CARB LIBRARY
14992

Sonoma Technology, Inc.

1360 Redwood Way, Suite C
Petaluma, CA 94954-1169
707 665-9900
Fax 707 665-9800
www.sonomatech.com

SPATIAL AND TEMPORAL CHARACTERISTICS OF CALIFORNIA PAMS AND LONG-TERM TREND SITE VOC DATA (1990-1997)

FINAL REPORT
STI-998241-1883-FR

By:
Hilary H. Main
Siana B. Hurwitt
Paul T. Roberts
Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169

Prepared for:
Mark Schmidt
U.S. Environmental Protection Agency
OAQPS/EMAD MD-14
Research Triangle Park, NC 27711

May 17, 1999

TD
885.5
V5
M3
1999

TABLE OF CONTENTS

Section	Page
LIST OF FIGURES	v
LIST OF TABLES	ix
1. INTRODUCTION	1-1
1.1 BACKGROUND	1-1
1.2 STUDY OBJECTIVES	1-1
1.3 GUIDE TO THIS REPORT	1-2
2. DATA VALIDATION	2-1
2.1 PAMS PROGRAM	2-1
2.2 DATA QUALITY OVERVIEW	2-1
2.3 DATA VALIDATION PROCEDURE	2-2
2.4 DATA VALIDATION SUMMARY	2-3
2.5 CONTINUOUS NMHC DATA	2-4
2.6 USE OF ARB AUDIT DATA	2-4
3. DATA ANALYSIS RESULTS	3-1
3.1 SPATIAL, TEMPORAL, AND STATISTICAL DISTRIBUTION OF THE PAMS MEASUREMENTS IN CALIFORNIA	3-1
3.1.1 Overall Characteristics	3-1
3.1.2 Abundant Hydrocarbons	3-3
3.1.3 Fingerprints	3-4
3.2 RELATIONSHIPS AMONG SPECIES	3-9
3.2.1 Changes in Benzene	3-9
3.2.2 Scatter Plot Matrices	3-10
3.2.3 Correlation Matrices	3-11
3.2.4 Cluster Analyses	3-11
3.2.5 Factor Analyses	3-12
3.3 ASSESSING THE AGE OF AN AIR MASS	3-13
3.4 NMHC/NO_x	3-15
3.5 DAY-OF-WEEK ISSUES	3-16
3.6 DISCUSSION OF SAMPLE SIZE	3-17
3.6.1 Addressing the PAMS Data Quality Objectives	3-17
3.6.2 Other Considerations	3-18
3.6.3 Summary	3-19
3.7 CONTINUOUS NMHC DATA	3-19
3.8 SOURCE APPORTIONMENT	3-20
3.8.1 Overview	3-20
3.8.2 Uncertainties	3-21
3.8.3 Model Limitations	3-22

TABLE OF CONTENTS (Concluded)

Section	Page
4. CONCLUSIONS AND RECOMMENDATIONS	4-1
4.1 EMISSION SOURCE ISSUES	4-1
4.2 DIURNAL VARIATION IN THE DATA	4-2
4.3 YEAR-TO-YEAR VARIATION IN THE DATA	4-4
4.4 SITING CHARACTERISTICS	4-4
4.5 ASSESSING VOC VERSUS NO_x CONTROLS	4-5
4.6 DAY-OF-WEEK ISSUES	4-5
4.7 CONTINUOUS NMHC DATA	4-6
4.8 SAMPLE SIZE ISSUES	4-7
4.9 DATA VALIDATION AND MEASUREMENT ISSUES	4-8
5. REFERENCES	5-1
APPENDIX A: LIST OF FLAGGED SAMPLES FOR THE CALIFORNIA PAMS AND TRENDS SITES, 1990-1997	A-1
APPENDIX B: HYDROCARBON ABBREVIATIONS USED IN THIS REPORT	B-1
APPENDIX C: SUMMARY OF COEFFICIENT OF VARIATION SCREENING OF 1997 DATA BY SITE	C-1
APPENDIX D: LINEAR REGRESSION RESULTS FOR CLOVIS 1995-1997 CONTINUOUS NMHC AND PAMS NMHC CONCENTRATIONS	D-1
APPENDIX E: AMBIENT DATA FILES FOR CMB8.0	E-1

LIST OF FIGURES

Figure		Page
2-1.	Map of California PAMS and long-term trend sites investigated in this project	2-6
2-2.	Scatter plot of benzene (y-axis) versus acetylene (x-axis) concentrations (ppbC) at a San Diego County PAMS site	2-7
2-3.	Continuous NMHC concentrations (ppbC) at Clovis during 1996	2-8
2-4.	Continuous NMHC concentrations (ppbC) at Clovis during 1997	2-9
3-1.	Annotated box-whisker plot with outliers	3-23
3-2.	Box plots of total non-methane hydrocarbon concentrations (ppbC) by site and time of day (PST) in 1997	3-24
3-3.	Box plots of total non-methane hydrocarbon concentrations (ppbC) by size and time of day (PST) in 1997	3-25
3-4.	Box plots of total non-methane hydrocarbon concentrations (ppbC) by site and time of day (PST) in 1997 (1996 data used for El Rio)	3-26
3-5.	Median composition at Del Paso Manor in 1997 at 0500, 1200, 1600, and 2300 PST	3-27
3-6.	Median composition at Parlier in 1997 at 0500, 1200, 1600, and 2300 PST	3-28
3-7.	Median composition at Arvin in 1997 at 0500, 1200, 1600, and 2300 PST	3-29
3-8.	Median composition at Los Angeles in 1997 at 0500 and 1200 PST	3-30
3-9.	Median composition at San Diego in 1997 at 0500 and 1200 PST	3-31
3-10.	Median composition at Clovis in 1997 at 0500, 1200, 1600, and 2300 PST	3-32
3-11.	Median composition at Bakersfield (Golden State) in 1997 at 0500, 1200, 1600, and 2300 PST	3-33
3-12.	Median composition at El Rio in 1996 at 0200, 0500, 0800, 1200, and 1600 PST	3-34
3-13.	Median composition at Fresno in 1997 at 0500, 1200, 1600, and 2300 PST	3-35
3-14.	Median composition at Elk Grove in 1997 at 0500, 1200, 1600, and 2300 PST	3-36

LIST OF FIGURES (Continued)

Figure **Page**

3-15. Median composition at New Folsom in 1997 at 0500, 1200, 1600, and 2300 PST... 3-37

3-16. Median composition at Simi Valley in 1997 at 0200, 0500, 1200, and 1600 PST.... 3-38

3-17. Median composition at Emma Wood in 1997 at 0200, 0500, 1200, and 1600 PST.. 3-39

3-18. Median composition at Los Angeles from 1992 through 1997 normalized over the identified hydrocarbons 3-40

3-19. Median composition at San Diego from 1992 through 1997 normalized over the identified hydrocarbons 3-41

3-20 Median composition at Fresno from 1992 through 1997 normalized over the identified hydrocarbons 3-42

3-21. Median composition at Clovis from 1993 through 1997 normalized over the identified hydrocarbons 3-43

3-22. Median composition at Bakersfield (Golden State) from 1995 through 1997 normalized over the identified hydrocarbons 3-44

3-23. Median composition at Del Paso Manor from 1994 through 1997 normalized over the identified hydrocarbons 3-45

3-24. Median composition at Parlier from 1995 through 1997 normalized over the identified hydrocarbons 3-46

3-25. Median composition at Arvin from 1995 through 1997 normalized over the identified hydrocarbons 3-47

3-26. Median composition at Simi Valley from 1995 through 1997 normalized over the identified hydrocarbons 3-48

3-27. Scatter plots of benzene to toluene concentrations (ppbC) by year at a) Del Paso Manor (1994-1997) and b) Clovis (1995-1997) 3-49

3-28. Scatter plots of benzene to toluene concentrations (ppbC) by year at a) Fresno (1994-1997), b) Los Angeles (1994-1997) and c) Bakersfield (1994-1997)..... 3-50

3-29. Scatter plots of benzene to toluene concentrations (ppbC) by year at a) San Diego (1994-1997) and b) Simi Valley (1995-1997)..... 3-51

LIST OF FIGURES (Continued)

Figure		Page
3-30.	Scatter plot matrix of ten abundant hydrocarbons at Del Paso Manor during 1997	3-52
3-31.	Scatter plot matrix of ten most hydrocarbons at Clovis during 1997	3-53
3-32.	Scatter plot matrix of 11 abundant hydrocarbons at Fresno 1st Street during 1997	3-54
3-33.	Scatter plot matrix of ten abundant hydrocarbons at Bakersfield during 1997	3-55
3-34.	Scatter plot matrix of ten abundant hydrocarbons at Los Angeles during 1997	3-56
3-35.	Scatter plot matrix of ten abundant hydrocarbons at San Diego during 1997	3-57
3-36.	Scatter plot matrix of ten abundant hydrocarbons at El Rio during 1996	3-58
3-37.	Scatter plot matrix of ten abundant hydrocarbons at Folsom during 1997	3-59
3-38.	Scatter plot matrix of ten abundant hydrocarbons at Elk Grove during 1997	3-60
3-39.	Scatter plot matrix of ten abundant hydrocarbons at Parlier during 1997	3-61
3-40.	Scatter plot matrix of ten abundant hydrocarbons at Emma Wood during 1997	3-62
3-41.	Scatter plot matrix of nine abundant hydrocarbons at Arvin during 1997	3-63
3-42.	Scatter plot matrix of ten abundant hydrocarbons at Simi Valley during 1997	3-64
3-43.	Cluster analysis results for Clovis in 1997 (morning data only)	3-65
3-44.	Cluster analysis results for Del Paso Manor in 1997 (morning data only)	3-66
3-45.	Cluster analysis results for Bakersfield in 1997 (morning data only)	3-67
3-46.	Cluster analysis results for Elk Grove in 1997 (morning data only)	3-68
3-47.	Factor analysis results for Fresno in 1997 (all data)	3-69
3-48.	Median toluene to benzene ratios by time of day (PST) in 1997 at a) Type 2 and b) Type 1 and 3 sites	3-70
3-49.	Median xylenes to benzene ratios by time of day (PST) in 1997 at a) Type 2 and b) Type 1 and 3 sites	3-71

LIST OF FIGURES (Continued)

Figure		Page
3-50.	Morning (0500-0700 PST) NMHC/NO_X ratios by site in 1997	3-72
3-51.	Annotated notched box-whisker plot with outliers	3-73
3-52.	Notched box plots of morning (0500 PST) toluene (TOLU) and NMHC (shown as TNMOC) concentrations (ppbC) at Bakersfield by day of week in 1997 (Monday = 1, Sunday = 7)	3-74
3-53.	Notched box plots of morning (0500 PST) NMHC (shown as TNMOC), toluene (TOLU), acetylene (ACETY), and total xylenes concentrations (ppbC) at San Diego by day of week in 1997 (Monday = 1, Sunday = 7)	3-75
3-54.	Results from a two-sample t-test on morning benzene concentrations grouped by year at Los Angeles	3-76
3-55.	Illustration of the effect of sample size on the 95 percent confidence intervals of the Los Angeles morning benzene concentrations in 1995	3-77
3-56.	Box plots of the benzene concentration (ppbC) data shown in Figure 3-55	3-78
3-57.	Mean NMHC concentrations and confidence intervals (CI) for the hourly and PAMS data collected at Clovis from 1995 through 1997	3-79

LIST OF TABLES

Table		Page
2-1.	Site information for PAMS and long-term trends sites in four California districts from 1990 through 1997	2-10
2-2.	Hydrocarbon and carbonyl compound data available at PAMS and long-term trends sites in four California districts from 1990 through 1997	2-11
2-3.	Number (n) of valid hydrocarbon samples available at PAMS and long-term trends sites in four California districts from 1990 through 1997	2-12
3-1.	Median NMHC concentrations at PAMS and long-term trends sites in four California districts from 1990 through 1997	3-80
3-2.	Median values based on 1997 valid data collected during all time periods	3-80
3-3.	Median values of each species group based on 1997 valid data collected during all time periods	3-81
3-4.	Ten most abundant hydrocarbons on a weight percent basis at the California PAMS sites in 1997	3-82
3-5.	Ten most abundant hydrocarbons on a reactivity weighted basis at the California PAMS sites in 1997	3-83
3-6.	Key species and their major sources	3-84
3-7.	Summary of hydrocarbons that exhibited a coefficient of variation greater than 50 percent for 1997 weight fraction data	3-85
3-8.	Selected hydrocarbon ratios measured at western urban U.S. tunnels	3-86
3-9.	Median 0500-0800 PST ambient hydrocarbon ratios at California PAMS sites in 1997	3-86
3-10.	Threshold values of VOC- or NO_x-limited indicators (afternoon concentrations or ratios)	3-87
3-11.	Summary statistics for Del Paso Manor in 1995	3-87
3-12.	Summary statistics for hourly NMHC and PAMS concentrations collected during 0500-0700 PST, 1995-1997 at Clovis	3-88

1. INTRODUCTION

1.1 BACKGROUND

In accordance with the 1990 Clean Air Act Amendments, the U.S. Environmental Protection Agency (EPA) initiated the Photochemical Assessment Monitoring Stations (PAMS) program for serious, severe, and extreme ozone nonattainment areas. The PAMS networks monitor for volatile organic compounds (VOCs), ozone, oxides of nitrogen (NO_x), and meteorological parameters. The PAMS networks were designed to provide data for the assessment of population exposure, ozone formation, and evaluation of ozone control strategies. The EPA, California Air Resources Board (ARB), and the Sacramento, San Joaquin Valley, and Ventura Districts sponsored research into the validation and analysis of the VOC data collected at the Districts' PAMS sites and the ARB's long-term trend sites located in Fresno, Los Angeles, and San Diego. This report summarizes VOC data analyses performed for these entities.

1.2 STUDY OBJECTIVES

This project was comprised of five tasks:

- Task 1: Provide technical support for VOCDat software. Support has included minor modifications to the software and assisting users via telephone and e-mail.
- Tasks 2 and 3: Conduct two VOCDat demonstration/training sessions (one in Research Triangle Park, North Carolina and one in Sacramento, California). These sessions were carried out in May and June of 1998, respectively.
- Task 4: Validate 1990-1997 PAMS and ARB long-term trend data and provide feedback to reporting agencies. This task was documented separately by Main (1998).
- Task 5: Investigate the PAMS and long-term trend VOC data; this is the focus of this report.

We began our investigation of the VOC data with a description and display of the spatial and temporal patterns of the data. Next, we performed interpretative analyses that included assessing the ozone formation potential, investigating the relative age of an air mass, and performing multivariate analysis of the data to estimate possible VOC sources. We performed the following tasks in this regard:

- Assessed general characteristics of VOCs including the importance of biogenic versus anthropogenic hydrocarbons.
- Assessed spatial and temporal issues including comparisons among PAMS site types.
- Assessed air mass age using ratio techniques.
- Reviewed day of week distributions and composition differences.
- Investigated source identification (qualitative).

1.3 GUIDE TO THIS REPORT

The remainder of this report provides a summary of the data used in the analyses (Section 2), a discussion of data analysis results (Section 3), and a summary of conclusions and recommendations for further work (Section 4). References are provided in Section 5 and a list of samples that were flagged during data validation efforts is provided in Appendix A. Supporting information for the analyses in this report is provided in Appendices B, C, D, and E.

2. DATA VALIDATION

We used VOCDat software to quality assure 1990 through 1997 PAMS (and PAMS-like) VOC data for four California air districts: Sacramento, San Joaquin Valley, San Diego, and Ventura. We summarized our validation efforts in a technical memorandum (Main, 1998). The Districts and ARB reviewed the data that we flagged as "odd". This section describes the data used in the analyses described in this report and provides the final results of the validation efforts.

2.1 PAMS PROGRAM

There are up to four types of PAMS sites in an area. The site type is based on its location relative to the emissions and transport pathways in an area, defined as follows:

- Type 1 - located to provide upwind and background characterization of ozone and precursors being transported into an area.
- Type 2 - located to document the maximum ozone precursor emissions impact. These sites are typically located downwind of the central business district and operate on the most intensive monitoring schedule of the PAMS sites.
- Type 3 - located to measure the maximum ozone concentration and typically situated further downwind than the Type 2 sites.
- Type 4 - located downwind of the nonattainment areas to assess the extreme downwind conditions. In some parts of the country, a Type 4 PAMS site may also be defined as a Type 1 site for another non-attainment area well downwind.

In California, most PAMS sites employ canisters for data collection. PAMS sites in Ventura County employ an automatic gas chromatography system (auto-GC) that reports hydrocarbon concentrations of about 60 target hydrocarbon species and the total nonmethane hydrocarbon (NMHC). Also at most sites in California, two to five 3-hr samples are collected every third day. Sampling details for the sites considered in this project are provided in **Table 2-1**. Carbonyl compounds are collected at the PAMS sites using commercially available dinitrophenylhydrazine (DNPH)-impregnated silica gel cartridges; however, results from most of the PAMS sites are reported only for formaldehyde, acetaldehyde, and acetone.

2.2 DATA QUALITY OVERVIEW

In general, the data were of good quality and reflect the hard work of the monitoring and analytical staff at the ARB and Districts. Problems typical of PAMS data sets from other states were not commonly observed including widespread peak misidentification, calibration gas carryover, and calibration data left in the database as ambient data. The District and ARB staff addressed a few minor problems in the database that we noted in the memorandum. The only remaining serious issue with the data is the quality of the acetylene data, particularly prior

to 1997. Laboratory staff meet annually to discuss improvements in techniques and are actively addressing this aspect of the data as well. The problems with acetylene may reduce the applicability of some analyses such as source apportionment.

Table 2-2 lists the sites, years, and dates for which we received data from the ARB[1]. The sites are shown in **Figure 2-1**. We performed data validation procedures on all the data listed in the tables. Subsequent data analyses were not performed on the San Diego County PAMS site data because the County chose not to participate in the project.

2.3 DATA VALIDATION PROCEDURE

After processing and investigating the supplied database (see Main, 1998), we then performed several validation steps. Our process of validation is to catch any remaining data not representative of ambient air conditions before we perform our analyses. To this end, we used VOCDat software (Main et al., 1998a) to perform the following:

- Applied automated screening checks for abundant hydrocarbons, hydrocarbon relationships, and outliers. We then inspected samples that failed the screening checks.
- Inspected time series and scatter plots of every hydrocarbon and hydrocarbon functional group (i.e., aromatics, paraffins, olefins, and unidentified). For example, we prepared and inspected scatter plots of each hydrocarbon against the total NMHC.
- Inspected every "fingerprint" for all sites. A fingerprint shows the concentration or weight percent of each hydrocarbon in a sample in the order in which the hydrocarbons elute from the chromatograph. We quickly "paged" through every sample and more closely examined fingerprints for samples that were flagged as suspect and the samples collected before and after the suspect samples.
- Reviewed summary statistics.

During these reviews and inspections, we flagged data that did not meet our conceptual model of ambient hydrocarbon relationships. Note that high (or low) absolute concentrations are not the primary reason for a flag. For example, **Figure 2-2** shows a scatter plot of benzene versus acetylene at a San Diego County PAMS site. Acetylene data collected on October 3-4, 1997 at Alpine, Overland, and El Cajon were anomalous compared to all other data collected, even though the acetylene concentrations were not abnormally high or low when compared to the overall database. However, the figure shows that these data stand out significantly and should be further investigated.

[1] Data were provided by the ARB via an Access database on CD-ROM. We did not compare these data to what is available on the EPA's Aerometric Information Retrieval System (AIRS).

2.4 DATA VALIDATION SUMMARY

The analytical laboratories reviewed the flagged data, corrected a few samples that contained misidentification problems, and commented on the rest of the data. The laboratories verified that most of the data we flagged were analytically correct with "no reasons to doubt the veracity of the data". As analysts, we retained the flags on these data and did not include them in "over-arching" analyses such as trends or summary statistics; however, these data can be used in case studies and other analyses using individual samples. Specific data validation findings are summarized in Appendix A. The tables list the species that were flagged and provide reasons why the data were flagged.

Table 2-3 summarizes the number of valid samples (i.e., samples without flags). We estimated the data completeness by dividing the number of samples collected by the number of samples planned (based on the schedule). In some cases, samples were collected on a schedule of every third day, for example, and on every day during some episodes. In these cases, data completeness may exceed 100 percent.

Key items for data analysts to note about the data sets are summarized below:

- Acetylene data are troublesome and a great deal of effort has been spent in an attempt to improve recovery of this hydrocarbon. Additional information contained in annual PAMS/hydrocarbon quality control (QC) reports should be reviewed.
- Formaldehyde concentrations at Del Paso Manor were extremely high for all samples between August 13, 1997 and September 29, 1997. The District monitoring staff noted that construction, including asphalt pouring, was conducted at the site during this time period. Thus, the formaldehyde concentrations are probably not representative of concentrations across a wider area.
- Decane and undecane concentrations were often relatively high at several sites including Elk Grove, Folsom, Arvin, and Clovis. Since no analytical reasons were discovered, the possibility of nearby sources, such as diesel emissions, should be considered.
- The 1990 data at Fresno are probably not useful for most analyses because of documented problems. In addition, 33 samples in 1992 did not have valid total NMHC values thus reducing the usefulness of these data.
- The 1990 and 1991 data at Los Angeles are probably not useful for most analyses because of documented problems.
- Data collected in Ventura were subjected to rigorous review and extensive reprocessing by the District. We used the reprocessed data "as is" without further review.

2.5 CONTINUOUS NMHC DATA

As a part of the PAMS program, California reports continuous (every hour) NMHC values using method 8202. We obtained these data from the ARB through their CD-ROM. We cursorily inspected the data and found a few issues that could affect subsequent analyses:

- At several sites, resolution of the concentrations was pphmC instead of ppbC.
- At Del Paso Manor, much of the data showed 0 concentrations. Also at this site, the concentrations drop by two orders of magnitude on April 1, 1997, from previous concentrations.
- The sample start hours on the CD-ROM we received were incorrect at all sites. We obtained corrected data.
- During 1995 in the Clovis data, we observed a large number of 0 concentrations. We assume these are below the apparent detection limit of 100 ppbC. In 1996 and 1997, the data were reported in 10 ppbC increments.
- During several time periods in 1996 and 1997 at Clovis, the continuous VOC data show high concentrations that decline slowly with time for the next few hours. In the PAMS database, this usually is indicative of column contamination that cleans out over a few hours. An example is provided in **Figure 2-3**.
- In 1997, the continuous VOC data at Clovis show a period with a higher "baseline" than surrounding periods (**Figure 2-4**). This time period should be inspected.

The continuous data from Clovis were used for only a couple of analyses. Throughout the report NMHC values refer to the PAMS canister 3-hr values unless explicitly stated otherwise. Before these data are placed in AIRS, the continuous NMHC data should be carefully validated.

2.6 USE OF ARB AUDIT DATA

We obtained the ARB laboratory audit data for 1995-1997, through the probe (TTP) audit results for 1997, and laboratory performance audit summaries for 1997 (Miguel, 1998). We reviewed the results and discussed them with ARB staff. These data were useful in illustrating the following:

- Laboratory analyses of individual hydrocarbon concentrations were generally well within ARB specifications of ±20 percent.
- Acetylene and ethane concentrations were often problematic. We discussed the need for also checking the total NMHC, sum of PAMS target species, and unidentified concentrations as well as individual hydrocarbon concentrations during audits.
- TTP checks of precision indicated problems with ethane concentrations at one of the laboratories. We discussed the need for TTP challenges to include acetylene as well.

- One of the laboratories had trouble with one of their auto-GC's during the laboratory performance audit resulting in low n-pentane concentrations. The district recommends that we not use the data from this auto-GC in our analyses.
- The laboratory performance audit showed that Ventura reports 1-butene while other laboratories reported i-butene.
- All the laboratories used the audit results to continue to improve their hydrocarbon analyses.

The ARB has thorough documentation and understanding of the precision and accuracy of their PAMS VOC data based on their audit program. (We inspected only a sampling of their documentation.) They also conduct regular meetings to discuss possible improvements in analytical and sampling techniques. From the documentation that we had, the uncertainty associated with the PAMS measurements appears to be about 10 to 20 percent for most hydrocarbons.

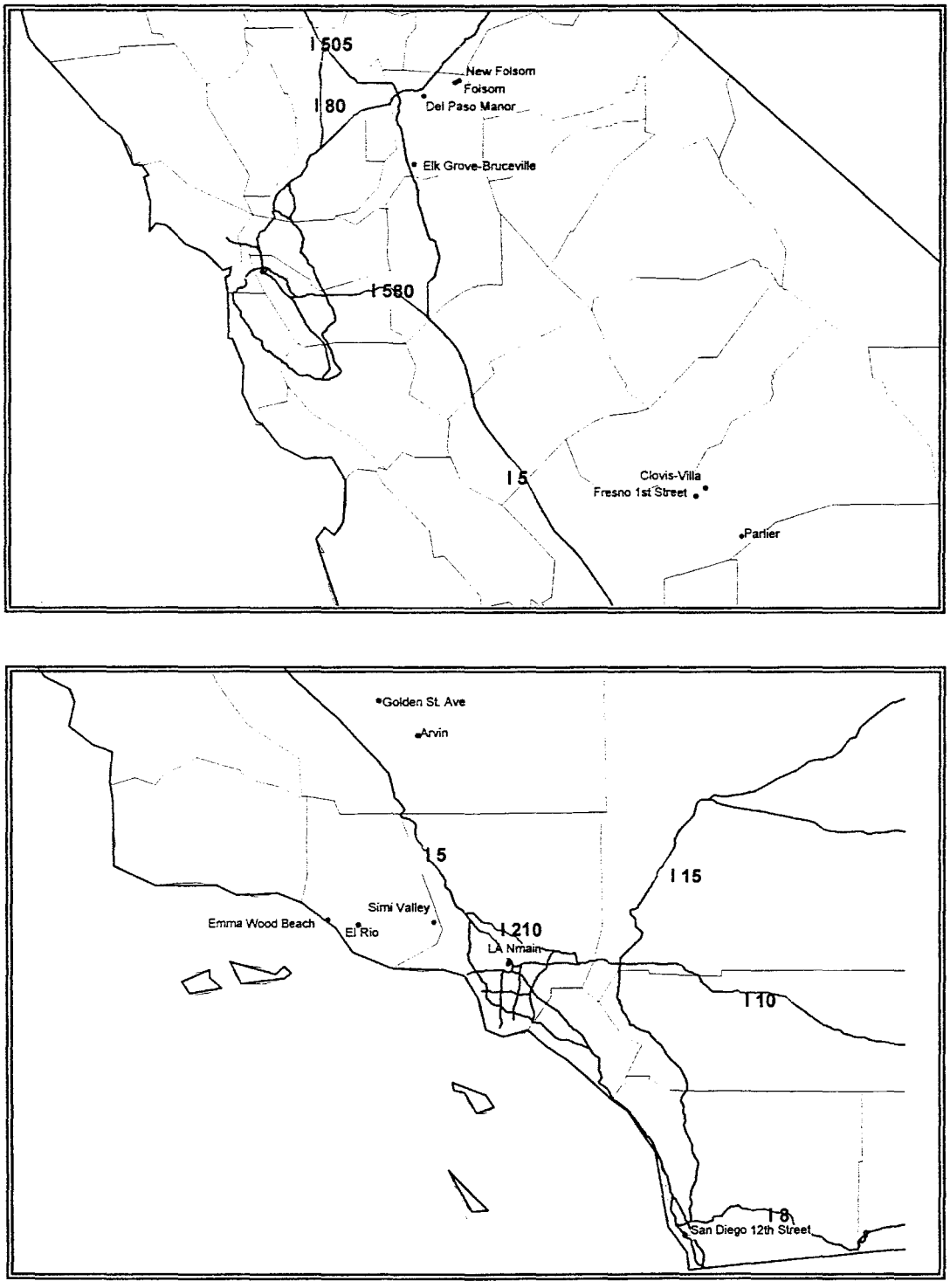


Figure 2-1. Map of California PAMS and long-term trend sites investigated in this project.

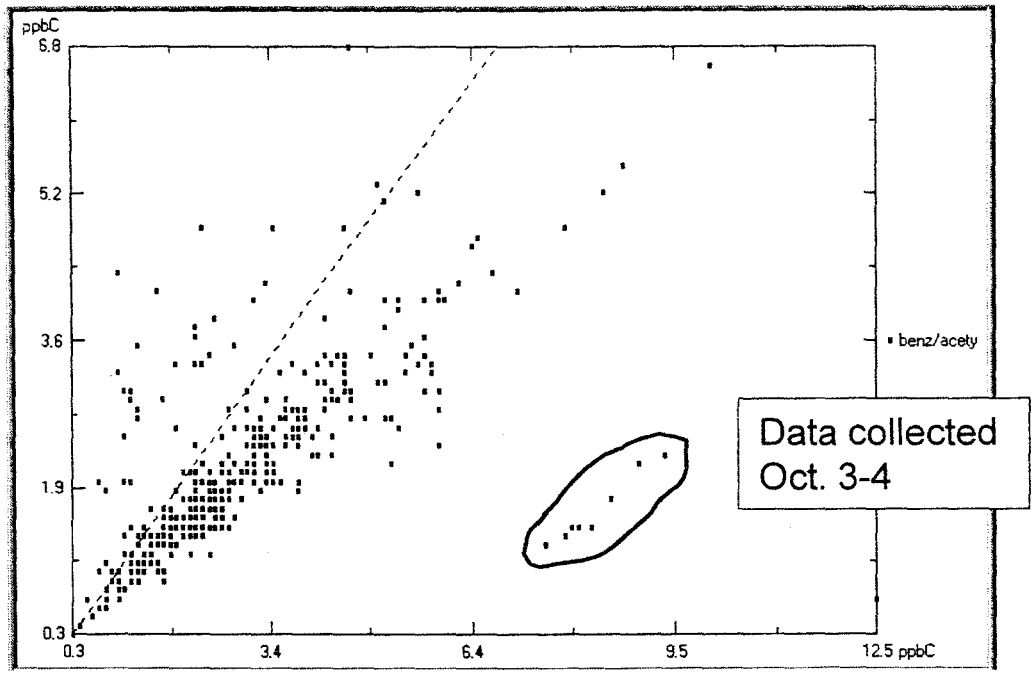


Figure 2-2. Scatter plot of benzene (y-axis) versus acetylene (x-axis) concentrations (ppbC) at a San Diego County PAMS site. Acetylene data collected on October 3-4, 1997, at Alpine, Overland, and El Cajon were anomalous compared to all other data.

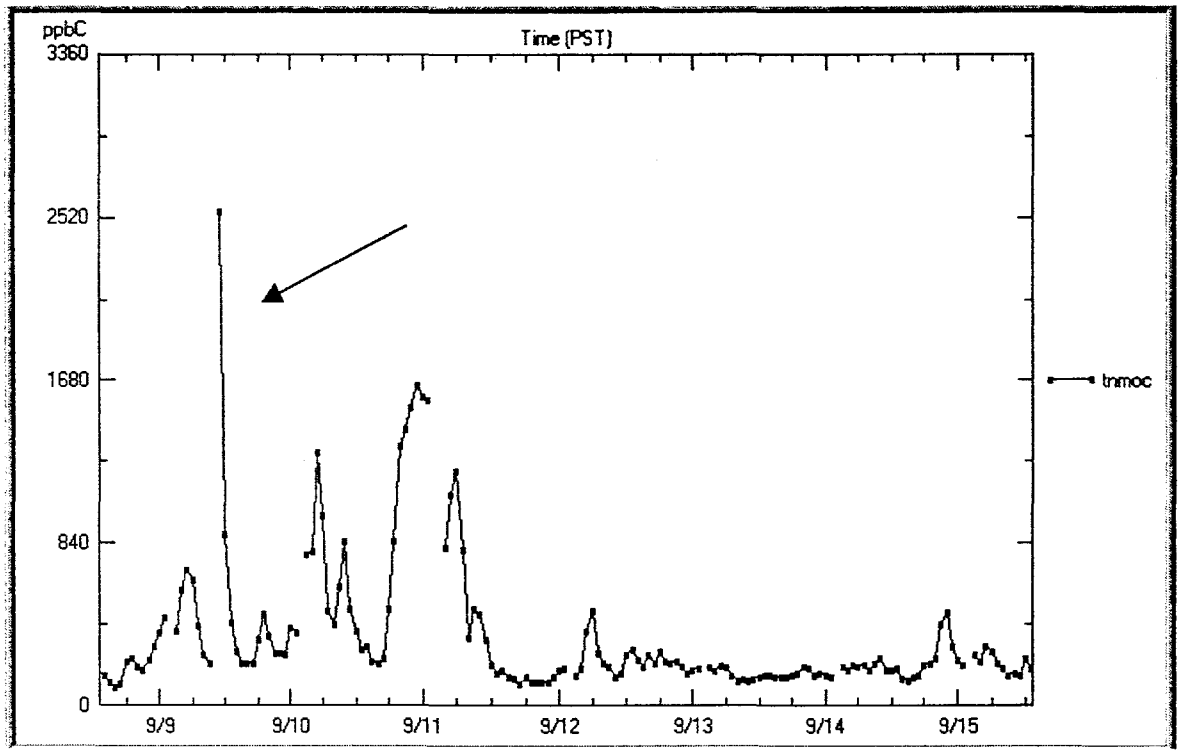


Figure 2-3. Continuous NMHC concentrations (ppbC) at Clovis during 1996. The arrow points out data that appear odd.

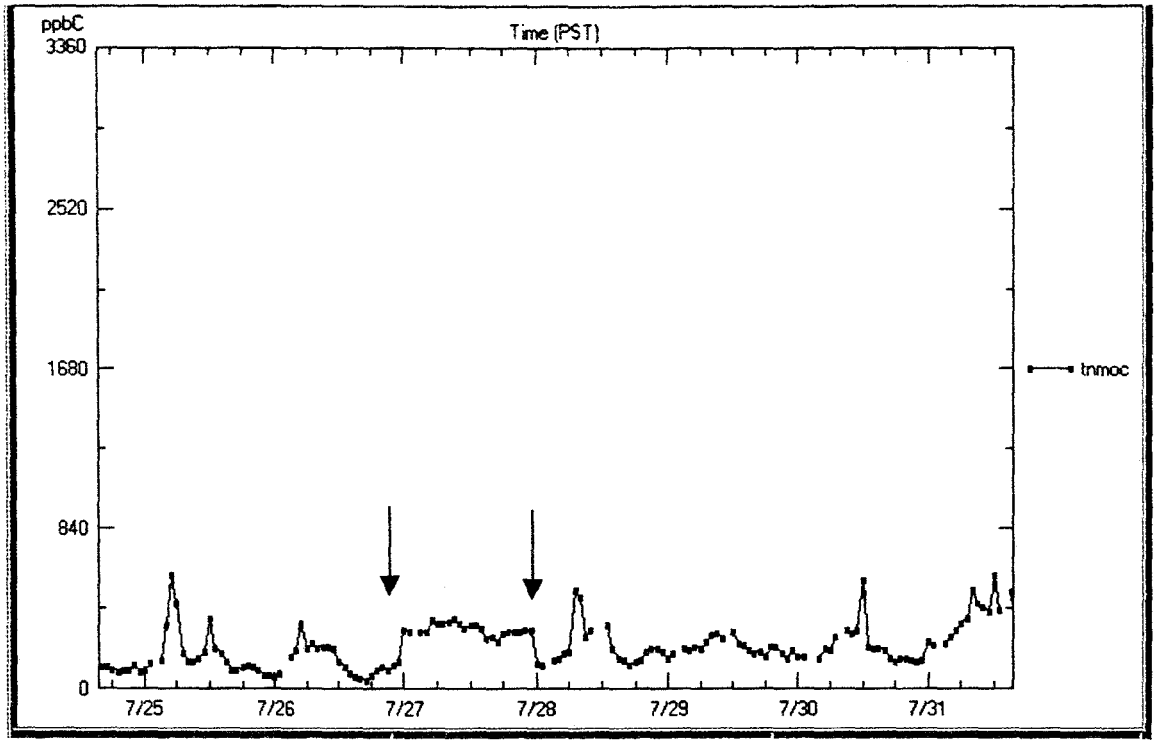


Figure 2-4. Continuous NMHC concentrations (ppbC) at Clovis during 1997. The arrows indicate a time period during which the data appear odd compared to surrounding time periods.

Table 2-1. Site information for PAMS and long-term trends sites in four California districts from 1990 through 1997. Districts include Ventura County, Sacramento, San Diego, and the San Joaquin Valley.

AIRS ID	City	MSA[a]	Sampling Frequency	Lat.	Long.	PAMS Site Type[b]	Hours (PST)
060670011	Elk Grove-Bruceville	Sacramento	3rd day	38.303	-121.4207	1	23,5,12,16
060670006	Del Paso Manor	Sacramento	3rd day	38.614	-121.3669	2A	23,5,12,16
060671001	Folsom	Sacramento	3rd day	38.675	-121.1858	3	23,5,12,16
060670012	New Folsom	Sacramento	3rd day	38.684	-121.1628	3	23,5,12,16
060290010	Bakersfield Golden St. Ave.	Bakersfield	3rd day	35.385	-119.0147	2	23,5,12,16
060295001	Arvin	Bakersfield	3rd day	35.209	-118.7763	1/3	23,5,12,16
060195001	Clovis-Villa	Fresno	3rd day	36.819	-119.7166	2	23,5,12,16
060190008	Fresno 1st Street	Fresno	3rd day	36.782	-119.7732	ARB	23,5,12,16
060194001	Parlier	Fresno	3rd day	36.597	-119.5042	3	5,12,16
060730003	El Cajon	San Diego	3rd day	32.791	-116.942	2	2,5,12,16
060730006	Overland	San Diego	3rd day	32.836	-117.129	2A	2,5,12,16
060731006	Alpine	San Diego	3rd day	32.842	-116.768	3	2,5,12,16
060731007	San Diego 12th St	San Diego	3rd day	32.710	-117.1526	ARB	5,12
060371103	Los Angeles N. Main	Los Angeles	3rd day	34.067	-118.242	ARB	5,12
061112003	Emma Wood State Beach	Ventura	3rd day	32.280	-119.3153	1	2,5,12,16
061113001	El Rio	Ventura	3rd day	34.252	-119.1545	2	2,5,12,16
061112002	Simi Valley	Ventura	3rd day	34.278	-118.6847	3	2,5,12,16

[a] MSA = Metropolitan statistical area

[b] Site types: 1= upwind and background site, 2= maximum ozone precursor emissions impact, 3= maximum ozone concentration, ARB = ARB long term trend site

Table 2-2. Hydrocarbon and carbonyl compound data available at PAMS and long-term trends sites in four California districts from 1990 through 1997. Districts include Ventura County, Sacramento, San Diego, and the San Joaquin Valley. Four samples were collected per day on every third day unless otherwise noted.

City	1990	1991	1992	1993	1994	1995	1996	1997
Elk Grove-Bruceville							7/8-9/30	7/2-10/1
Del Paso Manor					7/1-9/29	7/1-9/30	7/2-9/30	7/2-10/1
Folsom						7/28-10/30		
New Folsom							7/5-9/30	7/2-10/1
Bakersfield Golden St. Ave.					7/1-9/29	7/2-9/30	7/1-9/30	7/2-10/1
Arvin						7/1-9/30	7/1-9/30	7/2-10/1
Clovis-Villa				6/24-10/16[2]	7/6-9/29	7/1-9/30	7/1-9/30	7/2-10/1
Fresno – 1st Street	6/10-9/29[1]		6/5-10/15[2]	6/15-10/31[3]	7/1-9/30	7/1-9/30	7/1-9/30	7/1-10/1
Parlier						7/1-9/30	7/3-9/30	7/2-10/1
El Cajon						7/8-10/30	7/2-9/29	7/3-10/4
Overland						7/8-10/30	7/2-10/27	7/3-10/4
Alpine						7/8-10/27	7/2-9/29	7/3-10/4
San Diego - 12th Street	6/1-9/26[4]	6/14-10/15[1]	6/5-10/15[2]	6/15-10/1[2]	7/1-9/29[2]	7/2-10/3[2]	7/2-9/30[2]	7/3-10/1[2]
L.A.-North Main St.	6/4-9/26[4]	6/18-10/15[1]	6/5-9/30[2]	6/15-10/4[2]	7/7-10/5[2]	7/2-10/6[2]	7/2-9/30[2]	7/3-10/1[2]
Emma Wood State Beach							7/26-9/24	6/3-9/29
El Rio						6/5-10/30	6/2-9/24	6/3-10/31
Simi Valley						6/5-10/30	6/20-9/24	6/3-10/31

1 = one sample per day, every third day
2 = two samples per day, every third day
3 = eight samples per day, every third day
4 = one sample per day, every sixth day

Table 2-3. Number (n) of valid hydrocarbon samples available at PAMS and long-term trends sites in four California districts from 1990 through 1997. Districts include Ventura County, Sacramento, San Diego, and the San Joaquin Valley. Percent completeness (%) greater than 100 percent indicates that additional samples were collected beyond the sampling schedule.

City	1990		1991		1992		1993		1994		1995		1996		1997	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Elk Grove-Bruceville													100	89	117	98
Del Paso Manor									99	83	107	89	120	100	120	100
Folsom											106	86				
New Folsom													83	69	93	75
Bakersfield Golden St. Ave.									106	88	145	121	143	119	119	99
Arvin											124	100	112	90	118	95
Clovis-Villa							76	100	111	99	137	111	134	108	124	100
Fresno - 1st Street	27	71			79	88	230	72	132	107	130	105	145	117	129	104
Parlier											121	98	117	94	116	94
El Cajon											136	90	108	90	121	98
Overland											116	76	80	50	126	102
Alpine											140	92	105	88	109	88
San Diego - 12th Street	9	45	18	90	71	81	33	83	46	77	43	69	39	65	51	85
L.A.-North Main St.	20	100	29	73	59	78	67	84	49	79	44	71	42	68	55	89
Emma Wood													49	58	151	77
El Rio											105	54	60	40	105	54
Simi Valley											135	69	69	52	151	77

3. DATA ANALYSIS RESULTS

PAMS data are collected to be used to corroborate precursor emission inventories, assess changes in emissions due to control strategy implementation, assess ozone and precursor trends, provide input to models, and evaluate population exposure. Control strategy assessment includes an investigation of NO_x versus hydrocarbon controls, hydrocarbon source identification, and identification of hydrocarbons that are specifically important to ozone formation. To aid in these assessments, useful analyses include the statistical description of the validated data including the spatial and temporal variations, analysis of the age of the air mass using ratio techniques, and an evaluation of the ozone formation potential (photochemical reactivity) of the VOCs. Section 3 presents example analyses using the PAMS and long-term trend site data described in Section 2.

3.1 SPATIAL, TEMPORAL, AND STATISTICAL DISTRIBUTION OF THE PAMS MEASUREMENTS IN CALIFORNIA

3.1.1 Overall Characteristics

It is helpful to have an overall understanding of the database before proceeding to a more detailed analysis. The analyses should progress from the "big picture" (e.g., total NMHC) to the details (e.g., hydrocarbon groups and the individual hydrocarbons). We chose to focus on 1997 data for several analyses since these data were most recent, a comprehensive analysis of all data was beyond the scope and resources of the project, and significant yearly changes in individual hydrocarbons were observed such that we did not want to combine years for many of the analyses. To place the 1997 data in historical perspective, **Table 3-1** shows the median NMHC concentrations by year and site. Note that this table represents all data collected; for example, at Los Angeles, some site-years contain samples collected in the morning only, all day, or twice daily. In general, the 1997 total NMHC values are consistent with previous years.

Table 3-2 summarizes the median total NMHC and unidentified concentrations, isoprene percentages, and toluene-to-benzene ratios at each site (grouped by location). Our observations follow (these topic areas are further discussed later in this chapter):

- Total NMHC values are an indicator of emission source strength impacting the site. The NMHC concentrations were highest at the Type 2 sites. NMHC concentrations at Type 1 sites were lower than concentrations at Type 3 sites. These observations are consistent with siting criteria.
- Interestingly, the median unidentified hydrocarbon concentrations were similar among all non-Ventura County sites, ranging from 53 to 70 ppbC, regardless of site type. The unidentified values were computed by taking the difference between the reported NMHC and the sum of the reported identified hydrocarbons. Unidentified compounds can include polar (often oxygenated) species when drying of the samples is not

performed, PAMS species that were missed or misidentified, or VOCs not on the target list such as higher molecular weight alkanes and aromatics.

- Isoprene is the only tracer for biogenic emissions that is reported in the PAMS program. We would expect the contribution of isoprene to the NMHC to be higher for more rural sites than urban sites, even on a 24-hr basis. Overall, the isoprene percentage is low; less than 2 percent at most sites. In the Mid-Atlantic region, isoprene comprises a larger percentage, generally 3 to 4 percent (Main et al., 1999).
- The toluene/benzene ratio is an indicator of source strength and the proximity of fresh emissions. Toluene reacts away faster in the atmosphere than benzene and, thus, a lower ratio can be an indication of a more aged air mass. We would expect higher ratios for the Type 2 sites and lower ratios for the more rural sites; this is generally the case.

