New Directions in Air Quality Model Evaluation: Probabilistic Model Evaluation

Kristen Foley, Robert Pinder, Sergey Napelenok

2009 AMAD Peer Review: Session 2
Model Evaluation: Establishing Model's Credibility

Environmental Issue

❖ Model simulations are used to fill in the gaps when observations are not available. This includes providing information for future or hypothetical scenarios such as for risk assessment of different air quality control programs, or for modeling climate change impacts on air pollutants such as ozone and fine particulate matter.

❖ When weighing the societal benefits of different air quality management options, policy-makers need quantitative information about the **relative risks** and **likelihood of success** of different options to guide their decisions. A key component in such a decision support system is an air quality model that can estimate both a single "best-estimate" of pollutant levels and a credible probabilistic range of concentrations.

❖ An **ensemble** of deterministic simulations is frequently used to create probabilistic estimates that account for uncertainty in the modeling system. A challenge with applying these approaches for simulations of air pollutant concentrations is that chemistry-transport models require significant amounts of input data and computational resources to complete a single simulation.

Research Objectives

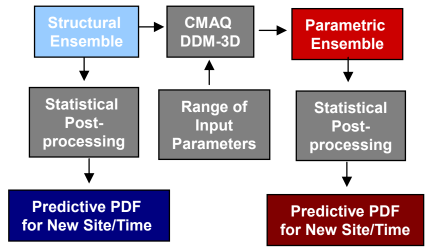
❖ Develop a computationally efficient approach to create an ensemble using multiple configurations of the CMAQ air quality model and quantitative estimates of the uncertainty in the model inputs.

❖ Apply and test a statistical method, known as Bayesian Model Averaging (BMA) (Raftery et al., 2005), for post-processing the ensemble of model runs based on observed pollutant levels.

❖ Use the range of predicted values provided by the ensemble of simulations to quantify the uncertainty in the model simulations (e.g. provide 90% confidence intervals for pollutant levels, rather than a single point estimate).

❖ Demonstrate the utility of such approaches for informing air quality management decisions. Use the ensemble of simulations to estimate the probability of exceeding a given threshold of pollutant concentration under current and future emission scenarios.

Approach



❖ **Structural Ensemble**: Change the structural form of the air quality model system. Differences in the six distinct model simulations included the chemical mechanism used in the air quality model and the land-surface model and planetary boundary layer model options in the meteorological model runs.

❖ **Parametric Ensemble**: Change inputs to the model. CMAQ-DDM-3D (Cohan et al., 2005) was used to estimate a reduced form model of ozone concentration to calculate the change in ozone as a result of changes in NO_x emissions, VOC emissions, and ozone boundary conditions. Previous work has shown these to be among the inputs that have the most impact on ozone concentration. These inputs were varied over a range based on previous studies and available data for this domain.

❖ **Post-processing Ensemble Simulations**: Weight each ensemble member based on observed ozone concentrations. The BMA predictive probability distribution function (PDF) model for the ozone concentration at one location, *s*, and time, *t*, y(s,t), is expressed as a weighted average of normal distributions with means equal to the *K* different ensemble member predictions at that site and time, $\{m_k(s,t),\ k=1,\ldots,K\}$:

$$p(y(s,t)\mid m_1(s,t),\ldots,m_K(s,t)) = \sum_{k=1}^{K} w_k N(m_k(s,t),\sigma^2)$$

Observed ozone concentrations were used to find the maximum likelihood estimates for σ^2 and the weight parameters for each ensemble member to reflect the "best" performing model runs. These parameter estimates can then be used to estimate a probability distribution function for any unmonitored location or time (e.g. Fig. 1).

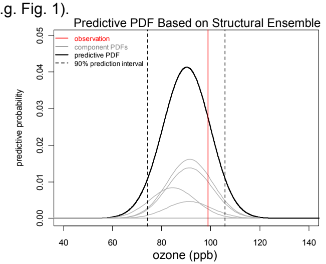


Figure 1. Predictive PDF and its components based on the 6 member structural ensemble for the average daily ozone concentrations at a site outside of Atlanta, GA on July 2nd, 2002.

Results and Discussion

Evaluating the Performance of an Ensemble of Simulations

This analysis is based on daily average ozone (9am-5pm) at 38 AQS stations in the SE US during July 2002.

