Appendix K

Report - Evaluation of the CAMx Base Case Model Performance for the August 31 - September 2, 1993 COAST Episode

May 27,1999

EVALUATION OF CAM$_X$ BASE CASE MODEL PERFORMANCE FOR THE AUGUST 31 - SEPTEMBER 2, 1993 COAST EPISODE

Prepared for:

the Texas Natural Resource Conservation Commission
Mr. James Red, Project Manager
Under Work Order No. 9800693000-04

Prepared by:

Neil Wheeler
Saravanan Arunachalam
Pat Dolwick
Jeff Vukovich

Environmental Programs
MCNC–North Carolina Supercomputing Center
P.O. Box 12889
Research Triangle Park, NC 27709-2889

Table of Contents

1. INTRODUCTION

Before an air quality model may used to guide policy decisions on future-year emissions control strategies, the ability of the model to accurately characterize ozone formation must be evaluated. Typically, this is done by comparing the simulated pollutant fields against the ambient observations from a historical ozone episode. This report summarizes the evaluation results from the Comprehensive Air Quality Model with Extensions (CAM_X) over the Beaumont-Port Arthur (B-PA) region during the August 31st – September 2nd 1993 episode.

According to prevailing thought in the air quality modeling community, it is possible to separate a meaningful model performance evaluation into two components (Tesche, 1990):

1. the *operational evaluation* which consists primarily of comparisons between the model and observed data, both qualitative and statistical, and

2. the *scientific evaluation* which attempts to determine if the physical processes leading to ambient pollutant concentrations are represented properly in the model.

This assessment of CAM_X model performance over the August 31st – September 2nd episode consists mostly of an operational evaluation primarily because of the absence of ambient data aloft and schedule constraints that work against a rigorous model evaluation.

Several other measures, however, were taken to ensure the highest possible quality model output fields. At each step prior to conducting base case simulations, the input fields were quality assured for consistency and obvious errors. Additionally, several diagnostic tests were completed to check the model formulation and response to various inputs. The goal of diagnostic testing is to determine whether a model exhibits expected behavior under extreme changes to its input. Diagnostic tests check for spurious behavior that might indicate problems in the model formulation. The results of the five diagnostic simulations are summarized in Section 3 of this report.

As outlined in the modeling protocol (TNRCC, 1998a), several methods were used to assess the model precision in reproducing observed ozone over the August 31st – September 2nd 1993 episode. Statistical methods can provide a quantitative measure of model performance. The results of these methods must be considered carefully, especially in cases where there are not a large number of monitors. Graphical displays comparing predicted to observed concentrations can provide qualitative information on model performance. The following techniques will be used in this analysis:

- Unpaired Peak Accuracy: This measure calculates the difference between the highest observed value and the highest predicted value independent of time and space. (USEPA recommended range: $\pm$15-20%)

- Normalized Bias: This test measures the ability of the model in reproducing observed patterns. Since there are many time periods when relatively low levels of ozone are predicted, and statistics from these periods are not very meaningful, this test will be limited to pairs where the observed concentration is greater than 60 ppb. (USEPA recommended range: $\pm$5-15%)
- Gross Error: This test will compare the difference between all pairs of predictions and observations that are greater than 60 ppb. This is a measure of model precision. (USEPA recommended range: 30-35%)
- Time-Series Plots: For each monitoring station in the domain and for each hour in the episode, the predicted concentration will be compared with the monitored concentration. This will determine if the model can predict the peak concentrations and if the timing of ozone generation in the model agrees with that found with the monitoring. Because modeled volumetric concentrations are compared with data from monitoring sites, that are specific points in space, perfect agreement should not be expected.
- Surface-Level Isopleths: For selected daytime hours, surface-level isopleths (lines of equal concentration) will be drawn. This shows how the model is predicting the extent, location, and magnitude of ozone formation. This information can be compared to monitoring results.
- Scatter plots: Scatter plots of predictions compared to observations depict the extent of bias in the ensemble of hourly data pairs. Systematic positioning of data points around the perfect correlation line indicates bias. The distribution of points over the area is an indication of error.

2. STATISTICAL ASSESSMENT OF MODEL PERFORMANCE

When viewing these results, remember that CAMx predicts a volumetric one-hour average over an entire grid cell. Monitoring data provides a measure of air quality at a specific point in space. To provide an accurate comparison with model predictions, the monitoring data would have to be transformed into volumetric one-hour averages over the same grid cells used in CAMx. However, monitoring networks are not dense enough to provide this information even for the most intensive studies that have been performed. Thus, comparison between the CAMx volumetric predictions and the monitored point measurements are the only recourse. This can provide insight into the model prediction accuracy, but does not provide precise measures of model performance. Also, since the two days of each modeling episode are initialization days during which ozone exceedances were not recorded, performance measures will be applied only to modeling results subsequent to the ramp-up days.

Table 1 displays some of the first-order evaluation statistics generated from the B-PA 1993 base case as calculated over the entire domain. The statistics represent only observed-model pairs where the observation is greater than 60 ppb. As can be seen from the table and

Figure 1a, model performance for this 1993 COAST episode meets the criteria outlined by USEPA for an acceptable base case. The unpaired peak predictions exhibit a negative (underprediction) bias on all both episode days and range from to 1 to 5 percent. While the maximum ozone is generally underpredicted, the mean ozone greater than 60 ppb is overestimated in the model on the 1st. Normalized bias values range from near zero on the 2nd to 12% on the other episode day. Gross error ranges from 18 to 29 percent on the episode days. Statistical model performance on 2 September is the best of any of the eight COAST episode days modeled to this point.

Date	Max Model Ozone Domainwide (ppb)	Max Model Ozone Stationwide (ppb)	Max Obs Ozone (ppb)	Unpaired Peak Accuracy	Normalized Bias	Normalized Gross Error	Mean Simulated (ppb)	Mean Obs (ppb)
8/31/93	104.1	86.4	96	.085	-.185	.317	59.1	75.1
9/01/93	**161.9**	**131.5**	**164**	**-.013**	**.125**	**.286**	**89.7**	**84.3**
9/02/93	**132.5**	**123.9**	**139**	**-.047**	**.014**	**.184**	**80.8**	**81.8**

Table 1. Model performance statistics. Values in bold are for the primary episode days.

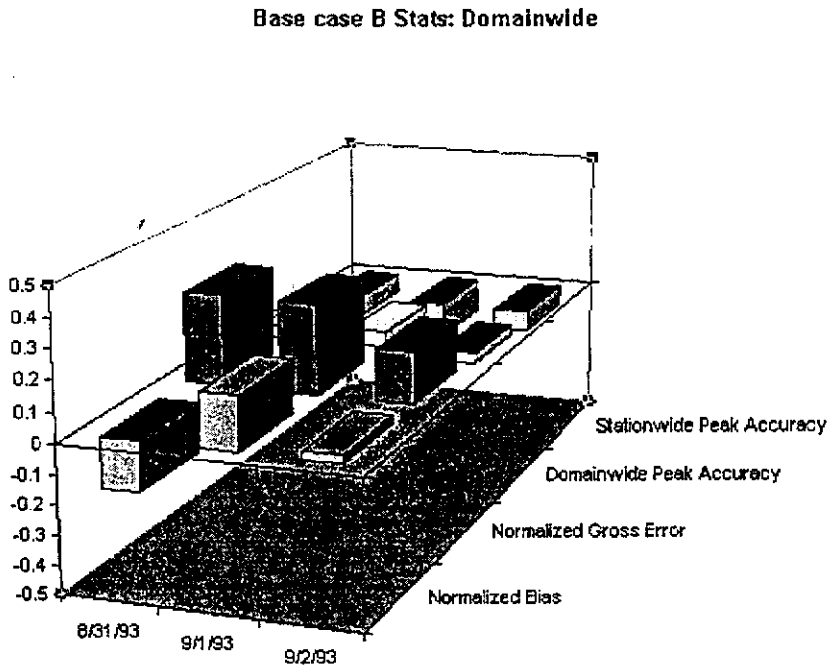

Figure 1a. COAST CAMx domainwide performance statistics for the 31 August - 2 September time period.

