Uncertainty Apportionment for Air Quality Forecast Models

Haiyan Cheng
Virginia Polytechnic Institute and State University
Computer Science Department
Blacksburg, Virginia, USA
hcheng04@vt.edu

Adrian Sandu
Virginia Polytechnic Institute and State University
Computer Science Department
Blacksburg, Virginia, USA
sandu@cs.vt.edu

ABSTRACT

Effective environmental protection policy making depends on comprehensive and accurate Air Quality Model (AQM) prediction results. The confidence level associated with the model prediction, as well as the uncertainty sources that contribute to the prediction uncertainty are important information that should not be neglected when interpreting simulation results. In this work, we explore the capability of the polynomial chaos (PC) method for uncertainty quantification (UQ) and propose a uncertainty apportionment (UA) approach that can be easily applied to any forecast models. The numerical tests on the STEM (Sulfur Transport Eulerian Model) for the northeast region of the United States provide a categorization for the major uncertainty sources that contribute to the uncertainty in the ozone concentration prediction. This information can be used to guide the optimal investment decisions as to which input measurement accuracy should be improved to make the maximum impact on reducing the uncertainty in the prediction result.

Categories and Subject Descriptors

I.6 [**Simulation and Modeling**]: Applications

Keywords

Uncertainty apportionment, Polynomial chaos, Collocation, STEM.

1. INTRODUCTION

To reduce the air pollution and protect the environment is one of the most important problems facing modern society. An effective environmental protection policy depends on accurate AQM forecast results and proper interpretations to the results. Although current AQM has adequately described the physical and chemical processes in the atmosphere, due to the limitations of human knowledge, uncertainties in parameters, initial and boundary conditions, assuming a perfect model is not realistic. Uncertainty quantification is the process to determine the degree/level of the doubt about some results. The outcome of the UQ analysis usually include added "error bars" to the original model forecast results indicating the upper and lower bounds of the uncertainties. More complete UQ information can be described by probability density functions (PDF) of the uncertainties, where the statistical higher moments can be computed, and the integrations over the specific area provide confidence level information. The UQ results can be used in both decision making and as a guideline to reduce uncertainties in order to improve accuracy of the model results.

In recent years, researches about UQ techniques become more and more attractive. Scientists seek effective ways to represent uncertainties, propagate them through the simulation, and analyze the causes and effects of the uncertainties. So far, the statistical representation of uncertainties remain popular, with the traditional Monte Carlo (MC) method [1] being the easiest to implement. MC generates an ensemble of random realizations of each uncertain parameter drawn from a prescribed distribution. The deterministic solvers are applied to each member to obtain an ensemble of results. Post-processing the ensemble results produces the probability density function (PDF), as well as the statistical mean and variance. MC is computationally expensive and inefficient due to its slow convergence rate ($1/\sqrt{n}$). However, with the improvement of the computation power, the MC result can always be used as a reference solution for other UQ approaches.

The polynomial chaos (PC) method has gained popularity in recent years as an effective UQ technique. By representing uncertainties with spectral approximations, PC is suitable for large degree of uncertainties for a relative small number of uncertainties. PC originates from the Wiener [2] homogeneous chaos that represents stochastic processes using Gaussian random variables. It was further developed by Ghanem and co-workers [3] in the finite element context. Karniadakis and Xiu [4] generalized and expanded the concept by matching the probability distribution of the random inputs with the PC expansion basis to reach the optimal exponential convergence rate. In recent years, PC has been successfully applied in many numerical models, however, most of the UQ for AQMs are currently implemented using the costly MC or improved MC with Latin Hypercube sampling [5, 6].

