Design, Application, and Recommendations for Including Inventory Uncertainties in Emission Inventory Preparation for Modeling

Marc R. Houyoux, Daniel H. Loughlin, and Andrew P. Holland
University of North Carolina at Chapel Hill, Carolina Environmental Program,
c/o MCNC, 3021 Cornwallis Road, RTP, NC, 27709
mhouyoux@unc.edu

H. Christopher Frey and Amr Abdel-Aziz
North Carolina State University, Department of Civil Engineering
CB 7908, Mann Hall, Raleigh, NC, 27695-7908

ABSTRACT

The air quality modeling community widely acknowledges that there are considerable uncertainties in the emission inventories used for modeling. Nonetheless, these uncertainties are typically ignored in air quality modeling applications. Three reasons for ignoring these uncertainties are (1) uncertainties in model inputs typically are not well quantified, (2) emissions and air quality modeling codes are not equipped to support uncertain inputs, and (3) methodologies are lacking that address the previous two issues. In this paper, we describe a methodology for integrating some types of emissions uncertainties into emissions modeling. The methodology is applied to an air quality modeling application, allowing the effects of emissions uncertainties on air quality simulations to be characterized. To facilitate this application, the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system has been modified to accept statistical or empirical distributions describing emissions factors as well as multiple realizations of hour-specific emissions data. SMOKE was used with these data to simulate many alternative realizations of the emissions inventory. These realizations were then modeled with the Multiscale Air Quality Simulation Platform (MAQSIP) and the uncertainty in the air quality model predictions was characterized. Computational intensity is addressed by distributing the model runs over a cluster of inexpensive Linux computers. The paper concludes with recommendations for improving the characterization of uncertainty in emissions inventories, with the goal of making probabilistic assessments more practical.

1 INTRODUCTION

A major concern in air quality modeling is uncertainty in emission inventories (EIs). There are fundamental questions regarding whether EIs, when used as inputs to air quality models (AQMs), are sufficiently accurate and precise to provide a basis for the development and evaluation of effective pollutant control strategies. Uncertainties are an unavoidable aspect of emission inventories and emissions modeling; one reason is that emissions data are often estimated by multiplying an emissions factor specific to the type of source by an activity value. For example, county-total emissions for household solvents can be estimated by multiplying a per-capita emission factor (tons/person) by the county population. In some cases, the emissions factor may have been determined through analyzing a small number of emissions samples from only a few sources. The results therefore may not be representative for a particular source. There may also be uncertainties due to measurement error and variability due to changing operating conditions and natural fluctuations (e.g., fuel impurities). One important detail of our work is that to simplify the methodology described here, we have not disassociated uncertainty and variability, although techniques for doing so are available. The activity data are another source of uncertainty. For some sources, activity values are easily measured or computed; for example, the amount of gasoline discharged from a gas station can be determined as a function of sales. In other cases, activity values are not so easily determined. An example is vehicle

miles traveled (VMT), which is used for estimating vehicular emissions. VMT is not measured explicitly for each day but rather is based on a limited sample of observations from particular days. Clearly, a large number of assumptions are used in developing an emissions inventory; however, the effect of these assumptions on air quality modeling efforts is not necessarily clear.

The work described here is part of a larger effort to develop, refine, and demonstrate methods for quantification of the uncertainties in EIs. Our goal is to illustrate the benefits of characterizing uncertainty, particularly with respect to data quality management, research planning, air quality modeling, and control strategy development. While uncertainties in meteorological data or model formation could also be accounted for, these are not the focus of the work described here. In this paper, we describe and demonstrate an emissions modeling system that is capable of utilizing quantitative representations of EI uncertainties in developing probabilistic air quality predictions. Such an analysis is useful in characterizing how lack of knowledge affects modeling results. This information can in turn be used in evaluating the accuracy of model predictions, which assesses the likelihood of achieving desired air quality improvements, and identifies how resources can be allocated most efficiently to reduce uncertainties in air quality predictions.

2 OVERVIEW

Performing a probabilistic air quality assessment involves multiple steps. An initial step is to determine the source categories to be considered. A practical approach is to identify the source categories (e.g., by source classification code) that contribute the highest quantity of emissions. Knowledge about the relative effect of alternative pollutants is also valuable in identifying source categories. The next step is to quantify the uncertainties in emissions factors and activity data for the selected categories. Ideally, the data from which the emissions factors were calculated is readily available. Unfortunately, this availability is not very likely for many source categories, since much of the data was generated years or decades ago and may not have been permanently recorded or maintained. If the data can be obtained, they can be fit with a statistical or empirical distribution. When using statistical distributions, the lack of a large number of data points often introduces uncertainty in the estimates of the distributions parameters. Techniques such as parametric bootstrapping can be used to quantify the uncertainty due to a lack of data. If the data used to calculate emissions factors are not readily available, expert elicitation for characterizing uncertainty is sometimes used. Finally, characterization of uncertainty for activity data can use approaches analogous to those for emission factors.

Once uncertainties have been characterized in an emission inventory, the next step is to propagate the uncertainties through the emissions and air quality models to obtain probabilistic estimates of air quality model predictions. Monte Carlo (MC) simulation techniques are most commonly used for this purpose. In MC simulation, a sample is taken from the distribution for each uncertain parameter. The combination of sampled values across the inventory provides a “realization” of the inventory. In this work, the emissions realization is generated using the emissions model and evaluated using the air quality model to obtain the corresponding realization of air quality predictions. During Monte Carlo simulation, this process of generating and evaluating realizations is repeated typically 50 to several hundred times.

After the output realizations have been generated, they can be evaluated to gain a variety of insights into the air quality modeling results. For example, one can estimate the mean and standard deviation of model results such as the peak predicted ozone concentrations. One can also analyze the air quality model outputs to evaluate the likelihood that air quality standards will or will not be met. Further, sensitivity analyses can be carried out to apportion the uncertainty in model outputs to various uncertain

inputs. This allows the determination of which input uncertainties have the largest impact on the model results, which is information that can be used to allocate resources improve the precision of modeling.