Another overview of the data is provided in **Table 3-3**; this table lists the median composition at each site by species groups. Observations include:

- The paraffins were the most abundant species group at Bakersfield, San Diego, Los Angeles, Emma Wood, and El Rio. The paraffin fraction was highest at the Bakersfield, Ventura, and Southern California sites, possibly indicating more influence from oil production, natural gas, or other sources of primarily paraffins.
- The olefin and aromatic fractions were higher at the more urban sites than at the more rural sites. This is consistent with our conceptual model of aging of the air mass - in general, the olefins and aromatic hydrocarbons are more reactive than the paraffins.
- The unidentified fractions were greater than 50 percent at the Sacramento and Fresno area sites. The practical implication of this relatively high fraction is that analyses of possible sources and sample ozone formation issues are less meaningful because more than half the sample carbon is unaccountable.

Another way to compare sites and obtain an overall understanding of the data is to inspect various stratifications of selected hydrocarbons and species groups. The data may be stratified in many different ways: by site, year, month, day of week, weekday/weekend, and time of day. One useful plot is a box-whisker plot (an example is shown in **Figure 3-1**). The box shows the 25th, 50th (median), and 75th percentiles. The whiskers always end on a data point, so when the plots show no data beyond the end of a whisker, the whisker shows the value of the highest or lowest data point. The whiskers have a maximum length equal to 1.5 times the length of the box (the interquartile range). If there are data outside this range, the points are shown on the plot and the whisker ends on the highest or lowest data point within the range of the whisker. The "outliers" are also further identified with asterisks representing the points that fall within three times the interquartile range from the end of the box and circles representing points beyond this. These plots are also useful for data validation.

For example, **Figures 3-2 through 3-4** show the diurnal variation of NMHC concentrations at all sites in 1997 using box plots. Observations include the following:

- Generally, the NMHC concentrations were lower midday than in the 0500 and 2300 PST samples. This temporal pattern is consistent with differences in mixing height between the early morning/nighttime and midday (i.e., low at night, higher in the day allowing for more dilution of emissions).
- In the Sacramento area, NMHC concentrations ranged form 50 to 200 ppbC and were generally highest at Del Paso Manor and lowest at Elk Grove. This is consistent with the site types.
- In the Bakersfield area, NMHC concentrations were about the same midday at both Bakersfield and Arvin, but much higher at Bakersfield in the 0500 and 2300 PST sampling periods. This temporal pattern may indicate higher emissions near the Bakersfield site at night.
- Unlike most other sites in this project, NMHC concentrations at 2300 PST at Clovis and Fresno were similar to midday. All three sites in the Fresno area exhibited similar concentrations at 0500 PST. The morning emissions appear to be similar in magnitude near the three monitoring sites.
- High NMHC concentrations at Emma Wood show that nearby sources sometimes influence this site. Investigation of the fingerprints at this site shows that high concentrations of propane, butanes, and pentanes were frequently observed.

3.1.2 Abundant Hydrocarbons

There are approximately (depending upon the year and the number of coelutions, etc.) 60 PAMS target species with many of these species frequently at or below the detection limit in most samples. Therefore, individual species analyses should focus on the most abundant species based on concentration and on reactivity-weighted composition. The maximum incremental reactivity (MIR) scale developed by Carter (1994) was used to characterize the reactivity of the samples. The MIR scale provides an estimate of moles ozone formed per mole carbon of each hydrocarbon measured, where the ozone formation estimates are intended to be used in a relative rather than absolute manner. The reactivity of each species is estimated by multiplication of its weight percent or concentration by its MIR factor. Note that the unidentified mass can be a significant contributor to the total carbon and its potential reactivity is unknown.

Tables 3-4 and 3-5 list the ten most abundant hydrocarbons on a weight percent basis and on a reactivity-weighted basis, respectively. To determine these lists, the median values of each species over the data set are sorted. We focused on the most recent year of data (1997) as the most up-to-date assessment of ambient VOC conditions. We also inspected the ten most abundant hydrocarbons for other years and found similar results. Observations from this analysis include the following:

- Many of the same species are abundant on a concentration basis at all sites in the table including ethane (except San Diego), propane, toluene, n-pentane, n-butane (except San Diego), i-pentane, and m-&p-xylenes (except El Rio and Emma Wood).

- All Type 1 sites also had the relatively less reactive i-butane as an abundant species. All Type 2 sites also had the highly reactive i-butene or 1-butene as abundant species. These findings are consistent with the site types.
- Benzene was less abundant than in the past (e.g., Stoeckenius et al., 1998); it was not present on the top 10 list at any of the sites. Benzene's decrease in importance probably reflects changes in fuel content (e.g., Main et al., 1998b; Hammond, 1996).
- On a concentration basis, the biogenic tracer isoprene is important at only a few sites (i.e., Arvin, Del Paso Manor, Clovis, and Folsom).
- When reactivity is considered, the increased importance of olefins and aromatic hydrocarbons is illustrated. Ethene, m-&p-xylenes, propene (except Simi Valley), toluene (except Emma Wood), i-pentane, and 1,2,4-trimethylbenzene were in the top ten at every site. At most of the San Joaquin Valley/Sacramento sites, isoprene was abundant. Many sites added 1-butene, n-butane, and 1-pentene to the abundant reactive species list.

Some chemical species are uniquely (or nearly uniquely) associated with a particular source and may be viewed as a "tracer of opportunity". If an indicator species can be identified for a major source within an urban area, information concerning source contributions may be obtained without the need for sophisticated approaches. Several examples of species which are identified in the PAMS measurements and may be useful in this regard include (Stoeckenius et al., 1994): butane for gasoline evaporative emissions; acetylene, benzene, or ethene for motor vehicle exhaust; ethane and propane for natural gas; and isoprene for biogenic emissions (**Table 3-6**). The most abundant species identified at the California sites indicate that motor vehicle exhaust was very important.

3.1.3 Fingerprints

Ultimately, the compositions of the VOC samples are used to try and assess what hydrocarbons are most important in ozone formation and then to try and determine the likely sources of those hydrocarbons. Control strategies can then be developed to target these sources. Questions of interest regarding the composition of samples at PAMS sites include: Are there significant changes in the composition with time of day? Are there significant changes in the composition from year to year? What sources likely contribute to important hydrocarbons? How important are the speciated measurements? How often do they need to be performed?

We investigated the composition of hydrocarbons (i.e., weight percent) at each site and how the composition varied with time of day in 1997 using fingerprint plots. In these plots, the hydrocarbons are listed in the order in which they elute from the chromatograph. The median weight percents instead of concentrations are used because it is easier to compare the relative differences among time periods and among sites. Also note that the data were screened to include only those hydrocarbons with more than five samples above the detection

limit. In most cases, this excluded the 0800 PST samples (only four samples were collected at a few sites in 1997). Observations from **Figures 3-5 through 3-17** include:

- The weight percent of isoprene was higher at midday and in the afternoon than in the morning at most sites (e.g., Figure 3-5). This is consistent with biogenic emission patterns.
- More-reactive species declined relative to less-reactive species at midday and in the afternoon at most non-Type 2 sites (e.g., Figure 3-6).
- The C2-C4 paraffins were dominant hydrocarbons at many of the non-Type 2 sites (e.g., Figure 3-7). As the air parcels travel downwind of the emission sources (e.g., from a Type 2 site to a Type 3 site), the heavier paraffins and aromatic hydrocarbons would react away faster relative to the lighter paraffins.
- The composition at Los Angeles (Figure 3-8) and San Diego (Figure 3-9), Type 2-like sites, showed little change with time of day for anthropogenic hydrocarbons indicating fresh emissions impacted the sites during both sampling periods. At Los Angeles, the n-butane contribution was higher at midday, possibly due to increased evaporative emissions.
- At Clovis (Figure 3-10), the composition showed relatively little change with time of day for many species except isoprene and toluene; these hydrocarbons showed higher contributions midday. The higher isoprene is consistent with midday peaks in isoprene emissions from biogenic sources. The higher toluene fraction midday may indicate a midday source of toluene unrelated to the other aromatic hydrocarbons. Possible sources of the toluene should be investigated.
- At Bakersfield (Figure 3-11), the composition showed relatively little change with time of day for the C5+ hydrocarbons. The C2-C4 paraffins were more abundant during the 0500 PST sampling period and the pentanes were higher during the 1200 PST sampling period. Possible sources of these hydrocarbons should be investigated.
- At El Rio (Figure 3-12), the C5+ hydrocarbons showed little change with time of day; the C2-C4 hydrocarbons showed the most change with time of day. One of the more interesting changes was the higher contribution from 1-butene during the 1200 and 1600 PST sampling periods. This is counter to our understanding of reactivity; we would expect this highly reactive hydrocarbon to be depleted midday relative to the morning. The contributions of ethane and propane were highest during the 0500 PST sampling period.
- At Fresno (Figure 3-13), many of the hydrocarbons had their highest fractions in the 0500 PST samples and their lowest contributions in the midday samples. The isoprene fraction was highest at 1600 PST. The diurnal behavior of the hydrocarbons at this site was more typical of a downwind site (i.e., Type 3) than a city-center site (i.e., Type 2).
- At Del Paso Manor (Figure 3-5), the weight fractions were higher for most hydrocarbons during the 0500 or 2300 PST sampling periods. Midday contributions from the aromatic hydrocarbons were typically lower than the night or morning time

periods; this is consistent with a more aged air mass. This site, like the Fresno site, shows many features commonly associated with a Type 3 site instead of a Type 2 site.

- The composition at Elk Grove (Figure 3-14) and New Folsom (Figure 3-15) were highly variable with time of day; likely due to the low concentrations observed at the sites. For many of the hydrocarbons, concentrations were below detection in more than half the samples. The more abundant anthropogenic hydrocarbons were highest during the 0500 and 2300 PST sampling periods while isoprene was highest during the 1600 PST period. The more reactive species, such as the xylenes, were mostly below detection during the daytime. The composition at Parlier (Figure 3-6) was similar to these two sites except that isoprene concentrations were typically below detection all day. These observations are consistent with the site types.
- The composition at Arvin (Figure 3-7) changed relatively little with time of day, but appeared relatively aged as many of the hydrocarbons were often below detection. Ethane, propane, n-butane, and i-pentane dominate the composition. The contribution from C5+ hydrocarbons is very small. The composition at this site appears to be well aged.
- The composition at Simi Valley (Figure 3-16) was about the same all day except for acetylene and ethane; these species showed their highest contribution midday. This is likely due to the accumulation of these two relatively less reactive hydrocarbons rather than new sources. The composition at this site was more consistent with a Type 2 site than a Type 3 site.
- At Emma Wood (Figure 3-17), the C3-C4 paraffins dominated the composition. These species were generally more important in the 0200 and 0500 PST sampling periods. The unidentified contribution was much higher during the 1200 and 1600 PST periods than in the morning coinciding with the decrease in contribution from the C3-C4 paraffins. This may be an indicator of more aged air impacting the site in the afternoon. The small contributions from C5+ hydrocarbons showed little change with time of day.
- The composition was very similar among the Clovis, Del Paso Manor, Bakersfield, Fresno, Los Angeles, and San Diego Type 2 sites. At these sites, the two dominant species are i-pentane and toluene, followed by the xylenes, ethane, and propane. The composition is typical of urban sites throughout the United States (see for example, Main et al., 1999; Stoeckenius et al., 1998).
- At the Simi Valley Type 3 site, the composition changed relatively little between 1995 and 1997 (acetylene differences may be analytical).

We also investigated the median composition as a function of year (**Figures 3-18 through 3-26**). This analysis is one way to investigate changes in the emissions. To lessen the differences caused by unidentified fraction changes among the years, the median compositions were normalized over the identified hydrocarbons. We found the following:

- The benzene fraction decreased significantly after 1995 at the Los Angeles (Figure 3-18), San Diego (Figure 3-19), Fresno (Figure 3-20), Clovis (Figure 3-21), Bakersfield (Figure 3-22), and Del Paso Manor (Figure 3-23) Type 2 and Type 2-like sites and at the Parlier (Figure 3-24) Type 3 site. Many C5-C7 paraffins showed increased fractions in 1996 and 1997. Also at Los Angeles, San Diego, Fresno, and Bakersfield, there was a significant decrease in the n-butane fraction. These changes in benzene, C5-C7 paraffins, and n-butane are consistent with fuel changes to reduce benzene, aromatics, and fuel volatility to conform to reformulated gasoline (RFG) regulations (e.g., http://www.epa.gov/OMSWWW/rfgnew.htm). Caldecott Tunnel (San Francisco Bay area) measurements of vehicle exhaust by Kirchstetter et al. (1999) showed similar trends in these hydrocarbons.
- At Los Angeles and San Diego, there was a significant increase in the i-butene fraction after 1994. This is understandable since i-butene is a thermal decomposition product of methyl-tert-butyl-ether (MTBE); MTBE is an oxygenated species that was added to RFG to meet oxygen guidelines. Zielinska et al. (1997) showed that i-butene increases were mirrored by MTBE increases in ambient air in the Los Angeles Basin. The Caldecott tunnel measurements also confirmed this observation (Kirchstetter et al., 1999).
- With low concentrations at the Arvin (Figure 3-25) Type 1/3 site, annual changes were harder to distinguish. However, small decreases in benzene and other aromatic hydrocarbons were observed.
- Benzene fractions did not appear to decrease between 1995 and 1997 at Simi Valley (Figure 3-26). Small increases in the relative amounts of 1-butene[1] and decreases in the heavier aromatic hydrocarbons were observed, and this is consistent with fuel changes.

These discussions show that significant changes in the composition with time of day and by year were observed. These changes corroborate emissions changes and our conceptual model of transport and air mass aging; these analyses show the importance of sample speciation.

To investigate the variation in the composition from day to day, we tabulated the hydrocarbons that had coefficient of variation (COV) values (based on the mean and standard deviations in the 1997 weight fractions) greater than 50 percent at each site (**Table 3-7** and Appendix C). An analysis of the COV illustrates the variability in the hydrocarbon weight fraction. This analysis was performed using weight fractions because we expect concentrations

[1] In audit results obtained from the ARB, Ventura County sometimes reported 1-butene when their analytical systems were challenged with i-butene. Most of the other California PAMS sites report i-butene.

to vary significantly. COV values were inspected overall for 1997 and by sampling period. Observations include:

- The two sites with the lowest variability in the overall hydrocarbon composition were Del Paso Manor and Los Angeles, where most COV values were below 30 and 40 percent, respectively. These results reflect the impact of fresh emissions of similar composition during all sampling periods; this is consistent with their locations. Based on the composition, the fresh emissions are likely mostly due to motor vehicles.
- Greater variability in sample composition (e.g., 17 hydrocarbons exhibited more than 50 percent COV) was observed at Elk Grove because of generally low concentrations (i.e., more hydrocarbons near the detection limit where measurement uncertainty is higher). This finding is also consistent with the site's rural location.
- The moderate variability (e.g., 10 to 14 hydrocarbons with more than 50 percent COV) observed at Bakersfield, Clovis, Fresno, and San Diego Type 2 and Type 2-like sites possibly reflects the impact of fresh emissions during the morning (likely from motor vehicle emissions), contrasted with a different composition at midday and in the afternoon. The midday and afternoon samples may be influenced by more aged air impacting the sites (i.e., the relative loss of more reactive species compared to less reactive species), increased biogenic emissions (e.g., isoprene), and increased evaporative emissions (e.g., butanes).
- The moderate variability (e.g., 9 to 10 hydrocarbons with more than 50 percent COV) observed at Folsom, Parlier, and Arvin sites is likely a function of changes in the composition with time of day due to aging of the air mass, biogenic emissions patterns, and sometimes low concentrations.
- In general, Ventura County hydrocarbons showed greater variability than the other sites. For example, we had to set the screening limit for COV values greater than 100 percent to reduce the number of hydrocarbons to include in the summary table. The county reported instrument difficulties during 1995-1997 that may have hampered their ability to obtain the best quality data.

We also investigated the COV as a function of the time of day data were collected. Observations from this analysis include:

- Typically, the COV was greater in the 1200 and 1600 PST samples at San Diego, Parlier, Arvin, Clovis, Del Paso Manor, Folsom, and Elk Grove than in the other samples. In the daytime, the sample composition is probably more affected by shifts in wind speed and direction, mixing with aloft air, photochemistry, and changes in emissions than in the early morning or nighttime.
- At Bakersfield, nighttime weight fractions of styrene, 1,3,5-trimethylbenzene, decane, and undecane varied significantly possibly indicating the impact of emissions under varying meteorological conditions. Hydrocarbon weight fractions of decane and undecane at Clovis showed similar nighttime variably.

- On an individual hydrocarbon basis, variability in the weight fractions at each site was likely from changes between sampling periods in emissions (e.g., biogenic isoprene emissions, evaporative butane emissions), in reactivity (the relative decrease in olefins and aromatic hydrocarbons), analytical difficulties (e.g., documented problems with acetylene, column recovery issues with decane and undecane), or combinations of these.

3.2 RELATIONSHIPS AMONG SPECIES

We investigated the relationship among the abundant species through the use of scatter plots, correlation matrices, and cluster and factor analyses. These analyses help the analyst understand which species potentially come from the same source, and what the typical ratios are between species or species groups.

3.2.1 Changes in Benzene

During data validation (Main, 1998), we observed a bifurcation in the benzene to toluene relationship at Del Paso Manor. Samples collected during 1994 and 1995 showed a different benzene to toluene ratio than samples collected in 1996 and 1997 (**Figure 3-27a**). The change shown in the figure is probably due to fuel changes between 1995 and 1996 with the implementation of California RFG in early 1996 (e.g., www.arb.ca.gov/cbg/kgo/arb71896.htm). To further investigate changes in fuel formulation and our ability to observe the change in ambient data, we prepared scatter plots of benzene to toluene for all years and sites. Results at the other sites varied as discussed below:

- At Clovis (**Figure 3-27b**), we observed higher benzene concentrations in 1995 with respect to toluene than in 1996 and 1997 which is consistent with the fuel changes. However, this site appears to have a separate toluene source from the benzene thus introducing more scatter in the relationship than was observed at Del Paso Manor.
- The benzene and toluene relationships at Fresno (**Figure 3-28a**) and Bakersfield (**Figure 3-28c**) were generally similar to the relationships observed at Del Paso Manor with a higher ratio in 1994 and 1995 than in 1996 and 1997.
- The benzene and toluene relationship at Los Angeles (**Figure 3-28b**) showed different ratios in 1994, 1995, and 1996/1997. These differences are in agreement with Federal Phase I RFG implementation in Los Angeles in 1995 and California RFG implementation in 1996.
- The benzene and toluene relationships at San Diego (**Figure 3-29a**) and Simi Valley (**Figure 3-29b**) do not show a clear trend with year.

3.2.2 Scatter Plot Matrices

We prepared scatter plot matrices of abundant hydrocarbons at each site (**Figures 3-30 through 3-42**). To interpret a scatter plot matrix, locate the row variable (e.g., ispna in Figure 3-30 near the top left) and the column variable (tolu) on the bottom. The intersection is the scatter plot of the row variable on the vertical axis against the column variable on the horizontal axis. Each column and row are scaled so that data fill each frame. Scale information is omitted for clarity. The diagonal plots contain histograms of the data for each row variable. Observations from this analysis include the following:

- Hydrocarbons at Type 2 and Type 2-like sites (i.e., Del Paso Manor, Clovis, Fresno, Bakersfield, Los Angeles, San Diego, and El Rio - Figures 3-30 through 3-36) show stronger correlations among hydrocarbons than sites at more downwind locations (i.e., Folsom, Elk Grove, Parlier, Emma Wood, and Arvin - Figures 3-37 through 3-41). For example, at Del Paso Manor, Clovis, Fresno, Bakersfield, and Los Angeles, hydrocarbons commonly present in motor vehicle emissions (e.g., toluene, xylenes, i-pentane, n-pentane) generally correlate well with hydrocarbons present in motor vehicle exhaust (e.g., acetylene, ethene) indicating that they are probably emitted from the same source.
- Biogenic isoprene does not correlate well with other abundant hydrocarbons, as expected.
- When i-butene is present, it generally correlates well with other exhaust-related hydrocarbons (see Del Paso Manor, Fresno, Clovis, Bakersfield, Los Angeles, and San Diego). Since i-butene has been linked to the combustion of MTBE, its relationship to other hydrocarbons formed during combustion is consistent.
- More scatter is observed in the El Rio data than in data from other Type 2 sites, in the San Diego data than in the Los Angeles data, and in the Clovis data than in the Fresno data. The sites with more scatter in the relationships may experience more frequent "hits" from nearby point sources or from more aged air than the other sites.
- At Bakersfield, ethane, propane, and n-butane correlate well indicating that they are emitted from similar sources.
- In Los Angeles and San Diego, n-butane concentrations do not correlate as well with other motor vehicle-related hydrocarbons (e.g., toluene, i-pentane, xylenes) as was observed at other sites. This may indicate unique sources of n-butane.
- At Clovis, the toluene concentrations showed little correlation with other abundant hydrocarbons indicating a separate, nearby source(s) impacting the site.
- At Simi Valley (Figure 3-42), hydrocarbon relationships are more similar to relationships observed at Type 2 sites than at non-Type 2 sites.

3.2.3 Correlation Matrices

We performed a Pearson correlation analysis on the morning rush "hour" data (0500 to 0800 PST). This analysis provides correlation statistics to corroborate relationships observed in the scatter plot matrices. We identified hydrocarbons that exhibited a correlation coefficient, r, of ≥ 0.707 (i.e., accounts for more than 50 percent of the variance). This analysis showed some interesting results:

- At Del Paso Manor, El Rio, Clovis, Bakersfield, and Fresno, hydrocarbons associated with exhaust (e.g., acetylene, benzene) and evaporative emissions (e.g., n-butane, isopentane) generally showed significant correlations. Many of the hydrocarbons appear to be from a common source: motor vehicles.
- The following relatively abundant hydrocarbons did not correlate with other abundant hydrocarbons indicating a separate emission source: decane (Clovis), i-butane, n-butane, and undecane (Los Angeles), and i-butane (Fresno).
- At Arvin, many of the hydrocarbons showed significant correlations with each other even though the air masses reaching this site tend to be well aged. More scatter (and thus lower correlation coefficients) was observed in the relationships of more reactive species.

3.2.4 Cluster Analyses

Cluster analysis is a multivariate procedure for detecting natural groupings in data. This analysis provides a graphical depiction of the relationships among data groupings, such as individual hydrocarbon species, samples collected at different sites or times of day, etc. Depending upon the complexity of the hydrocarbon mix at a site, one to several clusters or factors may be needed to account for a majority of the variability in the data. Multivariate analyses are useful for inferring the mix of hydrocarbon sources impacting a receptor location. As shown in Table 3-6, there are several key species (such as acetylene, ethene, ethane, i-butane, isoprene, benzene, and xylenes) that are helpful tracers of sources. One of the drawbacks to these analyses is that if there is little variability in a hydrocarbon concentration (e.g., the hydrocarbon is typically below detection), then that hydrocarbon cannot be included in the analysis.

We ran cluster analyses on most sites including all data and morning-only data. Example cluster analysis results for Clovis, Del Paso Manor, Bakersfield, and Elk Grove are shown in **Figures 3-43 through 3-46**. Clustering is computed using normalized Euclidean distance (root mean squared distances). To interpret the figures, consider a vertical line drawn at an arbitrary distance of about 5 in Figure 3-43 (Clovis). Dissecting the figure at 5 appears to be a logical break in the number of clusters. (We could just as easily select 2 in this example to increase the number of clusters to consider.) The species from m-&p-xylenes to n-decane (a key to species abbreviations is provided in Appendix B) on the figure are clustered together suggesting a common source. These species probably represent motor vehicle emissions, both exhaust (e.g., benzene, ethene, acetylene) and evaporative (e.g., n-pentane).

The propane and ethane (natural gas), i-pentane and toluene (motor vehicle and solvent use) concentrations do not cluster with the rest of the data indicating that the species did not behave similarly. Isoprene, a biogenic hydrocarbon, is expected to be in a different cluster from the anthropogenic hydrocarbons, however, its contribution to NMHC in the morning at this site is low. Later in the day, isoprene is in a separate cluster (not shown).

At Del Paso Manor (Figure 3-44), we chose an arbitrary line at about 1.5. The hydrocarbons from acetylene to i-butane form a cluster with motor vehicle exhaust components (e.g., acetylene, ethene) and evaporative emissions (e.g., butanes). Other groupings include i-pentane and toluene (motor vehicle emissions and solvent use), propane (natural gas), and n-decane (unknown, possibly industrial source or diesel emissions). Another cluster, containing ethane and m-&p-xylenes, is odd since ethane is typically associated with natural gas and the xylenes are typically associated with exhaust.

At Bakersfield (Figure 3-45), we chose an arbitrary line at about 8. Clusters at this site were quite different from other Type 2 sites including separate clusters for (and thus different behavior of) propane, acetylene, i-pentane, ethane, and n-butane than the rest of the hydrocarbons.

Figure 3-46 shows the cluster tree for Elk Grove, a Type 1/4 site. The clusters are similar to those previously discussed, with most hydrocarbons present in one cluster representing motor vehicle emissions and with separate clusters for the propane, i-pentane, ethane, and toluene. Propene was also in a separate cluster at this site. The Parlier cluster tree (not shown) was very similar to Elk Grove.

The i-pentane/toluene cluster seen at Del Paso Manor was also observed at Los Angeles, Folsom, Fresno, and Simi Valley. The ethane/toluene cluster may be common among the more downwind sites (e.g., Parlier and Elk Grove). The propane or propane/ethane cluster was common among most sites (Del Paso Manor, Arvin, Los Angeles, Clovis, Folsom, El Rio, Elk Grove, Fresno, Emma Wood, and Simi Valley). In the afternoon at Los Angeles, i-butene was in a separate cluster. Similarly, isoprene was in a separate cluster in the afternoon at Folsom.

This analysis is helpful as a first step in determining which sources are likely important contributors to the ambient VOC concentrations at a site. This analysis is not quantitative, but, rather, helps to begin to qualitatively assess which sources to include in quantitative assessments including chemical mass balance modeling.

3.2.5 Factor Analyses

Principal component analysis (PCA) has been applied as an objective method of removing the highly intercorrelated nature of variations in atmospheric concentrations. The PCA receptor model classifies variables into groups that can then be associated with factors that contribute to pollutant levels at receptors. These factors can be identified as emissions sources, chemical interactions, or meteorological phenomena, depending on the data sets that

have been submitted to PCA. Many of these factors indicate more than one possible cause. The analysis for California data was inhibited because many of the hydrocarbons were below the detection limit of about 1 ppbC.

Factor analysis was performed on the same data set as the cluster analysis; and factor analysis results were similar to those for the cluster analysis. At Fresno (**Figure 3-47**), four factors accounted for about 76 percent of the variance in the data. The first factor contained benzene, toluene, m-&p-xylenes, and acetylene (with more than 50 percent of the variance in these species explained by the factor) and is probably representative of motor vehicle exhaust and evaporative emissions. The second and third factors contained propane, n-butane, ethane, 1,2,4-trimethylbenzene, and i-butane and may be indicative of solvent, evaporative, and natural gas emissions. The last factor contains benzene, xylenes, acetylene, propene, and ethene and is indicative of motor vehicle exhaust. These cluster and factor analyses can be repeated for different data sets including: data collected at different times of day rather than all data combined, data collected under selected wind directions or ozone concentrations, weight percent data rather than concentration data, a larger number of hydrocarbons, the addition of air quality parameters such as NO_x or NO, and sample comparisons rather than species comparisons. Additional analyses were beyond the scope of this project. Note that a recent paper by White (1999) sheds some doubt on the usefulness of factor analyses of ambient concentration data. White's paper is likely to generate significant discussion about the usefulness of this type of analysis.

3.3 ASSESSING THE AGE OF AN AIR MASS

Many of the hydrocarbon species are used as indicators of ozone formation potential and tracers of urban emissions. Assuming that the ratio of these species of interest in the emissions are relatively constant throughout the day, the relative abundance of the more-reactive species (olefins and more-reactive aromatic hydrocarbons such as xylenes) should decrease with time during the daylight hours, and the relative abundance of less-reactive species (paraffins and the less-reactive aromatic hydrocarbons) will appear to increase. The ratios of more-reactive species concentrations to less-reactive species concentrations are also useful as indicators of the relative changes in species composition and age. Comparisons of the ratios among sites can be made to estimate the relative age of air parcels and help provide evidence of transport. In addition, this analysis may present evidence of the presence of fresh emissions or the presence of unique regional sources for a species. Commonly used ratios include xylenes/benzene, toluene/benzene, and ethane/acetylene (Nelson and Quigley, 1983). In this type of analysis, it is important to assess several different species ratios and look for consensus among the results. Analytical problems, nearby emissions of selected species, and other factors make it risky to rely upon the results of a single ratio.

While we investigated several ratios, we have summarized results from two of them: toluene to benzene (T/B) and total xylenes to benzene (X/B). To put these ratios in perspective, consider ratios of these species as measured in urban tunnels shown in **Table 3-8**. We expect ratios measured in tunnels to be relatively representative of fresh motor vehicle

emissions. Table 3-8 illustrates that there are significant differences among the ratios by year and by location. Fuel changes, fuel differences, fleet age, analytical or other measurement differences, and other factors contribute to these differences. The ambient ratios are a function of these factors and of the impact of other sources of the hydrocarbons considered in the ratio. For example, Zielinska et al. (1997) reported ambient morning T/B ratios of 3.9 at Los Angeles in 1996. This ambient ratio is significantly higher than the 1996 Sepulveda tunnel ratio of 2.5. We would expect an ambient site to measure a more aged air mass than a tunnel and we would expect the more aged air mass to experience a lower ratio as the toluene reacts away faster than the benzene. It is likely that non-motor vehicle toluene sources, such as solvent use, impact the Los Angeles site resulting in a higher ratio.

We summarized the median 0500-0800 PST T/B and X/B ratios at the California PAMS sites in 1997 in **Table 3-9**. The data are organized by region (i.e., Sacramento, Bakersfield, Fresno, Southern California, and Ventura). By assessing the data on a regional basis, some of the differences discussed above may be minimized. In general, the Type 2 and urban sites (e.g., Del Paso Manor) had higher ratios than the more downwind or upwind sites (e.g., Elk Grove), indicating fresher emissions impacting the Type 2 sites as expected. Exceptions included:

- The Folsom and Del Paso Manor sites have similar X/B ratios. There may be a separate toluene source impacting the Folsom site that causes a higher T/B ratio than expected at this site.
- The morning T/B ratios in San Diego and Los Angeles are similar to morning ratios observed at other urban sites in Sacramento or the San Joaquin Valley; however, the X/B ratios are slightly lower.
- The Ventura County sites have generally lower T/B and X/B ratios than other California sites. It is unclear whether these differences are due to fleet/fuel differences or analytical differences.
- The Simi Valley and El Rio sites have similar ratios; the Simi Valley data appear more typical of a Type 2 than a Type 3 site.

Figure 3-48 shows the median T/B ratios at the California sites in 1997 as a function of time of day. We expect the ratios to remain about the same throughout the day at the Type 2 sites due to the impact of fresh emissions all day. In contrast, we expect the ratios to decrease in the midday and afternoon at the more downwind/upwind sites due to aging of the air mass at these sites. One difficulty in this analysis is the relatively poor temporal resolution of the data. Each data point represents a 3-hr average beginning at that time. Observations from this figure include the following:

- The T/B ratios stay relatively constant with time of day at El Rio, Bakersfield, Los Angeles, Del Paso Manor, and Fresno as expected. Higher ratios midday were observed at Clovis and San Diego indicating a possible non-motor vehicle source of toluene in the daytime.

- The T/B ratios at the Type 1 and 3 sites decreased midday reflective of aging. Data are not available for all the sites and times because concentrations were often below detection, another indication of aging.

Figure 3-49 shows the median X/B ratios at the California sites in 1997 as a function of time of day. We have found in previous studies that the xylenes to benzene ratio can be more sensitive to changes in air mass age than the T/B ratio (e.g., Main et al., 1999). With X/B ratios, nearly all sites showed some degree of aging midday. Overall, however, the ratios were lower (more aged) at the Type 1 and 3 sites compared to the Type 2 sites as expected. The San Diego site showed higher ratios midday, possibly indicating a non-motor vehicle xylenes source during the daytime.

3.4 NMHC/NO_x

Emission control strategies are based on assessments of whether an area is "VOC-limited" or "NO_x-limited." Ozone is a secondary pollutant formed by the reaction of nitrogen oxides and hydrocarbons. In the presence of oxygen, and with the stimulus of solar radiation, nitrogen oxides react in the atmosphere to form ozone and nitric oxide (NO). Resultant ozone, however, is quickly reacted away to form nitrogen dioxide (NO_2). Ozone, therefore, cannot accumulate unless VOCs are present to consume or convert NO back to NO_2. The precursors to ozone (i.e., NO_x and VOCs) have both anthropogenic and biogenic origins.

There are several data analysis methods to evaluate the relative sensitivities of peak ozone concentrations to emissions of VOCs and NO_x. These methods all rely on the use of ambient measurements and have been deemed "observation-driven methods" (ODMs). ODMs are used in a diagnostic sense to assess the qualitative response of ozone concentrations to changes in VOC or NO_x concentrations. Specifically, the methods characterize expected responses of peak ozone concentrations at a given time and place to reductions in upwind VOC or NO_x emissions by determining if ozone formation is VOC- or NO_x-limited.

No single analysis should form the basis for emission control strategy decisions. Rather, several analyses of indicators, the results from previous data analysis tasks in this project, review of results from other projects, and modeling should provide concurrent evidence. Potential indicators include those listed in **Table 3-10**. Other methods for assessing this issue have been investigated for California under separate contract, including the smog-production (SP) algorithm (e.g., Blanchard performed SP analyses of the San Joaquin Valley as shown in Main et al., 1997). Ambient ozone, NO, and NO_x (or NO_y) concentrations are used to compute the extent of reaction. For this report, we focused only on the morning NMHC/NO_x ratios.

The ratio of NMHC to NO_x in the morning is an important parameter for photochemical systems. The ratio characterizes the efficiency of ozone formation in NMOC-NO_x-air mixtures. At low ratios (less than 10), ozone formation is slow and inefficient (VOC-limited). Decreasing NO_x levels may result in increased ozone, due to less NO titration of ozone. At

high ratios (greater than 15), ozone formation is limited by availability of NO_x rather than VOC. Under transitional regimes, both NO_x and VOC controls may be effective.

Figure 3-50 shows the morning NMHC/NO_x ratios for the sites in California. Data were screened for NO_x concentrations greater than 10 ppb and NMHC concentrations greater than 100 ppbC. These screens are used to ensure that the data are well above the detection limit of the measurements and that the sites are dominated by fresh emissions. Most of the sites (except Emma Wood and El Rio) indicate that the bulk of the data are in the VOC-limited regime. These results are not surprising for the urban sites because other recent investigations of PAMS sites in the United States have shown similar morning ratios (e.g., Main et al., 1999). Additional analyses using other indicator methods and the SP algorithm should be performed to obtain consensus on these results.

3.5 DAY-OF-WEEK ISSUES

Since we are finding that many of the hydrocarbons measured in the PAMS program are most likely associated with motor vehicle emissions, we would expect the concentrations to be a function of driving patterns. Since traffic patterns largely depend on the Monday-Friday workweek, we might expect to see large differences in hydrocarbon concentrations as a function of day of the week. We investigated the day-of-week patterns of the concentrations and weight fractions of several hydrocarbons including n-butane, propene, i-butene, ethene, benzene, toluene, acetylene, isoprene, xylenes, 2,2,4-trimethylpentane, i-pentane, 1,2,4-trimethylbenzene, and ethylbenzene. We also investigated the day-of-week patterns of the total NMHC. Box plots by day of week for every site were used to assess patterns over all time periods, during the morning, and during midday. In order to include information regarding sample size, we used box plots with notches indicating the confidence intervals (**Figure 3-51**). In these plots, boxes are notched (narrowed) at the median and return to full width at the lower and upper confidence interval values. Significant changes with day were only observed at a few sites for a few hydrocarbons. These are summarized below:

- Bakersfield data (**Figure 3-52**) show that the morning (0500 PST) toluene and NMHC concentrations were lower on Sunday than on the weekdays in 1997. (Similar results were observed for all years combined). This may indicate that Sunday morning emissions are lower than Monday through Saturday morning emissions. On a weight percent basis, there was a less significant difference among the days indicating a similar source mixture every day.
- San Diego data (**Figure 3-53**) show that the morning (0500 PST) concentrations of NMHC, acetylene, xylenes, and toluene were lower on weekends than on weekdays. This may indicate that the weekend morning emissions are lower than the weekday emissions. On a weight percent basis, there was no difference among the days indicating a similar source mixture every day.

3.6 DISCUSSION OF SAMPLE SIZE

3.6.1 Addressing the PAMS Data Quality Objectives

California has implemented a unique approach to PAMS sampling. At most sites, four 3-hr samples are collected every third day. These samples are then analyzed for the PAMS target species. In addition, continuous total NMHC measurements are performed at each PAMS site (daily, hourly). Data quality objectives (DQOs) for the PAMS program are discussed in the PAMS technical assistance document (TAD - U.S. Environmental Protection Agency, 1998) and the PAMS implementation manual (U.S. Environmental Protection Agency, 1994). One issue of concern to California, is whether or not there are enough samples with which to meet the requirements for PAMS. We have used the California data to illustrate how some of the DQOs might be met.

DQO 1.1 The data for any given pollutant measured at a PAMS site must be able to show the presence of a diurnal pattern, if a pattern exists, with an 80 percent confidence level.

This DQO can be addressed using the data set. We need to show a change in diurnal pattern with an 80 percent confidence level when the daily mean is compared to the mean at individual sampling times. **Table 3-11** shows the statistics for Del Paso Manor in 1995. These data show about a 20 percent difference in means and 80 percent confidence intervals (CI) that barely overlap. These data appear to meet the DQO. Based on the equations for computing a confidence interval, either an increase in confidence level (e.g., 95 percent) or a decrease in sample number causes an increase in the interval. Thus, if fewer samples had been collected in the morning, for example, the confidence intervals of the data would probably overlap significantly and thus these data would probably not have met the DQO.

DQO 3.2 The speciated VOC data collected at a Type 2 site, when composited into categories, must be able to demonstrate a 20 percent change in the seasonal average between two consecutive years with an 80 percent confidence level.

This DQO can also be addressed using the data set. **Figure 3-54** shows a t-test comparing benzene concentrations at Los Angeles between 1996 and 1997. Although the Los Angeles North Main site is not a PAMS site, the data are representative of a Type 2 site and the sampling frequency is similar to PAMS sites in California. In addition, these data happened to meet the DQO information discussed here. These data show that there was a 21 percent difference between the mean benzene concentrations in 1996 and 1997. The t-test p value of 0.208 indicates that there was a 20.8 percent chance of observing a difference as large as this even if the population means were identical. These data are close to meeting the DQO with 18 samples in 1996 and 25 samples in 1997.

To further expand upon this DQO, we explored the sensitivity of a comparison between two years with a large difference in means to the number of samples included in the comparison. We computed the mean benzene concentrations and 95 percent confidence

intervals for morning data collected in Los Angeles in 1995 and 1996. The original 1995 data set contained 23 benzene values. We then selected 13 of these values and recomputed the mean and confidence intervals. We repeated this for only five samples. For each case, we performed a t-test to compare the means from 1995 and 1996 to see if they were statistically significantly different. The process described here is shown in **Figure 3-55**. Box plots of the data shown in Figure 3-55 are shown in **Figure 3-56**.

In the example, decreasing the sample size increases the confidence interval. In general, if the 95 percent confidence intervals of two groups do not overlap, then there is good evidence that the group population means are different. If there is considerable overlap, then the population means of both groups might be the same. The t-test results indicate that the full 1995 and 1996 data sets are statistically significantly different. However, the means from the smaller 1995 data sets are not necessarily different from the 1996 data. This example was for a rather large (32 percent) difference between the means from the two years. These data show that the number of samples was adequate to observe a significant change in benzene concentration between 1995 and 1996 and that fewer samples would probably not be sufficient. Note that the change in benzene concentration was very large and that to observe a smaller change, such as the 3 percent change in total VOC concentrations required in DQO 3.1 (not listed here), more samples would probably be required.

3.6.2 Other Considerations

While the actual sample size for the annual changes in benzene appeared adequate, we need to consider other analyses that can be performed using the hydrocarbon data and the effects of sample size on those analyses:

- *Emission inventory evaluations.* When the VOC data are compared to the emission inventory, we need an adequate number of samples in order to assess the comparisons for different wind quadrants. We also need a sufficient number of samples above selected concentration thresholds (e.g., 100 ppbC NMHC) to ensure that we are comparing the inventory to fresh emissions well above background concentrations.
- *Day-of-week evaluations.* Currently, for an every-third-day sampling regimen, most of the California sites will have only three or four samples per day of the week in an ozone season. As shown in the example above, this will not be an adequate number of samples with which to investigate day-of-week issues. Combining several years of data for this type of analysis is probably not feasible given the significant year-to-year changes observed (e.g., benzene).
- *Receptor modeling.* To assess the possible contributions from various sources of hydrocarbons, particularly by wind quadrant, it is likely that there are insufficient samples on an every-third-day basis. The number of samples necessary for the analysis will be a function of the estimated error in the source profiles and the standard deviation in the ambient data. Based on the results from other studies, it is likely that daily samples are required for this type of analysis.

3.6.3 Summary

We used examples from the data to illustrate two of the PAMS data quality objectives. The examples discussed in this section seem to indicate that 20 to 25 samples are the minimum number of samples required to observe significant changes in an individual hydrocarbon between years or to observe a diurnal pattern. Statistically, one can determine the number of samples required to observe a given percent change using a power calculation. The results of the power calculation will give the minimum change that can be statistically detected given the sample size. The calculation is a function of the sample mean, sample standard deviation, confidence interval, and expected change and is discussed by Walpole and Myers (1989) and other statisticians. For all analyses, the adequate number of samples is a function of the magnitude of the expected change, the site, the individual hydrocarbon or ratio, and the time of day (e.g., more variability was observed in the midday samples than the morning samples).

Sample size assessments should be made for data from each site for selected analyses (e.g., day-of-week issues and annual trends) and for several parameters (e.g., NMHC, benzene, butane and the benzene-to-toluene ratio) to more fully assess the DQOs.

3.7 CONTINUOUS NMHC DATA

We investigated the continuous NMHC data at Clovis to address two issues:

- Are the concentrations measured on PAMS sampling days (e.g., every third day) representative of all summer days?
- How do the 8202 (hourly) and PAMS (3-hr canister) NMHC concentrations compare?

To address the first question, we investigated the early morning data (0500, 0600, 0700 PST). Morning data are probably most representative of the nearby emission sources and have been used in other analyses in this report. We prepared summary statistics for the following data sets:

- All hourly data collected year round from June 1, 1995 through November 30, 1997 (Hrly All).
- The 3-hr average data computed from the hourly values collected year round from June 1, 1995 through November 30, 1997 (3 hr All).
- Hourly data collected from July through September, 1995-1997 (Hrly Summer).
- The 3-hr average data computed from the hourly values collected from July through September, 1995-1997 (3 hr Summer).
- The 3-hr average data computed from the hourly values collected on days for which PAMS NMHC data exist (3 hr matching). Only days for which both data sets were available are considered.
- Canister, 3-hr average NMHC values form the PAMS data set (PAMS NMHC).

The morning mean values and 95 percent confidence limits are plotted in **Figure 3-57** for all data sets. Summary statistics are provided in **Table 3-12**. In comparing the entire hourly data set (Hrly All) and the 3-hr averages from the same data set (3hr All), note that the confidence interval is larger for the 3-hr averaged data. The mean and median values are similar between the two data sets while the standard deviation and coefficient of variation for the 3-hr averages are lower. The same trend is found in a comparison of the summer hourly data and 3-hr averages (Hrly Summer and 3hr Summer). Averaging tends to smooth the data and reducing sample size increases confidence intervals. These results are as expected. Other observations include the following:

- The summer morning concentrations are lower than the year-round concentrations. We have observed this seasonal variation in the past.
- The PAMS concentrations are lower than the hourly measurements. This is not a surprising finding as there are analytical differences between the two measurements. For example, a dryer is used prior to analysis of the PAMS samples; Nafion dryers have been shown to remove oxygenates and some hydrocarbons (Gong and Demerjian, 1995).
- For the PAMS sampling days, the 3-hr average NMHC concentrations (from the continuous data) have significantly lower mean and median values than the entire summer sampling period. This means that for the data available at Clovis (continuous data were not available for every PAMS sample), NMHC concentrations on PAMS days do not appear to represent the summer morning concentrations. Further investigation using other times of day and other sites should be made to see if this is a general conclusion or a function of the missing continuous data.
- Overall, the PAMS NMHC concentrations correlate relatively well with the 3-hr averaged continuous data ($r^2 = 0.63$). When the data are inspected by sampling period, however, the correlation between the two measurements is poor for the 1200-1400 PST and 1600-1800 PST sampling periods (see Appendix D).

