Results of the 2005 NATA Model-to-Monitor Comparison*

Regi Oommen
Data Analysis Workshop
Dallas, TX
April 6, 2011

*Final Report posted at:
http://www.epa.gov/ttn/atw/nata2005/05pdf/nata2005_model2monitor.pdf

Acknowledgements

- **EPA**
  - Barbara Driscoll
  - Ted Palma
  - Anne Pope

- **ERG Staff**
  - Stacie Enoch
  - Robin Weyl
  - Heather Perez
  - Jaime Hauser

Overview

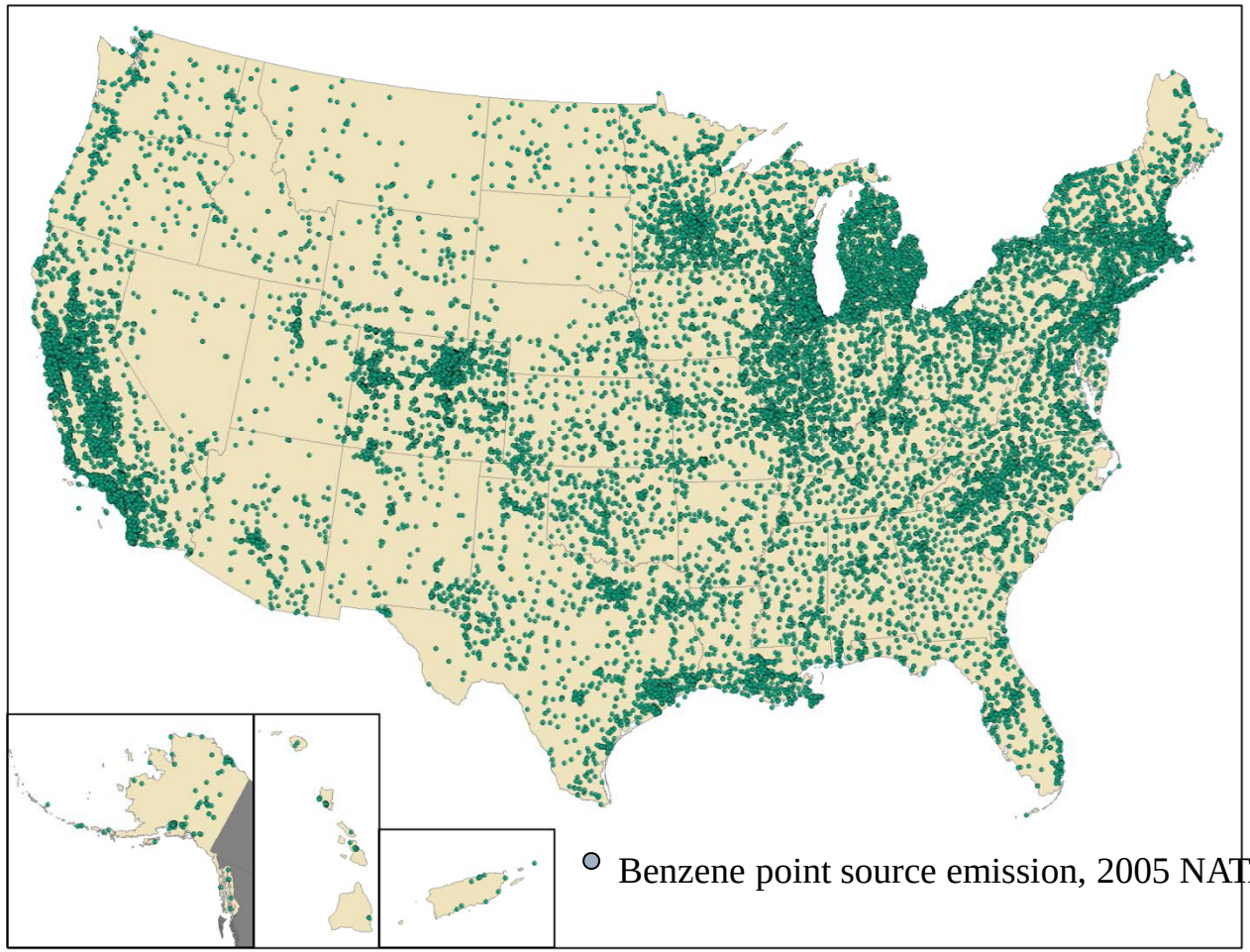
- **Background on NATA**
- **Data Sources**
- **Methodology**
- **Preliminary Results**
- **Conclusions**

Importance of Air Toxics Modeling

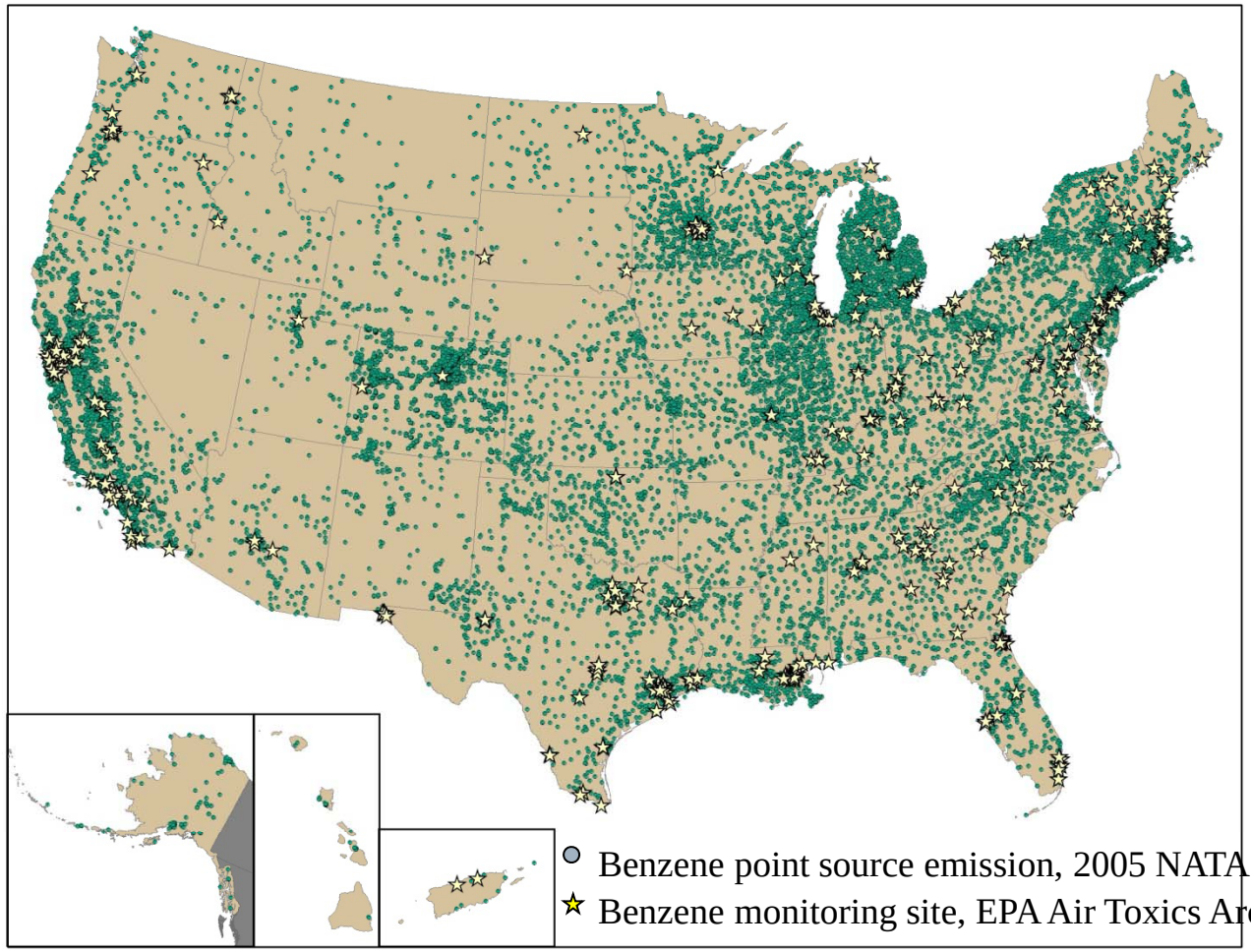
- Some Uses
  - Helps identify “hot spots”
  - Trends/Accountability
  - Validate air toxics monitoring

