9. EVALUATION OF SOURCE ATTRIBUTION METHODS

A variety of tests were performed to evaluate the performance of many of the methods that were used for source attributions that are described in Chapter 10 and 11. (The methods themselves were described in Chapter 8.) Some of the meteorological and trajectory information that was used for those attribution analyses was also evaluated. These evaluations are described in this chapter.

Specifically, the following evaluations and sensitivity tests were carried out:

- Meteorological model predictions were compared with routine surface meteorological measurements, radar profiler measurements of conditions aloft, and satellite observations of cloud fields.
- Effects of different wind fields and trajectory calculation methods on trajectory calculations were assessed.
- Predictions of some trajectory-based methods were compared with perfluorocarbon and sulfate measurements at the surface.
- Taking advantage of the multiple attribution methods available in the BRAVO Study, the attribution capabilities of the Trajectory Mass Balance (TrMB) and Forward Mass Balance Regression (FMBR) methods were evaluated using the "artificial reality" represented by the concentration field generated by the REMSAD model and the emissions and boundary conditions inputs into it.
- Results of simulations by both the REMSAD and CMAQ regional air quality models were compared with surface perfluorocarbon tracer concentration measurements at and near Big Bend National Park.
- Results of simulations by both the REMSAD and CMAQ-MADRID regional air quality models were compared with ambient SO2 and sulfate concentration measurements at the surface throughout the study area.
- Finally, the performance of the Synthesized REMSAD and CMAQ approaches (described in Section 8.4.4) was evaluated by comparison with ambient sulfate concentrations measurements throughout the study area.

These evaluation and sensitivity tests and the insights they provide concerning the performance of the source attribution methods and their meteorological inputs are discussed in this chapter.

9.1 Evaluation of Simulations of Meteorological Fields

The source attribution methods used in the BRAVO Study (except for TAGIT) depend on meteorological information. All rely on wind speed and direction information, and the regional air quality models also rely on temperature, humidity, and liquid cloud water information. Accordingly, the first evaluations described in this chapter are for the simulations of meteorological variables.

The meteorological fields generated by the various meteorological models were evaluated in three ways:

- MM5 predictions of surface temperature, water vapor mixing ratio, wind speed, and wind direction were compared with routinely measured values
- MM5, EDAS and FNL predictions of winds aloft were compared with radar profiler measurements
- Predictions of precipitation and cloud fields, as developed by the REMSAD model from MM5 temperature and moisture mixing ratio predictions, were compared with surface precipitation measurements and satellite observations of cloud cover.

We discuss these evaluations below.

9.1.1 Evaluation of MM5 Surface Meteorology Simulations

The MM5 model output was evaluated by visual inspections of plotted fields for all predictive variables and selected diagnostic variables, plus extensive statistical evaluation of simulations against measured values of surface temperature, water vapor mixing ratio, wind speed and direction, and precipitation. The meteorological model was evaluated using hourly upper-air wind measurements from the 10 radar wind profilers in the BRAVO database. In addition, upper-level winds from 8 radiosondes were included, but were available only at 12-hr intervals. Hourly surface observations used in the evaluations consisted of standard data plus supplemental data in the BRAVO database for wind, temperature, and water vapor mixing ratio. The evaluation is described in the report by Seaman and Stauffer (2003), a copy of which is included in the Appendix. The main points of the evaluation are summarized below.

Visual inspections were performed on wind, temperature, and pressure fields to ensure that solution did not contain obviously unphysical features, as might occur if the specified lateral boundary files were inconsistent with internal model solutions or if clearly erroneous data had been assimilated. However, for this evaluation, emphasis was placed on statistics -- the root mean square error (RMSE), mean absolute error (MAE) and bias. Because the RMSE gives greatest weight to the largest errors ("outliers"), it is useful for identifying situations where the model may have intermittent large errors that could otherwise be masked by longer periods of good performance. The MAE gives the magnitude of the most typical error, without considering whether the error is positive or negative. Finally, the bias allows for cancellation of positive and negative errors, which may be quite acceptable in certain situations. For example, small bias errors in the wind speed and direction may under many circumstances indicate that long-term transport is reasonably accurate, even though the instantaneous winds may have considerably larger errors.

Determination of the "best" model performance requires evaluating the picture provided by a combination of performance statistics. Reliance solely on any one of these statistics is apt to give a distorted view of overall model performance.

In order to interpret the statistical results that were derived, it is useful to provide benchmarks that reflect the state of the art for meteorological model performance. ENVIRON Corp. has analyzed results from over 30 modeling studies to formulate typical standards for meteorological skill (Emery et al., 2001), which are summarized in Table 9-1. Because of limited data aloft, benchmarks have been developed only for the surface layer, but they will also be used here as initial indicators of performance aloft.

Table 9-1. Ad hoc benchmarks of surface-layer meteorological model accuracy for air-quality applications (from Emery et al., 2001).

	Temperature (C)	Mixing Ratio (g/kg)	Wind Speed (m/s)	Wind Dir. (deg.)
Benchmarks	MAE< 2.0	MAE< 2.0	RMSE < 2.0	MAE< 30.0
	|Bias| < 0.5	|Bias| < 1.0	|Bias| < 0.5	|Bias| < 10.0

The corresponding BRAVO model performance statistics for MM5 were calculated over the 36-km, 12-km and 4-km domains for each 5 1/2-day segment and were compiled in tables by segment, intensive periods, month and for the full four-month BRAVO study period (Seaman and Stauffer, 2003). That evaluation focused on predictions in the surface layer, but also addressed wind predictions aloft. Since the final modeling performed in the BRAVO Study used the 36-km meteorological fields exclusively, our discussion of the evaluation results will focus on the 36-km fields and will compare them with the 12-km fields. For consistency, results for both grid scales are averaged over the area of the 12-km modeling domain, which is where the additional information for evaluating model performance was most readily available. The average 36-km grid performance over the area of the larger 36-km modeling domain is not addressed here, but may be different from that over the 12-km domain.

Table 9-2 summarizes the performance of the MM5 predictions of the meteorological fields at the surface and for winds aloft. These conclusions apply for both the 12-km and 36-km grid scales over the geographic region encompassed by the 12-km modeling domain shown in Figure 8-1; the overall performance with the 36-km grid differed when evaluated over the entire 36-km domain. Values in red exceed the *ad hoc* performance benchmarks in Table 9-1. Note, though, that benchmarks are provided there for only two performance metrics for each variable.

Review of the information in Table 9-2 leads to the following major conclusions concerning the performance of the MM5 model over the area of the 12-km modeling domain:

1. In most cases, the 12-km fields have the lowest errors for wind and they contain sufficient detail to capture the regional flow. The 12-km biases in the wind speeds and directions for all layers and in all model run segments are small and fall within the ad hoc benchmark values for accuracy. The small values of MAE for wind direction and RMSE for wind speed in the nominal planetary boundary layer (30-1500 m AGL) and the lower free troposphere (1500-5000 m) are well

within the benchmarks for accuracy at all levels. The RMSE and MAE 12-km wind speed and direction statistics reveal that the instantaneous local wind errors remain quite modest over all averaging periods from 5 1/2 days to 4 months. This is a favorable result for modeling regional and inter-regional transport because the majority of trace gases and aerosols are transported in these layers.

Table 9-2. Performance metrics for the 12-km and 36-km grid MM5 simulations over the area of the 12-km modeling domain. The bold values are for the 4-month BRAVO Study period; the two numbers below them indicate the range of the four monthly values. The red values exceed the ad hoc benchmarks in Table 9-1.

	Elevation AGL, m	RMS Error		Mean Absolute Error		Bias	
		12 km	36 km	12 km	36 km	12 km	36 km
Temperature, °C	Surface (15 m)	**1.71** 1.57, 1.82	**1.50** 1.23, 1.66	**1.37** 1.24, 1.48	**1.26** 1.16, 1.34	**-0.49** -0.60, -0.31	**-0.53** -0.65, -0.46
Mixing Ratio, g/kg	Surface (15 m)	**1.62** 1.17, 2.18	**1.69** 1.12, 2.29	**1.22** 0.95, 1.56	**1.26** 0.92, 1.67	**0.24** -0.19, 0.42	**0.22** -0.14, 0.40
Wind Speed, m/s	Surface (15 m)	**1.48** 1.43, 1.53	**1.80** 1.77, 1.82	**1.22** 1.18, 1.27	**1.51** 1.45, 1.55	**0.38** 0.20, 0.49	**0.73** 0.53, 0.82
	30-1500	**1.36** 1.21, 1.44	**2.64** 2.58, 2.71	**1.14** 1.10, 1.17	**2.18** 2.14, 2.25	**-0.20** -0.28, -0.12	**0.17** -0.22, 0.37
	1500-5000	**1.33** 1.19, 1.44	**2.16** 1.89, 2.37	**1.04** 0.90, 1.17	**1.74** 1.52, 1.91	**-0.46** -0.51, -0.43	**-0.77** -0.91, -0.67
Wind Direction, degrees	Surface (15 m)	**32.3** 30.2, 35.6	**41.5** 39.5, 43.9	**22.1** 21.5, 24.8	**29.5** 27.8, 31.6	**-0.2** -2.2, +0.9	**-2.4** -4.0, -1.2
	30-1500	**21.0** 19.0, 23.7	**37.1** 32.0, 42.6	**14.9** 13.7, 17.1	**28.1** 23.8, 32.8	**0.0** -0.9, 0.8	**-3.1** -4.9, -0.3
	1500-5000	**16.2** 13.2, 19.3	**32.2** 25.5, 35.8	**11.1** 9.3, 13.0	**23.6** 18.2, 26.4	**0.2** 0.1, 0.5	**-0.7** -1.6, 0.5

2. The errors in the 36-km wind fields over the 12-km domain are substantially greater than the 12-km grid wind errors. Although the 36-km wind directions still lie within the *ad hoc* benchmark values for most layers and time segments, the errors in the wind speed above the surface layer are greater than the benchmarks in Table 9-1. In the surface layer, however, the 36-km RMS error for wind speed is within the benchmark range.

3. Errors in surface temperature and moisture in the 12-km and 36-km modeling are well within the *ad hoc* benchmark values for accuracy. For both temperature and moisture mixing ratio, the differences in performance at the two grid scales are insignificant. The 12-km fields have more fine scale detail than the coarser 36-km fields, of course, and the more spatially more accurate temperatures may be important for calculating chemical reaction rates.

Overall, then, except for wind speed aloft, the quality of the 36-km meteorological fields approximated the *ad hoc* benchmarks within the area of the 12-km domain, which means that the wind direction, temperature, and mixing ratio fields approximate the state of the art of meteorological modeling. The error in 36-km wind speed aloft was greater than its benchmarks and its effect on regional and interregional transport assessments must be considered. (See Section 9.1.2 below for additional evaluation of wind performance aloft.)

9.1.2 Comparison of MM5, EDAS, and FNL Wind Fields to Radar Wind Profiler Measurements

Wind fields generated by the MM5 and FNL models and the EDAS analysis system (see Section 8.2 for descriptions) were used with several source and receptor models applied in the BRAVO study. REMSAD and CMAQ modeling used MM5 fields. The receptor models used wind fields from either MM5 or a combination of EDAS (July-September) and FNL (October). For some receptor models, both MM5 and the EDAS/FNL combination were used, with somewhat differing attribution results. Here we compare the various wind fields and evaluate them against wind profiler measurements. A more comprehensive discussion of this comparison is provided in an appendix of the CIRA/NPS report on the BRAVO Study (Schichtel et al., 2004). The CIRA/NPS report is included in the Appendix of this report.

For this evaluation, wind measurements by radar wind profilers at Big Bend, Eagle Pass, Llano, and Brownsville were compared to wind predictions by MM5, EDAS, and FNL four times per day (6 am, noon, 6 pm, and midnight Central Standard Time). (See Figure 3-4 for map of profiler locations.) The radar wind profiler provided data for 60-m thick layers up to about 2000-2500 m AGL, then every 100 m up to about 3500-4000 m AGL. The single radar wind profile level closest to each model layer height was used for comparison. Since FNL has few layers in the lower atmosphere, the number of comparisons was less than for the other two methods.

Several metrics were used to compare model and radar wind profiler (RWP) winds:

- Average model wind and average RWP wind speed;
- Magnitude of vector difference between modeled and RWP winds;
- Average difference in wind direction (degrees);
- Average absolute value of wind direction difference;
- Percent of periods where model and RWP wind directions were within 20 degrees and 30 degrees; and
- Percent of model and RWP wind directions from 8 general directions (N, NE, E, SE, S, SW, W, NW) by layers -- 0-500 m, 500-1500 m, >1500 m AGL.

Results for some of these metrics are described here. The others are addressed in the CIRA/NPS report in the Appendix.

Because the EDAS fields were not available for October 1999, the summary of MM5 fields was done separately for the July- September and October periods. This facilitates the

comparison of the EDAS and MM5 performance measures for the same temporal periods. The performance of the MM5 fields can be compared to that of the FNL fields for the month of October.

Table 9-3. Comparison of modeled winds to radar wind profiler measurements for July-September 1999. Data are averaged over all vertical levels and over all observation times.

Location	Model	Average Model WS, m/s	Average RWP WS, m/s	Average WD Difference, deg.	% Within ±20 deg	% Within ±30 deg
Big Bend	MM5	5.53	5.57	-3.19	37	52
	EDAS	5.32	5.70	-6.97	33	47
	FNL	6.05	5.69	-6.33	25	37
Llano	MM5	6.16	5.87	-6.17	52	67
	EDAS	5.86	5.96	-3.19	53	67
	FNL	5.55	6.22	-4.38	34	49
Brownsville	MM5	7.16	7.04	-3.62	67	81
	EDAS	6.84	7.11	-2.80	67	79
	FNL	6.73	7.19	-4.58	45	60
Eagle Pass	MM5	6.75	7.63	-3.37	56	71
	EDAS	6.87	7.83	0.93	60	76
	FNL	6.47	7.97	4.09	44	61

Some summary statistics for the MM5, EDAS, and FNL comparison to the radar wind profilers for the period July-September 1999 are shown in Table 9-3. (Differences are modeled values minus measured values.)

Both MM5 and EDAS model average wind speeds were quite close to the RWP average wind speed, although there was a slight underestimation of wind speed by both models at Eagle Pass. The FNL predictions differed more from the RWP measurements than did the other models. A slight counterclockwise bias of a few degrees in wind direction was noted at all sites for both models (except EDAS and FNL at Eagle Pass). It will be shown below that much of the difference at Big Bend is at the lower levels, likely due to channeling of flow by local terrain features. The fraction of EDAS and MM5 model wind directions within ±20 degrees ranged from one-quarter at Big Bend for FNL to two-thirds at Brownsville. The fraction of model winds within ±30 degrees ranged from about three-eighths for FNL at Big Bend to about four-fifths at Brownsville for EDAS and MM5.

Over all, FNL performed worse than EDAS and MM5 for all criteria except average wind direction difference (bias), where it was about the same. Some of this difference may be due to comparing measurements at a single radar wind profiler level to calculated values

averaged over the comparatively very coarse FNL vertical layers. Perhaps averaging the RWP data within each FNL layer would give better results, but this was not tried.

Table 9-4 provides the same comparisons for October. (There are no EDAS entries for October because EDAS data were not available then.) Comparing Tables 9-3 and 9-4, it can be seen that MM5 performance in October was much worse than for the July to September period, while the FNL performance was slightly poorer in October than for July-September. MM5 bias in wind directions increased substantially, to 16-23 degrees counter-clockwise. The fraction of wind directions within ± 20 degrees and ± 30 degrees decreased substantially.

Table 9-4. Comparison of modeled winds to radar wind profiler measurements in October 1999. Data are averaged over all vertical levels and over all observation times.

Location	Model	Average Model WS, m/s	Average RWP WS, m/s	Average WD Difference, deg.	% Within ±20 deg	% Within ±30 deg
Big Bend	MM5	5.20	6.15	-21.30	22	31
	FNL	6.12	6.67	-12.00	22	32
Llano	MM5	6.07	7.00	-16.31	33	45
	FNL	6.46	7.73	2.22	36	46
Brownsville	MM5	6.77	6.47	-16.51	35	47
	FNL	6.59	6.68	6.99	31	43
Eagle Pass	MM5	6.12	7.60	-22.86	34	46
	FNL	6.86	8.28	4.55	33	42

Even though the FNL performance degraded during October, the FNL performed about as well as MM5 did in October. Wind speed differences were somewhat higher for FNL than MM5, but wind direction bias was less for FNL. The fractions of wind directions within 20 and 30 degrees were very similar for the models, with a slight edge to MM5.

The percentages of modeled and RWP wind directions from 8 general directions (N, NE, E, SE, S, SW, W, NW) were then compared by layers — 0-500 m, 500-1500 m, and >1500 m. The comparisons were done for MM5 and EDAS the period July-September 1999. No comparison was done for October because EDAS data do not exist for October.

Figure 9-1 portrays the comparison at Big Bend. The RWP winds at the lowest level are mainly from the south, while the models show winds predominantly from the SE and E. The low level winds at Big Bend are most likely due to channeling by local terrain; the modeled winds would not be expected to resolve this channeling. At the 500-1500 m level, there is better agreement as the RWP winds shift more toward the southeast and east. The best agreement is for the layer above 1500 m, although both models (especially MM5) are a bit low on the frequency of northerly winds (i.e., winds from the north) and high for easterly and southeasterly winds.

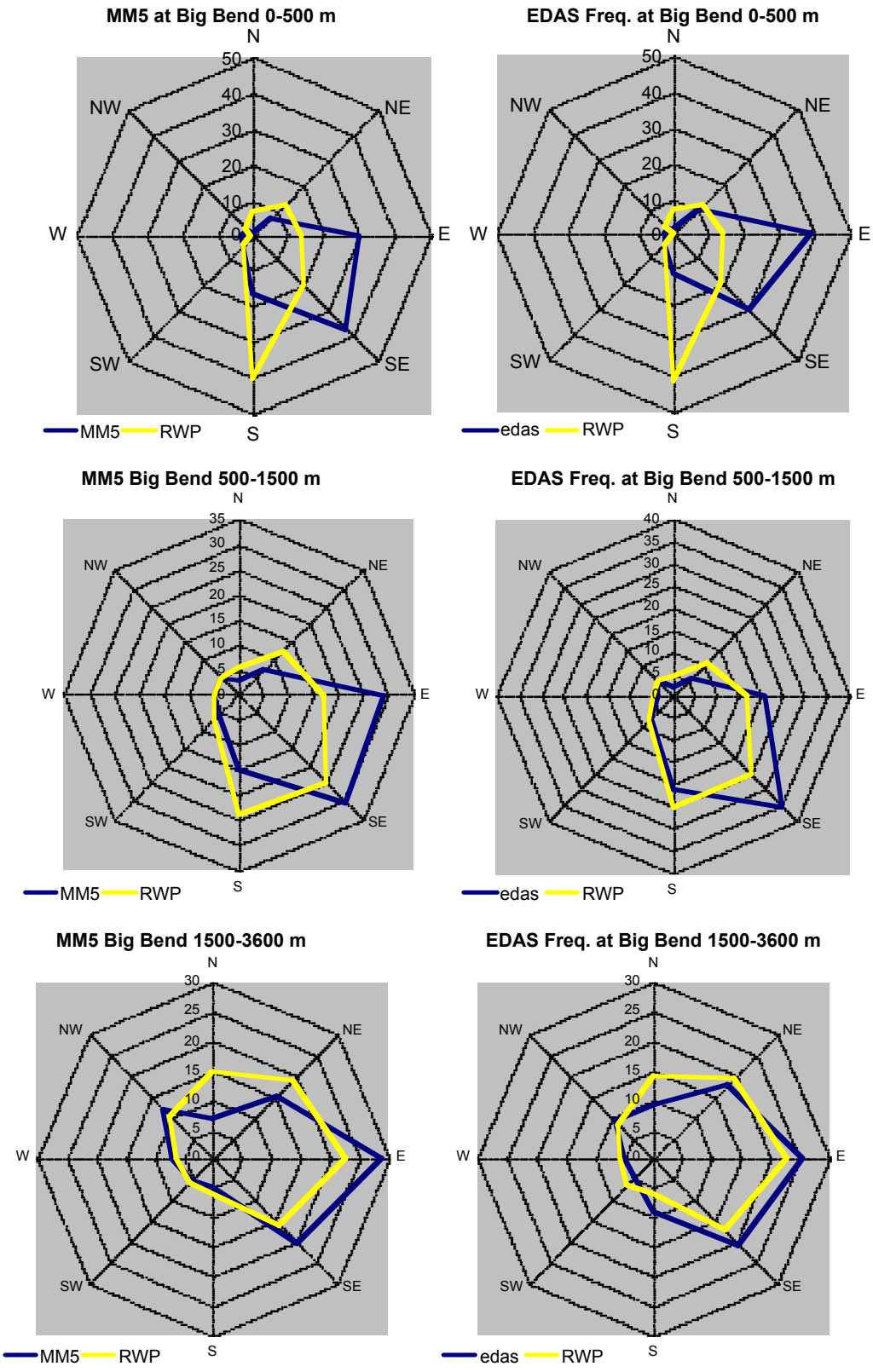

Figure 9-1. *Comparisons of MM5 and EDAS wind direction frequencies versus radar wind profiler measurements, by level above ground, for July-September 1999 at Big Bend National Park.*

For the Llano, Texas site, which does not have the complex terrain of the Big Bend area, the comparison of frequency of MM5 and EDAS winds to radar wind profiler winds is shown in Figure 9-2. In the lowest 500 m, the wind direction frequency compared well to the radar wind profiler for both models. For the 500-1500 m layer, the EDAS frequencies were quite similar to the RWP frequencies, while MM5 showed more E and SE winds and fewer S and SW winds than did the RWP. For the 1500-4000 m layer, both models differed substantially from the RWP, with the models indicating more easterly winds and fewer northerly winds.

Similar comparisons for Eagle Pass and Brownsville are shown in the CIRA/NPS report in the Appendix. Because the statistics in Table 9-4 showed MM5 performed more poorly in October than during the July-September period, similar plots that compare MM5 wind direction frequency distribution to RWP measurements for October are also included in the appendix. Although, the October MM5 frequency distribution is close to the observations, when averaged over all levels, there are substantial differences in the 0-500m layer.

In summary, MM5 and EDAS performance was similar for the July to September period. In October, MM5 performed similar to or slightly better than FNL, but its performance was much worse than during the July-September period. A clear preference for using the MM5 or EDAS/FNL combination of wind fields over the 4-month study period has not been established by this analysis.

9.1.3 Evaluation of REMSAD/MM5 Simulations of Precipitation and Clouds

The REMSAD and CMAQ-MADRID air quality models both estimated the particle sulfate and SO_2 concentrations for the BRAVO Study. Each of these models estimated the precipitation and constructed simulated cloud fields, based on the temperature and moisture mixing ratio information provided by the MM5 meteorological model. Simulation of precipitation is important because washout is the primary removal mechanism for particulate sulfate as well as an important removal mechanism for SO_2. Simulation of the existence of clouds and their locations are important to sulfate/SO_2 modeling because the aqueous phase chemical conversion of SO_2 to particulate sulfate occurs in cloud droplets is much faster than the dry chemistry that occurs when clouds are not present.

The following describes a semi-quantitative assessment of the ability of the REMSAD model to simulate precipitation and clouds. The initial purpose of the assessment was to see if the simulated cloud and precipitation fields were responsible for some of the poor performance at simulating sulfate during specific episodes, and is some cases that seemed to be the case. The work was expanded to survey all of the BRAVO Study period to assess the frequency of major discrepancies. This work was performed too late in the study to allow models to be revised and rerun, and more quantitative assessments were viewed as beyond the scope of the BRAVO Study. The focus was on the REMSAD/MM5 simulations; although cursory checks were made, a similar assessment was not performed for comparable CMAQ model simulations.

