Understanding Relationships Between Changes in Ambient Ozone and Precursor Concentrations and Changes in VOC and NOx Emissions from 1990 to 2004 in Central California

Phase I Interim Report

Contract No. 05-6CCOS

May 1, 2006

prepared for

San Joaquin Valleywide Air Pollution Study Agency
c/o
California Air Resources Board
1001 "I" Street
Sacramento, California 95814

submitted by

Charles L. Blanchard
Shelley Tanenbaum
Envair

Eric Fujita
Desert Research Institute

Jim Wilkinson
Alpine Geophysics

Disclaimer

ABSTRACT

We carried out a preliminary meteorological classification of central California ozone-season (May 1 to October 31) days during 1990 through 2004. The statistical method was K-means clustering using daily-average pressure gradients from San Francisco to Medford, Reno, and Fresno, 850 mb temperatures, heights, wind speed, and wind direction at 4 am and 4 pm at Oakland, and surface wind speed and direction at nine monitoring locations throughout central California. Although the clustering analysis is preliminary and subject to revision, it successfully identified five clusters that had higher mean peak 8-hour ozone levels than did other clusters, and these high-ozone clusters had differing meteorological characteristics and differing levels of ambient ozone precursor concentrations. Mean transport distances and directions varied among the high-ozone clusters, depending upon the monitoring location. Three of the clusters occurred most frequently in July and August, while the remaining two clusters occurred most frequently in September and October.

As a first step in trends analysis, we identified the top 60 peak 8-hour ozone days during each year within each of nine subregions of the CCOS domain. We then compiled several air quality metrics from hourly measurements made on the top 60 days. Decreasing concentrations of ozone precursors were observed not only in the morning, when ambient concentrations are expected to be dominated by fresh, local emissions in many locations, but also throughout the daytime hours when ozone formation occurs. Morning or mid-day CO and NO_x trends were statistically significant (downward, $p< 0.05$) at 20 of 24 sites and 22 of 25 sites, respectively. Trends in top 60 average peak 8-hour ozone concentrations were similar to trends in annual 4^{th}-highest daily ozone maxima, but with typically smaller rates of change, indicating that conclusions based on analyses of top 60 days are relevant to regulatory metrics.

The next steps are to revise and refine the meteorological classifications and to determine trends in ozone and ozone precursor metrics within each meteorological class.

CONTENTS

I. INTRODUCTION

The objective of this project is to clarify the relationships between trends in emissions of ozone precursors, trends in ambient concentrations of ozone precursors, and trends in ozone levels in central California from 1990 through 2004. The geographic domain of interest includes the entire region of the Central California Ozone Study, which incorporates the San Joaquin Valley, the Sacramento Valley, the San Francisco Bay area, and the central California coast.

The analysis of trends in air pollutant levels provides one (but not the only) method for evaluating the effectiveness of emission control programs. Because different types of meteorological conditions yield movements of ozone precursors from different source areas to particular receptors, the methods employed for analyzing trends could either reveal or obscure important links between emission sources and receptors. Trend analyses that are capable of linking subregional or site-specific trends in ozone levels to emission changes occurring in different areas of emission-source influence are likely to be of great utility for enhancing the effectiveness of emission control programs.

This project is divided into two phases, and this interim report documents the first portion of Phase I. Draft work plans were submitted in December 2005 and January 2006, and and suggestions were incorporated into the final version of the work plan that was submitted February 1, 2006.

II. DATA ACQUISITION

Meteorological Data

For our purposes, the most useful meteorological data are long-term measurements that cover most of the 1990 – 2004 period and that were made using consistent methods. The meteorological data are used here to classify days into different types, where the goal is to obtain types that represent different transport regimes. The primary classification method is clustering of ozone-season days (May 1 to October 31) based on sea-level pressure gradients, upper-level (850 mb) temperature, height, wind speed, and wind direction, and on surface wind speed and direction. Accordingly, the data that we obtained are:

- 24-hour mean surface pressure at San Francisco, Medford, Reno, and Fresno, from which we calculated sea-level pressure gradients from San Francisco to each of the other three locations;
- 850 mb temperature, height, wind speed, and wind direction at Oakland (00 and 12 UT, or 4 am and 4 pm local standard time);
- hourly wind speed and direction at 53 stations, from which we calculated 24-hour wind speed and direction (Table 1 and Figure 1).

The surface wind speed and wind direction measurements from three networks provided sufficient coverage of the CCOS domain for our classification analyses. Construction of diagnostic wind fields (which was not part of our work) would benefit from acquisition of meteorological data from additional networks.

Table 1. Meteorological monitoring sites in the CCOS region that have 10 or more years (1990-2004) of wind speed and wind direction. Networks are the Bay Area Air Quality Management District (BAAQMD), California Irrigation Management Information System (CIMIS), and National Climate Data Center – National Oceanic and Atmospheric Administration (NCDC).

NAME	SITE CODE	NETWORK CODE	BASIN	LATITUDE	LONGITUDE
ALVISO	ALV	BAQ	SFB	37.435	-121.95
ARROYO SECO	ECO	CIM	NCC	36.359	-121.29
BAKERSFIELD MEADOWS FIELD		NCDC		35.433	-119.05
BETTERAVIA	BET	CIM	SCC	34.924	-120.512
BISHOP	BIC	CIM	GBV	37.358	-118.404
BLACKWELLS CORNER	BKC	CIM	SJV	35.65	-119.958
BRENTWOOD	BTW	CIM	SFB	37.929	-121.659
BROWNS VALLEY	BRV	CIM	SV	39.271	-121.305
CAMINO	CMO	CIM	MC	38.754	-120.733
CARNEROS	CNR	CIM	SFB	38.219	-122.354
CASTROVILLE	CTV	CIM	NCC	36.768	-121.774
CHABOT	CHA	BAQ	SFB	37.721	-122.1
CHINA LAKE		NCDC		35.683	-117.7
COLUSA	CLU	CIM	SV	39.226	-122.024
CUYAMA	CUY	CIM	SCC	34.932	-119.605
DAVIS	DAV	CIM	SV	38.536	-121.776
DE LAVEAGA	AGA	CIM	NCC	36.998	-121.996
DURHAM	DUR	CIM	SV	39.609	-121.823
FIREBAUG	FRB	CIM	SJV	36.851	-120.59
FIVEPOINTS	FIV	CIM	SJV	36.336	-120.113
FRESNO STATE	FNS	CIM	SJV	36.821	-119.742
FRESNO YOSEMITE AIRPORT		NCDC		36.783	-119.717
GOLETA FOOTHILLS	GOL	CIM	SCC	34.472	-119.868
GREEN VALLEY ROAD	GVR	CIM	NCC	36.941	-121.767
HOPLAND	HFS	CIM	NC	39.014	-123.089
KESTERSON	KES	CIM	SJV	37.233	-120.88
KETTLEMAN CITY	KET	CIM	SJV	35.869	-119.894
KING CITY	KCO	CIM	SCC	36.121	-121.084
KREGOR PEAK	KRE	BAQ	SFB	37.943	-121.9
LINDCOVE	LCV	CIM	SJV	36.357	-119.059
LOS BANOS	NOS	CIM	SJV	37.092	-120.76
MANTECA	MAN	CIM	SJV	37.835	-121.223
MODESTO	MDT	CIM	SJV	37.636	-121.186
NAPA	NAP	BAQ	SFB	38.25	-122.3
NICOLAUS	NIC	CIM	SV	38.871	-121.545
OAKVILLE	OKV	CIM	SFB	38.434	-122.41
ORLAND	ORL	CIM	SV	39.692	-122.152
PANOCHE	PSV	CIM	SJV	36.889	-120.731
PARLIER	PRL	CIM	SJV	36.598	-119.503
SACRAMENO EXECUTIVE AIRPORT		NCDC		38.5	-121.5
SALINAS NORTH	SNC	CIM	NCC	36.717	-121.691
SALINAS SOUTH	SSC	CIM	NCC	36.621	-121.545
SAN LUIS OBISPO	SLB	CIM	SCC	35.306	-120.66
SANEL VALLEY	SVC	CIM	NC	38.983	-123.092
SANTA BARBARA	ARA	CIM	SCC	34.438	-119.736
SANTA ROSA	SRA	CIM	NC	38.401	-122.799
SANTA YNEZ	SYC	CIM	SCC	34.583	-120.078
STRATFORD	STF	CIM	SJV	36.158	-119.85
SUNOL	SUN	BAQ	SFB	37.594	-121.88
VISALIA	VLA	CIM	SJV	36.301	-119.223
WESTLAND	WLC	CIM	SJV	36.633	-120.382
WINDSOR	WDC	CIM	NC	38.668	-122.828
ZAMORA	ZOR	CIM	SV	38.808	-121.908

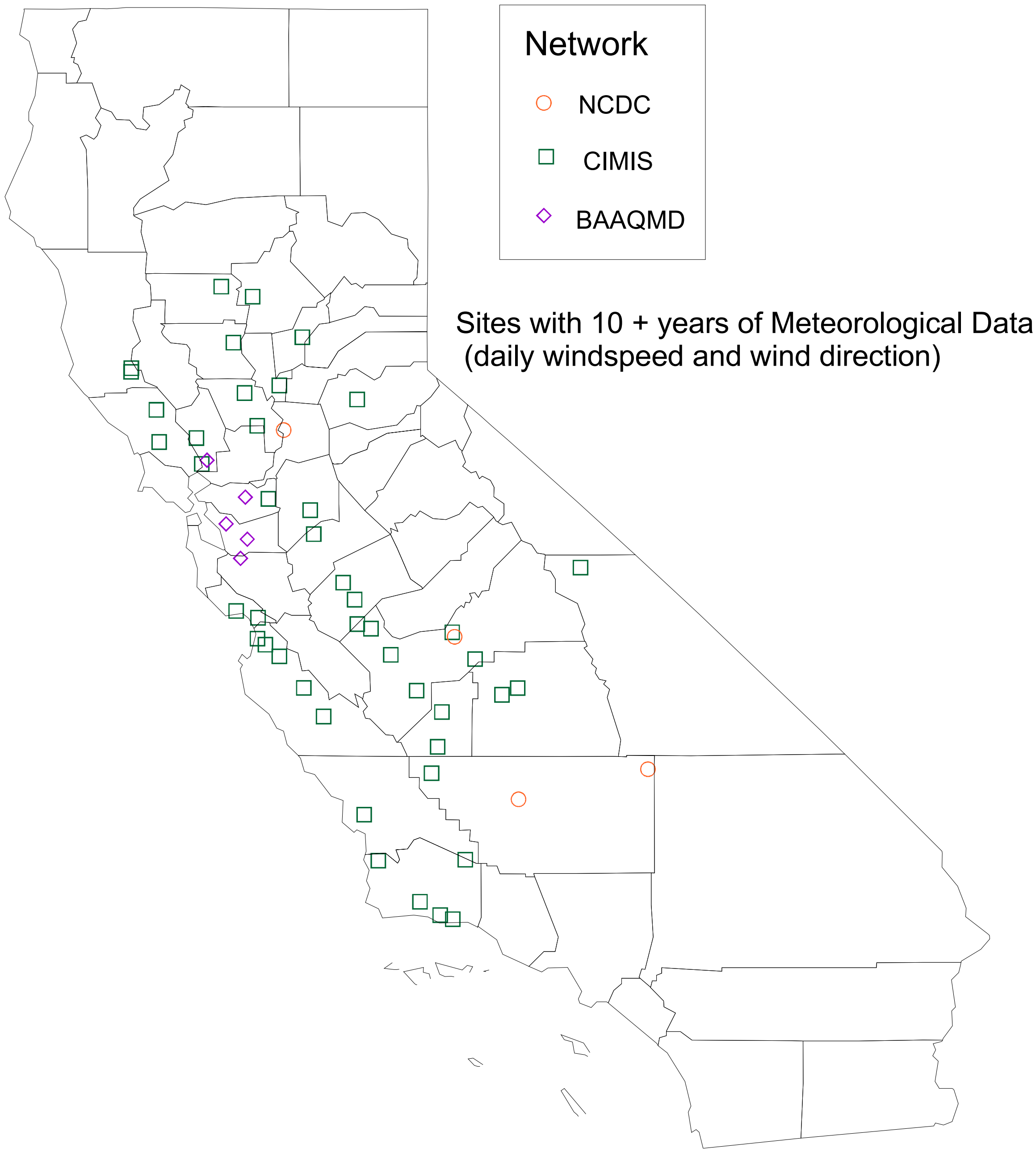

Figure 1. Meteorological monitoring sites in the CCOS region that have 10 or more years (1990-2004) of wind speed and wind direction data. Networks are the Bay Area Air Quality Management District (BAAQMD), California Irrigation Management Information System (CIMIS), and National Climate Data Center – National Oceanic and Atmospheric Administration (NCDC).

