Computer Science Technical Report
TR-07-37
Oct 23, 2007

Haiyan Cheng and Adrian Sandu

" Uncertainty Quantification and Apportionment in Air Quality Models using the Polynomial Chaos Method"

Computer Science Department
Virginia Polytechnic Institute and State University
Blacksburg, VA 24060
Phone: (540)-231-2193
Fax: (540)-231-9218
Email: sandu@cs.vt.edu
http://eprints.cs.vt.edu

Uncertainty Quantification and Apportionment in Air Quality Models using the Polynomial Chaos Method

Haiyan Cheng and Adrian Sandu
Virginia Polytechnic Institute and State University,
Computer Science Department,
Blacksburg, Virginia, 24061, USA
hcheng04@vt.edu, sandu@cs.vt.edu

Abstract

Simulations of large-scale physical systems are often affected by the uncertainties in data, in model parameters, and by incomplete knowledge of the underlying physics. The traditional deterministic simulations do not account for such uncertainties. It is of interest to extend simulation results with "error bars" that quantify the degree of uncertainty. This added information provides a confidence level for the simulation result. For example, the air quality forecast with an associated uncertainty information is very useful for making policy decisions regarding environmental protection. Techniques such as Monte Carlo (MC) and response surface are popular for uncertainty quantification, but accurate results require a large number of runs. This incurs a high computational cost, which maybe prohibitive for large-scale models. The polynomial chaos (PC) method was proposed as a practical and efficient approach for uncertainty quantification, and has been successfully applied in many engineering fields. Polynomial chaos uses a spectral representation of uncertainty. It has the ability to handle both linear and nonlinear problems with either Gaussian or non-Gaussian uncertainties.

This work extends the functionality of the polynomial chaos method to Source Uncertainty Apportionment (SUA), i.e., we use the polynomial chaos approach to attribute the uncertainty in model results to different sources of uncertainty. The uncertainty quantification and source apportionment are implemented in the Sulfur Transport Eulerian Model (STEM-III). It allows us to assess the combined effects of different sources of uncertainty to the ozone forecast. It also enables to quantify the contribution of each source to the total uncertainty in the predicted ozone levels.

1 Introduction

The Chemical Transport Models (CTMs) simulate the chemical reactions and transportation of trace species in the atmosphere. With a given initial condition, the chemical concentrations in the air can be computed for later times. The CTMs are driven by meteorological data, such as the temperature, wind fields, etc, from numerical weather prediction models. The deterministic CTMs assume the meteorological data, the boundary conditions, the initial conditions and the model itself are all exact. However, this is not true in reality. The uncertainty in the model and data will lead to uncertainty in the simulation results. An environmental policy decision based on the results without considering the uncertainties will not be an optimal policy. Uncertainty quantification is therefore necessary for air quality simulations. Finding a good approach to represent, quantify, and reduce those uncertainties has become an important task for model developers.

The uncertainty quantification problem is to assess how incomplete representations of the physics, and uncertainties in parameters, initial conditions, and boundary conditions, impact the model results. The uncertainty reduction problem is to use additional information (e.g., from observations) to lower the uncertainty. Several uncertainty quantification techniques have been explored and applied successfully in many engineering research fields. The traditional Monte Carlo method [1, 2] generates an ensemble of random realizations of each parameter drawn from its uncertainty distribution. The deterministic solvers are then applied to each member to obtain an ensemble of results. The ensemble is post-processed to obtain the relevant statistics of the results, such as the mean and standard deviation. The probability density function (PDF) can also be estimated from the ensemble of results. However, Monte Carlo method is computationally expensive due to the slow convergence rate, which is in the order of the inverse square root of the number of runs ($O(1/\sqrt{n})$). For large scale models that require long computational time, it is almost prohibitive. Although techniques such as Latin hypercube sampling [3], the quasi-Monte Carlo (QMC) method [4], and the Markov Chain Monte Carlo methods (MCMC) [5] can improve the Monte Carlo convergence rate, their applications are limited. For a smaller model, the solution obtained by sampling techniques, such as the Monte Carlo method, can always be used as a reference for other techniques.

The polynomial chaos method represents uncertainties with a spectral approximation that allows high order representations. The concept of the polynomial chaos originates from the homogeneous chaos defined by Wiener [6]. Ghanem and co-workers [7] have demonstrated that polynomial chaos is a feasible computational tool for scientific and engineering studies in the finite element context. Karniadakis and Xiu [8] generalized and expanded the concept by matching the probability distribu-

tion of the random input with the polynomial chaos expansion basis to reach the optimal exponential convergence rate. In recent years, the polynomial chaos method has been successfully applied in many numerical models, such as transport in porous media [9], solid mechanics [10, 11], structural mechanics [12], thermo-fluid systems [13], fluid-structure interaction and diffusion problems [14, 15, 16, 17], computational fluid dynamics [18], chemistry [19], and electrochemical microfluid system [20]. However, we are not aware of any applications of polynomial chaos method to quantify uncertainty in air quality models.

