UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards (OAQPS)
Research Triangle Park, North Carolina 27711

June 30, 2005

MEMORANDUM

SUBJECT: Analyses of Particulate Matter (PM) Data for the PM NAAQS Review

FROM: Mark Schmidt, OAQPS
Neil Frank, OAQPS
David Mintz, OAQPS
Tesh Rao, OAQPS
Lance McCluney, OAQPS

TO: file

The purpose of this memorandum is to describe and summarize multiple sets of analyses conducted for the review of the Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS). $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$ data (the latter generally estimated via difference method from collocated $PM_{2.5}$ and PM_{10} instruments) were analyzed, as well as PM composition information. Most $PM_{2.5}$ and PM_{10} data, and some corresponding meteorological information, were extracted from EPA's Air Quality System (AQS) database on various dates in July and August of 2004. $PM_{2.5}$ composition data from urban sites in the EPA Speciation Trends Network (STN) were retrieved from AQS in July 2004. PM mass and $PM_{2.5}$ composition data, from typically rural sites in the Interagency Monitoring of PROtected Visual Environmental (IMPROVE) aerosol monitoring network, were acquired from the National Park Service in October 2004. Additional PM composition data were obtained from EPA's "Supersites" program (for the Los Angeles metropolitan area; data were obtained from the principal investigator) in June 2004, and also from the SouthEastern Aerosol Research and Characterization Study (or SEARCH, for four monitoring in Georgia and Alabama) on various dates. Additional raw meteorological data were obtained from the National Weather Service; a database of 10-year average relative humidity-related measures was provided by Science Applications International Corporation (SAIC), an EPA contractor. Meteorological data were needed for visibility-related analyses, and also to convert AQS PM_{10} samples reported at 'standard conditions' (25° C, 760 mm Hg) to 'local conditions' (actual temperature and pressure). The conversion was necessary to calculate accurate estimates for $PM_{10-2.5}$; PM_{10} data are generally reported to AQS at standard conditions, and $PM_{2.5}$ data are reported at local conditions.

There are four attachments to this memo, each corresponding to the different types of data analyzed: Attachment A describes the AQS-based, 24-hour duration PM analyses;

Attachment B describes the AQS-based hourly PM characterization analyses; Attachment C describes the PM speciation (STN, IMPROVE, Supersite, and SEARCH) data analyses; and Attachment D describes the PM visibility-related analyses. Each attachment itemizes specific analysis tasks and notes related goals, assumptions, caveats, and processing methodology. Additional pertinent details are provided in the included presentation-format outputs, which include text, tables, maps, and graphs.

All AQS-based 24-hour duration PM (10 and 2.5 micron size cuts) data and hourly PM_{10} data used in the analyses were sampled with Federal Reference Methods (FRM) or Federal Equivalent Methods (FEM). Hourly AQS $PM_{2.5}$ data and particle data collected in the ESpN, IMPROVE, and Supersite networks (Attachment C) generally utilized non-FRM/FEM techniques.

Some analysis results are summarized at a broad regional level using the geographic regions specified below. The area definitions correspond to the regions utilized in previous EPA reports. The origin of the PM region definitions can be traced back to Figure 6-30 of EPA's 1996 PM Criteria Document, which identified regions on the basis of "uniqueness in aerosol trends, seasonality, size distribution, or chemical composition." Some sites (e.g., those in Alaska, Hawaii, Puerto Rico, and the Virgin Islands) were not assigned to a PM region. For these analyses, these sites were placed in a category labeled as 'Not in PM Region'. Data for these sites are excluded from charts shown 'by region' but are included elsewhere. Some analyses compare the eastern U.S. ('East') to the western U.S. ('West'); PM Regions 1, 2, and 3 are considered the 'East' and PM regions 4, 5, 6, and 7 are defined as the 'West'.

PM REGION CODE	PM REGION DESCRIPTION	HOW DEFINED
1	Northeast	ME, NH, VT, MA, RI, CT, NJ, DE, MD*, PA*, NY*, VA*, WV* (*east of -78.50° W longitude)
2	Southeast	NC, SC, TN, GA, FL, AL, MS, LA, AR, OK*, TX* (*east of -97.70° W longitude)
3	Industrial Midwest	NY*, PA*, WV*, VA*, KY, OH, MI, IN, IL, WI#, MN#, IA#, MO# (*west of -78.50° W longitude, #east of -91.50° W longitude)
4	Upper Midwest	MN*, WI*, IA*, MO*, ND, SD, NE, KS, CO# (*west of -91.50° W longitude, #east of -104.05° W longitude)
5	Southwest	OK*, TX*, NM, AZ, NV#, CA# (*west of -97.70° W longitude, #south of 37.00° N latitude and east of -115.50° W longitude)
6	Northwest	WA, ID, MT, WY, UT, OR, CO*, CA#, NV# (*west of -104.05° W longitude, #north of 37.00° N latitude)
7	Southern California	CA*, NV* (*west of -115.50° W longitude and south of 37.00° N latitude)

For additional information on the analyses documented in the attachments, please contact Mark Schmidt at (919) 541-2416.

4 Attachments

Attachment A

AQS-Based, 24-Hour Duration PM Analyses

General / Background:

This attachment describes the analyses of 24-hour duration $PM_{2.5}$ and PM_{10} data obtained from AQS. It also documents the analyses of 24-hour duration $PM_{10-2.5}$ estimates which were largely derived from the aforementioned AQS datasets; a limited amount of directly reported $PM_{10-2.5}$ data (via dichotomous samplers) were also obtained directly from AQS.

Construction of $PM_{2.5}$ database

The database utilized for most 24-hour $PM_{2.5}$ PM Staff Paper (SP) analyses is a hybrid of the one used to construct 2001-2003 production design values (PDV's) and, hence, used in the $PM_{2.5}$ designations process. Although the raw data are exactly the same, there are several core differences in the definition and determination of 'complete' sites. For the SP analyses, any site with 11 or more observations in all 12 quarters (2001-2003) was considered 'complete' and usable for general characterization. For PDV processing, 11 or more observations per quarter (all 12 quarters) was initially only sufficient (i.e., deemed a site 'complete') to prove nonattainment of the annual standard. To initially be deemed complete, in order to show that the annual standard was being met, a site needed at least 75% data capture in all 12 quarters; the 75% cut-point was based on the required sampling frequency. Additionally for the PDV processing, sites that initially did not meet the required completeness goals (11+ samples or 75%+ capture) but were close were then subjected to several conservative data 'substitution' routines to see if there was a good likelihood that they would have shown attainment or nonattainment of the standard had they actually met the completeness goals. These substitution routines included the substitution (for evaluation purposes only, not for actual modification of their PDV's) of low values for missing data to show nonattainment, and high values for missing data to show attainment. Sites that passed one or more of these tests were then deemed complete and their PDV considered valid. For SP analyses, data substitution was not implemented. One additional difference between PDV and SP processing is the treatment of flagged data. For PDV processing, regionally-concurred event-flagged data were excluded from the official design values, although such data did count towards completeness requirements. Unless otherwise specified, all data including flagged event (exceptional and/or natural) data were used for general SP characterization analyses (i.e., SP Chapter 2); DV's excluding regionally-concurred flags were used to generate tables of estimated number/percent of counties not meet alternative standard. It should be noted that in both PDV processing and the SP analyses, the 3-year average metrics (annual means, 98th percentiles, and 99th percentile for SP analyses) are referred to as design values (DV's). Separately, the 3-year DV's are frequently referred to simply as 'annual means' or '98th percentiles'. To reiterate and elaborate, for general SP characterization analyses, the following bullets are applicable (*unless otherwise noted for specific analyses*):

- 24-hour duration data for the time period 2001 to 2003 were polled from AQS for parameter 88101 [$PM_{2.5}$, local temperature and pressure conditions (LC)] on July 6, 2004.
- Only Federal Reference Method (FRM) or Federal Equivalent Method (FEM) data were considered. The following AQS method codes are considered FRM/FEM: 116, 117, 118, 119, 120, 123, 142, 143, 144, and 145.
- DV's were computed with and without event-flagged data. DV's excluding event-flagged data were utilized for Analyses 3, 12 and 13. (For Analysis 3, all event-flagged data were excluded; for Analyses 12 and 13, only regionally-concurred event-flagged data were excluded.)
- Data were processed on a site basis. The monitor with first occurring parameter occurrence code (POC) was considered the 'primary' monitor. If an additional monitor (POC) at the site (i.e., a 'collocated' one) contained an FRM sample on a day in which one was not present at the primary POC, then those data were used for the site. Essentially, all POC's were merged but only one sample per day maximum utilized, precedence given to the lowest POC number.
- SAS code ('pmfinemacro part1.sas') was used to pull the raw data from AQS; weed out non-FRM measurements; merge collocated monitor data to a site basis; ascertain data capture statistics; and compute means, percentiles, and corresponding DV's.
- To be considered complete and hence, usable for SP analyses, a site needed at least 11 samples in each of the 12 quarters (irregardless of sampling frequency). 827 sites met the completeness goal.
- Unless otherwise noted, the SP $PM_{2.5}$ database was used to generate the $PM_{2.5}$ plots, tables, and related outputs. [Occasionally, the $PM_{2.5}$ component of the $PM_{10-2.5}$ database was used in order to enhance the $PM_{2.5}$ versus $PM_{10-2.5}$ comparisons.]

Construction of $PM_{10-2.5}$ databases

Two 24-hour $PM_{10-2.5}$ databases (db's) were generated for Staff Paper (SP) related analyses, a core database (termed the 'regular' db) and a somewhat larger database (called the 'extended' db). The regular db was utilized for all $PM_{10-2.5}$ characterization analyses (i.e., for SP Chapter 2). The extended db was used for: 1) estimating the number/percentage of counties that would potentially not meet alternative NAAQS levels, 2) approximating potential $PM_{10-2.5}$ NAAQS levels (98th or 99th percentile) that would be 'equivalent' to the current PM_{10} daily NAAQS (expected exceedance form, 150 $\mu g/m^3$ level); and 3) evaluating possible network design scenarios.

In general, the regular db was constructed largely from collocated, same-day FRM/FEM PM_{10} and $PM_{2.5}$ measurement pairs utilizing a simplistic difference computation. This element of the processing was very similar to that implemented for previous SP processing. However, this time an additional $PM_{10-2.5}$ component was included, that being direct measurements of the size cut emanating from dichotomous ('dichot') samplers; only a small amount of dichot data were available / used. Data for verified micro/middle-scale source-impacted $PM_{2.5}$ sites were eliminated from consideration into the potential $PM_{10-2.5}$ database; these sites were not considered to be appropriate candidates for future $PM_{10-2.5}$ network sites. [The nine such sites are (AQS

ID): ‘180890022’, ‘180970066’, ‘180970043’, ‘170311016’, ‘171190023’, ‘170990007’, ‘440070020’, ‘481410053’, and ‘291250001’.]

The extended db includes the ‘regular’ db plus data pairs from non-collocated (but nearby - up to 5 miles away) FRM/FEM sites. The $PM_{10-2.5}$ estimate was anchored at the PM_{10} site. The assumption is that $PM_{2.5}$ is fairly spatially homogenous, but PM_{10} is not. [The rationale for expanding the $PM_{10-2.5}$ db to included non-collocated pairs of data is as follows: Many ‘high’ PM_{10} sites do not have collocated $PM_{2.5}$ because of disparate monitoring objectives. For PM_{10} the central objective is ‘highest concentration’; for $PM_{2.5}$ the main NAAQS objective is ‘population exposure’. Hence, by not including these non-collocated pairs, we would be ignoring many potentially high $PM_{10-2.5}$ locations.] Several PM_{10} sites identified as source-oriented and not also population exposure were omitted from the extended database because it was felt that they were not likely candidates for a future $PM_{10-2.5}$ network. [These sites, identified by EPA regional staff, are (AQS Site ID’s): ‘090090018’, ‘290970003’, ‘295100092’, ‘401010167’, ‘440070020’, ‘450430006’, ‘450630009’, ‘560050874’, ‘560050885’, ‘560050891’, ‘560050894’, and ‘560050907’.] Analysis 11 documents characterization of areas as urban based on several potential measures of urbanization. These measures were used in characterizing $PM_{10-2.5}$ concentrations in urban areas, as applied in Analyses 12 and 13.

The following statements detail the $PM_{10-2.5}$ db’s construction (both ‘regular’ and ‘extended’):

- 24-hour duration data for the time period 2001 to 2003 were retrieved from AQS for the following parameters on August 24, 2004: parameter 88101 ($PM_{2.5}$, LC); parameter 81102 [PM_{10}, standard temperature and pressure conditions (STP)]; parameter 85101 (PM_{10}, LC); parameter 86101 ($PM_{10-2.5}$, STP); and parameter 81103 ($PM_{10-2.5}$, STP)
- Summary daily data (which includes hourly measurements aggregated within AQS to a 24-hour period) were extracted from AQS (also on August 24, 2004) for parameter 81102 and parameter 85101. AQS maintains the raw hourly data and also aggregates the hourly information into summary daily records. A summary record is only deemed ‘valid’ if 75% or more of the hourly data (≥18) are present.
- For the difference method, only FRM/FEM PM_{10} and $PM_{2.5}$ data were utilized. All AQS PM_{10} data (except for a lone site in Alabama, ID ‘010970030’) were assumed to be FRM/FEM. $PM_{2.5}$ data were determined to be FRM/FEM based on method code (as indicated above for $PM_{2.5}$ db development).
- For the difference method, no effort was made to account for differences in sampling instruments or protocols between the collocated PM_{10} and $PM_{2.5}$ monitors. Because of these differences (and other factors), occasionally the calculated $PM_{10-2.5}$ values were negative; this is not unexpected for two independent observations, and negative $PM_{10-2.5}$ concentrations were not censored from the analyses.
- For the difference method, both the PM_{10} and $PM_{2.5}$ data used in the difference calculation were in units of $\mu g/m^3$ at local conditions, thus the calculated $PM_{10-2.5}$ values also are in those units. Parameter 81102 data, both summary and daily, were converted to local conditions using collocated temperature and pressure information. If collocated temperature and/or pressure data were not available, meteorological data from the nearest NWS station were used. If collocated data

were not available and the NWS data were missing, the STP data were not converted to LC and not used in the analyses.

- For the difference method, $PM_{10-2.5}$ estimates were constructed from all site-day pairs of collocated PM_{10} and $PM_{2.5}$ measurements. For example, if, for a particular site day, there were two readings of PM_{10} ('1' and '2') and two readings of $PM_{2.5}$ ('a' & 'b', then four total $PM_{10-2.5}$ estimates were generated ('1a', '1b', '2a', and '2b').
- In situations where multiple site-day estimates of $PM_{10-2.5}$ existed (combination of difference method pair estimates and/or direct dichot measurements), they were averaged to obtain an average $PM_{10-2.5}$ measurement for the site-day. This average was considered the actual $PM_{10-2.5}$ estimate or 'sample' for that site-day (and counts as only one observation towards data completeness). Thus, data were essentially processed on a 'site' basis.
- To be used in the SP analyses, a site needed 4, 8, or 12 consecutive quarters (2001-2003) of 11+ samples. This requirement is in contrast to the individual $PM_{2.5}$ and PM_{10} analyses which both required 'completeness' in all 12 quarters; the $PM_{10-2.5}$ criteria are more relaxed, in order to maximize the number of usable sites. Though nationally and regionally there are a sufficient number of 12-quarter complete $PM_{2.5}$ sites and also a sufficient number of 12-quarter complete PM_{10} sites, there are <u>not</u> a sufficient number of *collocated* 12-quarter complete $PM_{2.5}$ and PM_{10} sites, Specifically, the $PM_{10-2.5}$ analyses utilized <u>the most recent</u> 4, 8, or 12 consecutive quarters of 11 or more samples. A simple example is shown below. For this example site, the quarters that would have been utilized are shaded. Since the selection criterion evaluates available data in increments of 4 quarters, previous quarters could not be used due to the shortfall in 2002, quarter 1. An additional increment of 4 consecutive quarters meets the 11 minimum sample threshold (2001, quarters 1-4), but would not have been used since a more recent band of data (shaded) were available. Although the utilized selection criteria do not guarantee a calendar year(s) of data, they do provide at least one full year consisting of four quarters, thus reducing seasonal bias. Data present in quarters not part of the 4-, 8-, or 12-quarter period of interest were deleted and thus, not included in subsequent analyses.

Year / Quarter =	2001, Q1	2001, Q2	2001, Q3	2001, Q4	2002, Q1	2002, Q2	2002, Q3	2002, Q4	2003, Q1	2003, Q2	2003, Q3	2003, Q4
N=	12	13	14	15	10	15	16	14	15	13	11	9

- 489 sites (located in 351 counties) are in the $PM_{10-2.5}$ 'regular' database: 137 with 4 complete quarters, 122 with 8 complete quarters, and 230 with all 12 complete quarters. 712 sites (located in 382 counties) are in the $PM_{10-2.5}$ complete 'extended' database: 201 encompass 4 complete quarters, 177 have 8 complete quarters, and 334 have all 12 complete quarters.
- 'Annual' means and percentiles (e.g., 98^{th}, 99^{th}) were computed from 'annualized' (4-quarter increment) statistics. For example, if a site had 8 complete quarters starting with 2001-Q3 and ending with 2003-Q2, then two 'annual' 98th percentiles were computed, one for 2001-Q3 through 2002-Q2 and the other for 2002-Q3 through 2003-Q2. Likewise, two 'annual means' were calculated (according to standard weighted mean processing protocol in which data are first averaged by quarter, and then the 4 quarterly means are averaged together to obtain an 'annual

mean' figure). The 2 'annual' numbers (2 means and 2 98th percentiles) were then averaged to obtain the site's DV-type metrics. Hence, the DV-type metrics might represent 4, 8, or 12 quarters of data. Separately, the (4, 8, or 12-quarter) DV-type metrics are frequently referred to simply as 'annual means' or '98th percentiles'.

- For both db's ('regular and extended'), DV's were computed two ways: including all data, and excluding event-flagged data. A daily $PM_{10-2.5}$ estimate was considered flag if any of the constituent $PM_{2.5}$ or PM10 data were flagged. Concurrence was not a factor.
- SAS code was used to pull the raw 24-hour data from AQS ('raw from aqs.sas'); extract the summary daily data from AQS ('daily from aqs.sas'); process the AQS and NWS meteorological data needed to convert STP PM_{10} and $PM_{10-2.5}$ dichot data to LC ('gettemppress0103.sas'); filter out non-FRM $PM_{2.5}$ data, create $PM_{10-2.5}$ difference records, and create an interim db of all site-day record ('calc coarse 0103.sas'); average multiple site-day measurements, evaluate completeness requirements, compute means, compute percentiles, compute DV's , and generate raw and summary db's for complete sites only ('coarse comp final.sas').

Construction of PM_{10} databases

For SP analyses, the PM_{10} database utilized was the official 2001-2003 design value database with one addition. Official PM_{10} DV's exclude regionally-concurred natural and exceptional event flags. For comparability with $PM_{2.5}$ and $PM_{10-2.5}$ general characterization analyses (i.e., SP Chapter 2), PM_{10} DV's were also calculated using all data, flagged or not. The PM_{10} db creation relied on daily summary AQS extractions; SAS code ('airs_dailysum_pm10dv.sas') was used for the extraction. The AQS daily summaries table encompasses 24-hour filter measurements and hourly data aggregated to a 24-hour basis (as noted above in the $PM_{10-2.5}$ database discussion). Of the latter type, only the valid data (those with DAILY_CRITERIA_IND='Y', signifying 18+ hourly observations per day) were used.

Boxplot Figures

Many of the generated analyses figures are boxplots. Unless otherwise noted, in all of the AQS-based, 24-hour average duration boxplots, the following definitions apply:

- The bottom of the box depicts the 25th percentile of the plotted distribution.
- The top of the box depicts the 75th percentile of the plotted distribution.
- The line through the box identifies the distribution median.
- The top whisker cap identifies the 95th percentile of the plotted distribution.
- The bottom whisker cap identifies the 5th percentile of the plotted distribution.
- If shown, the distribution maximum and minimum are shown as asterisks.

Analysis 1 – Summaries and boxplots of $PM_{2.5}$ and $PM_{10\text{-}2.5}$ annual mean and 98th percentile DV's, by region

Goals:
- ? To characterize the typical average concentration levels of PM_{10} and $PM_{2.5}$ for different U.S. regions.
- ? To make comparisons of the size cuts.

Outputs:
- o Summary statistics were generated by region. See tables in Output A.1a.
- o Boxplots were generated of the distribution of site-level annual means and 98th percentile by region. See Output A.1b.

Methods:
- The SAS procedure UNIVARIATE was used to generate the summary statistics.
- SAS code ('inputbox mean 98p.sas' and 'boxplot pmf pmc.sas') was used generate the boxplots.

Analysis 2 – Maps of $PM_{2.5}$, $PM_{10\text{-}2.5}$, and PM_{10} county maximum annual mean and 98th percentile DV's, by region

Goals:
- ? To identify specific geographic areas with high and low annual mean and 98th percentile concentration levels.

Outputs:
- o $PM_{2.5}$ maps are shown in Output A.2a.
- o $PM_{10\text{-}2.5}$ maps are shown in Output A.2b.
- o PM_{10} maps are shown in Output A.2c.

Methods:
- Each county (with a complete site) was assigned the value of the site with the highest stated statistic (annual mean or 98th percentile DV).
- SAS code, 'map4shade.sas', was used to generate the $PM_{2.5}$ and $PM_{10\text{-}2.5}$ maps.
- SAS code, 'bwfammap.sas' and 'bwcntymap2.sas', was used to generate the PM_{10} maps.

Analysis 3 - Event-flagged data, $PM_{2.5}$ and $PM_{10\text{-}2.5}$

Goals:
- ? To identify the types of events which are flagged in AQS.
- ? To determine if there are significant amounts of event-flagged PM data.
- ? To determine if 'high' sites flag more data than 'low' sites.
- ? To see if events impact DV's.
- ? $PM_{2.5}$: To ascertain whether any DV's changed from 'violating the standard' to 'meeting the standard' after removing their event-flagged data
- ? To see if the impacts are different for 'high' versus 'low' sites
- ? To determine whether data distributions are similar for sites that flag data compared to those that do not flag data.

- ? $PM_{2.5}$: To evaluate the specific impact of episodic events on various air quality statistics (case studies).

Outputs:

- o Various tables, plots, and related discussion; see Output 3a for $PM_{2.5}$ and Output 3b for $PM_{10-2.5}$.

Methods:

- For the $PM_{2.5}$ flag analysis, the raw database was re-evaluated and DV's recalculated; any data point flagged for an event was excluded from the new DV computations *irrespective of the AQS concurrence indicator*. Unlike in production design value (PDV) processing for $PM_{2.5}$ and PM_{10} (and also for Analysis 12 and 13), the AQS regional concurrence indicator was not evaluated. Thus, the concurrence being set to 'yes' was not a requisite for flagged data to be excluded.
- SAS code ('ex events fine.sas' and 'quebec.sas') was used to evaluate the $PM_{2.5}$ events.
- SAS code ('ex events coarse.sas) was used to evaluate the $PM_{10-2.5}$ events.

Analysis 4 - Comparisons of site-level annual means to 98th percentiles, $PM_{2.5}$ and $PM_{10-2.5}$

Goals:

- ? To evaluate the relationship between site-level annual means and site-level 98th percentiles.

Outputs:

- o See Output A.4.

Methods:

- The distributions of site-level 98^{th} percentiles were plotted by intervals of site-level mean levels.
- SAS code was used to generate the $PM_{2.5}$ and $PM_{10-2.5}$ plots ('pmf boxplot p98 intmean.sas' and 'pmc boxplot p98 intmean.sas').

Analysis 5 – Regional correlations of $PM_{2.5}$, $PM_{10-2.5}$, and PM_{10}

Goals:

- ? To evaluate the correlation among the three size cuts.

Outputs:

- o See Output A.5.

Methods:

- Because the represented periods are different for $PM_{2.5}$, $PM_{10-2.5}$, and PM_{10} (e.g., For $PM_{10-2.5}$, the most recent 12, 8, or 4 quarters were utilized; for PM_{10} and $PM_{2.5}$, all 12 quarters were needed) and also because completeness was applied independently, the selected time periods did not necessarily match. If the common time periods of constituent raw data (for the sites that met the parameter selection criteria) were used for this analysis, some sites common to multiple parameters would not have any matches (by site-day) and others would have a seasonal bias (only have matches in certain quarters). To avoid this situation, the raw data used in this analysis were culled from the $PM_{10-2.5}$ database. This insured an equal number of each quarter for each site and also insured a minimum of 44 samples for each site (4 quarters * 11 samples each).

- A Pearson correlation coefficient was calculated for each site fraction pair (PM_{10} versus $PM_{2.5}$, $PM_{2.5}$ versus $PM_{10-2.5}$, and PM_{10} versus $PM_{10-2.5}$). The site correlation coefficients for each fraction were then averaged by region.
- SAS code was used to calculate the correlations and produce the plot ('procorr.sas').

Analysis 6 – Distribution of ratios of 24-hour average $PM_{2.5}$ to PM_{10}, by region

Goals:

? To identify typical site average 24-hour ratios of $PM_{2.5}$ to PM_{10}, by region.

Outputs:

o See Output A.6.

Methods:

- The ratio of $PM_{2.5}$ to PM_{10} was first calculated for each site-day. The site-day ratios of $PM_{2.5}$ to PM_{10} were then averaged by site and the distribution of the site ratios plotted by region.
- SAS code ('ratio of pmf to pmt.sas') was used for the analysis.

Analysis 7 – Evaluation of spatial averaging (SA) for $PM_{2.5}$

Goals:

? To determine if there are large differences between 'regular' DV's (based on highest site in area) and DV's calculated with SA.

? To evaluate various issues with spatial averaging. (e.g., are the would-be violating sites that could utilize SA located in lower-income, high percentage-minority, and/or lower education area (based on Census tract information) than the overall area?]

? To evaluate the current criteria for using SA.

Outputs:

o See Output A.7.

Methods:

- Initially started with the default SP $PM_{2.5}$ database (all sites with 11+ samples in each of the 12 quarters 2001-2003). Eliminated sites that are not (officially) compared to annual standard. (AQS Site ID's: '180890022', '180970066', '180970043', '170311016', '171190023', '170990007', '440070020', '481410053', and '291250001')
- Initially enforced the CFR spatial criteria of: 1) 0.6 overall correlation between sites, and 2) no more than 20% difference in site annual mean and spatial annual mean. The criterion that all SA sites should be impacted by similar emissions was not evaluated.
- Enforced CFR data handling requirement that if SA annual mean is less than or equal to the annual standard, then only SA sites with 75%+ capture each of the 4 Q's would have their annual mean included in the spatial annual average. (Only 11+ samples required in each of the 4 Q's if the spatial annual mean was greater

than the evaluated annual standard.) Changed level of standard (and completeness check) from 15 to 14 for accurate evaluation of SA effect on those standard levels.
- For 'area' definitions, utilized OMB definitions for Core-Based Statistical Areas (CBSA's) and Combined Statistical Areas (CSA's). If multiple sites were not located in a defined area, then area was assumed to be the county.
- Constructed SA set of sites by initially considering all sites in the area. If a site-pair correlation was less than cutoff, the lower DV site was eliminated. If a remaining set did not meet annual mean difference criterion, then the lowest DV site was omitted from the set and the revised set tried. This continued until the reduced set of sites met criteria, or until less than 2 sites were left. Note: Undoubtedly, different combinations of sites (selected with some rationale and/or at random) could/would meet criteria and yield different results.
- Only considered (for SA) areas with: 1) a regular DV greater than the evaluated annual standard level, and 2) a spatial DV greater than any (valid) non-SA site DV in the area.
- Evaluated appropriateness of 0.6 (correlation) and 20% (max difference in annual means) levels by comparing to typical universe values.
- Tightened the correlation criterion to 0.9 and the annual mean difference criterion to 10% to evaluate changes in results. SAS code ('spatial avg.sas') was used for the analysis.
- SAS code ('spatial avg.sas' and 'simple spatial.sas') was used to conduct the analyses.

Analysis 8 – Evaluation of 'high' $PM_{2.5}$ values

Goals:
- ? To identify the minimum number of days per year a site is permitted to exceed the annual 98th, 99th, and other percentiles.
- ? To evaluate the (entire) daily distributions of data plotted by 98th (and 99th) percentile-level intervals.
- ? To evaluate the daily distributions of data exceeding site-level 98th (and 99th) DV's plotted by 98th (and 99th) percentile intervals.
- ? To ascertain the actual number and percentage of days (site average, minimum, & maximum), for the 3-year period 2001-2003, where the concentration was significantly above the site 98th or 99th percentiles. [Significant defined as 5+ $\mu g/m^3$.]

Outputs:
- o See Output A.8.

Methods:
- SAS code ('dist above p98.sas') was used for the analysis.

Analysis 9 – Monthly patterns of urban $PM_{2.5}$ and $PM_{10-2.5}$

Goals:
- ? To identify monthly patterns, by region, in concentrations of $PM_{2.5}$ and $PM_{10-2.5}$

Outputs:

- PM$_{2.5}$ boxplots are plots are shown in Output A.9a. PM$_{10-2.5}$ boxplots are shown in Output A.9b.

Methods:

- Only data from monitors with AQS 'location setting' of 'URBAN AND CENTER CITY' or 'SUBURBAN' were used. Hence, the term 'urban' actually encompasses 'suburban' sites as well.
- All 24-hour average values (for complete 'urban' sites from the 'regular' PM$_{2.5}$ and PM$_{10-2.5}$ db's) were averaged together by region-month.
- In these boxplots, the boxes represent the interquartile range (25th to 75th percentiles) of each monthly distribution and the line inside the box is the median of the distribution. The trend line represents the mean, and the number above each box represents the number of 24-hour average observations that were used to generate each box plot. Whiskers (95th and 5th percentiles) were not plotted.

Analysis 10 – Comparison of urban and rural PM$_{10-2.5}$ mass levels

Goals:

? To compare urban PM$_{10-2.5}$ mass levels to corresponding rural levels.

Outputs:

- See Output A.10.

Methods:

- This analysis compared urban and rural mass levels *within* large metro areas (i.e., each area independently)
- Lotus was used to process data; Freelance was used to generate the bar-charts.

Analysis 11 – Characterization of 'urban' areas

Goals:

? To characterize areas as urban or non-urban using various measures of urbanization.

Outputs:

- See Output A.11.

Methods:

- This evaluation focused on non-pollutant measures of urbanization, specifically population, population density, vehicle mile traveled (VMT), and VMT density.
- SAS code ('pop vmt.sas') was used for the population:VMT analysis. SAS-SQL commands were used to generate additional information.

Analysis 12 – $PM_{10\text{-}2.5}$ equivalence to PM_{10} NAAQS (daily standard)

Goals:

? To estimate concentration levels for various $PM_{10\text{-}2.5}$ design value type metrics that would correspond to the 150 µg/m3 level for the current PM_{10} (expected exceedance based) 24-hour standard.

Outputs:

- See Output A.12.

Methods:

- Actual PM_{10} site-level DV's were evaluated against estimated $PM_{10\text{-}2.5}$ DV's. The analysis utilized DV's that exclude event-flagged data.
- Only 12-quarter $PM_{10\text{-}2.5}$ sites were used in this analysis (in order to eliminate differences due to different time periods for $PM_{2.5}$ and PM_{10} measurements).
- Used only sites characterized as urban, based on analysis 11.
- SAS code ('pmc equivalence to pmt.sas') was used for the analysis.

Analysis 13 – Estimated percentage of counties not likely to meet alternative $PM_{2.5}$ and $PM_{10\text{-}2.5}$ standards and existing PM_{10} NAAQS

Goals:

? To estimate the number, percentage and population of counties in the U.S. not likely to meet alternative PM standards.

? To estimate the percentage of counties on a regional basis not likely to meet alternative PM standards.

Outputs:

- See Output A.13.

Methods:

- DV's excluding event flagged data were used in this evaluation.
- For the annual $PM_{2.5}$ standard level evaluation (by itself, and in tandem with a daily standard), the sites officially exempted from the annual standard (AQS Site ID's): '180890022', '180970066', '180970043', '170311016', '171190023', '170990007', '440070020', '481410053', and '291250001') were not considered to be in violation of the annual standard no matter the level. Essentially, the annual mean DV was set to zero for these sites. These sites *were* compared to the alternative daily standard levels.
- For PM_{10} and $PM_{10\text{-}2.5}$, results were tabulated for the entire db's (using extended db for $PM_{10\text{-}2.5}$) as well as for the respective 'urban' portions (as in analysis 11)
- SAS code ('whatif county counts pmf.sas', 'whatif county counts pmc.sas', and 'whatif county counts pmt.sas') was used for the analyses.

Analysis 14 – $PM_{2.5}$ and $PM_{10\text{-}2.5}$ spatial homogeneity

Goals:

? To investigate the spatial homogeneity/heterogeneity of $PM_{2.5}$ and $PM_{10\text{-}2.5}$ within urban areas.

Outputs:

- See Output A.14a for $PM_{2.5}$ and Output A.14b for $PM_{10-2.5}$.

Methods:

- Within CSA's with 2 or more sites, annual mean levels were compared, and inter-sites correlations were computed. An additional indicator of spatial homogeneity, 'P90', the 90th percentile of the distribution of differences in 24-hour averages between two sites in the same urban area, also was calculated.
- To minimize temporal bias for $PM_{10-2.5}$ (for annual mean comparisons), the analyses only utilized 12-quarter sites or 8-quarter sites that operated the same 8 quarters.
- SAS code ('pmf csa.sas' and 'pmc csa.sas') was used to generate the tabular output and compute relevant statistics.

PM2.5: Summary Statistics for Site-Level Annual Mean DV

PMREG	PMREGDE	n	mean	max	p95	p75	median	p25	p05	min
0	Not in PMF	17	6.629412	11.9	11.9	7.4	6.4	5.1	3.9	3.9
1	Northeast	121	13.20248	17.3	16.4	14.6	13.3	12	9.6	6.5
2	Southeast	216	12.52407	18	15.7	13.9	12.55	11.2	9.1	7.4
3	Industrial M	217	14.60461	21.2	17.4	15.7	14.7	13.5	11.4	6.6
4	Upper Midw	71	9.988732	13.9	12.6	11.3	10.5	9	6	5.5
5	Southwest	33	8.515152	16.9	14.4	10.7	7.8	6.6	4.4	4
6	Northwest	110	9.37	17	13.4	10.8	9.1	7.8	5.6	4.5
7	Southern C	42	16.63333	27.8	25.2	21.3	16.9	12	6.9	6.2
	U.S.	827	12.45961	27.8	17.2	14.6	12.6	10.4	6.6	3.9

PM2.5: Summary Statistics for Site-Level 98th Percentile

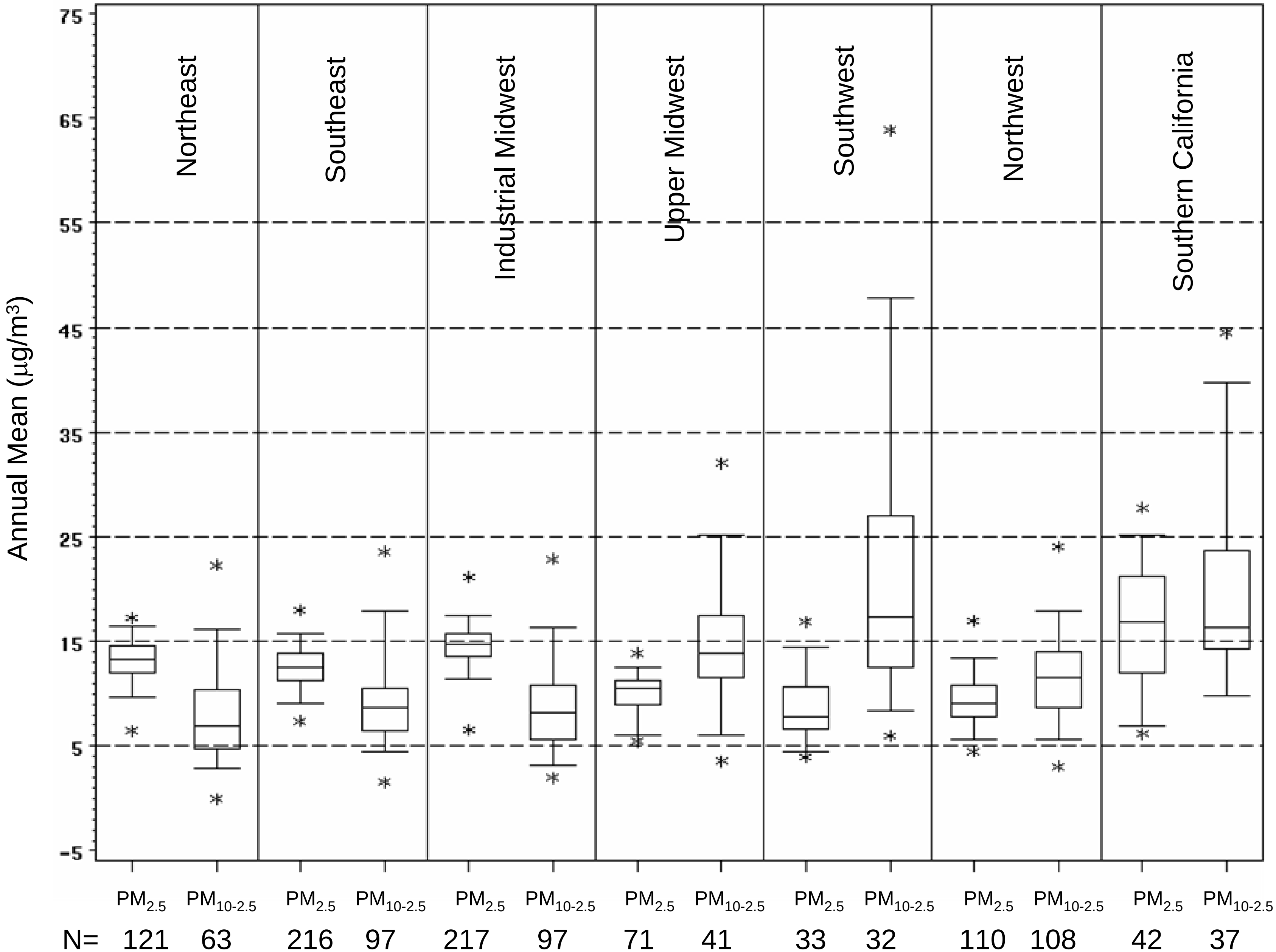
PMREG	PMREGDE	n	mean	max	p95	p75	median	p25	p05	min
0	Not in PMR	17	16.47059	40	40	17	15	13	9	9
1	Northeast	121	36.3719	48	43	40	37	33	29	25
2	Southeast	216	28.5787	40	36	31	29	26	20	17
3	Industrial M	217	36.29954	63	43	39	36	34	29	18
4	Upper Midw	71	25.08451	33	30	28	26	23	16	14
5	Southwest	33	22.09091	49	46	28	20	16	10	10
6	Northwest	110	31.87273	62	54	40	30.5	23	15	11
7	Southern C	42	45.38095	76	72	62	45.5	29	20	18
	U.S.	827	32.22854	76	46	37	32	27	17	9

PM10-2.5: Summary Statistics for Site-Level Annual Mean DV

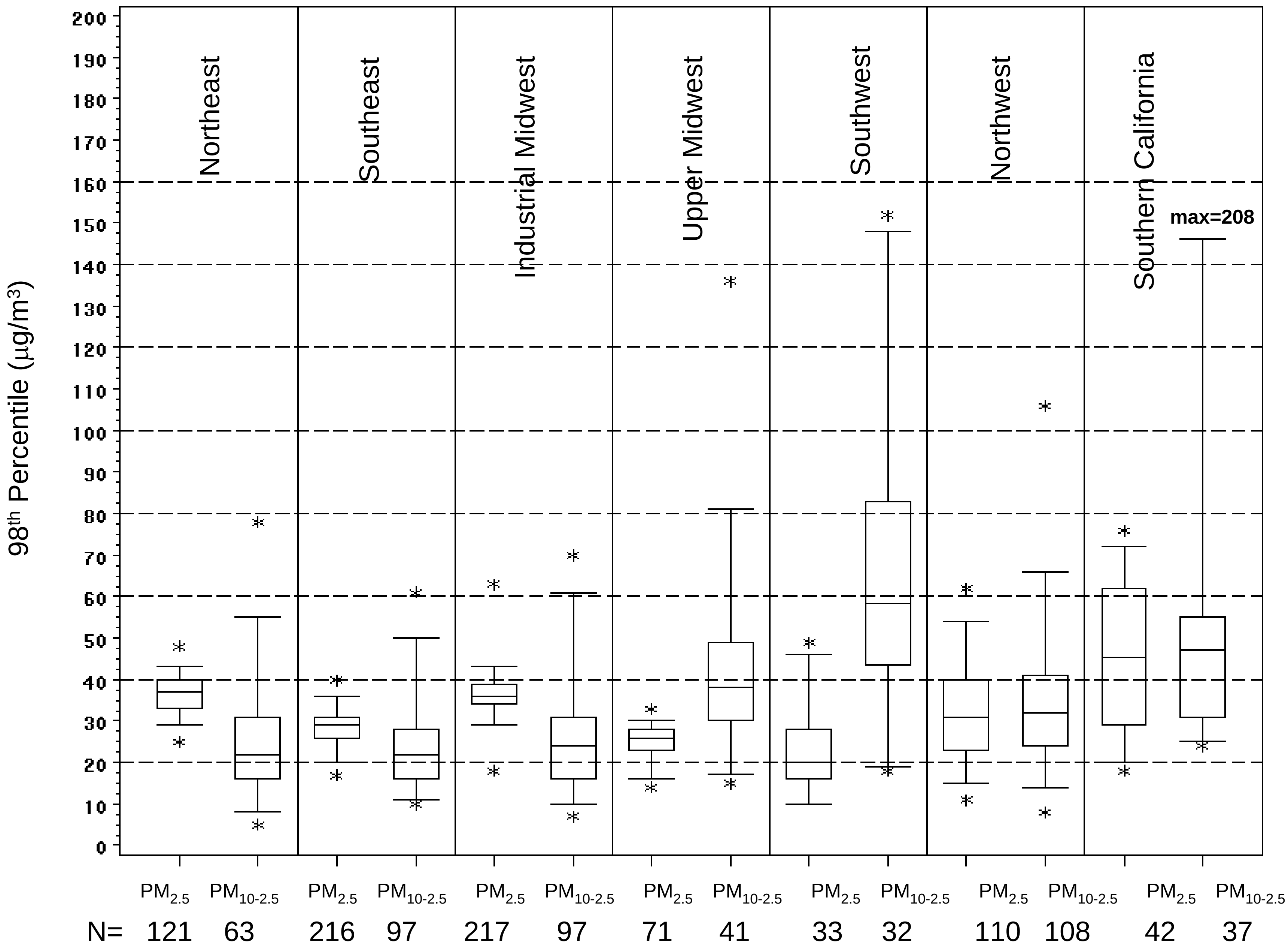
PMREG	PMREGDE	n	mean	max	p95	p75	median	p25	p05	min
0	Not in PMF	14	16.75259	30.2	30.2	24.6	15.0	11.0	1.8	1.8
1	Northeast	63	7.877622	22.3	16.2	10.4	6.9	4.8	2.8	0.0
2	Southeast	97	9.311192	23.6	17.9	10.6	8.7	6.5	4.5	1.6
3	Industrial M	97	8.842588	22.9	16.3	10.8	8.2	5.6	3.1	2.0
4	Upper Midv	41	14.37395	32.1	25.2	17.5	13.8	11.6	6.1	3.6
5	Southwest	32	21.1939	63.9	47.8	26.9	17.3	12.6	8.3	6.0
6	Northwest	108	11.58091	24.1	17.9	14.0	11.6	8.6	5.6	3.1
7	Southern C	37	19.80212	44.5	39.8	23.7	16.3	14.3	9.8	9.8
	U.S.	489	11.74376	63.9	24.9	14.7	10.5	7.0	4.1	0.0

PM10-2.5: Summary Statistics for Site-Level 98th Percentile

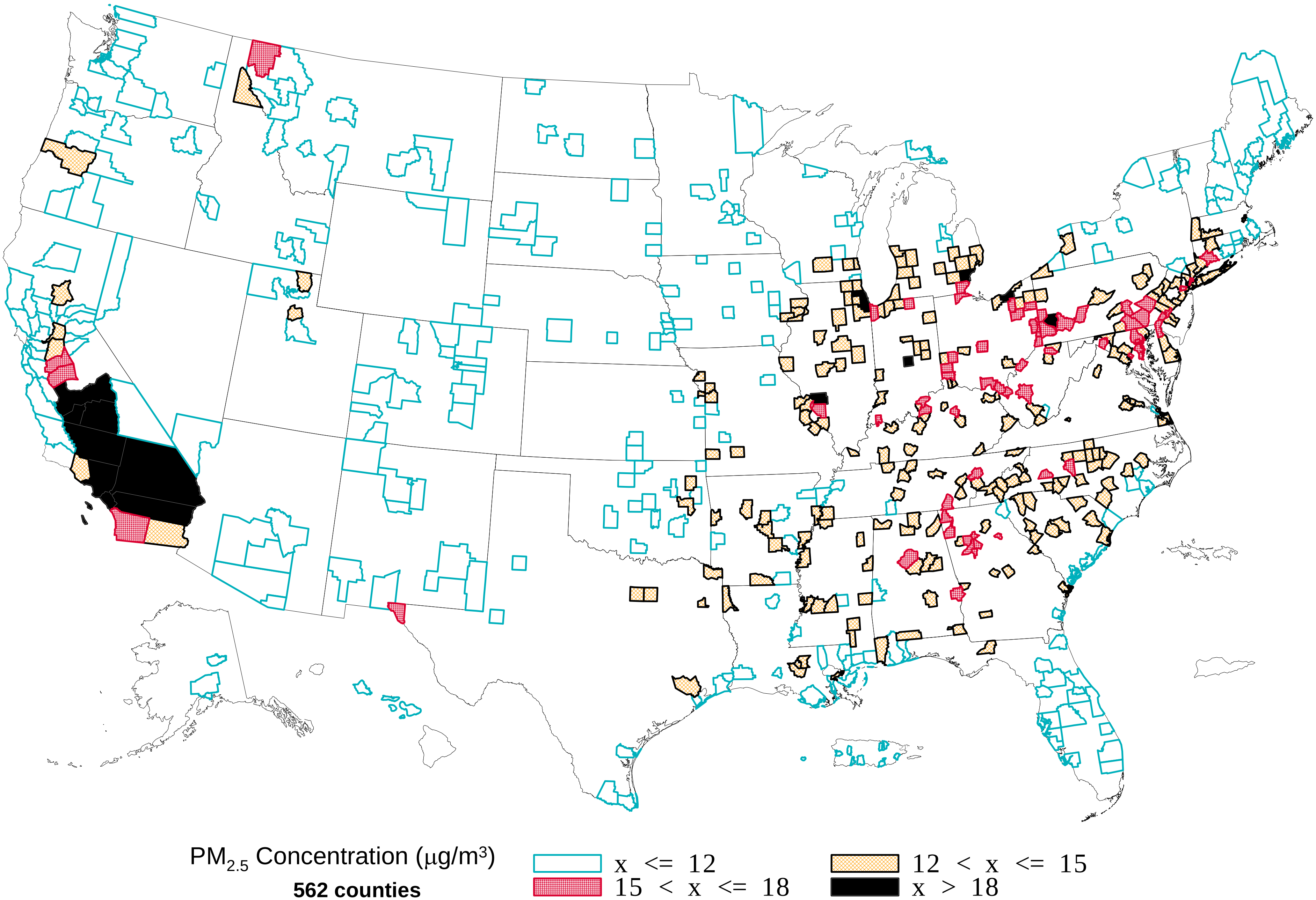
PMREG	PMREGDE	n	mean	max	p95	p75	median	p25	p05	min
0	Not in PMR	14	48	89	89	67	50.5	22	10	10
1	Northeast	63	25.39683	78	55	31	22	16	8	5
2	Southeast	97	24.5567	61	50	28	22	16	11	10
3	Industrial M	97	25.47423	70	61	31	24	16	10	7
4	Upper Midw	41	42.4878	136	81	49	38	30	17	15
5	Southwest	32	64.75	152	148	83	58.5	43.5	19	18
6	Northwest	108	33.88889	106	66	41	32	24	14	8
7	Southern C	37	52.97297	208	146	55	47	31	25	24
	U.S.	489	33.86299	208	74	41	28	20	11	5



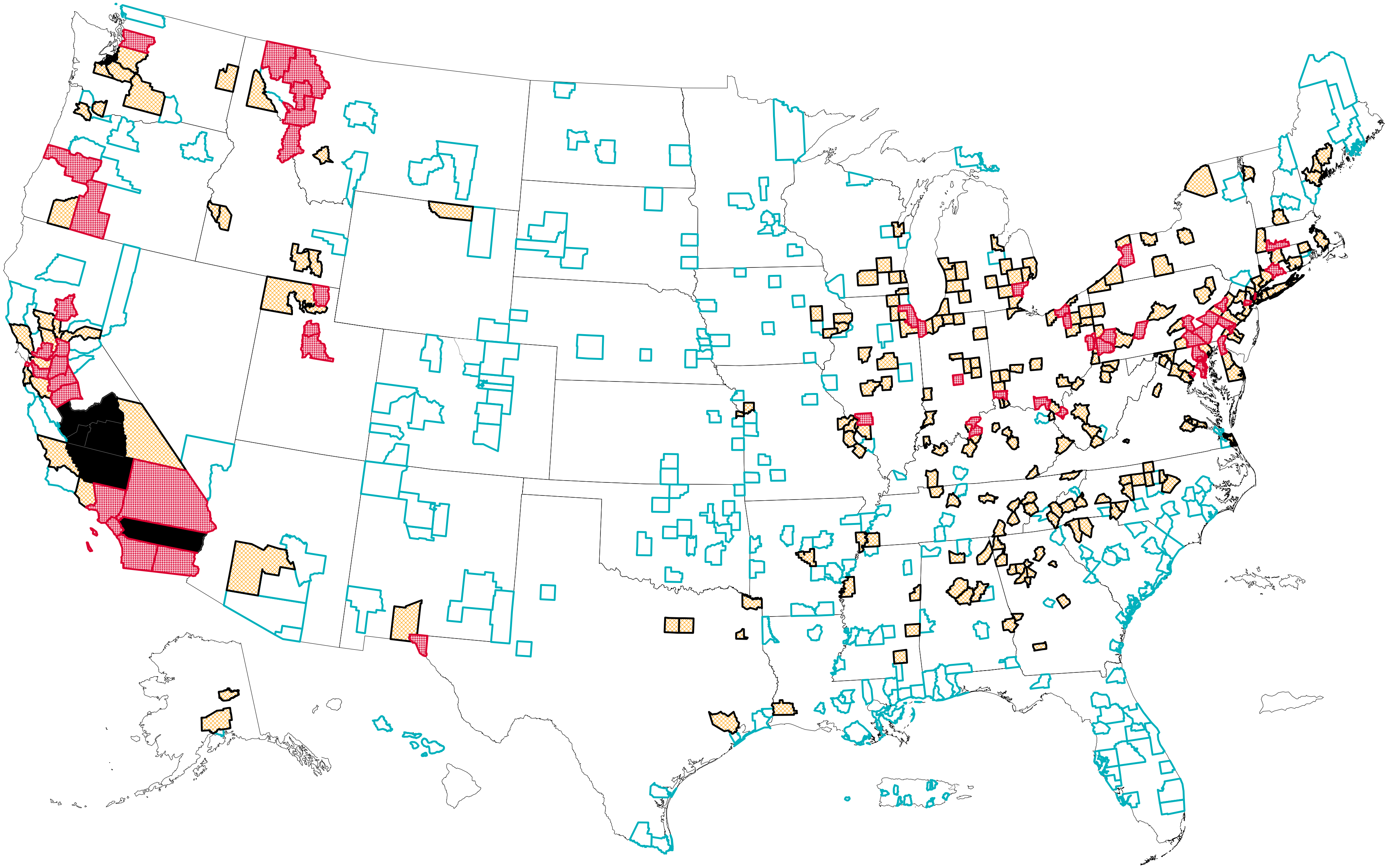
Distribution of annual mean $PM_{2.5}$ and estimated annual mean $PM_{10-2.5}$ concentrations by region, 2001-2003.
N = number of sites.



Distribution of 98th percentile 24-hour average $PM_{2.5}$ and estimated $PM_{10-2.5}$ concentrations by region, 2001-2003.
N = number of sites.



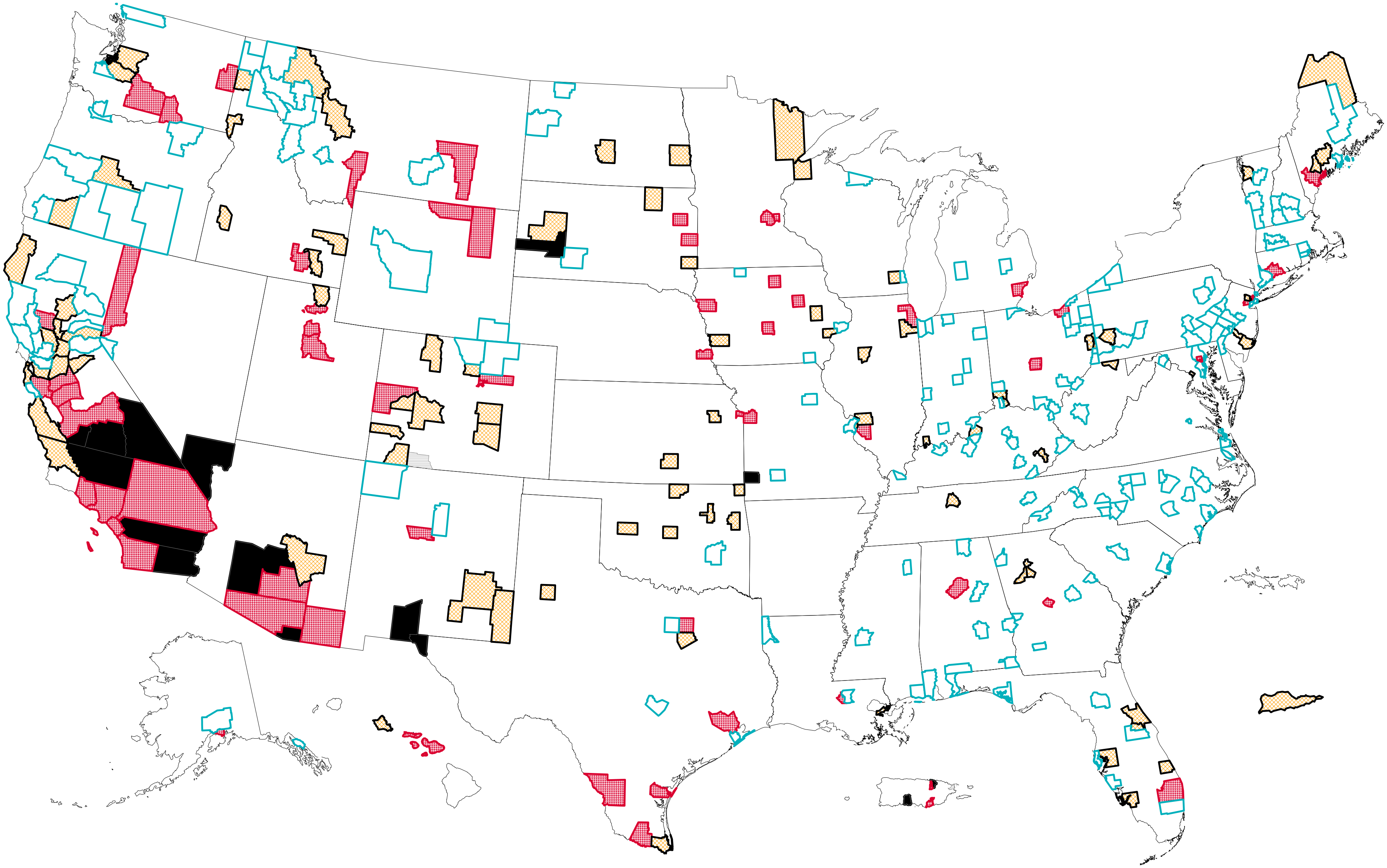
County-level maximum annual mean $PM_{2.5}$ concentrations, 2001-2003.



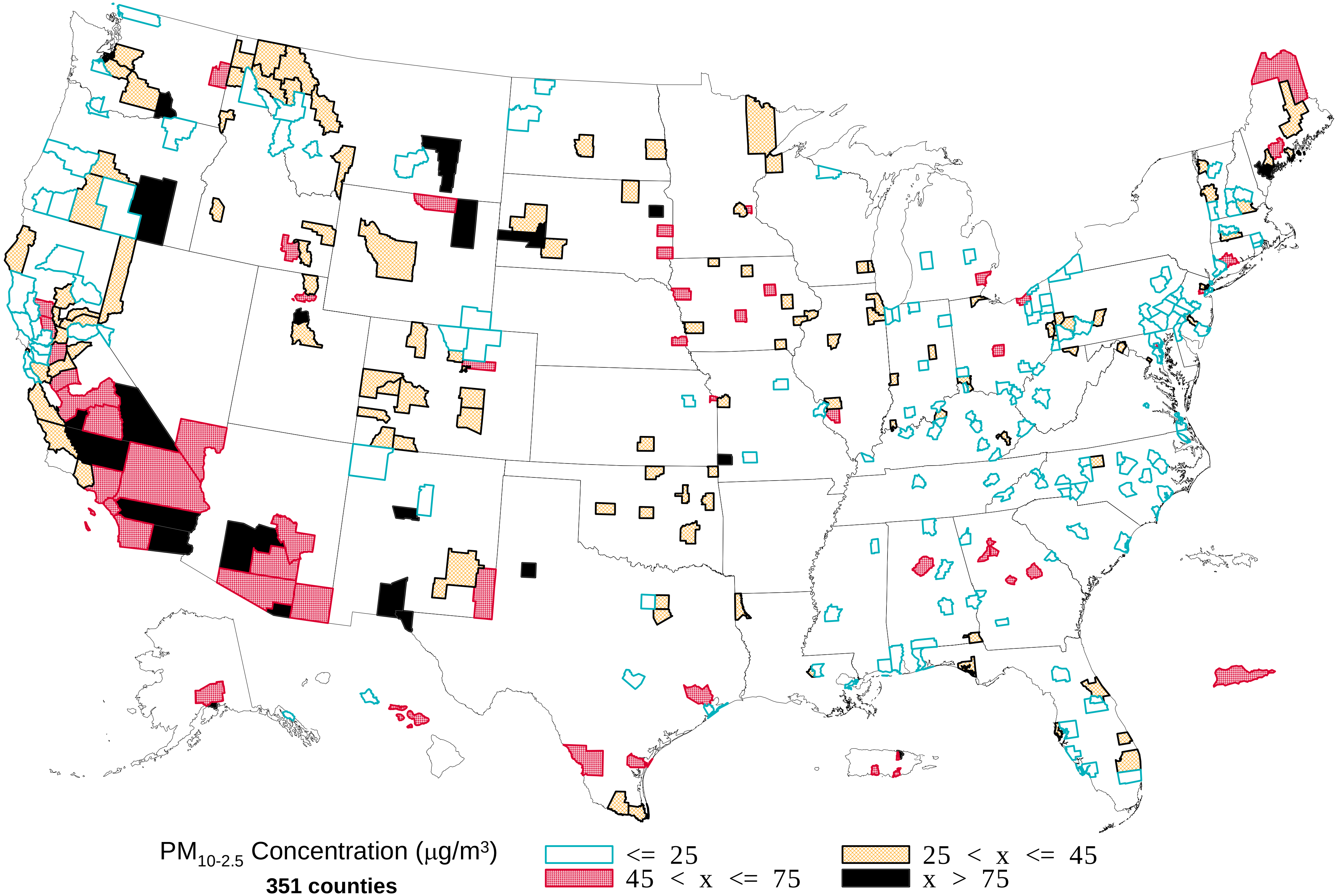
$PM_{2.5}$ Concentration (µg/m³)
562 counties

x <= 30
40 < x <= 65
30 < x <= 40
x > 65

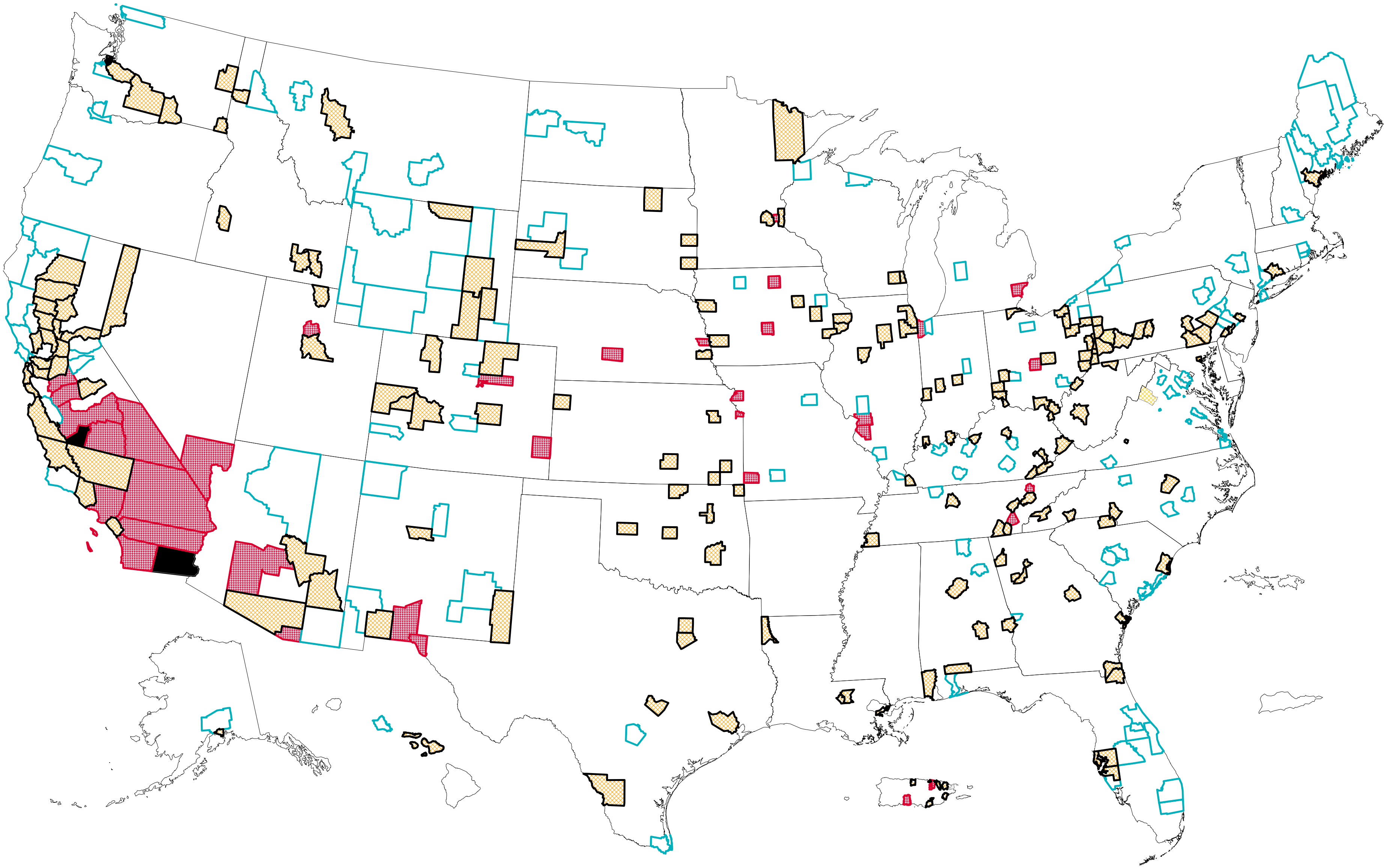
County-level maximum 98th percentile 24-hour average $PM_{2.5}$ concentrations, 2001-2003.



Estimated county-level maximum annual mean $PM_{10-2.5}$ concentrations, 2001-2003.



Estimated county-level maximum 98th percentile 24-hour average $PM_{10\text{-}2.5}$ concentrations, 2001-2003.

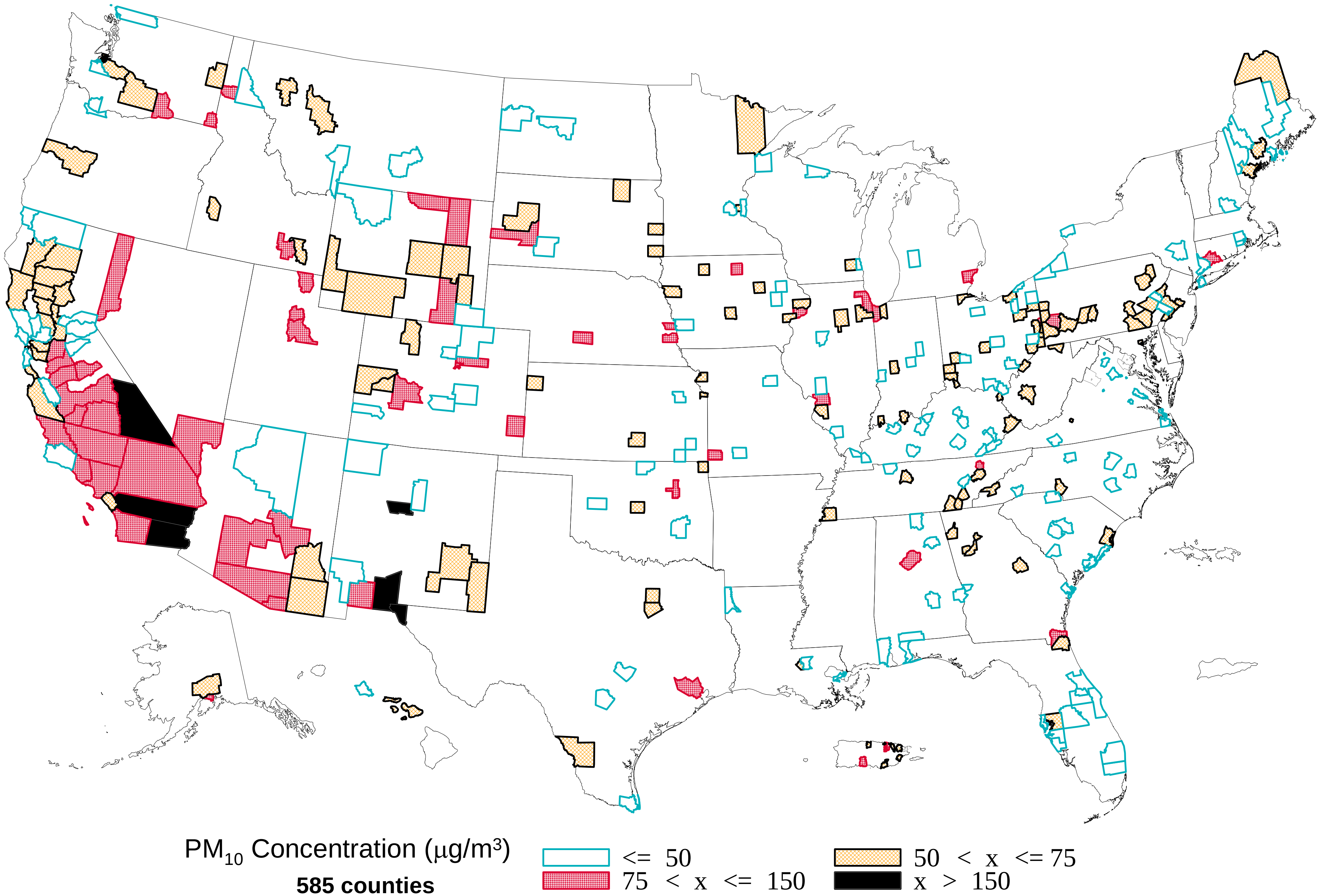


PM_{10} Concentration (µg/m^3)

585 counties

<= 20	20 < x <= 30
30 < x <= 50	x > 50

County-level maximum annual mean PM_{10} concentrations, 2001-2003.



County-level maximum 98th percentile 24-hour average PM_{10} concentrations, 2001-2003.

Episodic Events – PM2.5

- Questions:
  1. What types of events are flagged in AQS?
  2. Are there a significant amount of event-flagged data in AQS?
  3. Do ‘high’ sites (> 15 annual DV) flag more data than ‘low’ sites ($\leq$ 15)?
  4. How do events impact DV’s?
  5. Did any DV’s go from ‘violates the std’ to ‘meets the std’
  6. Are the impacts different for ‘high’ vs ‘low’ sites?
  7. Are data distributions similar for sites that flag data vs. sites that don’t flag data?
  8. What is specific impact in select areas (case studies)

1. What types of events are flagged in AQS?

PM2.5 flag counts, 2001-2003 - All data (meets or not meets completeness)

Flag Description	Event Class	Flag Count	Percent of Event Flags	Number of Sites Reporting
Forest Fire	natural	661	39.9%	273
Sahara Dust	natural	306	18.5%	12
Construction/Demolition	except.	253	15.3%	13
Highway Construction	except.	133	8.0%	4
Volcanic Eruptions	natural	99	6.0%	10
Roofing Operations	except.	51	3.1%	5
Structural Fire	except.	29	1.8%	4
Clean Up After A Major Disaster	except.	29	1.8%	24
Rerouting Of Traffic	except.	27	1.6%	1
High Winds	natural	25	1.5%	14
Infrequent Large Gatherings	except.	17	1.0%	6
Prescribed Burning	except.	17	1.0%	10
Agricultural Tilling	except.	8	0.5%	4
Seismic Activity	natural	1	0.1%	1
Total		1,656		339

PM2.5 Flag Counts, 2001-2003 - Data for sites that meet completeness

Flag Description	Event Class	Flag Count	Percent of Event Flags	Number of Sites Reporting
Forest Fire	natural	490	34.6%	197
Sahara Dust	natural	296	20.9%	10
Construction/Demolition	except.	239	16.9%	10
Highway Construction	except.	133	9.4%	4
Volcanic Eruptions	natural	99	7.0%	10
Roofing Operations	except.	51	3.6%	5
Rerouting Of Traffic	except.	27	1.9%	1
Clean Up After A Major Disaster	except.	26	1.8%	21
High Winds	natural	18	1.3%	10
Infrequent Large Gatherings	except.	14	1.0%	4
Structural Fire	except.	9	0.6%	2
Prescribed Burning	except.	8	0.6%	5
Agricultural Tilling	except.	6	0.4%	2
Seismic Activity	natural	1	0.1%	1
Total		1,417		249

- Most flagged data relate to natural events (~64%)
- Forest fires is the most common event flagged (looking by flag or by site)
- 30% of all complete sites (249 / 827) flagged at least one concentration
- For data from all sites (complete or not): 6 of the 10 highest PM2.5 values were flagged; 41 of the top 100; and 194 of the top 1000
- For data from complete sites: 7 of the 10 highest PM2.5 values were flagged; 39 of the top 100; and 148 of the top 1000

***Complete sites defined as those with 12 quarters of 11+ samples.**

2. Are there a significant amount of event-flagged data in AQS?

Percentage of event-flagged data - for complete sites (827)

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	0.3%	2.6%	3.2%	3.8%	4.8%	7.2%

All complete sites…. But some sites may not flag. (Hence, perhaps biased low). Reference as 'complete sites'

Percentage of event-flagged data - at sites with at least 1 flagged point [249 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	1.1%	8.8%	10.5%	12.7%	15.8%	23.8%

Complete sites with at least one event flag 2001-2003. But some other sites may have flagged if event occurred. (Hence, perhaps biased high). Reference as 'flag sites'

Percentage of event-flagged data at complete sites where RO has at least 1 flagged datapoint (not necessarily at all sites) [423 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	0.7%	5.2%	6.2%	7.5%	9.3%	14.0%

Complete sites where the reporting organization (RO) has flagged at least one event 2001-2003, though not necessarily at all sites. Reference as 'RO flaggers'. [The RO knows how to flaf.]

- Event-flagged data only account for .3% to 1.1% of all reported observations
- However, they account for considerably higher percentage of high values (i.e., values $\geq$ 95th, 96th, 97th, 98th, 99th percentile) ~ 5 %– 14%

3. Do ‘high’ sites (> 15 annual DV) flag more data than ‘low’ sites (< 15)?

Percentage of event-flagged data at complete sites where RO has at least 1 flagged datapoint (not necessarily at all sites) [423 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	0.7%	5.2%	6.2%	7.5%	9.3%	14.0%

Same as previous page (bottom). Break out by high / low.

Percentage of event-flagged data - RO flaggers, sites > 15.0 [58 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	1.0%	4.3%	5.2%	6.8%	8.3%	12.3%

High sites

Percentage of event-flagged data - RO flaggers, sites ≤ 15.0 [365 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	0.6%	5.3%	6.4%	7.6%	9.5%	14.3%

Low sites

- Not much difference

4. How do events impact DV's?

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - complete sites [827 sites]

	Reduction (ug/m3) in Stated Metric					
Site change	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	1.5	9	9	15	18	21
95th percentile	0.2	1	1	1	2	4
75th Percentile	0	0	0	0	0	0
Average	0.04	0.19	0.22	0.28	0.39	0.64
Median	0	0	0	0	0	0
25th percentile	0	0	0	0	0	0
5th Percentile	0	0	0	0	0	0
Minimum	-1.8	0	0	0	0	0

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - flag sites [249 sites]

	Reduction (ug/m3) in Stated Metric					
Site change	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	1.5	9	9	15	18	21
95th percentile	0.4	2	3	3	4	8
75th Percentile	0.2	1	1	1	2	3
Average	0.15	0.62	0.74	0.94	1.30	2.12
Median	0.1	0	0	0	1	1
25th percentile	0.1	0	0	0	0	0
5th Percentile	0	0	0	0	0	0
Minimum	-1.8	0	0	0	0	0

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - RO flaggers [423 sites]

	Reduction (ug/m3) in Stated Metric					
Site change	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	1.5	9	9	15	18	21
95th percentile	0.3	2	2	3	3	6
75th Percentile	0.1	0	1	1	1	1
Average	0.09	0.37	0.43	0.56	0.77	1.25
Median	0	0	0	0	0	0
25th percentile	0	0	0	0	0	0
5th Percentile	0	0	0	0	0	0
Minimum	-1.8	0	0	0	0	0

- The last table (RO flaggers) probably represents the best guess at national average effect.
- On average, removing flagged data would reduce annual DV's by about .1 ug/m3, 98^{th} percentiles by about .8 ug/m3, and 99^{th} percentiles by 1.3 ug/m3
- 25%+ sites would have 1 ug/m3 lower percentiles (96^{th}-99^{th}) if flagged data were omitted

5. Did any DV's go from 'violates the std' to 'meets the std'

- Three complete sites that violate the annual std of 15.0 would meet the std if event-flagged data were excluded
- However, in all 3 situations there exists additional sites that violate the std with or without event data (with much higher DV's)

site	DV all data	DV minus flags	state_name	county_name	csa_name
100031012	15.2	15.0	Delaware	New Castle	Philadelphia-Camden-Vineland, PA-NJ-DE-MD
245100007	15.1	15.0	Maryland	Baltimore (City)	Washington-Baltimore-Northern Virginia, DC-MD-VA-WV
470654002	15.2	15.0	Tennessee	Hamilton	Chattanooga-Cleveland-Athens, TN-GA

- No sites violate the daily std of 65 but would meet it if flagged data were excluded.

6. Are the impacts different for 'high' vs 'low' sites?

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - RO flaggers, sites > 15.0 [58 sites]

Site change	Reduction (ug/m3) in Stated Metric					
	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	0.3	3	3	3	4	7
95th percentile						
75th percentile						
Average	0.05	0.47	0.45	0.47	0.83	1.26
Median	0.1	0	0	0	0	0
25th percentile						
5th Percentile						
Minimum	-1.8	0	0	0	0	0

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - RO flaggers, sites < 15.0 [365 sites]

Site change	Reduction (ug/m3) in Stated Metric					
	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	1.5	9	9	15	18	21
95th percentile						
75th percentile						
Average	0.09	0.35	0.43	0.57	0.76	1.25
Median	0	0	0	0	0	0
25th percentile						
5th Percentile						
Minimum	0	0	0	0	0	0

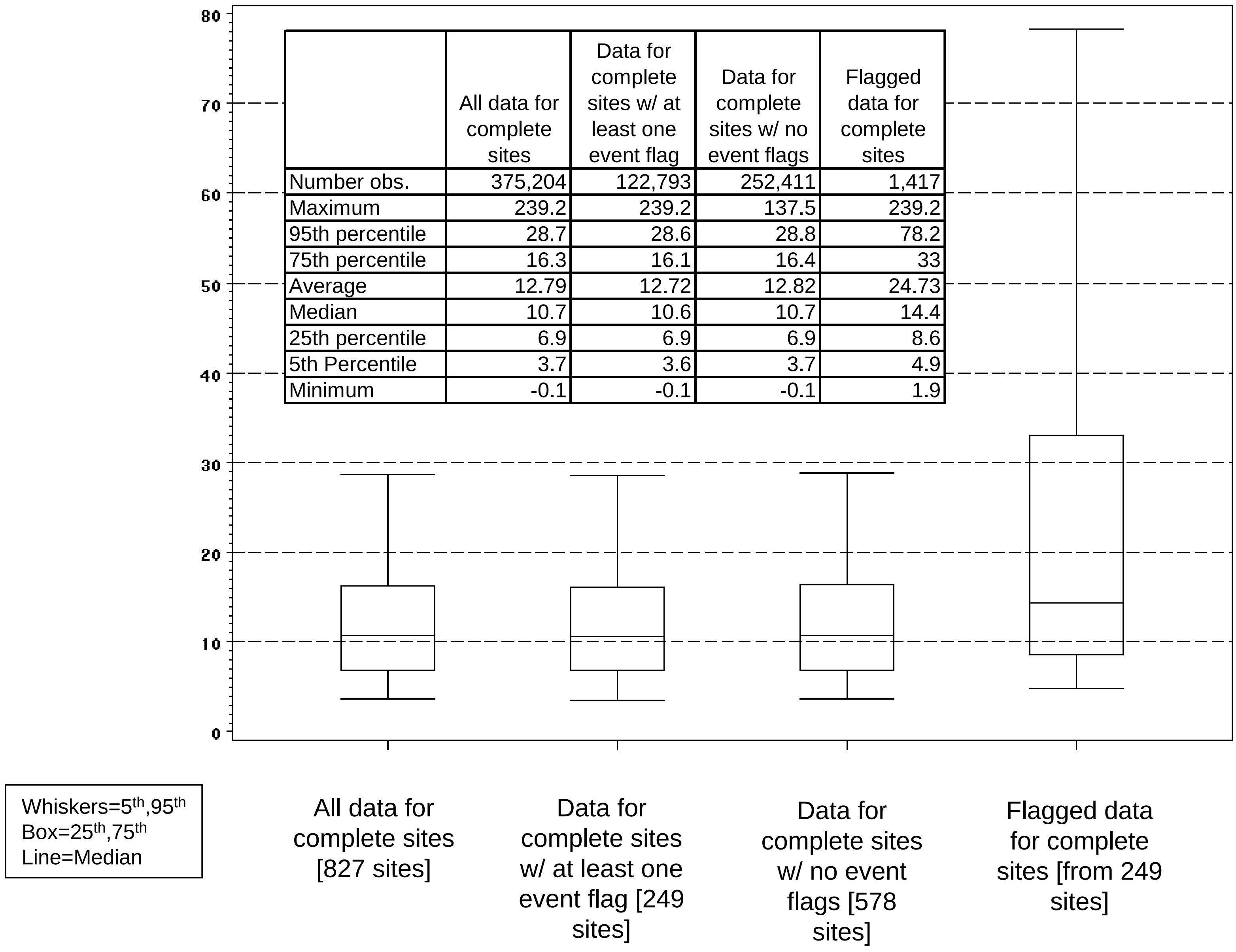
- Not much difference in effect on percentiles DV's
- High sites have a about double the reduction in annual DV's… but still small effect on average (less than .1 ug/m3)
- Some sites have considerable effects

7. Are data distributions similar for sites that flag data vs. sites that don't flag data?

- See next 2 slides
- 2nd slide more accurate comparison
- Not much difference in distributions for flag sites vs. no flag sites.
- But, there are obvious differences in data distributions of all data vs. flagged data.
  - Flagged data generally higher, average concentrations 12.5 - 12.8 for all data (at comp sites) vs. 24.8 for flagged data

7. Are data distributions similar for sites that flag data vs. sites that don't flag data?

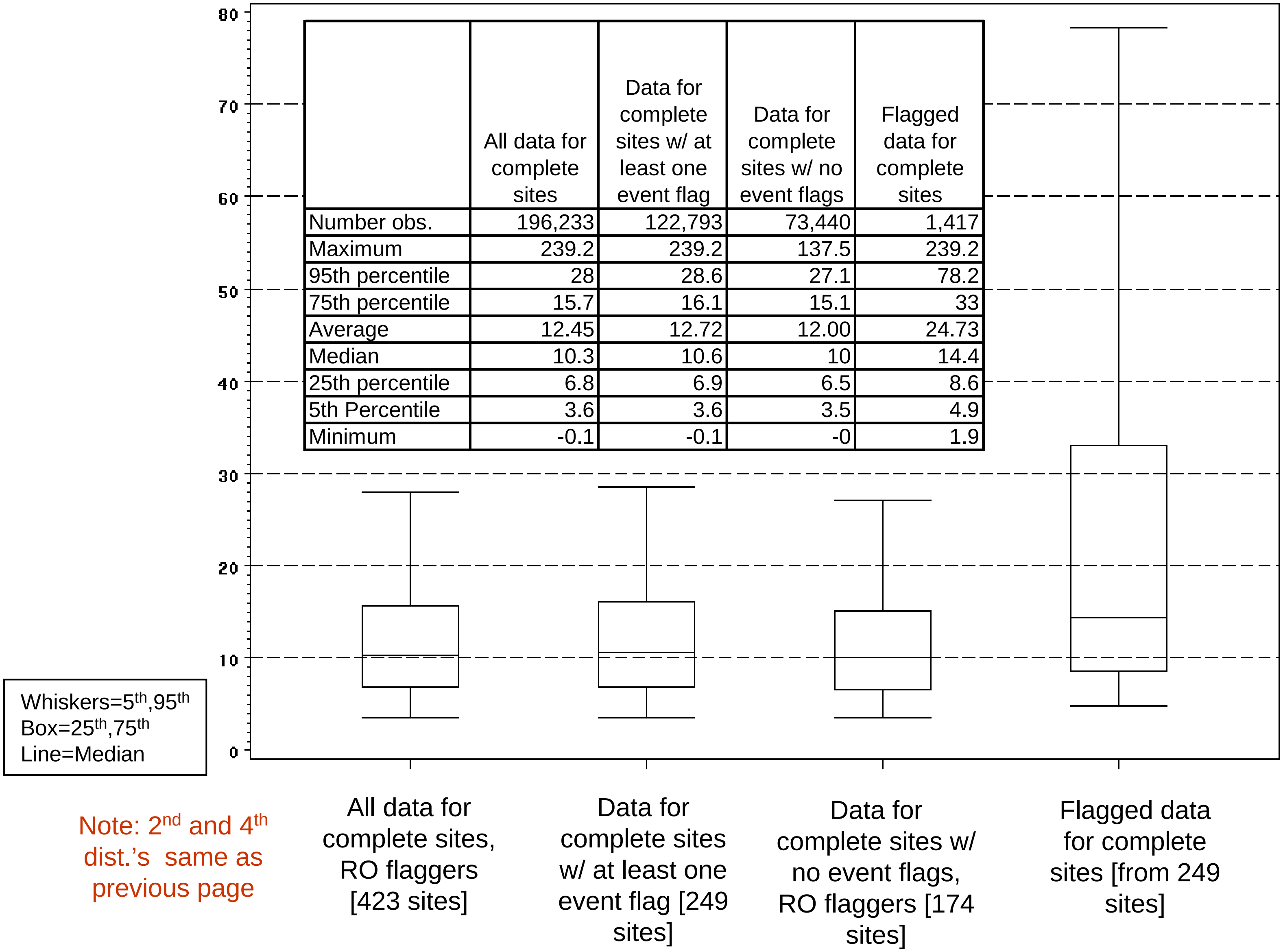
Distribution of PM2.5 concentrations: All data at complete sites, data for complete sites w/ event flags, data for complete sites w/out flags, flagged data from complete sites



	All data for complete sites	Data for complete sites w/ at least one event flag	Data for complete sites w/ no event flags	Flagged data for complete sites
Number obs.	375,204	122,793	252,411	1,417
Maximum	239.2	239.2	137.5	239.2
95th percentile	28.7	28.6	28.8	78.2
75th percentile	16.3	16.1	16.4	33
Average	12.79	12.72	12.82	24.73
Median	10.7	10.6	10.7	14.4
25th percentile	6.9	6.9	6.9	8.6
5th Percentile	3.7	3.6	3.7	4.9
Minimum	-0.1	-0.1	-0.1	1.9

7. Are data distributions similar for sites that flag data vs. sites that don’t flag data?

Distribution of PM2.5 concentrations: All data at complete sites for RO flaggers, data for complete sites w/ event flags (RO flaggers), data for complete sites w/out flags (RO flaggers), flagged data from complete sites



	All data for complete sites	Data for complete sites w/ at least one event flag	Data for complete sites w/ no event flags	Flagged data for complete sites
Number obs.	196,233	122,793	73,440	1,417
Maximum	239.2	239.2	137.5	239.2
95th percentile	28	28.6	27.1	78.2
75th percentile	15.7	16.1	15.1	33
Average	12.45	12.72	12.00	24.73
Median	10.3	10.6	10	14.4
25th percentile	6.8	6.9	6.5	8.6
5th Percentile	3.6	3.6	3.5	4.9
Minimum	-0.1	-0.1	-0	1.9

Note: 2nd and 4th dist.’s same as previous page

8. What is specific impact in select areas (case studies)

San Diego (County, MSA, CSA ~ all same)

Site	060730001			060730003			060730006		
Sample freq.	1-3			1-1			1-1		
# of event flags	2			0			2		
# obs.	340			954			330		
% flagged	0.6%			0.0%			0.6%		
	All data	Minus events	Diff.	All data	Minus events	Diff.	All data	Minus events	Diff.
Annual DV	14.6	14.0	0.6	15.7	15.7	0	12.8	12.4	0.4
95th%ile DV	29	28	1	30	30	0	24	24	0
96th%ile DV	30	30	0	32	32	0	26	25	1
97th%ile DV	33	31	2	33	33	0	27	27	0
98th%ile DV	35	34	1	35	35	0	29	28	1
99th%ile DV	38	38	0	38	38	0	31	30	1

Site	060731002			060731007		
Sample freq.	1-1			1-1		
# of event flags	2			4		
# obs.	978			981		
% flagged	0.2%			0.4%		
	All data	Minus events	Diff.	All data	Minus events	Diff.
Annual DV	15.9	15.9	0	15.9	15.6	0.3
95th%ile DV	33	33	0	33	31	2
96th%ile DV	35	34	1	36	34	2
97th%ile DV	36	36	0	38	37	1
98th%ile DV	38	38	0	41	40	1
99th%ile DV	40	40	0	46	45	1

site	value	date	flag
060730001	239.2	10/27/2003	E
060730006	170.2	10/27/2003	E
060731007	170.1	10/27/2003	E
060731007	104.6	10/26/2003	E
060731002	69.2	10/27/2003	E
060731007	42.9	10/29/2003	E
060731007	8.3	10/30/2003	E
060731002	7.3	10/30/2003	E
060730006	5.9	10/30/2003	E
060730001	5.7	10/30/2003	E

•The 3 highest concentrations (in all US) reported by complete sites 2001-2003 were at SD sites and were flagged for forest fires. [Note: 1 site was ‘down’ during the 10/03 episode]

•Removing the flagged data reduces the annual DV’s from 0 to.6ug/m3 (.4 at DV site); percentile DV’s went down 1 to 2 ug/m3 at the high site.

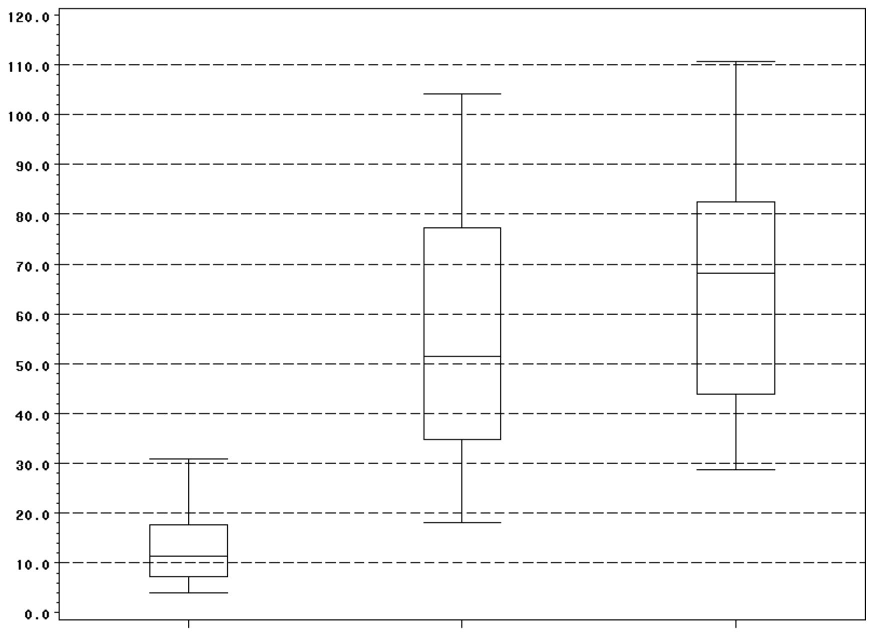
8. What is specific impact in select areas (case studies)

Quebec Fire Event, Effect on PM2.5

- Background:
  - Large smoke plume(s) originating from fires in Quebec, Canada circulated throughout northeast United States in early July 2002 elevating PM2.5 concentrations.
  - General consensus (e.g., CAIR rule modeling) that PM2.5 effects were seen during period, 7/6/2002 through 7/9/2002 in States, NH, VT, CT, RI, MA, NY, NJ, PA, MD, and DE. (Though NC and VA also identified some isolated site-days having effects.) For this analyses, all data during the stated time period and for the stated States were *assumed* to be flagged.
  - Most PM2.5 concentrations for those dates, those states were State flagged (with an 'E' for Forest Fire). Confusion in interpretation of flagging guidance might have prompted some States to not flag. [194 assumed flagged obs, 130 of which were flagged by States]
- Analyses Details:
  - This analyses focuses on the episode effects on 'complete' sites (PM2.5 sites with 11+ samples per Q, all 12 Q's 2001-2003.)
  - 110 complete PM2.5 sites with assumed flags.

8. What is specific impact in select areas (case studies)

Distribution of PM2.5 Concentrations



Concentrations at sites with assumed flags, excluding assumed flagged data (1)

Assumed flagged data (2)

Assumed flagged data also flagged by State (3)

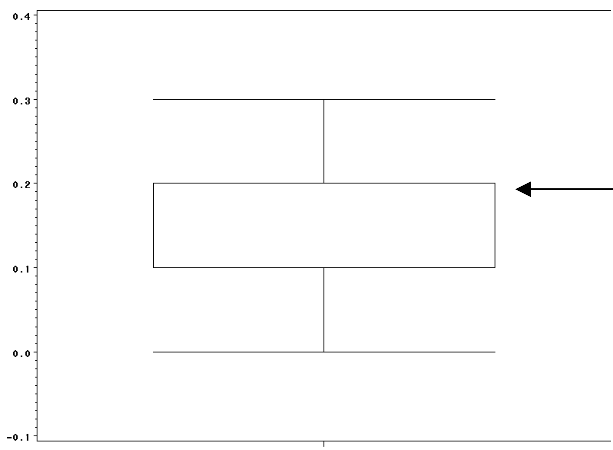
	(1)	(2)	(3)
n	54,716	194	130
MAX	101.7	118.4	118.4
PCT95	30.9	104.1	110.7
PCT75	17.7	77.3	82.5
MEAN	13.7	55.5	66.3
MEDIAN	11.4	51.6	68.3
PCT25	7.3	34.7	43.9
PCT05	4.0	18.1	28.7
MIN	0.0	8.7	11.6

- '2' and '3' distributions *somewhat* similar, hence flag assumption (dates, States) is OK
- 95% of Quebec fire data higher than 75% of other ('normal') data. [But some of 'other' data flagged for other events!]
- Average Quebec episode (assumed) concentration of 55.5ug/m3; median concentration of 51.6 ug/m3
- The ten highest concentrations reported 2001-2003 for the 110 complete (Quebec impacted) sites <u>were</u> associated with the Quebec event: 104.1 - 118.4 ug/m3.

8. What is specific impact in select areas (case studies)

Distribution of Annual Mean DV Differences,

(Mean_alldata minus Mean_minusQuebec)

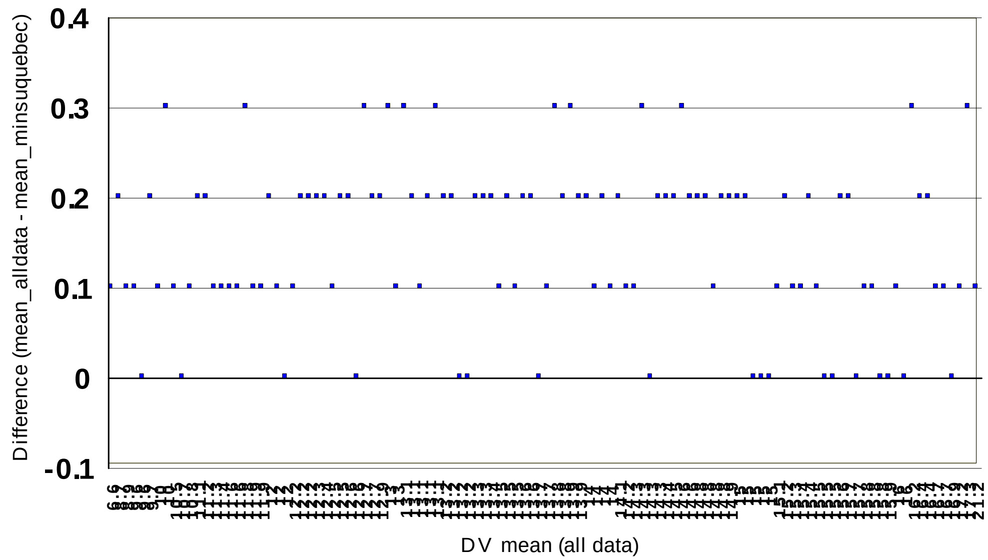


75th and median

- If Quebec fire data were excluded, Annual DV's would drop by 0 to .3 ug/m3.
- The average difference is .15 ug/m3
- The median difference is .2 ug/m3
- Sites with higher DV levels do not have bigger differences; see below

n	110
MAX	0.3
PCT95	0.3
PCT75	0.2
MEAN	0.15
MEDIAN	0.2
PCT25	0.1
PCT05	0
MIN	0

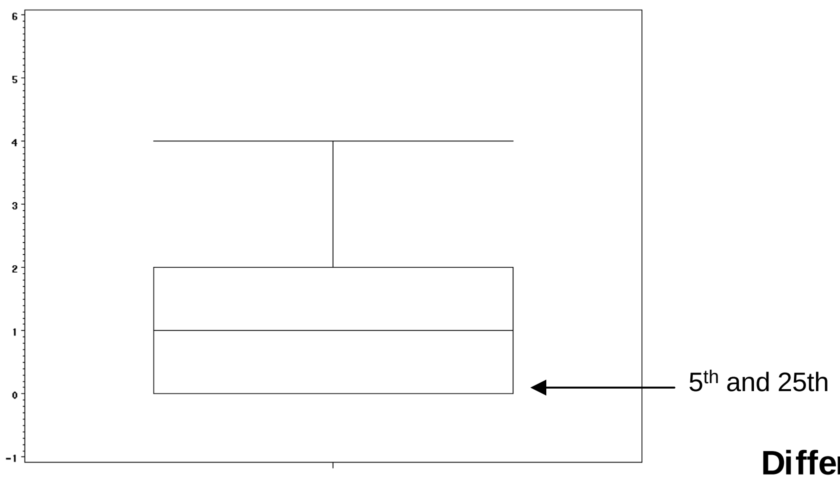
Difference in Annual Mean DV vs. Mean Level



8. What is specific impact in select areas (case studies)

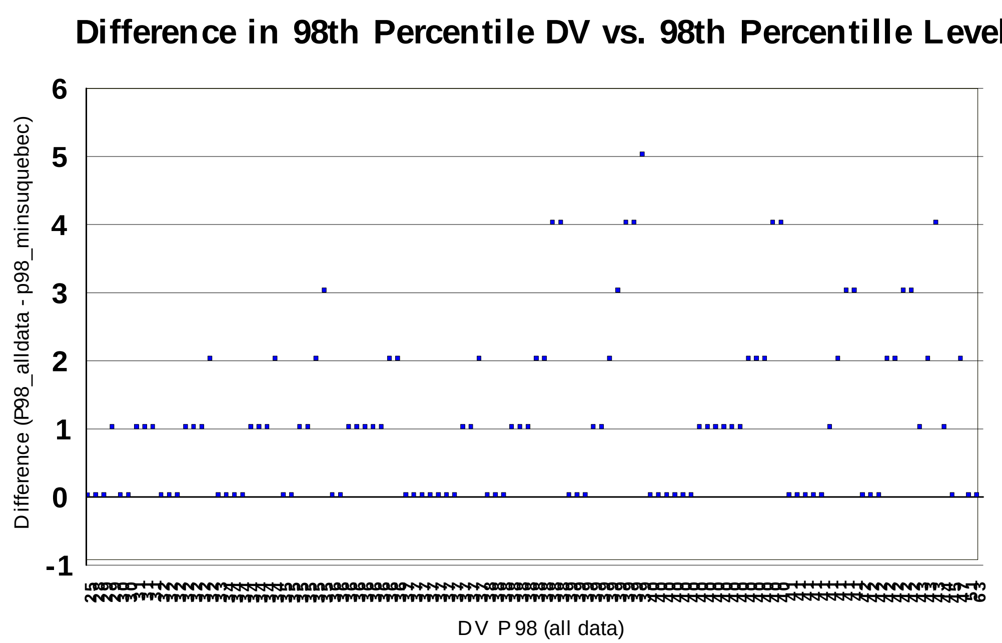
Distribution of 98th Percentile Differences,

(P98_alldata minus P98_minusQuebec)



- If Quebec fire data were excluded, 98th percentile DV's would drop by 0 to 5 ug/m3
- The average difference is 1.07 ug/m3
- The median difference is 1 ug/m3
- Sites with higher DV levels have slightly bigger differences; see below

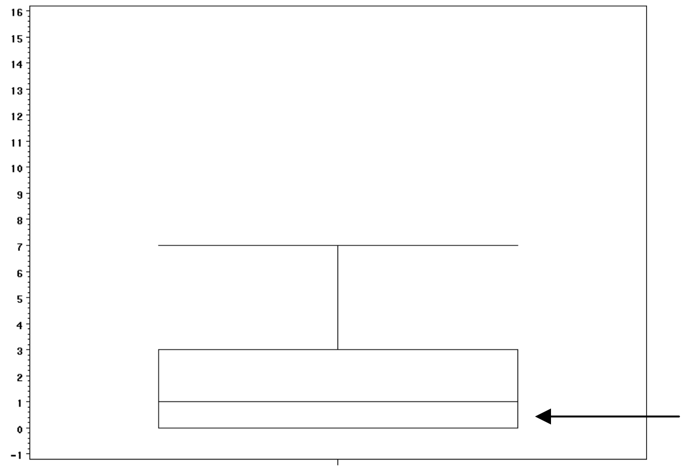
n	110
MAX	5
PCT95	4
PCT75	2
MEAN	1.07
MEDIAN	1
PCT25	0
PCT05	0
MIN	0



8. What is specific impact in select areas (case studies)

Distribution of 99th Percentile Differences,

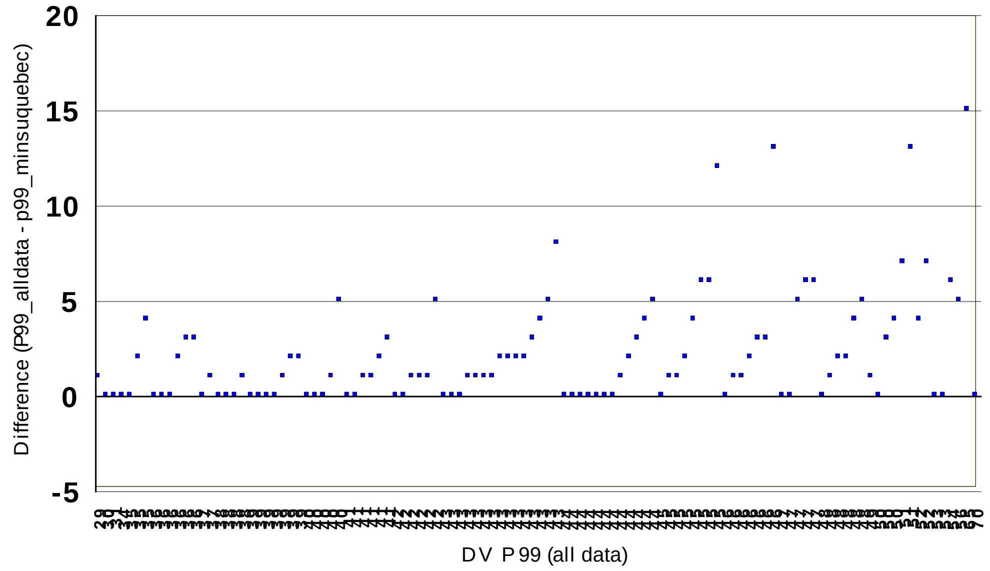
(P99_alldata minus P99_minusQuebec)



- If Quebec fire data were excluded, 99th percentile DV's would drop by 0 to 15 ug/m3
- The average difference is 2.19 ug/m3 (double the avg. diff of 98th)
- The median difference is 1 ug/m3
- Sites with higher DV levels have bigger differences; see below

n	110
MAX	15
PCT95	7
PCT75	3
MEAN	2.19
MEDIAN	1
PCT25	0
PCT05	0
MIN	0

Difference in 99th Percentile DV vs. 99th Percentille Level



8. What is specific impact in select areas (case studies) - Quebec

Crossing Thresholds

Annual Mean DV

	Threshold			
	15	14	13	12
Number of sites with Annual Mean DV (all data) > threshold	26	46	73	87
Number of sites with Annual Mean DV (all data) > threshold but Annual Mean DV (minus Quebec) ≤ threshold	2	2	7	2

98th Percentile DV

	Threshold					
	65	60	55	50	45	40
Number of sites with 98th Percentile DV (all data) > threshold	0	1	1	2	3	24
Number of sites with 98th Percentile DV (all data) > threshold but 98th Percentile DV (minus Quebec) ≤ threshold	0	0	0	0	1	9

99th Percentile DV

	Threshold					
	65	60	55	50	45	40
Number of sites with 99th Percentile DV (all data) > threshold	1	2	3	10	32	79
Number of sites with 99th Percentile DV (all data) > threshold but 99th Percentile DV (minus Quebec) ≤ threshold	0	1	2	6	14	16

- 98th percentile more stable of indicator than 99th percentile

Episodic Events – PM10-2.5

- Questions:
  1. What types of events are flagged in AQS?
  2. Are there a significant amount of event-flagged data in AQS?
  3. Do ‘high’ sites flag more data than ‘low’ sites?
  4. How do events impact DV’s
  5. Are the impacts different for ‘high’ vs ‘low’ sites?
  6. Are data distributions similar for sites that flag data vs. sites that don’t flag data?

1. What types of events are flagged in AQS?

PM10-2.5 Flag Counts, 2001-2003 - Data for sites that meet completeness

Flag Description	Event Class	Flag Count	Percent of Event Flags	Number of Sites Reporting
Forest Fire	natural	228	24.3%	102
Highway Construction	except.	208	22.2%	3
Sahara Dust	natural	189	20.1%	6
Construction/Demolition	except.	107	11.4%	14
High Winds	natural	90	9.6%	34
Volcanic Eruptions	natural	60	6.4%	6
Roofing Operations	except.	15	1.6%	3
Infrequent Large Gatherings	except.	14	1.5%	7
Rerouting Of Traffic	except.	13	1.4%	1
Agricultural Tilling	except.	7	0.7%	3
Clean Up After A Major Disaster	except.	4	0.4%	4
Sandblasting	except.	1	0.1%	1
Prescribed Burning	except.	1	0.1%	1
Seismic Activity	natural	1	0.1%	1
Total		938		146

Notes:
- **Complete sites defined as those with 12, 8, or 4 consecutive quarters of 11+ samples. [489 sites total]**
- **PM10-2.5 flag was set to the PM10 flag if it exists, else to the PM2.5 flag if it exists.**

- Most flagged data relate to natural events (~ 60%)
- Forest fires is the most common event flagged (looking by flag) but ‘highway construction, ‘Sahara dust’, ‘Construction/Demolition’, and ‘High Winds’ area also common (over 100+ site days)
- Looking by site, ‘Forest Fires’ affected the most sites, followed by ‘High Winds’
- 30% of all sites (146 / 489) reported at least one event flag.
- 2 of the 10 highest PMc values were flagged; 37 of the top 100; and 99 of the top 1000
- All ‘Volcanic Eruptions’ and ‘Sahara Dust’ flags are associated with monitoring site in Puerto Rico and the Virgin Islands

2. Are there a significant amount of event-flagged data in AQS?

Percentage of event-flagged data - for complete sites [489]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	0.9%	3.8%	4.2%	4.8%	5.3%	7.2%

All complete sites…. But some sites may not flag. (Hence, perhaps biased low). Reference as ‘complete sites’

Percentage of event-flagged data - at sites with at least 1 flagged point [146 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	3.1%	12.8%	14.0%	16.0%	17.9%	24.1%

Complete sites with at least one event flag 2001-2003. But some other sites may have flagged if event occurred. (Hence, perhaps biased high). Reference as ‘flag sites’

Percentage of event-flagged data at complete sites where RO has at least 1 flagged datapoint (not necessarily at all sites) [327 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	1.4%	5.7%	6.2%	7.2%	8.0%	10.8%

Complete sites where the reporting organization (RO) has flagged at least one event 2001-2003, though not necessarily at all sites. Reference as ‘RO flaggers’. [The RO knows how to flag.]

- Event-flagged data account for .9% to 3.1% of all reported observations. (about triple the PM2.5 rate)
- They account for considerably higher percentage of high values (i.e., values ≥ 95th, 96th, 97th, 98th, 99th percentile)

3. Do ‘high’ sites (≥ 50 98th percentile*) flag more data than ‘low’ sites (< 50 98th percentile*)?

Percentage of event-flagged data at complete sites where RO has at least 1 flagged datapoint (not necessarily at all sites) [327 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	1.4%	5.7%	6.2%	7.2%	8.0%	10.8%

Same as previous page (bottom), ‘RO flaggers’. Break out by high / low.

