TSD-1e

# CMAQ Model Performance and Assessment

# 8-h OTC Ozone Modeling

**Bureau of Air Quality Analysis and Research**

**Division of Air Resources**

**New York State Department of Environmental Conservation**

**Albany, NY 12233**

**March 19, 2006**

## Air quality model evaluation and assessment

One of the tasks that is required as part of demonstrating attainment for the 8-hr ozone NAAQS is the evaluation and assessment of the air quality modeling system that has been utilized to predict future air quality over the region of interest. As part of the attainment demonstration, the SMOKE/CMAQ modeling system was applied to simulate the pollutant concentration fields for the base year 2002 emissions with the corresponding meteorological information. The modeling databases for meteorology using MM5 (TSD-1a), the emissions using SMOKE (TSD-1b and TSD-1c), and application of CMAQ (TSD-1d) provides simulated pollutant fields that are compared to measurements, in order to establish the credibility of the simulation. In the following sections a comparison between the measured and predicted concentrations is performed and results are presented, demonstrating on an overall basis the utility of the modeling system in this application.

The results presented here should serve as an illustration of some of the evaluation and assessment performed on the base 2002 CMAQ simulation. Additional information can be made available by request from the New York State Department of Environmental Conservation.

## Summary of measured data

The ambient air quality data, both gaseous and aerosol species, for the simulation period of May through September 2002 were obtained from the following sources:

- EPA Air Quality System (AQS)
- EPA fine particulate Speciation Trends Network (STN)
- EPA Clean Air Status & Trends Network (CASTNet)
- Interagency Monitoring of PROtected Visual Environments (IMPROVE)
- Pinnacle State Park, NY operated by Atmospheric Science Research Center, University at Albany, Albany, NY
- Harvard Forest, Petersham, MA operated by Harvard University, Boston, MA
- Atmospheric Investigation, Regional Modeling, Analysis and Prediction (AIRMAP) operated by University of New Hampshire, Durham, NH
- NorthEast Ozone & Fine Particle Study (NE-OPS), led by Penn State University and other research groups in Philadelphia, PA
- Aircraft data obtained by the University of Maryland, College Park MD
- Wet deposition data from the National Atmospheric Deposition Program/National Trends Network (NADP/NTN), Atmospheric Integrated Research Monitoring Network (AIRMoN), and the New York State Department of Environmental Conservation (NYSDEC)

Measured data from sites within the Ozone Transport Region (OTR) plus the rest of Virginia were included here. The model-based data were obtained at the grid-cell corresponding to the monitor location; no interpolation was performed.

Ozone ($O_3$)

Hourly $O_3$ is measured at a large number of State, Local, and National Air Monitoring Stations (SLAMS/NAMS) across the US on a routine basis, and the data from 208 sites were extracted from the AQS database (http://www.epa.gov/ttn/airs/airsaqs/aqsweb/aqswebhome.html). Hourly $O_3$ concentrations from the Harvard Forest Environmental Management Site in Petersham, MA (http://www.as.harvard.edu/data/nigec-data.html); Pinnacle State Park in Addison, NY (http://www.asrc.cestm.albany.edu); and the four University of New Hampshire AIRMAP sites (http://airmap.unh.edu) were also included in this database. The EPA CASTNet program collects hourly $O_3$ at generally rural locations across the US (http://www.epa.gov/castnet); data from 22 sites, including two from West Virginia, were used in the model evaluation.

Fine particulate matter ($PM_{2.5}$)

The 24-hour average Federal Reference Method (FRM) $PM_{2.5}$ mass data collected routinely at SLAMS/NAMS sites across the US were extracted from AQS (257 sites). Hourly $PM_{2.5}$ mass was also included in this database, primarily extracted from AQS (54 sites). Hourly $PM_{2.5}$ mass were also taken from the Thompson Farm, NH AIRMAP site, Pinnacle State Park, and the NE-OPS site in Philadelphia, PA (http://lidar1.ee.psu.edu).

Fine particulate speciation

The 24-hour average $PM_{2.5}$ and fine particulate speciation (sulfate ($SO_4$), nitrate ($NO_3$), elemental carbon (EC), organic carbon/organic mass (OC/OM), and soil/crustal matter) from Class I areas across the US, collected every 3$^{rd}$ day, were obtained from the IMPROVE web site (http://vista.cira.colostate.edu/IMPROVE/Default.htm). In addition to these parameters, the EPA STN (http://www.epa.gov/ttn/amtic/speciepg.html) also reports ammonium ($NH_4$) to AQS; data from this network are collected every 3$^{rd}$ or 6$^{th}$ day. Data from 49 STN sites, generally in urban areas and often collocated with FRM monitors, and 21 IMPROVE sites (including Dolly Sods, WV) were used in this analysis. Organic mass is assumed to equal 1.8×OC, and soil/crustal matter is assumed to consist of oxides of Al, Ca, Fe, Si, and Ti. The STN OC data are blank-corrected by removing a monitor-specific, constant blank, and these values are available from http://www.epa.gov/airtrends/aqtrnd03/pdfs/2_chemspec0fpm25.pdf; the IMPROVE OC blanks are assumed to equal zero.

Criteria gaseous pollutants

Hourly carbon monoxide (CO; 97 sites), nitric oxide (NO; 75 sites), nitrogen dioxide ($NO_2$; 97 sites) and sulfur dioxide ($SO_2$; 134 sites) are also included in this model evaluation database. A large majority of these sites are SLAMS/NAMS monitors located primarily in urban in suburban areas, but data from the Harvard Forest, Pinnacle State Park, and AIRMAP sites are also included here.

Non-methane hydrocarbons

While there are several dozen hydrocarbon species measured routinely, for this model evaluation database the focus was on Carbon Bond IV species groups that consist of a single primary species. For this reason only ethene ($C_2H_4$), isoprene ($C_5H_8$), and formaldehyde (HCHO) concentrations were extracted from AQS. Hourly $C_2H_4$ and $C_5H_8$ data from 19 Photochemical Assessment Monitoring Stations (PAMS) sites and 24-hour average HCHO from 18 air toxics sites are included in this database.

University of Maryland aircraft data

The University of Maryland performed 144 aircraft spirals at 41 regional airport locations over 26 days from May-August 2002 (http://www.atmos.umd.edu/~RAMMPP). Spirals are approximately 20-45 minutes in duration, over which time the atmosphere from about 0-3 km is sampled. The concentrations of $O_3$, CO, and $SO_2$ from these spirals were included in this database, and help provide a semi-quantitative evaluation of CMAQ performance above the ground surface. Minute average aircraft data were compared to the nearest instantaneous 3-dimensional CMAQ output.

Wet deposition

The NADP (http://nadp.sws.uiuc.edu) collects wet deposition samples across the US, through the NTN and the AIRMoN. Weekly wet deposition samples are collected by the NTN, while daily or event-based samples were collected by the AIRMoN. The NYSDEC (http://www.dec.state.ny.us) also collects weekly wet deposition samples independently from the NADP. The wet deposition of $SO_4^{2-}$, $NO_3^-$, and $NH_4^+$ from 43 NADP/NTN sites, 7 NADP/AIRMoN sites, and 19 NYSDEC sites are included in this model evaluation database.

## Evaluation of CMAQ predictions

The following sections provide model evaluation information for the above referenced pollutants over the OTR portion of the 12-km modeling domain. The statistical formulations that have been computed for each species are as follows: $P_i$ and $O_i$ are the individual (daily maximum 8-hour $O_3$ or daily average for the other species) predicted and observed concentrations, respectively; $\overline{P}$ and $\overline{O}$ are the average concentrations, respectively, and N is the sample size.

Observed average, in ppb:

$$\overline{O} = \frac{1}{N}\sum O_i$$

Predicted average, in ppb (only use $P_i$ when $O_i$ is valid):

$$\overline{P} = \frac{1}{N}\sum P_i$$

Correlation coefficient, $R^2$:

$$R^2 = \frac{\left[\sum (P_i - \overline{P})(O_i - \overline{O})\right]^2}{\sum (P_i - \overline{P})^2 \sum (O_i - \overline{O})^2}$$

Normalized mean error (NME), in %:

$$NME = \frac{\sum |P_i - O_i|}{\sum O_i} \times 100\%$$

Root mean square error (RMSE), in ppb:

$$RMSE = \left[\frac{1}{N}\sum (P_i - O_i)^2\right]^{1/2}$$

Fractional error (FE), in %:

$$FE = \frac{2}{N}\sum \left|\frac{P_i - O_i}{P_i + O_i}\right| \times 100\%$$

Mean absolute gross error (MAGE), in ppb:

$$MAGE = \frac{1}{N}\sum \left|P_i - O_i\right|$$

Mean normalized gross error (MNGE), in %:

$$MNGE = \frac{1}{N}\sum \frac{\left|P_i - O_i\right|}{O_i} \times 100\%$$

Mean bias (MB), in ppb:

$$MB = \frac{1}{N}\sum (P_i - O_i)$$

Mean normalized bias (MNB), in %:

$$MNB = \frac{1}{N}\sum \frac{(P_i - O_i)}{O_i} \times 100\%$$

Mean fractionalized bias (MFB), in %:

$$MFB = \frac{2}{N}\sum \left[\frac{P_i - O_i}{P_i + O_i}\right] \times 100\%$$

Normalized mean bias (NMB), in %:

$$NMB = \frac{\sum (P_i - O_i)}{\sum O_i} \times 100\%$$

Daily maximum 8-hour $O_3$ concentrations

Model evaluation statistics, based on daily maximum 8-hour average $O_3$ levels on those days having (1) at least 18 valid observations, or (2) fewer than 18 valid observations but the observed daily maximum $O_3$ concentration was at least 85 ppb, are presented here for all sites across the OTR and all of VA. The data covered the period May 15 through September 29, excluding July 6-9, when many sites across the eastern US were affected by large forest fires in Quebec. There are 208 SLAMS/NAMS sites and 28 special sites.

These model evaluation statistics were computed using two different threshold values for observed daily maximum 8-hour $O_3$. First, the statistics were computed using only those days when the observed daily maximum 8-hour $O_3$ concentration exceeded 40 ppb. Second, the statistics were computed using only those days when the observed daily maximum 8-hour $O_3$ exceeded 60 ppb. This latter method focuses on the highest $O_3$ days.

Figures 1-4 display time series of observed and predicted daily maximum 8-hour $O_3$ concentrations averaged over all sites across the OTR, at SLAMS/NAMS and special sites and for the daily maximum two thresholds. These averages were computed for each day considering all sites that met the corresponding threshold criteria. In general the observed and predicted composite average $O_3$ concentrations track each other rather well, although there was fairly substantial underprediction during the mid-August period. Also, the model performance tends to be better when the lower cutoff (40 ppb) was considered.

Figures 5-8 display spatial maps of fractional error and mean fractionalized bias for the two threshold levels. At each site the statistics were computed over the entire modeling season. Both the SLAMS/NAMS and special monitors are displayed here. In general, the model performance was better in the vicinity of urban areas and along the northeastern corridor, compared to the performance in rural areas where the model tended to underpredict daily maximum concentrations. The other statistical metrics yielded similar results to FE and MFB.

Table 1 lists the median and range in fractional error, and the mean fractionalized bias of daily maximum 8-hour $O_3$ calculated at each site over the season, for both observed thresholds (40 and 60 ppb), as well as all sites versus just the SLAMS/NAMS sites. Considering just SLAMS/NAMS sites, FE was always less than 32% for the 40 ppb threshold, and less than 40% for the 60 ppb threshold. Similarly, the MFB at SLAMS/NAMS sites ranged from -29 to +23% for the 40 ppb threshold, and ranged from -40 to +22% for the 60 ppb threshold. Adding the special sites did not affect the statistics substantially.

Diurnal variations of gases

Figures 9-17 display the composite diurnal variations of the species reported hourly – $O_3$ (SLAMS/NAMS and other/special sites, displayed separately), continuous $PM_{2.5}$, CO, NO, $NO_2$, $SO_2$, ethene, and isoprene. The average diurnal variations are for the period of May 15-September 30 – again excluding July 6-9 – considering all sites in the OTR. Note that the $O_3$ diurnal variations were computed from running 8-hour averages, with hours denoting the start of the 8-hour block. The number of monitors used to compute each composite diurnal variation is shown in each figure.

For $O_3$, the composite diurnal pattern predicted by CMAQ is fairly similar to that observed, especially at the more urban SLAMS/NAMS monitors. However, on average CMAQ predicts the daily maximum about an hour earlier than observed. For most of the other species presented here, CMAQ tends to predict two daily peaks, one morning and one late afternoon. For some species, such as $PM_{2.5}$ mass the observed concentration on a composite basis has very little diurnal variation. On the other hand, primary pollutants like CO, NO, and ethane, CMAQ exhibits qualitative agreement with the observations.

Daily average concentrations of co-pollutant trace gases

Composite daily average predicted and observed concentrations of CO, NO, $NO_2$, $SO_2$, $C_2H_4$, HCHO, and $C_5H_8$ across the OTR are displayed in Figures 18-24. Daily average concentrations of the criteria gases, $C_2H_4$ and $C_5H_8$ were computed from hourly averages, and only those days having at least 12 hours of valid observed data were considered here. The HCHO data shown here are based on 24-hour average values every $6^{th}$ day. The criteria gas data cover the period May 15 – September 30, whereas the NMHC data only cover the June 1 – August 31 period, since these data are predominantly PAMS data; however, excluded from this analysis is the July 6-9 period when many sites across the eastern US were affected by large forest fires in Quebec.

Table 2 lists the median and range in mean fractionalized bias calculated at each site over the season used in this analysis. The values listed in Table 2 were computed at each site over the entire season. While the range in MFB is rather large for each species across all sites, the median MFB was below 50% for all species except $C_2H_4$, which is substantially overpredicted by CMAQ. It should be noted that these species can vary substantially from day to day, and days with very low modeled or observed values can contribute to high MFB.

$PM_{2.5}$ mass and speciation

Composite daily average predicted and observed concentrations of $PM_{2.5}$ mass (both daily average FRM data and continuous data), as well as major speciation –$SO_4$, $NO_3$, $NH_4$, EC, OM (defined here operationally as 1.8×blank-corrected organic carbon), and crustal mass (sum of oxides of Al, Ca, Fe, Si, and Ti) – across the OTR were compared in this analysis. The data cover the period May 15 – September 30, and again the July 6-9 period was excluded, when numerous sites in the eastern US were affected

by large forest fires in Quebec. The continuous and FRM $PM_{2.5}$ data are shown every day, since there are ample daily FRM sites across the OTR. The speciation data included here are daily averages every third day, and consist of the largely urban EPA STN and the largely rural IMPROVE network. The two speciation networks collect $PM_{2.5}$, $SO_4$, $NO_3$, EC, OM, and crustal mass, while only the STN reports $NH_4$ at a sufficient number of locations.

Table 3 lists the median and range in mean fractionalized bias calculated at each site over the season used in this analysis. The values listed in Table 3 were computed at each site over the entire season. Figures 25-39 display time series of composite average observed and predicted daily concentrations; in these figures, for each day the statistics were computed using all monitors with valid data. The best qualitative agreement between observed and modeled concentrations is exhibited for $PM_{2.5}$ and $SO_4$. Note that in the case of crustal mass, the data from July 4 are also not included since this day is greatly affected by fireworks. On July 4, the composite average observed and predicted crustal concentrations were 4.59 $\mu g\ m^{-3}$ and 1.74 $\mu g\ m^{-3}$, respectively at the STN monitors, and 4.46 $\mu g\ m^{-3}$ and 0.99 $\mu g\ m^{-3}$, respectively at the IMPROVE monitors.

As with the gaseous co-pollutant data, there is a substantial spread in MFB across the sites. However, the median MFB for $PM_{2.5}$ mass and $SO_4$ was generally small (<12%) for both urban and rural sites. CMAQ tends to overpredict $NO_3$, more so at the IMPROVE sites. CMAQ also tends to underpredict OM at both urban and rural sites, although some of this discrepancy may be attributed to the fact that OM is operationally defined and is highly dependent on the blank correction and multiplier to account for other components of OM not directly measured. CMAQ tends to overpredict both EC and crustal mass, especially at urban sites; similar to OM, the crustal mass overprediction is related to the fact that this parameter is operationally defined.

Wet deposition of sulfate, nitrate, and ammonium

Observed and predicted wet deposition of $SO_4$, $NO_3$, and $NH_4$ were compared over the period May 14 – September 30. For this analysis, weekly or event-based wet deposition amounts from the NADP/NTN (43 sites), NADP/AIRMoN (7 sites), and New York State DEC (19 sites) covering the entire OTR plus all of VA and WV were integrated over the four-and-a-half months. Because the observed weekly wet deposition samples did include July 6-9, the corresponding CMAQ predictions also include this period. Table 4 lists the model evaluation statistics for integrated wet deposition of $SO_4$, $NO_3$, and $NH_4$ at each site over the season, while Figures 40-42 compare the observed and predicted weekly values relative to the 1:1 line.

Overall CMAQ tended to overpredict wet deposition of these ions. On a percentage basis, the overprediction was least for $SO_4$ and highest for $NO_3$. The NME, MNGE, MNB, and NMB were less than 50% for the three ions. Given that precipitation is very difficult to predict, especially during the summer months when rainfall can vary tremendously over a 12 km by 12 km area represented by this model grid, CMAQ did a rather good job reproducing seasonal wet deposition over the OTR.

