4.1 A MONTE CARLO STUDY OF UNCERTAINTIES IN BENZENE AND 1,3-BUTADIENE CONCENTRATIONS CALCULATED BY AERMOD AND ISC IN THE HOUSTON SHIP CHANNEL AREA

Steven Hanna*[1], Robert Paine[2], David Heinold[2], Elizabeth Kintigh[2], and Dan Baker[3]
[1]Hanna Consultants, Kennebunkport, ME; [2]ENSR, Westford, MA; [3]Shell Global Solutions, Houston, TX

1. INTRODUCTION

This paper describes a Monte Carlo probabilistic uncertainty analysis related to air toxics studies in urban regions, expanding upon the strategy of the EPA (2000) Integrated Urban Air Toxics Study. That study was a deterministic one and presented a framework for addressing air toxics in urban areas using census-tract population centroids for receptors. A test example for the Houston area was described by the EPA (2002), using a straight-line Gaussian plume model (ISCST3). Given the complexities of studying air toxics in urban environments, an investigation of uncertainty was considered a useful extension of the original work.

The Monte Carlo (MC) probabilistic uncertainty approach is used here because it allows the combined influences of the uncertainties in many model inputs and parameters to be assessed. The resulting total uncertainty in the model outputs can be determined as well as correlations between uncertainties in inputs and outputs. Basic explanations of the MC procedures are provided in several books (e.g., Cullen and Frey, 1999) and examples of applications to atmospheric transport and dispersion models have also been published (e.g., Irwin et al. 1987 and Hanna et al. 2001 and 2005b).

The current uncertainty study began with the Houston example for the year 1996 as described by the EPA (2002), who used a 150 km by 150 km urban geographic domain, and used ISCST3 to calculate annual average concentrations of five toxic pollutants. The new study focuses on a smaller 15 km by 15 km Houston domain (see the inner square in Figure 1) covering the area around the Houston Ship Channel, and including many oil refineries and chemical processing plants, as well as numerous major highways. The modeling addresses the pollutants benzene and 1,3-butadiene, whose emissions are distributed among mobile sources, industrial sources, and area sources such as service stations. The concentrations in the inner square in Figure 1 have been calculated using emissions information from sources on a 30 km by 30 km domain, indicated by the outer square in the figure.

Annual averaged concentrations at 46 receptors (43 at census tract population centroids and three at monitoring sites) have been calculated on the Houston 15 km by 15 km receptor domain. The locations of the receptors are shown in Figure 2.

Two alternate straight-line Gaussian plume dispersion models have been run: ISCST3 (EPA, 1995) and AERMOD (Cimorelli et al., 2005). The model base runs for the Houston domain in Figure 1 were available from an earlier sensitivity study by Heinold et al. (2003). However, the emissions files that were ultimately used were provided by the EPA for the calendar year 1996 and are consistent with those used by the EPA for their similar Monte Carlo uncertainty study.

In the MC exercise, the models have been run 100 times for random choices of variations from the distributions characterizing the input parameters, and the responses of the key model output parameters are analyzed. 100 MC runs have been made for the combined emissions and dispersion model for the AERMOD, ISCST3 (all urban), ISCST3 (all urban grid sources), and ISCST3 (mixed rural/urban grid sources) dispersion models.

The first step in any uncertainty analysis is to clearly define the scientific questions being asked and the model outputs to be analyzed (Cullen and Frey, 1999). The following questions were addressed in this study, and answers are given is Section 7:

---

* *Corresponding author address*: Steven R. Hanna, Hanna Consultants, 7 Crescent Ave., Kennebunkport, ME 04046-7235; e-mail: hannaconsult@adelphia.net

Question Set 1 (uncertainty of modeling system): What is the total uncertainty in the annual average concentration (averaged over the 43 census tract population-centroid receptors) of benzene and 1,3-butadiene concentrations in the Houston region, and which input variables and model parameters have the most influence on this total uncertainty? What is the total uncertainty in the magnitudes and the locations of the 100 maximum annual-averaged concentrations calculated by each MC run at the 43 population centroid receptors and the three monitors?

Question 2 (contributions of emissions vs dispersion models to uncertainty): What is the relative uncertainty between the emissions and the transport and dispersion model?

Question 3 (contributions of emissions source classes): How do the total uncertainties and correlations differ for different source classes, such as mobile versus point, or industrial major point source versus industrial area and volume? And how would these differences impact conclusions regarding source apportionment?

Question 4 (dependence on model): Do the conclusions concerning uncertainty depend on model used (e.g., ISCST3 versus AERMOD or ISCST3 with EPA rural-urban designations versus ISCST3 with all-urban designations)?

The current paper is a shortened version of the project report by Hanna et al. (2005a). That report contains detailed tables and figures and additional justifications for the assumptions regarding input uncertainties.

2. ESTIMATES OF INPUT UNCERTAINTIES FOR EMISSIONS

There are uncertainties and mean biases in the emissions estimates for any region, but quantifying these is rather difficult since few data are available. In an earlier MC study by Hanna et al. (2001) involving photochemical grid models, the emissions uncertainties were assumed to have approximately a factor of three uncertainty (i.e., a 95 % range) and a log-normal distribution.

Uncertainties in emissions can be estimated by a combination of two approaches: analysis of available data, and expert elicitation (Cullen and Frey 1999). As an example of the first approach, emissions data available for benzene were used to derive probability density functions describing the uncertainties in emissions for several categories in Houston (Frey and Zhao, 2003). Fortunately, the emissions inventory in Houston for benzene and 1,3-butadiene is more complete than in most other parts of the country (Frey and Zhao, 2003). Furthermore, a key element of the current study has been the August 2003 joint EPA-API Workshop in Houston on uncertainties of emissions in benzene and 1,3-butadiene. For this study, the participants defined the benzene categories listed in Table 1, where the emission percentage for each category is listed. There are other emissions in other categories, but those emissions are relatively small. Most of the individual category uncertainties discussed at the Workshop were found to be in the range of plus and minus a factor of 1.5 to 3. Therefore it was decided, for the purposes of the current study, to simply assume a plus and minus factor of three uncertainty, with a log-normal distribution, for each category. This factor of three is assumed to cover the 95 % range of the uncertainty (or plus and minus two standard deviations). Also, after much discussion at the Workshop and afterwards, it was decided that there was not enough information to assume anything other than zero for the mean biases.

Using the same process, the Workshop participants discussed the 1,3-butadiene emissions inventory for the Houston Ship Channel domain and decided on the 13 emissions categories listed in Table 2, which including a description in words plus the percentage of the total. A smaller percentage (about 15 %) of the 1,3-butadiene emissions comes from mobile sources, compared to about 40 % for benzene.

There are no correlations assumed between any of the categories in Tables 1 and 2. To prevent unrealistical extremes in emissions from being selected by the MC random number process, it is assumed that there are no emissions that depart from the median by more than plus and minus five standard deviations (i.e., plus and minus a factor of 7.5 for the emissions uncertainties).

For all categories in Tables 1 and 2, the uncertainty in the MC study is represented by a distribution function (log-normal) and a definition of the 95 % range (i.e., plus and minus two standard deviations). None of the inputs is assumed to have a mean bias. That is, the medians

of the distributions are assumed to equal the value of that variable used in the base model run.

3. ESTIMATES OF METEOROLOGICAL AND DISPERSION MODEL INPUT UNCERTAINTIES FOR ISCST3 AND AERMOD

The uncertainties of the meteorological and dispersion model inputs and model parameters must also be estimated for ISCST3 (EPA, 1995) and AERMOD (Cimorelli et al, 2005). The averaging time for this modeling exercise is one year, since that is the appropriate averaging time for determining health effects of benzene and 1,3-butadiene. However, it is the standard procedure with both ISCST3 and AERMOD to run the model for every hour of the year, and then use a postprocessor to calculate the one year averaged concentration.

The uncertainties in transport and dispersion model inputs and parameters are difficult to estimate and there is much discussion about problems with various methods (Freeman et al. 1986, Irwin et al. 1987, Hanna 2002). This is a fairly new field of research in the atmospheric sciences. Fortunately, in most Monte Carlo uncertainty exercises, the final estimates of total uncertainty in model outputs are influenced substantially by only a subset of model inputs (Cullen and Frey, 1999). For meteorological inputs and dispersion models, a suggestion is used that was made by John Irwin of the EPA during collaborations on the current project: two components of the total uncertainty are defined: 1) the hour-to-hour variations at a single site, and 2) the variations in annual averages between sites. These two components are considered to be independent.

Data from many field experiments reported by Draxler (1984) can be used to estimate uncertainties in Gaussian plume model variables such as σ_y and σ_z. It appears that both σ_y and σ_z have a 95% uncertainty range of about plus and minus a factor of two (Hanna, 2002). Furthermore, minimum and maximum limits can be determined for the distributions of inputs so that they cannot be selected outside of a known physical range. For example, a σ_y less than 10 m is unlikely to occur at a downwind distance of 1000 m. The data on σ_y and σ_z suggest that their range is plus and minus a factor of five (Draxler, 1984), and there is no evidence to justify assuming a correlation between the two. Wind speed should also be constrained to be within known physical bounds when selected by a random number generator. Most meteorological inputs are limited to be within $\pm 5\sigma$ of the median. Some variables, such as cloud cover (CC), are bounded at both the low and high ends. A normal distribution is assumed for the uncertainty in CC, but CC cannot be less than 0.0 or greater than 1.0.

