EPA 22nd Annual National Conference on MANAGING ENVIRONMENTAL QUALITY SYSTEMS

Session: Ambient Air II

PM Coarse Data Quality Objective Tool Development

Presented at:

Sheraton New Orleans Hotel
New Orleans, Louisiana

Presented on:

April 15-17, 2002

Presented by:

Basil W. Coutant

Battelle

Acknowledgments

Rona Boehm,
Chris Holloman,
Michele Morara,
Kristen Swinton, &
of Battelle

Shelly Eberly &
Mike Papp
of USEPA/OAQPS

Funded by EPA, Contract Number 68-D-02-061

Battelle

Overview

- The DQO goal
- The Simulation Models
- The *DQO Companion for* PM_{coarse} software
- Parameter estimates from AQS Data
- Progress report

INTRODUCTION

- Data Quality Objective Process
  - The goal is to ensure that the data collected are relevant to and meet decision-maker needs.
  - The hardest part - finding out decision maker needs.
  - Once the needs are specified (and quantified) statistical models (simulation models in this case) can be used to quantify data quality that ensure the decision-maker needs or demonstrate how various data quality issues affect the quality of the end product.

PM_{coarse} Measurement Goals

- First, it is not expected that PM_{coarse} will be measured directly. Instead, $PM_{coarse} = PM_{10} - PM_{2.5}$. So the DQOs need to be in terms of PM_{10} & $PM_{2.5}$.
- The need is to measure PM_{10} & $PM_{2.5}$ for a yet to be determined annual standard and a yet to be determined daily standard.

The Annual Standard (we hope!)

- The annual standard is to be based on the mean of three consecutive annual means of PM_{coarse} not the means of the PM_{10} & $PM_{2.5}$ annual means.
  - Missing data will happen! The difference of the means from different sampling days will over weight one or the other.

The Daily Standard (we hope!)

- The daily standard is to be based on the mean of three consecutive annual percentiles (again of $PM_{coarse,}$ not the percentiles of the PM_{10} & $PM_{2.5}$.)
  - What percentile? - Hopefully the software will be used in making that decision. What we have seen from the tool is that the 98th percentile is not a good choice. Europe is using the 90^{th}.
  - How you calculate the percentile can make a big difference, especially out past the 90th percentile.

The Modeling Process

- First specify a model for the quantities of interest:
  - Long term seasonal patterns are represented by sinusoidal patterns for both PM_{coarse} and $PM_{2.5}$ with a phase shift between them allowed. (This implies a sinusoidal pattern for the PM_{10}. Everything that will go into the measurement process needs to be simulated.)
  - Random Log-normal deviations from the seasonal patterns are assumed. These deviations are allowed to be correlated in time and between the two fractions.

