

# PM<sub>2.5</sub> Design Values

- Data inputs
- Completeness requirements
- Calculations
- Current levels

Mark Schmidt

EPA-OAQPS

[Schmidt.mark@epa.gov](mailto:Schmidt.mark@epa.gov)

(919) 541-2416

# PM<sub>2.5</sub> Design Values - Data Inputs

- Only data present in EPA's Air Quality System (AQS) are used in DV calculations.
- Only Federal Reference Method (FRM), Federal Equivalent Method (FEM), and Approved Regional Method (ARM) PM<sub>2.5</sub> data are used.
  - Duration '7' for filter (FRM/FEM) samplers.
  - Duration 'X' for continuous (FEM/ARM) samplers.
    - From 'NAAQS averages' AQS table.
    - Valid if 18+ hours or pass data substitution test (using zeros for missing hours).
- DVs are calculated for all monitors / all monitor types. However,
  - Only DVs for population-oriented monitors are compared to the NAAQS. [All monitors are presumed to be population-oriented unless otherwise specified / notified.]
  - SPMs with less than two years of data are weeded out with completeness requirements
  - Monitors that are micro/middle scale and represent a unique area can be exempted from comparisons to the annual standard.
    - Must be documented in monitoring network plan

# PM<sub>2.5</sub> Design Values - Data Inputs

*continued*

- PM<sub>2.5</sub> DVs are calculated on a site basis. Site data area aggregated as follows:
  - The default dataset for a site consists of the measured concentrations recorded from the designated primary FRM/FEM/ARM monitor. The primary monitor should be designated in the appropriate monitoring network plan. [Until a “primary” designation is made, the lowest POC # with data is the presumed primary monitor.] All daily values produced by the primary sampler are considered part of the site record (i.e., that site's daily value).
  - Data for the primary monitor are augmented as much as possible with data from collocated FRM/FEM/ARM monitors. If a valid 24-hour measurement is not produced from the primary monitor for a particular day (scheduled or otherwise), but a valid sample is generated by a collocated FRM/FEM/ARM instrument (and recorded in AQS), then that collocated value shall be considered part of the site data record (i.e., that site's daily value). If a valid 24-hour measurement is not produced from the primary monitor for a particular day and more than one valid collocated FRM/FEM/ARM value is available, the average of those valid collocated values shall be used as the daily value.

# PM<sub>2.5</sub> Design Values - Data Inputs

*continued*

## Creation of composite site record - Example

- In this example, there are 3 FRM/FEM monitors collocated at one site.
- On days where there is a measurement from the designated primary monitor (Jan. 1, Jan. 2, Jan. 4, Jan. 7, Jan. 13, Jan. 16), that measurement alone is considered the daily site value (i.e., part of the composite site record).
- On days in which there is not a primary measurement, but there is one collocated measurement (Jan. 11, Jan. 12), that collocated value is considered part of the site record.
- On days in which there is not a primary measurement, but there are two collocated measurements (Jan. 10), the rounded average of the collocated values is considered the site's daily value.

Date	Required sampling frequency	Daily value for the designated primary monitor (µg/m <sup>3</sup> )	Daily value for a collocated FRM/FEM monitor (µg/m <sup>3</sup> )	Daily value for a 2nd collocated FRM/FEM monitor (µg/m <sup>3</sup> )	Site daily value (µg/m <sup>3</sup> )	Data source
<b>1-Jan</b>	3	13.6			13.6	primary
2-Jan	3	15.1			15.1	primary
3-Jan	3					
<b>4-Jan</b>	3	10.2	10.5	10.6	10.2	primary
5-Jan	3					
6-Jan	3					
<b>7-Jan</b>	3	16.5			16.5	primary
8-Jan	3					
9-Jan	3					
<b>10-Jan</b>	3		36.5	35.8	36.2	<b>avg. of 2 collocated</b>
11-Jan	3			37.5	37.5	<b>collocated</b>
12-Jan	3		36.5		36.5	<b>collocated</b>
<b>13-Jan</b>	3	28.4			28.4	primary
14-Jan	3					
15-Jan	3					
<b>16-Jan</b>	3	12.5	12.4	13.1	12.5	primary
17-Jan	3					

Note: Scheduled sampling days are bolded. However, when constructing the site record, sampling frequency and scheduled sampling days are irrelevant.

