

**Recommended Procedures for
Development of Emissions Factors and
Use of the WebFIRE Emissions Factor Database**

Office of Air Quality Planning and Standards
Office of Air and Radiation
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

December 17, 2010

Disclaimer

This report has been approved for publication by the Office of Air Quality Planning and Standards (OAQPS), U.S. Environmental Protection Agency. Mention of trade names or commercial products in this document does not constitute endorsement by the Agency.

EPA-454/R-XX-XXX

DRAFT

TABLE OF CONTENTS

Section		Page No.
Section 1.0	What is the Purpose of This Document?.....	1-1
Section 2.0	What is an Emissions Factor?.....	2-1
2.1	Emissions Data.....	2-2
2.2	Activity Data.....	2-3
Section 3.0	How Have We Historically Developed Emissions Factors?.....	3-1
Section 4.0	How are Emissions Factors Used?.....	4-1
Section 5.0	What are EPA’s Revised Procedures for Developing Emissions Factors?.....	5-1
5.1	Data Collection.....	5-1
5.2	Test Data Evaluation.....	5-3
5.3	Detection Limit Procedures.....	5-4
5.4	Identification of Outlier Data.....	5-5
5.5	Emissions Factor Derivation and Quality Assessment.....	5-5
Section 6.0	EPA’s Interactive Database for the Emissions Factors Program – What is WebFIRE?.....	6-1
6.1	What is WebFIRE?.....	6-1
6.2	How is WebFIRE Used?.....	6-1
6.3	Who Uses WebFIRE?.....	6-4
6.4	How Does WebFIRE Improve Emissions Factor Identification and Development?.....	6-4
Section 7.0	How Do I Find an Emissions Factor?.....	7-1
7.1	How Do I Identify and Retrieve an Emissions Factor from WebFIRE?.....	7-1
7.2	How Do I Obtain Background Information for My Selected Emissions Factor?.....	7-5
7.3	How Do I Identify the Data That Are Used to Derive the EPA-Recommended Emissions Factor?.....	7-6
Section 8.0	What Parameters Should I Consider When Deriving a User-Defined Emissions Factor?.....	8-1

8.1	Source Category and Process Considerations.....	8-1
8.2	Control Device Considerations.....	8-3
8.3	Pollutant Test Method Considerations.....	8-3
Section 9.0	How Do I Develop a User-Defined Emissions Factor?	9-1
9.1	How Do I Use WebFIRE to Create a User-Defined Emissions Factor?.....	9-1
9.2	What are the Potential Impacts Associated with Applying a User-defined Emissions Factor?	9-3
Section 10.0	How Do I Submit Data to WebFIRE?	10-1
10.1	What is the ERT and How is it Used to Document Emissions Tests?.....	10-2
10.2	What is the CDX and What is Its Role in Submitting Data to WebFIRE?.....	10-4
Section 11.0	What is the Data Review and Public Participation Process for Emissions Factor Development?	11-1
Appendix A – Guidance for Using Emissions Factors for Non-Inventory Applications		
Appendix B – Procedures for Determining Individual Test Report Quality Ratings		
Appendix C – Procedures for Handling Test Data That Are Below the Method Detection Limit		
Appendix D – Procedures for Determining Statistical Outliers		
Appendix E – Emissions Factor Development and Data Quality Characterization Procedures		
Appendix F – Statistical Procedures for Determining Valid Data Combinations		
Appendix G – Source Classification Codes for Source Categories Containing 15 or Fewer Units		

LIST OF FIGURES

Figures	Page No.
Figure 5-1. EPA's Revised Procedures for Developing Emissions Factors	5-2
Figure 6-1. WebFIRE Overview	6-2
Figure 7-1. Procedures for Retrieving Emissions Factors from WebFIRE	7-2
Figure 9-1. Emissions Factor Derivation in WebFIRE.....	9-2
Figure 10-1. Typical Work Flow When Using the ERT.....	10-3
Figure 11-1. Overview of the WebFIRE Public Participation and Emissions Factor Development Process	11-2

DRAFT

LIST OF TABLES

Tables	Page No.
Table 7-1. Data Fields Reported by WebFIRE Emissions Factor Search	7-3
Table B-1. Test Report Quality Rating Tool	B-3
Table C-1. Summary of WebFIRE Procedures for Handling Test Data That are Below Detection Limits.....	C-3
Table C-2. Example Data Set A.....	C-4
Table C-3. Example Data Set B.....	C-4
Table C-4. Calculations for Example Data Set B	C-5
Table C-5. Example Data Set C.....	C-5
Table C-6. Calculations for Example Data Set C	C-6
Table D-1. Table of Critical Values of Q_{crit}	D-3
Table F-1. Emissions Factor Characteristics for Group A and B	F-2
Table F-2. T Statistics for Group A and B.....	F-2
Table F-3. Emissions Factor Characteristics for Group C and D	F-3
Table F-4. T Statistics for Group C and D.....	F-4
Table F-5. Emissions Factor Characteristics for Group A and B	F-4
Table F-2. T Statistics for Group E and F.....	F-5
Table G-1. Source Classification Codes for Source Categories Containing 15 or Fewer Units.....	G-1

LIST OF ACRONYMS

Acronym	Term
AFS	Air Facility System
AFSEF	AIRS Facility Subsystem Emission Factor
AIRS	Aerometric Information Repository System
AMS	Area and Mobile Source
APTI	Air Pollution Training Institute
AP 42	Compilation of Air Pollutant Emission Factors, Volume I: Stationary Point and Area Sources
ASCII	American Standard Code for Information Interchange
BDL	Below Minimum Detection Limit
CAA	Clean Air Act
CAAA	Clean Air Act Amendments
CAS	Chemical Abstracts Service
CATC	Clean Air Technology Center
CDX	Central Data Exchange
CEMS	Continuous Emissions Monitoring System
CFR	Code of Federal Regulations
CO ₂	Carbon Dioxide
CSV	Comma Separated Values
CTM	Conditional Test Method
CTR	Composite Test Rating
EIS	Emission Inventory System
ELCD	Electrolytic Conductivity Detector
EMC	Emission Measurement Center
EPA	Environmental Protection Agency
ERT	Electronic Reporting Tool
FAQs	Frequently Asked Questions
FID	Flame Ionization Detector
FIRE	Factor Information Retrieval
FTIR	Fourier Transform Infrared
GC/MS	Gas Chromatography/Mass Spectroscopy
HAP	Hazardous Air Pollutants
HTML	Hypertext Markup Language
L&E	Locating and Estimating
MACT	Maximum Achievable Control Technology
MDL	Minimum Detection Limit
NDIR	Nondispersive Infrared Analyzer
NEET	New and Emerging Technology
NEI	National Emissions Inventory
NESHAP	National Emission Standard for Hazardous Air Pollutants

LIST OF ACRONYMS (Continued)

Acronym	Term
NMOC	Nonmethane Organic Compound
NO _x	Oxides of Nitrogen
NSPS	New Source Performance Standard
OAQPS	Office of Air Quality Planning and Standards
O&M	Operation and Maintenance
Pb	Lead
PDS	Project Data Set
PM	Particulate Matter
PM ₁₀	Particulate Matter with an aerodynamic diameter of 10 microns or less
PM _{2.5}	Particulate Matter with an aerodynamic diameter of 2.5 microns or less
QAPP	Quality Assurance Project Plan
QA/QC	Quality Assurance/Quality Control
RATA	Relative Accuracy Test Audit
SCC	Source Classification Code
SO ₂	Sulfur Dioxide
SO _x	Sulfur Oxides
TOC	Total Organic Compounds
TRI	Toxic Release Inventory
XATEF	Crosswalk/Air Toxics Emission Factor System
XML	Extensible Markup Language

Section 1.0
WHAT IS THE PURPOSE OF THIS DOCUMENT?

This guidance document describes the procedures, data evaluation criteria, and associated tools and data management systems that the U.S. Environmental Protection Agency (EPA) recommends for developing air pollutant emissions factors for stationary emissions units or processes. This document supersedes the previous EPA guidance document for emissions factor development (*Procedures for Preparing Emission Factor Documents (EPA-454/R-95-015, November 1997)*).

This document presents an introduction to emissions factors and provides the historical background for how and why the EPA has developed recommended emissions factors for stationary emissions units or processes. This document also describes the approach and procedures followed by EPA when developing new or revising existing recommended emissions factors.

This document provides an overview of EPA's WebFIRE – an online data storage and emissions factor retrieval and development tool. The EPA's Electronic Reporting Tool (ERT), a Microsoft Access[®] application that facilitates development and documentation of emissions test reports and the procedures that must be followed by individuals and entities to submit emissions and related process data to WebFIRE are also presented in this document. Finally, this document provides an overview of the data review and public participation process that the EPA follows when developing new or revised recommended emissions factors.

This document is organized as follows:

This section . . .	Contains or describes . . .
2.0	An overview of the characteristics that define an emissions factor.
3.0	A brief summary of EPA's historical procedures used to develop emissions factors and the various support programs prepared by the Agency.
4.0	A discussion of the various uses and limitations of emissions factors.
5.0	An overview of the Agency's revised approach for developing EPA-recommended emissions factors.

This section ...	Contains or describes ...
6.0	An overview of WebFIRE, EPA's online application for storage, retrieval, and development of emissions factors.
7.0	Considerations that should be evaluated by end users when developing user-defined emissions factors.
8.0	The steps users must follow to identify and retrieve emissions factors from WebFIRE.
9.0	The procedures users must follow to develop a user-defined emissions factor from a collection of related data contained in WebFIRE.
10.0	The steps to follow to submit emissions and related process data to WebFIRE.
11.0	The process by which the public can participate in the periodic development of EPA's recommended emissions factors.

This document also contains the following appendices:

This appendix ...	Contains or describes ...
A	Guidance for Using Emissions Factors for Non-Inventory Applications
B	Procedures for Determining Individual Test Report Quality Ratings
C	Procedures for Handling Test Data That Are Below the Method Detection Limits
D	Procedures For Determining Statistical Outliers
E	Emissions Factor Development and Data Quality Characterization Procedures
F	Statistical Procedures for Determining Valid Data Combinations
G	Source Classification Codes For Source Categories Containing 15 or Fewer Units

Section 2.0 WHAT IS AN EMISSIONS FACTOR?

An emissions factor is a tool used to estimate air pollutant emissions from a normally-operating process or activity (e.g., fuel combustion, chemical production). An emissions factor relates the quantity of pollutants released to the atmosphere from a process to a specific activity associated with generating those emissions. For most application purposes, users typically assume that an emissions factor represents the long-term average emissions for all facilities in a particular source category (i.e., the emissions factor represents a population average).

The simplest form of an emissions factor is an expression of the mass of pollutant emitted per unit of activity generating the emissions (e.g., pounds of particulate matter (PM) emitted per ton of coal burned). Typically, emissions factors are used to estimate process emissions as follows:

$$E = A \times EF \times [1 - (ER/100)]$$

Where:

E = emissions,
A = activity rate,
EF = emissions factor, and
ER = overall emissions reduction achieved by controls (%).

Emissions factors for more complex processes or activities (e.g., paved and unpaved roads, organic liquid storage tanks) are typically expressed using empirical equations. The empirical equation relates independent variables to the source emissions and typically provides for improved predictive accuracy when compared to a simple emissions factor. For example, the following emissions factor for vehicles traveling on unpaved surfaces at industrial sites was taken from EPA's *Compilation of Air Pollutant Emission Factors, Volume I: Stationary Point and Area Sources* (AP 42) (5th Edition, Section 13.2.2):

$$E = k (s/12)^a (W/3)^b$$

Where:

- E = particle size-specific emissions factor (lb/vehicle miles traveled);
- k = particle size multiplier (lb/vehicle miles traveled);
- s = surface material silt content (%);
- a, b = particle size-specific empirical constants, and
- W = mean vehicle weight (tons).

Emissions factors are developed using two related components: (1) emissions data, and (2) activity data.

2.1 EMISSIONS DATA

Typically, emissions data are obtained through direct measurement of releases from a process or activity (i.e., a sample of the process emissions is collected and analyzed). The emissions rate for the source, expressed in terms of mass of pollutant emitted per time unit (e.g., lb PM/hr), is calculated as the arithmetic average of the available, quality-assured test data. Depending on the sampling location and configuration of the process and associated control devices (if any), emissions data can reflect controlled or uncontrolled emissions.

Direct measurements of facility or process emissions are conducted for a variety of reasons such as:

- Characterize process emissions and/or control device performance;
- Assess changes in process or control device operation on emissions; and
- Demonstrate compliance with federal, state, local, or tribal air regulations.

Emissions testing may also be conducted for purposes such as conducting relative accuracy test audits (RATAs), linearity checks, and routine calibrations of continuous emissions monitoring system (CEMS) equipment.

The emissions rate for a specific process can also be determined by accounting for all of the materials entering and exiting that process and the process operating parameters. Using this material balance approach, pollutant emissions are calculated as the difference in process inputs and outputs. For certain processes, a mass balance provides an easier and less expensive

estimate of emissions than would be obtained by direct measurement. For example, carbon dioxide (CO₂) emitted from a fuel combustion process can be estimated from the stoichiometric relationship of the chemical reactants (i.e., carbon contained in the fuel and oxygen in the combustion air), the amount of each reactant that is consumed in the combustion process, and the amount of carbon remaining in any residual material (e.g., ash). In general, material balances are appropriate for use in situations where material is lost to the atmosphere (e.g., solvent evaporation in an uncontrolled coating process). Furthermore, a material balance may not be appropriate to estimate emissions from a process or activity in which material is consumed or chemically combined in the process.

2.2 ACTIVITY DATA

The composition and magnitude of emissions generated by a process unit are affected by a variety of process parameters such as raw materials and fuels used; process operating conditions; equipment configuration and age; and the skill and experience of process operators. Activity data for use in developing emissions factors are the parameter(s) that directly influence the quality and quantity of emissions from a process unit. Generally, activity data are collected during an emissions test to verify that the process is operating at the desired production level (e.g., to satisfy an operating permit emission limit). Activity data are typically expressed either in terms of a process input or output per time unit (e.g., gallons of oil burned per hour, tons of cement produced per day). For example, the activity data for a PM emissions factor for plywood manufacturing processes could be expressed in terms of the square feet of plywood produced per day. For an emissions rate determined using a mass balance approach, the activity data would typically include one or more process parameters used in the mass balance.

Section 3.0

HOW HAVE WE HISTORICALLY DEVELOPED EMISSIONS FACTORS?

The Clean Air Act of 1970 (CAA) defined EPA's responsibilities with regard to protecting and improving the nation's air quality. In response to the CAA, the EPA needed a method with which to characterize and quantify air pollutant emissions from processes and activities on a nationwide basis. Because there were a large number of diverse emissions sources, developing national estimates based upon site-by-site emissions testing was not feasible. Consequently, we developed criteria and non-criteria pollutant emissions factors for certain industrial processes or source categories for use in preparing emissions inventories. These emissions factors were based upon emissions test data, material balance calculations, modeling, and engineering judgment.

In 1972, the EPA's Office of Air Quality Planning and Standards (OAQPS) published the first document containing EPA's recommended emissions factors and supporting documentation (*Compilation of Air Pollutant Emission Factors, Volume I: Stationary Point and Area Sources* (AP 42)). As an aid to end users, OAQPS developed relative quality ratings for the AP 42 emissions factors, based upon EPA's analysis of the quality of the underlying test data and how representative the emissions factor was for the particular source category for which it was developed. The letter-grade ratings (e.g., A for excellent, E for poor) were based primarily on engineering judgment and did not incorporate statistical error bounds or confidence intervals.

Since its initial publication, we have periodically revised and updated AP 42 to incorporate new data and emissions estimating methodologies. The last hard copy version of AP 42 (fifth edition) was published in 1995; although, we have released five supplements through 2000. We provided electronic copies of AP 42 and the supplements (e.g., CD-ROM, downloadable files) to improve availability of the document. Currently, the fifth edition of AP 42, the supplements, and related information are available at:

<http://www.epa.gov/ttn/chief/ap42/>.

In addition to AP 42, we developed several other compilations of recommended emissions factors. To provide the user community with additional emissions factor information for air toxic pollutants beyond what was available in AP 42 at the time, we initiated the *Locating & Estimating (L&E)* document series in 1984. Unlike AP 42, which is organized by source category, the majority of the *L&E* documents focused on a specific pollutant (e.g., arsenic, benzene) or related group of pollutants (e.g., polycyclic organic matter). The *L&E* documents made use of AP 42 emissions factors, where available; however, in some cases, the AP 42 emissions factors were revised or supplemented to present the most complete assessment of the emissions for the specific air pollutant. A total of 36 individual *L&E* documents were produced through 1998.

We also compiled the Aerometric Information Retrieval System (AIRS) Facility Subsystem Emission Factors (AFSEF) and the Crosswalk/Air Toxics Emission Factors (XATEF) databases in 1990. The AFSEF database documented all emissions factors for criteria pollutants that existed in the AIRS mainframe look-up tables as of March 1990. The XATEF database contained emissions factors for toxic air pollutants that were developed based upon data available to EPA through October 1990. Ultimately, the EPA retired the AFSEF and XATEF databases and created the Factor Information Retrieval (FIRE) Data System. The FIRE database contains emissions factors from all AP 42 sections posted by September 1, 2004, the *L&E* document series, and the retired AFSEF and XATEF databases.

In 1997, we provided guidance materials (*Procedures for Preparing Emission Factor Documents, EPA-454/R-95-015, November 1997*) that described the procedures, technical criteria, and standards and specifications for developing and reporting air pollutant emissions factors for publication in either AP 42 or the *L&E* document series. This guidance document covered the compilation, review, and analyses of new data and information and preparation of supporting documentation for emissions factor development.

Although OAQPS has focused significant effort and resources on developing emissions factors, the procedures and guidance we have historically followed (*documented in EPA's Procedures for Preparing Emission Factor Documents, November 1997*) have not kept pace with

the increased volume of available emissions data or advances in information technology. For example, although AP 42 is available online, the format is analogous to a hard-copy document which is not conducive to incorporating new data, making corrections to data, or conducting data analyses. Also, because of their complex and somewhat subjective nature, the past emissions factor development procedures were slow to incorporate new emissions test data and also did not encourage active public participation. Additionally, our emissions factor development process and guidance materials have not considered non-inventory uses for emissions factors (see section 4.0 below). To address these short comings, we have revised our approach for developing emissions factors to be more responsive and transparent. Section 5.0 discusses our revised approach to developing and documenting emissions factors.

DRAFT

Section 4.0

HOW ARE EMISSIONS FACTORS USED?

Emissions factors are used to develop emissions estimates for processes and activities in cases where direct measurements are unavailable. Emissions factors are typically developed to represent long-term (e.g., annual) average emissions and, accordingly, data used for developing the emissions factors is usually based on emissions testing collected during normal process operating conditions. Short-term emissions from a particular process will vary significantly over time (i.e., within-process variability) because of fluctuations in normal process operating conditions, control device operating conditions, raw materials, ambient conditions, and other factors. Because of the relatively short duration of emissions tests and the limited range of conditions they represent, the available emissions and process data used to develop an emissions factor are not sufficient to account for these short-term emissions fluctuations.

The recommended emissions factors we have published are intended for use in preparing regional and national emissions inventories and the first part of the development of a regional or national control strategy to reduce area wide emissions. These inventories are essential tools in air quality management because they are used to estimate ambient pollutant concentrations; to model pollutant dispersion and transport in the atmosphere; and to develop and assess control strategies. Despite their original purpose, we are aware that emissions factors have been applied by other entities (e.g., federal, state, tribal, and local agencies; consultants; industries) for purposes beyond the intended use of supporting national and regional emissions inventory programs. Examples of non-inventory applications include:

- Developing site-specific emissions estimates;
- Identifying and evaluating site-specific control strategies and implementation plans;
- Determining applicability of permit and regulatory requirements;
- Conducting risk assessments;
- Demonstrating compliance;
- Verifying emissions offsets/emissions banking credits;
- Establishing permit limits;
- Collecting emissions statements/fees;
- Toxic Release Inventory (TRI) reporting; and

- International treaty reporting.

We are concerned that emissions factors have been applied to these non-emissions inventory uses without consideration of the limitations inherent in the use of emissions factors (e.g., factors are not particularly suitable to developing short-term or site-specific emissions estimates). We recommend that entities that choose to use emissions factors in non-inventory applications consider the impact of the reliability of emissions factors on their non-inventory programs (e.g., apply statistical procedures to account for variability). We also recommend that facilities conduct periodic retesting to confirm and adjust, if necessary, both the emissions factor and any applicable emissions limit derived from an emissions factor.

In light of the fact that the use of emissions factors has expanded beyond support of national and regional emissions inventory programs, we have developed additional guidance that addresses the use of emissions factors for non-inventory purposes (see Appendix A).

Section 5.0

WHAT ARE EPA'S REVISED PROCEDURES FOR DEVELOPING EMISSIONS FACTORS?

Beginning in 2003, OAQPS, the National Academy of Sciences, and EPA's Office of Inspector General conducted a review of the Agency's emissions factors program. Based upon the feedback received from stakeholders (e.g., industry, state/local/tribal entities, EPA program offices, environmental action groups), we revised our historical approach to developing emissions factors to reduce the level of subjectivity involved in the emissions factor development process. Our revised approach is also intended to improve the transparency and responsiveness of the process and to encourage meaningful public participation. Figure 5-1 provides an overview of our revised approach to developing new or to revising existing recommended emissions factors. The key revisions that we have implemented in our approach regarding the collection of emission data and supporting documentation, the evaluation of data, and the development and assessment of emissions factors are described in the following sections.

5.1 DATA COLLECTION

Based upon the review of our emissions factor program, we found that most emissions testing information and associated data are currently generated electronically. To take advantage of advances in information technology and the more widespread availability of electronic data production, our revised approach focuses on collecting new emissions data available in an electronic format.

