

Introduction to the Emissions Modeling Framework

This course introduces participants to the main parts of the Emissions Modeling Framework (EMF). The EMF is a user-friendly tool for running the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system or other software.

The “Introduction to the Emissions Modeling Framework” training covers the following:

- Data Management for importing (loading), viewing, documenting and revising SMOKE input data
- Quality Assurance for checking SMOKE inputs before they are used in SMOKE. The QA Manager is covered in more detail in the EMF Advanced Training.
- Case Management for setting up and running SMOKE jobs

Background

This section provides some additional description about the EMF that can help those who are not taking this training in a classroom setting. In the classroom, these materials are covered in the form of Powerpoint® slides in a file named “Introduction to the EMF.ppt”.

EPA created the EMF for several purposes:

- Improve timeliness and quality of data used in air quality models
- Provide transparency & tracking
 - Data with versions, metadata, QA status
 - Emissions modeling applications and associated data
- Make it easier to run SMOKE and other emissions modeling tools
- Support multi-pollutant modeling: criteria & toxics
- Create tools that can be used by EPA and others (e.g., RPOs, states)

The main parts of the EMF are:

Data Manager: Stores and organizes input and output data. Includes metadata (Data Properties), Data Viewer, and Data Editor. Keeps both old and new versions of datasets for better tracking.

Case Manager: Stores and organizes software runs for SMOKE, Surrogate Tool, and Speciation Tool. Provides graphical user interface (GUI) for SMOKE or other emissions modeling tools.

QA Manager: Includes QA Templates, QA Steps and status. Runs SQL queries. Users define steps and record progress. Separate parts for QA of Data and Cases.

EMF Database: PostgreSQL heart of EMF. Contains actual emissions data, metadata, and information needed to drive EMF interface.

Control Strategy Tool (CoST):

- Users can build control strategies and create controlled SMOKE inputs.
- Contains control measures database
- More information in next training

Surrogate Tool: Creates spatial surrogate SMOKE input file from Geographic Information System (GIS) shapefiles using Java and the Multimedia Integrated Modeling System (MIMS) Spatial Allocator (Note: this is not yet covered by EMF training modules).

Speciation Tool: Creates speciation profile SMOKE input files for user-defined chemical mechanisms using PostgreSQL and PERL. Supports CMAQ toxic mechanism and CAP/HAP VOC integration (Note: this is not yet covered by EMF training modules).

SMOKE: Sparse Matrix Operator Kernel Emissions modeling system, which can be used to create emissions inputs to air quality models. The EMF provides an interface to SMOKE input data and runs.

Figure 1 below shows how these components relate and what is covered in the EMF Basic Training.

Figure 1: Major EMF components and scope of this training.

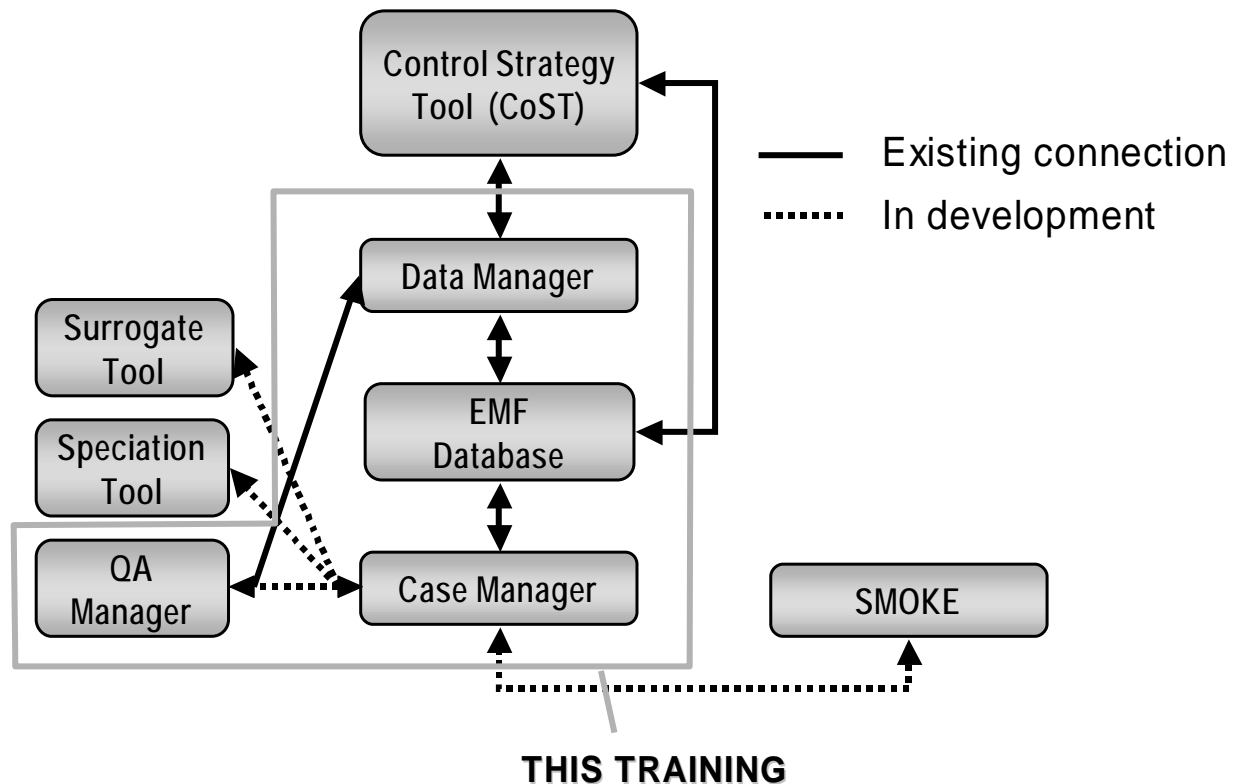
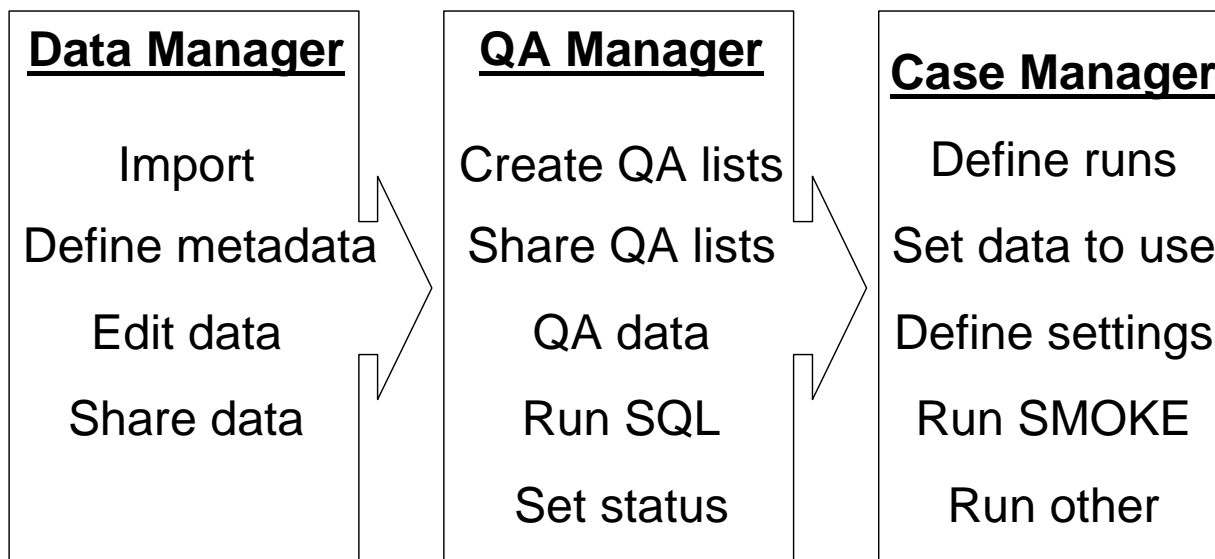


Figure 2 below helps to show the steps involved with using the EMF for running SMOKE or other applications.

Figure 2: Steps to use EMF for emissions modeling.



The critical information that you need to know about installing the EMF is available below. This list references additional online instructions for installing the EMF.

- Both Server and Client need to be installed
 - Require PostgreSQL 8.2, Java 6, and Tomcat 5.5
- Instructions available at:
<http://www.ie.unc.edu/cecpd/projects/emf/install/>
- Server: install 1 per site (shared database)
 - Mailing list will notify site admins of server updates
- Client: install 1 per user computer
 - EMF “Installer” handles initial install and updates. Simply double-click on icon and installation runs.
 - EMF Client tells you when its inconsistent with the version of the server.

A. Starting the EMF and logging in

The goal of this section is for you to (1) start the EMF application and (2) register as an EMF user.



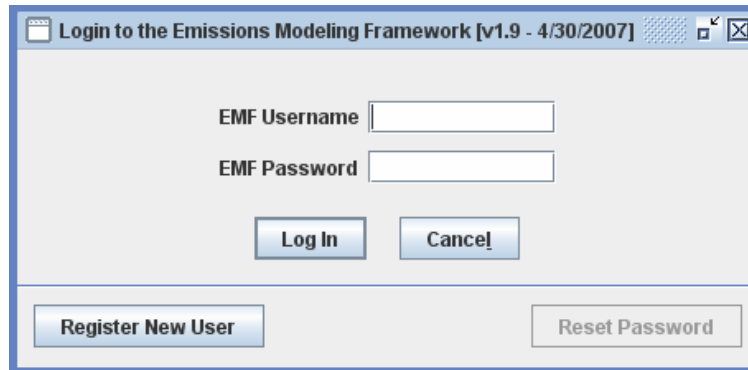
1. **Double click the EMF Client icon on your desktop:**

EMF Client

2. On Linux, you will next see the EMFlogin window
 - OR -

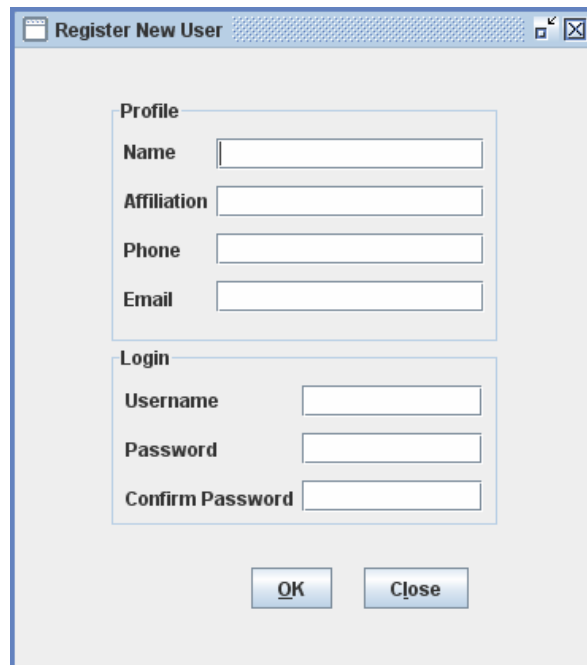
On Windows, you will see a command window (used to start Java). This window can be minimized or ignored.