In this paper, we explore the polynomial chaos method for uncertainty quantification in CTMs. We apply the polynomial chaos method to the regional large-scale Sulfur Transport Eulerian Model (STEM-III) [21]. We model the main sources of uncertainty in air quality models, namely, the emissions, the boundary conditions, and the wind field. We investigate the effect of those uncertainties on the predicted concentrations of the chemical species of interest, specifically the ozone. We perform the source apportionment to attribute the uncertainties in the predicted ozone concentrations to specific sources. The paper is organized as follows: Section 2 presents the polynomial chaos method and the source uncertainty apportionment approach. Section 3 describes the STEM model and the experimental setup. Section 4 presents the numerical results and discussions.

2 Polynomial Chaos Method and Source Uncertainty Apportionment

Let $(\Omega, \mathcal{F}, P)$ be probability space, where Ω is a sample space, $\mathcal{F}$ a σ-algebra of subsets of Ω, and P a probability measure. A general second-order random process $X(\theta) \in L_2(\Omega, \mathcal{F}, P)$ can be represented as:

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

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 $w(\xi)$. The family $\{\Phi^i\}$ satisfies the orthogonality relations:

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

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

$$\langle f, g \rangle = \int f(\xi)\, g(\xi)\, w(\xi) d\xi.$$

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 $w^{(k)}$ [22]:

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

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 polynomial chaos expansion with S terms. With the number of random variables denoted by d, and the maximum degree of the polynomials by p, S can be computed by [7]:

$$S = \frac{(d+p)!}{(d!\, p!)}. \tag{1}$$

With the growth of the polynomial order and the number of the random variables, the total number of terms in the expansion increases rapidly.

To illustrate the use of polynomial chaos method, consider the deterministic model:

$$y' = f(y), \qquad t^0 \leq t \leq T, \qquad y(t^0) = y^0, \tag{2}$$

with each state a vector of length n. Assume, without loss of generality, that the initial state is uncertain. To express the resulting uncertainty in the state vector y, we expand the state vector along the polynomial chaos basis $\Phi_1(\xi_1, \ldots, \xi_d), \ldots, \Phi_S(\xi_1, \ldots, \xi_d)$:

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

and insert it into the deterministic system (2) to obtain:

$$\sum_{i=1}^{S}(y^i)'\Phi^i(\xi) = f\left(\sum_{i=1}^{S} y^i\Phi^i(\xi)\right). \tag{3}$$

Using the polynomial chaos expansion, the original ODE is replaced by an ODE for the polynomial chaos coefficients, and the uncertainty information is embedded in these coefficients. The evolution equation for the stochastic coefficients can be obtained by either Galerkin or collocation approaches as explained below.

In the Galerkin polynomial chaos approach, we project the system (3) on the space spanned by the orthogonal polynomials, i.e., we take the inner product of (3) with $\Phi^i(\xi)$ and divide $\left\langle \Phi^i, \Phi^i \right\rangle$ at both sides to obtain:

$$\left(y^i\right)' = \frac{1}{\langle \Phi^i, \Phi^i \rangle}\left\langle \Phi^i(\xi), f\left(\sum_{i=1}^{S} y^i\Phi^i(\xi)\right)\right\rangle, \quad 1 \le i \le S. \tag{4}$$

This system has $n \times S$ components and their evolution is coupled. (Each of the n variables in (2) is represented by S polynomial chaos coefficients.)

The Galerkin approach requires the modification (expansion) of the deterministic system to a stochastic system, and the entire stochastic system needs to be evolved in time. Although only one system integration is needed, the computation time is significantly increased. At the end of integration, solving the linear system coefficients causes a computational cost increasing from $\mathrm{O}(n^3)$ in the deterministic case to $\mathrm{O}(n^3S^3)$ for the stochastic case. The S^3-fold increase can make the computation prohibitive [23], especially for large-scale simulation models.

The collocation approach imposes the system (3) to be satisfied at a given set of points μ^j $(1 \le j \le Q)$:

$$\sum_{i=1}^{S}(y^i)'\Phi^i(\mu^j) = f\left(\sum_{i=1}^{S} y^i\Phi^i(\mu^j)\right), \quad 1 \le j \le Q. \tag{5}$$

With matrix A defined using the basis function values at the collocation points

$$A_{j,i} = \Phi^i(\mu^j), \qquad 1 \le i \le S, \quad 1 \le j \le Q, \tag{6}$$

the collocation points in the system state space are:

$$Y^j(t) = \sum_{i=1}^{S} y^i(t)\Phi^i(\mu^j) = \sum_{i=1}^{S} A_{j,i}\, y^i(t), \quad 1 \le j \le Q. \tag{7}$$

With this notation, the equation (5) becomes:

$$(Y^j(t))' = f(Y^j(t)), \quad t^0 \le t \le T, \quad 1 \le j \le Q. \tag{8}$$

These are Q independent integrations of the deterministic system (2) starting from the initial conditions:

$$Y^j(t^0) = \sum_{i=1}^{S} A_{j,i}\, y^i(t^0), \quad 1 \le j \le Q. \tag{9}$$

After integration, we recover the stochastic solution coefficients at the final time T by solving the linear system of equations for y^i:

$$y^i(T) = \sum_{j=1}^{Q} \left(A^{\#}\right)_{i,j} Y^j(T), \quad 1 \le i \le S, \tag{10}$$

where $A^{\#}$ is the pseudo-inverse ($A^{\#} = A^{-1}$ if $Q = S$ and A is invertible).