To aid facilities in planning and reporting the results of emissions tests, we developed the Electronic Reporting Tool (ERT) (see Section 10.1). The ERT replaces time-intensive manual methods for test planning, test data compilation and reporting, and data quality assurance evaluations. Because of the prevalence of electronic data, we believe that our transition from the use of predominantly hard-copy resources (e.g., test reports, technical publications) for emissions factor development to the use of data in an electronic format will be relatively easy. The use of an electronic format will facilitate the ongoing collection, incorporation, and analysis of new test data and supporting documentation. Also, use of the ERT will enable us to streamline the

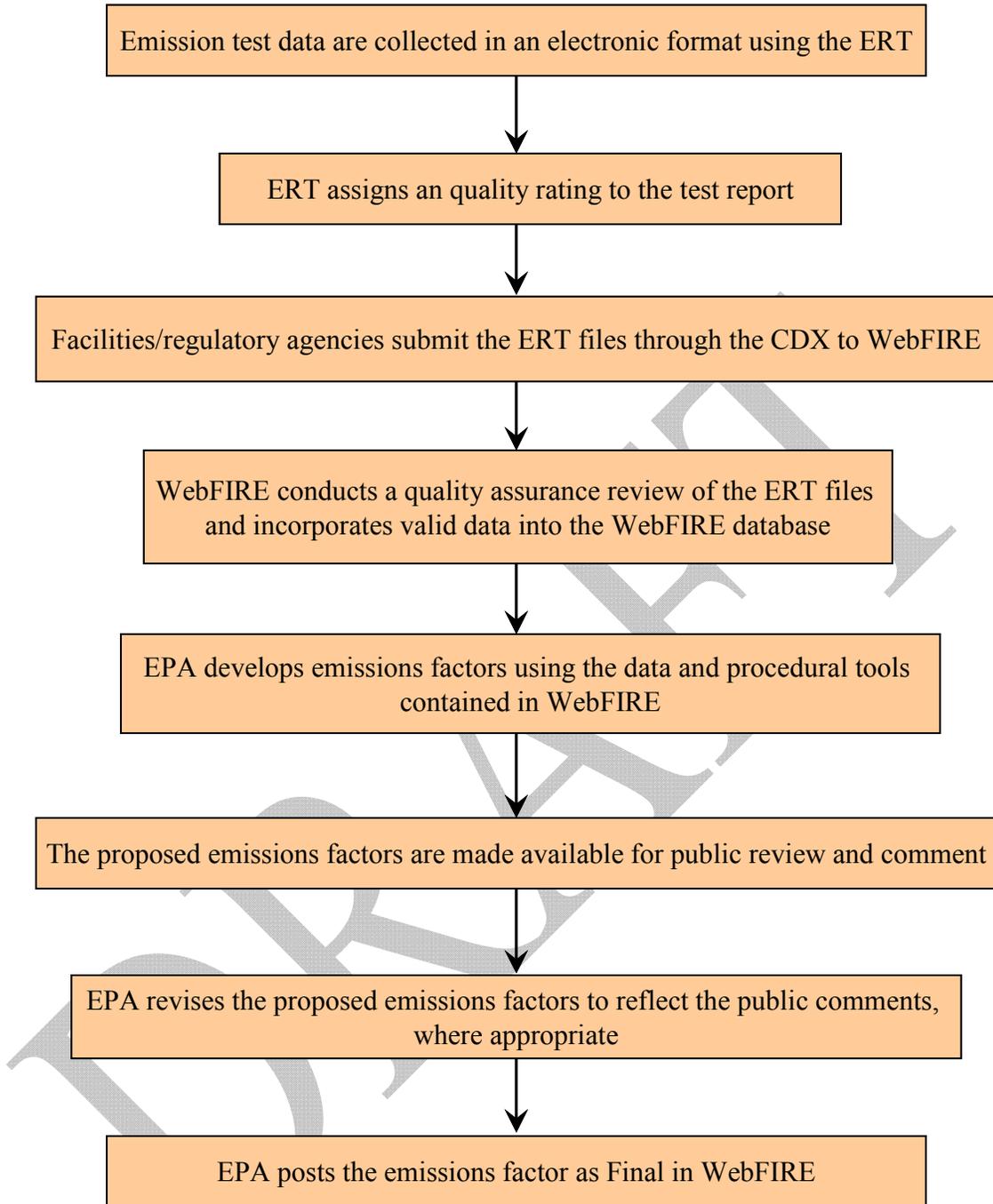


Figure 5-1. EPA's Revised Procedures for Developing Emissions Factors

emissions factor development process through more rapid data handling and quality assurance checks.

5.2 TEST DATA EVALUATION

Historically, EPA's quality ratings of emissions test data and test reports were largely subjective because each test program presented different issues (i.e., no two facilities, their operation, or the tests conducted at those facilities are exactly alike). Typically, EPA developed letter-grade quality ratings (A through D) for test reports based upon the Agency's review of the following criteria areas:

- Process operation;
- Test method and sampling procedures;
- Process information; and
- Analysis and calculations.

To reduce the subjectivity of our qualitative assessment of the emissions, process, and control device data collected during an emissions test, we have developed a more objective rating system for test reports (see Appendix B). The rating system is intended to produce unbiased and consistent assessments of the information included in test reports which, in turn, will help us to better characterize the process and the quality of emissions values.

The rating system consists of a set of objective review questions developed for EPA's for manual and instrument test methods that assess the quality of the process, control device, and measurement data collected during an emissions test in the following criteria areas:

- General information;
- Process and control device information;
- Sampling locations;
- Test methods and reporting requirements;
- Sampling equipment calibrations;
- Sample recovery; laboratory analysis; and
- Documentation.

A numeric score (the Individual Test Rating (ITR)) is determined for each test report as the prorated sum of the individual scores assigned to each review question based upon the answers provided (see Appendix B).

Our rating system also encourages facilities to improve the quality of underlying measurement data and potentially increase the ITR value by submitting their emissions test data to a state regulatory agency for their technical review. In cases where the state agency affirms the original responses provided to the review questions, additional points are awarded to the ITR value originally assigned by ERT when the measurement data were initially recorded by the testing contractor. If the state regulator determines that the initial review points were incorrectly assigned, the points originally assigned to a particular review question are deducted from the ITR.

5.3 DETECTION LIMIT PROCEDURES

After the candidate data set has been established, we must determine if any of the new data are based upon test results that were below the minimum detection limit (MDL) of the test method used to collect the emissions measurements. The MDL is defined by EPA as “the minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero and is determined from an analysis of a sample in a given matrix containing the analyte.” Essentially, the MDL is the smallest amount of a substance that an analytical method can reliably distinguish from zero, at a specified confidence level, from the instrument signal produced by a blank sample.

We have developed specific data handling procedures for cases where some or all of the emissions data collected during a test are below the MDL (BDL) and where the average data from the BDL tests are to be included in the candidate data set for use in developing an emissions factor. Appendix C contains a more detailed discussion of the procedures that we follow for handling BDL data.

5.4 IDENTIFICATION OF OUTLIER DATA

After the BDL data have been properly addressed, we subject the candidate data set (i.e., the new data that have been subjected to the BDL procedures and the existing AP 42 data) to statistical outlier tests to determine if any data values should be eliminated from further consideration. A statistical outlier refers to one or more data points that do not conform to the statistical pattern established by other data under consideration for the same process. These outlier values can be caused by an unusual process condition or circumstance that produced an unexpected and unrepresentative variation in the process emissions.

For the purposes of identifying outliers, our revised approach for developing emissions factors uses the Dixon Q test or the Grubbs test, depending on the number of data values to be evaluated. If there are fewer than three data points in the subject data set, an outlier analysis is not conducted. Appendix D contains a detailed discussion of the statistical tests we use to determine outlier data values. If data are determined to be outliers, the procedure is to flag these data as outliers and omit them in developing the EPA-recommended emissions factor while retaining them in the WebFIRE database.

5.5 EMISSIONS FACTOR DERIVATION AND QUALITY ASSESSMENT

After evaluating the candidate data set for BDL data and outlier values, we recommend a step-wise procedure to: (1) calculate an emissions factor value using the individual data points that result in the highest quality rating and most representative factor for the source category of interest, and (2) assign the quality rating of the resulting emissions factor. The procedures for calculating the emissions factor value and assessing factor quality are based upon an evaluation of the number of individual sources in the source category for which the emissions factor is being developed, the quality rating of individual test data (ITR), and the number of individual data points used to calculate the recommended emissions factor. Appendix E contains a detailed description of the emissions factor development and data quality characterization procedures.

Section 6.0

EPA'S INTERACTIVE DATABASE FOR THE EMISSIONS FACTORS PROGRAM – WHAT IS WEBFIRE?

6.1 WHAT IS WEBFIRE?

WebFIRE is EPA's online emissions factors repository, retrieval, and development tool. The WebFIRE database contains EPA's recommended emissions factors for criteria and hazardous air pollutants (HAP) for industrial and non-industrial processes. In addition, WebFIRE contains the individual data values used to develop the recommended factors and other data submitted to EPA by federal, state, tribal, and local agencies; consultants; and industries. For each recommended emissions factor and individual data value, WebFIRE contains descriptive information such as industry and source category type, control device information, the pollutants emitted, and supporting documentation. The home page for WebFIRE and links to Frequently Asked Questions (FAQs) and background information on data contained in the WebFIRE system can be found at:

<http://cfpub.epa.gov/webfire/>.

At this time, WebFIRE does not contain CEMS data. Although the WebFIRE system could accept and store CEMS data as emissions records, WebFIRE does not yet incorporate the corresponding process data and calculation algorithms necessary to develop activity-based emissions factors using CEMS data. We intend to provide this expanded capability in future releases of WebFIRE because we recognize the importance and potential value of CEMS data to emissions factor development.

6.2 HOW IS WEBFIRE USED?

WebFIRE's two primary functions are to provide: (1) storage and retrieval of recommended emissions factors and individual data points, and (2) tools for calculating and assessing the representativeness of a user-defined emissions factor derived from a set of individual data points. Figure 6-1 provides an overview of WebFIRE and its basic functionality.

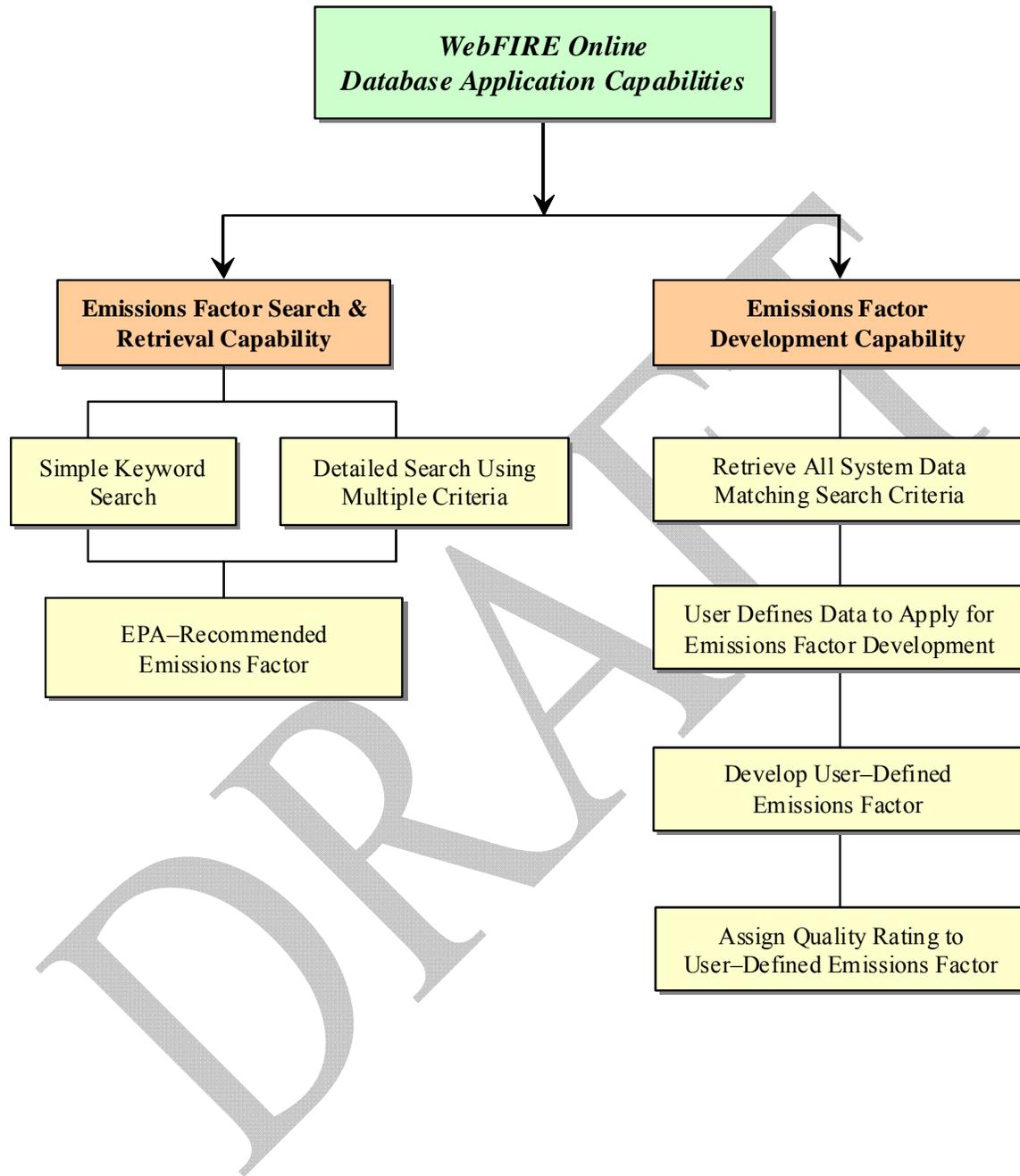


Figure 6-1. WebFIRE Overview

To retrieve a recommended emissions factor, WebFIRE provides for either a simple or detailed search. The simple search (denoted on the WebFIRE page as “Simple Keyword Search”) allows the user to search for emissions factor information in cases where the user has limited knowledge of the emissions process of interest (e.g., the emissions process is a wood-fired boiler). The simple search can be used as a starting point in WebFIRE; however, refining the search to determine the most useful and applicable emissions factor requires an iterative progression through the database that can be time-intensive. The detailed search (denoted on the WebFIRE page as “Detailed Emission Factor Search”) allows users to search and retrieve emissions factors in cases where they have detailed knowledge of emissions process of interest (e.g., the process is a wood-fired boiler that is controlled by a scrubber and electrostatic precipitator in series). Although one needs more informational inputs to initiate the detailed search, there are fewer iterative steps required (i.e., WebFIRE returns a useful emissions factor in less time).

Both the simple and detailed searches also provide a link that returns the data values used to derive the selected emissions factor and all other data values contained in WebFIRE that meet the search criteria. These other data values may include new data submitted to WebFIRE that have not been reviewed by EPA. Section 8.0 provides a more detailed discussion of the WebFIRE emissions factors search and retrieval tools.

WebFIRE also provides tools that allow a user to calculate an emissions factor from a set of individual data points contained in WebFIRE. These WebFIRE tools incorporate our revised approach for developing EPA-recommended emissions factors (see Section 5.0). In general, the user selects the individual data values to be used in developing an emissions factor. After the user selects the preliminary data set, WebFIRE evaluates the data set to identify and address BDL data and outlier values. Following the BDL and outlier value analyses, WebFIRE calculates an emissions factor value from the data set that best represents the process of interest. WebFIRE also assigns a relative quality rating to the user-defined emissions factor. Section 9.0 discusses WebFIRE's emissions factor development tools in more detail. Appendices C through E contain the BDL and outlier analyses and the calculations and procedures for deriving a user-defined emissions factor.

6.3 WHO USES WEBFIRE?

The data storage, retrieval, and emissions factor development capabilities of WebFIRE are available online to all public and private entities. Examples of WebFIRE users include, but are not limited to:

- Federal, state, local, or tribal air pollution control and regulatory agency personnel (e.g., for emission inventory development, preparation of emissions estimates for dispersion modeling, comparison of a site-specific emissions factor to an EPA-recommended emissions factor for a given process), and
- Environmental staff at industrial facilities (e.g., for emissions and process data submittal; comparison of process emissions to an EPA-recommended emissions factor or other related data).
- Environmental action groups (e.g., for air emissions and air permit oversight).
- Engineering consultants, university researchers, and international air agencies.

Periodically, the EPA will use the data and development tools contained in WebFIRE to revise existing and derive new recommended emissions factors as discussed in Section 11.0.

6.4 HOW DOES WEBFIRE IMPROVE EMISSIONS FACTOR IDENTIFICATION AND DEVELOPMENT?

The emissions factor repository, retrieval, and development tools in WebFIRE allow EPA to progress towards our goal of developing an interactive emissions factors program that will incorporate new data as it becomes available and produce high-quality emissions factors in a timely manner. We also believe that the benefits of online data access and electronic data submittal provided by WebFIRE will allow for easier, more effective involvement by the public interested in developing and improving emissions factors.

WebFIRE will also allow EPA to shift the role of OAQPS from that of sole developer of emissions factors to that of a facilitator. This shift will allow us to focus more resources on overseeing the emissions factor program, ensuring that more high-quality emissions factors are developed, and on developing policies for the appropriate use of emissions factors in non-

inventory applications where there are no policies currently available, or where existing policies are inadequate.

DRAFT

Section 7.0 HOW DO I FIND AN EMISSIONS FACTOR?

7.1 HOW DO I IDENTIFY AND RETRIEVE AN EMISSIONS FACTOR FROM WEBFIRE?

You have two options in WebFIRE to search for and retrieve EPA's recommended emissions factors: a Simple Keyword Search, and a Detailed Emissions Factor Search. WebFIRE also allows you to expand your simple or detailed search to include previously-recommended emissions factors that have been revoked by EPA. Figure 7-1 provides an overview of the factor retrieval process. Table 7-1 lists the data fields that are provided for each emissions factor record.

Using the Simple Keyword Search (Step 1 in Figure 7-1), you can retrieve emissions factor records by entering one or more simple terms such as: source category name (e.g., dry cleaning, wood combustion, boilers), process description (e.g., spreader stoker, catalytic cracking), SCC, or any other viable search term likely to be found in an emissions factor record. For example, if you enter in the phrase "spreader stoker," the simple search results page will display every EPA-recommended emissions factor that includes the complete phrase "spreader stoker" anywhere in the entire record. To make your search more specific, you can use the "AND" operator. For example, "spreader stoker AND PM₁₀" will limit the results to the pollutant PM₁₀. The "AND" operator must be capitalized. Do not use punctuation in the search window.

The SCCs are used by EPA to organize data for anthropogenic air pollutant sources that have similar production and emissions characteristics (e.g., gasoline storage tanks, polymer manufacturing facilities) into related groups or source categories. An overview of the SCC system is provided in Section 7.1 and the current list of SCCs and their descriptions can be downloaded from the WebFIRE website (<http://cfpub.epa.gov/webfire/>). At the WebFIRE website, clicking on the link for "All WebFIRE Source Classification Codes (SCC)" will take you to a Microsoft Excel[®] spreadsheet that lists various SCCs contained in WebFIRE. When searching using an SCC, do not use dashes, spaces, or other punctuation when entering the codes

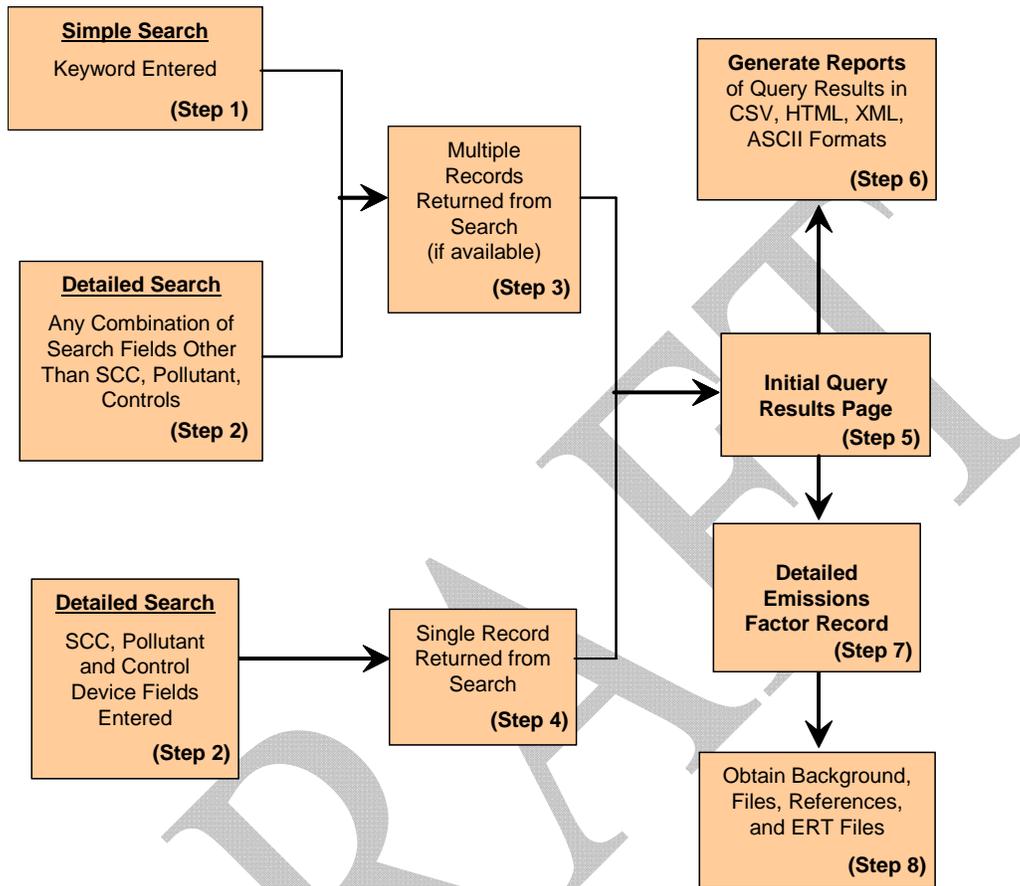


Figure 7-1. Procedures for Retrieving Emissions Factors from WebFIRE

Table 7-1. Data Fields Reported by WebFIRE Emissions Factor Search

Emissions Factor Record Data Elements	Description
Emissions factor	Numerical value and units of the emissions factor
SCC	Source Classification Code
SCC Levels	SCCs are comprised of four levels (starting with the most general source classification to the most specific). The definition of each level for the SCC is provided.
Pollutant name	Chemical name of pollutant factor
CAS number	Chemical Abstract Service (CAS) number assigned to the pollutant
NEI code	Identification number assigned to the pollutant in the NEI
Quality score	Individual Testing Rating (ITR) for process test data or Composite Test Rating (CTR) for recommended factors
Emissions Factor Representativeness	Qualitative characterization of how well an emissions factor statistically represents the population of similar facilities in a source category.
Primary control device	The first control device applied to the process
Second control device	The second control device applied to the process
Third control device	The third control device applied to the process
Fourth control device	The fourth control device applied to the process
Fifth control device	The fifth control device applied to the process
Sixth control device	The sixth control device applied to the process
Status	Identifies emissions factors as individual data value, EPA recommended factor, or proposed recommended emissions factor undergoing review
Data source type	Refers to the original document(s) from which factors were obtained
Restriction type	Refers to caveats or special considerations prior to use of the emissions factor
References	Test report or citation where the factor was derived
AP 42 Section	Identifies the specific AP 42 chapter and section where the process data and emission factor can be found
Formula	Empirical equation used to express an emissions factor
Date	Represents the date the emissions test data was collected and the date the factor was developed/revised.
Notes	Additional information to assist the user in understanding and applying an emissions factor

into the search window. For example, the SCC 1-01-001-01 would be entered as 10100101. If you want to conduct a search using a partial SCC code (e.g., SCC beginning with “101”), you should place an asterisk (*) after the partial code (e.g., “101*”). This format ensures that the search will retrieve all records where the SCC field begins with 101 but can also return other

records where the term “101” occurs in other data fields. For example, the “Notes” field may contain SCCs and other numerical values that match the “101” term.

