Beltsville PBL Air Quality Modeling – Atmospheric Chemistry
3-D Air Quality Modeling and Data Assimilation
William Stockwell, Rosa Fitzgerald

Objective: In this unit students will learn about how data assimilation may be used to determine the initial state of the atmosphere for meteorological modeling. Data assimilation remains under development for air quality modeling. Sensitivity analysis is closely related to data assimilation and students will learn about some basic concepts in sensitivity analysis. Students will perform some simple numerical experiments using brute force sensitivity analysis with the model of their choice.

Data Assimilation

A accurate representation of the initial three-dimensional meteorological state is a necessary starting point for a meteorological forecast. An air quality forecast requires the initial chemical state of the atmosphere as well. Time and spatially dependent boundary conditions at the top and lateral sides of the model need to be determined. Boundary conditions may be supplied from measurements or larger scale simulations, for example, a combination of global and stratospheric models could be used to supply boundary conditions for a regional scale model.

Data assimilation is used routinely for determining initial conditions for meteorological models. Daley (1991), Kalnay (2002), Lewis et al. (2006) and Sandu and Chai (2011) present reviews of several established methods for the assimilation of meteorological data. Objective analysis or spatial analysis was one of the first methods used to estimate meteorological fields for modeling (McRae et al. 1982). Methods of spatial analysis include optimum interpolation (OI) and 3-dimensional variational analysis (3DVAR). 3DVAR may be formulated in Bayesian terms (Lewis et al., 2006). OI and 3DVAR do not account explicitly for atmospheric dynamics so the produced meteorological fields are not necessarily consistent with the atmosphere's continuity equations of motion. However, the expectation is that the meteorological fields produced by spatial analysis will combine optimally forecast and observation error covariance.

A data assimilation method that produces meteorological fields consistent with those employed by a meteorological model is 4-dimensional variational assimilation (4DVAR; Lewis et al., 2006). The 4DVAR method uses a meteorological model as its forward operator. 4DVAR uses observational data and the model to estimate meteorological fields over an interval. These fields can then be used as the initial conditions to make a forecast. In this case, the initial fields produced by 4DVAR will be consistent with the model's dynamics. An understanding of measurement errors and those in the models (known as representativeness errors) is needed in the development of data assimilation methods. However, forecast errors are predicted implicitly in 4DVAR and not explicitly. The extent that the model's dynamics is used as a constraint on the assimilation process is not fixed, rather it may be weak or strong depending on what is the appropriate degree of forcing in a particular application.

Adjoint models are used in 4DVAR and the main function of the adjoint model is to make a connection between the numerical model variables and the observed quantities. Adjoint models provide a first-order approximation to sensitivity in a nonlinear model (AMS, 2012a). An adjoint model is a model consisting of adjoint equations that map a sensitivity gradient vector from time t_1 backward to a previous time t_0. The time t_0 would correspond to the initial time for a forecast. An adjoint version of a meteorological forecast model could be applied to make simulations while varying the initial state until differences between the adjoint simulations and observations are minimized.

There exist ensemble approaches to meteorological forecasting and data assimilation. An ensemble of forecasts may be made with one or several models and with the models having different initial conditions, boundary conditions or other model run parameter settings (AMS, 2012b). Ensemble approaches may be used to mitigate uncertainties in the preparation of the initial state for a forecast. It is assumed that the mean of many simulations, made with reasonable initial conditions and model setup, will be more accurate than a single forecast made with only one set initial conditions. Ensemble forecasts are especially useful if there are some input fields or model properties with a high degree of uncertainty that strongly affect the forecast. In any case, ensemble forecasts provide directly a measure of forecast uncertainty (Kalnay, 2002).

Ensemble Kalman filtering is one of the more advanced assimilation approaches. Kalman filtering is a sequential procedure that may be applied to ensemble approaches. Kalman filtering works well with ensemble approaches because it utilizes explicit predictions of forecast errors (Lewis et al., 2006). An ensemble Kalman filter consists of the 3DVAR method and an appropriately generated ensemble. Kalman filtering uses an ensemble forecast to deduce forecast error covariances and uses these to assimilate data from a given time. One of the most important advantages of Kalman filtering is that it avoids the necessity of creating a tangent linear model and an adjoint model of the forecasting meteorological model.