The criterion for choosing the collocation points is that the system matrix A (6) is not singular or near singular. Several options have been explored in [23], such as the random data sets, the Hammersley/Halton data sets [24, 25], and the sparse grid Smolyak data set [26, 27, 28]. In lower dimensions (small d), Hammersley/Halton and Smolyak data sets generate more accurate results than a randomly generated data set. However, in higher dimensional stochastic spaces (large d), both the Hammersley/Halton and Smolyak data sets have their data points aligned, causing the system matrix A to become singular or near singular. In this case, a randomly generated data set that leads to a non-singular system matrix should be used.

The comparison of the Galerkin PC approach and the collocation PC approach against the traditional Monte Carlo method for uncertainty quantification purpose is made in [23]. Obviously, without the requirement to modify the deterministic model, the non-intrusive collocation approach

provides an easier implementation with similar accurate result compared with the Galerkin approach.

In general, the implementation of 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 polynomial chaos expansion of the uncertain parameters (or uncertain initial conditions). (4) Select $Q \geq S$ collocation points and generate the system matrix A. (5) Run Q deterministic system simulations with the expansion obtained from (3). (6) Recover the polynomial chaos 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:

$$y(t) = y^0(t) + y^1(t)\Phi^1(\xi) + y^2(t)\Phi^2(\xi) + \ldots + y^d(t)\Phi^d(\xi) + H.O.T. \tag{11}$$

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

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 (11):

$$\langle y_\ell(t) \rangle = y_\ell^0(t), \quad 1 \leq \ell \leq n. \tag{12}$$

The variance is computed as:

$$s_\ell^2(t) = \Big\langle y_\ell(t) - \langle y_\ell(t) \rangle, y_\ell(t) - \langle y_\ell(t) \rangle \Big\rangle = \sum_{i=1}^{S} y_\ell^i(t) y_\ell^i(t) \langle \Phi^i, \Phi^i \rangle, \quad 1 \leq \ell \leq n. \tag{13}$$

The standard deviation is thus $s_\ell(t)$.

The covariance matrix of the model state at any time is computed by:

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

Using these measures, the model output can be visualized with an "error bar" representation for the uncertainty. In order to generate the PDF at any time, random samples with an appropriate distribution need to be drawn and plugged into the PC representation of the time-dependent state. With the known coefficients and the random numbers, an ensemble of states can be generated and represend by a PDF. In PC approach, as long as the PC coefficients and basis are known, a large state ensemble can easily be generated to form a smooth PDF curve. However, in Monte Carlo approach, each member of the ensemble states requires a full system run.

With the polynomial chaos representation of the system state, we can perform the Source Uncertainty Apportionment (SUA) by considering only the first order representations in (11) and attribute the corresponding uncertainty in the source uncertainty towards the total uncertainty. We note that in the Ozone Source Apportionment Technology (OSAT) [29, 30], the apportionment is performed on the mean value of the system state, while here, the apportionment is performed on the total uncertainty of the state vector. The proposed source uncertainty apportionment is carried out based on the polynomial chaos representation of the stochastic states (11). In the variance formula (13), we separate the terms corresponding to the linear terms from the higher order terms:

$$s_\ell^2 = \sum_{i=1}^{d} (y_\ell^i)^2 \langle \Phi^i, \Phi^i \rangle + \sum_{d+1}^{S} (y_\ell^i)^2 \langle \Phi^i, \Phi^i \rangle. \tag{14}$$

$(y_\ell^i)^2 \langle \Phi^i, \Phi^i \rangle$ in the linear portion is the part of the total variance s_l^2 that can be attributed to the i-th source of uncertainties (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 polynomial chaos representation, we not only compute the total variance, but also assess the percentage of the total variance coming from each individual source.

We implement the uncertainty quantification and the uncertainty source apportionment with a chemical transport model-the STEM model. The implementation details and the numerical results are discussed in the following sections.

3 Application of the Polynomial Chaos to the STEM Model

3.1 STEM Model

STEM [21] is a large-scale state-of-the-art regional CTM. It is used to predict the chemical concentrations at future times by solving the following material balance equations:

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

The first three parts in the first equation are the advection term (the transportation due to the wind field), the diffusion term (occurs due to the gas particle motion), and the chemical reaction term (causes the change of the chemical concentration). The last term, emission E_i, is the source of the air pollution.

We denote u the wind field vector, K the turbulent diffusive tensor, ρ the air density with the unit $moles/cm^3$, and c_i the mole-fraction concentration of chemical species ($1 \le i \le s$). For a specific chemical species, V_i^{dep} is the deposition velocity, Q_i is the surface emissions rate, and E_i is the elevated emissions rate. The third component $f_i(\rho c)$ in the above equation is the rate of the mole-fraction concentration change due to chemical reactions.

