Uncertainties in Benzene and 1,3-Butadiene Emissions in Houston and their effects on Uncertainties in Concentrations Calculated by AERMOD and ISC

Steven Hanna
Hanna Consultants, 7 Crescent Ave., Kennebunkport, ME 04046
hannaconsult@adelphia.net

Robert Paine, David Heinold, Elizabeth Kintigh
ENSR, 2 Technology Park Drive, Westford, MA 01886-3140
rpaine@ensr.com, dheinold@ensr.com, ekintigh@ensr.gov

H.Christopher Frey
Department of Civil, Construction, and Environmental Engineering, Campus Box 7908, North Carolina State Univ., Raleigh, NC 27695-7908, frey@eos.ncsu.edu

Dan Baker
Shell Global Solutions, 3333 Highway 6 South, Houston, TX 77082-3101
dan.baker@shell.com

Richard Karp
American Petroleum Institute, 1220 L Street, Northwest, Washington, DC 20005-4070
karpr@api.org

ABSTRACT

Because of the interest in air quality studies of toxics in urban areas, a Monte Carlo (MC) probabilistic uncertainty study is being conducted for a 15 km by 15 km domain centered on the Houston Ship Channel. This American Petroleum Institute (API) study is related to a similar MC uncertainty study being carried out over a much broader domain in the Houston metropolitan area by the EPA. The focus of the current study is on uncertainties in ISC3ST and AERMOD predictions of annual averaged concentrations of benzene and 1,3-butadiene, due to uncertainties in emissions and meteorological inputs. The uncertainties in emissions components are being estimated based on observed data variability supplemented by guidance from an API-EPA workshop held on this topic (typical emissions uncertainties are about +/- a factor of three (i.e., covering the 95 % range) for 21 benzene emissions categories and 13 1,3-butadiene emissions categories). The uncertainties in meteorological inputs (such as wind speed and stability) are also being determined from analysis of the field data used to develop parameterizations of dispersion coefficients, plus consultation with experts. However, the current paper addresses only the emissions uncertainties. ISC3ST and AERMOD are being run 100 times in MC mode, using random and independent perturbations of all inputs in order to estimate 1) the total uncertainty of the annual averaged concentrations, and 2) the inputs with uncertainties that are most strongly correlated with uncertainties in predicted concentrations. The results of the MC runs with ISC3ST and AERMOD will be discussed in a later paper.

INTRODUCTION

This paper describes a Monte Carlo probabilistic uncertainty analysis related to air toxics studies in urban regions. The EPA (2000) Integrated Urban Air Toxics Study (IUATS) presents a framework for addressing air toxics in urban areas. A test example for the Houston and Phoenix areas was given by the EPA (1999), and various enhancements including improved emissions models for mobile sources were tested for Houston in a later report (EPA, 2002).

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).

