Appendix G-9: Uncertainty in CMAQ and Over-predictions of Future Year Ozone Design Values

Uncertainty in CMAQ and Over-predictions of Future Year Ozone Design Values

March 15, 2007

Jeffrey Stehr, Charles Piety, Dale Allen

Department of Atmospheric and Oceanic Science
University of Maryland
College Park, MD

1. **Why is this analysis important?**
   CMAQ has been evaluated by using measures that reflect its ability to represent average conditions instead of its ability to respond to changes in emissions. This represents a disconnect between how the model is evaluated and how it is used. It also means that CMAQ was developed with its static performance in mind, not its dynamic performance. It is therefore likely that even though CMAQ meets traditional performance measures such as mean error and bias, it will underpredict the magnitude of ozone changes due to emissions changes. This analysis quantifies some of the uncertainties associated with CMAQ predictions and explains why future year ozone will likely be lower than CMAQ predicts.

2. **What questions are answered by this analysis?**
   - Will Maryland attain the 8-hour standard in 2009?
   - What is the evidence that CMAQ underpredicts changes in ozone due to changes in emissions?
   - How large are those underpredictions?

3. **What are the key take-away messages of this analysis?**
   CMAQ will meet performance criteria but underpredict changes in ozone due to emissions changes. CMAQ's predictions are given as ranges based on conservative estimates of uncertainty and representativeness, tempered by an understanding of CMAQ's apparent shortcomings. Other uncertainties are difficult to quantify. Maryland monitors are likely to be cleaner than CMAQ predicts.

4. **What conclusions are reached in this analysis with respect to Maryland's attainment demonstration?**
   Maryland is likely to meet the ozone standard in 2009 with some margin for error. CMAQ's response to changes in emissions is too rigid, so CMAQ will underestimate the magnitude of changes. Even with its demonstrated shortcomings, when applied to attainment modeling exercises, CMAQ predicts Maryland will attain the 8-hour ozone standard by 2012 and CMAQ predicts the entire Northeast will attain the 8-hour ozone standard by 2018.

Abstract

Several different methods have been used to compare the measured effects from historical changes in emissions to those forecast by the CMAQ model, and all affirm the idea that the response of CMAQ is less than that seen in nature. A study of the August 2003 Northeast Blackout shows that the electrical blackout caused a change of at least 7 ppbv ozone, and likely more, while a CMAQ simulation of the same event predicted only a 2.2 ppbv change. Other studies suggest similar CMAQ underpredictions of ozone changes in response to emissions changes. An ongoing study by EPA reveals that the NO_x SIP call likely produced double the benefit predicted by CMAQ. Meanwhile, the State of New Jersey reports that its ozone monitors appear to have reached their projected 2009 ozone concentrations already, three years ahead of schedule. Even when compared to results from within the 2002 ozone season, CMAQ underpredicts daily ozone variability, and shows important performance shortcomings in areas just upwind of Maryland on high ozone days, namely in the Ohio River Valley and central Virginia. CMAQ is therefore likely underpredicting changes in ozone due to emissions changes and underpredicting ozone transport, so its results should be viewed with such an understanding.

In an effort to intelligently represent CMAQ results, uncertainties have been estimated for two types of errors in CMAQ modeling. One source of uncertainty is the likely meteorological variability of future years. This is not to say that 2002 was not representative, but instead that meteorological variability from year to year is well known, and any future projections must account for this to achieve a reasonable margin of safety so a particularly bad future year will not result in numerous exceedances of the standard. A second type of uncertainty comes from the model and its emissions. This uncertainty was partially estimated by examining several different 2009 scenarios and determining the range of possible 2009 ozone values from those scenarios. These two sources of uncertainty do not cover all the possible sources of uncertainty in CMAQ projections, since errors in the inventory, meteorology, and model formulation all play a role, but are significantly more difficult to estimate. Most of the remaining errors are systematic, not random, in nature, so they should not be accounted for by expanding the uncertainty in the future year projection, but by altering CMAQ's prediction of future year ozone accordingly. As demonstrated earlier, CMAQ is known to underpredict change, likely underpredicting change by 100%. To account for CMAQ's underprediction, CMAQ changes were increased by a conservative 50%, and probable future ozone levels were calculated, along with probable ranges to account for uncertainty in representativeness and some model errors. The picture for the future is that the best estimates of 2009 ozone concentrations are very much like the 2012 concentrations modeled directly by CMAQ. This is in line with current observations from New Jersey (home to the highest ozone in the Northeast) that show modeled 2009 ozone concentrations turning up in 2006.

Introduction

Air quality models have generally been evaluated by their ability to reproduce average concentrations. Statistical measures such as mean error and bias show good performance when the average or median is well captured. Some performance measures that emphasize peak pollution performance have also been used, but seldom if ever has the model's ability to reproduce a change in pollution in response to a change in emissions been evaluated. Rapid changes in emissions are rare, so relatively few opportunites for such comparisons have existed. Most of the changes in ozone from day to day are caused by meteorology, with emissions remaining relatively constant. Because static performance measures have been used in the past, CMAQ's development has been geared toward improving these performance measures instead of toward improving CMAQ's ability to respond to changes in emissions. Therefore, it is entirely possible, indeed likely, that CMAQ would pass traditional performance criteria while underpredicting changes in air pollution in response to a change in emissions. In this section, we use several independent evaluations to demonstrate, and to some extent quantify, CMAQ's response to changes in emissions. Since CMAQ has been evaluated against, and therefore is developed with an eye toward, performance measures that emphasize the mean (e.g. mean bias, error, median, etc.) its dynamic response and its ability to respond to a change in emissions have not been evaluated, nor have they been incorporated into the development of the model. CMAQ, therefore, can do a good job of meeting traditional performance criteria that emphasize average performance while still failing to capture ozone extremes and ozone in certain areas.

Several studies combine to suggest that CMAQ, and likely photochemical models in general, underpredict the change in ozone concentrations that result from a change in NO_x emissions, particularly those from power plants. In this Appendix, the reasons behind the idea that CMAQ underpredicts change are explored, and then a plausible range of future design values is determined, examining several sources of variability and uncertainty.

The 2003 Northeast Blackout is examined from three observational perspectives and compared with the lone air quality modeling study to date to show that CMAQ falls short of simulating the changes expected on this day. Next, an upcoming paper (in preparation) is explored that models the effects of the NOx SIP call on air quality, and finds that CMAQ captures roughly half the benefits that were actually observed between 2002 and 2004. CMAQ predictions from the attainment modeling in this SIP are then examined against current trends at ozone monitors in New Jersey (home to the highest design values in the Northeast), showing that the projected 2009 future year design values have already been achieved. CMAQ's performance is then examined in more detail, and its inability to follow the daily range of ozone values is explored. CMAQ also does not perform as well in rural and upwind areas as in cities and suburban areas, especially in central Virginia and the Ohio River Valley, two known source regions for ozone on high ozone days in Maryland. This rural underperformance is likely a reflection of CMAQ's difficulty in handling ozone transport, as discussed in Appendix G-8. Finally, CMAQ's performance is analyzed as a function of NOx emissions from nearby sources, revealing that CMAQ does best when nearby sources are abundant, and worst when few sources are nearby. This again points to CMAQ's tendency to underpredict the magnitude of transported ozone.

Having laid the basis for why it appears that CMAQ will under represent changes in ozone, a second section explores sources of uncertainty in future year design values, and lays out ranges of probable design values. In light of these issues, a single future year design value cannot be accurately predicted. Instead, a range of values is given.

1. Dynamic Evaluation of CMAQ Leads to the Conclusion that CMAQ Ozone Predictions Lack the Dynamic Range Found in Nature

Air Quality Changes Due to the August 14, 2003 Northeast Blackout

On August 14, 2003, a chain reaction triggered the shutdown of much of the generating capacity in the northeastern U.S. and southeastern Canada. On August 14, 2003, just after 4:00 PM Eastern Daylight Time, the power grid of the United States and Canada experienced the largest blackout ever. The blackout affected an estimated 50 million people, and more than 61,800 megawatts (MW) of electrical load was lost in parts of Ohio, Michigan, New York, Pennsylvania, New Jersey, Connecticut, Massachusetts, Vermont and the province of Ontario. The North American Electric Reliability Council (NERC) reported that many units shut down completely at the start of the blackout, and that a cascade of events continued for several minutes until maximum impact. This resulted in over 263 power plants with 531 units in the U.S. and Canada (NERC reports that all of the fossil-fuel units in Ontario were blacked out) being shut down. Investigation of hourly CEM (Continuous Emissions Monitoring) data from the power plant smokestacks indicate that many of the units affected by the blackout reported no heat input at all during this 24-hour period until attempting to start up.