To minimize the potentially large number of emissions factor records retrieved when using a simple search, you can use the Detailed Emissions Factor Search (Figure 7-1, Step 2). The detailed search allows you to focus the factor retrieval process by entering multiple terms for the search criteria including:

- SCC (complete code or individual SCC level descriptions),
- Control device type,
- Pollutant or pollutant group type, and
- Specific AP 42 Section.

Whether you enter a complete SCC (8- or 10-digit), or the four individual descriptions for each SCC level, WebFIRE will return the same search results, provided the descriptions are correctly selected to match a valid SCC. For example, using SCC 10200203 will produce the same search result as using the following SCC level descriptions:

- Level 1: External Combustion Boilers,
- Level 2: Industrial,
- Level 3: Bituminous/Subbituminous Coal, and
- Level 4: Cyclone Furnace.

For the detailed search criteria, you are provided a drop-down menu of choices from which to select. After a search is conducted, you have the option to refine your search, as necessary.

For either the simple or detailed search (Figure 7-1, Step 5), the results page for the recommended emissions factor provides the following information:

- SCC,
- Level 1, 2, 3, and 4 SCC Descriptions,
- Pollutant name,
- Pollutant NEI number,
- Pollutant Chemical Abstract Service (CAS) number,
- Control device(s),

- Emissions factor,
- Emissions factor quality rating,
- Emissions factor representativeness,
- Data source type,
- Restriction type,
- Date of factor development,
- Factor status,
- Emissions factor reference(s),
- Applicable AP 42 section,
- Formula, and
- Notes.

At this stage of the search, you have the option of: (1) creating a summary report of the information shown on the results page (Figure 7-1, Step 6), or (2) obtaining additional background information for the recommended emissions factor that you selected (see Section 7.2). To accommodate various end uses of the retrieved data (e.g., emissions calculations, incorporation into a text file), WebFIRE provides you with the following reporting formats:

- Comma Separated Values (CSV) format (for importation into a spreadsheet or database),
- Extensible Markup Language (XML) format (for importation into XML parsing applications),
- American Standard Code for Information Interchange (ASCII) format (for importation into other applications), and
- Hypertext Markup Language (HTML) format (for printing).

7.2 HOW DO I OBTAIN BACKGROUND INFORMATION FOR MY SELECTED EMISSIONS FACTOR?

At the search results page, WebFIRE provides you the option of retrieving additional detailed information for the recommended emissions factor that you selected (Figure 7-1, Step 7). Clicking on the “Details” button located at the right-hand side of the search results page provides you with information such as the citation for the data; the applicable AP 42 section; formulas and equations that are applicable to the factor; and information on process configurations, operating conditions, control device configurations, and test conditions relevant to the emissions factor that you selected. This information is intended to give you a better understanding of your specific factor so you can make better decisions regarding its applicability.

From the “Emissions Factor Details” page, you can also retrieve additional supporting documentation for an emissions factor (Figure 7-1, Step 8). Links to web-based files are provided that allow you to obtain items such as factor background information documents, individual emissions test reports and data, and any other available documentation materials that may help you to better understand a factor’s derivation.

7.3 HOW DO I IDENTIFY THE DATA THAT ARE USED TO DERIVE THE EPA-RECOMMENDED EMISSIONS FACTOR?

In addition to the emissions factor data retrieval tools described in Sections 7.1 and 7.2, WebFIRE allows you to identify the specific emissions test data that were used to calculate the recommended emissions factors, as well as any other data contained in WebFIRE that met the search criteria (e.g., SCC, pollutant, control device) used to retrieve the recommended emissions factor. When you click on the “Factor Derivation Data” link on the “Emissions Factor Details” page, WebFIRE will return: (1) a list of the individual data values used to calculate the selected EPA-recommended emissions factor, and (2) a list of all the other individual data values contained in WebFIRE that match the original search criteria. For the individual data values retrieved, you are provided with the numeric value, the quality rating of the test report upon which the individual data value is based (see Appendix B), the date that the test was conducted, and a link (labeled “Details”) that allows you to obtain additional background and documentation for a particular individual data value. For example, if the recommended emissions factor you selected was originally obtained from AP 42, clicking on the “Factor Derivation Data” option provides you with a list of all the individual data values used to derive that AP 42 factor and any other data in the WebFIRE system that meets those same search criteria.

Section 8.0

WHAT PARAMETERS SHOULD I CONSIDER WHEN DERIVING A USER-DEFINED EMISSIONS FACTOR?

When you are selecting an emissions factor for use in developing an emissions estimate for a particular process or activity, the primary considerations are:

- How well the emissions factor represents the process for which the emissions estimate is being developed;
- The affect on emissions due to the presence (or absence) of a control device or technique; and
- The underlying test method used to measure the pollutant(s) represented by the emissions factor.

8.1 SOURCE CATEGORY AND PROCESS CONSIDERATIONS

EPA uses Source Classification Codes (SCCs) to classify different types of anthropogenic emissions activities. Each SCC represents a unique source category specific process or function that emits an air pollutant. The SCCs are used as a primary identifying data element in EPA’s WebFIRE, the National Emissions Inventory (NEI), and other EPA databases. The SCCs are also used by many regional, state, local, and tribal agency emissions data systems.

There are two types of SCCs: point and nonpoint. Point source SCCs have 8 digits and follow the pattern 1-22-333-44. The codes use a hierarchical system in which the definition of the emissions process gets increasingly more specific as you move from left to right. The first level of description provides the most general information on the category of emissions. The fourth category is the most detailed and describes the specific emitting process. Nonpoint SCCs represent area and mobile sources emissions and have 10 digits which follow the pattern 11-22-333-444.

The current list of SCCs and their descriptions can be downloaded from EPA’s Emission Inventory System (EIS) website: (<http://www.epa.gov/ttn/chief/eiinformation.html>). At this website, clicking on the link for “EIS Code Tables (including SCCs)” will take you to a

Microsoft Access[®] database that lists various files. Scroll down through the list of files until an entry called “Source Classification Code” is reached. Clicking on that file will reveal the current SCC listing.

EPA is updating and improving the *point* source SCCs. As technologies have changed over the years, EPA has recognized the need to remove out-dated SCCs and add SCCs for new emissions processes. A review of existing SCCs has shown several instances of duplicate SCCs for the same process. Duplicate SCCs are being eradicated to ensure that each emissions process has a unique SCC. In addition, EPA is working to assign SCCs to emission sources which are currently regulated but do not have SCCs. Other changes are being made to ensure that the assignment of an SCC is consistent with the descriptions associated with the hierarchy of digits that comprise each SCC.

The SCC revisions are intended to improve the overall organization of the SCC list by reducing the likelihood of a user choosing an incorrect SCC for their particular process. The SCCs are designed to categorize processes that create emissions. Therefore, another objective of revising the SCCs is to remove the description of control devices from the current SCC list.

Another objective of the SCC revision process is to reduce the use of miscellaneous SCCs, such as those including “99”s. Often these are labeled in the SCC list as “other not classified,” “specify in comments field,” or “miscellaneous.” These types of labels are not sufficient to classify emissions processes. Therefore, EPA intends to remove these SCCs by flagging their use in WebFIRE, thereby informing EPA to assign a new SCC. The new methodology will allow for SCC users to propose new SCC(s) for their emissions processes in an effective and logical way. Upon receipt of a request to establish a new SCC, EPA will perform an analysis to determine if the proposed SCC is unique or if an existing SCC should be used. The analysis will be based upon the uniqueness of the emissions profile of the process and other relevant considerations.

It is important to note that the revisions that are currently being made to the SCC process do not change the fundamental role that SCCs play in the emissions factor program or the way

that users will be able to search for specific emissions factors. These changes will improve the overall data quality of the emissions factors by ensuring that the data upon which the emissions factors are based are grouped in the appropriate SCC. In addition, a cross-walk will be provided so that revised SCCs can be identified by their old SCC number.

8.2 CONTROL DEVICE CONSIDERATIONS

In addition to assessing the production process or activity for which you are selecting or developing an emissions factor, you should have a clear understanding of the operation and performance characteristics of any control techniques or technologies that are used to reduce emissions from the process. When you are selecting or developing a controlled emissions factor, you must determine if the control device reflected in the emissions factor record is comparable to the type and configuration of any control device that is applied to the process for which you are developing the emissions estimate. You may also need to assess whether the pollutant of interest is reduced or eliminated by a particular type of control device, or determine whether a piece of equipment functions as an integral part of the process (e.g., a cyclone that separates product from a pneumatic conveying system, cooling coils in a vapor degreaser that reduce solvent loss) or whether it is a control device (e.g., a cyclone that reduces PM emissions from a wood sawmill, a thermal oxidizer that reduces organic emissions from a process vent). You may also find that a clear understanding of control device operation is useful when assessing the performance of control devices that are operated in series (WebFIRE accommodates up to six control devices for a single emissions factor record).

8.3 POLLUTANT TEST METHOD CONSIDERATIONS

The selection of a test method and how the method is applied to measure emissions from the process can affect the representativeness of the emissions data and the resulting emissions factor developed from the data. The majority of the recommended emissions factors contained in WebFIRE are based upon direct emissions measurements. In most cases, these measurements were obtained using EPA reference test methods that were created to support development, implementation, and compliance with federal standards (e.g., New Source Performance Standards (NSPS), National Emission Standards for Hazardous Air Pollutants (NESHAP)). In

addition, some emissions factors are based upon data collected using non-EPA test methods (e.g., methods developed by the California Air Resources Board (CARB)).

EPA reference test methods provide direct measurement of specific chemical species (e.g., carbon monoxide (CO), sulfur dioxide (SO₂)), emissions from a process or control device. The EPA reference test methods for measuring PM or total hydrocarbons (THC) measure the emissions of a group or class of pollutants rather than an individual compound or chemical species. In these cases, for example, the term “filterable PM” is considered to apply to the material that is captured upstream and on the sampling train filter maintained at a specific temperature. Consequently, the temperature at which the sampling train is operated affects the amount of “filterable” material collected (e.g., operating the sampling train at a lower temperature would tend to capture more compounds that have high vapor pressures).

When you are considering an emissions factor developed from PM or THC data, you should be aware of the underlying test method and conditions under which the test was conducted to determine if the emissions factor is appropriate for the pollutant for which you are preparing the emissions estimate. Often an understanding of how the method is applied can overcome confusion about applying the data and in comparing emissions from different facilities.

Section 9.0 HOW DO I DEVELOP A USER-DEFINED EMISSIONS FACTOR?

9.1 HOW DO I USE WEBFIRE TO CREATE A USER-DEFINED EMISSIONS FACTOR?

WebFIRE allows you to develop a user-defined emissions factor using the same procedures that the EPA follows to develop new or revise existing recommended emissions factors (see Section 5.0). Figure 9-1 shows the steps that you must follow to develop a user-defined emissions factor. First, you must select a set of candidate data values from the collection of individual data values underlying a recommended emissions factor and other related data values (see Section 8.3) by highlighting the check box next to each record. After you have selected the candidate data set, WebFIRE calculates the emissions factor value using the outlier, BDL, factor derivation, and quality assessment tools discussed in Section 5.0. At this time, these development tools are not applicable to the recommended emissions factors that are expressed as empirical equations because they contain more than one variable.

After the user-defined emissions factor has been calculated by WebFIRE, you can generate a report to provide documentation of the emissions factor development (Figure 8-1, Step 6). The report provides a summary of the user-defined emissions factor, the number of data points used to derive the factor, the corresponding SCC for the emissions factor, applicable control devices, the composite test rating (CTR) for the factor (see Appendix E), and how well the emissions factor represents the SCC. The report also shows the values and supporting information for the individual test data points that were used to derive the emissions factor. Because user-defined emissions factors are not retained in the WebFIRE database after they are created, we recommend a report be prepared for any user-defined emissions factor that you develop.

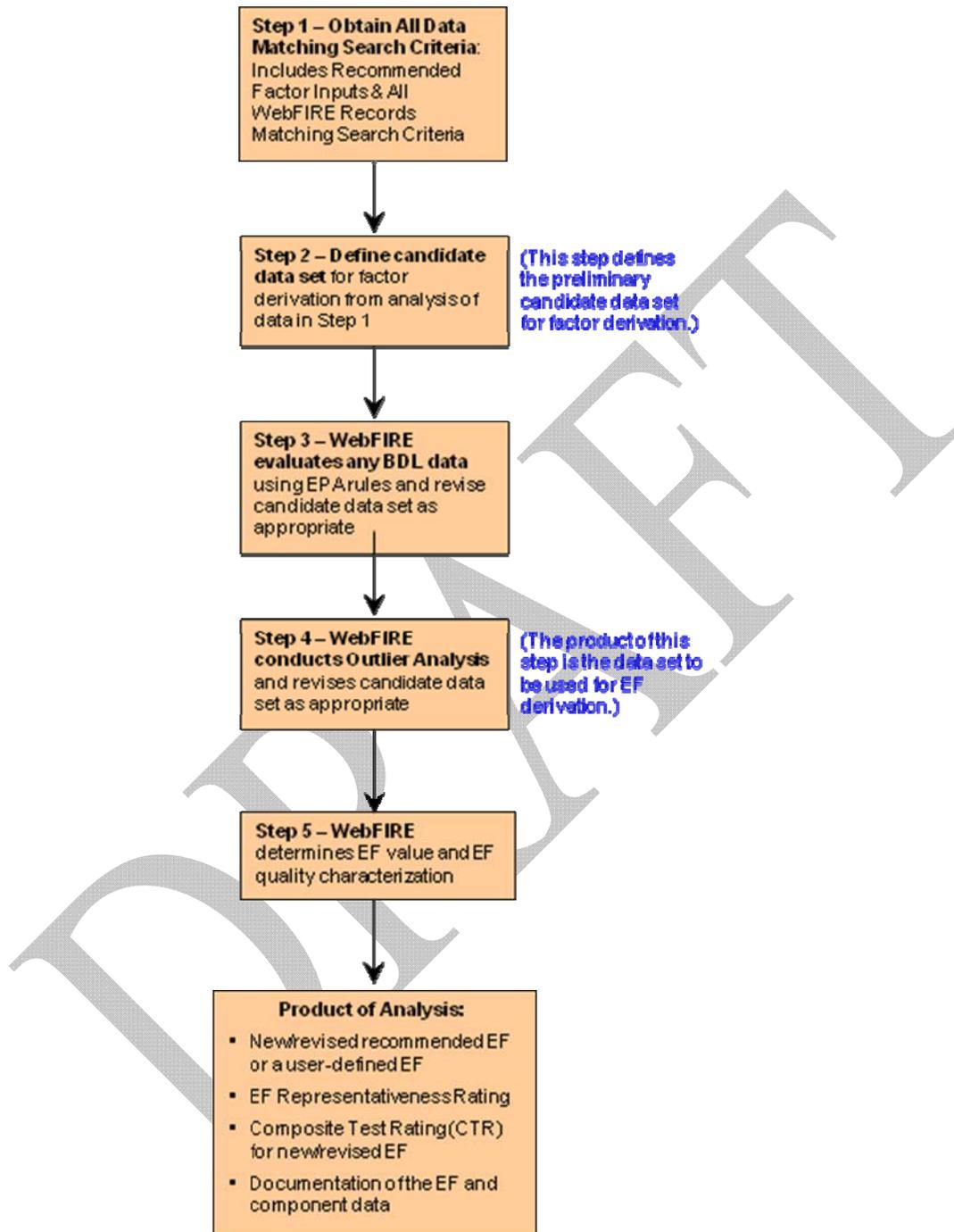


Figure 9-1. Emissions Factor Derivation in WebFIRE

9.2 WHAT ARE THE POTENTIAL IMPACTS ASSOCIATED WITH APPLYING A USER-DEFINED EMISSIONS FACTOR?

Since 1970, the EPA has developed and published air pollutant emissions factors, primarily for use in developing large-scale emissions inventories. As noted in Section 4.0, we recognize that the use of our recommended factors has expanded beyond the original intent of supporting inventory development. For example, recommended emissions factors have been used to prepare site-specific emissions estimates that are used to determine permit or regulation applicability and for emissions trading.

WebFIRE provides tools that allow users to develop emissions factors based upon individual data points selected by the user. Applying a user-defined emissions factor may affect whether or not your source is subject to certain regulations. For example, applying a user-defined emissions factor to a site-specific emissions estimate could show that a facility is not subject to a particular emissions standard where the previous use of a recommended emissions factor indicated that the emissions standard was applicable. For this reason, we encourage you to be judicious and responsible in your application of a user-defined emissions factor. We also encourage you to create and maintain the WebFIRE report (see Section 9.1) that documents the development of the user-defined emissions factor. WebFIRE does not retain user-defined emissions factors in the database after they have been created.

Section 10.0 HOW DO I SUBMIT DATA TO WEBFIRE?

To ensure consistency of data submittals from many different facilities and entities, we require that you submit data to WebFIRE using the EPA's Electronic Reporting Tool (ERT), or the alternative Microsoft® Excel® spreadsheet template:

http://www.epa.gov/ttn/chief/ert/ert_tool.html#webfirespreadsheet. The ERT (see Section 10.1) is an electronic alternative to submitting paper test reports and supporting documentation. After you have completely filled out the ERT (or spreadsheet), you must submit the information through the EPA's Central Data Exchange (CDX) at: <https://cdx.epa.gov/SSL/cdx/login.asp>. The CDX (see Section 10.2) is part of the Environmental Information Exchange Network and provides industry, states, tribes, and other stakeholders a fast, easy, and secure reporting service.

If you have an existing CDX account (e.g., you submit reports for the EPA's Toxic Release Inventory (TRI) Program), you can use your current user name and password to log into the CDX. You will need to follow the instructions provided after the log-in page to obtain approval from EPA to access the WebFIRE Data Upload program. After you obtain approval, CDX will add the data upload program to the list of CDX applications that you routinely use. If you do not have a current account with the CDX, you must complete the on-line registration process at: <https://cdx.epa.gov/SSL/CDX/regwarning.asp?Referer=registration>. After completion of the on-line registration application, EPA signature agreement and verification forms will be sent to you by mail. You must complete these forms and return them to EPA to obtain authorization for access to the CDX. The on-line registration process requires approximately 15 to 20 minutes. Agency approval of the signature agreement and verification forms typically takes 5 to 10 days.

For any questions regarding the CDX, the CDX Help Desk (<http://www.epa.gov/cdx/contact.htm>) is available for data submission technical support between the hours of 8:00 am and 6:00 pm (EST) at 1-888-890-1995 or helpdesk@epacdx.net. The CDX Help Desk can also be reached at 970-494-5500.

10.1 WHAT IS THE ERT AND HOW IS IT USED TO DOCUMENT EMISSIONS TESTS?

The EPA's ERT is a Microsoft Access[®] application developed by the Agency to aid facilities in planning and reporting the results of emissions tests. The ERT replaces time-intensive manual test planning, test data compilation and reporting, and data quality assurance evaluations. When properly applied, the ERT also facilitates coordination among the facility, the testing contractor, and the regulatory agency (e.g., for compliance demonstrations) in planning and preparing for the emissions test. The current version of the ERT, a list of the EPA test methods that are currently supported by the ERT, and guidance on the use of the ERT can be found at: http://www.epa.gov/ttn/chief/ert/ert_tool.html. An alternative to the Access[®] database tool for emissions test data reporting is available in the form of a Microsoft Excel[®] spreadsheet template found at: (http://www.epa.gov/ttn/chief/ert/ert_tool.html#webfirespreadsheet). Information regarding EPA's test methods can be found at EPA's Emission Measurement Center (EMC): <http://www.epa.gov/ttn/emc/>.

The ERT documents the following key information; some of which are required by EPA reference test methods for stationary sources:

- Four-level SCC specification,
- Process data from existing air permits (e.g., process throughput rates),
- Process rate levels during actual testing,
- Descriptions of the source, unit process, and control devices associated with the test,
- Process upsets or malfunctions during testing,
- Process flow diagram,
- Sampling locations,
- Test methods used,
- Deviations made to any test method, and
- Output flow rates and pollutant concentrations.

Figure 10-1 shows the typical steps followed when using the ERT. The ERT consists of: (1) a database application, (2) the project data set (PDS), and (3) a data upload spreadsheet. The application is a Microsoft Access[®] database that contains all of the data input screens, reports, calculations, and other items necessary to create and distribute a test plan and test report. The

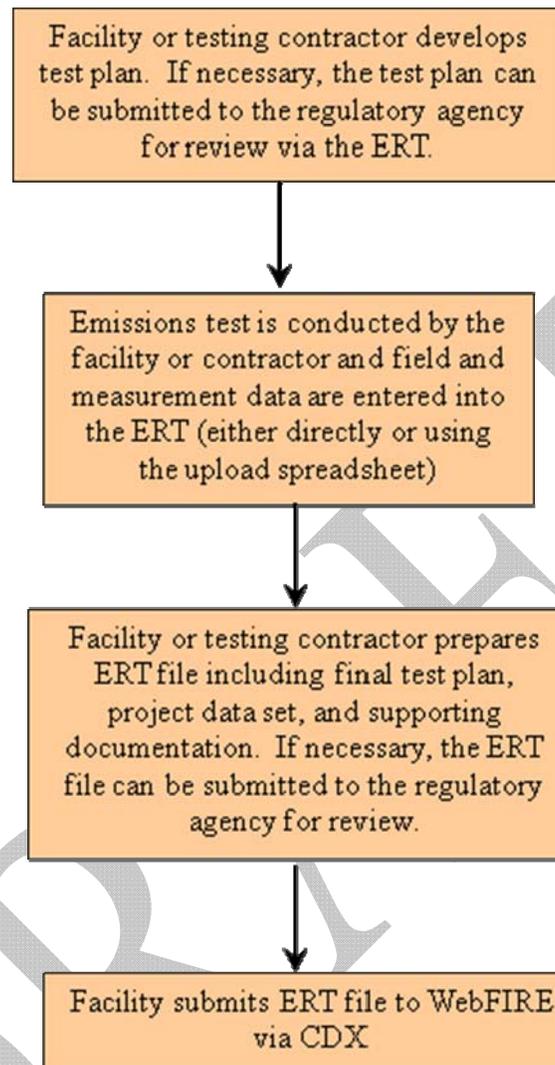


Figure 10-1. Typical Work Flow When Using the ERT

application also incorporates our evaluation system (see Section 5.2 and Appendix B) so that each test is assigned a numeric score that assesses the quality of the measurement data and associated information collected during an emissions test. A standalone version of the application is available that includes a setup routine that installs the ERT application database and the Microsoft Access[®] runtime program. The PDS is also a Microsoft Access[®] database that contains the test plan and measurement data for a single test report. This file is exchanged between the source test contractor, the client, and the regulatory agency, if necessary (e.g., for a compliance test). To provide flexibility to ERT users, the Microsoft Excel[®] spreadsheet can be

used to upload the sampling hardware and field measurement data recorded during a test into the PDS rather than entering the data directly into the PDS through the application.

