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

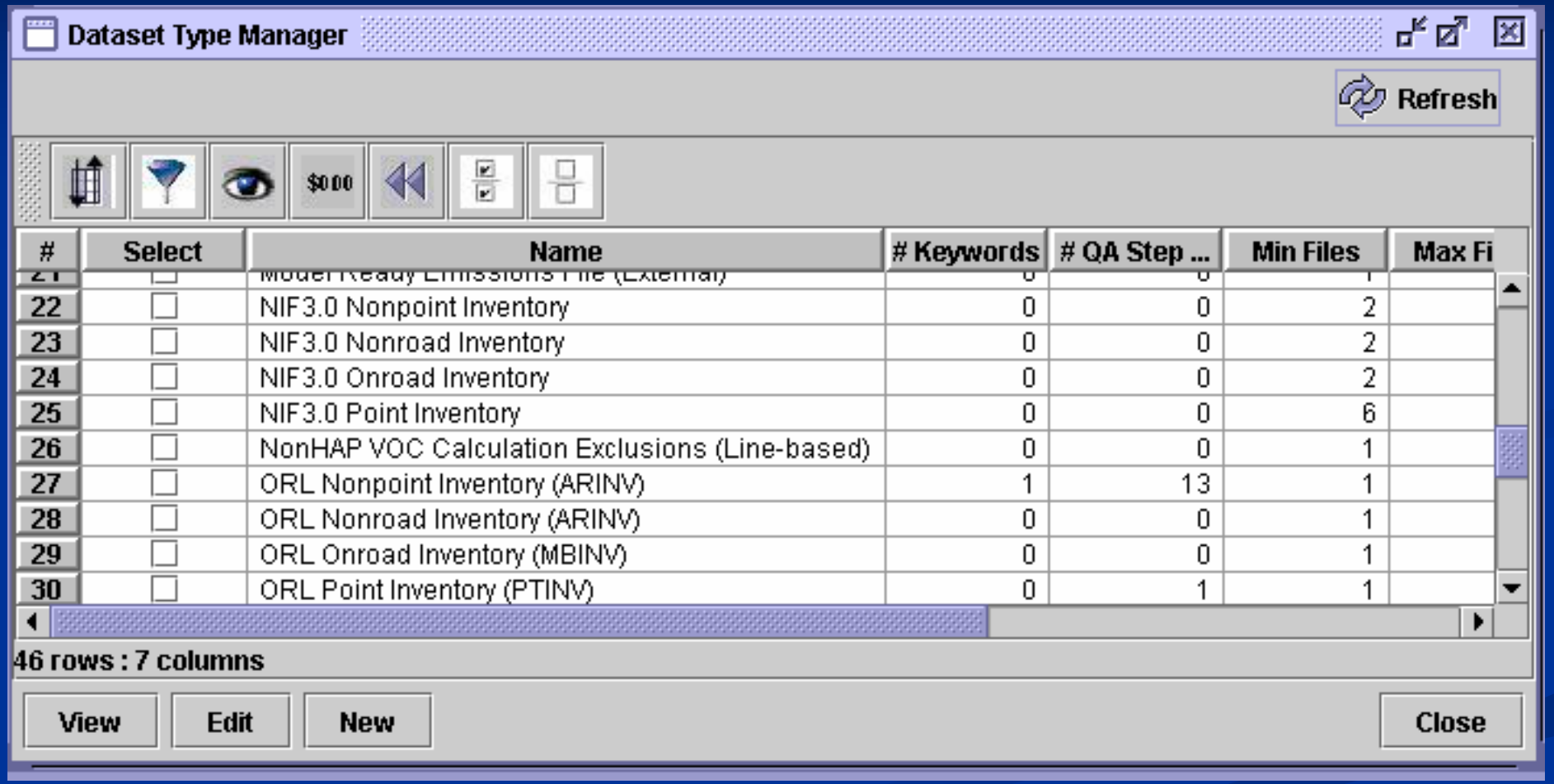
Example Inventory QA Steps

- 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

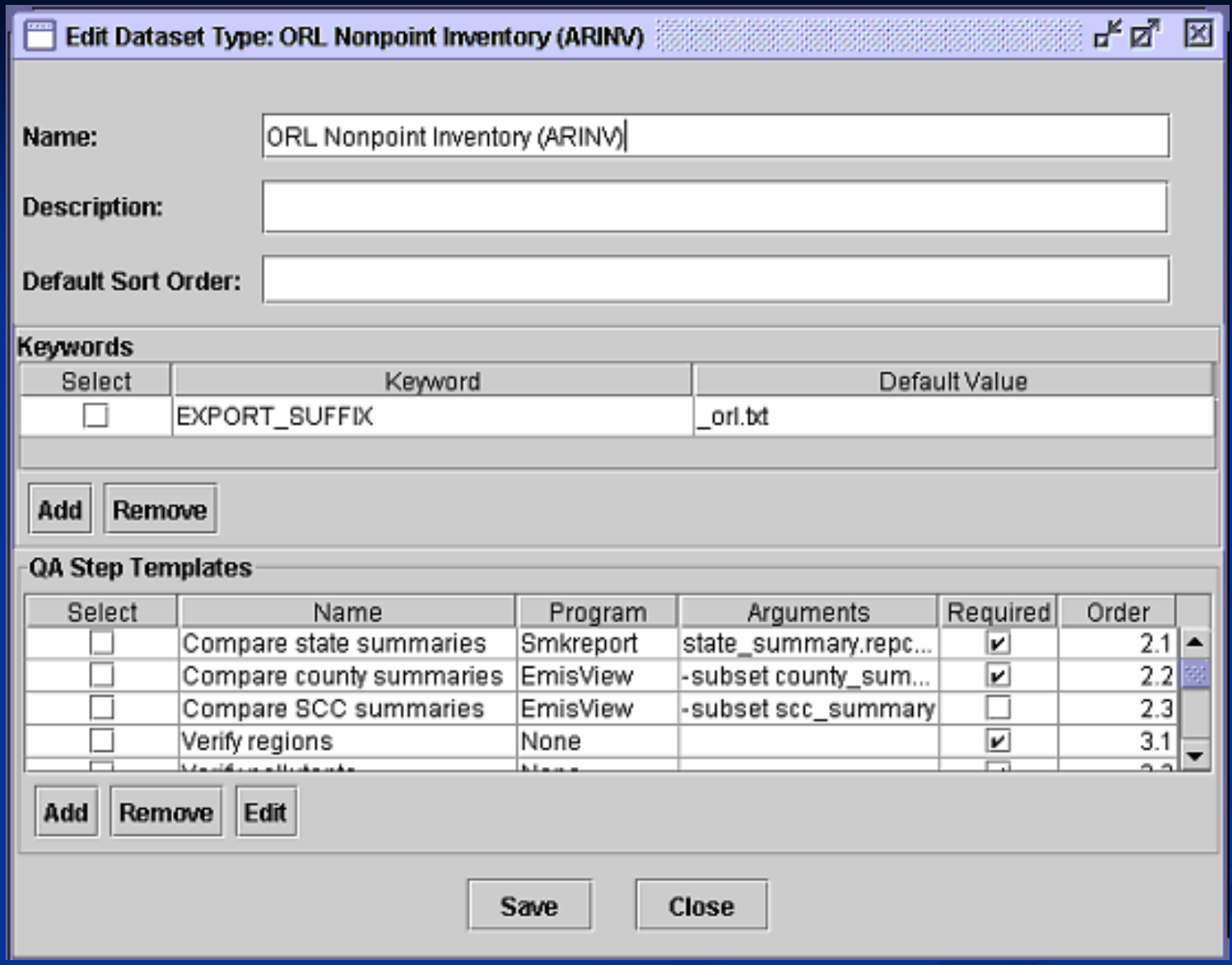
Tracking QA Steps in EMF

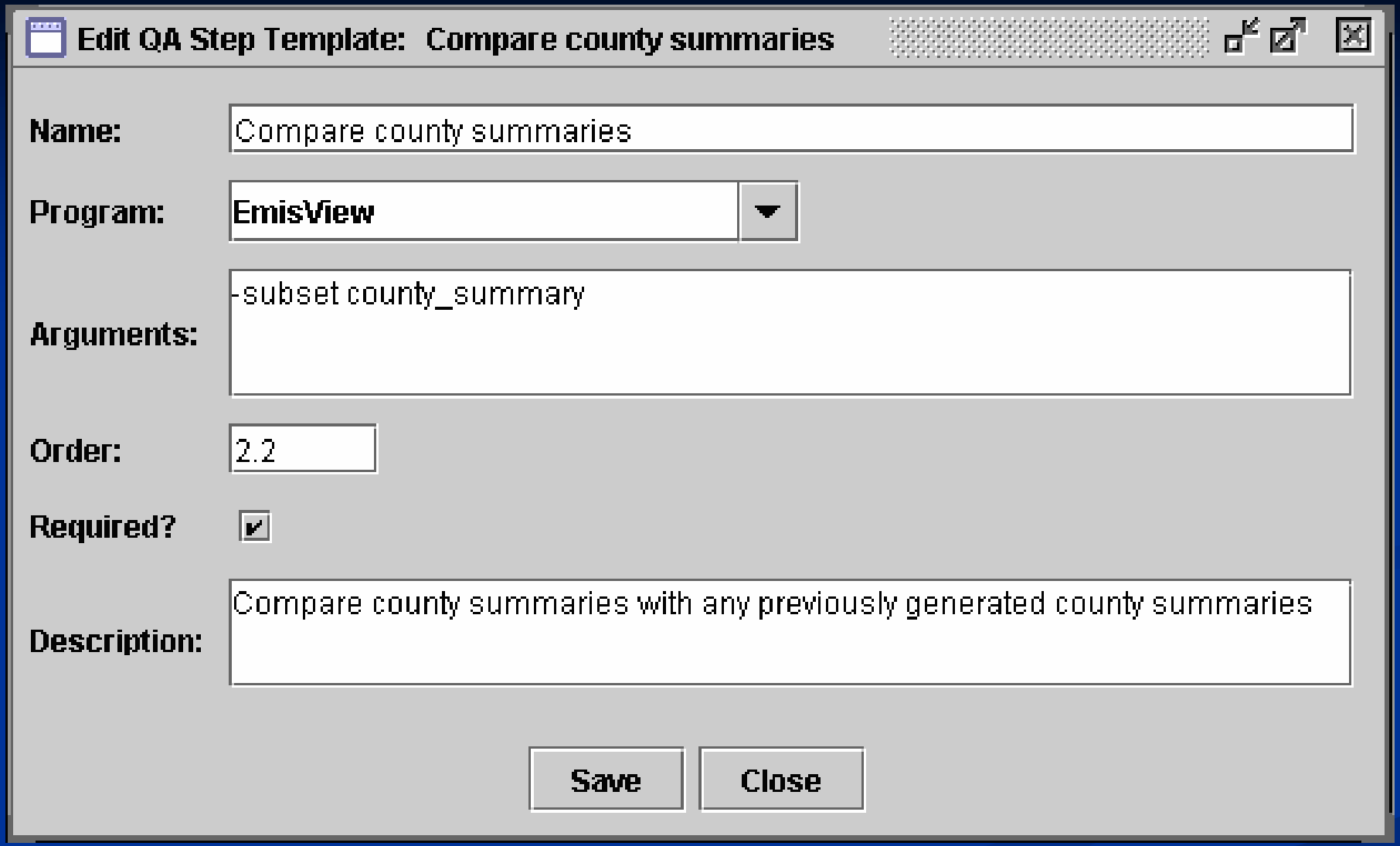
- First, set up Step Templates for Dataset Types
- Copy Steps to Dataset metadata from Templates
- Add any ad-hoc steps (not from templates) to Dataset metadata
- Record results of the steps for each version of a Dataset

Dataset Type Manager

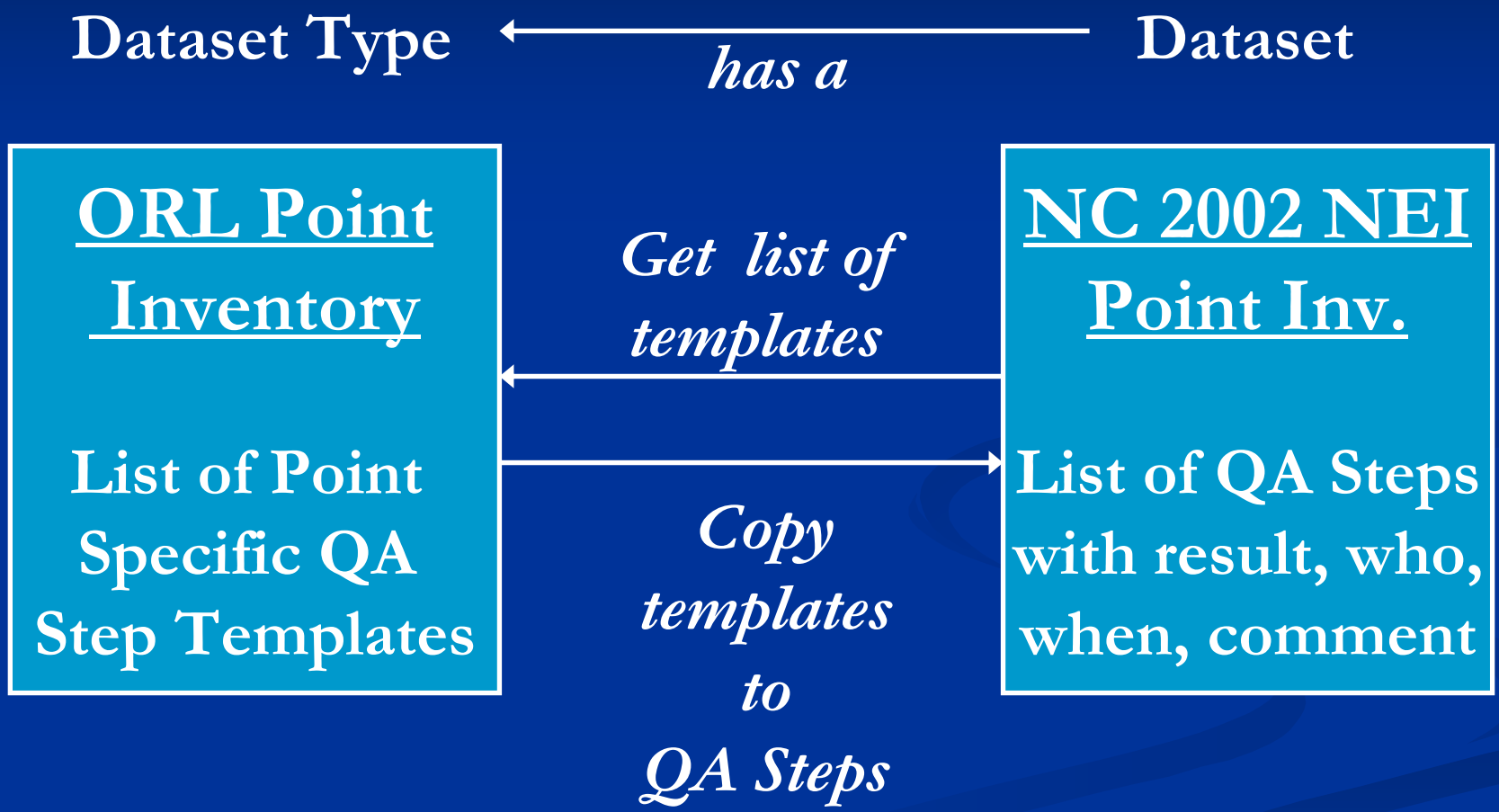


#	Select	Name	# Keywords	# QA Step ...	Min Files	Max Fi
22	☐	NIF3.0 Nonpoint Inventory	0	0	2	
23	☐	NIF3.0 Nonroad Inventory	0	0	2	
24	☐	NIF3.0 Onroad Inventory	0	0	2	
25	☐	NIF3.0 Point Inventory	0	0	6	
26	☐	NonHAP VOC Calculation Exclusions (Line-based)	0	0	1	
27	☐	ORL Nonpoint Inventory (ARINV)	1	13	1	
28	☐	ORL Nonroad Inventory (ARINV)	0	0	1	
29	☐	ORL Onroad Inventory (MBINV)	0	0	1	
30	☐	ORL Point Inventory (PTINV)	0	1	1	