In the work described here, we focus on the propagation step. In practice, this step has not been practical for large-scale modeling because of (1) the computational constraints associated with running complex emissions and air quality models up to hundreds of times, and (2) models have not generally been developed to accept probabilistic inputs. To some extent, computational concerns can be addressed using high performance or distributed computing. We have used the Condor software package released by the University of Wisconsin to distribute computations over a network of computers at locations, connected by the Internet. A distributed computing environment does not address how quantifications of uncertainty in inputs can be handled within a model, however. For example, most emissions models are not equipped to take as input a statistical distribution and its parameters for an emissions source.

One approach for propagation is to develop MC simulation code external to the model. Such code would generate multiple realizations of raw emissions files to the emissions model, which would be used by the emissions model to generate the gridded, hourly emissions data needed by the air quality model. The process of invoking and performing emissions modeling possibly hundreds of times is computationally challenging, and more efficient approaches exist. One such approach is to modify the emissions model to accept probabilistic specifications for emissions and activity factors. MC simulation can then be carried out during the emissions modeling instead of during the inventory building process. Because this latter approach does not require building many versions of emission inventory files that could be hundreds of Megabytes each, it also provides a much more streamlined and tractable solution.

In this context, we have implemented the ability to accept and simulate uncertain inputs into the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. The original version of SMOKE has been more fully documented elsewhere (Coats and Houyoux, 1996). In addition to uncertainties in emissions and activity factors, the modified SMOKE also accepts descriptions of uncertainties in spatial allocation factor assignments. SMOKE propagates these uncertainties through main stages of emissions data processing: inventory import, temporal allocation, spatial allocation, chemical speciation, and merging. The resulting model-ready emissions data is output as a collection of files instead of a single file, in which each file is used for one realization of the Monte Carlo simulation.

We have applied our approach in a real-world case study involving local-scale ozone air quality modeling in Charlotte, North Carolina. The redesigned SMOKE system and our data collection efforts have included VOC and NOx emissions for utility sources and non-utility point sources, on-road mobile sources, nonroad mobile sources, and stationary area sources. In the application described here, we evaluate the effects of uncertainties in utility emissions only. The emissions data were modeled using the Multiscale Air Quality Simulation Platform (MAQSIP) (Odman and Ingram, 1996). No changes needed to be made to the MAQSIP model because it was run in its standard form individually for each Monte Carlo simulation. The model changes were all done in the SMOKE system, as explained next.

3 SMOKE DESIGN AND APPROACH

3.1 SMOKE system overview

The SMOKE emissions processing system efficiently processes emissions data using matrix-vector multiplication. It performs the core functions of emissions processing: the spatial allocation, temporal allocation, chemical speciation, and control of area-source, mobile-source, and point-source emissions, and the generation of biogenic emission estimates. It also computes the elevated plume rise needed for point sources using a modified Briggs algorithm (Briggs, 1984). The version of SMOKE that we used in this project contains a driver for the MOBILE5a and MOBILE5b models (U.S. Environmental Protection Agency (EPA), 1994; U.S. EPA, EPA/OMS MOBILE5 Vehicle Emission Modeling

Software: MOBILE5b, available at http://www.epa.gov/oms/m5.htm#m5b, 1996), and it also uses a reorganized version of the Urban Airshed Model–Biogenic Emission Inventory System, version 2 (UAM-BEIS2) (Pierce et al., 1998). The efficient processing of SMOKE made it an appropriate choice for handling the large number of emissions cases needed for this project. Houyoux et al. (1996) describe this efficiency in more detail.

Figure 1 shows the major SMOKE processing steps for area and point sources. As shown in figure, the temporal allocation, chemical speciation, and gridding steps can be performed in parallel. They are performed before the final merge step, which computes the model-ready emissions. For point sources only, the layer fraction computation causes the merge step to create 3-d emissions for MAQSIP. That step is shown as an optional step in the figure because it applies only to point sources. For both area and point sources, emissions controls can optionally be applied in the control step. The basic mobile source processing is shown in Figure 2. In this effort, mobile sources included computing MOBILE5b emission factors using SMOKE and using them to compute emissions based on vehicle-miles-traveled (VMT) data. SMOKE uses the gridded temperatures to compute source-specific temperatures, which are used during temporal allocation (as shown in figure) to determine the appropriate MOBILE5b emission factors to use for each source and hour. For mobile sources, the temperature preprocessing depends on the gridding step as shown in the figure. The temporal allocation step includes both temporal allocation of VMT and computing emissions from VMT and the MOBILE5b emission factors. MOBILE5b must model the on-road mobile controls in this version of SMOKE. Otherwise, the on-road mobile source processing is the same as for other sources, and the merge step creates the model-ready emissions. We did not include uncertainties in biogenic sources for this project, so we will not discuss the biogenic processing in SMOKE here.

3.2 Approach and design for SMOKE enhancements

For this work, we developed a revised version of SMOKE that permits the propagation of statistical and empirical descriptions of inventory uncertainties for area (including nonroad mobile), on-road mobile, and point sources. Support for uncertainties in biogenic emissions may be added in the future.

Some major sources of emissions-modeling uncertainty relevant to this work included:

- Inventory emissions for area and point sources
- VMT
- Temporal profiles used for temporal allocation
- Speciation profiles used for chemical speciation
- Spatial surrogates used for spatial allocation
- The assignments of the temporal profiles, speciation profiles, and spatial surrogates (e.g., not the temporal profiles themselves, but which monthly, weekly, and diurnal profiles are selected for a given source)
- Layer fraction calculation

For this work, we designed changes to SMOKE to facilitate analysis and air quality modeling that includes two of these uncertainties. First, we considered the uncertainty in the inventory emissions values. The emissions values in the inventory (e.g., annual emissions) were assigned parametric and empirical probability distributions to describe the uncertainty about them. Therefore, SMOKE needed to be able to propagate these uncertainties through the emissions modeling steps to create multiple realizations of the model-ready emissions.

