COMPARISON OF 2002 MODEL-PREDICTED CONCENTRATIONS TO MONITORED DATA

As part of the 2002 National-Scale Air Toxics Assessment, EPA compared ASPEN-modeled concentrations with available, but geographically limited, ambient air quality monitoring data for the years 2002 through 2005. For each monitor-pollutant combination, EPA compared the annual average concentration estimated by the ASPEN model in the census tract where the ambient monitor is located to the annual average monitored value to get a point-to-point comparison between the model and monitor concentrations. EPA used an approach similar to that used for comparing the ASPEN model-to-monitor data for the 1996 and 1999 national-scale assessment except that EPA used updated emissions and monitor input data for the 2002 assessment; there were no major changes to model formulation. For more details about the model-to-monitor analysis for the 1996 national-scale assessment, see http://www.epa.gov/ttn/atw/nata/mtom_pre.html. For more details about the model-to-monitor analysis for the 1999 assessment, see http://www.epa.gov/ttn/atw/nata1999/99compare.html. Note that in this assessment, ambient chromium concentrations were compared to the sum of modeled chromium III and chromium IV concentrations. Chromium VI was measured at too few sites to provide a valid comparison for the model.

Table 1 shows the number of monitoring sites used in the 2002 comparison and the median ratio of model-to-monitor annual average concentrations by pollutant, on a point-to-point basis. The number of sites is the number of monitors with valid data. A large number of monitors means that more data are available which, in turn, facilitates an assessment of the degree of agreement between model and monitor data. The $PM_{2.5}$ metals (manganese, lead, arsenic, nickel, and selenium), benzene, toluene, and xylenes have the highest number of monitors. The number of available sites has increased substantially since the 1999 analysis. The median ratio is based on the model-to-monitor ratios for a given pollutant. A median close to 1 implies that the model overestimates the ambient concentrations about as often as it underestimates them. Methyl tert-butyl ether, acetaldehyde, and chloromethane all had median ratios between 0.9 and 1.1. The percent of sites estimated "within a factor of 2" is the percent of sites for which the model estimate is somewhere between half and double the monitor average. The "percent of sites estimated within 30%" is the percent of sites for which the model-to-monitor ratio is between 0.7 and 1.3. The "percent of sites underestimated" is the percent of sites for which the model-to-monitor ratio is below 1.

The degree of agreement between model-to-monitor data can be attributed to the following five uncertainties (which are the same identified in the 1996 and 1999 model-to-monitor comparison):

1. emission characterization uncertainties (e.g., specification of source location, emission rates, and release characterization);
2. meteorological characterization uncertainties (e.g., representativeness);
3. model formulation and methodology uncertainties (e.g., characterization of dispersion, plume rise, deposition,);
4. monitoring uncertainties; and

5. uncertainties in background concentrations.

ASPEN's limited ability to address the complex chemical transformation mechanisms needed to estimate ambient concentrations for highly reactive pollutants results in additional uncertainty for acetaldehyde and formaldehyde concentrations.

Figures 1 and 2 are box plots showing the distribution of the model-to-monitor ratios shown in Table 1. For example if there are 284 monitors measuring benzene, there are 284 model-to-monitor ratios to compute. EPA then computed the median of these 284 ratios as well as the percentiles to create the plot. The bottom of the box is the 25^{th} percentile, the top of the box is the 75^{th} percentile, and the horizontal line in the middle of the box is the median (i.e., 50^{th} percentile). If the model consistently agrees with the monitored data for the pollutant, the boxes will be narrow and centered at 1. Pollutants are organized alphabetically in two groups according to whether they are gases or embedded in particles. This side-by-side display of pollutants facilitates comparison to indicate which pollutants are being overestimated and underestimated, and which are estimated consistently. As in the 1996 comparison, the box plots do not show extreme percentiles (e.g., 10^{th} and 90^{th}) of the ratios because the extreme percentiles were far from the center of the distribution.

In this comparison, several assumptions about the monitoring data were made. Pollutants measured by fewer than 50 monitors and in limited geographical coverage (located in only one state) were excluded from the comparisons because the ability to assess model-to-monitor agreement is limited to that state or geographical area and does not extend nationwide. If annual average concentrations (e.g., >85% of the data were below the method detection limit) were not quantifiable using the monitor data, EPA also excluded the pollutant.

