8. ATTRIBUTION ANALYSIS AND MODELING METHODS

The BRAVO Study used a combination of data analysis methods and regional air quality models to provide information concerning the emission sources and source areas that are responsible for haze impacts at Big Bend National Park. The decision to use multiple complementary source attribution methods was made to develop a more comprehensive understanding of the causes of haze at Big Bend than could be done using any one method. Multiple methods make use of a larger fraction of the data collected by the study and afford the opportunity to reconcile results from the various approaches.

This chapter describes the data analysis approaches, the meteorological fields required as input to trajectory and air quality modeling, and the trajectory and air quality models that were used for source attribution by the BRAVO Study. Subsequent chapters evaluate the performance of the attribution methods (Chapter 9), display and discuss results (Chapters 10 and 11), and reconcile those results (Chapter 12). More detailed descriptions of the source attribution models and their results are contained in the appendices.

8.1 Data Analysis Methods

Source attribution by data analysis methods, also known as receptor modeling, includes a variety of approaches with the common feature being that ambient monitoring data are the basic source of information utilized in inferring sources or source areas. This is distinguished from air quality simulation modeling, also known as source-oriented modeling that relies primarily on predicting the meteorological properties or air pollution concentrations using mathematical descriptions of influential physical and chemical processes. Some of the data analyses methods described below include the use of modeled trajectories as though they were monitoring data. Because atmospheric transport models determine the trajectories, these techniques are more correctly considered hybrid approaches. They were included in the data analyses section because of their dependence on ambient monitoring data to infer associations between haze forming particulate matter and transport pathways.

8.1.1 Tracer Methods – TAGIT

Tracer-Aerosol Gradient Interpretive Technique (TAGIT) is a receptor model that can attribute primary or secondary species associated with a source whose emissions are “tagged” by a conservative tracer. The approach is straightforward. For each sample period, the background concentration of the species of interest, such as particulate sulfur, is determined by averaging the concentrations of that species at nearby monitoring sites that do not have tracer concentrations that are significantly above background levels. These sites are presumed to be unaffected by the tracer-tagged source and thus represent the average background in the area. This background for each sample period is then subtracted from the concentration of the species of interest at the tracer-impacted receptor sites for the corresponding sample period. The difference is the concentration attributable to the tagged source. It should be noted that the mass concentration data of the species of interest is normalized to a common atmospheric density for all sites. This eliminates the component of

spatial gradients that are a function of air density but would be equivalent in terms of volumetric ratios (e.g., ppb).

Uncertainties in the method arise from variations in the "background" concentration and from the assumption that there is no impact from the tagged source if the tracer concentration were less than the level considered to be "significantly" above its background. Uncertainty associated with both of these assumptions can be quantified. Unless the background concentrations of the species varies systematically in space (as it may in some circumstances), the first type of uncertainty is random, rather than a systematic bias. The second uncertainty mentioned, assuming no impact from the tagged source when tracer is not statistically above background, would lead to an underestimation of attribution. This is because background sites may have some impact from the tagged source, but with the uncertainty in the tracer concentration we cannot be sure that the site is impacted and it is thus denoted a "background" site. Its species concentrations are then used when computing the average background to be subtracted from the concentration at the receptor sites.

TAGIT computes attribution on a sample period by sample period basis. Some periods will have a negative concentration attributed to the tagged source; in Project MOHAVE for example, these negative values were generally within the calculated uncertainty of zero (Kuhns et al., 1999). The results are most meaningful when averaged over a number of sampling periods, such as the average over the field study. The assumptions, uncertainties, and limitations of TAGIT are discussed in detail elsewhere (Kuhns et al., 1999).

In the BRAVO Study, TAGIT was used to attribute particulate sulfur, SO_2, and total sulfur (particulate sulfur + SO_2 sulfur) associated with the *Carbón* I/II power plants. TAGIT was applied to two different tracers, the Eagle Pass synthetic tracer and SO_2. Neither of these tracers is ideal for TAGIT use in source attribution of the *Carbón* power plants. As discussed in Chapter 3, access to the *Carbón* power plant stacks to release tracer was not given, so Eagle Pass, TX was selected for tracer release as the closest location and surrogate for the *Carbón* facilities. Perfluorocarbon tracer released from a tower in Eagle Pass, some 32km northeast of the *Carbón*, does not necessarily result in coincident atmospheric distributions of tracer and *Carbón* power plant emissions as is assumed for TAGIT. If the tracer and power plant emission plumes are substantially separated, then sites that are considered to be at background, based upon tracer concentrations, may actually be impacted by the source. This would have the effect of improperly elevating background for subtraction. In addition, a site with tracer significantly above background may not actually be impacted by the source during that time period. Both of these effects tend to underestimate the true attribution to the source. Use of SO_2 as the TAGIT tracer for *Carbón* power plant emissions can be a problem because SO_2 is neither uniquely emitted by the *Carbón* power plants nor is it conserved in the atmosphere, both of which are assumed in TAGIT. It is more appropriate to label the use of SO_2 in TAGIT as an attribution of impacts by local SO_2 sources. The *Carbón* I/II power plants are by far the largest local emitters of SO_2, so the TAGIT results should overestimate the *Carbón* impact to a relatively minor extent. The extents and limitations of these problems are discussed in Chapter 9.

The TAGIT attribution by Eagle Pass tracer was done for the five 6-hour sites only; these sites have by far the greatest number of tracer samples. All particulate sulfur, SO_2, and total sulfur mass concentrations were standardized to the average atmospheric density at the BIBE site of 1.042 Kg m^{-3} that was obtained from NOAA meteorological measurements at K-Bar Ranch in Big Bend National Park. Sites were determined to be at background for sample periods where the measured ocPDCH tracer concentrations are less than 2 times the concentration uncertainty above global background for the tracer, as reported by Brookhaven National Laboratory (BNL). The concentration needed to be significantly above background was usually about 0.13 femtoliters per liter (fL L^{-1}) or parts per quadrillion (ppq). This tracer concentration corresponds to about 330 ng m^{-3} of SO_2 sulfur from the *Carbón* powerplants if the assumed SO_2 emission rate of 240,000 tpy (6900 g s^{-1}) is correct.

Attribution was calculated for each 6-hour site and sample period with an ocPDCH concentration greater than twice the uncertainty above background, provided at least one of the other 4 sites used had ocPDCH concentrations not significantly above background. There were five 6-hour sampling periods when all 5 sites had tracer concentrations above background; no calculations were made for these periods.

The TAGIT analysis done using SO_2 instead of ocPDCH for the same five monitoring sites required a definition of background SO_2 value. As was noted above for the Eagle Pass tracer plume, the *Carbón* plume is also not likely to be impacting all five sites simultaneously very often. Therefore the site with the lowest SO_2 was defined as the background site. By this approach, typical background SO_2 concentrations during the BRAVO Study are about 0.3 μg/m^3, while at impacted sites SO_2 concentrations are typically greater than 2 μg/m^3. The particulate sulfur for this site was then set as the background particulate sulfur and attribution was determined by subtracting this background from the particulate sulfur at each of the other sites.

Additional information concerning the TAGIT method and its results are contained elsewhere -- in Green et al. (2003) and in Chapter 10 of this report.

8.1.2 Empirical Orthogonal Function Analysis

Spatial and temporal patterns in the BRAVO particulate data have been examined using a statistical method known as Empirical Orthogonal Function (EOF) analysis. This analysis method simplifies the dozens of daily spatial patterns into a few that, when linearly re-combined, explain most of the variance in the data.

EOF analysis requires a time by site matrix of concentrations of a single aerosol species. No missing values are allowed, so times and sites with many missing values were eliminated, leaving July 26-Oct. 30 with sufficient data. Remaining missing values were filled in using either spatial or temporal interpolation and concentrations below detectable limit were set to zero. EOF analysis was applied to the daily data for the major and a few trace aerosol species measured at the BRAVO Study sites.

Examination of the EOF patterns is often useful to qualitatively determine dominant source-receptor relationships and to form hypotheses about significant source areas and

physical factors likely to influence transport and transformation of aerosols in a region. Plotting the site loadings for each factor on separate maps facilitates the interpretation of EOF patterns. Values are contoured so that gradients from high to low values can be easily seen. Source areas associated with each factor are presumed to be in the vicinity of the high factor loading sites if they are within the monitoring area, or in the direction of increasing factor loadings if the source areas are beyond the monitoring area.