Percentage of event-flagged data - RO flaggers, sites ≥ 50 [68 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	3.7%	14.9%	16.5%	19.4%	21.4%	29.6%

High sites

Percentage of event-flagged data - RO flaggers, sites < 50 [259 sites]

	Percent of All Samples	Percent of Samples Equal or Above Percentile Metric				
		95th	96th	97th	98th	99th
Site Average	0.7%	3.3%	3.5%	3.9%	4.5%	5.9%

Low sites

- High sites flag more data. (The flagged data makes them ‘high’ sites.). They flag about 5 times in total (on average) and also, 5 times the number of extreme values

** Approximately 20% of the 489 sites in the 2001-2003 PM10-2.5 database have a 98th percentile ≥ 50.*

4. How do events impact DV's?

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - complete sites [489 sites]

Site change	Reduction (ug/m3) in Stated Metric					
	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	10.8	58	137	215	202	189
95th percentile	1.3	3	3	5	8	21
75th Percentile	0.11	0	0	0	0	0
Average	0.17	0.68	0.97	1.41	1.65	3.30
Median	0.01	0	0	0	0	0
25th percentile	-0.06	0	0	0	0	0
5th Percentile	-0.29	0	0	0	0	0
Minimum	-1.39	0	0	0	0	0

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - flag sites [146 sites]

Site change	Reduction (ug/m3) in Stated Metric					
	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	10.76	58	137	215	202	189
95th percentile	2.33	6	11	20	23	80
75th Percentile	0.51	2	2	2	3	8
Average	0.52	2.29	3.25	4.72	5.53	11.05
Median	0.09	0	0	0	0	0
25th percentile	-0.02	0	0	0	0	0
5th Percentile	-0.23	0	0	0	0	0
Minimum	-1.39	0	0	0	0	0

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - RO flaggers [327 sites]

Site change	Reduction (ug/m3) in Stated Metric					
	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	10.8	58	137	215	202	189
95th percentile	1.6	5	5	10	12	26
75th Percentile	0.14	0	0	0	0	0
Average	0.25	1.02	1.45	2.11	2.47	4.93
Median	0.02	0	0	0	0	0
25th percentile	-0.05	0	0	0	0	0
5th Percentile	-0.28	0	0	0	0	0
Minimum	-1.39	0	0	0	0	0

- The last table (RO flaggers) probably represents the best guess at national average effect.
- On average, removing flagged data would reduce annual DV's by about .25 ug/m3, 98^{th} percentiles by about 2-3 ug/m3, and 99^{th} percentiles by 4-5 ug/m3.
- Some sites would have very large changes in in percentiles (95^{th}-99^{th}) if flagged data were omitted; see max and 95^{th}%ile site change rows

5. Are the impacts different for 'high' vs 'low' sites?

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - RO flaggers, sites ≥ 50 [68 sites]

	Reduction (ug/m3) in Stated Metric					
Site change	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	10.8	58	137	215	202	189
95th percentile						
75th percentile						
Average	0.88	3.82	5.94	8.76	10.28	19.65
Median	0.24	0.5	0	0.5	0	0
25th percentile						
5th Percentile						
Minimum	-1.39	0	0	0	0	0

Reductions (µg/m3) in annual and 24-hour design values as a result of exempting event-flagged data - RO flaggers, sites < 50 [259 sites]

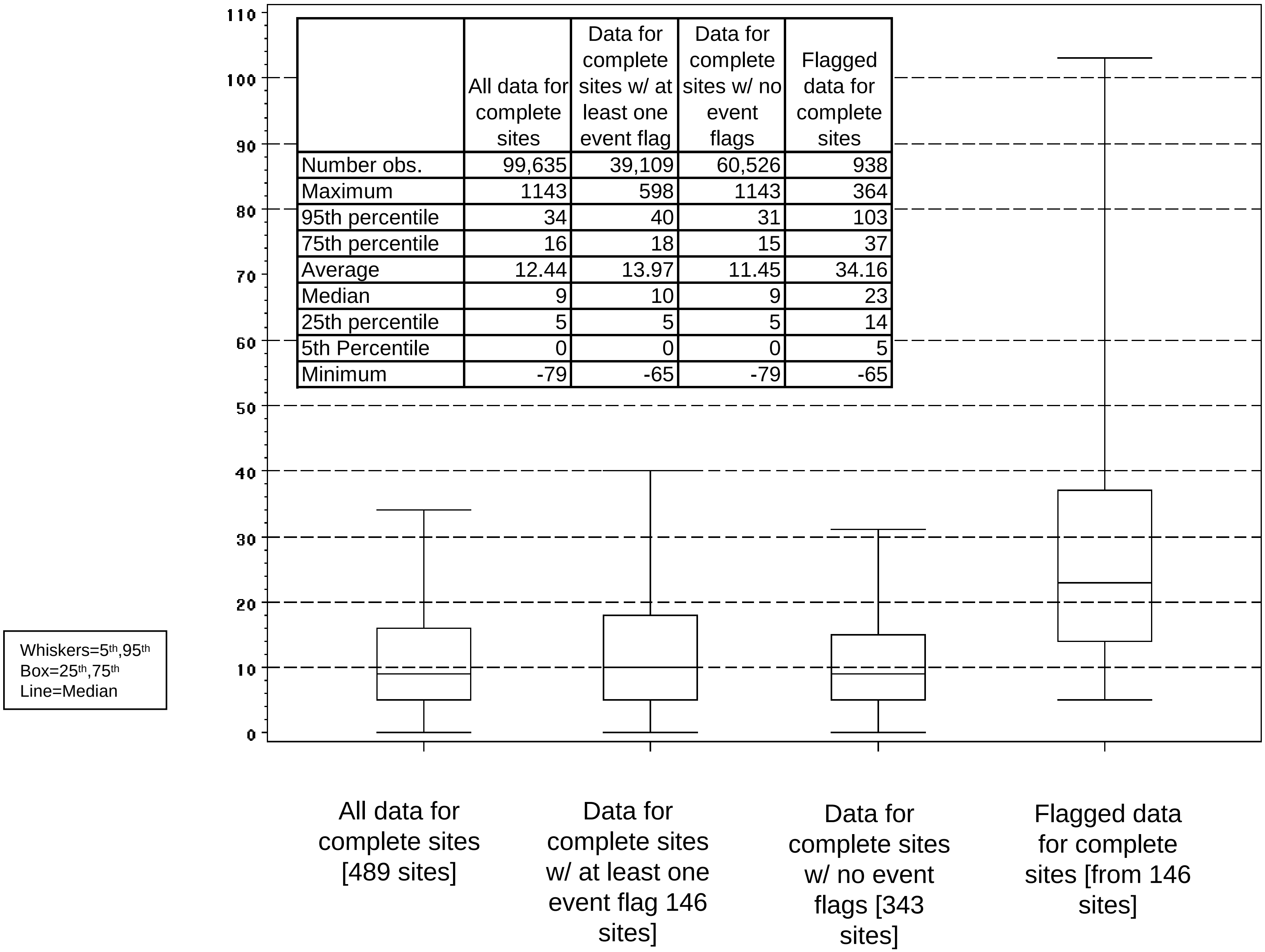
	Reduction (ug/m3) in Stated Metric					
Site change	Annual DV	95th Percentile	96th Percentile	97th Percentile	98th Percentile	99th Percentile
Maximum	3.35	24	24	24	24	37
95th percentile						
75th percentile						
Average	0.09	0.29	0.27	0.36	0.42	1.07
Median	0.01	0	0	0	0	0
25th percentile						
5th Percentile						
Minimum	-1.11	0	0	0	0	0

- Apparent differences in effect on annual DV and percentile DV's
- High sites have about ten times the reduction in annual DV's… about .9ug/m3 on average
- High sites have 10-20 times the reduction in percentile DV's
- Some sites (high and low) have considerable effects

6. Are data distributions similar for sites that flag data vs. sites that don't flag data?

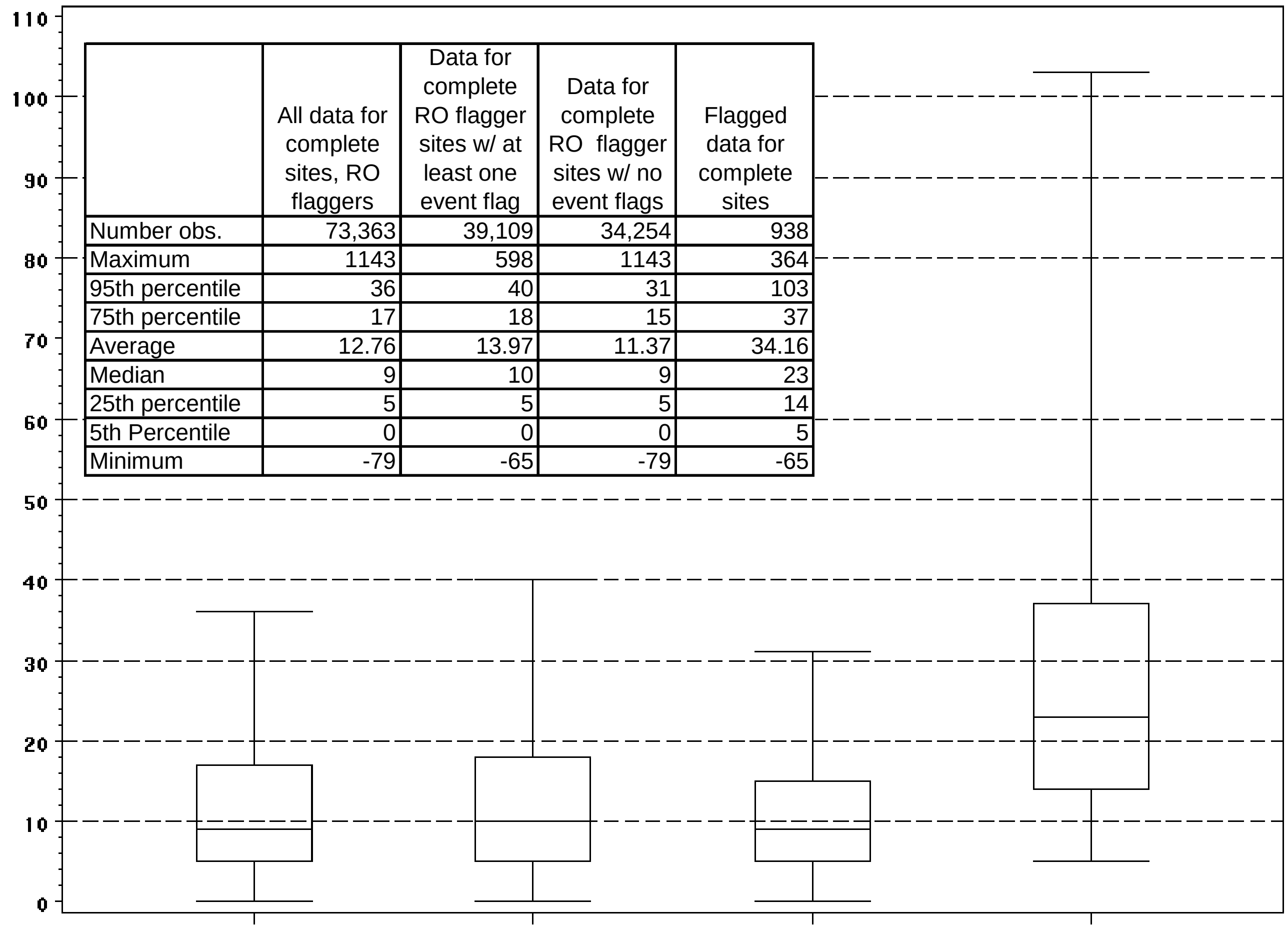
- See next 2 slides
- 2nd slide more accurate comparison (RO flaggers)
- Some differences on high end of distributions (flag sites vs. no flag sites). Flag site data higher.
- Obvious differences in data distributions of all data vs. flagged data.
    - Flagged data generally higher, average concentration is 12.4 - 12.8 for all data (at comp sites) vs. 34.1 for flagged data

7. Distribution of PM10-2.5 concentrations: All data at complete sites, data for complete sites w/ event flags, data for complete sites w/out flags, flagged data from complete sites



	All data for complete sites	Data for complete sites w/ at least one event flag	Data for complete sites w/ no event flags	Flagged data for complete sites
Number obs.	99,635	39,109	60,526	938
Maximum	1143	598	1143	364
95th percentile	34	40	31	103
75th percentile	16	18	15	37
Average	12.44	13.97	11.45	34.16
Median	9	10	9	23
25th percentile	5	5	5	14
5th Percentile	0	0	0	5
Minimum	-79	-65	-79	-65

7. Distribution of PM10-2.5 concentrations: All data at complete sites for RO flaggers, data for complete sites w/ event flags (RO flaggers), data for complete sites w/out flags (RO flaggers), flagged data from complete sites



	All data for complete sites, RO flaggers	Data for complete RO flagger sites w/ at least one event flag	Data for complete RO flagger sites w/ no event flags	Flagged data for complete sites
Number obs.	73,363	39,109	34,254	938
Maximum	1143	598	1143	364
95th percentile	36	40	31	103
75th percentile	17	18	15	37
Average	12.76	13.97	11.37	34.16
Median	9	10	9	23
25th percentile	5	5	5	14
5th Percentile	0	0	0	5
Minimum	-79	-65	-79	-65

Whiskers=5th,95th
Box=25th,75th
Line=Median

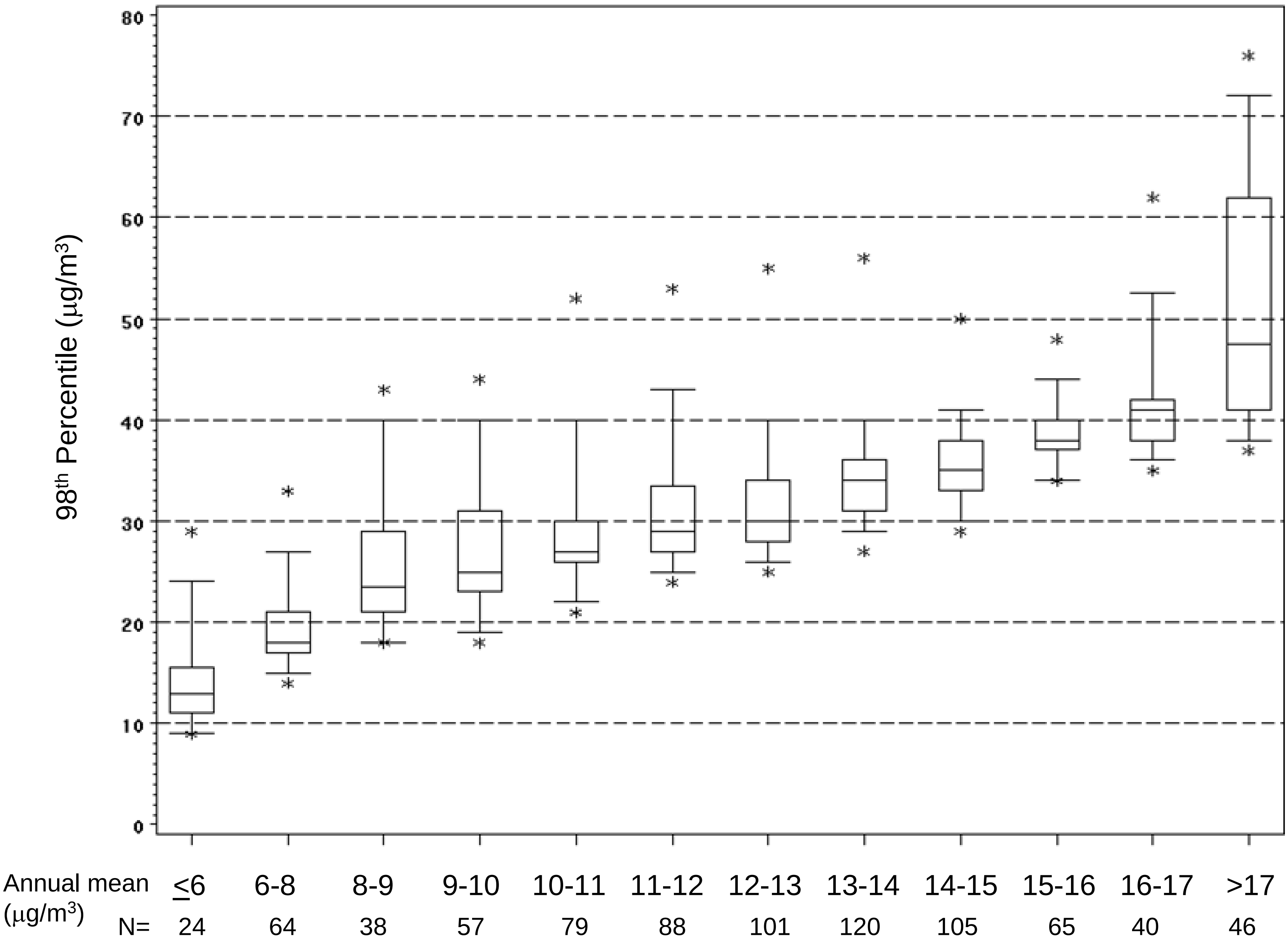
All data for complete sites, RO flaggers [327 sites]

Data for complete sites w/ at least one event flag [146 sites]

Data for complete sites w/ no event flags, RO flaggers [181 sites]

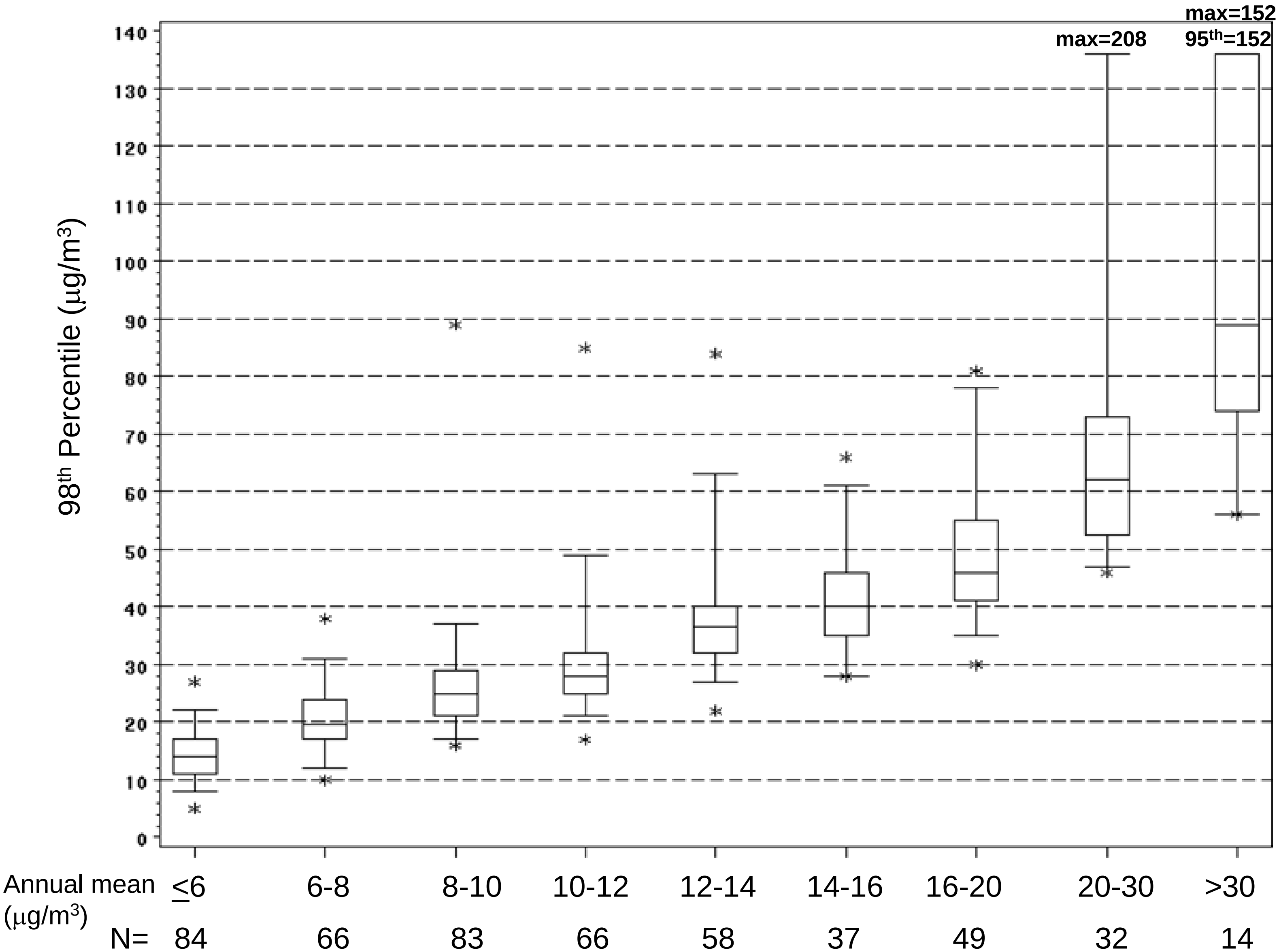
Flagged data for complete sites [from 146 sites]

Note: 2nd and 4th dist.'s same as previous page



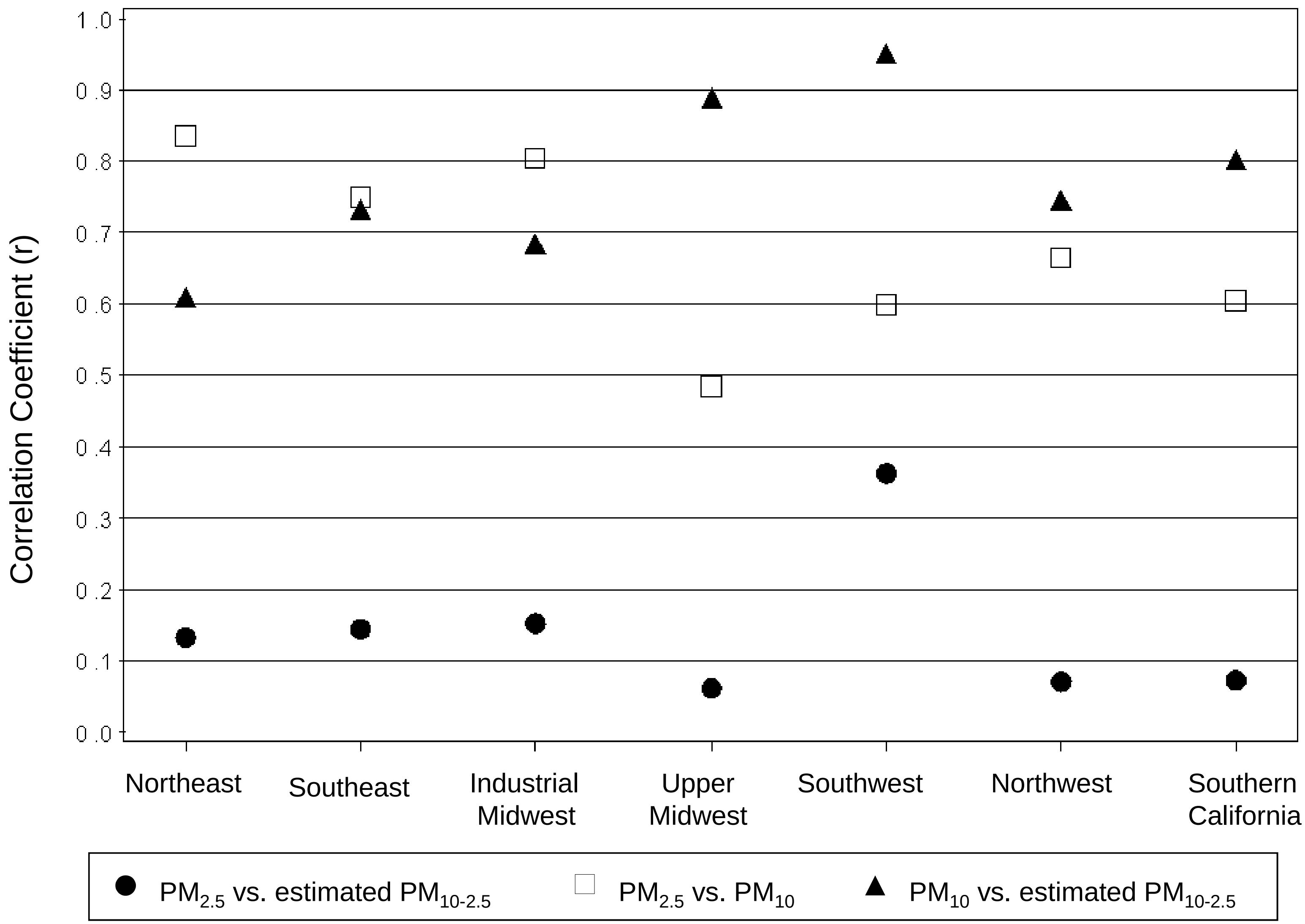
Distribution of annual mean vs. 98th percentile 24-hour average $PM_{2.5}$ concentrations, 2001-2003.

Box depicts interquartile range and median; whiskers depict 5th and 95th percentiles; asterisks depict minima and maxima. N= number of sites.

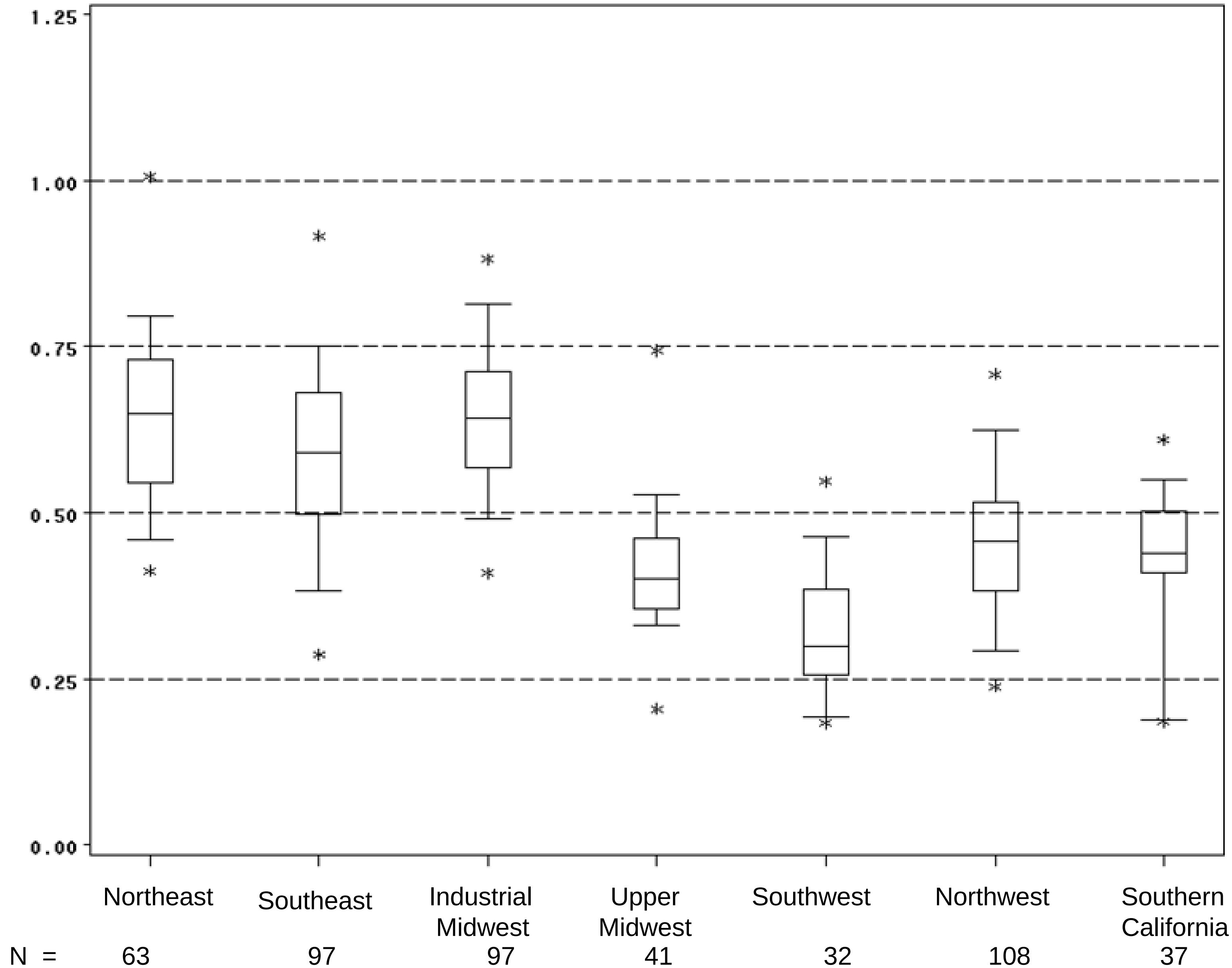


Distribution of estimated annual mean vs. 98th percentile 24-hour average $PM_{10\text{-}2.5}$ concentrations, 200- 2003.

Box depicts interquartile range and median; whiskers depict 5th and 95th percentiles; asterisks depict minima and maxima. N= number of sites.



Regional average correlations of 24-hour average PM by size fraction.



Distribution of ratios of $PM_{2.5}$ to PM_{10} by region, 2001-2003.

Box depicts interquartile range and median; whiskers depict 5th and 95th percentiles; asterisks depict minima and maxima. N = number of sites.

PM2.5 Spatial Averaging

- Questions:
  - Are there large differences between ‘regular’ (highest site in area) DV’s and spatial average (SA) DV’s?
  - What is the population in areas that could use SA (utilizing current criteria).
  - Would tightening the criteria provide more protection?
  - Are the would-be violating sites in an area that could utilize SA located in lower-income, high percentage-minority, and/or lower education locations than the overall area?
- Analyses details:
  - Started with the default SP PM2.5 database (all sites with 11+ samples in each of the the 12 quarters 2001-2003). Eliminated microscale sites that are not (officially) compared to annual std.
  - Initially enforced the CFR spatial criteria of 1) .6 overall correlation between sites, and 2) no more than 20% difference in site annual mean and spatial annual mean. Did not check criterion that all SA sites should be impacted by similar emissions.
  - Enforced CFR data handling requirement that if SA annual mean is less than or equal the annual std, then only SA sites with 75%+ capture each of the 4 Q’s would have their annual mean included in the spatial annual average (Only 11+ samples required in each of the 4 Q’s if the spatial annual mean was greater than the annual std.) Changed level of std (and completeness check) from 15 to 14 for accurate evaluation of SA effect on those std levels.
  - Constructed SA set of sites by initially considering all sites in the area (CSA, CBSA, or STCOU). If a site-pair correlation was less than cutoff, the lower DV site was eliminated. If e remaining set did not meet annual mean difference criterion then lowest DV site was omitted from set and revised set tried. Continued until reduced set of sites met criteria or less than 2 sites left. Note: Undoubtedly, different combinations of sites (selected w/ rationale and/or at random) could/would meet criteria and yield different results.
  - Only considered (for SA) areas with a regular DV > annual std. level <u>and</u> spatial DV $\geq$ any (valid) non-SA site DV in the area
  - Evaluated appropriateness of .6 (correlation) and 20% (max difference in annual means) levels
  - Tightened the correlation criterion to .9 and the annual mean difference criterion to 10% to evaluate changes in results.

Statistics for Areas that Qualify for Spatial Averaging;

Current Criteria (.6 corr., 20% diff in means), NAAQS Levels of 15, 14, 13

Using criteria of .6 correlation and +/- 20 % difference in annual means. Using annual std level of 15.0				Using criteria of .6 correlation and +/-20 % difference in annual means. Using annual std level of 14.0			
		Could use spatial averaging	Could use SA to meet 15.0 annual std			Could use spatial averaging	Could use SA to meet 14.0 annual std
Number of areas		32	10	Number of areas		45	7
Total population		50,645,671	14,254,268	Total population		63,848,777	8,932,198
Area distribution statistics:				Area distribution statistics:			
Difference in area DV's (ug/m3)	mean	1.06	0.84	Difference in area DV's (ug/m3)	mean	1.01	1.21
	max	2.8	1.5		max	2.8	2.1
	p95	2.7	1.5		p95	2.6	2.1
	p75	1.5	1.2		p75	1.5	2.0
	med	0.9	0.8		med	0.8	1.1
	p25	0.5	0.5		p25	0.4	0.6
	p05	0.2	0.2		p05	0.2	0.2
	min	0.2	0.2		min	0.0	0.2

- Under existing criteria (only considering minimum site correlation and maximum difference in annual means) and considering NAAQS levels of 15 and 14, 32-45 metropolitan areas with a combined population of 51-64 million could qualify for spatial averaging (SA). Note that most of these areas would only lower their area DV and still not attain the standard. But, a lower DV would help these areas attain more quickly, and there are also data capture (less stringent) benefits.
- Assuming these areas could pass (required) additional scrutiny, they would lower their areas DV's by up to 2.8 ug/m3. (Average reduction in area DV = 1 to 1.1 ug/m3)
- 7-10 of these areas would meet the annual std NAAQS (15 or 14 level) with their spatial average when they couldn't with their regular site-based DV. Average reduction in DV for these areas is .8-1.2 ug/m3. 9-14 million people live in these areas.

See area listings 1 & 2 next…..

Listing 1: Areas that Qualify for Spatial Averaging; Current Criteria (.6 corr., 20% diff in means), NAAQS Level of 15

Area	Pop.	Number of Sites in Area	Number of Sites in CMZ	Design value without SA	Design value with SA	Difference in DV's	Minimum area site DV	Maximum between-site difference in means	Minimum between-site correlation (annual)	High Site Census Tract Information					Other Site Census Tract(s) Information (avg.)					Area (CSA/CBSA) Information				
										Percent minority	Per capita income	Median Household Income	Median Family income	Average Education Level Attained*	Percent minority	Per capita income	Median Household Income	Median Family income	Average Education Level Attained*	Percent minority	Per capita income	Median Household Income	Median Family Income	Average Education Level Attained*
CBSA_Bakersfield, CA	661,645	5	3	21.8	21.0	0.8	20.3	7.9%	0.98	46%	$11,843	$18,777	$22,669	4.5	32%	$15,947	$33,390	$37,965	5.2	38%	$15,780	$38,858	$42,458	5.0
CBSA_Canton-Massillon, OH	406,934	2	2	17.3	16.6	0.7	15.8	5.6%	0.99	9%	$12,577	$24,205	$30,833	4.4	37%	$14,201	$10,457	$25,000	4.2	9%	$20,154	$36,917	$43,005	5.1
CBSA_Charleston, WV	309,635	2	2	17.1	16.3	0.8	15.5	5.3%	0.97	9%	$16,667	$20,929	$32,167	4.7	20%	$28,021	$27,217	$50,690	5.7	7%	$19,090	$29,508	$35,875	5.0
CBSA_Evansville, IN-KY	342,815	3	3	15.5	15.3	0.2	15.2	3.7%	0.96	11%	$12,773	$29,033	$36,446	4.9	8%	$23,162	$31,037	$46,836	5.1	8%	$20,026	$38,956	$46,128	5.1
CBSA_Hagerstown-Martinsburg, MD-WV	222,771	2	2	16.3	15.1	1.2	14.0	11.0%	0.80	17%	$14,688	$25,423	$35,591	4.3	73%	$21,284	$0	$0	8.9	9%	$19,222	$36,997	$42,510	5.1
CBSA_Huntington-Ashland, WV-KY-OH	288,649	3	3	16.6	15.8	0.8	15.0	8.3%	0.87	12%	$4,312	$6,624	$5,357	7.4	3%	$19,748	$32,969	$38,206	4.9	4%	$16,631	$29,341	$36,169	4.9
CBSA_San Diego-Carlsbad-San Marcos, CA	2,813,833	5	5	15.9	15.0	0.9	14.6	16.4%	0.66	46%	$10,278	$21,021	$23,870	5.1	29%	$16,989	$40,702	$46,701	5.1	33%	$22,928	$51,773	$57,106	5.6
CBSA_Weirton-Steubenville, WV-OH	132,008	4	4	17.8	17.0	0.8	16.2	7.8%	0.86	5%	$15,980	$30,000	$40,181	4.8	6%	$17,242	$33,295	$40,576	4.6	5%	$16,909	$32,335	$39,252	4.9
CBSA_Wheeling, WV-OH	153,172	2	2	15.7	15.4	0.3	15.2	1.9%	0.95	1%	$17,077	$31,836	$39,033	4.8	14%	$8,072	$7,663	$23,214	3.7	4%	$16,749	$29,113	$36,899	5.0
CSA_Birmingham-Hoover-Cullman, AL	1,129,721	8	4	18.0	16.0	2.0	14.7	13.5%	0.77	99%	$12,938	$16,995	$23,333	4.3	16%	$21,918	$45,552	$51,117	5.2	28%	$20,390	$36,593	$43,526	5.1
CSA_Chattanooga-Cleveland-Athens, TN-GA	629,561	3	3	15.6	15.4	0.2	15.2	3.6%	0.87	6%	$14,092	$23,713	$29,183	4.1	24%	$13,257	$22,338	$35,768	4.2	14%	$19,278	$33,613	$39,509	5.0
CSA_Chicago-Naperville-Michigan City, IL-IN-WI	9,312,255	28	2	17.7	17.5	0.2	17.3	4.6%	0.84	10%	$12,368	$31,156	$30,189	4.7	29%	$20,950	$45,553	$53,509	5.1	33%	$24,491	$52,263	$59,135	5.4
CSA_Cincinnati-Middletown-Wilmington, OH-KY-I	2,050,175	12	12	17.8	16.0	1.8	14.5	13.9%	0.90	13%	$19,121	$27,364	$36,667	4.4	22%	$17,950	$31,444	$38,807	4.9	14%	$22,786	$43,248	$49,355	5.3
CSA_Cleveland-Akron-Elyria, OH	2,945,831	13	11	18.3	15.9	2.4	14.2	19.2%	0.84	31%	$15,270	$25,221	$26,850	5.6	41%	$15,278	$28,755	$32,732	4.7	21%	$22,321	$46,452	$53,471	5.3
CSA_Columbus-Auburn-Opelika, GA-AL	420,965	3	3	15.3	14.6	0.7	14.3	8.2%	0.78	65%	$7,295	$10,121	$11,949	3.3	78%	$11,574	$18,636	$23,013	4.2	42%	$17,184	$31,978	$37,256	5.1
CSA_Columbus-Marion-Chillicothe, OH	1,835,189	3	3	16.7	16.2	0.5	15.9	5.9%	0.95	88%	$14,293	$21,486	$27,560	3.9	39%	$15,184	$28,309	$30,408	4.7	17%	$22,256	$45,186	$51,028	5.5
CSA_Dayton-Springfield-Greenville, OH	1,085,094	3	3	15.2	14.7	0.5	14.7	8.5%	0.93	6%	$17,457	$32,708	$40,117	5.3	7%	$16,186	$26,815	$34,558	4.9	16%	$21,263	$42,919	$49,338	5.3
CSA_Detroit-Warren-Flint, MI	5,357,538	14	6	19.5	16.8	2.7	15.1	18.5%	0.83	29%	$7,573	$19,713	$24,031	3.9	43%	$17,486	$35,422	$40,804	5.1	27%	$24,353	$53,256	$60,632	5.4
CSA_Fairmont-Clarksburg, WV	148,742	2	2	15.4	14.7	0.7	14.0	5.6%	0.96	3%	$13,328	$21,839	$28,906	4.6	8%	$14,417	$16,590	$30,031	4.3	4%	$16,094	$28,602	$34,255	4.9
CSA_Fresno-Madera, CA	922,516	2	2	19.7	19.5	0.2	19.2	3.1%	0.97	45%	$12,781	$31,131	$34,440	4.6	57%	$10,976	$16,842	$20,804	4.0	45%	$15,388	$36,870	$39,680	4.7
CSA_Greensboro--Winston-Salem--High Point, N	1,283,856	4	4	15.8	14.6	1.2	14.0	8.6%	0.93	50%	$19,691	$28,094	$34,320	4.6	42%	$25,501	$35,913	$47,006	5.1	25%	$21,090	$38,066	$45,213	5.0
CSA_Harrisburg-Carlisle-Lebanon, PA	629,401	2	2	15.8	14.9	0.9	15.8	13.5%	0.92	35%	$15,752	$31,557	$37,679	4.9	1%	$18,897	$44,341	$50,259	5.1	12%	$21,939	$42,855	$50,094	5.3
CSA_Indianapolis-Anderson-Columbus, IN	1,843,588	6	6	16.7	15.2	1.5	13.6	12.0%	0.92	40%	$9,869	$18,988	$20,417	4.4	17%	$18,785	$36,313	$41,702	5.3	16%	$22,715	$46,925	$53,537	5.4
CSA_Knoxville-Sevierville-La Follette, TN	779,013	5	5	16.7	15.6	1.1	14.2	11.4%	0.85	35%	$7,364	$11,305	$13,239	3.4	11%	$17,905	$35,858	$42,976	5.0	8%	$20,034	$33,904	$40,386	5.1
CSA_Lexington-Fayette--Frankfort--Richmond, K'	602,773	4	4	15.7	14.3	1.4	13.5	9.3%	0.75	28%	$10,418	$17,111	$18,679	5.6	30%	$17,721	$28,083	$36,300	4.8	12%	$20,520	$37,223	$43,417	5.3
CSA_Louisville-Elizabethtown-Scottsburg, KY-IN	1,292,482	6	6	16.9	15.6	1.3	14.1	12.4%	0.85	11%	$13,959	$25,315	$35,469	4.4	11%	$17,611	$27,800	$33,539	4.7	16%	$20,919	$41,171	$46,815	5.2
CSA_Philadelphia-Camden-Vineland, PA-NJ-DE-	5,833,585	14	14	16.4	14.9	1.5	14.3	13.3%	0.90	14%	$42,815	$42,000	$83,904	7.4	31%	$20,897	$40,182	$46,803	5.3	28%	$23,807	$51,473	$59,295	5.3
CSA_Pittsburgh-New Castle, PA	2,525,730	13	3	21.2	18.4	2.8	16.9	17.6%	0.79	2%	$19,491	$35,264	$42,857	4.9	16%	$16,873	$30,404	$38,243	4.8	10%	$20,635	$35,540	$43,510	5.2
CSA_St. Louis-St. Charles-Farmington, MO-IL	2,777,132	12	3	17.5	16.3	1.2	15.2	11.7%	0.79	6%	$17,556	$33,045	$37,313	4.8	37%	$24,136	$39,416	$47,776	5.0	21%	$22,267	$40,513	$47,145	5.3
CSA_Toledo-Fremont, OH	720,980	3	3	15.1	14.9	0.2	14.7	5.3%	0.94	94%	$6,662	$10,171	$10,104	2.9	33%	$14,752	$25,944	$32,969	5.0	16%	$20,529	$41,666	$49,237	5.3
CSA_York-Hanover-Gettysburg, PA	473,043	2	2	17.3	15.4	1.9	13.5	16.1%	0.83	3%	$21,145	$39,962	$47,045	5.3	7%	$18,471	$43,979	$47,042	5.0	7%	$20,603	$43,604	$49,414	5.1
CSA_Youngstown-Warren-East Liverpool, OH-PA	715,039	3	3	15.2	14.8	0.4	14.3	5.1%	0.93	45%	$9,869	$18,150	$30,556	5.3	28%	$16,142	$28,939	$37,757	4.6	12%	$18,399	$34,124	$40,480	5.1

- Areas that could use SA to meet NAAQS are underlined.
- Socioeconomic data from 2000 Census.
- Education Level defined as follows
  - Focused on ‘education level attained’ (left/lower column)
  - Created ‘education average’ variable as follows (right/lower formula):
    - (Weighted populations of each category)

```
55. P037001     : pop_mf - Total: Population 25+
  56. P037002     : Male 25+:
  57. P037003     : pop_m1 - Male No schooling completed
  58. P037004     : pop_m2 - Male Nursery-4th grade
  59. P037005     : pop_m3 - Male 5th and 6th grade
  60. P037006     : pop_m4 - Male 7th and 8th grade
  61. P037007     : pop_m5 - Male 9th grade
  62. P037008     : pop_m6 - Male 10th grade
  63. P037009     : pop_m7 - Male 11th grade
  64. P037010     : pop_m8 - Male 12th grade, no diploma
  65. P037011     : pop_m9 - Male High school grad (inc equivalency)
  66. P037012     : pop_m10 - Male Some college, under 1 year
  67. P037013     : pop_m11 - Male Some college, 1+ years, no degree
  68. P037014     : pop_m12 - Male Associate degree
  69. P037015     : pop_m13 - Male Bachelor's degree
  70. P037016     : pop_m14 - Male Master's degree
  71. P037017     : pop_m15 - Male Professional school degree
  72. P037018     : pop_m16 -Male Doctorate degree
  73. P037019     : Female 25+:
  74. P037020     : pop_f1 - Female No schooling completed
  75. P037021     : pop_f2 - Female Nursery-4th grade
  76. P037022     : pop_f3 - Female 5th and 6th grade
  77. P037023     : pop_f4 - Female 7th and 8th grade
  78. P037024     : pop_f5 - Female 9th grade
  79. P037025     : pop_f6 - Female 10th grade
  80. P037026     : pop_f7 - Female 11th grade
  81. P037027     : pop_f8 - Female 12th grade, no diploma
  82. P037028     : pop_f9 - Female High school grad (inc equivalency)
  83. P037029     : pop_f10 - Female Some college, under 1 year
  84. P037030     : pop_f11 - Female Some college, 1+ years, no degree
  85. P037031     : pop_f12 - Female Associate degree
  86. P037032     : pop_f13 - Female Bachelor's degree
  87. P037033     : pop_f14 - Female Master's degree
  88. P037034     : pop_f15 - Female Professional school degree
  89. P037035     : pop_f16 - Female Doctorate degree
```

```
avg_ed=
((pop_m1+pop_f1*1)+(pop_m2+pop_f2*2)+
(pop_m3+pop_f3*3)+(pop_m4+pop_f4*4)+
(pop_m5+pop_f5*5)+(pop_m6+pop_f6*6)+
(pop_m7+pop_f7*7)+(pop_m8+pop_f8*8)+
(pop_m9+pop_f9*9)+(pop_m10+pop_f10*10)+
(pop_m11+pop_f11*11)+(pop_m12+pop_f12*12)+
(pop_m13+pop_f13*13)+(pop_m14+pop_f14*14)+
(pop_m15+pop_f15*15)+(pop_m16+pop_f16*16))
/pop_mf;
```

Listing 2: Areas that Qualify for Spatial Averaging; Current Criteria (.6 corr., 20% diff in means), NAAQS Level of 14

										High Site Census Tract Information					Other Site Census Tract(s) Information (avg.)					Area (CSA/CBSA) Information				
Area	Pop.	Number of Sites in Area	Number of Sites in CMZ	Design value without SA	Design value with SA	Difference in DV's	Minimum area site DV	Maximum between-site difference in means	Minimum between-site correlation (annual)	Percent minority	Per capita income	Median Household Income	Median Family income	Average Education Level Attained*	Percent minority	Per capita income	Median Household Income	Median Family income	Average Education Level Attained*	Percent minority	Per capita income	Median Household Income	Median Family Income	Average Education Level Attained*
CBSA_Allentown-Bethlehem-Easton, PA-NJ	740,395	3	3	14.8	14.4	0.4	14.6	4.8%	0.91	12%	$17,983	$44,297	$48,333	4.8	7%	$21,244	$34,187	$48,800	4.6	10%	$21,867	$44,922	$52,674	5.2
CBSA_Augusta-Richmond County, GA-SC	499,684	2	2	14.7	13.2	1.5	12.4	9.6%	0.84	40%	$14,902	$29,783	$32,813	4.4	53%	$17,757	$36,991	$40,950	4.9	39%	$18,496	$37,529	$43,751	5.1
CBSA_Bakersfield, CA	661,645	5	3	21.8	21.0	0.8	20.3	7.9%	0.98	46%	$11,843	$18,777	$22,669	4.5	32%	$15,947	$33,390	$37,965	5.2	38%	$15,780	$38,858	$42,458	5.0
CBSA_Canton-Massillon, OH	406,934	2	2	17.3	16.6	0.7	15.8	5.6%	0.99	9%	$12,577	$24,205	$30,833	4.4	37%	$14,201	$10,457	$25,000	4.2	9%	$20,154	$36,917	$43,005	5.1
CBSA_Charleston, WV	309,635	2	2	17.1	16.3	0.8	15.5	5.3%	0.97	9%	$16,667	$20,929	$32,167	4.7	20%	$28,021	$27,217	$50,690	5.7	7%	$19,090	$29,508	$35,875	5.0
CBSA_Evansville, IN-KY	342,815	3	3	15.5	15.3	0.2	15.2	3.7%	0.96	11%	$12,773	$29,033	$36,446	4.9	8%	$23,162	$31,037	$46,836	5.1	8%	$20,026	$38,956	$46,128	5.1
CBSA_Hagerstown-Martinsburg, MD-WV	222,771	2	2	16.3	15.1	1.2	14.0	11.0%	0.80	17%	$14,688	$25,423	$35,591	4.3	73%	$21,284	$0	$0	8.9	9%	$19,222	$36,997	$42,510	5.1
CBSA_Huntington-Ashland, WV-KY-OH	288,649	3	3	16.6	15.8	0.8	15.0	8.3%	0.87	12%	$4,312	$6,624	$5,357	7.4	3%	$19,748	$32,969	$38,206	4.9	4%	$16,631	$29,341	$36,169	4.9
CBSA_Roanoke, VA	288,309	2	2	14.7	14.4	0.3	14.2	2.4%	0.96	14%	$15,721	$29,774	$37,699	4.7	6%	$22,330	$41,331	$50,891	5.5	15%	$21,006	$38,681	$45,437	5.1
CBSA_San Diego-Carlsbad-San Marcos, CA	2,813,833	5	5	15.9	15.0	0.9	14.6	16.4%	0.66	46%	$10,278	$21,021	$23,870	5.1	29%	$16,989	$40,702	$46,701	5.1	33%	$22,928	$51,773	$57,106	5.6
CBSA_South Bend-Mishawaka, IN-MI	316,663	3	3	14.3	14.1	0.2	14.0	2.6%	0.99	64%	$12,615	$25,466	$27,993	4.5	36%	$14,681	$32,046	$35,594	4.8	17%	$19,728	$39,967	$45,577	5.3
CBSA_Terre Haute, IN	170,943	2	2	14.6	14.0	0.6	13.4	6.2%	0.96	7%	$16,572	$32,321	$39,474	5.2	5%	$19,748	$38,281	$45,710	5.1	7%	$17,342	$35,029	$41,115	5.2
CBSA_Weirton-Steubenville, WV-OH	132,008	4	4	17.8	17.0	0.8	16.2	7.8%	0.86	5%	$15,980	$30,000	$40,181	4.8	6%	$17,242	$33,295	$40,576	4.6	5%	$16,909	$32,335	$39,252	4.9
CBSA_Wheeling, WV-OH	153,172	2	2	15.7	15.4	0.3	15.2	1.9%	0.95	1%	$17,077	$31,836	$39,033	4.8	14%	$8,072	$7,663	$23,214	3.7	4%	$16,749	$29,113	$36,899	5.0
CSA_Birmingham-Hoover-Cullman, AL	1,129,721	8	6	18.0	15.4	2.6	13.8	18.1%	0.77	99%	$12,938	$16,995	$23,333	4.3	16%	$21,918	$45,552	$51,117	5.2	28%	$20,390	$36,593	$43,526	5.1
CSA_Charlotte-Gastonia-Salisbury, NC-SC	1,897,034	5	5	14.9	14.4	0.5	14.0	6.1%	0.92	92%	$12,094	$26,829	$28,413	4.5	27%	$20,137	$37,554	$44,614	5.0	26%	$22,291	$39,740	$45,842	5.2
CSA_Chattanooga-Cleveland-Athens, TN-GA	629,561	3	3	15.6	15.4	0.2	14.6	3.6%	0.87	6%	$14,092	$23,713	$29,183	4.1	24%	$13,257	$22,338	$35,768	4.2	14%	$19,278	$33,613	$39,509	5.0
CSA_Chicago-Naperville-Michigan City, IL-IN-WI	9,312,255	28	2	17.7	17.5	0.2	17.3	4.6%	0.84	10%	$12,368	$31,156	$30,189	4.7	29%	$20,950	$45,553	$53,509	5.1	33%	$24,491	$52,263	$59,135	5.4
CSA_Cincinnati-Middletown-Wilmington, OH-KY-I	2,050,175	12	12	17.8	16.0	1.8	14.5	13.9%	0.90	13%	$19,121	$27,364	$36,667	4.4	22%	$17,950	$31,444	$38,807	4.9	14%	$22,786	$43,248	$49,355	5.3
CSA_Cleveland-Akron-Elyria, OH	2,945,831	13	11	18.3	15.9	2.4	14.2	19.2%	0.84	31%	$15,270	$25,221	$26,850	5.6	41%	$15,278	$28,755	$32,732	4.7	21%	$22,321	$46,452	$53,471	5.3
CSA_Columbus-Auburn-Opelika, GA-AL	420,965	3	3	15.3	14.6	0.7	14.3	8.2%	0.78	65%	$7,295	$10,121	$11,949	3.3	78%	$11,574	$18,636	$23,013	4.2	42%	$17,184	$31,978	$37,256	5.1
CSA_Columbus-Marion-Chillicothe, OH	1,835,189	3	3	16.7	16.2	0.5	15.9	5.9%	0.95	88%	$14,293	$21,486	$27,560	3.9	39%	$15,184	$28,309	$30,408	4.7	17%	$22,256	$45,186	$51,028	5.5
CSA_Dayton-Springfield-Greenville, OH	1,085,094	3	3	15.2	14.7	0.5	14.7	8.5%	0.93	6%	$17,457	$32,708	$40,117	5.3	7%	$16,186	$26,815	$34,558	4.9	16%	$21,263	$42,919	$49,338	5.3
CSA_Detroit-Warren-Flint, MI	5,357,538	14	6	19.5	16.8	2.7	15.1	18.5%	0.83	29%	$7,573	$19,713	$24,031	3.9	43%	$17,486	$35,422	$40,804	5.1	27%	$24,353	$53,256	$60,632	5.4
CSA_Fairmont-Clarksburg, WV	148,742	2	2	15.4	14.7	0.7	14.0	5.6%	0.96	3%	$13,328	$21,839	$28,906	4.6	8%	$14,417	$16,590	$30,031	4.3	4%	$16,094	$28,602	$34,255	4.9
CSA_Fort Wayne-Huntington-Auburn, IN	548,416	2	2	14.3	14.3	0.0	14.3	1.1%	0.99	14%	$15,132	$24,423	$36,659	4.3	13%	$19,343	$39,929	$44,730	5.4	11%	$20,468	$43,571	$49,877	5.3
CSA_Fresno-Madera, CA	922,516	2	2	19.7	19.5	0.2	19.2	3.1%	0.97	45%	$12,781	$31,131	$34,440	4.6	57%	$10,976	$16,842	$20,804	4.0	45%	$15,388	$36,870	$39,680	4.7
CSA_Greensboro--Winston-Salem--High Point, N	1,283,856	4	4	15.8	14.6	1.2	14.0	8.6%	0.93	50%	$19,691	$28,094	$34,320	4.6	42%	$25,501	$35,913	$47,006	5.1	25%	$21,090	$38,066	$45,213	5.0
CSA_Greenville-Anderson-Seneca, SC	791,895	2	2	14.5	12.5	2.0	10.6	18.1%	0.88	11%	$20,873	$47,161	$54,688	5.4	2%	$16,573	$30,429	$36,127	4.9	19%	$19,843	$36,301	$43,552	5.1
CSA_Harrisburg-Carlisle-Lebanon, PA	629,401	2	2	15.8	14.9	0.9	15.8	13.5%	0.92	35%	$15,752	$31,557	$37,679	4.9	1%	$18,897	$44,341	$50,259	5.1	12%	$21,939	$42,855	$50,094	5.3
CSA_Houston-Baytown-Huntsville, TX	4,815,122	6	6	14.2	12.1	2.1	10.9	19.4%	0.61	97%	$10,236	$24,353	$24,457	4.0	45%	$15,390	$38,444	$42,128	4.7	37%	$21,519	$41,701	$47,600	5.3
CSA_Huntsville-Decatur, AL	488,243	2	2	14.1	13.9	0.2	13.7	2.7%	0.88	54%	$13,252	$17,589	$23,000	5.2	12%	$19,520	$48,507	$54,079	4.2	23%	$21,033	$38,629	$45,429	5.4
CSA_Indianapolis-Anderson-Columbus, IN	1,843,588	6	6	16.7	15.2	1.5	13.6	12.0%	0.92	40%	$9,869	$18,988	$20,417	4.4	17%	$18,785	$36,313	$41,702	5.3	16%	$22,715	$46,925	$53,537	5.4
CSA_Johnson City-Kingsport-Bristol, TN-VA	480,091	2	2	14.7	14.5	0.2	14.3	2.4%	0.96	4%	$18,538	$25,522	$31,715	4.3	12%	$15,781	$24,412	$27,723	4.0	4%	$17,800	$31,032	$37,582	4.9
CSA_Knoxville-Sevierville-La Follette, TN	779,013	5	5	16.7	15.6	1.1	14.2	11.4%	0.85	35%	$7,364	$11,305	$13,239	3.4	11%	$17,905	$35,858	$42,976	5.0	8%	$20,034	$33,904	$40,386	5.1
CSA_Lexington-Fayette--Frankfort--Richmond, K'	602,773	4	4	15.7	14.3	1.4	13.5	9.3%	0.75	28%	$10,418	$17,111	$18,679	5.6	30%	$17,721	$28,083	$36,300	4.8	12%	$20,520	$37,223	$43,417	5.3
CSA_Little Rock-North Little Rock-Pine Bluff, AR	785,024	5	5	14.1	13.0	1.1	11.9	13.0%	0.78	89%	$8,205	$18,099	$21,758	4.0	16%	$15,474	$33,680	$40,409	5.3	26%	$19,069	$35,771	$41,537	5.2
CSA_Louisville-Elizabethtown-Scottsburg, KY-IN	1,292,482	6	6	16.9	15.6	1.3	14.1	12.4%	0.85	11%	$13,959	$25,315	$35,469	4.4	11%	$17,611	$27,800	$33,539	4.7	16%	$20,919	$41,171	$46,815	5.2
CSA_Nashville-Davidson--Murfreesboro--Columb	1,381,287	3	3	14.4	13.4	1.0	13.5	7.4%	0.88	23%	$20,803	$40,781	$49,598	5.3	11%	$21,017	$41,519	$50,386	5.1	20%	$22,287	$42,067	$48,075	5.3
CSA_Philadelphia-Camden-Vineland, PA-NJ-DE-	5,833,585	14	14	16.4	14.9	1.5	14.3	13.3%	0.90	14%	$42,815	$42,000	$83,904	7.4	31%	$20,897	$40,182	$46,803	5.3	28%	$23,807	$51,473	$59,295	5.3
CSA_Pittsburgh-New Castle, PA	2,525,730	13	3	21.2	18.4	2.8	16.9	17.6%	0.79	2%	$19,491	$35,264	$42,857	4.9	16%	$16,873	$30,404	$38,243	4.8	10%	$20,635	$35,540	$43,510	5.2
CSA_St. Louis-St. Charles-Farmington, MO-IL	2,777,132	12	9	17.5	15.3	2.2	14.5	19.2%	0.76	6%	$17,556	$33,045	$37,313	4.8	37%	$24,136	$39,416	$47,776	5.0	21%	$22,267	$40,513	$47,145	5.3
CSA_Toledo-Fremont, OH	720,980	3	3	15.1	14.9	0.2	14.7	5.3%	0.94	94%	$6,662	$10,171	$10,104	2.9	33%	$14,752	$25,944	$32,969	5.0	16%	$20,529	$41,666	$49,237	5.3
CSA_York-Hanover-Gettysburg, PA	473,043	2	2	17.3	15.4	1.9	13.5	16.1%	0.83	3%	$21,145	$39,962	$47,045	5.3	7%	$18,471	$43,979	$47,042	5.0	7%	$20,603	$43,604	$49,414	5.1
CSA_Youngstown-Warren-East Liverpool, OH-PA	715,039	3	3	15.2	14.8	0.4	14.3	5.1%	0.93	45%	$9,869	$18,150	$30,556	5.3	28%	$16,142	$28,939	$37,757	4.6	12%	$18,399	$34,124	$40,480	5.1

- Areas that could use SA to meet NAAQS are underlined.

Issues w/ Spatial Averaging

- Are the would-be violating ('high') sites in an area that could use SA located in lower-income, high percentage-minority, and/or lower education locations than the overall area?

Comparison of High-Site Census Tract Socioeconomic Data to Area Average

NAAQS Level of 15

	Areas that could use spatial averaging			Areas that could attain the standard using spatial averaging (subset of left columns)		
Variable	Total	Number where indicated metric is higher for the metro area than in the high-site census tract	Number where indicated metric is lower for the metro area than in the high site census tract	Total	Number where indicated metric is higher for the metro area than in the high-site census tract	Number where indicated metric is lower for the metro area than in the high site census tract
Percentage Minority	32	13	19	10	3	7
Per Capita Income	32	29	3	10	9	1
Median Family Income	32	31	1	10	10	0
Median Household Income	32	29	3	10	9	1
Education Level Attained	32	25	7	10	6	4

NAAQS Level of 14

	Areas that could use spatial averaging			averaging (subset of left columns)		
Variable	Total	Number where indicated metric is higher for the metro area than in the high-site census tract	Number where indicated metric is lower for the metro area than in the high site census tract	Total	Number where indicated metric is higher for the metro area than in the high-site census tract	Number where indicated metric is lower for the metro area than in the high site census tract
Percentage Minority	45	15	30	7	1	6
Per Capita Income	45	40	5	7	6	1
Median Family Income	45	43	2	7	6	1
Median Household Income	45	40	5	7	5	2
Education Level Attained	45	36	9	7	5	2

In most areas that could use SA (15 or 14 NAAQS level), the high site is located in an area populated by lower income, higher percentage minority, and less-educated people when compared to the overall metro area.

Issues w/ Spatial Averaging

- Is there a relationship between the magnitude of the DV disparity and the disparity in the socioeconomic variables?
- See computations below for NAAQS level of 14.

Area	Difference in DV's	Difference Between High-Site Census Tract and Other-Site Census Tract(s)					Difference Between High-Site Census Tract and Area (CSA/CBSA) Average				
		Percent minority	Per capita income	Median Household Income	Median Family Income	Average Education Level Attained*	Percent minority	Per capita income	Median Household Income	Median Family Income	Average Education Level Attained*
CBSA_Allentown-Bethlehem-Easton, PA-NJ	0.4	6%	-$3,261	$10,110	-$467	-0.4	3%	-$3,884	-$625	-$4,341	-0.4
CBSA_Augusta-Richmond County, GA-SC	1.5	-13%	-$2,855	-$7,208	-$8,137	-0.7	1%	-$3,594	-$7,746	-$10,938	-0.7
CBSA_Bakersfield, CA	0.8	14%	-$4,104	-$14,613	-$15,296	-0.5	8%	-$3,937	-$20,081	-$19,789	-0.5
CBSA_Canton-Massillon, OH	0.7	-28%	-$1,624	$13,748	$5,833	-0.7	0%	-$7,577	-$12,712	-$12,172	-0.7
CBSA_Charleston, WV	0.8	-11%	-$11,354	-$6,288	-$18,523	-0.3	2%	-$2,423	-$8,579	-$3,708	-0.3
CBSA_Evansville, IN-KY	0.2	3%	-$10,389	-$2,004	-$10,390	-0.2	3%	-$7,253	-$9,923	-$9,682	-0.2
CBSA_Hagerstown-Martinsburg, MD-WV	1.2	-56%	-$6,596			-0.8	9%	-$4,534	-$11,574	-$6,919	-0.8
CBSA_Huntington-Ashland, WV-KY-OH	0.8	9%	-$15,436	-$26,345	-$32,849	2.5	8%	-$12,319	-$22,717	-$30,812	2.5
CBSA_Roanoke, VA	0.3	9%	-$6,609	-$11,557	-$13,192	-0.4	-1%	-$5,285	-$8,907	-$7,738	-0.4
CBSA_San Diego-Carlsbad-San Marcos, CA	0.9	18%	-$6,711	-$19,681	-$22,832	-0.5	13%	-$12,650	-$30,752	-$33,237	-0.5
CBSA_South Bend-Mishawaka, IN-MI	0.2	28%	-$2,066	-$6,580	-$7,601	-0.8	48%	-$7,113	-$14,501	-$17,584	-0.8
CBSA_Terre Haute, IN	0.6	2%	-$3,176	-$5,960	-$6,236	0.0	0%	-$770	-$2,708	-$1,641	0.0
CBSA_Weirton-Steubenville, WV-OH	0.8	-1%	-$1,262	-$3,295	-$395	-0.2	-1%	-$929	-$2,335	$929	-0.2
CBSA_Wheeling, WV-OH	0.3	-13%	$9,005	$24,173	$15,819	-0.2	-3%	$328	$2,723	$2,134	-0.2
CSA_Birmingham-Hoover-Cullman, AL	2.6	83%	-$8,980	-$28,557	-$27,784	-0.8	70%	-$7,452	-$19,598	-$20,193	-0.8
CSA_Charlotte-Gastonia-Salisbury, NC-SC	0.5	65%	-$8,043	-$10,725	-$16,201	-0.7	66%	-$10,197	-$12,911	-$17,429	-0.7
CSA_Chattanooga-Cleveland-Athens, TN-GA	0.2	-19%	$835	$1,376	-$6,585	-0.8	-9%	-$5,186	-$9,900	-$10,326	-0.8
CSA_Chicago-Naperville-Michigan City, IL-IN-WI	0.2	-19%	-$8,582	-$14,397	-$23,320	-0.7	-23%	-$12,123	-$21,107	-$28,946	-0.7
CSA_Cincinnati-Middletown-Wilmington, OH-KY-I	1.8	-9%	$1,171	-$4,080	-$2,140	-0.9	-1%	-$3,665	-$15,884	-$12,688	-0.9
CSA_Cleveland-Akron-Elyria, OH	2.4	-10%	-$8	-$3,534	-$5,882	0.3	10%	-$7,051	-$21,231	-$26,621	0.3
CSA_Columbus-Auburn-Opelika, GA-AL	0.7	-13%	-$4,279	-$8,515	-$11,064	-1.8	23%	-$9,889	-$21,857	-$25,307	-1.8
CSA_Columbus-Marion-Chillicothe, OH	0.5	49%	-$891	-$6,823	-$2,848	-1.6	72%	-$7,963	-$23,700	-$23,468	-1.6
CSA_Dayton-Springfield-Greenville, OH	0.5	-1%	$1,271	$5,894	$5,560	0.1	-10%	-$3,806	-$10,211	-$9,221	0.1
CSA_Detroit-Warren-Flint, MI	2.7	-14%	-$9,913	-$15,709	-$16,773	-1.5	1%	-$16,780	-$33,543	-$36,601	-1.5
CSA_Fairmont-Clarksburg, WV	0.7	-5%	-$1,089	$5,249	-$1,125	-0.3	-1%	-$2,766	-$6,763	-$5,349	-0.3
CSA_Fort Wayne-Huntington-Auburn, IN	0.0	1%	-$4,211	-$15,506	-$8,071	-1.0	3%	-$5,336	-$19,148	-$13,218	-1.0
CSA_Fresno-Madera, CA	0.2	-12%	$1,805	$14,289	$13,636	-0.1	1%	-$2,607	-$5,739	-$5,240	-0.1
CSA_Greensboro--Winston-Salem--High Point, N	1.2	7%	-$5,810	-$7,819	-$12,686	-0.4	25%	-$1,399	-$9,972	-$10,893	-0.4
CSA_Greenville-Anderson-Seneca, SC	2.0	10%	$4,300	$16,732	$18,561	0.3	-8%	$1,030	$10,860	$11,136	0.3
CSA_Harrisburg-Carlisle-Lebanon, PA	0.9	34%	-$3,145	-$12,784	-$12,580	-0.3	23%	-$6,187	-$11,298	-$12,415	-0.3
CSA_Houston-Baytown-Huntsville, TX	2.1	52%	-$5,154	-$14,091	-$17,671	-1.3	60%	-$11,283	-$17,348	-$23,143	-1.3
CSA_Huntsville-Decatur, AL	0.2	42%	-$6,268	-$30,918	-$31,079	-0.2	31%	-$7,781	-$21,040	-$22,429	-0.2
CSA_Indianapolis-Anderson-Columbus, IN	1.5	23%	-$8,916	-$17,325	-$21,285	-1.0	24%	-$12,846	-$27,937	-$33,120	-1.0
CSA_Johnson City-Kingsport-Bristol, TN-VA	0.2	-8%	$2,757	$1,110	$3,992	-0.5	0%	$738	-$5,510	-$5,867	-0.5
CSA_Knoxville-Sevierville-La Follette, TN	1.1	24%	-$10,541	-$24,553	-$29,737	-1.8	27%	-$12,670	-$22,599	-$27,147	-1.8
CSA_Lexington-Fayette--Frankfort--Richmond, K'	1.4	-2%	-$7,303	-$10,972	-$17,621	0.3	16%	-$10,102	-$20,112	-$24,738	0.3
CSA_Little Rock-North Little Rock-Pine Bluff, AR	1.1	74%	-$7,269	-$15,581	-$18,651	-1.2	63%	-$10,864	-$17,672	-$19,779	-1.2
CSA_Louisville-Elizabethtown-Scottsburg, KY-IN	1.3	0%	-$3,652	-$2,485	$1,930	-0.8	-5%	-$6,960	-$15,856	-$11,346	-0.8
CSA_Nashville-Davidson--Murfreesboro--Columb	1.0	12%	-$214	-$738	-$788	0.0	3%	-$1,484	-$1,286	$1,523	0.0
CSA_Philadelphia-Camden-Vineland, PA-NJ-DE-	1.5	-17%	$21,918	$1,818	$37,101	2.1	-14%	$19,008	-$9,473	$24,609	2.1
CSA_Pittsburgh-New Castle, PA	2.8	-14%	$2,618	$4,860	$4,614	-0.3	-8%	-$1,144	-$276	-$653	-0.3
CSA_St. Louis-St. Charles-Farmington, MO-IL	2.2	-32%	-$6,580	-$6,371	-$10,463	-0.5	-15%	-$4,711	-$7,468	-$9,832	-0.5
CSA_Toledo-Fremont, OH	0.2	61%	-$8,090	-$15,773	-$22,865	-2.3	78%	-$13,867	-$31,495	-$39,133	-2.3
CSA_York-Hanover-Gettysburg, PA	1.9	-4%	$2,674	-$4,017	$3	0.2	-4%	$542	-$3,642	-$2,369	0.2
CSA_Youngstown-Warren-East Liverpool, OH-PA	0.4	18%	-$6,273	-$10,789	-$7,201	0.2	34%	-$8,530	-$15,974	-$9,924	0.2
Correlation between DV difference column and socioeconomic variable difference columns		**-0.045839**	**0.061794**	**-0.0862136**	**0.022738**	**0.0460458**	**-0.037935**	**0.031241**	**-0.0253302**	**-0.011631**	**0.0460458**

- **There does not appear to be a relationship between magnitude of DV disparity and the disparity in the socioeconomic variables.**
- **There are obviously many other factors that determine differences in the socioeconomic variables <u>across areas.</u>**

Issues w/ Spatial Averaging

- Within an area, is there a relationship between DV level and the socioeconomic variable level?
- Assume other factors cause differences across areas. Look for relationships within areas. Look in all areas with multiple sites, not just areas where SA is applicable.

Correlation of Within-Area Monitoring Site Tract Data - DV versus Percent Minority

Areas with multiple Sites			Areas with 2 Sites			Areas with 3+ Sites				
Number	Number w/ positive correlation	Percent areas w/ negative correlation	Number	Number w/ positive correlation	Percent areas w/ positive correlation	Number	Number w/ positive correlation	Percent areas w/ positive correlation	Mean Correlation (where positive)	Median Correlation (where positive)
125	84	67%	50	24	48%	75	60	80%	0.6175	0.659

Correlation of Within-Area Monitoring Site Tract Data - DV versus Per Capita Income

Areas with multiple Sites			Areas with 2 Sites			Areas with 3+ Sites				
Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Mean Correlation (where negative)	Median Correlation (where negative)
125	93	74%	50	35	70%	75	58	77%	-0.5906	-0.5967

Correlation of Within-Area Monitoring Site Tract Data - DV versus Median Household Income

Areas with multiple Sites			Areas with 2 Sites			Areas with 3+ Sites				
Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Mean Correlation (where negative)	Median Correlation (where negative)
125	92	74%	50	32	64%	75	60	80%	-0.5791	-0.5615

Correlation of Within-Area Monitoring Site Tract Data - DV versus Median Family Income

Areas with multiple Sites			Areas with 2 Sites			Areas with 3+ Sites				
Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Mean Correlation (where negative)	Median Correlation (where negative)
125	96	77%	50	33	66%	75	63	84%	-0.599	-0.6246

Correlation of Within-Area Monitoring Site Tract Data - DV versus Average Education

Areas with multiple Sites			Areas with 2 Sites			Areas with 3+ Sites				
Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Number	Number w/ negative correlation	Percent areas w/ negative correlation	Mean Correlation (where negative)	Median Correlation (where negative)
125	71	57%	50	27	54%	75	44	59%	-0.5662	-0.5886

- **In most areas, there appears to be a negative relationship between DV and 1) education level attained, 2) per capita income, 3) median household income, and 4)median family income**
- **In most areas, there is a positive relationship between DV and percentage minority.**

Is Adjustment of SA Criteria Appropriate?

- The 2 considered SA criteria --- .6 minimum correlation and 20% +/- maximum difference in annual means --- were initially suggested in 1997 with limited knowledge of actual conditions (lack of data). Now that we have several years of monitoring data available, should we consider adjustments to these criteria?
- 3 simple evaluations were conducted:

1. Benchmark typical within-area correlation (of daily PM2.5 concentrations). [If SA requires a minimum of .6 correlation, but .6 is only average or worse, shouldn’t areas/sites need to show better (higher R) to be permitted to use SA?]
2. Compared annual correlations to seasonal correlations. [If there is significant differences between annual and seasonal correlations, shouldn’t the minimum criterion be applied on a seasonal basis?]
3. Benchmarked average percent difference in annual site means to annual spatial means. [If SA requires a maximum of 20% difference in annual means (site vs. spatial) but 20% is only average or worse, shouldn’t areas/sites need to show better (lower % difference) to be permitted to use SA?]

Is Adjustment of SA Criteria Appropriate?

1. Benchmark typical within-area correlation (of daily PM2.5 concentrations)
   - Procedure:
     - Utilized SP PM2.5 database (11+samples, all 12 quarters ’01-’03).
     - Calculated correlation between all site pairs in each area (CSA or CBSA)
     - Calculated univariate statistics for site correlations at national level
     - Also averaged correlation to area level then calculated univariate statistics for area averages at national level
     - Reran using only sites pairs where DV’s were within 20% tolerance

All Site Pairs

	Site Stats.	Area Average Stats.
N	2227	129
Maximum	0.9899	0.9899
95th	0.9701	0.9732
75th	0.9343	0.9473
Median	0.8993	0.8999
Mean	0.8764	0.8609
25th	0.8521	0.8228
5th	0.7353	0.6019
Minimum	-0.0854	0.3669

Site Pairs Where DV w/in 20%

	Site Stats.	Area Average Stats.
N	1914	122
Maximum	0.9899	0.9899
95th	0.9712	0.9494
75th	0.9397	0.9494
Median	0.9055	0.9044
Mean	0.8942	0.885
25th	0.8618	0.8462
5th	0.7785	0.7172
Minimum	0.3569	0.5217

- **More than 95% of all site pairs have a correlation greater then .7**
- **The median site correlation is about .9**
- **More than 95% of all areas have an average correlation greater than .6**
- **The median area average correlation is about .9**

Is Adjustment of SA Criteria Appropriate?

2. Compared annual correlations to seasonal correlations.
   - Procedure:
     - •Utilized SP PM2.5 database (11+samples, all 12 quarters ’01-’03).
     - •Calculated correlation between all site pairs in each area (CSA or CBSA)
       - •Calculated correlation for all paired data points (‘annual’)
       - •Calculated correlation for all paired data points by aggregate quarter (e.g., ‘Q1’= all pairs in 2001-Q1, 2002-Q1, and 2003- Q1) [‘Seasonal’]

•Of the 2227 site pairs:

- **•There was an average difference of about 13% between the annual correlation and the minimum seasonal correlation.**
- **•The median difference is about 6%.**
- **•More than 25% of the pairs had a difference of more than .11 R**
- **•In about 8% of the situations where the ‘annual’ R was > .6, the minimum seasonal R was < .6.**

Is Adjustment of SA Criteria Appropriate?

2. Compared annual correlations to seasonal correlations, cont.

			Correlations						
Area	Site 1	Site 2	Aggregate	Q1	Q2	Q3	Q4	Minimum Quarterl Correlation	Difference (Annual - Min_Q)
CSA_Salt Lake City-O	490353007	490030003	0.869	0.931	0.829	0.218	0.917	0.218	0.650
CSA_Seattle-Tacoma-C	530670013	530330037	0.740	0.793	0.863	0.156	0.755	0.156	0.584
CSA_Seattle-Tacoma-C	530670013	530330057	0.750	0.716	0.868	0.185	0.763	0.185	0.564
CSA_Las Vegas-Paradi	320031019	320030022	0.843	0.309	0.856	0.862	0.813	0.309	0.535
CSA_Seattle-Tacoma-C	530611007	530330037	0.807	0.778	0.905	0.315	0.833	0.315	0.492
CSA_Salt Lake City-O	490571003	490353007	0.865	0.921	0.581	0.379	0.909	0.379	0.486
CBSA_Portland-Vancou	410671003	410090004	0.762	0.743	0.769	0.288	0.838	0.288	0.474
CSA_Salt Lake City-O	490350003	490030003	0.912	0.932	0.811	0.445	0.919	0.445	0.466
CSA_Atlanta-Sandy Sp	130670003	130630091	0.800	0.346	0.842	0.869	0.914	0.346	0.454
CSA_Atlanta-Sandy Sp	132230003	130670003	0.816	0.372	0.877	0.912	0.905	0.372	0.444
CSA_Seattle-Tacoma-C	530611007	530330057	0.786	0.769	0.814	0.346	0.780	0.346	0.440
CBSA_Provo-Orem, UT	490495010	490494001	0.922	0.937	0.861	0.510	0.970	0.510	0.412
CSA_Atlanta-Sandy Sp	131210032	130670003	0.862	0.459	0.945	0.963	0.930	0.459	0.403
CSA_Seattle-Tacoma-C	530610005	530330057	0.874	0.857	0.863	0.474	0.883	0.474	0.400
CSA_Atlanta-Sandy Sp	131390003	130670003	0.774	0.381	0.843	0.892	0.782	0.381	0.393
CSA_Atlanta-Sandy Sp	131210039	130670003	0.703	0.317	0.823	0.900	0.627	0.317	0.386
CSA_Seattle-Tacoma-C	530670013	530330080	0.648	0.654	0.852	0.264	0.694	0.264	0.384
CSA_San Juan-Caguas	720610005	720530003	0.707	0.324	0.830	0.848	0.671	0.324	0.383
CBSA_Provo-Orem, UT	490495010	490490002	0.946	0.959	0.872	0.564	0.983	0.564	0.381
CBSA_Portland-Vancou	410510246	410090004	0.741	0.752	0.873	0.361	0.879	0.361	0.380
CSA_Omaha-Council B	310550052	310250002	0.739	0.870	0.938	0.918	0.359	0.359	0.380
CBSA_Tucson, AZ	040191028	040190011	0.793	0.419	0.892	0.829	0.924	0.419	0.375
CSA_Omaha-Council B	311530007	310250002	0.744	0.950	0.950	0.854	0.372	0.372	0.372
CSA_Salt Lake City-O	490570007	490353006	0.936	0.943	0.877	0.564	0.952	0.564	0.372
CSA_Salt Lake City-O	490571003	490350003	0.909	0.914	0.541	0.550	0.929	0.541	0.367
CSA_New York-Newark	340273001	090011123	0.787	0.420	0.863	0.923	0.780	0.420	0.367
CSA_Atlanta-Sandy Sp	131210039	130630091	0.764	0.398	0.705	0.881	0.856	0.398	0.365
CSA_Omaha-Council B	310550019	310250002	0.775	0.909	0.956	0.946	0.416	0.416	0.358
CSA_New York-Newark	340392003	340210008	0.876	0.521	0.903	0.967	0.904	0.521	0.355
CSA_Milwaukee-Racine	550790099	550790010	0.857	0.968	0.502	0.991	0.987	0.502	0.355
CSA_Salt Lake City-O	490571003	490353006	0.912	0.917	0.673	0.561	0.925	0.561	0.351
CSA_Omaha-Council B	310250002	191550009	0.751	0.866	0.901	0.931	0.400	0.400	0.351
CSA_Milwaukee-Racine	550790043	550790010	0.812	0.927	0.461	0.992	0.952	0.461	0.351
CSA_Little Rock-Nort	050690006	050450002	0.793	0.443	0.664	0.855	0.857	0.443	0.350
CSA_Atlanta-Sandy Sp	132230003	131210039	0.706	0.356	0.747	0.868	0.561	0.356	0.350
CSA_Washington-Balti	240030019	110010043	0.870	0.521	0.973	0.830	0.892	0.521	0.349
CBSA_Pocatello, ID	160770011	160050015	0.755	0.785	0.412	0.720	0.763	0.412	0.343
CSA_Salt Lake City-O	490571003	490570007	0.955	0.977	0.613	0.679	0.956	0.613	0.342
CBSA_Honolulu, HI	150031001	150030010	0.436	0.790	0.581	0.722	0.095	0.095	0.341
CSA_Washington-Balti	511071005	110010043	0.850	0.509	0.965	0.809	0.847	0.509	0.341
CBSA_Albuquerque, NI	350439004	350010024	0.606	0.267	0.730	0.814	0.696	0.267	0.339
CSA_Oklahoma City-SI	401091037	400819005	0.801	0.467	0.814	0.947	0.826	0.467	0.334
CSA_Salt Lake City-O	490571003	490350012	0.915	0.948	0.582	0.623	0.894	0.582	0.333
CSA_Salt Lake City-O	490570007	490030003	0.947	0.970	0.835	0.616	0.941	0.616	0.331
CSA_Washington-Balti	240251001	110010043	0.819	0.491	0.936	0.750	0.863	0.491	0.328
CSA_Salt Lake City-O	490353006	490030003	0.921	0.934	0.788	0.595	0.927	0.595	0.327
CSA_New York-Newark	340270004	340171003	0.843	0.520	0.924	0.942	0.908	0.520	0.323
CSA_Washington-Balti	245100007	110010043	0.847	0.526	0.900	0.846	0.878	0.526	0.322
CSA_Washington-Balti	510130020	110010043	0.904	0.587	0.981	0.890	0.948	0.587	0.318
CSA_San Juan-Caguas	720690001	720610005	0.670	0.353	0.873	0.900	0.717	0.353	0.317
CSA_Omaha-Council B	310550051	310250002	0.741	0.935	0.820	0.958	0.424	0.424	0.317
CSA_Seattle-Tacoma-C	530610005	530330037	0.828	0.807	0.920	0.511	0.825	0.511	0.316
CSA_New York-Newark	340270004	340230006	0.893	0.577	0.923	0.948	0.928	0.577	0.316
CSA_New York-Newark	340273001	340270004	0.931	0.616	0.961	0.978	0.927	0.616	0.315
CSA_Little Rock-Nort	051191004	050450002	0.809	0.494	0.665	0.894	0.818	0.494	0.315
CSA_Oklahoma City-SI	401090035	400819005	0.816	0.502	0.840	0.967	0.794	0.502	0.314
CSA_Washington-Balti	240313001	110010043	0.862	0.549	0.977	0.816	0.880	0.549	0.313

- **The table on the left shows examples of where there are large differences between the ‘annual’ correlations and the ‘seasonal’ correlations**
- **There are instances where the ‘annual’ correlation is more than 4 times the minimum ‘seasonal’ correlation.**
- **In most of these extreme cases, the ‘annual’ still meets the current suggested minimum of .6**

Is Adjustment of SA Criteria Appropriate?

3. Benchmarked average percent difference in annual site means versus annual spatial means)
   - Procedure:
     - Utilized SP PM2.5 database (11+samples, all 12 quarters '01-'03).
     - Calculated average difference between annual site mean and annual area spatial mean. Note that all complete sites in the area were included in the analyses even though this would often not be the case in 'real world' (since there are many situations where real low sites would not be included based on correlation, etc.) Thus, the differences shown below are biased high.

Average percent difference in annual site mean versus annual spatial mean

N	1722
Maximum	151.5%
95th	23.7%
75th	9.8%
Median	5.0%
Mean	8.1%
25th	2.0%
5th	0.4%
Minimum	0.0%

•The median (absolute) difference is 5%
•The average difference is 8%
•In less than 25% of all cases is the difference greater than 10%
•The current SA criterion of 20% is between the 90th and 95th percentile.

What would adjustment of the criteria yield?

Using criteria of .9 seasonal correlation and +/-10 % difference in annual means. Using annual std level of 15.0				Using criteria of .9 seasonal correlation and +/-10 % difference in annual means. Using annual std level of 14.0			
		Could use spatial averaging	Could use SA to meet 15.0 annual std			Could use spatial averaging	Could use SA to meet 14.0 annual std
Number of areas		12	2	Number of areas		18	1
Total population		22,327,531	1,233,836	Total population		27,499,635	1,381,287
Area distribution statistics:				Area distribution statistics:			
Difference in area DV's (ug/m3)	mean	0.52	0.50	Difference in area DV's (ug/m3)	mean	0.44	0.70
	max	1.2	0.7		max	1.2	0.7
	p95	1.2	0.7		p95	1.2	0.7
	p75	0.8	0.7		p75	0.7	0.7
	med	0.4	0.5		med	0.3	0.7
	p25	0.2	0.3		p25	0.2	0.7
	p05	0.1	0.3		p05	0.0	0.7
	min	0.1	0.3		min	0.0	0.7

•By tightening the annual mean difference criterion and the correlation criterion, much fewer areas would qualify for SA. Using a .9 quarterly correlation cutoff (as shown above) would narrow the option to 18 or fewer areas. The average difference in area means (SA versus regular) would also decline to about .4-.5 ug/m3. Total population for these areas is 22-27 million.

•Only 1 or 2 of these areas could use SA to meet the annual std NAAQS with their spatial average when they couldn't with their regular site-based DV. The realized reduction in DV for these areas would be .5-.7 ug/m3. 1 million people live in those areas.

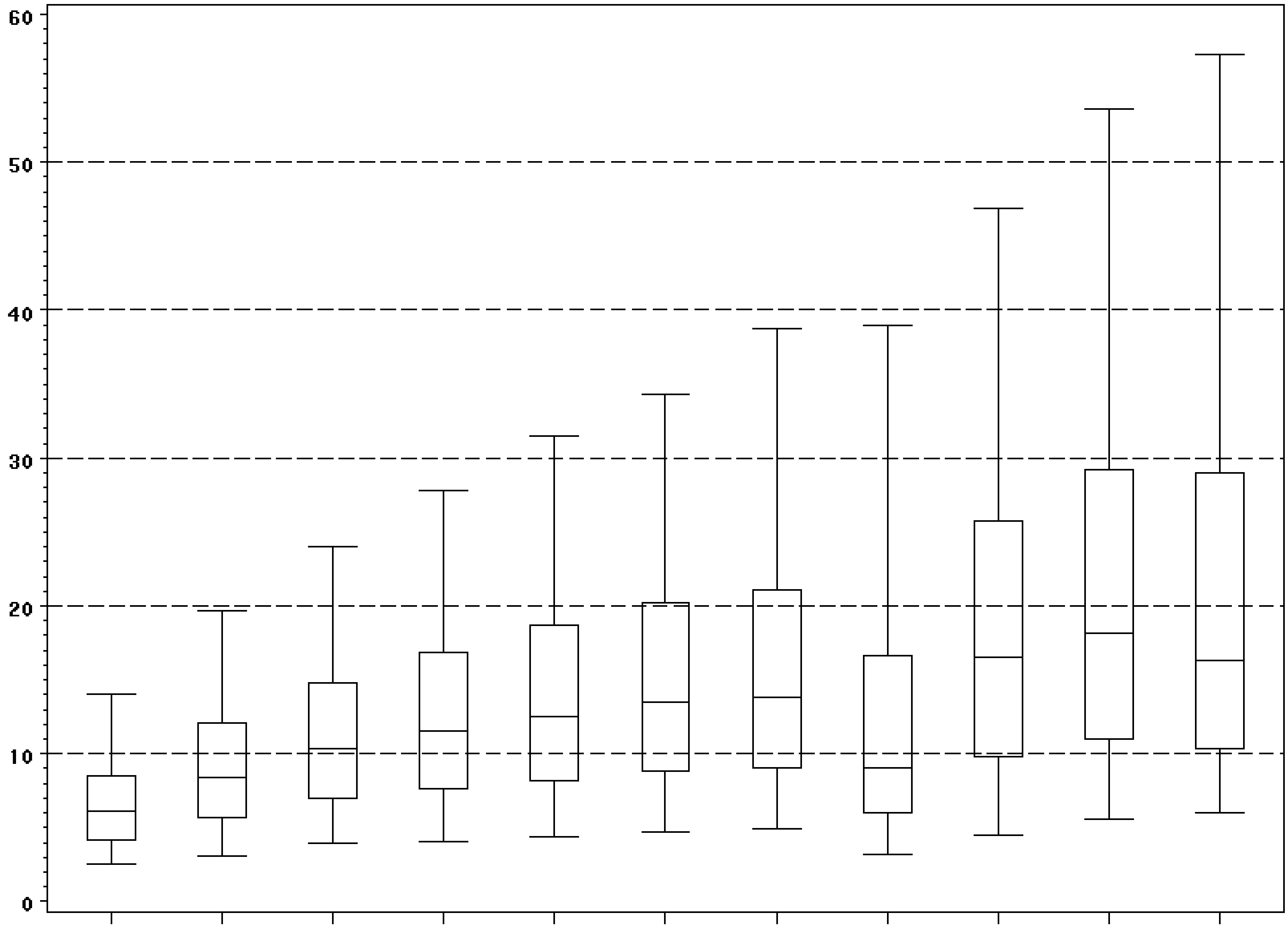
PM2.5 Evaluation of High Concentrations

- Purpose:
  - To identify the minimum number of days permitted per year to exceed the annual 98th, 99th, etc. percentiles.
  - To evaluate the (entire) daily distributions of data plotted by 98th (and 99th) percentile level intervals.
  - To evaluate the daily distributions of data exceeding site-level 98th (and 99th) DV's plotted by 98th (and 99th) percentile intervals.
  - To ascertain the actual number and percentage of days (site average, minimum, & maximum), for the 3-year period 2001-2003, where the concentration was significantly above the site 98th or 99th percentiles. [Significant defined as 5+ ug/m3.]

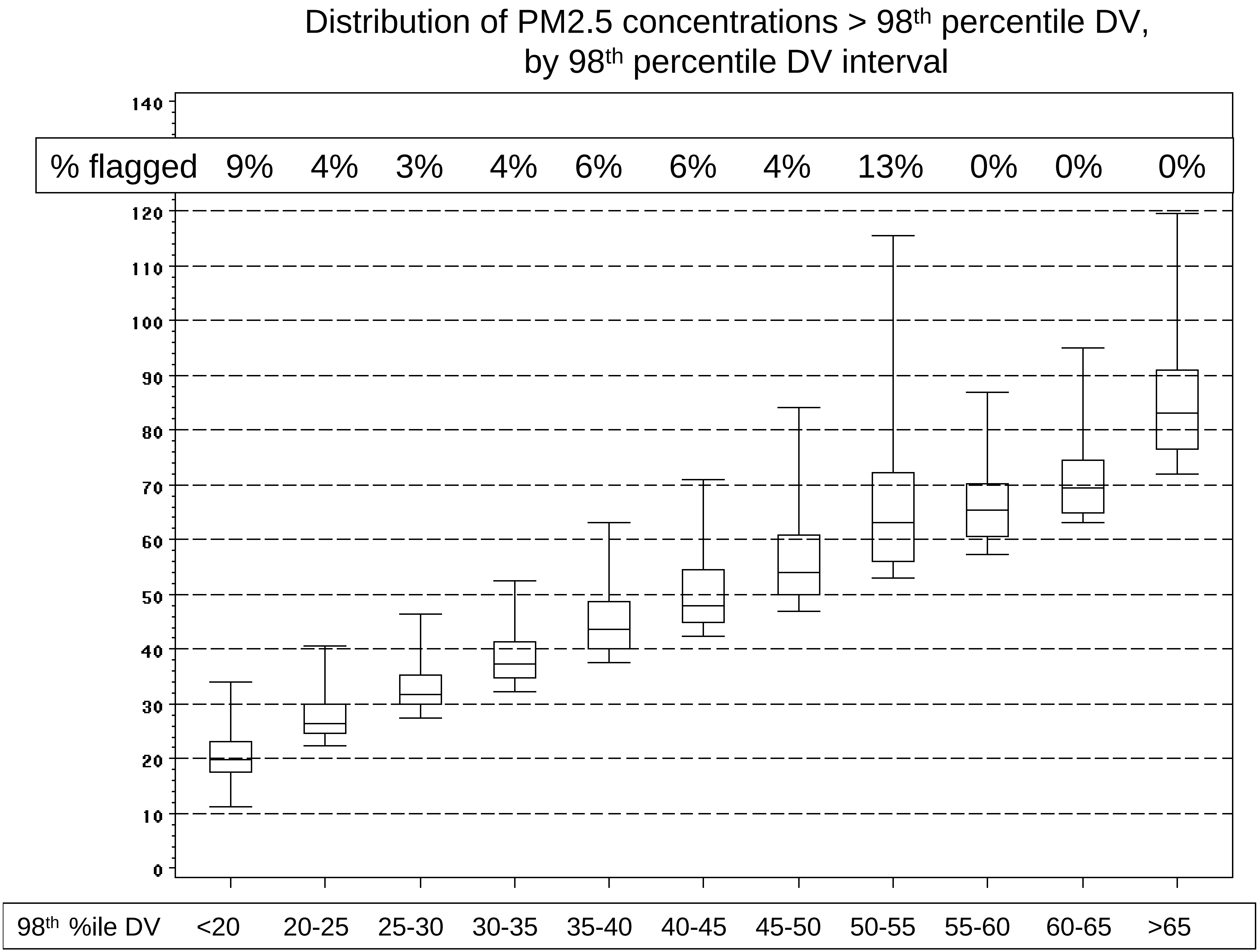
Number of Exempted Days Per Year for Percentile Metrics

Sampling Frequency	Number of Sample Days	Number of Sample Days Above Stated Percentile				
		95th	96th	97th	98th	99th
Every Day	~ 365	18	14	10	7	3
Every 3rd Day	~ 122	6	4	3	2	1
Every 6th Day	~ 61	3	2	1	1	0

Distribution of PM2.5 concentrations by 98th percentile DV interval



98th %ile DV	<20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	>65


Distribution of PM2.5 concentrations > 98th percentile DV,
by 98th percentile DV interval
% flagged 9% 4% 3% 4% 6% 6% 4% 13% 0% 0% 0%
140
120
110
100
90
80
70
60
50
40
30
20
10
0
98th %ile DV <20 20-25 25-30 30-35 35-40 40-45 45-50 50-55 55-60 60-65 >65

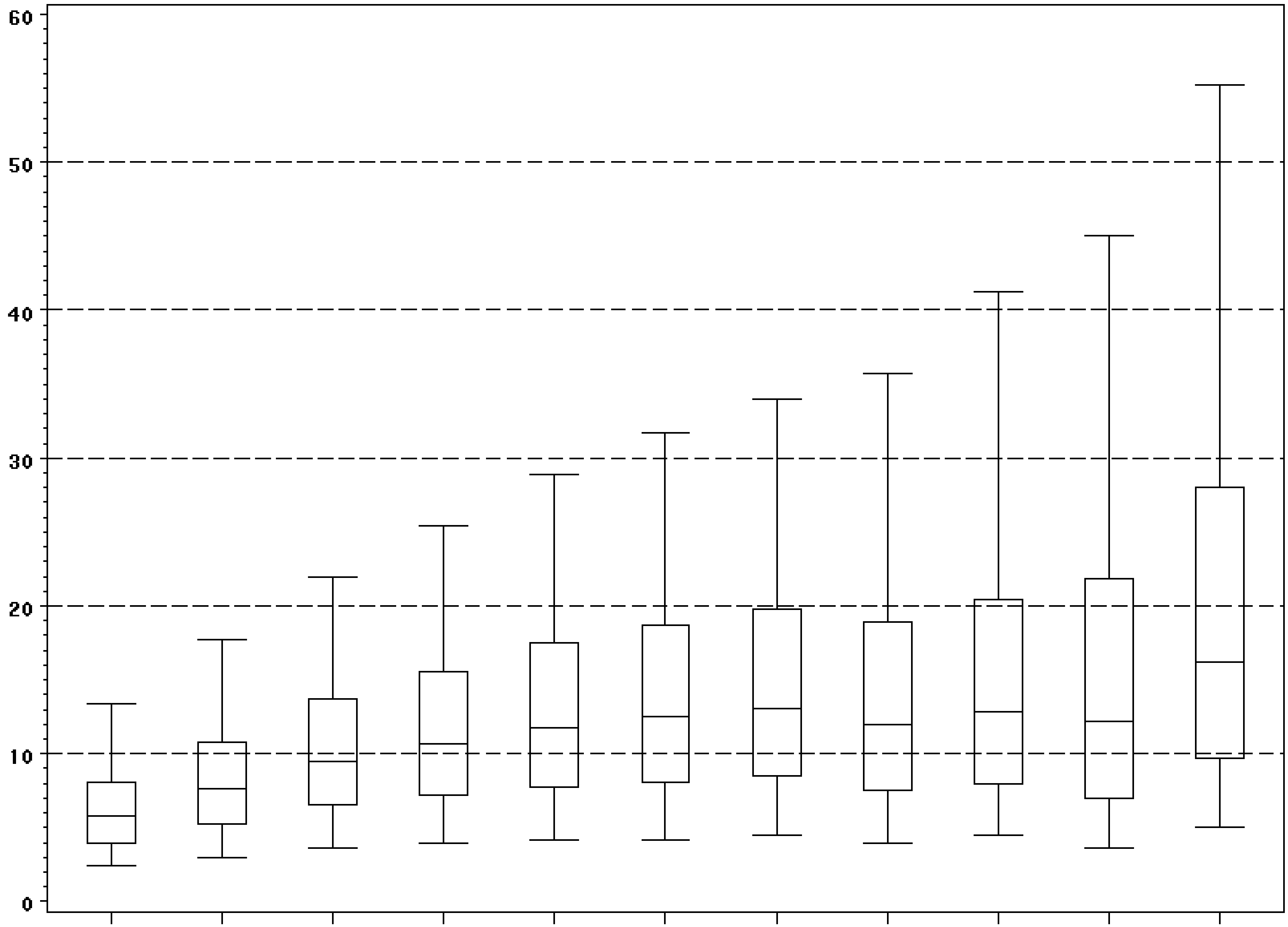
Percent and number of days PM2.5 concentrations exceeded the site 98th percentile DV by more than 5ug/m3, 2001-2003

		Days ≤ P98DV + 5 (but ≥ P98DV)						Days > P98DV + 5					
P98 DV	# sites	Minimum # Days	Mean # Days	Maximum # Days	Minimum % Days	Mean % Days	Maximum % Days	Minimum # Days	Mean # Days	Maximum # Days	Minimum % Days	Mean % Days	Maximum % Days
all	827	0	5.7	24	0.0%	1.3%	4.6%	0	4.1	20	0.0%	0.9%	2.8%
<20	81	0	6.3	21	0.0%	1.6%	3.7%	0	2.2	7	0.0%	0.6%	1.8%
20-25	81	0	7.1	22	0.0%	1.6%	3.4%	0	3.0	11	0.0%	0.7%	1.7%
25-30	192	1	5.6	24	0.3%	1.4%	4.6%	0	3.2	12	0.0%	0.8%	2.8%
30-35	197	0	5.7	19	0.0%	1.3%	3.6%	0	4.0	15	0.0%	0.9%	2.5%
35-40	179	0	5.4	18	0.0%	1.1%	2.3%	1	5.1	16	0.3%	1.1%	2.2%
40-45	53	0	5.7	16	0.0%	1.0%	2.0%	0	6.0	19	0.0%	1.1%	2.8%
45-50	12	0	4.1	13	0.0%	0.7%	1.3%	3	5.5	13	0.4%	1.3%	2.8%
50-55	11	0	4.0	9	0.0%	0.7%	1.7%	2	8.3	20	0.9%	1.4%	2.1%
55-60	7	0	4.9	9	0.0%	0.7%	1.2%	4	7.3	12	1.0%	1.3%	1.7%
60-65	7	3	4.9	7	0.7%	1.3%	2.1%	3	6.0	12	0.9%	1.5%	2.6%
>65	7	1	2.7	5	0.3%	0.4%	0.6%	4	9.9	18	1.2%	1.9%	2.8%

Maximum number of days in any one year (2001-2003) that a site exceeded it's 3-year 98th or 99th percentile DV.

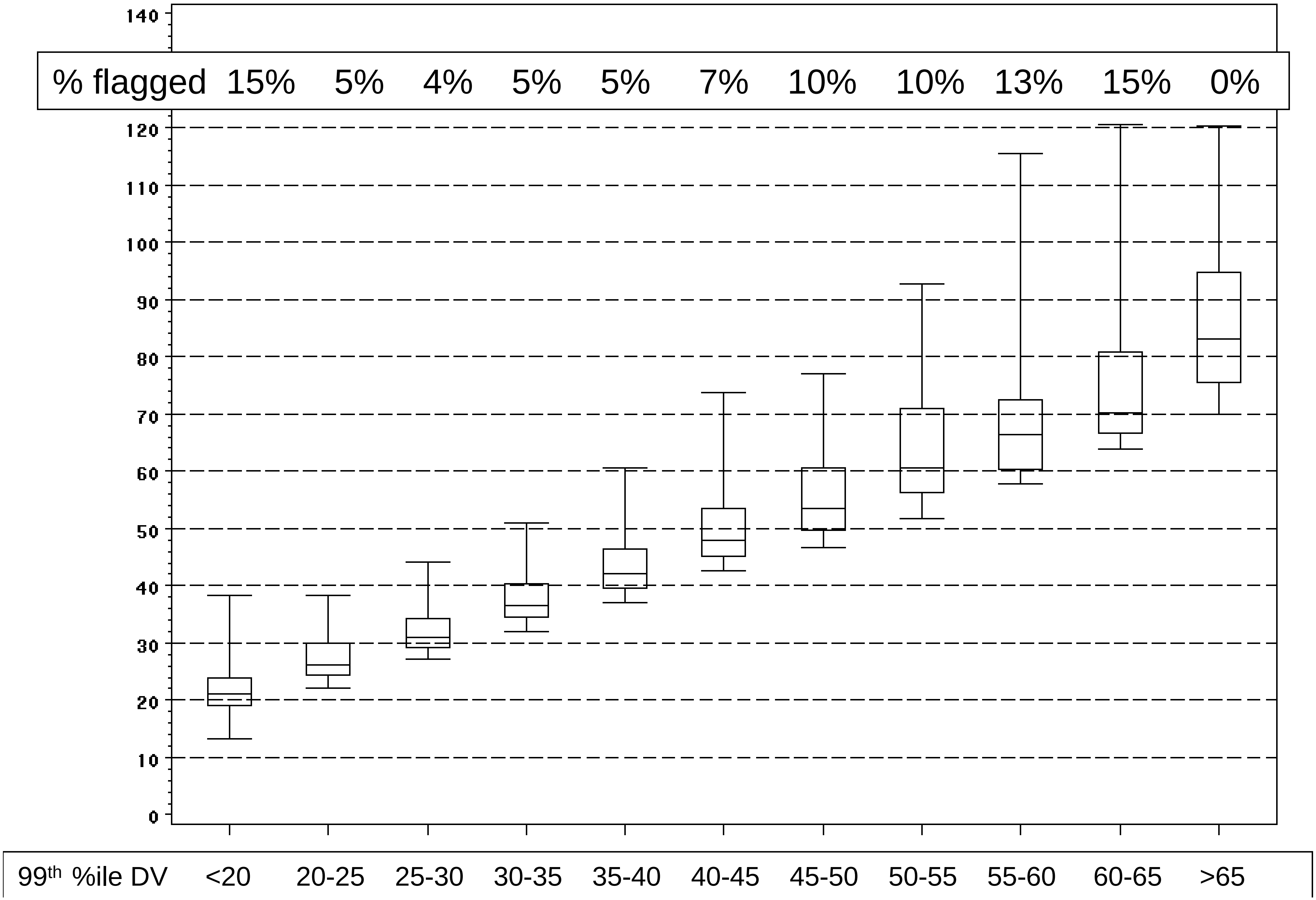
- Site 410290133 exceeded its 98th percentile DV (of 37ug/m3) 20 times in 2002.
- Site 410350004 exceeded its 99th percentile DV (of 65ug/m3) 13 times in 2002.
- The theoretical answer for both is 365 (or 365 for leap-year)!

