G-35-803
#20 (New)

Inverse Modeling for Atlanta, Georgia

Final Progress Report (GTRC contract #771-390279)

Submitted to
Air Protection Branch
Environmental Protection Division
Georgia Department of Natural Resources
4224 International Parkway, Suite 120
Atlanta, Georgia 30354

Prepared by
School of Earth and Atmospheric Sciences
Georgia Institute of Technology
Atlanta, Georgia 30332-0340

Carlos Cardelino
Co-Principal Investigator

Dana Hartley
Co-Principal Investigator

Other Investigators / Students
Michael E. Chang
Wen-Ling Chang
Danielle Haas-Laursen

July 1, 1996

Table of Contents

I. Executive Summary

In our proposal of April 25, 1995, we planned to continue the inverse modeling project to resolve the Urban Airshed Model (UAM) predictions of isoprene, carbon monoxide (CO), volatile organic compounds (VOCs), and nitrogen oxides (NOx) with observations taken during the 1992 Southern Oxidants Study (SOS) Atlanta Intensive. Isoprene has been completed. Due to an earlier error in our inverse code, previous findings regarding isoprene are superseded by the results presented here. These revised results have been submitted for peer review and publication in *Geophysical Research Letters*. Using pseudo-data, we were able to successfully run one inverse for CO and one for NO. However using real data, we have not been able to use the inverse method to resolve the emissions of either of these species. We believe that the difficulty lies in the spatial distribution of the emission sources. Whereas isoprene sources (all biogenic) are numerous, small, and ubiquitous, sources of CO and NO (and anthropogenic sources of VOCs too) are large, relatively concentrated, and highly spatially dependent. A small error in the spatial allocation of these sources in the emission inventory may cause a large difference between model predicted ambient concentrations and observations at any one point in space and time. Because the Kalman filter is driven by the differences between observations and model predictions, these small errors are translated into large fluctuations in the emissions. The resultant emissions reflect the errors in the spatial allocation of the inventory rather than the magnitude of the emissions. Unfortunately, there is not enough observation data to "unlock" the spatial constraints and use the inverse method to solve for the magnitude *and* spatial distribution of the anthropogenic emissions.

The remainder of this report describes the updated isoprene results and documents the CO and NOx inverse attempts. Anthropogenic volatile organic compounds share many of the same characteristics as CO and NOx. Thus, due to the problems associated with CO and NOx, we did not attempt to run an inverse with VOCs.

II. Inverse Modeling for Atlanta, Georgia

1. Isoprene - Revised Results

Since our last presentation, we discovered coding errors in our inverse application. These errors were in the initialization of isoprene at each hour and in some of the details of the error matrices. The code corrections resulted in emission correction factors that differed from the values we originally reported, however the conclusions are the same. The table below summarizes the original values and the revised values.

Inverse Derived Relative Scaling Factors w.r.t. BEIS or BEIS2

	BEIS	BEIS2	Split-BEIS		Split-BEIS2	
			urban	rural	urban	rural
Original	18	4	17	20	3	5
Revised	7	2	10	6	2	2+

Due to the correction, negative emissions occur at various times throughout the day as shown in figure 1. At hour 7, the inverse method adjusted the emissions more to compensate for errors in the initial conditions than for errors in the emission inventory. The physical implication of the negative emissions at hours 11, 12, 16, 17 and 18 indicate that the model could be lacking a key isoprene removal mechanism, or that the inverse method is correcting for an overestimation of earlier emission adjustments. For hours 16 and 17, the model assumes clear sky although rain was recorded in Atlanta. The impact of this assumption on comparisons between observations and model predictions is not clear. Because of this uncertainty, the data are omitted from the isoprene emission inventory comparisons when rain may have interfered (hours 16, 17, and 18). No physical explanation corresponding to the negative emissions at hours 11 and 12 has been found. These emissions cannot be omitted because it is uncertain whether an unknown removal mechanism acted at these hours, or if the method is correcting itself for an over estimation of isoprene emissions during the morning hours (8, 9, and 10).

In addition to the change in the magnitude of emissions, results from the split-inventory runs show that the spatial allocation of emissions is also different than the distribution originally reported. Emissions in the urban areas are higher than previously shown. This finding is in better agreement with the latest emissions research (Geron et al. 1995).

The code changes and different emissions results also affected the model performance statistics. The table below shows the original and revised values.

Model Performance (Normalized Bias) relative to Observations

	BEIS	BEIS2	Inverse-adjusted BEIS	Inverse-adjusted BEIS2	Inverse-adjusted Split BEIS	Inverse-adjusted Split BEIS2
Original	-85%	-30%	+3%	+7%	+1%	+2%
Revised	-83%	-19%	-15%	-9%	-8%	-5%

The change in emissions did not affect the ozone simulation. Although the increase in emissions is smaller than originally reported, they are still sufficient to saturate the region with hydrocarbons. Additional isoprene (as in the original run) has no effect on the ozone beyond the stoichiometric level.

All of the above results are described in detail in the Appendix: a copy of our revised paper "Inverse Modeling of Biogenic Isoprene Emissions" submitted to *Geophysical Research Letters* on 6/15/96.

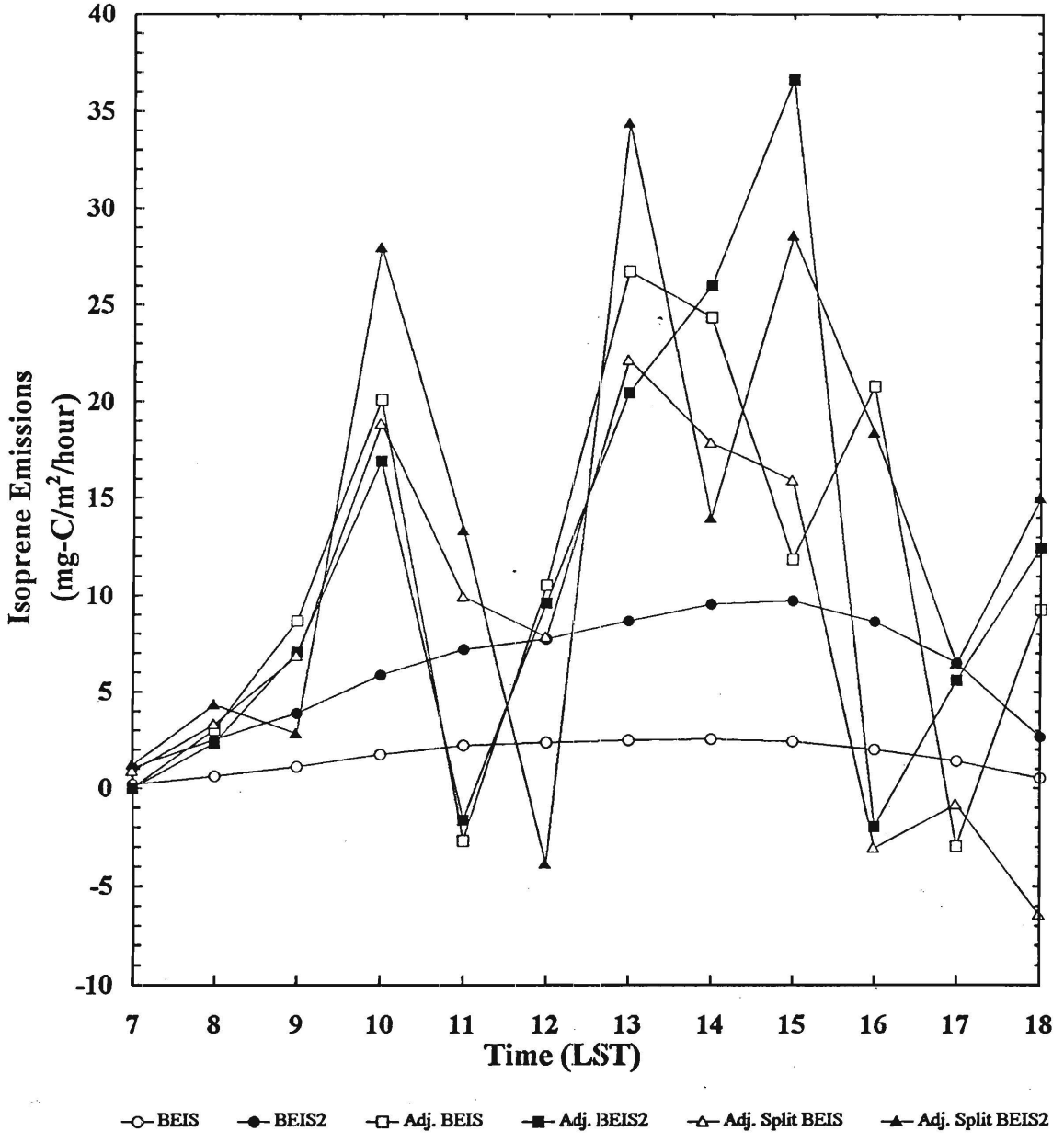

Figure 1. Revised inverse derived average isoprene emissions for Atlanta, Georgia (August 10, 1992).

2. Carbon Monoxide (CO)

On August 10, 1992, there were four (4) carbon monoxide (CO) monitors operating in the Atlanta Area. The Georgia Tech site (TECH) was the only true urban site. Both South Dekalb (SDEK) and the Holcomb Bridge Road (HOLC) sites may be considered as suburban. Finally the Yorkville site (YORK) was the only rural station. Figure 2 shows the observed concentrations and initial UAM predicted (prior to any adjustments) concentrations. At the TECH site, CO concentrations are overpredicted by 270%; at the SDEK site, the UAM overpredicts CO by 6%; at HOLC, the model underpredicts CO by 41%; and at YORK, CO concentrations are overpredicted by 15%. In total, the UAM overpredicts CO concentrations by 64% (mostly due to the large error at TECH).