3.8 SOURCE APPORTIONMENT

3.8.1 Overview

Receptor models are used to resolve the composition of ambient VOC into components related to emission sources. A commonly used receptor model, the chemical mass balance (CMB) model, uses an effective least squares solution to a set of linear equations which express each measured chemical species concentration as a linear sum of products of source profile species and source contributions. Model inputs include source profiles (i.e., fractional amount of hydrocarbons in the VOC emissions from each source type), receptor (ambient) concentrations, and realistic uncertainties for both the source and receptor values. Uncertainties are used to weigh the relative importance of input data to model solutions and to estimate uncertainty of the source contributions. Model output includes the contributions

(quantitative) from each source type to the total ambient VOC and individual hydrocarbons, the uncertainty in the estimates, and performance parameters. Though the CMB solution is identical to some statistical inference methods, it is not entirely dependent on statistical principles. The basic model equations that represent the source receptor relationship, the effective variance weighting, and the error propagation are all based on physical principles that are then incorporated into a statistical estimation procedure. Pertinent references for CMB modeling include: Fujita et al., 1995; Fujita et al., 1994; Gertler et al., 1996; Henry, 1992; Lin and Milford, 1994; Lu and Fujita, 1995; Main et al., 1997; Pace and Watson, 1987; Scheff and Wadden, 1993; and U.S. Environmental Protection Agency, 1987.

As a part of this project, we prepared ambient data files for input to CMB8.0 for all Type 2 and Type 2-like sites. These ambient data files were created using a template created with Microsoft Access database software. The CMB8.0 preferred file-naming conventions were used. All species abbreviations used in the ambient data records are identical to those found in the species selection file, SOINV.SEL, which is included with the CMB8.0 input package. The species abbreviations and corresponding names can be found in Appendix E.

We tested the files with source profiles available to us to make sure that the files operated correctly; however, these source profiles were not suitable for use in California. Up-to-date and location specific source profiles will need to be prepared in order to perform a high quality source apportionment study using the PAMS data. Selection of proper source profiles is critical to a defensible source apportionment analysis. Recent CMB analyses performed in California indicated the need for more up-to-date source profiles (Magliano, 1998).

3.8.2 Uncertainties

Estimation of uncertainty in the ambient data is a function of the minimum detection limit, coefficient of variation of measurements, and the concentration. An example method for estimating uncertainty is provided below (Fujita et al., 1994):

$$\sigma\,(C) = ((2MDL)^2 + (COV*C)^2)^{1/2}$$

where:

σ	=	root mean square error for concentration value (C)
MDL	=	minimum detection limit for auto-GC (0.1 to 0.2 ppbC)
COV	=	coefficient of variation of measurements (5 to 10 percent)
C	=	concentration

In the files that we created, we assumed a 15 percent uncertainty for all hydrocarbons at all sites. This uncertainty estimate may be somewhat low for some hydrocarbons but should serve as a reasonable starting point for exploration of the CMB model. Once suitable source profiles are available for each District, more representative uncertainty estimates for the ambient data should be computed based on the District's actual MDL and COV values for each hydrocarbon.

3.8.3 Model Limitations

In future CMB model analyses, analysts need to be fully aware of the uncertainties and limitations of the model. We have provided references to several papers and reports that provide excellent and thorough discussions of receptor model limitations. Three key limitations are as follows (Fujita et al., 1994):

- Many emitters have similar species composition profiles. This means that one cannot directly distinguish between gasoline vapor from a vehicle and from a storage tank at a refinery as both emissions have the same source profile. However, this distinction may be inferred using wind direction.
- Species composition profiles change between source and receptor. This means that one cannot currently account for changes due to photochemistry; we know that hydrocarbons do react during transport from a source to a receptor. Modelers minimize this limitation by using less reactive hydrocarbons in the model. Research into ways to incorporate reactivity is also being performed.
- Receptor models cannot predict the consequences of emissions reductions. This means that one cannot take estimated source profiles resulting from changes in emissions and predict ambient concentrations. Urban airshed modeling or other more detailed modeling is required to assess the effects of emission changes.

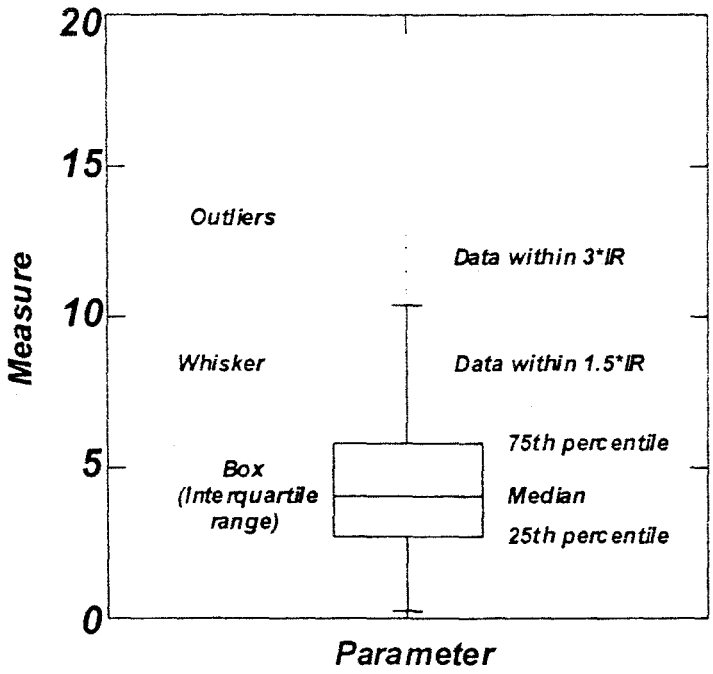


Figure 3-1. Annotated box-whisker plot with outliers.

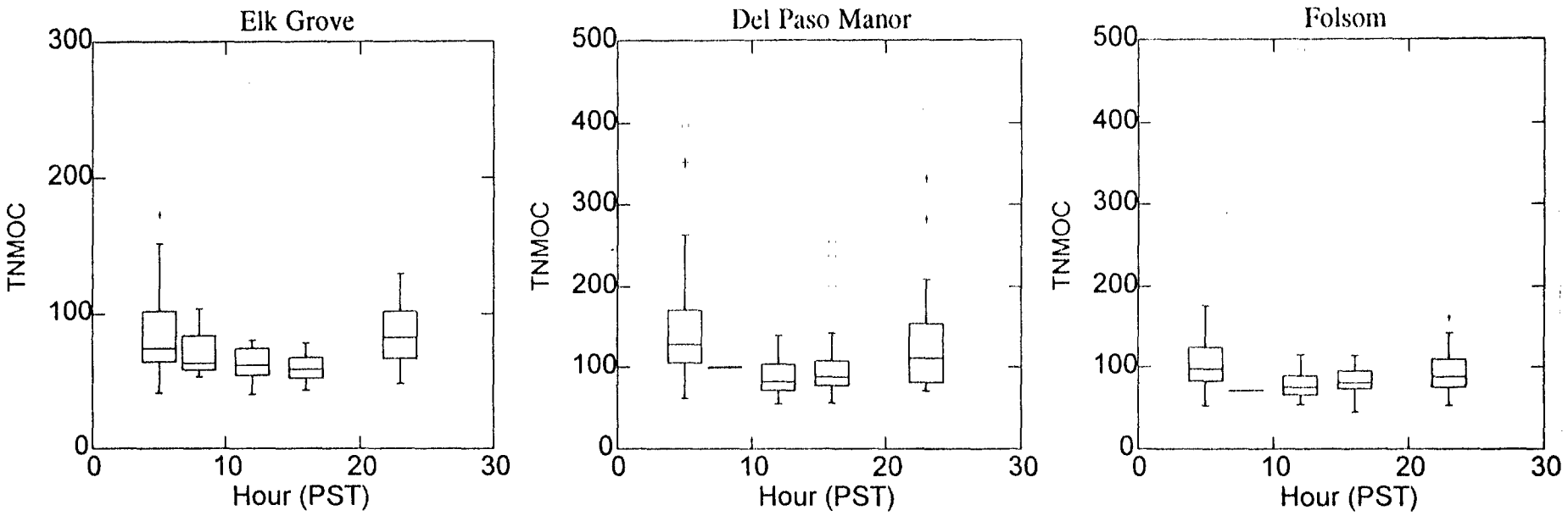


Figure 3-2. Box plots of total non-methane hydrocarbon concentrations (ppbC) by site and time of day (PST) in 1997. Some data points were excluded in order to show sufficient detail. The concentration scale may vary between sites.

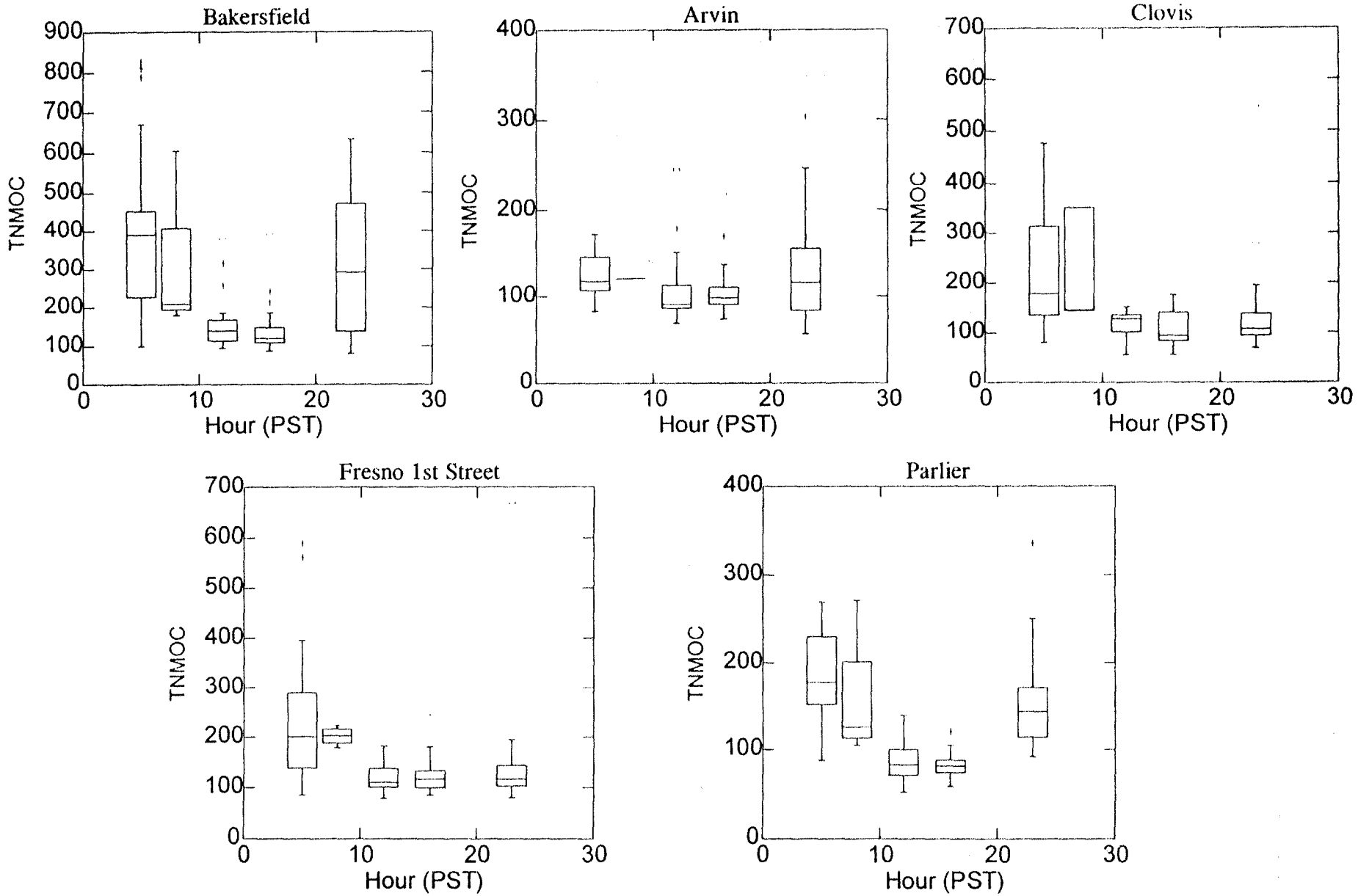


Figure 3-3. Box plots of total non-methane hydrocarbon concentrations (ppbC) by site and time of day (PST) in 1997. Some data points were excluded in order to show sufficient detail. The concentration scale may vary between sites.

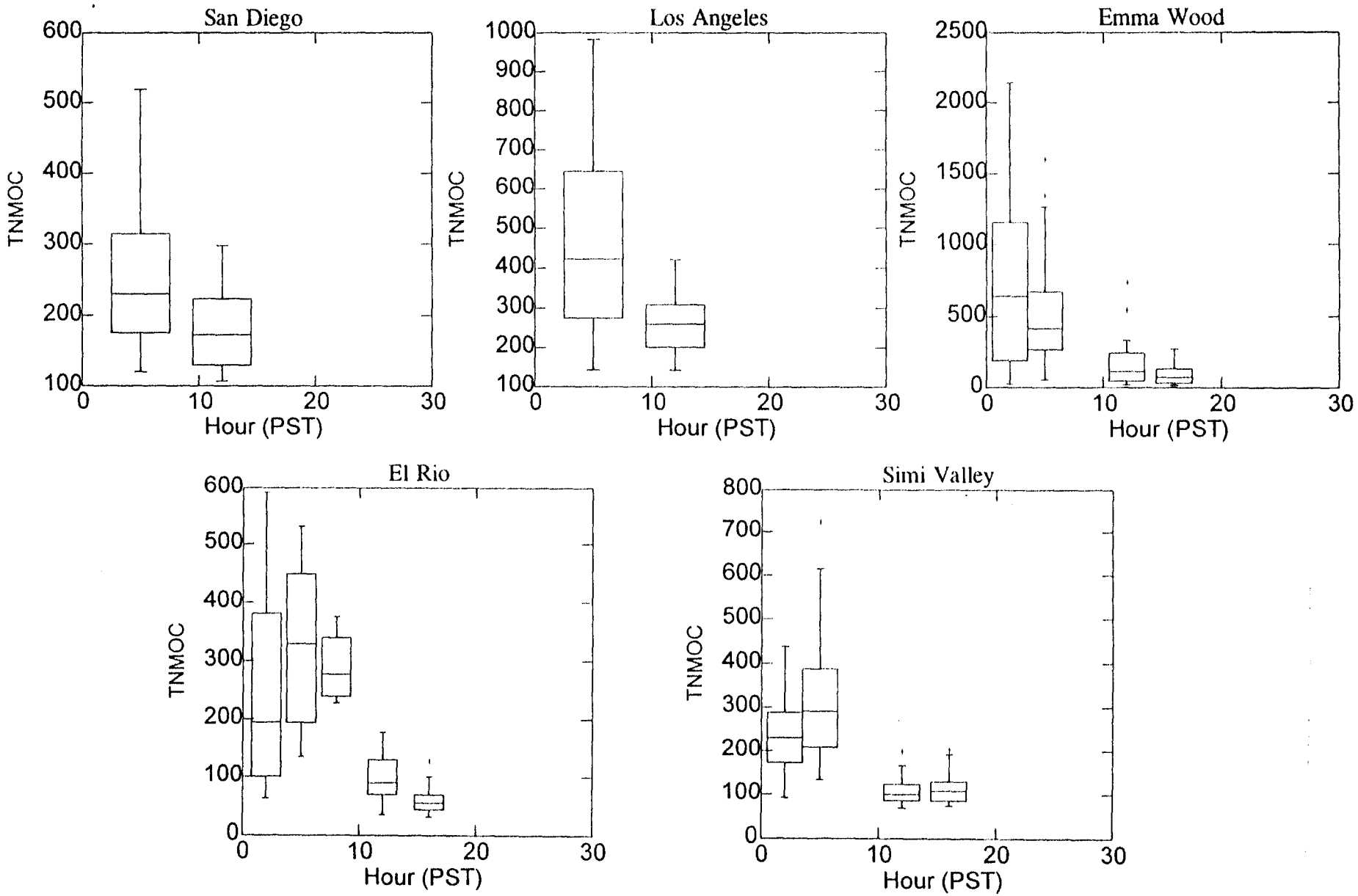


Figure 3-4. Box plots of total non-methane hydrocarbon concentrations (ppbC) by site and time of day (PST) in 1997 (1996 data used for El Rio) . Some data points were excluded in order to show sufficient detail. The concentration scale may vary between sites.

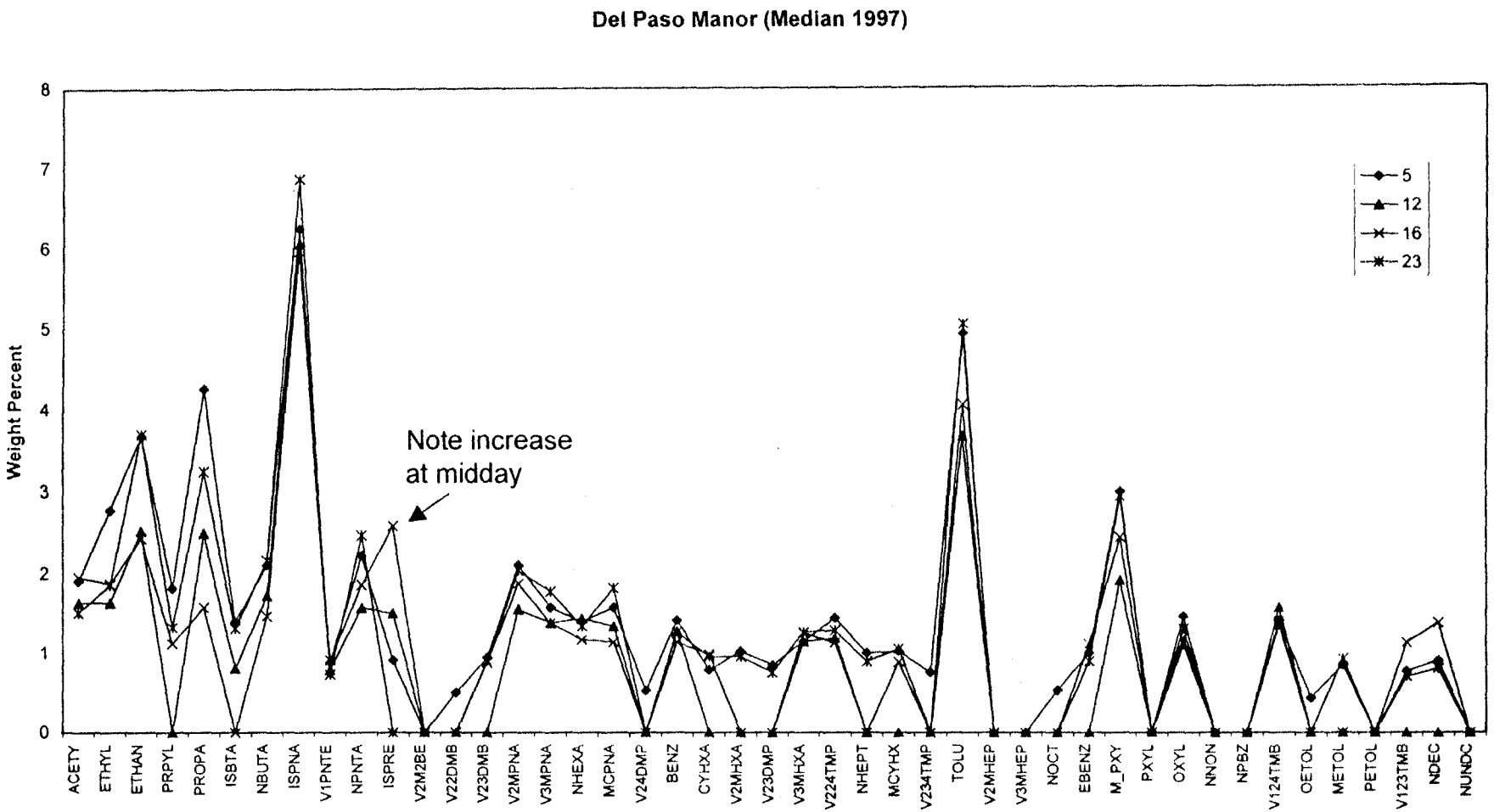


Figure 3-5. Median composition at Del Paso Manor in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

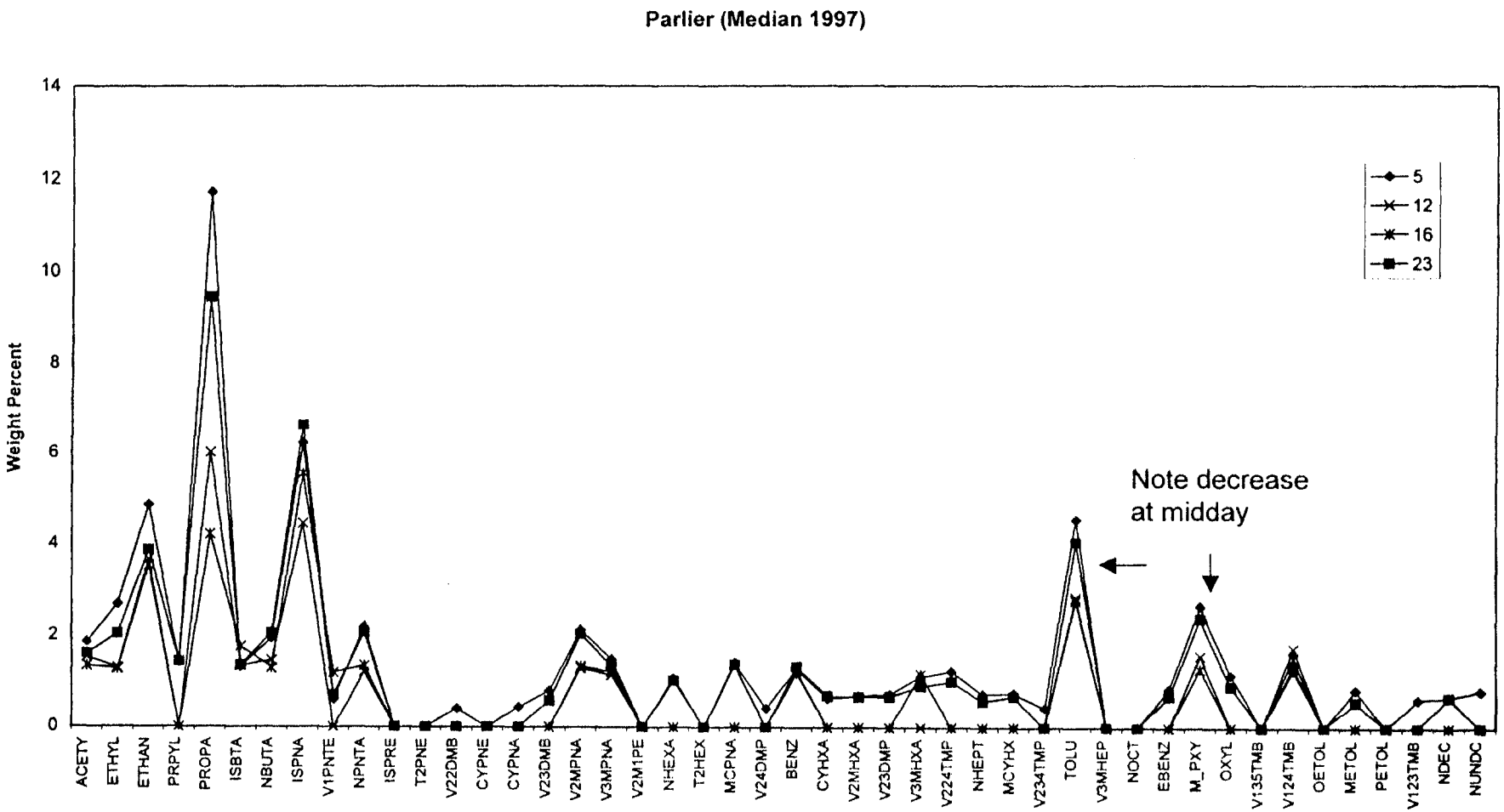


Figure 3-6. Median composition at Parlier in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

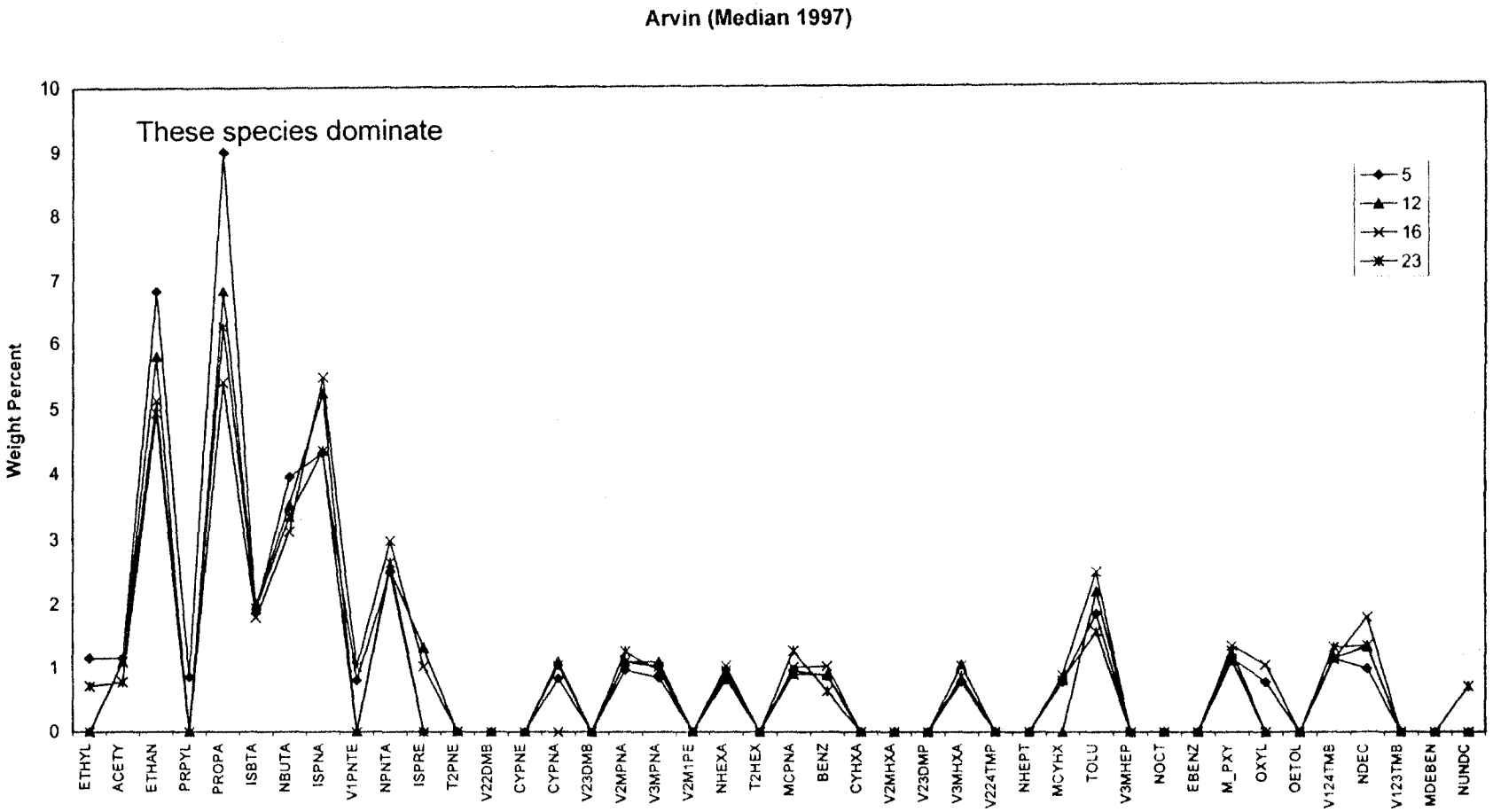


Figure 3-7. Median composition at Arvin in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

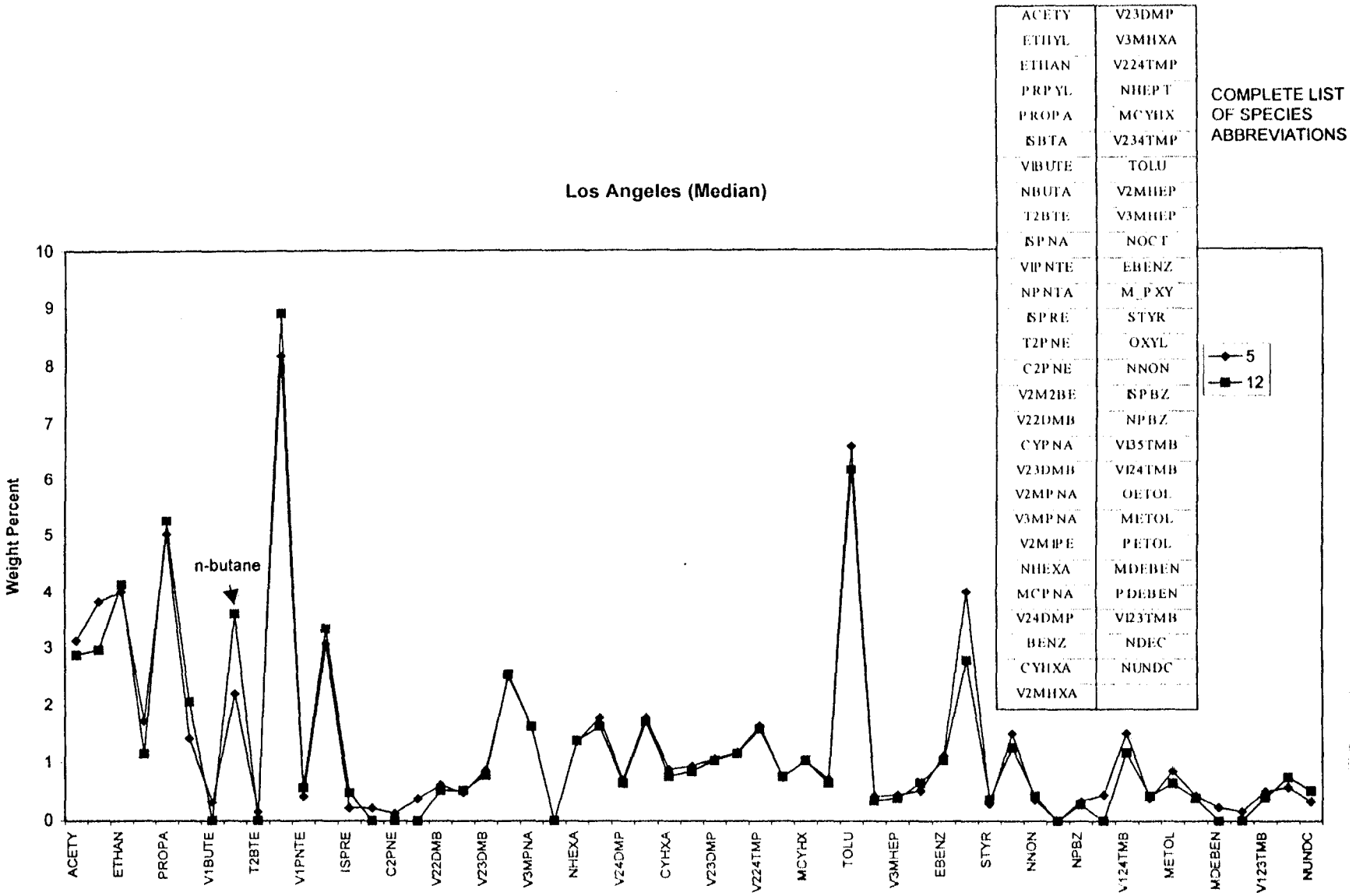


Figure 3-8. Median composition at Los Angeles in 1997 at 0500 and 1200 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

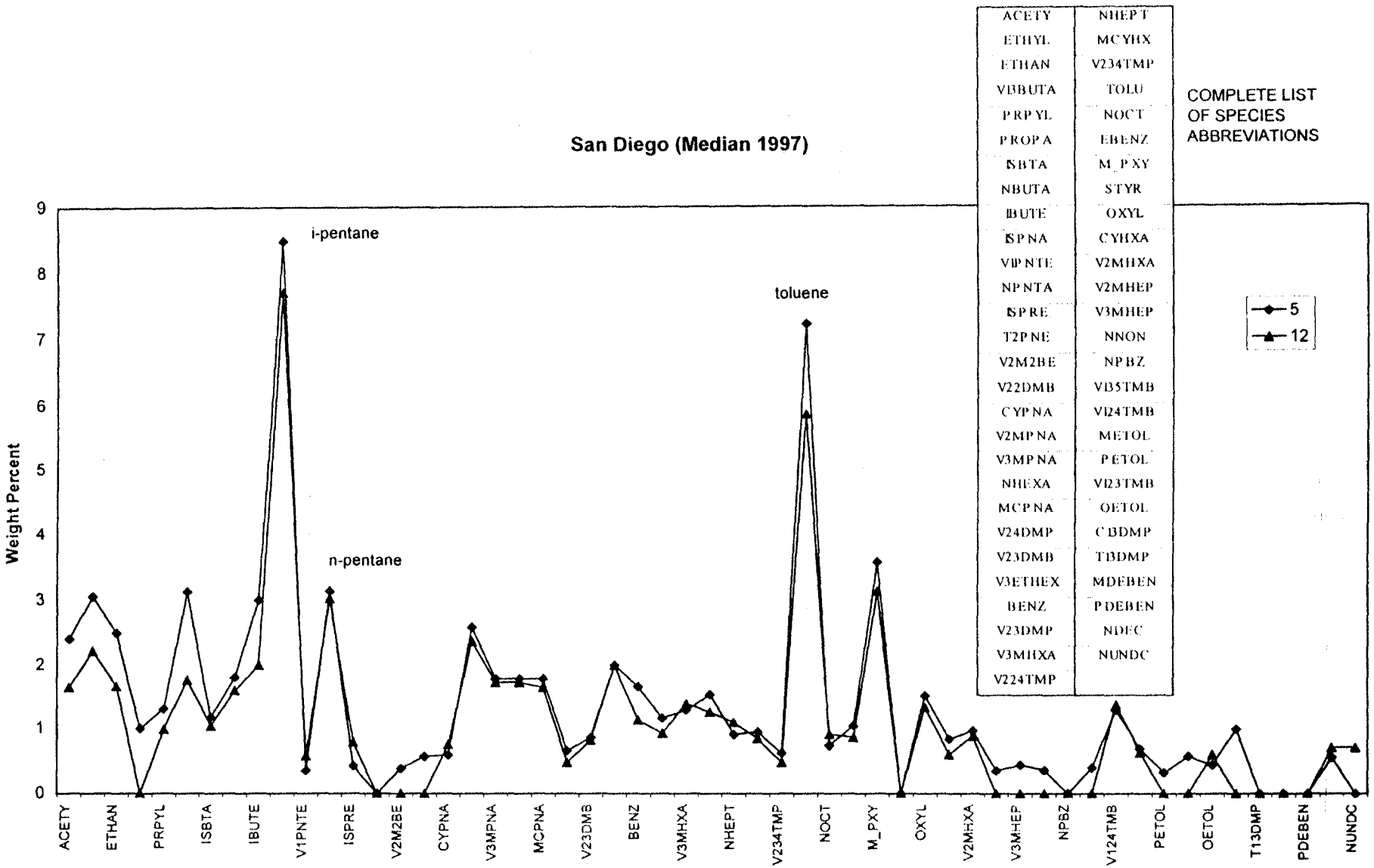


Figure 3-9. Median composition at San Diego in 1997 at 0500 and 1200 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

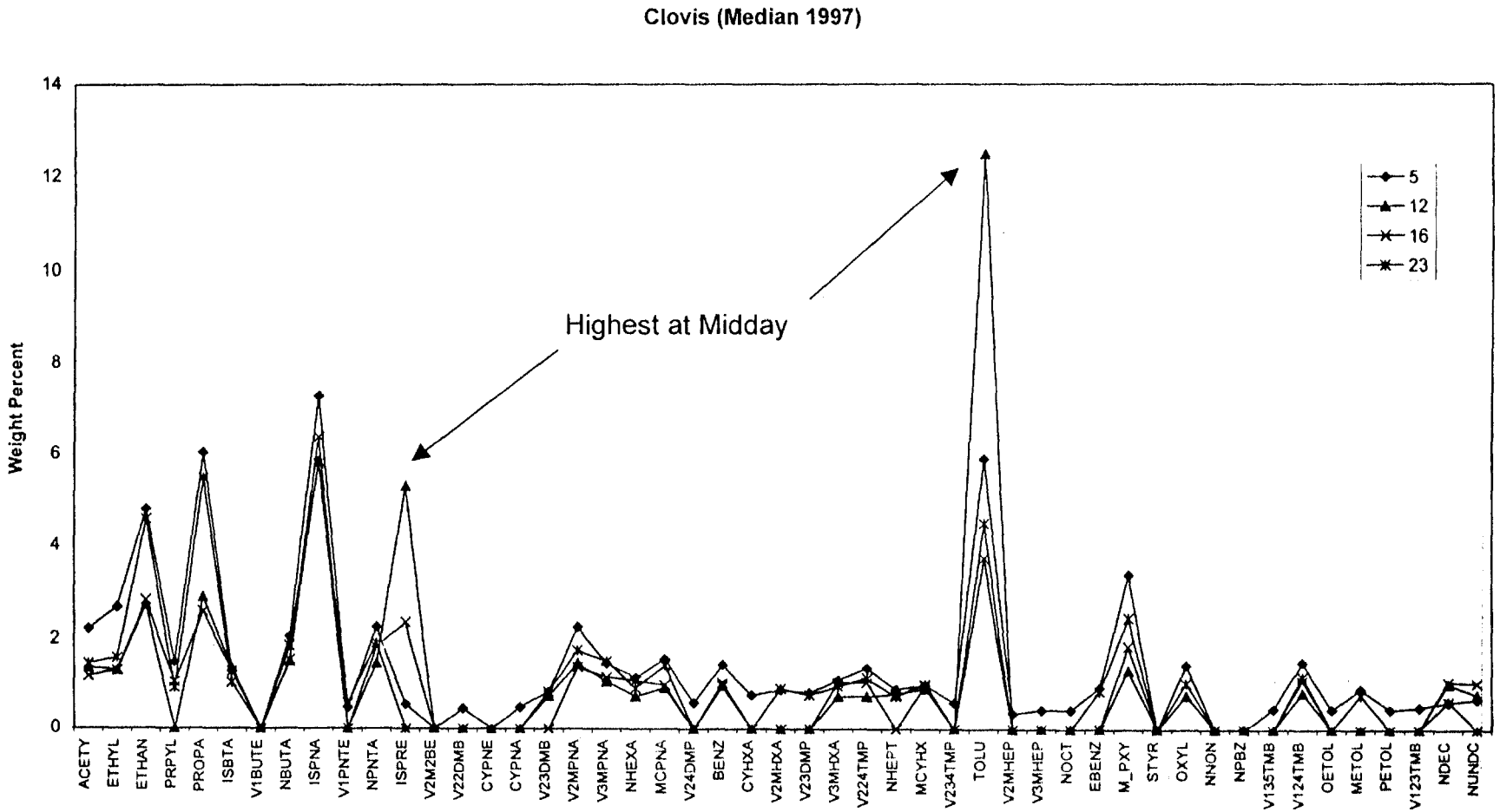


Figure 3-10. Median composition at Clovis in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

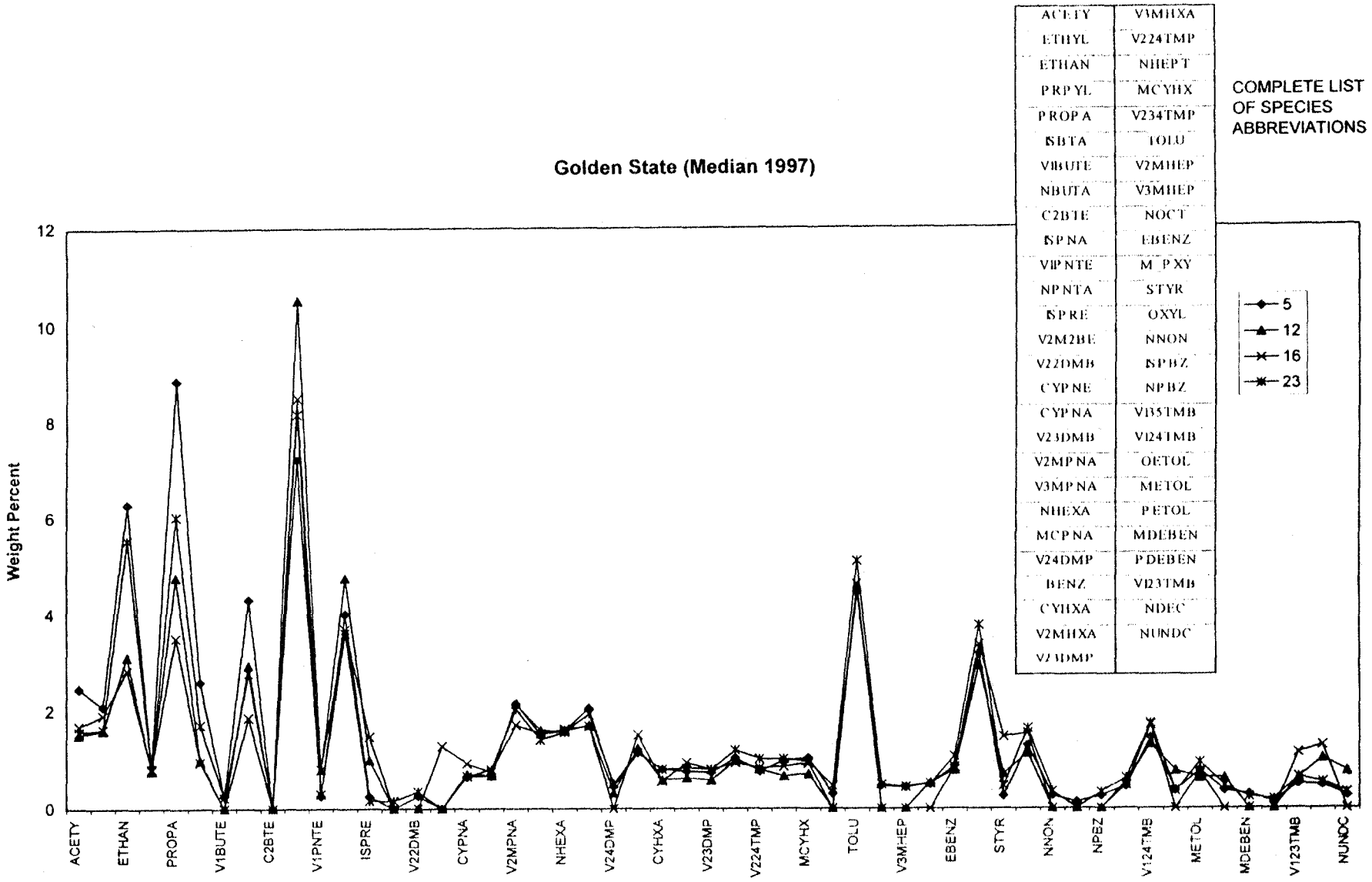


Figure 3-11. Median composition at Bakersfield (Golden State) in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