Second, we modified SMOKE to address uncertainties in the assignment of spatial surrogates for spatial allocation of area, non-road mobile, and on-road mobile sources. We permitted each surrogate assignment to include probability information about which spatial surrogate is the best representation of

the spatial allocation of emissions for each assignment. For example, the standard assignment for dry cleaning emissions is to use the population spatial surrogate in all cases. In our approach, one could assign a 50% probability that the population surrogate is the most appropriate surrogate and a 50% probability that the housing surrogate is the most appropriate. As many probabilities and surrogates can be assigned as there are data available to support such assignments.

To create a version of SMOKE that supports inventory emission uncertainties, we addressed three major design elements: (1) developing an approach for providing the uncertainties to SMOKE since they are not available with standard inventories, (2) devising a new emissions processing arrangement that would be both effective and efficient when needing to generate 50 or more emissions realizations, and (3) ensuring a resulting system works both for standard emissions processing and for processing for uncertainty. We describe each of these elements in the paragraphs below.

For the first design element, we devised a new emissions file to allow emissions modelers to provide the quantitative descriptions of uncertainty about the inventory. This approach allows emissions modelers to assign uncertainty information about any existing inventory, without having to change the actual inventory files at all. The same inventories can be used for both deterministic (i.e., without uncertainty) modeling and stochastic modeling (i.e., with uncertainty), and the type of modeling that is done depends on the presence of the additional inventory uncertainty file. This file is called the AUNCERT, MUNCERT, or PUNCERT file, depending on whether the file is for area, mobile, or point sources. The file includes sections for parametric uncertainty profiles, empirical uncertainty profiles, and assignment of those profiles to the inventory by country/state code, county code, and/or source category code (SCC). For point sources, plant identification information can also be used to assign the uncertainty profiles. The parametric information provided in the file defines statistically characterized uncertainties for the normal, lognormal, gamma, and Weibul distributions. The empirical assignments, by definition, support any shape distribution to describe uncertainty about the inventory emissions.

Figure 3 provides an example of an AUNCERT file. In the first section, the /FACTOR ASSIGNMENT/ packet assigns the parametric (Meth=P) and empirical (Meth=E) to several SCCs in the inventory. It is not necessary to provide uncertainty information for all SCCs. In this example, country/state/county code (CFIP) is not used as a basis for making the assignments, because no country, state, or county-specific data were available to justify such assignments. The distributions are assigned by name; for example, the SOLUTILMISC distribution is assigned to SCC 2301030000 for the ROG pollutant. The Approach column indicates when to resample the emissions value from the probability distribution. The four settings one can elect to include for the Approach are the following:

- S: Uses the same emissions adjustment for all inventory sources that match the assignment, and different adjustments for each realization. Adjustment of emissions for each realization therefore occurs once per assignment (e.g., by SCC) only once for the entire episode.
- I: Uses independent adjustments of each inventory source that matches the assignment. In other words, a new sample is taken from the distribution for each source in the inventory. Adjustment of emissions for each realization occurs once for each source that matches an assignment and only once for the entire episode.
- ST: The same as the “S” adjustments, but emissions are readjusted for each hour of the episode instead of once per episode.
- IT: The same as the “I” adjustments, but emissions are readjusted for each hour of the episode, instead of once per episode.

We have created Figure 4 through Figure 6 to help explain our approach for the second major design element: devising a new emissions processing arrangement that would be both effective and efficient when needing to generate 50 or more emissions realizations. These figures illustrate the flow of SMOKE programs and some input, intermediate, and output files that are relevant to this explanation. In

Figure 4, we provide the revised area-sources processing approach for handling uncertain inventories. The Smkinven program imports the inventory from the inventory file, ARINV, and it optionally reads the uncertainty information from the AUNCERT file described above. Since the uncertainty input information is optional, all uncertainty-based steps in Figure 4 include the optional designation; nevertheless, they are required if one wants to include uncertainty in the emissions processing. The AUCOUT file identifies the uncertain sources, the AUCEOUT file contains the empirical distribution information, and the AUCPOUT file contains the parametric distribution detailed. The Grdmat and Spcmat programs perform respectively the spatial allocation and chemical speciation steps as in the original version of SMOKE. The Temporal program performs the temporal allocation and performs the sampling of emissions for each source, based on the instructions in the AUNCERT file. The Temporal program generates both the standard hourly emissions file ATMP and N_T number of ATMPU files, where N_T is the number of realizations defined by the SMOKE user. To conserve disk space, the ATMPU files contain only those sources that are being treated as uncertain. The ATMP file has the emissions for the uncertain sources zeroed out, to prevent double counting when the ATMP data and ATMPU data are recombined for each realization. The Smkmerge program creates the hourly, gridded, speciated model-ready files with all "certain" sources in the AGTS_L file shown in the figure at the top right. The Smkmerge program is also run in "uncertain" mode and iterates through N_T realizations to create an AGTS_L file including the uncertain source emissions for each realization. SMOKE's Mrggrid program, not shown on the diagram, is then used for each realization to create the model-ready emissions for input to the MAQSIP model.

Figure 5 provides a similar flow diagram for processing point sources. The differences between this figure and Figure 4 are (1) the file names (start with "P" for point instead of "A" for area), (2) it includes the optional hour-specific input file PTHOUR and the resulting hour-specific intermediate file PHOUR, and (3) it includes the Laypoint program for computing the layer-fractions in each hour. The 3-d PGTS_L files are computed by the Smkmerge program for the "certain" inventory sources at the top right as well as the uncertain sources for realizations 1 through N_T, shown at the bottom right. Mrggrid also combines these files for each realization to provide the model-ready point source inventories.