You will see the main login window. When you are an existing user, you can simply enter your EMF username and password and click the “Log In” button.

A screenshot of a login window titled "Login to the Emissions Modeling Framework [v1.9 - 4/30/2007]". The window has a light gray background and a blue border. It contains two text input fields: "EMF Username" and "EMF Password". Below these fields are two buttons: "Log In" and "Cancel". At the bottom of the window, there are two more buttons: "Register New User" on the left and "Reset Password" on the right.

3. **Click the “Register New User” button** at the bottom left of the login window.

You will see a *Register New User Window*.

A screenshot of a "Register New User" window. The window has a light gray background and a blue border. It is divided into two main sections: "Profile" and "Login". The "Profile" section contains four text input fields: "Name", "Affiliation", "Phone", and "Email". The "Login" section contains three text input fields: "Username", "Password", and "Confirm Password". At the bottom of the window, there are two buttons: "OK" and "Close".

4. **Enter your Name, Affiliation, Phone Number, and E-mail address** in the “Profile” section of the *Register New User Window*.
5. **Enter a Username and Password** that you would like to use for your training:
 - a. Password must be at least 8 characters
 - b. Password must contain at least 1 numeric character
6. **Click the “OK” button**
7. You will now see the *EMF Main Window*

B. Main EMF window

The goal of this section is for you to (1) learn about the menu choices on the EMF *Main Window* and (2) learn about the *Status Window*.

The main window is what you see after you log into the EMF.

B.1. Main Menu

The *Main Window* also contains the *Main Menu* at the top. These menus allow you to access the EMF features.

The “File” Menu

File	
Import	Goes directly to the Importer part of the Data Manager
Logout	Logs out of the EMF Client
Exit	Closes the EMF Client

The “Manage” Menu

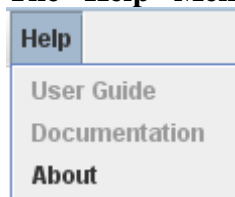
Manage	
Datasets	Opens the Data Manager
Cases	Opens the Case Manager
Dataset Types	Opens the Dataset Type Manager (for definition of dataset types)
Sectors	Opens the Sector Manager (not covered by training)
Control Measures	Opens the Control Measures Manager (CoST/Advanced training)
Control Strategies	Opens the Control Strategies Manager (CoST/Advanced training)
My Profile	Opens a window to edit user profile (cannot change user name). For other settings, the EMF uses a file called “EMFPrefs.txt”. On UNIX, this file is located in /home/training. On Windows, this file is in the C:\Documents and Settings\<username> directory.

When the Speciation Tool and Surrogate Tool are integrated with the EMF, they will be treated by the EMF as “Cases”; therefore, they will be available from the Case Manager. The Case Manager will run these tools and organize multiple runs. These tools may also be available via separate menu items.

The “Window” Menu

Window	
Close All	Closes all open EMF Windows, except for the main Window and the Status Window
Status	Brings up the EMF Status Window

The “Help” Menu



Opens the EMF User Guide web pages (not yet implemented)

Opens the Document Manager (not yet implemented)

Shows the version of the EMF being run

B.2. Status Window

The *Main Window* contains the *Status Window* at the bottom.

- The EMF uses the Status Window to provide you messages, including
 - Import and export status
 - QA status (when you run a QA program, e.g., SQL)
 - Error messages
- The *Status Window* “Refresh” button is used to force the window to get updated. Otherwise, the EMF Client checks for new messages from the server every two minutes.
- The “Trash” button allows you to clear messages from the window.
- A scroll bar allows you to scroll through messages when the small window is filled up.

The Status Window



C. Data Manager

The purpose of this section is for you to learn how to use the EMF Data Manager. The Data Manager has the following major features:

- Stores SMOKE (or other) emissions modeling input data in a PostgreSQL database
- Facilitates storing and maintaining metadata (information about data)
- Facilitates users changing data and maintaining multiple versions of datasets
- Facilitates multiple users accessing the same data using a Client-Server approach
- Records who, what, when, and why of data changes and make that history available
- Records which versions of data were used in which SMOKE (or other) applications
- Support access of QA Manager functions

C.1. The *Dataset Manager Window*

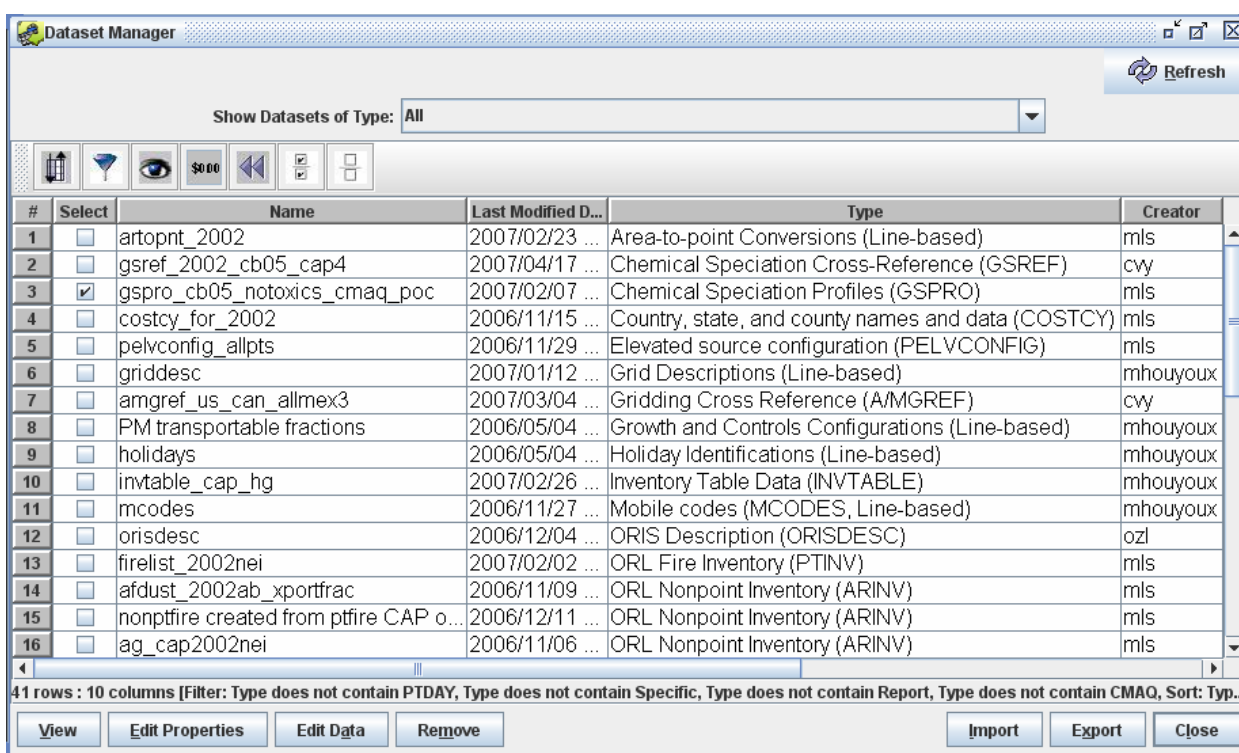
In this subsection, we will:

- View, sort, and filter a list of datasets from the *Dataset Manager Window*
- Learn about the EMF Toolbar functions and practice using it
- Use the Dataset Manager to find SMOKE emissions summaries in the EMF

C.1.a. View list of datasets

The Dataset Manager provides a list of datasets. To access the Dataset Manager:

1. On the Main Menu, **click the Manage → Datasets menu option**. You will next see the *Dataset Manager Window*.
2. To see a list of Datasets, you must specify a type of dataset.
Using the “Show Datasets of Type:” pull-down list, choose “All”. The interface will pause momentarily while the Client accesses the Data list from the EMF Server (recall that for the training, the laptops are each running the Client and the Server).



#	Select	Name	Last Modified D...	Type	Creator
1	<input type="checkbox"/>	artopnt_2002	2007/02/23 ...	Area-to-point Conversions (Line-based)	mls
2	<input type="checkbox"/>	gsref_2002_cb05_cap4	2007/04/17 ...	Chemical Speciation Cross-Reference (GSREF)	cvy
3	<input checked="" type="checkbox"/>	gspro_cb05_notoxics_cmaq_poc	2007/02/07 ...	Chemical Speciation Profiles (GSPRO)	mls
4	<input type="checkbox"/>	costcy_for_2002	2006/11/15 ...	Country, state, and county names and data (COSTCY)	mls
5	<input type="checkbox"/>	pelvconfig_allpts	2006/11/29 ...	Elevated source configuration (PELVCONFIG)	mls
6	<input type="checkbox"/>	griddesc	2007/01/12 ...	Grid Descriptions (Line-based)	mhouyoux
7	<input type="checkbox"/>	amgref_us_can_allmex3	2007/03/04 ...	Gridding Cross Reference (AMGREF)	cvy
8	<input type="checkbox"/>	PM transportable fractions	2006/05/04 ...	Growth and Controls Configurations (Line-based)	mhouyoux
9	<input type="checkbox"/>	holidays	2006/05/04 ...	Holiday Identifications (Line-based)	mhouyoux
10	<input type="checkbox"/>	invtable_cap_hg	2007/02/26 ...	Inventory Table Data (INVTABLE)	mhouyoux
11	<input type="checkbox"/>	mcodes	2006/11/27 ...	Mobile codes (MCODES, Line-based)	mhouyoux
12	<input type="checkbox"/>	orisdesc	2006/12/04 ...	ORIS Description (ORISDESC)	ozl
13	<input type="checkbox"/>	firelist_2002nei	2007/02/02 ...	ORL Fire Inventory (PTINV)	mls
14	<input type="checkbox"/>	afdust_2002ab_xportfrac	2006/11/09 ...	ORL Nonpoint Inventory (ARINV)	mls
15	<input type="checkbox"/>	nonptfire created from ptfire CAP o...	2006/12/11 ...	ORL Nonpoint Inventory (ARINV)	mls
16	<input type="checkbox"/>	ag_cap2002nei	2006/11/06 ...	ORL Nonpoint Inventory (ARINV)	mls

41 rows : 10 columns [Filter: Type does not contain PTDAY, Type does not contain Specific, Type does not contain Report, Type does not contain CMAQ, Sort: Typ...