❖ **Reducing bias**: The structural ensemble predictions tended to be biased high for this study period. The ensemble generated using CMAQ-DDM-3D and the post-processing technique both reduced the mean absolute bias. Applying BMA to the parametric ensemble produced the best results in terms of lower absolute bias and higher correlation with observations.

❖ **Ensemble calibration:** Statistical post-processing provides ensembles that are very well calibrated. For example, when the predictive PDF based on the parametric ensemble was used to estimate 90% prediction intervals for a set of cross validation sites not used in the statistical estimation, the relative frequency of the observed concentrations falling within these bounds was 91%, i.e. near the nominal value. In contrast, using the min/max values predicted by the 6 member structural ensemble only captured the observations 24% of the time. By creating well-calibrated ensembles, we can calculate the probability of exceeding a threshold concentration under different emissions control strategies (e.g. Fig. 2).

Estimating the Probability of Exceeding a Threshold

Figure 3 below is an example of the ensemble performance in predicting the probability of the ozone concentration exceeding a threshold value of 60ppb. The best results were from the post-processed parametric ensemble; this approach showed a large improvement over the original structural ensemble.

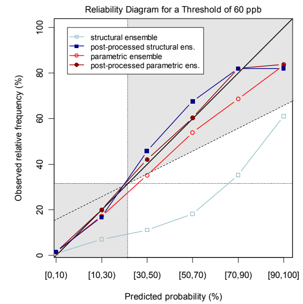


Figure 3. Reliability diagram for a threshold of 60 ppb. The grey shaded area represents ensemble estimates that have overall skill greater than a climatological estimate, denoted by the dotted lines as the observed relative frequency of exceeding the threshold across all sites/days in the study (0.31).

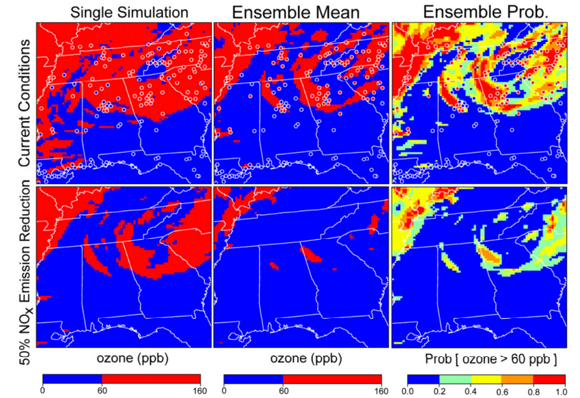


Figure 2. Spatial plots of ozone and probability of exceeding the threshold concentration, for current conditions (top) and with a 50% reduction in NO_x emissions (bottom). These plots are for July 8, 2002 at 5pm EDT. Observations are shown in white circles.

Conclusions

❖ In the absence of quantitative estimates of the uncertainty in the inputs, Bayesian Model Averaging and observed concentrations can be used to significantly reduce bias and improve model skill in predicting the probability of exceeding a threshold concentration.

❖ Likewise, in the absence of observed values, CMAQ-DDM-3D can be used to efficiently generate ensemble members and improve the model's skill.

❖ These probabilistic methods allow air quality managers to quantitatively compare the relative risks and benefits of air quality control options and to select the emissions control strategy that has the largest probability of success.

Future Directions

❖ Use emissions data by sector to create a more realistic range of input parameters.

❖ Use observed NO_2 data (in addition to ozone data) to help weight ensemble members.

❖ Extend the methodology to $PM_{2.5}$.

❖ Use to assess uncertainty in regional downscaling runs and provide insight when designing simulations for climate impact studies.

Future Impact

❖ **Model evaluation**: Use the information to uncover cases and locations that cannot be explained using the known range of model inputs.

❖ **Risk management tool**: Provide probabilistic information for epidemiological studies and exposure models to estimate air quality impacts on sensitive sub-groups and individuals with lifestyles or work environments in high concentration areas.

❖ **Risk Mitigation tool**: Use credible estimates of uncertainty to assess the risks and benefits of control strategies (e.g. Do some emission control strategies deliver larger air quality benefits with lower uncertainty?) Include analysis as part of the "weight of evidence" involved in the SIPs process.

COLLABORATORS:

H. Christopher Frey, North Carolina State University

Haluk Özkaynak, USEPA/ORD/NERL/HEASD