Hourly Air Pollutant Measurements

We compiled hourly gas-phase air-pollutant measurements made from 1990 through 2004, as summarized in Table 2, for each long-term air-quality monitoring site in nine subregions of the CCOS domain[1], as follows:

- Southern Bay Area
  - Fremont
  - Gilroy
  - Los Gatos
  - San Jose - 4th Street
  - San Jose - Piedmont
  - San Martin

- Northeastern Bay Area
  - Bethel Island
  - Concord
  - Pittsburg
  - Vallejo

- Eastern Bay Area
  - Livermore - Old First Street
  - Livermore – Rincon

- Urban/suburban Sacramento
  - Folsom - City Yard
  - Folsom - Natoma Street
  - North Highlands
  - Rocklin-Sierra College
  - Roseville-N Sunrise Blvd
  - Sacramento - Del Paso Manor
  - Sacramento - El Camino and Watt
  - Sacramento - T Street
  - Sloughhouse
  - Elk Grove

- Northern Sierra Foothills
  - Auburn - DeWitt C Avenue
  - Cool
  - Grass Valley - Litton Building
  - Placerville - Gold Nugget Way

---

[1] As specified in the work plan and the RFP.

- Southern Sierra Foothills
  - Five Mile Learning Center
  - Jackson
  - Jerseydale
  - San Andreas
  - Sonora
  - Yosemite

- Northern San Joaquin Valley
  - Merced-S Coffee Avenue
  - Modesto - 14th Street
  - Stockton - Hazleton
  - Stockton - Mariposa
  - Tracy-24371 Patterson Pass Road
  - Turlock - S Minaret Street

- Central San Joaquin Valley
  - Fresno - 1st Street
  - Fresno - Drummond
  - Fresno - Sierra Skypark #2
  - Madera
  - Clovis
  - Parlier
  - Visalia

- Southern San Joaquin Valley
  - Arvin - Bear Mountain Road
  - Bakersfield - 5558 California Avenue
  - Bakersfield - Golden State Highway
  - Edison
  - Maricopa
  - Oildale - 3311 Manor Street

Table 1. Summary of data availability for designated sites, 1990-2004 (75 percent daily and annual completeness).

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
MC	Northern Sierra Foothills	Grass Valley	3126	1994-2004	10										
		Placerville	3017	1993-2004	11									1993-2004	12
	Southern Sierra Foothills	Five Mile	3164	1996-2000	4										
		Jackson	2993	1993-2004	12									1993-2004	12
		San Andreas	3144	1995-2004	10									1995-2004	10
		Sonora	2968	1993-2004	12									1993-2004	12
		Yosemite Wawona	2183	1991-1996	5										
		Yosemite Camp Mather	2949	1992-1994	3										
		Yosemite Turtleback	3018	1993-2004	12										
		Yosemite HQ	3045	1990-1994	5										
		Yosemite Merced River	3664	2004-2004	1	2004-2004	1			2004-2004	1			2004-2004	1
SFB	Northern Bay Area	Bethel Is.	2804	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Concord	2831	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Pittsburgh	2102	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Vallejo	2410	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
	Southern Bay Area	Fremont	2293	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Gilroy	2320	1990-1997	8									1990-1993	4
		Los Gatos	2613	1990-2004	12										
		San Jose 4th St.	2413	1990-2001	12	1990-2001	12	1990-2001	12	1990-1993	3			1990-2001	12
		San Jose Piedmont	2969	1993-1997	5										
		San Martin	3140	1995-1997	3										
	Eastern Bay Area	Livermore Old First St	2372	1990-2000	11	1990-2000	11	1990-2000	11	1990-1993	3			1990-1999	10
		Livermore Rincon	3490	2000-2004	5	2000-2004	5	2000-2004	5					2000-2004	5
SJV	Northern San Joaquin	Merced	3022	1992-2004	13	1992-2004	11	1992-2004	11	1992-2004	11				
		Modesto	2833	1990-2004	14	1990-2004	14	1990-2004	14	1990-2004	14			1990-2004	14
		Stockton Hazleton	2094	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	15
		Stockton Mariposa	2553	1990-2000	11										
		Tracy	3159	1996-2003	8	1996-2003	8	1996-2003	8	1996-2003	8				
		Turlock	2996	1993-2004	12	1993-2004	12	1993-2004	12	1993-2004	12			1993-2004	12
	Central San Joaquin	Clovis	3026	1991-2004	14	1991-2004	14	1991-2004	14	1991-2004	14	2003-2004	2	1991-2004	14
		Fresno 1st St.	3009	1991-2004	14	1990-2004	15	1990-2004	15	1990-2004	15	2003-2004	2	1990-2004	15
		Fresno Drummond	2013	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	15

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
		Fresno Sierra Skypark #2	2844	1992-2004	13	1991-2004	14	1991-2004	14	1991-2004	14			1990-2004	15
		Madera Health Dept.	2591	1991-1996	6										
		Madera Pump Yard	3211	1998-2004	7	1998-2004	7	1998-2004	7	1998-2004	7	2004-2004	1		
		Parlier	2114	1990-2004	15	1994-2004	11	1994-2004	11	1994-2004	11	2003-2004	2	1990-1993	4
		Visalia	2032	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	15
	Southern San Joaquin	Arvin	2941	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15	2003-2004	2		
		Bakersfield Golden State Hwy	3145	1995-2004	10	1995-2003	8	1995-2003	9	1995-2003	9	2003-2004	2	1995-2004	9
		Bakersfield 5558 CA Ave	3146	1995-2004	10	1995-2004	10	1995-2004	10	1995-2004	10			1995-2004	10
		Edison	2312	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-1993	4
		Maricopa	2919	1990-2004	14										
		Oildale	2772	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-1993	4
SV	Northern Sierra Foothills	Auburn	2891	1990-2004	9										
	Urban/ Suburban Sacramento	Elk Grove	2977	1993-2004	12	1993-2004	10	1993-2004	10	1993-2004	10				
		Folsom	2472	1990-1996	7	1990-1995	6	1990-1995	6	1990-1995	6				

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
		Folsom	3187	1997-2004	8	1997-2004	7	1997-2004	7	1997-2004	7	2003-2004	2		
		North Highlands	2123	1990-2004	15	1990-2004	12	1990-2004	12	1990-2004	12			1990-2004	15
		Rocklin	2656	1990-1990	1										
		Roseville	2956	1993-2004	12	1993-2004	12	1993-2004	12	1993-2004	12			1993-2004	12
		Sacramento Del Paso Manor	2731	1990-2004	15	1990-2002	12	1990-2002	12	1990-2002	12	2003-2003	1	1990-2004	14
		Sacramento El Camino & Watt	2840											1990-2004	14
		Sacramento T St.	3011	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	14

PAMS VOC Data

Although other research groups have created validated hydrocarbon databases, all such databases covered limited periods of time or spatial domains. No combination of previously created databases covered the full CCOS domain and time period of interest in this study. Therefore, we acquired hydrocarbon data directly from EPA archives and conducted our own validation.

The EPA data included hourly hydrocarbon levels, multi-hour PAMS hydrocarbon and carbonyl data, and 24-hour resolution air-toxics data (of which certain species, such as benzene and toluene, are of principal interest). We found no discrepancies between the PAMS measurements that are available from EPA and those that are available from CARB (some measurements, such as the air-toxics data, were included in the EPA database but not in the CARB archives). Our level 2 data validation process includes the following steps:

- Determine the following values for each sample: sum of 55 target PAMS hydrocarbons, NMHC/NMOC, PAMS/NMOC ratios, % ethene, % acetylene, ethene/acetylene ratios, % ethane plus propane, % isopentane, % benzene, % toluene, % xylenes, sum of xylenes/benzene ratios, % n-C_{10} plus nC-^{11} , formaldehyde/NMOC ratios. Compare summary statistics (mean, maximum, minimum, standard deviation) for these values by sampling site, year and sampling period with expected ranges in values consistent with the site characteristics for each site and flag samples that depart significantly from these expectations.

- Derive summary statistics by sampling site, year and sampling period (mean, maximum, standard deviation) for all species and species groups, both for concentration and weight fractions. Data for each species will be stratified by site and time of day and values above 5 ppbC for hydrocarbons and 2 ppbv for carbonyl compounds will be flagged if the ratio of the absolute difference between sample and mean values to the standard deviation among values within the stratified group is greater than three (i.e., |value-mean|/sigma > 3).

- Prepare mean diurnal time-series plots several of the key parameters identified in the first bullet and examine consistency of diurnal patterns with site characteristics.

- Prepare scatterplots of total NMOC from speciation data versus collocated continuous measurements (e.g, preconcentration direct injection FID or TEI 55) if data are available.
- Remove and document data identified as suspect.

Emission Inventories

We obtained final versions of the ARB emissions inventories in March and carried out a preliminary data review.[2] All files were readable, and are now installed among our databases. No obvious errors or discrepancies were found. We will carry out additional consistency checks as a part of our emission trends analyses.

---

[2] These inventories were provided for use in the present project only, and may undergo additional review and revision by CARB.

III. METEOROLOGICAL CLASSIFICATION

Approach

We applied K-means clustering using daily-average pressure gradients, 850 mb temperatures, heights, wind speed, and wind direction, and surface wind speed and direction. K-means clustering is a statistical procedure that divides objects (days, in the present application) so that days in different clusters exhibit different values of the pressure gradients, 850 mb data, surface wind speed and direction, or other variables specified by the analyst. The procedure is analogous to a multivariate analysis of variance in which the groups are not known in advance. Consequently, the user must specify the number of groups, which can require consideration of additional criteria (not necessarily of a statistical nature). Clustering is an iterative procedure, which begins by selecting one "seed" for each cluster, then assigning and re-assigning days to clusters by minimizing the within-groups sums of squares.

