ELSEVIER

Extension and evaluation of sensitivity analysis capabilities in a photochemical model

S.L. Napelenok [a,*], D.S. Cohan [b], M.T. Odman [c], S. Tonse [d]

[a] *Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration,
In partnership with the United States Environmental Protection Agency, 109 T.W. Alexander Drive, U.S. EPA, Mail Drop E243-01,
Research Triangle Park, NC 27711, USA*
[b] *Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, MS 317, Houston, TX 77005, USA*
[c] *Department of Civil and Environmental Engineering, Georgia Institute of Technology, 311 Ferst Drive, Atlanta, GA 30332, USA*
[d] *Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA*

Abstract

The decoupled direct method in three dimensions (DDM-3D) provides an efficient and accurate approach for probing the sensitivity of atmospheric pollutant concentrations to various changes in photochemical model inputs. The implementation of DDM-3D for the widely used Community Multiscale Air Quality (CMAQ) model was updated to account for recent changes in the base model and to include additional chemical mechanisms and advection schemes. The capabilities of CMAQ–DDM-3D were extended to enable execution using multiple processors in parallel and the computation of sensitivities to chemical reaction rate constants. The resulting direct sensitivity modeling system was tested for statistical agreement with the traditional difference method for calculating sensitivities, considering a summer episode in a domain covering the continental United States. Sensitivities to domain-wide and sector specific emissions, initial/boundary conditions, and chemical reaction rates were compared and found to be in good correlation for both primary and secondary air pollutants. The scalability of CMAQ–DDM-3D to the number of processors used in parallel was also examined. Sensitivity calculations were found to scale in a similar way to the base model, where the benefit to model runtime of adding more processors diminished for simulations that used more than eight processors.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Sensitivity analysis; Decoupled direct method; CMAQ–DDM-3D; Regional atmospheric modeling

1. Introduction

In both regulatory and research communities, it is often useful to determine not only the state of an environmental system, but also its response to perturbations in the various parameters that define it. Sensitivity analysis is an important component of environmental modeling that is able to quantify such responses and is important for both model evaluation and uncertainty analysis (Refsgaard et al., 2007). For air quality management applications, comprehensive modeling systems have been developed in order to track the evolution of pollutants

as these are emitted from sources, chemically altered, transported, and deposited. The Community Multiscale Air Quality Modeling System (CMAQ) is a widely used and continuously updated air quality model with detailed physical and chemical treatments of both gaseous and particulate pollutants using a "one atmosphere" approach (Byun and Schere, 2006). The decoupled direct method (Dunker, 1981; Yang et al., 1997) provides an efficient and accurate approach to probe the sensitivity of atmospheric concentrations to perturbations in model inputs. When compared to the most commonly used technique for calculating sensitivities, the brute-force (difference) method where input parameters are varied one-at-a-time, DDM-3D is less prone to numerical noise particularly for small perturbations. The initial implementation of DDM-3D in CMAQ was described earlier (Cohan et al., 2005;

* Corresponding author. Tel.: +1 919 541 1135; fax: +1 919 541 1379.
E-mail address: napelenok.sergey@epa.gov (S.L. Napelenok).

Napelenok et al., 2006). Here, DDM-3D was updated to be functional within version 4.5.1 of the CMAQ model. The capabilities of CMAQ−DDM-3D were extended to enable parallel process execution and the calculation of sensitivities to chemical reaction rate constants. The performance of new and existing CMAQ−DDM-3D features was then evaluated.