# PM<sub>2.5</sub> Design Values - Data Inputs *continued*

## Creation of composite site record

- Procedure specified in (10/2006) CFR.
- AQS is being enhanced to generate composite site record data (24-hour and summary)

# PM<sub>2.5</sub> DVs – Completeness Requirements

- Data completeness is determined on a site basis ...using the composite site record
- Data capture rates are computed on a site-year-quarter. "Quarter" is a calendar quarter. Quarterly data capture rates (expressed as a percentage) are calculated as: ((credible number of samples for the quarter) ÷ (scheduled number of samples for the quarter))\* 100, and rounded to the nearest integer.
- "Creditable samples" are summed by site-year-quarter. Creditable samples are samples that are given credit for data completeness. They include samples collected on required sampling days and valid "make-up" samples taken for missed or invalidated scheduled samples. Make-ups can be made by either the primary or the collocated instruments. The composite site record is developed before potential make-ups are validated (as credible samples).
- "Total samples" are summed by site-year-quarter and also by site-year. Total samples include credible samples plus "extra" samples. Extra samples are reported samples in the composite site record that are not deemed "credible".
- For sites that sample every day, the credible number of samples will always be the same as the total number of samples. Also, sites that sample every day can not have make-ups since every day is a scheduled day.

# PM<sub>2.5</sub> DVs – Completeness Requirements

*continued*

- **The data completeness requirements for the annual PM<sub>2.5</sub> NAAQS are:**
  - For single site comparisons to the NAAQS, three valid annual means are required.
  - For spatial averaging, three valid spatially averaged annual means are required. A spatially averaged annual mean encompasses all sites (of those to be averaged) with a valid annual (site) mean. Thus, for each of the three years, at least one site must have a valid annual (site) mean.
  - A site's annual mean is deemed valid ("complete") if it meets one of the following:
    - (CFR) 75% or greater data capture for each quarter, or ...
    - (CFR) 11 or more samples (*total number*) per quarter and the annual mean (or corresponding spatially averaged annual mean) is > 15.0 µg/m<sup>3</sup>, or ...
    - (CFR) 11 or more samples (*total number*) per quarter and the annual design value is > 15.0 µg/m<sup>3</sup>, or ...
    - (CFR) less than 11 (*total number*) per quarter and pass the "over NAAQS" data substitution test, or ...
    - (guidance) 50% or greater data capture each quarter and pass one of two "under NAAQS" data substitution tests.
  
- **The data completeness requirements for the 24-hour PM<sub>2.5</sub> NAAQS are:**
  - A site must have three valid annual 98th percentile values.
  - A site's annual 98th percentile value is deemed valid ("complete") if it meets one of the following:
    - (CFR) 75% or greater data capture for each quarter, or ...
    - (CFR) 1 or more samples for the year and the annual 98th percentile is > 35 µg/m<sup>3</sup>, or ...
    - (CFR) 1 or more samples for the year and the 24-hour design value is > 35 µg/m<sup>3</sup>, or ...
    - \*(guidance) 50% or greater data capture each quarter and pass one of two "under NAAQS" data substitution tests.

# PM<sub>2.5</sub> DVs – Annual DV Calculations

## Procedure for calculating single-site annual DVs (not spatially averaged)

1. Calculate the quarterly means for the site. Add all of the 24-hour sample concentrations; then, divide by the number of samples (total number, not creditable number). Do not round quarterly means!
2. Calculate the annual means from the quarterly means. For each year, add the four (or less) quarterly means; then, divide by 4 (or the number of quarters with data). Do not round annual means!
3. Calculate the (unrounded) 3-year average annual mean. Add the three (or less) annual averages and then divide by 3 (or the numbers of years with data)
4. Round to 1 decimal place.
5. Determine validity of design value and compare to the standard. The design value is valid if all three years (annual means) meet completeness criteria