Finally, Figure 6 provides the on-road mobile source processing steps. The differences between this figure and Figure 4 are (1) the file names (start with "M" for mobile instead of "A" for area), and (2) the figure includes the alternative approach to computing emissions using MOBILE5b emission factors. As mentioned in describing Figure 2 as well, the Grdmat spatial allocation step is necessary prior to preprocessing the temperature in the meteorology files, which is done using the Premobl program. This is because the Grdmat program creates an "ungridding matrix," MUMAT, which facilitates the computation of source-based temperatures from the gridded meteorology to use in MOBILE5b. The Premobl program creates two intermediate files (not shown) that are used by the Emisfac program, which drives the MOBILE5b program – running it as many times as necessary for the temperatures and MOBILE5b input scenarios provided in a separate input file, which is also not shown on the figure. The Emisfac program stores the emission factors in the MEFSND and MEFSD files. The Temporal program computes the on-road emissions from these emission factors and from the VMT in the MOBL inventory file, and then it applies emissions adjustments dictated by the MUCOUT, MUCEOUT, and MECPOUT uncertainty intermediate files. As was done for area and point sources, the Smkmerge program is used twice: once to compute emissions for the "certain" sources, and once for all realizations of the uncertain sources. The Mrggrid program is used to combine the area/nonroad, biogenic, point, and mobile files for each realization for the single 3-d emissions file input to the MAQSIP model.

4 ADDRESSING COMPUTATIONAL REQUIREMENTS

Carrying out a Monte Carlo simulation involving regulatory-scale emissions and air quality models is a computationally challenging task. For example, the air quality modeling episode described later in

this paper required approximately 11 hours to execute on a 2 GHz Linux computer with 512 megabytes of memory. Evaluating 50 realizations on that computer would therefore require approximately 23 days. Our goal was to evaluate 50 realizations for up to 12 different scenarios. Clearly, this is impractical on a single computer configured as we have described. Another option is to use multiple computers to carry out the runs. In its simplest form, this requires installing the models and data on each computer, setting up batch files, starting execution, monitoring execution, then compiling all the results to a single computer for analysis. This process is very tedious and labor intensive.

As an alternative, we chose to implement a distributed computing approach based upon the University of Wisconsin's Condor Project and software (http://www.cs.wisc.edu/condor). We installed the Condor software on 9 Linux computers. These computers were located at NC State University in Raleigh, NC, and at MCNC, in Research Triangle Park, NC (approximately 20 miles apart). The two clusters were connected only via the Internet. Five of the computers were 450 MHz computers with 256 megabytes of memory. The other four were more powerful, ranging from 1.8 to 2.4 GHz, each with 512 megabytes of memory. Some of the computers were dual-boot systems, and these were unavailable for use during certain times of the day; this was easily handled by the Condor software.

One of the computers was configured to be the master node of the ad-hoc cluster. This computer had a large disk drive and was able to store the model inputs and outputs. From this machine, we submitted the modeling jobs to the Condor cluster, and the Condor software then determined which of the cluster's computers were available and delivered the model and requisite data to the available machines. The Condor software monitored the MAQSIP execution and retrieved the MAQSIP results to the master node upon completion of each modeling "job." If execution was interrupted on a given machine, the job was resubmitted when another machine became available. The Condor software supports a feature called break-pointing, in which interrupted jobs can be restarted where they were interrupted, although we did not utilize this feature in this project. Our Condor-based distributed computing approach was successful in decreasing the time requirements for evaluating 50 emissions realizations from 23 days to less than 4 days.

5 APPLICATION

In this section, we describe an application of our approach to a problem involving uncertainties in utility emissions. This section includes descriptions of the modeling scenario, the emission inventory, and the MAQSIP simulations.

5.1 Example modeling case

We performed our modeling on a local-scale ozone modeling application for Charlotte, North Carolina. Charlotte has been close to noncompliance for the 1-hour ozone standard, and is expected to be out-of-compliance with the 8-hour standards. The modeling episode used in this exercise had previously been used by the state of North Carolina in its 1-hour State Implementation Plan (SIP) modeling efforts. The domain is represented using a 4-km Lambert conformal grid, as shown in Figure 7. The detailed parameters of the case are the following:

Grid projection: Lambert Conformal with Alpha=30, Beta=60, Gamma=-90, and center at (-90,40).

Domain: Origin at (672, -564) kilometers with 63 rows by 63 columns and 4-km square grid cells. (Figure 7)

Episode: Six days, from July 10th to July 16th, were modeled using one six-day (145-hour) period. The episode starts and ends at 12 GMT.

Meteorology data: 1995 meteorology files in MCIP output format were generated originally by MCNC using the Pennsylvania State University (PSU)-National Center for Atmospheric Research (NCAR) Fifth-generation Mesoscale Model (MM5) (Grell et al., 1994).

5.2 Emission Inventory

The emission inventory was the 1995 North Carolina (NC) Department of Environment and Natural Resources (DENR) inventory for SIP modeling. The inventory includes the typical four emissions source categories (area, biogenic, mobile, and point sources), and data for nitrogen oxides (NOx), reactive organic gases (ROG), and carbon monoxide (CO). The area-source inventory in this case includes the nonroad mobile sources, while the mobile inventory includes on-road mobile sources only, using VMT and MOBILE5b emission factors computed using SMOKE. The point source inventory contains hour-specific emissions from the 1995 Continuous Emissions Monitoring (CEM) database of NOx for North Carolina and South Carolina electric generating utilities (EGUs). Table 1 lists the EGU facilities for which we used the CEM data.

To this inventory, we collected the following information about uncertainty in the inventory:

- Parametric and empirical probability distributions of uncertainty on the average-day emission values for NO_x and VOC (Frey and Bammi, 2002; Frey and Li; Frey and Zheng, 2002; Frey and Zheng, 2002]
- Empirical probability distribution for uncertainty in the hourly NO_x CEM data to use for both base-year and future-year emission estimates (Abdel-Aziz and Frey, 2002)

The case that is described in this paper includes only the uncertainty characterization from analysis of the CEM data for the EGUs. We did not complete the testing of the SMOKE system and the MAQSIP runs for the other data available for area, nonroad mobile, on-road mobile, and non-utility point sources in time to include in this paper.

