

H12.2004.8HRB: ROLE OF MODELING ASSUMPTIONS IN THE HOUSTON MID- COURSE REVIEW (UNC)

EXECUTIVE SUMMARY

BACKGROUND

The Texas Commission on Environmental Quality (TCEQ) was required by the US EPA to submit a Mid-Course Review (MCR) of the 2000 Houston-Galveston 8-county One-hour Ozone Non-attainment State Implementation Plan (SIP) by the end of 2004. For this task, TCEQ chose to develop completely new ozone modeling databases for a new 2000 episode. This modeling effort reflected new insights into the ozone problem in the Houston-Galveston area (HGA) that had been gained from the many investigations following the 2000 SIP and the analysis of data from the TexAQS-2000 extensive field-monitoring program. The new modeling completely replaced the previous 1993 ozone modeling that supported the 2000 SIP and resulted in significant modifications of the rules for NO_x and VOC emissions. Nevertheless, new TCEQ Commissioners spoke publicly about their concerns over uncertainties in the modeling and some of the rules. They wanted more assurance that, given the rules, HGA would show attainment in 2007. Put another way: Could the model be biased in the future simulations and thus mislead them about attainment status in 2007? These concerns generated the H-12.2004.8HRB project. The overall goals of the entire H-12.2004.8HRB project were to explore answers to two main questions:

Do the input and operational assumptions used in the MCR SIP modeling introduce bias into predicted 2007 ozone concentrations?

Do any potential biases in model predicted 2007 ozone concentrations impact the effectiveness of proposed control strategies?

Answers to these questions were not intended to generate replacement of the proposed SIP model nor its results, but rather to provide support for “weight of evidence” arguments in support of the modeling and the derived rules and to provide insight into the most uncertain areas that might then be narrowed as part of future efforts in other SIPs.

TASKS

The tasks, which occurred between July and December 2004, were divided among four research groups: University of North Carolina, University of Texas at Austin, University of Houston, and ENVIRON Corp. These project teams investigated the effects of several input assumptions on the base case and future case ozone predictions; these were: wild-fire emissions estimates, biogenic emissions estimates by different land-use and land cover (LULC) characterizations, choices of industrial point source (NEGU) VOC inventory for highly reactive VOC (HRVOCs) and other VOCs (OVOCs), effects of grid resolution on ozone production, and different chemical mechanism assumptions.

1.1.1 ENHANCED DAY-BY-DAY, SITE-BY-SITE, MODEL PERFORMANCE ANALYSIS

A major task in the UNC project was to greatly expand the model performance analysis beyond the usual three EPA ‘stat tests’ which are: unpaired peak O₃ accuracy, average O₃ bias, and average gross O₃ error. Because the same O₃ concentration can be formed from many different combinations of NO_x and VOC, and each of these same O₃ concentrations will respond differently to reductions in its precursors (see Chapter 1), it is important to establish that both the concentrations of the precursors and those of O₃ are similar to those in the ambient environment so that the model’s O₃ would be formed for reasons similar to those in the observed world. Then one has a reason to believe that the model’s response to precursor reductions would be similar to that in the observed world and that effective real-world control measures could be found with the model. For effective policies involving point sources, it is important to estimate the accuracy of simulating the source receptor relationships in the observed world. Investigating the accuracy of the predicted wind flow and vertical mixing fields in the model can accomplish a major part of this task. Thus, UNC set about assessing the quality of ‘history matching’ for the model using both physical (e.g., wind conditions) and chemical observations in a monitor-site-by-monitor-site, day-by-day approach based on graphical presentations (that meet certain criteria). UNC then supplemented these with a variety of summary statistics.

For the comprehensive model performance assessment task, UNC reviewed the current literature on the subject of evaluation of regulatory numerical models and the use of statistical measures (Chapter 1). The UNC team developed new graphical evaluation procedures (Chapter 2) and used them for UNC’s part of H12.2004.8HRB project (Chapter 3, Chapter 4, and Appendix I). Additionally, UNC calculated the EPA’s performance measures as well as alternative statistical measures (Chapter 5 and Appendix II).