We report our initial clustering results. We expect to revise and refine the clustering analysis, so the preliminary results will likely change.

For our initial clustering evaluation, we specified 12 clusters. Our expectation was that about half would prove to be populated largely by days having lower ozone levels, and therefore would not be of interest. Ideally, we would obtain approximately 4 to 6 high-ozone clusters representing different transport directions and/or distances. Since we will be disaggregating the top 60 ozone days per year into different clusters, we would expect approximately 10 to 20 days per year per cluster on average.[3]

We used the following variables for clustering:

- Sea-level pressure gradients from San Francisco to Medford, Reno, and Fresno

---

[3] The clustering was applied to all days from May 1 through October 31 of each year. Subsequently, days within the top 60 subsets will be identified by cluster. Each subregion's top 60 subsets consist of all days that are among the top 60 peak ozone days at any site in the subregion, so the full set of high-ozone days for each subregion will usually number more than 60 per year.

- 850 mb height, temperature, wind speed, and wind direction at Oakland at 4 am and 4 pm local standard time
- Surface wind speed and direction at nine locations: Bakersfield, Brentwood, Chabot, Davis, Fresno, Modesto, Parlier, Sacramento, and Santa Rosa. For this initial clustering, we used only nine surface sites to allow us to test cluster differences at locations whose data were not used as part of the clustering input. We picked locations throughout the geographical domain.

All wind speeds and directions were converted to vector coordinates (u and v components) before clustering. Because the variables used substantially different metrics (meters per second, degrees Celsius, meters elevation, and millibars), we standardized[4] variables before clustering.

We analyzed 2760 days (184 days per year). The resulting clusters ranged in size from 25 to 417 days; excluding the largest and smallest clusters, the cluster sizes ranged from 94 to 402 days. Nineteen days were missing one or more pressure gradients; 618 days were missing 850 mb data. For seven of the sites, surface wind speed and direction data were missing on 14 to 78 days; Parlier and Chabot data were missing on 596 and 1474 days, respectively. Days were classifiable even when one or more measurements were missing; however, we have not yet tested the effects of missing data on the "goodness of splits."

Meteorological Characteristics of Clusters

The 12 clusters display differences in the 850 mb wind directions, as shown in Figure 2. Cluster 4 has a strong northerly component; cluster 11 has a strong southerly component. For the 850 mb winds, the interquartile ranges (25^{th} to 75^{th} percentiles) span a range of roughly 5 m sec^{-1}, on average. The clusters also show differences in surface winds, as illustrated in Figure 3 for Bakersfield. In this case, the interquartile ranges are about 1 m sec^{-1} for most of the clusters.

---

[4] Standardization involves subtracting means and dividing by standard deviations, so that all variables are in units of standard deviations above or below their means.

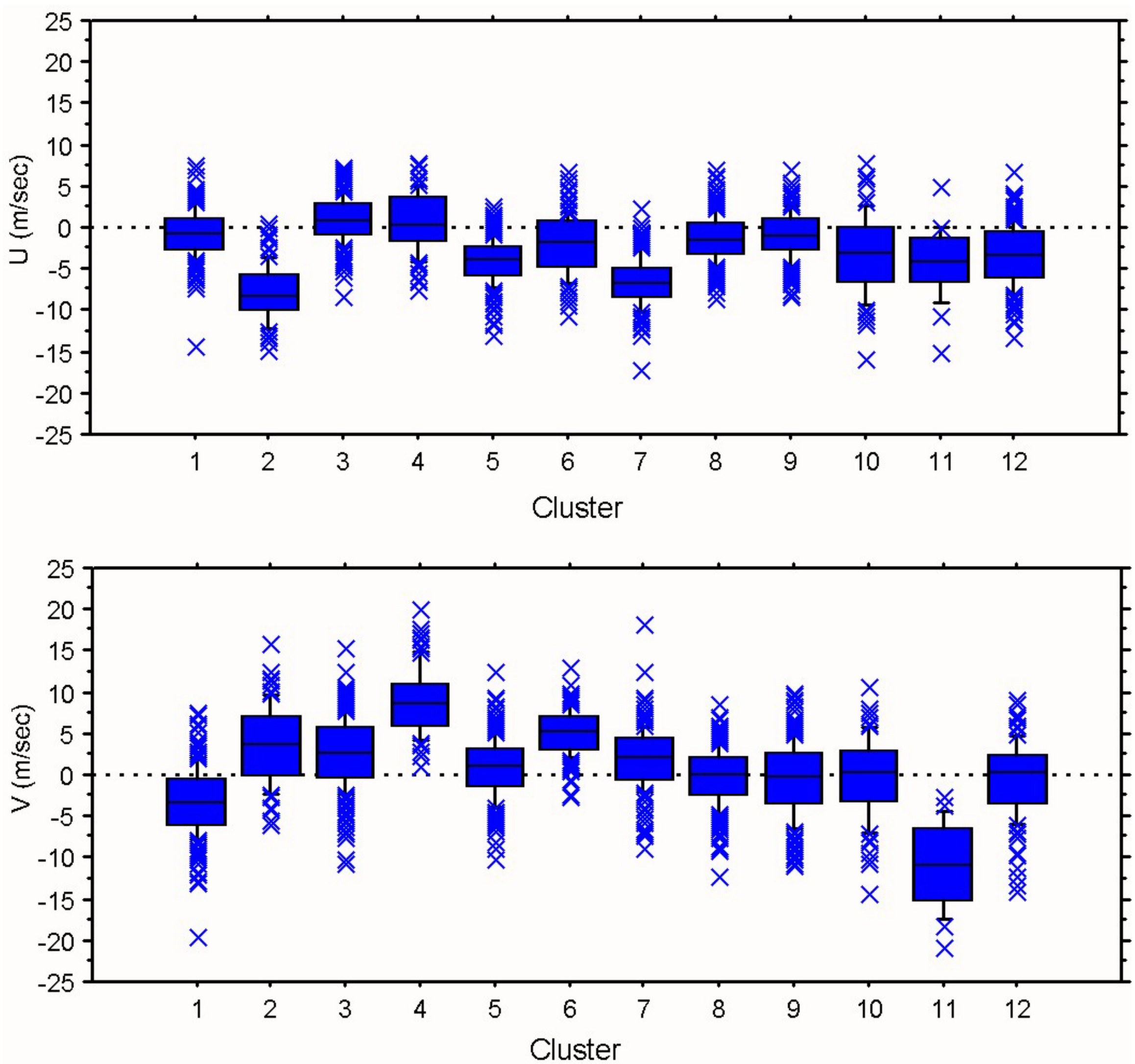

Figure 2. Vector component 850 mb winds at 4 am at Oakland (top) easterly (u, with positive from the east) and (bottom) northerly (v, with positive from the north). The box-and-stem graphics depict the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles, with values below the 10 percentile or above the 90^{th} percentile marked by "x's."

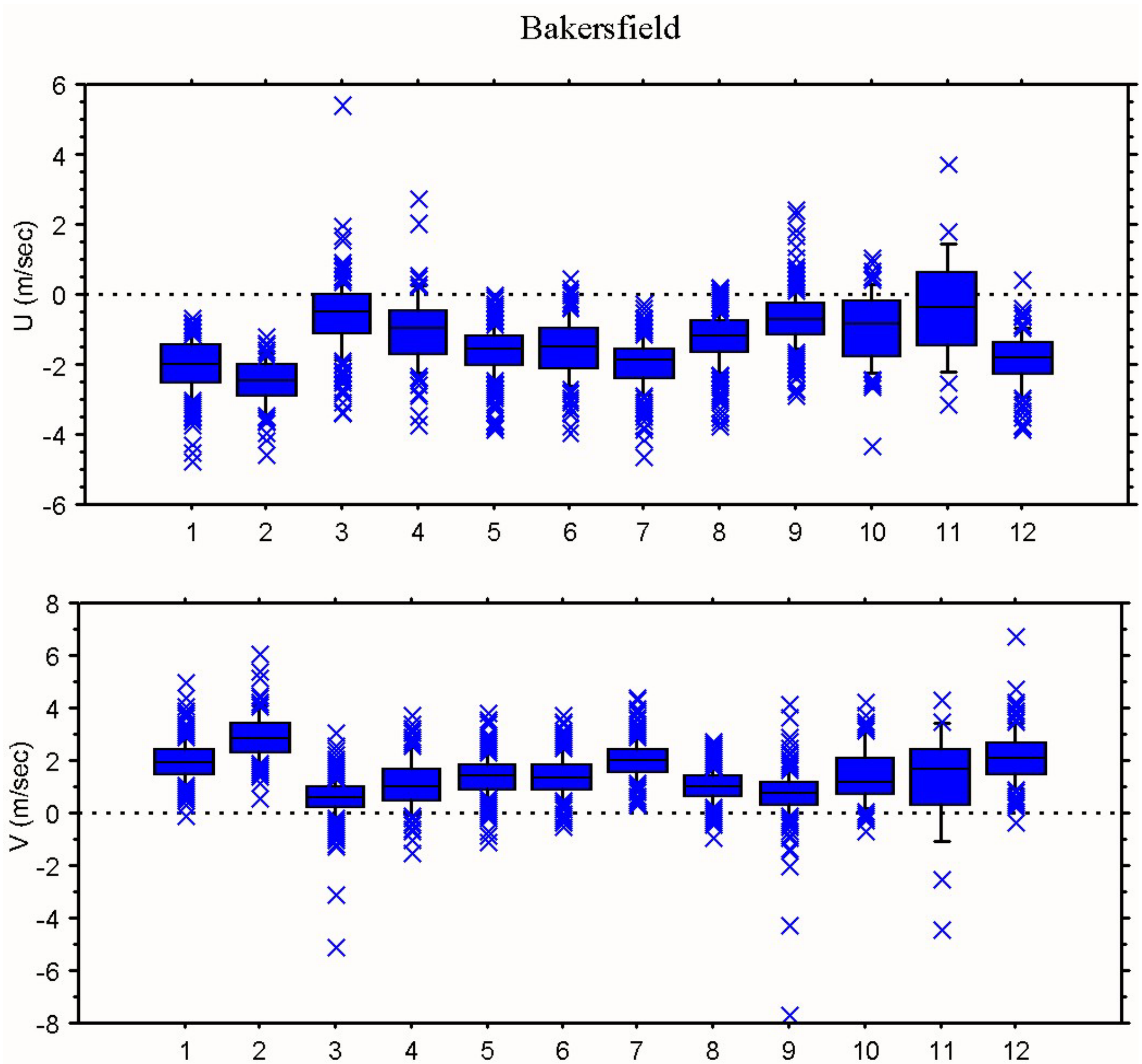

Figure 3. Vector component 24-hour surface winds at Bakersfield (top) easterly (u, with positive from the east) and (bottom) northerly (v, with positive from the north). The box-and-stem graphics depict the 10th, 25th, 50th, 75th, and 90th percentiles, with values below the 10 percentile or above the 90th percentile marked by "x's."

