# EXECUTIVE SUMMARY

Due to violations of the 0.08 parts per million (ppm) 8-hour ozone National Ambient Air Quality Standard (NAAQS) based on 2005-2007 air quality data, in November 2007 the Denver Metropolitan Area (DMA) reverted to an 8-hour ozone nonattainment area. This requires the DMA to develop an 8-hour ozone State Implementation Plan (SIP) that demonstrates the area will achieve the 0.08 ppm 8-hour ozone NAAQS by 2010. The Denver Regional Air Quality Council (RAQC), in consultation with the Colorado Department of Health and Environment (CDPHE) Air Pollution Control Division (APCD) contracted with ENVIRON International Corporation and their subcontractor Alpine Geophysics, LLC to develop the photochemical modeling databases necessary to demonstrate that the DMA will achieve the 0.08 ppm 8-hour ozone NAAQS by 2010.

## OVERVIEW OF APPROACH

The Comprehensive Air-quality Model with extensions (CAMx; www.camx.com) was set up for a June-July 2006 episode on a 36/12/4 km grid with the 4 km domain focused on Colorado. Meteorological inputs were prepared using the MM5 meteorological model whose results and evaluation are discussed by McNally and co-workers (2008). An initial emissions inventory was prepared using the SMOKE emissions modeling system and a preliminary 2006 base case was performed. A preliminary model performance evaluation was conducted and diagnostic sensitivity tests performed to identify an optimal model configuration for simulating ozone formation in the DMA (Morris et al., 2008b).

A revised CAMx 2006 base case (Run 17) simulation was conducted that included the following emission updates from the preliminary 2006 base case simulation reported by Morris and co-workers (2008):

- Corrections and enhancements to the 2006 emissions for Colorado provided by the CDPHE/APCD.
- Application ConCEPT MV mobile source emissions modeling system that uses diurnally varying link-based Vehicle Miles Traveled (VMT), fleet mix and other data to provide more detailed on-road mobile source emissions in the DMA.
- Use of the WRAP Phase III oil and gas production emissions inventory for the Denver-Julesburg Basin.
- Biogenic emission estimates from the MEGAN biogenic emissions model (Guenther and Wiedinmyer, 2004).

This report presents the model performance evaluation for the final 2006 base case simulation performed as part of the Denver 8-hour ozone attainment demonstration modeling.

## MODEL PERFORMANCE EVALUATION

The model performance evaluation of the Denver final 2006 base case simulation performed both an operational evaluation that evaluated how well the model predicted the ozone observations and a diagnostic evaluation that evaluated the model for ozone precursors, key indicator species, particulate matter (PM) and ozone aloft.

**Operational Model Performance Evaluation**

The operational model performance evaluation focused on how well the model predicted the observed surface ozone concentrations in the DMA and included graphical displays of model performance as well as statistical evaluation metrics and comparisons with model performance goals.

Comparison against Model Performance Goals

One element of an ozone SIP model performance evaluation is to test how well the model reproduces the observed ozone concentrations. As part of this, statistical performance metrics are calculated that are compared against model performance goals as a method to gauge model performance, compare it against other studies and to assist in determining whether the model is getting the right answer. EPA's latest modeling guidance (EPA, 2007) emphasizes using graphical and diagnostic evaluation techniques to assure that the photochemical rid model is capturing the correct chemical regimes and emissions sources that lead to the high ozone (i.e., assuring that the model is getting the right answer for the right reason).

EPA's 1991 1-hour ozone guidance included three performance goals that have been used for over two decades to assist in evaluating ozone models as part of the ozone SIP modeling process:

- Unpaired Accuracy of the Peak ≤±20%;
- Normalized Mean Bias ≤±15%; and
- Normalized Mean Gross Error ≤35%.

The Mean Normalized Bias and Gross Error statistical measures are calculated using all predicted and observed hourly ozone pairs matched by time and location for which the observed ozone is 60 ppb or greater.

Figure ES-1 compares the Denver final 2006 base case CAMx hourly ozone model performance against these three EPA model performance goals for each day during the June-July 2006 episode. The CAMx final 2006 base case simulation achieves the Unpaired Accuracy of the Peak performance goal of ≤±20% for 58 of the 60 simulation days of the episode (i.e., 97% of the modeled days). There are 58 days with bias and error comparisons during the episode as two days had no observed ozone greater than 60 ppb so no statistics could be calculated. Of these 58 days, 50 of them (86%) achieved EPA's ≤±15% performance for Mean Normalized Bias and all of them achieves EPA's performance goal for Mean Normalized Gross Error.

