electric Power Co., Inc.
  Columbus Southern Power
  Picway Unit 9
- General Public Utilities Corp.
  Metropolitan Edison Co.
  Titus Unit 1
- Allegheny Power System, Inc.
  West Penn Power Co.
  Mitchell Unit 3

Table 5-1 shows key characteristics of the sites for the additional wall effects tests. See Table 1-2 for corresponding key characteristics of DeCordova, Lake Hubbard, and Homer City.

Table 5-1. Key Characteristics of Sites for Additional Wall Effects Tests

Station	**Conesville**			
Location	Conesville, OH			
Stack No.	1/2		3	5
Stack Inside Diameter	17.2 ft		19.0 ft.	34.0 ft.
Test Dates	5/16/98		3/25/98	1/15-16/98
Unit	1	2	3	5
Fuel	Bituminous Coal	Bituminous Coal	Bituminous Coal	Bituminous Coal
Boiler Type	B&W 1.000 x 10^6 lb/hr cyclone	B&W 1.000 x 10^6 lb/hr cyclone	Riley Stoker 1.100 x 10^6 lb/hr front-fired	CE 3.131 x 10^6 lb/hr tangential-fired
Nameplate Capacity	148 MWe	136 MWe	162 MWe	444 MWe
Load Range During Tests[a]	79-84 MWe	79-88 MWe	75-77 MWe	Day: 379-433 MWe Night: 218-433 MWe

Station	**Picway**	**Mitchell**	**Titus**
Location	Lockbourne, OH	Courtney, PA	Reading, PA
Stack No.	9	3	1
Stack Inside Diameter	14.5 ft	22.0 ft	13.9 ft
Test Dates	1/13/98	12/17/97	2/4-5/98
Unit	9	3	1
Fuel	Bituminous Coal	Bituminous Coal	Bituminous Coal
Boiler Type	Riley Stoker 0.900 x 10^6 lb/hr front-fired	CE 1.978 x 10^6 lb/hr tangential-fired	CE 0.570 x 10^6 lb/hr tangential-fired
Nameplate Capacity	106 MWe	299 MWe	75 MWe
Load Range During Tests[a]	82-84 MWe	Day: 16-point: 112-117 MWe 12-point: 212-217 MWe Night: 204-267 MWe	Day: 66-80 MWe Night: 55-80 MWe

[a] Load range is across all tests and does not necessarily indicate the range for any single test. Refer to Exhibit J-1 in Appendix J for an explanation of the velocity stability criterion that applies for individual tests.

The data from the wall effects tests at DeCordova, Lake Hubbard, Conesville 1/2, Picway, and Mitchell that met the completeness and velocity stability criteria, discussed earlier, are reported in this section. The following paragraphs provide summary descriptions of the wall effects tests conducted at each site reported in this section.

DeCordova. Seventeen wall effects tests were conducted at DeCordova during the day, seven using manual probes and ten using Autoprobes. Fourteen additional wall effects tests were conducted at DeCordova at night using Autoprobes. The data from all of these tests were used in developing the findings reported in this section, except for one test where modified Kiel probes were used and one test where spherical and DAT probes were used together (with DAT probes used in two ports and spherical probes in the other two). The data from the modified Kiel probe test were excluded because measurements could not be taken beginning at 1 in. from the stack wall. The data from the spherical/DAT probe test were excluded from analysis in this section because the results could not be ascribed to a single type of probe. The spherical probe data (obtained at two of the four ports during the test) were used in deriving the wall effects adjustment factor shown in Table 4-1 in Section 4.

Lake Hubbard. At Lake Hubbard, three wall effects tests were conducted during the day: one using DAT probes and one each using Autoprobes operated in the straight-up and yaw-nulled modes. Eighteen additional wall effects tests were conducted at Lake Hubbard at night using Autoprobes. The data from all of these tests were used in developing the findings reported in this section.

Conesville Unit 1 and 2 Stack. Twelve wall effects tests were conducted at the Conesville Unit 1 and 2 shared stack during the day using Autoprobes. The data from all of these tests were used in developing the findings reported in this section.

Picway. Twelve wall effects tests were conducted at Picway during the day using Autoprobes. The data from all of these tests were used in developing the findings reported in this section.

Mitchell. Sixteen wall effects tests were conducted at Mitchell during the day using Autoprobes. Six tests (three 12-point straight-up and three 12-point yaw-nulled) did not meet the velocity stability criterion described in Appendix J, and results for these tests were not included in Table J-1 in Appendix J or used in developing the findings reported in this section.

Other Test Sites. No data sets from wall effects tests conducted at Homer City, Conesville Unit 3, Conesville Unit 5, or Titus contained a complete set of measurements beginning at 1 in. from the wall for all ports. Thus, the data from these sites were excluded from the analyses reported in this section.

Exhibits J-3 through J-11 in Appendix J provide additional information on the wall effects tests conducted at each site, including descriptions of the nature and limitations of the data collected at the sites not reported in this section. The Series 7 tables in Appendix A of the DeCordova, Lake Hubbard, and Homer City site data reports contain the data listed below for all the day-time tests at these sites. Tables O1 through O3 in Appendix O of the site data reports contain analogous data for the over-night tests at DeCordova, Lake Hubbard, and Homer City. Tables J-2(A-C) through J-7(A-C) in Appendix J of this report contain these data for tests at the other sites. These tables contain the following data: (1) velocity values for wall effects traverse points; (2) original (unadjusted) and wall effects-adjusted velocity values for Method 1 points; and (3) percent differences between the original (unadjusted) and wall effects-adjusted average cross-stack velocity values.

Wall effects calculation forms for all tests (day-time and over-night) at DeCordova, Lake Hubbard, and Homer City are located in Appendix P of the DeCordova and Lake Hubbard site data reports and Appendix R of the Homer City site data report. The wall effects calculation forms for tests at the other sites are located in Attachments A through F of Appendix J of this report.

5.4 Analysis of Wall Effects Data

Table 5-2 summarizes the wall effects data for tests where measurements could be made beginning at 1 in. from the stack wall and that met the previously described velocity stability criterion. Unless otherwise indicated, the results in Table 5-2 are based on 16-point baseline traverses. In columns 4 through 12 of Table 5-2, results are tabulated by probe type and, for manual Type S probes and Autoprobes, by operational mode (straight-up and yaw-nulled). The Autoprobes 16-point data for DeCordova and Lake Hubbard shown in columns 6 and 7 of Table 5-2 are the combined results for Matrix C tests conducted during the day and the additional wall effects tests conducted at night. The day-time and over-night 16-point Autoprobes results are shown separately in Table J-8 in Appendix J. The last two columns in Table 5-2 show results aggregated across probe types and operational modes for 16-point and 12-point tests.

Table 5-2. Summary of Percent Difference Between Original and Wall Effects-Adjusted Average Velocity

Site	Stack Mat'l	Statistics	Manual Type S		Baseline Autoprobes				DAT	Prandtl	French	Aggregate Results[a]	
					16 Point		12 Point						
			Straight-Up	Yaw-Nulled	Straight-Up	Yaw-Nulled	Straight-Up	Yaw-Nulled				16 Point	12 Point
TU DeCordova	Steel	No. Tests	1	1	13	11			1	1	1	7	
		Average	-1.83%	-1.93%	-1.50%	-1.51%			-1 57%	-1 59%	-1.13%	-1.58%	
		s.d.			0.52%	0.59%						0.26%	
TU Lake Hubbard	Steel	No. Tests			10	10			1			3	
		Average			-1.41%	-1.27%			-1.56%			-1.41%	
		s.d.			0.93%	0.97%						0.15%	
AEP Conesville Unit 1/2	Brick and mortar	No. Tests			3	3	3	3				2	2
		Average			-1.94%	-1.92%	-4 23%	-4.28%				-1.93%	-4 26%
		s d.			0 24%	0.17%	0 12%	0.25%				0 01%	0.04%
AEP Picway	Brick and mortar	No. Tests			3	3	3	3				2	2
		Average			-1.92%	-1.95%	-2.37%	-2.53%				-1.94%	-2 45%
		s.d.			0.18%	0 23%	0 39%	0 40%				0.02%	0.11%
Allegheny Mitchell	Brick and mortar	No Tests			4	4	1	1				2	2
		Average			-1.75%	-1.69%	-2 01%	-2.03%				-1.72%	-2 02%
		s d.			1 83%	1 66%						0 04%	0 01%
Averages[b]	Steel	No. Sites	1	1	2	2			2	1	1	2	
		Average	-1.83%	-1.93%	-1.46%	-1.39%			-1.57%	-1.59%	-1 13%	-1 50%	
		s d.			0 06%	0.17%			0 01%			0.12%	
	Brick and mortar	No. Sites			3	3	3	3				3	3
		Average			-1.87%	-1 85%	-2.87%	-2.95%				-1.86%	-2 91%
		s d.			0.10%	0.14%	1 19%	1 18%				0 12%	1 19%
	Steel and brick and mortar	No. Sites	1	1	5	5	3	3	2	1	1	5	3
		Average	-1.83%	-1.93%	-1 70%	-1 67%	-2.87%	-2 95%	-1 57%	-1.59%	-1.13%	-1 72%	-2.91%
		s d.			0 24%	0.29%	1.19%	1.18%	0 01%			0 23%	1.19%

[a] Based on equal weighting of each probe type/mode of operation for site-designated rows and equal weighting of each site for stack material-designated rows.

[b] Based on equal weighting of average results for each site.

The first five three-row blocks in Table 5-2 present results by site. The averages and standard deviations shown in columns 4 through 12 of these blocks are based on equal weighting of each test. The averages and standard deviations shown in the last two columns of these blocks are based on equal weighting of average results for each probe type and operational mode. For example, the average aggregate 16-point result for Lake Hubbard (-1.41%) is based on equal weighting of the average of the ten 16-point Autoprobes straight-up tests (-1.41%), the average of the ten 16-point Autoprobes yaw-nulled tests (-1.27%), and the single DAT test (-1.56%). This approach ensures that results aggregated across probe types and modes of operation are not artificially weighted toward those probe types/modes of operation for which a greater number of tests were performed.

The last three three-row blocks in Table 5-2 present results aggregated according to type of stack, with the first of these three-row blocks presenting results for steel stacks, the second presenting results for brick and mortar stacks, and the third presenting results for all stack types. Results in all columns of these last three three-row blocks are based on equal weighting of average results for each site. For example, the average aggregate 16-point result for brick and mortar stacks (-1.86%) is based on equal weighting of the average result for each of three sites: Conesville Unit 1/2 (-1.93%), Picway (-1.94%), and Mitchell (-1.72%). This approach ensures that results aggregated according to stack type are not artificially weighted toward those sites where a greater number of tests were performed.

5.4.1 General Findings

Weighting the results for each probe type/mode of operation and site equally, as described above, the average percent difference in velocity across all probe types and sites for tests using 16-point Method 1 traverses was -1.72% (s.d. 0.23%, n = 5 sites). With the exception of one 16-point test where there was no (0%) change in velocity, three 16-point tests where the wall effects-adjusted velocity ranged from 0.15% to 1.13% higher than the unadjusted velocity, and four 16-point tests where the percent difference values (-4.33% and -3.87% for brick and mortar stacks and -2.82% and -2.82% for steel stacks) were substantially larger than the next largest percent difference values (see Section 5.4.10), the remaining 62 16-point wall effects tests analyzed for this report resulted in a difference in velocity ranging from -0.61% to -2.36%. For 12-point traverses, the average percent difference in velocity across all probe types and sites was -2.91% (s.d. 1.19%, n = 3 sites), with results across 14 tests ranging from -2.00% to -4.49%.