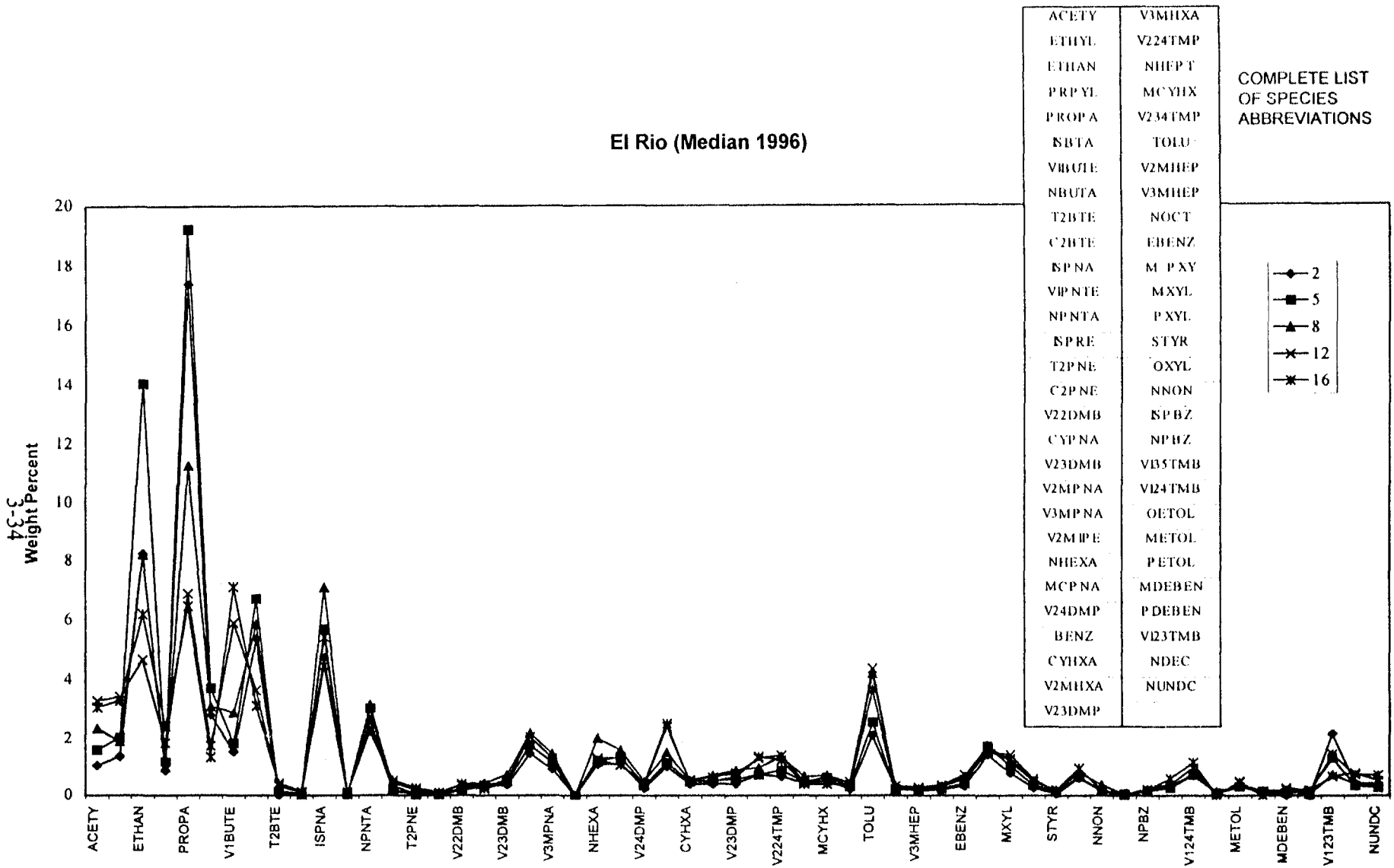


Figure 3-12. Median composition at El Rio in 1996 at 0200, 0500, 0800, 1200, and 1600 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

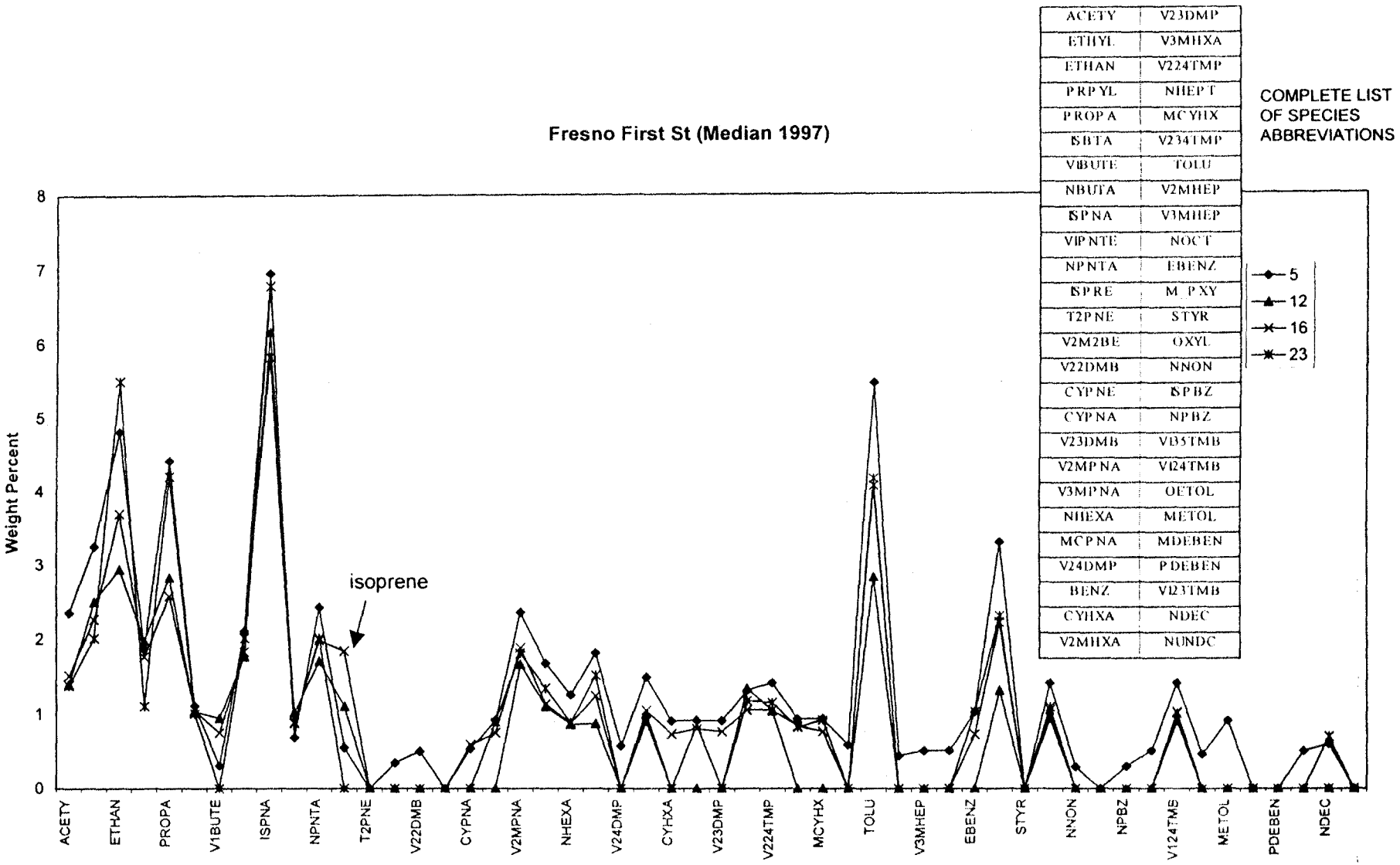


Figure 3-13. Median composition at Fresno in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

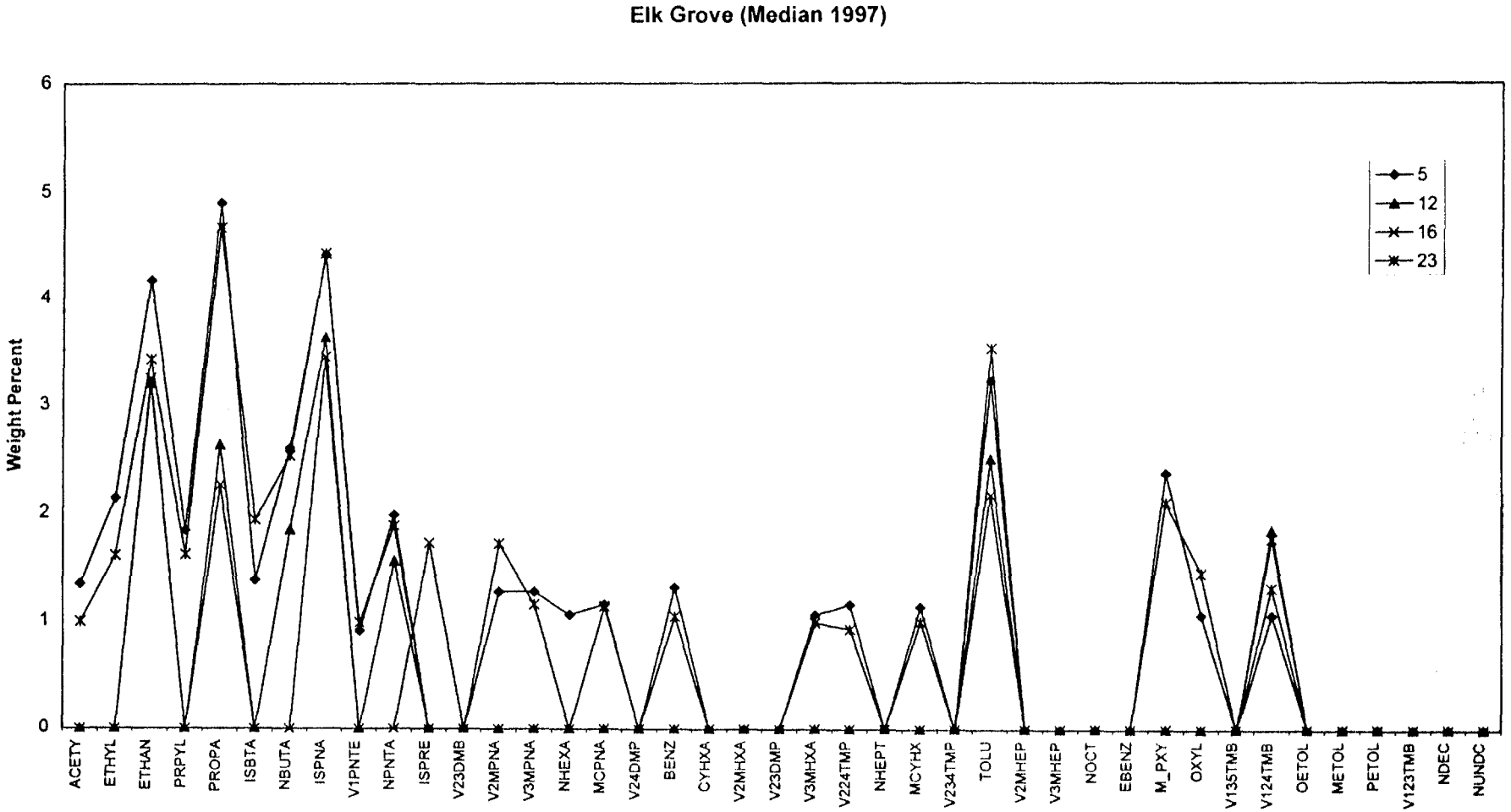


Figure 3-14. Median composition at Elk Grove in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

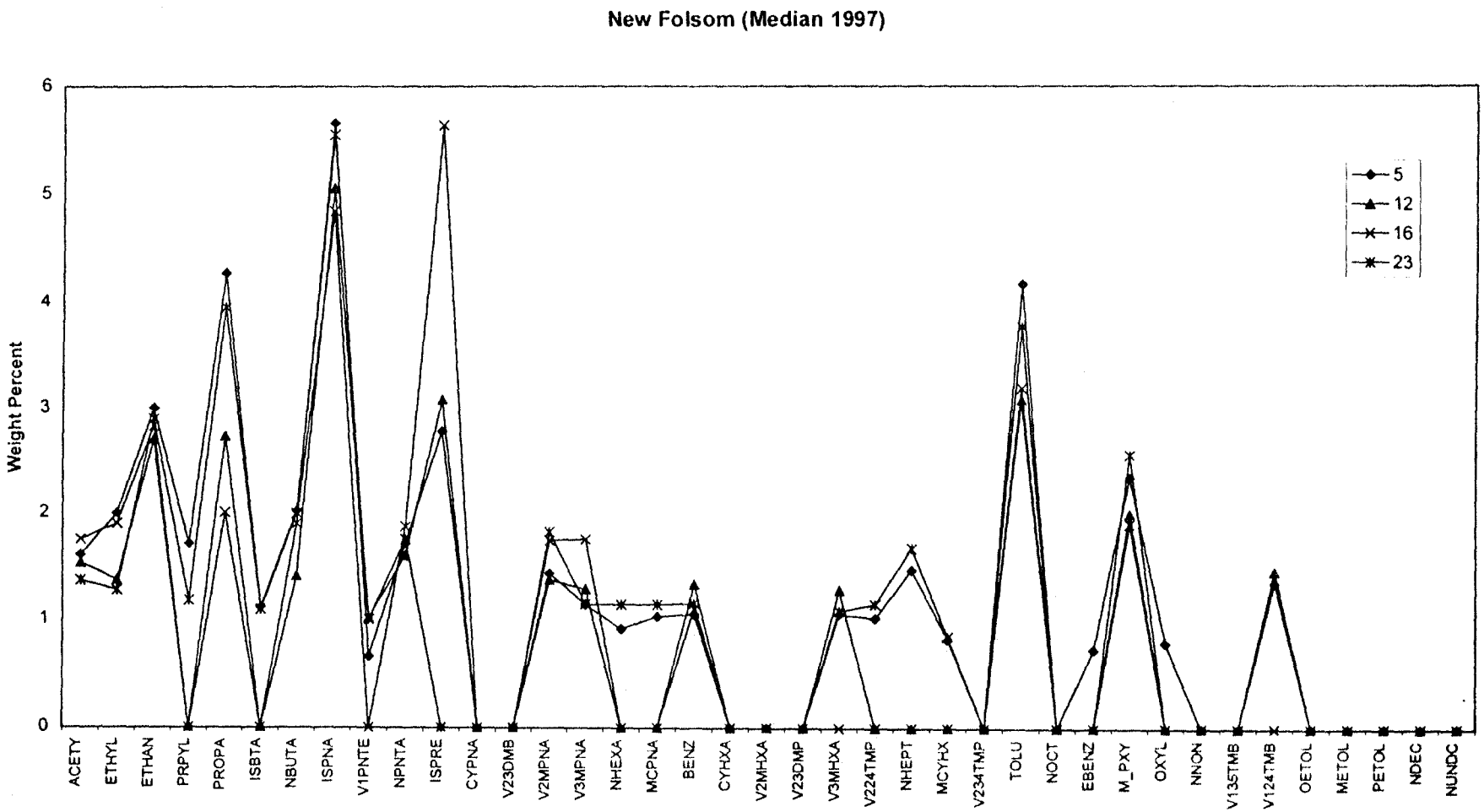


Figure 3-15. Median composition at New Folsom in 1997 at 0500, 1200, 1600, and 2300 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

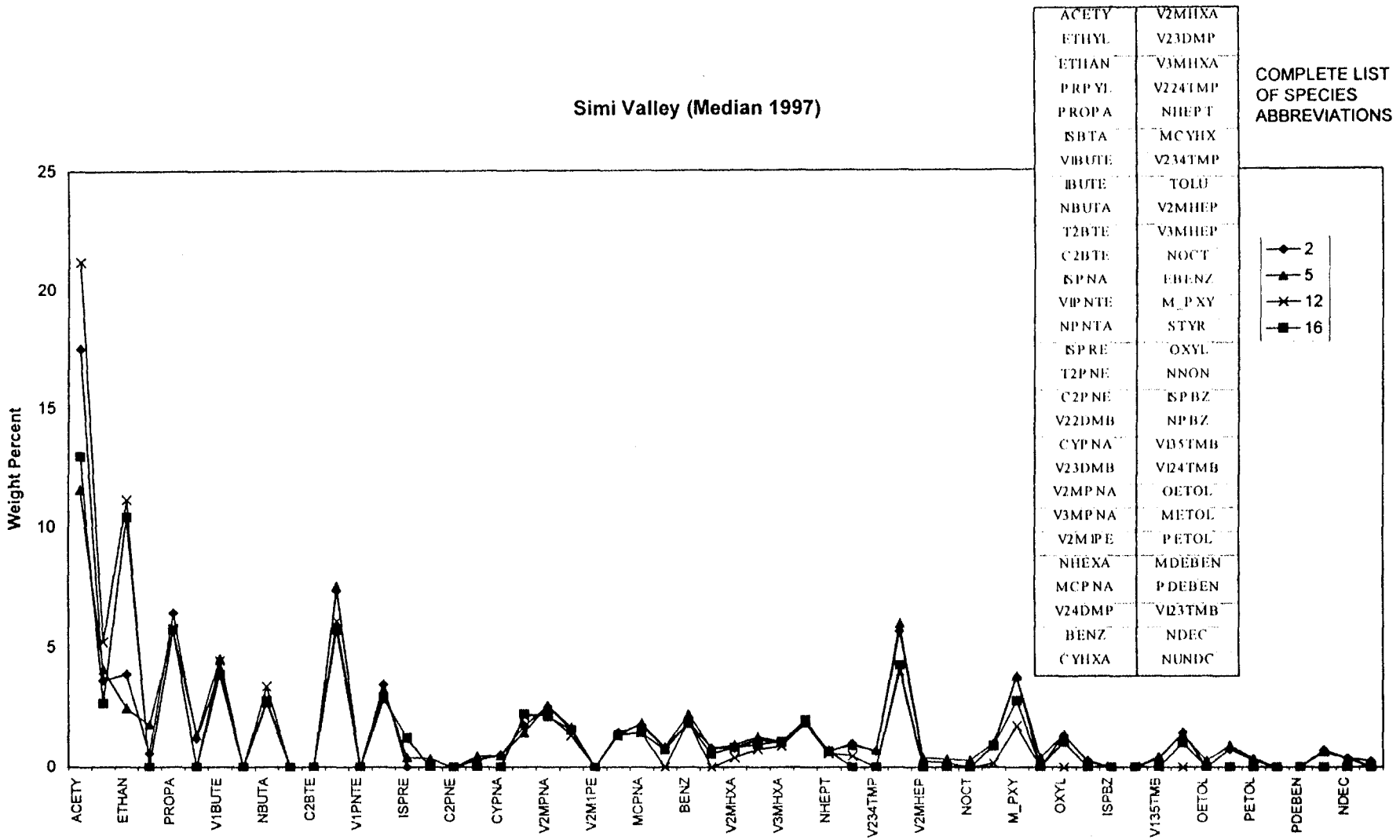


Figure 3-16. Median composition at Simi Valley in 1997 at 0200, 0500, 1200, and 1600 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

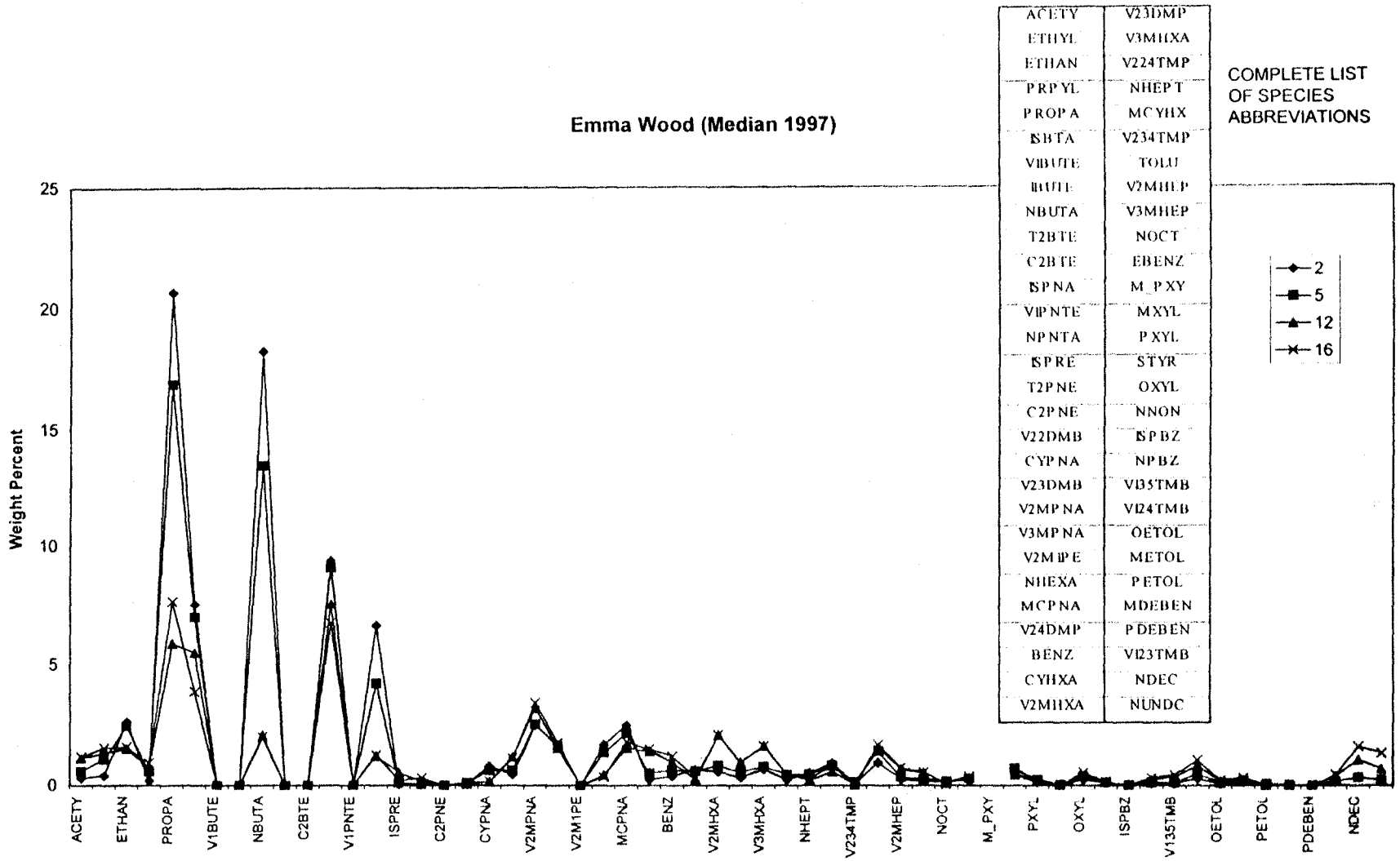


Figure 3-17. Median composition at Emma Wood in 1997 at 0200, 0500, 1200, and 1600 PST. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

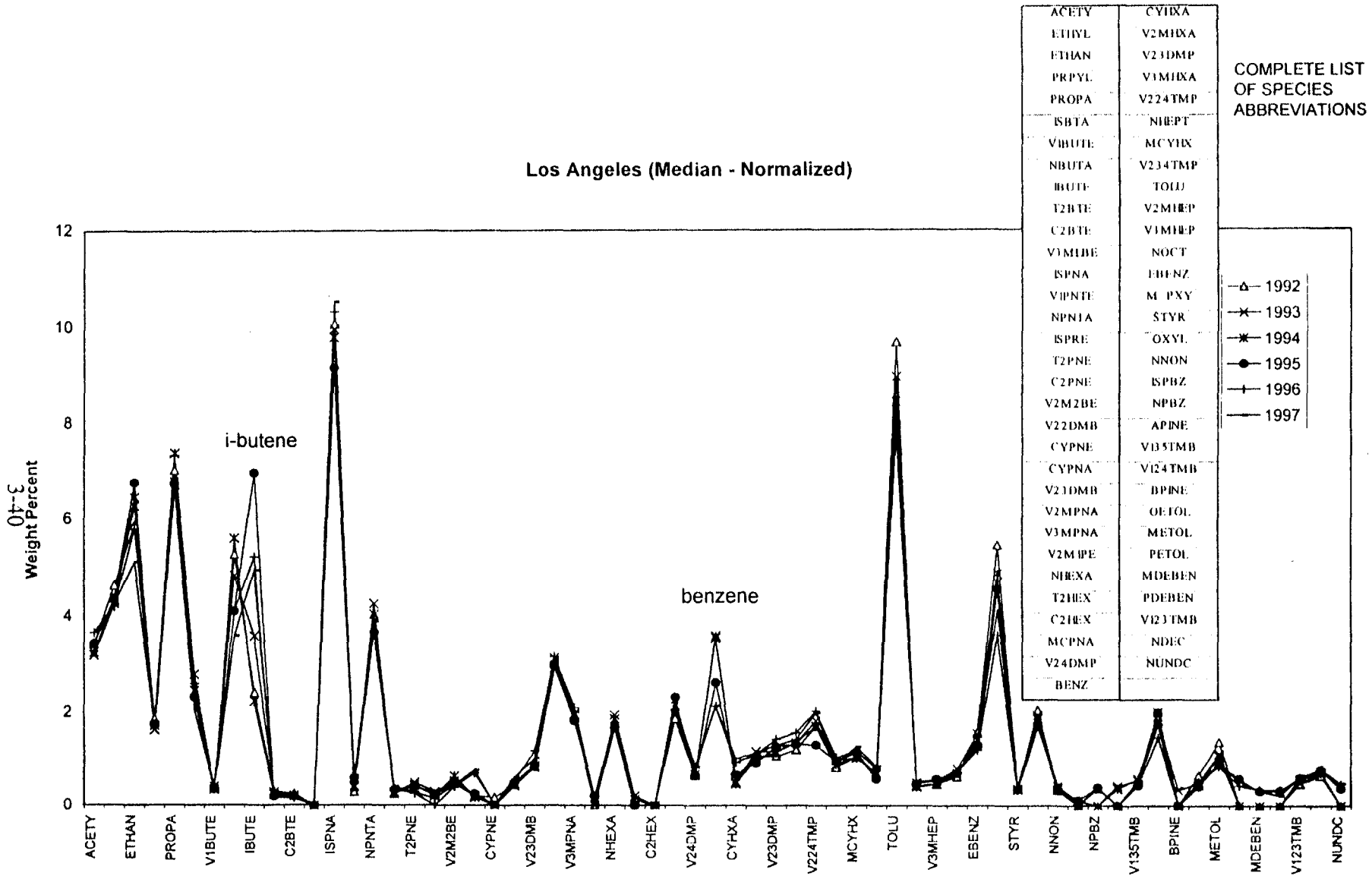


Figure 3-18. Median composition at Los Angeles from 1992 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

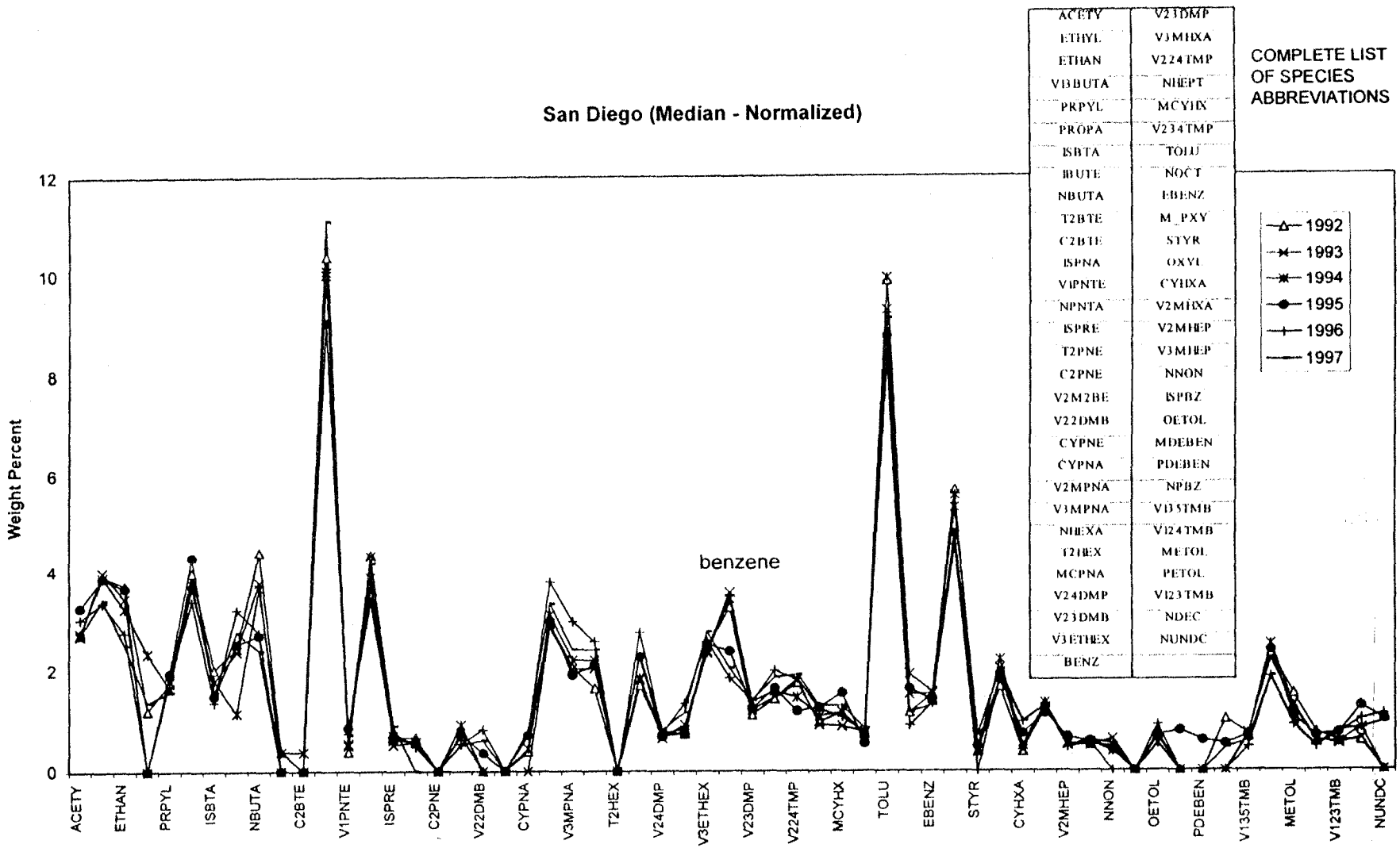


Figure 3-19. Median composition at San Diego from 1992 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

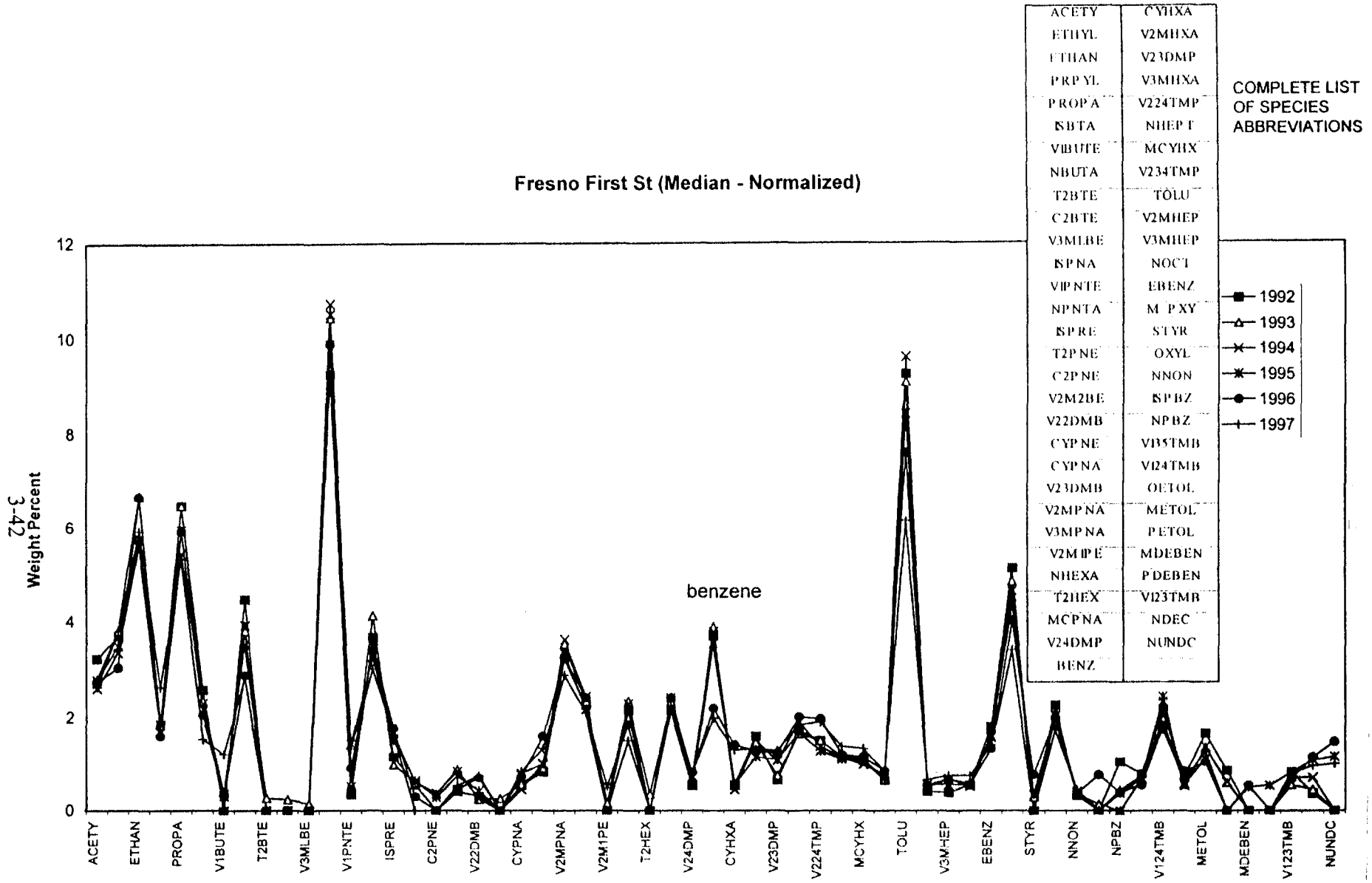


Figure 3-20. Median composition at Fresno from 1992 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

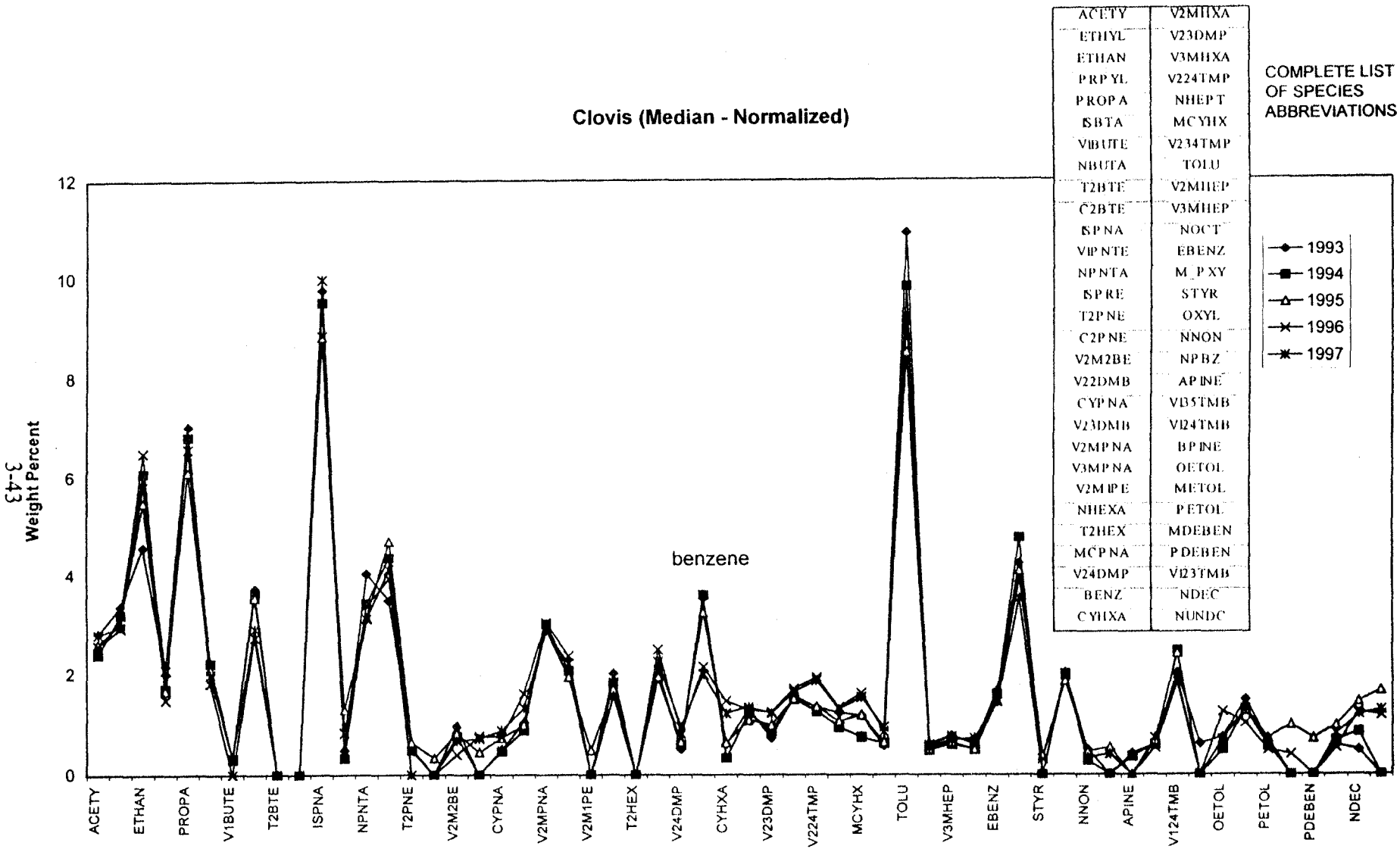


Figure 3-21. Median composition at Clovis from 1993 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

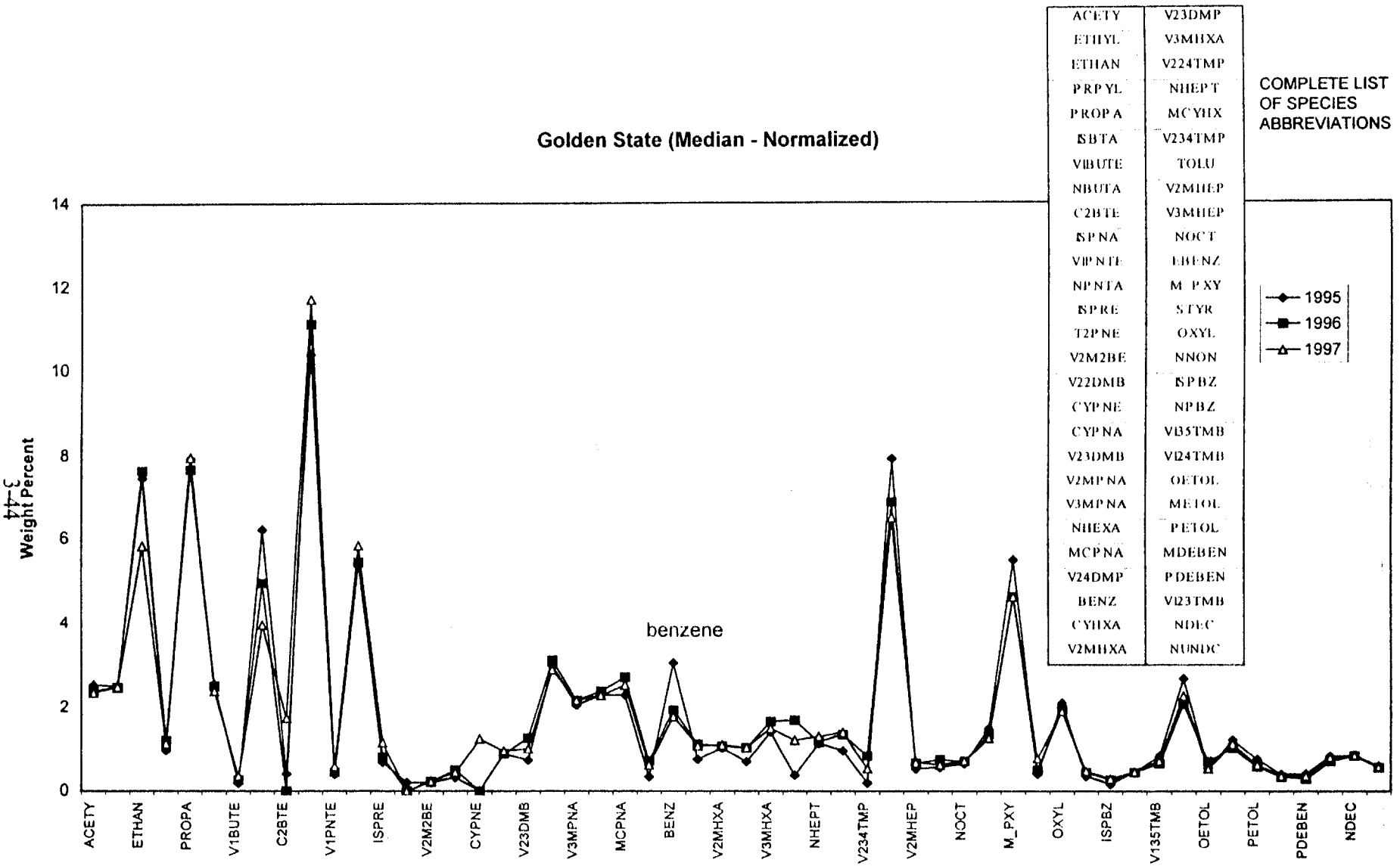


Figure 3-22. Median composition at Bakersfield (Golden State) from 1995 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

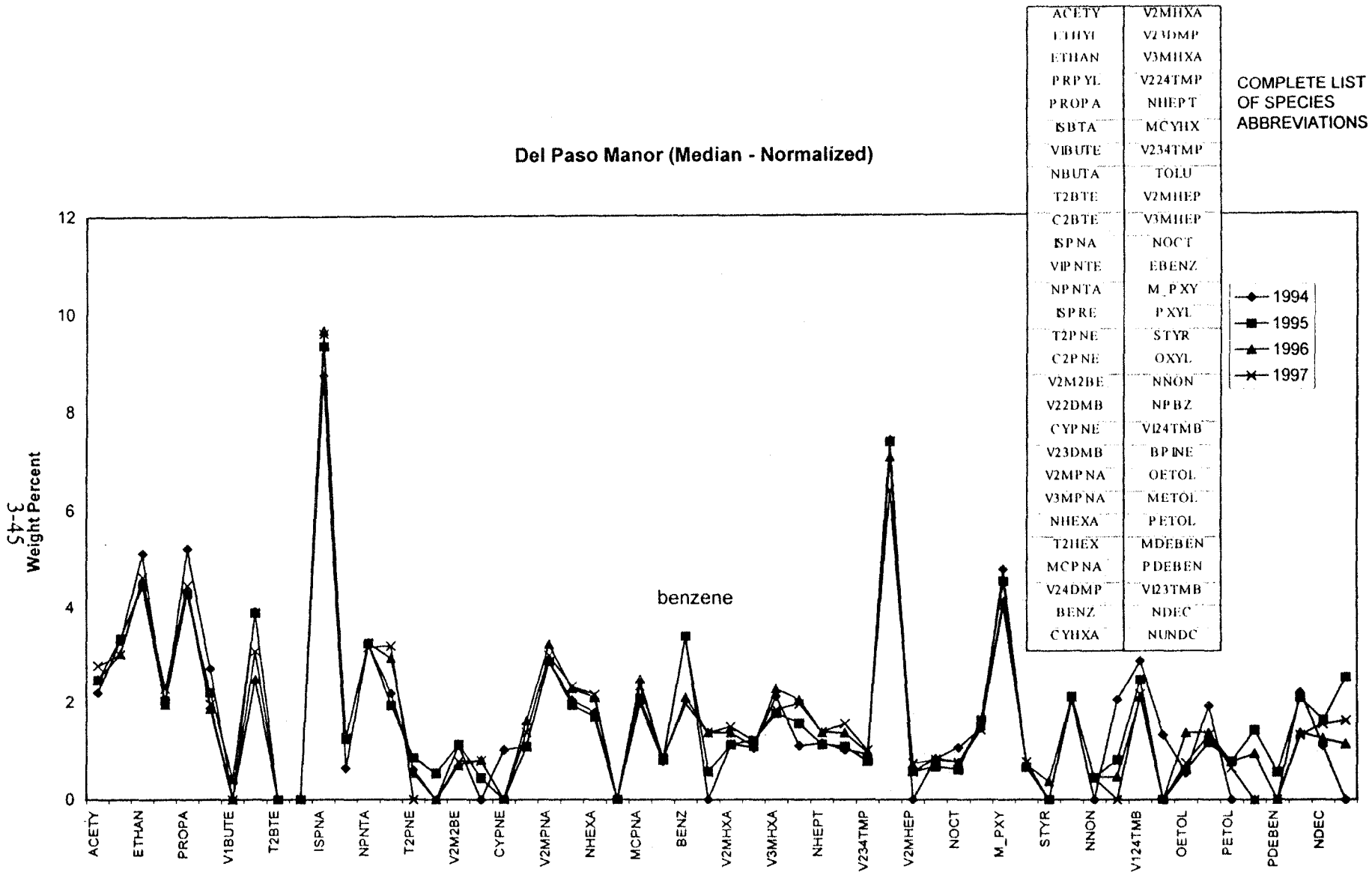


Figure 3-23. Median composition at Del Paso Manor from 1994 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