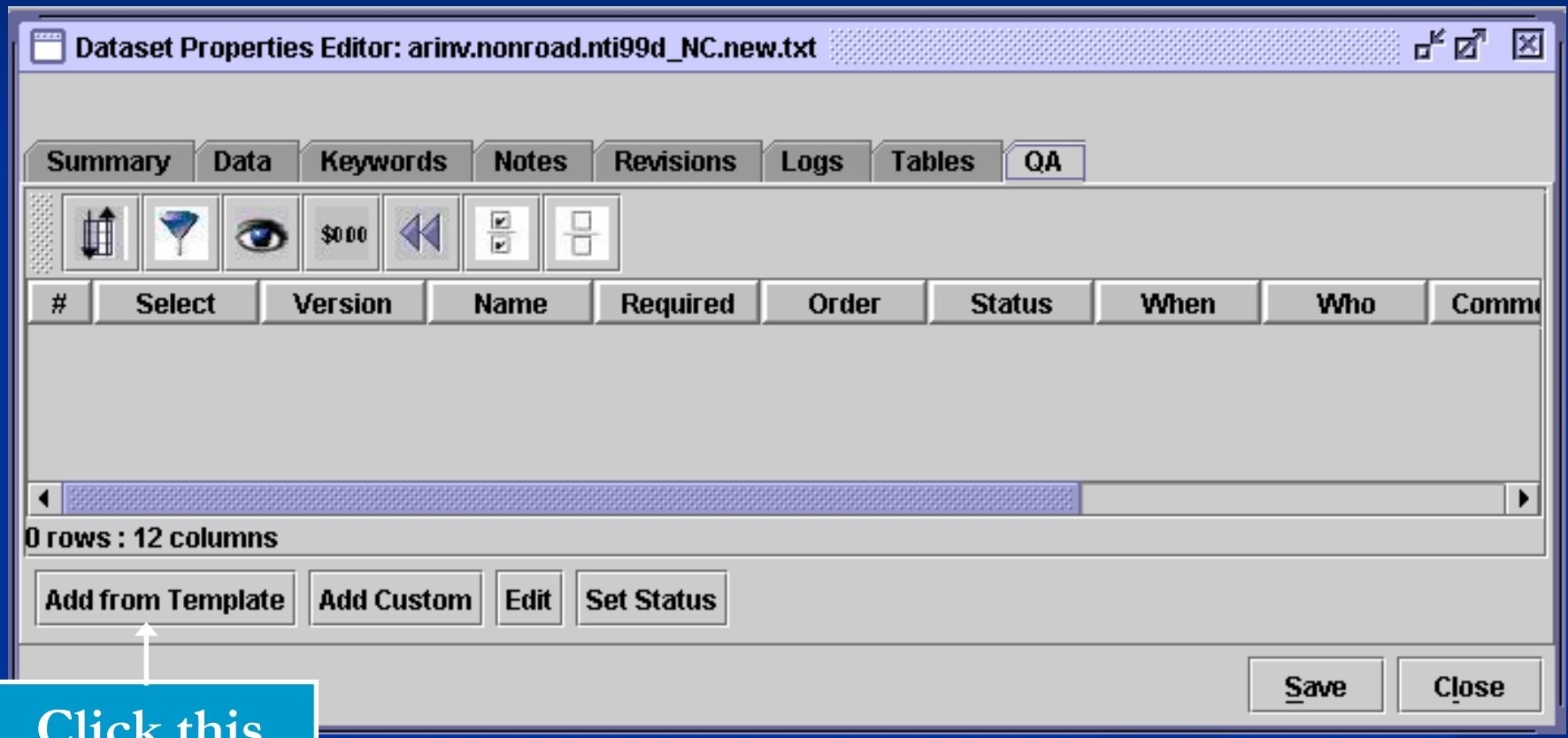

Edit Dataset Type: ORL Nonpoint Inventory (ARINV)
Name:
ORL Nonpoint Inventory (ARINV)
Description:
Default Sort Order:
Keywords
Select
Keyword
Default Value
EXPORT_SUFFIX
_orl.txt
Add
Remove
QA Step Templates
Select
Name
Program
Arguments
Required
Order
Compare state summaries
Smkreport
state_summary.repc...
2.1
Compare county summaries
EmisView
-subset county_sum...
2.2
Compare SCC summaries
EmisView
-subset scc_summary
2.3
Verify regions
None
3.1
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