Results of the EOF analysis for particulate sulfate are presented and discussed in Chapter 10. More complete description of the EOF analysis method and factor loading contour maps for all of the aerosol components are in Barna et al. (2004) (in the Appendix).

8.1.3 Factor Analysis

The composition of fine particles ($PM_{2.5}$) sampled during the BRAVO Study was summarized and explored using a statistical method known as Factor Analysis (FA). FA was applied to the elemental analysis data for the 32 BRAVO Study sites in Texas (see Table 3-3 and site map in Figure 3-1), and at Big Bend monitoring site to the more complete composition data set including elements, ions and carbon composition data.

FA attempts to reduce the complexity of multivariate data sets by identifying a smaller number of underlying factors that explain the covariation in the data set. The Variables (e.g., elements and other component species) are assumed to be linear combinations of the underlying factors. Variables loading on a factor are highly correlated to each other and not well correlated to variables in the other factors. Often when applied to particle composition data sets the resulting factors can be identified with specific pollution emission source types or activities by the species that they include. For example the elemental components that make up suspended crustal material (e.g., Al, Si, Ca, Ti, Fe, etc.) usually are strongly associated with a single factor that is identified with crustal emissions. Not all factors are unambiguously associated with a single source type or activity.

Application of factor analysis to the Texas elemental data reduced the 22 elemental variables for each site to four factors, and reduced the 37 composition variables for the Big Bend monitoring site to six factors. For the Texas elemental data factors, spatial patterns of the resulting factors were examined to further explore plausible source associations with each. Sites with the highest factor loadings are generally those nearest the source(s) that contributes to that factor. Additional information about the method and its application are available in Chapter 10 and Mercado et al. (2004) (in the Appendix).

8.2 Meteorological Fields

Trajectory and air quality models used for attribution assessment require meteorological data as input. Three meteorological models provided the input information required by the trajectory and air pollution models used in the BRAVO Study. EDAS and FNL are routinely run by the National Weather Service's National Centers for Environmental Prediction (NCEP) for use in weather forecasting. MM5 was run specifically for the BRAVO study by the Pennsylvania State University's Department of Meteorology.

8.2.1 EDAS

NCEP maintains operational meteorological models for weather forecasting. One modeling system is the Eta Data Assimilation System (EDAS), which generates initial conditions for the Eta forecast model (Black 1994; Parrish et al., 1996). The EDAS meteorological fields are generated using a three-dimensional variational objective data assimilation analysis scheme. This is essentially a sophisticated data interpolation scheme that uses the Eta forecast model to optimally merge and spatially interpolate measured meteorological fields. In 1999, EDAS incorporated 34 different data types from 26 data sets including land and marine surface observations, upper air data from the rawinsonde and wind profiler networks, Aircraft Communications Addressing and Reporting System (ACARS) and meteorological fields derived from satellite data.

NCEP operates EDAS on a 32-km grid on a terrain-following, vertical-coordinate system with 45 levels, and generated meteorological field every three hours. These data are interpolated to a 40-km Lambert Conformal grid and isobaric levels. NOAA's Air Resource Laboratory (ARL) has direct access to the EDAS data stream and saves out a subset of the EDAS data suitable for input into dispersion models. These data are available from ARL's READY website (READY, 2003). The ARL archive contains the EDAS data interpolated to a 40-km Lambert Conformal grid with every other grid point saved out on 22 isobaric surfaces.

8.2.2 FNL

The Global Data Assimilation System (GDAS) (Kanamitsu, 1989; Derber et al., 1991; Parrish and Derber, 1992) is another of the operation systems that NCEP runs to generate inputs into the meteorological forecast models. GDAS uses a Spectral Statistical Interpolation (SSI) scheme coupled with the spectral Medium Range Forecast model (MRF) forecast model. The SSI scheme is closely related to the three dimensional variational analysis system used in the EDAS system (Parrish and Derber, 1992), and it incorporates similar data as EDAS. However, GDAS is the final run in the series of NCEP operational model runs, and includes late arriving data that can not be incorporated into EDAS.

NCEP runs GDAS four times a day at 00, 06, 12, and 18 UTC. Model output is for the SSI analysis time and a 6-hour forecast. NCEP's post-processing of the GDAS converts the data from spectral coefficient form to 1 degree latitude-longitude (360 by 181) grids and from the 42 sigma level vertical coordinate system to isobaric levels. NOAA's ARL takes these fields and converts them to polar stereographic grids with ~180 km resolution and saves out 13 of the pressure levels. Some fields such as precipitation and surface fluxes are only available at the forecast time so ARL merges the GDAS and forecast runs to create a complete archive. Since GDAS is the last operational model run, it is known as the Final Run at NCEP and ARL calls this archive FNL. The data used in the report were downloaded from ARL's READY website (READY, 2003).

8.2.3 MM5

The non-hydrostatic 3-D MM5 mesoscale model (Grell et al., 1994) was used for the numerical modeling of the meteorology during the BRAVO study period. The MM5 uses a terrain-following sigma vertical coordinate (non-dimensionalized pressure) and supports multiple grid nesting, a full array of physical parameterizations, and four-dimensional data assimilation. The latest public-release version of the modeling system available from NCAR at the time of the BRAVO numerical modeling study (MM5v3) was used as the basis for all model runs. With very few exceptions, all codes used in the BRAVO meteorological modeling were selected from standard options of the NCAR-supported publicly available software.

BRAVO's application of MM5 adopted a triply nested domain configuration. The nested domains had horizontal grid resolutions of 36-km, 12-km and 4-km, and are shown in Figure 8-1. Each domain was applied with 35 vertical layers. The middle of the first layer (first computation level) was at 18 m above ground level (AGL). Above the first level, the layer thickness was gradually increased with height, so that the greatest resolution was in the boundary layer. The model top was placed at 50 mb, instead of the usual 100-mb top often used in air-quality applications involving MM5. The high model top allowed more of the lower stratosphere to be included in the domains to ensure that overshooting updrafts in deep thunderstorms could not approach or reach the model's lid, which could cause numerical instability. Objective analyses for initial and lateral boundary conditions were generated by horizontally interpolating archived NCEP Eta-model fields onto the 36-km MM5 domains as the background. The finer grids of the MM5 were initialized by interpolating from the 36-km-grid fields. In addition three-hourly objective surface analyses were generated from the surface observations for use in the model's data assimilation system.

The MM5 runs were produced as a series of segments, with most segments being 5 1/2 days in length. Re-initialization of the model at the end of every 5 1/2 days ensured that there would not be an accumulation of numerical errors that could degrade the usefulness of the model fields. A 12-h overlap was provided between each segment, representing a "spin-up" period during which the model can come into dynamical balance from the somewhat unbalanced initial states. The spin-up periods at the beginning of each segment was discarded prior to input into the air quality models. The 36- and 12-km domains were run simultaneously using two-way interactive nested grids for the entire four-month BRAVO period. Because of the heavy computational burden imposed by very high grid resolution, the 4-km domain was run for two limited intensive-study periods of about 10 days each that were selected by the BRAVO committee. Complete MM5 output fields were written and archived for all domains at 1-h intervals for the appropriated study periods.

As discussed in subsequent sections of this report, ultimately the CAPITA Monte Carlo, REMSAD, and CMAQ models employed only the 36-km MM5 output.