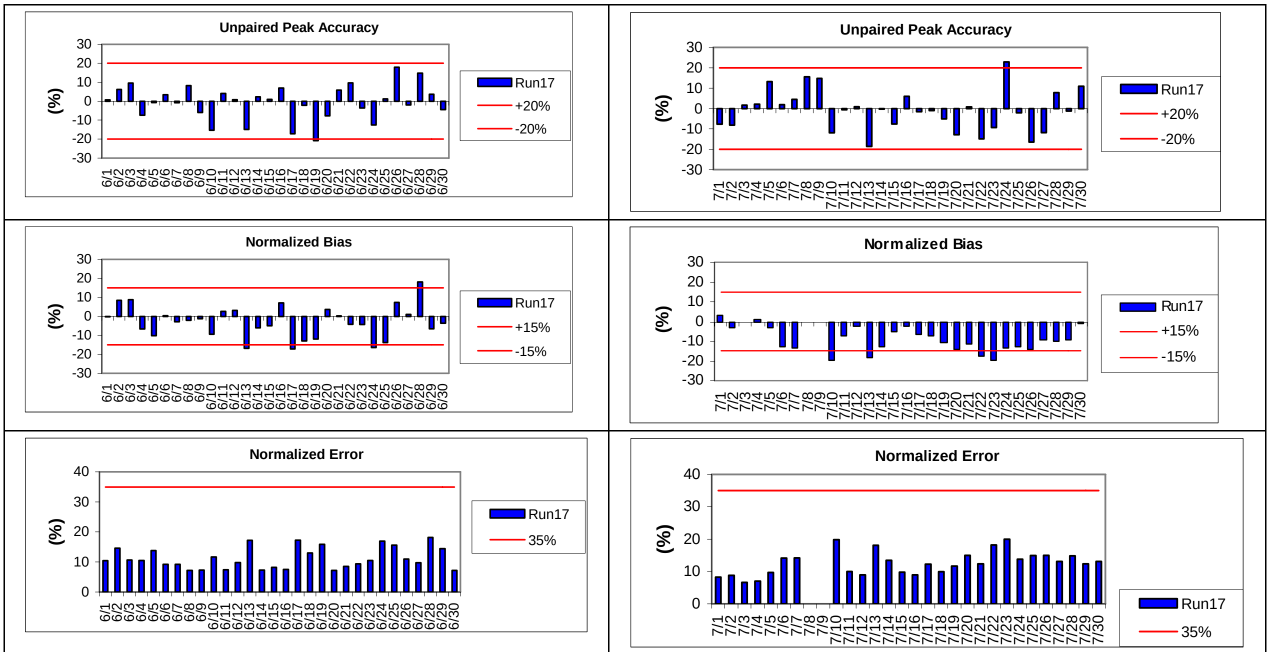


**Figure ES-1.** Hourly ozone performance statistics for June (left) and July (right) 2006 and Unpaired Peak Accuracy (top), Mean Normalized Bias (middle) and Mean Normalized Gross Error (bottom) and comparison against EPA's model performance goals (red lines).

EPA's draft 1999 8-hour ozone modeling guidance has a performance goal for daily maximum 8-hour ozone concentrations that the predicted value near the monitor be within ±20% on most monitor-days. This is a particularly important performance metric as it is these exactly same predicted daily maximum 8-hour ozone concentrations near the monitor that are used to make future year ozone projections. By "near the monitor" we used the same 7 x 7 array of 4 km grid cells centered on the monitor as used in the ozone projections and have made this analysis three ways by selecting the Maximum predicted value in the array, the predicted value Closest to the observed value, and the predicted value Co-Located at the monitor, with this last comparison being a particularly stringent test of model performance. Table ES-1 summarizes the results of this performance test, with a scatter plot of the predicted and observed daily maximum 8-hour ozone concentrations Co-Located at the monitor shown in Figure ES-2. Using the Maximum, Closest and Co-Located predicted daily maximum 8-hour ozone concentration near the monitor we see that the model achieves the within ±20% of the observed value performance goal 76%, 89% and 82% of the monitor-days. Thus, the Denver final 2006 base case CAMx simulation (Run 17) daily maximum 8-hour ozone predictions achieves EPA's ≤±20% of the observed value on most monitor-days performance goal.