These calculations were also made looking specifically at the nonattainment subregions (H-G[1], and Beaumont-Port Arthur[2]). Figures 1b and 1c show the results. All of the subregional

[1] Includes all monitors in Harris, Galveston, Chambers, and Brazoria counties.
[2] Includes all monitors in Orange, Jefferson, and Hardin counties.

statistics are indicative of adequate model performance. Appendix A provides the complete performance tables for each day of the episode and all three regions.

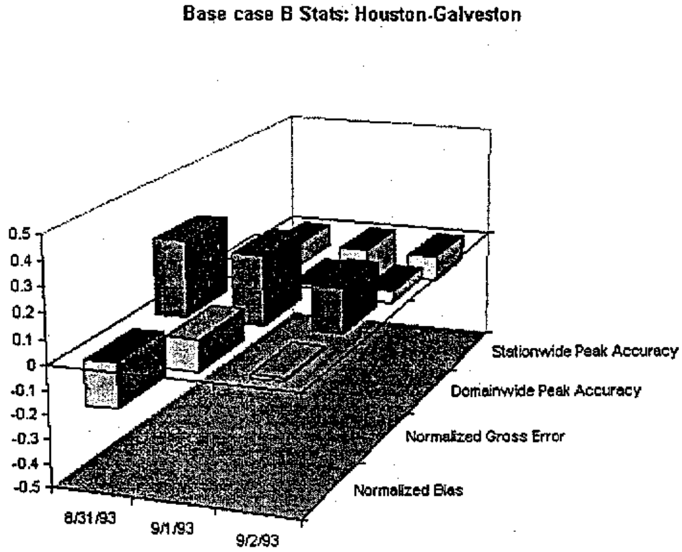

Figure 1b. COAST CAMx performance statistics in Houston-Galveston nonattainment counties.

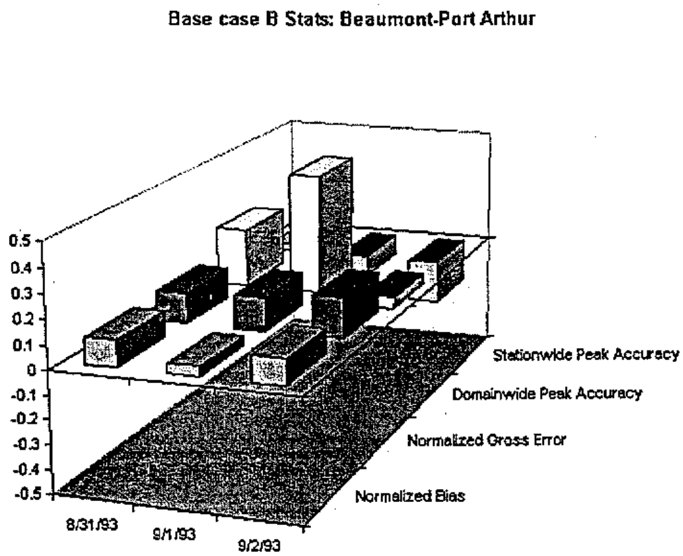

Figure 1c. Comparisons of COAST CAMx performance statistics in Beaumont-Port Arthur .

3. QUALITATIVE EVALUATION OF MODEL PERFORMANCE

Figure 2 displays a scatter plot of model-observed pairs for all hours of the 8/31 through 9/02/93 episode. Generally, the performance of the CAMx COAST modeling for this episode is comparable to many other photochemical modeling exercises. In particular, the inability of the model to reproduce the highest ozone levels at specific sites is a feature often seen in this type of modeling. While underestimating the (relatively limited) highest ozone concentrations, the model tends to overestimate observations in the less than 60 ppb range. This leads in part to the occasional positive bias seen in the performance statistics and is also not uncommon in modeling applications such as this.

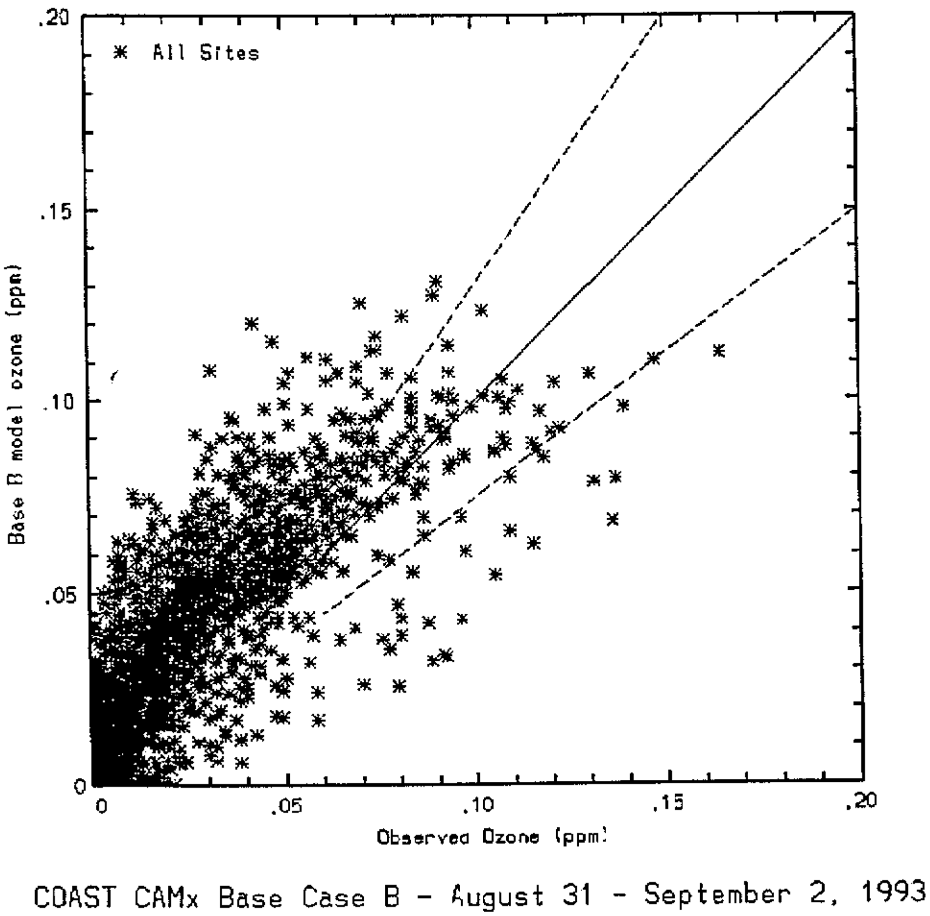

Figure 2. Scatter plot of ozone observations vs. model ozone for the 8/31 to 9/02/93 COAST episode. Pairs were generated for all hours and monitoring data points. The solid line represent non-biased model predictions. The dashed lines bound a region of acceptable model bias.

