

Quality Assurance of Emission Inventory Data using the EMF and EmisView

Alison Eyth, R. Partheepan

Carolina Environmental Program

University of North Carolina at Chapel Hill

Marc Houyoux

Emissions Inventory and Analysis Group

U.S. EPA OAQPS

Goals for Quality Assurance in EMF

- Formalize list of QA Steps to be performed on each type dataset
- Track information about progress of steps: status, who, when, etc.
- Integrate with EMF data management
- Support tracking of QA steps for multiple versions of datasets
- Speed (automate) the QA process



Typical QA Process for Emission Model Inputs

- *Level of formalization of QA processes varies between modeling studies and organizations*
- Typically, there is a list of steps to perform that vary for each type of data
- Some part of each step may be automated and other parts are manual
- If a step does not pass, it may need to be repeated if data values were changed
- Ideally, once steps are completed successfully, data is then used for modeling



Example Inventory QA Steps

- Validate data format
- Compare high level summaries to previous summaries (by state, SCC)
- Verify that the expected regions, codes, and pollutants exist in the data
- Check for duplicate records
- Verify that the time period is correct
- Validate outliers for emissions records
- Check for within-record consistency (e.g., validate coords., stack params.)
- Check consistency across records



Tracking QA Steps in EMF

- First, set up Step Templates for Dataset Types
- Copy Steps to Dataset metadata from Templates
- Add any ad-hoc steps (not from templates) to Dataset metadata
- Record results of the steps for each version of a Dataset



Dataset Type Manager

Dataset Type Manager Refresh

#	Select	Name	# Keywords	# QA Step ...	Min Files	Max Fi
21	<input type="checkbox"/>	ModelReady Emissions File (External)	0	0	1	
22	<input type="checkbox"/>	NIF3.0 Nonpoint Inventory	0	0	2	
23	<input type="checkbox"/>	NIF3.0 Nonroad Inventory	0	0	2	
24	<input type="checkbox"/>	NIF3.0 Onroad Inventory	0	0	2	
25	<input type="checkbox"/>	NIF3.0 Point Inventory	0	0	6	
26	<input type="checkbox"/>	NonHAP VOC Calculation Exclusions (Line-based)	0	0	1	
27	<input type="checkbox"/>	ORL Nonpoint Inventory (ARINV)	1	13	1	
28	<input type="checkbox"/>	ORL Nonroad Inventory (ARINV)	0	0	1	
29	<input type="checkbox"/>	ORL Onroad Inventory (MBINV)	0	0	1	
30	<input type="checkbox"/>	ORL Point Inventory (PTINV)	0	1	1	

46 rows : 7 columns

View Edit New Close



Edit Dataset Type: ORL Nonpoint Inventory (ARINV)

Name:

Description:

Default Sort Order:

Keywords

Select	Keyword	Default Value
<input type="checkbox"/>	EXPORT_SUFFIX	_orl.bt

Add

Remove

QA Step Templates

Select	Name	Program	Arguments	Required	Order
<input type="checkbox"/>	Compare state summaries	Smkreport	state_summary.repc...	<input checked="" type="checkbox"/>	2.1
<input type="checkbox"/>	Compare county summaries	EmisView	-subset county_sum...	<input checked="" type="checkbox"/>	2.2
<input type="checkbox"/>	Compare SCC summaries	EmisView	-subset scc_summary	<input type="checkbox"/>	2.3
<input type="checkbox"/>	Verify regions	None		<input checked="" type="checkbox"/>	3.1
<input type="checkbox"/>	Verify all data	None		<input type="checkbox"/>	3.2

Add

Remove

Edit

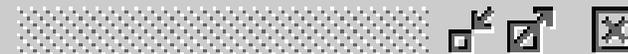
Save

Close





Edit QA Step Template: Compare county summaries



Name:

Compare county summaries

Program:

EmisView



Arguments:

-subset county_summary

Order:

2.2

Required?