A. CO Inverse with Pseudo-data

A pseudo-data CO inverse run was used to determine the optimum settings for the Kalman filter. Area, nonroad mobile, and mobile sources were all merged into one gridded and temporally allocated source. Point sources were ignored since their contribution to the CO inventory was minor. This single source was then initially set to 160% of the baseline. The filter was then used to determine the base emissions using only the differences between the predicted CO concentrations (with initial emissions at 160% of the baseline) and the observed concentrations (pseudo-data - CO concentrations extracted from a baseline run). Figure 3 shows the pseudo-data observations and the initially perturbed concentrations at each of the 4 sites. Figure 4 shows the inverse derived factors necessary to force the model to match the pseudo-data. In this case, a value of 1.00 implies a perfect convergence. Deviations from 1.00 are the percent error in emissions from true convergence. Thus, for a 60% error in the emissions, we concluded from this pseudo-data run that the inverse method is capable of resolving a single domain-wide CO source.

B. CO Inverse with Observations from the 1992 Atlanta Intensive

Using observations of CO from the 1992 Atlanta Intensive (see figure 2) and the August 10, 1992 (day specific) CO emission inventory, an inverse was run to estimate the temporal distribution of CO emissions assuming that the spatial distribution of the emissions is correct. Figure 5 shows that the inverse was successful in forcing the UAM predictions to match the observations. Overall, the bias was reduced from + 64% to -18%. In particular, the adjusted emissions vastly improved the simulation for the TECH site (at the expense of the SDEK site). The biases at the TECH, SDEK, HOLC, and YORK sites were +16%, -43%, -52%, and +8% respectively.

The inverse-derived CO emission adjustment factors are shown in figure 6. It is clear that the oscillations in the emissions are unrealistic. In an attempt to dampen these oscillations, we adjusted several filter input parameters and ran the inverse again. In this new run, the emissions covariance was decreased from 50 to 0.2. This parameter is a measure of the uncertainty in the emissions inventory (a large value indicates a large uncertainty and a willingness to allow the emissions to fluctuate freely to ensure the best fit between observations and model predictions; a small value indicates confidence in the emissions inventory and allows only small "fine-tune" emissions adjustments). The threshold for convergence was also increased to 5% from 1%. The method finds the best solution by iteration. It stops when the change in emissions is less than the threshold value. Increasing the threshold limits the amount that the final solution may differ from the initial guess by stopping the iterative process early (and perhaps prematurely). By "tuning" these parameters, a satisfactory and stable solution is obtained. Overall at all stations, the bias is reduced to just +3%. The biases at the TECH, SDEK, HOLC, and YORK sites were +73%, -24%, -47%, and +12% respectively (see figure 7). Further, the emissions seem to be more well behaved. Figure 8 shows that the CO inventory is overestimating CO emissions in the morning and late evening hours and is about correct during the daylight hours.

Originally, the parameters that were changed in the last run were initially set to optimize the performance of the pseudo-data run. They were then changed to optimize the performance of the real-data run. To determine if the run with the new parameters is a legitimate solution, the pseudo-data test was run again with the new parameters. Again, the emissions were perturbed to 160% of the baseline value. Rather than the relatively perfect solution seen in the initial pseudo-data run (see figure 4), figure 9 shows that the solution begins to degrade at hour 8 and could

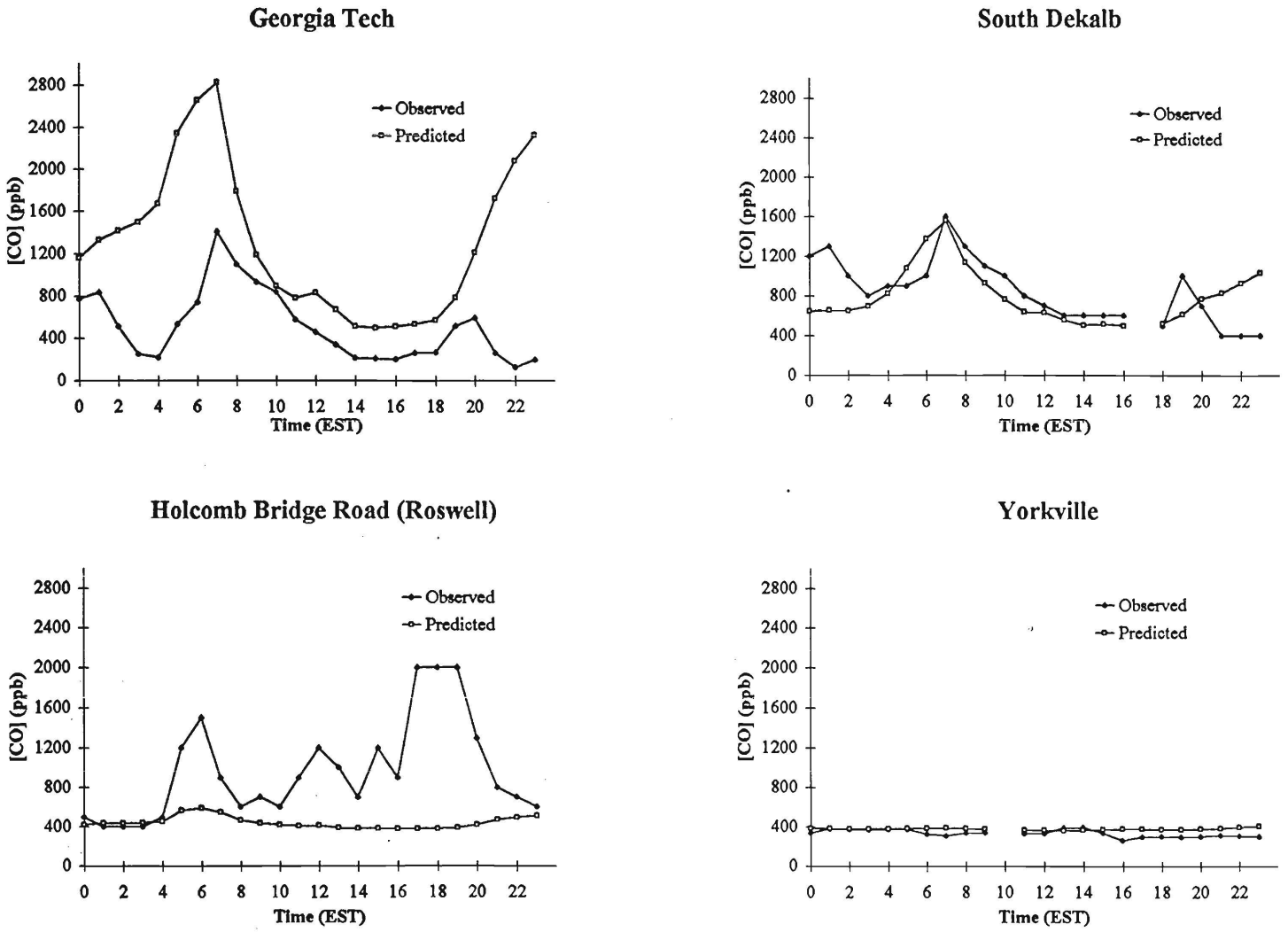

Figure 2. Observed and initial UAM predicted CO concentrations in Atlanta, Georgia (August 10, 1992).

