

A BAYESIAN APPROACH FOR THE REDUCTION OF UNCERTAINTY IN THE INDUSTRIAL SOURCE COMPLEX-SHORT TERM MODEL VERSION 3 (ISCST3)

Hwong-wen Ma,^{1,*} Wayne Wei-Yuan Tu,² Douglas Crawford-Brown² and Chi-Feng Chen¹

¹Graduate Institute of Environmental Engineering
National Taiwan University
Taipei 106, Taiwan

²Department of Environmental Sciences and Engineering
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

Key Words: Bayesian approach, ISCST-3 model, uncertainty analysis, risk assessment

ABSTRACT

The Industrial Source Complex-Short Term Model-Version 3 (ISCST3) is a steady-state Gaussian dispersion model designed to support the Environmental Protection Agency's regulatory modeling programs. It is often employed to estimate air concentrations and depositions associated with air emissions from a source of air toxics, and thus to provide exposure information for risk assessment. Combining a prior distribution described by the ISCST3 model output concentration and the associated model uncertainty with a likelihood distribution consisting of measurement data and its associated measurement uncertainty, a posterior distribution describing the uncertainty in the pollutant concentration was produced. The resulting Bayesian methodology was used to determine 95% confidence intervals for pollutant concentrations under a variety of scenarios. The methodology can also determine if the inclusion of the modeling results could significantly reduce the uncertainty associated with estimates of air concentration used in exposure assessment. This study has found that given a large standard deviation of modeling results (geometric standard deviation 5 and above), the Bayesian approach does not result in reduction of overall uncertainty. The Bayesian technique could be very useful for determining where and how modeling and measurement can be combined to more accurately identify pollutant concentrations and in turn, more accurately determining human health risks.

INTRODUCTION

The development of a model for an incinerator risk assessment includes the collection of site-specific information on environmental factors such as air, water and land use conditions, exposure factors, and release characteristics. These characteristics, associated with the incinerator air emissions, are used to develop a multimedia transfer, transport, and transformation model to estimate the concentration distribution in various environmental media and food items with which people come into contact. From this, estimates of the dose distribution that people receive and finally the risk distribution associated with the concentrations are calculated.

Risks arise because of the existence of a hazard and the uncertainty associated with the resulting effect. Uncertainties in risk analyses can arise from incom-

plete knowledge about the system one is studying, including uncertainty due to model formulation and uncertainty due to parameter estimation conditional upon adoption of a particular model form. A better understanding of the uncertainties associated with the model can help to determine the accuracy of the model as a tool in risk assessment. Comprehension of the model's accuracy gives a better understanding, in turn, of the accuracy in estimates of people's exposure to incinerator emissions. From this, the accuracy of estimates of risks of adverse health effects associated with exposure to these emissions can be estimated. Finally, the reliability of decisions as to whether the emission levels are safe for overall human health can be determined. The goal of the decision-maker is to develop policies that will produce the desired risk-based goals with reasonable confidence. Thus, quantifying uncertainty in risk calculations for the model is essential to

*Corresponding author
Email: hwma@ntu.edu.tw

the decision making process.

According to Morgan and Henrion [1], uncertainties can be classified into several sources, including random error and precision, systematic error and accuracy, variability, approximations, and model form uncertainty. All of these sources of uncertainty appear in the practice of calculating risks to individuals and populations from air toxics emitted by a source. Characterization of these uncertainties requires consideration of the uncertainties due to measurements of air concentrations (issues of precision and accuracy), stochastic variation in physical parameters (issue of variability), and model uncertainty (issues of approximations and model form). Reduction of these uncertainties can best be accomplished by making maximal use of both the model predictions and measurement results.

The issue posed now is how modeling results and measurement data can be combined to reduce the overall uncertainty and help in the decision making process. This study accomplishes that goal through use of a Bayesian methodology to investigate how to combine modeling results with field observations to provide more reliable estimates of exposure and risk based on model predictions in ISCST3 (The Industrial Source Complex-Short Term Model-Version 3). The approach taken is to first determine the uncertainty associated with characterization of exposure at a single point in space and time within the exposure field surrounding a source of air toxics. This procedure makes use of Bayes' Theorem to combine measurement and modeling results. Having established this procedure, the research further considers the application of the methodology to more complex problems involving multiple spatial locations and time-weighted average exposures. Throughout the discussion, an example of exposure to air toxics from an incinerator in Taiwan is used. It is important to note, however, that the results of the study are broadly applicable and not restricted to this example.

MODEL UNCERTAINTY ANALYSIS

1. The Exposure Scenario

The exposure scenario used in this study to develop and demonstrate the Bayesian methodology is part of a research to perform a full risk assessment on an incinerator facility located in Taipei County, Taiwan. For that assessment, the ISCST3, a steady-state Gaussian dispersion model designed to support the US Environmental Protection Agency's regulatory modeling programs, will be used to perform air dispersion modeling of the incinerator facility [2].

