

AIR TOXICS EMISSION FACTORS FOR COMBUSTION SOURCES USING PETROLEUM-BASED FUELS

VOLUME 2 DEVELOPMENT OF EMISSION FACTORS USING API/WSPA APPROACH

Health and Environmental Affairs Department
Publication Number 348

**DEVELOPMENT OF TOXICS EMISSION FACTORS FOR PETROLEUM
INDUSTRY COMBUSTION SOURCES**

**Draft Final Report
Volume 2 - Development of Emission Factors
Using CARB Approach**

Prepared for

**Scott Folwarkow
Western States Petroleum Association
2300 Clayton Road, Suite 1440
Concord, Ca 94520**

and

**Karen Ritter
American Petroleum Institute
2020 L Street, N.W.
Washington, DC 20005**

Prepared by

**Energy and Environmental Research
18 Mason
Irvine, CA 92718**

July 14, 1997

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
Disclaimer	i
Acknowledgments	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Acronyms	vi
1.0 Introduction	1
2.0 Data Collection	2
3.0 Screening	2
3.1 Procedures	2
3.2 Results	3
3.3 Source Prioritization	3
4.0 Detailed Validation	4
4.1 Detailed Validation Results	5
4.1.1 Validation Problems and Calculation Check Failures	6
4.1.2 Detailed Validation Flags	8
5.0 Data Extraction	12
6.0 Emission Factors	13
6.1 Design and Operating Parameters	13
6.2 Normalizing Units	13
6.3 Run Specific Method Rating	14
6.4 Run Specific Emission Factor Calculation	15
6.5 Major and Sub Group Evaluation Parameters	16
6.6 Detailed Data Listings	17
6.7 Outlier Analysis	17
6.8 Sub Group Evaluation	18
6.8.1 Boiler - Fuel Oil	20
6.8.2 Heater - Refinery Gas	20
6.8.3 Boiler - Refinery Gas	21
6.8.4 Internal Combustion Engine - Diesel	22
6.8.5 Internal Combustion Engine - Field Gas	22
6.8.6 Internal Combustion Engine - Natural Gas	23
6.8.7 Steam Generator - Crude Oil	23
6.8.8 Steam Generator - Natural/Casing Gas	24
6.8.9 Steam Generator - Natural Gas	24
6.8.10 Turbine - Natural Gas	24
6.8.11 Turbine - Natural/Refinery Gas	25
6.9 Sub Group Emission Factor Calculation	25
6.10 Sub Group Method and Population Ratings	26
6.11 CARB Overall Quality Rating	26
6.12 EPA Overall Quality Rating	27

LIST OF TABLES

<u>Table</u>	<u>Title</u>	<u>Page</u>
1	Data Collection Summary for CARB Project	29
2a	Report Information Extracted for Screening	30
2b	Device Information Extracted for Screening	31
2c	Substance Information Extracted for Screening	32
3a	Screening Summary by Source Type (Report Information)	33
3b	Screening Summary by Source Type (Substance Information)	34
4	Source Ranking Summary	35
5	Listing of Sources Considered for Emission Factor Development	36
6	Listing of Sources Not Considered for Emission Factor Development	37
7	Summary of Major Validation and Extraction problems	38
8	Key Design and Operating Parameters	40
9	Assigned Source Classification Codes and Units	41
10	Listing of Secondary and Primary Validation Checks for CARB and EPA Methods Applicable to Project	42
11	Method Rating Summary	43
12	Major Groups and Comparison Matrix	44
13	Outlier Action Report by Device and Substance	46
14	Outlier Action Report by Major Group and Substance	48
15	Emission Factor Groups	51
16	Emission Factor Group Comparison	52
17	Substance Specific Uncertainty and Relative Standard Deviation	54
18	CARB Overall Rating	56
19	EPA Overall Rating That Would be Assigned to California Data	57
20	Emission Factors SCC Units	58
21	Emission Factors lb/MMBtu	76

ACRONYMS

4D 2378	Same as 2,3,7,8-Tetrachlorodibenzo-p-dioxin.
4D Other	Same as Tetrachlorodibenzo-p-dioxin other.
4D Total	Same as Tetrachlorodibenzo-p-dioxin total.
4F 2378	Same as 2,3,7,8-Tetrachlorodibenzofuran.
4F Other	Same as Tetrachlorodibenzofuran other.
4F Total	Same as Tetrachlorodibenzofuran total.
5D 12378	Same as 1,2,3,7,8-Pentachlorodibenzo-p-dioxin.
5D Other	Same as Pentachlorodibenzo-p-dioxin other.
5D Total	Same as Pentachlorodibenzo-p-dioxin total.
5F 12378	Same as 1,2,3,7,8-Pentachlorodibenzofuran.
5F 23478	Same as 2,3,4,7,8-Pentachlorodibenzofuran.
5F Other	Same as Pentachlorodibenzofuran other.
5F Total	Same as Pentachlorodibenzofuran total.
6D 123478	Same as 1,2,3,4,7,8-Hexachlorodibenzo-p-dioxin.
6D 123678	Same as 1,2,3,6,7,8-Hexachlorodibenzo-p-dioxin.
6D 123789	Same as 1,2,3,7,8,9-Hexachlorodibenzo-p-dioxin.
6D Other	Same as Hexachlorodibenzo-p-dioxin other.
6D Total	Same as Hexachlorodibenzo-p-dioxin.
6F 123478	Same as 1,2,3,4,7,8-Hexachlorodibenzofuran.
6F 123678	Same as 1,2,3,6,7,8-Hexachlorodibenzofuran.
6F 123789	Same as 1,2,3,7,8,9-Hexachlorodibenzofuran.
6F 234678	Same as 2,3,4,6,7,8-Hexachlorodibenzofuran.
6F Other	Same as Hexachlorodibenzofuran other.
6F Total	Same as Hexachlorodibenzofuran total.
7D 1234678	Same as 1,2,3,4,7,8-Heptachlorodibenzo-p-dioxin.
7D Other	Same as Heptachlorodibenzo-p-dioxin other.
7D Total	Same as Heptachlorodibenzo-p-dioxin total.
7F 1234678	Same as 1,2,3,4,6,7,8-Heptachlorodibenzofuran.
7F 1234789	Same as 1,2,3,4,7,8,9-Heptachlorodibenzofuran.
7F Other	Same as Heptachlorodibenzofuran other.
7F Total	Same as Heptachlorodibenzofuran total.
8D	Same as Octachlorodibenzo-p-dioxin.
8F	Same as Octachlorodibenzofuran.
AB2588	Air Toxics "Hot Spots" Information and Assessment Act of 1987
AB	Afterburner
APC	Air Pollution Control
API	American Petroleum Institute
APS	Air Pollution Control System
ATEDS	Air Toxic Emission Data System
B	Benzene
BAAQMD	Bay Area Air Quality Management District
BD	Blank Data
BF	Baffle Filter
BTX	Benzene, Toluene, and Xylene
C	Cyclone or Carbon
CD	Calibration Data
CARB	California Air Resources Board
CO	Carbon Monoxide
COC	Carbon Monoxide Oxidation Catalyst