Description:

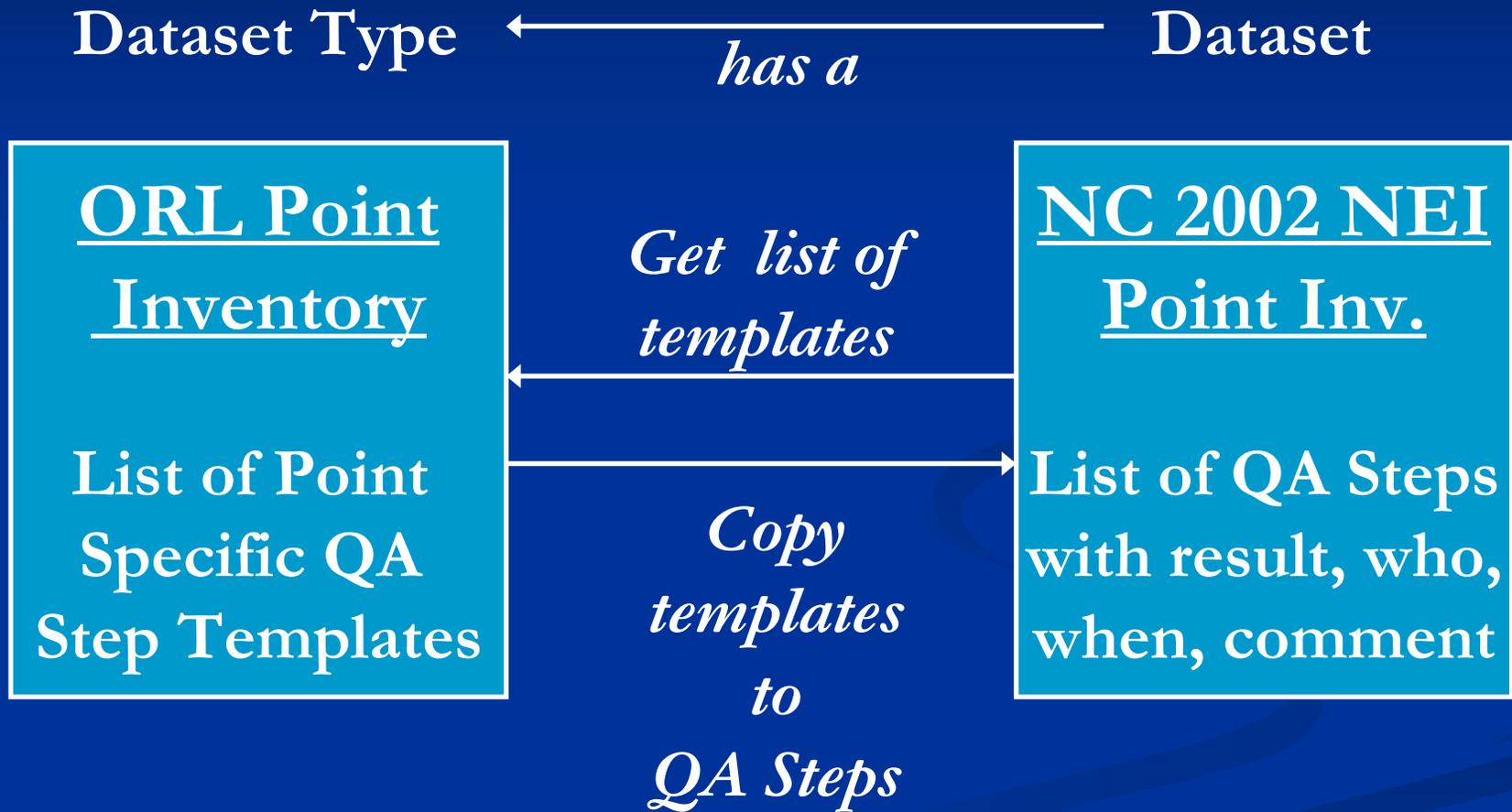
Compare county summaries with any previously generated county summaries