The facility is a point source with one stack and is considered to be located in a rural setting. The dispersion model requires facility specific input parameters such as the emission rate (1 g s^{-1} for the present scenario), physical release height (120 m), stack tem-

perature (418 K), stack gas exit velocity (9 m s^{-1}), and the inside stack diameter (2 m). In addition, it requires terrain data for the area surrounding the facility as well as hourly meteorological data for calculations of pollutant concentrations. The meteorological data consist of the date, time, wind speed, wind direction, ambient temperature, stability class, mixing height, friction velocity, Monin-Obukhov length, surface roughness length, precipitation code, and precipitation rate. The air concentrations are estimated every 250 m away from the facility until 5 km and at every ten degrees around the facility from 0 to 350° .

2. Determination of Output Uncertainty as A Result of Assigned Uncertainty for Model Parameters

The uncertainty analysis for the model concentrated on using the uncertainties in meteorological data and input source parameters to determine overall model uncertainty. The analysis began with the meteorological data from April 13, 1997 as the baseline scenario. Following Dabberdt and Miller [3], uncertainties were defined for the following variables: wind speed, wind direction, stability, plume rise, and emission rate. Uncertainties were assigned to each variable as included in Table 1. Note that plume rise is dependent on multiple factors: stack gas exit velocity, wind speed, stability, mechanical rise, plume buoyancy, and mixing depth. Since there is a direct linear relationship between plume rise and stack gas exit velocity, the $\pm 25\%$ change in stack height was reflected by using stack gas exit velocity as the adjusted variable [4].

Model runs then were performed using the baseline, upper end and lower end of the confidence interval for each parameter. This produced three estimates of the concentration for each spatial location in the exposure field. The method for identifying output uncertainty was to find the fractional error (FE) associated with a particular alternative scenario (i.e. a scenario with specified input uncertainty). The general equation for FE used in the present study is:

$$FE = \frac{\text{Baseline scenario output} - \text{Alternative scenario output}}{\text{Baseline scenario output}} \quad (1)$$

The above calculation was performed for all directions and all distances from the origin for all of the alternative scenarios associated with a particular input variable to determine output uncertainty (summarized by FE) for each spatial location in the exposure field for the input uncertainty. Then the overall uncertainty associated with that input uncertainty was identified through a histogram. For example, Fig. 1 shows a histogram for output uncertainty associated with input uncertainty of $+15\%$ in wind speed. It was produced

Table 1. Summary of the overall output uncertainties to within one standard deviation given a specific uncertainty for 5 input parameters. The assigned values of input uncertainty are given by Dabberdt and Miller [3]. The Overall Uncertainty shows the uncertainty associated with using a variety of random input uncertainties. The Range column shows the range of values that provide the output uncertainty values for a specified parameter to within one standard deviation. The midpoint value shows the midpoint between the low and high end of the range.

Parameter	Assigned Parameter Uncertainty	Range of Output Uncertainty (%)	Midpoint of Output Uncertainty (%)
Wind direction	± 5 Degrees	100-120	110
Wind speed	± 15%	0-10	5
Stability class	± 1 category	180-200	190
Plume rise	± 25%	20-100	60
Emission rate	± 15%	20-40	30
Overall uncertainty		200-350	275

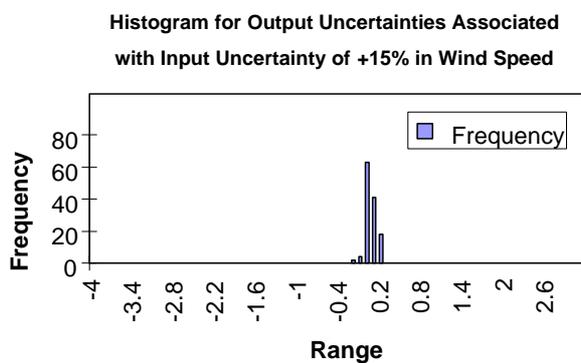


Fig. 1. Output uncertainties associated with an input uncertainty of 15 % in wind speed. The x-axis gives the range of uncertainty values from -4 to 3 in increments of 0.1. The y-axis shows the frequency with which output uncertainties fall within a particular range.

using bins of width 0.1 for the FE (i.e. bins of: 0-0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4, 0.4-0.5, etc.)

A calculation was then made to determine the 95% percent confidence interval for the FEs in the same population for each source of uncertainty (i.e. the value of the FE for which at least 95% of the spatial locations produce a FE equal to or less than this value; this yields an upper bound estimate of the FE likely to be associated with a prediction at a randomly selected spatial location). Taking into account both the uncertainty introduced when wind speed is underestimated and when wind speed is overestimated, an overall uncertainty in predicted air concentration due to uncertainty in wind speed of 10% is indicated.