Upper-air $O_3$, CO, and $SO_2$ data

The University of Maryland operated an instrumented light aircraft during the summer of 2002. On 26 days from May-August meteorological, trace gas, and particle scattering/absorption data were collected during ascent or descent spirals over 41 regional airports. In all, 144 spirals were performed from near the surface to about 3 km above ground level. For this analysis, composite average profiles of $O_3$, CO, and $SO_2$ were created over three time periods: "morning" (08-11 EST), "afternoon" (12-16 EST), and "evening" (17-19 EST). The minute average observed concentrations were aggregated into layer averages, which correspond to the lowest 15 model layers. Model layers are increasingly thick away from the surface; the surface layer is about 20 m thick while the $15^{th}$ layer is about 500 m thick (and centered about 2.8 km above the ground). Figures 43-51 display the observed and predicted composite vertical profiles of $O_3$, CO, and $SO_2$ for the three time periods. In terms of profile shape, CMAQ was in good qualitative agreement for all three species above the surface during the afternoon hours. For CO, the model tends to greatly underpredict observed levels near the surface, whereas the predicted $O_3$ and $SO_2$ concentrations are closer to the respective observed values.

## Summary

Various model evaluation statistics are presented here for a variety of gaseous and aerosol species in addition to $O_3$. In general, the CMAQ results were best for daily maximum $O_3$ and daily average $PM_{2.5}$ and $SO_4$ mass. Many other species vary tremendously over the course of a day, or from day to day, and small model over- or underprediction at low concentrations can lead to large biases on a composite basis. It is important to demonstrate that the model performs reasonably over the diurnal cycle, not just in terms of daily maximum or average values. Also, it is important to demonstrate that the model can reproduce concentrations above the ground level.

Table 1. Median and range in fractional error (FE, %) and mean fractionalized bias (MFB, %) for daily maximum 8-hour $O_3$ using the 40 ppb and 60 ppb observed thresholds. The values using only SLAMS/NAMS sites are boldfaced, the values using all sites are in regular font.

| Metric, threshold | Range (%) | Median (%) |
|---|---|---|
| FE, 40 ppb | +10 to +34%<br>**+10 to +32%** | +15%<br>**+15%** |
| MFB, 40 ppb | -34 to +23%<br>**-29 to +23%** | -6%<br>**-6%** |
| FE, 60 ppb | +9 to +40%<br>**+9 to +40%** | +15%<br>**+15%** |
| MFB, 60 ppb | -40 to +22%<br>**-40 to +22%** | -12%<br>**-11%** |

Table 2. Median and range in mean fractionalized bias (%) for daily average CO, NO, $NO_2$, $SO_2$, $C_2H_4$, HCHO, and $C_5H8$.

| Pollutant | Range in MFB (%) | Median MFB (%) |
|---|---|---|
| CO (97 sites) | -128 to +144% | -10% |
| NO (75 sites) | -182 to +116% | -46% |
| $NO_2$ (97 sites) | -125 to +107% | +13% |
| $SO_2$ (134 sites) | -139 to 140% | +3% |
| $C_2H_4$ (19 sites) | +28 to +168% | +86% |
| HCHO (18 sites) | -66 to +96% | -13% |
| $C_5H_8$ (19 sites) | -54 to +165% | +43% |

Table 3. Median and range in mean fractionalized bias (%) for daily average $PM_{2.5}$, $SO_4$, $NO_3$, $NH_4$, EC, and OM.

| Pollutant | Range in MFB (%) | Median MFB (%) |
|---|---|---|
| $PM_{2.5}$ (FRM; 257 sites) | -59 to +119% | -4% |
| $PM_{2.5}$ (continuous; 57 sites) | -39 to +85% | +5% |
| STN $PM_{2.5}$ (49 sites) | -45 to +102% | -9% |
| IMPROVE $PM_{2.5}$ (21 sites) | -36 to +19% | -10% |
| STN $SO_4$ (49 sites) | -21 to +60% | +12% |
| IMPROVE $SO_4$ (21 sites) | -26 to +16% | -7% |
| STN $NO_3$ (49 sites) | -73 to +406% | +25% |
| IMPROVE $NO_3$ (21 sites) | -57 to +358% | +64% |
| STN $NH_4$ (49 sites) | -36 to +112% | +16% |
| STN EC (49 sites) | -42 to +269% | +34% |
| IMPROVE EC (21 sites) | -60 to +146% | -27% |
| STN OM (49 sites) | -82 to -25% | -58% |
| IMPROVE OM (21 sites) | -60 to +7% | -40% |
| STN crustal (49 sites) | +2 to +546% | +182% |
| IMPROVE crustal (21 sites) | -18 to +163% | +38% |

Table 4. Model evaluation statistics for integrated wet deposition of $SO_4$, $NO_3$, and $NH_4$

.

| Parameter | $SO_4$ | $NO_3$ | $NH_4$ |
|---|---|---|---|
| Observed average, mg $m^{-2}$ | 1063 | 704 | 185 |
| Predicted average, mg $m^{-2}$ | 946 | 367 | 117 |
| Correlation coefficient, $R^2$ | 0.17 | 0.22 | 0.12 |
| NME, % | 34 | 49 | 48 |
| RMSE, mg $m^{-2}$ | 490 | 417 | 109 |
| FE, % | 36 | 62 | 57 |
| MAGE, mg $m^{-2}$ | 365 | 344 | 89 |
| MNGE, % | 36 | 45 | 46 |
| MB, mg $m^{-2}$ | -118 | -337 | -68 |
| MNB, % | -3 | -44 | -28 |
| MFB, % | -13 | -61 | -44 |
| NMB, % | -11 | -48 | -37 |

Figure 1.

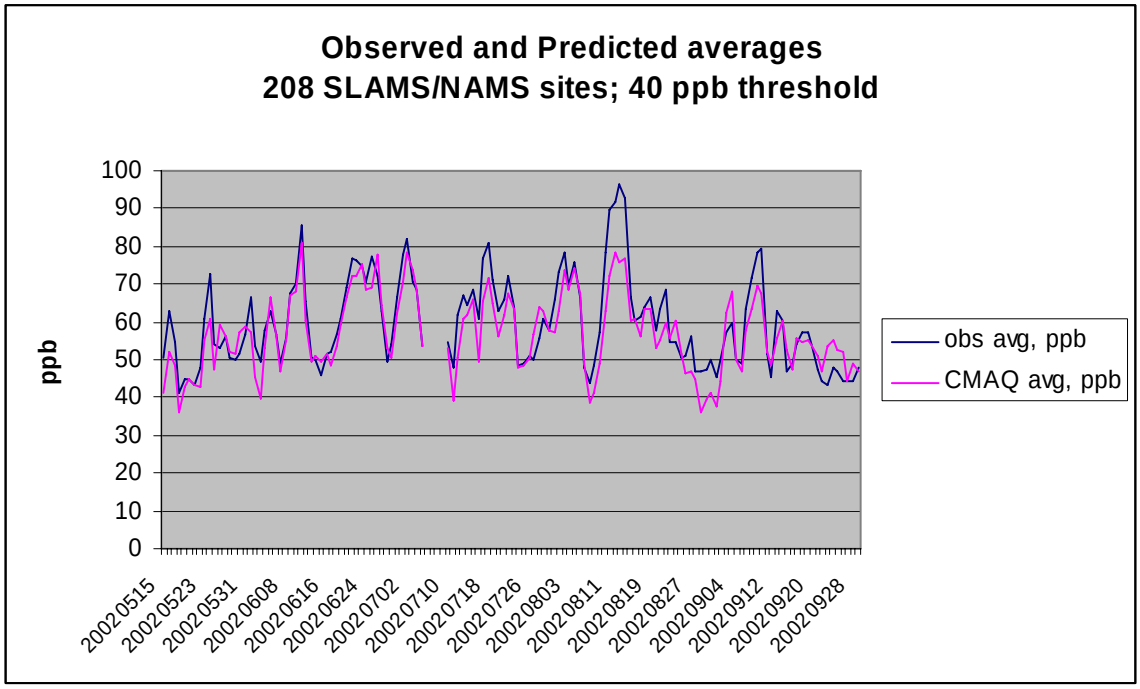



Figure 2.

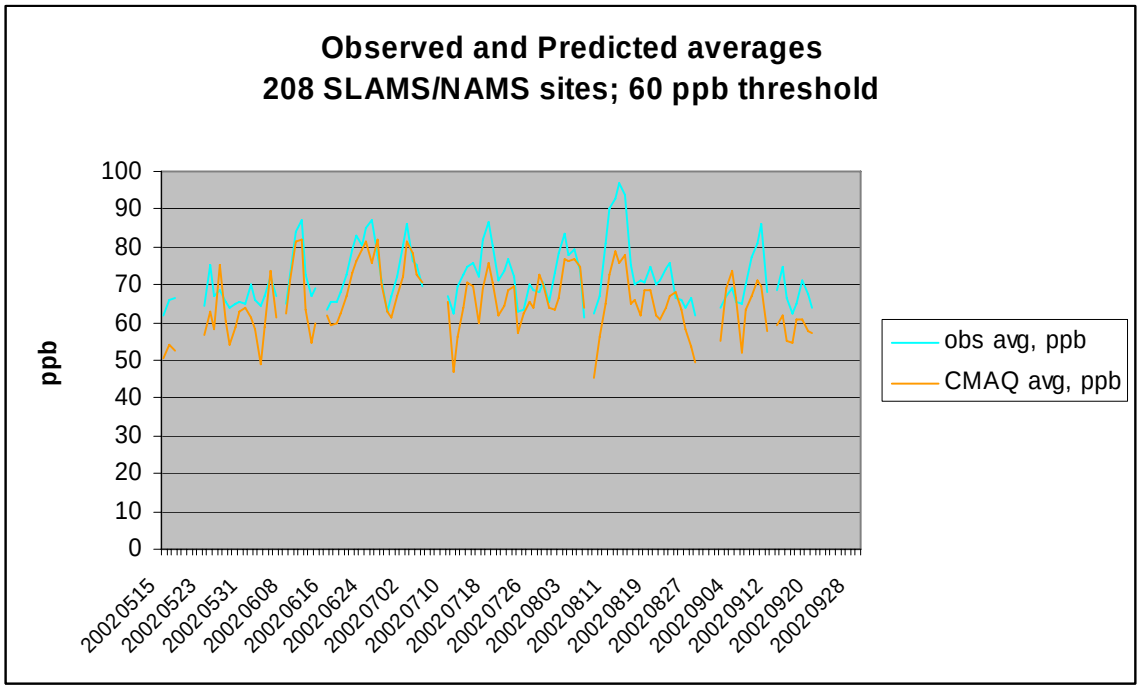


Figure 3.

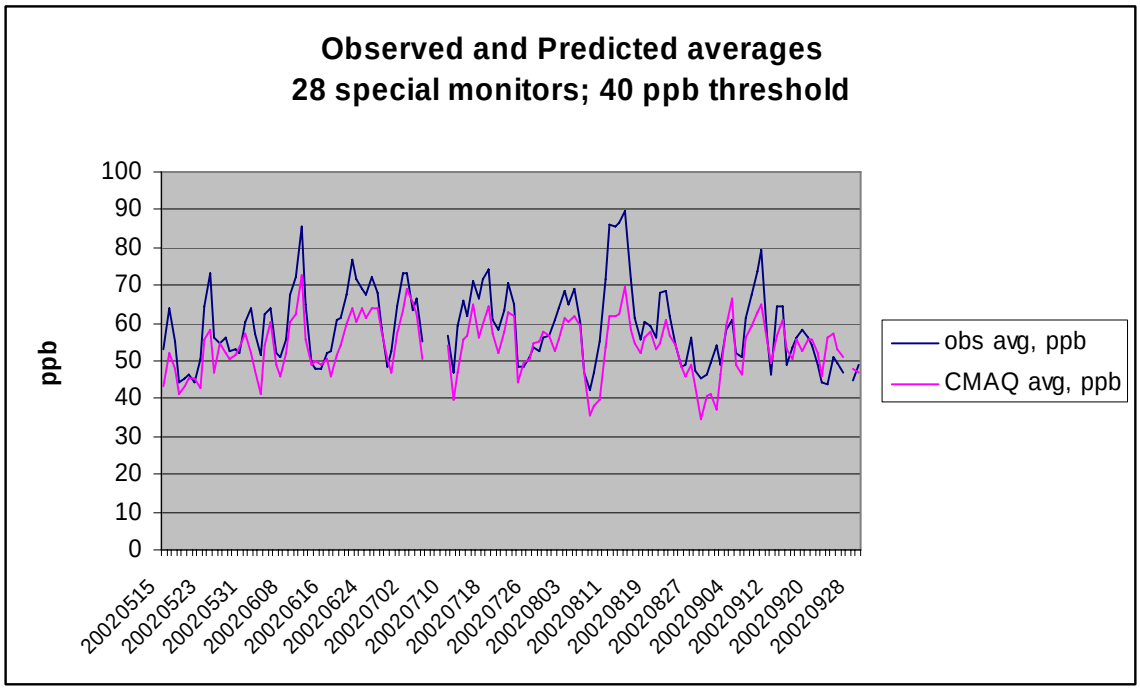



Figure 4.

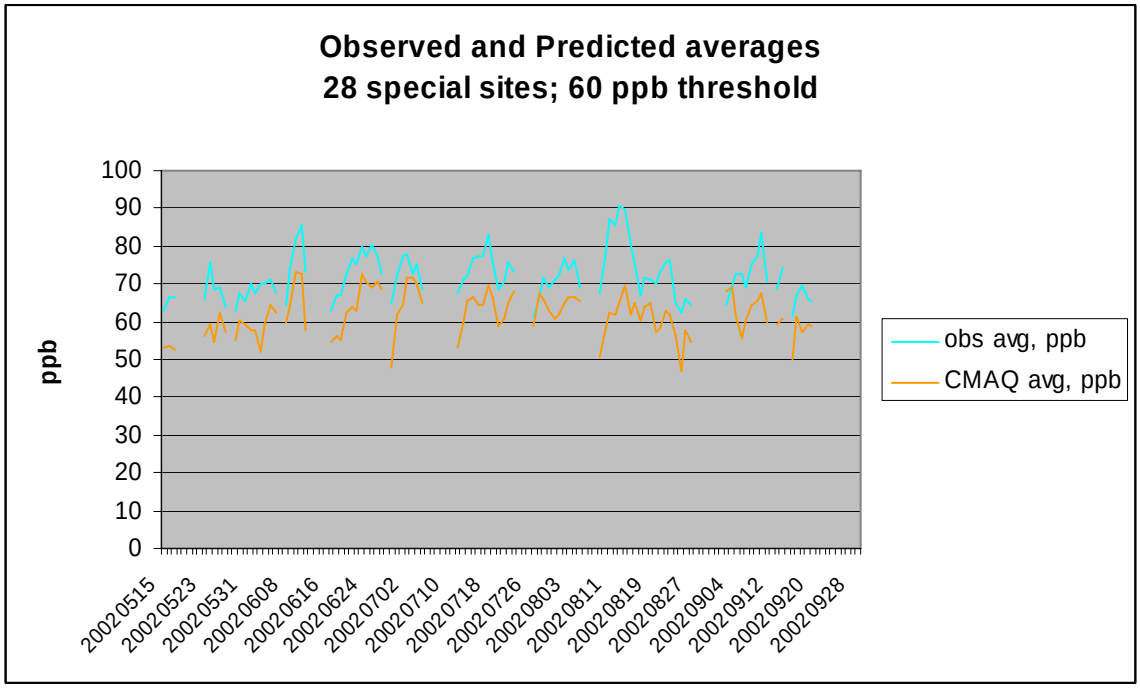


Figure 5.

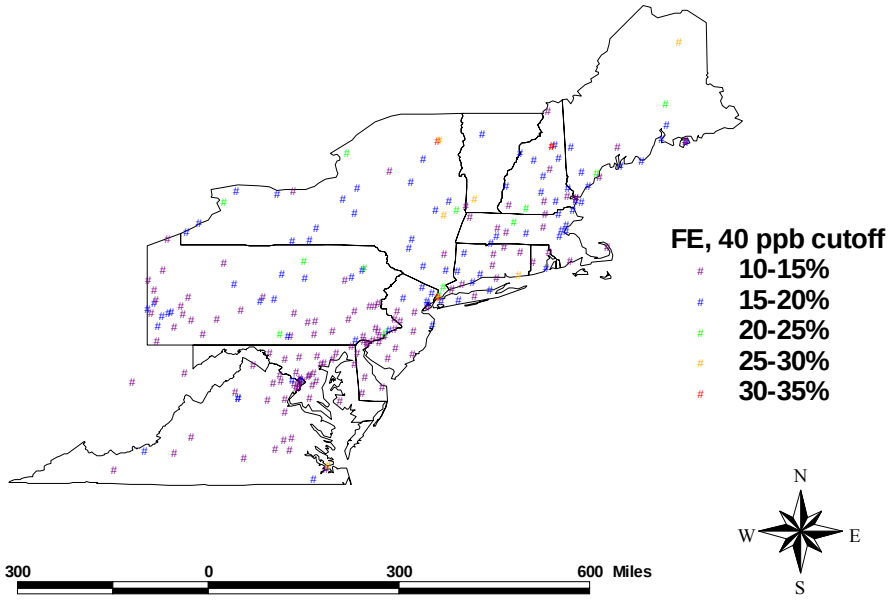



Figure 6.