Save

Close



QA Step Templates vs. QA Steps



Adding QA Steps to a Dataset

Dataset Properties Editor: arinv.nonroad.nti99d_NC.new.txt

Summary Data Keywords Notes Revisions Logs Tables QA

Select Version Name Required Order Status When Who Comm

0 rows : 12 columns

Add from Template Add Custom Edit Set Status

Save Close

Click this button



Adding Steps to Dataset from Templates

Add QA Steps: ORL Nonpoint Inventory (ARINV)

Version Initial Version (0) ▼

Required

- Initial Version (0)
- Version 1 (1)
- Version 2 (2)

Optional

- Compare SCC summaries
- Verify SCCs
- Check for duplicate records
- Check for unique attributes

OK Cancel



Summary of QA Steps for all Versions of a Dataset

Dataset Properties Editor: orl_arinv.nonpoint.nti99_NC.txt *

Summary Data Keywords Notes Revisions Logs Tables QA

\$0.00

#	Select	Versi...	Name	Required	Order	Status	When	Who	Con
1	<input type="checkbox"/>	0	Compare state summaries	<input checked="" type="checkbox"/>	2.1	Complete	05/16/2006 20:07	Alison Eyth	good
2	<input type="checkbox"/>	0	Compare county summaries	<input checked="" type="checkbox"/>	2.2	Complete	05/07/2006 21:34	Alison Eyth	
3	<input type="checkbox"/>	0	Compare SCC summaries	<input type="checkbox"/>	2.3	Complete	05/07/2006 21:34	Alison Eyth	
4	<input type="checkbox"/>	0	Verify regions	<input checked="" type="checkbox"/>	3.1	Complete	05/07/2006 21:34	Alison Eyth	
5	<input checked="" type="checkbox"/>	0	Verify pollutants	<input checked="" type="checkbox"/>	3.2	Not Started	05/17/2006 11:39	Alison Eyth	
6	<input checked="" type="checkbox"/>	0	Verify SCCs	<input type="checkbox"/>	3.3	Not Started	05/17/2006 11:39	Alison Eyth	
7	<input type="checkbox"/>	0	Check for duplicate records	<input type="checkbox"/>	4.1	Failed	05/07/2006 21:35	Alison Eyth	
8	<input type="checkbox"/>	0	Verify time period	<input checked="" type="checkbox"/>	5.1	Failed	05/07/2006 21:35	Alison Eyth	
9	<input type="checkbox"/>	0	Check for emission outliers	<input checked="" type="checkbox"/>	6.1	In Progress	05/17/2006 11:38	Alison Eyth	
10	<input type="checkbox"/>	0	Check for other outliers	<input checked="" type="checkbox"/>	7.1	In Progress	05/17/2006 11:38	Alison Eyth	
11	<input type="checkbox"/>	0	Check within record consistency	<input checked="" type="checkbox"/>	8.1	Skipped	05/07/2006 21:35	Alison Eyth	
12	<input type="checkbox"/>	0	Check across-record consistency	<input checked="" type="checkbox"/>	9.1	Skipped	05/07/2006 21:35	Alison Eyth	
13	<input type="checkbox"/>	1	Compare state summaries	<input checked="" type="checkbox"/>	2.1	In Progress	05/07/2006 21:35	Alison Eyth	

21 rows : 12 columns

Add from Template Add Custom Edit Set Status

Save Close



Edit QA Step: Compare state summaries - orl_arinv.nonpoint.nti99_NC.txt (v0)

Name: Compare state summaries

Version: Initial Version (0)

Program: Smkreport

Arguments: state_summary.repconfig

Order: 2.1

Required?

Description: Compare the state summaries of the inventory with any previously created state level summaries.