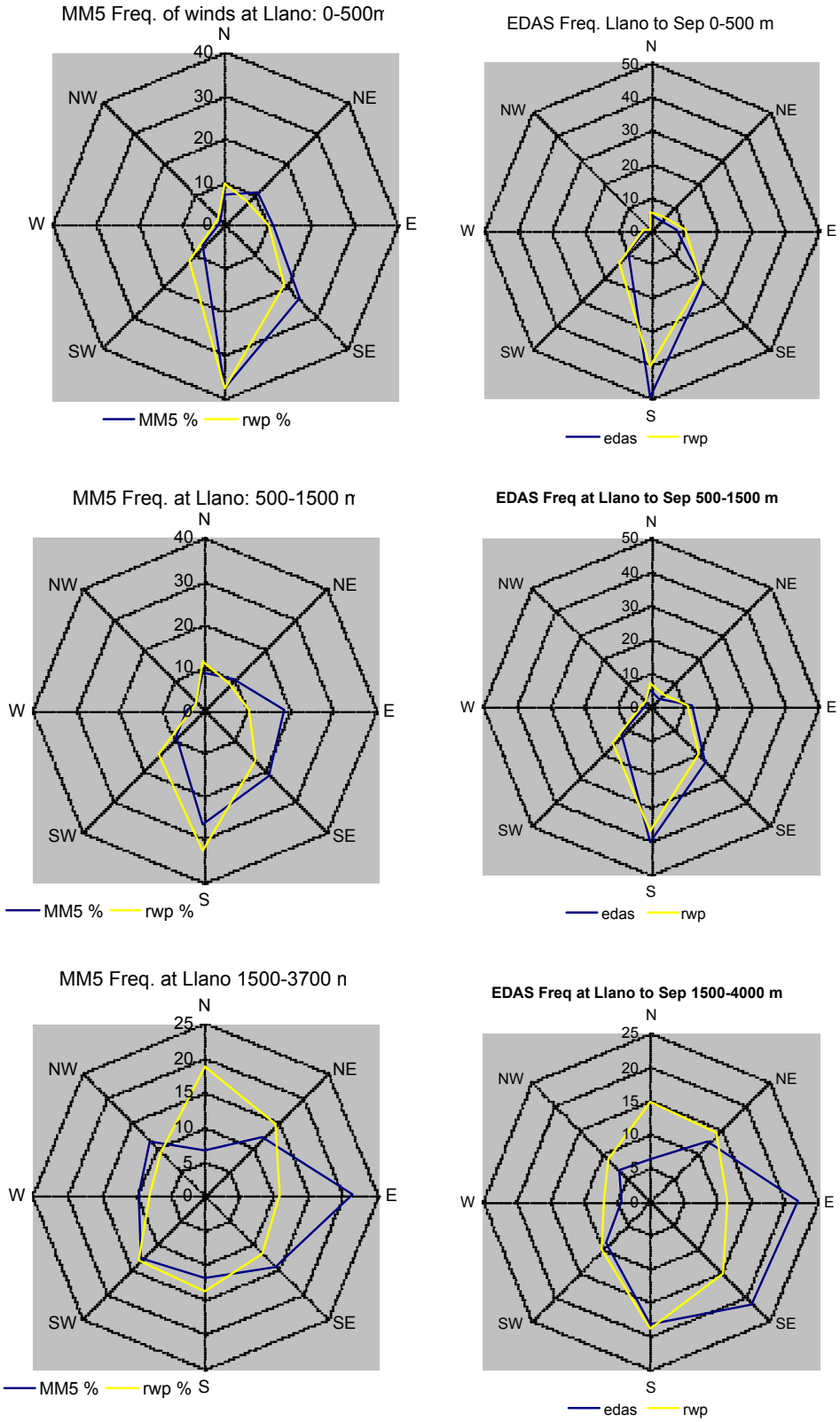

Figure 9-2. *Comparisons of MM5 and EDAS wind direction frequencies versus radar wind profiler measurements, by level above ground, for July-September 1999 at Llano, Texas.*

The National Weather Service publishes a weekly weather summary that includes maps for each 24-hour period that show areas that received precipitation. These were compared to corresponding maps showing the accumulated 24-hour precipitation predicted by the REMSAD model. Examples of the precipitation maps are shown in Figure 9-3. The comparison was limited to examining the correspondence of geographic patterns of predicted and observed precipitation. No attempt was made to compare precipitation quantities.

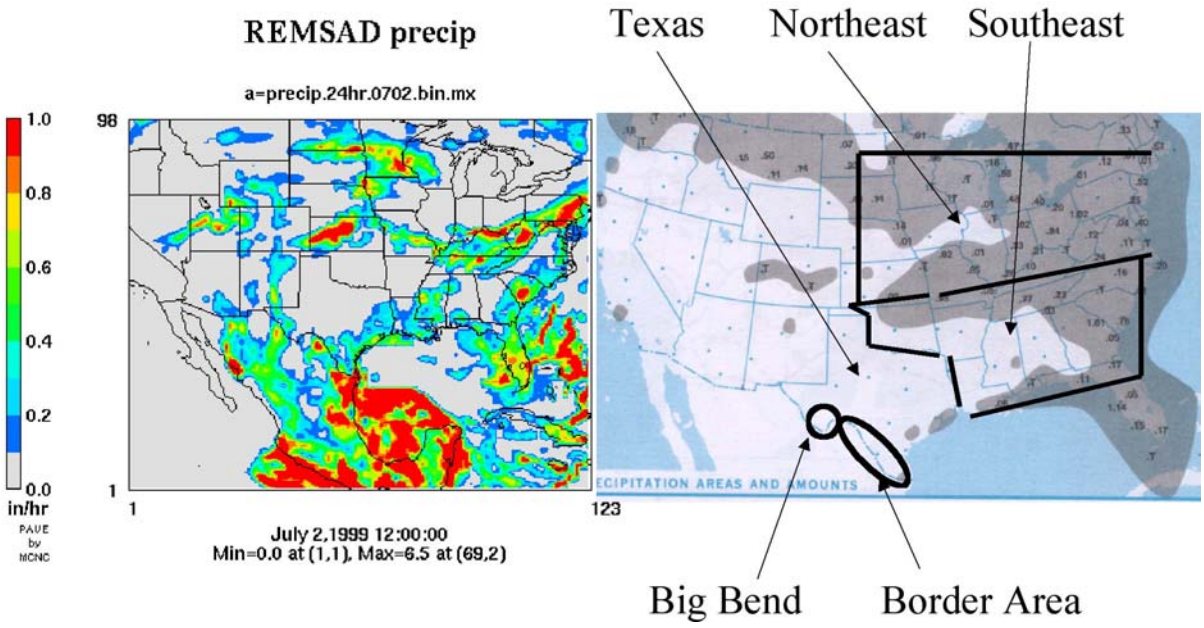

Figure 9-3. *Examples of REMSAD precipitation prediction and observed precipitation maps (for July 2, 1999). The right hand figure shows the five regions within which comparisons were made.*

Similarly, the 00Z and 12Z photos from the GOES satellite were compared with the corresponding REMSAD-predicted clouds. This approach cannot assess whether simulated clouds are at the correct heights, because cloud heights are not available in satellite photos. However, it can unambiguously identify incidents (times and locations) where the model is incorrect in its predictions, either predicting clouds when none are seen or predicting cloud-free sky where there are clouds. Examples of the cloud maps that were used are shown in Figure 9-4.

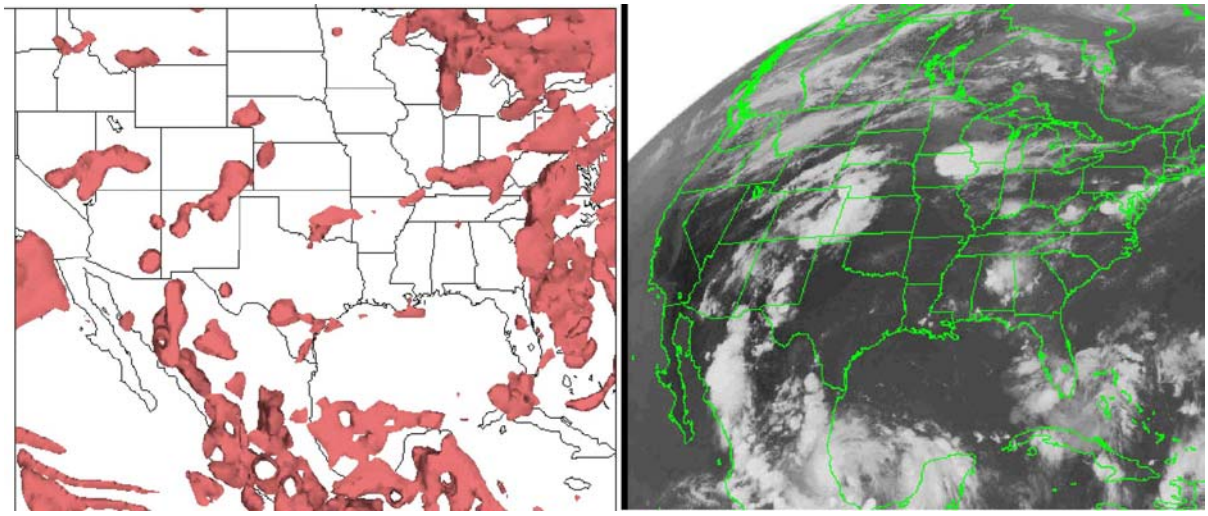

Figure 9-4. *Examples of REMSAD predicted cloud cover and corresponding GOES satellite picture (for July 3, 1999).*

For both the precipitation and cloud comparisons, five geographic areas were designated to be of particular interest, namely the area around Big Bend National Park, the Texas - Mexico border area, the State of Texas, the southeastern U.S., and the northeastern U.S., as indicated in Figure 9-3. The maps of both modeled and measured precipitation and clouds in each region were examined to qualitatively assess the fraction of the area of each region that was covered with precipitation or clouds. Results were assigned scores using a quintile grading system in which none is assigned the value of 0; 1% to 20% = 1; 20% to 40% = 2; 40% to 60% = 3; 60% to 80% = 4; and 80% to 100% = 5, and the model and measured scores were compared to assess model performance. An example of the precipitation scoring, based on the maps in Figure 9-3, is given in Table 9-5.

Table 9-5. Example of the qualitative ratings of predicted and observed precipitation for the five regions, for the maps shown in Figure 9-3. The scoring scheme is 0 = no precipitation, 1 = 1% to 20%, 2 = 20% to 40%, 3 = 40% to 60%, 4 = 60% to 80%, and 5 = 80% to 100% of the region showing precipitation.

	Predicted Score	Observations Score
Big Bend	2	0
Border Area	4	0
Texas	1	1
Southeastern US	3	2
Northeastern US	2	4

The assessment was conducted for every day of the BRAVO Study and the results were compiled as differences between the ratings for the predicted minus the observed ratings for each of the five regions.

The main conclusions of this precipitation comparison are as follows:

- REMSAD generally overestimates precipitation near Big Bend especially in July, August & September
- REMSAD generally overestimates precipitation along the Rio Grande border area between Big Bend and the Gulf Coast especially in July and August
- REMSAD tends to underestimate precipitation in the Northeast
- No bias is evident in the precipitation estimates for Texas and the Southeast.

The cloud cover comparisons produced the following principal conclusions:

- REMSAD cloud estimates for Big Bend are frequency wrong, but are unbiased over the duration of the study
- REMSAD tends to overestimate presence of clouds in the border region.
- REMSAD predictions of clouds over Texas are generally unbiased.

These results indicate limitations on the ability of the REMSAD and MM5 combination to faithfully reproduce the meteorological conditions that are influential in determining particulate sulfate and SO_2 concentration fields. The overestimation of precipitation near Big Bend and along the U.S.–Mexico border region in the first three months of the study may have been a contributor to the underestimation of particulate sulfate concentrations at and near Big Bend by both REMSAD and CMAQ-MADRID, as described later in Sections 9.9 and 9.11.

9.2 Sensitivity of Trajectories to Wind Fields

A comprehensive evaluation of the effects of wind fields on back trajectories was carried out in the BRAVO Study and is described in the CIRA/NPS report (Schichtel et al., 2004), which is included in the Appendix

The goals of the evaluation of wind fields and back trajectory models were to determine:

- Whether there are biases between the various available back trajectory models and available input meteorological data;
- Whether past studies using the older Atmospheric Transport and Diffusion (ATAD) model with the usual rawinsonde data as input are valid in the light of new models and new gridded wind fields;
- Whether the unavailability of the EDAS gridded wind field during October caused any biases during that month;

- Whether any findings of biases in trajectory models and wind fields are valid for Big Bend during summer and fall 1999 only (the BRAVO study period) or whether similar biases exist for other regions of the country and for other seasons and other years in south Texas.

The models examined were the Atmospheric Transport and Dispersion (ATAD), HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) ver. 4.5, and CAPITA Monte Carlo models. Input meteorological data included rawinsonde data, rawinsonde data plus the four BRAVO wind profilers, and three modeled gridded wind fields (EDAS, FNL, and MM5). Details of the input meteorology are discussed in Section 8.2 and details of the three back-trajectory models are given in Section 8.3.

Big Bend is a particularly challenging receptor site at which to generate back trajectories due to its proximity to both complex terrain and to data-sparse areas. Analyses of back trajectories by several methods, including plotting trajectories by all methods for each day and examination of the residence time probability density functions, showed that there are some systematic differences between the results of different back trajectory model/input data combinations at Big Bend National Park during BRAVO.

9.2.1 Effect of Wind Field Choice

Most of the differences appear to be due to the choice of input wind field rather than the choice of back trajectory model, although there are also some differences due to model alone. Using only rawinsonde data (ATAD model only) gives trajectories arriving at Big Bend from the most southerly direction, EDAS data results in the most easterly trajectories, and those generated using MM5 and FNL data fall between these extremes. These general directional biases are consistent across models and are more pronounced during summer than fall. The BRAVO tracer data clearly indicate that ATAD with rawinsonde data alone results in trajectories that are too southerly, especially during the summer. This is most likely due to wind sounding data scarcity in the region. Supplementing the rawinsonde data with information from the four BRAVO wind profilers causes the ATAD/rawinsonde trajectories to be more similar to those generated using the gridded wind fields. The MM5 data have a higher mean wind speed than EDAS/FNL and will therefore be somewhat more likely to attribute concentrations to source areas farther away from Big Bend than will the EDAS/FNL wind fields.

Figure 9-5 illustrates the effect of choice of wind field on the back trajectories generated by one model, the CAPITA Monte-Carlo model. Each line represents the path of one of the particles that is followed by the model. The left panel shows that on September 15th the trajectories generated using the EDAS and MM5 wind fields are similar, although the previously-mentioned greater reach of the MM5 trajectories is apparent. The lower panel below the map shows the height histories of the particle trajectories, starting with Time = 0 at Big Bend, and reveals that the MM5 wind field places the particles at higher altitudes, where the wind speeds tend to be greater.

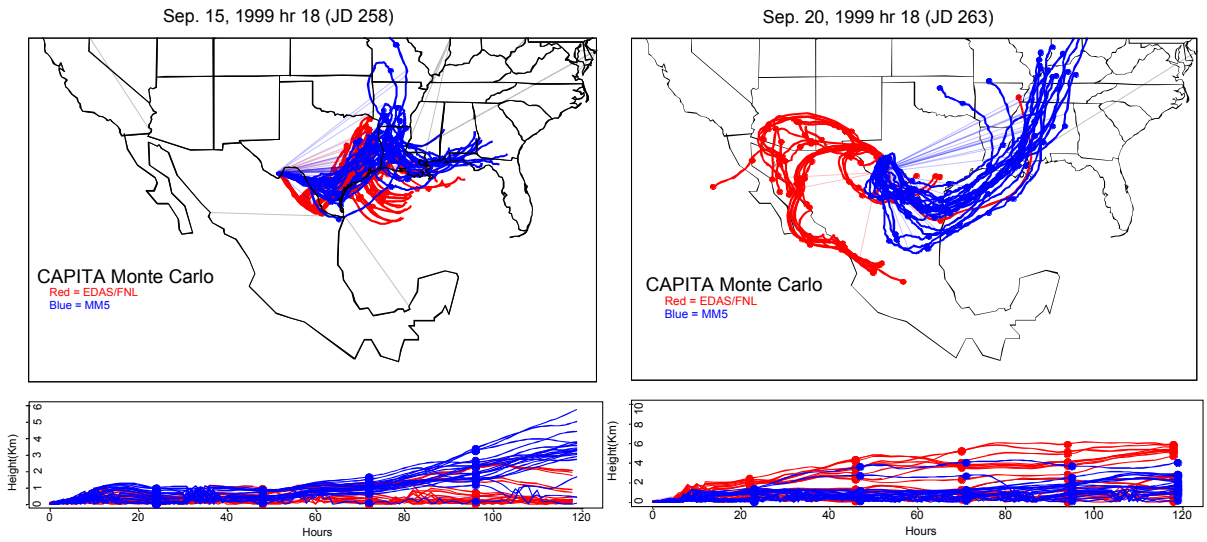

Figure 9-5. *Examples of the effects of EDAS and MM5 wind fields on back trajectories produced by the CAPITA Monte Carlo model. The lower panels are side views of the back trajectories, with Big Bend National Park at Time = 0.*

On September 20th, however, the trajectories based on the two wind models are quite different, as shown in the right panel of Figure 9-5. Here, almost all of the EDAS back trajectories lie to the west of Big Bend and all of the MM5 back trajectories lie to the east of Big Bend. This time the bottom panel shows that the EDAS wind field lofted the particles to higher altitude where, because of wind shear on this day, most of those Monte-Carlo particles were west of Big Bend.

Over the duration of the study, agreement between wind fields was generally more like that represented in the left panel, with rough agreement between trajectories, but situations such as shown in the right panel point out the need for critical analysis when interpreting any trajectory results based on a limited number of days.

9.2.2 Effect of Trajectory Model Choice

There are also some differences between results that can be attributed to the choice of model. In general the ATAD and Monte Carlo models give results more similar to each other than to HYSPLIT. This is primarily because both ATAD and Monte Carlo, though different in mechanism, average the horizontal transport within the mixed layer. The predominant wind direction calculated by HYSPLIT is less southerly and more along the Texas-Mexico border than that calculated by ATAD and Monte Carlo using the same input wind fields. This is true with either EDAS/FNL or MM5 input.

HYSPLIT, which uses the gridded wind fields without the averaging of ATAD and without the random vertical movements of the Monte Carlo model, is more sensitive to start location, start height, and placement of individual endpoints. This is because slight changes in position can affect the grid cell and layer for which data is used for the next time step. Slight differences in trajectory height can result in large differences in horizontal placement

of trajectories in HYSPLIT during time periods when there is significant directional wind shear with height. Trajectories produced by the Monte Carlo model can also be somewhat sensitive to height if a particle moves from within the mixed layer to above. Average differences caused by occasional bumping of a trajectory from one grid cell and/or horizontal layer to a neighboring one will be minimized as more trajectories are aggregated. Thus, longer analysis periods, more start heights and times are all desirable. Both HYSPLIT and the Monte Carlo model allow trajectories to get as high as 10 km. Due to its different design, ATAD does not have trajectories above the mixed layer, with a maximum of 3 km.

9.2.3 Effect of Back Trajectory Start Height

Figure 9-6 illustrates the effect of start height (i.e., the chosen trajectory height at Big Bend), which affects the mean trajectory height and hence the mean trajectory speed and trajectory direction, for back trajectories in HYSPLIT. The higher speeds typically associated with greater heights will tend to implicate more distant source areas, while lower speeds and heights give more weight to nearer sources, but this generalization may not always hold. For example, on July 15th (the left panel of Figure 9-6), trajectories group according to start height. Back trajectories from 10, 100, and 200 m all follow the same general path into the Gulf of Mexico, with the effect of increasing wind speed with height reflected in the trajectory lengths. Trajectories from 500 and 1000 m are farther south and pass over the Yucatan Peninsula. The 200-m trajectory starts out like the lowest trajectories but has a completely different looping path in the last 30 hours. On September 20th (the right panel), on the other hand, the low altitude back trajectories all follow the same path to the

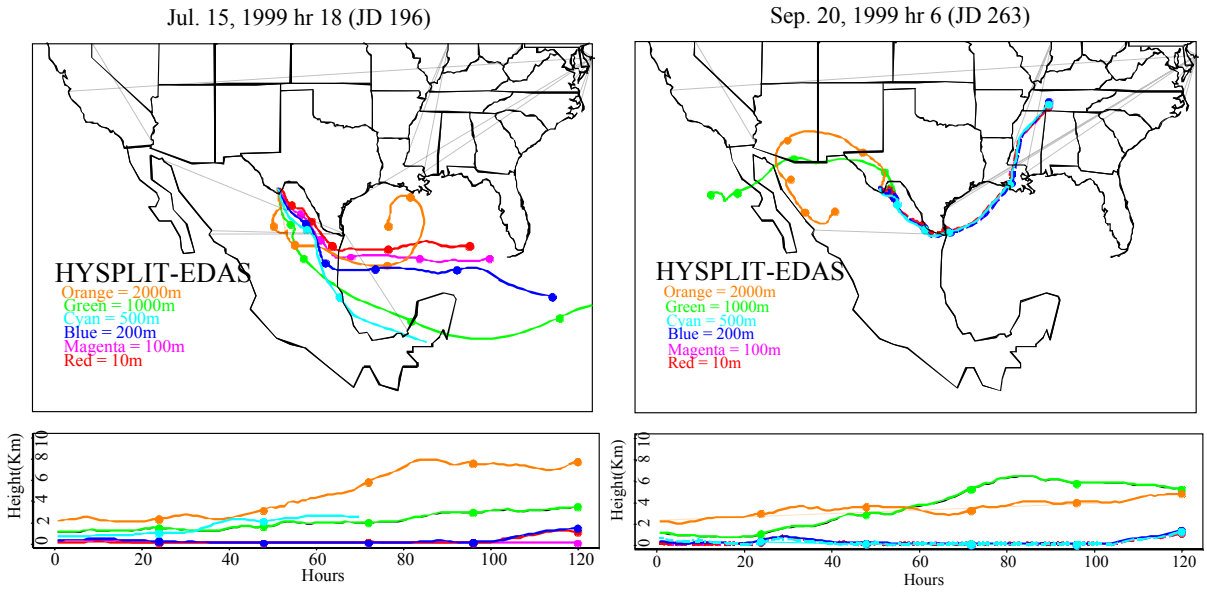

***Figure 9-6**. Examples of the effect of start height on the trajectories produced by the HYSPLIT model with the EDAS wind field.*

east while the higher trajectories are to the west, again illustrating the same wind shear that was implicated in the right panel of Figure 9-5. (Note that the September 20th back trajectories in Figure 9-6 start 12 hours earlier than those on the same date in Figure 9-5.)

In general, the lowest start heights result in the fewest endpoints in the Western U.S. (i.e., west of Big Bend) during BRAVO. Also for EDAS/FNL input, lower start heights result in fewer endpoints in areas of Mexico to the west and southwest of Big Bend. The fractions of endpoints in these areas increase as the start heights increase.

9.2.4 Effect of Using FNL Winds in October

Since EDAS data were unavailable during October 1999, the consequences of having to substitute the lower-resolution FNL data during this month were investigated. During five other Octobers, using FNL data rather than EDAS resulted in fewer air masses arriving from west and southwest Mexico and the southwestern United States, and more air masses arriving from eastern Texas and from that part of Mexico just south of the Texas border. During these five Octobers, HYSPLIT trajectories generated using FNL data were on average, much lower than those generated with EDAS data and consequently had lower average wind speeds.

Cursory investigation of trajectory differences for other regions and other seasons indicates that the magnitude of the directional differences between the ATAD (rawinsonde), EDAS and FNL wind fields is both geographically and seasonally dependent. So, though there are also likely to be biases in other regions, the directions of these biases cannot necessarily be inferred from the BRAVO data.

When evaluated against the tracer concentration data, it was found that the 36-km MM5 and EDAS/FNL wind fields had approximately equal skill at predicting the tracer release locations.

Overall, the above analyses provided comparable confidence in the both the MM5 and EDAS/FNL wind fields, with no clear indication that either was superior to the other, and therefore both data sets were used for further trajectory-based source attribution analyses.

9.3 Evaluation of the Trajectory Mass Balance (TrMB) Method using Perfluorocarbon Tracer Measurements

The performance of the various trajectory-based source attribution methods described in Section 8.3 was evaluated using the data provided by the perfluorocarbon tracer measurements described in Section 3.2. Since the tracer release locations and source strengths were known, the tracer data set provided a solid basis for evaluating the source attribution capabilities of these methods.

The evaluation of the TrMB method is described in this section. Sections 9.4 and 9.5 describe tracer-based evaluations for the CAPITA Monte Carlo model and the FMBR approach, respectively. Further evaluation of the TrMB and FMBR methods was carried out using the REMSAD description of the sulfur concentration field as a “synthetic reality” in

which receptor concentrations and source locations and strengths were again known. These evaluations are described in Sections 9.6 and 9.7.

To test the performance of the Trajectory Mass Balance (TrMB) approach to source attribution (described in Section 8.3.5), it was exercised using the total perfluorocarbon tracer concentration at Big Bend as the dependent variable and back trajectory endpoints from the four tracer release sites as the independent variables. The results were compared with the known tracer emissions to see how well the method attributed the tracer concentrations at Big Bend to the four tracer release locations. The analysis is described in detail in the CIRA/NPS report (Schichtel et al., 2004), which is included in the Appendix.

The total tracer contribution by each source was calculated by summing the 24-hr average concentrations of that source's perfluorocarbon tracer at the K-Bar site in Big Bend National Park for each day during the second half of the study (when the four tracers were released from four different sites). The four tracer release locations were shown in Figure 3-2 and the source areas used to represent each for the TrMB evaluation are shown in Figure 9-7.