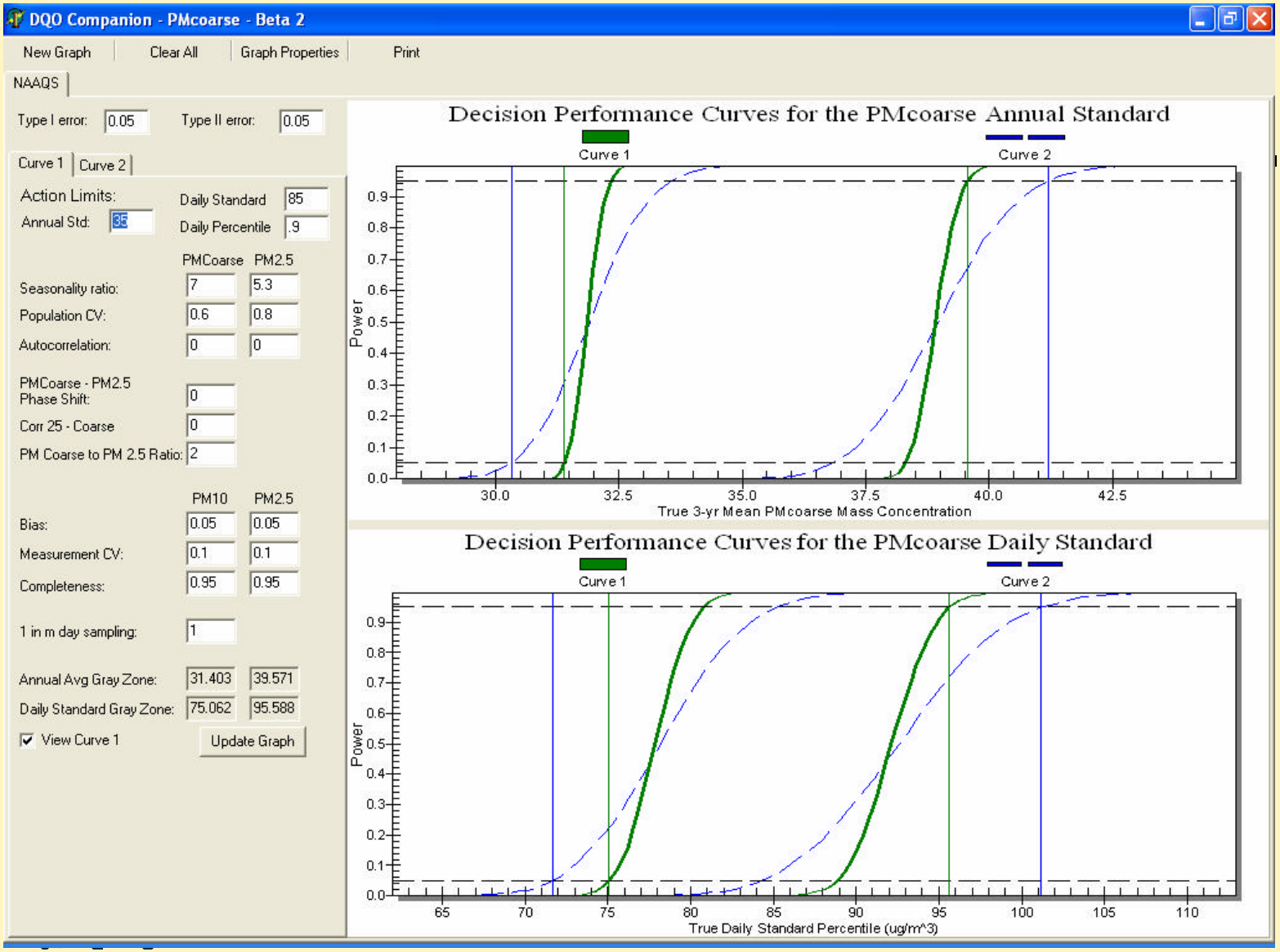
Battelle

The Modeling Process (cont.)

- Next each of the parameters in the model(s) need to be estimated.
  - The estimation is at the site level. To get national level DQOs, the range of the parameter is examined to get a "worst" case estimate for each.
  - We have done this - details to follow later.
- With the above, we can simulate true concentrations and measured concentrations with varying levels of completeness, precision, and bias. This allows us to examine how often decision errors occur.

The DQO Companion for PM_{coarse} Software.

- The software plots decision performance curves. These show likely a site would be declared in violation of the (selected) standard.
- The software can plot up to 5 different scenarios at a time.
- The user interface is very similar to the *DQO Companion* software developed for $PM_{2.5}$ (Has anyone here used it?)
- Values shown are not intended to suggest DQOs or a standard.