2. Method

2.1. Decoupled direct method in 3D

Photochemical models follow emissions, transport, and chemical transformations of gaseous and particulate pollutants by solving the atmospheric diffusion equation:

$$\frac{\partial C_i}{\partial t} = -\nabla(\mathbf{u}C_i) + \nabla(\mathbf{K}\nabla C_i) + R_i + E_i, \tag{1}$$

where C_i is the concentration of species i, $\mathbf{u}$ is the wind field, $\mathbf{K}$ is the diffusivity tensor, R_i is the net rate of chemical production, and E_i is the emissions rate. DDM-3D efficiently calculates first order sensitivities based on an analogous equation:

$$\frac{\partial S_{ij}}{\partial t} = -\nabla(\mathbf{u}S_{ij}) + \nabla(\mathbf{K}\nabla S_{ij}) + \mathbf{J}S_{ij} + E_i, \tag{2}$$

where S_{ij} is the semi-normalized sensitivity of species i to perturbations in parameter p_j, scaled by its nominal value P_j ($S_{ij} = P_j(\partial C_i/\partial p_j)$; repetition of index j does not imply summation), and $\mathbf{J}$ is the Jacobian matrix, ($J_{ij} = \partial R_i/\partial C_j$) representing the chemical interactions between species. The nominal value P_j is the parameter value used in the base case model simulation. The derivation is presented in more detail elsewhere (Hakami et al., 2003).

CMAQ is a state-of-the-science Eulerian photochemical model widely used for regulatory and research applications. It is developed and maintained by the NOAA Atmospheric Sciences Modeling Division (ASMD) in partnership with US EPA and collaboration from numerous other academic and government agencies (Byun and Schere, 2006). DDM-3D

Table 1
Linear regression analysis of the comparison between first order sensitivities of gaseous species calculated using the brute-force and DDM for Aug 2−9, 2004 (Aug 1st was discarded as a spin-up day.)

Gaseous species (ppb)	Concentration (90%)	Concentration (base case)	Concentration (110%)	Sensitivity[a]
	X, σ	X, σ	X, σ	X, σ
Hourly SO_2 to SO_2 emissions	0.44, 0.11	0.49, 0.12	0.54, 0.13	0.51, 0.14
R^2				**0.998**, 0.00
Slope				0.99, 0.01
Intercept				0.00, 0.00
Hourly NH_3 to NH_3 emissions	0.51, 0.28	0.59, 0.32	0.66, 0.36	0.82, 0.42
R^2				**0.945**, 0.04
Slope				0.94, 0.04
Intercept				−0.01, 0.02
Hourly monoterpene to monoterpene emissions	0.36, 0.24	0.41, 0.27	0.45, 0.30	0.46, 0.31
R^2				**0.999**, 0.00
Slope				0.99, 0.00
Intercept				0.00, 0.00
Hourly NO_2 to NO_x emissions	1.2, 0.7	1.3, 0.8	1.5, 0.8	1.3, 0.7
R^2				**0.977**, 0.02
Slope				0.92, 0.06
Intercept				0.03, 0.04
8-h O_3 to NO_x emissions	37.0, 3.4	37.4, 3.6	37.8, 3.8	4.3, 2.1
R^2				**0.919**, 0.09
Slope				1.00, 0.05
Intercept				−0.20, 0.20
8-h O_3 to mobile NO_x emissions	37.3, 3.6	37.4, 3.6	37.6, 3.7	1.2, 0.6
R^2				**0.913**, 0.10
Slope				1.01, 0.06
Intercept				−0.05, 0.06
8-h O_3 to NO_2 photolysis rate[b]	36.8, 3.5	37.4, 3.6	38.0, 3.8	6.1, 2.1
R^2				**0.985**, 0.01
Slope				1.00, 0.02
Intercept				0.09, 0.05

The presented summary includes a ''cell-by-cell'' comparison for each simulation hour. Brute-force results are regressed against DDM, and the mean, X, and standard deviation, σ, are calculated for the 8 days of the simulation averaged across all grid cells.

[a] Sensitivities are shown to domain-wide emissions from the DDM calculation normalized to a 100% increase in emissions.

[b] From reaction: $NO_2 + h\nu \xrightarrow{k_1} NO$.

has been integrated into the CMAQ model for gaseous (Cohan et al., 2005) and particulate species (Napelenok et al., 2006) according to prior developments of the method (Dunker, 1981; Yang et al., 1997).