Distribution of PM2.5 concentrations by 99th percentile 'DV' interval



99th %ile DV	<20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	>65

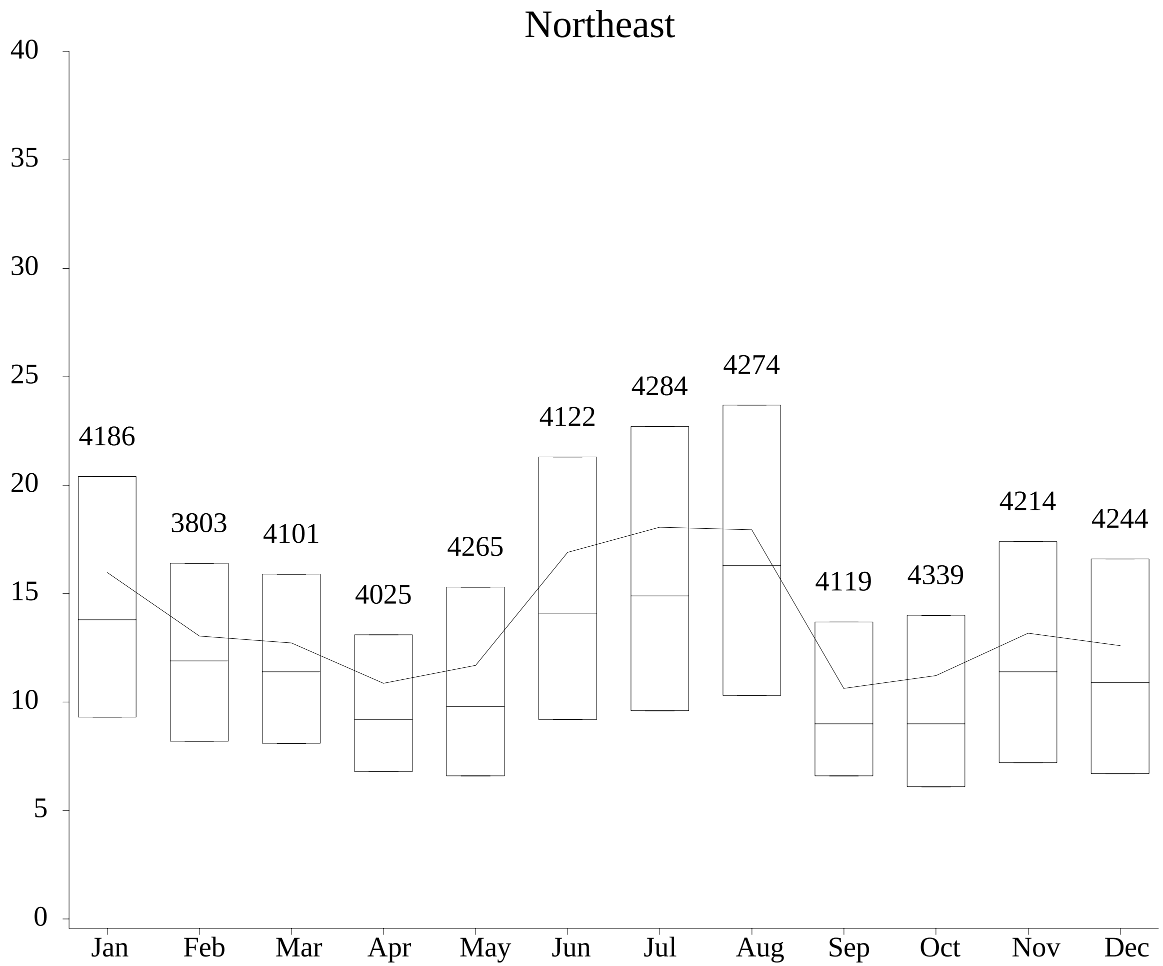
Distribution of PM2.5 concentrations > 99th percentile 'DV', by 99th percentile DV interval



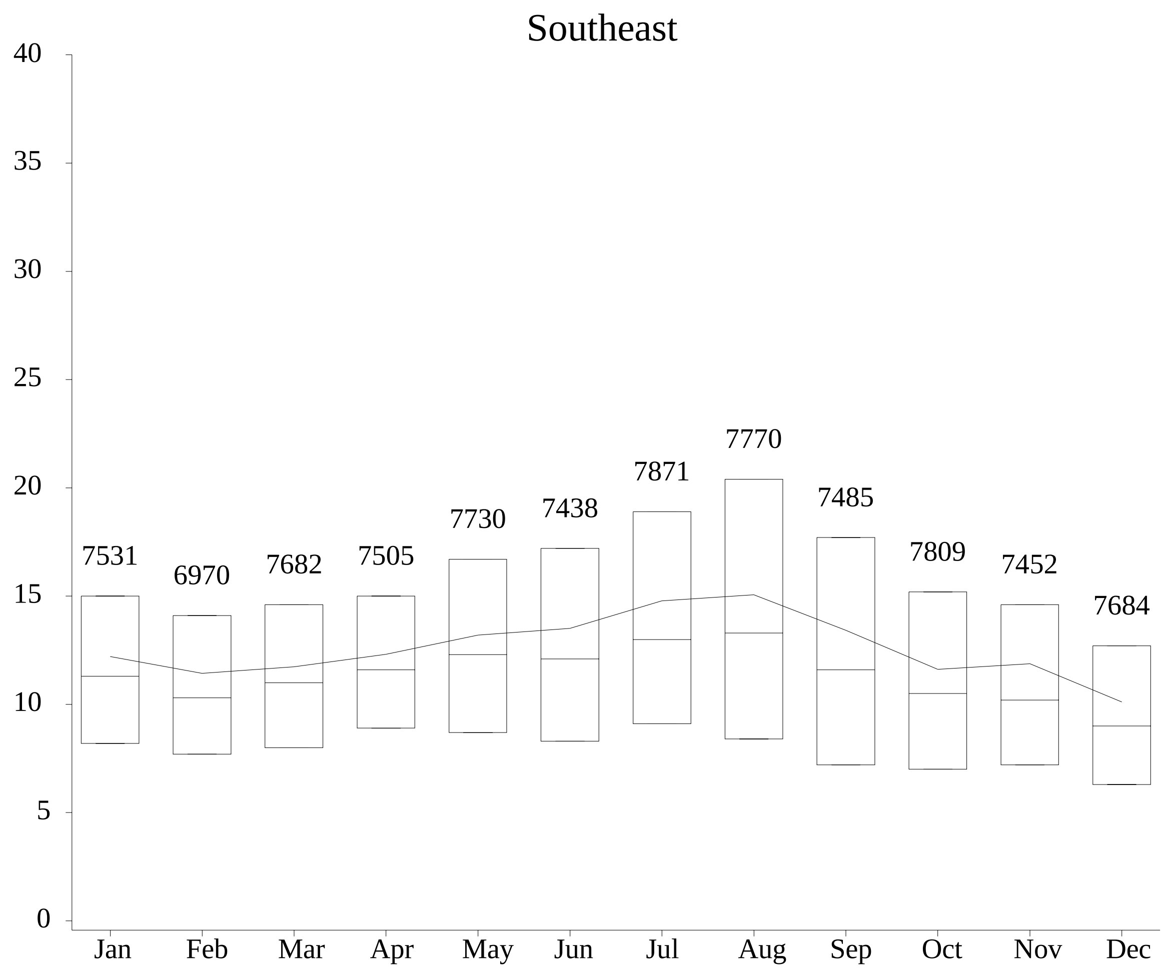
Percent and number of days PM2.5 concentrations exceeded the site 99th percentile 'DV' by more than 5ug/m3, 2001-2003

P99 DV	# sites	Days ≤ P99DV + 5 (but ≥ P99DV)						Days > P99DV + 5					
		Minimum # Days	Mean # Days	Maximum # Days	Minimum % Days	Mean % Days	Maximum % Days	Minimum # Days	Mean # Days	Maximum # Days	Minimum % Days	Mean % Days	Maximum % Days
all	827	0	3.4	14	0.0%	0.8%	3.9%	0	2.3	12	0.0%	0.5%	2.4%
<20	53	0	4.0	12	0.0%	1.0%	2.4%	0	1.4	6	0.0%	0.3%	0.9%
20-25	58	0	3.5	11	0.0%	0.9%	2.8%	0	1.4	6	0.0%	0.4%	1.0%
25-30	121	0	4.1	11	0.0%	1.1%	2.6%	0	1.6	7	0.0%	0.4%	2.4%
30-35	183	0	3.6	13	0.0%	0.9%	3.9%	0	1.9	7	0.0%	0.4%	1.1%
35-40	161	0	3.3	12	0.0%	0.8%	1.8%	0	2.3	8	0.0%	0.5%	1.4%
40-45	145	0	3.2	14	0.0%	0.6%	1.9%	0	3.0	8	0.0%	0.6%	2.1%
45-50	53	0	2.9	10	0.0%	0.5%	1.3%	0	3.4	7	0.0%	0.6%	1.4%
50-55	19	0	2.1	5	0.0%	0.4%	1.1%	1	3.7	12	0.3%	0.7%	1.5%
55-60	7	0	1.9	5	0.0%	0.4%	0.9%	2	3.6	7	0.4%	0.8%	1.1%
60-65	13	0	3.2	7	0.0%	0.6%	1.3%	1	4.0	10	0.3%	0.8%	1.3%
>65	14	0	2.1	6	0.0%	0.4%	0.9%	2	5.5	10	0.6%	1.2%	2.3%

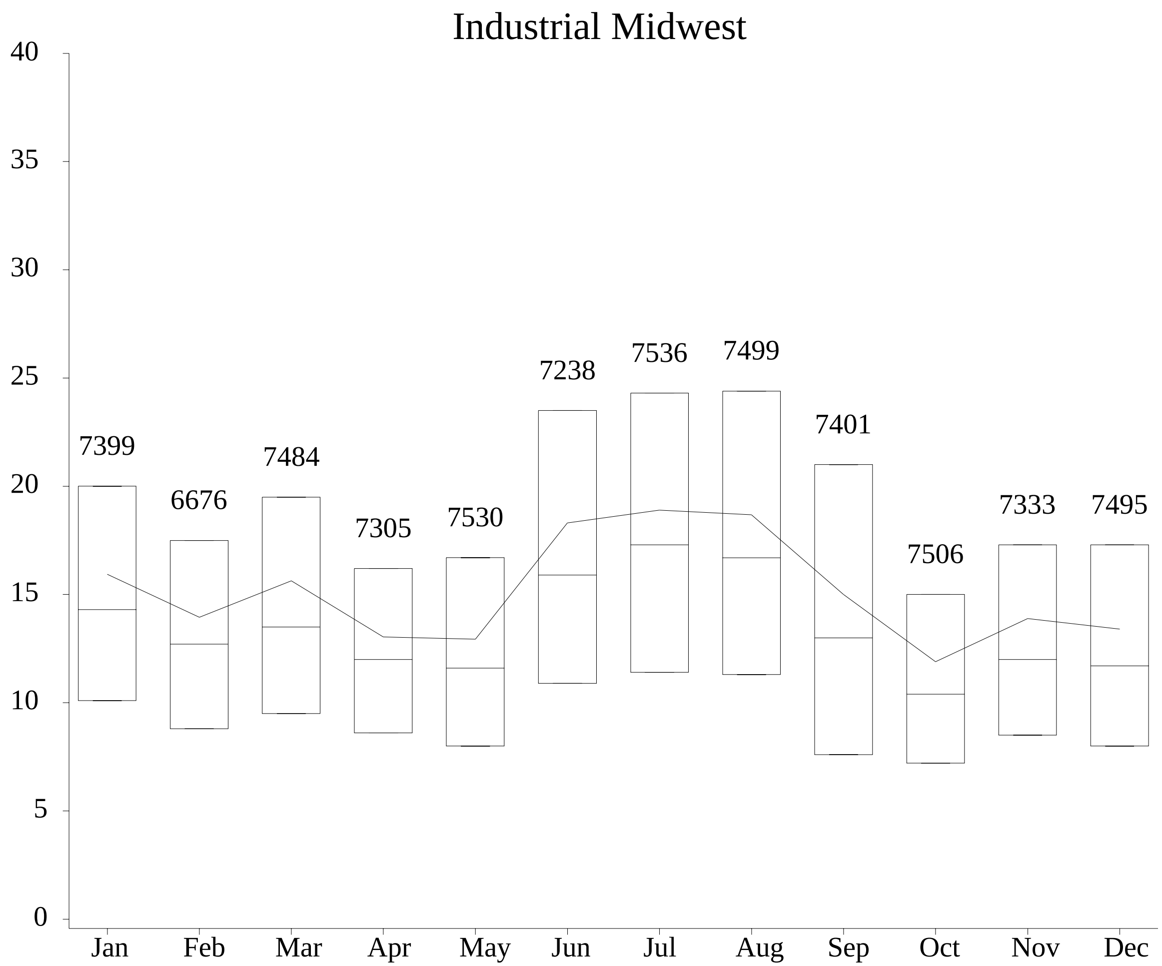
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



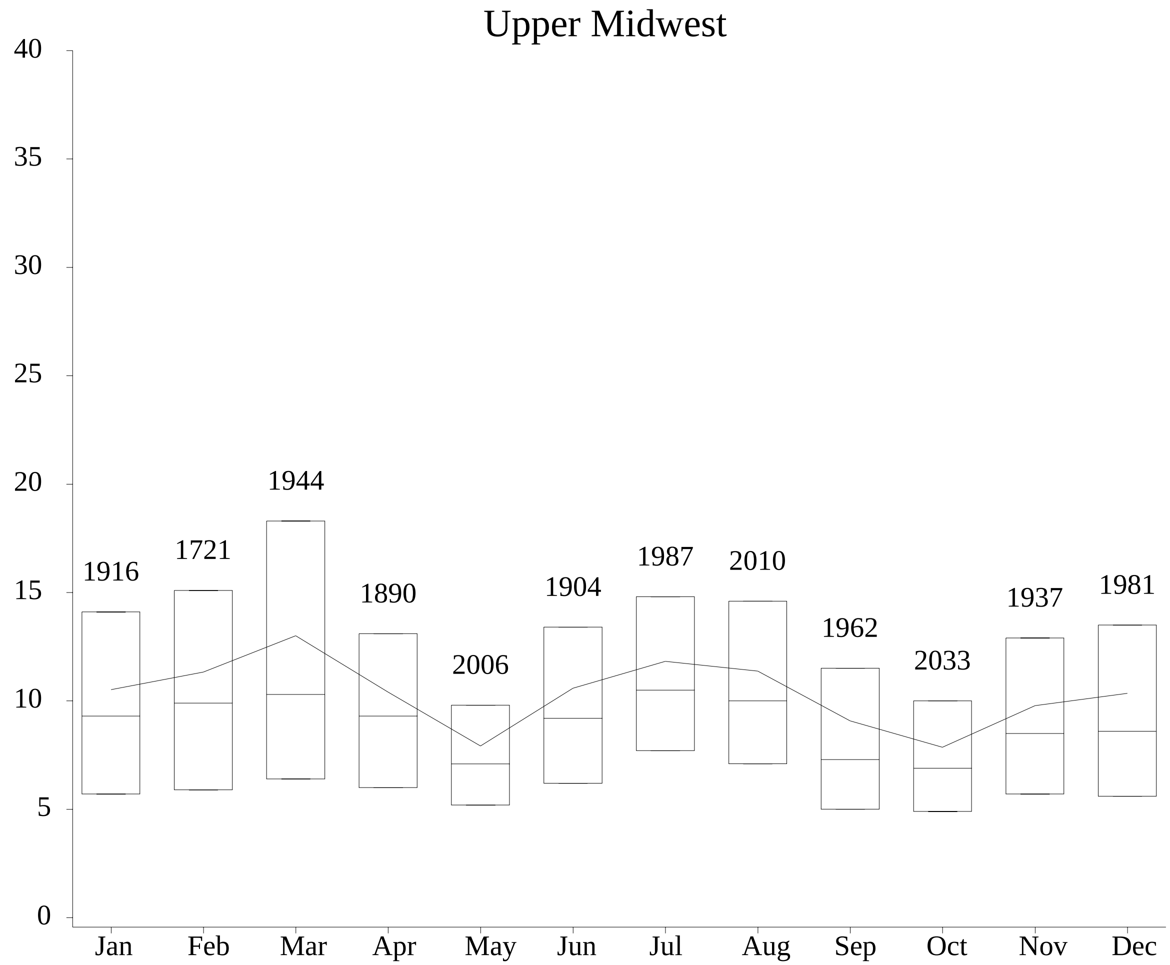
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



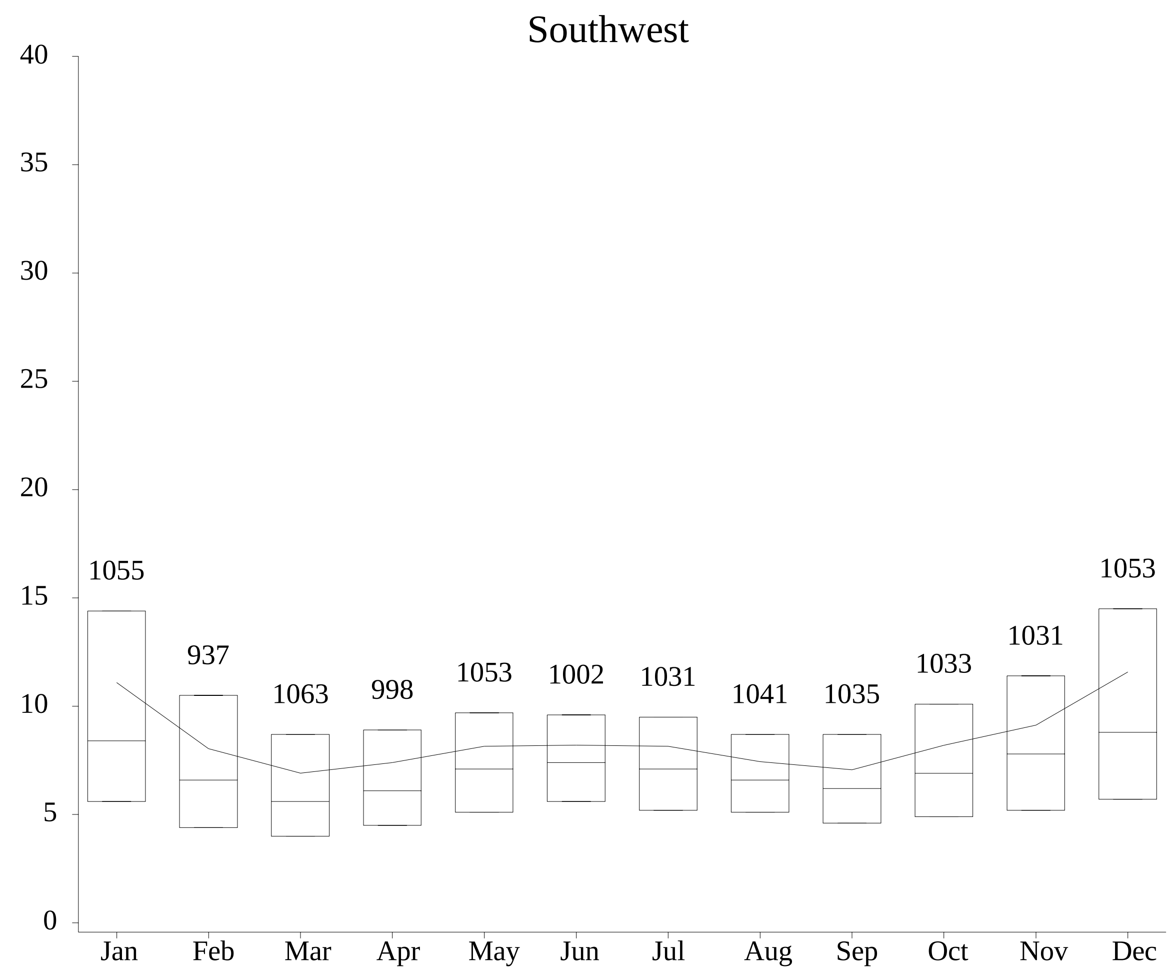
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



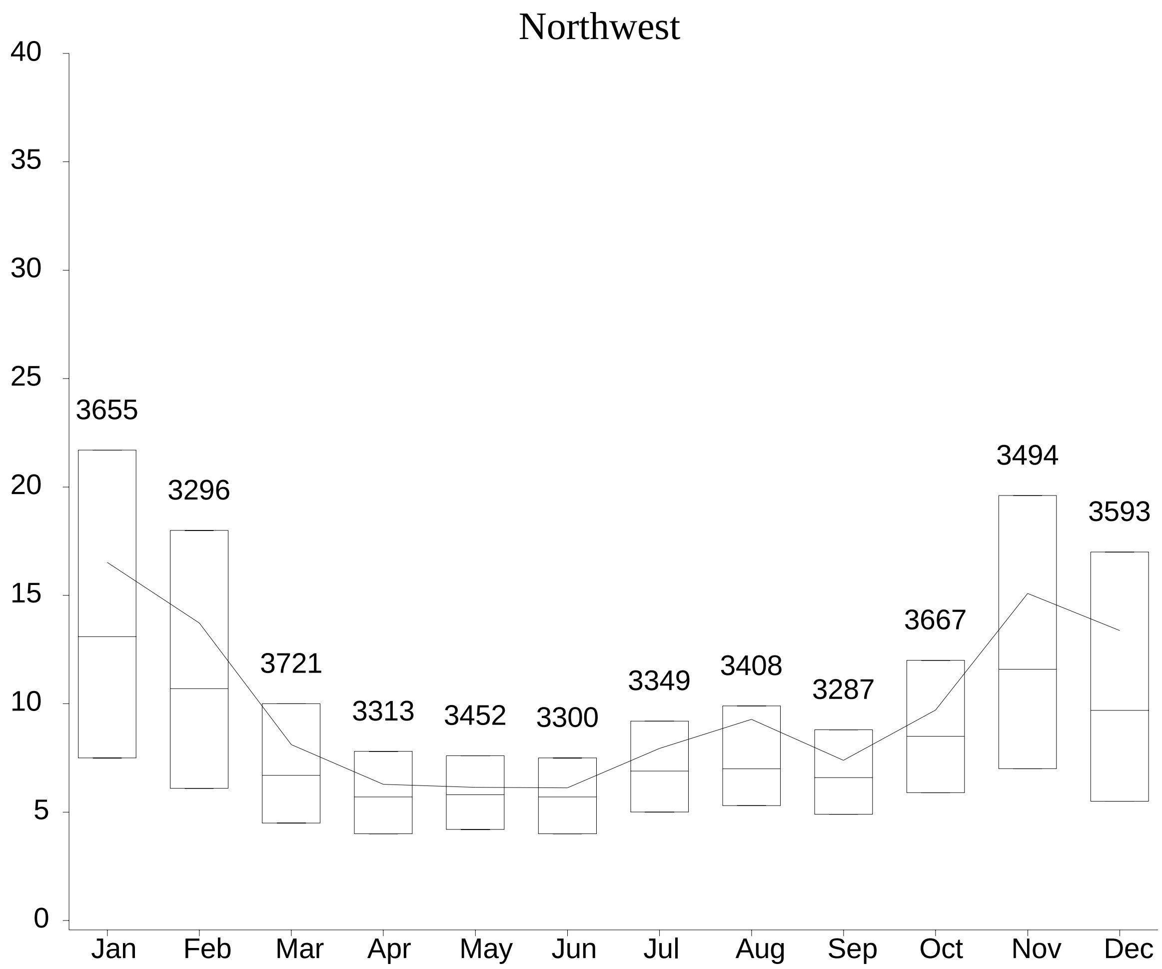
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



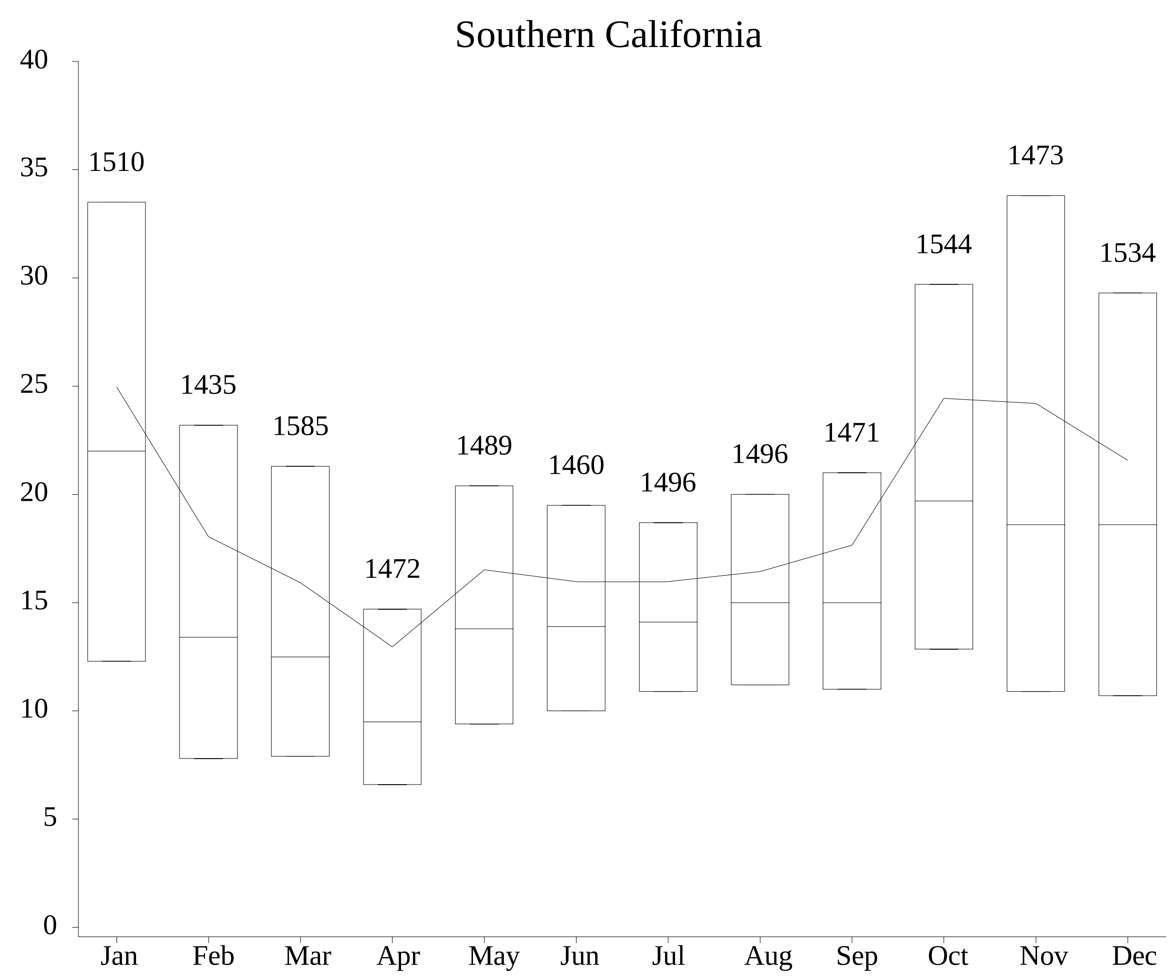
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



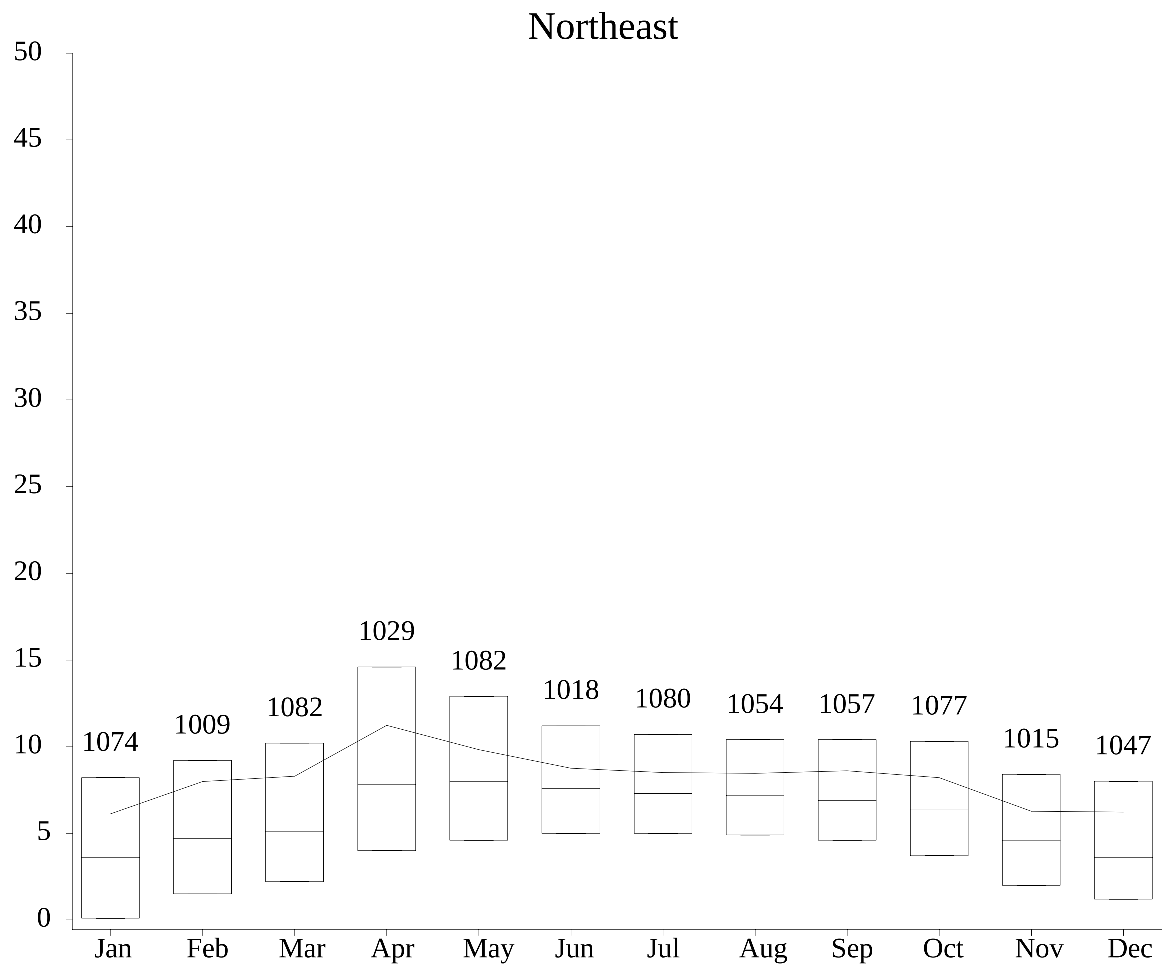
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



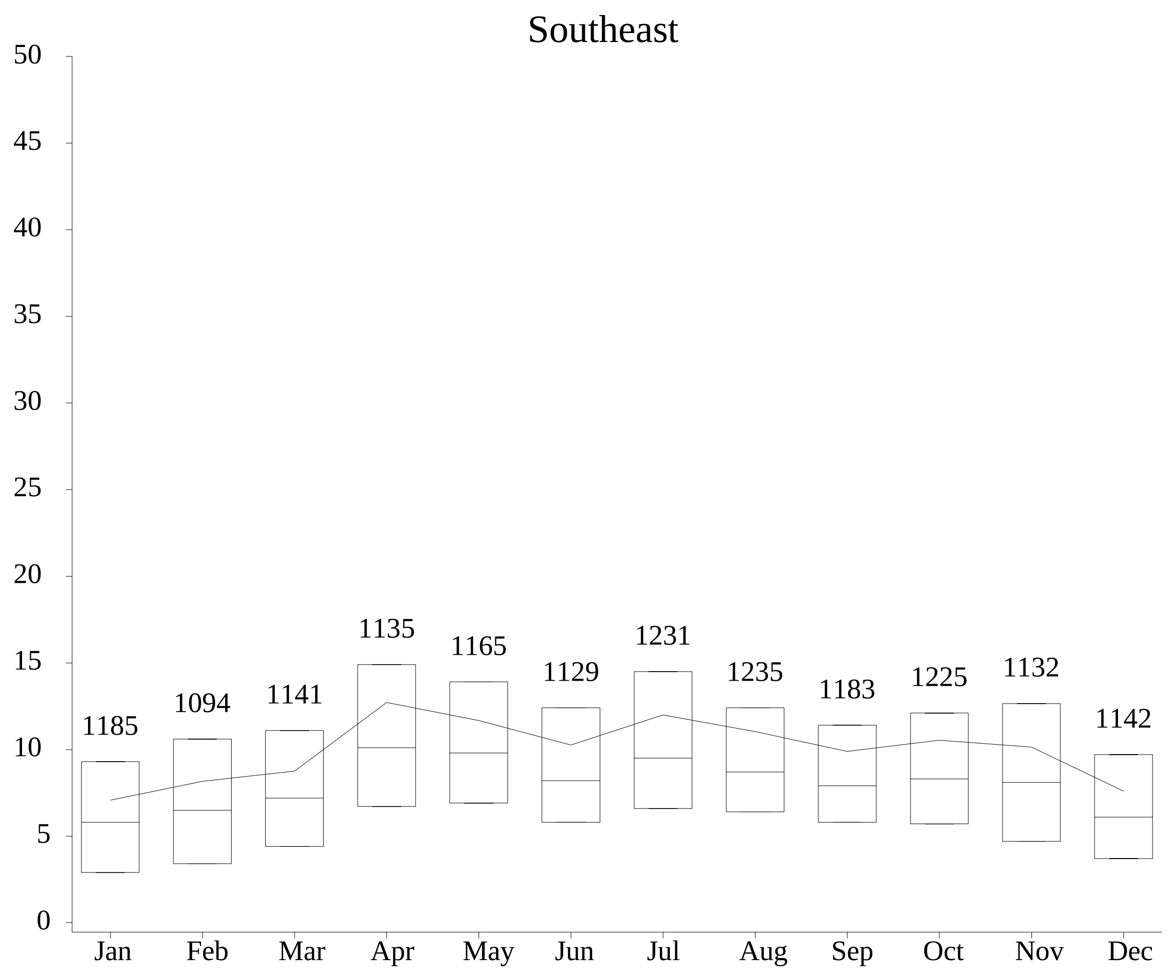
Urban 24-hour average $PM_{2.5}$ concentration distributions by region and month, 2001-2003.



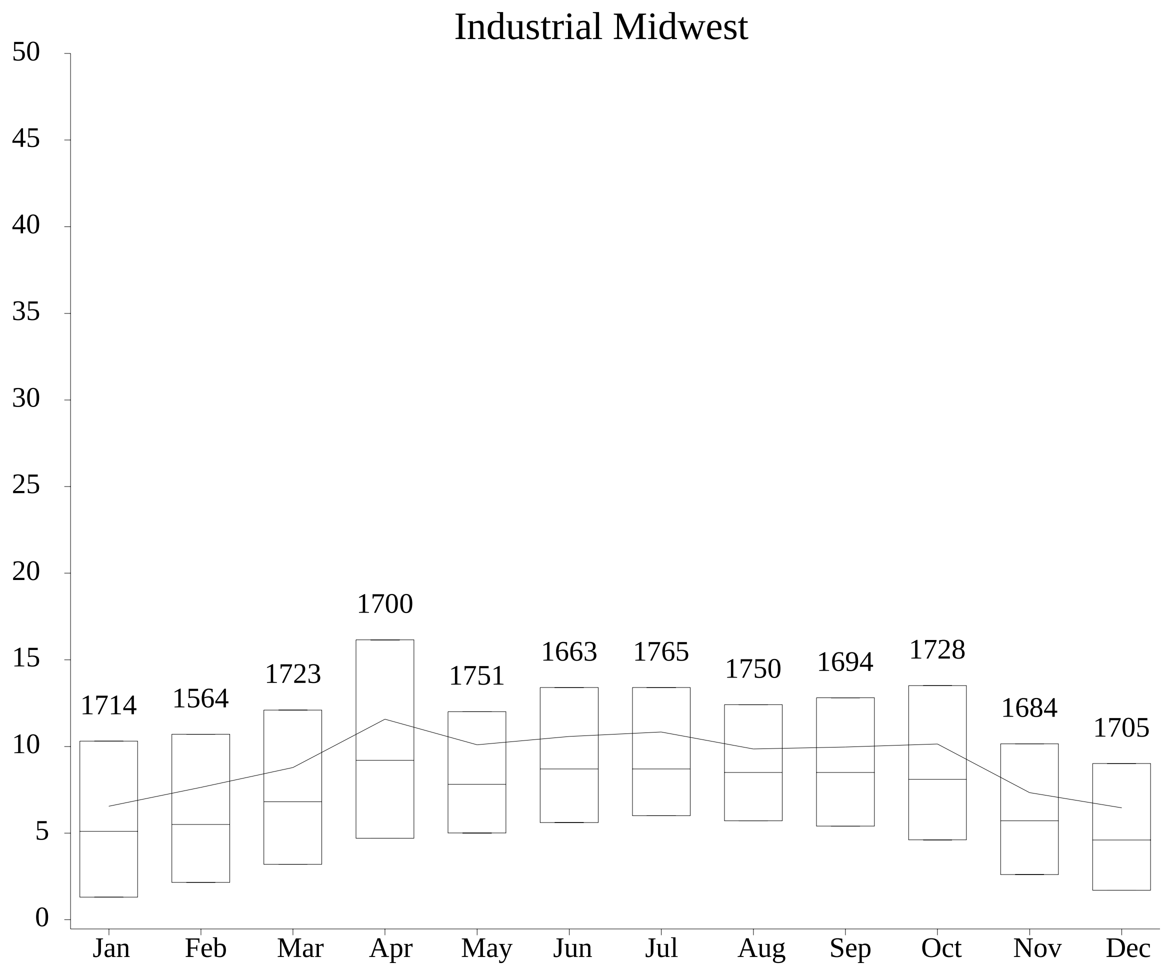
Urban 24-hour average $PM_{10\text{-}2.5}$ concentration distributions by region and month, 2001-2003.



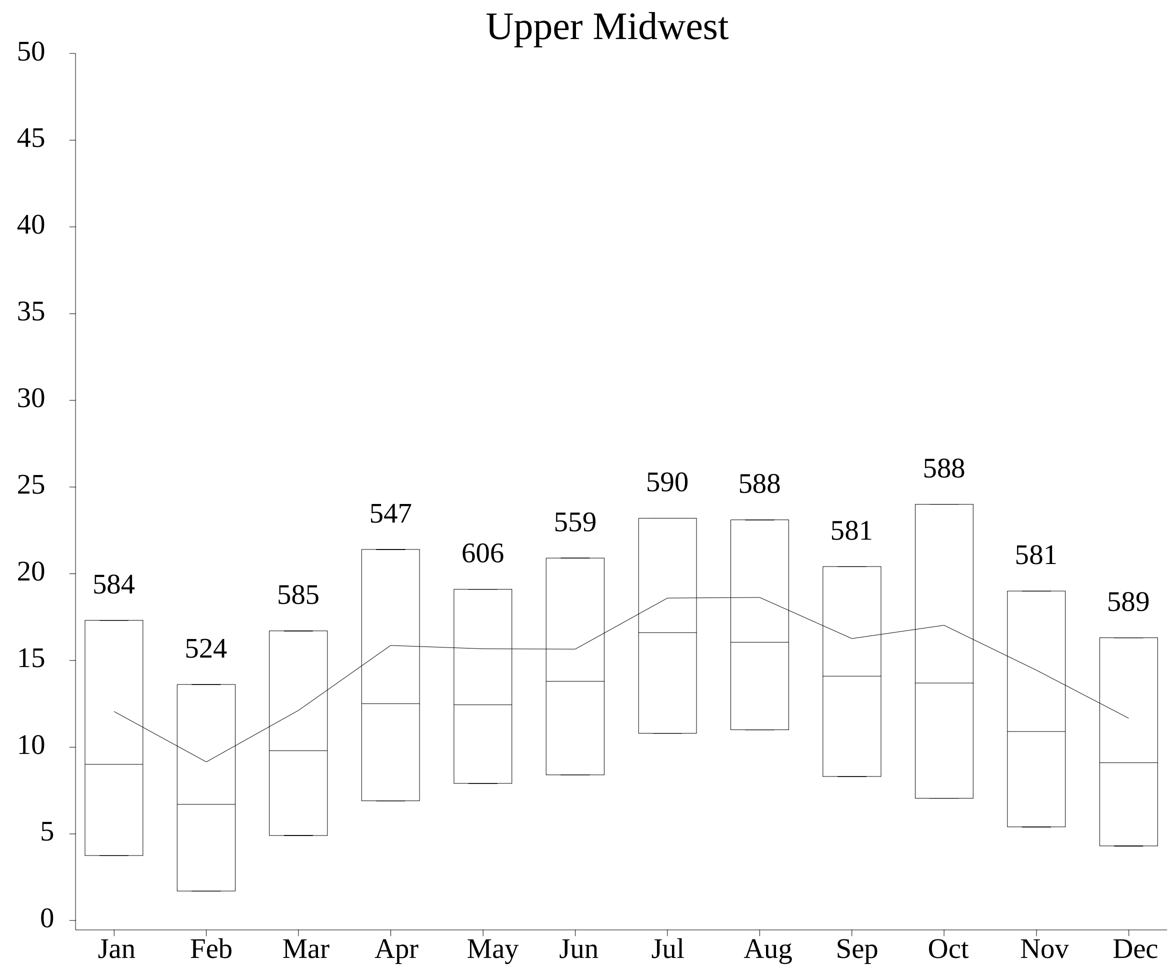
Urban 24-hour average $PM_{10\text{-}2.5}$ concentration distributions by region and month, 2001-2003.



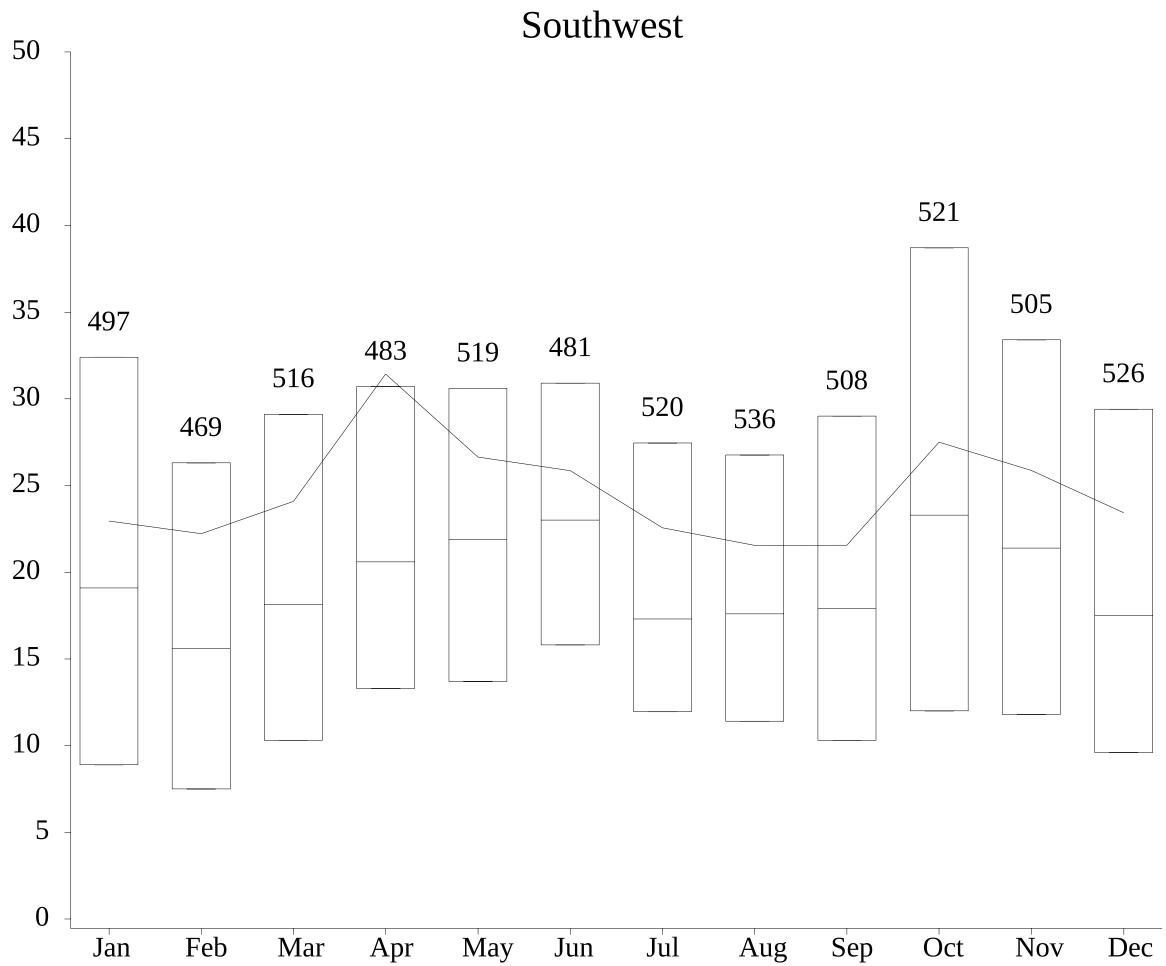
Urban 24-hour average $PM_{10\text{-}2.5}$ concentration distributions by region and month, 2001-2003.



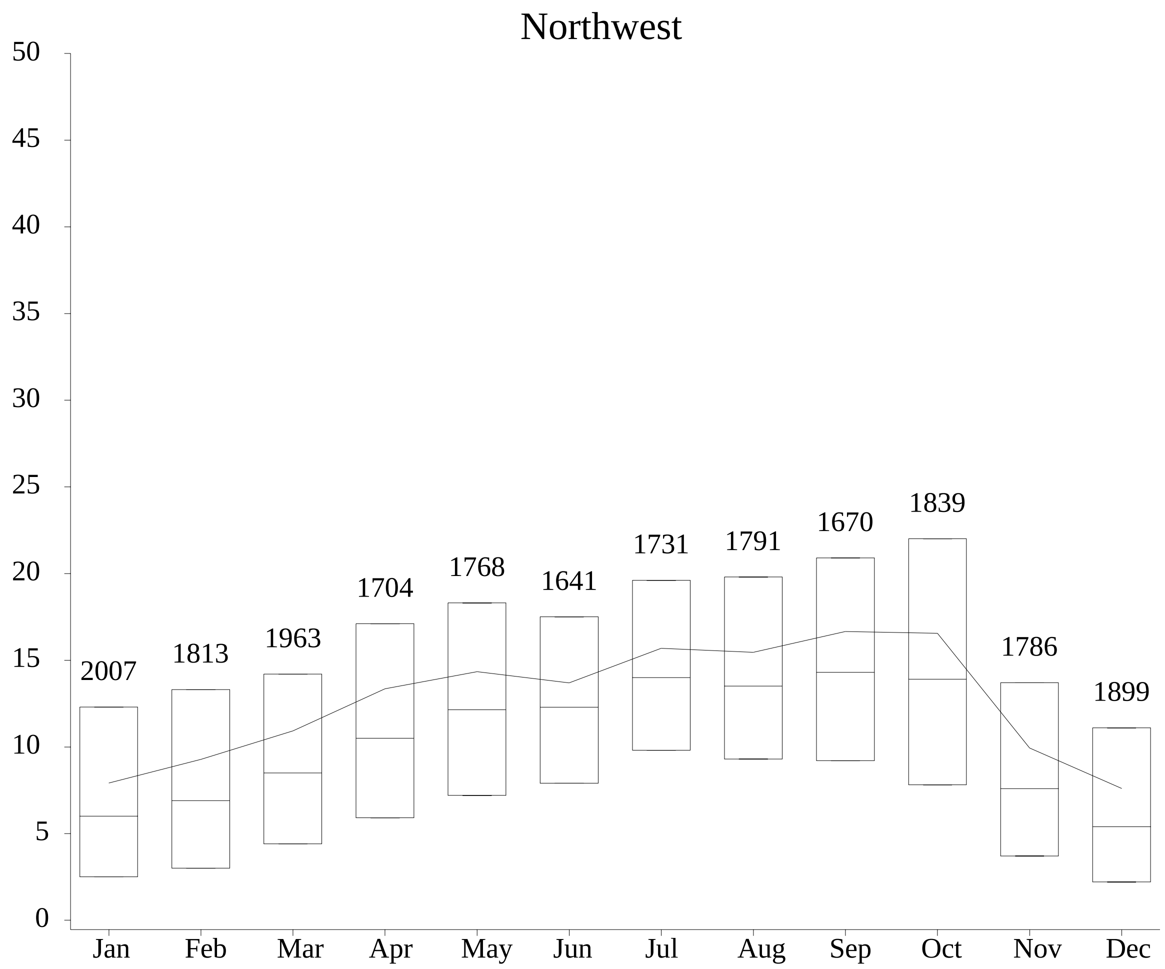
Urban 24-hour average $PM_{10\text{-}2.5}$ concentration distributions by region and month, 2001-2003.



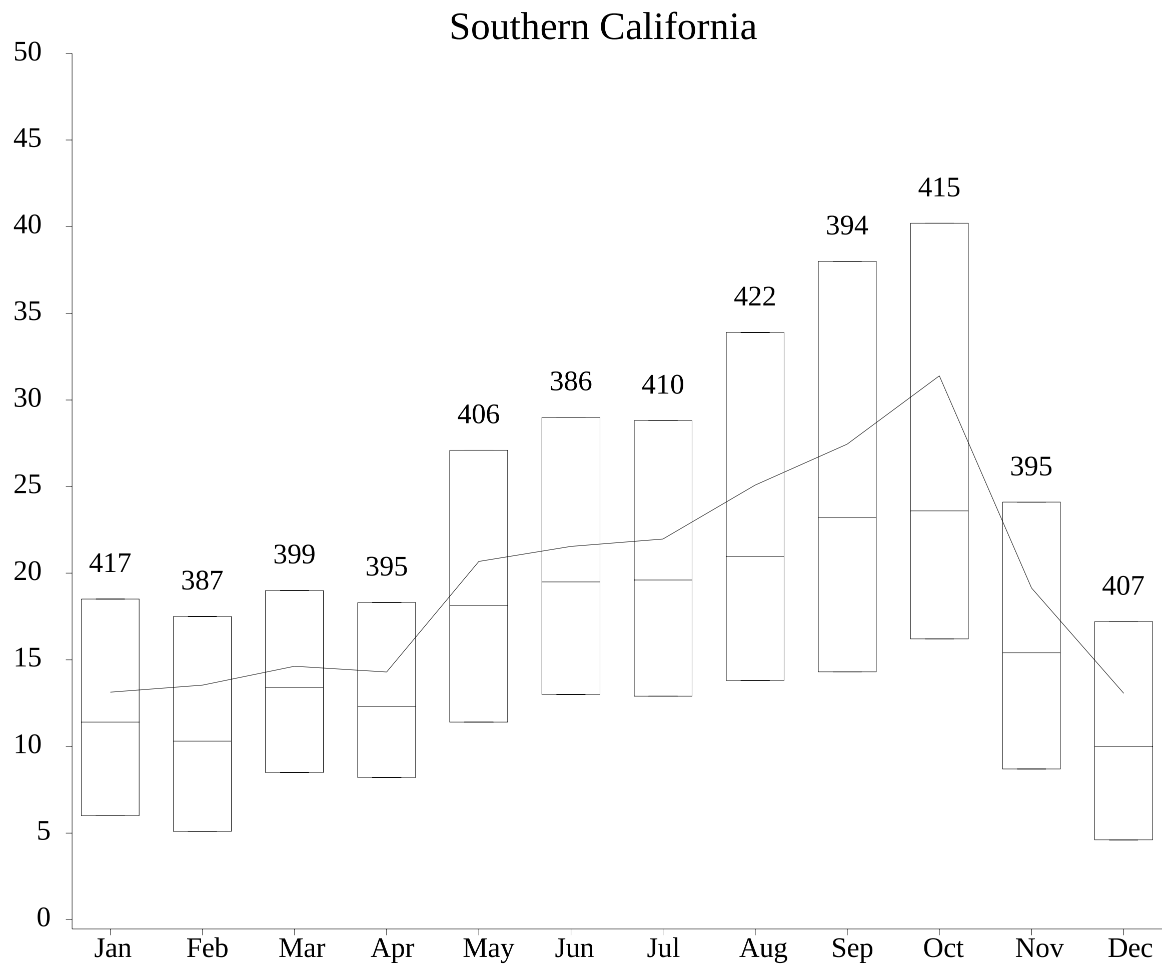
Urban 24-hour average $PM_{10-2.5}$ concentration distributions by region and month, 2001-2003.



Urban 24-hour average $PM_{10\text{-}2.5}$ concentration distributions by region and month, 2001-2003.



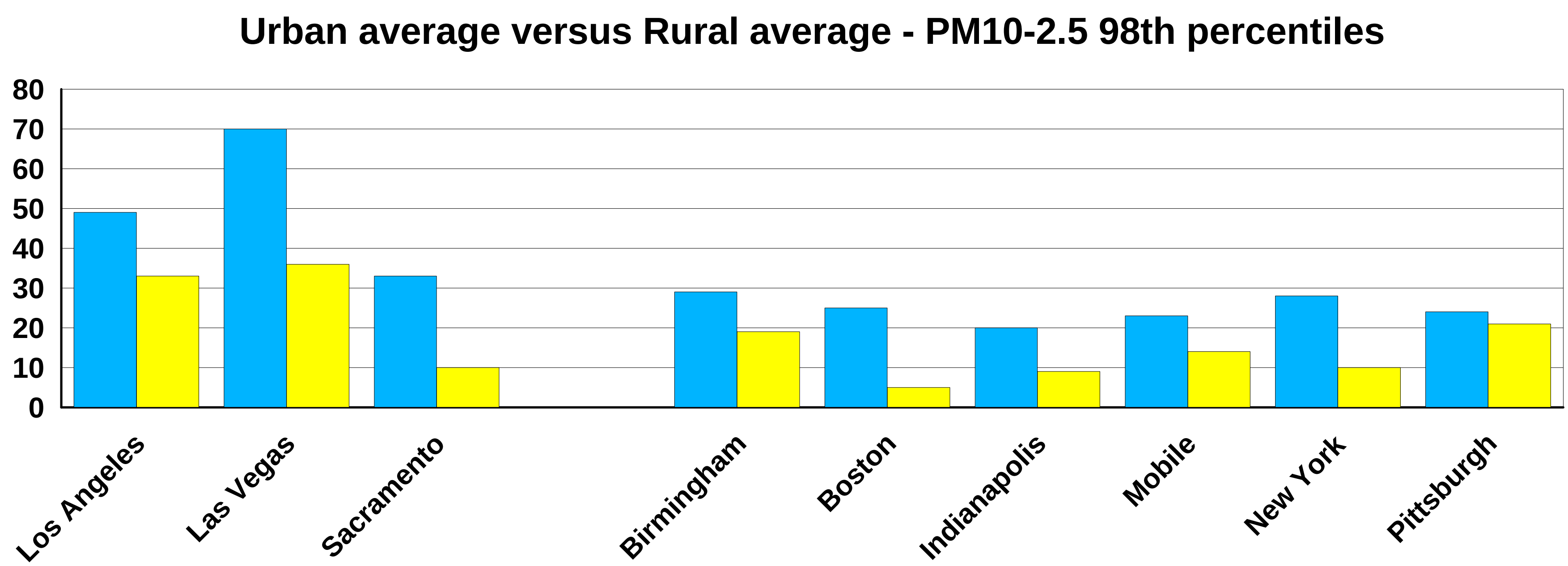
Urban 24-hour average $PM_{10\text{-}2.5}$ concentration distributions by region and month, 2001-2003.

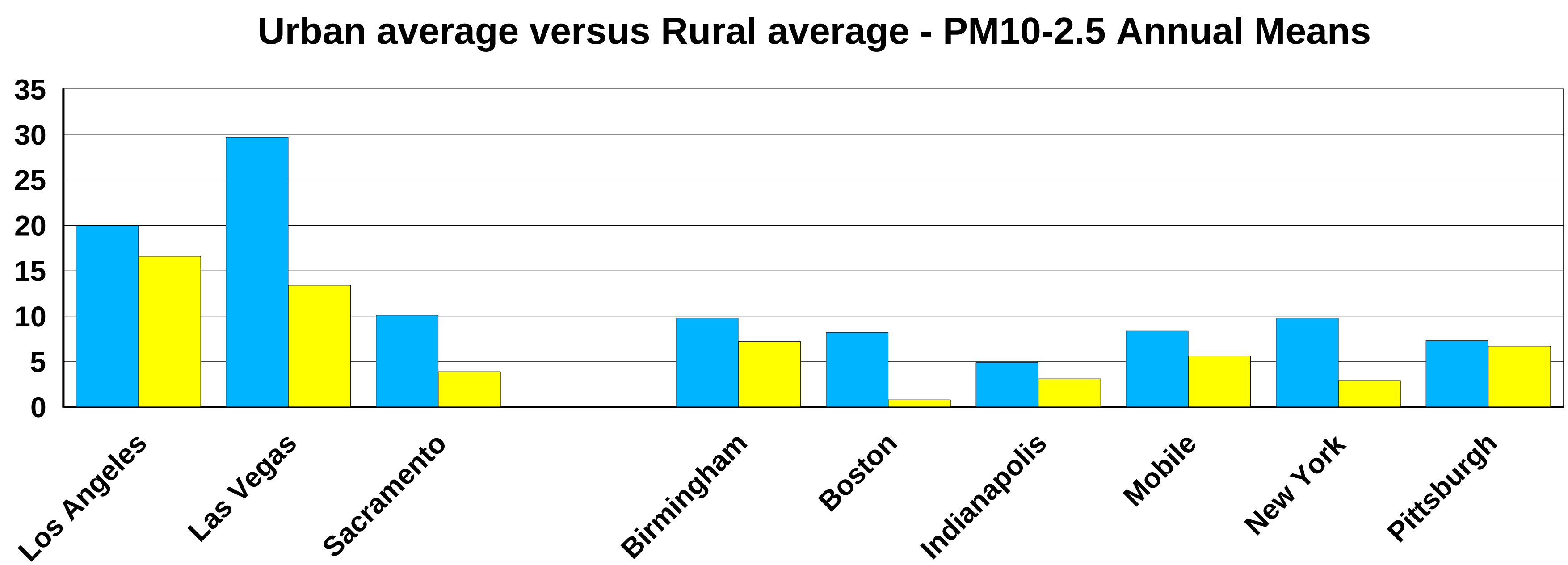


PM10-2.5 Urban / Rural Mass Comparison

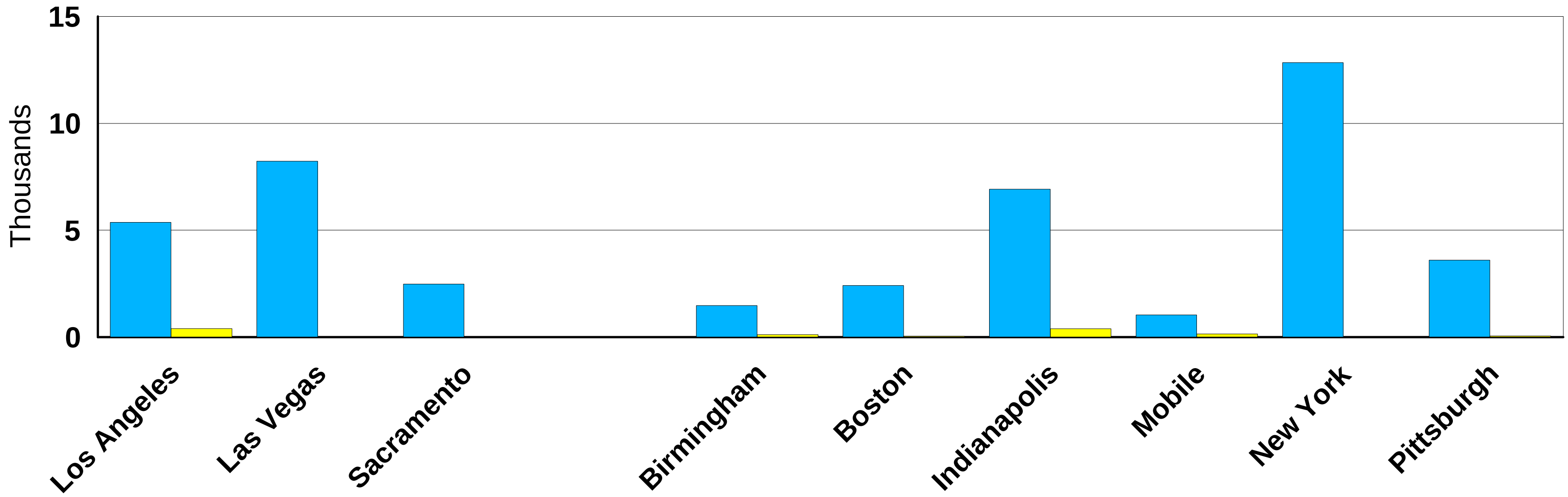
Analysis Details:

- Compared annual mean and 98th percentile levels for 'rural' sites in large metro areas to corresponding 'urban' ones
- Used CSA area definitions
- Used AQS 'Location Setting' (LS) field as 'rural'/'urban' indicator
  - 'rural' if LS is ''RURAL'
  - 'urban' if LS is 'URBAN AND CENTER CITY' or 'SUBURBAN'
- Looked in all CSAs with at least one 'rural' and 1 'urban' site
- Poulation density assigned to site by Census block group.


Urban average versus Rural average - PM10-2.5 98th percentiles
80
70
60
50
40
30
20
10
0
Los Angeles
Las Vegas
Sacramento
Birmingham
Boston
Indianapolis
Mobile
New York
Pittsburgh


Urban average versus Rural average - PM10-2.5 Annual Means
35
30
25
20
15
10
5
0
Los Angeles
Las Vegas
Sacramento
Birmingham
Boston
Indianapolis
Mobile
New York
Pittsburgh

Urban average versus Rural average - Population Density



Number of sites in CSA

	Los Angeles	Las Vegas	Sacramento		Birmingham	Boston	Indianapolis	Mobile	New York	Pittsburgh
Urban	10	3	7		5	3	2	1	12	8
Rural	2	2	2		2	1	1	1	1	1

Summary

Area	Urban Avg.	Rural Avg.	urban is x% larger
98th perentile			
Los Angele	49	33	48%
Las Vegas	70	36	94%
Sacrament	33	10	230%
Birminghar	29	19	53%
Boston	25	5	400%
Indianapoli	20	9	122%
Mobile	23	14	64%
New York	28	10	180%
Pittsburgh	24	21	14%
Annual mean			
Los Angele	20	16.6	20%
Las Vegas	29.7	13.4	122%
Sacrament	10.1	3.9	159%
Birminghar	9.8	7.2	36%
Boston	8.2	0.8	925%
Indianapoli	4.9	3.1	58%
Mobile	8.4	5.6	50%
New York	9.8	2.9	238%
Pittsburgh	7.3	6.7	9%
Pop Density			
Los Angele	5368	404	1229%
Las Vegas	8230	4	205650%
Sacrament	2479	9	27444%
Birminghar	1476	117	1162%
Boston	2413	42	5645%
Indianapoli	6920	387	1688%
Mobile	1033	150	589%
New York	12842		
Pittsburgh	3596	50	7092%

- **In all large metro areas (CSA's) with at least one 'rural' and one 'urban' sites. ...**
  - **The urban average 98th percentile is larger then the rural average 98th percentile**
  - **The urban average annual mean is larger than the rural average annual mean**
- The rural sites are in located in block groups with significantly lower population density

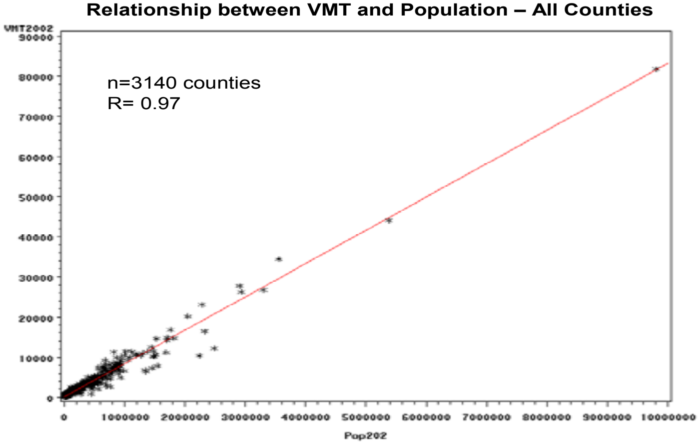
Characterizing areas as "urban"

- Goal:
  - Characterization of areas as urban (or non-urban) using different data sources as measures of urbanization
- Data used include
  - Population (e.g., CBSA/CSA size; density)
  - Traffic (e.g., vehicle miles traveled)
  - Location of industrial sources of PM10-2.5
- Note – AQS has a field (location_setting) that delineates monitoring sites into 3 categories: 'urban and center city,' 'suburban,' and 'rural.' This field, although historically utilized for urban/rural comparative analysis (including in the PM SP) is often inaccurate or misleading.

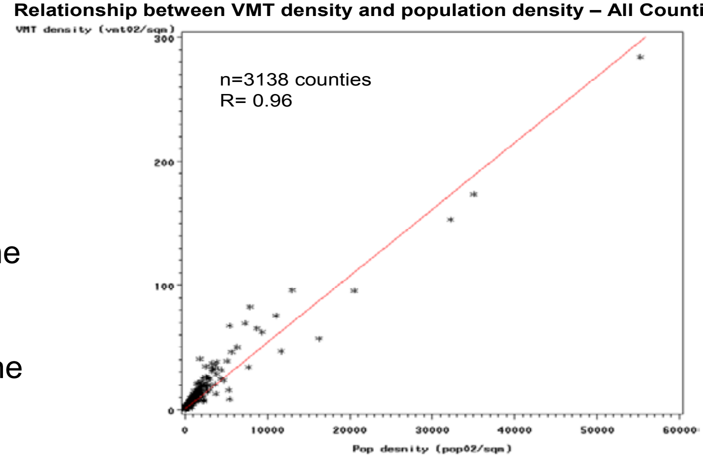
Vehicle Miles Traveled (VMT) Data

- As part of the National emission inventory process, VMT data are estimated for every U.S. county, but generally not for smaller geographic areas. Because PM10-2.5 is somewhat spatially heterogeneous within metro areas, demographic data for geographic entities smaller than counties would be more useful than county level info.
- The relationships between VMT and population, and between VMT density and population density were evaluated to see if population (and pop density) could be used as a surrogate for VMT (and VMT density). 2002 population and 2002 VMT were used. As seen below, population <u>can</u> be used as surrogate for VMT (level) and population density <u>can</u> be used as a surrogate for VMT density.

Relationship between VMT and Population – All Counties

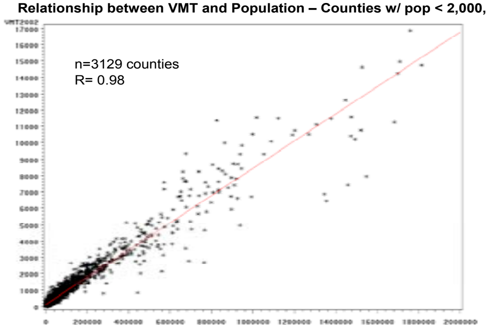


Relationship between VMT density and population density – All Counties

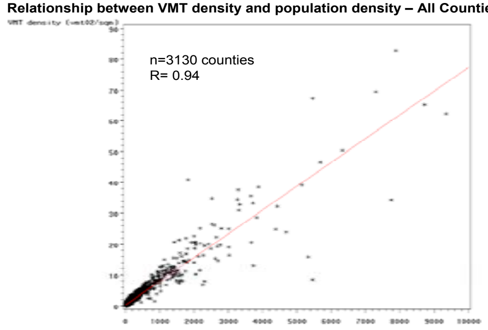


Since there appear to be some outliers that could be driving the relationship, the analysis was rerun for the lower end of the spectrum. Similar results were achieved.

Relationship between VMT and Population – Counties w/ pop < 2,000,000



Relationship between VMT density and population density – All Counties



CBSA and CSA Definitions

- In the December 27, 2000 Federal Register, OMB announced new standards for defining metropolitan areas. The new standards replace the previous MSA/CMSA definitions. Below are some key aspects of the new standards .

- The new standards will consider statistical rules only when defining Metropolitan and Micropolitan Statistical Areas.
- The Metropolitan and Micropolitan Statistical Area Standards do not equate to an urban-rural classification. All counties included in Metropolitan and Micropolitan Statistical Areas and many other counties contain both urban and rural territory and populations. OMB recognizes that formal definitions of settlement types such as inner city, inner suburb, outer suburb, exurb, and rural are useful to researchers, analysts, and other users of federal data. However, such settlement types are not considered for the statistical areas in this classification.
- Metropolitan and Micropolitan Statistical Areas will be called collectively Core Based Statistical Areas (CBSAs). Metropolitan Statistical Areas will be based on urbanized areas of 50,000 or more population and Micropolitan Statistical Areas will be based on urban clusters of at least 10,000 but less than 50,000 population. The location of these cores will be the basis for identifying the central counties of CBSAs. The use of urbanized areas as cores for Metropolitan Statistical Areas is consistent with current practice. Urban clusters, used to identify the Micropolitan Statistical Areas, will be identified by the Census Bureau following Census 2000 and will be conceptually similar to urbanized areas.
- Counties will be the geographic building blocks. Counties will be the geographic building blocks for defining CBSAs throughout the United States and Puerto Rico.
- Commuting patterns will determine how many counties are part of the CBSA. Journey to work, or commuting, will be the basis for grouping counties together to form CBSAs. A county qualifies as a CBSA county if (a) at least 25 percent of the employed residents of the county work in the CBSA's central county or counties, or (b) at least 25 percent of the jobs in the potential outlying county are accounted for by workers who reside in the CBSA's central county or counties. Measures of settlement structure, such as population density, will not qualify outlying counties for inclusion in CBSAs.
- 2 or more related CBSA's can be combined into a consolidated Statistical Area (CSA)

•*Although being part of (within) a CBSA does not necessarily indicate urbanization (as noted in bullet 2 above), we will assume that highly densely populated regions of larger metropolitan CBSA's (in regards to population) are, in fact, 'urban'.*

•*Because PM10-2.5 is somewhat spatially heterogeneous, it makes more sense to consider overall area size according to CBSA rather than CSA definitions.*

CBSA Counts (CBSA's as potential default PM10-2.5 NA areas)

- Note: Numbers below include VI and PR. Population is from 2000 Census. Total U.S. pop = 285M.
- There are 935 total CBSA's.
  - They account for 265M total population (93% of U.S.)
- There are 370 'metropolitan' CBSA's
  - They account for 236M total population (83% of U.S.)
- There are 340 'metropolitan' CBSA's with population > 100K
  - They account for 234m total population (82% of U.S.)
- There are 195 'metropolitan' CBSA's with population > 200K
  - They account for 213M total population (75% of U.S.)

•*As an initial criterion for prospective PM10-2.5 NA area evaluations, **a CBSA population cutoff of 100K** or 200K appears reasonable.*
•*Only 'metropolitan' CBSA's will initially be considered.*

Census 'Tract'/'Block'/'Block Group' Definitions

Census tract

A small, relatively permanent statistical subdivision of a county delineated by a local committee of census data users for the purpose of presenting data. Census tract boundaries normally follow visible features, but may follow governmental unit boundaries and other non-visible features in some instances; they always nest within counties. Designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment, census tracts average about 4,000 inhabitants. They may be split by any sub-county geographic entity.

Block

A subdivision of a census tract (or, prior to 2000, a block numbering area), a block is the smallest geographic unit for which the Census Bureau tabulates 100-percent data. Many blocks correspond to individual city blocks bounded by streets, but blocks - especially in rural areas - may include many square miles and may have some boundaries that are not streets. The Census Bureau established blocks covering the entire nation for the first time in 1990. Previous censuses back to 1940 had blocks established only for part of the nation. Over 8 million blocks are identified for Census 2000.

Block group (BG)

A subdivision of a census tract (or, prior to 2000, a block numbering area), a block group is the smallest geographic unit for which the Census Bureau tabulates sample data. A block group consists of all the blocks within a census tract with the same beginning number. Example: block group 3 consists of all blocks within a 2000 census tract numbering from 3000 to 3999. There are approximately 208 thousand BG's in the U.S. (excluding PR and VI). A block group typically contains about 1,100 inhabitants

- *Using Census geographic entities provides inherent confidence in corresponding population estimates*
- *For evaluating population density, counties are generally considered too big and too variable. Block groups are a compromise between blocks (the smallest Census entity) and tracts (larger areas).*
- *BG population density will be used (in tandem w/ CBSA population) for prospective PM10-2.5 NA area evaluations*
  - ***A value of 500 people per square mile was selected for use** in subsequent 'urban' SP analyses. About 85% of the sites in the PM10-2.5 extended db that fit the CBSA criteria of 'metro' and 100+K population are in BG's with > 500 pop density. 579 of the 712 total sites in the extended db meet the CBSA criterion alone and 491 meet both the CBSA and BG density criteria*
    - *Using AQS location setting field, only 7 of the 491 'urban' sites have a location setting of 'RURAL'*

PM10-2.5 Equivalence to PM10 (urban emphasis)

Analysis Goal:

- Estimate a PM10-2.5 daily standard level (for 98th and 99th percentile forms, urban environments) that would be ‘equivalent’ to the current PM10 exceedance-based daily NAAQS of level, 150 mg/m3.

Estimate a PM10-2.5 daily standard level (98th percentile form, urban environments) that would be 'equivalent' to the current PM10 exceedance-based daily NAAQS of level, 150 $\mu g/m^3$.

Notes:

- Both techniques utilize collocated site-level PM10 and PM10-2.5 DV's.
- All sites used meet the 'urban' criteria of 100+K CBSA pop, 500+ pop density for Census block group
- All results were aggregated to regional, 'east' / 'west', and U.S. levels

Methods:

1. Model (linear regression) PM10-2.5 98th and 99th percentile DV's as function of PM10 expected exceedance (ee) concentration-equivalent DV. Use relationship (intercept & slope) and current PM10 NAAQS level of 150 to estimate an associated PM10-2.5 NAAQS levels.
2. Calculate key distribution stats (quartiles, medians, means) for ratios of: (PM10-2.5 98th percentile) / (PM10 ee concentration-equivalent DV) and (PM10-2.5 99th percentile) / (PM10 ee concentration-equivalent DV). Multiply stats by 150 to estimate ranges of PM10-2.5 NAAQS levels

1. Model (linear regression) $PM_{10-2.5}$ 98^{th} (and 99^{th}) percentile DV as function of PM_{10} expected exceedance (ee) concentration-equivalent DV. Use relationship (intercept & slope) and current PM_{10} NAAQS level of 150 to estimate associated $PM_{10-2.5}$ NAAQS levels.

Notes:

- The PM10 ee concentration-based DV's are valid (for official DV's) even if capture is incomplete or there are missing years. 'Complete' sites (as ascertained for the annual std) should have more reliable daily DV's.
- For general PM10-2.5 SP characterization, we utilized a db representing 4, 8, or 12 quarters (not necessarily synonymous with calendar years); DV's from the 12-quarter sites will match better temporally with the collocated 'complete' PM10 data (though the number of sites will be less).
- Hence, only data for 'complete' PM10 sites that are also 12-quarter PM10-2.5 sites were used.

98^{th} percentile

PMREG	PMREGDESC	Intercept	slope	dof	R-square		PMc_P98 (for 150 PMt)
1	Northeast	6.24	0.21	16	0.07		**38**
2	Southeast	6.08	0.31	36	0.44		**52**
3	Industrial Midwest	-3.24	0.44	61	0.59		**63**
4	Upper Midwest	-6.35	0.58	15	0.58		**80**
5	Southwest	1.61	0.40	12	0.85		**62**
6	Northwest	0.08	0.38	37	0.39		**57**
7	Southern California	0.47	0.37	14	0.69		**55**
	East	-0.64	0.39	117	0.49		**58**
	West	2.01	0.38	84	0.69		**59**
	U.S.	-0.20	0.39	203	0.68		**59**
	average of 7 regions (level)						**58**

99^{th} percentile

PMREG	PMREGDESC	Intercept	slope	dof	R-square		PMc_P98 (for 150 PMt)
1	Northeast	11.64	0.21	16	0.03		**43**
2	Southeast	3.30	0.44	36	0.62		**69**
3	Industrial Midwest	-0.46	0.47	61	0.54		**70**
4	Upper Midwest	-5.26	0.68	15	0.62		**97**
5	Southwest	11.17	0.44	12	0.80		**77**
6	Northwest	1.62	0.44	37	0.45		**68**
7	Southern California	-4.51	0.50	14	0.71		**71**
	East	0.35	0.45	117	0.50		**68**
	West	2.78	0.47	84	0.70		**73**
	U.S.	-0.56	0.48	203	0.70		**72**
	average of 7 regions (level)						**71**

1. Model (linear regression) PM10-2.5 98th (and 99th) percentile DV as function of PM10 expected exceedance (ee) concentration-equivalent DV. Use relationship (intercept & slope) and current PM10 NAAQS level of 150 to estimate associated PM10-2.5 NAAQS levels.

95% confidence intervals for Method 1 estimates for independent variable value of 150

98th Percentile

Area	predicted (98th percentile UPM10-2.5)	lclm	uclm
Reg1	37.5	6.7	68.3
Reg2	52.1	42.4	61.9
Reg3	62.5	55.1	69.9
Reg4	80.0	59.6	100.3
Reg5	61.5	54.0	69.1
Reg6	56.5	46.8	66.2
Reg7	55.4	45.5	65.3
East	57.6	51.8	63.4
West	59.2	55.3	63.0
U.S.	58.6	55.8	61.4

99th Percentile

Area	predicted (99th percentile UPM10-2.5)	lclm	uclm
Reg1	42.8	-2.9	88.5
Reg2	68.9	59.2	78.6
Reg3	70.0	61.3	78.7
Reg4	96.9	75.0	118.8
Reg5	77.0	66.9	87.0
Reg6	67.5	57.4	77.7
Reg7	70.5	57.5	83.5
East	68.0	61.4	74.5
West	72.9	68.3	77.4
U.S.	71.6	68.3	74.9

2 Calculate key distribution stats (quartiles, medians, means) for ratios of: (PM10-2.5 98th percentile) / (PM10 ee concentration-equivalent DV) and (PM10-2.5 99th percentile) / (PM10 ee concentration-equivalent DV). Multiply stats by 150 to estimate ranges of PM10-2.5 NAAQS levels.

Note:

- Using only data for 'complete' PM10 sites that are also 12-quarter PM10-2.5 sites were used.

98th percentile

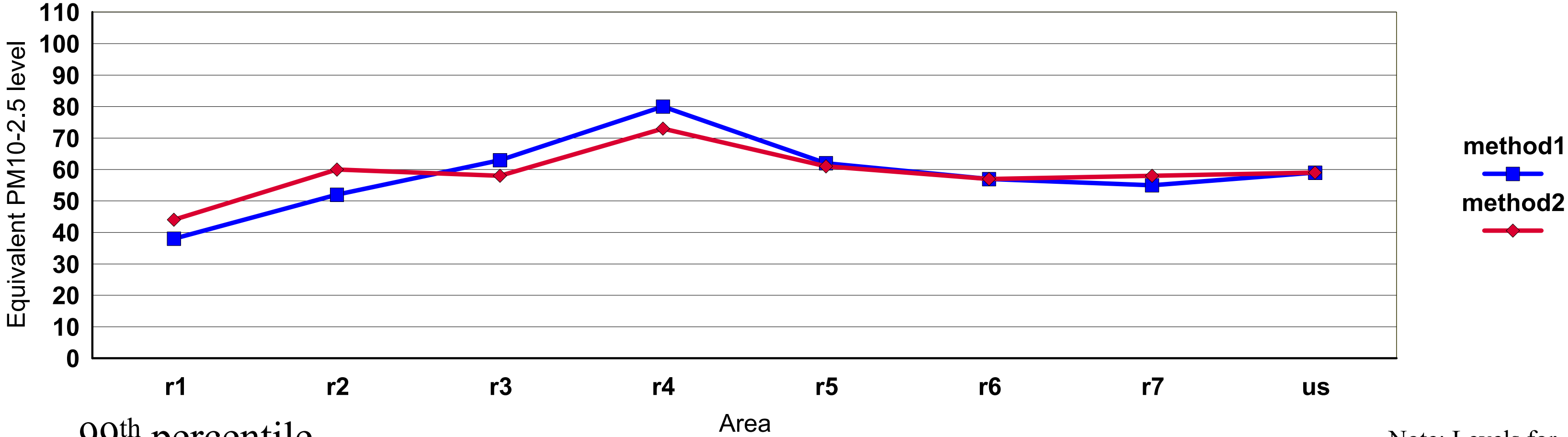
PMREG	PMREGDESC	n	ratio_ p75	ratio_ mean	ratio_ median	ratio_ p25		ratio_ p75 (*150)	ratio_ mean (*150)	ratio_ median (*150)	ratio_ p25 (*150)
1	Northeast	18	0.34	0.30	0.27	0.22		51	**44**	**41**	33
2	Southeast	38	0.46	0.40	0.39	0.32		69	**60**	**58**	48
3	Industrial Midwest	63	0.46	0.39	0.37	0.30		69	**58**	**56**	44
4	Upper Midwest	17	0.57	0.49	0.48	0.42		85	**73**	**72**	63
5	Southwest	14	0.43	0.41	0.40	0.35		64	**61**	**60**	53
6	Northwest	39	0.47	0.38	0.37	0.28		71	**57**	**56**	43
7	Southern California	16	0.46	0.39	0.39	0.33		69	**58**	**59**	50
	East	119	0.44	0.38	0.36	0.29		67	**57**	**55**	43
	West	86	0.48	0.41	0.40	0.33		72	**61**	**60**	50
	U.S.	205	0.46	0.39	0.38	0.31		69	**59**	**57**	46
	Average of 7 regions							68	**59**	**57**	48

99th percentile

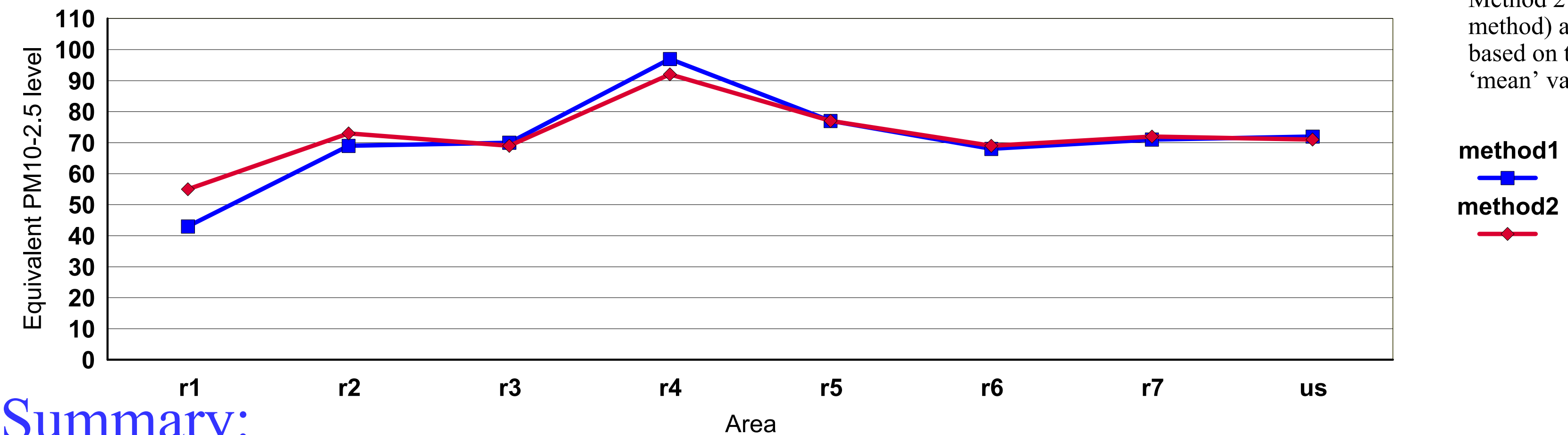
PMREG	PMREGDESC	n	ratio_ p75	ratio_ mean	ratio_ median	ratio_ p25		ratio_ p75 (*150)	ratio_ mean (*150)	ratio_ median (*150)	ratio_ p25 (*150)
1	Northeast	18	0.44	0.37	0.31	0.28		65	**55**	**47**	42
2	Southeast	38	0.59	0.49	0.48	0.41		88	**73**	**72**	62
3	Industrial Midwest	63	0.56	0.46	0.43	0.36		83	**69**	**64**	54
4	Upper Midwest	17	0.68	0.61	0.61	0.53		103	**92**	**91**	79
5	Southwest	14	0.57	0.52	0.50	0.44		86	**77**	**75**	66
6	Northwest	39	0.56	0.46	0.44	0.35		84	**69**	**66**	53
7	Southern California	16	0.56	0.48	0.51	0.39		83	**72**	**77**	58
	East	119	0.55	0.46	0.43	0.35		83	**68**	**65**	53
	West	86	0.59	0.50	0.52	0.40		89	**75**	**78**	59
	U.S.	205	0.57	0.48	0.47	0.38		85	**71**	**71**	57
	Average of 7 regions							85	**73**	**70**	59

Comparison of method 1 and method 2 results:

98th percentile



99th percentile



Note: Levels for Method 2 (ratio method) are based on the 'mean' values.

Summary:

Average of both methods (at U.S. levels) is around 60 µg/m^3 for 98th percentile and 70 µg/m^3 for 99th percentile

Predicted Percentage of Counties w/ Monitors Not Likely to Meet Alternative PM Standards

PM2.5, PM10-2.5, and PM10

Estimated Number/Population/Percentage of Counties Violating PM2.5 Alternative NAAQS, Annual Only & Combination Annual / 98th Percentile

Alternative Standards and Levels (□g/m^3)	Population in monitored counties not likely to meet stated standard and level (1000's)	Percent population (county based) not likely to meet stated standard and level	Number of counties not likely to meet stated standard and level	Percent number of counties not likely to meet stated standard and level*								
	U.S. Total				Northeast	Southeast	Industrial Midwest	Upper Midwest	Southwest	Northwest	Southern CA	Outside Regions**
Annual only												
15	55,855	30%	78	14%	19%	7%	29%	0%	0%	4%	60%	0%
14	76,934	41%	140	25%	28%	21%	51%	0%	5%	5%	67%	0%
13	102,444	55%	224	40%	47%	40%	76%	4%	5%	7%	67%	0%
12	122,454	66%	304	54%	70%	61%	89%	12%	5%	12%	67%	0%
Combined Annual / 24-hour												
15 / 65	55,855	30%	78	14%	19%	7%	29%	0%	0%	4%	60%	0%
15 / 50	58,391	31%	82	15%	19%	7%	29%	0%	0%	9%	60%	0%
15 / 45	60,757	33%	87	15%	19%	7%	29%	0%	10%	12%	60%	0%
15 / 40	65,296	35%	94	17%	20%	7%	30%	0%	10%	19%	60%	0%
15 / 35	89,779	48%	153	27%	45%	8%	47%	0%	10%	36%	60%	7%
15 / 30	133,216	72%	289	51%	78%	29%	87%	6%	19%	51%	80%	13%
15 / 25	159,187	86%	441	78%	98%	77%	99%	51%	43%	65%	80%	13%
14 / 65	76,934	41%	140	25%	28%	21%	51%	0%	5%	5%	67%	0%
14 / 50	79,470	43%	144	26%	28%	21%	51%	0%	5%	10%	67%	0%
14 / 45	81,129	44%	147	26%	28%	21%	51%	0%	10%	12%	67%	0%
14 / 40	84,919	46%	153	27%	28%	21%	52%	0%	10%	19%	67%	0%
14 / 35	101,327	55%	191	34%	45%	22%	58%	0%	10%	36%	67%	7%
14 / 30	134,420	72%	296	53%	78%	33%	88%	6%	19%	51%	80%	13%
15 / 25	159,187	86%	441	78%	98%	77%	99%	51%	43%	65%	80%	13%
13 / 65	102,444	55%	224	40%	47%	40%	76%	4%	5%	7%	67%	0%
13 / 50	103,759	56%	226	40%	47%	40%	76%	4%	5%	10%	67%	0%
13 / 45	105,418	57%	229	41%	47%	40%	76%	4%	10%	12%	67%	0%
13 / 40	108,257	58%	234	42%	47%	40%	76%	4%	10%	19%	67%	0%
13 / 35	115,814	62%	255	45%	53%	40%	77%	4%	10%	36%	67%	7%
13 / 30	137,807	74%	318	57%	78%	43%	90%	8%	19%	51%	80%	13%
13 / 25	159,187	86%	441	78%	98%	77%	99%	51%	43%	65%	80%	13%
12 / 65	122,454	66%	304	54%	70%	61%	89%	12%	5%	12%	67%	0%
12 / 50	122,454	66%	304	54%	70%	61%	89%	12%	5%	12%	67%	0%
12 / 45	123,910	67%	306	54%	70%	61%	89%	12%	10%	14%	67%	0%
12 / 40	126,750	68%	311	55%	70%	61%	89%	12%	10%	20%	67%	0%
12 / 35	132,384	71%	325	58%	70%	61%	89%	12%	10%	36%	67%	7%
12 / 30	144,722	78%	362	64%	84%	62%	94%	14%	19%	51%	80%	13%
12 / 25	159,243	86%	442	79%	98%	78%	99%	51%	43%	65%	80%	13%
Total number of monitored counties (w/ data) ---->				*562*	*83*	*168*	*130*	*49*	*21*	*81*	*15*	*15*
Total population of monitored counties (1000's) ---->				*185,780*	*38,730*	*43,574*	*39,000*	*7,793*	*8,617*	*22,948*	*22,467*	*2,652*

* Based on 2001-2003 data for sites with at least 11 samples per quarter for all 12 quarters. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Estimated Number/Population/Percentage of Counties Violating PM2.5 Alternative NAAQS, Annual Only & Combination Annual / 99th Percentile

Alternative Standards and Levels ($\mu g/m^3$)	Population in monitored counties not likely to meet stated standard and level (1000's)	Percent population (county based) not likely to meet stated standard and level	Number of counties not likely to meet stated standard and level	Percent number of counties not likely to meet stated standard and level*								
	U.S. Total				Northeast	Southeast	Industrial Midwest	Upper Midwest	Southwest	Northwest	Southern CA	Outside Regions**
Annual only												
15	55,855	30%	78	14%	19%	7%	29%	0%	0%	4%	60%	0%
14	76,934	41%	140	25%	28%	21%	51%	0%	5%	5%	67%	0%
13	102,444	55%	224	40%	47%	40%	76%	4%	5%	7%	67%	0%
12	122,454	66%	304	54%	70%	61%	89%	12%	5%	12%	67%	0%
Combined Annual / 24-hour												
15 / 65	55,946	30%	79	14%	19%	7%	29%	0%	0%	5%	60%	0%
15 / 50	61,520	33%	89	16%	19%	7%	29%	0%	10%	15%	60%	0%
15 / 45	65,834	35%	101	18%	24%	7%	32%	0%	10%	21%	60%	0%
15 / 40	86,303	46%	150	27%	47%	9%	42%	0%	10%	36%	67%	7%
15 / 35	126,468	68%	247	44%	72%	17%	77%	0%	19%	51%	80%	13%
15 / 30	151,550	82%	383	68%	96%	54%	97%	35%	38%	59%	80%	13%
15 / 25	165,619	89%	475	85%	100%	86%	99%	69%	48%	73%	87%	13%
14 / 65	77,025	41%	141	25%	28%	21%	51%	0%	5%	6%	67%	0%
14 / 50	81,892	44%	149	27%	28%	21%	51%	0%	10%	15%	67%	0%
14 / 45	84,236	45%	157	28%	30%	21%	52%	0%	10%	21%	67%	0%
14 / 40	99,235	53%	195	35%	48%	23%	57%	0%	10%	36%	73%	7%
14 / 35	129,387	70%	266	47%	72%	27%	78%	0%	19%	51%	80%	13%
14 / 30	151,550	82%	383	68%	96%	54%	97%	35%	38%	59%	80%	13%
15 / 25	165,619	89%	475	85%	100%	86%	99%	69%	48%	73%	87%	13%
13 / 65	102,535	55%	225	40%	47%	40%	76%	4%	5%	9%	67%	0%
13 / 50	106,181	57%	231	41%	47%	40%	76%	4%	10%	15%	67%	0%
13 / 45	108,360	58%	238	42%	49%	40%	76%	4%	10%	21%	67%	0%
13 / 40	116,019	62%	262	47%	59%	40%	77%	4%	10%	36%	73%	7%
13 / 35	135,204	73%	302	54%	75%	40%	85%	4%	19%	51%	80%	13%
13 / 30	152,684	82%	391	70%	96%	58%	97%	35%	38%	59%	80%	13%
13 / 25	165,619	89%	475	85%	100%	86%	99%	69%	48%	73%	87%	13%
12 / 65	122,454	66%	304	54%	70%	61%	89%	12%	5%	12%	67%	0%
12 / 50	124,673	67%	308	55%	70%	61%	89%	12%	10%	16%	67%	0%
12 / 45	126,634	68%	314	56%	71%	61%	89%	12%	10%	22%	67%	0%
12 / 40	132,537	71%	331	59%	75%	62%	89%	12%	10%	36%	73%	7%
12 / 35	143,294	77%	354	63%	80%	62%	92%	12%	19%	51%	80%	13%
12 / 30	154,844	83%	409	73%	96%	68%	98%	35%	38%	59%	80%	13%
12 / 25	165,619	89%	475	85%	100%	86%	99%	69%	48%	73%	87%	13%
Total number of monitored counties (w/ data) ---->				*562*	*83*	*168*	*130*	*49*	*21*	*81*	*15*	*15*
Total population of monitored counties (1000's) ---->				*185,780*	*38,730*	*43,574*	*39,000*	*7,793*	*8,617*	*22,948*	*22,467*	*2,652*

* Based on 2001-2003 data for sites with at least 11 samples per quarter for all 12 quarters. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Note: 'Annual only' data same as preceding slide

Estimated Number/Population/Percentage of Counties Violating PM10-2.5 Alternative NAAQS Levels, 98th Percentile *Encompassing All Sites in 'Extended' Database*

Alternative Levels	Percent of counties, total and by region, (and total percent population) not likely to meet alternative 24-hour (98th percentile form) $PM_{10-2.5}$ standards*								
	Total Counties (population)	Northeast	Southeast	Industrial Midwest	Upper Midwest	Southwest	Northwest	Southern CA	Outside Regions**
Number of counties with monitors (Population, in thousands)	*382 (150,595)*	*57*	*82*	*73*	*33*	*20*	*88*	*15*	*14*
100	3 (4)	0	1	0	3	16	1	25	0
95	3 (4)	0	1	0	3	16	1	25	0
90	3 (4)	0	1	0	3	21	1	25	0
85	3 (5)	0	1	0	3	26	2	25	0
80	5 (7)	0	1	1	9	37	3	31	7
75	8 (8)	2	2	3	9	37	8	31	14
70	9 (15)	2	4	4	12	37	8	44	14
65	11 (17)	2	4	7	15	37	10	44	29
60	14 (22)	2	4	14	18	47	13	44	43
55	15 (23)	5	5	14	21	47	14	44	43
50	19 (33)	9	6	15	30	53	16	56	57
45	25 (41)	12	15	16	42	63	22	69	64
40	33 (47)	14	20	21	55	68	38	69	79
35	39 (52)	19	20	27	64	74	50	75	86
30	51 (63)	26	30	41	70	79	67	88	86
25	61 (70)	37	44	56	85	84	72	94	86

* Based on 2001-2003 data for sites with 4, 8, or 12 consecutive quarters with at least 11 samples per quarter. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Estimated Number/Population/Percentage of Counties Violating PM10-2.5 Alternative NAAQS Levels, 99th Percentile *Encompassing All Sites in 'Extended' Database*

Alternative Levels	**Percent of counties, total and by region, (and total percent population) not likely to meet alternative 24-hour (99th percentile form) $PM_{10-2.5}$ standards+B45***								
	Total Counties (population)	**Northeast**	**Southeast**	**Industrial Midwest**	**Upper Midwest**	**Southwest**	**Northwest**	**Southern CA**	**Outside Regions****
Number of counties with monitors (Population, in thousands)	***382 (150,595)***	***57***	***82***	***73***	***33***	***20***	***88***	***15***	***14***
100	4 (8)	4	2	0	3	21	3	25	7
95	6 (9)	4	2	3	3	32	6	25	7
90	7 (10)	5	2	4	6	37	6	25	7
85	10 (13)	5	2	5	12	42	11	25	14
80	11 (17)	5	4	7	15	47	11	31	21
75	13 (20)	5	5	10	15	47	11	44	21
70	14 (22)	5	6	10	21	47	13	44	21
65	18 (28)	12	9	14	33	53	14	44	43
60	23 (34)	14	10	15	36	58	22	63	50
55	27 (43)	14	13	18	48	63	26	69	64
50	33 (48)	18	17	23	52	68	36	81	64
45	38 (52)	18	24	26	55	79	48	81	71
40	45 (56)	23	27	32	70	79	55	88	86
35	54 (67)	32	37	44	82	79	66	94	86
30	63 (72)	42	48	55	85	84	74	94	86
25	76 (83)	58	66	68	94	95	88	100	86

* Based on 2001-2003 data for sites with 4, 8, or 12 consecutive quarters with at least 11 samples per quarter. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Estimated Number/Population/Percentage of Counties Violating PM10-2.5 Alternative NAAQS Levels, 98th Percentile *Encompassing Only 'Urban' Sites in 'Extended' Database*

Alternative Levels	Percent of counties, total and by region, (and total percent population) not likely to meet alternative 24-hour (98th percentile form) $PM_{10-2.5}$ standards*								
	Total Counties (population)	**Northeast**	**Southeast**	**Industrial Midwest**	**Upper Midwest**	**Southwest**	**Northwest**	**Southern CA**	**Outside Regions****
Number of counties with monitors (Population, in thousands)	***259 (141,859)***	***44***	***60***	***57***	***18***	***13***	***45***	***15***	***7***
100	3 (5)	0	2	2	0	23	2	20	0
95	3 (5)	0	2	2	0	23	2	20	0
90	3 (5)	0	2	2	0	23	2	20	0
85	3 (5)	0	2	2	0	31	2	20	0
80	5 (6)	0	2	2	0	46	2	20	14
75	7 (8)	2	3	4	0	46	4	20	29
70	7 (9)	2	3	4	0	46	4	33	29
65	9 (11)	2	3	5	6	46	9	40	29
60	12 (16)	2	5	7	6	62	13	40	43
55	13 (18)	5	5	7	17	62	13	40	43
50	16 (27)	5	7	9	22	62	16	53	57
45	23 (40)	9	13	14	50	69	20	67	57
40	30 (46)	11	18	21	50	77	36	67	71
35	37 (52)	18	18	28	72	77	51	73	86
30	48 (61)	23	32	42	78	77	64	87	86
25	58 (68)	36	42	58	89	77	67	93	86

* Based on 2001-2003 data for sites with 4, 8, or 12 consecutive quarters with at least 11 samples per quarter. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Estimated Number/Population/Percentage of Counties Violating PM10-2.5 Alternative NAAQS Levels, 99th Percentile *Encompassing Only 'Urban' Sites in 'Extended' Database*

Alternative Levels	Percent of counties, total and by region, (and total percent population) not likely to meet alternative 24-hour (99th percentile form) $PM_{10-2.5}$ standards*								
	Total Counties (population)	**Northeast**	**Southeast**	**Industrial Midwest**	**Upper Midwest**	**Southwest**	**Northwest**	**Southern CA**	**Outside Regions****
Number of counties with monitors (Population, in thousands)	***259 (141,859)***	***44***	***60***	***57***	***18***	***13***	***45***	***15***	***7***
100	4 (6)	5	2	0	0	23	2	20	14
95	5 (7)	5	2	2	0	31	2	20	14
90	6 (8)	7	2	4	0	38	2	20	14
85	8 (10)	7	2	4	6	46	11	20	14
80	10 (11)	7	3	4	6	54	11	20	29
75	12 (14)	7	5	5	6	54	11	40	29
70	13 (15)	7	7	5	17	54	13	40	29
65	16 (19)	9	8	7	33	54	16	40	43
60	19 (27)	11	10	7	39	62	16	60	43
55	22 (40)	11	13	11	44	62	18	67	57
50	29 (46)	16	17	19	56	62	33	73	57
45	36 (52)	16	23	26	56	77	44	80	71
40	42 (56)	23	27	32	72	77	51	87	86
35	53 (65)	32	33	47	94	77	62	93	86
30	61 (71)	45	47	54	94	77	71	93	86
25	75 (82)	57	68	70	100	92	82	100	86

* Based on 2001-2003 data for sites with 4, 8, or 12 consecutive quarters with at least 11 samples per quarter. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Number/Population/Percentage of Counties Violating PM10 NAAQS

Database	Percent of counties, total and by region, (and total percent population) not meeting the current PM_{10} standards								
	Total Counties (population)	**Northeast**	**Southeast**	**Industrial Midwest**	**Upper Midwest**	**Southwest**	**Northwest**	**Southern CA**	**Outside Regions****
All PM_{10} sites *: [Number of counties with monitors (Population, in thousands)]**	***585 (170,118)***	***84***	***120***	***115***	***52***	***33***	***142***	***18***	***21***
Percent violating	8 (13)	0	3	3	6	27	10	61	10
PM_{10} sites that meet 'urban' criteria *: [Number of counties with monitors (Population, in thousands)]*	***309 (153,546)***	***59***	***70***	***67***	***21***	***17***	***50***	***15***	***10***
Percent violating	6 (12)	0	1	3	0	29	4	53	10
Urban PM_{10} sites, also $PM_{10-2.5}$ *[Number of counties with monitors (Population, in thousands)]*	***259 (141,859)***	***44***	***60***	***57***	***18***	***13***	***45***	***15***	***7***
Percent violating	7 (11)	0	2	4	0	38	2	47	14

* Based on official EPA design values for 2001-2003; see http://epa.gov/airtrends/values.html.

** "Outside Regions" includes Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

Summary of $PM_{2.5}$ FRM Data Analyses in 49 Metropolitan Areas, 2001-2003

Area *	N Sites	3-year Average Annual Mean						24-Hour P_{90} (μg/m^3) **		
		Levels (μg/m^3)			Percent Difference		r (Max site versus Min site)	Max Pair	Min Pair	r (Max Pair)
		Area Avg	Max Site	Min Site	Largest diff., any site versus Area Avg	Max site versus Min site				
Albuquerque, NM	4	7.0	10.2	5.0	31%	51%	0.42	10.9	2.6	0.42
Atlanta-Sandy Springs-Gainesville, GA	8	15.9	18.0	14.1	12%	22%	0.71	9.4	3.5	0.71
Bakersfield, CA	5	15.3	21.8	6.7	56%	69%	0.00	44.8	6.0	0.16
Baton Rouge-Pierre Part, LA	5	12.3	13.1	10.8	12%	18%	0.85	7.7	2.4	0.62
Birmingham-Hoover-Cullman, AL	8	14.8	18.0	12.6	18%	30%	0.78	12.7	3.5	0.78
Charlotte-Gastonia-Salisbury, NC-SC	5	14.3	14.9	14.0	4%	6%	0.94	4.1	1.7	0.92
Chicago-Naperville-Michigan City, IL-IN-WI	28	14.7	17.7	11.7	20%	34%	0.77	13.6	2.2	0.73
Cincinnati-Middletown-Wilmington, OH-KY-IN	12	16.0	17.8	14.5	10%	19%	0.95	7.0	2.4	0.95
Cleveland-Akron-Elyria, OH	13	15.5	18.3	13.4	15%	27%	0.87	11.4	3.2	0.87
Dallas-Fort Worth, TX	7	12.8	13.9	11.7	9%	16%	0.92	5.2	2.3	0.92
Denver-Aurora-Boulder, CO	6	8.7	10.8	4.5	48%	58%	0.40	11.4	4.0	0.42
Detroit-Warren-Flint, MI	14	15.2	19.5	12.6	22%	35%	0.85	14.1	3.2	0.85
Eugene-Springfield, OR	4	9.4	13.4	6.6	30%	51%	0.57	19.3	4.8	0.57
Grand Rapids-Muskegon-Holland, MI	4	13.0	13.8	12.3	6%	11%	0.91	5.8	3.2	0.90
Greensboro--Winston-Salem--High Point, NC	4	14.6	15.8	14.0	8%	11%	0.94	5.5	2.5	0.93
Houston-Baytown-Huntsville, TX	6	11.7	14.2	9.6	18%	32%	0.78	8.9	6.2	0.64
Indianapolis-Anderson-Columbus, IN	6	15.3	16.7	13.6	11%	19%	0.93	6.8	2.0	0.93
Kansas City-Overland Park-Kansas City, MO-KS	10	12.0	13.9	10.8	14%	22%	0.76	9.1	1.4	0.76
Knoxville-Sevierville-La Follette, TN	5	15.3	16.7	14.2	8%	15%	0.86	6.2	2.7	0.86
Las Vegas-Paradise-Pahrump, NV	5	7.1	11.0	4.0	44%	64%	0.03	17.6	2.5	-0.03
Lexington-Fayette--Frankfort--Richmond, KY	4	14.4	15.7	13.5	8%	14%	0.86	5.9	3.3	0.86
Little Rock-North Little Rock-Pine Bluff, AR	5	13.0	14.1	11.9	8%	16%	0.79	7.6	5.1	0.78
Los Angeles-Long Beach-Riverside, CA	22	19.0	27.8	9.9	48%	64%	0.50	39.6	5.3	0.50
Louisville-Elizabethtown-Scottsburg, KY-IN	6	15.6	16.9	14.1	10%	17%	0.85	8.2	3.9	0.85
Memphis, TN-MS-AR	6	13.1	14.0	11.7	11%	16%	0.86	6.3	2.2	0.82
Miami-Fort Lauderdale-Miami Beach, FL	6	8.2	9.5	7.4	14%	22%	0.73	5.5	1.7	0.73
Milwaukee-Racine-Waukesha, WI	6	13.1	13.2	12.5	5%	5%	0.96	4.1	2.2	0.93
Minneapolis-St. Paul-St. Cloud, MN-WI	12	10.5	12.0	9.7	13%	19%	0.79	8.0	2.6	0.79
New Orleans-Metairie-Bogalusa, LA	4	11.5	12.2	10.4	10%	15%	0.91	4.0	2.8	0.90
New York-Newark-Bridgeport, NY-NJ-CT-PA	29	13.5	16.4	11.2	18%	32%	0.85	12.5	2.0	0.84
Omaha-Council Bluffs-Fremont, NE-IA	7	10.4	10.7	9.8	6%	8%	0.86	5.2	2.1	0.78
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	14	14.9	16.4	13.8	9%	16%	0.94	7.6	3.1	0.94
Phoenix-Mesa-Scottsdale, AZ	5	9.3	11.4	6.3	32%	45%	0.22	14.0	4.2	0.22
Pittsburgh-New Castle, PA	13	15.8	21.2	13.2	25%	38%	0.75	21.8	3.2	0.69
Portland-Vancouver-Beaverton, OR-WA	6	8.2	9.5	6.1	26%	36%	0.84	9.5	3.0	0.76
Provo-Orem, UT	4	9.8	10.9	8.8	10%	19%	0.88	6.5	3.0	0.92
Raleigh-Durham-Cary, NC	5	13.3	13.9	12.2	8%	12%	0.93	5.7	2.4	0.88
Richmond, VA	5	13.4	14.0	12.8	4%	9%	0.88	5.8	3.2	0.88
Sacramento--Arden-Arcade--Truckee, CA-NV	5	9.9	12.5	7.6	23%	39%	0.37	16.0	6.0	0.21
Salt Lake City-Ogden-Clearfield, UT	7	11.4	14.0	9.0	21%	36%	0.92	11.0	3.8	0.92
San Diego-Carlsbad-San Marcos, CA	5	15.0	15.9	12.8	15%	19%	0.89	10.6	4.6	0.69
San Jose-San Francisco-Oakland, CA	9	10.8	11.8	8.4	22%	29%	0.67	13.5	4.7	0.67
San Juan-Caguas-Fajardo, PR	5	7.2	9.3	5.1	29%	45%	0.71	6.8	1.7	0.71
Seattle-Tacoma-Olympia, WA	10	9.4	11.1	5.3	44%	52%	0.30	19.1	2.9	0.30
St. Louis-St. Charles-Farmington, MO-IL	12	15.0	17.5	14.0	14%	20%	0.82	10.3	2.2	0.76
Virginia Beach-Norfolk-Newport News, VA-NC	5	12.5	13.0	11.9	5%	8%	0.93	4.6	2.7	0.90
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV	20	14.5	16.7	12.2	16%	27%	0.82	9.7	2.6	0.82
Weirton-Steubenville, WV-OH	4	17.1	17.8	16.2	5%	9%	0.87	8.3	6.1	0.86
Wichita-Winfield, KS	4	10.9	11.1	10.2	6%	8%	0.96	2.9	1.3	0.91

* 'Area' is the larger of a Combined Statistical Area (CSA) or a Core Based Statistical Area (CBSA). See http://www.whitehouse.gov/omb/bulletins/fy05/b05-02.html.

** 'P_{90}' is the 90th percentile of the distribution of differences in 24-hour averages between two sites in the same urban area.

Summary of Estimated $PM_{10-2.5}$ Analyses in 21 Metropolitan Areas, 2001-2003

Area *	N Sites	3-year Average Annual Mean						24-Hour P_{90} (µg/m^3) **		
		Levels (µg/m^3)			Percent Difference		r (Max site versus Min site)			
		Area Avg	Max Site	Min Site	Largest diff., any site versus Area Avg	Max site versus Min site		Max Pair	Min Pair	r (Max Pair)
Anchorage, AK	3	14.8	23.7	9.6	38%	59%	0.13	52.3	22.5	0.13
Birmingham-Hoover-Cullman, AL	5	7.0	9.0	5.6	22%	38%	0.76	10.0	3.0	0.55
Cleveland-Akron-Elyria, OH	8	11.6	16.3	5.6	52%	66%	0.55	26.0	8.0	0.64
Denver-Aurora-Boulder, CO	3	15.5	22.1	7.7	50%	65%	0.54	29.3	14.5	0.54
Detroit-Warren-Flint, MI	3	15.3	18.7	8.8	42%	53%	0.60	30.5	25.0	0.32
El Paso, TX	4	23.2	28.3	13.9	40%	51%	0.89	31.0	15.0	0.92
Las Vegas-Paradise-Pahrump, NV	5	23.2	33.3	9.0	61%	73%	0.65	40.0	17.0	0.65
Los Angeles-Long Beach-Riverside, CA	11	21.6	44.5	13.7	51%	69%	0.38	57.5	8.5	0.03
Miami-Fort Lauderdale-Miami Beach, FL	4	10.2	15.3	8.4	33%	45%	0.63	14.0	3.0	0.63
Minneapolis-St. Paul-St. Cloud, MN-WI	3	19.1	23.6	15.5	19%	34%	0.62	23.0	19.5	0.38
New York-Newark-Bridgeport, NY-NJ-CT-PA	5	8.7	22.3	2.9	67%	87%	0.21	35.3	6.5	0.21
Orlando-The Villages, FL	3	9.5	10.2	8.5	11%	17%	0.71	6.0	4.0	0.71
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	3	5.5	6.4	4.3	22%	33%	0.48	10.0	6.0	0.48
Pittsburgh-New Castle, PA	6	6.4	8.5	3.5	45%	59%	0.67	13.0	5.0	0.46
Sacramento--Arden-Arcade--Truckee, CA-NV	3	10.4	12.0	8.2	21%	32%	0.38	17.5	6.5	0.25
Salt Lake City-Ogden-Clearfield, UT	3	17.9	24.1	14.4	26%	40%	0.72	24.0	9.0	0.72
San Jose-San Francisco-Oakland, CA	7	10.8	13.4	7.8	28%	42%	0.69	13.5	4.5	0.53
San Juan-Caguas-Fajardo, PR	3	24.4	30.2	18.0	26%	40%	0.64	22.0	17.0	0.64
Virginia Beach-Norfolk-Newport News, VA-NC	3	4.2	4.5	4.0	7%	11%	0.54	5.0	3.0	0.54
Weirton-Steubenville, WV-OH	4	12.4	13.8	10.7	14%	22%	0.53	15.0	11.5	0.43
Wichita-Winfield, KS	3	11.9	13.7	10.3	13%	25%	0.81	11.0	5.0	0.69

* 'Area' is the larger of a Combined Statistical Area (CSA) or a Core Based Statistical Area (CBSA). See http://www.whitehouse.gov/omb/bulletins/fy05/b05-02.html.

