P1B.1

QUANTIFYING UNCERTAINTY IN CONCENTRATION OUTPUT GENERATED BY ATMOSPHERIC AND DIFFUSION MODELS

Michael K. Atchison[1], ENSCO Inc., Melbourne, FL,
Brian J. Provoncha, Stephanie L. Seely, and Dr. Alfred J. Vanderzanden

1. INTRODUCTION

A frequent problem encountered in air-pollution meteorology is how to quantify uncertainty in concentration data produced by atmospheric transport and diffusion models (ATDM). Sources of uncertainty can arise from many model inputs (e.g., different weather or model settings). One of the primary sources of uncertainty arises from error in the horizontal wind data used by models to calculate advection and diffusion of pollutants. Since winds are a critical model input in determining source-to-sampler transport and ultimately the predicted concentrations, it is very important to assess the uncertainties due to these data. This paper will describe a system called CLUES (Concentration Level Uncertainty Ensemble System), which has been developed to provide information about the uncertainty in concentration output generated by atmospheric transport and diffusion models.

CLUES uses a variant of the classical Monte Carlo method to generate an ensemble of ATDM outputs from which uncertainty information is obtained. This is accomplished in four main steps. The first is to make use of a statistical model to generate simulated errors for the wind input. The next step is to use these simulated errors to perturb the winds. Depending on the type of wind data, this may be done dynamically while the ATDM is running, or otherwise is done prior to running the ATDM. In either case, the ATDM is run with the perturbed winds to generate an ensemble of concentration fields. Once the ensemble is created, statistics (standard deviations, confidence limits, and other quantities) are generated and displayed graphically to help the user assess overall model output uncertainty. Finally, CLUES also provides a way to check for convergence of the ensemble statistics.

This paper will present the overall structure of CLUES, as well as how the system has been applied to a transport and diffusion model.

2. DESIGN OF CLUES

The CLUES software system includes Fortran 90 programs, Ada packages, C-shell scripts, and the commercial statistical software SAS®. At ENSCO, a 16-processor SGI with 4GB of memory is used to perform perturbation of the winds and execution of the ensemble of transport and dispersion simulations. Processing of the concentration model output takes place on a Sun Microsystems platform using SAS®. CLUES may be considered computationally intensive because it requires significant computer resources to generate an ensemble of modest size (for example, 30 to 100 members). However, even in the case of gridded wind input from a mesoscale meteorological model, CLUES is still practicable since the approach taken is to perturb wind output from a single run of a mesoscale model, rather than running the mesoscale model multiple times. Using repeated execution of a mesoscale model would require significantly more computer resources for a comparable size ensemble. The system also conserves disk space by allowing the user to define a subset of the mesoscale model gridded winds so that perturbations are produced only for the time interval and pressure levels that are needed to calculate the advection and dispersion of the pollutant.

In 1999, the initial prototype was designed to perturb mesoscale model data and to use the transport and dispersion algorithms of the Short-range Layered Atmospheric model (SLAM) to produce downwind concentrations (ENSCO, Inc. 1994). Since then, the software has been upgraded to include the ability to perturb surface and upper-air observational data. A feature still currently in development will allow the calculation of uncertainties in pollutant concentrations due to errors in estimated mixing depths. Version 1.0 remains officially in development as further options are added and its utility is explored through case studies.

3. CLUES METHODOLOGY

The main principle behind the design of CLUES is to enable certain field variables or transport model parameters to be selected with the intent of propagating their uncertainties into the transport model output. In theory, the approach may be applied to any input parameter of the ATDM, so long as a method exists for estimation of the uncertainty in the parameter setting. As previously stated, the first step is to simulate errors in some ATDM model input. Currently the software accepts either wind fields from the Regional Atmospheric Modeling System (RAMS)

[1] Corresponding author's address: ENSCO, Inc., 4849 North Wickham Road, Melbourne, FL 32940

version 4.3 (Pielke et al., 1992) or observed surface and upper-air wind data.