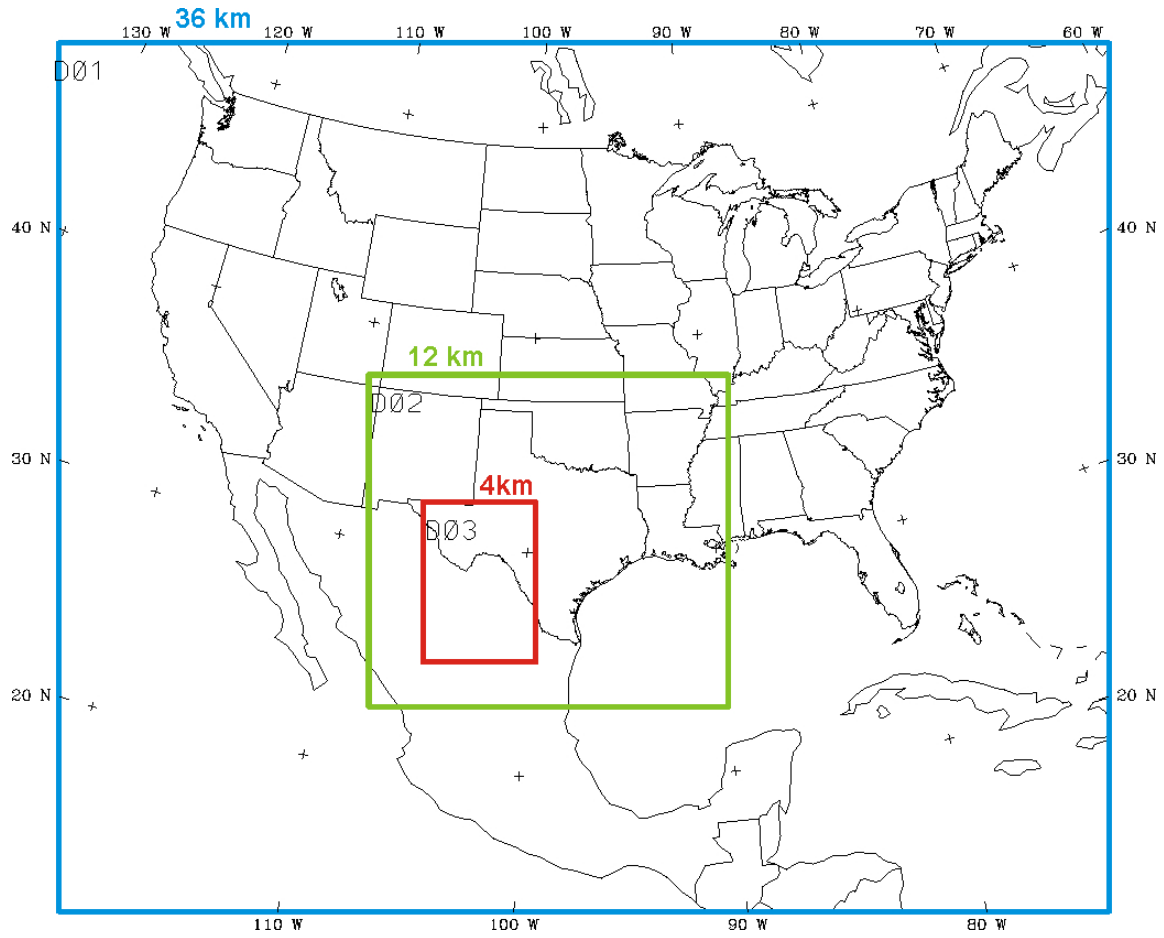

Figure 8-1. *Locations of the triply-nested MM5 domains, with grid resolutions of 36, 12, and 4 km.*

8.3 Trajectory-Based Analyses

Three trajectory models used for the BRAVO Study are described below. This is followed by sections describing numerous methods to employ trajectory information along with air quality data to infer source attribution. More information on these methods is described in the CIRA/NPS report on the BRAVO Study (Schichtel et al., 2004), which is in the Appendix.

8.3.1 ATAD Trajectories

The Air Resource Laboratory Atmospheric Transport and Dispersion (ATAD) model (Heffter, 1980) is a Lagrangian parcel model with a single variable depth transport layer. The base of the transport layer is generally 300 meters above the ground. For most time periods, the top of the transport layer is the lowest level within a critical inversion at which the potential temperature is 2°C above that at the inversion base. A critical inversion is defined as an inversion with a potential temperature lapse rate of at least 5°C km^{-1}. When no critical inversion exists, the transport layer top is assumed to be 3,000 meters above the

ground. For trajectories that begin at night, the initial transport layer depth is approximated by two σ_z, where σ_z is the standard deviation of the vertical dispersion of a Gaussian plume for stable conditions. This is used only until the first daytime period of the trajectory.

Average winds within the transport layer are interpolated spatially (inverse of squared distance weighting) and temporally from all available radiosonde data within 250 km. If there are no stations within 250 km, data from stations within 600 km are used. The trajectory is terminated if there are no upper air data within 600 km. Complex terrain is not explicitly considered in the model, although the transport layer is always at least 300 meters above the terrain near each radiosonde station. A back trajectory is started from the receptor every six hours. An air parcel position, or "endpoint", is determined for every three hours backward in time for a maximum of 120 hours (five days).

8.3.2 HYSPLIT Trajectories

The HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess, 1998) was developed by the NOAA Environmental Research Laboratories. It can compute a range of outputs from simple air parcel trajectories (advection of a single particle) to dispersion and deposition simulations. For BRAVO, version 4.5 of the model was used in simple back trajectory mode.

In its trajectory mode, HYSPLIT can do computations forward or backward in time. Default vertical motion, which was employed for BRAVO, is calculated using the input vertical velocity field. Required input is a gridded meteorological data set on a polar, Lambert or Mercator map projection with data at regular intervals. Back trajectory positions or "endpoints" are calculated hourly. For BRAVO, trajectories were calculated for up to 10 days backward in time.

The advection of a particle or puff is computed from the average of the three-dimensional velocity vectors at the initial position and at the first guess of the next position. Velocity vectors are linearly interpolated in both space and time. Trajectories terminate if they exit the model top (specified as 10 Km AGL for BRAVO), but advection continues along the surface if they intersect the ground. The integration time step can vary during the simulation and is computed such that it is less than 0.75 of the meteorological grid spacing. A simple integration method is employed. Higher order integration schemes were found to add no precision because data observations are linearly interpolated from the grid to the integration point (Draxler, 1998).

8.3.3 CAPITA Monte Carlo Trajectories

The CAPITA Monte Carlo model described in Section 8.4.1 determines transport and dispersion by releasing numerous virtual particles every time step into any of the meteorological field described above, and by tracking their horizontal and vertical locations estimate the transport and dispersion. The path of each of the virtual particles can be tracked either forward or backward in time to generate forward and backward trajectories. Along with ATAD and HYSPLIT, Monte Carlo back-trajectories were used in the trajectory source attribution and air mass history methods described below.

8.3.4 Ensemble Air-Mass History Analyses

An air-mass history identifies the 2-D or 3-D pathway an air-mass took enroute to the receptor, a trajectory, and other meteorological parameters along this transport pathway. The implication is that sources, or lack of sources, along the pathway and near the receptor are responsible for the receptor's air quality. These are powerful techniques that potentially link source regions to the receptor concentrations. However, individual air-mass histories can have large errors (Kahl and Samson, 1986; Rolph and Draxler, 1990; Stohl, 1998). In fact, different model assumptions or wind fields can generate air-mass trajectories that are 180 degrees out of phase. So long as the trajectory method is not subject to systematic biases, the large uncertainty that can manifest itself in individual trajectories can be reduced through the use of ensemble air-mass history techniques, which aggregate large numbers of air-mass histories together. Such methods have proved useful in identifying source regions in previous investigations (Gebhart et al., 2001; White et al., 1994; Malm, 1992; Gebhart and Malm, 1991, 1994; Bresch et al., 1986; Ashbaugh et al., 1985).

There are a variety of methods to summarize and use ensemble air mass histories. The methods used for the BRAVO Study are summarized in Table 8-1. A more technically rigorous description of each is included in Schichtel et al. (2004) (in the Appendix).

8.3.5 Trajectory Mass Balance (TrMB)

The trajectory mass balance regression is a receptor modeling technique in which the measured concentration of a species of interest, in this case 24-hour average fine particulate sulfur measured at Big Bend (K-Bar), is the dependent variable and numbers of back-trajectory endpoints in each source region (i.e., the residence time as defined in Table 8-1) are the independent variables. Ordinary Least Squares (OLS) regression was used to solve for the "transfer coefficients" for each source. The transfer coefficients, with units of concentration per endpoint (i.e., the same as concentration per hour residence time within a source region), are estimates of the average relationship between air mass residence time in the source area and measured concentration at the receptor. This relationship depends on average emissions, dispersion, deposition, and chemical transformation.

