3.17 USING CURRENT-GENERATION METEOROLOGICAL/PHOTOCHEMICAL MODELING SYSTEMS FOR REAL-TIME OZONE FORECASTING

Christian Hogrefe*, S. Trivikrama Rao, and Kenneth L. Demerjian
Atmospheric Sciences Research Center, University at Albany, State University of New York, Albany, NY

1. INTRODUCTION

Until recently, grid-based photochemical modeling systems have been used predominantly to simulate historic high-ozone events (typically lasting a few days), i.e. for hindcasting purposes. Recent developments in computing technology have now made it possible to apply these modeling systems for providing real-time air quality forecasts (McHenry et al., 2000; Chang and Cardelino, 2000). However, since real-time numerical air quality modeling is still in its infancy, operational air quality predictions by federal, state and local agencies are based on a combination of weather predictions, statistical analyses, and expert judgment (Ryan et al., 2000; Dye et al., 2000), i.e. traditional techniques. Cardelino et al. (2001) describe a forecasting program for Atlanta, Georgia that incorporated both traditional techniques and a photochemical model. Numerical models can provide higher spatial and temporal resolution than the traditional methods, but it is necessary to evaluate the quality of these predictions and estimate the modeling systems’ uncertainty before photochemical models can be more widely used for real-time air quality predictions.

In this paper, ozone predictions from a hindcast simulation for the summer of 1995 generated by the RAMS/UAM-V modeling system (Sistla et al., 2001a) are first analyzed to establish a “best case” scenario for model performance (in hindcast simulations, meteorological observations are routinely assimilated using 4DDA techniques). We compare the forecasting skill of this modeling system to air quality forecasts generated by the traditional methods (e.g., statistical analyses, weather forecasts, expert judgment). We then discuss the presence of inherent uncertainty associated with the outputs of grid-based models, and present a method to transform the deterministic model predictions into a probabilistic form that takes into account known sources of model uncertainty. The method is applied to the summer 1995 hindcast simulation as well as a 2001 real-time air quality forecasting pilot project described in a companion paper (Cai et al., 2002).

2. MODEL DESCRIPTION AND DATA BASE

For this study, ozone observations for the summer of 1995 were extracted from the United States Environmental Protection Agency’s (EPA) Aerometric Information Retrieval System (AIRS) database. Preliminary ozone observations for the summer of 2001 were obtained from the EPA’s AIRNOW system for the purpose of model evaluation (R. Wayland, EPA, personal communication). As stated above, we utilize the three-months hindcast model simulation carried out for the time period June 4 – August 31, 1995 covering the eastern United States to establish a best-case scenario and assess model uncertainty. The photochemical model used was the Urban Airshed Model - Variable Grid Version (UAM-V) (Systems Applications International, 1995). The meteorological input for this simulation was prepared using the Regional Atmospheric Modeling System (RAMS3b) (Walko et al., 1995). Details on this simulation have been described elsewhere (Sistla et al., 2001a,b; Rao et al., 2000; Biswas et al., 2001; Biswas and Rao, 2001; Hogrefe et al., 2000, 2001a,b). In addition, we analyze ozone concentrations predicted by a real-time numerical air quality forecasting pilot study for the summer of 2001. In this project, meteorological fields from both MM5 (Grell et al., 1994) and SKIRON (Nickovic et al., 2001; http://forecast.uoa.gr/charactnew.html) were used to drive the CAMx (Environ, 2000) photochemical model. In this study, we only analyze predictions from the MM5/CAMx system. Details on the model setup for this pilot study can be found in a companion paper (Cai et al., 2002). For the analysis presented here, model predictions are extracted for the grid cells that contain observational sites.

3. METHOD OF ANALYSIS

In addition to statistical techniques commonly used to evaluate hindcast simulations (such as mean normalized bias error, mean normalized gross error, systematic and unsystematic error), air quality forecasts are often evaluated through so-called “Forecast Verification Statistics” as defined in U.S. EPA (1999). These statistics measure the forecasting skill in terms of correctly/incorrectly predicting concentration levels above/below a certain threshold and are presented in Table 1.