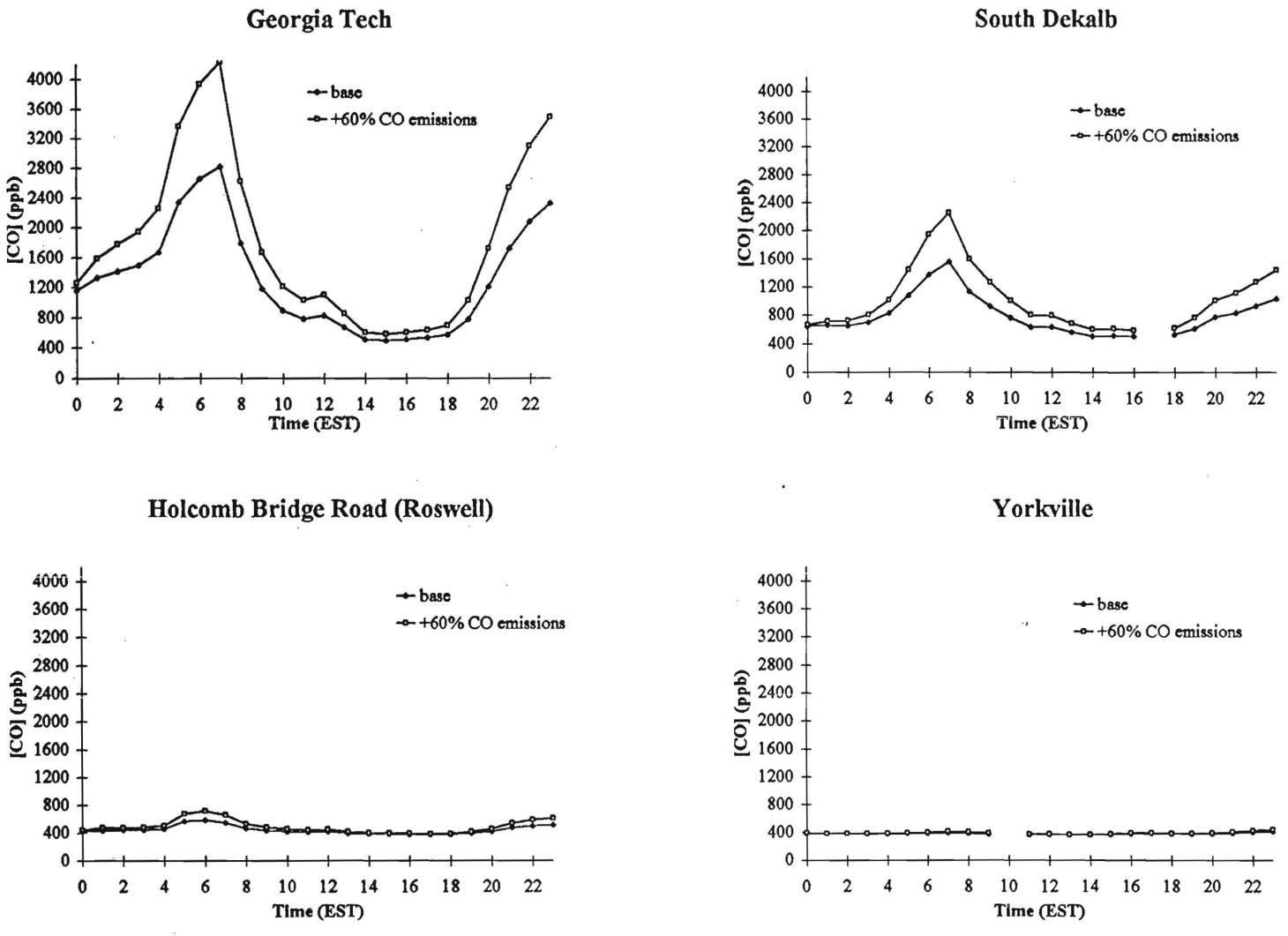

Figure 3. Base (pseudo-observations) and perturbed (+60% CO emissions) UAM predicted CO concentrations in Atlanta, Georgia (August 10, 1992).

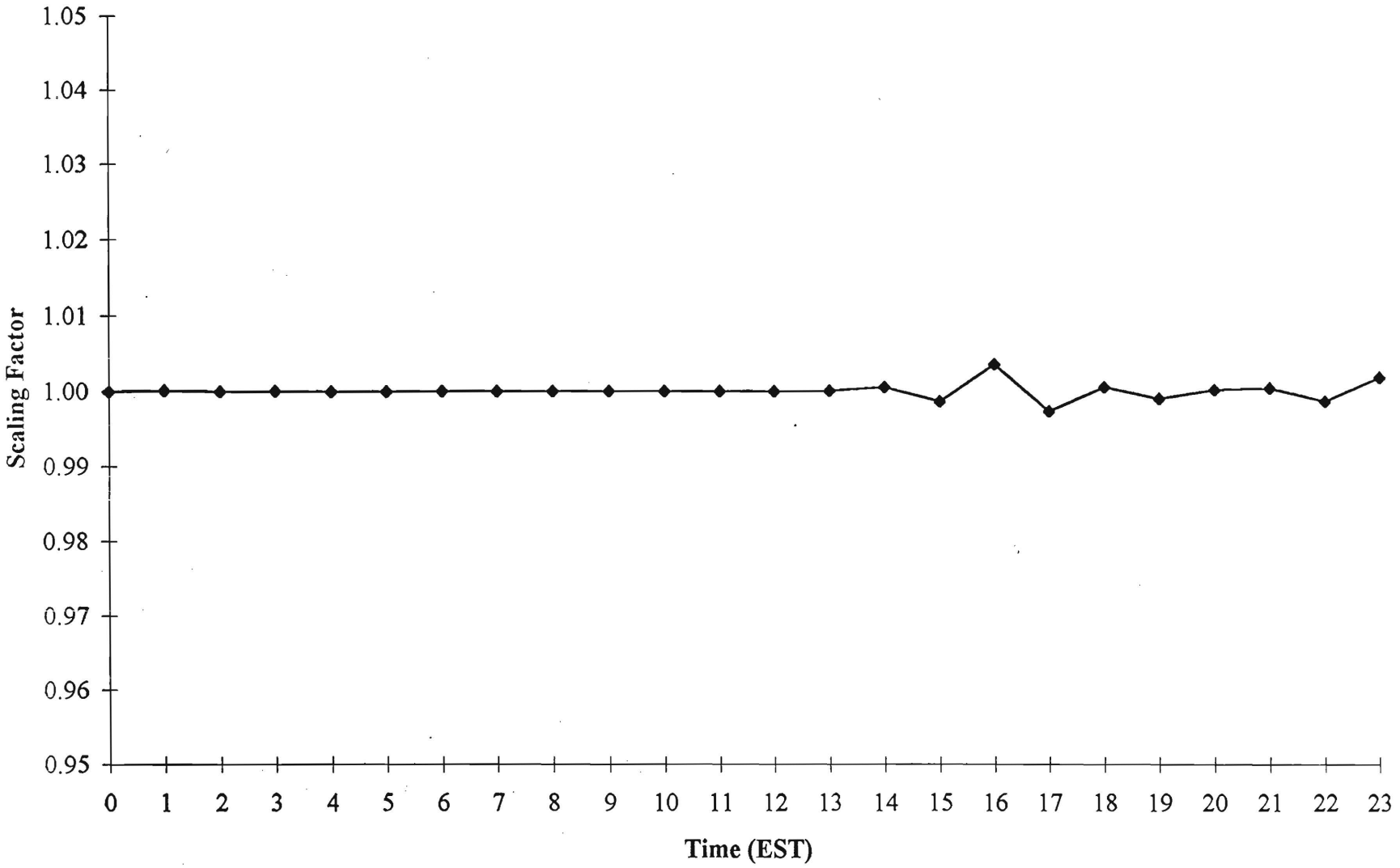

Figure 4. Inverse derived CO emission scaling factors for pseudo-data (60% perturbation).

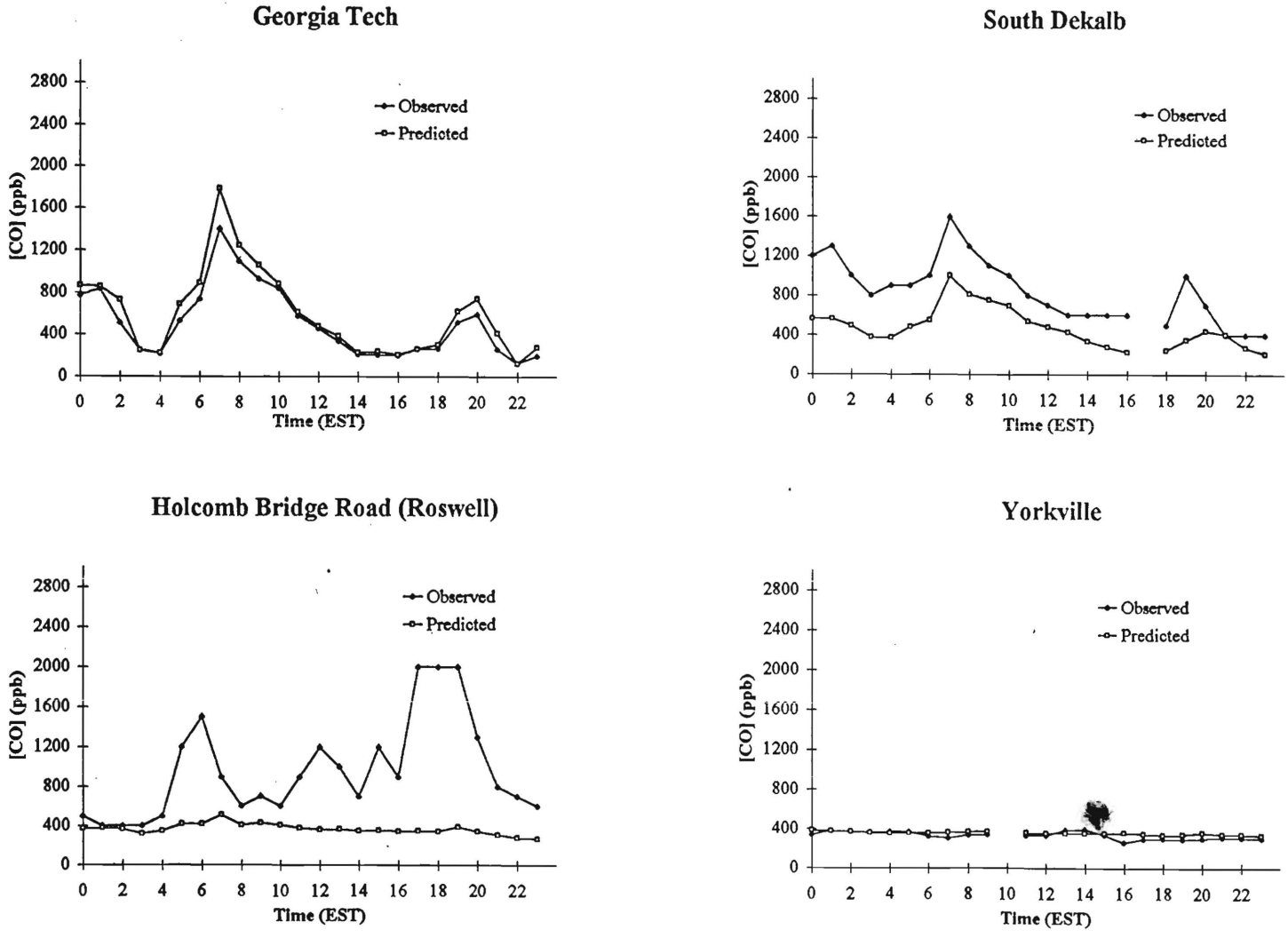

Figure 5. Observed and inverse adjusted - UAM predicted CO concentrations in Atlanta, Georgia (August 10, 1992).