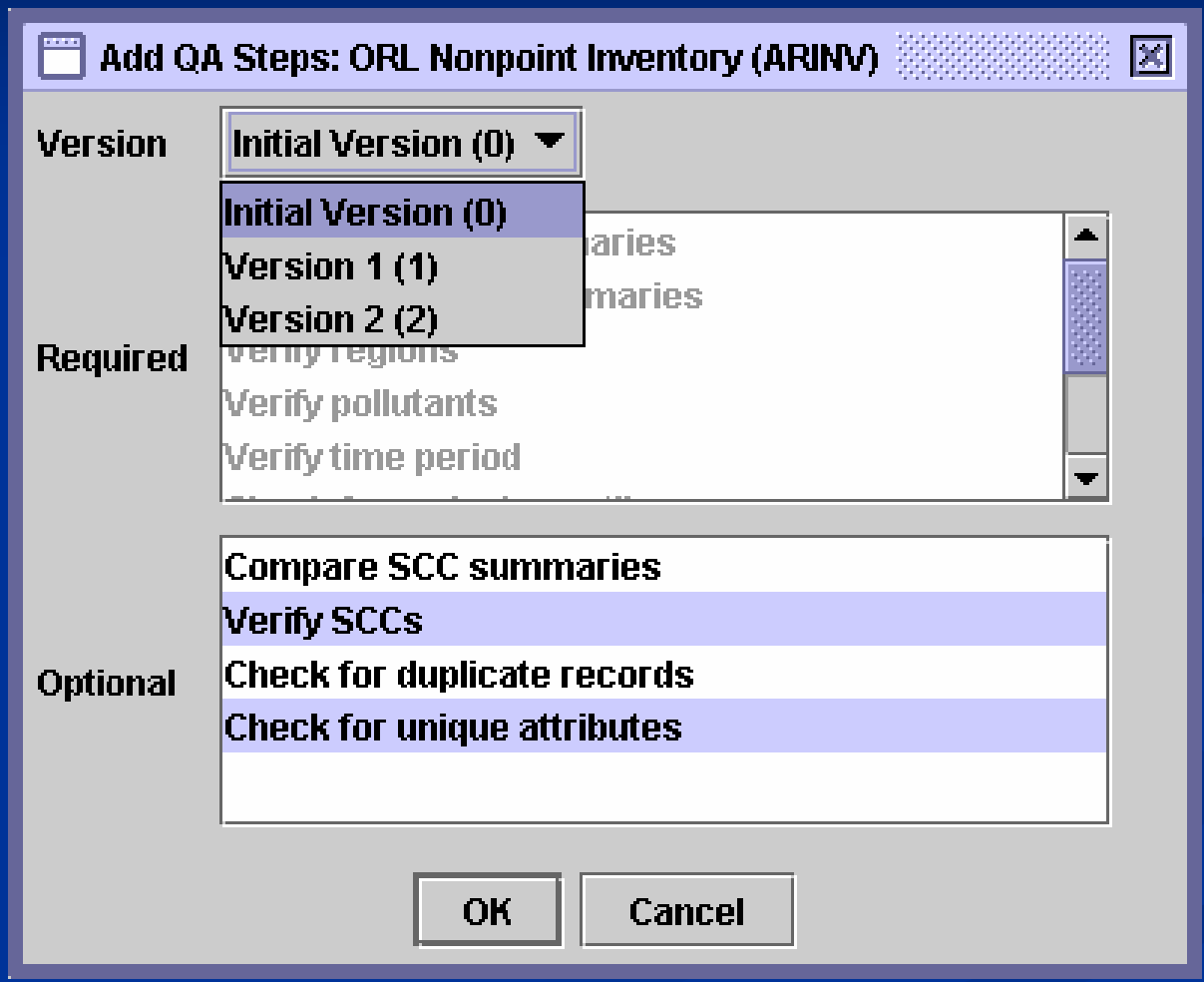


Adding QA Steps to a Dataset

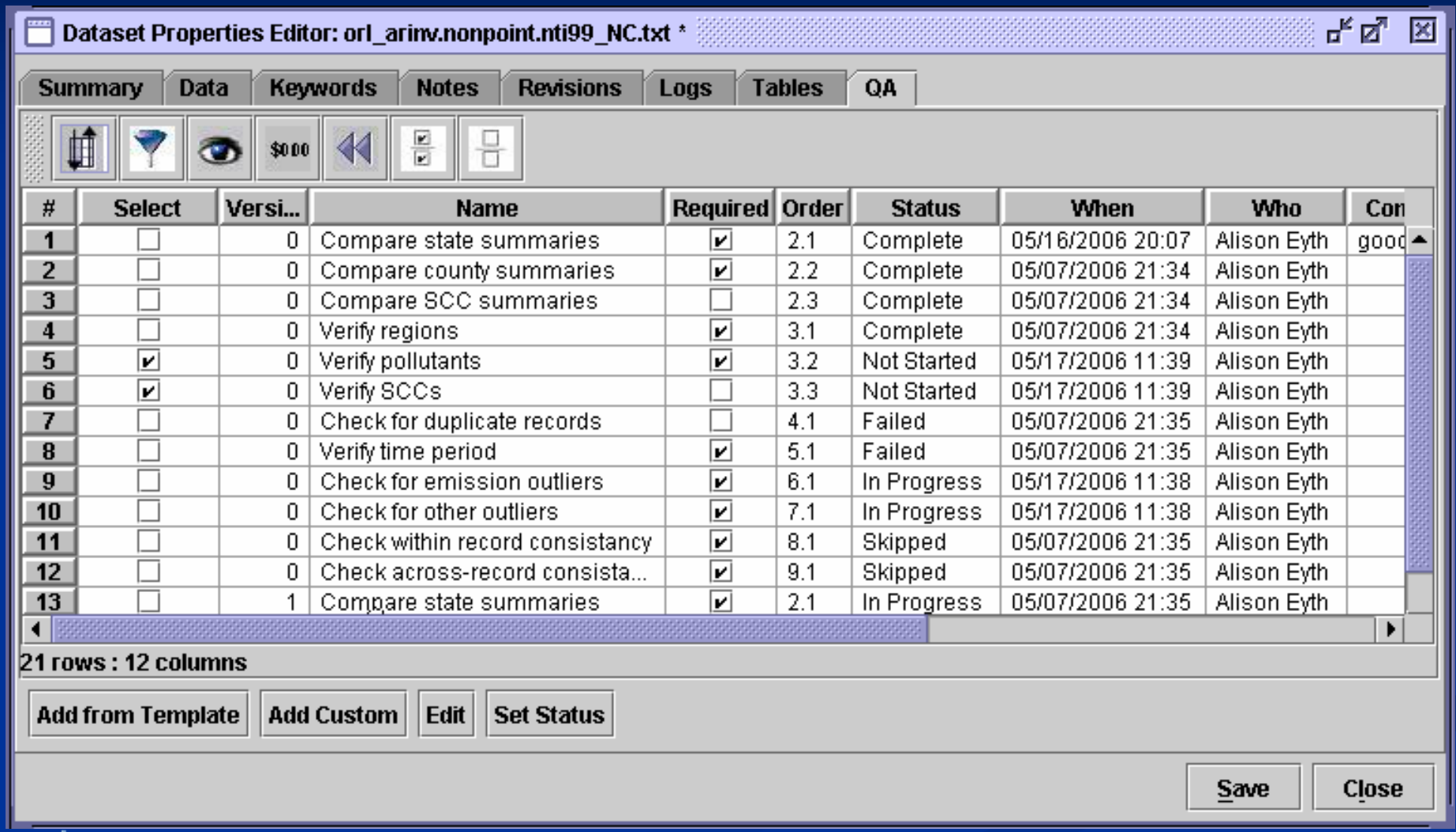


Click this button

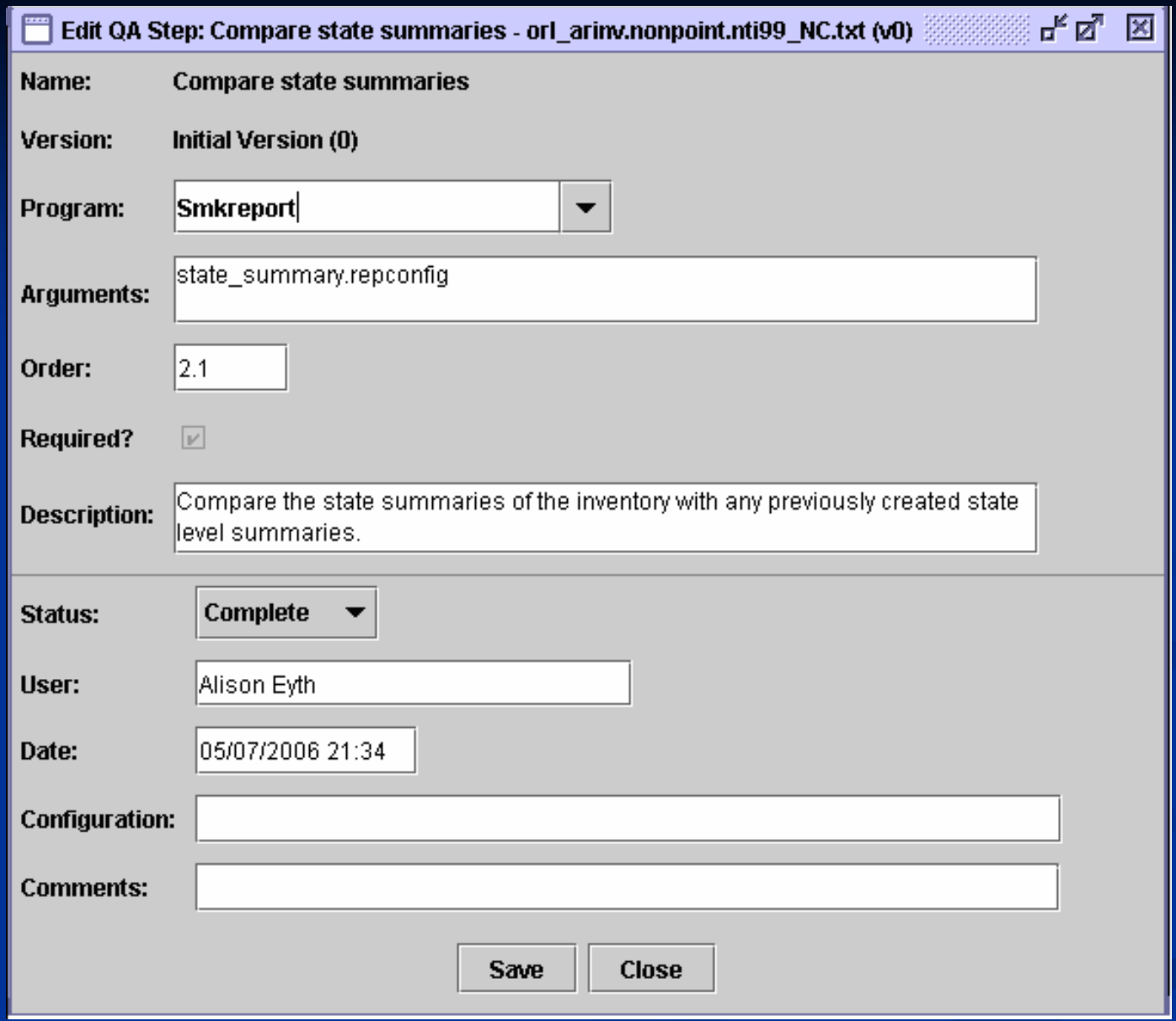
Adding Steps to Dataset from Templates



Summary of QA Steps for all Versions of a Dataset