1.1.2 BIOGENIC VOCs

Although UNC actively participated in reviewing and deciding research directions in the biogenic VOC emissions work and in the subsequent investigation of the effects of vertical mixing on surface layer concentrations, UNC’s main simulation work was on elements of, and alternative assumptions about, the NEGU VOC emission inventories (i.e., industrial point sources). This work had to await the outcome of the biogenic inventory investigations, because biogenic emission choices could impact significantly the HRVOC and OVOC choices. For this study topic, UNC also had to perform several sensitivity analyses for horizontal grid resolution, vertical diffusivity, and wind speeds because UNC found the effects of VOC inventory choices were likely sensitive to physical conditions based on the findings from the evaluation study for biogenic emissions choices and the previous H13 project results.

For the choice of biogenic VOC emissions, the issue was whether biogenic emissions estimated using the TFS-LULC dataset would be better than emissions estimated using the TCEQ-LULC dataset. The former was based on recent high-resolution satellite data that when used in CAMx produced less O₃ over western Houston, while the latter was based on older spatial—but much more tree-speciated—data that when used in CAMx produced higher O₃ over western Houston. For this task, UT performed most of simulations with input help from the UH team for the TFS-LULC data. The UNC team conducted evaluations against ambient data and the ENVIRON team conducted evaluations against the UH modeling that used EPA’s CMAQ model with a different vertical transport formulation. The conclusion on the LULC element of the biogenic inventory was that the details of tree specification were more important than spatial resolution of tree distri-

bution, and such a conclusion heavily favored the TCEQ-LULC data, which produced more O₃. The teams, however, were unable to demonstrate that any of the biogenic VOC inventories predictions could ‘match history’ at several monitored sites, including the research-quality site at the LaPorte Airport that had been part of the TexAQS study. In all cases, isoprene concentrations were over-predicted by factors of two to five at the surface sites, but were slightly under-predicted compared to concentrations measured by aircraft at 640 meters (Chapter 6, Section 6.1.1, Figure 6-6, and the UT H12 report). Comparisons of isoprene measurements at TCEQ sites over several years showed that the concentrations measured during the model episode were not unusual. All teams explored the cause of these large over-predictions at the surface and concluded that the model had a vertical mixing problem in the lowest layers of the CAMx model. Cell-to-cell exchange rates, or Kvs, control the mixing in the model. That is, the best explanation appears to be that the exchange rates were too low in the first eight or nine model layers and thus not enough of the surface emissions were being transported aloft. Clearly an inaccurate vertical mixing representation would affect all surface emitted species, including NO_x and thus impact other subsequent choices in the project. Because the H12 teams lacked the resources, time, and charge to ‘fix’ this problem, to make progress in this project, the H12 teams decided to use the biogenic emissions based on TCEQ-LULC as the best biogenic emission input for use in other tests. A consequence of this assumption was that the episode model would likely have a significant amount of surface VOC reactivity compared to the observed world. Consequently, UNC, in an effort to mitigate this effect, decided to include in its subsequent work both the original ‘1xKv’ CAMx values and a linearly scaled ‘2xKv’ and even a few ‘4xKv’ set of values for many of the other sensitivity tests (see the Tables in Chapter 3).

1.1.3 NEGU VOCs

The emissions inventories for the biogenic sources, mobile sources, area sources, and EGU sources were the same in all of UNC’s simulations and were those that TCEQ used in its MCR SIP modeling (they were supplied to the H12 teams by TCEQ emissions modelers). Five different base case and four different future case NEGU VOC emission inventories created by UNC were merged with the other sources “SIP” inventories to produce versions of the total inventory that differed in the mass and spatial distribution of the industrial point source emissions. These inventories are described in Chapter 3.