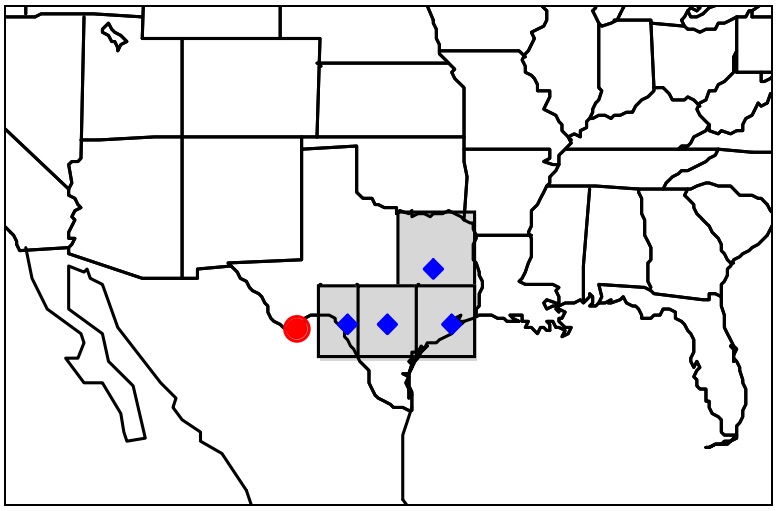

Figure 9-7. *The tracer release regions (shaded boxes) defined for the TrMB evaluation. Blue diamonds represent the actual tracer release locations and the Big Bend National Park location is at the red circle.*

A challenge with using tracer concentrations in TrMB is how to handle the many reported concentration values that were negative. The most severe problem is with iPPCH, the tracer that was released in northeastern Texas (the northernmost release location in Figure 9-7), whose mean reported concentration at K-Bar was negative. After some experimentation, it was decided to (1) not use the northeast Texas tracer data because the concentration of the iPPCH tracer was generally so low as to be mostly undetectable at K-Bar, and (2) set any negative 24-hr tracer component concentration values elsewhere to zero before summing them to calculate the total tracer concentration for each day.

Another option considered for making the tracer data more robust was to average the measured concentrations at K-Bar, Persimmon Gap, and San Vicente, three sites all within or near the boundaries of Big Bend National Park. The model results and performance did not

change substantially when data from one or both of the other sites were averaged with the K-Bar data, so the analysis was continued using data from K-Bar only.

The time period evaluated is September 17 through October 28. Releases of PDCB and PTCH were begun at San Antonio and Houston on Sept. 17. Tracer release of PTCH from Houston was stopped on Oct. 25. Emissions for the other three sites continued until Nov. 1, but release of ocPDCH from Eagle Pass became erratic after Oct. 28. A constant release of tracers throughout this time period would have been ideal for TrMB analysis. This was not the case during the BRAVO Study, which will contribute to some error in TrMB results.

Table 9-6 is a summary of the results of this analysis. TrMB applied with several trajectory model and wind field combinations is able to reproduce the known attributions of all three tracers (given in the top part of the table) to within the error in the measured concentration and the standard errors of the regression coefficients. The combinations

Table 9-6. Results of TrMB attribution of the tracer material that arrived at K-Bar from 9/17 to 10/28/99. The row above the double line gives the percent of the total measured concentration due to each tracer. The remaining rows give the modeled percent attributions for several combinations of back-trajectory model, input meteorological data, and trajectory lengths. Attributions that are accurate within the uncertainty of the measurement and standard error of the regression coefficients are shown in bold red type.

	ocPDCH Eagle Pass	PTCH San Antonio	PDCB Houston	R^2
Tracer Measurements				
Mean Concentration (ppq)	6.52±0.99	23.47±3.39	2.62±0.33	
Mean percentage of total tracer	20±4	72±13	8±1	
Attributions of Tracer Material Arriving at K-Bar (%)				
ATAD Raw 5-day	**35 ± 12**	**65 ± 14**	**0 ± 9**	0.495
ATAD EDAS/FNL 5-day	**16 ± 8**	33 ± 9	51 ± 9	0.708
HYSPLIT EDAS/FNL 5-day	**28 ± 12**	**67 ± 13**	**5 ± 9**	0.640
7-day	**29 ± 13**	**68 ± 15**	**3 ± 11**	0.603
10-day	**27 ± 13**	**73 ± 16**	**0 ± 11**	0.612
Monte Carlo EDAS/FNL 5-day	**30 ± 9**	43 ± 10	27 ± 8	0.721
7-day	**33 ± 10**	**50 ± 10**	18 ± 8	0.643
10-day	**30 ± 9**	39 ± 10	31 ± 8	0.689
ATAD MM5 5-day	**34 ± 12**	**60 ± 13**	**6 ± 9**	0.564
HYSPLIT MM5 5-day	82 ± 18	18 ± 18	**0 ± 11**	0.484
7-day	81 ± 18	19 ± 19	**0 ± 11**	0.489
10-day	81 ± 18	19 ± 18	**0 ± 12**	0.502
Monte Carlo MM5 5-day	**23 ± 16**	**77 ± 19**	**0 ± 12**	0.643
7-day	**26 ± 17**	**74 ± 21**	**0 ± 13**	0.616
10-day	**24 ± 17**	**76 ± 21**	**0 ± 14**	0.626

include HYSPLIT with EDAS/FNL input, the CAPITA Monte Carlo model with MM5, ATAD with MM5, and ATAD with raw sounding data. In general, choice of 5, 7, or 10-day back trajectories made little difference for this evaluation and the results were usually the same within the standard error of the regression coefficients, irrespective of the trajectory length. (Note that the results given here depend somewhat on the geometries of the tracer release regions shown in Figure 9-7. Different boundaries for the regions surrounding each tracer source may produce different perceptions of performance.)

The worst performance is by HYSPLIT with MM5 input, a combination that dramatically overstates the Eagle Pass tracer contribution and understates that from San Antonio. Eagle Pass is approximately 250 km from K-Bar, while San Antonio is almost twice as distant at approximately 450 km. This may be an indication that HYSPLIT with MM5 has too many back-trajectory endpoints close to the receptor at the expense of too few farther away.

Other problem combinations were ATAD with EDAS/FNL input and the Monte Carlo model with EDAS/FNL input, both of which overestimated tracer arriving from Houston and underestimated the contribution from San Antonio. Houston is the most distant of the three modeled release sites at approximately 750 km from K-Bar, so these latter combinations are overestimating the most distance source area.

Results of this evaluation suggest that the best model/wind field combination for tracer attributions, and thus for accurate back trajectory placement within south Texas, is HYSPLIT with EDAS/FNL input. The combinations of CAPITA Monte Carlo or ATAD with MM5 input and ATAD with raw sounding input also were able to re-create the known tracer attributions. Because HYSPLIT with MM5 input performed poorly in both the tracer test described in this section and the simulated sulfate tests described in Section 9.6, this combination was not used for BRAVO sulfate source attribution. Attribution modeling using back trajectories from either the Monte Carlo or ATAD models with EDAS/FNL input is also suspect.

9.4 Evaluation of the CAPITA Monte Carlo Model Using Perfluorocarbon Tracer Measurements

The CAPITA Monte Carlo particle dispersion model was tested against the perfluorocarbon tracer measurements, in order to evaluate that model's ability to simulate near field and synoptic scale transport and to further explore the effects of different wind fields. The approach used for this evaluation and the results it produced are described here; details can be found in the CIRA/NPS BRAVO report (Schichtel et al., 2004), which is included in the Appendix.

As described in Section 8.4.1, the CAPITA Monte Carlo model is a long-range transport model that simulates air mass transport and diffusion by tracking the movement of multiple particles released from a source. A modeled wind field is used to advect the particles in three-dimensional space. The vertical mixing that takes place within the atmospheric boundary layer is simulated using a Monte Carlo technique that evenly distributes the particles between the surface and the mixing height.

For this evaluation, the Monte Carlo model was driven by two sets of wind fields, the 36 km MM5 winds developed for BRAVO (described in Section 8.2.3) and a combination of data from the National Centers for Environmental Prediction (NCEP) Eta Data Assimilation System (EDAS; see Section 8.2.1) and Global Data Assimilation System (GDAS; see Section 8.2.2). The 36 km wind fields of MM5 and EDAS data at 80-km spacing were used whenever available, except that FNL data on a 190 km grid were used instead of EDAS data in October, when the EDAS data were not available.

For these Monte Carlo simulations, 100 particles were released every hour from each of the four tracer release locations at a fixed effective stack height. The particles were tracked for five days or until they left the grid. Simulated tracer concentration fields were generated every hour by weighting each particle by the actual tracer emission rate for its source and summing the weights of all particles that fell in a given 36 km grid over most of North America.

The observed and modeled tracer data were compared at Big Bend by averaging the results for the three Big Bend sites (San Vicente, K-Bar and Persimmon Gap) and over southwest Texas (the average of the six 6-hr tracer sites extending from San Vicente north 265 km to Monahans Sandhills). (See Figure 3-2 for site locations.) The Big Bend sites were aggregated together since the wind fields cannot resolve transport to individual sites 30 km apart. Aggregating the six sites allows us to test whether the model transports the tracer to the vicinity of the monitoring sites, which is a less stringent and more appropriate test for the Monte Carlo model and its intended use.

Figures 9-8 and 9-9 display the day-to-day measured and modeled tracer concentrations, for both the MM5 and EDAS/FNL wind fields, over the duration of the BRAVO Study. Each figure includes three plots — the top one is for the tracer released at Eagle Pass, the one on the lower left is for the tracer released at San Antonio, and the one on the lower right is for the tracer released at Houston (the Parish power plant). Error bars on the data points indicate estimates of the uncertainties in the measured concentrations. As described in Section 3.2, the Eagle Pass tracer release took place over the entire 4-month study period, while the releases at San Antonio and Houston lasted from the middle of September to the end of the study. As discussed in Section 5.4, the quality of the measurements of the tracer released in northeastern Texas was poor except for occasional episodes and so results based on that tracer are not shown here. Statistics concerning the data in these figures are given in Table 9-7. Figures 9-8 and 9-9 show that Monte Carlo simulations using either the MM5 or EDAS/FNL wind fields were able to reproduce the major features of the timing of the largest Eagle Pass (ocPDCH) tracer impacts at both sets of sampling sites. The durations of the simulated tracer impacts tended to be comparable to, or shorter than, those measured. The simulation based on MM5 meteorology was better able to simulate the Eagle Pass tracer pattern, particularly from August 10 to September 12. Neither simulation was able to reproduce the day-to-day variability of the observed concentrations of Eagle Pass tracer with great skill, with r^2 between 0.18 and 0.31, although many estimates were within the ranges of uncertainty of the corresponding measurements. Also, as indicated in Table 9-7, the EDAS simulation underestimated the Eagle Pass tracer concentration by about a factor of two at Big Bend while the MM5 simulation underestimated the average concentration by only 5%.

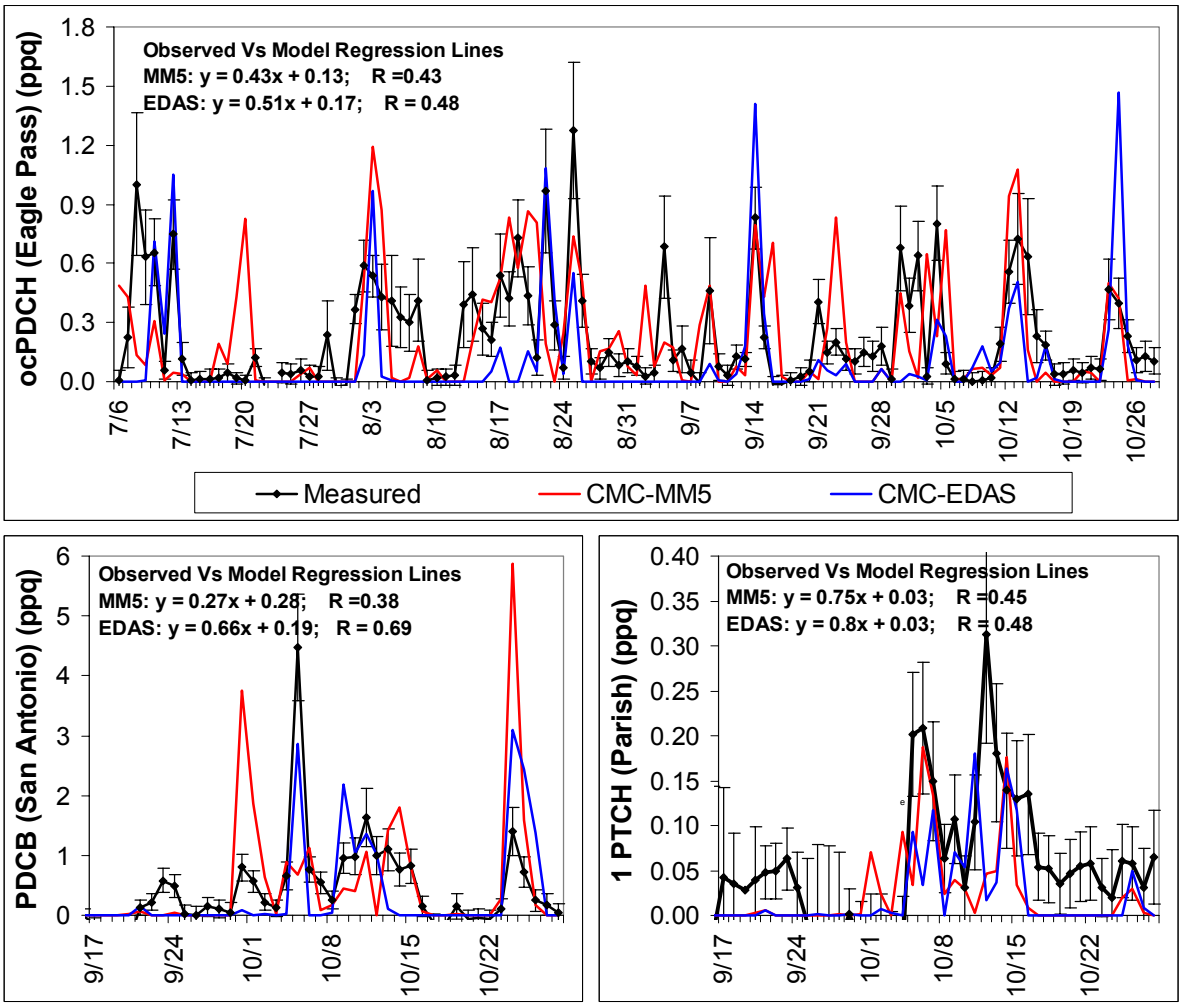

Figure 9-8. *Comparison of observed and modeled tracer concentrations, for the Eagle Pass, San Antonio, and Houston tracers, at Big Bend National Park over the duration of the study. The measured and modeled tracer concentrations were both averaged over the K-Bar, Persimmon Gap and San Vicente monitoring sites. Error bars represent the uncertainties of the measurements.*

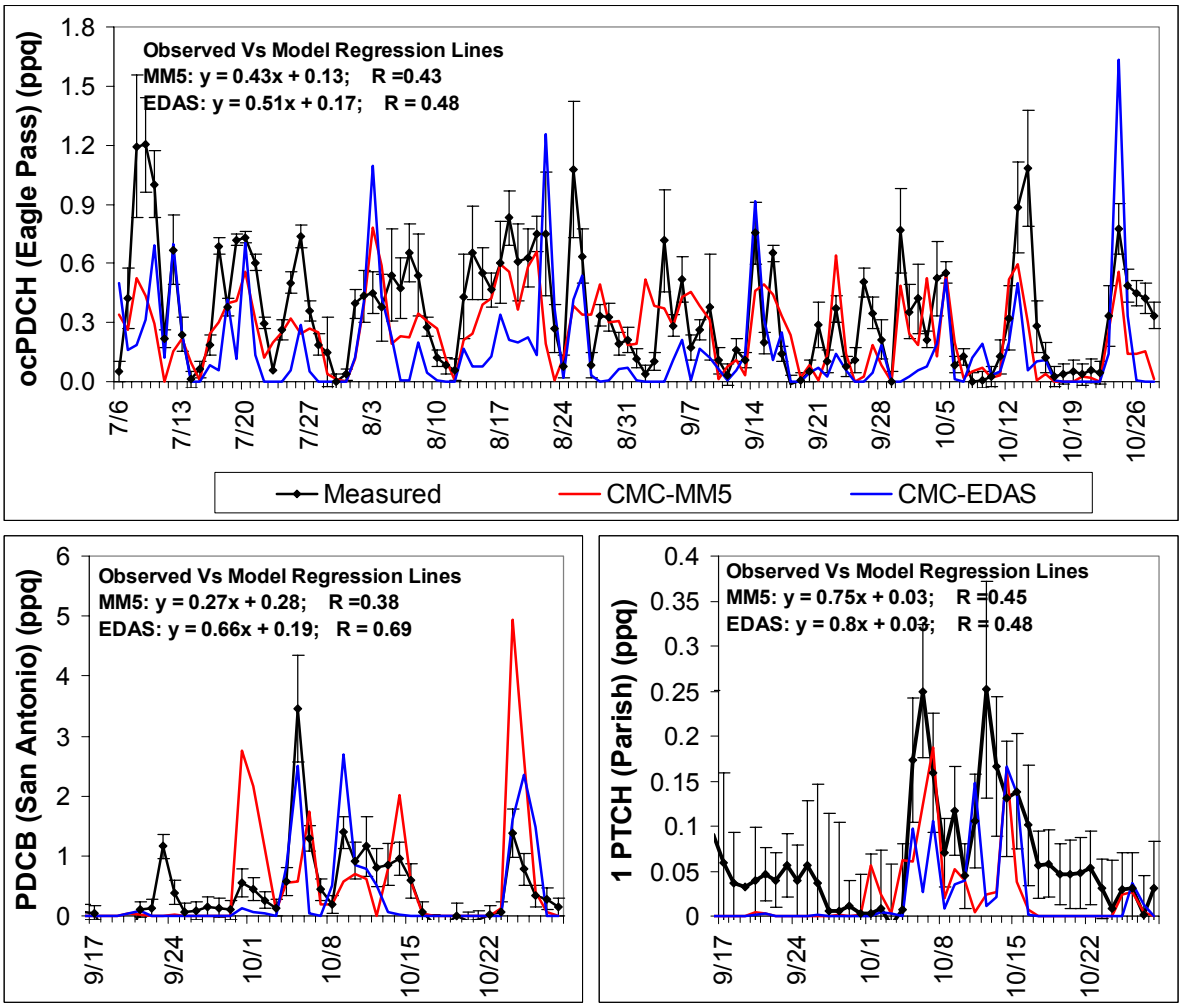

Figure 9-9. *Comparison of spatially-averaged observed and modeled tracer concentrations in southwest Texas (in the same format as Figure 9-8). The tracer concentrations were averaged over the six 6-hour monitoring sites: San Vicente, K-bar, Persimmon Gap, Marathon, Fort Stockton and Monahans Sandhills.*

Table 9-7. Model performance statistics for the Monte Carlo simulation of tracer concentrations averaged over the Big Bend sites (San Vicente, K-bar and Persimmon Gap) and over all six-hour tracer sites (San Vicente, K-bar, Persimmon Gap, Marathon, Fort Stockton and Monahans Sandhills) for the entire 4-month study period. Concentrations are in parts per quadrillion (ppq).

	MM5 Wind Fields -- Big Bend Sites (K-bar, Persimmon Gap, San Vicente)								
	Average		*Bias*	*Standard Dev.*		*RMS*		*Regression Line*	
	Obs	*Model*	*Mod/Obs*	*Obs*	*Model*	*Error*	r^2	*Intercept*	*Slope*
Eagle Pass	0.219	0.207	0.95	0.28	0.29	0.29	0.18	0.13	0.43
San Antonio	0.430	0.542	1.26	0.79	1.12	1.09	0.14	0.28	0.27
Houston	0.052	0.024	0.45	0.08	0.05	0.07	0.20	0.03	0.75

	MM5 Wind Fields -- 6-hour Tracer Sites								
	Average		*Bias*	*Standard Dev.*		*RMS*		*Regression Line*	
	Obs	*Model*	*Mod/Obs*	*Obs*	*Model*	*Error*	r^2	*Intercept*	*Slope*
Eagle Pass	0.344	0.247	0.72	0.29	0.19	0.26	0..31	0.13	0.85
San Antonio	0.418	0.534	1.28	0.67	1.01	0.94	0..18	0.27	0.28
Houston	0.058	0.022	0.38	0.06	0.04	0.06	0.31	0.04	0.81

	EDAS/FNL Wind Fields -- Big Bend Sites (K-bar, Persimmon Gap, San Vicente)								
	Average		*Bias*	*Standard Dev.*		*RMS*		*Regression Line*	
	Obs	*Model*	*Mod/Obs*	*Obs*	*Model*	*Error*	r^2	*Intercept*	*Slope*
Eagle Pass	0.219	0.103	0.47	0.28	0.26	0.30	0.23	0.17	0.51
San Antonio	0.430	0.368	0.86	0.79	0.82	0.63	0.48	0.19	0.66
Houston	0.052	0.022	0.43	0.08	0.05	0.07	0.23	0.03	0.80

	EDAS/FNL Wind Fields -- 6-hour Tracer Sites								
	Average		*Bias*	*Standard Dev.*		*RMS*		*Regression Line*	
	Obs	*Model*	*Mod/Obs*	*Obs*	*Model*	*Error*	r^2	*Intercept*	*Slope*
Eagle Pass	0.344	0.166	0.48	0.29	0.26	0.34	0.21	0.26	0.51
San Antonio	0.418	0.340	0.81	0.67	0.71	0.55	0.46	0.20	0.63
Houston	0.058	0.020	0.34	0.06	0.04	0.07	0.26	0.04	0.74

The tracer release sites at San Antonio and Houston are farther from Big Bend than Eagle Pass, approximately 450 km and 750 km distant respectively. Therefore, comparisons with those tracers test the Monte Carlo model's ability to simulate more regional scale transport. As was shown in Figure 9-8, tracer from San Antonio impacted Big Bend and southwest Texas several times in the first half of October and simulated tracer impacts also occurred during this time period. The tracer from Houston also had the largest impacts at Big Bend and southwest Texas in the first half of October.

As indicated in Table 9-7, the model run using EDAS/FNL meteorology had the most skill at reproducing the day-to-day variability of San Antonio tracer impacts at Big Bend, with $r^2 = 0.35$, compared to $r^2 = 0.14$ using the MM5 winds. However, the EDAS simulation underestimated the San Antonio tracer concentration at Big Bend by 14% while the MM5 simulation overestimated the tracer concentration by 26%. Monte Carlo modeling of the Houston tracer with either wind field simulated the measured concentrations with about equal skill, underestimating the average concentration by about a factor of 2 and with correlation coefficients around 0.45.

In summary, Monte Carlo model simulations with both the MM5 and EDAS/FNL wind fields properly transported tracer to Big Bend during the measured multi-day tracer events, and did not transport tracer to Big Bend when the measured tracer concentration was near the background levels for multi-day periods. The Monte Carlo model with MM5 winds appeared to be somewhat better at simulating the near field transport (as represented by the Eagle Pass tracer), while the EDAS winds produced better simulation of more distant scale transport, particularly of the tracer from San Antonio.

9.5 Evaluation of the Forward Mass Balance Regression (FMBR) Method Using Perfluorocarbon Tracer Measurements

The Forward Mass Balance Regression (FMBR) technique that was described in Section 8.3.6 combines information from the Monte Carlo modeling with a multivariate analysis. To test the performance of the FMBR method, it was applied to the perfluorocarbon tracer concentrations measured throughout Texas. The analysis method and results are summarized here; details are provided in the CIRA/NPS report (Schichtel et al., 2004), which is included in the Appendix.

Two analyses were carried out. The first analysis tested whether the technique was able to retrieve the tracer release site locations and tracer release rates using only the measured ambient tracer concentrations plus estimates of air mass transport by Monte Carlo modeling from nearly 700 potential release sites distributed throughout most of the US and Mexico. The second analysis tested the ability of FMBR to properly attribute the contributions from tracer source regions to total tracer concentrations at Big Bend. This approach is similar to the test used to evaluate the TrMB method in Section 9.3.

The tracer concentrations used for these analyses are those reported from the ambient tracer measurements, which reflect the measured concentrations minus an assumed background from sources other than the four BRAVO release locations. Some of the resulting concentrations are negative, which reflects the uncertainty in the measurement and background concentration. For the northeast Texas tracer there may have been a systematic bias in the adjusted concentrations. For example, the July average i-PPCH tracer at Big Bend was negative. All negative values were included in this analysis, however.

9.5.1 Retrieval of Tracer Release Locations and Rates

The FMBR inversion technique was first used to retrieve the tracer release locations and release rates of the ocPDCH tracer that was continuously released from Eagle Pass and

the i-PPCH tracer that was continuously released from northeast Texas during the last six weeks of the study. The inversion was conducted using singular value decomposition, which can invert an under-determined system and dampen instabilities that occur in least square regressions of ill-conditioned systems, such as the source receptor relationship.

The resulting reconstruction of the locations of the Eagle Pass and northeast Texas tracer release sites and rates is displayed in Figure 9-10. (The results shown here are based on analyses using the MM5 wind fields, but similar results were achieved with EDAS/FNL winds.) Benefiting from a data rich system with about 800 data points and multiple monitoring sites, the reconstruction was able to properly identify the Eagle Pass tracer release site location (the red cell with the highest emission rate) and was very close to the northeast Texas release site location despite the limited amount of usable tracer data there.