** 'P_{90}' is the 90th percentile of the distribution of differences in 24-hour averages between two sites in the same urban area.

Attachment B

AQS-Based, Hourly PM Characterization Analyses

General / Background:

This attachment describes the characterization analyses of hourly $PM_{2.5}$ data obtained from AQS. It also documents the analyses of hourly $PM_{10-2.5}$ estimates which were derived from the aforementioned $PM_{2.5}$ AQS dataset and a corresponding PM_{10} AQS dataset. Meteorological data from the nearest NWS site was used to convert the PM_{10} data to local temperature and pressure conditions.

Construction of $PM_{2.5}$ database

The database (db) utilized for all hourly $PM_{2.5}$ SP analyses was based on almost *all available hourly AQS $PM_{2.5}$ data*. The following statements document the creation of the db:

- Hourly duration (AQS duration code = '1') data for the time period 2001 to 2003 were polled from AS for parameter $88_{10}1$ ($PM_{2.5}$, local temperature and pressure conditions, LC) on August 24, 2004. [Deleted data with method codes of 740 or 741, per Tim Hanley of Ambient Air Monitoring (AAMG).]
- Data were processed on a monitor basis.
- To be used, a monitor had to meet the completeness goals of at least 75% of hours in a day (18+) at least 75% days in a quarter (68+). The most recent 4, 8, or 12 consecutive quarters that met those goals were utilized. 264 monitors met the completeness criteria: 128 monitors had 4 usable quarters, 72 had 8 usable quarters, and 64 had 12 'complete' quarters. Only data for those monitors, and for the corresponding days with 18+ hours, were kept; data for other monitors and/or days that did not have 18+ samples were discarded.
- SAS code ('raw from AQS.sas') was used to extract the raw data from AQS.

Construction of $PM_{10-2.5}$ database

The db utilized for all hourly $PM_{10-2.5}$ SP analyses was based on the $PM_{2.5}$ db specified above plus corresponding hourly PM_{10} data. A simplistic difference method (PM_{10} - $PM_{2.5}$) was used to generate the $PM_{10-2.5}$ estimates. PM_{10} data were retrieved for both 'local temperature and pressure conditions' (LC) and 'standard temperature and pressure' (STP) conditions. National Weather Service (NWS) data were used to convert the STP data to LC. Since $PM_{2.5}$ and PM_{10} data were then all in LC $\mu g/m^3$ units, resultant estimated $PM_{10-2.5}$ estimates were in the same units. $PM_{10-2.5}$ estimates were generated on site basis. The following statements provide additional detail:

- As noted above, hourly duration data for the time period 2001 to 2003 were polled from AQS for parameter $88_{10}1$ on August 24, 2004. Data for method codes of 740 or 741 were deleted.
- Hourly data for parameters $81_{10}2$ (PM_{10}, STP) and $85_{10}1$ (PM_{10} LC) were also retrieved from AQS on August 24, 2004.

- Raw NWS hourly data for 2001-2003 were obtained from Bill Cox of Air Quality Modeling Group (AQMG) on March 19, 2004. Utilized fields were relative humidity (RH), barometric pressure (BP), and temperature (T).
- PM_{10} STP data were converted to LC by using the corresponding (same date, same hour) met data from the nearest NWS site.
- Multiple site-date-hour measurements (from collocated monitors) of PM_{10} and/or $PM_{2.5}$ were averaged (independently)
- Hourly $PM_{10-2.5}$ was estimated by subtracting the $PM_{2.5}$ concentration from the PM_{10} (LC based) concentration.
- To be used, a site had to meet the completeness goals of at least 75% of hours in a day (18+) at least 75% days in a quarter (68+). The most recent 4, 8, or 12 consecutive quarters that met those goals were utilized. 31 sites met the completeness criteria: 14 sites had 12 usable quarters, 14 sites had 8 complete quarters, and 3 had all 12 quarters complete. Only data for those sites, and for the corresponding days with 18+ hours, were kept; data for other sites or days that did not have 18+ samples were discarded.
- SAS code was used to extract the raw data from AQS ('raw from AQS.sas') and to convert PM_{10} STP data to LC ('calc hourly coarse.sas').

<u>Analysis 1 – Hourly versus 24-hour, $PM_{2.5}$ and $PM_{10-2.5}$</u>

<u>Goals:</u>

- ? To determine how well correlated is the hourly daily maximum with the 24-hr average?
- ? How well do/would daily and annual standards control hourly peaks?
- ? How do the 1-hr distributions compare to the 24-hr distributions

<u>Outputs</u>:

- o Various tables and box-plots were generated for $PM_{2.5}$; Summary statistics were generated; see Output B.1a.
- o Various tables and box-plots were generated for $PM_{10-2.5}$; Summary statistics were generated; see Output B.1b.

<u>Methods</u>:

- 24-hour data were calculated from the hourly data and pseudo DV's (annual and 98th percentile) were constructed from the hourly-based daily averages. This technique was utilized instead of matching to collocated FRM data for 2 reasons: 1) To avoid sampler bias that would have resulted from comparing the continuous to the filter-based FRM measurements, and 2) To maximize the number of observations: FRM instruments may not have been collocated with the continuous monitors and even if they were, the FRM might have only sampled every 3rd day or every 6th day.
- SAS code was used for all of the analyses: 'correlations_maxvmean3.sas', 'correlations_maxvmean3_pmc.sas', 'hourly v 24 pmf.sas', 'hourly v 24 pmc.sas', 'hourly peak to mean pmf.sas', and 'hourly peak to mean pmc.sas'.

Analysis 2 – Diurnal distributions, seasonal plots, and episodic events of hourly measurements of $PM_{2.5}$ and $PM_{10-2.5}$ concentrations, 2001-2003.

Goals:

- ? To characterize and contrast short-term (diurnal) patterns of $PM_{2.5}$ and $PM_{10-2.5}$.
- ? To characterize differences in seasonal diurnal patterns.
- ? To investigate (and contrast) the effect of episodic events on hourly $PM_{2.5}$ and $PM_{10-2.5}$.

Outputs:

- o Diurnal boxplots, representing 4, 8, or 12 quarters of 2001 to 2003, were generated for every hourly $PM_{2.5}$ monitor and for every hourly $PM_{2.5}$ site. Regional aggregation plots were also included. Two 'example' sites were selected from the pools, one to basically represent 'eastern' sites and the other, to generally depict 'western' sites.
- o Seasonal line-plots were created for one of the two selected sites.
- o The effect of an episodic event on $PM_{2.5}$ and $PM_{10-2.5}$ concentrations over a 2-day period were plotted for an additional 'western' location.
- o All plots (except for the universe pools) are shown in output B.2.

Methods:

- The data hourly data were adjusted for daylight savings time (both size cuts). SAS code ('hour_boxplot20012003_daylight_savings.sas', 'seasonal hour avg line plots.sas', and 'elpaso_gso_20012003.sas'), was used to make the adjustment and generate all of the plots.

Analysis 3 – Evaluation of hour-to-hour changes (increases) in $PM_{2.5}$

Goals:

- ? To characterize typical (median) monitor-level hour-to-hour increases in $PM_{2.5}$

Outputs:

- o See SAS output screen capture in Output B.3

Methods:

- Only 'increases' from one hour to the next were evaluated. SAS code ('hour difference distribution.sas') was used for the evaluation.

Hourly vs. 24-hr – PM2.5

- Questions:
  1. How well correlated is the hourly daily max with the 24-hr average?
  2. How well do the daily and annual stds control hourly peaks?
  3. How do the 1-hr distributions compare to the 24-hr distributions
- Analyses details:
  1. Hourly data from AQS. TEOM, BAM, whatever. May or may not be 'adjusted' to be more FRM like.
  2. Only used sites that met completeness criteria of 75% hours in a day; 75% days in a quarter; most recent 4, 8, or 12 consecutive quarters. 264 sites met criteria (64 had 12 Q's, 72 had 8 Q's, and 128 had 4 Q's.) Only used data for those sites… days w/ 75%+
  3. 24-hr data calculated from hourly data.
  4. Pseudo DV's (annual and 98th percentile) constructed from hourly daily averages. … Instead of matching to collocated FRM DV. ….Rationale – Avoid sampler bias, continuous vs. filter

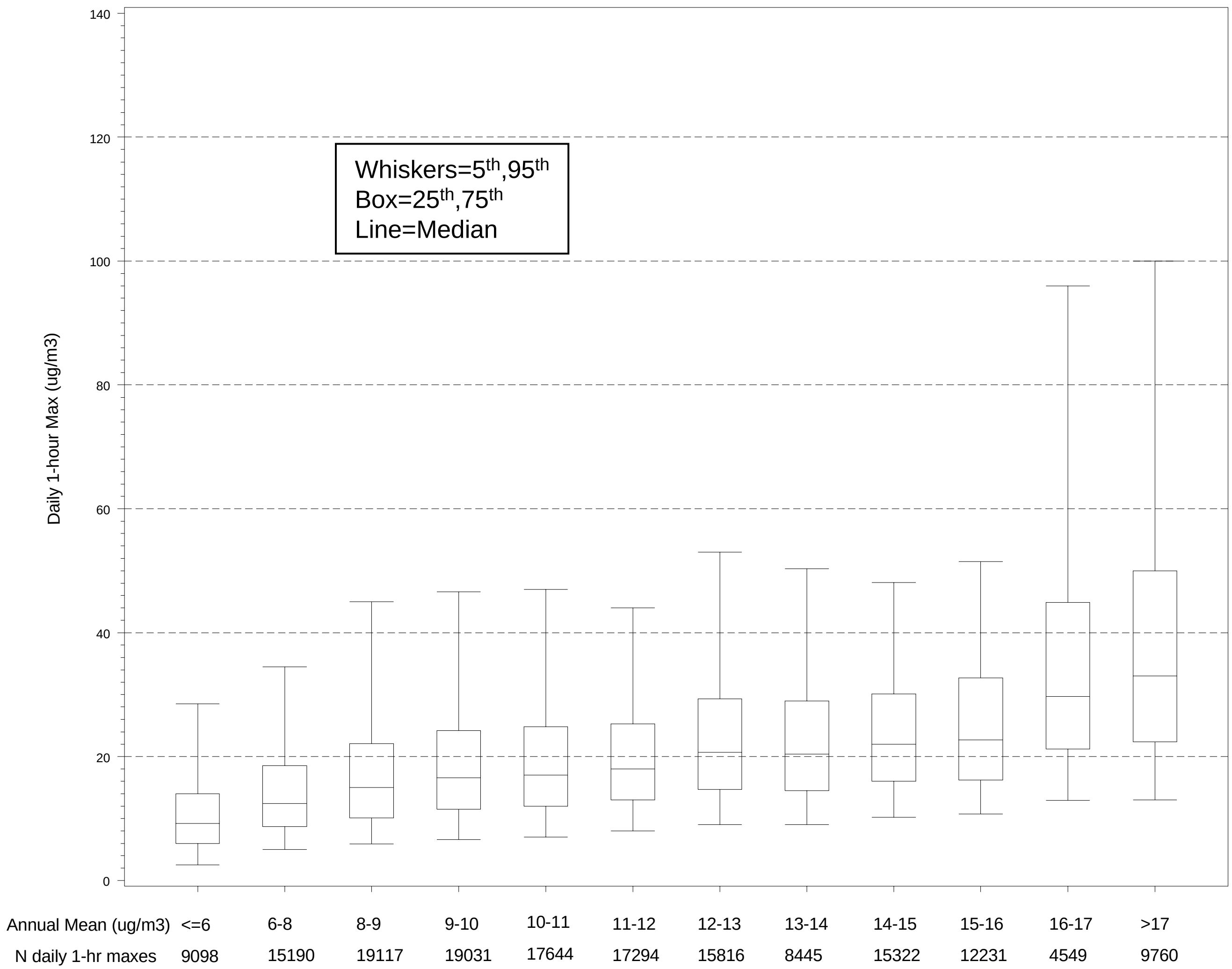
1. How well correlated is the hourly daily max with the 24-hr avg?

		Site Correlation			
HEI region	# sites	mean	median	minimum	maximum
National	264	0.82	0.84	0.53	0.95
Industrial Midwest	41	0.80	0.85	0.55	0.92
Northeast	51	0.89	0.90	0.72	0.95
Northwest	57	0.84	0.85	0.66	0.93
Southeast	65	0.78	0.80	0.53	0.91
Southern California	5	0.82	0.80	0.77	0.92
Southwest	26	0.81	0.83	0.71	0.94
Upper Midwest	17	0.74	0.76	0.56	0.87
Not in PMREG Regi	2	0.80	0.80	0.77	0.83

•Good correlation; consistent across geographic regions.

2a. How well does an annual standard control hourly peaks?

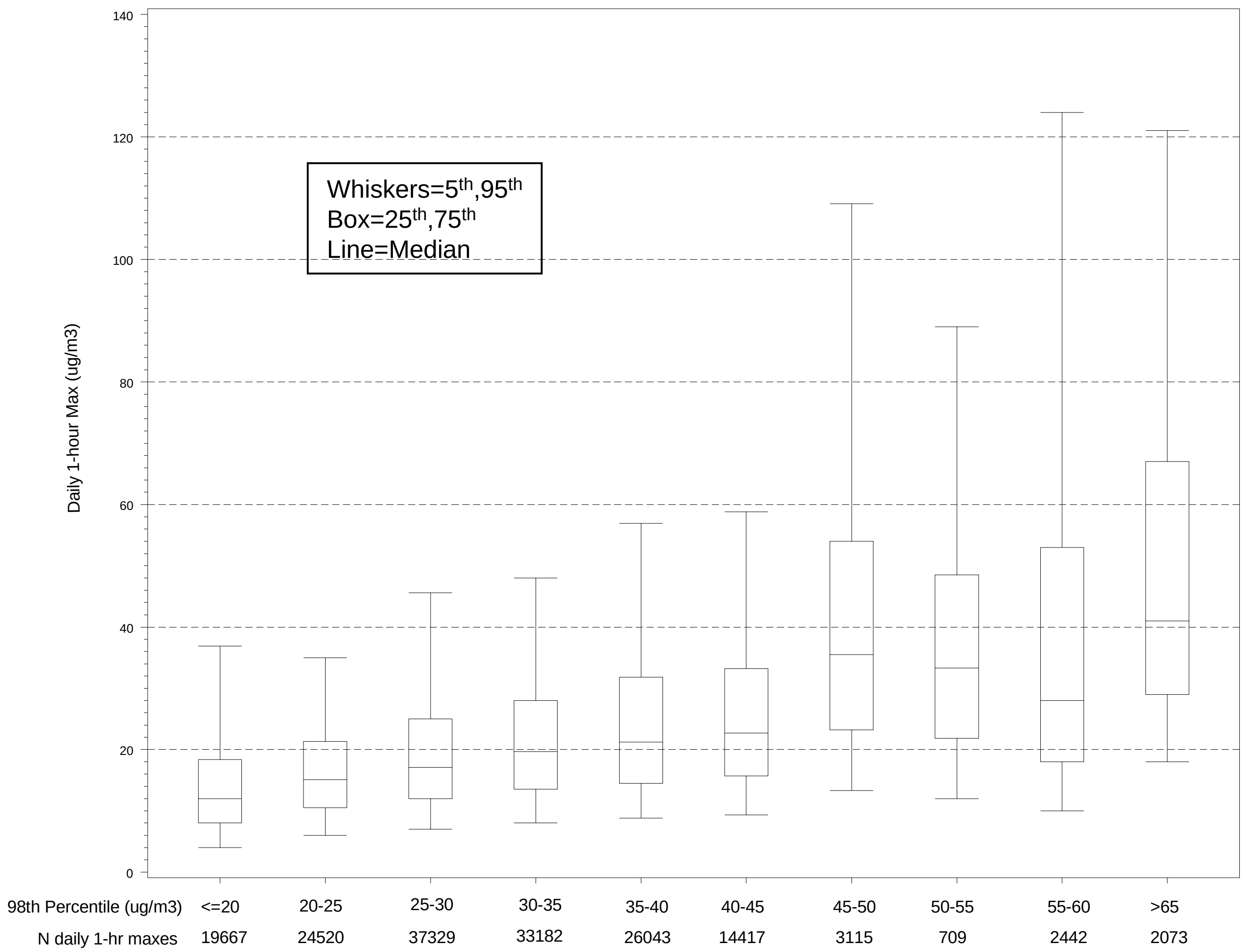
Distribution of daily 1-hour maxes vs. annual mean PM2.5 concentrations, 2001-2003



- More than 95% of daily max 1-hr's are $\leq$ 50 ug/m3 when annual DV $\leq$ 16
- [The 95th percentile (daily max 1-hr) is $\leq$ 50 for *most* of the annual mean intervals $\leq$ 16]

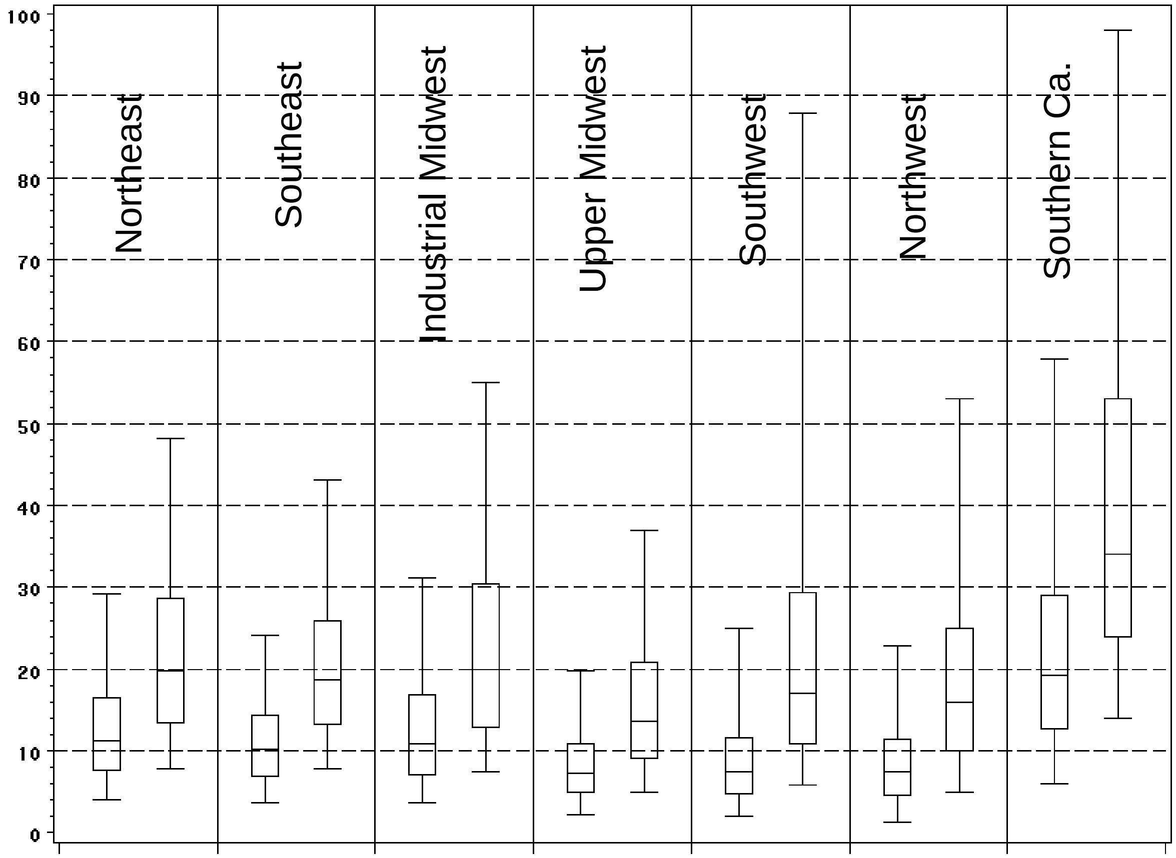
2b. How well does a daily standard control hourly peaks?

Distribution of daily 1-hour max's vs. 98th percentile 24-hour average PM2.5 concentrations, 2001-2003



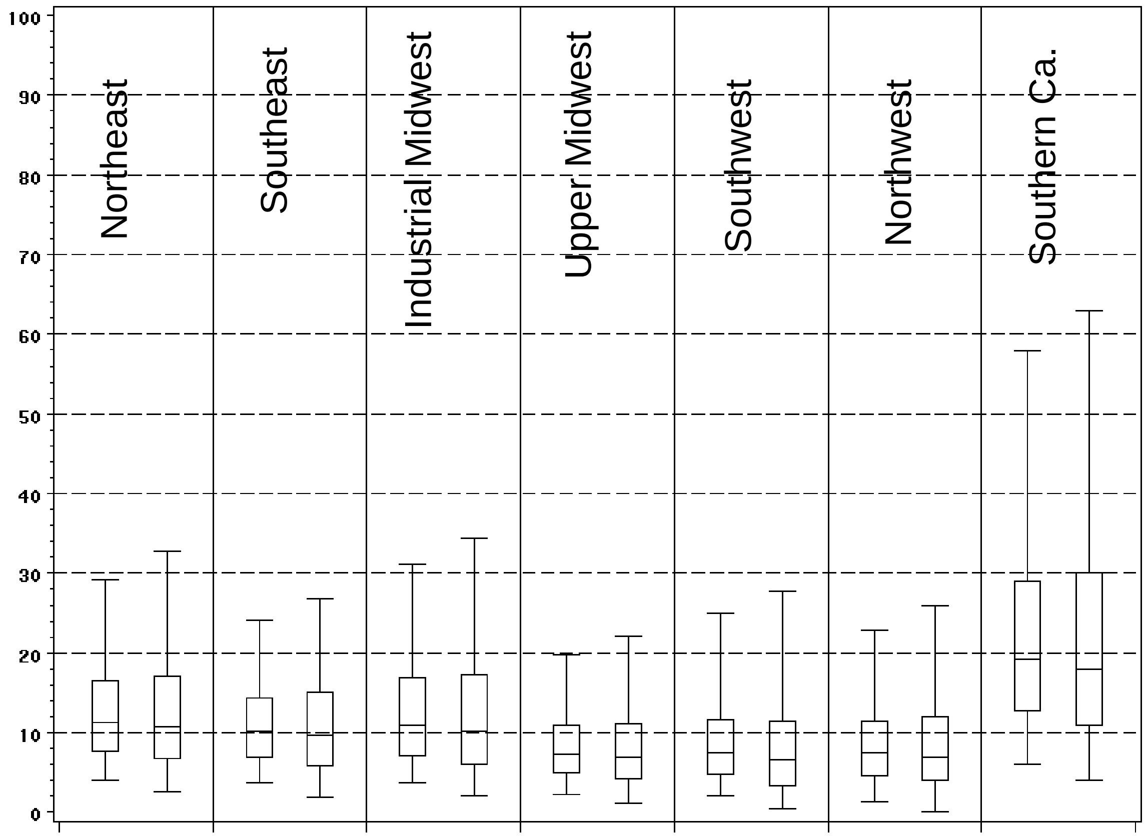
- More than 95% of daily max 1-hr's are ≤ 50 ug/m3 when daily DV (98th percentile) ≤ 45
- [The 95th percentile (daily max 1-hr) is ≤ 50 for *most* of the 98th percentile intervals ≤ 45]
- [The 95th percentile (daily max 1-hr) is ≤ 60 for *all* of the 98th percentile intervals ≤ 45]

3a. How does the max 1-hr distribution compare to the 24-hr distribution?



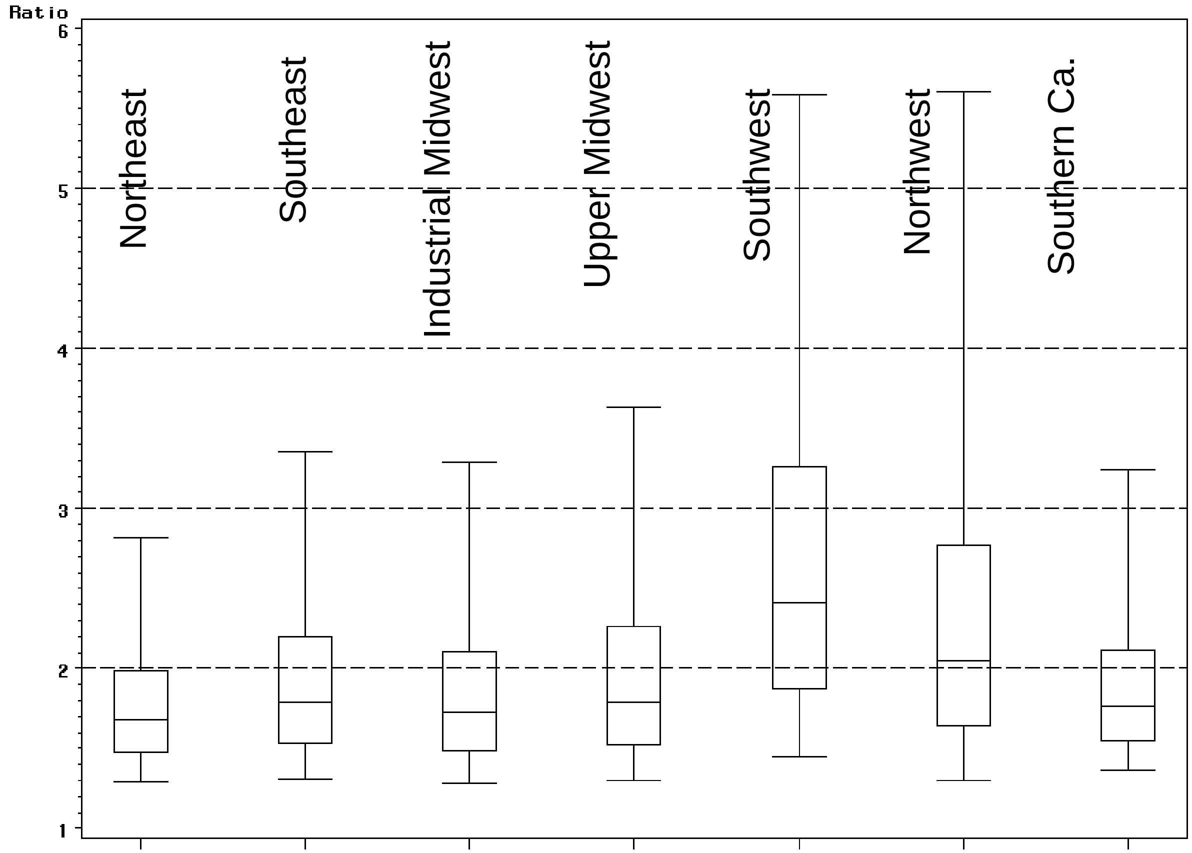
24hr / Max:1hr

3b. How does the 1-hr distribution (all hrs) compare to the 24-hr distribution?



24hr / 1hr

3c. How does the peak-to-mean ratio (max 1hr / 24hr avg) compare by region?



Ratio: max_1hr / 24hr_avg

Hourly vs. 24-hr – PM10-2.5

- Questions:
  1. How well correlated is the hourly daily max with the 24-hr average?
  2. How well would daily and annual standards control hourly peaks?
  3. How do the 1-hr distributions compare to the 24-hr distributions?

- Analyses details:
  1. Hourly data constructed by difference method from (AQS) collocated continuous PM10 and PM2.5.
  2. Only used sites that met completeness criteria of 75% hours in a day; 75% days in a quarter; most recent 4, 8, or 12 consecutive quarters. 31 sites met criteria (3 had 12 Q's, 14 had 8 Q's, and 14 had 4 Q's.) Only used data for those sites… days w/ 75%+
  3. 24-hr data calculated from hourly data.
  4. Constructed psuedo DV's (annual and 98^{th} percentile) from hourly daily averages.

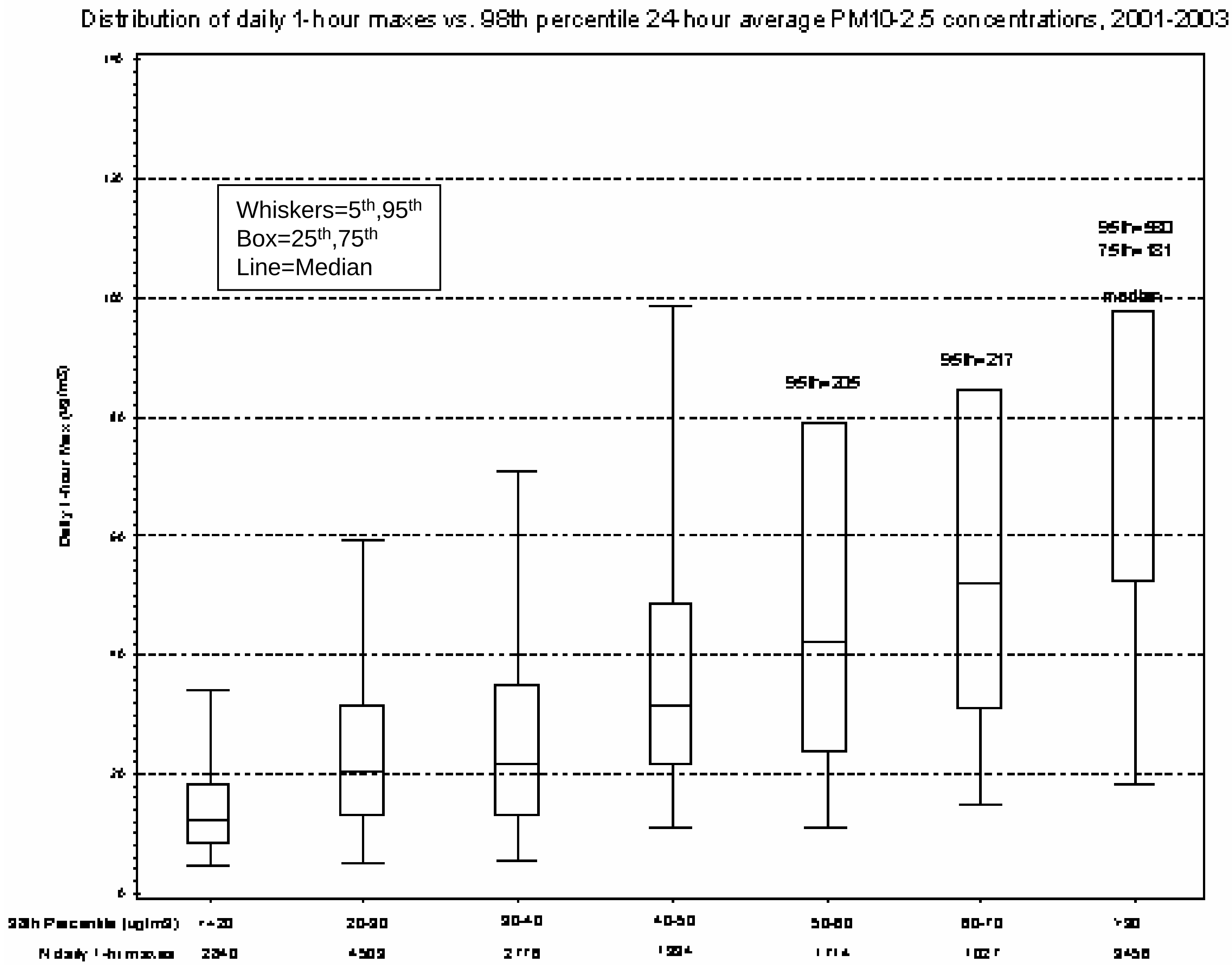
1. How well correlated is the hourly daily max with the 24-hr avg?

HEI Region	# sites	Site Correlation			
		mean	median	minimum	maximum
National	31	0.80	0.81	0.67	0.91
Industrial Midwest	9	0.81	0.83	0.75	0.86
Northeast	3	0.78	0.81	0.67	0.85
Northwest	5	0.77	0.77	0.69	0.88
Southeast	6	0.79	0.78	0.70	0.91
Southwest	5	0.84	0.85	0.78	0.91
Upper Midwest	3	0.83	0.84	0.80	0.86

No data (meeting completeness) for Southern California

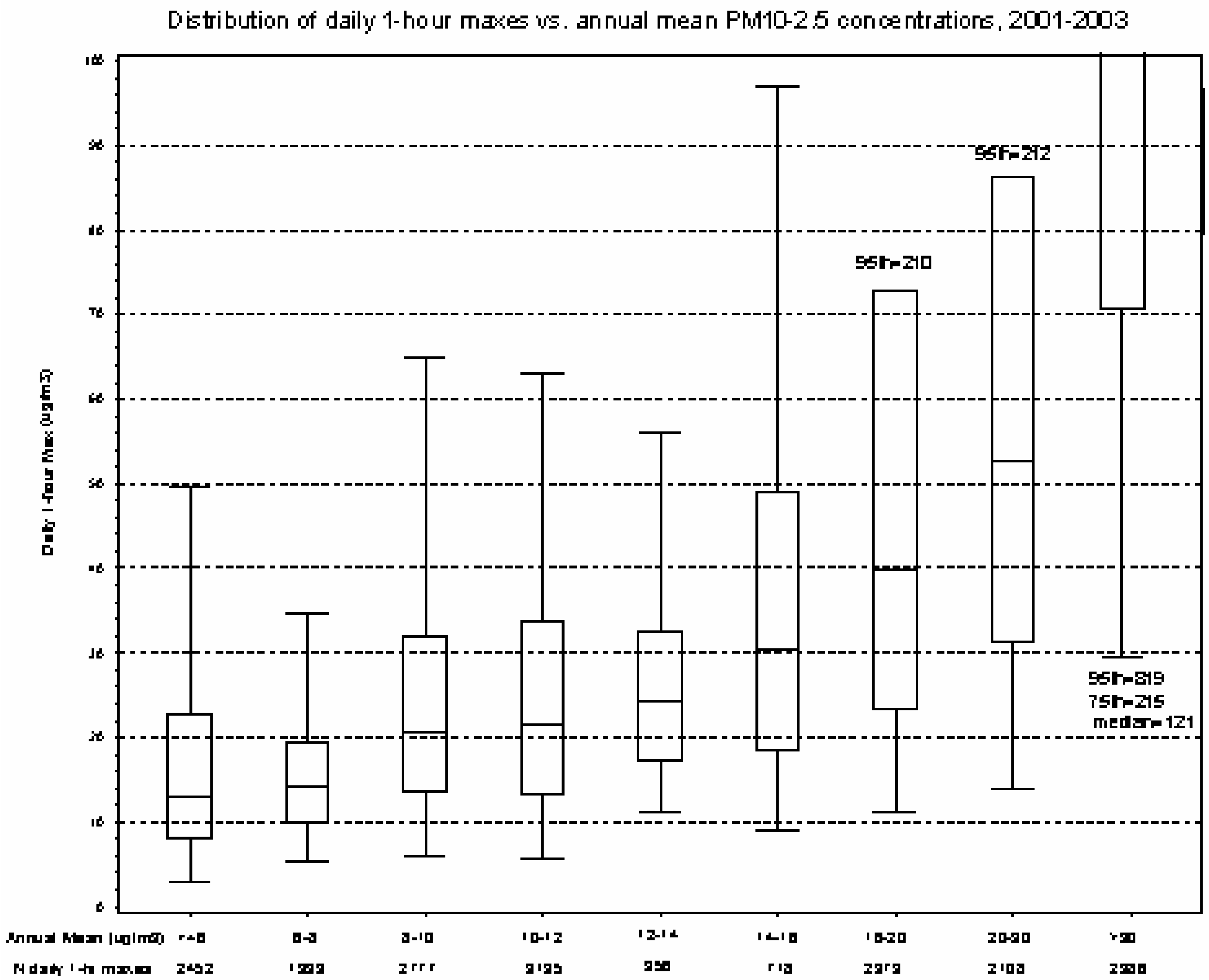
•Good correlation; consistent across geographic regions.

2a. How well would a daily standard control hourly peaks?



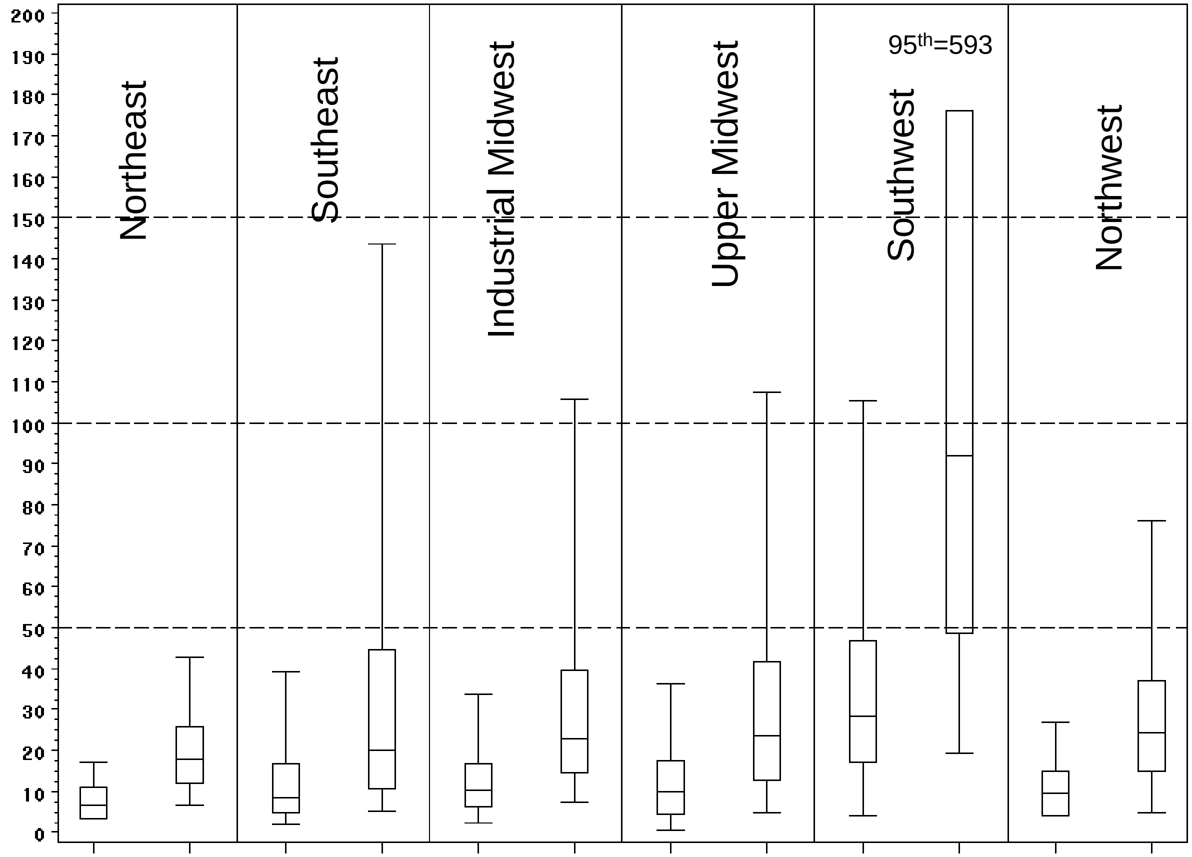
- A daily PM10-2.5 standard would appear not to control hourly peaks unless set on the low end (of the intervals shown here)

2b. How well would an annual standard control hourly peaks?



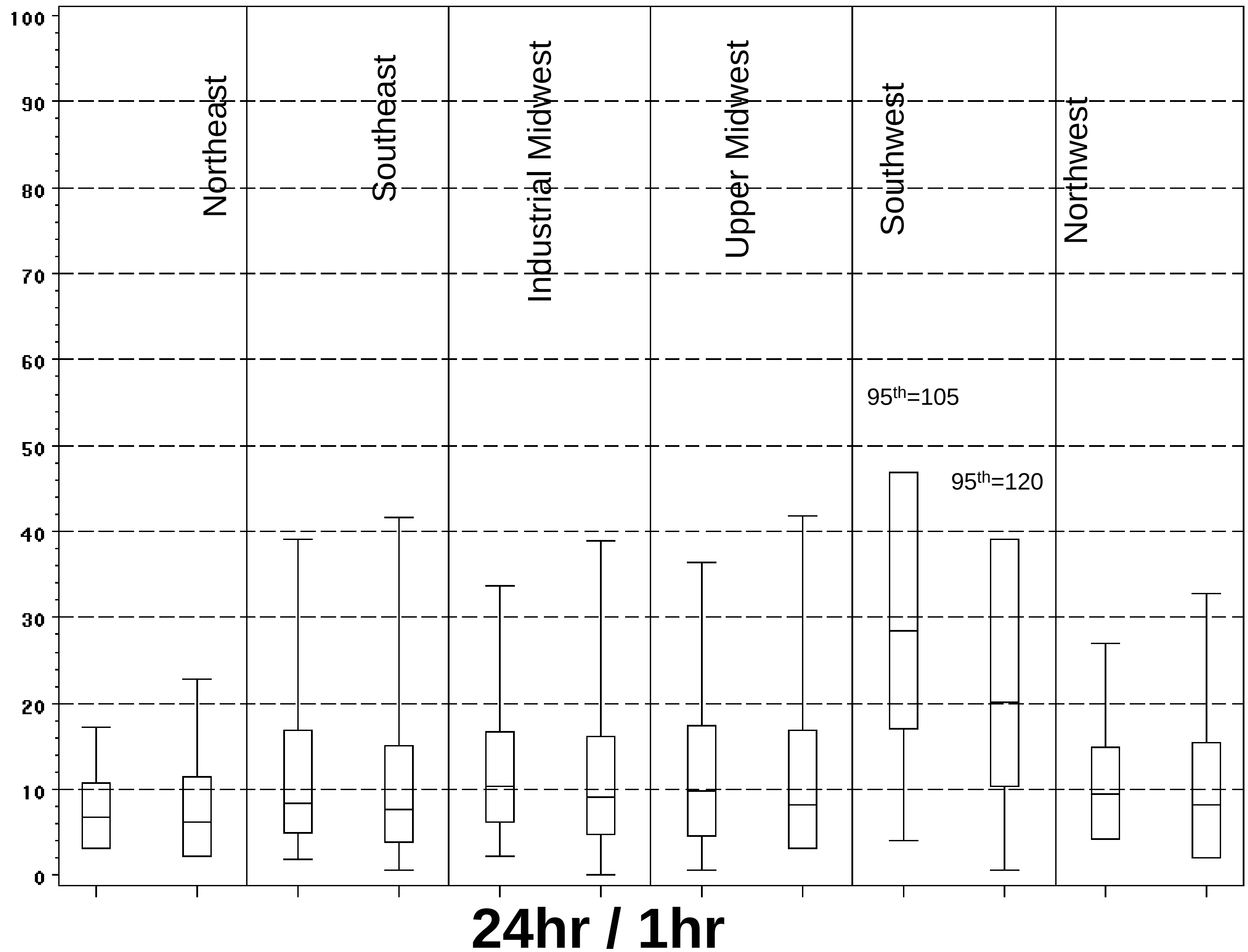
- An annual PM10-2.5 standard would appear not to control hourly peaks unless set on the low end (of the intervals shown here)

3a. How does the max 1-hr distribution compare to the 24-hr distribution?

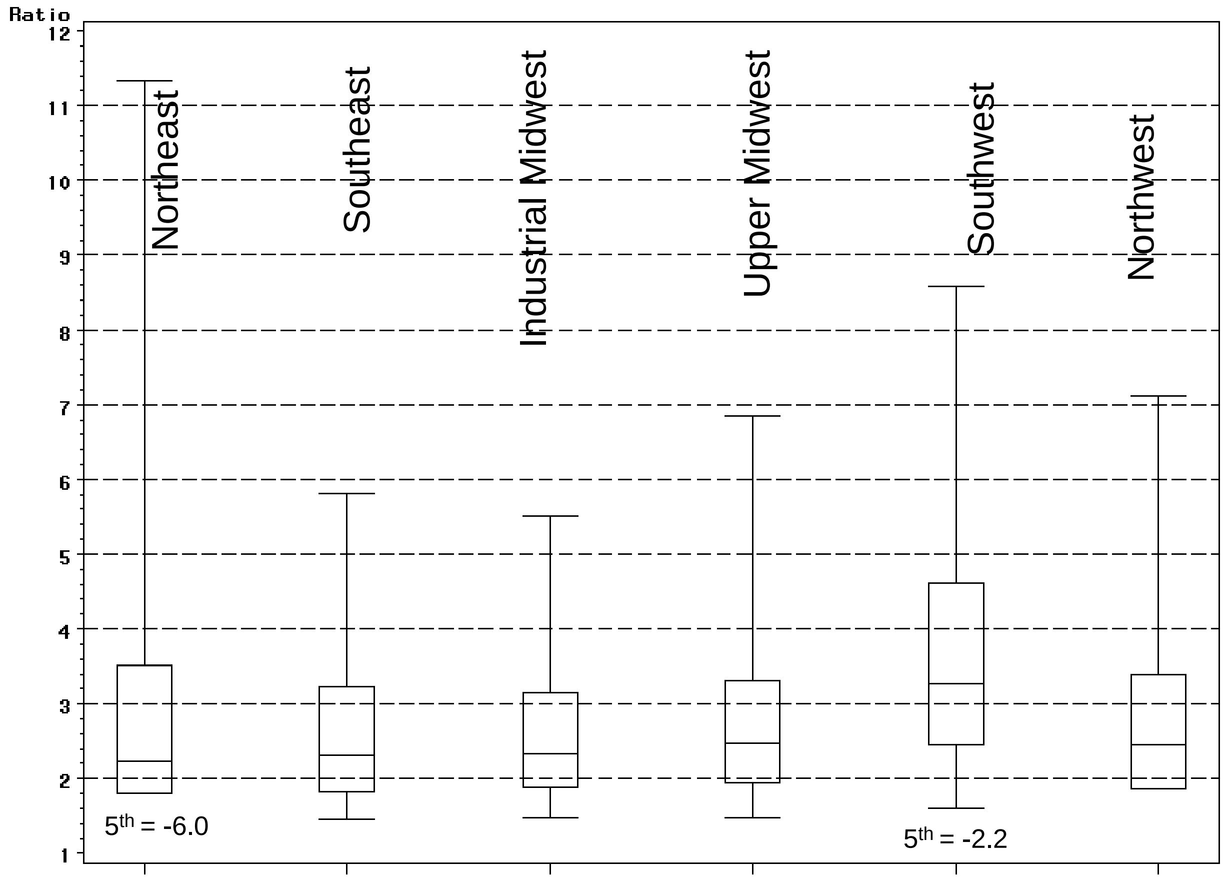


24hr / Max:1hr

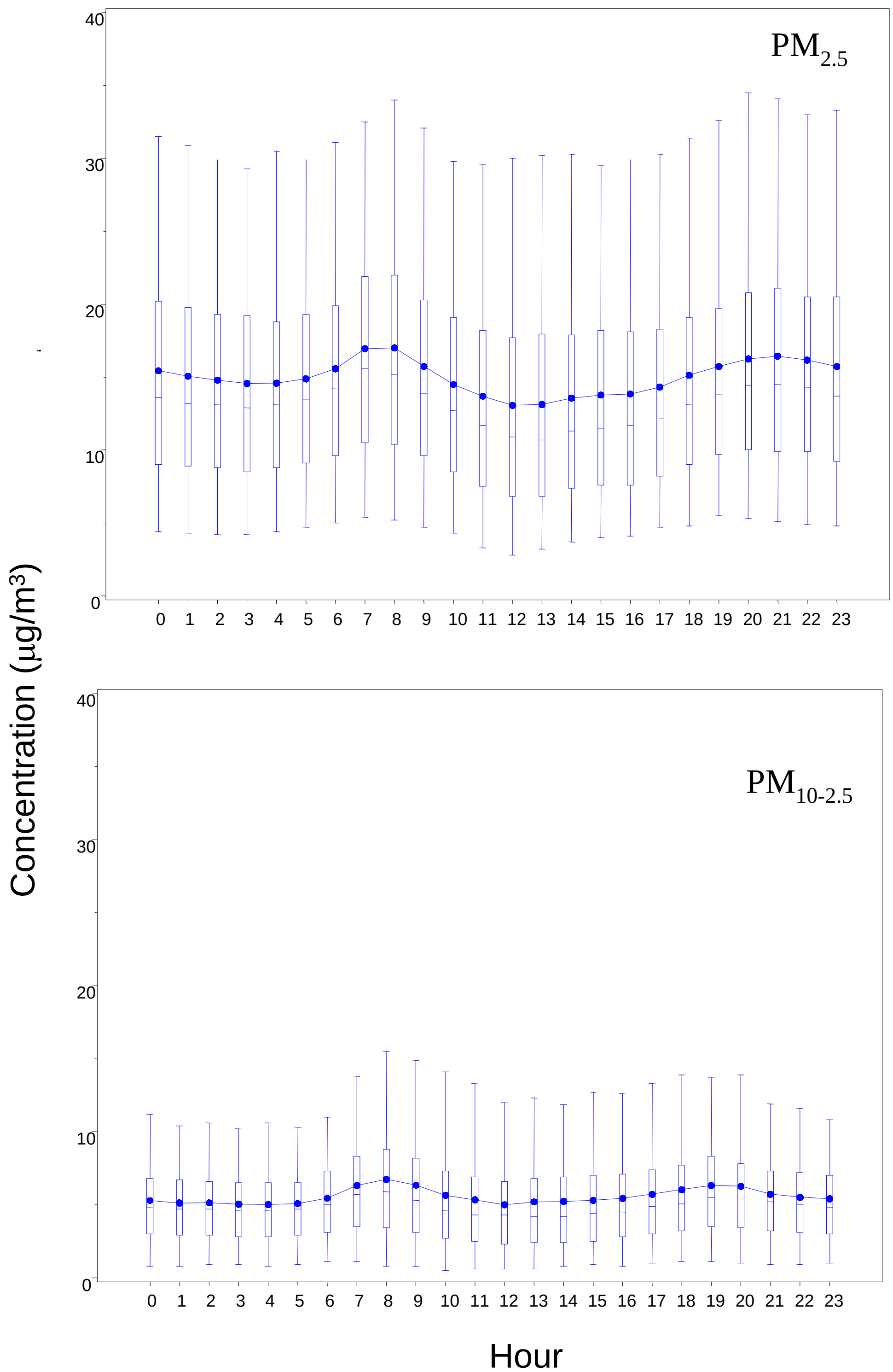
3b. How does the 1-hr distribution (all hrs) compare to the 24-hr distribution?



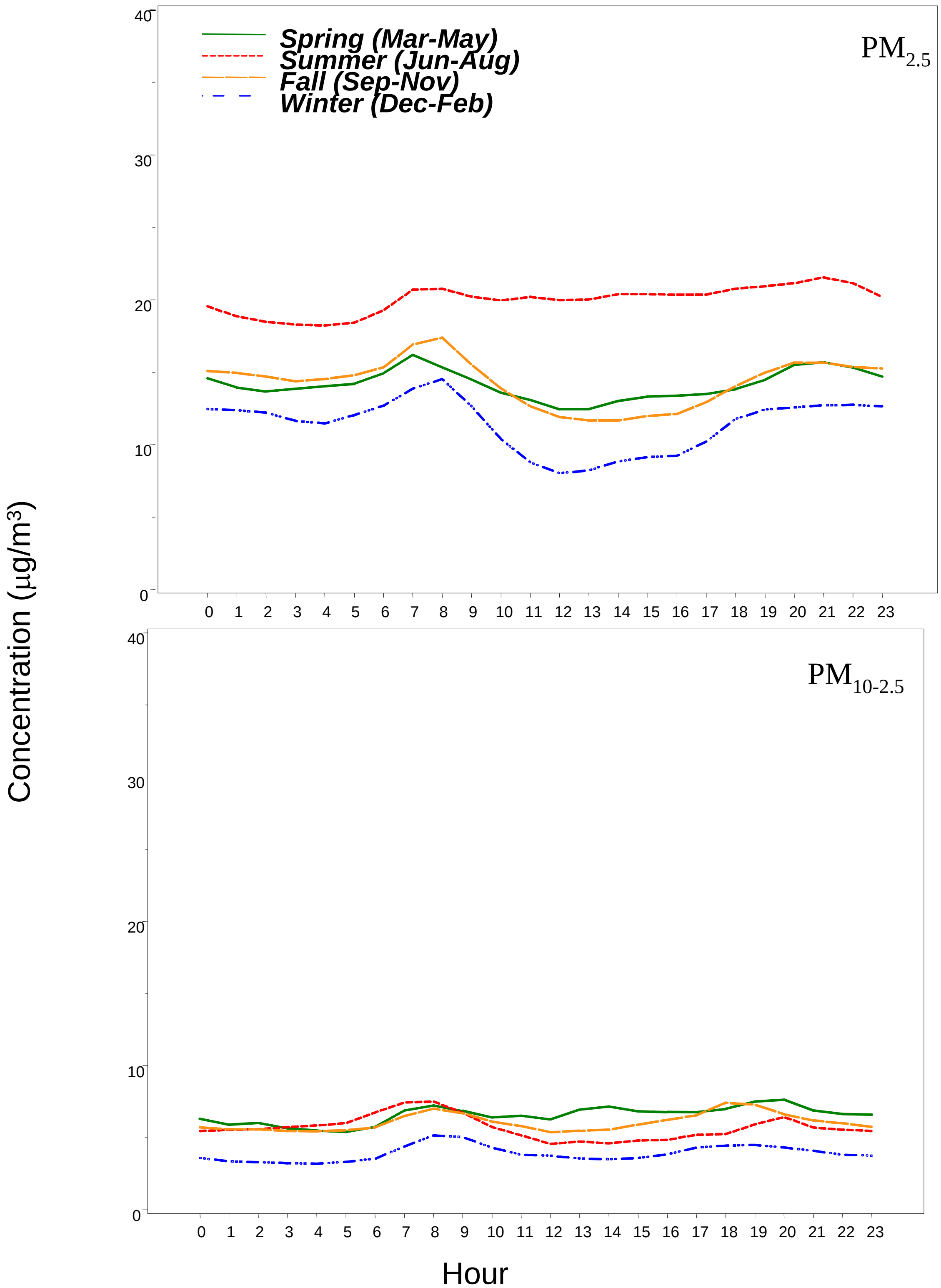
3c. How does the peak-to-mean ratio (max 1hr / 24hr avg) compare by region?



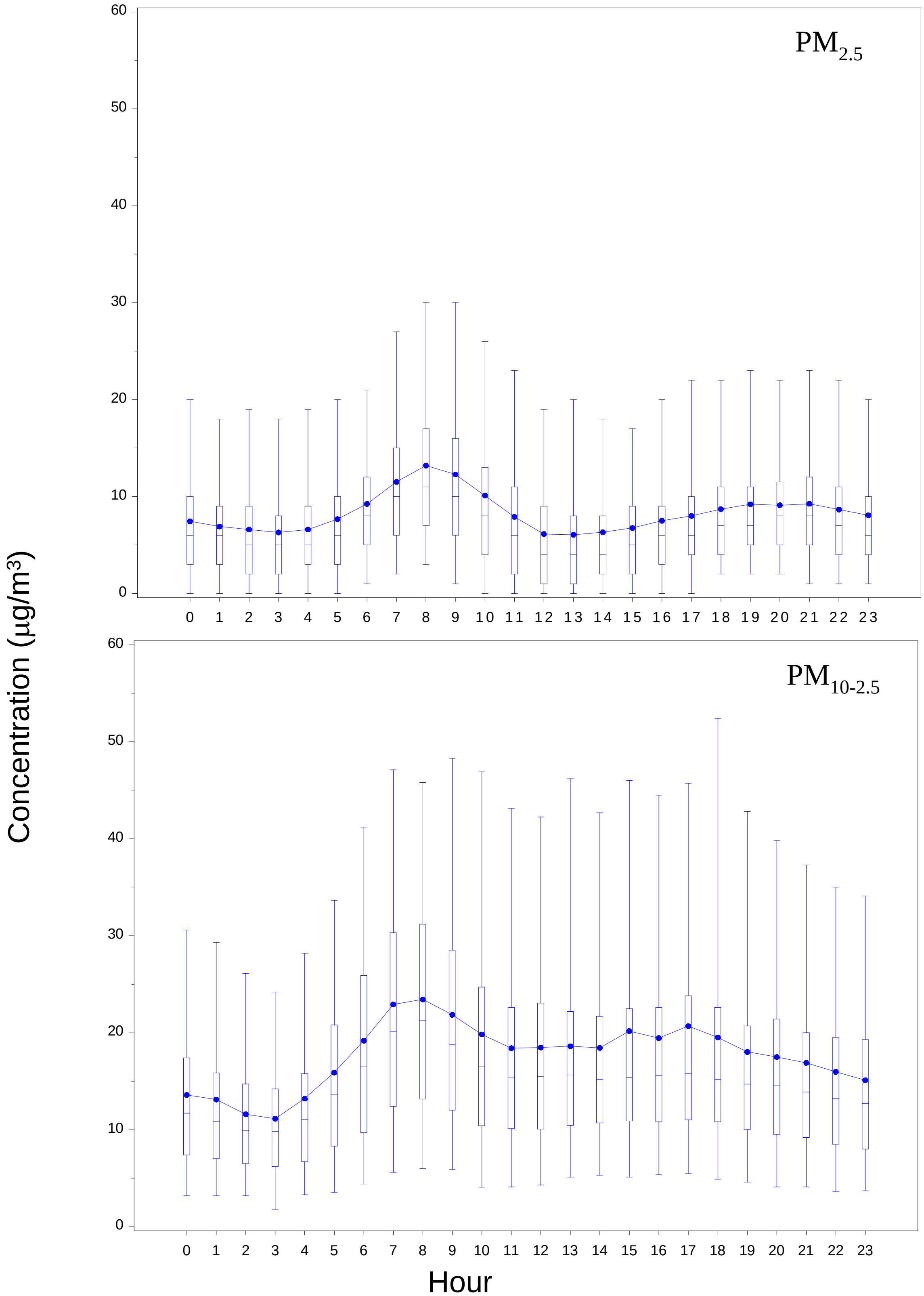
Ratio: max_1hr / 24hr_avg



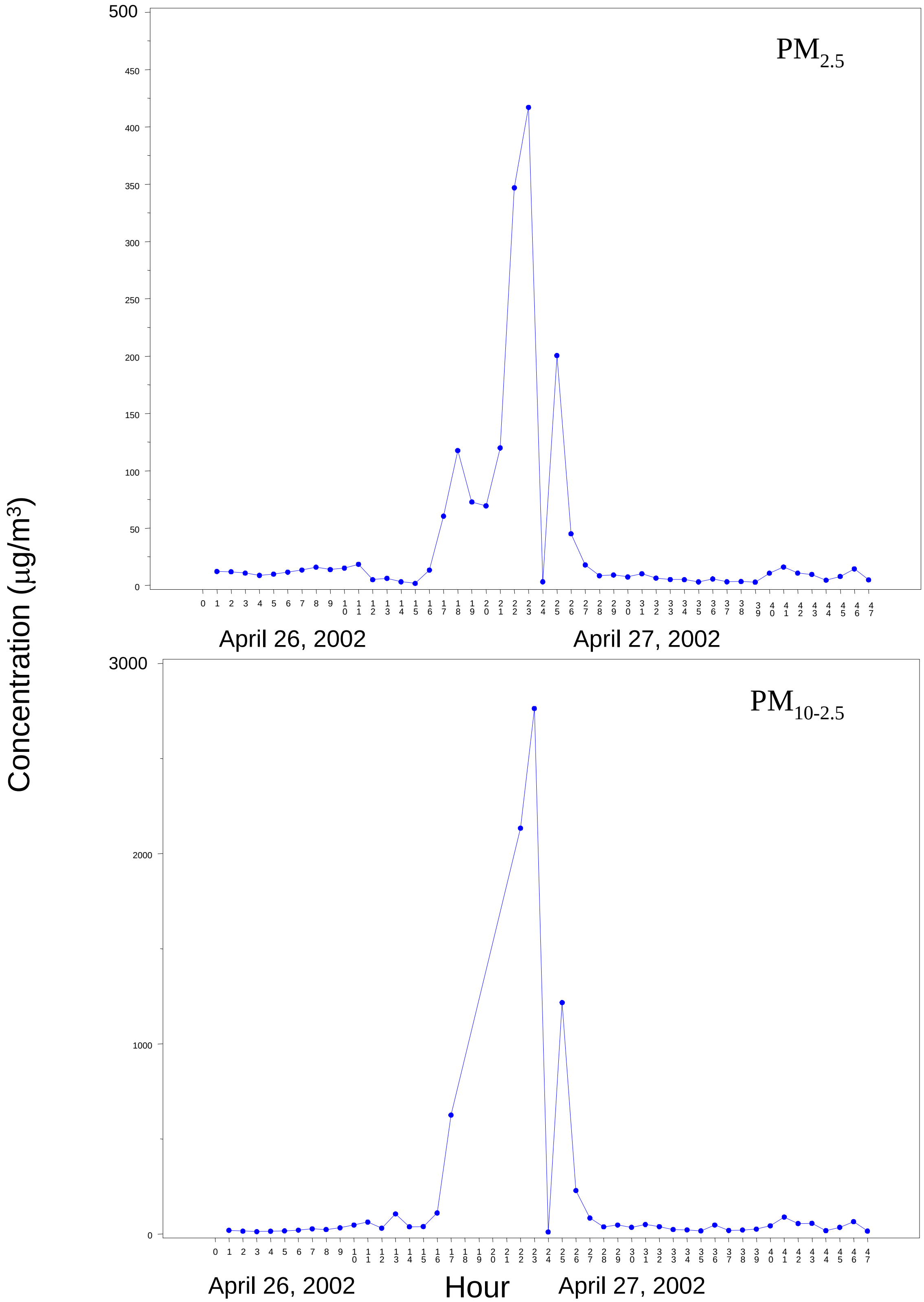
Hourly average $PM_{2.5}$ and $PM_{10\text{-}2.5}$ concentrations at a Greensboro, NC monitoring site, 2001-2003. Upper panel shows the distribution of $PM_{2.5}$ concentrations and the lower panel shows the distribution of $PM_{10\text{-}2.5}$ concentrations (box plot of interquartile range, mean, median, 5th and 95th percentiles)



Seasonal hourly average $PM_{2.5}$ and $PM_{10\text{-}2.5}$ concentrations at a Greensboro, NC monitoring site, 2001-2003. Upper panel shows the $PM_{2.5}$ concentrations and the lower panel shows the $PM_{10\text{-}2.5}$ concentrations.



Hourly average $PM_{2.5}$ and $PM_{10\text{-}2.5}$ concentrations at a Denver, CO monitoring site, 2001-2003. Upper panel shows the distribution of $PM_{2.5}$ concentrations and the lowe panel shows the distribution of $PM_{10\text{-}2.5}$ concentrations. (Box plots of interquartile ranges, means, medians, 5^{th} and 95th percentiles.)



Hourly $PM_{2.5}$ and $PM_{10\text{-}2.5}$ concentrations at a El Paso, TX monitoring site, April 26, 2002-April 27, 2002. Upper panel shows the hourly $PM_{2.5}$ concentrations and the lower panel shows the hourly $PM_{10\text{-}2.5}$ concentrations. Note the different scales.

Source: Schmidt et al. (2005)

The UNIVARIATE Procedure
Variable: median (the median, diff)

Moments

N	264	Sum Weights	264
Mean	1.81780303	Sum Observations	479.9
Std Deviation	0.99451218	Variance	0.98905447
Skewness	2.0698422	Kurtosis	4.43380155
Uncorrected SS	1132.485	Corrected SS	260.121326
Coeff Variation	54.7095676	Std Error Mean	0.06120799

Basic Statistical Measures

Location		Variability	
Mean	1.817803	Std Deviation	0.99451
Median	1.500000	Variance	0.98905
Mode	2.000000	Range	5.50000
		Interquartile Range	0.80000

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	29.69878	Pr > |t|	<.0001
Sign	M	132	Pr >= |M|	<.0001
Signed Rank	S	17490	Pr >= |S|	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	6.0
99%	5.5
95%	4.0
90%	3.0
75% Q3	2.0
50% Median	1.5
25% Q1	1.2
10%	1.0
5%	1.0
1%	0.7
0% Min	0.5

The UNIVARIATE Procedure
Variable: median (the median, diff)

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
0.5	228	5.3	36
0.6	252	5.3	40
0.7	224	5.5	48
0.8	116	5.7	47
0.8	144	6.0	14

The UNIVARIATE Procedure
Variable: pct95 (the 95th percentile, diff)

Moments

N	264	Sum Weights	264
Mean	8.46041667	Sum Observations	2233.55
Std Deviation	4.47284534	Variance	20.0063455
Skewness	2.91291781	Kurtosis	13.3914067
Uncorrected SS	24158.4325	Corrected SS	5261.66885
Coeff Variation	52.8679085	Std Error Mean	0.2752846

Basic Statistical Measures

Location		Variability	
Mean	8.460417	Std Deviation	4.47285
Median	7.000000	Variance	20.00635
Mode	5.000000	Range	38.00000
		Interquartile Range	4.40000

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	30.73335	Pr > |t|	<.0001
Sign	M	132	Pr >= |M|	<.0001
Signed Rank	S	17490	Pr >= |S|	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	41.0
99%	27.0
95%	16.0
90%	13.5
75% Q3	10.0
50% Median	7.0
25% Q1	5.6
10%	5.0
5%	4.8
1%	4.0
0% Min	3.0

The UNIVARIATE Procedure
Variable: pct95 (the 95th percentile, diff)

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
3.0	226	23.0	161
3.9	116	26.6	106
4.0	257	27.0	13
4.0	256	28.4	189
4.0	232	41.0	14

Attachment C

PM Speciation Data Analysis

General/Background:

This attachment describes the PM speciation-related analyses. The PM speciation data were obtained from a variety of sources, including AQS, the VIEWS website (for IMPROVE data), from $PM_{2.5}$ Supersite Principal Investigators, and from the SouthEastern Aerosol Research and Characterization Study (SEARCH).

Construction of $PM_{2.5}$ Urban Speciation Database:

$PM_{2.5}$ speciation data were extracted from the AQS in July 2004. The following post-processing of the extracted data was done to arrive at the final database.

- All event-flagged data were removed.
- Only data from 'suburban' or 'urban' sites were retained; sites designated 'RURAL' in AQS were omitted from the database.
- Completeness was checked to ensure that a minimum of 11 observations were present for all major chemical components for each of the 4 quarters that make up a year. Quarters are defined as: January-March; April-June; July-September; and October-December. Major species include: organic carbon, elemental carbon, ammonium, all components of crustal material (as listed below in the Crustal Material equation), nitrate, and sulfate.
- New variables were computed as follows:
  - OCM = (Measured organic carbon, blank corrected) * 1.4.
  - Ammonium Sulfate = 1.375 * Measured Sulfate.
  - Ammonium Nitrate = 1.29 * Measured Nitrate.
  - Crustal Material = [2.2 * Al] + [2.48 * Si] + [1.63 * Ca] + [2.42 * Fe] + [1.94 * Ti].
  - Total Carbonaceous Mass (TCM) = Organic Carbon Mass (OCM) + Elemental Carbon (EC).
- Data were reduced into quarterly and annual averages. Note that annual averages must have four complete quarters of data.
- Excel spreadsheets were developed of the extracted and reduced data.
- The source of this database is either referred to as the EPA Speciation Network (ESpN) or the Speciation Trends Network (STN).

Construction of $PM_{2.5}$ Rural Speciation Database:

IMPROVE $PM_{2.5}$ speciation data were extracted from the VIEWS website, http://vista.cira.colostate.edu/views/, in October 2004. A methodology similar to that outlined above (for the urban data) were used to reduce the IMPROVE data. It should be noted that VIEWS already does some of the computations outlined above.

- Data were used *as reported* to VIEWS website
- Only data for rural sites and the Washington, DC site were retained.
- Completeness was checked to ensure that a minimum of 11 observations were present for all major species for each of the 4 quarters for the year in question. Quarters are defined as before. Major species are also the same as before with the exception of ammonium, which is not routinely measured in the IMPROVE protocol.
- All the variables calculated above are automatically computed by and reported in VIEWS for the IMPROVE data. In addition, the PM_{10} measured values were retained as an additional parameter for all observations.
- The data were reduced into quarterly and annual averages. Note that annual averages must have four complete quarters of data.
- As before, Excel spreadsheets were developed of all the extracted and reduced data.

Creation of PM Supersite Database:

A summary of speciation data collected at the Los Angeles Supersite (USC site) was sent to EPA by its Principal Investigator, Constantinos Sioutas. These data were used as delivered. Mass and speciation data were available for $PM_{1.0}$ (ultrafine), $PM_{2.5}$ (fine), and $PM_{10\text{-}2.5}$ (coarse). These data spanned the one year 10/2002 to 9/2003.

Creation of SEARCH Database

$PM_{10\text{-}2.5}$ data for calendar year 2003 were retrieved from the public archive of the SouthEastern Aerosol Research and Characterization (SEARCH) network (http://www.atmospheric-research.com/) for four monitoring in Georgia and Alabama.

Analysis 1 – Rural $PM_{2.5}$ and chemical constituent trends for 1993-2003.

Because there aren't enough urban $PM_{2.5}$ speciation data to construct a trend line, rural IMPROVE data were used to generate an 11-year trend. Note that the Washington, DC IMPROVE data can only be used to investigate urban trends in that single location.

Goals:

- To show $PM_{2.5}$ mass and chemical constituent trends by region.
- To better explain how mass and components vary spatially and temporally.

Outputs:

- See Output C.1.

Methods:

- Using the $PM_{2.5}$ rural database constructed above, sites that were complete for the entire time period of 1993-2003 (by quarter) were retained and binned into East, West, and DC.
- Annual averages were computed by site for each year then averaged for the entire region. Only one site went into the Washington, DC trend line.
- Line graphs (to represent $PM_{2.5}$ mass, ammonium sulfate, ammonium nitrate, total carbonaceous mass (TCM), and crustal material) were generated in a spreadsheet (Lotus 1-2-3)

Analysis 2 – Rural $PM_{10-2.5}$ Trends for 1993-2003

As with Analysis 1 outlined above, IMPROVE data were used to construct a trend line of the difference between PM_{10} and $PM_{2.5}$ (which represents the coarse fraction of PM).

<u>Goals</u>:

- ? To show $PM_{10-2.5}$ trends by Region of the United States.
- ? To show and contrast East, West, and Washington, DC (urban) trends.

<u>Outputs</u>:

- o See Output C.2.

<u>Methods</u>:

- Using the $PM_{2.5}$ rural database constructed above, sites that were complete for the entire time period of 1993-2003 were retained and binned into East, West, and DC. Only two variables were checked for completeness, $PM_{2.5}$ and PM_{10}.
- Annual averages were computed by site for each year and then averaged for the entire region. Note that only one site went into the trend line for Washington, DC.
- Line graphs were generated in a spreadsheet (Lotus 1-2-3).

Analysis 3 – 2003 Annual Average urban and rural $PM_{2.5}$ speciation patterns

<u>Goals</u>:

- ? To show urban and rural speciation patters for the year 2003 by Region of the United States. These regions were: Northeast, Southeast, Industrial Midwest, Upper Midwest, Southwest, Northwest, and Southern California.

<u>Outputs</u>:

- o See Output C.3.

Methods:

- Using the rural and urban $PM_{2.5}$ speciation databases developed above, sites that were complete for year 2003 were retained
- Annual averages were computed for each site for the year 2003.
- Sites were binned into each of the Regions mentioned above and regional averages were computed.
- Urban and rural stacked bar charts for each region [using the major species ammonium sulfate, ammonium nitrate, total carbonaceous mass (TCM), and crustal material] were generated in Lotus.

Analysis 4 – 2003 Seasonal urban and rural $PM_{2.5}$ speciation patterns

<u>Goals</u>:

- ? To ascertain if there are seasonal variations in regional speciation profiles.

<u>Outputs</u>:

- o See Output C.4.

<u>Methods</u>:

- For those sites that were complete for the entire year of 2003, quarterly averages were computed for all the major components and binned by region.

- Urban and rural stacked bar charts (for the major components) for each region, for each of the four seasons were generated via Lotus.

Analysis 5 - Evaluation of (10/2002 to 9/2003) Los Angeles 'Supersite' multiple size-cut speciation data

Goals:

? Using one year of available data, as developed above, for the USC Los Angeles supersite, to evaluate speciation patterns of ultrafine, fine, and coarse fraction PM.

Outputs:

o See Output C.5

Methods:

- Using data as directly received from the Los Angeles Supersite's principal investigator, speciation patterns were depicted (generated in Lotus 1-2-3 spreadsheets) using pie charts for each of the three modes mentioned above.
- Major species displayed in the pie charts include sulfate, nitrate, organic carbon, elemental carbon, and crustal elements. The latter includes Fe, Ca, Si, Al, K and trace metals. These data are the same as published in Figures 3, 4 and 5 as bar charts in Sardar et al.[1] In the Sardar paper, crustal elements are labeled metals. For consistency with the data presented by Sardar, the pie charts represent reported measurements and do not represent their corresponding PM mass.

Analysis 6 –Coarse mass and its composition at the LA 'Supersite' compared with two pairs of urban rural data in the Southeastern U.S.

Goals:

? Using one year of available $PM_{10-2.5}$ data at the USC Los Angeles Supersite and the most recent data for two urban sites and nearby rural sites in the southeastern US from the SEARCH network, evaluate $PM_{10-2.5}$ mass and particle composition differences.

Outputs:

o Stacked bar charts, by concentration and percentage mass; see Output C.6.

Methods:

- The USC measurement data are presented in terms of their estimated $PM_{10-2.5}$ particle mass. Inorganic nitrate and sulfate concentrations were assumed to be solely associated with their ammonium salts, the crustal component reflect the measured elements plus their common oxides and also include additional trace metals. Organic carbon mass was estimated by multiplying the measured organic carbon by a factor of 2.5 to account for the mass of H, O, and other elements in the coarse particle organic compounds. This factor, which corresponds to the OC to OC mass multiplier for cellulose, results in better mass closure than the use of the traditional OC multiplier of 1.4. It is the same factor used for the SEARCH data as described below.

[1] Satya B. Sardar, Philip M. Fine, and Constantinos Sioutas, "Seasonal and spatial variability of the size-resolved chemical composition of particulate matter (PM10) in the Los Angeles Basin", Journal of Geophysical Research, VOL. 110, (2005)

- Four sites were selected from the SEARCH network to characterize $PM_{10-2.5}$ mass and its composition for two urban sites and their nearby rural locations: Birmingham (BHM, urban), Centerville, AL (CTR, rural), Atlanta, GA (ATL, urban) and Yorkville, GA (YRK, rural). Because January - March data were missing for 2003, only the April 2003-December 2003 data was used for these comparisons. The Atlanta site is labeled JST by SEARCH.
- The SEARCH $PM_{10-2.5}$ data only includes SO_4, NO_3, NH_4 and six crustal elements: Al, Si, Ca, Fe, Ti and K. The mass associated with the crustal elements is presented in terms of their common oxides: Al_2O_3, SiO_2, K_2O, CaO, TiO_2, and Fe_2O_3. SEARCH labels this summary term 'major metal oxides' which is explicitly defined as [Al * 1.89] + [Ca * 1.40] + [Fe * 1.43] + [K * 1.21] + [Si * 2.14] + [Ti x 1.67]. For this analysis, sulfate and nitrate mass is estimated as fully ammoniated salts as done for the USC data.
- As suggested by Edgerton et al.[2], total $PM_{10-2.5}$ carbonaceous mass is estimated as the difference between measured $PM_{10-2.5}$ mass and its estimated inorganic constituents. To derive the estimated EC and OC portions of this carbon mass, the ratio of OC to EC (BC) are used as reported by Edgerton et al. from special carbon measurement studies during 2000, 2001, 2003 and 2004. The mean $PM_{10-2.5}$ EC and OC concentrations (in $\mu g/m^3$) are presented below.

SEARCH carbon special study data			
site	n	EC ($\mu g/m^3$)	OC ($\mu g/m^3$)
CTR	92	0.27	1.17
BHM	98	2.70	2.18
YRK	60	0.08	1.07
JST	121	0.21	1.60

- To account for the mass associated with OC, a 2.5 OC factor is used. This is consistent with a large portion of the coarse fraction particle OC as vegetative material and provides better coarse mass closure as observed by Edgerton et al.
- This processing results in the following estimated fractional portions of carbon mass which are then used in this analysis.

Fractional portions of carbon mass in $PM_{10-2.5}$		
site	EC	OCM
CTR	0.08	0.92
BHM	0.33	0.67
YRK	0.03	0.97
JST	0.05	0.95

[2] Eric S Edgerton, Gary S. Casuccio, John J. Jansen, Benjamin E. Hartsell. "Measurement of Carbonaceous Material in PMcoarse. poster presented September, 2004 at the 8th International Conference on Carbonaceous Particles in the Atmosphere.

Analysis 7 – Evaluation of $PM_{2.5}$ speciation on high mass ('dirty') days

Goals:

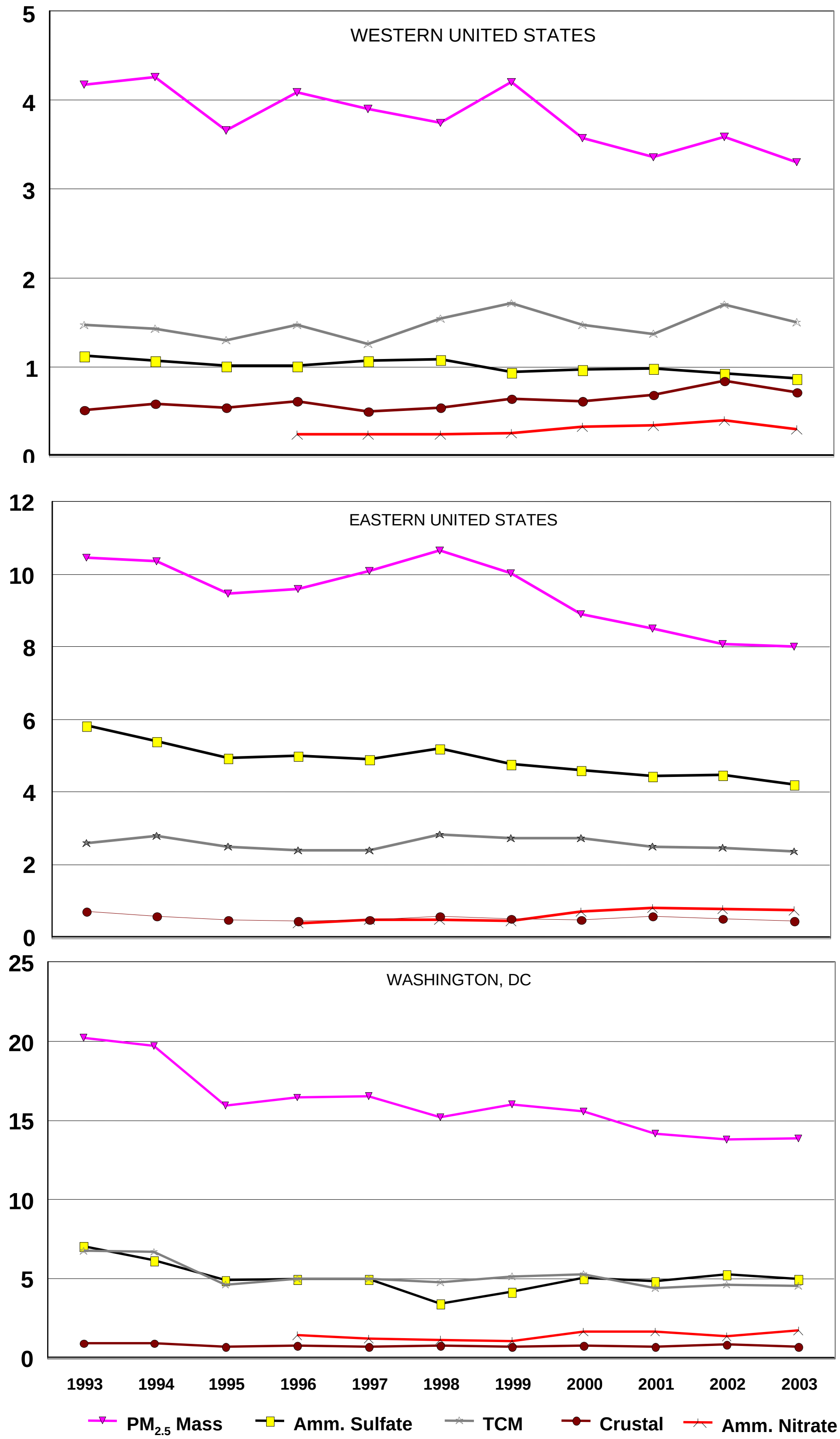
- ? To compare site-level component profiles on high $PM_{2.5}$ mass (a.k.a., 'the dirtiest') days to annual average profiles. (To see if there are differences in the relative proportions of the major speciation components.)

Outputs:

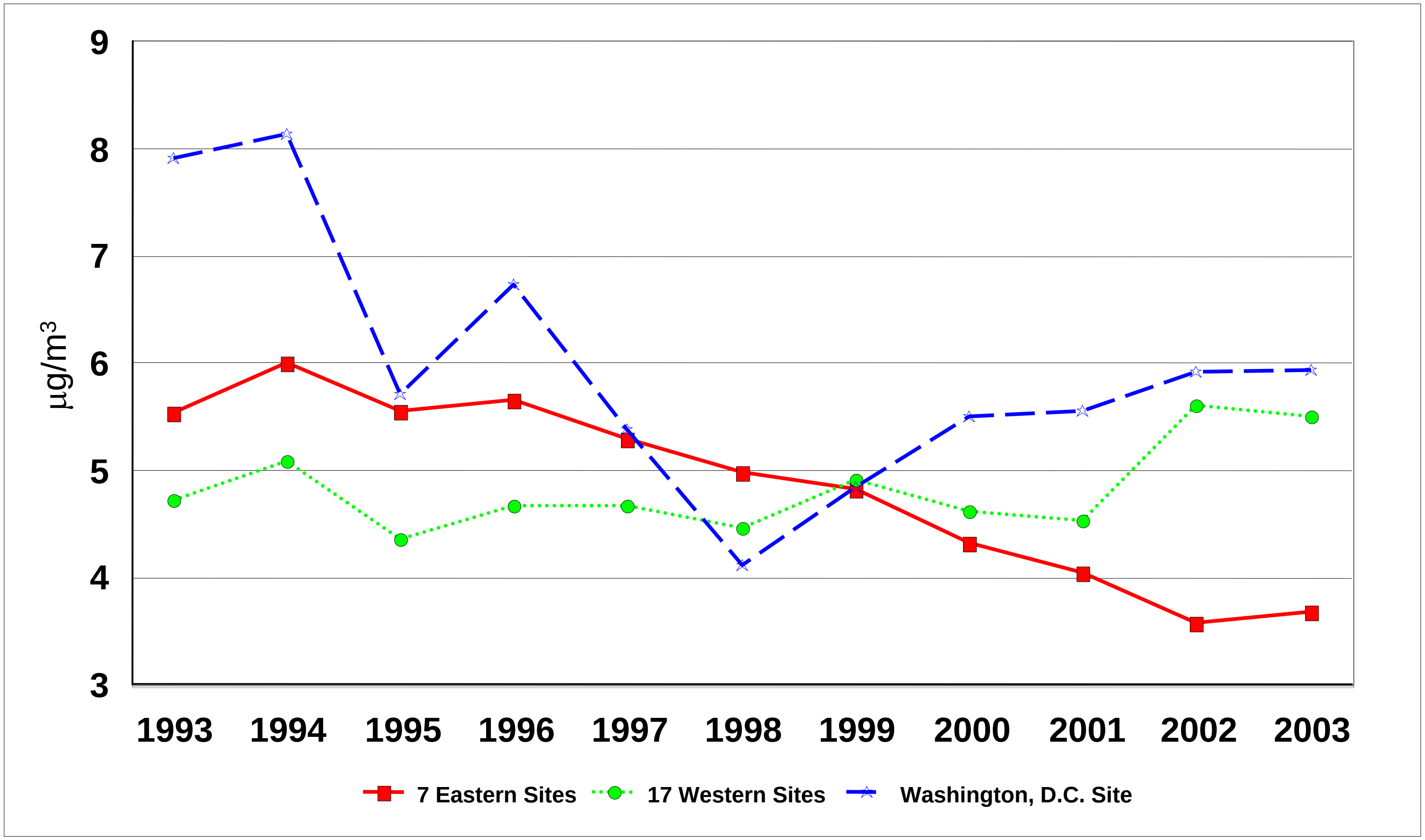
- o For 8 urban speciation sites, the chemical composition of $PM_{2.5}$ is shown for the 2003 annual average and the 5 highest mass days, the latter both individually and in aggregate. Results are shown in tabular form and in pie charts; see Output C.7.

Methods:

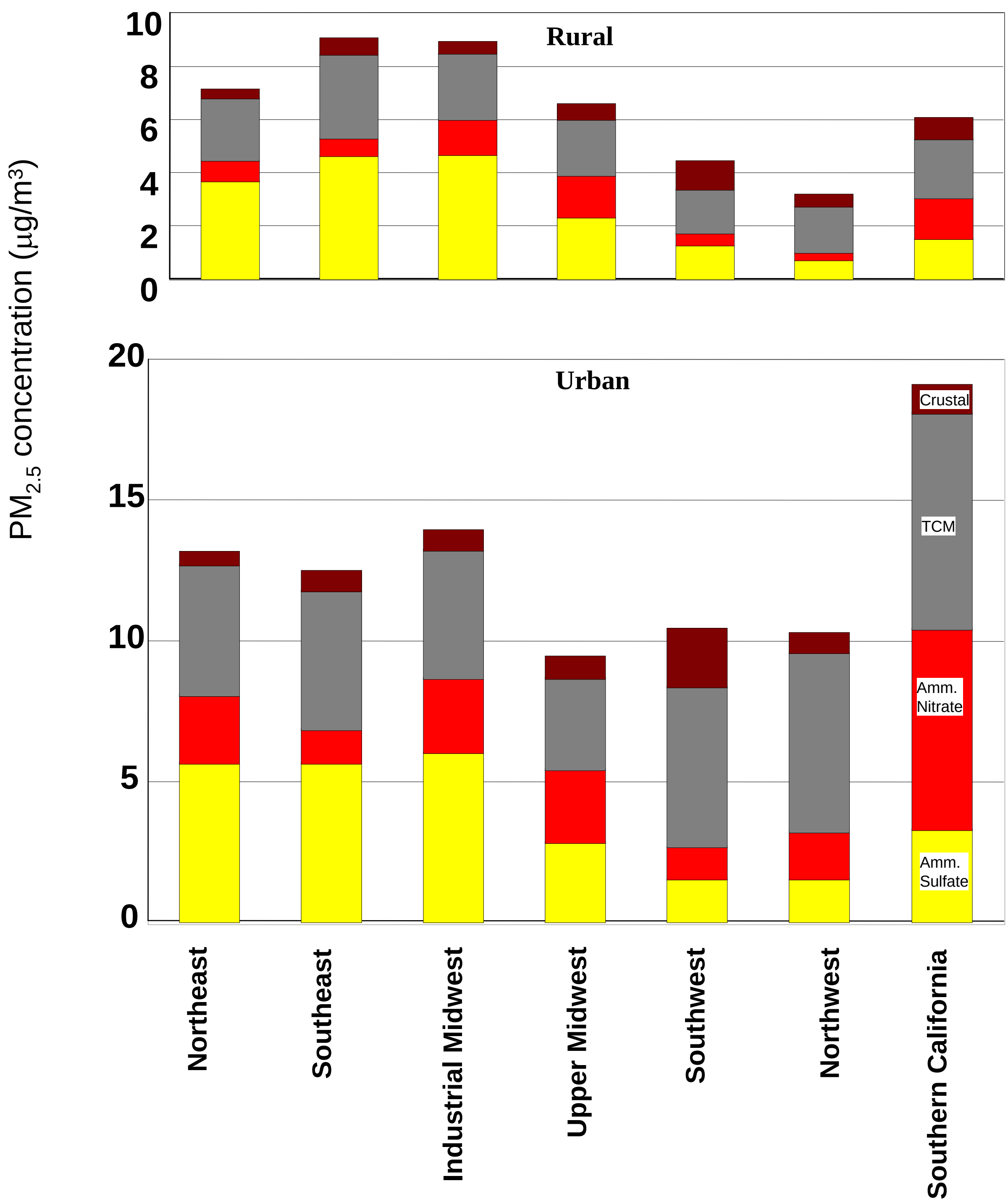
- Data for eight urban monitoring sites, representing major metropolitan areas, were culled from the 2003 urban speciation network database.
- Analysis focused on the four speciation components: ammonium sulfate, ammonium nitrate, total carbonaceous mass (or TCM, the sum of EC and OCM), and crustal material.
- For identifying the highest $PM_{2.5}$ ('dirtiest') days, collocated FRM mass was used as the indicator. For a very few days where collocated FRM mass wasn't available, gravimetric mass was substituted. The top 5 days (in 2003, at each of those 8 sites) were used as the metric for the 'dirty' end of the spectrum. [For a site operating on a 1-in-3 day sampling schedule, the 98^{th} percentile is the 3^{rd} highest value.]
- Lotus was used to process the data; Lotus and Freelance were used to produce the charts and tables.



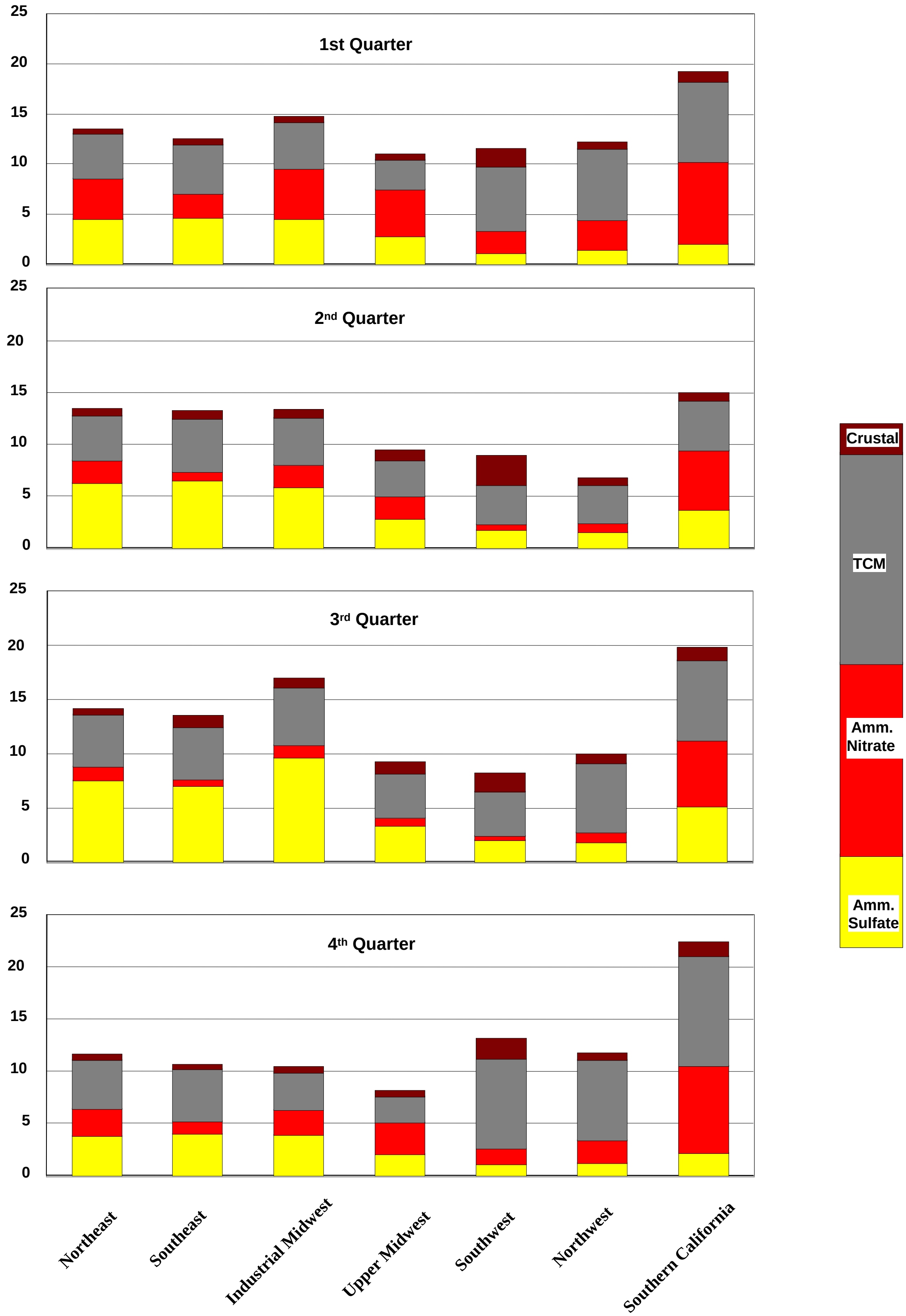
Average annual average trend in $PM_{2.5}$ mass, ammonium sulfate, ammonium nitrate, total carbonaceous mass, and crustal material at IMPROVE sites, 1993-2003.



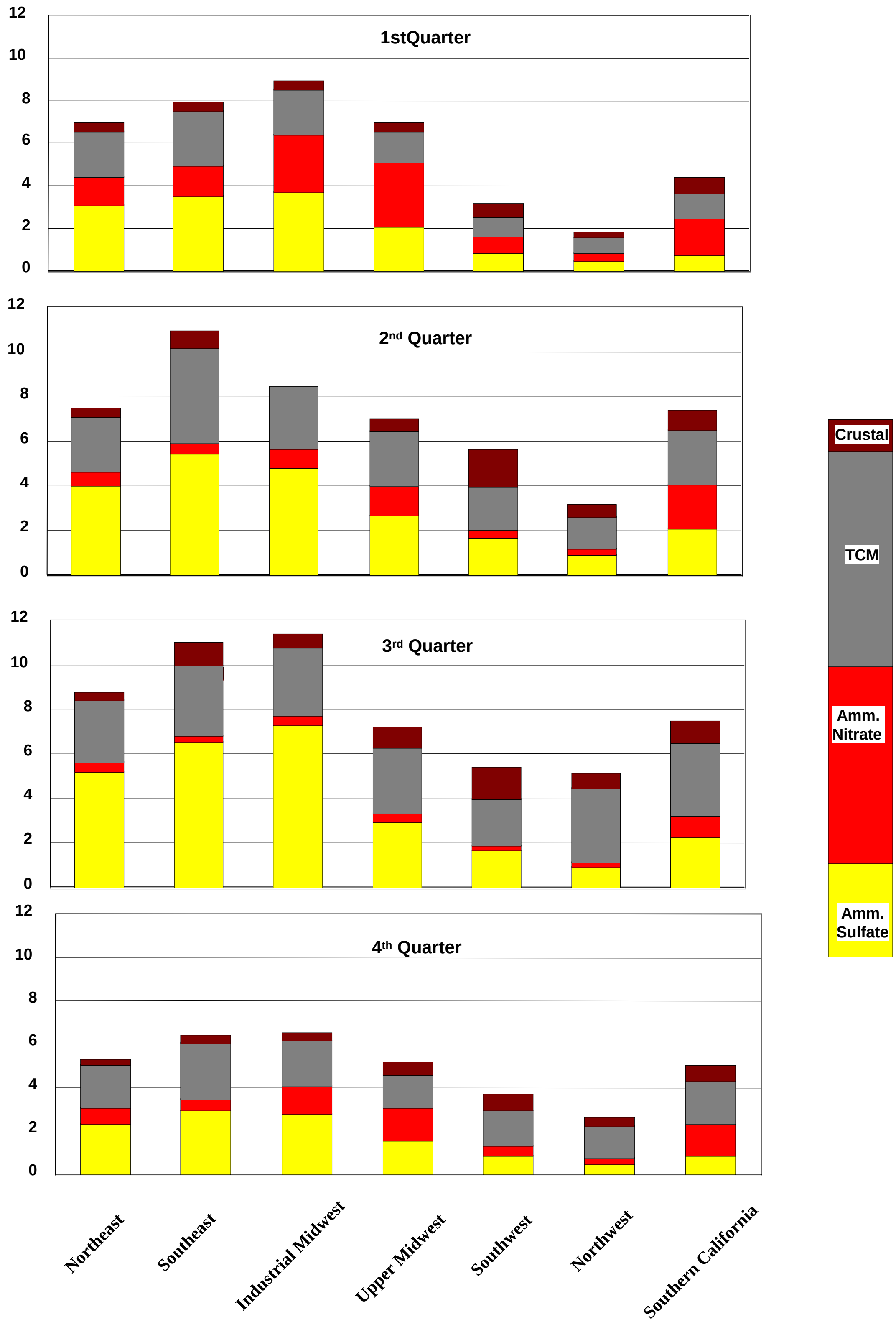
Average measured annual average $PM_{10\text{-}2.5}$ concentration trend at IMPROVE sites, 1993-2003.



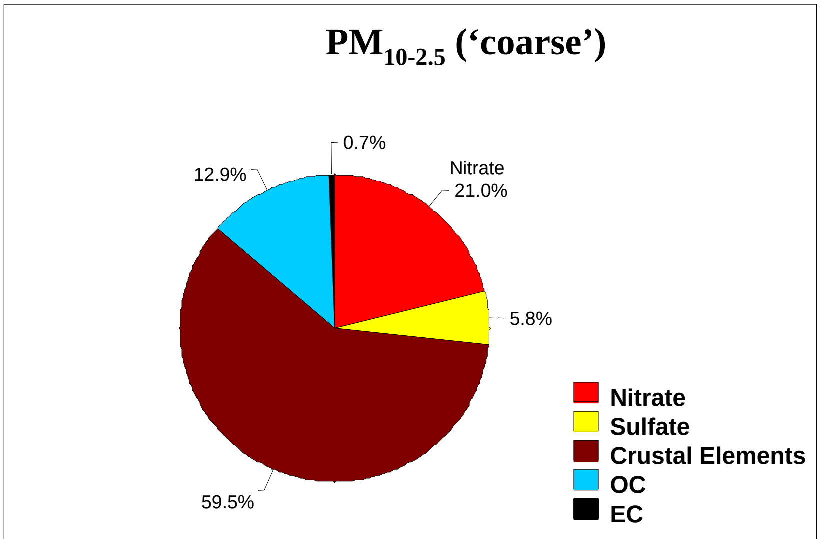
Annual average composition of $PM_{2.5}$ by region, 2003. Rural data (top panel) from IMPROVE network, urban data (bottom panel) from EPA Speciation Network. Components (from top to bottom) are crustal material, total carbonaceous mass (TCM), ammonium nitrate, and ammonium sulfate.

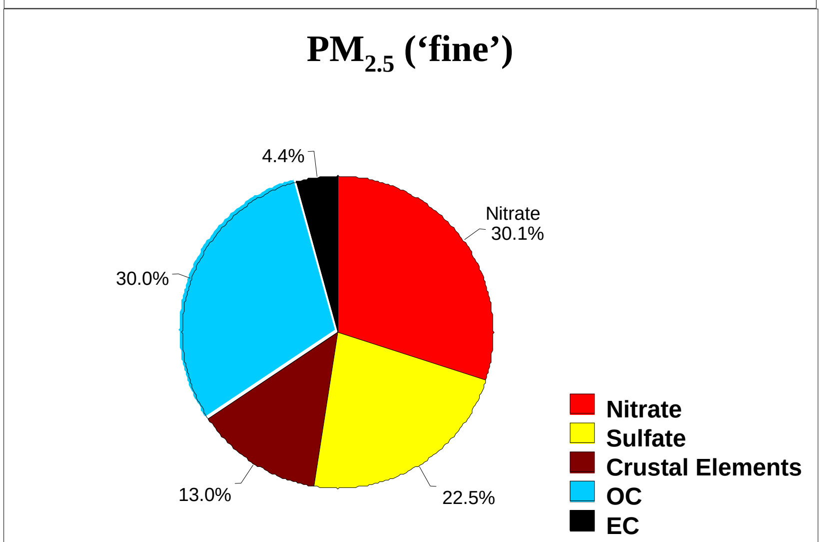


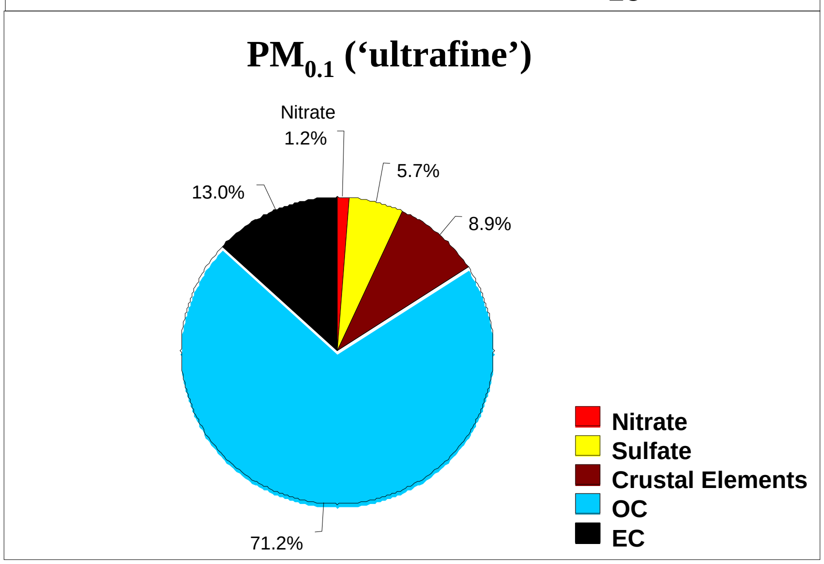
Seasonal average composition of urban $PM_{2.5}$ by region, 2003. Data from EPA Speciation Network. Components (from top to bottom) are crustal material, total carbonaceous mass (TCM), ammonium nitrate, and ammonium sulfate.



Seasonal average composition of rural $PM_{2.5}$ by region, 2003. Data from IMPROVE Network. Components (from top to bottom) are crustal material, total carbonaceous mass (TCM), ammonium nitrate, and ammonium sulfate.

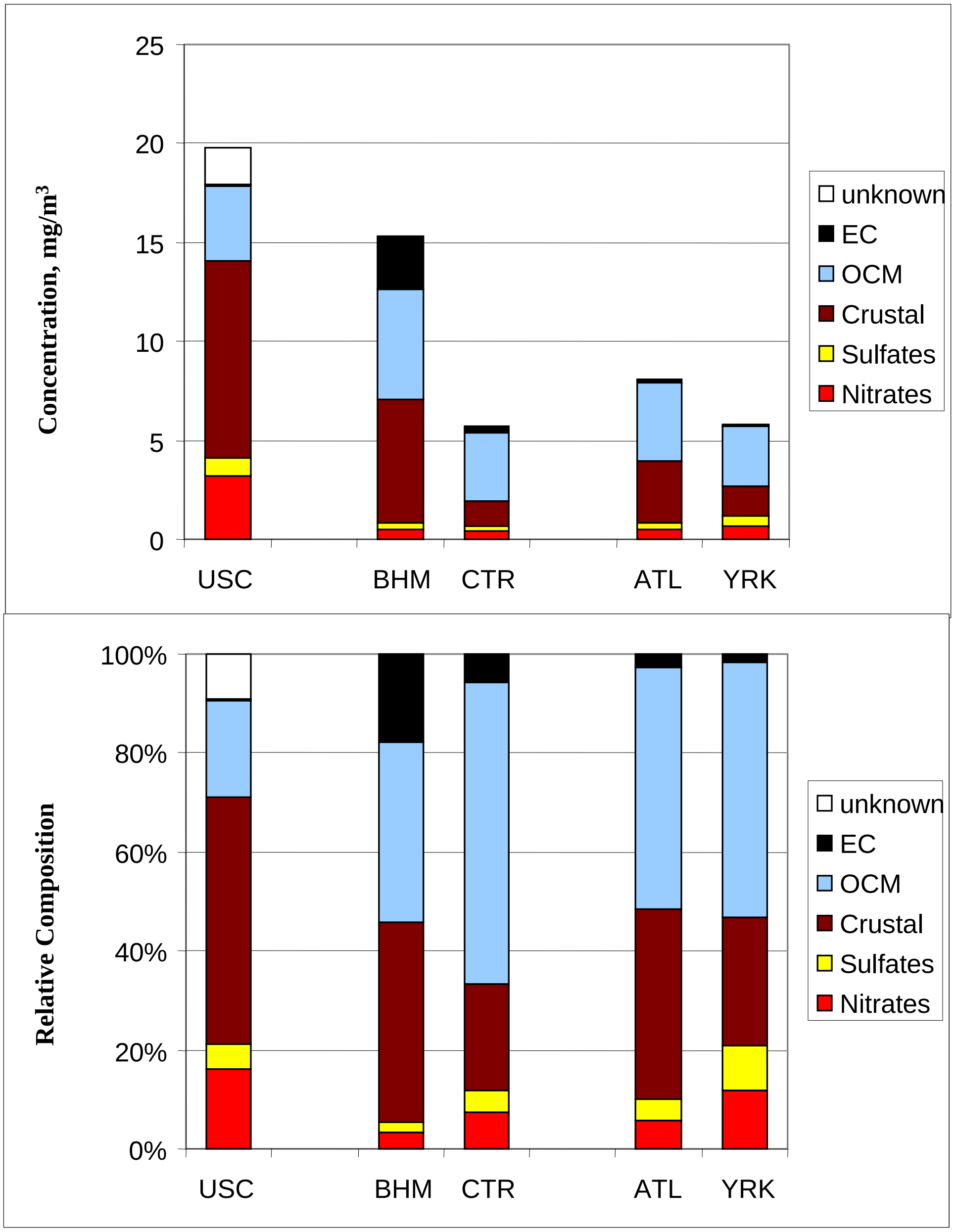






Average $PM_{10-2.5}$, $PM_{2.5}$, and $PM_{0.1}$ (ultrafine) chemical composition at theUSC EPA 'supersite' monitor in Los Angeles, CA, 10/2001 to 9/2002. Components represent measured ions, carbon and crustal elements including trace metals and are shown in clockwise order (starting with nitrate) as listed in legend from top to bottom.

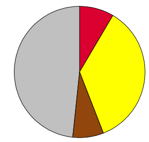
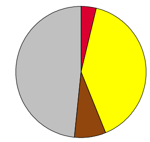
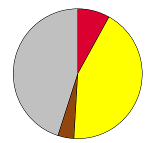
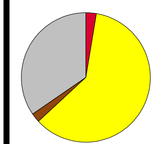
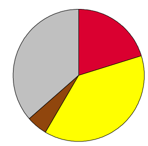
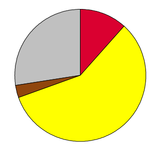
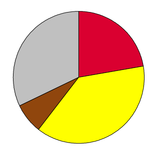
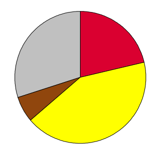
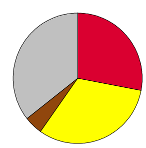
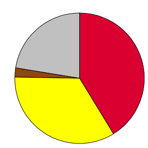
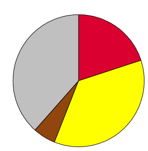
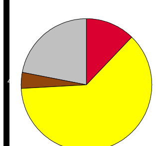
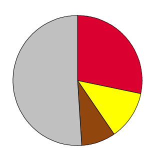
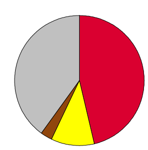
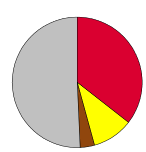
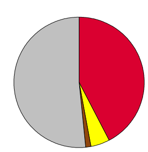
Source: Sardar et al. (2005)



Average PM10-2.5 composition for Los Angeles and two eastern urban-rural pairs. Based on USC Supersite data (10/2002 to 9/2003), and Birmingham, AL (BHM, urban), Centerville, AL (CTR, rural), Atlanta, GA (ATL, urban) and Yorkville, GA (YRK, rural) monitoring sites in the Southeastern Aerosol Research and Characterization (SEARCH) Study, 4/2003-12/2003. The top panel shows mass concentration, $\mu g/m^3$ and the bottom panel shows composition as percent of measured mass.