Status: Complete

User: Alison Eyth

Date: 05/07/2006 21:34

Configuration:

Comments:

Save Close



Generating QA Reports

- Not currently automated, but...
- EmisView, Smkreport, Smkinven, and other utilities can generate information needed for inventory QA
- Commonly needed reports:
 - Summaries by state, county, SCC, ...
 - Outlier [threshold] reports for emissions values and stack parameters
 - Top N reports [by state] (needs sorting)
 - Reports of particular slices of inventory



Goals for EmisView

- Preserve (and reproduce) analyses that have already been performed
- Subselect and summarize inventories for each of the major inventory formats
- Present same subset of data in multiple ways (e.g., tables, plots, maps)
- Automatically generate consistent set of tables and plots for similar datasets
- Integrate with EMF, but developed before EMF Data Management was mature



Mix and Match Components

- Between sessions, EmisView stores:
 - Datasets - inventories to analyze
 - Subsets - ways to slice & summarize data
 - Products - ways to present data
 - Analyses
- Each Analysis uses one Dataset, Subset, and Product – you can mix & match the components



Main EmisView GUI - Datasets

The screenshot shows the EmisView application window with the 'Datasets' tab active. The window title is 'EmisView' and it has a menu bar with 'File', 'Options', 'Tools', and 'Help'. Below the menu bar are tabs for 'Datasets', 'Subsets', 'Products', and 'Analyses'. A toolbar contains icons for home, funnel, eye, currency (\$0.00), left arrow, refresh, and window management. The main area is a table with 6 rows and 6 columns. Row 6 is selected. Below the table, it says '6 rows : 6 columns'. At the bottom are buttons for 'New', 'Copy', 'Rename', 'Delete', 'Configure', and 'Help'.

#	Select	Name	Type	Region	Start
1	<input type="checkbox"/>	Area fugitive dust	IDA Nonpoint/Nonroad Inventory	US	01/01/01 01
2	<input type="checkbox"/>	CT NIF3 Nonroad	NIF3 Nonpoint Inventory	CT	01/01/99 01
3	<input type="checkbox"/>	CT NIF3 Point 1999	NIF3 Point Inventory	CT	01/01/99 01
4	<input type="checkbox"/>	CT NIF3 Nonpoint	NIF3 Nonpoint Inventory	CT	01/01/99 01
5	<input type="checkbox"/>	Point Fugitive Dust	IDA Point Inventory	US	01/01/01 01
6	<input checked="" type="checkbox"/>	orl_arinv.nonpoint.nti99_NC.txt	ORL Nonpoint Inventory	NC	01/01/99 01



Main EmisView GUI - Subsets

EmisView

File Options Tools Help

Datasets Subsets Products Analyses

Bar Chart Funnel Eye \$0.00 Left Arrow Window Icons

#	Select	Name	Region	SCC	Columns
1	<input type="checkbox"/>	All Records	All		0
2	<input type="checkbox"/>	By SCC	All		13
3	<input type="checkbox"/>	PM 10 > 1000	All		17
4	<input type="checkbox"/>	by NAICS	All		11
5	<input type="checkbox"/>	by SIC	All		11
6	<input type="checkbox"/>	Triangle	NC Triangle		0
7	<input type="checkbox"/>	By FIPS	All		8

