

3 - Year Fine Particle (PM_{2.5}) Data Quality Assessment

**Mike Papp, Shelly Eberly,
Mark Schmidt**





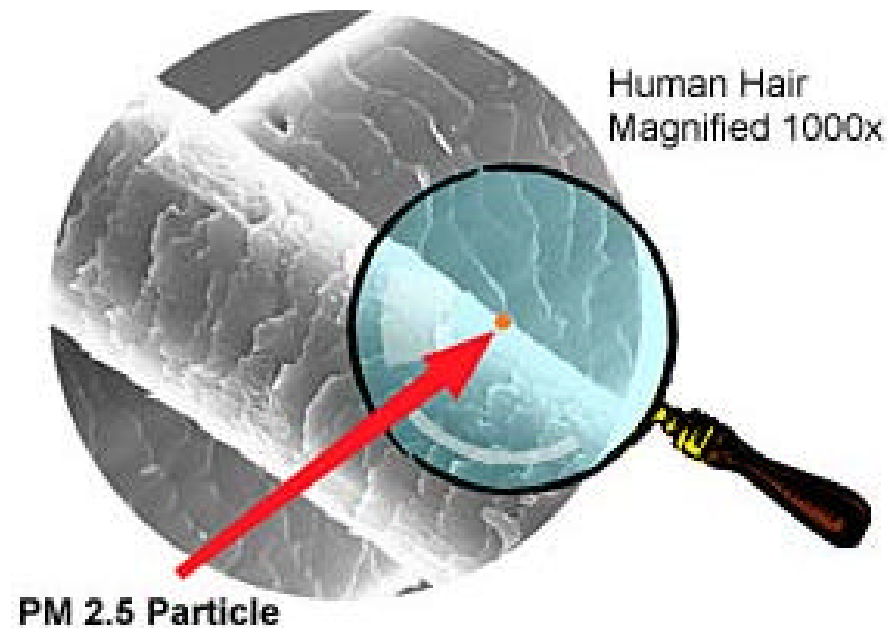
Presentation Topics

- **Program Background**
- **Data Quality Objectives**
- **Data Quality Assessment**
- **Next Steps**

Fine Particles (PM_{2.5})- what are they?

A complex mixture of extremely small particles and liquid droplets with aerodynamic diameters of \leq [a nominal] 2.5 μm

PM_{2.5} particles are so small that 30 of them side-by-side would barely equal the width of a human hair (graphic courtesy of U.S. Department of Energy)



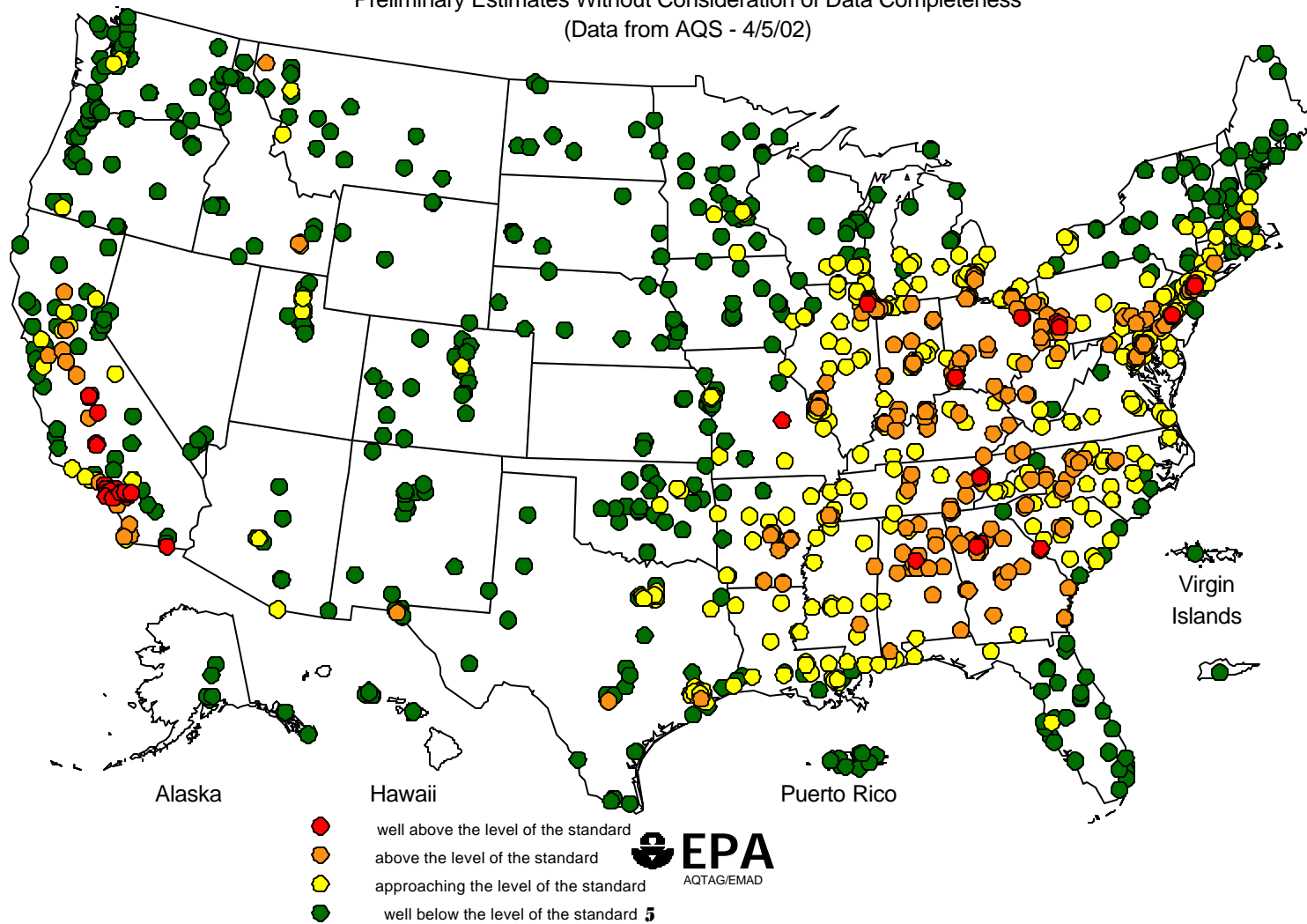
Why Collect PM_{2.5} Data?

- **Comparison with annual PM_{2.5} NAAQS (15 ug/m³)**
- **Comparison with daily PM_{2.5} NAAQS (65 ug/m³)**
- **Information for sensitive groups (AQI)**
- **General information to public (mapping)**
- **Support health studies, evaluation of emission inventories, simulation models, ...**
- **General understanding/characterization (temporal and spatial) of air quality**

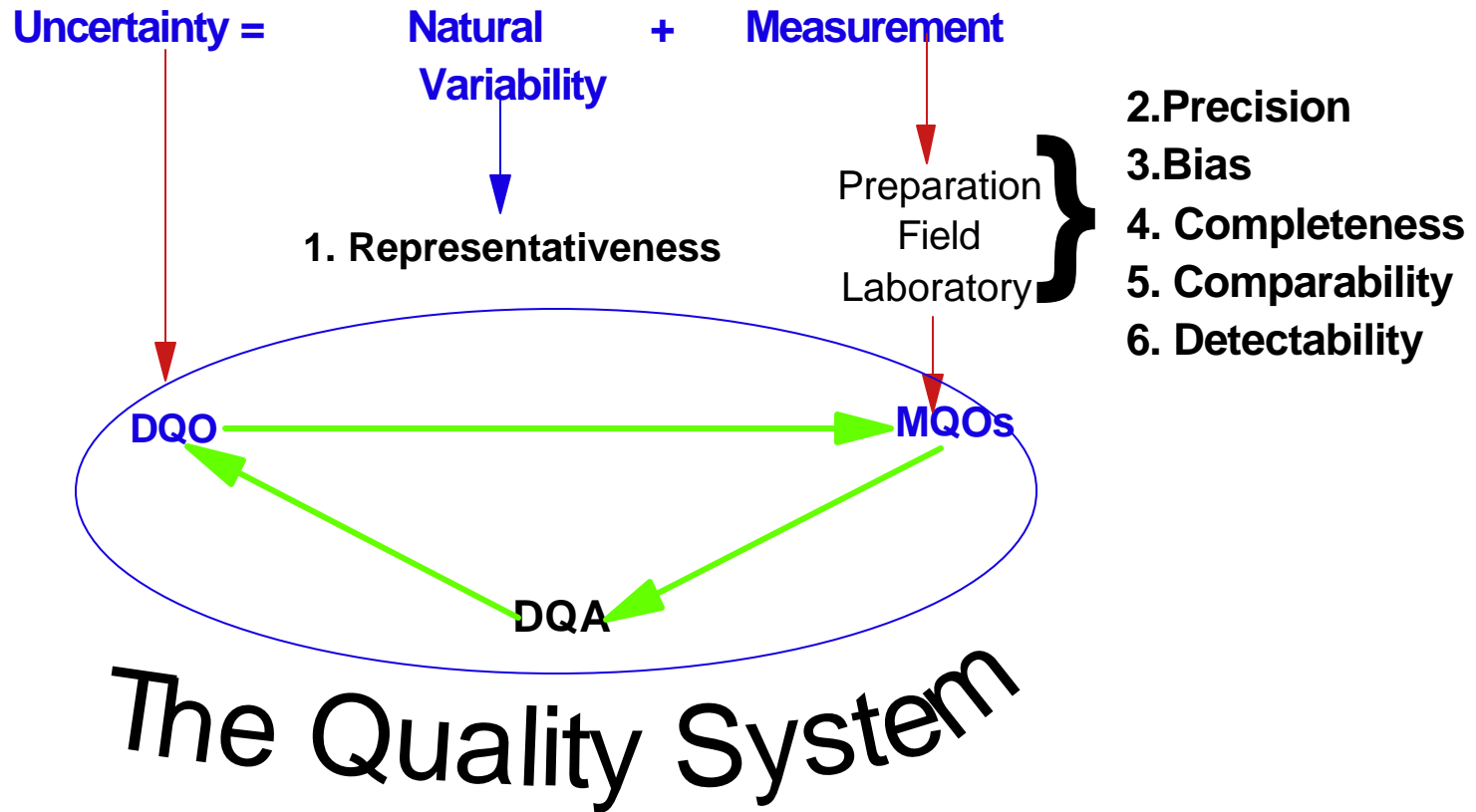
*Data can be used for all these analyses... BUT...
real question is how confident are we in the results?*

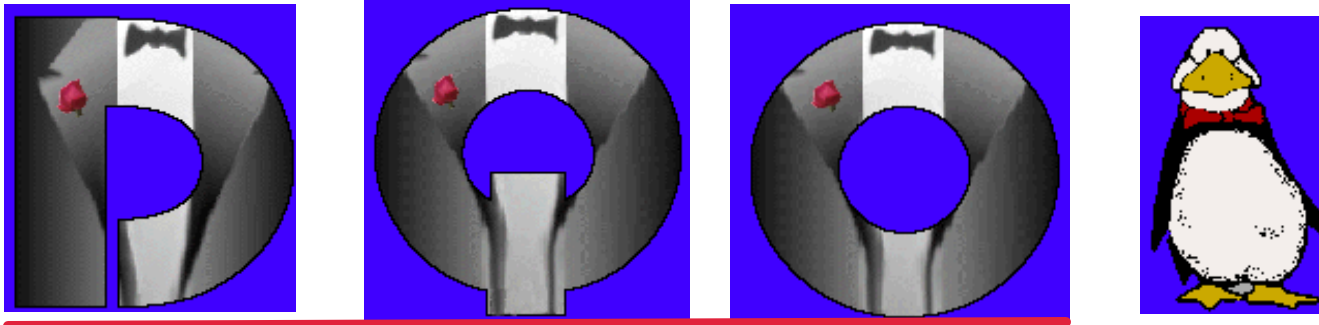
1999-2001 Annual Mean PM_{2.5}

Preliminary Estimates Without Consideration of Data Completeness
(Data from AQS - 4/5/02)



Understanding and Controlling Uncertainty





A process for ensuring that environmental data will be adequate for their intended use.