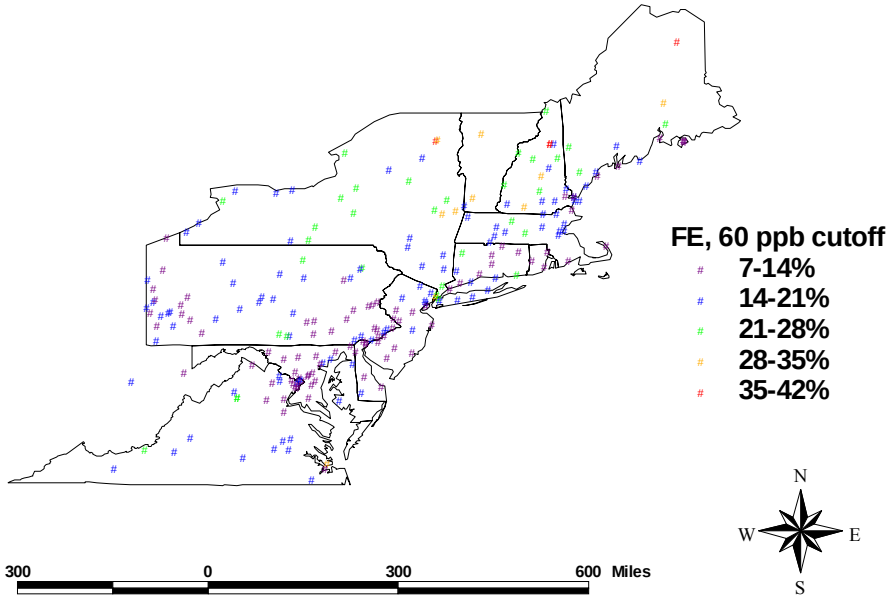


Figure 7.

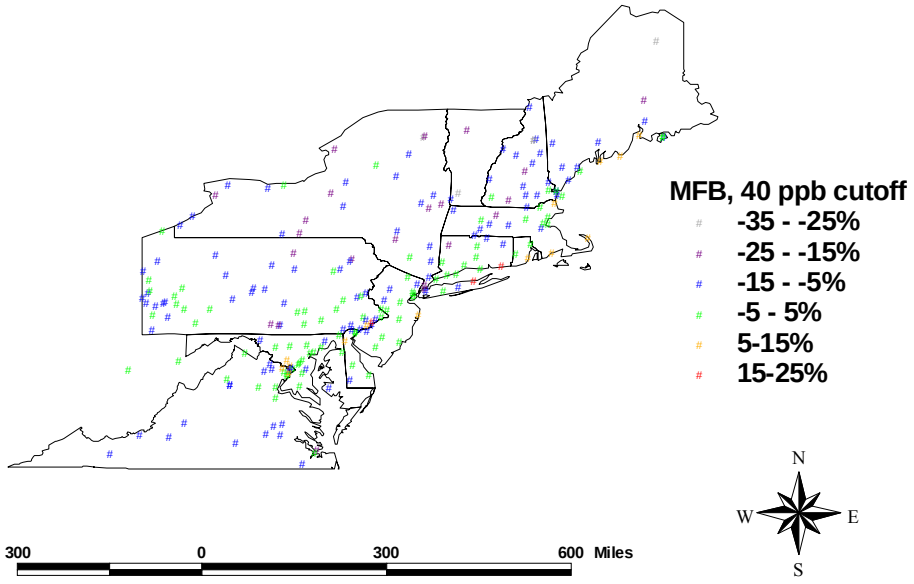



Figure 8.

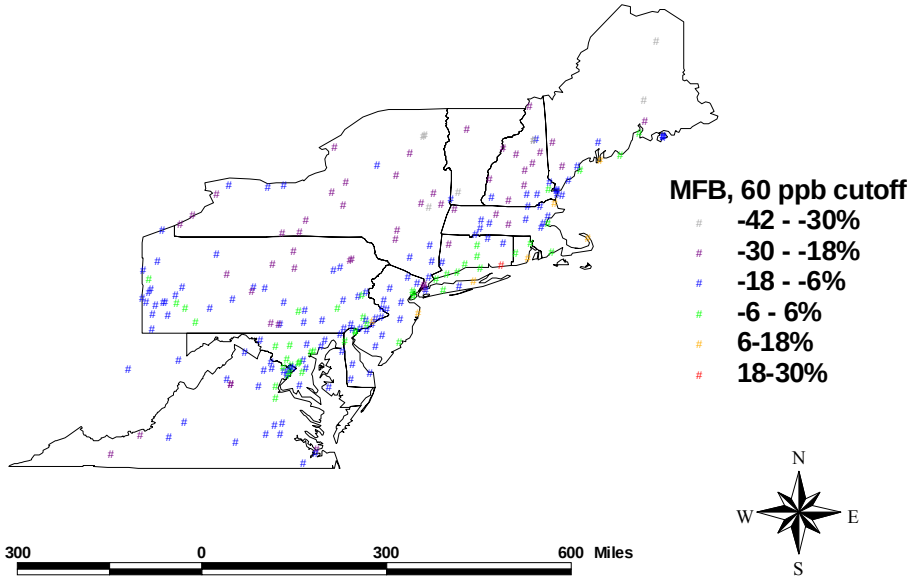


Figure 9.

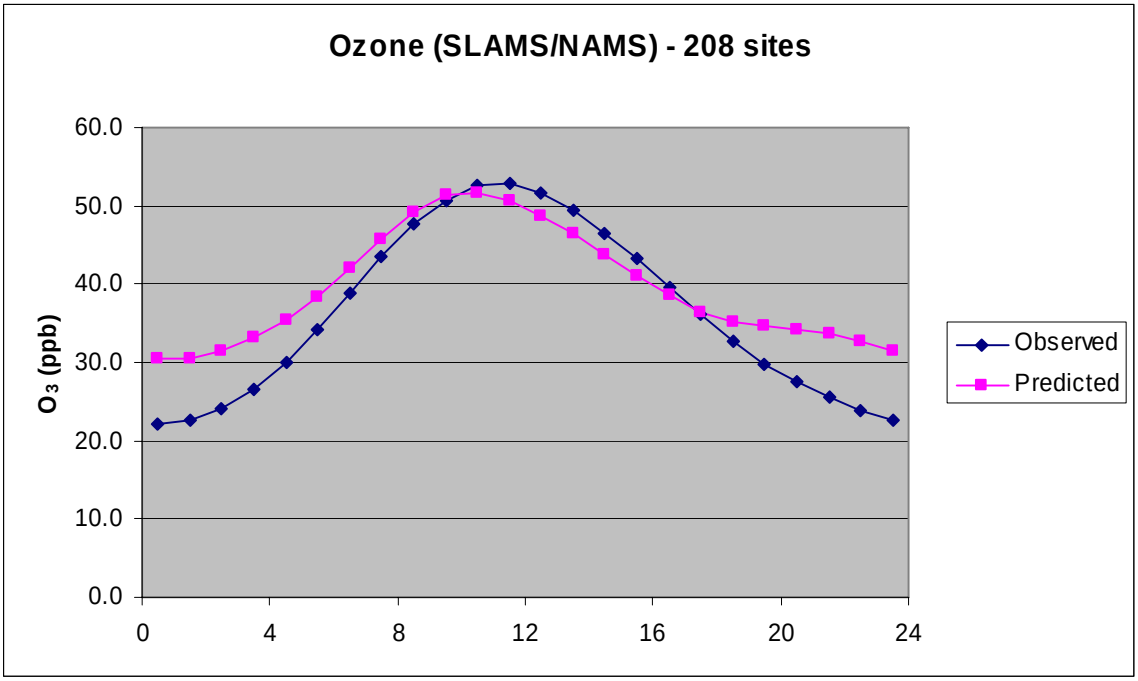



Figure 10.

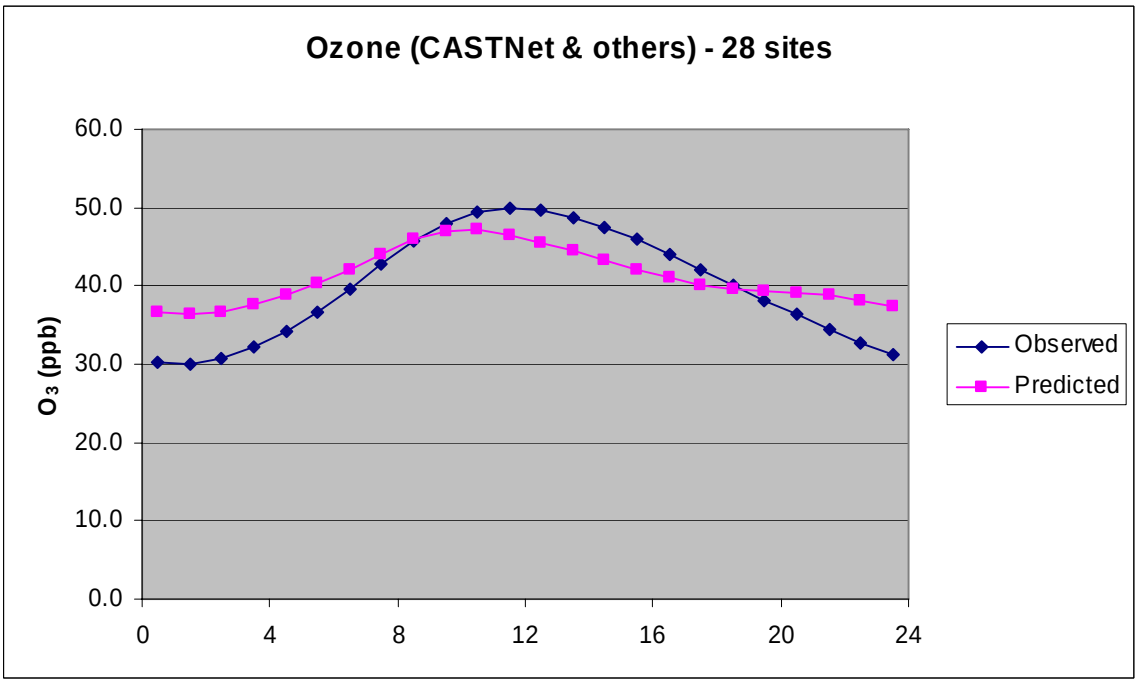


Figure 11.

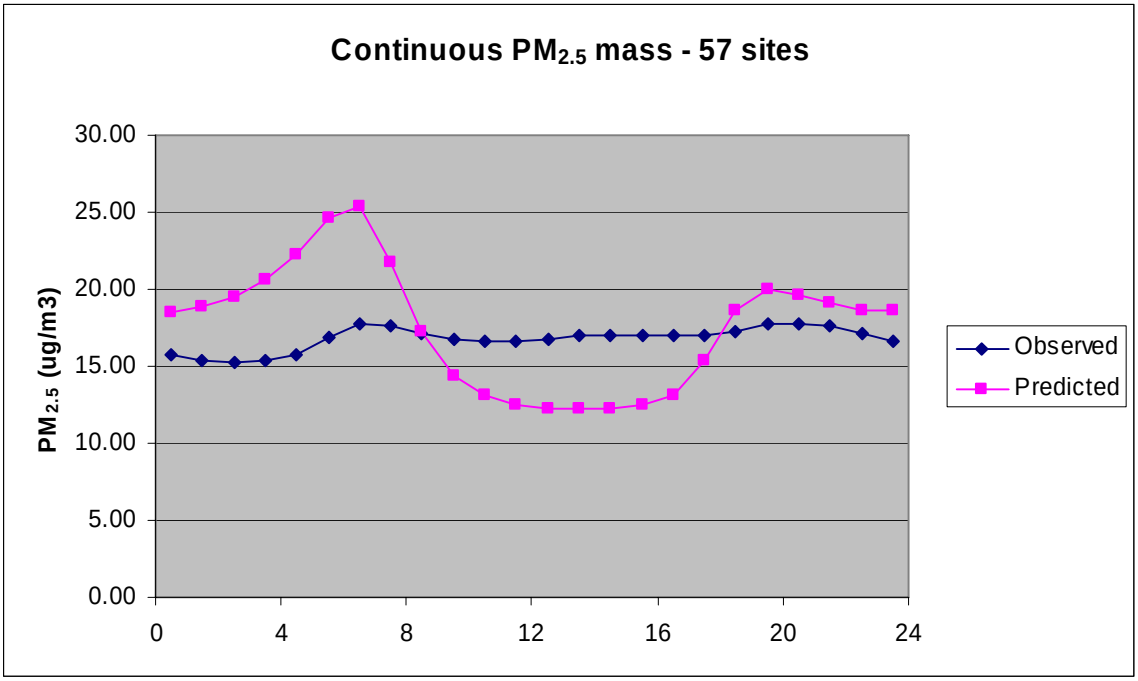



Figure 12.

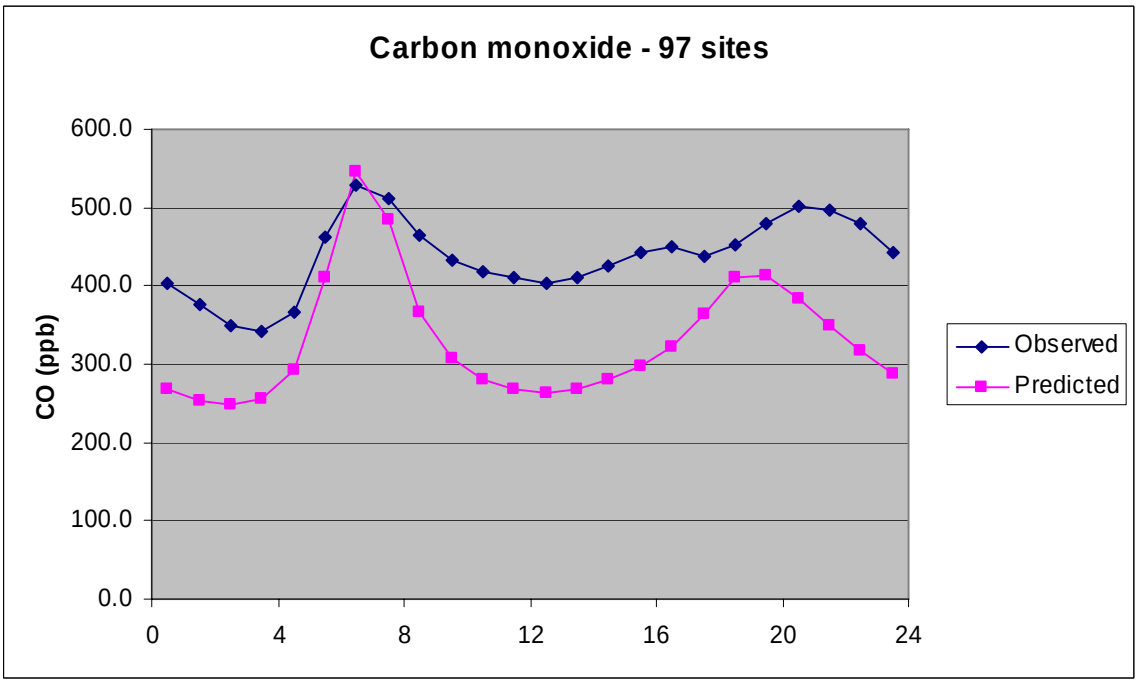


Figure 13.

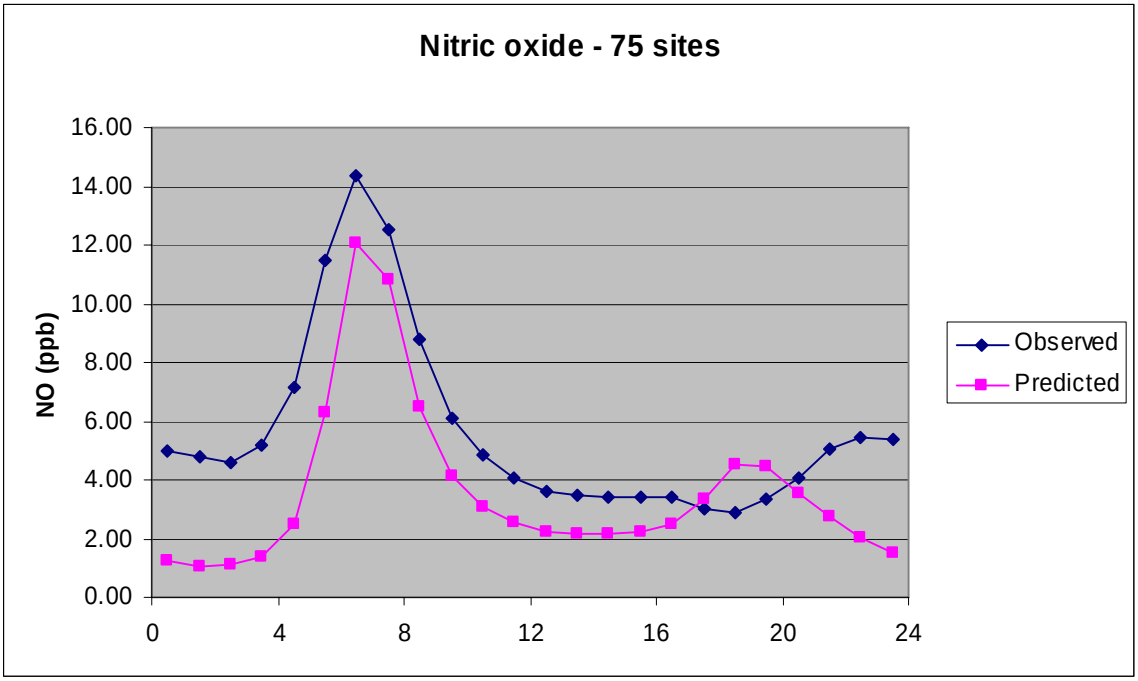



Figure 14.