Table 8-1. Summary of ensemble air mass history analysis methods.

Term (abbreviation)	Definition	Comments
Residence Time (RT)	Fraction of transport time air parcels spends in a predefined geographic area (typically a map grid cell) on their way to the receptor site.	Maps of RT shows transport pathways to the receptor site.
Overall RT (ORT)	Residence time for transport during an extended period of time (e.g., a month, or a year).	Maps of ORT shows transport pathways to the receptor site for the specified period of time.
High condition RT (HRT)	Residence time for transport during high concentration or haze level conditions. Typically the top 10% to 20% of conditions.	Maps of HRT shows transport pathways to the receptor site during highly impacted periods.
Low condition RT (LRT)	Residence time for transport during low concentration or haze level conditions. Typically the lowest 10% to 20% of conditions.	Maps of LRT shows transport pathways to the receptor site during un-impacted periods.

Table 8-1. (continued)

Term (abbreviation)	Definition	Comments
Probability Density Function (PDF)	Residence time divided by the area of the geographic cells, done to normalize for cells of uneven size.	Permits assessment of transport pathways for arbitrary source or political regions of unequal size.
Source Contribution (SC)	Residence time normalized by distance. Calculated by multiplying RT by the distance between the receptor site and grid cell.	This removes the bull's eye affect of RT contour maps caused by the necessarily high density of trajectories near the receptor location.
Overall SC (OSC)	ORT times the distance between the receptor site and grid cell.	Maps of OSC shows transport pathways to the receptor site for the specified period of time without the bull's eye pattern.
High condition SC (HSC)	HRT times the distance between the receptor site and grid cell.	Maps of HRT shows transport pathways to the receptor site during highly impacted periods without the bull's eye pattern.
Low condition SC (LSC)	LRT times the distance between the receptor site and grid cell.	Maps of LRT shows transport pathways to the receptor site during un-impacted periods without the bull's eye pattern.
High Conditional Probability (HCP)	The amount of time that an air parcel spends in a grid cell that arrives at the receptor site during high concentrations or haze levels divided by the amount of time the air parcel spends in the same grid cell for all conditions at the receptor site for the same overall period of time.	Maps of HCP show the potential for areas to contribute to impacted conditions at the receptor location when the area is in the transport path.
Low Conditional Probability (LCP)	The amount of time that an air parcel spends in a grid cell that arrives at the receptor site during low concentrations or haze levels divided by the amount of time the air parcel spends in the same grid cell for all conditions at the receptor site for the same overall period of time.	Maps of HCP show potential for areas to contribute to un-impacted conditions at the receptor location when the area is in the transport path.
High Incremental Probability (HIP)	High residence time (HRT) minus overall residence time (ORT) for corresponding periods of time.	Maps of HIP show the most probable areas for transport when receptor site conditions are impacted.
Low Incremental Probability (LIP)	Low residence time (LRT) minus overall residence time (ORT) for corresponding periods of time.	Maps of LIP show the most probable areas for transport when conditions at the receptor site are un-impacted.
Accumulation Potential (AP)	Accumulation potential is one of two components into which the PDF can be subdivided. It is a measure of the time an air parcel spends in each grid cell.	Maps of AP show areas where air parcel speed tends to be slow thus allowing more time to accumulate pollutants from emission sources that can be transported to the receptor site.
Normalized Directional Frequency (D)	Normalized directional frequency is one of two components into which the PDF can be subdivided. It is a measure of how often trajectories traverse each grid cell.	Maps of D show the frequency that transport traverses an area.
Trajectory Maximum Source Contribution (SC)	Trajectory maximum is an extreme limiting case source attribution method that credits each grid cell traversed by a trajectory with responsibility for all of the pollution measured at the receptor site when the air parcel arrives.	While this is certainly an unrealistic attribution value for areas, it does set an upper limit, which some attribution methods might exceed. Its sole use is to check for such gross inconsistencies.

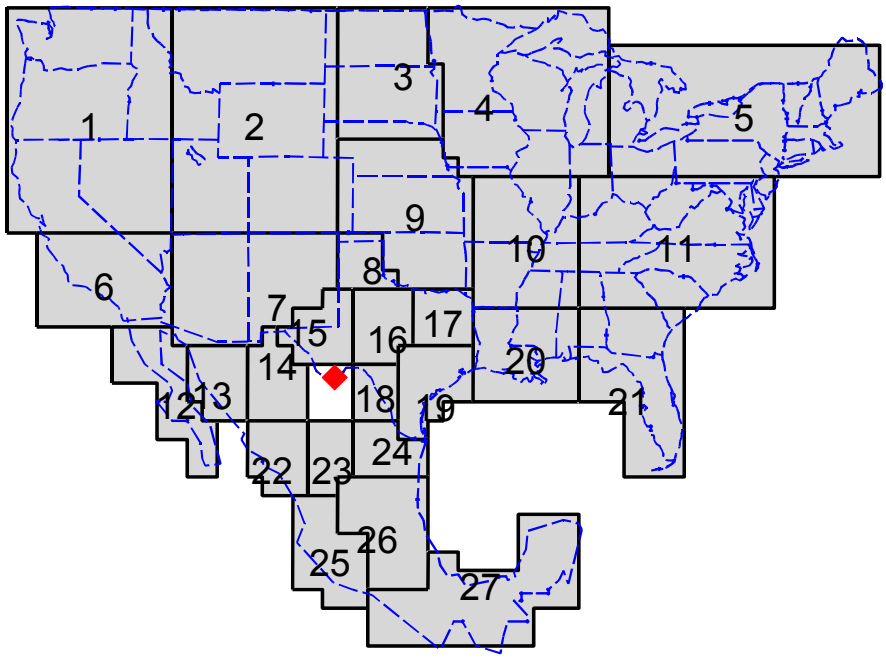

Figure 8-2. *Source areas with map identification numbers used for Trajectory Mass Balance Modeling.*

Back trajectories were started from Big Bend National Park every hour with an endpoint or air mass location calculated hourly for five, seven, or ten days. Two different trajectory models, HYSPLIT and the CAPITA Monte Carlo (CMC) model were used. Each was run with two input wind fields, MM5 on 36-Km grids and the standard wind fields used by the National Weather Service for weather prediction, EDAS when available and the lower resolution FNL when EDAS was missing. HYSPLIT was started at heights of 100, 200, 500, and 1000 m above ground. Source areas used to classify the trajectories are shown in Figure 8-2 and identified in Table 8-2.

Source areas were chosen based on several criteria. First, areas were chosen partially based on interest in the attributions from the area (i.e., separating the influence of sources in Mexico from sources in Texas). Second, source areas near the receptor can be smaller than sources farther away due to the inherent error in trajectory endpoint locations as the time from the receptor increases. Third, model performance is better (i.e., R^2 is larger, fewer statistically insignificant regression coefficients, superior simulations of synthetic tracer) if the source areas have significant emissions of the pollutant of interest and if all or most trajectories passing through the source region would be expected to have similar exposure to emissions, dispersion, and transformation enroute to the receptor; and finally, to avoid collinearities between source regions, the timing and number of trajectories passing through each region should be reasonably independent from other regions. It is often difficult to choose areas that simultaneously satisfy all criteria.

Table 8-2. Names and map identification numbers for individual source areas composite source regions.