The above process was repeated for all input parameters. Results are summarized in Table 1. In each case, the magnitude of the overall uncertainty due to uncertainty in a given parameter reflects the 95% confidence interval for the FEs introduced by variations in that parameter within the range noted by Dabberdt and Miller [3]. It may be assumed, therefore, that the actual uncertainty associated with any given spatial

location will be at or below this value. The uncertainty shown in this table, therefore, is likely to be an upper end estimate of that uncertainty (i.e. a conservative estimate, as is common in risk assessment).

In addition to calculating the individual output uncertainties associated with particular inputs, the overall uncertainty associated with simultaneous uncertainty in all parameter values was calculated. This was performed in Monte Carlo fashion by randomly selecting input parameter values from the upper and lower ends of their uncertainty distributions, followed by running ISCST3. The same calculation of FE described previously then was performed for each spatial location and a histogram developed. The 95% confidence interval on FE was determined using the set of all of these Monte Carlo simulations. Table 1 shows the range of FEs to be expected from each source (95% confidence interval), the midpoint of this range, and the overall uncertainty to within one standard deviation.

BAYESIAN ANALYSIS

1. Bayesian Methodology

In estimating the concentration, both the measurement result and the modeling result provide information, but both also have associated uncertainties. The present study uses a Bayesian methodology to combine the two results and reduce the overall uncertainty. The framework for the Bayesian analysis was based on Bayes' theorem [5].

Before obtaining the monitoring data, the user of the Bayesian methodology needs to know the confidence with which it may be stated that the pollutant concentration is at a particular value, C [6]. That confidence is provided by a model result and its associated uncertainty (here, the result of the ISCST3 model run under the same conditions as the measurement). The uncertainty in the model result may be summarized by a probability density function (PDF) with

mean equal to the model result and variance described by the results of Section 2. This PDF is referred to in the Bayesian methodology as the prior since it is generated prior to the measurement result [7].

Next, observations are made in the field to provide measurements of pollutant concentrations. Whether it is due to human error in measurement, improper calibration of the instrument or other cause, the measurement result will also have an uncertainty associated with it. This results in a PDF describing the uncertainty in the concentration based solely on the measurement. This PDF is referred to as the likelihood PDF in the Bayesian methodology.

Taking into account both the modeling uncertainty (prior PDF) and the measurement uncertainty (likelihood PDF), the Bayesian methodology can be used to generate a composite PDF describing the overall uncertainty in the estimate of concentration at a particular spatial location where both the model result and measurement result are available. This combined PDF confidence is called the posterior.

The following equation calculates the PDF for the prior, which was determined from the analysis in Section 2 to be approximately normal distribution:

$$\text{Prior}(C) = \frac{\exp\left(\frac{-(C - C')^2}{2\sigma^2}\right)}{\sigma(2\pi)^{0.5}} \quad (2)$$

where

$\sigma = F \times C'$

C' = concentration value based on model predictions

C = the value at which the model concentration is being evaluated (i.e. the value for which the prior is being estimated)

σ = standard deviation associated with the uncertainty in the predictions of concentration from the model

F = FE associated with a particular model concentration

Equation 3 calculates the PDF of the likelihood describing the probability that a measured concentration, M , would be produced if the true concentration were C_m [8]. This PDF was chosen here to be Normal since the PDF for measurement uncertainties typically is Normal.

$$L(M | C_m) = \frac{\exp\left(\frac{-(C_m - M)^2}{(2\sigma)^2}\right)}{\sigma(2\pi)^{0.5}} \quad (3)$$

where

$\sigma = F_m \times M$

M = the measured concentration at the spatial location

C_m = the "true" concentration value at the spatial location

F_m = FE (standard error of the mean) associated with the measurement of concentration at this spatial location

Equation 4 calculates the posterior PDF representing the confidence that the "true" concentration lies in the neighborhood of any value, C .

$$\text{Posterior}(C | \text{Measurement}) = \left(\frac{\text{Prior}(C) \times L(M | C)}{\int \text{Prior}(C) \times L(M | C)} \right) \quad (4)$$

2. An Example

The above PDFs was approximated by a discrete distribution. The interval (or bin) width was selected to be $1E-4$ (10^{-4}) units of air concentration to provide reasonable accuracy in the approximation. The process was performed through an EXCEL spreadsheet. For the purpose of demonstration, Fig. 2 shows an example of the PDF calculation result. Figure 2a displays the discrete prior PDF associated with the model estimate in a particular spatial location. The dots represent the height of the prior in each interval. Figure 2b displays the discrete likelihood PDF for the case where the measurement mean equals the model result. The dots represent the height of the likelihood in each interval. Figure 2c displays the discrete posterior PDF for the case where the measurement mean equals the model result. The dots represent the height of the posterior in each interval. The lines in Fig. 2 represent a smoothed function fitting the dots (i.e. the approximation to a continuous PDF).

