J4.4 LINKING AIR TOXICS CONCENTRATION FROM CMAQ TO THE HAPEM5 EXPOSURE MODEL AT NEIGHBORHOOD SCALES FOR THE PHILADELPHIA AREA

Jason Ching*[1], Thomas Pierce*[1], Ted Palma[2], William Hutzell[3],
Ruen Tang[4], Alan Cimorelli[5], and Jerold Herwehe[6]

[1]Atmospheric Sciences Modeling Division, ARL, NOAA, RTP, NC
[2]Office of Air Quality Planning and Standards, USEPA, RTP, NC
[3]National Exposure Research Laboratory, USEPA, RTP, NC
[4] Computer Sciences Corporation, RTP, NC
[5]Air Programs Division, Region III, USEPA, Philadelphia, PA
[6]Atmospheric Turbulence and Diffusion Division, ARL, NOAA, Oak Ridge, TN

1. INTRODUCTION

Historically, Gaussian plume models have provided estimates of ambient concentrations of air pollutants for input to human exposure models. However, most Gaussian-based modeling systems do not account for complex chemical reactions and struggle to account for background concentrations. The USEPA is developing the capability to link air toxics (AT) concentrations from an advanced photochemical grid model to the Hazardous Air Pollutant Exposure Model (HAPEM). The basis for AT modeling is the Community Multi-scale Air Quality (CMAQ) modeling system (Byun and Ching, 1999), a "one-atmosphere" chemical transport model. Because the AT model must simulate the spatial distribution of "toxic hot spots" across an urban area, the CMAQ system needs to account for specific toxic compounds at a fine-scale grid resolution. Recently, the HAPEM4 (www.epa.gov/ttn/atw/nata/modelexp.html) model was extended to consider concentration variability (HAPEM5). We are proposing a neighborhood-scale modeling paradigm that will couple AT concentration estimates from CMAQ at relatively fine grid resolutions to estimates of within-grid variability obtained from ambient concentration distribution functions (CDFs) developed for each grid cell. Under this paradigm, the CMAQ system will be linked to HAPEM5, which has the capability to incorporate information on the statistical variability of the ambient air pollutant concentrations. For the pilot study described in this paper, information provided to HAPEM5 will include the mean, median, and the 90th percentile of the concentration distribution. Because HAPEM5 typically requires at least one year of pollutant concentration values, CMAQ must provide a simulation over at least a one year period.

The CMAQ modeling system has been configured with a modified version of the Carbon Bond IV chemical mechanism that explicitly treats a number of gas-phase air toxic compounds. The system, known as CMAQ-AT, has been run for an annual period in a nested mode at 36, 12, and 4 km grid mesh resolutions using the 1999 National Emission Inventory (www.epa.gov/ttn/chief/net/) and meteorological outputs from 2001 simulations with the Penn State/NCAR Mesoscale Meteorological Model (MM5) (box.mmm.ucar.edu/mm5/). The 36 km grid mesh encompasses the continental United States, while the 12 and 4 km grid meshes encompass Philadelphia and Delaware.

In prior investigations of Philadelphia, Ching et al. (2004) employed a 1.3 km nest to produce finely-resolved concentration fields of photochemical pollutants. Because our pilot study requires annual simulations, the grid resolution has been initially restricted to a 4 km grid mesh to explore the feasibility of linking CMAQ-AT with human exposure models. Eventually, we plan to explore the use of finer-grid meshes and the use of the Industrial Source Complex (ISC) model with a high-resolution receptor network capable of providing within-grid concentration distributions of slow reacting species.

This paper highlights the following interim results: (a) outputs of CMAQ-AT for several toxic air pollutants (formaldehyde, acetaldehyde, acrolein, 1-3 butadiene, and benzene); (b) comparisons of model outputs to observations from an available monitoring site; and, (c) linkages of annual concentrations from CMAQ-AT to HAPEM-5.