CVAAS	Cold Vapor Atomic Absorption Spectrometry
CVR	Case Vapors Recovered
Cr	Chromium
D	Dioxins
DD	Device Description
DM	Demister
dscfm	Dry Standard Cubic Feet per Minute
dscf	Dry Standard Cubic Feet
EER	Energy and Environmental Research Corporation
EF	Emission Factor
EPA	Environmental Protection Agency
USG	United States Government
F	Formaldehyde
GFAAS	Graphite Furnace Atomic Absorption Spectrometry
H ₂ S	Hydrogen Sulfide
HC	Hexavalent Chromium
HCHO	Formaldehyde
HCl	Hydrogen Chloride
HF	Hydrogen Fluoride
HI	Hydrogen Chloride
HNO ₃	Nitric acid
HO	Halogenated Organics
Hp	Horse Power
HS	Hydrogen Sulfide
ICAP	Inductively Coupled Argon Plasma
ICE	Internal Combustion Engine
IS	Internal Standards
LD	Laboratory Data or Location Data
lbs/MMcf	Pounds per Million Cubic Feet
lbs/Mgal	Pounds per Thousand Gallons
lbs/ton coke	Pounds per Ton Coke
lbs/1000 Barrels	Pounds per Thousand Barrels
LP	Liquid Petroleum
M	Metals
MMBtu	Million British Thermal Units
MMcf	Million Cubic Feet
MD	Method Description
MDL	Method Detection Limit
Mgal	Thousand Gallons
MMT	Multiple Metals Train
Ni	Nickel
NIOSH	National Institute of Occupation Safety and Hazard
NO _x	Nitrogen oxides
O ₂	Oxygen
O	Oxygen
PAH	Polycyclic Aromatic Hydrocarbons
PCB	Polychlorinated Biphenyls
PCDD	Polychlorinated Dibenzo-p-dioxin
PCDF	Polychlorinated Dibenzofuran
PH	Polycyclic Aromatic Hydrocarbons
ppbv	Parts per Billion Volume

PQL	Practical Quantitation Limit
PR	Process Rate
QA/QC	Quality Assurance/Quality Control
QD	Quality Assurance/Quality Control Data
RFG	Refinery Fuel Gas
ROC	Reactive Organic Compound
RSD	Relative Standard Deviation
S	Strokes per Cycle
SCAQMD	South Coast Air Quality Management District
SCC	Source Classification Code
SCR	Selective Catalytic Reduction
SD	Sample Data or Spray Dryer
SG	Steam Generator
SIC	Standard Industrial Classification
SO ₂	Sulfur Dioxide
SVOC	Semi-Volatile Organic Compounds
THC	Total Hydrocarbons
Ti	Titanium
TO	Thermal Oxidizer
VC	Vinyl Chloride
VOC	Volatile Organic Compound
WSPA	Western States Petroleum Association

1.0 Introduction

A wide range of air toxics source test results have been generated since the California Air Toxics "Hot Spots" Information and Assessment Act of 1987 (AB 2588) was initiated. The Western States Petroleum Association (WSPA) funded a project in the early 1990's to develop an emission factor database from WSPA funded tests. Emission factors resulting from this project are provided in the document titled "Pooled Source Testing of Combustion Devices: Database Users Guidance Manual" which was released on January 13, 1994. The California Air Resources Board (CARB) recently funded a project to collect all AB2588 source test reports including petroleum and nonpetroleum sources. Over 800 tests were collected by CARB. CARB developed over 2000 emission factors from 200 of the collected tests. These emission factors are available in the report titled "Development of Toxics Emission Factors from Source Test Data Collected under the Air Toxics Hot Spots Program" which was released on April 11, 1996. CARB did not develop emission factors for petroleum industry sources because they believe the petroleum industry has already developed and is continuing to develop emission factors.

As the primary contractor for the CARB emission factor project, Energy and Environmental Research (EER) identified several tests collected by CARB which were not included in the WSPA emission factor database. Under joint funding from WSPA and the American Petroleum Institute (API), EER was contracted to update the existing WSPA database with the additional CARB test reports. CARB agreed to provide WSPA and API access to the source test reports provided the emission factors were developed using CARB procedures. In addition to providing access to the test data, CARB allowed WSPA to use its emission factor graphical user interface (GUI). The GUI permits users to sort, list, print, and export emission factors from personal computers. This project benefited the petroleum industry in the following ways:

- Districts, CARB, EPA, and the petroleum industry have the most representative and complete database available for petroleum industry combustion devices.
- All source test results were reviewed using procedures developed by CARB. These procedures were developed to assess the quality of source test results allowing emission factors of known quality to be developed. Using emission factors developed from CARB procedures and approved by CARB add credibility to the resulting emission inventories.
- All emission factors are available in an easy to use GUI. The GUI allows users to search for emission factors, generate reports, and export data for use in spreadsheets and other programs. Industry could not access the previous WSPA emission factor database by computer.

This report describes the procedures used to validate and develop emission factors for petroleum industry combustion sources. The report also provides the results of the validation efforts and emission factor calculation procedures. It should be noted that the procedures were developed by CARB for the project discussed previously. Section 2 provides a brief description of the CARB data collection efforts. A description of the screening procedures applied by CARB are provided in Section 3 along with a list of priority sources for emission factor development. The validation procedures and results are described in Section 4. Data extraction procedures are described in Section 5. Emission factor calculation procedures and results are described in Section 6. Supporting documentation such as statistical analyses are provided in Volume 3. The GUI is described in a document titled the "Database User's Manual".

2.0 Data Collection

The California Air Resource Board (CARB) recently completed a project to collect all air toxics source test reports prepared in response to the California Air Toxics "Hot Spots" Information and Assessment Act of 1987 (AB 2588). The CARB data collection efforts included the following activities:

- **District survey.** The survey was designed to locate key sources of test data. In addition, copies of the tests were requested from the smaller districts. Imperial, Glenn, Lassen, North Coast Unified, Kern, and Great Basin mailed copies of their test reports; and
- **District data collection.** Teams were sent to the major districts identified through the survey. In a similar program, the EPA also was collecting air toxics test data at districts throughout the State. The EPA sent teams to Ventura, Santa Barbara, Sacramento, and the Bay Area. EER sent teams to San Joaquin, South Coast, San Diego, Mojave, Monterey, Mendocino, Colusa, Yolo, and Placer.

Through these activities, most of the test data available in the State has been collected. Table 1 lists the number of tests collected at each district. 799 tests were collected at 19 out of 31 districts. A test includes the quantification of air toxics and other emissions from a device or group of interconnected devices operating under one condition. A report provides the results of one or more tests.

3.0 Screening

One of the key goals of this project was to develop emission factors which can be used to accurately assess air toxics emissions from key sources. Before developing emission factors it is critical to eliminate test results of unknown accuracy and device types which are not of primary interest. This allows for the best use of project resources. This section describes the procedures used to screen tests. Results from the initial screening also are provided. It should be noted that the initial screening of petroleum sources was conducted by CARB.

3.1 Procedures

Table 2a-c lists the data types extracted from each test report for the screening analysis. The information is divided into three categories including:

- **Report Information (Table 2a).** This information describes the complete report.
- **Device Information (Table 2b).** This information describes each device type tested. Each report can have multiple devices. For example, a report may include results for a boiler test and an IC engine test.
- **Substance Information (Table 2c).** This information describes measurements conducted on the fuel and air emissions for each device. For example, a boiler test may have included metals analysis of the fuel and PAH analysis of the stack emissions.

3.2 Results

All the information described in Section 3.1 has been extracted from each petroleum industry test report collected by CARB. Section 2 of Volume 3 lists the device type, test documentation information available, and substances quantified for the fuel and air emissions.

Tables 3a and b summarize the test documentation information available and substances quantified for several device/process and material categories. Table 3a lists the percent of tests in each device/process and material category which provided device, QA/QC and method descriptions, sample, lab, calibration, location and blank data, and included enough information to develop emission factors. The devices and processes listed as required are those given in AB 2588 appendix D. Appendix D specifies source testing requirements for key device and process types. The greatest number of required tests were conducted on catalytic crackers. It should be noted that catalytic reformers were included with the catalytic crackers. Of the 9 tests for this category, 22% provided enough information to develop emission factors as shown in Table 3a. The fuel or raw material feed rate or production rate and emission rate are required to calculate an emission factor. The main substances quantified for catalytic crackers are metals, formaldehyde, and benzene as shown in Table 3b.

Tables 3a and b also list several device and process types for which source testing is not required. Testing may have been conducted on these sources because of district requirements, a lack of accurate estimation techniques, or other facility specific concerns. Of the 161 tests evaluated, 134 are non-required tests. Gas fired internal combustion engines have the most source test data of any non-required category. Formaldehyde was quantified in 29% of the internal combustion engine tests.