# PM<sub>2.5</sub> DVs – 24-hour DV Calculations

- The 24-hour standard DV is simply a rounded average of three annual 98th percentile values. Annual 98th percentile values are actual, discreet, reported values; they are not statistically interpolated estimates. There are two prescribed procedures in Appendix N for calculating annual 98th percentile values: 1) the “regular” procedure, which will be used by most sites, and 2) a special formula for sites that operate on an approved seasonal sampling frequency. Once you have identified your three annual 98th percentile values (using the appropriate technique(s)), follow the steps below:
  1. Calculate the (unrounded) 3-year average of the (three) annual 98th percentile values. Add the 3 annual 98<sup>th</sup> percentile values and divide the sum by 3.
  2. Round to the nearest integer. The result is the 24-hour standard design value.
  3. Determine validity of the 24-hour standard design value and compare to the standard. A 24-hour standard design value is valid if the three constituent annual 98th percentile values are valid.

# PM<sub>2.5</sub> DVs – 24-hour DV Calculations

*continued*

## Procedure for identifying “regular” (non-seasonal) annual 98<sup>th</sup> percentile values

- *Appendix N (10/2006 version ) prescribes a particular procedure and formula (i.e., Equation 5) for identifying “regular” annual 98<sup>th</sup> percentile values. However, there is an inadvertent omission in the procedure/formula which would cause the results (the annual 98<sup>th</sup> percentiles) to be calculated erroneously low whenever the annual creditable number of samples is not equal to the annual total number of samples (i.e., when there are extra samples in the year). OAQPS is in the process of correcting this rule error. The correction will involve the replacement of the currently-stated statistical formula and instructions with a simpler look-up table approach. The table look-up method is not new; it was described in the previous PM data handling guidance as an acceptable alternative to the previous CFR annual 98<sup>th</sup> percentile formula (1997 version).*
- Identification of annual 98<sup>th</sup> percentile values using the regular procedure will be based on the *creditable number of samples* (as described below), rather than on the actual (*total*) number of *samples*. Extra samples, however, are candidates for selection as the annual 98<sup>th</sup> percentile. [The creditable number of samples will determine how deep to go into the data distribution, but all samples (creditable and extra) will be considered when making the percentile assignment.] The annual creditable number of samples is the sum of the four quarterly creditable number of samples.