*On assignment to the National Exposure Research Laboratory, U.S. Environmental Protection Agency. Corresponding author's email address: ching.jason@epa.gov

2. RESULTS

The CMAQ-AT results are presented here in a manner consistent with the input requirements for HAPEM5. Specifically, HAPEM5 requires that concentration values be grouped into eight 3-hourly annualized diurnal time blocks. For example, the first time interval represents the first three hours of each day (00-03 LT) averaged over the year, the second interval represents the next three hours (03-06 LT) averaged over the year and so on. HAPEM5 attempts to account for concentration variability by accepting as inputs the mean, median and 90th percentile values of each diurnal time interval. Thus, some of the CMAQ-AT results shown below will illustrate the temporal variability captured by the mean, median and 90th percentile values. In addition, we show results that illustrate the contributions from primary and secondary sources for formaldehyde, acetaldehyde and acrolein. We will focus on results extracted from two 4-km grid cells over central Philadelphia. Although HAPEM5 is designed to perform assessments on a census tract basis, we will assume as a first step that the concentration in each census tract can be associated with the grid cell overlaying the centroid of a census tract.

2.1 Results of CMAQ annual simulations at 4 km grid size

In Figure 1, formaldehyde concentrations from a single 4-km grid cell are grouped to show the annual time series for the eight 3-hourly time diurnal time intervals. The modeled results are taken from layer 1 of grid cell (26,47). Layer 1 is approximately 38 m deep. Figure 1 shows similar seasonal patterns for all diurnal time periods, although the differences appear slightly more pronounced for the nighttime periods.

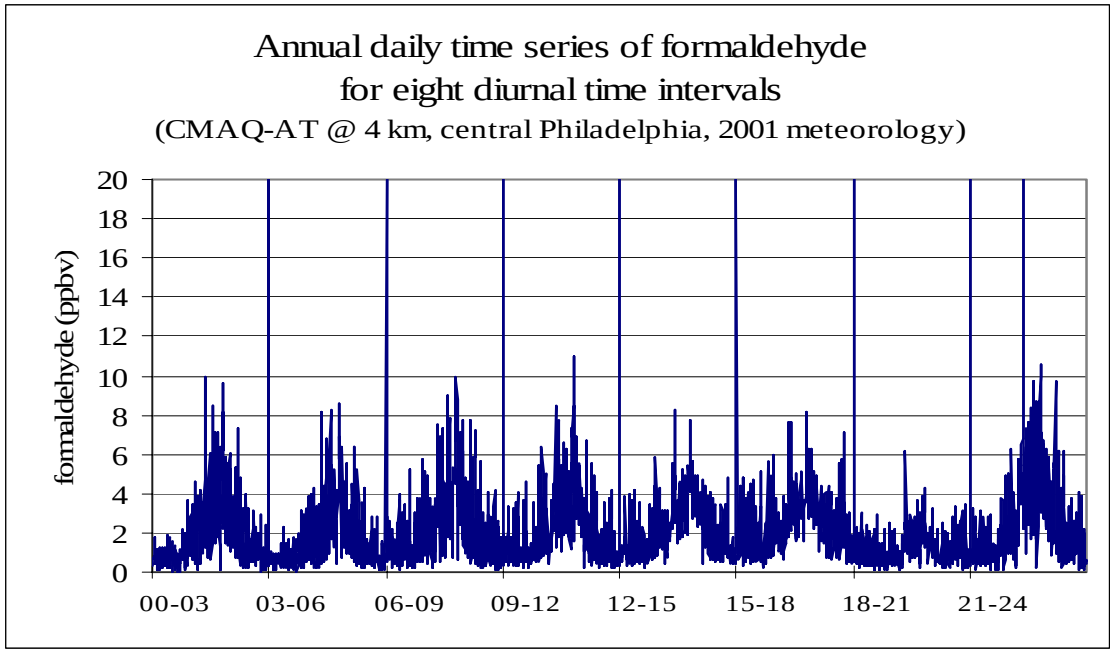

Figure 1. Daily concentrations of formaldehyde for 2001 as simulated in layer 1 of CMAQ-AT over central Philadelphia (26,47). Each of the eight time intervals contains three-hour averages for each day of the year; 00-03 corresponds to midnight to 3 a.m.