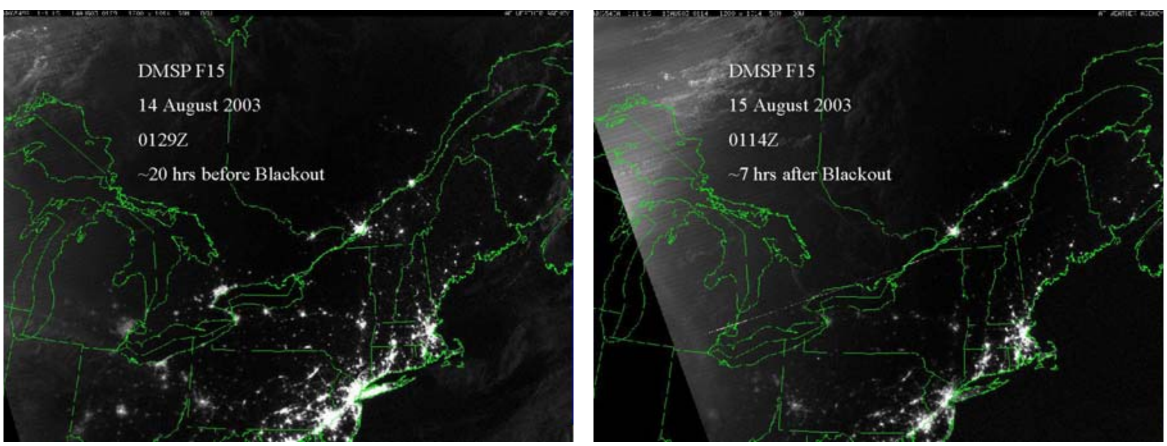


Figure 1. Visible satellite photos twenty hours before the blackout (left) and seven hours after the blackout (right), indicating near total loss of power around Lake Erie and Lake Ontario, and substantial power loss in Detroit and Long Island, NY, among others.

The obvious problem with studying air quality due to a blackout is that many ground-based monitoring stations were without power as well. Fortunately, the UMD aircraft was flying on this day (having been forecast as a high ozone day—a forecast that did not pan out), and was rerouted to Selinsgrove, PA, downwind from the heart of the blackout. Airborne measurements were taken over Maryland and Virginia (outside the blackout area) and Pennsylvania (in the center of the area affected by the blackout) on August 15, 2003, 24 hours into the blackout [Marufu et al., 2004].

This flight, in conjunction with others, provides an excellent opportunity to test the response of air quality models to a large, sudden drop in emissions. Several different comparisons based on observations suggest that the 2003 Northeast Blackout produced a large change in ozone, while the lone modeling study performed to date [Hu et al, 2006] shows much smaller changes in ozone.

Airborne measurements on August 15, 2003, the day after the 2003 Northeast Blackout began, show that ozone was 30 ppbv lower throughout the lowest 1.5 km of the atmosphere on that day, and 38 ppbv lower at ground level, than on a meteorologically similar day, August 4, 2002 (Figure 2).

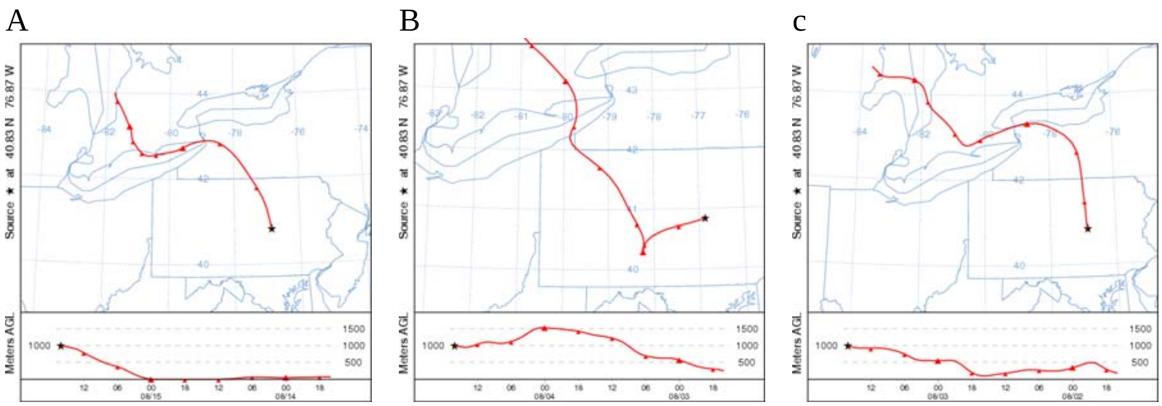


Figure 2. Back trajectories ending at 1000 m over Selinsgrove, PA, in the heart of the area affected by the blackout on a) August 15, 2003, one day into the blackout, b) August 4, 2002, the first reference day, and c) August 3, 2005, the second reference day. The newer reference day is meteorologically more like the day of the blackout, but it occurs after controls mandated by the NO_x SIP Call were installed.

Reference Days for Comparison with the Day of the 2003 Northeast Blackout.

A subsequent comparison to August 3, 2005, which was, from a transport standpoint (Figure 3), more like the blackout day than the day in 2002 was, found smaller changes in ozone. Unfortunately, the August 3, 2005 flight took place after the large reductions from the NO_x SIP Call had been put in place, so the comparison is not as direct as one would like. Comparisons with the August 3, 2005 reference day therefore represent changes that are smaller because the NO_x SIP Call and other emissions reduction programs reduced ozone on that day, so changes are less dramatic. Regardless, on the 2005 reference day, ozone concentrations were 7 ppbv higher than on the blackout day. Since this flight took place after NO_x SIP Call controls were installed, the actual change in ozone due to the blackout must have been substantially larger. Therefore, the actual baseline must be higher than the ozone profile measured on August 3, 2005.

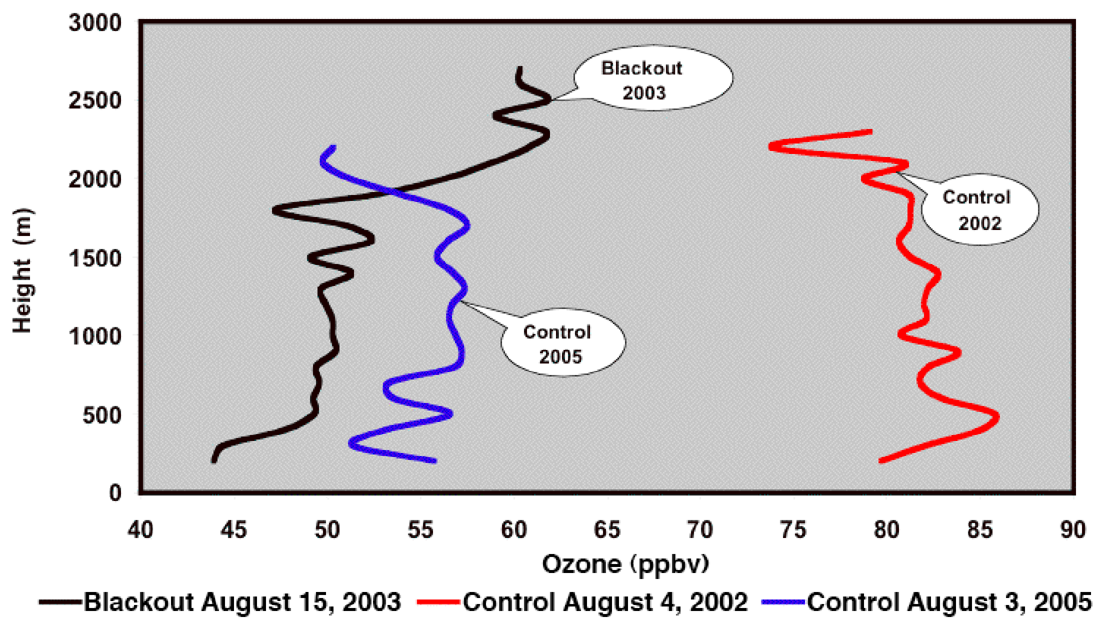


Figure 3. Ozone profiles taken from the UMD aircraft on the day of the blackout (leftmost, in black), the 2005 control day (middle, in blue), and the 2002 control day (rightmost, in red). As shown in Figure 2, the meteorology on the reference day in 2005 was more like that on the day of the blackout, but 2005 was also after the installation of controls from the NO_x SIP Call, so the smaller effect is in part due to controls installed between 2002 and 2005.

