2) Ambient Monitoring Data, Air Quality Models and Source Apportionment Methods

2.1 Air Quality, Emissions, and Meteorological Data and Processing

The BRAVO data analyses examined and integrated aerosol, optical, tracer, emission, and meteorological data sets from a number of different sources. This section describes all of the data sets used in the data analyses as well as specialized data processing that was performed.

2.1.1 Air Quality Monitoring During BRAVO

In order to accomplish the goals of the BRAVO study, an extensive air quality monitoring network was deployed from July–October 1999 by several cooperating agencies including the National Park Service (NPS), the U.S. Environmental Protection Agency (EPA), the Texas Commission on Environmental Quality (TCEQ), and the utility sector. This network collected speciated fine (<2.5 μm, $PM_{2.5}$) aerosol and SO_2 gaseous data throughout Texas. Unique perfluorocarbon tracers were also released from four industrial/urban sites within Texas and monitored throughout Texas. In addition, aircraft measurements of high time resolution of light scattering, sulfur dioxide, nitrogen oxides, sulfate, and meteorological parameters were taken over eastern Texas and the Texas Gulf Coast.

An intensive monitoring site was deployed at the K-Bar Ranch, a site within Big Bend National Park (Figure 2-1). At this site aerosol and gaseous species were measured including: speciated PM_{10} (<10 μm) and $PM_{2.5}$ (<2.5 μm) aerosol, aerosol size distributions, size-resolved aerosol data, gaseous nitric acid, ammonia and hydroperoxides, perfluorocarbon tracer concentrations, as well as aerosol scattering, absorption, total extinction and hygroscopic properties.

In this report, data from the study-wide fine aerosol and tracer networks and some additional sulfate concentrations collected at K-Bar are used. These data are more fully described below. More detailed information on the intensive aerosol and gaseous monitoring at K-Bar can be found in *Green et al.* [2000], *Hand et al.* [2002], *Malm et al.* [2003], and *Lee et al.* [2004].

2.1.1.1 Aerosol and SO_2 Monitoring Network

The monitoring network for collecting speciated $PM_{2.5}$ aerosol and SO_2 gaseous data included 37 sites, all located in Texas except Wichita Mountains in Oklahoma (Figure 2-1, Table 2-1). The locations of the monitoring sites were chosen to satisfy several purposes: general gradient sites in Texas (about 100 km apart), border gradient sites at the Texas/Mexico border, Texas/other U.S. states border sites, coastal gradient sites, Big Bend area gradient sites, Class I areas, and sites predominantly downwind of tracer release locations. The purpose of each site is listed in Table 2-1. Thirty-seven sites collected 24-hour integrated samples. Six sites collected 6-hour samples in Big Bend National Park and to its north (Figure 2-2). At the K-Bar monitoring site, located in the center of Big Bend National Park, collocated samplers were collecting 24-, 12-, and 6-hour integrated samples. It should be noted that 19 of the 37 sites did not begin monitoring until July 22 or 23. Also, initially the Big Bend monitor was located outside of K-Bar, but was moved to the K-Bar location on July 22.

All sites collected aerosols using the Interagency Monitoring of Protected Visual Environments (IMPROVE) sampling system [*Malm et al.*, 1994] to collect the $PM_{2.5}$ samples. These samples were analyzed for mass and elemental composition. In addition, fine particulate

ions and carbon as well as SO_2 were measured at some of these sites. Figure 2-1 and Table 2-2 summarize the type of measurements and the parameters for which the filters were analyzed at each monitoring site.

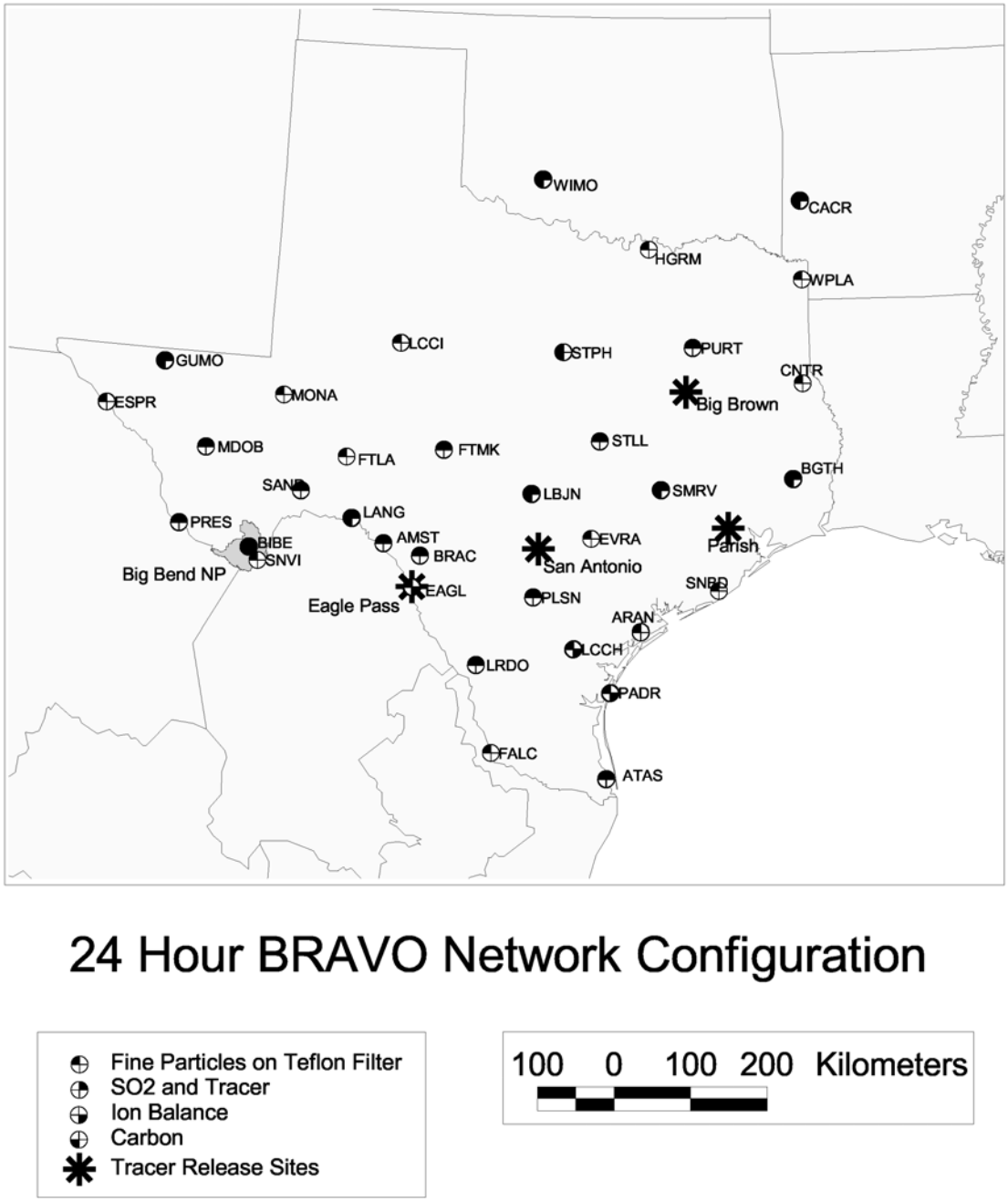

Figure 2-1. 24-hour network of gas and aerosol sampling locations.

Table 2-1. Aerosol and tracer monitoring site abbreviations, names, latitude, longitude, elevation, and purpose.

Site	Name	Latitude	Longitude	Elevation (m)	Purpose
AMST	Amistad	29.47	-101.02	351	Downwind of Carbón/Eagle Pass
ARAN	Aransas	28.32	-96.83	0	Coastal gradient
BIBE	Big Bend (K-Bar)	29.30	-103.18	1052	Receptor/Big Bend area gradient
BGTH	Big Thicket	30.48	-94.35	38	Texas/Louisiana border/gradient
BRAC	Brackettville	29.32	-100.42	335	Downwind of Carbón/Eagle Pass
CNTR	Center	31.83	-94.17	24	Texas/Louisiana border/gradient
EPS	Eagle Pass	28.87	-100.52	274	Mexico border/near Carbón
ESPR	Esperanza	31.17	-105.72	1067	Mexico border gradient
EVRA	Everton Ranch	29.63	-97.65	244	Gradient
FALC	Falcon Dam	26.55	-99.17	61	Border gradient
FTLA	Fort Lancaster	30.67	-101.70	762	Gradient
FTMK	Fort McKavett	30.83	-100.10	671	Gradient
FTST	Ft Stockton	30.92	-102.90	983	Big Bend area gradient
GUMO	Guadalupe Mtns	31.83	-104.82	1659	Class I area
HGRM	Hagerman	33.73	-96.75	244	Texas/Oklahoma border
ATAS	Laguna Atascosa	26.22	-97.35	4	Coastal and Mexico border gradient
LCCI	Lake Colorado City	32.32	-100.90	640	Gradient
LCCH	Lake Corpus Christi	28.07	-97.90	91	Inland ion balance
LANG	Langtry	29.80	-101.55	396	Mexico border/downwind of Carbón
LRDO	Laredo	27.80	-99.45	148	Mexico border gradient
LBJN	LBJ	30.25	-98.63	518	Gradient/downwind of San Antonio
MARA	Marathon	30.20	-103.23	1280	Big Bend area gradient
MDOB	McDonald Observatory	30.67	-104.02	2043	Gradient
MONA	Monahans Sandhills	31.48	-102.80	831	Big Bend area gradient
PADR	North Padre Island	27.45	-97.30	0	Coastal/ion balance
PRSG	Persimmon Gap	29.67	-102.18	915	Big Bend area gradient
PLSN	Pleasanton	28.78	-98.57	122	Gradient
PRES	Presidio	29.57	-104.35	838	Mexico border gradient
PURT	Purtis Creek	32.35	-98.00	187	Gradient/downwind of Big Brown
SNBD	San Bernard	29.90	-95.58	0	Coastal gradient
SNVI	San Vicente	29.12	-103.03	549	Big Bend area gradient
SAND	Sanderson	30.18	-103.22	610	Gradient/downwind of Carbón
SMRV	Somerville Lake	30.33	-96.52	84	Gradient
STPH	Stephenville	32.27	-98.17	274	Gradient
STLL	Stillhouse Lake	31.02	-97.53	213	Gradient
WIMO	Wichita Mtns	34.70	-98.58	488	Class I area
WPLA	Wright Patman Lake	33.30	-94.15	9	Texas/Arkansas/Louisiana border

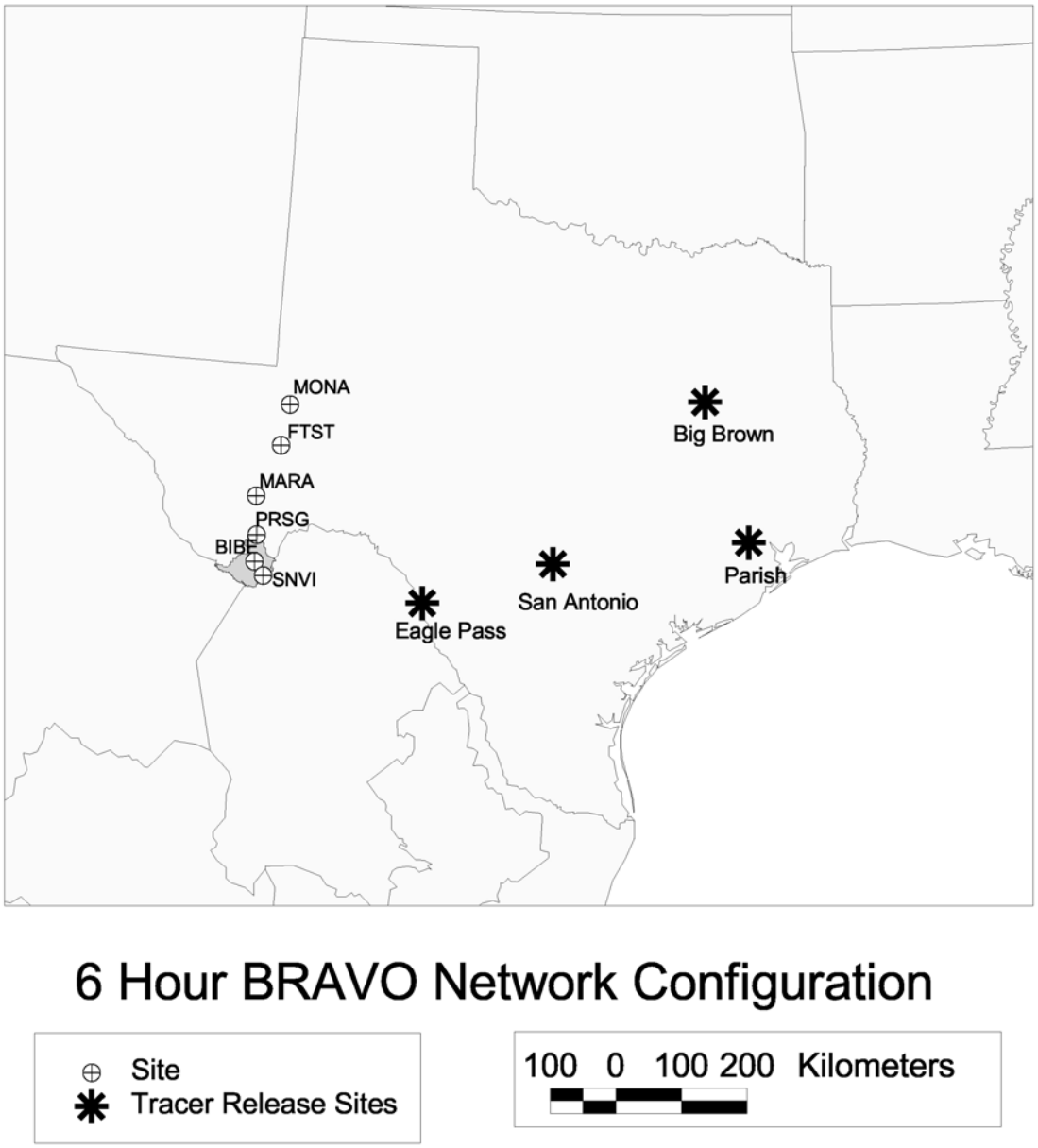

Figure 2-2. 6-hour network of gas and aerosol sampling locations.

Table 2-2. Number of measurement sites by measurement type.

Measurement Type	Number of Sites
24-hour $PM_{2.5}$ elements (H, Na-Pb, mass, b_{abs}) (Teflon filter)	37
24-hour SO_2 and tracer	18
24-hour $PM_{2.5}$ carbon (quartz filter)	7
24-hour $PM_{2.5}$ ions (nylon filter)	4
6-hour $PM_{2.5}$ elements, SO_2, tracer	6
24-hour PM_{10} elements, ions, carbon	1 site at K-Bar
12-hour $PM_{2.5}$ elements, ions, carbon	1 site at K-Bar
Collocated 24-hour $PM_{2.5}$ elements, ions, carbon, SO_2, tracer	1 site at K-Bar
Collocated 24-hour PM_{10} elements, ions, carbon	1 site at K-Bar
Collocated 6-hour $PM_{2.5}$ elements, SO_2, tracer	1 site at K-Bar

The IMPROVE monitoring system used to collect the samples was designed for the IMPROVE network and has been operated extensively in the network and during field studies since 1988 [*Malm et al.,* 1994]. The IMPROVE sampler can employ up to four independent modules, three for monitoring fine particles and the fourth module to collect particles less than 10 μm. Each module incorporates a separate inlet array, filter pack, and pump assembly; however, all modules are controlled by one timing mechanism. It is convenient to consider a particular module, its associated filter, and the parameters measured from the filter as a channel of measurement (e.g., channel A).

Channels A, B, and C were each equipped with a 2.5 μm cyclone inlet. Channel A, which utilizes a Teflon filter, was analyzed for fine mass ($PM_{2.5}$) gravimetrically, nearly all elements with atomic mass number >11 (Na) and <82 (Pb) by proton induced x-ray emission (PIXE) and/or by x-ray florescence (XRF), elemental hydrogen by proton elastic scattering analysis (PESA), and light absorption.

Channel B utilized a single Nylasorb filter as a collection substrate. A denuder, coated with a mixture of Na_2CO_3 and reagent grade glycerol, before the nylon filter removes nitric acid vapors. The material collected on the filter was extracted ultrasonically in an aqueous solution that was subsequently analyzed by ion chromatography for the anions sulfate, nitrate, nitrite, and chloride. Sodium, ammonium, potassium, magnesium, and calcium ion concentration were also measured using extracts from these filters.