The standard deviation of the posterior then was determined by progressively summing the confidences in intervals surrounding the mean until a value of approximately 95% was obtained. For the particular posterior distribution used in this example, this 95% confidence interval is $0.75-0.95 \text{ pg m}^{-3}$, which indicates a standard deviation of 0.05 pg m^{-3} .

3. Summary of Posterior Distributions under 48 Conditions

The method outlined in the example considered above was applied to a broad range of "true" concentrations and ratios of the mean measurement results to model results. Based on the analysis of Venkatram [9], it was assumed that the model result was unlikely to differ from the measurement by more than an order of magnitude. Specifically, values of measurement concentration equal to the model concentration (C_m), equal to twice the model concentration (called here C_{m*2}), equal to five times the model concentration (called here C_{m*5}), and equal to ten times the model concentration (called here C_{m*10}) were considered.

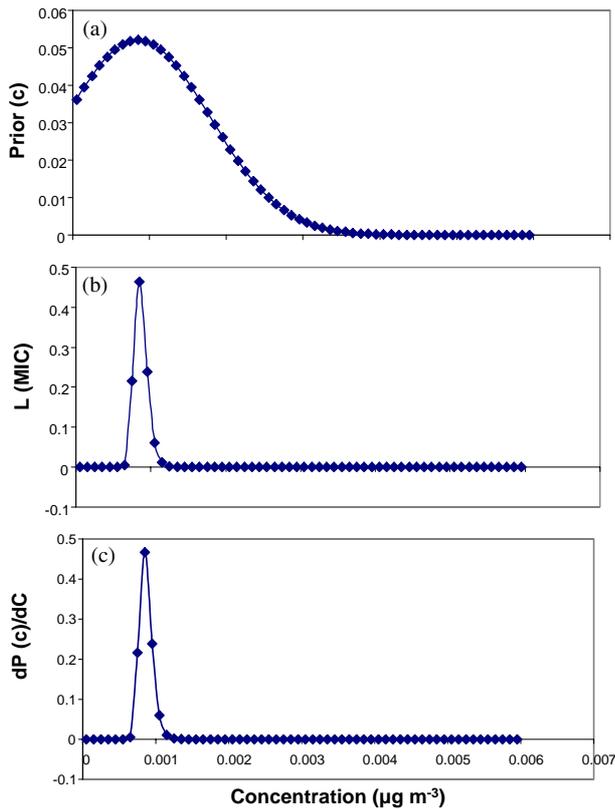


Fig. 2. Example graphs of the PDFs of the prior, the likelihood, and the posterior. (a) Graph of the prior PDF. The x-axis gives the possible concentration values and the y-axis shows the probability corresponding to each pollution concentration. The total confidence is shown as the area underneath the curve. (b) The likelihood PDF. The x-axis gives concentration and the y-axis shows the probability corresponding to each pollution concentration. The total confidence is shown as the area underneath the curve. (c) Posterior confidence that concentration is of a particular value. The x-axis shows the concentration in $\mu\text{g m}^{-3}$ and the y-axis shows the probability corresponding to each pollution concentration. Here, the 95% confidence interval shows that concentration lies between 0.75–0.95 pg m^{-3} , with standard deviation of 0.05 pg m^{-3} , given that model concentration and measurement concentration are the same and that model uncertainty is 110% and measurement uncertainty is 10%.

Table 2 summarizes the 95% confidence intervals and their associated standard deviations on estimates of air concentration for a pollutant under different conditions of model/measurement ratios (1, 2, 5 and 10) and measurement uncertainty (standard errors of 10 and 50%). The first column identifies the particular input parameter examined. The second column shows whether the PDF examined showed model concentration equal to measurement concentration, twice, five times, or ten times model concentration. The third

column shows the 95% confidence interval given 10% measurement error. The fourth column shows the standard deviation associated with the values in column three. The fifth column shows the 95% confidence interval for 50% measurement error. The final column gives the standard deviation for the confidence intervals in column five.

DISCUSSION

1. Discussion of the Results

By using a Bayesian technique to combine ISCST3 modeling results (the prior) with measurement data (the likelihood), a posterior PDF is produced describing the uncertainty in the concentration of a pollutant at a location in space and time. However, as will be shown below, we found that the inclusion of the modeling results may or may not result in significant reduction of the uncertainty associated with estimates of air concentration used in exposure assessment. As expected, the results of Section 3 indicate that as measurement concentration deviates further and further from the model concentration, the posterior distribution shifts and broadens. Specifically, the mean of the posterior moves to a point between the centroids of the prior and likelihood distributions. In conjunction with the shifting and broadening of the curve, the range of concentration values that constitute the 95% confidence interval increases.