Comparisons with Air Quality in Areas not Affected by the Blackout.

Parts of the Northeast were not affected by the 2003 blackout. On the day of the blackout, the University of Maryland aircraft flew through those areas in addition to areas (e.g. near Selinsgrove, PA) that were affected by the blackout. When compared with observations taken outside the area of the blackout, ozone was minimally 10 ppbv lower, and more likely 20-25 ppbv lower in the area around Selinsgrove. The areas around Selinsgrove are indicated in Figure 4 by dashed lines leading from “SEG”, the symbol for Selinsgrove’s airport.

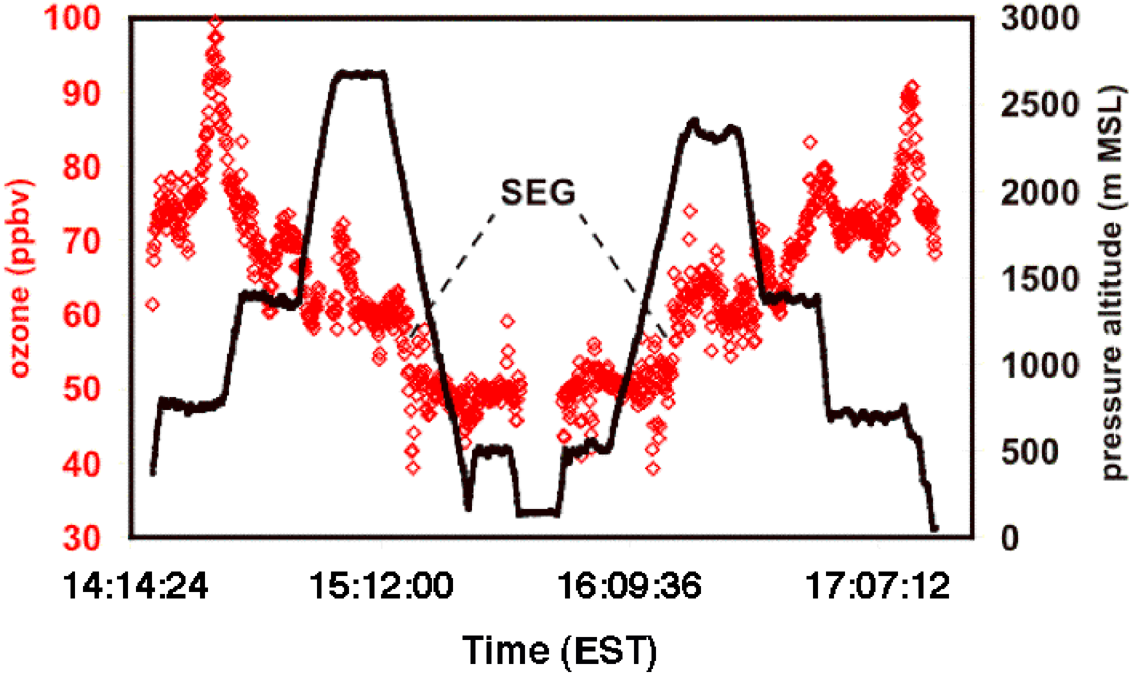


Figure 4. Ozone (diamonds, in red) and aircraft altitude (black line) over the course of the flight from Cumberland, MD to the airport in Selinsgrove, PA (SEG), and on to Ft. Meade, MD one day into the August, 2003 blackout. The area indicated by the lower ends of the dashed lines is within the area affected by the blackout, where ozone was approximately 50 ppbv. Outside the area affected by the blackout, though not at the same altitude, ozone is roughly 60 ppbv, and well outside, ozone is roughly 75 ppbv. The effect of the blackout therefore ranges from a conservative estimate of 10 ppbv to 20 or 25 ppbv.

Comparisons Between the Blackout and Climatological Ozone from Past Flights.

To examine the possibility that the day of the blackout was a low-ozone day, but one well within normal ranges, the data from the blackout flights have been compared to the climatological data set of all UMD aircraft flights. The University of Maryland has flown hundreds of flights throughout the Northeast on high ozone days. Back-trajectories (48 hour) were calculated for all the spirals flown by the UMD aircraft and grouped into clusters, with each cluster therefore representing similar meteorological conditions. When compared with flights falling within the same cluster, the ozone observations taken below ~2000 m over Selinsgrove during the blackout are the lowest of all past observations. Air quality observations on the blackout flight therefore represent a highly unusual day. The climatology of past ozone flights is indicated below in Figure 5, with the ozone profile also superimposed.

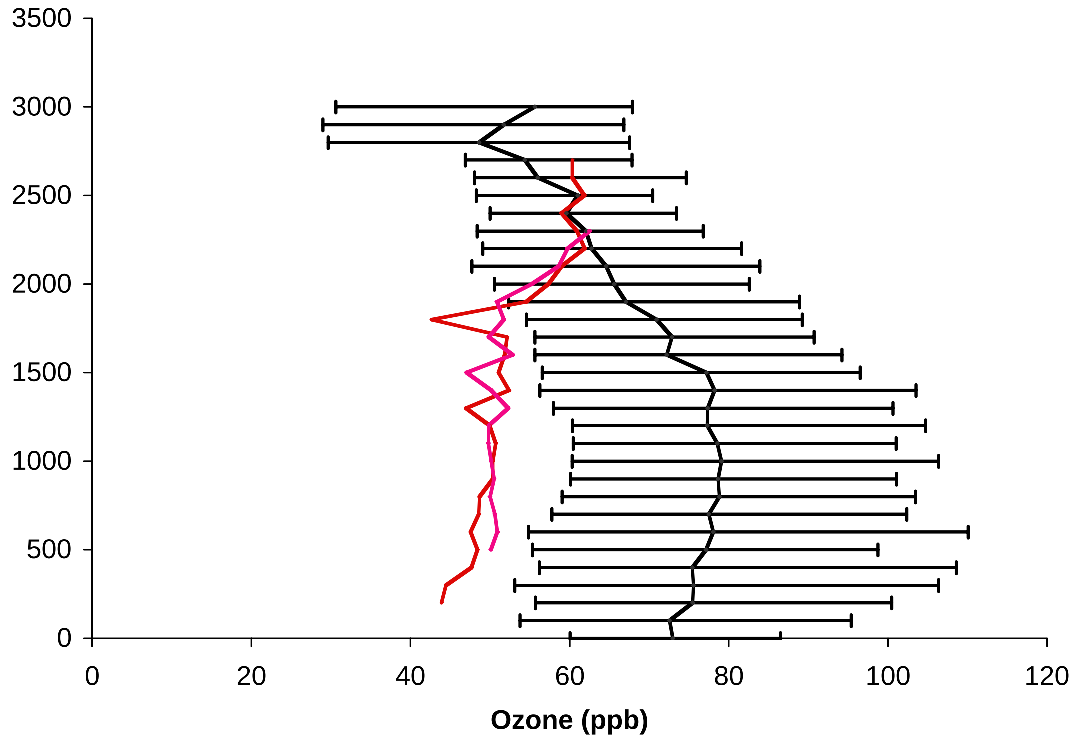


Figure 5. Climatology of ozone flights. Back trajectories were clustered, and all 62 flights in this cluster are compared against the blackout profiles (solid lines in color) of ozone taken over Selinsgrove on August 15, 2003. Data from all 62 profiles were grouped into 100 m altitude bins, and analyzed statistically. The median is shown by the solid vertical black line, and the 10th and 90th percentiles for all flights in this cluster are shown as horizontal error bars. The blackout flight (one spiral upward and one downward) stands out as distinct, and those data points below 2000 m represent the lowest ozone of all flights in this cluster.