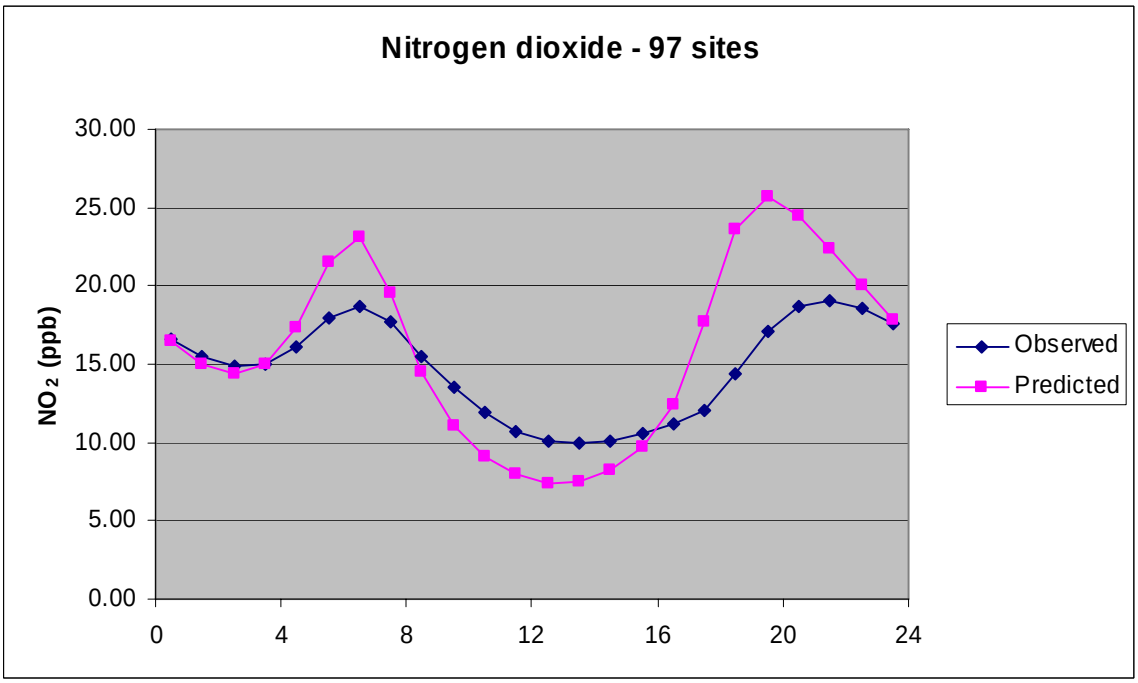


Figure 15.

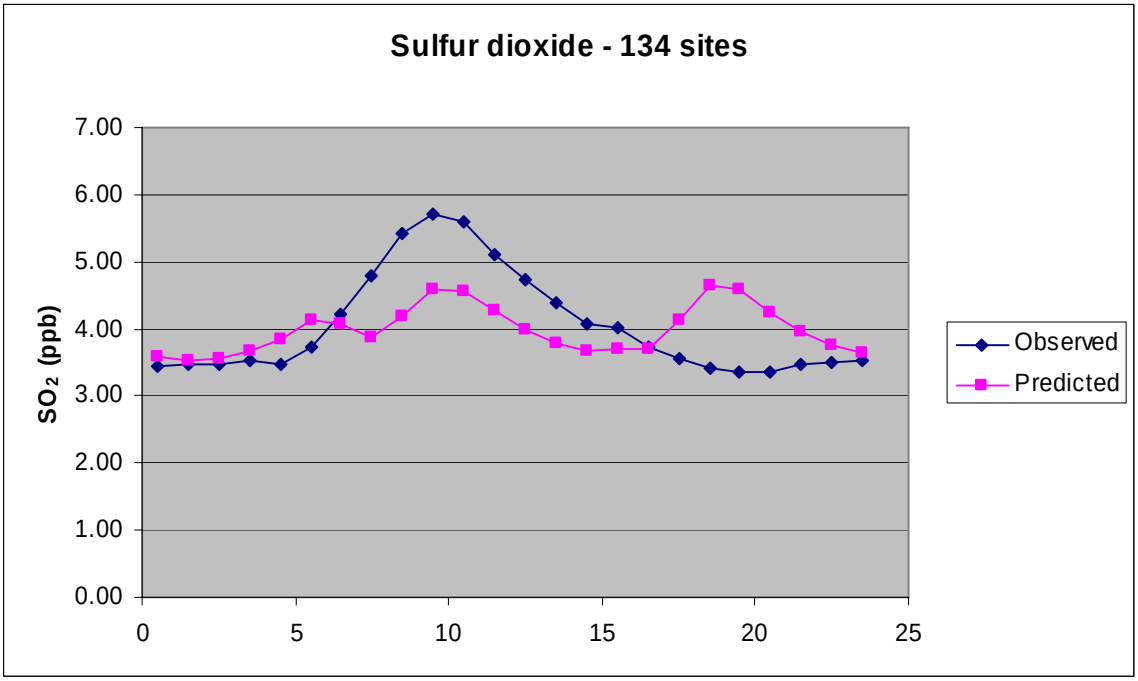



Figure 16.

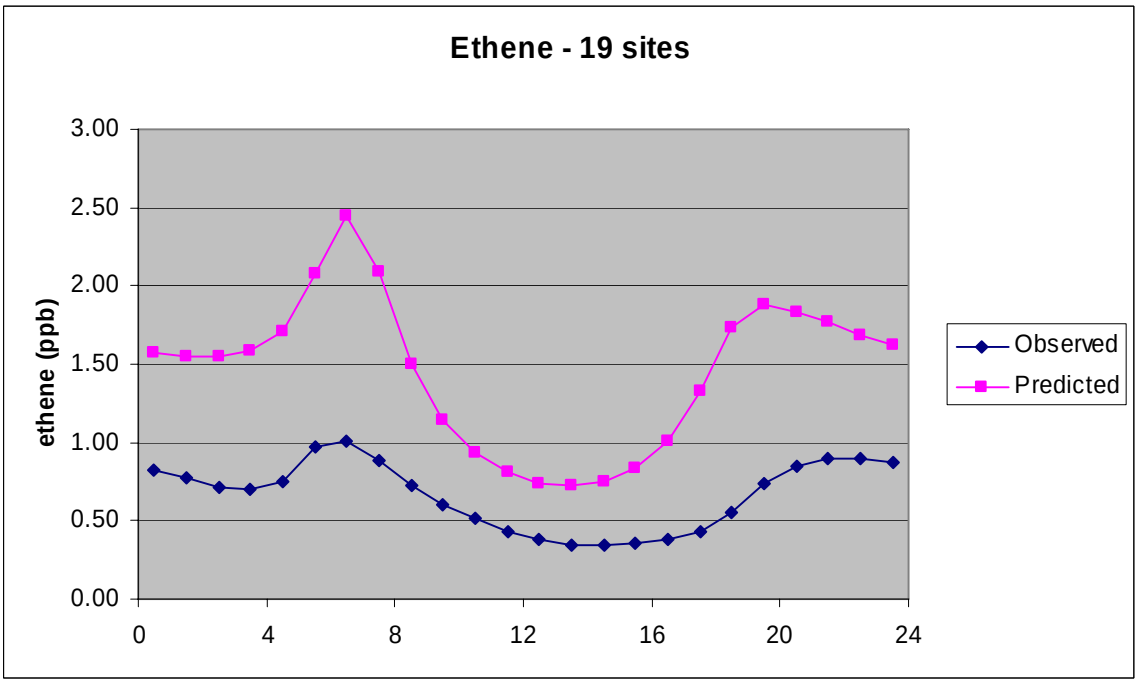


Figure 17.

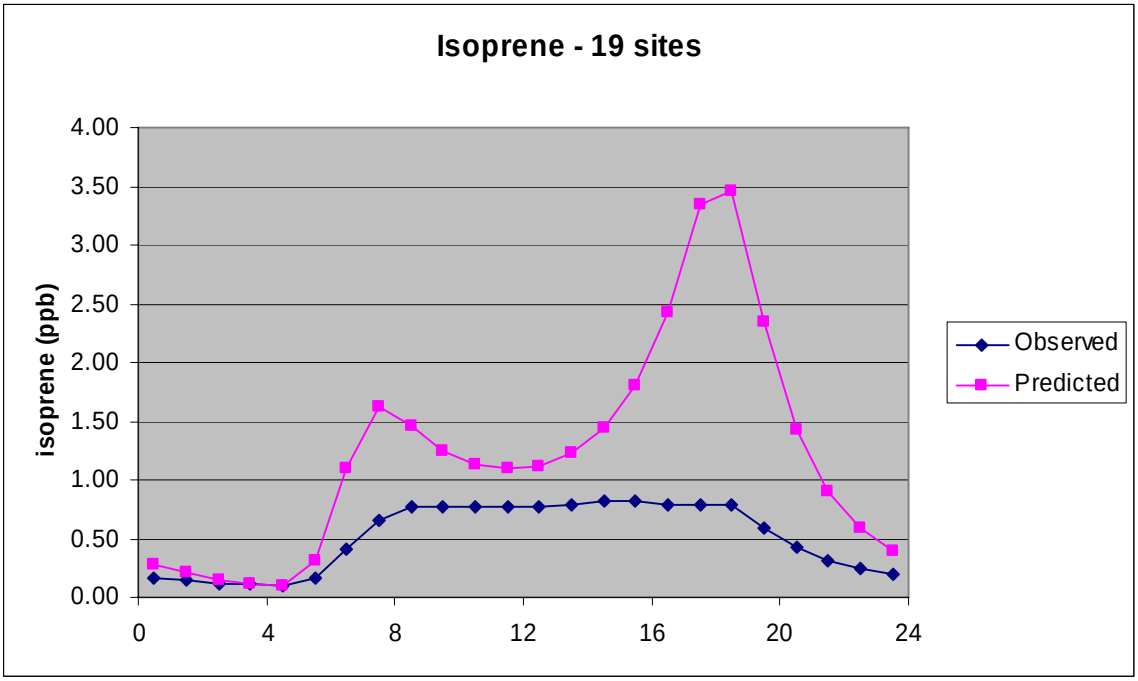



Figure 18.

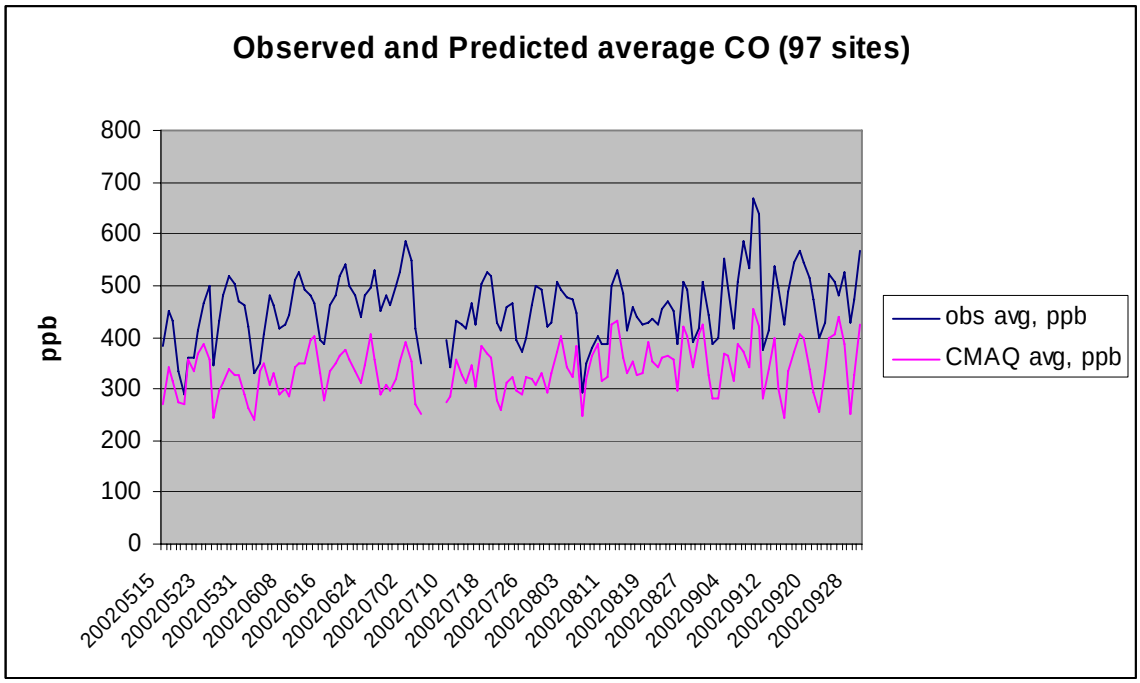


Figure 19.

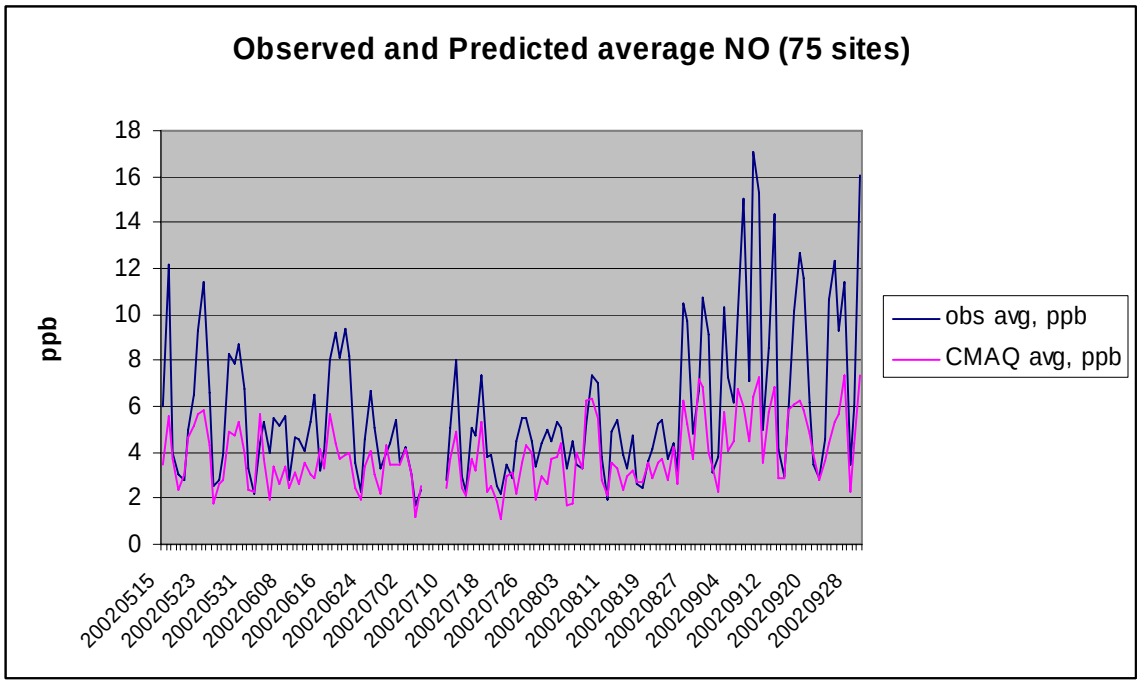



Figure 20.

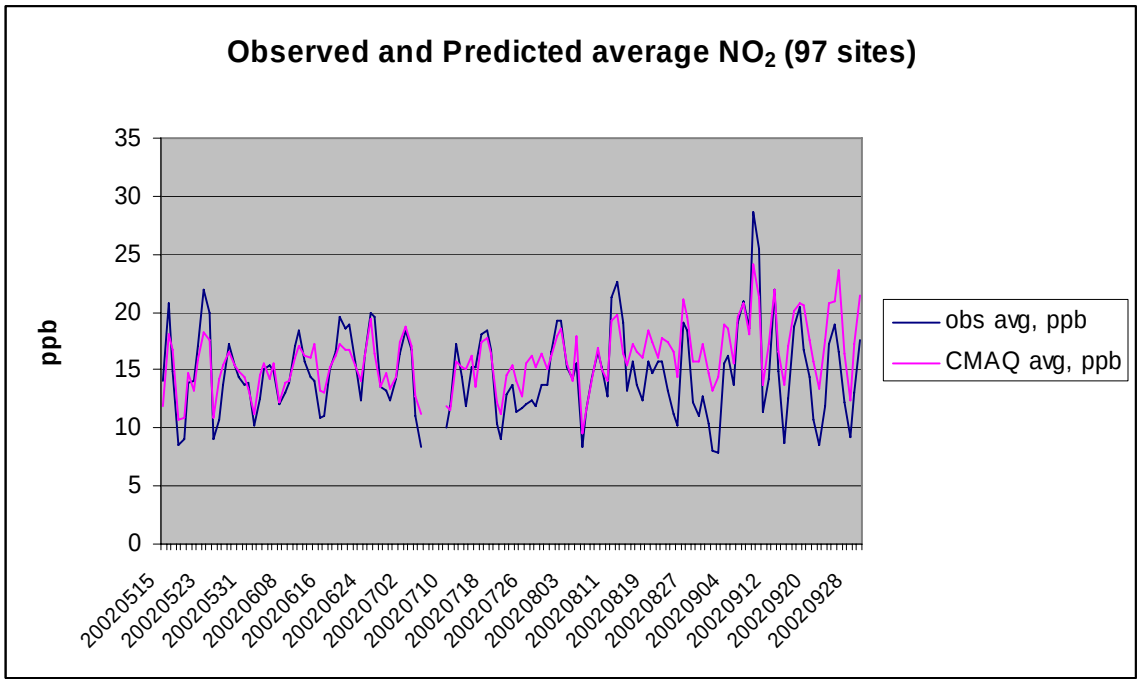


Figure 21.