The mean u and v surface wind components at each monitoring site show differences among clusters that are related first to transport distance and second to direction (Figures 4 and 5). On these plots, the monitoring site may be considered to be at the center, so that the dots then show the mean points of origin of air masses for each cluster. Bakersfield and Parlier show strong correlation between the mean u and v components, implying that the principal differences among the present clusters have to do with transport distances, rather than directions. Note that these plots depict the 24-hour mean u and v vector components in units of meters per second; appropriate scaling generates a mean transport distance and direction over a specified time period (e.g., multiplication by 3.6 yields mean transport distances over one hour in units of kilometers). Definition of a time interval (e.g., 24 hours) can generate a data-driven estimate of an approximate zone of influence for each location, which will subsequently be useful for comparison with emission rates determined from appropriate aggregations of emission-inventory grid cells near each monitoring site.

Cluster 11 is unusual in exhibiting a mean southeasterly origin at several locations (Parlier, Sacramento, Manteca, Modesto), and southerly at Chabot. Cluster 4 shows strong flow from the northwest at Sacramento, more moderate northwesterly flow at Manteca and Modesto, northerly-to-northeasterly flow at Napa and Chabot, and limited net flow at Bakersfield and Parlier.

Later in this section we will show that clusters 1, 3, 5, 8, and 9 are characterized by higher mean-ozone concentrations than other clusters. Transport distances and directions exhibit differing degrees of variation among these high-ozone clusters, depending upon the monitoring location (Figures 4 and 5). One objective of a second, more refined clustering analysis will be to obtain high-ozone clusters with greater between-cluster variations of transport distances and directions.

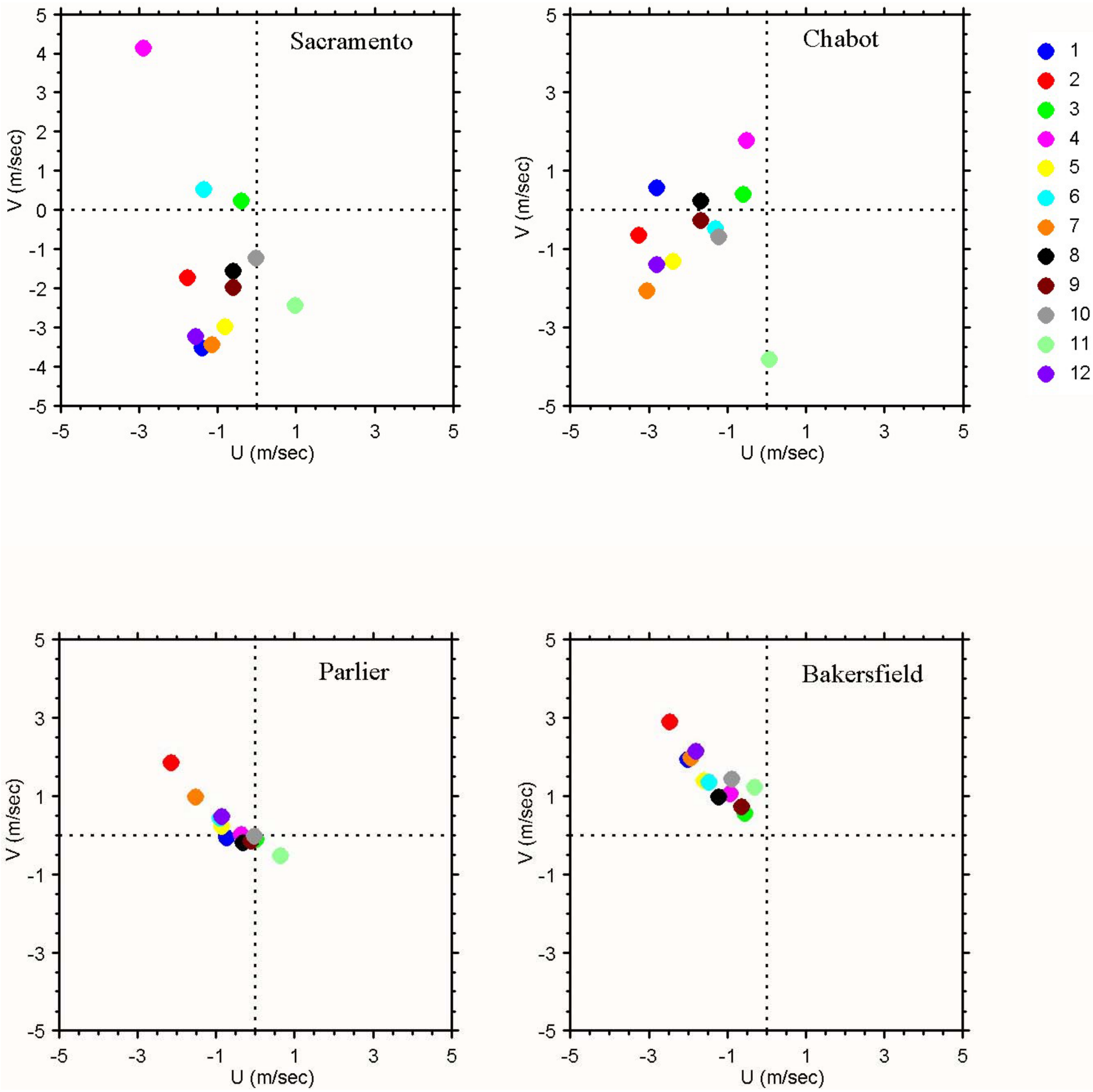

Figure 4. Mean vector component surface winds at four locations, by meteorological cluster. The monitoring site is located at the center of each plot and each dot shows the average point of origin of air masses reaching the monitor. Data from each of the sites shown here were used in the cluster analysis.

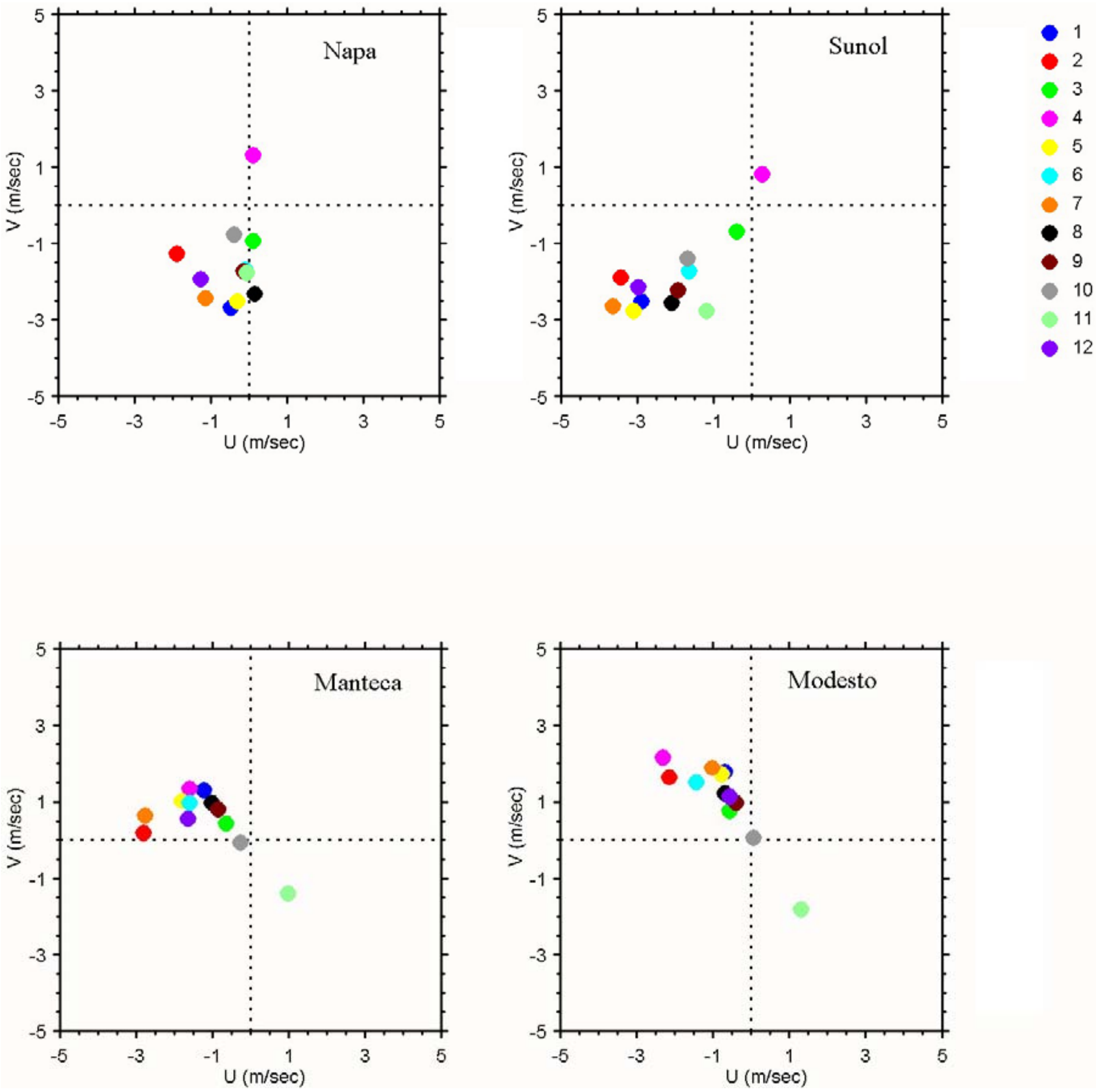

Figure 5. Mean vector component surface winds at four locations, by meteorological cluster. The monitoring site is located at the center of each plot and each dot shows the average point of origin of air masses reaching the monitor. Data from the sites shown here were not used in the cluster analysis.

The data indicate that a given type of cluster tends to persist over periods of one to several days (Figure 6), presumably reflecting synoptic-scale conditions. This result is both expected and desirable, because the cluster analysis was based upon meteorological variables linked to synoptic variations (e.g., pressure gradients) and is intended to group days into clusters having different transport characteristics. Figure 6 suggests the existence of some seasonal variations in cluster frequencies, which is more strongly identified in Figure 7. As shown in Figure 7, clusters 1, 5, and 8 predominate in July and August, whereas clusters 3 and 9 are most common in September and October. The seasonal variation of cluster frequencies reflects real meteorological changes that occur within each year. At the same time, the cluster divisions are more complex than a simple splitting of the data by month, because the mean transport distances and directions at some sites differ among the three July-August clusters (1, 5, and 8), as well as between the two September-October clusters (3 and 9) (Figures 4 and 5). Cluster frequencies do not show obvious trends over time, though there is substantial year-to-year variation (Figure 7).