# PM<sub>2.5</sub> DVs – 24-hour DV Calculations

*continued*

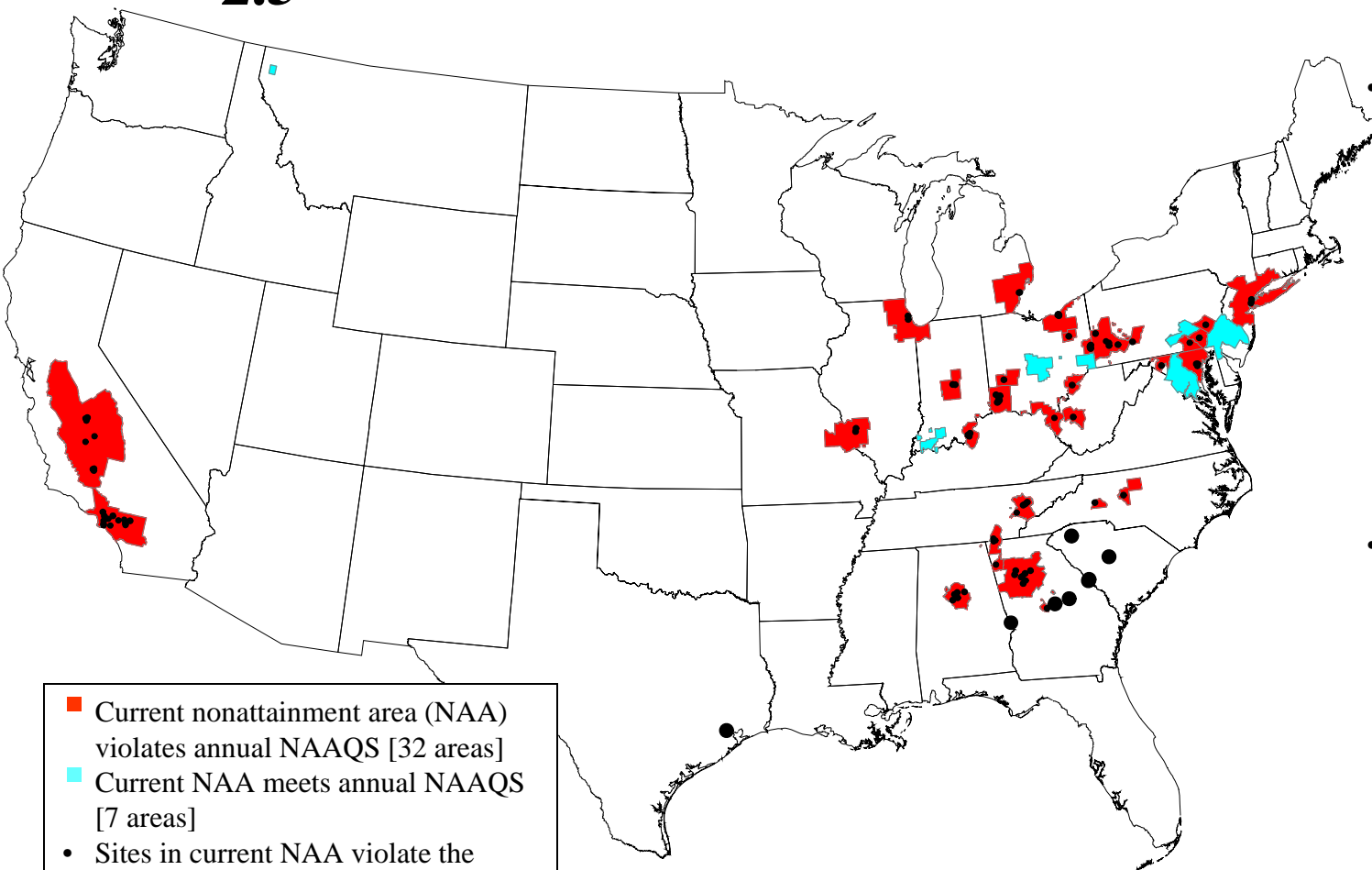
## Procedure for identifying “regular” (non-seasonal) annual 98<sup>th</sup> percentile values

- Sort/array all the daily values from a particular site and year by descending value. (For example: (x[1], x[2], x[3], ..., x[n]). In this case, x[1] is the largest number and x[n] is the smallest value.) The 98th percentile is determined from this sorted series of daily values which is ordered from the highest to the lowest number. [Since an annual 98<sup>th</sup> percentile will never be further than 8th on such a list (i.e., no lower in concentration than the 8th highest value), a shortened list (e.g., top 8) is sufficient.]
- The 98th percentile value for year y,  $P_{0.98, y}$ , is identified via the table below. Using the left column of the table, determine the appropriate range (i.e., row) for the annual *creditable number of samples* for year y ( $cn_y$ ). The corresponding “n” value in the right column identifies the rank of the annual 98th percentile value in the descending sorted list of daily site values for year y. Thus,  $P_{0.98, y} =$  the n<sup>th</sup> largest value.

**98th percentile look-up table**

Creditable number of samples for year "y" ( $cn_y$ )	$P_{0.98, y}$ is the n <sup>th</sup> maximum value for the year, where n is the listed number
1 - 50	1
51 - 100	2
101 - 150	3
151 - 200	4
201 - 250	5
251 - 300	6
300 - 350	7
351 - 366	8