We computed the value of these metrics for the hindcast simulation in order to compare them to the performance of current operational air quality forecasting techniques reported in other studies.

To estimate the uncertainty of ozone predictions from air quality models, we utilize the model evaluation results presented in Hogrefe et al. (2001a,b). A spectral decomposition technique was used in their studies, and we apply the same technique to estimate the magnitude of fluctuations on time scales that are not captured by...
the modeling systems. Further details on this technique can be found in Hogrefe et al. (2001a,b).

4. RESULTS AND DISCUSSION

4.1 Comparison of Model Predictions to Current Operational Air Quality Forecasts

In a past evaluation study for regression-type forecasting (Dye et al., 2000) in California, values of 125 ppb for 1-hr ozone concentrations and 85 ppb for 8-hr ozone concentrations were used as thresholds to distinguish between events/non-events for the statistics listed in Table 1. This study reported values of ~85-90% for accuracy, 1.0-1.6 for bias, ~70% for the probability of detection, and ~40% for the false alarm rate. Dye et al. (2000) further reported that these values were comparable to those for a similar ozone-forecasting program established for the Mid-Atlantic Region (Ryan et al., 2000). It should be noted that these statistics were based on a comparison of a predicted county-wide daily maximum concentration and the actual daily maximum concentration observed anywhere within the county.

In an initial evaluation of the MM5/MAQSIP real-time air quality forecasting project, McHenry et al. (2000) reported a critical success index of ~30% for a threshold of 70 ppb for 1-hr ozone concentrations. They did not report values for the other statistics or for higher thresholds.

Table 2 presents the forecast statistics for the 3-months RAMS/UAM-V hindcast simulation. In this analysis, observed and predicted daily maximum ozone concentrations were paired in space, i.e. Observations from roughly 500 monitoring locations in the northeastern United States were compared to model predictions at the corresponding grid cells. Therefore, this analysis is somewhat more stringent than the one reported in Dye et al. (2000) who performed the evaluation on a county-wide basis. In addition to using 125/85 ppb for daily maximum 1hr/8hr average ozone concentrations as thresholds distinguishing events/non-events, the dependence of the forecast statistics on the choice of the threshold is also examined. Additional threshold levels of 65 ppb, 85 ppb, and 105 ppb were also chosen for this purpose. For the Air Quality Index (AQI), which is based on 8hr average ozone concentrations, the thresholds separating “good” from “moderate”, “moderate” from “unhealthy for sensitive groups”, “unhealthy for sensitive groups” from “unhealthy”, and “unhealthy” from “very unhealthy” are 65 ppb, 85 ppb, 105 ppb, and 125 ppb, respectively.

Table 2 illustrates that evaluation statistics depend

### TABLE 1A.

<table>
<thead>
<tr>
<th>Scheme used for defining the quantities “A” – “D” used in Table 1b.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictions</td>
</tr>
<tr>
<td>No Exceedance</td>
</tr>
<tr>
<td>Exceedance</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>No Exceedance</td>
</tr>
<tr>
<td>A (Model correctly predicted no exceedance)</td>
</tr>
<tr>
<td>B (Model predicted an exceedance that did not occur)</td>
</tr>
<tr>
<td>Exceedance</td>
</tr>
<tr>
<td>C (Model failed to predict an exceedance that occurred)</td>
</tr>
<tr>
<td>D (Model correctly predicted an exceedance)</td>
</tr>
</tbody>
</table>

### TABLE 1B.

Forecast evaluation metrics as defined in EPA (1999).