The 3-dimensional computational domain Ω covers a region of the atmosphere. Denote the outward normal vector on each point of the boundary $\partial\Omega$ by $\vec{n}$, the boundary is partitioned into three parts: the inflow, outflow, and ground level portion: $\partial\Omega = \Gamma^{IN} \bigcup \Gamma^{OUT} \bigcup \Gamma^{GR}$, where Γ^{IN} is the set of boundary points where $\vec{u} \cdot n \le 0$ and Γ^{OUT} the set of points where $\vec{u} \cdot n > 0$.

The STEM model is solved using the operator splitting [31] technique to handle the transport and the chemistry successively. The advection terms are solved using a third order one-dimensional upwind finite difference scheme. The diffusion terms are discretized using a second order central difference scheme. The chemical time integrator is implemented using the Kinetic Preprocessor (KPP) [32]. The gas phase chemical mechanism SAPRC-99 [33] is used to account for the gas-phase atmospheric reactions of volatile organic (VOCs) and nitrogen oxides (NO_x). 93 chemical species

(88 variables and 5 constants) and 213 chemical reactions are involved. If the initial concentration, the boundary condition, and the emissions are given, then the solution of the model is uniquely determined [34].

With the STEM model, we can forecast the chemical concentrations at future times. We are especially interested in the concentration of the ozone (O_3) and its precursors, such as Nitric Oxide (NO), Nitrogen Dioxide (NO_2), and Formaldehyde ($HCHO$).

The 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. A stochastic version of the STEM is needed to account for the uncertainties in the parameters and data. Due to the large scale and the complexity of the model, the traditional full Monte Carlo approach is not possible. Based on our knowledge about the most important uncertainty sources and the polynomial chaos method, we develop a stochastic version of the STEM that considers the uncertainties associated with the emissions, the boundary conditions, and the wind field.

To explicitly model the main sources of uncertainty, we introduce nine independent random variables, one for each source of uncertainty:

$$\xi_1 \rightarrow \text{NO ground emissions (NO)} \qquad \xi_2 \rightarrow \text{NO2 ground emission } (NO_2)$$

$$\xi_3 \rightarrow \text{HCHO ground emissions (HCHO)} \qquad \xi_4 \rightarrow \text{West Dirichlet B.C. for ozone (WBC)}$$

$$\xi_5 \rightarrow \text{East Dirichlet B.C. for ozone (EBC)} \qquad \xi_6 \rightarrow \text{South Dirichlet B.C. for ozone (SBC)}$$

$$\xi_7 \rightarrow \text{North Dirichlet B.C. for ozone (NBC)} \qquad \xi_8 \rightarrow \text{Top Dirichlet B.C. for ozone (TBC)}$$

$$\xi_9 \rightarrow \text{Wind field (Wind)}$$

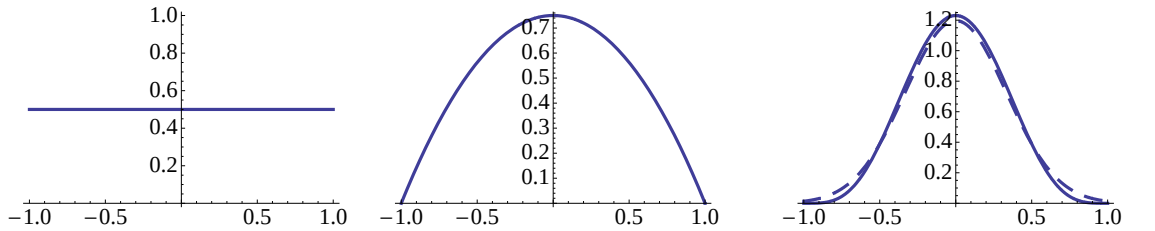
3.2 Experimental Setup

Following the seven-step procedure described in Section 2, we first select the type of the random variables, then derive a corresponding orthogonal basis to achieve the optimal convergence rate. The most common random variables are the Gaussian and uniform distributed random variables. The Gaussian distribution assumes that the majority of the possibilities occur around the mean value, but there might be some extreme cases where the occurrence is nowhere close to the mean value. Since Gaussian PDF has an infinite tail, it does not provide a bound for the uncertainties. Uniform distribution is a special case of the Beta distribution, it assumes that the uncertainty has an equal probability between the minimum and the maximum bound. Uniform distribution does not provide a most likely value. However, uncertainties in air quality models have well defined bounds. For example, all emissions have positive values and can not exceed a maximum value. We have

knowledge of the most likely value. To capture this kind of knowledge we model uncertainty sources in the CTM by Beta distributions with a most likely value information and a finite support. The Beta PDF is defined as follows:

$$w(\xi) = \frac{\Gamma(a+b+2)}{2^{a+b+1}\Gamma(a+1)\Gamma(b+1)} \cdot (1-\xi)^a(1+\xi)^b, \quad \text{where } \Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt. \tag{15}$$

In Figure 1(a), we show the PDF of uniform random variables. The PDF of a Beta function with parameters $a = 1, b = 1$ is shown in Figure 1(b). The comparison of a Beta PDF with parameters $a = 4, b = 4$ (solid) and Gaussian distribution (dashed) is shown in Figure 1(c). In our experiment, we use Beta distributed random variables with parameters $a = 1, b = 1$.