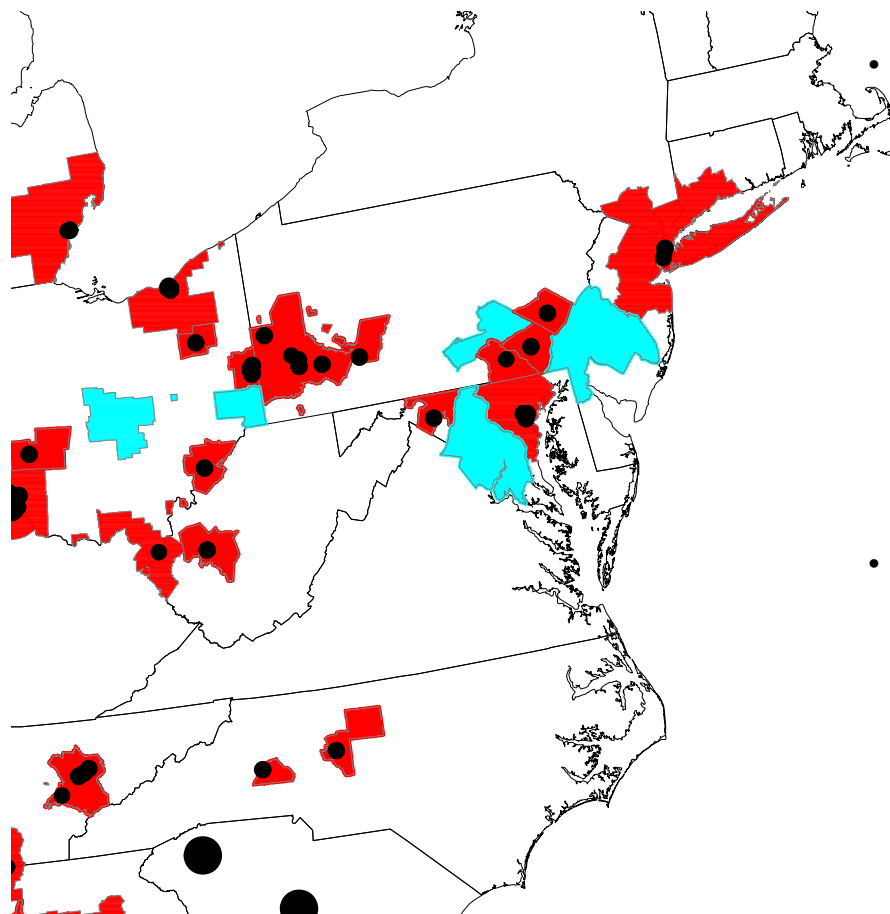
# PM<sub>2.5</sub> DV Levels – 2004-2006 Annual Std



■ Current nonattainment area (NAA) violates annual NAAQS [32 areas]  
■ Current NAA meets annual NAAQS [7 areas]  
 • Sites in current NAA violate the annual NAAQS (94 sites)  
 • Sites not in a current NAA violate the annual NAAQS (8 sites)

- 7 NAA clean for annual NAAQS: Libby, Columbus (OH), Evansville, Harrisburg, Philadelphia, Washington, Wheeling, but ...
  - Philadelphia has designated NA sites that are incomplete.
  - Only Evansville and Wheeling are also clean for the 24-hr NAAQS
- 8 additional violating sites (not in NAA) are located in 7 areas:
  - Greenville, SC (Unclassifiable area); Augusta, GA
  - Columbia, SC
  - Columbus, GA
  - Houston, TX
  - Washington County, GA
  - Wilkinson County, GA