Based on extensive analysis of aircraft measurements and the new event emission reports required by TCEQ after 2003, the previous H13 project conducted by UT/UNC found that there were highly significant *episodic* event emissions in the real world. In fact, the results of the H13 project were used to set the “short-term limit” cap in the MCR SIP. The TCEQ modeling, with its “every-day-imputed” HRVOC emissions (called the psito2n2 inventory), however, did not produce high enough peak ozone at the highest observed days and locations, while at the same time, it significantly over-predicted O₃ concentrations on the low ozone days in the episode. The psito2n2 adjusted inventory also resulted in very large over-predictions of measured olefins at surface stations for all days (see Chapter 4 and 6, Figures 6-7, 6-14, and 6-15). The H13 work had also shown that simulations at a 1-km grid resolution were necessary to produce the high concentrations of O₃ observed (see Chapter 3, Section 3.1.2 and Figure 3.3) but the TCEQ SIP modeling had been done only at a 4-km grid resolution.

It was clear from observations that VOCs other than HRVOCs were often elevated in HGA (called OVOCs here). These are mostly alkanes, which react much slower than the

olefins in the HRVOCs. There were some questions raised by some Commissioners as to whether these too should be increased in the emission inventories and UNC conducted sensitivity analyses to demonstrate the effect on O₃ from such increases. UNC first investigated simulations using an inventory without the “imputed” adjustments (called the regular inventory) and found that these were in reasonable agreement with observations on the low O₃ days, but failed to produce O₃ above 125 ppb on the highest days observed. Increasing the fugitive fraction of the regular inventory, which was mostly alkane, by adding 60% or 180% more fugitive VOCs produced average increases of less than 0.5 ppb O₃ and did not help improve any of the performance problems. These simulations were included in the statistical analyses reported in Chapter 5. Given the other performance problems described below, UNC decided not to further investigate the effects of OVOCs.

UNC decided to test the hypothesis that the regular inventory in combination with “likely” daily HRVOC emissions events could produce comparable (or even better) model performance to that in the TCEQ’s psito2n2 modeling case with its essentially constant HRVOC emissions adjustment. If this were the case, we wished to contrast the relative magnitudes of the HRVOC emissions in the two simulations. This would provide some bounds on the “long-term cap” for HRVOCs that was in the MCR SIP and which was an increase over the emissions in the regular inventory. Thus, for each day, the UNC team explored possible HRVOC emission events at different places and put the best event simulation for each modeling day together with the regular inventory as the UNC’s best base case inventory and this was called the regevt1 inventory. Then, the UNC team compared the performance of the regevt1 base case with that of TCEQ’s psito2n2 base case. The same was done for the future case of each inventory. These results are described in Chapters 4 and 5 and are discussed in Chapter 6.

PRODUCTS

Altogether UNC preformed 26 multi-day long simulations investigating nine VOC inventories, two grid resolutions, and four meteorological variations (see Chapter 3, Section 3.2). In addition, to explore the role of event emissions verses TCEQ imputed emissions, UNC performed six dispersion-only simulations on four days to investigate source receptor relationships, followed by 30 day-long simulations with full chemistry involving combinations of 10 industrial sites (see Chapter 3, Section 3.3).

There were an enormous number of plots produced: more than 175,000 occupying nearly 1.2 gigabytes of storage. The vast majority of these were visualizations of spatial fields of primary and secondary species concentrations (O₃, NO, NO₂, NO_x, ETH, OLE, ALD2, FORM) each hour of each day. Others included various species observed and predicted time series and scatter plots (with special versions for wind speed and direction), and peak concentration bar charts. All of these plots, with a viewer program, are available on a DVD disk. The source codes for all the processing and graphics production are available on a CD-ROM. These were written in Python, a free, Open Source, multi-platform language that is widely supported for use in scientific applications.

Descriptions of the most critical graphs and detail discussions of the days with high ozone exceedances are presented in Chapter 4. In addition to the highlights shown in Chapters 4 and 6, Appendix I MPP provides for each day from August 22 to September 6, 2000 a 66-page side-by-side and site-by-site comparison of the most significant graphical outputs for psito2n2 case and the regevt1 case. Another document in Appendix I MPP provides comparisons of three base case simulations with three future case simula-

tions. Appendix II STAT provides a 175-page description of the statistical analysis results in the form of plots and tables.