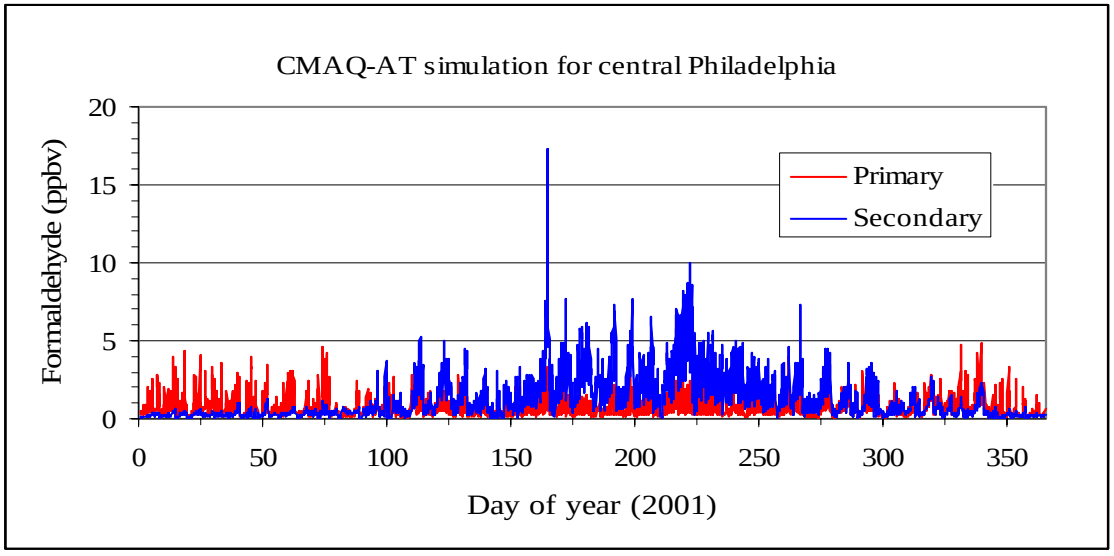

Figure 2a. Annual variability in primary and secondary formaldehyde as modeled by CMAQ-AT for a 4 km grid cell (26,47) over central Philadelphia. The data are grouped into three-hour averages from the 2001 simulation.

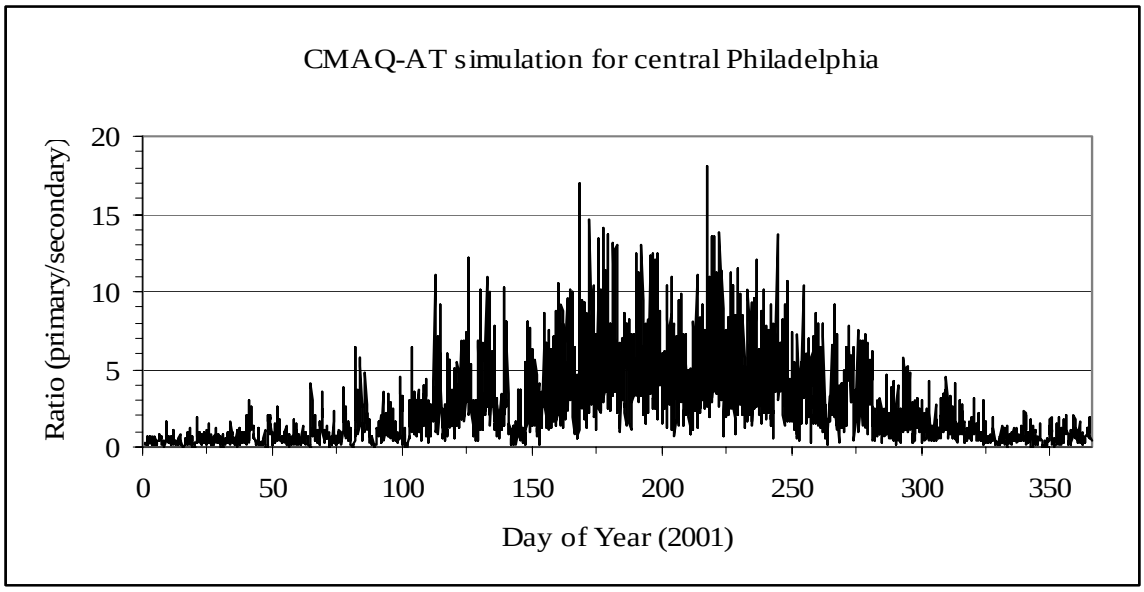

Figure 2b. Ratio of secondary formaldehyde to primary formaldehyde using the data shown in Figure 2a.