**Table ES-1.** Percent of the monitor-days that the model predicted daily maximum 8-hour ozone concentrations near the monitor is within ±20% of the observed value (total monitor-days = 1008).

| Maximum Near the Monitor | | |
|---|---|---|
| Percent Difference | # Days | % Days |
| <-20 | 9 | 1% |
| -20% to +20 | 769 | 76% |
| > +20 | 230 | 23% |
| **Closest Near the Monitor** | | |
| Percent Difference | # days | % |
| <-20 | 23 | 2% |
| -20% to +20 | 902 | 89% |
| > +20 | 83 | 8% |
| **Co-Located At the Monitor** | | |
| Percent Difference | # days | % |
| <-20 | 48 | 5% |
| -20% to +20 | 829 | 82% |
| > +20 | 131 | 13% |

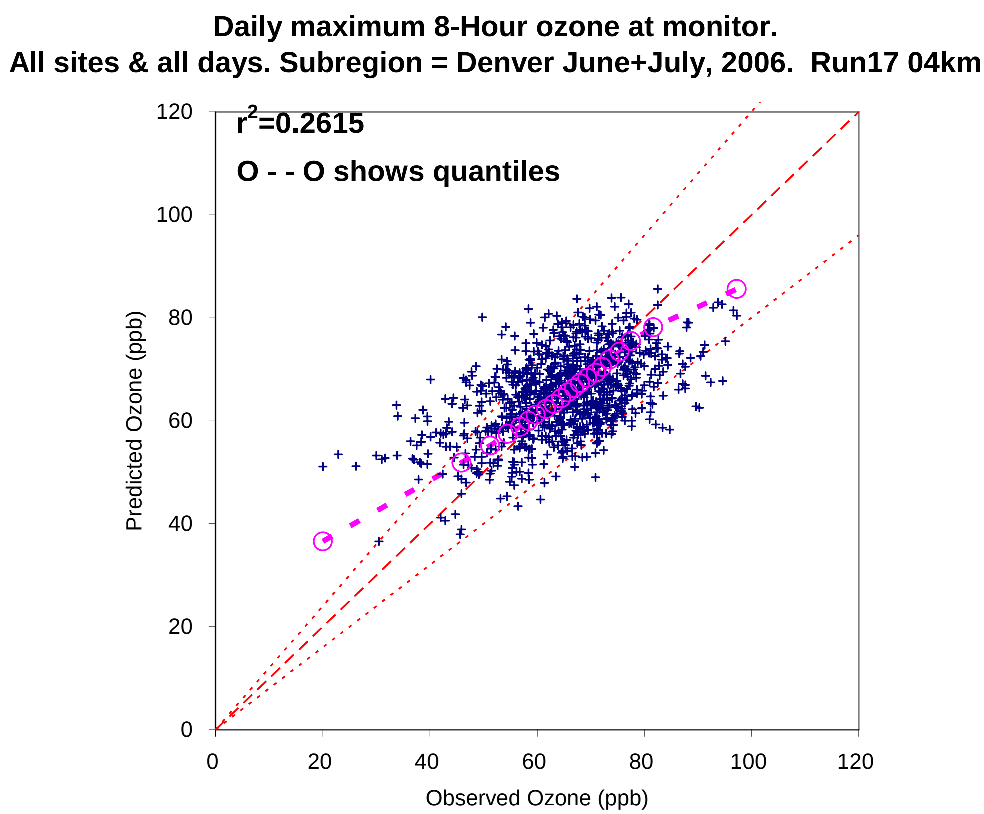


**Figure ES-2.** Comparison of predicted and observed daily maximum 8-hour ozone concentrations for the June-July 2006 modeling period and the final CAMx Run 17 2006 base case simulation using the Co-Located predicted value near the monitor.