In this paper, we explore PC applied to AQMs for UQ purpose. We use the regional large-scale Sulfur Transport Eulerian Model (STEM-III) [7] to simulate the atmospheric chemical transport, advection, diffusion and reactions. Fol-

SAC'09 March 8-12, 2009, Honolulu, Hawaii, U.S.A.

lowing environmental expert's experiences, we model the main sources of uncertainties in the AQM, namely, the emissions, the boundary conditions, and the deposition velocities. We investigate the effect of those uncertainties on the predicted concentrations of the chemical species of interest, specifically the ozone. We propose a novel approach - uncertainty apportionment (UA) to attribute the uncertainties in the predicted ozone concentrations to specific input uncertainties. The paper is organized as follows: Section 2 presents the PC method and the UA approach. Section 3 shows the numerical test result on the STEM model. Conclusions are drawn in section 4.

2. UNCERTAINTY APPORTIONMENT WITH POLYNOMIAL CHAOS METHOD

Using Wiener's homogeneous chaos representation [2], a general second-order random process can be represented as:

$$X(\theta) = \sum_{i=0}^{\infty} a^i \, \Phi^i(\xi(\theta)),$$

in which θ is a random event, $\Phi^i(\xi(\theta))$ are polynomial functionals defined in terms of the multi-dimensional random variable $\xi(\theta) = (\xi_1(\theta), \ldots, \xi_d(\theta))$ with the joint probability density function of $w(\xi)$. a^i are the corresponding coefficients. The family $\{\Phi^i\}$ satisfies the orthogonality relations:

$$\langle \Phi^i, \Phi^j \rangle = 0 \quad \text{for} \quad i \neq j, \tag{1}$$

where the inner product on the Hilbert space of random functionals is the ensemble average $\langle \cdot, \cdot \rangle$:

$$\langle f, g \rangle = \int f(\xi)\, g(\xi)\, w(\xi) d\xi. \tag{2}$$

If the uncertainty sources are modeled as independent random variables $\xi = (\xi_1, \cdots, \xi_d)$ with a joint probability distribution $w(\xi) = w^{(1)}(\xi_1) \cdots w^{(d)}(\xi_d)$ (d is the number of uncertain parameters), then a multi-dimensional orthogonal basis is constructed from the tensor products of one-dimensional polynomials $\{P_m^{(k)}\}_{m \geq 0}$ orthogonal with respect to the density $w^{(k)}$ [8]:

$$\Phi^i(\xi_1, \cdots \xi_d) = P_{i_1}^{(1)}(\xi_1) P_{i_2}^{(2)}(\xi_2) \cdots P_{i_d}^{(d)}(\xi_d).$$

In this case, the evaluation of d-dimensional scalar products is reduced to d independent one-dimensional scalar products. In practice, we consider a truncated PC expansion with S terms [3]:. With the number of random variables denoted by d, and the maximum degree of the polynomials by p, S can be computed by:

$$S = \frac{(d+p)!}{(d\,!\,p\,!)}. \tag{3}$$

With the growth of the polynomial order and the number of the random variables, the total number of terms in the expansion increases rapidly. This is a major limitation of the PC method. For a simple ODE deterministic model:

$$y(t)' = f(y(t)), \qquad t^0 \leq t \leq T, \qquad y(t^0) = y^0, \tag{4}$$

with the uncertain states represented by PC expansion:

$$y(t) = \sum_{i=1}^{S} a^i(t) \Phi^i(\xi), \tag{5}$$

we generate the stochastic system where the uncertainties are embedded in the coefficients $a^i(t)$:

$$\sum_{i=1}^{S} (a^i(t))' \Phi^i(\xi) = f\left(\sum_{i=1}^{S} a^i(t) \Phi^i(\xi) \right). \tag{6}$$

Both an intrusive Galerkin and a non-intrusive collocation approaches can be used to solve for the coefficients [9]. Without the requirement to modify the deterministic model, the non-intrusive collocation approach provides an easier implementation with similar accurate results compared with the Galerkin approach.