The second set of figures explores the differences between primary and secondary formaldehyde. In these figures, the concentrations are plotted as an annual time series, unlike the time interval values shown in Figure 1. CMAQ-AT allows formaldehyde to be tracked separately as a primary and a secondary species. Primary formaldehyde can be linked to the direct emission of formaldehyde, while secondary formaldehyde results from photochemical reactions, especially the reaction of isoprene with the hydroxyl radical. Figure 2a shows the annual time series of the primary and secondary formaldehyde species for a central Philadelphia grid cell. The primary species

varies less on a seasonal basis than the secondary species. Further, the contribution of the primary species to total formaldehyde is typically larger than that of the secondary species during the colder months, but the opposite is true during the warmer months. Figure 2b shows the ratio of the secondary to primary contributions for the same time period and grid cell. Figure 2b confirms that the secondary contribution is smaller than the primary contribution during the colder months, but the secondary contribution eventually greatly exceeds the primary contribution during the hotter summer months. This behavior is attributed to the increase in photochemical activity during warm, sunny periods.

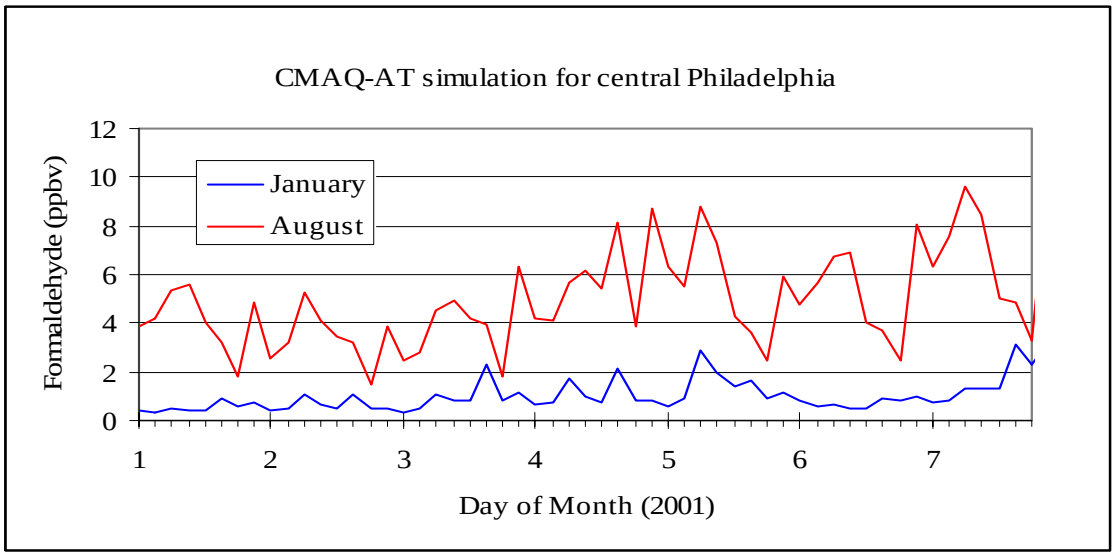

Figure 3. Formaldehyde concentrations from CMAQ-AT at a 4 km central Philadelphia grid cell (26,47) for a week in January and in August. Simulated concentrations are three-hour averages.