As for the emissions inputs, in all cases, the uncertainty is represented by a distribution function (normal or log-normal) and a definition of the 95 % range (i.e., plus and minus two standard deviations). None of the inputs is assumed to have a mean bias. The normal or log-normal functions are defined as:

Normal:

$$p(x) = ((2\pi)^{1/2}\sigma_x)^{-1}\exp(-(x-\mu)^2/(2\sigma_x^2)) \quad (1)$$

Log-Normal:

$$p(\ln x) = ((2\pi)^{1/2}\sigma_{\ln x})^{-1}\exp(-((\ln x-\mu_{\ln x})^2/(2\sigma_{\ln x}^2)) \quad (2)$$

where p is the probability density function (pdf), x is the variable of interest, μ is the mean of x, $\mu_{\ln x}$ is the mean of lnx, σ_x is the standard deviation of x, and $\sigma_{\ln x}$ is the standard deviation of lnx. The cumulative distribution function, $CDF_1(x_1)$ or $CDF_1(\ln x_1)$ can be defined as the integral of p(x) or p(lnx) from minus infinity to x_1 or $\ln x_1$. CDF_1 ranges from 0.0 to 1.0. The log-normal distribution is consistent with the statement that "the 95 % range of variable x is plus and minus a factor of Y". The log-normal distribution is a good choice for many air quality and meteorological variables which have uncertainties with magnitudes larger than a factor of ± 50 %.

It is assumed that, for the log-normally distributed variables, such as wind speed, the relative (normalized by the median) perturbation expressing the total uncertainty is the product of the relative perturbations for the local hour-to-hour and site-to-site (annual average) uncertainties. The local uncertainties are randomly selected each hour, and the site-to-site uncertainties are randomly selected to apply for the entire year. For each hour, the total relative perturbation in a given log-normally distributed input variable or parameter is given by the product of the two relative perturbations. In the case of a normally-distributed variable, such as wind

direction, the total relative perturbation is the sum of the two relative perturbations.

The specific recommendations for hour-to-hour and for site-to-site perturbations are given below. In most cases, the two types of perturbations are assumed to be nearly equal, which can be refined as additional data are analyzed in future studies. In all cases, the distribution is assumed to be bounded at ± five σ. That is, if a random number is drawn that is outside this bound, then draw again. Note that, for a Gaussian distribution, the 95 % range is equivalent to ± two σ.

Wind Speed (log-normal) - Wind speed is assumed to have a ± 30% uncertainty (covering the 95% range) for each of the hourly and site-to site components. According to NWS procedures, it is assumed that any wind speed less than 2.5 knots is listed as "calm". A procedure was devised so that the fraction of calms before and after the MC perturbations were applied would remain the same, on average. If the final randomly-selected wind speed is below the threshold value of 2.5 kt it is set to calm or zero wind speed, as required by the ISCST3 and AERMOD methodologies.

Wind Direction (normal) - The wind direction (degrees, ^{o}) is assumed to have a ± 30^o uncertainty (covering the 95% range) for each of the hourly and site-to site components. The sum of the original wind direction and the hourly perturbation plus the site-to-site Selected wind direction values greater than 360^o and less than 0^o were corrected to be in the range of 0^o to 360^o (e.g., -15^o is corrected to 345^o).

Cloud Cover (normal) – Cloud cover, ranging from 0 (clear) to 1.0 (overcast), is assumed to have an uncertainty of ± 0.1 (covering the 95% range) for each of the hourly and site-to site components. If the randomly-selected cloud cover is greater than 1.0, it is reset to 1.0. If it is less than 0.0, it is reset to 0.0.

Mixing height (log-normal) (ISC and AERMOD for some conditions) – A log-normal distribution with a 95 % range of ± 20 % (σ_{lnx} of 0.091) is assumed for both hour-to-hour and for site-to-site perturbations.

Surface Roughness (log-normal) (AERMOD only) - The surface roughness length, z_o, is assumed to have a log-normal distribution with a 95 % range of a factor of plus and minus a factor of 3 (σ_{lnzo} of 0.55) for site-to-site perturbations. No hour-to-hour perturbations are assumed. Note that, since ISC does not directly input z_o, and instead assumes either urban or rural surface conditions, those uncertainties are being handled in the current study by running ISC in two modes – 1) assuming all sources are surrounded by urban terrain, and 2) assuming sources are surrounded by either rural or urban terrain, following EPA (1995) criteria.

Bowen Ratio (log-normal) (AERMOD only) – The Bowen Ratio (BR) is the ratio of the sensible heat flux to the latent heat flux at the ground surface. Assume a log-normal distribution with a 95 % range of plus and minus a factor of 2 (σ_{lnx} of 0.347) for both hour-to-hour and site-to-site perturbations. Note that cloud cover, wind speed, BR, and surface roughness are used in AERMOD to estimate the Monin-Obukhov length, L.

dT/dz (log-normal) (ISC and AERMOD) – Vertical temperature gradient is assumed to have a log-normal distribution with a 95 % range of plus and minus a factor of 2 (σ_{lnx} of 0.347) for both hour-to-hour and for site-to-site perturbations. Note that variations in dT/dz are expected to primarily affect the plume rise calculations. The perturbations to dT/dz are applied inside the code after the temperature gradients have been calculated by the internal modules.

σ_y and σ_z (log-normal) (ISC and AERMOD) – The perturbations to σ_y and σ_z are applied inside the code after the dispersion parameters have been calculated by the internal modules. Assume a log-normal distribution with a 95 % range of ± 50 % (σ_{lnx} of 0.203) for both hour-to-hour and also for site-to-site perturbations. No correlation is assumed between σ_y and σ_z (i.e., they are varied randomly and independently).

Processing of Meteorological Parameters

AERMOD – The meteorological preprocessor, AERMET, was applied along with special pre-processing programs to generate the 100 meteorological data files and profile data used as input to AERMOD, which included both site-to-site and hour to hour perturbations. After this step, further modifications to the 100 surface

meteorological files were made to account for perturbations of σ_y and σ_z, and dT/dz that were read by the modified AERMOD program. The result was 100 surface and profile data files perturbed for wind speed, wind direction, cloud cover, Bowen ratio, surface roughness, σ_y and σ_z, and dT/dz.

ISCST3 - MPRM was applied along with special pre-processing programs to generate "onsite" meteorological data files for ISCST3. Added to the ISCST3 meteorological files were hourly perturbations for σ_y and σ_z, and dT/dz.

4. MONTE CARLO SAMPLING AND ANALYSIS METHODS

4.1 *Monte Carlo Sampling Methods*

The Monte Carlo (MC) sampling procedure is straightforward. With simple random sampling from the input distributions with no assumed correlations among input fluctuations, the number of needed MC runs is manageable since it is not dependent on the number of variables. The number of MC runs used in this study is 100, which is a reasonable compromise between the desire to have more runs to narrow the confidence bounds in the results, and the desire to have less runs to save computer time.

For the emissions uncertainties, a given perturbation or random number applies for the entire year of ISCST3 or AERMOD runs. For the meteorological and dispersion model uncertainties, a site-to-site variability (averaged over a year) and an hour-to-hour variability are assumed. Separate model runs were carried out for benzene and 1,3-butadiene.

For the site-to-site component, as for the emissions, a random number is selected that applies for the entire year. Before the set of m = 100 MC runs is carried out for each of the three combinations of emissions and transport and dispersion models, a set of "n" random numbers is generated (one for each of the n input variables or parameters). In order to make the model-to-model comparisons more meaningful in the subsequent analysis, the same sets of n random numbers for m = 100 MC runs have been used for the runs for each model. For the purpose of subsequent calculations of correlations between fluctuations in inputs and fluctuations in model outputs, the three sets of MC inputs for the three model options have each been saved in a large n x m matrix. Since the correlation is calculated using the rank of the variable rather than its exact value, the final statistical analysis does not recall the value of the physical variable (e.g., u component of wind speed in m/s) selected for the MC run, since the value of the initial random number that was selected serve for the correlation calculation. The rank correlation method is chosen as a basis for analysis because it is not as influenced by extreme values as the standard correlation coefficient calculated using the actual magnitudes of the variables.

For the hour-to-hour input variations, the random number selection procedures are similar to those for the site-to-site, except that each new dispersion model run has a new set of random numbers for each input for each of the 8760 hours of the year. These random numbers have also been saved for use in correlation calculations, although, as expected, the correlations are minimal for the hour-to-hour variations. This is because, over a year, the plus and minus deviations during each hour will tend to cancel out. The same sets of random numbers have been used for the same MC run for each dispersion model.