Channel C utilized tandem quartz fiber filters for the collection of fine particles and the estimation of an organic carbon artifact from a secondary filter. The positive organic artifact is the results of adsorption of organic gases on the quartz filter. These filters were analyzed by thermal optical reflectance (TOR) for elemental and organic carbon [*Chow et al.,* 1993].

Additional channel A, B, and C samplers were run with PM_{10} rather than $PM_{2.5}$ inlets so that coarse particle species mass concentrations could be estimated by differencing PM_{10} and $PM_{2.5}$ concentrations. The 24-hour filter samples were collected each day starting at 08:00 Central Daylight Time. Exposed cassettes from channels A, B and C were stored in sealed plastic bags and shipped for storage and analysis.

In order to assure high quality data, experiments were designed such that observables could be estimated or modeled in a number of different ways. Mass was gravimetrically determined for both D_{aero} <10 and 2.5 μm, which can be compared to reconstructed mass based on the summation of measured aerosol species. Dry and ambient scattering coefficients were measured, which in turn can be compared to modeled scattering coefficients that are based on aerosol species mass and size measurements.

2.1.1.2 Tracer Data

As part of the BRAVO study, four perfluorocarbon tracers were released from four locations in Texas and monitored at 24 sites throughout Texas. Each tracer monitoring site was collocated with a PM2.5 module A aerosol monitor. The objectives of the tracer component of the study were to tag source areas with the potential for transport of aerosols and gases resulting in significant visibility impairment at Big Bend National Park, evaluate the dispersion (transport and diffusion) capabilities of Eulerian and Lagrangian models deployed for understanding the causes of haze during BRAVO, and to evaluate the receptor-oriented source attribution techniques. In this report the tracer data were only used to evaluate the transport and dispersion

mechanisms in the airmass history and air quality models and the receptor source attribution techniques.

The four tracer release points are Eagle Pass, the Big Brown coal-fired power plant, San Antonio, and the Parish coal-fired power plant in Houston (Figure 2-2). The Eagle Pass release point is about 250 km southeast of Big Bend National Park. The releases at Eagle Pass were made from the top of a 107 m (350 feet) tower to simulate the releases from the Carbón I & II coal-fired power plant stacks located 30 km to the south. At Eagle Pass, ocPDCH was continuously released during the four-month study. In addition, during the first two months of the study PDCB was released on alternate days and PTCH was released during the daytime. The timing tracers were designed to allow estimation of the transport times from Eagle Pass to Big Bend National Park. Big Brown is located at the northern end of the Lignite Belt near Fairfield, TX, and operated by Texas Utilities. The tracer was introduced into the exhaust duct just before the point where the duct enters the stack. The iPPCH tracer was continually released from the Big Brown stack throughout the four-month study.

The two timing tracers, PDCB and PTCH, released from Eagle Pass during the first half of the study, were moved to San Antonio (PDCB) and to the Parish Power Plant in Houston, TX, (PTCH) and were continuously released from these sites from September 17, 1999, to the end of October. The San Antonio release system was located at an air quality monitoring station run by the utility company, City Public Service. The site was inside the beltway and was chosen to represent the San Antonio area source. The release in the Houston area was made from the Parish Power Plant operated by Reliant Energy, located about 64 km southwest of Houston. Tracer was introduced into the stack of Boiler 7 through a sampling port at the 100 ft. level.

Tables 2-3 and 2-4 present the design tracer release rates and schedule for tracer release from the four sites. However, as shown in Figures 2-3 to 2-6 the tracer release rates fluctuated from day to day. At times the fluctuations were significant; for example, the Big Brown tracer was not released from 8/15–17 and 10/8–16. Therefore, the tracer release rates need to be considered when interpreting the tracer concentrations.

Table 2-3. Tracer release schedule first phase of study.

Location	**Eagle Pass**	**Eagle Pass**	**Eagle Pass**	**Big Brown (NE Texas)**
Tracer	Perfluoro-1,2-dimethylcyclo hexane	1,1,2,2,3,4-hexafluoro-3,4-bis(trifluoromethyl) cyclobutane	Perfluoro-1,3,5-trimethylcyclo hexane	perfluoro-isopropylcyclohexane
Tracer Abbrev	**ocPDCH**	**PDCB**	**PTCH**	**iPPCH**
Release period	7/5/99-11/1/99	7/5/99-9/13/99	7/5/99-9/13/99	7/9/99-11/1/99
Design Release Rate (kg/hr)	0.155 (2.6 g/min)	0.525 (8.75 g/min) alternate days (8am-8am) CDT	0.184 (3 g/min) 8am – 8pm CDT only	0.092 (1.5 g/min)

Table 2-4. Tracer release schedule for the second half of the study.

Location	**Eagle Pass**	**San Antonio**	**Parish (Houston)**	**Big Brown (NE Texas)**
Tracer	Perfluoro-1,2-dimethylcyclo hexane	1,1,2,2,3,4-hexafluoro-3,4-bis(trifluoromethyl) cyclobutane	Perfluoro-1,3,5-trimethylcyclo hexane	perfluoro-isopropylcyclohexane
Tracer Abbrev	**ocPDCH**	**PDCB**	**PTCH**	**iPPCH**
Release period	7/5/99-11/1/99	9/17/99-11/1/99	9/17/99-10/25/99	7/9/99-11/1/99
Design Release Rate (kg/hr)	0.155 (2.6 g/min)	0.442 (7.4 g/min)	0.115 (1.9 g/min)	0.092 (1.5 g/min)

The tracer concentrations were monitored at 24 of the 37 aerosol receptor sites (Figure 2-1, Table 2-1). Most of the tracer, aerosol, and SO_2 sampling sites collected 24-hour averaged data; however, six 6-hour tracer sites were located along an approximately 300 km long arc running north-northeast from the Rio Grande River in Big Bend National Park (Figure 2-2). The K-Bar monitoring site also had a collocated 1-hour sampler. The 6-hour monitors had a 4 times larger flow rate than the 24-hour samplers. Therefore, both the 6- and 24-hour samples had equal sample volumes for analysis. More details of the tracer release including the reasons for each of the deviations from the optimum release rates can be found in *Watson et al*. [2000].

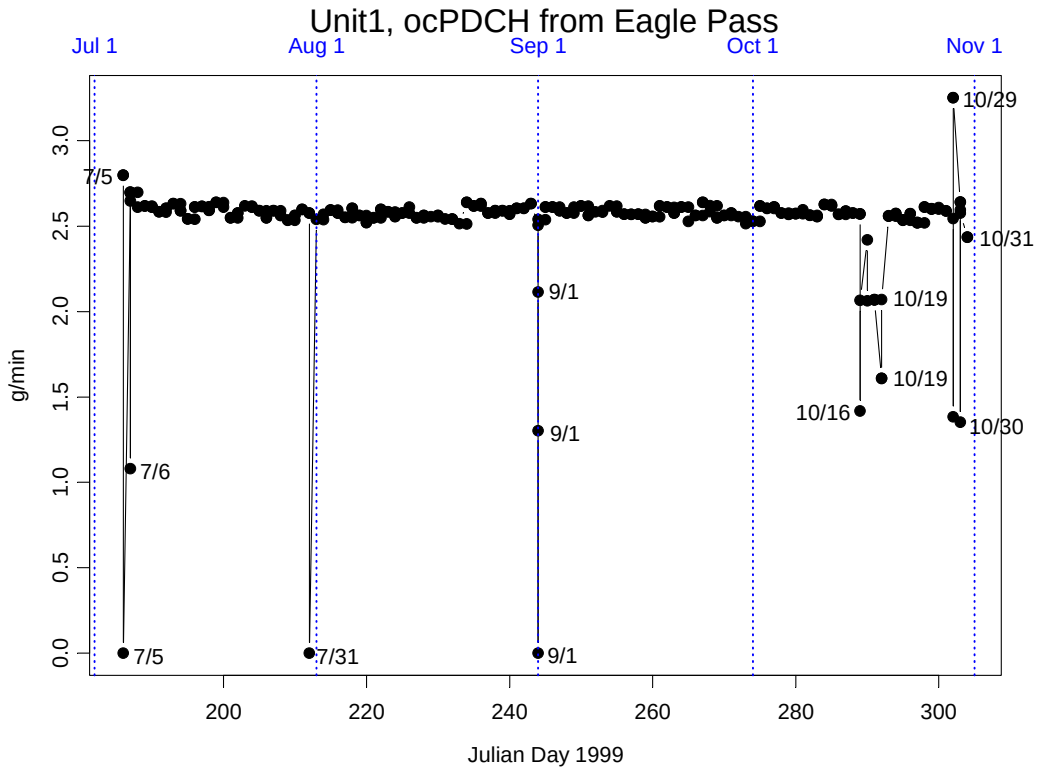

Figure 2-3. Release rates of ocPDCH from Eagle Pass (g/min) during the BRAVO study.

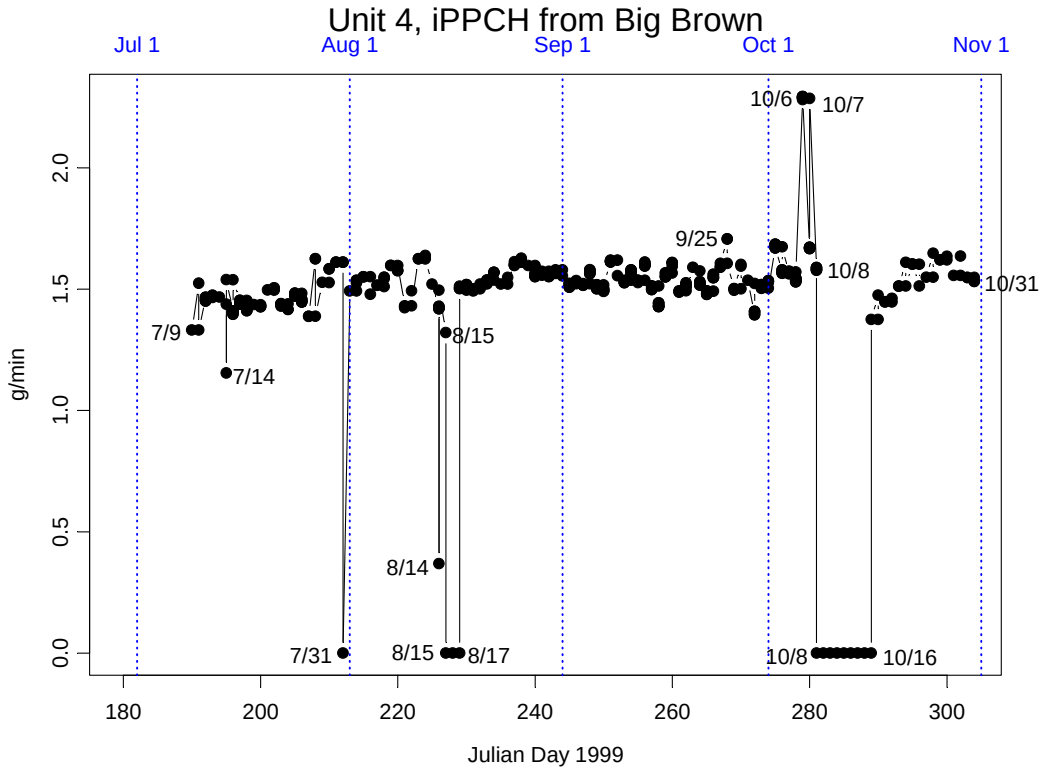

Figure 2-4. Release rates of iPPCH from Big Brown (g/min) during the BRAVO study.

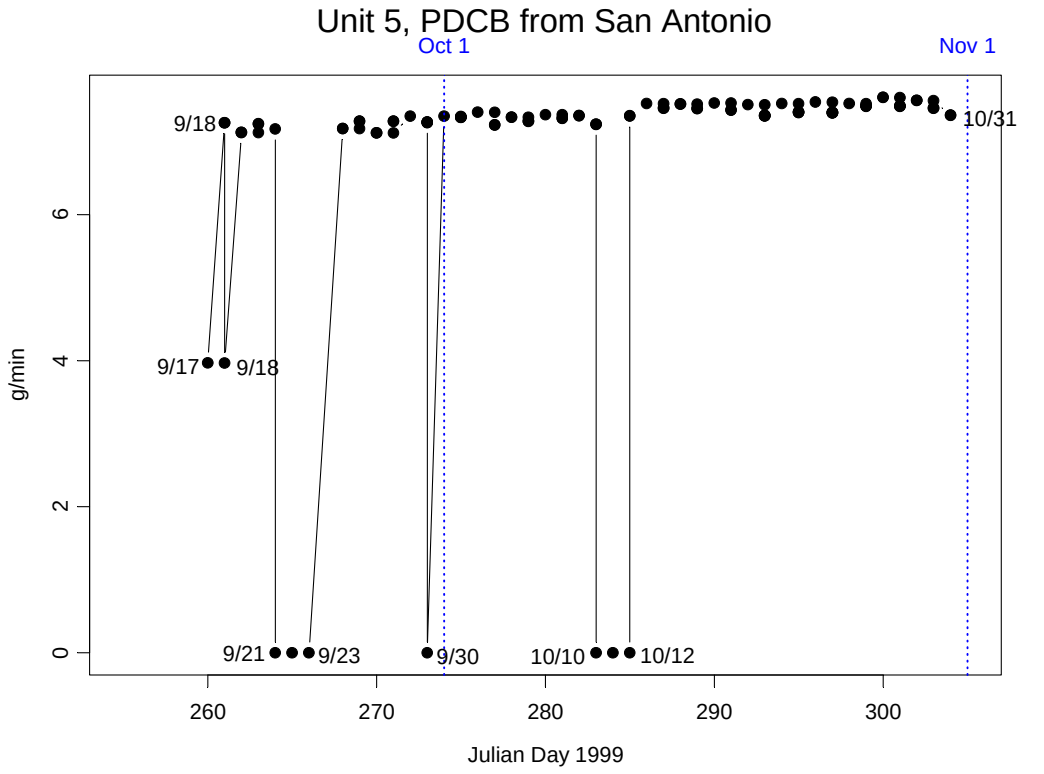

Figure 2-5. Release rates of PDCB from San Antonio (g/min) during the second half of the study.

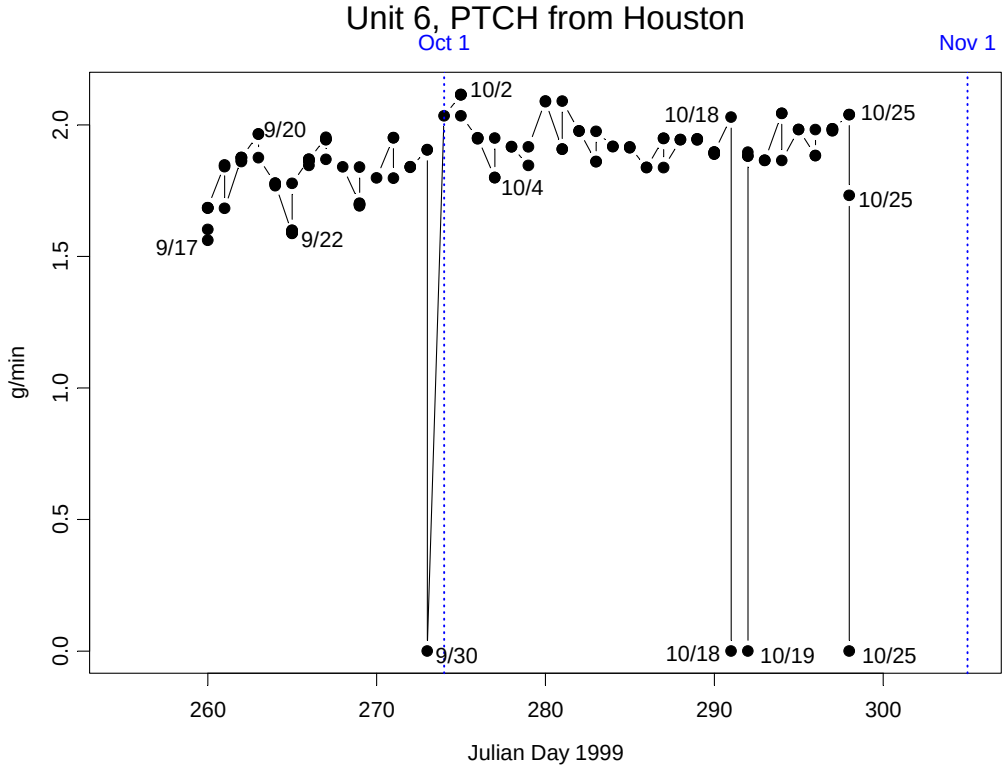

Figure 2-6. Release rates of PTCH from Houston (g/min) during the second half of the study.