Source: USC site data (Sardar et al. 2005), eastern data (SEARCH website) adjusted as described in Schmidt et al. (2005)

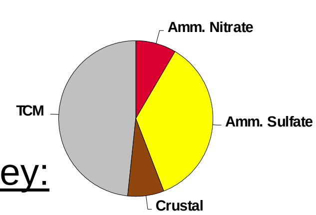
$PM_{2.5}$ composition on high mass days in select urban areas, 2003.

Urban Area	Statistic*	Composition Percents (%)				$PM_{2.5}$ mass** (µg/m3)	Annual average	Average of 5 highest days
		Amm. Nitrate	Amm. Sulfate	Crustal	TCM			
Birmingham, AL	• Annual average	8.5	35.6	7.6	48.3	**17.9**		
	• Average of 5 highest $PM_{2.5}$ mass days	3.8	40.0	7.8	48.3	**40.7**		
	• Highest $PM_{2.5}$ mass day	1.9	55.1	5.5	37.4	**46.6**		
	• 2nd highest $PM_{2.5}$ mass day	4.2	26.9	11.0	57.9	**40.4**		
	• 3rd highest $PM_{2.5}$ mass day	15.3	15.7	10.7	58.4	**39.2**		
	• 4th Highest $PM_{2.5}$ mass day	2.7	51.1	7.4	38.7	**39.1**		
	• 5th Highest $PM_{2.5}$ mass day	2.6	34.6	6.4	56.3	**38.3**		
Atlanta, GA	• Annual average	8.1	42.8	4.0	45.0	**15.2**		
	• Average of 5 highest $PM_{2.5}$ mass days	2.6	60.1	2.3	34.3	**35.2**		
	• Highest $PM_{2.5}$ mass day	2.0	70.5	1.9	25.6	**37.8**		
	• 2nd highest $PM_{2.5}$ mass day	2.0	47.8	2.5	47.8	**37.1**		
	• 3rd highest $PM_{2.5}$ mass day	2.4	67.6	2.1	27.9	**36.8**		
	• 4th Highest $PM_{2.5}$ mass day	3.2	50.8	2.9	43.1	**35.0**		
	• 5th Highest $PM_{2.5}$ mass day	3.6	67.5	1.9	27.0	**29.3**		
New York City, NY	• Annual average	20.2	38.3	5.1	36.4	**13.1**		
	• Average of 5 highest $PM_{2.5}$ mass days	11.6	57.9	3.0	27.4	**40.5**		
	• Highest $PM_{2.5}$ mass day	3.6	58.3	5.5	32.6	**45.9**		
	• 2nd highest $PM_{2.5}$ mass day	5.0	69.0	1.4	24.6	**45.8**		
	• 3rd highest $PM_{2.5}$ mass day	27.8	42.1	3.1	27.0	**38.2**		
	• 4th Highest $PM_{2.5}$ mass day	5.1	59.4	4.6	30.9	**36.4**		
	• 5th Highest $PM_{2.5}$ mass day	9.7	62.2	2.0	26.1	**36.0**		
Cleveland, OH	• Annual average	22.3	38.3	7.4	32.1	**17.6**		
	• Average of 5 highest $PM_{2.5}$ mass days	21.4	42.5	6.3	30.0	**44.1**		
	• Highest $PM_{2.5}$ mass day	32.7	43.2	2.3	21.7	**57.9**		
	• 2nd highest $PM_{2.5}$ mass day	25.1	41.5	4.0	29.3	**46.4**		
	• 3rd highest $PM_{2.5}$ mass day	4.8	64.4	8.7	22.1	**45.5**		
	• 4th Highest $PM_{2.5}$ mass day	8.8	37.5	14.7	39.0	**35.7**		
	• 5th Highest $PM_{2.5}$ mass day	31.4	20.5	4.0	44.0	**35.0**		
Chicago, IL	• Annual average	28.0	31.8	4.6	35.6	**15.2**		
	• Average of 5 highest $PM_{2.5}$ mass days	41.2	34.0	2.3	22.4	**34.4**		
	• Highest $PM_{2.5}$ mass day	46.0	30.7	1.2	22.1	**38.3**		
	• 2nd highest $PM_{2.5}$ mass day	49.2	36.4	0.8	13.6	**35.3**		
	• 3rd highest $PM_{2.5}$ mass day	51.8	27.7	1.2	19.3	**35.1**		
	• 4th Highest $PM_{2.5}$ mass day	5.6	61.7	3.8	28.9	**32.5**		
	• 5th Highest $PM_{2.5}$ mass day	47.8	16.1	5.3	30.8	**30.7**		
St. Louis, MO	• Annual average	20.0	36.0	5.6	38.4	**14.5**		
	• Average of 5 highest $PM_{2.5}$ mass days	12.2	61.9	3.9	22.0	**35.9**		
	• Highest $PM_{2.5}$ mass day	6.2	69.1	3.6	21.0	**50.6**		
	• 2nd highest $PM_{2.5}$ mass day	5.0	67.0	2.0	26.0	**36.0**		
	• 3rd highest $PM_{2.5}$ mass day	6.4	69.2	3.2	21.3	**33.1**		
	• 4th Highest $PM_{2.5}$ mass day	5.0	58.9	8.2	28.1	**30.8**		
	• 5th Highest $PM_{2.5}$ mass day	40.2	42.3	2.7	14.7	**28.9**		
Salt Lake City, UT	• Annual average	28.3	12.2	8.5	51.1	**10.0**		
	• Average of 5 highest $PM_{2.5}$ mass days	46.3	10.8	2.9	40.0	**40.6**		
	• Highest $PM_{2.5}$ mass day	50.6	6.3	2.5	40.5	**59.5**		
	• 2nd highest $PM_{2.5}$ mass day	43.5	11.9	2.6	42.0	**52.1**		
	• 3rd highest $PM_{2.5}$ mass day	42.4	13.5	3.7	40.4	**34.2**		
	• 4th Highest $PM_{2.5}$ mass day	48.2	5.9	4.7	41.3	**28.7**		
	• 5th Highest $PM_{2.5}$ mass day	45.4	20.2	1.5	32.8	**28.4**		
Fresno, CA	• Annual average	35.5	10.2	3.6	50.7	**18.0**		
	• Average of 5 highest $PM_{2.5}$ mass days	42.4	4.7	1.3	51.6	**54.2**		
	• Highest $PM_{2.5}$ mass day	55.2	4.6	2.1	38.2	**59.0**		
	• 2nd highest $PM_{2.5}$ mass day	58.4	8.5	0.9	32.2	**56.3**		
	• 3rd highest $PM_{2.5}$ mass day	17.5	1.5	1.3	79.7	**54.4**		
	• 4th Highest $PM_{2.5}$ mass day	35.1	5.3	1.0	58.6	**52.6**		
	• 5th Highest $PM_{2.5}$ mass day	44.6	3.7	1.3	50.3	**50.0**		

* The 5 highest days shown (and aggregated) for each site actually represent the 5 highest days (based on collocated FRM mass; see next bullet) that the speciation monitor sampled. FRM monitors at different locations in the metropolitan area and/or collocated FRM measurements on days that the speciation sampler did not record valid data may have had higher values than some or all of the 5 high values shown. Event-flagged data were omitted from this analyses.

** '$PM_{2.5}$ mass' concentration represents the collocated (w/ speciation monitor) same-day FRM measurement unless not available, in which case the speciation monitor gravimetric mass was substituted.

Key:



Attachment D

Visibility-Related PM Analyses.

General / Background:

This attachment describes the SP visibility-related PM analyses. The following analyses were undertaken to address statements made in, and subsequent comments to, the first draft (2003) EPA Particulate Matter (PM) Staff Paper (SP). Specifically, the draft PM SP noted in Chapter 6 that "EPA recognized that the selection of an appropriate level for a national secondary [$PM_{2.5}$] standard to address visibility protection was complicated by regional differences in visibility impairment due to several factors, including background and current levels of PM, the composition of PM, and average relative humidity." Preliminary analyses indicate that these regional differences are not significant when considering urban areas and daylight averaging periods. A rapidly growing national database of Federal Reference Method (FRM) $PM_{2.5}$, continuous $PM_{2.5}$ and $PM_{2.5}$ chemical speciation measurements now provides better opportunities to explore these factors.

General Goals:

Goals of these analyses were:

- Compare regional levels (including East versus West) of urban $PM_{2.5}$ mass, all hours and select daylight periods.
- Compare regional composition (including East versus West) of urban $PM_{2.5}$.
- Compare regional relative humidity (RH) (including East versus West) for select time periods (e.g., 24-hour average and shorter daylight periods).
- Compare regional relationships of urban visibility versus $PM_{2.5}$ mass levels for select time periods (using reconstructed extinction as the indicator for visibility).
- Estimate levels of $PM_{2.5}$ that equate to various visual range goals
- Evaluate the 'worst 20%' visibility days using both reconstructed extinction estimates and $PM_{2.5}$ as a surrogate.
- Estimate the rate of exeedance for various $PM_{2.5}$ levels (by county) for various different time periods.
- Estimate the percentage of counties that would violate potential NAAQS levels.

Formulae and Assumptions:

The visibility analyses carried out make wide use of several formulas and assumptions:

- Reconstructed light extinction formula from Malm et al (1999)[1]:

[1] Malm, W.C.; Day, D.E.; Kreidenweis, S.M. (1999). Light scattering characteristics of aerosols as a function of relative humidity: a comparison of measured scattering and aerosol concentrations using the theoretical models. J. Air Waste Manage. Assoc. 50: 686-709

$$
\begin{aligned}
b_{ext} = \; & [(3) * f(RH) * PM_{2.5} \text{ mass of sulfates}] + \\
& [(3) * f(RH) * PM_{2.5} \text{ mass of nitrates}] + \\
& [(4) * PM_{2.5} \text{ mass of OCM}] + \\
& [(10) * PM_{2.5} \text{ mass of EC}] + \\
& [(1) * PM_{2.5} \text{ mass of 'soil' or 'crustal'}] + \\
& [(0.6) * PM_{10-2.5} \text{ mass, a.k.a. PMc}] + \\
& 10 \quad \text{(Rayleigh scattering by gases)}
\end{aligned}
$$

Where …….

- b_{ext} is the calculated total light extinction in inverse megameters (Mm^{-1}). This is also referred to as reconstructed light extinction (RE).
- $PM_{2.5}$ component masses are in units of $\mu g/m^3$.
- 'Sulfates', 'nitrates', 'OCM', 'EC', and 'crustal' are defined as specified in Attachment C. Assumption is the 5 components comprise total $PM_{2.5}$ mass.
- *f(RH)* is the relative humidity adjustment factor that accounts for the relative humidity effects on hygroscopic aerosols. See Output A.i for a table of conversions. RH levels were capped at 95%, reflecting the lack of accuracy in higher relative humidity values and their disproportionately high impact on reconstructed light extinction; hence, *f(RH)* was capped at 7.4.

- Koschmieder relationship between visible range (VR) and RE:

  $$VR\ (km) = 3912 / RE\ (Mm^{-1})$$

  Or, restated

  $$RE\ (Mm^{-1}) = 3912 / VR\ (km)$$

- Assumption 1: Speciation profiles are fairly consistent over time (i.e., on a daily basis). Hence, a 24-hour speciation profile can be applied to corresponding hourly mass data. This assumption was made due to: a) the desire to evaluate sub-daily time periods (i.e., hourly increments) of visibility and PM relationships, and b) the lack of a sufficient amount and quality of hourly speciation data. Continuous speciation instruments are still undergoing further development and refinement. Continuous speciation data from a pilot study were evaluated (on a limited basis) to check this assumption. The nominal evaluation found some credibility in the assumption.
- Assumption 2: Speciation profiles are fairly consistent over space (e.g., within 50 miles). Hence, a speciation profile from a 'nearby' site can be applied to a non-source-oriented $PM_{2.5}$ continuous site (for the same day). Previous analyses by OAQPS have shown that multiple speciation sites in the same metropolitan area have similar profiles. In fact, there are considerable similarities at regional levels.
- Assumption 3: Hourly $PM_{10-2.5}$ levels can be estimated by applying regional ratios of 24-hr size cut ratios to hourly $PM_{2.5}$ data. Because of the desire to conduct the visibility analyses on an hourly (or multi-hour) time block, hourly $PM_{10-2.5}$ data were needed (per the RE formula). Nationally, there are *some* collocated hourly PM_{10} and hourly $PM_{2.5}$ monitors, but to limit the analyses to those sites would have produced a considerably small database. Thus, to make a more robust database, where collocated (PM_{10} and$PM_{2.5}$) hourly measurements were not available, the coarse fraction mass was estimated from hourly

$PM_{2.5}$ using 24-hour based size fraction ratios. These ratios were computed via Analyses 6 (and corresponding Output A.6) in Attachment A.

Construction of visibility database

When these visibility analysis were first initiated (early 2004), the most current available urban $PM_{2.5}$ speciation and PM continuous data (that represented at least a full year, and was seasonally unbiased) were for the timeframe April 2002 through March 2003. (RH data were available for longer and more current periods but the PM data needs drove the time period selection). The first visibility database (db) was constructed for that noted period. Later, around October 2004, the database was updated to represent the full year 2003 (January through December). A minor portion of the analyses referenced in this attachment reflect the older (3/2002 - 4/2003) db, but the majority reflect the newer db (1/2003 – 12/2003). If not otherwise specified, assume the newer db was utilized. The following statements document the creation of the newer db (though the processing steps taken for the early db were almost identical):

- EPA speciation network (ESpN) data for 2003 were provided by Tesh Rao of OAQPS on 9/21/2004. Data were retrieved from the EPA's national ambient air quality database, the Air Quality System (AQS), on May 17, 2004. See Attachment C for more details. For the visibility analyses, only the major profile component *percentages* (of the total of those 5 major components) were used. [The component percentages were multiplied by the $PM_{2.5}$ hourly data in order to estimate an hourly speciated dataset. Henceforth, let CP represent the component percentage (in mass) of the major component mass sum. That is, CP_sulf = component percentage of sulfates; CP_nit= component percentage of nitrates; CP_ocm = component percentage of organic carbon mass; CP_ec = component percentage of elemental carbon; and CP_cr = component percentage of crustal material.]
- $PM_{2.5}$ continuous data for 2003 were polled from AQS on August 24, 2004. [Data with method codes of 740 or 741 were excluded, per Tim Hanley of OAQPS.]
- PM_{10} continuous data for 2003 were retrieved from AQS on August 24, 2004.
- Raw National Weather Service (NWS) hourly data for 2003 (and also for 2001 and 2002; the additional 2 years were used in Analysis 4) were obtained from Bill Cox of OAQPS on March 19, 2004. Utilized fields were relative humidity (RH), barometric pressure (BP), and temperature (T).
- 10-year meteorological (relative humidity-related) database constructed by Ken Walsh of Science Applications International Corporation (SAIC) was received on February 18, 2004 and an addendum on October 5, 2004. The 10-year database contains averaged NWS site-level hourly RH and *f(RH)* (by site X month X hour). The 10 years of data encompassed 1988-1997.
- Database estimates for visibility (RE) were anchored at the continuous $PM_{2.5}$ sites.
- Hourly $PM_{2.5}$ and hourly PM_{10} data (for $PM_{10-2.5}$ estimation) were matched by site X date X hour; collocation was required. As noted above (assumption 3), if a continuous $PM_{2.5}$ site did not have collocated continuous PM_{10}, then hourly $PM_{10-2.5}$ value were estimated using regional 24-hour ratios of size fractions. Regional size fraction ratios were matched to the hourly $PM_{2.5}$ data by region.
- Hourly $PM_{2.5}$ data and daily speciation data were matched by site X date; either the speciation monitor had to be collocated or within 50 miles of the continuous $PM_{2.5}$ monitor.

Note implementation of assumptions 1 and 2. [More than half of the observations in the visibility db had either collocated speciation data or speciation data from within 3 miles.]

- Raw meteorological (met) data and 10-year average NWS RH-related data *from the nearest NWS site* were matched to the $PM_{2.5}$ continuous data by date X hour. [More than 75% of the NWS data used were within 21 miles of the $PM_{2.5}$ continuous monitor; more than 50% of the NWS data used were within 11 miles of the $PM_{2.5}$ continuous monitor.]
- Using the merged inputs identified above and the RE formula specified above, RE was computed (for each site-date-hour of 2003 where all components were available) according to the following general formula:

$$\begin{aligned} RE = \; & [(3) * f(RH) * (\text{hourly } PM_{2.5} * CP\_sulf)] + \\ & [(3) * f(RH) * (\text{hourly } PM_{2.5} * CP\_nit] + \\ & [(4) *(\text{hourly } PM_{2.5} * CP\_ocm)] + \\ & [(10) * (\text{hourly } PM_{2.5} * CP\_ec)] + \\ & [(1) * (\text{hourly } PM_{2.5} * CP\_cr)] + \\ & [(0.6) * (\text{hourly } PMc)] + \\ & 10 \end{aligned}$$

  Where …….
  - The *f(RH)* used in the formula either corresponded to the actual, same date-hour *f(RH)* value (from RH table look-up), or to the 10-year average *f(RH)*.
- In addition to computing hourly RE with either 'actual' *f(RH)* or 10-year average *f(RH)*, additional variations of the RE formula were also calculated. Examples of RE computational variations utilized in SP visibility analyses include:
  - RE_real = RE as above using the 'actual' *f(RH)*.
  - RE_avg10 = RE as above using the 10-year average *f(RH)*.
  - RE_real_avgc = same as RE_real except that regional size fraction ratios were always used to estimate hourly $PM_{10-2.5}$ (not just when collocated hourly PM_{10} was available).
  - RE_avg10_avgc = same as RE_avg10 except that regional size fraction ratios were always used to estimate hourly PMc (not just when collocated hourly PM_{10} was available).
  - RE_real_NC = same as RE_real except that the $PM_{10-2.5}$ component was omitted.
  - RE_avg10_NC = same as RE_avg10 except that the $PM_{10-2.5}$ component was omitted.
- SAS code was used to create the db's:
  - 'match hourly to nws and spec - 1b.sas' was used to ascertain a list of available collocated and/or nearby sites for $PM_{2.5}$ continuous, $PM_{2.5}$ speciation, and NWS.
  - 'merge espn nws cont - 2b.sas' merged the three input data files noted above and derived *f(RH)* for 'actual' RH.
  - 'pm10 - 3b.sas'merged continuous PM10 data (to estimate hourly $PM_{10-2.5}$ when available)

- ‘add regional pmc ratio - 4b.sas’ added the regional ratios of 24-hour size fractions (to estimate hourly $PM_{10-2.5}$ if hourly data PM10 were available at the site.
- ‘final calcs - 5b.sas’ made the RE computations.

Analysis 1 – Comparison of regional levels (including East versus West) of urban $PM_{2.5}$ mass and components

Goals:

? To characterize and compare regional and East and West urban $PM_{2.5}$ concentration levels.

Outputs:

- Annual averages of $PM_{2.5}$ (24-hour and shorter time-frame) were calculated from different networks and/or portions of networks. Various tables and graphs show the comparisons. See Output D.1.

Methods:

- SAS procedures (MEANS and SUMMARY) were used to compute averages. Freelance Graphics was used to make the plots.

Analysis 2 – Evaluation of relative humidity (RH) data.

Goals:

? To characterize the diurnal pattern of RH and FRH in various areas (e.g., East versus West).

? To compare RH and *f(RH)* from different db’s

Outputs:

- Various plots and tables were generated. See Output D.2.

Methods:

- SAS code (‘rh boxplots.sas’ and ‘RH boxplots for visdb.sas’) was used to generate the diurnal distribution box-plots.
- SAS code (‘diurnal final.sas’) generated the diurnally summarized *average* data; corresponding line plots were created in Freelance Graphics.
- The RH table was created in Excel.

Analysis 3 – Evaluation of relationship between RE and $PM_{2.5}$; evaluation of diurnal patterns in RE and model components; and evaluation of different averaging time blocks

Goals:

? To assess the relationship between RE and $PM_{2.5}$

? To characterize the diurnal pattern of RE and related components by region.

? To ascertain the cause of diurnal/model variations (‘outliers’).

? To investigate different time periods as possible candidates for an averaging time.

Outputs:

- Various line-plots and tables were generated; see Output D.3.

Methods:

- SAS code ('diurnal final.sas') generated the diurnally summarized *average* data.
- The line-plots were created with Freelance Graphics
- Tables were made in Excel.

Analysis 4 – Estimation of $PM_{2.5}$ levels needed to achieve various visual range goals

Goals:

? Help inform decision regarding 'level' of possible secondary standard (i.e., estimate the $PM_{2.5}$ levels needed to achieve various visual ranges).

Outputs:

o Regional boxplots for various visual range goals were generated; see Output D.4.

Methods:

- The methodology utilized: a) the formulas identified above for 'construction of visibility database'; b) estimates of annual regional speciation profiles as percents of total (using all available 2003 ESpN data, not just sites used in the visibility database); c) estimates of regional ratios of $PM_{10-2.5}$ to $PM_{2.5}$ (from Analysis A.6 in Attachment A; term 'ratio_pmc/pmf')); and estimates of the regional distributions of hourly RH (from all NWS site for the period 2001-2003, not just sites in the visibility db). Given those formulae and inputs, we then solved for hourly $PM_{2.5}$ for various visual range targets. SAS code ('target pmf for visual range goals.sas') was used to process the inputs and generate the boxplots. Below is a walk-through of the formulae; the underlined fields are the 'knowns' (using inputs b, c, & d):

As above ...	RE =	[(3) * *f(RH)* * (hourly $PM_{2.5}$ * CP_sulf)] + [(3) * *f(RH)* * (hourly $PM_{2.5}$ * CP_nit] + [(4) *(hourly $PM_{2.5}$ * CP_ocm)] + [(10) * (hourly $PM_{2.5}$ * CP_ec)] + [(1) * (hourly $PM_{2.5}$ * CP_cr)] + [(0.6) * (hourly $PM_{10-2.5}$)] + 10
As above ...	RE =	3912 / VR
Substituting ...	'$PM_{2.5}$ * ratio_pmc/pmf' for '$PM_{10-2.5}$'	
Hence ...	$PM_{2.5}$ =	(RE -10) / (3* *f(RH)* * CP_sulf) + (3 * *f(RH)* * CP_nit) + (4 * CP_ocm) + (10 * CP_ec) + (1 * CP_cr) + (.6* ratio_pmc/pmf))
Thus ...	For each hour and each region, 100 estimates of $PM_{2.5}$ were calculated (*f(RH)* being the varying field, $f(RH)_1$ to $f(RH)_{100}$ representing the regional distribution). Hourly $PM_{2.5}$ estimates were then averaged to the desired averaging time (24-hours or 4-hour block, 12 p.m. - 4 p.m.)	

Analysis 5 – Evaluation of the 'worst 20%' days

Goals:

? To estimate visual ranges in the 'east' and 'west' for the worst 20% days in a year (based on 24-hour average and 4-hour block average).

- ? To ascertain the approximate annual percentile value that would represent the (average of) worst 20% days. A CASAC comment noted that 'average of 20% worst' is actually closer to 92^{nd} than 90^{th} percentile. Hence, this analysis helps inform considerations of the 'form' of the standard.

Outputs:
- o Table of visual ranges for worst 20% days; see Attachment D.5a.
- o SAS output showing the distribution of annual site percentile moments that correspond to the mean of the worst 20% days. See Attachment D.5b

Methods:
- Table of visual ranges: For the 2003 visibility database, site-level 80^{th} percentile values were calculated for both 24-hour and 4-hour average RE estimates. The 24-hour and 4-hour RE averages were only considered valid if they included at least 18 or 3 hours each, respectively. All estimated RE values greater or equal to those 80^{th} percentile values ('i.e., the 'worst' 20%) were then converted to visual ranges and averaged. The site-level average visual ranges were then average by region and by 'east' / 'west'. The analysis was conducted with SAS code ('20haziest.sas').
- Percentile representing worst 20%: This analysis was conducted with SAS code ('mean of worst 20percent.sas'). First, site level means of the sites' 20% worst days were calculated. 'Worst days' (and thus, means of) were based on two metrics, estimated RE (using the 2003 visibility database), and $PM_{2.5}$ concentration (using available continuous data for 2001-2003). The site-level (worst 20%) means where then compared to the entire site-level distributions to ascertain where approximately (by rounded percentile) they would fall. The national distribution of these site-level moments were then computed.

Analysis 6 – Estimation of the exceedance levels of various $PM_{2.5}$ levels

Goals:
- ? Estimate county exceedance levels (as percentage of days) for varying $PM_{2.5}$ levels for different timeframes.
- ? Estimate the number, percentage and population of counties in the U.S. not likely to meet alternative levels/forms of a secondary $PM_{2.5}$ standard; estimate the percentage of counties on a regional basis not likely to meet alternative secondary standards.

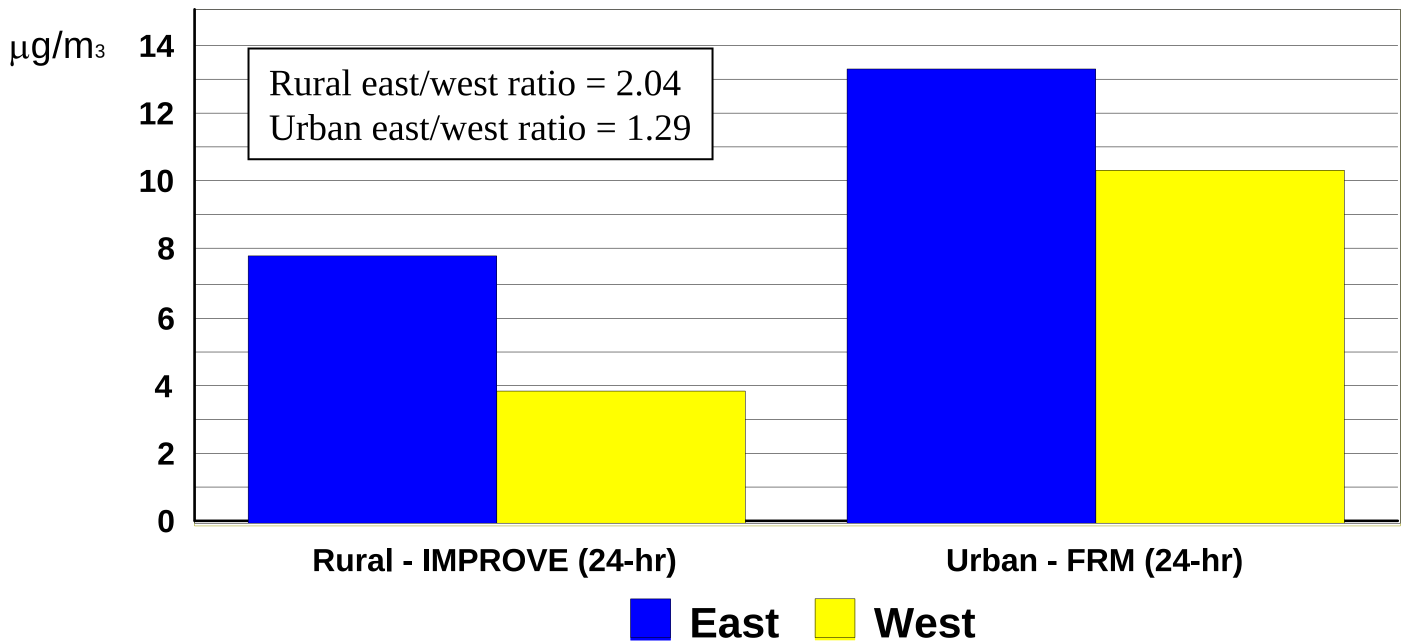
Outputs:
- o Line-plots, distinguishing between 'eastern' and western' counties, were generated for a 24-hour period and 12 p.m. - 4 p.m. block; see Attachment D.6a.
- o A table showing results of the second goal for the 12 p.m. - 4 p.m. block for 90^{th}, 92^{nd}, 95^{th}, and 98^{th} percentile metrics was produced; see Attachment D.6b.

Methods:
- Plots: All available continuous 2001-2003 $PM_{2.5}$ data were used. Data were aggregated to desired averaging time (by site-day). The county level maximum for each day (of the possible 1095 days) for the specified time-frame was identified. This maximum was compared to $PM_{2.5}$ levels from 5 to 65 $\mu g/m^3$ in 5 $\mu g/m^3$ increments. The results were plotted as a percentage of the total available

(monitored in the county) days. SAS code ('violating thresh bigdb.sas') for the data manipulation and plotting.

- Table: All available continuous 2001-2003 $PM_{2.5}$ data were considered. Data were aggregated by site-year-day for the 12 p.m. - 4 p.m. time frame. At least 3 of the 4 hourly values had to be present for the 4-hour average to be valid. If a site had 275 days (in a calendar year) of valid 4-hour averages, the site-year was valid. Annual 90^{th}, 92nd, 95^{th}, and 98^{th} percentile values were calculated. If a site had 2 or 3 valid years, then the 2 or 3 annual percentile values were averaged together. For each county with at least one valid site, the highest site level (1-, 2-, or 3-year) percentile values were identified. These county level maximum values (one value per county for 90^{th}, 92^{nd}, 95^{th}, and 98^{th} percentiles) were compared to possible NAAQS levels of 20, 25, and 30 $\mu g/m^3$. Results were tabulated as a percentage of counties violating the stated values (as compared to the total number of monitored counties) for U.S. and regional (SP) levels. Additionally, for the U.S. level, the percentage of the population in those violating counties (as compared to total monitored population) was calculated. SAS code ('table of estimated county violation secondary.sas') was used to generate the table inputs; the table was constructed in EXCEL.

Year 2003

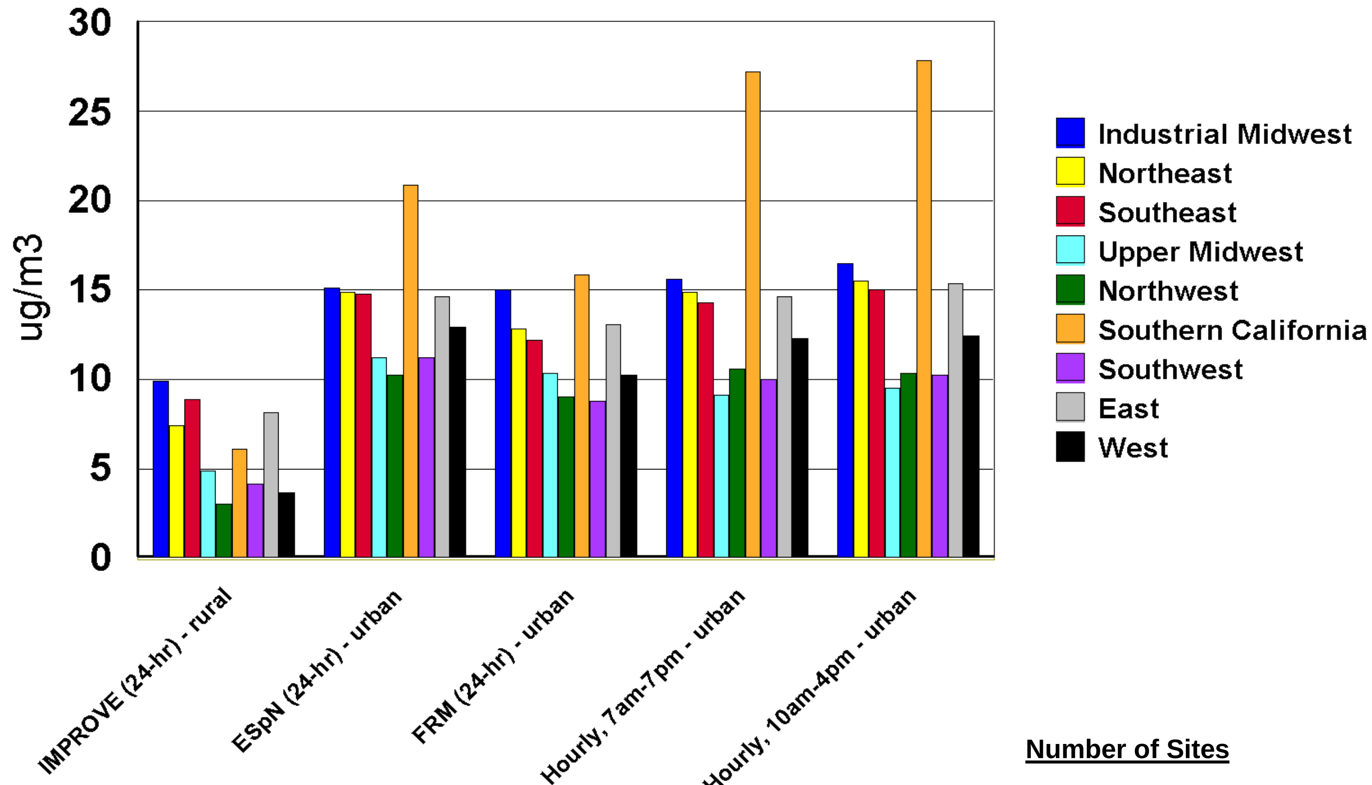


Note: Urban IMPROVE sites and rural FRM sites excluded.

$PM_{2.5}$ concentration differences between eastern and western areas and between rural and urban areas for 2003 (Jan 2003 - Dec 2003).

PM2.5 Concentrations, Annual Averages

(4/01/02 - 3/31/03) ('old' vis. db)



east/west ratio =	2.23 IMP	1.13 ESpn	1.27 FRM	1.19 Hrly 7-7	1.24 Hrly 10-4

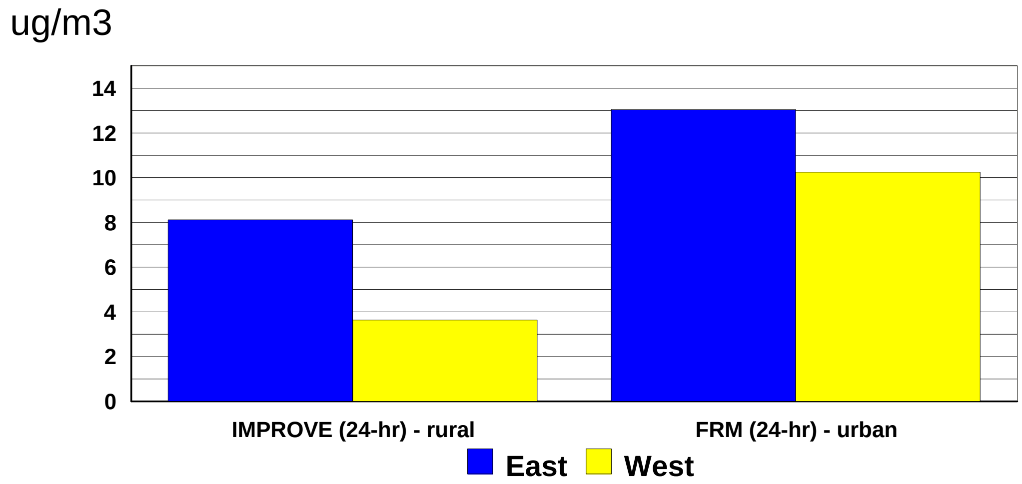
Larger differences in rural east/west mass than urban east/west mass!

Number of Sites

Region	IMPROVE	Speciation	FRM	Continuous
Industria	12	35	249	19
Northeast	15	26	191	25
Southeast	13	37	264	33
Upper Mid	6	8	79	2
Northwest	49	13	154	9
Southern	9	5	48	2
Southwest	22	2	56	6
East	46	106	783	79
West	80	20	258	17

East/West Rural/Urban differences in PM2.5 Mass

(4/01/02 - 3/31/03)



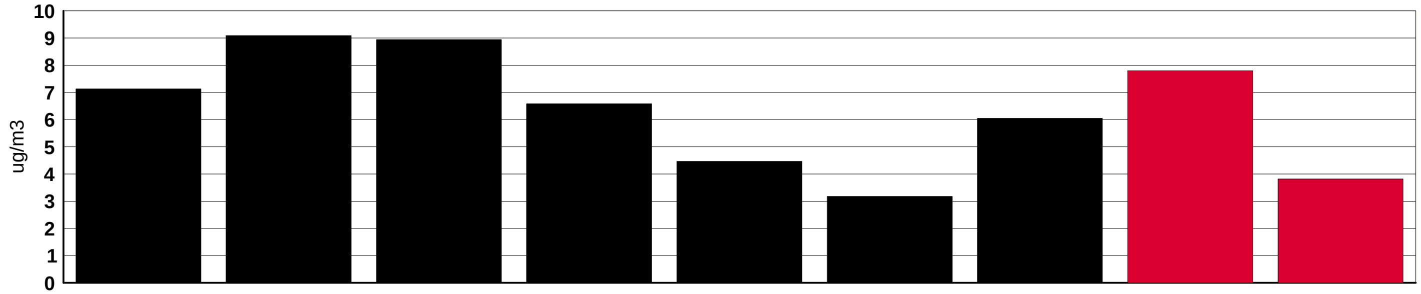
East/West urban PM2.5 concentrations more similar than rural east/west concentrations.

Rural east/west ratio = 2.23
Urban east/west ratio = 1.27

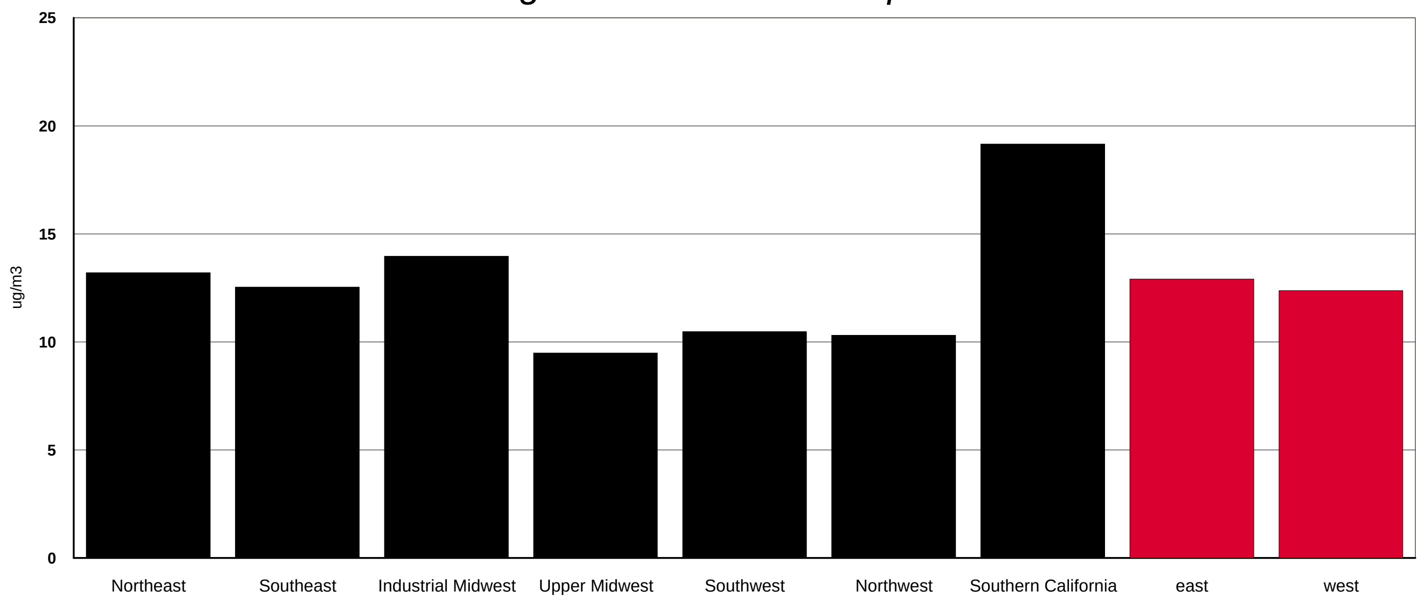
Note: Urban IMPROVE sites and rural FRM site excluded.

Urban East / West $PM_{2.5}$ Mass Levels are More Similar than Rural East / West Mass Levels

2003 Annual average PM2.5 at rural IMPROVE sites East/West=2.04

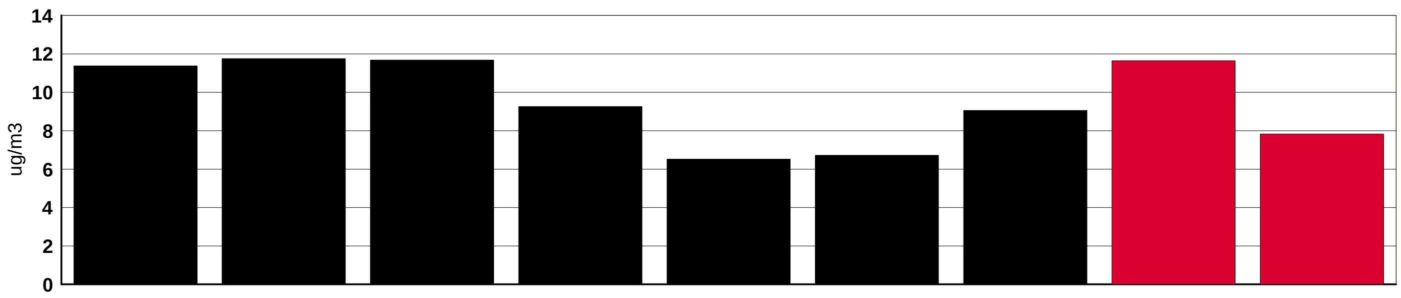


2003 Annual average PM2.5 at urban ESpN sites East/West=1.04

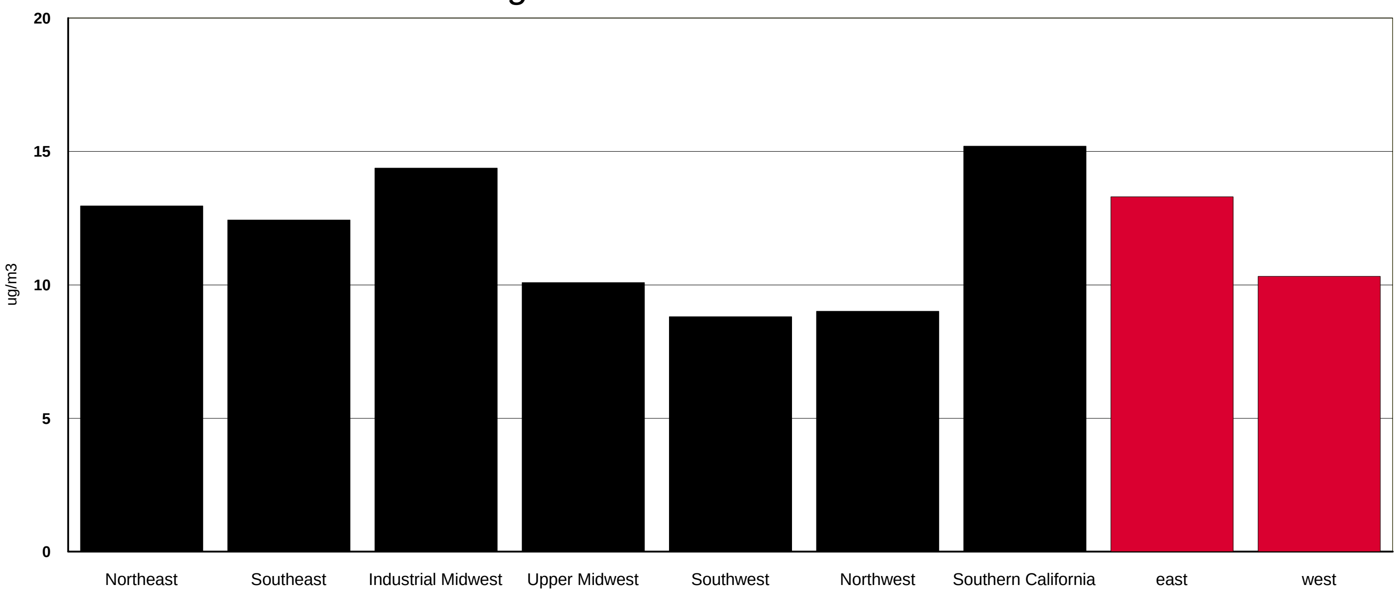


Urban East / West $PM_{2.5}$ Mass Levels are More Similar than Rural East / West Mass Levels

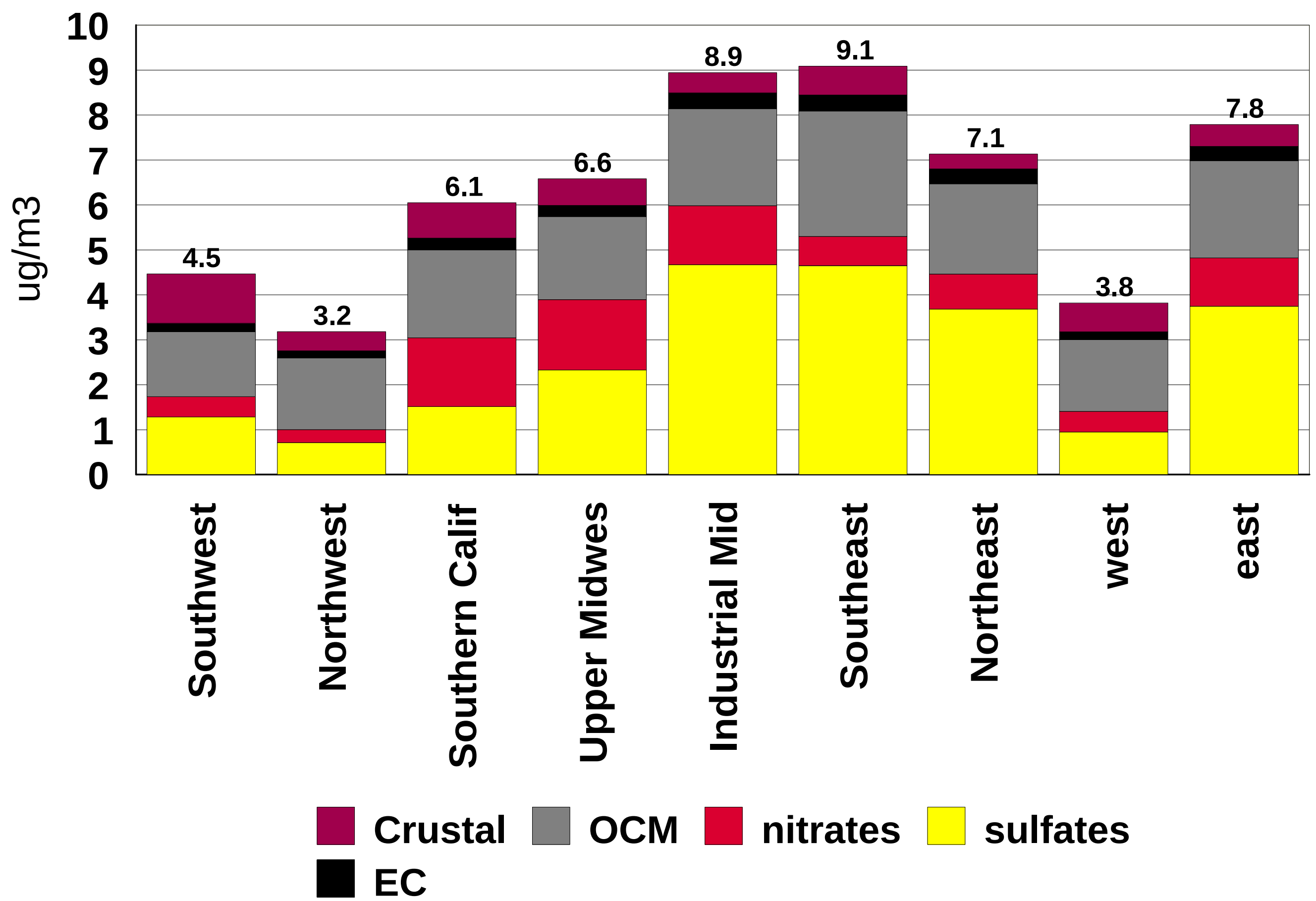
2003 Annual average PM2.5 at rural FRM sites East/West=1.48

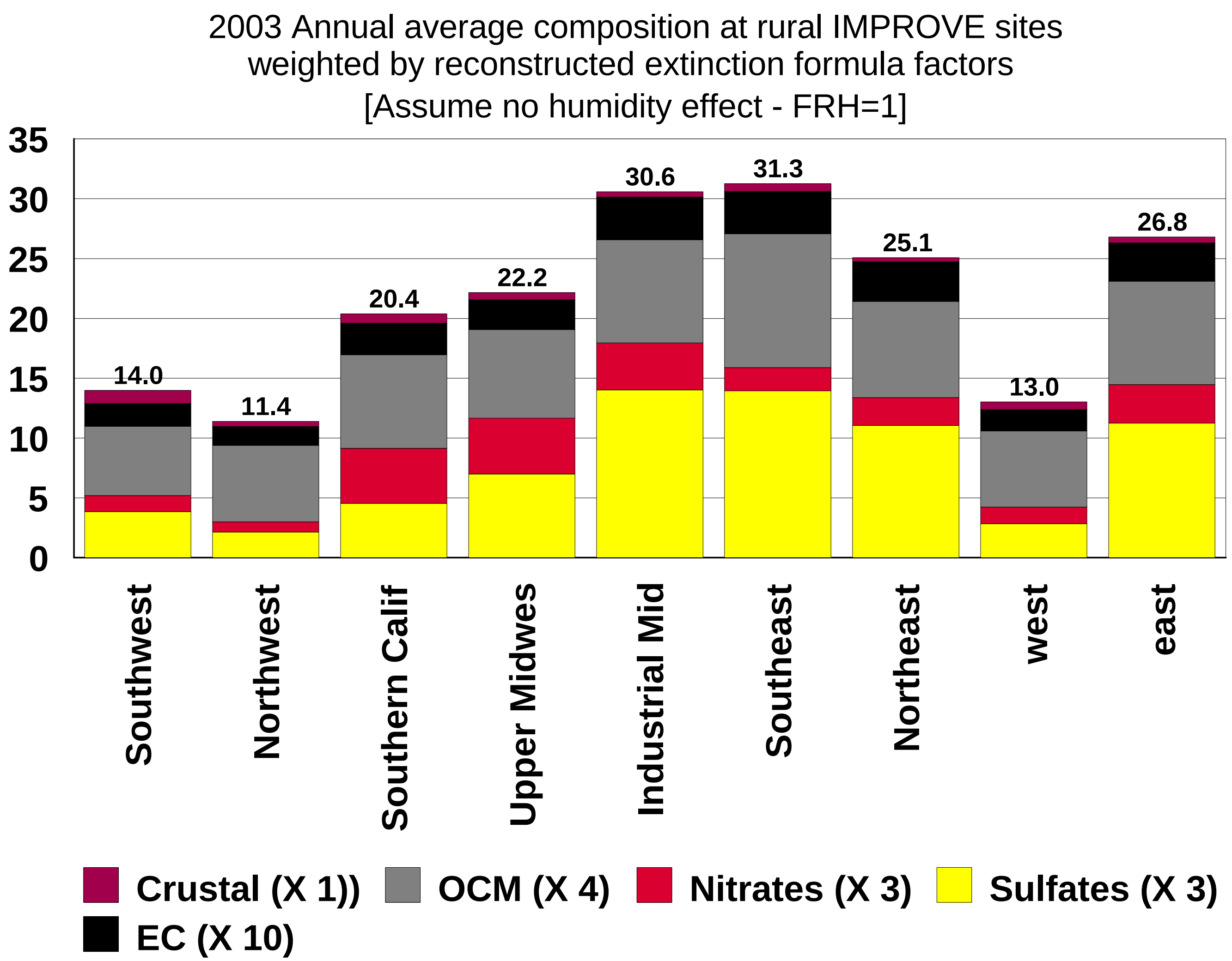


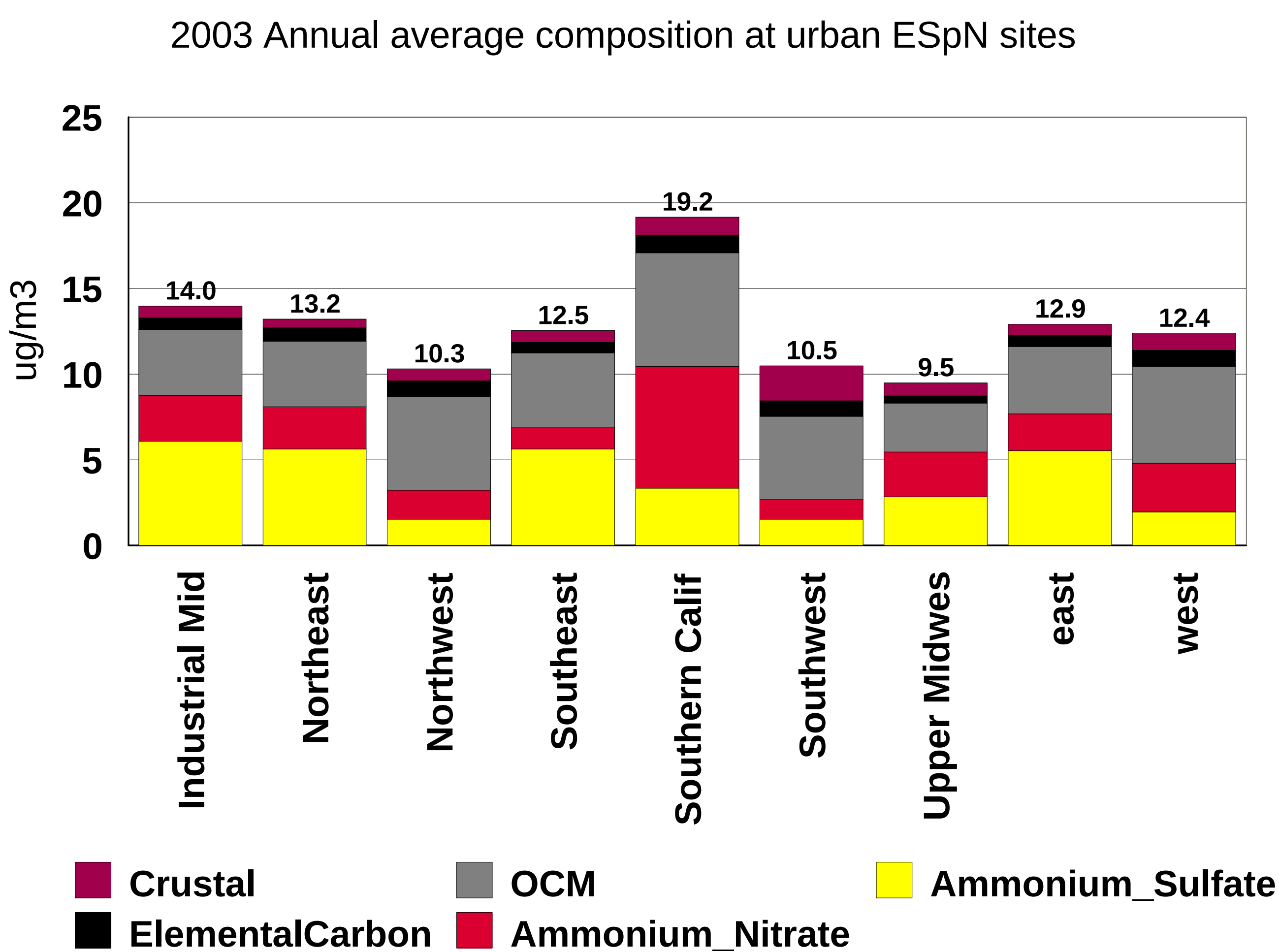
2003 Annual average PM2.5 at urban FRM sites East/West=1.29

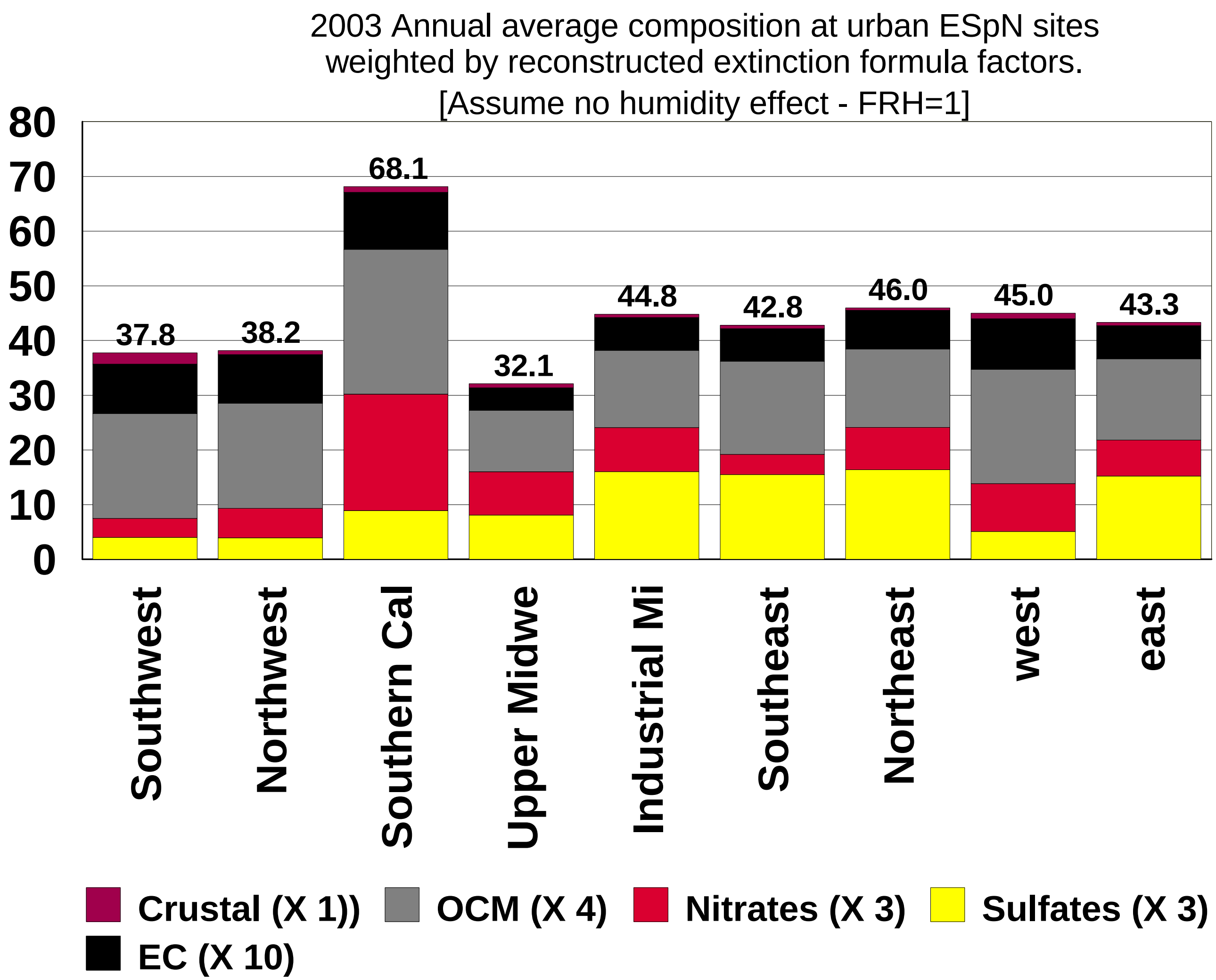


2003 Annual average composition at rural IMPROVE sites




2003 Annual average composition at rural IMPROVE sites
weighted by reconstructed extinction formula factors
[Assume no humidity effect - FRH=1]
35
30
25
20
15
10
5
0
14.0
11.4
20.4
22.2
30.6
31.3
25.1
13.0
26.8
Southwest
Northwest
Southern Calif
Upper Midwes
Industrial Mid
Southeast
Northeast
west
east
Crustal (X 1))
OCM (X 4)
Nitrates (X 3)
Sulfates (X 3)
EC (X 10)

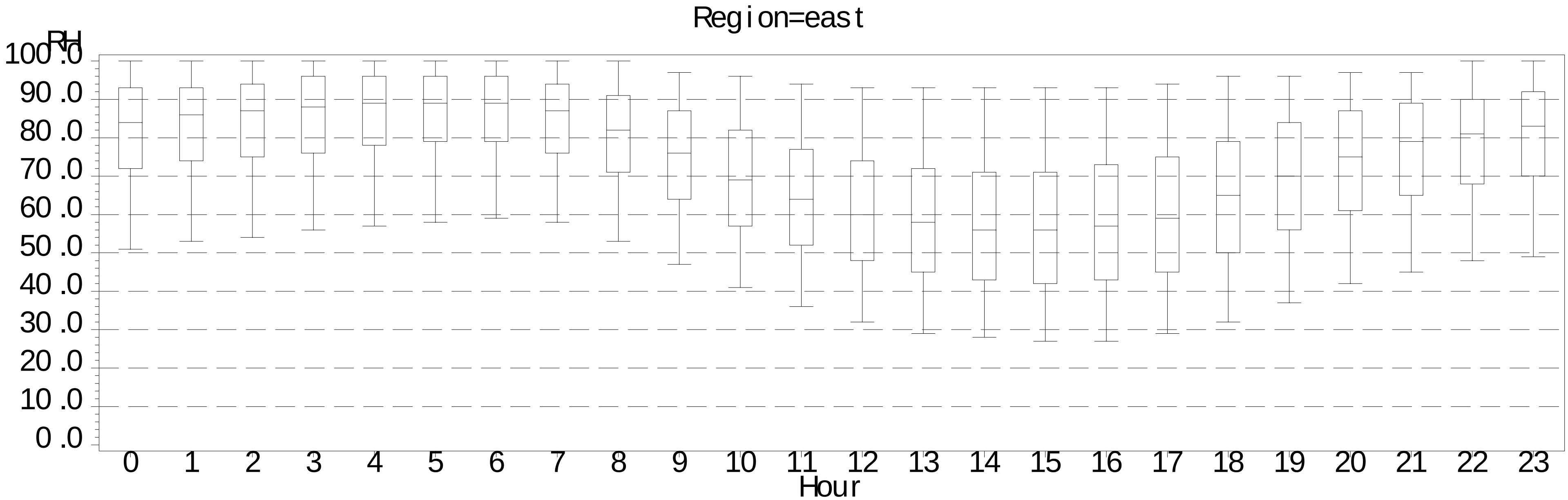

2003 Annual average composition at urban ESpN sites
ug/m3
25
20
15
10
5
0
14.0
13.2
10.3
12.5
19.2
10.5
9.5
12.9
12.4
Industrial Mid
Northeast
Northwest
Southeast
Southern Calif
Southwest
Upper Midwes
east
west
Crustal
OCM
Ammonium_Sulfate
ElementalCarbon
Ammonium_Nitrate

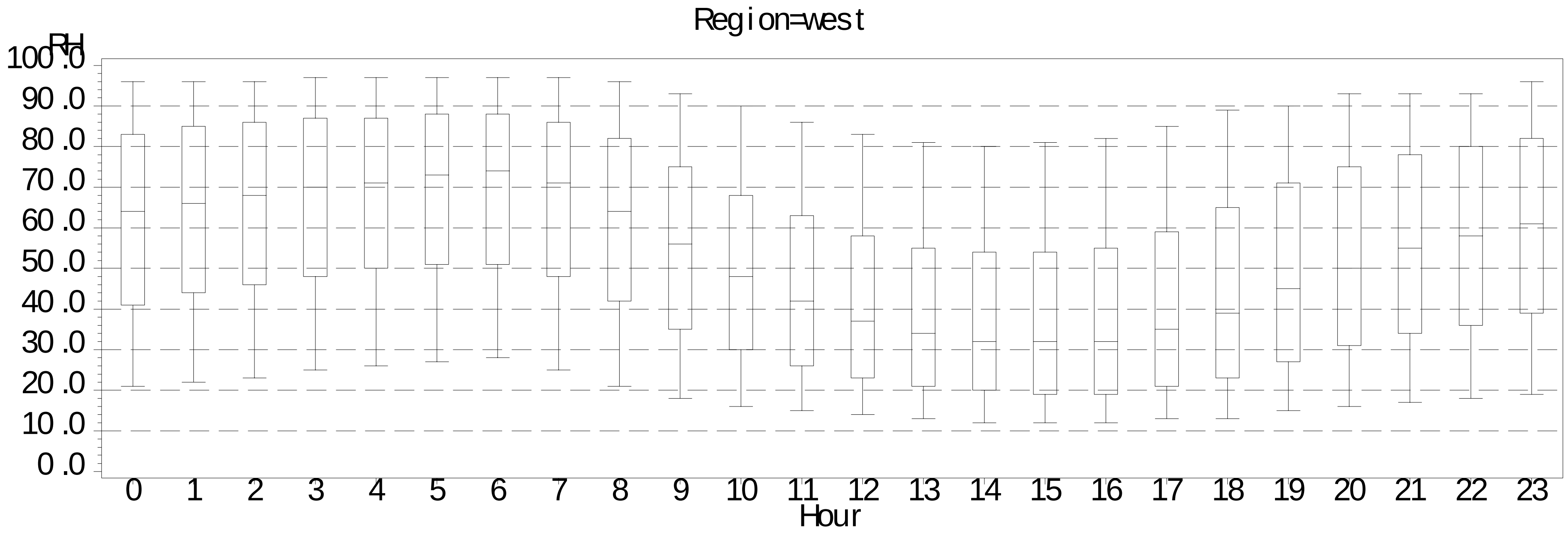

2003 Annual average composition at urban ESpN sites
weighted by reconstructed extinction formula factors.
[Assume no humidity effect - FRH=1]
80
70
60
50
40
30
20
10
0
37.8
38.2
68.1
32.1
44.8
42.8
46.0
45.0
43.3
Southwest
Northwest
Southern Cal
Upper Midwe
Industrial Mi
Southeast
Northeast
west
east
Crustal (X 1))
OCM (X 4)
Nitrates (X 3)
Sulfates (X 3)
EC (X 10)

Average 12-4pm PM2.5,PM2.5 profile, and PM10-2.5 in 2003 visibility db

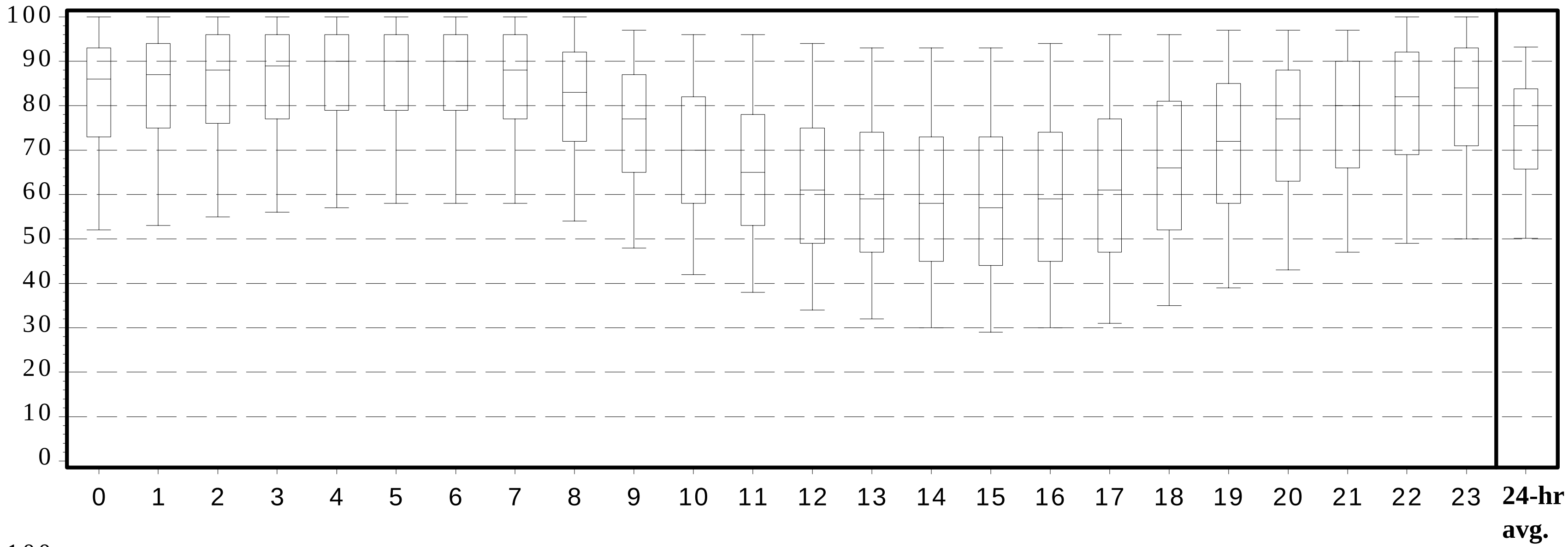
PM profile 12-4pm	Industrial Midwest	Northeast	Southeast	Upper Midwest	Northwest	Southern CA	Southwest
	Avg.	Avg.	Avg.	Avg.	Avg.	Avg.	Avg.
PM2.5	13.3	13.1	11.3	9.7	9.2	21.5	10.3
PM25_ocm	3.65	3.70	3.85	3.38	4.80	8.46	3.38
PM25_ec	0.67	0.90	0.53	0.45	0.67	0.93	0.61
PM25_nitrates	2.86	2.35	1.07	2.36	1.73	5.79	0.51
PM25_sulfates	5.46	5.65	5.02	2.83	1.36	2.86	1.61
PM25_crustal	0.63	0.52	0.84	0.66	0.65	3.42	4.18
PM10-2.5	8.0	6.9	7.2	10.6	10.3	31.3	44.0

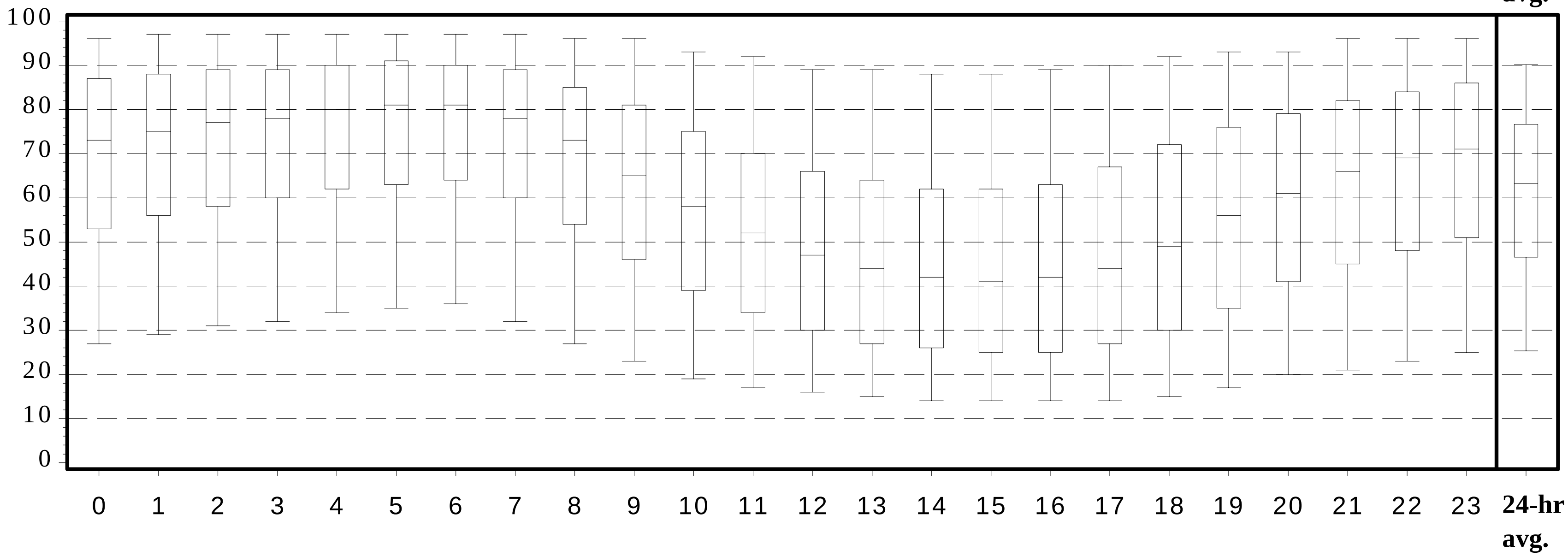
2003 RH at NWS sites, by region and hour



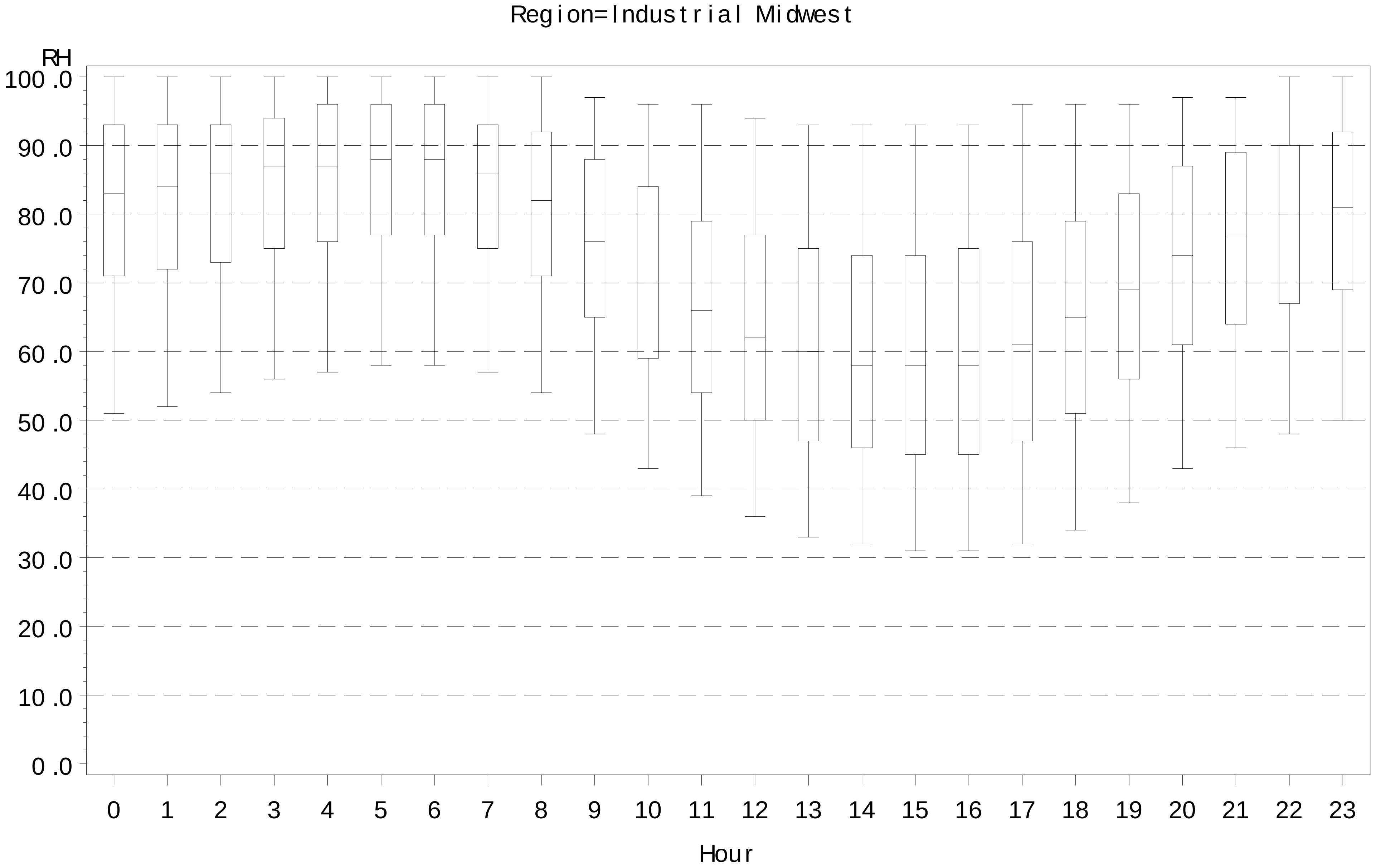


2003 RH at NWS sites, by region and hour - add 24-hr avg

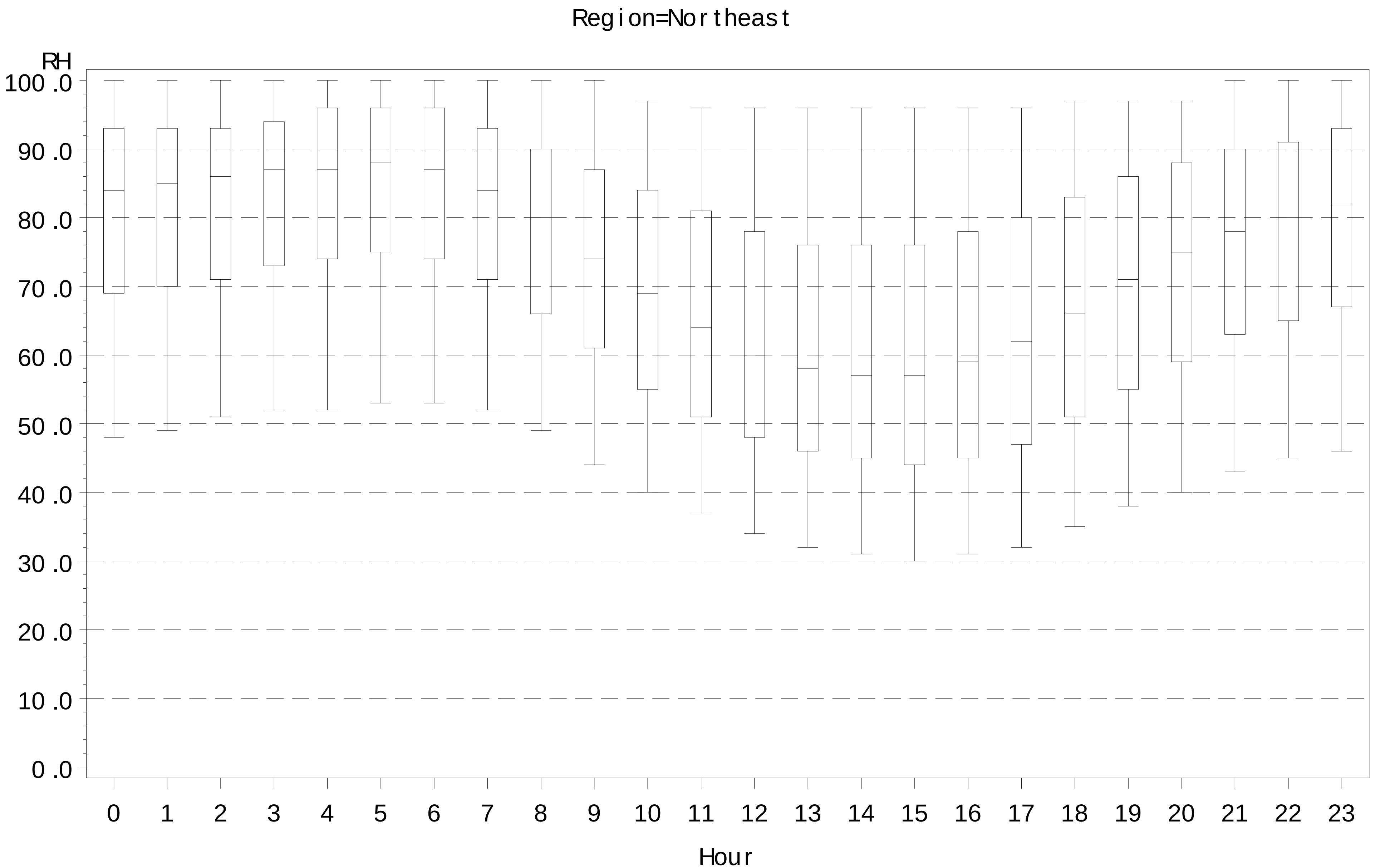




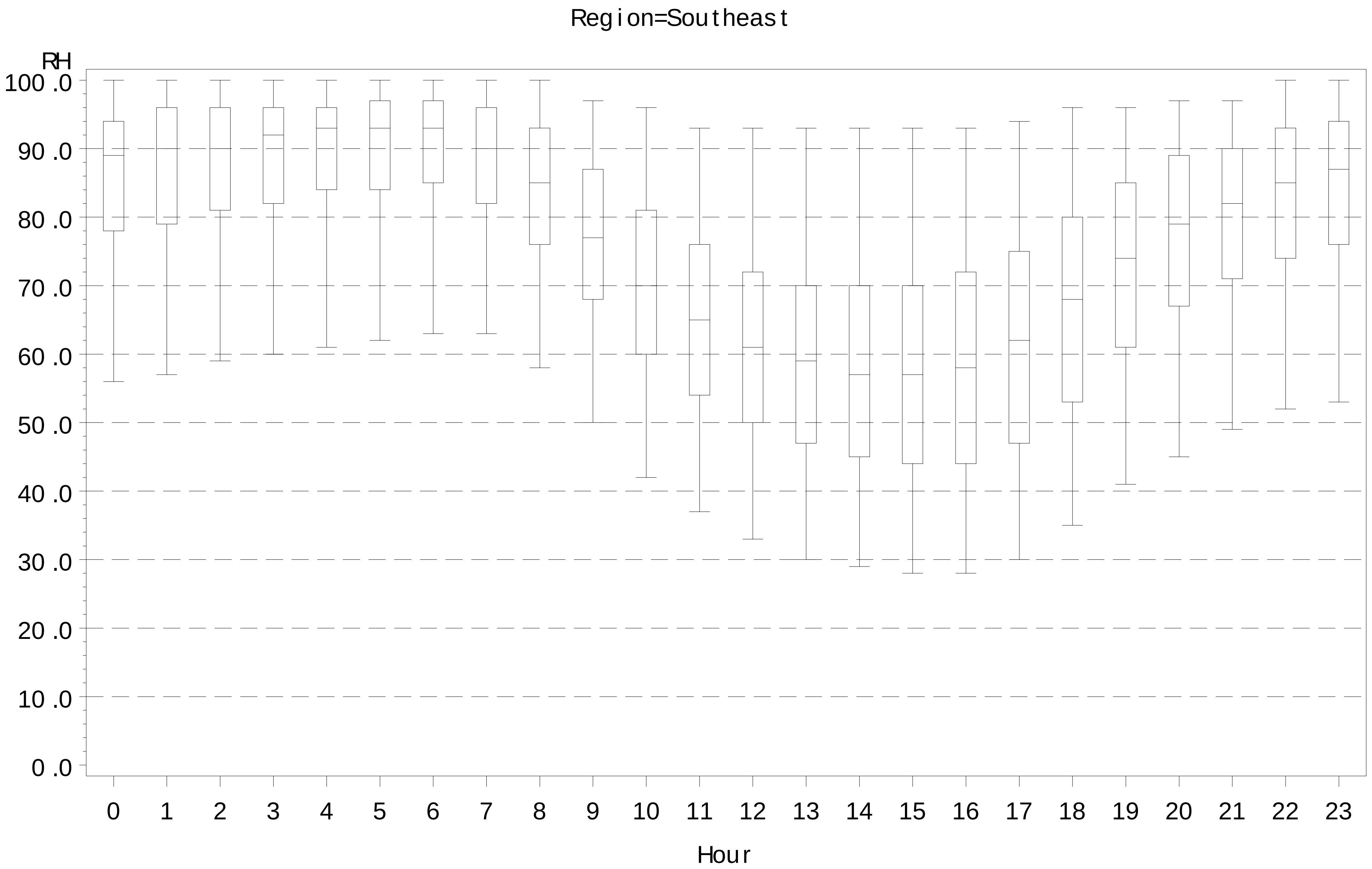
2003 RH at NWS sites, by region and hour



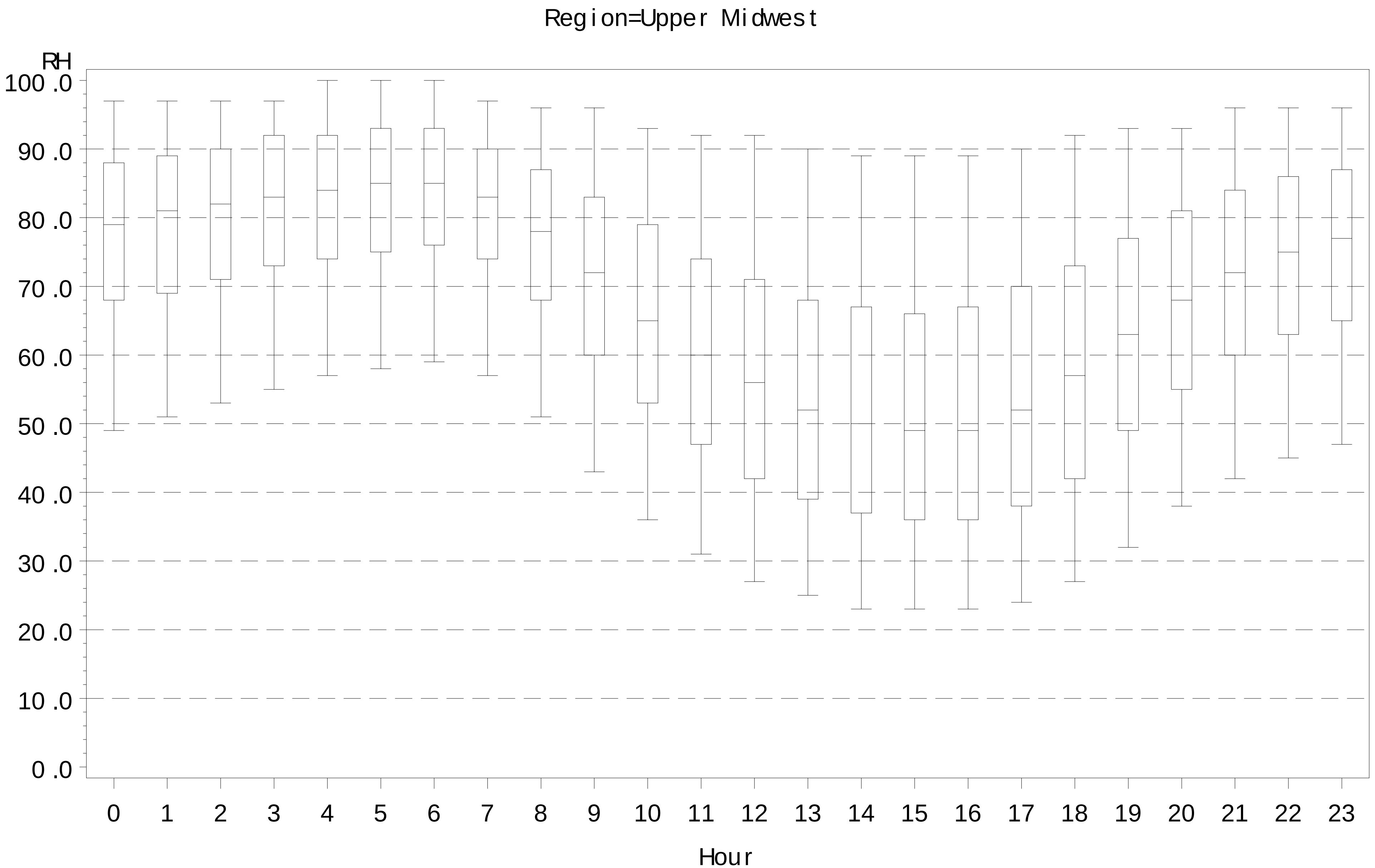
2003 RH at NWS sites, by region and hour



2003 RH at NWS sites, by region and hour

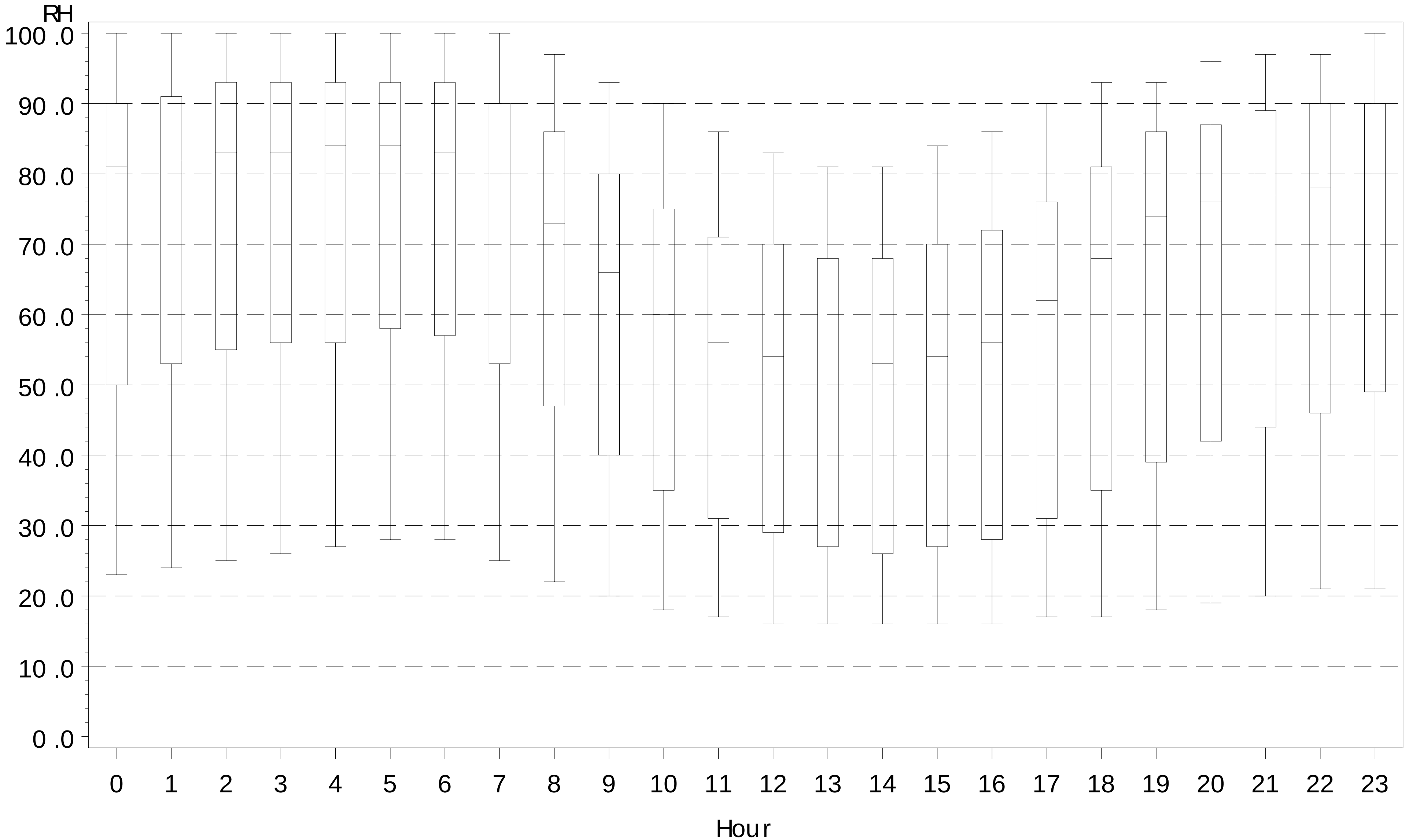


2003 RH at NWS sites, by region and hour

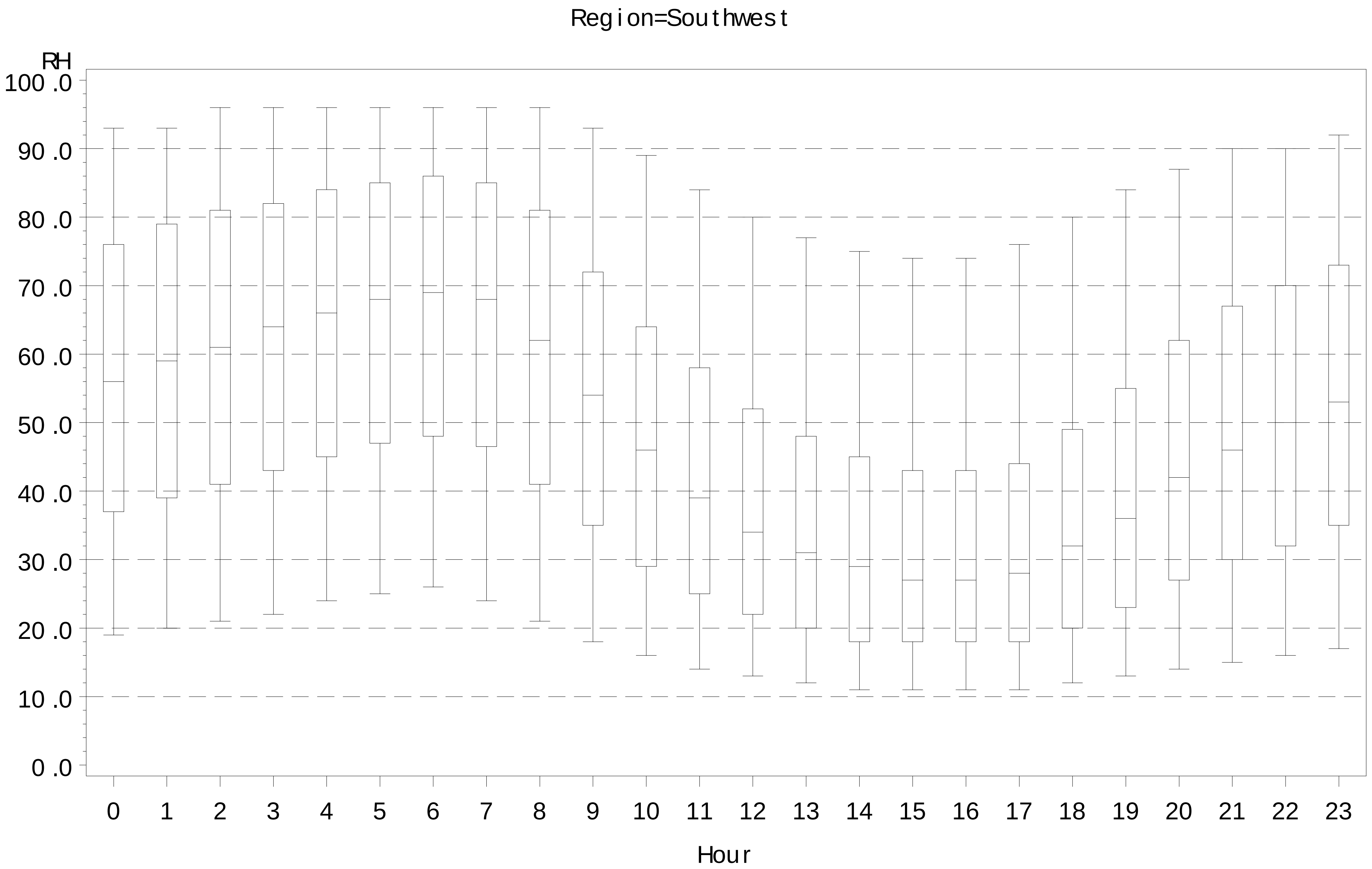


2003 RH at NWS sites, by region and hour

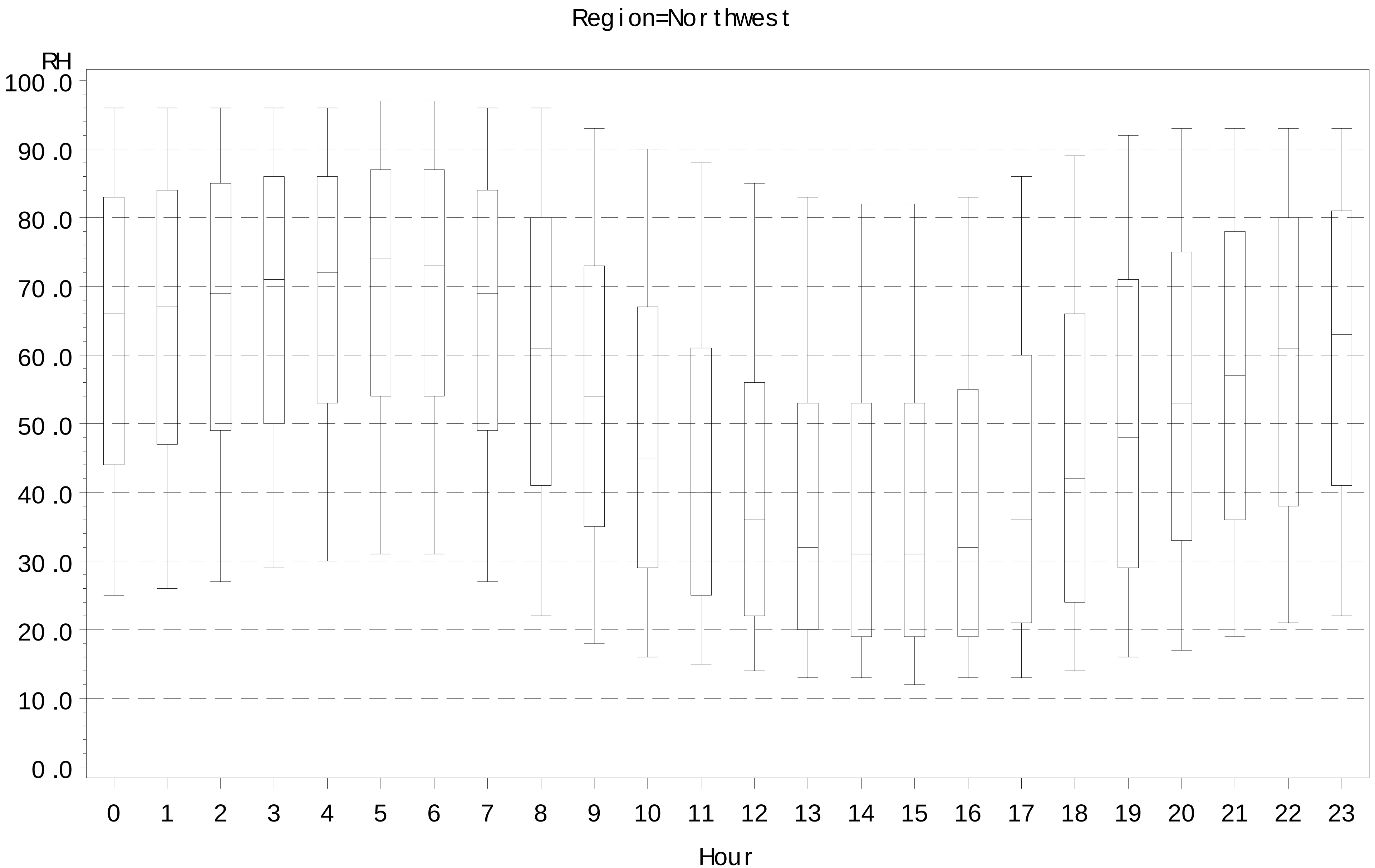
Region=Southern California



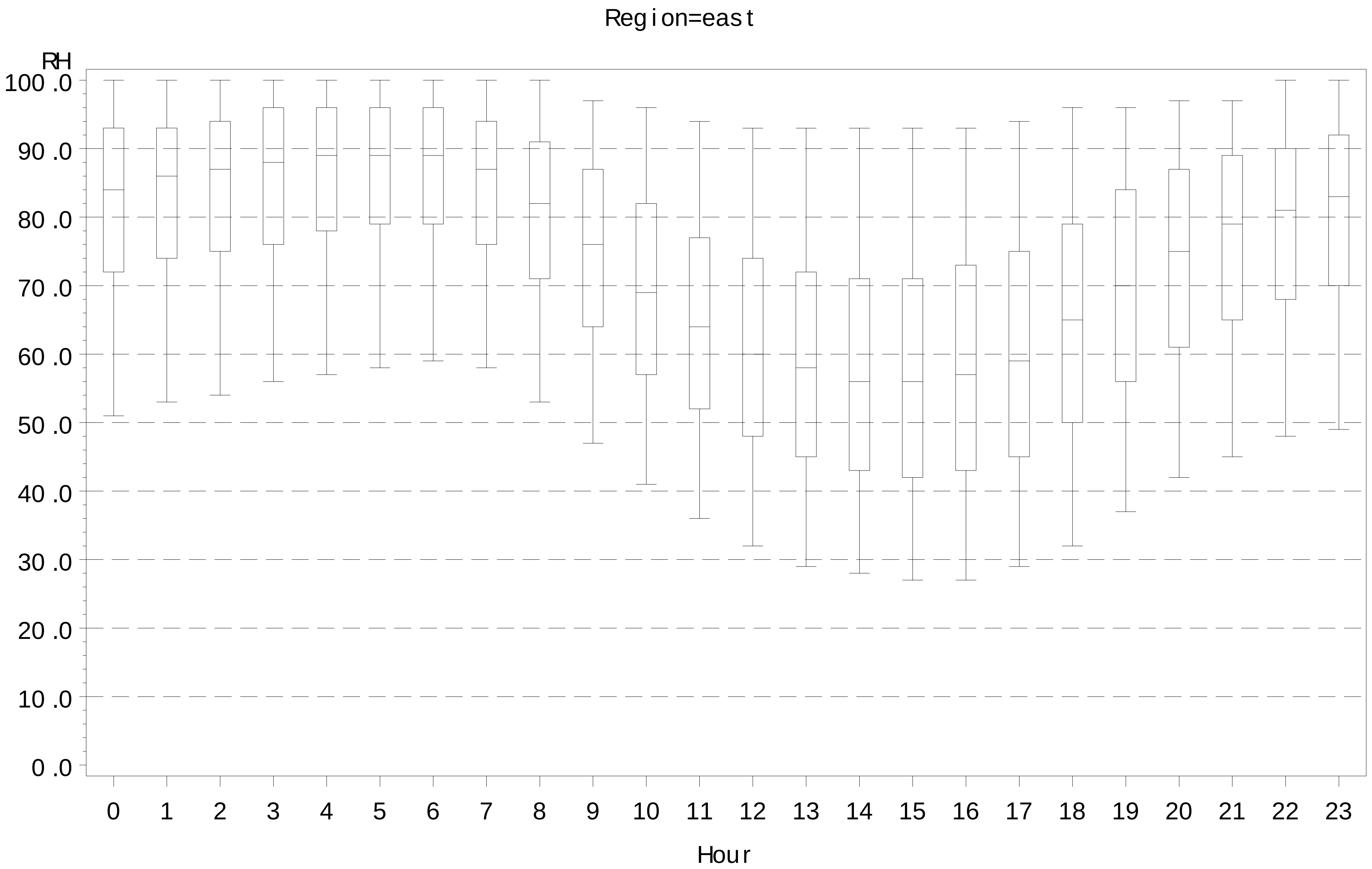
2003 RH at NWS sites, by region and hour



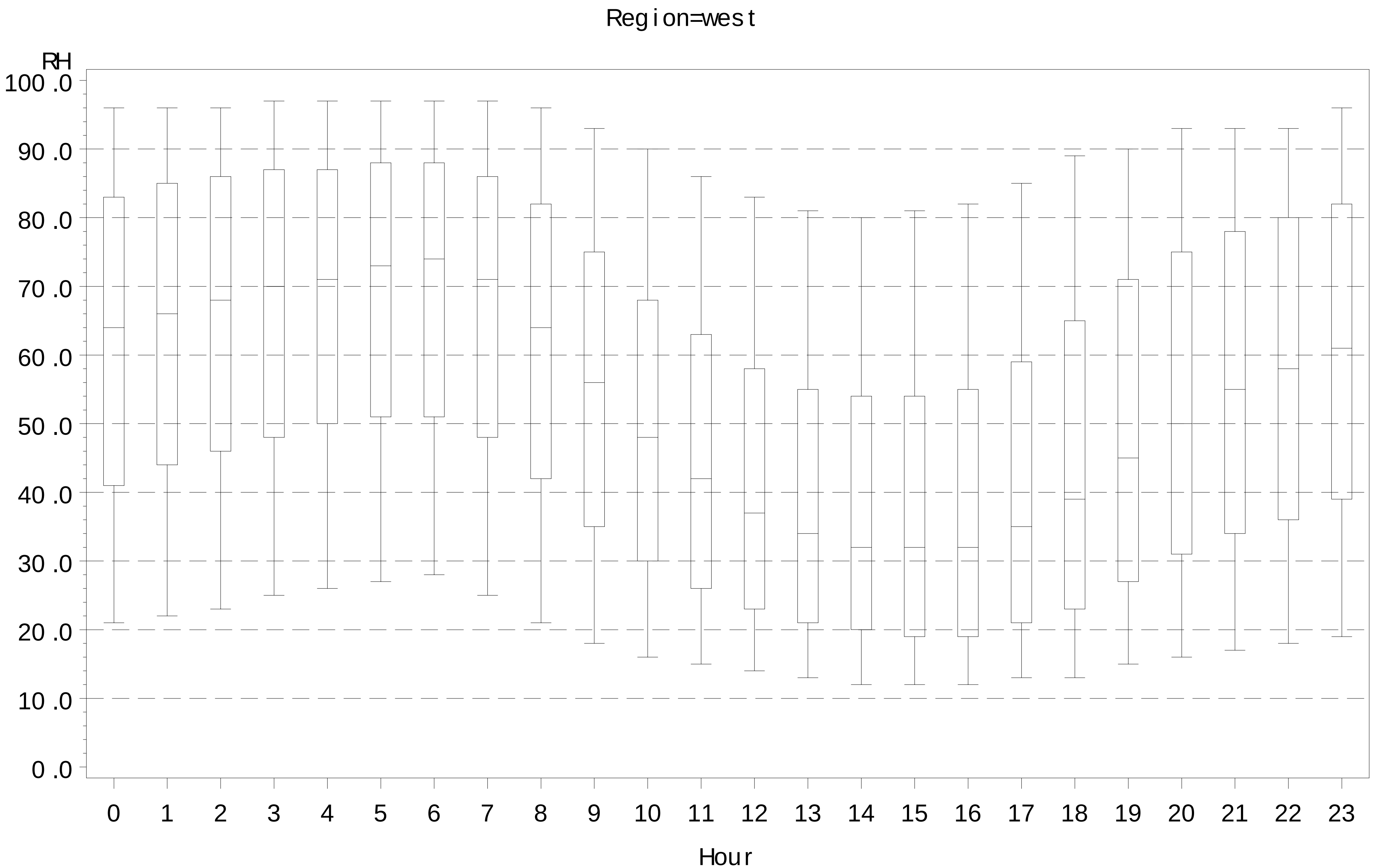
2003 RH at NWS sites, by region and hour