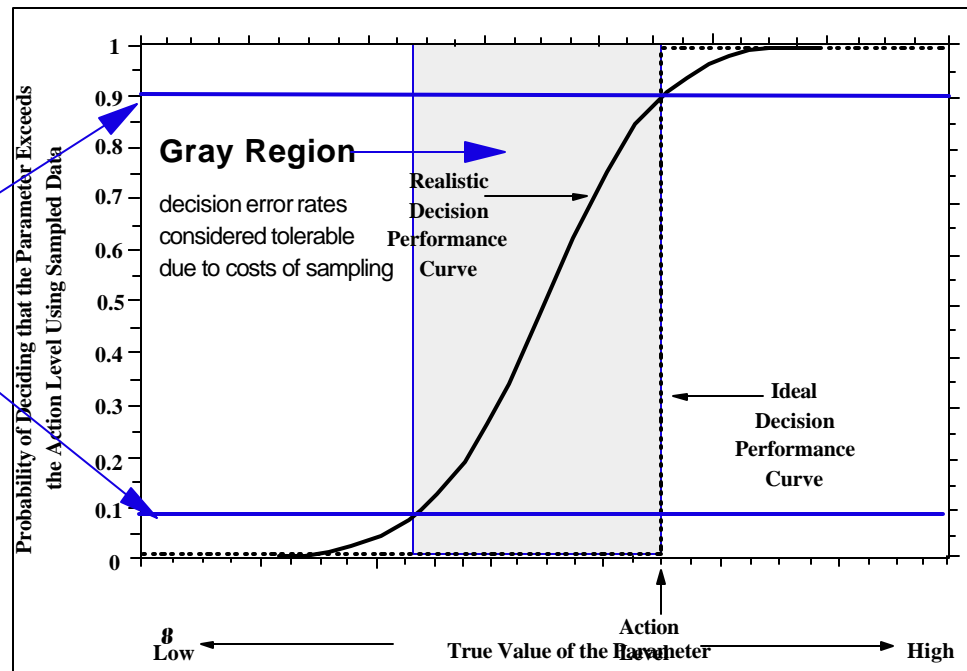
- ▶ Clarifies study objectives
- ▶ Defines appropriate types of data to collect
- ▶ Specifies the tolerable levels of potential decision errors

What is a power curve?

- Graphically represents the quality of the decision process
- Shows the probability that environmental data will lead us to a given decision, as function of unknown truth
- Stipulate the decision makers tolerable risk for decision errors
- Assists in understanding the magnitude of uncertainties and optimizing sampling designs

Decision error limits

The risk the decision maker is willing to assume of making an incorrect decision








What do you use to
feed a power curve?

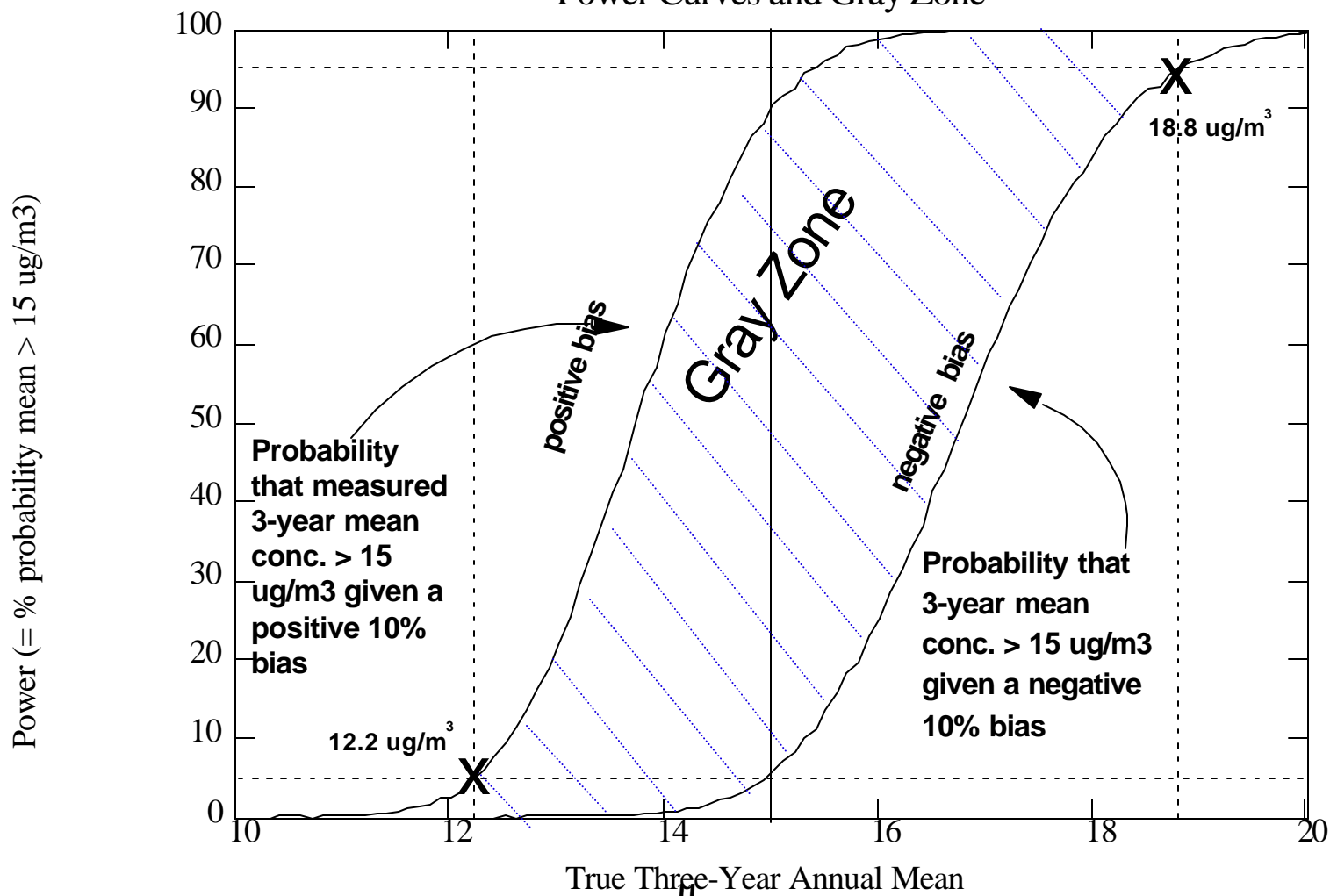


Parameters in Developing PM_{2.5} Mass DQOs- the Conservative Approach

<i>2001 Assumptions</i>	
3-Year Bias = \pm 10%	
3-Year Precision = 10%	
Annual NAAQS is controlling standard	
No spatial uncertainty and each monitor stands on its own (no spatial averaging)	
1 in 6 sampling with 75% completeness (144 days)	
3-year annual average is truth, (every day sampling and 100% comp.) up to bias and measurement variability	
Season ratio = 5.3	
Lognormal distribution for population variability, 80% CV	
Normal distribution of measurement uncertainty	
No auto correlation	
Decision errors 5%	

A Fed PM_{2.5} DQO Power Curve (based on conservative assumptions)

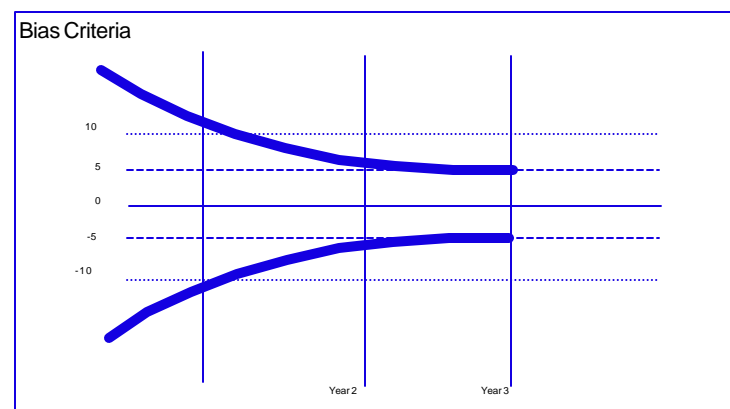
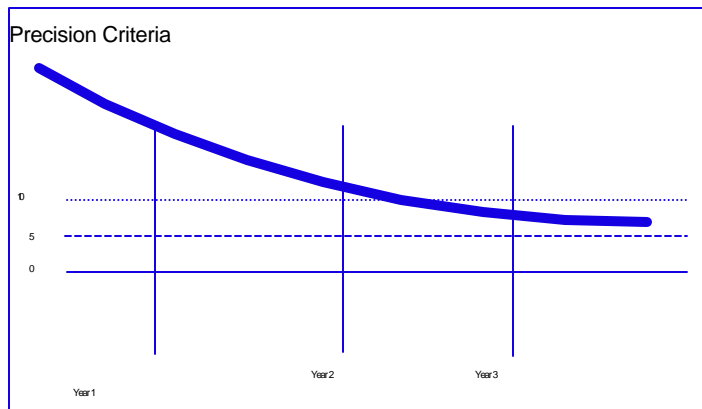
Power Curves and Gray Zone