5.4.2 Point-to-Point Percent Change in Velocity

Figures 5-1 through 5-5 illustrate the percent change in velocity between adjacent traverse points (averaged across all ports) for all probes in the Matrix C (day-time) wall effects tests at DeCordova and Lake Hubbard (Figures 5-1 and 5-2) and for Autoprobes tests at Picway, Mitchell, and Conesville Unit 1/2 (Figures 5-3 through 5-5). The figures indicate that at distances farther than 5 to 7 in. from the stack wall, the average percent change in velocity between adjacent points either varied positively and negatively from point to point (Lake Hubbard, Picway, and Mitchell) or changed very little from point to point (DeCordova and Conesville Unit 1/2 stack). Beginning at a distance of 5 to 7 in. from the stack wall, the average percent change in velocity between adjacent points became significantly and consistently more negative when moving from point to point toward the stack wall, indicating an acceleration in velocity decay in this region.

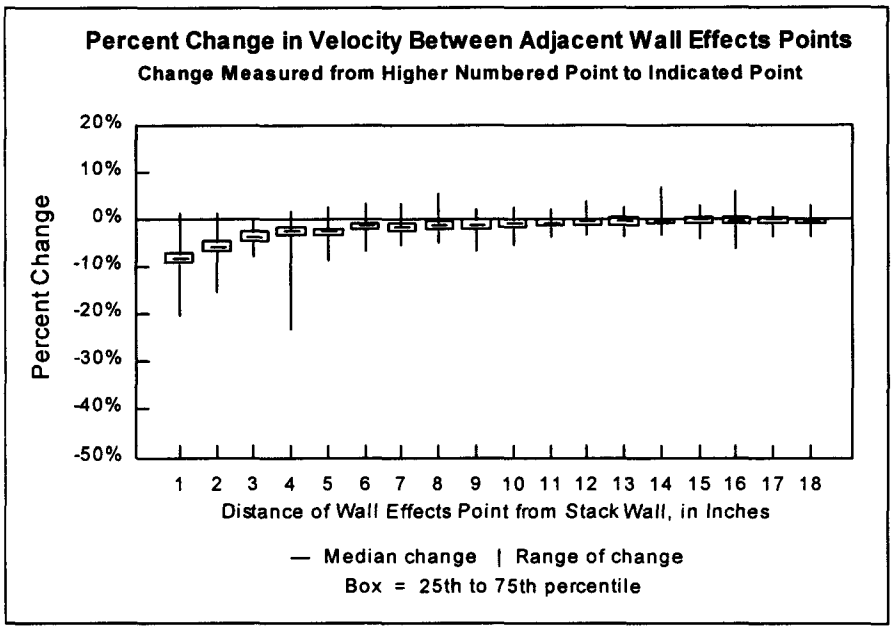

Figure 5-1. Point-to-point percent change in velocity across all probes and ports at DeCordova.

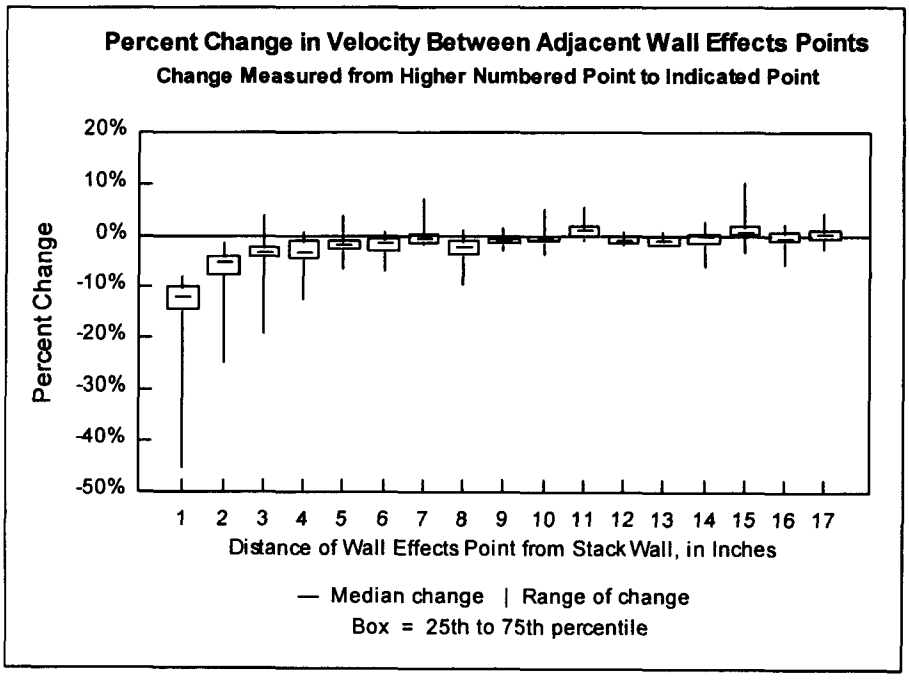

Figure 5-2. Point-to-point percent change in velocity across all probes and ports at Lake Hubbard.

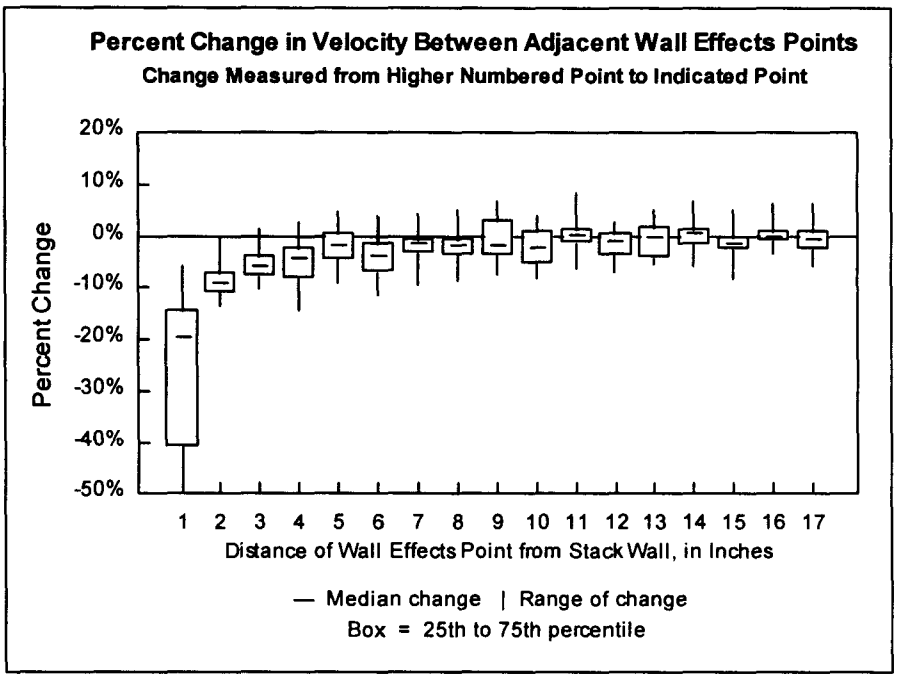

Figure 5-3. Point-to-point percent change in velocity across all ports at Picway.

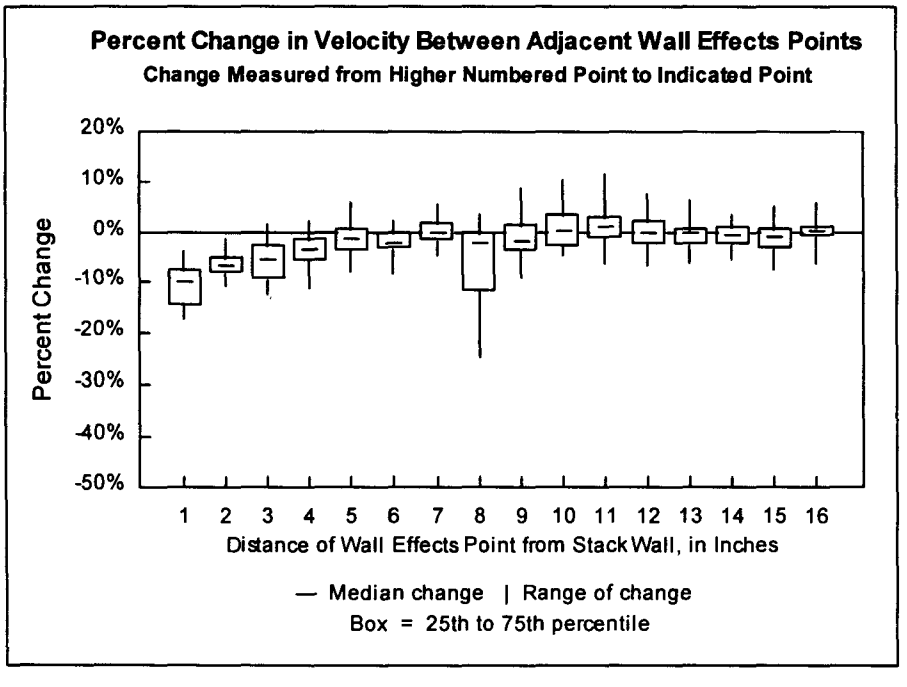

Figure 5-4. Point-to-point percent change in velocity across all ports at Mitchell.

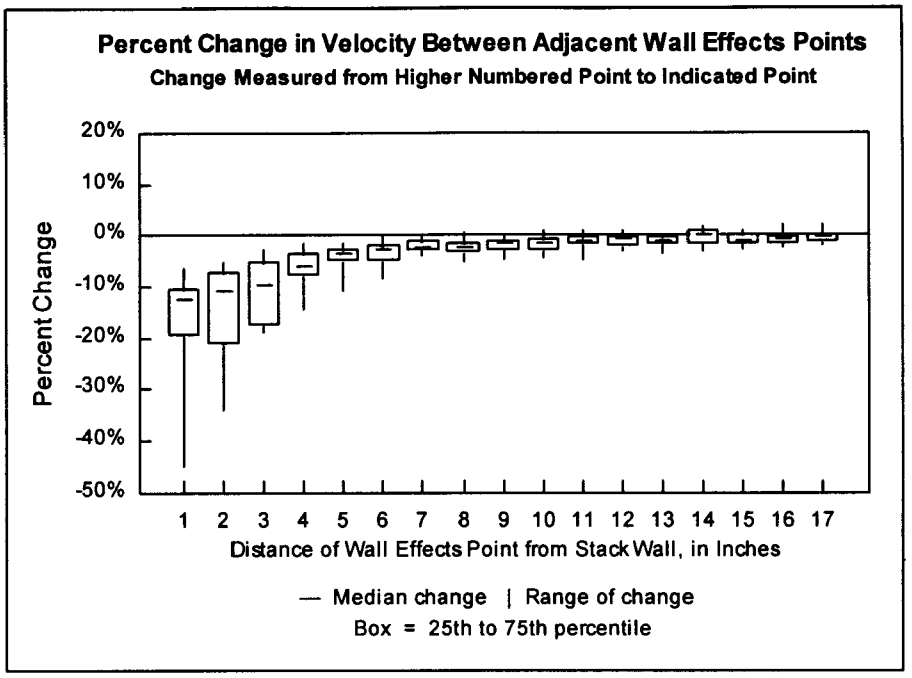

Figure 5-5. Point-to-point percent change in velocity across all ports at Conesville Unit 1/2 stack.