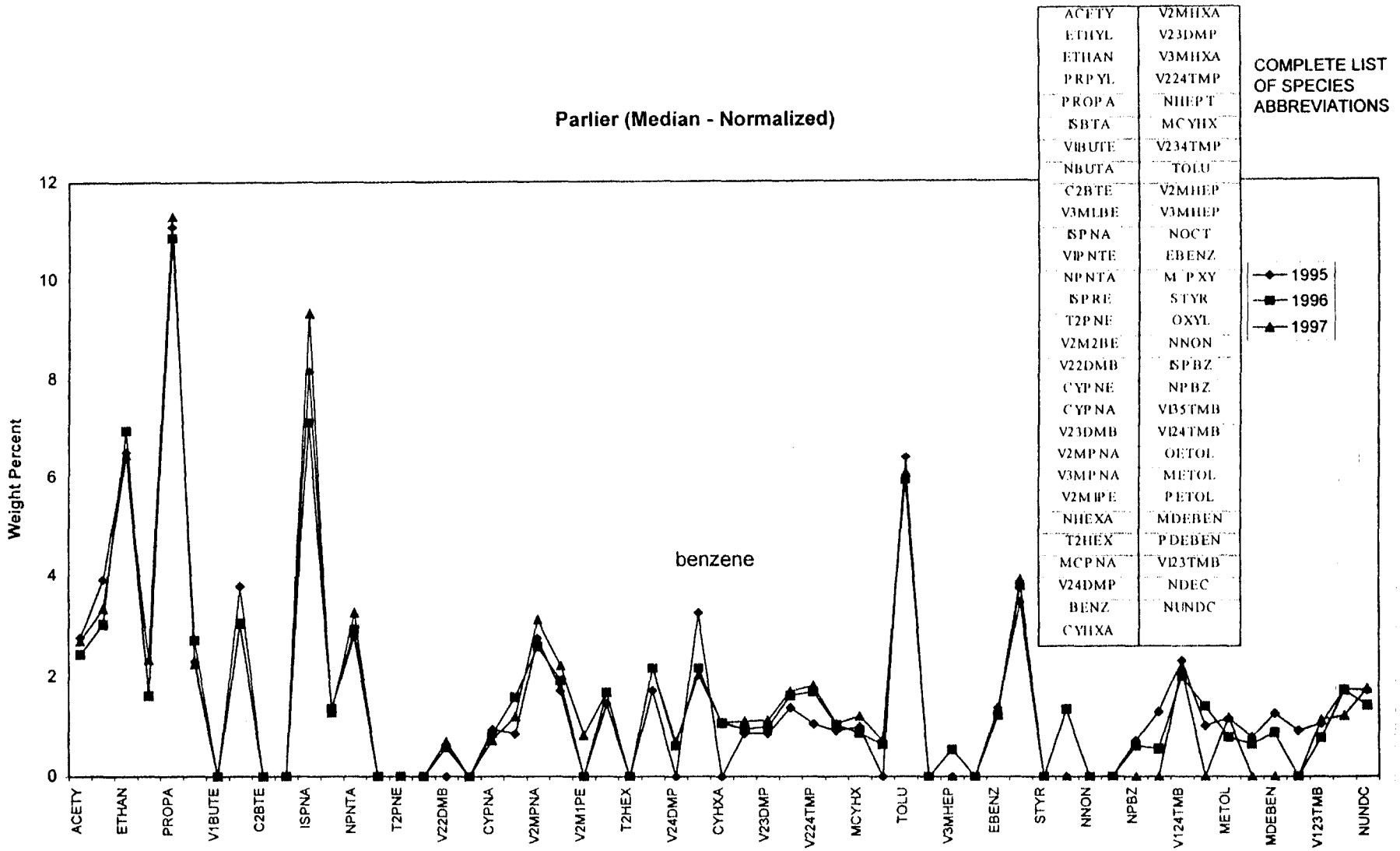


Figure 3-24. Median composition at Parlier from 1995 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

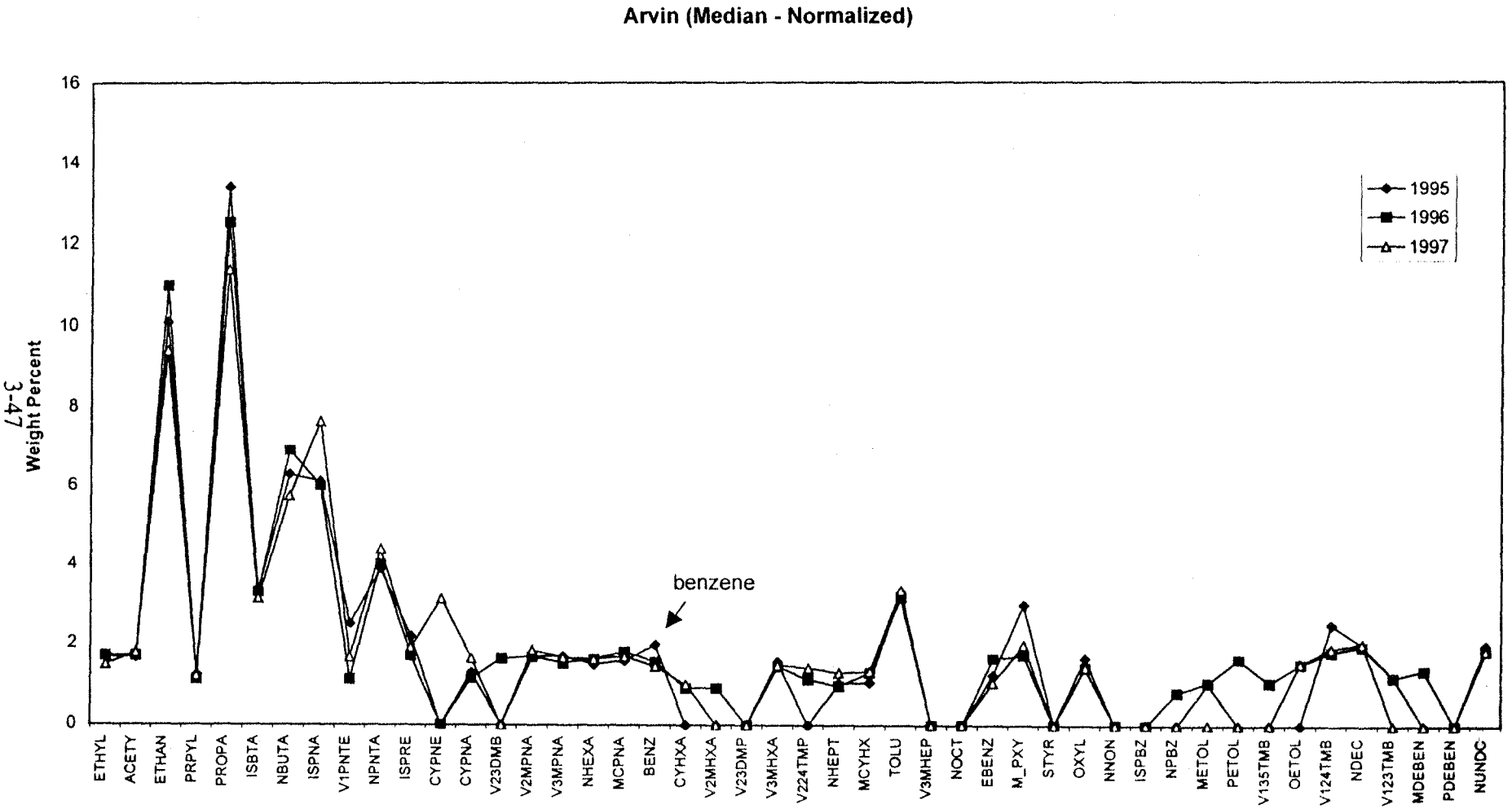


Figure 3-25. Median composition at Arvin from 1995 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

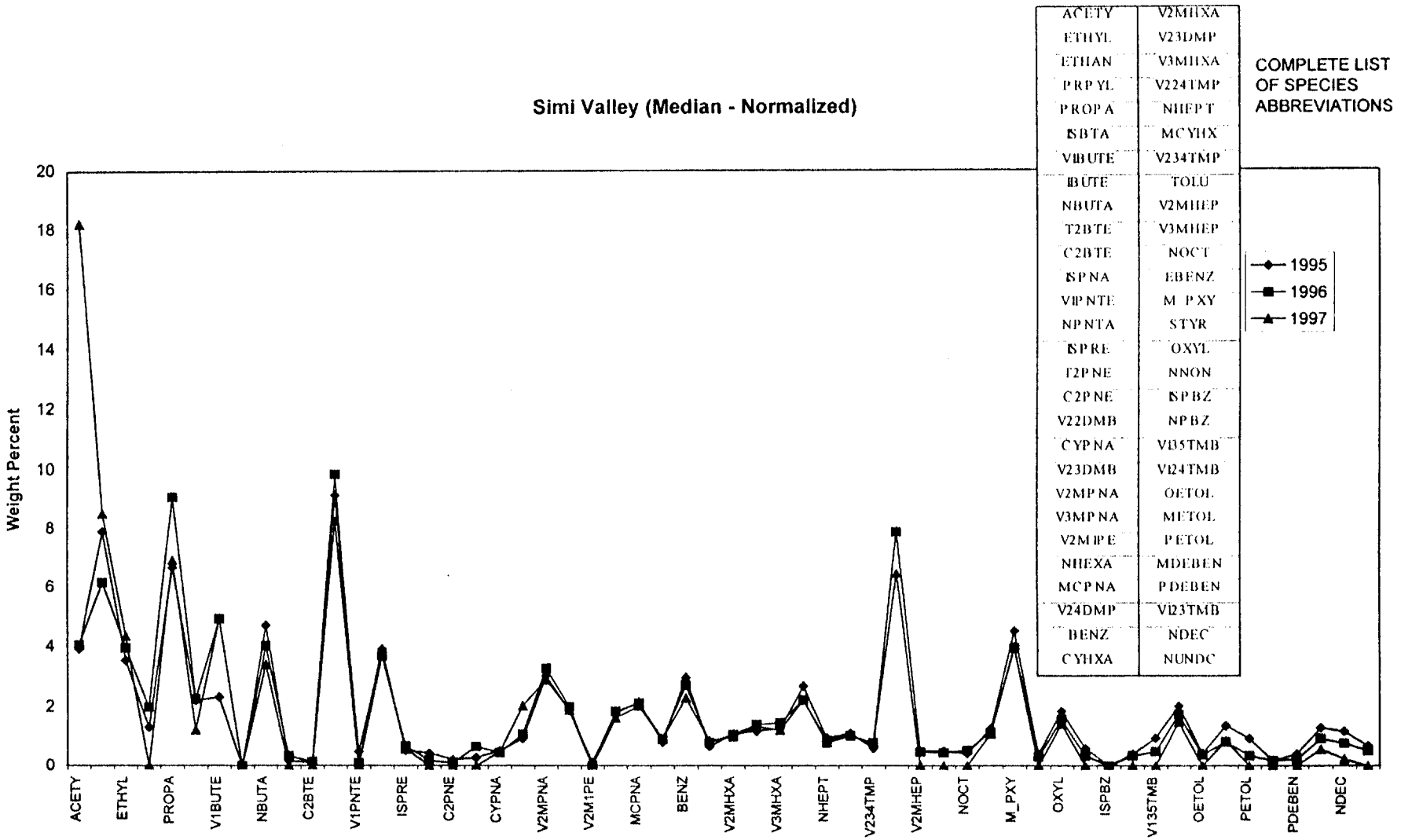


Figure 3-26. Median composition at Simi Valley from 1995 through 1997 normalized over the identified hydrocarbons. A key to the species abbreviations is provided in Appendix B. Data were plotted if more than five samples were above detection.

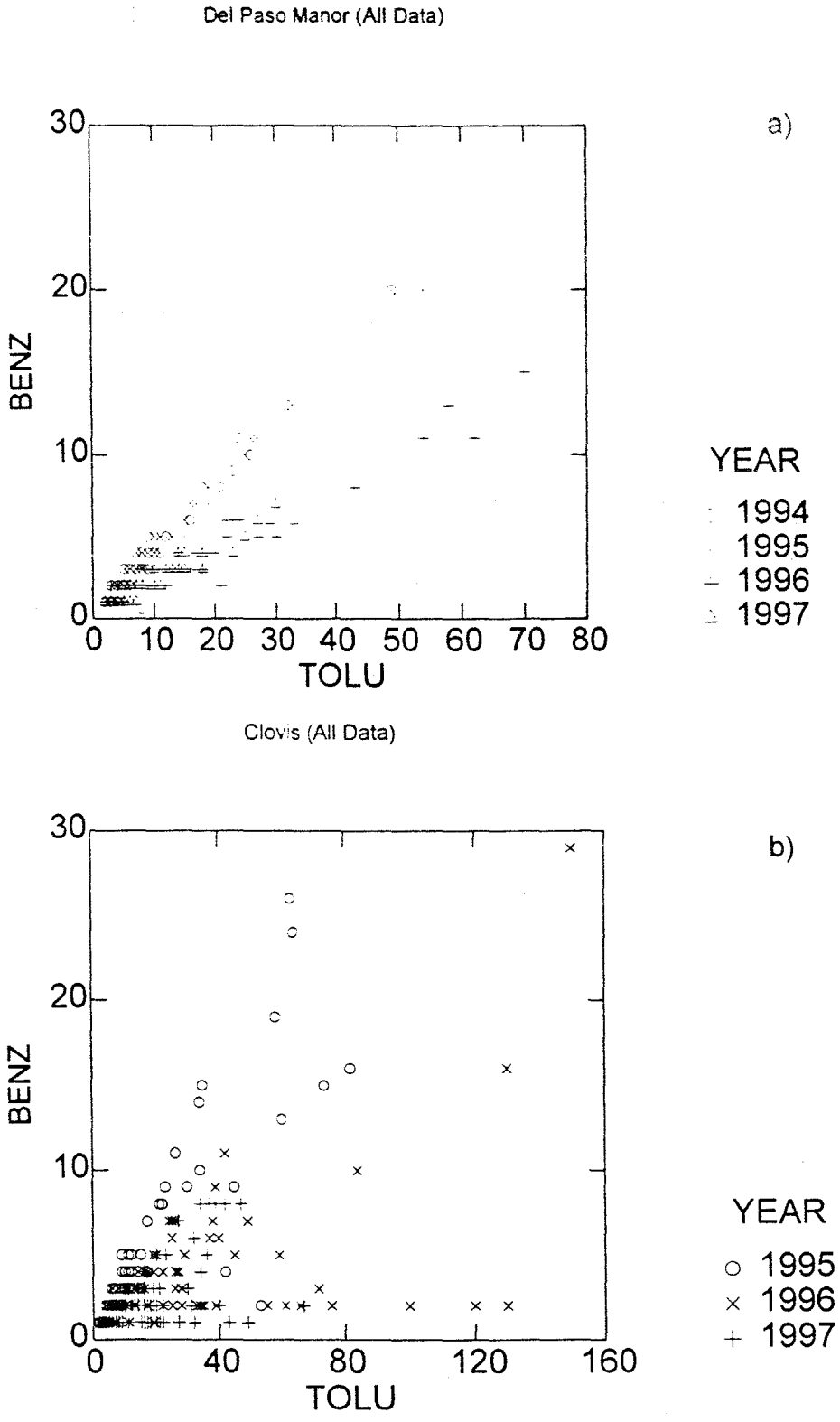


Figure 3-27. Scatter plots of benzene to toluene concentrations (ppbC) by year at a) Del Paso Manor (1994-1997) and b) Clovis (1995-1997).

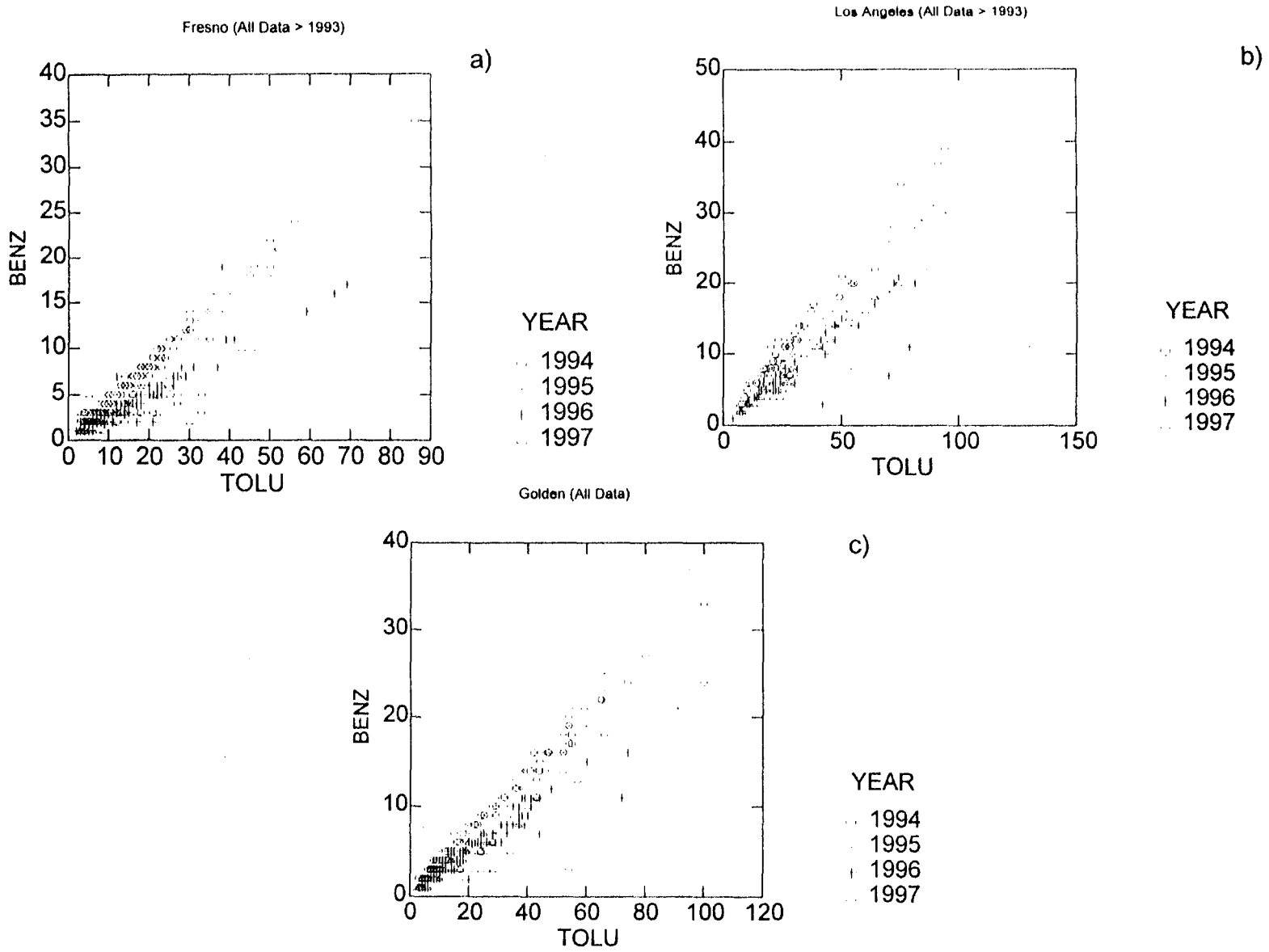


Figure 3-28. Scatter plots of benzene to toluene concentrations (ppbC) by year at a) Fresno (1994-1997), b) Los Angeles (1994-1997), and c) Bakersfield (1994-1997).

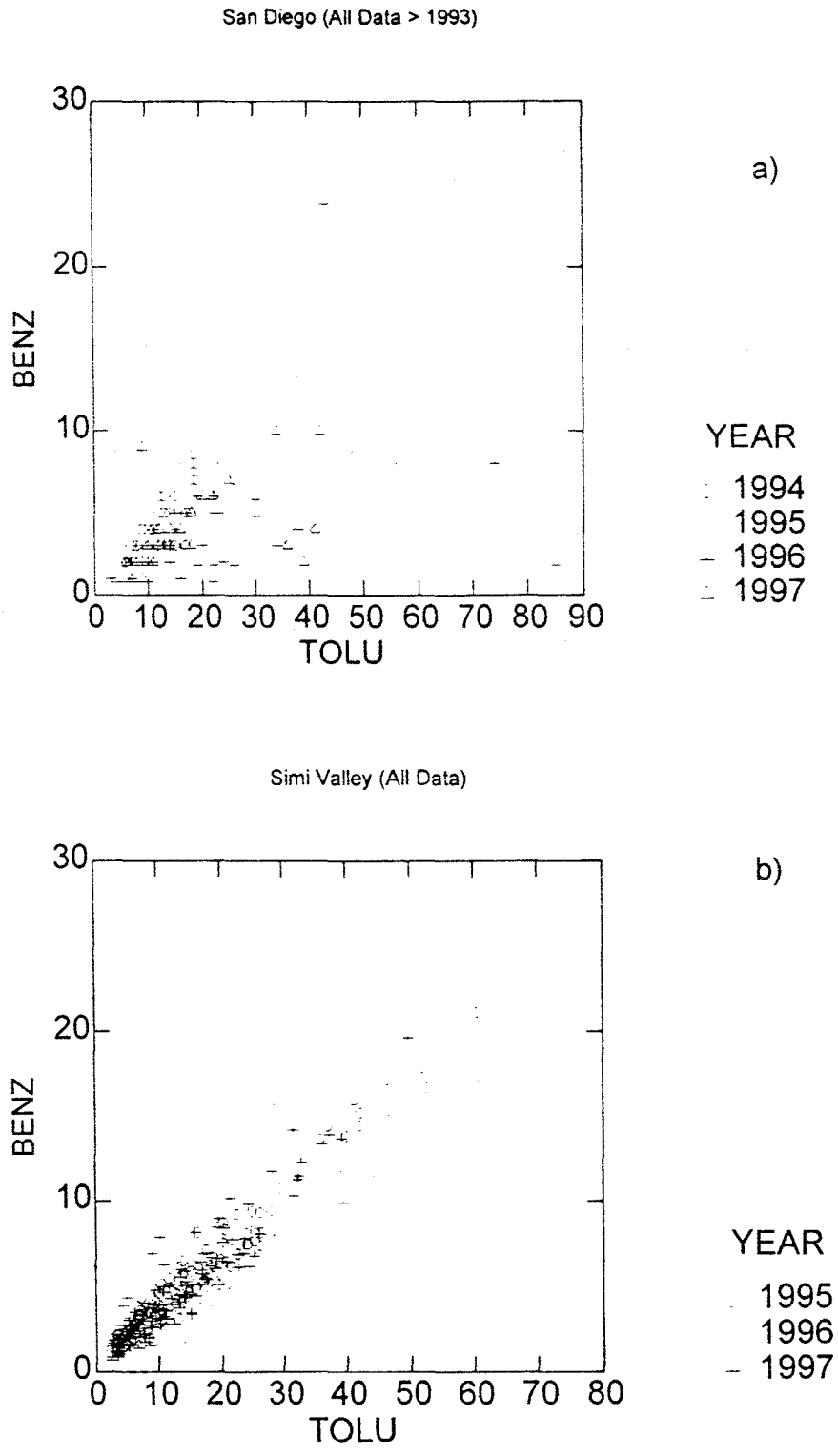


Figure 3-29. Scatter plots of benzene to toluene concentrations (ppbC) by year at a) San Diego (1994-1997) and b) Simi Valley (1995-1997).

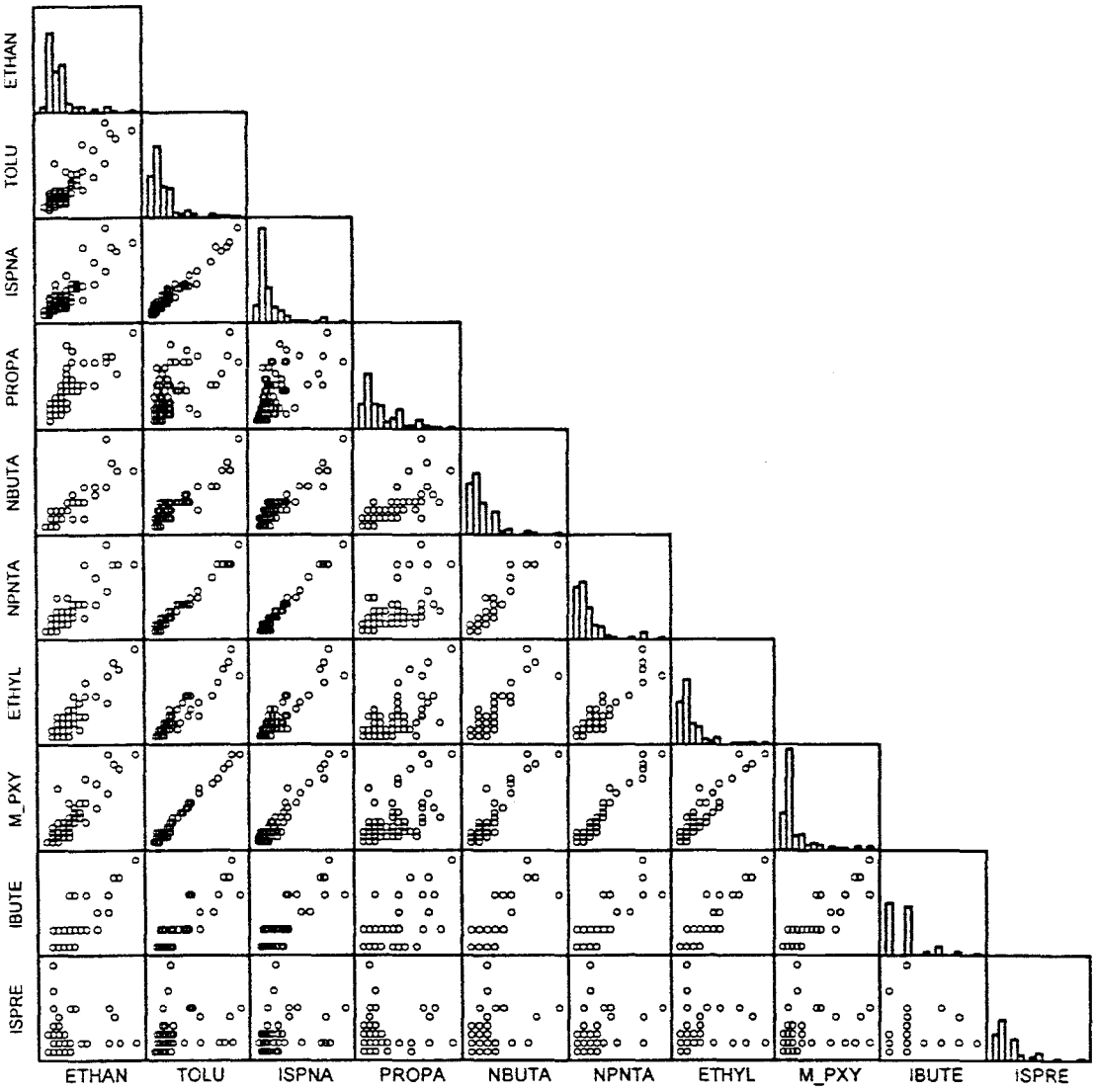


Figure 3-30. Scatter plot matrix of ten abundant hydrocarbons at Del Paso Manor during 1997.

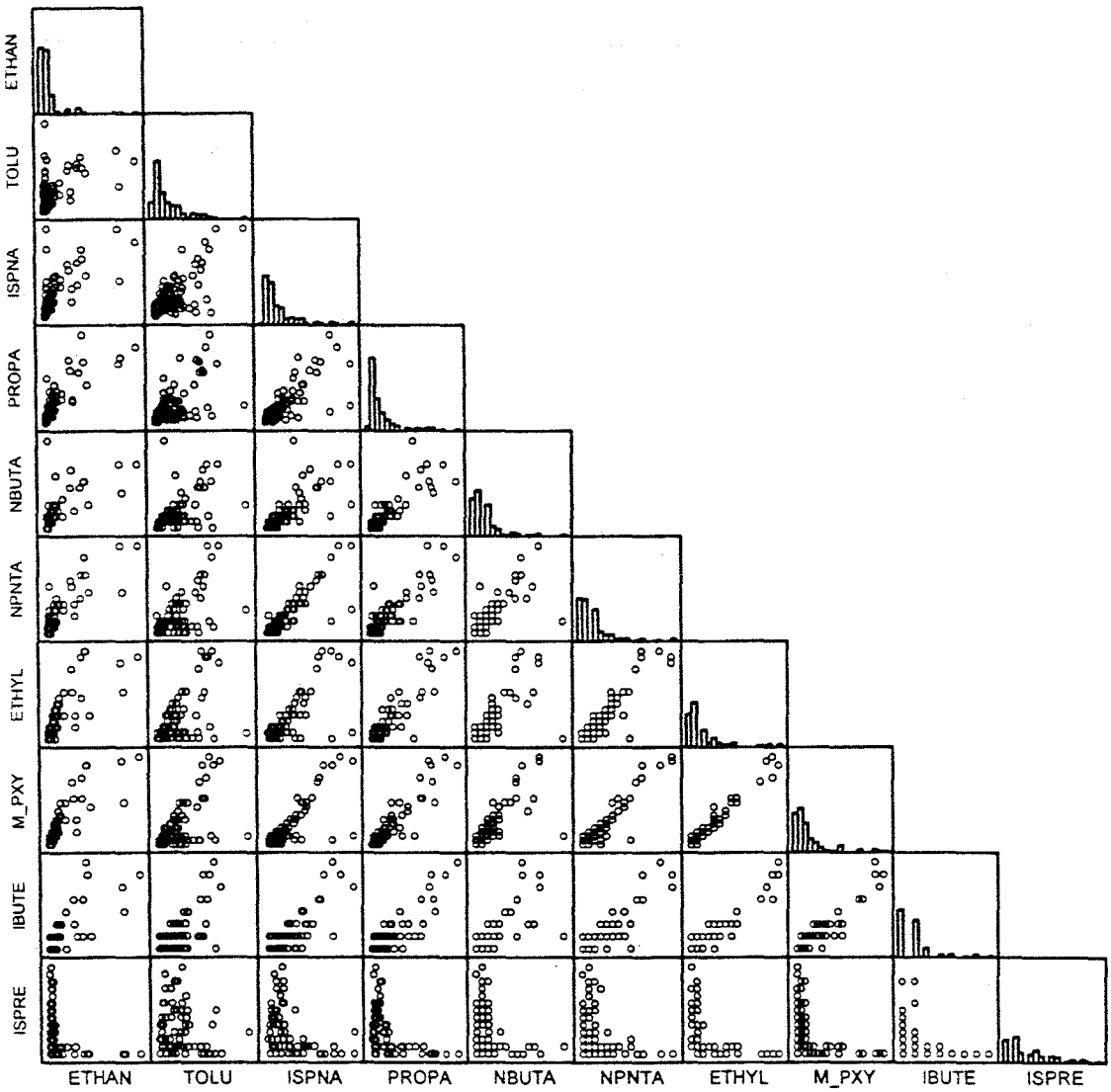


Figure 3-31. Scatter plot matrix of ten most hydrocarbons at Clovis during 1997.

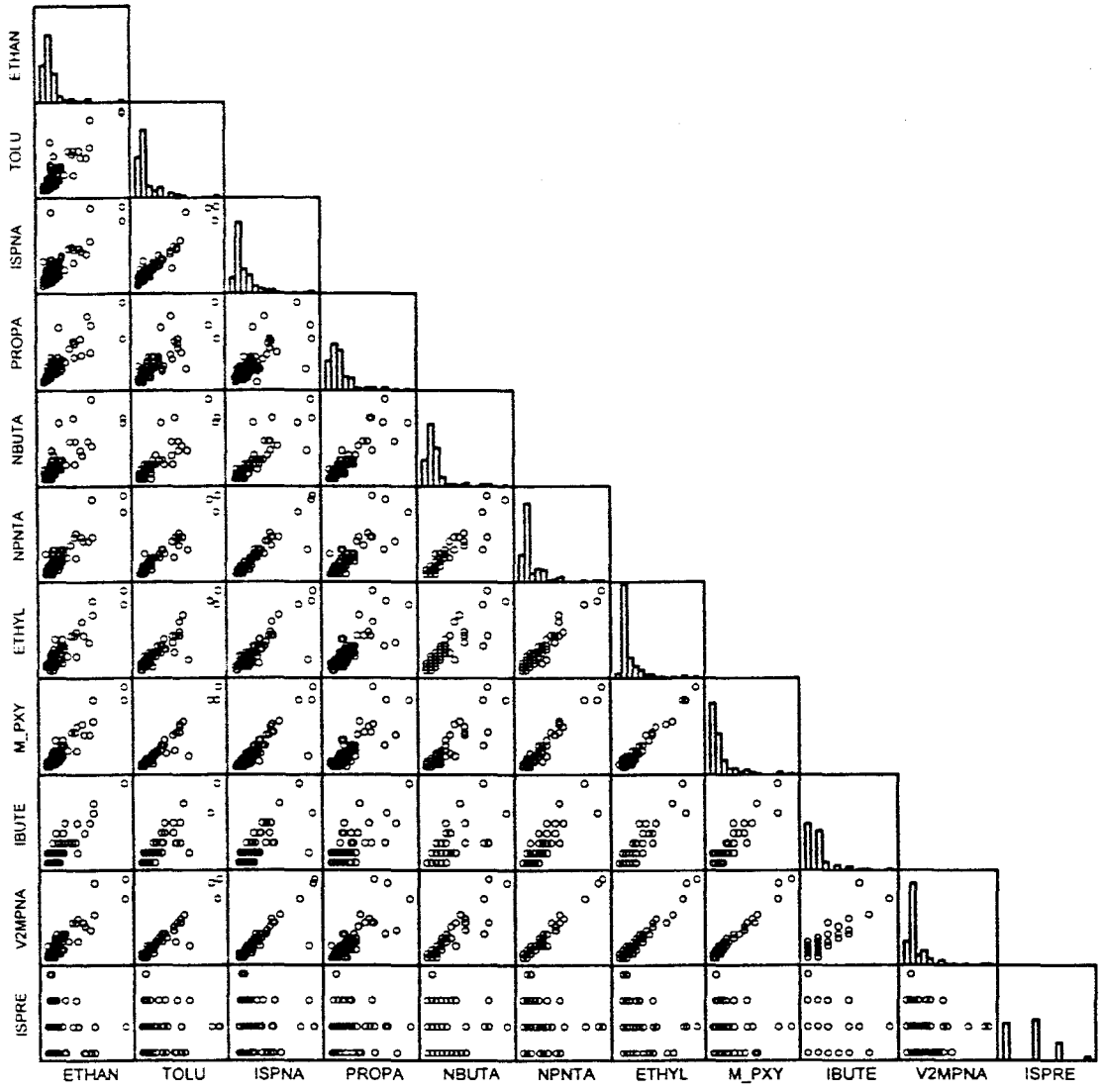


Figure 3-32. Scatter plot matrix of 11 abundant hydrocarbons at Fresno 1st Street during 1997.

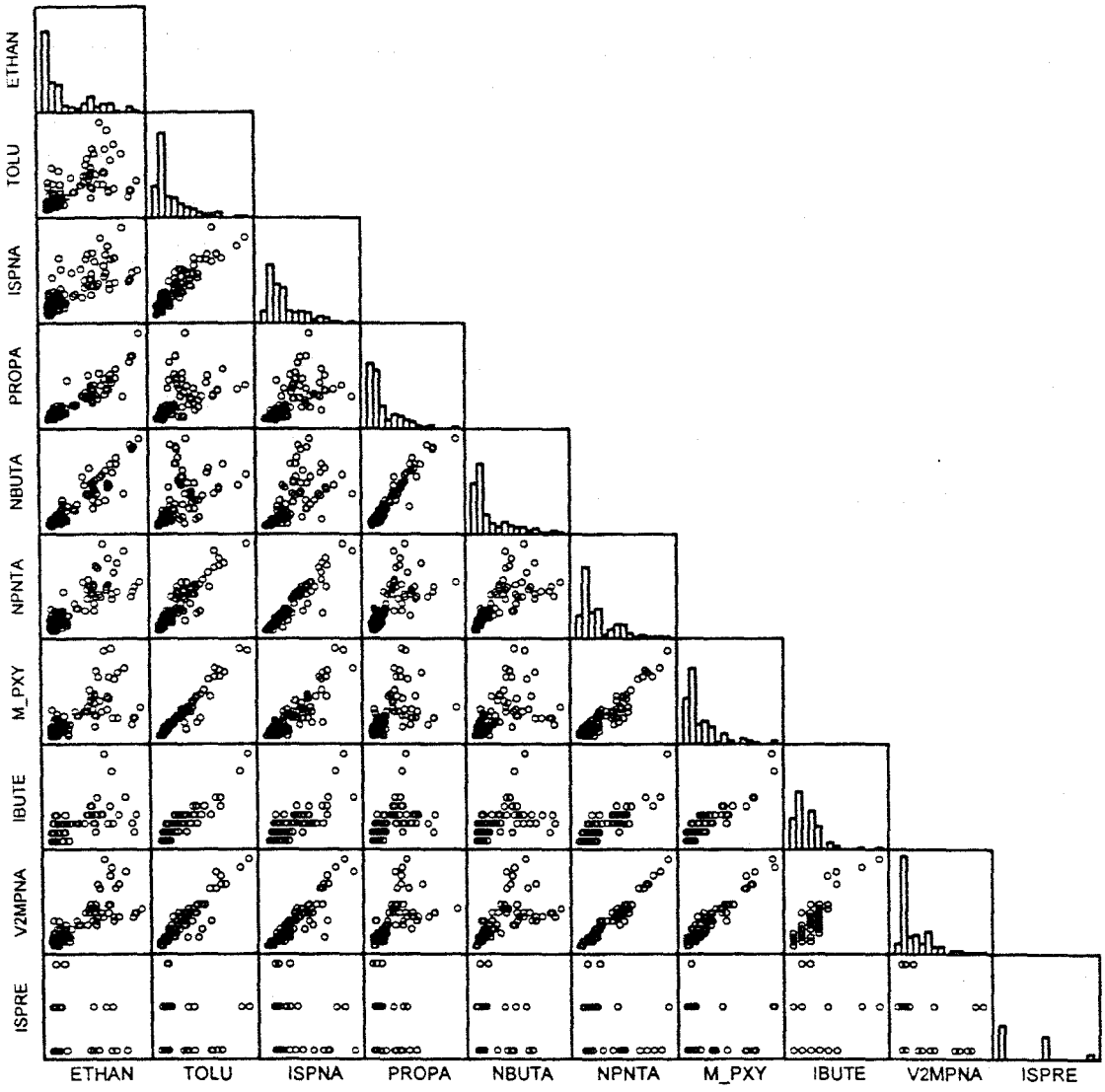


Figure 3-33. Scatter plot matrix of ten abundant hydrocarbons at Bakersfield during 1997.

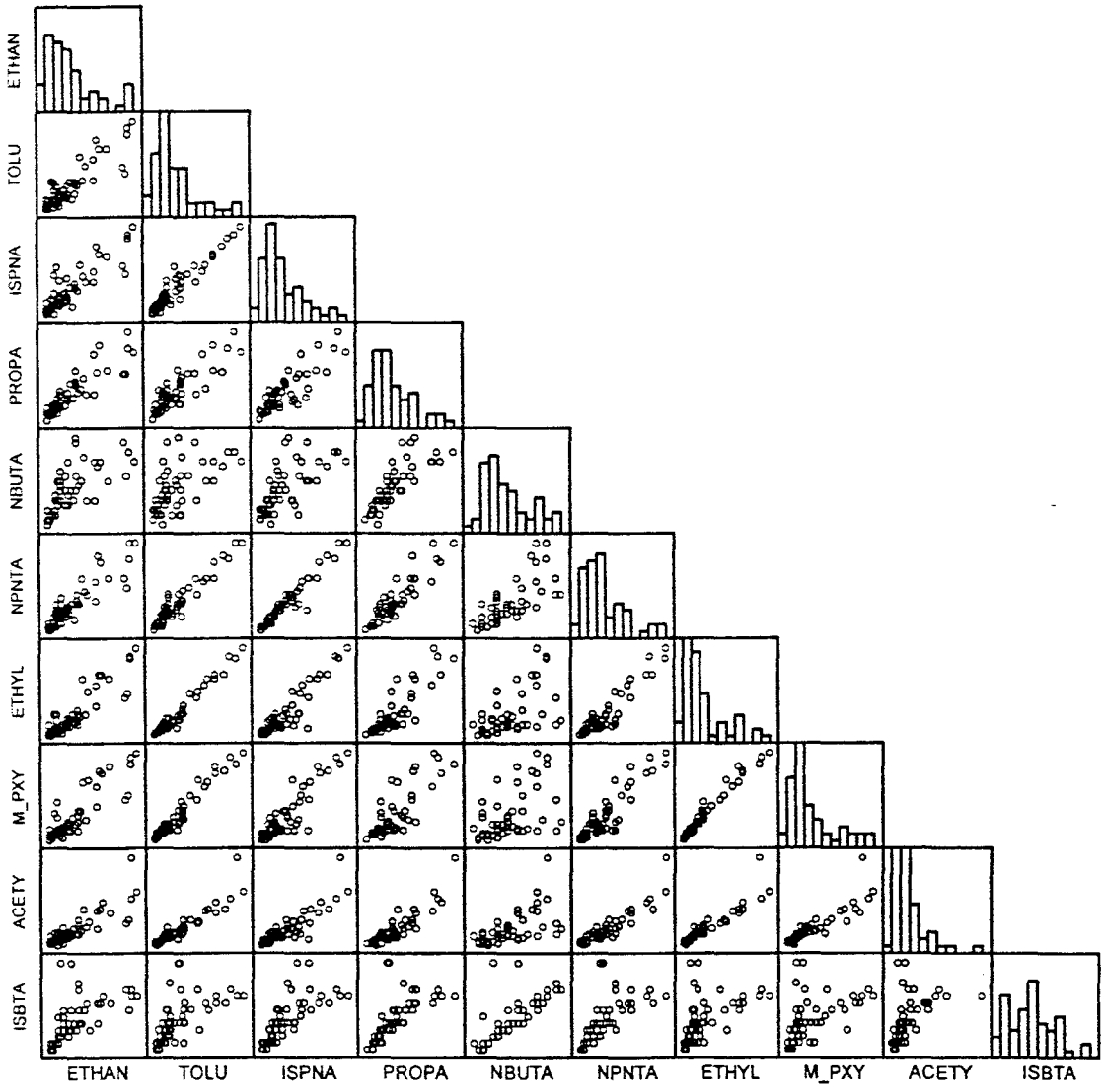


Figure 3-34. Scatter plot matrix of ten abundant hydrocarbons at Los Angeles during 1997.

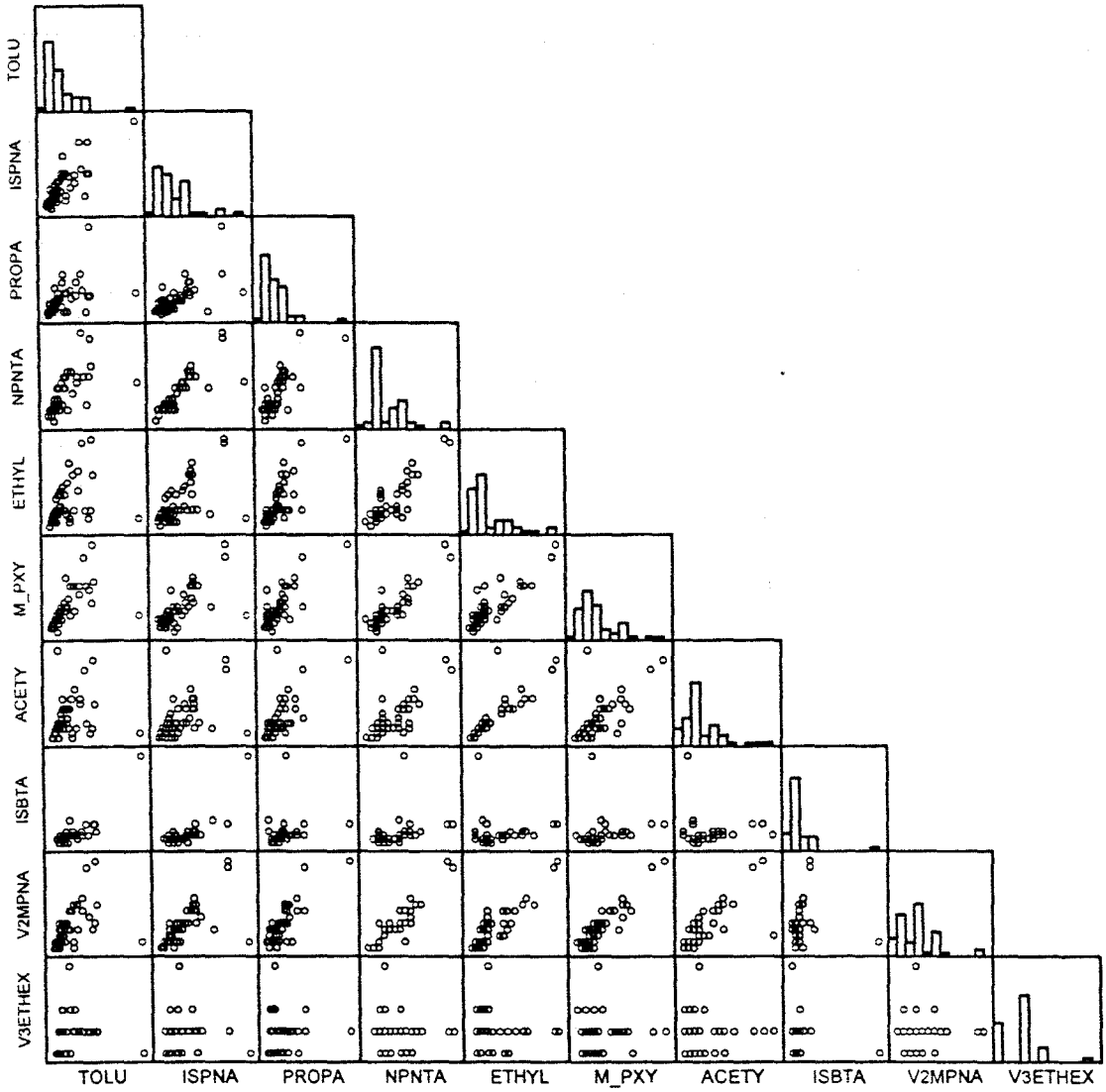


Figure 3-35. Scatter plot matrix of ten abundant hydrocarbons at San Diego during 1997.