View Edit Properties Edit Data Remove Import Export Close

3. Using the dataset list that appears, **scroll through the list of datasets that have been pre-loaded onto this EMF installation**.

SMOKE users should recognize many of the types of datasets listed. The names in parentheses are the SMOKE files environment variables.

4. **Review the columns available in the list by using the left-right scroll bar**. These fields are some of the Metadata available for each Dataset. We will show you how to edit this information later in this training.

5. **Sort the datasets by Dataset Type by clicking on the Dataset Type column. Try sorting using other columns.**
6. Note that this window list cannot be edited by typing directly into the window.

C.1.b. The EMF toolbar

The Dataset Manager also provides a toolbar for sorting and filtering, as shown here:



This toolbar also appears on many other EMF Windows. The purpose of each icon is:

	Sort
	Apply a filter to limit which rows are shown.
	Show/Hide columns and apply filter
	Format by column the font and display of numbers
	Reset to initial table (clear sorts, filters, and formats)
	Select all rows
	Deselect all rows

Exercise 1

With the “Show Datasets of Type” menu set to “All”, use the toolbar to show all datasets that are in ORL format. Hint: all ORL Dataset Types have the letters “ORL” in the Type name.



Question: How many ORL inventory datasets are loaded into the EMF? _____

Exercise 2

Find all of the SMOKE output reports loaded into the database (the Dataset Type is called “SMOKE Report”).



Question: What are the two ways that you can do this?

1. _____
2. _____

If you have time, feel free to explore the Dataset Manager some more.

C.2. Importing datasets

The purpose of this section is for you to learn the following:

- The difference between creating a Dataset and Dataset Type
- The difference between internal and external datasets
- How to import internal datasets
- How to import external datasets
- How to use the multiple-file feature when importing datasets

C.2.a. Dataset Definitions

Dataset Type: A specific file format and data structure (fields and their types) that the EMF expects when importing, storing, and exporting data. The Dataset Types relate directly to the files and the formats needed by SMOKE or other programs run from the EMF.

Dataset: Data and associated metadata stored in the EMF database, which has a specific Dataset Type. The EMF database can store Datasets in two ways:

- Internal datasets: Data values are stored inside the database as a PostgreSQL table
- External datasets: Data values are stored as a file (on your hard drive), but the EMF knows where the file is and keeps some of the Metadata as with an internal dataset. The change histories are not kept for external datasets. External datasets are used for binary files (including SMOKE output files) and any dataset for which a user decides not to store internally, such as large datasets or those that rarely change.

C.2.b. Creating Datasets versus creating Dataset Types

Datasets are created in one of the following ways. Creating Datasets relies on a Dataset Type to already have been created for the Dataset of interest.

- Import a data file from disk (Basic Training)
- Run CoST to create a detailed control strategy result, summary tables, and a controlled inventory (Advanced Training)
- Run SMOKE or another program to create output files and automatically import them to the EMF. (Implementation ongoing)
- Run the Speciation Tool or the Surrogate Tool to create SMOKE input files, and automatically import them to the EMF (Implementation ongoing)

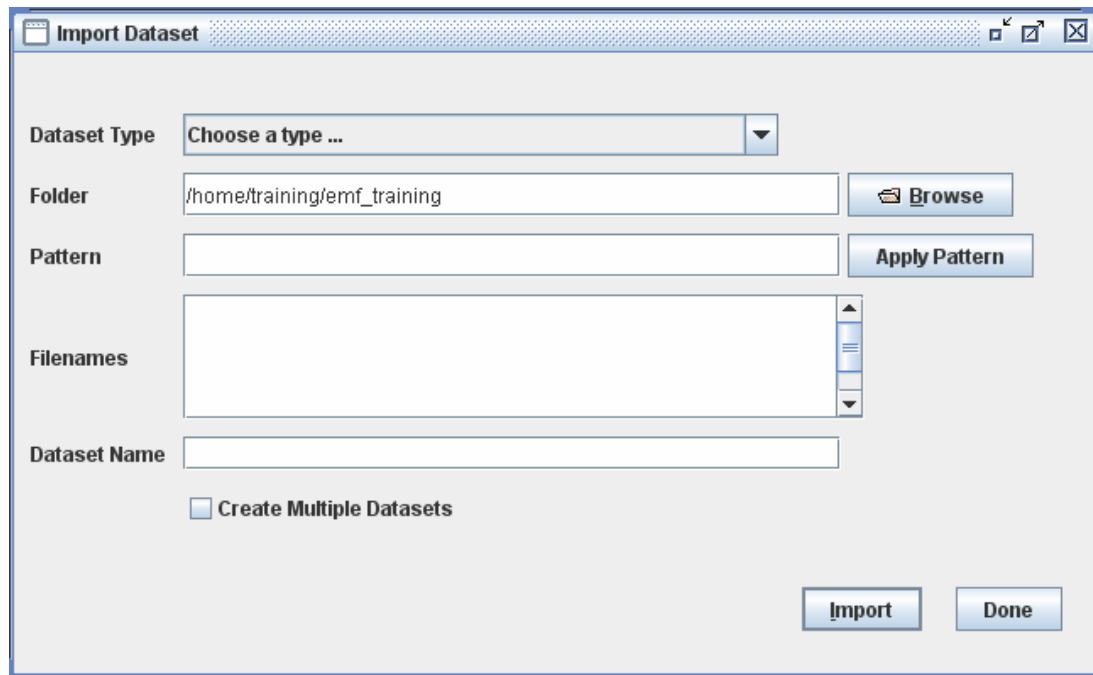
Dataset Types are created using the *Main Menu*: Manage → Dataset Types. We will look at Dataset Types more closely as part of the EMF Advanced Training.


C.2.c. Importing an “Internal” dataset

We will now import an annual on-road mobile dataset as an internal dataset. This Dataset will have a (predefined) Dataset Type called “ORL Onroad Inventory (MBINV)”. Note, the dataset file must reside on the server computer, not the client’s desktop.

1. Start the *Import Dataset Window* in one of two ways:
Click the “Import” button on the Dataset Manager.
---- OR ----
Select the File → Import menu item from the Main Menu

The *Import Dataset Window* will appear.



2. From the “Show Datasets of Type” menu, select “ORL Onroad Inventory (MBINV)”. Since this Dataset Type is an internal type, the EMF will import the data values into the database.
3. **Click “Browse”** to select the file to import.
 - a. The *File Browser Window* will appear.
 - b. The *Browser* will start you in the directory /home/training/emf_training/tmp. You can navigate by either typing the explicit path into the Folder, or by double-clicking on the subfolder. Note “..” subfolder will move you up a directory.
 - c. **Select the checkbox**  **next to the file named “NC_SC_VA_2020_National_Controls_w_NLEV_Crxns_OnRoad.txt”**
 - d. **Click “OK” on the File Browser Window**

4. **Change the Dataset Name** field to a name that does not contain a date, version number, or file extension.

The Dataset Name is automatically populated by the EMF using the data file name. However, since the EMF Exporter adds the date last edited, the version number, and the file extension (e.g., .txt), these should be removed from the dataset name manually during import.

For example, if the data file were named mbinv_onroad_2002ann_04mar2007_v4.txt, then you might want to make the dataset name “onroad_2002ann”.

We cover exporting datasets and the export naming convention in a later section of this training.

5. **Click “Import” on the *Import Dataset Window*.** The *Status Window* will show the status of the import. It will send a message when the import process has started and when it has finished. You are free to do other things in the EMF while the import is ongoing, it will not “lock up” the EMF Client interface.
6. Note - The *File Browser* shows you files on the EMF server, so when the EMF Server and Client are running on different machines, your data files to import need to reside on the Server machine.
7. Do not click the “Done” button (yet). This button closes the *Import Dataset Window*, but we are not done importing yet. If you happen to click “Done” by mistake, reopen the Import window to be ready for the next step of the training.
8. Now, **find the Dataset you imported in the *Dataset Manager*.** You will need to (a) have the “ORL Onroad Inventory” Dataset Type selected and (b) **click the Refresh button** for the newly imported dataset to appear in the list.

Note: If there is a problem with the format of the dataset, for example, wrong number of fields or a character value in a numeric field, or a mix of different delimiters (commas& semicolons), the dataset will not import and an error will be shown in the status bar, with an explanation of the problem.



Question: What is the “Last Modified Date” that shows in the table for your new file?

C.2.d. Importing an “External” dataset

We will now import a set of spatial surrogates as an external dataset. This Dataset will have a (preloaded) Dataset Type called “Spatial Surrogates (External Multifile)”. Although the spatial surrogates are separated into 197 individual files, we will import all of these files as a single EMF dataset, using the multiple-file import feature. Note, the external dataset files must reside on the server computer, not the client’s desktop.

1. From the “Dataset Type” pull-down menu, **select “Spatial Surrogates (External Multifile)”**. Since this Dataset Type is an external type, the data will be registered with the EMF database, but the data will be present only as ASCII files on the EMF server.








2. Click **“Browse”** to select the directory from which you will import.
 - a. The *File Browser Window* will appear.
 - b. Choose the subfolder **36km_surg**, i.e. the directory
/home/training/emf_training/tmp/36km_surg
 - c. Click **“OK”** on the *File Browser Window*
3. In the **“Pattern”** box of the *Import Dataset Window*, enter:
***.txt**
Alternatively, you can use the **“Browse”** button to select all of the *.txt files.
4. Click the **“Apply Pattern”** button. The EMF will populate the list of all of the surrogate files being imported. If needed, you can manually edit this list; however, we do not need to edit the list for this step.
5. Enter a dataset name for your dataset. Since the exporter will not use this dataset name, you are free to include date and version information in the Dataset Name. The exporter does not actually export these files, since the data values have not actually been imported.
6. Click **“Import”**. Check the *Status Window* for the import status of this dataset.
7. Find the Dataset in the *Dataset Manager* table.



Question: How do you know which Dataset Types are “external”?

C.3. Dataset Manager buttons

The *Dataset Manager Window* has many buttons at the bottom of the window. This section simply describes the purpose of each of the buttons.