Based on analysis of the CEM data (Abdel-Aziz and Frey, 2002), 50 realizations of the PTHOUR input file were created for input to SMOKE. These files were used to create suites of 50 realizations of model-ready inputs with uncertainty characterized for the EGU sources only; area, nonroad mobile, on-road mobile, non-utility point sources, and biogenic sources were modeled deterministically using the values from NC DENR's SIP modeling efforts.

5.3 MAQSIP simulations

We performed two applications of the probabilistic modeling approach using MAQSIP with a Carbon Bond 4 (CB4) configuration, the grid shown in Figure 7, and 12 model layers. The vertical structure with a nonhydrostatic sigma-P configuration with and model top at 100 mbar and a sigma level configuration of 1.0, 0.995, 0.987, 0.974, 0.956, 0.936, 0.913, and 0.

In the first application, 50 realizations were evaluated for a case in which uncertainties in utility NOx emissions were evaluated using a time-series analysis approach. It was assumed that there was no spatial correlation between stack emissions (including those at the same plant). In the second application of 50 realizations, spatial correlations were considered. Each evaluation required approximately 5 days to carry out using the 9 computers that were available within the Condor network. The outputs were then compiled and analyzed to evaluate how considering uncertainties in utility emissions affected the likelihood of achieving the 1- and 8-hour ozone standards.

6 RESULTS

In Figure 8, we provide a plot of the maximum hourly ozone concentrations across the episode for the dependent-uncertainties case. The analysis performed on these results concluded that the dependent case gave significantly different results from those of the independent case. Since the dependent case is more refined and the results were substantially different from the independent case, we have concluded that including the uncertainty dependencies in such analyses is necessary. In addition, including the dependencies among the plants played a critical role in assessment of the overall uncertainty in ozone for analyses involving the 1-hour ozone standard. Further analysis showed that the range of uncertainty in ozone values solely attributable to utility NO_x emissions can be as large as large as 20 ppb or more. Analysis of the results of attaining the 8-hour ozone standard showed that the uncertainty in the utility NO_x emissions did not have an impact on whether NC was out of attainment or not. The predicted ozone exceedances were so widespread that the variation in utility NO_x emissions did not affect the 8-hour nonattainment designation. These results are presented in full in (Abdel-Aziz and Frey, 2003).

Since a difference of up to 20 ppb has been observed by including the emission inventory uncertainties in utility point sources, we can infer that considering emission uncertainties are quite important. A difference of 20 ppb in predicted ozone concentrations could mean the difference between demonstrating 1-hour ozone attainment and not demonstrating it. Decision makers who need to assess attainment designations could be much better informed by taking such information into consideration.

Additionally, this 20 ppb difference resulted from uncertainties in one of our most certain emissions data. This is because the CEM data are measured on an hourly basis and provided as input to air quality models. Uncertainties in area sources, nonroad sources, on-road mobile sources, non-utility point sources, and biogenic emissions are likely to be even larger. In many cases, emissions from these other sources contribute to the overall prediction of ozone at least as much as do utility NO_x emissions. Therefore, consideration of these uncertainties may lead to even greater ranges of predicted ozone concentrations.

7 CONCLUSIONS

We conclude with recommendations for improving the characterization of uncertainty in emissions inventories, with the goal of making probabilistic assessments more practical. We have demonstrated that a system now exists to include some sources of emission uncertainties in air quality modeling. These uncertainties can have an impact on demonstrating attainment of the 1-hour ozone standard. The uncertainties in the CEM data that we considered in this work are likely to be the smallest of any emissions category, because they are measured emissions values. We anticipate that when we use the system for uncertainties on other source categories, much larger impacts on air quality modeling results will be observed.

Now that we have established a practical approach to including uncertainties in emissions processing, and other work has demonstrated that the impact of including these can be significant, the next reasonable step is gaining more widespread use of such an approach. More widespread use would lead to air quality modelers questioning the sources of uncertainty that cause uncertainty in the outputs of their models. Such questioning would ultimately lead to identifying inventory uncertainties and targeting those inventory improvements that can best reduce uncertainty in air quality model predictions. It would also lead to more informed decision-making about air quality improvements that can be expected by emission changes resulting from emission controls.

A significant roadblock to applying the techniques described here to other cases is the quantification of uncertainties in other regions. We have targeted inventory uncertainties that are specific to the sources that play a major role in the Charlotte air quality modeling. Other regions of the country may have very

different key sources, and emissions from those sources will need to have quantitative assessments of uncertainties to be able to use our approach. Those efforts are likely to find the same problems with data collection that was found in this project. To help ameliorate this problem in the future, all data collection efforts undertaken to create revised emission factors and activity totals should be performed in such a way that the data can be used for analysis of data uncertainties. This includes developing specifications that result in the recording of information to describe what data were collected, how the data were collected, and how the data were documented, stored, and made available to the public. If the data can be collected and maintained in such a way that facilitates calculation of uncertainty estimates about the data, an uncertainty-based approach in modeling applications will be much more practical in the future.

Finally, we have not included all emission modeling uncertainties in this effort. The uncertainties associated with temporal allocation, chemical speciation, emission controls, and computation of layer fractions could be just as large if not larger than the uncertainties considered in this work. Once the use of the uncertainty-enabled SMOKE system becomes more widespread and the benefits of including uncertainties in air quality modeling more widely realized, it may be useful to make additional modifications to the system to include these other uncertainties. In addition, the existing system is based on MOBILE5b, which has since been replaced by MOBILE6 in newer versions of SMOKE. The two versions should be integrated to make uncertainty analysis using MOBILE6 possible as well. Also, the uncertainties in biogenic emissions, which play a large role in ozone formation in the Eastern U.S., should be included to allow air quality modelers to establish a more realistic picture of the range of impact biogenic emissions have on air quality models.

8 REFERENCES

Abdel-Aziz, A., and H.C. Frey, "Quantification of Variability and Uncertainty in Hourly NOx Emissions from Coal-Fired Power Plants," *Proceedings, Annual Meeting of the Air & Waste Management Association*, Pittsburgh, PA, June 2002.