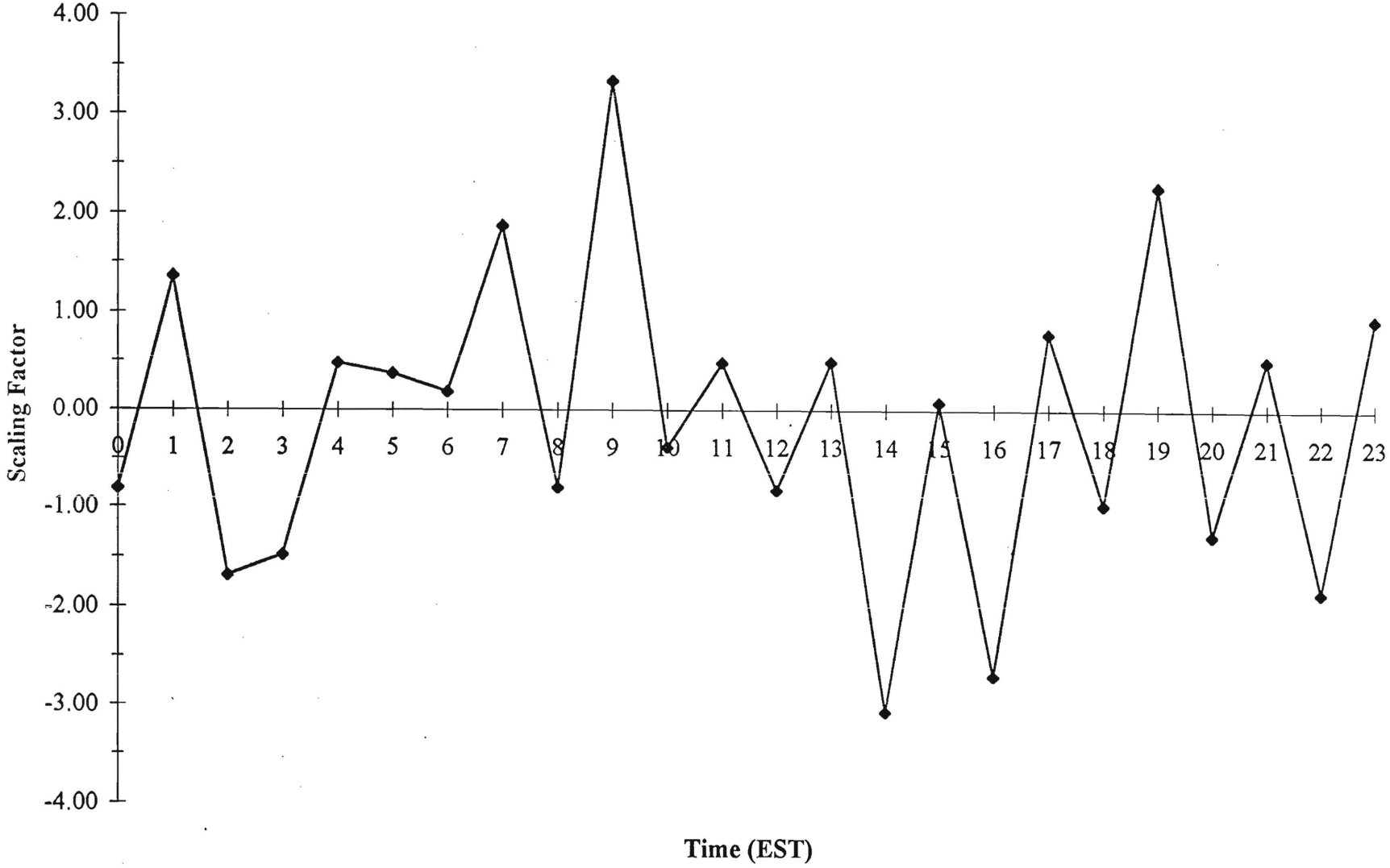

Figure 6. Inverse derived CO emission scaling factors.

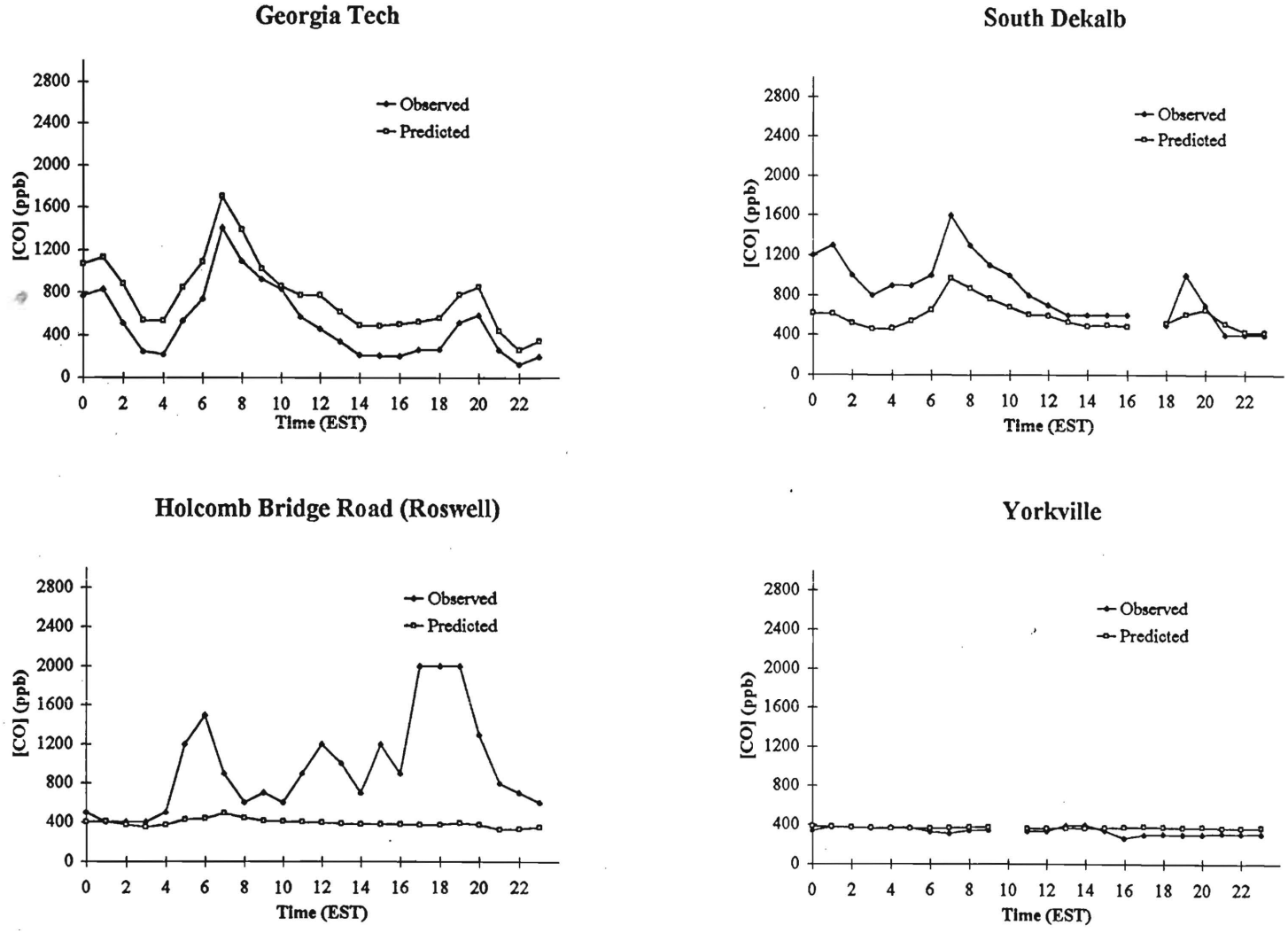

Figure 7. Observed and "tuned" inverse adjusted - UAM predicted CO concentrations in Atlanta, Georgia (August 10, 1992).