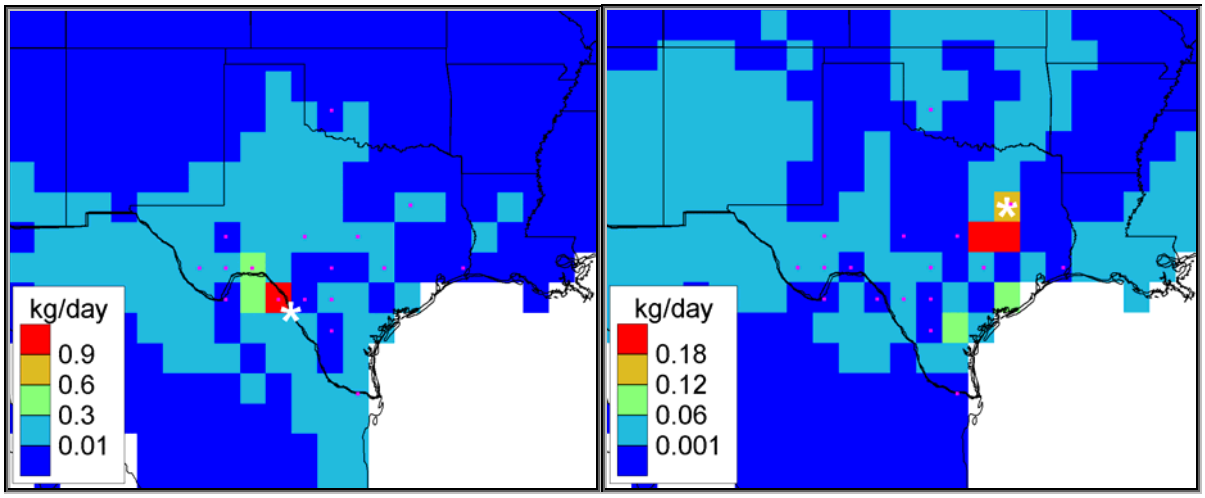

Figure 9-10. *FMBR reconstructions of the Eagle Pass and northeast Texas tracer release locations and rates. The colors indicate the estimated tracer emissions from each cell and the asterisks indicate the actual tracer release locations.*

The estimated tracer release rates are compared with the actual rates in Table 9-8. For the Eagle Pass tracer release, when all tracer concentration data were used, the reconstructed total tracer release rate of 4.2 kg/day is close to the actual rate of 3.7 kg/day. The release rate for the northeast Texas tracer was underestimated by about a factor of 6, though.

Note that the method did not work as well if only the tracer concentrations at Big Bend were used, particularly for the northeast Texas tracer. For that limited data set, the estimated location of the Eagle Pass tracer release was still quite good. The northeastern Texas release location was not identified correctly, though, but rather the reconstruction indicated tracer releases along the pathway from the actual tracer release location to Big Bend.

Table 9-8. Performance of the FMBR method in estimating tracer release rates.

Tracer	Receptors Used	Release Rate (kg/day)	
		Actual	Retrieved
Eagle Pass (ocPDCH)	All	3.7	4.2
	Big Bend	3.7	2
N.E. Texas (i PPCH)	All	2	0.34
	Big Bend	2	0.016

The FMBR technique relies on varied transport directions from the source to the receptor to minimize the collinearities between the transport pathways from the various sources to the receptor. The fact that the northeast tracer release site could not be isolated indicates that there was little variability in the transport pathways from northeastern Texas to Big Bend. Perhaps the FMBR method could have better isolated the tracer release location if data had been available for a longer time period, allowing for more tracer impacts to be included in the analysis. Alternatively, the northeast tracer may have been released at too low a rate to be enable consistent quantitative measurement its impacts at Big Bend, and increasing the amount of data in the analysis would not have improved the results.

9.5.2 Source Apportionment of Tracer Concentrations

We now address a different question: Given the locations of a small number of source regions, can FMBR properly estimate the contribution of each source region to the ambient concentration at Big Bend? To investigate this question, the contribution of each individual tracer to the total was estimated using FMBR, in much the same manner as was done for TrMB in Section 9.3.

Since tracers were released from the Houston and San Antonio sites only after mid September, this analysis included data from September 19 through October 30. The results of the FMBR attribution are compared against the actual tracer proportions in Figure 9-11. As shown, on average the FMBR source attributions properly identified the San Antonio tracer as the largest contributor and northeast Texas as the smallest.

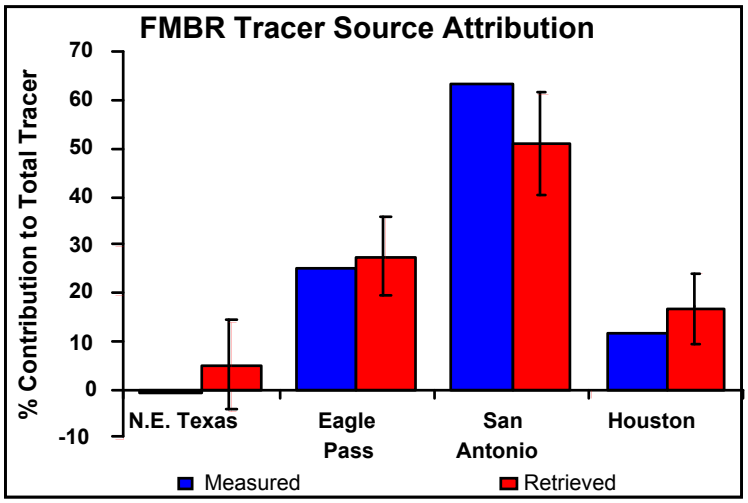

Figure 9-11. *Average percentage source contributions to measured tracer at Big Bend, as estimated by FMBR (red) and as measured (blue). The standard errors of the estimates are shown.*

9.5.3 Discussion of FMBR Tracer Evaluation

The results of these tests illustrate that, given enough data and a sufficiently accurate transport model, FMBR can properly identify the tracer release sites and rates with no *a-priori* information on the tracer release site locations. However, using only the Big Bend data, the resolution of the results diminish and the method only identifies the common transport pathways associated with the highest tracer concentrations at the Big Bend monitoring site. When the locations of the tracer release sites were provided to the FMBR, the technique was able to decompose the tracer time series into its contributions from the four individual tracer gases. Consequently, FMBR can be a useful technique for estimating average source contributions of inert species.

9.6 Evaluation of the Tracer Mass Balance (TrMB) Method Using REMSAD-Modeled Sulfate Concentrations

One test of the accuracy of a source attribution method is to determine how well it can reproduce known source attributions. One known set of attributions is that of sulfate sources simulated by the REMSAD model. In this case, there are known concentrations of the predicted 24-hr average sulfate at K-Bar and known attributions of this simulated sulfate to the BRAVO source areas. For purposes of the test, there is no assertion that the REMSAD-simulated sulfate source attributions are accurate in the real world (nor any requirement that they be so), only that in the simulated REMSAD world we have a consistent set of known emissions, chemistry, and meteorology. We would expect, if TrMB were a reliable approach for source attribution, that it could reproduce the REMSAD-simulated source attributions.

Since this juxtaposition of models and their results is unusual, it is important to understand that the goal of the exercise is not to test whether the REMSAD-simulated sulfur field is correct. The results need not be correct, but just plausible. The purpose of the test is to determine whether the receptor approach being evaluated can, with its assumptions and

limitations, come close to reproducing the known average attributions created by the REMSAD modeling. In the case of trajectory-based methods, such as the TrMB approach under discussion here, a specific desire is to see whether using trajectory endpoints alone is enough to explain average sulfate source attribution.

For this application, several trajectory models were used with TrMB, all based on the MM5 meteorological fields that were used in the REMSAD modeling, and the analysis was carried out for the period July 6 to October 28, 1999. The evaluation is described in detail in the CIRA/NPS report (Schichtel et al., 2004), which is contained in the Appendix. Key points are summarized below.

Table 9-9 summarizes the TrMB attributions of REMSAD sulfate to each of the four large BRAVO source regions — western U.S., Mexico, Texas, and eastern U.S. — that were described in Table 8-3. The attributions to these large areas were generated by aggregating the attributions of the 27 smaller source areas that were listed in Table 8-2 and illustrated in Figure 8-2. The bottom two rows of the table give the actual REMSAD attributions, first without the effects of boundary conditions and non-linearity in the apportionment process (from Figure 11-8), and second if the boundary conditions and non-linear fractions are added proportionally. (The REMSAD attribution assigns 7% to boundary conditions and 2% to non-linearity.) TrMB attribution values within 10 percentage points of the "correct" answer (the REMSAD attributions with the redistributed boundary conditions and non-linearity in the last row) are shown in bold red type for easy identification.

Table 9-9. Percent attributions of REMSAD-simulated sulfate by TrMB using MM5 winds and various trajectory models, compared to the corresponding REMSAD attributions in the two bottom rows show. TrMB attributions within 10 percentage points of the REMSAD attributions in the last row are shown in bold red type. The last three columns compare the TrMB predicted concentrations to REMSAD-simulated sulfate concentrations.

	ATTRIBUTIONS TO SOURCE REGIONS (%)				**TrMB PERFORMANCE**		
Trajectory Model used with TrMB	Texas	Mexico	Eastern U.S.	Western U.S.	R^2	Mean $SO_4^=$ overestimate	
						$\mu g/m^3$	%
CAPITA 5-day	**19**	**31**	**39**	**11**	.778	-0.023	-1%
CAPITA 7-day	**20**	**24**	**36**	**20**	.798	-0.012	-1%
CAPITA 10-day	**21**	**21**	**37**	21	.775	-0.013	-1%
HYSPLIT 5-day	43	**25**	16	**17**	.768	0.011	+1%
HYSPLIT 7-day	43	**23**	16	**18**	.820	-0.026	-1%
HYSPLIT 10-day	46	**21**	19	**13**	.801	-0.014	-1%
ATAD 5-day	**25**	**33**	**36**	8	.735	-0.024	-1%
REMSAD REFERENCE							
REMSAD w/o BC and nonlinearity	16	23	42	9			
REMSAD w/ BC and nonlinearity	18	25	46	10			

The best reproduction of the REMSAD sulfate attributions occurred when the TrMB method was used with the CAPITA Monte Carlo Model using 5-day back trajectories, although the 7-day and 10-day trajectories were nearly as good. The ATAD model with MM5 input also attributed the sulfate correctly to all four source regions within 10 percentage points of the correct values. The HYSPLIT model with MM5 input was able to reproduce the correct source attributions for Mexico and the western U.S. but was unable to correctly apportion sulfate from Texas and the eastern U.S., attributing much more to Texas and much less to the eastern U.S. than REMSAD.

For the best-performing CAPITA Monte Carlo model, the trajectory length made little difference in the TrMB-predicted attributions, except for the western U.S., which is considered to be culpable for twice as much sulfate with 7- or 10-day trajectories as with 5-day ones. This is intuitively reasonable since air masses rarely arrived at Big Bend directly from the western U.S. during BRAVO, but more often traversed across the eastern U.S. and/or Texas prior to arrival. Thus, on average, the travel time from the western U.S. was longer than from the other large source areas and so longer trajectories attributed more to this region. (It is also possible that the greater attribution associated with longer trajectories is a consequence of over-attributing emissions from distant sources when the trajectories are typically so high that they would not pick up emissions from within the mixed layer.)

Looking at the trajectories can provide insight into why the HYSPLIT simulation did not accurately reproduce the REMSAD sulfate attributions even though it uses essentially the same input meteorological data as CAPITA MC and ATAD There is a large difference between the CAPITA MC and HYSPLIT back trajectories on Sept. 1 (Julian day 244), which was the day with the highest measured sulfate concentration at Big Bend and so is an influential point in the multiple linear regression of TrMB. The CAPITA MC model has most of the endpoints in the eastern U.S. on this date (as does ATAD with MM5 input), while HYSPLIT puts most of them in Texas.

While all three models have trajectories with the same general direction on Sept. 1, the HYSPLIT trajectories are much lower in height and are essentially on the ground. This is true even for trajectories with a start height of 1000 m. Due to the lower transport height, they also have much lower wind speeds and so remain in Texas while the CAPITA MC and ATAD trajectories extend into the Eastern U.S. Because this was an influential day, the differences in trajectory heights on this day alone may explain why HYSPLIT was unable to reproduce the REMSAD source attributions.

The TrMB method assumes that the concentration attributable to a specific source area is proportional to the number of trajectory endpoints that fall in that area. For simulated REMSAD sulfate, where the concentration of sulfate attributed to each source area is known, it is possible to examine whether the assumed linearity of the relationship between endpoints and sulfate concentrations is valid. Figure 9-12 displays scatter plots of the daily number of endpoints in each of the four large areas versus the sulfate concentrations at Big Bend that were attributed by REMSAD to the area for each day. The endpoints were generated using the CAPITA MC model. The solid blue line on each graph is the mean ratio of sulfate attributed by REMSAD to the source region to number of endpoints in that region. This is

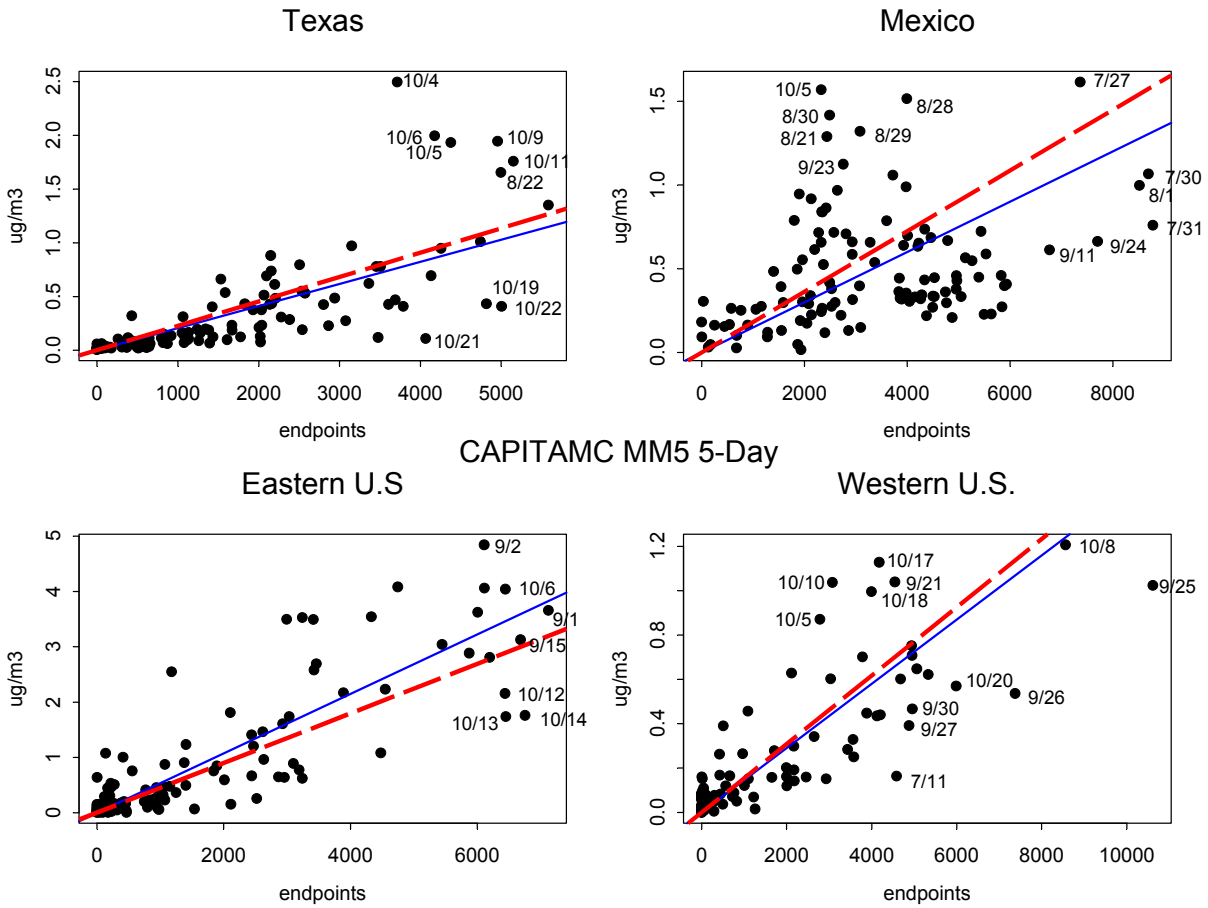

Figure 9-12. *Scatter plots of REMSAD sulfate concentration attributions vs. number of endpoints for each of four large source areas using 5-day back trajectories from the CAPITA MC Model and MM5 winds. The solid blue lines represent the mean ratio of REMSAD-attributed sulfate to number of endpoints for each source area and the dashed red lines represent the corresponding mean ratio of the TrMB analysis. Note that the concentration scale is different for each plot.*

the "correct" relationship that the TrMB model would have to re-create in order to give the same mean attributions as REMSAD. The scatter of the points about the blue line gives an idea of how much the true relationship between endpoints and concentrations deviates from the mean from day to day, and thus how far the attribution for an individual day could deviate from the mean attribution. The dashed red line indicates the mean ratio derived by the TrMB analysis. The angle between the red and blue lines shows how well the TrMB method was able to reproduce the mean REMSAD attribution of sulfate; the best that TrMB can do is match the red line to the blue line.

Figure 9-12 shows that the assumption of linearity seems to be fairly good when using the Monte Carlo model, because the mean attribution is quite good (the red line is close to the blue line), though some large deviations from the mean relationship exist on individual days. For example, for the Texas source area, most points lie close to the blue line. The

main exception is a period of a few days in October when REMSAD sulfur concentrations from all source areas are under-predicted by TrMB. This indicates that there was something non-average about the chemistry and/or deposition on those days. In the case of October 4 (the point with the highest sulfate attribution), the deviation from the line indicates that the sulfate attribution to Texas on this day could be underestimated by TrMB by nearly a factor of 5, even though on average, the mean attribution to Texas was within one percentage point of being correct. In the REMSAD model results, which are "truth" for this test, both October 4 and 5 have higher than average ozone concentrations over southeast Texas and an unusual period of no precipitation over Texas, which could cause the relationship between sulfate at Big Bend and transport from areas within Texas to be different from average. In fact, REMSAD assigns large attributions to Texas during that period.

In summary, the TrMB method, using either the Monte Carlo or ATAD trajectory model run with the MM5 wind fields, can reproduce the average REMSAD sulfate attributions for the four large source areas to within ten percentage points. Examination of the relationship between sulfate attributions and endpoints indicates, though, that caution should be exercised when using TrMB to estimate source attributions on individual days. HYSPLIT can reproduce the attributions to Mexico and the western U.S., but it overestimates the attribution to Texas and underestimates that of the eastern U.S. This may be due to trajectory heights that are too low on a single highly influential day, thus giving too much influence to Texas and not enough to the eastern U.S.

9.7 Evaluation of the Forward Mass Balance Regression (FMBR) Method Using REMSAD-Modeled Sulfate Concentrations

In Section 9.5, the FMBR source attribution technique was evaluated using inert tracer concentrations. Sulfate source attribution has the complicating factor that the emitted sulfur undergoes transformation and removal processes during transport from the source to the receptor. To evaluate the FMBR method for a reactive species, it was applied using the REMSAD-predicted 24-hour sulfate concentrations at Big Bend. As in the TrMB evaluation described in the section above, the REMSAD results, provided an artificial reality in which both the concentrations and the contributions from 10 large source regions and the model boundary conditions were known. Therefore, the FMBR source attribution results could be compared against "known" source attributions.

The main points concerning the evaluation of FMBR using the REMSAD sulfate concentrations are presented below. Details can be found in the CIRA/NPS report (Schichtel et al., 2004), which is included in the Appendix.

For this evaluation, air mass transport from 10 source regions was estimated using the CAPITA Monte Carlo Model driven by the MM5 wind fields (which are the same wind fields used for the REMSAD simulation). Nine of these source regions, listed in Table 9-10, are combinations of the 17 regions shown in Figure 8-3. Monte Carlo transport of concentrations from the REMSAD domain boundaries was not simulated, so the REMSAD-calculated contribution from the boundary conditions was added as an additional source contribution for the FMBR analyses.

Table 9-10. Source attribution estimates from application of the FMBR method using a set of REMSAD-predicted sulfate concentrations, and comparison with a sample set of REMSAD source attributions. Results are shown for three trajectory lengths. Estimated attributions within one standard error of the REMSAD results are in bold, red type.

	FMBR % Contribution ± Standard Error			
	5 day	7 day	10 day	*REMSAD Model
SOURCE REGIONS				
Carbón Plant	**16.1 ± 4.2**	**16.5 ± 4.1**	**15.4 ± 4.2**	14.1
Rest of Mexico	**13.6 ± 9.6**	**12.2 ± 10.2**	**11.6 ± 11.1**	10.0
NE Texas	**7.8 ± 2.8**	9.5 ± 3.0	10.2 ± 3.1	5.4
SE Texas	14.0 ± 4.1	**11.9 ± 4.3**	**11.8 ± 4.6**	8.8
Rest of Texas	**5.0 ± 4.2**	**4.0 ± 4.5**	**3.7 ± 4.8**	2.3
LA/MS & MO/IL/AR	33.8 ± 5.6	**28.7 ± 7.6**	**27.6 ± 8.2**	22.3
East Central U.S.	0.6 ± 2.1	7.4 ± 4.4	**8.8 ± 4.8**	13.6
Rest of Eastern U.S.	2.1 ± 2.9	2.7 ± 4.1	**4.0 ± 5.1**	7.4
Western U.S.	0.0 ± 7.8	0. 0± 8.5	**0.0 ± 9.3**	9.3
Boundary Conditions	7	7	7	7.2
AGGREGATED REGIONS				
Mexico (all)	**29.6 ± 10.4**	**28.8 ± 11.0**	**27.0 ± 11.8**	24.0
Texas (all)	26.8 ± 6.5	25.4 ± 7.0	25.7 ± 7.3	16.5
Eastern U.S. (all)	36.5 ± 6.7	**38.8 ± 9.7**	**40.4 ± 10.8**	43.2
Western U.S. (all)	0. 0± 7.8	0.0 ± 8.5	**0.0 ± 9.3**	9.3
Boundary Conditions	7	7	7	7.2

* REMSAD source attributions were increased by 2.7% to account for missing mass due to nonlinearities in the apportionment simulations. Note that these attributions are a sample set and do not necessarily represent the findings of the BRAVO Study, which are presented in Chapter 11.

The FMBR-estimated daily concentrations of sulfate over each of the source regions compared favorably with the REMSAD sulfate data, with r^2 about 0.8, an RMS error of 40%, and bias of 4%. Consequently, the sulfur transport from 10 large source regions, multiplied by a constant, explains 80% of the variance in the REMSAD sulfate daily time series.

The percentages of sulfate attributed by the FMBR approach to each of the source regions, and for larger aggregate regions, are presented in Table 9-10 for three different lengths – 5, 7, and 10 days – of the simulated source plumes. REMSAD attributions for the same source regions are shown for comparison. Each FMBR value is followed by its standard error; apportionments within one standard error of the REMSAD values are shown in bold red type. The boundary conditions accounted for about 7% of the total REMSAD simulated sulfate at Big Bend, which is included in all columns. (Note that the attributions in Table 9-10 are not the final REMSAD attributions developed by this study, but rather an earlier set of REMSAD results that provide an artificial reality for evaluating the performance of the FMBR method. The actual attribution results of the BRAVO Study will be presented in Chapters 10 and 11.)

For the 10-day simulations, apportionment estimates for all but the northeast Texas source region are within the standard error of the REMSAD results. Simulations with shorter trajectories reproduced the REMSAD apportionments in fewer source regions.

In the lower part of Table 9-10, it is interesting to note that the attribution to all of Texas is overestimated by about the same amount that the attribution to the western U.S. is underestimated, while the attributions for the eastern U.S. and Mexico source regions properly reproduce the REMSAD result (i.e., they're within one standard error). The cause for the Texas estimating error was explored, and it was found that the FMBR technique has a tendency to increase the attributions to closer source regions at the expense of more distant source regions. Also, when included, boundary condition contributions tend to be incorrectly attributed to the closer source regions. When these potential biases are taken into account, the FMBR method can be a useful method for regional source apportionment.

9.8 Evaluation of REMSAD Simulations of Perfluorocarbon Tracers

We now turn to assessing the performance of the regional air quality models used for source attribution in the BRAVO Study. These models were evaluated against the perfluorocarbon concentrations measured during the tracer study and against ambient sulfur measurements at Big Bend and throughout Texas. The tracer evaluations assessed the performance of the regional modeling system (i.e., the air quality model and the meteorological field) when simulating transport and diffusion. Since the perfluorocarbon tracers are inert gases, the tracer evaluations did not assess the ability of the modeling system to simulate chemical transformation or deposition.

This section describes the evaluation of the REMSAD system against the perfluorocarbon measurements. A similar evaluation for the CMAQ system is described in Section 9.10. Evaluations of the REMSAD and CMAQ-MADRID modeling systems against ambient sulfur measurements, which also test the chemical transformation capabilities of the models, will be presented in Sections 9.9 and 9.11, respectively.