Appendix B contains model isopleth plots with overlaid observations for 1000, 1200, 1400, 1600, 1800, and 2000 CST for each day of the episode. These plots allow for a limited evaluation of the spatial accuracy of the base case simulation results. Descriptions of the qualitative differences between the afternoon observed and model fields follow.

Observations on the afternoon of the 1st day of the episode are quite low. Most sites in the Beaumont-Port Arthur area topped out in the 50-60 ppb range. The ambient ozone was also low in Harris County where the highest observed value was near 80 ppb. The model accurately predicts the relatively low ozone over COAST region on the 31st. As seen in Figure 3, the model generates a plume of ozone greater than 80 ppb to the north of the B-PA area in a region with only limited monitoring data. The available data does indicate that the ozone was higher in the northern parts of the domain.

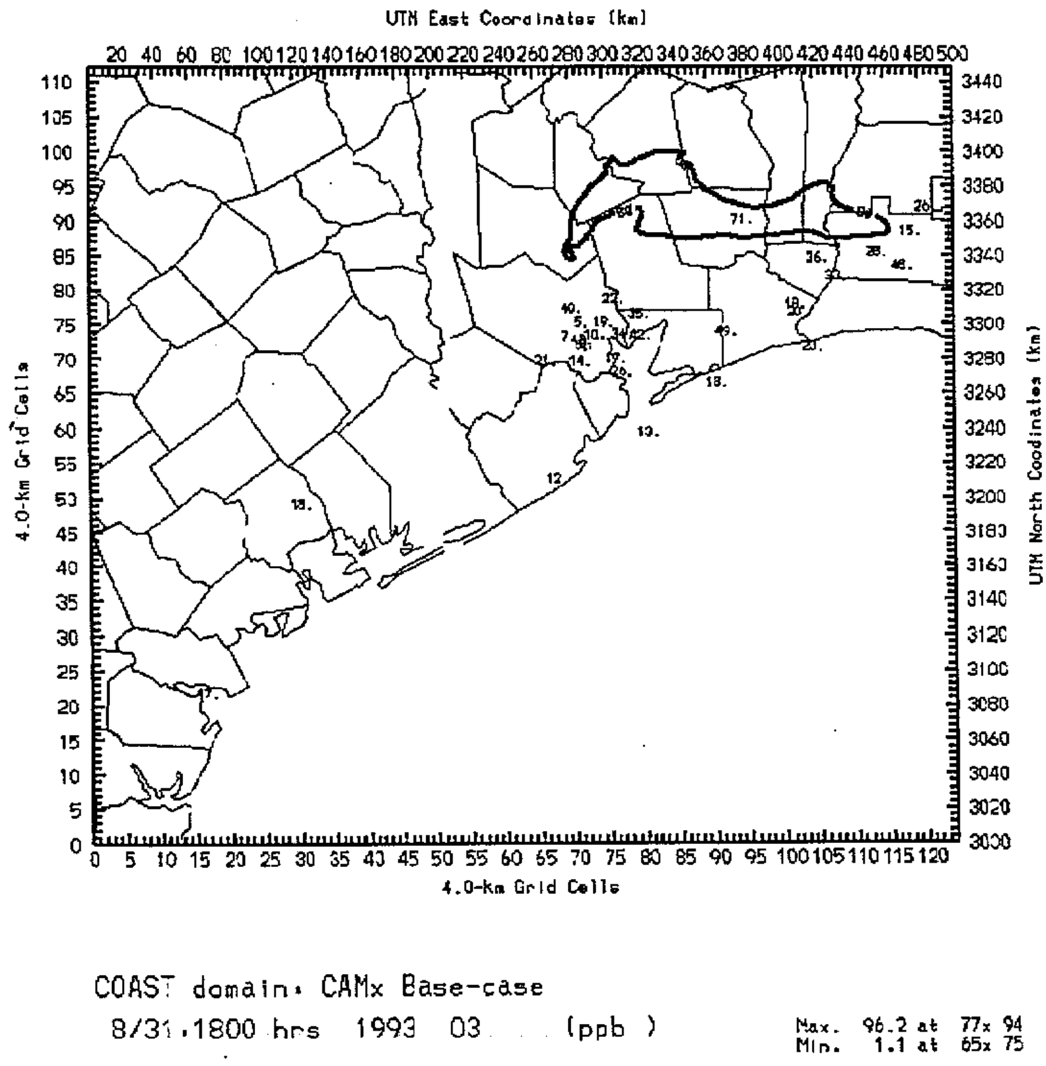

Figure 3. Model ozone isopleth plot with observations overlaid for 6 p.m. CST on 31 August. Contours of 80, 120, and 160 ppb are plotted. Maximum and minimum model predictions for this hour are plotted in the lower right portion of the graph.

By the afternoon of the 1st, ambient ozone concentrations are considerably higher than the preceding day. At noon, there is a peak of 136 ppb at one of the Galveston Bay sites (Smith Point). By 4 p.m. CST, this peak has migrated to the northeast and is near Gilchrist TX with a magnitude of 122 ppb. Finally at 6 p.m., the peak had advected out of Chambers County and into Jefferson County, part of the three-county B-PA nonattainment area, with a value of 118 ppb. There is a very sharp ambient ozone gradient between the location in southwest Jefferson County and those sites close to Beaumont TX which are in the 40-70 ppb range.

Figure 4 shows the model ozone fields at 4 p.m. on 1 September with the ambient observation data overplotted. As discussed in Section 2, this day features predominantly model overestimates. The model simulates a large are of ozone greater than 80 ppb from near Freeport TX, north through the eastern half of Harris County and then east through southern Jefferson and Orange counties into the Gulf of Mexico. There is a closed contour of 120 ppb or higher ozone over far eastern Harris County. In general, the model is displaying a large positive bias for those sites in the Houston region while simultaneously reproducing ozone along the COAST shoreline sites fairly accurately.