Notice that the model parameters with smaller associated output uncertainties produced narrower posterior distributions. As the model output uncertainty increased, the posterior distribution characterized by the 95% confidence interval for this distribution broadened as expected. And therefore, the 95% confidence interval also became larger. The same statement is applicable to measurement uncertainty. As measurement uncertainty increased from 10 to 50%, the posterior distributions again became broader with larger 95% confidence intervals. (However, notice that for wind speed, one of the standard deviations was not identifiable.)

The data produced in section 3 present a specific case in the Bayesian methodology, using the numerical values for wind speed, etc, from the exposure scenario studied. Table 3 shows a more generalized scenario utilizing the same Bayesian methodology described above, but without reference to specific wind speeds, etc. Using a lognormal prior and a normal likelihood, the table compares nine sets of posterior distributions. A prior median concentration value of 1 was used. However, the geometric standard deviations (GSD) of these priors were set at 1.1, 2, and 5 (these represent potential model uncertainties encountered in practice, respectively). Each of these priors was compared to three sets of likelihoods. The likelihoods were normal and their means were set equal to the

Table 2. Summary of the 95 % confidence intervals and their associated standard deviations. The first column identifies the particular input parameter examined. The second column shows whether the PDF examined showed model concentration equal to measurement concentration, twice, five times, or ten times model concentration. The third column shows the 95% confidence interval given 10% measurement error. The fourth column shows the standard deviation associated with the values in column three. The fifth column shows the 95% confidence interval for 50% measurement error. The final column gives the standard deviation for the confidence intervals in column five.

		95% Confidence with 10% Measurement Error (pg m ⁻³)	Standard Deviations (pg m ⁻³)	95% Confidence with 50% Measurement Error (pg m ⁻³)	Standard Deviations (pg m ⁻³)
Wind direction	PDF C _m	0.75-0.95	0.05	0.05-1.45	0.35
	PDF C _{m*2}	1.35-1.95	0.15	0.15-2.35	0.55
	PDF C _{m*5}	3.35-4.35	0.25	1.55-3.35	0.45
	PDF C _{m*10}	5.95-7.55	0.4	2.75-4.55	0.45
Wind speed	PDF C _m	0.75-0.95	0.05	0.75-0.95	0.05
	PDF C _{m*2}	0.95-1.15	0.05	0.75-0.95	0.05
	PDF C _{m*5}	1.35-1.55	0.05	0.85-1.05	0.05
	PDF C _{m*10}	Not Measurable	N/A	1.05-1.25	0.05
Stability Class	PDF C _m	0.75-0.95	0.05	0.5-1.45	0.35
	PDF C _{m*2}	1.35-1.95	0.15	0.05-2.65	0.65
	PDF C _{m*5}	3.35-4.75	0.35	1.25-4.45	0.8
	PDF C _{m*10}	6.35-8.35	0.5	2.95-5.95	0.75
Emission Rate	PDF C _m	0.75-0.95	0.05	0.35-1.15	0.2
	PDF C _{m*2}	1.25-1.65	0.1	0.75-1.35	0.15
	PDF C _{m*5}	2.55-2.95	0.1	1.25-1.85	0.15
	PDF C _{m*10}	3.95-4.35	0.1	1.85-2.45	0.15
Plume Rise	PDF C _m	0.75-0.95	0.05	0.05-1.45	0.35
	PDF C _{m*2}	1.45-1.85	0.1	0.55-1.75	0.3
	PDF C _{m*5}	2.95-3.75	0.2	1.35-2.55	0.3
	PDF C _{m*10}	4.95-5.75	0.2	2.35-3.35	0.25
Overall Uncertainty	PDF C _m	0.75-0.95	0.05	0.05-1.45	0.35
	PDF C _{m*2}	1.35-1.95	0.15	0.05-2.65	0.65
	PDF C _{m*5}	3.35-4.95	0.4	0.95-5.15	1.05
	PDF C _{m*10}	6.75-8.75	0.5	2.65-7.25	1.15

prior mean (C_m), 1.2 times the prior mean (C_{m*1.2}), twice the prior mean (C_{m*2}), and five times the prior mean (C_{m*5}). In addition, three standard deviations for the likelihood were used (10, 25 and 50%). Again, these represent the range of measurement uncertainties one is likely to encounter in practice. For each set of posterior distributions, the central tendency estimate (CTE), standard deviation (68% confidence interval), and standard error were examined.

Notice from Table 3 that given a prior GSD of 1.1 and likelihood standard deviations of 10, 25, and 50%, producing a posterior PDF is effective in reducing uncertainty in all but one scenario. The only instance where the posterior PDF is ineffective is with a prior GSD of 1.1, a likelihood standard deviation of 10%, and the likelihood mean set at five times the prior median. Here, the standard error is 0.129 or

12.9%. For each of the other scenarios where prior GSD is 1.1, the posterior standard error ranges between 0.07 and 0.099, a slight improvement over only having either modeling results or measurement data. In addition, notice that the posterior CTE given a prior GSD of 1.1 is approximately equal to the measurement mean when the likelihood standard deviation is low (10%) but is almost entirely driven by the prior median value as the likelihood standard deviation increases.