These results show that the interquartile range of model-to-monitor comparisons was within a factor of two for 1,3-butadiene, formaldehyde, acetaldehyde, chloromethane, carbon tetrachloride, benzene, toluene, xylenes, lead $PM_{2.5}$, and nickel $PM_{2.5}$. The remainder of the pollutants show various degrees of agreement. These results are an improvement over those found in the 1996 and 1999 national-scale assessment comparisons. However, the model is still underestimating several pollutants, most noticeably, acrylonitrile, chlorobenzene, isopropylbenzene, antimony, arsenic $PM_{2.5}$, manganese TSP, mercury $PM_{2.5}$, and selenium $PM_{2.5}$ all have 75^{th} percentile median ratios below 0.5. There are five possible reasons that ASPEN underestimates pollutant concentrations; these reasons also applied to the 1996 and 1999 assessments):

1. The National Emissions Inventory (NEI) may be missing specific emissions sources (for many of the sources in the NEI some of the emissions parameters are missing).

2. The emission rates may be underestimated. EPA believes the ASPEN model itself contributed only in a minor way to the underestimation. The modeled results from the ASPEN predecessor compared favorably to monitoring data in cases where the emissions and meteorology were accurately characterized and the monitors made more frequent readings.

3. There is uncertainty in the accuracy of the monitor averages, which, in turn, have their own sources of uncertainty. Sampling and analytical uncertainty, measurement bias, and temporal variation can all cause the ambient concentrations to be inaccurate or imprecise representations of the true atmospheric averages.

4. Model-to-monitor spatial comparisons are imprecise. The results suggest that the model estimates are uncertain on a local scale (i.e., at the census tract level). EPA believes that the model estimates are more reliably interpreted as being a value likely to be found within 30 km of the census tract location.

5. Background concentrations are poorly characterized. Most of the pollutants for which the model underestimated ambient concentrations were those for which background concentrations were not estimated. If background concentrations are a large fraction of ambient concentrations, the result would be large underestimations in model predictions.

Table 1. Agreement of 2002 model-predicted concentrations and ambient monitored concentrations on a point-by-point basis. Pollutants listed were monitored in at least 50 locations in several states.

Parameter	Number of Sites	Median of Model: Monitor	Percent Within Factor of 2	Percent Within 30%	Percent Underestimated
Manganese $PM_{2.5}$	343	0.73	64%	30%	67%
Lead $PM_{2.5}$	339	0.67	70%	32%	71%
Benzene	284	1.47	69%	29%	23%
Toluene	270	1.53	66%	28%	22%
Arsenic $PM_{2.5}$	260	0.09	12%	4%	92%
Xylenes	256	1.21	65%	32%	39%
Chloromethane	251	1.02	97%	81%	45%
Chromium $PM_{2.5}$	230	0.51	63%	32%	67%
Nickel $PM_{2.5}$	228	0.75	48%	23%	61%
Selenium $PM_{2.5}$	226	0.02	0%	0%	100%
Carbon Tetrachloride	224	1.17	97%	74%	17%
Styrene	217	0.46	35%	16%	76%
1,3-Butadiene	191	0.78	69%	29%	63%
Dichloromethane	187	0.75	65%	43%	79%
Formaldehyde	165	0.65	75%	32%	84%
Acetaldehyde	164	0.97	84%	52%	52%
N-Hexane	163	0.60	47%	23%	70%
Lead TSP	147	0.32	27%	12%	90%
Mercury $PM_{2.5}$	142	0.01	0%	0%	100%
Tetrachloroethylene	125	0.63	59%	26%	77%
Propionaldehyde	122	0.81	61%	29%	59%
2,2,4-Trimethylpentane	122	1.48	58%	25%	34%
1,4-Dichlorobenzene	120	0.41	29%	7%	74%
Chlorobenzene	115	0.05	12%	7%	81%
Methyl Tert-Butyl Ether	109	0.94	49%	20%	54%
Methyl Chloroform	102	1.99	46%	10%	14%
Isopropylbenzene	94	0.03	4%	4%	98%
Chloroform	86	0.82	63%	31%	60%
Chromium TSP	85	0.20	25%	12%	88%
Manganese TSP	80	0.14	8%	3%	94%
Trichloroethylene	76	0.48	37%	20%	71%
Methyl Isobutyl Ketone	75	2.57	28%	13%	25%
Antimony $PM_{2.5}$	63	0.07	0%	0%	100%
Nickel TSP	56	0.37	41%	18%	84%
Acrylonitrile	50	0.03	0%	0%	100%