4.2 *Description of Outputs and Analysis Techniques*

For ISCST3, two sets of modeled concentrations were produced for analysis: 1) all sources as urban (denoted as ISCST3-urban), and 2) mixed urban and rural sources (denoted as ISCST3). For each pollutant, a total of 300 ISCST3 runs were made (100 runs for all sources assuming urban terrain, 100 runs for those sources surrounded by rural terrain, and 100 runs for those sources surrounded by urban terrain). The "mixed urban and rural" ISCST3 predicted concentrations at any location were then assumed to be the sum of the predicted concentrations for the "urban sources" and "rural sources" runs just mentioned. Note that the "urban" and "rural" designations are prescribed by procedures outlined by the EPA (1995) and are the same us used in the EPA (2002) studies of Houston. For AERMOD, one set of 100 MC runs was carried out for each pollutant, automatically representing mixed rural and urban sources through its input of a spatially-varying surface roughness length. Consequently, a total of 800 runs was carried out for the two pollutants and the various model combinations.

The outputs are analyzed to determine the characteristics of the total variability, and to determine the inputs whose variations have the most effect on the variations in the outputs. The specific outputs to be analyzed partly depend on the set of relevant questions being asked of the study. As explained above, the primary outputs are 1) the annual average concentration averaged over all 46 receptors, and 2) the maximum annual average concentration at any of the 46 receptors. The 46 receptors consist of 43 census tract population centroids, plus three routine monitoring sites.

Total Variability - As mentioned earlier, receptor locations are specified at 43 census tract population centroids and at three monitoring locations shown in Figure 2. Individual annual-averaged concentrations at each of these 46 locations for each MC run are archived. In addition, the following outputs are estimated for each MC run: 1) Spatial-averaged annual-average concentrations over the 43 population-based receptors, and 2) Peak annual-average concentration over all 46 receptors and receptor location.

The 100 sets of MC outputs averaged over the entire domain and the maximum at any receptor across the domain are rank-ordered and used to define 95 % confidence intervals (the "total variability") and maxima and minima. The shape of the output pdf is studied to determine if it is lognormal. The 100 locations of the maximum concentration are studied to determine possible reasons for those maxima. The results are studied to see if there is much difference from one pollutant to another (benzene or 1,3-butadiene), and whether the emissions inputs or the meteorology/dispersion model inputs contribute the most to the total uncertainty.

Scatter Plots, Correlation Coefficients, and Multiple Linear Regression - After the MC outputs are created, a variety of statistical analysis techniques are applied to identify key contributors to output uncertainty. These include scatter plots of inputs versus outputs, calculations of correlation coefficients, and application of multiple linear regression analysis (Cullen and Frey 1999; Hanna et al. 1998, 2001, 2005).

In most probabilistic MC assessments, the majority of the uncertainty in the output distribution (annual averaged concentrations in this scenario) is attributable to uncertainty in a small subset of the inputs (Cullen and Frey, 1999). An identification of this subset of highly significant contributors to output uncertainty can help guide future research. The rank correlation coefficient is used to identify important inputs. The analysis also allows the relative contributions of the uncertainties in the emissions model and the transport and dispersion model to the uncertainties in the output concentrations to be assessed. For m = 100 MC runs, the correlation can be said to be significantly different from 0.0, with 95 % confidence, if the magnitude of the correlation coefficient exceeds about $2/m^{1/2}$, or 0.2.

The five to ten parameters with the largest correlation coefficients (significant at the 95 % confidence level) are included in a multiple linear regression analysis. The output is a linear regression equation where the perturbations in predicted concentrations are expressed as linear combinations of the perturbations in input variables or model parameters. The fraction of the variance explained by each parameter is also estimated.

5. RESULTS OF ANALYSIS OF MONTE CARLO RUNS

This section includes examples of the Monte Carlo (MC) results for the three models (AERMOD, ISCST3 (mixed urban/rural), and ISCST3-all urban), for the maximum (peak) annual-averaged concentrations at any of the 46 receptors (43 census tract population centroids and 3 monitors), and for the spatial-averaged annual average concentrations over the 43 census tract population centroid receptors. The locations of these receptors were shown in Figure 2. The project report (Hanna et al, 2005) contains full sets of tables and figures.

5.1 *Variations in Locations of Maximum (Peak) Annual-Averaged Concentrations*

The location of the maximum (peak) annual-averaged concentration among the 46 receptors (43 population centroids and 3 monitors) was determined for each of the 100 MC runs for each model and pollutant. Given the variability of emission sources categories and other model inputs, it is expected that this maximum location will shift. Figures were prepared showing these locations, with legends indicating the number of

runs that each receptor was the highest for the three models (AERMOD, ISC, and ISC-urban).

For benzene, the frequency of occurrence is more or less split between two groups of receptors – the first group is the single monitor in the northern part of the domain next to a major highway (I-10), where mobile sources dominate the emissions; the second group consists of three census-tract population-centroid receptors in the western part of the domain within the industrial region along the ship channel, where chemical processing plants and oil refineries dominate the emissions. There does seem to be a small difference between the frequencies of locations for AERMOD and ISC-urban compared with ISC – that is, there are slightly more AERMOD and ISC-urban maxima near the ship channel than next to I-10, while the reverse is true for ISC.

For 1,3-butadiene, over 90 % of the maxima for all models occur at the monitor adjacent to the busy highway I-10. Very few maxima occur in the industrial ship channel area, despite the fact that most of the 1,3-butadiene emissions are from industrial sources. This apparent inconsistency can be explained as a combination of several factors, such as the very close proximity of the receptor/monitor to I-10, and the tendency for the industrial sources to be associated with stacks and/or plume rise.

5.2 *Probability Density Functions of Monte Carlo Concentrations*

For each set of 100 MC runs (for the three models and the two pollutants), the 100 values of predicted concentrations have been ranked from low to high, and probability density functions (pdfs) or frequency distributions calculated. Figure 3 is an example (for benzene from AERMOD runs) of histograms showing the frequency distributions of the maximum (peak) annual-averaged concentrations found at a single receptor among all the receptors (top part of figure), and the annual averaged concentration averaged over the 43 census tract population centroids (bottom part of figure). The frequency distributions all have longer "tails" at the high-concentration end of the distribution, which is characteristic of a log-normal distribution. Even though the distributions of the 100 MC outputs in Figure 3 appear somewhat ragged (i.e., not smooth), the distributions are likely to be satisfactorily capturing the main output of interest, namely the 95 % range of the MC outputs. The references, such as IAEA (1988), NCRP (1996) and Cullen and Frey (1999), agree that 100 MC runs are sufficient to capture the primary aspects of the spread of the distribution. If more MC runs were carried out (say 500 or 1000), the distributions would smooth out, but there would be only minor improvement in estimation of the total spread.

The distributions are presented in a somewhat different way in Figure 4 and Table 3. Here, each set of 100 ranked MC outputs is used to estimate certain values and percentiles (min, 2.5, 5, 10, 25, 50, 75, 90, 95, 97.5, and max) of the Cumulative Distribution Function (CDF). Figure 4 contains plots of these various percentile levels for the CDFs of the 100 MC outputs of maximum (peak) concentrations (top), and concentrations averaged over all census tract population centroid receptors (bottom), for benzene (part a) and 1,3-butadiene (part b). The quantitative concentrations at the plotted points are listed in Table 3. Our discussion focuses on the range from the 2.5^{th} and the 97.5^{th} percentiles, since that defines the 95 % range, which is of interest in many statistical analyses.

Figure 4 suggests that the relative ranges (i.e., the 95 % range divided by the median or the 50^{th} percentile) are fairly constant from one model to the other, from one pollutant to the other, and from the max (peak) C to the spatial-averaged C. The 95 % uncertainty range appears to be "about plus or minus a factor of two or three". The average range in the table is found to be ± factor of 2.41 for the maximum (peak) and ± factor of 2.03 for the census tract population centroid average. To arrive at this general result in Table 4, we calculate and list the ratio $50^{th}/2.5^{th}$ (C(50)/C(2.5)) and the ratio $97.5^{th}/50^{th}$ (C(97.5)/C(50)).

Note that the 50^{th} percentile (median) predictions of ISCST3(mixed urban/rural mode) are about 25 % higher than the corresponding predictions for ISCST3 (urban) for benzene and about a factor of 1.5 to 2 higher for 1,3-butadiene. The difference between the "mixed rural/urban" and "all urban" predictions is because there is more turbulence and hence dispersion for urban conditions. These differences between the two ISC modes reflect a difference in assumed surface roughness conditions.

Table 3 contains calculations of the geometric mean (GM) of the $50^{th}/2.5^{th}$ and $97.5^{th}/50^{th}$ ratios, in order to investigate whether there is a trend

with model or pollutant. GM is used to define the "± factor of GM" covering the 95 % range. It is concluded that there is usually less than a 10 % difference between the relative 95 % uncertainties (normalized by the median) for benzene and for 1,3-butadiene. It is also concluded that the 95 % uncertainty range is about 20 % larger (± factor of 2.41 versus ± factor of 2.03) for the maximum (peak) concentration over the domain compared to the spatially-averaged centroid concentration over the domain.

Generally, the 95 % range is about 25 % larger for AERMOD than for ISCST3 or for ISCST3-Urban. This 25 % difference can be postulated to be due to the fact that the AERMOD MC runs include uncertainties in surface roughness, which is shown in later sections to have a significant correlation with variations in AERMOD-predicted concentrations, while the individual ISCST3 MC runs do not once the terrain surface is selected. If the difference between the predictions of ISCST3 (mixed urban/rural) and ISCST3 (urban) is considered to be caused by the effects of uncertainties in surface roughness, then this accounts for the 25 % difference in range.