Spatial and Temporal Evaluation of Ozone Model Performance

A detailed analysis of the ozone model performance was assessed by examining the ability of the model to predict the observed ozone concentrations for three 3-day episodes during the June-July 2006 modeling period when 7 of the 9 8-hour ozone exceedance days occurred. The detailed ozone model performance revealed that at some days the model failed to reproduce the elevated ozone concentrations and model performance was poor. Whereas, on other days the ozone model performance was quite good. In general, there were days with good model performance at the high southern (e.g., Chatfield) and northwestern (e.g., Rocky Flats North) ozone monitors, but there was a general underprediction bias at the northern Fort Collins West monitor. Figure ES-3 compares the spatial distribution of the predicted and observed daily maximum 8-hour ozone concentrations in the two highest days during the two month modeling period: July 14 and 29, 2006. On July 14th, the model is correctly predicting elevated ozone concentrations to the north-northwest of the DMA with the highest modeled values near the highest observed value at Fort Collins West monitor, although the modeled peaks are below the observed ozone peaks. On July 29, 2006 the model correctly predicts that the very highest ozone concentrations occur in and near the DMA with both the modeled and observed 8-hour ozone peaks in excess of 90 ppb. Time series of predicted and observed hourly ozone concentrations for the July 13-16 and July 27-29 periods and the two highest monitoring sites (RFNO and FTCW)are shown in Figure ES-4. In general the model underestimates the observed ozone peaks; the exception to this is at Rocky Flats North monitor for July 27-29, 2006 that exhibits very good ozone model performance.

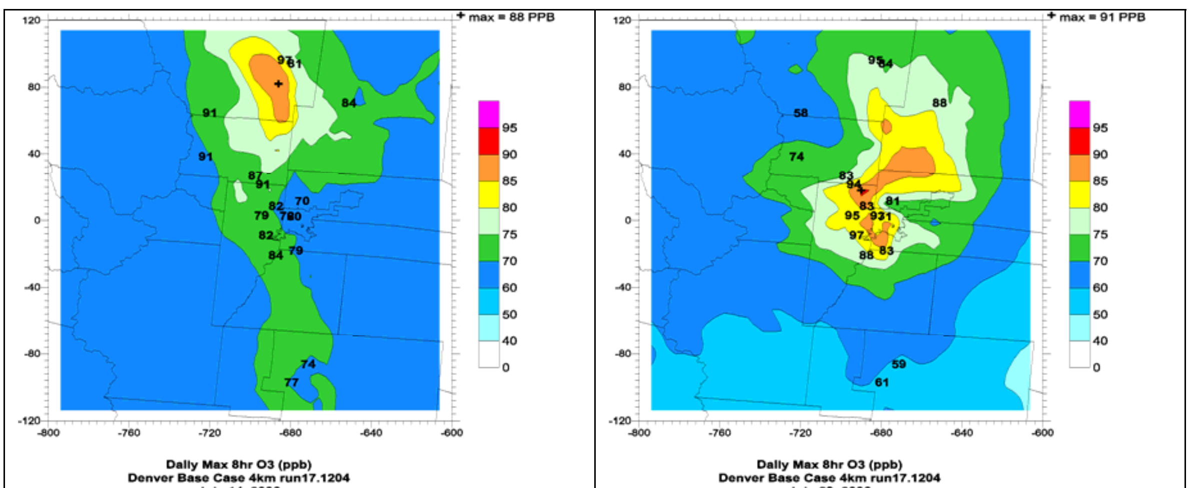


**Figure ES-3.** Comparison of the predicted and observed daily maximum 8-hour ozone concentrations on the two highest observed ozone days during the June-July 2006 modeling period using the final 2006 base case CAMx simulation (Run 17) results.