In order to evaluate its performance at simulating the transport and diffusion of inert emissions, the REMSAD regional air quality model was used to simulate the four perfluorocarbon tracers that were released during the BRAVO study (as described in Section 3.2. The configuration of REMSAD for the tracer simulation was similar to that used for the base emissions simulation described in Section 11.1, except that 1) the chemistry mechanism was not invoked, since the tracer did not undergo chemical transformation, 2) loss via depositional settling was not considered, since it was assumed that the tracers have very low deposition velocities, and 3) background concentrations were set to zero. The grid scale of the model was 36 km. Horizontal winds, temperature, and other meteorological fields were simulated by MM5. Details of the evaluation are provided in the CIRA/NPS report (Schichtel et al., 2004), which is included in the Appendix.

Each of the four tracer releases was treated as a point source emission. Two of the tracers were released in the stacks of power plants, the northeast Texas tracer (PPCH) at the Big Brown power plant and the Houston tracer (PTCH) at the Parish power plant, and thus were lofted higher into the boundary layer due to buoyant and momentum plume rise.

REMSAD simulates this initial plume rise. Tracer releases at the other two locations were released directly in the ambient air and thus did not have any plume rise. An example of a REMSAD simulation of the transport and dispersion of the four tracers is illustrated in Figure 9-13.

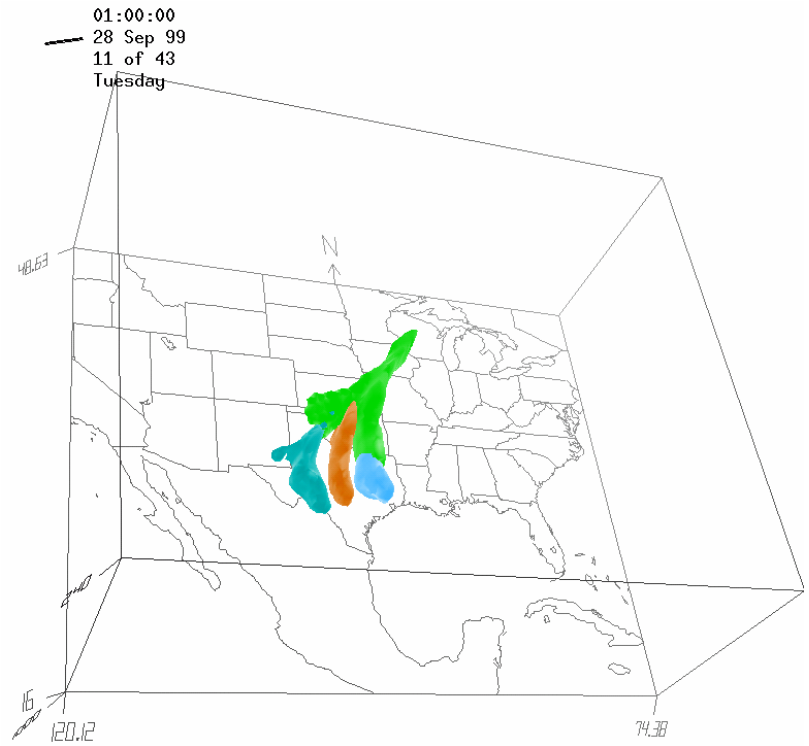

***Figure 9-13**. Example of the REMSAD-simulated dispersion of the four tracer plumes on 28 September 1999. The colored areas in this 3-dimensional view of the REMSAD modeling domain represent concentrations above 0.1 ppq. The wind is generally toward the north, so the southern ends of the colored regions are in the vicinity of the release locations. The plumes represent the tracer released from (from left to right across the bottom) Eagle Pass, San Antonio, Houston and (the green area) northeast Texas.*

Because of the coarse grid scale of the model and the complex topography in the Big Bend region, both modeled estimates and measured tracer concentrations were averaged over all three measurement sites in Big Bend National Park. (San Vicente, K-Bar, and Persimmon Gap). These three-site averages of 24-hr measurements and modeled estimates are compared for each tracer in Figures 9-14 through 9-17. Results for the Eagle Pass and northeast Texas tracers are shown for the entire four month BRAVO period, while results for the San Antonio and Houston tracers are shown for those last six weeks of the BRAVO period when the tracers were released from these sites. Note that the vertical scales of these figures are not all the same. The "mixing ratio" is equivalent to the tracer concentration in ppq by volume. Performance statistics corresponding to these figures are tabulated in Table 9-11.

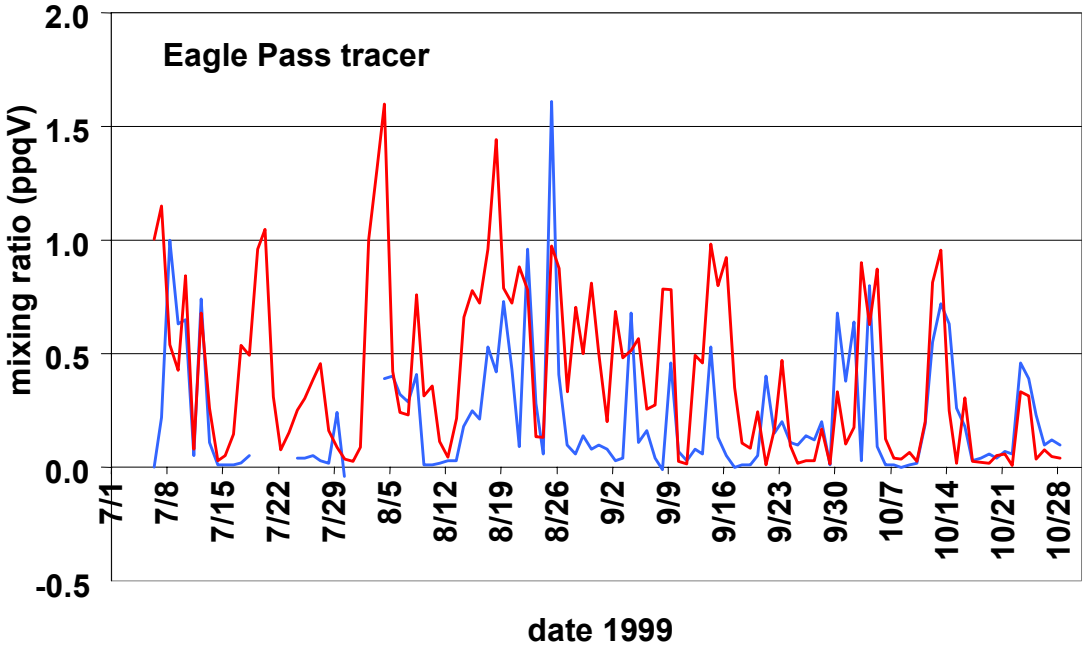

Figure 9-14. *Comparison of 24-hr three-site average REMSAD predictions (red) of the spatially-averaged concentrations in Big Bend National Park versus three-site average measurements (blue) for the continuous Eagle Pass tracer.*

At Big Bend National Park, the spatially-averaged measured concentration of the tracer released from Eagle Pass (Figure 9-14) generally ranges between 0 and 1 ppq during the four month study period, with a peak of 1.6 ppq observed on 25 September. Tracer "hits" are generally well replicated by REMSAD, although the average predicted value of 0.39 ppq is nearly twice the observed value of 0.22 ppq, and from mid-August through September there is a clear tendency for REMSAD to overstate tracer concentrations. Also, the modeled concentrations do not vary as much as the measured ones.

In contrast to the Eagle Pass tracer, the model underestimated the mean concentrations of the Houston and San Antonio tracers (Figures 9-16 and 9-17). For the northeast Texas tracer, the measurements of which were barely above the analytical noise and for which the background-adjusted concentrations were often negative, the correlation coefficient was a surprising 0.34 (r^2 = 0.12), but the model significantly overstated the measured concentrations.

In general, it appears that the performance of the model was best for the tracers released closest to the park (at Eagle Pass and San Antonio), and became poorer as the transport distance increased (as to Houston and northeast Texas). The distance also affected the concentrations of the actual tracers at the park, so the more distant tracer concentrations were the least reliable. Interestingly, the poorer performance corresponded with the tracers released into power plant stacks (at Houston and northeast Texas) instead of near ground level (at Eagle Pass and San Antonio). Whether the resulting more-elevated transport played a role in the performance comparison is an open question.

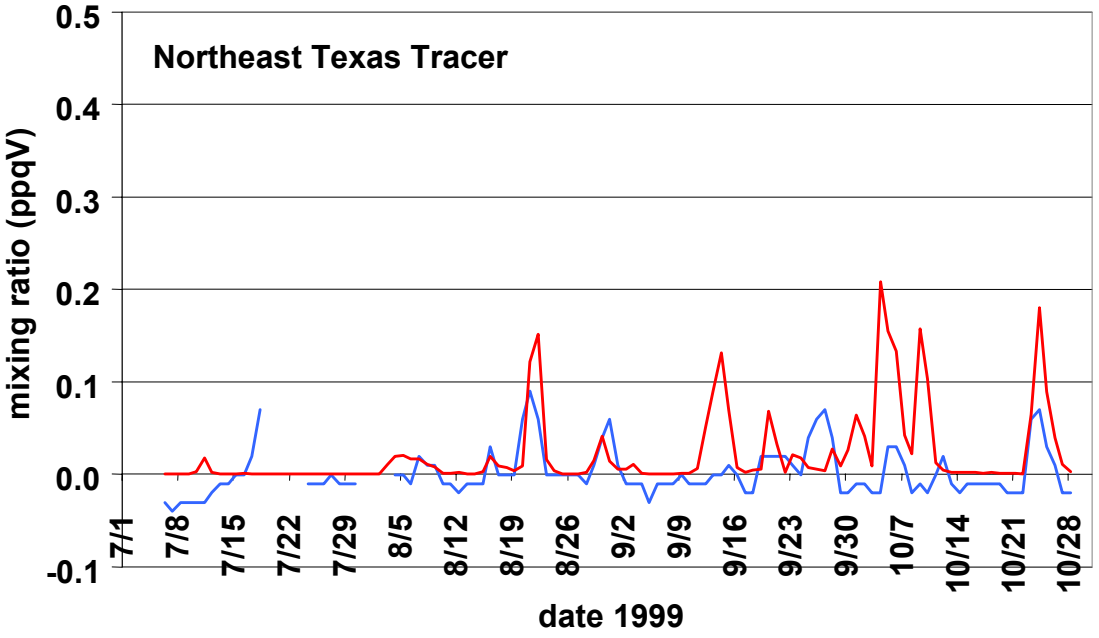

Figure 9-15. *Comparison of 24-hr three-site average REMSAD predictions (red) of the spatially-averaged concentrations in Big Bend National Park versus three-site average measurements (blue) for the northeast Texas tracer.*

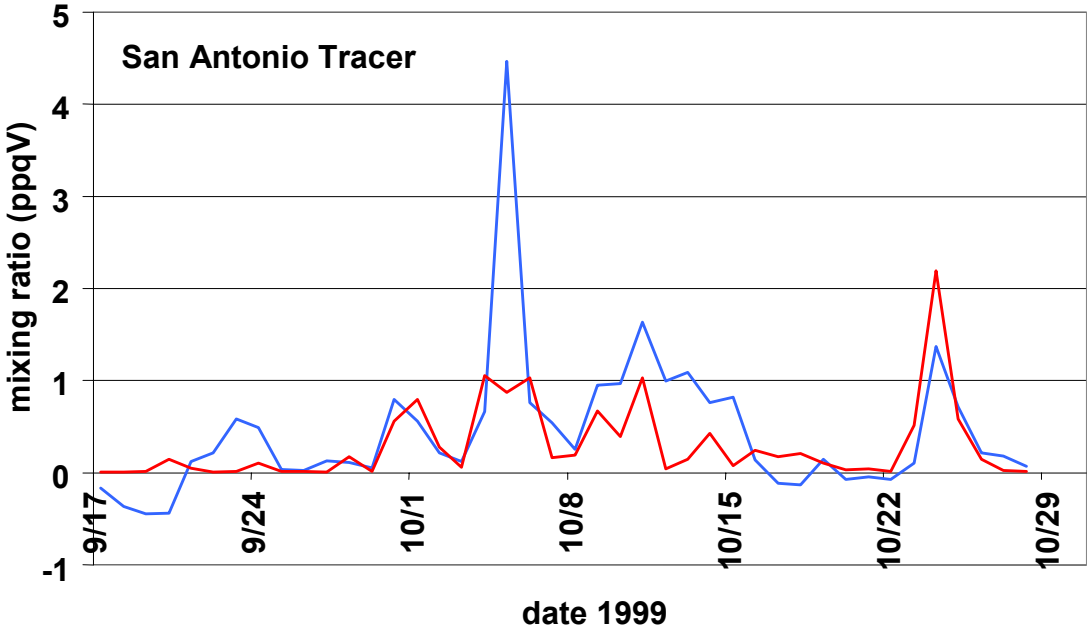

Figure 9-16. *Comparison of 24-hr three-site average REMSAD predictions (red) of the spatially-averaged concentrations in Big Bend National Park versus three-site average measurements (blue) for the San Antonio tracer.*

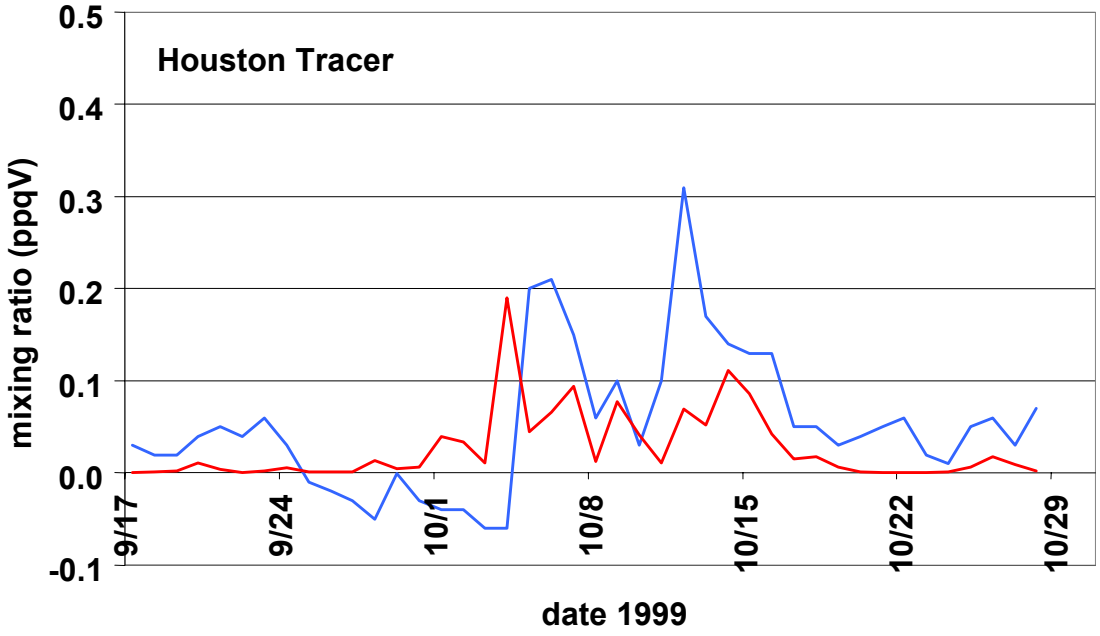

Figure 9-17. *Comparison of 24-hr three-site average REMSAD predictions (red) of the spatially-averaged concentrations in Big Bend National Park versus three-site average measurements (blue) for the Houston tracer.*

Table 9-11. REMSAD model performance statistics for the four tracers. The variables compared are the spatially-averaged observed and modeled 24-hr averages of the concentrations at the three monitoring sites in Big Bend National Park.

	Average (ppq)		Bias	Std. Dev. (ppq)		Coeff. of Variation		RMS Error			Regression Line	
	Obs	Mod	Mod/ Obs	Obs	Mod	Obs	Mod	Abs. (ppq)	Rel (%)	r^2	Int. (ppq)	Slope
Eagle Pass (ocPDCH)	0.214	0.386	1.80	0.27	0.35	1.24	0.91	0.37	171	0.22	0.25	0.62
San Antonio (PDCB)	0.522	0.327	0.63	0.78	0.44	1.50	1.35	0.70	133	0.27	0.17	0.30
Houston (1PTCH)	0.055	0.029	0.53	0.08	0.04	1.46	1.37	0.08	149	0.10	0.02	0.16
NE Texas (iPPCH)	0.003	0.025	8.43	0.03	0.04	8.61	1.79	0.05	1638	0.12	0.02	0.59

9.9 Evaluation of REMSAD Simulations of SO_2, and Sulfate Measurements

The complete REMSAD regional air quality model (described in Section 8.4.2) was used to create a base simulation of air quality for every hour of the July–October 1999 BRAVO Study period. The concentrations estimated in this base simulation were compared with measurements to evaluate the model's skill. Emphasis was placed on REMSAD's ability to simulate particulate sulfate concentrations, since sulfate aerosol is the leading contributor to haze in Big Bend NP. Complete results of this evaluation are described in the CIRA/NPS report (Schichtel et al., 2004), which is included in the Appendix, and the main points are summarized here.

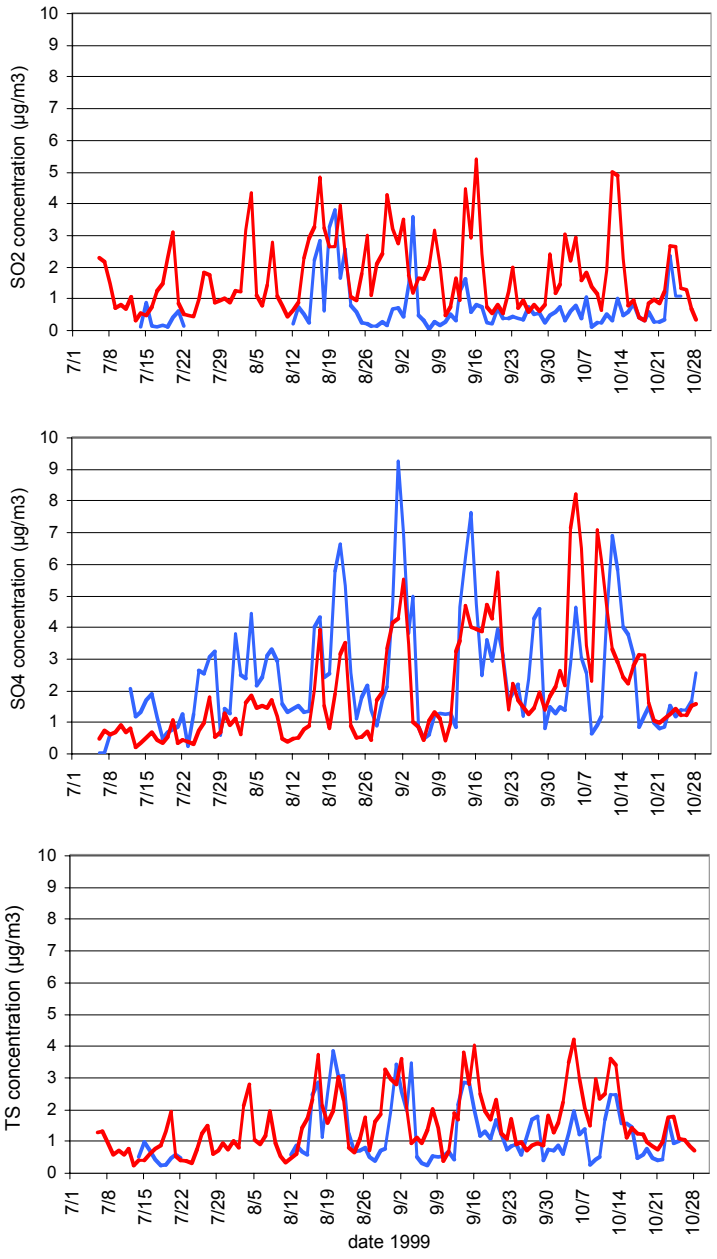

Figure 9-18. *Comparisons of observed (blue) and modeled (red) time series of 24-hr averaged concentrations of SO_2, sulfate, and total sulfur at the K-Bar sampling site in Big Bend National Park.*

Time series analyses of 24-hr average modeled vs. observed sulfur species, i.e., sulfur dioxide (SO_2), particulate sulfate ($SO_4^{=}$), and total sulfur (TS, the sum of the sulfur in SO_2 and $SO_4^{=}$), were prepared for the three air quality monitors within Big Bend National Park (K-Bar, San Vicente, and Persimmon Gap). An example, for K-Bar, is shown in Figure 9-18. (Plots for all three locations are included in the CIRA/NPS BRAVO Study report (Schichtel et al., 2004).

Performance statistics for the same data at all three locations individually are given in Table 9-12. Definitions of those statistics, which are used in this report for describing both REMSAD and CMAQ performance, are given in Table 9-13. In Table 9-12, note that the San Vicente and K-Bar sites, which are approximately 30 km apart, fall within the same REMSAD 36 km grid cell, and hence the modeled concentrations for these two sites will be the same, even though the measured concentrations at the two locations may differ. The averages in Table 9-12 show small differences, though, because differing numbers of valid measurement samples dictate different numbers of both measured and corresponding modeled daily values for each location and for each variable. Also, the observed total sulfur calculation requires concentrations of both SO_2 and sulfate, and thus encompasses fewer days than either of these components.

Table 9-12. Performance statistics over the full BRAVO Study period for REMSAD simulations of SO_2, sulfate, and total sulfur at the three sampling locations within Big Bend National Park.

	K-Bar			San Vicente			Persimmon Gap		
	SO_2	$SO_4^=$	TS	SO_2	$SO_4^=$	TS	SO_2	$SO_4^=$	TS
Avg. Observed (ug/m^3)	0.69	2.47	1.19	0.60	2.59	1.19	1.43	2.71	1.63
Avg. Simulated (paired) (ug/m^3)	1.84	2.02	1.69	1.75	2.05	1.66	2.00	2.30	1.80
r^2	0.09	0.25	0.31	0.12	0.27	0.34	0.15	0.25	0.29
Regression Slope	0.47	0.48	0.61	0.51	0.50	0.66	0.31	0.51	0.50
Intercept (ug/m^3)	1.52	0.83	0.96	1.45	0.75	0.88	1.56	0.91	0.99
Mean Normalized Error (%)	357	62	106	413	59	103	200	56	62
Mean Normalized Bias (%)	352	1	91	404	-5	86	186	1	40
Fractional Error (%)	99	60	57	112	59	57	71	52	42
Fractional Bias (%)	92	-28	38	102	-32	35	53	-24	16

Table 9-12 shows a tendency for REMSAD to overstate SO_2 concentrations at K-Bar and San Vicente, with very large mean normalized biases of 352 and 404%, respectively (i.e., on average, REMSAD overestimates the measured SO_2 concentrations at these two locations by roughly a factor of five). The average bias is less, but still large, at Persimmon Gap, where the mean normalized bias is 186%. The coefficient of determination (r^2) between measured and modeled SO_2 concentrations is uniformly low at these sites, ranging between 0.09 and 0.15.

Table 9-13. Definitions of performance statistics used in this report.

	Simulated or Predicted	Observed
Mean or Average	$\frac{1}{N}\sum_{i=1}^{N} P_i$	$\frac{1}{N}\sum_{i=1}^{N} O_i$
Coefficient of Determination	R^2	
Mean Bias	$\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)$	
Mean Error	$\frac{1}{N}\sum_{i=1}^{N}\lvert(P_i - O_i)\rvert$	
Fractional Bias (%)	**Mean Normalized Bias (%)**	
$\frac{2}{N}\sum_{i=1}^{N}\frac{(P_i - O_i)}{(P_i + O_i)}$	$\frac{1}{N}\sum_{i=1}^{N}\frac{(P_i - O_i)}{O_i}$	
Fractional Error (%)	**Mean Normalized Error (%)**	
$\frac{2}{N}\sum_{i=1}^{N}\left\lvert\frac{P_i - O_i}{P_i + O_i}\right\rvert$	$\frac{1}{N}\sum_{i=1}^{N}\frac{\lvert P_i - O_i\rvert}{O_i}$	
Root Mean Square (RMS) Error	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}$	

Note, however, that the intersite correlations between the SO_2 measurements themselves are also low, as shown in Table 9-14. For example, the coefficient of determination between the K-Bar and San Vicente observed SO_2 time series is $r^2 = 0.18$, and between K-Bar and Persimmon Gap it is only 0.01, even though these sites are within relatively close proximity. (K-Bar is approximately 40 km south of Persimmon Gap and 30 km north of San Vicente.) Also, the average and peak SO_2 concentrations recorded during the four-month field study were considerably higher at Persimmon Gap than at the other two sites. This lack of correlation between the SO_2 time series at the different Big Bend monitoring sites may be due to plumes from local sources (i.e., *Carbón* I/II) that are affected by the complex terrain in the area of the park. For example, the San Vicente monitor lies in a valley floor near the Rio Grande, and is 400 to 500 m lower than the K-Bar and Persimmon Gap monitors; San Vicente may be impacted by plumes traveling up the river valley that do not reach the higher-elevation sites. In any case, the low correlations between measured concentrations in the same or contiguous 36-km cells needs to be taken into account when evaluating model performance for SO_2.

Table 9-14. Intersite correlations of 24-hr average concentrations of SO2, sulfate, and total sulfur measured at the three Big Bend National Park monitoring sites.