5.3 *Analysis of Monte Carlo Outputs to Calculate Correlations and Apply Multiple Regression Methods*

The current section presents correlations between variations in the individual inputs and variations in the predicted concentration outputs. In addition, multiple linear regression methods are applied to the several inputs whose variations are significantly correlated with the variations in the outputs.

"Rank" correlation coefficients were calculated. This helps reduce the influence of outliers that would occur if actual magnitudes were used. Correlation coefficients that are significantly different from 0.0 at the 95% confidence level can be estimated. The magnitude of the correlation coefficient that defines the 95 % confidence level for 100 pairs of data is 0.20 (i.e., $2/(100)^{1/2}$).

For benzene, the variables with significant correlations include: source categories 1 (motor vehicles), 3 (refineries), 10 (landfills), and 11 (sewage treatment plants), wind speed, surface roughness (in AERMOD) and σ_z (vertical dispersion parameter). The inputs with relatively large correlations (magnitudes exceeding 0.4) are:

Source Category 1 (motor vehicles) has a correlation exceeding 0.6 for benzene for all models and for the maximum (MAX) and the spatial-average over the census tract population centroid receptors (AVG).

Source Category 3 (refineries) has a magnitude exceeding 0.4 for benzene from AERMOD runs.

Surface roughness has a correlation of -0.55 for benzene from AERMOD runs for the spatial average of the centroid receptors.

σ_z has a correlation of -0.49 and -0.43 for benzene for the spatial average of the centroid receptors from ISC runs and from ISC-Urban runs.

The large correlations for benzene for motor vehicles and for refineries are consistent with the locations of the maximum concentrations for the 100 MC runs as discussed in Section 5.1. The locations of the maxima were roughly equally split between the site next to the busy highway I-10, and a few sites in the industrial area near the Houston Ship Channel.

For 1,3-butadiene, the variables with significant correlations include source categories 1 (industrial), 2 (rubber and latex production), 7 (cooling towers), and 10 (road segments), as well as wind speed, surface roughness, and σ_z. By far the largest correlations (about 0.9 for the maximum concentration) are for category 10 (road segments). This result may be related to the fact that nearly all of the maximum 1,3-butadiene concentrations occur at the monitor near the busy highway I-10. The meteorological and dispersion model variables with significant correlations are the same as for benzene.

The multiple linear regression coefficients were estimated for benzene and for 1,3-butadiene, starting with the variables with significant correlations. A variable was included if it was significant for at least one of the models. Table 4 lists the multipliers of the regression equation, the estimated intercept (i.e., the concentration for the situation where all perturbations are zero), and the final correlation coefficient of the regression equation. Benzene is in part a and 1,3-butadiene is in part b of the table.

The regression equation assumes that X_i represents the normalized MC variation for independent parameter, i, and that Y represents the difference between the predicted concentration and the concentration that would occur if there

were no variations in any of the parameters (the intercept). Concentrations are given in $\mu g/m^3$. For example, the following equation is inferred from the regression coefficients in Table 4a for benzene from AERMOD runs for the maximum concentration:

$$Y_{AER\_MAX} = C(AERMOD) - 7.94\ \mu g/m^3 = 0.573X_{cat1} + 0.379X_{cat3} + 0.248X_{cat11} - 0.204\ X_{WS} - 0.269\ X_{ROUGH} - 0.115\ X_{SIGMAZ} \quad (3)$$

As an example of the use of this equation, assume that the normalized perturbations happen to have values of 1.5 for emission category 1, 0.9 for emission category 3, 1.1 for emission category 11, 0.95 for wind speed (WS), 1.3 for surface roughness (ROUGH), and 1.0 for sigma z. With these assumptions, the above regression equation indicates that the resultant maximum concentration predicted by AERMOD would increase by 0.815 $\mu g/m^3$. Since the intercept concentration is 7.94 $\mu g/m^3$, then the final estimated concentration would be 8.76 $\mu g/m^3$, or an increase of about 10 %. Of course, it must be recognized that these are not "real" measurement data used to develop the regression relations, but are model predictions from the 100 sets of MC perturbations of the assumed input variables.

The magnitudes of the regression coefficients listed in Table 4 more or less parallel the individual correlation coefficients. We note that, whereas some of the variables with relatively large coefficients make intuitive sense (e.g., CAT1 is mobile source emissions, which have the largest fraction of the total emissions), others are less so (e.g., CAT10 is landfills, which have a much smaller fraction). The reasons for some of these higher correlations are not obvious and more study is needed.

The relative influence on the total uncertainties of the emissions category group versus the meteorological input group can be estimated by comparing the sums of the squares of the regression coefficients in Table 4 for these groups. The numbers listed in Table 5 are the resulting fraction of total explained variance due to the emissions category group. It is seen that, in all but one case (AER_AVG for 1,3-butadiene), the emissions category group has the largest fraction. For the maximum or peak concentration receptor (MAX in the tables), the emissions category group (specifically the mobile source category), dominates the explained variance, primarily because the maximum occurs almost exclusively at the monitor adjacent to highway I-10. For the spatially-averaged census tract population centroids (AVG in the tables), there is more of a balance between the contributions of the emissions and the meteorology groups. For the AVG data, AERMOD has a larger fraction of explained variance than ISC in the meteorology group primarily because AERMOD accounts for variations in surface roughness, while the two individual ISC options do not. However, as mentioned earlier, if we consider the differences between ISCST3-Urban and ISCST3-Mixed to be due to differences in terrain roughness (urban vs rural), then the surface roughness could be considered to have more of an effect on the ISC uncertainty than indicated in the table.

6. CONCLUSIONS, INCLUDING ANSWERS TO SPECIFIC QUESTIONS

Four key question sets were asked in Section 1, and are repeated below, followed by the answers, which are based on the results of the MC analysis.

Question Set 1 (uncertainty of modeling system): What is the total uncertainty in the annual average concentration (averaged over the 43 census tract population-centroid receptors) of benzene and 1,3-butadiene concentrations in the Houston region, and which input variables and model parameters have the most influence on this total uncertainty? What is the total uncertainty in the magnitudes and the locations of the 100 maximum annual-averaged concentrations calculated by each MC run at the 43 population centroid receptors and the three monitors?

Answer 1: First it is noted that the census-tract population centroid receptors studied here were chosen to be identical to those used in the EPA (2002) study. The Monte Carlo analysis shows that the total uncertainty (defined as the 95 % range) in the averaged-annual concentration for the spatially-averaged census-tract population centroid receptors is about ± a factor of 2.0, with only slight variation with model or with pollutant. When the maximum annual-average concentration at any receptor location is analyzed, the uncertainty increases by about 20 % to about ± a factor of 2.4. In general the uncertainties for AERMOD are about 20 to 30 % greater than the uncertainties for either ISCST3-Urban or ISCST3-

Mixed Urban/Rural, largely due to AERMOD directly incorporating surface roughness while ISCST3 requires a pre-selection of one of two classes (urban or rural) of surface roughness..

The uncertainties in benzene and 1,3-butadiene concentration are found to be most strongly related to variations in motor vehicle (mobile source) emissions. Some industrial emission categories are also found to be significantly correlated with uncertainties in concentrations. The location that shows the strong correlations with motor vehicle emissions is the monitor receptor sited adjacent to a busy highway (I-10). The uncertainties in benzene and 1,3-butadiene concentrations are also significantly related to variations in three meteorological and dispersion model parameters – wind speed, surface roughness (for AERMOD), and σ_z (vertical dispersion). However, surface roughness (implicitly parameterized by assigning either "urban" or "rural" land use to the area around a given source) is also important for ISCST3, since predicted concentrations for the ISCST3-Mixed Urban/Rural option are about 20 % to 100 % greater than for the ISCST3-Urban option.

Question 2 (contributions of emissions vs dispersion models to uncertainty): What is the relative uncertainty between the emissions and the transport and dispersion model?

Answer 2: Generally there were about four to six of the emissions categories and three of the transport and dispersion model categories whose uncertainties had significant (at the 95 % confidence level) correlations with uncertainties in predicted benzene and 1,3-butadiene concentrations. The relative influence on the total uncertainties of the emissions category group versus the meteorological input group was estimated by comparing the sums of the squares of the regression coefficients the two groups. It is seen that, in all but one case (AERMOD averaged over the census tract centroids for 1,3-butadiene), the emissions category group has the largest fraction. For the maximum concentration receptor, the emissions category group (specifically the mobile source category), dominates the explained variance, primarily because this occurs at the monitor adjacent to highway I-10. For the spatially-averaged census tract population centroids, there is more of a balance between the contributions of the emissions and the meteorology groups.

Question 3 (contributions of emissions source classes): How do the total uncertainties and correlations differ for different source classes, such as mobile versus point, or industrial major point source versus industrial area and volume? And how would these differences impact conclusions regarding source apportionment?