3.3 Source Prioritization

To develop emission factors of known quality requires a detailed evaluation of each test collected. To conduct an effective evaluation, the test should include device and method descriptions, and sample, laboratory, QA/QC, calibration, sample location, and blank data. In addition, the test must provide representative process rate information so that emission rates can be normalized to develop emission factors. To evaluate the quality of the collected tests, seven queries or search cases were conducted as shown in Table 4. Search cases 1-3 were for required tests and search cases 4-6 were for non-required tests. Search cases 1-6 required that process rates be available. Tests with process rates can be used to develop emission factors. Search case 7 included tests without process rates. Search cases 1 and 4 required that all supporting information be provided. These tests have the information necessary to develop emission factors of known quality. Search cases 2 and 5 did not require that QA/QC, calibration, or location data be provided. In general, this information is not of primary concern when assessing the accuracy of emission data. Search cases 3, 6, and 7 did not require any supporting information for the emission results. Test data matching these conditions cannot be effectively evaluated for quality. Table 4 shows that the number of tests matching the search cases ranged from 2 (search case 1) to 60 (search case 6). Overall search cases 1, 2, 4 and 5 would provide the best pool of data for emission factor development. In general, search cases 3 and 6 do not provide enough information to validate the results.

As mentioned previously, 161 petroleum tests were collected by CARB. Each of these tests is listed in Section 2 of Volume 3. Search case 7 tests cannot be used to develop emission factors and will not be considered for this project. In addition, fugitive sources such as bulk terminals were not considered because the focus of this project is on combustion sources. Several

data sets also were collected which did not provide any useful air toxic emissions and thus were eliminated from further consideration. Table 5 lists the final set of sources which be validated in detail and included in the emission factor calculations. It should be noted that some of these tests may be eliminated based on the detailed validation results. As shown in the Search Case column of Table 5, some search case 3 and 6 sources were selected. In general, these sources were selected to increase the size of the sample for emission factor development or to ensure that a wide range of source types were represented. Search case 3 and 6 tests do not provide all of the information necessary to evaluate data quality. Table 5 also lists supporting information including Number of Tests, Control Device Used, Report Information, Report Year, Sample Location, and Substance Group as described in Tables 2a-c. In the Report Information columns, a yes (Y) or no (N) is provided if the report provides the specific type of information requested such as Device Description (DD). In the Substance Group columns, a yes (Y) is provided if substances in the group were quantified. For example, if Arsenic and Chromium were quantified in a test, a Y would be listed in the Metals (M) Substance Group column. A blank in the substance group column indicates no substances in the group were quantified.

Table 6 lists tests which will not be included in the emission factor development process. These tests were eliminated for one of the following reasons: emission factors cannot be developed (search case 7); not combustion sources; or no air toxics information.

4.0 Detailed Validation

The detailed validation procedures include checking to ensure the correct sampling and analysis procedures were used, qualifying significant problems such as high field blanks, checking calculations, and evaluating the accuracy of the test results. All methods needed to quantify the substances listed in AB2588 appendix D have been reviewed including:

CARB Methods Reviewed

- 11 - Hydrogen Sulfide (1983)
- 12 - Inorganic Lead (March, 1986)
- 15 - Hydrogen Sulfide (June, 1983)
- 101A - Mercury (1986)
- 104 - Beryllium (1986)
- 106 - Vinyl Chloride (June, 1983)
- 410A/B - Benzene (March, 1986)
- 421 - Hydrogen Chloride (January, 1987 and December, 1991)
- 422 - Volatile Halogenated Organics (January, 1987 and December, 1991)
- 423 - Inorganic Arsenic (January, 1987)
- 424 - Cadmium (1987)
- 425 - Total and Hexavalent Chromium (January, 1987 and September, 1990)
- 428 - PCDD/PCDF and PCB (March, 1988 and September, 1990)
- 429 - PAH (September, 1989)
- 430 - Aldehydes (September, 1989 and December, 1991)
- 433 - Nickel (1989)
- 436 - Trace Metals (March, 1991 and 1992)
- EPA MMT - Trace Metals (1989)

Section 3 of Volume 3 provides validation procedures for each method. These procedures were developed using experience gained conducting air toxic source tests, and reviewing AB2588 test reports, EPA and CARB test method documentation and CARB method review sheets.

Primary parameters were identified to ensure critical data quality indicators were checked. The primary parameters provide an overall assessment of data quality but may not provide an indication of why a particular problem occurred. For example, if a method required field, reagent, and method blanks, only the field blank was considered a primary parameter because it indicates the total interference and/or contamination resulting from the field and laboratory activities. However, the field blank does not indicate if the contamination resulted from the field and/or laboratory activities. For this project, it was more important to evaluate the overall quality of the emissions data.

Only those parameters provided in the test reports in the form required by the method were checked. For example, if the method required that field blank levels over 20% be flagged, the flags were transferred from the test report to the emission factor database. However, if the field blank levels were reported but not divided by the sample value, the ratios were not calculated. Instead, a notation was made to indicate that field blanks were collected and analyzed but the results were not flagged appropriately. The only exceptions to this rule were for CARB Methods 430 and 436. For these two methods, field blank ratios were calculated because they were rarely reported by the contractors.

All versions of the methods effective during the AB2588 program have been reviewed. This allowed review of source test data which was gathered using old test methods. Review sheets were not prepared for the fuel analyses because these results were not extracted from the reports. Fuel analyses do not provide a direct measurement of emissions from a source but are used instead to estimate emissions. The emission factors developed during this program had to provide accurate assessments of emissions.

4.1 Detailed Validation Results

Summary

- 72 tests were selected for detailed data validation and extraction.
- Of the 72 tests, 33 “old” tests evaluated in a previous project funded by the CARB to develop emission factors.
- 2 of the 39 “new” tests were eliminated from the detailed review process.
- 70 tests had data validated and extracted.

This section presents the results of the method validation and calculation checks discussed above for the 72 tests selected for air toxic emission factor development. A test includes the quantification of air toxics and other emissions from a device or group of interconnected devices operating under one condition. Specifically, this section first chronicles the major problems associated with the data validation and extraction procedures and then discusses the details of the method validation results.

Overall, of the 39 “new” tests that passed the initial screening and were selected to undergo the method validation and calculation check procedures, 2 tests were eliminated from the process. One test was excluded from the detailed validation screening because emission factors could not be calculated from the reported data. Originally, this test passed the initial screening because at first glance the test appeared to report all the necessary data to develop a validated emission factor, but a more detailed review of the reports revealed data deficiencies. The other test was dropped because it was determined the device tested was not a petroleum device. Considering the 2 tests that were dismissed from the original group of 72, a total of 70 tests (33 “old” tests and 37 “new”) had data validated, checked, and extracted.

4.1.1 Validation Problems and Calculation Check Failures

- Of all tests validated, the most common problem was the lack of a full set of internal standards used during Method 429 (PAH) analyses.
- Of 5 calculation check failures documented, all of them were Method 430 (HCHO) failures.

Major problems encountered during the method validation and extraction procedures are documented in Table 7. The table lists all 72 tests according to Report ID, Device ID, Number of Tests, Contractor ID, Device Type, Material Used, Review Date, Comment, and Calculation Check Status. These criteria are discussed in detail below:

- **Report ID:** This is the number that was assigned to a device or similar group of devices in each document during the initial screening phase. Similar devices all have the same primary characteristics such as an internal combustion engine. The report ID is a four digit number followed by a letter. The four digit number distinguishes different documents. A unique letter is assigned to each device or group of devices in a document. If, for example, a document contained results for two boilers and an internal combustion engine, the devices would be given the same four digit number (####), but each would have its own letter identifier (e.g., ####A for the two boilers and ####B for the ICE).
- **Device ID:** This three digit number is assigned to each device or group of interconnected devices upon entry into the database. Some facilities have a group of devices which emit to a common stack. For example, a facility may have six steam generators exhausting to one stack. These six steam generators would receive a single device ID. Each engineer entering data had his/her own assigned set of device ID numbers so the person responsible for validating and extracting the results from a particular test could be tracked. In many cases, the report ID and device ID can be used to reference a device or group of interconnected devices. In some cases, however, a report ID references multiple devices. For example, Report ID 2409A references 14 devices (Device IDs 114 to 127) as shown in Table 7.
- **Number of Tests:** As mentioned earlier, a test includes the quantification of air toxics and other emissions from a device or group of interconnected devices operating under one condition. A condition is defined as a set of operating constraints which are fixed during a test. For example, one condition would be a boiler fired on natural gas under normal load. Another condition might be the same boiler fired on fuel oil under normal load. In this case, a single device ID would be assigned and 2 tests would be listed in Table 7.
- **Contractor ID:** This field lists the ID of the contractor responsible for conducting the tests. The name of the contractor is not provided to protect the source of the data.
- **Device Type:** This field displays the type of device whose emissions are characterized in the test.
- **Material Used:** This field describes the materials consumed by each device or group of interconnected devices during all tests. In other words, if the device is a combustion source, this field lists the fuels that were fired during each test. If the device is not a non-combustion source, this field lists the materials processed by the device.