5.4.3 Findings With Respect to Stack Material

Based on the 16-point aggregate results summarized in Table 5-2, the average percent difference in velocity due to wall effects was -1.50% (s.d. 0.12%, n = 2 sites) for smooth stacks (DeCordova and Lake Hubbard) and -1.86% (s.d. 0.12%, n = 3 sites) for rough stacks (Conesville Unit 1/2, Picway, and Mitchell). The observed larger average percent difference for rough stacks is qualitatively consistent with theoretical models of the effects of stack walls on velocity.

Further analysis was performed to determine whether the average percent differences in velocity due to wall effects in smooth and rough stacks were statistically significantly different from each other. For this analysis, the aggregated data used to obtain the results in Table 5-2 were not suitable. The results appearing in Table 5-2 were aggregated to avoid the possibly confounding effects on the overall average that could arise from (1) different numbers of tests using each probe type/mode of operation at any given site and (2) different numbers of tests from site to site. However, the aggregating process reduces the variability associated with the overall average, as reflected in the small standard deviations noted above and shown in the last two columns in Table 5-2. Retaining such variability is critical in evaluating whether two means are statistically significantly different. For this reason, in this phase of the analysis, the average and standard deviation for the 16-point steel stack data and brick and mortar stack data were calculated using the unaggregated percent difference values (n = 50 tests for steel stacks and n = 20 tests for brick and mortar stacks). The upper and lower 95% confidence limits around the means also were calculated, as follows:

$$UCL = \overline{PD} + cc \qquad \textbf{Eq. 5-1}$$

$$LCL = \overline{PD} - cc \qquad \text{Eq. 5-2}$$

where:

$$cc = t_{0.025,\, n-1} \times \frac{S_d}{\sqrt{n}} \qquad \text{Eq. 5-3}$$

UCL	=	upper 95% confidence limit (%),
LCL	=	lower 95% confidence limit (%),
PD	=	average percent difference for steel stacks or brick and mortar stacks (%),
cc	=	95% confidence coefficient (%),
$t_{0.025,n-1}$	=	t-distribution critical value for a probability of 0.025 and n- 1 degrees of freedom,
S_d	=	standard deviation of percent difference values for steel stacks or brick and mortar stacks (%), and
n	=	number of tests used in developing the summary statistics.

The results of the analysis for the unaggregated data are shown in Table 5-3. For purposes of comparison, Table 5-3 also shows the results for data aggregated by probe type/mode of operation at each site and then averaged across sites, as presented in the next-to-last column of the sixth and seventh three-row blocks in Table 5-2. Figure 5-6 displays the average percent difference in velocity due to wall effects for steel stacks and for brick and mortar stacks, along with the corresponding upper and lower 95% confidence limits around these mean values, for the unaggregated data.

Table 5-3. Summary Statistics on Percent Difference in Velocity for Steel Stacks and for Brick and Mortar Stacks Based on Unaggregated Data and Data Aggregated by Site

Stack Type	Steel		Brick and Mortar	
Data Aggregation	**None[a]**	**Probe/site[b]**	**None[a]**	**Probe/site[b]**
No. of Tests Averaged (n)	50	2	20	3
Average	-1.45%	-1.50%	-1.85%	-1.86%
Standard Deviation	0.70%	0.12%	1.00%	0.12%
95% Confidence Coefficient	0.20%	1.08%	0.47%	0.30%
95% Confidence Interval	-1.25% to -1.65%	-0.42% to -2.58%	-1.38% to -2.32%	-1.56% to -2.16%

[a] Based on equal weighting of each test.

[b] Based on equal weighting of the averages for each site, where site averages are based on equal weighting of the average results for each probe type/mode of operation at the site; e.g., for steel stacks, the values in Table 5-3 are the same as the values in the 16-point aggregate results column of the sixth three-row block of rows (steel stack) in Table 5-2.

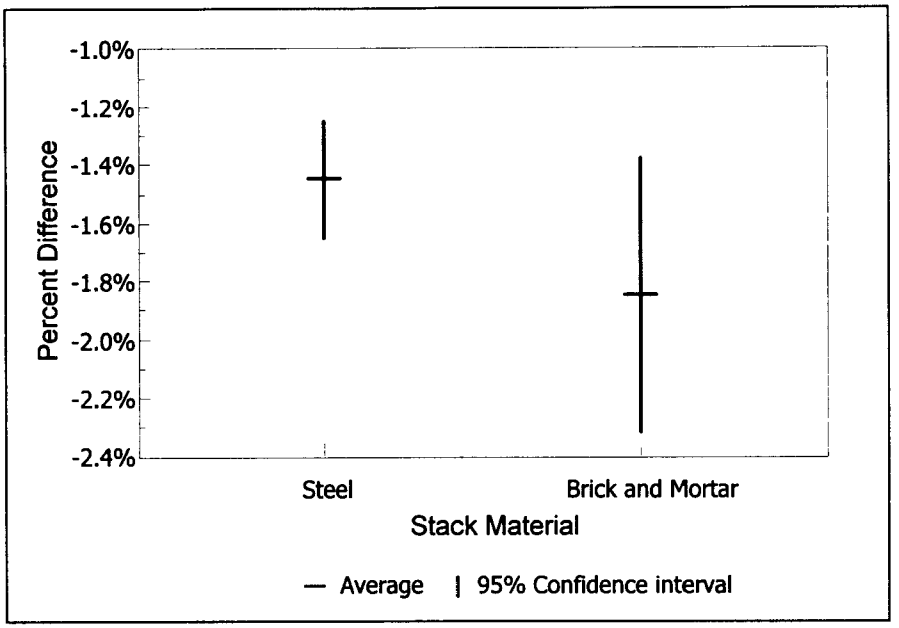

Figure 5-6. Average and 95% confidence intervals of percent difference in velocity, by stack material, based on equal weighting of each test.

Table 5-3 shows that while data aggregation has little effect on the average percent differences for steel stacks and brick and mortar stacks, the standard deviations for the unaggregated data are larger than those for the aggregated data.

Figure 5-6 suggests that the wall effect for brick and mortar stack walls is likely to be greater than that for steel walls, which is qualitatively consistent with theoretical models of wall effects for different types of wall material and attendant roughness. However, the magnitude of the differences is small: 0.40% for the unaggregated data and 0.36% for the aggregated data. In addition, using a t-test to determine whether the average percent difference in velocity for steel stacks is statistically significantly different from the average percent difference for brick and mortar stacks revealed that the difference was not statistically significant at the 95% confidence level. This is reflected in the overlap of the confidence interval lines shown in Figure 5-6.

5.4.4 Site-to-Site Comparisons

Across all sites, the average percent difference in velocity due to wall effects for any site was always within 0.53% of that of every other site, based on 16-point aggregate results. There was only a 0.17% difference in the average percent difference in velocity due to wall effects between the two sites with steel stacks. For the three sites with brick and mortar stacks, the difference in average percent difference in velocity due to wall effects between any two sites was always less than or equal to 0.22%.

5.4.5 Findings Relating to the Number of Points in the Original Traverse

The 12- and 16-point Autoprobes data collected at the Conesville Unit 1/2 stack, Picway, and Mitchell shown in Table 5-2 indicate that the percent difference between original and wall-effects adjusted average velocity becomes larger as the number of points in the original traverse decreases.

This finding is attributed primarily to the way wall effects adjustments were calculated in the study. For the calculational procedure used in the study, as the number of points in the original traverse increases, a smaller proportion of the traverse points are replaced with a wall effects-adjusted value. That is, in a 12-point traverse, one third (four out of 12) of the baseline traverse points are replaced by wall effects-adjusted values; whereas, in a 16-point traverse, one fourth (four out of 16) are replaced. Thus, using the procedure employed in the study, the lower the number of baseline traverse points, the greater is the impact of any velocity decay near the wall. As shown by the aggregate 16- and 12-point percent difference values in the last two columns of the next-to-last three-row block in Table 5-2, the average percent difference in velocity due to wall effects across the three brick and mortar stack sites was -2.91% (s.d. 1.19%, n = 3 sites) for 12-point traverses and -1.86% (s.d. 0.12%, n = 3 sites) for 16-point traverses.

Not only was the average percent difference in velocity due to wall effects larger for 12-point traverses than for 16-point traverses, but the average velocities after adjustment for wall effects for 12-point traverses were lower, on average, than the average velocities after adjustment for wall effects for 16-point traverses. This finding can be seen in Table 5-4, which shows the average velocities for the 16- and 12-point Autoprobes tests at the Conesville Unit 1/2 stack, Picway, and Mitchell. Table 5-4 presents the average unadjusted and wall effects-adjusted velocities for 16- and 12- traverses for tests in which both 16- and 12-point traverses were conducted.[8] Across the three sites, the average unadjusted velocities for 12-point original traverses were 0.79% higher, on average, than the unadjusted velocities for 16-point traverses[9]; whereas, after adjustment for wall effects, the average velocities for 12-point traverses were 0.40% lower than the average velocities for 16-point traverses. It should be noted that the average wall effects-adjusted velocity for 12-point traverses was lower than that for 16-point traverses due to the results for one of the three sites (Conesville 1/2). For the other two sites, the wall-effects adjusted velocity for 12-point traverses was essentially the same as that for 16-point traverses.

Table 5-4. Summary of Unadjusted and Wall Effects-Adjusted Velocities for 16- and 12-Point Traverses

Site	Average Baseline Velocity (ft/sec)			Average Adjusted Velocity (ft/sec)		
	16-Point	12-Point	% Difference[a]	16-Point	12-Point	% Difference[a]
Conesville 1/2	54.49	55.02	0.97%	53.43	52.68	-1.40%
Picway	37.85	38.06	0.55%	37.11	37.12	0.03%
Mitchell	36.54	36.85	0.85%	36.05	36.11	0.17%
Average			0.79%			-0.40%

[a] (12-point - 16-point)/16-point

5.4.6 Probe Type Comparisons

Findings regarding differences in wall effects attributable to probe type are somewhat tenuous because data from different types of probes were not obtained at a sufficient number of sites to allow

[8] Only one of four 12-point traverses conducted at Mitchell met the previously described velocity stability criterion. Thus, data from only the one of four corresponding 16-point tests at Mitchell shown in Table 5-2 are reflected in Table 5-3.

[9] This is qualitatively consistent with the relationship between 12- and 16-point Autoprobes data observed throughout the flow study in which 12-point traverse data are almost always higher than 16-point data.

definitive findings to be derived. While data were obtained at all sites for Autoprobes, data were obtained at only one site (DeCordova) for Type S, Prandtl, and French probes and at only two sites (DeCordova and Lake Hubbard) for DAT probes. Table I1-1 in Appendix I1 shows wall effects adjustment factors (i.e., percent differences) by probe type based on those Matrix C tests at DeCordova and Lake Hubbard where measurements could be made beginning at 1 in. from the stack wall in all ports. As can be seen from this table, Type S probes operated in the yaw-nulled mode produced the largest wall effects adjustment factor, followed in decreasing order of wall effects adjustment factor by Type S probes operated in the straight-up mode, Prandtl, DAT, and Autoprobes operated in the yaw-nulled mode.[10] Spherical probes produced the smallest wall effects adjustment factor, which was based on results for two of four ports sampled in the test at DeCordova. It is emphasized that these findings regarding wall effects adjustment factors and the relative order of probe types with respect to producing wall effects adjustment factors are somewhat tenuous due to the small number of sites where data were obtained for the Type S, DAT, Prandtl, French, and, especially, the spherical probes.