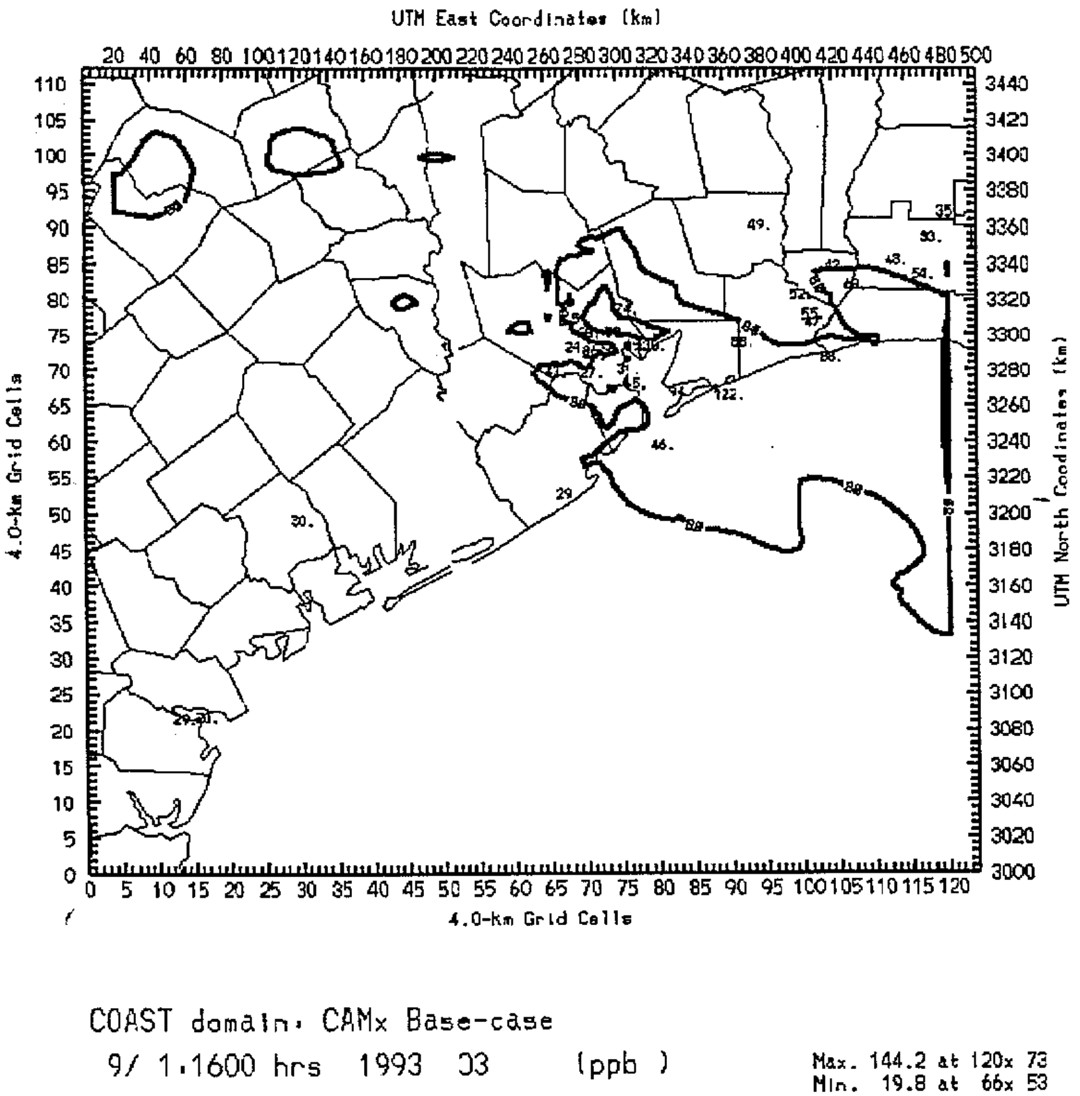

Figure 4. Model ozone isopleth plot with observations overlaid for 4 p.m. CST on 1 September. Contours of 80, 120, and 160 ppb are plotted.

Model performance is exceptional on the last day of the episode. The simulated fields were relatively non-biased with low error values. On 2 September, there was a relatively large region of ambient ozone ranging from 80 to 110 ppb from just north of the Houston area through Beaumont-Port Arthur and into western Louisiana. The model captures this pollutant fields over this region very well as is seen in Figures 5a, 5b, and 5c.

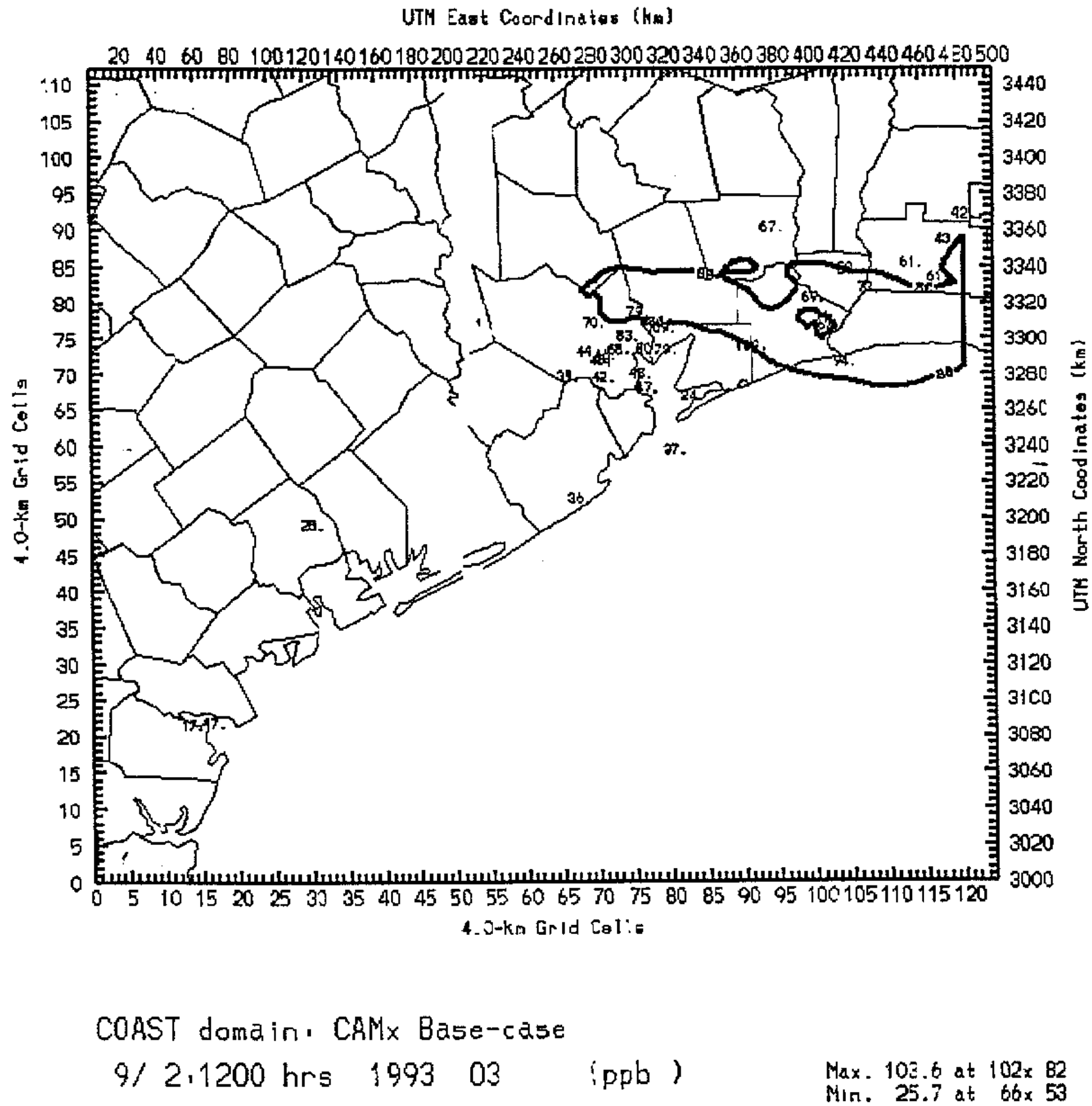

Figure 5a. Model ozone isopleth plot with observations overlaid for 12 p.m. CST on 2 September. Contours of 80, 120, and 160 ppb are plotted.