Comparison with the CMAQ Model.

These three comparisons (ozone on the blackout day vs. two reference days, ozone inside the area affected by the blackout vs. outside, and blackout profiles over Selinsgrove vs. climatology) suggest that the blackout produced a substantial change in ozone over central Pennsylvania. Unfortunately, these changes are not reflected in CMAQ modeling. Hu et al. [2006] used the Direct Decoupled Method (DDM), which their group embedded in CMAQ, to calculate changes in ozone on the blackout due to a variety of sources from several regions. DDM calculates sensitivities to emissions changes as well as the usual pollutant concentrations that are calculated in CMAQ. Using this method, Hu et al. calculated the sensitivity of ozone to changes in power plant emissions and another sensitivity to changes in mobile source emissions, calculating an ozone reduction of only 4.2% (2.2 ppbv) to power plant NO_x, and very little change due to changes in mobile source NO_x, which is far less than the response calculated by any of the above observation-based methods. The mobile source NO_x emissions changes used in this modeling exercise were not calculated by any direct means, but determined by assuming that weekend travel patterns existed on the day of the blackout. This simulation was performed using an updated version of CMAQ 4.3 with the SAPRC99 chemical mechanism, which is thought to be more responsive to changes in emissions than the CB4 chemical mechanism that has been used for OTC/MANE-VU modeling, owing to its more detailed chemistry. Since even the smallest calculated change in ozone (from observational methods) was seven ppbv (14%), and the changes were likely considerably higher (e.g. exceeding 30%), it appears that the model is underrepresenting the change in ozone due to changes in emissions. The comparison of changes simulated in the model to changes deduced from a number of observation-based analyses suggests that the model is not appropriately capturing the response in ozone due to changes in power plant emissions.

EPA Modeling Study of the Effects of the NO_x SIP call.

EPA, in association with several university colleagues, is performing a CMAQ simulation of 2002 and 2004 summertime air quality to determine the benefits of the NO_x SIP call [Gilliland et al., manuscript in preparation, 2007]. Since this represents a large change in emissions over a relatively short time, the benefits should be noticeable. Emissions and meteorology were simulated for both 2002 and 2004. The final results are not out yet, and the paper is currently in preparation, but preliminary results suggest that though observed median 8-hour ozone levels changed by about 18 ppbv, the CMAQ model only simulates a change of 8 ppbv. It is possible that the model's lack of responsiveness might be due to some quirk of the meteorological simulation, whereby the meteorological simulation was in error in such a way as to render the photochemical simulation unresponsive to emissions changes or that the meteorological simulation could have produced an artificial environment that would have yielded higher ozone in reality, but these results suggest that the CMAQ model underpredicts changes in ozone, especially where power plant emissions are concerned. The only mobile emissions changes that would have occurred from 2002 to 2004 would have been from the modest vehicle fleet turnover in those years.

Critical monitors in New Jersey have already reached their 2009 design values in 2006.

Analysis of the latest 8-hour ozone design values (based on observations from the summers of 2004, 2005, and 2006) for the Colliers Mill ozone monitoring location in central New Jersey show that this monitor has already reached its projected 2009 design value in 2006 [R. Papalski, NJDEP, personal communication]. These calculations use a base year of 2002, so 2006 is only four years into the seven-year period, and New Jersey DEP calculates that automobile fleet turnover alone will still generate some substantial reductions in NO_x emissions over the rest of the time period. This is a critical location, since its 2002 design value was a colossal 106 ppbv. Its projected 2009 design value is 93, and its current design value is only 92. Furthermore, this is not a fluke result at an isolated monitor. All other New Jersey ozone monitoring locations are now close to their projected 2009 design values (Table 1), and the average value at all of the ozone monitoring locations is the same as the 2009 average projected value. Lastly, none have 2004-2006 design values that are higher than Colliers Mills, so that is still the high ozone monitoring location for the state of New Jersey, and indeed the entire Mid-Atlantic and Northeast.

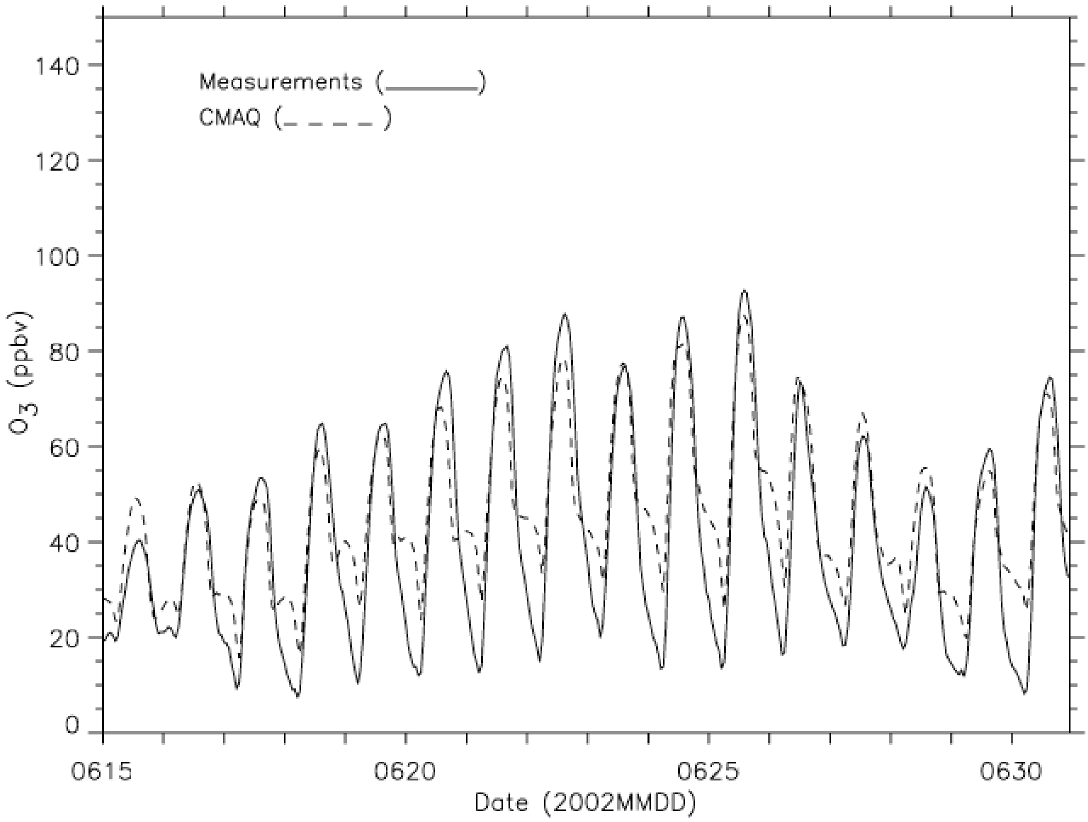
Table 1. Predicted and Measured Design Values in New Jersey

CCOUNTY	OZONE MONITOR LOCATION	2004-2006 DESIGN VALUE (ppbv)	Modeled Predicted O_3 Concentration (BOTW-v.3) (ppbv)
Atlantic	**Nacote Creek**	78	75
Passaic	**Ramapo**	79	79
Monmouth	**West Long Beach (Monmouth Univ)**	81	83
Morris	**Chester**	82	85
Camden	**Camden Lab**	83	89
Bergen	**Teaneck**	85	87
Cumberland	**Millville**	85	82
Gloucester	**Clarksboro**	85	88
Hudson	**Bayonne**	85	80
Mercer	**Rider U**	87	87
Middlesex	**Rutgers U**	88	85
Hunterdon	**Flemington**	88	84
Camden	**Ancora**	88	90
Ocean	**Colliers Mills**	92	93
	Average (ppbv)	**85**	**85**

CMAQ Does Not Capture the Diurnal Cycle, and Therefore the Dynamic Range of Ozone Values.