10 rows : 6 columns

Copy Rename Delete Configure



A Threshold Subset

Subset: PM 10 > 1000 (IDA Nonpoint/Nonroad Inventory)*

Summary Region SCCs Columns Filter Report By

Apply Filter? Match using: ALL criteria ANY criteria

Add Criteria Delete Criteria

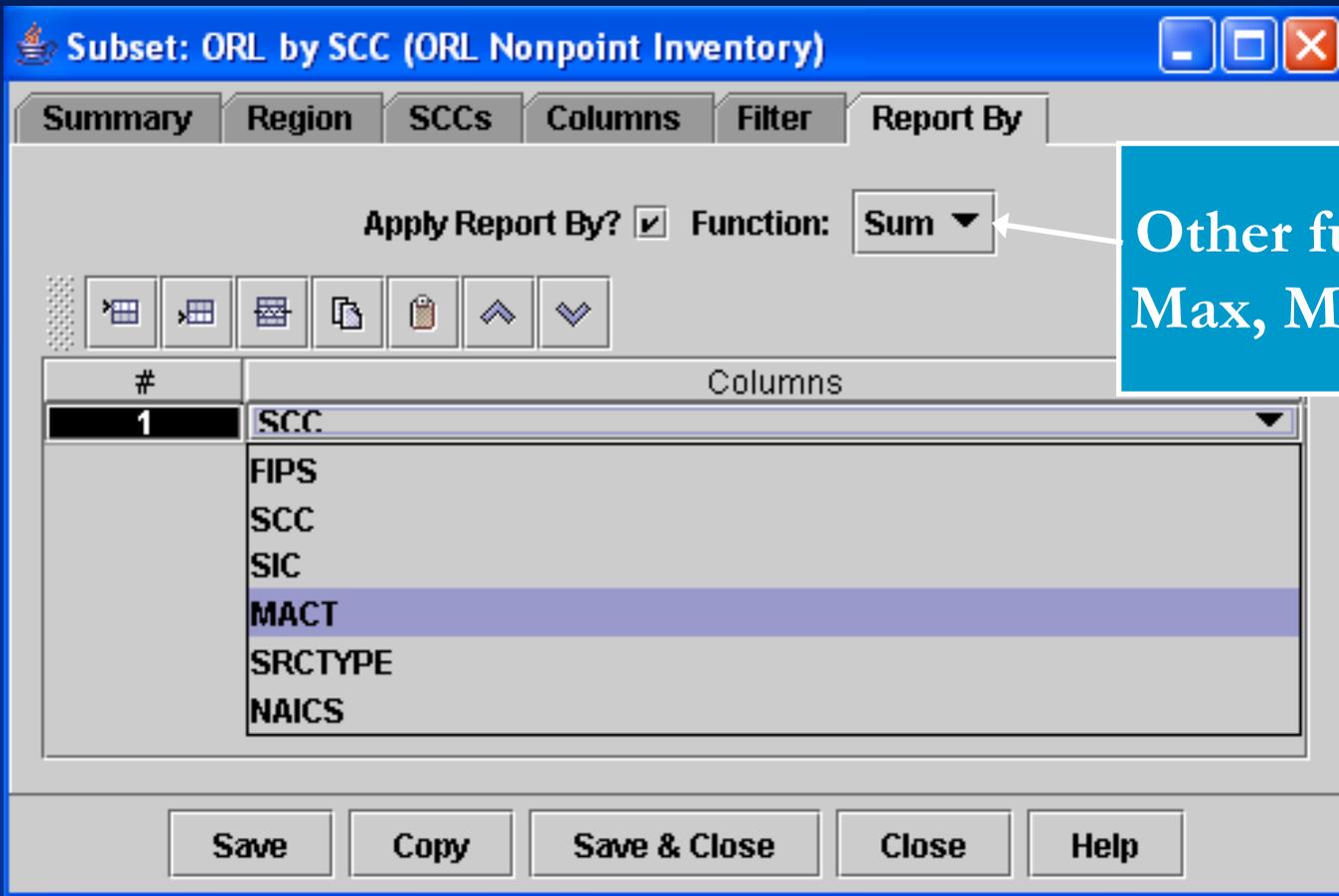
Column Name	Operation	Value
PM10	>	1.00E03
SCC	starts with	

STID
CYID
FIPS
State
SCC
PM10
PM2_5
PM10_AVD

Save Copy Save & Close Close Help



A Summarizing Subset



Other features of subsets: Set of states or counties, list of SCCs, choose columns – features are cumulative



EmisView Main GUI - Products

EmisView

File Options Tools Help

Datasets Subsets **Products** Analyses

↑ ↓ \$0.00 ◀ ▶ ☑ ☐

#	Select	Name	Configuration File	System Defined
1	<input type="checkbox"/>	Table and Plots		true
2	<input type="checkbox"/>	Top 10 PM 2.5	D:\My Documents\top10pm25.cfg	false
3	<input type="checkbox"/>	Sort by PM10	D:\My Documents\sortbyPM10.cfg	false

3 rows : 4 columns

New Copy Rename Delete Configure Help

Currently, products are differentiated based on their Configuration file (saved from Table to be shown later)