Answer 3: The differences in total uncertainty by source class are difficult to derive from the results, since all inputs were varied simultaneously and independently. However, it was possible to study the correlations for the various source classes (21 for benzene and 13 for 1,3-butadiene). It was found that uncertainties in mobile sources were the major contributor to uncertainties in predicted concentrations of both benzene and 1,3-butadiene. The second-largest contributor to the total uncertainty was refineries and the chemical industry. This specific result may be partly attributed to the fact that the receptor with the strong correlation with mobile source emissions was not a census-tract centroid, but was an ambient air monitor located adjacent to a busy highway (I-10). This study used the same census-tract population centroids as the EPA, as well as the three monitors on the domain. Also, it should be noted that the source categories used here closely parallel those from the EPA study (Frey and Zhou, 2003), but are not necessarily a one-to-one match with the source classes defined by EPA for regulatory assessments (e.g., MACT Area Source Standards and Mobile Source Air Toxics).

Question 4 (dependence on model): Do the conclusions concerning uncertainty depend on model used (e.g., ISCST3 versus AERMOD or ISCST3 with EPA rural-urban designations versus ISCST3 with all-urban designations)?

Answer 4: There are few differences among the models regarding conclusions concerning total uncertainty and the variables whose uncertainties have the strongest correlation with the modeled concentration uncertainties. The absolute concentrations predicted deterministically by the base case runs from the three models may vary by a factor of as much as two, but the relative uncertainties (normalized by the base run concentrations) show much less variation. The

largest variation by model in the MC results is that the total uncertainty for AERMOD is about 20 % larger than the total uncertainty for the two ISCST3 options. However, since AERMOD accounts for surface roughness and the individual ISCST3 options do not, the inclusion of uncertainties in surface roughness inputs in the MC uncertainty analysis for AERMOD would represent an additional contribution to the uncertainty. Since our results show that the choice of urban or rural terrain in ISCST3 can cause 20 % to 100 % differences in predicted concentrations, the differences between ISCST3-Urban and ISCST3-Mixed Urban/Rural can be implicitly considered to be a measure of the uncertainty due to surface roughness variations. Consequently, accounting for this difference in ISCST3 options, it can be concluded that there is very little difference between the total uncertainty of AERMOD and ISCST3.

ACKNOWLEDGEMENTS

This research has been sponsored by the American Petroleum Institute (API), with Howard Feldman and Richard Karp as the program managers over the course of the study. Prof. H.C. Frey of North Carolina State University was a consultant on the project and contributed many useful ideas and comments. The authors appreciate the suggestions and data sets provided by Joseph Touma, James Thurman, and John Irwin of the U.S. EPA.

REFERENCES

Cimorelli, A.J., S.G. Perry, A. Venkatram, J.C. Weil, R.J. Paine, R.J. Wilson, R.F. Lee, W.D. Peters, R.W. Brode, and J.O. Paumier, 2005: *AERMOD – Description of Model Formulation.* USEPA, RTP, NC 27711, 91 pages. http://www.epa.gov/scram001/7thconf/91aermod/aermod_mfd.pdf.

Cullen, A.C.; and H.C. Frey, 1999: *The Use of Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum: New York, 335 pages.

Draxler, R.R., 1984: Diffusion and Transport Experiments. Chapter 9 in *Atmospheric Science and Power Production* (D. Randerson, ed.), DOE/TIC-27601, U.S. Dept. of Energy. NTIS, Springfield, VA 22161, 367-422.

EPA, 1995: *User's Guide for the Industrial Source Complex (ISC3) Dispersion Model (Revised);* Volume II - Description of Model Algorithms. EPA-454/b-95-0036, USEPA, RTP, NC 27711.

EPA, 2000: *National Air Toxics Program: The Integrated Urban Strategy*. EPA-453/R-99-007. OAQPS/EPA, RTP, NC 27711.

EPA, 2002: *Example Application of Modeling Toxic Air Pollutants in Urban Area*. EPA-454/R-02-003, OAQPS/EPA, RTP, NC 27711, 91 pages, http://www.epa.gov/scram001/tt25htm#toxics.

Freeman, D.L., R.T. Egami, N.F. Robinson and J.G. Watson, 1986: A method for propagating measurement uncertainty through dispersion models. *J. Air Poll. Control Assoc*., **36**, 246-253.

Frey, H.C. and Y. Zhao, 2003: *Development of Probabilistic Emission Inventories of Benzene, Formaldehyde and Chromium for the Houston Domain*. Prepared by North Carolina State University for Carolina Environmental Program and U.S. EPA, Research Triangle Park, NC.

Hanna, S.R., 2002: *Meteorological Modeling in MACCS2*. Final Report P047, Hanna Consultants, 7 Crescent Avenue, Kennebunkport, ME 04046, Prepared for the U.S. Nuclear Regulatory Commission Two White Flint North, 11545 Rockville Pike, Rockville, MD 20852-2738.

Hanna, S.R., Z. Lu, H.C. Frey, N. Wheeler, J. Vukovich, S. Arumachalam and M. Fernau, 2001: Uncertainties in Predicted Ozone Concentration Due to Input Uncertainties for UAM-V Photochemical Grid Model Applied to the July 1995 OTAG Domain. *Atmos. Environ*., **35**, 891-903.

Hanna, S.R., R.J. Paine, D. Heinold, and E. Kintigh, 2005a: Uncertainties in Benzene and 1,3-Butadiene Emissions in Houston and their Effects on Uncertainties in Concentrations Calculated by AERMOD and ISC. Report P055 by Hanna Consultants, Kennebunkport, ME 04046, Prepared for the API, 1220 L Street NW, Washington, DC 20005.

Hanna, S.R., J. Wilkinson, A.G. Russell, J. Vukovich and A. Hansen, 2005b: Monte Carlo estimation of uncertainties in BEIS3 emission outputs and their effects on uncertainties in Chemical Transport Model predictions. *J. Geophys. Res.,* **110**, D01302, doi: 10.1029/2004JD004986.

Heinold, D., R. Paine, and H. Feldman, 2003: *Quantitative evaluation of the EPA urban air toxics modeling strategy: Results of sensitivity studies*. Paper number 69639, Proceedings of AWMA Annual Meeting,

IAEA, 1989: *Evaluating the Reliability of Predictions made using Environmental Transfer Models*. IAEA Safety Series No. 100, International Atomic Energy Agency, Vienna, Austria.

Irwin, J.S., S.T. Rao, W.B. Petersen and D.B. Turner, 1987: Relating error bounds for maximum concentration estimates to diffusion meteorology uncertainty. *Atmos. Environ*., **21**, 1927-1937.

NCRP, 1996: *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*. (ed., F.O. Hoffman), NRCP Commentary Number 14, Nat. Coun. on Radiation Prot. and Meas., 7910 Woodmont Ave., Bethesda, MD.

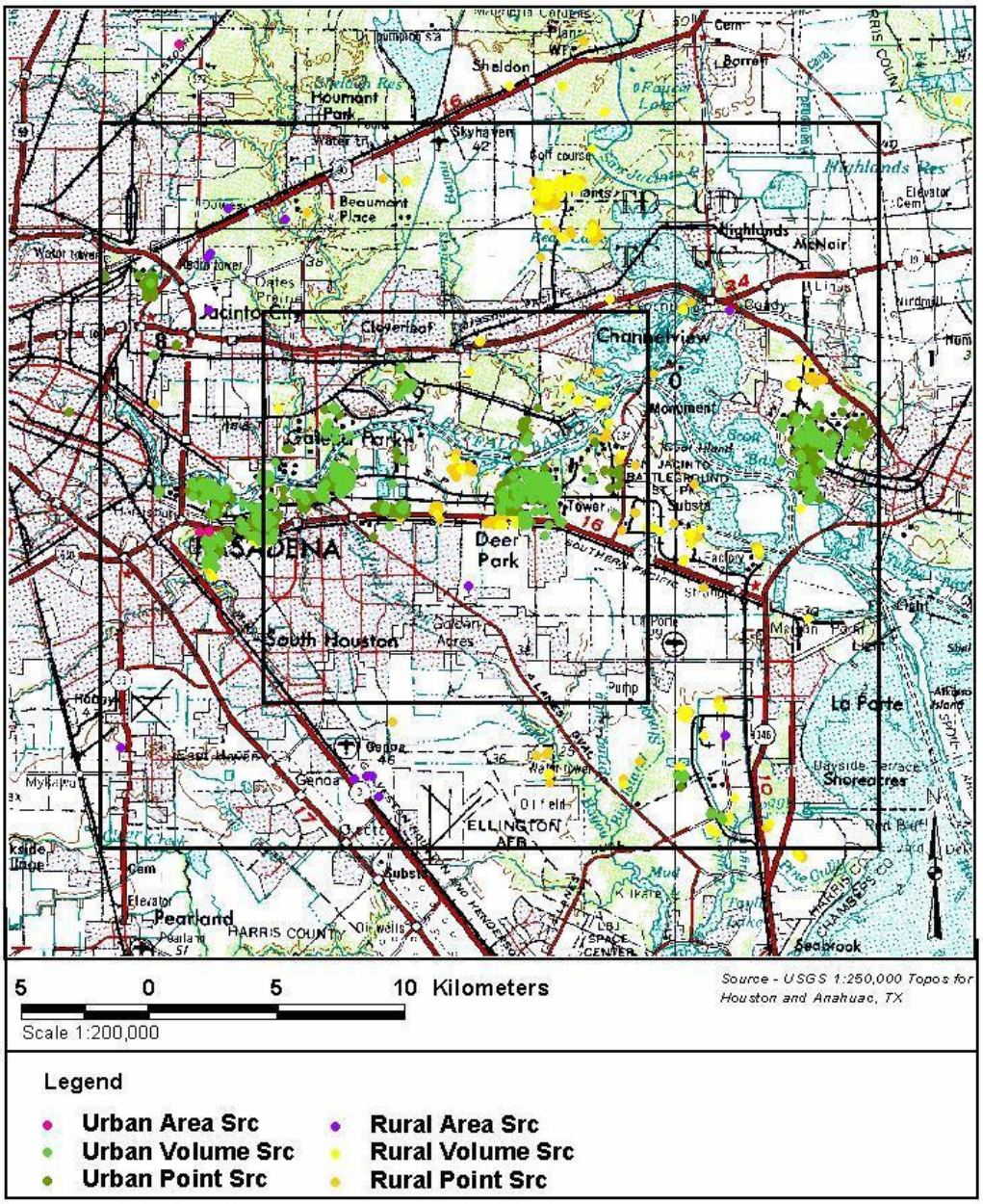


Figure 1. Map of the Houston domain used for the Monte Carlo study. The receptor domain is the 15 km x 15 km inner square and the source domain is the 30 km x 30 km outer square. The sources assumed to be in urban or rural terrain are indicated, and separate colors are used for area, volume and point source categories.