Map Number	Individual Sources	Composite Source Regions	REMSAD Source Regions
1	Pacific Northwest	Western U.S.	Western U.S.
2	Northern Rockies	Western U.S.	Western U.S.
3	Dakotas	Western U.S.	Western U.S.
4	North Central States	Eastern U.S.	North Central States
5	Northeast	Eastern U.S.	Other East
6	Southern California	Western U.S.	Western U.S.
7	AZ / NM	Western U.S.	Western U.S.
8	Texas Panhandle	Texas	Other Texas
9	Central Plains	Western U.S.	Western U.S.
10	MO/IL/AR	Eastern U.S.	MO/IL/AR
11	East Central States	Eastern U.S.	East Central States
12	Baja California	Mexico	Mexico
13	Northwest Mexico	Mexico	Mexico
14	North Central Mexico	Mexico	Mexico
15	West Texas	Texas	Other Texas
16	North Central Texas	Texas	Other Texas
17	Northeast Texas	Texas	Northeast Texas
18	*Carbón* I/II	Mexico	*Carbón* I/II
19	Southeast Texas	Texas	Southeast Texas
20	LA/MS	Eastern U.S.	LA/MS
21	FL/GA	Eastern U.S.	Other East
22	West Central Mexico	Mexico	Mexico
23	Central Mexico	Mexico	Mexico
24	Monterrey Region, MX	Mexico	Mexico
25	SW Coast of Mexico	Mexico	Mexico
26	Mexico City and Volcano	Mexico	Mexico
27	S. Mexico / Yucatan	Mexico	Mexico

8.3.6 Forward Mass Balance Regression (FMBR)

FMBR is identical in concept to the TrMB technique. However instead of using backtrajectories to assess the transport between the source regions and receptors, forward plumes that include information about both transport and dispersion were calculated for various source regions. The CAPITA Monte Carlo model (described in Section 8.4.1), run using EDAS/FNL meteorological fields, was used to generate plumes for a uniformly spaced grid of virtual sources every 100 km. This was done in the model by releasing 100 points every two hours from each location and tracking them for five days. As shown in Figure 8-3, the 100-km grid virtual sources were organized into source regions similar to those used in the TrMB (see Section 8.3.5).

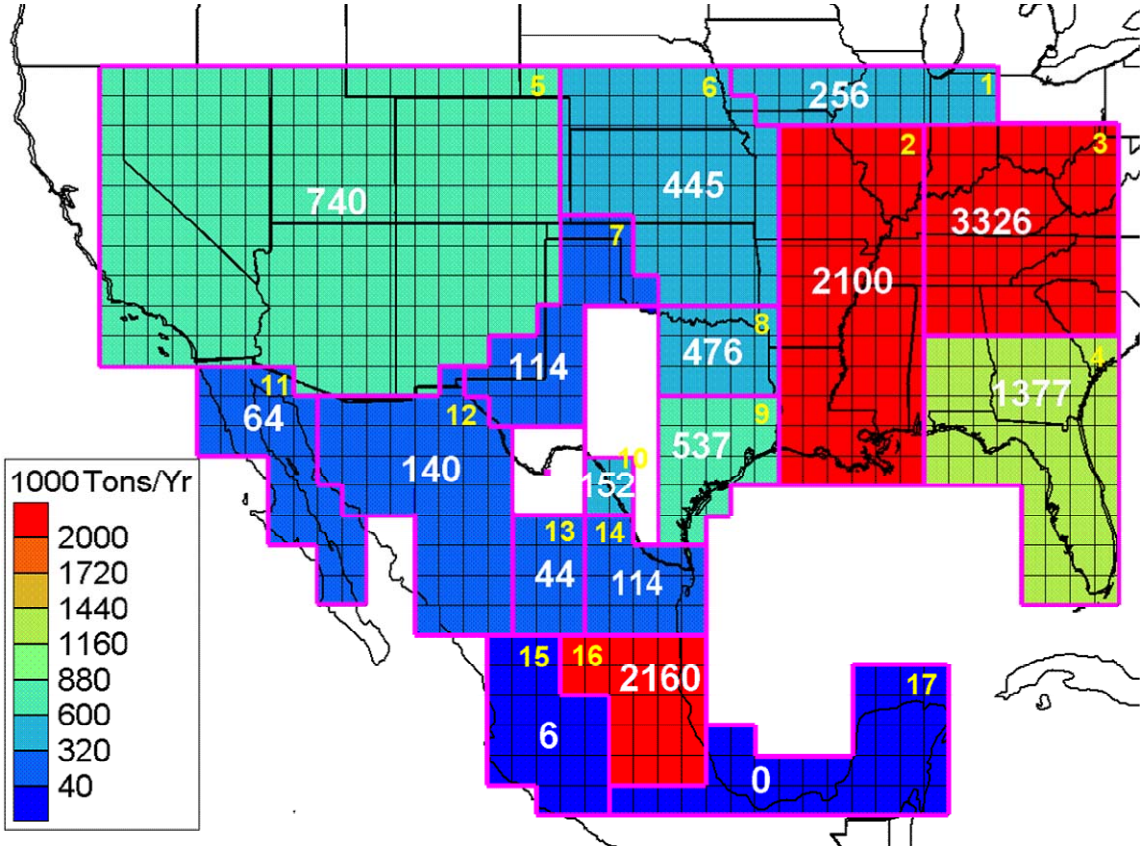

Figure 8-3. *Locations of 17 source areas composed of 100-km grid virtual sources used to track transport and dispersion in the Forward Mass Balance Regression method. The values and colors of each source region are the SO_2 emission rates in 1,000 tons per year from the BRAVO and National Emission Trend inventories. Open regions surrounding Big Bend and in central Texas were not included in the assessment and have near-zero SO_2 emissions. Composite source regions consistent with those of the other attribution analyses are composed of the numbered source areas as follows: Eastern US - 1-4; Western US - 5 and 6; Texas - 7-9; Carbón - 10; and Mexico - 10-17.*

The numbers of particles from the various source regions that arrive at the Big Bend receptor site during any period of time are related to the transit probabilities for each region. A linear least squares technique is used to identify a set of regression coefficients which when multiplied by their respective transit probabilities for each source region for each of the sample periods produces a "best" estimate of the measured particulate sulfate concentrations at Big Bend. SO_2 was also attributed to these source regions in this manner.

The regression coefficients determined by this method account for the combined effects of emission rates, plus deposition and chemical transformation during transport for each source region over study period. These factors undoubtedly change over time, as do transport and dispersion. However, this approach, just like TrMB, uses a single regression coefficient value for each source region for the entire BRAVO Study period, so it cannot be expected to attribute accurately for periods of highly unusual emissions, deposition or chemical transformation.

8.4 Air Quality Simulation Modeling

8.4.1 CAPITA Monte Carlo

The Monte Carlo model is a particle dispersion model capable of simulating regional scale transport, transformation, and dry and wet removal of aerosols (Schichtel and Husar, 1996, 1997). In this study, only the atmospheric transport module was used.

The simulation of the regional scale transport and diffusion in the Monte Carlo models is conducted by moving inertialess particles in the Eulerian frame. The advection of the particles is accomplished by multiplying the mean three dimensional wind vectors at each particle's location in space and time by the time step Δt. The mean wind vector is obtained from meteorological models, such as MM5 (described above). These gridded wind fields are interpolated to the particle's position by using bilinear interpolation in the horizontal and linear interpolation in the vertical and time. The model uses a Δt on the order of 20 minutes. Perfect reflection is assumed at the surface but if a particle reaches the grid edge or top of the model domain, it is removed from the analysis.

The random components of the velocity field occur at a resolution smaller than the meteorological model grids used to generate wind fields. Therefore, these components are derived using simplified models of the atmospheric turbulence driven by the available meteorological variables. The vertical and horizontal fluctuating components of the wind fields are modeled separately since the vertical diffusion primarily depends on the intense convective and mechanical mixing process in the atmosphere and horizontal dispersion primarily depends on the divergence of the horizontal wind fields.

In the case of vertical diffusion, during the day, for the intense convective mixing within the mixed layer the turn over time of a large eddy is about 15 minutes. In the Monte Carlo model this mixing is simulated by evenly distributing the particles between the surface and the mixing height. The vertical mixing is only applied to particles below the mixing height. Particles enter or leave the mixing layer as the mixing height grows or contracts and due to vertical motion from subsiding and rising air-masses. This simple vertical diffusion mechanism has been incorporated into several dispersion models (Patterson et al., 1981; Anfossi et al., 1995; Saltbones et al., 1998).

Horizontal diffusion is based upon the Prandtl mixing length model. This allows for the parameterization of the diffusion by an effective eddy diffusion coefficient K. The diffusion process is implemented as a random walk displacement of radius $\sqrt{2K\Delta t}$ for each time step in the model. The position from 0 to $\sqrt{2K\Delta t}$ is chosen from a Gaussian distribution with a mean of 0 and unit standard deviation. The value of the diffusion coefficient is a geographical function of the time of day and season. It is recognized that using an eddy diffusion coefficient to represent horizontal diffusion is a crude approximation. However, it has been shown that horizontal diffusion has a small effect on multi-day regional scale dispersion (Uliaz, 1994; Schichtel and Husar, 1996).