FINDINGS

1.1.4 FLAWED HISTORY MATCHING FOR WIND FIELDS-PART 1

The EPA “Meteorological Monitoring Guidance for Regulatory Modeling Applications” (EPA 454/R-99-005, Feb. 2000) says that the system performance expected for monitoring site observations is a wind speed and direction starting speed of 1.8 kilometers per hour (km/h) [0.5 m/s] with a wind speed accuracy of $\pm(0.72 \text{ km/h} + 5\% \text{ of observed})$, and a direction accuracy of ± 5 degrees. On our wind speed visualization plots, we have included a 2:1 (representing a predicted wind speed of twice the observed) and 0.5:1 (representing a predicted wind speed of half the observed) dotted line. Any hourly points falling outside these two dotted lines we describe as representing serious flaws in history matching. We consider good agreements to be within $\pm 50\%$. On UNC’s wind direction visualization plots, we mark each 30 degrees (1/3 of a quadrant) and consider any hourly points falling outside of ± 60 degrees of the monitored resultant hourly direction to represent serious flaws in history matching. We consider good agreements to be within ± 30 degrees.

The new wind speed and direction graphical visualizations quickly permitted the UNC team to determine that there were significant flaws in the wind speed and direction fields provided to the model for the periods August 22– August 24, 2000 and September 1– September 6, 2000. These CAMx meteorology inputs were derived from the ATMET MM5 meteorological simulations that were done for the entire period August 18 to September 6, 2000. TCEQ replaced the middle period, August 25—August 31, 2000, of the ATMET simulations with MM5 simulations from Texas A&M University (TAMU) that were based on GOES-satellite-corrected soil moisture that permitted the model to predict morning temperature rise rates better. These adjustments also produced significantly better wind fields for the period August 22-August 31, 2000. To have an extended episode, however, the TCEQ blended the ATMET MM5 and TAMU MM5 meteorology files and called the result a “hybrid” meteorology that covered the period August 22 to September 6, 2000. Unfortunately, the UNC analysis presented in Chapter 4 and discussed in Chapter 5, shows that, at the majority of the wind monitoring sites, the ATMET MM5 predictions are too fast by factors of 2 to 5 and the morning and afternoon resultant wind directions errors were large and frequently between 30 and 90 degrees in error.

While these ATMET days were included in the day-by-day statistical analysis results presented in Chapter 5, UNC judged that the poor quality of the simulated physical conditions so distorted the source receptor relationships for these days that it would be fruitless to perform any day-by-day, site-by-site analyses for them. Further, it is highly likely that any policies evaluated with these results would be unreliable. The graphical results for these days are included in Appendix I MPP on a day-by-day basis; UNC did not attempt any interpretation of these. Thus, the UNC team concentrated on extensive graphical performance analyses only for the seven-day period of August 25 to August 31 covered by the TAMU MM5 GOES-corrected meteorology for the various base case and future case conditions.

1.1.5 FLAWED HISTORY MATCHING FOR WIND FIELDS-PART 2

Within the August 25 to August 31 time period (i.e., the TAMU MM5 GOES-correct meteorology), UNC’s graphical analysis revealed that there were also periods and

locations with serious mismatches among the predicted and observed wind fields at the monitoring sites, however, these mismatches were quite different from those in the AT-MET MM5 meteorology case. Unlike the ATMET speed predictions, which were generally much *higher* than the observations, in the case of the TAMU MM5 GOES-corrected meteorology, during the daytime hours, the speed predictions were generally *less than* the observations even when the observations were well above the recommended starting speed (in some important cases, even when the observations were 4-8 km/h). In fact, over large areas, the model incorrectly became calm for as long as five hours (see Figure 6-4 and 6-5), while the monitored wind speeds were well above the starting speed thresholds of the sensors. During the nighttime hours, this condition was reversed, i.e., the predicted wind speeds were often a factor of two higher than the observations. For 8/25, 8/26, and 8/29, there were large areas of virtually no flow predicted in the morning while the observations were 4 to 8 km/h; on 8/27 and 8/28, the predictions and observations were mostly above 4 km/h and in the afternoon above 16 km/h and the model and observations were in reasonably agreement over most of the domain during the daylight hours for these higher flows; on 8/30 and 8/31, there were again large areas incorrectly predicted as calm or very low flow, while the observations were 2 to 6 km/h. For these days, however, the problems were in the afternoon rather than the morning.