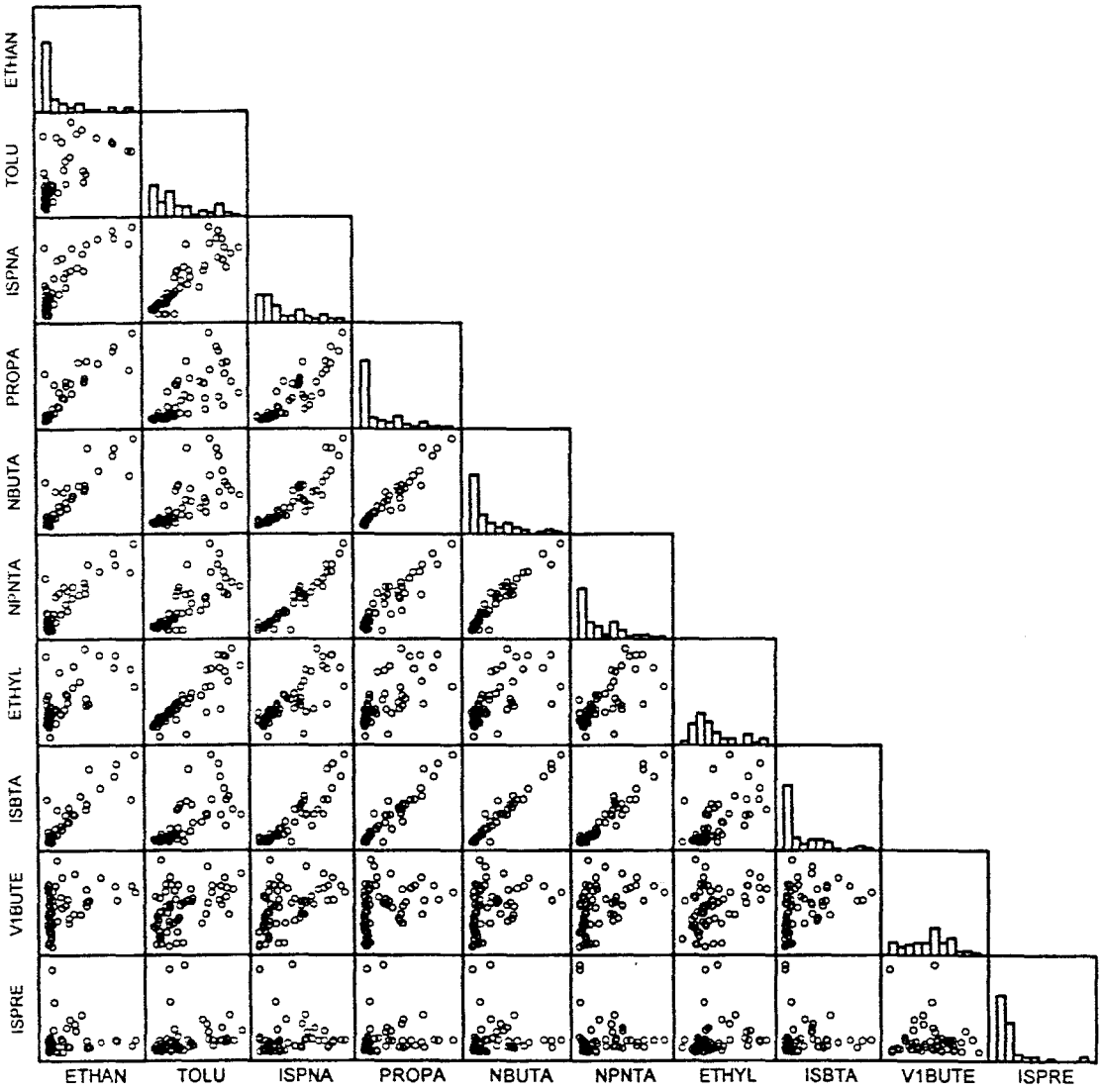


Figure 3-36. Scatter plot matrix of ten abundant hydrocarbons at El Rio during 1996.

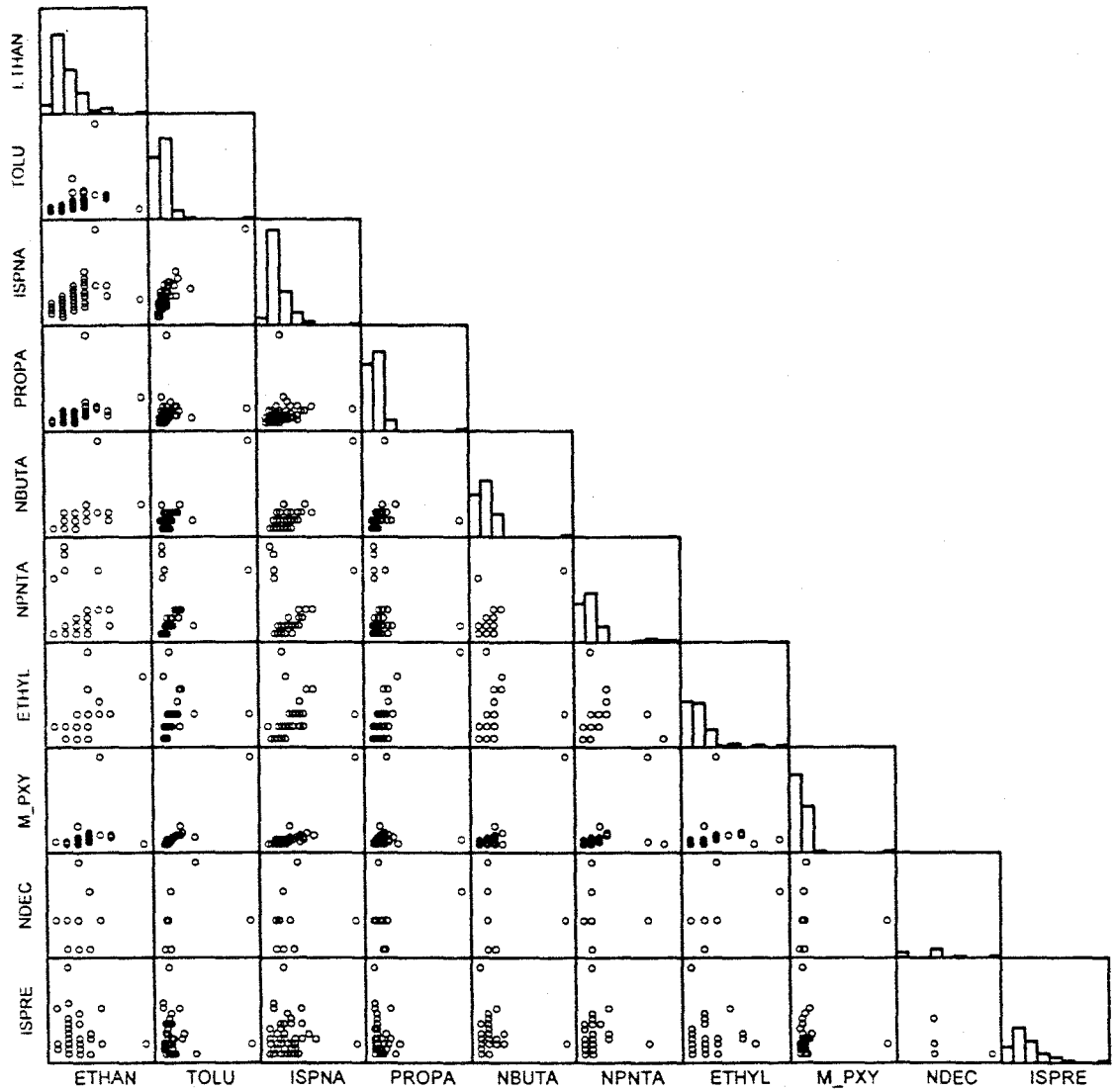


Figure 3-37. Scatter plot matrix of ten abundant hydrocarbons at Folsom during 1997.

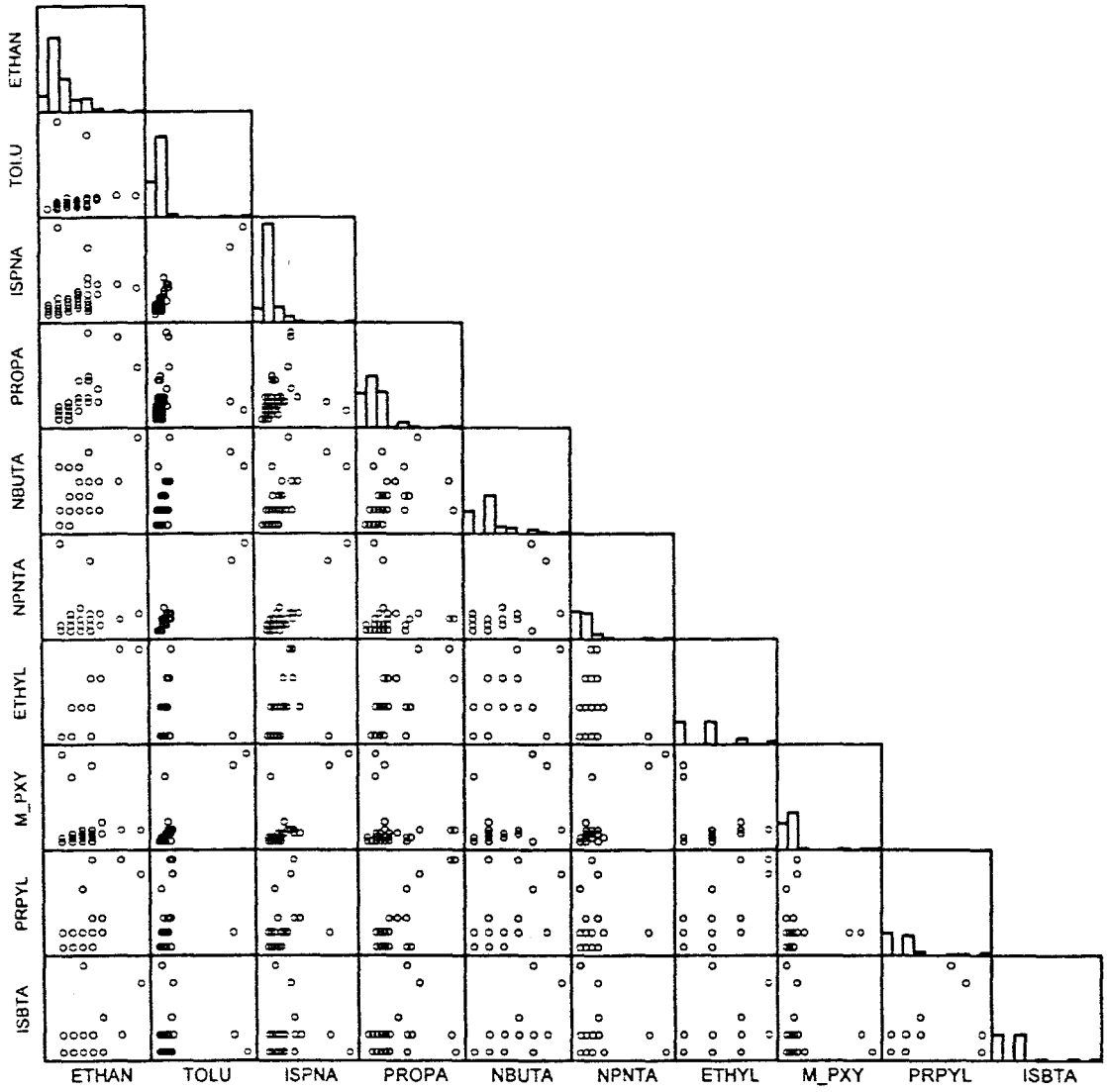


Figure 3-38. Scatter plot matrix of ten abundant hydrocarbons at Elk Grove during 1997.

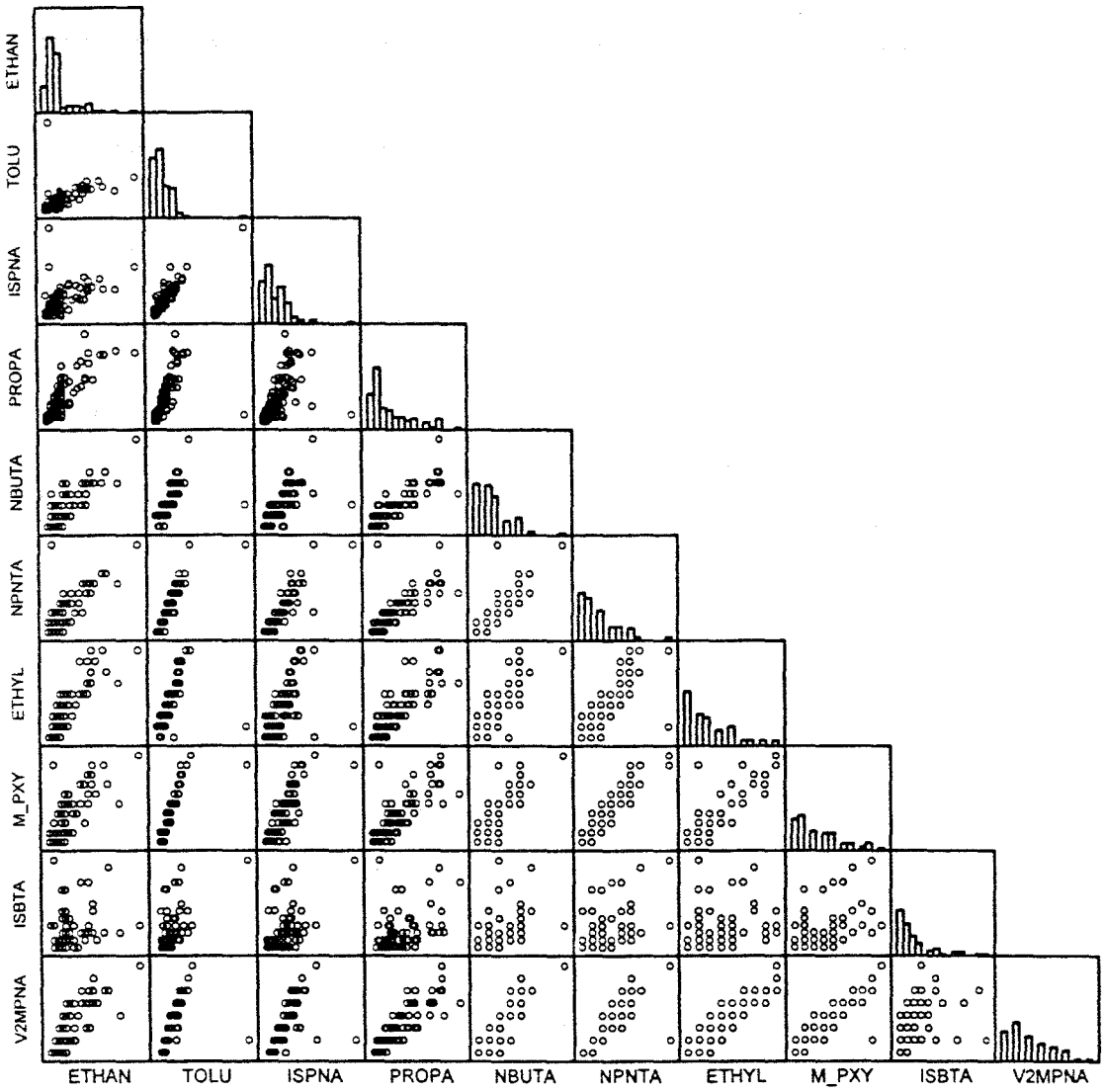


Figure 3-39. Scatter plot matrix of ten abundant hydrocarbons at Parlier during 1997.

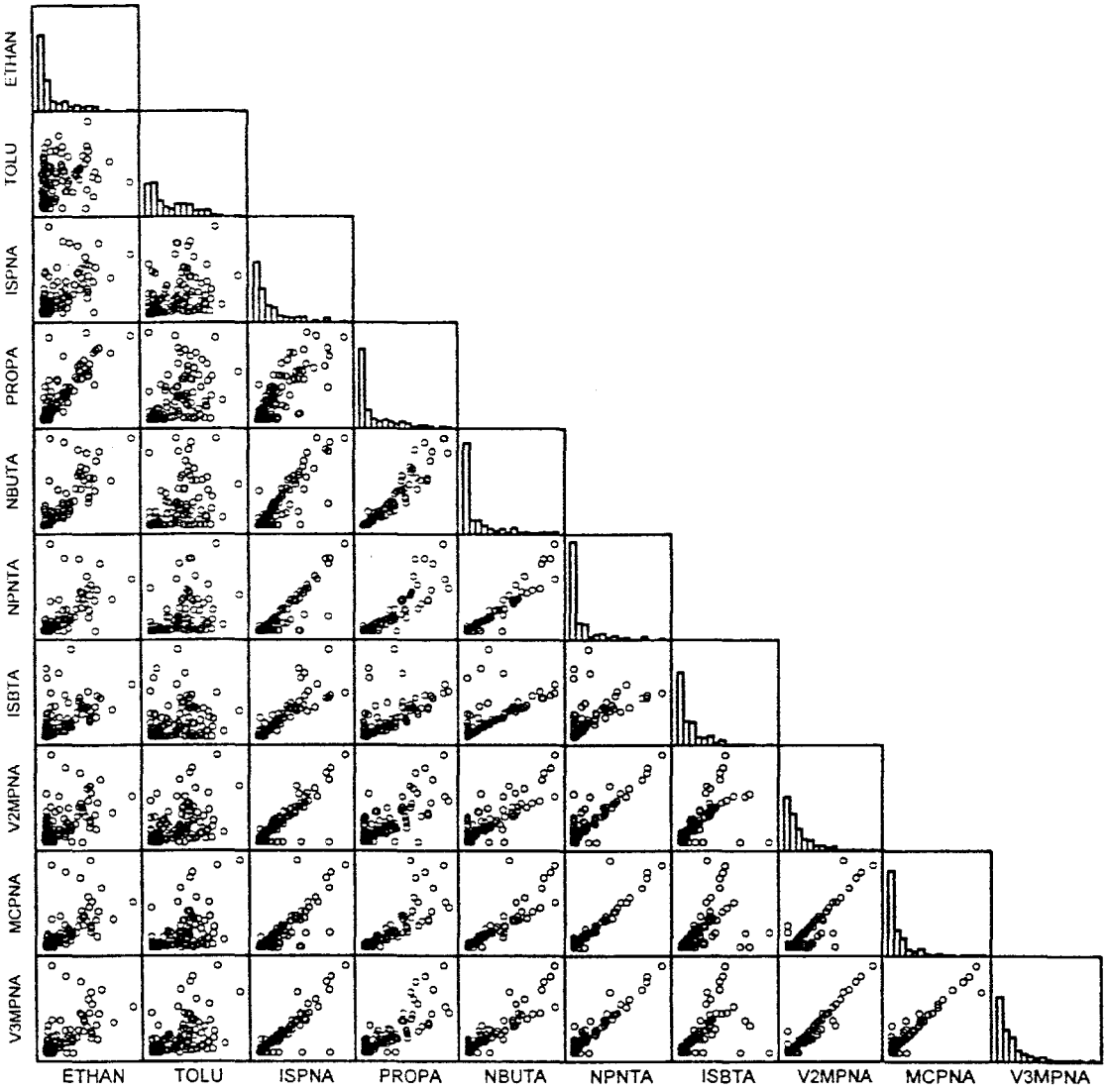


Figure 3-40. Scatter plot matrix of ten abundant hydrocarbons at Emma Wood during 1997.

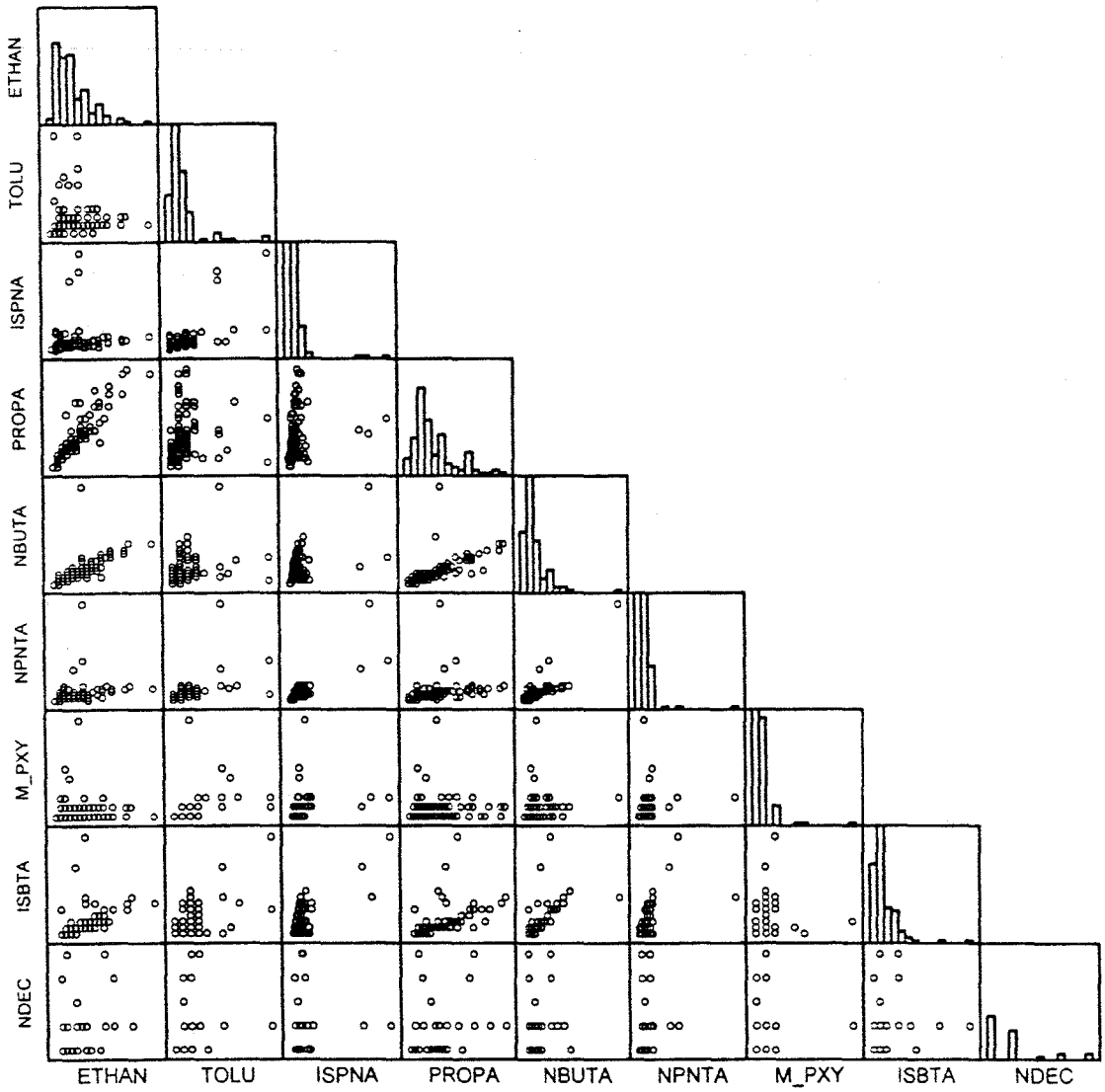


Figure 3-41. Scatter plot matrix of nine abundant hydrocarbons at Arvin during 1997.

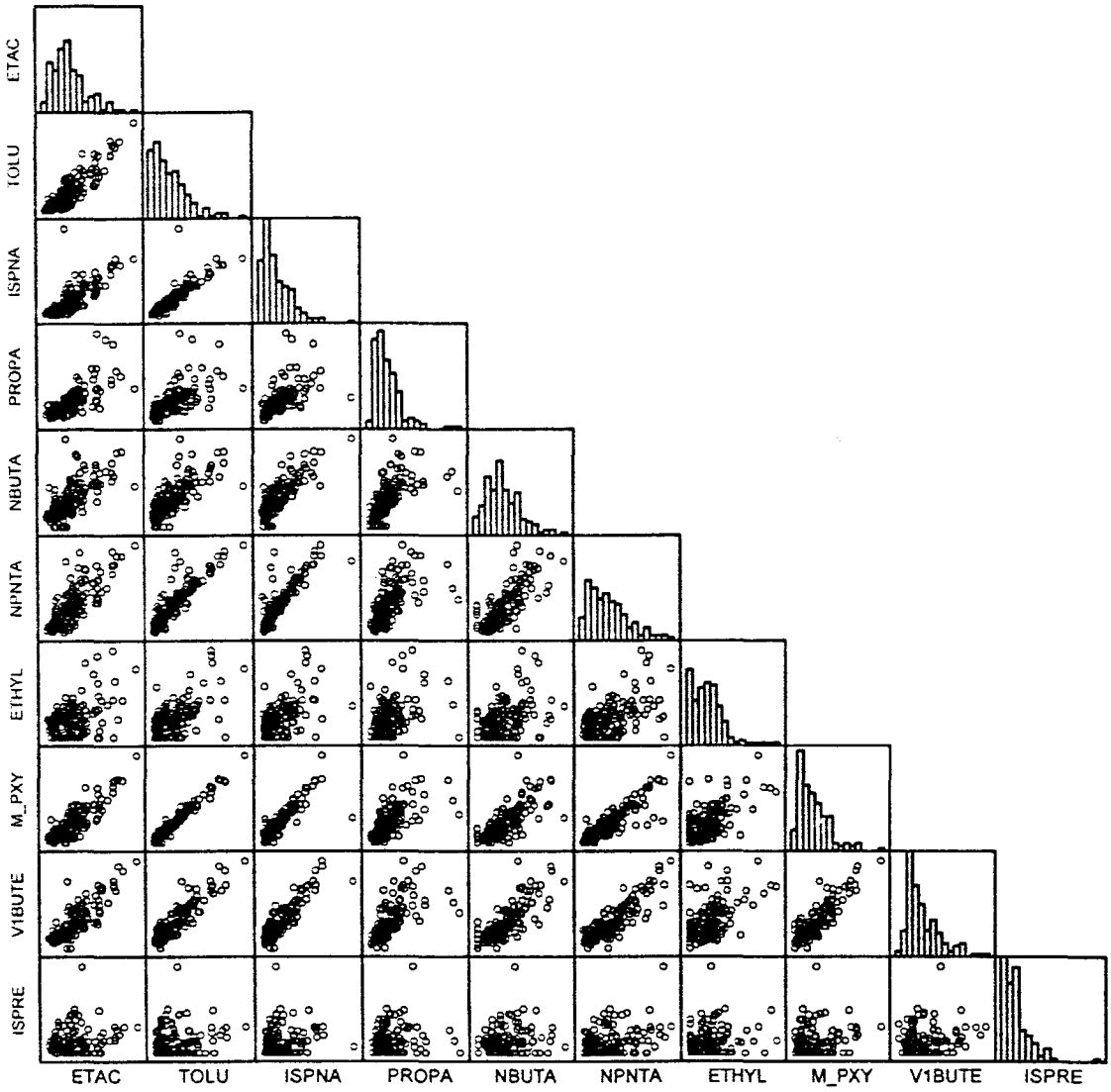


Figure 3-42. Scatter plot matrix of ten abundant hydrocarbons at Simi Valley during 1997.

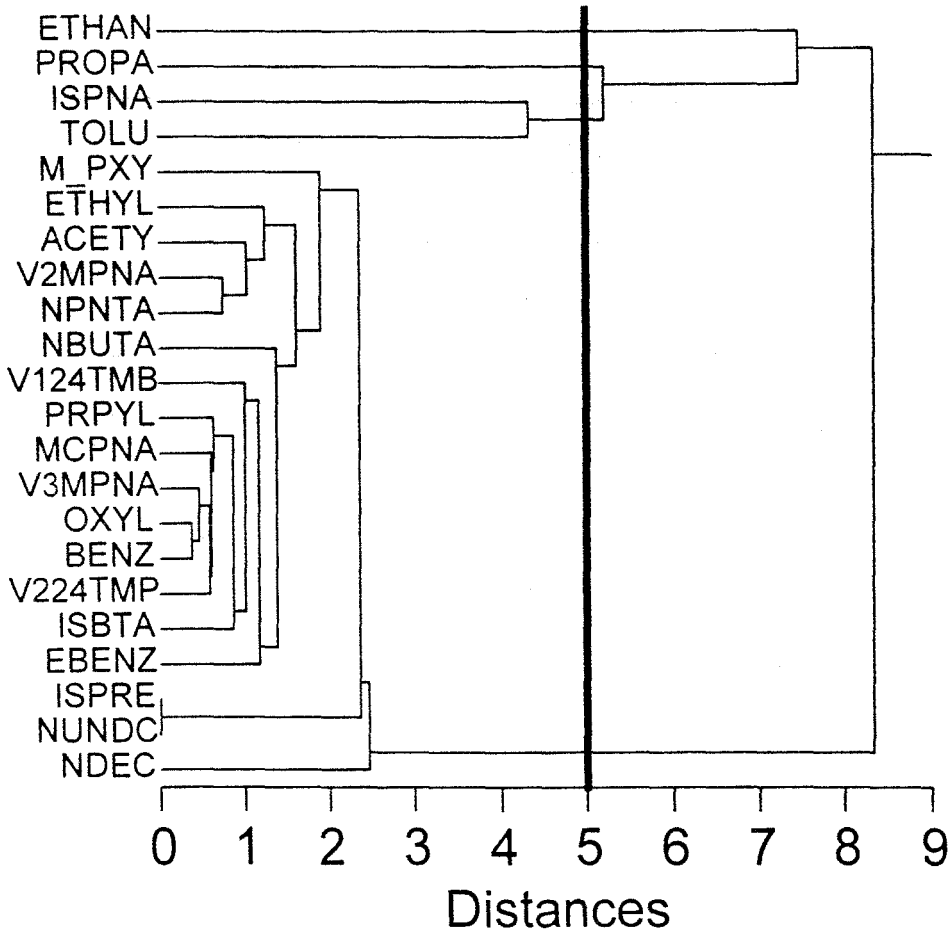


Figure 3-43. Cluster analysis results for Clovis in 1997 (morning data only). The bold vertical line at a distance of 5 is provided for discussion purposes.

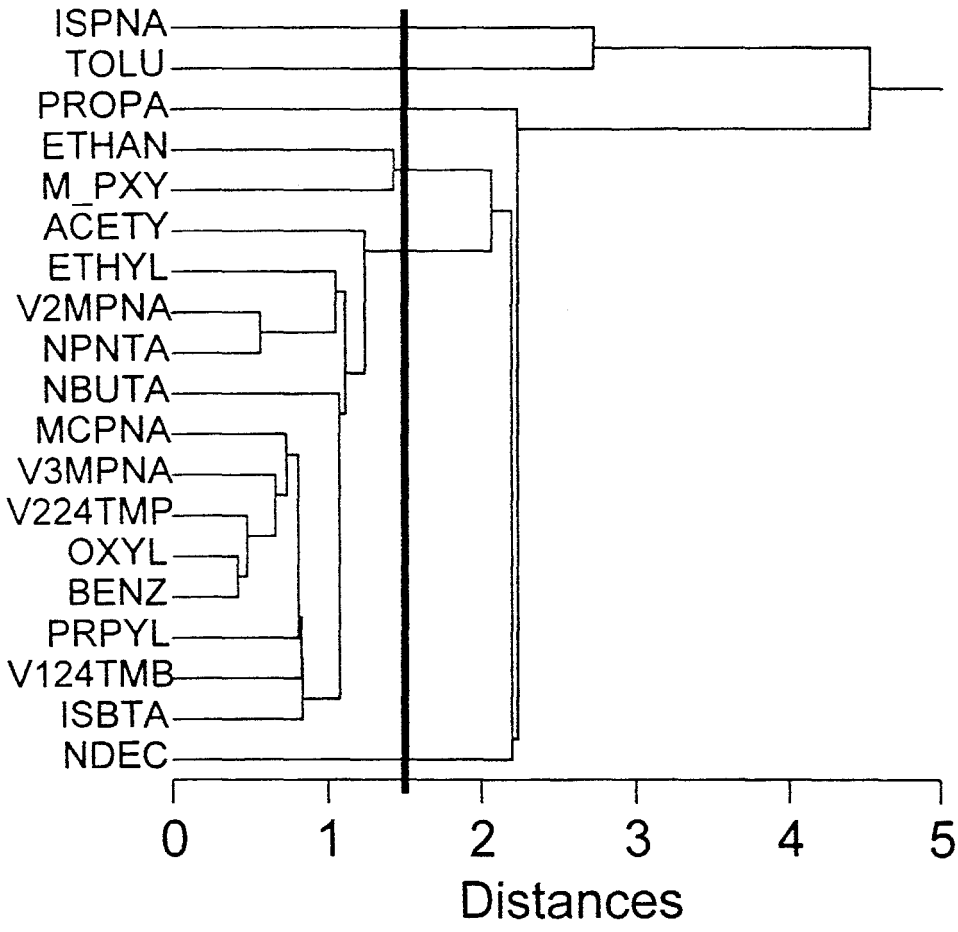


Figure 3-44. Cluster analysis results for Del Paso Manor in 1997 (morning data only). The bold vertical line at a distance of 1.5 is provided for discussion purposes.

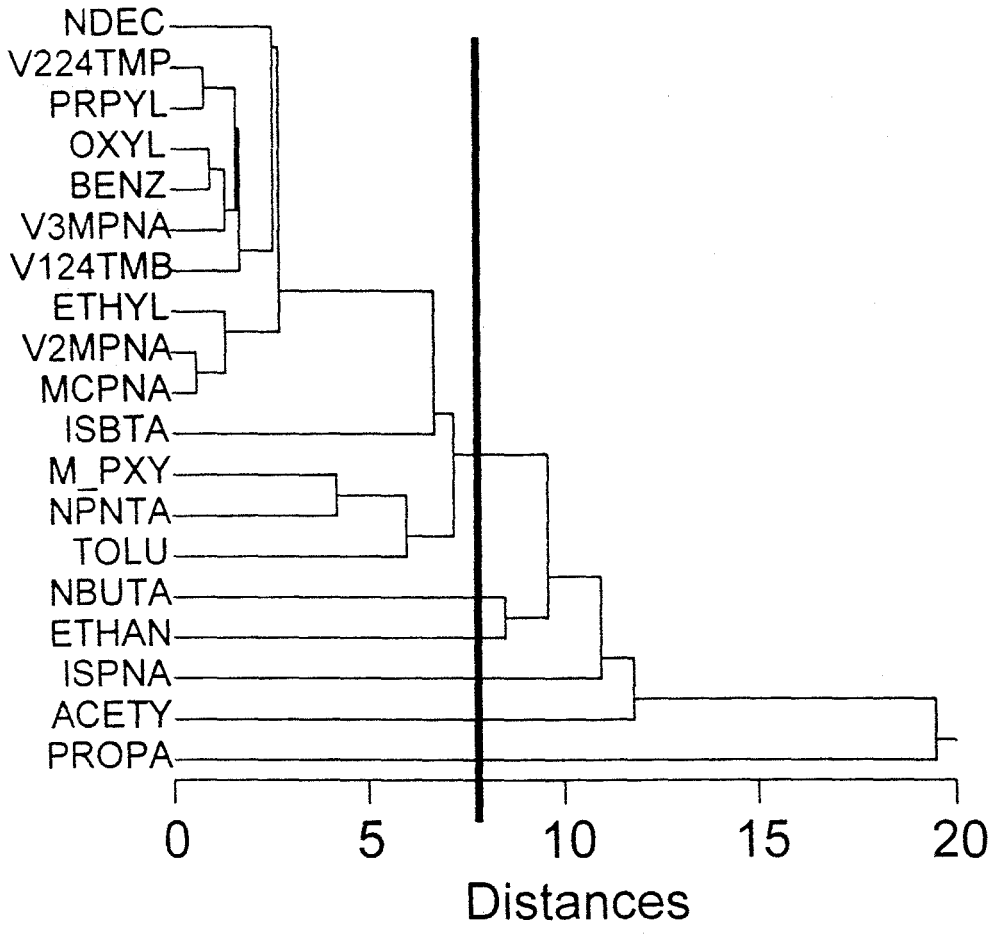


Figure 3-45. Cluster analysis results for Bakersfield in 1997 (morning data only). The bold vertical line at a distance of 8 is provided for discussion purposes.

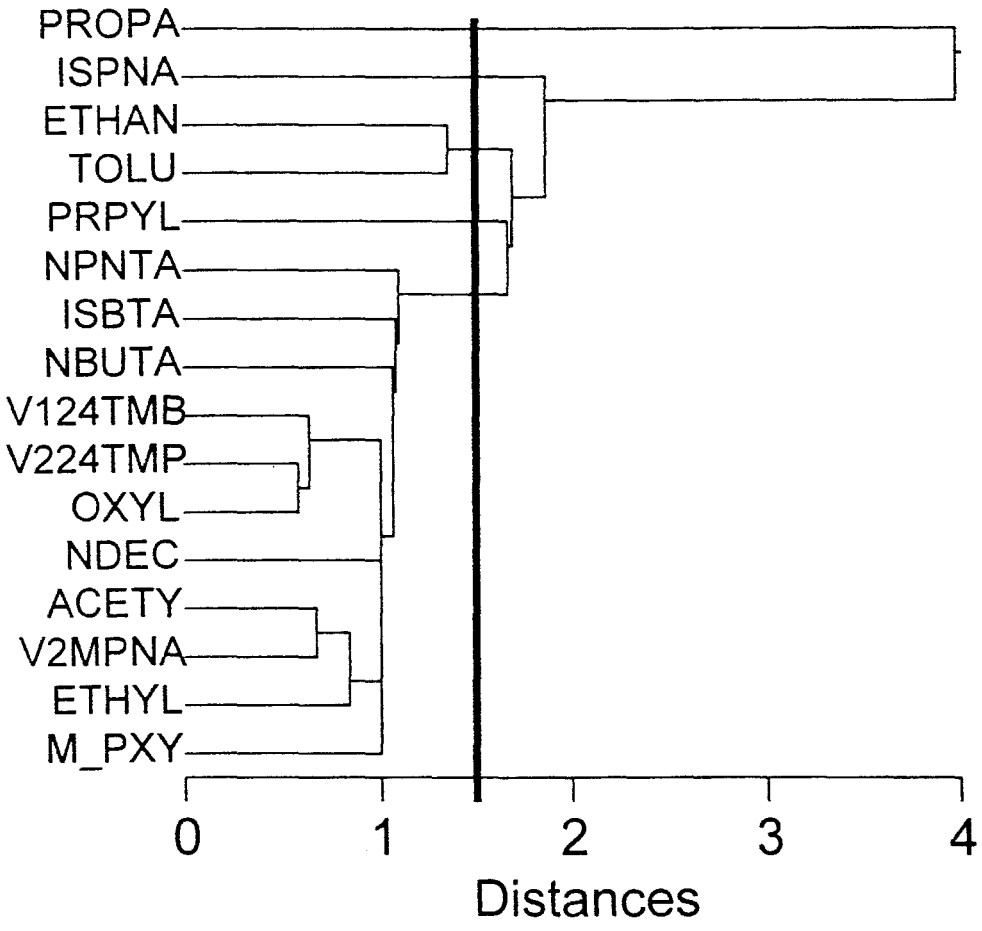


Figure 3-46. Cluster analysis results for Elk Grove in 1997 (morning data only). The bold vertical line at a distance of about 1.5 is provided for discussion purposes.

Rotated Loading Matrix (VARIMAX, Gamma = 1.0000)

	1	2	3	4
NPNTA	0.829	0.194	0.222	0.094
MCPNA	0.818	0.176	0.082	0.430
ISPNA	0.789	-0.089	0.078	0.257
V2MPNA	0.779	0.110	-0.052	0.482
OXYL	0.766	0.141	-0.310	0.228
TOLU	0.754	0.074	0.025	0.219
V224TMP	0.748	-0.068	0.266	0.143
M_PXY	0.708	0.140	0.196	0.567
NHEXA	0.680	0.445	-0.013	-0.034
BENZ	0.634	0.028	-0.361	0.537
V3MPNA	0.540	-0.027	0.342	0.573
ACETY	0.527	0.150	-0.125	0.708
PROPA	-0.026	0.843	0.240	-0.099
ISBTA	0.014	0.839	-0.206	0.060
NBUTA	0.263	0.757	0.097	0.002
V124TMB	0.058	0.573	-0.598	0.108
ETHAN	0.279	0.287	0.652	0.259
PRPYL	0.057	-0.007	0.147	0.886
ETHYL	0.477	-0.193	-0.044	0.804

"Variance" Explained by Rotated Components

1	2	3	4
6.617	2.795	1.448	3.576

Percent of Total Variance Explained

1	2	3	4
34.824	14.713	7.622	18.820

Figure 3-47. Factor analysis results for Fresno in 1997 (all data). Factor loadings greater than 0.5 have been highlighted. Factors 1, 2, and 3 accounted for 35, 15, and 8 percent of the total variance, respectively.

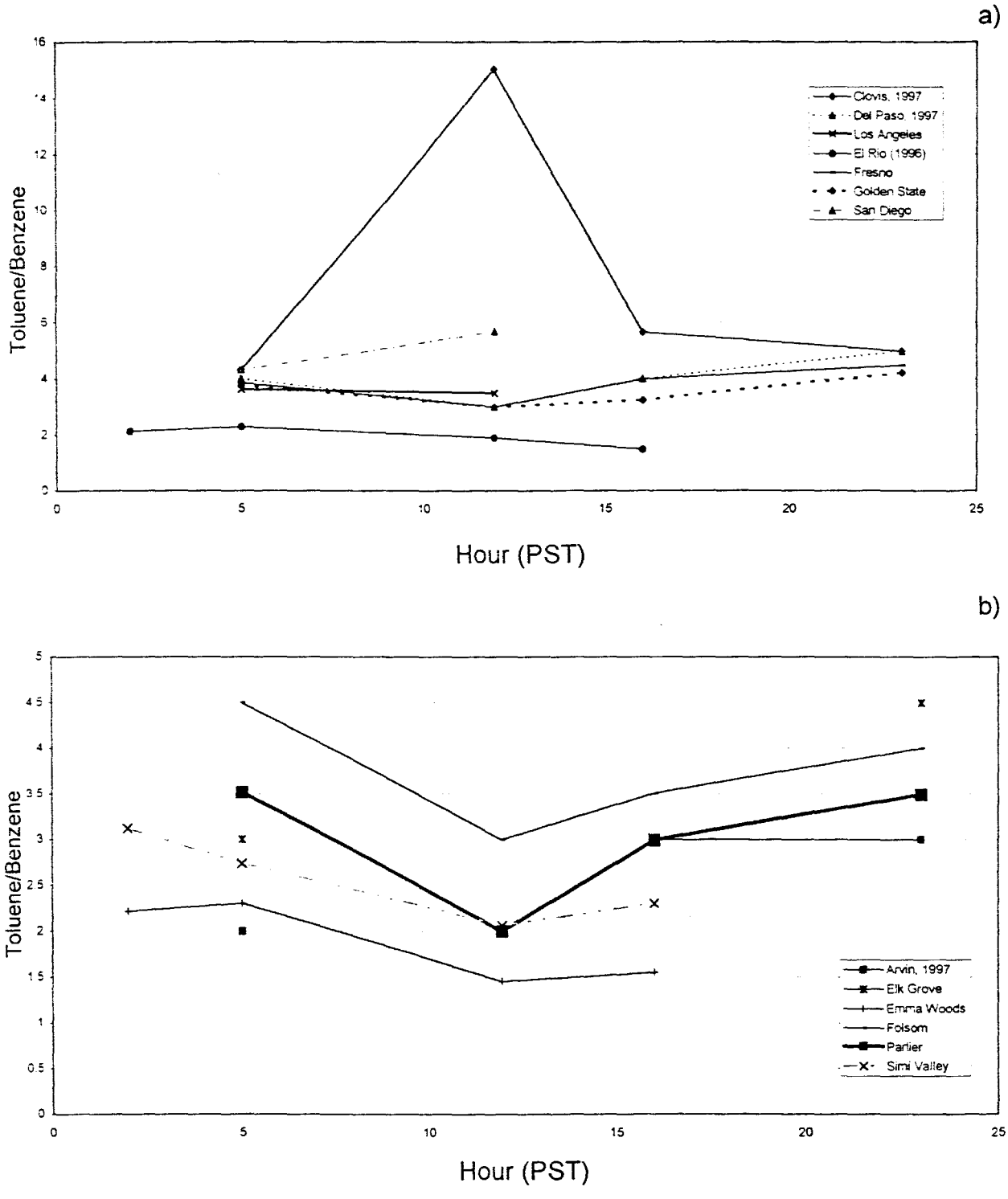


Figure 3-48. Median toluene to benzene ratios by time of day (PST) in 1997 at a) Type 2 and b) Type 1 and 3 sites. Data were plotted if more than five samples were above detection.