8.4.2 REMSAD

REMSAD, the Regional Modeling System for Aerosols and Deposition, is a prognostic, Eulerian-grid air quality model designed to simulate the formation and long-range transport of aerosols and their precursors (Seigneur et al., 1999; SAI, 2002). REMSAD has been optimized to be computationally efficient, allowing the simulation of long time periods (e.g., monthly or yearly) over large model domains (e.g., continental-scale). This is achieved in part through the highly simplified treatment of organic species in the chemistry mechanism. REMSAD is based on the numerical solution of the atmospheric diffusion equation (e.g., Seinfeld and Pandis, 1998), which expresses the physical and chemical processes that affect atmospheric pollutants and their precursors, including advection, diffusion, wet and dry deposition, and chemical transformation.

The REMSAD model domain for the BRAVO Study covers most of the contiguous United States and northern Mexico, and is approximately centered on Texas. A geodetic (latitude/longitude) horizontal coordinate system is used, with a model grid resolution of approximately 36 km. The domain extends to 74^{o} W and 120^{o} W at the eastern and western boundaries, respectively, and to 49^{o} N and 16^{o} N for the northern and southern boundaries, respectively. The vertical dimension is defined in terrain-following sigma-pressure coordinates. Thirteen vertical layers are used, with thinner layers specified near the surface and thicker layers aloft. The top of the model domain is set to 50 mb.

The REMSAD chemistry mechanism treats gas-phase, aqueous phase, and aerosol equilibrium processes. Gas phase chemistry is calculated with the Micro Carbon Bond IV mechanism (μCB-IV), which is based on a reduced formulation of the widely used Carbon Bond IV mechanism (SAI, 2002). The μCB-IV contains a simplified treatment of organic compounds, with three lumped-species variables representing anthropogenic organic compounds, carbonyls and biogenic organic compounds. The inorganic and radical chemistry portions of μCB-IV are identical to CB-IV. The representation of aqueous chemistry, which is particularly important to SO_4 formation, treats the reactions of dissolved SO_2 with H_2O_2, O_3 and atomic oxygen catalyzed by iron and manganese. The MARS-A thermodynamics module is used to calculate the equilibrium between nitrate, SO_4 and ammonia (Saxena et al., 1986).

SO_2 and SO_4 boundary concentrations were created using results from the GOCART global climate model (Chin et al., 2000). GOCART data were not available for 1999; instead, GOCART results for 2000 were used to construct a four-month average boundary concentration field for REMSAD, with the presumption that the GOCART predictions between 1999 and 2000 were similar on a seasonal basis. REMSAD meteorological fields were from the 36-km grid modeling domain output of the MM5 model as described above.

REMSAD simulation for the base case used emissions and boundary conditions as described above to predict pollutant concentrations for every location within the modeling domain for every hour of the four-month study period. Comparisons between base-case predicted and measured sulfate and SO_2 at Big Bend and the other study monitoring sites provided a means to evaluate the performance of REMSAD (see Chapter 9). Since

REMSAD does not tag the emission source identifications, the base case simulation provided no source attribution information.

Source attribution was accomplished with REMSAD by re-running the model with a modified emissions inventory for each source or source region that is to be assessed. This was done two ways for sulfate/SO_2 attribution for each of nine composite source regions shown in Table 8-2 plus for the domain boundary conditions. For the first approach, known as "emission-in" only SO_2 emissions in the specific source region of interest are used, and elsewhere they are set to zero. For the second approach, known as "emissions-out", SO_2 emissions outside of the specific source region of interest are maintained and those within are set to zero. The model results for the emissions-in case for a specific source region predict its impact at any location within the modeling domain. The predicted concentration fields for the emission-out case needs to be subtracted from the corresponding base-case concentration fields in order to produce predictions of the source area of interest. The difference in results between the emissions-in and emissions-out attribution approach is interpreted as a measure of non-linearity of atmospheric properties. Source attribution values for each source region and the domain boundary conditions were determined in this way for every sample period for all monitoring sites throughout the study area.

8.4.3 CMAQ

The Community Multiscale Air Quality (CMAQ) model was used to perform simulations representative of ambient conditions during the BRAVO study. CMAQ (Byun and Ching, 1999) is a modular air quality model developed for simulating the fate and transport of atmospheric gases and particulate matter. The CMAQ model was employed to simulate the tracer release experiments using the following configuration options: Bott scheme for horizontal and vertical advection, the eddy diffusion scheme for vertical diffusion, and the Smagorinsky scheme for horizontal diffusion. (The Smagorinsky scheme was found to perform better at simulation of the tracer plumes than the diffusion algorithms that are regularly contained in CMAQ and therefore it was used for all BRAVO attributions simulations.)

For chemical transport simulations, the CMAQ-MADRID (Model for Aerosol Dynamics, Reaction, Ionization, and Dissolution) model was applied to the BRAVO study. The MADRID atmospheric aerosol modules were implemented within the 2000 version of CMAQ and replaced the aerosol code in the EPA-distributed version of CMAQ. Gas-phase chemistry was simulated using the regional acid deposition mechanism, version 2 (RADM 2) with the 4-product isoprene chemistry and aqueous processes were simulated using the RADM cloud module.

The MADRID modules were formulated after a review of currently available data and algorithms for simulating processes that govern the chemical composition and size distribution of ambient particulate matter (EPRI, 2002). For example, whereas the original CMAQ aerosol code tracks a modal representation of the particle size distribution, the MADRID modules track a discrete sectional representation of the particle distribution. Several configuration options are available within MADRID to represent the particle size distribution, the physics of aerosol particles and the thermodynamics of their organic and

inorganic aerosol constituents. The set of modules selected for the BRAVO application offers a reasonable compromise between numerical accuracy and computational efficiency. Two size sections represent fine ($PM_{2.5}$) and coarse particles. Each section is composed of ammonium, chloride, nitrate, sodium, sulfate, other (unclassified) inorganic species, water, elemental (black) carbon, primary organic compounds, and 38 surrogates of anthropogenic and biogenic secondary organic aerosol (SOA) compounds.

The thermodynamics of inorganic aerosol species is simulated by the ISORROPIA module, modified in order to improve numerical stability via on-line calculation of activity coefficients. The RADM2 mechanism was extended when used with the MADRID modules to include detailed reactions of volatile organic compounds with atmospheric oxidants leading to the formation of low volatility products. In accordance with their vapor pressure, these semi-volatile organic compounds (SVOC) undergo reversible absorption and evaporation between the gas phase and the particle phase resulting in secondary organic aerosol. The MADRID–1 option selected for this study uses an empirical approach, based on the results of smog chamber data, for the calculation of the SOA yields and partitioning of SVOC into aerosol particles comprised of complex mixtures of primary and secondary organic compounds.

Additional CMAQ modules were modified for compatibility with MADRID, including cloud processes and dry/wet deposition. Detailed descriptions of the CMAQ-MADRID modules are provided elsewhere (EPRI, 2002).

CMAQ was run for the entire BRAVO Study period at 36-km spatial resolution over a domain that corresponds to the 12-km grid domain for the MM5 modeling as shown in Figure 8-1. Temporally and spatially varying boundary conditions were derived from REMSAD estimates that were adjusted using measurements to reduce biases found in the REMSAD SO_2 and particulate sulfate concentrations. For source apportionment the domain was divided into four geographical source areas consistent with the selection of source regions for attribution analysis with REMSAD: Mexico, Texas, Eastern U.S., and Western U.S. (excluding Texas). A total of five sensitivity simulations were conducted. Since modified REMSAD estimates serve as an outer nest to the CMAQ-MADRID simulations, the first sensitivity simulation evaluated the contribution from the boundary conditions of the REMSAD domain (as estimated from the GOCART model) to the fine particulate sulfate load in the CMAQ domain. Each of the ensuing sensitivity simulations evaluated the contribution of a particular source region to the fine particulate sulfate load by removing all other primary sulfur emissions (SO_2 and primary sulfate) within the CMAQ domain and the corresponding contribution to the inner nest boundary conditions.