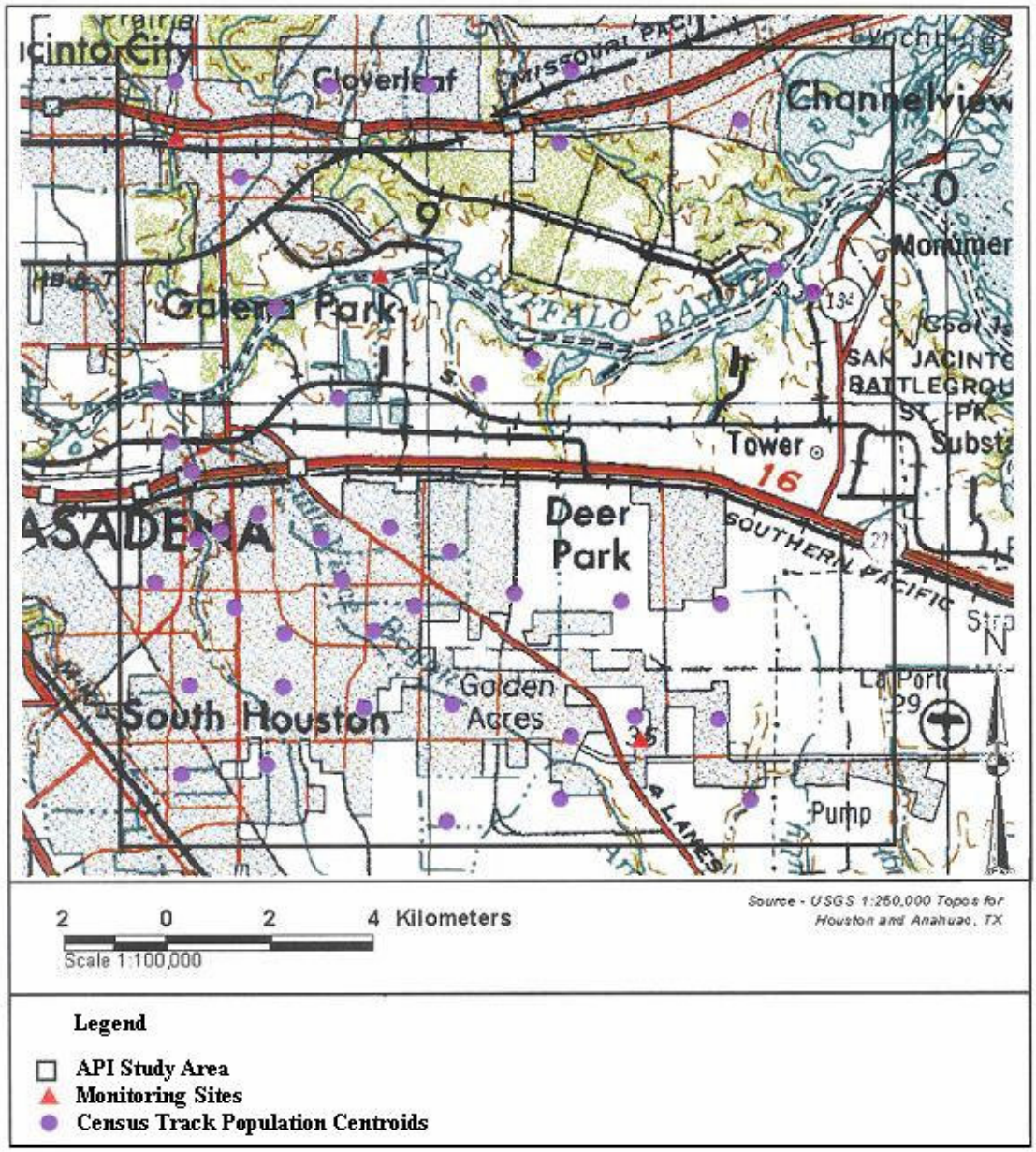
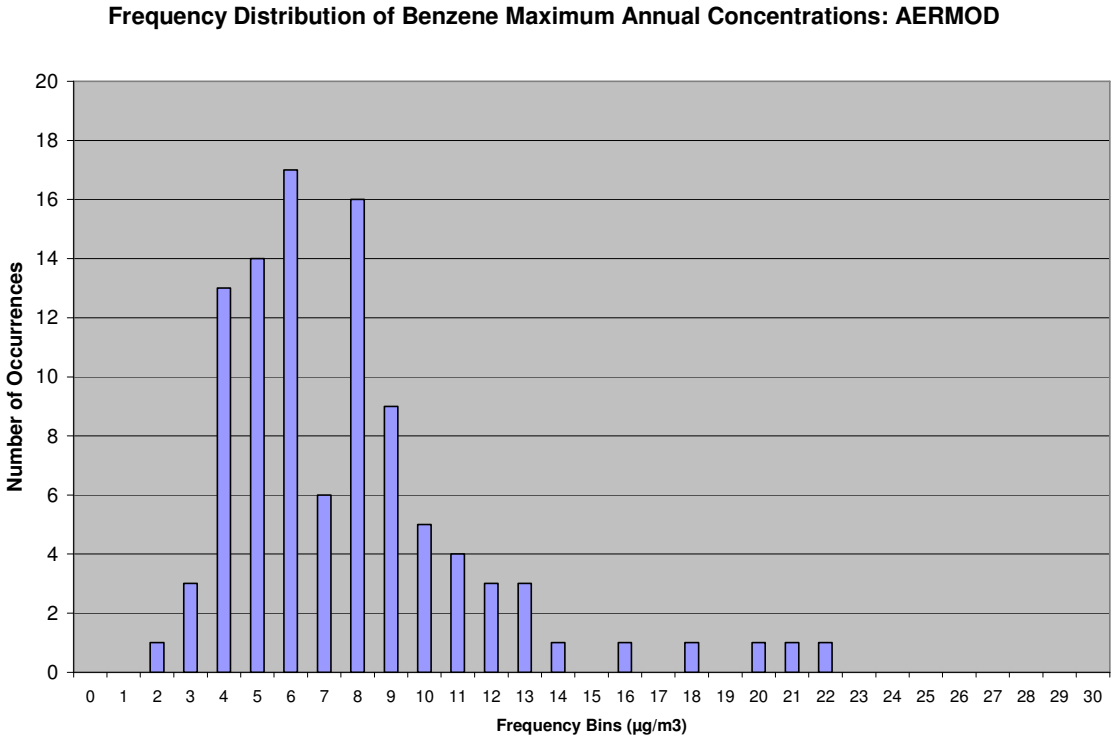


Figure 2 – Map of 15 km by 15 km receptor domain (the inner square in Figure 1), showing the 43 census tract population centroids (round dots) and the three monitoring sites (triangles).



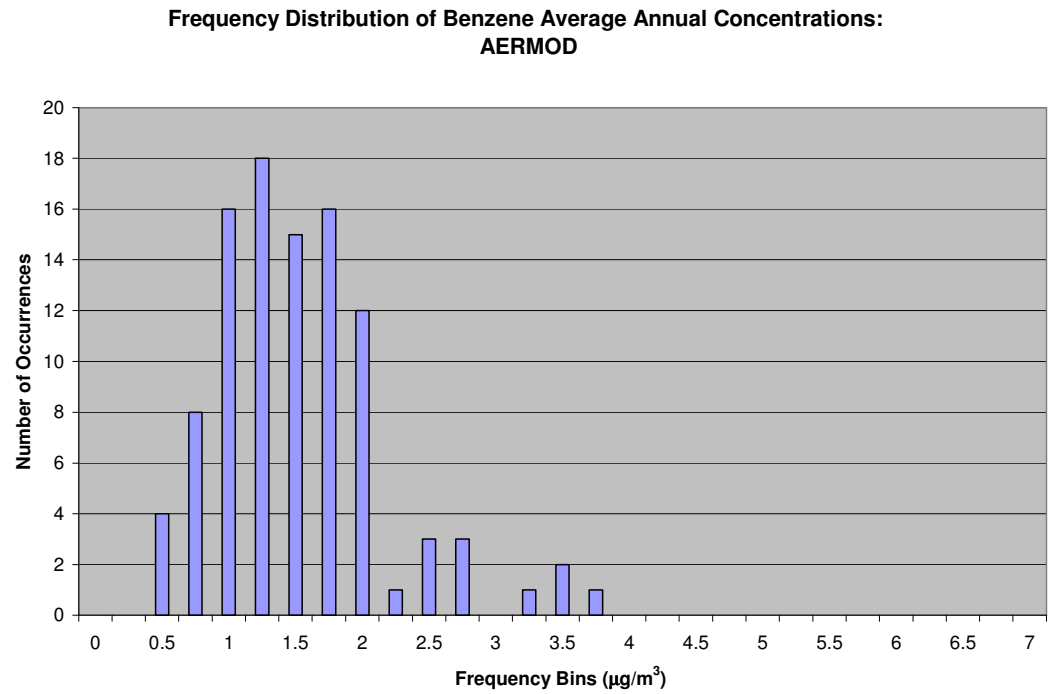


Figure 3. AERMOD benzene frequency distributions (occurrences per interval) of annual averaged concentrations for the 100 MC runs for the maximum or peak at any individual receptor anywhere on the domain (top panel) and for the spatial-average over all 43 census tract population centroids (bottom panel).