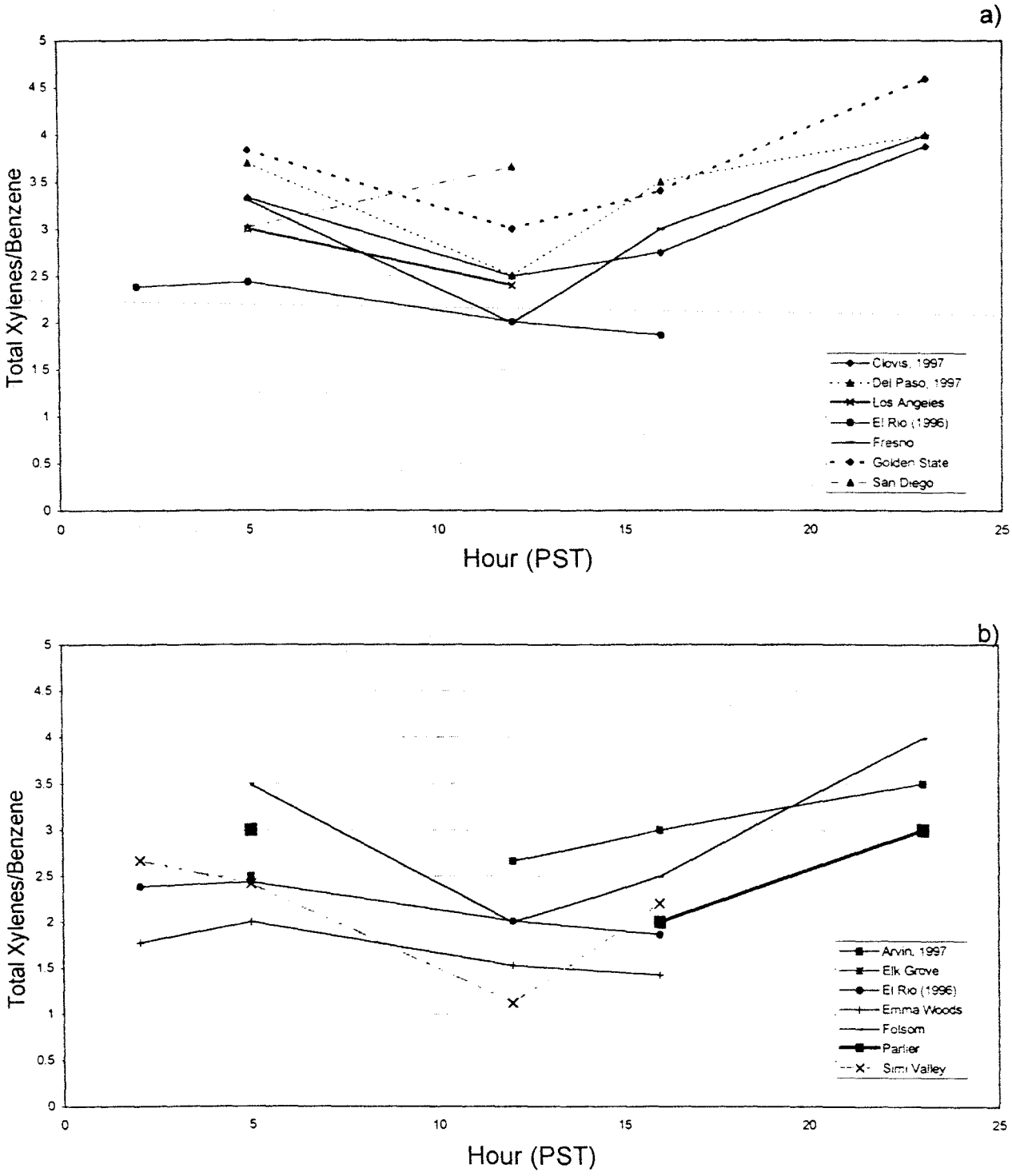


Figure 3-49. Median xylenes to benzene ratios by time of day (PST) in 1997 at a) Type 2 and b) Type 1 and 3 sites. Data were plotted if more than five samples were above detection.

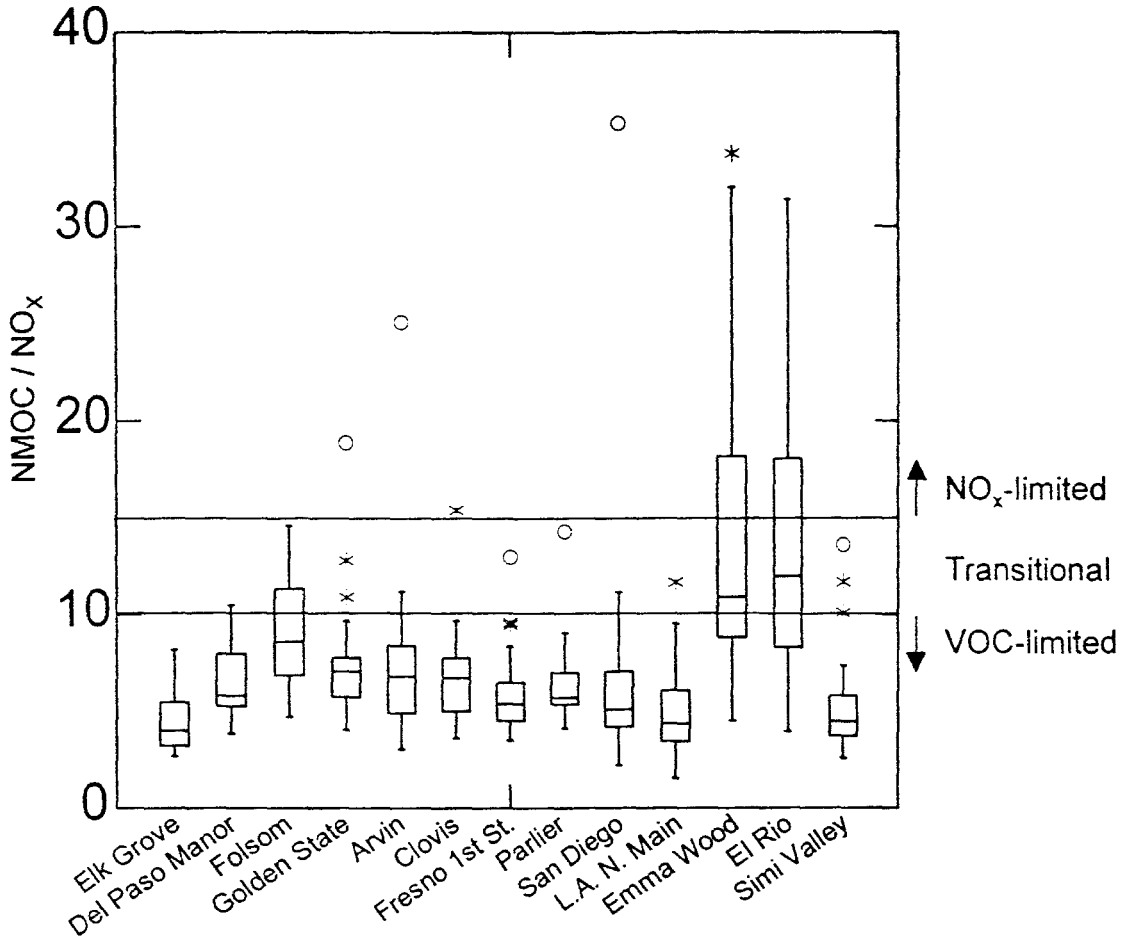


Figure 3-50. Morning (0500-0700 PST) NMHC/NO_x ratios by site in 1997. Data shown meet the following criteria: NO_x concentrations > 10 ppb and NMHC >100 ppbC. Some data points were excluded in order to show sufficient detail.

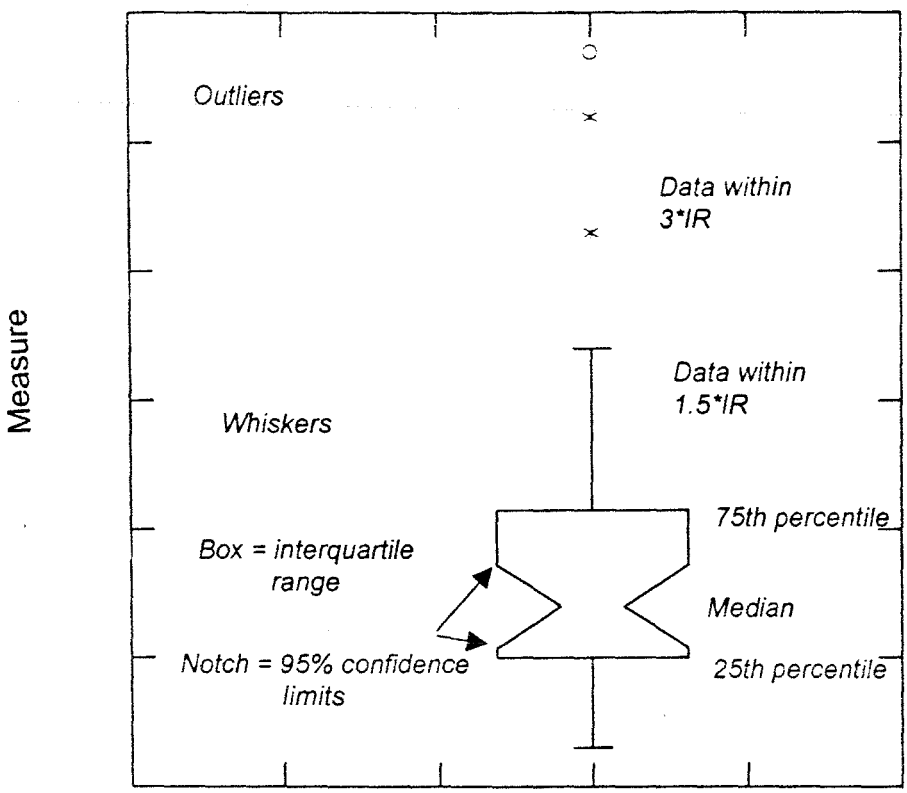


Figure 3-51. Annotated notched box-whisker plot with outliers.

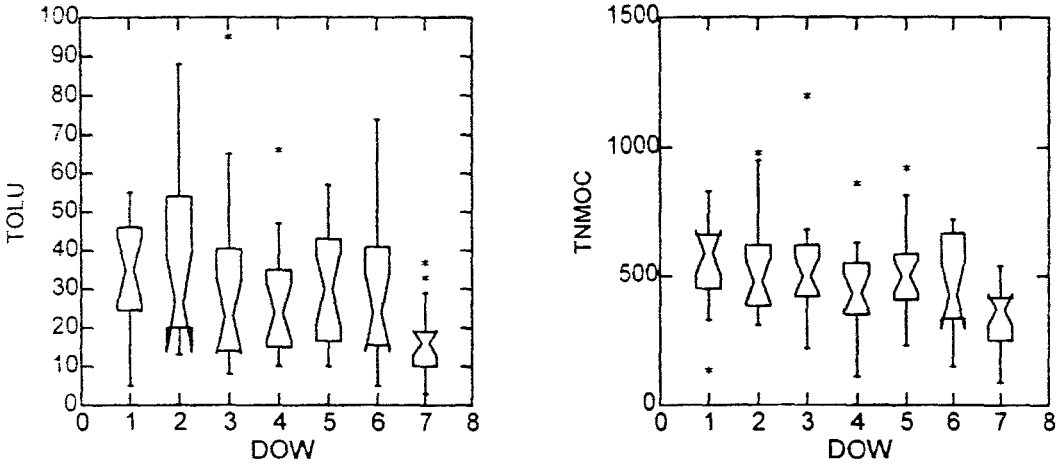


Figure 3-52. Notched box plots of morning (0500 PST) toluene (TOLU) and NMHC (shown as TNMOC) concentrations (ppbC) at Bakersfield by day of week in 1997 (Monday = 1, Sunday = 7).

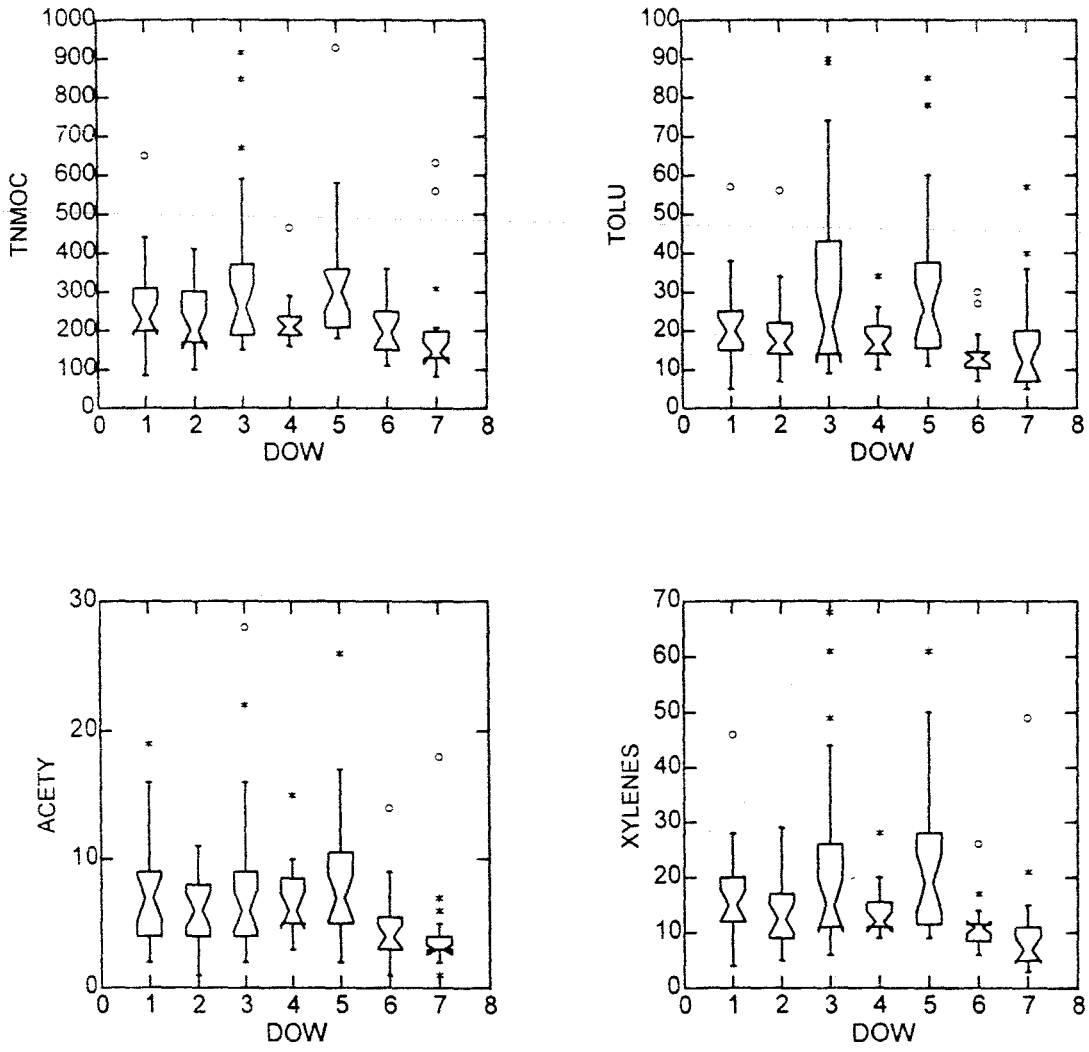


Figure 3-53. Notched box plots of morning (0500 PST) NMHC (shown as TNMOC), toluene (TOLU), acetylene (ACETY), and total xylenes concentrations (ppbC) at San Diego by day of week in 1997 (Monday = 1, Sunday = 7).

Results:

Group	N	Mean	SD
1996	18	11.167	6.271
1997	25	8.800	5.515

Separate Variance t = 1.283 df = 33.8 Prob = 0.208
Difference in Means = 2.367 95.00% CI = -1.382 to 6.116

Pooled Variance t = 1.311 df = 41 Prob = 0.197
Difference in Means = 2.367 95.00% CI = -1.279 to 6.013

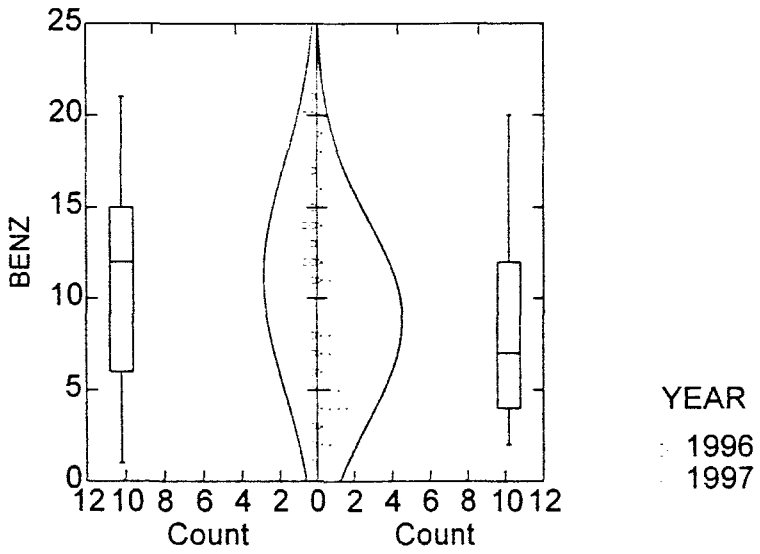
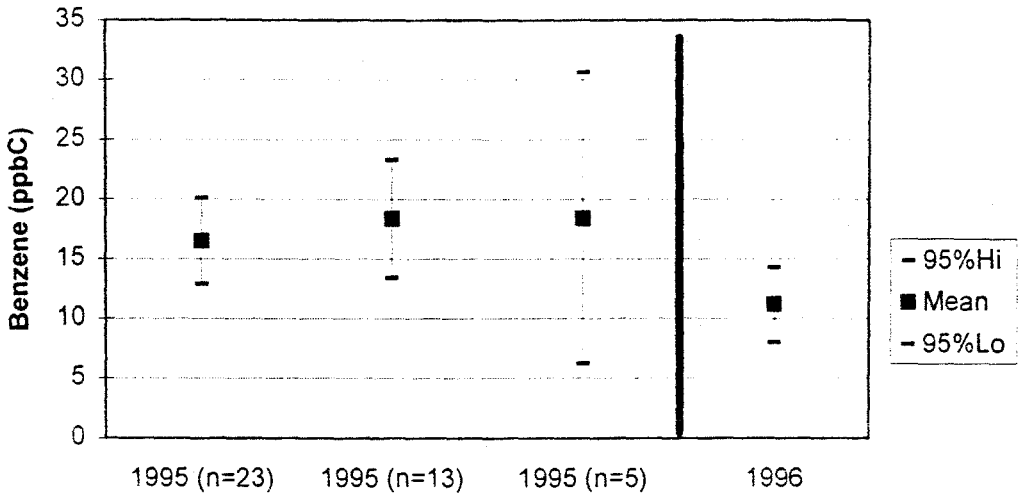


Figure 3-54. Results from a two-sample t-test on morning benzene concentrations grouped by year at Los Angeles. There was a 21 percent difference between the mean benzene concentrations in 1996 and 1997. The t-test indicates that there was a 20.8 percent chance (p value) of observing a difference as large as this (21 percent) even if the population means were identical.



T-test results (in comparison to 1995 data set):

1995 samples	23	13	5
Diff. of Means	5.4	5.2	5.2
p value	0.02	0.14	0.3

Figure 3-55. Illustration of the effect of sample size on the 95 percent confidence intervals and mean values of the Los Angeles morning benzene concentrations in 1995. Initially, the 1996 mean concentration was 32 percent lower than the 1995 mean. The confidence interval increases with a decreasing number of samples. The 1996 data set contained 25 samples. The t-test results for comparison of the three 1995 data sets with the 1996 data set are provided.

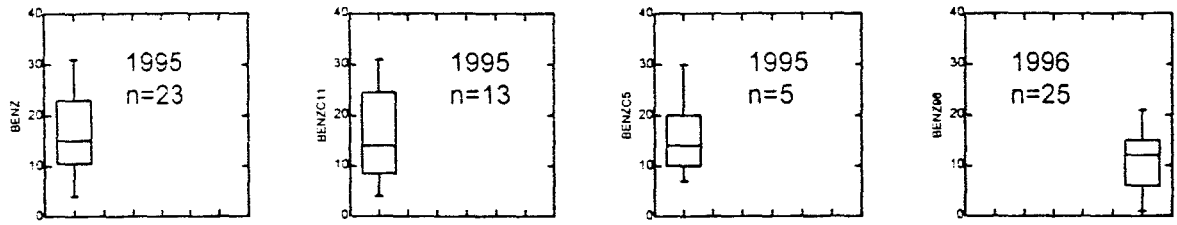


Figure 3-56. Box plots of the benzene concentration (ppbC) data shown in Figure 3-55. Each box plot indictes the year and the number of samples (n).

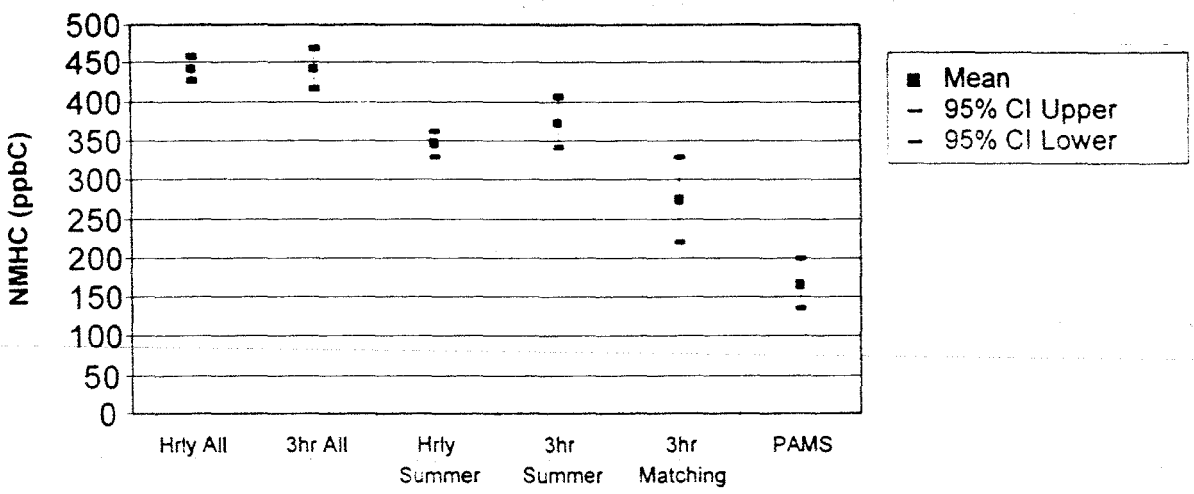


Figure 3-57. Mean NMHC concentrations and confidence intervals (CI) for the hourly and PAMS data collected at Clovis from 1995 through 1997. Hrly All = all hourly data collected year round from June 1, 1995 - November 30, 1997. 3hr All = 3-hr average data computed from the hourly values collected year round. Hrly Summer = hourly data collected from July - September, 1995-1997. 3hr Summer = 3-hr average data computed from the hourly values collected from July - September, 1995-1997. 3hr Matching = 3-hr average data computed from the hourly values collected on days for which PAMS NMHC data exist. PAMS = Canister, 3-hr average NMHC values from the PAMS data set.

Table 3-1. Median NMHC concentrations at PAMS and long-term trends sites in four California districts from 1990 through 1997. Districts include Ventura County, Sacramento, San Diego, and the San Joaquin Valley. See Table 2-2 for sampling details.

City	1990	1991	1992	1993	1994	1995	1996	1997
Elk Grove							73	67
Del Paso Manor					110	110	110	102
Folsom						100	93	87
Bakersfield					150	205	210	168
Arvin						100	100	107
Clovis						130	130	119
Fresno	220		120	170	130	120	140	125
Parlier						100	100	108
San Diego			211	219	160	190	180	196
Los Angeles	660	690	442	301	315	460	370	302
Emma Wood							172	199
El Rio						124	141	231
Simi Valley						169	144	191

Table 3-2. Median values based on 1997 valid data collected during all time periods. Exceptions are: El Rio data were of better quality in 1996. Solid lines separate Sacramento, Bakersfield, Fresno, Southern California, and Ventura areas.

Site	Site Type	NMHC (ppbC)	Unidentified (ppbC)	Isoprene %	Toluene/ Benzene
Elk Grove	1	67	53	1.5	3.0
Del Paso Manor	2	102	60	2.1	4.0
Folsom	3	87	55	3.2	4.0
Bakersfield	2	168	64	0.8	3.7
Arvin	1/3	107	69	1.2	3.0
Clovis	2	127	62	2.6	5.1
Fresno	Urban	127	64	1.1	4.0
Parlier	3	108	63	0.68	3.3
San Diego	Urban	230	70	0.43	4.3
Los Angeles	Urban	259	65	0.50	3.5
Emma Wood	1	74	19	0.22	1.6
El Rio	2	231	88	0.27	2.1
Simi Valley	3	108	45	1.2	2.7

Table 3-3. Median values of each species group based on 1997 valid data collected during all time periods. Exceptions are: El Rio data were of better quality in 1996. Solid lines separate Sacramento, Bakersfield, Fresno, Southern California, and Ventura areas.

Site	Site Type	NMHC (ppbC)	Paraffins %	Olefins %	Aromatics %	Unidentified %
Elk Grove	1	67	14	3.8	4.3	69
Del Paso Manor	2	102	24	6.5	9.5	57
Folsom	3	87	19	6.3	7.6	65
Bakersfield	2	168	41	5.6	14	39
Arvin	1/3	107	31	2.0	4.5	62
Clovis	2	127	27	6.8	13	51
Fresno	Urban	127	28	9.0	9.8	53
Parlier	3	108	28	5.5	7.8	59
San Diego	Urban	230	42	9.0	19	30
Los Angeles	Urban	259	50	9.1	16	24
Emma Wood	1	74	47	5.0	8.4	34
El Rio	2	231	46	7.1	9.4	37
Simi Valley	3	108	37	11	13	38

Table 3-4. Ten most abundant hydrocarbons on a weight percent basis at the California PAMS sites in 1997. Major sources of these hydrocarbons are provided in Table 3-6.

VOC	Elk Grove	Emma Wood	Arvin	Del Paso Manor	Golden State Ave.	Clovis	Fresno First	San Diego	Los Angeles	El Rio	Simi Valley	Folsom	Parlier
Site Type[a]	1	1	1/3	2A	2	2	Urban	Urban	Urban	2	3	3	3
Ethane	✔	✔	✔	✔	✔	✔	✔		✔	✔	✔	✔	✔
Toluene	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
i-Pentane	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Propane	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
n-Butane	✔	✔	✔	✔	✔	✔	✔		✔	✔	✔	✔	✔
n-Pentane	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Ethylene	✔			✔		✔	✔	✔	✔	✔	✔	✔	✔
m&p-xylenes	✔		✔	✔	✔	✔	✔	✔	✔		✔	✔	✔
Acetylene								✔	✔	✔	✔		
i-Butene				✔	✔	✔	✔	✔	✔				✔
Other[b]	prpyl isbta	isbta 2mpna mcpna 3mpna	ibuta ndec ispre	ispre	3ethex 2mpna	ispre	2mpna	2mpna 3ethex		1bute isbta	1bute	ispre ndec	2mpna

[a] 1 = upwind and background site, 2 = maximum ozone precursor emissions impact, 3 = maximum ozone concentration

[b] ibuta = i-butane; ispre = isoprene; ndec = n-decane; prpyl = propene; isbta = i-butane; 1bute = 1-butene; 2mpna = 2-methylpentane; mcpna = methylcyclopentane; 3mpna = 3-methylpentane; 3ethex = 3-ethylhexane; ndec = n-decane

Table 3-5. Ten most abundant hydrocarbons on a reactivity weighted basis at the California PAMS sites in 1997. Major sources of these hydrocarbons are provided in Table 3-6.

VOC	Elk Grove	Emma Wood	Arvin	Del Paso Manor	Golden State Ave.	Clovis	Fresno First	San Diego	Los Angeles	El Rio	Simi Valley	Folsom	Parlier
Site Type[a]	1	1	1/3	2A	2	2	Urban	Urban	Urban	2	3	3	3
Isoprene			✔	✔	✔	✔	✔				✔	✔	
Toluene	✔		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
i-Pentane	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Propene	✔	✔		✔	✔	✔	✔	✔	✔	✔		✔	✔
1,2,4-Trimethylbenzene	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
o-Xylene	✔		✔	✔	✔	✔	✔	✔	✔		✔	✔	✔
Ethylene	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
m&p-xylenes	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
i-Butene	✔		✔	✔	✔	✔	✔	✔	✔			✔	✔
Other[b]	mcpna nbuta	isbta propa nbuta mcpna 2mpna	1pnte nbuta	123tmb		13butad		mcpna 2mpna	mcpna 2mpna	1bute 123tmb propa nbuta	1bute mcpna 123tmb	1pnte	1pnte mcpna

[a] 1= upwind and background site, 2= maximum ozone precursor emissions impact, 3= maximum ozone concentration

[b] 13butad = 1,3-butadiene; mcpna = methylcyclopentane; 2mpna = 2-methylpentane; 1pnte = 1-pentene; nbuta = n-butane; 123tmb = 1,2,3-trimethylbenzene; mcpna = methylcyclopentane; 1bute = 1-butene; propa = propane; nbuta = n-butane

Table 3-6. Key species and their major sources (from Stoeckenius et al., 1994).

Species	Major Sources	Comments
acetylene, benzene	Motor vehicle exhaust, combustion	Tracer for vehicle exhaust
ethene	Motor vehicle exhaust, petrochemical industry	Tracer for vehicle exhaust
ethane	natural gas use	Non-reactive
propane	Natural gas use, oil and gas production	Relatively non-reactive, often underestimated in emission inventory.
i-butane	Consumer products, gasoline evaporative emissions, refining	Used as replacement of chlorofluorocarbons in consumer products
n-butane	Gasoline evaporative emissions	Tracer of gasoline use
isoprene	Biogenics	Tracer of biogenics, highly reactive
benzene	Motor vehicle exhaust, combustion processes, refining	Tracer for vehicle exhaust
toluene	Solvent use, refining, mobile sources	One of the most abundant species in urban air
olefins	Gasoline evaporative emissions, plastics	Reactive
xylenes	Solvent use, refining, mobile sources	Reactive

Table 3-7. Summary of hydrocarbons that exhibited a coefficient of variation greater than 50 percent for 1997 weight fraction data. Only hydrocarbons that were highly variable at at least three sites are included in this table; however, the totals by site represent all data. The comprehensive list by site is provided in Appendix C.

Hydrocarbon	Total by VOC	Elk Grove[a]	Del Paso Manor[b]	Folsom	Bakersfield	Arvin	Clovis	Fresno	Parlier	San Diego	Los Angeles[c]	Emma Wood	El Rio[a]	Simi Valley[a]
Site type[d]		1	2	3	2	1/3	2	"2"	3	"2"	"2"	1	2	3
Acetylene	3				✔					✔		✔		
Propene	5	✔				✔						✔	✔	✔
Propane	4	✔	✔	✔	✔									
i-Butane	7	✔		✔	✔	✔	✔	✔	✔					
n-Butane	3				✔		✔			✔				
c-2-Butene	3											✔	✔	✔
1-Pentene	7	✔			✔		✔		✔			✔	✔	✔
Isoprene	9		✔	✔	✔		✔	✔			✔	✔	✔	✔
t-2-Pentene	3							✔				✔		✔
c-2-Pentene	3											✔	✔	✔
Cyclopentene	3			✔	✔				✔					
Cyclopentane	4								✔	✔			✔	✔
2-Methyl-1-pentene	3											✔	✔	✔
2,3,4-Trimethylpentane	3	✔							✔			✔		
Toluene	3	✔					✔	✔						
n-Octane	5				✔			✔	✔	✔				✔
o-Xylene	3	✔		✔					✔					
Styrene	5				✔			✔		✔		✔		✔
n-Nonane	6			✔			✔	✔		✔			✔	✔
i-Propylbenzene	3											✔	✔	✔
1,3,5-trimethylbenzene	3	✔										✔		✔
1,2,4-trimethylbenzene	5	✔		✔		✔	✔	✔						
1,2,3-trimethylbenzene	5				✔	✔		✔	✔					✔
o-Ethyltoluene	3		✔									✔	✔	
m-Diethylbenzene	4							✔				✔	✔	✔
p-Diethylbenzene	5				✔						✔	✔	✔	✔
n-Decane	8		✔		✔	✔	✔	✔	✔	✔				✔
Undecane	10	✔	✔	✔	✔	✔	✔	✔	✔		✔			✔
Total by site (all data)		17	5	10	14	9	12	13	10	10	4	19	14	23

[a] COV values were greater than 100 percent.
[b] Most COV values were below 40 percent.
[c] Most COV values were below 30 percent.
[d] Site types: 1 = upwind and background site, 2 = maximum ozone precursor emissions impact, "2" = Type 2-like, 3 = maximum ozone concentration

Table 3-8. Selected hydrocarbon ratios measured at western urban U.S. tunnels.

Location	Ratio	1994	1995	1996	1997
Sepulveda Tunnel - Los Angeles[a]	benzene/acetylene		0.80	0.53	
	toluene/benzene		2.1	2.5	
	xylenes/benzene		2.2	2.3	
	propene/ethene		0.46	0.50	
Caldecott Tunnel - San Francisco Bay Area[b]	benzene/acetylene	2	1.7	1.1	1.6
	toluene/benzene	1.6	1.6	2.6	2.3
	xylenes/benzene	1.5	1.4	2.0	1.8
	propene/ethene	0.45	0.55	0.59	0.51
Cassiar Tunnel - Vancouver, BC[c]	benzene/acetylene		1.3		
	toluene/benzene		1.9		
	xylenes/benzene		1.4		
	propene/ethene		0.56		

a O'Connor et al., 1998
b Kirchstetter et al., 1999
c Rogak et al., 1998

Table 3-9. Median 0500-0800 PST ambient hydrocarbon ratios at California PAMS sites in 1997. Solid lines separate Sacramento, Bakersfield, Fresno, Southern California, and Ventura areas.

Location	Site Type	toluene/benzene	xylenes/benzene
Elk Grove	1	3.0	2.5
Del Paso Manor	2	4.0	3.7
Folsom	3	4.5	3.5
Bakersfield	2	3.8	3.8
Arvin	1/3	2.0	2.5
Clovis	2	4.3	3.3
Fresno	Urban	3.9	3.3
Parlier	3	3.5	3.0
San Diego	Urban	4.3	3.0
Los Angeles	Urban	3.6	3.0
Emma Wood	1	2.3	2.0
El Rio (1996)	2	2.3	2.4
Simi Valley	3	2.7	2.4

Table 3-10. Threshold values of VOC- or NO_x-limited indicators (afternoon concentrations or ratios).

Indicator	Threshold for NO_x-Limitation	Reference
NO_y	< 10 to 25 ppb	Milford et al., 1994
NO_z	< 5 to 20 ppb	Milford et al., 1994
Ozone/NO_y	> 5 to 10	Sillman, 1995; Jacob et al., 1995
Ozone/NO_z	> 6 to 11	Sillman, 1995; Jacob et al., 1995
(Ozone - 40 ppb)/NO_y	> 4	Sillman, 1995; Jacob et al., 1995
Formaldehyde/NO_y	> 0.2 to 0.4	Sillman, 1995; Jacob et al., 1995
Hydrogen Peroxide/Nitric Acid	>0.3 to 0.5	Sillman, 1995; Jacob et al., 1995
Hydrogen Peroxide/NO_y	> 0.2 to 0.4	Sillman, 1995; Jacob et al., 1995
Hydrogen Peroxide/NO_z	> 0.2	Sillman, 1995; Jacob et al., 1995
Morning NMOC/NO_x	> 10 to 15	National Research Council (1991)
Afternoon NMOC/NO_x	> 15 to 20	National Research Council (1991)

NO_x = NO + NO_2 + poorly defined fraction of other NO_y species (given conventional analyzers)
NO_y = NO_x + HNO_3 + organic nitrates + inorganic nitrates = NO_x + NO_z
NO_z = Oxidation products of NO_x = NO_y* (1 - NO_x/NO_y)
NMOC = nonmethane organic compounds

Table 3-11. Summary statistics for Del Paso Manor in 1995.

Statistic	NMHC	
	Daily	0500-0800 PST
No. of Samples	107	27
Mean	153	198
80 % CI Upper	169	236
80 % CI Lower	136	161

Table 3-12. Summary statistics for hourly NMHC and PAMS concentrations collected during 0500-0700 PST, 1995-1997 at Clovis.

Statistics	Hourly All[a]	3-hr Avg. All[b]	Hourly summer[c]	3-hr Avg. Summer[d]	Matching 3-hr from 1-hr[e]	PAMS[f]
No.	2540	851	933	237	73	73
Minimum	0	0	0	0	0	49
Maximum	3770	2300	1770	1660	1167	640
Median	300	333	280	313	207	110
Mean	444	443	346	374	275	167
95% CI Upper	459	468	363	407	330	199
95% CI Lower	428	418	329	341	219	135
Std. Dev.	399	374	262	260	238	139
CV	90	84	76	69	87	83

[a] Hrly All = all hourly data collected year round from June 1, 1995 - November 30, 1997.
[b] 3 hr All = 3-hr average data computed from the hourly values collected year round.
[c] Hrly Summer = hourly data collected during July - September, 1995-1997.
[d] 3 hr Summer = 3-hr average data computed from the hourly values collected during July - September, 1995-1997.
[e] 3 hr matching = 3-hr average data computed from the hourly values collected on days for which PAMS NMHC data exist.
[f] PAMS = Canister, 3-hr average NMHC values from the PAMS data set.

4. CONCLUSIONS AND RECOMMENDATIONS

The EPA, California ARB, and the Sacramento, San Joaquin Valley, and Ventura Districts sponsored research into the validation and analysis of the VOC data collected at the District's PAMS sites and the ARB's long-term trend sites located in Fresno, Los Angeles, and San Diego. This report summarizes VOC data validation and analyses performed for these entities. This section provides conclusions and recommendations based on the analyses.

4.1 EMISSION SOURCE ISSUES

Ultimately, the compositions of VOC samples are used to assess which hydrocarbons are most important in ozone formation and then to determine the likely sources of those hydrocarbons. Control strategies can then be developed to target these sources. We used assessments of the abundant hydrocarbons, relationships among hydrocarbons, temporal and spatial variations among the hydrocarbons, and key tracers of typical VOC sources to assess qualitatively what sources were important on a site-by-site basis. Conclusions from these analyses include the following:

- Nearly all analyses provided evidence that motor vehicle exhaust is very important at every site in California.
- While the isoprene percentage is low (less than 2 percent at most sites), it is still an important ozone precursor when reactivity is considered. Isoprene is the only tracer for biogenic emissions that is reported in the PAMS program.
- Benzene is less abundant than in the past; benzene's decrease in importance reflects changes in fuel content due to RFG.
- The olefin i-butene, a combustion product of MTBE, is more abundant in ambient air after the implementation of RFG. When i-butene is present, it generally correlates well with other exhaust-related hydrocarbons as found in other studies.
- Not all the VOCs appear to be related solely to motor vehicle exhaust. Unique sources of some of the abundant hydrocarbons are apparent at several sites including the following:
    - Oil production, natural gas, and other sources of primarily paraffins appear important at the Bakersfield, Ventura, and Southern California sites.
    - High NMHC concentrations due to propane, butanes, and pentanes at Emma Wood show that nearby sources sometimes influence this site.
    - In Los Angeles and San Diego, n-butane appears to have a significant source (or sources) other than motor vehicle exhaust. Also at Los Angeles, toluene, i-butane, and undecane appear to have non-exhaust sources.
    - At Clovis, unique nearby sources of toluene (especially midday) and decane appear to impact the site.

- At Fresno, a unique nearby source of i-butane appears to impact the site.
- At Folsom, a toluene source separate from the xylenes and benzene appears to impact the site.

We did not perform a quantitative assessment of source apportionment using the Chemical Mass Balance Model because of the lack of suitable source profiles and limitations in contract resources. We recommend that future source apportionment efforts using the California PAMS data consider the following:

- As a useful analysis prior to the more labor-intensive CMB analyses, cluster and factor analyses should be repeated for different data subsets including: data collected at different times of day, data collected under selected wind directions or ozone concentrations, concentration and weight percent data, a larger number of hydrocarbons, and the addition of air quality parameters such as NO_x or CO. For example, cluster and factor analyses of the sample compositions indicated common natural groupings of the hydrocarbons including benzene, ethene, acetylene (motor vehicle exhaust); butanes and n-pentane (motor vehicle evaporative emissions); propane and ethane (natural gas); i-pentane and toluene (motor vehicle and solvent use); and n-decane (an unknown, possibly industrial source or diesel emissions). These analyses help the analyst determine possible sources to include in the source profile database.
- Results from this report and from additional cluster/factor analyses should be used to make sure that all probable source types are included in future CMB analyses and that all important and unique hydrocarbons are included in at least one source profile.
- Recent CMB analyses performed in California indicated the need for more up-to-date source profiles (Magliano, 1998). Up-to-date and location specific source profiles will need to be prepared in order to perform a high quality source apportionment study using the PAMS data. Selection of proper source profiles is critical to a defensible source apportionment analysis. CMB modeling using PAMS data is planned for a South Coast Air Quality Management District contract during 1999.

4.2 DIURNAL VARIATION IN THE DATA

Significant changes with time of day were observed in the hydrocarbon composition at most sites, even the Type 2 sites. This finding indicates the need for the collection of speciated samples during all sampling periods. Specific conclusions from our investigation of the diurnal variation in the data include the following:

- On an individual hydrocarbon basis, variability in the weight fractions at each site is likely due to changes in emissions between sampling periods (e.g., biogenic isoprene and evaporative butane emissions midday are greater than in the morning), changes in reactivity (e.g., some olefins and aromatic hydrocarbons react away faster midday relative to paraffins), analytical difficulties (e.g., documented problems with acetylene and column recovery issues with decane and undecane), or combinations of these.