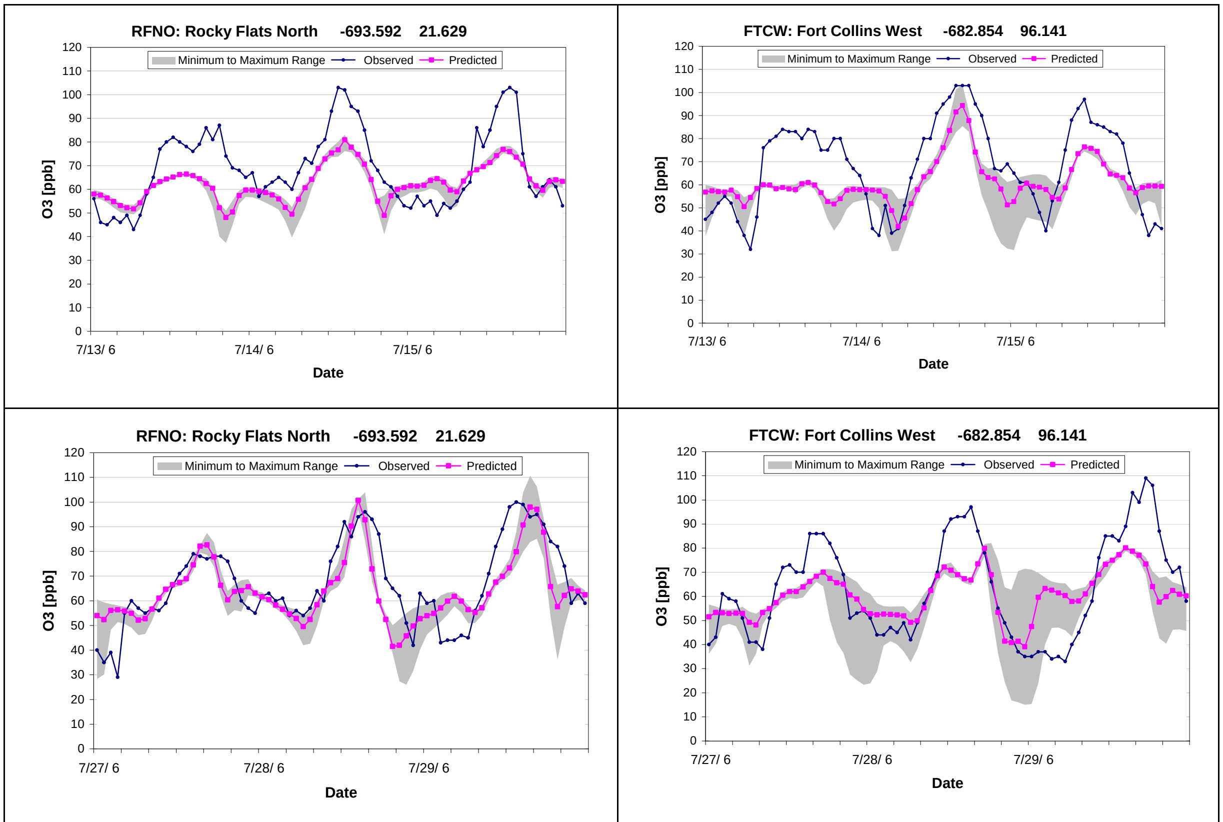


**Figure ES-4.** Predicted and observed hourly ozone concentrations at the Rocky Flats North (left) and Fort Collins West (eight) monitoring sites during July 13-16, 2006 (top) and July 27-29, 2006 (bottom) periods.

## Diagnostic Model Performance Evaluation

In addition to the many diagnostic tests performed to test the sensitivity of the model to model inputs and options performed as part of the preliminary model performance evaluation discussed in previous reports (Morris et al., 2008b; McNally, et al., 2008), the diagnostic model performance evaluation also included comparisons against ozone precursors, key indicator species, PM species and ozone aloft.

### <u>Ozone Precursor and Indicator Species Comparisons</u>

The CDPHE/APCD collected 3-hour VOC samples at several sites during a few days of the June-July 2006 modeling period. Figure ES-5 displays the example comparisons of ozone precursor and their ratios at the downtown Denver CAMP monitoring site. In these comparisons, the observed VOC species were converted to the CB05 lumped VOC species that is the chemical mechanism used in CAMx. Then the modeled and observed CB05 species were summed to get the total predicted and observed VOC concentrations, respectively. Note that ethane (ETHA) was not included in the total VOC species when summing the CB05 lumped species. Also note that since the VOC sampling did not collected measurements for methanol and ethanol, those species were also not accounted for when summing the CB05 species.

The model is systematically underpredicting the observed VOC concentrations at the CAMP monitoring site, whereas for NOx and CO there are days with underpredictions and days with overpredictions, although on average NOx is underpredicted as well. The CAMP monitor is located in downtown Denver where we would expect VOC, NOx and CO emissions to be dominated by on-road mobile sources. Thus, these comparisons provide a good evaluation of the on-road mobile source emissions. However, there are incommensurability differences between the observed point measurements at the 4 km grid cell average model predictions. Thus, a better indication of the accuracy of the on-road mobile source emissions are the VOC/NOx and CO/NOx key indicator ratio comparisons. Comparison of the predicted and observed VOC/NOx ratios also provide an indication of whether the model is reproducing the correct chemical regime in the DMA. Of the 15 days with morning VOC/NOx comparisons, very good comparisons are seen on 11 (73%) of the days. Of the four days in which the model is underpredicting the observed VOC/NOx ratio one is July $4^{th}$ that has atypical traffic patterns that were not simulated in the CONCEPT modeling. Two of the other VOC/NOx ratio overprediction days were weekend days, although on two other weekend days we saw good performance. The CO/NOx ratios appear to be slightly overpredicted on most days which may indicate that MOBILE6 is overestimating the on-road mobile source CO emissions, which has also been noted by Pollack and co-workers (2004). Ethane (ETHA) is underpredicted by approximately a factor of 10 at the CAMP monitor with observed values of ~25 ppbC and predicted values of ~2.5 ppbC. The underprediction of ethane is even greater at the two Weld County monitoring sites where the CB05 paraffin species is also underpredicted. As ethane is primarily associated with natural gas, its underprediction could indicate that organic emissions from natural gas related and oil and gas development sources are understated in the inventory.