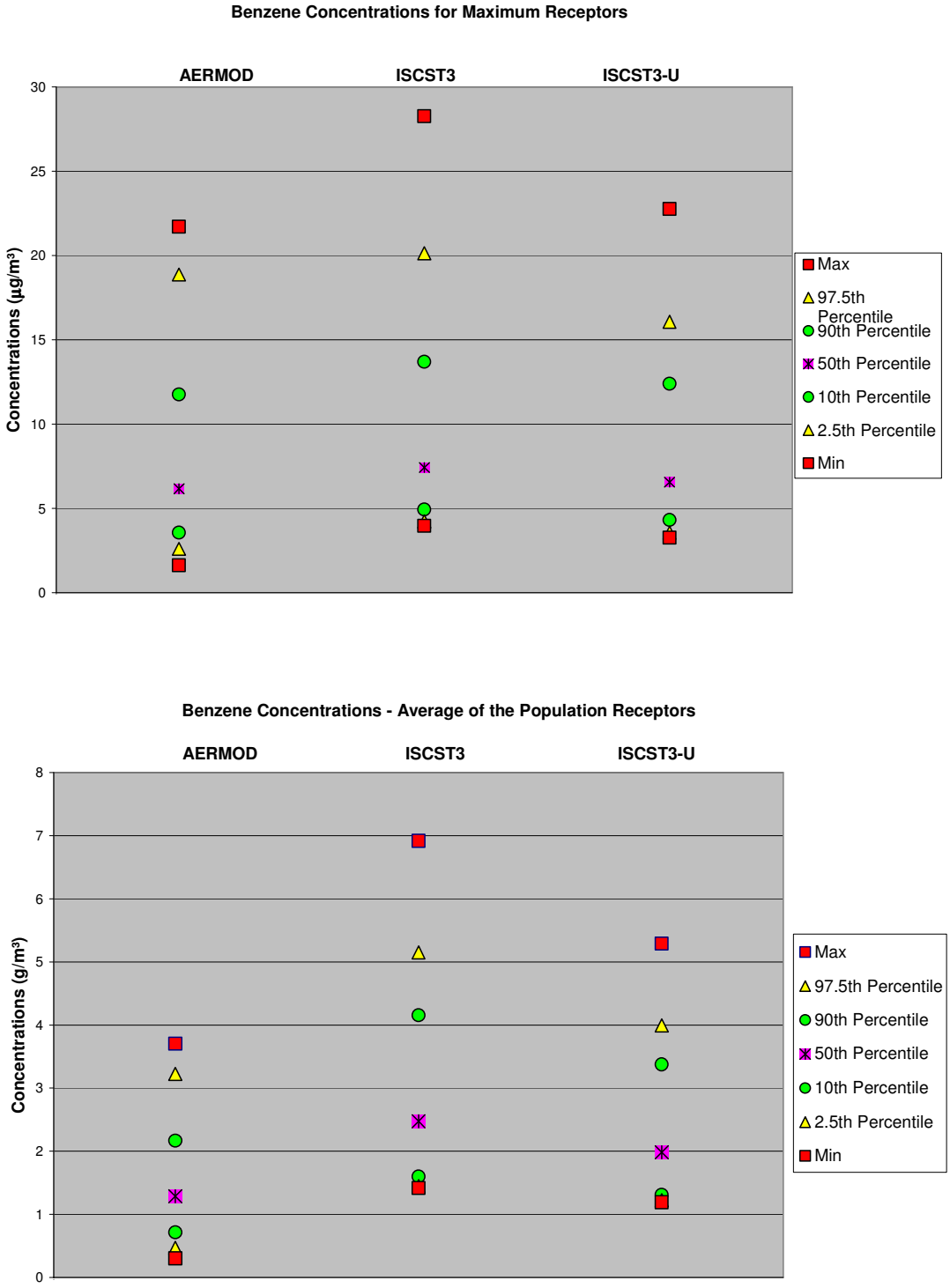


Figure 4a. Significant points on the CDF based on 100 MC runs for predicted annual-averaged benzene concentrations (µg/m^3) for the maximum or peak at any individual receptor anywhere on the domain (top panel) and for the spatial-average over all 43 census tract population centroids (bottom panel). Quantitative values for the points are listed in Table 3

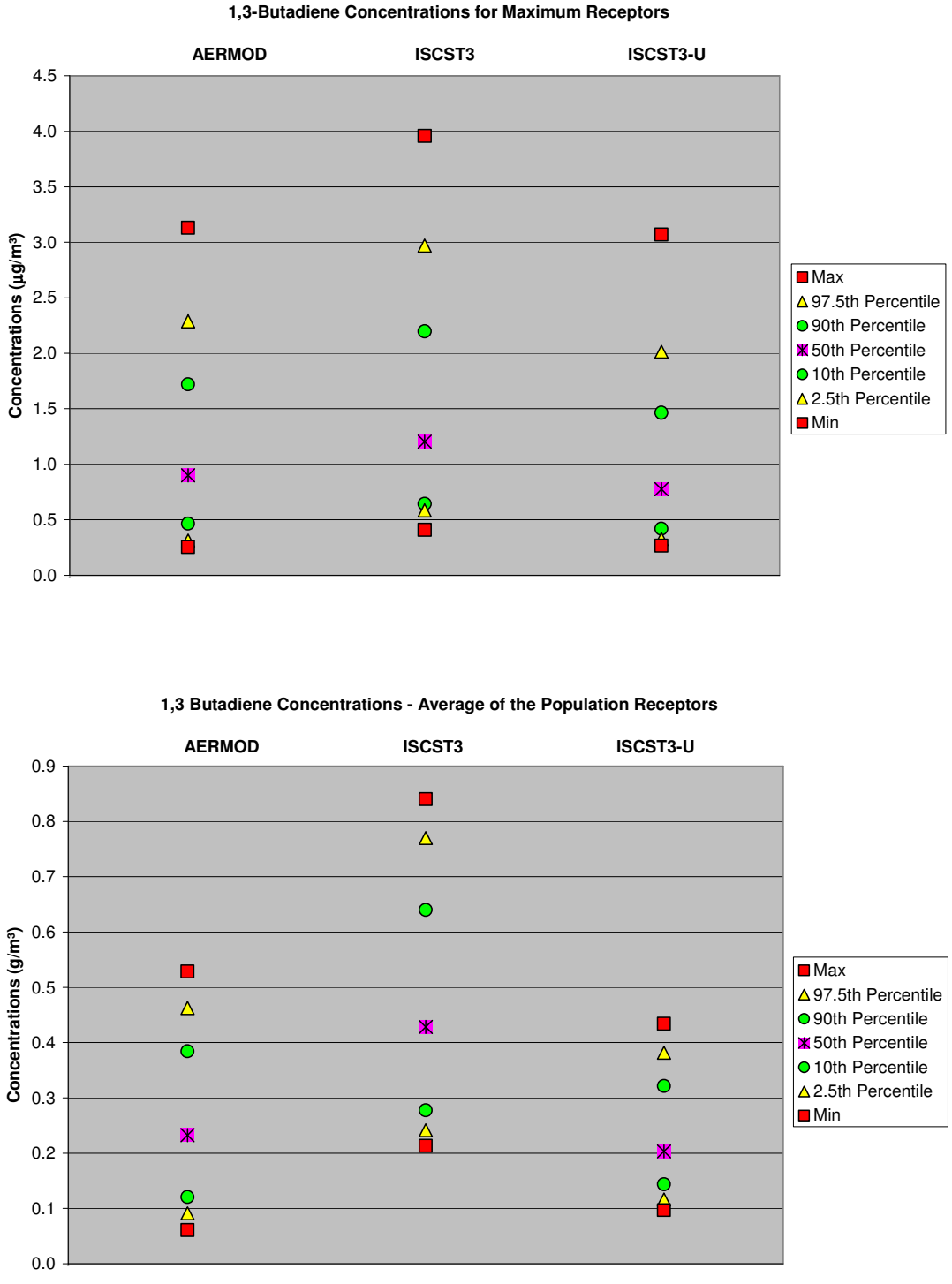


Figure 4b. Significant points on the CDF for 100 MC runs for predicted annual-averaged 1,3-butadiene concentrations (µg/m^3) for the maximum or peak at any individual receptor anywhere on the domain (top panel) and for the spatial-average over all 43 census tract population centroids (bottom panel). Quantitative values for the points are listed in Table 3

Table 1. Benzene Emission Categories in the Houston Source Domain. Initially, there were 24 Categories Recommended by the Houston Emissions Workshop (see Frey and Zhao, 2003). A Few Categories were Combined that were Similar or that had Very Small Emissions.