- **Review Date:** This is the date when the method validation and extraction were completed.
- **Comment:** This field lists the problem number. If a significant problem was discovered while validating a test, a problem number was placed in the "Comment" field. If the problem was new and had not already been encountered, then a new number was added to the list. This was continued until all major problems had been documented. There were a total of 5 major problems encountered during the data validation and extraction procedures.
- **Calculation Check Status:** This field describes each test's performance during the calculation checks. This field lists four codes: F - Failed calculation check; P - Passed calculation check; NR - Calculation check not required because calculations were checked for another device in the report; and, ND - Not enough data to check calculations. If a test failed a calculation check, the "F" is followed by the method number that failed the check. A "P" is listed only if all calculation checks passed. The ND code means one of two things, either there is enough data in the test report to conduct a calculation check or there is not enough information in the test report to verify whether the checked passed or failed. Lastly, an "NR" implies that there was more than one device in the test report. In such cases, it was assumed that the contractor conducted all calculations in the same fashion and to perform calculation checks for the same method on all devices would be redundant.

Listed below is an itemized summary of Table 7. It lists all major problems (#1-5) and which reports experienced them. Also, a brief explanation of EER's course of action following the problem is given. Problems associated with calculation checks are also discussed. It should be noted that each line in Table 7 describes a device, group or interconnected devices, or group of similar devices. If a device ID is given, the line describes a device or group of interconnected devices.

- (1) An emission factor cannot be calculated for reports 2599M and 2511A. The results were not extracted or validated.
- (2) Dioxins/Furans and PAH were sampled using a single train for reports 2484A and 2541A. CARB approved the use combined sample train results for emission factor development in the previous project.
- (3) Naphthalene was sampled using Method 410 for report 2409A. Since naphthalene was quantified using an incorrect method, these results were not validated or extracted.
- (4) Method 429 sampling was not isokinetic for reports 2375B and 2375C. The sampling methodology for these tests was modified for non-isokinetic testing by eliminating the glass nozzle and probe from the sampling train. The diesel IC engines were sampled using a 3/8 inch diameter glass probe placed in the center of the exhaust stack. The glass probe was connected directly to a Teflon sample line. No mention of CARB approval was given for these modification. The results were validated, extracted, and noted in the database.
- (5) A full set of Method 429 internal standard recoveries was not reported for reports 2599L, 2599A, 2599N, 2599B, 2599C, 2599D, 2599E, 2599J, 2599G, 2599H, 2599I, 2513C, 2513A, 2598A, 2375B, and 2375C. CARB Method 429 requires the spiking of 14 internal standards into each sample. For no known reason, these tests did not spike all 14 standards. Instead, most spiked and reported recoveries for about half of them. The results were validated and extracted and noted in the database.

(F) Reports 2460B (Method 430) and 2007A (430) failed calculation checks.

Method 430: The two reports which failed Method 430 did not provide the reporting limit when the sample to blank ratio was less than 5. The results for these reports were not corrected.

4.1.2 Detailed Validation Flags

Summary

- The majority of detailed flags complied denote poor procedures by the contractors, not actual parameter failures.

After all the test results were validated and extracted, the method validation sheets were compiled and the validation flags were entered into the database. The validation flags were then condensed and exported to a spreadsheet for tabular summation. The results of the detailed data validation are summarized in the tables located in Section 4 of Volume 3. There is one table per method. Each test method has a set of validation parameters that are used to verify proper sampling and analytical procedures. These parameters are organized into sections by type of sampling or analytical check. The sections are in boldface and shaded gray in the tables. A detailed review of a source test report can produce three basic responses with the corresponding flag notations for each validation parameter:

- Pass: A blank cell in the table
- Insufficient information to report a parameter: R, V, N, P, RN, PN
- Fail: Y, RF, PF

Consider Method 428 (1988) as an example. This method is applicable to the determination of Polychlorinated Dibenzo-p-dioxin (PCDD), Polychlorinated Dibenzofuran (PCDF) and Polychlorinated Biphenyl (PCB) emissions from stationary sources. Section 4 of Volume 3 shows a total of 2 stacks (14210, 14610) were sampled and analyzed using this method. The first three digits of the stack ID are the device ID and the last two digits identify each stack on the device. The stack ID is used in Section 4 of Volume 3 because some devices have multiple stacks. Each of these stacks may have been tested and therefore validation was conducted on each stack. "Sample runs < 3" parameter shows a blank cell for each stack. This means that 3 Method 428 runs were conducted at the stack listed. The following is a brief explanation of each notation.

- R This notation is used when it cannot be determined whether the parameter was conducted or not.
- V If values were not provided for a parameter, this notation was used.
- N This notation is used when a parameter was not conducted.
- P This notation is used specifically for the Pitot tube semi-annual calibration sheet parameter.
- RN* Similar to the V notation but applied to more detailed parameters that require run quantification.

PN* Again, similar to the V notation but applied to more detailed parameters that require both run and substance quantification.

Y This notation is used when a parameter was conducted and failed.

RF* Similar to the Y notation but applied to more detailed parameters that require run quantification.

PF* Again, similar to the Y notation but applied to more detailed parameters that require run and substance quantification.

* Numbers before these notations represent how many times a parameter failed or could not be checked.

The following is a list summarizing the validation tables for each method. The list contains those parameters which were flagged for 50% or more of the devices in each table unless noted. By far, the most prevalent type of flags found in the tables are those associated with reporting. Still, there are plenty of failures which are noteworthy, but they are much less frequent. Note, primary validation parameters are underlined and failures are in italics. For more details on specific parameter failures, please refer to the tables in Section 4 of Volume 3.

Method 11 (1983): Single stack tested - see Section 4 of Volume 3

Method 12 (1986): Insufficient reporting - Swirl check
Insufficient reporting - Semi-annual pitot tube calibration
Failure - Field reagent blank not used to correct samples
Insufficient reporting - Atomic absorption spectrometry not used
Insufficient reporting - Atomic absorption spectrometry not conducted in triplicate

Method 15 (1983): No parameters flagged 50% or more

Method 101A (1986): Insufficient reporting - Swirl check
Insufficient reporting - Method 1 not used for sample location
Insufficient reporting - Dry gas meter pre- and post-test check
Insufficient reporting - Semi-annual pitot tube calibration
Insufficient reporting - Filter temperature
Failure - Field reagent blank not used to correct samples
Failure - Combined analysis not used

Method 104 (1986): Insufficient reporting - Swirl check
Insufficient reporting - Nozzle size check
Insufficient reporting - Dry gas meter pre- and post-test check
Insufficient reporting - Semi-annual pitot tube calibration
Insufficient reporting - Field reagent blank not conducted for acetone
Failure - Field reagent blank not used to correct samples