In general, the PC collocation method includes the following steps: (1) Model the sources of uncertainty by random variables with appropriate PDFs. (2) Build the S orthogonal polynomials (expansion basis). (3) Generate the PC expansion of the uncertain parameters (or uncertain initial conditions). (4) Select collocation points. (5) Run an ensemble of the deterministic system simulations with the expansion obtained from (3) on each collocation point. (6) Recover the PC coefficients of the results by formulating and solving the linear equation systems. (7) Extract the mean and the standard deviation of the final solution, and generate the PDFs. In this procedure, the majority of the computation time will be spent on step (5) for repeated deterministic system runs.

After the coefficients are recovered, the system solution can be represented by a linear combination of stochastic coefficients times basis functions. After rearranging the terms from lower orders to higher orders, the component ℓ of the state solution is:

$$\begin{aligned} y_\ell(t) \;=\; & \underbrace{a_\ell^0(t)}_{\text{0th order term}} \\ & + \underbrace{a_\ell^1(t)\Phi^1(\xi_1) + a_\ell^2(t)\Phi^2(\xi_2) + \ldots + a_\ell^d(t)\Phi^d(\xi_d)}_{\text{linear order terms}} \\ & + \; H.O.T. \quad 1 \leq \ell \leq n. \end{aligned} \tag{7}$$

In the above representation, the superscripts for the system coefficients represent the stochastic modes. $\Phi^1(\xi_1), \ldots, \Phi^d(\xi_d)$ are linear polynomials in variables $\xi_1, \ldots, \xi_d$ respectively. The H.O.T. represents the terms of orders 2 and up, which include the higher order for ξs and the cross terms.

From the PC representation we can derive the statistics of the output uncertainty. The mean value is given by the 0^{th} order term in the stochastic representation (7):

$$\langle y_\ell(t) \rangle = \overline{y_\ell(t)} = a_\ell^0(t), \quad 1 \leq \ell \leq n. \tag{8}$$

The variance is computed as:

$$\begin{aligned} s_\ell^2(t) \;&=\; \Big\langle y_\ell(t) - \langle y_\ell(t) \rangle, y_\ell(t) - \langle y_\ell(t) \rangle \Big\rangle \\ &=\; \sum_{i=1}^{S} a_\ell^i(t) a_\ell^i(t) \langle \Phi^i, \Phi^i \rangle, \quad 1 \leq \ell \leq n. \end{aligned} \tag{9}$$

The standard deviation is thus $s_\ell(t)$. The covariance matrix of the model state at any time is computed by:

$$R_{k,\ell}(t) = \langle y_k(t), y_\ell(t)\rangle = \sum_{i=1}^{S} a_k^i(t) a_\ell^i(t) \langle \Phi^i, \Phi^i\rangle.$$

With the PC representation of the system states, we can perform the uncertainty apportionment (UA) by considering only the first order representations in (7) and attribute the corresponding uncertainties in the source uncertainty towards the uncertainty in the model result. The apportionment related researches, such as the Ozone Source Apportionment Technology (OSAT) [10, 11] and the receptor-oriented source apportionment [12, 13] are performed on the mean value of the system state, while here, the apportionment is performed on the variance of the state vector. The proposed UA approach is carried out based on the PC representation of the stochastic states (7). In the variance formula (9), we separate the terms corresponding to the linear terms from the higher order terms:

$$s_\ell^2 = \sum_{i=1}^{d} (a_\ell^i)^2 \langle \Phi^i, \Phi^i\rangle + \sum_{i=d+1}^{S} (a_\ell^i)^2 \langle \Phi^i, \Phi^i\rangle. \qquad (10)$$

$(a_\ell^i)^2 \langle \Phi^i, \Phi^i\rangle$ in the linear portion is the part of the total variance s_ℓ^2 that can be attributed to the i-th source of uncertainty (modeled by variable ξ_i). The higher order terms represent the mixed contribution resulting from the interaction of multiple sources.

Using this feature of the PC representation, we not only compute the total variance, but also assess the percentage of the total variance coming from each individual sources. The numerical test results are described in the following section.