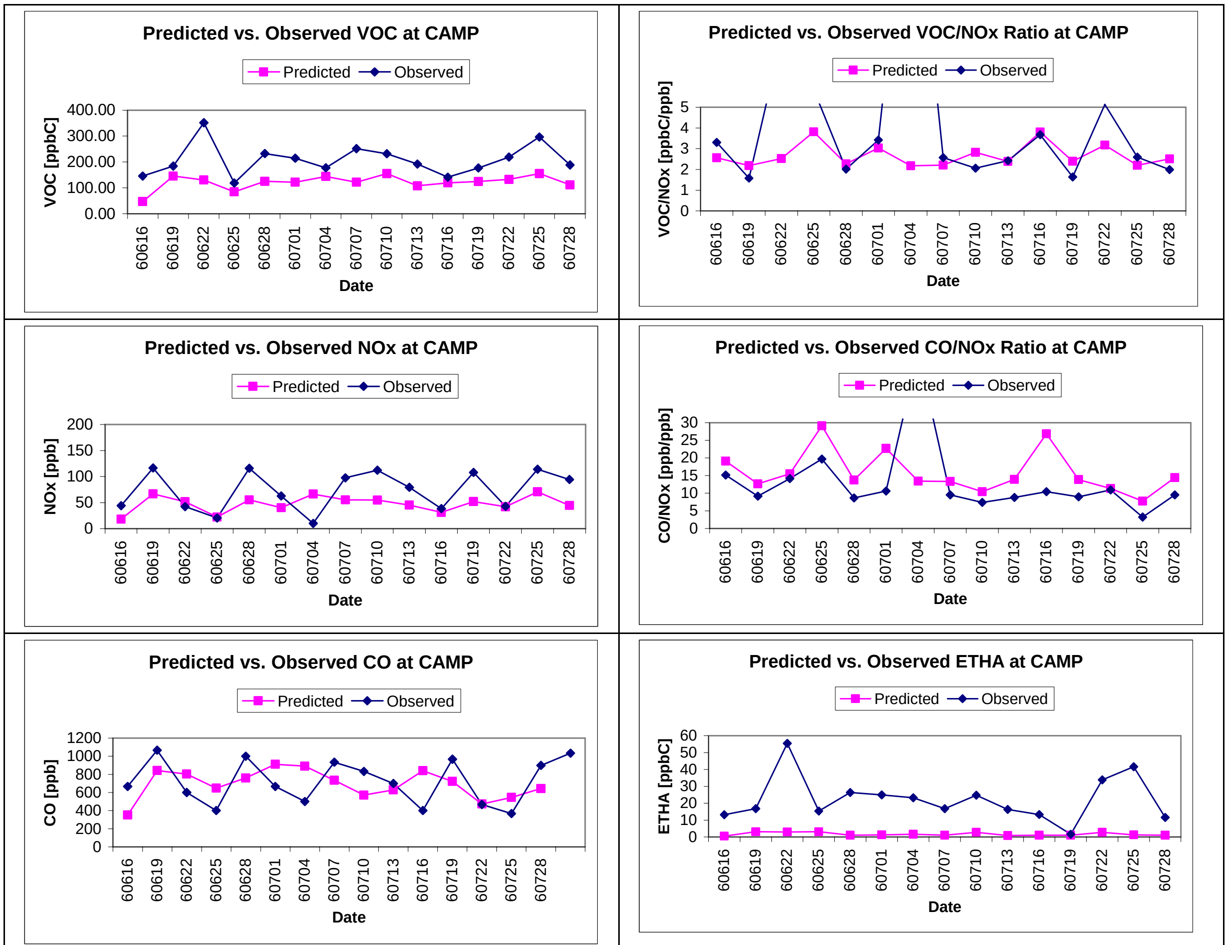


**Figure ES-5**. Comparison of predicted and observed early morning (5-8am MST) VOC, NOx and CO concentrations and VOC/NOx and CO/NOx ratios at the CAMP monitoring in downtown Denver.