- EPA uses the National-Scale Air Toxics Assessment (NATA)

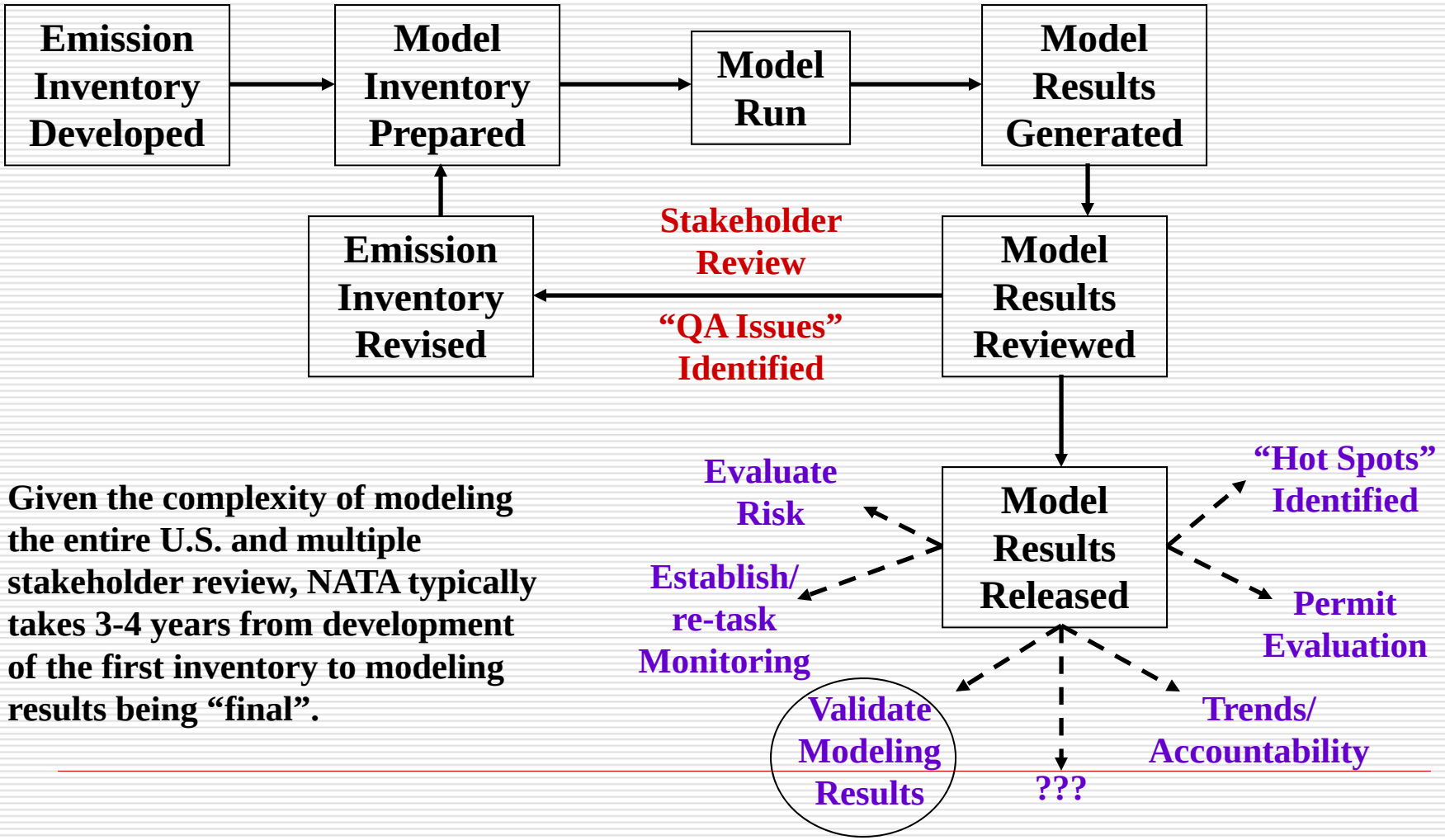
Importance of Air Toxics Modeling – "Bridging the Gap"



Importance of Air Toxics Modeling – "Bridging the Gap"



"Typical" Steps in Air Toxics Modeling



Given the complexity of modeling the entire U.S. and multiple stakeholder review, NATA typically takes 3-4 years from development of the first inventory to modeling results being "final".

Assessing Model Results – Quality Assurance

- School Air Toxics Monitoring Initiative Example
  - The Risk Screening Environmental Indicators (RSEI) Model was used by USA Today to overlay risk with school locations.
  - Model used 2007 Toxic Release Inventory (TRI).

Assessing Model Results – Quality Assurance

- School Air Toxics Monitoring Initiative Example

  - A number of the “sources” had reported incorrect emissions
  - A number of the “sources” had significant emission reductions not reflected in TRI.
  - RSEI Model not accurately modeling certain pollutants (e.g., diisocyanates).

Assessing Model Results – Quality Assurance

- School Air Toxics Monitoring Initiative Example
  - EPA also used NATA 2002 model results in assessing monitoring locations.
  - A number of “potential areas” were not chosen when ground-truthing of emission sources was performed.

Assessing Model Results – Quality Assurance

- NATA 2002 Example
  - In June 2009, EPA released the results of the 2002 NATA
  - Results indicated that census tracts in a small city in a western state were the highest in the country for cancer risk.

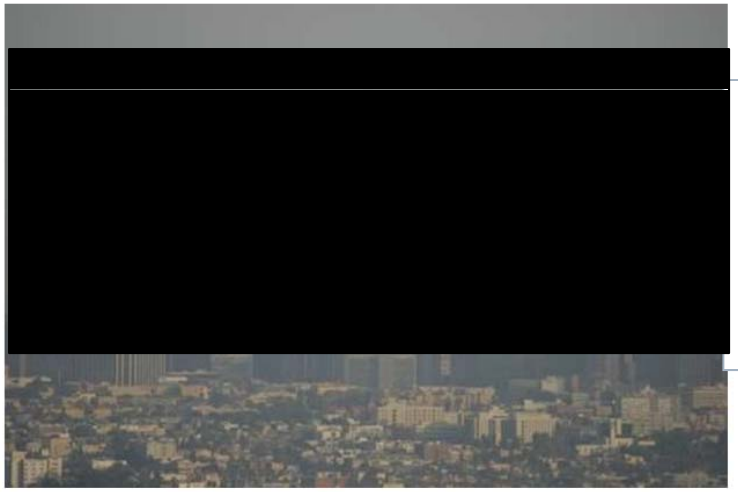
Assessing Model Results – Quality Assurance

- "Infamous" City:
  - Described as a progressive city with strong retail base
  - Median household income = $91,476 (2007)
  - City sends parade floats to the Tournament of Roses parade
  - Tourism/retail major industry

Assessing Model Results – Quality Assurance

[redacted] air pollution may endanger babies, people in general

June 25, 2009 | 8:18 pm (0) (7) Comments (6)



It looks like [redacted]air could be killing us in more ways than one.