3. NUMERICAL TESTS ON STEM

STEM is a large-scale state-of-the-art regional AQM. It is used to predict the chemical concentrations at future times by solving the following material balance equations:

$$\begin{aligned}
\frac{\partial c_i}{\partial t} &= -u\cdot\nabla c_i + \frac{1}{\rho}\nabla\cdot(\rho K \nabla c_i) + \frac{1}{\rho} f_i(\rho c) + E_i, \\
c_i(t^0, x) &= c_i^0(x), \\
c_i(t, x) &= c_i^{IN}(t, x) \qquad \text{for} \quad x \in \Gamma^{IN}, \\
k_{nn}\frac{\partial c_i}{\partial n} &= 0 \quad \text{for} \quad x \in \Gamma^{OUT}, \\
k_{nn}\frac{\partial c_i}{\partial n} &= V_i^{dep} c_i - Q_i \quad \text{for all} \quad 1 \le i \le n.
\end{aligned}$$

The details about the STEM can be found in [7]. We are especially interested in the forecasted ozone (O_3) concentration. We would like to quantify the uncertainty in the predicted ozone concentration and assess the impact to the forecast uncertainty coming from different uncertainty sources.

The current STEM model is deterministic, that is, it uses one emission profile and assumes no errors in the parameters and inputs, as well as no errors in the representation of physical processes. However, this is not true in reality. Based on environmental experts' experiences [14] about the most important uncertainty sources for atmospheric model, we choose seven independent random variables $\xi_1, \ldots, \xi_7$ (11) to explicitly model the major uncertainties that are associated with the model. Although other factors are important, such as the chemical reaction coefficients, they are not characterized in our test case. Our test case is designed to show the power of UQ and UA approaches. It is easy to be modified to reflect other interesting cases. For a complete UQ analysis, please refer to [14].

$$\begin{aligned}
\xi_1 &- NO_x \text{ ground em. } (NO, NO_2) \quad (-20\% - +20\%) \\
\xi_2 &- AVOC \text{ ground em. } (HCHO, ALK, OLE, ARO) \\
&\quad (-50\% - +50\%) \\
\xi_3 &- BVOC \text{ ground em.} \\
&\quad \text{(ISOPRENE, TERPENE, ETHENE)} \\
&\quad (-40\% - +40\%) \\
\xi_4 &- \text{Dep. veloc. for } O_3 \quad (-50\% - +50\%) \\
\xi_5 &- \text{Dep. veloc. for } NO_2 \quad (-50\% - +50\%) \\
\xi_6 &- \text{West B.C. for } O_3 \quad (-5\% - +5\%) \\
\xi_7 &- \text{West B.C. for } PAN \quad (-5\% - +5\%)
\end{aligned} \qquad (11)$$

To show the capability of modeling non-Gaussian uncertainties with PC, we use the Beta distributed random variables with parameters $a = 1, b = 1$, and the Jacobi order 2 polynomials as PC basis. PC collocation approach with the Hammersley/Halton points are used. As a result, $S = 36$ deterministic model runs are needed.

The computational region used in this study covers the northeastern part of the United States, as shown in Figure 1, with a total size of $1500\,km \times 1320\,km \times 20\,km$. Using the horizontal resolution of $60\,km \times 60\,km$ and a variable vertical resolution, a 3-dimensional grid with $25 \times 22 \times 21$ points is used. The initial conditions, meteorological field data, boundary conditions, and emissions are obtained from the ICARTT (International Consortium for Atmospheric Research on Transport and Transformation) campaign. The data corresponds to the time window from July 20-22, 2004. We run a 48-hour forecast starting at 12 noon GMT (8am EDT) on July 20, 2004. We are interested in the forecasted ozone concentration as well as the associated uncertainties at ground level for the following four major cities: 1. Columbus, Ohio; 2. Washington D.C.; 3. Boston, Massachusetts, and 4. New York City. We compute the total uncertainty associated with the predicted concentrations and attribute the total uncertainty to different sources.