In addition to the underestimation of observed ethane and paraffin VOC species noted above, carbonyl VOC species were also systematically underpredicted by the model. The underprediction of acetaldehyde may be due in part to the SMOKE VOC speciation profile for on-road mobile sources not accounting for ethanol blended gasoline whose combustion produces higher acetaldehyde than conventional gasoline. However, the reasons for the large underprediction of formaldehyde are less clear and since formaldehyde is an important VOC species that initiates the radical cycle its underprediction may help explain why the model tends to form ozone too slowly and does not obtain as high ozone peaks as observed. A review and evaluation of the VOC speciation profiles in the Denver modeling is recommended.

<u>Ozone Aloft</u>

During six days of the June-July 2006 modeling period, ozonesondes were released from Boulder that obtained a measured vertical ozone profile in the atmosphere. One hypothesis for the conceptual model on ozone formation in the DMA is that on some days there is a reservoir of ozone above the ground that is entrained and mixed to the ground as the mixing height rises. The comparison of the modeled vertical ozone profile with the ozonesonde measurements would provide an indication of whether the model is capturing this phenomenon. As the ozonesonde rises it will move downwind with the prevailing wind. In these comparisons we used the modeled vertical ozone profile in the grid cell at the time of the ozonesonde launch, so did not account for the horizontal and temporal displacement of the ozonesonde measurements from its launch point and start time. Figure ES-6 displays two of the comparisons of the predicted and observed vertical ozone profiles for June 15 and July 21, 2006. In general, the model does a better job reproducing the observed vertical ozone profiles for the three June days than the three July days. For example, on June 15 the model and observed ozone agree well in the lowest 2 km of the atmosphere and then deviate from each other. On the other hand, on July 21, 2006 the model is underpredicting the observed ozone in the lowest 2 km of the atmosphere by 20 ppb.

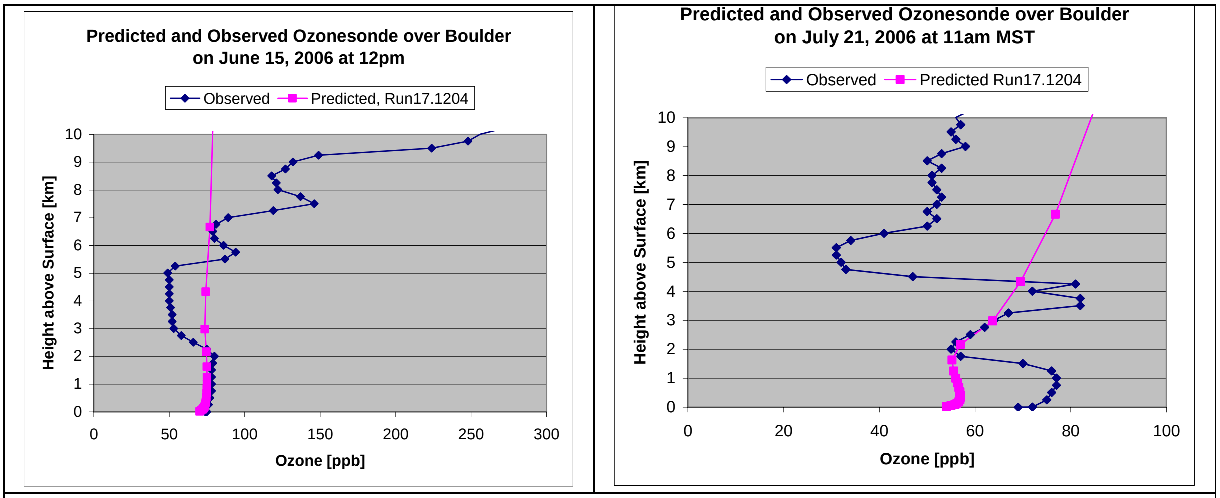


**Figure ES-6.** Comparison of predicted and observed vertical ozone profiles at Boulder for June 15 and July 21, 2006.

PM Performance

The Denver June-July 2006 photochemical modeling database was developed for demonstrating attainment of the 8-hour ozone standard. Consequently, the focus was on ozone and ozone precursor performance. Thus, no effort was spent on optimizing model performance for particulate matter (PM) species. In particular, the CDPHE/APCD only provided complete ozone precursor emissions (i.e., VOC, NOx and CO) for the state of Colorado. When modeling many sources categories, such as on-road mobile using SMOKE-MOBILE6 and CONCEPT MV models and processing the CEM data for point sources, we pick up the PM precursor emissions. However, for area and non-road mobile source emissions for sources in Colorado we only have ozone precursor emissions so are missing many PM related species (e.g., SO2, primary PM and ammonia). Thus, we expect to underpredict PM mass and PM species in the model performance evaluation.