(a) Uniform distribution (a=0, b=0)

(b) Beta distribution (a=1, b=1)

(c) Gaussian distribution (dashed) and Beta distribution (solid) (a=4, b=4)

Figure 1: Probability densities for uniform, Gaussian and Beta distributions.

The polynomial family orthogonal with respect to Beta PDF (15) are Jacobi polynomials. For optimal convergence, we use them as basis of the PC expansion. The normalized Jacobi polynomial basis on $[-1, 1]$ with $a = 1, b = 1$, up to order 2 for one variable are:

$$\begin{aligned} J_0(x) &= \frac{\sqrt{3}}{2}; \\ J_1(x) &= \frac{\sqrt{15}}{2}x; \\ J_2(x) &= \frac{3\sqrt{7}}{4\sqrt{6}}(-1+5x^2). \end{aligned} \tag{16}$$

The 9-dimensional basis set is formed from tensor products of these one-dimensional Jacobi polynomials.

We consider uncertainties in a certain range of the base value, with base value being the most likely. In our experiment, we introduce 30% uncertainty for emissions, 10% for boundary conditions

and 3% for wind field. With the added uncertainties, the deterministic state vectors become the stochastic state vectors (17):

$$\begin{aligned} EM_{NO} &= EM_{NO} + 0.3 \times \xi_1 \times EM_{NO} \\ EM_{NO_2} &= EM_{NO_2} + 0.3 \times \xi_2 \times EM_{NO_2} \\ EM_{HCHO} &= EM_{HCHO} + 0.3 \times \xi_3 \times EM_{HCHO} \\ WBC &= WBC + 0.1 \times \xi_4 \times WBC \\ EBC &= EBC + 0.1 \times \xi_5 \times EBC \\ NBC &= NBC + 0.1 \times \xi_6 \times NBC \\ SBC &= SBC + 0.1 \times \xi_7 \times SBC \\ TBC &= TBC + 0.1 \times \xi_8 \times TBC \\ Wind &= Wind + 0.03 \times \xi_9 \times Wind. \end{aligned} \tag{17}$$

As ξ_1 takes values in [-1,1], the uncertain NO emissions take values in the range of $\pm 30\%$ from the base value.

The stochastic state vector is expanded using the order 2 Jacobi polynomial chaos expansion. Applying (1) with $p = 2$ and $d = 9$, there are $S = 55$ uncertain parameters (expansion coefficients) for each uncertainty state. In order to recover the 55 parameters, we need to run the deterministic system for at least 55 times, each on a different collocation point. The collocation points are randomly generated such that the system matrix A is non-singular.

4 Numerical Results

We implement the polynomial chaos method on STEM for real life simulation. The computational region used in this study covers the northeastern part of the United States, as shown in Figure 2, with a total size of $1500km \times 1320km \times 20km$. Using the horizontal resolution of $60km \times 60km$ 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 modeled

ozone concentration and the associated uncertainty at ground level in 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 ozone concentration and attribute the total uncertainty to different sources.

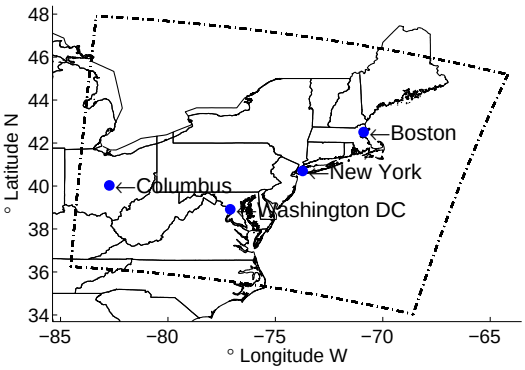
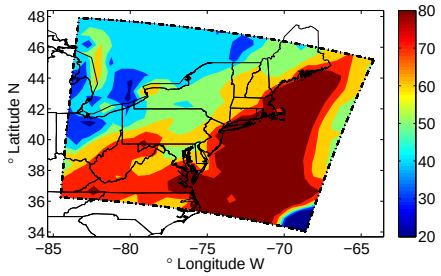


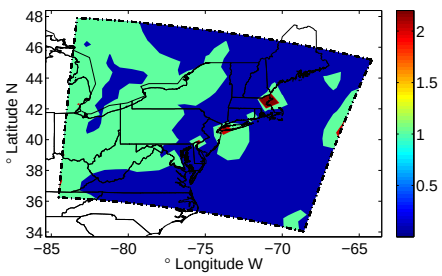
Figure 2: Simulation region and the location of four major cities where uncertainty in ozone predictions will be assessed.

We plot the ozone concentrations at different time. Figure 3(a) shows the predicted mean ozone concentration after 33 hours, at 5pm (EDT) on July 21, 2004; the associated standard deviation is shown in Figure 3(b). Predicted mean ozone concentration after 48 hours (at 8am (EDT) on July 22, 2004) and the standard deviations are shown in Figure 3(c) and Figure 3(d) respectively. From the figures, we can see that due to the wind from the west boundary, the ozone is formed and transported towards the east. After 48 hours, the majority of the ozone is located above the Atlantic ocean. The figures also show that major cities have significant emissions that contribute to the ozone formation. From Figure 3(b) and 3(d), we observe that the uncertainties propagate from the west to the east. Comparing 3(a) and 3(c), we see that ozone concentration is much higher in the afternoon than in the early morning due to the diurnal concentration of the NO_x emissions.