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


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

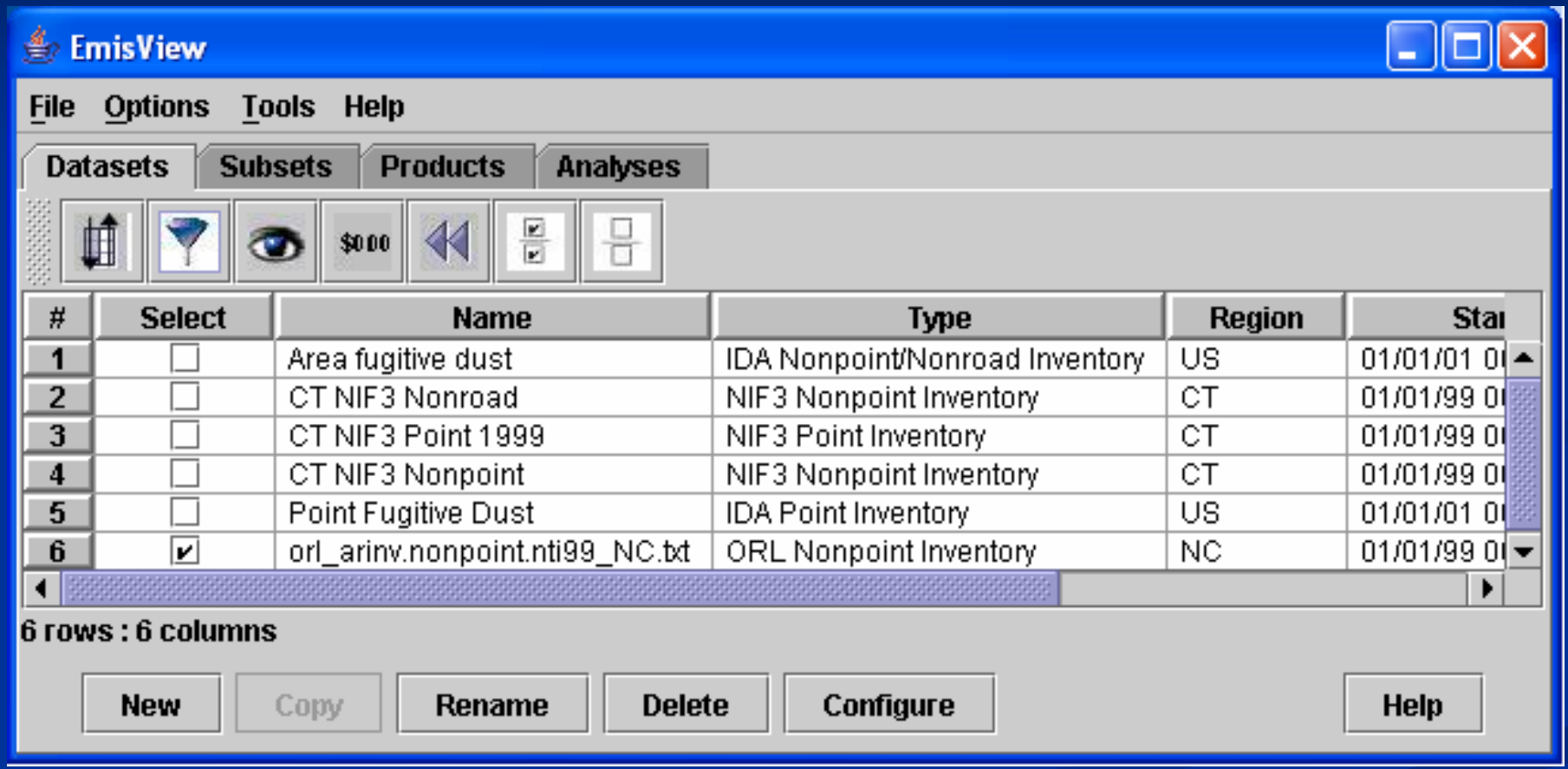
Goals for EmisView

- 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