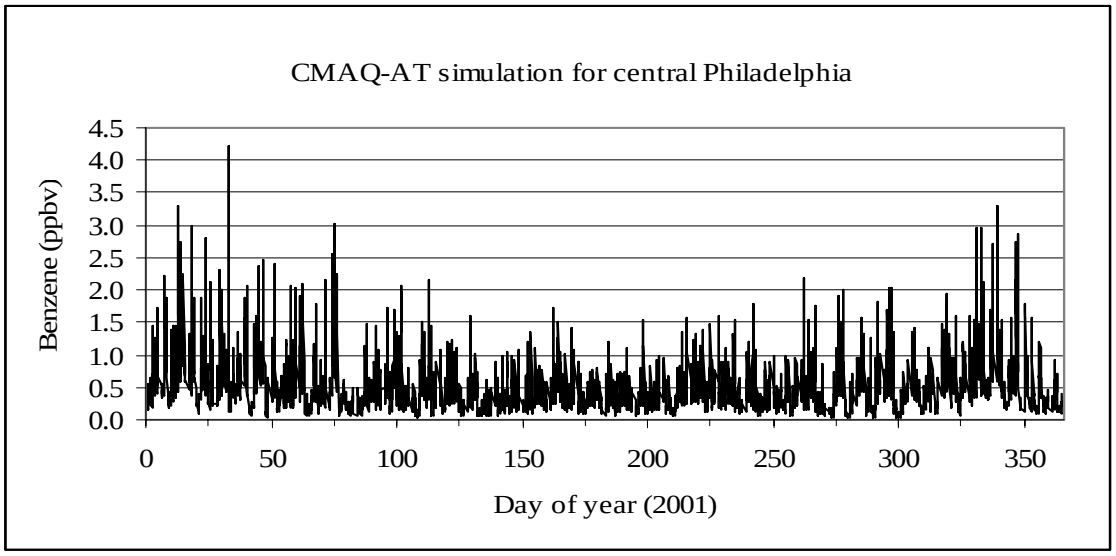

Figure 4. Annual time series of benzene simulated by CMAQ-AT for a 4 km grid cell (26,47) over central Philadelphia. Concentrations are taken from layer 1 and grouped into 3-hour averages.

Figure 3 compares the variation in modeled formaldehyde (in 3-hour averages) for a week in January 2001 and a week in August 2001. Diurnal variations appear relatively small for both months, especially for August. This suggests that the temporal variability for modeled formaldehyde may be due primarily to seasonal changes upon which are superimposed finer temporal variations due to synoptic events.

Results for a slower reacting compound, benzene, are shown in Figure 4. The annual time series for benzene is similar to that seen for the primary contribution of formaldehyde species, with peak values occurring during the colder months. This is attributed to the trapping of the pollutants during periods of lower mixing heights.

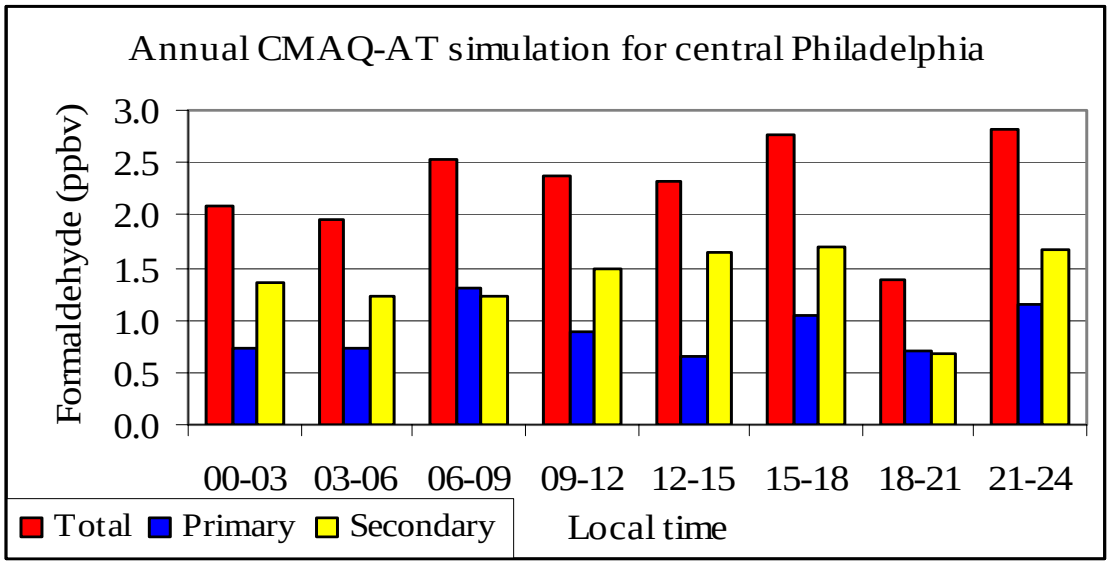

Figure 5. Diurnal variations of primary and secondary contributions to total formaldehyde for grid cell (27,46). Concentrations are averaged into 3-hour averages from the 2001 CMAT-AT simulation.