Two studies released Wednesday have linked toxic air pollution in Southern California to cancer and complications with birth.

Exposure to local traffic-generated pollution increased the risk of major complications and preterm birth, concluded a report published online in Environmental Health Perspectives. Local scientists studied the relationship of traffic pollution, preterm birth and a complication called preeclampsia that can lead to maternal and perinatal morbidity.

Meanwhile, an Environmental Protection Agency study found that [redacted] has some of the highest levels of cancer-related toxic air pollutants in the country. For residents of [redacted], located at the heart of the [redacted], the EPA estimated the cancer risk due to air toxics at 1,200 in 1 million, the highest in the country and more than 33 times the national average. The statistic represents the expected number of additional deaths per million people, based on a lifetime exposure to the chemicals.

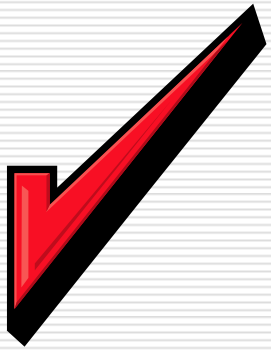
For much of the rest of the [redacted], cancer risks ranged from 50 per million to 75 per million, according to the EPA.

Assessing Model Results – Quality Assurance

- NATA 2002 Example
  - A review of the emissions data revealed that a small industrial facility contributing to the elevated cancer risk was:
    - Shutdown prior to 2002
    - Not in the town of question
- Incorrect (old) data were in State's 2002 NEI submittal

Assessing Model Results – Quality Assurance

- NATA 2002 Example
  - Emissions were removed from the 2002 emission inventory
  - Model was rerun for census tracts in question
  - Results reposted in August 2009
  - Ensured that data was not carried-forward into 2005 NATA!



Assessing Model Results – Quality Assurance

EPA erred in figuring [redacted] cancer risk

Agency used 1989 data for a plant that's really in [redacted]

July 11, 2009 | [redacted]

Recommend 0 Tweet Submit 0 Digg

The city of [redacted] is used to raking in awards and distinctions. Named an All-America City in [redacted] and a Tree City USA for [redacted] bills itself as an [redacted] A fountain adjacent to City Hall belches up clean, chlorinated water to the delight of area children, who romp in their bathing suits in the sun reflected off the city's award-winning, titanium-clad library.

But [redacted] was tagged with a more noxious distinction two weeks ago, when the federal Environmental Protection Agency declared it the city with the highest cancer risk in the country -- more than 33 times the national average. News crews descended, e-mail from alarmed residents flooded in, and city officials were forced to answer hard questions at a town hall meeting.

When he heard the news, Mayor [redacted] said, his first thought was: "It had to be a mistake."

Two weeks later, it looks like the mayor was right.

The EPA said its data on the [redacted] plant, which it said was spewing about 1,250 pounds of the cancer-causing chemical hydrazine, was from 1989.

Twenty years later, [redacted] is emitting less than two pounds of hydrazine per year, according to air monitoring data released Thursday by the [redacted]

That pushed the lifetime cancer risk from hydrazine exposure to about 1 in 2 million. The EPA had calculated the area's cancer risk at about 1,200 in 1 million, largely due to hydrazine exposure.

The EPA defended the study, saying that it got its data from a 2002 national emissions inventory that used data from the state.

Why the inventory may have had 20-year-old data, and whether the facility even was emitting 1,250 pounds of hydrazine in 1989, remain a mystery.

Questions to Guide the Study

- **Which pollutants are in good agreement between the ambient concentrations and the NATA model?**

- **Which pollutants are under-predicted between the ambient concentrations and the NATA model?**

- **Which pollutants are in over-predicted between the ambient concentrations and the NATA model?**

Background on NATA

- Conducted every three years
  - Began with 1996 assessment
  - Recently finished 4th assessment based on 2005 emissions (results just made public March 11, 2011).
  - http://www.epa.gov/ttn/atw/nata2005

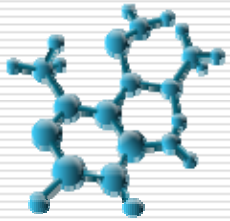
- Assesses cancer and/or noncancer risk for over 170 pollutants at the census tract-level (>66,000 census tracts in U.S.).

Background on NATA

- Sector specific results (point, area nonpoint, onroad, nonroad, background, etc.)
- Starting point is the 2005 NEI
- Other data sources include:
  - ambient monitoring data
  - chemical transformation/reactivity information
  - topography
  - population

Data Sources - NATA

- **2005 NATA**
  - For this analysis, specific receptor locations were modeled for over 100 HAPs.
  - Improved understanding of secondary formation and transformation of important HAPs (formaldehyde, acetaldehyde, acrolein, 1,3-butadiene).
  - Coke oven facilities: Emissions buoyancy accounted for.
  - Dose-response factors/unit risk estimates updated.



Data Sources – NATA NEI

- **2005 NATA National Emissions Inventory (NEI)**
  - All sectors (point, area nonpoint, onroad, nonroad, biogenic)
  - Criteria and HAPs
  - Primarily state/local/tribal data. Also integrates emissions data from EPA/other federal programs:
    - Clean Air Markets Division (CAMD)
    - Risk and Technology Review (RTR)
    - Toxic Release Inventory
    - Other studies (trade associations, Bureau of Ocean Energy Management, Regulatory, and Enforcement)

Data Sources – NATA NEI

- **2005 NATA National Emissions Inventory**
  - Several improvements compared to previous inventories
    - RTR and Lead NAAQS revisions were incorporated.
    - Certain nonpoint source categories disaggregated to point sources inventory (chrome plating, forest and wildfires).
    - Data from airports (19,000+) were added
    - MOVES model used for certain mobile source HAPs
    - Landfill emissions adjusted/removed

Data Sources – Ambient Monitoring

- **Phase VI Ambient Monitoring Archive**
  - Ambient monitoring archive of over 26 million HAP records
  - Timeframe: 1973-2007
  - 2005 year: 2.9 million HAP records. Composed of data from:
    - EPA's Air Quality Subsystem (92%)
    - Interagency Monitoring of Protected Visual Environments (IMPROVE) (7%)
    - Phase V historical archive (1%)

Methodology

- **Calculating Annual Averages**
  - Emission estimates are for an entire year
  - NATA model develops annual average concentrations
  - Procedure
    - Step 1: Extract 2005 ambient HAP data from the Phase VI archive.
    - Step 2: For sub-daily measurements (hourly, etc.), calculate valid daily measurements.
    - Step 3: Identify daily concentrations (by HAP, by site) which represent an entire year.
    - Step 4: Calculate annual average by HAP by site from the valid daily averages.