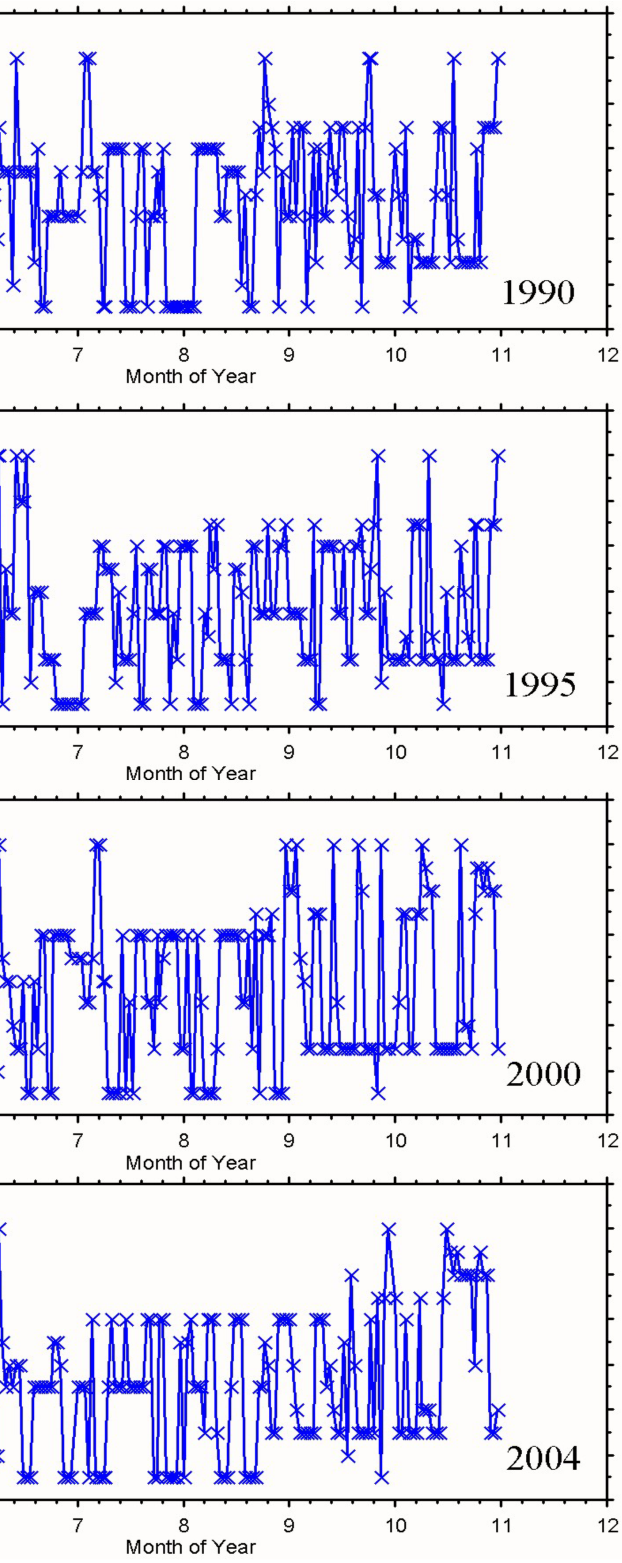

Figure 6. Daily occurrence of cluster types during four years.

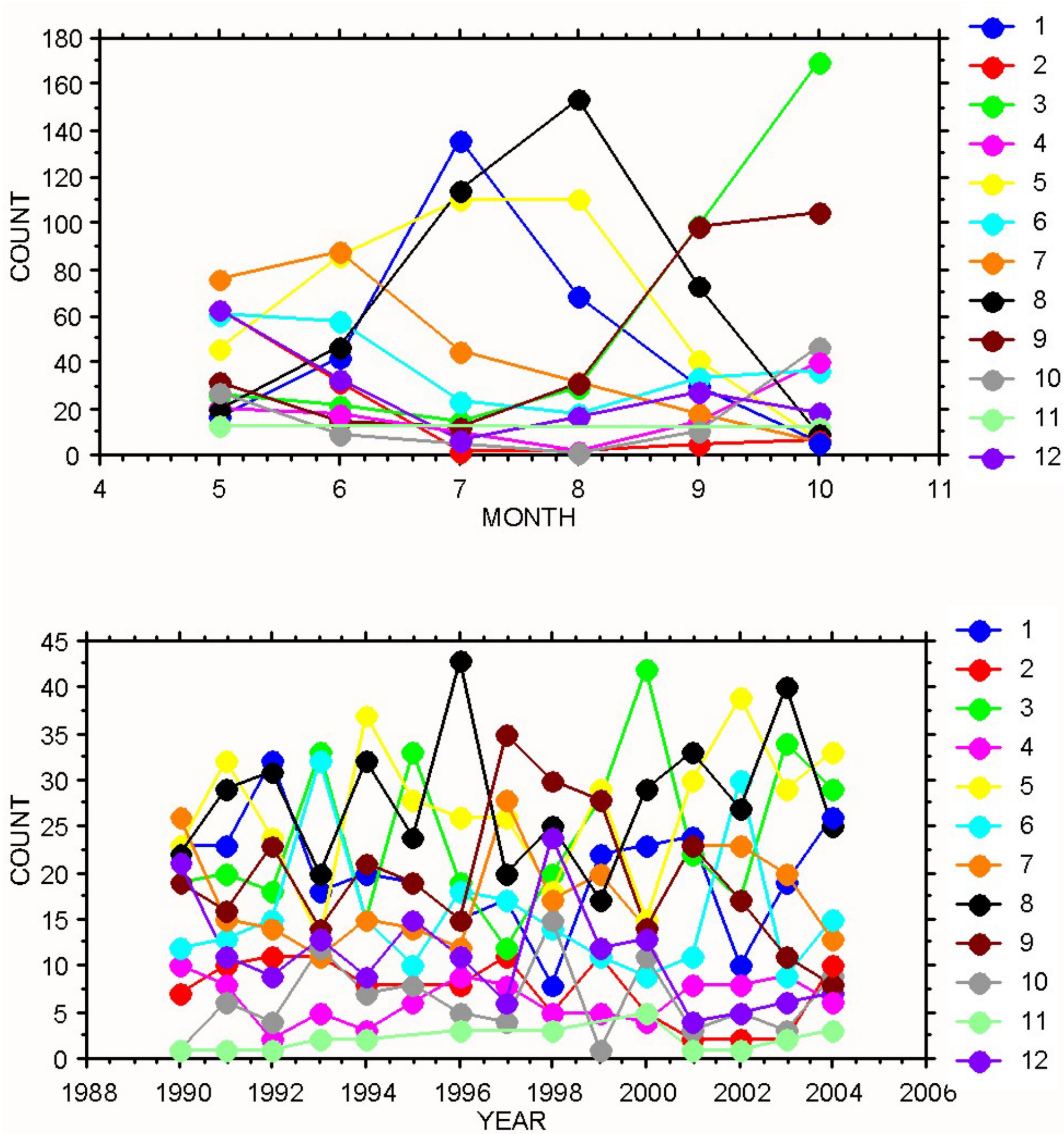

Figure 7. Frequencies of occurrence of days by cluster type versus (top) month and (bottom) year.

Variations of Air Pollutant Concentrations Among Clusters

The different clusters exhibited differences in ambient concentrations of air pollutants (Figures 8, 9, and 10). Clusters 3 and 9 (most prevalent in September and October) tended to show higher ambient concentrations of CO and NO_x than did many other clusters. High mean CO and NO_x was also associated with clusters 8 and 10. Mean ratios of CO/NO_x did not show strong patterns of variation among clusters, though some differences did occur. The mean CO/NO_x ratios varied among sites, with Sacramento showing lower mean ratios than Fresno, for example, and both sites exhibiting lower mean ratios than the overall averages across all monitoring sites.

Clusters 1, 3, 5, 8, and 9 had higher mean peak 8-hour ozone levels than did other clusters. These higher-ozone clusters will therefore be of greater interest in subsequent analyses. As shown in Figure 8, the higher-ozone clusters exhibited different mean concentrations of CO and NO_x. Although the clustering analysis is preliminary and subject to revision, it successfully identified the target number of high-ozone clusters (i.e., 4 to 6), and it identified high-ozone clusters having different levels of ambient ozone precursor concentrations. Three high-ozone clusters (1, 5, and 8) are most frequent during July and August, whereas two (3 and 9) are most frequent during September and October). The frequencies of occurrence differ in a way that is more complex than a simple seasonal difference (Figure 7).

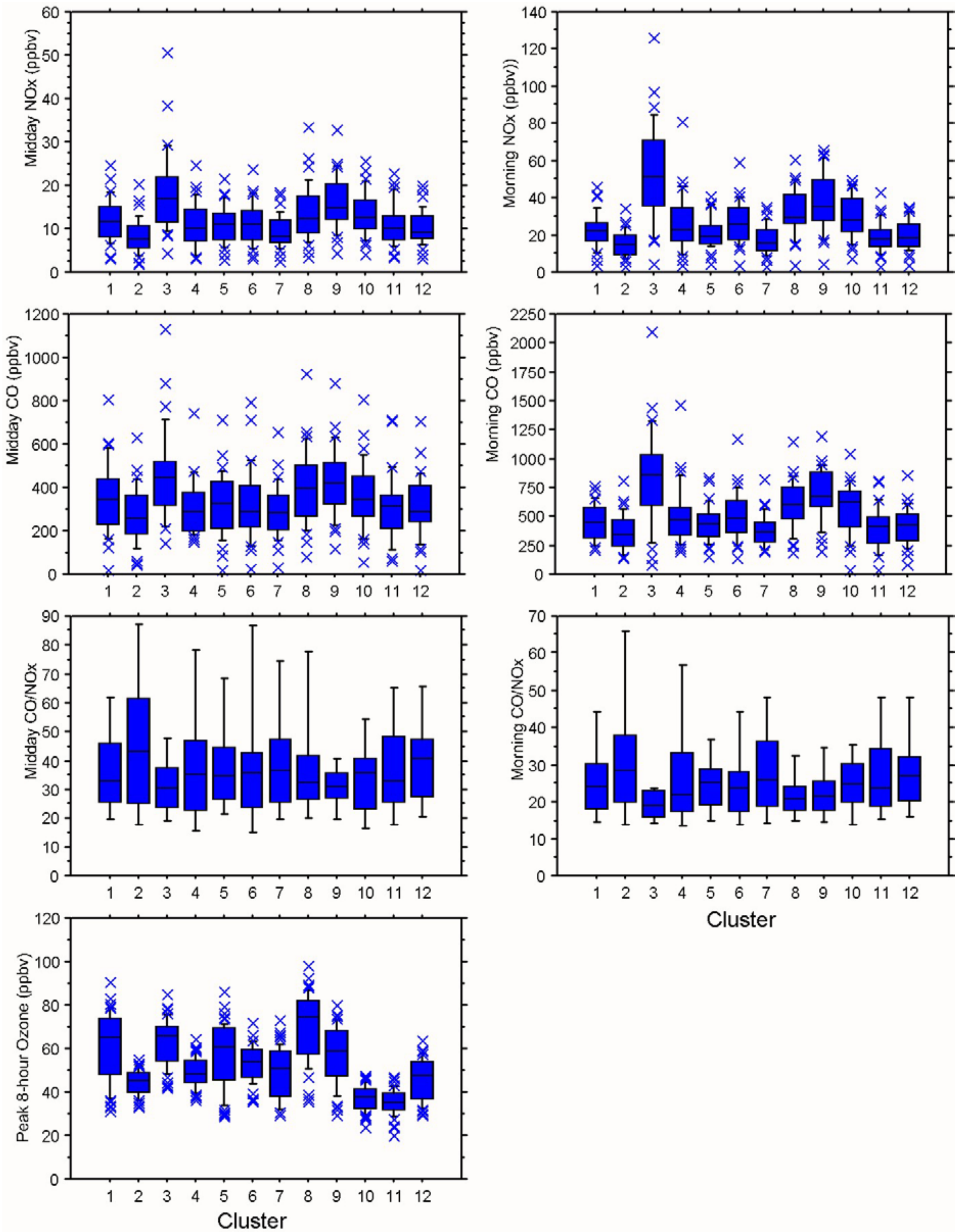

Figure 8. Statistical distributions of NO_x, CO, and CO/NO_x at the time of the peak 8-hour ozone (midday) and from 5 to 10 am (morning), and distributions of peak 8-hour ozone. The data consist of mean values disaggregated by cluster and determined for each monitoring location. The statistical distributions show the ranges of the sites' mean values.