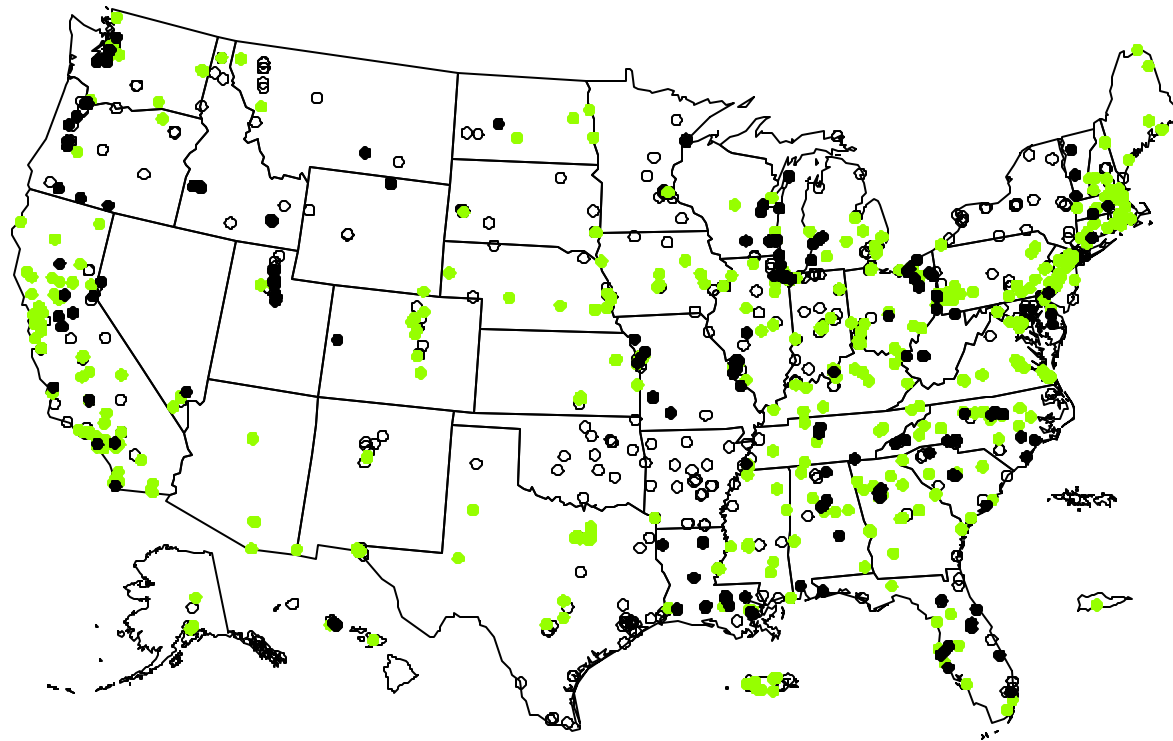
Data Quality Objective

Decision around the gray zone can be made with 95% confidence if:

- ✓ **Completeness** can be maintained at 75% or above
- ,
- ✓ **Precision** can be controlled to 10% CV, and
- ✓ **Bias** can be controlled to $\pm 10\%$

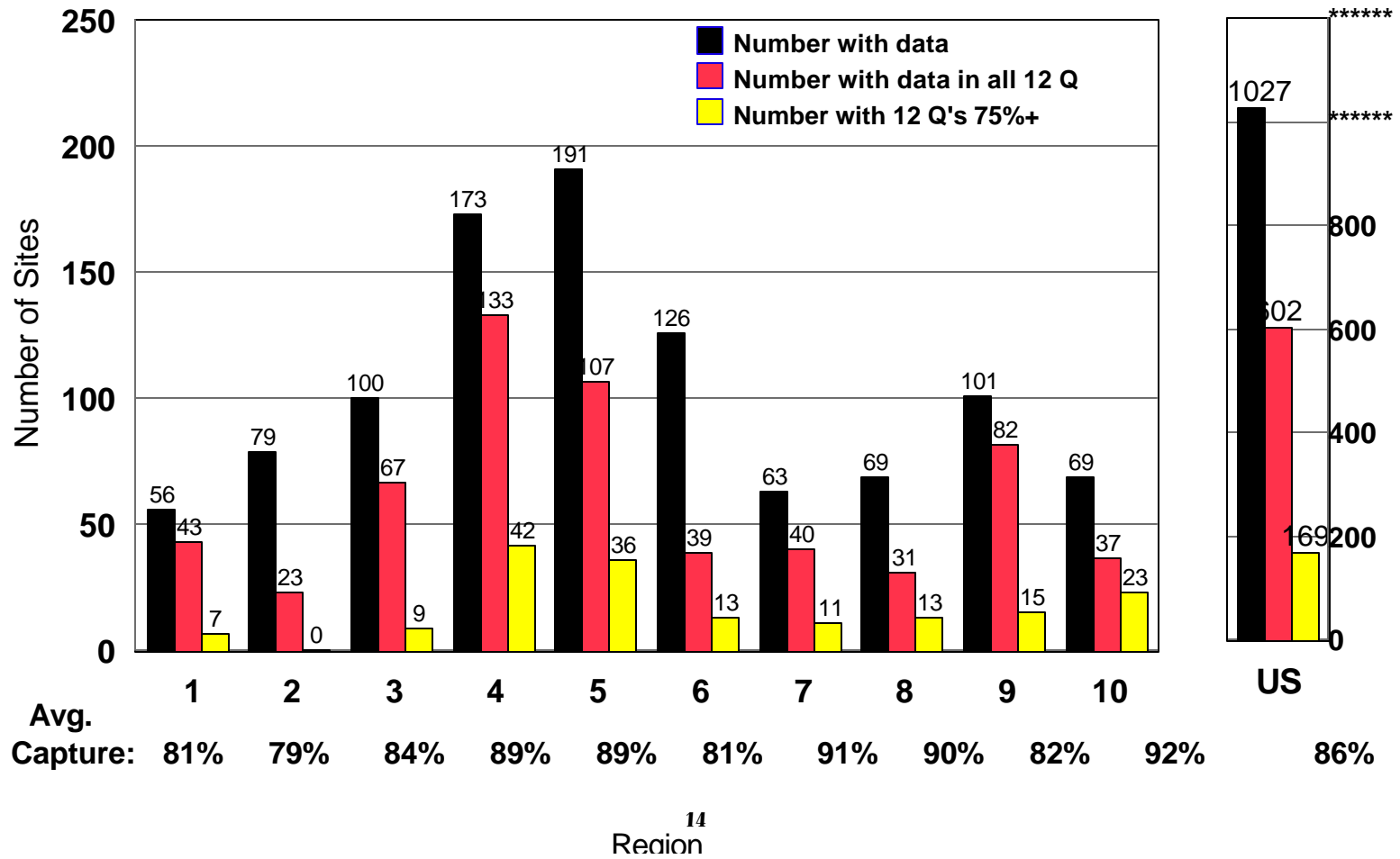


PM_{2.5} Completeness- Routine Data

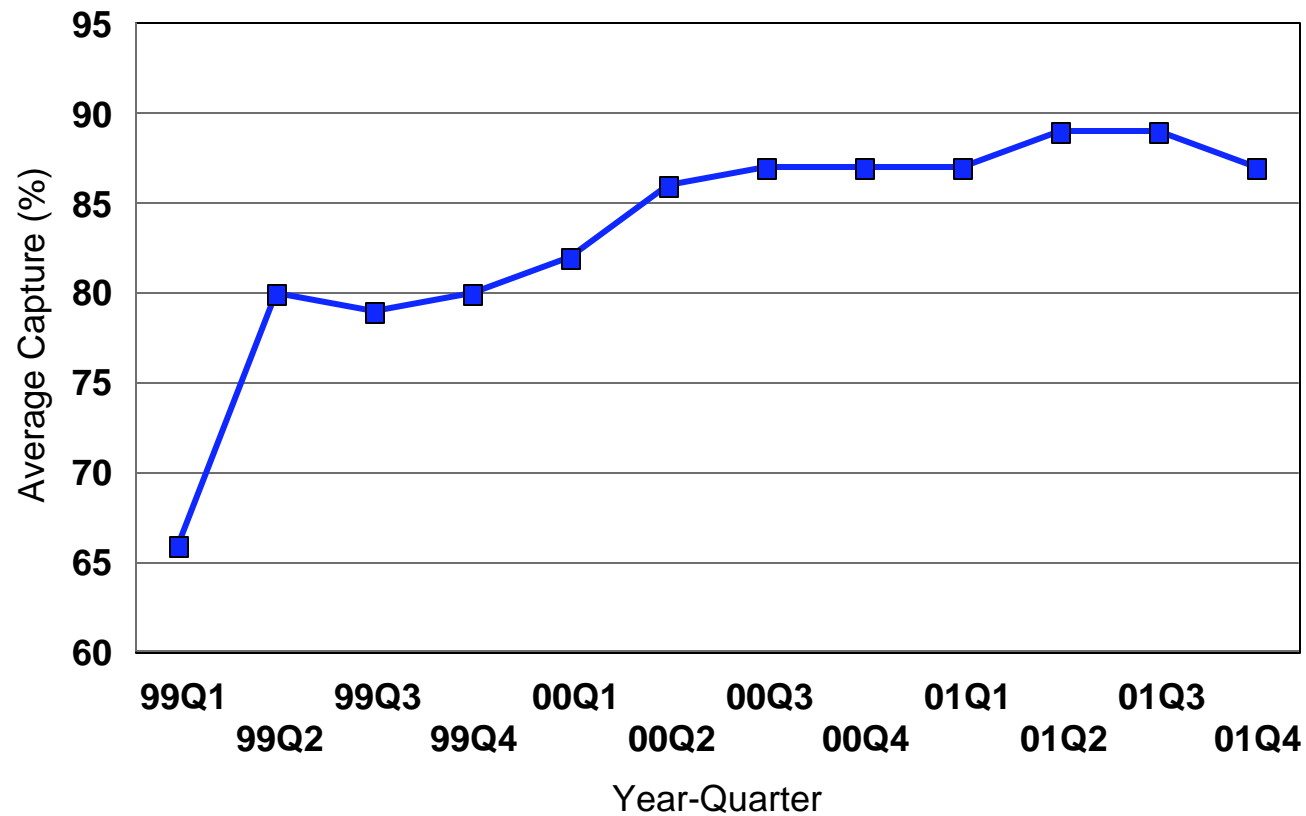


- 75%+ complete in all 12 quarters ('99-'01) [169]
- Data in all 12 quarters, but not 75% in all [433]
- Other sites with data [425]