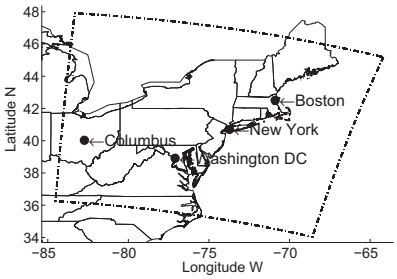
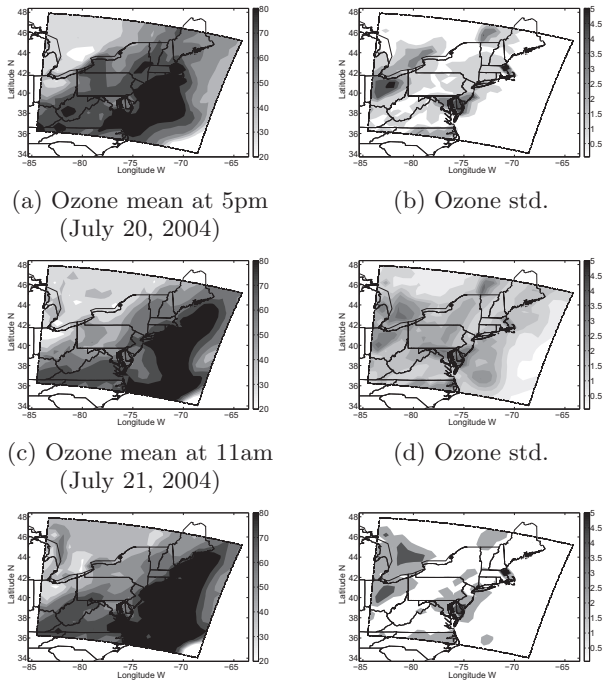


Figure 1: Simulation region and the location of four major cities where uncertainty in ozone predictions will be assessed.

Figure 2(a) shows the predicted ozone concentration mean value after 9 hours, at 5pm (EDT) on July 20, 2004; the

associated standard deviation is shown in Figure 2(b). Predicted ozone concentration mean value after 27 hours (at 11am (EDT) on July 21, 2004) and the standard deviations are shown in Figure 2(c) and Figure 2(d). Predicted mean ozone concentration after 33 hours (at 5pm (EDT) on July 21, 2004) and the standard deviations are shown in Figure 2(e) and Figure 2(f) respectively. The figures show that major cities have large amount of emissions that contribute to the ozone formation. Comparing Figure 2(b), 2(d) and 2(f), we see that during the mid of the day, the ozone uncertainty is spread out, but around evening time, the uncertainties are more clustered at the major cities, and the uncertainty magnitude are larger.



(a) Ozone mean at 5pm (July 20, 2004)

(b) Ozone std.

(c) Ozone mean at 11am (July 21, 2004)

(d) Ozone std.

(e) Ozone mean at 5pm (July 21, 2004)

(f) Ozone std.

Figure 2: Ozone concentration mean and standard deviation at 5pm EDT July 20, 11am EDT July 21, and 5pm EDT July 21, 2004.