2003 RH at NWS sites, by region and hour

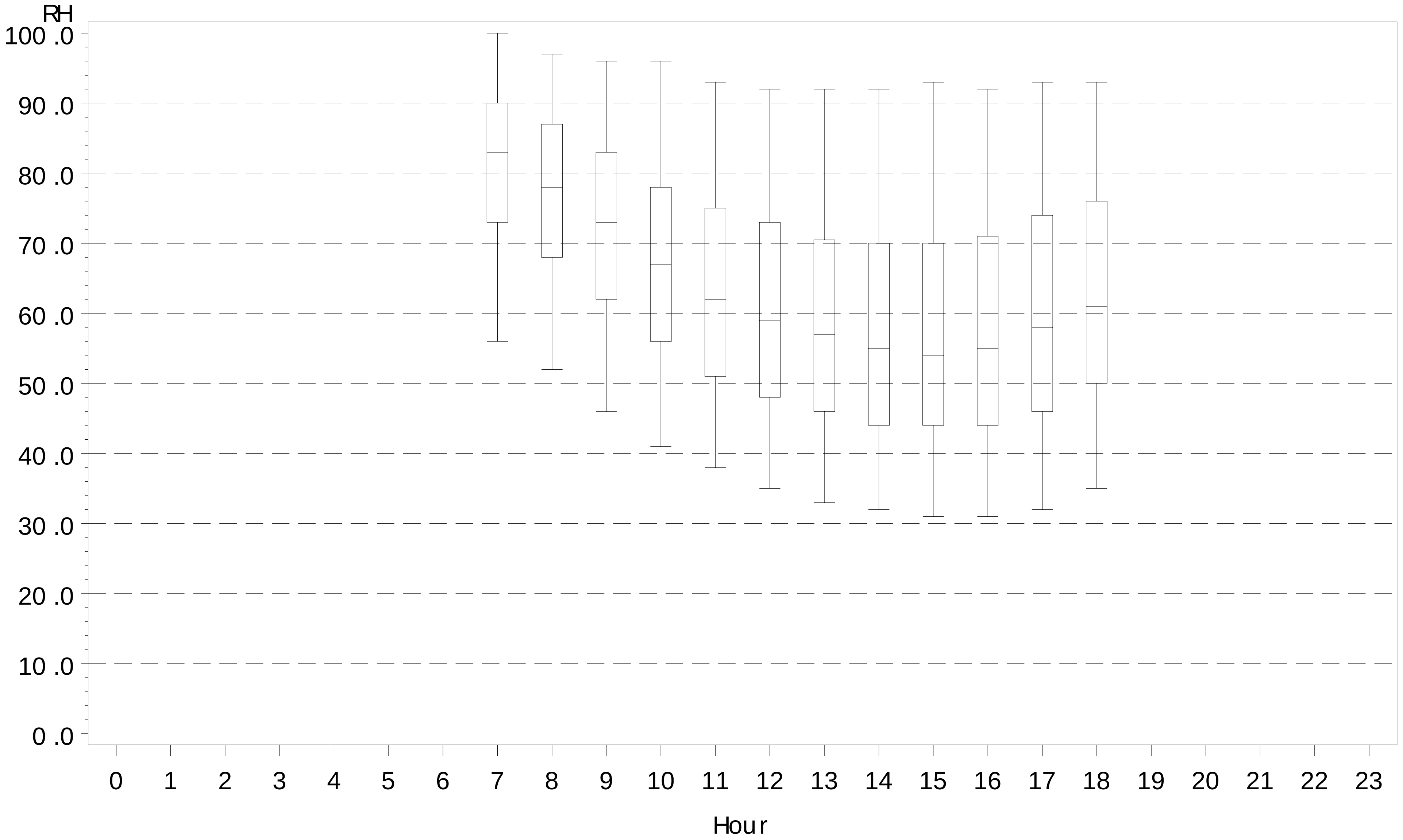


2003 RH at NWS sites, by region and hour



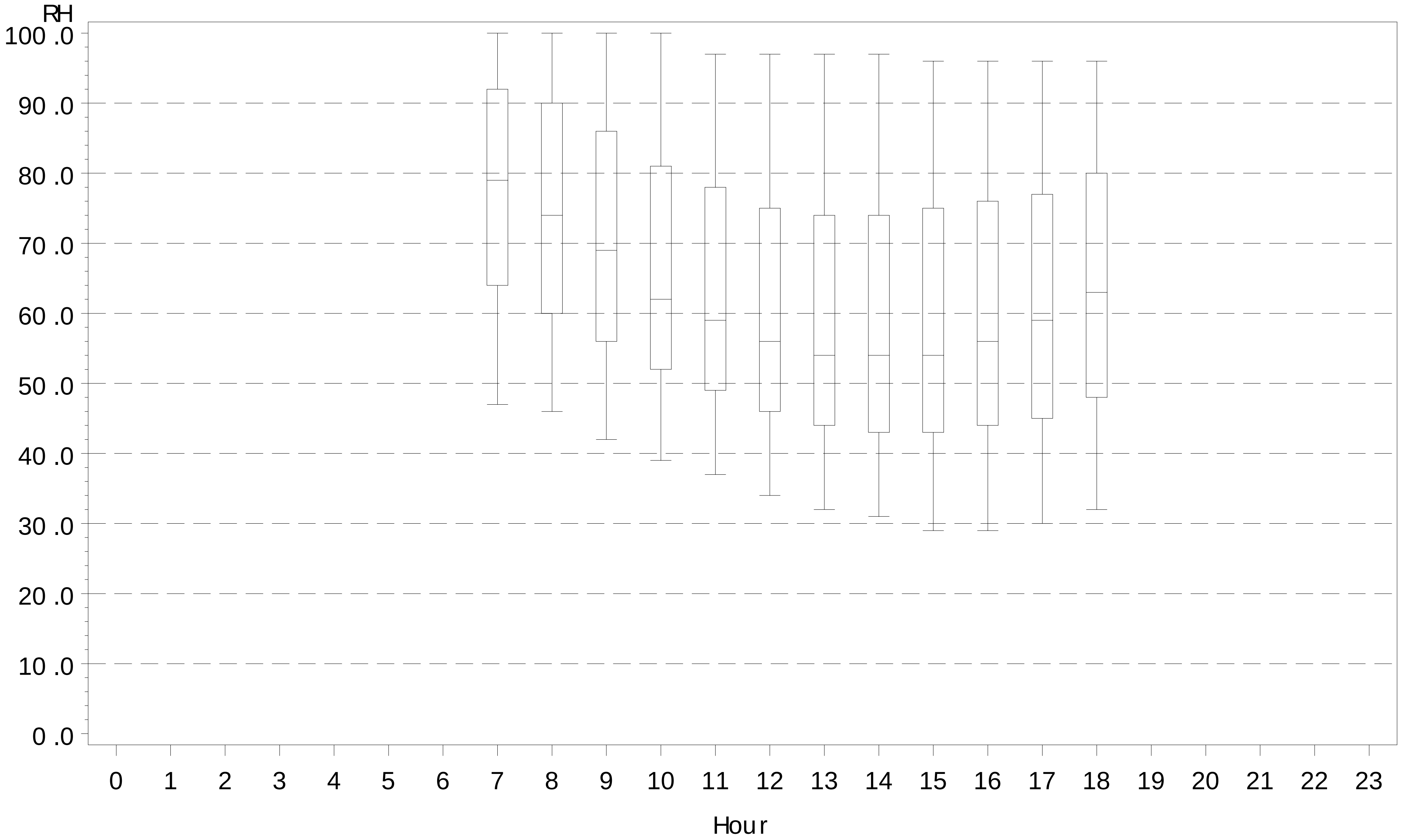
2003 RH at NWS sites for site-days used in visibility analyses (new db), by region and hour

region=Industrial Midwest



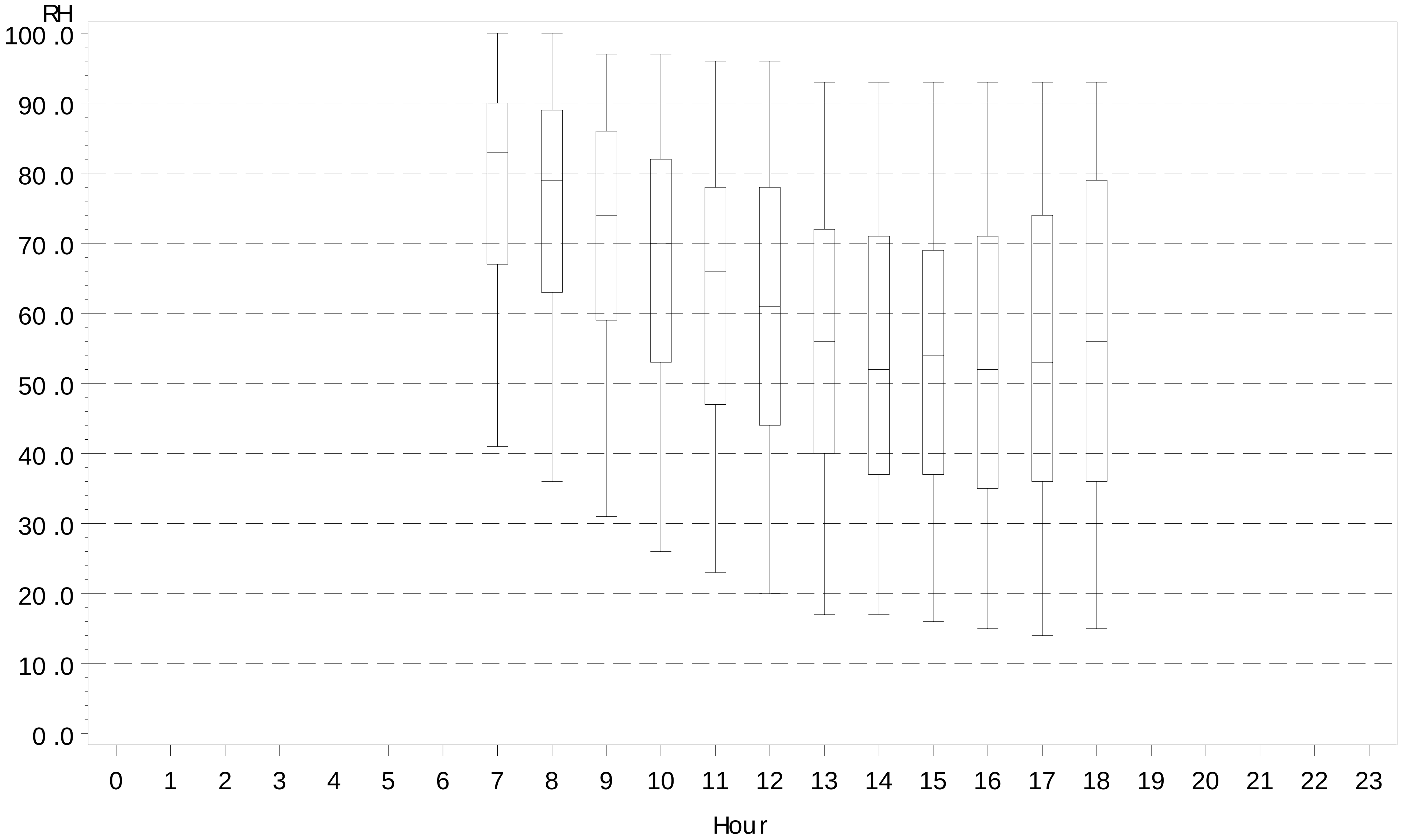
2003 RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Northeast



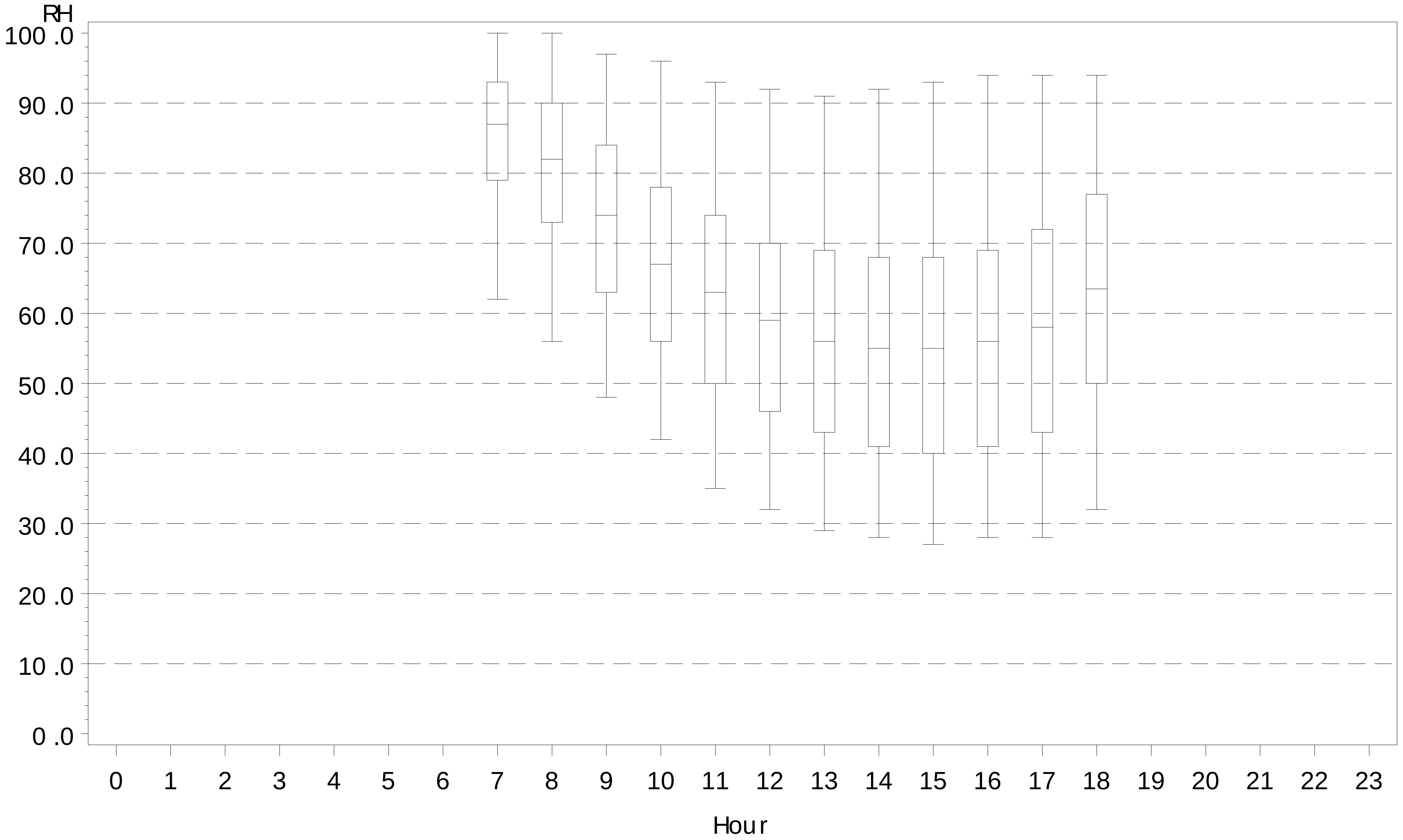
2003 RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Northwest

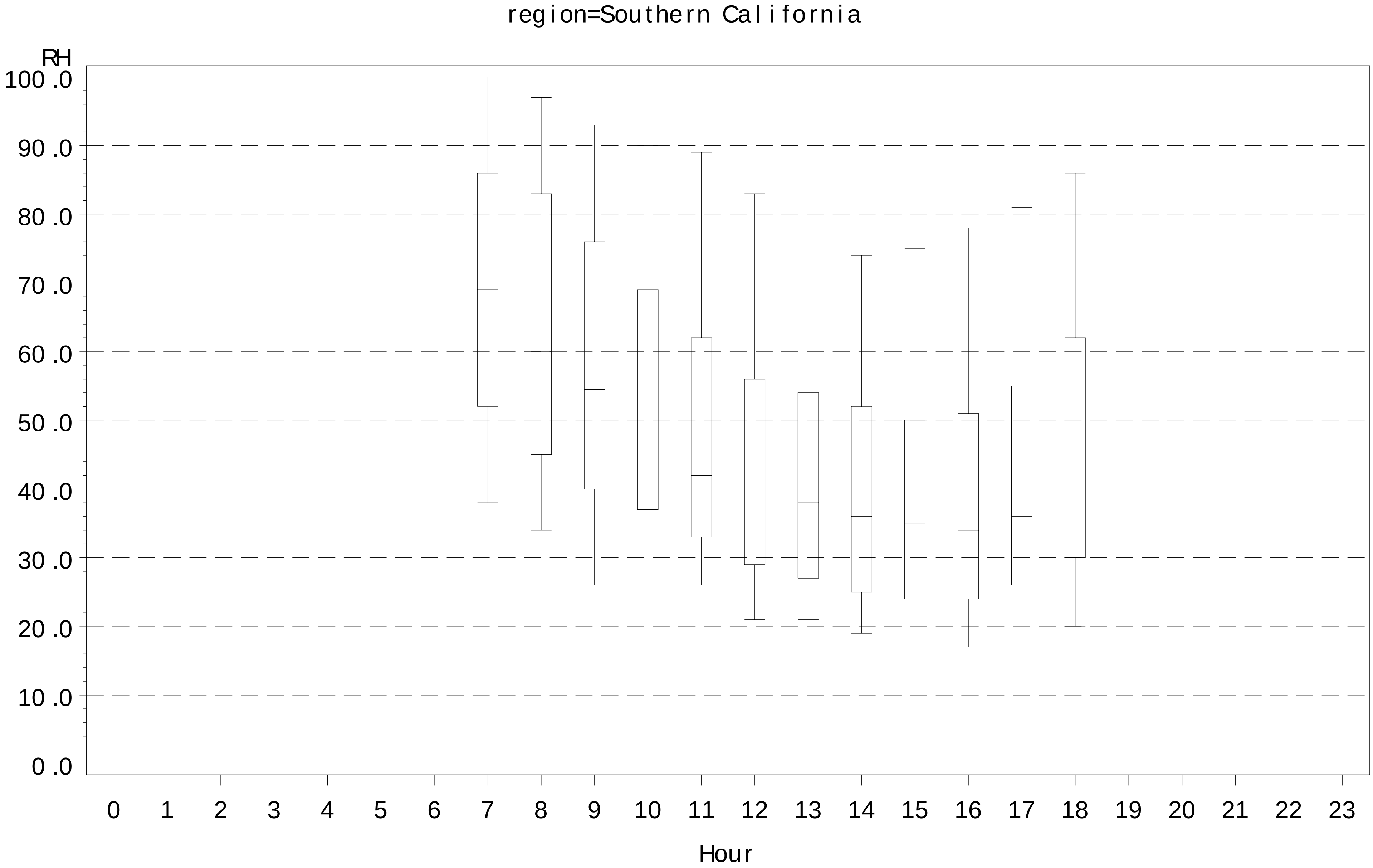


2003 RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Southeast

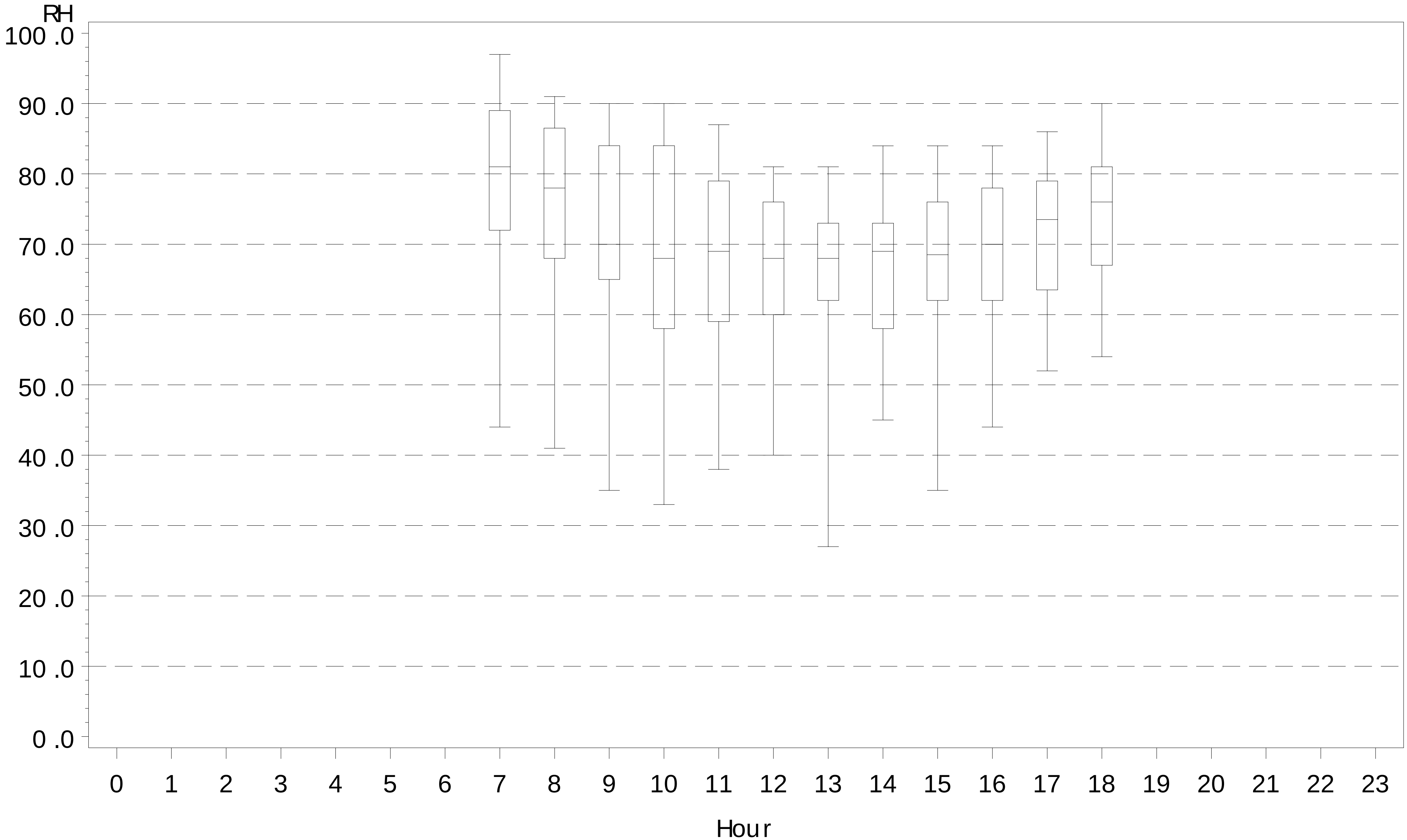


2003 RH at NWS sites for site-days used in visibility analyses, by region and hour



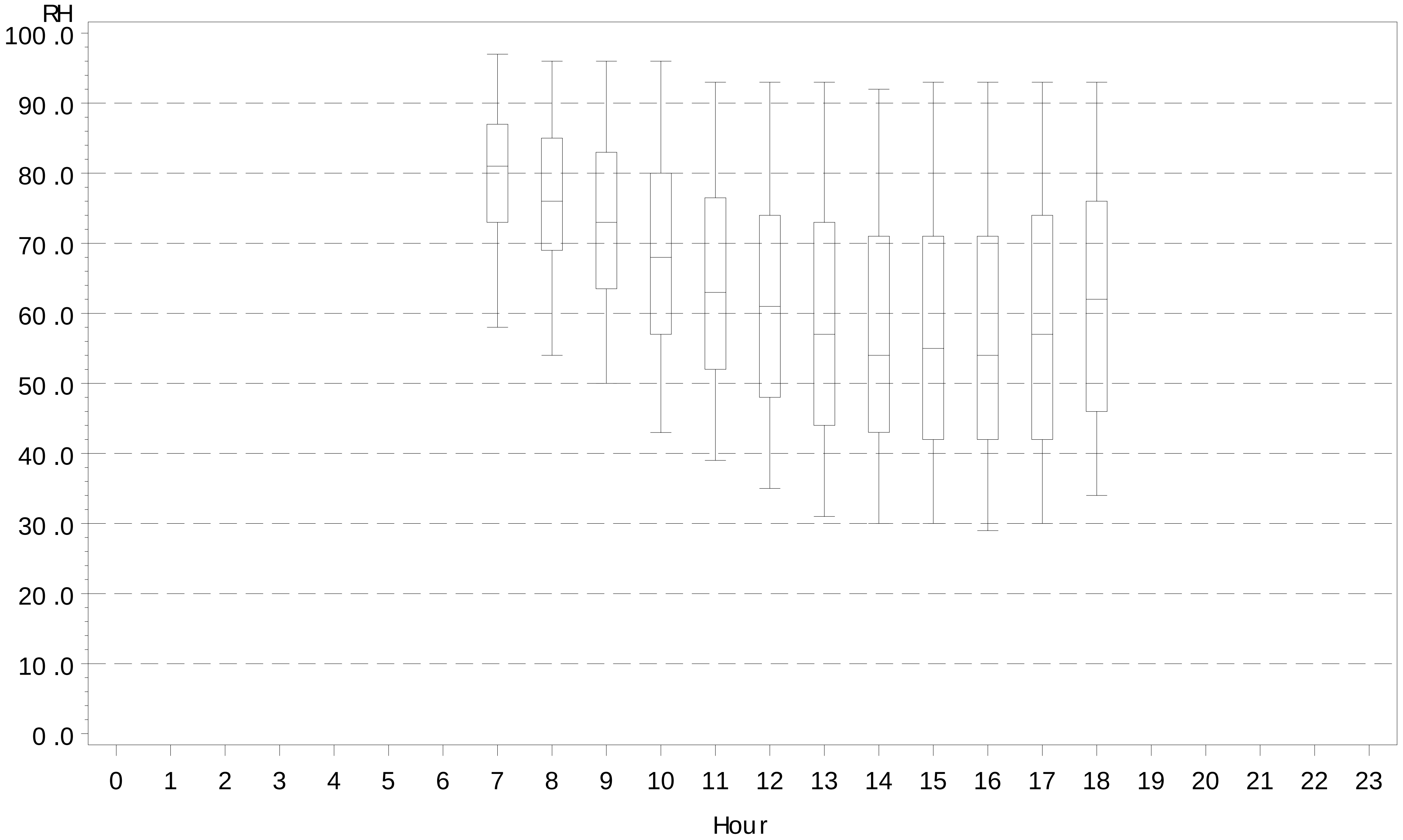
2003 RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Southwest



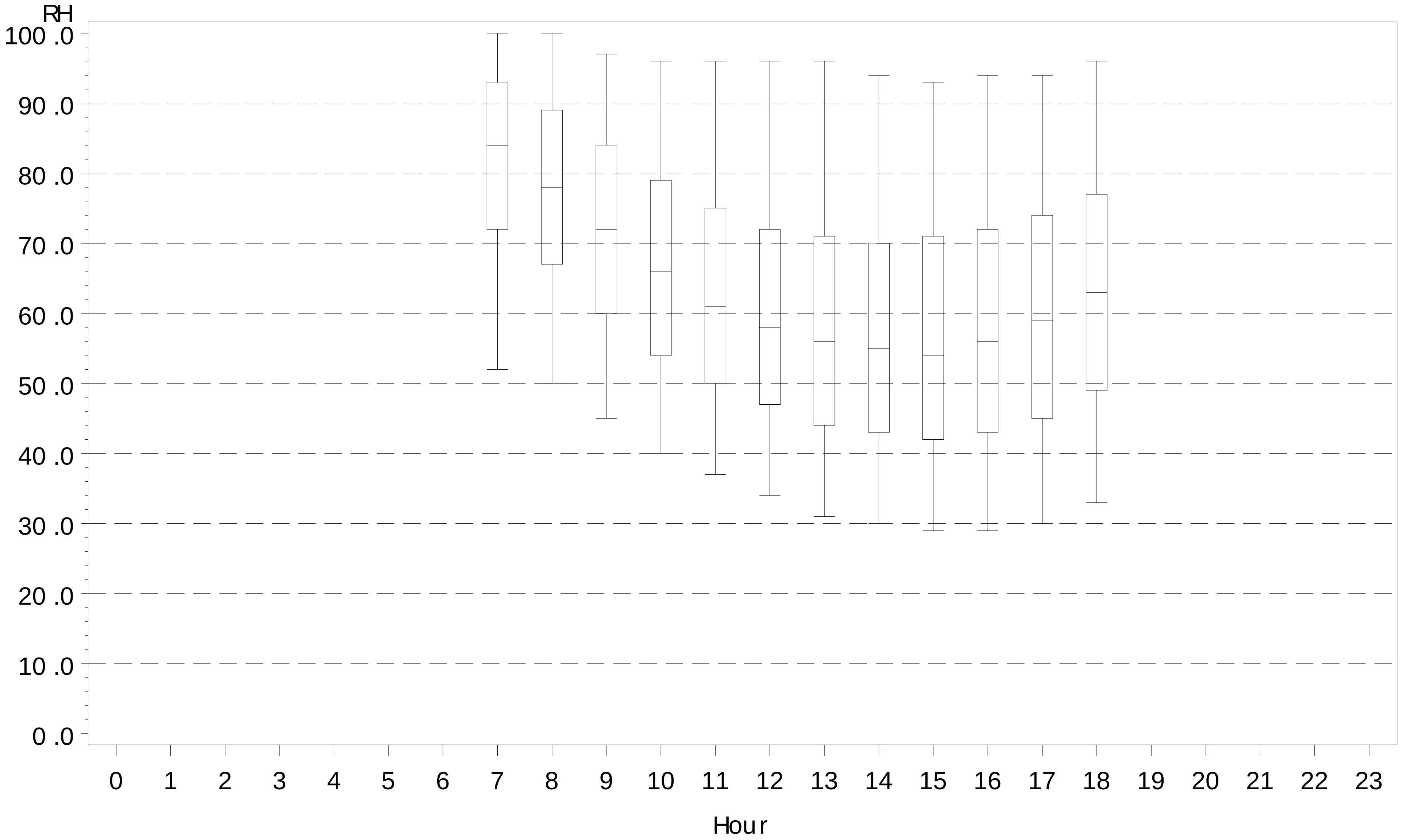
2003 RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Upper Midwest

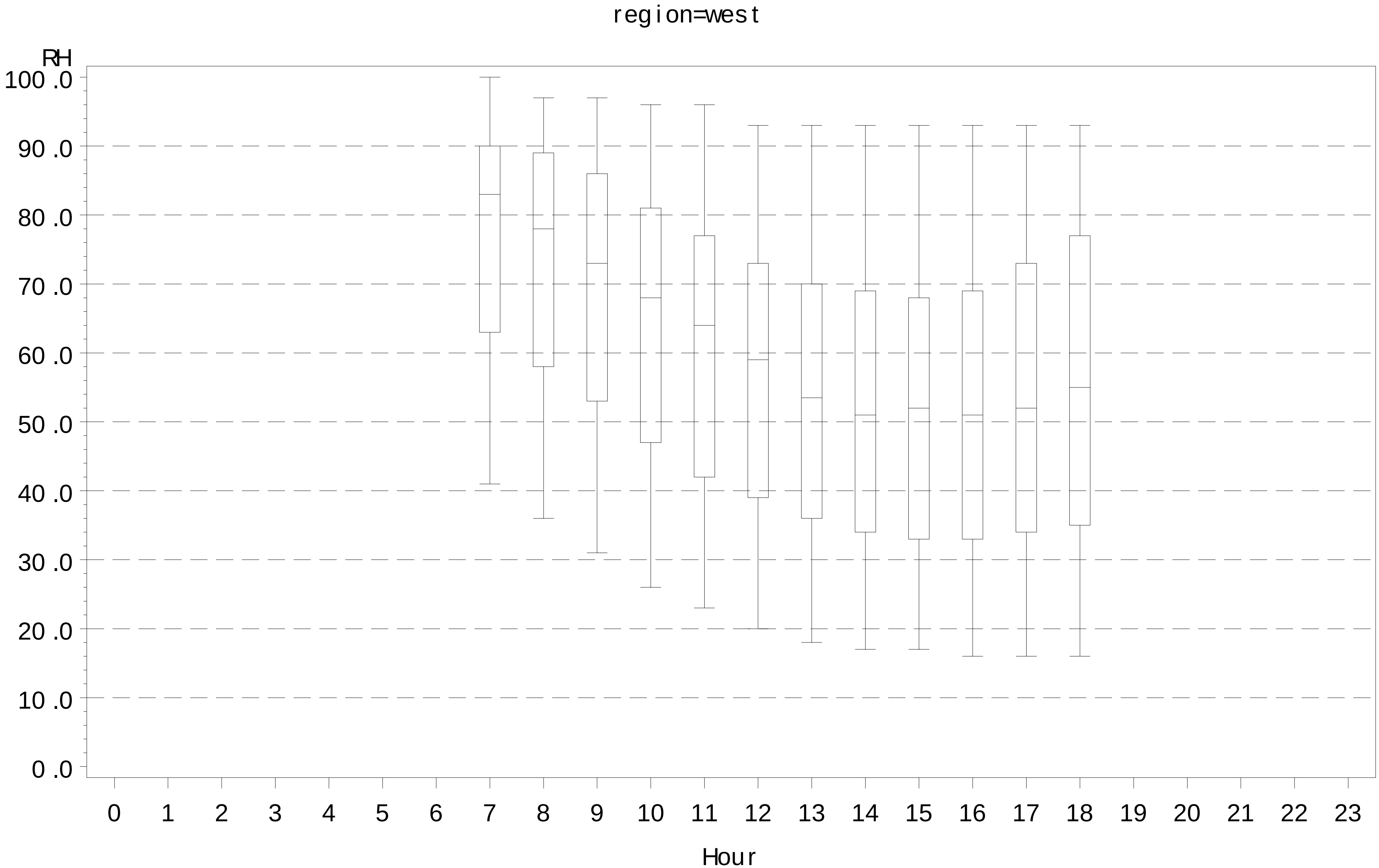


2003 RH at NWS sites for site-days used in visibility analyses, by region and hour

region=east

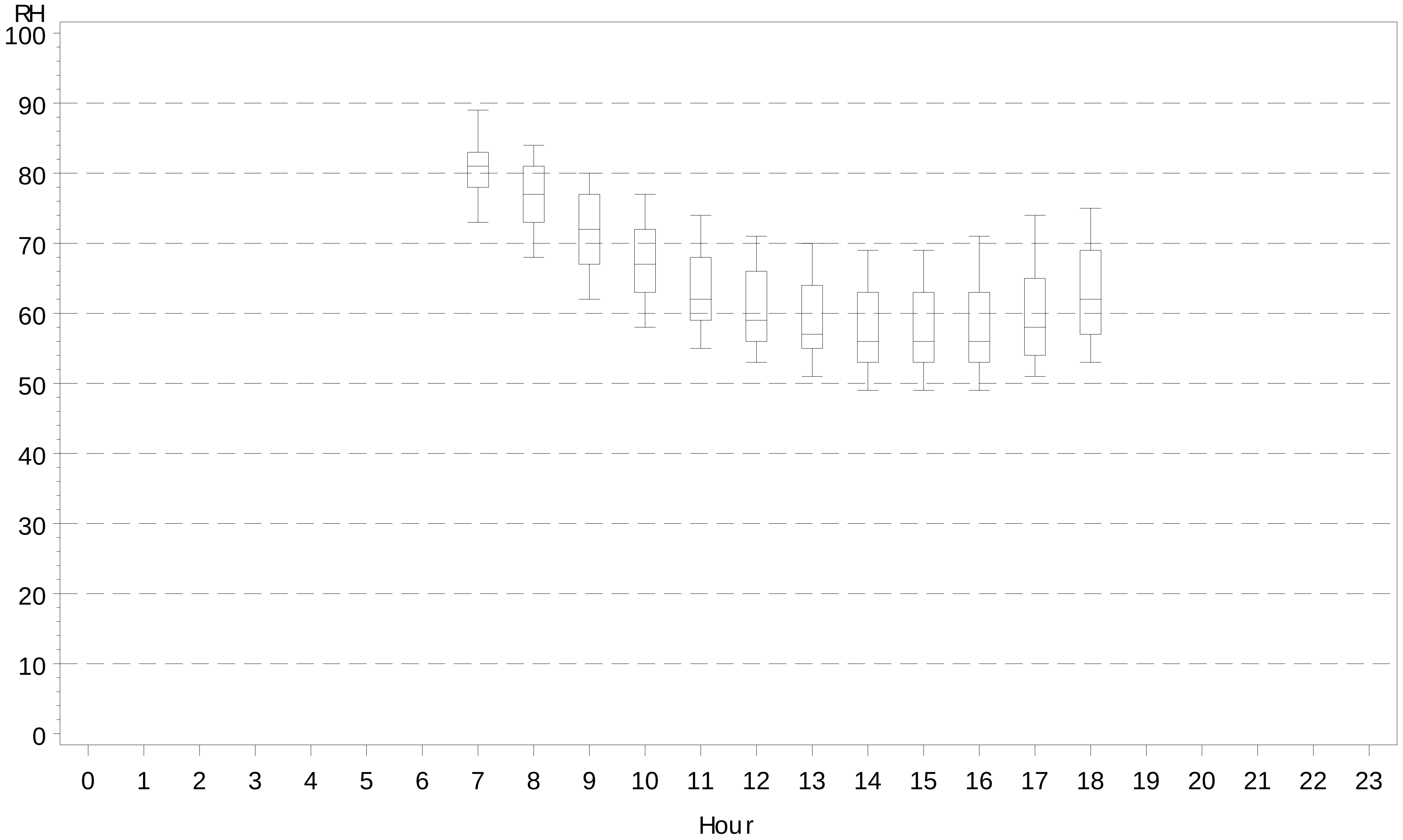


2003 RH at NWS sites for site-days used in visibility analyses, by region and hour



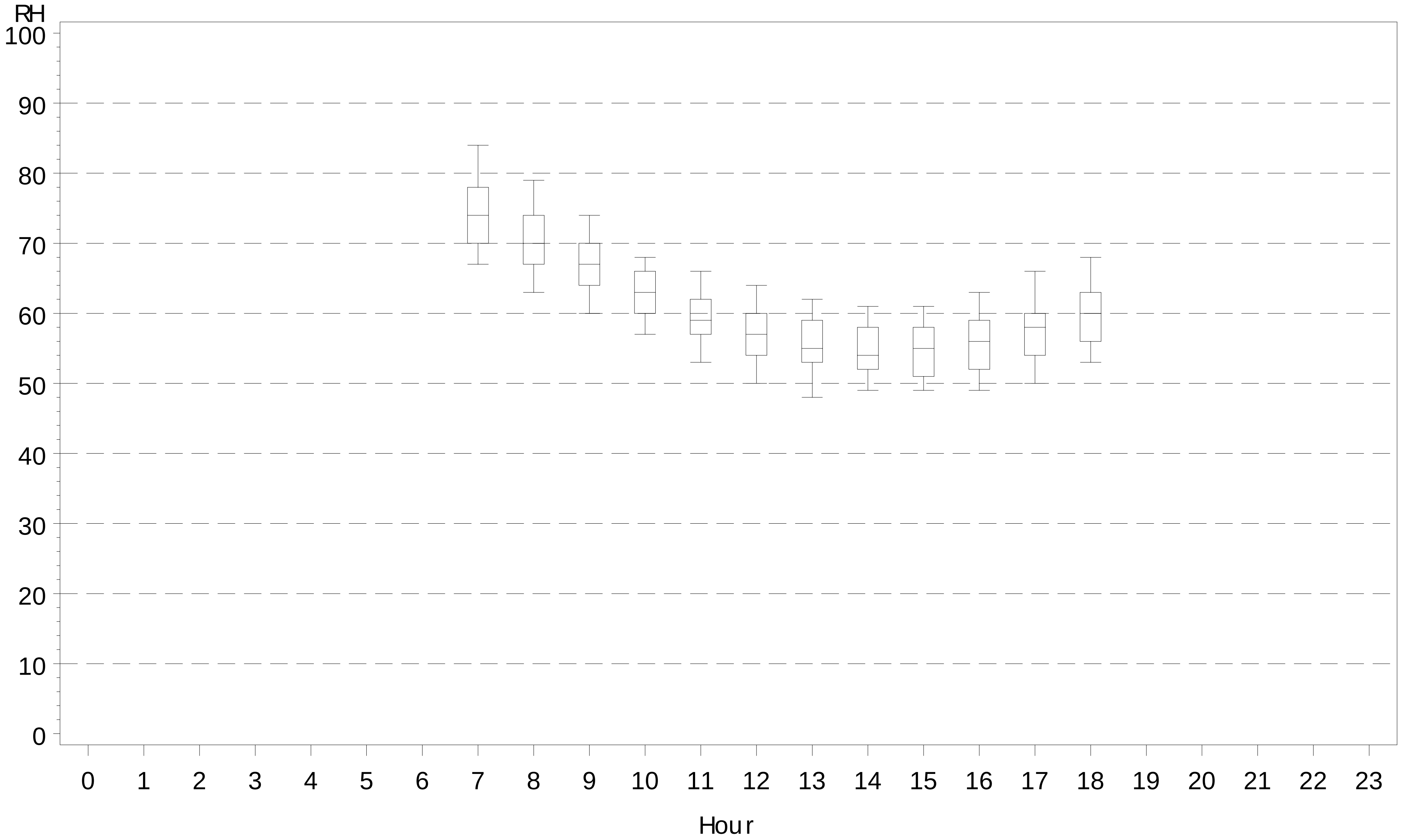
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Industrial Midwest



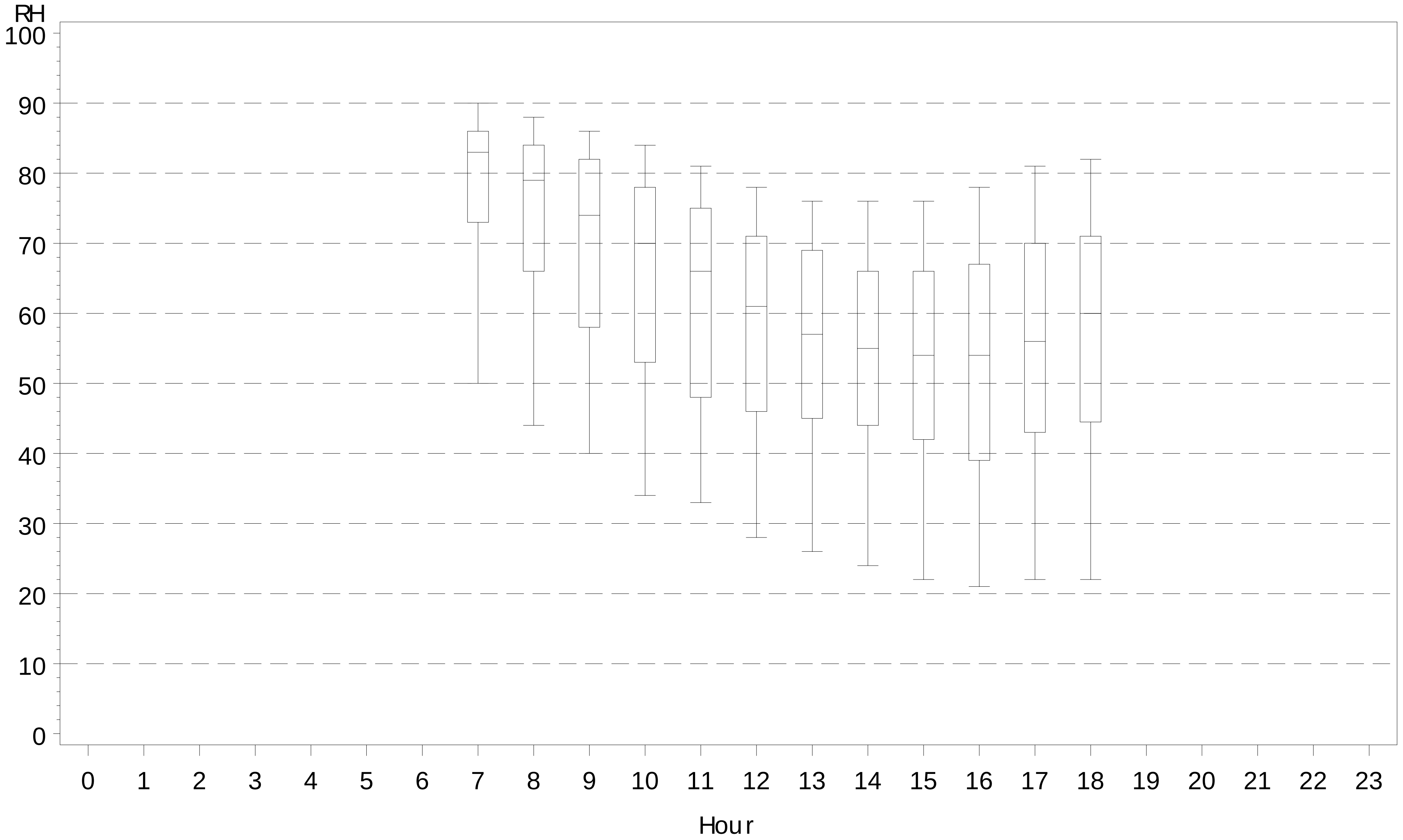
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Northeast



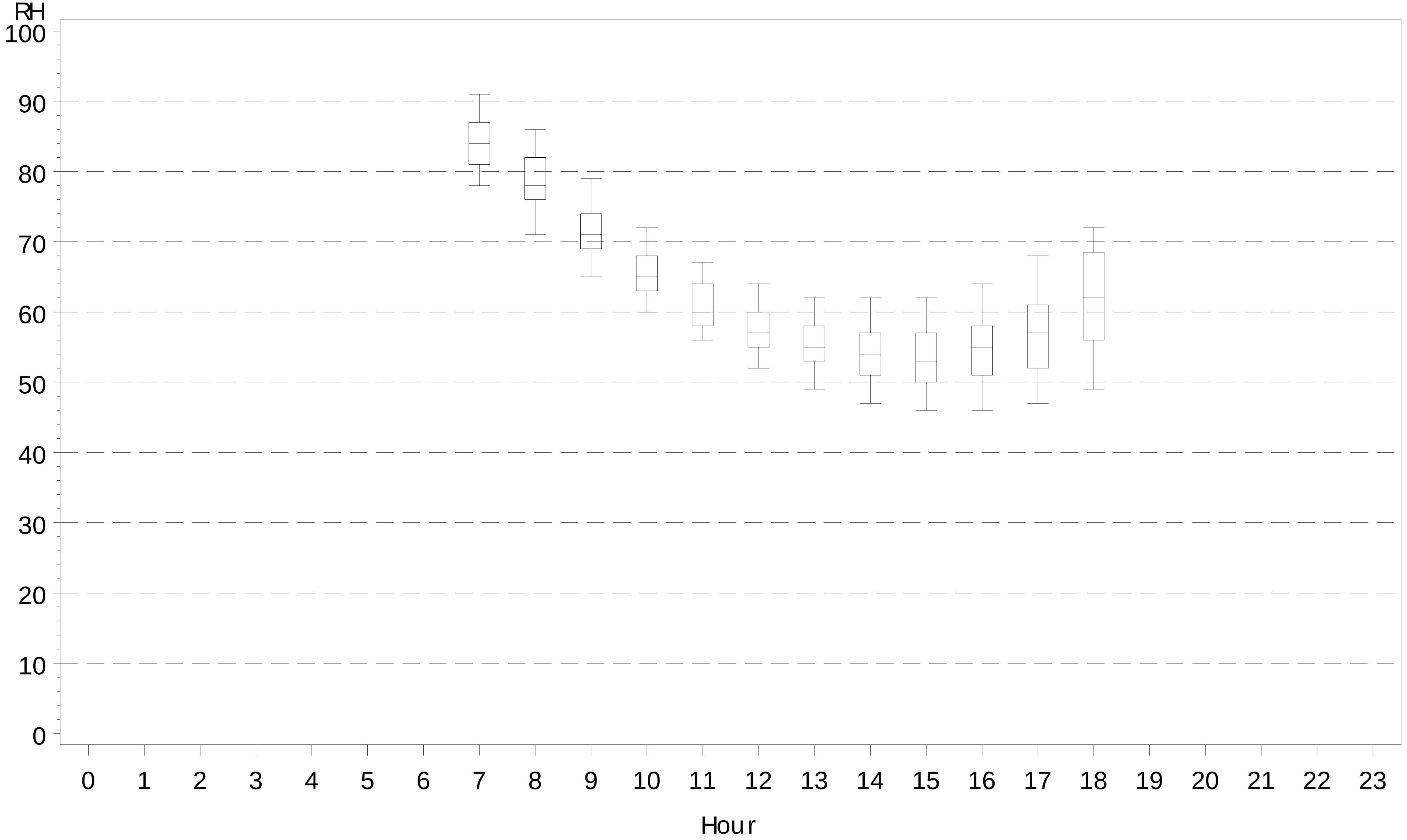
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Northwest



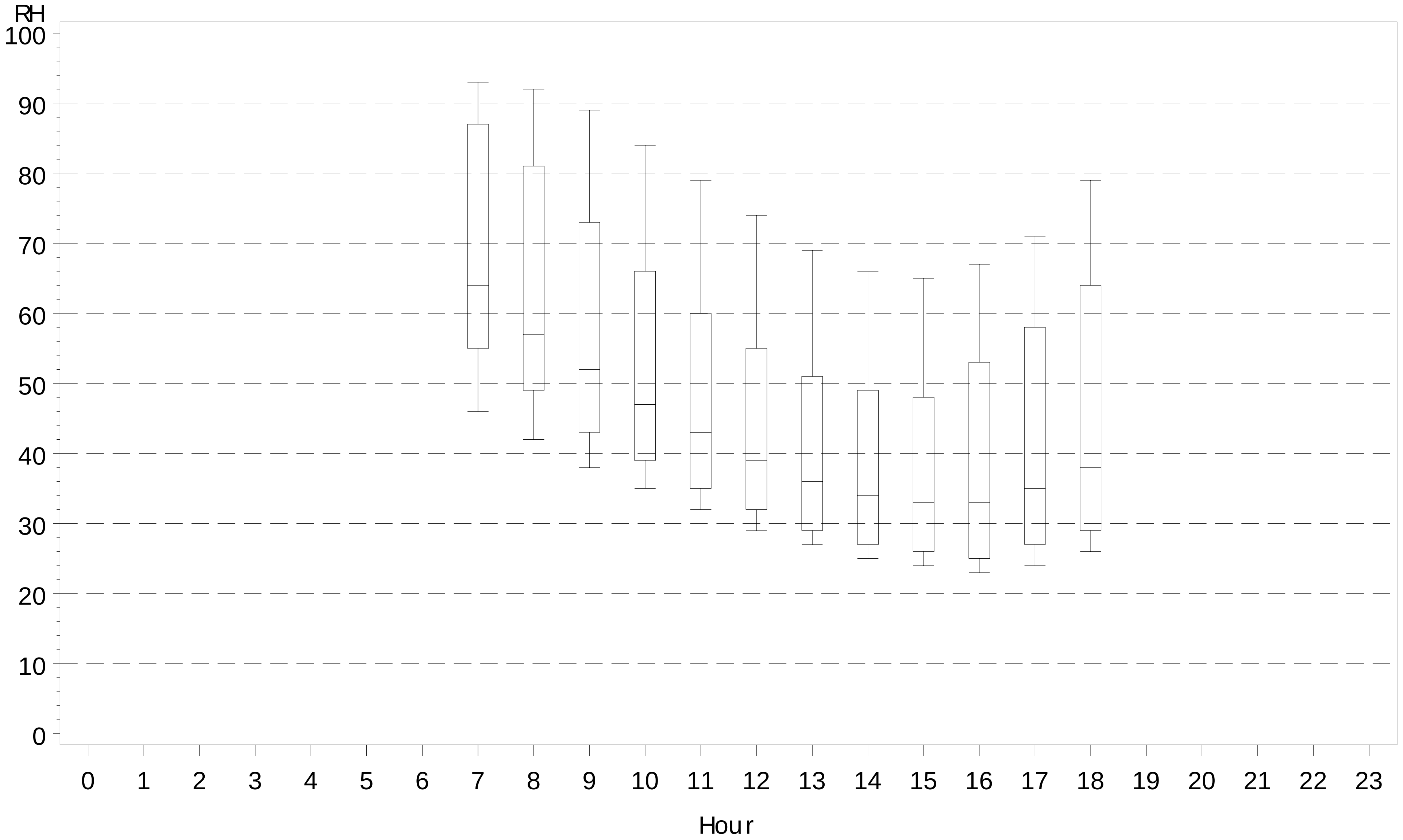
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Southeast



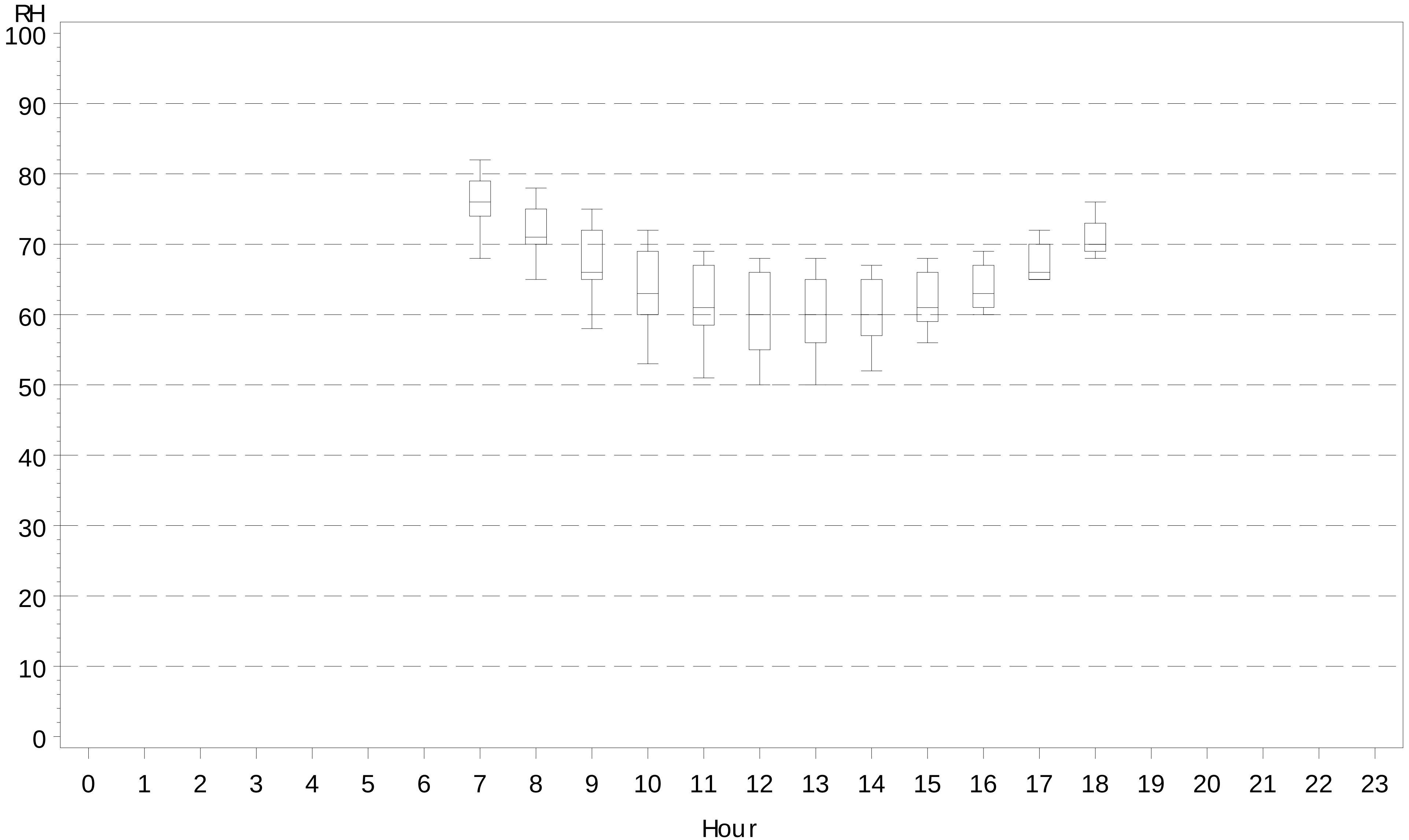
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Southern California



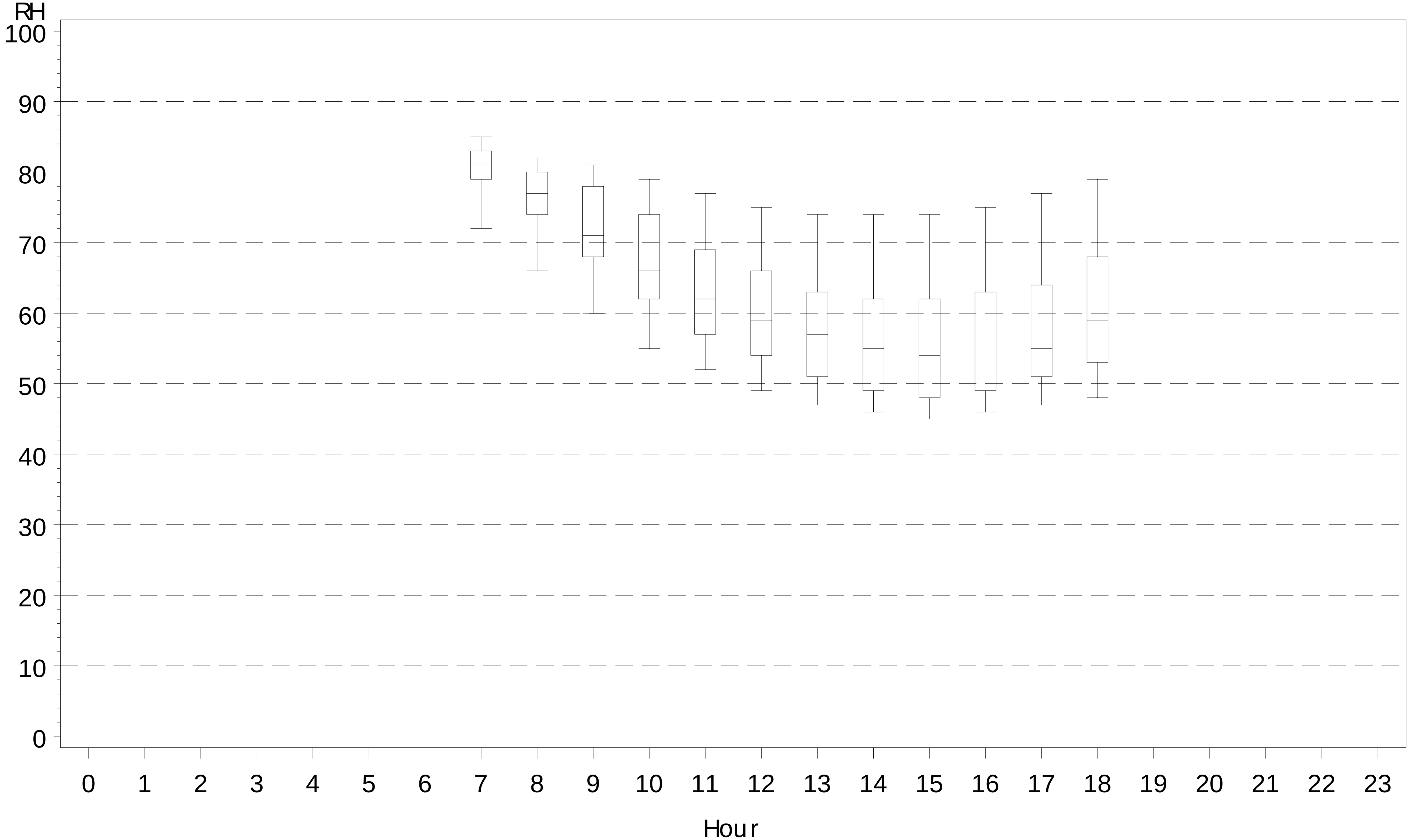
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Southwest



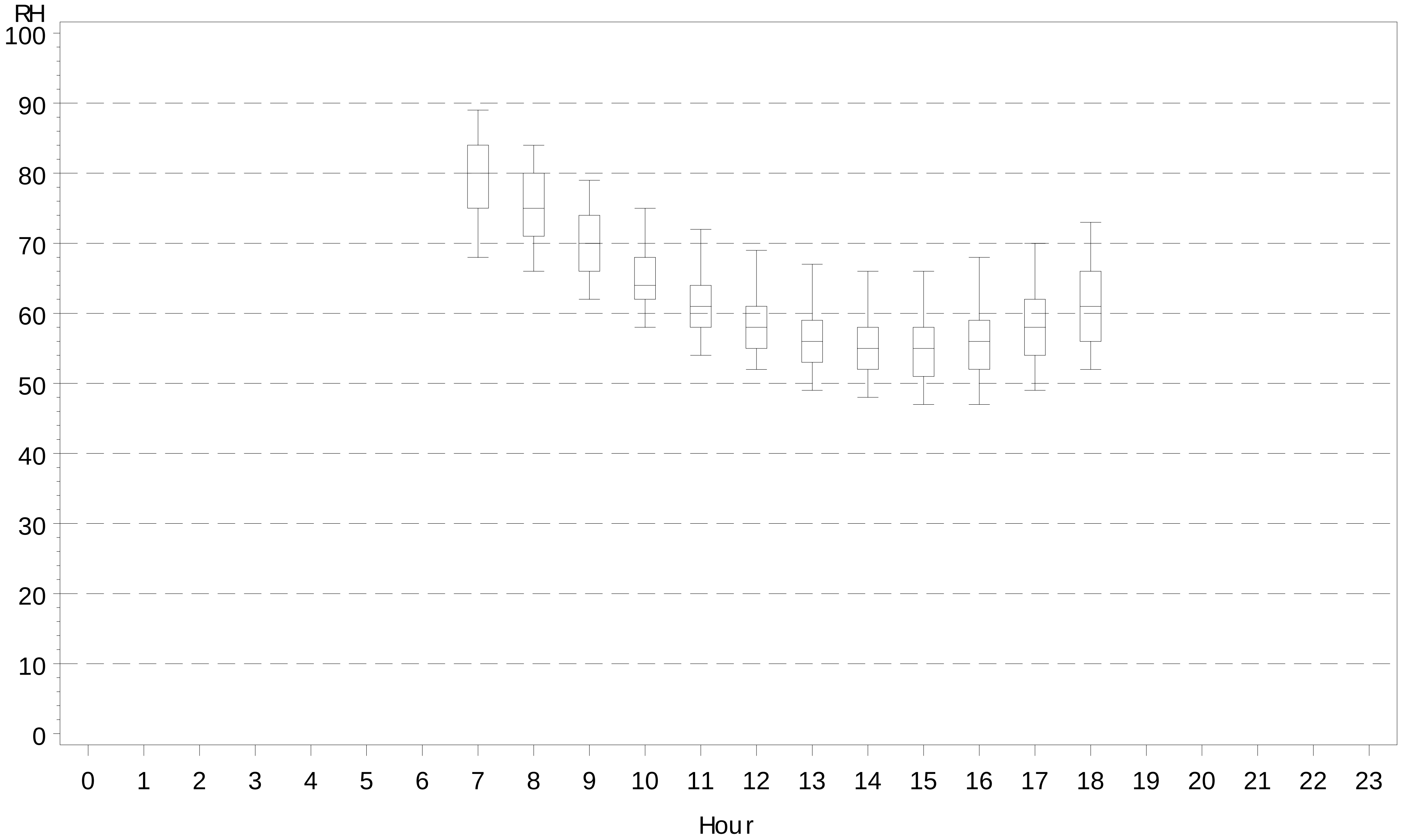
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=Upper Midwest



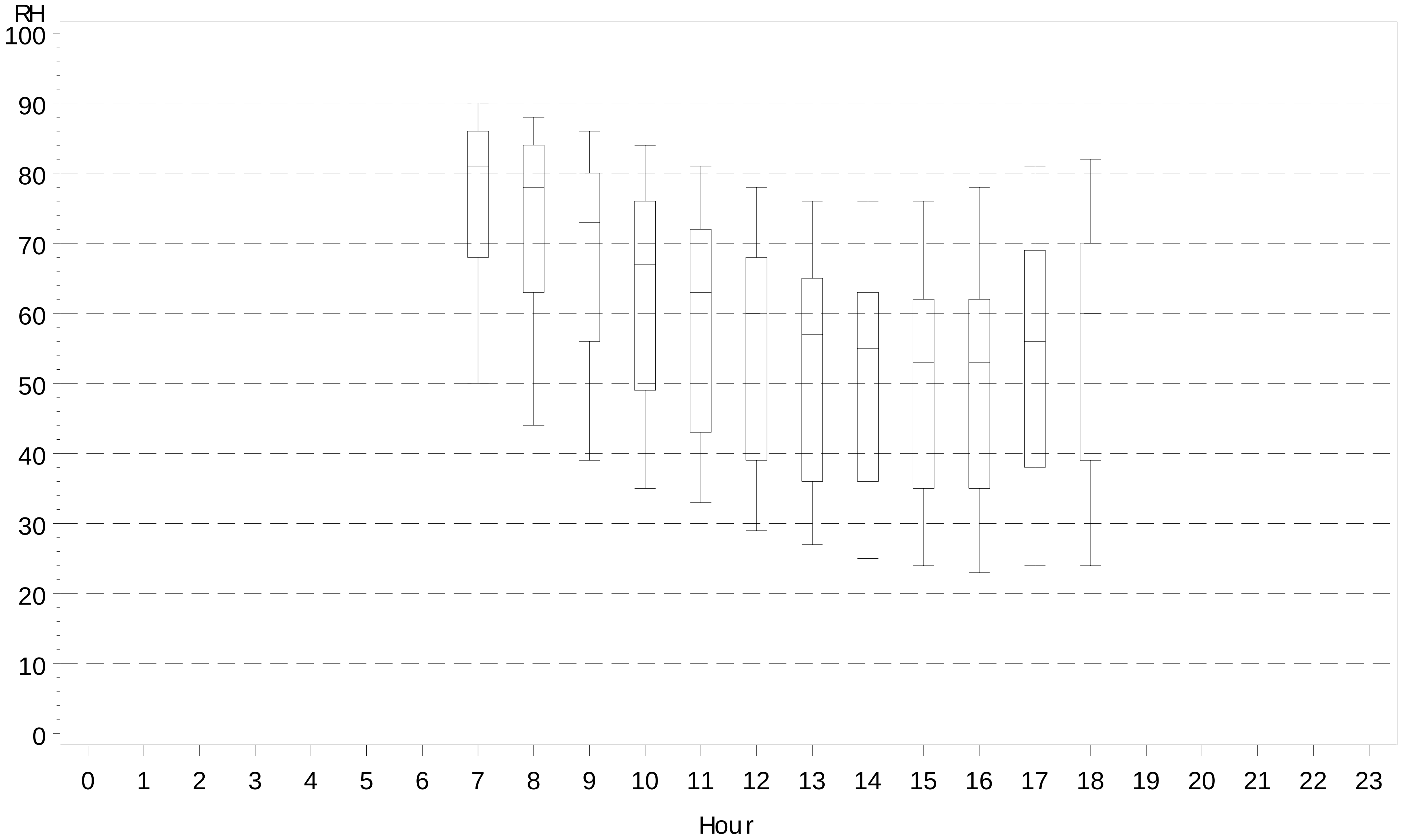
10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=east

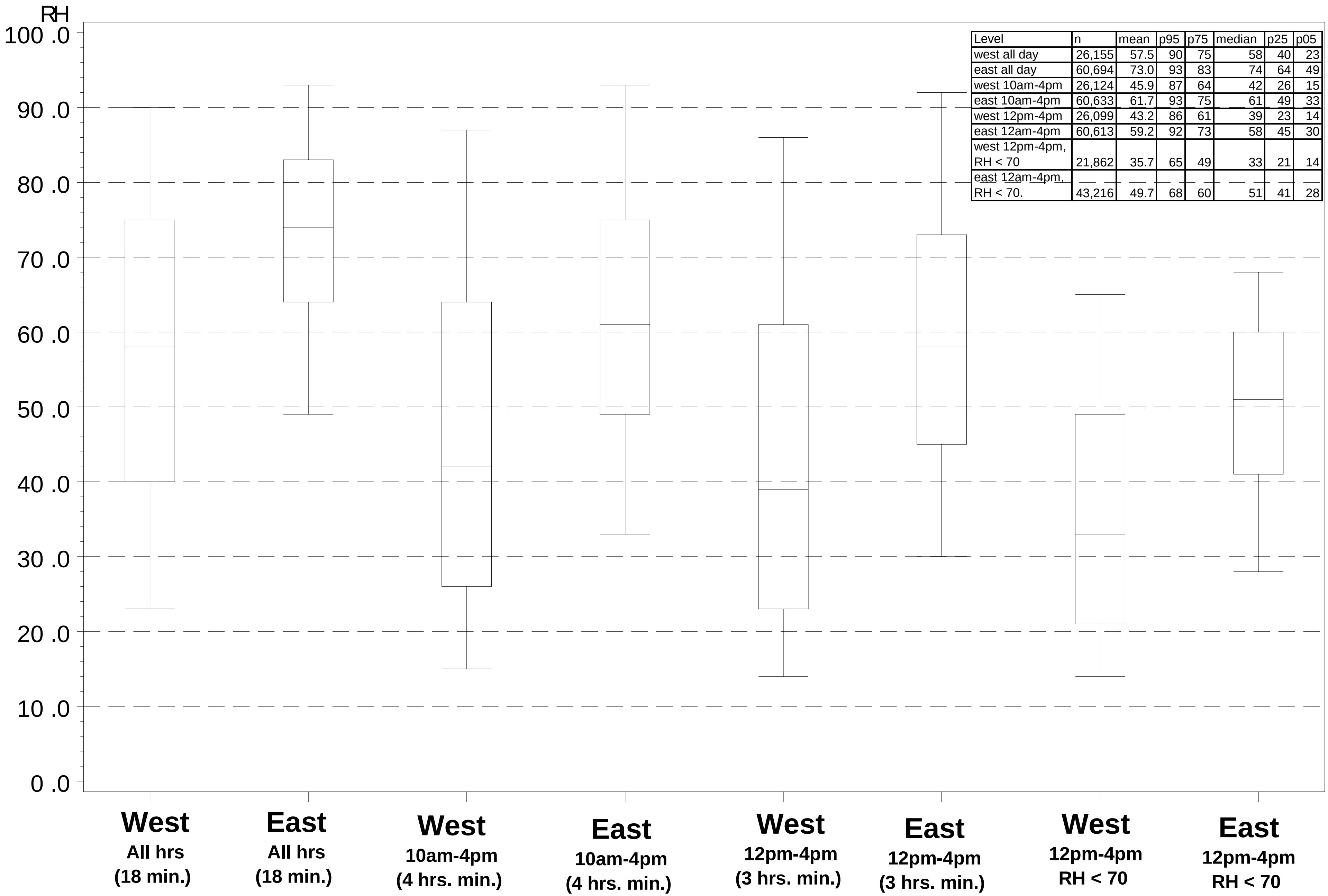


10-year average month-hour RH at NWS sites for site-days used in visibility analyses, by region and hour

region=west

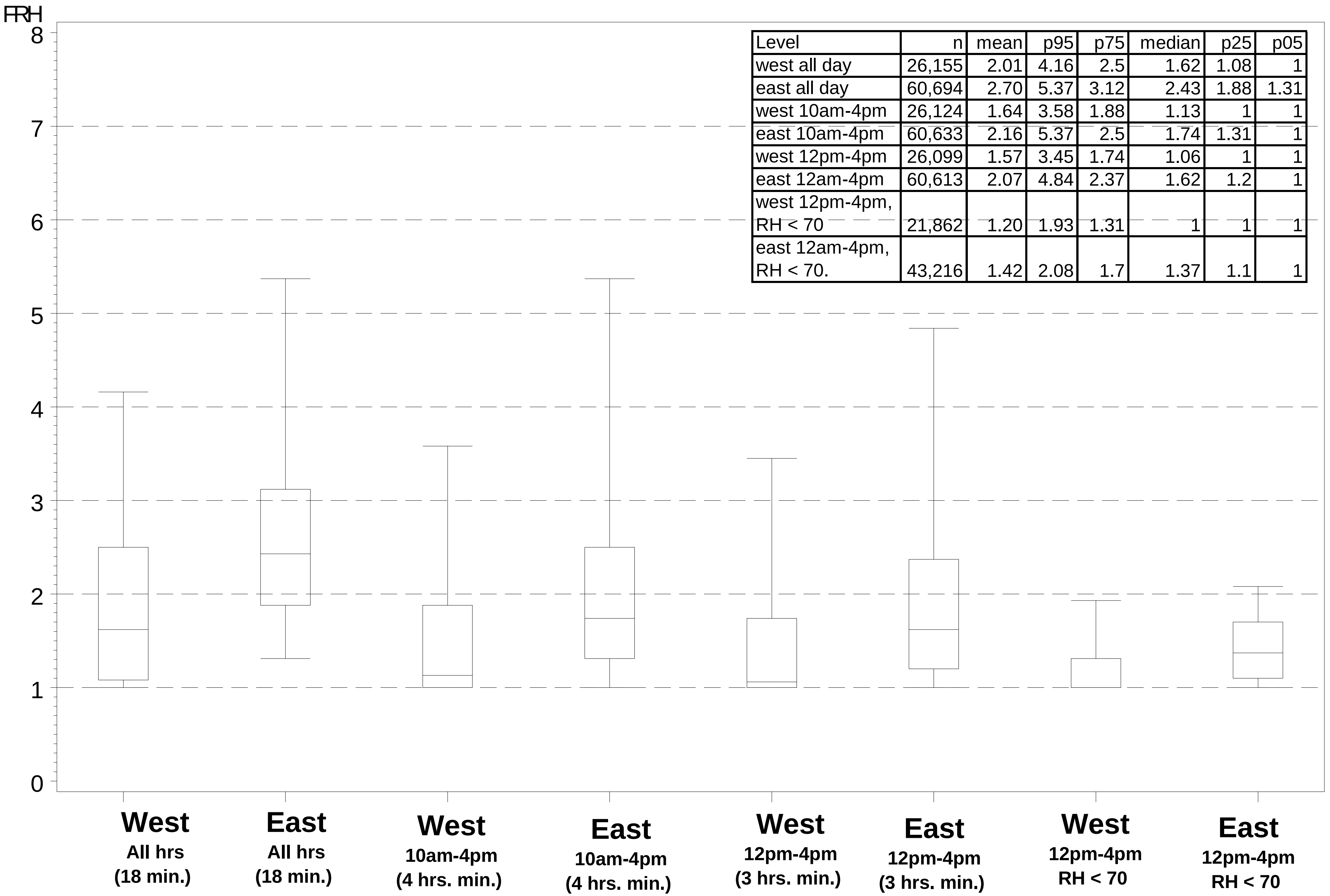


2003 NWS average RH for east and west, various time periods



Level	n	mean	p95	p75	median	p25	p05
west all day	26,155	57.5	90	75	58	40	23
east all day	60,694	73.0	93	83	74	64	49
west 10am-4pm	26,124	45.9	87	64	42	26	15
east 10am-4pm	60,633	61.7	93	75	61	49	33
west 12pm-4pm	26,099	43.2	86	61	39	23	14
east 12am-4pm	60,613	59.2	92	73	58	45	30
west 12pm-4pm, RH < 70	21,862	35.7	65	49	33	21	14
east 12am-4pm, RH < 70.	43,216	49.7	68	60	51	41	28

2003 NWS average FRH for east and west, various time periods



Level	n	mean	p95	p75	median	p25	p05
west all day	26,155	2.01	4.16	2.5	1.62	1.08	1
east all day	60,694	2.70	5.37	3.12	2.43	1.88	1.31
west 10am-4pm	26,124	1.64	3.58	1.88	1.13	1	1
east 10am-4pm	60,633	2.16	5.37	2.5	1.74	1.31	1
west 12pm-4pm	26,099	1.57	3.45	1.74	1.06	1	1
east 12am-4pm	60,613	2.07	4.84	2.37	1.62	1.2	1
west 12pm-4pm, RH < 70	21,862	1.20	1.93	1.31	1	1	1
east 12am-4pm, RH < 70.	43,216	1.42	2.08	1.7	1.37	1.1	1

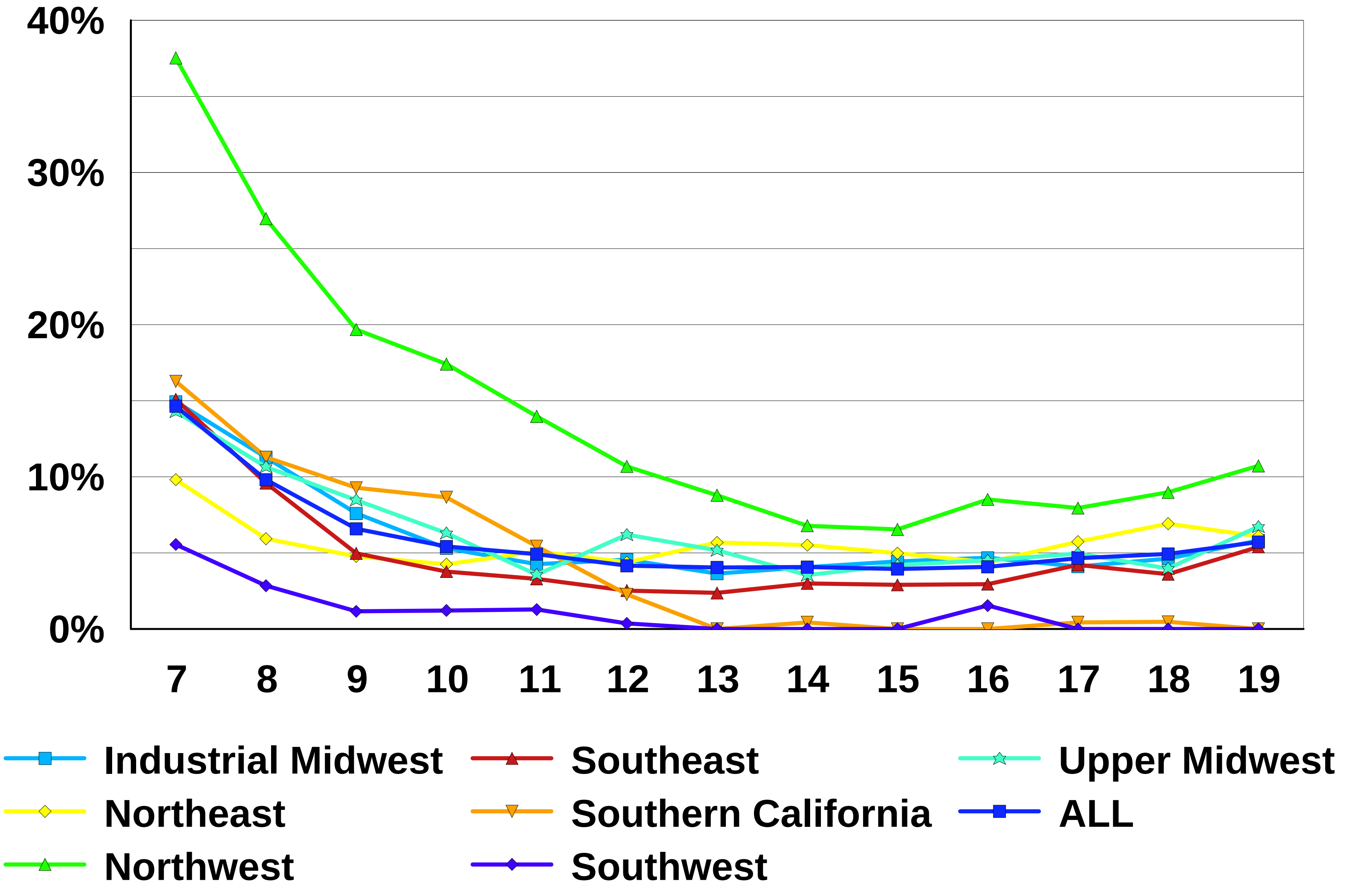
Percent of Hourly NWS RH Observations Under Select RH Thresholds, 2003

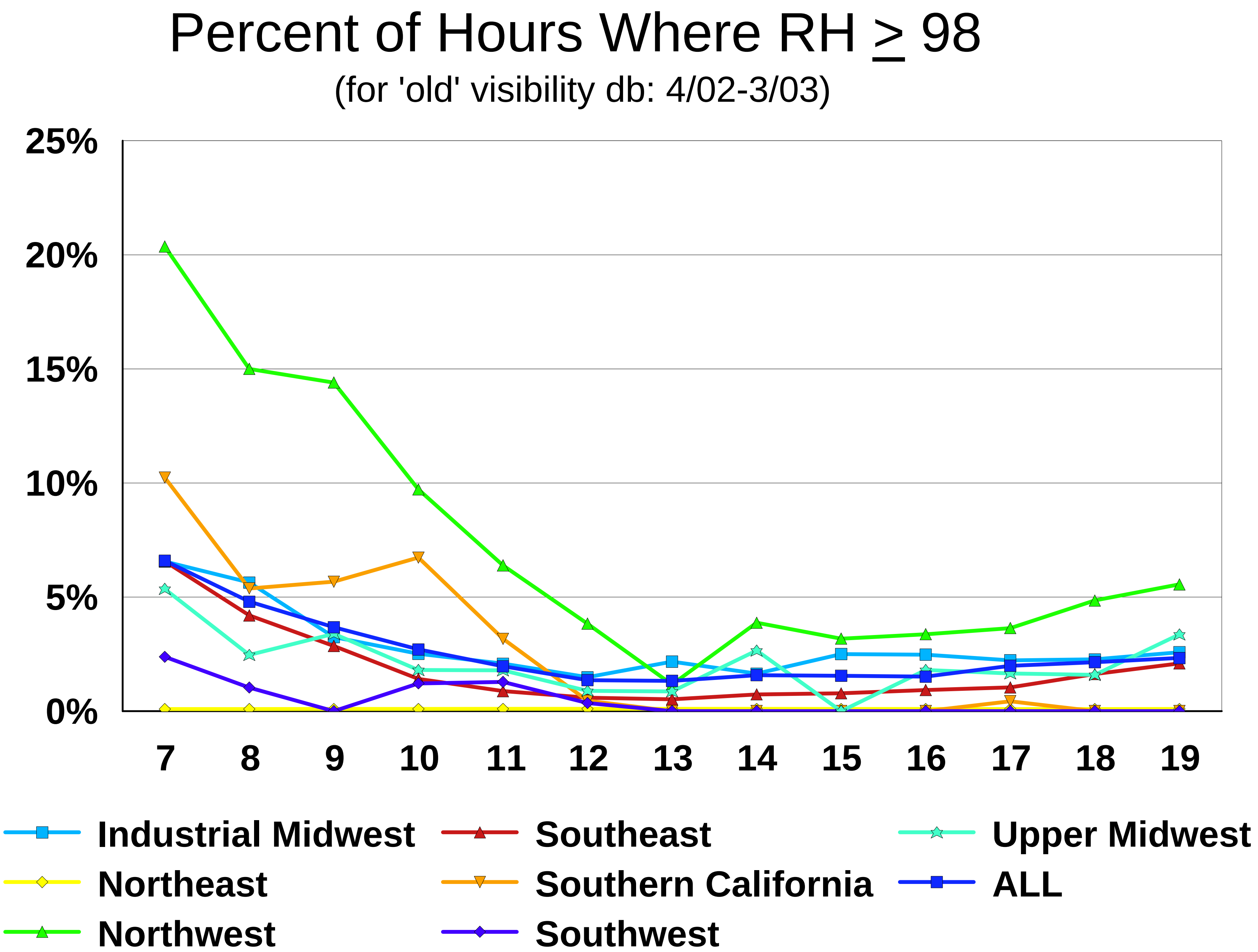
	Northwest				Southern CA				Southwest				Upper Midwest				Industrial Midwest				Northeast				Southeast			
	% N where RH less than				% N where RH less than				% N where RH less than				% N where RH less than				% N where RH less than				% N where RH less than				% N where RH less than			
Hour	95	90	80	70	95	90	80	70	95	90	80	70	95	90	80	70	95	90	80	70	95	90	80	70	95	90	80	70
0	90%	81%	62%	47%	88%	71%	49%	39%	97%	90%	79%	67%	91%	79%	52%	29%	83%	68%	43%	23%	79%	63%	41%	26%	76%	51%	28%	15%
1	89%	79%	59%	45%	86%	69%	48%	38%	96%	89%	76%	63%	90%	76%	48%	25%	81%	65%	39%	20%	76%	60%	39%	24%	72%	47%	25%	13%
2	88%	78%	57%	42%	86%	68%	47%	37%	95%	87%	74%	60%	88%	73%	44%	22%	79%	62%	36%	18%	75%	59%	37%	23%	69%	43%	22%	11%
3	87%	76%	54%	39%	84%	67%	46%	36%	94%	86%	71%	57%	87%	71%	41%	20%	77%	59%	33%	17%	73%	57%	35%	21%	66%	40%	21%	10%
4	86%	74%	52%	37%	83%	65%	45%	34%	93%	84%	68%	55%	85%	69%	38%	18%	75%	57%	31%	15%	72%	55%	33%	20%	63%	38%	19%	9%
5	85%	73%	50%	35%	83%	65%	45%	34%	92%	82%	66%	52%	83%	66%	35%	16%	74%	55%	30%	14%	71%	54%	32%	19%	61%	36%	18%	9%
6	86%	74%	52%	37%	83%	66%	45%	33%	92%	81%	64%	50%	84%	66%	34%	15%	74%	56%	29%	14%	73%	56%	34%	19%	61%	36%	17%	8%
7	88%	78%	57%	42%	86%	73%	50%	37%	93%	83%	66%	52%	88%	73%	41%	18%	79%	62%	33%	15%	78%	63%	40%	23%	70%	43%	20%	9%
8	90%	82%	65%	50%	91%	82%	62%	45%	94%	88%	74%	59%	92%	82%	55%	29%	85%	71%	44%	22%	84%	71%	49%	30%	82%	63%	33%	14%
9	93%	87%	73%	60%	95%	90%	75%	57%	96%	93%	84%	71%	94%	87%	68%	46%	90%	80%	58%	35%	88%	78%	61%	42%	90%	79%	57%	28%
10	95%	90%	79%	68%	97%	94%	83%	66%	98%	95%	90%	81%	96%	91%	77%	59%	92%	85%	69%	48%	91%	83%	69%	52%	94%	87%	73%	47%
11	97%	93%	84%	74%	99%	96%	89%	72%	99%	97%	93%	86%	97%	93%	82%	68%	94%	88%	75%	58%	92%	86%	74%	59%	96%	90%	81%	62%
12	97%	94%	86%	78%	99%	98%	92%	76%	99%	98%	95%	90%	97%	94%	85%	73%	95%	90%	79%	64%	93%	87%	76%	64%	97%	92%	84%	69%
13	98%	95%	88%	81%	99%	98%	93%	78%	99%	98%	96%	92%	98%	95%	87%	76%	96%	91%	81%	68%	93%	88%	78%	66%	97%	93%	86%	73%
14	98%	96%	89%	82%	99%	98%	93%	78%	99%	98%	96%	93%	98%	96%	88%	78%	96%	92%	82%	70%	93%	87%	78%	67%	97%	93%	86%	74%
15	98%	95%	89%	82%	99%	98%	92%	76%	99%	98%	96%	94%	98%	95%	88%	78%	96%	92%	82%	70%	93%	87%	77%	66%	97%	92%	86%	73%
16	98%	95%	88%	80%	99%	97%	89%	71%	99%	98%	96%	94%	98%	95%	87%	77%	96%	91%	81%	69%	93%	87%	76%	64%	96%	92%	84%	70%
17	97%	94%	85%	77%	99%	95%	82%	64%	99%	98%	96%	93%	98%	94%	85%	74%	95%	90%	80%	66%	92%	85%	73%	60%	96%	91%	81%	65%
18	96%	92%	82%	72%	98%	90%	72%	55%	99%	98%	95%	91%	97%	93%	83%	69%	94%	89%	76%	60%	90%	83%	68%	54%	95%	88%	74%	53%
19	95%	90%	78%	67%	96%	85%	62%	48%	99%	97%	93%	87%	96%	92%	78%	62%	93%	86%	71%	52%	89%	79%	62%	46%	93%	82%	63%	40%
20	94%	88%	74%	62%	94%	80%	57%	46%	99%	96%	91%	83%	95%	90%	72%	52%	91%	82%	63%	42%	87%	75%	56%	39%	90%	75%	51%	29%
21	93%	86%	71%	57%	92%	77%	55%	44%	98%	95%	88%	78%	94%	87%	66%	44%	89%	78%	57%	35%	84%	72%	51%	34%	87%	68%	43%	23%
22	92%	84%	68%	53%	91%	75%	53%	42%	98%	94%	85%	74%	93%	84%	61%	38%	87%	74%	51%	30%	83%	69%	47%	30%	83%	62%	37%	19%
23	91%	83%	65%	50%	89%	73%	51%	41%	97%	92%	81%	70%	92%	82%	56%	33%	85%	71%	47%	26%	81%	66%	44%	28%	80%	57%	32%	17%

Red indicates less than 75%

Percent of Hours Where RH ≥ 95

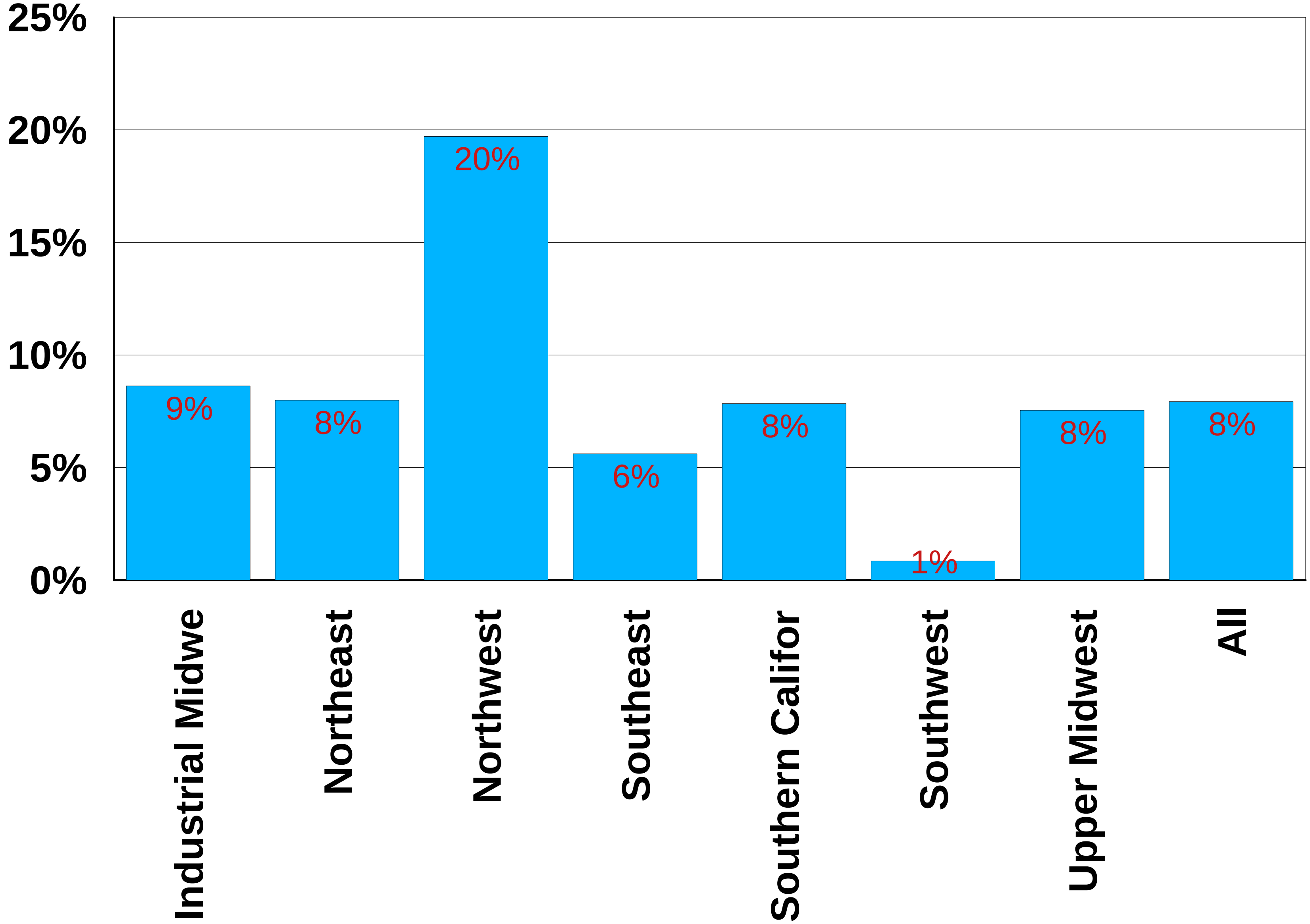
(for 'old' visibility db: 4/02-3/03)



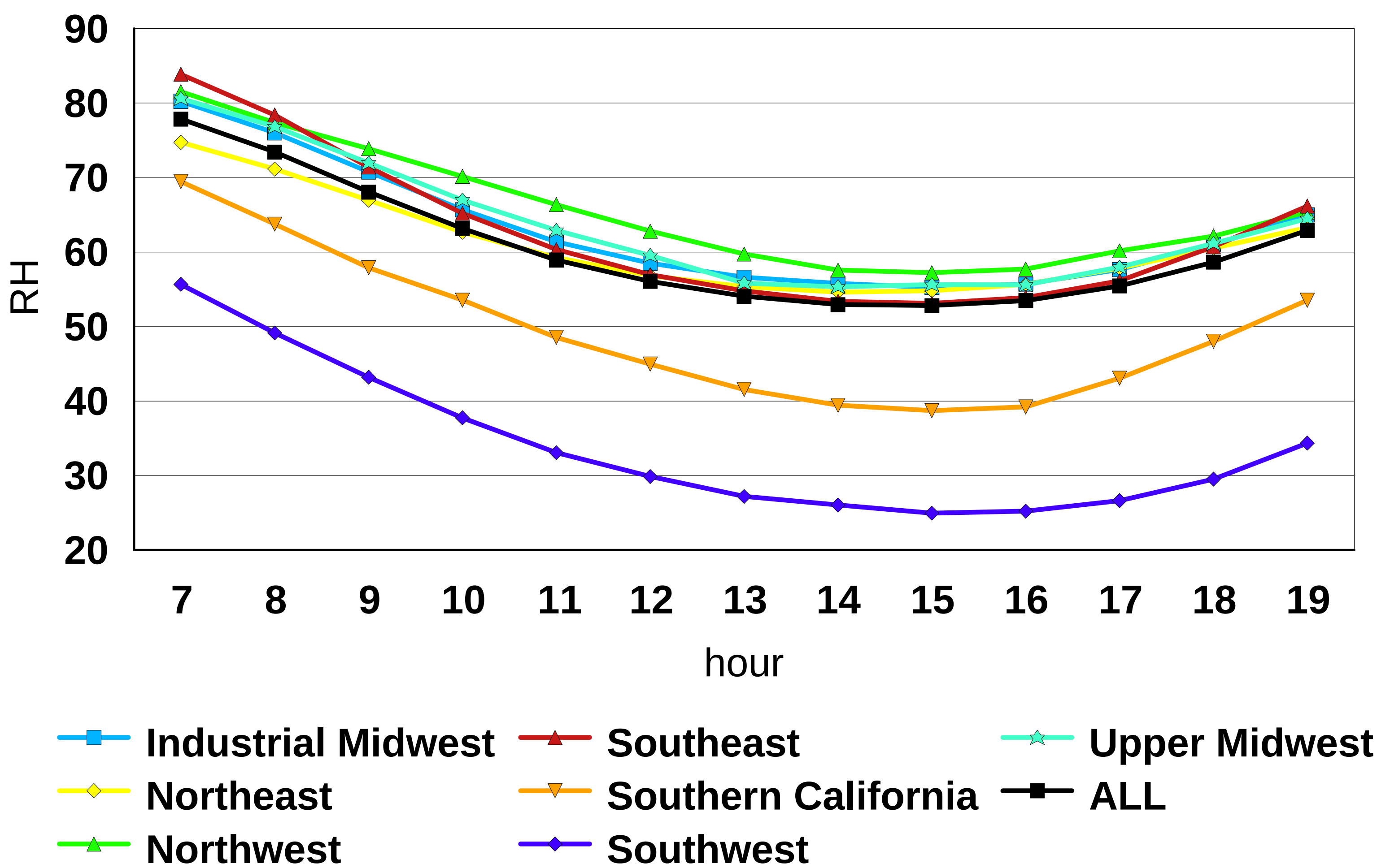

Percent of Hours Where RH ≥ 98
(for 'old' visibility db: 4/02-3/03)
25%
20%
15%
10%
5%
0%
7
8
9
10
11
12
13
14
15
16
17
18
19
Industrial Midwest
Southeast
Upper Midwest
Northeast
Southern California
ALL
Northwest
Southwest

Percent of 10am-4pm blocks where RH ≥ 95 (for 1+ hrs)

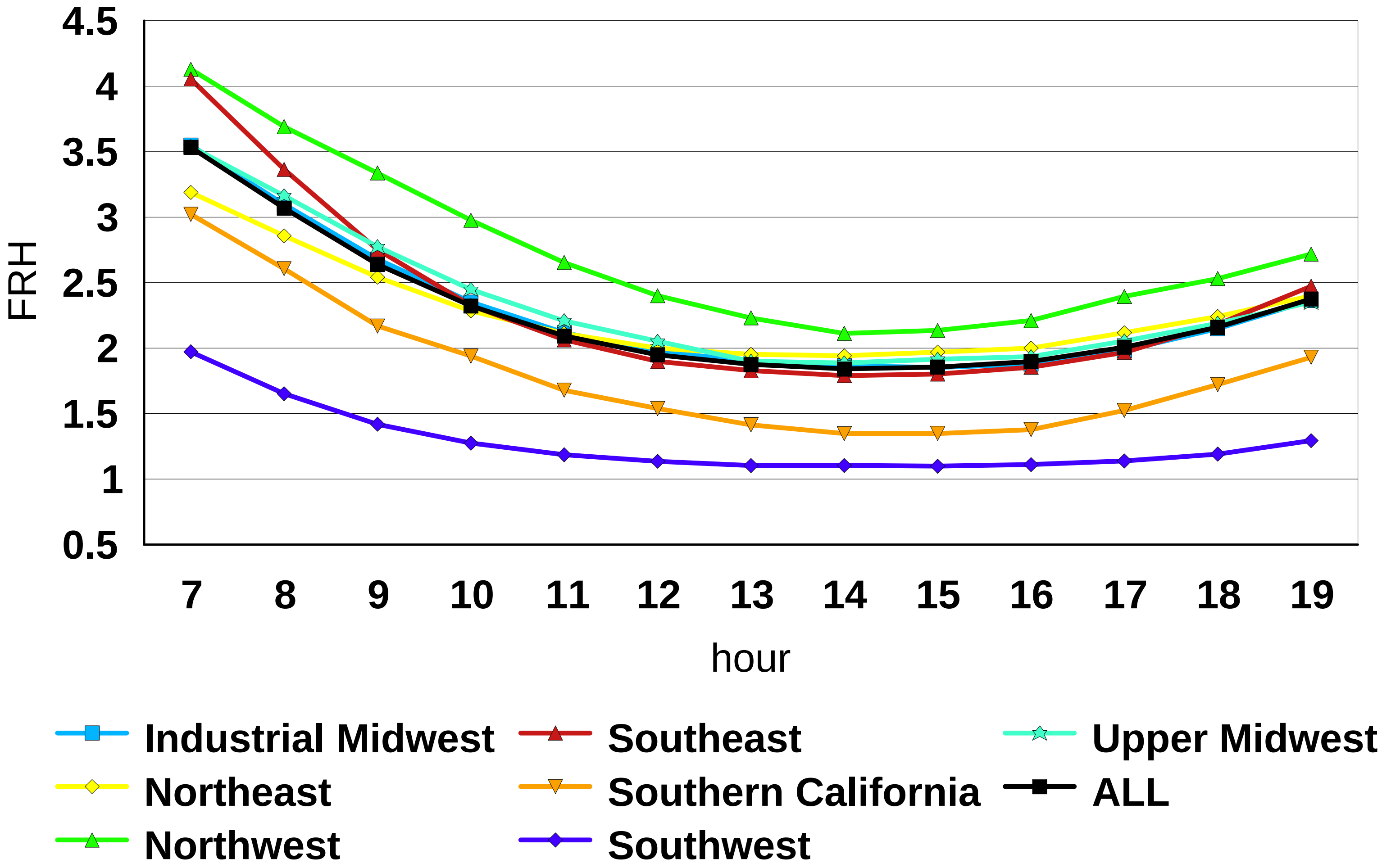
(for 'old' visibility db: 4/02-3/03)



RH - 10yr Avg
(for 'old' vis. db)

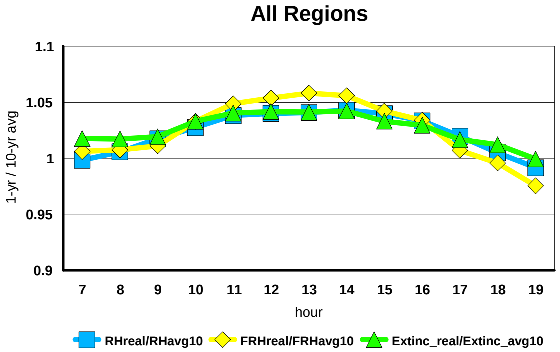


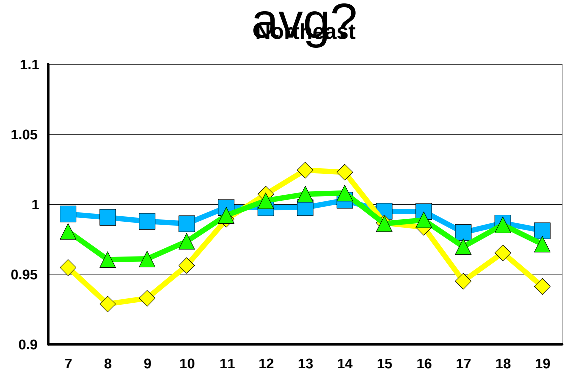
FRH - 10yr Avg (for 'old vis. db)

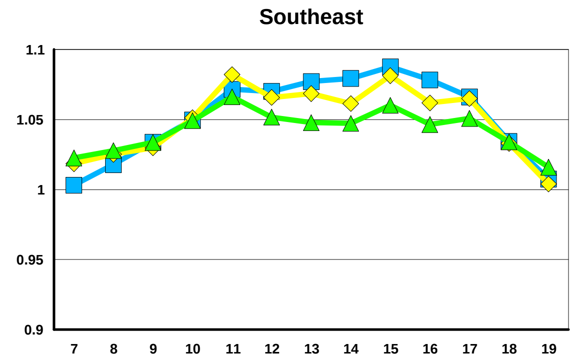


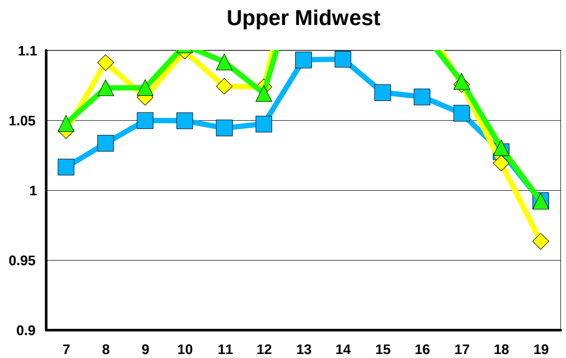
RH, FRH, Recon Extinction: Actual RH vs. 10yr Avg

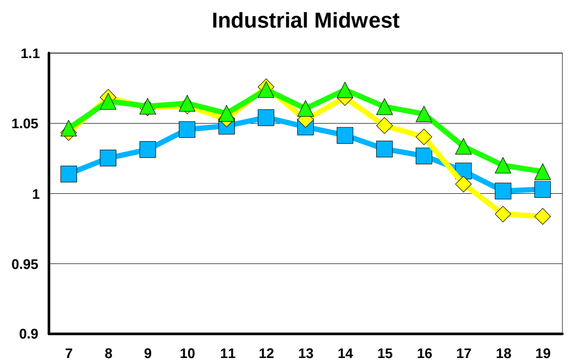
How Representative was 4/02 - 3/03 ('old' vis db) ..... compared to 10-yr avg?