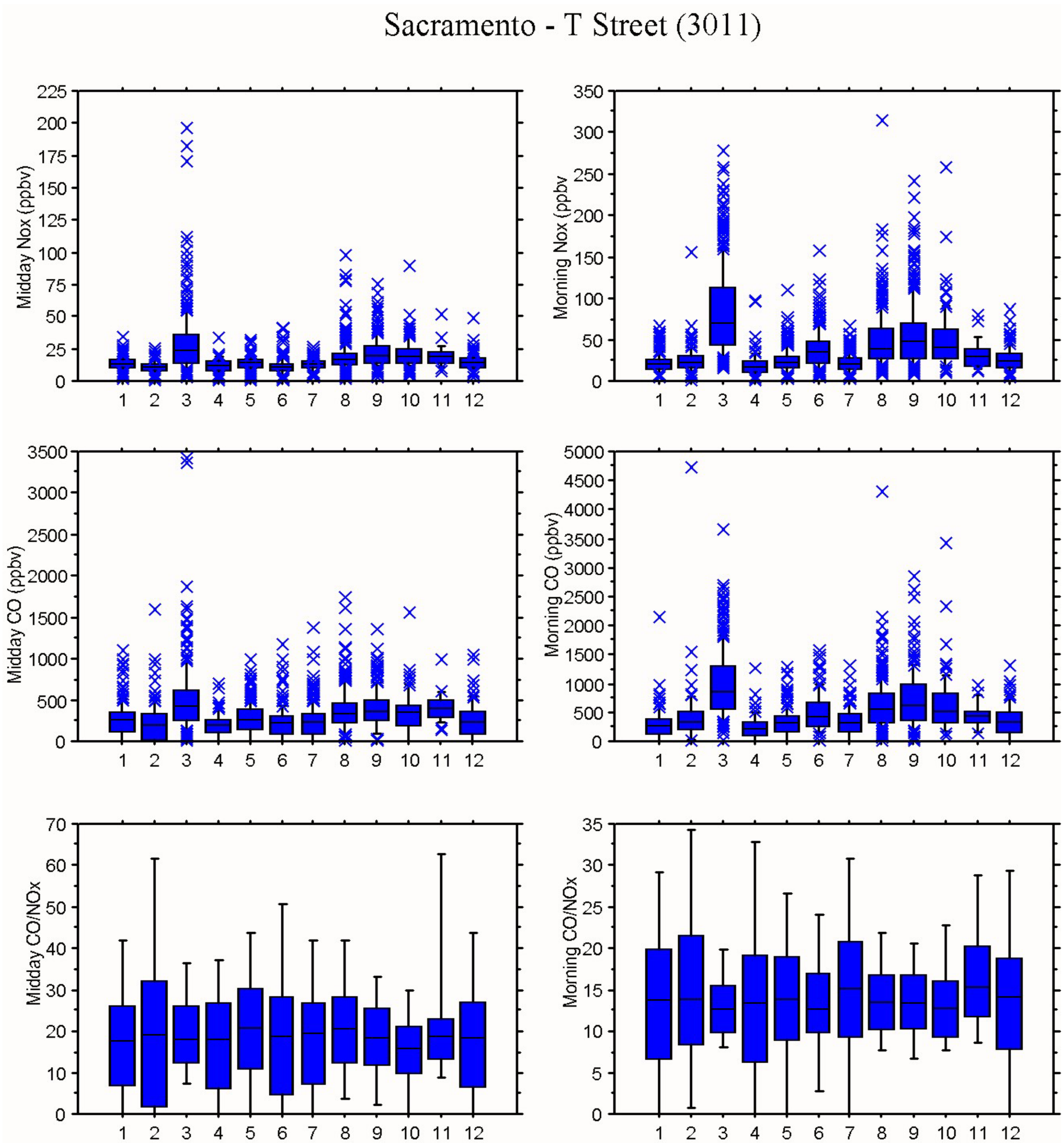

Figure 9. Statistical distributions of NO_x, CO, and CO/NO_x at the time of the peak 8-hour ozone (midday) and from 5 to 10 am (morning). The data consist of Sacramento T-Street daily values disaggregated by cluster.

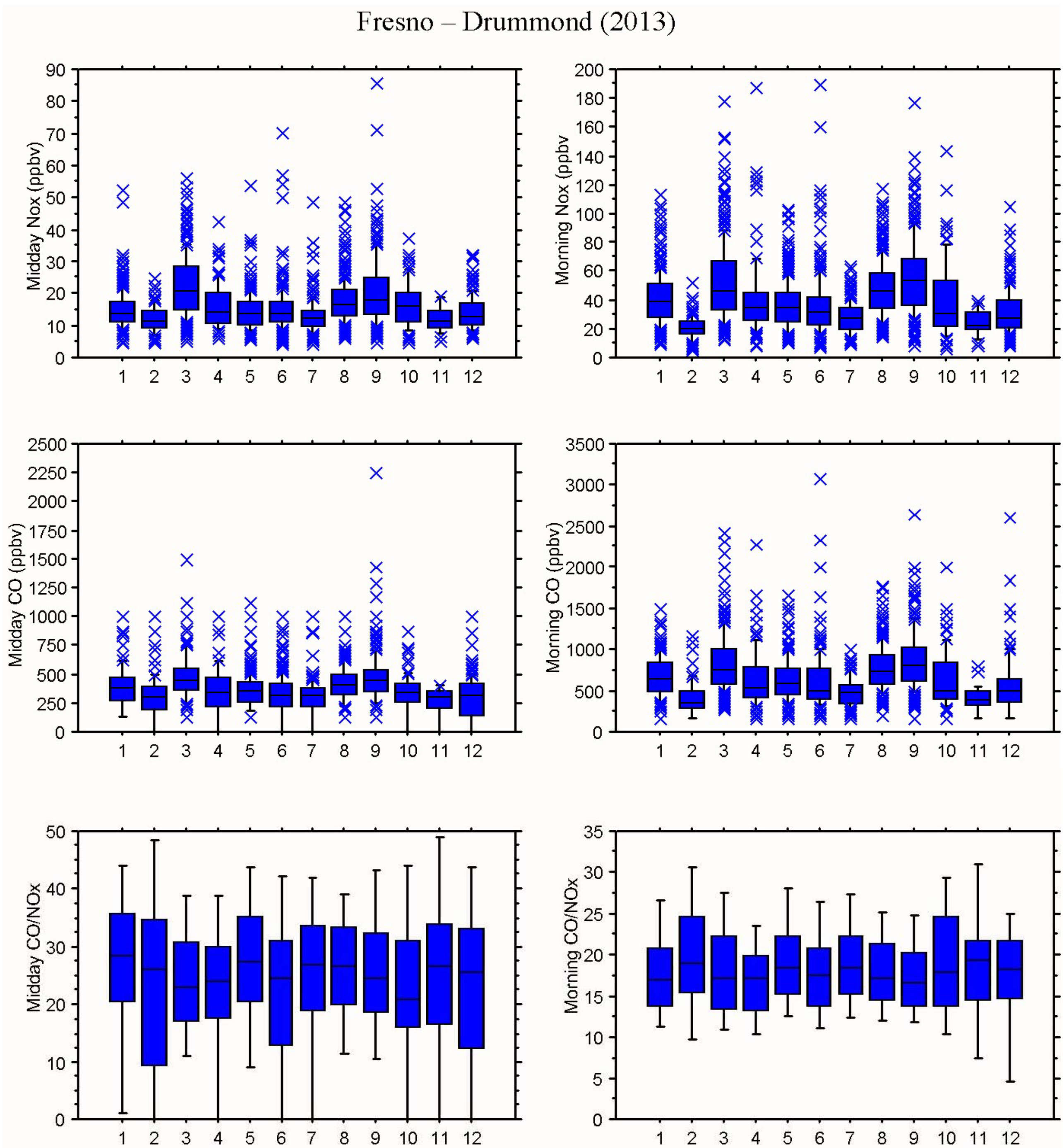

Figure 10. Statistical distributions of NO_x, CO, and CO/NO_x at the time of the peak 8-hour ozone (midday) and from 5 to 10 am (morning). The data consist of Fresno – Drummond Street daily values disaggregated by cluster.

IV. TRENDS

Ozone and Precursor Metrics

We identified the top 60 peak 8-hour ozone days during each year within each subregion. We then compiled several air quality metrics from hourly measurements made on the top 60 days. A two-step process was followed:

1. For each of the top 60 days, we compiled the average 8-hour ozone maxima at each site. We also compiled average concentrations of ozone precursors (NO, NO_x, and CO) from (a) measurements with start hours of 5 through 10 am and (b) the hours making up each day's peak 8-hour ozone maximum at each site.
2. For each site, we determined an average annual concentration of ozone (8-hour maxima) and ozone precursors (5 – 10 am and times of ozone maxima). For simplicity, we refer to the precursor metrics as morning and mid-day averages, respectively.

In addition to these metrics, we also determined the annual 4th-highest 8-hour ozone maximum at each site.

Simple Trends

We determined trends in the annual-average ozone and ozone precursor metrics using least-squares linear regression and t-tests of statistical significance of regression slopes (Tables 2 and 3).[5] Key observations about the observed trends are:

- Trends in mid-day precursor concentrations were similar to trends in morning precursor concentrations, but with typically smaller rates of change (Figures 11 and 12).
- Trends in top 60 average peak 8-hour ozone concentrations were similar to trends in annual 4th-highest daily ozone maxima, but with typically smaller rates of change (Figure 13).
- Morning or mid-day CO trends were statistically significant (downward) at 20 of 24 sites.
- Morning or mid-day NO_x trends were statistically significant (downward) at 22 of 25 sites.

---

[5] We did not combine data from relocated sites. Some increase in the lengths of site records may be possible.

The first observation is important because it signifies that decreasing concentrations of ozone precursors were observed not only in the morning, when ambient concentrations are expected to be dominated by fresh, local emissions in many locations, but also throughout the daytime hours when ozone formation occurs.

The second observation is important because it implies that trends in the ozone metric that we will use as the basis for later analyses (average of top 60 peak daily concentrations) parallel the trends in a regulatory statistic (annual 4^{th}-highest daily 8-hour maximum).