Examining the second column where the prior GSD increases to 2, producing a posterior PDF is only effective in reducing uncertainty for some of the scenarios. With a likelihood standard deviation of 10%, uncertainty is slightly reduced until the measurement mean becomes 5 times the model concentration (C_{m*5}). Notice here, the posterior CTE is driven mostly by the

Table 3. Comparison of posterior central tendency estimate (CTE), the standard deviation, and the standard error (Std Error) for different combinations of lognormal prior and normal likelihood. A prior median concentration value of 1 with geometric standard deviations (GSD) of 1.1, 2, and 5. These three sets of priors are compared to three sets of likelihood standard deviations (SD) of 10%, 25%, and 50% to produce nine separate sets of posterior results. The likelihood means are set equal to the prior mean (C_m), 1.2 times the prior mean ($C_{m*1.2}$), twice the prior mean (C_{m*2}), and five times the prior mean (C_{m*5}).

		Prior Median = 1 Prior GSD = 1.1			Prior Median = 1 Prior GSD = 2			Prior Median = 1 Prior GSD = 5		
		CTE	Standard Deviation	Std Error	CTE	Standard Deviation	Std Error	CTE	Standard Deviation	Std Error
Likelihood SD = 10%	C_m	0.995	0.07	0.07	0.995	0.09	0.09	0.985	0.1	0.101
	$C_{m*1.2}$	1.075	0.08	0.074	1.185	0.11	0.093	1.185	0.12	0.101
	C_{m*2}	1.225	0.11	0.09	1.955	0.19	0.097	1.975	0.2	0.101
	C_{m*5}	1.165	0.15	0.129	4.775	0.5	0.105	4.915	0.5	0.102
Likelihood SD = 25%	C_m	0.995	0.08	0.08	0.945	0.23	0.243	0.935	0.26	0.278
	$C_{m*1.2}$	1.015	0.09	0.089	1.105	0.28	0.253	1.115	0.31	0.278
	C_{m*2}	1.025	0.1	0.098	1.695	0.49	0.289	1.835	0.52	0.283
	C_{m*5}	1.105	0.1	0.099	3.365	1.27	0.377	4.445	1.32	0.297
Likelihood SD = 50%	C_m	0.995	0.09	0.09	0.825	0.37	0.448	0.685	0.47	0.686
	$C_{m*1.2}$	0.995	0.09	0.09	0.895	0.43	0.48	0.785	0.56	0.713
	C_{m*2}	1.005	0.09	0.09	1.005	0.67	0.667	1	0.995	0.995
	C_{m*5}	0.995	0.09	0.09	0.805	0.8	0.994	0.095	N/A	N/A

likelihood mean. Increasing the likelihood standard deviation to 25%, the posterior is only effective in reducing uncertainty when measurement and model concentrations are the same (C_m). Notice that the mean looks to be driven equally by the prior median and the likelihood mean. When the likelihood standard deviation increases to 50%, the posterior is effective in reducing uncertainty until measurement concentration is twice model concentration (C_{m*2}). In addition, notice that the posterior central tendency is almost entirely driven by the prior median.

The third column where the prior GSD is increased to 5 shows that producing posterior PDFs are ineffective in reducing uncertainty if one has measurement results with errors of 10, 25 and 50%. In addition, notice that given likelihood standard deviations of 10 and 25%, the posterior CTE is driven mostly by the likelihood mean. However, when the likelihood standard deviation increases to 50%, the CTEs become driven by the prior median. Notice that when measurement concentration becomes five times model concentration (C_{m*5}), the CTE drops significantly to 0.09 with an undetectable 68% confidence interval for the standard deviation.

Overall, a few general conclusions can be drawn. First, the Bayesian approach did not produce a large reduction in uncertainty for exposure estimates under most circumstances. Second, for each of the three prior GSDs, given a small likelihood standard deviation (10%), the posterior CTE was approximately equal to the likelihood mean. However, as the likelihood standard deviation increased to 25%, the CTE became increasingly controlled by the modeling result.

With a 50% likelihood standard deviation, the CTE was entirely driven by the modeling result. Third, as the prior GSD increases, fewer and fewer of the posterior PDFs reduced uncertainty, especially with a prior GSD of 5 where none of the posterior PDFs showed a reduction of uncertainty compared to having only measurement data. Here, the recommendation would be made that given a large GSD (5 and above), the Bayesian approach should be avoided as it results in no reduction of overall uncertainty.