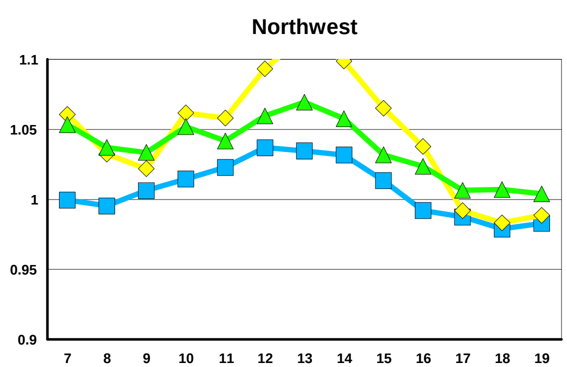


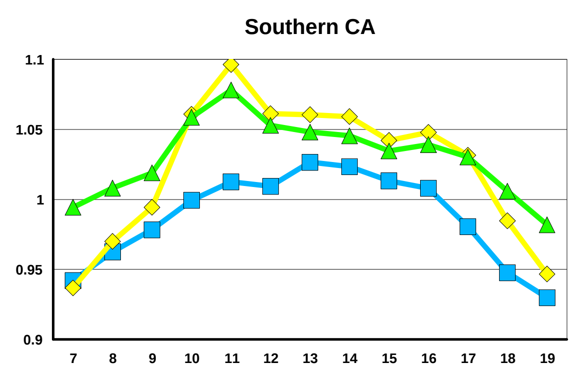


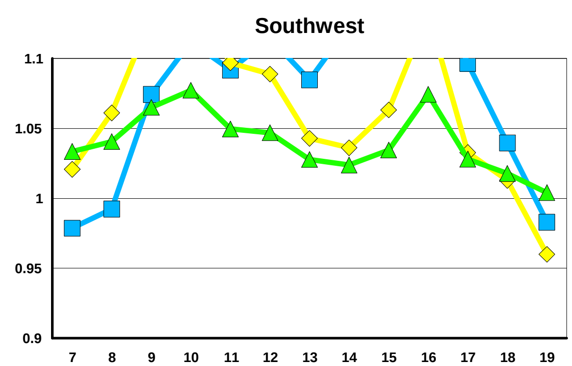


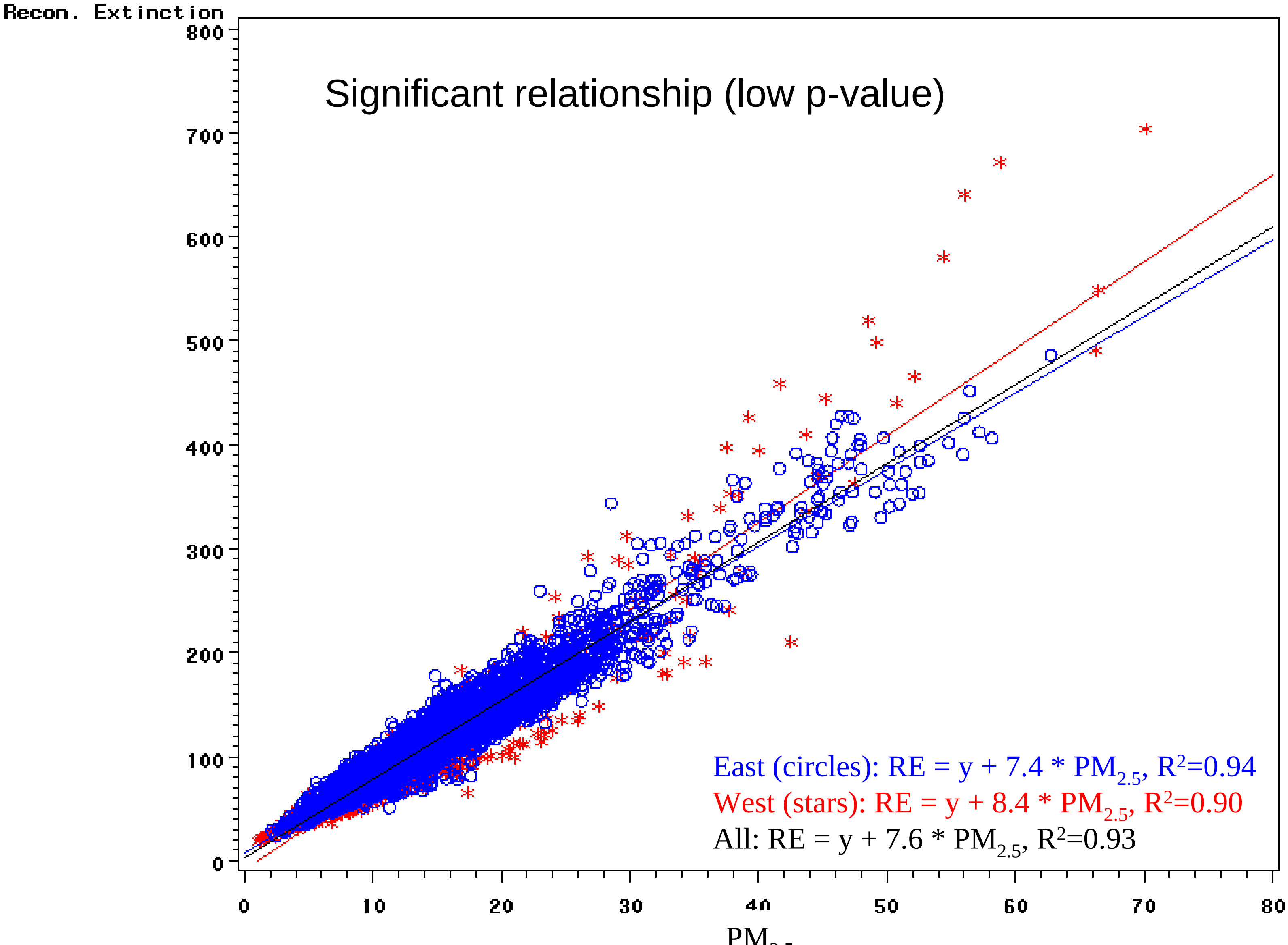


- Most areas had higher midday humidity during 4/02-3/03 than 'average' (~4% higher for all regions). Hence, FRH and reconstructed extinction values are higher than those calculated w/ 10-yr avg FRH's.... for same PM2.5 levels. Slopes for Recon. Extinc. = PM2.5 would be greater

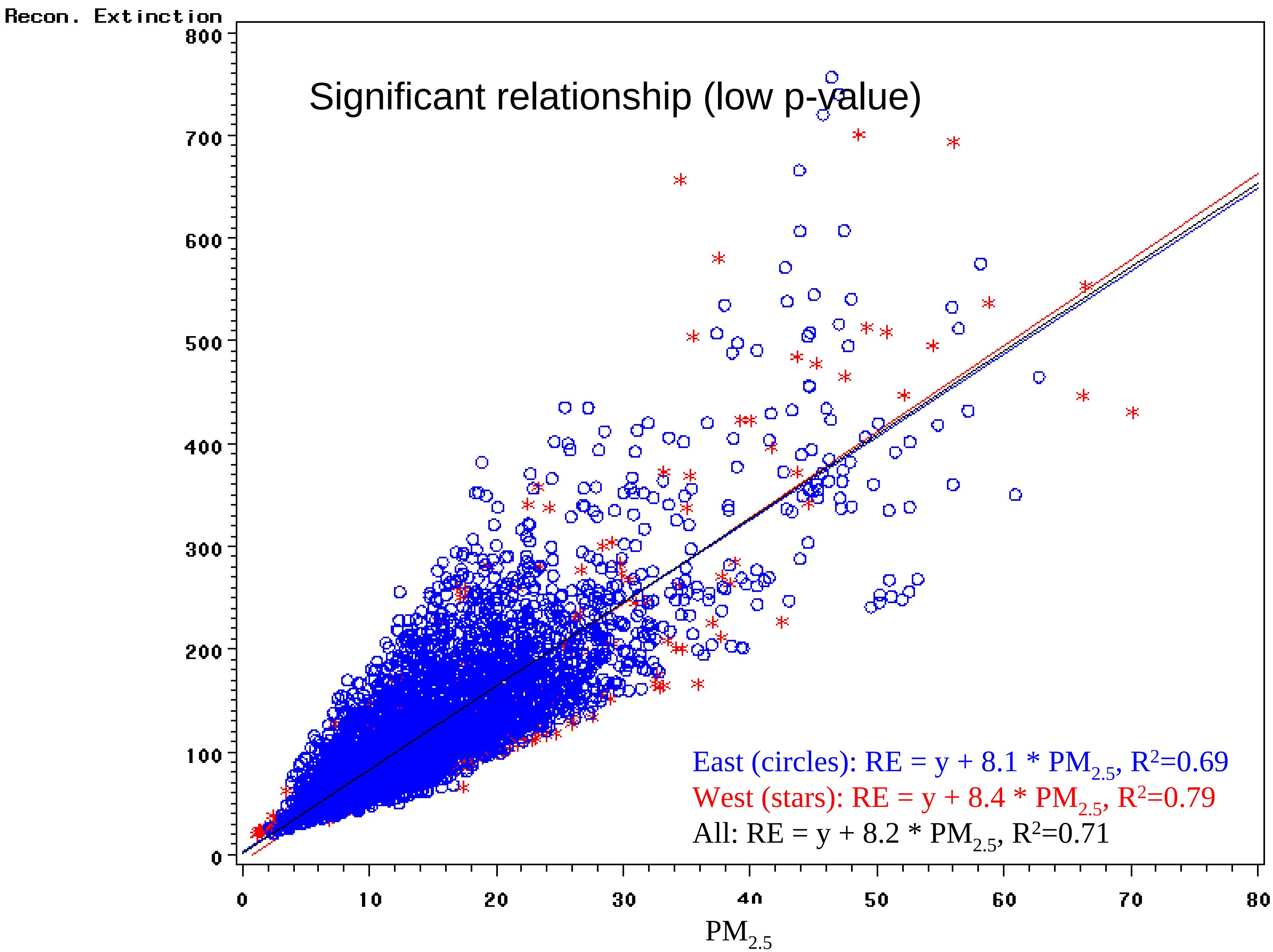








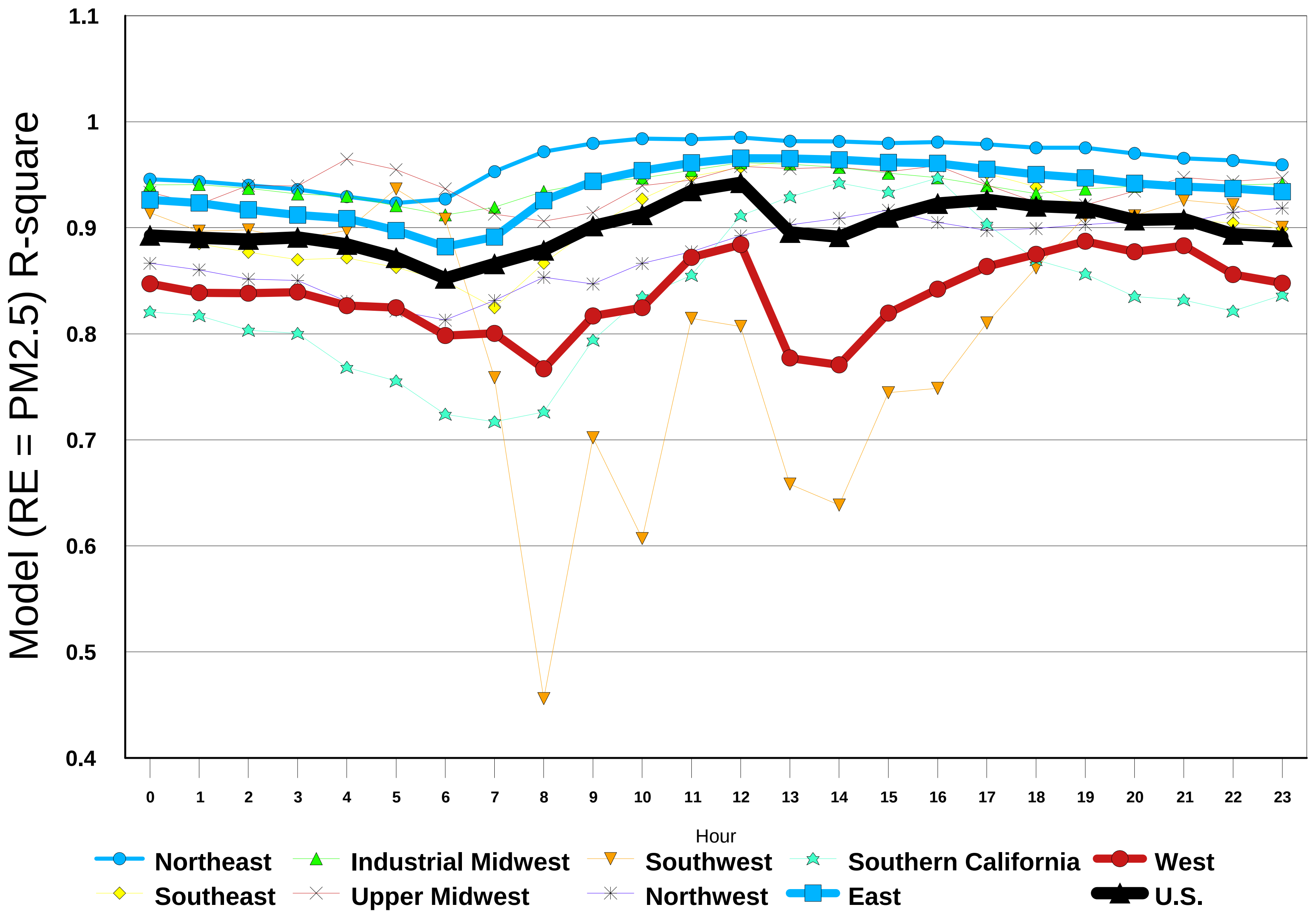
Relationship between reconstructed light extinction (RE) and 24-hour average $PM_{2.5}$. 2003.
Using 10-year average *f(RH)*.



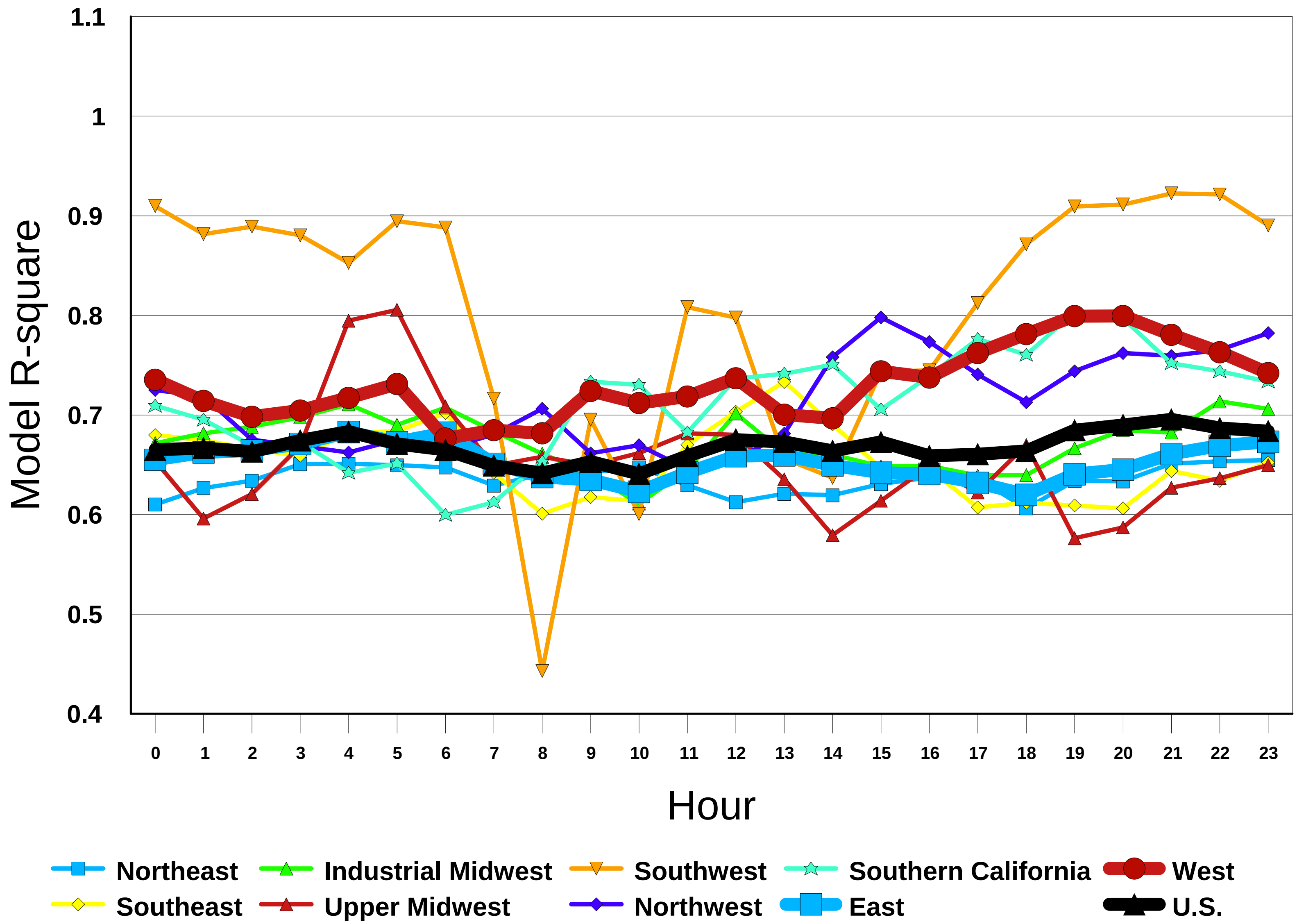
Relationship between reconstructed light extinction (RE) and 24-hour average $PM_{2.5}$, 2003. Using actual *f(RH)*

Model R^2 for relationship between reconstructed light extinction (RE) and hourly PM2.5 (increase in RE due to incremental increase in PM2.5), 2003.

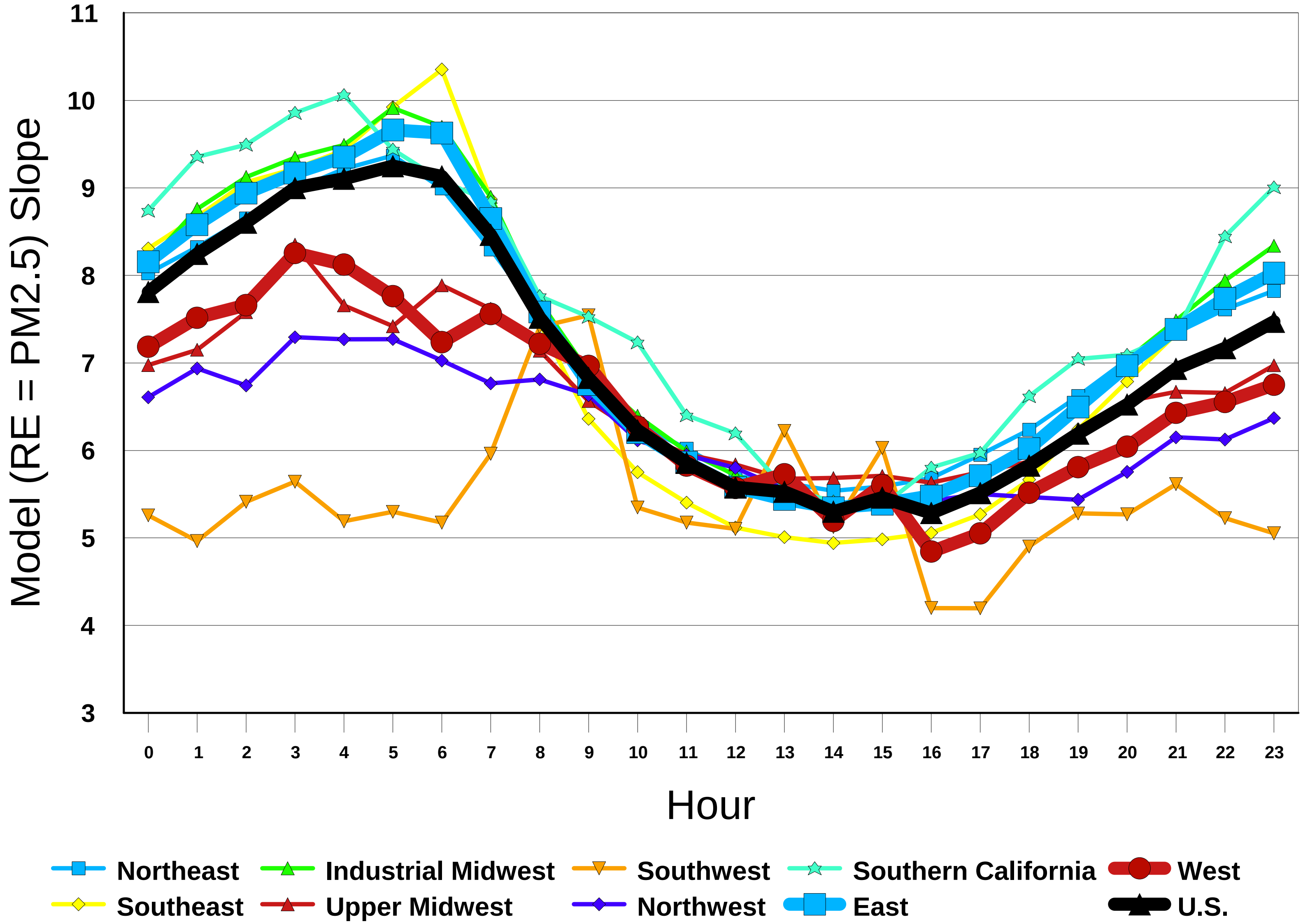
RE computed using 10-year average *f(RH)*.



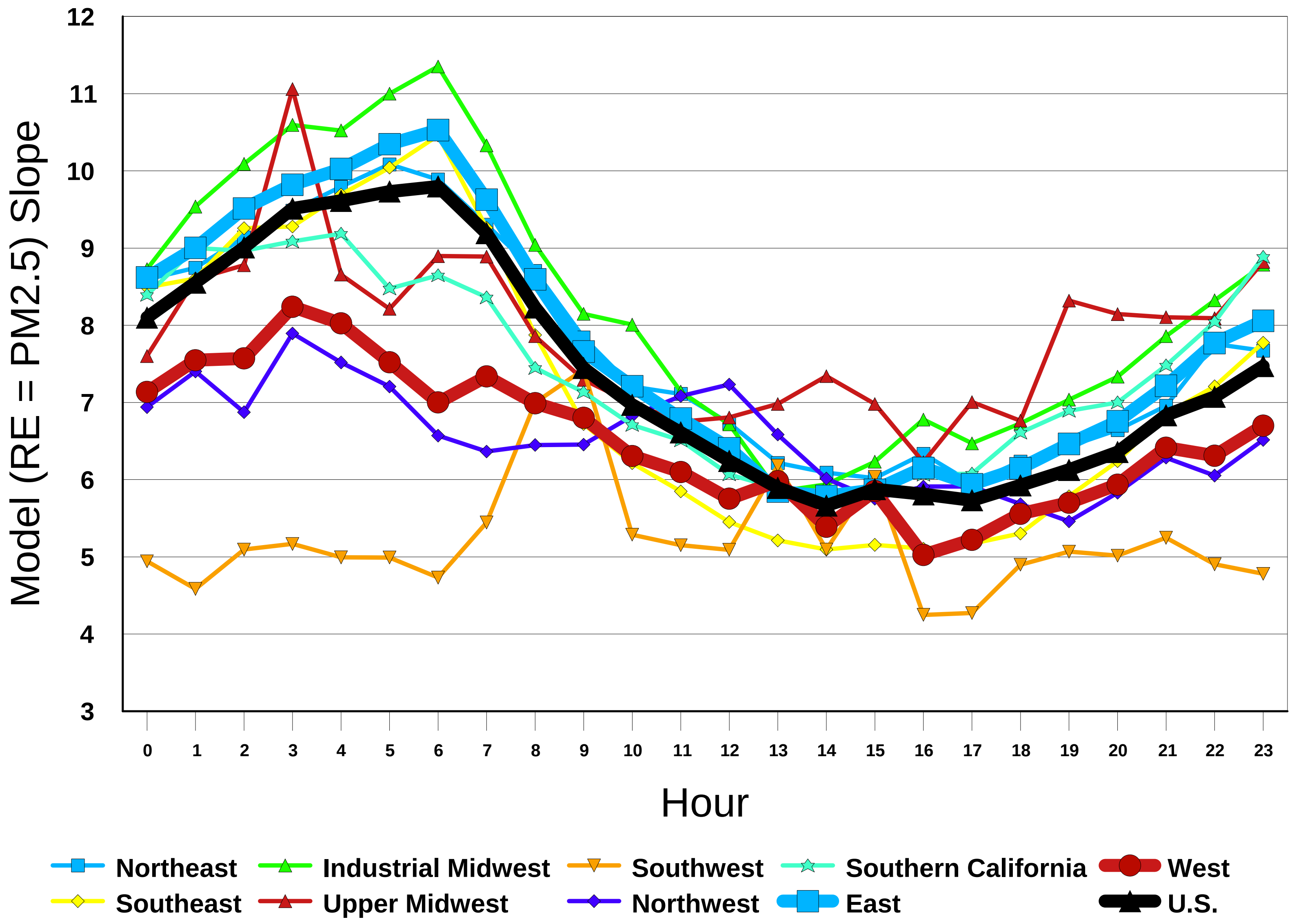
Model R^2 for relationship between reconstructed light extinction (RE) and hourly PM2.5 (increase in RE due to incremental increase in PM2.5), 2003.
RE computed with actual *f(RH)*.

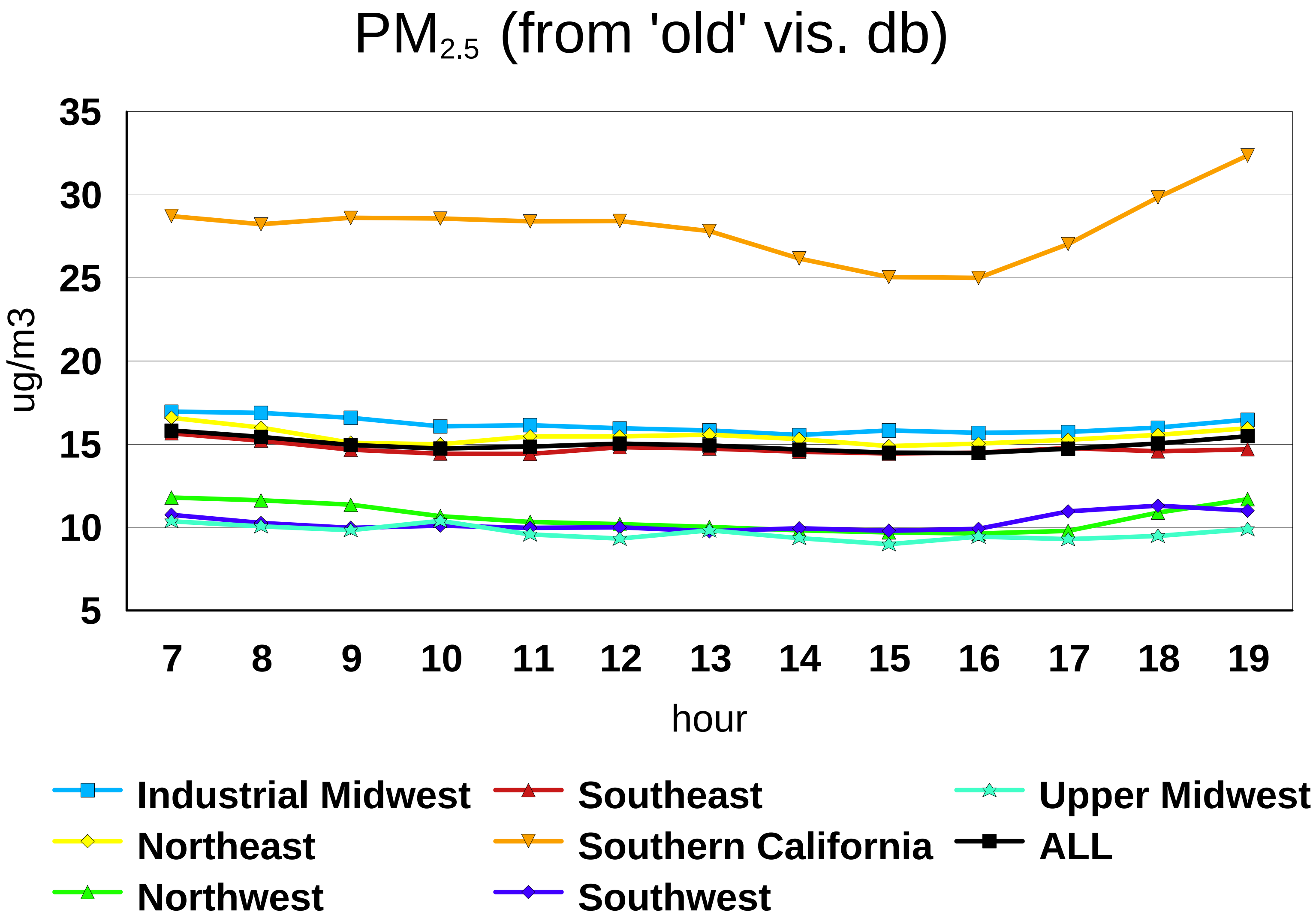


Model slope for relationship between reconstructed light extinction (RE) and hourly PM2.5 (increase in RE due to incremental increase in PM2.5), 2003.
RE computed with 10-year average *f(RH)*.



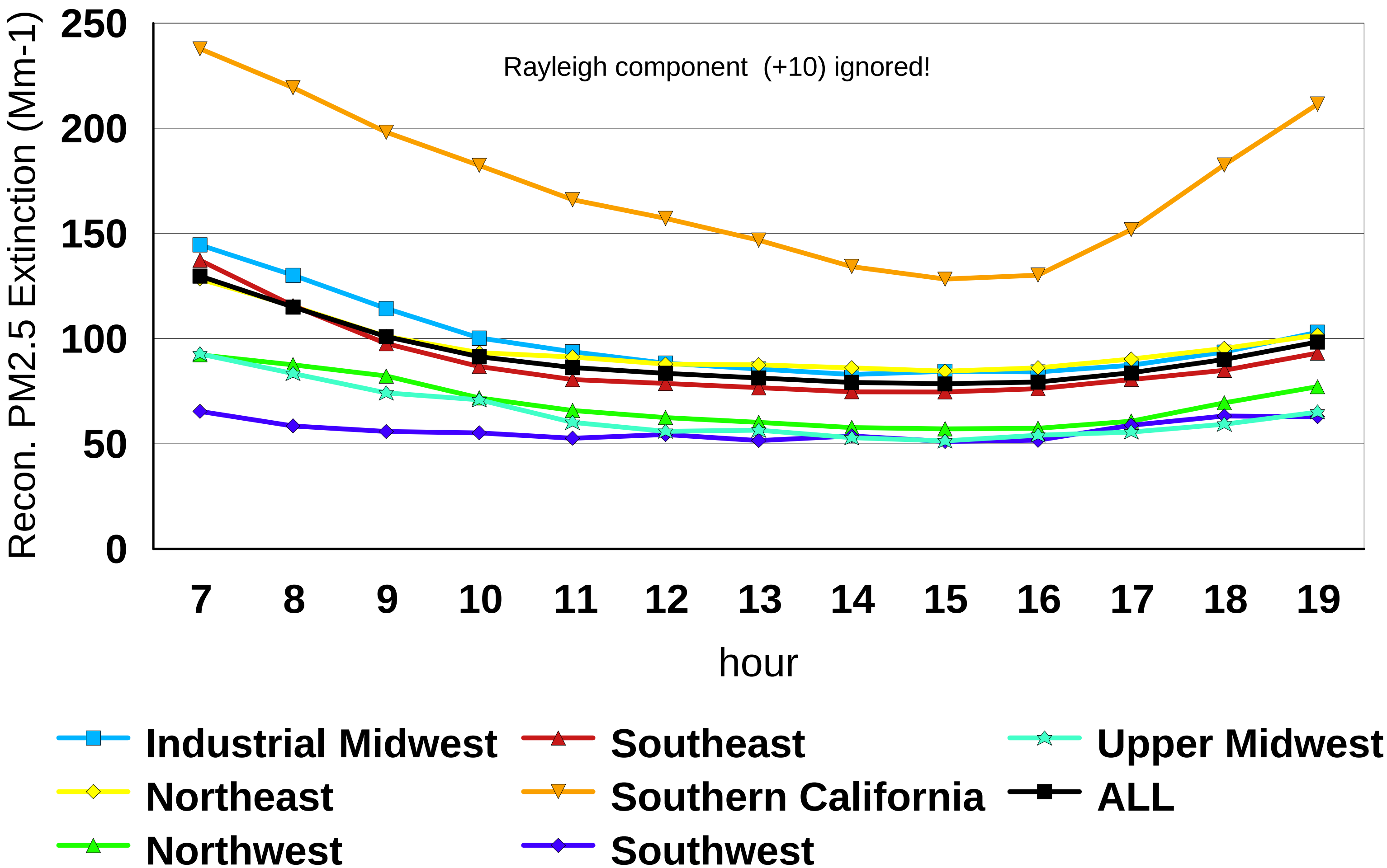
Model slope for relationship between reconstructed light extinction (RE) and hourly PM2.5 (increase in RE due to incremental increase in PM2.5), 2003.
RE computed with actual *f(RH)*.





Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

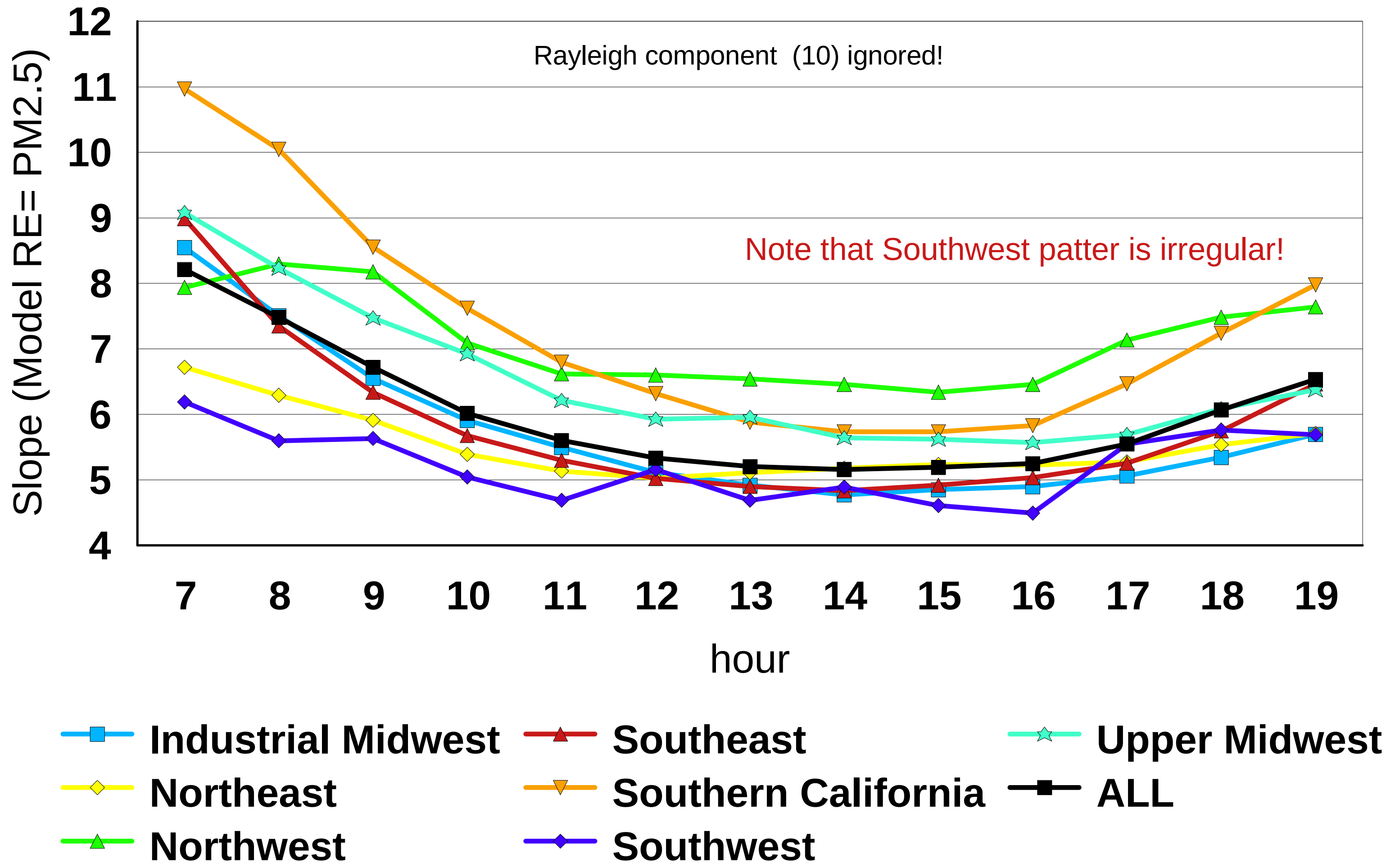
Reconstructed Extinction ('old' db) - using 10-year avg. met and combo PMc*



[*If collocated hourly PM10 was not present (for a site-day), a PMcoarse value was estimated using regional PM2.5-to-PM10 ratios]

Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

Slope (Recon. Extinction = $PM_{2.5}$) - using 10-year avg. met and combo PMc*

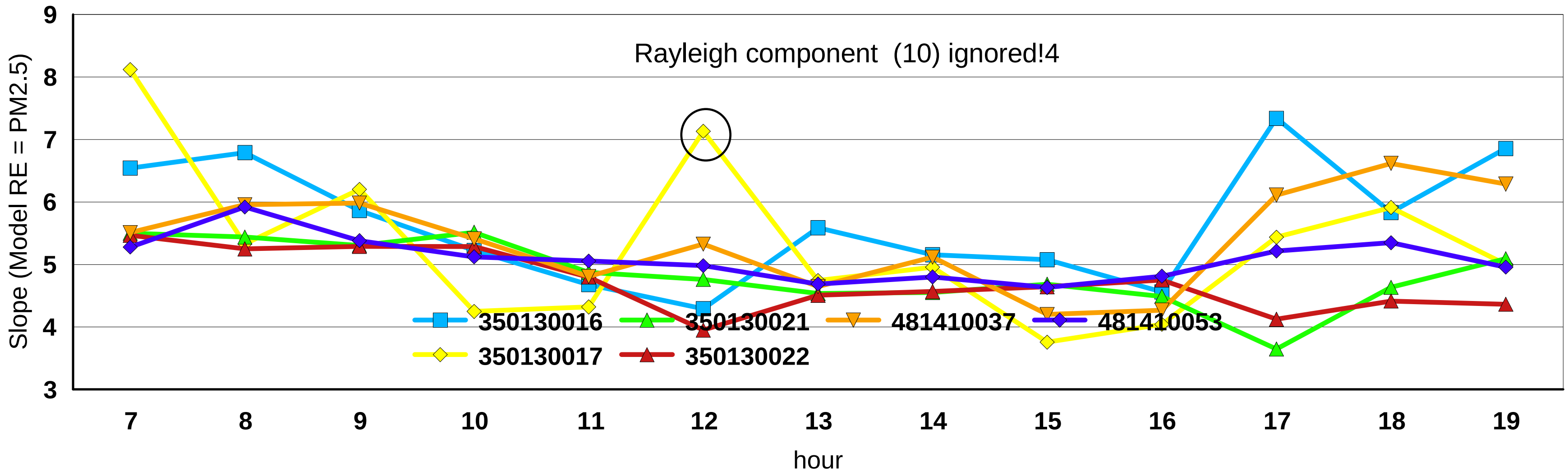


[*If collocated hourly PM10 was not present (for a site-day), a PMcoarse value was estimated using regional PM2.5-to-PM10 ratios]

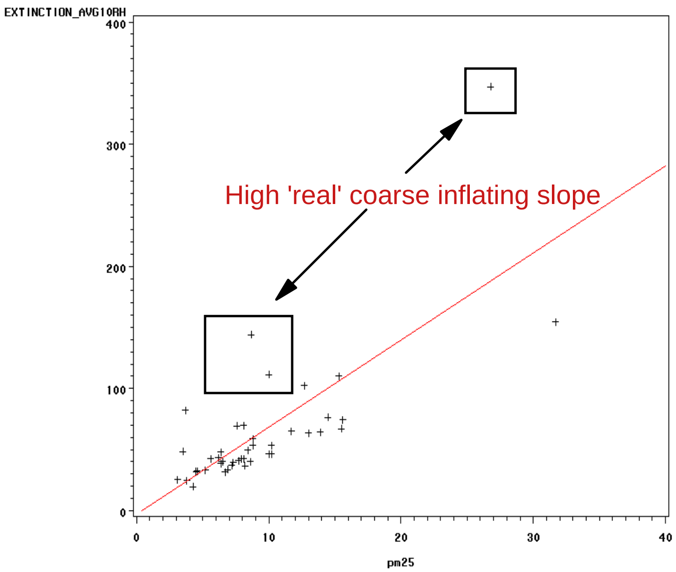
Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

Slope (Recon. Extinction = $PM_{2.5}$) - using combo PMc

Southwest Sites



Site 350130017 (yellow above), Hr=12

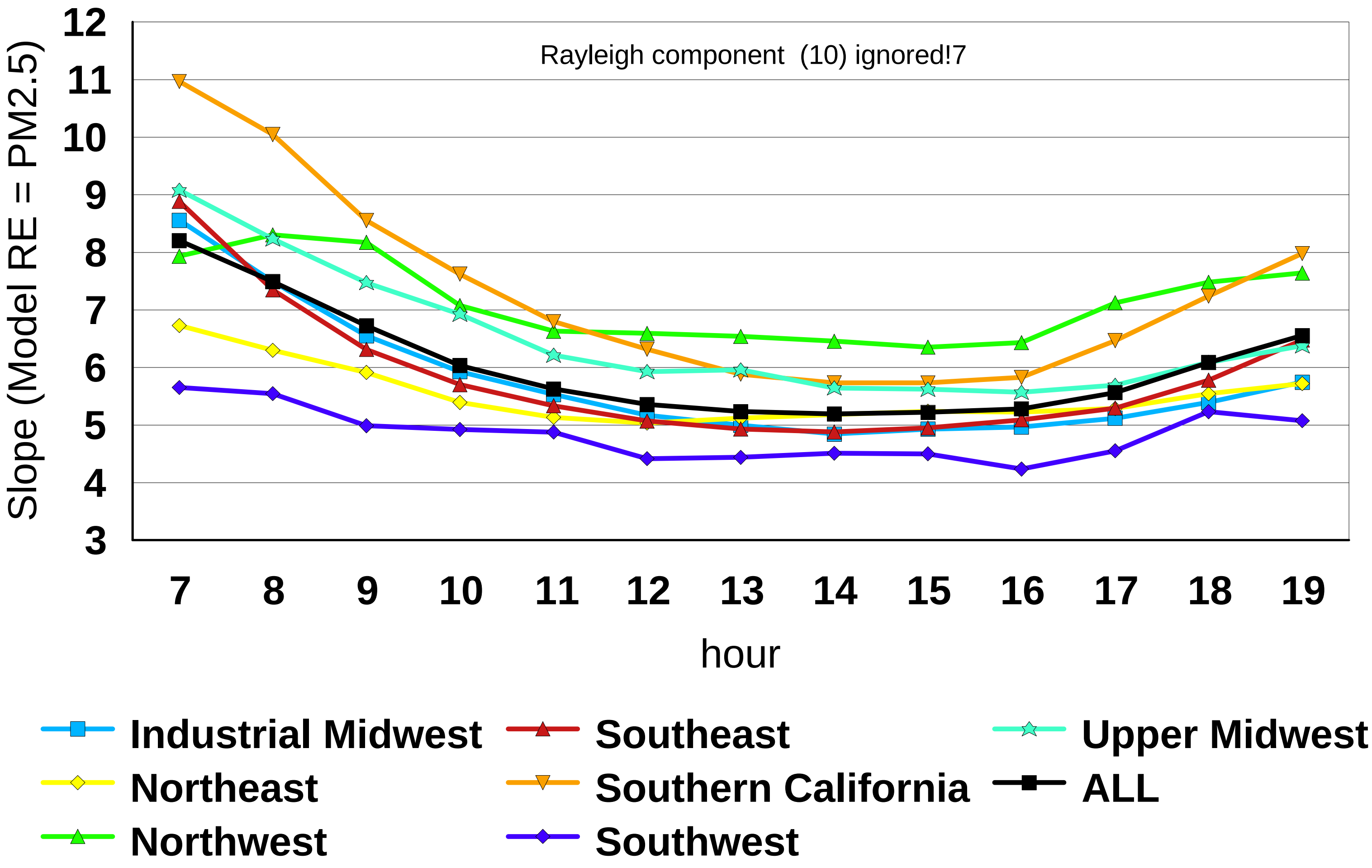


EXTINCTION_AVG10RH=
(3*FRH10AVG*PM25_NITRATES) +
(3*FRH10AVG*PM25_SULFATES) +
(4*PM25_OCM) +
(10*PM25_EC) +
(PM25_CRUSTAL) +
(.6* PMC);

Rayleigh component (10) ignored!

Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

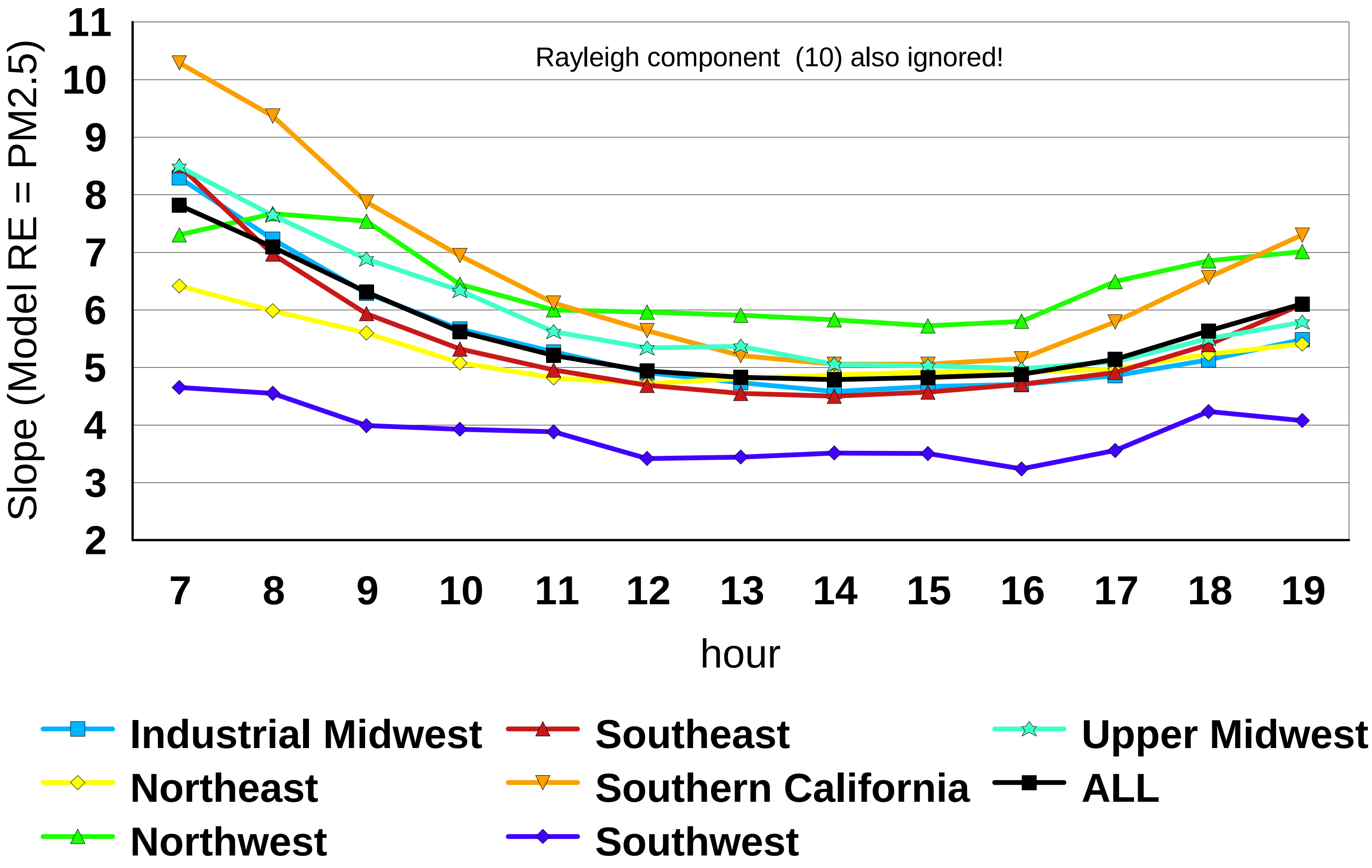
Slope (Recon. Extinction=$PM_{2.5}$) - using PMC from Reg. Ratios*



[* All PMcoarse data were estimated using regional PM2.5-to-PM10 ratios]

Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

Slope (Recon. Extinction = $PM_{2.5}$) - ignoring PMc component

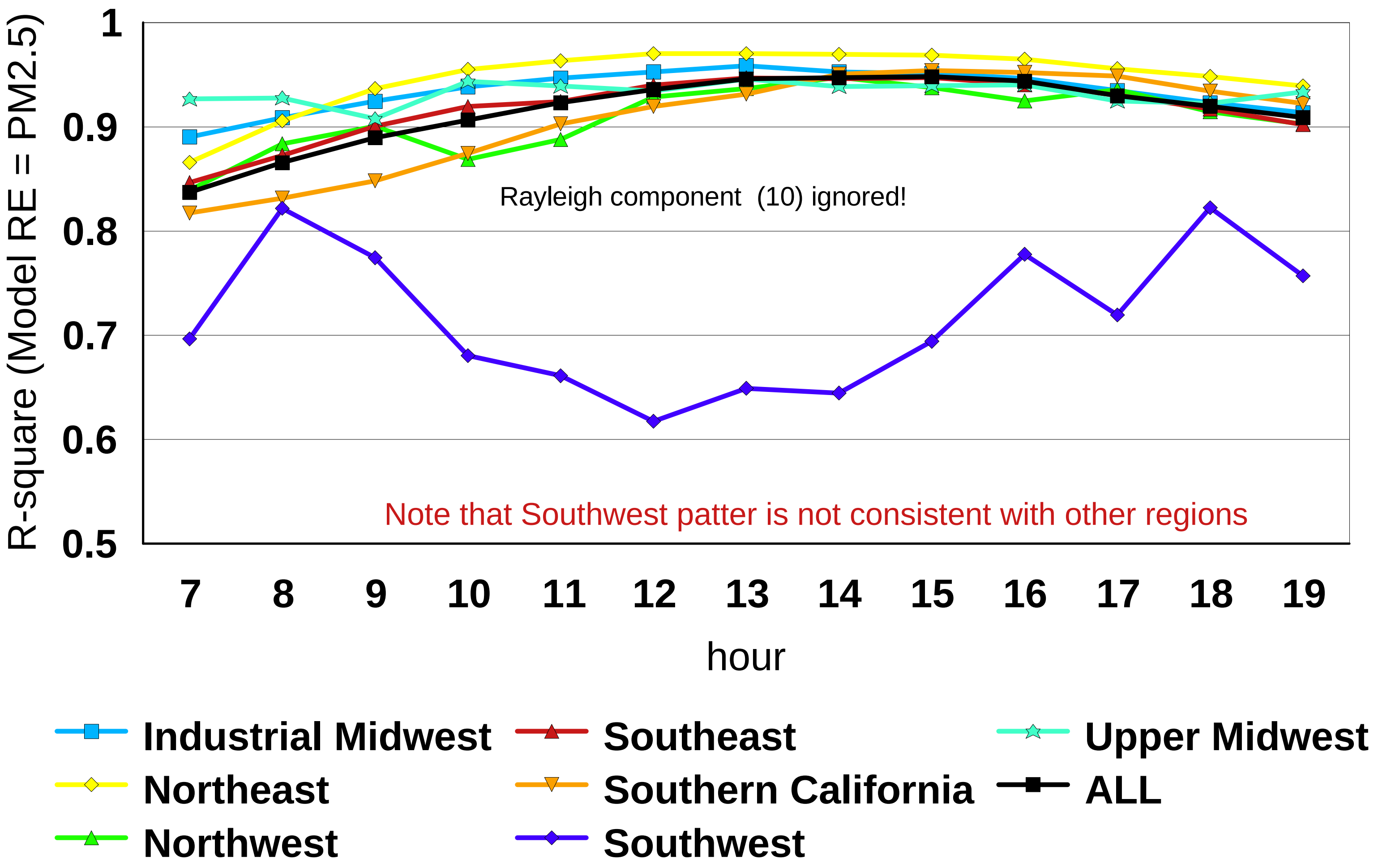


EXTINCTION_AVG10RH=
(3*FRH10AVG*PM25_NITRATES) +
(3*FRH10AVG*PM25_SULFATES) +
(4*PM25_OCM) +
(10*PM25_EC) +
(PM25_CRUSTAL) +
~~(.6* PMC);~~

Rayleigh component (10) ignored!

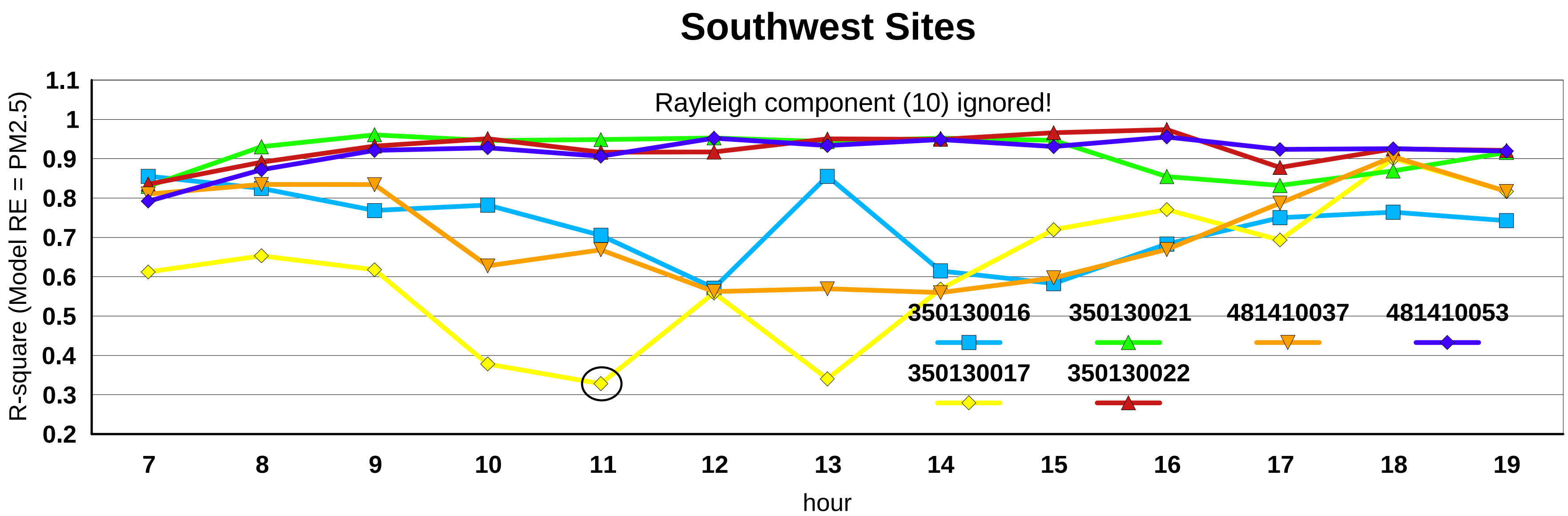
Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

R^2 (Recon. Extinction = $PM_{2.5}$) - using combo PMc

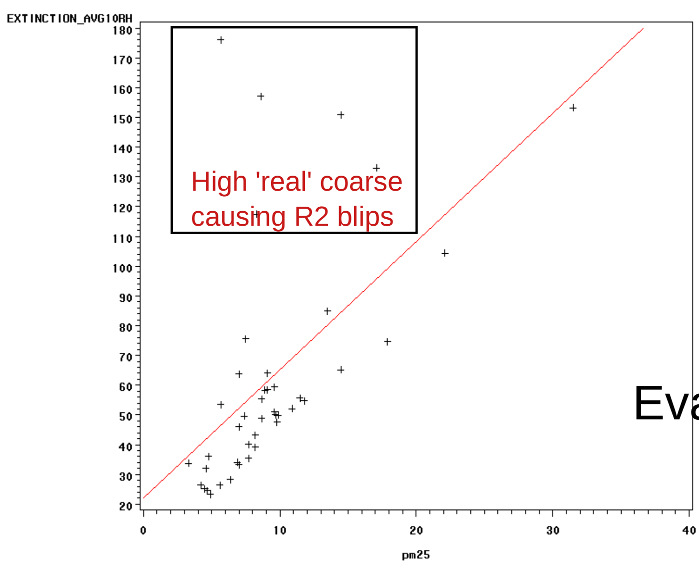


Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

R^2 (Recon. Extinction = $PM_{2.5}$) - <u>using combo PMc</u>



Site 350130017 (yellow above), Hr=11

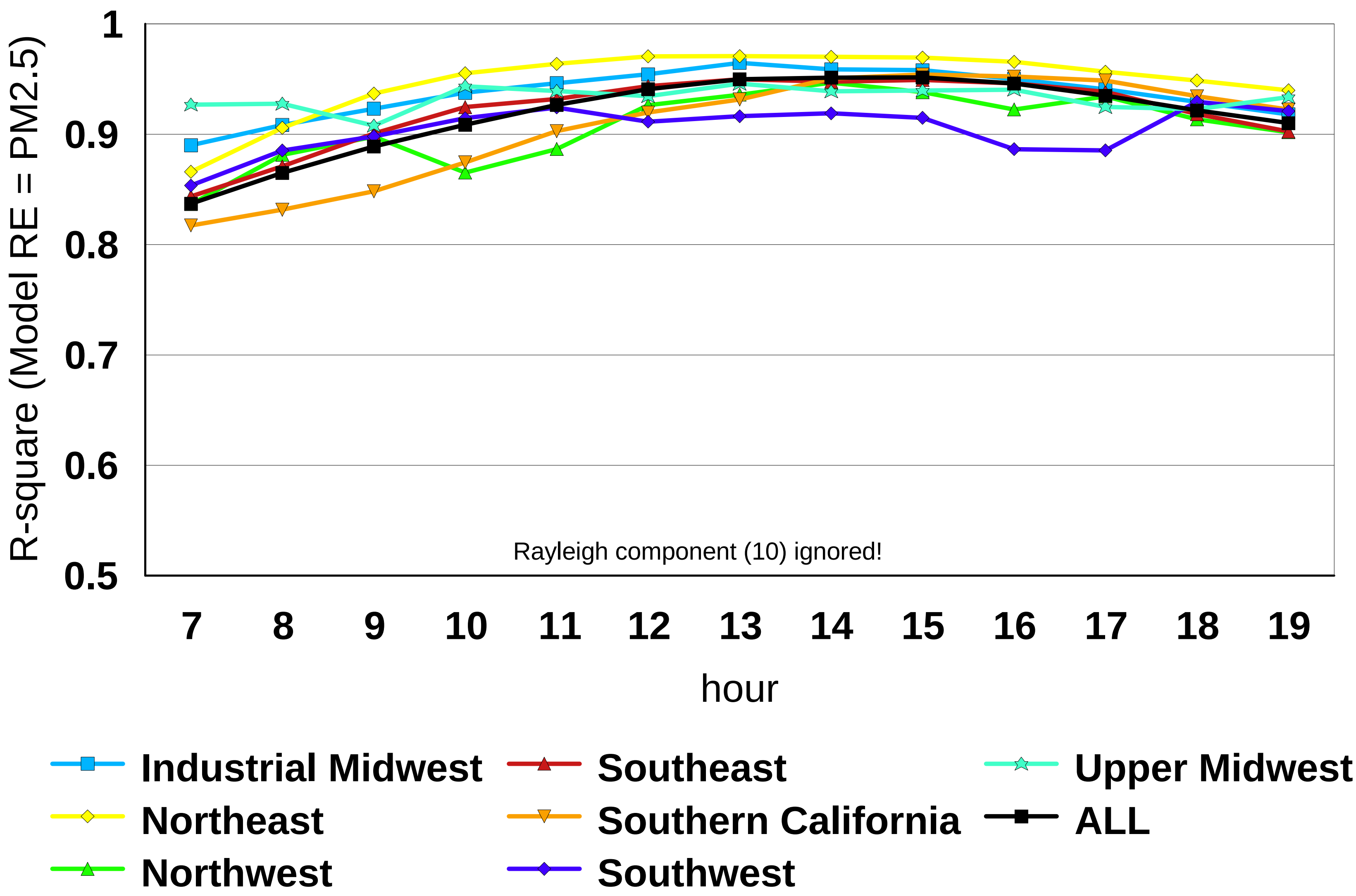


EXTINCTION_AVG10RH=
(3*FRH10AVG*PM25_NITRATES) +
(3*FRH10AVG*PM25_SULFATES) +
(4*PM25_OCM) +
(10*PM25_EC) +
(PM25_CRUSTAL) +
(.6* PMC);

Rayleigh component (10) ignored!

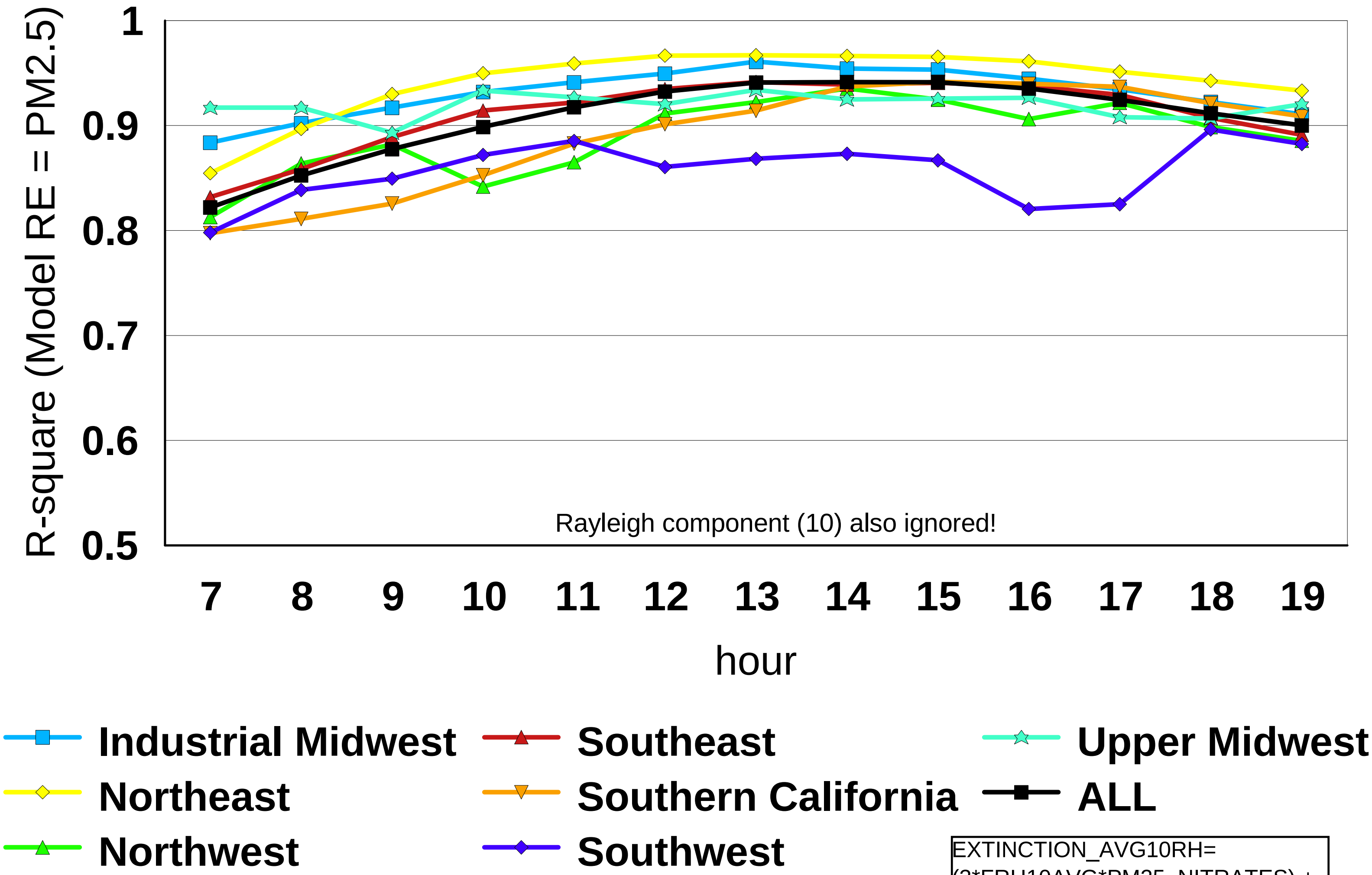
Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

R^2 (Recon. Extinction = $PM_{2.5}$) - using PMc from Reg. Ratios



Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

R^2 (Recon. Extinction = $PM_{2.5}$) - <u>ignoring PMc component</u>



EXTINCTION_AVG10RH=
(3*FRH10AVG*PM25_NITRATES) +
(3*FRH10AVG*PM25_SULFATES) +
(4*PM25_OCM) +
(10*PM25_EC) +
(PM25_CRUSTAL) +
~~(.6* PMC);~~

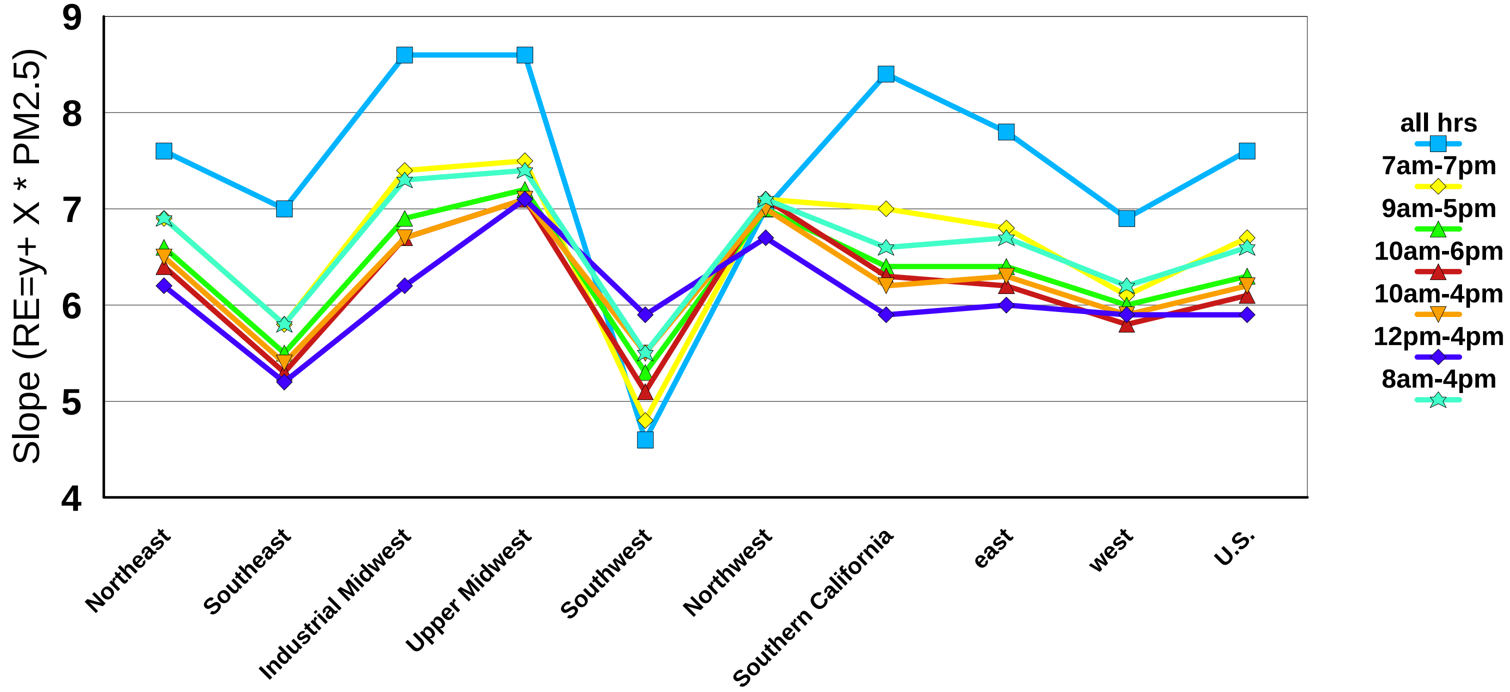
Rayleigh component (10) ignored!

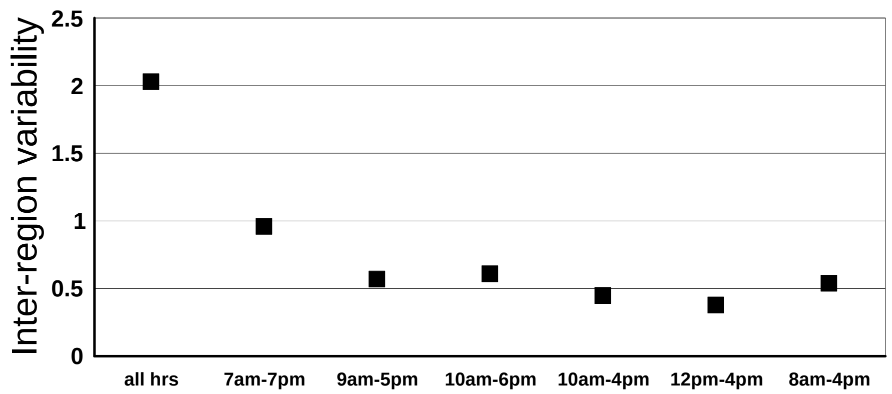
Evaluation of diurnal patterns -'old' (4/'02 -3/'03) db

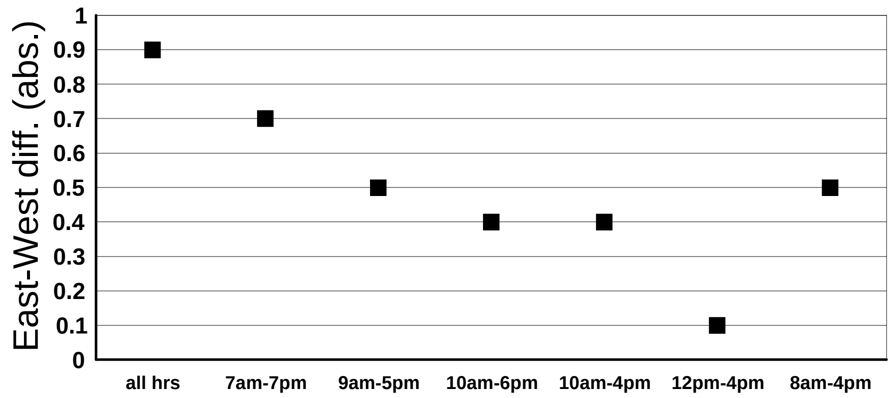
Comparison of Different Averaging Periods – 2003 db 10-year average *f(RH)*

				All hours (all RH)						
Timeframe	# Hours	Min. Hrs Rqrd	Area	Mean 10-yr Avg. RH	Mean 10-yr Avg. FRH	Mean PM2.5	Mean RE	Corr.	Inter-cept	Slope
All hrs	24	18	Northeast	66.2	2.66	14.0	108	0.98	5.5	7.3
			Southeast	71.2	3.05	12.1	98	0.94	12.0	7.1
			Industrial Midwest	71.1	2.88	14.6	118	0.97	8.1	7.5
			Upper Midwest	69.7	2.74	10.2	80	0.96	8.8	7.0
			Southwest	41.0	1.51	11.3	73	0.86	19.1	4.8
			Northwest	67.0	2.84	10.1	76	0.91	9.4	6.6
			Southern California	61.7	2.39	24.1	168	0.91	-37.4	8.5
			east	69.5	2.88	13.4	106	0.97	8.5	7.3
			west	60.1	2.43	12.3	88	0.91	4.5	6.8
			Total	67.8	2.80	13.2	103	0.95	7.8	7.2
			Inter-region variability	113.45	0.27	23.78	1109.6	0.002	346.19	1.30
7a.m. to 7p.m.	12	9	Northeast	60.8	2.28	14.1	99	0.99	10.6	6.3
			Southeast	62.9	2.39	12.0	84	0.96	16.0	5.7
			Industrial Midwest	64.7	2.41	14.3	103	0.97	13.3	6.3
			Upper Midwest	63.6	2.32	9.8	72	0.96	10.1	6.3
			Southwest	34.3	1.29	10.8	70	0.83	17.9	4.8
			Northwest	59.9	2.37	9.2	67	0.92	9.4	6.2
			Southern California	53.7	1.93	22.7	142	0.92	-14.3	6.9
			east	62.7	2.36	13.2	94	0.97	12.4	6.1
			west	53.3	2.04	11.5	78	0.91	11.0	5.9
			Total	61.0	2.30	12.9	91	0.96	12.2	6.1
			Inter-region variability	114.52	0.17	21.17	715.1	0.003	116.14	0.43
9a.m. to 5p.m.	8	6	Northeast	58.2	2.12	13.7	92	0.99	11.3	5.9
			Southeast	59.2	2.10	11.4	74	0.97	14.4	5.2
			Industrial Midwest	62.0	2.21	13.9	95	0.97	13.6	5.8
			Upper Midwest	61.1	2.15	9.5	67	0.97	10.1	6.0
			Southwest	31.4	1.19	9.8	64	0.85	12.6	5.3
			Northwest	57.1	2.17	8.6	61	0.93	9.8	6.0
			Southern California	50.3	1.72	22.4	133	0.94	-7.5	6.3
			east	59.6	2.13	12.8	85	0.98	11.9	5.7
			west	50.4	1.87	10.9	73	0.92	10.5	5.7
			Total	57.9	2.08	12.5	83	0.96	11.6	5.7
			Inter-region variability	115.84	0.14	22.29	646.8	0.002	57.04	0.15
10a.m. to 6p.m.	8	6	Northeast	57.1	2.07	13.8	92	0.99	11.4	5.8
			Southeast	57.2	1.99	11.3	72	0.97	13.6	5.1
			Industrial Midwest	60.5	2.13	13.8	92	0.97	13.7	5.7
			Upper Midwest	59.4	2.06	9.3	64	0.97	10.2	5.9
			Southwest	29.3	1.15	10.2	66	0.86	13.9	5.1
			Northwest	55.1	2.06	8.5	59	0.94	9.7	5.8
			Southern California	48.7	1.65	22.0	128	0.95	-6.1	6.1
			east	58.0	2.05	12.8	83	0.98	11.6	5.6
			west	48.6	1.79	10.8	71	0.92	11.2	5.5
			Total	56.3	2.00	12.4	81	0.96	11.6	5.6
			Inter-region variability	119.08	0.13	21.09	583.6	0.002	50.35	0.15
10a.m. to 4p.m.	6	4	Northeast	57.3	2.07	13.6	90	0.99	11.5	5.8
			Southeast	57.9	2.01	11.2	71	0.97	13.6	5.1
			Industrial Midwest	61.0	2.15	13.7	92	0.97	13.4	5.7
			Upper Midwest	60.1	2.08	9.3	65	0.97	10.2	5.9
			Southwest	30.3	1.16	9.6	64	0.85	10.6	5.6
			Northwest	55.6	2.08	8.3	59	0.93	10.2	5.8
			Southern California	49.1	1.66	22.3	130	0.94	-4.5	6.0
			east	58.5	2.06	12.6	83	0.98	11.7	5.6
			west	49.2	1.80	10.6	70	0.92	10.1	5.7
			Total	56.8	2.01	12.2	80	0.96	11.3	5.6
			Inter-region variability	115.45	0.13	22.80	626.3	0.002	38.84	0.08
12p.m. to 4p.m.	4	3	Northeast	55.4	1.99	13.7	89	0.99	12.0	5.6
			Southeast	55.0	1.88	11.1	68	0.98	13.0	5.0
			Industrial Midwest	58.8	2.03	13.7	89	0.98	13.5	5.5
			Upper Midwest	57.4	1.96	9.1	62	0.98	9.8	5.7
			Southwest	27.6	1.12	10.3	68	0.85	7.3	5.9
			Northwest	52.8	1.93	8.0	55	0.95	10.6	5.6
			Southern California	46.4	1.55	21.4	121	0.96	-0.9	5.7
			east	56.1	1.95	12.6	80	0.98	11.7	5.4
			west	46.4	1.68	10.5	68	0.90	8.6	5.7
			Total	54.3	1.90	12.2	78	0.96	10.8	5.5
			Inter-region variability	118.26	0.11	20.19	508.8	0.002	25.06	0.08
8a.m. to 4p.m.	8	6	Northeast	60.2	2.23	13.8	95	0.99	11.0	6.1
			Southeast	62.4	2.29	11.7	80	0.96	15.6	5.5
			Industrial Midwest	64.3	2.35	14.2	100	0.97	13.4	6.1
			Upper Midwest	63.6	2.29	9.9	72	0.97	10.3	6.2
			Southwest	34.3	1.26	9.6	65	0.82	11.4	5.5
			Northwest	59.9	2.34	8.9	65	0.92	9.5	6.2
			Southern California	52.9	1.86	23.0	141	0.92	-9.9	6.6
			east	62.1	2.29	13.0	90	0.98	12.3	6.0
			west	53.2	2.01	11.1	76	0.91	9.6	6.0
			Total	60.5	2.24	12.6	88	0.96	11.7	6.0
			Inter-region variability	112.73	0.16	23.43	742.5	0.003	72.12	0.15

Comparison of Slopes (RE=y + X * PM2.5) for Select Timeframes, by Region Using actual FRH

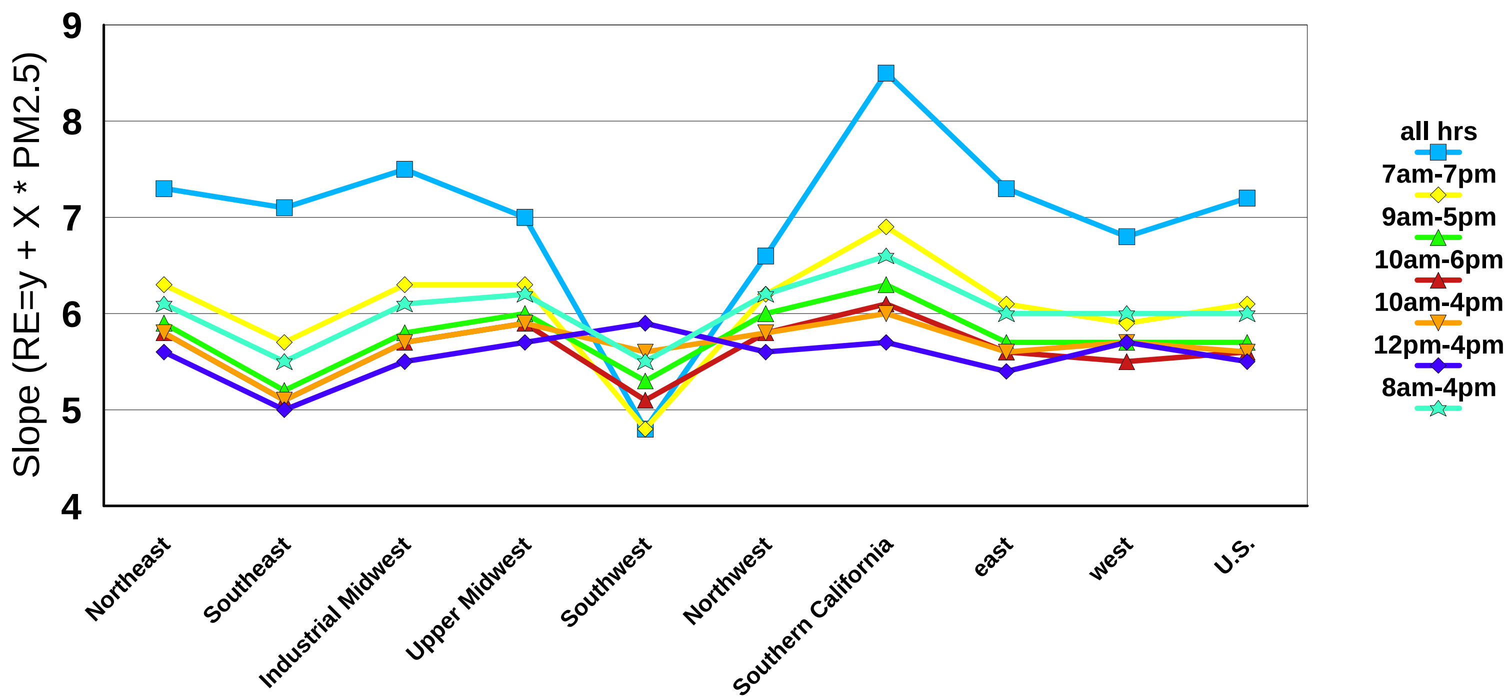


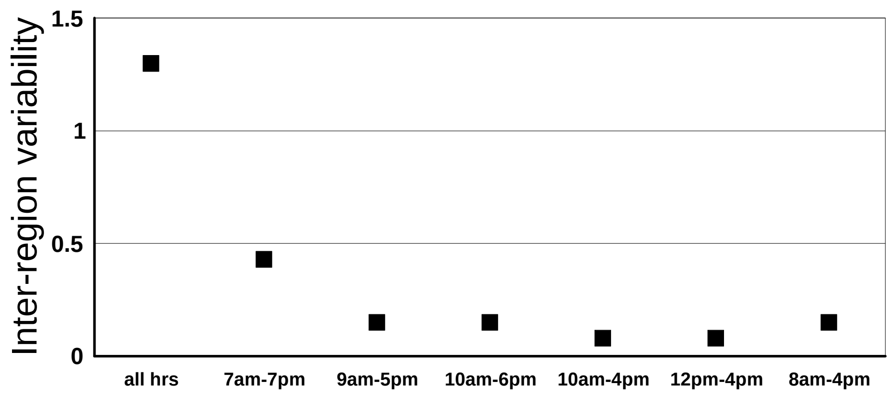


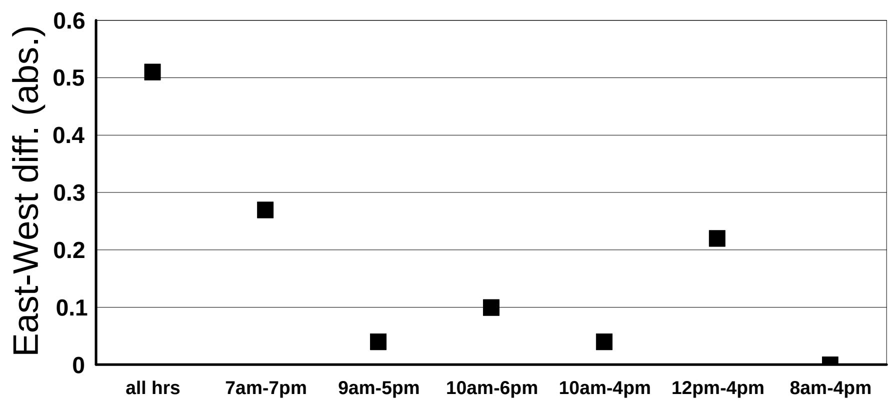


Source: 2003 visibility database.

Comparison of Slopes (RE=y + X * PM2.5) for Select Timeframes, by Region Using 10-Year Avg. FRH

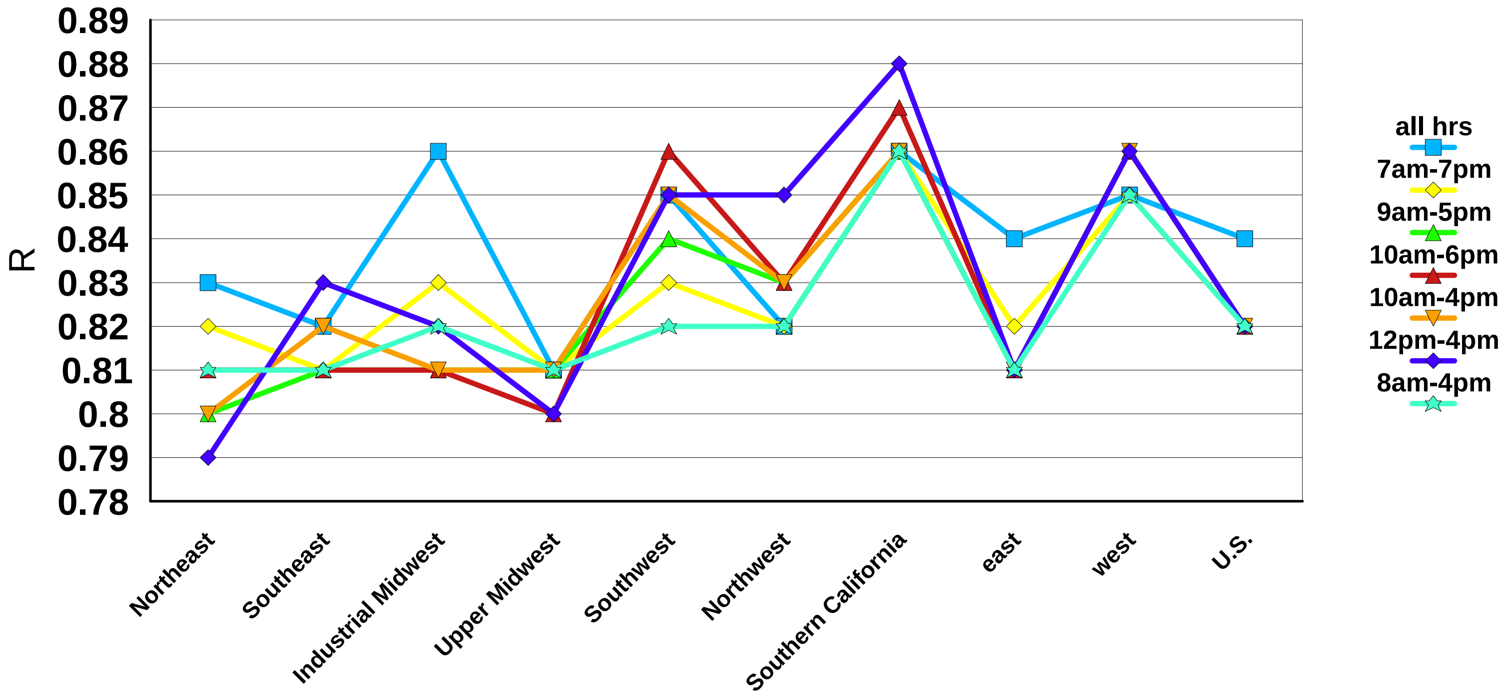


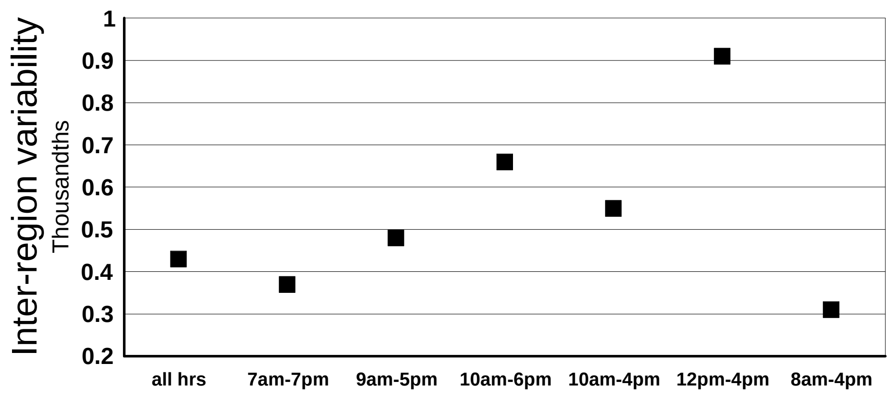


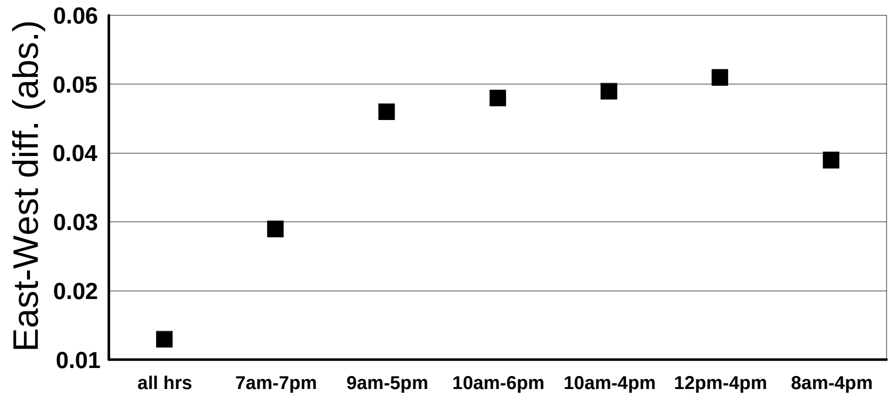


Source: 2003 visibility database.

Comparison of Correlations (RE vs. PM2.5) for Select Timeframes, by Region Using Actual FRH

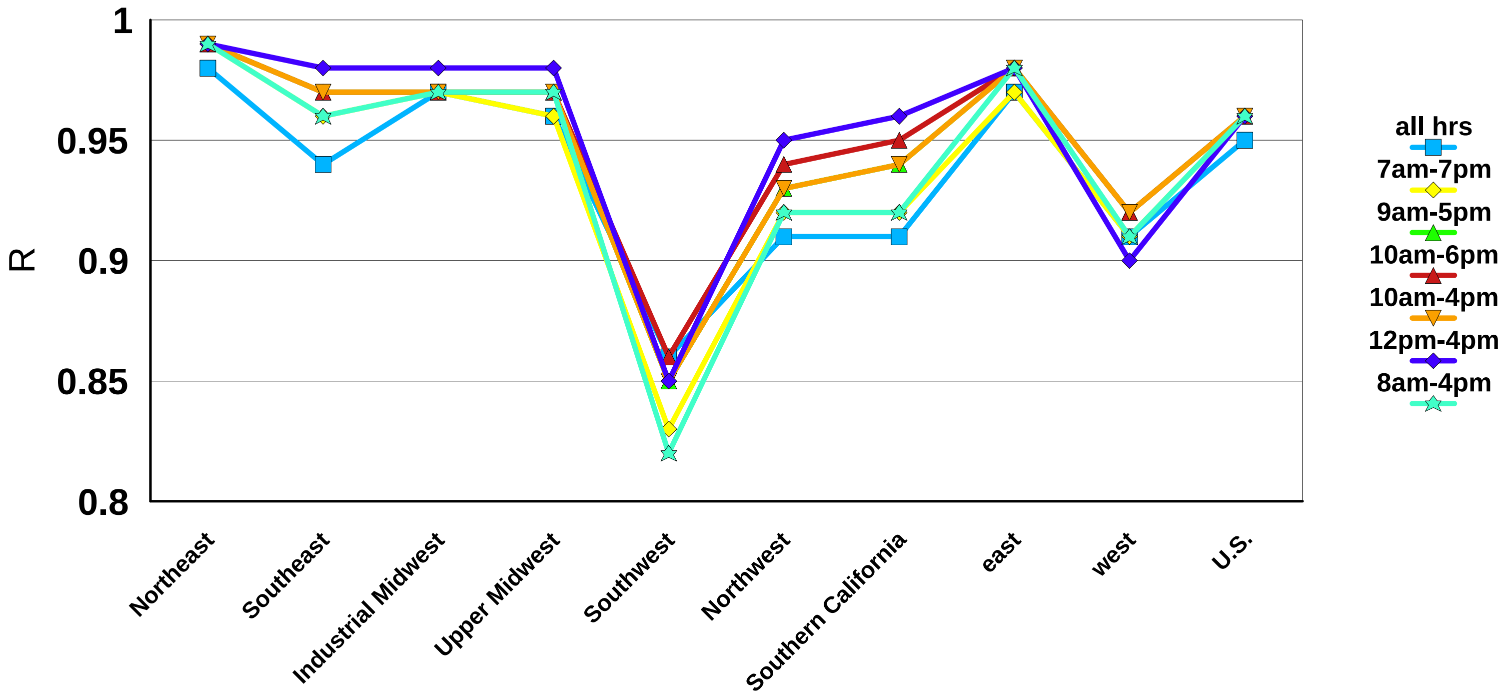


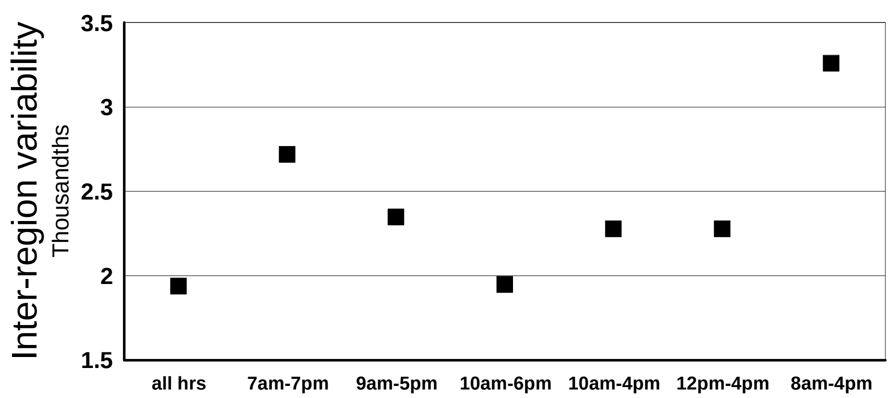


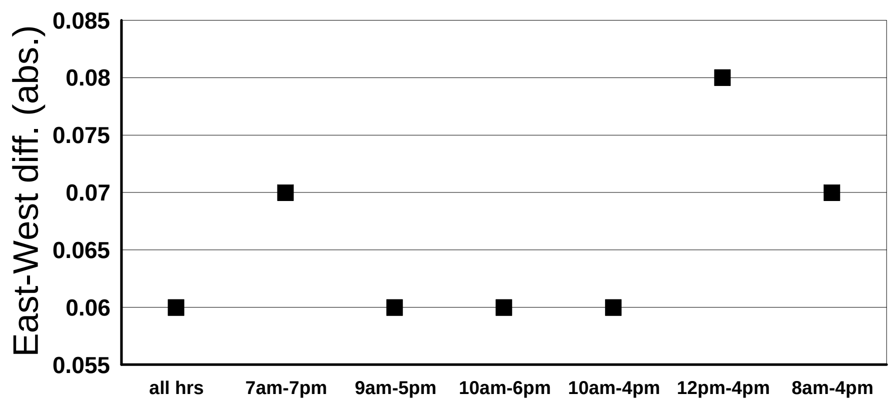


Source: 2003 visibility database.

Comparison of Correlations (RE vs. PM2.5) for Select Timeframes, by Region Using 10-Year Avg. FRH







Source: 2003 visibility database.

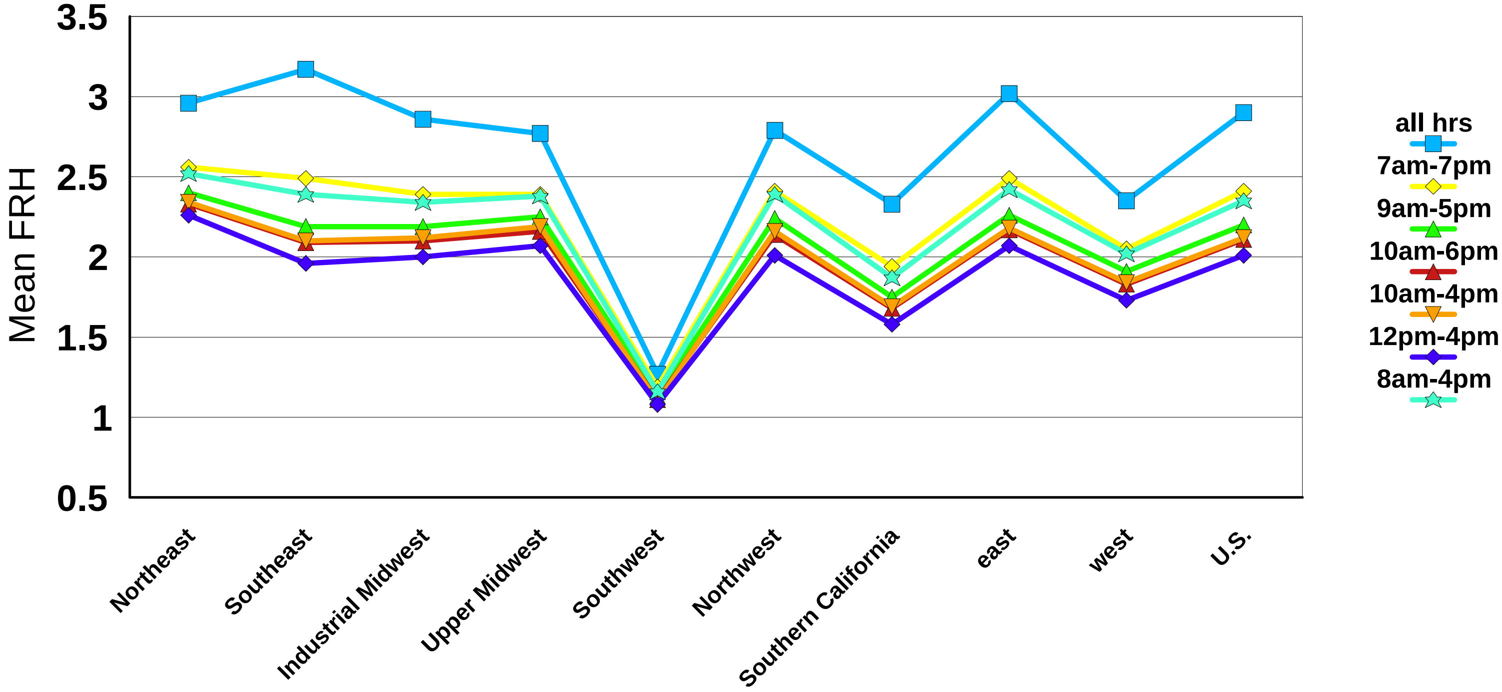
Comparison of Different Averaging Periods – 2003 db, using actual *f(RH)*

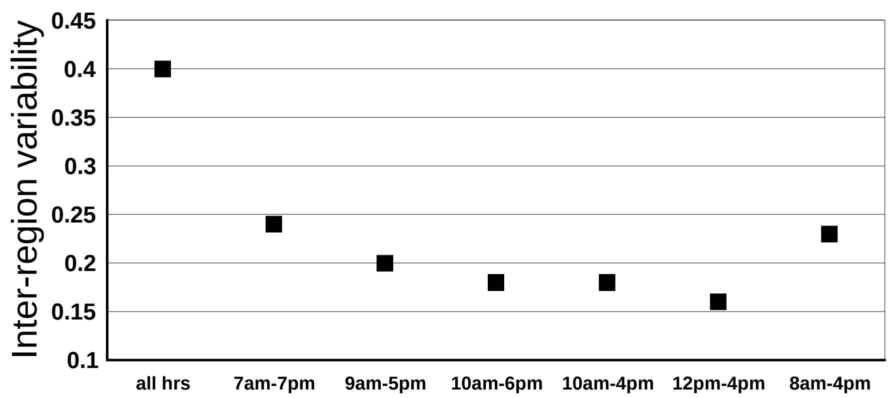
Ranking of time-periods based on slope and r-square

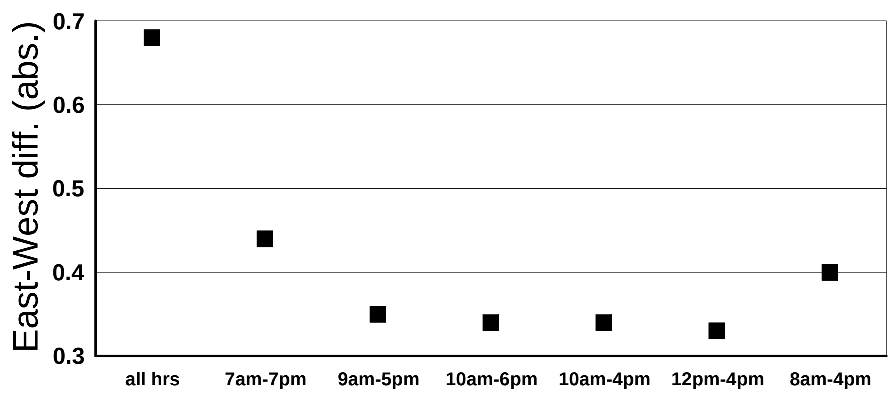
	national slope		inter-region slope variability		east-west slope diff		national corr		inter-region corr variability		east-west corr diff		
	national slope	rank (low to high)	inter-region slope variability	rank (low to high)	east-west slope diff	rank (low to high)	national corr	rank (high to low)	inter-region corr variability	rank (low to high)	east-west corr diff	rank (low to high)	6-Factor Average
all hrs	7.6	7	2.03	7	0.9	7	0.84	1	0.000	3	0.01	1	4.3
7am–7pm	6.7	6	0.96	6	0.7	6	0.82	2	0.000	2	0.03	2	4.0
9am-5pm	6.3	4	0.57	4	0.5	4	0.82	5	0.000	4	0.05	4	4.2
10am-6pm	6.1	2	0.61	5	0.4	3	0.82	4	0.001	6	0.05	5	4.2
10am-4pm	6.2	3	0.45	2	0.4	2	0.82	7	0.001	5	0.05	6	4.2
12pm-4pm	5.9	1	0.38	1	0.1	1	0.82	3	0.001	7	0.05	7	3.3
8am-4pm	6.6	5	0.54	3	0.5	5	0.82	6	0.000	1	0.04	3	3.8

Based on this analysis, the 12-4 period is 'best'!

Comparison of RH Effects (Avg. FRH) for Select Timeframes, by Region Using Actual RH/FRH

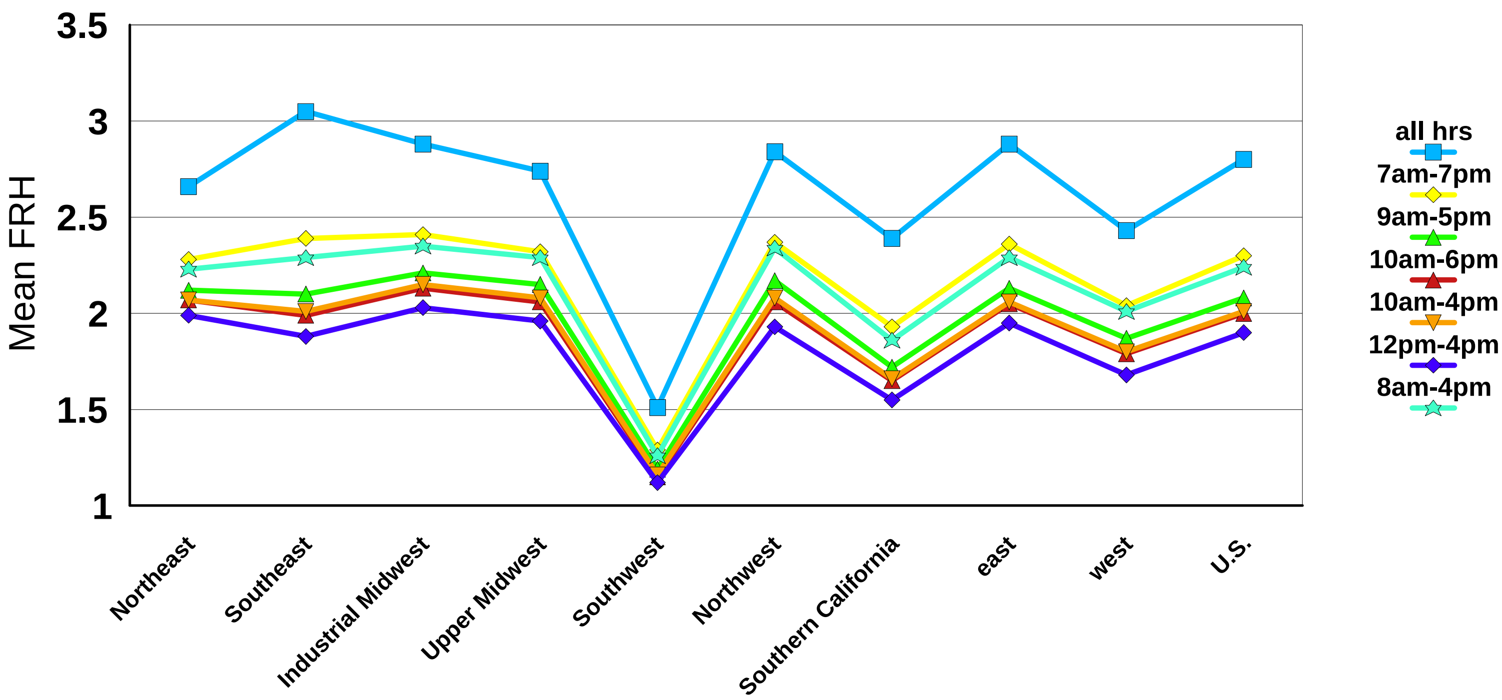


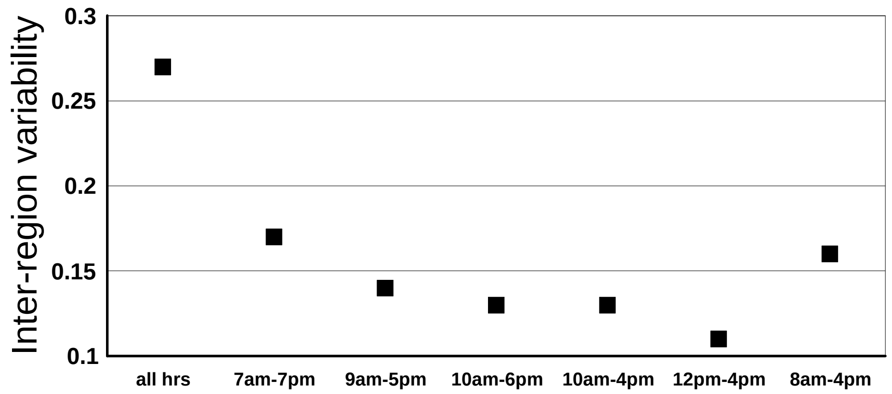


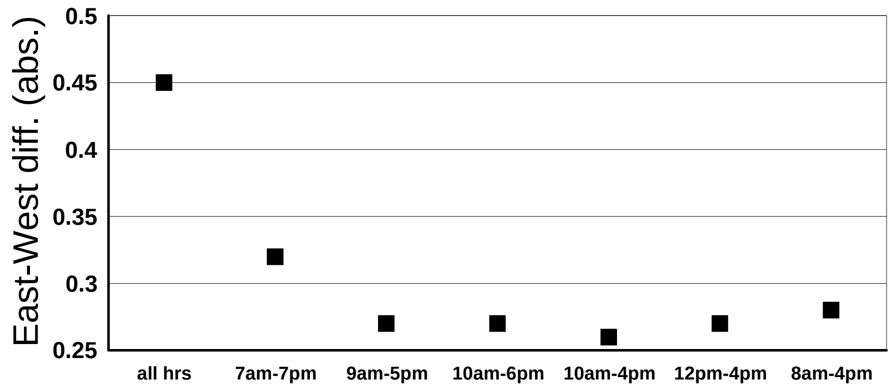


Source: 2003 visibility database.

Comparison of RH Effects (Avg. FRH) for Select Timeframes, by Region Using 10-Year Avg. FRH

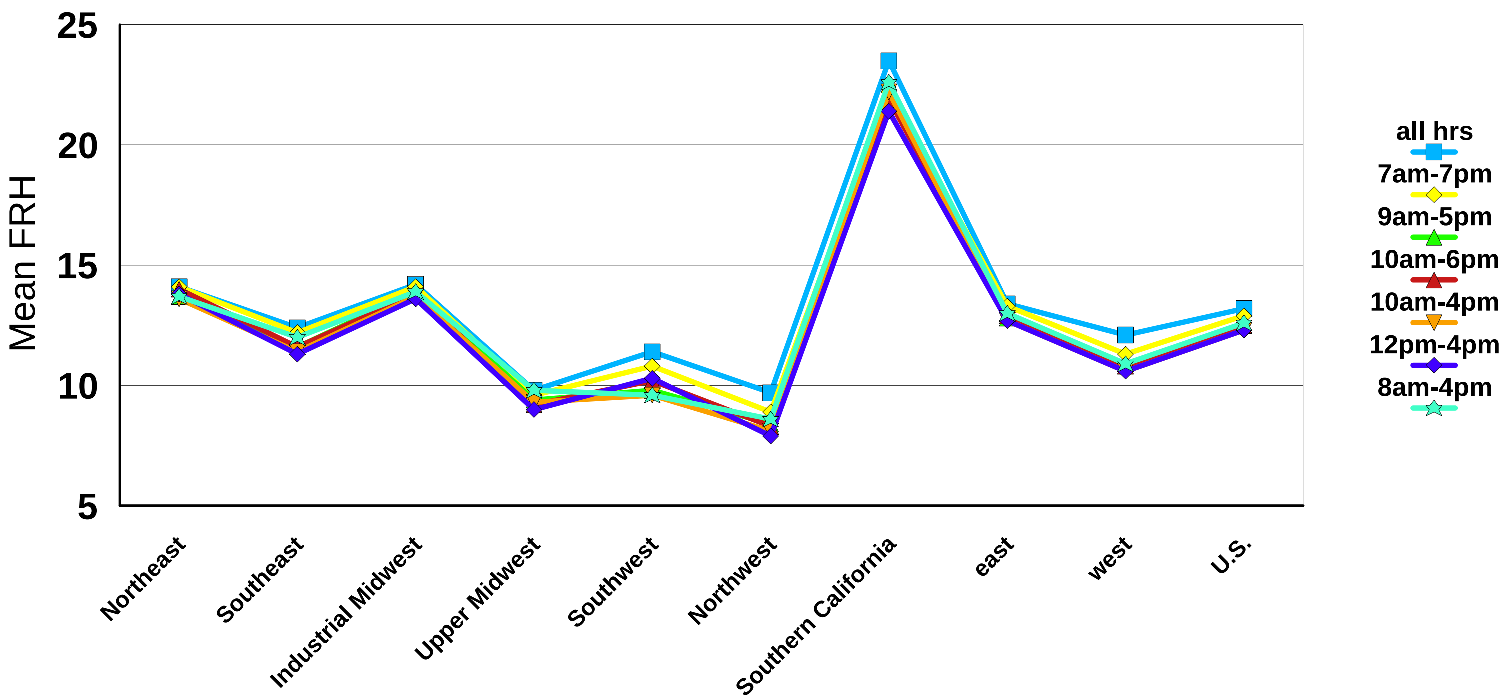


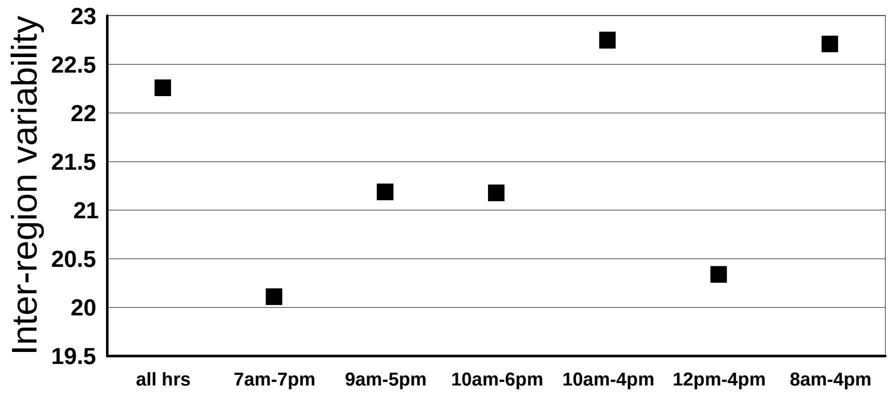


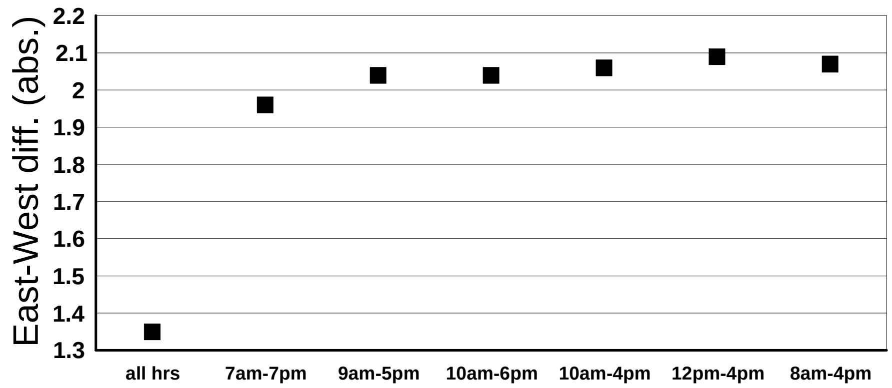


Source: 2003 visibility database.