Ideally, initial and boundary conditions (and data assimilation) should be developed from high resolution 3-D meteorological measurements. Satellites, sondes and ground-based meteorological profilers help in providing this data, but an extensive data is seldom available and only during special research studies for limited areas. This makes meteorological data assimilation a very underdetermined problem from a mathematical point of view. Observing Systems Simulations Experiments (OSSEs; Arnold and Dey, 1986, Lord et al., 1997, Baker et al., 1995) can be used to create simulated datasets for the development of improved data assimilation methods and observing systems.

Data assimilation is not widely used for air quality modeling due to the much larger number of variables than required for meteorological modeling. There are a very number of chemical species concentrations that need to be represented by 3-D fields. The concentrations of the chemical species are subject to strong, highly nonlinear coupling. Finally, there are no observing systems that provide the necessary complex chemical measurements in 3-D. However, 4DVAR has been investigated for use with air quality models. For example, an adjoint module for a complex chemical scheme (Stockwell et al., 1990) was developed (Elbern et al., 1997) and this

module was incorporated into a full D-d air quality model by linking the chemical adjoint module with the adjoint dynamical component of an air quality model (Elbern et al., 1997; Elbern and Schmidt, 1999). They showed that 4DVAR may be a feasible method for air quality modeling. Today it is typical that the air quality model is initialized with idealized profiles based on field measurements and "best guesses". The model spins-up to find its own initial state by simulating a few days before the forecast period. The results of the spin-up period are discarded.

Sensitivity and Uncertainty Analysis

Although there are many different kinds of sensitivity analysis techniques that may be applied to meteorological and air quality modeling, most can be grouped into local sensitivity analysis and global analysis methods. For example, first-order ozone local sensitivity analysis coefficients give the response of chemical concentrations, such as ozone, to a model parameter (such as a rate coefficient, emission rate or meteorological input). They are local in the sense that they are valid incrementally near nominal input and parameter conditions. Local sensitivity analysis methods involve finding the response of a simulated quantity with respect to a parameter change. For example, a local, first order sensitivity analysis could be used to find the change in an ozone concentration [O_3] with respect to a change in aggregate VOC emissions E_{VOC}. This sensitivity would be the derivative, $\frac{d[O_3]}{dE_{VOC}}$.

Sensitivity coefficients may be estimated by varying model inputs and parameters one-by-one and this method is known as "brute force". Brute force methods may provide useful insight into the effect of uncertainties such as those in emissions inventories (Fujita et al., 2016). However, the brute force method does not provide the more comprehensive analysis that can be achieved by local and global sensitivity analysis combined with uncertainty analysis (Gao et al.,1995; Yang et al., 1995; Russell et al., 1995; Saltelli et al.; 2005). Although brute force methods can be used to estimate quantities, such as $\frac{d[O_3]}{dE_{VOC}}$, they are very computationally inefficient because brute force methods require at least two simulations with a small (incremental) parameter variations for each sensitivity coefficient evaluated. The accuracy of brute force methods is a concern because the size of the parameter variations must be large enough so that differences between the simulated values are large enough to avoid numerical noise limits while small enough to avoid the nonlinearities involved with chemistry. Another serious problem is that a complete sensitivity analysis by brute force methods requires a very large number of simulations. The decoupled direct method (DDM; Dunker 1984) allows sensitivities to be calculated directly without the need of performing multiple simulations. The set of differential equations for an air quality model that include sensitivity differential equations is very stiff and therefore very difficult to solve numerically, the decoupled direct method (DDM) and its extension, the high-order decoupled direct methods (HDDM) allow these differential equations to be solved.

A normalized sensitivity coefficient for chemical species X to the ith rate coefficient, $S_{X,i}$, is given by the following equation where $[\widehat{k_1}]$ is the nominal value of the ith rate coefficient, $[\widehat{X(t)}]$ is the nominal value of the concentration of species X at time t, and $\frac{d\{X(t)]}{dk_i}$ is the derivative of the concentration, *[X(t)]*, with respect to the rate coefficient k_i.

$$S_{X,i} = \frac{[\widehat{k_i}]}{[\widehat{X}]} \frac{d[X(t)]}{dk_i}$$

For example, the normalized sensitivity coefficients for the production of ozone and hydrogen peroxide are given by the following equations.