Figure ES-7 displays an example evaluation for total PM2.5 mass from the more urban-oriented FRM and more rural-oriented IMPROVE monitoring networks. As expected, PM2.5 is underestimated most of the time. This underprediction is prevalent across all PM species with the lowest underprediction for SO4 and highest for NO3. These results are consistent with missing PM precursor emissions from area and non-road sources that would affect NO3 the most.

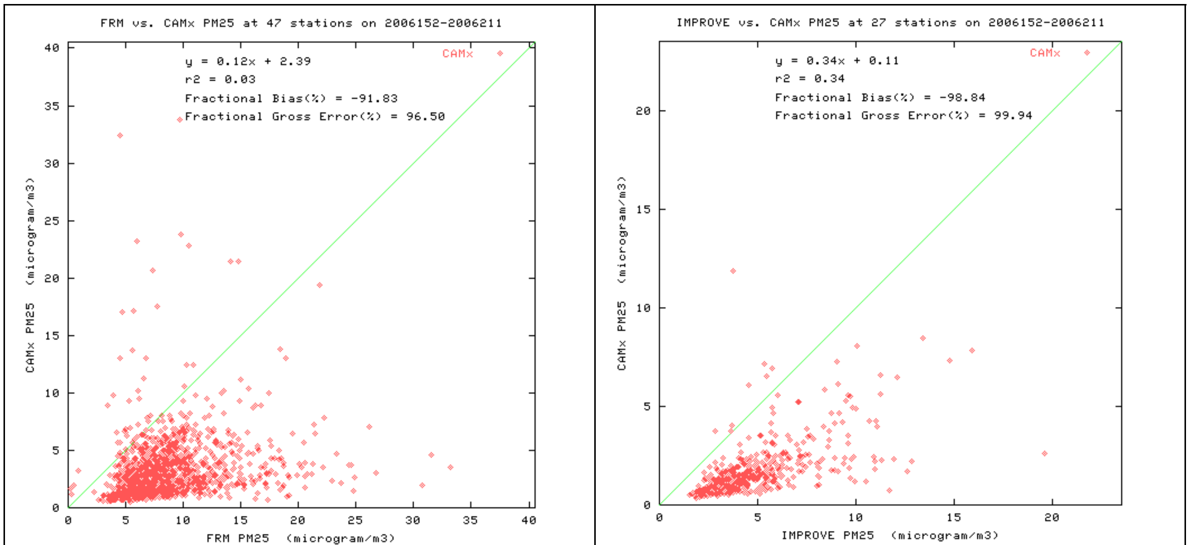


**Figure ES-7.** Comparison of predicted and observed total $PM_{2.5}$ mass model performance at FRM (left) and IMPROVE (right) sites in the 12 km modeling domain.

## CONCLUSIONS OF MODEL PERFORMANCE EVALUATION

As noted in EPA's latest air quality modeling guidance, "by definition, models are simplistic approximations of complex phenomena" that "…contain many elements that are uncertain". Consequently, achieving perfect model performance is unattainable and some uncertainties will always exist. The Denver final 2006 base case (Run 17) CAMx simulation achieves EPA's performance goals on a vast majority of days during the June-July 2006 modeling episode. The model is also exhibiting very good agreement for VOC/NOx ratios in Denver on most days, suggesting that the model is simulating the correct chemical regimes. The model performance is as good as or better than past ozone SIP modeling in the Denver and many other areas whose SIPS have been approved by EPA. Based on the model performance evaluation presented in this report, we conclude that the model is performing well enough to reliable project future-year ozone concentrations within the normal uncertainties of photochemical grid modeling. Although care should be taken that the ozone projections are not unduly affected by the few poor performing days.

The model performance evaluation has identified several areas of future research that could improve model performance including a focus on VOC speciation and the presence of aldehydes in the inventory, improvements in oil and gas emissions and other sources of natural gas related emission sources and better simulation of ozone aloft.