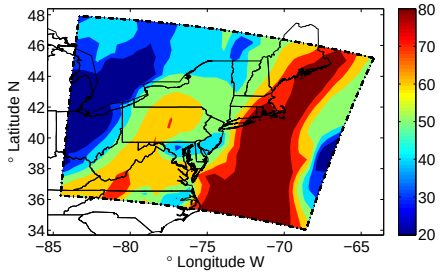
Using the uncertainty apportionment technique discussed in Section 2, the contribution percentage of each source to the total uncertainty is represented in Table 1 for 33 hours and Table 2 for 48 hours. The "-" in the tables represents contributions that are less than 1%. From the table, we observe that the major uncertainty contributors are NO emissions, NO_2 emissions, the west boundary condition, and the wind field. The Columbus's uncertainty is most affected by the uncertainty in the west boundary condition. Boston is most affected by the uncertainty in NO emissions. Comparing the "WBC" (west boundary condition) column for Washington, D.C. between the two tables, the



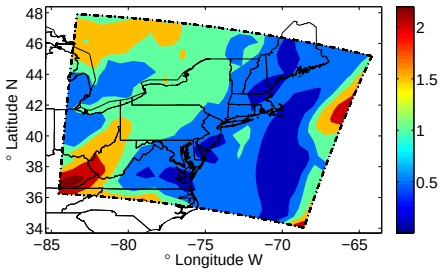
(a) Ozone mean value after 33 hours [ppbv] (5pm EDT on July 21, 2004)



(b) Ozone standard deviation after 33 hours [ppbv]



(c) Ozone mean value after 48 hours [ppbv] (8am EDT on July 22, 2004)



(d) Ozone standard deviation after 48 hours [ppbv]

Figure 3: Ozone concentration mean, and standard deviation at 5pm EDT July 21, 2004 and 8am EDT July 22, 2004. Standard deviation is a measure of the uncertainty magnitude.

Cities	NO	NO_2	$HCHO$	WBC	EBC	SBC	NBC	TBC	Wind	H.O.T
CB	5.3%	-	-	94%	-	-	-	-	-	-
DC	67%	-	-	30%	-	-	-	-	2.8%	-
NY	97%	-	-	-	-	-	-	-	1.1%	-
BT	99%	-	-	-	-	-	-	-	-	-

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

uncertainty contribution from the west boundary conditions has increased from 30% to 84% due to the wind blowing from the west to the east. After 48 hours the west boundary condition uncertainty is being felt on the Atlantic coast.

The 48-hour ozone concentration time series and source uncertainty apportionment are plotted in Figure 4 for four cities. The upper panels display the time series with the "error bars" indicating the uncertainty magnitude. The size of the "error bars" in the figure is two times 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

Cities	NO	NO_2	$HCHO$	WBC	EBC	SBC	NBC	TBC	Wind	H.O.T
CB	-	-	-	99.7%	-	-	-	-	-	-
DC	13%	-	-	84%	-	-	-	-	2.5%	-
NY	55%	-	-	12%	-	-	-	-	32%	-
BT	91%	8.3%	-	-	-	-	-	-	-	-

Table 2: Source uncertainty apportionment. Percentage contributions of different sources for the 4 cities at 8 am EDT on July 22, 2004 ("-" indicates $< 1\%$).

panels of Figure 4 display the source 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". From Figure 4, we see that the uncertainty in the predicted Columbus ozone concentration is dominated by the west boundary condition uncertainty, while for the other three cities, most of their uncertainty comes from the NO emissions. In Figure 4(b), we observe that it takes about 24 hours for the uncertainties in the west boundary condition to have an effect on the Washington, D.C. area.