2.1.1.3 University Research Glassware (URG) Cyclone/Annular Denuder/Filter Pack Sampling System

Colorado State University (CSU) sampled aerosol composition by collecting 24-hour integrated filter samples using a URG denuder/filter-pack [*Lee and Collett*, 2002; *Lee et al.* 2004] assembly with a $PM_{2.5}$ inlet. Samples were run from 08:00 (CDT)–08:00 (CDT). The filters were analyzed for major ions by Ion Chromatography (IC) using analytical equipment located on site. CSU samplers were located at K-Bar throughout the BRAVO study period (July 1–Oct. 31, 1999).

Two parallel sampling trains were used to facilitate a rapid change between samples. Air was drawn through the sample train at a nominal value of 10 lpm (actual volumetric flow) in order to provide the desired 2.5 μm aerodynamic diameter size cut in the URG cyclone. Flow was provided by a vacuum pump and controlled by a mass flow controller. A dry gas meter was installed in-line between the filter pack and mass flow controller in order to provide an integrated measure of total sample flow. The sample flow read by the dry gas meter was corrected to ambient conditions by correcting for the pressure drop through the denuders and filters. An inline pressure measurement was used to make this correction. The URG samplers were mounted on the top of the scaffolding with the pumps located on the ground below.

Ambient air was drawn through a cyclone (D_{50} = 2.5 μm) and along etched glass denuder walls (air passes through two denuders in series), which were coated with chemicals that absorb the gaseous species of interest. NaCl (0.1%) coated the first denuder for collection of HNO_3 and the second denuder was coated with 0.5 g citric acid in 50 mL of methanol to collect ambient NH_3. Pre-filter collection of NH_3 helps preserve acidic aerosol samples [*Perrino et al.*, 1990].

The remaining air stream was then filtered through 47 mm diameter Teflon, nylon, and coated glass fiber filters in series. The Teflon filter (Gelman Teflo, 2.0 μm pore size) was used to collect particulate matter.

The nylon membrane filter (Gelman Nylasorb) and citric acid impregnated glass fiber filter were used to capture nitric acid and ammonia, respectively, which may be volatilized from particulate matter on the Teflon filter.

The primary $PM_{2.5}$ ionic species, Cl^-, SO_4^{2-}, NO_3^-, Na^+, NH_4^+, K^+, Mg^{2+}, and Ca^{2+}, and gas phase concentrations of HNO_3 and NH_3 were measured. Also, $PM_{2.5}$ aerosol acidity was measured on-site for all study days.

2.1.1.4 Integrating Nephelometers

The details of ambient nephelometer measurements were covered in *Malm et al.* [1994], *Molenar* [1997], *Malm and Day* [2000], and *Day et al.* [2000] and will only be briefly reviewed here. Five Optec NGN-2 integrating nephelometers, in various configurations, were operated during the study. One nephelometer was fitted with an Anderson PM_{10} inlet, while two Optec nephelometers were operated at reduced flow rates (113 l min^{-1}) and were fitted with a Bendix-240 cyclone inlet with a 2.5 μm cutpoint. Two nephelometers utilized the open-air configuration and were operated using standard IMPROVE protocols [*Air Resource Specialists*, 1995]. Thermistors were placed inside the inlet, where sample air was assumed to be at ambient temperature, and at the sample exit, where sample air temperature should have reflected any heating of the aerosol. Monitoring the difference in sample temperature between the inlet and the outlet of each nephelometer allowed determination of whether heating of the sample had occurred and if there was a subsequent change in sample RH. Generally, the degree to which the sample was heated was less than 0.5°C. The accuracy of the nephelometer measurements in this configuration and the degree of heating were discussed in some detail by *Day et al.* [2000].

A sixth and seventh nephelometer (Radiance Research M903 integrating nephelometers) were configured to measure dry and controlled humidity (D_{aero} <2.5 μm) scattering. In the controlled humidity configuration, prior to measurements by the nephelometer, air was drawn through a temperature-controlled humidity conditioner consisting of Perma Pure Nafion dryers, while temperature was controlled by placing the dryers in a constant temperature water bath. Because temperature changes in the sampling plumbing can cause unwanted and unknown RH changes, temperatures were monitored throughout the system.

2.1.1.5 Transmissometer

Transmissometers are calibrated to measure the irradiance, at 550 nm, of a light source after the light has traveled over a finite atmospheric path. The transmittance of the path is calculated by dividing the measured irradiance at the end of the path by the calibrated initial intensity of the light source. The average extinction of the path is calculated using Bouger's law from the transmittance and length of the path. The measurement is ambient in that air samples are not passed through an enclosed chamber.

Transmissometers used in this study are the OPTEC, Inc., LPV-2 instruments, which have been in use since 1986. Their use in remote locations such as national parks is discussed by *Molenar et al.* [1989] while their use in urban settings is presented by *Dietrich et al.* [1989]. Careful operation of the transmissometer (daily cleaning of optics and pre and post calibrations) should result in extinction measurements with an accuracy of about 10% *Molenar et al.* [1989].

The transmissometer used in this study was located near the particle samplers but on the rim of the canyon. Therefore, the separation between the transmissometer and all other instruments was about 1.5 km.

2.1.1.6 Aethalometer

The aethalometer collects aerosol continuously on a quartz fiber filter tape, while measuring the optical transmittance through the filter [*Hansen et al.*, 1982; *Hansen et al.*, 1984]. The rate of decrease of optical transmittance as a function of the rate of increase of filter loading has been found to be proportional to the amount of absorbing carbon loading. The aethalometer can be operated at a flow rate of 5-10 l min^{-1} and is purported to have an accuracy of about 10%. Its sample air was extracted from a sampling plenum that was also fitted with a 2.5 μm cyclone inlet.

2.1.2 BRAVO Emission's Inventory

The BRAVO emissions inventory consists of hourly, gridded emissions of CO, NH_3, NO_x, PM_{10}, $PM_{2.5}$, SO_2, and various VOCs for the U.S. and Mexico [*Kuhns et al*., 2001]. Several different sources were used to determine emissions within the Regional Modeling System for Aerosols and Deposition (REMSAD) model domain, including EPA's National Emission Inventory (NEI) database [*U.S. Environmental Protection Agency*, 2003b], emission factors for Mexican point and area sources from the National Institute of Ecology of Mexico, and biogenic VOC emissions from the Biogenic Emissions Inventory System (BEIS) [*Geron et al*., 1994]. Emissions were processed with the Sparse Matrix Operator Kernel Emissions (SMOKE) model [*MCNC*, 2002].

The concentration of sulfate aerosol is largely determined by regional emissions of SO_2 (Figure 2-7). Large SO_2 sources are prevalent in the eastern U.S. (e.g., power plants along the Ohio River Valley) and eastern Texas. There are also several large SO_2 point sources in northern Mexico within 250 km of Big Bend National Park. The SO_2 sources in Texas include coal-fired power plants, oil refineries, and carbon black producers. Several coal-fired power plants are sited in and near a lignite belt which runs from northeast Texas towards the Mexican state of Coahuila. The principal SO_2 emitters in Mexico consist of coal-fired power plants and fuel oil refining and combustion operations. The Carbón I & II power plants, located near the U.S./Mexico border and 210 km southeast of Big Bend National Park, are the largest coal-combustion facilities in Mexico [*Green et al*., 1999]. SO_2 emissions for the largest point sources in the U.S. were estimated using continuous emissions monitoring (CEM) data, with attendant stack parameter data. These data were generally not available for Mexican point sources, and hence their emissions are estimated using plant through-put calculations [*Kuhns et al*., 2001]. In addition to these anthropogenic SO_2 sources, the Popocatepetl volcano, located 100 km southeast of Mexico City, intermittently releases large amounts of SO_2.

SO_2 emission in southern Mexico, Central America, Cuba, and shipping lanes were not included in the inventory. Contributions from these source areas are expected to be small.

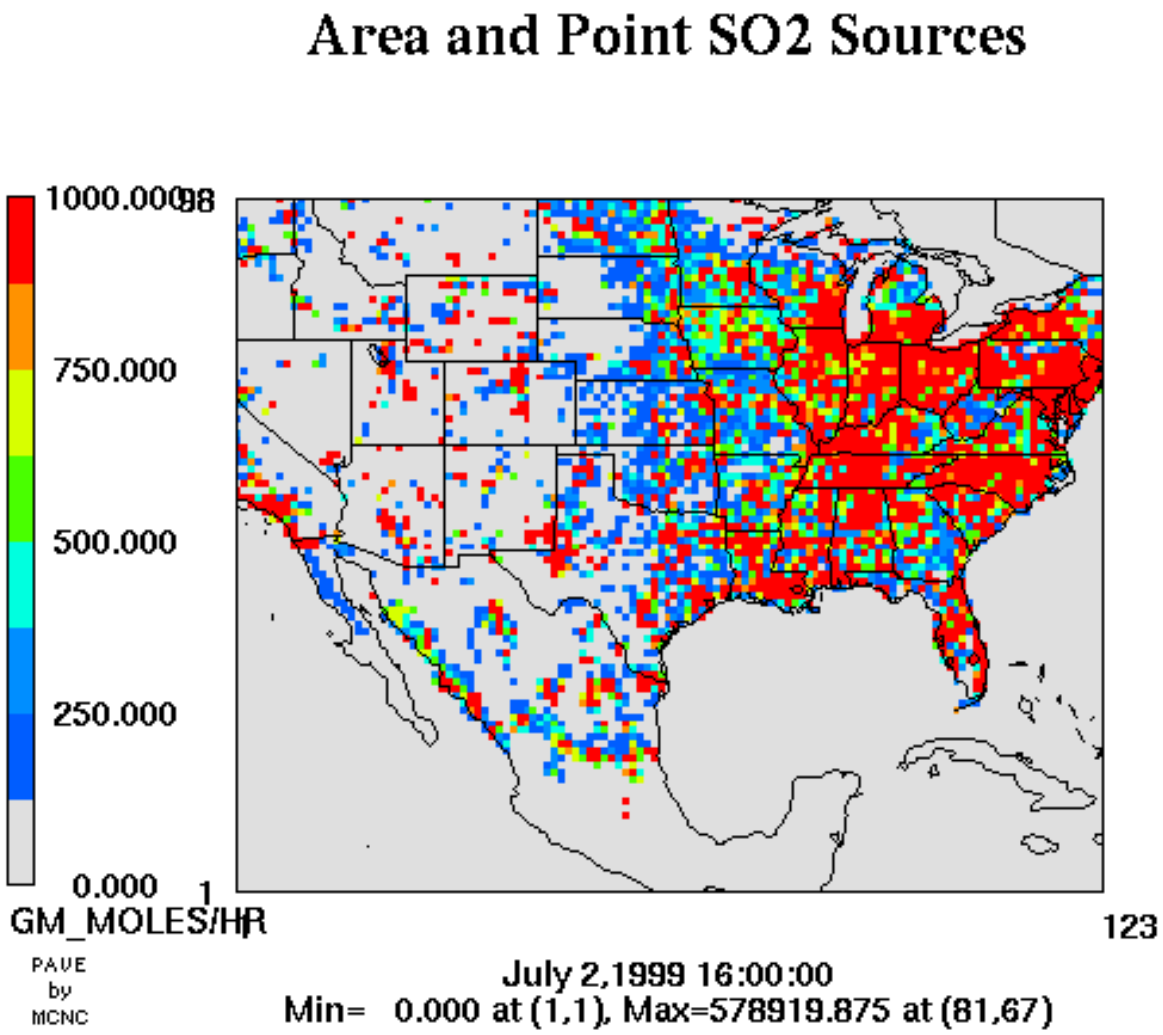

Figure 2-7. Example SO_2 emissions (moles hour^{-1}) from point and area sources within the REMSAD model domain.

Regional-scale SO_2 emission rates for the base emissions scenario are shown in Figure 2-8. SO_2 sources are defined as either elevated point sources (e.g., coal-fired power generation) or area sources (e.g., diesel combustion from mobile sources). Point sources constitute the bulk of the inventory at 73%. Aggregating all of the sources within the REMSAD domain yields a total annual SO_2 emission rate of approximately 19 Tg/yr. Of this, 14 Tg/yr of SO_2 is contributed by the eastern U.S., followed by Mexico at 2.5 Tg/yr, the western U.S. at 1.8 Tg/yr, and Texas at 1.0 Tg/yr.

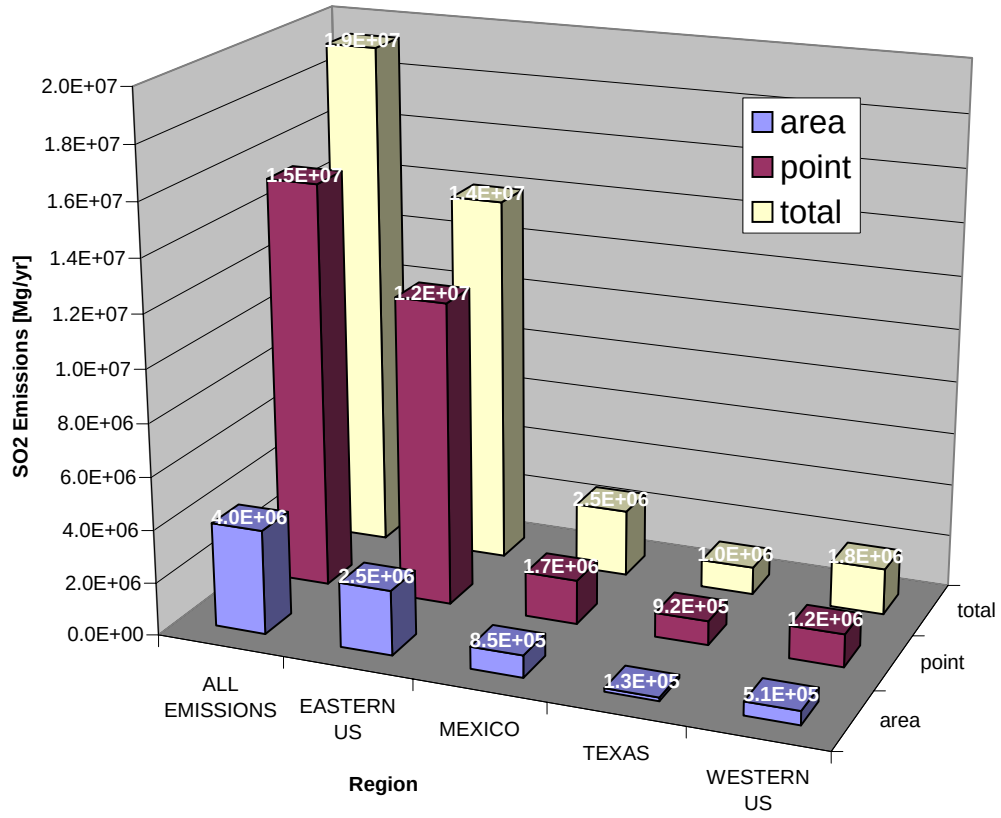

Figure 2-8. SO_2 emissions (Mg/yr) for each source region.

2.1.3 Meteorological Data

2.1.3.1 MM5

The MM5 model was applied for the BRAVO study period to provide custom metrological data fields for use in the air quality modeling. The MM5 model was run in non-hydrostatic mode using a continental-scale 36 km domain with nested 12 km and 4 km domains. MM5 employed observational nudging by incorporating the available surface and upper air measured wind data from the National Weather Service and data from the four BRAVO radar wind profilers. Only the 36 km wind fields were used in the following analyses. See *Seaman and Stauffer* [2003] for details of the MM5 modeling for BRAVO.

2.1.3.2 EDAS

The National Weather Service's National Centers for Environmental Prediction (NCEP) maintain 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 [*Zapotocny et al.,* 2000; *Zapotocny et al.,* 2002] 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 generates meteorological fields every 3 hours. These data are interpolated to a 40 km Lambert Conformal grid and isobaric levels. The National Oceanic and Atmospheric Administration's (NOAA) 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. The grid in ARL's EDAS archive is presented in Figure 2-9.