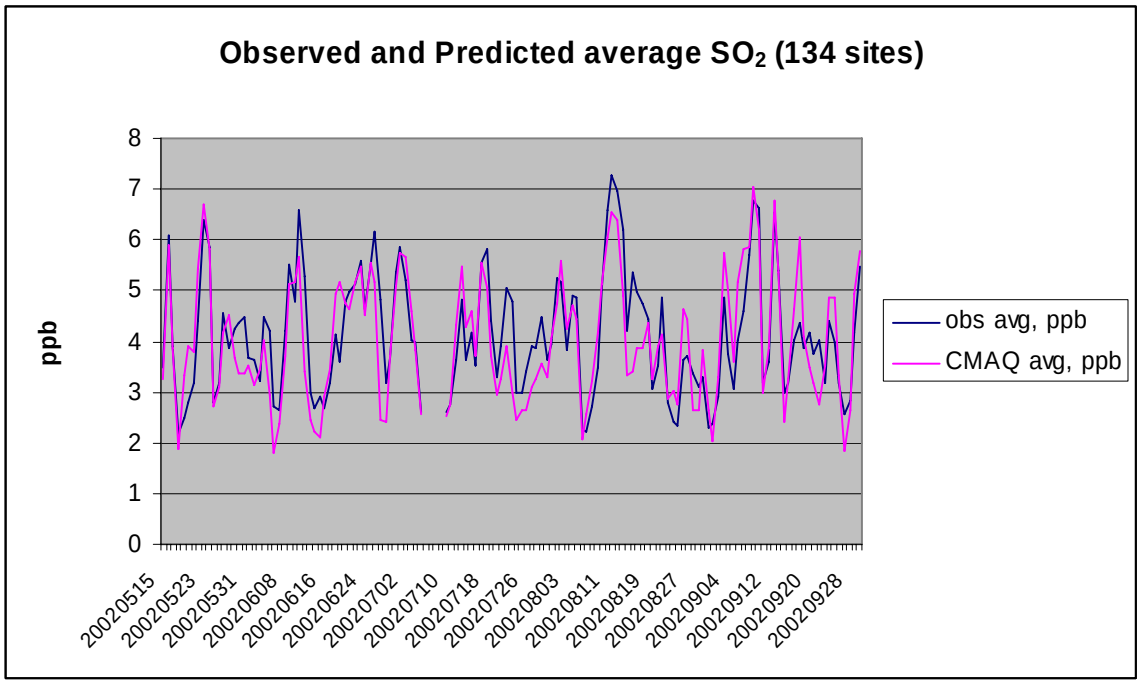



Figure 22.

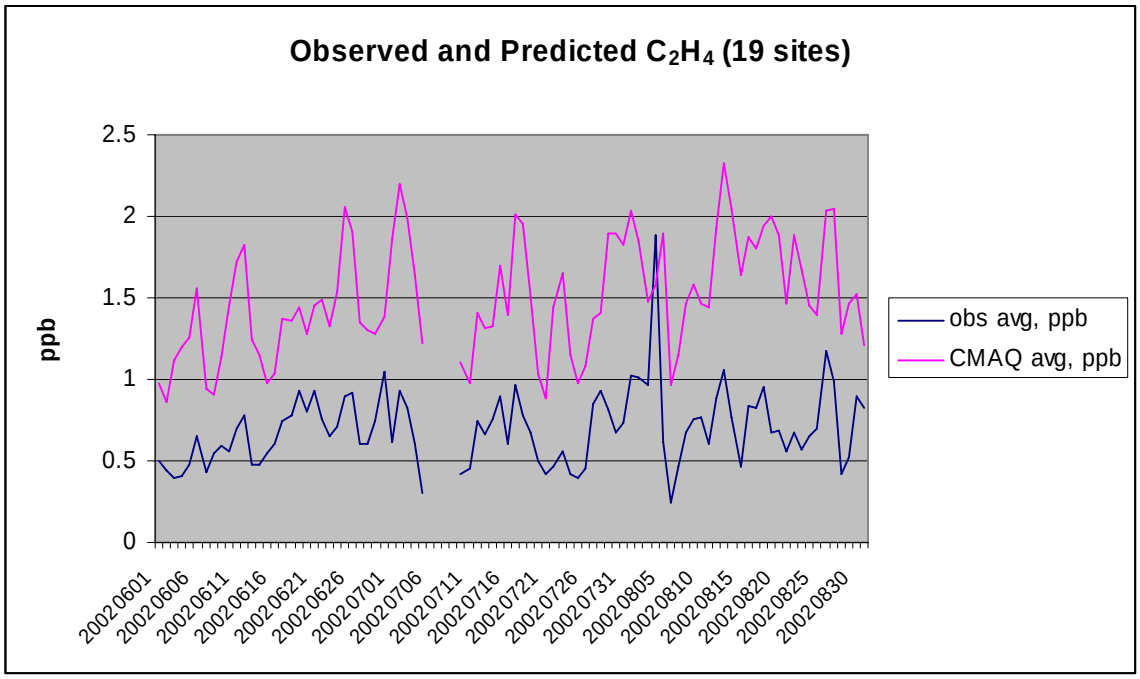


Figure 23.

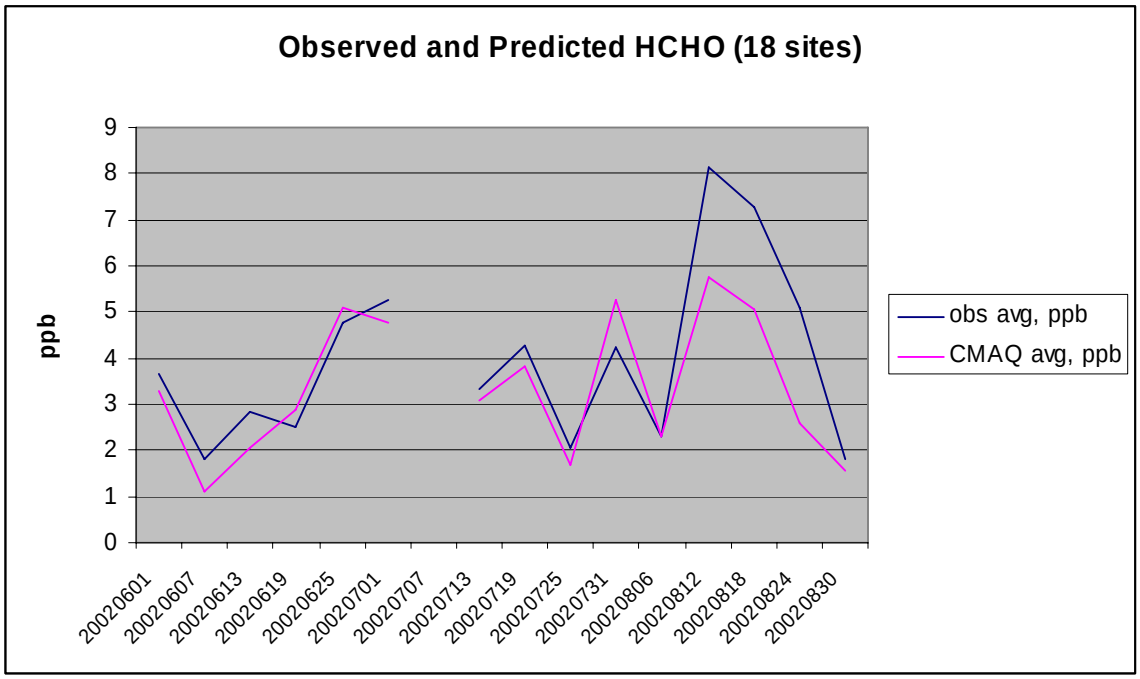



Figure 24.

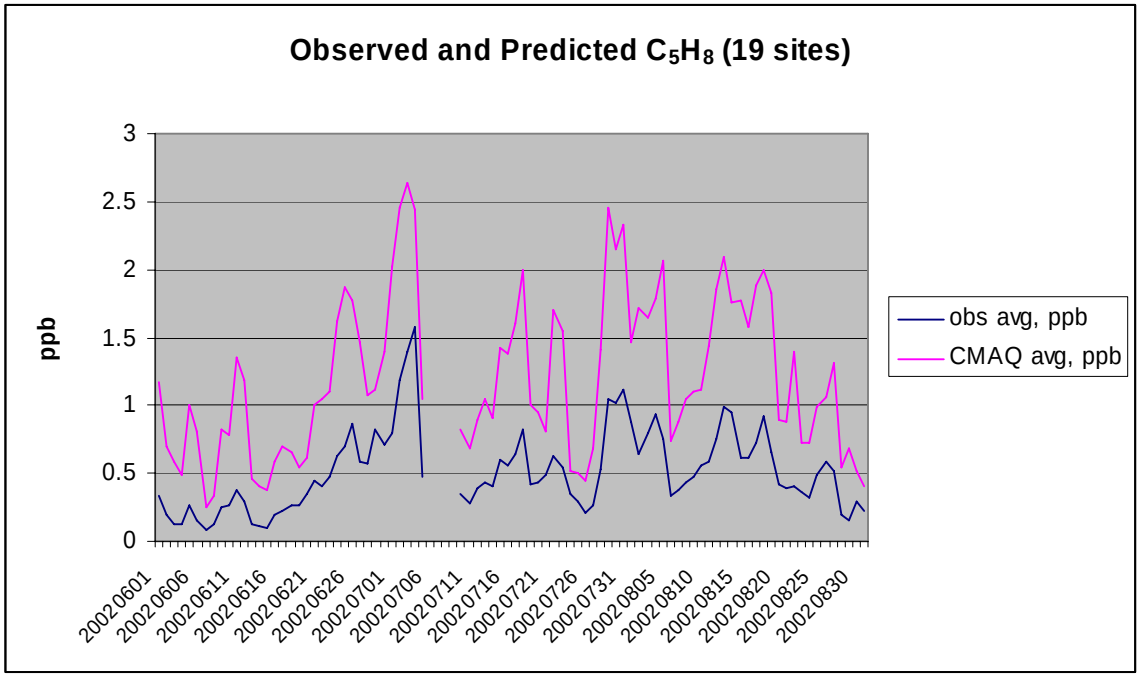


Figure 25.

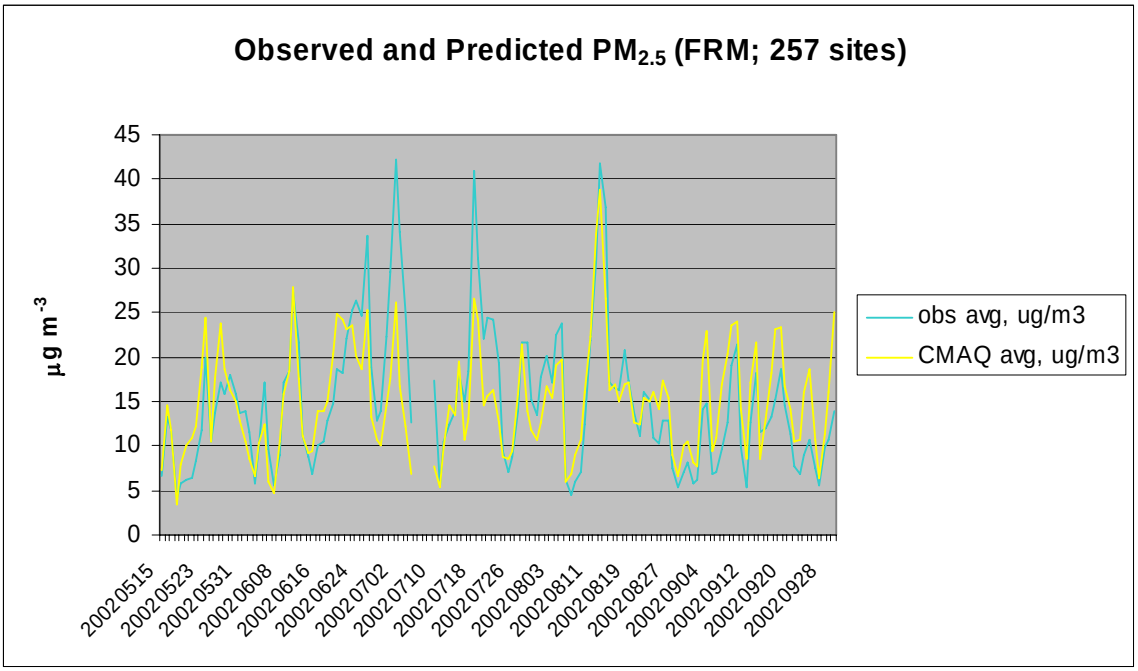



Figure 26.

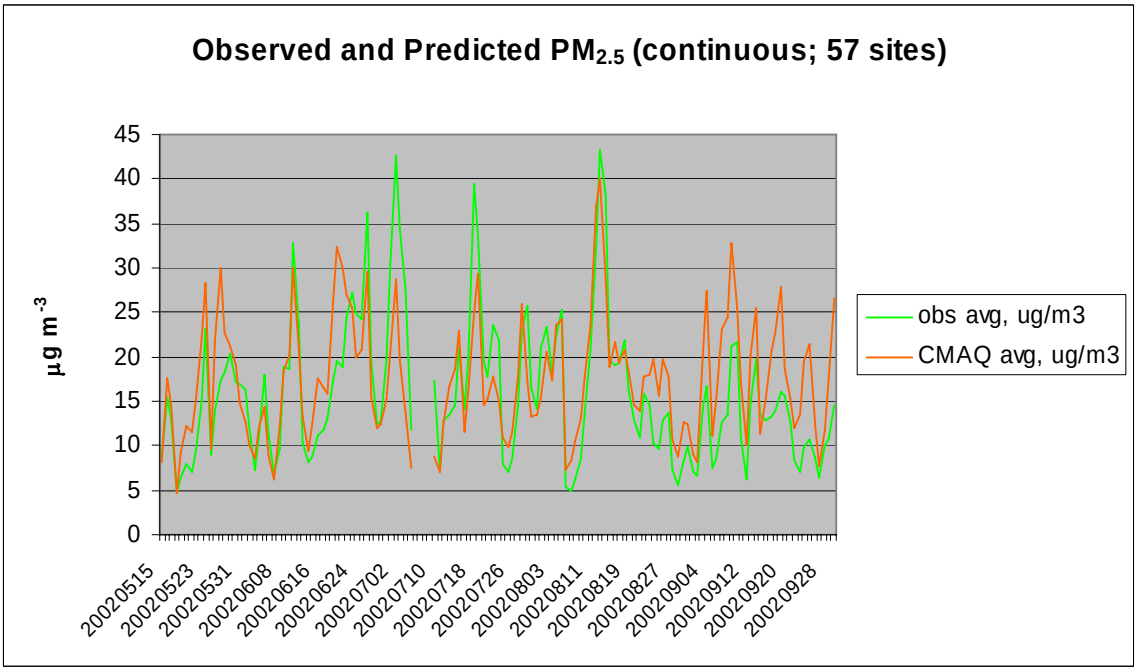


Figure 27.

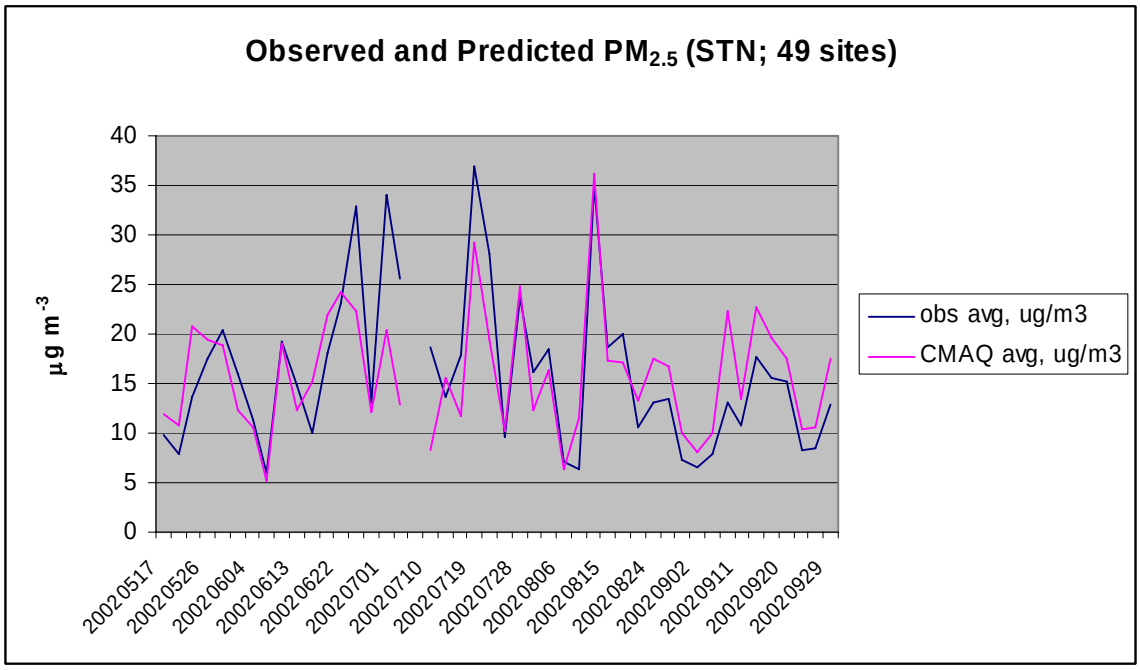



Figure 28.