5.4.7 Effect of Stack Gas Velocity

As shown in Figure 5-7, which is based on all tests involving 16-point original traverses, there is no obvious relationship between the percent change in velocity due to wall effects and the average velocity in the baseline traverse.

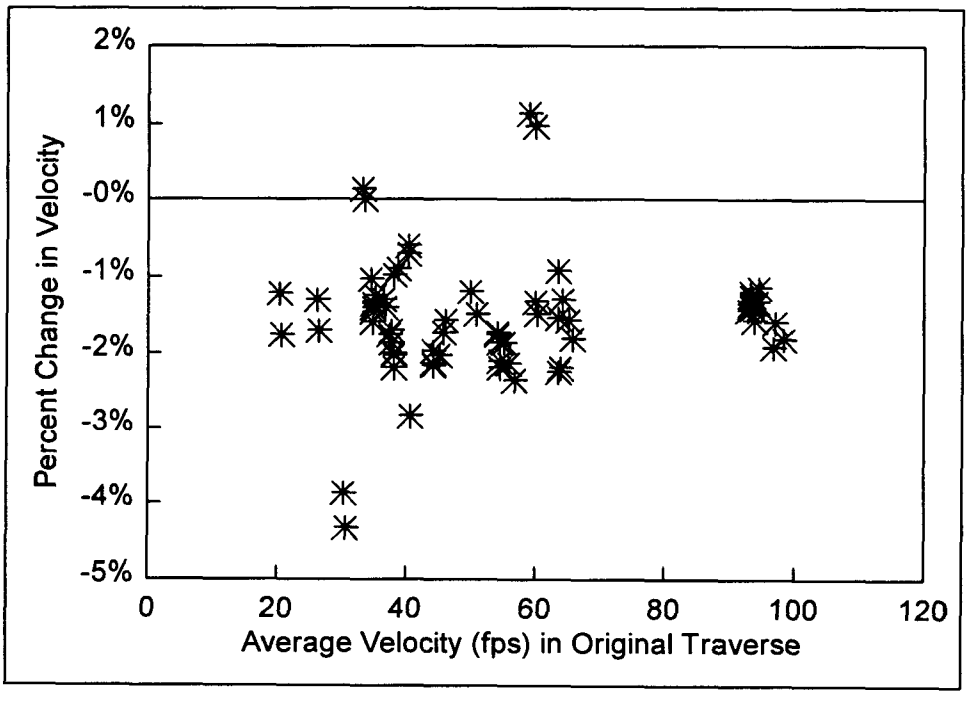

Figure 5-7. Percent change in velocity due to wall effects vs. average velocity in the original traverse.

[10] Table I1-1 does not include a wall effects adjustment factor for Autoprobes operated in the straight-up mode because this probe type was not included in the analyses for which Table I1-1 was prepared. The Autoprobes operated in the straight-up mode produced approximately the same wall effects adjustment factor as the DAT probes.

5.4.8 Findings on Practical Aspects of Wall Effects Testing

Due to the size and shape of the probe head, it was not possible to obtain measurements using modified Kiel probes at distances closer than 4 in. from the stack wall, which is the region where most of the decay in velocity occurs. For this reason, modified Kiel probes are not suitable for use in wall effects testing. (See Sections 6 and 7 for additional findings pertaining to equipment and procedures that are relevant to wall effects testing.)

5.4.9 Wall Effects Calculations Using Minimum Number of Traverse Points

While as many 1-in. incremented points as possible were sampled in the field tests in order to gather information about the velocity decay close to the stack wall, a simpler implementation of the wall effects procedure could involve sampling as few as two points per port: one 1-in. incremented point and d_{rem}. For comparison with the results obtained using all data, the calculational procedure was carried out using only the velocity measurements made at the point located at 1 in. from the stack wall and d_{rem} in each of the four Method 1 sectors adjacent to the stack wall. The percent differences between average stack velocities before and after adjustment for wall effects, calculated both using all data and using only two data points, are presented for each test in Tables J-9 (steel stacks) and J-10 (brick and mortar stacks) in Appendix J. These tables list results for only those tests that met the completeness and velocity stability criteria noted above. Table 5-5 provides summary statistics for the data listed in Tables J-9 and J-10.

Table 5-5 shows that the percent differences calculated using only two data points tend to be smaller in magnitude, or less negative, than their counterparts calculated using all data points. Across all tests at all stacks, percent difference values calculated using all data points ranged from -4.33% to +1.13%, with an average of -1.56% (s.d. 0.81%, n=70 tests), whereas the values calculated using only two data points ranged from -1.79% to +2.41%, with an average of -0.72% (s.d. 0.66%, n=70 tests).

5.4.10 Maximum and Minimum Percent Differences

The maximum and minimum percent differences (decreases) in average velocity due to adjustment for wall effects were -2.83% and +1.13%, respectively, for tests at steel stacks (see Table J-9 in Appendix J), and -4.33% and +0.15%, respectively, for tests at brick and mortar stacks (see Table J-10 in Appendix J). The positive percent differences clearly are anomalous. These anomalous values may be due to a slight overall increase in the stack velocity between the times when measurements were made in the baseline traverse and in the wall effects traverse (even though the data met the previously discussed velocity stability criterion) or to imprecision (random fluctuations) in the baseline and/or wall effects traverse measurements. The first hypothesis is supported by the fact that the anomalous minimum value for steel stacks was obtained during overnight tests at Lake Hubbard when the load was not as closely controlled as during the day-time tests. Similarly, the anomalous minimum value for brick stacks was obtained during a test at Mitchell, where the load may not have been controlled as closely as during the DeCordova and Lake Hubbard day-time tests. The minimum *negative* percent differences in average velocity for steel and brick and mortar stacks were -0.61% (see Table J-9 in Appendix J) and -1.23% (see Table J-10 in Appendix J), respectively.

The two largest percent differences for tests at steel stacks (-2.82% and -2.82%) (see Table J-9 in Appendix J) and for tests at brick and mortar stacks (-4.33% and -3.87%) (see Table J-10 in Appendix J) also are suspicious in that they are significantly larger than the next largest percent

difference values. These anomalous maximum values also may be due to imprecision in the baseline and/or wall effects traverse measurements or to a slight overall decrease in the stack velocity between the times when measurements were made in the baseline traverse and in the wall effects traverse. This latter hypothesis is supported by the fact that the anomalous maximum values for steel stacks were obtained during overnight tests at DeCordova when the load was not controlled as closely as during the day-time tests. Similarly, the anomalous minimum values for brick stacks were obtained during tests at Mitchell where anomalous minimum values also were obtained. Excluding the possibly anomalous maximum and next-to-maximum values, the next largest percent differences were -2.36% for tests at steel stacks (see Table J-9 in Appendix J) and -2.21% for tests at brick and mortar stacks (see Table J-10 in Appendix J).

Table 5-5. Comparison of Percent Differences Between Original and Wall Effects-Adjusted Average Velocities, Using All Data Points and Using Only Two Data Points per Port

All Stacks	% Difference Using All Data Points	% Difference Using Two Data Points
Smallest (Least Negative) % Difference	1.13%	2.41%
Smallest Negative % Difference	-0.61%	-0.08%
Largest (Most Negative) % Difference	-4.33%	-1.79%
Average (Excluding Non-negative % Differences)	-1.69%	-0.87%
Standard Deviation	0.62%	0.37%
Average (Using All % Differences)	-1.56%	-0.72%
Standard Deviation	0.81%	0.66%

Steel Stacks	% Difference Using All Data	% Difference Using Two Data Points
Smallest (Least Negative) % Difference	1.13%	0.20%
Smallest Negative % Difference	-0.61%	-0.08%
Largest (Most Negative) % Difference	-2.82%	-1.71%
Average (Excluding Non-negative % Differences)	-1.55%	-0.74%
Standard Deviation	0.23%	0.36%
Average (Using All % Differences)	-1.45%	-0.71%
Standard Deviation	0.70%	0.39%

Brick and Mortar Stacks	% Difference Using All Data Points	% Difference Using Two Data Points
Smallest (Least Negative) % Difference	0.15%	2.41%
Smallest Negative % Difference	-1.23%	-0.39%
Largest (Most Negative) % Difference	-4.33%	-1.79%
Average (Excluding Non-negative % Differences)	-2.06%	-1.08%
Standard Deviation	0.80%	0.37%
Average (Using All % Differences)	-1.85%	-0.75%
Standard Deviation	1.00%	1.08%

SECTION 6
FINDINGS

Drawing on the analyses presented in Sections 3 and 4, the findings for each of the tested probes and related issues are presented below.

6.1 DAT PROBE

The DAT probe gave favorable results with respect to approaching the central tendency of the data (within ±1.0% of the grand mean at DeCordova, Lake Hubbard high load, and Homer City and 2.3% below the grand mean at Lake Hubbard low load). This finding is supported by the refined central tendency analysis in Section 4 (DAT probes were always within ±1.3% of the grand mean across all sites). At Lake Hubbard and Homer City (the sites with non-axial flow), the flow values measured with the 3-D probes (DAT and spherical) statistically distinguishable from flow measured with the 1-D probes (Type S straight-up, French, and Prandtl), while they were not at DeCordova (the site with axial flow).

The DAT probe's variability was typical of manual probes (the coefficient of variation ranged from 2.08% at Homer City to 3.14% for the low-load runs at Lake Hubbard), but was significantly higher than that typical of automated probes. Conversely, the DAT probe had the largest range of ranks in the rank-order analysis of volumetric flow, which suggests high run-to-run variability. Such variability is also evident in the confidence interval plots of the DAT probe data. This variability is not unexpected for a probe that measures three separate pressure drops for each velocity determination. In addition, the differences in probe copies and test team operation described below contribute to probe variability.

The ANOVA analyses of volumetric flow indicate that at two sites probe copy was a significant factor in volumetric flow obtained by the DAT probes; that is, some variation in flow is attributable to different copies of the probe. In addition, at one site the test team was a significant factor in flow variation. The differences in results produced by different copies of the DAT probe are not surprising because, unlike with all other probes, which were provided by Cadmus, each test team supplied its own DAT probe.

6.2 PRANDTL PROBE

Although the coefficient of variation of the Prandtl probe at the two gas sites where it was tested was the lowest of any of the manual probes (0.63% and 1.27%), mean flow readings at both sites were higher than the central tendency. The Prandtl probe had the highest deviation from the central tendency (4.3%) at the near-axial site (DeCordova) and a less severe excursion (2.2%) at the moderate yaw/moderate pitch site (Lake Hubbard). This finding is consistent with the refined central tendency analysis in Section 4 (3.53% from the central tendency at DeCordova and 2.08% at Lake Hubbard). Because the Prandtl is used as a standard to calibrate other probes in wind tunnels with axial flow, the high deviation at the near-axial site is unexpected. On the other hand, its more central behavior at Lake Hubbard suggests that the flow reading taken with the probe may not be adversely affected by moderate yaw and pitch. At Decordova, the only site where a Matrix B test was performed on the Prandtl probe, a test team effect but no probe copy effect was detected. It is interesting to note that the flow values determined by the Prandtl probe are similar to those found by the manual Type S probe operated in the yaw-null mode.