Methodology

- **Averaging Criteria – Valid Daily Averages**
  - Sub-daily measurements must have minimum 75% temporal coverage within a day:
    - Minimum eighteen 1-hour detected measurements
    - Minimum six 3-hour detected measurements
    - Minimum five 4-hour detected measurements
    - Use zero as a surrogate for non-detects

Methodology

- **Averaging Criteria – Valid Quarterly coverage**
  - Calendar quarter must have minimum 75% temporal coverage within a quarter:
    - Quarters are: January-March, April-May, June-August, September-December.
    - Minimum six pre-described sub-quarter zones with a valid daily average.
    - Sites sampling 1-in-12 days will have 7 or 8 samples within a quarter. More intensive sampling (1-in-6 days or 1-in-3 days) will have more opportunity to meet this criteria.

Methodology

- **Averaging Criteria – Annual Average**
  - Annual average must have minimum 75% temporal coverage within a year (i.e., three valid quarters)
  - If all criteria are met, average the valid daily concentrations and non-detects using zero as a surrogate.

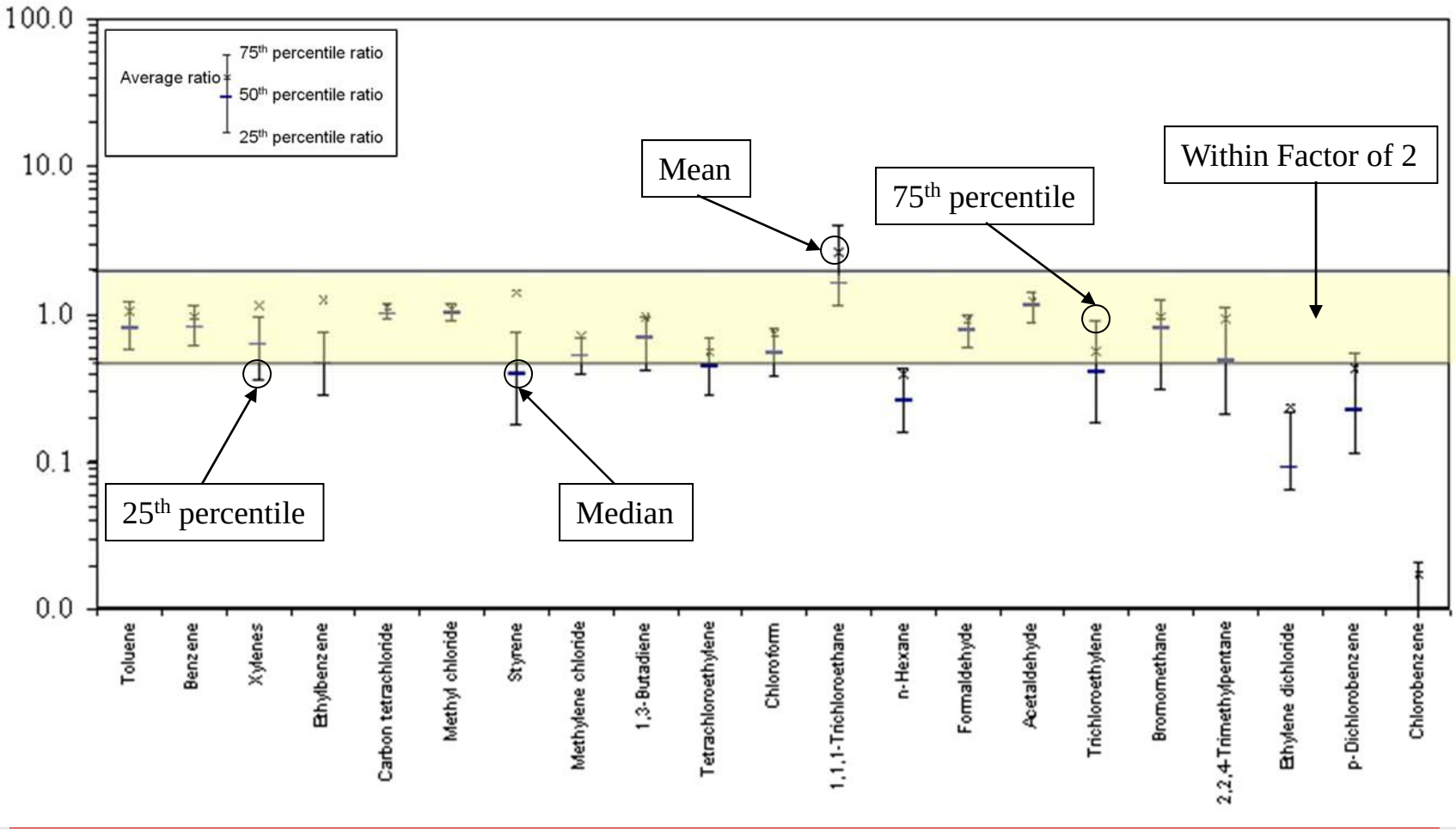
Methodology

- **Model-to-Monitor Comparison**
  - Simply divide model concentration by annual average concentration for each HAP and monitor.
  - Statistical distributions (minimum 25 monitors by HAP):
    - 25^{th}, 50^{th}, and 75^{th} percentiles
    - Average
    - Percent monitors within 10%, 20%, and 30%
    - Percent monitors within Factor of 2
    - Percent monitors under-estimated
    - Percent monitors over-estimated