Category	Description	% of Total
1 (includes 1, 2, 9 and 18)	Light Duty Gas Vehicles (LDGV), Light Gas Trucks (LDGT), Road Segments, Heavy Duty Gas Vehicles (HDGV), Heavy Duty Diesel Vehicles (HDDV)	28.5
3	Petroleum refineries	24.7
4	Non-road 4-stroke gas engines, Internal Combustion Engines	8.7
5	Non-road 2-stroke gas engines	2.1
6	Non-road diesel (construction, farm, and industry)	1.6
7	Oil and gas production	0.6
8	Natural gas transmission and marine transport	3.8
10	Forest wildfires, Municipal landfills	0.4
11	Solid waste disposal (sewage treatment, aeration tanks)	3.5
12	Acetylene production (butylenes, ethylene, propylene, olefin)	2.9
13	Fuel oil external combustion, External combustion boilers	2.3
14	Typical ethylene plant	1.0
15	Gas service stations stages I and II	0.6
16	Petroleum industry fugitives	1.6
17	Managed burning, prescribed	0.04
19 (includes 19 and 24)	Chemical manufacturing; fugitive emissions and leaks, general processes	1.0
20	Aircraft	0.4
21	Petroleum industry; fugitive emissions; misc. Petroleum & solvent evap.	7.3
22	Process vents in refinery production	0.9
23	Loading, ballasting, transit losses from marine vessels	1.3
25	Industrial Processes	6.8

Table 2. 1,3-Butadiene Emissions Source Categories in the Houston Domain.

Category	Description	% of Total
1	Fuel oil external combustion, petroleum and solvent evaporation, organic solvent evaporation, fuel fired equipment, natural gas, flares, industrial processes, petroleum industry, process gas	40.1
2	Styrene-butadiene rubber and latex production, nitrile butadiene rubber production	15.6
3	Chemical manufacturing fugitive emissions, industrial processes, general processes, fabricated metal products fugitive emissions, plastics production	17.5
4	Industrial processes, chemical manufacturing, butadiene fugitive emissions	2.5
5	Ethylene plant, industrial processes chemical manufacturing butylenes. Ethylene propylene, olefin production fugitives emissions	3.9
6	Loading, ballasting, transit losses from marine vehicles	1.6
7	Industrial processes, petroleum industry cooling towers and fugitive emissions from flanges and all streams	2.0
8	Aircraft	0.8
9	Unknown	0.9
10	Road segments	6.3
11	On-road gridded	4.4
12	Non-road	2.6
13	Non-point	1.9

Table 3. Percentiles (from Figure 4) on the CDFs from the 100 MC runs for the Maximum Receptor (Left Side) and the Spatial Average over all Census Tract Population Centroid Receptors (Right Side) for the Annual-Average Benzene (Top) and 1,3-Butadiene Concentrations (Bottom). In Addition, the Geometric Mean of the ± Factor Defining the 95 % Range is Listed. Various Averages for the Groups are Presented at the Very Bottom of the Table.

Percentile of Modeled Benzene Concentrations (µg/m^3)

	Maximum Receptor			Spatial Average		
Percentile	**AERMOD**	**ISCST3**	**ISCST3-U**	**AERMOD**	**ISCST3**	**ISCST3-U**
Min	1.62	3.96	3.27	0.30	1.42	1.19
2.5	2.59	4.22	3.55	0.47	1.45	1.22
5	3.24	4.51	3.82	0.54	1.49	1.25
10	3.55	4.93	4.31	0.71	1.60	1.31
25	4.61	6.11	5.28	0.96	1.99	1.68
50	6.16	7.42	6.55	1.28	2.48	1.98
75	8.60	11.30	8.66	1.70	3.11	2.50
90	11.77	13.69	12.40	2.17	4.15	3.38
95	13.70	18.89	15.11	2.61	5.09	3.78
97.5	18.86	20.13	16.07	3.23	5.15	3.99
Max	21.70	28.25	22.75	3.70	6.92	5.29
50th/2.5th	2.38	1.76	1.85	2.72	1.71	1.62
97.5th/50th	3.06	2.71	2.45	2.52	2.08	2.02
geom mean	2.70	2.18	2.13	2.62	1.89	1.81

Percentile of Modeled 1,3-Butadiene Concentrations (µg/m^3)

	Maximum Receptor			Spatial Average		
Percentile	**AERMOD**	**ISCST3**	**ISCST3-U**	**AERMOD**	**ISCST3**	**ISCST3-U**
Min	0.25	0.41	0.27	0.06	0.21	0.10
2.5	0.31	0.59	0.33	0.09	0.24	0.12
5	0.37	0.61	0.37	0.11	0.26	0.13
10	0.46	0.64	0.42	0.12	0.28	0.14
25	0.60	0.91	0.60	0.18	0.34	0.17
50	0.90	1.20	0.78	0.23	0.43	0.20
75	1.28	1.58	1.09	0.29	0.52	0.25
90	1.72	2.20	1.47	0.38	0.64	0.32
95	2.15	2.61	1.89	0.43	0.70	0.33
97.5	2.29	2.97	2.01	0.46	0.77	0.38
Max	3.13	3.96	3.07	0.53	0.84	0.43
50th/2.5th	2.9	2.03	2.36	2.56	1.79	1.67
97.5th/50th	2.54	2.48	2.58	2	1.79	1.9
geom mean	2.71	2.24	2.47	2.26	1.79	1.78
ben&but avg	2.71	2.21	2.3	2.44	1.84	1.8
3 model avg		2.41			2.03	

Table 4a. Multiple Regression Coefficients for Benzene (e.g., see Equation 3). Bold Numbers are Significant at the 95 % Level.

Variable	AER_MAX	AER_AVG	ISC_MAX	ISC_AVG	ISC_MAXU	ISC_AVGU
CAT_1	**0.573**	**0.433**	**0.777**	**0.689**	**0.639**	**0.676**
CAT_3	**0.379**	**0.389**	**0.169**	**0.206**	**0.262**	**0.266**
CAT_10	0.078	0.063	0.063	0.019	0.053	0.024
CAT_11	**0.248**	**0.167**	**0.183**	**0.109**	**0.327**	**0.14**
WS	**-0.204**	**-0.186**	**-0.220**	**-0.247**	**-0.238**	**-0.285**
ROUGH	**-0.269**	**-0.478**	NA	NA	NA	NA
SIGMA_Z	**-0.115**	**-0.201**	**-0.137**	**-0.347**	**-0.214**	**-0.284**
Intercept	7.940	2.095	10.582	4.522	9.719	3.398
Correlation Coef.	0.924	0.931	0.943	0.942	0.928	0.942

* All bolded values are significant.

Table 4b. Multiple Regression Coefficients for 1,3-Butadiene (e.g., see Equation 3). Bold Numbers are Significant at the 95 % Level.

Variable	AER_MAX	AER_AVG	ISC_MAX	ISC_AVG	ISC_MAXU	ISC_AVGU
CAT_1	**0.064**	**0.264**	**0.114**	**0.406**	**0.053**	**0.370**
CAT_2	0.049	**0.329**	0.046	**0.137**	0.063	**0.272**
CAT_3	-0.015	**0.249**	0.037	**0.347**	0.012	**0.323**
CAT_7	-0.010	0.035	-0.023	0.036	-0.019	0.068
CAT_10	**0.895**	**0.253**	**0.943**	**0.512**	**0.949**	**0.392**
CAT_11	0.045	**0.212**	0.019	**0.120**	0.023	**0.214**
WS	**-0.133**	**-0.210**	**-0.156**	**-0.213**	**-0.145**	**-0.235**
CLD_CVR	-0.041	-0.022	-0.011	-0.021	-0.021	-0.067
ROUGH	**-0.376**	**-0.685**	**NA**	**NA**	**NA**	**NA**
SIGMA_Z	**-0.076**	**-0.319**	**-0.053**	**-0.519**	**-0.130**	**-0.545**
Intercept	1.101	0.394	1.058	0.618	0.875	0.314
Correlation Coef.	0.966	0.920	0.988	0.958	0.993	0.956

*All bolded values are significant

Table 5. Fraction of Explained Variance Contributed by the Group of Emissions Inputs (with Prefix CAT) in Table 4. The Remaining Fraction is Contributed by the Group of Meteorological Inputs.

	AER_MAX	AER_AVG	ISC_MAX	ISC_AVG	ISC_MAXU	ISC_AVGU
Benzene	0.81	0.55	0.91	0.74	0.85	0.77
1,3-But	0.83	0.36	0.97	0.65	0.96	0.59
Average	0.82	0.45	0.94	0.70	0.91	0.68