6.3 SPHERICAL PROBE

The four original spherical probes (damaged at DeCordova and subsequently repaired and re-calibrated before Lake Hubbard) produced moderately high flow values at DeCordova and by far the lowest flow values for both the high- and low-load tests at Lake Hubbard (12.3% and 10.6%, respectively, below the grand mean). If Run 3 is excluded, as suggested by peer reviewers, the average flow measurement for spherical probes for Lake Hubbard high-load tests is lower than the grand mean by only 3.45%. At Lake Hubbard low-load, the percent deviation from the grand mean was three times greater than that of any other method. On the other hand, at Homer City where both the new set and old set of spherical probes were tested, the volumetric flow values were very close to each other and much closer to the central tendency of the data (2.0% and 2.1% below the grand mean). The rank order of the spherical probes was on the high end at DeCordova, lowest at Lake Hubbard, and in the middle at Homer City. Consistent results were observed from the refined central tendency analysis in Section 4, except that the deviation from the central tendency was smaller across all sites.

At Lake Hubbard, the site with the highest traverse point-to-traverse point variation in flow velocity and flow angle, the spherical probe's variability as reflected in the CV of the flow measurements was 19.70% at high load (nearly four times that of any other probe type) and 7.39% at low load (nearly twice that of any other probe type). Again, if Run 3 is excluded, the coefficient of variation for the high-load test drops significantly from 19.70% to 1.35%. At the other test sites, the variability of flow values was comparable to those of other manual probes. An inspection of the confidence interval plots in Appendix D reveals that the high variability at Lake Hubbard high load was due to a dramatically lower flow value determined by one spherical probe operated by one test team during a single run, whereas at Lake Hubbard low load it was due to swings in the spherical probe's flow values from run to run. Inspection of the dispersion analysis graphs in Appendix G also indicates greater swings and spread among the flow values obtained by different copies of the spherical probe across all test sites, as compared to other probe types. Of all probes tested, test team effect was most pronounced for the spherical probes. In all Matrix B (intraprobe comparison) runs at Lake Hubbard and Homer City, the choice of test team was a significant factor in measured volumetric flow. In addition, one of the new spherical probes was found to produce flow measurements that were statistically different from the other three probes even though all four probes had the same design and dimensions and were fabricated by the same manufacturer.

6.4 AUTOPROBES YAW-NULLED

Of all the tested probe types, the baseline Type S Autoprobe system operated in the yaw-nulled mode with 16-point traverses was closest to the central tendency across all three sites and load levels, as found in both the original and refined central tendency analyses (0.3%, -0.1%, -0.4%, and -1.6% for DeCordova, Lake Hubbard high load and low load, and Homer City, respectively, from the central tendency in the original analysis, and -0.23%, -0.20%, 0.61%, and -0.53% for DeCordova, Lake Hubbard high load and low load, and Homer City, respectively, from the central tendency in the refined analysis). Results for 12- and 48-point traverses and the single manual Autoprobe were comparable. The baseline 12-, 16-, and 48-point Autoprobes yaw-nulled and the manual Autoprobe yaw-nulled displayed the lowest variability among all the methods tested, with coefficients of variation consistently below 1%. For example, the coefficients of variation of the baseline Autoprobes 16-point yaw-nulled were 0.32%, 0.42%, 0.63%, and 0.99% at DeCordova, Lake Hubbard high load and low load, and Homer City, respectively. These central tendency and

variability results are confirmed by visual inspection of the confidence interval plots in Appendix D and dispersion plots in Appendix G.

6.5 AUTOPROBES STRAIGHT-UP

Straight-up operation of the Autoprobes consistently produced higher flow values than yaw-nulled operation. As expected, the difference between the straight-up and yaw-nulled modes increased from near-axial to moderate yaw/moderate pitch to high-yaw flow conditions. For example, the percent differences between the yaw-nulled and straight-up 16-point baseline Autoprobes were 0.14%, 1.08%, 4.33% for DeCordova, Lake Hubbard, and Homer City, respectively. For 16-point traverses, the deviation from the grand mean of straight-up operation increased from 0.3% at DeCordova, to 1.2% and 0.9% for Lake Hubbard high load and low load, to 2.7% for Homer City. At the near-axial site (DeCordova) the deviation from the grand mean for straight-up operation (for 16-point, 48-point, or manual 16-point) was no more than 0.2% larger than the deviation from the grand mean for yaw-nulled operation. At the other sites, differences between the straight-up and yaw-nulled values were substantially larger (up to 1.4% at Lake Hubbard and 4.4% at Homer City). The low variability observed for yaw-nulled operation was also found for straight-up operation. The coefficient of variation of the manual and baseline Autoprobes 12-, 16- and 48-point straight-up never exceeded 1.23% at any site. The central tendency and variability results for the straight-up operation with 48 traverse points at the two gas-fired sites (i.e., within 0.3% of the central tendency with coefficients of variation less than 0.31%) were comparable to the results discussed in the previous section for the Autoprobes operated in yaw-nulled mode.

6.6 TYPE S PROBE YAW-NULLED

The Type S probe operated in yaw-nulled mode consistently produced flow results 2.2% to 2.9% higher than the grand mean. The results of the refined central tendency analysis in Section 4 are consistent with these results (1.75% to 3.10% above the central tendency). As stated in Section 6.2, the percent difference from the grand mean and the average volumetric flow values for the Type S probe yaw-nulled were very similar to those for the Prandtl probe at the two gas-fired sites, where Prandtl probes were tested. At each site the average volumetric flow values measured by the two probes differed by less than 1%. The coefficients of variation for the Type S probe yaw-nulled across all three sites ranged from 1.76% to 3.23%, which was typical for the manual probes tested. No significant differences among probe copies were found, but at Lake Hubbard some test team effect on flow measurement was detected. As with the Autoprobes, yaw-nulled operation consistently produced lower flow values than straight-up operation. As expected, the difference between the two modes increased from near-axial to moderate yaw/moderate pitch to high-yaw flow conditions, where the differences between the two modes, derived in terms of the percent difference from the baseline Autoprobes 16-point straight-up data, were 0.50%, 1.93%, 4.60% for DeCordova, Lake Hubbard, and Homer City, respectively). These differences between the two modes are similar to those observed for the Autoprobes.

6.7 TYPE S PROBE STRAIGHT-UP

The Type S probe operated in the straight-up mode has been the standard method for measuring volumetric flow for over 20 years. The rank order analysis showed the median rank of its volumetric flow values to be highest or second highest at each of the three field test sites. The Type S probe straight-up had the highest positive difference (6.9%) from the central tendency at Homer City and in low-load operation at Lake Hubbard (3.4%), second highest for high-load operation at Lake

Hubbard (6.0%), and third highest at DeCordova (3.5%). In the refined central tendency analysis in Section 4, the Type S probe straight-up had the highest positive deviation from the central tendency at Lake Hubbard high load and low- load, and Homer City (5.69%, 4.15%, and 7.62%, respectively), and the second highest at DeCordova (2.52%). At the same time, its measurements at the near-axial site (DeCordova) were slightly lower than those of the Prandtl, which is the accepted standard pitot used to calibrate other probes in wind tunnels with axial flow. The coefficient of variation of the Type S straight-up measurements was 2.31% or less at all three sites. Some probe copy effect was detected at DeCordova and some test team effect was detected at Lake Hubbard, but neither of these effects was detected at the coal-fired site (Homer City).

6.8 FRENCH PROBE

At the moderate yaw/moderate pitch and high-yaw angle sites (Lake Hubbard and Homer City), the French probe produced volumetric flow values that were lower than the central tendency (-2.0% and -3.3%). At Homer City, the French probe produced the lowest measurements of all the tested in-stack methods (3.3% below the grand mean). The French probe's coefficient of variation ranged from 1.62% at the gas-fired site with axial flow (DeCordova) to 4.18% at the coal-fired site, where it was the most variable of all tested methods. Its variation in rank order relative to the central tendency, from +6 at DeCordova to -6 at Lake Hubbard and -8 at Homer City, suggests a strong dependence of flow measurement capability on yaw angle. The variability of the French probe is confirmed by visual inspection of the confidence interval plots (in Appendix D) and dispersion analysis plots (in Appendix G). At the two sites where Matrix B tests were performed on the French probe, no test team or probe copy effects were detected.

6.9 MODIFIED KIEL PROBE

The modified Kiel probe produced the highest volumetric flow values of all the tested methods at Lake Hubbard, high load (7.3% above the grand mean) and second highest at DeCordova (3.6% above the grand mean). However, at Homer City, the modified Kiel probe had the second smallest deviation from the grand mean (0.3%) of all tested methods. At Lake Hubbard, low load it had the smallest deviation (2.0%) from the grand mean of all the tested in-stack manual methods. The modified Kiel probe's CV at Homer City (1.24%) was lowest of any of the tested manual probes. At DeCordova, the CV (1.39%) was in the middle of the range for manual probes. The probe had the second highest CV at Lake Hubbard (5.19% at high load and 4.04% at low load). The high variability at Lake Hubbard is evident in the confidence interval plots (in Appendix D). Some probe copy effect was detected at Homer City, but not at DeCordova. No test team effect was detected at either site.

6.10 COMPARISON TO ENGINEERING BASELINE

One of the collateral goals of the field study was to reduce the disparity reported by industry between in-stack and combustion-based calculations of heat rate and flow. To evaluate this reported disparity, the differences between flow values measured by the in-stack probes and baseline values determined using the MMBtu method were calculated. At DeCordova, Lake Hubbard high load and low load, and Homer City, respectively, the following percent differences from the MMBtu method flow values were found: 0.39%, -1.88%, -3.55%, and -1.44% for the Autoprobes 16-point yaw-nulled; 2.45%, 0.04%, -1.16%, and 1.88% for the Type S probe yaw-nulled; and 0.37%, -2.71%, -5.33%, and -0.86% for the DAT probe.

6.11 Comparison of Manual and Electronic Pressure Measuring Devices

From a practical standpoint, differences in pressure readings between manual and electronic pressure devices were generally small. Where statistically significant differences were detected, the differences were not consistent for different types of probes and different velocity conditions. When used with Type S probes, however, manual devices tended to read higher pressures than electronic transducers in field tests, but read lower than electronic transducers in the wind tunnel tests. Generally, the differences between the manual and the electronic devices are smaller on average and less variable in the wind tunnel tests than in the field tests.

6.12 Wall Effects

Wall-effects adjustments to flow generally resulted in a decrease in calculated velocity. The average percent difference in average velocity across all probe types and field test sites for which complete wall effects data sets were obtained is -1.72% (s.d. = 0.23%, n = 5 sites). (Complete wall effects data sets were obtained at DeCordova and Lake Hubbard and at three additional sites that were tested solely for wall effects.) No significant difference was found among probe types. A small difference was observed between the percent difference in the average velocities measured in smooth (steel) stacks (-1.50%) and rough (brick and mortar) stacks (-1.86%). However, the variance of the wall effects measurements was sufficiently large that these differences were not found to be statistically significant at the 95% confidence level. As expected, the percent decrease in average velocity between unadjusted and wall-effects adjusted traverses becomes smaller as the number of points in the Method 1 traverse increases.