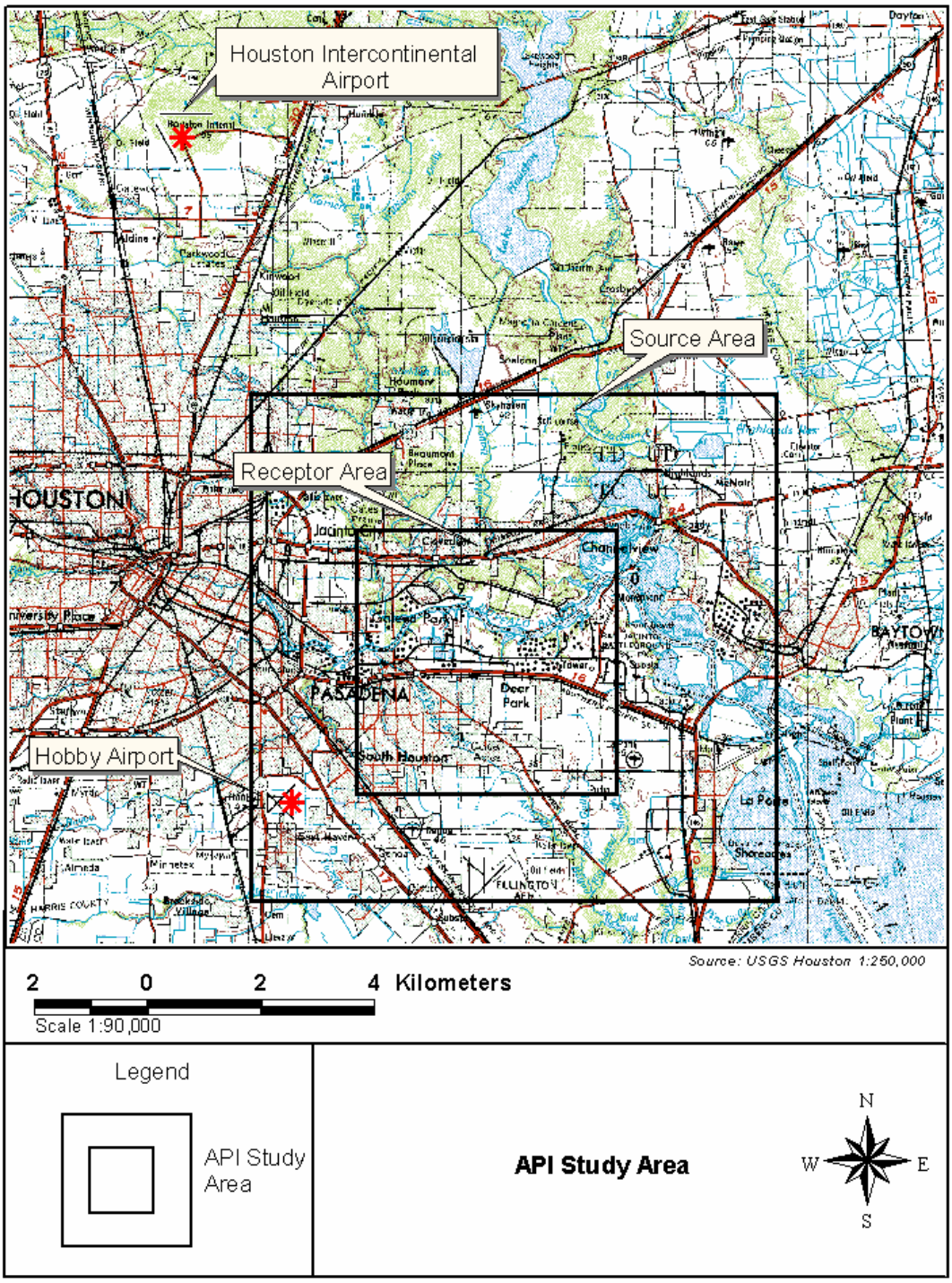
The specific uncertainty study that is being discussed here began with the Houston example described by the EPA (1999, 2002), which used a 150 km by 150 km urban geographic domain including Houston and made use of an emissions model and a transport and dispersion model (ISC3ST) to calculate annual average concentrations of five toxic pollutants. The current plan focuses on a smaller 15 km by 15 km Houston domain (see the inner square in Figure 1) concerning the area around the Houston Ship Channel. The concentrations in the inner square in Figure 1 are calculated using emissions information from sources on a 30 km by 30 km domain, indicated by the outer square in Figure 1. The Houston ship channel domain includes many oil refineries and chemical processing plants, as well as numerous major highways.

Annual averaged concentrations at 49 receptors (46 at census tract centroids and three at monitoring sites) are being studied in the Houston 15 km by 15 km receptor domain. The current MC uncertainty study focuses on two pollutants – benzene and 1,3-butadiene, whose primary sources are mobile sources and industrial sources.

Two alternate transport and dispersion models are being run in the Monte Carlo exercise: ISC3ST(EPA, 1995) and AERMOD (Cimorelli et al., 1998). The necessary emissions files and the major base runs for the Houston domain in Figure 1 are available from an earlier sensitivity study by Heinold et al. (2003) of ISC3ST and AERMOD on this same geographic area.

In the MC probabilistic uncertainty methodology, the modeling system is run 100 times for random choices of variations in the input parameters and the responses of the key model output parameters are analyzed. 100 MC runs are being made for the combined emissions and dispersion model for the AERMOD, ISC3ST (all rural), ISC3ST (urban grid sources), and ISC3ST (rural grid sources) dispersion models.

Figure 1. Domain of the API uncertainty study



In parallel with the uncertainty study being described in this paper, there is an EPA-sponsored MC uncertainty study currently underway for the larger (about 150 km by 150 km) Houston domain. Benzene is also one of the pollutants being studied by the EPA, and its emissions uncertainties were recently estimated (Frey and Zhao, 2003). The two groups are collaborating in several ways, such as by participated in the emissions uncertainty workshop on 26-27 August 2003 (Hanna, 2003).