	Persimmon Gap	K-Bar	San Vicente
SO_2			
Persimmon Gap	1.00	0.01	0.06
K-Bar	0.01	1.00	0.18
San Vicente	0.06	0.18	1.00
$SO_4^=$			
Persimmon Gap	1.00	0.81	0.84
K-Bar	0.81	1.00	0.89
San Vicente	0.84	0.89	1.00
TS			
Persimmon Gap	1.00	0.30	0.35
K-Bar	0.30	1.00	0.70
San Vicente	0.35	0.70	1.00

Turning now to sulfate, Figure 9-18 and Table 9-12 showed that the REMSAD sulfate estimates more closely matched the measurements than did those of SO_2. Mean normalized errors are about 60%, mean normalized biases are smaller than 5%, and correlation coefficients are about 0.25 ($r^2 \sim 0.06$). In both the estimated and observed $SO_4^=$ time series, the lowest concentrations generally occur during July and the first half of August, where trajectory analysis shows flow predominately from Mexico. Although the timing of the REMSAD $SO_4^=$ peaks appear to be in good agreement with the observed peaks during this period (e.g., 4 August and 17 August), there is a clear bias for the predicted peak SO_4 concentrations to be too low at all three monitors. This may indicate that Mexican SO_2 emissions are underestimated, or perhaps that the sulfate conversion rates calculated by REMSAD are too low, or perhaps the model overestimates deposition. From the latter half of August until the end of October, SO_4 concentrations are higher, with a peak SO_4 concentration of 9 $\mu g/m^3$ SO_4 observed on 1 September at K-Bar.

Although the timing of the predicted peak sulfate concentrations generally appears to be in good agreement with the observations during the second half of the BRAVO Study period, there are a few notable exceptions, such as a 5 $\mu g/m^3$ peak that was observed at all three locations around 27-29 September. This late September episode is not evident in the modeled time series at any of the three monitors, although concentrations of this magnitude were predicted within a 100 km of Big Bend NP, as shown in Figure 9-19. Also, there appears to be an error in the timing of the mid-October SO_4 peak, with the predicted maximum occurring 9 October while the actual maximum was observed on 12 October.

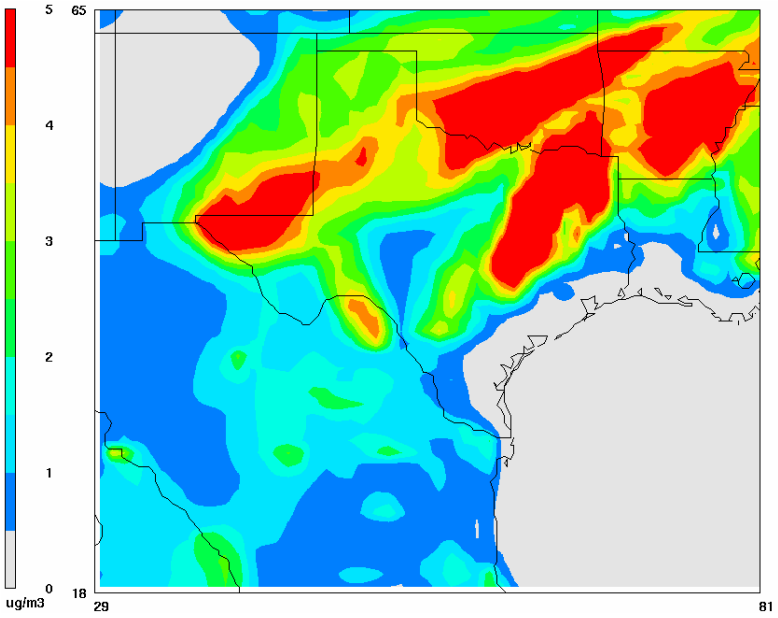

Figure 9-19. *Smoothed REMSAD simulation of ground-level sulfate concentrations at 1700 CDT on 28 September 1999. Note the small cloud of elevated sulfate concentrations to the east of Big Bend.*

Table 9-12 showed that intersite correlations of observed sulfate, ranging between 0.81 and 0.89, are considerably better than those for SO_2. This is to be expected, as SO_4 has a relatively long residence time in the atmosphere as compared to SO_2, and the $SO_4^{=}$ plume that impacts Big Bend NP is likely to be more dispersed with weaker spatial gradients.

The final variable considered here is TS, which is the concentration of sulfur that is contributed by both SO_2 and $SO_4^{=}$. According to Table 9-12, mean normalized biases for TS at Big Bend range from 40 to 91%, a consequence of the overestimation of SO_2, and mean normalized errors are 62 to 206%. Correlation coefficients for TS are between 0.29 and 0.34, values that are slightly higher than those for SO_2 and SO_4. The magnitudes and timing of the predicted TS concentrations appear to match the observations relatively well (although there is a significant amount of missing data in July) except for October, when the amount of sulfur predicted at Big Bend NP is significantly greater than the observed concentrations. Also, it is interesting that the timing of the mid-October TS peak is correct, contrasted with the premature occurrence of the simulated SO_4 peak, possibly indicating that estimated SO_4 oxidation rates are too rapid during this period.

A similar performance analysis was carried out for air quality monitors located along the northern, eastern, southern, and western peripheries of the BRAVO monitoring network. The monitoring sites considered include 1) the Wichita Mountains site, a Class I area in southern Oklahoma, 2) the Big Thicket site in eastern Texas near the Louisiana border, 3) the Laguna Atascosa site at the southern end of Texas near the Mexico border, and 4) the Guadalupe Mountains National Park site, another Class I area in western Texas. SO_2 concentrations are overstated by REMSAD at all four sites, especially at Big Thicket, while the $SO_4^{=}$ time series shows much better agreement between the observations and predictions. Details are provided in the CIRA/NPS report on the BRAVO Study

Finally, the overall performance of REMSAD was evaluated by aggregating the measured and simulated concentration data at all 37 BRAVO monitoring locations. Performance statistics are shown in Table 9-15, and scatter plots of observed and predicted SO_2, SO_4, and TS are shown in Figure 9-20. There is a clear tendency for SO_2 concentrations to be overestimated, with a mean normalized bias of 263% during the four-month BRAVO period. The mean normalized error for SO_2 is 274% and the correlation coefficient is 0.13. The performance of the $SO_4^=$ simulations is better, with a mean normalized bias of 19%, a mean normalized error of 65%, and an R^2 of 0.38. The TS predictions are also too high. A more detailed performance evaluation of the REMSAD simulation, including time series analyses at available CASTNet monitors, is provided in the CIRA/NPS report on the BRAVO Study (Schichtel et al., 2004), which is included in the Appendix.

Table 9-15. Consolidated performance statistics over the full BRAVO Study period for REMSAD simulations of SO_2, sulfate, and total sulfur at 37 BRAVO network measurement locations.

	37 BRAVO Sites		
	SO_2	**$SO_4^=$**	**TS**
Average Observed (µg/m^3)	1.67	3.10	1.83
Average Simulated (paired) (µg/m^3)	3.29	3.32	2.70
R^2	0.13	0.38	0.25
Regression Slope	0.41	0.84	0.64
Intercept (µg/m^3)	2.61	0.72	1.53
Mean Normalized Error (%)	274	65	95
Mean Normalized Bias (%)	263	19	77
Fractional Error (%)	89	56	56
Fractional Bias (%)	75	-10	33

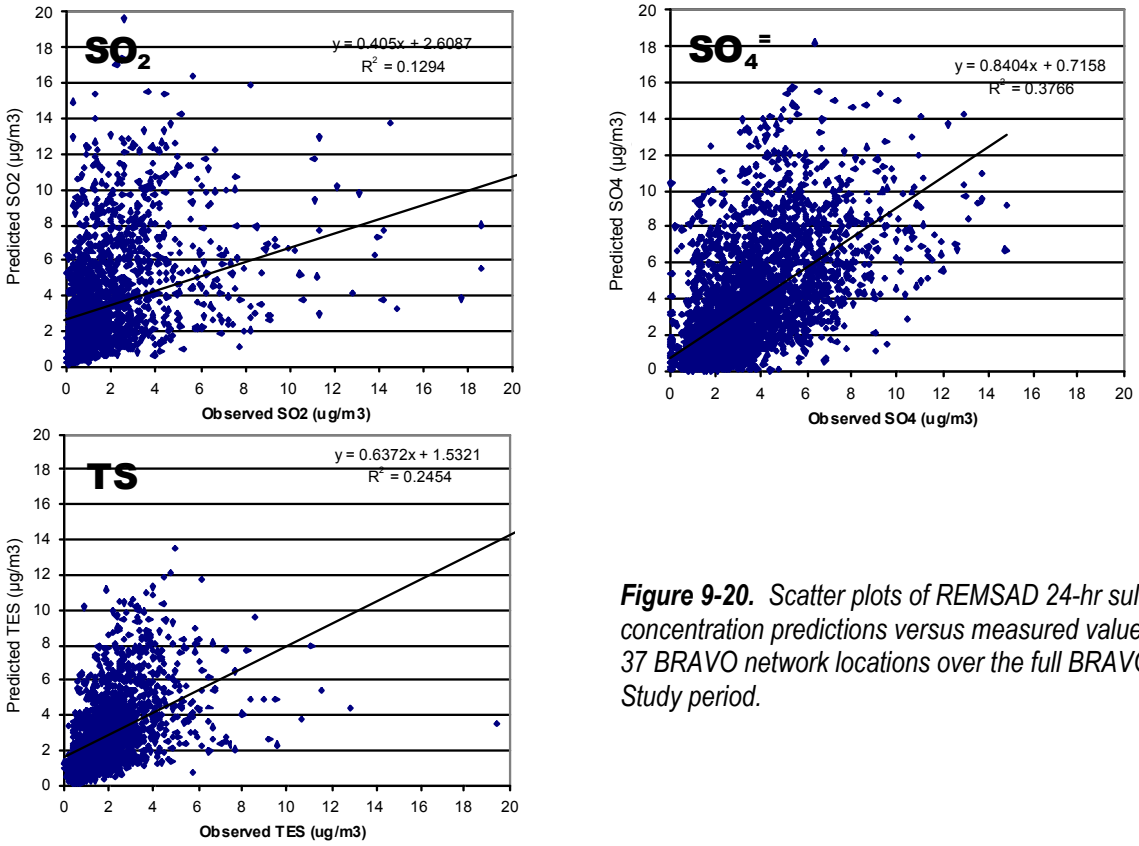

Figure 9-20. *Scatter plots of REMSAD 24-hr sulfur concentration predictions versus measured values at 37 BRAVO network locations over the full BRAVO Study period.*

9.10 Evaluation of CMAQ Simulations of Perfluorocarbon Tracers

The dispersion simulation capability of the CMAQ three-dimensional Eulerian air-quality model was tested by simulating the transport of the inert perfluorocarbon tracers from specific point sources to receptor sites. For the inert tracer simulation, the chemistry, aerosol, and dry deposition modules were disabled. Meteorological fields were generated by processing the MM5 output with MCIP version 2.1 and a four-month simulation was conducted from July 2 to October 30, 1999. The approach and results are summarized here. For details see the EPRI report (Pun et al., 2004), which is contained in the Appendix.

For these tracer simulations, CMAQ was configured with the Bott advection scheme for horizontal and vertical advection, the eddy diffusion scheme for vertical diffusion, and the non-standard Smagorinsky scheme for horizontal diffusion. This is the same configuration that was used subsequently for the aerosol attribution modeling, as described in Section 8.4.4. CMAQ was initially evaluated against the tracer measurements using the horizontal diffusion scheme that is standard in the model and the alternative Smagorinsky scheme (which is the scheme used in REMSAD) that produces greater diffusion. Since the Smagorinsky scheme gave slightly better model performance when compared to the tracer measurements (Pun et al., 2003), it was selected for use in the subsequent aerosol modeling

for source apportionment and thus is the only configuration whose performance will be addressed here.

The four continuously-released perfluorocarbon tracers used for this evaluation were released from Eagle Pass (oPDCH), San Antonio (PDCB), Houston (PTCH), and northeast Texas (i-PPCH), as described in Section 3.2. From early July to mid September, PTCH and PDCB "timing" tracers were released intermittently from Eagle Pass in addition to the continuous tracer. As was shown in Figure 3-2, six monitoring stations – San Vicente, K-Bar, Persimmon Gap, Marathon, Fort Stockton and Monahans Sandhills – measured tracer concentrations near and north of Big Bend with 6-hour samples during the study.

The time series of the measured and simulated concentrations at K-Bar for the four continuously-released tracers are shown in Figures 9-21 to 9-24. The concentrations in these plots are in parts per quadrillion (ppq), which is equivalent to femtoliters per liter (fl/l).

For the northeast Texas tracer, "negative" or zero concentrations were observed during much of the 4-month period, as shown in Figure 9-21. Indeed, the average concentration was slightly negative at K-Bar, which suggests that the estimated background concentration that was subtracted from the laboratory-measured concentration may have been too large. Only a few measurements were observed where the tracer signal was clearly distinguishable from the noise. The highest observed concentrations (~0.2 ppq) occurred in the last part of October and several lower peaks (0.05 to 0.1 ppq) were observed in July,

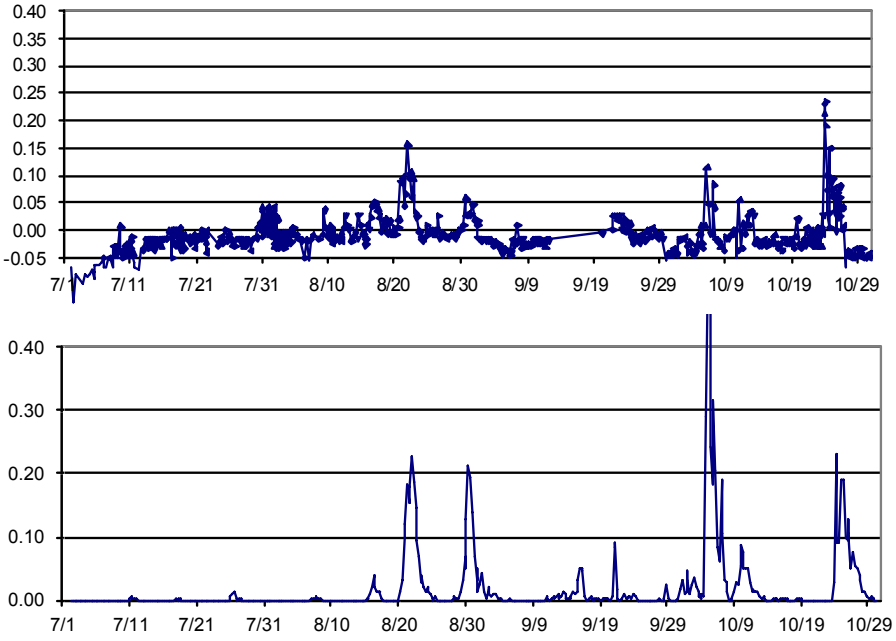

Figure 9-21. *Measured (top) and CMAQ-simulated (bottom) time series of NE Texas tracer concentrations (in ppq) at K-Bar.*

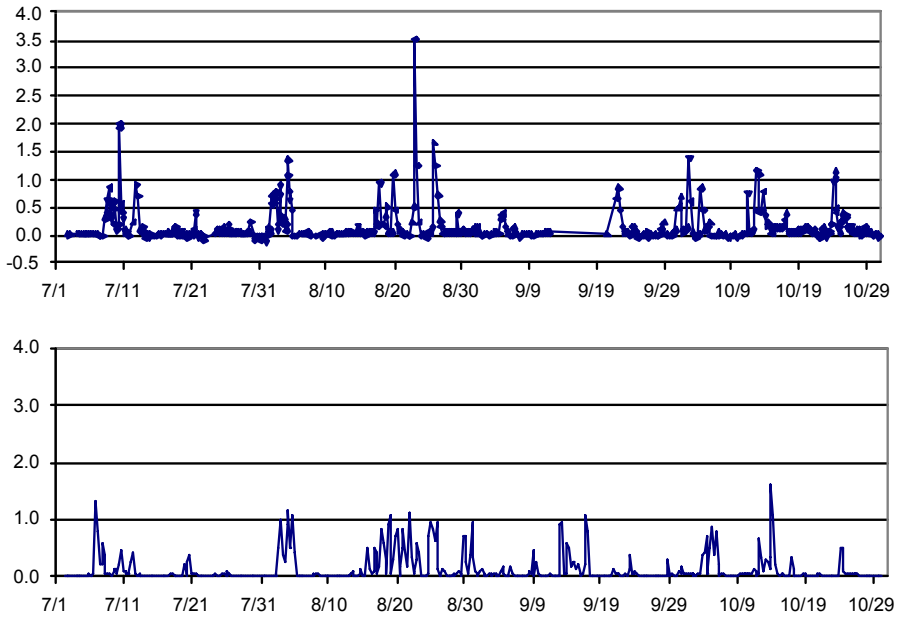

Figure 9-22. *Measured (top) and CMAQ-simulated (bottom) time series of Eagle Pass PDCH tracer concentrations (in ppq) at K-Bar. Note that the concentration scale here is 10 times as large as that in Figure 9-21.*

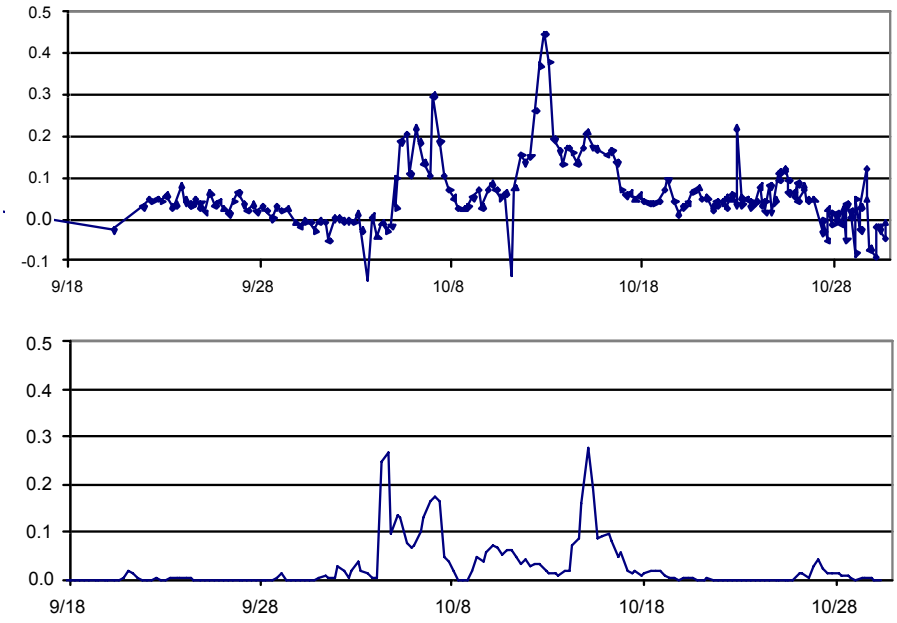

Figure 9-23. *Measured (top) and CMAQ-simulated (bottom) time series of Houston tracer concentrations (in ppq) at K-Bar.*

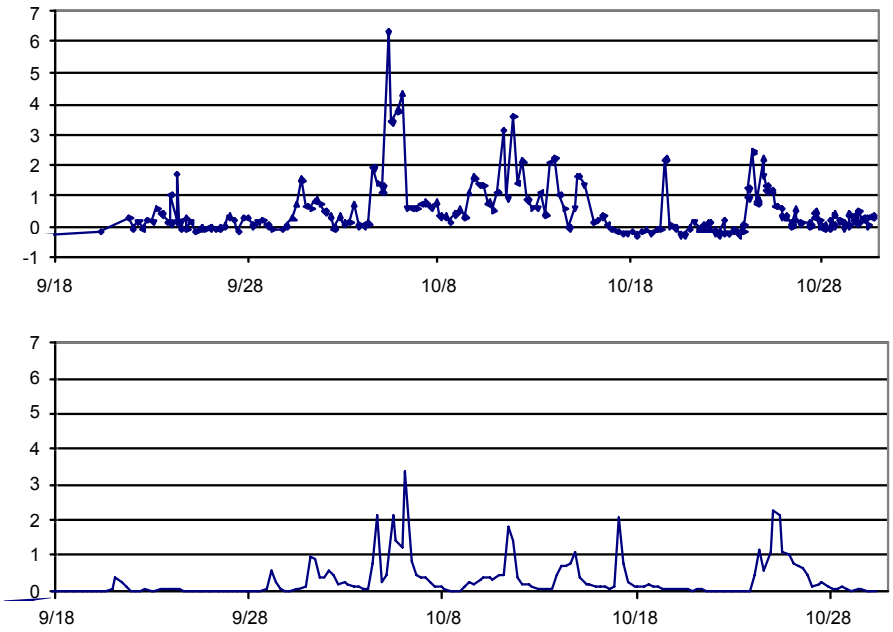

Figure 9-24. *Measured (top) and CMAQ-simulated (bottom) time series of San Antonio tracer concentrations (in ppq) at K-Bar.*

August, and October. CMAQ was able to reproduce the timing of all but the July 31 tracer peak. However, except for the late October peak, the model overestimated tracer concentrations. CMAQ also predicted concentrations in excess of 0.4 ppq on October 5, when significant measured concentration was not observed.

The continuous Eagle Pass tracer, PDCH, whose release location was the closest to Big Bend, was observed at Big Bend in several clusters, as shown in Figure 9-22. Several matching clusters were predicted by CMAQ, although the magnitudes of the predicted concentrations were not always consistent with the observed values. For example, the highest simulated concentrations were approximately 1 ppq whereas a peak in the observations was observed at 3.5 ppq. In addition, the model also predicted concentrations near 1 ppq at times when above-zero concentrations were not observed.

Houston tracer measurements at K-Bar showed two periods of elevated concentrations around October 7 and October 13, as shown in Figure 9-23. CMAQ also predicted elevated concentrations near those dates, although the simulated peaks were early for the October 7 event and late for the October 12 event.

The timing of several predicted San Antonio tracer peaks at K-Bar matched observations, as shown in Figure 9-24. However, a peak was simulated on October 16, but observed concentrations were negligible then. Maximum simulated concentrations were lower than the observed peak concentrations.

The number of tracer measurements that exceeded the analytical uncertainty varied with the choice of tracer and the transport distance. Over all six 6-hour tracer measurement

sites, tracer concentration measurements greater than twice the analytical uncertainty (2σ) accounted for less than 10% and 12% of the measurements for the distant northeast Texas and Houston tracers, respectively. For the San Antonio tracer, the percentages of values greater than 2σ ranged from 15 to 34%, and for the continuous Eagle Pass tracer, the percentages of values greater than 2σ ranged from 16 to 59%. At a specific site, the K-Bar site, the percentages of values greater than 2σ for the San Antonio, Eagle Pass, Houston and northeast Texas tracers were 30%, 22%, 10% and 1%, respectively. Due to this relatively infrequent occurrence of measurable tracer impacts, normalized performance metrics that gauge error between the magnitudes of simulated tracer concentrations and observations (e.g., mean normalized error and mean normalized bias) are not appropriate. Rather, the metrics chosen here to test CMAQ tracer simulation performance relate more to the characteristics of the distribution and timing of the tracer (e.g., ratio of means and coefficient of variation).

Such statistical relationships between the tracer concentrations predicted by CMAQ and measured concentrations were calculated for the six 6-hour tracer monitoring sites. As an example, statistics for the K-Bar site in Big Bend National Park are summarized in Table 9-16. Major differences at other sites from the behavior at K-Bar are mentioned below. Tables for all six sites are in the EPRI Modeling Report (Pun et al., 2004), which is included in the Appendix.

Table 9-16. Performance of CMAQ at simulating tracer concentrations at K-Bar for all tracers released during the study.

	Eagle Pass PDCH	NE Texas PPCH	Houston PTCH	San Antonio PDCB	Eagle Pass PTCH	Eagle Pass PDCB
Observed Mean, ppq	0.14	-0.005	0.07	0.67	0.12	0.52
Predicted Mean, ppq	0.13	0.02	0.04	0.31	0.09	0.78
Ratio of Means	0.96	-4.30	0.49	0.46	0.77	1.52
Std. Dev. (Observed), ppq	0.30	0.03	0.09	1.01	0.35	2.49
Std. Dev. (Predicted), ppq	0.25	0.06	0.06	0.53	0.20	1.86
Coeff. of Variation (Obs)	2.19	-5.85	1.26	1.51	2.98	4.84
Coeff. of Variation (Pred)	1.89	3.19	1.62	1.71	2.16	2.37
RMS Error	0.33	0.07	0.10	0.88	0.36	2.78
RMSE/Observed Mean	2.37	-13.88	1.35	1.30	3.03	5.40
Coeff. of Correlation (r)	0.31	0.36	0.33	0.63	0.25	0.22

These statistics show that concentrations of the continuous tracer (PDCH) released at Eagle Pass were generally understated by CMAQ at all 6-hour monitoring sites, especially at the four northernmost sites. Most correlation coefficients (r) were about 0.3, except for a lower value (0.19) at Persimmon Gap. The best mean prediction was obtained at the K-Bar

site, where the ratio of mean predicted to mean observed concentrations approached unity (0.96).