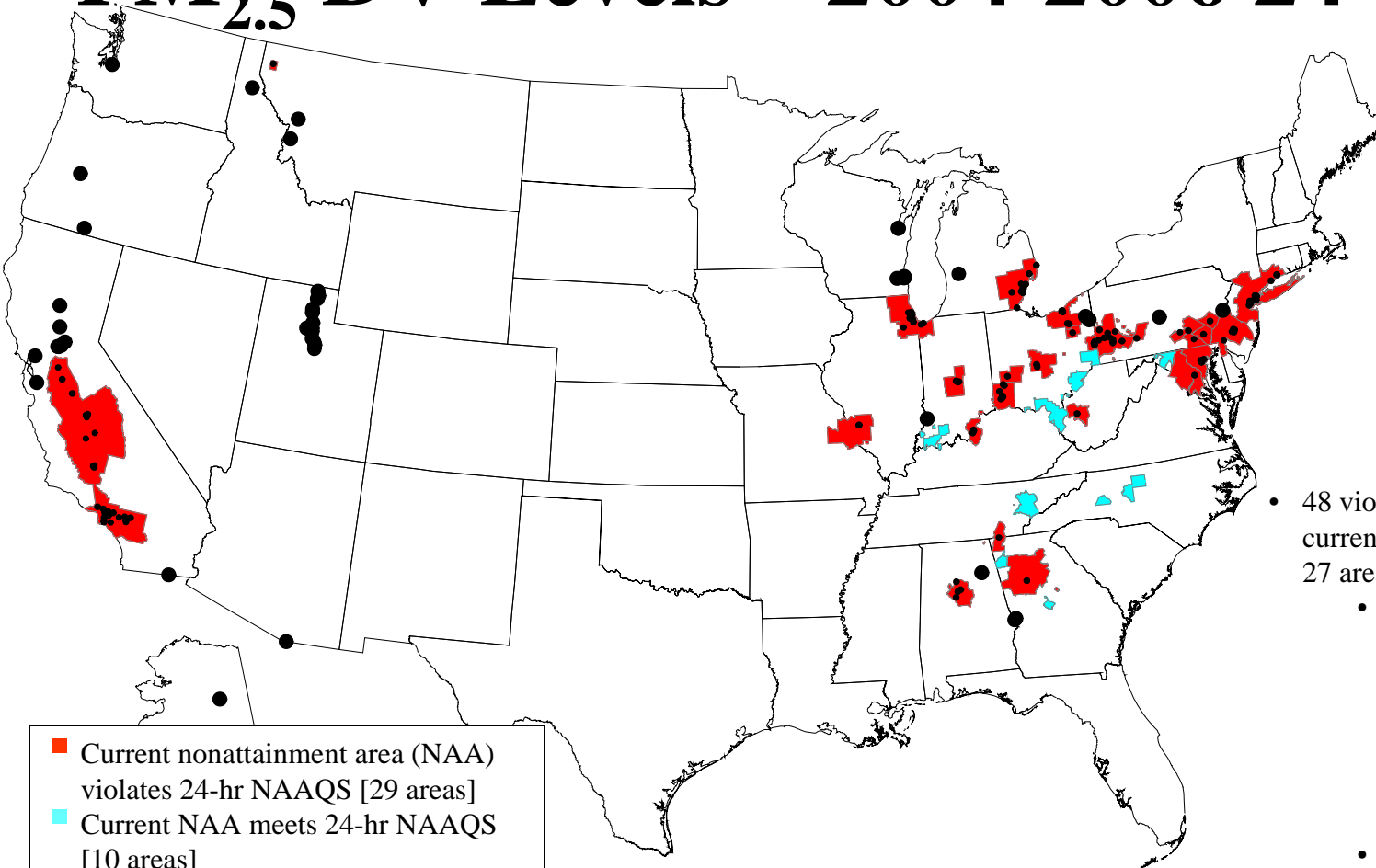
# PM<sub>2.5</sub> DV Levels – 2004-2006 Annual Std



- Current nonattainment area (NAA) violates annual NAAQS [32 areas]
- Current NAA meets annual NAAQS [7 areas]
- Sites in current NAA violate the annual NAAQS (94 sites)
- Sites not in a current NAA violate the annual NAAQS (8 sites)

- 7 NAA clean for annual NAAQS: Libby, Columbus (OH), Evansville, Harrisburg, Philadelphia, Washington, Wheeling, but ...
  - Philadelphia has designated NA sites that are incomplete.
  - Only Evansville and Wheeling are also clean for the 24-hr NAAQS
- 8 additional violating sites (not in NAA) are located in 7 areas:
  - Greenville, SC (Unclassifiable area); Augusta, GA
  - Columbia, SC
  - Columbus, GA
  - Houston, TX
  - Washington County, GA
  - Wilkinson County, GA