The nighttime prediction problems are very likely associated with the model's very low predicted nighttime mixing depth, almost always confined to layer 1, which is about 34 meters in depth. The nighttime depth that was used in the models was determined by examining the supplied Kv profile values. Wind observations were at 10 meters. Even if we used a 'power law' profile to correct the mid-level (17 m) predictions to 10 meters (which may or may not be proper and can not be done with just information included in the model's meteorology input files), the corrections would only reduce the wind speeds by about 20%. The very large over-predictions of the nighttime observed species concentrations strongly imply that a more appropriate urban nighttime mixing height would be closer to 200 meters, consistent with an urban heat island. Changes to the planetary boundary layer (pbl) scheme at the lowest levels in the MM5 meteorology model to introduce a heat island effect over Houston would perhaps reduce significantly the predicted lowest level nighttime wind speed due to an enhanced drag imparted by the surface.

1.1.6 FLAWED PRECURSOR CONCENTRATIONS HISTORY MATCHING

One would expect that large differences between wind field predictions and wind field observations would be reflected in differences in the predictions and observations of primary and perhaps secondary species concentrations. This was true for NO, NO₂, individual HRVOCs, biogenic VOCs and products like formaldehyde (FORM) and ozone. Many of these history-matching problems were common to both the psito2n2 and regevent1 inventory simulations, which implies that their causes are associated with either the predicted physical conditions (i.e., wind field or vertical mixing) that determine the volume into which emissions occur or they are associated with the emissions mass rates in the components of the emissions inventories that were the same in these two cases (i.e., the mobile source, area source, biogenic source, or EGU source inventories).

Especially troubling were the erroneously predicted high NO_x concentrations that occurred throughout the simulations. For example, on 8/25 at the Haden Road (H03H) site located between the Ship Channel and I10, the observed NO was about 10 ppb while predicted NO was over 100 ppb, see Figure 6.7; on 8/30, at the Deer Park 2 (C35) site, ob-

served NO₂ was less than 15 ppb and falling while the predicted NO₂ peaked at 40 ppb and then declined to 25 ppb, but then rose again in the afternoon, causing depressed ozone to be predicted (see Figure 6.8). This over predicted NO_x problem clearly extends into the future cases. At 10 of 14 sites measuring NO_x on 8/25, 11 of 14 sites on 8/29, 8 of 14 sites on 8/30, and 6 of 14 sites on 8/31 the ***future case NO₂ predicted concentrations are higher than the base case observed NO₂ values, even after NO_x control strategy reductions*** (see Appendix I MPP). The sites that did show future case NO₂ reductions below the base case observations were those near the ship channel.

Already mentioned above were the factor of 2 to 5 in over-predictions in the very reactive biogenic VOC, isoprene (ISOP). Also mentioned before was that ETH, OLE and ALD2 in psito2n2 cases are also over predicted during the daylight hours and are severely over predicted—more than a factor of 10—at night (see Figures 6-7, 6-14, and 6-15). These VOCs are more accurately predicted in the regevent1 case, but are still somewhat over-predicted at night. Both cases miss some important event emissions after dark.

One would normally think that conditions of both high NO_x and high VOCs might lead to over prediction of ozone, but this was not the case on 8/25, 8/29, 8/30, and 8/31. Instead we found that there were systematic under predictions of FORM (formaldehyde, a reaction product of OH radicals with ETH), which indicates that, in the model, the sites predicted to have large concentrations of NO₂ are likely very NO_x inhibited. That is, new OH radicals were reacting to terminate with NO₂ making HNO₃ rather than reacting with ETH to propagate and oxidize NO to NO₂ and recycle OH radicals as needed for significant ozone formation.