DQO Companion - PMcoarse - Beta 2
New Graph
Clear All
Graph Properties
Print
NAAQS
Type I error: 0.05
Type II error: 0.05
Curve 1
Curve 2
Action Limits:
Annual Std: 35
Daily Standard 85
Daily Percentile .9
PMCoarse PM2.5
Seasonality ratio: 7 5.3
Population CV: 0.6 0.8
Autocorrelation: 0 0
PMCoarse - PM2.5 Phase Shift: 0
Corr 25 - Coarse 0
PM Coarse to PM 2.5 Ratio: 2
PM10 PM2.5
Bias: 0.05 0.05
Measurement CV: 0.1 0.1
Completeness: 0.95 0.95
1 in m day sampling: 1
Annual Avg Gray Zone: 31.403 39.571
Daily Standard Gray Zone: 75.062 95.588
View Curve 1
Update Graph
Decision Performance Curves for the PMcoarse Annual Standard
Curve 1
Curve 2
Power
True 3-yr Mean PMcoarse Mass Concentration
Decision Performance Curves for the PMcoarse Daily Standard
Curve 1
Curve 2
Power
True Daily Standard Percentile (ug/m^3)

The Two Example Senarios

- Annual standard = 35
- Daily = 85 for the 90th percentile
- Bias = Up to +/-5% for both PM_{10} & $PM_{2.5}$ The curves show the worst case combination for positive and negative bias.
- Precision = 10% Coefficient of Variation (CV)
- Completeness = 95% for both PM_{10} & $PM_{2.5}$
- *Green = Daily sampling and Blue = every third day.*

Parameter Estimates

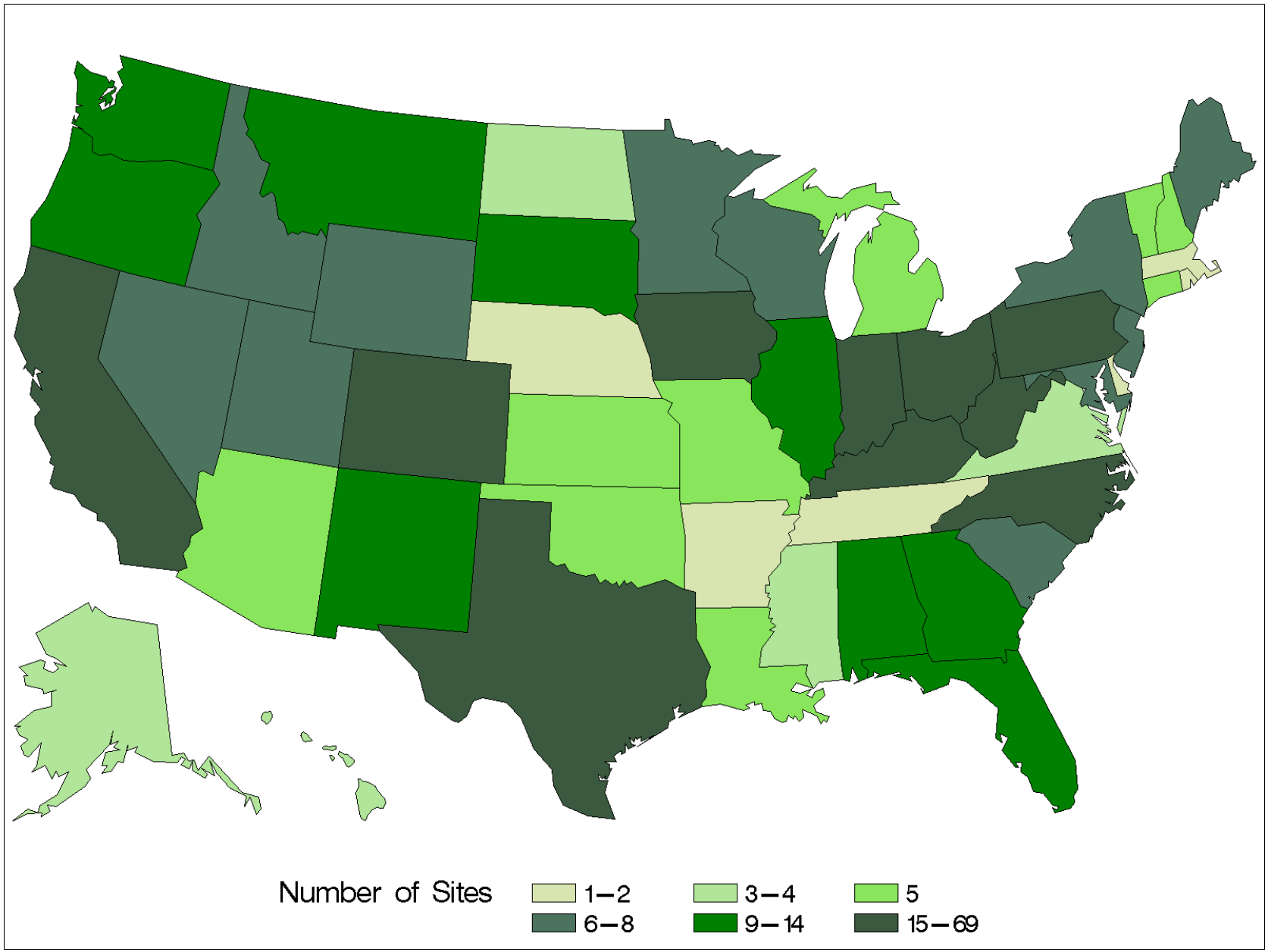
- AQS data has been used to develop site specific parameter estimates for:
  - The seasonality ratios (high to low points in the sinusoidal curve.)
  - The "population CV" (a measure of how much the true values deviate from the seasonal pattern.)
  - Autocorrelation (a measure of how similar the deviations are from day-to-day.)
  - The correlation between the PM_{coarse} & $PM_{2.5}$ deviations.
  - PM_{coarse} to $PM_{2.5}$ ratio (the ratio of the annual means)
  - The phase shift between the seasonal patterns.