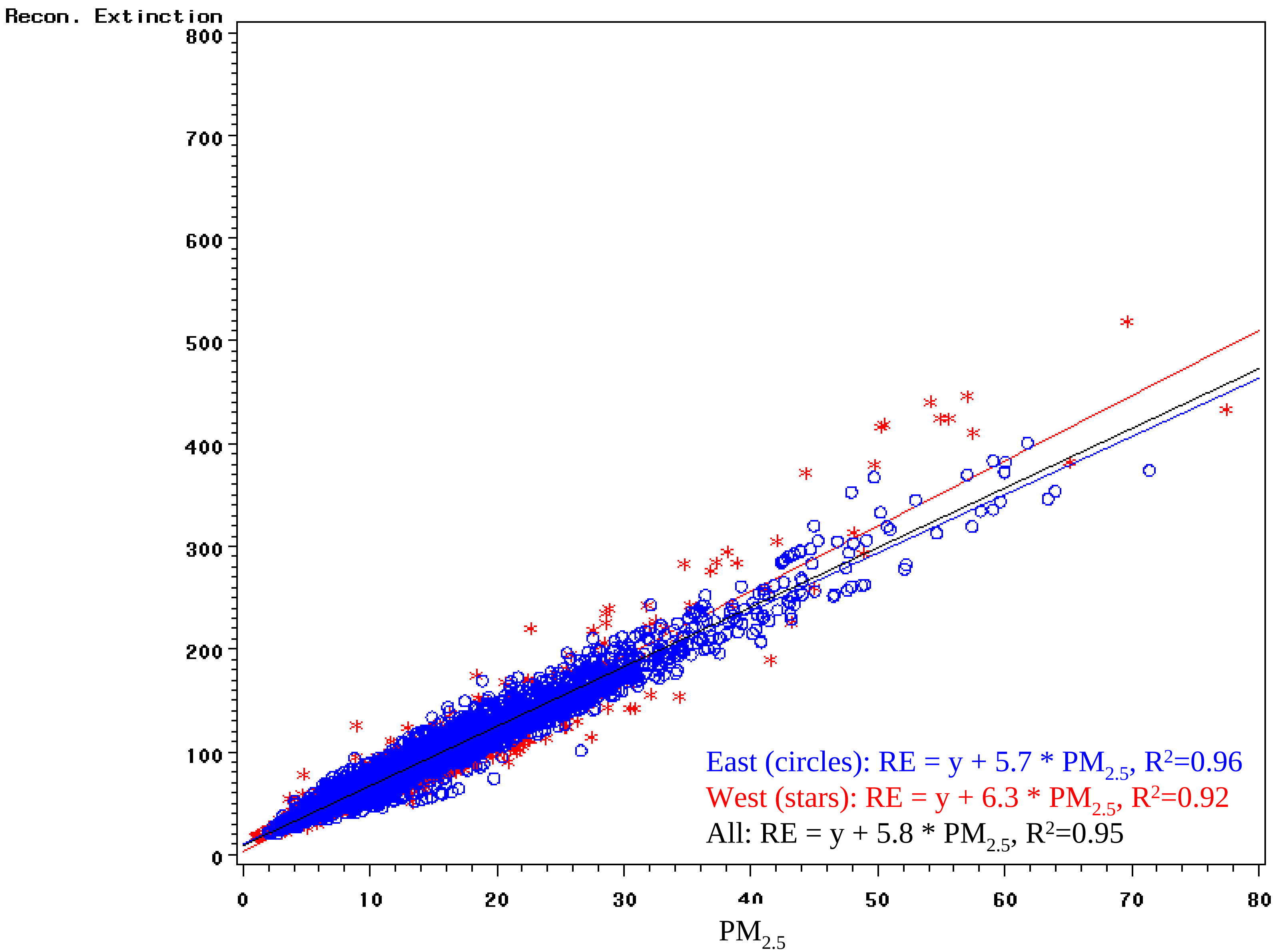
Comparison of PM2.5 Select Timeframes, by Region



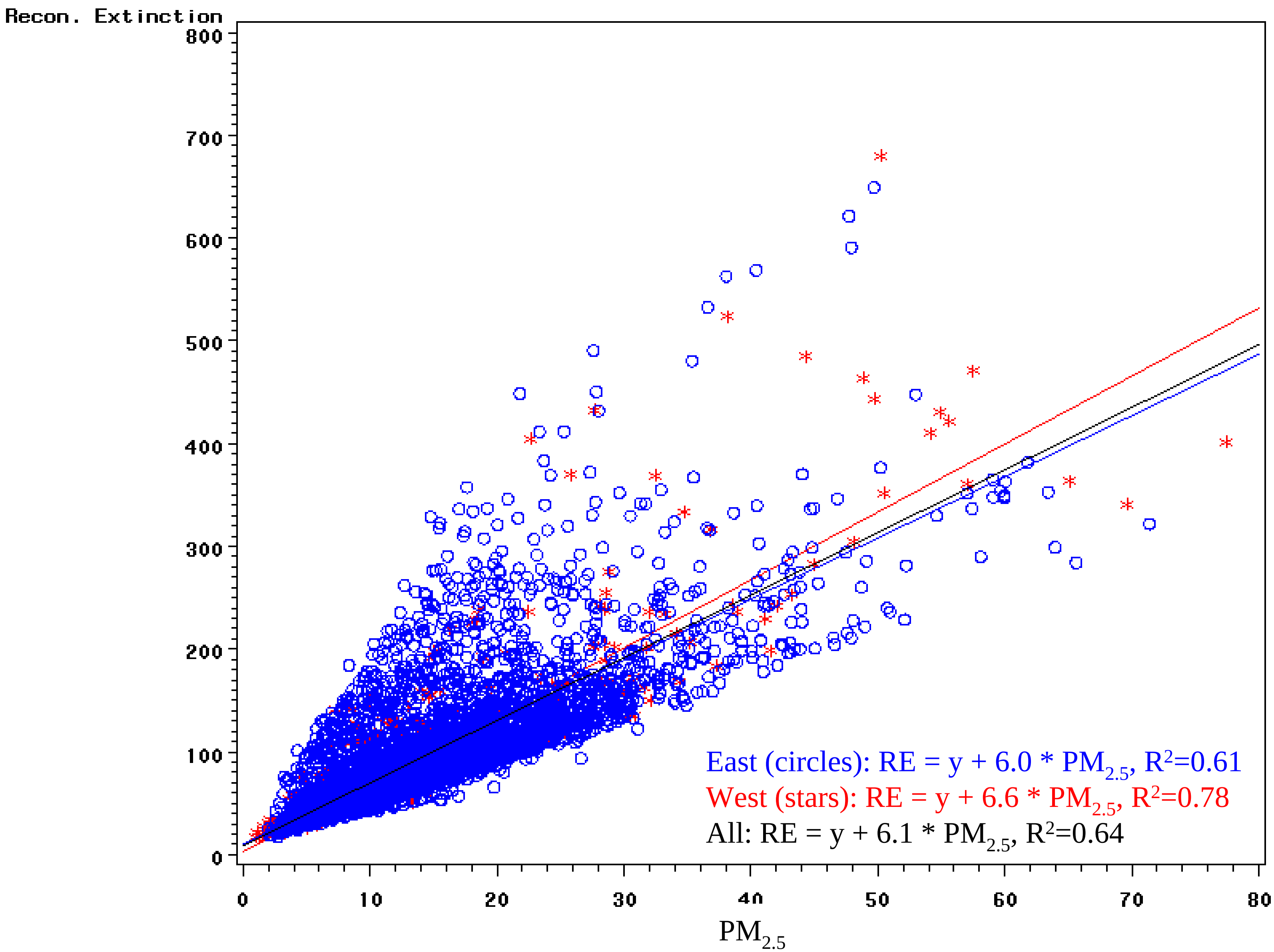




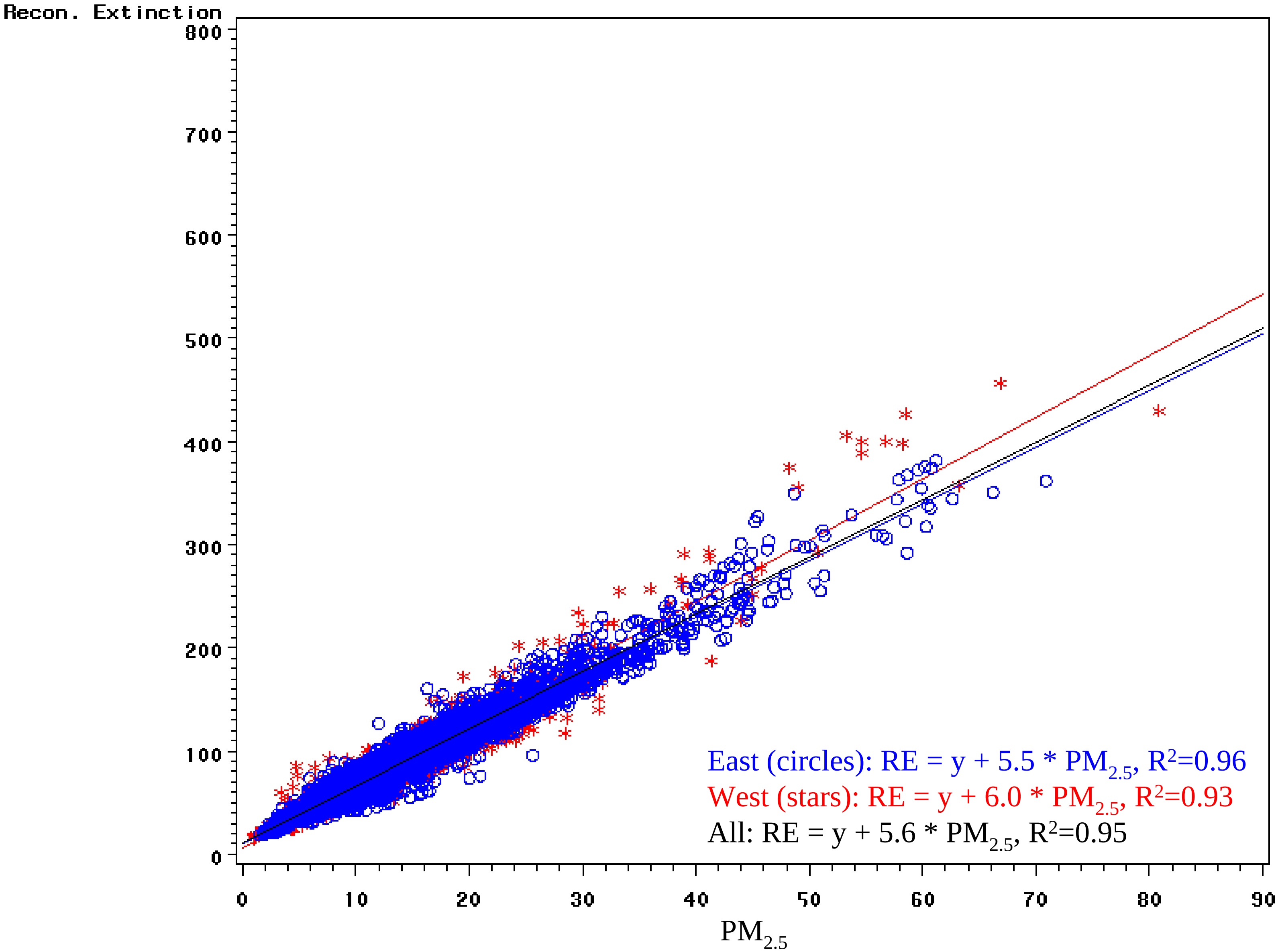
Source: 2003 visibility database.



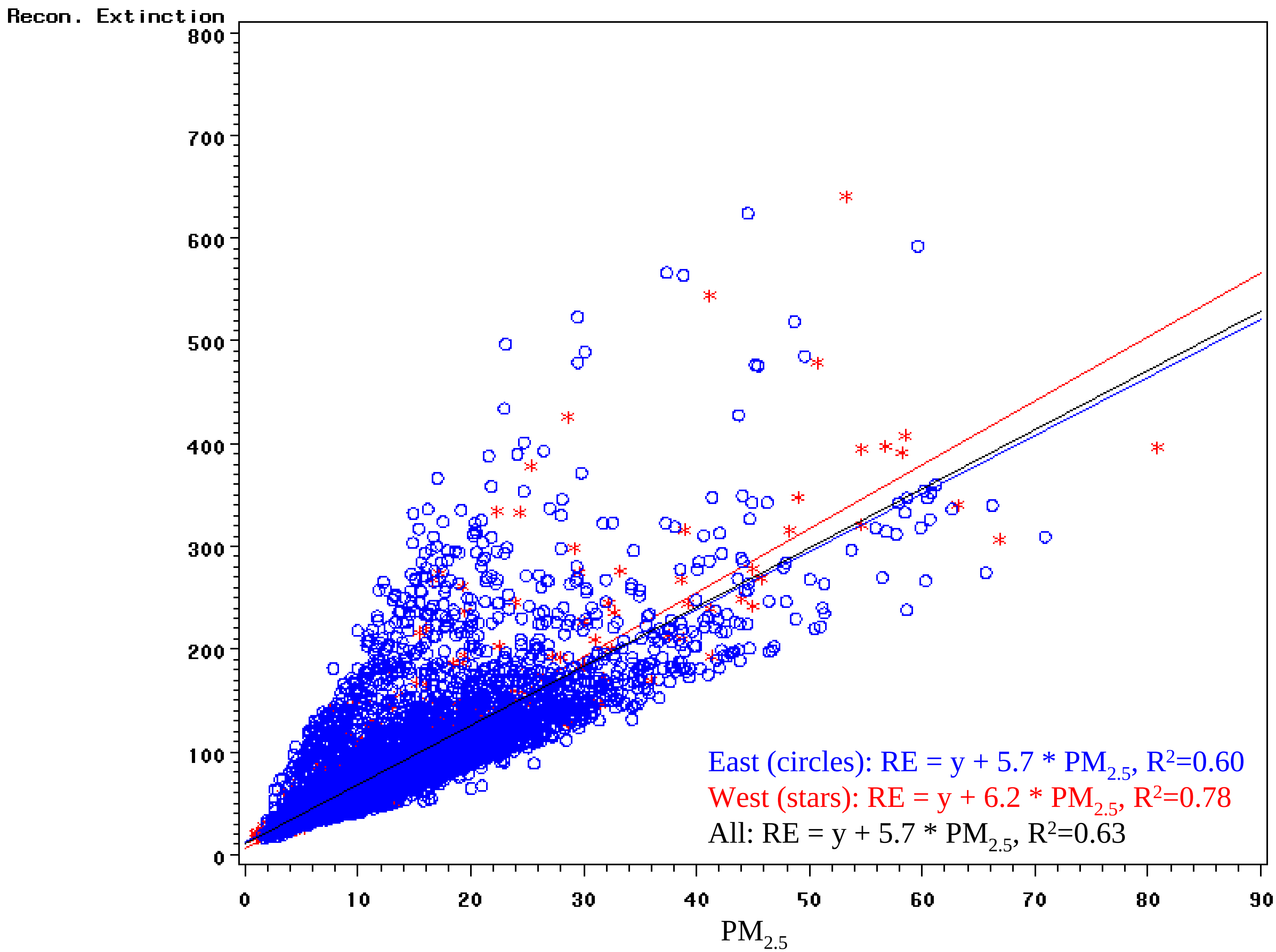
Relationship between reconstructed light extinction (RE) and 10 a.m. – 6 p.m. average $PM_{2.5}$, 2003.
Using 10-year average *f(RH)*.



Relationship between reconstructed light extinction (RE) and 10 a.m. – 6 p.m. average $PM_{2.5}$, 2003.
Using actual *f(RH)*



Relationship between reconstructed light extinction (RE) and 12 p.m. – 4 p.m. average $PM_{2.5}$, 2003.
Using 10-year average *f(RH)*.



Relationship between reconstructed light extinction (RE) and 12 p.m. – 4 p.m. average $PM_{2.5}$, 2003.
Using actual *f(RH)*.

Correlations for 12-4pm Average PM2.5 Concentration and Reconstructed Extinction, 2003

Area	N	Correlation
All Regions	97,563	.767
East	80,413	.746
West	17,150	.848
Northeast	26,592	.759
Southeast	32,791	.710
Ind. Midwest	19,142	.748
Upper Midwest	1,888	.786
Southwest	3,112	.793
Northwest	10,789	.822
Southern CA	3,249	.831

Source: 2003 visibility database.

SP Visibility Analyses

‘Level of Secondary Standard’

Goal: Help inform decision regarding 'level of standard'

- Outputs: Show/summarize $PM_{2.5}$ levels needed to meet various visual range goals
- Focus:
  - 24-hr & 12pm-4pm timeframes
  - VR's of 10, 20, 30, 40, 50, 60 kilometers
- Method: Assume most variation in RE caused by RH. Used fixed composition - use average regional PM2.5 composition; PMc as function of PM2.5 (from regional ratios); vary RH using regional (hourly-specific) distribution.
- Inputs / Processing:
  - 1 year ('03) complete ESpN $PM_{2.5}$ speciation data: EC, OCM, Sulfates, Nitrates, Crustal. Sites needed 4 quarters of 11+ samples.
    - Components (in ug/m3) averaged by site X quarter; then by site (annual), then across PMreg.
    - Regional component percentages (of Remass) computed from PMregion average mass compositions
  - PMc/PMf ratio, by PMregion, from 'regular' SP $PM_{10-2.5}$ db.
    - Db represents most recent consecutive 12-, 8-, or 4-quarter period from '01-'03.
    - Regional ratios (of PMc/PMf) computed at site level then averaged by PMregion
  - 3-year average ('01-'03) hourly RH data from all NWS sites
    - Assigned NWS sites to PMregion
    - Averaged RH by ID X month X day X hr. [Average of 3 years] NOTE THAT THIS STEP WAS SUBSEQUENTLY CHANGED TO USE THE 3-YEAR (RAW) DISTRIBUTION INSTEAD OF THE DISTRIBUTION OF THE 3-YEAR AVERAGES.
    - Computed univariate distribution p1-p100, by PMregion X hr. [Dist. of all sites' 365/6 hrly avgs.]
    - For each percentile value of RH (by PMregion V hr), identified table-look-up value for FRH. RH capped at 95. ([.e., Same FRH value (7.4) assigned to all RH's $\geq$ 95.]

Inputs / Processing - Continued

- Converted VR levels (10-60) to Reconstructed Extinction (RE) levels by formula

  RE = 3912 / VR

- Using RE formula of:

  RE =
  (**3** * FRH * PM25_NITRATES) +
  (**3** * FRH * PM25_SULFATES) +
  (**4** * PM25_OCM) +
  (**10** * PM25_EC) +
  (PM25_CRUSTAL) +
  (**.6** * PMc)
  +**10** [Rayleigh l.s.]

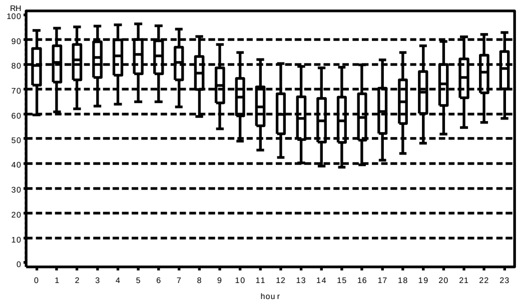
  Solved for $PM_{2.5}$ for each VR level, each hour, each percentile, by PMregion, using formula:

  PM_{25} =
  (RE -**10**) / ((**3***frh*sulfates%)+(**3***frh*nitrates%)+ (**4***ocm%)+(**10***ec%)+(**1***crustal%)+(**.6***$ratio_{pmc/pmf}$))

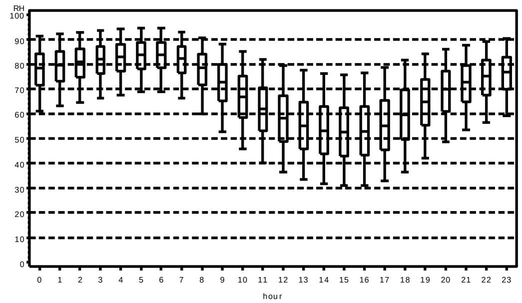
- Averaged computed hourly $PM_{2.5}$ values (by VR level, by PMregion, by percentile)… for 4-hr and 24-hr periods.

Regional Distributions of RH, by Hr

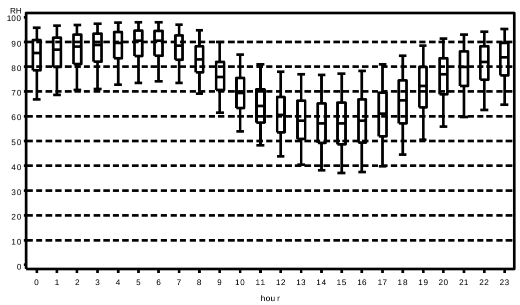
Northeast



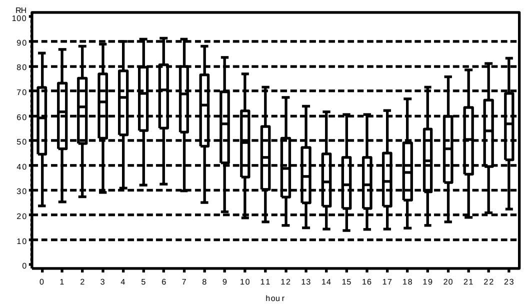
Upper Midwest



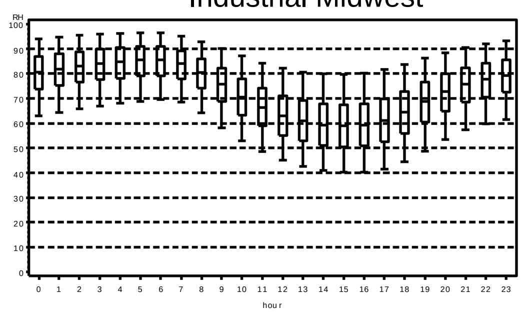
Southeast



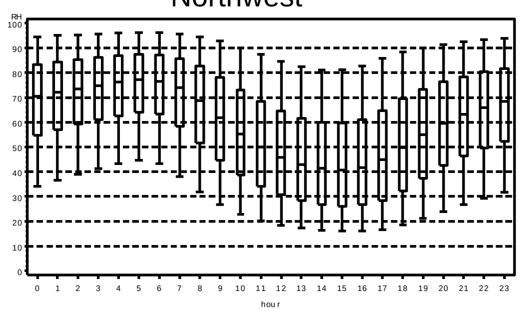
Southwest



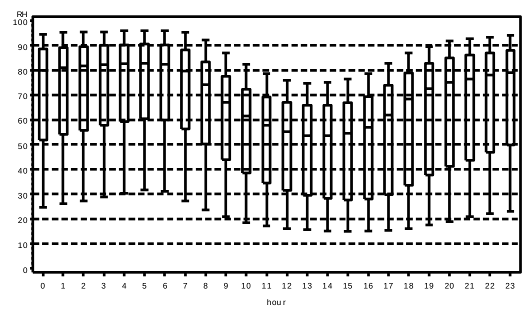
Industrial Midwest



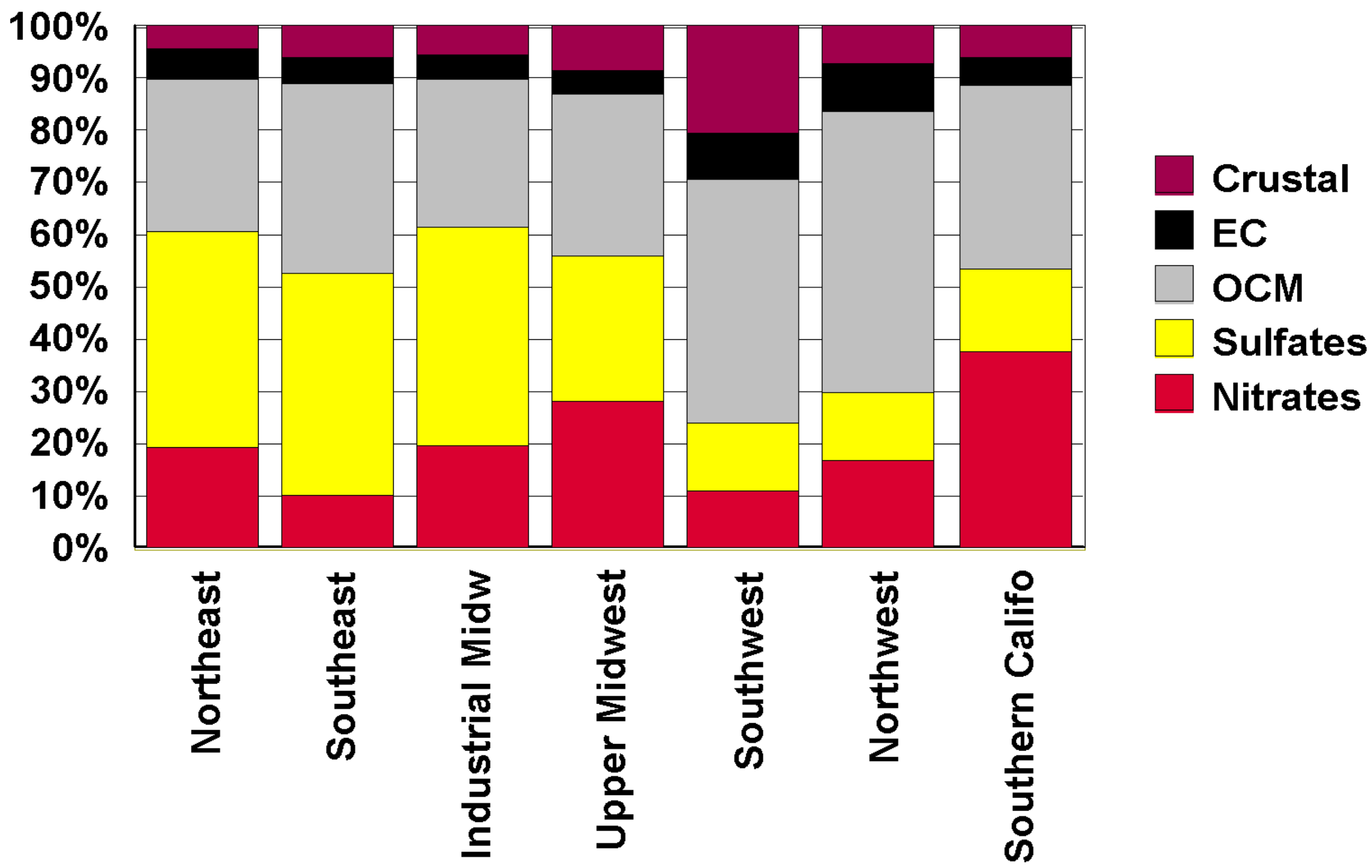
Northwest



Southern California



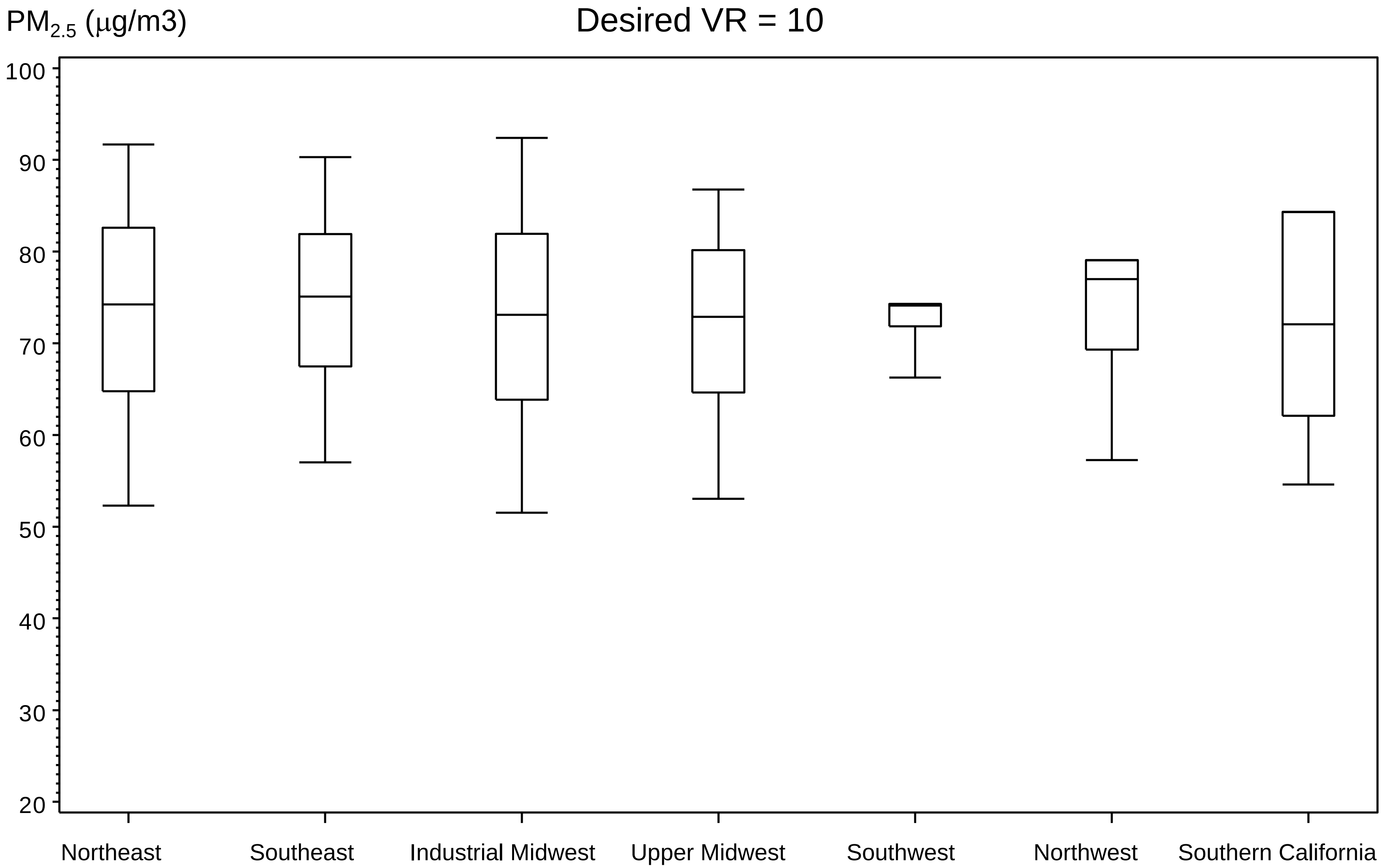
Regional Average Compositions



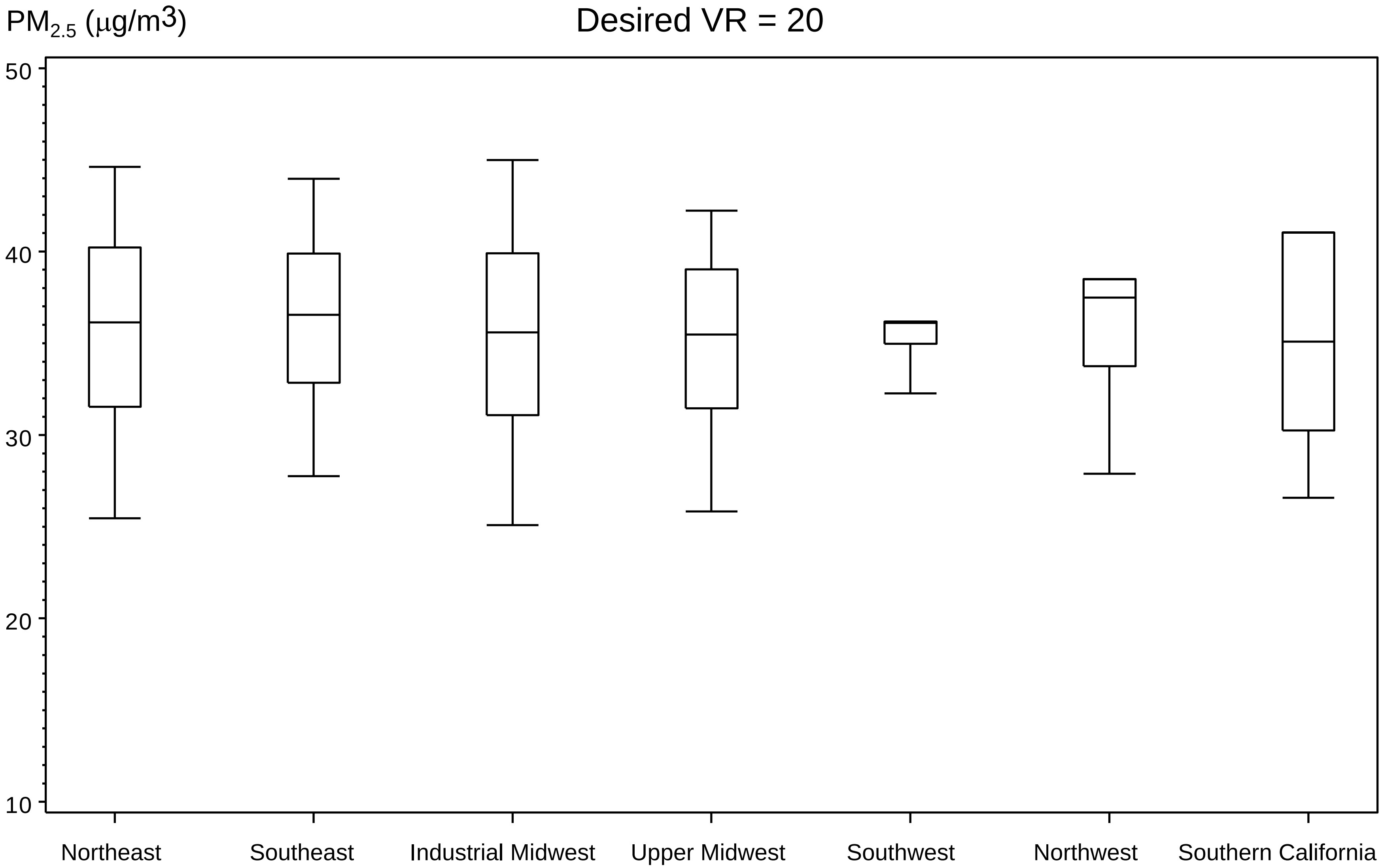
Regional PMc / PMf Ratios

Northeast	0.620
Southeast	0.794
Industrial Midwest	0.587
Upper Midwest	1.545
Southwest	2.417
Northwest	1.305
Southern Californi	1.485

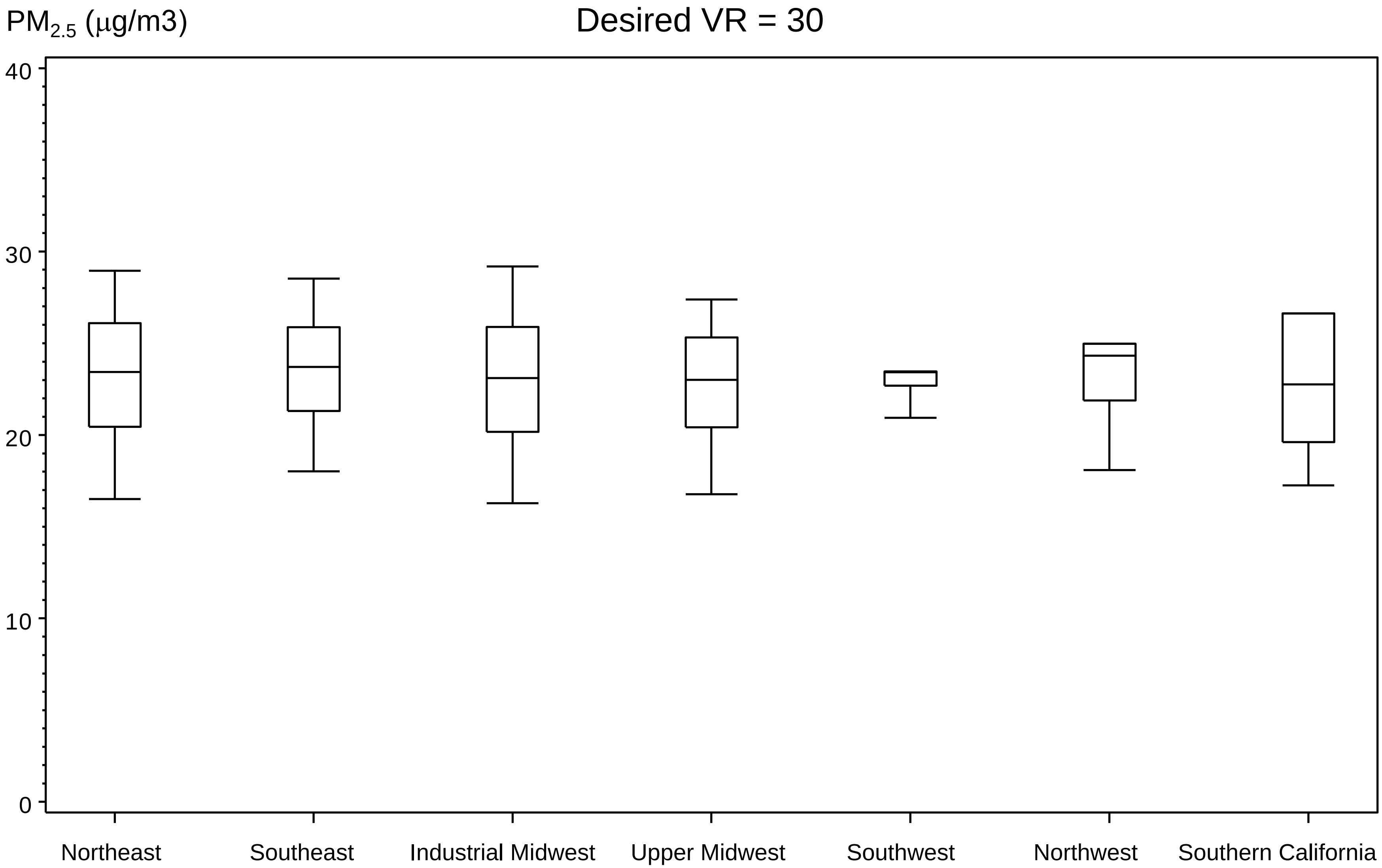
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe

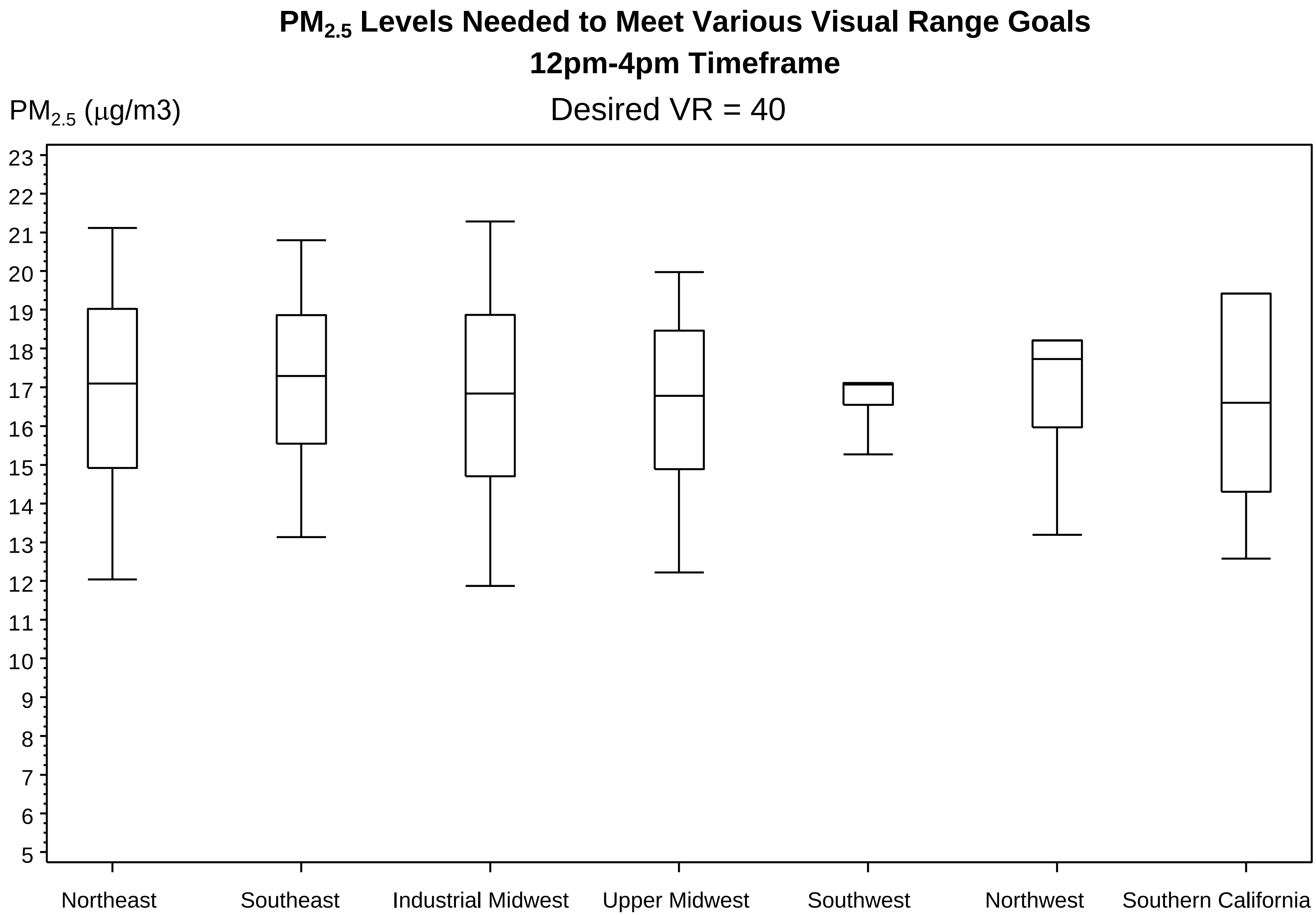


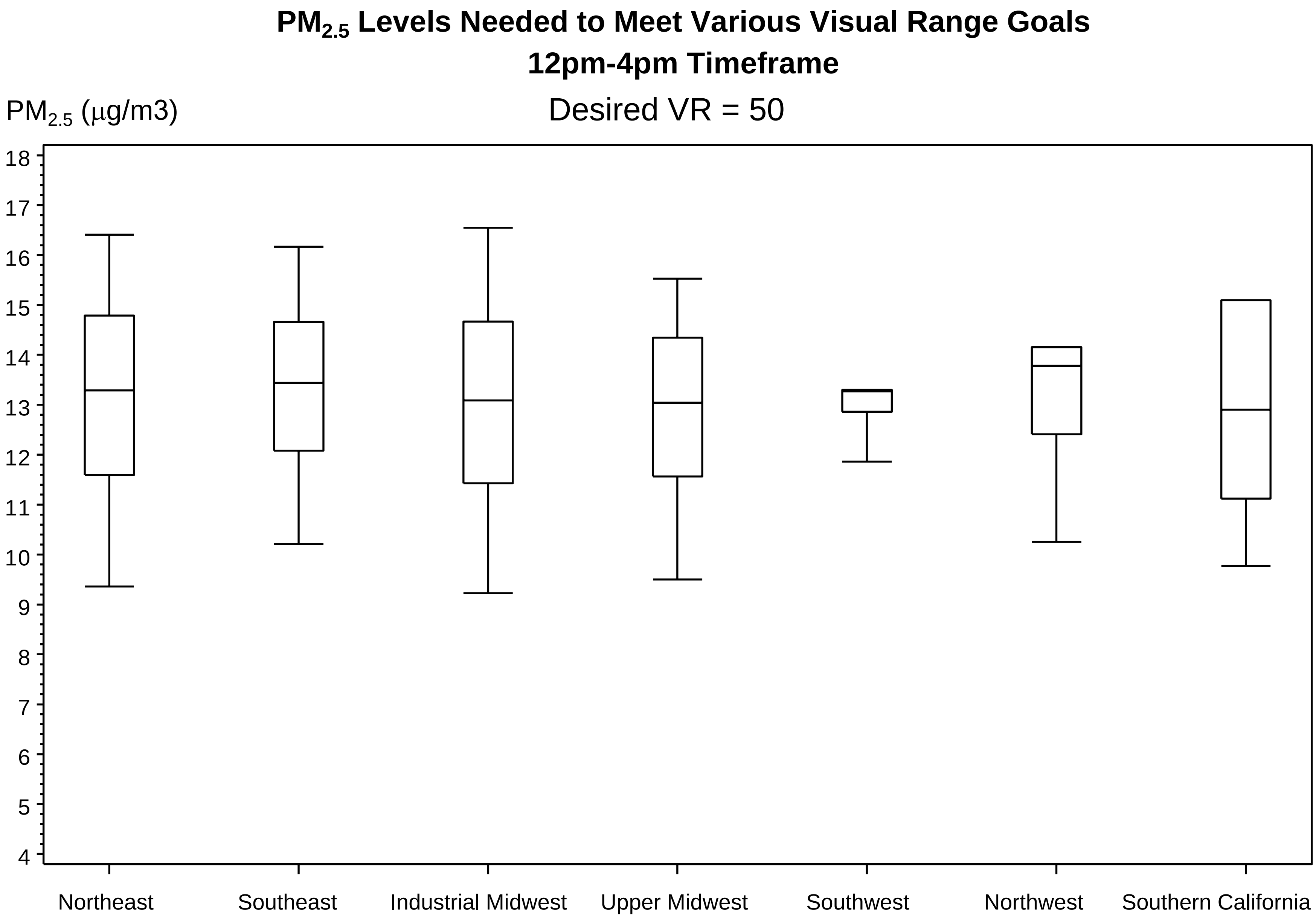
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe



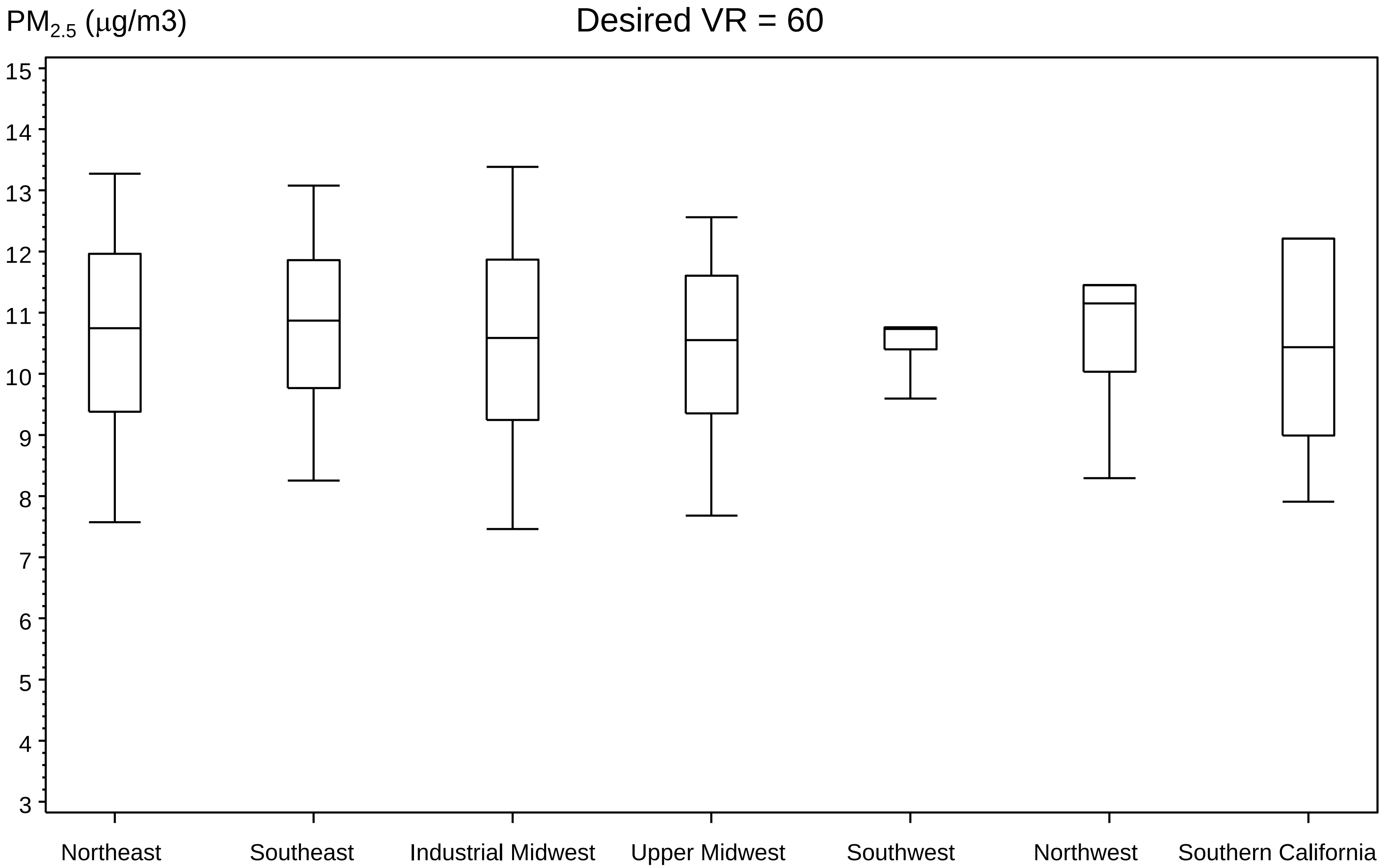
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe

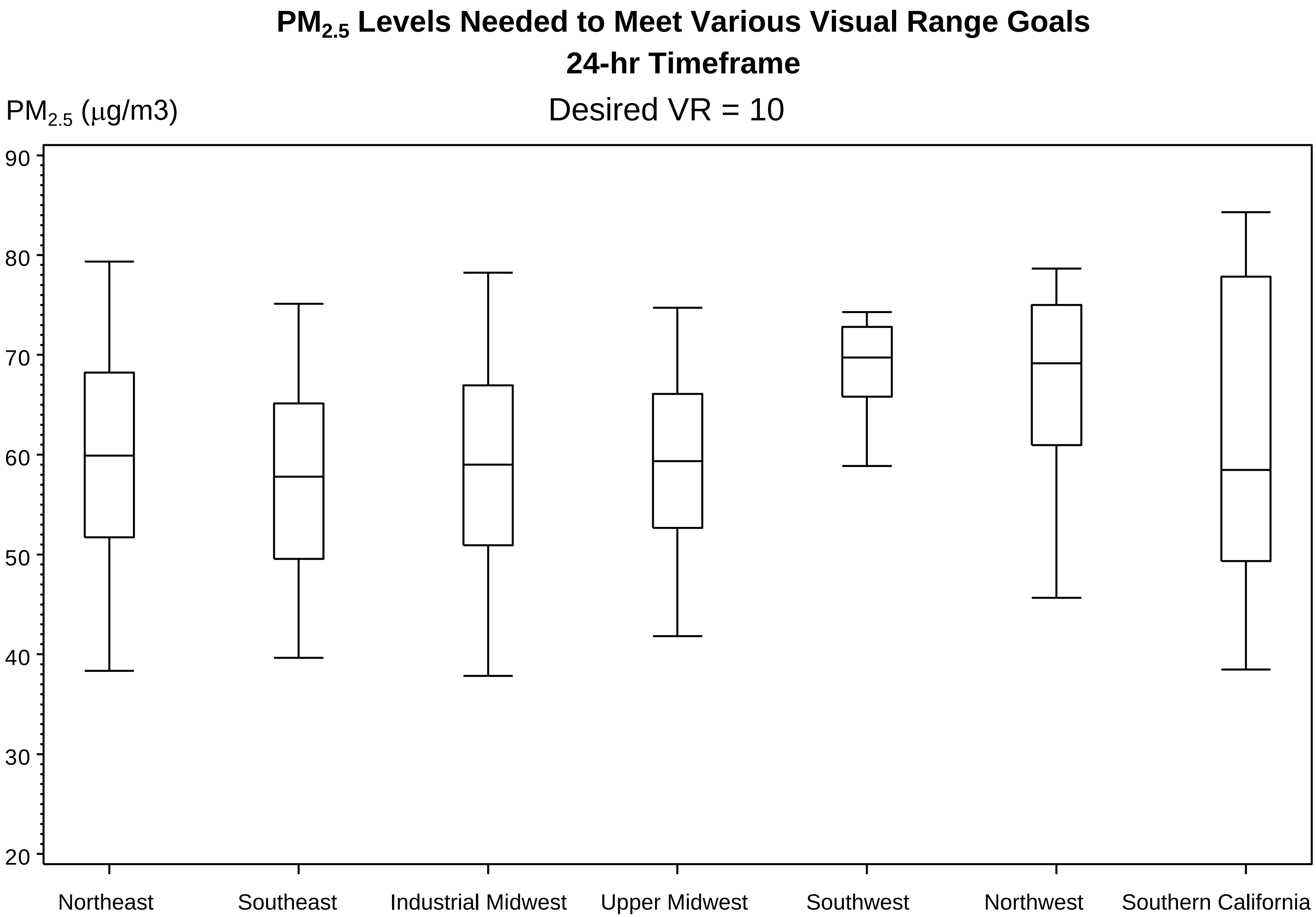



PM2.5 Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe
PM2.5 (µg/m3)
Desired VR = 40
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
Northeast
Southeast
Industrial Midwest
Upper Midwest
Southwest
Northwest
Southern California

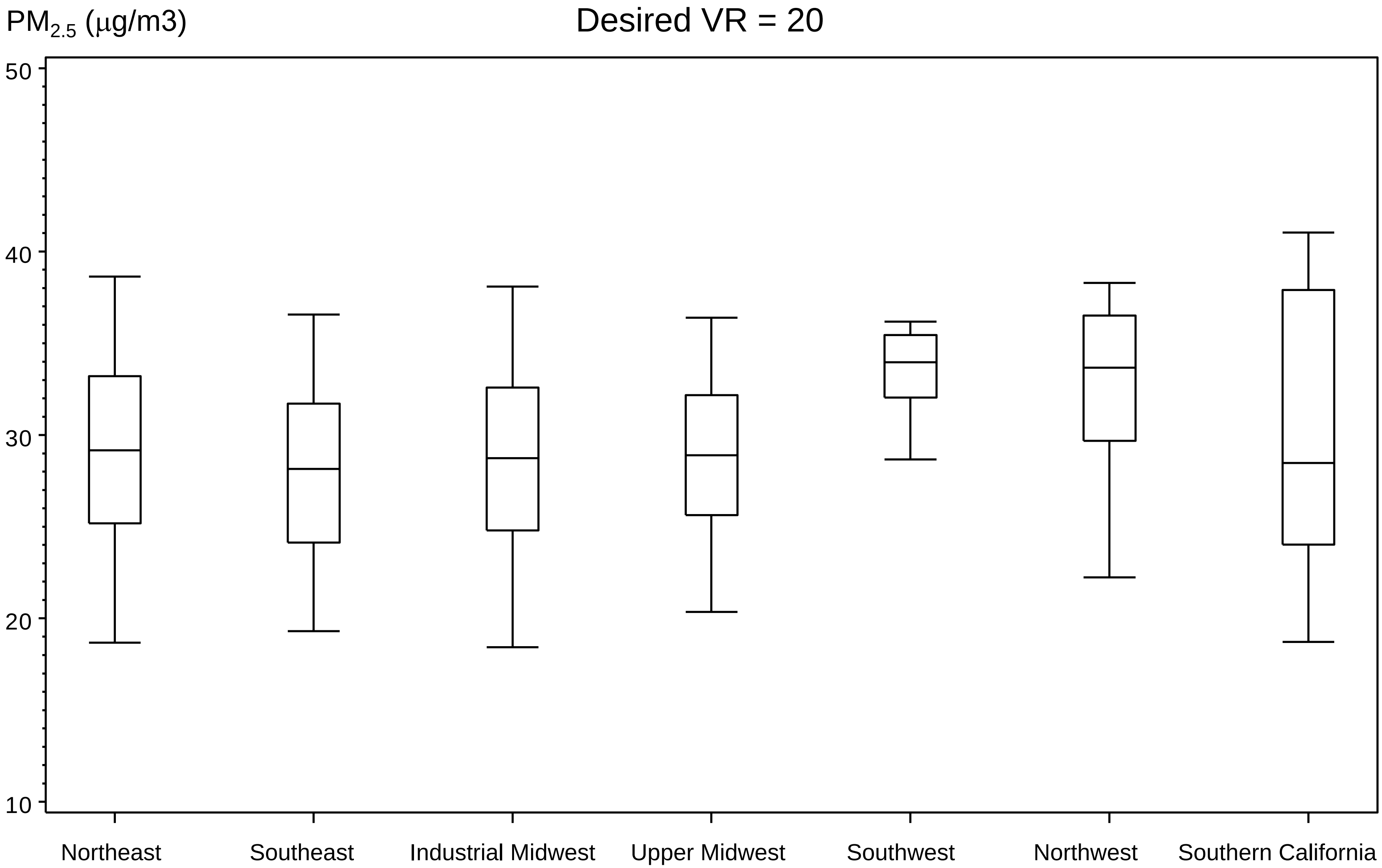

PM2.5 Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe
PM2.5 (μg/m3)
Desired VR = 50
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
Northeast
Southeast
Industrial Midwest
Upper Midwest
Southwest
Northwest
Southern California

$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe

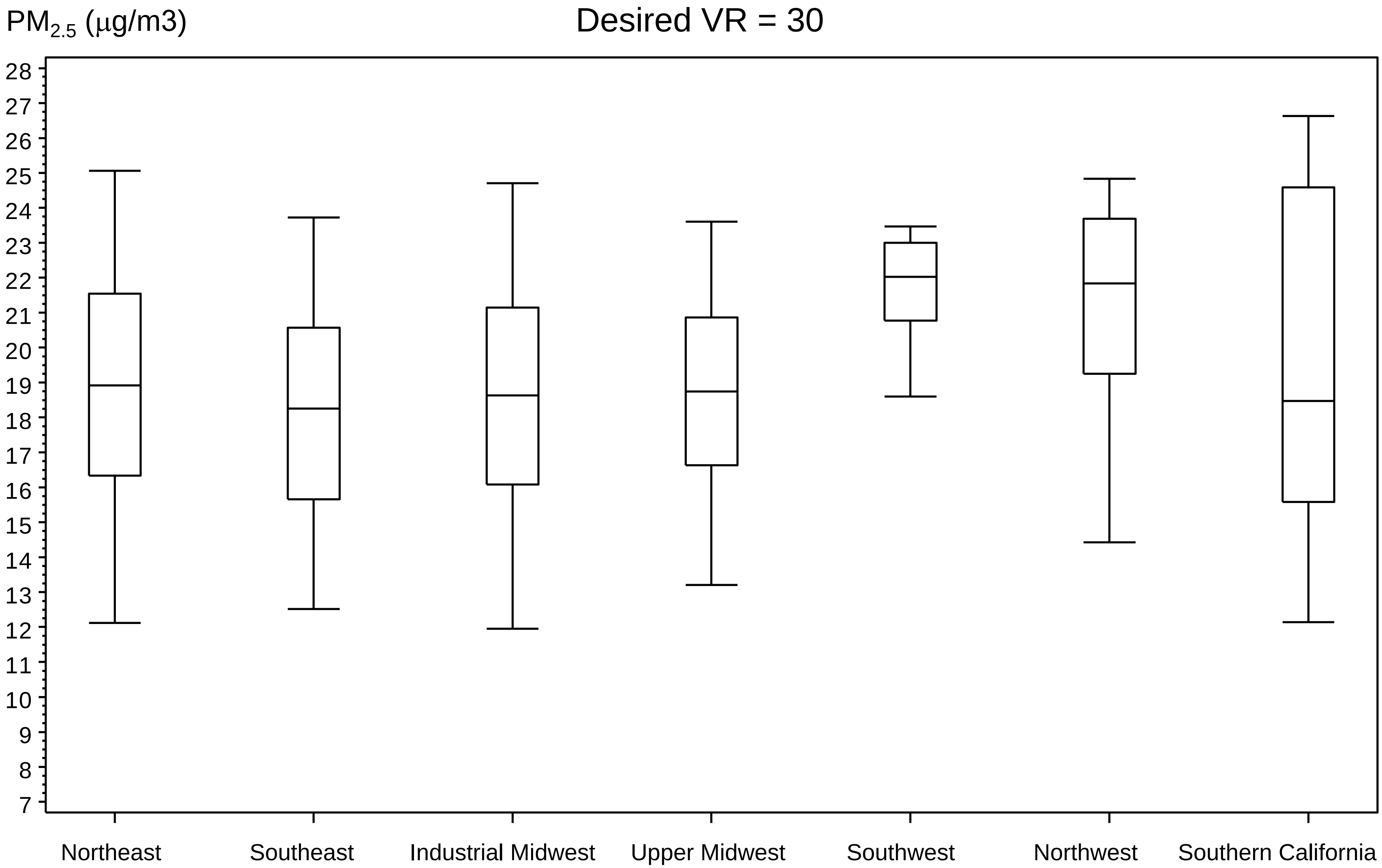



PM2.5 Levels Needed to Meet Various Visual Range Goals
24-hr Timeframe
PM2.5 (µg/m3)
Desired VR = 10
90
80
70
60
50
40
30
20
Northeast
Southeast
Industrial Midwest
Upper Midwest
Southwest
Northwest
Southern California

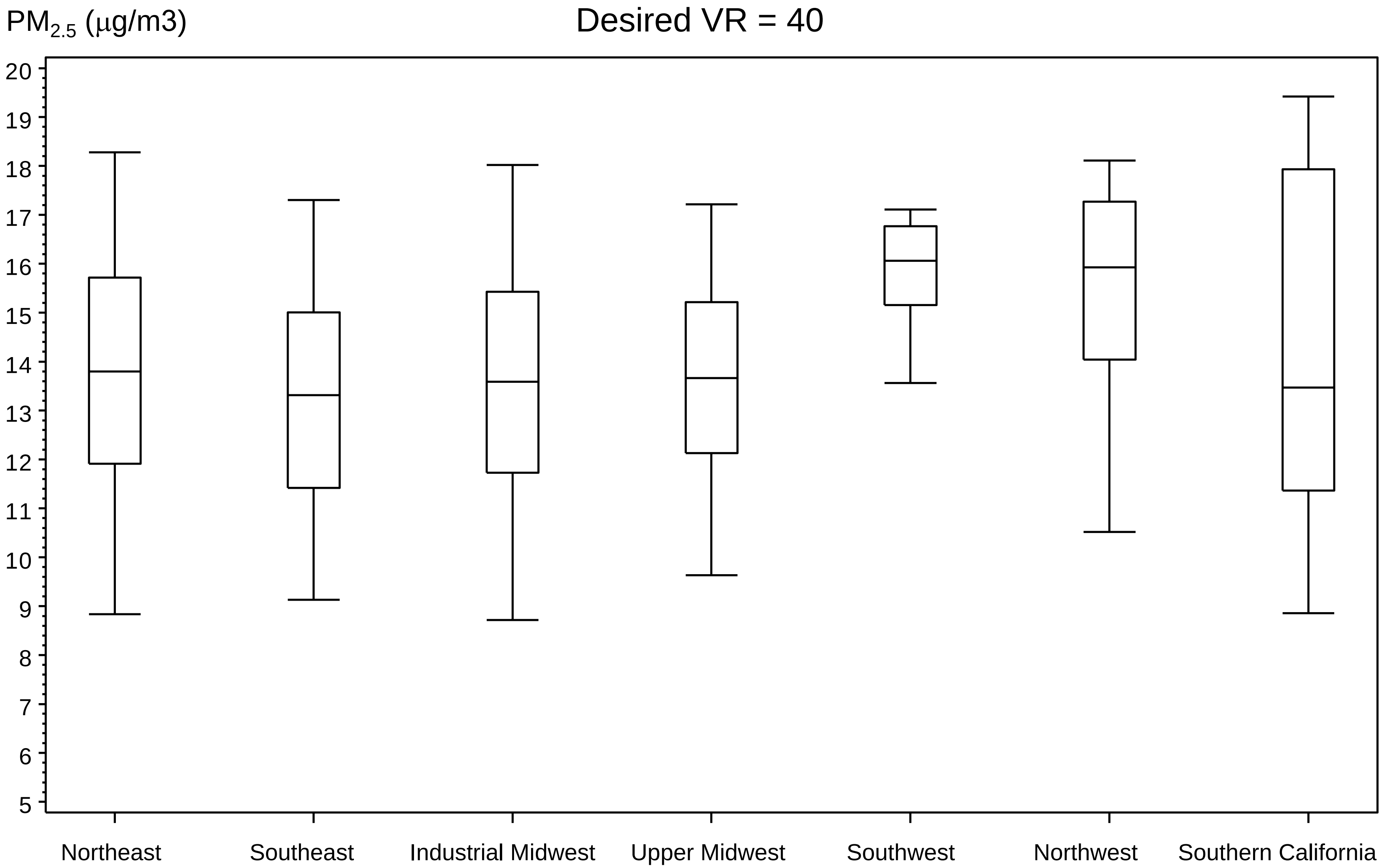
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-hr Timeframe



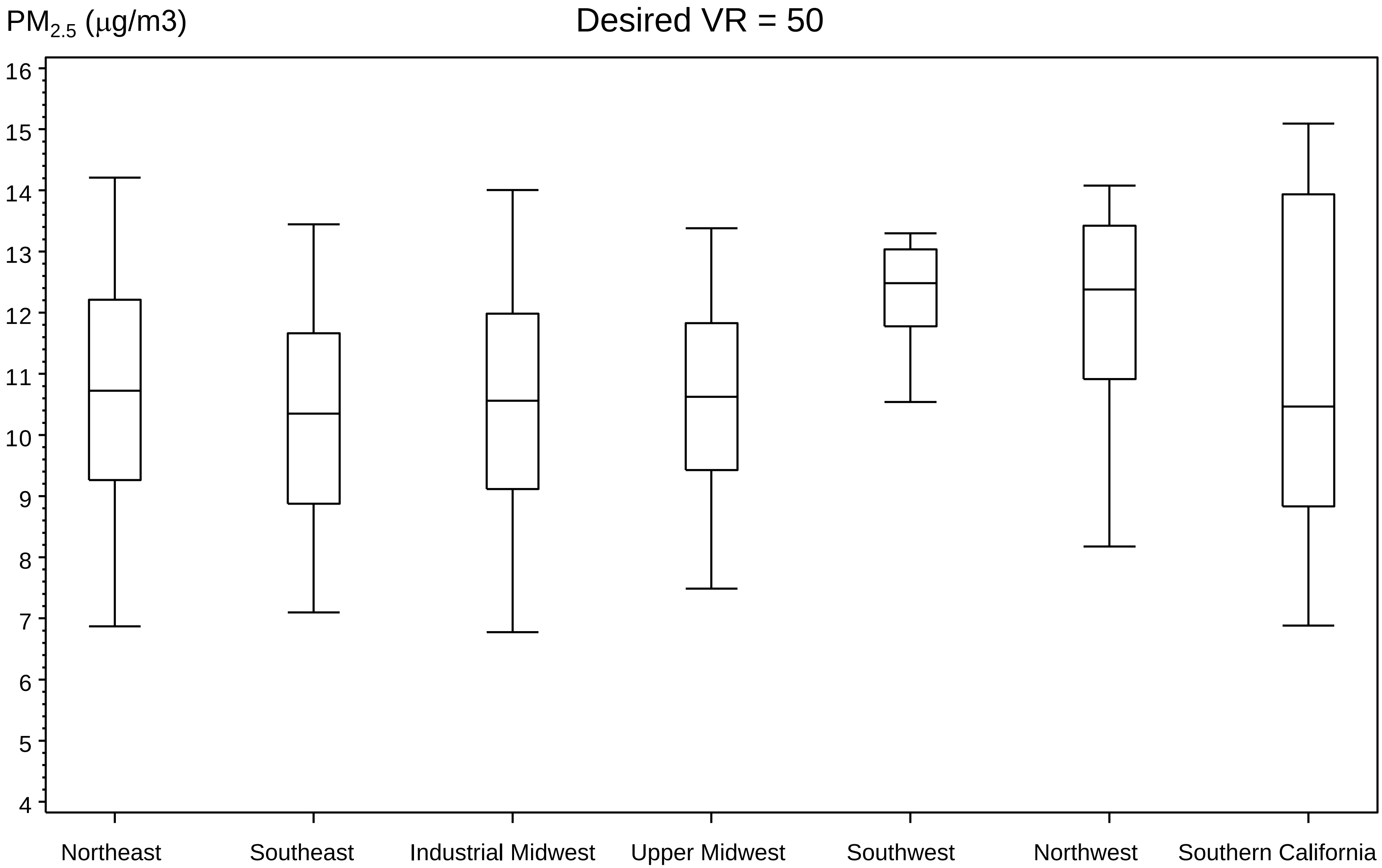
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-hr Timeframe



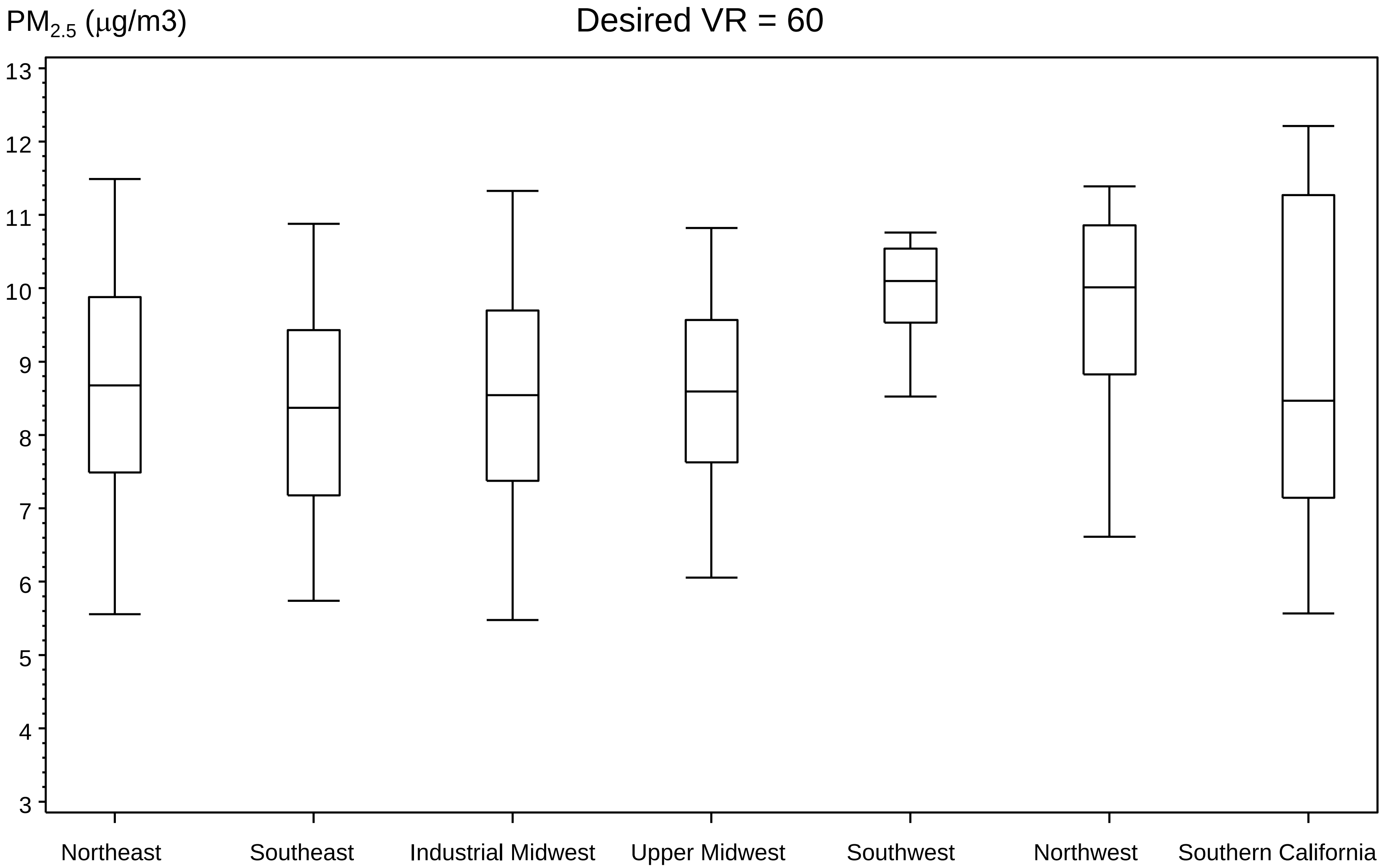
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-hr Timeframe



$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-hr Timeframe



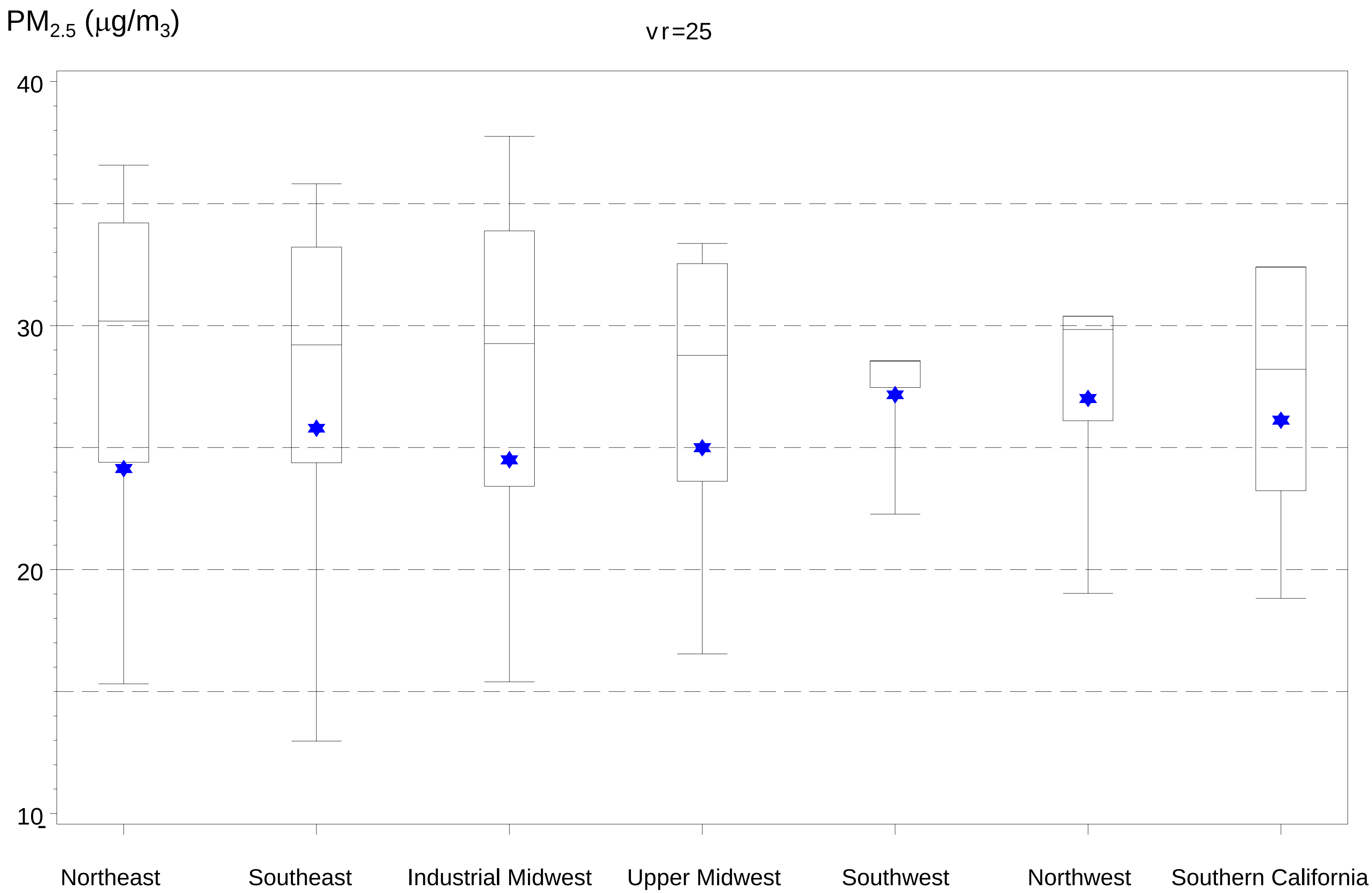
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-hr Timeframe



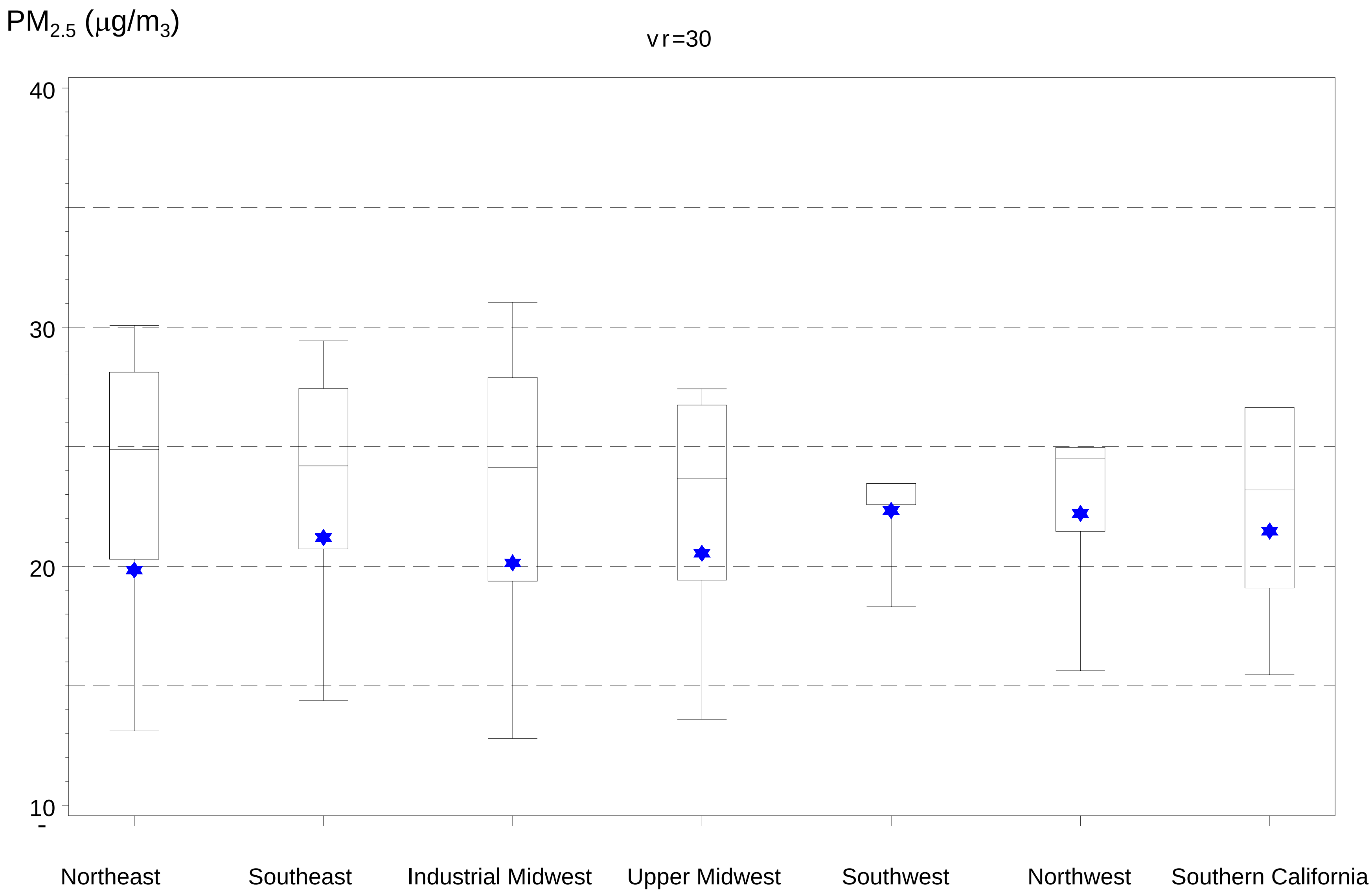
Methodology was tweaked to:

1. Use 3-year regional (raw) distribution of hourly NWS RH values instead of distribution of 3-year averages of hourly NWS RH.
2. Solve for visual range targets of 25, 30, and 35 km.
3. Add mean (utilizing regional, hourly FRH means) to graphs. Shown as an asterisk.
4. Scales were made common.

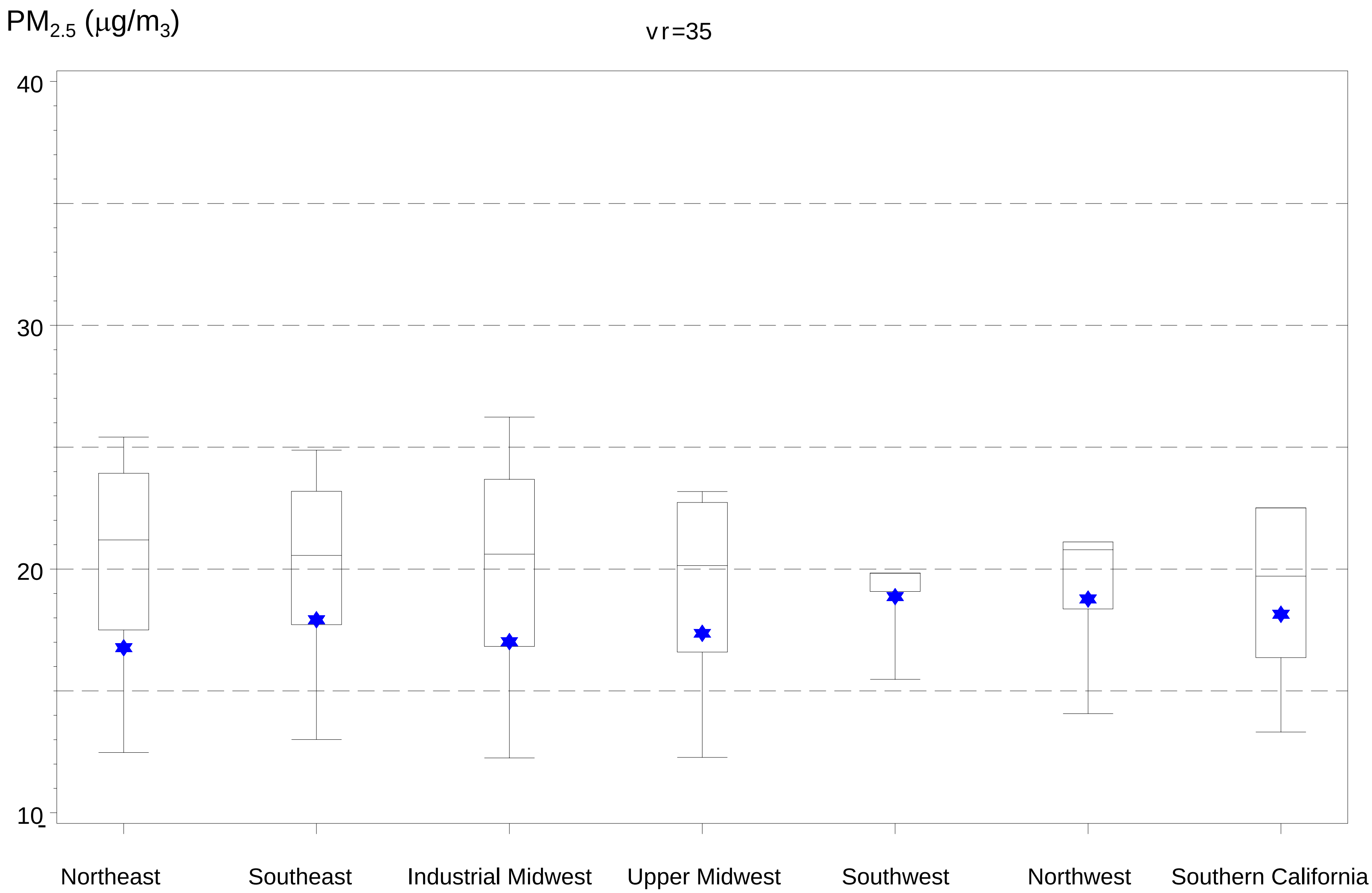
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe



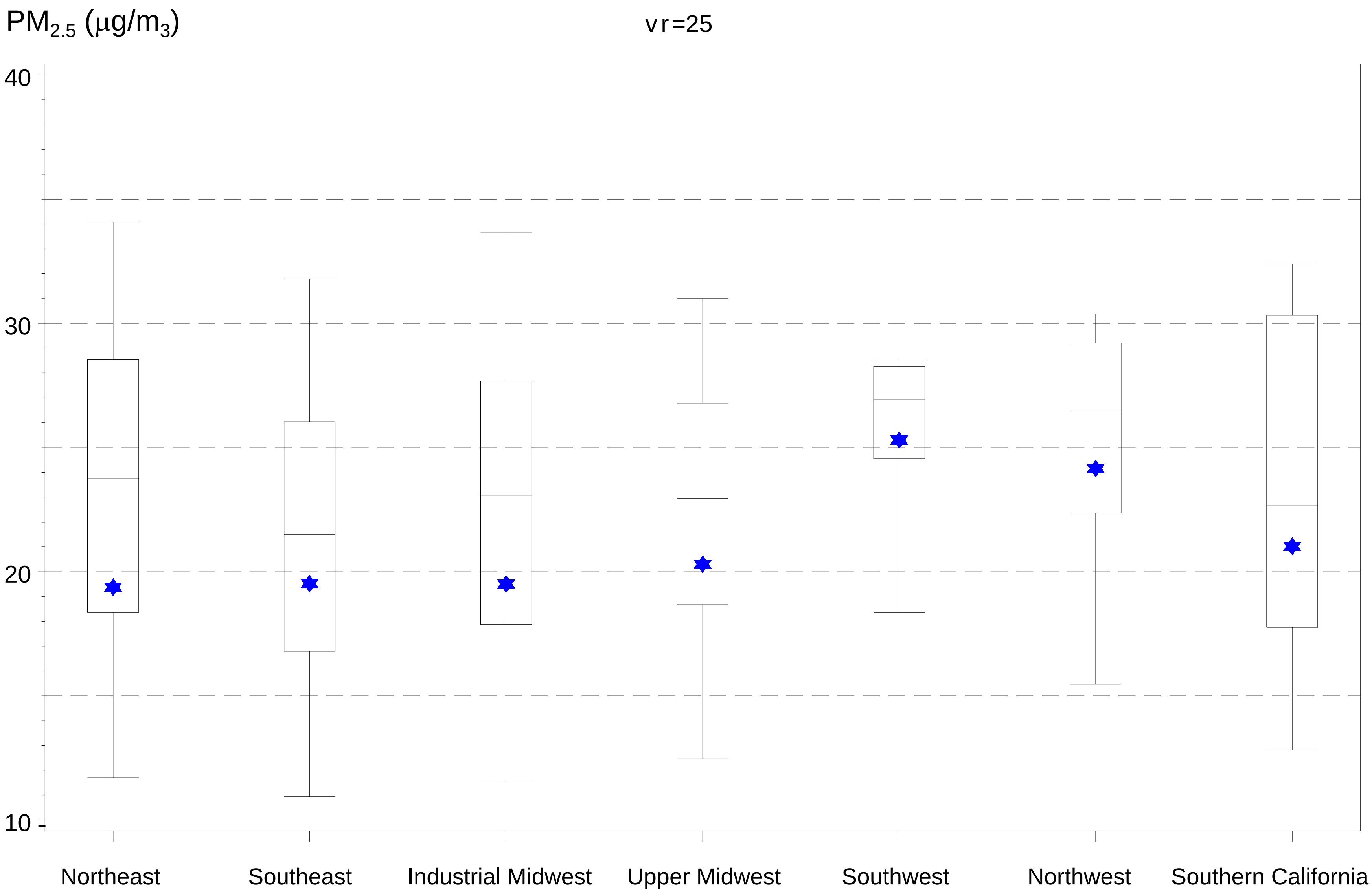
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe



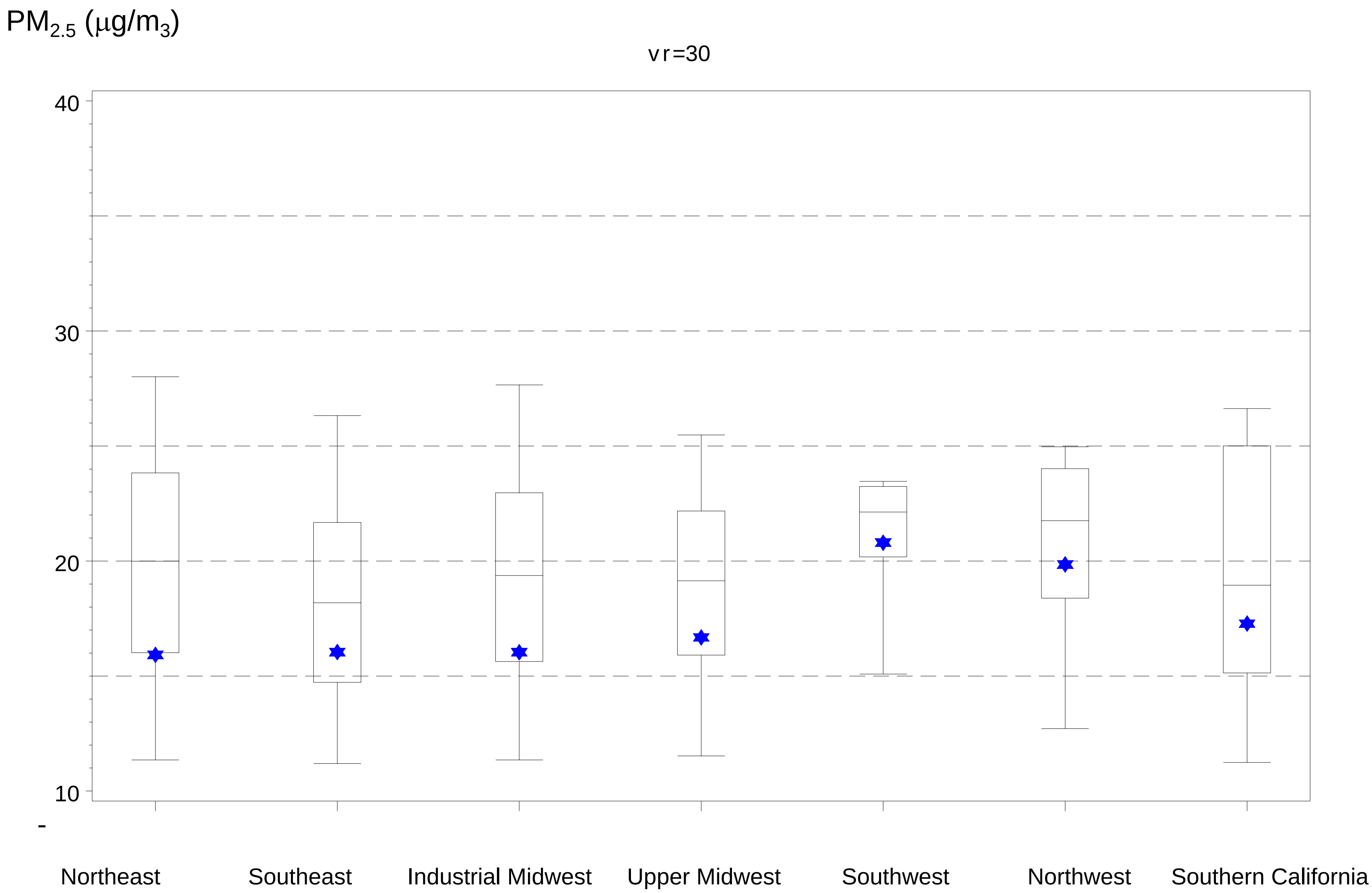
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
12pm-4pm Timeframe



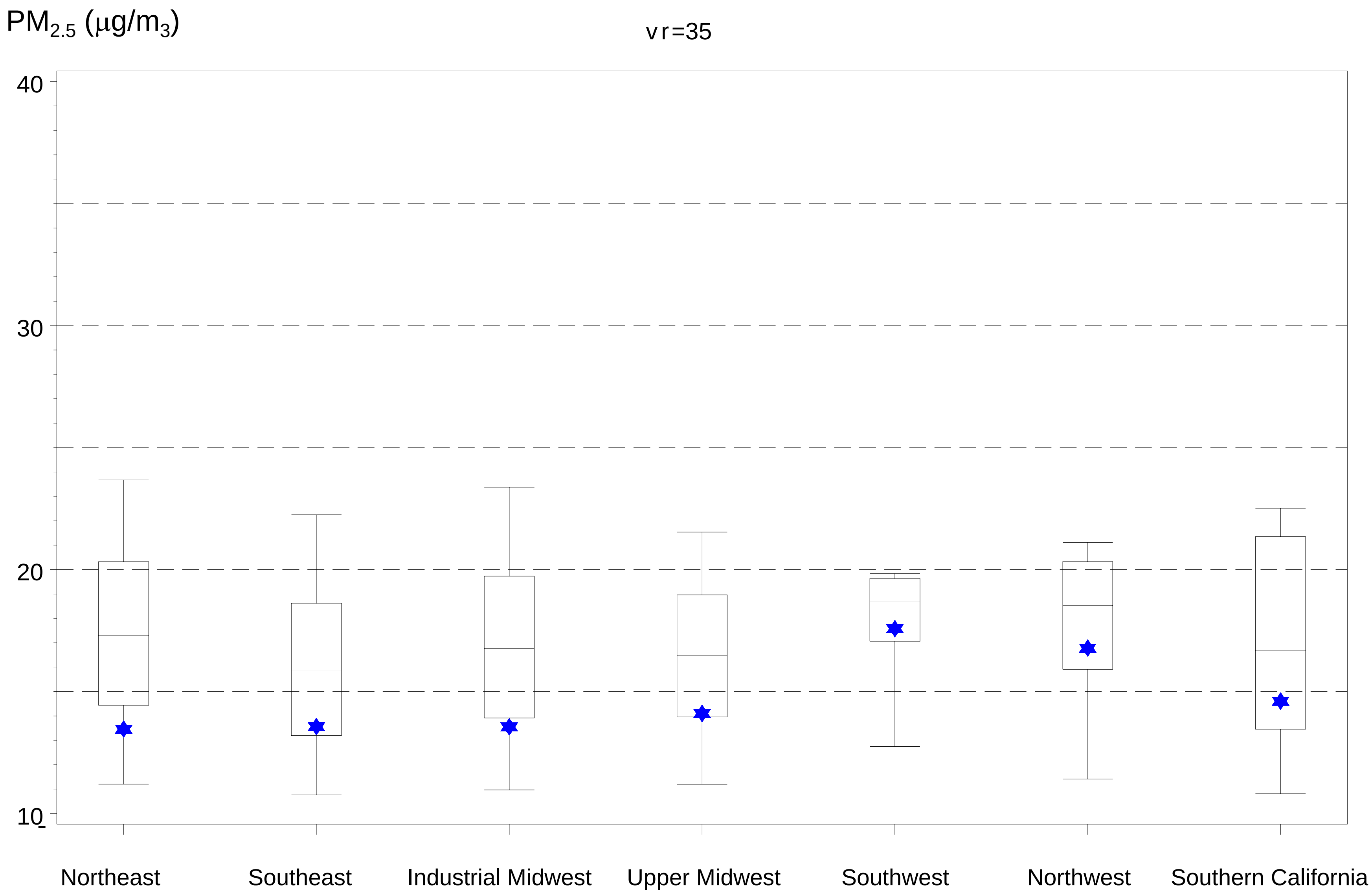
$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-Hour Timeframe



$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-Hour Timeframe



$PM_{2.5}$ Levels Needed to Meet Various Visual Range Goals
24-Hour Timeframe



Worst 20% VR Days – 2003 Visibility Database
Estimate of Visual Ranges

24-hour (min 18 hours required)

EW	PMREGDESC	#Sites	VR		EW	PMREGDESC	#Sites	VR
east	Industrial Midwest	37	17.9		west	Northwest	19	29.0
east	Northeast	40	19.2		west	Southern California	5	13.4
east	Southeast	51	23.3		west	Southwest	6	34.6
east	average	128	**20.1**		west	Upper Midwest	3	30.5
					west	average	33	**26.9**

4-hour, 12-4pm (min 3 hours required)

EW	PMREGDESC	#Sites	VR		EW	PMREGDESC	#Sites	VR
east	Industrial Midwest	37	23.2		west	Northwest	19	39.0
east	Northeast	40	21.8		west	Southern California	5	18.6
east	Southeast	51	32.0		west	Southwest	6	32.6
east	average	128	**25.7**		west	Upper Midwest	3	33.5
					west	average	33	**30.9**

$PM_{2.5}$ Secondary NAAQS – Form of Std

Comparability to Regional Haze rule:

- Regional Haze rule uses average of 20% worst days
- SP Jan05 draft suggests 90th percentile for secondary NAAQS (using $PM_{2.5}$ as surrogate for visibility)
- CASAC comment notes that ‘average of 20% worst’ is closer to 92nd than 90th percentile

Use $PM_{2.5}$ as surrogate for worst days

- Used $PM_{2.5}$ continuous database
- Calculated 24-hr and 4-hr avg's; required 18, 3 obs min per day respectively
- Required 275+ (75%) min days per year
- Identified 80th percentile $PM_{2.5}$ (24- and 4-hr avg) for each site-year
- Computed mean of 20% worst days (by site-year) [days that exceeded 80^{th}]
- Identified where in the distribution (of site-year-days) the mean_20%worst fell
    - Call estimated-site-year-moment_num
- Output shown is distribution of estimated-site-year-moment_num's

24-hour

n:	420
mean:	91.945238095
max:	96
p95:	94
p75:	93
med:	92
p25:	91
p05:	90
min:	90

4-hour (12-4pm)

n:	417
mean:	92.122302158
max:	97
p95:	95
p75:	93
med:	92
p25:	91
p05:	90
min:	90

Use Estimated RE to ascertain worst days

- Used 2003 visibility database
- Calculated 24-hr and 4-hr avg's; required 18, 3 obs min per day respectively
- Required 100+ min days per year
- Identified 80th percentile RE (24- and 4-hr avg) for each site-year
- Computed mean of 20% worst days (by site-year) [days that exceeded 80th]
- Identified where in the distribution (of site-year-days) the mean_20%worst fell
    - Call estimated-site-year-moment_num
- Output shown is distribution of estimated-site-year-moment_num's

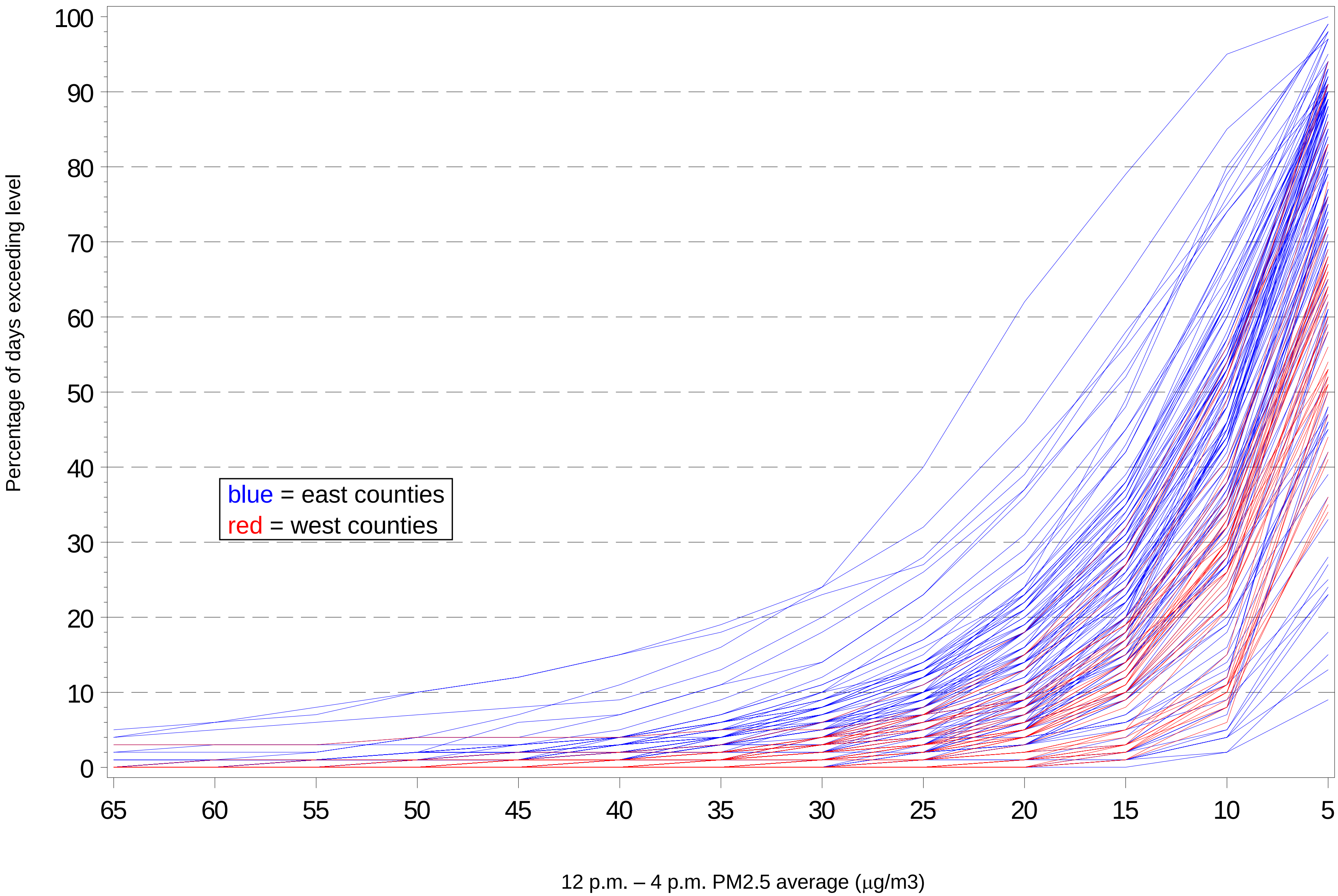
24-hour

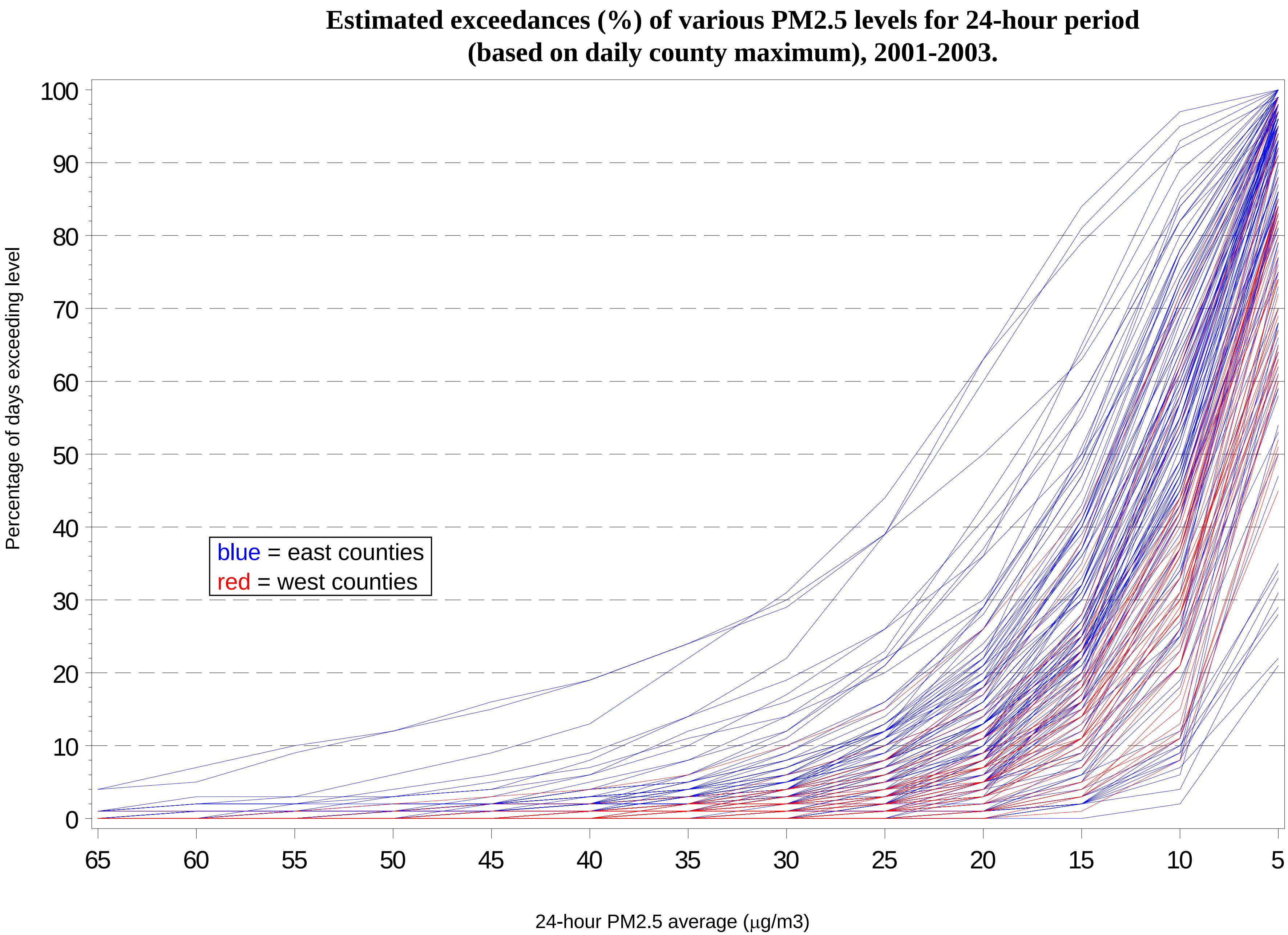
n:	47
mean:	90.765957447
max:	93
p95:	92
p75:	91
med:	91
p25:	90
p05:	89
min:	88

4-hour (12-4pm)

n:	42
mean:	91.261904762
max:	94
p95:	93
p75:	92
med:	91.5
p25:	90
p05:	89
min:	88

Estimated exceedances (%) of various PM2.5 levels for 12 p.m. - 4p.m. (based on daily county maximum), 2001-2003.





Estimated exceedances (%) of various PM2.5 levels for 24-hour period
(based on daily county maximum), 2001-2003.
Percentage of days exceeding level
100
90
80
70
60
50
40
30
20
10
0
blue = east counties
red = west counties
65
60
55
50
45
40
35
30
25
20
15
10
5
24-hour PM2.5 average (µg/m3)

Table 7-1. Predicted percent of counties with monitors (and percent of population in counties with monitors) not likely to meet alternative 4-hour (12 p.m. - 4 p.m.) $PM_{2.5}$ secondary standards

Alternative Standards and Levels		Percent of counties, total and by region, (and total percent population) not likely to meet stated standard and level*								
Form (percentile)	Level (μg/m3)	Total Counties (population)	Northeast	Southeast	Industrial Midwest	Upper Midwest	Southwest	Northwest	Southern CA	Outside Regions**
Number of counties with monitors (Population, in thousands)		**168 (78,419)**	**33**	**45**	**30**	**16**	**14**	**25**	**3**	**2**
90th	20	44 (63)	76	31	83	25	7	4	100	50
90th	25	18 (33)	39	7	33	0	7	4	67	0
90th	30	4 (12)	9	0	3	0	7	0	67	0
92nd	20	51 (67)	76	36	90	38	29	8	100	100
92nd	25	27 (46)	52	16	53	0	7	4	100	0
92nd	30	8 (17)	15	0	17	0	7	0	67	0
95th	20	70 (83)	88	73	97	50	50	24	100	100
95th	25	47 (67)	79	31	87	19	29	4	100	100
95th	30	24 (43)	52	7	47	0	14	4	100	0
98th	20	85 (96)	100	100	100	63	57	48	100	100
98th	25	70 (81)	94	62	100	50	57	28	100	100
98th	30	56 (73)	85	38	90	19	57	24	100	100

* Based on 2001-2003 data for sites with at least 1 year of complete data. Completeness criteria (per year) = Minimum of 3 hours per day (in 4-hour 12-4pm window), 275+ days per year. As such, these estimates are not based on the same air quality data that would be used to determine whether an area would attain a given standard or set of standards. These estimates can only approximate the number of counties that are likely not to attain the given standards and should be interpreted with caution.