Figure 1 - Model-to-Monitor Ratios for Gaseous HAPs

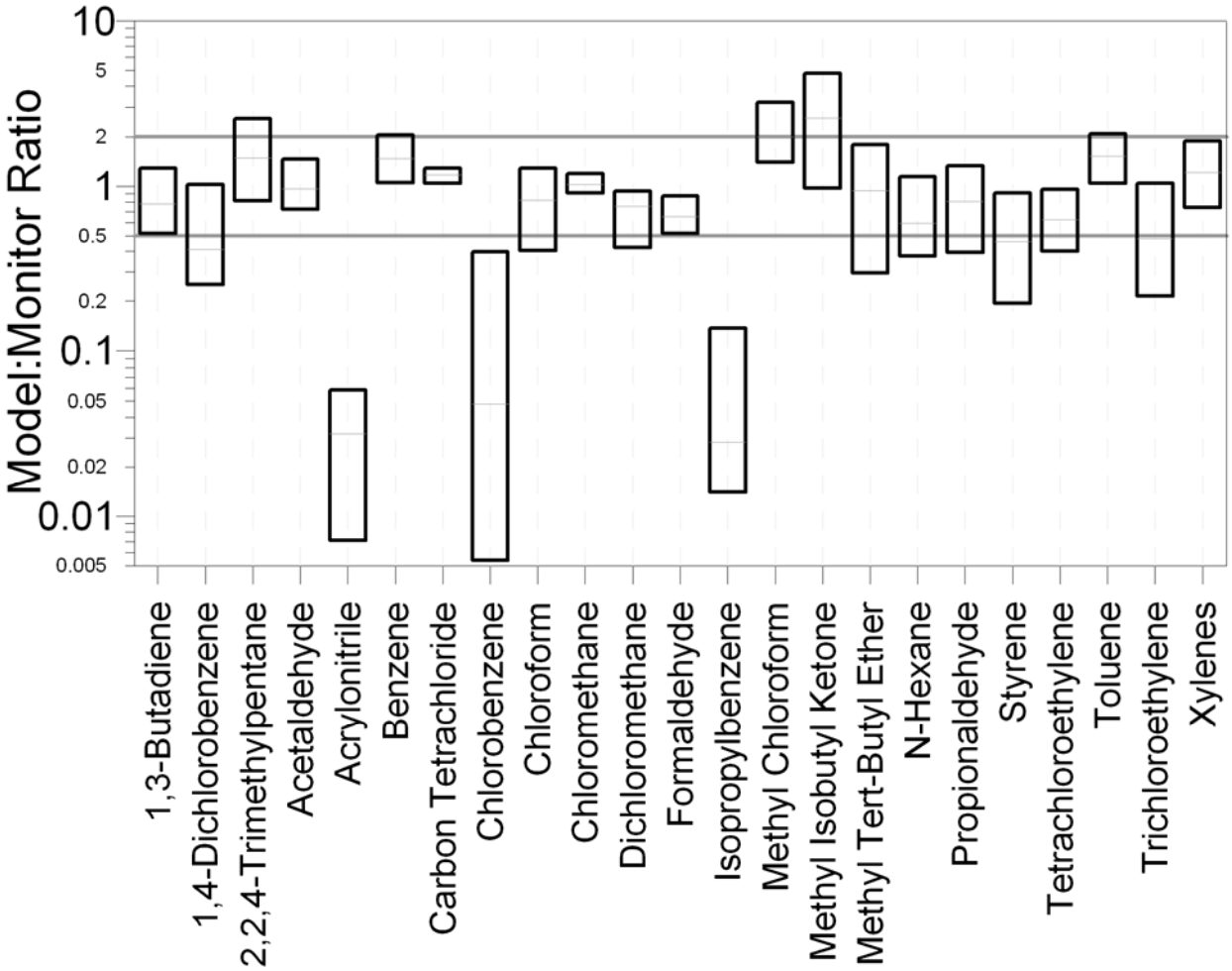

Figure 2 - Model-to-Monitor Ratios for Particulate HAPs