We could be more definitive about the cause of the low ozone production in the presence of so much NO_x and high VOCs if we had performed process analysis runs over many of the monitor site areas; the scope, effort, and short time duration prevented us from doing these runs and analyses in this project.

1.1.7 OZONE PASSES STATISTICAL TESTS

Even with the wind field and precursor history matching problems, the psito2n2 inventory modeling case could pass EPA's statistical performance evaluation for ozone. Likewise, the regevent1 inventory case statistical tests results were comparable to TCEQ's modeling. That is, the results of standard EPA statistical evaluation showed that there were no significant differences in ozone model performance between the psito2n2 case and the regevent1 case. Moreover, the regevent1 case showed better performance on August 29, 2000 by eliminating the false positive at H04H and, for August 30 – August 31, 2000, by predicting peak ozone higher and closer to the observations.

But, of course, there was a very big difference in the precursor conditions between these two cases: in the TCEQ psito2n2 case, a total of 1,092 tons of HRVOC were added for the August 25-31 time period, while in the UNC regevent1 case, only 67 tons of HRVOC were added for the same period. That is, TCEQ added, proportional to the regular emission inventory, 156 tons of HRVOC per day throughout the entire modeling period, while UNC added 8.7 tons on the lowest day to 29 tons of HRVOC on the highest day, but only added HRVOC for 13 hours during the four days. Conventional wisdom might suggest that there should be large differences in ozone outcomes in these two scenarios, but as was shown in Chapters 4 and 5, this was not the case: essentially the same O₃ performances were obtained.

Similar or better performance with regevent1, however, did not ensure UNC's approach was the best achievable because the UNC team found that there were significant discrepancies in the wind fields at the most critical times at the exceedance monitors and these significantly undermined the apparent good performance of regevent1. That is, based on the wind field in the model, one can infer "where" to insert an emissions event, so as to "reproduce" the observed transient high ozone event (THOE) and greatly improve the model's statistical performance. But detailed hourly graphical analysis of the predicted and observed wind field at the monitor reporting the THOE strongly suggested that the real-world event source was in an entirely different direction from the one used in the model to obtain good performance. Thus, the 'right answer' was obtained, but for the 'wrong reasons'.

UNC also performed alternative statistical evaluations with new statistics in use in other modeling fields and with skill scores often used in numerical weather predictions. These were performed for both O₃ and NO_x concentrations, day-by-day and for spatial sub-sets of monitors and for all the various emissions inventories we ran. The results showed that, on the days with exceedances, the simulation with the best agreement for ozone was the one with HRVOCs everywhere and with the most diluted condition, i.e., the psito2n2 case at 4-km grid with 2xKv. These are conditions needed to lessen the incorrectly predicted NO_x inhibition conditions and to provide a source of HRVOCs regardless of the wind direction. Clearly being 'best' statistically does not mean being 'right'.

These alternative statistical analyses also showed that TCEQ's psito2n2 case results were better than UNC's regevent1 case on August 25, 2000, but that the two were comparable for the rest of the analysis period. This was in part due to UNC's dependency on the using the model's dispersion field to locate likely event emission sources; the very slow wind field on August 25 would not transport the sources we thought were the most likely far enough to the west. If we had more time to correct our choices, we could improve the performance of regevent1 case, but as discussed above, this might be more of getting the 'right answer', but for the 'wrong reasons'.