Method 410 (1986): No parameter flagged 50% or more

Method 421 (1987): Single stack tested - see Section 4 of Volume 3

- Method 422 (1987): *Failure - Analysis date not within 24 hours of collection*
- Method 422 (1991): Single stack tested - see Section 4 of Volume 3
- Method 423 (1987): Single stack tested - see Section 4 of Volume 3
- Method 424 (1987): Single stack tested - see Section 4 of Volume 3
- Method 425 (1987): Insufficient reporting - Swirl check
 Insufficient reporting - Nozzle size check
 Insufficient reporting - Dry gas meter pre- and post-test check
 Insufficient reporting - Semi-annual pitot tube calibration
 Insufficient reporting - Matrix spike not conducted once per test (Hexavalent Chromium)
 Insufficient reporting - Matrix spike percent recovery (Hexavalent Chromium)
 Insufficient reporting - Duplicates not conducted for every 10 samples (Total Chromium)
- Method 425 (1990): Insufficient reporting - Swirl check
 Insufficient reporting - Probe proof not conducted per probe
 Insufficient reporting - Probe proof total chrome > detection limit
 Insufficient reporting - Matrix spike percent recovery (Hexavalent Chromium)
 Insufficient reporting - Matrix spike not conducted daily (Total Chromium)
- Method 428 (1988): Insufficient reporting - Swirl check
 Insufficient reporting - Nozzle size check
 Insufficient reporting - Dry gas meter pre- and post-test check
 Insufficient reporting - Semi-annual pitot tube calibration
 Insufficient reporting - Surrogate standards not conducted one per run
 Insufficient reporting - Surrogate standards percent recovery (37Cl-2,3,7,8-TCDD)
 Insufficient reporting - Surrogate standard percent recovery (13C-1,2,3,4,6,7,8-HpCDF)
 Insufficient reporting - Laboratory control spike percent accuracy
 Insufficient reporting - Internal standards not conducted one per sample
 Insufficient reporting - Internal standards percent recovery
- Method 428 (1990): Single stack tested - see Section 4 of Volume 3
- Method 429 (1989): Insufficient reporting - Swirl check
 Insufficient reporting - Dry gas meter pre- and post-test check
 Insufficient reporting - Semi-annual pitot tube calibration
 Insufficient reporting - Surrogate standards percent recovery
 Insufficient reporting - Laboratory control spike percent accuracy
 Insufficient reporting - Internal standards percent accuracy
 Insufficient reporting - Analysis date

- Method 430 (1989): Insufficient reporting - Pre- and post-test sample flow
 Insufficient reporting - Matrix spike not conducted per test
 Failure - Matrix spike not reported
- Method 430 (1991): Insufficient reporting - Rotometer calibration check
 Insufficient reporting - Pre- and post-test sample flow
 Insufficient reporting - Sampling date not within 2 days of reagent blank check
 Insufficient reporting - Extraction date
 Insufficient reporting - Analysis date
- Method 433 (1989): Insufficient reporting - Swirl check
 Insufficient reporting - Semi-annual pitot tube calibration
 Insufficient reporting - Field reagent blank not conducted for two filters and 0.1N HNO₃
 Failure - Field reagent blank not used to correct sample
 Failure - Atomic absorption spectrometry not used
 Insufficient reporting - Atomic absorption spectrometry not used
 Insufficient reporting - Atomic absorption spectrometry not conducted in triplicate
- EPA MMT (1989): Insufficient reporting - Swirl check
 Insufficient reporting - Nozzle size check
 Insufficient reporting - Dry gas meter pre- and post-test check
 Insufficient reporting - Semi-annual pitot tube calibration
 Failure - Field reagent blank not used to correct samples
 Insufficient reporting - Duplicates percent difference (ICAP)
 Insufficient reporting - Duplicates not conducted per run (GFAAS)
 Insufficient reporting - Duplicates not conducted per run (CVAAS)
 Insufficient reporting - Matrix spikes percent recovery (GFAAS)
 Insufficient reporting - Matrix spikes percent recovery (CVAAS)
- Method 436 (1992): Insufficient reporting - Swirl check
 Insufficient reporting - Dry gas meter pre- and post-test check
 Insufficient reporting - Pre-test leak check
 Insufficient reporting - Post-test leak check
 Failure - Field blank not collected once per test
 Insufficient reporting - Sample/field blank ratio
 Failure - Sample/field blank ratio
 Insufficient reporting - Field reagent blank not used to correct samples
 Insufficient reporting - Analysis dates (ICAP)
 Failure - Analysis dates (CVAAS)
 Insufficient reporting - Duplicates percent difference (ICAP)
 Insufficient reporting - Duplicates not conducted per run (GFAAS)
 Failure - Duplicates not conducted per run (CVAAS)
 Insufficient reporting - Matrix spikes not conducted on front- and back-half (GFAAS)
 Insufficient reporting - Matrix spikes percent recovery (GFAAS)
 Failure - Matrix spike not conducted on one nitric impinger (CVAAS)
 Insufficient reporting - Matrix spike not conducted on one nitric impinger (CVAAS)

Insufficient reporting - Matrix spike percent recovery

Lastly, the tables in Section 4 of Volume 3 do not specify validation results for individual hazardous air pollutants. Consequently, one cannot determine if any one substance failed method parameters more than others by using these tables alone. Section 5 of Volume 3, however, contains a table which provides such information. The table lists how many times a particular substance failed a validation parameter. It includes the Method, Version (Year), Failed Check, Substance, and Count. The table presents results for both isokinetic and non-isokinetic trains.

5.0 Data Extraction

Data extraction is the process of entering design and operating information, and emission results into a database. After the validation activities were completed as described in Section 4.0, the emissions data were extracted. If a critical validation parameter was not satisfied for a method, such as analyzing the front- and back-half component of a CARB 436 train, the emissions data for the method were not extracted. If several methods were not suitable for extraction in a single test report, the complete test report was rejected. If the test report did not provide sufficient information to develop emission factors, it also was rejected. For each test that was not rejected, the following information was extracted.

Device Information

- 1.) Source classification code (SCC)
- 2.) Standard industrial code (SIC)
- 3.) Control device type
- 4.) Fuel type or material processed
- 5.) Capacity
- 6.) Company
- 7.) Location (City, Address, Zip, Facility ID, District Code)
- 8.) Report Date

Sample and Analysis Procedures

- 1.) Sampling method
- 2.) Analysis method
- 3.) Contractor
- 4.) Detection limit based on MDL or PQL

Run Information

- 1.) Process rate and unit (must be appropriate for emission factor development)
- 2.) Site run ID
- 3.) Date of Run
- 4.) Fuel/Material type burned during test
- 5.) Description of operation during test
- 6.) Stack flow rate (dscfm)
- 7.) Stack moisture (%)
- 8.) Stack temperature (F)
- 9.) Stack oxygen (%)

Emission Information

- 1.) Substance
- 2.) Detection indicator (Detected or Not Detected)
- 3.) Data quality flags
- 4.) Concentration value and unit
- 5.) Emission rate value and unit
- 6.) Emission factor value and unit

6.0 Emission Factors

A procedure was developed to provide emission factors of known quality for a wide range of air toxics and source types. This procedure considers the design and operation of the sources, process stream characteristics, data quality, source population size, and emission factor variability. The procedure includes the following steps:

- Identify Design and Operating Parameters
- Identify Normalizing Units
- Assign Run Specific Method Ratings
- Calculate Run Specific Emission Factors
- Identify Major and Sub Group Evaluation Parameters
- Compile Detailed Data Listings
- Conduct Outlier Analysis
- Identify Sub Groups
- Calculate Emission Factors for each Sub Group
- Assign Sub Group Method and Population Ratings
- Assign CARB Overall Quality Rating
- Assign EPA Overall Quality Rating

Each of these steps is discussed below. The discussion includes a brief background which describes why the step is needed and the approach used. The background is followed by a presentation of the results of applying the subject step to the data.

6.1 Design and Operating Parameters.

Background. To develop emission factors, sources must be grouped by their design and operating characteristics. Ideally, emissions from devices in each group should be similar or have low variability when normalized. To define design and operating parameters, a literature review was conducted. AP-42 was one of the best sources of information identified. In addition to AP-42, EER used its experience in past programs such as the WSPA air toxic emission factor project to define design and operating parameters. Table 8 lists key parameters which may affect emissions from external combustion, internal combustion, and gas turbines.