<table>
<thead>
<tr>
<th>Accuracy (%)</th>
<th>Percent of forecasts that were correct</th>
<th>100 * (A+D)/(A+B+C+D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Alarm Rate (FAR) (%)</td>
<td>Percent of forecasted exceedances that did not occur</td>
<td>100 * B/(B+D)</td>
</tr>
<tr>
<td>Probability of Detection (POD) (%)</td>
<td>Percent of observed exceedances that were forecasted correctly</td>
<td>100 * D/(C+D)</td>
</tr>
<tr>
<td>Critical Success Index (CSI) (%)</td>
<td>Measures how well high ozone events are predicted (not influenced by number of correct non-exceedance forecasts)</td>
<td>100 * D/(B+C+D)</td>
</tr>
<tr>
<td>Bias</td>
<td>Number of forecasted exceedances / Number of observed exceedances</td>
<td>(B+D)/(C+D)</td>
</tr>
</tbody>
</table>
critically on the threshold considered. For higher thresholds, the accuracy of the (exceedance/no exceedance) predictions increases, but the false alarm rate also rises while the probability of detection and the critical success index decrease. Compared to the regression-type forecasting, the photochemical models show a comparable or higher accuracy, a comparable bias for lower thresholds but higher bias for higher thresholds ("moderate" to "unhealthy for sensitive groups" threshold), and a lower probability of detection and higher false alarm rate.

In summary, it appears that operational air quality forecasts outperform photochemical models in terms of the metrics listed in Tables 1a and 1b. This is consistent with the findings reported in Cardelino et al. (2001). However, the evaluation of air quality forecasts should not be limited to a small set of statistical measures. For example, the potential ability of the photochemical models to provide valuable information about the temporal and spatial evolution of ozone plumes cannot be quantified by such exceedance-based statistical measures. On the other hand, it becomes clear that it is essential to quantify the uncertainty in air quality forecasting. In the following section, we suggest a method to estimate the uncertainty associated with the model predictions.

4.2. Quantifying the Uncertainty of Model Predictions

Several recent studies investigated the uncertainty of ozone predictions from photochemical modeling systems. Using Monte Carlo simulations, Hanna et al. (2000) found that ozone predictions have a typical uncertainty of about 60 % due to the uncertainty in input parameters to the photochemical model. Biswas and Rao (2001) and Ku et al. (2001) found that different treatment for meteorological modeling introduces an uncertainty of 20-30% to the predictions of individual daily maximum ozone concentrations. Hogrefe et al. (2001a,b) have used time series analysis to evaluate photochemical modeling systems. In their studies, they introduced the concept of ‘inherent’ and ‘reducible’ uncertainty. The ‘inherent’ uncertainty was defined as the inability of the grid-based models to capture the observed fluctuations that are caused by processes acting on scales that are not resolvable with the grid cell size used in the model simulation. ‘Reducible’ uncertainty arises from imperfect scientific understanding on how to best describe certain atmospheric processes that can be resolved by the models (e.g. cloud parameterization and prediction, the uncertainty about the proper parameterization for the PBL evolution in mesoscale models, inadequacies in the model input data, etc.). It manifests itself in the model-to-model and model-to-observation differences of ozone concentrations predicted by state-of-science modeling systems using different scientifically-credible process formulations; model users are confronted with this ‘reducible’ uncertainty when applying a model to forecast ambient air quality. While the ‘inherent’ uncertainty is the theoretical lower bound for the model’s uncertainty even for a ‘perfect’ model, the sum of ‘inherent’ and ‘reducible’ uncertainties is still a lower bound for the total modeling uncertainty in practical applications (Hogrefe et al., 2001a,b).

As discussed by Hogrefe et al. (2001), model performance is poorest for high-frequency (intra-day) fluctuations and for the day-to-day variations of the diurnal amplitude. Thus, we treat the magnitude of these two components (for a description on how to estimate these components using time scale analysis, see Hogrefe et al., 2001a,b) during the afternoon hours (when the daily maximum ozone concentration occurs) as an estimate for the lower bound of uncertainty in predicting the daily maximum ozone concentrations with current-generation models. The magnitude is measured through the standard deviation ("sigma") of the sum of the observed intra-day and diurnal components during afternoon hours (1200 - 1700 EST), reflecting the model’s inability to capture the day-to-day fluctuations of the magnitude of these components. Thus, we calculated this magnitude ("sigma") at each station for the summer of 1995 (June – August), and then

---

**TABLE 2:** Performance statistics for the 1995 RAMS/UAM-V hindcast simulation for different threshold values separating "exceedance" from "non-exceedance" for daily maximum 1-hr and 8-hr ozone concentrations. The first value in each cell is calculated for daily maximum 1-hr concentrations, the second value is calculated for daily maximum 8-hr concentrations.