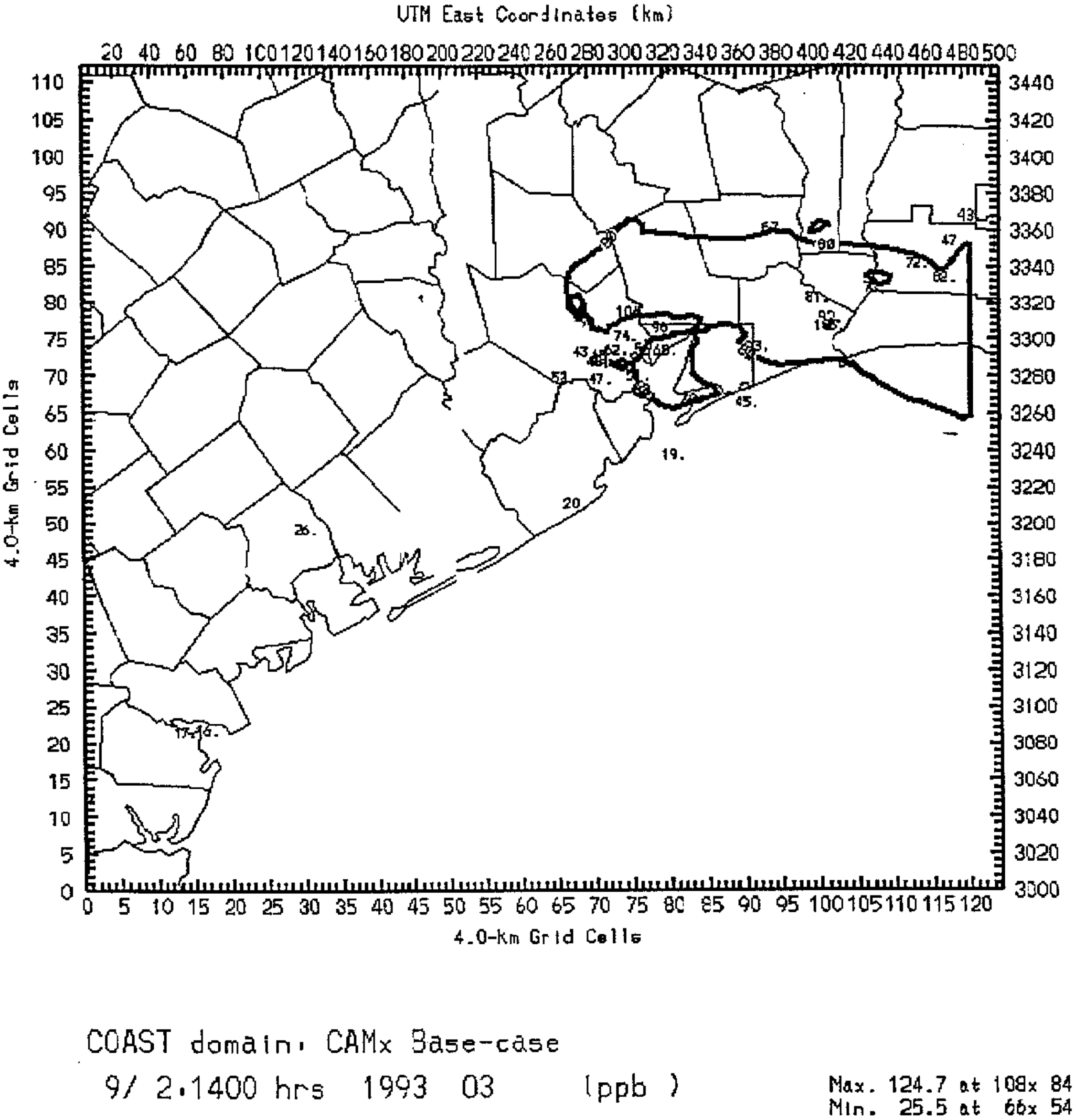

Figure 5b. Model ozone isopleth plot with observations overlaid for 2 p.m. CST on 2 September. Contours of 80, 120, and 160 ppb are plotted.

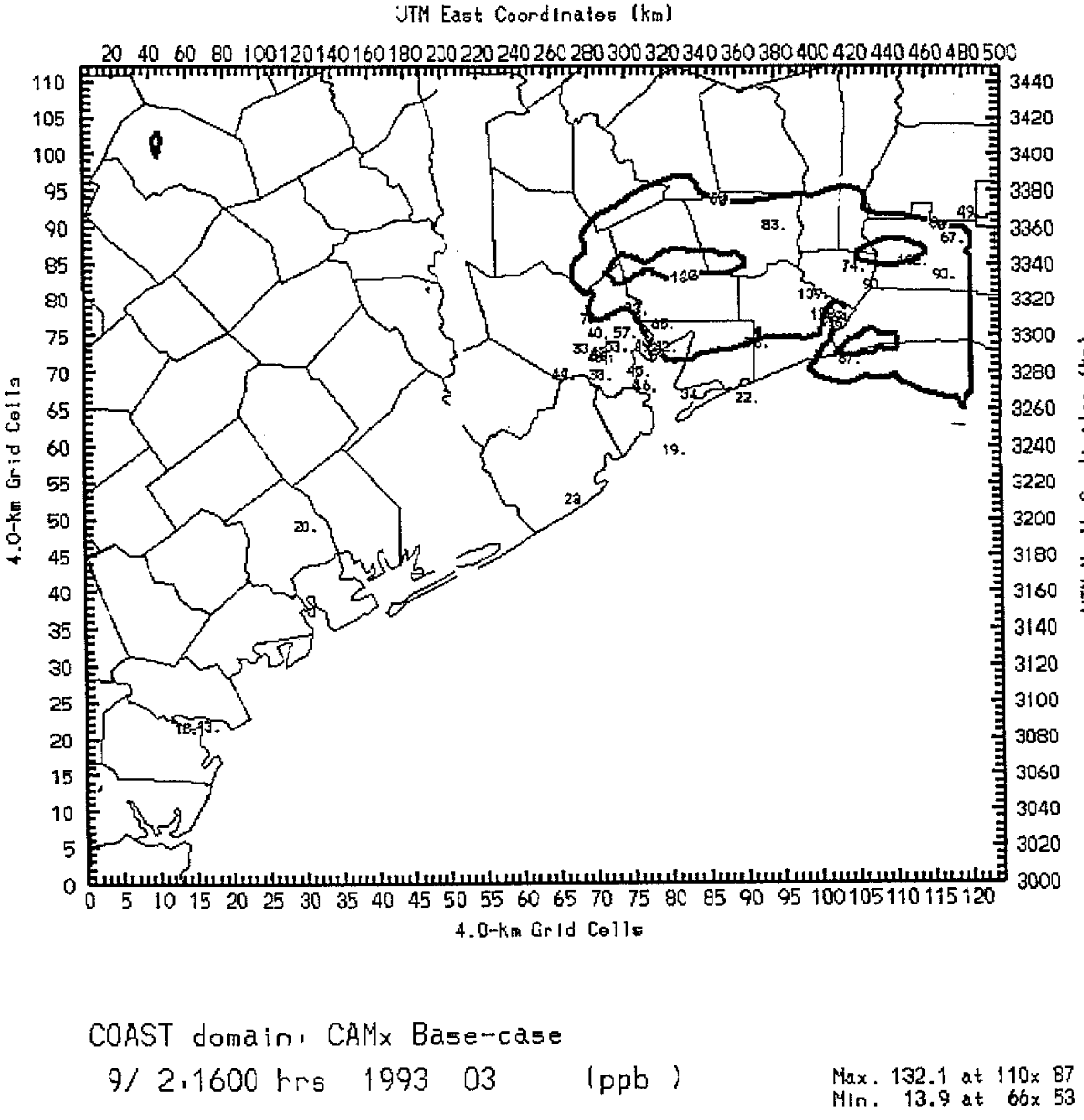

Figure 5c. Model ozone isopleth plot with observations overlaid for 4 p.m. CST on 2 September. Contours of 80, 120, and 160 ppb are plotted.