Upon completion, the ERT contains all of the emissions data and supporting information (e.g., test plan, equipment calibration documentation) prepared and collected for the test. In addition, an electronic copy (PDF) of the entire report documenting the emissions test and other supporting information is attached to the ERT and submitted as a zip file.

The ERT also provides you the ability to create an XML export file for the WebFIRE emissions factor database. The format of this ERT output file is specifically designed to provide inputs for the data fields contained in WebFIRE (e.g., emissions value and units; SCC; ITR). To facilitate incorporation of the data into WebFIRE, the output file is configured to accept emissions values expressed in terms of mass of pollutant emitted per unit of activity. The output file also accepts emissions test results that are expressed as a concentration or an emissions rate (i.e., mass emitted per time unit) subject to further evaluation to determine if the data could be expressed in units that are suitable for use in emissions factor development.

Use of the ERT will provide for consistent criteria to quantitatively assess the quality of the data collected during the emissions test and to standardize the test report contents. The use of the ERT also improves the availability of the supporting documentation necessary to conduct such an evaluation. Additionally, the ERT lays the groundwork for future capabilities to electronically exchange information in the test reports with facility, state, local, or federal data systems.

10.2 WHAT IS THE CDX AND WHAT IS ITS ROLE IN SUBMITTING DATA TO WEBFIRE?

Electronic environmental data submissions to EPA, including submission of emissions data for use in WebFIRE, must be made through the CDX. The CDX is part of the Environmental Information Exchange Network that was developed by the EPA and the states to facilitate online sharing of environmental information among EPA, states, tribes, localities, and other entities. The CDX is a broad-based tool that offers industry, states, tribes, and other

stakeholders a fast, easy, and secure reporting service. As part of EPA's e-government initiative, the CDX helps to ensure that both the public and regulatory agencies can access the information needed to document environmental performance, understand environmental conditions, and make sound decisions to protect the environment.

The benefits of the CDX to the EPA and related program offices include:

- Elimination of redundant infrastructure and its associated costs,
- Facilitation of faster, lower-cost implementation of new or modified data flows,
- Integration of data to Agency data repositories,
- Establishment of consistent procedures for electronic signatures,
- Reduction in the time needed to make information publicly accessible,
- Reduction in the record management costs by elimination of redundant recordkeeping, and
- Compliance with the Cross-Media Electronic Reporting Regulation (CROMERR).

The benefits to the industry, states, local agencies, and tribes associated with the CDX include:

- Reduction of overall reporting burden,
- Improvement in data accessibility,
- Electronic confirmation that information was received and that the electronic form was filled out correctly,
- Reduction in the time and costs associated with environmental data submission requirements,
- Simplification of reporting to a single point in the Agency instead of many separate programs,
- Faster securing of submission through built-in edit and data quality checks,
- Improvement of security and transmission of confidential business information (CBI) through registration and authentication,
- Reduction of burden of complying with new or changing requirements, and
- Streamlining of reporting through the Exchange Network and Web Services.

EPA expects facilities to produce an increased amount of new emissions test data to be provided in response to new regulations that require the submission of emissions tests to demonstrate compliance with federal air regulations. In the recent regulatory proposals (e.g., Standards of Performance for New Stationary Sources and Emission Guidelines for Existing Sources: Commercial and Industrial Solid Waste Incineration Units (CISWI, 75 FR 31938)), the EPA has proposed that source owners and operators collect certain compliance data by using the

ERT and submit the file to EPA through the Agency's CDX network for storage in the WebFIRE database.

In the future, we anticipate that the Agency will use the capabilities of the CDX to provide for electronic exchange of information in test reports with facility, state, and federal data systems. For example, the ERT allows sources to document facility-specific information that may also be required under other regulatory data systems, such as the Air Facility System (AFS). Such systems contain compliance, enforcement, and permit data for stationary sources of air pollution regulated by EPA and state/local/tribal agencies. In the future, this information could be exported to the AFS via the CDX, reducing the burden associated with duplicate entries. Transfers to other data systems such as the National Emissions Inventory, Toxic Release Inventory, and Title V reporting may also be desirable.

DRAFT

Section 11.0
**WHAT IS THE DATA REVIEW AND PUBLIC PARTICIPATION PROCESS FOR EMISSIONS
FACTOR DEVELOPMENT?**

An overview of the public participation and data review process used by EPA when implementing Section 130 for source test and/or emissions factor data is shown in Figure 11-1. The CAA contains provisions that encourage EPA to obtain public participation and review of the development of recommended emissions factors. Section 804 of the 1990 CAA Amendments (CAAA) addressed the issue of emissions factor revisions and public participation by adding Section 130 to Part A of Title I of the Act. Section 130 states that:

“Within 6 months after enactment of the Clean Air Act Amendments of 1990, and at least every 3 years thereafter, the Administrator shall review and, if necessary, revise the methods (emission factors) used for the purposes of this Act to estimate the quantity of emissions of carbon monoxide, volatile organic compounds, and oxides of nitrogen from sources of such air pollutants (including area and mobile sources). In addition, the Administrator shall establish emission factors for sources for which no such methods have previously been established by the Administrator. The Administrator shall permit any person to demonstrate improved emissions estimating techniques, and following approval of such techniques, the Administrator shall authorize the use of such techniques. Any such technique may be approved only after appropriate public participation. Until the Administrator has completed the revision required by this section, nothing in this section shall be construed to affect the validity of emission factors established by the Administrator before the date of the enactment of the Clean Air Act Amendments of 1990.”

Periodically, EPA will review, compile, and analyze the data contained in WebFIRE for the purposes of revising existing and developing new recommended emissions factors, as appropriate. We do not have an established schedule upon which the development of new and/or revised emissions factors will take place. Rather, we will consider the following parameters to determine if emissions factor development is warranted:

- The amount of new source test/emissions factor data that have been received;
- The degree of variability with existing emissions factors in WebFIRE; and
- EPA programmatic needs related to new rules, policies, and other Agency tools.

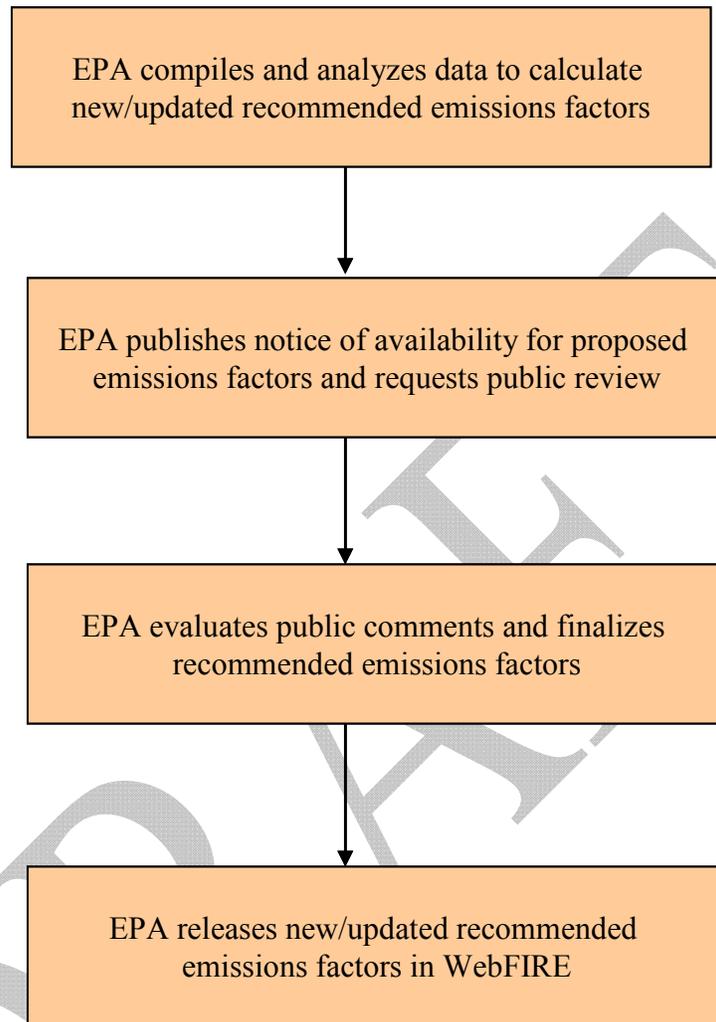


Figure 11-1. Overview of the WebFIRE Public Participation and Emissions Factor Development Process

Section 11.0 What is the Data Review and Public Participation Process for Emissions Factor Development?

If we receive a substantial amount of new information for a given process type and that process is a significant emitter of one or more pollutants, new factor review and development activities could be prompted. If we receive only a few new data points for a process type, it is less likely that the new data alone would initiate the extensive factor review and development process. Another point that we consider is the difference and variability between the existing recommended emissions factors in WebFIRE and the newer data. If the newer data do not significantly change the existing factor(s), the need to revise the factor would be less urgent. Lastly, decisions to initiate factor review and development may be tied to programmatic issues and schedules occurring within EPA. For example, new data or the need for improved emissions factors may be driven by new regulations that are under development or that were recently promulgated. Also, emissions inventory requirements may be in place that will demand new emissions factors.

When one or more of these considerations warrant, EPA will initiate the emissions factor review and development activities. As a result of this process, EPA will propose new and/or revised emissions factors for specific processes (i.e., SCCs). The draft or proposed emissions factors will be flagged within WebFIRE as “Proposed” to identify their status. EPA will publicly announce the availability of these proposed emissions factors and invite public review and comment. The public announcement may take the form of an EPA Listserv email notification (e.g., NEI Listserv, InfoCHIEF Listserv) or, in the event of a large and/or very important release, a formal Federal Register notice. These notifications would describe the nature of the new emissions factors that have been developed and their associated source categories. Typically, the public will have a 60-day review and comment period for the proposed factors. Examples of some topics to consider include, but are not limited to:

- The validity and accuracy of the test methods applied to obtain sample measurements;
- The validity and accuracy of the analytical procedures used to quantify measurements;
- The completeness, thoroughness, and transparency of the source test documentation;
- The correlations made between process parameters and test data conditions;
- The accuracy of the assigned SCC and control device codes; and

Section 11.0 What is the Data Review and Public Participation Process for Emissions Factor Development?

- The adequacy and accuracy of the process description for the source category and the associated documentation.

The process for submitting comments (e.g., format and method of submittal, due dates, submittal address) will be described in the data availability announcements. To facilitate our review of emissions data submitted by commenters, the data must be submitted in electronic format via CDX. We welcome comments on the newly-posted data and on any other aspect of the WebFIRE database, associated emissions test data, and emissions factors at any time. Commenters should review all information pertinent to the correct calculation of emissions factors from the underlying test data. The review should address how well the mass or concentration measurement data were combined with process operating data (e.g., fuel use, material throughput, item production, power output) to yield an emissions factor. If controls are in place, control device operating conditions should be correctly associated with process conditions and factored into the emissions factor development. It is particularly important that reviewers confirm the process and source category associations made for the data. New or revised process flow diagrams and/or schematics should be submitted if an industry has undergone significant changes since the last revision. These process associations should be made using SCCs, recognizing that in, some cases, new SCCs may be required.

At the conclusion of the 60-day review period, EPA evaluates the comments received and makes any appropriate modifications to the data in WebFIRE. If commenters provide new emissions test data for use in emissions factor development, we will consider combining the newer data with the existing data for a given source type or category. When determining valid combinations of existing and new data, we use statistical analyses that are based upon the Student's t-test (see Appendix F). If the comments identify issues or raise questions that EPA cannot address, the original submitter will be contacted for reconciliation. After all comments are appropriately addressed and EPA is satisfied with the quality of the emissions factor data, the "Proposed" emissions factor status flag in WebFIRE is changed to "Recommended" and the previous emissions factor, if any, is flagged as "Revoked." The new and/or revised recommended emissions factors are then made available to the public in WebFIRE (<http://cfpub.epa.gov/webfire/>).

DRAFT

APPENDIX A

GUIDANCE FOR USING EMISSIONS FACTORS FOR NON- INVENTORY APPLICATIONS

DRAFT

1.0 Introduction

It is commonplace for federal, state, tribal, and local agencies and consultants and industry to use published default emissions factors in emissions estimations for a variety of non-inventory uses, including, but not limited to: determining permit and regulatory applicability and emissions reporting for calculating permit fees and assessing compliance determinations. The application of even highly-rated emissions factors may yield inaccurate facility-level emissions estimates due to the inherent variability of emissions factors. Verification of emissions factors used in non-inventory applications through emissions testing can yield multiple benefits for the public and the environment, and reduce the liability of sources.

Emissions factors include inherent variability as these values are averages of individual data points. Published emissions factors are often based upon a particular control technology or application that may not be representative of the specific operation of a particular process at a specific facility. Additionally, because emissions factors are developed from stack test data collected during periods when process equipment were operating in a normal, steady-state condition, using emissions factors for start-up, shutdown or malfunction conditions is inappropriate. As a result, alternative means of estimating emissions are required during these conditions.

While AP 42 includes ratings to provide a general sense of the quality of the emissions factor, issues still exist in the application of even highly-rated emissions factors for facility-specific use. For example, the owner or operator of a facility may decide that, because AP 42 identifies a pollutant of concern as having an A- rated factor, use of an emissions factor would be appropriate in calculating compliance with a *Plantwide Applicability Limit* (PAL) in a permit. The use of an industry average emissions factor to determine compliance with a PAL could lead to decision making based upon inaccurate data. The quality rating is primarily used as an indicator of how representative the emissions factor is in estimating the national average based upon a specific population of sources. The quality rating does not provide an indication of the reliability of predicting single source emissions within a stated accuracy.

Furthermore, emissions factor development is a dynamic process that is continuously evolving to allow for improvement. Published emissions factors are subject to revision based upon changes in industrial processes, measurement techniques and referenced test methods, and the availability of additional test data or further critical review of previously-published data. Obviously, changes to an emissions factor will impact any calculations based upon the emissions factor and may have an impact on decisions that were made based upon those calculations (e.g., whether a specific source meets the major source threshold for a permit, whether additional risk analyses should be conducted, whether additional permit fees are due). Depending upon the decisions previously made and the outcome of new calculations using revised emissions factors, additional administrative and technical costs may result (e.g., costs associated with revising permit limits and conditions, rerunning risk models).

Because of these implications, site-specific emissions factors developed from data obtained by conducting field tests are preferred over the use of AP 42 factors for facility-level applications. Source owners or operators can gain benefits through site-specific emissions

testing to obtain data in lieu of reliance on AP 42 or other emissions factors that are not site specific.

Through the emissions testing verification process, source owners or operators are able to reduce potential liability issues that exist due to circumstances where emissions factors are inappropriate for use by a particular operation. In cases where measured emissions rates are lower than those calculated using published emissions factors, facilities may also see financial benefits through lower permit fees or through the creation of excess emissions credits that can be sold in emissions markets. The following sections summarize some non-inventory uses of emissions factors, general approaches to consider when applying emissions factors for non-inventory purposes, and the potential results and implications of emissions factor use for non-inventory applications.

2.0 Approaches to Consider

Several alternative approaches to using the published emissions factor should be considered for facility-level applications other than inventory development. The two basic approaches include: (1) evaluation of the data comprising the emissions factor data set and selection of a parameter other than the mean (emissions factor) from the data set; and (2) site-specific emissions testing. Each of these approaches is briefly described in the following sections. General criteria to consider in deciding whether to use the published emissions factor or to consider these other approaches include:

- The perceived representativeness of the emissions factor for the source emissions being calculated (e.g., representativeness relative to fuel type, process design and operation)
- The importance of accuracy for the emissions being calculated using the emissions factor relative to the overall emissions calculation and purpose (e.g., based upon the available information, is the source a relatively minor or a major contributor of emissions in a calculation of total emissions for a PAL?).

2.1 Evaluation and Utilization of Emissions Factor Data Set Parameters

Depending upon the intended use of the emissions factor, it may be desirable to use a parameter of the emissions factor data set other than the mean value. For example, selection of the maximum value in the data set contained in WebFIRE, in lieu of the emissions factor (mean value) for input into a risk model will provide a more conservative risk estimate (extra margin of safety). Two data parameters to consider for use are the upper range of the data and an upper percentile/rank. Table A-1 presents an example emissions factor data set for benzene emissions from drum mix hot mix asphalt plants. The data set is comprised of 19 emissions tests and the emissions factor is A-rated. The published emissions factor is 0.00039 lb benzene per ton of asphalt loaded.

2.1.1 Upper Range of Data

In this approach, the user estimates the emissions from the source based upon the maximum test value contained in the emissions factor data set. Without site-specific testing, this approach can yield a very conservative emissions estimate when compared to the average emissions factor. For the example in Table A-1, the value used would be 0.0012 lb benzene per ton of asphalt loaded, or 3 times the published emissions factor.

Table A-1. Benzene Emissions Drum Mix Hot Mix Asphalt Plants

Test No.	Emissions (lb/ton of asphalt loaded)
1	0.000063
2	0.000092
3	0.00012
4	0.00012
5	0.00015
6	0.00022
7	0.00026
8	0.00027
9	0.00029
10	0.00030
11	0.00036
12	0.00038
13	0.00040
14	0.00041
15	0.00044
16	0.00056
17	0.00069
18	0.0011
19	0.0012

Reference: AP 42, Chapter 11.1

2.1.1 Nearest Rank/Percentile

As an alternative to using the maximum test value reported in the development of the emissions factor, you could use a percentile value based upon the ranking of the emissions test data. For percentiles over the 50th percentile, this approach is not as conservative as using the maximum value, but it is more conservative than using the mean emissions factor. The term “percentile” is used to define the proportion of data points that fall under the selected value. Equation 1 can be employed to identify the data value that corresponds to a specified percentile value.

$$\text{Equation 1: } n = \frac{N}{100}p + \frac{1}{2}, \text{ rounding to the nearest integer.}$$

Where: n = nearest rank (e.g., 5th value).
 p = percentile value (e.g., 80th percentile).
 N = number of ordered values (i.e., number of data points).

Under this approach, you would select the percentile to be used based upon the data set and the intended application of the emissions factor, as well as any available information that might be used to select the value deemed to be representative of the operation of the source. For example, using the data set in Table A-1, the data value corresponding to the 80th percentile is calculated using Equation 1 as follows:

$$n = (19/100) * 80 + 1/2 = 15.7 = 16^{\text{th}} \text{ value}$$

The 16th value is 0.00056 lb benzene per ton of asphalt loaded, corresponding to the 80th percentile for this data set. This value is approximately 1.5 times the mean value. Similarly, for the data set in Table A-1, the 90th percentile is calculated using Equation 1, as follows:

$$n = (19/100) * 90 + 1/2 = 17.6 = 18^{\text{th}} \text{ value (rounded to closest integer).}$$

The 18th value is 0.0011 lb benzene per ton of asphalt loaded which is approximately 2.8 times the mean value for this data set.

For certain applications, such as the use of emissions factors for emissions trading or offsetting, it may be appropriate to use a value taken from the lower end of the emissions factor range. For the data set in Table A-1, the 20th percentile is calculated using Equation 1, as follows:

$$n = (19/100) * 20 + 1/2 = 4.3 = 4^{\text{th}} \text{ value.}$$

The 4th value is 0.00012 lb benzene per ton of asphalt loaded which is approximately 1/3 the mean value for this data set.

3.0 Site-Specific Emissions Testing

Site-specific emissions testing offers several advantages to using published emissions factors including, but not limited to:

- Information gathered on the actual emissions from the source under representative operating conditions;
- Increased confidence that the reported emissions are representative of actual emissions; and
- Ability to update and improve emissions estimates based upon periodic retesting.

Where multiple emissions tests are conducted over a period of time (e.g., annual tests), these data provide information that can be used to develop a site-specific emissions factor (i.e., an average

emissions rate over time). This site-specific emissions factor can provide a means to monitor changes in emissions that may be indicative of improvements in performance due to operational changes or, conversely, deterioration of performance due to control device degradation or within-source variability of the process and associated controls.

The obvious disadvantage of site-specific emissions testing is the associated cost. Consequently, the decision to conduct site-specific emissions testing should be considered in the context of the relative importance of the emissions estimate to the intended use of the data (e.g., risk assessment, permit applicability determination, establishing a permit limit). Factors to consider include:

- Availability of highly-rated emissions factors that are expected to be representative of the source emissions;
- Relationship of the emissions from the source to which an emissions factor is being applied to the total emissions being calculated for the overall use (e.g., based upon available information the particular source is expected to represent less than 10 percent of the facility's emissions for a PAL calculation); and
- Consequences of erring.

The frequency of testing is another factor to consider. A single test may be sufficient to confirm that the emissions from the source do, in fact, represent less than 10 percent of the total facility emissions for a PAL. On the other hand, multiple, periodic tests might be appropriate to establish a representative site-specific emissions factor where a facility applies for and is granted a permit as a synthetic minor, but where the calculated emissions approach the major source threshold (e.g., the calculated emissions are 90 percent of the major source threshold). In this case, it may be appropriate to conduct periodic emissions testing to develop a site-specific emissions factor and confirm the facility continues to operate below the major source threshold on a routine basis.

The decision regarding when to test and how often to test encompasses many factors and is not subject to firm criteria nor does it follow a specific decision tree. The following sections provides some example scenarios where emissions factors are applied to a non-inventory use and where use of other emissions factor data parameters and/or emissions testing may be appropriate.