Previously, CMAQ–DDM-3D included only the SAPRC99 chemical mechanism (Carter, 2000). To give users more flexibility, the Jacobian matrix for the Carbon Bond 4 (CB4) mechanism (Gery et al., 1989) was also calculated with possibility to include newer versions of these and other chemical mechanisms as they become available and widely used. Likewise, the capability for using the inverse donor-cell advection scheme (Odman and Russell, 2000; Hu et al., 2006) in addition to the piecewise parabolic method (PPM) was included. As a previously unavailable option, direct sensitivity calculation to perturbations of chemical reaction rates was added, following the method of Yang et al. (1997). As the major attraction of using DDM-3D is the convenience of being able to calculate several different sensitivities to various combinations of input parameters, the sensitivity code was made compatible with parallel execution of the CMAQ model.

2.2. Evaluation

The updated CMAQ–DDM-3D implementation was tested for a 9 day episode spanning from August 1 to 9, 2004. Concentrations and sensitivities were calculated on a 36 km resolution grid covering the continental United States ($148 \times 112 \times 14$ cells). The fifth generation mesoscale model (MM5) version 3.6.3 (Grell et al., 1995) was used to develop meteorological fields and the Sparse Matrix Operator Kernel Emissions (SMOKE) version 2.0 (US-EPA, 2004) was used to process emissions. While both CB4 and SAPRC99 mechanisms are fully functional with the current implementation of DDM-3D, the results for only the latter are presented. All simulations were done using PPM advection scheme, RADM cloud module, and EBI chemical solver.

DDM-3D sensitivities were compared to those computed by the central difference method:

$$S_{ij}^{\text{BF}} = \frac{C_i^{110\%j} - C_i^{90\%j}}{0.2}, \tag{3}$$

where $C_i^{110\%j}$ and $C_i^{90\%j}$ represent the concentration of species i resulting from 110% and 90% perturbations in the parameter j, respectively. This difference in concentrations is scaled by the total magnitude of the perturbation (20%) to arrive at semi-normalized traditional difference method or "brute-force" sensitivities that are directly comparable to the sensitivities computed by DDM-3D. It was demonstrated previously that the 20% difference achieved the balance of being appropriately large to avoid numerical noise in the comparison and small enough to capture the local derivate for base-level parameter in the case of emissions sensitivities (Napelenok et al., 2006). PM species were evaluated based on 24-h averages, while both hourly and 8-h running averages were analyzed for gaseous species. Brute-force and direct sensitivities
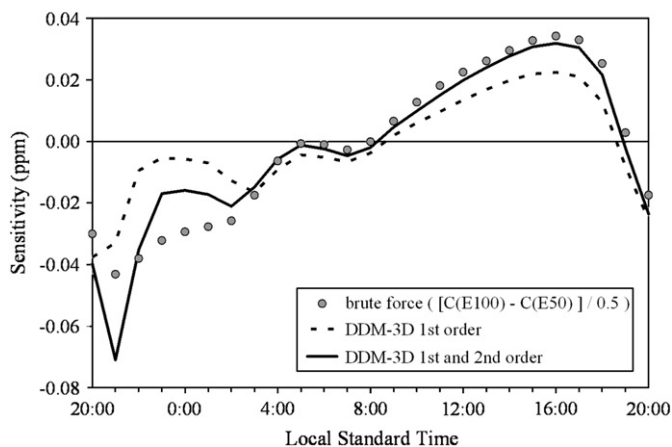

Fig. 1. Hourly ozone sensitivity in Atlanta, GA to domain-wide reduction in NO$_x$ emissions on August 6, 2004. Brute-force sensitivity is computed for a 50% reduction. The initial decrease then increase of brute-force sensitivity, which is due to complex and highly non-linear night-time chemistry in the urban boundary layer including ozone titration by NO$_x$, is not seen with first order DDM. Non-linearities are better captured by higher order DDM-3D.

for all time averages were compared in each cell of the domain for the entire episode using simple linear regression.