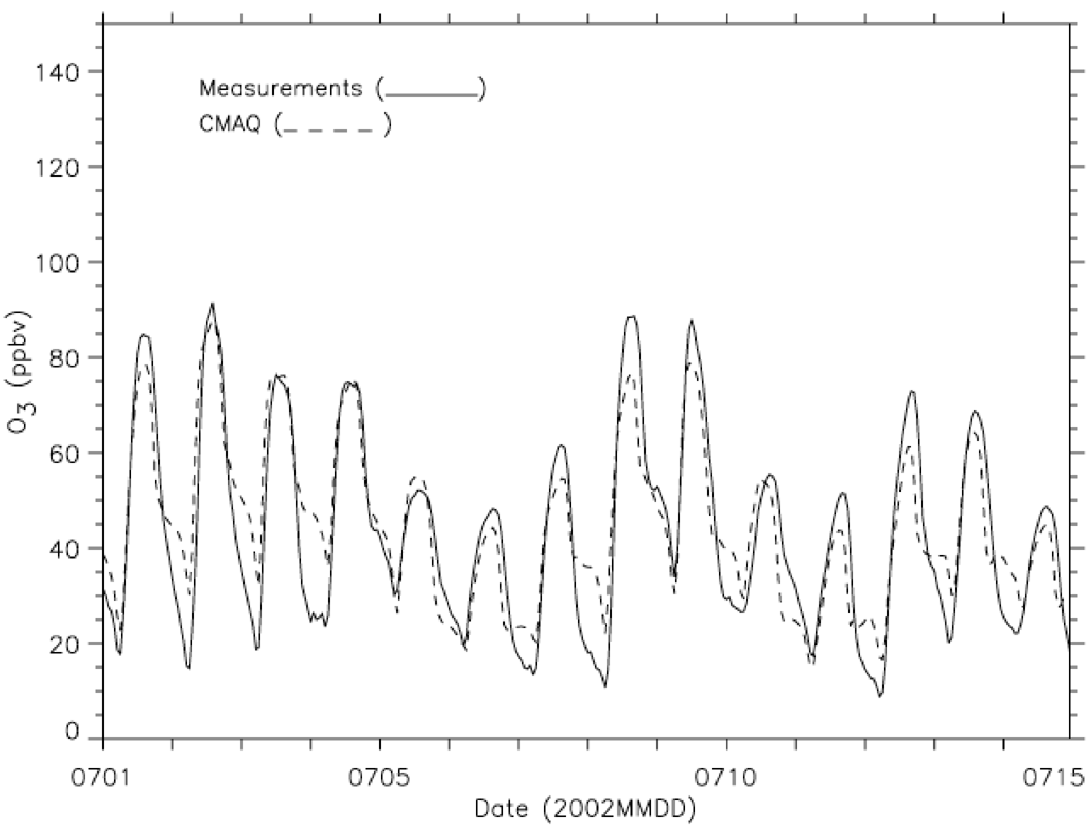
CMAQ's response lacks the diurnal and longer-time scale cycles that are in the observations. CMAQ generally falls short of peak ozone and overpredicts low ozone at night. These shortcomings likely have specific causes (e.g. NO_x titration has been blamed in the past for failures in predicting ozone at night) but those specific causes generally add up to a lack of response to change in the model. Figure 6 below compares CMAQ predictions with surface ozone measurements taken across Maryland, Washington, D.C., Virginia, and the Philadelphia nonattainment area. CMAQ consistently falls short of peak ozone values, especially on the highest ozone days. CMAQ's performance is therefore at its worst when it is needed the most. As will be shown in the following section, CMAQ's performance is also worse in the geographic locations where it is most needed.

a)



Measurements (______)
CMAQ (_ _ _ _ _)
0
20
40
60
80
100
120
140
O_3 (ppbv)
0615
0620
0625
0630
Date (2002MMDD)

b)



Measurements (______)
CMAQ (_ _ _ _ _)
0
20
40
60
80
100
120
140
O_3 (ppbv)
0701
0705
0710
0715
Date (2002MMDD)

c)

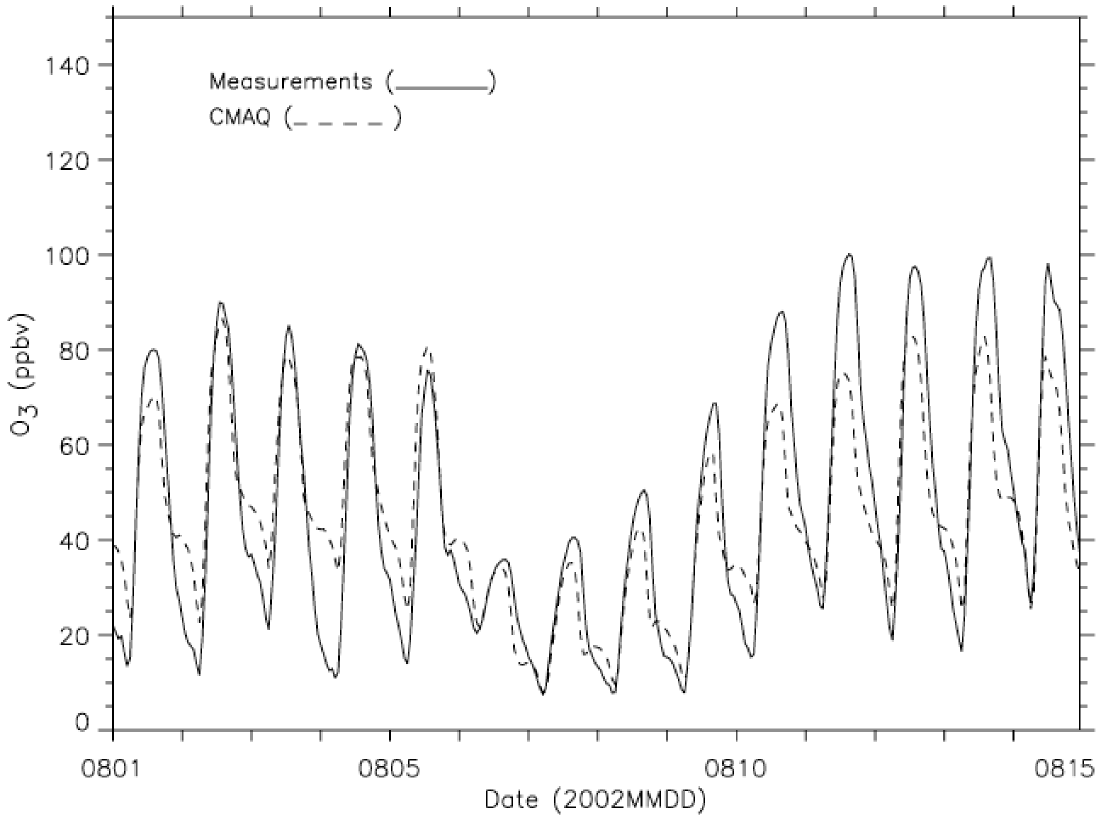


Figure 6. Ozone CMAQ simulations (dashed line) and measurements (solid line) across Washington, D.C., Maryland, Virginia and the Philadelphia nonattainment area. Observations were averaged across all monitor locations in Washington, D.C., Maryland, Virginia, and the Philadelphia non-attainment area. Model outputs were sampled from the grid cell containing each monitored location and also averaged. Ozone is consistently overpredicted by CMAQ (dashed line) at night and underpredicted during the day. CMAQ's low bias is especially pronounced on high ozone days. a) June 15-30, 2002 b) July 1-14, 2002 c) August 1-15, 2002.

Performance in Upwind and Rural Areas is Worse than in Urban and Suburban Areas.