3.1 Approaches for Using Emissions Factor Parameters and Site-Specific Testing For Estimating Emissions for Non-Inventory Purposes

This section discusses some of the potential implications to sources, permitting agencies, and the environment when using published emissions factors for a number of common site-specific, non-inventory applications. Approaches that may be considered to minimize the potential impacts also are presented.

3.1.1 Permitting: Title V Applicability

An implication of using a default emissions factor to determine whether the requirement of a Title V permit is applicable to a source is that the actual emissions are underestimated and the owner or operator incorrectly determines that a permit is not required. A facility might find itself subject to Title V if subsequent emissions testing conducted at the facility indicates that emissions exceed an applicability threshold but a Title V permit application was not submitted by the facility. On the other hand, if a facility uses a default emissions factor that overestimates its actual emissions, the facility may be required to submit a Title V permit application unnecessarily, adding an administrative burden to both the facility and the permitting agency. In instances where a published emissions factor is revised, resulting in changes to emissions estimates for a facility, uncertainty and confusion with respect to the accuracy of the emissions estimate and whether a Title V permit is required may occur. Use of a site-specific emissions factor based upon emissions testing becomes more critical as emissions estimates based upon default emissions factors approach applicability threshold limits. An accurate emissions estimate via site-specific emissions testing provides assurance of correct major or minor source determinations. In addition, use of the 80th percentile of the emissions factor data, for the activity or process of interest, in lieu of the emissions factor can help minimize these negative consequences.

3.1.2 Permitting: Establishing Plant-wide Applicability Limits

The need for a facility to more accurately assess its emissions under a PAL stems from the flexibility which is provided by the PAL. Where default emissions factors overestimate actual emissions from the facility, the owner or operator is foregoing potential emissions increases from process or activity changes that cannot be performed if the facility is nearing the emissions cap. Alternatively, if an emissions factor underestimates actual emissions, the facility may be in danger of exceeding the facility PAL. In certain cases, even the upper range of the published AP 42 or WebFIRE emissions data may underestimate the actual emissions of an activity, and subsequent, more accurate data obtained by emissions testing may severely impact the ability of a facility to maintain the desired flexibility under a PAL. Limiting the use of emissions factors to sources included in a PAL to those sources that are estimated to be less than 10 percent of the total PAL emissions can help improve source flexibility under a PAL. In addition, in the absence of site-specific data, you could utilize the 80th percentile of the emissions factor data for the activity or process of interest to estimate the emissions for establishing the PAL. For emissions sources that are estimated to comprise more than 50 percent of the PAL, site-specific emissions testing is an alternative approach that can provide increased flexibility to owners and operators of sources that must comply with the PAL.

3.1.3 Permitting: NSR/PSD Applicability

According to the NSR Workshop Manual (Draft 1990), the use of emissions factors is identified as an approach for calculating facility emissions. As discussed above with respect to Title V applicability determinations, accurate emissions estimates provide greater assurance of making the correct applicability determination and can reduce unnecessary administrative burdens associated with the use of emissions factors that differ from actual emissions. One

approach to minimizing these concerns is to utilize the 80th percentile of the emissions factor data for the activity or process of interest. Another approach is to incorporate a requirement into the NSR/PSD permit to conduct site-specific emissions testing to verify the emissions level. This latter approach is particularly useful for emissions sources that are estimated to comprise more than 50 percent of the total facility emissions.

3.1.4 Risk Assessment (Air Toxics)

Risk assessment applications may include regulatory development, applicability determinations, and risk management planning. Typically, risk assessors use emissions factors as a screening tool to determine order-of-magnitude risk potential. For example, a regulatory agency may use emissions factors in models to determine if a state air toxic program applies to a facility based upon modeled fence-line concentrations. However, the results based upon such modeled concentrations may result in the improper determination of risk. An overestimate of emissions based upon a higher emissions factor than actual operations may result in a facility unnecessarily being required to take corrective actions to address the risk, such as the installation of additional pollution controls. One approach to minimizing this concern is to utilize the 95th percentile of the emissions factor data. While the upper range of the data may provide the most conservative value for a risk assessment, the use of the 95th percentile would eliminate the most extreme values. For those sources for which the estimated risk is nearing the unacceptable threshold level, emissions testing or ambient monitoring could be employed to provide the most accurate estimates of risk levels.

3.1.5 Emissions Trading Programs

Typically, accurate emissions estimates obtained by continuous emissions monitoring, mass balance calculations, or site-specific emissions factors are needed for emissions trading programs. For low-emitting units, emissions factors may be allowed by the program. However, in this case, a “discount” (i.e., an adjustment for uncertainty) is typically applied to both the credit being taken and the offset being purchased to account for the uncertainty. Another approach to reducing the uncertainty is to develop site-specific emissions factors via testing, except in cases where emissions factors based upon material balances are used.

APPENDIX B

PROCEDURES FOR DETERMINING INDIVIDUAL TEST REPORT QUALITY RATINGS

DRAFT

1.0 Introduction

Historically, EPA's quality reviews of emissions test data and test reports were largely subjective because each test program presented different issues (i.e., no two facilities, or the tests conducted at those facilities, are exactly alike). Typically, EPA developed quality ratings (letter grades of A through D) for test reports based upon the Agency's review of the following criteria areas:

- Source operation;
- Test method and sampling procedures;
- Process information; and
- Analysis and calculations

To reduce the subjectivity of quality reviews, the individual test rating (ITR) assigned by the Electronic Reporting Tool (ERT) is based upon responses to questions that assess the quality of the process, control device, and emissions data collected during a source test. The methodology used by the ERT for assessing the quality of emissions test data follows the same basic principles as EPA's historic methodology. However, the new ERT procedure provides a consistent objective framework for test contractors to follow when collecting data, and for independent reviewers to follow when assessing data quality.

The test report quality rating methodology consists of two components: (1) the source test review, and (2) the state regulatory agency review. Table B-1 shows the questions that are used to evaluate the quality of data submitted to the ERT. The information requested in the table is indicative of a complete and well-documented test report. The documentation is verified by the ERT and potentially by the state regulatory agency. It is important to note that state review is not required; however, the review can improve the quality rating of the test report.

2.0 Source Test Review

For the source test component (usually administered by the testing contractor or the facility), the score is calculated by ERT based upon report completeness and includes criteria related to: process data, control device information, test method performance, and quality assurance. The maximum possible score for a test report that is not reviewed by the state is 75 points. Rather than force the question scores to total 75 points, the source test review score is prorated using the following approach:

Prorated score for test method-related questions = (total points awarded)/(max total points)*(75);

3.0 State Review

The quality of an emissions factor is only as good as the source data upon which it is based. In the majority of cases, the test report, which is typically prepared by the testing contractor, is the only documentation available for evaluating whether emissions data are

suitable for use in emissions factor development. In all cases, the quality of the underlying source data is enhanced when the test report is reviewed by a third party, in this case the state.

Under the ERT quality rating procedure, the state reviewer evaluates the responses to certain questions (shown in Table B-1) contained in the QA Review section of the ERT. If the reviewer determines that the information requested by the question is present in the ERT file and is correct, additional points are added to the score given by the ERT. If the state reviewer determines that points were incorrectly assigned (i.e., the information contained in the ERT file is incomplete or erroneous), points are deducted from the original score. The facility can address the concerns identified by the state reviewer and resubmit the ERT file to the CDX. The net points awarded as a result of the third-party review are prorated to a 25-point basis, as follows:

$$\text{Prorated score for state review} = (\text{net total points awarded})/(\text{max total points})*(25).$$

Therefore, the maximum quality rating for a test report that is reviewed by the state is 100 points.

DRAFT

Table B-1. Test Report Quality Rating Tool

Criteria Area	Review Questions	ERT Scoring	State Reviewer Scoring	
		Points for "Yes"	If "Yes," Add . . .	If "No," Deduct . . .
Completeness Review	Is a description of test location provided?	6		
	Is a drawing of the test locations provided?	6		
	Has a description of test methods used, including deviations from standard procedures, been provided?	6		
	Has a detailed discussion of sampling conditions been provided?	3		
	Is a schematic of each sampling train attached?	3		
	Is a full description of the facility provided?	2		
	Were the operating parameters for the process being tested reported?	5		
	Are the parameters monitored by the facility to assess the performance of the control device described?	5		
	Have all emissions tests specified in the test plan been performed?	3		
Calibration Reports	Are the following reports attached?			
	<i>Manual Test Methods</i>			
	Dry gas meter pre-test calibration	6		
	Dry gas meter post-test calibration	6		
	Thermocouple and display calibration	4		
	Pitot inspection records	6		
	Nozzle inspection records	4		
	Flow meter calibrations	6		
	<i>Instrumental Test Methods</i>			
	Instrumental method analyzer interference tests	5		
	Response time tests	6		
	System bias tests	6		
	Converter efficiency tests	6		
Raw Data Reports	Are the following reports attached?			
	<i>Manual Test Methods</i>			
	Method 1 sample point evaluation	5		
	Cyclonic flow checks	5		
	Exhaust gas conditions	5		
	Raw sampling data	5		
	Field notes	5		
	Description of the recovery procedures	3		
	Complete laboratory report	5		
Chain of custody forms	2			

Table B-1. Test Report Quality Rating Tool (Cont.)

Criteria Area	Review Questions	ERT Scoring	State Reviewer Scoring	
		Points for "Yes"	If "Yes," Add . . .	If "No," Deduct . . .
	Flow diagram of sample recovery	1		
	Flow diagram of sample analysis	1		
	<i>Instrumental Test Methods</i>			
	Stratification checks	6		
	Raw data	6		
	Sampling system flow and temperature logs	6		
	Calibration gas certifications	5		
	Calibration reports	5		
	Complete description of the sampling system	3		
	<i>Process & Facility Operation</i>			
	Process operating parameter data	5		
	Control device operating and monitoring parameter data	5		
	A detailed discussion of the process and control device operation	5		
	Quality Assurance (QA) Review	<i>Manual Test Method QA</i>		
Was the DGM pre-test calibration within the criteria specified by the test method?			1	2
Was the DGM post-test calibration within the criteria specified by the test method?			1	2
Were thermocouple calibrations within method criteria?			1	1
Was pitot inspection acceptable?			1	1
Were nozzle inspections acceptable?			1	1
Were flow meter calibrations acceptable?			1	2
Did the cyclonic flow evaluation show acceptable flow?			1	2
Were the appropriate number and location of sampling points used?		4		
Were pre- and post-test leak checks performed?			2	2
Were leak-checks performed each time the sample port was changed?			1	1
Was the correct procedure used to perform the leak-check?			1	1
Was the entire sampling train leak-checked?			2	2
Did post-test leak checks meet method requirements?			2	3
Did probe and filter temperatures meet method criteria?		5		
Did iso-kinetic sampling rates meet method criteria?		5		
Was the required minimum sample volume collected?		5		
<i>Laboratory QA</i>				
Was the recovery process consistent with the method?		1	2	

Table B-1. Test Report Quality Rating Tool (Cont.)

Criteria Area	Review Questions	ERT Scoring	State Reviewer Scoring	
		Points for "Yes"	If "Yes," Add . . .	If "No," Deduct . . .
	Were field blanks collected?		1	1
	Were recovered sample volumes measured and recorded?		2	2
	Were received sample volumes measured and recorded?		2	2
	Was there a loss of sample during shipping?		2	3
	If sample loss occurred, was an approved method used to compensate for the loss?		2	3
	Was sample pH checked and recorded?		1	2
	Was pH within method specifications?		2	2
	Were recovered sample fractions consistent with the method?		2	3
	Was the laboratory certified/accredited to perform these analyses?		2	3
	Does the laboratory report document the analytical procedures?		1	1
	Was the analytical technique the same as specified in the test plan?		1	1
	Was the analytical detection limit specified in the test plan met?		1	2
	Is the detection limit adequate for the purposes of the test program?		2	3
	Were any problematic analytical conditions encountered during the test?		1	2
	Were all laboratory QA requirements performed?		2	3
	Were method required analytical standards analyzed?		2	3
	Were samples maintained within the proper temperature ranges?		1	2
	Were sample hold times within method requirements?		2	2
	Were laboratory duplicates within acceptable limits?		2	2
	Were spike recoveries within method requirements?		2	3
	Were method specified analytical blanks analyzed?		1	1
	<i>Instrumental Test Method QA</i>			
	Did calibration standards meet method criteria?	5		
	Have the calibration standards expired?	5		
	Did system bias checks meet method requirements?	5		
	Was the NOX converter test acceptable?	5		
	Did interference checks meet method requirements?	5		
	Did the analyzer meet pre-/post-test bias and drift requirements?	5		
	Were instrument flow rates within limits?		2	3
	Were recorded test trailer temperatures within limits?		2	3
	Was the duration of each sample run within method criteria?		1	2
	Did sample point times meet the method requirements?		1	2

Table B-1. Test Report Quality Rating Tool (Cont.)

Criteria Area	Review Questions	ERT Scoring	State Reviewer Scoring	
		Points for "Yes"	If "Yes," Add . . .	If "No," Deduct . . .
	Was stratification present?		2	3
	Was a traverse performed during sample collection?		2	2
	<i>Process Data</i>			
	Were process monitors calibrated?	5		
	Was process data concurrent with testing?	5		
	Was the process stable during testing?	5		
	<i>Other QA Indicators</i>			
	Was a Qualified Source Test Individual (QSTI) on-site and in charge of the field team?		1	1
	Was the test team familiar with the test methods and understood procedures?		1	2
	Was a representative of the regulatory agency on-site during the test?		1	2

DRAFT

Some of the information requested in Table B-1 is specific to certain test methods. For example, nozzle inspection records (listed under “Calibration Reports”) are only applicable when the test method requires the use of a nozzle. In cases like this, a test report will not be given a lower rating if the test method used does not require the use of a nozzle. Instead, quality ratings are dependent upon the testing requirements. For example, if an instrumental test method is used, only those questions that pertain to the method will be used to evaluate the quality of the test. Because the overall score is normalized based upon the maximum score that can be assigned for any given method, the fact that some questions are not scored does not reduce the overall maximum score possible for one test method relative to another method.

4.0 Rationale for Evaluation Criteria

The rationale for including the specific information considered in calculating the ITR are provided below.

1. Completeness Review – The documentation requirements specified in the “Completeness Review” are used to assess certain aspects of the test program impacting the quality (e.g., accuracy, precision, reliability) of the test data. A complete test report should include: information typically contained in a test plan, identification of the facility, a description of the test location(s), a schematic or drawing of the test location(s), and the test methods to be used. Documentation of the conduct of the test methods, deviations from required test methods, and laboratory reports describing the analysis of the test samples are indicative of the precision and accuracy of emissions data. The conditions during the time of sampling and the operating parameters for the process and any air pollution controls are indicative of the reliability and representativeness of the emissions measured during the test period. If the various pieces of information listed here are not provided, conformance to the test method cannot be determined and the precision and accuracy of the data cannot be verified.
2. Calibration Reports – Calibration reports provide documentation that equipment has been inspected, properly maintained, and is operating correctly during testing. If no calibration data are present, or if the calibration data have expired, the results of testing cannot be considered accurate. Calibration errors will lead to inaccurate measurements and therefore inaccurate emissions rates.
 - Manual Test Methods – Equipment used to measure flow rate and temperature must be properly inspected and calibrated to ensure accurate results. Flow rate and temperature are important factors in source testing and have a direct impact on the calculation of emissions rates. Faulty or mis-calibrated equipment can lead to inaccurate readings and inaccurate results.
 - Instrumental Test Methods – Similar to the manual methods, this information is used to determine if analyzers are operating correctly for each test. This data includes pre-test calibration checks, bias determinations for each test run, and equipment operational checks. If the information in this section is missing, the data contained in the test report cannot be considered accurate.

3. Raw Data Reports

- Manual Test Methods – The documentation in this section verifies information reported in the test program and confirms that field QA activities have been performed. This section provides documentation of stack characteristics, exhaust gas conditions, and sample point evaluation, all of which are important for properly characterizing emissions. A complete laboratory report, recovery procedures, and chain of custody forms give a good indication of how well the samples were recovered, handled, and analyzed.
- Instrumental Test Methods – With the exception of raw data, this information is required by the reference methods and is used to verify that operating limits for instrumentation are within acceptable ranges. Stratification checks are now required by the EPA reference methods in some instances and this documentation verifies that sampling procedures were appropriate for the exhaust conditions at the time of the test.
- Process & Facility Operation – Process and operating data are key components in demonstrating that the facility is operating within normal conditions and that the data collected are representative of normal operation. This information also allows for the calculation of production-based emission factors. Documentation of control devices and their monitoring parameters verifies that devices are working properly, provides information that can later be used as indicators of continued performance, and assures that testing was done under typical control conditions.

4. Quality Assurance Review – The evaluation criteria listed in the next five sections are based upon the QA requirements of the reference methods, NSPS, and NESHAP.

- Manual Test Method QA – Calibration criteria evaluated in this section are specified in the reference methods and address field measurement equipment calibrations and inspections. These criteria establish the minimum operating limits for measurement equipment that provide confidence in the accuracy and precision of the test results. This information addresses the critical elements of the test equipment that have a direct impact on measurement and subsequent calculation of sample volumes, effluent flow rates, and pollutant concentrations.
- Laboratory QA – Laboratory information evaluated in this section is directly related to the accuracy of the laboratory analysis of pollutant samples collected in the field. Listed items have a direct impact on the analysis of the samples and the reliability of the test data. For example, sample integrity during transport is assessed by comparing sample volumes to the recorded values prior to shipping, which may indicate potential loss of sample media. Another example is analytical detection limits, which must be sensitive enough to measure the pollutant of interest at concentrations appropriate for the test plan.
- Instrumental Test Method QA – Instrumental test methods have specific QA checks specified in the reference methods. These checks are designed to demonstrate that the sampling system and analyzers are:

- i. Capable of meeting minimum acceptance criteria for acquiring a representative effluent sample, and
- ii. Operating in a stable environment.

This information verifies that the analytical accuracy and precision of the measurement results are acceptable for regulatory programs.

- Process Data QA – The evaluation criteria listed here are based upon the instrumental test method evaluations for data accuracy and representativeness. Process disruptions may have a negative impact on the accuracy of the data. Calibration information establishes the reliability and accuracy of the values used to calculate emission rates.
- Other QA Indicators – Among other factors that will increase the assurance of high-quality data from a source emissions test is the participation of qualified individuals during the field testing. A qualified source testing individual (e.g., someone recognized by the Source Evaluation Society or meeting the criteria outlined in ASTM D7036-04) is someone who has demonstrated a high level of knowledge and ability consistent with an experienced field test team leader responsible for emissions test planning, preparation, conduct, and reporting. Another factor is the presence of a qualified observer during the field emissions testing. Such an observer may be an independent technical expert or a representative of the state, local, or federal agency familiar with source emissions testing and who was on site during the test to monitor progress.

APPENDIX C

PROCEDURES FOR HANDLING TEST DATA THAT ARE BELOW THE METHOD DETECTION LIMITS

DRAFT

1.0 Introduction

In some cases, the result of a process emissions test is not an emission rate, but a determination that the target pollutant was not present at or above the minimum detection limit of the test method (MDL). We define an MDL as the minimum concentration of a substance that can be measured and reported with a given level confidence that the analyte concentration is greater than zero. The MDL is determined from an analysis of a sample in a given matrix containing the analyte. For purposes of this program, that level of confidence is 99 percent. Stated another way, the MDL is the smallest amount of a substance that an analytical method can reliably distinguish from zero, at a specified confidence level, from the signal produced by a blank sample.

It is important to understand that the MDL is a statistical parameter and not a chemical one. An MDL can vary from substance to substance and from measurement process to measurement process. Variability is introduced into MDLs by the analysts conducting the measurements, the equipment and chemicals used in the measurements, and the QA/QC procedures used. A separate MDL should be generated for each run of a test program. After MDLs have been developed, the results of the testing can be compared. Results that are less than the MDL are referred to as below the MDL (BDL).

2.0 Description of Procedures

We have developed specific procedures that are to be used when some or all data collected during an emissions test are BDL, and are possibly to be included in the candidate data set used for developing emissions factors. Note that MDLs are to be determined prior to conducting any data outlier tests so that appropriate values can be assigned for BDL data when they are used in outlier analyses.

It is not unusual for environmental data to contain some values that are below the detection limits that can be achieved by current analytical techniques. Because such values are expected, data users have developed calculation techniques to account for these BDL values that exist but are difficult to quantify with the accuracy typically associated with values found above MDLs. Generally, these calculation techniques recognize that small and large sample sizes do not warrant rigorous mathematical approaches to provide a numerical value that replaces a value found to be BDL. On the other hand, medium sample sizes warrant mathematical approaches that provide numerical values associated with a maximum likelihood estimator (MLE), a value found via calculation to be between $\frac{1}{2}$ the MDL and the MDL.

These approaches work well for programs managed by other Agency offices tasked with establishing regulatory emissions limits and determining compliance for specific individual facilities in narrow source categories. However, such rigor is overly complicated for the WebFIRE emissions factor development program because emissions factors are, by design, representative of generic facilities in broad source categories. As a result, the procedures adopted for handling non-detect data in the derivation of emissions factors are more straightforward and are based upon two general principles. First, as emissions test values generally represent the average of three test runs, a data set containing more than 10 test values is

based upon more than 30 individual samples. Such a data set is important, for according to the central limit theorem, as one obtains 30 or more individual samples, the distribution of those samples approaches that of a normal distribution, whose statistical characteristics are obtained readily. Second, the use of actual measured data is preferred over the use of BDL data when an adequate amount of measured data is available. This generally reduces the uncertainty associated with emissions factors derived, in part, from data that are BDL.

In understanding the recommended BDL data procedures, note that a run refers to the net period of time during which an emissions sample is collected, as well as to the amount of pollutant emitted during that time period. Likewise, a test refers to the net period of time over which separate runs, typically three, are conducted, as well as to the average amount of pollutant emitted over the test period. When a test produces all non-detect values, that information will be flagged in the ERT as being BDL; the ERT will also require that the MDL be provided. For purposes of emissions factors development in WebFIRE, values identified as BDL will be handled as follows:

1. When the candidate data set contains only BDL values, WebFIRE will return the code “BDL” and identify in parentheses the highest MDL in the data set.
2. When the candidate data set has fewer than 11 values above the MDL, WebFIRE will replace the data values identified as BDL with values equivalent to $\frac{1}{2}$ their MDL. If a replacement value exceeds the highest test value, WebFIRE will not include that replacement value in calculating an average emissions factor.
3. When the data set has 11 or more values above MDL, WebFIRE will calculate an average emissions factor using only the values above MDL.