First order sensitivities of gaseous species to emissions showed good agreement between the traditional difference method and DDM-3D (Table 1). For all gaseous species, the
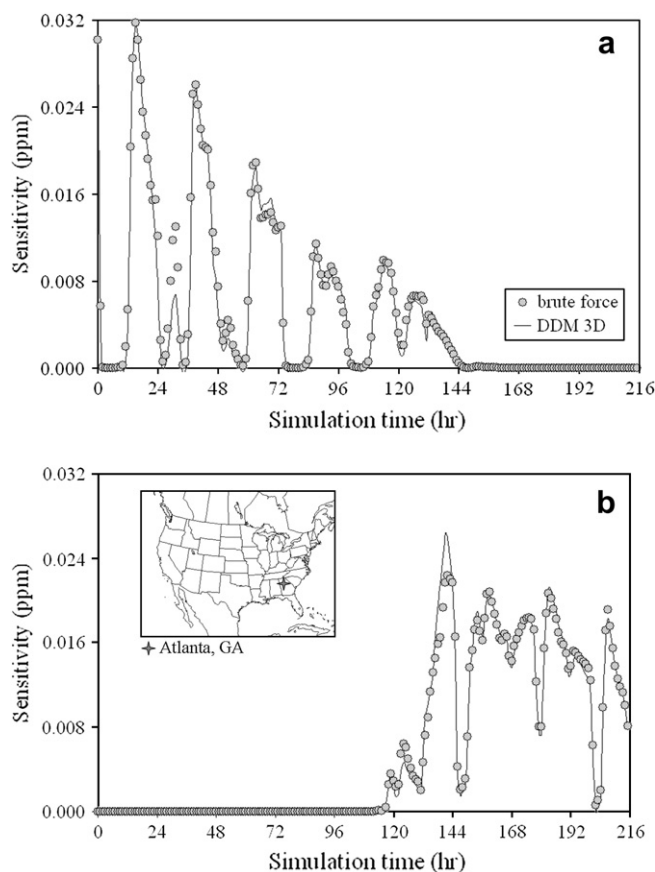

Fig. 2. Hourly ozone sensitivity in Atlanta, GA to initial (a) and boundary (b) conditions of ozone during the Aug 1–9, 2004 episode (ppm). Both initial and boundary conditions of ozone were set to 0.035 ppm for this simulation.

correlation coefficients were near 1.0, the regression slopes were near 1.0, and the intercepts were near 0.0. Sensitivities of primary pollutants to their own emissions had slightly higher correlations (R^2 of 0.998 for SO_2) than those formed through complex chemical interaction (R^2 of 0.919 for O_3 from NO_x emissions), although O_3 sensitivities to NO_x emissions showed very strong performance during daytime hours as reported elsewhere (Cohan et al., 2005). Since DDM-3D captures only the local derivative at the base-level of input parameters, first order sensitivities are not enough to capture large perturbations for non-linear relationships. To better estimate large changes (e.g. a 50% reduction of NO_x), higher order sensitivity functionality (Hakami et al., 2003) was added to the implementation for gaseous species. It was found that including second order sensitivity greatly improves the comparison to brute-force when evaluating ozone formation (Fig. 1). Higher order sensitivities are not yet available for PM species, but work is underway to include these in the future.

CMAQ–DDM-3D is also able to compute sensitivities to initial and boundary conditions, and here, it has been extended to compute sensitivities to chemical reaction rate constants. In order to verify the consistency of direct sensitivities for these perturbations with those calculated using the traditional difference method, sensitivities of ozone to initial and boundary conditions of ozone (Fig. 2) and to the NO_2 photolysis rate were calculated during the episode (Table 1). For Atlanta, it was found that initial conditions can have a lasting impact on concentrations with over 0.024 ppm of ozone attributed as late as the afternoon of the second day of the simulation. This site was significantly far away from the boundaries that it took several days to register the boundary condition sensitivity. Both initial and boundary sensitivities showed good statistical agreement between brute-force and DDM-3D for Atlanta and other locations.

The comparison of sensitivities for individual PM species to emissions was similar to the one performed for the previous version of CMAQ–DDM-3D (Napelenok et al., 2006). Direct relationships between the aerosol species and its gaseous precursor led to better agreement, while indirect relationships with higher degree of non-linearity and susceptibility to numerical error, such as sulfate sensitivity to ammonia, showed a relatively poor comparison (Table 2).