	Views the Dataset Properties and data values (internal datasets only). After opening in “View” mode, with an extra click, you can also edit the properties and data values.
	Opens the <i>Dataset Properties Editor</i> so you can edit the Dataset Properties ¹ .
	Opens the <i>Data Editor</i> so you can edit the data values for a dataset ² .
	Removes a dataset from the Dataset Manager ³ .
	Opens the <i>Import Dataset Window</i> for importing datasets
	Opens the <i>Export Dataset Window</i> for exporting datasets
	Closes the <i>Dataset Manager Window</i>

¹ The EMF locks the Dataset Properties (for that dataset) for 2 hours, which prevents other users from viewing or editing the Dataset Properties for that dataset. The EMF releases the lock in 2 hours or after you close the Dataset Properties Editor.

² The EMF locks the Dataset (values) for 2 hours. The EMF renews the lock each time you save changes to the dataset while editing. The EMF releases the lock 2 hours after the last change or after you close Data Editor.

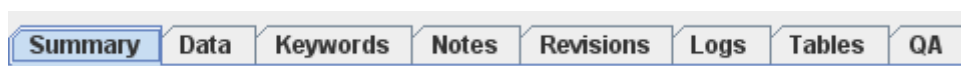
³ Notes on the “Remove” button (1) at this time, a removed Dataset still resides in the EMF Database, but you can no longer access it through the EMF interface, (2) datasets used in Cases cannot be deleted, (3) only the owner of a dataset (or an administrator) can delete a given dataset.

C.4. Dataset properties

The purpose of this section is for you to learn about Dataset Properties and how they are different from the data values.

“Dataset Properties” refers to a dataset’s metadata. A few key points about Dataset Properties:

- Using the EMF, Dataset Properties are stored and accessed separately from the data values.
- Some Dataset Properties are not version-specific, such as the summary, keywords, and tables.
- Internal datasets have both Dataset Properties and data values. External datasets just have Dataset Properties – the “values” simply list the paths/names of the files on disk.
- Dataset properties include the following groups of properties, which are listed on separate tabs of the Dataset Properties Viewer and Editor.



- The “Data” tab allows you to access the data values provided in rows and columns.
- The other tabs give access to the Metadata, which we describe in the next several subsections.

C.4.a. Viewing Dataset Properties

To view Dataset Properties, you would select one or more datasets to view from the *Dataset Manager*, and then click the “View” button. This is different from the *Editor* only in that you cannot change any of the values. It is useful to use the *Viewer* when you want to see the information and not risk accidentally changing it. We also plan to implement dataset permissions to allow users to set limits on who can edit their datasets (e.g., user, group, and all write permissions).

Since the *Dataset Properties Viewer* and *Editor* are nearly the same, we will tour them using the *Dataset Properties Editor* so that we can also add to the example metadata.

C.4.b. Editing Dataset Properties

In this section, we will explore each of the tabs in the *Dataset Properties Editor*, except for the “Data” tab.

1. From the *Dataset Manager*, **select the dataset named “gsref_2002_cb05_cap4” by clicking the checkbox next to the dataset.** This is the SMOKE CB05 speciation cross-reference file.
2. **Click the “Edit Properties” button on the *Dataset Manager Window*.** After a short time, you will then see the Dataset Properties Editor, with the tabs that are shown in the picture above. The *Summary* tab will be selected.

Dataset Properties Editor: gsref_2002_cb05_cap4

Summary | Data | Keywords | Notes | Revisions | Logs | Tables | QA

Name: gsref_2002_cb05_cap4

Description:
 # SMOKE GSREF file for CB05 for CAPS
 # CB05 chemical mechanism for the CMAQ model, including PM2.5 and Hg profiles.
 #

Project: [Dropdown]

Creator: cvy

Dataset Type: Chemical Speciation Cross-Reference (GSREF)

Time Period Start: [Text Box]	Status: Imported
Time Period End: [Text Box]	Last Modified Date: 04/17/2007 17:44
Temporal Resolution: Choose a resolution [Dropdown]	Last Accessed Date: 03/20/2007 12:06
Sector: Choose a sector [Dropdown]	Creation Date: 03/20/2007 12:06
Region: [Dropdown]	Intended Use: [Dropdown]
Country: Choose a country [Dropdown]	Default Version: 0 (Initial Version) [Dropdown]

Save **Close**

3. **Summary Tab**

Review the different Metadata in the *Summary* tab. Note that the only item here that is version-specific is “Default version”. Thus, the values selected need to apply to all versions of a dataset. If they do not, then you need two datasets instead of one.

You can edit any of the fields that you wish. Experiment and see how things work.

Name: The name of the dataset. Same as the name that appears in the Dataset Manager.

Description: This information will be written to the header of the exported ASCII version of the data. It can be used to document anything about the dataset that you’d like to see go into the actual SMOKE file header.

Project: User-defined project name. Can type an entry and reuse it later.

Creator: The EMF user who imported the dataset

Time Period Start and End: User-defined time period start and end. Some dataset types (such as annual inventories) have default time periods set.

Temporal Resolution: User-selected (annual, monthly, etc.)

Sector: User-selected sector of the emissions. Sectors are defined by the *Sector Manager*.

Region: User-defined region name. Can type an entry and reuse it later.

Country: User-defined county name.

Status: EMF-set status (either “Imported” or “Created by control strategy” at this time)

Last Modified: Last modified date of the dataset. This is initialized based on the data file date for imported files.

Last Accessed: The import date or the last exported date of the dataset, if the dataset has been exported.

Creation Date: Date that the dataset was created (imported, created by CoST, SMOKE, Surrogate Tool, or Speciation Tool)

Intended Use: User-defined intended use, which is meant for now to let friendly users see that other users may have “private” data. The EMF does not yet enforce “private” versus “public” data, but we have plans to support that in the future.

Default Version: User-defined version to be treated by all users as the official version. Helps clarify intent of versions when some users may have versions in development or have special sensitivity versions. For now, the Dataset Manager can only export the “default version”, but the Case Manager can export any version.

4. **Keywords Tab**

Keywords are a place for you to define supplemental metadata for your dataset.

Additionally, the EMF has “special” keywords that support various features. Note that the keywords are not version-specific, thus all keywords need to apply to all versions of a dataset.

- a. **Click on the Keywords tab**
- b. To add a note, **click on the “Add” button**. The EMF then adds a blank line to the table on the tab.
- c. In the left column (not the checkbox) on the new line, **enter your own keyword or select from one of the preset keywords**. You will be able to reuse this keyword in other datasets once you have entered it for one dataset.
- d. In the right column on the new line, **enter the value for the keyword**.

Some important preset keywords are as follows:

EXPORT_SUFFIX: The EMF appends the value of this keyword to the dataset name when exporting data. At EPA, we use _orl.txt, _ida.txt, and .txt as export suffixes.

EXPORT_PREFIX: The EMF starts the name of exported datasets with the value of this keyword. At EPA, we have not used this as much, but it could be set to the beginning of file names, such as “arinv” for the SMOKE area-source files.

REPLACED_BY_DATASET: Users can set this keyword to indicate when a dataset has been replaced by another dataset. Sometimes its useful to reimport a dataset rather than edit it in the EMF. If you need to document this transition, this keyword can help you do that.

REPLACES_DATASET: This is similar to REPLACED_BY_DATASET, but it is a keyword used for the dataset that has replaced another dataset, to indicate what dataset has been replaced by it.

5. **Notes Tab**

The purpose of notes is to record information about datasets that you observe and want to document. Notes can include observations, problems, and references to supporting documentation for a dataset. The “notes” can also be added from the Data Editor, as you will see later.

a. **Click on the Notes tab**

- b. To add a note, **click on the “Add” button**. You will see a *Create New Note Window* appear. **Enter any note that you wish by filling in the following fields with whatever you like:**

Summary: This is like a name for the note. It’s what you will see when you look at the list of notes on the tab.

Type: A number of preset note types are available. As you can see, notes can be also used to flag problems and post resolutions to them. Notes can also be used to document references for the dataset.

Details: This space is for you to write the details about the note (a description of the problem, resolution, question, observation, or reference materials).

In the case of additional data documentation, the path and name of an external file can be entered. In fact, these reference materials could be added to the EMF as their own dataset, and the Note could reference the EMF dataset that has the materials.

References: This space allows you to select another note that your note reference. A classic example is listing the “problem” as the reference for a “resolution” Note.

Version: The Dataset version to which this note applies.

6. **Revisions Tab**

This tab stores the “who”, “what”, “when” and “why” about dataset revisions. Initially, it is blank, but rows will be added as users make revisions to data using the *Dataset Editor*. This tab allows you to view the revision information in summary (the table on the tab) and in detail (by clicking the “View” button). We’ll look at this more later when we use the Dataset Editor.

7. **Logs Tab**

This tab contains the “usage history” of the dataset. It currently records dataset exports (who, when, optionally why) and documents how many lines were exported (for use by users for QA purposes).

8. **Tables Tab**

This tab is primarily useful for datasets with multiple tables, such as the SMOKE temporal profiles or NIF datasets. It shows what tables are in the dataset and how large they are.

9. **QA Tab**

This tab provides the main access point to the QA Manager. You can define QA steps for both Dataset Types and Datasets.

- QA Steps on Dataset Types provide a template that can be used by all Datasets of that type
- QA Steps added to Dataset only can be customized to an individual dataset, but cannot be shared among other datasets of that type.
- QA Steps are version-specific. Once the step has been added for a specific version, it can be added and repeated in a later version.

You will learn more about the QA tab in the QA Manager section of the training.

Use the information provided above to answer the following questions. All questions refer to the dataset for which you have been reviewing the properties.