Results. Few reports contained all of the information listed in Table 8. Basic system type, feed material or fuel, and air pollution control system type were available for most sources. Unit capacity, load, and burner type were available for most external combustion sources. If a primary parameter was not described in a report, the appropriate facility was contacted.

6.2 Normalizing Units.

Background. An emission factor characterizes air toxic emissions as a ratio of the amount of pollutant released to a process related parameter. Emission factors are typically expressed in terms of mass of emission per mass or volume of fuel or material fed or product produced. Thus, the emission rate is normalized by the production rate or by the feed rate of fuel or material. This method of normalizing assumes that emissions are directly proportional to production rate or fuel or material feed rate. Based on established procedure, normalizing units were assigned based on the source classification codes (SCC).

Results. The first step in determining the appropriate normalizing units is to assign SCCs.

Using a SCC look-up table from the EPA and descriptions provided in the test reports, SCCs were assigned for each test. Table 9 lists the SCCs assigned. For several tests, no SCC was available which adequately described the test. For example, several of the reciprocating internal combustion engines were fired on field gas. For these sources, a natural gas SCC was used as noted in Table 9. In addition, five steam generators, one heater, and three gas turbines fired natural and process gas simultaneously. SCCs were available for natural gas and process gas separately. However, SCCs were not available for simultaneous burning of natural and process gas. Therefore, dummy SCCs were assigned as shown in Table 9. For the internal combustion engines, steam generators, gas turbines and heaters, additional SCCs must be requested from the EPA. The required normalizing units for each SCC also are provided in Table 9. Emission factors have been expressed in these units.

6.3 Run Specific Method Rating.

Background. To compare and evaluate test results, it is important to denote the test methods that were used and the level of documentation that was provided. Various systems have been developed to categorize test methods and the level of documentation. For example, the EPA developed the system described below.

EPA Method Rating

- A When tests are performed by a sound methodology and are reported in enough detail for adequate validation.
- B When tests are performed by a generally sound methodology but lack enough detail for adequate validation.
- C When tests are based on an untested or new methodology or are lacking a significant amount of background data.
- D When tests are based on a generally unacceptable method but the method may provide an order-of-magnitude value for the source.
- U Unrateable.

The EPA method ratings were used for the Factor Information Retrieval (FIRE) system which includes criteria and air toxic emission factors. For the CARB emission factor project the EPA system was modified to distinguish between EPA and CARB methods as well as tests which provide sufficient documentation and those that do not. The method rating system used for CARB emission factor project is described below.

CARB Method Rating

- A Test was performed using a new or old CARB methodology and sufficient documentation was provided to validate the results.
- B Test was performed using a new or old EPA methodology and sufficient documentation was provided to validate the results.
- C Test was performed using a new or old CARB methodology and insufficient documentation was provided to validate the results.

- D Test was performed using a new or old EPA methodology and insufficient documentation was provided to validate the results.
 - E An assumption was made in the emission factor calculation which could significantly affect the accuracy of the results. Methods which do not have validation check procedures also were rated under this category.
 - F Emission data is unacceptable for inclusion in emission factor database. If a sampling problem or process upset occurred which significantly impacted the emission results, the emission results were excluded from emission factor calculations. A statistical test was used to identify outliers as described in section 6.7.
- ** It should be noted that the EPA methods are not considered inferior. However, the majority of the test methods were CARB's because they were mandated by the Hot Spots program. An EPA method could be used if there was no corresponding CARB test method or if the source asked for an equivalency determination. CARB and EPA test methods are different in many cases and can lead to different results. CARB test methods were rated higher than EPA's to provide consistent test result comparisons.

A test received an A or B (C or D) rating if a specified number of primary validation parameters could be checked. Primary validation parameters are those that indicate overall contamination, poor recovery, and imprecision. Primary parameters are identified in Table 10 for CARB and EPA methods. The table also provides the number of primary parameters per method and the number of primary parameters required to determine if sufficient documentation was provided. The number of parameters required to determine if sufficient documentation was provided was based on the following criteria.

Primary Parameters	Sufficient Documentation if
0 to 2	0 Missing
3 to 4	1 Missing
5 to 6	2 or fewer Missing

Results. Table 11 summaries the method ratings by method. The table lists the number of times the method was used for the tests extracted and validated during this project. Several tests received E ratings because no validation check procedures have been developed. Data from two tests were assigned F ratings because they were identified as outliers as discussed in Section 6.7.

6.4 Run Specific Emission Factor Calculation.

Background. A source test usually includes three runs per sample method. Emission factors must be calculated for each of these runs. Once appropriate groups have been defined (see Sections 6.5 and 6.8), run emission factors from one or more source tests are averaged together. In general, emission factors are calculated on a run basis using feed or production rates and air toxic emission rates. For combustion sources, when feed rates are not available, F-Factors can be used in combination with the stack oxygen and air toxic emission concentration using the following equation.

$$EF = FF \cdot HV \cdot (21 / (21 - O_2)) \cdot C$$

For Gas Fired:

EF = Emission Factor, lb/MMcf
FF = Fuel F-Factor, dscf @ No Excess Air/MMBtu
HV = Fuel Heating Value, MMBtu/MMcf
O₂ = Stack Oxygen
C = Substance Concentration, lb/dscf @ Stack Oxygen

For Liquid Fired:

EF = Emission Factor, lb/Mgal
FF = Fuel F-Factor, dscf @ No Excess Air/MMBtu
HV = Fuel Heating Value, MMBtu/Mgal
O₂ = Stack Oxygen
C = Substance Concentration, lb/dscf @ Stack Oxygen

For Solid Fired:

EF = Emission Factor, lb/ton
FF = Fuel F-Factor, dscf @ No Excess Air/MMBtu
HV = Fuel Heating Value, MMBtu/ton
O₂ = Stack Oxygen
C = Substance Concentration, lb/dscf @ Stack Oxygen

Results. Run specific emission factors were calculated for each source in the appropriate unit (see Table 9).

6.5 Major and Sub Group Evaluation Parameters

Background. A key step in the emission factor development process is to identify devices which have similar designs and operation. The design and operating parameters selected to categorize the devices should impact air toxic emissions. If the parameters are defined appropriately and correct normalizing units are assigned, emission factors developed for each group of devices will be distinctive and will have low variability. These emission factors can be used to accurately assess emissions from similar devices. The first step in the categorization process is to divide the sources into major groups based on their primary design characteristics. Primary design characteristics are those parameters which are known to impact emissions such as basic system and feed type. For this study, emission data from different major groups were not combined when calculating emission factors.

The second step in the grouping process is to identify if any sub groups are present within each major group. Sub group identification is based on an evaluation of secondary design and operating parameters. Before sub groups can be established, secondary design and operating parameters must be identified, detailed data listings must be prepared, and outliers must be identified and eliminated from the analysis if sampling problems occurred. In addition, guidelines and statistical tests should be established to determine if sub groups are needed and appropriate. Secondary design and operating parameters are discussed and listed in this section. Detailed data listings are described in Section 6.6. The outlier analysis is discussed in Section 6.7 and guidelines for establishing sub groups are presented in Section 6.8.

Results. Major groups have been identified for all devices as shown in Table 12. The first column of this table describes each major group. Horizontal lines are used in Table 12 to segregate the sources contained in each major group. For example, the "Boiler, Ref. Gas" major group contains 5 tests.

Secondary parameters which were considered when developing sub groups are listed in the table. For the major group "Boiler, Ref. Gas", the impact of air pollution control (APC) system, load, and burner type was investigated. NC indicates no comparison was made. For example, the only SCC for the major group "Boiler, Ref. Gas" is 10200701, therefore no comparison of SCC was made.

It should also be noted that the oxygen level was examined for all combustion sources. Potential divisions based on an analysis of the data have been listed in Table 12. For the gas fired internal combustion engines, lean refers to oxygen levels greater than 2%. Due to the limited operating data present in the reports, the oxygen level compared in all cases was the stack oxygen level. If a source has significant air dilution after the combustion chamber, the stack oxygen will not reflect the true combustion environment.