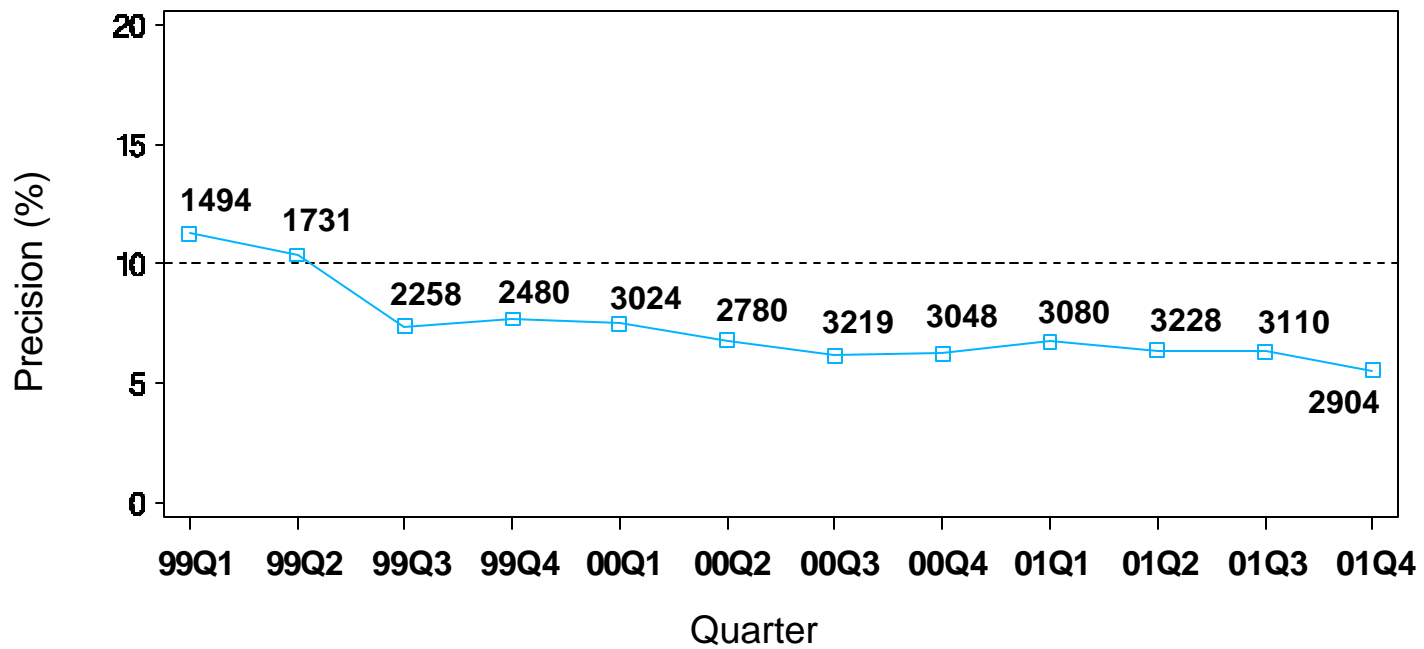
PM_{2.5} Completeness (Requirement) & Capture Rate (Performance)



Routine Data Completeness - Average Capture Rates

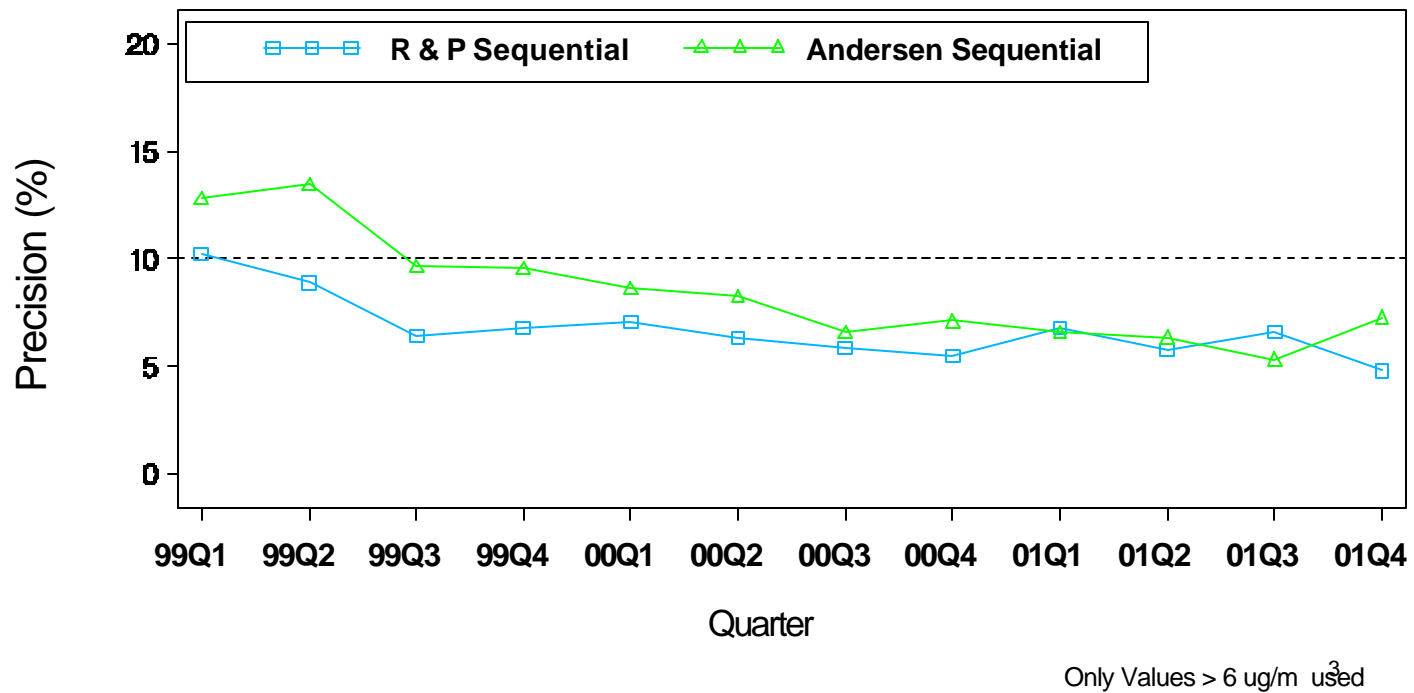


PM_{2.5} Precision- National Estimates

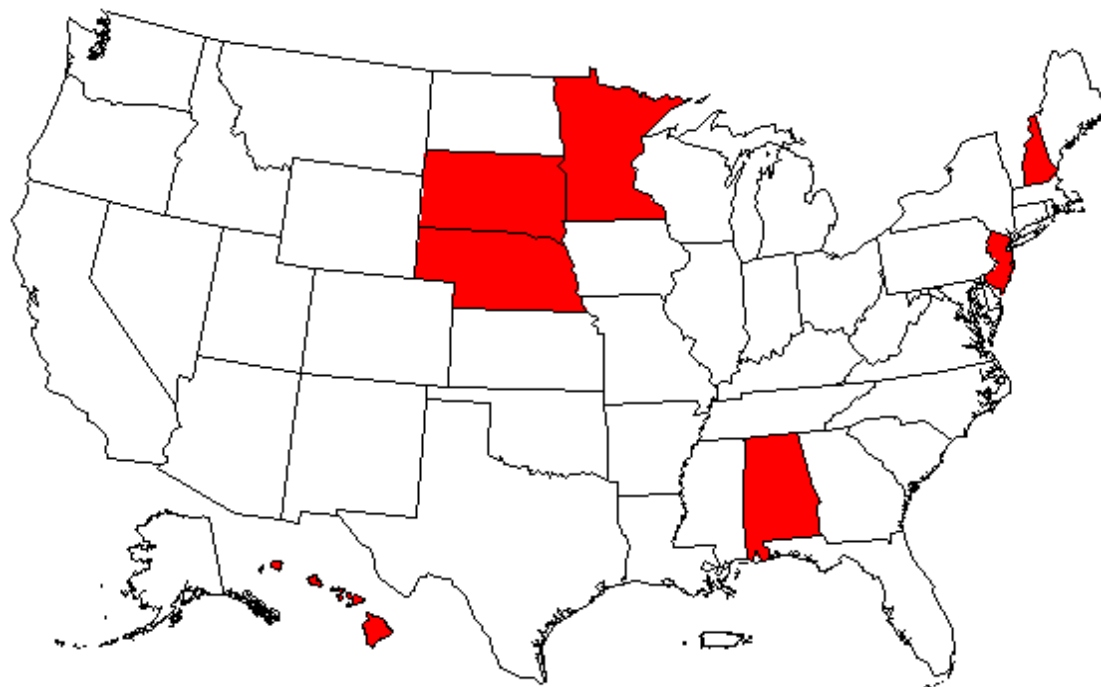




Points are labeled with the number of observations in each quarter
Only values > 6 ug/m³ used

PM_{2.5} Precision - Major Method Designations



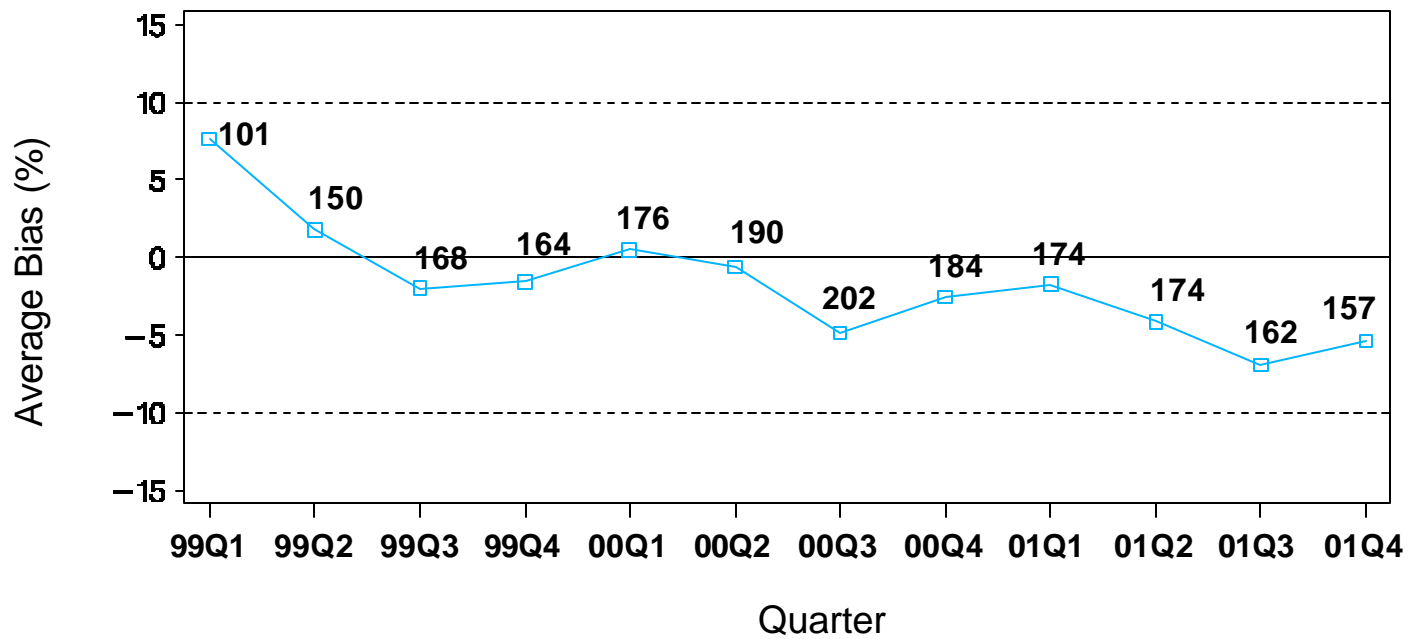
PM_{2.5} Precision- National Perspective