4. SITE-BY-SITE ASSESSMENT OF MODEL PERFORMANCE

This section will attempt to assess the quality of the B-PA base case modeling on a site-by-site basis. As discussed above, assessing model performance statistics can be difficult when comparing a model grid volume average against an observation at a single point in space. The station time series plots analyzed in this section are based on a bilinear interpolation from the four model grid cells surrounding the monitor location. Please remember when viewing these plots that there can be instances when a monitor data point is not representative of an entire grid cell. For the purpose of brevity only a handful of plots will be shown in the body of this document, but model versus observed comparisons can be found for all COAST locations in Appendix C (time series plots) and Appendix D (scatter plots).

There are seven monitoring locations within the three-county B-PA nonattainment area as discussed below.

<u>BEAUMONT C02/JEFFERSON CO TX:</u> As shown in Figure 6a, model ozone is overestimated on 1 September as discussed earlier. The model performs fairly well at this during the remainder of the episode, with slight overprediction during the early morning on the 2nd but a relatively accurate peak prediction.

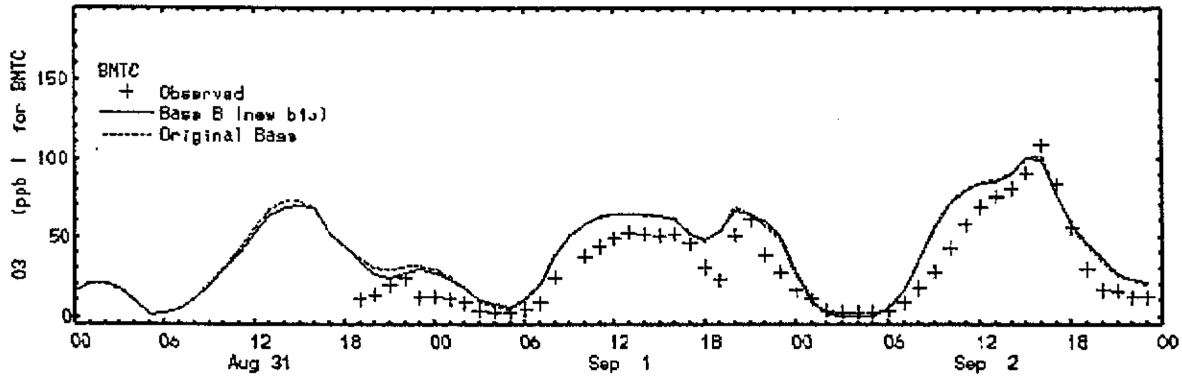

Figure 6a. Time series of observed (plus signs) and modeled ozone (solid line) at the BMTC site for the 31 August – 2 September COAST episode.

<u>SETRPC SITE 43 TX:</u> At this site in Port Arthur TX, the model overpredicts observed ozone during the evening of the 31st by about 40-60 ppb due to an overabundance of ozone transport from the Houston-Galveston area. The model is also unable to replicate the sharp observed peaks of the 1st or 2nd. See Figure 6b for the time series display.

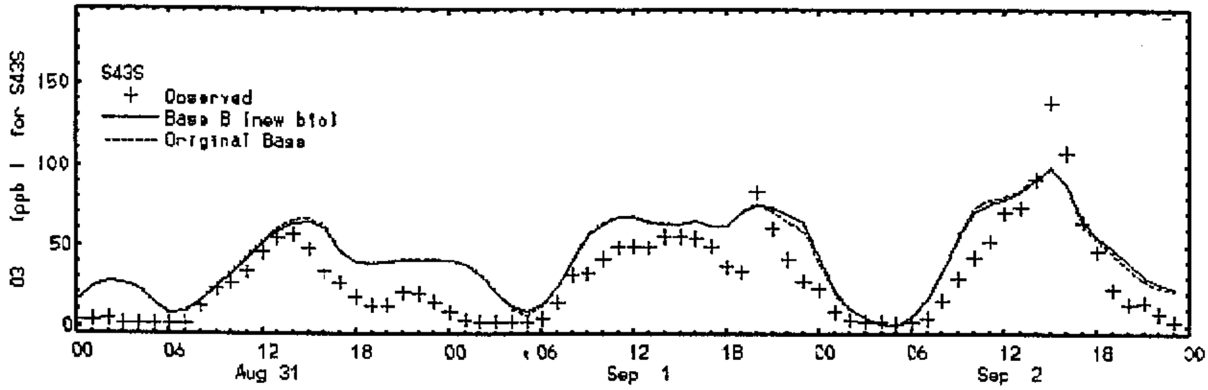

Figure 6b. Time series of observed (plus signs) and modeled ozone (solid line) at the S43S site for the 31 August – 2 September COAST episode.

PORT ARTHUR WEST C28 TX: Model performance at this site (see Figure 6c) is very similar to the other Port Arthur site. The model has difficulty reproducing the early evening peak of about 80 ppb on the 1st and the mid-afternoon peak of approximately 120 ppb on the 2nd.

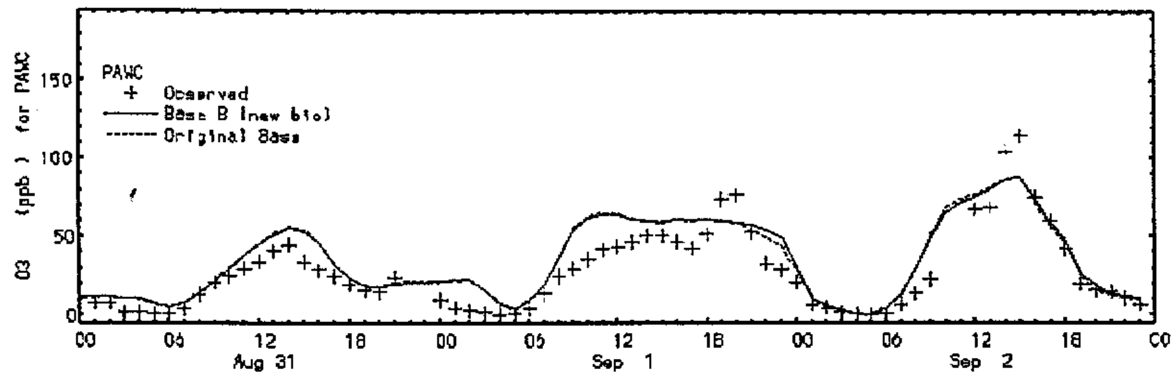

Figure 6c. Time series of observed (plus signs) and modeled ozone (solid line) at the PAWC site for the 31 August – 2 September COAST episode.

SETRPC SITE 40 TX: At this site located near Sabine Pass TX along the Gulf of Mexico shoreline, ozone is strongly overestimated on 31 August. The remainder of the episode features accurate model performance with a slight tendency for model ozone to stay higher longer into the evening than observed. See Figure 6d for more details.

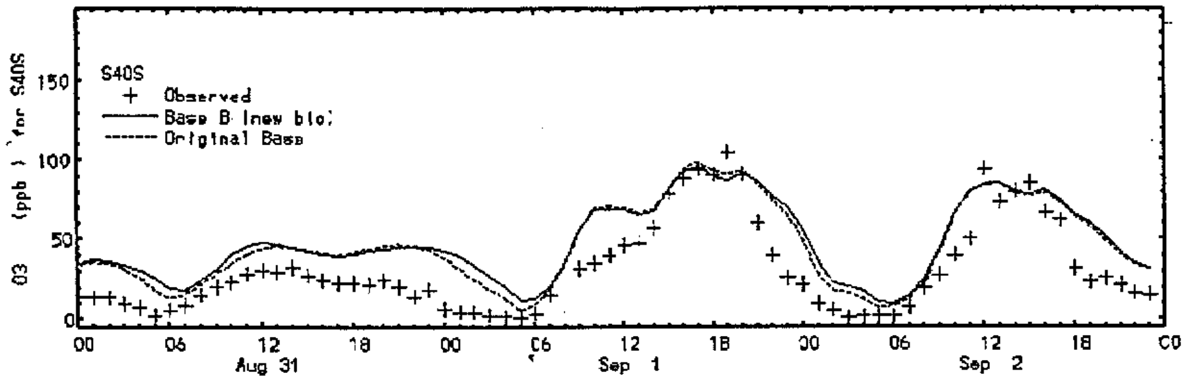

Figure 6d. Time series of observed (plus signs) and modeled ozone (solid line) at the S40S site for the 31 August – 2 September COAST episode.