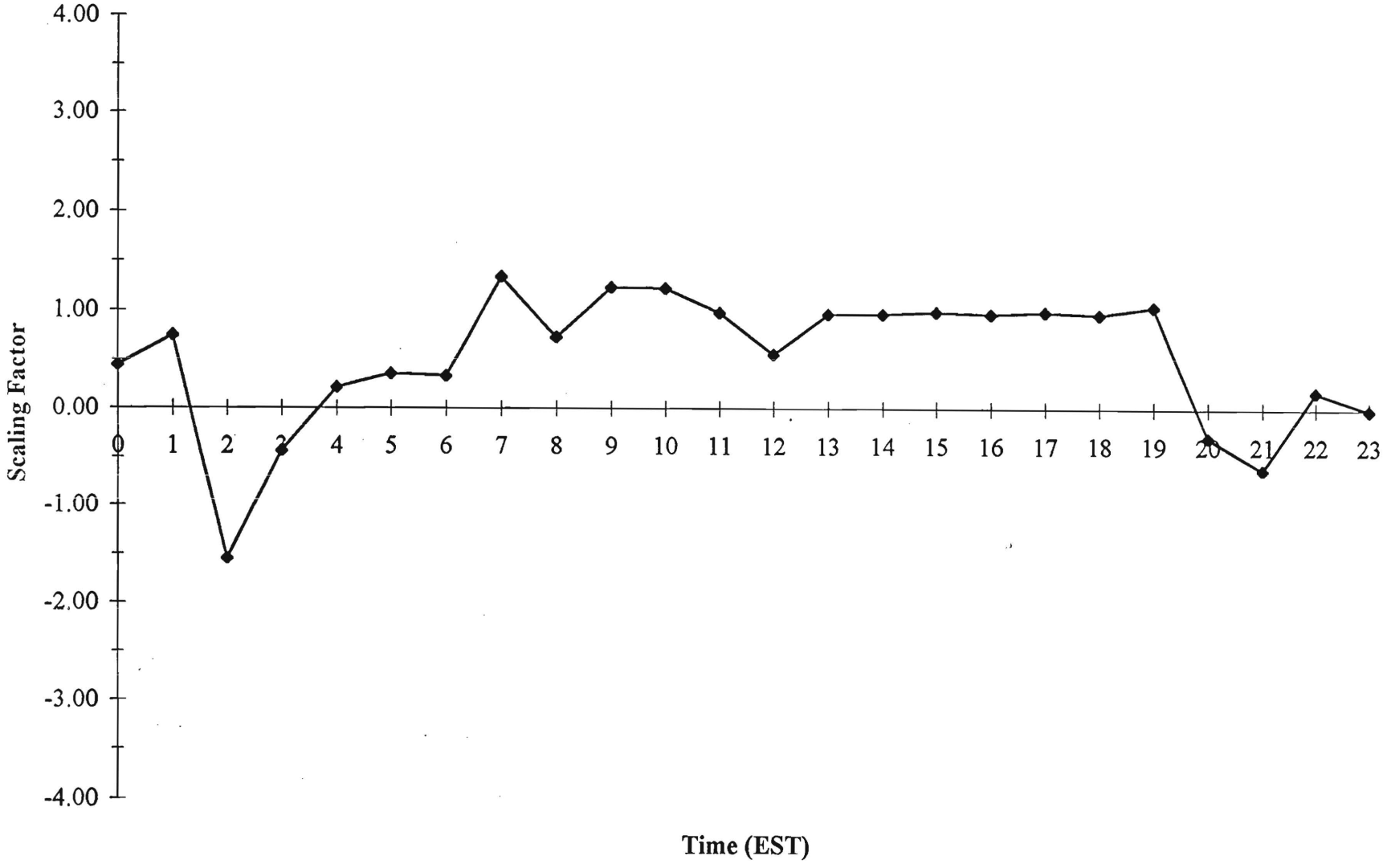

Figure 8. "Tuned" inverse derived CO emission scaling factors.

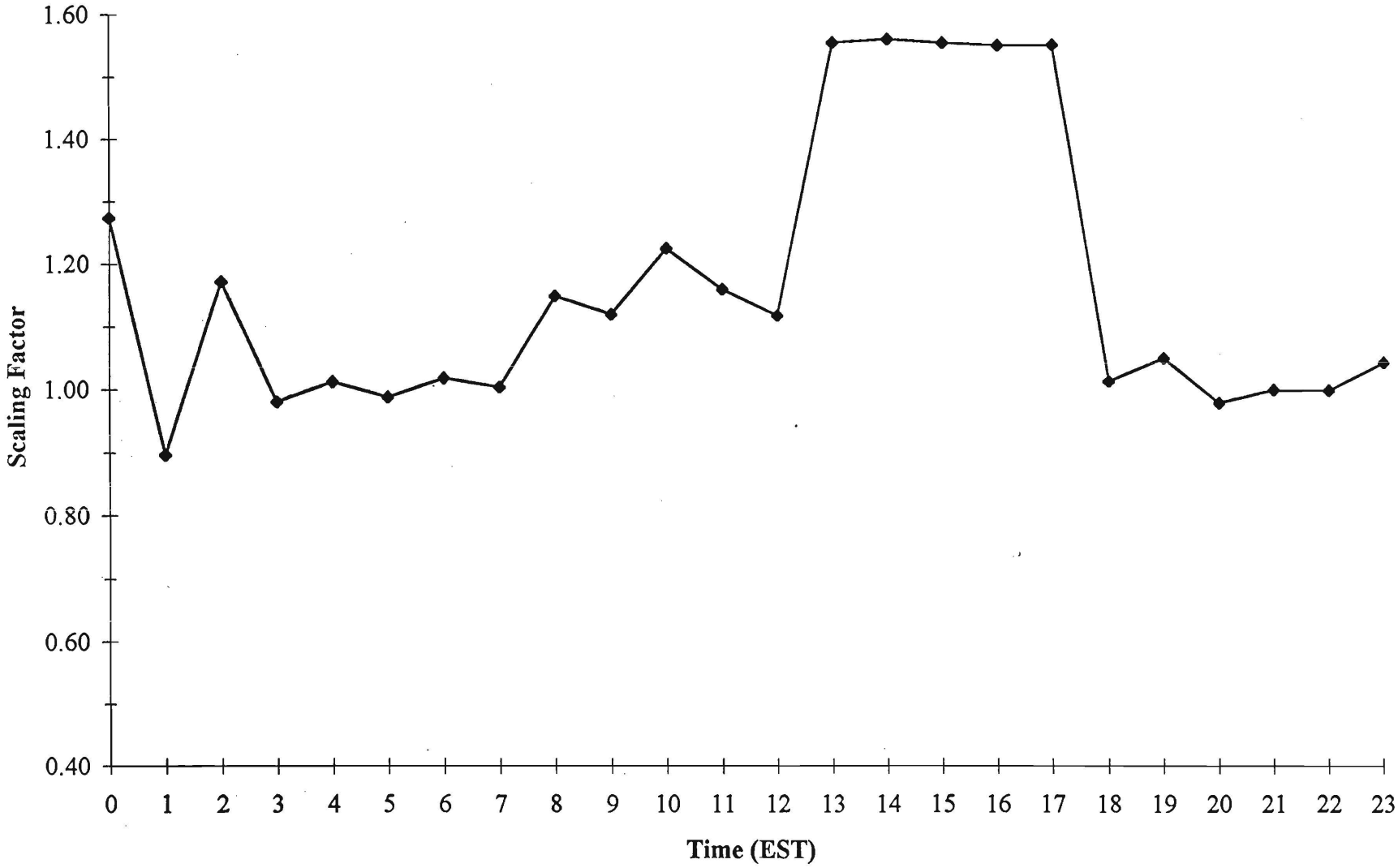

Figure 9. "Tuned" inverse derived CO emission scaling factors for pseudo-data (60% perturbation).

not resolve the emissions at all between hours 13 and 17. This would indicate that the damping effect was too severe. By over-restricting the amount by which the emissions were allowed to change, the emissions could not change sufficiently so that model predictions match the pseudo-observations. This must also be true for the real-data run. Thus, it is difficult to conclude that the encouraging results seen in the "tuned" CO inverse run represent the true CO emissions. Several more runs were made in an attempt to find a compromise between the need to dampen the oscillations and the diametric need to allow sufficient emissions change. However, damping the oscillations proved to be the more difficult to control; they returned as soon as the parameters were relaxed. No satisfactory solution was found.

We suspect that the difficulty in running a CO inverse lies in the spatial distribution of the CO emissions in the modeling inventory. Before we begin an inverse problem, we assume that all the emissions are approximately correctly spatially distributed. Spatial distributions may be determined either directly (e.g. emissions along a roadway) or by surrogate (e.g. CO emissions follow population patterns). Further, while we do not assume that the magnitude of the emissions are correct, we do assume that the relative spatial distribution is correct. For example, if it is known that two times more vehicles travel across the northern arc of I-285 than across the southern arc, then a good first assumption is that the mobile emissions across the northern arc are twice those across the southern arc. Given the spatial distribution, we then solve for the magnitude of the emissions at each hour, thus determining the temporal allocation. In obtaining this solution, we also assume that the meteorology (primarily wind speed, wind direction, and mixing height) is also approximately correct. This plan worked well for the biogenic isoprene emissions. It does not seem to work well for the anthropogenic CO emissions.

In Atlanta, sources of isoprene (all biogenic) are numerous, small, and ubiquitous. There are no highly concentrated isoprene sources. At any isoprene monitor , about the same concentration of isoprene will be measured regardless of the direction of the wind. On the other hand, some sources of CO emissions may indeed be extremely concentrated (e.g. along a roadway). Thus, a CO receptor could measure high concentrations of CO if it is near a source or the wind is blowing from the location of a source. The same receptor could also measure low concentrations if the wind is blowing from another direction. In the model, small errors in the spatial distribution of CO sources or in the speed and direction of the wind could cause large discrepancies between model predictions and observations. Since the biogenics are more homogeneously distributed, this was not a problem during the isoprene inverse.

Last year, we reported (Cardelino et al. 1995) a similar finding using a different approach. For the anthropogenic emissions, we compared observed ratios of HC to NOx and observed ratios of CO to NOx to the corresponding ratios derived from emission estimates. The results showed that the ambient ratios of HC/NOx and CO/NOx were higher than the emission ratios by 20% and 73% respectively. We also compared UAM predicted concentrations of HC, NOx, and CO to observed concentrations and found that the ambient concentrations were greater than the corresponding UAM concentrations by 14%, 40% and 58%, respectively. We concluded then that the variability shown by the different comparisons suggested that a possible source of the discrepancies was the spatial distribution of the anthropogenic emissions. Unfortunately with only 4 sites, we do not have enough CO observations to attack the spatial distribution problem with the inverse method.