6.6 Detailed Data Listings.

Background. To investigate the impact of secondary design and operating parameters and evaluate outliers, lists of emission factors, design and operating parameters and data quality parameters must be compiled. These lists are used to identify trends and as inputs to the statistical evaluation procedures.

Results. The comparison parameters listed in Table 12 and normalized emissions data were compiled into 9 tables, one for each device type. Each of these tables was sorted by major group, category, substance, and normalized emissions. The number and type of design and operating parameters listed depended on the device type. For example, Fuel type, SCC, strokes per cycle, capacity, APC system, stack oxygen, and stack flow were listed for internal combustion engines. Because the detailed data listings contain confidential information they have not been provided in this report.

6.7 Outlier Analysis

Background. Before establishing subgroups, outliers must be identified and evaluated. If an outlier results from a calculation or data entry error, it can be corrected. Outliers resulting from sampling and analysis problems may be eliminated from data analysis activities. There are many approaches for identifying and handling outliers. For this study, the outlier analysis was conducted in two steps as indicated below.

- i.) Conduct an outlier analysis per substance per test and per substance per major group. The Dixon test was used to identify outliers per substance per test and per substance per major group. To use the Dixon test, a group of data is selected and sorted from lowest to highest emissions. Then the high and low points are examined statistically in relation to the other points in the data set. The test will identify if the high and low points are outliers at a prescribed level of confidence. For this study the confidence level was 95%. It should be noted that when applying the Dixon test to samples with three points many outliers are identified where two of the points in the data set have approximately equal values and the third point is slightly higher. This commonly occurs when two points are not

detected and the third point is detected. For this analysis, if the other two points in a data set had similar values and the outlier was within _ order of magnitude of their value, no other checks or action was taken. These values were accepted as being within the expected precision of the test method.

- ii.) Evaluate outlier points identified in Step I to determine if sampling problems, calculation errors, and process upsets occurred. Outliers with calculation errors were corrected and outliers with sampling problems were assigned a method rating of F for unacceptable. Emissions with method ratings of F were not used to develop emission factors. Outliers were not eliminated unless a sampling or process problem occurred. A major component of the outlier evaluation are problems found during the detailed validation of the test reports. These problems are indicated by the flags described in Section 4 of Volume 3.

Results. Table 13 summarizes the results of the outlier analysis by device and substance. Each data point in the table is an outlier. An "x" in the calculation, process or method column indicates that no problems of that type could be found in the test report or database. A "0" indicates that a problem was located and some action was taken. The problem and action are described in the comment column. 22 outliers were identified in the analysis by device and substance. These outliers were assigned an F method rating and excluded from the emission factor calculations.

The results of the outlier analysis by major group and substance are given in Table 14. The outliers in this analysis were due to suspect data resulting from apparent contamination of samples. A total of 4 outliers were identified in the analysis by major group and substance. These outliers will receive an F method rating and will be excluded from the emission factor calculations.

6.8 Sub Group Evaluation

Background. Subgroups may be developed for major groups with two or more sources. Major groups are discussed and identified in Section 6.5. As the number of sources increases, the potential for subgroup development also increases. Subgroups are developed when a secondary design or operating parameter is identified which impacts emissions. Engineering judgement and statistical analysis can be used to determine if the secondary parameters have a significant impact on emissions. If a secondary parameter does impact emissions, subgroups are established resulting in lower emission variability than present across the major group. If the statistical analysis contradicts commonly accepted knowledge about emission behavior, subgroups should not be developed.

In cases where a secondary parameter impacts one substance but not another, the data for both substances could be segregated into different subgroups. Another approach would be to segregate the data for the substance which was impacted by the parameter into different subgroups and combine the data for the substance which was not impacted. This approach can generate a large number of subgroups with high variability. To reduce the number of subgroups and the variability of emissions data in each subgroup, subgroups were identified in this project using the following two step process.

- I.) Identify which secondary parameters (comparison parameters) identified in Table 12 impact the emissions data by reviewing data listings and using the t-Test. The t-Test uses the t distribution to determine if two samples are from the same population when the variances are unknown but equal. The test is applicable to

samples containing less than 30 data points. A sample is a group of data with a distinct value or range of values of the secondary parameter considered. If the t-Test indicates that two samples are not from the same population, the secondary parameter which the samples were grouped by has a significant impact on emissions. It should be noted that the t-Test was only used to support the development of subgroups. In no case was the t-Test used to blindly develop subgroups. Before developing subgroups, the results of the t-Test were examined to ensure they were reasonable based on engineering judgement.

- ii.) Segregate tests in each major group into subgroups based on those secondary parameters identified in Step I which impact the emissions data. Results from one device were not split into different subgroups. This approach is appropriate when a substance is impacted by the secondary parameter and when it is not impacted.

Results. The impact of the secondary parameters identified in Table 12 is quantified statistically in Section 6 of Volume 3. Section 6 of Volume 3 includes a series of similar table containing the results of the t-Test evaluation of each comparison listed in Table 12. The tables are listed below.

Volume 3 Section 6 Statistical Analysis Reference Tables:

Table 6-1.	Source Classification Code comparison
Table 6-2.	Air pollution control system comparison
Table 6-3.	Internal combustion engine strokes per cycle comparison
Table 6-4.	Internal combustion engine oxygen comparison
Table 6-5.	Internal combustion engine capacity comparison
Table 6-6.	Internal combustion engine strokes per cycle and oxygen comparison
Table 6-7.	Internal combustion engine strokes per cycle and capacity comparison
Table 6-8.	Internal combustion engine strokes per cycle, oxygen and capacity
Table 6-9.	Gas turbine duct burner comparison
Table 6-10.	External combustion source load comparison
Table 6-11.	External combustion source burner type comparison
Table 6-12.	External combustion source load and burner comparison

Each of these tables includes a description of the data sets being compared, number of points, average, standard deviation, and detection limit ratio (Detect Ratio). The detect ratio is the ratio of the sum of detected values to the sum of detected and non-detected values. A detect ratio of one indicates all of the data was detected. A detect ratio of zero indicates all of the data was not detected. If the difference between the data sets is significant based on the t-Test, a "Yes" is provided in the last column of the table. If the comparison includes more than three runs (two tests or more), the significantly higher data set is shaded. For comparisons where the difference is not significant or the difference is significant but contains three or fewer runs, the higher data set is underlined. It should be noted that if the difference is significant but none of the data was detected, no shading is provided. If the lower data set includes all non-detect data and the higher data set includes detected data, then the higher data set is shaded provided each data set in the comparison includes more than three runs. If the sample sizes are too small, one run only, for statistical comparison, an "NA" is given in the last column.

No Subgroup Development Possible

Major Group	Reason
*Asphalt Blowing - Blow Cycle	Single Source

*Asphalt Blowing - No Blow Cycle	Single Source
*Coke Calcining	Single Source
*Heater - Natural Gas	Single Source
*Heater - Natural/Refinery Gas	Single Source
*Heater - Oil	Single Source
*Reboiler - Ethylene Glycol	Single Source
*Reboiler - Triethylene Glycol	Single Source
*Turbine - Natural/Refinery/Liquid Petroleum Gas	Single Source
*Turbine - Refinery Gas	Single Source

For the remaining major groups, subgroup development is possible. Each section below provides a brief description of the subgroup analysis for the remaining major groups. A complete list of final subgroups is provided in Table 15. Subgroups are compared in Table 16.

6.8.1 Boiler - Fuel Oil

Number of Tests - 2
Significant Parameters - None
Subgroups - None

Comment - There is no observed difference in design and operation of the two units tested so no subgroups were developed.