CONCLUSIONS

1.1.8 BASE CASE IS BIASED HIGH FOR PRECURSORS, LOW FOR OZONE

Direct comparisons of the model predictions with concentration observations for the NO_x species and the biogenic isoprene (ISOP) shows that, for the Harris County monitors, the model predictions are on the average biased high, sometimes by large factors. The normalized biases for NO and NO₂ on the exceedance days were more than +100% on three of the days and were around +80% on the fourth day. At the two observation sites where the HRVOC species ETH and OLE were measured, model predictions for the psito2n2 inventory case were over-predicted by factors of 2 to as large as 10; the model predictions for the regular and regevent1 inventory case, however, were mostly within a factor less than 2. For ISOP, which was measured at the same sites, all the models using the TCEQ-LULC dataset over-predicted observed concentrations by factors of 2-5. Secondary product species, O₃ and formaldehyde, are frequently under-predicted by the model. The normalized bias for O₃ on exceedance days is -20% or more, with the psito2n2 inventory case having the lower values. On the non-exceedance days, however, the psito2n2 case over-predicts the ozone.

These results are consistent with a base case that is strongly NO_x inhibited well beyond the observed condition. Such a condition results in radical termination greatly exceeding radical propagation and consequently limiting formation of O₃. On an ozone-

to-precursor response diagram (i.e., O₃ isopleth diagram, Figure 1-1), this corresponds to moving the real-world operating point from near the O₃ ridgeline to a modeling point that is well off the ridgeline and in the upper left quadrant of the diagram, which has lower ozone. While the NO_x is over-predicted in all tested inventories, the use of the regular emissions inventory resulted in VOC predictions that were more consistent with the observations. If the NO_x were predicted correctly, the VOCs from the regular emissions inventory could have produced much more ozone. In the psito2n2 case, the every-day-adding of 156 tons of HRVOC to a condition that was predicting excess NO_x resulted in moving the model's operating point to the right and toward the ridgeline and higher ozone, consistent with having to pass the EPA peak ozone accuracy tests.

Thus, the base case MCR SIP model appears to have an incorrect precursor to ozone operating point that is too high in both NO_x and VOCs, but is not reactive enough in producing ozone because of the non-linearity of the system. Adding large 'event' HRVOCs to parcels in these conditions moves the parcel's operating point much further to the right (well beyond the ozone maximum ridgeline).

1.1.9 FUTURE CASE IS ALSO BIASED HIGH FOR PRECURSORS, UNKNOWN FOR OZONE

In the future case emissions inventories both NO_x and VOCs were reduced from the base case values in psito2n2 case and the regevent1 case. The NO_x reductions were the same in the two cases. In the psito2n2 case, the future VOC reductions were not sufficient to remove all the "added VOC due to imputation" meaning that there was some growth of HRVOCs over the regular inventory, whereas in the regevent1 case all the added event VOCs were removed. The rest of the control strategy was essentially the same for psito2n2 and regevent1 cases. TCEQ emissions modelers supplied all cases.

The NO_x reductions used in the future cases, however, were not sufficient to result in future predicted NO_x concentrations that were lower than the base case measured NO_x at many locations on each day, meaning that the future case model precursor-to-ozone operating points were not predicted close to what the future, real-world NO_x conditions will be when the control strategy NO_x reductions are fully implemented. This is not because the reductions were not large enough; it was because the starting predicted concentrations were many times larger than the observations. Thus, in the future condition, the model is still operating in a non-linear portion of the response surface that is not close to where the future real world will be given implemented reductions and therefore the model is unlikely to provide accurate simulations of likely outcomes to various emissions reduction strategies.

The NO_x reduction relieves the model of some of its NO_x inhibition and permits it to make more ozone with the same VOC, but it clearly is still deep into inhibition. This is why, in the H13 work, a 90% NO_x reduction showed more O₃ production in the future case than did a 80% NO_x reduction. Likewise, operating the future case model with 4-km cells, or with 2xKv inputs or with both (i.e., conditions that result in more instant dilution of the NO_x emissions, giving lower NO_x concentrations) all gave higher ozone predictions than the standard inputs 1-km configuration.

1.1.10 FUTURE CASE BIASES IMPACTS EFFECTIVENESS OF CONTROLS

The demonstrated attainment with the MCR SIP modeling based on the psito2n2 inventory or on the alternative regevent1 inventory cannot be accepted at face value: the model has significantly too much NO_x in future cases to be reliable for estimating the ef-

fectiveness of a “long-term cap” on HRVOC as a means of limiting O₃ production in the future.