3. Nitric Oxide (NO)

On August 10, 1992, there were seven (7) nitric oxide (NO) monitors operating in the Atlanta Area. Once again, the Georgia Tech site (TECH) was the only true urban site. The South Dekalb (SDEK), Fort McPherson (FORT), and Tucker (TUCK) sites may be considered as suburban. The rural sites were Yorkville (YORK), Buzzard Flapper (BFCK) and Clover Lake (CLOV). Figure 10 shows the observed concentrations and initial UAM predicted (prior to any adjustments) concentrations.

When observed concentrations are extremely small, the normalized bias (the statistic we have used to measure model performance relative to observations) can become extremely exaggerated:

$$\text{normalized bias} = \{\Sigma(\text{predicted} - \text{observed})/\text{observed}\}/N$$

where N is the number of prediction-observation pairs.

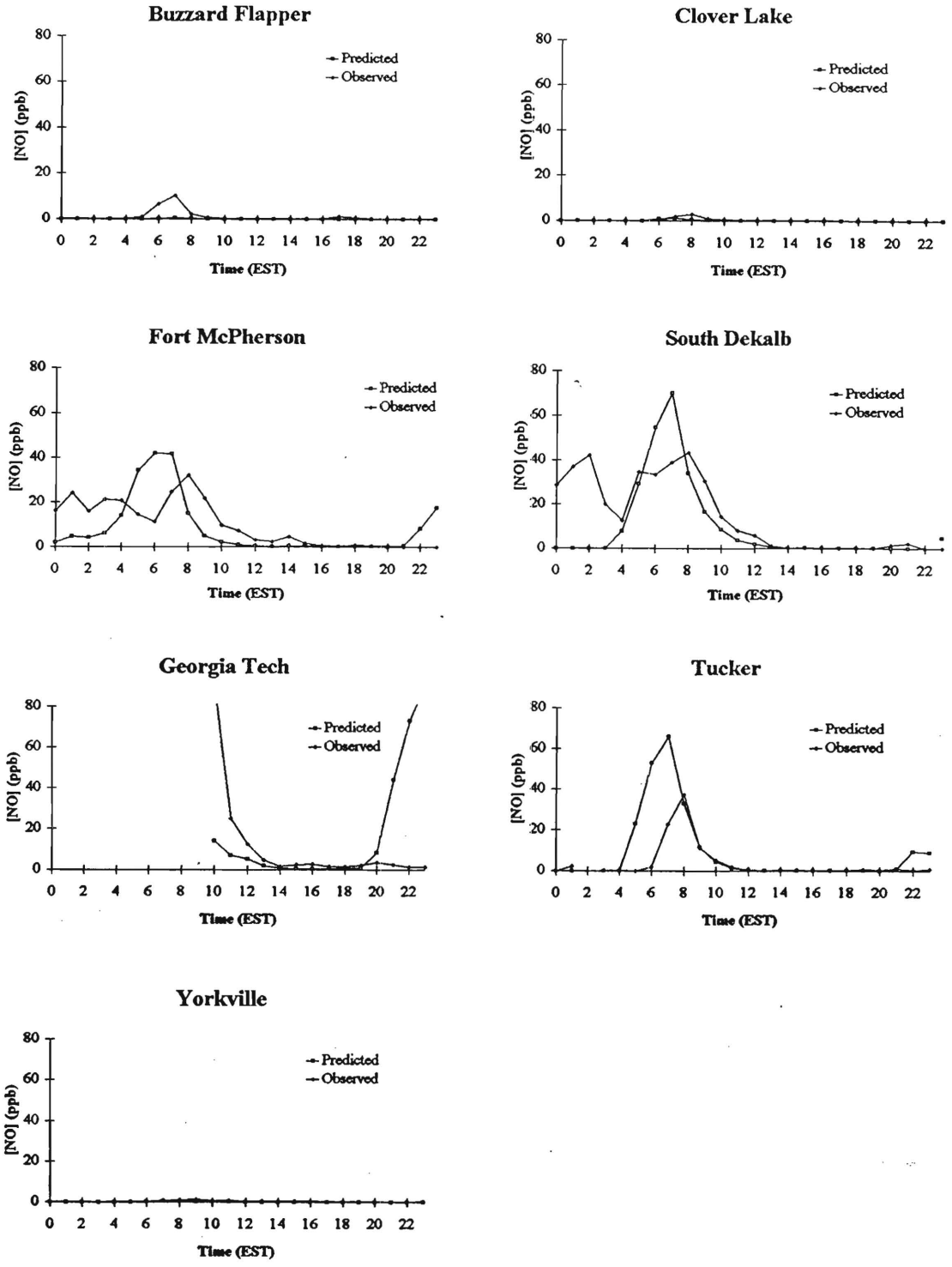

Figure 10. Observed and initial UAM predicted NO concentrations in Atlanta, Georgia (August 10, 1992).

In these cases, it is more appropriate to use the bias to assess model performance:

$$\text{bias} = \{\Sigma(\text{predicted} - \text{observed})\}/N$$

The latter is actually what the inverse method is attempting to minimize. Overall, before any adjustments, the UAM slightly overpredicts NO concentrations by 0.18 ppb. Performance is the worst at the TECH site; at this site, on average the model overpredicts NO by 6.91 ppb. There are also differences at the suburban sites. The model underpredicts by 1.42 ppb at the FORT site and by 5.24 ppb at the SDEK site. It overpredicts by 5.20 ppb at the TUCK location. The model predicts NO relatively well for the rural sites. The biases at BFCK, CLOV, and YORK are +1.04 ppb, -0.20 ppb, and -0.34 ppb respectively.

Since the model is already performing relatively well for most hours as evidenced by the small overall bias, it will be difficult to improve the simulation using the inverse method. In particular, the afternoon hours will be extremely difficult since concentrations of NO do not seem to be affected by emissions. Further, NO emissions also share some of the same spatial attributes (and thus some of the same problems) as CO emissions. Nevertheless, we did attempt to run an inverse for NO.

A. NO Inverse with Pseudo-data

All of the area, nonroad mobile, and mobile NO emissions sources were merged into one gridded and temporally allocated source. Although point sources are a significant source of NO in the Atlanta area, it was assumed that their emissions were well known in the 1992 inventory. This is a valid assumption since day specific data was used for 90% of the total point source NOx emissions (Chang et al. 1996). For the inverse, it must also be assumed that the UAM adequately simulates point source plumes. (This is a widely noted deficiency in version IV of the UAM, and thus another obstacle for the NO inverse. Plumes are more aptly treated in version V of the model with a plume-in-grid mechanism.) The single low-level source was initially set to 150% of the baseline. As with the CO run, the filter was then used to determine the base emissions using only the differences between the predicted NO concentrations (with initial emissions at 150% of the baseline) and the observed concentrations (pseudo-data - NO concentrations extracted from a baseline run). Figure 11 shows the pseudo-data observations and the initially perturbed concentrations at each of the 7 sites. Figure 12 shows the inverse derived factors necessary to force the model to match the pseudo-data. Using pseudo-data, we expected to get a near perfect solution (recall that a value of 1.00 implies a perfect convergence), yet the solution was less than perfect. This can indicate that the NO signal is not strong enough to overcome the difficulties associated with running an inverse for NO. From this test with an induced 50% error in the emissions, we decided to cautiously proceed with a real-data NO inverse run. Any results obtained must be viewed in the context of the uncertainties described above.

B. NO Inverse with Observations from the 1992 Atlanta Intensive

NO observations from the 1992 Atlanta Intensive (see Figure 10) were used to determine hourly factors needed to adjust the August 10, 1992 (day specific) NO emission inventory to minimize the differences between model predictions and observations. This run was terminated after hour 10 when it was noticed that an unusually large number of iterations were required before the solution would converge. Typically, the method will converge on a solution in less than 10 iterations. For hour 3, the method required 268 iterations. Nevertheless, this abbreviated run does provide some insight. Figure 13, when compared to figure 10 shows that the inverse was functioning properly - model predictions at each station tended to follow the observed profiles more closely than in the initial unadjusted run. For hours 0-10, the bias at most sites decreased. For BFCK, CLOV, FORT, SDEK, TECH, TUCK, and YORK the inverse adjusted bias statistics (and original unadjusted bias statistics) are -1.90 (-1.91), -0.30 (-.32), +1.61 (-3.75), -13.63 (-10.32), -53.69 (-76.03), +5.81 (+10.95), -0.32 (-0.33) respectively. Even at the South Dekalb site where the statistics indicate that the simulation was made worse by the adjustment, visual inspection suggests that the inverse simulation better represents the observations.