2.3. Computational efficiency

The computational efficiency of DDM-3D has been discussed in detail elsewhere (Napelenok et al., 2006; Yang

Table 2
Linear regression analysis of the comparison between first order sensitivities of particulate matter species calculated using the brute-force and DDM for Aug 2–9, 2004 (Aug 1st was discarded as a spin-up day.)

Particulate matter species ($\mu g/m^3$)	Conc. (base case)	Sensitivity[a] to E(SO_2)	Sensitivity[a] to E(NO_x)	Sensitivity[a] to E(NO_x) mobile[b]	Sensitivity[a] to E(NH_3)	Sensitivity[a] to E(Xyl)	Sensitivity[a] to E(Terp)
	X, σ	X, σ	X, σ	X, σ	X, σ	X, σ	X, σ
Sulfate	1.85, 0.26	1.07, 0.27	0.19, 0.07	0.06, 0.03	0.02, 0.00	~0.00	~0.0
R^2		**0.998**, 0.00	**0.985**, 0.01	**0.964**, 0.04	**0.291**, 0.12		
Slope		1.00, 0.01	1.00, 0.03	0.98, 0.04	0.47, 0.18		
Intercept		0.00, 0.01	0.01, 0.00	0.00, 0.00	0.01, 0.00		
Nitrate	0.06, 0.02	−0.04, 0.02	0.19, 0.07	0.02, 0.00	0.07, 0.03	~0.00	~0.00
R^2		**0.236**, 0.00	**0.863**, 0.05	**0.776**, 0.08	**0.835**, 0.04		
Slope		0.14, 0.22	1.35, 0.14	1.36, 0.16	1.03, 0.18		
Intercept		0.01, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00		
Ammonium	0.43, 0.08	0.14, 0.05	0.04, 0.02	0.01, 0.01	0.17, 0.04	~0.00	~0.00
R^2		**0.979**, 0.01	**0.941**, 0.02	**0.883**, 0.05	**0.983**, 0.01		
Slope		1.10, 0.05	1.18, 0.06	1.17, 0.09	1.03, 0.04		
Intercept		0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00		
Anth. org.	0.02, 0.01	~0.00	~0.00	~0.00	~0.00	0.004, 0.00	~0.00
R^2						**0.998**, 0.00	
Slope						1.07, 0.02	
Intercept						0.00, 0.00	
Bio. org.	0.64, 0.08	~0.00	0.06, 0.00	0.02, 0.00	~0.00	~0.00	0.61, 0.08
R^2			**0.870**, 0.05	**0.884**, 0.05			**0.792**, 0.17
Slope			1.39, 0.13	1.39, 0.15			1.51, 0.26
Intercept			0.00, 0.00	0.00, 0.00			0.01, 0.07
Secondary PM[c]	3.00, 0.44	1.17, 0.30	0.35, 0.12	0.10, 0.05	0.25, 0.06	~0.00	0.58, 0.07
R^2		**0.991**, 0.00	**0.947**, 0.02	**0.916**, 0.03	**0.910**, 0.03		**0.790**, 0.17
Slope		1.03, 0.02	1.15, 0.03	1.16, 0.04	1.02, 0.10		1.51, 0.27
Intercept		0.02, 0.01	0.02, 0.01	0.01, 0.00	0.00, 0.01		0.02, 0.07

The presented summary includes a "cell-by-cell" comparison for each simulation hour. Brute-force results are regressed against DDM, and the mean, X, and standard deviation, σ, are calculated for the 8 days of the simulation averaged across all grid cells.

[a] Sensitivities are shown as 24-h averages to domain-wide emissions from the DDM calculation normalized to a 100% increase in emissions.

[b] Emissions from only mobile sources of NO_x.