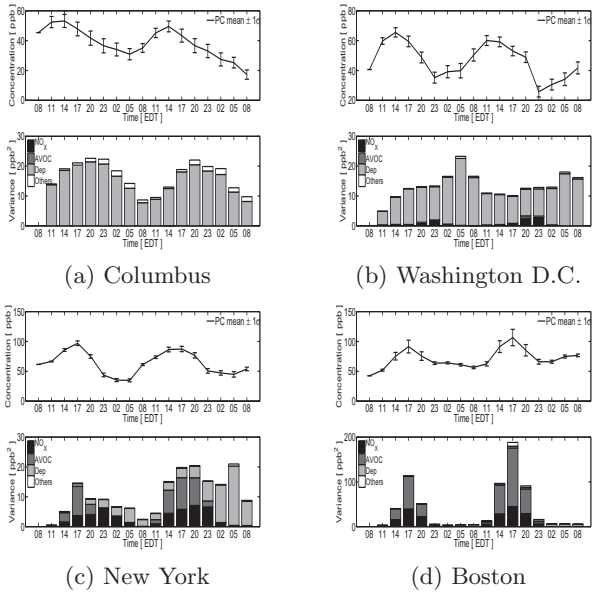
We further analyze the total uncertainties with the UA technique discussed in Section 2. The contribution percentage of each source to the total uncertainty is represented in Table 1 for 5pm, July 20 and 11am, July 21. The "-" in the table represents contributions that are less than 1%. From the table, we observe that NO_x, AVOC emissions, and the deposition velocity uncertainties are the major contributors for the predicted ozone uncertainty. For Columbus and Washington D.C., the predicted ozone uncertainty is mostly affected by the uncertainties in the ozone deposition velocity. While for New York and Boston, NO_x and AVOC emissions contribute significantly. We also observe that AVOC uncertainty contribution increases from around 30% to 60% for both New York and Boston from 11am to 5pm, indicating a significant impact of the traffic emission to the air pollution.

The 48-hour ozone concentration time series and UA results are plotted in Figure 3 for four cities. The upper panels

5pm, July 20					
Cities	NO_x	AVOC	Dep(O_3)	WBC(O_3)	H.O.T
CB	-	-	96%	1%	3 %
DC	3%	1%	96%	-	-
NY	25%	68 %	6%	-	1%
BT	35%	64%	-	-	1%
11am, July 21					
Cities	NO_x	AVOC	Dep(O_3)	WBC(O_3)	H.O.T
CB	-	-	93%	5%	2%
DC	1%	1%-	96%	-	1%
NY	22%	31%	47%	-	-
BT	43%	33%	24%	-	-

Table 1: Ozone uncertainty apportionment. Percentage contributions of different sources for the 4 cities at 5 pm EDT on July 20, 2004 and 11 am EDT on July 21, 2004 ("-" indicates $< 1\%$).

display the time series with the "error bars" indicating the uncertainty magnitude. The size of the "error bars" in the figure is the standard deviation. The prominent two-peak shape indicates that the ozone concentration is low at night and high during the day, reaching the maximum concentration around 5pm in the afternoon. The lower panels of Figure 3 display the uncertainty apportionment results. Since there are mainly four sources that contribute to the total uncertainty, we only plot those sources. Contributions from other sources are all added up and represented by "Others". Figure 3 provides us the categorization of the input uncertainties that impact most to the predicted ozone uncertainty.



(a) Columbus

(b) Washington D.C.

(c) New York

(d) Boston

Figure 3: Ozone concentration for 4 cities over 48 hours (upper) and the uncertainty apportionment for the total variance (lower).

Using PC for UQ provides important insights about the average ozone concentration levels. Using the forecasted ozone concentrations in Boston from 11am to 11pm (first peak in the upper panel of Figure 3(d)), we perform an

interpolation and compute an 8-hour average ozone concentration for the time window from 12 noon to 8pm. The PDF plot in Figure 4 shows the uncertainty information associated with this average concentration. The shaded area in the figure indicates the probability of the 8-hour average ozone concentration in Boston exceeding 75 ppbv. The probability of 8-hour average ozone concentration exceeding 70 ppbv is about 88%, exceeding 75 ppbv is about 68%, exceeding 80 ppbv is about 43%, and exceeding 85 ppbv is about 22%. This visual uncertainty illustration provides a very effective way to communicate uncertainties in the model results to the policy makers, and allow them to make informed decisions.

Furthermore, the UA results provide insights about the impact of different input uncertainties on the model result. This information is essential to guide efforts to reduce the measurement errors. For example, Figure 3 shows that both Columbus and Washington DC. ozone uncertainties are dominated by ozone deposition velocity uncertainty. Investing more to increase the AVOC and NO_X measurement accuracy for these two cities might not have a significant impact on reducing the predicted ozone uncertainty.