The inverse derived emission factors are shown in figure 14. Although the run was terminated prematurely, results from hours 7-10 indicate that the solution may again oscillate from positive to negative emissions similar to the CO inverse. When the run was terminated during hour 11, emissions were moving back towards a negative

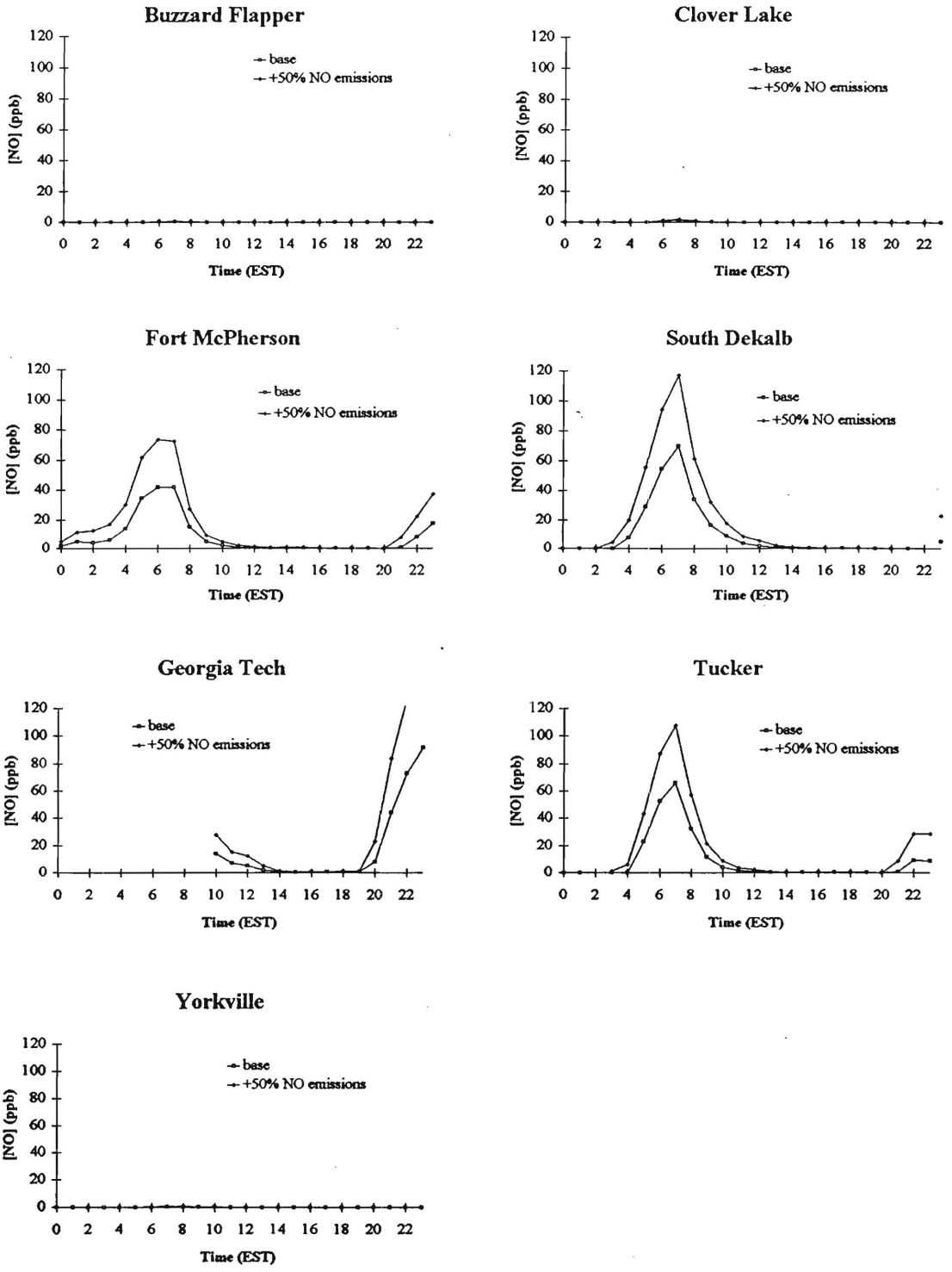

Figure 11. Base (pseudo-observations) and perturbed (+50% NO emissions) UAM predicted NO concentrations in Atlanta, Georgia (August 10, 1992).

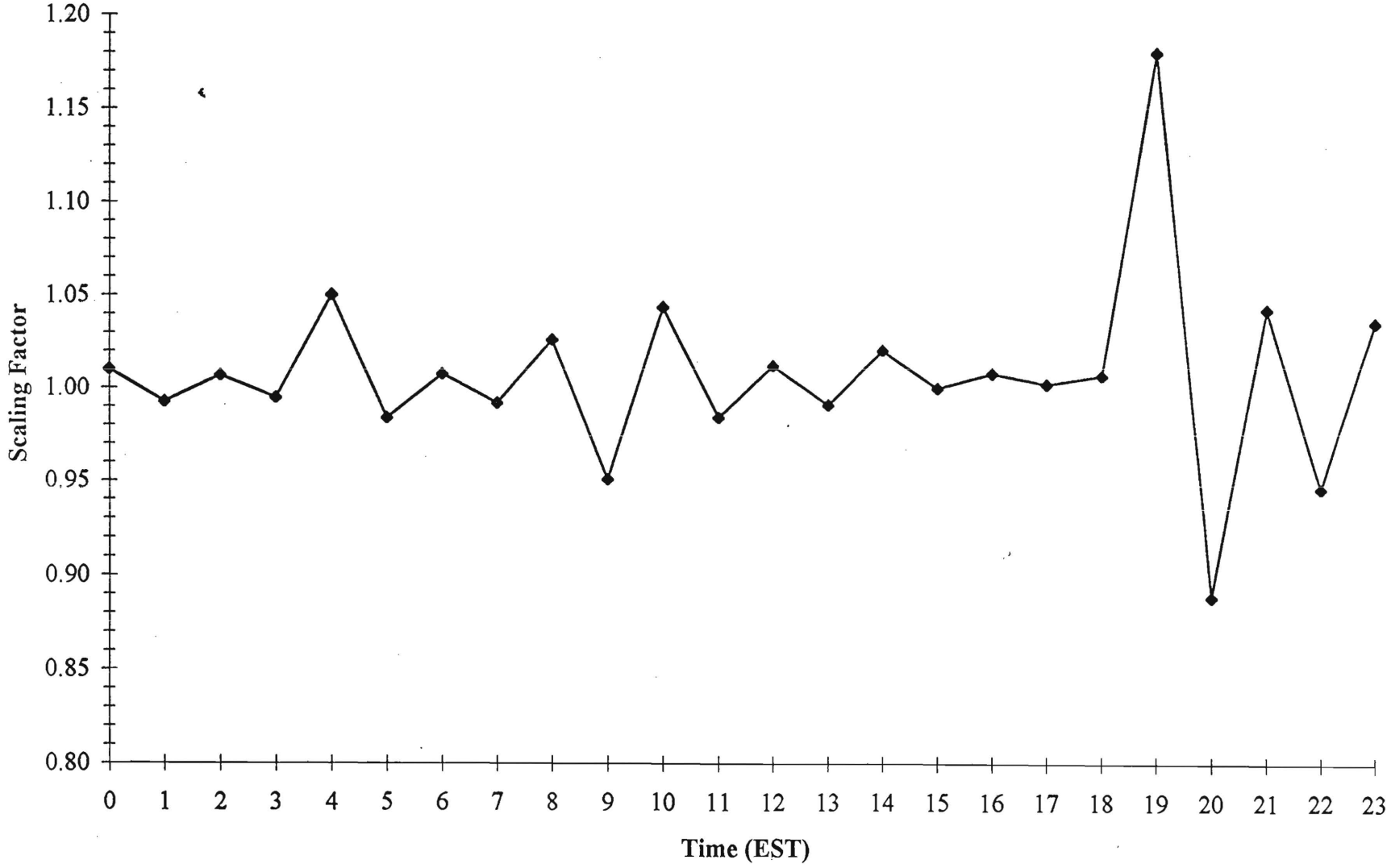

Figure 12. Inverse derived NO emission scaling factors for pseudo-data (50% perturbation).