# PM<sub>2.5</sub> DV Levels – 2004-2006 24-hour Std

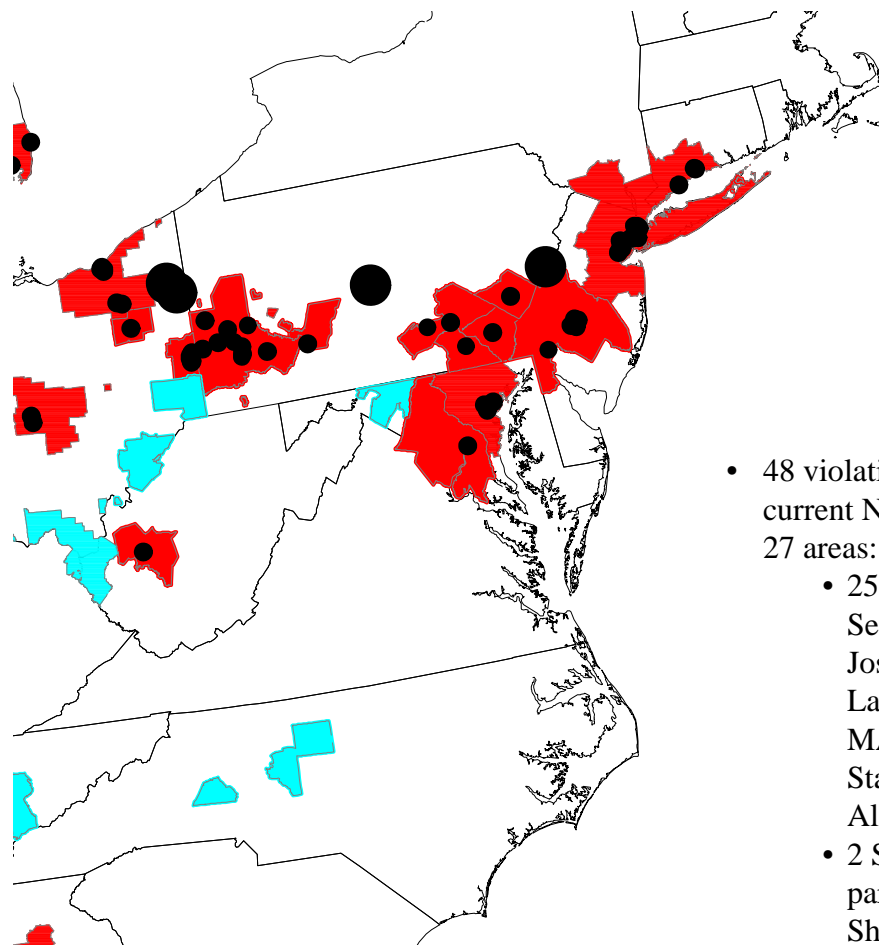


■ Current nonattainment area (NAA) violates 24-hr NAAQS [29 areas]  
■ Current NAA meets 24-hr NAAQS [10 areas]

- Sites in current NAA violate the 24-hr NAAQS (131 sites)
- Sites not in a current NAA violate the 24-hr NAAQS (48 sites)

- 48 violating sites outside of current NAA are located in 27 areas:
  - 25 CBSAs: Largest are Seattle, Sacramento, San Jose, Milwaukee, Salt Lake City. Two in MARAMA region are State College, PA and Allentown, PA.
  - 2 State-Counties (not part of a CBSA): Shoshone ID; Ravalli MT

# PM<sub>2.5</sub> DV Levels – 2004-2006 24-hour Std



- Current nonattainment area (NAA) violates 24-hr NAAQS [29 areas]
- Current NAA meets 24-hr NAAQS [10 areas]