6.13 Calibrations

6.13.1 One- and Two-Dimensional Probes

All probes were calibrated at North Carolina State University (NCSU) before and after field testing. Probes were also calibrated at National Institute of Standards and Technology (NIST) after the field testing was completed. Comparison of the NCSU pre-and post-test results showed that all the post-test C_p values derived at 60 ft/sec were within ±1.5% of the pre-test values (the values used to calculate flow in the field tests). The largest differences between pre-test and post-test C_p values occurred at 30 ft/sec. For all probes, the post-test 30 ft/sec C_p values were higher than the pre-test coefficients, ranging from 0.4% higher for one Prandtl probe to 5.1% higher for one Type S probe. Pre- and post-test C_p values obtained at 60 and 90 ft/sec were generally consistent. Even with the comparatively large changes observed at 30 ft/sec, C_p values averaged over all three velocities changed by less than 1.5% between pre- and post-test calibrations. Although the C_p values measured at NIST were generally higher than the corresponding NCSU post-test values, the NIST post-test C_p values were within ±2.2% of the NCSU values.

6.13.2 Three-Dimensional Probes

DAT Probes

For three of the four DAT probes, differences in velocities calculated between pre- and post-test NCSU calibrations were less than 2% in the -20° to +20° pitch angle range. In the pitch angle range of -10° and +10°, which is comparable to the range of pitch angles measured at the three utility stacks during this field study, the differences in velocity were less than 1%. For reasons that are unclear, the velocity differences ranged from 3.35% to 4.18% within the -10° to +10° pitch angle range for the fourth DAT probe. A comparison of calculated velocities between the NIST and post-test NCSU calibrations shows that, within the -10° to +10° pitch angle range, calculated velocities over the three

nominal wind tunnel velocity settings (i.e., 30, 60, and 90 ft/sec) were 1.7% or less for each probe. In the -30° to +30° pitch angle range, the average percent difference of calculated velocity between the NIST and post-test NCSU calibrations was less than 3.8% for each probe.

Spherical Probes

For the four original spherical probes, the results of the four sets of calibrations at NCSU showed that the calibration values varied by up to 5% over all pitch angles. For the second set of probes, the change in calibration was approximately 4% to 5% within the -10° to +10° pitch angle range. A comparison of the calculated velocities for the NIST and NCSU post-test spherical probe calibrations shows that, within the -30° to +30° pitch angle range, the calculated NCSU velocity for one probe was low relative to NIST (-1.2%). For all other probes, the NCSU calculated velocities were 1.9% to 6.4% higher than those obtained by NIST.

SECTION 7
RECOMMENDED EQUIPMENT AND PROCEDURES FOR MEASURING VOLUMETRIC FLOW

A variety of equipment and procedures for measuring volumetric flow was evaluated during this study. This section describes the equipment and procedural revisions and their intended purpose, actual field performance, and potential value in improving measurements of volumetric flow.

7.1 BACKGROUND

The recommendations contained in this section include proposed performance standards for the equipment and procedures used in performing yaw, pitch, and velocity calibrations. The performance standards are expressed as tolerance limits (e.g., ±2°). Tables 7-1 through 7-3 show the proposed tolerances for each constituent factor and the resulting cumulative tolerance limits obtained by summing the tolerances of all constituent factors. The "recommended" performance specifications represent tolerances that the field tests indicated are achievable through careful performance of the test procedures using currently available technology.

Two principles were followed in establishing the "recommended" specifications on the constituent factors:

- Field and laboratory tests had to indicate that the specified error limits were well within the capabilities of commonly available equipment and typical test personnel.

- The error limits on each constituent factor were set so that the cumulative yaw and pitch angle tolerances would not exceed 10° and the cumulative velocity calibration tolerance would not exceed 8.5%. Figure 7-1 shows the test method error resulting from various errors in angle measurements.

7.2 EQUIPMENT REVISIONS

Recommended equipment revisions based on the field study results include specifications for physical features of probes and probe components and the use of probe supports and stabilization devices. Additional recommendations involve the use of digital inclinometers for determining yaw angles, the use of electronic transducers during probe calibrations in wind tunnels and during field tests, and cross-sectional area requirements for wind tunnels used for determining probe-specific calibration coefficients.

7.2.1 Probes and Probe Components

Any probe used to determine the yaw angle of the total velocity vector should have certain physical specifications, such as effective length of the probe, inside diameter (I.D.) of the pressure tubes, a probe sheath, and a scribe line.

Physical Dimensions

Maintaining the probe head in a horizontal position is an important consideration in limiting error in angle and velocity measurements. For probes that measure pitch, the degree of deflection from horizontal results in an equivalent displacement in pitch with the resulting error in velocity shown

Table 7-1. Proposed Yaw Angle Tolerances

Factor	Tolerances
Wind tunnel axial flow	±3°
Reference scribe line precision	±1°
Inclinometer precision	±1°
Protractor wheel measurement resolution	±1°
Pointer needle resolution	±1°
Rotational position check	±1°
Yaw angle calibration	±1°
Cumulative Tolerance[a]	7° - 8°

[a] The tolerances associated with the protractor wheel and pointer needle are included in the cumulative tolerance only when these devices are used for yaw angle measurement.

Table 7-2. Proposed Pitch Angle Tolerances

Factor	Tolerances
Wind tunnel axial flow	±3°
Pitch plate resolution	±1°
Horizontal straightness of probe	4°
Pitch angle calibration curve	±2°
Cumulative Tolerance	10°

Table 7-3. Proposed Velocity Calibration Tolerances

Factor	Tolerances
Wind tunnel velocity drift during calibration	2%
Error due to wind tunnel blockage	0.5%
Calibration coefficient variation	
2D probes	3%
3D probes	
Within ±10° pitch range	3%
Beyond ±10° pitch range	5%
Cumulative Tolerance[a]	5.5% - 7.5%

[a] Velocity drift must be factored into the cumulative tolerance when a probe is calibrated in a wind tunnel where the standard and tested pitots take successive pressure measurements during calibration, not in tunnels where they can take concurrent measurements during calibration.

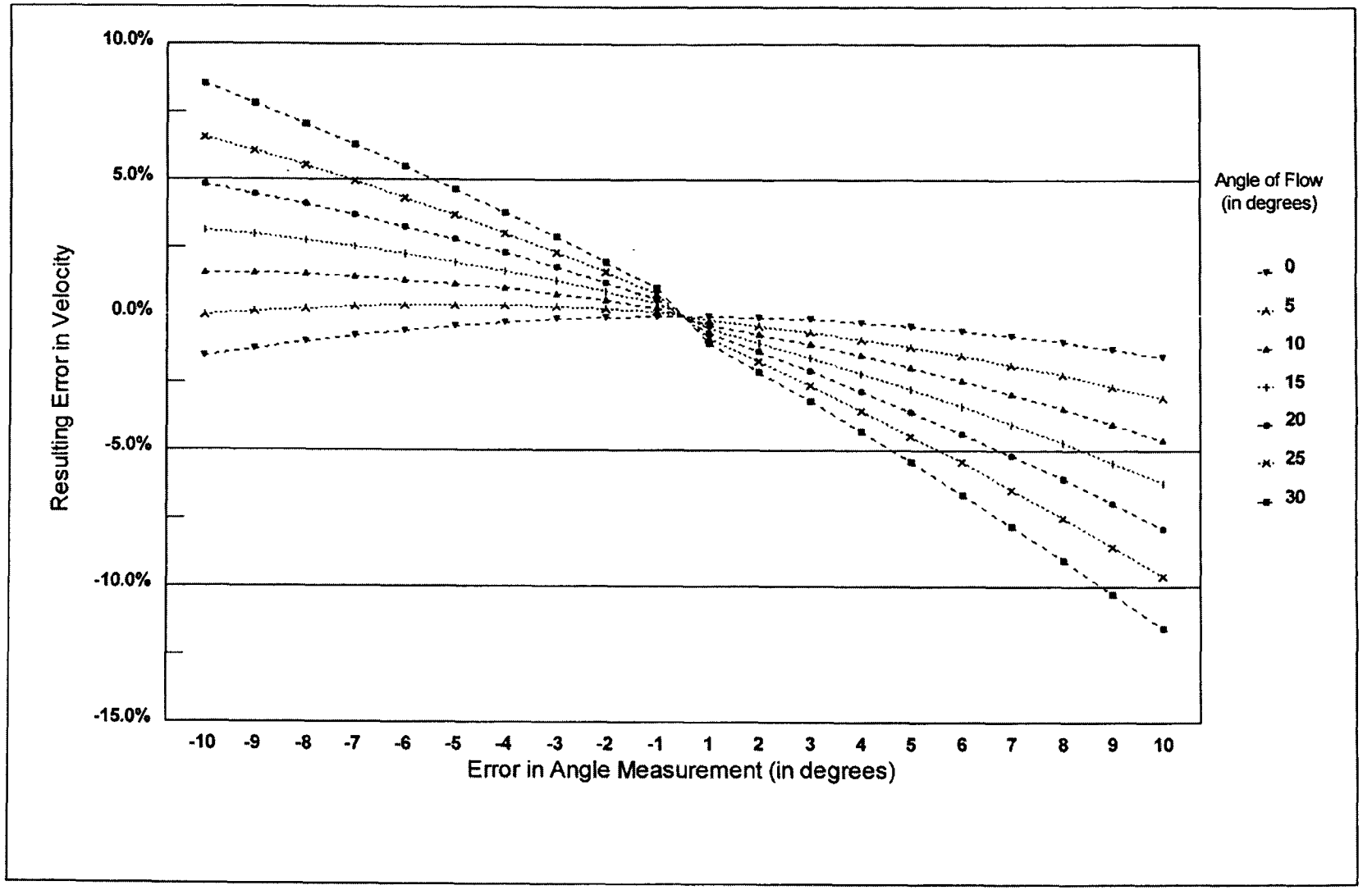

Figure 7-1. Error in test method produced in error angle measurements. {Note: Percentage error in velocity = [cos(flow angle + error) - cos(flow angle)]}.

in Figure 7-1. In addition, even for probes that do not measure the pitch angle of flow, deflection from horizontal affects the applicable velocity calibration coefficient and could affect the yaw angle measurements. For example, wind tunnel tests[11] that were performed in preparation for the field studies whose results are reported here showed that, at a pitch displacement of -5°, the calibration coefficient for a Type S probe changed by as much as 2.4%.

Experience gained during the three field tests suggested a simple equipment requirement for maintaining the horizontal position of manual probes: the effective length of the probe (coupled with a cylindrical probe extension, if necessary) should be at least 3 ft longer than the last traverse point marking on the probe shaft. An extension of this length will enable an operator to maintain the probe's horizontal stability when it is fully inserted into the stack. Other physical mechanisms are also available for maintaining the horizontal position of the probe. For example, a comparable horizontal position could be maintained by a probe inserted into a bushing sleeve installed on a test port or an automated system that includes an external probe casing with a transport system.