As mentioned above, CMAQ's low bias on high ozone days is larger (more negative) at rural sites than in suburban or urban areas. This topic has already been covered in Appendix G-8, so only a brief review will be given here. Suburban performance is somewhat poorer than that in urban areas, though not as bad as in rural areas. Since these rural areas are generally upwind of sites with high design values, CMAQ is underpredicting transport in crucial areas at crucial times. Furthermore, CMAQ's performance is poor in the Ohio River Valley and central Virginia, both of which are often upwind of Maryland on high ozone days. To further illustrate this point, the correlation coefficient, *R*, was calculated by performing a least squares fit between ozone observations and CMAQ predictions for each monitoring location. These correlation coefficients were then compared with emissions data for point sources within an 80 km radius of each observing site. In this part of the country, and on this scale, point source emissions are correlated with emissions from automobiles and other sources, so areas with fewer point source emissions also tend to have fewer emissions of other types. The results are presented in Figure 7, and show that CMAQ performs poorly in areas with few local emissions. As has been found earlier, CMAQ has some trouble in properly capturing transported ozone, and this is likely another reflection of that problem.

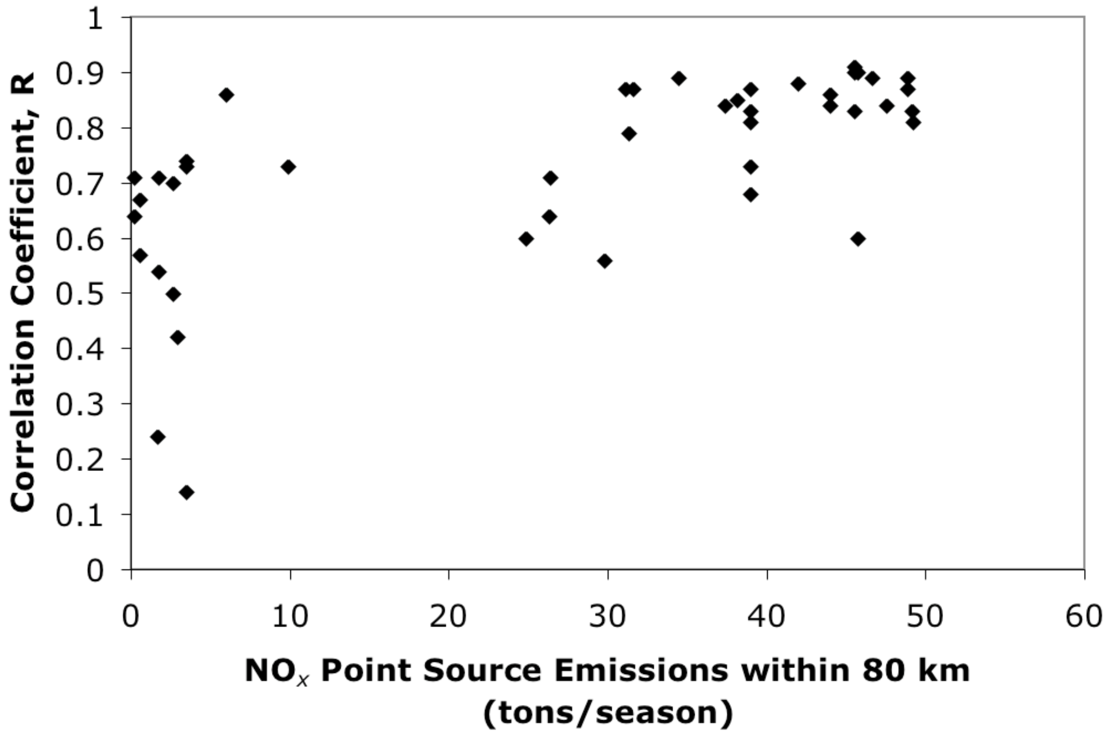


Figure 7. Correlation coefficient between surface ozone observations and CMAQ ozone predictions vs. ozone season point source emissions of NO$_x$ within an 80 km radius of the ozone monitor. Point source emissions in this region tend to be co-located with other emissions, and serve as a proxy for total emissions. Performance at sites with lower nearby NO$_x$ emissions is notably poorer than in areas where NO$_x$ is relatively abundant. Since areas with lower local nearby NO$_x$ tend to be upwind rural areas, CMAQ is underperforming in upwind areas.

2. Uncertainties and Probability in Future Year Design Values.

Two sources of uncertainty in projecting future year design values are relatively straightforward to quantify: representativeness of 2002 design values as indicators of current air quality, and uncertainty in the range of future year design values as calculated from several different simulations of base year and future year design values (e.g. similar runs with different versions of the model and different versions of the inventory as well as a run with small changes in the future year emissions would all produce somewhat different results, but would not represent the changes expected from the implementation of a major control program). Other sources of uncertainty include model formulation and the degree to which the meteorological fields represent actual conditions in 2002. These additional sources of uncertainty are extremely difficult to quantify, since they represent unknowns in the formulation of the model or sensitivity to differences between modeled and actual meteorology that cannot easily be determined. Therefore, the approach taken here is to generate reasonable estimates of uncertainty (note that this is different from model *error*) due to the things that are readily quantified and to modify them only slightly based on the preceding discussion.

Range of Base Year Design Values.

EPA's recommended procedure for calculating design values calls for creating 3-year averages of the fourth highest 8-hour average ozone reading for the individual years. Since there is some variability in these 3-year averages, EPA further recommends averaging three such values from successive years to obtain a design value that is centered on the base year (e.g., for 2002, one would take the 3-year averages from 2000-2002, 2001-2003 and 2002-2004, thereby giving 2000 and 2004 single weight, 2001 and 2003 double weight, and 2002 triple weight in a 5-year average). Since variations in meteorology lead to substantive variations in year-to-year peak ozone values, the degree to which the base year, or any of these 3-year periods, is representative of overall conditions in the area is one source of uncertainty in determining whether or not an area will come into attainment in the future.

Currently, most ozone monitoring locations throughout the Mid-Atlantic and Northeast show improving trends in ozone concentrations over the years that went into the 5-year weighted average, though the design values at some have risen modestly. The average difference between the highest and lowest 3-year design values is 6 ppbv. Similarly, the average standard deviation for each site over this time period is +/- 3 ppbv (using standard deviation only as an estimate of variability and not suggesting that a 3-data point standard deviation is adequate for any individual station). Both these measures suggest that variations in meteorology can reasonably be expected to produce substantial variability in the design values themselves.

It appears reasonable that the representativeness of the current year can therefore produce a variation of 3 ppbv about some central value. Further analysis of these design values shows that the vast majority of sites show substantial improvements in ozone design values over this time period, likely due to the combination of NO_x reductions from the NO_x SIP call and meteorological conditions in 2003 and 2004 that were less conducive to ozone production.

The largest improvement was a 14 ppbv reduction at Colliers Mills, the ozone monitor location with the highest current and future year design value in the entire Northeast, and the worst was a 5 ppbv increase in the Bronx, NY. No Maryland monitors show any increases in design values, with all showing decreases, as shown in Table 2. Beyond the scope of the analysis time period, the decreases continued in Maryland through 2005, and experienced a slight increase in 2006. The summer of 2006 was a very warm summer, with conditions highly conducive to making ozone, so the *slight* increase from 2005 likely represents progress, since ozone levels would have risen somewhat higher if not for the control programs that were going into place.

Table 2. Trends in Design Values at Maryland Monitors

Site Name	AIRS ID	Design Values by Year			Change from 2002 to 2004 (ppbv)
		2002 (ppbv)	2003 (ppbv)	2004 (ppbv)	
Davidsonville	240030014	102	98	94	-8
Ft. Meade	240030019	101	97	93	-8
Padonia	240051007	92	89	85	-7
Essex	240053001	93	93	88	-5
South Carroll	240130001	92	89	85	-7
Fair Hill	240150003	104	98	91	-13
S Maryland	240170010	94	94	91	-3
Frederick Airport	240210037	91	88	83	-8
Edgewood	240251001	104	103	94	-10
Aldino	240259001	100	98	93	-7
Millington	240290002	102	95	89	-13
Rockville	240313001	89	88	83	-6
Greenbelt*	240330002	95	93		*
PG Equestrian**	240338003			94	*
Hagerstown	240430009	87	86	83	-4
*Monitor discontinued in 2003 (Greenbelt) due to loss of permission to use location. **Monitoring began in 2002 at this location.					

Ranges of Future Year Predictions.