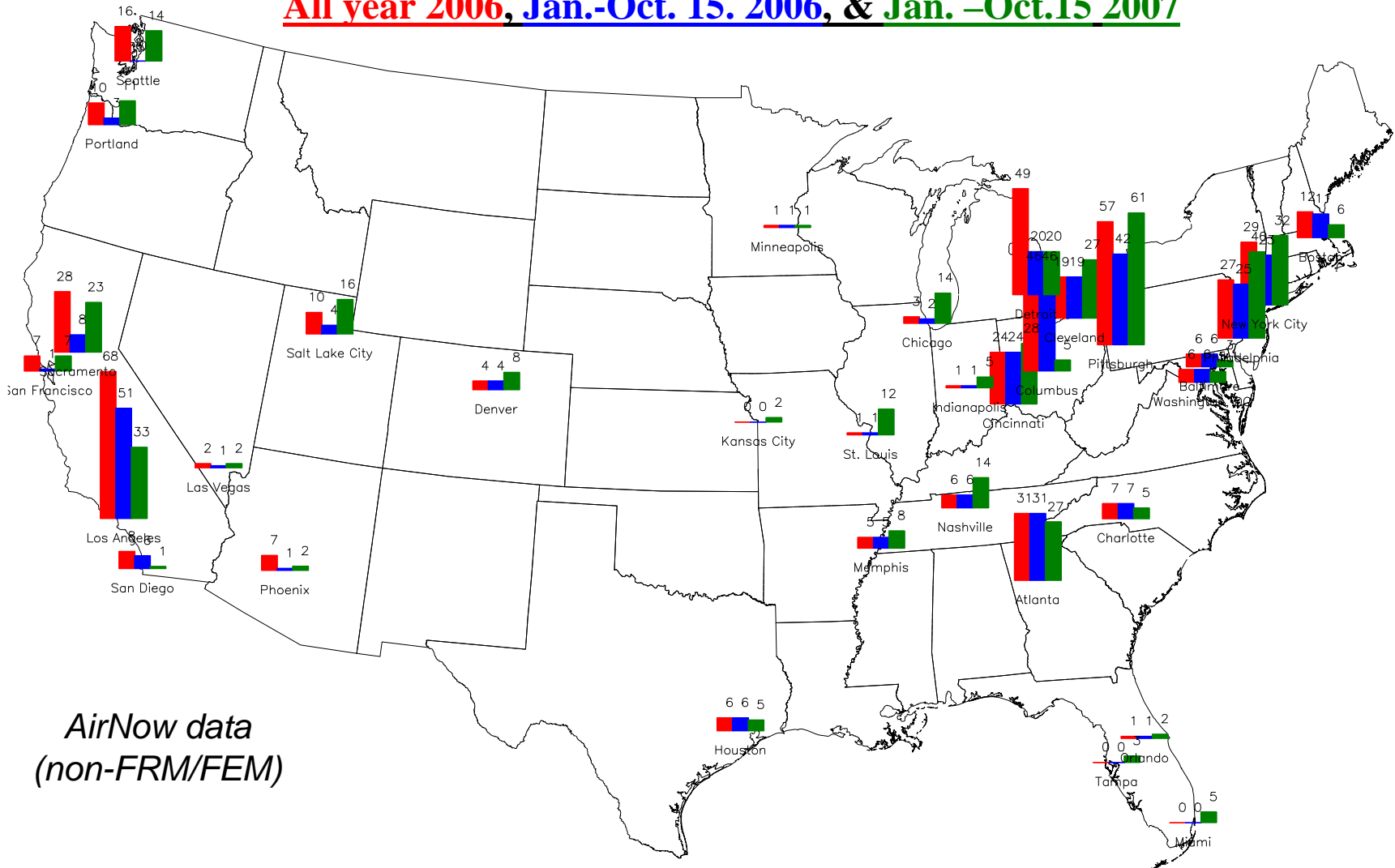
- Sites in current NAA violate the 24-hr NAAQS (131 sites)
- Sites not in a current NAA violate the 24-hr NAAQS (48 sites)

- 48 violating sites outside of current NAA are located in 27 areas:
  - 25 CBSAs: Largest are Seattle, Sacramento, San Jose, Milwaukee, Salt Lake City. Two in MARAMA region are State College, PA and Allentown, PA.
  - 2 State-Counties (not part of a CBSA): Shoshone ID; Ravalli MT

# Preliminary 24-Hour PM<sub>2.5</sub> - 2007 vs. 2006

## MSA Exceedance Day Counts – USA Today Cities

All year 2006, Jan.-Oct. 15. 2006, & Jan. –Oct.15 2007



*AirNow data  
(non-FRM/FEM)*