This paper addresses the emissions uncertainties. The meteorological and dispersion model uncertainties will be described in a separate paper.

ESTIMATES OF INPUT UNCERTAINTIES FOR EMISSIONS

There are uncertainties and probably also biases in the emissions estimates for Houston, but quantifying these is rather difficult given the difficulty in getting sample data. In an earlier MC study by Hanna et al. (2001) involving photochemical grid models applied over the eastern U.S., the emissions uncertainties were assumed to have approximately a factor of three uncertainty (i.e., 95 % range) and a log-normal distribution. The earlier study assumed that the mean bias was zero, since there was not a good scientific reason for assuming otherwise. Fortunately, the emissions inventory in Houston for benzene and 1,3-butadiene is more complete than in most other parts of the country.

Uncertainties in emissions can be estimated by a combination of two approaches: 1) analysis of available data, and 2) 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).

A key element of the current study has been the 26-27 August 2003 Workshop in Houston on uncertainties of emissions in benzene and 1,3-butadiene (Hanna, 2003). The Workshop employed a combination of the data analysis and expert elicitation approaches. A major recommendation arising from the Workshop was that almost the same 24 benzene emissions categories, suggested by Frey and Zhao (2003) and being used by the EPA for their Houston uncertainty study, should be used for the current uncertainty study. Furthermore, it was recommended that the 1,3-butadiene emissions follow the same strategy, but with different categories, as detailed below. It should be noted that most of the discussions of 1,3 butadiene emissions took place after the Workshop.

The Workshop participants discussed the uncertainties for the 21 benzene categories defined in Table 1. The emission amount and the percentage for each category are listed. There are other emissions in other categories, but it is rationalized that those emissions are relatively small and will not significantly impact the total model uncertainty. The categories in Table 1 are somewhat different from those in the Frey and Zhao (2003) report and in the Hanna (2003) Workshop summary, since subsequent discussions resulted in some realignment and replacement. This is mainly because the Ship Channel

Table 1. Emission categories used for benzene in the uncertainty assessment

Cate-Gory	Description	TPY	% of Total	Emissions Type and Source
1	Light Duty Gas Vehicles (LDGV), Light Gas Trucks (LDGT), Road Segments	475.0	28.5	On-Road - HDDV, HDGV, LDGT, LDGV, MC, and All Road Segments
2	Petroleum refineries	412.7	24.7	Point - Petr Refineries, Catalytic Cracking, and Sulfur Plants; Non-Point - Petr Refining-Nat Gas Support
3	Non-road 4-stroke gas engines, Internal Combustion Engines	145.8	8.7	Non-Road - Res. Heat-Distillate Oil; Point - Internal Comb Engines; Non-Pt - Station Inter. Comb. Eng. Diesel & Nat. Gas
4	Non-rd 2-stroke gas engines	34.3	2.1	Non-Road - Off Highway Gas. 2-stroke
5	Non-road diesel (const-ruction, farm, and industr)	26.2	1.6	Non-Road - Off Highway Diesel
6	Oil and gas production	10.2	0.6	Non-Pt Oil & Nat. Gas Prod and support
7	Natural gas transmission and marine transport	63.7	3.8	Non-Point - Nat Gas Transmission & Storage; Marine Cargo Handling
8	Forest wildfires, Municipal Landfills	5.8	0.4	Non-Point - Open Burning-Scrap Tires, Forests & Wildfires, POTWs,; Point - Municipal Landfills
9	Solid waste disposal (sewage treatment, aeration tanks)	59.2	3.5	Point - Waste Disposal and Solid Waste Disposal
10	Acetylene prod (butylenes, ethylene, propylene, olefin)	47.8	2.9	Point - Acetylene Production
11	Fuel oil external combustion, External Combustion Boilers	37.9	2.3	Point - Fuel Oil External Comb, External Comb Boilers; Non-Point - POTW Digest Gas, Res. Heat. Distillate Oil,
12	Typical ethylene plant	17.0	1.0	Point - Ethylene Plant
13	Gas service stations stage 1	9.6	0.6	Non-Point - Gasoline Distribution Stage I & II
14	Petroleum industry fugitives	26.8	1.6	Point - Petroleum Industry Fugitives
15	Managed burning, prescribed	0.6	0.04	Non-Point - Open Burning: Prescribed
16	Chemical manufacturing; fugitive emissions	16.7	1.0	Pt - Chem Manuf: Fug Emis; Non-Pt - Indus Org & Inorg Chem Manuf, Misc. Org Chem Proc; On-Rd - LDDV
17	Aircraft	6.5	0.4	Point – Aircraft
18	Petr ind; fug emis; misc. Petr & Solvent Evap.	121.8	7.3	Point - Petroleum and Solvent Evaporation
19	Process vents in refinery production	15.0	0.9	Point - Process Vents in Refinery Production
20	Loading, ballasting, transit losses from marine vessels	21.6	1.3	Point - Loading, Ballasting, Transit Losses from Marine Vessels; Non-Road - Commercial Marine Vessels
21	Industrial Processes	113.3	6.8	Pt - Ind Proc; Non-Pt - Consumer Prod Usage, Architect Surface Coatings, Asphalt Concrete and Roofing Manuf
	Total Emissions	1667.6	100.0	