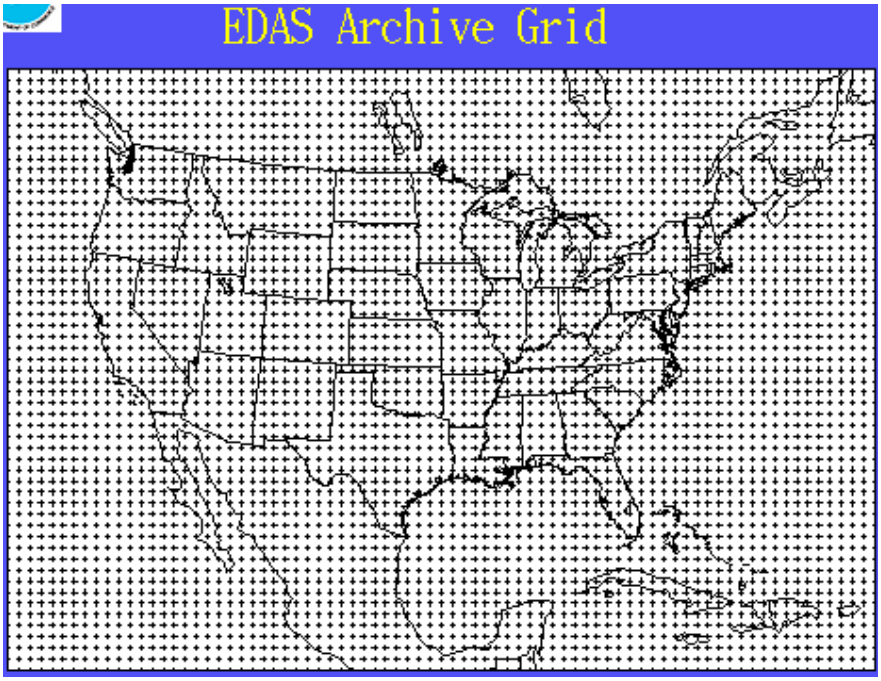

Figure 2-9. The spatial domain and horizontal grid used in ARL's EDAS Archive.

2.1.3.3 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 [*Sela*, 1980]. 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 to 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 [*Petersen and Stackpole*, 1989].

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 (Figure 2-10) and saves out 13 of the pressure levels [*Stunder*, 1997]. 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].

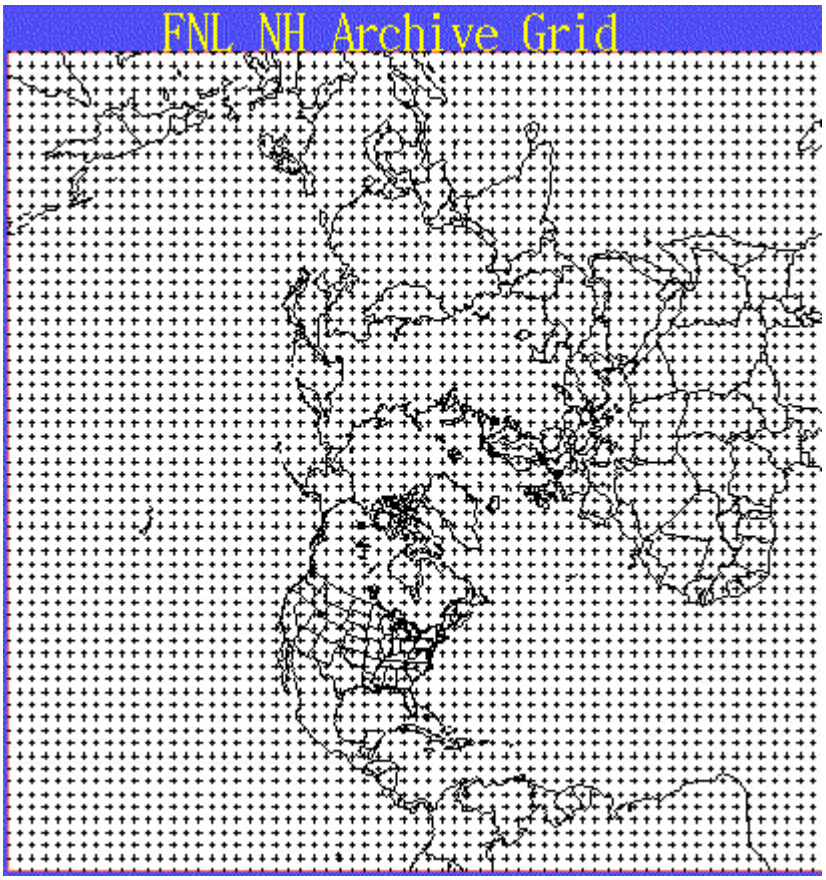

Figure 2-10. The spatial domain and horizontal grid used in ARL's FNL Archive.

2.1.3.4 Rawinsonde Data

The usual upper air data used as input to the Atmospheric Transport and Dispersion (ATAD) model are data from the FSL/NCDC Radiosonde Data Archive. These data are available on compact disk for 1946–1997 from the National Climatic Data Center (NCDC). For later years, including the BRAVO time period, the data can be obtained from a web site maintained by the Forecast Systems Laboratory (FSL), a division of NOAA at http://raob.fsl.noaa.gov/. Data include station information, height, pressure, wind speed, wind direction, temperature, and dew point temperature.

The real-time data archive consists of all observations transmitted via the North American Global Telecommunications Service (GTS), primarily from North and Central American civilian and some military rawinsonde sites. Since late 1998, the archive has been upgraded to include all international observations transmitted via GTS. All of the raw data goes through quality control analyses including extensive gross error and hydrostatic consistency checks. At most rawinsonde sites, weather balloons are launched twice daily at 0 and 12 Greenwich Mean Time (GMT). There are generally 1–3 launch sites per state. Figure 2-11 below shows the locations and number of observations at each site during BRAVO.

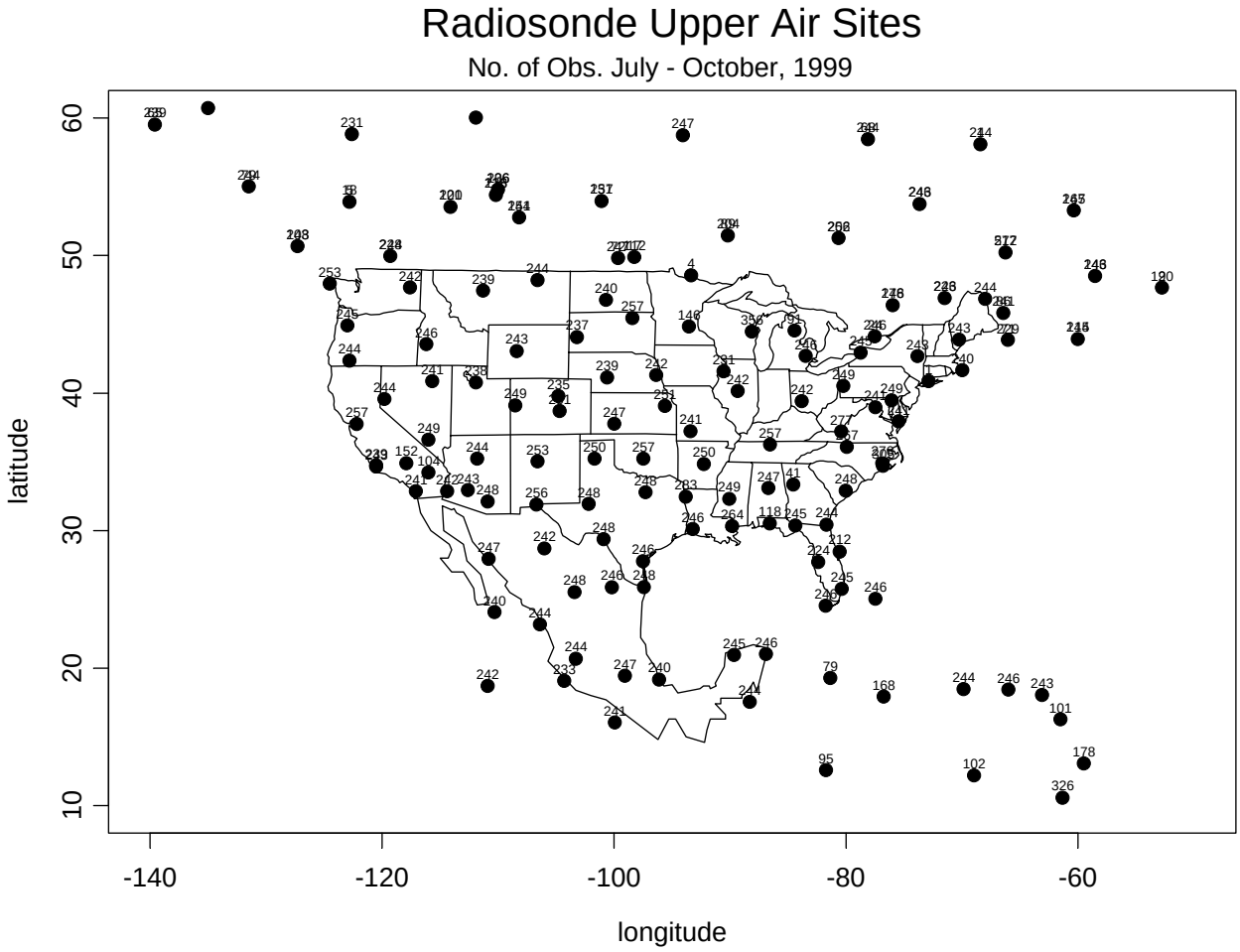

Figure 2-11. Radiosonde upper air sites.

2.1.3.5 BRAVO Wind Profilers

The Bravo study deployed four radar wind profilers in eastern and southern Texas (Figure 2-12). These wind profilers collected wind speed and direction as a function of height, and the data was generally reported every hour. Radar wind profilers equipped with a radio acoustic sounding system (RASS) also obtained vertical profiles of virtual temperature, although usually only to 500–1500 m.

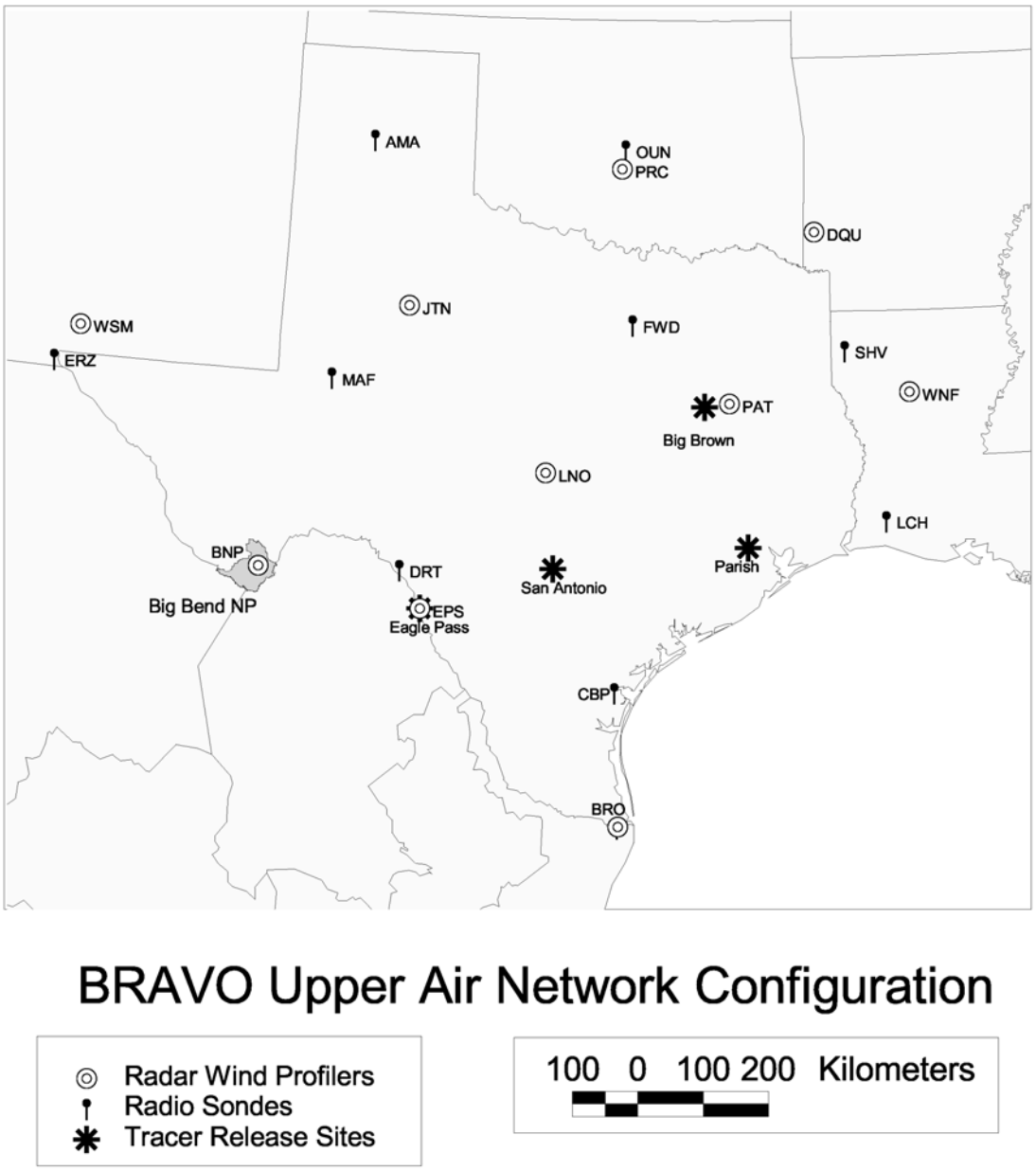

Figure 2-12. BRAVO wind profiles and other upper air meteorological sites.

Wind profiler data were collected in two modes. The 60 m mode gives better vertical resolution but less height coverage while the 100 m mode can extend to higher heights. Generally, the 60 m mode data are used for the lowest 700 mb, with the 100 m data used above that level. See Appendix 2A for a sampling of some of these data presented in graphical form.

2.1.3.6 Surface Meteorological Data

NOAA's Environmental Technology Laboratory (ETL) collected 10-minute and 60-minute averaged data from standard 10 m meteorological towers at the same four sites as the wind profilers from July 1 to October 31, 1999. Data collected included surface pressure, 2 m temperature and relative humidity, wind speed, wind direction, down-welling solar radiation, net

radiation, mixing ratio, precipitation, and peak gusts. Appendix 2B contains some of these data summarized graphically.

2.1.3.7 Relative Humidity Sensors

Three Rotronics mp 100f combination relative humidity/temperature sensors were housed in PVC holders and aspirated by a fan. The reported accuracy of the relative humidity sensor is ±2%. The flow rate through the holder was approximately 120 l min^{-1}. The sensors were approximately six feet above ground level, six feet from each other, and mounted near the inlets of the nephelometers.

2.2 Transport and Air Quality Models

The BRAVO study source attribution analyses, conducted by the NPS/CIRA group, used the Regional Modeling System for Aerosols and Deposition (REMSAD) air quality model, a prognostic Eulerian grid model capable of simulating some of the atmospheric chemical and physical processes. Also, the three transport models: Air Resource Laboratory Atmospheric Transport and Dispersion (ATAD); HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) and the CAPITA Monte Carlo model, were used to calculate back trajectories and source plume dispersion estimates.

This section describes each model. The REMSAD model is evaluated against tracer data in chapter 5 and sulfur simulation and source attribution results are presented in chapter 6. Back trajectory estimates from all three models driven by MM5, EDAS/FNL, and measured RAOB data are presented in chapter 4. The CAPITA Monte Carlo source plume dispersion simulation is evaluated against the tracer data in chapter 5. Back trajectory analyses based on the residence time analysis using the CAPITA Monte Carlo model are presented in section 8.1.3. The CAPITA Monte Carlo model was also used to simulate the dispersion of a source's plume which was then used in the FMBR source apportionment technique. This source apportionment technique is described in section 2.3.2.2, evaluated in section 7.1.2 and 7.2.2 and the sulfate source apportionment results presented in section 8.2.3. All three transport models were used in the TrMB source apportionment technique described in section 2.3.2.1, evaluated in section 7.1.1 and 7.2.1 and source apportionment results are presented in section 8.2.2.

2.2.1 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 [*SAI*, 2001; *Seigneur et al.,* 1999]. 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 (see, for example, *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 covers most of the contiguous U.S. 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 and 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 only "VOC" (most anthropogenic organics), "CARB" (carbonyls), and "ISOP" (biogenic organics) represented. The inorganic and radical chemistry portions of μCB-IV are identical to CB-IV. 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 [*Kim et al*., 1993].