$$S_{O_3,i} = \frac{[\widehat{k_i}]}{[\widehat{O_3}]} \frac{d[O_3]}{dk_i} \qquad S_{H_2O_2,i} = \frac{[\widehat{k_i}]}{[\widehat{H_2O_2}]} \frac{d[H_2O_2]}{dk_i}$$

Figure 1 shows an example of a time series of normalized sensitivity coefficients for the production of hydrogen peroxide to the rate coefficients for the reactions: $CH_3CH\text{-}(OH)\text{-}CH_2O_2$ + NO → Products; HCHO + hv → 2 HO_2 + CO; Toluene + HO → Products; HO_2 + NO → Products and NO_2 + HO → HNO_3. Note that $CH_3CH\text{-}(OH)\text{-}CH_2O_2$ is an organic peroxy radical produced from alkenes.

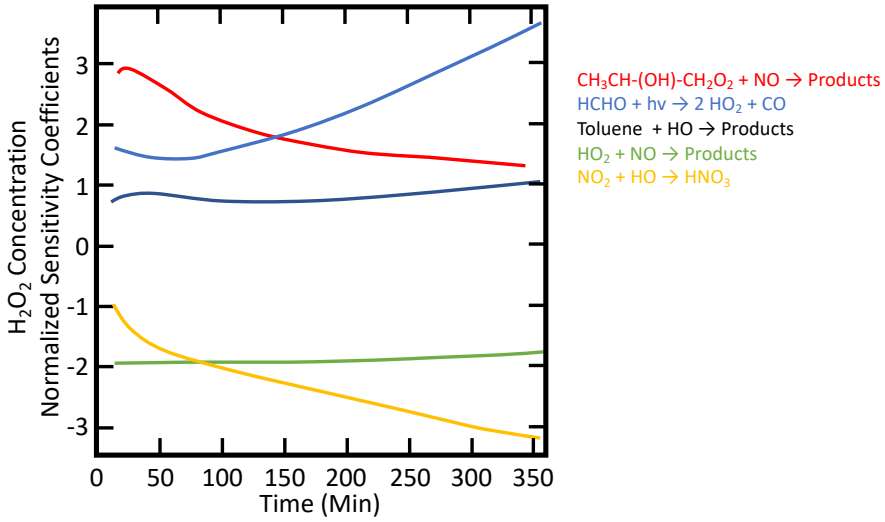

Figure 1. Selected normalized sensitivity coefficients for hydrogen peroxide concentrations with respect to rate coefficients as described in Stockwell (1986).

Note that the normalized sensitivity coefficients are time dependent; some increase while others decrease with time. If a sensitivity coefficient increases with time that rate coefficient value become more important in determining the final concentration of a substance.

High-order sensitivity analysis coefficients extend local methods to provide measures of the response of ozone concentrations over a wider range of conditions but within some limits. Local sensitivity methods have been extended to the higher-order HDDM for the analysis of O_3 due to its complex nonlinear chemical formation processes and have been used successfully to evaluate the effectiveness of control measures and quantify O_3 formation potential (Hakami et al., 2003; Cohan et al.,2005; Kim et al., 2009, Itahashi et al., 2013). Sensitivities may be calculated for higher

order derivatives, such as $\frac{\partial^2[O_3]}{\partial E_{VOC}^2}$ and $\frac{\partial^2[O_3]}{\partial E_{VOC}\partial E_{NOx}}$. These higher order derivatives extend the range of emission changes over which the sensitivity coefficients apply and coefficients such as $\frac{\partial^2[O_3]}{\partial E_{VOC}\partial E_{NOx}}$ allow cross terms between the VOC emissions and the NO_x emissions, E_{NOx} to be calculated. Codes for the calculation of these HDDM coefficeints, similar to DDM, have be developed (Hakami et al., 2003; Byun and Schere, 2006) and it have been incorporated in to CMAQ (Byun and Schere, 2006).