Examining the results from similar 2009 scenarios gives insight into how the model responds to changes in emissions. Different future year scenarios have been modeled to support the Maryland SIP modeling effort. Taken together, the variations among these attempts at representing future year emissions give some idea as to the precision, though not the total error in the model, since the largest errors in the modeling are systematic, and will not show up in tests of model and emissions inventory precision. As discussed above, the CMAQ model likely does not respond as much to changes in emissions as the atmosphere does. Therefore, it is not surprising that its future year predictions do not vary much between different emissions scenarios. The scenarios used to examine this source of uncertainty were: OTC base A and base B modeling, VISTAS model outputs at overlapping monitors, and OTC's "beyond on the books on the way" modeling run. The VISTAS modeling represents a different, partially independent attempt at modeling future year design values using somewhat different emissions, different meteorology, and a slightly different modeling platform. The OTC base A and base B cases represent two different versions of the CMAQ model (4.4 and 4.5) and different versions of the base year inventory. The "beyond on the books on the way" run was also examined because its results are similar to the other 2009 future base scenarios, except for a few additional emissions control strategies, so in this context it represents perturbations to the emissions projected for 2009. Examining this run allows one to gain some insight into the sensitivity of the model to errors and uncertainties in the emissions inventories. There are not enough scenarios to generate a proper standard deviation at each site, or to give much meaning to a range generated from those individual predictions, but the average range and the average standard deviation give an estimate of variability between runs at sites across the modeling domain. The average range was +/- 0.83 ppbv, and the average standard deviation was +/- 0.75 ppbv. Splitting the difference gives roughly +/-0.8 ppbv.

This represents the variability to be expected from different attempts at modeling future year air quality, and some of the variability expected from small errors in the emissions inventory. The range of 2009 projections does not represent the full uncertainty in future year results, but the sensitivity of the model to small variations in emissions. Therefore, it represents only part of the uncertainty in the modeled result. Emissions are likely more uncertain than these simple estimates would suggest, with errors in some emissions inventory categories as high as 50% [e.g. Choi et al., 2006].

The two measures above can be combined to give a conservative estimate of the uncertainty in future year projections. Since the representativeness of the base year and variations in future year projections are uncorrelated, standard error propagation techniques can be used, namely by squaring and adding the uncertainties, and taking the square root of the sum to get the combined uncertainty. The uncertainties do not add because they are not correlated, so one is as likely to be positive as the other is to be negative. The combination gives an uncertainty in future year design values of 3.1 ppbv.

The sizes of other sources of uncertainty are considerably more difficult, if not impossible (since they represent unknown aspects of the science of air pollution), to estimate reasonably. Many, such as CMAQ's lack of responsiveness to changes in emissions, are systematic, and therefore cannot reasonably be represented with error bars, but should instead be represented with a shift of the projected value. The discussion in the previous part of this appendix centered on several comparisons between modeled values and measurements around times of change. As discussed, CMAQ underrepresents change, and should be corrected for that tendency. The previous discussion hints that the benefits are likely to be double (or more) what CMAQ predicts, so any benefit from CMAQ should be increased 100%. To allow for considerable margin for error, the likely 100% increase in benefit is cut in half, and changes predicted by CMAQ will be increased by only 50% to account for its lack of sensitivity. Doubling the effects predicted by CMAQ more closely represents the previous discussion, but the more conservative route has been chosen.

Results are presented below (Table 3, Figures 8 and 9) as base year (2002) values and CMAQ predictions for 2009 and 2012, along with a projection of the likely 2009 design value, based on the conservative 50% underresponse estimate, and the upper and lower bounds of that 2009 design value, combining the estimates of uncertainty established before with the estimate of CMAQ's lack of responsiveness.

Table 3. Current and Projected Design Values and Their Uncertainties for Maryland Monitors

Monitor Name	AIRS-ID	Observed	Modeled		Probable		
		2002 Design Value	2009 Design Value	2012 Design Value	2009 Design Value	2009 Lower Bound	2009 Upper Bound
Davidsonville	240030014	98.0	84	78	77.0	73.9	80.1
Ft. Meade	240030019	97.0	84	78	77.5	74.4	80.6
Padonia	240051007	88.7	77	72	71.2	68.1	74.3
Essex	240053001	91.3	80	76	74.4	71.3	77.5
South Carroll	240130001	88.7	75	69	68.2	65.1	71.3
Fair Hill	240150003	97.7	81	75	72.7	69.6	75.8
S. Maryland	240170010	93.0	76	70	67.5	64.4	70.6
Frederick Airport	240210037	87.3	74	68	67.4	64.3	70.5
Edgewood	240251001	100.3	85	80	77.4	74.3	80.5
Aldino	240259001	97.0	82	76	74.5	71.4	77.6
Millington	240290002	95.3	80	74	72.4	69.3	75.5
Rockville	240313001	86.7	76	71	70.7	67.6	73.8
Greenbelt	240330002	94.0	82	76	76.0	72.9	79.1
P. G. County Equestrian	240338003	94.0	81	76	74.5	71.4	77.6
Hagerstown	240430009	85.3	73	67	66.9	63.8	70.0

Table 3 illustrates the best estimates for future design values and their uncertainties. Even with a broad range of uncertainty due to representativeness issues and some model uncertainty, given CMAQ's under-response to changes in emissions, it is highly probable that Maryland will attain the 8-hour ozone standard in 2009, even allowing for adverse weather conditions. The upper bounds are all below 81 ppbv, with most well below. Probable 2009 design values are also very similar to calculated design values for 2012. Therefore, much like the current (2006) New Jersey design values presented in Table 1, which showed that 2009 values had arrived three years early, these projections suggest that 2012 design values will arrive early for Maryland monitors as well. Regardless, by 2012, CMAQ predicts that Maryland will be well in attainment at all of its monitors, with only Edgewood remaining as high as 80 ppbv.

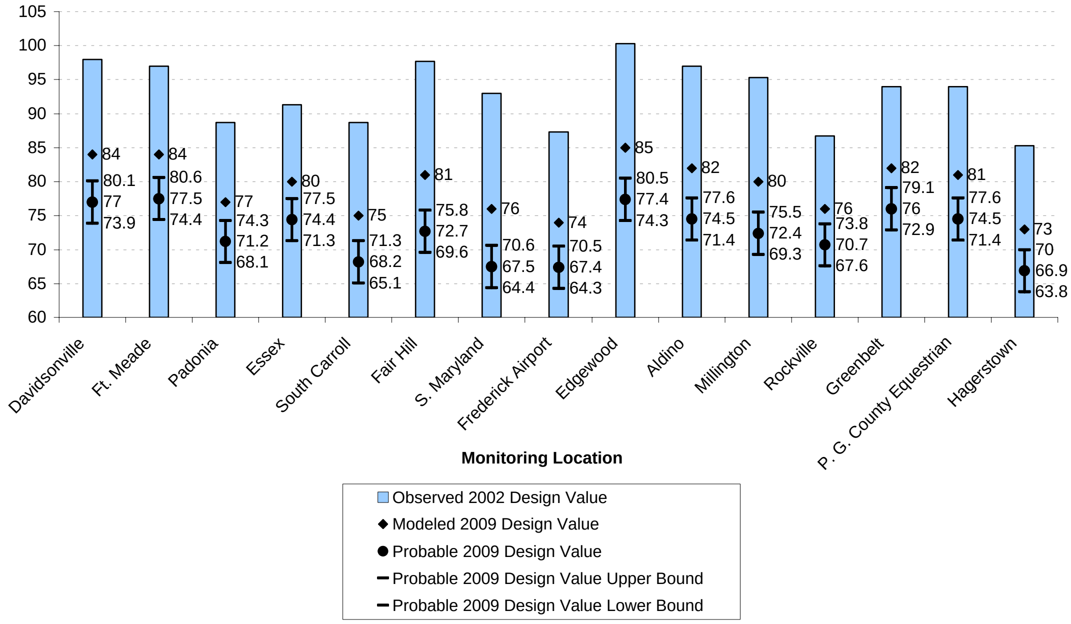


Figure 8. A graphical depiction of the data from Table 3, depicting 2002 base year design values (blue columns), modeled 2009 design values (black diamonds), and the most probable future year design values along with the upper and lower bounds for those future year values (round circles and associated error bars, respectively). All Maryland monitoring locations are shown.