Precision (%)  < 10, Within DQO  > 10, Outside DQO

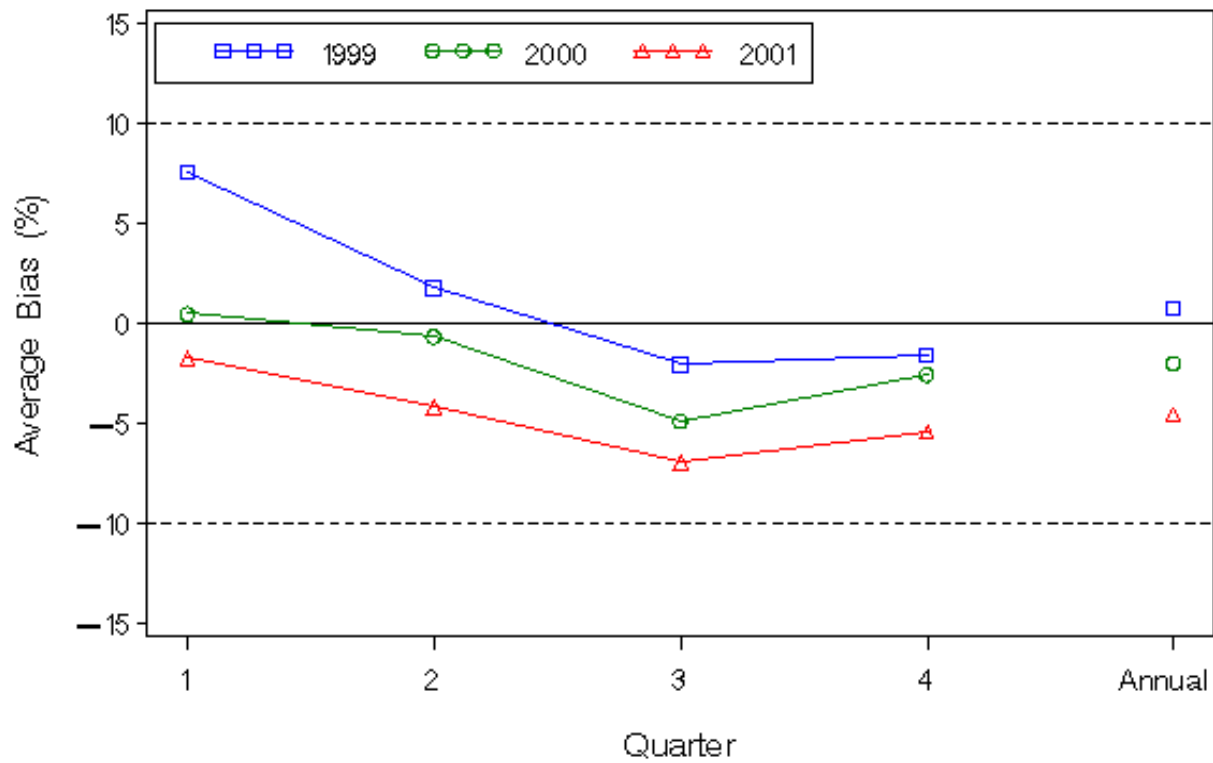
Aggregated over all Reporting Organizations within each state.
Only Values $> 6 \mu\text{g}/\text{m}^3$ Used.

PM_{2.5} Bias- National Estimates by Quarter

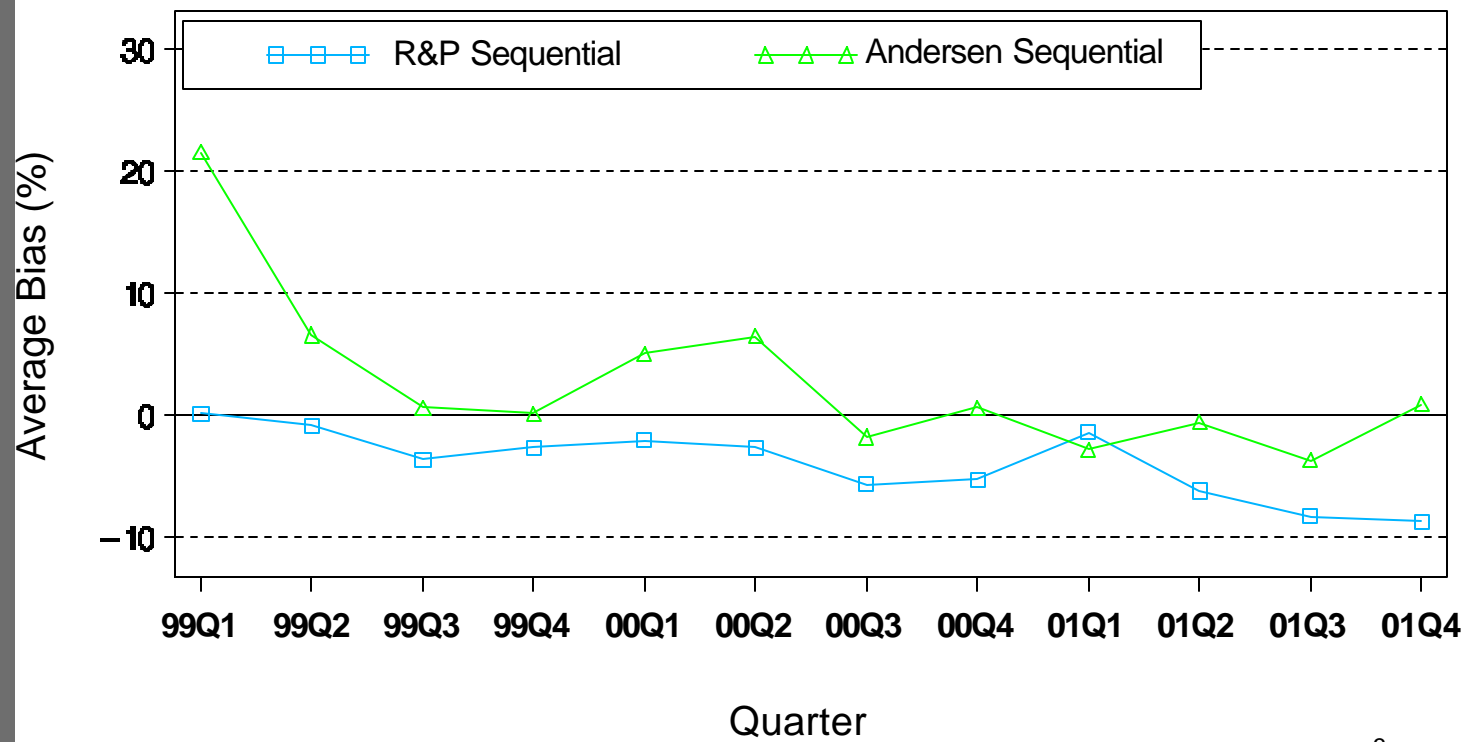


Points are labeled with the number of observations in each quarter
Only values > 6 ug/m³ used

PM_{2.5} Bias... A trend?