area has a larger fraction of industrial sources than the larger Houston domain studied by Frey and Zhao (2003).

Since most of the individual category uncertainties were found to be in the range of a factor of 1.5 to 3, it was decided, for the purposes of the current study, to simply assume a 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. 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.

There are no correlations assumed between any of the categories in Table 1. 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 a factor of 7.5 (2.5 times the 95 % range). Because the probability of such large extremes is very small (less than 0.1 %), the truncation of the extremes should not significantly bias the simulated distribution..

Similarly, the 1,3-butadiene emissions inventory was studied for the Houston Ship Channel domain and 13 emissions categories were assumed to cover most of the emissions. Table 2 contains the list of the 13 emissions categories, including a description, a total "Tons Per Year" (TPY) for that category, and the associated percent of the total. A much smaller fraction (about 15 %) of the 1,3-butadiene emissions comes from mobile sources, compared to the fraction (about 40 %) for benzene.

In all categories in Tables 1 and 2, the uncertainty is represented by a distribution function (log-normal) and a definition of the 95 % range (i.e., plus and minus three 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.

MONTE CARLO SAMPLING METHODS

The Monte Carlo (MC) sampling procedure is straightforward. By using simple random sampling, standard statistical confidence limit formulas can be applied to the results. Also, with simple random sampling with no assumed correlations among input fluctuations, the number of needed MC runs is not dependent on the number of variables. The number of MC runs is recommended to be 100, and 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. Previous MC studies suggest that 100 runs are sufficient to provide useful results on total uncertainty and on correlations (e.g., Hanna et al, 2001).

Random selections of variables must be constrained to be within known physical bounds. Since the normal and lognormal distributions can go to infinity, there is a small chance that an extreme large or small value may be chosen. For example, an hourly-averaged

emission 1000 times the expected median value is not physically possible. In the current study, MC-selected data are constrained to be within about $\pm 5\sigma$ of the median.

Table 2. 1,3-butadiene emissions source categories

Category	Description	TPY	% of Total
1	Fuel oil external comb, petr and solvent evap, organic solvent evap, fuel fired equip, natural gas, flares, indust proc, petr ind, process gas	271.8	40.1
2	Styrene-butadiene rubber and latex production, nitrile butadiene rubber production	105.8	15.6
3	Chemical manuf fugitive emis, industl processes, general processes, fabricated metal products fugitive emissions, plastics production	118.8	17.5
4	Industrial processes, chemical manufacturing, butadiene fugitive emissions	17.1	2.5
5	Ethylene plant, industrial processes chemical manufacturing butylenes. Ethylene propylene, olefin production fugitives emissions	26.3	3.9
6	Loading, ballasting, transit losses from marine vehicles	10.7	1.6
7	Indust proc, petr industry cooling towers and fugitive emissions from flanges and all streams	13.9	2.0
8	Aircraft	5.1	0.8
9	Unknown	6.4	0.9
10	Road Segments	42.4	6.3
11	On-road Gridded	30.0	4.4
12	Non-road	17.6	2.6
13	Non Point	12.6	1.9
	Total Emissions	**678.4**	**100.0**

The recommendations for emissions uncertainties assumed that a given random number would apply for the entire year of ISC3ST or AERMOD runs. 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) from a Gaussian distribution with mean 0.0 and variance 1.0 using any random number software package. In order to make the comparisons more meaningful in the subsequent analysis, the same sets of n random numbers for m = 100 MC runs will be used for the runs for each model. The exact value of a given input that is used in the model for a given MC run is obtained by converting the random number selected from a Gaussian or normal distribution with mean 0.0 and variance 1.0 to the actual normal or log-normal distribution prescribed for that variable.