Amr Abdel-Aziz and H. Christopher Frey, "*Propagation of Uncertainty in Hourly Utility NO_x Emissions Through a Photochemical Grid Air Quality Model: A Case Study for the Charlotte, NC Modeling Domain*," submitted to Environmental Science and Technology, 2003.

Briggs, G. A., Plume rise and buoyancy effects, in *Atmospheric Science and Power Production*, edited by D. Randerson, *DOE/TIC-27601 (DE84005177)*, Off. of Sci. and Tech. Inf., U.S. Dep. of Energy, Oak Ridge, Tenn., 1984.

Coats, C. J., Jr., and M. R. Houyoux, Fast emissions modeling with the Sparse Matrix Operator Kernel Emissions modeling system, paper presented at *The Emission Inventory: Key to Planning, Permits, Compliance, and Reporting*, Air and Waste Manage. Assoc., New Orleans, La., Sept. 4-6, 1996.

Houyoux, M. R.; Vukovich, J. M.; Coats, C. J. Jr.; Wheeler, N. M.; Kasibhatla, P. S. "Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project", *J. Geophys. Res.*, 2000, 105, D7, pp 9079-9090.

Houyoux, M. R., C. J. Coats Jr., A. Eyth, and S. Lo, Emissions modeling for SMRAQ: A seasonal and regional example using SMOKE, paper presented at Computing in Environmental Resource Management, Air and Waste Manage. Assoc., Research Triangle Park, N. C., Dec. 2-4, 1996.

Frey, H.C., and S. Bammi, "Quantification of Variability and Uncertainty in Lawn and Garden Equipment NO_x and Total Hydrocarbon Emission Factors," *Journal of the Air & Waste Management Association*, 52(4):435-448 (April 2002).

Frey, H.C., and S. Li, "Quantification of Variability and Uncertainty in Natural Gas Engine Emission Factors," *Journal of the Air & Waste Management Association*, accepted pending minor revisions.

Frey, H. C., and J. Zheng, "Quantification of Variability and Uncertainty in Utility NO_x Emission Inventories," *J. of Air & Waste Manage. Association*, 52(9):1083-1095 (September 2002).

Frey, H.C., and J. Zheng, "Probabilistic Analysis of Driving Cycle-Based Highway Vehicle Emission Factors," *Environmental Science and Technology*, 36(23):5184-5191 (December 2002).

Grell, G. A., J. Dudhia, and D. R. Stauffer, A description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), *Rep. NCAR/TN-389+STR*, 138 pp., Natl. Cent. for Atmos. Res., Boulder, Colo., 1994.

Odman, M. T., and C. L. Ingram, Technical report: Multiscale Air Quality Simulation Platform (MAQSIP): Source code documentation and validation, *Rep. ENV-96TR002-v1.0.*, 83 pp., MCNC Environ. Programs, Research Triangle Park, N. C., 1996.

Pierce, T., C. Geron, L. Bender, R. Dennis, G. Tonnesen, and A. Guenther, Influence of increased isoprene emissions on regional ozone modeling, *J. Geophys. Res.*, *103*, 25,611-25,629, 1998.

U.S. Environmental Protection Agency (EPA), User's guide to MOBILE5 (Mobile Source Emission Factor Model), *Rep. EPA-AA-AQAB-94-01*, 174 pp., Off. of Air and Radiat./Off. of Mobile Sources, Ann Arbor, Mich., May 1994.

9 ACKNOWLEDGEMENTS

We acknowledge the EPA's STAR Grant program that has provided the funding for this work.

10 TABLES

Table 1: Average daily emissions and locations for power plants

Name	State	Location (Row, Col)	Average Daily Emissions (t/d)
Allen	NC	(23, 33)	60.4
Buck	NC	(38, 45)	24.4
Cliffside	NC	(22, 16)	50.5
Belews Creek	NC	(55, 50)	286.9
Dan River	NC	(61, 57)	15.4
WS Lee	SC	(6, 2)	22.2
Marshall	NC	(34, 33)	145.4
Riverbend	NC	(27, 33)	24.9
HB Robinson	SC	(3, 54)	10.8