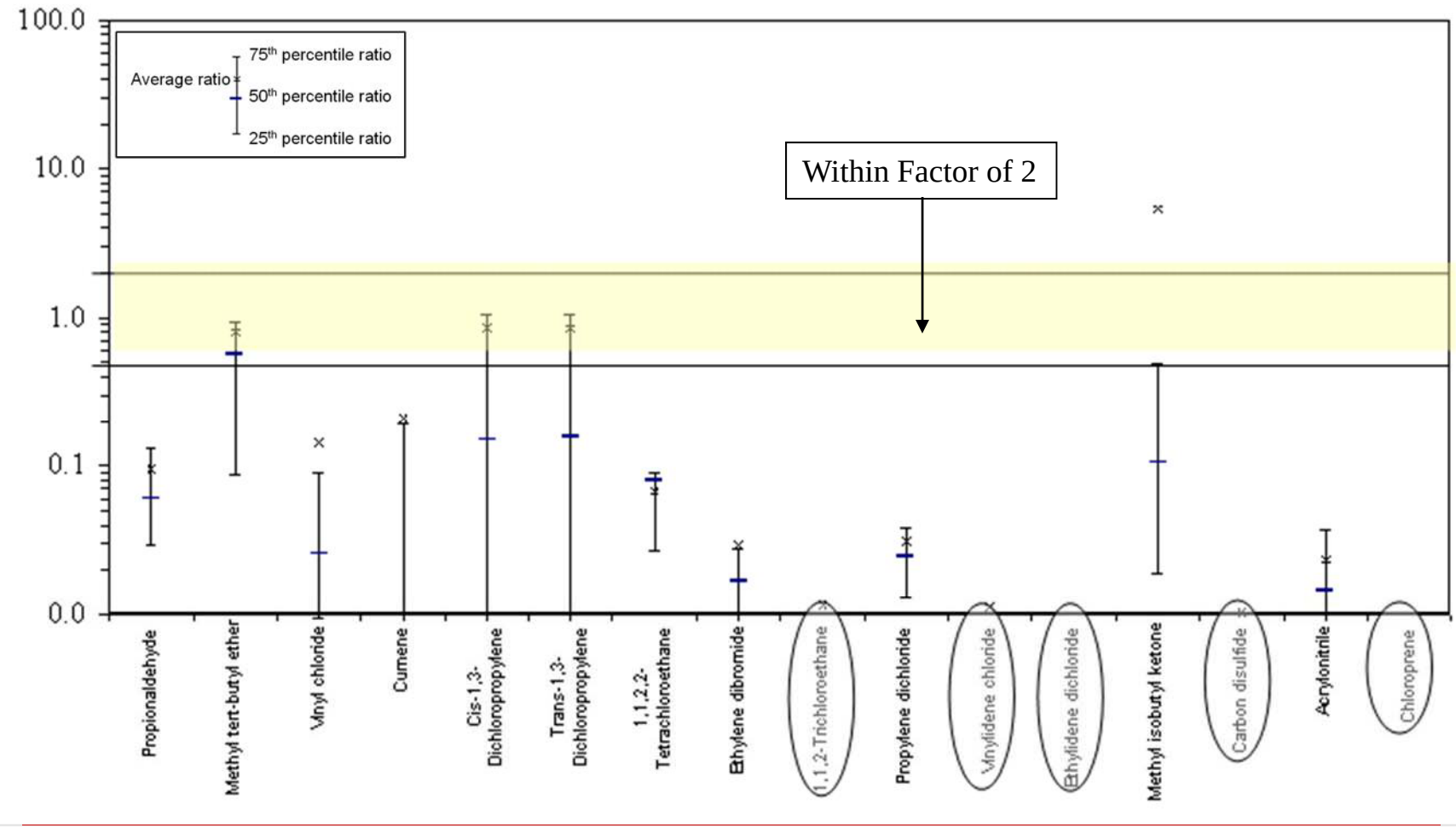
Results – Top 10 by Monitor Count

HAP	# Monitor	Median	Average	Average % Difference	% Mon. within 30%	% Mon. within 20%	% Mon. within 10%	% Mon. within Factor of 2	% Under-estimated
Toluene	297	0.826	0.984	7	39	25	11	75	67
Benzene	296	0.812	1.071	-2	48	32	15	82	66
Xylenes	266	1.284	3.465	16	28	20	8	59	77
Ethylbenzene	244	0.471	1.275	27	20	13	6	41	85
Carbon Tetrachloride	222	1.018	1.135	13	87	75	52	95	48
Methyl Chloride	206	1.030	1.083	8	85	70	35	98	43
Styrene	195	0.397	1.402	40	15	11	6	32	83
Methylene Chloride	190	0.524	0.726	-27	11	4	2	48	85
1,3-Butadiene	176	0.697	0.962	-4	31	17	8	56	76
				Within 10%		**>50%**		**>50%**	**<50%**

Gaseous HAPs (>100 monitors)

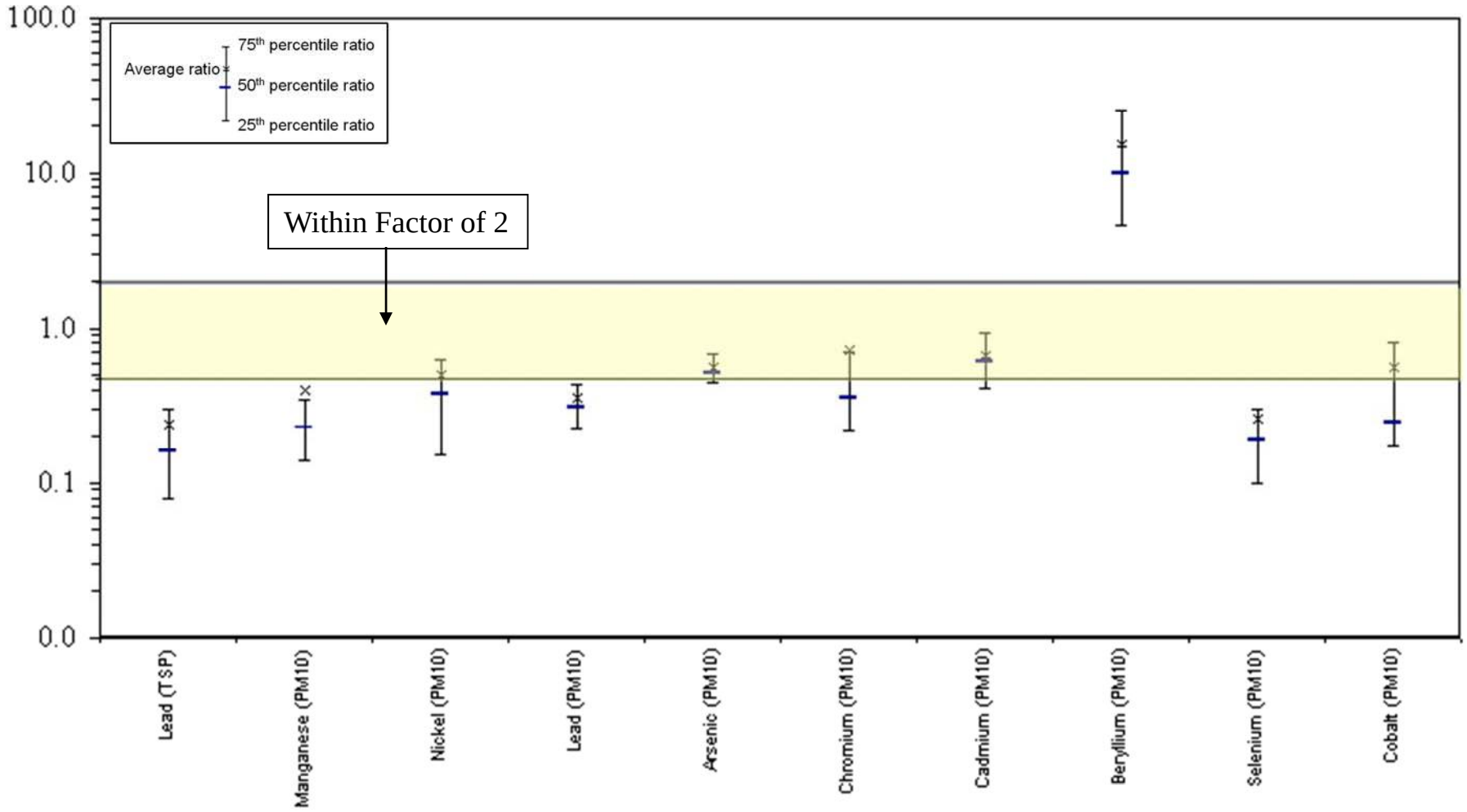


Gaseous HAPs (25-100 monitors)



Upper ends and interquartile ranges of the model-to-monitor ratios are below 0.010 and not presented.