Battelle

AQS Data

- 622 sites with co-located data between 1999-2001
- Limited to sites with at least 3 (paired) concentrations in each month. (502 sites left.)
- Estimates for everything except the autocorrelation for these 502 sites. (See map.)
- Autocorrelation estimates require daily sampling. (65 sites used.) The lowest non-negative value is the conservative choice. The 10th percentiles were 0.05 and 0.11 for $PM_{2.5}$ & PM_{coarse} respectively. (So 0 is good in general.)

Battelle


Number of Sites
1–2
3–4
5
6–8
9–14
15–69

Disclaimer

- The PM_{10} & $PM_{2..5}$ measurements are in "different" units. PM_{10} is reported in units of "standard conditions," while $PM_{2.5}$ is reported under local conditions. (Under standard conditions, the volume used to calculate the concentrations is adjusted to 1 Atm and 25 C.) Presumably this will not continue.
- This can be treated as an extra source of noise in the data used. (i.e. Don't use the most extreme estimates. Use the 10th and 90th percentiles.)
- Or use the methods described in the paper to recalculate the estimates.
- *The software allows ranges beyond the estimated values just in case.*

The parameter estimates

Quantile	2.5 ratio	Coarse ratio	2.5 CV	Coarse CV	2.5 autocor.	Coarse autocor.	Coarse / 2.5	C-22.5 cor.
2.5	1.46	1.68	0.35	0.4	0	0	0.28	-0.23
10	1.63	2.05	0.41	0.49	0.05	0.11	0.37	-0.05
40	2.02	3.24	0.51	0.66	0.36	0.24	0.72	0.19
50	2.14	3.82	0.53	0.71	0.38	0.27	0.87	0.25
60	2.28	4.42	0.56	0.76	0.41	0.38	1.04	0.31
90	4.01	14.34	0.69	1.08	0.58	0.54	2.22	0.56
97.5	5.72	52.52	0.8	1.39	0.8	0.69	3.29	0.69

Progress

- A national database of ambient behavior estimates has been built with existing data. Reasonable ranges can be inferred, in spite of the disclaimer. Report due by the end of the month.
- The software tool should be released by the end of the month. Remaining items include minor edits to the user's guide and putting everything together into a self-installing package.
- Decision-makers still need to decided what the NAAQS will be and the right balance between decision errors and DQO characteristics.