The contribution from a source region to fine particulate sulfate at Big Bend National Park was defined as the difference between the base case fine sulfate concentration and the fine sulfate concentration predicted by the attribution sensitivity simulation. Source attribution values at the Big Bend monitoring site were estimated for the duration of the BRAVO study period, for each month of the BRAVO study period, and for seven selected episodes. In spite of the potential for large perturbation in emissions to produce non-linear responses to SO_2 oxidation and aerosol sulfate dynamics, the contributions from the source

regions and the outer nest boundary typically aggregated to within 2% of the total final particulate sulfate predicted by the base case simulation.

8.4.4 Synthesized REMSAD and CMAQ

In order to objectively assess and account for any potential source attribution biases from the REMSAD and CMAQ-MADRID models, the model-estimated sulfate attribution results were compared to measured sulfate concentrations throughout Texas. In many respects the synthesis inversion method is similar to Forward Mass Balance Regression (Section 8.3.6), which used Monte Carlo transit probabilities as independent parameters in a linear best-fit model. Modeled source attribution results developed by REMSAD or CMAQ-MADRID as described above were used as the independent parameters for this attribution approach. The synthesis inversion generates a set of source-region specific and temporally varying regression coefficients, that when multiplied by the appropriate model-determined attribution results for each source region produces a "best fit" linear representation of the measured concentration at one or more monitoring sites (Enting, 2002).

If the modeled attribution results are unbiased, all of the regression coefficients would be near 1 and the sulfate source attribution would be similar to those from the original model. Regression coefficients that are significantly different from 1 would indicate that the emissions data and/or the model's treatment of transport, dispersion, chemistry, and/or deposition are biased for the sources regions with non-unity regression coefficients. In that case the synthesized attribution results would be different from those of the original model.

Application of the synthesis inversion method for the BRAVO Study involved regression analysis to determine the daily regression coefficients for each source area by fitting each day's observed particulate sulfate concentrations at most of the BRAVO Study monitoring sites using the REMSAD or CMAQ-MADRID attribution results for those sites. Additional information about the synthesis inversion method and its application for the BRAVO Study is available in Chapters 11 and 12 and the CIRA/NPS report on the BRAVO Study (Schichtel et al., 2004), which is in the Appendix.

8.5 Attribution Source Regions

A set of common source regions was defined as a basis for comparison of the attribution analyses results by various approaches. The source regions include the eastern U.S., western U.S., Texas, and Mexico, where the eastern U.S includes Louisiana, Arkansas, Missouri, Iowa, and Minnesota plus all states further east, and the western states include all other U.S. states except Texas. In addition, most methods (all but CMAQ-MADRID and Synthesized CMAQ-MADRID) accounted for the *Carbón* power plants contributions to particulate sulfate separately from the contributions from other sources in Mexico. Table 8-3 summarizes SO_2 emissions from the BRAVO emissions inventory for the source regions within the REMSAD domain.

Table 8-3. Annual SO_2 emissions (units of million metric tons per year) from the BRAVO emissions inventory for area and point sources within the four attribution source regions of the REMSAD model domain (see Figure 8-4).

	All Regions	Eastern US	Mexico	Texas	Western US
Area sources	4.0	2.5	0.9	0.1	0.5
Point sources	15.4	11.5	1.7	0.9	1.2
Total	19.3	14.0	2.5	1.0	1.7

The various attribution approaches have differing domains and spatial resolutions that affect the extents to which they can accommodate the definitions of the common source regions.

Figure 8-2 shows the source areas used with the TrMB analyses. To simplify the implementation of the TrMB method, source area boundaries were defined by the fewest distinct north-south and east-west lines that reproduce the main features of the political boundaries. These source areas are composited as indicated in the third column of Table 8-2 to form the four source regions and the area surrounding the *Carbón* power plants. The fact that the *Carbón* power plants are situated in an area of otherwise low SO_2 emissions makes this approach for attributing their impact reasonable.

The source regions for the FMBR shown in Figure 8-3 are somewhat different because of the use of virtual source locations on a 100km grid and the model domain. The Texas source region for FMBR excludes a region in the center Texas that has very low SO_2 emissions to reduce the chance of collinearity errors in separately attributing northeast, southeast and west Texas source area contributions to Big Bend particulate sulfate.

The REMSAD modeling used the TrMB boundaries for the four source regions except that they are restricted at the outer edges by the REMSAD domain as shown in Figure 8-4.

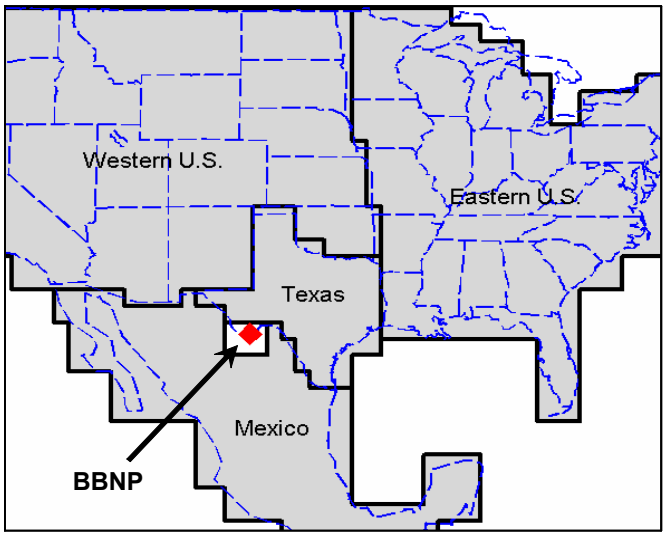

Figure 8-4. *Map of the REMSAD attribution source regions.*

The CMAQ-MADRID model uses a much smaller computational domain than the REMSAD model. As shown in Figure 8-5, this smaller domain is the primary limitation for application of CMAQ-MADRID with the common attribution source regions defined earlier. In order to deal with this limitation, the CMAQ-MADRID attribution sensitivity simulations take advantage of the nesting of its domain within the REMSAD domain. As described above (Section 8.4.3), the base case CMAQ-MADRID simulation uses observation-scaled REMSAD concentration fields as boundary conditions. When performing a CMAQ-MADRID attribution simulation for a specific source region, concentration fields from the corresponding REMSAD attribution simulation are used to derive the appropriate boundary conditions. For example, an attribution simulation for the Western U.S. removes emissions from the Western U.S. cells in Figure 8-5 and removes the Western U.S. influence from the concentrations throughout the boundary of the domain. In order to maintain mass balance, the observation scaling factors used at the boundaries of the base case simulation are applied to all attribution simulations.

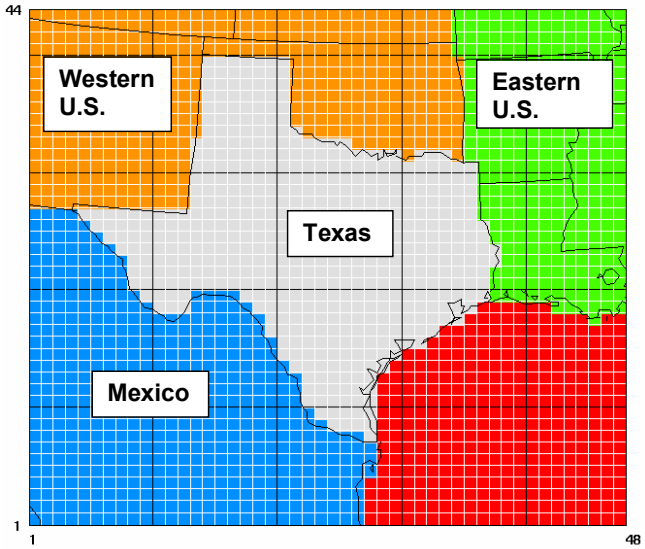

Figure 8-5. *The CMAQ-MADRID attribution source regions.*

None of the attribution approaches exactly matches the boundaries of the set of common source regions. However most of the boundary imperfections and differences are in regions of low emissions density or in some cases the boundaries are at such a great distance from Big Bend (e.g., the northern boundary differences for FMBR and REMSAD compared with TrMB) that impacts from beyond them are not thought to be of practical significance. The source-region boundaries within the CMAQ-MADRID domain are less immune to these concerns because they are much nearer to Big Bend and in some locations the domain boundary is over areas of relatively high emissions density. The use of the scaled REMSAD results as the CMAQ-MADRID boundary conditions for the base case and the attribution simulations is, in essence, a one-way nesting of the CMAQ-MADRID model within REMSAD, so the REMSAD definitions of the source regions beyond its domain are also applicable for CMAQ-MADRID attribution simulations.