The model performance for the timing tracers PTCH and PDCB released from Eagle Pass exhibits a greater degree of variability. (Recall that these two tracers were released only during the first half of the study.) For these tracers, the ratios of mean simulated concentrations to mean observed concentrations were as high as 1.52 for the PDCB tracer at K-Bar and as low as 0.31 for PTCH at Persimmon Gap. The timing tracer correlation coefficients (r) varied from a low of 0.15 for PDCB at Monahans Sandhills (the most northerly site) up to 0.40 and 0.42 for the two timing tracers at Fort Stockton and 0.36 and 0.41 for the two tracers at Marathon.

For the northeast Texas tracer (PPCH), the measured standard deviations compared to the average observed concentrations (the coefficient of variation in Table 9-15) are quite large since high concentrations when the plume directly "hits" the monitoring site were observed infrequently. Low average concentrations and high standard deviations are also found at the other Big Bend sites. The correlation coefficients (r) for this tracer ranged from 0.25 to 0.55, depending on site -- surprisingly good values considering the low signal levels of the measurements.

The correlation coefficients for the Houston tracer ranged from 0.14 to 0.33, depending on site, and those for the San Antonio tracer ranged from 0.38 to 0.63. The correlation of the San Antonio tracer measurements and predictions was typically better than that of the Houston tracer, although the trend was not consistent for other performance metrics, such as the ratio of means.

The CMAQ tracer simulations reveal that the simulation of transport processes alone can contribute significant uncertainty in the overall application of an air pollution model. The performance of the model here is indicative of the difficulty inherent in the simulation of narrow plumes from single point sources in 3-D models. Small errors in the dispersion algorithms or in the wind fields can cause the simulated plume to miss the receptor when the actual tracer reached a receptor, or vice-versa.

It was found that coarse grid scales provide better performance for predicting the transport – with respect to distribution and timing – of the inert BRAVO tracers. Simulated plumes are generally more dispersed with a coarse grid and thus show higher probability of impacting the monitoring site as the plume travels to the vicinity of the monitor. This is partly due to the artificial dilution that occurs as a point source emission is instantly dispersed over the area of a grid cell. (Diffusion in three-dimensional air-quality models results from both turbulent diffusion and the artificial numerical diffusion.) In addition to the artificial dilution that occurs in a grid cell, differences in the wind speed and wind direction in the meteorological input used to drive the transport modules can contribute to differences in the location of the tracer plumes.

It is reasonable to infer that the simulation of inert tracers from isolated point sources may require a higher accuracy in meteorological inputs than the simulation of regionally distributed pollutant emissions because of the more pronounced gradients in the concentrations of the point source tracer. While higher spatial resolutions reduce

uncertainties in the numerical representation of various physicochemical processes in Eulerian models (e.g., plume chemistry, cloud effects), finer resolutions may not always yield better results in situations where transport plays a dominant role in the distribution of the simulated species of interest.

The BRAVO study provided an opportunity to evaluate the transport components of Eulerian models independently of chemical processes. For the San Antonio tracer, CMAQ was able to explain 40% of the variance at K-Bar. For the other tracer releases, the CMAQ model was only capable of explaining 5 to 13% of the variance (as estimated by the coefficient of determination, r^2) there. It appears that, for inert gases at least, the CMAQ modeling system is restricted in its ability to represent transport and diffusion from a single point source to specific receptors (a conclusion that applies also to REMSAD).

9.11 Evaluation of CMAQ Simulations of SO_2, Sulfate, and Other Aerosol Components

For particulate matter simulations, the MADRID aerosol module was incorporated into the CMAQ model, as described in Section 8.4.3. The performance of the resulting model, called CMAQ-MADRID here, is discussed below. (More detail is provided in the EPRI report by Pun et al. (2004), which is included in the Appendix.)

9.11.1 Development of Base Case

For the CMAQ-MADRID modeling, the boundary conditions, emissions inputs, and model configuration that were used for the source attribution estimates (to be described in Section 11.2) were established after evaluating the model's performance in preliminary and sensitivity simulations. We summarize these initial simulations and their results here and describe the performance of the resulting modeling system. (The initial simulations and performance are described in detail in the EPRI report (Pun et al., 2004).)

The preliminary simulation used the BRAVO emissions inventory that was described in Chapter 4. Boundary conditions were derived from the REMSAD modeling results at the boundary of the final CMAQ domain. The performance of this simulation was appraised by comparing 24-hour predictions of several species concentrations with observations throughout the BRAVO network.

Model performance for this preliminary simulation was not satisfactory for SO_2 and sulfate over the eastern half of the CMAQ domain. To improve model performance, two potential causes for underestimates of sulfur concentrations were explored – possible understatement of Mexican SO_2 emissions and possible incorrect boundary conditions resulting from overstatement of SO_2 and sulfate along the boundaries of the CMAQ domain.

Possible understatement of Mexican SO_2 emissions was suggested by systematic underestimation of sulfate at K-Bar during periods in July when there was a predominance of southerly winds. This hypothesis was reinforced by a September 2002 presentation by the Undersecretary of the Mexican Ministry of Energy, titled "Mexico and USA Power Plant Emissions in Perspective." that presented total SO_2 emissions for three Mexican states bordering the United States, which constitute a large portion of the CMAQ modeling domain.

Comparing SO_2 emissions from all sources other than the *Carbón* power plants, the BRAVO estimate (225,000 tons/year) disagreed with the estimate of the Mexican Ministry of Energy (398,000 tons/year), a ratio of about 1.8.

Incorrect boundary conditions for CMAQ could result from the overestimation of SO_2 and sulfate by REMSAD along the boundaries of the CMAQ domain, as identified in Section 9.9. Specifically, compared to measurements, REMSAD overestimated sulfate concentrations along the CMAQ domain boundary by approximately 40% and SO_2 concentrations there by a factor of about 3.8.

For one model configuration that explored the combined effects of these factors, emissions of SO_2 and sulfate within the Mexican portion of the CMAQ domain (except emissions from the *Carbón* power plants) were increased by a factor of two and the REMSAD-based lateral boundary conditions for SO_2 and sulfate were scaled with CASTNet and IMPROVE observations. When these adjustments were made, model performance for SO_2 and sulfate improved over that of the preliminary base case simulation, with bias and error reduced for both species throughout the BRAVO network, as shown in Table 9-17. In the end, this simulation was selected as the CMAQ-MADRID base case simulation and served as a basis for estimating the source attributions that will be presented in Chapter 11.[1] Consequently, the rest of the discussion below will address the performance of this configuration of CMAQ-MADRID.

Table 9-17. Network-average CMAQ-MADRID SO_2 and fine SO_4^{2-} performance statistics for preliminary and final base case simulations.

Statistic	Units	Weekly SO_2		24-hour fine SO_4^{2-}	
		Prelim. Base Case	Final Base Case	Prelim. Base Case	Final Base Case
Observed mean	µg/m³	1.48	1.48	3.1	3.1
Simulated mean	µg/m³	1.74	1.61	4.67	3.97
Coefficient of Determination (r^2)		0.45	0.46	0.47	0.47
Mean bias	µg/m³	0.26	0.12	1.57	0.87
Mean error	µg/m³	0.69	0.60	2.24	1.67
Mean Normalized Bias	%	48	40	57	37
Mean Normalized Error	%	71	64	82	65
Fractional Bias	%	15	20	18	8
Fractional Error	%	44	43	54	47
RMS error	µg/m³	1.09	0.97	3.58	2.65

[1] The CMAQ-MADRID base case simulation covers the time period July 2 – October 28, which is slightly shorter than the period of the REMSAD simulation. This causes slight variations in statistical data based on observations, e.g. mean observed concentrations, compared to the corresponding values in the tables shown in Section 9.9 in conjunction with the evaluation of REMSAD performance.

9.11.2 Base Case Performance

Focusing now solely on the new base case configuration, Table 9-18 shows statistics illustrating the performance of the CMAQ-MADRID base case simulation in simulating 24-hr fine particulate matter component concentrations at the K-Bar site.

As indicated by the coefficient of determination (r^2), the model accounts for 52% of the variance in fine sulfate mass and only 23% of the variance in total fine particulate mass at Big Bend. The mean normalized error in the predictions for all species is greater than 50%; however, the mean normalized error is species-dependent and ranges from 51% and 55% for ammonium and sulfate to 96% for nitrate at the K-Bar site.

Model performance over multiple monitoring locations in the BRAVO Study network is summarized in Table 9-19. Note that the number of monitoring locations differed for different components. (Ammonium and nitrate concentrations were measured only at Big Bend and thus are not included in this table.)

Thirty-seven stations measured fine sulfate and total fine particulate matter ($PM_{2.5}$). Consistent with the behavior at Big Bend, Table 9-19 shows that the model exhibited positive bias (both absolute and relative) in its predictions of fine particulate sulfate, together with negative absolute bias and positive normalized (relative) bias in its predictions of total fine particulate mass.

Table 9-18. CMAQ-MADRID performance for 24-hr averages of particulate matter components at K-Bar over the duration of the BRAVO study.[2]

Statistic	units	SO_4^{2-}	NO_3^-	NH_4^+	EC	OM	$PM_{2.5}$
Observed mean	μg/m³	2.49	0.23	0.82	0.23	1.25	6.33
Simulated mean	μg/m³	2.61	0.03	0.7	0.05	0.45	4.61
Coeff. of determination (r^2)		0.52	0.01	0.47	0.13	0.1	0.23
Mean bias	μg/m³	0.12	-0.19	-0.12	-0.18	-0.8	-1.72
Mean error	μg/m³	1.08	0.21	0.35	0.18	0.81	3.09
Mean Normalized Bias	%	20	-79	7	-76	-59	7
Mean Normalized Error	%	55	96	51	76	62	70
Fractional Bias	%	0	-166	-12	-135	-98	-31
Fractional Error	%	46	173	48	135	100	62
RMS error	μg/m³	1.51	0.24	0.45	0.14	1.08	4.05

[2] A threshold of 0.1 μg/m³ was used for all species in such statistical tables throughout this section, except for Table 9-22.

Table 9-19. CMAQ-MADRID performance for 24-hr averages of particulate matter components over all BRAVO sites over the duration of the BRAVO study.

Statistic	units	SO_4^{2-}	EC	OM	$PM_{2.5}$
Observed mean	µg/m³	3.10	0.17	2.36	9.92
Predicted mean	µg/m³	3.97	0.15	1.22	9.54
Coeff. of determination (r^2)		0.47	0.03	0.61	0.21
Mean bias	µg/m³	0.87	-0.03	-1.14	-0.39
Mean error	µg/m³	1.67	0.12	1.23	5.04
Mean Normalized Bias	%	37	4	-50	16
Mean Normalized Error	%	65	73	55	65
Fractional Bias	%	8	-34	-78	-56
Fractional Error	%	47	81	82	7
RMS error	µg/m³	2.65	0.2	1.55	7.06
Number of Stations		37	6	6	37

Organic and elemental carbon concentrations were measured at only six stations. Their locations traversed Texas, however, and thus should be fairly representative of the entire network. The model showed a small bias in EC prediction, but the miniscule correlation between predictions and observations shows that this apparent agreement is not reflected on a day-to-day basis. Organic matter (corresponding to OC x 1.4 for measured organic carbon) yielded a coefficient of determination of 0.61, but there is a consistent underestimation across the network.

Overall, Tables 9-17 and 9-18 show that the mean error in the CMAQ-MADRID modeled estimates of all particulate matter components is greater than 50%, both at Big Bend and when averaged over the entire network. Furthermore, Table 9-16 showed that the model is able to explain roughly half or less of the variance in both SO_2 and sulfate concentrations. The predictive performance for total sulfur (not shown here) is comparable, so the errors are not due solely to limitations in the simulation of the conversion of SO_2 to sulfate.

Sulfur Compounds. The coefficient of determination, r^2, for all modeled and measured 24-hr sulfate concentration values over the entire network was 0.47 and a scatter plot of this data shows large scatter, although the slope of the linear regression line is very close to unity.

In order to explore the scatter, the spatial distributions of the mean normalized bias and mean normalized error of the sulfate estimates are plotted in Figure 9-25. The mean normalized error is a positive number greater than or equal to the normalized bias. Regions with large normalized bias and a similarly large mean normalized error represent areas of consistent overestimates. At the eight stations that exhibited the highest bias in predicted sulfate normalized to observations (53-130%), the normalized bias represented a significant portion (on average ~80%) of the mean normalized error (66-159%). Of these stations, seven are located in eastern Texas, showing a marked propensity of the model to overstate sulfate

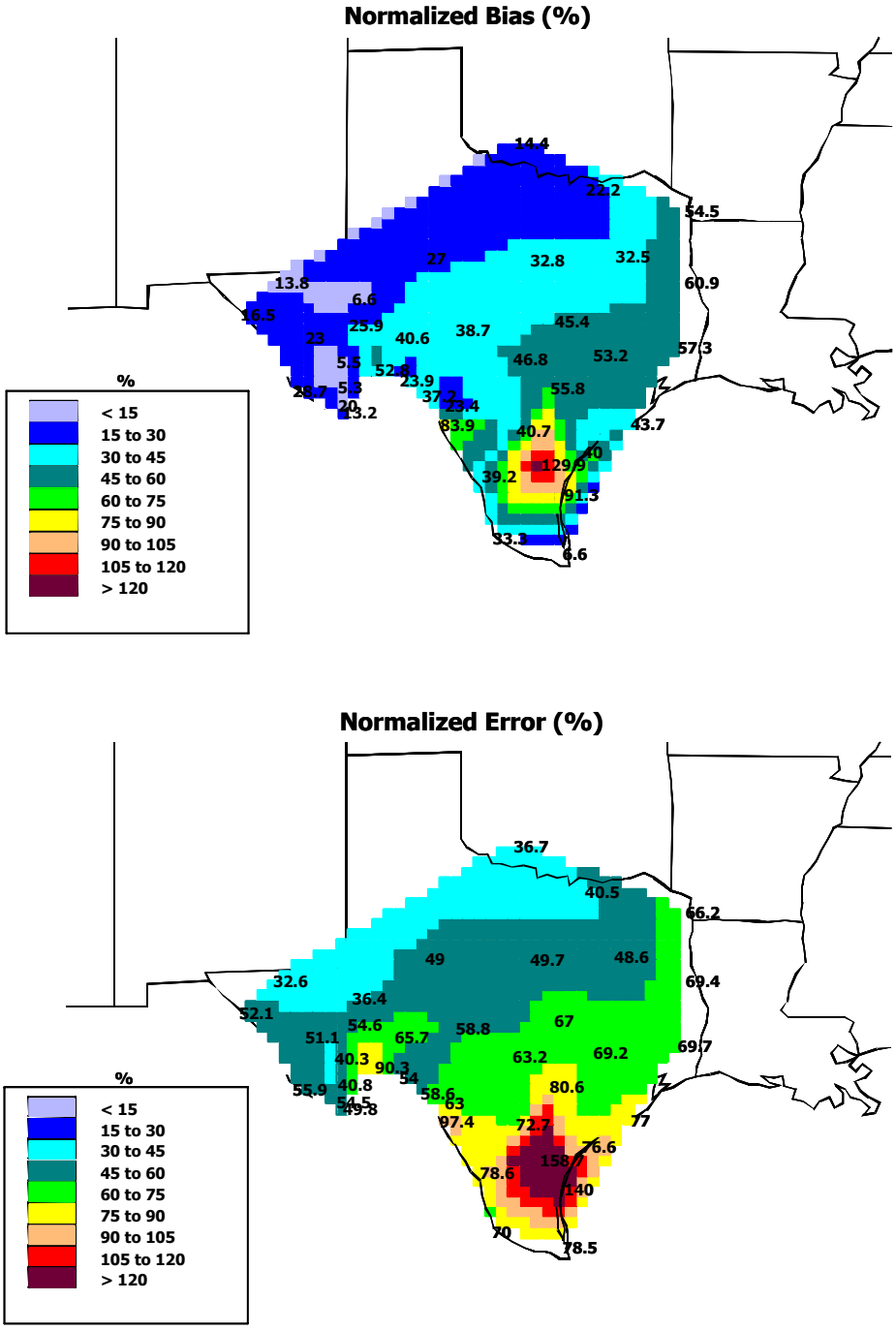

Figure 9-25. *Spatial distributions of mean normalized bias and mean normalized error in CMAQ-MADRID predictions of fine sulfate over the BRAVO network.*

concentrations in the eastern portion of the domain.[3] On the other hand, at the twelve stations where the mean normalized bias represented a lesser portion (on average ~30%) of the mean normalized error (33-78%), eleven are located in the vicinity of Big Bend National Park and/or near the US-Mexico border, showing a tendency of the model to underestimate more often in these regions compared to other regions of the domain.[4]

Particularly severe mean normalized bias and error are apparent in Figure 9-25 at two locations in southern Texas, quite distant from Big Bend. Since poor model performance in southern Texas is not necessarily indicative of poor model performance at Big Bend, the impact of these two locations on the statistics in Table 9-19 was explored. Since Table 9-19 is based on 37 stations, removing two stations would not be expected to have a large effect, and this turns out to be true. Specifically, for sulfate the effect is to slightly reduce the mean concentration to 3.91 $\mu g/m^3$, increase r^2 to 0.49, and slightly decrease the bias and the error metrics. The biggest relative change is a decrease in mean normalized error from 65% to 61%. The effects on $PM_{2.5}$ statistics are even smaller.

For the duration of the BRAVO study and across the entire BRAVO Network, particulate sulfate accounted for 62% of the observed total sulfur concentrations on average, comparable to a sulfate fraction of 61% for the predicted total. About 3/4 of the predicted sulfate fractions lie within 20 percentage points of the observed values. Both data sets show that high sulfate concentrations tend to be associated with high sulfate fractions.

This discussion indicates that model performance for sulfur differed in the eastern and western parts of the modeling domain. We can explore this difference by comparing performance at the K-Bar and Big Thicket stations. The K-Bar station at Big Bend National Park, apart from being the focal station of the study, is representative of the region where the model was more likely to understate concentrations of sulfur species. The Big Thicket station in eastern Texas is representative of the eastern portion of the domain, where the model showed high positive bias in simulated concentrations of sulfur species due to consistent overestimates in the region. Tables 9-20 and 9–21 provide performance statistics for these two locations.

[3] The predicted signal at the eighth station, Eagle Pass in southern Texas near the Mexican border, is affected detrimentally by inclusion of the *Carbón* power plants in the same model grid cell.

[4] At the twelfth station, Wichita Mountains, less than 30% of the potential data set was valid, which precluded any significant analysis.

Table 9-20. CMAQ-MADRID performance for sulfur species at the K-Bar site, BBNP, over the duration of the BRAVO study.

Statistic	Units	24-hr SO_2	Weekly SO_2	24-hr SO_4^{2-}	24-hr Total S	Weekly Total S
Observed mean	µg/m³	0.88	0.87	2.49	1.25	1.30
Predicted mean	µg/m³	0.85	0.87	2.61	1.33	1.41
Coeff. of determination (r^2)		0.03	0.29	0.52	0.30	0.38
Mean bias	µg/m³	-0.04	-0.01	0.12	0.08	0.11
Mean error	µg/m³	0.72	0.33	1.08	0.64	0.39
Mean Normalized Bias	%	85	16	20	34	9
Mean Normalized Error	%	134	42	55	64	28
Fractional Bias	%	0	5	0	8	4
Fractional Error	%	72	38	46	48	27
RMS error	µg/m³	1.24	0.46	1.51	0.95	0.50

Table 9-21. CMAQ-MADRID performance for sulfur species at the Big Thicket site over the duration of the BRAVO study.

Statistic	Units	24-hr SO_2	Weekly SO_2	24-hr SO_4^{2-}	24-hr Total S	Weekly Total S
Observed mean	µg/m³	1.03	1.04	4.18	1.91	1.94
Predicted mean	µg/m³	2.56	2.62	6.44	3.40	3.48
Coeff. of determination (r^2)		0.28	0.65	0.63	0.43	0.16
Mean bias	µg/m³	1.53	1.58	2.26	1.48	1.54
Mean error	µg/m³	1.73	1.63	2.67	1.60	1.54
Mean Normalized Bias	%	239	149	57	95	81
Mean Normalized Error	%	255	154	70	101	81
Fractional Bias	%	63	73	32	50	53
Fractional Error	%	88	79	47	58	53
RMS error	µg/m³	2.66	2.12	3.66	2.23	1.86

As illustrated by the time series in the top graph in Figure 9-26, the model is unable to capture the evolution of SO_2 concentrations at Big Bend, as reflected by the low coefficient of determination ($r^2 = 0.03$) given in Table 9-20. The model fails to capture the two principal peak SO_2 events and predicts only one of the three lesser peak SO_2 events. The model also predicts five sharp SO_2 peaks that were not observed in the field. Over all, the model generally understated high concentration events and overstated periods of typically low concentrations.

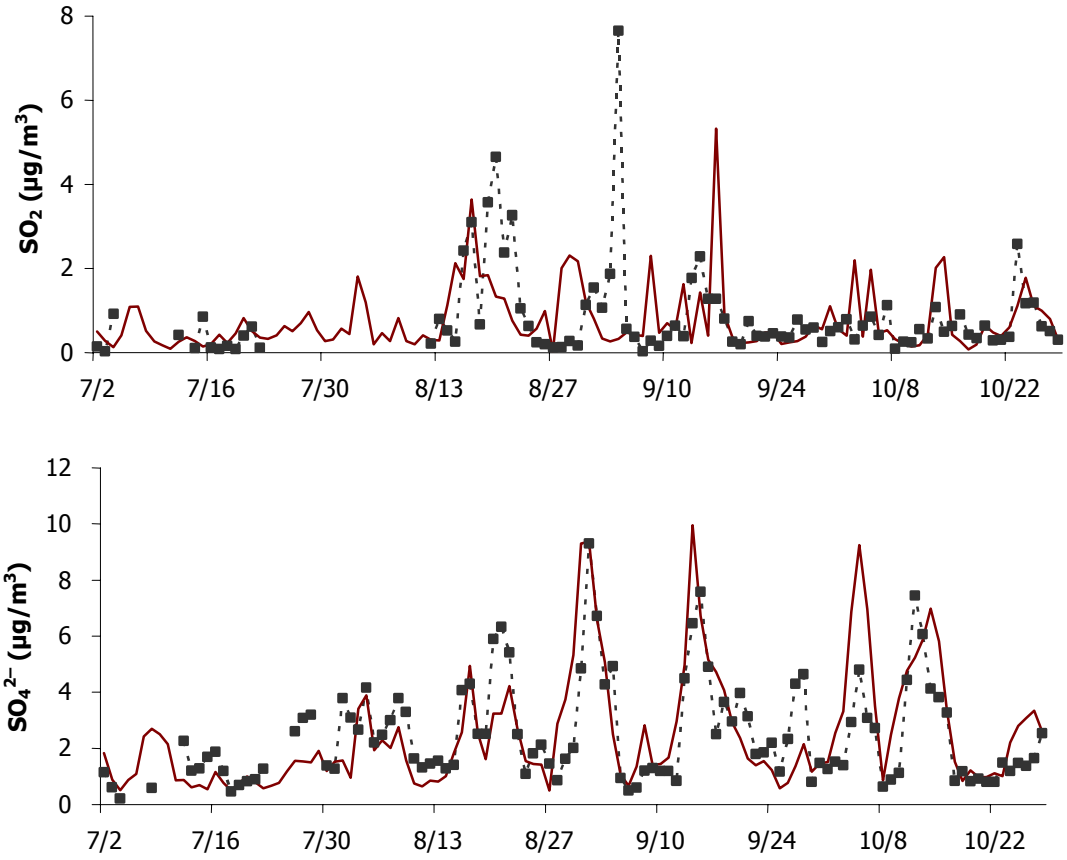

Figure 9-26. *Measured and simulated 24-hr SO_2 (top) and sulfate (bottom) concentrations at K-Bar. The points represent the measurements and the line represents the simulated concentrations.*

Table 9-20 also indicates that fine particulate sulfate concentrations at Big Bend are simulated by the CMAQ-MADRID modeling system with a lower bias and normalized error than the estimates for SO_2. The model can explain approximately half of the variance in fine sulfate observations (i.e., r^2=0.52 – nearly double the value of the coefficient of determination for SO_2) – and this is reflected in the qualitatively better correspondence in the bottom graph in Figure 9-26.

For total sulfur (the sum of the sulfur in gas-phase SO_2 and fine particulate sulfate), the model's simulation of the presence of peaks and valleys is better than for either SO_2 or sulfate, although temporal offsets of one or two days are common. The coefficient of determination, r^2, for 24-hr TS is much better than that for SO_2, but it is still not as good as that for sulfate. In order to illustrate the influence of minor temporal shifts in the predictions compared to observations, weekly averages of total sulfur concentrations were compared for the twelve weeks for which observational data could be computed. Of the twelve weeks, weekly predicted total sulfur values were within ±10% of the observations for four weeks.

Overestimates and underestimates were relatively evenly distributed over the other weeks, which resulted in the small average bias given in the last column of Table 9-20.