<table>
<thead>
<tr>
<th>Threshold value</th>
<th>45 ppb</th>
<th>65 ppb</th>
<th>85 ppb</th>
<th>105 ppb</th>
<th>125 ppb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>85% / 80%</td>
<td>70% / 70%</td>
<td>80% / 85%</td>
<td>90% / 95%</td>
<td>98% / 99%</td>
</tr>
<tr>
<td>FAR</td>
<td>15% / 20%</td>
<td>35% / 50%</td>
<td>60% / 70%</td>
<td>80% / 85%</td>
<td>85% / 90%</td>
</tr>
<tr>
<td>POD</td>
<td>95% / 90%</td>
<td>80% / 70%</td>
<td>60% / 50%</td>
<td>45% / 40%</td>
<td>40% / 35%</td>
</tr>
<tr>
<td>CSI</td>
<td>80% / 75%</td>
<td>55% / 45%</td>
<td>35% / 20%</td>
<td>20% / 15%</td>
<td>15% / 5%</td>
</tr>
<tr>
<td>Bias</td>
<td>1.0 / 1.0</td>
<td>1.1 / 1.2</td>
<td>1.2 / 1.6</td>
<td>1.5 / 2.5</td>
<td>2.0 / &gt;4</td>
</tr>
</tbody>
</table>
performed a spatial interpolation to determine the percentage uncertainty in ozone concentration contributed by these components at each model grid cell. Figure 1 presents a map of this uncertainty in predicting the daily maximum 8-hr concentration due to the models' inability to properly simulate the intra-day and diurnal components. The map illustrates that the uncertainty ranges from 15% to 30% in most areas; higher values tend to occur in the urban areas, while lower values tend to occur in the rural areas. This percentage uncertainty can then be added/subtracted to each model prediction to estimate the range of most likely ozone concentrations. The same analysis can be performed for time series of 1-hr average ozone concentrations to estimate the uncertainty associated with the prediction of daily maximum 1-hr ozone concentrations. It should be noted that this approach assumes that the model is unbiased, i.e. the uncertainty is purely due to the model's inability to simulate certain fluctuations rather than a systematic over- or under estimation. While past studies have shown that this is certainly not the case (Sistla et al., 2001; Hanna et al., 1996), once the presence and magnitude of such a bias has been quantify, it would be easy to empirically adjust for this bias for the purpose of ozone forecasting and, thus, the bias would not be considered 'uncertainty' as defined above.

4.3. Probabilistic Forecasting

In this section, we illustrate how the presence of uncertainty could be conveyed in the presentation of ozone predictions based on numerical models. One way of presenting probabilistic ozone forecasts would be to predict a range of concentrations rather than a single number. To this end, the estimated percentage uncertainty can be added/subtracted to each model prediction as described above.

Figure 2a illustrates how the total percentage of observed daily maximum 8-hr ozone concentrations which falls within the predicted range increases as the amount of added/subtracted uncertainty increases. As stated above, we measure the magnitude of the uncertainty by the standard deviation ("sigma") of the sum of the observed intra-day and diurnal components during afternoon hours (1200 - 1700 EST). As the amount of uncertainty is increased from 1 sigma to 3 sigma, the percentage of daily maximum 8-hr ozone concentrations that fall within the predicted range increases from 50% to 92% for the 1995 RAMS/UAM-V hindcast simulation. If the systematic bias of ozone predictions at each station is removed prior to this
analysis, the other curve in Figure 2a illustrates that for any given level of added/subtracted “sigma”, a larger percentage of daily maximum 8-hr ozone concentrations fall within the predicted range. The results of applying this analysis to the 2001 real-time MM5/CAMx ozone forecast simulation are displayed in Figure 2b. It can be seen that – both for the curves with and without bias adjustment – the percentage of observations that fall within the predicted range at any given level of “sigma” is lower than for the RAMS/UAM-V hindcast simulation. This confirms the hypothesis that the RAMS/UAM-V hindcast simulation should be viewed as a “best case” scenario, and that the performance of real-time forecast simulations is not that good.