[c] Secondary PM = sulfate + nitrate + ammonium + anthropogenic SOA + biogenic SOA.

et al., 1997; Dunker, 1981). It was found that calculating more than two sensitivity parameters simultaneously is more efficient using DDM-3D, with greater CPU saving as the number of parameters increases (Napelenok et al., 2006). Here, parallel processor implementation was also evaluated to check for computational scalability with increasing number of processors. All simulations were performed on the SGI Altix 4700 machine with dual-core Itanium2 processors and 4 GB of memory per core. The executables were built using the Intel Fortran compiler version 9.1 using mid-range optimization options (-O2). It was found that DDM-3D scales similarly to the base CMAQ code once the number of sensitivity parameters has been taken into account (Fig. 3). While there are large benefits of using multiple processors for both the CMAQ model and DDM-3D, they are quickly diminished from input/output operations as more CPUs are employed (at about eight for this system). Scalability was also found to be highly dependent on the computer system as a comparable set of simulations was found to experience large runtime benefits past eight processors on an Athlon Linux cluster (results not shown).

Napelenok et al. (2006) reported that advection is the second largest CPU time consumer in CMAQ after the gas-phase chemistry, ahead of diffusive transport by a small margin. CPU time spent in advection increased proportionally with the number of DDM parameters in their testing. To further the computational efficiency of the model, the advection code was reviewed for potential improvements. It was found that the computation of advective fluxes is conditional upon the direction of the wind velocity in CMAQ. Positive (in) and negative (out) fluxes are computed using different expressions depending on the sign of the velocity component along the direction

of advection, which is directionally split into three components (two horizontal and one vertical). The sign check was being done inside the species loop, unnecessarily increasing the number of time-consuming logical operations. Since the wind direction is the same for all species the sign could be checked only once, and the corresponding flux expression applied to all species, instead of once for every species. Accordingly, the logical (IF) check was moved outside the species loop. This required making the species loop the innermost loop with the grid cell (or directional zone) loop on the outside. Along with this coding change, the storage of the arrays within the advection module was changed from grid-cell-wise to species-wise with a simple swap of the orders of array dimensions and indexes. Note that these changes may further improve the performance on vector processing computers when the number of species is larger than the number of grid cells along any direction. As a result of these changes, the CPU time spent in the advection module decreased by 27% resulting in an overall speedup of about 11% (using one processor). This speedup impacts both concentration and sensitivity arrays since they are processed in the same way in the advection module. In the future, the species and DDM parameter loops can be combined to make the innermost loop much longer (if five DDM parameters are being computed, six times longer). This would significantly improve the performance not only because of the lengthening of the innermost loop, but because of the reduction in logical operation count as well. Note that the direction of the velocity component can be checked once for all species and all DDM parameters in that case, instead of once for each DDM parameter in our current implementation.

3. Conclusions

The implementation of the decoupled direct method in 3D in the CMAQ model was updated to reflect the recent evolution of the base code. More functionality to the direct sensitivity calculation was added in the form of allowing more model set-up options, calculating sensitivities to reaction rates, and allowing the use of parallel processors. A wide range of DDM-3D sensitivities was compared to traditional difference method computations during a summertime episode over the continental United States and showed reasonable statistical agreement with the traditional difference method.

As was discussed in the previous implementation of DDM-3D for PM species, the agreement between direct sensitivities and those calculated using the difference method is better for more direct relationships between aerosol species and the gaseous emissions (Napelenok et al., 2006). For example, correlations in the sensitivity of sulfate to SO_2 emissions and ammonium to NH_3 emissions are high (0.988 and 0.983). For more indirect relationships, such as the sensitivity of nitrate to SO_2 and sulfate to NH_3, the correlation is poor (0.236 and 0.291). Several factors are responsible for the poor comparison. Sensitivities from indirect relationships tend to be much smaller in magnitude than the corresponding concentration species and are therefore more susceptible to numerical
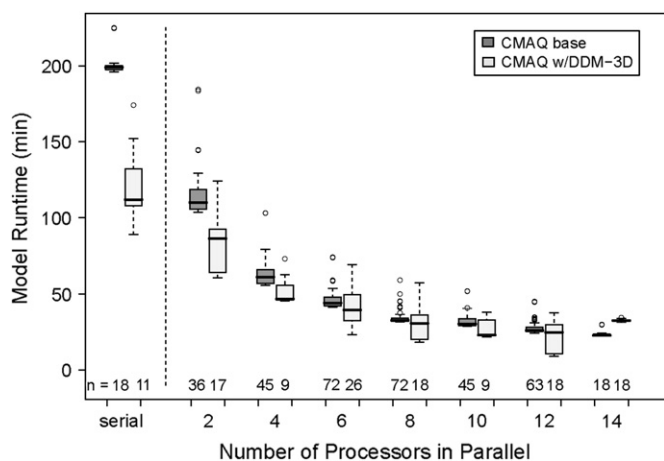