In the case of RAMS gridded wind fields, CLUES does not perturb the winds independently at each of the grid points, as might be done in a classical Monte Carlo implementation. Instead, an uncertainty model was formulated that allows the simulated RAMS wind errors to be spatially correlated across the grid. The nature of this correlation is similar to that used in the data assimilation process at Goddard Space Flight Center (Dee and Da Silva, 1999 and Gaspari and Cohn, 1999). Qualitatively, this correlation has the property of being large for pairs of grid points that are close to each other, but gradually decreases to zero for pairs of grid points separated by increasingly larger distances. The statistical model used for the simulated wind errors is that of a three-dimensional Gaussian random field having the correlation structure just mentioned. Efficient implementation of this uncertainty model is based on the weighted average method (Oliver, 1995). The parameters that specify the uncertainty model for RAMS winds are estimated empirically on the basis of spatial correlations computed from a selected RAMS data set and a corresponding set of radiosonde observations. The time period and region covered by these model-building data sets were, respectively, September through November 1997 and the continental United States (ENSCO, Inc. 2001). For reasons of computational efficiency the RAMS uncertainty model is applied repeatedly to a single set of RAMS wind data in order to generate an ensemble of such wind data sets. Each of these wind data sets is then used as input to the ATDM in order to generate an ensemble of ATDM concentration outputs.

In the case of surface and upper-air observational wind data, measurements are made independently from site to site. Consequently, it seems reasonable to use an uncertainty model that permits perturbation of observational data to be performed independently from site to site. However, for upper-air measurements, one should still expect wind errors to be correlated vertically at the various pressure levels. This is reasonable since a vertical wind sounding is obtained from the same instrument. In addition to measurement error, observational winds are subject to round-off error before they are placed in the wind data base. For example, surface wind directions are typically rounded to the nearest ten degrees and surface wind speeds are rounded to the nearest knot (approximately 0.5 m/s). The uncertainty models provided by CLUES for observed surface and upper-air winds attempt to account for all these things.

For observed surface winds, the uncertainty model requires a two-step perturbation for both wind speed and wind direction in order to simulate surface wind error. First, a perturbation is made independently for wind speed and wind direction to account for round-off error. In both cases, the statistical model for round-off error is a Uniform distribution over an appropriate range of values. Then, to account for measurement error, a second perturbation is applied. The statistical model used to perturb surface wind speed is the distribution of wind speed that results when the underlying eastward and northward components of wind error have a two-dimensional Gaussian distribution. Likewise, the surface wind direction is perturbed according to the distribution of wind direction that results when the underlying eastward and northward components of wind error have a two-dimensional Gaussian distribution. The two Gaussian distributions just mentioned are not the same; they differ in their standard deviations. This is necessary in order to obtain measurement perturbations that reflect published values of measurement errors for surface wind speed and direction. By default CLUES uses errors given in (Meteorological Handbook, 1982), but the user has the option to make his own specifications.

For upper-air winds from radiosondes, the uncertainty model used by CLUES also requires a two-step perturbation, similar to that used for surface wind observations. The first step applies a perturbation independently to wind speeds and wind directions at each pressure level to account for round-off error. As for surface winds, the statistical model for round-off error is a Uniform distribution over an appropriate range of values. The second step is somewhat different than that used for surface winds, but the purpose is the same: to account for measurement error. Basically, what happens is that the wind speeds and wind directions at each pressure level are converted to downwind and crosswind components. A Gaussian perturbation is applied to each of these components in such a way that the simulated errors of the wind components have vertical correlations that decrease as the pressure difference increases. This correlation model was estimated empirically from the same data sets that were used to estimate the RAMS uncertainty model. By default, the standard deviation parameters of the Gaussian model are those given in (Meteorological Handbook, 1982) for upper-air radiosondes. The crosswind error standard deviation is constant, but the downwind error standard deviation increases with wind speed.

CLUES applies the uncertainty models for both observed surface winds and observed upper-air winds dynamically within the ATDM. Perturbing the observed winds while the ATDM is running is more efficient than preparing an ensemble of wind data sets containing perturbed winds that would then be used as input to the ATDM. This is true for observed winds since there is no spatial (horizontal) correlation that needs to be modeled, as in the case of RAMS winds. The additional time required to perturb an individual radiosonde sounding or an individual surface observation is almost negligible compared to the time it takes to process the data when not making perturbations.