Global sensitivity methods provide insight into the effects of parameter variations on model simulations over a wider variety of conditions than local sensitivity analysis (Gao et al., 1995; Russel et al., 1985). Global sensitivity methods usually involve Monte Carlo methods where model input parameters are generated randomly within a range of physically possible values and the uncertainty range in conditions (Russell et al., 1995). Cumulative frequency diagrams or tables that show the relative probability of simulation outcomes may be constructed from Monte Carlo selected simulations. Figure 2 shows an example of cumulative frequency diagram for ozone control strategy based on a 25% reduction in initial concentrations of volatile organic compound emissions. A global sensitivity analysis made with a full set of Monte Carlo simulations can be considered as a very large ensemble simulation. A full set of Monte Carlo would be much too expensive in terms of computational cost so it must be constrained. Latin Hypercube Sampling (McKay et al., 1979) is often used to reduce the number of simulations while still sampling a representative set of input parameters. This method has been have used in air quality studies such as Gao et al. (1995) and Russell et al. (1995).

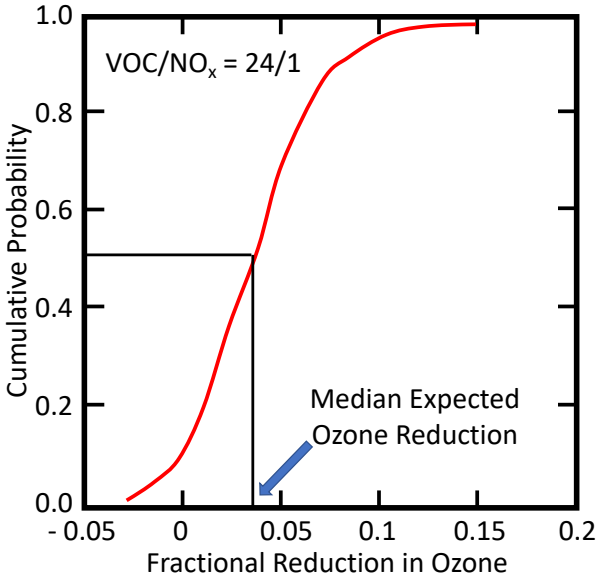

Figure 2. An example of a cumulative frequency diagram for ozone derived from a set of Monte Carlo simulations made with Latin hypercube sampling (Gao et al., 1995). This figure shows the probability of a fractional reduction in ozone due to a 25% reduction in initial concentrations of volatile organic compound emissions. The blue arrow shows the median expected ozone reduction for the proposed control strategy.

References

AMS (2012a) American Meteorological Society, Adjoint Model. Glossary of Meteorology. (http://glossary.ametsoc.org/wiki/Adjoint_model)

AMS (2012b) American Meteorological Society, Adjoint Model. Glossary of Meteorology. (http://glossary.ametsoc.org/wiki/Ensemble_forecast)

Arnold, C. P. and C. H. Dey (1986) Observing-Systems Simulation Experiments: Past, Present, Future, *Bull. Amer. Met. Soc., 67*, 687-695.

Baker, E. W., G.D. Emmitt, F. Robertson, R.M. Atlas, J.E. Molina, D.A. Bowdle, J. Paegle, R. M. Hardesty, M.J. Post, R.T. Menzies, T.N. Krishnamurti, R.A. Brown, J.R. Anderson, A.C. Lorenc and J. McElroy (1995) Lidar-Measured Winds from Space: A Key Component for Weather and Climate Prediction, *Bull. Amer. Met. Soc., 76*, 869-888.

Byun, D. and K.L., Schere (2006) Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Appl. Mech. Rev*., **59**, 51, 2006.

Cohan, D.S., A Hakami, Y. Hu, and A.G, Russell (2005) Nonlinear Response of Ozone to Emissions: Source Apportionment and Sensitivity Analysis, *Environ. Sci. & Techn.*, *39*, 6739-6748.

Daley, R. (1991) *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, 457 pp.

Dunker, A.M (1984) The Decoupled Direct Method for Calculating Sensitivity Coefficients in Chemical Kinetics, *J. Chem. Phys., 81*, 2385-2395.

Elbern, H., H. Schmidt and A. Ebel (1997) Variational Data Assimilation for Tropospheric Chemistry Modeling, *J. Geophys. Res., 102*, 15967-15985.

Elbern, H. and H. Schmidt (1999) A Four-Dimensional Variational Chemistry Data Assimilation Scheme for Eulerian Chemistry Transport Modeling, *J. Geophys. Res., 104*, 18583-18598,.