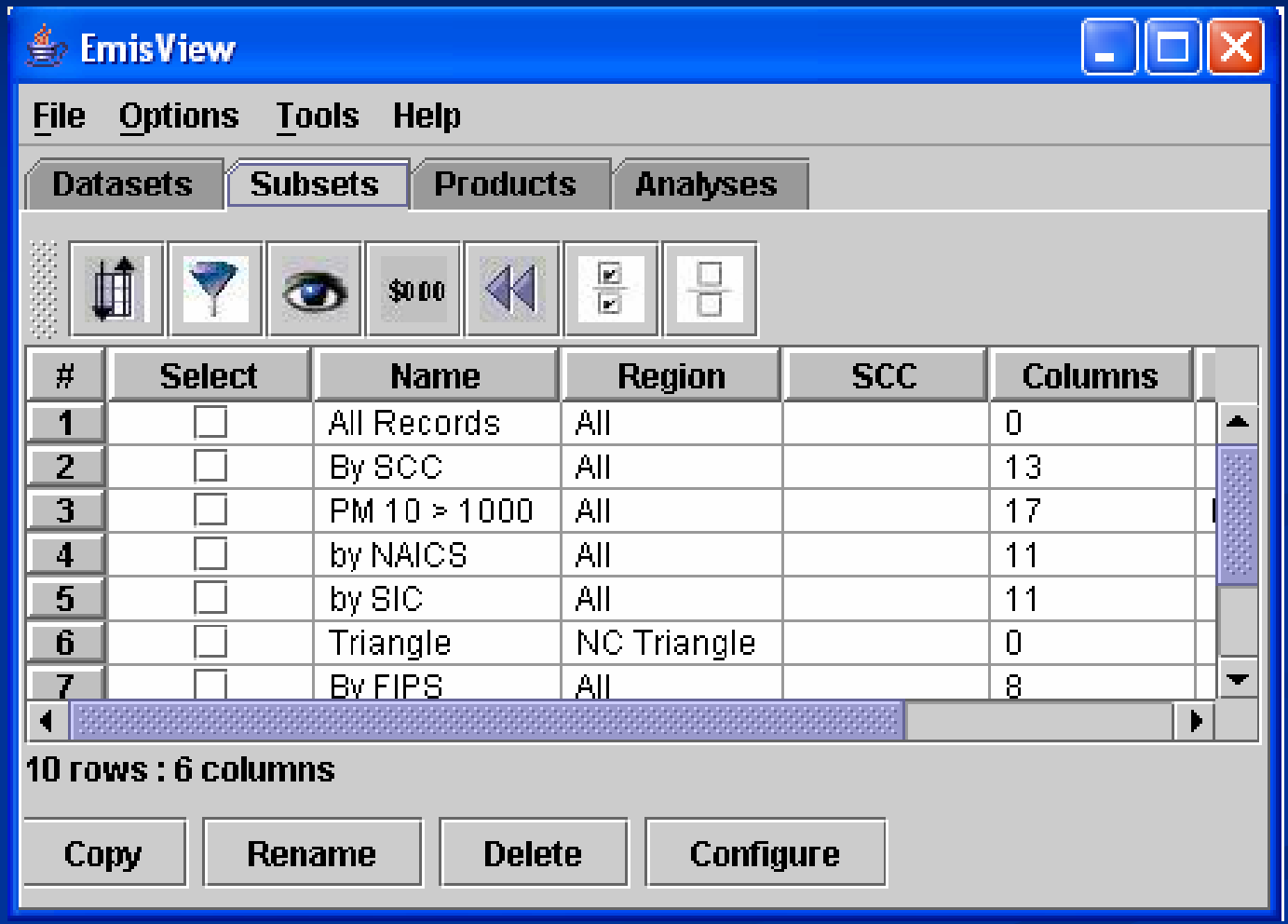
Mix and Match Components

- 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

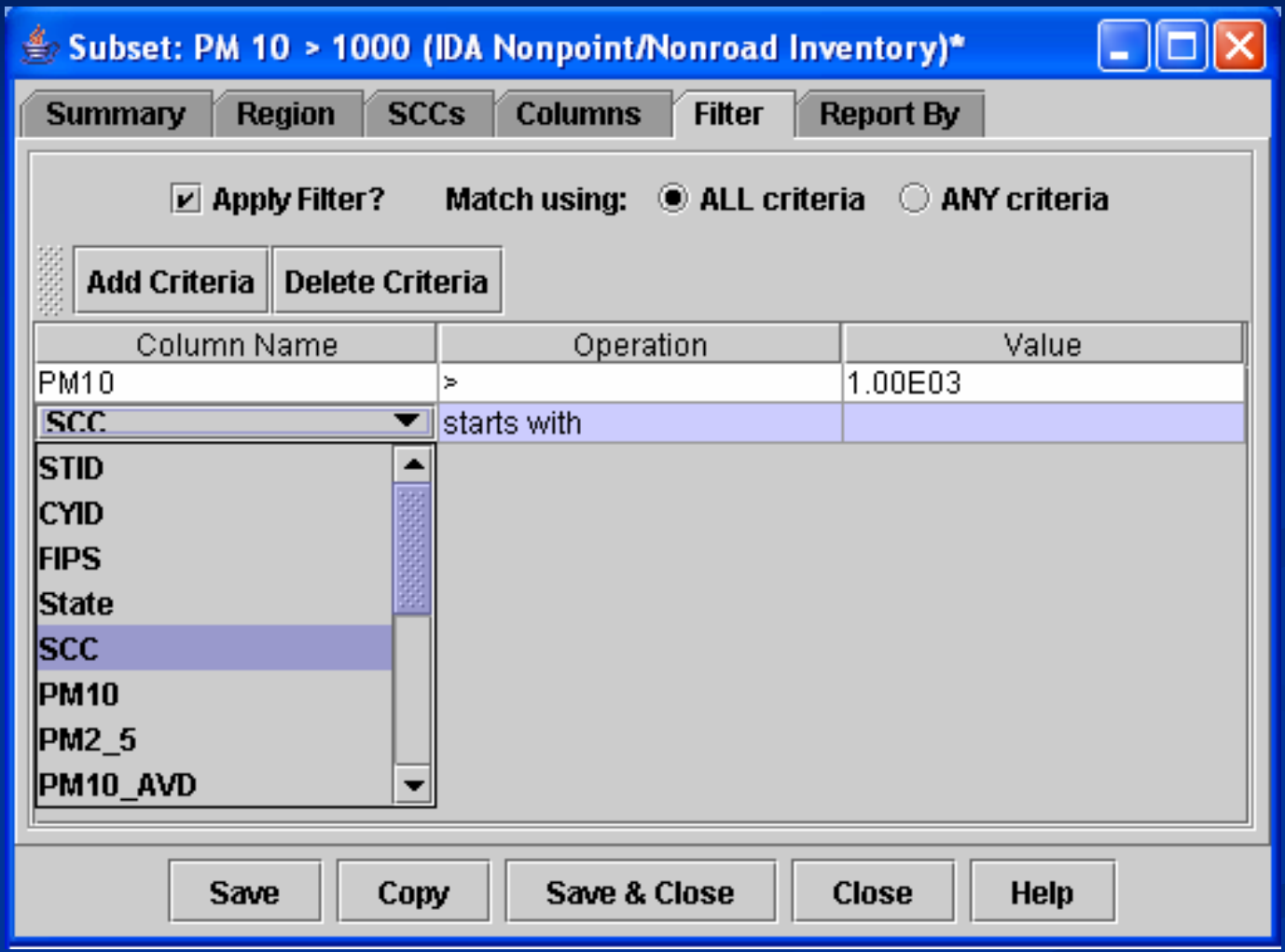
Main EmisView GUI - Datasets



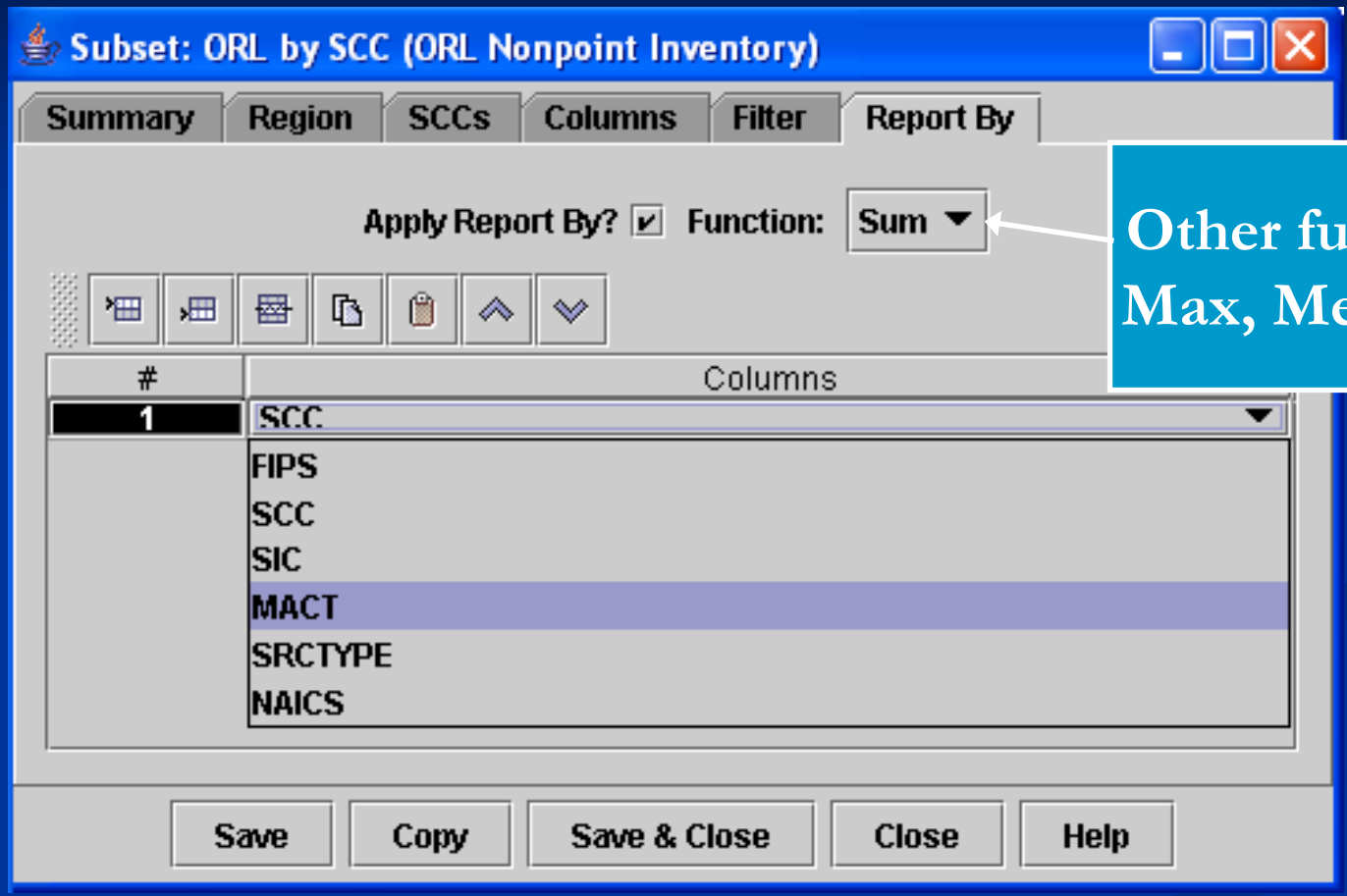
Main EmisView GUI - Subsets