Finally, it should be mentioned that before an ensemble of ATDM outputs can be generated, CLUES requires that the ATDM be executed once by the user (using the most appropriate settings) in order to produce a "baseline" simulation. Then the transport model is run repeatedly using this simulation as a template (once for each set of perturbed wind data) in order to generate an ensemble of concentration outputs. The variation among the members of this ensemble is the basis for information about the uncertainty in the concentration model output.

4. EXAMPLE OUTPUT

Once an ensemble of ATDM outputs has been generated, the next step is to create ensemble statistics to help the user assess uncertainty. The first example we show is a confidence limit for concentrations simulated by the ATDM SLAM at a sampler downwind of a source. Figure 1 shows the variation of a pollutant concentration (red line) at the sampler over time. The blue symbols represent average hourly concentrations, each from some member of the ensemble of concentrations produced by the ATDM. The purple line represents the ensemble mean concentration over time. The yellow area delimits the 3-sigma partial confidence limits around the modeled concentration (in red). The term 'partial confidence limit' indicates that only part of the overall uncertainty is being accounted for.

An important step is to verify convergence of the ensemble statistics. A lack of convergence can indicate the ensemble did not have enough members. In that case, additional members may be generated until convergence is observed. Figure 2 contains a plot of a convergence parameter that describes the adequacy of convergence for confidence limits. As the ensemble size increases, the convergence parameter will eventually approach zero. In practice, however, we only require that the convergence parameter get close to zero, which is indicated by the yellow-filled region at the bottom of the plot.

Frequently, ATDM users also desire the ability to assess uncertainty in the location of modeled plumes. This information is easily generated from the CLUES concentration ensemble, and may be displayed in a specialized Java mapping application. An example of such a display is shown in Figure 3. The black isopleths represent levels of pollutant concentration observed from the "baseline" SLAM run. The concentration data for this example were calculated by the ATDM over a regular grid of virtual samplers surrounding the pollutant source. For each of these virtual samplers, a certain percentage of the ensemble members produced concentrations above the value observed at that sampler in the baseline run. When these percentages are assigned to the sampler locations and contoured, a representation emerges of the spatial uncertainty in the plume location. The colored areas in this figure represent percentages, and the plot is referred to as a type of threshold frequency.

Another type of threshold frequency calculation assists the user in assessing the uncertainty in the timing of the arrival of a plume at a sampler location. For example, the plot in Figure 4 shows the hourly average concentration (yellow) observed at a sampler location as calculated in the baseline simulation. The red line indicates the percentage of ensemble members whose calculated concentrations at that same sampler location exceeded those of the baseline run. The user may also specify a fixed threshold for comparison, as shown in Figure 5.

5. THE FUTURE OF CLUES

Future development of CLUES is expected to be directed towards providing additional support for propagating uncertainty due to other meteorological parameters (for example, atmospheric mixing heights), providing support for other ATDMs, incorporating atmospheric chemistry and the associated uncertainty due to chemical parameters (for example, reaction rate coefficients), providing better graphics and displays, and improving the uncertainty models used within CLUES.

6. SUMMARY

We have described a software system called CLUES that was designed to generate estimates of uncertainty in ATDM concentration output. The system uses the Monte Carlo method to rapidly generate ensembles of ATDM simulations using a variety of uncertainty models. The current version can perturb wind input to the ATDM for either mesoscale model gridded wind fields or surface and upper-air observational data. The perturbation technique may be applied to any input parameter, as long as a model exists for estimating uncertainty in the parameter setting. The CLUES system also has a relatively sophisticated set of methods for interpreting ensemble data and displaying statistics for the user. Sample output was included to show how the user may assess uncertainty in the magnitude and timing of simulated concentrations at a sampler; uncertainty in the width, centerline, and plume boundaries, or the ensemble probability of exceeding a fixed threshold concentration at a sampler location. A feature currently in development will allow the calculation of uncertainties in pollutant concentrations due to errors in estimated mixing depths. Also, the system will eventually utilize other ATDMs in addition to SLAM. Version 1.0 remains officially in development as further options are added and its utility is explored through case studies.

7. REFERENCES

Dee, D.P. and A.M. da Silva, 1999: Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology. Mon. Wea. Rev., 127,1822-1834.