Fig. 3. CPU time required to complete one 24-h simulation on a $148 \times 112 \times 14$ cell domain. DDM-3D results are normalized by $(1 + npsen)$, where *npsen* is the number of sensitivity parameters calculated and ranged between 2 and 14 in this analysis. DDM-3D scaling with *npsen* was shown in Napelenok et al. (2006). Boxes contain 50% of the data points, the horizontal line denotes the mean, and the whiskers contain ± 1.5 times the interquartile range as defined by (Tukey, 1977). Outliers, shown as open circles, occur during periods of high network traffic. *n* Indicates the number of samples included in the mean. Mean DDM-3D runtime in the serial simulation (one processor) is significantly lower than base CMAQ due to the scaling by larger *npsen* values associated with these particular runs (7 runs used *npsen* = 5 and 4 runs used *npsen* = 10).

noise. Also, non-linearity plays a greater role in these cases and second and higher order terms become more important. These terms are not yet included in PM sensitivities.

The scalability of parallel processing was considered and found to be in line with the base model. Finally, some adjustments to base CMAQ advection code allowed for an 11% speedup in model runtime.

References

Byun, D.W., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Applied Mechanics Reviews 59, 51–77.

Carter, W.P.L., 2000. Documentation of the SAPRC99 Chemical Mechanism for VOC Reactivity Assessment. Air Pollution Research Center and College of Engineering, Center for Environmental Research and Technology, University of California, Riverside, CA.

Cohan, D.S., Hakami, A., Hu, Y., Russell, A.G., 2005. Nonlinear response of ozone to emissions: source apportionment and sensitivity analysis. Environmental Science & Technology 39 (17), 6739–6748.

Dunker, A.M., 1981. Efficient calculation of sensitivity coefficients for complex atmospheric models. Atmospheric Environment 15, 1155–1161.

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1989. A photochemical kinetics mechanism for urban and regional scale computer modeling. Journal of Geophysical Research 94 (D10), 12925–12956.

Grell, G., Dudhia, J., Stauffer, D., 1995. A description of the fifth-generations Penn State/NCAR mesoscale model (MM5). NCAR Technical Note, NCAR/TN-398 + STR.

Hakami, A., Odman, M.T., Russell, A.G., 2003. High-order, direct sensitivity analysis of multidimensional air quality models. Environmental Science & Technology 37 (11), 2442–2452.

Hu, Y., Odman, M.T., Russell, A.G., 2006. Mass conservation in the Community Multiscale Air Quality model. Atmospheric Environment 40 (7), 1199–1204.

Napelenok, S.L., Cohan, D.S., Hu, Y.T., Russell, A.G., 2006. Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM). Atmospheric Environment 40 (32), 6112–6121.

Odman, M.T., Russell, A.G., 2000. Mass conservative coupling of non-hydrostatic meteorological models with air quality models. In: Air Pollution Modeling and its Application, vol. XIII. Kluwer Academic/Plenum Publishers, New York.

Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling & Software 22 (11), 1543–1556.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley Pub. Co.

US-EPA, 2004. SMOKE v2.0 User's Manual. Available from: http://www.smoke-model.org/version2/index.cfm.

Yang, Y.J., Wilkinson, J.G., Russell, A.G., 1997. Fast, direct sensitivity analysis of multidimensional photochemical models. Environmental Science & Technology 31 (10), 2859–2868.