Questions:

- How many revisions have been made in the EMF to the dataset? _____
- How many versions are there (include version 0, which was the originally imported data)? _____
- What are the prefix and suffix that will be used when this dataset is exported?
Prefix: _____ Suffix: _____
- When was this dataset last exported and what user exported it?
User: _____ Date: _____
- Which Excel file created this dataset prior to it being imported into the EMF (hint: use the “Summary” tab). _____

C.5. Data Viewer

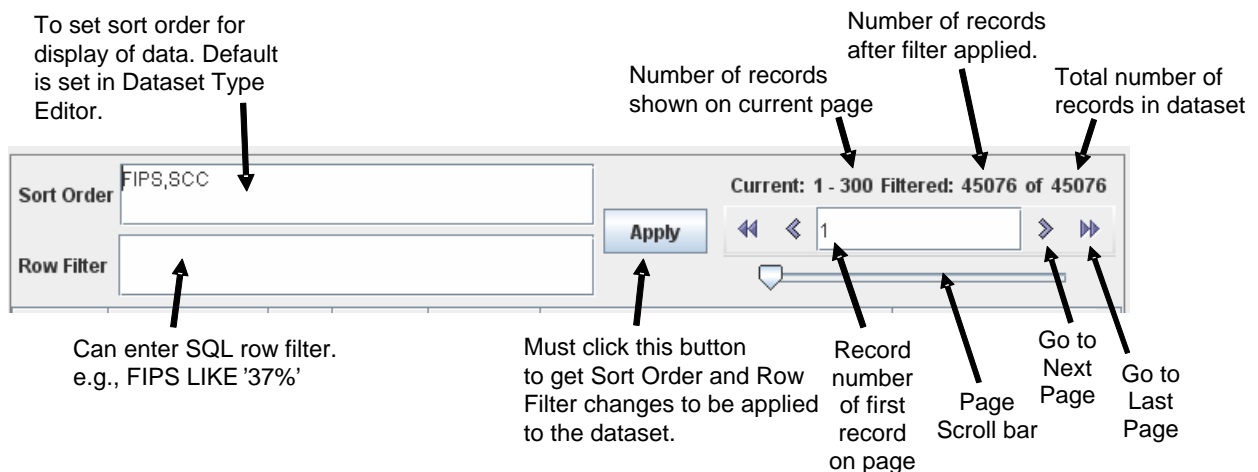
The purpose of this section is for you to become familiar with the Data Viewer’s main features, including:

- Dataset pagination
- Sort and filter
- Copy/paste values from EMF into other applications
- Viewing a multi-table dataset vs. a single-table dataset

The *Data Viewer* allows you to see (but not edit) the data values of your datasets. All of the Dataset Viewer’s features are also available on the *Data Editor*.

1. From the *Dataset Manager*, select the Dataset Type “**ORL Nonpoint Inventory (ARINV)**”. This is an example of a Dataset Type with a single table. An exercise will give you an example of a Dataset Type with multiple tables.
2. Next from the *Dataset Manager*, select the dataset named “**afdust_2002ab_xportfrac**” by finding and clicking on the checkbox next to the dataset name in the list.
3. Click the “**View**” button on the Dataset Manager. You will next see the *Dataset Properties Viewer Window* appear.
4. Click on the “**Data**” tab.
5. Note that the Initial Version (0) is selected from the pull-down menu, since this is the only available version. This is how you choose which version to view when there are multiple versions. Click the “**View**” button. You will next see the *Data Viewer*.

The top of the *Data Viewer* contains the interface for several of the features we will review next. The following is a annotated picture of the top of the *Data Viewer*.



6. Click the “Go To Next Page” button to see how scrolling through the dataset works.



Question: What is the value of PM2.5 on the last row of page 2? _____

7. Click the “Go To Last Page” button to go to the end of the dataset.



Question: What is the SCC shown on the first row of the last page? _____

8. In the Sort Order box, enter the following sort request and click the “Apply” button.

SCC desc, FIPS



Question: What did the “desc” do to the sort order?

9. In the Row Filter box, enter the following filter and click the “Apply” button.

FIPS = '42037' AND SCC LIKE '2311%'



Question: How many records are shown after you apply this filter? _____

10. Here are some more examples of filters:

- a. POLL = 'NOX'
- b. POLL IN ('CO', 'VOC', 'PM10')
- c. POLL <> 'NOX'
- d. FIPS LIKE '37%'
- e. ANN_EMIS > 1000 AND FIPS LIKE '37%'
- f. FIPS='37001' OR FIPS='37003'

11. The data values from a single page of the *Dataset Viewer* can be cut and paste as tab-delimited ASCII data into other applications. When running the EMF from Windows®, this includes pasting into Microsoft® Excel®. In this example, we will paste into a text editor called “gedit”.

- a. **From the menu at the top of the Linux Desktop: Applications→Accessories→Text Editor.** This will open up text editor window.
- OR -
In Windows, open the Notepad from the Windows Start→ Accessories menu.
- b. **Back in the EMF, use your mouse to highlight the rows resulting from the Filter applied in Step 9.**
- c. **With the rows selected, type <Ctrl>-C to copy the data.**
- d. **In the gedit window, type <Ctrl>-V to paste the data.**

Exercise 3

Use the Dataset Manager to find a “Country, State, County” SMOKE file (Dataset Type called “Country, state, and county names and data (COSTCY)”. Open the Dataset in the Data View. Use the “Data” tab to view the state table (note: you will have to figure out how to pick the state table).



Question: How many states are included in this table?_____

Exercise 4



Use the Dataset Manager to find and view the Dataset you imported in Section C.2.c (NC_SC_VA_2020_National_Controls_w_NLEV_Crxns_OnRoad). Find the largest NOX emissions value for a single record. _____

C.6. Data Editor

The purpose of this section is for you to become familiar with the following aspects of the Data Editor:

- The difference between “View” and “Edit”
- Creating versions and editing data
- Dataset Properties associated with data editing
 - The Revisions tab
 - Last Modified Date on the Summary tab

C.6.a. The difference between “View” and “Edit”

Editing a dataset in the EMF is more than just opening and changing. It involves the following steps:

1. Pick your dataset and open it in edit mode
(with viewing, you open in “view” mode and you are done)
2. Add a new version
3. Open a table of the new version in the Data Editor
4. Change the data
5. Enter metadata about “what” was changed and “why” it was changed.
6. Save and close the data
7. Set the changed version to “final”.

C.6.b. Creating versions and editing data

The following numbered items walk you through an example of each of the steps above.

1. Pick your dataset and open it in edit mode
 - a. From the *Dataset Manager*, **select the Dataset Type “Chemical Speciation Profiles (GSPRO)”**
 - b. Next from the *Dataset Manager*, **select the dataset named “gspro_cb05_notoxics_cmaq_poc” by finding and clicking on the checkbox next to the dataset name in the list.**
 - c. **Click the “Edit data” button on the *Dataset Manager*.** You will next see the *Dataset Versions Editor Window* appear. This is the window that you use to select which version you will modify and to add a new version.
2. **Click the “New” button to add a new version.** You will next see a *Create a New Version of <Name> Window*, with the name of your dataset included in the dialog window name. You use this window to define the version you are creating. In that window:
 - a. Select the version you are using as the basis for your new version. **In this case, select the only available option: Version 0.**
 - b. **Enter any name you wish in the “Name” box available in the *Create a New Version Window*.** This is the name of your version.
 - c. **Click the “OK” Button.** The *Create a New Version Window* will disappear and you will now see your version listed in the *Dataset Versions Editor Window*

3. In the *Dataset Versions Editor Window*, **select your new version by clicking the checkbox next to the Version Name. Then, click the “Edit” button.** You will now see your data in the *Data Editor*.

The top of the data editor is the same as the Data Viewer, with two differences.

- a. First, there is a message about a “lock” at the top right of the *Data Editor*:

Lock expires at 2007/04/20 17:14

This is the lock described in the table in section C.3. The lock prevents other users from accessing your version while you are editing it. The lock is updated each time you save a change, and it will expire after 2 hours. The expiration prevents others from being permanently locked out of the dataset in case you leave the editing session open when you leave the office or go on vacation!

- b. Under the main header is a new set of icons that you can use for helping to edit the data.



Use the following table to help you understand the purpose of these buttons.

	Insert row above <u>highlighted</u> row (click anywhere on row with mouse to highlight it). Note: the checkboxes do not affect this operation.
	Insert row below <u>highlighted</u> row (click anywhere on row with mouse to highlight it). Note: the checkboxes do not affect this operation.
	Delete <u>selected</u> rows (select row using checkbox)
	Select all rows of dataset
	Deselect all rows of dataset

4. Change the dataset in the following ways:

- a. **Add two rows and insert the following speciation profile:**

Code	Pollutant	Species	Split	Divisor	Massfrac
16007	TOG	FORM	0.58	30.0	0.58
16007	TOG	PAR	0.42	14.4	0.42

- b. **Select all rows with Code = ‘0003’ and delete these records.** Hint: the “filter” can help you make sure you select all rows with Code = ‘0003’.

5. **Save using the “Save” button and try to close your data using the “Close” button.**

Whoops! Unless you have entered information in the “What was changed” and “Why it was changed” fields, the *Data Editor* will not let you close the editing session without filling in these fields.



What was changed	Why it was changed
<input type="text"/>	<input type="text"/>

Question: Where on the *Data Editor* did the EMF tell you what you did wrong?

Fill in information about the change in the boxes provided. You can of course circumvent the process by entering meaningless information if you choose to do that, but it will be recorded for all too see that you didn't document your changes!

Note that you can only enter information about “what” and “why” if you actually made changes. Otherwise, these fields are inactive.

6. **Now close the Editor using the “Close” button** to return to the *Dataset Versions Editor*.
7. A few notes about making versions “final”:
 - Once you have finalized your changes for the current version, then you mark your version as “Final”.
 - A dataset can always be modified by adding another version – it's not the dataset that is final, just the version.
 - Once a version is “final” it cannot be changed. To change the dataset after making a version final, you must create the next version.

Mark your version as final as follows – **select your version with a checkbox and click the “Mark Final” button on the *Dataset Versions Editor***. The checkbox in the “Is Final?” column will change from empty to checked next to your dataset version.

C.6.c. Dataset Properties associated with data editing

Now that we have edited a dataset, you can see how that affected the metadata available from the Dataset Properties tab.

1. From the *Dataset Manager* **select the same GSPRO file you edited before and click the “Edit Properties” button.**
2. **Click on the “Revisions” tab.** You should see the information about your revision stored as the last entry in that tab. Note that the user who made the change and the date were stored automatically. The “what” and “why” contents you included show up here.
3. **Click on the “Summary” tab.**
 - a. Note that the “Last Modified Date” has been changed to reflect your new version.
 - b. Under the “Default Version” menu, note that you now have a choice – you can make your version the default if you wish. A few notes on the Default Version:
 - You can only make “Final” versions the Default Version.
 - The “Default Version” is the only version that you can export from the Dataset Manager.
 - The Case Manager can export any version – even if not marked Final.