EmisView Main GUI - Analyses

EmisView

File Options Tools Help

Datasets Subsets Products **Analyses**

\$000

#	Select	Name	Dataset	Subset	Product
2	<input type="checkbox"/>	All dust	Area fugitive dust	All Records	Table and Plots
3	<input type="checkbox"/>	NIF3 Point All Records	Area fugitive dust	PM 10 > 1000	Table and Plots
4	<input type="checkbox"/>	NIF3 Nonpoint	CT NIF3 Nonpoint	All Records	Table and Plots
5	<input type="checkbox"/>	NIF3 Nonroad	CT NIF3 Nonroad	All Records	Table and Plots
6	<input type="checkbox"/>	NIF 3 Point by SIC	CT NIF3 Point 1999	by SIC	Table and Plots
7	<input type="checkbox"/>	NIF 3 Point by NAICS	CT NIF3 Point 1999	by NAICS	Sort by PM10
8	<input type="checkbox"/>	NIF3 Point Top 10 PM2.5	Area fugitive dust	PM 10 > 1000	Top 10 PM 2.5
9	<input type="checkbox"/>	Analysis2	CT NIF3 Point 1999	All Records	Table and Plots
10	<input type="checkbox"/>	CT Nonroad by FIPS	CT NIF3 Point 1999	by SIC	Sort by PM10
11	<input type="checkbox"/>	CT Nonroad by County	CT NIF3 Nonroad	By FIPS	Table and Plots
12	<input type="checkbox"/>	NC Nonpoint NTI 1999 by SCC	orl_arinv.nonpoint.nti99_NC.txt	ORL by SCC	Table and Plots
13	<input type="checkbox"/>	NC Nonpoint NTI 1999 by FIPS	orl_arinv.nonpoint.nti99_NC.txt	ORL by FIPS	Table and Plots
14	<input type="checkbox"/>	CT NIF3 Point	CT NIF3 Point 1999	All Records	Table and Plots

14 rows : 5 columns

New Copy Rename Delete Configure Run Help



 **Analysis: Fugitive dust by SCC**   

Analysis Name

Description

Subset

Custom Query

Dataset

Products

All Records

By SCC

PM 10 > 1000

Area fugitive dust

Table and Plots

Top 10 PM 2.5



Analysis Results

Analysis Results: Fugitive dust by SCC

Toolbar:

	SCC	PM10	PM2_5	PM10_AVD	PM2_5_AVD	
1	2311010000	535.90	107.20	0.00	0.00	▲
2	2311020000	12080.00	2417.00	0.00	0.00	
3	2311030000	6383.00	1277.00	0.00	0.00	
4	2325000000	6795.00	1359.00	0.00	0.00	
5	2801000003	29620.00	5924.00	0.00	0.00	
6	2805001000	10520.00	1578.00	0.00	0.00	▼

6 rows : 13 columns

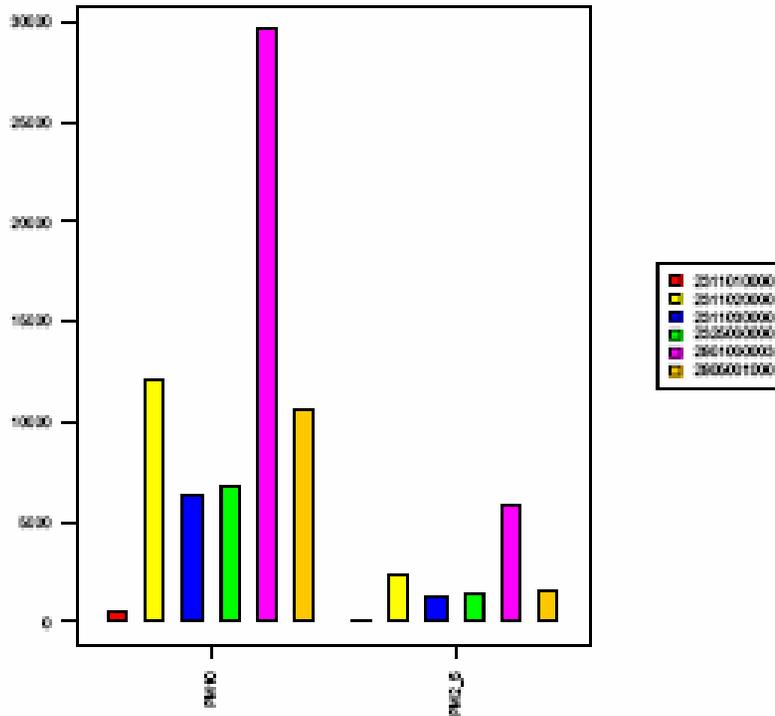
Description Load Configuration Export Close