A Threshold Subset



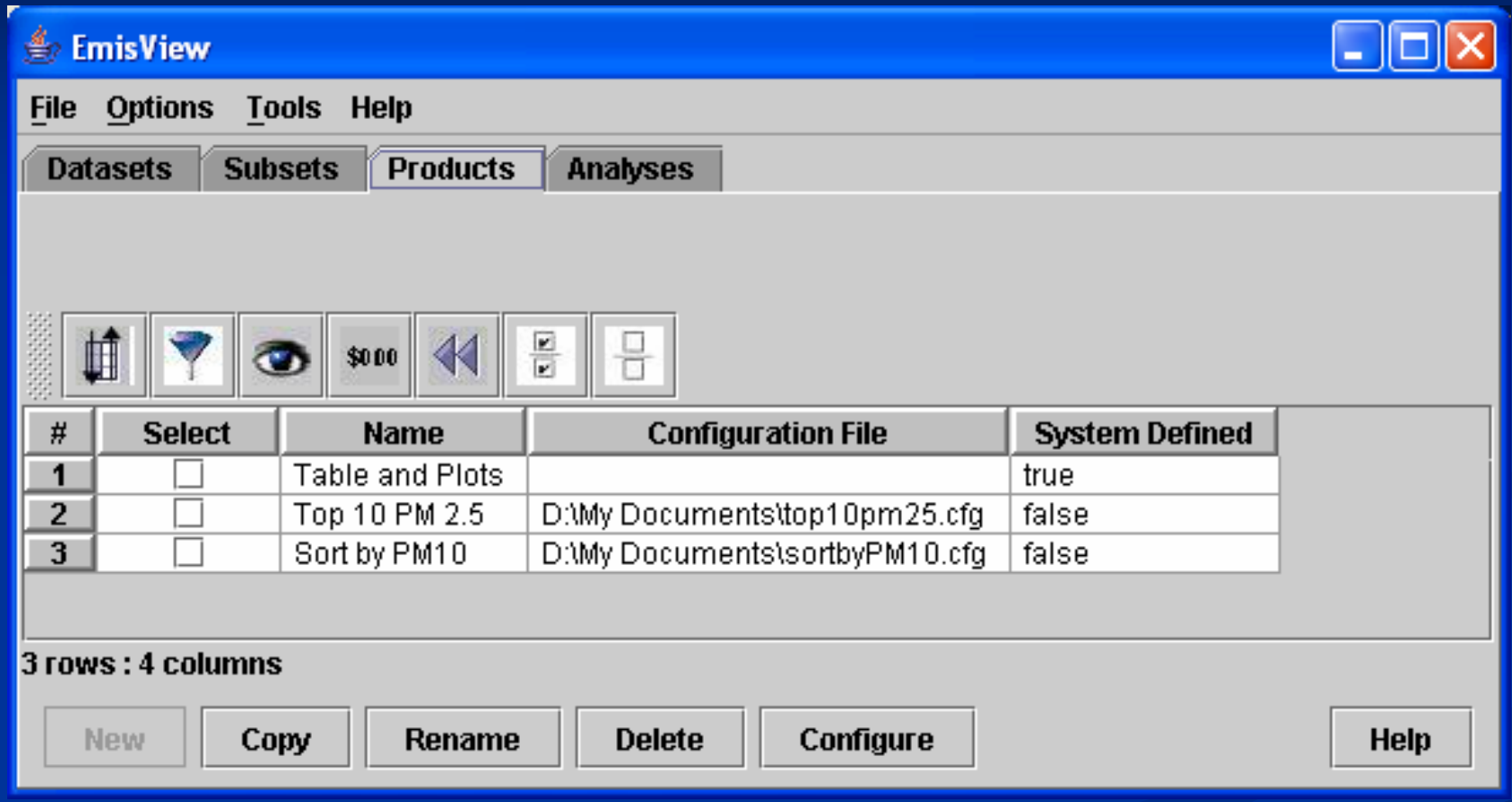
A Summarizing Subset



Other functions: Max, Mean, Min

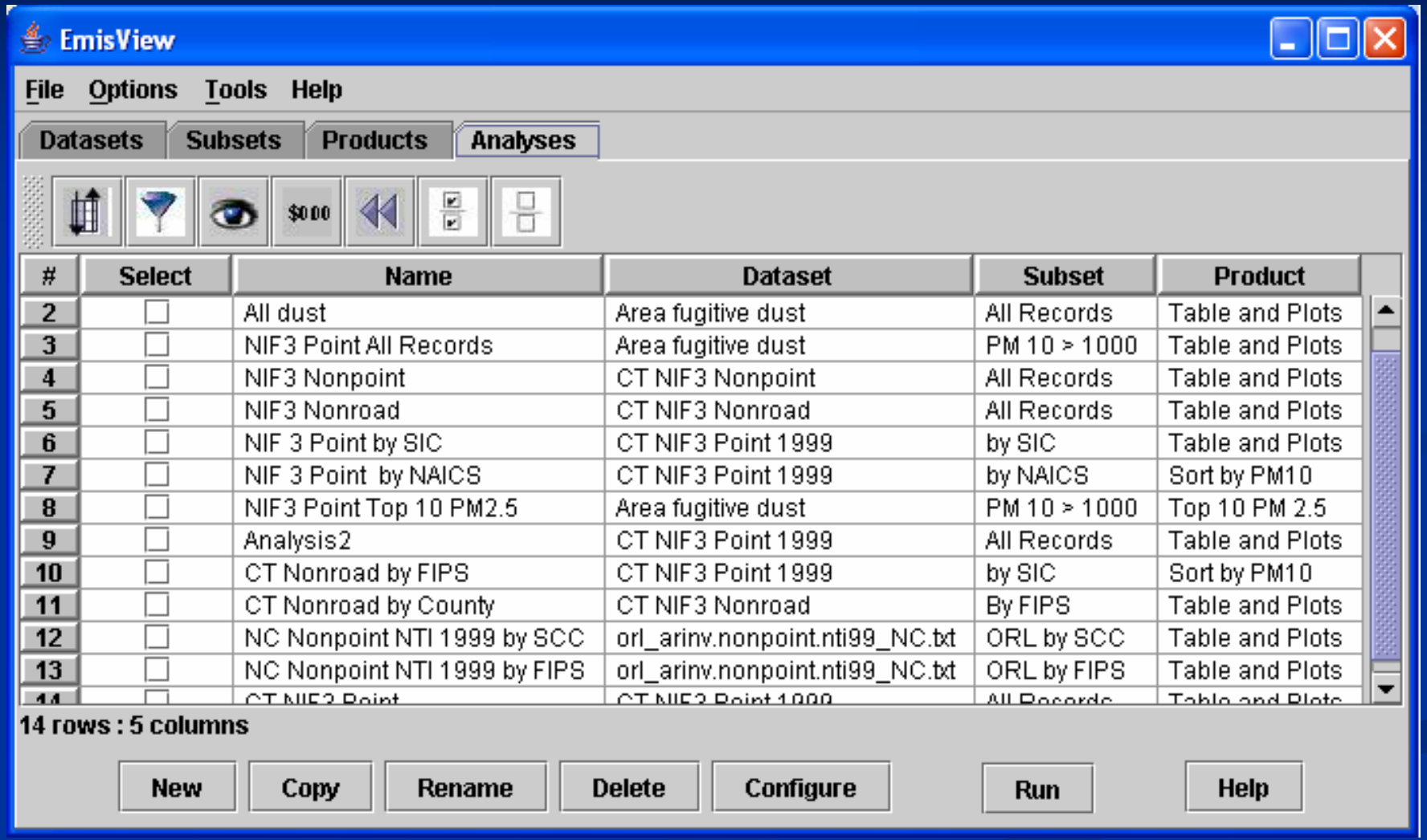
Other features of subsets: Set of states or counties, list of SCCs, choose columns – features are cumulative