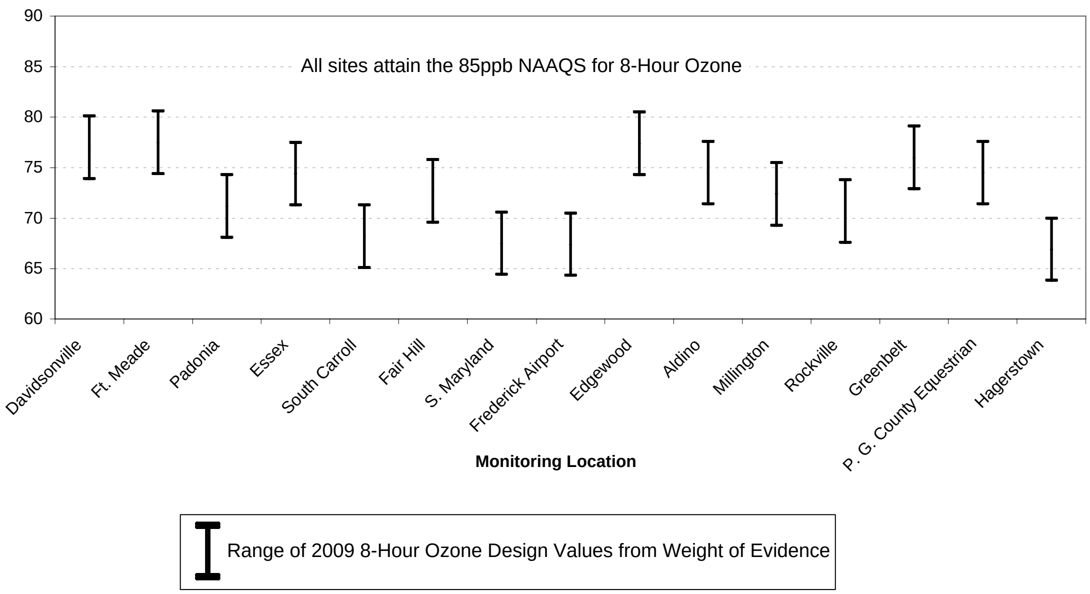


Figure 9. The ranges of probable future year design values from Table 3, given in a graphical format. The lower end of each bar represents the lower bound of the most likely future year design value, while the upper end of each bar represents an upper bound.

Conclusions

Several different methods have been used to compare the measured effects from changes in emissions to those predicted by the CMAQ model, and all affirm the idea that the changes in ozone will be larger (e.g. ozone will be better) than predicted by CMAQ. A study of the 2003 Northeast Blackout [Marufu et al., 2004] shows that the blackout caused a drop of at least 7 ppbv ozone, and likely considerably more, while a modeling study of the same event [Hu et al., 2006] used CMAQ to predict only a 2.2 ppbv change. An ongoing study by EPA reveals that the NO_x SIP call likely produced double the benefit that CMAQ predicted. Meanwhile, the State of New Jersey reports that its ozone monitor locations appear to have reached their 2009 design values in 2006, three years ahead of time. When compared to observations from the 2002 ozone season, CMAQ underpredicts diurnal variability, and shows important performance shortcomings in areas just upwind of Maryland on high ozone days, namely the Ohio River Valley and the state of Virginia. Furthermore, performance on high ozone days tends to be best in urban areas, next best in suburban areas, and worst in rural areas, so CMAQ is underpredicting ozone in upwind areas from which it would enter the largely urban and suburban nonattainment areas.

Uncertainties have been estimated for two types of errors in CMAQ modeling. One source of uncertainty is the range of possible meteorological conditions that might be encountered in future years. This is not to say that 2002 was not representative, but instead that meteorological variability from year to year is well known, and any future projections must account for this to achieve a reasonable margin of safety, so particularly bad future year meteorology will not result in numerous exceedances of the standard. Some of the uncertainty arising from the model and its emissions was estimated by examining several different 2009 scenarios and determining the range of possible 2009 ozone design values from those scenarios. These two sources of uncertainty do not cover all the possible sources of uncertainty in CMAQ projections; errors in the inventory, meteorology, and model formulation all play a role, but are significantly more difficult to estimate. The error estimate and the future year meteorological variability estimate were combined to generate an estimate of future year uncertainty in ozone design values.

To account for CMAQ's resistance to change, CMAQ changes were increased by 50%, and probable future ozone design values were calculated, along with probable ranges of ozone concentrations to account for meteorological variability and some model errors. The resulting picture of future ozone is that likely 2009 ozone design values correspond to 2012 design values calculated directly by CMAQ. This is in line with current observations from New Jersey (home to the ozone monitoring location with the highest design value in the Northeast) that show projected 2009 ozone concentrations turning up in 2006.

Future Work

A considerable ongoing effort is underway throughout the numerical modeling community to determine the proper way to calculate the uncertainty from a computer models. Currently, proper calculations of model uncertainties are difficult to determine, especially when the model is used in a relative sense by calculating changes from a baseline instead of absolute predictions of pollutant concentrations. For photochemical modeling, the largest errors are thought to be systematic, and should be correctable. Meteorological modeling, in contrast, often suffers from random errors and nonlinear behavior, so small changes are amplified. Since the two systems are different (though coupled), methods different from those that have worked so well in meteorological modeling (e.g. ensemble modeling techniques) will have to be developed to assess photochemical modeling uncertainties.

In Appendix G-10, some of the shortcomings of the CMAQ chemical mechanism were laid out, along with some possible suggestions for remedies. One possible course of action would be to perform some sensitivity tests with another model or another chemical mechanism such as SAPRC99, which is more detailed and therefore should be more representative of atmospheric chemistry. Some of the lack of response in CMAQ is likely due to problems with emissions inventories and their processing. Emissions from point sources are likely underestimated on high energy demand days, since peaking units come on line for only a short time in the heat of the day in response to peak demand. Scattering of light by aerosols is known to increase photolysis rates aloft and decrease them near the surface, thereby altering photochemistry. Likewise, clouds alter photochemistry, and remain a very difficult forecasting problem, so they too are likely poorly represented in both MM5 and CMAQ. If clouds and the effects of aerosols are misrepresented, then CMAQ's photochemistry will not perform properly, overpredicting ozone at the surface and underpredicting ozone, and therefore transport, aloft. Finally, the meteorology by itself is another source of uncertainty, and almost certainly plays a role in CMAQ's lack of response.

CMAQ is a state-of-the-art air quality modeling tool, and has proven its utility with impressive capabilities. The purpose of this appendix is not to disparage CMAQ, but to advocate using a tool responsibly, with a proper acknowledgement of its abilities and shortcomings. By assessing the model's performance and examining the implications of the model's shortcomings one can improve upon its predictions. A comparable approach is used in generating the MOS (Model Output Statistics) by comparing observations to outputs from National Weather Service forecast models. The skill of weather predictions increases substantially in going from raw model output to the MOS because MOS takes into account (through a more complex statistical procedure) model tendencies and corrects them by comparing the model to observations. Furthermore, the comparisons made in this and other appendices likely were not possible in the past, since large, sudden reductions in emissions like those from the NO_x SIP Call and the 2003 blackout are relatively rare. The prospect of dynamic model evaluation holds great promise for rapid model development in the coming years.

References

Choi, Y.-J., S. H. Ehrman, R. V. Calabrese, J. W. Stehr, R. R. Dickerson, A Combined Approach for the Evaluation of a VOC Emissions Inventory, *J. Air Waste Manage. Assoc.*, 56:169–178, February, 2006.

Gilliland, A., Personal communication, January 12, 2007; manuscript in preparation.

Hu, Y. M. T. Odman, and A. G. Russell, Re-examination of the 2003 North American electrical blackout impacts on regional air quality, *Geophysical Research Letters*, Vol. 33, L22810, doi:10.1029/2006GL027252, 2006.

Marufu, L.T., B. F. Taubman, B. Bloomer, C. A. Piety, B. G. Doddridge, J. W. Stehr, and R. R. Dickerson, The 2003 North American electrical blackout: An accidental experiment in atmospheric chemistry, *Geophysical Research Letters*, Vol. 31, L13106, doi:10.1029/2004GL019771, 2004.