The last two observations show that precursor reductions were large enough to be statistically significant at most sites, and that these reductions were geographically widespread. Nonetheless, a few sites did not exhibit significant downward trends, indicating that possible changes in local influences should be investigated especially at such sites.

Table 1. Summary of data availability for designated sites, 1990-2004 (75 percent daily and annual completeness).

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
MC	Northern Sierra Foothills	Grass Valley	3126	1994-2004	10										
		Placerville	3017	1993-2004	11									1993-2004	12
	Southern Sierra Foothills	Five Mile	3164	1996-2000	4										
		Jackson	2993	1993-2004	12									1993-2004	12
		San Andreas	3144	1995-2004	10									1995-2004	10
		Sonora	2968	1993-2004	12									1993-2004	12
		Yosemite Wawona	2183	1991-1996	5										
		Yosemite Camp Mather	2949	1992-1994	3										
		Yosemite Turtleback	3018	1993-2004	12										
		Yosemite HQ	3045	1990-1994	5										
		Yosemite Merced River	3664	2004-2004	1	2004-2004	1			2004-2004	1			2004-2004	1
SFB	Northern Bay Area	Bethel Is.	2804	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Concord	2831	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Pittsburgh	2102	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Vallejo	2410	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
	Southern Bay Area	Fremont	2293	1990-2004	15	1990-2004	15	1990-2004	15	1990-1993	3			1990-2004	15
		Gilroy	2320	1990-1997	8									1990-1993	4

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
		Los Gatos	2613	1990-2004	12										
		San Jose 4th St.	2413	1990-2001	12	1990-2001	12	1990-2001	12	1990-1993	3			1990-2001	12
		San Jose Piedmont	2969	1993-1997	5										
		San Martin	3140	1995-1997	3										
	Eastern Bay Area	Livermore Old First St	2372	1990-2000	11	1990-2000	11	1990-2000	11	1990-1993	3			1990-1999	10
		Livermore Rincon	3490	2000-2004	5	2000-2004	5	2000-2004	5					2000-2004	5
SJV	Northern San Joaquin	Merced	3022	1992-2004	13	1992-2004	11	1992-2004	11	1992-2004	11				
		Modesto	2833	1990-2004	14	1990-2004	14	1990-2004	14	1990-2004	14			1990-2004	14
		Stockton Hazleton	2094	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	15
		Stockton Mariposa	2553	1990-2000	11										
		Tracy	3159	1996-2003	8	1996-2003	8	1996-2003	8	1996-2003	8				
		Turlock	2996	1993-2004	12	1993-2004	12	1993-2004	12	1993-2004	12			1993-2004	12
	Central San Joaquin	Clovis	3026	1991-2004	14	1991-2004	14	1991-2004	14	1991-2004	14	2003-2004	2	1991-2004	14
		Fresno 1st St.	3009	1991-2004	14	1990-2004	15	1990-2004	15	1990-2004	15	2003-2004	2	1990-2004	15
		Fresno Drummond	2013	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	15
		Fresno Sierra Skypark #2	2844	1992-2004	13	1991-2004	14	1991-2004	14	1991-2004	14			1990-2004	15
		Madera Health Dept.	2591	1991-1996	6										
		Madera Pump Yard	3211	1998-2004	7	1998-2004	7	1998-2004	7	1998-2004	7	2004-2004	1		

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
		Parlier	2114	1990-2004	15	1994-2004	11	1994-2004	11	1994-2004	11	2003-2004	2	1990-1993	4
		Visalia	2032	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	15
	Southern San Joaquin	Arvin	2941	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15	2003-2004	2		
		Bakersfield Golden State Hwy	3145	1995-2004	10	1995-2003	8	1995-2003	9	1995-2003	9	2003-2004	2	1995-2004	9
		Bakersfield 5558 CA Ave	3146	1995-2004	10	1995-2004	10	1995-2004	10	1995-2004	10			1995-2004	10
		Edison	2312	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-1993	4
		Maricopa	2919	1990-2004	14										
		Oildale	2772	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-1993	4
SV	Northern Sierra Foothills	Auburn	2891	1990-2004	9										
	Urban/ Suburban Sacramento	Elk Grove	2977	1993-2004	12	1993-2004	10	1993-2004	10	1993-2004	10				
		Folsom	2472	1990-1996	7	1990-1995	6	1990-1995	6	1990-1995	6				
		Folsom	3187	1997-2004	8	1997-2004	7	1997-2004	7	1997-2004	7	2003-2004	2		
		North Highlands	2123	1990-2004	15	1990-2004	12	1990-2004	12	1990-2004	12			1990-2004	15
		Rocklin	2656	1990-1990	1										
		Roseville	2956	1993-2004	12	1993-2004	12	1993-2004	12	1993-2004	12			1993-2004	12
		Sacramento Del Paso Manor	2731	1990-2004	15	1990-2002	12	1990-2002	12	1990-2002	12	2003-2003	1	1990-2004	14

BASIN	Subregion	SITE	CODE	O3		NO		NO2		NOx		NMHC		CO	
				Years	No.	Years	No.	Years	No.	Years	No.	Years	No.	Years	No.
		Sacramento El Camino & Watt	2840											1990-2004	14
		Sacramento T St.	3011	1990-2004	15	1990-2004	15	1990-2004	15	1990-2004	15			1990-2004	14

Table 2. Regression slopes for annual averages of NO, NO_x and CO measurements (see text). Slopes are marked with bold typeface when the significance level is less than 0.05. Calculations are limited to sites with 10 or more years of data.

Region	Site	Slopes (ppbv year^{-1})					
		Hours of Maximum daily 8-hour Ozone			5-10 am		
		NO	NOx	CO	NO	NOx	CO
Eastern San Francisco Bay Area	2372	**0.26**	0.10	**-29.42**	0.04	-0.35	**-42.98**
	Median	0.26	0.10	-29.42	0.04	-0.35	-42.98
Northern San Francisco Bay Area	2102	-0.05	**-0.68**	**-20.76**	**-0.49**	**-1.09**	**-33.81**
	2410	0.01	**-0.73**	**-45.60**	**-0.80**	**-1.26**	**-52.72**
	2804	-0.08	**-0.71**	6.23	**-0.50**	**-1.11**	0.19
	2831	**-0.08**	**-0.90**	**-61.93**	**-1.10**	**-1.91**	**-84.60**
	Median	-0.06	-0.72	-33.18	-0.65	-1.18	-43.27
Southern San Francisco Bay Area	2293	-0.02	**-0.74**	**-47.61**	-0.45	-0.85	**-47.42**
	2413	**-0.26**	**-1.10**	**-55.75**	-0.31	-0.58	**-45.31**
	Median	-0.14	-0.92	-51.68	-0.38	-0.72	-46.36
Northern San Joaquin Valley	2094	-0.20	**-0.73**	**-35.82**	-0.52	**-1.05**	**-48.54**
	2833	**-0.50**	**-1.30**	**-36.27**	**-0.73**	**-1.45**	**-42.38**
	2996	**-0.23**	**-0.61**	**-26.18**	**-0.53**	**-0.94**	**-36.86**
	3022	-0.06	**-0.59**		-0.11	-0.47	
	3159	-0.18	**-0.53**		-0.70	-1.21	
	Median	-0.20	-0.61	-35.82	-0.53	-1.05	-42.38
Central San Joaquin Valley	2013	-0.11	**-0.70**	-5.57	**-0.55**	**-0.95**	**-27.72**
	2032	0.05	-0.20	-11.39	0.00	-0.26	**-32.47**
	2114	0.01	**-0.22**		-0.08	-0.39	
	2844	-0.04	**-0.43**	**-20.68**	-0.19	**-0.72**	**-45.64**
	3009	-0.05	**-0.32**	-20.48	**-0.41**	-0.66	**-42.97**
	3026	0.02	**-0.34**	**-27.03**	**-0.34**	**-0.86**	**-48.24**
	Median	-0.01	-0.33	-20.48	-0.26	-0.69	-42.97
Southern San Joaquin Valley	2312	**-0.27**	**-0.44**		**-0.29**	-0.36	
	2772	-0.09	**-0.50**		**-0.31**	**-0.80**	
	2941	0.00	**-0.31**		**-0.20**	**-0.45**	
	3145	**-0.09**	**-0.91**	8.52	-0.43	**-1.46**	-6.12
	3146	-0.10	-0.23	-13.73	-0.33	-0.60	**-41.10**
	Median	-0.09	-0.44	-2.60	-0.31	-0.60	-23.61
Sacramento Valley	2123	-0.16	-0.17	**-42.68**	-0.17	-0.26	-44.23
	2731	-0.03	-1.07	-19.49	-0.67	-1.54	-22.63
	2956	**0.15**	**-0.27**	**-16.12**	-0.17	-0.57	**-16.10**
	2977	0.01	0.08		0.04	0.18	
	3011	0.09	**-0.46**	0.41	-0.26	-0.68	**-26.91**
	Median	0.01	-0.27	-17.80	-0.17	-0.57	-24.77
Northern Sierra Foothills	3017			3.54			-3.61
	Median			3.54			-3.61
Southern Sierra Foothills	2968			-0.45			**-30.40**
	2993			-9.47			**-19.43**
	3144			1.87			1.13
	Median			-0.45			-19.43

Table 3. Regression slopes of 4th highest daily 8-hour maximum and annual average of top 60 8-hour maximum over time for the period 1990-2004. Slopes are in bold typeface when significance level is less than 0.05. Calculations are limited to sites with 10 or more years of data.

Region	**Site**	**Slope (ppbv year^{-1})**	
		4th Highest 8-hour maximum	**Top 60 8-hour maximum**
Eastern San Francisco Bay Area	2372	0.21	0.02
	Median	0.21	0.02
Northern San Francisco Bay Area	2102	-0.17	-0.03
	2410	0.03	-0.08
	2804	-0.33	-0.15
	2831	0.3	0.22
	Median	-0.07	-0.05
Southern San Francisco Bay Area	2293	-0.71	-0.23
	2320	-0.21	-0.3
	2413	-1.3	-0.53
	2613	-0.62	-0.09
	2969	**-1.65**	-0.96
	3140	-0.57	-0.52
	Median	-0.67	-0.41
Northern San Joaquin Valley	2094	**-0.69**	**-0.41**
	2553	**-0.71**	**-0.97**
	2833	-0.29	-0.18
	2996	-0.29	-0.11
	3022	0.33	0.22
	3159	-1.35	-0.63
	Median	-0.49	-0.3
Central San Joaquin Valley	2013	0.17	0.35
	2032	-0.45	-0.17
	2114	0.1	0.59
	2844	0	0.27
	3009	-0.58	-0.38
	3026	-0.03	-0.1
	Median	-0.02	0.08
Southern San Joaquin Valley	2312	-0.63	-0.55
	2772	0.07	0.24
	2919	-0.31	-0.05
	2941	0	0.11
	3145	-0.53	-0.22
	3146	-0.72	-0.57
	Median	-0.42	-0.13
Sacramento Valley	2123	0.13	0.13
	2731	-0.04	-0.1
	2956	-1.17	-0.18