8.6 Visualization of Spatial Patterns

Visualizing results from air quality and meteorological measurements and model outputs is both an important diagnostic tool and a useful method for communicating important features of the simulations. Animation of measurement data can provide a sense of the atmospheric flow and evolution of polluted air masses that may be suggestive of the influence of extended emission source regions. Data animation can also be directly compared with of air quality model output animations to qualitatively assess the ability of the model to reproduce the patterns and dynamics inferred from the measurements. For the BRAVO MM5 and REMSAD simulations, several visualization packages were used, including PAVE[1] and Vis5d.[2] Both PAVE and Vis5D are freely available via the Internet. Example visualization "snapshots" are shown in Figures 8-6 through 8-8. Several animations of model results and of measured concentrations can be found in the CIRA-NPS report (Schichtel et al., 2004) in the Appendix.

An important characteristic of such visualizations of model outputs is that they transform tens-of-gigabytes of numbers into coherent patterns that can be processed by the eye and compared with patterns observed in nature. For example, predicted cloud fields can be compared to satellite imagery, or spatial patterns of predicted and observed sulfate concentrations can be compared (as in the example shown in Figure 8-8). These qualitative comparisons of spatial patterns, combined with time series analyses of observations and predictions, provide a more complete understanding of model performance.

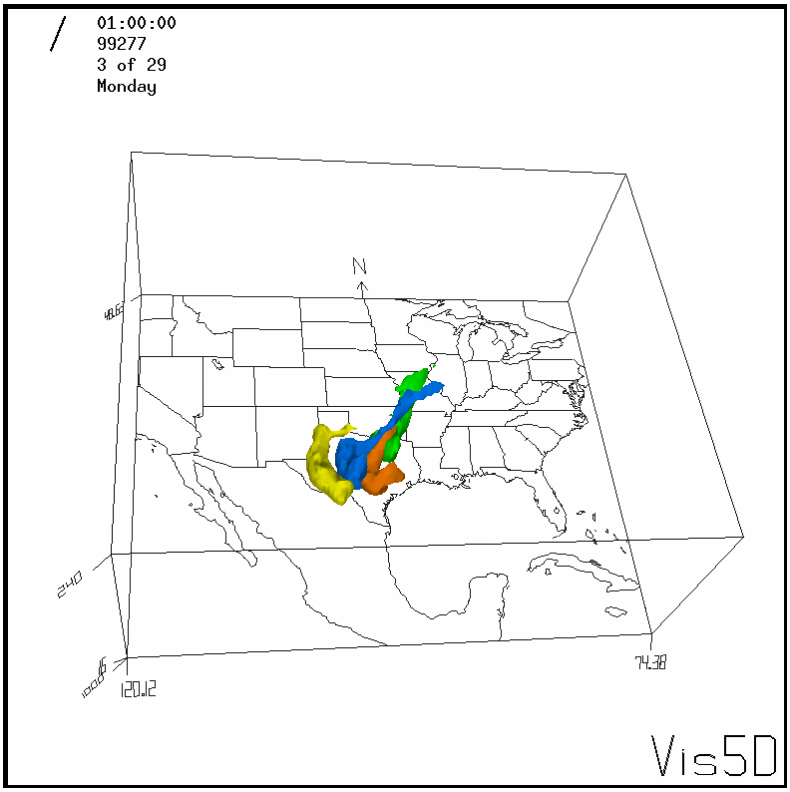

Figure 8-6. *Plumes of a conservative tracer, released from four sites within Texas, as simulated by the REMSAD air quality model (4 October 1999).*

[1] http://www.cmascenter.org/modelclear.shtml

[2] http://www.ssec.wisc.edu/%7Ebillh/vis5d.html

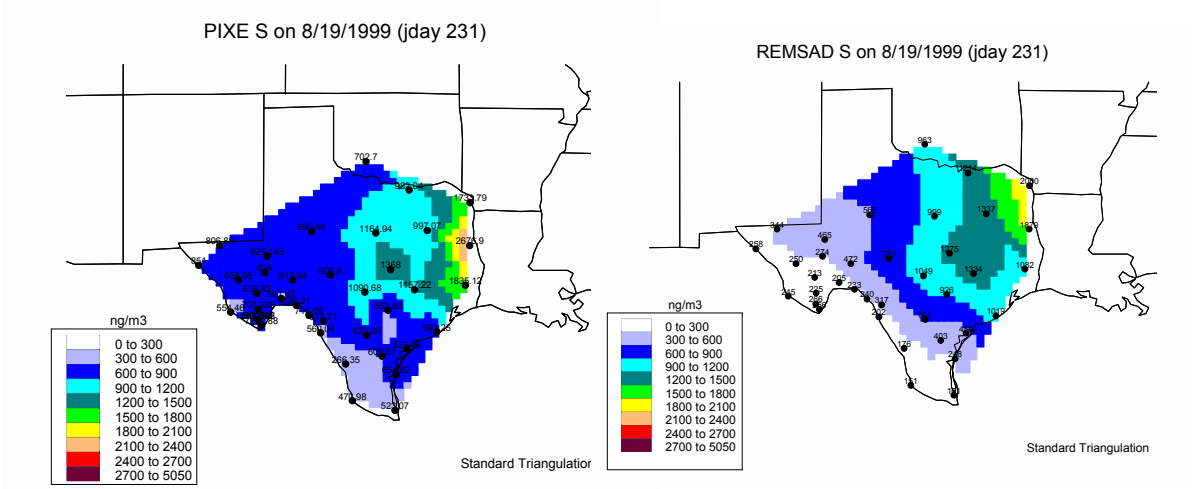

Figure 8-7. *Comparison of spatial patterns of observed sulfate (left) and simulated sulfate (right) (19 August 1999).*

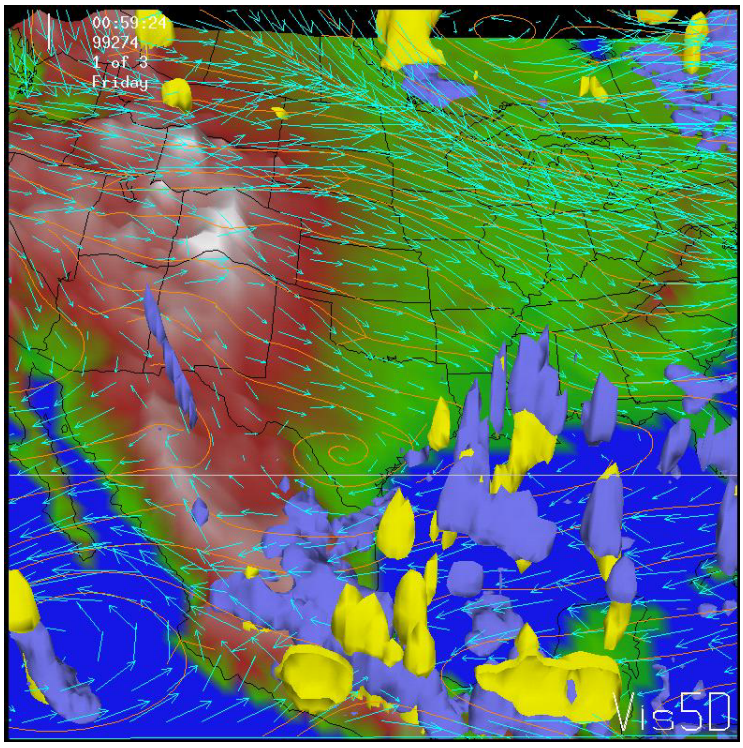

Figure 8-8. *Precipitating clouds (yellow), non-precipitating clouds (blue), wind vectors, and streamlines predicted by the MM5 weather model (1 October 1999).*