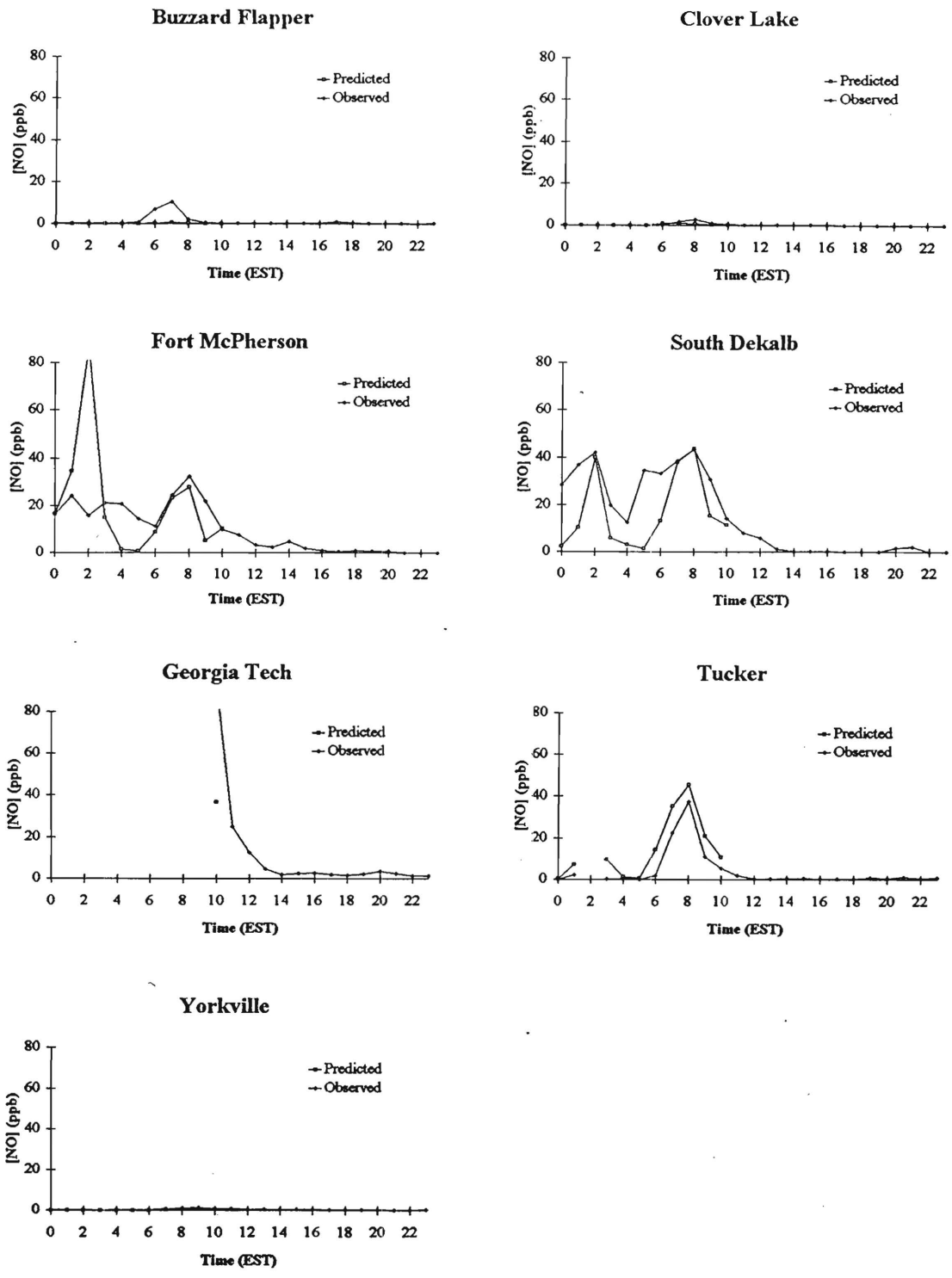

Figure 13. Observed and inverse adjusted - UAM predicted NO concentrations in Atlanta, Georgia (August 10, 1992).

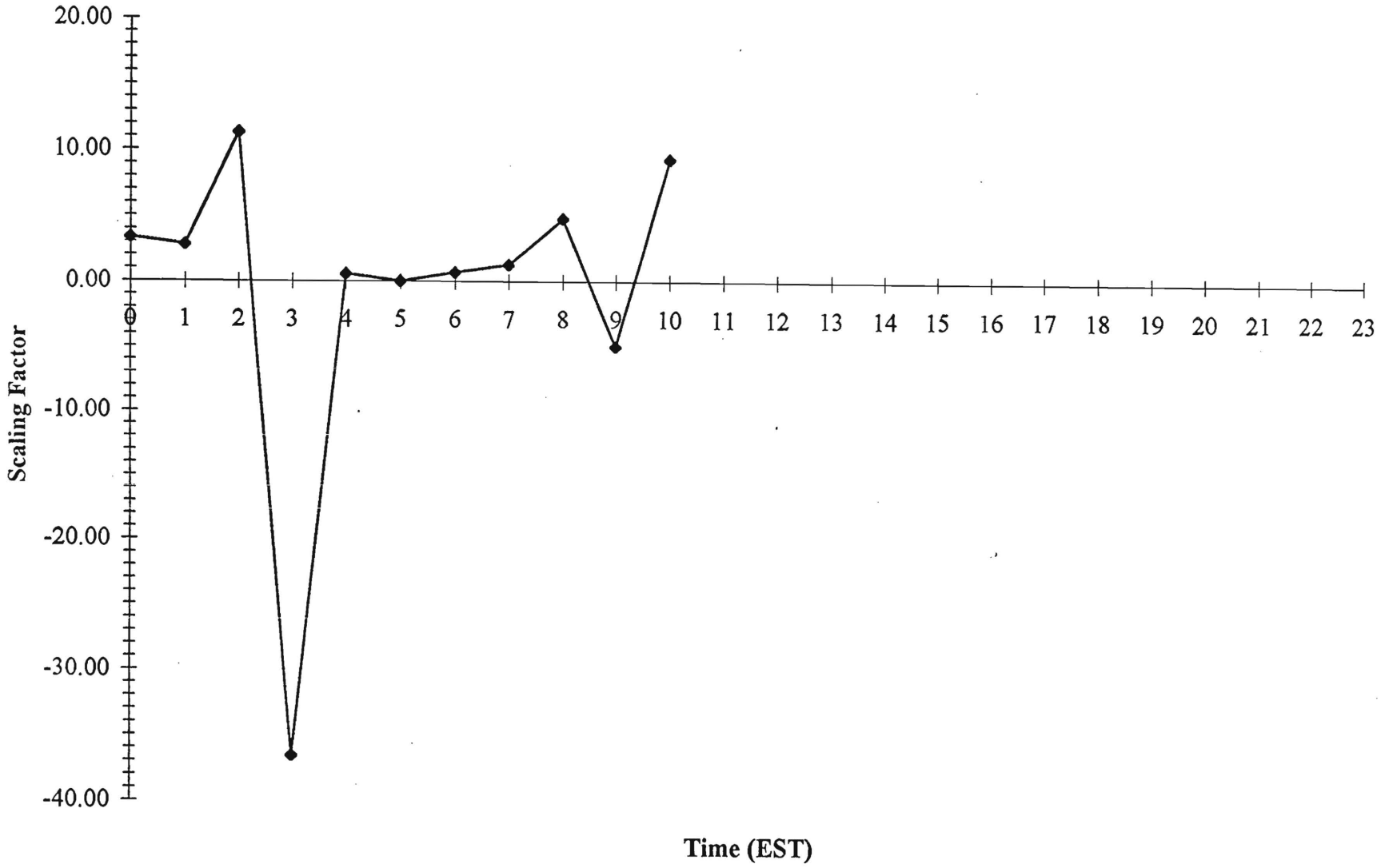

Figure 14. Inverse derived NO emission scaling factors.

solution (not shown) following the positive adjustment during hour 10. This should not be surprising given that the low-level NO emissions are dominated by the same sources (i.e. mobile) as the CO emissions. We would expect that the NO inverse would have many of the same problems as the CO inverse. Given this sign along with the other difficulties described previously, no further NO inverse runs were completed.

4. Volatile Organic Compounds (VOCs)

Relative to the inverse problem to estimate the emissions of anthropogenic hydrocarbons, the isoprene, carbon monoxide, and nitric oxide inverse problems should have been easy to solve. In each of the previous cases, the chemical specie studied was measured explicitly and modeled explicitly. Within Atlanta, the only significant source of isoprene is biogenic emission; the dominant sink is reaction with the hydroxy radical (OH). Unlike the other hydrocarbons, there is no in situ production due to the oxidation and reaction of other hydrocarbons. Thus, isoprene works well in the inverse method because there is very little interference from other mechanisms. Likewise, we did not expect to have much difficulty with carbon monoxide. On the time scale of 1 day, CO is virtually unreactive. Further, since the emission of CO is dominated by only one source (mobile), using the inverse to solve for the CO emissions should have been relatively straight forward. We did not anticipate the difficulties associated with the spatial allocation of the sources. We did anticipate that NO would provide additional problems. Although the majority of NOx emitted is in the form of NO, during the day it is quickly oxidized so that the majority of ambient NOx is in the form of NO_2. This property makes it difficult to measure the change in ambient NO concentrations due to a change in NO emissions. Still, since NOx chemistry is treated explicitly in the UAM, this should have been a minor obstacle.

Because anthropogenic hydrocarbons are neither explicitly measured nor explicitly modeled, substantial noise is introduced into the inverse method. Literally hundreds of hydrocarbons are present in the ambient air. Only 56 of these species were measured during the 1992 Atlanta Intensive. It is not clear if this small sample is representative of the total volatile organic compound loading that existed in Atlanta. Further, to facilitate comparison with model predictions, these observations must first be mapped into the Carbon Bond IV (CB-IV) chemical mechanism. This will further degrade the quality of the measurements by subjecting them to the vagaries of the chemical model. From the emissions side, estimates of total VOC must also undergo the CB-IV transformation. While none of these issues would have prevented us from conducting an inverse analysis of the anthropogenic VOCs, these together with the difficulties we discovered in trying to solve for the CO and NO emissions have led us to believe that it would be fruitless to attempt a VOC inverse. We believe that failure to find a satisfactory inverse solution for the CO and NO inventories is a sufficient reason to conclude that a suitable VOC inverse solution also does not exist.

5. Future Plans

Over the next several months, we will test other approaches to determine if we can minimize the problems associated with the spatial allocation of the anthropogenic emissions. Specifically, we will experiment with longer averaging times (e.g. 3 hours instead of 1 hour), and a time inclusive approach rather than a time iterative approach. This work will be submitted to a peer review journal and we will continue to keep the DNR informed of any progress.

References

Cardelino, C., D. Hartley, M.E. Chang, W.L. Chang, D. Haas-Laursen; Emissions and Urban Airshed Modeling Issues for Atlanta, Georgia - Final Report; School of Earth & Atm. Sci., Georgia Institute of Technology; June 12, 1995.

Chang, W.L., C. Cardelino, M.E. Chang; The Use of Survey Data to Investigate Ozone Sensitivity to Point Sources; in press *Atmospheric Environment*; 1996.

Geron,C.D., T.E. Pierce, A.B. Guenther; Reassessment of Biogenic Volatile Organic Compound Emissions in the Atlanta Area; *Atmospheric Environment* V29, p. 1569; 1995.

III. Appendix

Chang, M.E., D.E. Hartley, C. Cardelino, W.L. Chang; "Inverse Modeling of Biogenic Isoprene Emissions;" submitted to *Geopysical Research Letters*; June 15, 1996.