Turning now to the Big Thicket Station in the eastern part of the domain, performance statistics were given in Table 9-21. For all species and time scales of comparison, the mean normalized bias corresponds to a large portion of the normalized error, indicating the model rarely underestimated the measured concentrations. On average, the model's SO_2 concentration estimates are 3.4 times as great as the observations.

Daily SO_2 estimates are plotted alongside concentrations measured at Big Thicket in Figure 9-27. The model overestimates on 64 of the 75 days with valid observations and the coefficient of determination is low (r^2 = 0.28). However, as shown in Table 9-21, model performance improves substantially when weekly SO_2 concentrations were compared, with the r-squared value increasing to 0.65, although weekly SO_2 concentrations were estimated in almost all weeks.

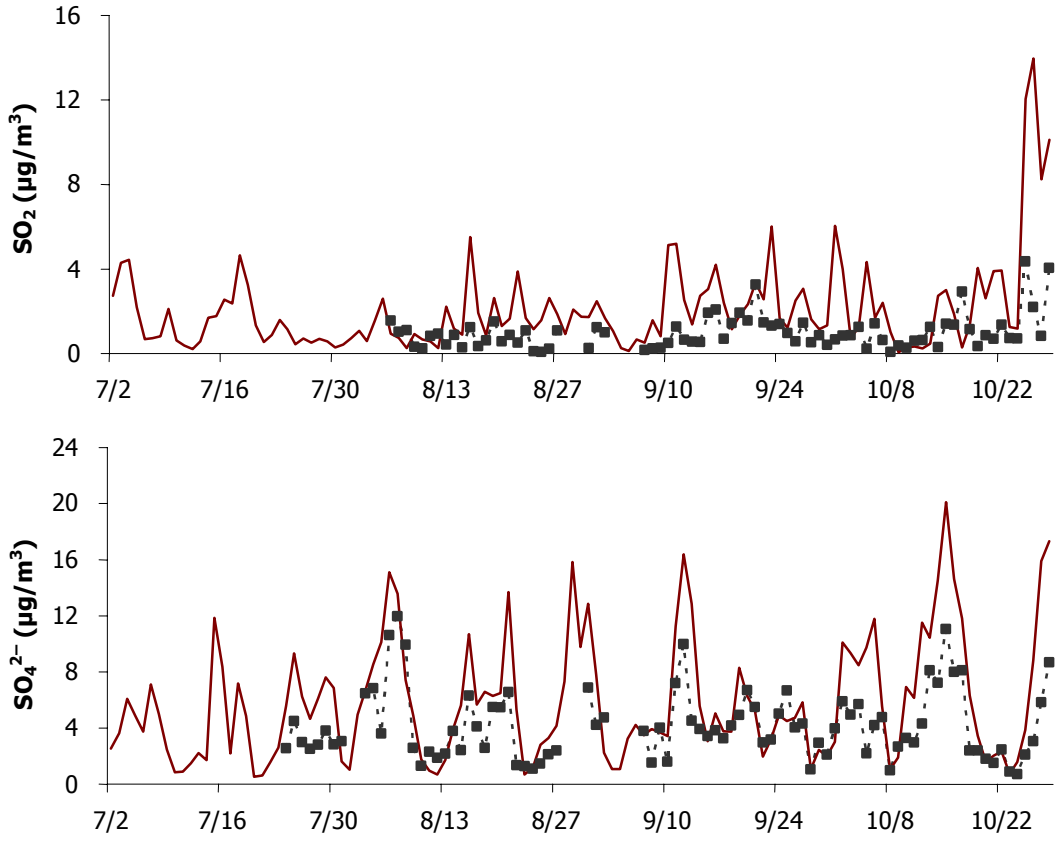

Figure 9-27. *Measured and simulated daily SO_2 concentrations (top) and fine sulfate concentrations (bottom) at Big Thicket. The points represent the observations and the line represents the simulated concentrations.*

For sulfate, the model overestimated the daily sulfate load at Big Thicket on 64 of the 88 days with valid measurement data. The statistics in Table 9-21 indicate that the model was able to capture daily fluctuation in fine particulate sulfate concentrations there, but it regularly overestimated their values by approximately 50%.

Daily and weekly total sulfur fluctuations at Big Thicket were fairly well captured by the model, albeit with regular overestimations of peaks at twice the observed values, but the model failed to represent the lower and more consistent longer-term average concentrations.

As was indicated in Table 9-20, the correlation between total sulfur predictions and observations deteriorates when weekly averages are considered, although the bias decreases. The weekly total sulfur concentrations observed at Big Thicket ranged from 1.2 to 2.4 $\mu g/m^3$ S, with a mean of approximately 2 $\mu g/m^3$ S and coefficient of variation (standard deviation/mean) of 0.20; in contrast, predicted total sulfur concentrations for Big Thicket ranged from 2.1 to 6.0 $\mu g/m^3$ S, with a mean concentration of approximately 3.5 $\mu g/m^3$ S and coefficient of variation (standard deviation/mean) of 0.35. Overall, at Big Thicket the ratio of weekly mean predicted total sulfur concentrations to the mean observed total sulfur concentrations was about 1.8.

Other Particulate Matter Components and $PM_{2.5}$ Mass Concentration. The ability of the CMAQ-MADRID model to reproduce daily variations of the components of $PM_{2.5}$ other than sulfate can be evaluated at K-Bar, which is the only location where all five major directly-measurable components of fine particulate matter – ammonium, elemental (black) carbon, nitrate, organic mass and sulfate – were determined. In addition to the five major components and total mass, an "other" category composed of soil elements, sodium and chloride, and all other unidentified species will be used in the following discussion.

Recall that, as shown in Table 9-18, the CMAQ-MADRID model reproduced total $PM_{2.5}$ mass throughout the network with a negative mean bias (-0.39 $\mu g/m^3$) and a positive mean normalized bias (16%).

As for the organic mass, the discussion there showed predictions versus measurements at six stations exhibited a coefficient of determination ($r^2 = 0.61$; see Table 9-19) that indicates that the model could reproduce fairly accurately the fluctuations in organic mass. The model predictions show a mean normalized bias (-50%) nearly equal in magnitude but opposite in sign to the normalized error, indicating that the model regularly predicted approximately half of the organic mass observed at the monitoring stations.[5]

Table 9-22, a modified version of Table 9-18, lists model performance statistics determined from comparing predictions to observations on the 103 days (out of 119 simulation days) when concentrations of all 5 major fine particulate components and total

[5] Consistent with the conventional IMPROVE methodology, a factor of 1.4 was used to convert organic carbon measurements to organic mass in order to account for other atoms such as hydrogen, oxygen and nitrogen. Recent studies suggest that this factor may need to be increased depending on location (Turpin and Lim, 2001), but doing so would further increase the degree of underestimation of organic mass by the model.

Table 9-22. CMAQ-MADRID performance for 24-hr concentrations at the K-Bar site over the duration of the BRAVO study.[6]

Statistic	units	SO_4^{2-}	NO_3^-	NH_4^+	EC	OM	Other	$PM_{2.5}$
Observed mean	μg/m³	2.52	0.20	0.82	0.12	1.27	1.56	6.49
Simulated mean	μg/m³	2.63	0.04	0.68	0.06	0.44	0.80	4.66
Coeff. of determination (r^2)		0.53	0.03	0.50	0.00	0.10	0.02	0.23
Mean bias	μg/m³	0.11	-0.16	-0.13	-0.06	-0.82	-0.76	-1.83
Mean error	μg/m³	1.08	0.20	0.34	0.09	0.84	1.47	3.13
Mean Normalized Bias	%	19	26	6	8	-60	-45	-13
Mean Normalized Error	%	54	194	51	89	6	160	51
Fractional Bias	%	-1	-154	-13	-13	-116	163	-31
Fractional Error	%	46	169	48	19	121	282	87
RMS error	μg/m³	1.51	0.24	0.45	0.14	1.10	2.37	4.11

particle mass were measured at the K-Bar site. The "Other" component represents the difference between the measured $PM_{2.5}$ mass concentration and the sum of the major components; this "other" component is likely to contain dust, crustal material, metal oxides, and sea-salt and may also represent water remaining in the sample at the time of weighing. In addition, since some components of $PM_{2.5}$ may volatilize during sample handling and measurement, the "other" component could also reflect negative contributions. Finally, "other" bears the burden of absorbing all of the measurement uncertainties of the other components.

The stacked bar charts in Figure 9-28 illustrate the composition information in Table 9-22. Over the analysis period, the mean $PM_{2.5}$ mass is underestimated by a factor of 1.4, equivalent to an underestimate of 1.83 μg/m³. On the other hand, the average simulated absolute concentrations of sulfate and ammonium agree quite well with the measured concentrations. However, due to the underestimation of all non-sulfate components, the model estimates a higher contribution of fine sulfate to $PM_{2.5}$ mass. The simulated proportional contributions of the various other components of $PM_{2.5}$ are also quite different from the measured proportions, except for ammonium (NH_4^+) and elemental carbon.

The difference between model estimates and observations of total fine particle mass cannot be explained solely by discrepancies in model predictions for sulfate. CMAQ-MADRID predicted a much tighter relation between fine sulfate and $PM_{2.5}$ mass ($r^2 = 0.91$)

[6] Statistics in Table 9-22 were computed without the lower cutoff value of 0.1 μg/m³ used in previous tables in this section, and thus these statistics test model performance at lower concentration levels than those that apply for Table 9-17. Also, only those days with valid measurements for all components, and thus with the ability to derive a value for "Other", were used.

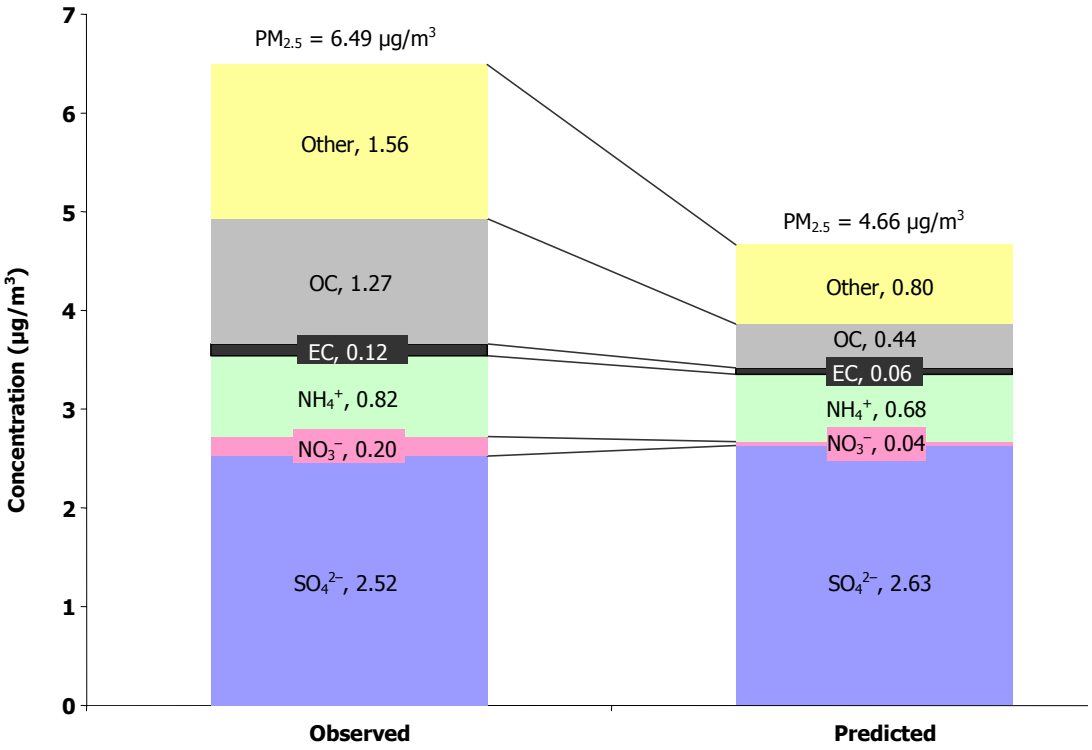

Figure 9-28. *Compositions of observed and simulated $PM_{2.5}$ at K-Bar, averaged over the duration of the BRAVO Study.*

than the observations (r^2=0.61). In addition, the model failed to simulate peak $PM_{2.5}$ mass events at times when sulfate was only a minor contributor. Finally, linear regression lines through the observed and modeled data in plots of $PM_{2.5}$ concentration versus sulfate concentration yield similar slopes, but the line representing the simulated results is offset 1.96 µg/m^3 below the one for the measurements. (This separation is consistent with the mean bias of -1.83 µg/m^3 shown for $PM_{2.5}$ in Table 9-22.)

Rather, it turns out that error in simulating the organic mass concentrations is the principal contributor to the underestimation of total fine particulate mass. The average mean simulated organic mass is approximately one-third of the observed organic mass. Even as a proportion of the simulated $PM_{2.5}$ mass concentration, organic mass only accounted for 9.5% of the total predicted $PM_{2.5}$ mass whereas organic material corresponded to 19.5% of the mass in the ambient data.

The contribution of organic material to the $PM_{2,5}$ mass concentration can be significant. On seventeen days, observed organic mass concentrations were greater than those of sulfate. In contrast, the model only predicted fine organic mass greater than *half* of the predicted fine sulfate mass on four occasions. The tendency to underestimate the contribution of organic mass was greatest at higher PM concentrations.

The following additional conclusions result from the evaluation of the CMAQ-MADRID modeling. Details are provided in the EPRI report in the Appendix (Pun et al., 2004).

- The concentrations of "other" components – including mechanically-induced and windblown dust, crustal material, metal oxides, and sea-salt – accounted for approximately 24% of the observed mass during the BRAVO study at Big Bend. Many of these are not represented reliably in the emissions inventory. The model underestimated these "other" components, particularly in the months of July and August when they accounted for large portions of the total fine particulate matter mass.
- Nitrate concentrations at Big Bend were underestimated by the model and their day-to-day variation was not represented well.
- Concentrations of organic compounds and elemental carbon were much more closely correlated with each other in the model simulations than in the observational data set at Big Bend National Park.

9.11.3 Discussion of CMAQ-MADRID Performance

The analyses above showed that, averaged over the BRAVO network and study period, sulfate estimates by the CMAQ-MADRID model were relatively unbiased. The model accounted for approximately half of the fluctuations in sulfate concentrations with a normalized error commensurate with expected performance of current models. However, as is the case with most regional models, the CMAQ-MADRID modeling system showed limitations in predicting temporal variations of sulfate concentrations.

Overestimates at Big Bend National Park, when they occurred, were influenced by a marked positive bias in predictions of both gas-phase and particulate sulfur concentrations throughout the eastern portion of the BRAVO Study domain. The exact causes for overestimation of sulfur species in this portion of the domain are unknown, although these overestimates are coincident with the positive bias in boundary conditions from the eastern United States as prescribed by the REMSAD model.

In contrast, underestimates of sulfur species occurred at the southern and southwestern portions of the BRAVO Network (in the vicinity of Big Bend National Park). At times of underestimation at Big Bend National Park, wind trajectories originated more often from Mexico than they did during periods of overestimation. This evaluation suggests shortcomings in the estimation of emissions of sulfur compounds and in the simulation of the conversion of SO_2 to fine particulate sulfate in the Mexican domain.

Also, concentrations of organic compounds were underestimated at Big Bend. This is relevant because organics constitute the second largest identifiable component of fine particulate matter mass there.

9.12 Synthesis Inversion – Merging Modeled Source Apportionment Results and Receptor Data

As has been described in Section 8.4.4, both the REMSAD and CMAQ-MADRID models were employed to simulate the sulfate concentrations over the BRAVO Study area. The REMSAD model was employed to predict the sulfate concentrations in most of North America, and estimate the contribution of emissions from major U.S., and Mexican source regions to sulfate concentrations at Big Bend and other BRAVO monitoring sites. The CMAQ model was similarly employed, but predicted sulfate over a smaller region centered on Texas and estimated the sulfate contributions from only Texas, Mexico, western US, and eastern US source regions and the boundary conditions. The CMAQ boundary conditions were generated from the REMSAD predicted sulfate concentrations, modified by measured sulfate and sulfur dioxide concentrations from the IMPROVE and CASTNet monitoring networks.

There were some differences in the emissions inputs used by the two models. Specifically, CMAQ used an upper estimate of the *Carbón* power plant's SO_2 emissions, of 241,000 tons/yr compared to the 152,000 tons/yr used by REMSAD. Also, the other Mexican SO_2 emissions in the CMAQ domain were doubled compared to what was used in REMSAD.

As presented in Sections 9.9 and 9.11, the simulation of sulfate at Big Bend National Park and throughout Texas by the REMSAD and CMAQ-MADRID models is consistent with the state of the art of regional modeling. However, spatial and temporal trends in the errors resulted in larger errors at some times and places. For example, as discussed in Section 9.9, REMSAD nearly uniformly underestimated the sulfate concentrations throughout Texas during July and overestimated them in October. The positive bias in eastern Texas was at least partly due to overestimation of contributions from eastern U.S. sources.

In order to account for these biases, the modeled sulfate source attribution results, including the boundary conditions, from both models were scaled with observed sulfate concentrations throughout Texas to derive alternative source attribution estimates that better fit the measured data. This technique is based on inverse modeling of a linearized version of the continuity equation (Enting, 2002). The theory for this technique and regression method is summarized in Section 8.4.4 and is described in detail in the CIRA/NPS report on the BRAVO Study (Schichtel et al., 2004). As will be discussed in Section 11.1, the attribution of sulfate by REMSAD was not significantly influenced by nonlinearities. Therefore, the use of a linear model was appropriate.

Daily scaling coefficients were derived using data from most of the BRAVO monitoring sites and three consecutive days of data. A moving three-day window was used to derive scaling coefficients for the center days, resulting in the generation of daily coefficients from 9 July through 28 October 1999.

The synthesized REMSAD and CMAQ simulated Big Bend sulfate compared better to the observed values than the original model runs, with correlation coefficients above 0.85 and RMS errors about 40%, as shown in Table 9-23. The synthesized results still

underestimated the observed concentration by more than 15%. This bias is at least partially due to the regression analysis, which forced the intercept through zero and did not account for errors in the source attribution results.

Table 9-23. Comparison of sulfate simulation performance of the synthesized REMSAD and CMAQ models at Big Bend over the period 9 July through 28 October 1999.

	Synthesized REMSAD	Synthesized CMAQ
r^2	0.85	0.89
Bias (%)	-18	-16
RMS Error (%)	42	36

When the daily sulfate concentrations resulting from the synthesis approach are compared with measured values, the simulated concentrations are not exactly correct, of course. Of the predictions by the two models, the synthesized CMAQ results are likely more trustworthy. The original CMAQ results compared better to the observed sulfate data and the synthesized CMAQ results were closer to the original model estimates than those of REMSAD. Therefore, the CMAQ source attributions required smaller bias corrections, and relied less on the regression analysis to fit the observed data.

As will be shown in Section 11.3, both synthesized models were quite consistent, on average, in their attributions of sulfate to different source regions throughout the United States and Mexico. This good agreement between the two synthesized model results is an improvement over the original model results. Therefore, the synthesized inversion appears to account for different biases in the two models. It is not known, however, whether these biases are due to underestimations in emissions or systematic errors in the modeled transport, chemistry or removal processes.

Comparing the two synthesized model attribution results on a day-by-day basis revealed some differences. Texas, Eastern U.S. and Western U.S. attribution results were quite similar, with correlation coefficients of 0.88 and above. However, the source attribution results for Mexico and the boundary conditions had correlation coefficients less than 0.7. Mexico is the source region for which the CMAQ simulations used higher emission rates than the REMSAD ones, an increase that was supported by more recent Mexican emission inventories.

Although there are differences between the two models' source attributions to Mexico and the boundary conditions, but they are not large enough to result in different attribution conclusions. Therefore, in Chapter 11 only the daily synthesized CMAQ results are discussed in detail. The CMAQ results were chosen for several reasons:

- They conform better to the more current northern Mexico emission estimates,
- The original CMAQ results compared better to the observed sulfate data,

- The synthesized CMAQ results are closer to the original CMAQ model estimates than those of synthesized REMSAD were to regular REMSAD results. Therefore, the CMAQ source attributions required smaller bias corrections, and relied less on the regression analysis to fit the observed data.
- The synthesized CMAQ results compare better to the observed data.

9.13 Conclusions Concerning Performance of the Source Attribution Methods

The many analyses of model performance discussed here produced the not-surprising conclusion that the meteorological and air quality models ultimately used for source apportionment in the BRAVO Study performed pretty much within the norms of state-of-the art model performance. Most of the methods were able to estimate the average – over the study area and over the period of the study – sulfate concentrations and, to a lesser degree, the SO_2 concentrations. The models that provided finer spatial and temporal resolution were less successful at capturing all of the day-to-day details of concentration variation, especially for SO_2 and for peaks in either sulfate or SO_2. The REMSAD and CMAQ air pollution models both overestimated sulfur concentrations in the eastern part of the study area. CMAQ used measurement-scaled REMSAD outputs to provide boundary conditions, which adjusted for this overestimate in the REMSAD inputs to CMAQ, so the CMAQ overestimation should be relatively independent of the REMSAD bias.

Quantitatively, the MM5 wind fields used for the modeling and trajectory analyses tended to estimate hourly-average wind direction within 30 degrees (mean absolute error) and speed within 2 m/s (RMS error). Long-term biases were better than 10 degrees and 0.5 m/s, except that wind speeds aloft were slightly worse. The better long-term statistics provide a first indication of why the trajectory methods and air quality models performed better on average than on any particular day. The EDAS wind fields used for some of the trajectory analyses appeared to be comparable in quality to those of MM5.

The MM5 meteorological model outputs, when processed by the REMSAD model, tended to overestimate precipitation in the BRAVO Study area and cloud cover along the Texas-Mexico border.

Accurately simulating the narrow perfluorocarbon tracer plumes was expected to be challenging, and it turned out to be so. Performance for the tracer that traveled the longest distance, from northeast Texas, could not be evaluated reliably by any of the methods because the measured sample concentrations were so small. For the other tracers, none of the methods could consistently reproduce the day-to-day variation of the tracer signals, but most methods had some success some of the time. Both the REMSAD and CMAQ models performed the best for the Eagle Pass and San Antonio tracers, and not as well for the tracer released from the more distant Houston. Both returned the highest coefficients of determination, r^2, for the San Antonio tracer, with the best $r^2 = 0.4$ obtained by CMAQ.

Several of the methods were able to attribute the tracers to their sources with varying degrees of accuracy. For the Tracer Mass Balance (TrMB) method, the combination of the HYSPLIT trajectory model and EDAS winds performed relatively well at attributing the

tracers to their sources, but the same HYSPLIT model did very badly when it was applied with MM5 winds. The role of the winds was reversed when TrMB was exercised with the CAPITA Monte Carlo model. Then the performance with MM5 was pretty good but that with EDAS was poor. This means that some evaluation of performance is required in order to use the TrMB method.

A similar sensitivity to choice of wind field was reflected when the CAPITA Monte Carlo (CMC) model was used alone. It was able to approximate the apportionment of tracer from the nearby sources using the MM5 wind fields, but did better at apportioning the more distant tracer sources using EDAS winds. The CMC method was not able to reproduce the observed day-today tracer concentration variation at the receptor sites, however.

The Forward Mass Balance Regression (FMBR) method was able to identify the Eagle Pass tracer release location from the K-Bar tracer measurements. It could also estimate the release rate if data from all tracer sampling sites were used.

Turning to performance at estimating the concentrations of sulfur compounds in an artificial "reality" defined by application of the REMSAD model, TrMB with CMC trajectories and the MM5 wind fields was able to attribute study-average sulfate to various source regions with an accuracy of about 10 percentage points. The FMBR method performed relatively well at this task, except for a 10-percentage-point overestimation of the simulated Texas contribution and a comparable underestimation of the contribution from areas west of the BRAVO Study area.

Actual sulfate concentrations were predicted with relatively small average bias by both CMAQ-MADRID and REMSAD. Both models tracked many of the sulfate peaks, but neither model was able to predict all of the major sulfate peaks. Tracking of SO_2 peaks was poorer for both models, presumably because the SO_2 plumes were less dispersed. The coefficient of determination, r^2, at Big Bend National Park was less than 0.27 for REMSAD predictions of sulfate, but it was 0.52 for CMAQ. When all of the monitoring sites in the study area were considered, r^2 for sulfate was 0.38 for REMSAD and 0.47 for CMAQ. Mean absolute gross errors for both models were in the vicinity of 65%. Both models greatly overestimated sulfur (especially SO_2) in the eastern part of the study area.

The synthesis inversion process, which scaled the CMAQ and REMSAD predictions using measurements, improved the performance of both models, to the extent that r^2 for sulfate at Big Bend was greater than 0.7 (which means that these hybrid models explained more than 70% of the variation in the concentrations) and the bias was about –18%.

Because the different methods have differing uncertainties, the final attribution assessment of the BRAVO Study involved taking into account the performance demonstrated here as the attribution findings of the various methods were merged into a reconciled assessment that reflected the weight of evidence produced by the varying methods. This final attribution assessment reconciliation is presented in Chapter 12, and is based on the individual assessments that are described in Chapters 10 and 11.