ENSCO, Inc., 1994: Software Requirement Specification for the Short-Range Layered Atmospheric Model (SLAM) CSCI of the Integrated Evaluation System. Prepared for the U. S. Air Force by ENSCO Inc., Melbourne FL.

ENSCO, Inc., 2001: Ensemble Uncertainty Analysis (EUA) User's Guide, Prototype Version 0.1. Prepared for the U. S. Air Force by ENSCO Inc., Melbourne FL.

Gaspari, G. and S. Cohn, 1999: Construction of Correlation Functions in Two and Three Dimensions. Quart. J. Roy. Meteor. Soc., 125, 723-757.

Meteorological Handbook, 1982: ESMCP 105-1, Eastern Space and Missile Center Pamphlet, Department of the Air Force, Headquarters Space and Missile Center, Patrick Air Force Base, FL.

Oliver, D.S., 1995: Moving Averages for Gaussian Simulation in Two and Three Dimensions. Math. Geol., 27, 939-960.

Pielke, R.A., W.R. Cotton, R.L. Walko, C.J. Tremback, W.A. Lyons, L.D. Grasso, M.E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee, and J.H. Copeland, 1992: A Comprehensive Meteorological Modeling System – RAMS. Meteor. Atmos. Phys., 49, 69-91.

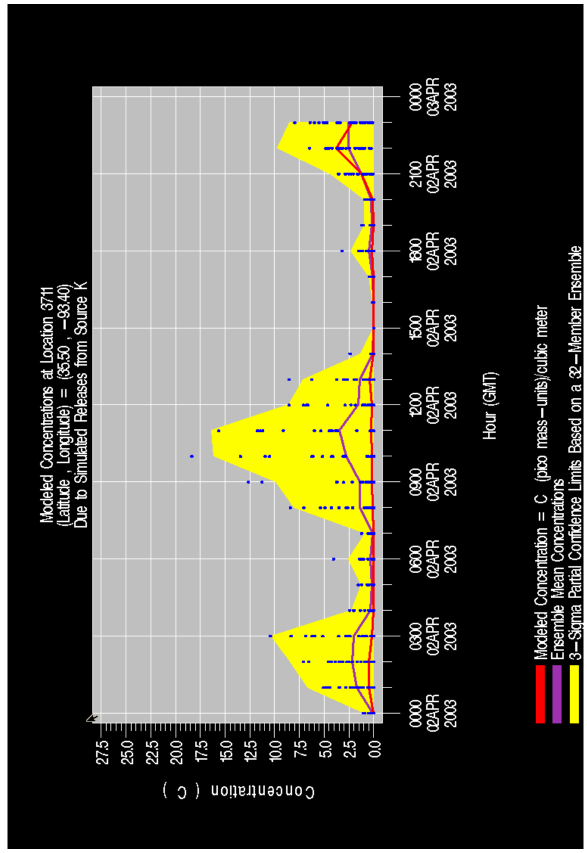

Figure 1. Time series of simulated hourly average concentration produced at a sampler by the ATDM SLAM (red) shown with ensemble members (blue), ensemble mean (purple) and 3-sigma partial confidence limits (yellow).