C.7. Export data from Dataset Manager

The purpose of this section is for you to learn how to export data from the Dataset Manager:

- Exporting one or more files
- Built-in file naming convention
- Using keywords for prefixes and suffixes

C.7.a. Exporting data files from the Dataset Manager

The Dataset Manager allows you to export any dataset into an ASCII dataset. Some key points about the Dataset Manager Export capabilities:

- You can also export multiple datasets of the same and different dataset types at the same time.
- You do not have to wait for the export to complete to continue using the EMF.
- Export continues even if you log out of the system
- External datasets are not exported (since the data has never been inside the database)
- The EMF exports SMOKE input data in SMOKE formats, and it export SMOKE reports in a semi-colon delimited format with some additional header information.
- The export is going to the server machine only. It's not going to your desktop machine.

You will now practice using the Export feature. Take the following steps:

1. **Open the *Dataset Manager Window***
2. Use the pull-down at the top of the *Dataset Manager* to **select Dataset Type of “SMOKE Report”**.
3. **Click the appropriate checkbox at left to select the Dataset named “rep_nonpt_inv_county_2002ac_nonpt.txt”**
4. **Click the “Export” button** at the bottom right of the *Dataset Manager* and you will see the *Export Window*, as shown below.

Export: invtable_caphap_ptfire

Datasets This is where the EMF automatically shows the Dataset Name

Folder Use the “Browse” button to select the output location **Browse**

Purpose You enter why you are exporting the data, and the EMF records this in the Logs tab of the Dataset Properties.

☐ **Overwrite files if they exist?**

Use this checkbox to overwrite existing files. Otherwise, the EMF will give an error if the files are there.

Exports Datasets **Export** Close window **Done**

5. Using the information in the *Export Window* figure for guidance, **set the output location to /home/training/emf_training/2002/smoke/export** (you'll need to create the "export" directory, which you can do from the EMF File Browser).
6. **Click the "Export" button.** Use the Status window to confirm that the export has started. This window will also tell you when the export has completed and how long it took.



Question: How long did your export take? (answer depends on computer's speed)

NOTE: If you had logged out of the EMF Client and logged back in during your export, the data would still have exported since that operation runs on the EMF server.



Question: If you forgot your export location days later, how could you find out to where you had exported?

7. Open the directory to which you exported your dataset, and look at the file name. On your Linux desktop, double click on the "training's Home" icon, and use the File Browser to navigate to the directory.

- OR -

In Windows, use the file browser to go to the directory.



Question: What is the name and how does it related to the Dataset Name in the EMF?

In the next section, we will learn more about how the EMF set the name of this exported dataset.

C.7.b. Naming convention

All datasets exported from the EMF use the following naming convention:

`<prefix>_<dataset name>_<date>_<version>_<suffix>`

Where:

`<prefix>`: a prefix set by the Dataset Properties Keyword EXPORT_PREFIX

`<dataset name>`: the name of the dataset, in lowercase, with characters that are not alpha-numeric replaced with underscores. For example, if the Dataset Name is "2002 Nonpoint NEI", then `<dataset name>` would be "2002_nonpoint_nei".

`<date>`: ddmmyyyy (e.g., 14may2007)

`<version>`: v# (e.g. v1), where # is the version number

`<suffix>`: a suffix set by the Dataset Properties Keyword EXPORT_SUFFIX

D. QA Manager

In this section, you will learn how to use the EMF QA Manager to take a systematic approach to data quality assurance. Specifically, you will learn about the following:

- The steps you will need to take while using the EMF quality assurance
- Adding QA steps to a dataset from a template

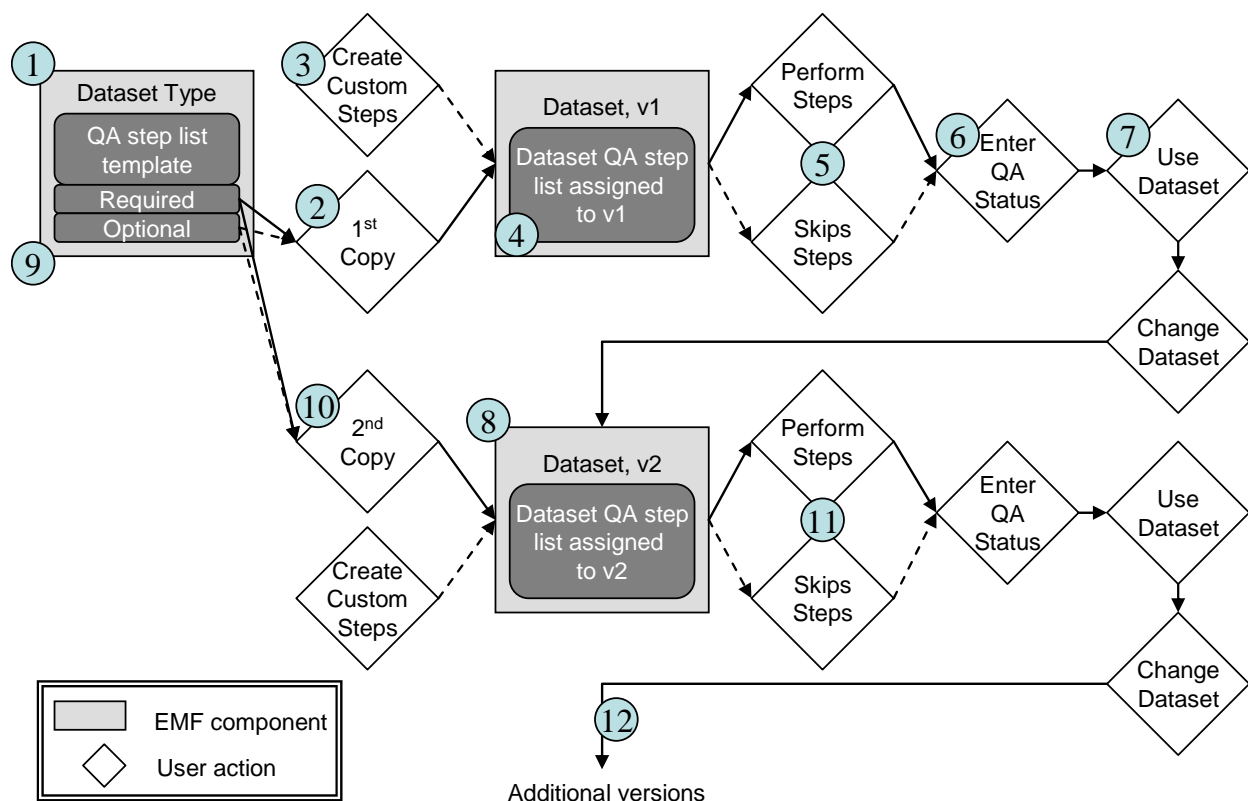
More of the QA Manager is included in the EMF Advanced Training.

D.1. EMF quality assurance overview

The EMF Quality Assurance Manager is an integral part of the Data Manager, which codifies an approach for quality assuring inputs and outputs of emissions modeling steps. The EMF supports a certain “way of doing things”, but allows users to enter and perform any QA steps on data that they believe is needed to do emissions modeling.

The diagram in Figure 2 below explains how to use the EMF to perform and document data QA. The dashed lines represent optional linkages, which depend on choices made by users. The step numbers shown in the figure correspond to the numbers in the subsequent list.

Figure 2: EMF approach to performing quality assurance of data



1. User sets up a QA step list template for each Dataset Type. This could also be done by an emissions modeling expert working at the user's site or entered from documentation provided by EPA.

2. User copies QA steps from template to an actual dataset
 - a. Steps labeled “Required” will be copied to actual datasets when using the template. Users do not have the option to not copy these steps.
 - b. Users can choose whether or not to copy “Optional” steps to actual datasets.
3. User optionally creates custom QA steps, which will only apply to that dataset/version and will not be able to be reused (copied to other datasets or versions) later.
4. Now the dataset has a list of steps that need to be performed.
5. Users have the choice to perform the QA steps or skip them. In either case, the result will be documented by the system. If a step is skipped, the user can document why. If a step is performed, the user can document the results.
 - a. Steps can be manually performed
 - b. Steps can include using SQL commands to create data summaries needed for a particular QA step. The EMF can run that SQL code and create new summary tables for the dataset/version.
6. Users enter a QA status for each step. For steps that involved running SQL commands, the EMF automatically updates the QA Status to “In Progress” when you run the query.
7. Once the steps are complete, the user can use the dataset.
8. Either based on the QA or based on using the data, the user may find that data changes are needed. The user would create version 2 of the dataset with the Dataset Editor, and would then start the QA process over again.
9. The user can update the QA Template for the Dataset Type at any time during the process.
10. When the user wants to redo the QA steps for a new version, the QA Template must be applied again, to that new dataset version.
11. Users can then perform the QA or skip steps as in (5) above. The results of the QA steps applied to Version 1 are not lost, since they are kept and maintained separately.
12. This process can continue for as many dataset versions as needed.

D.2. Adding QA steps to a dataset

There are two ways to add QA steps to a dataset (1) from a template and (2) as a custom step just to that dataset. In this section, you will learn how to add a QA step from a template.

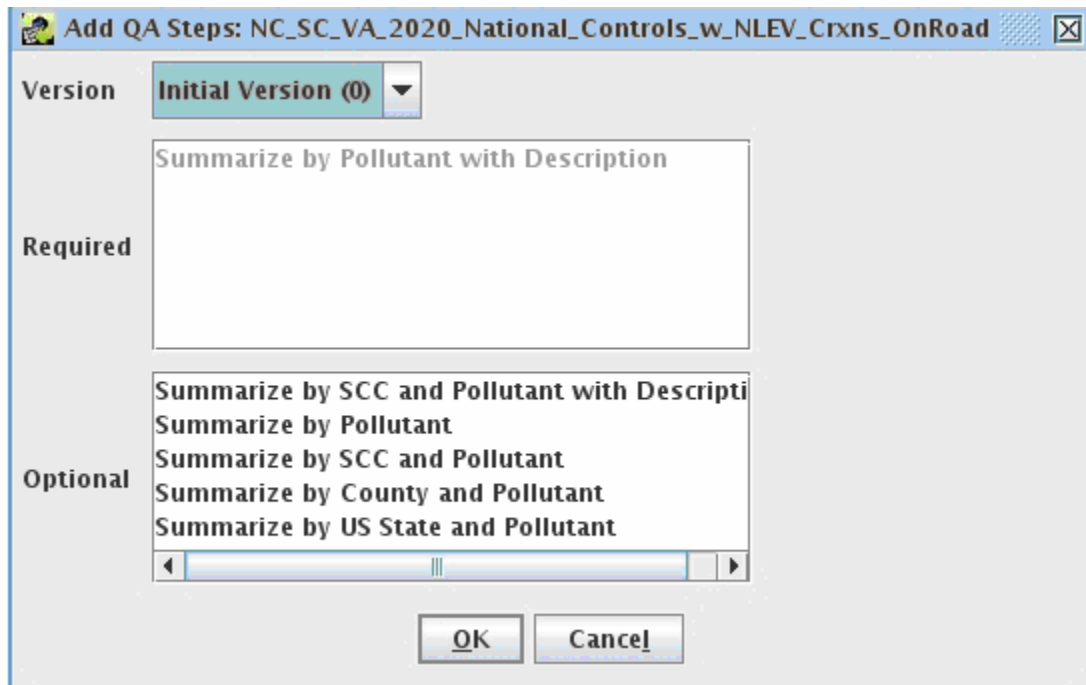
Unlike the Dataset Manager, there is no separate “main window” for the QA Manager. Instead, use the “QA” Tab on the Dataset Properties for quality assurance of a given dataset.