TSP/PM_{10} HAPs



Results

- **Overall:**
  - 5,400+ model-to-monitor comparisons for 69 HAPs
  - 9% of all median ratios were between 0.9 and 1.1
  - 17%....were between 0.8 and 1.2
  - 25%....were between 0.7 and 1.3
  - 44%....were within a Factor of 2

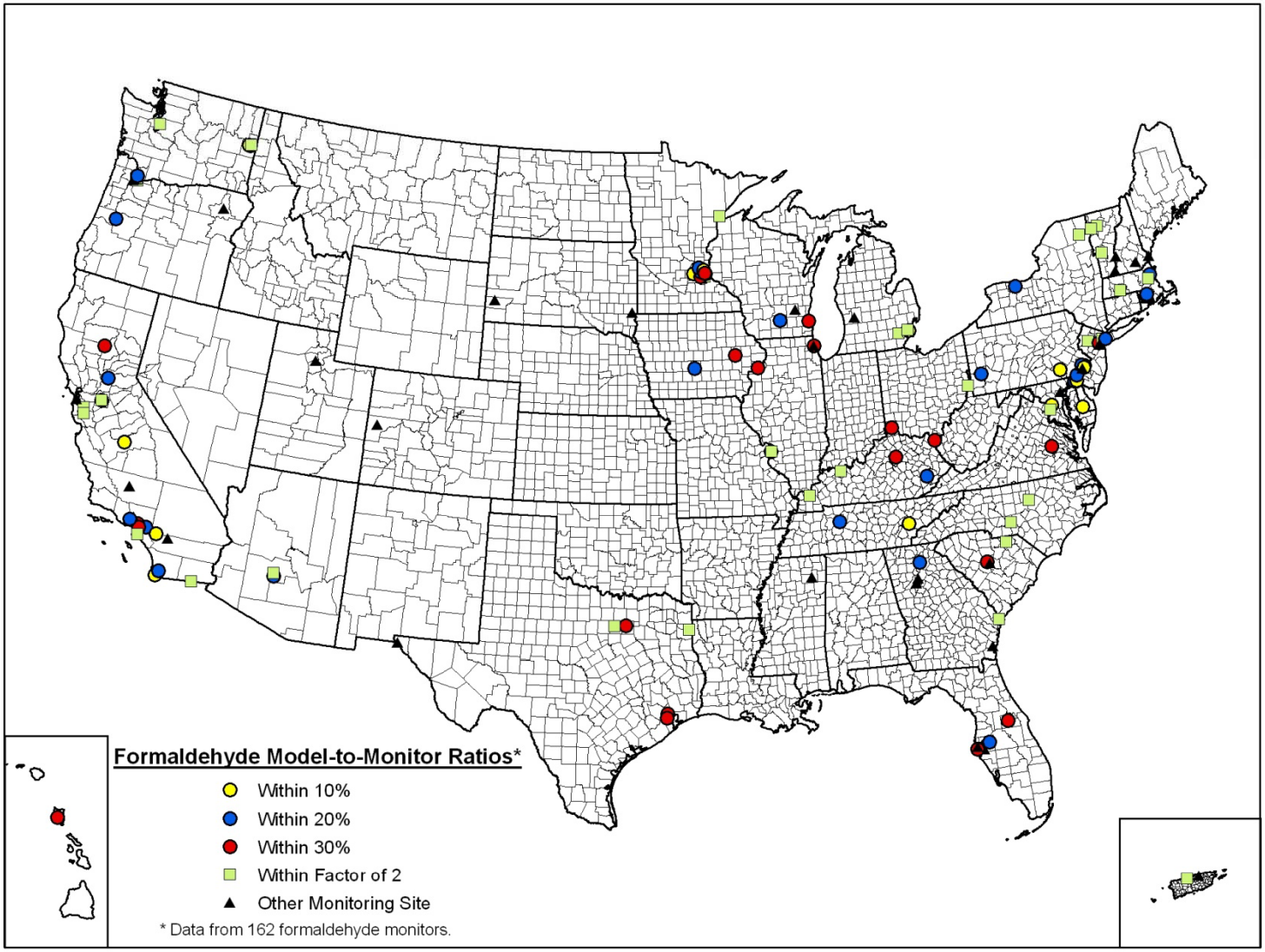
- **Carbon tetrachloride, methyl chloride, and arsenic (PM_{10}) had median ratios between 0.9 to 1.1**

- **Interquartile range within Factor of 2 for acetaldehyde, arsenic (PM_{10}), benzene, carbon tetrachloride, formaldehyde, methyl chloride, and toluene.**

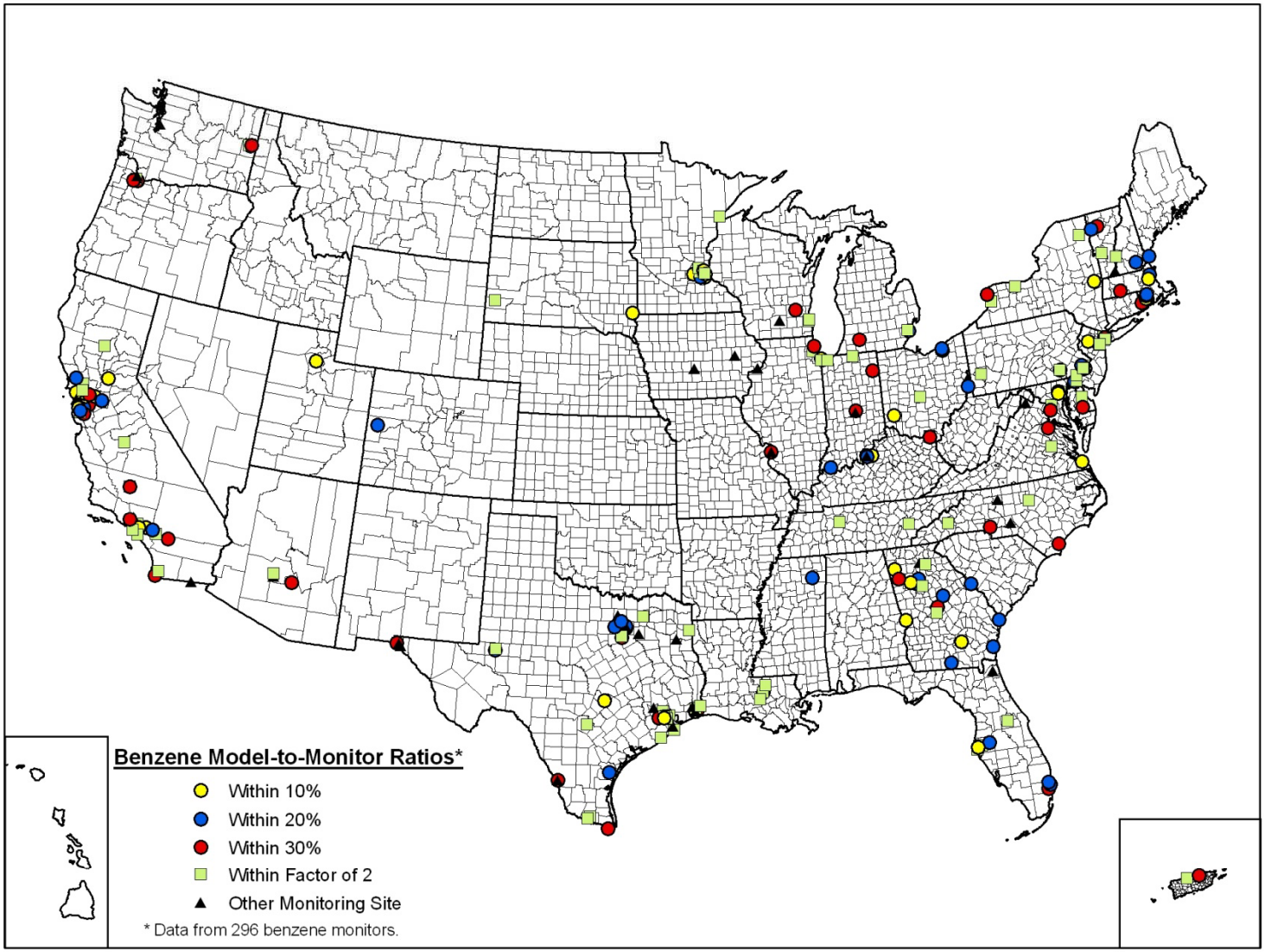
Results

- **Under-prediction (75th percentile < Factor of 2) for 19 HAPs:**
  - Ethylene dichloride
  - *n*-Hexane
  - 1,1,2-Trichloroethane
  - Acrylonitrile
  - Carbon disulfide
  - Chloroprene
  - Manganese (PM_{10})
  - Cumene
  - 1,1,2,2-Tetrachloroethylene
  - Vinyl chloride
  - Ethylene dibromide
  - Ethylidene dichloride
  - Methyl isobutyl ketone
  - Propionaldehyde
  - Propylene dichloride
  - Vinylidene chloride
  - Chlorobenzene
  - Selenium (PM_{10})
  - Lead (PM_{10} and TSP)
- **Over-prediction (25th percentile > Factor of 2) for 1 HAP:**
  - Beryllium (PM_{10})