The inside diameter (I.D.) of the pressure port connection tubing should be at least 1/4 in. to reduce the time required for pressure equilibration. This recommendation stems from the field test at the DeCordova site, where the response time for one of the DAT probes with 1/16-in. I.D. port connection tubing was approximately one minute longer than the response times of all other probes, which had 1/8-in. I.D. diameter connection tubing. To further reduce response time, the tubing that is used to connect the probe and the pressure-measuring device should be as short as practicable. An on-site response time check of the measurement system setup may indicate that the connection tubing should be replaced, if the response time appears to be too slow.

Sheath Requirements for Probes Used to Determine Yaw Angle

Velocity probes, particularly Type S pitot probes, are used by testing firms in a variety of stack test applications and configurations with other test equipment. Current test method requirements include design specifications for the geometry of the pitot head and pressure ports, but no specifications are prescribed for the probe shaft. Although the same type of probe head is used in these applications, a diverse variety of probe sheath configurations is often used, including cylindrical and rectangular configurations. In some applications, probes are used without sheaths, and the pitot tubes are simply welded together to add rigidity to the probe shaft.

Some of these probe shaft configurations are unsuitable for performing yaw-nulling during wind tunnel calibrations and field tests, because precise rotational movements and measurements are difficult to make. Two-dimensional (2-D) and 3-D (3-D) probes fabricated for this field test and others supplied by the test teams comprised a uniform set of probes with cylindrical sheaths that extended along the full length of the probe shaft. The cylindrical outer sheath provided a surface suitable for inscribing a yaw-null reference line and for attaching an angle-measuring device. The sheath also enabled precise rotational movement of the probe during the yaw-nulling procedure.

For yaw-nulling with 2-D and 3-D probes, the physical design of the probe should (1) provide a surface for inscribing a reference scribe line (i.e., a line corresponding to the probe's yaw-null rotational position); (2) accommodate attachment of an angle-measuring device to the probe shaft; and (3) facilitate precise rotational movement of the probe for determining yaw angles. The wind

[11] The Cadmus Group, Inc. 1997, "Flow Reference Method Testing and Analysis: Wind Tunnel Experimental Results," EPA/430/R-97-013.

tunnel and field tests reported here suggest that these capabilities can readily be provided by requiring probes to include a cylindrical sheath rigidly attached to the probe assembly and enclosing all pressure lines. The sheath should extend continuously from within 4 in. to 12 in. of the probe head to the farthest position away from the probe head, where an angle-measuring device may be attached during use in the field. The sheath should be fabricated of material that can be permanently scribed with a reference scribe line that is sufficiently precise to indicate rotational changes to within ±1°. The sheath of the fully assembled probe must be sufficiently rigid and straight at all rotational positions such that, when one end of the probe shaft is held in a horizontal position, the extent of the bend in the fully extended probe meets the specifications recommended in section 7.3.2 below.

Other sheath designs (e.g., a rectangular sheath mounted in a gimbal) should be considered if they can be shown to satisfy the three requirements described in the previous paragraph.

Reference Scribe Line

During the pre-test NCSU calibrations of the 2-D and 3-D probes, a calibration was performed to demonstrate that the yaw-nulling procedure used by each probe was accurate to within ±2° at the velocities where a calibration coefficient was determined. Following these checks, a reference position, representing the true zero yaw position, was marked on each probe shaft. Field test personnel then used this marking as a point of reference for installing the digital inclinometer.

This experience led to several observations. First, an indication of the zero yaw position should be permanently marked on the probe, not by visual inspection of the probe, but by yaw-nulling the probe in a wind tunnel. The wind tunnel tests showed that scribe lines placed on probes by visual inspection could not reliably meet the ±2° specification, whereas those placed by yaw-nulling could consistently meet this specification. Second, a permanent line inscribed on the probe shaft is very useful for lining up the angle-measuring device with the zero yaw position on any probe that will be used to determine yaw angles of flow. Finally, the zero yaw position that is permanently marked on the probe should be checked by yaw-nulling the probe in a wind tunnel before its initial use in the field. Any measured offset from the zero yaw position should be recorded for use in offsetting yaw measurements in the field. This check should be treated as a calibration record and then verified during each subsequent calibration.

7.2.2 Probe Supports and Stabilization Devices

When probes are used for determining flow angles, keeping the probe head in a relatively stable horizontal position is critical. Although little can be done to eliminate vibration of the probe head at a traverse point in the stack, test personnel can, at a minimum, secure the section of probe extending outside the test port. Two types of stabilization devices were used in the field tests. At DeCordova and Lake Hubbard, monorails installed above each port were used to support the probes. For the Homer City test, which was performed from within an annular space between the stack and outer shell, stands were provided to the test teams upon which the probe shafts were rested to help maintain horizontal position during flow angle determinations.

The physical characteristics of each test platform may dictate the most suitable type of stabilization device. For probes longer than 10 ft, the portion outside the stack should be secured. Three types of devices are recommended: (1) monorails installed above each port, (2) probe stands on which the probe shaft may be rested, (3) and bushing sleeves of sufficient length secured to the stack ports to maintain the probe in a horizontal position. Additionally, testers should place a level or an angle-

measuring device on the portion of the probe shaft that extends outside of the test port to ensure that the probe is horizontal before taking readings.

7.2.3 Yaw Angle-measuring Devices

During the field tests, digital inclinometers were used on each of the 2-D and 3-D probes instead of an analog gauge typically used by testers. The digital device proved superior to the analog gauge for two reasons. First, the digital device provided greater precision in angle measurements, having a manufacturer's rated precision of 0.1°. Second, it was easier to read the digital inclinometer during the field tests when the probe end and inclinometer extended well beyond the railing of the stack platform.

Digital inclinometers or a comparable angle-measuring device capable of measuring and displaying the rotational position of the probe to within ±1° should be used with manually operated probes. The device should be able to be locked into position on the probe shaft and retain its original physical location on the probe shaft during the full course of field testing. Angle readings should be recorded to the nearest whole number (i.e., the display readings of 0.1° should be ignored).

A collar or block that can be attached to the probe sheath may be required for locking the digital inclinometer or comparable angle-measuring device into position on the probe shaft. An engineering diagram of the block collar used in these field tests appears in Appendix H of the DeCordova field test plan.[12]

Alternatively, a protractor wheel and pointer assembly can be used to measure yaw angles. The protractor wheel can be attached to a port opening and set in a fixed rotational position to indicate the yaw angle position of the probe relative to the longitudinal axis of the stack or duct. A pointer assembly that includes an indicator needle mounted on a circular collar can be placed on the probe shaft and be locked into a rotational position on the probe's reference scribe line. The protractor wheel and pointer assembly must be able to read rotational angles to within 1°.

7.2.4 Pressure-measuring Devices

The pressure gauge comparison reported in Section 4.4 indicated that, in wind tunnel tests, consistent differences could not be detected among inclined manometers, Magnehelic® gauges, and electronic manometers. However, the NCSU principal investigator suggested that differences between the probe calibrations conducted at NCSU at different times (described in Section 2.2.1 of this report) could be due in part to a 2% to 3% error in the Magnehelic® gauge readings, particularly at the low pressures measured at 30 ft/sec. When used to measure velocity pressure in electric utility stacks, small differences were detectable between the electronic manometers and the other pressure-measuring devices under turbulent flow conditions, but not under more stable flow conditions (see Appendix I2 of this report).

Superior data capture, elimination of operator error, and ease and speed of use represent advantages of electronic manometers coupled with an electronic data capture device for both field testing and wind tunnel use. However, the absence of consistently detected differences between the electronic manometers and other pressure-measuring devices in most situations suggests that use of electronic manometers should be recommended rather than required.

---

[12] The Cadmus Group, Inc., 1997, "Flow Reference Method Testing and Analysis: Field Test Plan, DeCordova Steam Electric Station," EPA/430/R-97-024.

7.2.5 Wind Tunnel Cross-sectional Area Requirements

Questions have been raised regarding the quality of calibration factors derived in wind tunnels having small cross-sectional areas relative to the size of the probe head. During this study, calibration factors assigned to probes by vendors often differed substantially from the calibration factors derived in the NCSU wind tunnel tests and subsequently confirmed in the NIST independent calibrations. A primary cause of this disparity appears to be the size of the wind tunnels used by probe suppliers to calibrate their equipment.

Wind tunnels used to calibrate velocity probes must provide axial flow within the test section, stable velocity over time, and consistent velocity measurements made by standard and test probes. They should be capable of achieving velocities between 20 ft/sec and 100 ft/sec. The cross-sectional area of the tunnel must be large enough to ensure fully developed flow in the presence of both the tested probe and standard calibration pitot. To ensure this condition is met, the cross-sectional area of the test section should be of sufficient size that the volume of the probe head, shaft, and attached devices satisfies the following condition:

$$V \leq 0.005(A)^{1.5} \qquad \textbf{Eq. 7-1}$$

where:

V = volume of the probe head, shaft, and attached devices for both the test and calibration probes, if placed simultaneously in the wind tunnel, or the larger of the test and calibration probes, if placed separately in the wind tunnel; and

A = cross-sectional area of the test section.

Equation 7-1 is a generalized statement of the conditions necessary to ensure that the effect of the probe blockage on the flow velocity in a wind tunnel will not exceed 0.5%. For the probes used in the field tests, this inequality will be satisfied when the projected area of the probe head, shaft, and attached devices does not exceed 4% of the cross-sectional area of the tunnel, and the diameter of a circular wind tunnel, or width of the shorter side of a rectangular wind tunnel, is at least 12 in.

To further ensure that wind tunnels meet Agency standards, wind tunnel operators should be required to calibrate Agency-supplied audit probes on a periodic basis and report the results to the Agency.

7.3 Procedural Revisions

Based on the results of this study, several recommendations are being made regarding wind tunnel operation, probe calibration, and field test set-up and performance.

7.3.1 Wind Tunnel Procedures

Wind tunnel procedures recommended for revision include verifying axial flow, placing scribe lines on probes, and performing yaw-nulling and velocity calibrations; wind tunnel facility specifications are also recommended.

Wind Tunnel Axial Flow Verification

The design of the wind tunnel and the location of the test section relative to flow disturbances must ensure stable, fully developed axial flow patterns at the test point(s) where the probes are calibrated

and at the point where the calibration pitot tube is positioned for the reference pressure measurements (if different from the test probe location).

To verify that axial flow exists in these locations within the wind tunnel, a series of yaw- and pitch-angle measurements must be taken using either a 3-D probe or a 2-D wedge probe. A previously calibrated 3-D probe would allow the operator to obtain both the yaw- and pitch-angle measurements using one port in the wind tunnel. Using an uncalibrated 3-D probe or a 2-D wedge probe would require that measurements be made from two ports, that is, yaw angles would first be measured using the tested probe port and then measurements would be repeated using the 90° offset port that provides the pitch angle of flow. These angle measurements should be made at the lowest and highest velocities at which probes will be calibrated in the tunnel. To ensure that flow in the tunnel is sufficiently axial to maintain the precision requirements described earlier in this section, all such angle measurements must be within ±3° of 0°.

Wind Tunnel Velocity Drift Check

To confirm that flow within the wind tunnel remains stable, a velocity drift check should be performed each time a probe is calibrated in the tunnel. The drift check should show that, at each wind tunnel velocity setting, consecutive velocity pressure measurements taken by the calibration pitot remain within a predefined limit (e.g., a change of less than or equal to 2%). This check would not be necessary in cases when concurrent, paired calibration pitot tube and tested probe pressure measurements are taken.