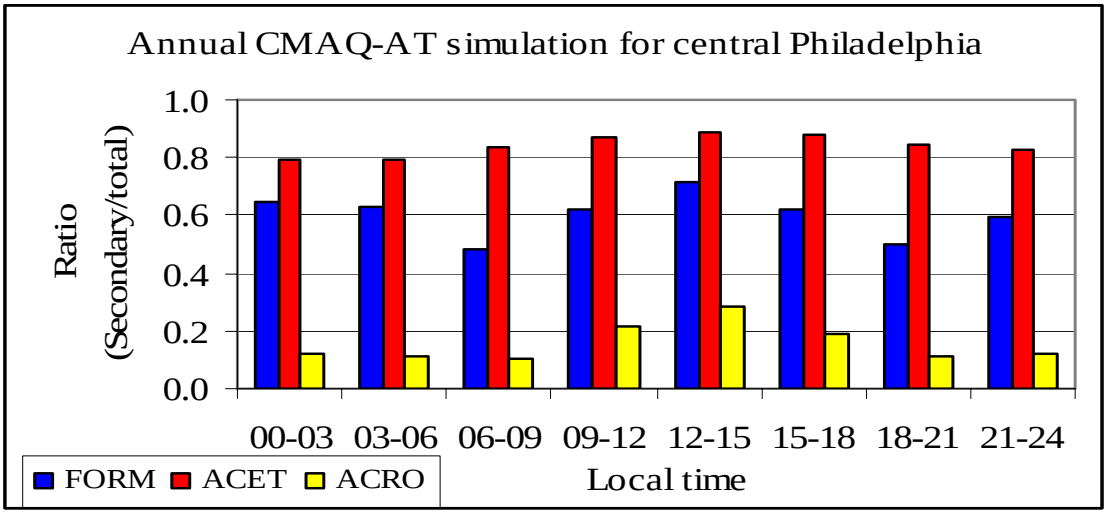

Figure 6. The secondary contribution of formaldehyde (FORM), acetaldehyde (ACET), and acrolein (ACRO) to total annual average concentrations from a CMAQ-AT simulation for central Philadelphia. Hourly concentrations are grouped into 3-hourly time intervals.

The next set of results provides relevant summary statistics of the annual simulations as needed for running HAPEM5. Results are shown from a 4 km grid cell over central Philadelphia, although these results are taken from a grid cell (27,46) just SE of the grid cell shown in Figures 1-4.

Figure 5 compares the primary and secondary contributions of formaldehyde annualized for the eight diurnal time periods. The secondary contribution to total formaldehyde is greater than the primary contribution for six of the eight time intervals. The two contributions are about equally divided during the morning (06-09 LT) and evening (18-20 LT) commuter traffic periods.

Figure 6 shows, on an annual basis for most diurnal time periods, that the secondary contribution is greater than the primary contribution for formaldehyde and acetaldehyde. For modeled values of acrolein, most of the contribution is primary. This suggests that reactive toxic compounds should be modeled explicitly since the role of atmospheric chemistry varies by chemical species.

A statistic that has direct relevance to HAPEM5 is the 90th percentile of the annual concentration. Figure 7 compares that 90th percentile value to the mean value for three toxic pollutants. Ratios are shown for the eight annualized 3-hour time intervals needed for input to HAPEM5. In general, the ratio of the 90th percentile to the mean is about a factor of two for formaldehyde, aceteldehyde and acrolein. Differences in the ratios are surprisingly small between the three pollutants. Diurnally, the ratio shows some variability with higher values for formaldeyde during the early morning (03-06) and for acrolein during the late afternoon (15-18).