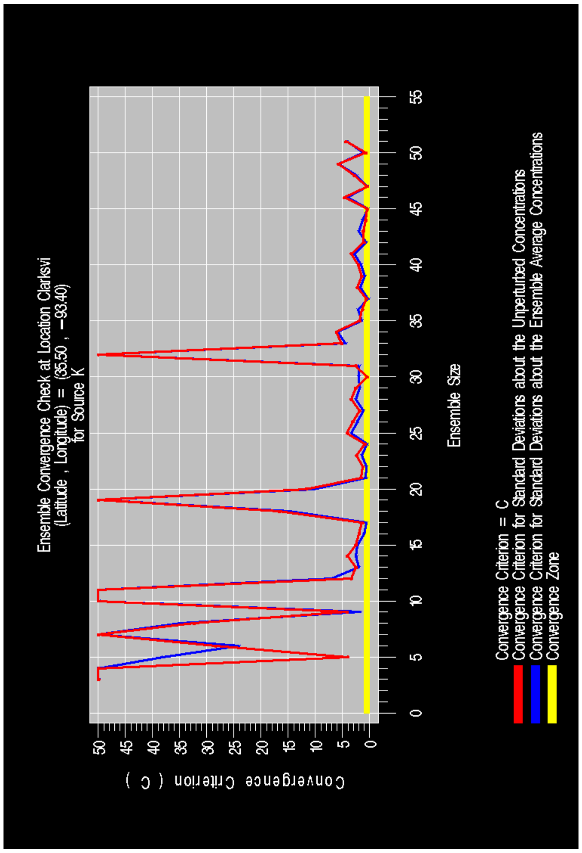

Figure 2. This plot is used to determine the adequacy of the ensemble size by showing the variation of some convergence criteria over time as the size of the CLUES ensemble is increased.

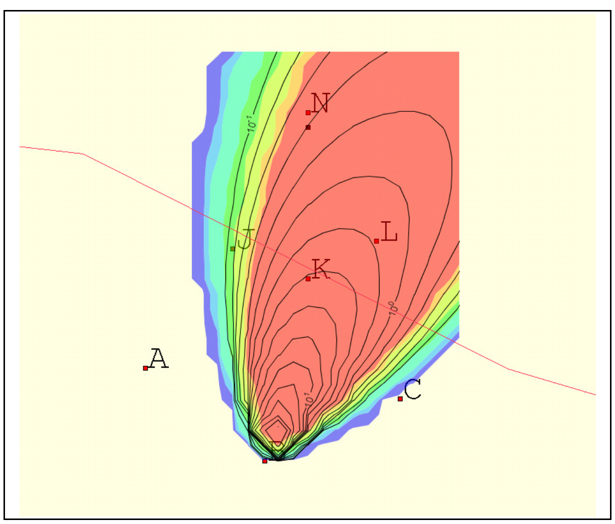

Figure 3. Depiction of baseline ATDM concentrations (black lines) paired with ensemble percentages (colored contours) for assessing the spatial uncertainty of plume location.

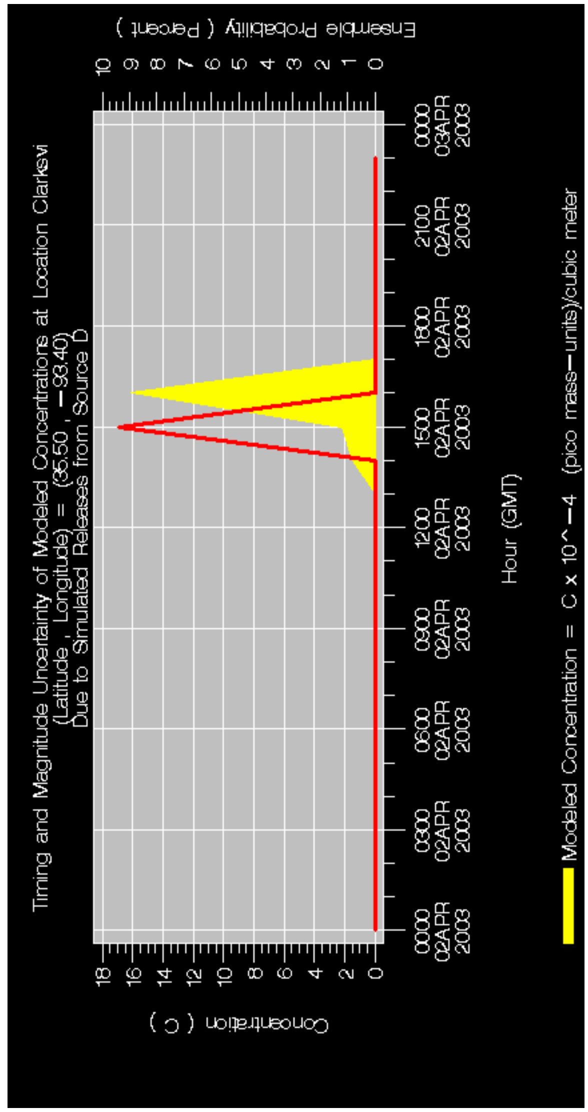

Figure 4. Time series of simulated hourly average concentration (yellow) produced at a sampler by the ATDM SLAM paired with ensemble percentages (red) in order to depict model-timing uncertainty. Concentrations from ensemble members were compared with baseline concentrations.