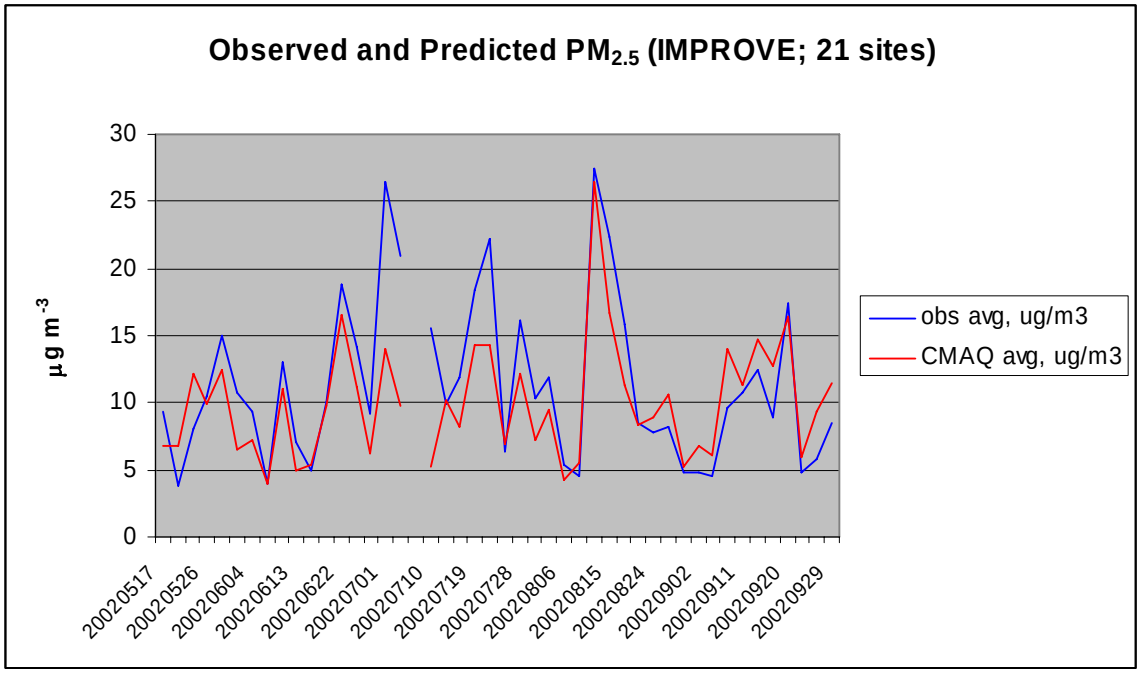


Figure 29.

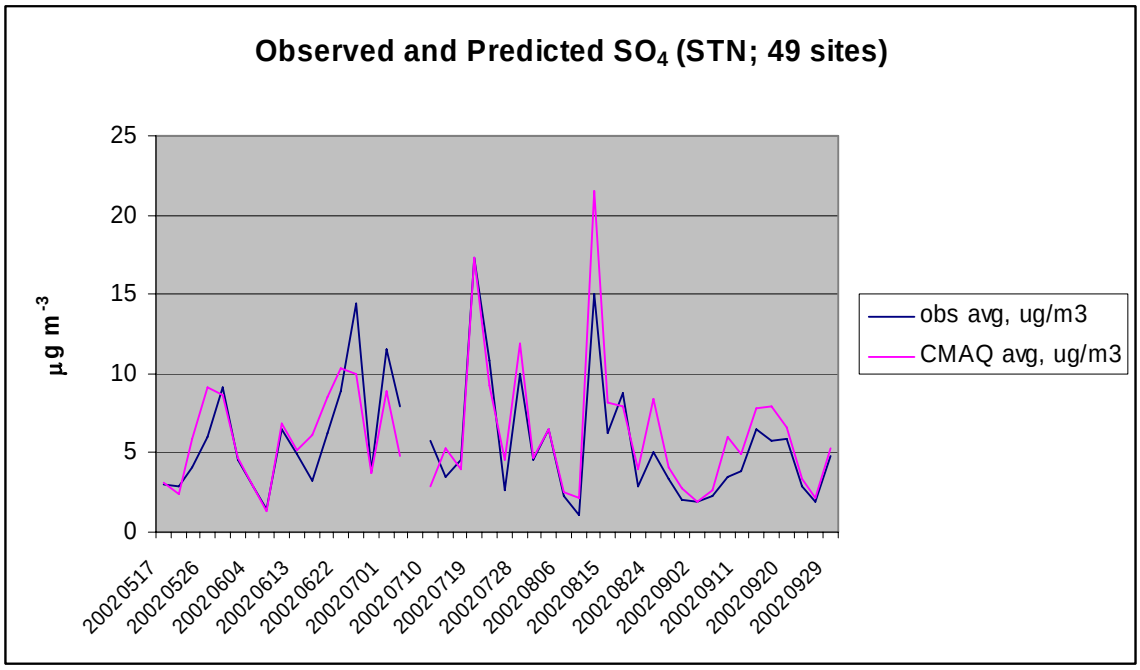



Figure 30.

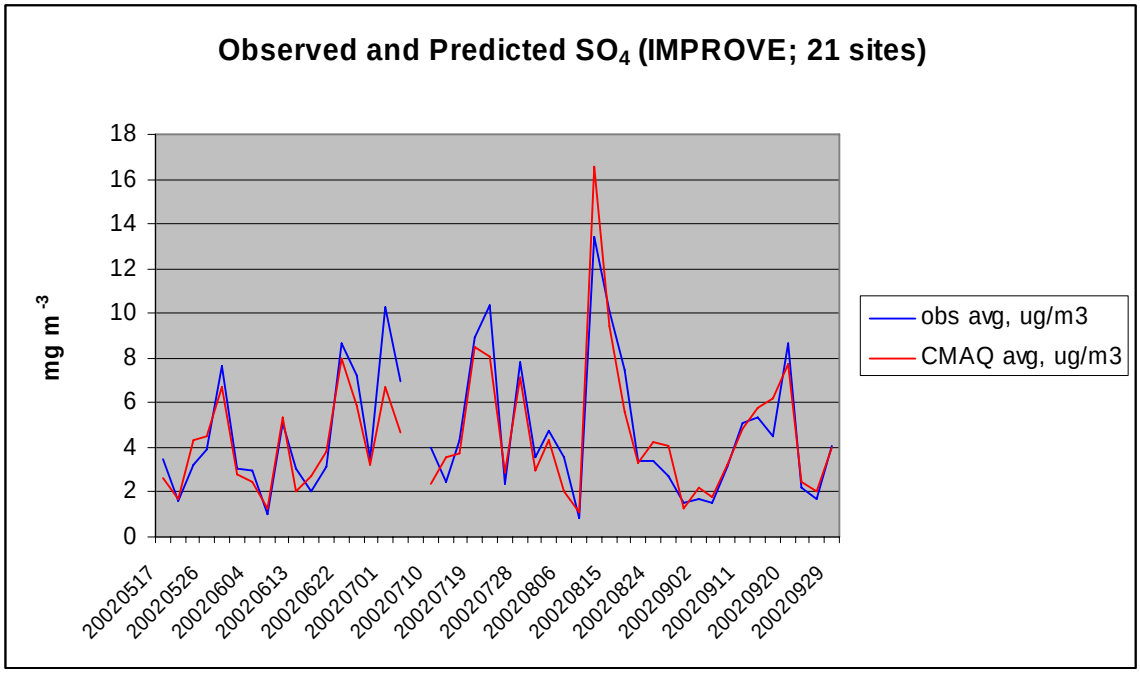


Figure 31.

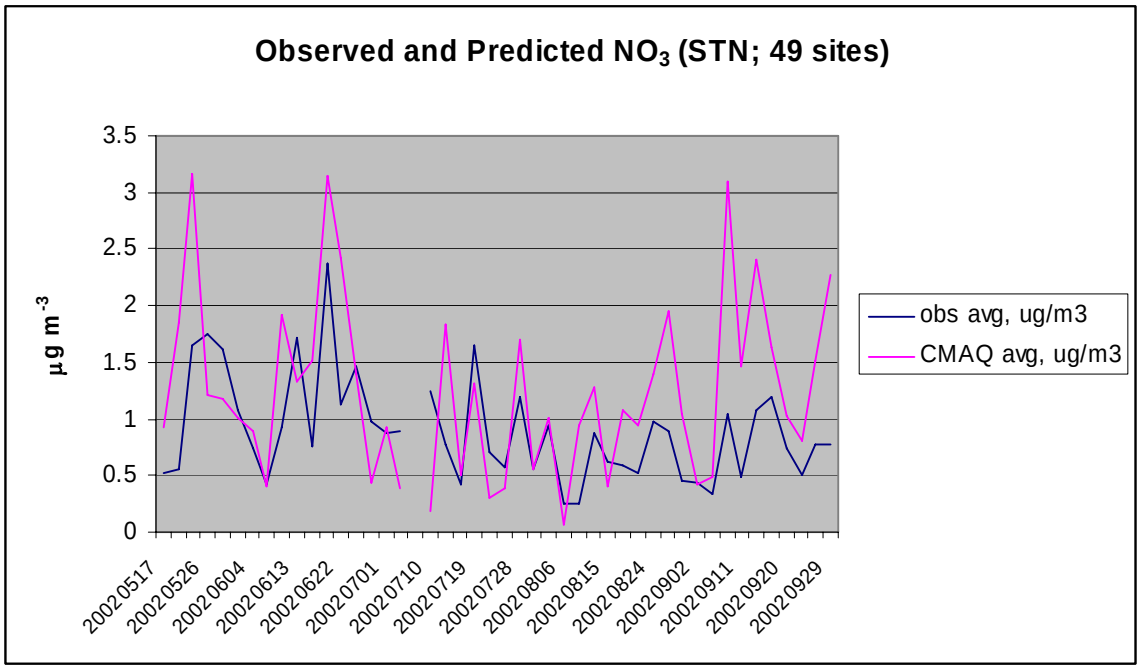



Figure 32.

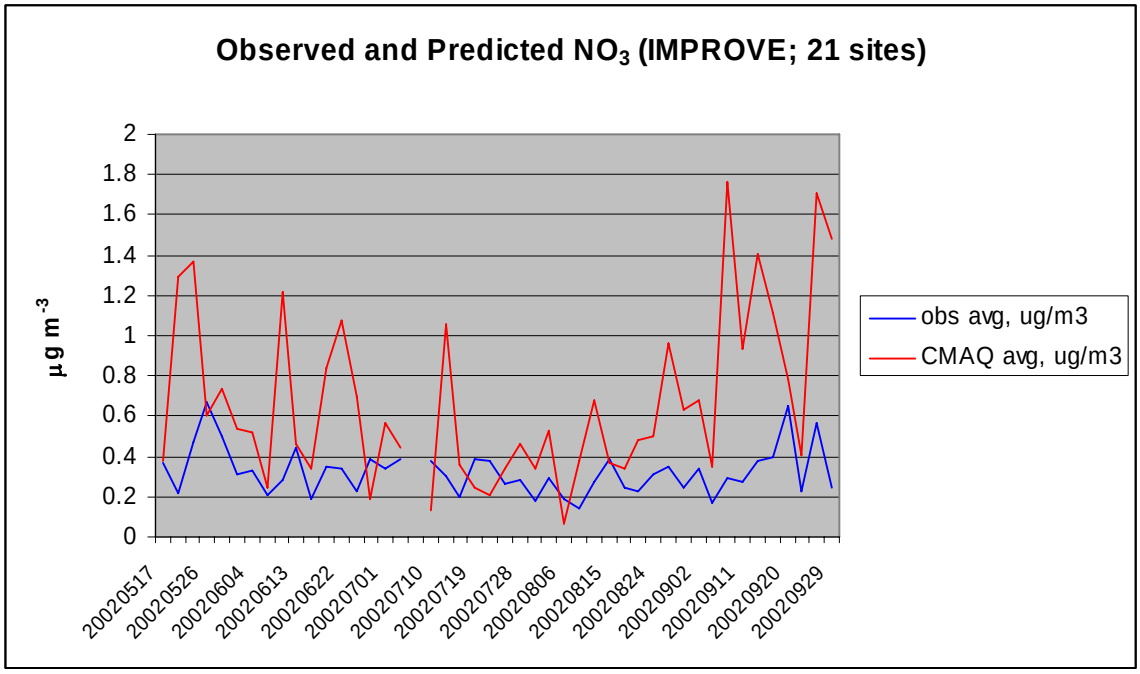


Figure 33.

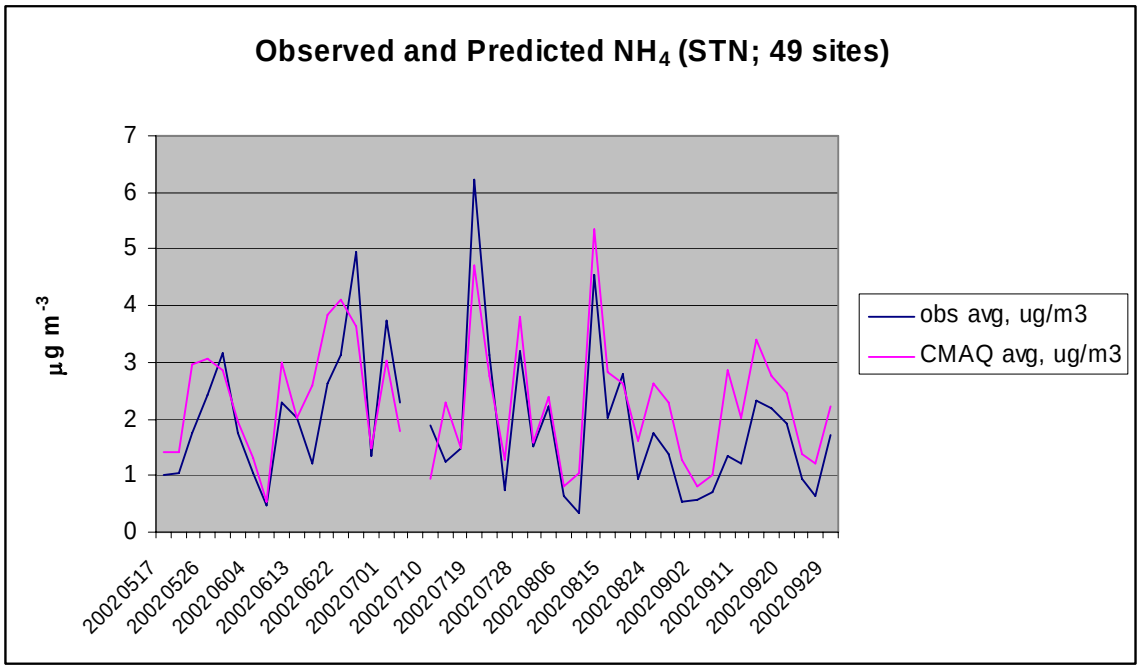



Figure 34.

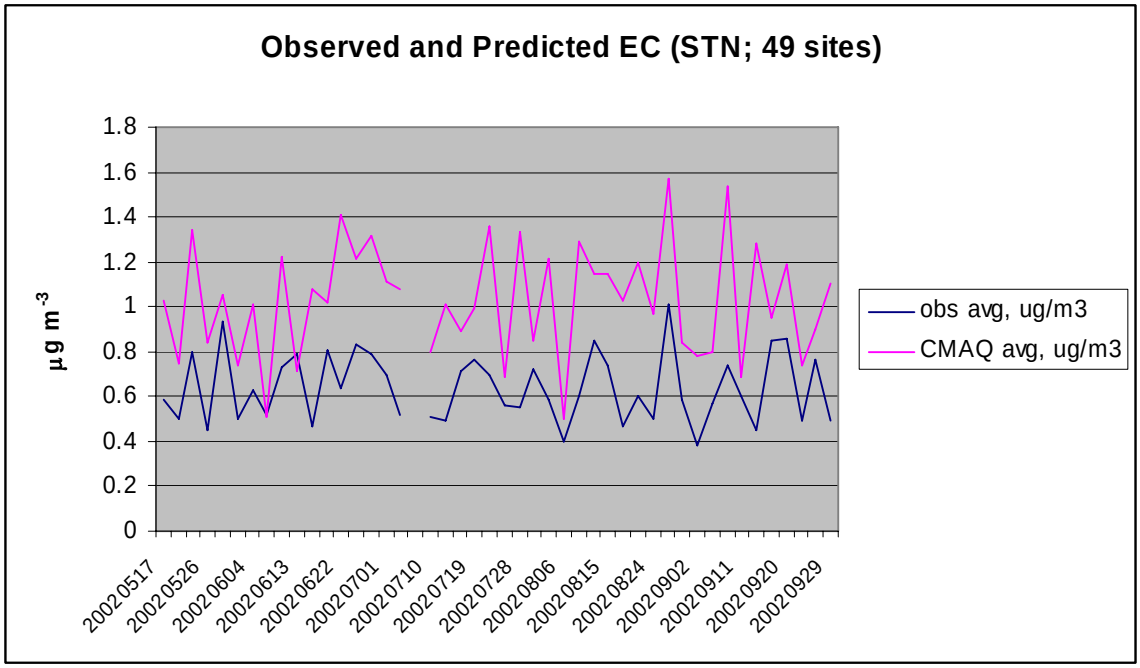


Figure 35.