SO_2 and SO_4 boundary concentrations were created using results from the Global Ozone Chemistry Aerosol Radiation Transport (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.

Wind fields, temperature fields, and other meteorological data for BRAVO were simulated by MM5 [*Grell et al.,* 1994]. MM5 simulations were provided by the Department of Meteorology at the Pennsylvania State University. MM5 is a non-hydrostatic, prognostic, regional-scale meteorological model. Physical parameterizations used for the BRAVO study include the *Kain-Fritsch* [1990] deep-convection parameterization, the *Dudhia* [1989] resolved-scale precipitation model, the *Shafran et al.* [2000] 1.5-order turbulence closure scheme, the *Zhang-Anthes* [1982] surface-flux model, and the *Dudhia* [1989] two-stream broad-band atmospheric radiation scheme. Four dimensional data assimilation (FDDA) [*Stauffer and Seaman*, 1994], using both analysis nudging and observational nudging, was employed. Recent studies have shown the importance of using FDDA when simulating the meteorology of air pollution events [*Seaman*, 2000; *Barna and Lamb*, 2000]. Meteorological observations available for FDDA included surface and upper-air measurements from the National Weather Service and several radar wind profilers located in Texas.

The REMSAD dispersion mechanism is evaluated against the tracer data in chapter 5. Results of the REMSAD modeling and source attribution estimates are presented and discussed in chapter 6.

2.2.2 ATAD

The Air Resource Laboratory Atmospheric Transport and Disperson (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 m 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 2K above that at the inversion base. A critical inversion is defined as an inversion with a potential temperature lapse rate of at least 5K/km. When no critical inversion exists, the transport layer top is assumed to be 3000 m above the ground. For trajectories that begin at night, the initial transport layer depth is approximated by 2 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 m above the terrain near each radiosonde station. A back trajectory is started from the receptor every 6 hours. An air parcel position, or "endpoint", is determined for every 3 hours backward in time for a maximum of 120 hours (5 days).

2.2.3 HYSPLIT

The HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model [*Draxler and Hess*, 1998] was developed by the NOAA Air Resources Laboratory. It can compute a range of outputs from simple air parcel trajectories (advection of a single particle) to dispersion and deposition simulations. For BRAVO, the model was used in simple back trajectory mode. Version 4.7 is now available, but HYSPLIT modeling for BRAVO was done initially using version 4.4 and later re-done using version 4.5. The described changes (see the READY HYSPLIT web site at http://www.arl.noaa.gov/ready/hysplit4.html) between versions 4.5 and 4.7 are not expected to result in major changes for the BRAVO output. Trajectories calculated from versions 4.4 and 4.5 were identical for most BRAVO time periods, though there were a few days with differences ranging from subtle to dramatic.

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 omega field. Other options include isosigma, isobaric, isentropic, and isopycnic. The model can be run with multiple nested input data grids, though this was not done for BRAVO. 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 backwards 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 [*Kreyszig*, 1968; *Pettersen*, 1940] 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].

2.2.4 CAPITA Monte Carlo Model

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,* 1997]. In this study, only the atmospheric transport module was used.

The simulation of the regional scale transport and diffusion in the Monte Carlo model is conducted by moving inertialess particles in the Eulerian frame according to the equation

$$\mathbf{x}(t+dt) = \mathbf{x}(t) + [\bar{\mathbf{u}}(\mathbf{x},t) + \mathbf{u}'(\mathbf{x},t)]dt \quad (2\text{-}1)$$

where **x** (=x,y,z) represents the particle position vector, and $\bar{\mathbf{u}}$ ***(=***$\bar{u},\bar{v},\bar{w}$***)*** and **u**' (=u',v',w'), represent the time-averaged and fluctuating components of the flow field, respectively. The basic meteorological information required for this consists of the mean wind field $\bar{u},\bar{v},\bar{w}$ to advect the particles and the fluctuating components u',v',w' to diffuse the particles at all times over the whole model domain.

2.2.4.1 Advection

The advection of the particles is accomplished by multiplying the mean three dimensional wind vectors at the particle's location in space and time by the time step Δt. The mean wind vector is obtained from meteorological models, such as those described in section 2.1.3. 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. This is a large time step so the acceleration of the particle due to curvatures and changes in wind speed needs to be taken into account. This is done by using the iterative Petterssen integration scheme [*Petterssen*, 1940; *Stohl*, 1998], which is equivalent to assuming constant acceleration within the time interval to account for changes in the wind vector from t_n to t_{n+1}. Perfect reflection is assumed at the surface but if a particle reaches the grid edge or top of the model domain, it is terminated.

2.2.4.2 Vertical Diffusion

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 is primarily dependent on the convective and mechanical mixing process in the atmosphere and horizontal diffusion primarily depends on the divergence of the horizontal wind fields.

In the case of vertical diffusion, during the day the intense convective mixing within the mixed layer causes the turnover time of a large eddy to be about 15 minutes. Therefore, within this time period, the eddy's position is independent of its previous location. 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 due to vertical motion from subsiding and rising airmasses. A graphical illustration of the implementation of the above processes is presented by the three trajectories in Figure 2-13. The intense vertical mixing within the mixing layer is indicated by the crossing of the trajectory lines. The stable layers, during nocturnal hours, are illustrated in the lines of near constant height for each quantum. In Texas, the summertime mixing height varies diurnally from 10s of meters at night to 1–2 km during a sunny afternoon. In the nocturnal mixing layer, the assumption of a well-mixed layer is an over-simplification. However, these layers are low, generally less than 200 m, and the error of the particles' vertical position in relation to the meteorological models vertical resolution is not significant. This simple vertical diffusion mechanism has been incorporated into several dispersion models [*Patterson et al.*, 1981; *Anfossi et al.*, 1995; *Saltbones et al.*, 1998].

This vertical diffusion mechanism is rather simple, but it does capture the essential features of the interaction between mixing layer and free troposphere. That is, within the mixing layer, the intense daytime mixing evenly distributes material from the surface to the top of the mixed layer, followed by the collapse of the mixed layer in the evening, allowing wind veer and sheer in the lower layers of the atmosphere to spread the pollutants. It has been shown that this simple dispersion mechanism produces similar regional scale transport estimates as the more complex vertical dispersion mechanism for simulating a passive tracer in the atmosphere [*Desiato et al.,* 1998].

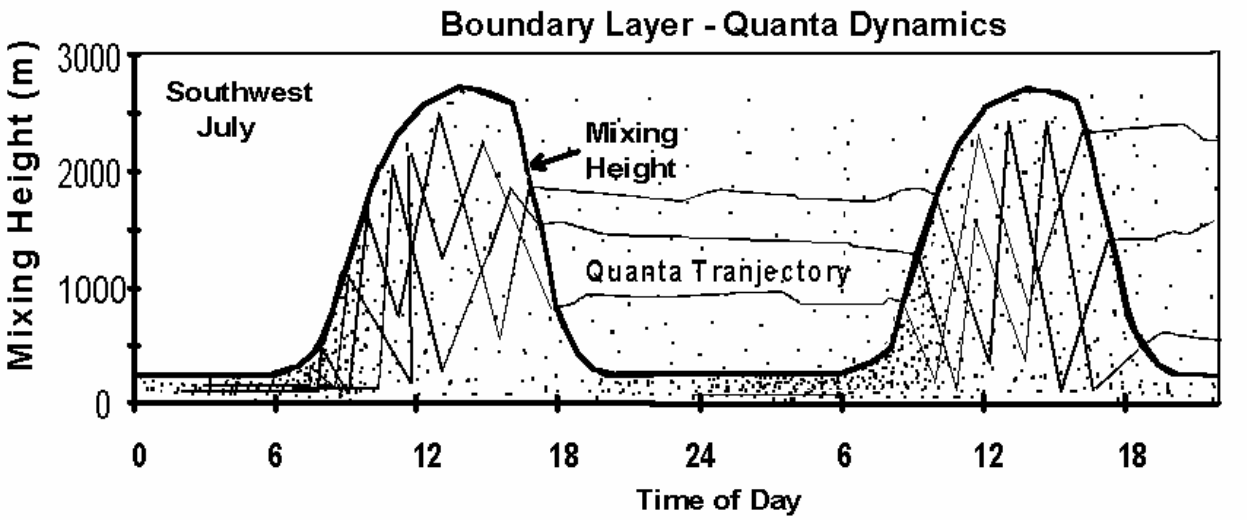

Figure 2-13. The dynamics of the planetary boundary layer and of pollutants being emitting into this layer as simulated by the CAPITA Monte Carlo model.

2.2.4.3 Horizontal Diffusion

A common approach for the calculation of the 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. Values of K ranging between 10^2 m^2/s in stable conditions and 10^6 m^2/s during intense convective activity have been used in the past [*Munn and Bolin* 1971; *Patterson et al*., 1981; *Gifford*, 1982]. Also, *Desiato et al.* [1998] tuned a dispersion model to tracer data and found a best fit constant K of $2.5\ 10^4$ m^2/s, while *Ishikawa* [1995] found a best fit K of $3.3*10^4 – 1.0*10^4$ m^2/s in dispersion simulations of Chernobyl dispersion data.

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 [*Schichtel and Husar*, 1996; *Uliaz*, 1994].

2.2.4.4 Meteorological Data Processing

The EDAS, FNL, and MM5 36 km wind fields were used to drive the CAPITA Monte Carlo model. The model requires these input data to have terrain following vertical coordinate

systems and an estimate of the mixing height. The MM5 meteorological was on a terrain following coordinate system and had an estimate of the mixing height, so no transformations were necessary. However, the vertical velocity was relative to a flat earth and this was transformed to be relative to the vertical coordinate system.

The EDAS and FNL data were on a constant pressure coordinate system. These data were reprocessed, transforming the constant pressure levels to 20 terrain following Cartesian coordinates from 10 to 10000 m. During this process, the 10 m surface level winds and the 2 m temperature and RH surface variables were incorporated into the first model upper layer. In addition, the vertical velocity was converted from mbar/sec to m/sec.

Neither meteorological data set had an estimate of the mixing height, so this was derived from available metrological parameters using a method based upon a bulk Richardson number approach [*Holtslag et al.*, 1995; *Vogelezang and Holtslag,* 1996]. The mixing height was defined as the height where the bulk Richardson number reaches a critical value of $Ri_c = 0.25$. To account for turbulence due to surface friction in a near neutral or stable mixing layer, the bulk Richardson number was modified to

$$Ri_{bk} = \frac{g\left(\Theta_{vk} - \Theta_{v1}\right)\left(z_k - z_1\right)}{\Theta_{v1}\left[\left(u_k - u_1\right)^2 + \left(v_k - v_1\right)^2 + 100u_*^2\right]} \tag{2-2}$$

where g is gravity, z_1, z_k are the heights of the lowest model level and of the model level k, u and v are the horizontal wind field components, Θ_v is the virtual potential temperature, and u_* is the friction velocity.

The term $100u_*^2$ accounts for turbulence due to friction [*Vogelezang and Holtslag*, 1996]. This equation holds for only neutral and stable conditions. However, it can be extended to convective situations by replacing Θ_{v1} with

$$\Theta_{v1}^{'} = \Theta_{v1} + 8.5\frac{H_{vs}}{w_*} \tag{2-3}$$

where w_* is the convective velocity scale and H_{vs} is the virtual surface heat flux, defined as $H_{vs} = \left(w^{'}\Theta_v^{'}\right)_s$. The second term on the right represents the excess temperature of rising thermals, which is a measure of the strength of the convective mixing [*Vogelezang and Holtslag*, 1996].

This parameterization of the mixing layer depth has been validated against measured data [*Holtslag et al.*, 1995; *Vogelezang and Holtslag,* 1996; *Sorensen,* 1998] where the optimum critical values were found to be between 0.15–0.35. The robustness of this technique has led it to be widely used in dispersion models [*Brandt et al.*, 1998; *Saltbones et al.*, 1998; *Sorensen*, 1998; *Stohl*, 1998].

2.3 Source Apportionment Methods

This section describes the source apportionment techniques applied during BRAVO to assess the contribution of U.S. and Mexican sources on the sulfate concentrations at Big Bend National Park. The methods discussed have been divided into qualitative and quantitative source attribution methods, where the qualitative analyses include trajectory and air quality data

analysis techniques. The quantitative techniques include receptor and source oriented modeling as well as hybrid techniques.

2.3.1 Qualitative

2.3.1.1 Space/Time Pattern Analysis (EOF)

Empirical Orthogonal Functions (EOF) analysis is essentially a spatial factor analysis or principal component analysis. Details can be found in many standard texts. [e.g., *Lawson and Hanson*, 1974, *Green* 1976]. There are several variations of these types of analyses, but all involve decomposing a single matrix into two orthogonal matrices. Here, a matrix of centered concentrations (site means subtracted from each concentration) of a single species is decomposed into two matrices as follows:

$$Z = A \bullet P \tag{2-4}$$

where Z has rows corresponding to observation times and columns to monitoring sites. A is a time by factor orthonormal matrix of dimensionless time weights and P is a factor by site orthogonal matrix of spatial patterns. The values of P are weighted deviations from the mean concentration at each site and have units of concentration. Each row of P can be visualized as a map. The P matrix is calculated by multiplying the eigenvectors of Z^TZ by a matrix which contains the square roots of the eigenvalues of Z^TZ on the diagonal. The A matrix can be found by inverting P. Each column of *A* is visualized as a timeline that shows the relative weight of the factor for each observation time. Optionally, rather than Z^TZ, there are other matrices that could similarly be decomposed. These include the raw data cross product matrix, the correlation matrix, and the covariance matrix. An advantage of using Z^TZ is that the results are easily physically interpretable.

The resulting factors or EOFs are orthogonal, statistically simple, and each explains a decreasing amount of the total variance. Typically, only a few spatial patterns, often only 2-6, are required to explain 80–90% of the covariance in the data. Other physical information is then used to qualitatively interpret the meaning of each of these factors. This is an exploratory data analysis technique. Usually an attempt is made to associate each spatial pattern with meteorological conditions and candidate source areas. For example, strong spatial gradients in one portion of the monitoring domain are indicative of a source in that area or transport from that direction [*Henry et al*., 1991]. Spatial patterns for different chemical species are often quite different and can be used to help determine regionally different trace element ratios. The time weights show when each pattern dominated and can then be used to help evaluate the reasonableness of wind fields and back trajectories. Occasionally an EOF is generated that contains a high value at a single site with a high time factor for only a single day. These are often useful for quickly spotting outliers in the data. Outliers can be due to real events such as fires or due to measurement problems.

EOFs as calculated above are not a unique solution to the decomposition of the data. There are many methods of orthogonal and oblique rotation, of which Varimax [*Kaiser*, 1958; *Richman*, 1986] is probably the best known. Varimax rotation forces each site to load as strongly as possible onto a single factor and has frequently been used in the past in similar field studies. [*Gebhart and Malm*, 1997; *Malm and Gebhart,* 1996, 1997; *Gebhart et al.*, 2000]. However, for the BRAVO study neither Varimax nor Quartimax [*Richman*, 1986] rotations added any information or interpretability to the analysis. The usual result is that after rotation the spatial patterns are essentially the same as before rotation, though the order, time weights, and variance

explained by each were shuffled a bit. Therefore, for simplicity, only the unrotated EOFs are reported here.

EOF analysis requires a time by site matrix of concentrations of a single aerosol species. No missing values are allowed and concentrations below minimum detectable limit are set to 0. All 6-hour data were averaged to 24-hour concentrations and then missing value were eliminated or estimated. First, time periods and then sites with more than 25% of data missing were eliminated. This reduced the analysis period of June 29–October 31 to July 26–October 30, with August 5 and 6 also being removed. Three sites, Padre Island, Guadalupe Mountains, and Wichita Mountains, were also removed, resulting in a matrix of 95 days by 34 sites. Next, all observations immediately preceded and followed by valid data at the same site were approximated by linear interpolation in time. Remaining missing data were filled by inverse distance-weighted spatial interpolation. For the sulfur matrix, only 0.8% of the data were filled by time interpolation and 2.8% by spatial interpolation. Similar fractions of the data matrices for other species were estimated in the same way. Due to the removal of most of July, results of EOF analyses are not expected to be directly comparable to other analyses of BRAVO data that include that time period. EOF analyses were completed for all measured species with adequate data. Results are discussed in section 8.1.1.