ANALYSIS TECHNIQUES

The analysis shall make use of the outputs of the m =100 sets of MC runs with each of the three linked emissions and transport and dispersion models on the Houston domain. Suppose there are 100 sets of predicted annual concentrations for benzene and 1,3-butadiene at each receptor location. There are also 100 sets of randomly-selected variations in each of the n input parameters or model parameters (i.e., the n x m matrix mentioned in the previous section, but only for the inputs that are annual-averaged). The outputs will be analyzed to determine the characteristics of the total variability. The paired outputs and inputs (for annual averaged inputs) will be subjected to a correlation analysis 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. The primary outputs are 1) the annual average concentration averaged over the 46 population-weighted centroids of census tracts, and 2) the annual average concentration at each of the 46 centroids and at the three receptor locations.

Total variability

The 100 sets of MC outputs averaged over the entire domain and at selected receptors shall be rank-ordered and used to define 95 % confidence intervals and maxima and minima. The 95 % confidence range is known as the "total variability" and will probably be close to a factor of two or three. The shape of the output pdf shall be estimated (usually these are lognormal). The results will be studied to see if there is much variability in space from one receptor to another or from one pollutant to another (benzene or 1,3-butadiene). Special attention will be paid to identifying receptors that may be overly influenced by a nearby source, so that these can be analyzed separately.

Statistical analysis techniques such as correlation coefficients

After the MC outputs are created, a variety of statistical analysis techniques will be applied to identify key contributors to output uncertainty. These include scatter plots of inputs versus outputs, correlation coefficient analysis, regression analysis, and principal component analysis (see Cullen and Frey, 1999). Scatter plots are most revealing when uncertainty in output is dominated by uncertainty in a small number of inputs. Linear and nonlinear relationships, as well as patterns and trends, can be identified readily during visual analysis of scatter plots, then tested using other techniques.

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[7]. An identification of this subset of highly significant contributors to output uncertainty can help guide future research. The most common statistical parameter used to identify important inputs is the rank correlation coefficient. In the Hanna et al. (2001) study (with 100 MC runs) of uncertainty in regional ozone

models, out of 128 input variables, there were only about five that yielded correlations greater than 0.5 with predicted maximum ozone concentration.

The analysis will rank the input variables and model parameters whose variations have the strongest correlation with the output variables. The analysis will also allow 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.

ACKNOWLEDGEMENTS

This research has been sponsored by the American Petroleum Institute. The authors appreciate the suggestions and data sets provided by Joseph Touma and John Irwin of the U.S. EPA.

REFERENCES

Cimorelli, A.J.; Perry, S.G.; Venkatram, A.; Weil, J.C.; Paine, R.J.; Wilson, R.J.; Lee, R.F.; and Peters, W.D. *AERMOD – Description of Model Formulation (Version 98314 (AERMOD and AERMET) and 98022 (AERMAP).* USEPA, RTP, NC 27711, 113 pages, 1998.

Cullen, A.C.; Frey, H.C. *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, 1999.

EPA. *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, 1995.

EPA. *Air Dispersion Modeling of Toxic Pollutants in Urban Areas. Guidance, Methodology, and Example Applications*. EPA-454/R-99-021, OAQPS/EPA, RTP, NC 27711, 1999.

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

EPA. *Example Application of Modeling Toxic Air Pollutants in Urban Area*. EPA-454/R-02-003, OAQPS/EPA, RTP, NC 27711, 91 pages + 6 page appendix, http://www.epa.gov/scram001/tt25htm#toxics, 2002.

Frey, H.C.; Zhao, Y. *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 September 2003.

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

Hanna, S.R. *Summary of 26-27 August 2003 Houston Emissions Uncertainty Workshop on Benzene and 1,3-Butadiene.* Prepared for the American Petroleum Institute by Hanna Consultants, 7 Crescent Ave., Kennebunkport, ME 04046, 31 pages, 2003.

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

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