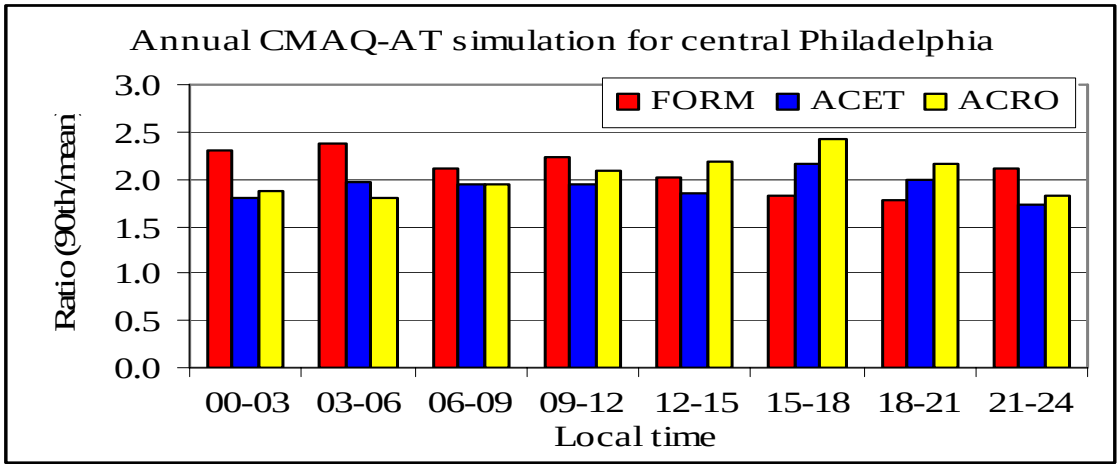

Figure 7. Comparison of 90th percentile versus mean values of formaldehyde (FORM), acetaldehyde (ACET), and acrolein (ACRO) as computed for each diurnal time period (3-hour averages) for grid cell (27,46).

2.2 Comparison of monitoring and model outputs

Although detailed field observations for the Philadelphia modeling domain are lacking, air toxic concentration measurements are available from a single monitor in Camden, New Jersey that is part of EPA's Urban Air Toxic Monitoring Program (UATMP) (ERG, 2002). Observations taken as part of the UATMP were made over a 24 h period every 6 to 12 days. The Camden site is located just east of Philadelphia in a semi-industrial area. The primary emission sources are located mainly to the west and to the north of the Camden site. Further details on the site and the sampling protocol are available from ERG (2002).

Concentrations from this monitoring site for a few toxic compounds were compared to matching modeled concentrations simulated with the 4-km and 36-km versions of CMAQ-AT for 2001. Concentrations from layer 1 of the grid cell overlaying the Camden site were extracted from the two CMAQ-AT simulations. Comparisons for 1,3-butadiene, formaldehyde, acetaldehyde, and benzene are shown in Table 1. In general, the modeled mean values compared reasonably well against the observed values, and the means, standard deviations, and correlations from the 4 km

version of CMAQ-AT compared more closely with the observed values than the modeled values extracted from the 36 km version of CMAQ-AT. While this limited comparison is encouraging, to properly assess and evaluate model performance will require observational data from more than one monitor location.

Table 1. Comparison of air toxic concentrations measured at the Camden, NJ, site to CMAQ-AT (layer 1). All samples, except benzene, are 24-hour averages; benzene is a 1-hour average.

		Mean (ug/m^3)			Std. deviation (ug/m^3)			Correlation	
Compound	n	Obs	4 km	36 km	Obs	4 km	36 km	4 km	36 km
1,3-Butadiene	28	0.33	0.18	0.12	0.34	0.12	0.08	0.07	0.09
Formaldehyde	44	3.68	2.91	2.25	3.21	2.13	1.52	0.42	0.38
Acetaldehyde	44	2.09	2.49	1.92	1.42	1.20	0.78	0.45	0.44
Benzene	1328	1.11	1.02	0.77	1.06	0.71	0.40	0.48	0.41

2.3 HAPEM5 Results

2.3.1 Model Linkage

An air quality dispersion model, such as the CMAQ model, estimates an ambient concentration for a given time period at a given location. If a human stayed at a fixed location for a specified time period (in this case, one year), then the ambient concentration predicted by an air quality dispersion model would equal the "apparent" exposure, or the concentration available for the human to breathe. In the real world, however, people generally move from location to location (e.g., from home to work, or home to school). Also, most people do not spend their entire day outdoors; a majority of time is spent in indoor locations (e.g., the home, workplace, school, or vehicle). Studies have shown that air quality concentrations in indoor environments can be quite different than those in the outdoor environment. Because of these factors, a human exposure model is generally employed to consider these factors and predict the "apparent" inhalation exposure.

The HAPEM5 is a screening-level exposure model designed to predict the "apparent" inhalation exposure for the general population, or a specific sub-population, over spatial scales ranging from urban environment to nationwide.

HAPEM5 uses the general approach of tracking representatives of specified demographic groups as they move among indoor and outdoor microenvironments and among geographic locations. The estimated pollutant concentrations in each microenvironment visited are combined into a time-weighted average concentration, which is assigned to members of the demographic group.

HAPEM5 uses four primary sources of information: population data from the US Census, population activity data from human diary data, microenvironmental data, contained within the model and air quality data that is provided for the study region.