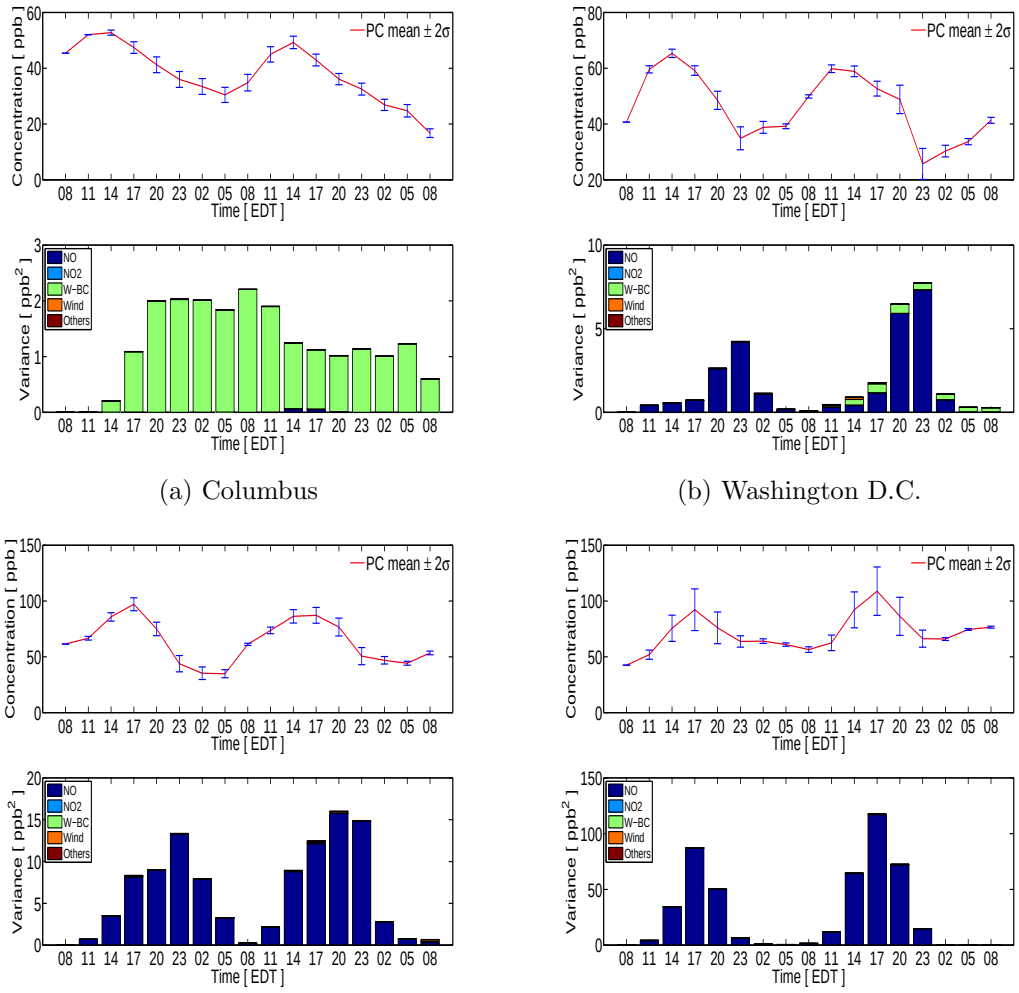
Figure 5 provides a snapshot of the source uncertainty apportionment for 4 cities after 33 and 48 hours. Figure 5(a) corresponds to the second peak in Figure 4. We can see from the plot that the major contributor to uncertainty in Boston, New York and Washington, D.C. are the NO emission. However, the Columbus uncertainty are mainly due to the west boundary condition uncertainty.

We generate the PDF plot at the second peak time with 50,000 sample draws, shown in Figure 6. The PDF plot has a similar shape to the original uncertainty Beta distribution we introduced into the model. This indicates that for this particular case, the uncertainty in the ozone concentration prediction responds nearly linearly to the uncertainties in the parameters. However, the proposed technique can account for fully nonlinear responses which are to be expected for other scenarios.

Using the polynomial chaos method for both uncertainty quantification and apportionment can provide "confidence level" information for air quality forecast. This information is especially useful for making policy decisions regarding environmental protection. We expect the proposed approach will have a positive impact on the environmental policy making, 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 Acknowledgments

This work is supported by the National Science Foundation through the awards NSF CAREER ACI-0413872, NSF ITR AP&IM 0205198. We gratefully acknowledge C. R. Carmichael, T. Chai,



(a) Columbus

(b) Washington D.C.

(c) New York

(d) Boston

Figure 4: Ozone concentration for 4 cities over 48 hours (upper) and the source uncertainty apportionment for the total variance (lower). The error bars show two times of the standard deviation. Only major contributors are shown. With the West boundary conditions dominates the uncertainty in Columbus, OH. The uncertainty in Washington, NY, Boston are mainly due to the uncertainty in NO emissions. The level of uncertainty changes with the diurnal cycle.

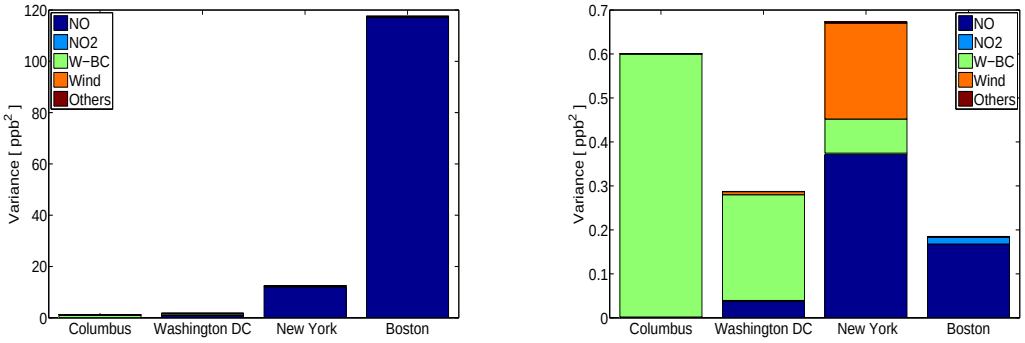
Y. Tang for the data and simulations setup, as well as for many fruitful discussions.

References

[1] G. Fishman. *Monte Carlo: concepts, algorithms and applications*. Springer-Verlag, New York, 1996.

[2] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer-Verlag, 2001.