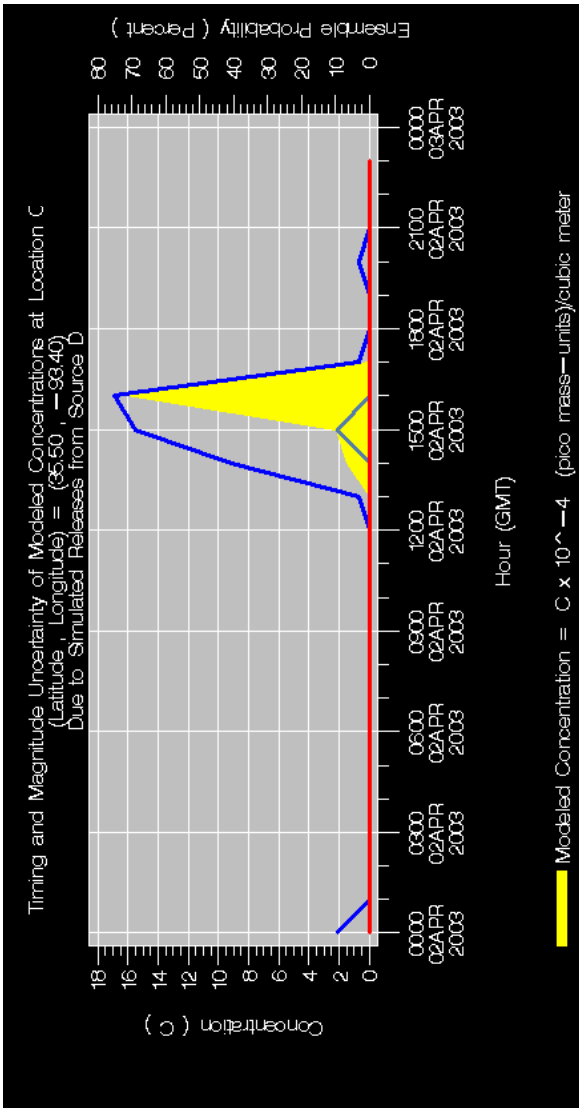

Figure 5. Time series of simulated hourly average concentration (yellow) produced at a sampler by the ATDM SLAM paired with ensemble percentages (blue) in order to depict model-timing uncertainty. Concentrations from ensemble members were compared with a fixed threshold.

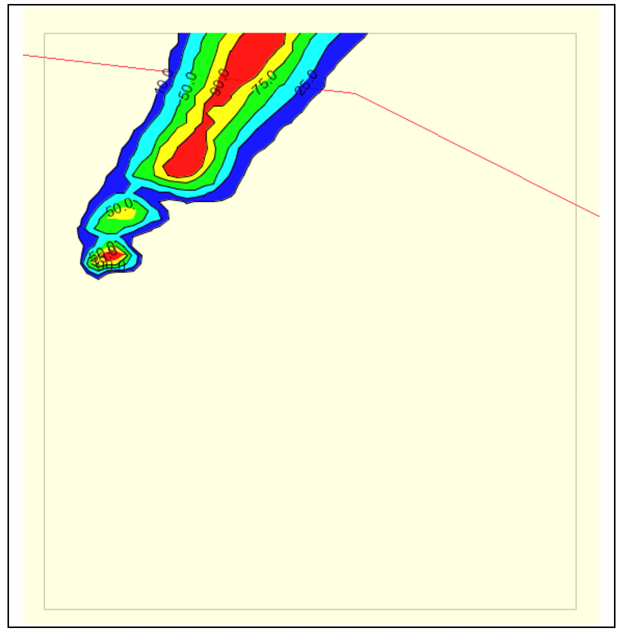

Figure 6. Ensemble probability of a concentration exceeding a threshold over a grid of virtual samplers. Concentrations from ensemble members were compared with a fixed threshold. This type of plot is used to spatially depict uncertainty in the width and position of the plume.