As human activity data generally exhibits a diurnal pattern the air quality data provided to HAPEM5 must capture the expected diurnal pattern. The hourly CMAQ results were utilized to build a diurnal temporal pattern by averaging the model results over 3-hour blocks (i.e., midnight - 3am, 3am-6am…) for the entire year. To help exhibit the range in the diurnal pattern, similar patterns were built utilizing the median and 90th percentile ambient levels. Such temporal information is useful for risk assessment in both bounding the range of potential exposure levels and in helping to define the annual variability in exposure.

As ambient predictions from CMAQ are calculated on a regularly spaced grid, the HAPEM5 model is designed to perform its exposure assessments on a census tract. Thus, for input to HAPEM5, the CMAQ air quality estimates were converted to a census tract resolution by selecting the grid concentration overlaying each census centroid. In an urban area, such as the Philadelphia study area, even with a 4 km grid, this generally results in multiple census tracts residing a single grid cell. The study area for HAPEM5 included the 381 census tracts in the Philadelphia area.

Ambient input files for HAPEM5 have been built from the CMAQ-AT diurnal distributions at the census tract level. For comparison in this paper, three CMAQ-AT diurnal distributions (mean, median, and 90th percentile) were examined for two pollutants: a reactive pollutant (formaldehyde) and a relatively non- reactive pollutant (benzene).

2.3.2 HAPEM5 Model Results

A total of six HAPEM5 simulations were made using the three diurnal distributions and two pollutants (benzene and formaldehyde). The average exposure for the Philadelphia area estimated from these simulations, along with the draft 1999 NATA HAPEM5 results which uses a Gaussian plume model (USEPA, 2004), is presented in Table 2. Future analyses will examine the spatial variability of these exposure estimates across the individual census tracts in Philadelphia.

As a benchmark, the CMAQ results are compared to the HAPEM5 model results from the 1999 National Air Toxic Assessment (NATA). The CMAQ-AT‘s estimates for benzene are about half of that predicted by NATA and about 85% of the NATA estimated formaldehyde exposure values.

Table 2. Average Philadelphia exposure levels (ug/m^3) computed with HAPEM5.

	CMAQ-AT 4 km			NATA
Compound	Mean DD	Median DD	90% DD	Mean DD
Benzene	1.26	0.96	2.63	2.23
Formaldehyde	2.15	1.63	4.60	2.57

DD= diurnal distributions pattern

An examination of the temporal variation of the diurnal patterns shows that exposure levels are almost doubled as compared to using average annual patterns. The 90th percentile estimate provides an upper estimate that is comparable to day-to-day variations in ambient levels. Toxicologists believe that capturing this variation is important in characterizing both acute as well as chronic risk (USEPA, 2001).

3. SUMMARY

This paper has shown that a sophisticated chemical grid model can be used to provide the air toxic concentration fields needed to drive an exposure model. A comparison of the model results with a limited set of observations suggests that the model performance is reasonable. For this pilot study, air toxic concentrations generated by the CMAQ-AT model for a 4 km grid mesh overlaying Philadelphia were successfully formatted for direct input to the human exposure model HAPEM5.

Disclaimer: *This paper has been reviewed in accordance with United States Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication.*

4. REFERENCES

Byun, D. and J. Ching, 1999: Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, EPA/600/R-99/030, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Ching, J., S. Dupont, J. Herwehe, T. Otte, A. Lacser, D. Byun, and R. Tang, 2004: Air quality modeling at coarse-to-fine scales in urban areas, In *Proceedings of the Sixth AMS Conference on Atmospheric Chemistry: Air Quality in Megacities*, American Meteorological Society, Seattle, Washington, January 11-15, 2004.

ERG, 2002: 2001 Urban Air Toxics Monitoring Program (UATMP), EPA/R-02-010, Environmental Protection Agency, Research Triangle Park, NC. www.epa.gov/ttn/amtic/airtxfil.html

USEPA, 2001: Evaluating the National-Scale Air Toxics Assessment 1996 Data - An Advisory by the EPA Science Advisory Board , U.S. Environmental Protection Agency Science Advisory Environmental Board (1400A), Washington, DC, EPA-SAB-EC-ADV-02-001, December 2001.

USEPA, 2004: The ASPEN Model, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. [Available from www.epa.gov/ttn/atw/nata/aspen.html]]