11 FIGURES

Figure 1: SMOKE emissions processing steps for area and point sources

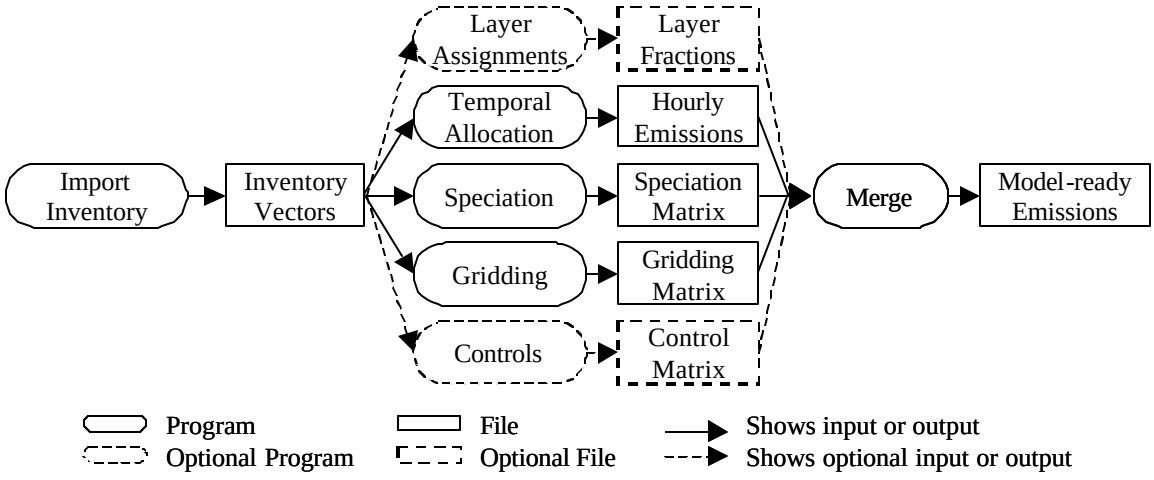


Figure 2: SMOKE processing steps for on-road mobile sources

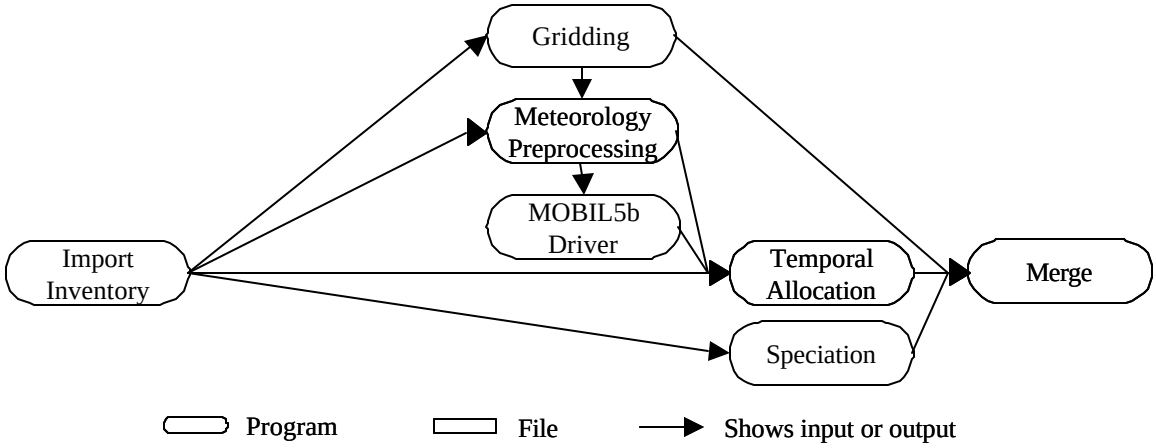


Figure 3: Example of an AUNCERT file

```
/FACTOR ASSIGNMENT/
# CFIP  SCC          CPOL  METH  NDIST        APPROACH
000000  2301030000   ROG   P     SOLUTILMISC  S
000000  2301040000   ROG   P     PETROLSTOR   S
000000  2401001000   ROG   E     ONE          ST
000000  2461021000   ROG   E     TWO          ST
000000  2465900000   ROG   E     THREE        ST
/END/

#            NDIST         TYPE  NUMP   PARM1    PARM2
/PARAMETRIC/ CBASPHALT     N     2      1.000    0.240
/PARAMETRIC/ OPENBURN      L     2      0.003    0.353
/PARAMETRIC/ MISCPRODS     N     2      1.000    0.040
/PARAMETRIC/ ARCHCOAT      N     2      1.000    0.058
/PARAMETRIC/ SOLUTILMISC   N     2      1.000    0.100
/PARAMETRIC/ SFCCOATWOOD   G     2     24.560    0.410
/PARAMETRIC/ GASSERVSTAT   N     2      1.000    0.140
/PARAMETRIC/ PETROLSTOR    N     2      1.000    0.150
/PARAMETRIC/ WASTEDISP     N     2      1.000    0.200

#           TYPE    NDIST
/EMPIRICAL/ S       ONE
#     EFVAL PROB
0.83    0.001
0.84    0.001
0.84    0.001
0.85    0.001
(more removed for brevity in document)
/END/
```

Figure 4: SMOKE programs and files for processing with uncertain inventories for area sources

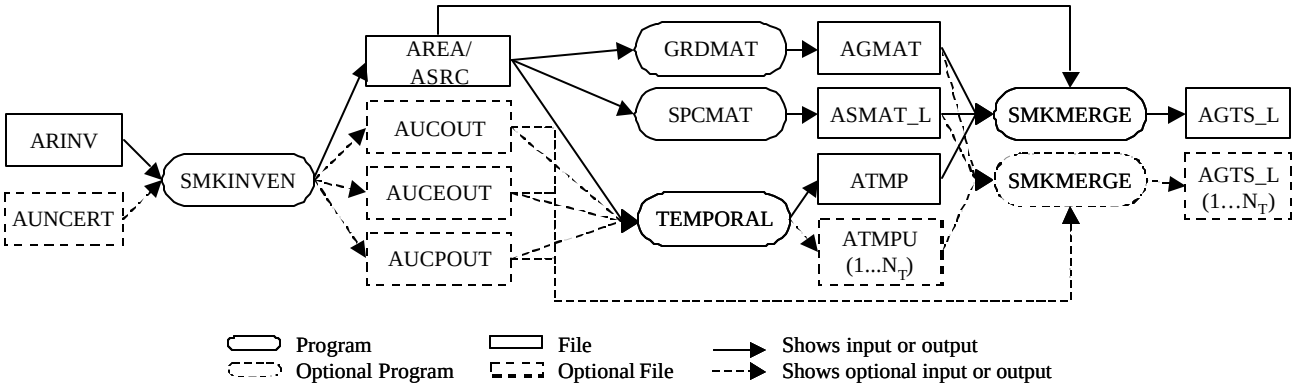


Figure 5: SMOKE programs and files for processing with uncertain inventories for point sources