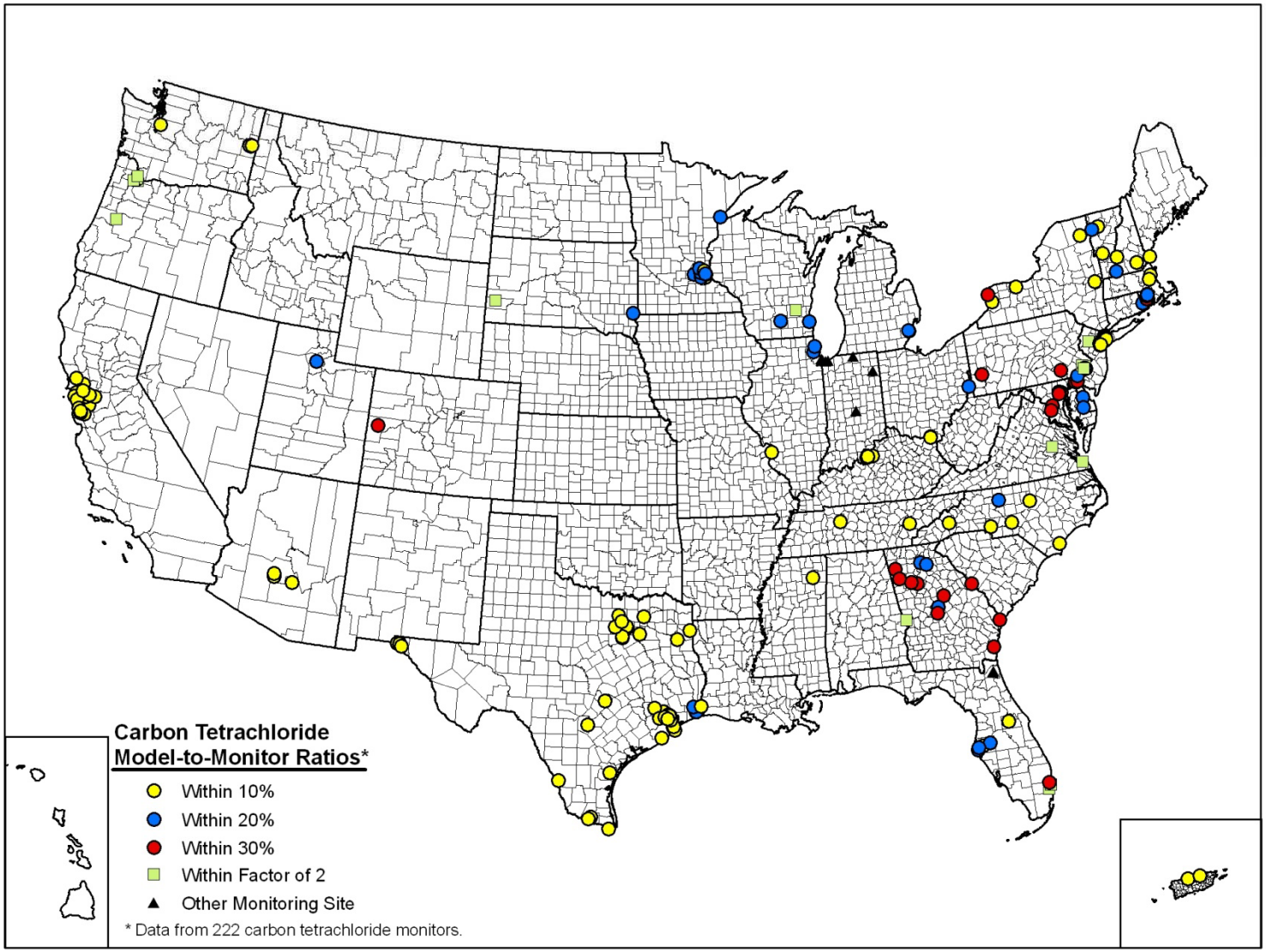
Geographic Dispersion - Formaldehyde



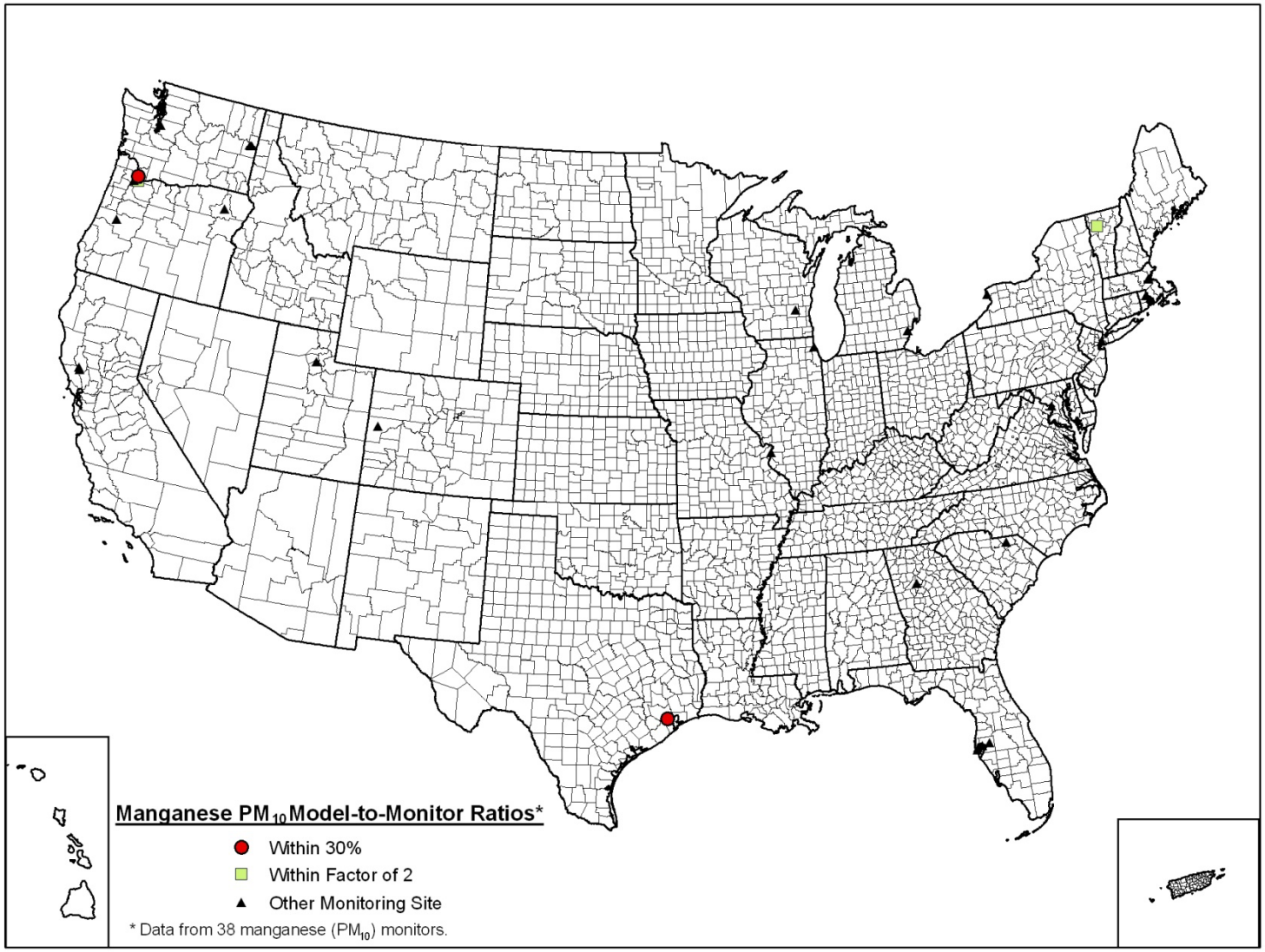
Geographic Dispersion - Benzene



Geographic Dispersion – Carbon Tetrachloride



Geographic Dispersion – Manganese (PM_{10})



Data Considerations

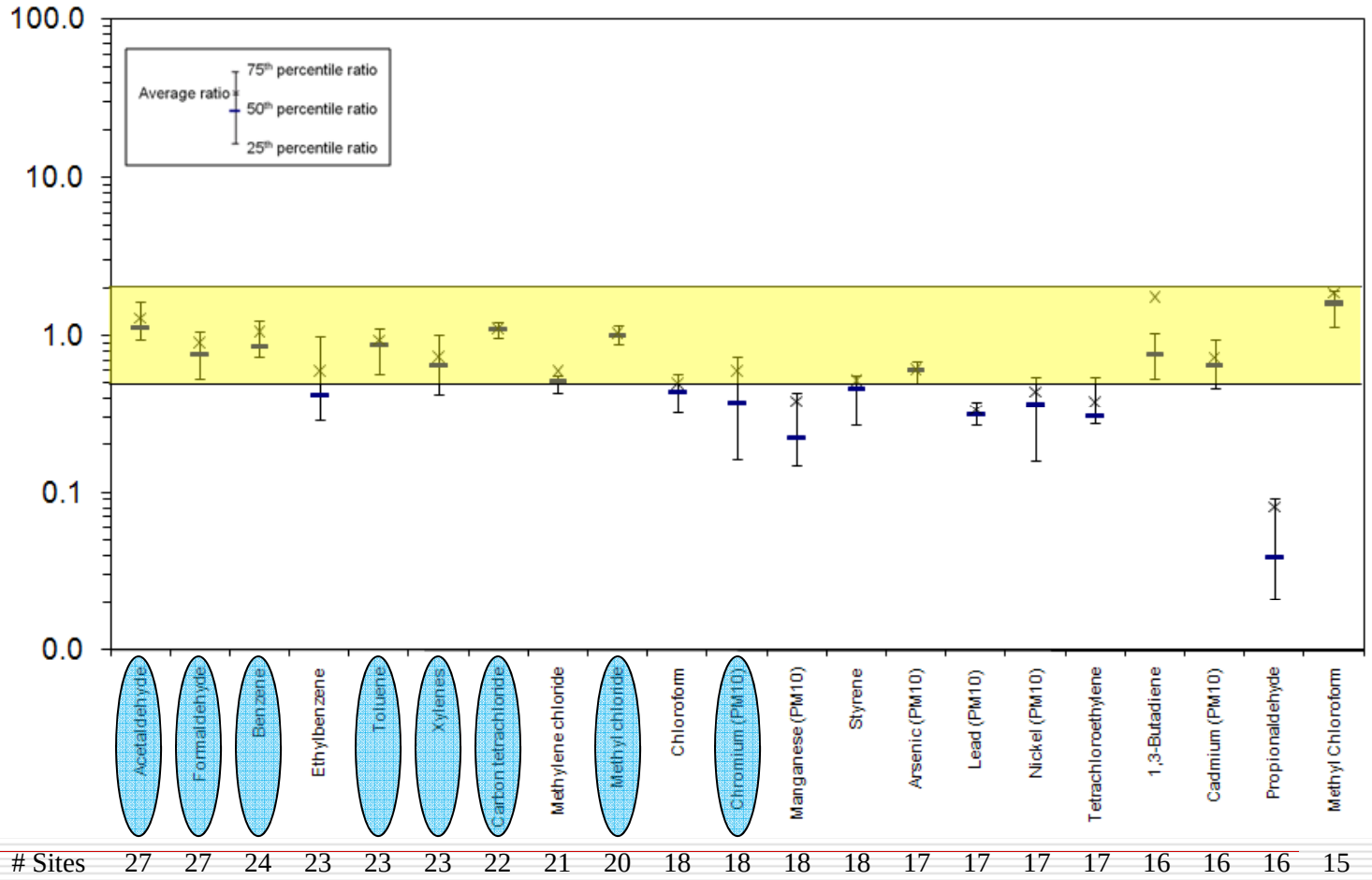
- **Uncertainties:**
  - Emissions characterization (i.e, location, emission rates, release parameters)
  - Meteorological characterizations (i.e., representativeness)
  - Model formulation and methodology (i.e., dispersion, plume rise, deposition)
  - Monitoring uncertainties (i.e., questions about acrolein, annual averaging techniques, etc.)
  - Background concentrations (i.e., representativeness)

Data Considerations

- **Under-estimation:**
  - NATA NEI may be missing specific emission sources
  - Emission rates may be under-estimated
  - Monitoring data/sampling methods/non-detects/averaging techniques
  - Background concentrations poorly characterized

Comparison to 2005 NATTS Only (min. 14 sites per pollutant)

At NATTS locations ("uniform" program sites), the Model-to-Monitor comparisons did well.