When the base case NO_x predictions are made more accurate by some change in the modeling system and these conditions are then brought forward to the future case with the required NO_x reductions, we may find that the future case is a) still NO_x inhibited, b) near the optimum condition to produce ozone given the available VOCs, that is, near the ridgeline or c) below the ridgeline (i.e., somewhat NO_x-limited). In each of these conditions, the future case could be in demonstrated attainment or not. If the future is not in demonstrated attainment, the condition indicates what types of additional reductions that are likely to be needed to reach attainment in the model. Given the present modeling results, however, we do not know how the model’s future predictions relate to the real world and so we cannot draw meaning conclusions from the modeling results about the effectiveness or the likely magnitude of a “long-term cap” on HRVOC. The safest current approach might be to set the “long-term cap” at the present level in the reported or regular inventory. Additional “weight of evidence” arguments are needed.

On the other hand, the “short-term limit rule” on HRVOCs that was estimated with this model is likely somewhat conservative. This is because the large amount of VOC added in parcels drives the parcels’ conditions to be well to the right of the optimum condition to produce O₃ given the available NO_x, i.e., far right of the ridgeline in the response diagram, which means they are very NO_x-limited. Since the modeling system is too NO_x - rich in the MCR SIP future case, any corrections would lower the NO_x making the ‘event’ parcels even more NO_x-limited and likely to produce less ozone than in the current system. In the real world, therefore events limited to the size estimated to be acceptable with the MCR SIP model are actually likely to produce less ozone.

1.1.11 FUTURE EFFORTS NEEDED TO REDUCE UNCERTAINTY

Clearly the cause of the large NO_x over-predictions in the base case CAMx model must be found and corrected. It is likely that these over-predictions are caused in part by:

- Simulated wind fields for critical time periods and over large spatial areas that are predicted to be calm, some sites for up to five hours, in contrast to the observations that suggest that the real world flow is several cells per hour
- Simulated vertical mixing rates that are too slow to mix surface emissions to model layer 8 or 9 (640 m) were the aircraft were flying
- Simulated nighttime mixing depths that are too shallow resulting in incorrect large build up of predicted concentrations from surface emissions into the shallow layers; this also likely results in large differences in predicted and observed nighttime surface layer wind speeds.

It is clear that choices of the MM5 simulation parameters and the evaluation of the resulting meteorological fields cannot be done independent of their effects on the air quality predictions; this must be an iterative process. Further, the MM5 simulation effort cannot be declared complete until the predicted fields have been used in an evaluation of the air quality model even if the emissions inventory is in a preliminary stage. For example, there has been a lot of effort made to assure that the MM5 model simulates the temporal profile of the pbl depth, but little effort at evaluating the mixing rates from the surface to 1 km. Even large errors in the maximum depth may only result in volume changes of 10-20%, while differences in mixing rates in common use in models today can result in factors of 2-5 in surface concentrations.

There is also a possibility that the NO_x emissions inventory rates are too high for one or more of the NO_x source sectors. This cannot be supported until we can use meteorological inputs that result in better history matching for the wind field and nighttime concentrations. Given the spatial distribution of the over-prediction problem the most likely candidate would be the mobile source inventory. It is clear already that there is a CO versus NO_x prediction problem in the mobile inventory for HGA in that the model's CO/NO_x ratio is greatly over-predicted relative to the observed CO/NO_x ratio. The TCEQ's monitoring network CO is nearly useless for evaluating CO inventories due to low accuracy and drift problems in the monitors; perhaps it is time for TCEQ to invest in CO monitors similar to those used in the TexAQS study.

Finally, it is clear that we cannot continue to judge ozone models by evaluating ozone concentrations alone. If more comprehensive efforts involving both the meteorological observations and the precursor species observations had been in use for the last two years, the current MCR SIP modeling results would be more compelling as an attainment demonstration.