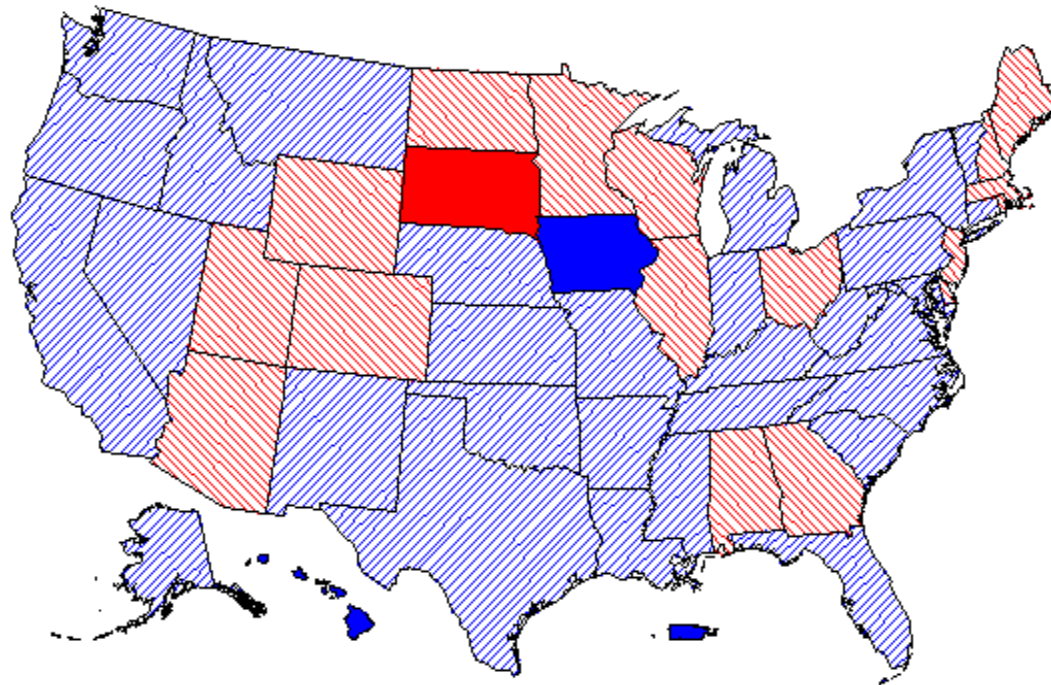


PM_{2.5} Bias by Major Method Designation







Only values > 6 ug/m³ used

PM_{2.5} Bias Estimates- National Perspective

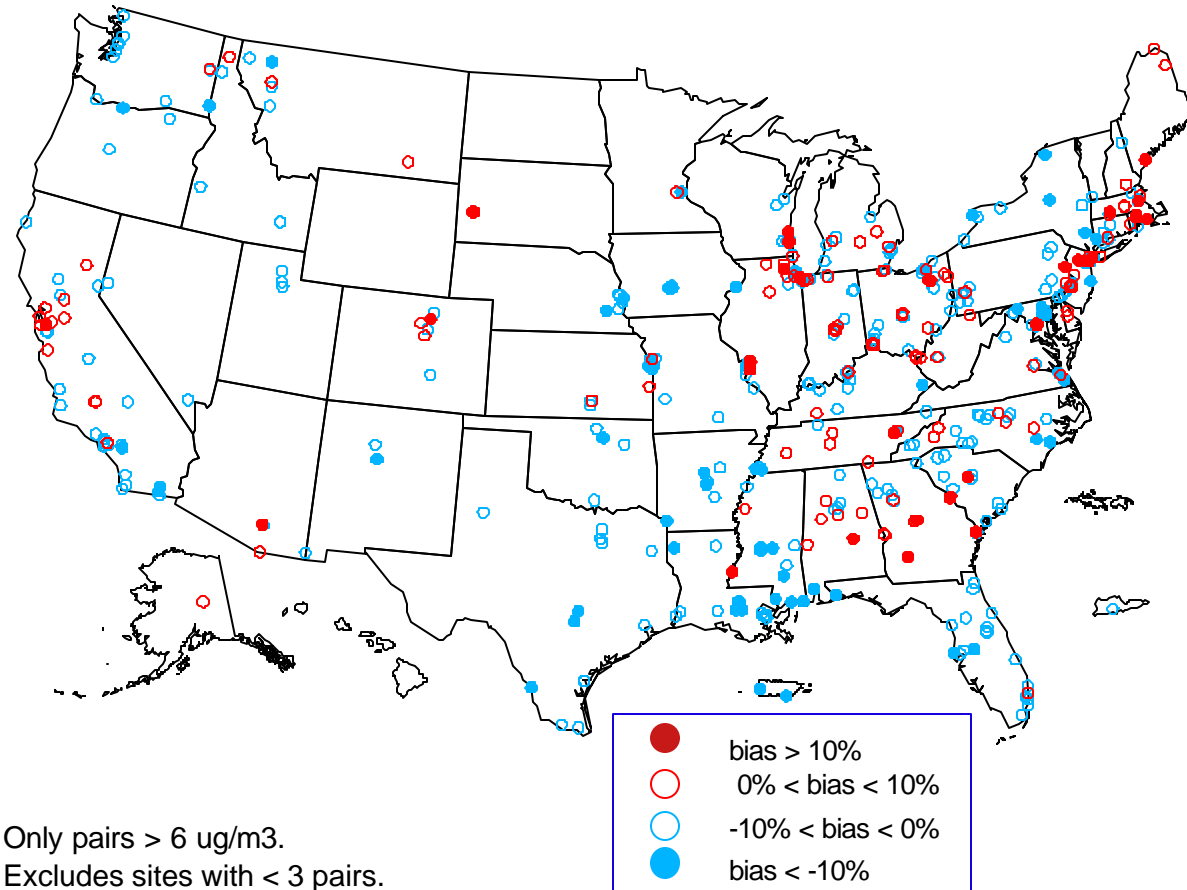


Average Bias (%)

 < -10, Outside DQO	 -10 to 0, Within DQO
 0 to 10, Within DQO	 > 10, Outside DQO

Aggregated over all Reporting Organizations within each state.
Only Values > 6 $\mu\text{g}/\text{m}^3$ Used.

PM_{2.5} Bias- Spatial Distribution of Site-Level 99-01





Well... What does the $PM_{2.5}$ data quality indicators tell us relative to the DQO? Can we feed the power curve?



Resulting DQOs for Annual NAAQS

- **Acceptable/achievable 3-yr average bias was 10% and 3-yr measurement precision was 10% CV.**
- **Associated gray zone is [12.2,18.8]. Recall this**
 - **is for comparison to annual NAAQS, and**
 - **is for one of the most extreme cases**
 - high seasonal ratio
 - high pop cv
 - 1-in-6 sampling with 75% completeness
- **Annual Standard Gray Zone**
 - **especially sensitive to: sampling frequency, bias, population variability, seasonal ratio**
 - **not sensitive to: measurement precision**

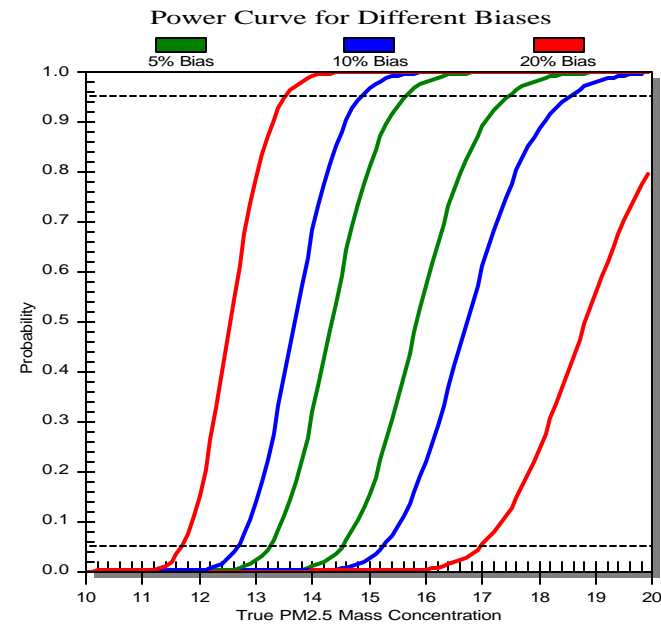
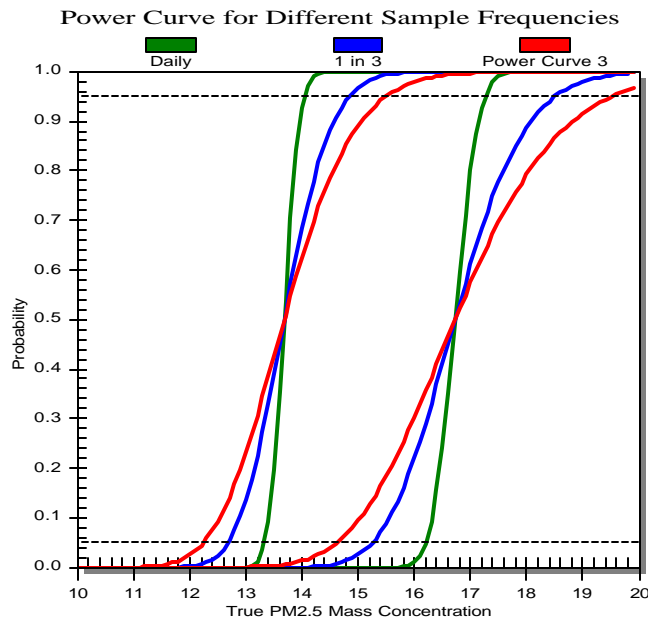
Examples of Sensitivity of Gray Zone

■ Sampling Frequency

- 1 in 6: [12.2,18.8]
- 1 in 3: [12.8,17.9]
- Daily: [13.5,17.1]

■ Bias

- 5% bias: [13.0,17.7]
- 10% bias: [12.2,18.8]
- 20% bias: [11.3,21.1]



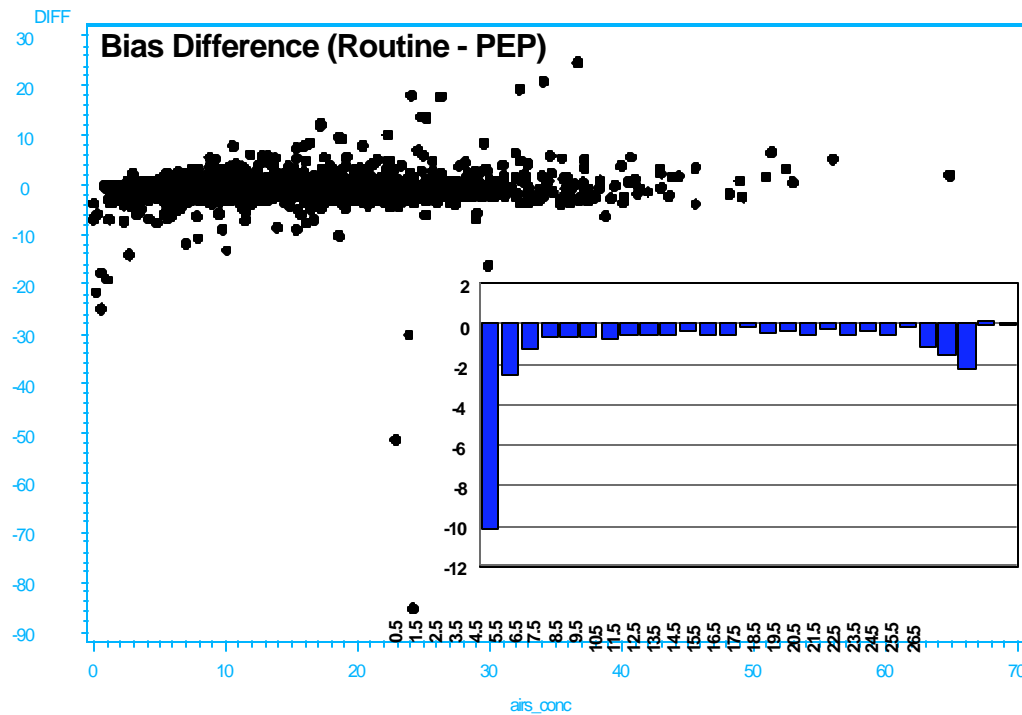
Next Steps

- **Develop DQO variables list at the Site Level**
 - available in QA Report
 - will provide 3-year performance as well as the last year (2001)
 - determines whether the site is within the DQO gray zone.
- **States can access DQO Software and plug their variables into the tool (<http://www.epa.gov/ttn/amtic/dqotool.html>)**

SITE LEVEL PARAMETERS FOR DQO TOOL and RESULTING GRAY ZONES														
-----Population Variables-----														
(not expected to change from 3-yr period to 3-yr period)														
99-'01 Estimates and Gray Zones														
"Future" 3-Year Estimates and Gray Zones (Based on)														
AIRS ID	Average Conc. (NOT DV)	Seasonal Ratio	Popn CV	Autocor-relation	Samp Freq	Completeness	Bias	Measurement CV	99-01 Gray Zone	Site Gray Zone within Goal Gray Zone?	2001 Samp Freq	2001 Completeness	2001 Bias	2001 Measurement CV
Site 1														
Site 2														
Site 3														

Next Steps (continued)