2.3.1.2 Factor Analysis

In general, factor analysis is used to attempt to explain the correlations between observed variables in terms of underlying factors which are not directly observable. For BRAVO, in addition to the spatial factor analysis or EOF discussed above, another factor analysis was used to explore the correlations between different measured fine particulate species at a single receptor site (K-Bar). The resulting factors, or groups of species that are correlated, were interpreted as different source types. For example, a factor for which organic carbon, fine mass, and potassium all loaded highly together would be qualitatively interpreted as a smoke factor. Factor analysis, like EOF and principal component analyses, is a method of reducing a large number of observed variables into a few linear combinations of those variables that can explain the data as well or nearly as well as the larger number of individual variables. Furthermore, we would like each factor to have some physical meaning. In this case, each factor represents a source type and each daily factor score represents the relative contribution of that factor type to the measured fine particulate concentrations on that day. Factor analysis is calculated identically to EOF analysis as shown in equation 2-4, except that the initial concentration matrix, Z, is a species by time matrix at one site, rather than a site by time matrix of one species at multiple sites. In this case, the two resulting decomposed matrices, A and P, are a species by factor matrix of loadings, representing the compositions of the derived source types, and a factor by time matrix of factor scores, representing the daily contributions of each source type. Details and results of the factor analysis are discussed in section 8.1.2.

2.3.1.3 Ensemble Airmass History Analyses

An airmass history identifies the 2-D or 3-D pathway an airmass 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 power techniques that potentially link source regions to the receptor concentrations. However, individual airmass histories can have large

errors [*Stohl*, 1998; *Rolph and Draxler*, 1990; *Kahl and Samson*, 1986]. In fact, different model assumptions or wind fields can generate airmass trajectories that are 180 degreesout of phase. The large uncertainty in individual trajectories can be reduced through ensemble airmass history techniques which aggregate large numbers of airmass histories together. These ensemble airmass techniques then identify the typical airmass transport pathways.

This section describes the residence time analysis, one class of ensemble airmass history analysis techniques. These methods are used in chapter 5 to help evaluate the various airmass history model and meteorological drives, and chapter 6 to evaluate the airmass transport to Big Bend National Park during the BRAVO study.

2.3.1.3.1 Residence Time

The residence time analysis [*Ashbaugh*, 1983; *Poirot & Wishinski*, 1986] aggregates airmass histories together by first overlaying a grid over a spatial domain, then counting the number of hours each airmass resided within each grid cell. Usually, the number of hours an airmass resided in a grid cell is approximated by the number of trajectory segment endpoints that fall in the grid cell. These grid cell residence times are then normalized by the total number of hours resulting in a probability field identifying the most likely regions in which an airmass will reside en route to the receptor. The most likely transport directions to a receptor are then along the ridges of the probability fields.

The overall residence time (ORT) for each grid cell is then defined as

$$ORT_{i,j} = \frac{1}{N}\sum_{t=1}^{T} n_{i,j,t} \tag{2-5}$$

where n is the number of back trajectory segment endpoints in the grid cell at longitude i and latitude j before the trajectory arrived at the receptor during measurement period t. N is the total number of endpoints for all time periods and T is the total number of time periods. A trajectory segment endpoint is the position of the airmass at a specific time prior to impacting the receptor. Airmass histories are typically calculated every hour or several times per day, depending on the model. If a grid is used which does not have a uniform area, then the residence time probability field is normalized by each grid cell's area, creating a residence time probability density function (PDF).

High (HRT) and low (LRT) concentration residence times are similar, except only trajectories that arrived at the receptor when the concentration was greater than or less than, respectively, a selected value are considered. The cutoffs for high and low concentrations are arbitrary, but are often the highest and lowest 10–20% of the measurements. HRT and LRT are calculated by

$$HRT_{i,j} = \frac{1}{H}\sum_{t=1}^{T} h_{i,j,t} \tag{2-6}$$

$$LRT_{i,j} = \frac{1}{L}\sum_{t=1}^{T} l_{i,j,t} \tag{2-7}$$

where H and L are the total number of trajectory endpoints arriving during high and low concentration time periods and h and l are the number of trajectory segment endpoints in each grid cell for these times.

If a grid with unequal grid cell areas is used, then the residence time analysis can be biased. For example, as grid cell increases in size, more trajectories are likely to traverse it and its residence time will increase. Normalizing the residence time probabilities by the grid cell removes this potential bias. The units of the residence time analysis are then inverse area. This creates a residence time PDF since if the area-normalized residence time analysis is integrated over all space it will equal one.

2.3.1.3.1.1 Decomposition of the Residence Time Analysis

The residence time probability is the result of the frequency of airmass transport over a grid cell and the length of time the airmasses spent over the grid cell. The residence time PDF was decomposed into two components, a transport directional frequency, the fraction of trajectories that passed over a grid cell, and the accumulation potential, the average time an airmass resided over the grid cells normalized by the grid cell length.

$$RT_{i,j}\,PDF = \frac{D_{i,j} * AP_{i,j}}{\tau} \tag{2-8}$$

$$D_{i,j} = \frac{1}{L} * \frac{1}{Tr} \sum_{t=1}^{T} tr_{i,j,t} \tag{2-9}$$

$$AP_{i,j} = \frac{1}{L} * \frac{1}{Tr} \sum_{t=1}^{T} \left(n_{i,j,t} * \Delta\tau \right) \tag{2-10}$$

where $D_{i,j}$ [1/m] is the normalized transport direction frequency for grid cell i,j, $AP_{i,j}$ [s/m] is the accumulation potential for grid cell i.j, τ [hr] is trajectory length, $\Delta\tau$ [hr] is the trajectory segment endpoint length, $tr_{i,j,t}$ are the number of trajectories in the airmass during time period t that traverses grid cell i,j, Tr = total number of trajectories for all time periods T, and L [m]= length scale; note L^2 is the area of the grid cell

The accumulation potential can be thought of as an inverse of a characteristic transport speed and is related to the inverse of the speed the airmasses are transported over a grid cell and the number of times it traverses the grid cell due to recirculation or flow reversals. This is a better index for determining the exposure of an airmass to a source region's emission than the measured wind speed. For example, fast recirculating winds would spend more time over a source region than slower but directionally persistent winds. Therefore the recirculating airmass would be exposed to a region's emissions for a longer period of time, increasing the potential for pollutants to accumulate.

The residence time analysis used three-dimensional airmass histories. To characterize the heights of the airmass histories, the height of all trajectory segment endpoints within a grid cell were average together. While the average trajectory segment endpoint height gives some indication of an airmass characteristic height, it is independent of the residence time analysis. In the future, it should be possible to create a three-dimensional residence time PDF taking into account the trajectory segment endpoint heights. In this case, the residence time decomposition

would contain a component accounting for the vertical distribution of the trajectory segment endpoints.

2.3.1.3.1.2 Source Contribution Function (Distance Weighted Residence Time)

The overall source contribution function (OSC) is the overall residence time (ORT) normalized by an equal probability surface (EPS). The EPS is defined as being an idealized ORT which would exist if all airmasses arrived at the receptor following a straight trajectory with constant speed with equal probability from all directions. Residence time fields always have a peak at the receptor because all back trajectories originate there. The source contribution function is the residence time with this central tendency removed. It is always proportional to the residence time multiplied by the distance of the grid cell from the receptor. For relative comparisons between grid cells, the proportionality constant is irrelevant, and in many applications it has been set to $1/r_{max}$, where r_{max} is the maximum distance from the receptor to any trajectory. However, the proportionality constant can be derived with a smaller radius if desired.

OSC is defined as

$$OSC_{i,j} = \frac{ORT_{i,j}}{EPS_{i,j}} \tag{2-11}$$

where ORT is the overall residence time and EPS is calculated by

$$EPS_{i,j} = \frac{1}{r_{i,j}} \frac{A_{i,j}}{2\pi R_{norm}} \tag{2-12}$$

where $r_{i,j}$ is the distance from the receptor to the center of the grid cell. The inverse distance factor results from the fact that all trajectories originate at the receptor. Thus a grid cell near the receptor is more likely to have an endpoint in it than a grid cell far away. $A_{i,j}$ is the area of the grid cell. In a grid system based on degrees of latitude and longitude, each grid cell does not have the same area, with cells to the north being smaller than those to the south. Grid cell area in square km for a 1 degree by 1 degree grid cell is calculated by

$$\begin{aligned} A_{i,j} &= \Delta X \ \Delta Y \\ &= (1\,^{\circ}Lat)^2 (111.1\ Km/^{\circ}Lat)^2 \, Cos(Lat). \end{aligned} \tag{2-13}$$

R_{norm} is the "normalization radius" or radius of the area in which the EPS integrates to 1, so $(2\,\pi\,R_{norm})$ is the total area in which the EPS is defined and is somewhat arbitrary. Because the EPS is zero beyond R_{norm}, OSC is undefined when $r_{i,j}$ is greater than R_{norm}. As an example, R_{norm} here is calculated such that the ratio of R_{norm} to the total number of endpoints is equal to the ratio of radius to endpoints in the central 9 grid cells. Thus R_{norm} is calculated by solving

$$\frac{R_{norm}}{N} = \frac{R_{cent}}{N_{cent}}$$

where (2-14)

$$N_{cent} = \sum_{i=iorg-1}^{iorg+1} \sum_{j=jorg-1}^{jorg+1} n_{i,j}$$

and iorg and jorg are the i and j coordinates of the grid cell containing the receptor (the origin). R_{cent} is calculated by solving for it in the following relationship.

$$\pi R_{cent}^2 = \sum_{i=iorg-1}^{iorg+1} \sum_{j=jorg-1}^{jorg+1} A_{i,j} \quad (2\text{-}15)$$

The functional equations for the overall high concentration and low concentration source contribution functions are then

$$OSC_{i,j} = \frac{2\pi R_{norm} r_{i,j}}{N A_{i,j}} \sum_{t=1}^{T} n_{i,j,t}$$

$$HSC_{i,j} = \frac{2\pi R_{norm} r_{i,j}}{H A_{i,j}} \sum_{t=1}^{T} h_{i,j,t} \quad (2\text{-}16)$$

$$LSC_{i,j} = \frac{2\pi R_{norm} r_{i,j}}{L A_{i,j}} \sum_{t=1}^{T} l_{i,j,t}$$

H and L are similar to N except they are the total number of back trajectory endpoints associated with high and low concentrations, respectively.

The source contribution function provides a formal means of limiting the extent that is considered in the analysis of the distance an airmass will travel. However, if this cut-off is ignored then the source contribution function is proportional to multiplying the residence time probability by the distance between the receptor and grid cell. In practice, equation 2-16 is reduced to the residence time multiplied by the distance to the grid cell

$$\begin{aligned} OSC_{i,j} &= r_{i,j} ORT_{i,j} \\ HSC_{i,j} &= r_{i,j} HRT_{i,j} \\ LSC_{i,j} &= r_{i,j} LRT_{i,j} \end{aligned} \quad (2\text{-}17)$$

Using a distance-weighted residence time probability to remove the central tendency in the residence time PDF was first put forth by *Poirot and Wishinski* [1986].

2.3.1.3.1.3 Conditional Probability

Conditional probability is the probability that if an airmass passed through a grid cell, it arrived at the receptor when the concentration satisfied a given condition. The selected condition is usually the measurement of a high or low concentration. The high concentration and low concentration conditional probabilities are calculated by

$$HCP_{i,j} = \frac{HRT_{i,j}}{ORT_{i,j}} * \frac{H}{N} = \sum_{t=1}^{T} h_{i,j,t} \; / \; \sum_{t=1}^{T} n_{i,j,t}$$
$$LCP_{i,j} = \frac{LRT_{i,j}}{ORT_{i,j}} * \frac{L}{N} = \sum_{t=1}^{T} l_{i,j,t} \; / \; \sum_{t=1}^{T} n_{i,j,t} \qquad (2\text{-}18)$$

where all variables are as defined in the residence time equations.

Grid cells which have few total endpoints may not have statistically significant conditional probabilities and are usually not reported. An extreme example is a grid cell with only one endpoint. The only possible values for the conditional probability are then 0 or 1. Such grid cells are usually at the edges of the domain.

2.3.1.3.1.4 Incremental Probability

The incremental probability identifies regions that are more or less likely to be traversed when the receptor concentrations satisfied a given condition compared to an average day [*Poirot et al.*, 1999; *Poirot et al.*, 2001]. For example, the high and low incremental probabilities are the difference between the high or low residence time probability and the overall residence time probability:

$$HIP_{i,j} = HRT_{i,j} - ORT_{i,j}$$
$$LIP_{i,j} = LRT_{i,j} - ORT_{i,j} \qquad (2\text{-}19)$$

The high incremental probability (HIP) field differs from the high condition probability (HCP) field because the HIP metric is determined by subtraction (the extent to which the high day probability is greater than the average day), while the HCP metric is determined by division (the fraction of total trajectory segment endpoints passing over a cell that results in high concentration days). Thus, the HCP indicates the potential for a location to contribute if that area is upwind of the receptor, while the HIP reflects the most probable upwind locations if the receptor concentration is high.

2.3.1.4 Trajectory Max

The trajectory max technique is a first order source apportionment technique meant to give an extreme upper bound on the possible contribution of a source region to a receptor's air quality. This is an airmass history technique where it is assumed that if an airmass traversed a source region, then the receptor concentration is solely due to emissions from that source region. Therefore the receptor site would have a 100% potential contribution, while a distant source region that an airmass never traversed would have a 0% contribution.

$$SC_{i,j} = \frac{\sum_{t=1}^{T} C_t \delta_{i,j,t}}{\sum_{t=1}^{T} C_t} \times 100 \qquad (2\text{-}20)$$

$$tr_{i,j,t} > 0, \delta_{i,j,t} = 1$$
$$tr_{i,j,t} = 0, \delta_{i,j,t} = 0$$

where $SC_{i,j}$ is the maximum percent source contribution of source region i,j to the receptor, C_t is the receptor concentration at time period t, and $tr_{i,j,t}$ are the number of trajectories in the airmass during time period t that traverses source region cell i,j.

The residence time analyses are used to evaluate the difference wind fields and airmass history models in chapter 4. Results of the trajectory residence time and related analyses are discussed in section 8.1.3, while results from the trajectory max analysis is presented in section 8.1.4.

2.3.2 Quantitative

2.3.2.1 Trajectory Mass Balance (TrMB)

The Trajectory Mass Balance (TrMB) model [*Pitchford and Pitchford*, 1985; *Iyer et al.*, 1987; *Gebhart et al*., 1988; *Malm et al*., 1989; *Gebhart and Malm*, 1989, 1994; *Malm*, 1992; *Gebhart et al*., 1993] is a special case of the general mass balance equation [*Iyer et al*., 1987; *Malm et al*., 1989) in which measured concentrations at a receptor are assumed to be linearly related to the frequency of airmass transport from a source area to the receptor by the following relationship:

$$C_{it} = \sum_{j=1}^{J} Q_{ijt} T_{ijt} N_{jt} E_{ijt} \qquad (2\text{-}21)$$

The subscripts i,j, and t refer to chemical species, source area, and time, respectively. C is the measured concentration, Q is the emission rate, T is a transformation and deposition factor to account for deposition, diffusion, and chemical conversion, N is the number of back trajectory endpoints and E is an entrainment factor to account for disassociation between the back trajectories and the transport of pollutants. An endpoint is defined as the calculated position of an air parcel that eventually will arrive at the receptor. Endpoints are calculated hourly for 5 days back in time. In this application the only variables used explicitly are C, the concentrations measured at Big Bend and N, the number of back trajectory endpoints in each source area for each day. The remaining term, QTE, called the "transfer coefficients" are estimated by Ordinary Least Squares (OLS) regression for each source such that equation 2-21 simplifies to equation 2-22 where

$$C_t = \bar{a}_0 + \sum_{j=1}^{J} \bar{a}_j N_{jt} + error_t \qquad (2\text{-}22)$$

The subscript i has been dropped for simplicity since we are now dealing with a single measured species. The species concentrations at the receptor per endpoint in the source region are a=QTE and are the regression coefficients with units of concentration per endpoint. These are estimates of the average relationship between airmass residence time in the source area and

measured concentration at the receptor. The intercept was fixed at 0, forcing all identified source areas to account for the measured concentrations. Although emissions and precipitation data are available for BRAVO, they have not been included in the modeling. Note that whenever a quantity is replaced by its average, the error increases.