Region	Site	Slope (ppbv year^{-1})	
		4th Highest 8-hour maximum	**Top 60 8-hour maximum**
	2977	-0.32	-0.13
	3011	-0.2	-0.24
	Median	-0.2	-0.13
Northern Sierra Foothills	2891	-0.82	**-0.68**
	3017	-0.57	**-0.56**
	3126	1.13	0.7
	Median	-0.57	-0.56
Southern Sierra Foothills	2968	-0.38	-0.2
	2993	-0.61	-0.5
	3018	-0.3	-0.36
	3144	-0.79	**-1.03**
	3161	**-1.06**	-0.54
	Median	-0.61	-0.5

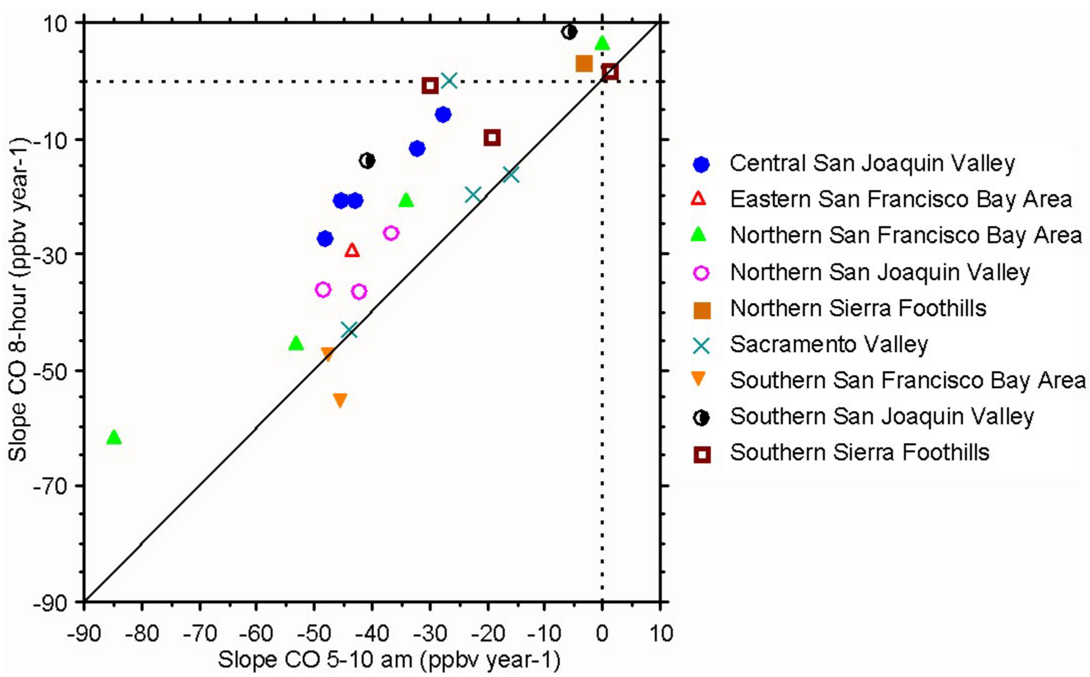

Figure 11. Comparison of mid-day and morning CO trends. Trends are expressed as rates of change in ppbv year^{-1}, with negative values representing downward trends.

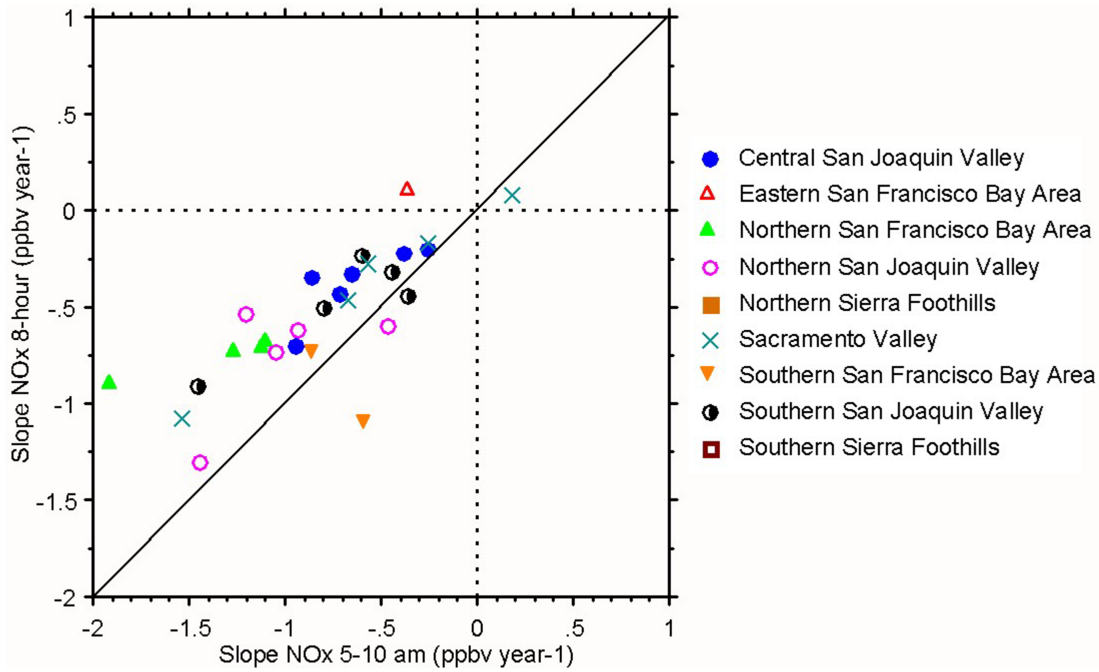

Figure 12. Comparison of mid-day and morning NO_x trends. Trends are expressed as rates of change in ppbv year^{-1}, with negative values representing downward trends.

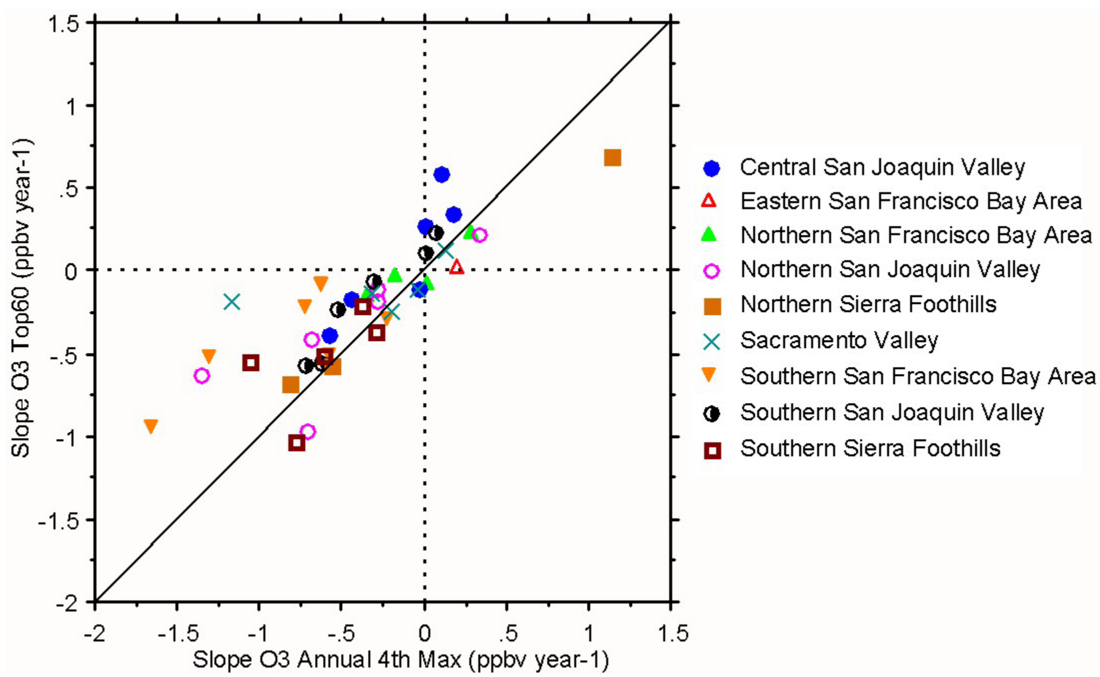

Figure 13. Comparison of trends in ozone metrics: annual-average top 60 and annual 4th-highest daily 8-hour maximum. Trends are expressed as rates of change in ppbv year^{-1}, with negative values representing downward trends.

V. CONCLUSION

We carried out a preliminary meteorological classification of ozone-season (May 1 to October 31) days during 1990 through 2004. The statistical method was K-means clustering using daily-average pressure gradients from San Francisco to Medford, Reno, and Fresno, 850 mb temperatures, heights, wind speed, and wind direction at 4 am and 4 pm at Oakland, and surface wind speed and direction at nine monitoring locations throughout central California. We intend to refine the classification procedure using additional surface stations from among the 53 sites for which we have at least ten years data. Additional evaluation is needed to test the effects of missing data on the classifications.

For our initial clustering evaluation, we specified 12 clusters. Our expectation was that about half would prove to be populated largely by days having lower ozone levels, and therefore would not be of interest. Our objective was to obtain approximately 4 to 6 clusters representing different transport directions and/or distances.

We obtained five clusters that had higher mean peak 8-hour ozone levels than did other clusters. These higher-ozone clusters exhibited different mean concentrations of CO and NO_x. Although the clustering analysis is preliminary and subject to revision, it successfully identified the target number of high-ozone clusters (i.e., 4 to 6), and it identified high-ozone clusters having different levels of ambient ozone precursor concentrations. Three of the clusters occurred most frequently in July and August, while the remaining two clusters occurred most frequently in September and October. The distributions of the cluster frequencies were more complex than a simple seasonal division, though. Mean transport distances and directions exhibit differing degrees of variation among the high-ozone clusters, depending upon the monitoring location. One objective of a second, more refined clustering analysis will be to obtain high-ozone clusters with greater between-cluster variations of transport distances and directions.

As a first step in trends analysis, we identified the top 60 peak 8-hour ozone days during each year within each of nine subregions of the CCOS domain. We then compiled several air quality metrics from hourly measurements made on the top 60 days.

Trends in mid-day precursor concentrations were similar to trends in morning precursor concentrations, but with typically smaller rates of change. This observation is important because it means that decreasing concentrations of ozone precursors were observed not only in the morning, when ambient concentrations are expected to be dominated by fresh, local emissions in many locations, but also throughout the daytime hours when ozone formation occurs.

Trends in top 60 average peak 8-hour ozone concentrations were similar to trends in annual 4th-highest daily ozone maxima, but with typically smaller rates of change. This second observation is important because it implies that trends in the average of top 60 peak daily concentrations parallel the trends in a regulatory statistic (annual 4th-highest daily 8-hour maximum). As a result, we may conclude that conclusions based on analyses of top 60 days are relevant to regulatory metrics.

Morning or mid-day CO trends were statistically significant (downward) at 20 of 24 sites. Morning or mid-day NO_x trends were statistically significant (downward) at 22 of 25 sites. These last two observations show that ozone precursor reductions were large enough to be statistically significant at most sites, and that these reductions were geographically widespread. A few sites did not exhibit significant downward trends, suggesting the possibility changes of countervailing local influences.

The next steps in this project are to revise and refine the meteorological classifications and to determine trends in ozone and ozone precursor metrics within each meteorological class.