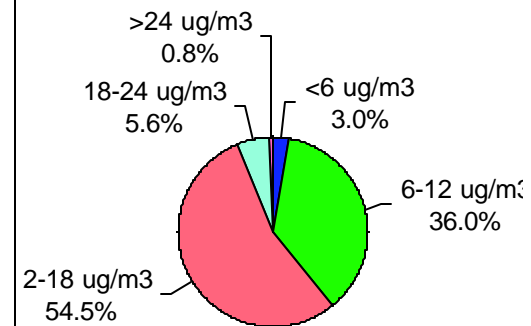
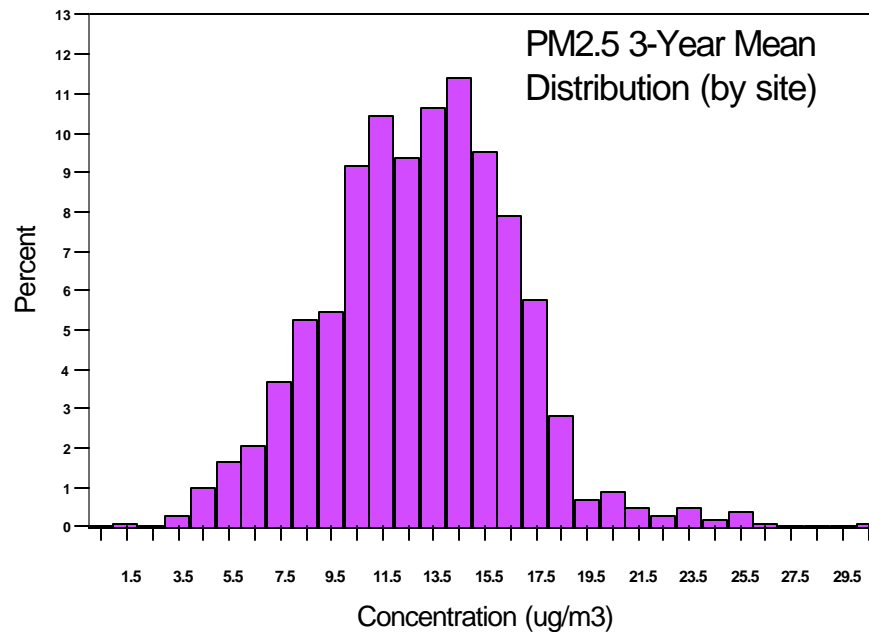
- Review and Revise Precision and Bias Statistics
 - May be able to keep data $< 6 \text{ ug/m}^3$



Next Steps (continued)

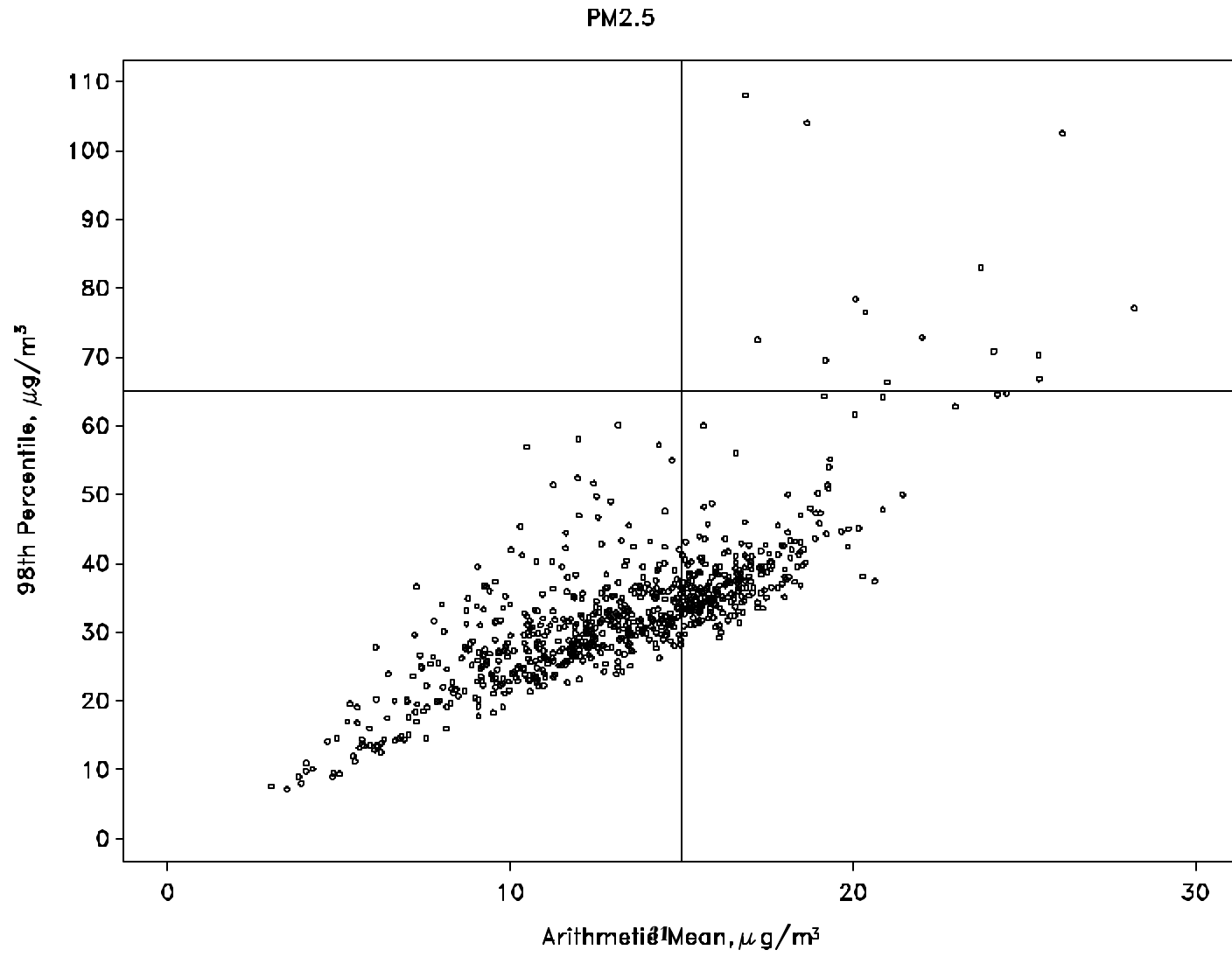
Pursue bias trend

- Work with State, Locals and Tribes
- Focus PEP around "important" sites
- Try to increase PEP completeness



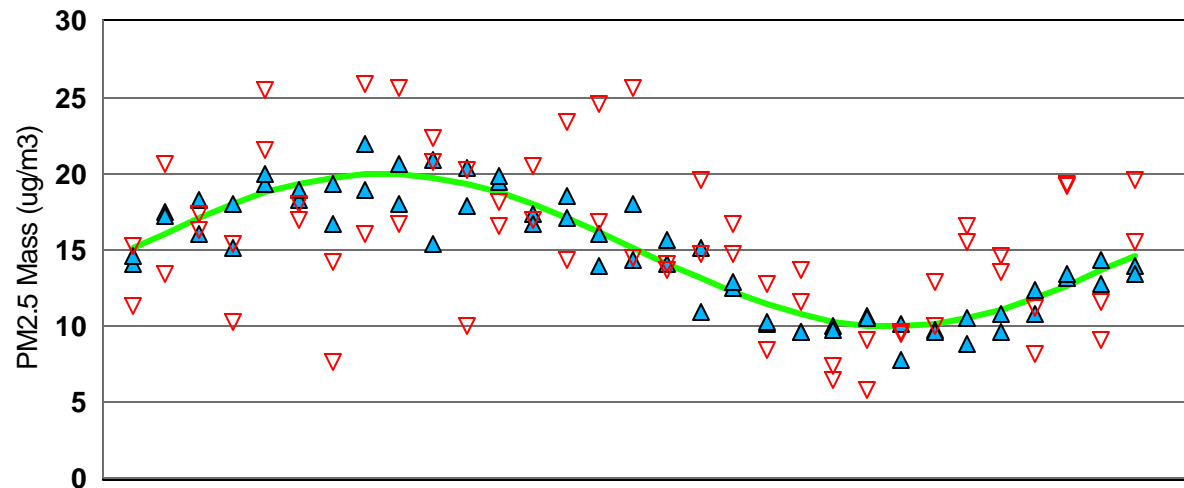
Supporting information for DQO Assumptions

The Annual Standard is the Controlling Standard



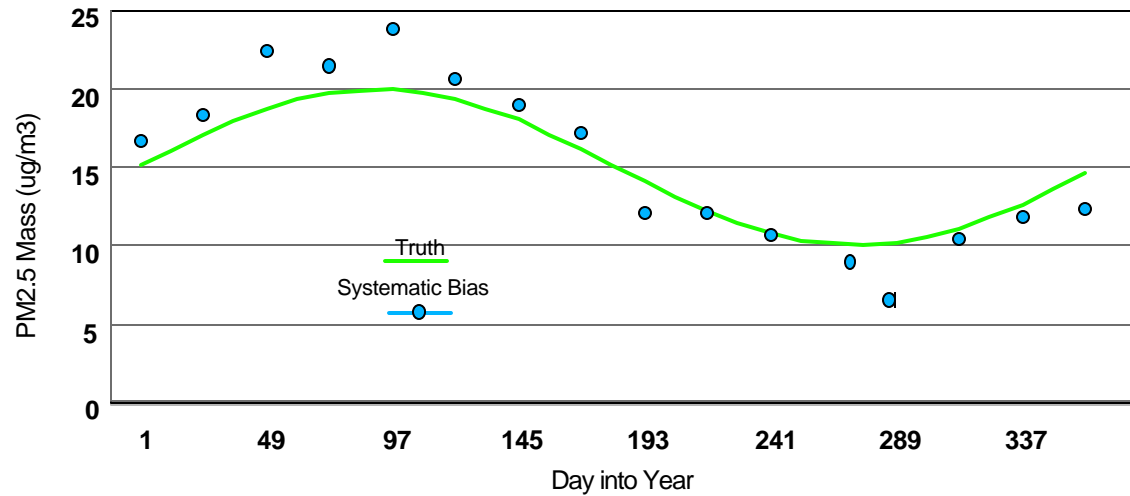
Terminology - Definition of Precision

- Precision - repeatability of a measurement system.
- Estimated using collocated instruments of same make.
 - 25% of sites in a reporting organization collocated. Sampled every 6 days
 - Precision based on 3 years of data at reporting organization level