Toolbar: multi-column sort, top N, bottom N, filter, hide cols, format, plot, save configuration, reset, [statistics]

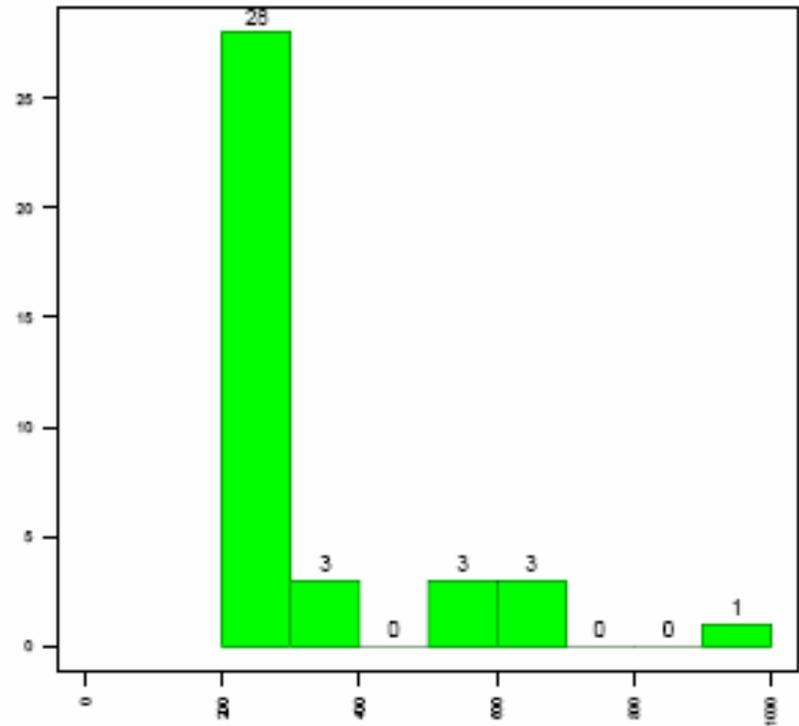


Examples of Plots

Fugitive Dust Emissions by SOC



Histogram of Exit Temperatures



Other plots: CDF, box-whisker, rank order, XY, line, time series, discrete category, tornado



Plot Configuration Options

Customize Bar Plot

File Help

Data sets

Bar Data Series 2 selected Set

Bar Plot Options

Plot Title	Fugitive Dust Emission...	<input type="button" value="Edit"/>
Plot Subtitle		<input type="button" value="Edit"/>
Plot Footer		<input type="button" value="Edit"/>
Bars		<input type="button" value="Edit"/>
Legend		<input type="button" value="Edit"/>
Category Axis		<input type="button" value="Edit"/>
Numeric Axis		<input type="button" value="Edit"/>
Borders		<input type="button" value="Edit"/>
Margin Size		<input type="button" value="Edit"/>
Text Boxes		<input type="button" value="Edit"/>

Page Options

File Name:

Plots can be saved as PDF, PNG, JPG, PS, and PTX



Software Requirements

- Open source software
- Runs on Linux, Unix, and Windows
- Requires Java 1.4 or 1.5 and R
- EmisView accesses local or remote DB server
 - Fall 2005 version supports MySQL or PostgreSQL
 - Later versions PostgreSQL only
- EMF is client-server PostgreSQL only