The comparison of the posterior standard errors with different likelihood standard deviations as well as different prior GSDs can be used to determine whether it would be worthwhile to employ the Bayesian methodology to reduce overall uncertainty. For example, Fig. 3a compares the posterior standard errors of the three prior GSDs (1.1, 2, and 5) given a likelihood standard deviation of 10%. Figure 3b compares the posterior standard errors of the three likelihood standard deviations (10, 25, and 50%) with a prior GSD of 1.1. The above figures (Figs. 3a and 3b) can be used to determine if a reduction in uncertainty can be expected when one has measurement data with known measurement error and modeling results with known geometric standard deviations. For example, if one has measurement data with measurement error of 10% and modeling results with a geometric standard deviation of 1.7, Fig. 3a can be used. Imagine that the measurement concentration is about twice that of the modeling result. Examining Fig. 3a, with the above given information, the posterior standard error for a GSD of 1.7 lies somewhere between the two series that show the standard errors for GSDs of 1.1 and 2.

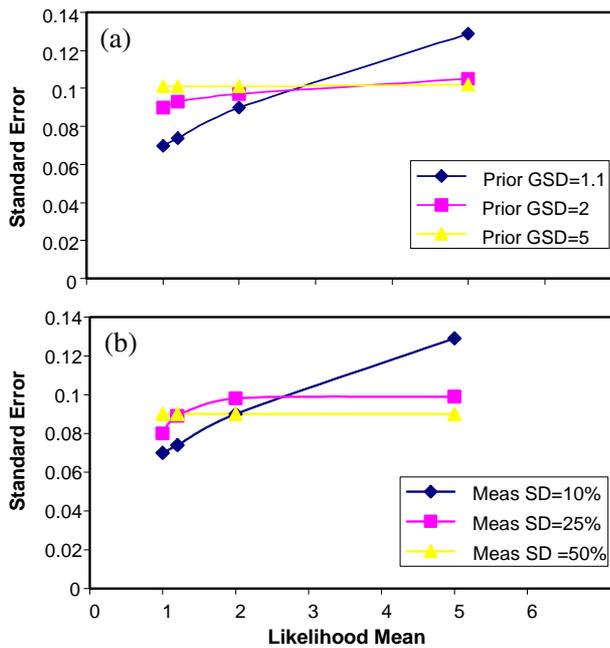


Fig. 3. Posterior standard error comparison for a particular likelihood standard deviation or prior GSD. (a) Posterior standard error comparison for likelihood standard deviation of 10%. The x-axis shows the increase of the likelihood mean and the y-axis shows the posterior standard error. The three series compare the standard errors associated with prior GSD's of 1.1, 2, and 5 respectively. (b) Posterior standard error comparison for prior GSD of 1.1. The x-axis shows the increase of the likelihood mean and the y-axis shows the posterior standard error. The three series compare the standard errors associated with measurement standard deviations (meas SD) of 10%, 25%, and 50% respectively.

At that location, the posterior standard errors are approximately 0.09 and 0.1 respectively. It is expected that given a GSD of 1.7 and 10% measurement error, the posterior standard error at the same location (measurement concentration approximately twice model concentration) is somewhere between 0.09 and 0.1. Thus, it would be worthwhile to employ the Bayesian methodology to reduce overall uncertainty in this instance.

2. Complex Scenarios and Model Calibration

The mathematics described in previous section considered concentration at a single spatial location given measurement data and model prediction with both quantities estimated for a relatively short time period (24-h time-weighted average concentrations). Additionally, the parameter values were relatively stable during the period of calculation and measurement. The methodology developed for this simple case may, however, be expanded to more complex cases involving measurement locations and longer time intervals

over which fluctuations of parameter values occur.

A more complex scenario may again involve concentration at a single point given a time-weighted average measurement but over a long time period (e.g. several years). Over this longer period of time, model uncertainty should be smaller, since fluctuations in wind speed, direction, etc, will be averaged or smoothed. The methodology developed here, however, will continue to apply, with replacement of the model's FE (F) by a smaller value characteristic of predictions of long-term time-weighted averages.

The second complex situation involves multiple spatial locations for measurements, each with measurement data and model predictions for a short time period. One may need to know the mean exposure or concentration to the population spread across several spatial locations. Given posterior PDFs for these separate spatial points, the standard deviation of the concentration estimate for each location can be determined. Knowing standard deviation, the equation for the total mean standard deviation (i.e. standard deviation of the population-weighted average exposure) is given below:

$$(\sigma_{\text{mean}})^2 = (\sigma_1)^2 \times (f_1)^2 + (\sigma_2)^2 \times (f_2)^2 + (\sigma_3)^2 \times (f_3)^2 + \dots + (\sigma_n)^2 \times (f_n)^2 \quad (5)$$

where f_i is the fraction of the population exposed at each spatial location, i , and σ_i is the standard deviation of the posterior at location i .