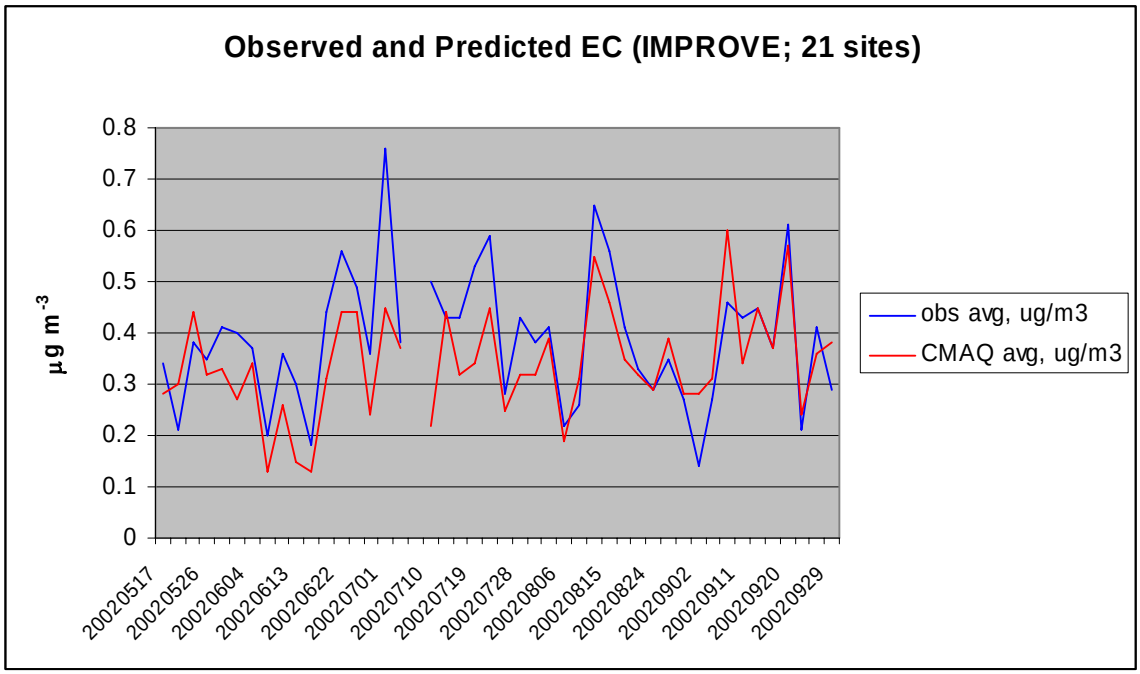



Figure 36.

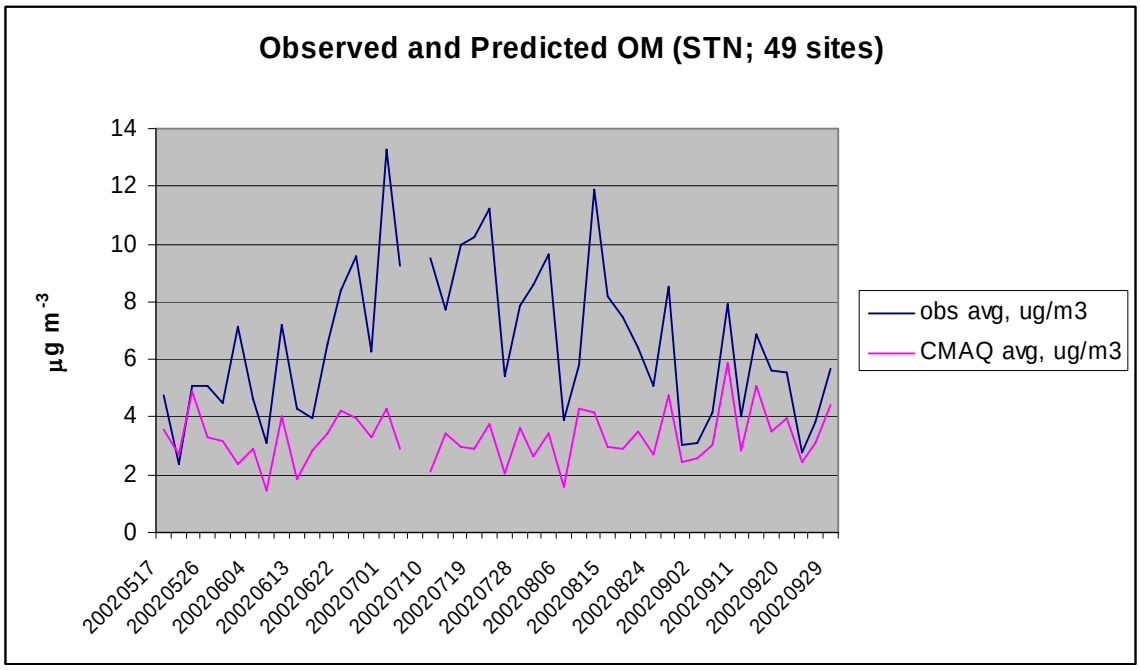


Figure 37.

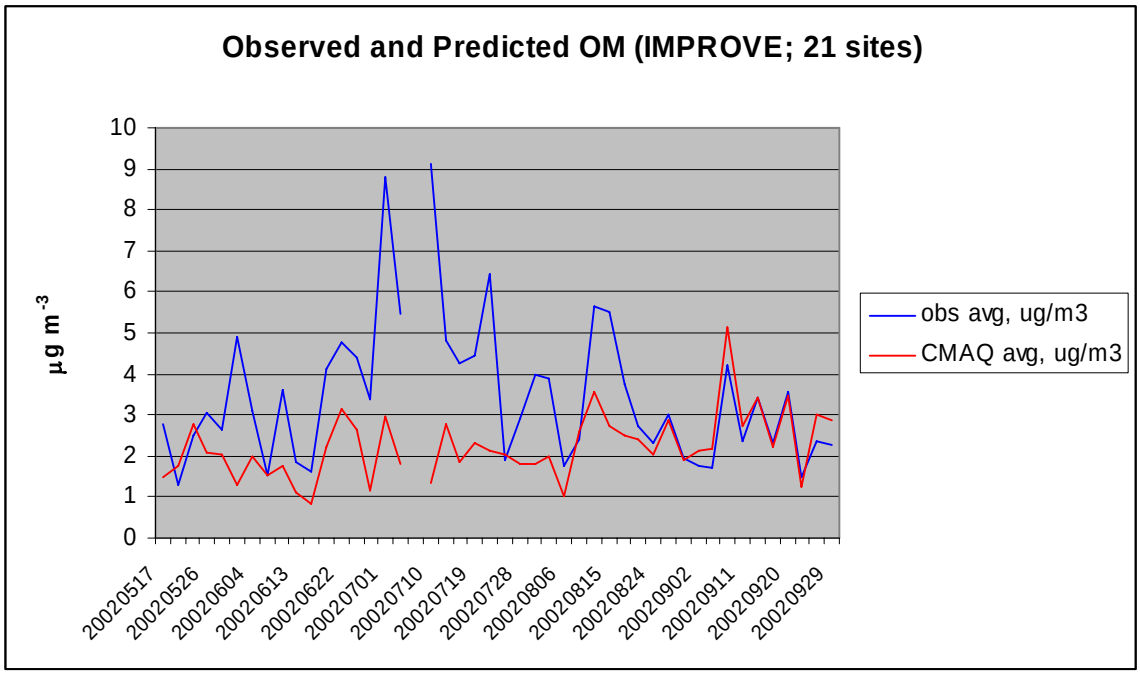



Figure 38.

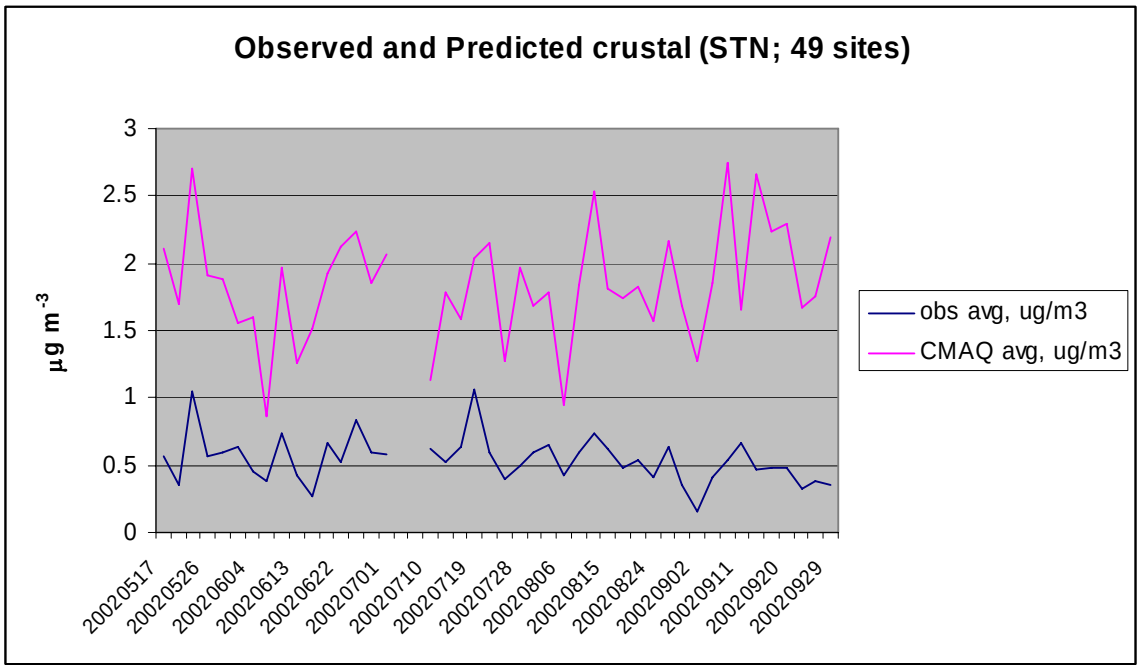


Figure 39.

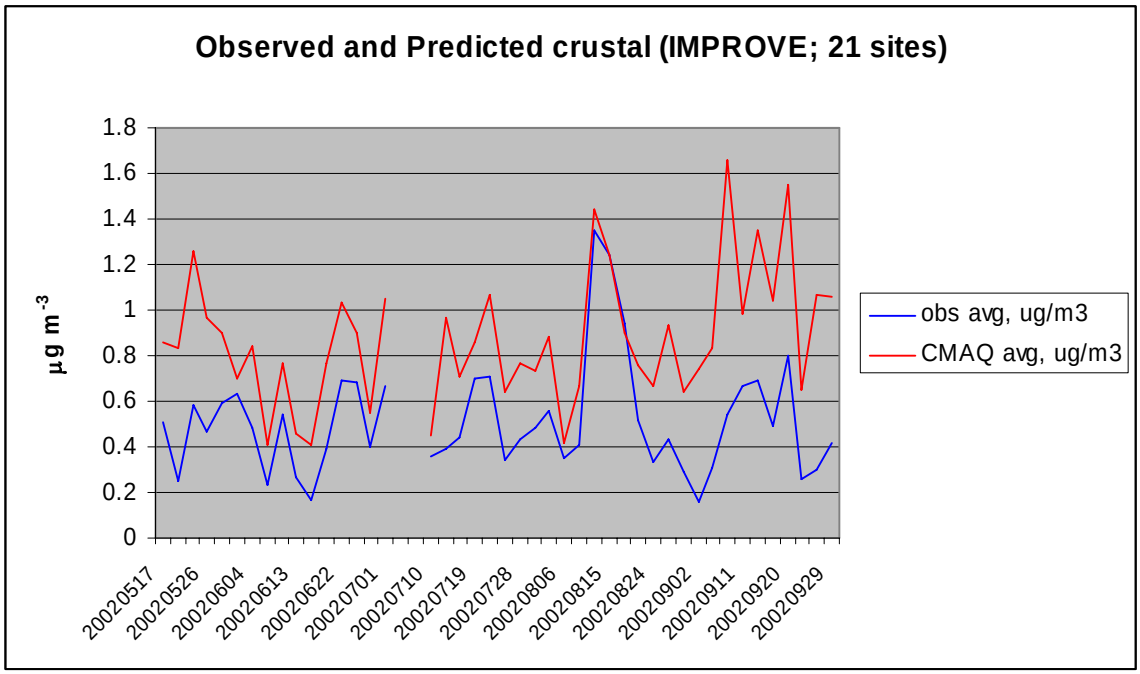



Figure 40.

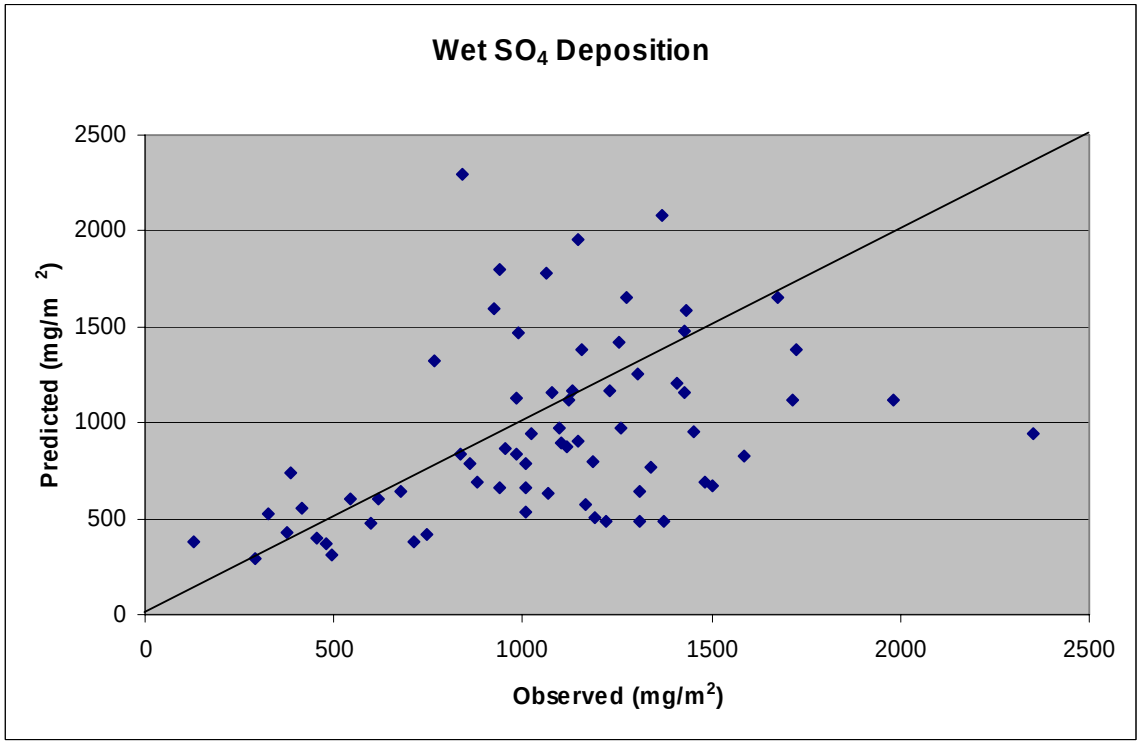


Figure 41.

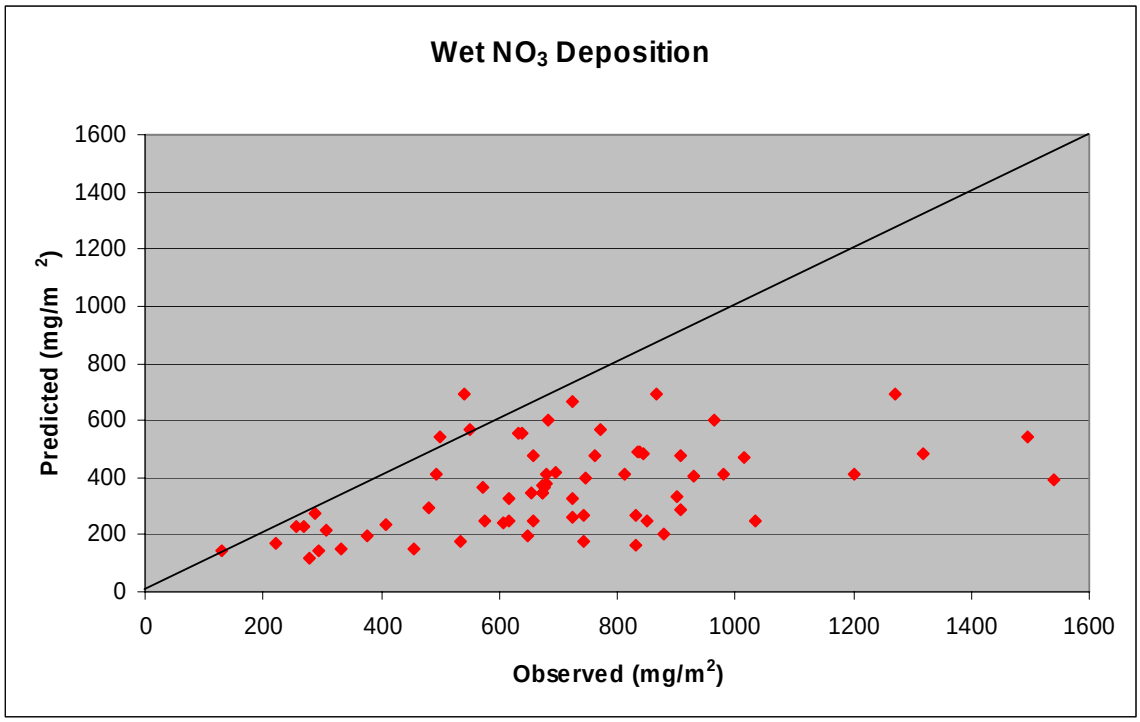



Figure 42.