EmisView Main GUI - Products

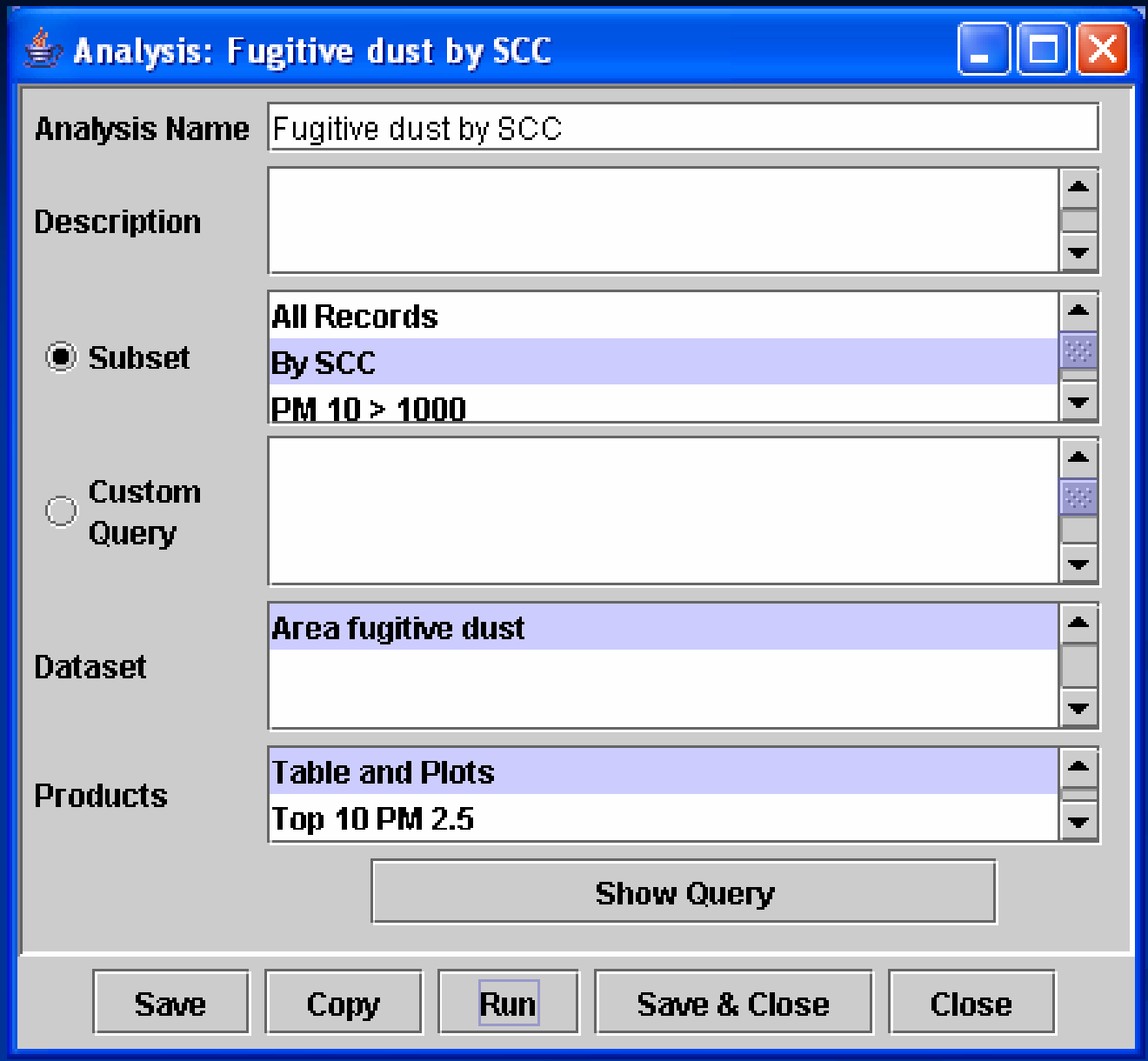


Currently, products are differentiated based on their Configuration file (saved from Table to be shown later)