The point with the highest product of $\sigma \times f$ contributes the most to the total mean standard deviation and also contributes the most to the calculation of risk for the population. Reduction of the overall uncertainty in mean (population-weighted) exposure would, therefore, be accomplished in the most cost-effective manner (i.e. in the lowest number of additional measurements) by focusing on additional measurements at the spatial location with the highest value of this product.

3. Assumptions and Conclusions

It is important to note the assumptions underlying the present analysis. The most evident assumption is that the dispersion of air pollutants can be properly described by a Gaussian distribution. In addition, along with the use of a Gaussian distribution come a host of assumptions and constraints, such as that stated in Beychok [10]. The use of Dabberdt and Miller's [3] article to assign uncertainties to the five input parameters (wind direction, wind speed, stability class, plume rise, emission rate) is also an important assumption. The assigned input uncertainties may not necessarily represent appropriate input uncertainties specific to the ISCST3 model.

The way in which Dabberdt and Miller's assigned uncertainties were used can also be questioned.

For example, in order to determine the output uncertainty associated with a 15% increase in wind speed as the input uncertainty, all wind speed values were uniformly increased by 15% prior to the model run. Given that the model run was for 24 h and ISCST3 required hourly wind speed data, a total of 24 wind speed values were increased by 15%. The assumption made is that the anemometer measuring wind speed is always overestimating wind speed by 15%. This is not necessarily the case. It is more likely that the anemometer could be off within the range of $\pm 15\%$. Thus, a more effective way of estimating output uncertainty could be to randomly select input uncertainties within the range of $\pm 15\%$ for each hourly measurement of wind speed in the 24-h period. The input uncertainties for the other variables could be manipulated similarly to the method.

The substitution of stack gas exit velocity as the input variable used to propagate uncertainty for plume rise is also of concern. As described previously, plume rise is an equation that is a function of multiple variables. Adjusting only stack gas exit velocity to determine the output uncertainty associated with plume rise may not produce a particularly accurate estimate of its uncertainty.

In addition, only five input parameters described above for model uncertainty were considered. Certainly, these are not the only sources of uncertainty within the model. These five input parameters chosen most likely contribute to the bulk of the uncertainty in ISCST3. However, a broader approach would include more parameters for the uncertainty analysis.

In estimating the measurement uncertainty, 10 and 50% were used because they were representative of typical measurement uncertainties. Applying this process to an actual site would certainly require knowing the measurement uncertainty associated with the particular monitor being used for measuring pollutant concentration.

Finally, calculating the prior and likelihood functions, the model output and measurement uncertainties were assumed to be normally distributed. Equations for the prior and likelihood reflected that assumption. Had the assumption been made that model output and measurement uncertainties were lognormally distributed, different prior and likelihood equations would have been used resulting in different posterior distributions.

In addition to the refining of the approach to address the above assumption and limitations, a subject of great interest would be a more complex scenario in which there are multiple points with some not having measurement data. The method of properly interpolat-

ing between points with posterior distributions and points without them would be of great use for decision-makers faced with limited funds. By identifying where the greatest need for monitors is, a significant amount of time and resources can be saved.

Overall, the method presented in the study has the potential to be very useful. By refining the process of uncertainty analysis as well as the Bayesian methodology, the Bayesian technique could be very useful for more accurately identifying pollutant concentrations and in turn, more accurately determining human health risks in a more efficient way.

REFERENCES

1. Morgan, M.G. and M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, New York, NY (1990).
2. U.S. Environmental Protection Agency (USEPA), *User's Guide for the Industrial Source Complex (ISC3) Dispersion Models: Volume I-User Instructions*. USEPA, Washington, DC (1995).
3. Dabberdt, W.F. and E. Miller, *Uncertainty, ensembles and air quality dispersion modeling: Applications and challenges*. *Atmos. Environ.*, 34(27), 4667-4673 (2000).
4. Davis, M.L. and D.A. Cornwell, *Introduction to Environmental Engineering*. 3rd Ed., WCB/McGraw-Hill, Boston, MA (1998).
5. Jeffrey, R.C., *The Logic of Decision*. 2nd Ed., University of Chicago Press, Chicago, IL (1983).
6. Crawford-Brown, D.J., *Mathematical Methods of Environmental Risk Modeling*. Kluwer Academic Publishers, Boston, MA (2001).
7. Schmitt, S.A., *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics*. Addison-Wesley Pub. Co., Reading, MA (1969).
8. Venkatram, A., *The expected deviation of observed concentrations from predicted ensemble means*. *Atmos. Environ.*, 13(11), 1547-1549 (1979).
9. Beychok, M.R., *Error Propagation in Air Dispersion Modeling*. <http://www.air-dispersion.com/feature.html> (Feb. 2001).

Discussions of this paper may appear in the discussion section of a future issue. All discussions should be submitted to the Editor-in-Chief within six months of publication.

Manuscript Received: April 30, 2007
Revision Received: June 13, 2007
and Accepted: June 20, 2007