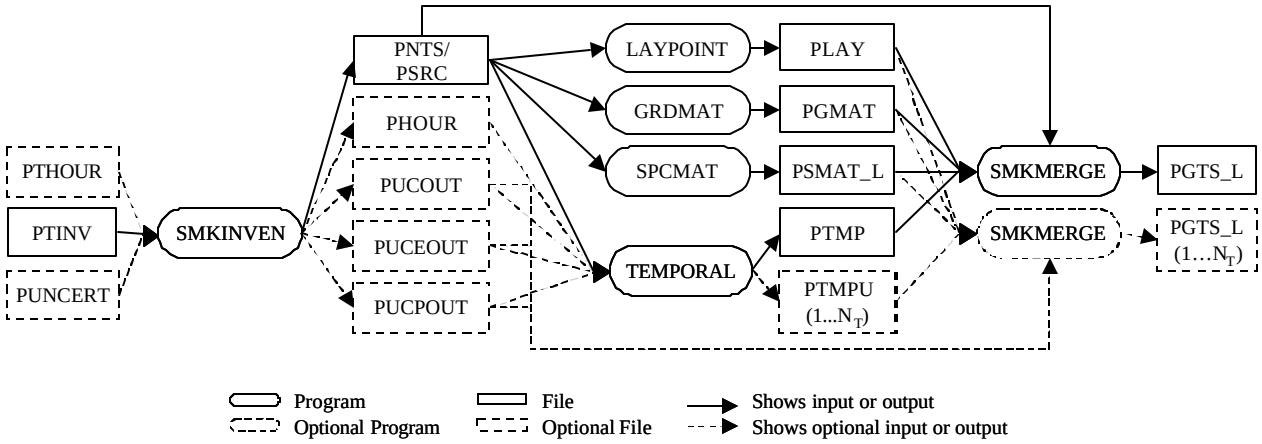


Figure 6 SMOKE programs and files for processing with uncertain inventories for on-road mobile sources

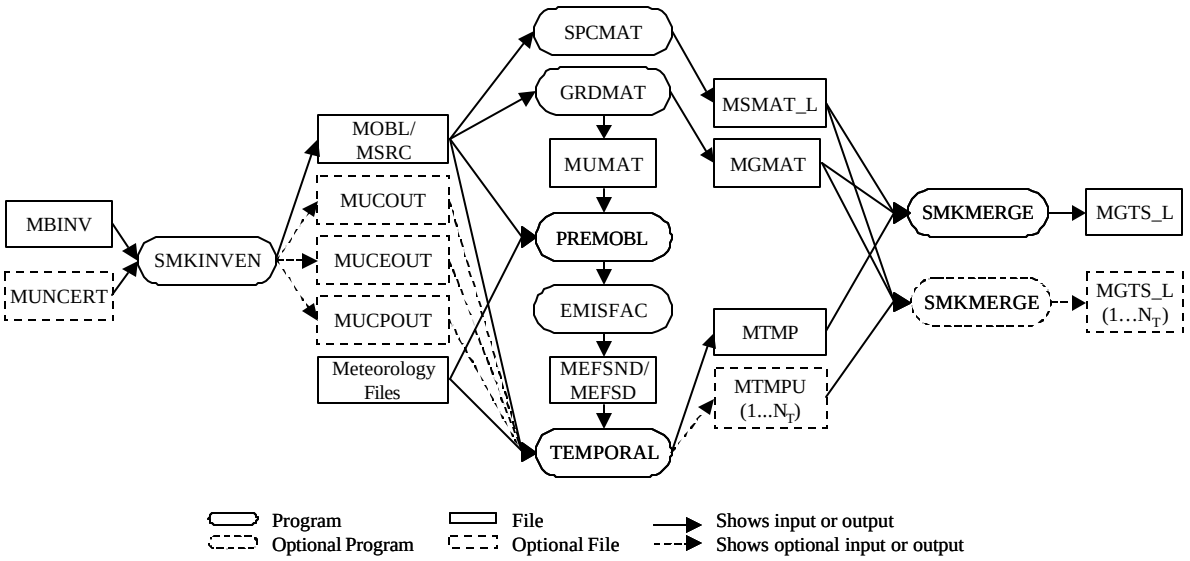


Figure 7: Charlotte test domain

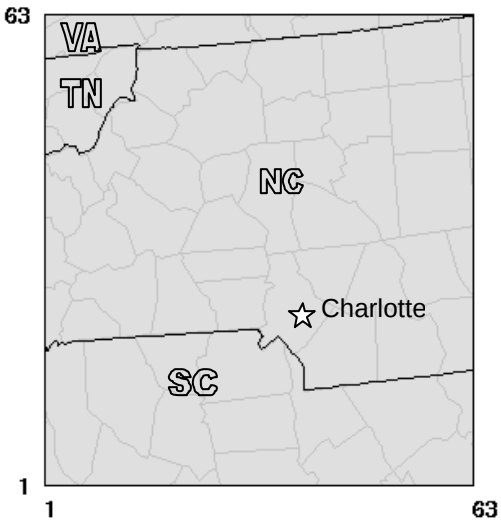
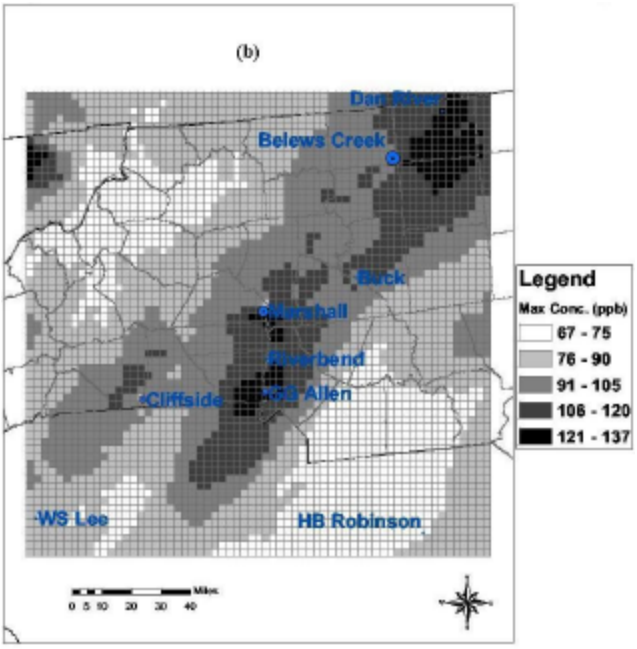


Figure 8: Maximum hourly ozone levels at each grid cell for dependent case



KEYWORDS
SMOKE
emissions modeling
uncertainty
Monte Carlo simulation
air quality modeling