D.2.a. Adding from a template

The following steps will allow you to add steps to a dataset from a template.

1. **Open the *Dataset Manager Window***
2. Use the pull-down at the top of the *Dataset Manager* to **select Dataset Type of “ORL Onroad Inventory (MBINV)”**.
3. **Click the appropriate checkbox at left to select the Dataset named “NC_SC_VA_2020_National_Controls_w_NLEV_Crxns_OnRoad.txt”**

4. **Click the “Edit Properties” button.** You will see the Dataset Properties Edit Window
5. **Select the “QA” Tab.** You will see that the tab has a table header, but no contents in the table. This is because no QA steps have yet been added to the dataset.
6. To add QA steps from the Template that has already been created for ORL Onroad Inventory files, **click the “Add from Template” button** at the bottom left of the QA Tab. You will see the window shown here, with the optional steps as shown.



7. **Select all optional QA steps by:**
 - a) **Click the first optional step with your mouse**
 - b) **While holding down the <shift> key, click the second last step.** The <shift> key allows you to select any range of steps from the list.
 - c) Note that using the <ctrl> key allows you to select and group of individual QA steps.
8. With all Optional steps selected, **click the “OK” button.** The *Add QA Steps* dialog will disappear, and you will see the 5 new steps added to your QA Tab table.

Finally, you will look at the details of one of the QA steps:

9. In the QA Tab, open the “Summarize by Pollutant with Description” step by **clicking on the checkbox to the left of the step and click the “Edit” button.** You will see the *Edit QA Step Window*.

The top part of the window has the details about running/performing the step.

Name: Summarize by Pollutant with Description

Version: Initial Version (0) **Program:** SQL

Arguments:

```
select POLL, p.pollutant_code_desc, sum(ann_emis) as ann_emis
from $TABLE[1], reference.pollutant_codes p where POLL=p.pollutant_code
group by POLL, p.pollutant_code_desc order by POLL
```

Order: 0.0 **Required?** ☒

Description:

The bottom part of the window has status information and will look like the following picture before a step has been run.

QA Status: Not Started **Output Name:**

QA User: **Run Status:**

QA Date: **Run Date:**

Current Output? ☐

Configuration:






Comments:

Export Folder: /home/training/emf_training/2002/smoke **Browse**

Save **Close** **Run** **View Results** **Export**

The table below explains each part of this window:

QA Status	User-defined QA Status. The EMF will set to “In Progress” once the “Run” button has been clicked on steps run by the EMF. If changed by you, this automatically sets the QA User and QA Date.
QA User	Three ways to set the user performing the QA: 1. For SQL steps, EMF sets to user who ran the SQL query. 2. Set when QA Status is set by user 3. Can manually type in that someone else did the step
QA Date	Records the date/time that the step was done. Three ways to set just like the QA User.

Configuration	Optional configuration file that can be used as input to the QA step. For example a Smkreport “REPCONFIG” file. A future version of the EMF will link this file to a Dataset in the database.
Comments	Any users comments on the results of the QA step. This is where you document your results.
Export Folder	For SQL steps, the export location for creating the CSV table of results from the query.
Output Name	Unique name EMF assigns automatically to tables created by SQL queries
Run Status	EMF sets to “Success” or “Failed” depending on whether the SQL query ran or not.
Run Date	EMF sets date when the query was run.
Current Output?	EMF determines if the query has been run since the dataset was last changed. If the box is checked, then
	Saves and closes the <i>Edit QA Step Window</i>
	Closes the <i>Edit QA Step Window</i> without saving
	Runs the QA step (for SQL queries only, for now). In the future, the EMF will be able to run other programs.
	(not yet implemented). When implemented, it will permit the results of a SQL query to be displayed in a on-screen table.
	Exports the results of a SQL step to a CSV file.

In the EMF Advanced Training, we will further discuss the QA, including creating and running QA steps.

E. Case Manager

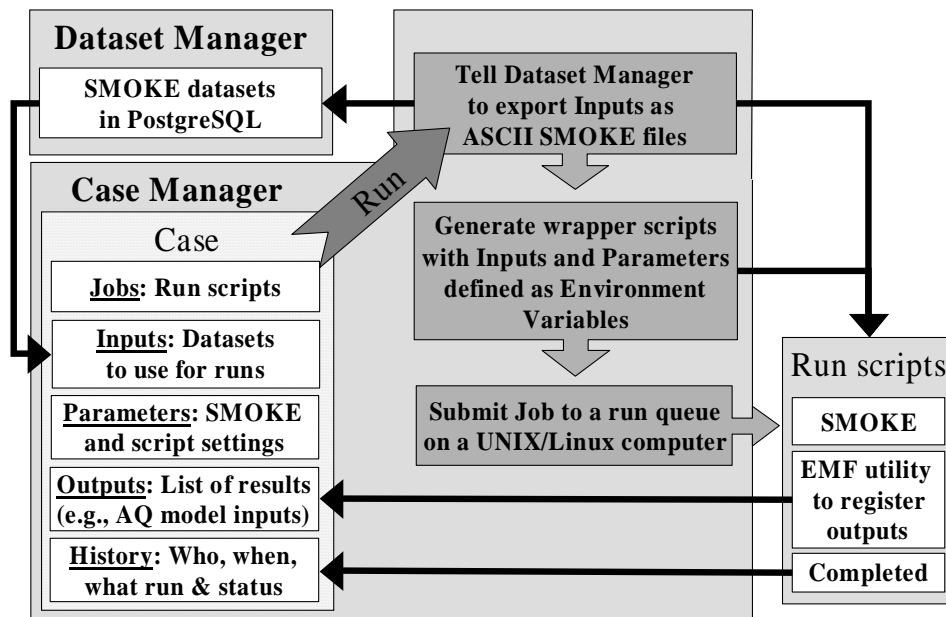
The EMF Case Manager runs SMOKE or other programs. EPA plans to use the Case Manager to run SMOKE, the Surrogate Tool, and the Speciation Tool. However, at this time the Case Manager is not yet complete.

At this time, the Case Manager can be used to setup batch jobs, but not actually run them.

In this training, we will use the Case Manager to learn about the following:

- The Case Manager main window and its buttons
- Opening a case and reviewing its main parts.
- Copying and editing a case
- Exporting data from a case
- Running a case
- Finding case outputs

The following figure provides an overview of how the Case Manager and Dataset Manager work together to run SMOKE.

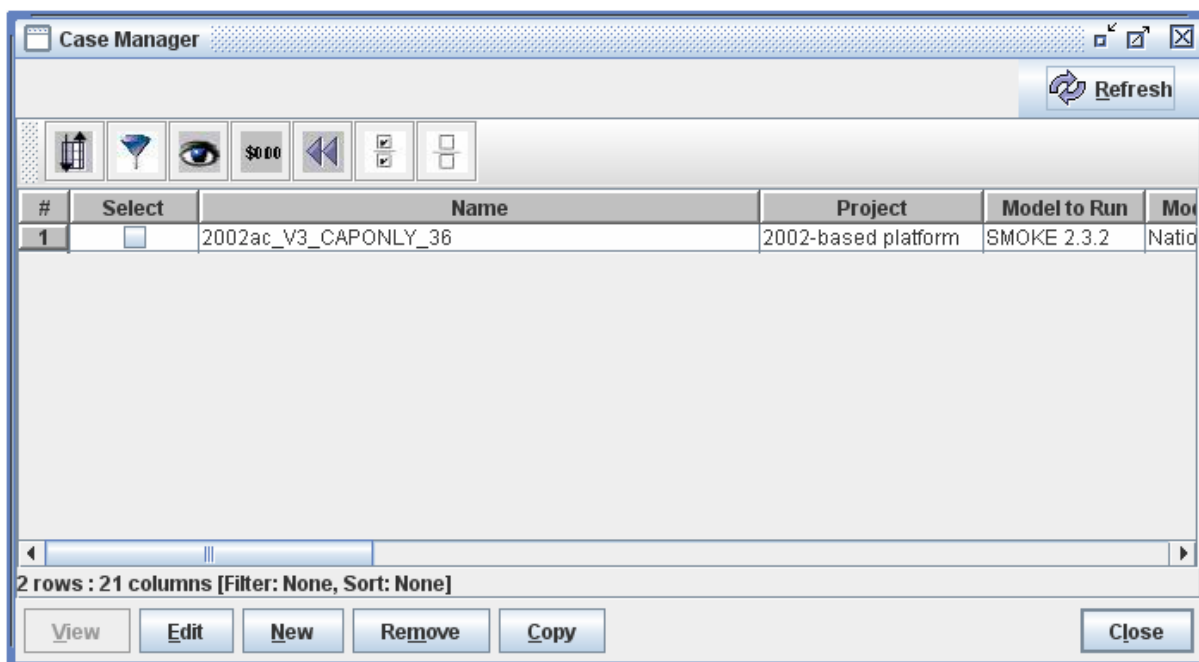


E.1. The Case Manager main window and its buttons

The Case Manager provides a list of all cases. You use the Case Manager as the main point of access for creating, editing, and running cases.

To access the Case Manager:







1. On the Main Menu, **click Manage → Cases menu option**. You will next see the *Case Manager Window*.



You will see only a single case loaded into the Case Manager, which has been created specifically for this training. Here are some notes about this case as compared to our full 2002-based platform case “2002ac”:

- Only the US sectors are included in this case. The sectors for Canada and Mexico have been excluded. Generally, a case includes all sectors needed for input on an air quality model run.
- For sectors with month-specific data (e.g., ptfire, ptipm, onroad, nonroad), only the January and July runs have been included.
- Not all of the “Parameters” have been set for running an actual case, so SMOKE users will notice that this part of the case is incomplete.