EmisView FY0? Enhancements

- Integrate with latest EMF importers
- Support more types of datasets
- New summaries: State (ORL), Overall
- Save Top-N as a configuration
- Simple script-based interface / API
- Integrate with EMF client
- Select and access Datasets and Versions directly from EMF DB
- Export Shapefiles for GIS analysis
- Access to existing statistical analyses



Availability of EmisView & EMF

- EmisView October 11, 2005 version is available from <http://emisview.sourceforge.net>
- Please report any issues
- Enhanced FY06 EmisView will be released by September 2006
- Public EMF release not yet funded, but direct arrangements could be made and code is on SourceForge



EmisView Summary

- EmisView should be able to produce the commonly desired reports for EI QA (summary, outlier, top N)
 - Needs testing on national scale
- EmisView will be able to produce standardized sets of tables and plots for similar types of inventories
- Addition of new data types should increase usefulness beyond inventories



Summary of EMF QA Features

- EMF supports definition of required and optional steps for each type of dataset (codifies the QA process)
- For a particular dataset, steps can be quickly copied from templates, or can be added as custom steps
- Tracking is performed for each step: who did it, when, status, comment
- Information is available to support future automation of steps



Analysis Results: CT Point 1999 All Records

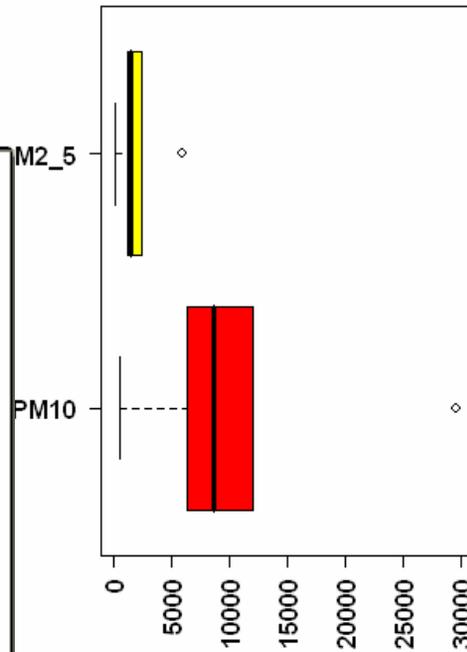
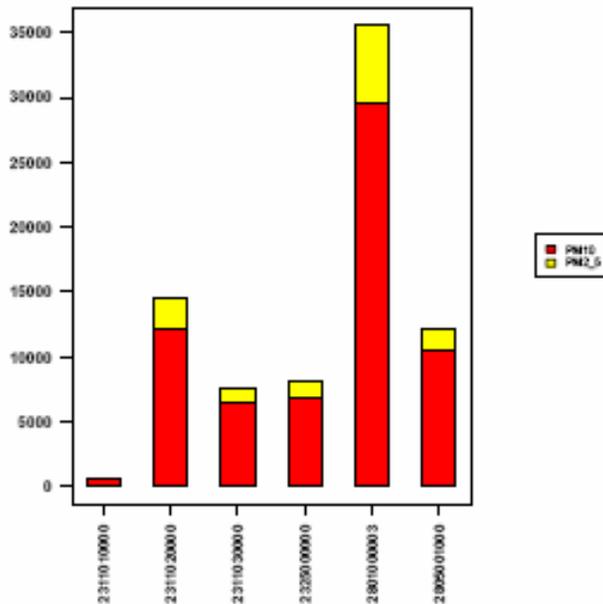


	State	FIPS	Facility	Unit	Process	Point	SCC	SIC	NAICS
1	CT	9001	0019	P0010	01	001	10200603	873	1
2	CT	9001	0019	P0012	01	002	20300101	873	1
3	CT	9001	0167	E0001	01	003	40301019	517	1
4	CT	9001	0167	E0002	01	004	40400101	517	1
5	CT	9001	0167	E0004	01	006	40400101	517	1
6	CT	9001	0167	E0005	01	007	40400101	517	1
7	CT	9001	0167	E0006	01	008	40301019	517	1

1113 rows : 27 columns

Description

Fugitive Dust Emissions by SCC



Connecticut Nonpoint Emissions by County