In addition, we examined the dependence of the percentage of observations falling within the predicted range as a function of both observed daily maximum concentration and the amount of uncertainty added/subtracted (as measured by “sigma”) to the predictions. The results, presented in Figure 3, illustrate that the percent of observations falling within the predicted range for any given level of “sigma” is almost independent of the observed daily maximum ozone concentration (the bias was not removed in this calculation). In other words, the method of providing a range of predicted ozone concentrations will lead to similar results regardless of the observed concentration except for very low ozone concentrations.

Figure 4 demonstrates the application of this method to the 2001 real-time ozone forecast project (Cai et al., 2002). Time series of observed 1-hr daily maximum concentrations and ranges of predicted concentrations are shown for two monitoring stations in New York State; the Loudonville monitor is at a suburban location, while the Piseco Lake monitor is at a rural location. For these plots, the predicted range was calculated by adding/subtracting one “sigma” from each prediction. Since “sigma” is calculated as percentage uncertainty, the predicted ranges are larger for high ozone concentrations. It can be seen that the general pattern of ozone concentration was captured by the model predictions, and that the observations frequently fall within the predicted range. However, there are many occasions on which the observations did not fall within the predicted range, indicating – as illustrated above – that adding one “sigma” is a too small an estimate for the modeling uncertainty. The observation that the longer-term fluctuations appear to be better captured by the model compared to individual days’ predictions is in agreement with the results of recent model evaluation.
In addition to presenting the ranges for the predicted ozone concentration, we can also assume that – for an unbiased model - the uncertainty of any given predicted value is described by a normal distribution characterized by a certain standard deviation. If we further assume that this standard deviation is the observation-derived quantity "sigma", we can then calculate the probability that a certain ozone concentration will be exceeded based on model predictions. For this purpose, the difference between the predicted concentration and the threshold (say, 84 ppb for 8-hr ozone concentrations) would be calculated at a given station and expressed in terms of “sigma” at this location. If, for example the predicted concentration was 80 ppb and “sigma” at this station was 10%, the difference between predicted concentration and threshold would be 0.5 “sigma”. Assuming a normal distribution centered at the predicted value, this would mean that there is a 31% chance of exceeding the threshold. Using this method, spatial maps of the probability that a certain threshold is exceeded can be constructed.

5. SUMMARY

In this study, the forecasting performance of a photochemical hindcast simulation was evaluated using a set of threshold-based statistical metrics. When comparing the performance of the traditional air quality forecasts with numerical models, regression techniques, 

**FIGURE 3**: Percentage of observed daily maximum 8-hr ozone concentrations falling within the RAMS/UAM-V predicted range as a function of both the amount of added/subtracted uncertainty as measured by “sigma” (described in the text) and a function of the observed daily maximum 8-hr ozone concentration.
weather prediction, and expert judgment seem to outperform the numerical predictions. However, photochemical models can provide useful temporal and spatial information that is not available with the other methods. To this end, it is essential that we quantify the uncertainty associated with model predictions. Therefore, we proposed a methodology that is based on past model evaluation studies and estimated a lower bound of uncertainty for model predictions at each grid cell derived from spectrally decomposed observed ozone time series.

The application of this method was illustrated using both a 1995 hindcast simulation and a 2001 real-time forecast simulation. As expected, the uncertainty in the real-time forecast simulation was larger than that of the hindcast simulation. Predicting a range of possible ozone concentrations rather than a single number takes into account the well-documented inability of photochemical modeling systems to correctly predict the observation at a single point in space and time while still providing useful information about the temporal and spatial patterns of ozone concentrations. The results of this study illustrate that photochemical modeling systems can be important tools for air quality forecasting.

6. ACKNOWLEDGMENTS

This work was supported in part by the New York State Energy Research and Development Agency (NYSERDA) under agreement number 4914-ERTER-ES99 and the U.S. EPA under grant number R826731476.

7. REFERENCES


FIGURE 4: Time series of observed daily maximum ozone concentrations and MM5/CAMx predicted ranges at Loudonville, NY and Piseco Lake, NY. Observations are marked with a “+” sign and connected by lines, predicted ranges are marked by the vertical bars.