The basic guidance for handling BDL test data in the ERT or evaluating BDL data in WebFIRE for use in emissions factor development is summarized below in Table C-1.

Table C-1. Summary of WebFIRE Procedures for Handling Test Data That are Below Detection Limits

Types of Data	Basis for Emissions Factors
All candidate data are BDL	No average emissions factor is determined; the emissions factor is reported as BDL and the highest MDL is provided in a comment field.
Candidate data contains BDL data and actual individual facility emissions factors; after removing superfluous BDL data, total number of values is 10 or fewer	Emissions factor average is calculated using the test values and $\frac{1}{2}$ the MDL for all BDL data, provided that $\frac{1}{2}$ the MDL is equal to or less than the data set’s highest test value. When $\frac{1}{2}$ the MDL is greater than the highest test value, that BDL value is excluded.
Candidate data contains BDL data and actual individual facility emissions factors; total number of actual individual emissions factor values is 11 or greater	All BDL data are excluded from consideration; only actual individual facility emissions factor values are used.

The following examples illustrate WebFIRE's procedures for handling data that are MDL when calculating emissions factor.

Example 1

Table C-2 shows a candidate data from WebFIRE in which all values are BDL. If, as shown in Table C-2, a WebFIRE search for candidate data to use in calculating an emissions factors returns a data set in which all values are BDL, WebFIRE will not determine an average emissions factor value or a quality rating. Rather, WebFIRE will return the following information: "BDL and the highest MDL in the data set is 88 mg/kg."

Table C-2. Example Data Set A

Test No.	Test Value	Test MDL
1	BDL	10 mg/kg
2	BDL	12 mg/kg
3	BDL	70 mg/kg
4	BDL	20 mg/kg
5	BDL	88 mg/kg
6	BDL	38 mg/kg

Example 2

Table C-3 shows a candidate data set that consists of a mix of data that are above the MDL and data that are BDL.

Table C-3. Example Data Set B

Test No.	Test Value	Test MDL
1	19 mg/kg	--
2	16 mg/kg	--
3	BDL	70 mg/kg
4	11 mg/kg	--
5	18 mg/kg	--
6	26 mg/kg	--
7	22 mg/kg	--
8	BDL	20 mg/kg
9	BDL	88 mg/kg
10	BDL	38 mg/kg

Table C-4 shows the calculations applied to data in Table C-3. For Test No. 9, $\frac{1}{2}$ the MDL is 44 mg/kg, which is greater than the highest individual test value in the data set (26 mg/kg in Test No. 6). Therefore, it would not be included in the subsequent outlier and emissions factor averaging analyses. The same holds true for Test No. 3 where $\frac{1}{2}$ the MDL equals 35 mg/kg, which is greater than 26 mg/kg. Test Nos. 8 and 10 would be included in the data set since $\frac{1}{2}$ the MDL values of 10 mg/kg and 19 mg/kg are less than the highest individual test value in the data set (26 mg/kg).

In this example, there are fewer than 11 values after those data whose replacement values are greater than or equivalent to the highest value above the detection limit are removed. As a result, WebFIRE assigns values to the remaining BDL runs equivalent to $\frac{1}{2}$ their MDL and then calculates the emissions factor for this data set (17.6 mg/kg) by averaging 19, 16, 11, 18, 26, 22, 10, and 19 mg/kg.

Table C-4. Calculations for Example Data Set B

Test No.	Test Value	Test MDL	$\frac{1}{2}$ MDL for BDL Data	$\frac{1}{2}$ MDL > Highest Test Value?	Value for Averaging Analysis
1	19 mg/kg	--	--	--	19 mg/kg
2	16 mg/kg	--	--	--	16 mg/kg
3	BDL	70 mg/kg	35 mg/kg	Yes	<i>Data Not Used</i>
4	11 mg/kg	--	--	--	11 mg/kg
5	18 mg/kg	--	--	--	18 mg/kg
6	26 mg/kg	--	--	--	26 mg/kg
7	22 mg/kg	--	--	--	22 mg/kg
8	BDL	20 mg/kg	10 mg/kg	No	10 mg/kg
9	BDL	88 mg/kg	44 mg/kg	Yes	<i>Data Not Used</i>
10	BDL	38 mg/kg	19 mg/kg	No	19 mg/kg
Average					17.6 mg/kg

Example 3

Table C-5 shows a candidate data set that consists of a mix of data that are above the MDL and data that are BDL. This scenario illustrates the recommended approach where the 13 values above the MDL are used to develop an emissions factor and the 6 BDL values are not used.

Table C-5. Example Data Set C

Test No.	Test Value	Test MDL
1	19 mg/kg	--
2	16 mg/kg	--
3	BDL	24 mg/kg
4	11 mg/kg	--
5	18 mg/kg	--
6	33 mg/kg	--
7	22 mg/kg	--
8	BDL	10 mg/kg
9	BDL	8 mg/kg
10	17 mg/kg	--
11	14 mg/kg	--
12	23 mg/kg	--
13	9 mg/kg	--
14	25 mg/kg	--
15	BDL	12 mg/kg

Table C-5. Example Data Set C (Cont.)

Test No.	Test Value	Test MDL
16	12 mg/kg	--
17	BDL	10 mg/kg
18	14 mg/kg	--
19	BDL	16 mg/kg

Table C-6 shows the calculations applied to the data in Table C-5. Because there are more than 11 values that are above the MDL in the candidate data set, none of the BDL data are used to calculate an emissions factor. WebFIRE excludes values associated with BDL and calculates the emissions factor for this data set (17.9 mg/kg) by averaging 19, 16, 11, 18, 33, 22, 17, 14, 23, 9, 25, 12, and 14 mg/kg.

Table C-6. Calculations for Example Data Set C

Test No.	Test Value	Test MDL	½ MDL for BDL Data	Value for Averaging Analysis
1	19 mg/kg	--	--	19 mg/kg
2	16 mg/kg	--	--	16 mg/kg
3	BDL	24 mg/kg	12 mg/kg	<i>Data Not Used</i>
4	11 mg/kg	--	--	11 mg/kg
5	18 mg/kg	--	--	18 mg/kg
6	33 mg/kg	--	--	33 mg/kg
7	22 mg/kg	--	--	22 mg/kg
8	BDL	10 mg/kg	5 mg/kg	<i>Data Not Used</i>
9	BDL	8 mg/kg	4 mg/kg	<i>Data Not Used</i>
10	17 mg/kg	--	--	17 mg/kg
11	14 mg/kg	--	--	14 mg/kg
12	23 mg/kg	--	--	23 mg/kg
13	9 mg/kg	--	--	9 mg/kg
14	25 mg/kg	--	--	25 mg/kg
15	BDL	12 mg/kg	6 mg/kg	<i>Data Not Used</i>
16	12 mg/kg	--	--	12 mg/kg
17	BDL	10 mg/kg	5 mg/kg	<i>Data Not Used</i>
18	14 mg/kg	--	--	14 mg/kg
19	BDL	16 mg/kg	8 mg/kg	<i>Data Not Used</i>
Average				17.9 mg/kg

APPENDIX D

PROCEDURES FOR DETERMINING STATISTICAL OUTLIERS

DRAFT

1.0 Introduction

After a candidate data set has been selected for emissions factor development by an end user and the BDL analysis has been performed (see Appendix C), EPA conducts a set of tests (i.e., the *Dixon Q Test* or the *Grubbs Test*) to identify data values in the candidate data set that are statistical outliers (i.e., a data point that does not conform to the statistical pattern established by other data under consideration). These statistical tests are incorporated into EPA's WebFIRE (see Section 6.1). Emissions data are usually log-normally distributed; therefore, for the purposes of evaluating outliers for emissions factor development, the assumption is made that all emission test data values in the candidate data set follow log normal distributions.

The Dixon Q Test is used to determine outliers for 3 to 6 average values that exhibit a normal or log normal distribution. Additional information and equations describing the Dixon Q Test are available at: http://www.chem.uoa.gr/applets/AppletQtest/Text_Qtest2.htm. The Grubbs Test is used to determine outliers for 7 or more average values that exhibit a normal or log normal distribution. Further background information for the Grubbs Test, including key equations, is available on the National Institute of Standards and Technology's (NIST's) website (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm>). An interactive version of the Grubbs Test for statistical outlier evaluation can be run from the following site: <http://www.graphpad.com/quickcalcs/Grubbs1.cfm>.

2.0 Description of Procedures

In WebFIRE, the outlier test is applied to the candidate data set in an iterative process. Each run of the outlier test may identify a single outlier value (if any in the data set) and the test is applied until all outliers have been identified and removed from the candidate data set. However, the data values removed from the candidate data set are not removed from the WebFIRE database because the outlier designation is relative to the population of values selected for the candidate data set (i.e., an outlier in one data set may be an acceptable value in a different data set).

All outlier tests in WebFIRE are performed using an alpha of 0.20, meaning that the developer is willing to accept a 20 percent risk of rejecting a valid observation. For purposes of the emissions factor program, include all values from individual test runs (even those named by others as outliers, provided the values remain identifiable) in calculating the test averages before conducting an outlier test.

The general approach to use for determining data outliers is shown in Figure D-1. If there are less than three average values to be evaluated, a statistical outlier test is not performed by WebFIRE because statistical analyses cannot determine outliers from such a small sample size. If there are 3 to 6 data values in the candidate data set, WebFIRE applies the Dixon Q Test to determine outliers. If there are 7 or more data values for analysis, the Grubbs Test is used to identify outliers.

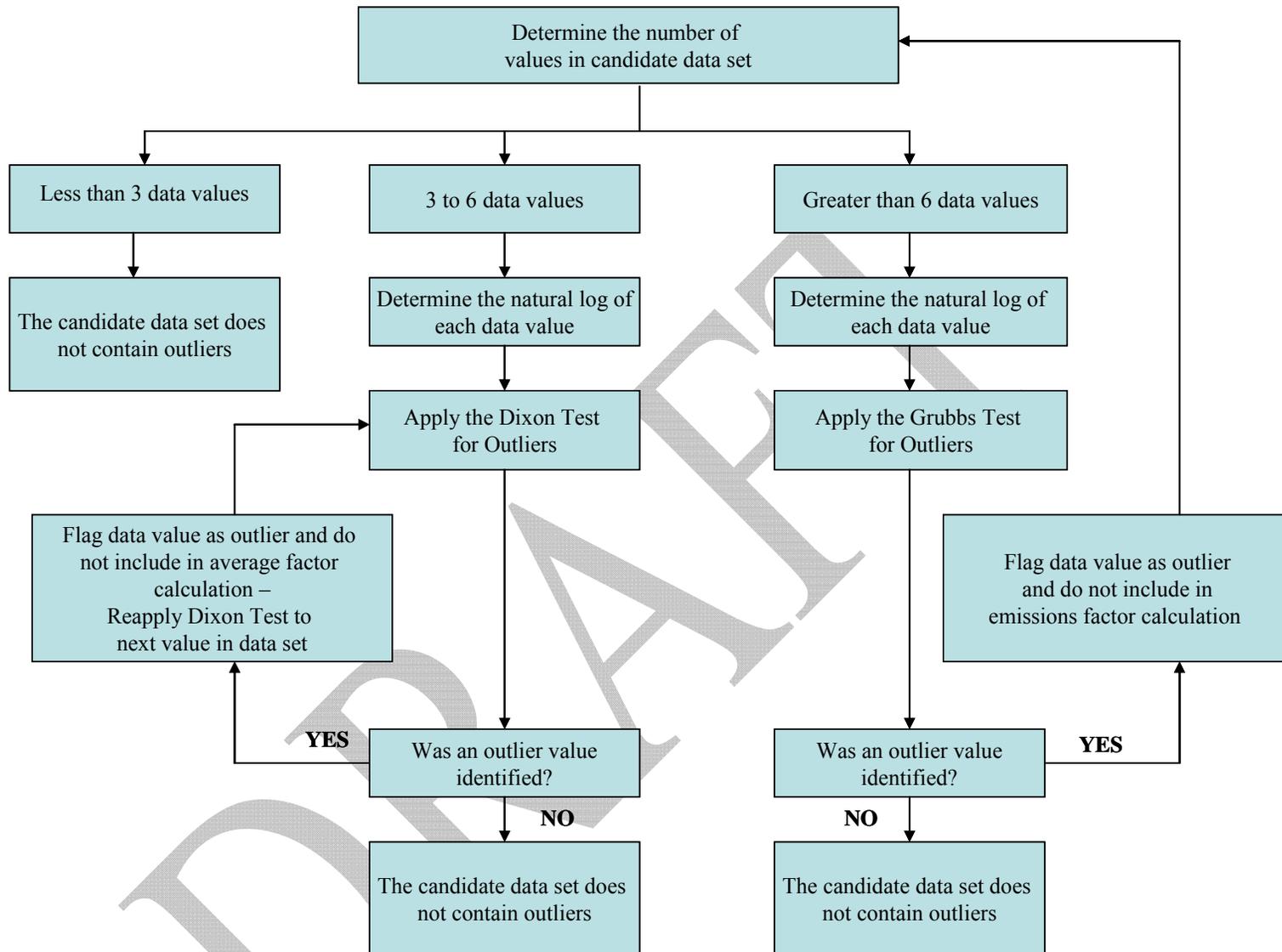


Figure D-1. Procedures to Identify Data Outliers for in a Candidate Data Set

The basic equation for the Dixon Q Test is: $Q_{\text{exp}} = \frac{X_2 - X_1}{X_N - X_1}$

where: Q_{exp} = The experimental value calculated from the data set,
 N = Number of values comprising the data set arranged in ascending order,
 X_1 = First value in data set,
 X_2 = Second value in data set, and
 X_N = The last value in the ascending order data set.

The value Q_{exp} is compared to a Q-critical (Q_{crit}) value found in reference tables for the Dixon Q Test (http://www.chem.uoa.gr/applets/AppletQtest/Text_Qtest2.htm). The Q_{crit} value used must correspond to the confidence level selected for the outlier analysis (80 percent in this case). The Q_{crit} table is shown below in Table D-1. For example, at an 80 percent confidence level and using five average data values, the Q_{crit} value is 0.557. Therefore, for a value to be considered an outlier, its Q_{exp} must be greater than the corresponding Q_{crit} (0.557). If it is, that average value is considered to be an outlier.

Table D-1. Table of Critical Values of Q_{crit}

N	Q_{crit} (CL:80%)	Q_{crit} (CL:90%)	Q_{crit} (CL:95%)	Q_{crit} (CL:99%)
3	0.886	0.941	0.970	0.994
4	0.679	0.765	0.829	0.926
5	0.557	0.642	0.710	0.821
6	0.482	0.560	0.625	0.740
7	0.434	0.507	0.568	0.680
8	0.399	0.468	0.526	0.634
9	0.370	0.437	0.493	0.598
10	0.349	0.412	0.466	0.568

If an outlier is detected by WebFIRE, it is removed from the candidate data set and the Dixon Q Test is performed again. Outliers are removed from the candidate data set until no outliers are detected. When the data set does not contain outliers, WebFIRE calculates a user-defined emissions factor. If there are three or fewer valid data points, no further outlier tests are performed.

The Grubbs Test statistic is defined as follows:

$$G = \frac{\max_{i=1, \dots, N} |Y_i - \bar{Y}|}{s}$$

Where: G = Grubbs test statistic,
 Y = sample mean,
 Y_i = test value, and
 s = data set standard deviation.

This form of the Grubbs Test is known as the two-sided version of the test. For this version, the hypothesis of no outliers is rejected if, at the specified significance level (0.01 in this case), the following is true.

$$G > \frac{(N-1)}{\sqrt{N}} \sqrt{\frac{t^2_{(\alpha/(2N), N-2)}}{N-2 + t^2_{(\alpha/(2N), N-2)}}$$

Where $t_{(\alpha/(2N), N-2)}$ denotes the upper critical value of the t-distribution with N-2 degrees of freedom and a significance level of $\alpha/(2N)$.

If an outlier is detected by WebFIRE, it is removed from the other average values and the number of valid data points remaining is determined. The Grubbs Test, or the appropriate outlier test as determined by the number of valid data points, is performed again. Outliers are removed from the other values using the Grubbs Test (or Dixon Test if the number of valid data points is less than 6) until no outliers are detected. When the data set does not contain outliers, WebFIRE calculates an emissions factor using the remaining data set.

APPENDIX E

EMISSIONS FACTOR DEVELOPMENT AND DATA QUALITY CHARACTERIZATION PROCEDURES

DRAFT

1.0 Introduction

The procedures used within WebFIRE to determine which individual data points (i.e., values from stack tests) to use in deriving an emissions factor are based upon two premises: (1) higher-quality data are preferred over lower-quality data, and (2) more data points are preferred over less data points. These concepts are combined with simple statistical procedures to derive the approach used by WebFIRE to assign a quality rating to the derived emissions factor. This quality rating indicates how well the derived factor represents the average of the emissions from a particular source category. These procedures are described in detail in the following sections.

2.0 Terms and Definitions

As a prelude to presenting these procedures, it is important to explain and define the parameters used for the emissions factor calculations and data quality characterizations:

1. Individual Test Rating (ITR) – The ITR value is the quality indicator assigned to individual source test reports. This value is based upon the level of documentation available in the test report, the use and conformance with established EPA reference test method (or other test methods with comparable precision and accuracy), and the operation of the source and associated emissions controls at known and representative conditions. The ITR ranges from a high of 100 to a low of 0. The procedures for calculating the ITR are presented in Appendix B.
2. Composite Test Rating (CTR) – The CTR is a weighted-average quality indicator for groups of test reports. An inverse square weighting of the ITR values for the test reports is used in calculating the CTR. As with the ITR, the CTR ranges from a high of 100 to a low of 0.
3. Factor Quality Index (FQI) – The FQI is a numerical indicator representing the derived emissions factors ability to estimate the entire national population. The FQI is dependent upon both the CTR and the number of sources used to develop the emissions factor. The FQI is analogous to the standard error of the mean (σ_M) in statistical calculations. In statistical calculations, σ_M provides an indication of the confidence associated with an estimate of the mean of a population when a given number of samples are obtained from the population. The σ_M is calculated from the standard deviation of the samples (or other estimate of the populations variability) divided by the square root of the number of samples. In the FQI, the parameter $100/CTR$ simulates the function of the standard deviation in that measurements with great variability (due to variations between sources in the population, variations with individual sources, precision and accuracy of the methods used for measurement, and other factors affecting variations in the measured values) are larger in value than measurements with less variability. In the FQI, the minimum value is associated with emissions tests that are judged to have the greatest precision and accuracy of sources operating at representative conditions. This is the appropriate data set selection for use in emissions factor derivation as increases in the σ_M and increases in the number of samples used to estimate the mean of the population serve to reduce the value of the FQI in proportion to the estimated reliability of the estimate of

the mean. In addition, like σ_M , equal values of FQI provide comparable reliability in the estimate of the population mean irrespective of differences in the CTR and the number of samples used for estimating the population mean.

4. Emissions factor quality indicator – There are three quality indicators used to characterize the calculated emissions factor:
 - *Highly Representative* is assigned to emissions factors having the lowest FQI rating.
 - *Moderately Representative* is assigned to emissions factors having intermediate FQIs.
 - *Poorly Representative* is assigned to emissions factors based upon tests that have the highest FQI rating.

5. Boundary criteria – Boundary criteria refers to the specific conditions that determine which quality rating (i.e., Poorly Representative, Moderately Representative, or Highly Representative) is assigned to an emissions factor. Based upon our experience with developing emissions factors, we determined that, for source categories containing more than 15 sources, three tests with a CTR of 100 (FQI = 0.5774) qualifies for a Moderately Representative rating. Likewise, more than 11 tests with a CTR of 100 (FQI = 0.3015) qualifies for a Highly Representative rating. These criteria are designed to allow for the development of highly-representative emissions factors without the burden of conducting an inordinate amount of stack tests. For source categories containing 15 or fewer sources, it is appropriate to allow fewer tests to attain a specific quality rating. Specifically, more than one test with a CTR of 100 (FQI = 1.000) qualifies for a Moderately Representative rating and more than three tests with a CTR of 100 (FQI = 0.5774) qualifies for a Highly Representative rating. For both population sizes, degradation of the CTR requires an increase in the number of tests to compensate for the decrease in the average test quality to achieve the same FQI. Table E-1 provides the boundary line equations for the two population sizes and Figures E-1 and E-2 provide the graphical relationship between the CTRs and the number of tests required for the boundary conditions.

Table E-1. FQI and boundary line equations.