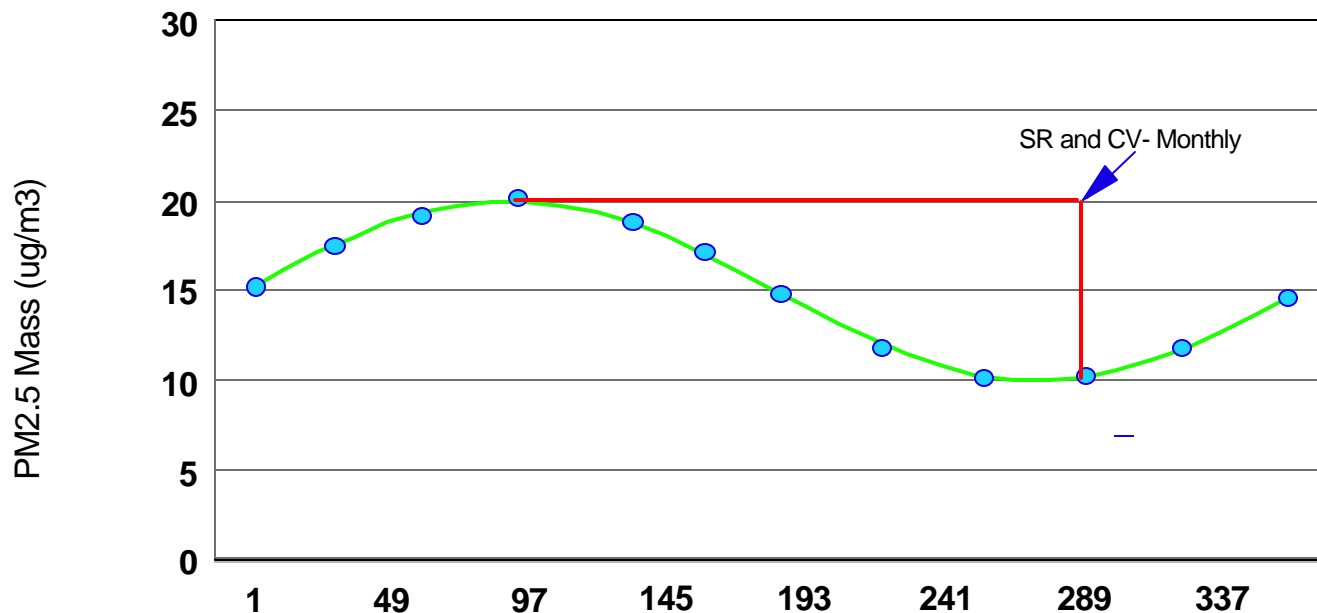
Terminology - Definition of Bias

- **Bias - deviation from "truth."**
- **Estimated using PEP $((FRM-PEP)/PEP)$.**
 - **25% of sites in a reporting organization collocated with PEP sampler 4 times a year**
 - **Bias based on 3 years of data at reporting organization level**



Terminology - Season Ratio & Population Variability (data set - sites with annual means between 10 -20 ug/m³)

- **Season Ratio-** ratio between high and low points on a curve on a monthly or bi-monthly basis
- **Population variability -** population variation about mean seasonal curve (CV) on a monthly or bi-monthly basis



Season Ratio and Population Variability

Distribution of ratios of highest to lowest monthly or bi-monthly mean at a site.

	Monthly	Bimonthly
# of sites	289	292
Mean	2.07	1.76
Percentiles		
Minimum	1.24	1.11
90.0	2.60	2.12
91.0	2.65	2.36
92.0	2.79	2.38
93.0	2.87	2.49
94.0	3.01	2.57
95.0	3.70	3.17
96.0	4.41	3.36
97.0	4.61	3.90
98.0	5.25	4.03
99.0	6.05	4.69
Maximum	6.54	4.89

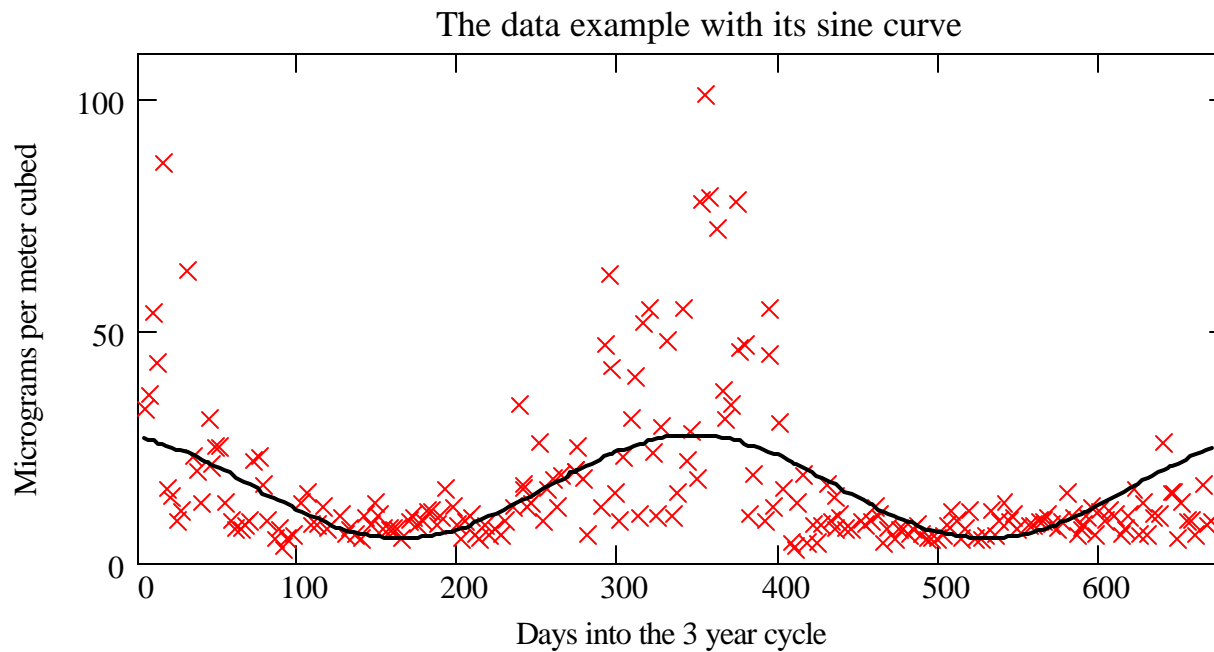
Distribution of CVs about monthly and bimonthly means

	Monthly	Bimonthly
# of sites	3,398	1,752
Mean	49.6	50.7
Percentiles		
Minimum	16.1	22.9
10	34.6	37.6
25.0	40.4	42.8
50.0	48.1	49.4
75.0	56.3	56.9
90.0	66.6	64.7
95.0	73.7	70.5
96.0	75.4	72.3
97.0	78.2	75.9
98.0	83.8	79.1
99.0	93.5	89.8

Season Ratio of 5.3 and Pop. CV of 80% chosen
(conservative but realistic)

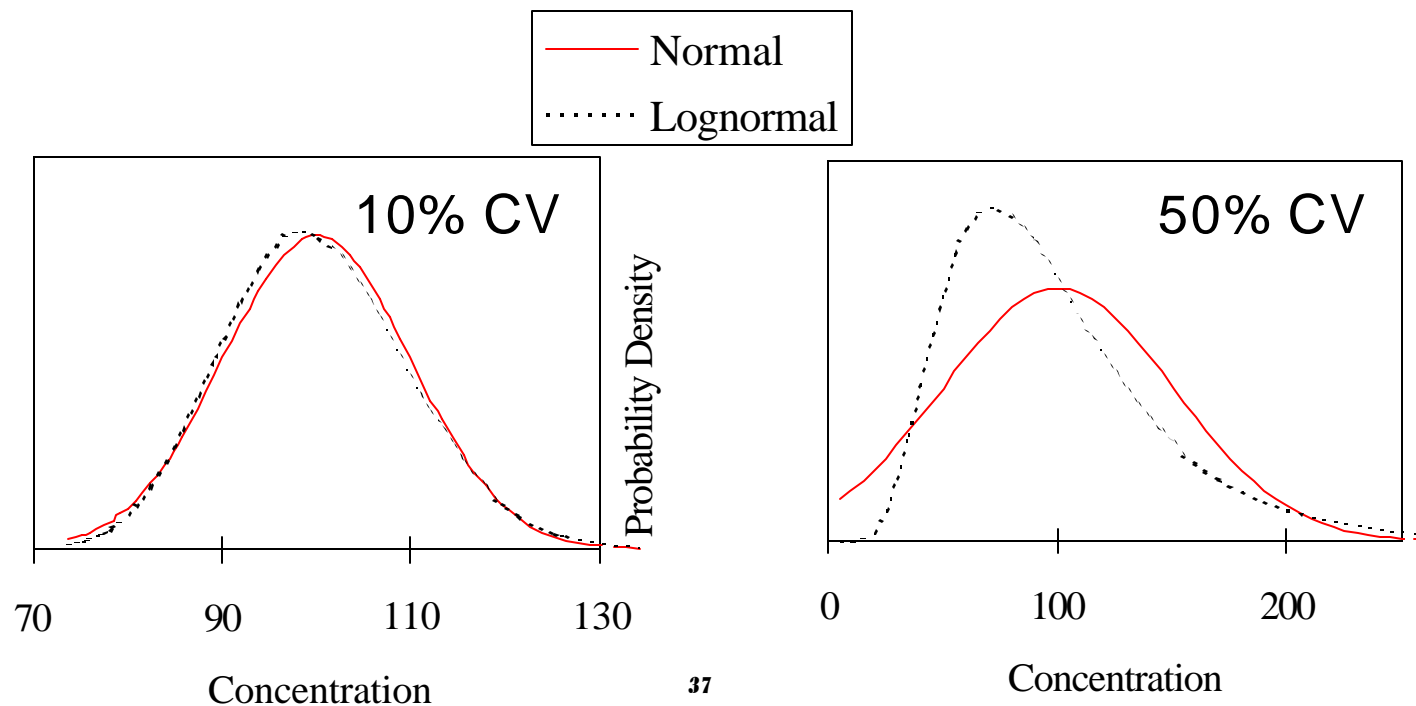
Normal vs Lognormal Distribution Around Sinusoidal Curve

Normal distribution with 80% pop. CV would result in about 10% negative values



Normal Distribution of Measurement Uncertainty

- Current PM2.5 precision estimates (CY99, 00, 01) are ~ 8% CV
- Normal and lognormal measurement uncertainty very similar at lower CV's
- Therefore; normal distribution assumption is appropriate.



Auto Correlation

- **How well 1 day can predict (correlates to) the next**
 - **There is auto correlation during everyday sampling**
 - **Since the DQO set at 1 in 6 day sampling auto correlation set to 0**

Now that we have all these #@*!assumptions
how do we use them?