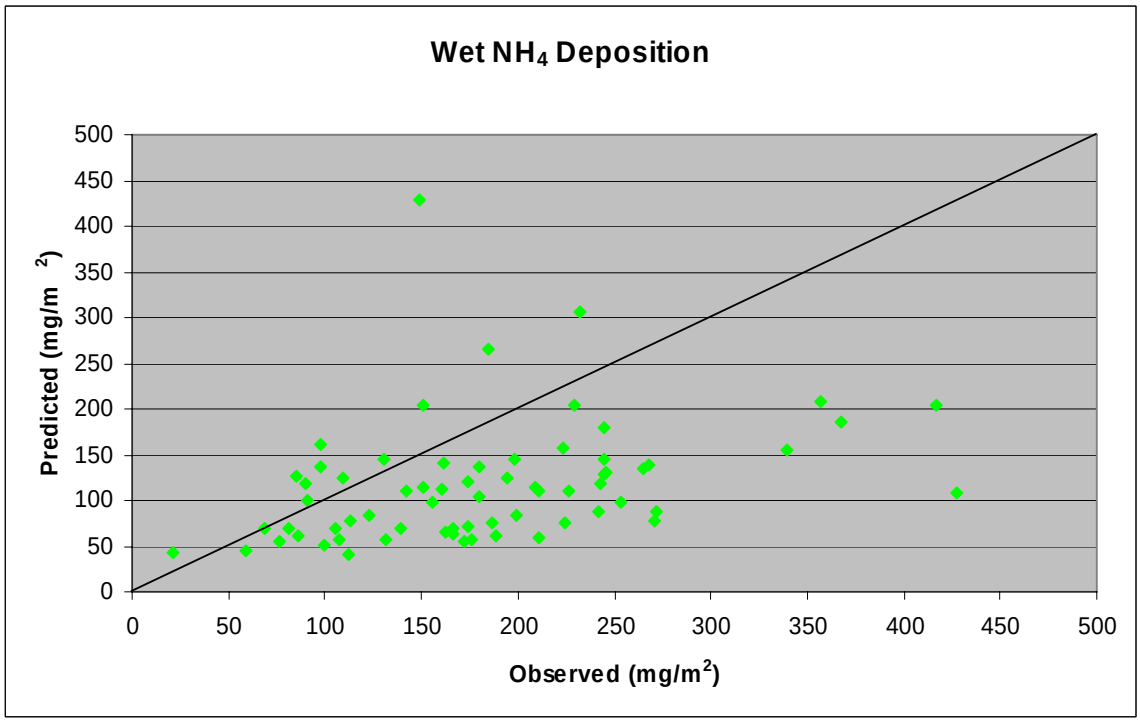


Figure 43.

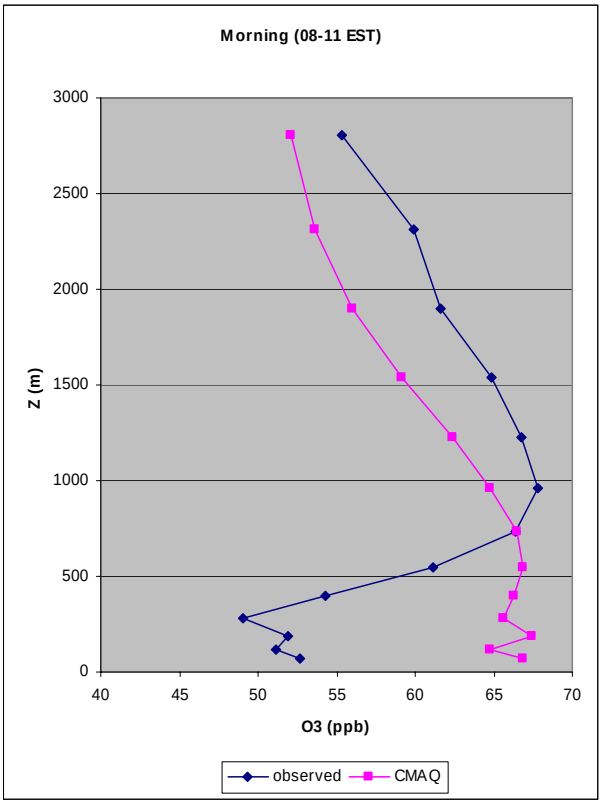



Figure 44.

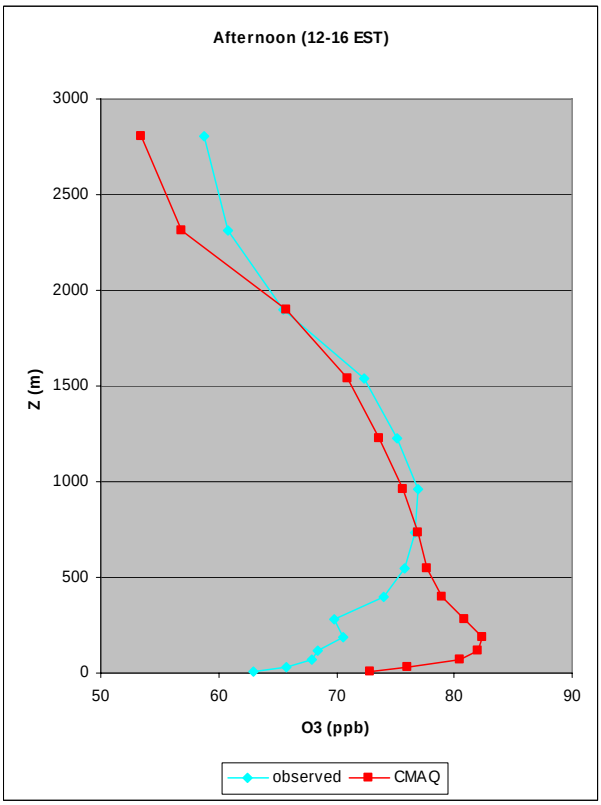


Figure 45.

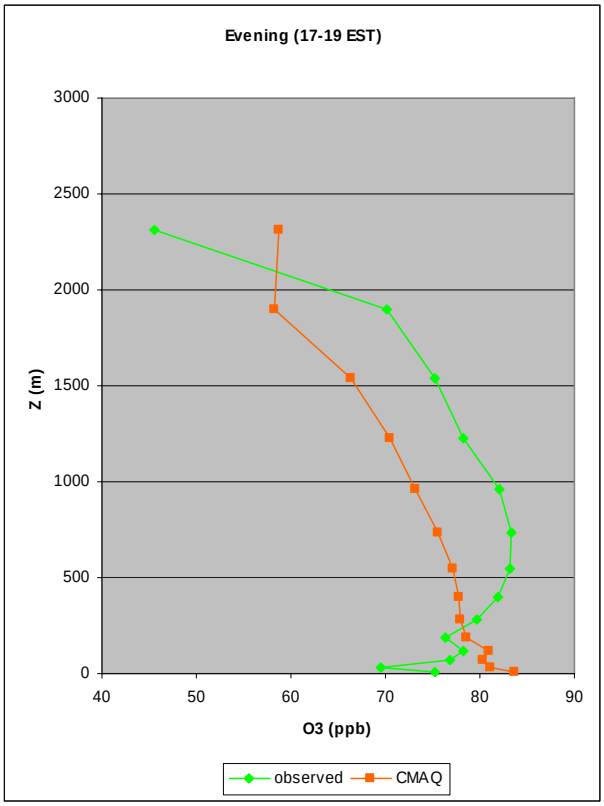



Figure 46.

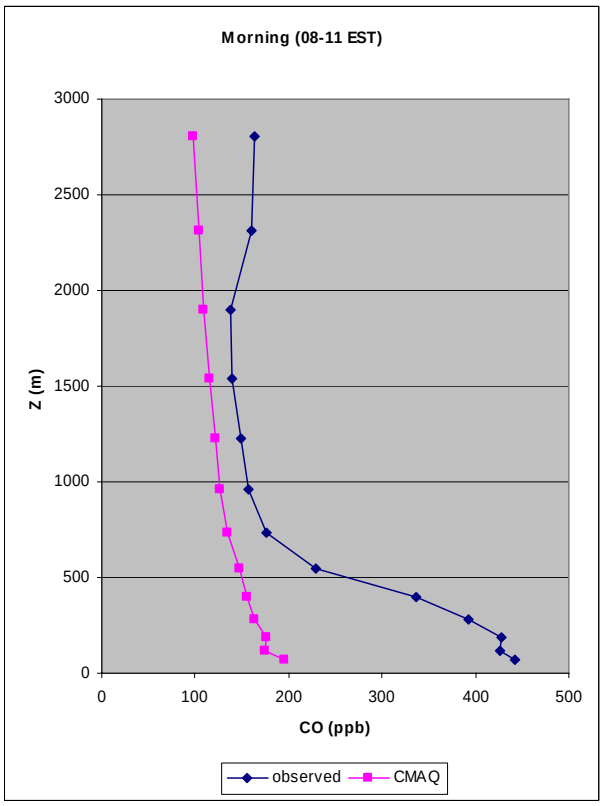


Figure 47.

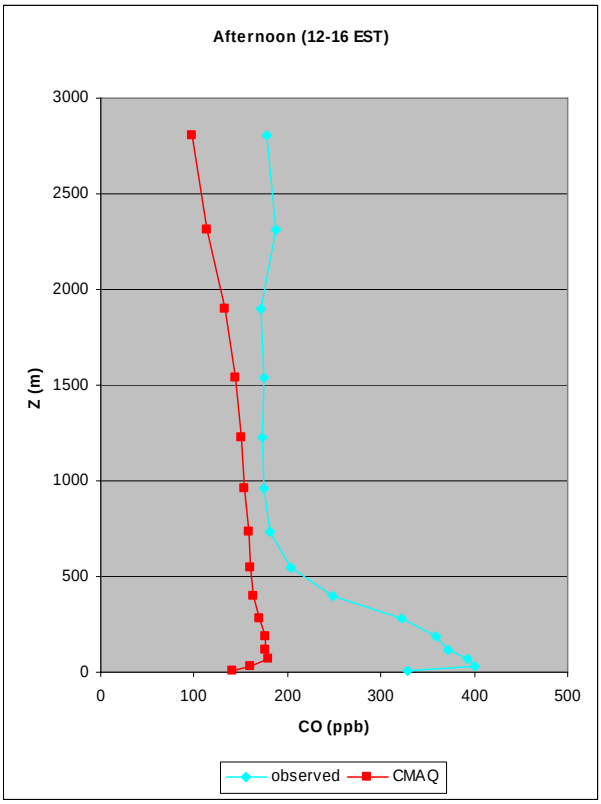



Figure 48.

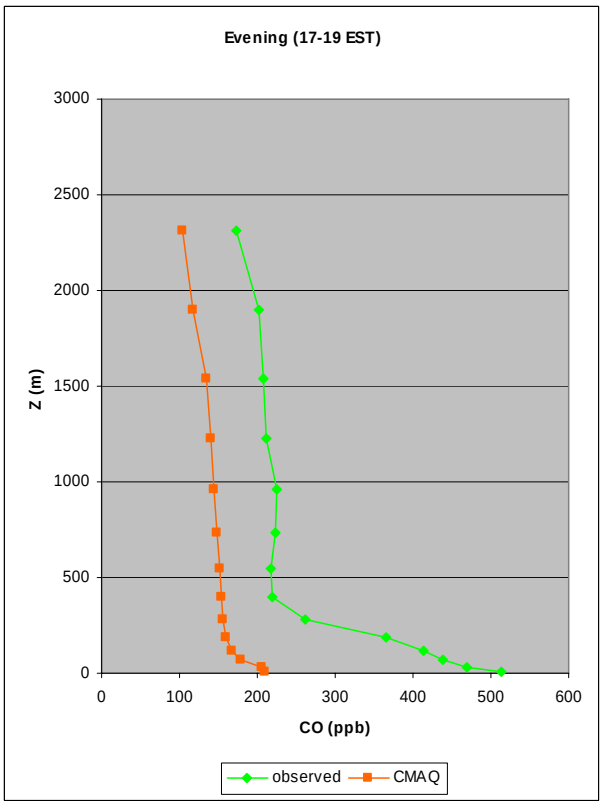


Figure 49.

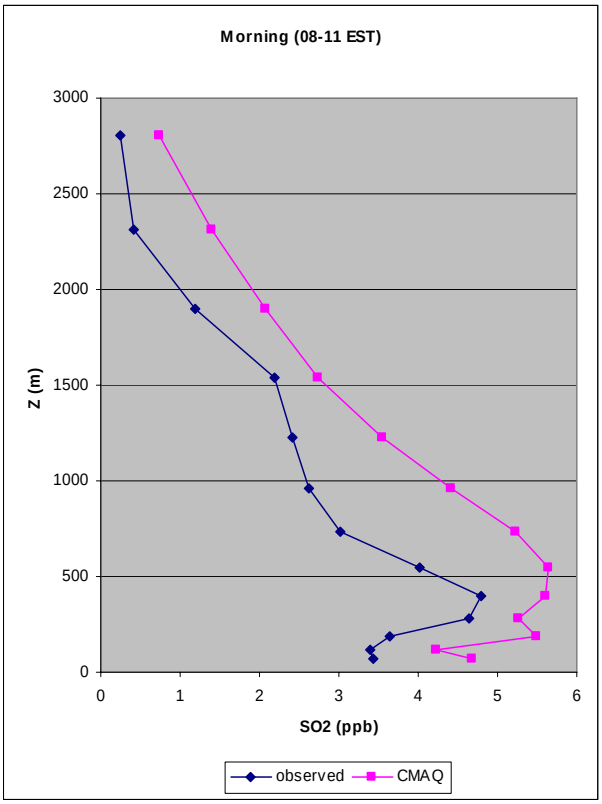



Figure 50.

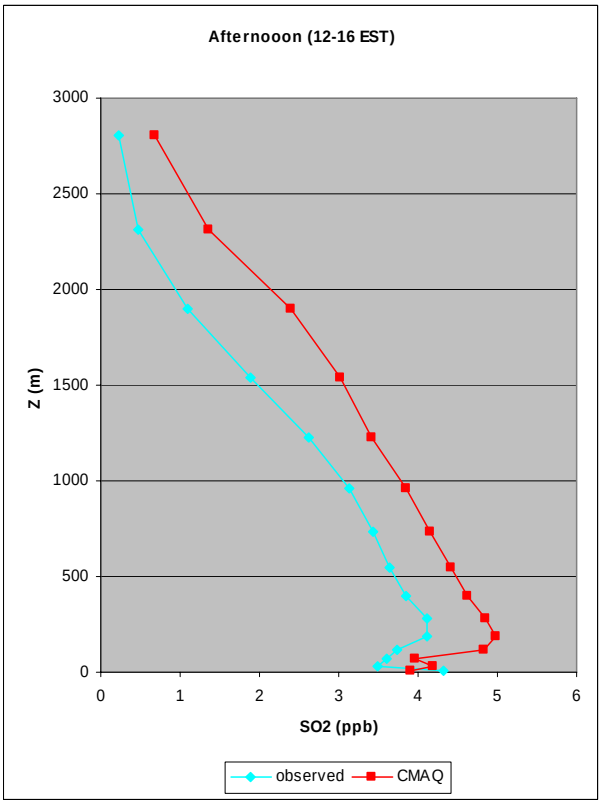
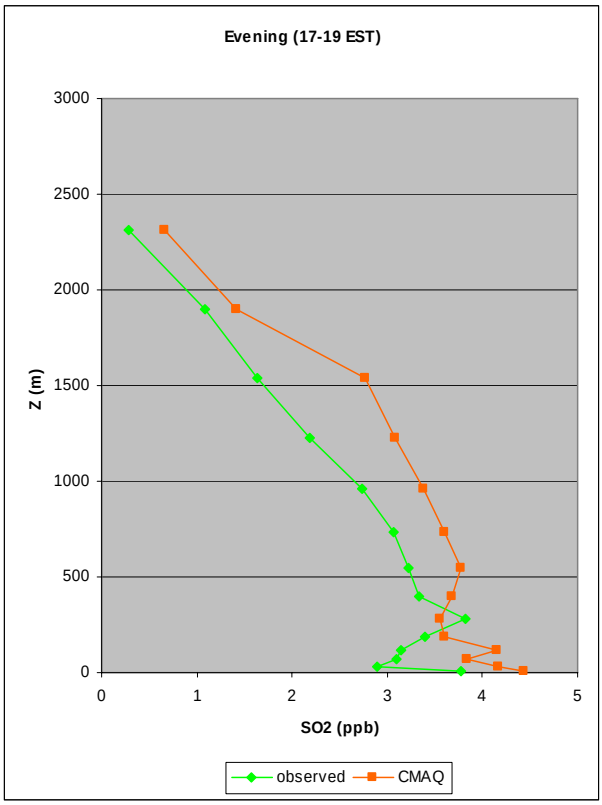


Figure 51.


Evening (17-19 EST)
3000
2500
2000
1500
1000
500
0
Z (m)
0
1
2
3
4
5
SO2 (ppb)
observed
CMAQ

THIS PAGE INTENTIONALLY BLANK