If the source category contains . . .	Then use these boundary line equations . . .	
	Poorly to moderately representative of SCC	Moderately to highly representative of SCC
More than 15 sources	FQI = 0.5774 $N = 30,000 * CTR^{-2}$	FQI = 0.3015 $N = 110,000 * CTR^{-2}$
15 or fewer sources	FQI = 1 $N = 10,000 * CTR^{-2}$	FQI = 0.5774 $N = 30,000 * CTR^{-2}$

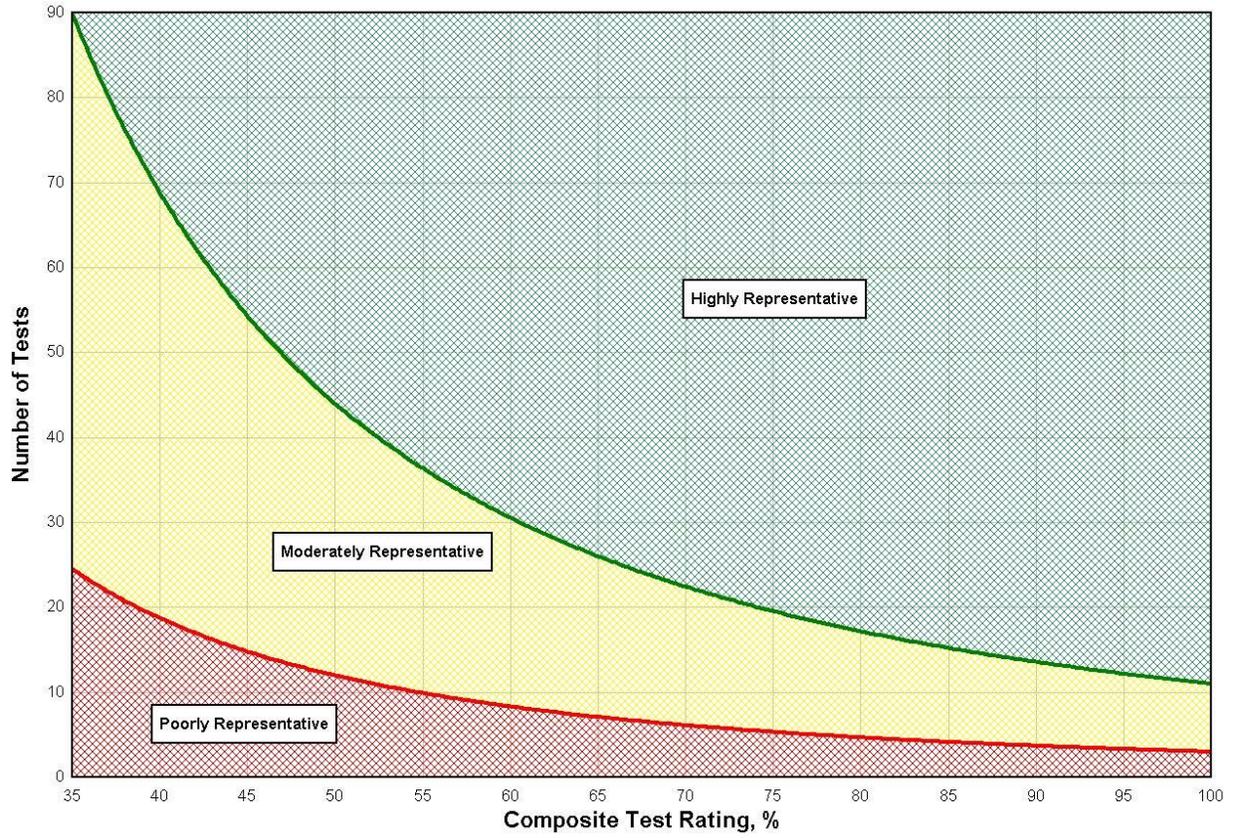


Figure E-1. Emissions Factor Representativeness Areas for Source Categories Containing More Than 15 Sources

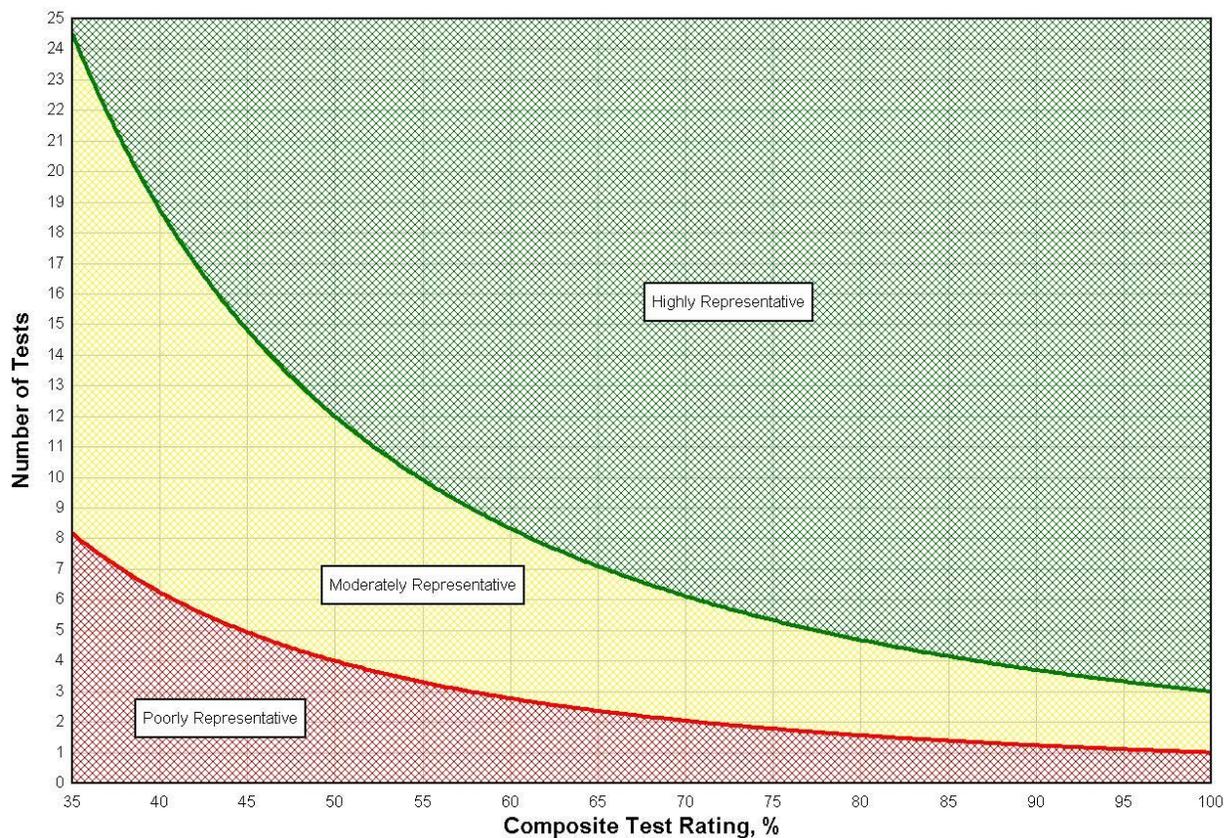


Figure E-2. Emissions Factor Representativeness Areas for Source Categories Containing 15 or Fewer Sources

3.0 Procedures

The following steps summarize the specific calculation and data quality characterization procedures used in WebFIRE to calculate a new or revise an existing recommended emissions factor:

- **Step 1** – WebFIRE groups the emissions test data that are being considered for emissions factor development by pollutant, SCC, and type of control device. All subsequent steps are performed individually on each pollutant/SCC/control device grouping.
- **Step 2** – WebFIRE subjects the candidate data set to the BDL and outlier analyses (See Appendices C and D, respectively).
- **Step 3** – WebFIRE arranges the individual data values being considered in descending order by: (1) the ITR and (2) the data value.

- Step 4 – Beginning with the second individual data value and continuing sequentially in order, WebFIRE calculates the CTR using the following equation:

$$CTR_n = \left[\frac{\sum_{i=1}^n \left(\frac{1}{ITR_i} \right)^2}{N} \right]^{-0.5}$$

Where:

- CTR = Composite Test Rating,
- ITR = Individual test rating (assigned by ERT), and
- N = Number of sources with ITRs equal or greater value as the source evaluated for inclusion.

It should be noted that a CTR is calculated for each combination of data sets potentially used to derive an overall emissions factor. For example, using a data set consisting of 10 source tests, WebFIRE would calculate 9 CTRs, beginning with the first two data points, then the first three data points, and so forth until a CTR is calculated for all 10 data points.

- Step 5 – For each calculated CTR, WebFIRE calculates the FQI using the equation:

$$FQI = \frac{100}{CTR \times N^{0.5}}$$

Where:

- CTR = Composite Test Rating associated with the data set selected for deriving the emissions factor, and
- N = Number of sources with ITRs equal or greater value as the source evaluated for inclusion.

- Step 6 – WebFIRE compares the calculated FQI with the FQI for the previous ITR grouping. If the FQI associated with the larger grouping (i.e., more data points) is less than the FQI with fewer data points, then WebFIRE proceeds back to Step 2 to perform the next sequence in calculations. If the FQI associated with the larger grouping is greater than the preceding FQI, then WebFIRE does not include the data responsible for the increase in the FQI in calculating the emissions factor and excludes the remaining data (with lower ITRs) from consideration.
- Step 7 – WebFIRE calculates the emissions factor using all source test data that were included in calculating the lowest FQI. This includes all test data with higher ITRs than the ITR value that resulted in an increased FQI value.

- **Step 8** – WebFIRE determines if the SCC corresponding to the individual data values selected by the end user contains 15 or fewer sources. Table E-2 lists the SCCs that we expect to contain 15 or fewer sources. Appendix G contains the descriptions for the SCCs shown in Table E-2.
- **Step 9** – WebFIRE compares the FQI for the sources used to calculate the emissions factor with the corresponding boundary criteria for assigning one of the three emissions factor quality ratings. Different boundary criteria are used for source categories containing 15 or fewer sources and for source categories containing greater than 15 sources.

Table E-2. SCCs Expected to Contain 15 or Fewer Sources

SCCs That Contain 15 or Fewer Sources ^a			
101011	301157	304009	316160
101019	301158	304010	360001
102003	301167	304040	390003
102011	301169	304049	401004
102016	301176	304051	402028
102017	301181	305004	501002
201003	301190	305013	625400
201013	301195	305022	631110
201900	301210	305024	631250
203009	301211	305026	631310
204002	301252	305029	631340
2810040	301253	305032	641300
301017	301254	305033	641301
301019	301301	305034	641302
301025	301302	305035	641310
301028	301303	305036	641320
301029	301304	305038	644200
301036	301305	305042	644500
301038	301401	305044	645200
301039	301402	305045	645210
301041	301403	305046	646100
301051	302003	305089	646150
301091	302012	305090	646200
301100	302022	305092	646300
301111	302028	314010	646320
301112	302039	315010	646330
301113	302042	315027	648200

Table E-2. SCCs Expected to Contain 15 or Fewer Sources (Cont.)

SCCs That Contain 15 or Fewer Sources ^a			
301114	303004	315031	648210
301121	303005	315040	648220
301124	303006	316030	649200
301126	303007	316040	651100
301133	303011	316050	651300
301137	303012	316060	651350
301140	303030	316120	651400
301152	303031	316130	685100
301153	303040	316140	
301156	304002	316150	

^a These 6-digit (point) or 7-digit (nonpoint) SCCs represent the source categories expected to have fewer than 15 sources. All SCCs starting with these code sequences are included.

Example 1

Table E-3 below contains an example set of 35 individual test data values selected to develop an emissions factor for SCC 303010 and their ITR, CTR, N, and FQI values, as well as an indication of whether or not the test data value should be used to calculate an emissions factor and the representativeness of the CTR.

Table E-3. Individual Test Data and Various Characteristics

Individual Test EF Value	ITR	CTR	N	FQI	Use for EF Average?	EF Representativeness
0.0108	98	98.00	1	1.0204	Yes	Poorly
0.1100	98	98.00	2	0.7215	Yes	Poorly
0.0917	92	95.87	3	0.6022	Yes	Poorly
0.0212	92	94.86	4	0.5271	Yes	Moderately
0.0339	91	94.05	5	0.4755	Yes	Moderately
0.0027	91	93.52	6	0.4365	Yes	Moderately
0.0563	89	92.83	7	0.4072	Yes	Moderately
0.0165	89	92.32	8	0.3829	Yes	Moderately
0.0158	88	91.81	9	0.3631	Yes	Moderately
0.0044	88	91.41	10	0.3460	Yes	Moderately
0.0675	88	91.08	11	0.3310	Yes	Moderately
0.0043	88	90.81	12	0.3179	Yes	Moderately
0.0449	74	89.10	13	0.3113	Yes	Moderately
0.0203	73	87.58	14	0.3052	Yes	Moderately

Table E-3. Individual Test Data and Various Characteristics (Cont.)

Individual Test EF Value	ITR	CTR	N	FQI	Use for EF Average?	EF Representativeness
0.0603	70	85.97	15	0.3003	Yes	Highly
0.0425	70	84.64	16	0.2954	Yes	Highly
0.0130	70	83.51	17	0.2904	Yes	Highly
0.1440	69	82.45	18	0.2859	Yes	Highly
0.0177	68	81.45	19	0.2817	Yes	Highly
0.0317	68	80.58	20	0.2775	Yes	Highly
0.0052	68	79.82	21	0.2734	Yes	Highly
0.1350	68	79.14	22	0.2694	Yes	Highly
0.0006	60	77.90	23	0.2677	Yes	Highly
0.0023	45	74.85	24	0.2727	No	Not applicable
0.0724	45	72.33	25	0.2765	No	Not applicable
0.0960	44	70.08	26	0.2799	No	Not applicable
0.0538	40	67.54	27	0.2850	No	Not applicable
0.0170	38	65.07	28	0.2904	No	Not applicable
0.0132	35	62.48	29	0.2972	No	Not applicable
0.0124	34	60.14	30	0.3036	No	Not applicable
0.0029	30	57.41	31	0.3128	No	Not applicable
0.0018	30	55.16	32	0.3205	No	Not applicable
0.0083	30	53.28	33	0.3268	No	Not applicable
0.0009	30	51.66	34	0.3319	No	Not applicable
0.0034	30	50.27	35	0.3362	No	Not applicable

Figure E-3 shows a plot of the CTR and N data in Table E-3 and the boundaries created by the line equations. In developing the emissions factor for the example data set, the first 23 values in Table E-3 are included in the emissions factor calculation because the FQI increases for the first time between the 23rd and 24th pair. Using the first 23 values yields an emissions factor of 0.0413 with a quality rating of Highly Representative.

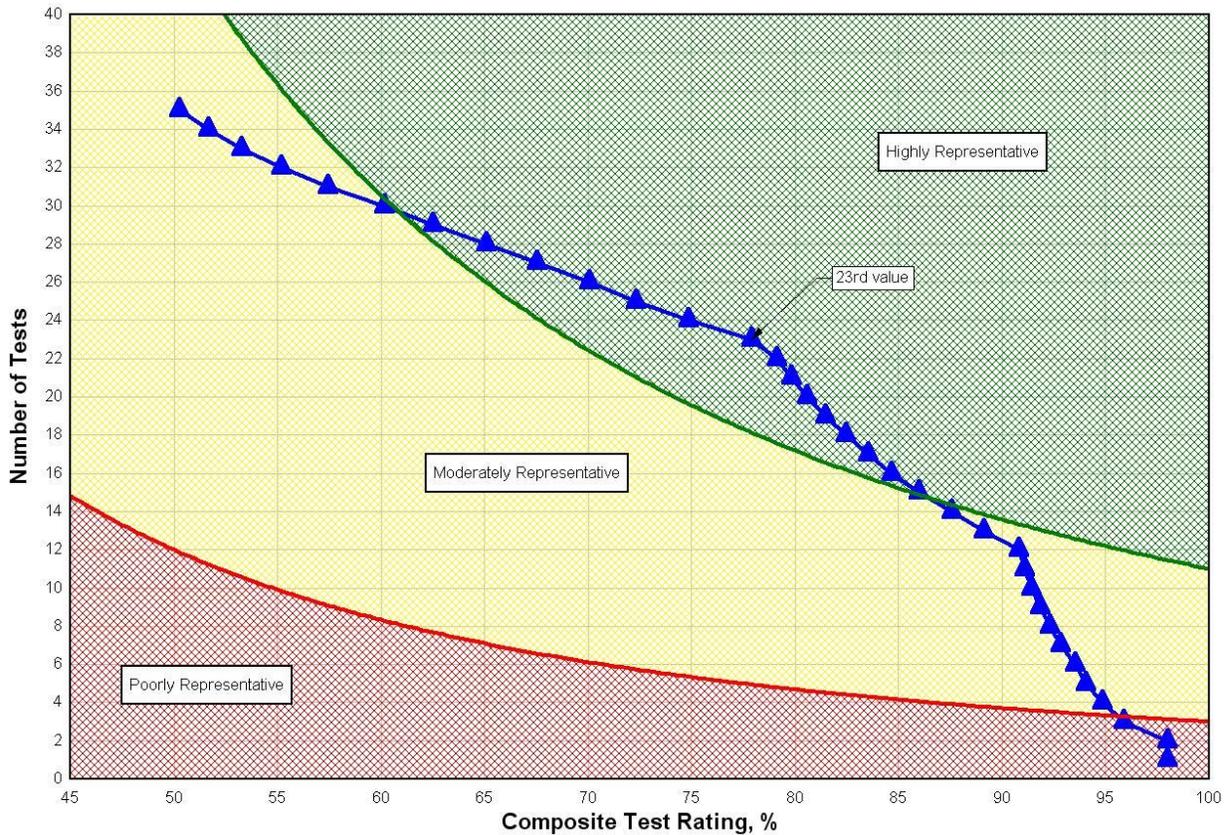


Figure E-3. Plot of CTR and N data from Table E-3

Example 2

Table E-4 contains another example set of individual test data values selected for use in developing a composite emissions factor for SCC 303011, which is expected to contain 15 or fewer sources per Table E-1.

Table E-4. Individual Test Data Values Selected for Developing an Emissions Factor for a Source Category Containing 15 or Fewer Sources

Individual Test Data Value	ITR
0.0015	45
0.0004	60
0.0055	30
0.0019	30
0.0012	30
0.0640	30
0.0113	30

Table E-4. Individual Test Data Values Selected for Developing an Emissions Factor for a Source Category Containing 15 or Fewer Sources (Cont.)

Individual Test Data Value	ITR
0.0088	30
0.0029	88
0.0611	92
0.0402	70
0.0299	74
0.0375	89
0.0118	68
0.0072	99

Table E-5 shows the same data after the table has been sorted; the N, CTR, and FQI values been calculated; and an indication of whether or not the test data value should be used to calculate an emissions factor and the representativeness of the CTR have been determined as specified in Steps 1 through 9.

Table E-5. Individual Test Data and Various Characteristics for a Source Category with 15 or Fewer Sources

Test EF Value	ITR	CTR	N	FQI	Use for EF Average?	EF Representativeness
0.0072	99	99.00	1	1.0101	Yes	Poorly
0.0611	92	95.31	2	0.7419	Yes	Moderately
0.0375	89	93.06	3	0.6204	Yes	Moderately
0.0029	88	91.71	4	0.5452	Yes	Highly
0.0299	74	87.16	5	0.5131	Yes	Highly
0.0402	70	83.42	6	0.4894	Yes	Highly
0.0118	68	80.56	7	0.4692	Yes	Highly
0.0004	60	76.80	8	0.4603	Yes	Highly
0.0015	45	69.75	9	0.4779	No	Not applicable
0.0012	30	58.11	10	0.5442	No	Not applicable
0.0019	30	51.97	11	0.5801	No	Not applicable
0.0088	30	48.12	12	0.6000	No	Not applicable
0.0113	30	45.45	13	0.6103	No	Not applicable
0.0640	30	43.48	14	0.6147	No	Not applicable
0.0055	30	41.97	15	0.6152	No	Not applicable

Figure E-4 shows a plot of the CTR and N data in Table E-5 and the boundaries created by the equations. In developing the emissions factor for the example data set, the first 8 values in Table E-5 are included in the emissions factor calculation because the FQI increases for the first time between the 8th and 9th pair. Using the first 8 values yields an emissions factor of 0.0239 with a quality rating of Highly Representative.

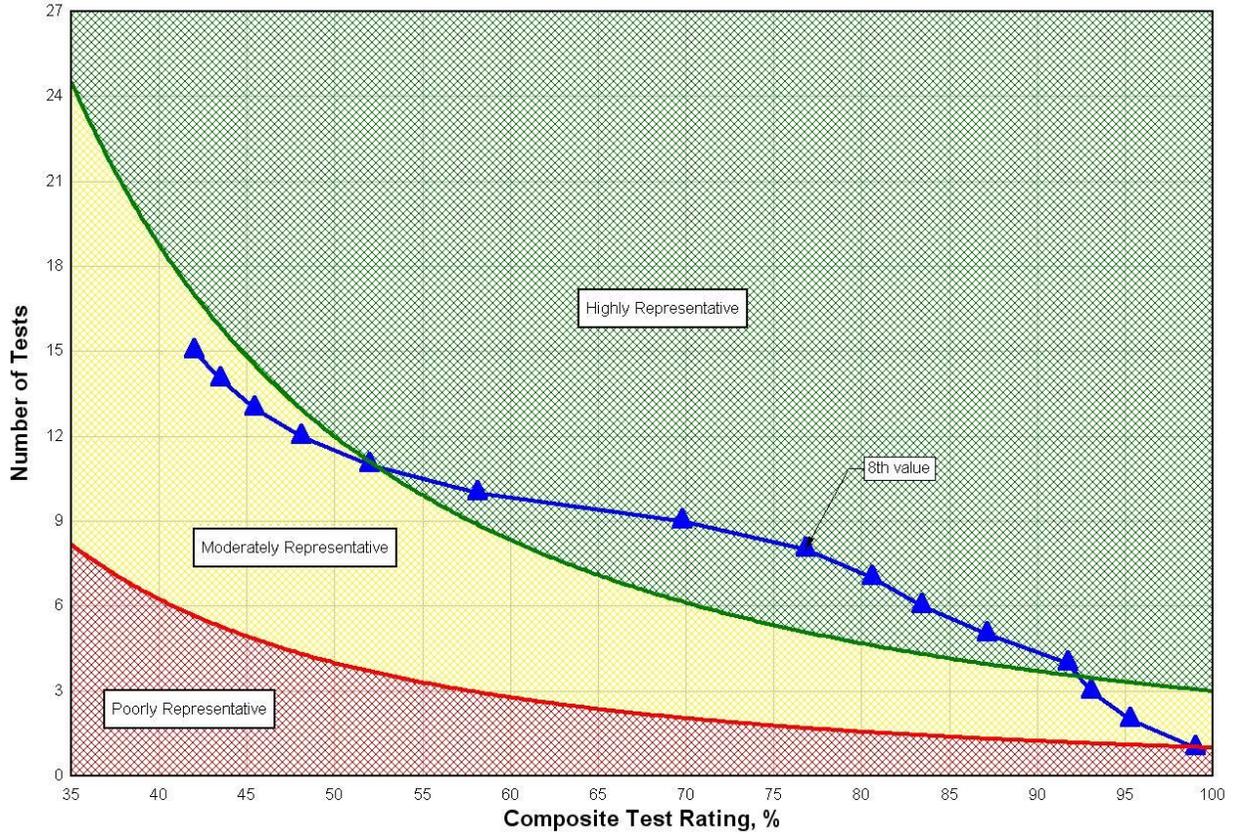


Figure E-4. Plot of Selected data from Table D-6.

For test data submitted to WebFIRE using ERT, a numerical ITR value will be assigned to the data by ERT prior to incorporation in WebFIRE. For data that were incorporated into WebFIRE prior to the development of ERT (e.g., the underlying data used to develop AP 42 emissions factors), the current subjective, letter-grade quality ratings have been converted to numerical values as follows:

Test Data Letter Grade	Equivalent ITR Score
A	80
B	60
C	45
D	30
U	0

For example, a previous test rated as a “B” that is part of the candidate data set for emissions factor development would have an ITR of 60 for use in calculating the CTR. We used this approach because it would be time intensive and prohibitively costly to reevaluate every previous test report to assign it an ITR based on the rating system contained in the ERT.