The error, as shown in equation 2-23, is due to the deviations of the unknown terms from the mean plus measurement error:

$$error_t = \sum_{j=1}^{J} (a_{jt} - \overline{a_j}) N_{jt} + measurement\ error \qquad (2\text{-}23)$$

Source areas are chosen based on several criteria. First, areas are chosen based on interest in the attributions from the area, e.g., separating the influence of sources in Mexico from sources in Texas. Second, sources 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 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 en route to the receptor. 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.

Collinearities between source areas are investigated using the Variance Inflation Factor (VIF) [*Belsley et al.*, 1980] where

$$VIF_i = \frac{1}{1 - R_i^2} \qquad (2\text{-}24)$$

R^2 is the multiple correlation coefficient of the endpoints in one source area regressed on the endpoints in the remaining source areas. A high value indicates that the endpoints in a single source region could be nearly explained by a linear combination of the endpoints in the remaining source regions. A VIF greater than 10 is considered strong collinearity. Another indication of collinearities between source areas is the correlation matrix of the endpoints, although correlations can only reveal collinearities between pairs of source areas, e.g., source 1 with source 2, and not cases where two or more source areas could be linearly combined to predict another. A further indication of collinearities is large standard errors for the regression coefficients.

A test of the TrMB method using tracer data is discussed in section 7.1.1, a test of TrMB using simulated sulfate is in section 7.2.1, and TrMB results for measured sulfate at Big Bend National Park are shown in section 8.2.2.

2.3.2.2 Forward Mass Balance Regression (FMBR)

The TrMB technique presented in the previous section is based upon a special case of the general mass balance equation for back trajectories. A similar relationship can be derived using the forward Lagrangian framework (forward airmass histories) such that for a single ambient species [*Lamb et al.*, 1971; *Iyer et al.*, 1987; *Schichtel*, 1996].

$$c_{ik} = \frac{1}{\Delta X_i} \sum_j \sum_\tau T_{i,k|j,k-\tau} K_{i,k|j,k-\tau} E_{j,k-\tau} \quad (2\text{-}25)$$

where c_{ik} is concentration at receptor j and time k, [g/m^3], $E_{j,k-\tau}$ is mass emitted from source j at time k-τ, [g], $T_{i,k|j,k-\tau}$ is transit probability, [ ], $K_{i,k|j,k-\tau}$ is kinetic probability [ ], ΔX_i is the averaging volume at the receptor I, $k-\tau$ is source release time, and τ is age of the source mass at the receptor.

The transit probability describes the dispersion of pollutants emitted into an airmass over time. Specifically, the transit probability is the likelihood of airmass at source j and time k-τ being transported to a receptor i at time k where τ is the time it takes the airmass to be transported from the source j to receptor i, or the age of the airmass. The kinetic probability accounts for the transformation and removal of pollutants during the transport from the source i to the receptor j during the time τ. The kinetic probability has a different interpretation for primary and secondary species. For primary species it is the probability that the mass is not removed during transport from the source to the receptor. For secondary species it is the probability that the primary emissions were transformed to the secondary species and not removed during transport from the source to the receptor.

This equation relates the receptor concentrations c_{ik} to the emissions $E_{j,k-\tau}$ via the transfer coefficient $Tr_{i,k|j,k-\tau} = T_{i,k|j,k-\tau} K_{i,k|j,k-\tau}$ and is also known as the source receptor relationship (SRR). The transfer coefficient is equivalent to the relative contribution of a source to a receptor.

If Tr and E are known, then equation 2-25 can be used to estimate the concentrations C_{ik}. However, if C_{ik} are known and one or more of the factors on the right hand side of equation 2-25 are unknown, then the relationship can be inverted to find these unknown factors. In this application, the concentrations are known and the transit probabilities are estimated from the CAPITA Monte Carlo model. Therefore the kinetic probabilities and emission are unknown. These two terms can be combined into a single source/sink term $(KE)_{i,k|j,k-\tau} = K_{i,k|j,k-\tau} E_{j,k-\tau}$, and equation 2-25 is then inverted to solve for this source/sink term.

The number of elements in the source/sink term is equal to the number observations multiplied by the number of source/time pairs, while the vector of concentration values only has the number of observations. Therefore this is a highly under-determined system and the number of unknowns must be reduced. In this application, this is accomplished by integrating the SRR over the source release time and averaging the source/sink term over the receptor times and sites. Therefore the SRR became

$$c_{ik} = \sum_j T_{i,k|j} (KE)_j + error_{ik} \quad (2\text{-}26)$$

where the term $(KE)_j$ is the source/sink term for each source j and the error term accounts for measurement and averaging errors in the source/sink term. The number of unknown terms is then equal to the number of sources j which can be made smaller than the number of observation giving us an over determined system. The source/sink term is then obtained by inverting equation 2-26, i.e., $(KE) = (T)^{-1} c$. The contribution of individual sources to the receptor concentration can then be estimated from $c_{ikj} = T_{i,k|j} (KE)_j$.

Aggregating the source/sink term over all receptors and times is equivalent to assuming that each source's emission rate and kinetic probability are constant over time and receptors. This may be an adequate assumption for emission rates, but the transformation and removal process driving K will vary dramatically in space and time. For example, K would be about 0 for a heavy rain event, while under dry conditions K could be near 1, i.e. no removed mass. Therefore caution needs to be applied when grouping receptor sites to estimate (KE) and the results are only applicable to the average across all receptors and times used in the analysis.

Equation 2-26 is inverted using linear least squares techniques. However, the source receptor relationship is an ill-conditioned system [*Newsam and Enting*, 1988; *Schichtel*, 1996] causing its inversion by least squares to amplify the errors in the system. This results in an unstable solution with large positive and negative values. To dampen these instabilities information needs to be added to the system. This was done using two approaches. The first approach used singular value decomposition (SVD) [*Jackson* 1972; *Press et al*., 1988; *Schichtel*, 1996]. SVD is equivalent to zeroth-order linear regularization which finds a solution by balancing the model agreement (minimization of the square of the residuals) and the magnitude of the solution vector. In this case, the added information is the magnitude of the solution vector, because as magnitude of the solution vector decreases the solution stability increases. A key to the SVD solution is selecting the appropriate balance or trade off between the model agreement and the solution stability. This was done by selecting the solution that equally maximized the model agreement and solution stability. If the solution was still unstable as evidenced by large positive and negative values, then the weight to the solution stability was increased.

In the zeroth order solution, the more weight given to the solution's stability, the lower the effective resolution of the solution space. This is similar to *a priori* grouping the source regions prior to the inversion. If the source regions are sufficiently grouped together, it is possible to use ordinary least squares techniques to obtain stable solutions. Therefore the second means of adding information to the system was by aggregating the source regions together. The aggregated source/sink regions are defined and discussed in section 2.3.2.6.

The basic idea of the Forward Mass Balance Regression is to determine the average source contribution from pre-defined source regions and measured receptor data, in this case particulate sulfate. FMBR does this by essentially comparing the receptor concentration levels and to the relative inert contribution, i.e. the transit probability, from all modeled source regions impacting the receptor. A source region that repeatedly has a large inert contribution when the concentrations are low will have a low source/sink term. That is, it has low emissions and/or the emitted mass is usually removed before impacting the receptor, or in the case of a secondary species such as sulfate, the primary species was not transformed to the secondary species. On the other hand if a source has a large inert contribution and the concentrations are repeatedly high, then it will have a high source/sink term. Due to the fact that multiple sources impact the receptor at any given time, a regression technique is used to decompose the receptor mass to the source regions, thus deriving multiple average source/sink terms.

The decomposition of the receptor mass among source regions relies on airmasses traversing each source region but following different transport direction to the receptor at different times. Therefore a different number and source regions will be impacting the receptor at different times. Any two source regions with highly correlated relative inert contributions cannot be reliably separated from each other. Therefore one of the key aspects of this technique

is to properly group the source regions so that they have varied source/sink terms and non collinear transit probabilities. The selection of source regions is presented in section 2.3.2.5

The FMBR technique is validated against tracer data in section 7.1.2. It is validated against simulated sulfate concentrations in section 7.2.2, and FMBR results for measured sulfate at Big Bend National are presented in section 8.2.3.

2.3.2.3 Synthesis Inversion – Merging Air Quality Model Source Apportionment Results and Receptor Data

The previous sections defined the source receptor relationship from the backward and forward Lagrangian framework and used these formulations to estimate source contributions to receptor sites by merging measured concentrations and airmass transport calculation. A similar technique can also be applied to Eulerian models. This section outlines a general inverse modeling technique based on the Eulerian model framework. Then it is shown how this can be used in conjunction with the REMSAD and CMAQ-MADRID source attribution results to develop an alternative source attribution of Big Bend's sulfate. CMAQ-MADRID is also an Eulerian grid model similarly employed as REMSAD by the Atmospheric and Environmental Research (AER) and EPRI work groups [*Electric Power Research Institute*, 2004].

2.3.2.3.1 Synthesis Inversion of the Source Receptor Relationship

The conservation of mass equation governing the concentration of trace species in the atmosphere can be written as

$$\frac{\partial}{\partial t} m(\ ,t) = q(\ ,t) + \mathrm{T}[m(\ ,t),t] + \mathrm{K}[m(\ ,t),t] \qquad (2\text{-}27)$$

where m(**r**,t) [g/m^3] is the concentration of the trace species at location **r** and time t, q(**r**,t) [g/m^3/s] is the local source emission rate, $\mathrm{T}[\mathrm{m}(\mathbf{r},\mathrm{t}),\mathrm{t}]$ [g/m^3/s] is a transport operator, and $\mathrm{K}[\mathrm{m}(\mathbf{r},\mathrm{t}),\mathrm{t}]$ [g/m^3/s] is a kinetic operator describing the physical/chemical removal and transformation processes.

Equation 2-27 expresses the rate of change of the concentration at a point **r** and time t as the sum of the local sources, the contribution due to trace-gas transport from other locations, plus net loss and gains due to removal, and chemical transformation processes at **r** and time t. The transport processes account for both advection and turbulent diffusion.

Enting [2000, 2002] showed that if the kinetic and transport operators are linearly dependent on the concentrations, then equation 2-27 is a linear differential equation that linearly relates the concentration field m(**r** ,t) to the sources q(**r** ,t). Therefore given a specified set of linear boundary conditions, equation 2-27 can be solved using the Green function method

$$m(\ ,t) = m_o(\ ,t) + \int_{r'}\int_{t_o}^{t} G(\ ,t,\ \ ,t')q(\ ,t')dt'd^3r' \qquad (2\text{-}28)$$

where m_o(**r**, t) [g/m^3] is the concentration that would result in the absence of emissions from the initial model simulation time t_o and G(**r**,t, **r**',t') [1/m^3] is the Green function. Mathematically, G is a mapping from the space-time distribution of the sources q(**r** ,t) onto the space-time distribution of the concentrations m(**r** ,t). Physically, G is the response of the concentration field due to a change in the source field and is equivalent to the continuous transfer coefficient in equation 2-25.

For inverse modeling purposes it is often useful to work with a discrete form of equation 2-28. This can be done by decomposing the source term into a set of processes and distributions such that

$$q(\ ,t)=\sum_j s_j\sigma_j(\ ,t) \tag{2-29}$$

where s_j is a set of unknown scale factors and $\sigma_j(\mathbf{r},t)$ is a set of known source distributions. Therefore, by integrating over the source space (**r'**) and time t'

$$m(\ ,t)=\sum_j s_j G_j(\ ,t) \tag{2-30}$$

where

$$G_j(\ ,t)=\int_{r'}\int_{t'} G(\ ,t,\ \ ,t')\sigma_j(\ \ ,t')dt'd^3r' \tag{2-31}$$

the concentrations m_o are assumed constant and incorporated into equation 2-29 as a pseudo-source. It is common to normalize the source distributions:

$$\int_{r'}\int_{t'}\sigma_j(\ \ ,t')dt'd^3r'=1 \tag{2-32}$$

Therefore the scale factors s_j have a direct interpretation as average source strengths. Finally, discretisation of equation 2-30 in the receptor space (**r**) and time t can be done by averaging over an appropriate area and time such that

$$c_i=\sum_j G_{ij}s_j+\varepsilon_i=m_i+\varepsilon_i \tag{2-33}$$

where c_i are the observed concentration values, m_i are the modeled concentration values, ε_i are the errors in c_i, s_j are the source strengths, and G_{ij} is the discretisation of G(**r**,t, **r**',t'), estimated through the integration of the air quality model via equation 2-31.

In the development of 2-33, the Green function method required a linear differential equation with respect to the concentrations m(**r**,t). However, from a practical stand point, the discretised G_{ij} is calculated by integrating the air quality model with the appropriate boundary conditions and source distributions $\sigma_j(\mathbf{r},t)$ and the model can contain nonlinearities. The development of source receptor relationship for inversion techniques in this way is known as synthesis inversion, since the discretisation of sources involved synthesizing the results as a linear combination of pre-specified source patterns [*Enting*, 2002].

2.3.2.3.1.1 Application to REMSAD and CMAQ-MADRID Source Attribution Results

The REMSAD model simulated the sulfate concentrations in most of North America and estimated the contribution of sulfate from ten major source regions and the model boundary conditions to Big Bend National Park and the other 36 BRAVO monitoring sites. The CMAQ-MADRID model was similarly employed to simulate the contribution from 4 major source regions and the boundary conditions. In both cases, $\sigma_j(\mathbf{r},t)$ are the actual source emission rates for each source region used in REMSAD and CMAQ-MADRID. The discrete Green function is

then equal to the contribution of the modeled source region q to the monitoring site p for the 24-hour average time period k. By letting the source/time pair q and k equal index j and the receptor/time pair p and k equal index i, the REMSAD or CMAQ-MDRIS source attribution results can be represented by equation 2-33, where G_{ij} is the daily contribution from each source region to each daily sulfate concentration at one or more BRAVO monitoring sites and s_j is the source attribution scaling coefficient. If $s_j = 1$, then $\Sigma_j G_{ij}$ is approximately equal to the predicted sulfate for the model observation $_i$ (m_i). It does not exactly equal the predicted concentrations due to non-linearities in the air quality models' chemistry modules, though the non-linearities were found to be small.

In this application, c_i are the measured sulfate concentrations and equation 2-33 is solved for s_j using Bayesian least squares regression. In the manner that equation 2-33 was developed, s_j represents biases in the source emission rates. However, s_j compensates for errors in the source attribution results which are due to errors in the emissions, transport, chemical transformation, and removal of sulfur. Therefore s_j represents the biases in the source attribution results and will be referred to as the source attribution scaling coefficient. If the model source attributions are unbiased, then the regression coefficients (s_j) will be equal to 1.

In section 7.2.3 the synthesis inversion approach is tested against synthetic sulfate data and source attribution estimates from the REMSAD model. In section 8.2.1 the results from the Synthesized REMSAD and CMAQ-MADRID are presented.

2.3.2.4 REMSAD Emission Sensitivity Simulations

Results from the base emissions simulation were used to assess REMSAD's skill in predicting sulfate concentrations within the domain, and in particular at Big Bend NP. Once this base simulation was established, it was compared to simulations in which regional sulfur emissions were modified. The change in predicted sulfate arising from these emission sensitivity simulations defined a source region's contribution at Big Bend NP. Four source regions were evaluated for their contribution to predicted sulfate at Big Bend NP: Mexico, Texas, the eastern U.S., and the western U.S. (Figure 2-14). Sulfur dioxide and primary sulfate emissions from the source regions of interest were modified by either 1) removing these emissions from the source region ("emissions out" scenario), or 2) retaining these emissions, but removing sulfur dioxide and primary sulfate emissions from the rest of the domain ("emissions in" scenario). Examples of the modified emission inventories are shown in Figure 2-15. These two approaches for treating emissions were designed to elucidate the potential nonlinear behavior in the chemical conversion of SO_2 to SO_4 due to the availability of oxidants (e.g., hydroxyl, hydrogen peroxide). For example, in the "emissions out" simulation for Texas (Figure 2-15a), more SO_2 is present within the model domain as compared to the "emissions in" simulation (Figure 2-15b). This additional SO_2 could have the effect of reducing the predicted oxidizing capacity of the atmosphere, as oxidants would be consumed by reacting with the available SO_2 outside of the source region of interest. In contrast, more oxidants will be available in the "emissions in" simulation, as they will not have reacted with SO_2 from areas outside of the source region of interest. The difference between these two emission scenarios should indicate the model's sensitivity to predicted oxidant concentrations, and whether the conversion of SO_2 to SO_4 is "oxidant limited". In addition to the four large regions discussed here, several smaller source regions were also evaluated in an effort to refine the source contributions to sulfate at Big Bend NP.