SETRPC SITE 42 TX: This monitor is located in central Orange county. The most noticeable feature is the failure of the model to reproduce the low ozone observed on the 1st. While the observations drop down to less than 50 ppb, the ozone concentrations remain about 70-80 in the model, as seen in Figure 6e. In general, this site is marked by an rather high positive model bias.

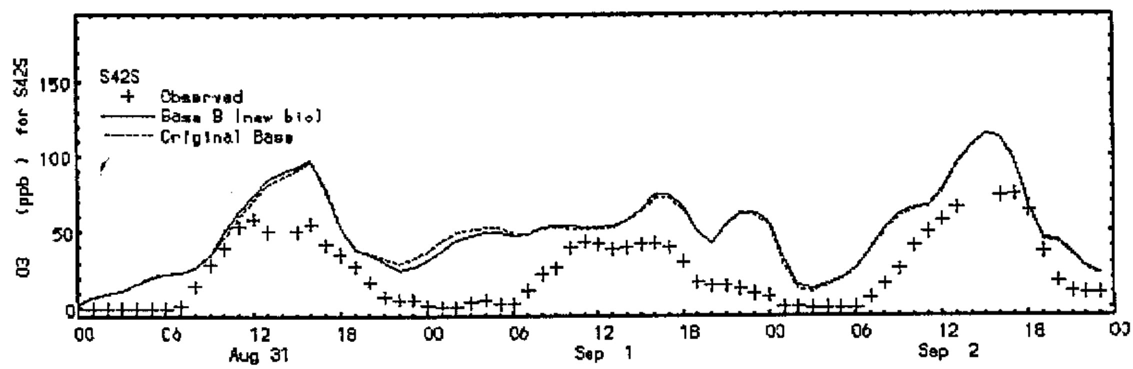

Figure 6e. Time series of observed (plus signs) and modeled ozone (solid line) at the S42S site for the 31 August – 2 September COAST episode.

KOUNTZE C85/HARDIN CO TX: The Kountze monitor is located in a rural region about 40 km north of Beaumont. This site observed high amounts of ozone on 2 September (approximately 120 ppb), as seen in Figure 6f. The model captures this peak reasonably accurately, although the ozone buildup occurs much more gradually in the model than in the ambient data. Note that the model does not reproduce the presumably local titration effects at night.

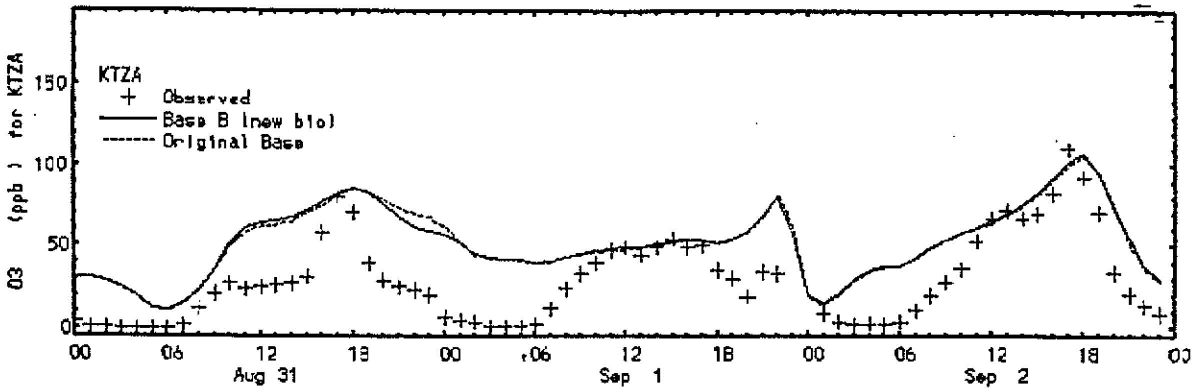

Figure 6f. Time series of observed (plus signs) and modeled ozone (solid line) at the KTZA site for the 31 August – 2 September COAST episode.

CAMS 9, West Orange, TX: Model behavior is similar at this Orange county site as in the SETRPC SITE 42. Note the positive model bias on both episode days shown in Figure 6g.

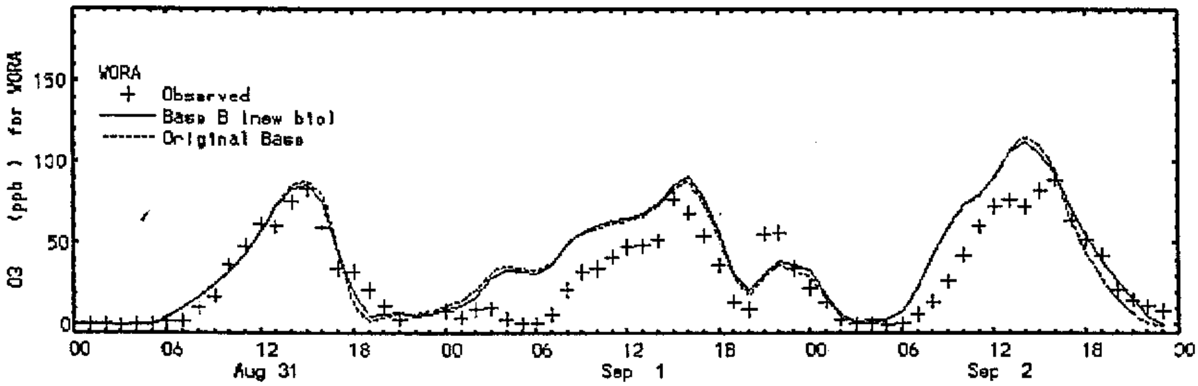

Figure 6g. Time series of observed (plus signs) and modeled ozone (solid line) at the WORA site for the 31 August – 2 September COAST episode.

5. DIAGNOSTIC EVALUATION AND TESTING

Diagnostic tests are designed to check the model formulation and the response to various inputs. Given an unlikely extreme input condition (e.g., zero emissions), the model results are checked to determine that the model produces an appropriate response. Generally, a series of tests are run on the emissions, meteorological inputs, and boundary conditions. Two simple diagnostic simulations were performed for the B-PA episode (8/31 to 9/2/93), as follows:

- zero all emissions (anthropogenic and biogenic)
- zero boundary and initial conditions
- reduce wind speeds by 50 percent

The goal of diagnostic testing is to determine whether the model exhibits expected behavior under extreme changes to its input. Diagnostic tests check for spurious behavior that might indicate problems in the model formulation. Overall, the results of the three CAM_X diagnostic tests were generally consistent with expectation.

Removing all emissions (anthropogenic and biogenic) from the COAST modeling data set sharply reduced ozone over the domain (up to 110 ppb). The ozone reductions were greatest as expected in the areas which local emissions were most culpable for the high simulated ozone. Figure 7 shows the simulated ozone fields for an afternoon hour on the third day of the episode before and after the removal of emissions.