Placement of Reference Scribe Line

As noted above in section 7.2.1, the zero yaw position should be permanently marked on the probe, not by visual inspection, but by yaw-nulling the probe in a wind tunnel. Two alternative procedures could be used to establish the location of this reference scribe line on the probe sheath. In the first approach, the orientation of the line is established by yaw-nulling the probe in the wind tunnel (i.e., yaw angle calibration) and then placing the line at some specified angular position relative to this determined yaw-null position. The second approach would be to arbitrarily place the reference scribe line on the sheath, and the offset would then be determined during the yaw angle calibration in the wind tunnel. This second approach may be preferable because it is likely to be easier to implement. With either approach, the offset of the line relative to the probe's yaw-null position would become part of the probe's calibration record and would be used by field test personnel in determining the correct placement of the angle-measuring device on the probe shaft. Without a reference scribe line, documented offset, and periodic yaw offset verifications (described below), the probe should not be used for measuring yaw angles and deriving yaw-adjusted velocity values.

Reference scribe lines may not be necessary for an automated probe system if a reference rotational position of the probe is built into the probe system design. For these probe systems, a "flat" (or comparable, clearly identifiable physical characteristic) should be provided on the probe casing or flange plate to ensure the reference position of the probe assembly remains in a vertical or horizontal position.

Yaw-nulling Calibration

If probes are used to determine the yaw angle of flow, a quality control check should be implemented to document the rotational position of the reference scribe line relative to the probe's yaw-null position to within ±1°.

This procedure should be performed before the probe is first used in a field test. It should also be performed during each subsequent wind tunnel calibration.

Velocity Calibration Procedures

The probe calibration procedures followed during the NCSU and NIST calibrations were based on the current specifications in Methods 1 and 2 and draft Method 2F. The calibration routine was performed at three wind tunnel velocity settings of 30, 60, and 90 ft/sec.

A minimum of three replicate measurements was taken at each velocity setting for 2-D probes and at each velocity and pitch angle setting for 3-D probes. Upon completion of the first set of calibrations at the three velocities, the wind tunnel was shut down. The operator then repeated the probe set-up procedures (i.e., verified the alignment of the digital inclinometer to the probe head and measured the probe tip insertion depth in the tunnel). After the wind tunnel velocity was again set to nominal, the yaw-nulling procedure and calibration routine were repeated at the three velocities.

The calibration procedures for 2-D and 3-D probes should include the provision to perform the calibrations at two wind tunnel velocity settings, the expected minimum and maximum velocities where the probes will be used. The calibration coefficients obtained for 2-D probes should remain stable to within ±3% over this range or the calibration should have to be repeated at two alternative velocity settings until the ±3% specification is met. The calibration curves for 3-D probes should meet similar specifications. At 5° increments over the ±15° pitch range, the percent difference between the velocity calibration coefficients, F_2, obtained at the two chosen nominal settings should not exceed ±3%. Beyond the ±15° pitch range, the percent difference should not exceed ±5%.

To build adequate quality control into the test method, probe recalibrations should be performed after a specified time period (e.g., 12 months from its first field use) or number of field tests (e.g., 10). A recalibration should also be performed whenever routine visual inspection of the probe indicates a physical change that may affect the previously derived calibration.

Calibration procedures for the 2-D and 3-D probes should be consistent, except for the procedures specifically affecting the determination of yaw and pitch angles. For example, the calibration procedure for 3-D probes should be the same as that used for the 2-D probes, except for the provisions pertaining to pitch-angle orientation during tests and the generation of calibration curves.

The calibration procedures outlined in the unpromulgated 1993 version of draft Method 2F include the requirement that the 3-D probe be calibrated in the pitch-angle range of -60 to +60°. Because pitch angles are typically within ±15° at utility stacks, a pitch-angle range of ±15° should suffice under most conditions. If larger pitch angles are anticipated, then the calibration must include the expected pitch angle range.

7.3.2 Field Test Procedures

Recommended revisions to field test procedures include procedures to (1) minimize bending of probe shafts, (2) verify correct placement of the yaw angle-measuring device, (3) account for the yaw component of flow, (4) account for near-wall velocity decay, (5) account for test port extension into the stack, and (6) measure velocity at individual traverse points.

Horizontal Straightness Test

Bends in the probe shaft can affect the measurement of differential pressure and determinations of yaw and pitch angles. A procedure was implemented in the field tests to ensure that a probe shaft is at least sufficiently rigid and straight to achieve a basic tolerance for straightness. For probes that are coupled with a probe extension, this check will verify that the probe head and extension are properly aligned when assembled. To verify that the probe shaft has not been bent beyond tolerance in the course of testing, the horizontal straightness test must be performed before and after use in field testing or wind tunnel calibration.

The procedure is performed on the fully assembled probe (including the probe head and all probe shaft extensions). The probe is secured in a horizontal position using a stationary support at a point along the probe shaft approximating the location of the stack entry port when sampling at the farthest traverse point from the stack wall. The unsupported length of probe should be at least the distance from the stack entry port to the traverse point farthest from the wall. An angle-measuring device is used to measure the horizontal position of probe shaft at the secured end. With the probe held in this position, the angle-measuring device is placed next to the probe head to measure the declination (sag angle) from horizontal.

Alternatively, the declination can be determined trigonometrically by measuring the distance from the floor to the centerline of the probe sheath at the controlled end and to the centerline at the probe tip, that is, at the free end. The sag angle can then be calculated from the relative difference between the two measurements, which is referred to as the sag distance and the distance along the probe shaft between the measurement points at the controlled and free ends of the probe.

Based on performance of the horizontal straightness check by the test teams that participated in the field tests and subsequent checks made on various test probes, the sag angle limits shown in Table 7-4 were found to be readily achievable.

Rotational Position Calibration

The probe's reference scribe line is permanently marked on the probe shaft, and any offset from the zero yaw position is documented during the wind tunnel calibration. The reference scribe line and offset are used by field test personnel to determine the correct placement of the angle-measuring device on the probe shaft for the field test. A procedure used during the field tests provided verification that the angle-measuring device, once attached to the probe and aligned with the reference scribe line (including any offset), correctly indicated the rotational position of the probe head to within ±2°.

Table 7-4. Horizontal Straightness Declination Limits

For Horizontal Distance (d_p) Between Controlled and Free End	Sag Angle (α) Specification
$d_p \leq 10$ ft	$\lvert\alpha\rvert \leq 2°$
10 ft. $< d_p \leq 20$ ft	$\lvert\alpha\rvert \leq 3°$
$d_p > 20$ ft	$\lvert\alpha\rvert \leq 4°$

With the fully assembled probe held horizontally, direct measurements were made of the rotational positions of the reference scribe line and the angle-measuring device to verify that the specification was met. A digital inclinometer was affixed to the probe sheath and locked into position on the reference scribe line. This inclinometer provided a measure of the rotational position of the reference scribe line. In the pre-test procedure, this angle was used to align a second digital inclinometer (i.e., the one attached to the other end of the probe to measure yaw angles during the field test) in the same rotational orientation as the reference scribe line (to within ±2°). The post-test check consisted of repeating the procedure to confirm that the rotational orientation of the digital inclinometer that was attached to the probe was still within ±2° of the rotational position of the reference scribe line.

A similar out-of-stack check can be performed on an automated probe system by placing an inclinometer on a "flat" part of the probe casing or a flange plate to establish a reference vertical or horizontal position of the probe assembly. The rotational position of the probe head then can be set to a series of yaw angles, as determined by the controller for the automated system, and checked by positioning the inclinometer on the side of the probe head.

Experience in the field tests indicated that a ±1° standard was readily achievable in the pre-test check, and a ±2° standard was readily achievable in the post-test check.

Although not experienced during the three field tests conducted for this project, testers may encounter physical constraints at a test location that prohibit full assembly of the probe and extension outside of the stack. In circumstances where a probe extension has to be added to the probe while remaining in the test port, the rotational position check should be performed each time that an extension is added to the probe and the angle-measuring device is re-positioned on the probe shaft.

Yaw-nulling Procedures

If probes are to be used to account for the yaw angle component of flow in a stack, suitable yaw-nulling procedures must be defined. For manual probes, the yaw angle is determined by rotating the probe until a null differential pressure reading is obtained. If impact and static ports are used for yaw-nulling (e.g., with the Type S probe), the probe head must be rotated 90° back to orient the impact opening into the direction of flow. In this position, the yaw angle is the rotational angle between the pitot impact port and the longitudinal axis of the stack, as indicated by the digital inclinometer. The differential pressure reading is then recorded at this probe rotational position.

3-D probes are yaw-nulled by rotating the probe until zero differential pressure between ports P_2 and P_3 is obtained. The yaw angle is then determined using the digital inclinometer.

An automated probe system that uses a curve-fitting routine for determining the yaw-null position of the probe head should have its yaw-nulling procedure verified in a wind tunnel as being equivalent to the manual procedure described above.

Wall Effects

Methods 1 and 2 do not include procedures that directly account for the decline in velocity near the stack wall. Currently, the only way to account for the wall effect is to increase the number of traverse points across the entire stack, rather than to increase the number of traverse points only close to the wall. A method was used during the field tests for taking a series of measurements close to

the stack wall that can be factored into the calculation of the average axial stack velocity in order to account for any velocity decay near the wall.

Using this procedure, the original velocity measurement at the Method 1 traverse point closest to the stack wall is replaced with a substitute wall effects-adjusted velocity value that captures any measured velocity decline close to the stack wall. This procedure (1) assumes that the velocity at the stack wall drops to zero, (2) takes measurements in 1-in. increments starting as close to the wall as the chosen probe allows and continuing out as far as the inside edge of the Method 1 near-wall equal-area segment, and (3) algebraically integrates over these measurements. Because wall effects measurements are made to within 1 in. of the stack wall, it is particularly important that test ports be adequately sealed to minimize leakage during wall effects tests.

Traverse Point Locations

The test ports at each of the three field test sites extended into the stack by 1 in. or less. This port lip extension, which could be greater at other sites, must be accounted for when determining the position of the traverse point markings on the velocity probe.

After accounting for such physical offsets, the probe length necessary to reach each traverse point should be marked directly on the probe sheath. Before beginning any field test, an out-of-stack verification should be performed to ensure that these position markings are correct. For manual probes, these point positions should be marked on the probe shaft and verified by measuring the distance of each marking from the probe pressure port. The automated probe system can extend the probe to each of the prescribed point positions, which can be verified by measuring the distance between the port flange and probe head pressure port.

Calculation of Average Velocity

Unlike draft Method 2F, Method 2 does not derive flow velocity values at individual traverse points. Instead, to derive the average stack gas velocity, Method 2 averages the differential pressures obtained at all the traverse points and then uses the resulting average differential pressure to calculate average stack gas velocity. Such an approach does not allow use of information from yaw-angle determinations, which require calculation of flow velocity at each individual traverse point.

If yaw angle is to be determined, the approach in Method 2 should be replaced by a procedure that allows differential pressure to be measured at individual traverse points and flow velocity to be calculated at each traverse point. The yaw-adjusted velocity is then obtained by multiplying this calculated impact velocity by the cosine of the yaw angle measured at that traverse point. Finally, the yaw-adjusted velocity values at all traverse points are averaged to obtain the average stack gas velocity. Such a procedure was used with the yaw-nulled probes involved in the field study.