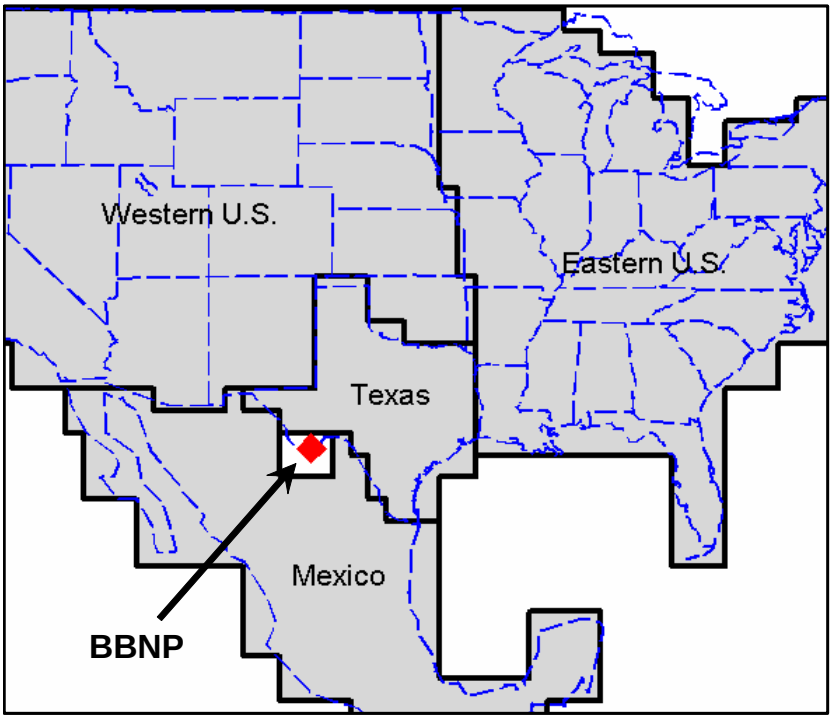

Figure 2-14. The four source regions evaluated for their impact on SO_4 predicted at BBNP: Mexico, Texas, the eastern U.S., and the western U.S. SO_2 emissions in the region immediately surrounding BBNP are near zero and not included in the emissions processing.

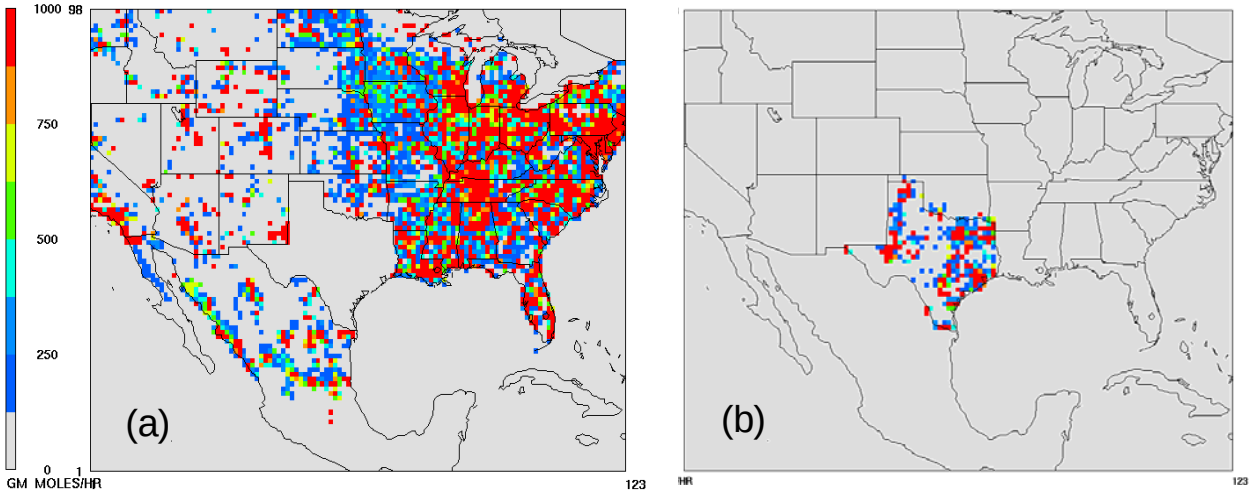

Figure 2-15. Example SO_2 emissions (moles/hr) for (a) the Texas "emissions out" scenario and (b) the Texas "emissions in" scenario.

2.3.2.5 BRAVO Source Areas

BRAVO source areas used for sulfate source attribution were chosen by group consensus in order to facilitate comparison of the results from different models.

Factors considered in the choice of source areas were:

1. Division of areas based on political or scientific desire to know the contributions from those areas to concentrations at Big Bend. For example, there was interest in knowing the relative contributions of sources in Mexico vs. sources in the United States to sulfate at Big Bend. There was also interest in estimating the contribution of the Carbón I & II electric generating stations and in knowing how much of the U.S. contribution was from Texas vs. how much was from outside Texas.
2. For the receptor models that utilize forward or backward trajectories, collinearities between source areas had to be minimized. Using those techniques, it's not possible to separate the contributions from two areas if airmasses always or almost always pass over both areas before arriving at the receptor.
3. Limitations of the modeling domain. For example, although Africa is a suspected source of dust during the summer, the modeling domain did not extend that far from Big Bend due to the geographic limits of the gridded wind fields.
4. For trajectory-based models, source areas should generally be larger as the distance from the receptor increases. This is to allow for increased uncertainty in trajectory locations and increased dispersion as the distance and travel time between source and receptor increase. Dispersion is explicitly calculated by some models, but not by others.
5. For trajectory-based regression models, the source areas should be small enough so that a single coefficient is sufficient to explain the mean relationship between source area and receptor no matter which portion of the source area the trajectories go through.
6. The regression techniques rely on the contribution from the source regions being larger than the errors in the model and observations. Therefore, each source region needed large SO_2 emission rates.

Sources chosen for sulfate source attribution after consideration of these factors are shown in Figure 2-16, with the names given in Table 2-5. Each of the 27 smaller source areas was also assigned to one of four large regions: Texas, Mexico, the eastern U.S., or the western U.S. These are also shown in Table 2-5.

Similar criteria were used to choose the four source areas representing the perfluorocarbon tracer release sites for Trajectory Mass Balance receptor modeling which are shown in Figure 2-17.

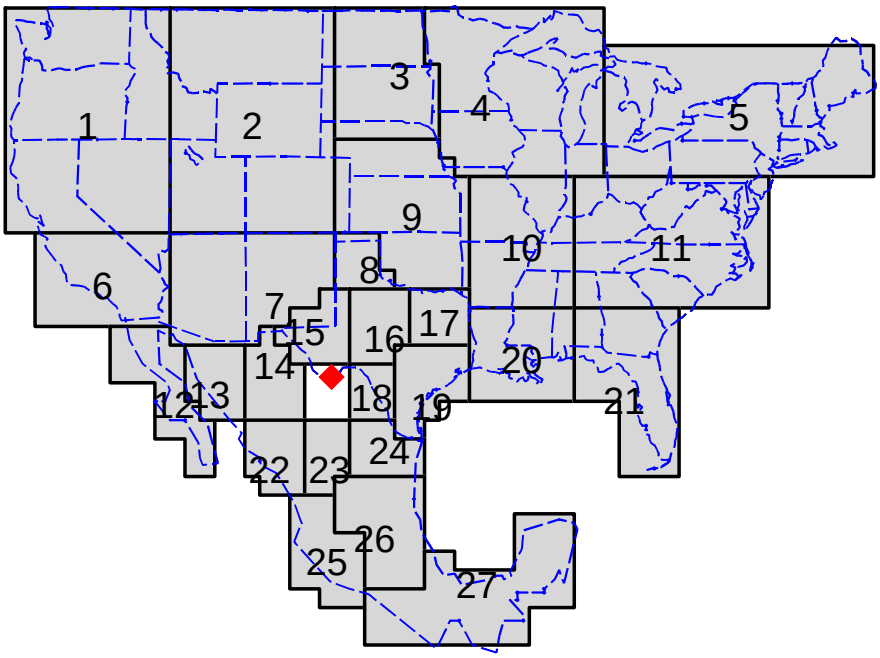

Figure 2-16. Consensus source areas used for source attribution of sulfate at Big Bend.

Table 2-5. Sulfate source area names.

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

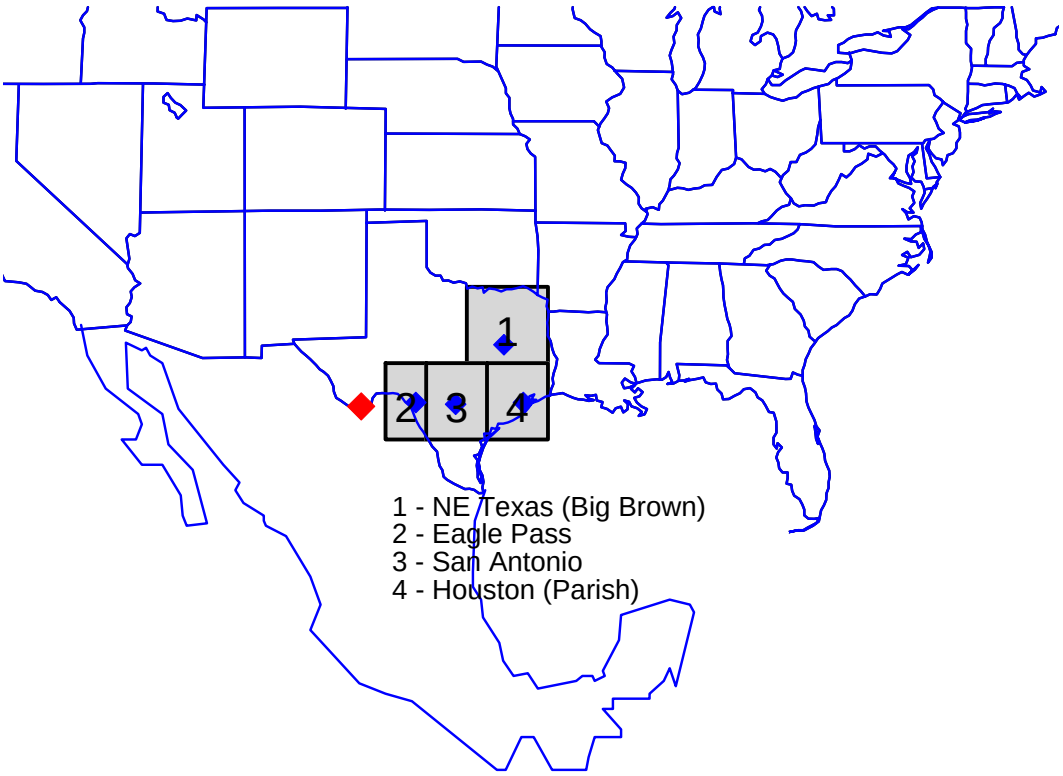

Figure 2-17. Source areas used for TrMB modeling of tracers.

The Forward Mass Balance Regression (FMBR) technique used a modified version of the source regions in Figures 2-16 and 2-17. This was necessary, because in the FMBR analysis forward airmass transport was simulated from 100 x 100 km areas throughout a domain smaller than depicted in Figure 2-16. The source areas could not transect one of the 100 x 100 km areas, so the source areas in Figure 2-16 were shifted up to 50 km. In addition, correlations of transport between some of the source areas were above 0.7, indicating a high degree of collinearity. These source areas were merged and adjusted until no correlation coefficient was above 0.7. This process reduced the 27 source regions to only 17 which are presented in Figure 2-18. Figure 2-19 presents the modified source areas representing the perfluorocarbon tracer release sites used in the FMBR modeling of the tracer data.

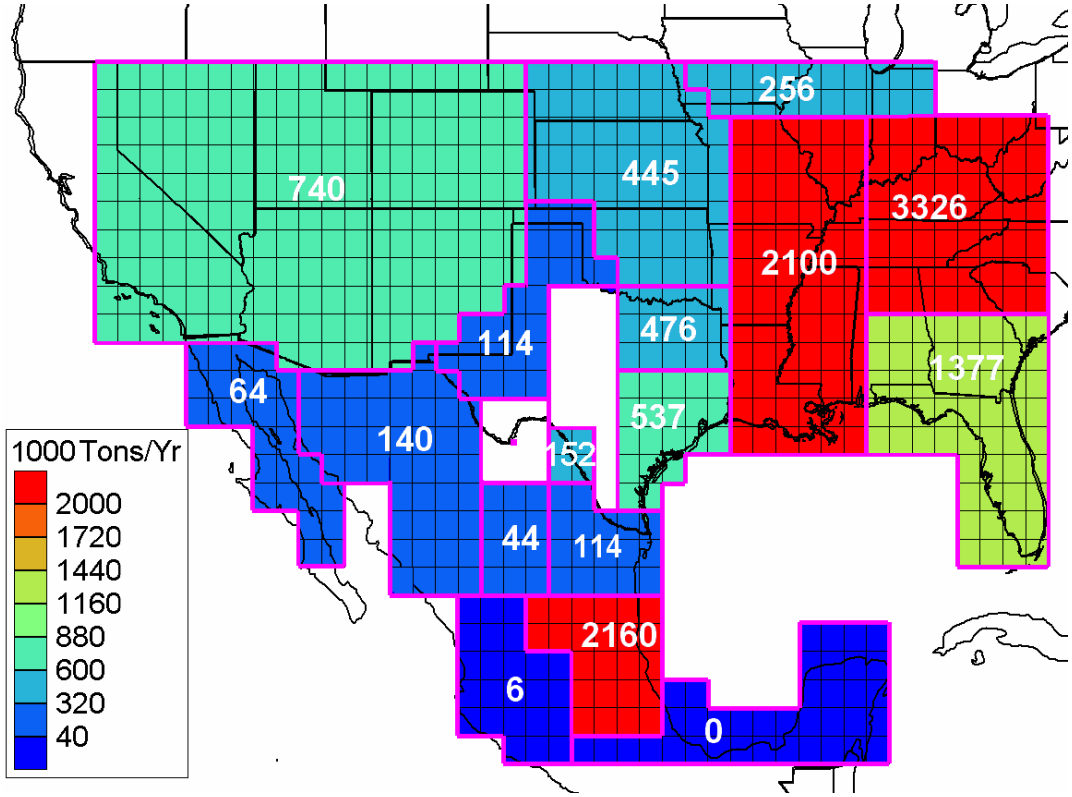

Figure 2-18. Modified consensus source areas used in the FMBR source attribution of sulfate at Big Bend. The grid represents the 100 x 100 km areas that forward airmass transport was simulated from. The values and colors of each source region are the SO_2 emission rates in 1000 tons/year from the BRAVO and National Emission Trend inventories.

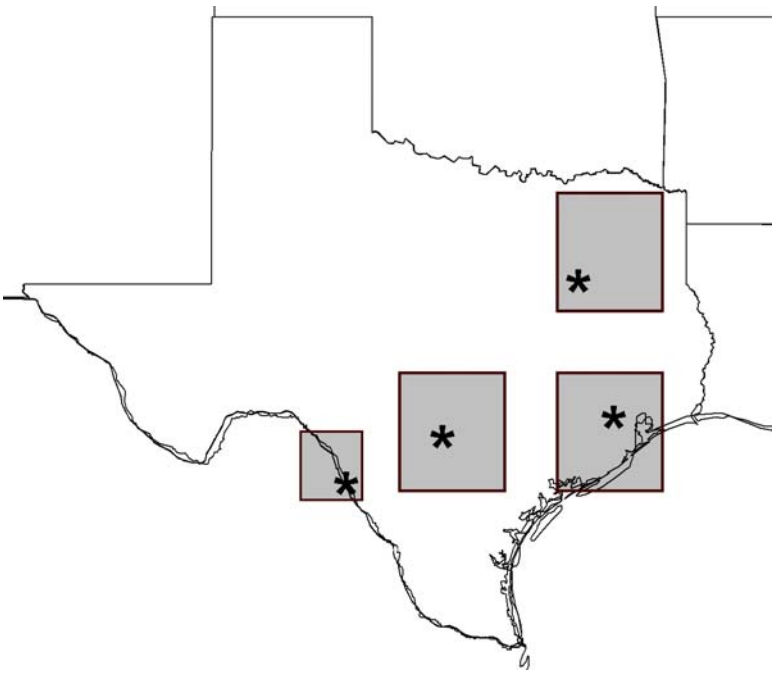

Figure 2-19. Source areas used for FMBR modeling of tracers.