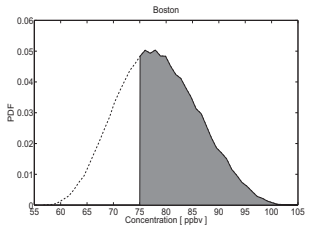


Figure 4: Boston average ozone PDF from 12pm-8pm EDT on July 20, 2004.

4. CONCLUSIONS

We propose a uncertainty apportionment approach based on the polynomial chaos expansion of the system uncertainty. The numerical tests with the STEM model show the ability of the UA technique to further analyze the impact of the input uncertainties towards the total uncertainty in the forecast states. Using PC method, we not only obtain "confidence level" information for the predicted chemical concentration in the air (even for non-Gaussian uncertainties), but also have knowledge about the most important uncertainty sources. We expect that the UA technique provides new insights for making policy decisions regarding environmental protection, because qualified policy decision based on models can be made only when one knows the level of confidence to be attributed to the model results.

5. REFERENCES

[1] G. Fishman. *Monte Carlo: Concepts, Algorithms and Applications*. Springer-Verlag, New York, 1996.

[2] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

[3] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Dover publications, 1991.

[4] D. Xiu and G. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24:619–644, 2002.

[5] P. T. Martien, R. A. Harley, J. B. Milford, and A. G. Russell. Evaluation of incremental reactivity and its uncertainty in southern california. *Environ. Sci. Technol.*, 37(8):1598–1608, 2003.

[6] J. S. Irwin and S. R. Hanna. Characterising uncertainty in plume dispersion models. *International Journal of Environment and Pollution*, 25(1-4):16 – 24, 2005.

[7] G. R. Carmichael, Y. Tang, G. Kurata, I. Uno, D. Streets, J. H. Woo, H. Huang, J. Yienger, B. Lefer, R. Shetter, D. Blake, E. Atlas, A. Fried, E. Apel, F. Eisele, C. Cantrell, M. Avery, J. Barrick, G. Sachse, W. Brune, S. Sandholm, Y. Kondo, H. Singh, R. Talbot, A. Brandy, D. Thorton, A. Clarke, and B. Heikes. Regional-scale chemical transport modeling in support of the analysis of observations obtained during the TRACE-P experiment. *J. Geophys. Res.*, 108(D21-8823):10649–10671, 2003.

[8] A. Sandu, C. Sandu, and M. Ahmadian. Modeling multibody dynamics with uncertainties. Part I: Theoretical and computational aspects. *Multibody System Dynamics*, 15(4):369–391, 2005.

[9] H. Cheng and A. Sandu. Numerical study of uncertainty quantification techniques for implicit stiff systems. In *Proceedings of the 45th annual southeast regional conference*, pages 367 – 372. ACM Press New York, NY, USA, 2007.

[10] A. M. Dunker, G. Yarwood, J. P. Ortmann, and G. M. Wilson. Comparison of source apportionment and source sensitivity of ozone in a three-dimensional air quality model. *Environ. Sci. Technol.*, 36:2953–2964, 2002.

[11] D. S. Cohan, A. Hakami, Y. Hu, and A. G. Russell. Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis. *Environ. Sci. Technol.*, 39:6739–6748, 2005.

[12] S. L. Miller, M. J. Anderson, E. P. Daly, and J. B. Milford. Source apportionment of exposure to volatile organic compounds: I. evaluation of receptor models using simulated exposure data. *Atmospheric Environment*, 36:3629–3641, August 2002.

[13] M. J. Anderson, E. P. Daly, S. L. Miller, and J. B. Milford. Source apportionment of exposure to volatile organic compounds: II. application of receptor models to TEAM study data. *Atmospheric Environment*, 36:3643–3658, August 2002.

[14] M. S. Bergin, G. S. Noblet, K. Petrini, J. R. Dhieux, J. B. Milford, and R. A. Harley. Formal uncertainty analysis of a Lagrangian photochemical air pollution model. *Environmental Science and Technology*, 33(1116-1126), 1999.