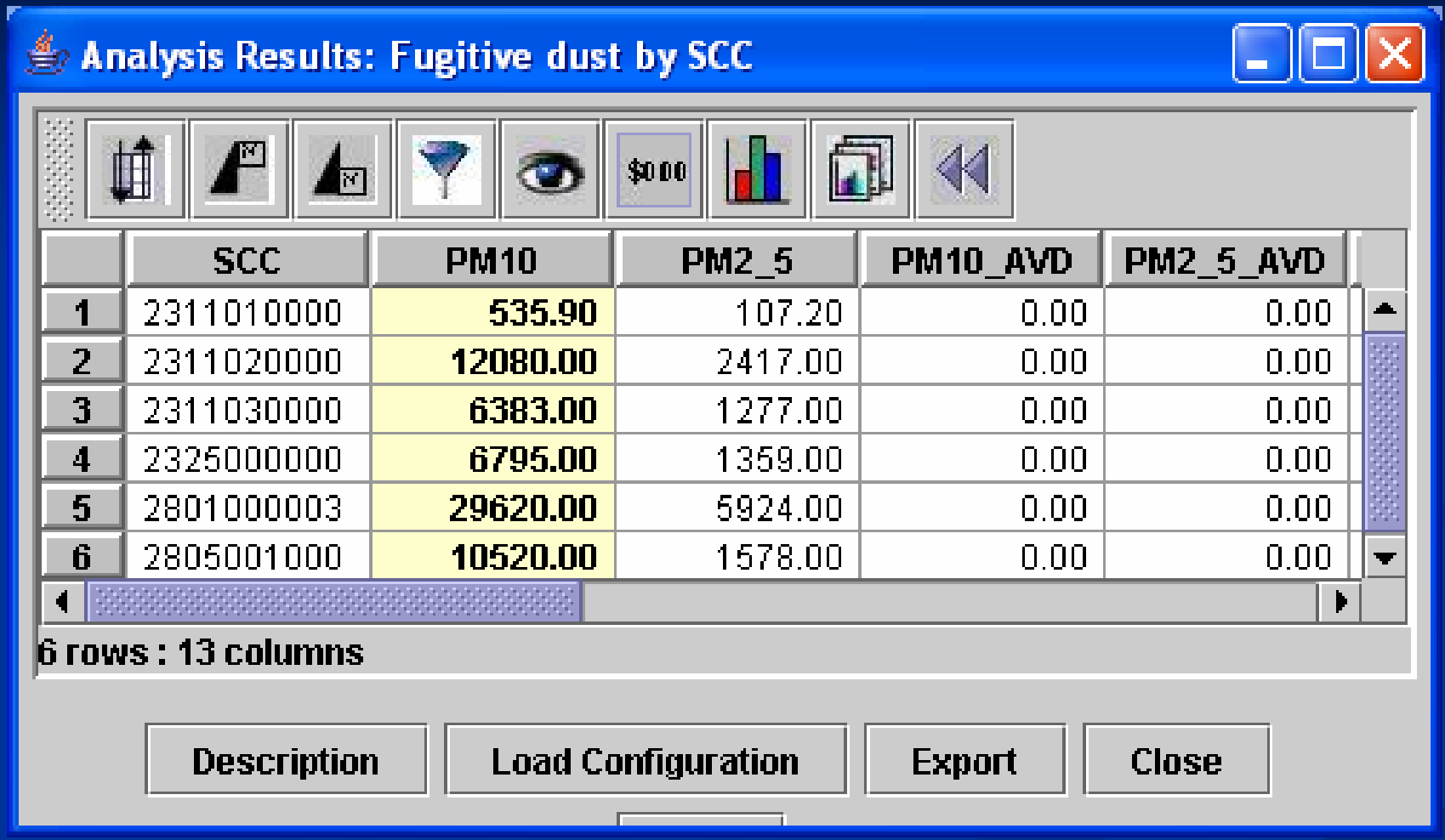
EmisView Main GUI - Analyses



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

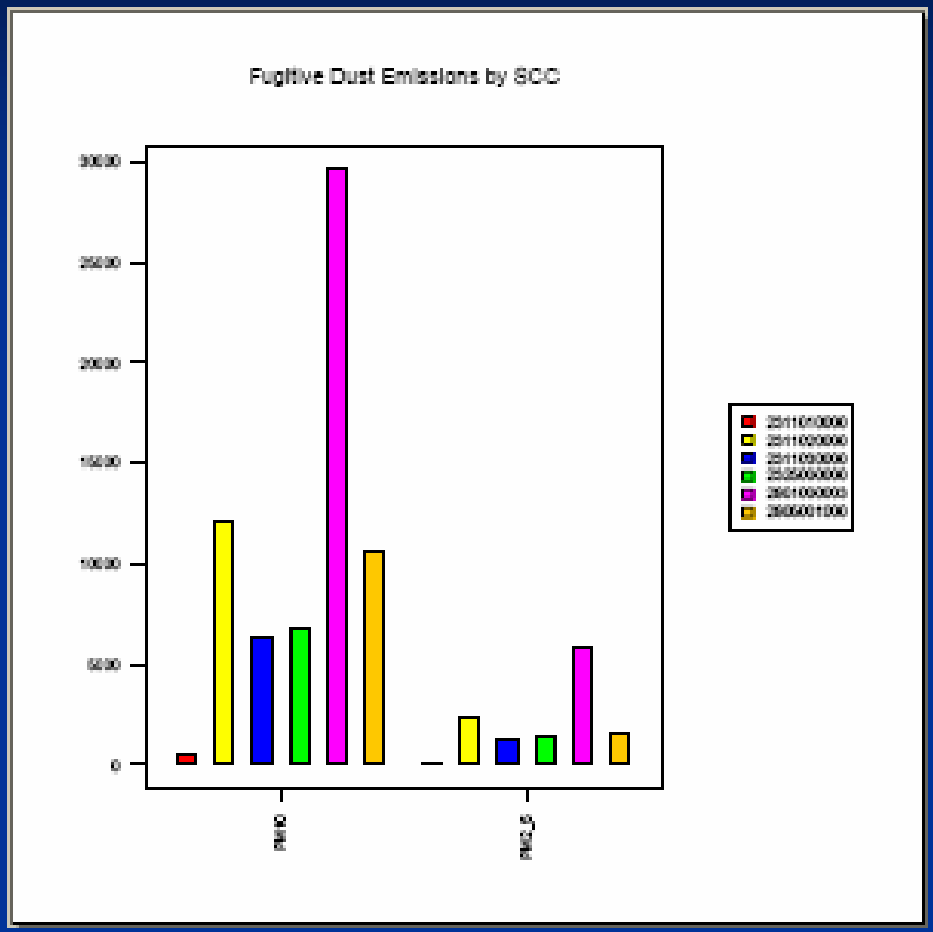

Analysis: Fugitive dust by SCC
Analysis Name
Fugitive dust by SCC
Description
Subset
All Records
By SCC
PM 10 > 1000
Custom Query
Dataset
Area fugitive dust
Products
Table and Plots
Top 10 PM 2.5
Show Query
Save
Copy
Run
Save & Close
Close

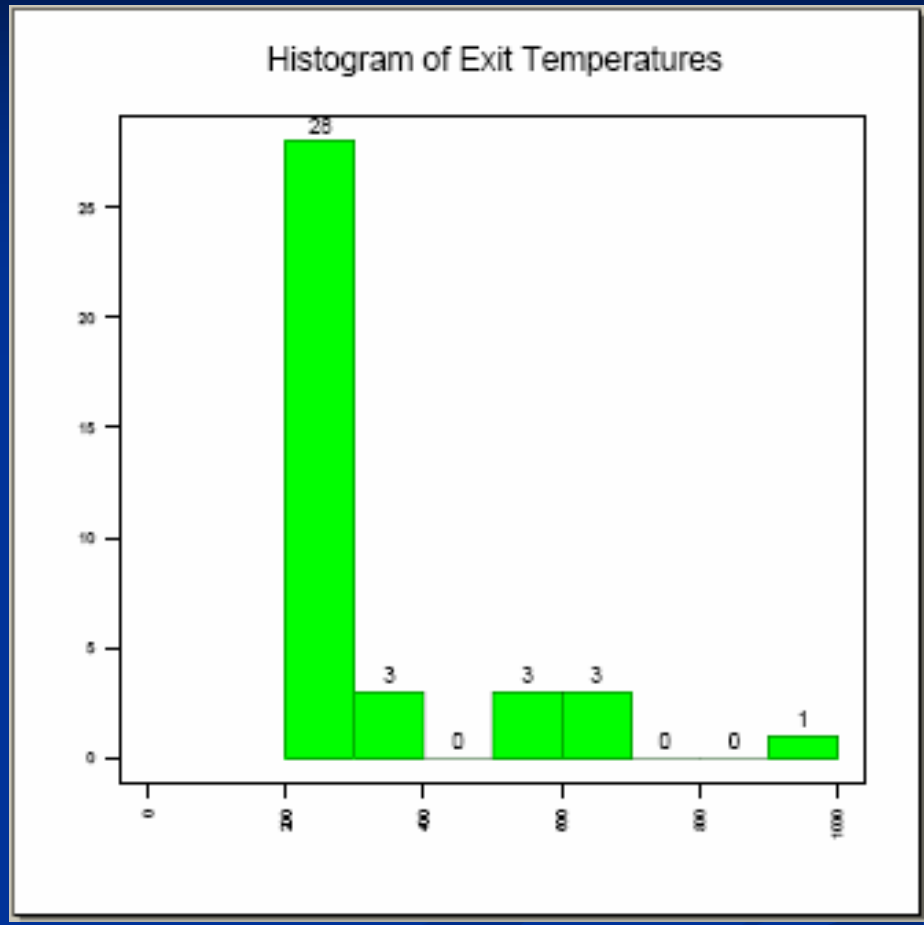
Analysis Results



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

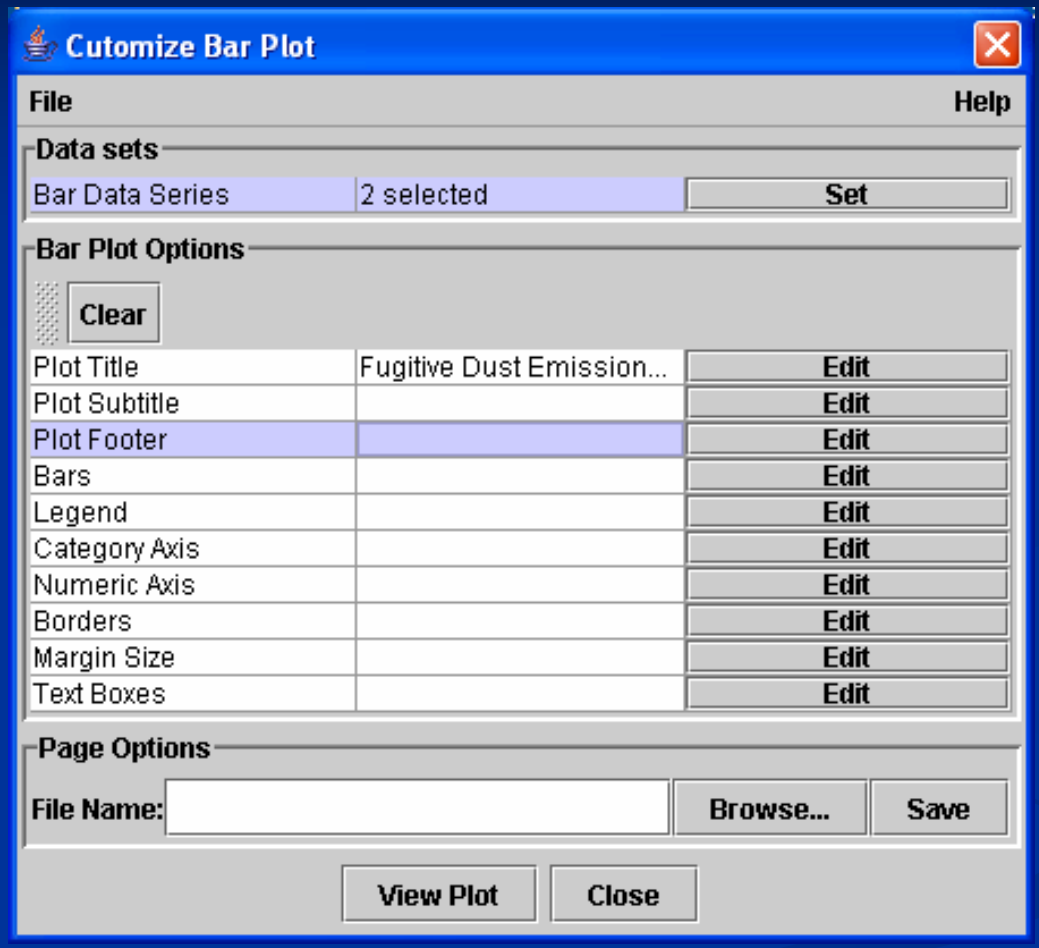
Examples of Plots





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

Plot Configuration Options



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