Fujita, E.M., D.E. Campbell, W.R. Stockwell and E. Saunders, R. Fitzgerald and R. Perea (2016) Projected Ozone Trends and Changes in the Ozone-Precursor Relationship in the South Coast Air Basin in Response to Varying Reductions of Precursor Emissions, *J. Air Waste Manage. Assoc., 66,* 201–214. doi: 10.1080/10962247.2015.1106991

Gao, D., W.R. Stockwell and J.B. Milford (1995) First Order Sensitivity and Uncertainty Analysis for a Regional Scale Gas-Phase Chemical Mechanism, *J. Geophys. Res.*, *100*, 23153-23166.

Hakami, A., M.T. Odman, and A.G Russell, High-order, Direct Sensitivity Analysis of Multidimensional Air Quality Models (2003) *Environ. Sci. & Techn., 37*, 2442-2452.

Itahashi, S., I. Uno, and S. Kim, Seasonal Source Contributions of Tropospheric Ozone over East Asia Based on CMAQ-HDDM (2013) *Atmos. Environ. 7*0, 204-217.

Kalnay, E. (2002) *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, 341 pp.

Kim, S., D.W. Byun, and D.S, Cohan (2009) Contribution of Inter- and Intra-State Emissions to Ozone over Dallas-Fort Worth, Texas. *Civil Engineering and Environmental Systems, 26*, 103-116,.

Lewis, J. et al. (2006) *Dynamic Data Assimilation: A Least Squares Approach*, Cambridge University Press, Cambridge, 654 pp.

Lord, S.J., E. Kalnay, R. Daley, G.D. Emmitt and R. Atlas (1997) Using OSSEs in the Design of the Future Generation of Integrated Observing Systems. *Preprint Volume, 1st Symposium on Integrated Observing Systems*, Long Beach, CA, 2-7 February.

McKay, M.D., R.J. Beckman and W.J. Conover (1979) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics, 21*, 239-245.

McRae, G.J. and Goodin, W.R. and Seinfeld, J.H. (1982) Mathematical modeling of photochemical air pollution, Environmental Quality Laboratory Report, 18. California Institute of Technology , Pasadena, CA. https://resolver.caltech.edu/CaltechEQL:EQL-R-18 see also https://authors.library.caltech.edu/25729/

Russell, A., J.B. Milford, M.S. Bergin, S. McBride, L. McNair, Y. Yang, W.R. Stockwell and B. Croes (1995) Urban Ozone Control and Atmospheric Reactivity of Organic Gases, *Science*, *269*, 491-495.

Sandu, A. and T. Chai (2011) Chemical Data Assimilation – An Overview, *Atmosphere, 2*, 426-463; doi:10.3390/atmos2030426

Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo (2005) Sensitivity Analysis for Chemical *Models, Chem. Rev., 105*, 2811-2828. DOI: 10.1021/cr040659d

Stockwell, W.R., A Homogeneous Gas Phase Mechanism for use in a Regional Acid Deposition Model (1986) *Atmos. Environ., 20,* 1615-1632.

Stockwell, W.R., P. Middleton, J.S. Chang and X. Tang (1990) The Second Generation Regional Acid Deposition Model Chemical Mechanism for Regional Air Quality Modeling, *J. Geophys. Res., 95*, 16343-16367.

Yang Y-J., W.R. Stockwell and J.B. Milford (1995) Uncertainties in Incremental Reactivities of Volatile Organic Compounds, *Envir. Sci. Technol., 29*, 1336-1345.

Exercises

Perform a sample local sensitivity study. Chose one of your models, chemical box, 1-D, Quick TUV, etc. Choose a base case and simulate it. Make additional simulations while varying the model parameters by some small amount (±5%). Use the results to calculate sensitivity coefficients $S = \Delta O / \Delta p$ where ΔO is the change in an observable and Δp is the change in a model parameter or input variable. What are the highest local sensitivity coefficients you find?

Can you normalize your local sensitivity coefficients as discussed above? If you can, please do, and compare the normalized sensitivity coefficients. What are the highest normalized local sensitivity coefficients now?

You might try investigating the effect of varying the magnitude of parameter variations on sensitivity coefficients. Try changing the parameters by 10%, 20% and 50%. Use the new simulations to determine the effect on your recalculated sensitivity coefficients.

Do your calculations have any implications for measurement systems / Observing Systems Simulations Experiments (OSSEs)?