The *Case Manager Window* has several buttons at the bottom of the window. The table below describes the purpose of each of the buttons:

	(not yet implement). This button will allow users to open a case in “View only” mode, to allow access without the possibility of accidentally changing the case. Once users permissions are added to the EMF, some users will only have View privileges on cases and not edit privileges.
	Opens the tabbed window for changing the details of a case.
	Creates an empty new case
	Removes a case from the Case Manager. Only the owner of a case can delete it.
	Copies a case, which can then be further edited.
	Closes the <i>Case Manager Window</i>

Finally, note that the *Case Manager Window* also has the EMF toolbar to help find cases when there are many listed in the main window. This toolbar works the same way as the one explained in Section C.1.b.

E.2. Open a case and learn about its main parts

The purpose of this section is to open a case, see the main parts of a case, and learn about their purposes. We will start with an example of a case that has been previously added to the EMF. This case is similar to one being developed to run SMOKE for OAQPS’s 2002-based modeling platform.

1. To open a case (in edit mode), **select the checkbox to the left of the Case Name and click the “Edit” button.** You will then see the *Case Editor Window*, with the “Summary” tab shown opened.

Case Editor: 2002ac_V3_CAPONLY_36

Summary Jobs Inputs Parameters Outputs History

Name: 2002ac_V3_CAPONLY_36 Abbreviation: 2002ac

Category: run SMOKE for CMAQ Is Final: ☐ Is Template: ☐

Description: Version 3 Platform (Spring) CAP only using NEI 2002 v3, CB05, SPECIATE4.0 for EVALUATION Sectors: ptfire, alm, none

Project: 2002-based platform Add Remove

Run Status: Not Started Copied From: 2002_Platformv2_CAPONLY_36

Last Modified Date: 2007/05/10 12:50 Last Modified By: Marc Houyoux

Model to Run: SMOKE 2.3.2 Air Quality Model: CMAQ

Modeling Region: National Speciation: cmaq_cb05pm

Control Region: National Meteorological Year: 2002

I/O API Grid Name: US36KM_148X112 Base Year: 2002

Grid Resolution: 36km Future Year: 2002

Start Date & Time: 2002/01/01 00:00 End Date & Time: 2002/12/31 23:59

Save Close

The *Case Editor Window* has a tab for each of the parts of a case as shown here:

Summary Jobs Inputs Parameters Outputs History

The following list summarizes the purpose of each Case component and tab:

- The “**Summary**” tab contains high-level information about the case, such as the case name, I/O API grid name, chemical mechanism, year and run period, run status, air quality model, and sectors included.
- The “**Jobs**” tab defines all of the individual script runs and options needed to create the data generated by the case.
- The “**Inputs**” tab defines all of the input datasets (from the Dataset Manager) needed as inputs for the case.
- The “**Parameters**” tab defines all of the settings used by the Jobs and programs used in the Jobs.
- The “**Outputs**” tab (when implemented) will list all of the output data generated by the Job. When we complete the run feature of the EMF, the outputs from a case will be automatically registered by the case as either External (e.g., CMAQ inputs or SMOKE reports) or internal (e.g., SMOKE reports) datasets.

- The “**History**” tab (when implemented) will give the history of who ran which jobs, when they were run, and their run status.
2. **Use the EMF Toolbar to explore the “Inputs” Tab** using sorting and filtering (note: this will be more interesting for SMOKE users).
 3. **Close the case by clicking the “Close” button.**

E.3. Copy and edit a case

In this section, you will use the *Case Manager* and *Case Editor*, to copy and then modify existing cases. We will modify the Summary, Jobs, Inputs, and Parameters tabs, to give you a better understanding about each part of the EMF and how they can be used. The next three exercises will step through the process of modifying these existing cases. We do not build a case from scratch because the number of inputs and parameters needed would take too long for this training session.

Exercise 5

In this exercise, we will copy the case and add a merge step to the jobs. First, create a copy and edit the name:

1. In the *Case Manager Window*, **click the checkbox next to the “2002ac V3 CAP for EMF training” case, then click the “Copy” button.** After a brief pause, the EMF will display a new case in the *Case Manager Window*, with a name that starts with “Copy of 2002ac V3 CAP”.
2. **Click the checkbox next to this copied case, and click the “Edit” button.** You will now see “Summary” tab of the *Case Editor Window*.
3. Under the “Summary” tab, **edit the Name field, change it to “2002ac V3 CAP Merged for EMF training”. Click on the “Save” button.** This will return you to the *Case Manager Window*.

Now we will add a merge script to the case:

4. **Reselect the new case, “2002ac V3 Cap Merged for EMF training”, and click the “Edit” button.** Go to the “Jobs” tab.
5. **Click the “Add” button** at the bottom of the window.
6. In the *Add Job Window*, set the **Name** to “Merge all”.
7. For the **Executable**, use the browse button to go to the scripts directory:
/home/training/emf_training/2002/smoke/subsys/v3/smoke23/scripts/cases/2002ac.
Select the smk_mrggrid_2002ac_36km.csh script and click “Ok”.
8. For **Arguments**, use: -q “1 2 3 4” Y. Set Order to 2.0. Then click “Save”.

Now we will add a new parameter called MRG_DIFF_DAYS that controls the SMOKE Mrggrid program. We will define this parameter only for this program and for the new job:

9. In the Case Editor Window, go to the “Parameters” tab.

10. Click the “Add” button at the bottom of the window.

11. In the *Add Parameter Window*, fill in the following values

- **Parameter name** = Merge different days
- **Program** = Mrggrid
- **Environmental Variable** = MRG_DIFF_DAYS
- **Type** = Boolean
- **Value** = Y
- **Sector** = All sectors
- **Job** = Merge All

12. Click “Save”

13. Review changes and then close the *Case Editor Window*.

Exercise 6

In this exercise, we will create a case with a grid cell resolution of 12 km instead of 36km:

1. Make a copy of the case “2002ac V3 Cap Merged for EMF training” and rename it “2002ac V3 Cap 12km for EMF training”.
2. In the “Summary” tab, **change the Grid resolution to 12km.**
3. We want to change the I/O API Grid Name to a real grid in our Grid description file (this file has a list of spatial domains, including their resolutions, size, and geographic projections):
 - Go to the “Inputs” tab
 - **Select the “GRIDDESC. Grid Description Information” input**, note the version number
 - **Click “View Data”**, then navigate to the “Data” tab and view the data for the appropriate version.
 - Go to the bottom of the dataset. Here are a series of grid names that have the appropriate resolution. Pick one (it doesn’t matter which) and type that name into the I/O API Grid Name field under the Case Editor Window’s “Summary” tab.
4. **Click “Save”** and clean up some of the additional windows that you opened.

Exercise 7

In this exercise, we will add another job and additional input datasets. In particular, we will add an “other onroad” job that includes mobile data from Canada and Mexico:

1. Make a copy of the case “2002ac V3 Cap Merged for EMF training” and rename it “2002ac V3 Cap North America for EMF training”.
2. In the “Summary” tab of your new case, **change the Modeling Region and Control Region** to US, Canada, Mexico.
3. In the “Jobs” tab, **“Add” a new job**:
 - **Name** = Onroad other
 - **Executable**, use the “Browse” to go back to the scripts directory and select the script, “smk_othon_2002ac.csh”. Scripts are in:
/home/training/emf_training/2002/smoke/subsys/v3/smoke23/scripts/cases/2002ac
 - **Arguments** = -q “1 2 3 4” Y
4. **Click the “Save” button to save the case.**

Now we want to add new datasets to include mobile data from Canada and Mexico:

5. Remember, to add a new dataset we should **go to the Dataset Manager**. Use the “Import” button to add data of type “IDA Mobile”. Select the files starting with “mbinv” from the input directory: /home/training/emf_training/tmp/. You can use “Create Multiple Datasets” if you want to select all 3 inventories at the same time.
6. **Return to the Case Manager Window**. Select our new case to edit.
7. Go to the “Inputs” tab and **click “Add”** to add an input:
 - **Input name** = onroad Canada 2000
 - **Program** = Smkinven
 - **Environmental Variable** = MBINV_A
 - **Dataset Type** = IDA Mobile
 - **Dataset** = mbinv_onroad_Canada2000_07nov2006_v0_ida.txt
 - **Job** = Onroad other
 - **Subdirectory** = othon
8. Repeat the above step for the 2 Mexican inventories, assigning the datasets to MBINV_B and MBINV_C.
9. Return to the “Summary” tab and update the **Description** to reflect this case. Click “Save”.

E.4. Exporting data from case

Since programs like SMOKE cannot read data directly from a database, the EMF exports the SMOKE input datasets to ASCII data files before running a case. You can also use the “Export Inputs” button on the “Inputs” tab to export all selected datasets to SMOKE input files in the correct format. Like all other buttons, this button operates on the datasets selected with checkboxes. All SMOKE inputs can be exported at the same time using this button.

To use this feature, take the following steps. Note that we will not export all of the datasets for this example case because it takes a bit too long to export all of the large datasets.

1. On the “Inputs” tab of your copied case, **select the three files that you added and the ARTOPNT file** using the checkboxes at left.
2. **Click the “Export Inputs” button.**
3. View the main Status Window for information regarding the status of your exports.
4. When the Status Window says that the datasets have exported, confirm this (if you’d like) by double clicking on the “training’s Home” icon on the Linux Desktop and using the File Browser to navigate to the directory (See Status Window for export directory).
- OR -
In Windows, using the Windows file browser to navigate to the directory.

E.5. Running a case

Although running a case is not yet implemented, this will be done using the “Run” button at the bottom of the “Jobs” tab. The basic steps for running a case are:

1. You select all of the jobs that you want to run on the “Jobs” tab using the checkboxes at left.
2. You click the “Run” button at the bottom of the “Jobs” tab.
3. The Case Manager exports all of the datasets needed for the selected jobs, and/or checks to see that they have already been exported previously.
4. The Case Manager builds wrapper scripts to run the executables based on the “Parameters” and “Inputs” tabs. It then submits all selected jobs to a queue.
5. The main Status window will tell you the start and complete status of each job, and whether any jobs have failed.
6. The EMF will update the Status of each job, which you can view on the “Jobs” tab and when you edit/open an individual job.
7. The run scripts (e.g., SMOKE scripts) call an EMF script that automatically logs each of the output files from the job.

E.6. Finding case outputs

As mentioned above, the outputs from your case can be automatically registered by the EMF. The registered outputs will show up in two places: (1) The Outputs tab of the Case Editor and (2) the Dataset Manager. You can use either approach to find the outputs.