6.8.2 Heater - Refinery Gas

Number of Tests - 16
Significant Parameters - None
Rejected Tests - PAH data for Reports 2599A, B, C, D, N
Subgroups - None

Comments - Based on a review of the PAH data it was found that PAH data from tests in document 2599 are two to three orders of magnitude higher than the other tests. PAH data from Report 2599 are significantly higher. Several reasons for the difference include: LRMS was used yielding high detection limits, none of the required internal standards were used in the analytical procedures, and high levels of contaminants were found in many samples. The omission of the required internal standards is major failure and can impact the emission results significantly. Because one of the objectives of this project is to develop accurate emission factors, PAH data from document 2599 were eliminated from the emission factor development procedures. Similar reasoning was used to eliminate PAH data from asphalt production devices in the project titled "Development of Toxics Emission Factors from Source Test Data Collected under the Air Toxics Hot Spots Program" (see Final Report prepared for California Air Resources Board dated April 11, 1996).

The statistical analysis of heaters included a comparison of post combustion air pollution control devices, load, burner type, and load/burner type. Statistical analyses for these parameters are provided in Tables 6-2, 6-10, 6-11, and 6-12 of Volume 3. Observations from the comparisons are provided below.

- The comparison of post combustion air pollution control devices indicated several significant differences. It is likely that these differences were a result of the small sample sizes compared.

- Several substances from low load sources do have significantly higher emissions. However, several substances from high load sources also have significantly higher emissions. This inconsistent impact of load indicates another parameter is probably responsible for the difference.
- Only one substance (Chrysene) in the CB/LNB burner comparison indicates a significant difference. This comparison is based on a small data set. If burner type was a key parameter it should impact the emissions of other PAH substances significantly.
- When grouping by burner type and load the number of significant differences increase but the number of sources also drops. Most significant differences are based on single source comparisons. In particular, one source with low load and CBs accounts for most of the instances where PAH emissions are significantly higher. For several PAH substances additional data for low load/ CBs is available. In these instances, the low load/CB sources do not have significantly higher emissions. This indicates the single source may have design/operating difference besides load and CBs which account for the significant difference.

Based on the above review all of the heater data were lumped into a single source category with no subgroups. Most significant differences are based on single test comparisons. In addition, these significant differences are not consistent from one HAP to the next. For example for acenaphthene low load sources to have higher emissions while for benzo (a) pyrene high load sources appear to have higher emissions. The lack of strong trends in the field data and results from recent laboratory evaluations which indicate no difference in emissions from conventional and low Nox burners operated under a range of typical refinery conditions (PERF project) support the lumping of the field data into a single group. In addition, a more robust emission factor can be developed by combining all data into a single group especially when the variability cannot be effectively explained by observed differences in operation.

6.8.3 Boiler - Refinery Gas

Number of Tests - 5

Significant Parameters - None

Rejected Tests - PAH data for Reports 2599B, E, G, H

Subgroups - None

Comments - The PAH data for reports 2599B, E, G, and H were eliminated for the reasons described in Section 6.8.2. The statistical analysis of boilers included a comparison of post combustion air pollution control devices, load, burner type, and load/burner type. Statistical analyses for these parameters are provided in Tables 6-2, 6-10, 6-11, and 6-12 of Volume 3. Observations from the comparisons are provided below.

- The comparison of post combustion air pollution control devices indicated no significant differences.
- Emissions of phenol are significantly higher for high load sources. This observation is based on the comparison of a single high load test to several low load tests.
- The CB/LNB burner comparison does not indicate any significant differences.

- When grouping by burner type and load the number of significant differences goes up but the number of sources also drops. All significant differences are based on single source comparisons.

Based on the above review all of the boiler data were lumped into a single source category with no subgroups. All significant differences are based on single test comparisons. In addition, these significant differences are not consistent from one HAP to the next. The lack of strong trends in the field data and results from recent laboratory evaluations which indicate no difference in emissions from conventional and low Nox burners operated under a range of typical refinery conditions (PERF project) support the lumping of the field data into a single group. In addition, a more robust emission factor can be developed by combining all data into a single group especially when the variability cannot be effectively explained by observed differences in operation.

6.8.4 Internal Combustion Engine - Diesel

Number of Tests - 3

Significant Parameters - Oxygen Level ($> 13\%$ or $< 13\%$)

Subgroups - 2

Comments - The statistical analysis of diesel internal combustion engines included a comparison of stack oxygen, size, and oxygen/size. Statistical analyses for these parameters are provided in Tables 6-4, 6-5, and 6-8 of Volume 3. In all of the comparisons conducted, one of the data sets compared included results from a single source test. In general, these comparisons are not considered reliable. However, when the differences are large and supported by process knowledge subgroups can be developed. In particular, it was found that the source with stack oxygen level $\leq 13\%$ has higher PAH emissions than the sources with a stack oxygen level $> 13\%$ as shown in Table 6-4 in Section 6 of Volume 3. This is not the case for formaldehyde and acetaldehyde emissions. These substances have higher emissions at higher oxygen levels. While the impact of oxygen on emissions is in general based on small data sets, the behavior is supported by theory and subgroups were developed.

6.8.5 Internal Combustion Engine - Field Gas

Number of Tests - 6

Significant Parameters - Strokes (2 or 4)

- Oxygen Level (Lean or Rich)

Subgroups - 3

Comments - The statistical analysis of field gas internal combustion engines included a comparison of strokes per cycle, stack oxygen, size, strokes per cycle/oxygen, strokes per cycle/size, and strokes per cycle/oxygen/size. Statistical analyses for these parameters are provided in Tables 6-3, 6-4, 6-5, 6-6, 6-7 and 6-8 of Volume 3. All 2 stroke engines have higher emissions than the 4 stroke engines as shown in Table 6-3 of Section 6 of Volume 3. It should also be noted that all of the 2 stroke engines also have high oxygen levels. For the 4 stroke engines, the data indicated that sources with oxygen levels of less than 2% have higher BTX emissions than sources with oxygen levels greater than 2% as shown in Table 6-6 of Section 6 of Volume 3. Sources with oxygen levels greater than 2% had higher formaldehyde emissions. Sources with oxygen levels greater than 2% are characterized as lean burning and sources with oxygen levels less than 2% are characterized as rich burning. An examination of capacity indicated no significant differences as shown in Table 6-5 of Section 6 of Volume 3. When strokes per cycle, and rich or lean combustion are used to develop subgroups, the following subgroups are

obtained as shown in Table 15.

- 1 - 2 stroke/lean
- 2 - 4 stroke/lean
- 3 - 4 stroke/rich

6.8.6 Internal Combustion Engine - Natural Gas

Number of Tests - 13

Significant Parameters - Oxygen Level (Lean or Rich)

Rejected Tests - 4 tests rejected from report 2460B

Subgroups - 2

Comments - Report 2460B included results from four units fired on natural gas. Results from this report were rejected because the strokes per cycle could not be verified. In addition, the formaldehyde results were much lower than all other internal combustion engines. After reviewing the report it was found that the results were reported at levels well below those found in the blank. In these cases the reporting limit should be provided. Insufficient information was available to determine the reporting limit so the results were eliminated.

All of the remaining tests are for 4 stroke engines with capacities less than 650 hp. The statistical analysis for lean and rich engines clearly shows that rich burning engines have higher BTX emissions as shown in Table 6-4 of Section 6 of Volume 3. Lean burn engines have higher formaldehyde emissions. Based on these observations, the following subgroups were established as shown in Table 15.

- 1 - 4 stroke/lean
- 2 - 4 stroke/rich

6.8.7 Steam Generator - Crude Oil

Number of Tests - 3

Significant Parameters - None

Subgroups - None

Comments - The statistical analysis of crude oil fired steam generators included a comparison of post combustion air pollution control devices, load, burner type, and load/burner type. Statistical analyses for these parameters are provided in Tables 6-2, 6-10, 6-11, and 6-12 of Volume 3. Observations from the comparisons are provided below.

- The comparison of devices without control to devices with SO₂ control indicated lead was significantly higher for devices without SO₂ control. This result is likely a result of another parameter and is based on a comparison of small data sets.
- Low load sources do have significantly higher metal emissions. This statement is based on a comparison of small data sets. It is likely that another parameter such as feed metal content is responsible for the observed difference. Organic emissions from high load sources are generally higher. It should be noted that the difference in load is small between the low load (75%) and high load (86%). It is not likely that this small change in load would account for the change in emissions.