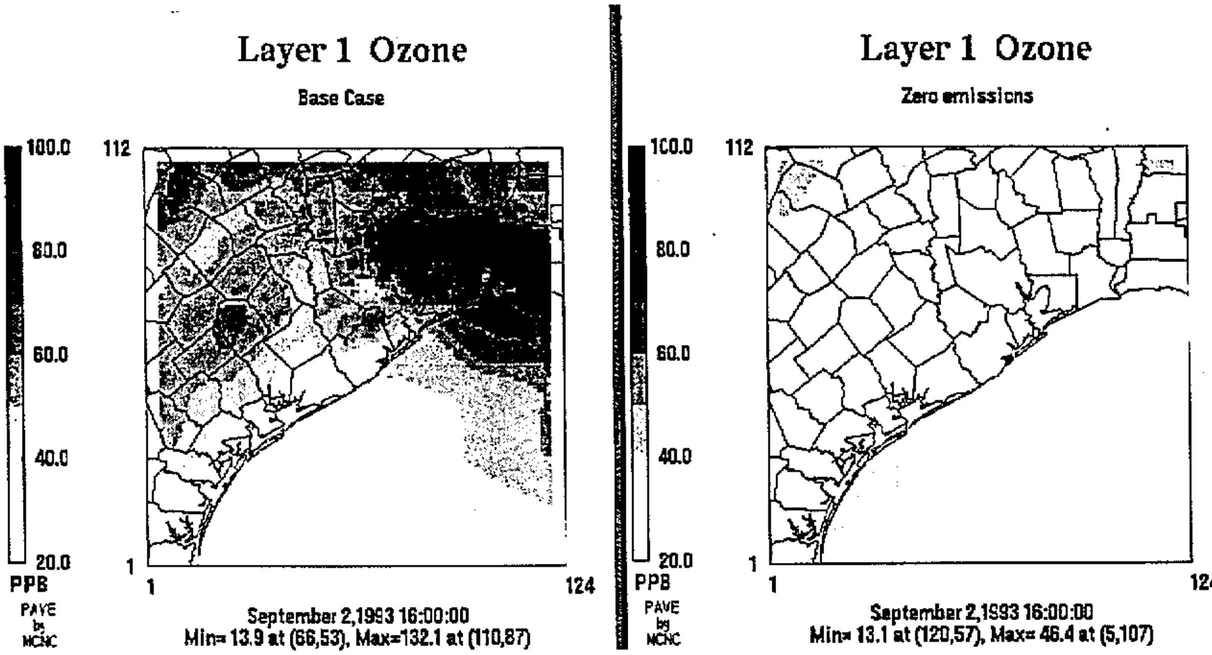

Figure 7. Simulated ozone concentrations for 1600 CST, 2 September 1993. Left side of the plot is the field from the base case simulation. The right side of the plot represents the simulation with zero emissions.

The zero initial and boundary condition test proved somewhat troublesome for CAM_X 1.13. The simulation crashed about 18 hours into the simulation with the message:

```
No Convergence in TRAP: kount, dt, time =   1  7.629E-06  0.000E+00
errbig, rerr = 9.864E-034.195E-05
```

The sharp gradient caused by zero boundary conditions along the border of the domain essentially caused the chemical mechanism to become unable to solve the necessary equations.

This same issue was also seen in previous diagnostic modeling over the Dallas-Fort Worth domain (Yarwood, 1999). One of the revisions in CAM_X version 2.00 was designed to accommodate tests of this nature. The inability of the model to handle this test is not expected to have any impact of the eventual intent of this modeling exercise (i.e., an attainment demonstration) as sharp artificial gradients such as these will not exist in that modeling.

The results of the half wind speed diagnostic test confirmed the findings from the same test over the 6-11 September episode (MCNC, 1999a). Daily peak ozone values increased (20-60 ppb), mostly in grid cells adjacent to the urban areas. Model performance is deteriorated when the wind velocities are adjusted in this manner. The sharp sensitivity of the modeling results to changes in scalar winds is expected and suggests the model is not mistreating advection.

6. CONCLUSIONS AND RECOMMENDATIONS:

In general, CAM_X **model performance is acceptable** for the August 31st to September 2nd 1993, COAST episode. This is based primarily on the statistical analyses presented in Section 2 of this report, in addition to the qualitative (Section 3) and site-by-site (Section 4) assessments. Additionally, the results from the diagnostic testing bolster the acceptability of the modeling exercise. While the model appears to be sufficient for use as technical support to guide air quality planning issues, there are a few modeling features that should be recognized as this work proceeds.

- Model performance on 2 September 1993 is very good. The statistics for this day indicate a more accurate replication of observed fields than any of the other COAST modeling days.

- There is generally a positive bias (overprediction) in the base year CAM_X results on 1 September.

- As is usually the case, the model underestimates ozone observations in the high end of the observation spectrum and overestimates ozone in the lower ranges (i.e., less than 60-70 ppb).

- Model performance statistics also indicate acceptable model performance when considered on a subregional (Beaumont-Port Arthur and Houston-Galveston) basis.

- None of the diagnostic/sensitivity tests indicated any major problems with the CAM_X model formulation. In general, the model performed as expected, as a function of prescribed input changes.

In terms of possible model improvements, additional evaluation and revision of the meteorological fields would likely yield the greatest improvements to CAM_X model performance. Previous meteorological evaluations have uncovered errors and biases in the original COAST meteorological modeling (TNRCC, 1998b). In particular, the thermal flows (land breeze/ sea breeze) along Galveston Bay and the Gulf of Mexico should be closely scrutinized.

Also, an additional element of uncertainty in a photochemical modeling exercise such as this is the emissions estimates. TNRCC has devoted considerable effort to developing an accurate emissions inventory. Sensitivity runs performed throughout the course of the study based on inventory revisions have shown that simulated ozone is highly sensitive to changes in urban emissions in particular (MCNC, 1999b). It is recommended that effort be made to "evaluate" the inventory estimates by reconciling them against ambient data whenever possible. Analyses such as these, in combination with a successful model performance evaluation, further bolster the confidence one has in using a model to guide air quality attainment planning.

7. REFERENCES

ENVIRON International Corporation, 1997: *User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx).* Novato, CA.

MCNC, North Carolina Supercomputing Center, 1999: *Evaluation of CAMx Base Case Model Performance for the September 6-11, 1993 COAST Episode.* Research Triangle Park, NC.

MCNC, North Carolina Supercomputing Center, 1999: *Incorporation of Emissions Improvements for the Beaumont-Port Arthur CAMx Modeling.* Research Triangle Park, NC.

Tesche, T.W., 1990: *Improvement of Procedures for Evaluating Photochemical Models.* Final Report. Prepared for the California Air Resources Board.

Texas Natural Resource Conservation Commission, 1998: *Protocol for Ozone Modeling of the Beaumont-Port Arthur area with the Comprehensive Air Quality Model with Extensions (CAMx) and COAST Data.* Austin, TX.

Texas Natural Resource Conservation Commission, 1998: *Revisions to the State Implementation Plan for the Control of Ozone Air Pollution. Attainment Demonstration for the Houston/Galveston Ozone Nonattainment Area.* Austin TX.

Yarwood, Greg, 1999: Personal communication.

Due to the length, the appendices associated with this report
are not available in electronic file.

Please contact Eve Hou, (512) 239-5838 or ehou@tceq.state.tx.us, of the
TCEQ to attain a hardcopy version.