Average Model-to-Monitor Ratio was within 30%

Conclusion

- **Which pollutants are in <u>good agreement</u> between the ambient concentrations and the NATA model?**
  - Acetaldehyde
  - Arsenic (PM_{10})
  - Benzene
  - Carbon tetrachloride
  - Formaldehyde
  - Methyl chloride
  - Toluene

Conclusion

- **Which pollutants are <u>under-predicted</u> between the ambient concentrations and the NATA model?**
  - Ethylene dichloride
  - *n*-Hexane
  - 1,1,2-Trichloroethane
  - Acrylonitrile
  - Carbon disulfide
  - Chloroprene
  - Manganese (PM_{10})
  - Cumene
  - 1,1,2,2-Tetrachloroethane
  - Vinyl chloride
  - Ethylene dibromide
  - Ethylidene dichloride
  - Methyl isobutyl ketone
  - Propionaldehyde
  - Propylene dichloride
  - Vinylidene chloride
  - Chlorobenzene
  - Selenium (PM_{10})
  - Lead (PM_{10} and TSP)
- **Which pollutants are <u>over-predicted</u> between the ambient concentrations and the NATA model?**
  - Beryllium (PM_{10})

Questions?

Regi Oommen
919-468-7829
regi.oommen@erg.com

Barbara Driscoll
919-541-1051
palma.ted@epa.gov

Ted Palma
919-541-5470
palma.ted@epa.gov