[3] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting



(a) 5pm EDT on July 21, 2004 (b) 8am EDT on July 22, 2004

Figure 5: Source uncertainty apportionment. Contributions of different sources for the 4 cities at (a) 5pm EDT on July 21, 2004 and (b) 8am EDT on July 22, 2004. (Only the major contributors are shown.)

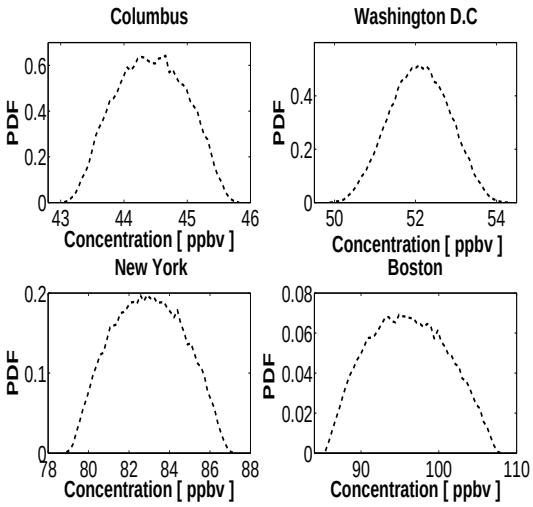


Figure 6: The probability distribution function of ozone concentrations at 5pm EDT on July 21, 2004. This gives a complete characterization of uncertainty in ozone concentration at different locations.

values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):237–240, 1979.

[4] H. G. Niederreiter. Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84(6):957–1041, 1978.

[5] R. L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32:1296–1308, 1984.

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

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

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

[9] R. G. Ghanem. Probabilistic characterization of transport in heterogeneous media. *Computational Methods in Applied Mechanics and Engineering*, 158:199–220, 1998.

[10] R. G. Ghanem. Ingredients for a general purpose stochastic finite element formulation. *Computational Methods in Applied Mechanics and Engineering*, 168:19–34, 1999.

[11] R. G. Ghanem. Stochastic finite elements for heterogeneous media with multiple random non-Gaussian properties. *ASCE Journal of Engineering Mechanics*, 125:26–40, 1999.

[12] R. G. Ghanem, J. Red-Horse, and A. Sarkar. Model properties of a space-frame with localized system uncertainties. In *8th ASCE Specialty Conference of Probabilistic Mechanics and Structural Reliability*, No. PMC200-269, 2000.

[13] O. P. Le Maitre. A stochastic projection method for fluid flow II. random process. *Journal of Computational Physics*, 181:9–44, 2002.

[14] D. Xiu. Efficient collocational approach for parametric uncertainty analysis. *Communications in Computational Physics*, 2(2):293–309, April 2007.

[15] D. Xiu, D. Lucor, C. H. Su, and G. Karniadakis. Stochastic modeling of flow-structure interaction using generalized polynomial chaos. *ASME Journal of Fluid Engineering*, 124:51–59, 2002.

[16] D. Xiu, R. G. Ghanem, and I. G. Kevrekidis. An equation-free, multiscale computational approach to uncertainty quantification for dynamical systems. *IEEE Computing in Science and Engineering Journal (CiSE)*, 7(3):16–23, 2005.

[17] D. Xiu and G. E. Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 191(43):4927–4948, 2002.

[18] L. Mathelin, M. Y. Hussaini, and T. A. Zang. Stochastic approaches to uncertainty quantification in CFD simulations. *Numerical Algorithms*, 38(1-3):209–236, 2005.

[19] M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545–555, 2003.

[20] B. Debusschere, H. N. Najm, A. Matta, O. M. Knio, R. G. Ghanem, and O. P. Le Maitre. Protein labeling reactions in electrochemical microchannel flow: Numerical prediction and uncertainty propagation. *Physics of fluids*, 15(8):2238–2250, 2003.

[21] 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.

[22] 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.

[23] 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.

[24] J. M. Hammersley. Monte Carlo methods for solving multivariable problems. *Annual New York Academy of Science*, 86:844–874, 1960.

[25] J. H. Halton and G. B. Smith. Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702, December 1964.

[26] S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963.

[27] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for elliptic partial differential equations with random input data. Technical Report MOX 85, Politecnico di Milano, June 2006.

[28] C. G. Webster. *Sparse Grid Stochastic Collocation Techniques for the Numerical Solution of Partial Differential Equations with Random Input Data*. PhD thesis, Florida State University, February 2007.

[29] 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.

[30] 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.

[31] P. Csomos. Operator splitting procedure for air pollution transport models. In *Large-scale Scientific Computing (LSSC 2005)*, pages 331–338. Springer-Verlag, 2006.

[32] V. Damian, A. Sandu, M. Damian, F. Potra, and G.R. Carmichael. The Kinetic PreProcessor KPP - a software environment for solving chemical kinetics. *Computers and Chemical Engineering*, 26:1567–1579, 2002.

[33] W. P. L. Carter. Implementation of the saprc-99 chemical mechanism into the model-3 framework. Technical report, United States Environmental Protection Agency, January 2000.

[34] A. Sandu, D. N. Daescu, G. R. Carmichael, and T. Chai. Adjoint sensitivity analysis of regional air quality models. *Journal of Computational Physics*, 204:222–252, 2005.