DRAFT

APPENDIX F

STATISTICAL PROCEDURES FOR DETERMINING VALID DATA COMBINATIONS

DRAFT

1.0 Introduction

As new emissions data are incorporated into WebFIRE, we expect that, periodically, the Agency will need to determine whether a new data set should be combined with an existing data set. In these cases, we will follow the procedures specified in this appendix. We anticipate that these procedures will be applied on a limited, case-by-case basis, most likely on data that are expected to be from the same type of emissions units, with the same type of emissions controls, and under the same type of operational process. These procedures would not be applied to source test data from processes or controls that are clearly separate and distinct (e.g., coke oven emissions and electric arc furnace emissions).

Simple statistical characteristics such as the number of values, the average of the values, and the variance of the values can be used to represent a data set for computational purposes. Comparison of similar characteristics between data sets can determine whether the data sets are from the same population of values. If the data sets are determined to be from the same population of values, the data sets can be combined into a single, combined data set, often referred to as a pool. Pooled values are preferred over individual values, because pooled values provide the best estimate of a population's variance. In some instances, a single value (computed as the average of an emissions test) is compared to other values. Under those circumstances, the single value must be adjusted to multiple values so that there are enough data to calculate a variance.

2.0 Description of Procedures

The data combination assessment procedures that the EPA will use to determine whether a new data set should be combined with an existing data set are based upon use of the statistical Student's t-test. The following steps are used to determine if proposed data can be combined acceptably.

1. Obtain all emissions test data (i.e., the number of values and the numerical values of the data set) that were used to calculate the existing emissions factor and that will be reviewed for inclusion in a new (or revised) emissions factor. If only a single average value from an emissions test exists and the underlying test run values are available, use the individual test run values from the emissions test to determine statistical characteristics. If only a single average value from an emissions test exists but the underlying test run values are not available, the average value is used three times (i.e., the single average value is used to represent the individual test run values).
2. Prepare null hypothesis (i.e., the means of the two sets are equal) and alternative hypotheses (i.e., the means of the two sets are unequal).
3. Conduct a Student's t-test on the data sets assuming unequal variances. Find t_{critical} values at the 0.2 significance level for the appropriate number of degrees of freedom. If the Student's t-test statistic is greater than the t_{critical} value, assume the means are unequal. Do not combine the data sets.

4. If the Student's t-test yields a value that is less than or equal to the t_{critical} value, assume the means are equal.
5. Combine the data sets after applying the appropriate processes for evaluating BDL data (if needed) and removing data outliers.

Examples illustrating the use of the data combination assessment procedures are shown below.

Example 1

Table F-1 presents two data sets: Group A, which is used to calculate the current emissions factor of 0.0118 pounds of pollutant per ton of fuel combusted, and Group B, which is from a similar source category with similar controls, and operated under a similar process. Group A and B do not contain any BDL values or values that are considered to be outliers.

Table F-1. Emissions Factor Characteristics for Group A and B

Group A Source Test Data	Group B Source Test Data
0.0015	0.0029
0.0004	0.0611
0.0055	0.0402
0.0019	0.0299
0.0012	0.0375
0.064	0.0118
0.0113	0.0072
0.0088	

Table F-2 shows the results of the Student's t-test for the two groups of data. These values yield a t-test statistic equivalent to -1.40 using an alpha of 0.2 and a critical value of t (0.870) for a one-tailed, heteroscedastic distribution with seven degrees of freedom for Group A and six degrees of freedom for Group B.

Table F-2. T Statistics for Group A and B

t-Test: Two-Sample Assuming Unequal Variances		
	Group A	Group B
Mean	0.011825	0.027229
Variance	0.00046	0.000443

Table F-2. T Statistics for Group A and B (Cont.)

t-Test: Two-Sample Assuming Unequal Variances		
Observations	8	7
Degrees of Freedom	7	6
t-test statistic ($\alpha = 0.2$)	-1.40138	
P(T<=t) one-tail	0.092261	
t Critical one-tail	0.870152	

Because the value of the t-test statistic (-1.40) is less than the critical value of t (0.870), the data are sufficient to show that the means of Group A and Group B are equal. Therefore, the null hypothesis (i.e., the means are equal) is accepted. Given that the means of Groups A and B are equal, the individual test data sets can be combined. If the means had been unequal, the Group A and B individual test data sets would not be combined.

Example 2

Table F-3 presents two data sets: Group C, which is based upon one emissions test whose individual run data are available (0.0005, 0.0015, and 0.0025), and Group D, which is based upon one emissions test whose individual run data are unavailable (the average value of 0.0029 is assumed for each of the three test runs). The data for the two groups are from a similar source category with similar controls, and operated under a similar process. None of the values in Group C or D are BDL and none of the values in Group C or D is an outlier.

Table F-3. Emissions Factor Characteristics for Group C and D

Group C Source Test Data	Group D Source Test Data
0.0005	0.0029
0.0015	0.0029
0.0025	0.0029

Table F-4 shows the results of the Student's t-test for the two groups of data. These values yield a t-test statistic equivalent to -2.42 using an alpha of 0.2 and a critical value of t (1.06) for a one-tailed, heteroscedastic distribution with two degrees of freedom for Group C and two degrees of freedom for Group D.

Table F-4. T Statistics for Group C and D

t-Test: Two-Sample Assuming Unequal Variances		
	Group C	Group D
Mean	0.0015	0.0029
Variance	0.000001	0
Observations	3	3
Degrees of Freedom	2	2
t-test statistic ($\alpha = 0.2$)	-2.424871	
P(T<=t) one-tail	0.068088	
t Critical one-tail	1.06066	

Because the value of the t-test statistic (-2.42) is less than the critical value of t (1.06), the data are sufficient to show that the means of Group C and Group D are equal. We have assumed for this analysis that additional data are not available; therefore, we accept the null hypothesis (i.e., the means are equal). Given that the means of Groups C and D are equal, the emissions factor data sets are combined. If the means had been unequal, the emissions factor data sets would not be combined.

Example 3

Table F-5 presents two data sets: Group E, which is used to calculate the current emissions factor of 0.0154 pounds of pollutant per ton of fuel combusted, and Group F, which is from a similar source category with similar controls, and operated under a similar process. Groups E and F do not contain any BDL values or values that are considered to be outliers.

Table F-5. Emissions Factor Characteristics for Group A and B

Group E Source Test Data	Group F Source Test Data
0.016	0.0145
0.017	0.015
0.015	0.013
0.015	0.012
0.016	0.0135
0.014	0.012
0.0145	0.0135
0.0155	

Table F-6 shows the results of the Student's t-test for the two groups of data. These values yield a t-test statistic equivalent to -1.40 using an alpha of 0.2 and a critical value of t

(0.870) for a one-tailed, heteroscedastic distribution with seven degrees of freedom for Group A and six degrees of freedom for Group B.

Table F-2. T Statistics for Group E and F

t-Test: Two-Sample Assuming Unequal Variances		
	Group E	Group F
Mean	0.015375	0.013357
Variance	9.11E-07	1.31E-06
Observations	8	7
Degrees of Freedom	7	6
t-test statistic ($\alpha = 0.2$)	3.678486	
P(T<=t) one-tail	0.001626	
t Critical one-tail	0.872609	

Because the value of the t-test statistic (3.68) is greater than the critical value of t (0.87), the data are not sufficient to show that the means of Group E and Group F are equal. We have assumed for this analysis that additional data are not available; therefore, we reject the null hypothesis (i.e., the means are unequal). Given that the means of Groups E and F are unequal, the emissions factor data sets are not combined.

APPENDIX G

SOURCE CLASSIFICATION CODES FOR SOURCE CATEGORIES CONTAINING 15 OR FEWER UNITS

DRAFT

Table G-1. Source Classification Codes for Source Categories Containing 15 or Fewer Units

Data Category	SCC L3	SCC L1 Description	SCC L2 Description	SCC L3 Description
POINT	101011	External Combustion Boilers	Electric Generation	Bagasse
POINT	101019	External Combustion Boilers	Electric Generation	Coal-based Synfuel
POINT	102003	External Combustion Boilers	Industrial	Lignite
POINT	102011	External Combustion Boilers	Industrial	Bagasse
POINT	102016	External Combustion Boilers	Industrial	Methanol
POINT	102017	External Combustion Boilers	Industrial	Gasoline
POINT	201003	Internal Combustion Engines	Electric Generation	Gasified Coal
POINT	201013	Internal Combustion Engines	Electric Generation	Liquid Waste
POINT	201900	Internal Combustion Engines	Electric Generation	Flares
POINT	203009	Internal Combustion Engines	Commercial/Institutional	Kerosene/Naphtha (Jet Fuel)
POINT	204002	Internal Combustion Engines	Engine Testing	Rocket Engine Testing
NONPOINT	2810040	Miscellaneous Area Sources	Other Combustion	Aircraft/Rocket Engine Firing and Testing
POINT	301017	Industrial Processes	Chemical Manufacturing	Phosphoric Acid: Thermal Process
POINT	301019	Industrial Processes	Chemical Manufacturing	Phthalic Anhydride
POINT	301025	Industrial Processes	Chemical Manufacturing	Cellulosic Fiber Production
POINT	301028	Industrial Processes	Chemical Manufacturing	Normal Superphosphates
POINT	301029	Industrial Processes	Chemical Manufacturing	Triple Superphosphate
POINT	301036	Industrial Processes	Chemical Manufacturing	Chromic Acid Manufacturing
POINT	301038	Industrial Processes	Chemical Manufacturing	Sodium Bicarbonate
POINT	301039	Industrial Processes	Chemical Manufacturing	Hydrogen Cyanide
POINT	301041	Industrial Processes	Chemical Manufacturing	Nitrocellulose
POINT	301051	Industrial Processes	Chemical Manufacturing	Animal Adhesives
POINT	301091	Industrial Processes	Chemical Manufacturing	Acetone/Ketone Production
POINT	301100	Industrial Processes	Chemical Manufacturing	Maleic Anhydride
POINT	301111	Industrial Processes	Chemical Manufacturing	Asbestos Chemical
POINT	301112	Industrial Processes	Chemical Manufacturing	Elemental Phosphorous
POINT	301113	Industrial Processes	Chemical Manufacturing	Boric Acid
POINT	301114	Industrial Processes	Chemical Manufacturing	Potassium Chloride
POINT	301121	Industrial Processes	Chemical Manufacturing	Organic Dyes/Pigments
POINT	301124	Industrial Processes	Chemical Manufacturing	Chloroprene
POINT	301126	Industrial Processes	Chemical Manufacturing	Brominated Organics

Table G-1. Source Classification Codes for Source Categories Containing 15 or Fewer Units (Cont.)

Data Category	SCC L3	SCC L1 Description	SCC L2 Description	SCC L3 Description
POINT	301133	Industrial Processes	Chemical Manufacturing	Acetic Anhydride
POINT	301137	Industrial Processes	Chemical Manufacturing	Esters Production
POINT	301140	Industrial Processes	Chemical Manufacturing	Acetylene Production
POINT	301152	Industrial Processes	Chemical Manufacturing	Bisphenol A
POINT	301153	Industrial Processes	Chemical Manufacturing	Butadiene
POINT	301156	Industrial Processes	Chemical Manufacturing	Cumene
POINT	301157	Industrial Processes	Chemical Manufacturing	Cyclohexane
POINT	301158	Industrial Processes	Chemical Manufacturing	Cyclohexanone/Cyclohexanol
POINT	301167	Industrial Processes	Chemical Manufacturing	Vinyl Acetate
POINT	301169	Industrial Processes	Chemical Manufacturing	Ethyl Benzene
POINT	301176	Industrial Processes	Chemical Manufacturing	Glycerin (Glycerol)
POINT	301181	Industrial Processes	Chemical Manufacturing	Toluene Diisocyanate
POINT	301190	Industrial Processes	Chemical Manufacturing	Methyl Methacrylate
POINT	301195	Industrial Processes	Chemical Manufacturing	Nitrobenzene
POINT	301210	Industrial Processes	Chemical Manufacturing	Caprolactum (Use 3-01-130 for Ammonium Sulfate By-product Production)
POINT	301211	Industrial Processes	Chemical Manufacturing	Linear Alkylbenzene
POINT	301252	Industrial Processes	Chemical Manufacturing	Etherene Production
POINT	301253	Industrial Processes	Chemical Manufacturing	Glycol Ethers
POINT	301254	Industrial Processes	Chemical Manufacturing	Nitriles, Acrylonitrile, Adiponitrile Production
POINT	301301	Industrial Processes	Chemical Manufacturing	Chlorobenzene
POINT	301302	Industrial Processes	Chemical Manufacturing	Carbon Tetrachloride
POINT	301303	Industrial Processes	Chemical Manufacturing	Allyl Chloride
POINT	301304	Industrial Processes	Chemical Manufacturing	Allyl Alcohol
POINT	301305	Industrial Processes	Chemical Manufacturing	Epichlorohydrin
POINT	301401	Industrial Processes	Chemical Manufacturing	Nitroglycerin Production
POINT	301402	Industrial Processes	Chemical Manufacturing	Explosives Manufacture – Pentaerythritol Tetranitrate (PETN)
POINT	301403	Industrial Processes	Chemical Manufacturing	Explosives Manufacture – RDX/HMX Production
POINT	302003	Industrial Processes	Food and Agriculture	Instant Coffee Products
POINT	302012	Industrial Processes	Food and Agriculture	Fish Processing
POINT	302022	Industrial Processes	Food and Agriculture	Cotton Seed Delinting

Table G-1. Source Classification Codes for Source Categories Containing 15 or Fewer Units (Cont.)

Data Category	SCC L3	SCC L1 Description	SCC L2 Description	SCC L3 Description
POINT	302028	Industrial Processes	Food and Agriculture	Mushroom Growing
POINT	302039	Industrial Processes	Food and Agriculture	Carob Kibble
POINT	302042	Industrial Processes	Food and Agriculture	Vinegar Manufacturing
POINT	303004	Industrial Processes	Primary Metal Production	Coke Manufacture: Beehive Process
POINT	303005	Industrial Processes	Primary Metal Production	Primary Copper Smelting
POINT	303006	Industrial Processes	Primary Metal Production	Ferroalloy, Open Furnace
POINT	303007	Industrial Processes	Primary Metal Production	Ferroalloy, Semi-covered Furnace
POINT	303011	Industrial Processes	Primary Metal Production	Molybdenum
POINT	303012	Industrial Processes	Primary Metal Production	Titanium
POINT	303030	Industrial Processes	Primary Metal Production	Zinc Production
POINT	303031	Industrial Processes	Primary Metal Production	Leadbearing Ore Crushing and Grinding
POINT	303040	Industrial Processes	Primary Metal Production	Alumina Processing - Bayer Process
POINT	304002	Industrial Processes	Secondary Metal Production	Copper
POINT	304009	Industrial Processes	Secondary Metal Production	Malleable Iron
POINT	304010	Industrial Processes	Secondary Metal Production	Nickel
POINT	304040	Industrial Processes	Secondary Metal Production	Lead Cable Coating
POINT	304049	Industrial Processes	Secondary Metal Production	Miscellaneous Casting and Fabricating
POINT	304051	Industrial Processes	Secondary Metal Production	Metallic Lead Products
POINT	305004	Industrial Processes	Mineral Products	Calcium Carbide
POINT	305013	Industrial Processes	Mineral Products	Frit Manufacture
POINT	305022	Industrial Processes	Mineral Products	Potash Production
POINT	305024	Industrial Processes	Mineral Products	Magnesium Carbonate
POINT	305026	Industrial Processes	Mineral Products	Diatomaceous Earth
POINT	305029	Industrial Processes	Mineral Products	Lightweight Aggregate Manufacture
POINT	305032	Industrial Processes	Mineral Products	Asbestos Milling
POINT	305033	Industrial Processes	Mineral Products	Vermiculite
POINT	305034	Industrial Processes	Mineral Products	Feldspar
POINT	305035	Industrial Processes	Mineral Products	Abrasive Grain Processing
POINT	305036	Industrial Processes	Mineral Products	Bonded Abrasives Manufacturing
POINT	305038	Industrial Processes	Mineral Products	Pulverized Mineral Processing
POINT	305042	Industrial Processes	Mineral Products	Clay processing: Ball clay

Table G-1. Source Classification Codes for Source Categories Containing 15 or Fewer Units (Cont.)

Data Category	SCC L3	SCC L1 Description	SCC L2 Description	SCC L3 Description
POINT	305044	Industrial Processes	Mineral Products	Clay processing: Bentonite
POINT	305045	Industrial Processes	Mineral Products	Clay processing: Fuller's earth
POINT	305046	Industrial Processes	Mineral Products	Clay processing: Common clay and shale, NEC
POINT	305089	Industrial Processes	Mineral Products	Talc Processing
POINT	305090	Industrial Processes	Mineral Products	Mica
POINT	305092	Industrial Processes	Mineral Products	Catalyst Manufacturing
POINT	314010	Industrial Processes	Transportation Equipment	Brake Shoe Debonding
POINT	315010	Industrial Processes	Photo Equip/Health Care/Labs/Air Condit/SwimPools	Photocopying Equipment Manufacturing
POINT	315027	Industrial Processes	Photo Equip/Health Care/Labs/Air Condit/SwimPools	Thermometer Manufacture
POINT	315031	Industrial Processes	Photo Equip/Health Care/Labs/Air Condit/SwimPools	X-rays
POINT	315040	Industrial Processes	Photo Equip/Health Care/Labs/Air Condit/SwimPools	Commercial Swimming Pools - Chlorination-Chloroform
POINT	316030	Industrial Processes	Photographic Film Manufacturing	Product Manufacturing - Substrate Preparation
POINT	316040	Industrial Processes	Photographic Film Manufacturing	Product Manufacturing - Chemical Preparation
POINT	316050	Industrial Processes	Photographic Film Manufacturing	Product Manufacturing - Surface Treatments
POINT	316060	Industrial Processes	Photographic Film Manufacturing	Product Manufacturing - Finishing Operations
POINT	316120	Industrial Processes	Photographic Film Manufacturing	Support Activities - Cleaning Operations
POINT	316130	Industrial Processes	Photographic Film Manufacturing	Support Activities - Storage Operations
POINT	316140	Industrial Processes	Photographic Film Manufacturing	Support Activities - Material Transfer Operations
POINT	316150	Industrial Processes	Photographic Film Manufacturing	Support Activities - Separation Processes
POINT	316160	Industrial Processes	Photographic Film Manufacturing	Support Activities - Other Operations
POINT	360001	Industrial Processes	Printing and Publishing	Typesetting (Lead Remelting)
POINT	390003	Industrial Processes	In-process Fuel Use	Lignite
POINT	401004	Petroleum and Solvent Evaporation	Organic Solvent Evaporation	Knit Fabric Scouring with Chlorinated Solvent
POINT	402028	Petroleum and Solvent Evaporation	Surface Coating Operations	Glass Optical Fibers
POINT	501002	Waste Disposal	Solid Waste Disposal - Government	Open Burning Dump
POINT	625400	MACT Source Categories	Food and Agricultural Processes	Cellulose Food Casing Manufacture

Table G-1. Source Classification Codes for Source Categories Containing 15 or Fewer Units (Cont.)

Data Category	SCC L3	SCC L1 Description	SCC L2 Description	SCC L3 Description
POINT	631110	MACT Source Categories	Agricultural Chemicals Production	2,4-D Salts and Esters Production
POINT	631250	MACT Source Categories	Agricultural Chemicals Production	Captan Production
POINT	631310	MACT Source Categories	Agricultural Chemicals Production	Chlorothalonil Production
POINT	631340	MACT Source Categories	Agricultural Chemicals Production	Dacthal Production
POINT	641300	MACT Source Categories	Styrene or Methacrylate Based Resins	Polymethyl Methacrylate Prod - Bulk Polymerization, Batch-cell Method
POINT	641301	MACT Source Categories	Styrene or Methacrylate Based Resins	Polymethyl Methacrylate Prod - Bulk Polymerization, Continuous Casting
POINT	641302	MACT Source Categories	Styrene or Methacrylate Based Resins	Polymethyl Methacrylate Prod-Bulk Polymeriz'n, Centrifugal Polymeriz'n
POINT	641310	MACT Source Categories	Styrene or Methacrylate Based Resins	Polymethyl Methacrylate Prod - Solution Polymerization
POINT	641320	MACT Source Categories	Styrene or Methacrylate Based Resins	Polymethyl Methacrylate Prod - Emulsion Polymerization
POINT	644200	MACT Source Categories	Cellulose-based Resins	Carboxymethylcellulose Production
POINT	644500	MACT Source Categories	Cellulose-based Resins	Cellulose Ethers Production
POINT	645200	MACT Source Categories	Miscellaneous Resins	Alkyd Resin Production, Solvent Process
POINT	645210	MACT Source Categories	Miscellaneous Resins	Alkyd Resin Production, Fusion Process
POINT	646100	MACT Source Categories	Vinyl-based Resins	Polymerized Vinylidene Chloride Production - Emulsion, Latex Prod.
POINT	646150	MACT Source Categories	Vinyl-based Resins	Polyvinyl Acetate Emulsions, Batch Emulsion Process
POINT	646200	MACT Source Categories	Vinyl-based Resins	Polyvinyl Alcohol Production, Solution Polymerization
POINT	646300	MACT Source Categories	Vinyl-based Resins	Polyvinyl Chloride and Copolymers Production - Suspension Process
POINT	646320	MACT Source Categories	Vinyl-based Resins	Polyvinyl Chloride and Copolymers Production - Solvent Process
POINT	646330	MACT Source Categories	Vinyl-based Resins	Polyvinyl Chloride and Copolymers Production - Bulk Process
POINT	648200	MACT Source Categories	Miscellaneous Polymers	Maleic Anhydride Copolymers Production - Bulk Polymerization
POINT	648210	MACT Source Categories	Miscellaneous Polymers	Maleic Anhydride Copolymers Production - Solution Polymerization
POINT	648220	MACT Source Categories	Miscellaneous Polymers	Maleic Anhydride Copolymers Production - Emulsion Polymerization
POINT	649200	MACT Source Categories	Fibers Production Processes	Rayon Fiber Production
POINT	651100	MACT Source Categories	Inorganic Chemicals Manufacturing	Antimony Oxides Manufacturing
POINT	651300	MACT Source Categories	Inorganic Chemicals Manufacturing	Fumed Silica Manufacturing
POINT	651350	MACT Source Categories	Inorganic Chemicals Manufacturing	Quaternary Ammonium Compounds Manufacturing
POINT	651400	MACT Source Categories	Inorganic Chemicals Manufacturing	Sodium Cyanide Manufacturing
POINT	685100	MACT Source Categories	Miscellaneous Processes (Chemicals)	Phthalate Plasticizers Production