- At nearly all sites, isoprene contributions are higher midday and in the afternoon, which is consistent with biogenic emission patterns.
- Many anthropogenic hydrocarbons in the sample composition showed little change with time of day at the Type 2 and Type 2-like sites indicating that fresh emissions impact the sites during all sampling periods. However, there were several interesting exceptions that should be considered in future source apportionment work including:
    - At Los Angeles, the n-butane contribution to the NMHC is higher at midday, possibly due to increased evaporative emissions.
    - At Clovis, the toluene contribution is higher midday possibly indicating a midday source of toluene unrelated to the other aromatic hydrocarbons.
    - At El Rio, 1-butene contributions are higher at midday and during the afternoon sampling periods. This is counter to our understanding of reactivity; we would expect this highly reactive hydrocarbon to be depleted at midday relative to the morning.
    - At Fresno and Del Paso Manor, many of the hydrocarbons had their highest fractions in the 0500 PST samples and their lowest contributions in the midday samples. The diurnal behavior of the hydrocarbons at these sites was more typical of a downwind site (i.e., Type 3) than a city-center site (i.e., Type 2).
    - At Bakersfield, nighttime weight fractions of styrene, 1,3,5-trimethylbenzene, decane, and undecane vary significantly possibly indicating the impact of emissions under varying meteorological conditions. Hydrocarbon weight fractions of decane and undecane at Clovis show similar nighttime variability.
- More-reactive species declined relative to less-reactive species at midday and in the afternoon at most non-Type 2 sites. For example, the less-reactive C2-C4 paraffins were dominant hydrocarbons at many of the non-Type 2 sites. This is consistent with our conceptual model of transport and VOC reactivity. As the air parcels travel downwind of the emission sources (e.g., from a Type 2 site to a Type 3 site), the heavier paraffins and aromatic hydrocarbons react away faster relative to the lighter paraffins. At Emma Wood, the unidentified contribution was much higher during the daytime than in the morning probably indicating more aged air impacting the site in the afternoon. At these sites, source apportionment of the midday and afternoon samples will have more uncertainty because the data are less likely to meet the CMB model assumption that the hydrocarbons do not change between source and receptor.
- In general, Ventura County hydrocarbons showed greater variability in hydrocarbon concentrations and weight fractions than the other sites. This variability will introduce more uncertainty in the source apportionment investigation, for example.

4.3 YEAR-TO-YEAR VARIATION IN THE DATA

Significant changes in individual hydrocarbon weight fractions, hydrocarbon ratios, and hydrocarbon concentrations from year to year were observed at most sites and these changes were related to documented changes in gasoline formulation. The most dramatic change in ambient VOC composition from year to year was a significant decrease in the benzene fraction. In contrast, we did not observe significant changes in total NMHC concentrations from year to year. These findings indicate the need for a sufficient number of speciated samples to be collected each year in order to assess these annual changes. Discussion of the trends in VOCs is provided separately (Wittig et al., 1999).

The changes in benzene, i-butene, C5-C7 paraffins, and n-butane that were observed in the California data are consistent with fuel changes to reduce benzene, aromatics, and fuel volatility to conform to RFG regulations. Federal Phase I RFG was implemented in Southern California (including Ventura County) before summer 1995. California Phase 2 RFG was implemented in all of California before the summer of 1996. In comparison of the annual median fingerprints and benzene to toluene scatter plots, we noted the following specific findings:

- The benzene fraction decreased significantly between 1995 and 1996 at the Los Angeles, San Diego, Fresno, Clovis, Bakersfield, and Del Paso Manor Type 2 and Type 2-like sites and at the Parlier Type 3 site most likely due to California Phase 2 RFG implementation. The fractions of many C5-C7 paraffins increased between 1995 and 1996/1997 at these sites. Also at Los Angeles, San Diego, Fresno, and Bakersfield, there was a significant decrease in the n-butane fraction during this time period.
- The benzene fraction decreased significantly between 1994 and 1995 in Los Angeles, most likely due to Federal Phase I RFG implementation.
- At Los Angeles and San Diego, there was a significant increase in the i-butene fraction after 1994. This is understandable since i-butene is a thermal decomposition product of MTBE; MTBE is an oxygenated species that was added to RFG to meet oxygen guidelines. These findings were consistent with other investigations (Zielinska et al., 1997 and Kirchstetter et al., 1999).
- Benzene fractions did not appear to decrease between 1995 and 1997 at Simi Valley. These results are not consistent with the results at other sites or with our understanding of RFG implementation.

4.4 SITING CHARACTERISTICS

We investigated the VOC data with respect to PAMS siting criteria. For example, Type 1 sites should provide upwind and background characterization of the ozone and its precursors, Type 2 sites should show fresh emissions, and Type 3 sites should represent maximum ozone concentrations downwind of maximum precursor emissions. In general, the

VOC data investigated as a part of this project were consistent with siting criteria. For example, NMHC concentrations were typically highest at Type 2 sites and lower at Type 1 than at Type 3 sites. We also found that the composition appeared to be more aged at Type 1 and Type 3 sites than at Type 2 sites. The principal exception in the database is that the Simi Valley VOC data appear more typical of a Type 2 than a Type 3 site. Hydrocarbon relationships, temporal behavior of the composition, and hydrocarbon ratios at this site are very similar to the El Rio Type 2 site data. While siting appears adequate for secondary pollutant formation, VOCs are influenced by fresh emissions. This needs to be taken into account when performing source apportionment.

4.5 ASSESSING VOC VERSUS NO_x CONTROLS

The ratio of NMHC to NO_x in the morning is an important parameter for photochemical systems. The ratio characterizes the efficiency of ozone formation in NMOC-NO_x-air mixtures. At low ratios (less than 10), ozone formation is slow and inefficient (VOC-limited). Decreasing NO_x levels may result in increased ozone, due to less NO titration of ozone. At high ratios (greater than 15), ozone formation is limited by availability of NO_x rather than VOC. Under transitional regimes, both NO_x and VOC controls may be effective. One of the important uses of PAMS VOC data is to assess the need for VOC versus NO_x controls in a region. Most of the sites (except Emma Wood and El Rio) indicate that the majority of the morning NMHC/NO_x ratios are in the VOC-limited regime. These results are not surprising for the urban sites because other recent investigations of PAMS sites in the United States have shown similar morning ratios (e.g., Main et al., 1999). However, no single analysis should form the basis for emission control strategy decisions. Rather, we recommend that several analyses of indicators, application of the SP algorithm, the results from other data analysis tasks in this project, review of results from other projects, and modeling should provide concurrent evidence.

4.6 DAY-OF-WEEK ISSUES

Since we found that many of the hydrocarbons measured in the PAMS program are most likely associated with motor vehicle emissions, we would expect the concentrations to be a function of driving patterns. Since traffic patterns largely depend on the Monday-Friday workweek, we might expect to see large differences in hydrocarbon concentrations as a function of day of the week. Given the relatively large annual differences noted in several of the important hydrocarbons, we primarily investigated the 1997 data for the analyses. Since there are few samples per day of week at each site, we used notched box plots as a visual tool to assess daily differences. These plots take into account the number of samples by drawing confidence intervals around the median values. Most sites show no significant difference in NMHC concentration, individual hydrocarbon concentration, or hydrocarbon weight fraction by day of week. A preliminary conclusion from this analysis is that morning emissions and ambient composition are not a strong function of day of the week. However, because of the small number of samples involved, other analyses should be performed to corroborate these

results (including day of week distributions of the continuous NMHC data and other air quality parameters). Exceptions were found at Bakersfield and San Diego:

- Ambient NMHC and toluene concentrations on Sunday mornings appear to be lower than on Monday through Saturday mornings at Bakersfield. On a weight percent basis, there was a less significant difference among the days indicating a similar source mixture every day.
- Concentrations of NMHC, xylenes, toluene, and other hydrocarbons were lower on the weekends than during the weekdays indicating that weekend morning emissions appear to be lower than weekday emissions in San Diego. On a weight percent basis, there was no difference among the days indicating a similar source mixture every day.

4.7 CONTINUOUS NMHC DATA

We made a preliminary assessment of the continuous NMHC data focusing on one site (Clovis). These conclusions may or may not apply to data from other sites. For Clovis, we found the following:

- The summer morning NMHC concentrations are lower than the year-round concentrations. We have observed this seasonal variation in the past.
- The morning PAMS NMHC concentrations are lower than the hourly NMHC concentrations. This is not a surprising finding as there are analytical differences between the two measurements. For example, a dryer is used prior to analysis of the PAMS samples; Nafion dryers have been shown to remove oxygenates and some hydrocarbons (Gong and Demerjian, 1995). A dryer is not used for the hourly NMHC measurements.
- Morning NMHC concentrations on days on which PAMS samples were collected do not appear to represent the summer morning concentrations. For the PAMS sampling days, the 3-hr average NMHC concentrations (from the continuous data) have significantly lower mean and median values than the entire summer sampling period. This means that for Clovis, we need to use caution in interpreting broad implications from PAMS data analyses because they may not be representative of the entire summer.
- The PAMS NMHC concentrations do not correlate with the 3-hr averaged continuous NMHC data during the daytime. It is possible that this is a function of the oxygenated hydrocarbons and their affect on the continuous measurements. The practical implications of this are: 1) it would be inappropriate at this time to apply an average fingerprint from the PAMS data to the continuous NMHC data during the daytime and 2) it would be inappropriate to apply any conclusions regarding daytime trends in the continuous NMHC data to the PAMS data.

Additional investigations of the continuous NMHC data and their relationship to the PAMS NMHC data should be made using other times of day and other sites to see if these conclusions are applicable to all sites or a function of a specific site.

4.8 SAMPLE SIZE ISSUES

One issue of concern to analysts using the California data sets is the effect of sample size on the analyses. To address this, we used various statistical representations of the data (such as t-tests or notched box plots) to take sample size into account throughout our analyses. A closely related issue of concern to the reporting agencies is the adequacy of the data to meet PAMS data quality objectives. To address this concern, we used examples from the data to illustrate two of the PAMS data quality objectives.

The examples in this project seem to indicate that a minimum of about 25 samples is required to observe significant changes in an individual hydrocarbon between years or to observe a diurnal pattern. This means that the collection and speciation of samples every third day during July through September appears to be the minimum for these types of analyses and that fewer samples than this would fail to meet data quality objectives.

While 25 samples was probably adequate to show the relatively large annual changes in ambient benzene, for example, other analyses may lead to more subtle differences and thus may require larger data sets. We need to consider other analyses (such as the following) that can be performed using the hydrocarbon data and the effects of sample size on those analyses.

- Emission inventory evaluations. When the VOC data are compared to the emission inventory, we need an adequate number of samples in order to assess the comparisons for different wind quadrants. We also need a sufficient number of samples above selected concentration thresholds (e.g., 100 ppbC NMHC) to ensure that we are comparing the inventory to fresh emissions well above background concentrations. Starting with a database containing only 25 samples will reduce the efficacy of these analyses.
- Day of week evaluations. Currently, for an every-third-day sampling regimen, most of the California sites will have only 3 or 4 samples per day of week in an ozone season. This did not appear to be an adequate number of samples with which to investigate day of week issues at most sites. Combining several years of data for this type of analysis is probably not feasible given the significant year to year changes observed (e.g., benzene). For this type of evaluation, daily samples are probably required.
- Ozone episode evaluations. Similar arguments apply to an analysis of VOC data for days with maximum ozone concentrations above selected thresholds. For this type of evaluation, daily samples are probably required.
- Receptor modeling. It is likely that there are insufficient samples on an every-third-day basis to quantitatively assess the possible contributions from various sources of hydrocarbons, particularly by wind quadrant. The number of samples necessary for the analysis will be a function of the estimated error in the source profiles and the standard deviation in the ambient data. Based on the results from other studies, it is likely that daily samples are required for this type of analysis. Further, the time resolution of the samples (i.e., 3-hr averages, four times per day) may not be adequate to discern

between various sources in different wind quadrants and during the entire 24-hr period. Hourly data are typically required for detailed source apportionment.

Further research into the issues of sample size should be made. Statistically, one can determine the number of samples required to observe a given percent change using a power calculation. The results of the power calculation will give the minimum change that can be statistically detected given the sample size. The calculation is a function of the sample mean, sample standard deviation, confidence interval, and expected change and is discussed by Walpole and Myers (1989) and other statisticians. For all analyses, the adequate number of samples is a function of the magnitude of the expected change, the site, the individual hydrocarbon or ratio, and the time of day (e.g., more variability was observed in the midday samples than in the morning samples). Sample size assessments should be made for data from each site for selected analyses (e.g., day of week issues, annual trends) and for several parameters (e.g., NMHC, benzene, butane, benzene to toluene ratio) to more fully assess the DQOs.

4.9 DATA VALIDATION AND MEASUREMENT ISSUES

We used VOCDat software to quality assure 1990 through 1997 PAMS (and PAMS-like) VOC data for four California air districts: Sacramento, San Joaquin Valley, San Diego, and Ventura (Main, 1998). In general, the data were of good quality and reflect the hard work of the monitoring and analytical staff at the ARB and Districts. Problems typical of PAMS data sets from other states were not commonly observed. The only remaining serious issue with the data is the quality of the acetylene data, particularly prior to 1997; problems with acetylene may reduce the applicability of some future analyses such as source apportionment.

With regard to data validation, we recommend that the ARB and Districts perform the following:

- Use graphical techniques to review data, in addition to current procedures, to catch any data not representative of ambient air conditions. Validation procedures should include applying automated screening checks for abundant hydrocarbons, hydrocarbon relationships, and outliers; inspecting time series and scatter plots of every hydrocarbon and hydrocarbon functional group (i.e., aromatics, paraffins, olefins, and unidentified); inspecting every "fingerprint" for all sites; and reviewing summary statistics for outliers. Once unusual samples have been identified, the samples should be rechecked to determine possible causes of the abnormalities.
- Put validated speciated VOC data into AIRS including data collected prior to the initiation of PAMS. These data are a valuable resource from which to assess trends in VOCs.
- Inspect the continuous NMHC data at all sites for (1) 0 concentrations (should be listed as below detection or missing); (2) abrupt changes in resolution or "baseline" concentrations; and (3) high values followed by a slow decline in concentration as this behavior may be indicative of some sort of contamination. Once the entire data set has

been carefully validated, we recommend that the continuous NMHC data be placed in AIRS.

- Laboratory audit results are critical for assessing analytical precision and accuracy. During laboratory audits, we recommend that samples be checked for the total NMHC, sum of PAMS target species, and unidentified concentrations as well as individual hydrocarbon concentrations.
- The through-the-probe audits are useful for assessing the entire sampling system. We recommend that future TTP challenges include acetylene. Again, total NMHC and unidentified concentrations should be assessed.
- The laboratories should continue to use the audit results as an effective feedback mechanism for improving their hydrocarbon analyses.

With regard to measurement issues, we recommend that the ARB and Districts perform the following:

- Investigate possible sources of decane and undecane, including diesel emissions, at several sites including Elk Grove, Folsom, Arvin, and Clovis. The ARB should consider taking samples of semi-volatile hydrocarbons to corroborate the canister data.
- The ARB should obtain more information regarding what species comprise the unidentified portion of the NMHC. The median unidentified hydrocarbon concentrations were similar among all non-Ventura County sites, ranging from 53 to 70 ppbC, regardless of site type. Further, the unidentified fractions were more than 50 percent at the Sacramento and Fresno area sites. The practical implication of this relatively high fraction is that analysis of possible sources and sample ozone formation issues are less meaningful because more than half the sample carbon is unaccountable. Unidentified compounds can include polar (often oxygenated) species when drying of the samples is not performed, PAMS species that were missed or misidentified, or VOCs not on the target list such as higher molecular weight alkanes and aromatics.
- The ARB and Districts should consider measuring ambient MTBE (or other fuel additives if MTBE is phased out) for additional tracking of RFG.
- The ARB and Districts should consider purchasing an auto-GC to operate continuously (i.e., hourly, daily) side-by-side with canister collection for a few weeks at each PAMS site. Data from this instrument would add more time resolution and provide better comparisons to the continuous NMHC values to better our understanding of emission sources impacting the sites and enable us to better address some of the measurement issues.

5. REFERENCES

Blanchard C.L., Lurmann F.W., Korc M.E., and Roth P.M. (1994) The use of ambient data to corroborate analyses of ozone control strategies. Final report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC by Sonoma Technology, Inc., Santa Rosa, CA and Envair, San Anselmo, CA, STI-94030-1433-FR, Contract No. 68D30020, December.

Carter W.P.L. (1994) Development of ozone reactivity scales for volatile organic compounds. *J. Air & Waste Manag. Assoc.* **44**, 881-899.

Fujita E.M., Watson J.G., Chow J.C., and Lu Z. (1994) Validation of the chemical mass balance receptor model applied to hydrocarbon source apportionment in the Southern California Air Study. *Environ. Sci. Technol.* **28**, 1633-1649.

Fujita E.M., Watson J.G., Chow J.C., and Magliano K.L. (1995) Receptor model and emissions inventory source apportionments of nonmethane organic gases in California's San Joaquin Valley and San Francisco Bay Area. *Atmos. Environ.* **29**, 3019-3035.

Gertler A.W., Fujita E.M., Pierson W.R., and Wittorff D.N. (1996) Apportionment of NMHC tailpipe vs non-tailpipe emissions in the Fort McHenry and Tuscarora Moutain tunnels. *Atmos. Environ.* **30**, 2297-2305.

Gong Q. and Demerjian K.L. (1995) Hydrocarbon losses on a regenerated Nafion dryer. *J. Air & Waste Manag. Assoc.* **45**, 490-493.

Hammond D. (1996) Ambient trends of benzene in California from 1990 through 1995. Paper presented at the *U.S. Environmental Protection Agency and Air & Waste Management Association International Symposium on Measurement of Toxic and Related Air Pollutants, Research Triangle Park, NC, May 7-9.*

Henry R.C. (1992) Dealing with near collinearity in chemical mass balance receptor models. *Atmos. Environ.* **26A**, 933-938.

Jacob D.J., Horowitz L.W., Munger J.W., Heikes B.G., Dickerson R.R., Artz R.S., and Keene W.C. (1995) Seasonal transition from NO_X to hydrocarbon-limited conditions for ozone production over the eastern United States in September. *J. Geophys. Res.* **100**, 9315-9324.

Kirchstetter T.W., Singer B.C., Harley R.A., Kendall G.R., and Hesson J.M. (1999) Impact of California reformulated gasoline on motor vehicle emissios. 2. Volatile organic compound speciation and reactivity. *Environ. Sci. Technol.* **33**, 329-336.

Lin C. and Milford J.B. (1994) Decay-adjusted chemical mass balance receptor modeling for volatile organic compounds. *Atmos. Environ.* **28**, 3261-3276.

Lu Z. and Fujita E.M. (1995) Volatile organic compound source apportionment for the coastal oxidant assessment for Southeast Texas Study. Final report prepared for Texas Natural Resource Conservation Commission, Austin, TX by Desert Research Institute, Reno, NV.

Magliano K. (1998) Chemical mass balance modeling of data from the 1995 Integrated Monitoring Study. Final report prepared for the San Joaquin Valleywide Air Pollution Study Agency, Sacramento, CA by the California Air Resources Board, Sacramento, CA, July.

Main H.H. (1998) Status report on VOCDat software technical support, demonstrations of the utility of VOCDat software, and related analyses. Technical memorandum to D. Hammond, California Air Resources Board, August 18, 1998.

Main H.H., Roberts P.T., and Chinkin L.R. (1997) PAMS data analysis workshop: illustrating the use of PAMS data to support ozone control programs. Prepared for U.S. Environmental Protection Agency, Research Triangle Park, NC, presented at California Air Resources Board and EPA Region IX, Sacramento, CA, STI-997100-1719-WD7, May.

Main H.H., Roberts P.T., and Prouty J.P. (1998a) VOCDat user's guide. Report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC by Sonoma Technology, Inc., Petaluma, CA, STI-997160-1763-DFR2, July.

Main H.H., Roberts P.T., Reiss R., and Chinkin L.R. (1998b) Analysis of PAMS data to evaluate reformulated gasoline effects. Paper presented at the *Air and Waste Management Association 1998 Annual Meeting and Exhibition, San Diego, CA, June 14-18.*

Main H.H., Hurwitt S.B., and Roberts P.T. (1999) Characteristics of volatile organic compounds in the Mid-Atlantic region. Report prepared for MARAMA, Baltimore, MD by Sonoma Technology, Inc., Petaluma, CA, STI-998482-1869-FR, March.

Miguel M.G. (1998) Laboratory audit results, through-the probe audits, and other material provided to H. Main. September 24, 1998.

Milford J.B., Gao D., Sillman S., Blossey P., and Russell A.G. (1994) Total reactive nitrogen (NO_y) as an indicator of the sensitivity of ozone to reductions in hydrocarbon and NO_x emissions. *J. Geophys. Res.* **99**, 3533-3542.

National Research Council (1991) *Rethinking the Ozone Problem in Urban and Regional Air Pollution*. National Academy of Sciences/National Research Council, National Academy Press, Washington, DC.

Nelson P.F. and Quigley S.M. (1983) The m, p-xylenes: ethylbenzene ratio, a technique for estimating hydrocarbon age in ambient atmospheres. *Atmos. Environ.* **17**, 659-662.

O'Connor C.M., Gertler A.W., and Sagebiel J.C. (1998) A comparison of ozone-forming potential of on-road emissions measured at the sepulveda tunnel between 1995 and 1996. Presented at the *Air & Waste Management Association's 91st Annual Meeting & Exhibition, San Diego, California, June 14-18.*

Pace T.G. and Watson J.G. (1987) Protocol for applying and validating the CMB model. Report prepared by Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-450/4-87-010, May.

Rogak S.N., Pott U., Dann T., and Wang D. (1998) Gaseous emissions from vehicles in a traffic tunnel in Vancouver, BC. *J. Air & Waste Manag. Assoc.* **48**, 604-615.

Scheff P.A. and Wadden R.A. (1993) Receptor modeling of volatile organic compounds: 1. emission inventory and validation. *Environ. Sci. Technol.* **27**, 617-625.

Sillman S. (1995) The use of NO_y, H_2O_2, and HNO_3 as indicators for ozone-NO_x-hydrocarbon sensitivity in urban locations. *J. Geophys. Res.* **100**, 14175-14188.

Stoeckenius T.E., Ligocki M.P., Cohen B.L., Rosenbaum A.S., and Douglas S.G. (1994) Recommendations for analysis of PAMS data. Final report prepared by Systems Applications International, San Rafael, CA, SYSAPP94-94/011r1, February.

Stoeckenius T., Yarwood G., Shepard S., Taylor A., Lee K., Main H.H., and Haste T.L. (1998) Analysis of data from the 1995 NARSTO-northeast air quality study. Volume I: data validation and statistical summaries of routine data. Report prepared for Coordinating Research Council, Atlanta, GA by ENVIRON International Corporation, Novato, CA, and Sonoma Technology, Inc., Petaluma, CA, December.

U.S. Environmental Protection Agency (1987) Protocol for reconciling differences among receptor and dispersion models. Report prepared by Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-450/4-87-008, March.

U.S. Environmental Protection Agency (1994) Photochemical assessment monitoring stations implementation manual. Report prepared by Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-454/B-93-051, March.

U.S. Environmental Protection Agency (1998) Technical assistance document for sampling and analysis of ozone precursors. Report prepared by the National Exposure Research Laboratory, Research Triangle Park, NC, EPA/600-R-98/161, September.

Walpole R.E. and Myers R.H. (1989) *Probability and statistics for engineers and scientists*, 4th Edition. McMillan Publishing Company, NY.

White W.H. (1999) Phantom spatial factors: an example. *J. Air & Waste Manag. Assoc.* **49**, 345-349.

Wittig A.E., Main H.H., Roberts P.T., Hurwitt S.H. (1999) Trends analyses of California PAMS and long-term trend air quality data (1987-1997). Report prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC by Sonoma Technology, Inc., Petaluma, CA, STI-998393-1885-FR, May.

Zielinska B., Shire J., Harshfield G., Sagebiel J., and Pasek R. (1997) VOC concentrations in the South Coast Air Basin (CA) following introduction of reformulated gasoline. Paper presented at the *Air & Waste Management Association Measurement of Toxic and Related Air Pollutants Symposium Conference, Raleigh Durham, NC, April 29-May 1.*

APPENDIX A

LIST OF FLAGGED SAMPLES FOR THE CALIFORNIA PAMS AND TRENDS SITES 1990-1997

This appendix contains tables that list the samples and individual hydrocarbons that were flagged as suspect (S) or invalid (I). Data validation, when done well, is labor-intensive. To assess the validity of the data, we inspected a time series plot of every species, prepared numerous scatter plots and scatter plot matrices, including a plot of every species and species group versus the total NMHC, and inspected the fingerprint plot of every sample. Our strategy is to flag entire samples when we identify a problem with two or more of the most abundant species (e.g., toluene, i-pentane, n-pentane, i-butane, n-butane, benzene, acetylene, ethane, xylenes, ethene). We also flag samples when only one hydrocarbon has been identified as problematic if that hydrocarbon represents a significant portion of the total NMHC (e.g., more than 20 percent). Individual species are flagged as suspect when there are problems noted, but the concentration of the hydrocarbon is low compared to other species in the sample. Data are flagged that do not fit our conceptual model of ambient VOC concentrations and relationships. In some cases, the samples may have an analytical error that needs to be corrected, a sampling problem that was inadvertently missed, or may just appear odd. Upon review of the data by the reporting agencies, some suspect data may be deemed invalid, some data may be reprocessed and re-reported by the agency, and other data may remain suspect (cause unknown). These latter samples remain flagged in our working database and in this report for the attention of data analysts.

The tables in this appendix are set up as follows:

- Date = Date of sample(s) flagged
- Time = Begin Hour(s) in standard time of the flagged sample(s)

- QC = the quality control code of S (suspect) or I (invalid). Invalid flags were typically only applied by the reporting agencies. For a sample or individual hydrocarbon to be flagged as invalid, a documented reason needed to be provided (e.g., instrument failure, known contamination, data not representative of ambient air such as calibration gas, etc.).

- Species or Sample = A designation of the individual hydrocarbon (HC) or the entire sample record (record) to which the flag applies.

- Flagged Samples = The number of samples flagged with the same code and reason (generally for the same date and used to shorten the list).

- Comments regarding why samples were flagged.

- Also included at the bottom of the files are comments for future data analysts that discuss additional observations of the data that may have a bearing on planned analyses or conclusions from the data.

The tables are provided as a check for data analysts who obtain their data from the California ARB CD-ROM to ensure that they have the most up-to-date data file available. The tables list the samples that we flagged for which no changes were made by the monitoring laboratory but that from a data analyst's point of view, do not fit our conceptual model of ambient hydrocarbon data. We typically recommend that these flagged data not be used in "over-arching" analyses (such as summary statistics) but be considered on a case study basis.

The tables include data quality information for the following sites: Elk Grove, Del Paso Manor, and Folsom; Bakersfield and Arvin; Clovis, Fresno, and Parlier; San Diego and Los Angeles; Emma Wood, El Rio, and Simi Valley; and Alpine, El Cajon, and Overland.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Clovis	8/2/93	5:00	S	Sample	1	Missing C2-C4 species; flame out
Clovis	8/2/93	12:00	S	Sample	1	Missing C2-C4 species; flame out
Clovis	8/9/94	16:00	S	Sample	1	Odd fingerprint, high concentrations
Clovis	8/1/95	5:00	S	Sample	1	High decane
Clovis	9/15/95	12:00	S	Sample	1	High concentrations of heavier species
Clovis	9/16/95	8:00	S	Sample	1	High concentration of 123tmb
Clovis	9/16/95	12:00	S	Sample	1	Odd fingerprint, high concentrations
Clovis	9/16/95	16:00	S	Sample	1	High decane
Clovis	9/23/95	23:00	S	Sample	1	High undecane
Clovis	9/24/95	12:00	S	Sample	1	High undecane
Clovis	9/26/95	23:00	S	Sample	1	High undecane
Clovis	7/6/96	5:00	S	Sample	1	High concentrations of heavier species
Clovis	7/11/96	16:00	S	Sample	1	High undecane
Clovis	8/9/96	12:00	S	Sample	1	High toluene
Clovis	8/9/96	16:00	S	Sample	1	ispnta QC not met
Clovis	9/3/96	12:00	S	Sample	1	High toluene, high 223tmbutane
Clovis	9/29/96	23:00	S	Sample	1	acetylene QC not met
Clovis	7/6/97	12:00	S	Sample	1	High decane
Clovis	7/9/97	5:00	S	Sample	1	High decane
Clovis	7/14/97	8:00	S	Sample	1	High concentrations of heavier species
Clovis	7/14/97	16:00	S	Sample	1	Missing C4+ species (no DB-1 column data)
Clovis	7/20/97	23:00	S	Sample	1	High decane
Clovis	7/24/97	5:00	S	Sample	1	ispnta QC not met
Clovis	7/30/97	12:00	S	Sample	1	High toluene, high 223tmbutane
Clovis	8/19/97	23:00	S	Sample	1	High toluene, high 223tmbutane
Clovis	8/22/97	23:00	S	Sample	1	High decane

Notes to data analysts: (608 samples)
Unidentified greater than 50% (203 of 586)

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Fresno First St.	6/28/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	7/1/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	7/19/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	7/22/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	7/25/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	7/28/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	8/3/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	8/9/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	8/12/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	8/27/90	500	S	Sample	1	PAMSHC > NMOC
Fresno First St.	9/5/90	500	S	Sample	1	PAMSHC > NMOC; acetylene did not meet QC
Fresno First St.	9/15/92	1200	S	Sample	1	PAMSHC > NMOC
Fresno First St.	9/27/92	1200	S	Sample	1	PAMSHC > NMOC
Fresno First St.	10/6/92	500	s	Sample	1	High 2,3,3-trimethylbutane
Fresno First St.	10/10/93	1200	s	Sample	1	Acetylene did not meet QC
Fresno First St.	7/1/95	2300	S	Sample	1	PAMSHC > NMOC
Fresno First St.	9/6/95	1200	s	Sample	1	High benzene relative to others; toluene < LOD
Fresno First St.	9/24/95	1600	s	Sample	1	High benzene relative to others; toluene < LOD
Fresno First St.	9/6/96	500	s	Sample	1	Acetylene did not meet QC
Fresno First St.	9/24/96	500	s	Sample	1	Acetylene did not meet QC
Fresno First St.	7/2/97	2300	s	Sample	1	High benzene relative to others
Fresno First St.	7/24/97	1200	S	Sample	1	High benzene relative to others
Fresno First St.	7/27/97	1600	S	Sample	1	i-pentane did not meet QC
Fresno First St.	9/30/97	2300	s	Sample	1	High benzene relative to others

Notes to data analysts: (896 samples)
In 1990, most of the samples are missing acetylene.
Overall, 176 of 846 samples have unidentified > 50 percent.
Missing NMOC for 33 samples, mostly in 1992; these data may not have met QC criteria.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Bakersfield Golden St.	6/30/94	2300	s	123tmb	1	High concentration
Bakersfield Golden St.	8/1/95	500	s	123tmb	1	High concentration
Bakersfield Golden St.	7/14/97	800	S	decane	1	High concentration
Bakersfield Golden St.	7/21/97	500	S	123tmb	1	High concentration
Bakersfield Golden St.	8/2/97	1600	s	sample	1	High concentration
Bakersfield Golden St.	8/14/97	1600	s	sample	1	propane did not meet QC criteria
Bakersfield Golden St.	9/3/97	2300	s	123tmb	1	High concentration
Bakersfield Golden St.	9/28/97	500	s	sample	1	m&p-xylenes below LOD

Notes to data analysts: (521 samples)
The acetylene concentrations were high compared to benzene in several samples.
The unidentified is greater than 50 % in 60 of 489 samples

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Arvin	7/7/95	23:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	7/11/95	5:00	S	Sample	1	High Unidentified concentration
Arvin	7/14/95	5:00	S	Sample	1	High propane concentrations
Arvin	7/28/95	5:00	S	Sample	1	High decane, undecane, Isopentane concentrations
Arvin	8/4/95	12:00	S	Sample	1	High undecane
Arvin	8/4/95	16:00	S	Sample	1	High undecane
Arvin	8/10/95	16:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	8/13/95	12:00	S	Sample	1	Toluene < LOD
Arvin	8/15/95	23:00	S	Sample	1	Only 124tmb, toluene, propane, ethane reported; others < LOD
Arvin	8/30/95	12:00	S	Sample	1	Toluene < LOD
Arvin	9/6/95	5:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	9/15/95	5:00	S	Sample	1	Low concentration of toluene vs Benzene
Arvin	9/17/95	16:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	7/1/96	23:00	S	Sample	1	High decane
Arvin	7/2/96	12:00	S	Sample	1	High decane, undecane
Arvin	7/6/96	8:00	S	Sample	1	High decane
Arvin	7/6/96	16:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	7/7/96	5:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	7/8/96	5:00	S	Sample	1	High concentrations of heavy species (C8+)
Arvin	7/17/96	16:00	S	Sample	1	High undecane
Arvin	7/23/96	12:00	S	Sample	1	High undecane
Arvin	7/26/96	16:00	S	Sample	1	High undecane
Arvin	8/13/96	12:00	S	Sample	1	High decane
Arvin	8/30/96	23:00	S	Sample	1	High decane
Arvin	9/11/96	23:00	S	Sample	1	Missing C4+ species; QC not met several species
Arvin	9/12/96	12:00	S	Sample	1	Missing C4+ species; QC not met several species
Arvin	9/14/96	23:00	S	Sample	1	Missing C4+ species; QC not met several species
Arvin	8/10/97	23:00	S	Sample	1	Missing C4+ species; QC not met several species
Arvin	8/16/97	23:00	S	Sample	1	Propane did not meet QC
Arvin	9/16/97	5:00	S	Sample	1	High 123tmb, Isopentane, Isobutane and n-hexane

Notes to data analysts: (384 samples)
Acetylene data appears troublesome - very low or missing in most samples.
Decane and undecane concentrations showed an "alligator plot" - indicating possible problems with the data or source issues.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Parlier	7/2/95	12:00	S	Sample	1	High decane
Parlier	7/17/95	12:00	S	Sample	1	Low acetylene, very high C7+, particularly decane, high unidentified
Parlier	7/20/95	12:00	S	NMOC	1	Sum of identified > NMOC
Parlier	8/25/95	12:00	S	Sample	1	High decane
Parlier	9/12/95	12:00	S	Sample	1	High decane
Parlier	9/12/95	16:00	S	Sample	1	Low acetylene, high unidentified, High decane, C7+ including oetol, 135tmb
Parlier	9/16/95	5:00	S	Sample	1	High mcyhex, nhept
Parlier	9/17/95	12:00	S	Unid	1	Low acetylene, high unidentified
Parlier	7/3/96	5:00	S	Sample	1	High 124tmb
Parlier	7/3/96	8:00	S	Sample	1	High C10+
Parlier	7/6/96	8:00	S	Sample	1	High undecane
Parlier	7/7/96	8:00	S	Sample	1	High C10+
Parlier	7/3/97	16:00	S	Sample	1	High 124tmb
Parlier	7/14/97	5:00	S	Sample	1	Missing C5+ data; no DB-1 data
Parlier	7/14/97	12:00	S	Sample	1	Missing C5+ data; no DB-1 data
Parlier	7/18/97	16:00	S	Sample	1	High C8+
Parlier	9/10/97	12:00	S	Sample	1	High decane
Parlier	9/28/97	16:00	S	Sample	1	High xylenes, ebenz

Notes to data analysts: (372 samples)
High propane concentrations observed frequently at this site.
186 of 351 samples had unidentified greater than 50 percent.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
12th Street	6/4/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	7/22/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	7/28/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	8/15/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	9/2/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	9/14/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	9/20/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	9/26/90	5:00	S	Sample	1	Abundant species missing or at LOD
12th Street	6/5/92	5, 12:00	S	Sample	2	Abundant species missing or at LOD
12th Street	6/23/92	5, 12:00	S	Sample	2	Abundant species missing or at LOD (ethane)
12th Street	7/2/92	5:00	S	Sample	1	High UNID / TNMOC
12th Street	7/11/92	12:00	S	Sample	1	Many species missing or at LOD
12th Street	8/28/92	5:00	S	Sample	1	Many species missing or at LOD
12th Street	9/24/92	5:00	S	Sample	1	Many species missing or at LOD
12th Street	10/9/92	5:00	S	Sample	1	Uncommonly high concentrations
12th Street	10/15/92	4:00	S	Sample	1	High concentration 224trimethylpentane
12th Street	6/21/93	5:00	S	Sample	1	Abundant species missing or at LOD (acetylene)
12th Street	6/24/93	5:00	S	Sample	1	Abundant species missing or at LOD (acetylene)
12th Street	6/27/93	5:00	S	Sample	1	Abundant species missing or at LOD (acetylene)
12th Street	9/28/93	5:00	S	Sample	1	Many species missing or at LOD
12th Street	7/4/94	12:00	S	Sample	1	Abundant species missing or at LOD (acetylene)
12th Street	7/27/94	5, 12:00	S	Sample	2	Missing ispna
12th Street	7/31/94	12:00	S	Sample	1	High concentration ispna
12th Street	8/17/94	5, 12:00	S	Sample	2	Abundant species missing
12th Street	8/18/94	5:00	S	Sample	1	High concentration ispna
12th Street	8/21/94	12:00	S	Sample	1	Abundant species missing or at LOD
12th Street	8/27/94	5:00	S	Sample	1	Toluene missing or at LOD
12th Street	8/30/94	5:00	S	Sample	1	UNID < 0
12th Street	9/2/94	5:00	S	Sample	1	High concentration ispna
12th Street	9/11/94	5:00	S	Sample	1	High concentration ispna
12th Street	9/11/94	12:00	S	Sample	1	High concentration ispna (>1ppmC)
12th Street	9/29/94	5:00	S	Sample	1	UNID < 0
12th Street	7/17/95	12:00	S	Sample	1	High concentration isbta
12th Street	8/31/95	5:00	S	Sample	1	High concentration ispna
12th Street	9/27/95	5:00	S	Sample	1	High concentration ispna
12th Street	9/27/95	12:00	S	Sample	1	Missing acetylene
12th Street	8/4/96	5, 12:00	S	Sample	2	Abundant species missing or at LOD
12th Street	8/7/96	12:00	S	Sample	1	High concentration isbta
12th Street	8/25/96	12:00	S	Sample	1	High concentration isbta
12th Street	9/18/96	5:00	S	Sample	1	Missing acetylene
12th Street	9/18/96	12:00	S	Sample	1	High concentration ispna , missing acetylene
12th Street	9/21/96	5, 12:00	S	Sample	2	Abundant species missing or at LOD
12th Street	7/6/97	12:00	S	Sample	1	acetylene < LOD
12th Street	7/12/97	5:00	S	Sample	1	Missing C4+ species
12th Street	7/24/97	5, 12:00	S	Sample	2	ispna did not meet QC criteria
12th Street	8/14/97	12:00	S	Sample	1	propane did not meet QC criteria
12th Street	9/22/97	5:00	S	Sample	1	High concentration ispna

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
L.A. N. Main	6/29/91	500	S	Sample	1	Several species odd concentrations
L.A. N. Main	6/5/92	500	S	Sample	1	Missing ethane
L.A. N. Main	6/5/92	1200	S	Sample	1	Missing ethane
L.A. N. Main	6/26/92	1200	S	Sample	1	Missing ethane
L.A. N. Main	7/2/92	500	S	Sample	1	Missing ethane
L.A. N. Main	7/2/92	1200	S	Sample	1	Missing ethane
L.A. N. Main	9/5/94	1200	S	Sample	1	High concentration several species
L.A. N. Main	9/23/94	500	S	Sample	1	high concentration several species
L.A. N. Main	9/12/96	1200	S	Sample	1	Missing acetylene, odd fingerprint
L.A. N. Main	9/21/96	500	S	Sample	1	Missing acetylene
L.A. N. Main	9/27/96	500	S	Sample	1	missing acetylene
L.A. N. Main	9/27/96	1200	S	Sample	1	missing acetylene
L.A. N. Main	9/30/96	500	S	Sample	1	missing acetylene
L.A. N. Main	9/30/96	1200	S	Sample	1	missing acetylene
L.A. N. Main	7/12/97	1200	S	Sample	1	Missing species above C4
L.A. N. Main	8/20/97	1200	S	Sample	1	High concentration several species
L.A. N. Main	8/23/97	500	S	Sample	1	missing acetylene
L.A. N. Main	8/26/97	500	S	Sample	1	missing acetylene

Notes to data analysts: (383 samples)
In 1990:
Many problems with the entire year including: possible misidentification (shown as split scatter plots and outliers) of 22dmb cypenta, 24dmp; many samples missing acetylene; nonane and o-xylene coelute.
In 1991:
Acetylene data problematic again - ethane&acetylene coelutions reported, but sometimes when ethane is reported separately.
In 1993:
A period of low acetylene concentrations from 6/15/93 through 7/24/93. Appears to affect only this species, but data look bad for this time period.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Emma Woods	7/20/96	2:00	S	Sample	1	High isoprene, other species
Emma Woods	7/23/96	16:00	S	Sample	1	High 124tmb
Emma Woods	7/26/96	5:00	S	Sample	1	Abundant species missing or at LOD
Emma Woods	7/26/96	12:00	S	Sample	1	High isoprene, other species
Emma Woods	7/26/96	16:00	S	Sample	1	High 124tmb
Emma Woods	7/29/96	2:00, 5:00	S	Sample	2	High isoprene, other species
Emma Woods	8/1/96	2:00	S	Sample	1	High isoprene, other species
Emma Woods	8/4/96	2:00	S	Sample	1	High isoprene, other species
Emma Woods	8/7/96	5:00	S	Sample	1	High isoprene, other species
Emma Woods	8/10/96	2:00, 5:00	S	Sample	2	High isoprene, other species
Emma Woods	8/10/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	8/13/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	8/22/96	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
Emma Woods	8/22/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	8/25/96	2:00	S	Sample	1	Odd fingerprint, high concentrations
Emma Woods	8/25/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	8/28/96	2:00	S	Sample	1	Odd fingerprint, high concentrations
Emma Woods	8/28/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	8/31/96	2:00	S	Sample	1	Odd fingerprint, high concentrations
Emma Woods	9/3/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	9/6/96	16:00	S	Sample	1	Benzene > Toluene
Emma Woods	9/9/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	9/12/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	9/15/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	9/18/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	9/21/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	9/24/96	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	6/9/97	12:00	S	Sample	1	Toluene missing
Emma Woods	6/12/97	5:00	S	Sample	1	High isoprene, other species
Emma Woods	6/12/97	12:00	S	Sample	1	Benzene > Toluene
Emma Woods	6/15/97	12:00	S	Sample	1	Toluene missing
Emma Woods	6/18/97	5:00	S	Sample	1	Toluene missing
Emma Woods	6/21/97	12:00	S	Sample	1	Toluene missing
Emma Woods	7/9/97	5:00	S	Sample	1	Toluene missing
Emma Woods	7/15/97	2:00	S	Sample	1	Toluene missing
Emma Woods	7/21/97	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	7/24/97	12:00	S	Sample	1	UNID / TNMOC >70%
Emma Woods	7/30/97	16:00	S	Sample	1	Benzene > Toluene
Emma Woods	8/14/97	12, 16:00	S	Sample	2	Benzene > Toluene
Emma Woods	8/26/97	5:00	S	Sample	1	Toluene missing
Emma Woods	8/29/97	12:00	S	Sample	1	Toluene missing
Emma Woods	9/19/97	12, 16:00	S	Sample	2	Benzene > Toluene

60

Notes (265 samples):
1996- n-Decane was often higher than expected. Samples were flagged when n-Decane was reported as the most abundant species. The 2:00 and 5:00 samples were unusually high in the -C4 species, including isoprene. Only samples that were obviously unrealistic were flagged as suspect
June and July of 1997 show odd values of Butane and Pentane
Ethane and Propane show a distinct relationship from late August through October of 1996

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
El Rio	7/12/93	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	9/4/93	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	10/23/94	5:00	S	Sample	1	UNID <0
El Rio	10/29/94	5:00	S	Sample	1	UNID <0
El Rio	6/17/95	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	6/20/95	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	6/23/95	5, 12:00	S	Sample	2	High concentration n-Decane
El Rio	6/24/95	5, 12:00	S	Sample	2	High concentration n-Decane
El Rio	6/26/95	5, 12:00	S	Sample	2	High concentration n-Decane
El Rio	7/13/95	5, 12:00	S	Sample	2	High concentration n-Decane
El Rio	7/14/95	5:00	S	Sample	1	High concentration n-Decane
El Rio	7/15/95	12:00	S	Sample	1	High concentration n-Decane
El Rio	7/23/95	12:00	S	Sample	1	High concentration n-Decane
El Rio	7/26/95	8, 12, 16:00	S	Sample	3	High concentration n-Decane
El Rio	7/28/95	5, 8:00	S	Sample	2	ETAC, Ethane, and Acetylene missing or at LOD
El Rio	7/29/95	2:00	S	Sample	1	ETAC, Ethane, and Acetylene missing or at LOD
El Rio	8/7/95	2:00	S	Sample	1	ETAC, Ethane, and Acetylene missing or at LOD
El Rio	8/10/95	12, 16:00	S	Sample	2	High concentration Isobutane
El Rio	8/13/95	5, 12:00	S	Sample	2	High concentration Isobutane
El Rio	8/16/95	2, 5, 12, 16:00	S	Sample	4	High concentration n-Decane
El Rio	8/19/95	2, 12:00	S	Sample	2	High concentration n-Decane
El Rio	8/22/95	2, 12, 16:00	S	Sample	3	High concentration n-Decane
El Rio	8/28/95	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	8/28/95	6:00	S	Sample	1	High concentration acet
El Rio	9/9/95	2:00	S	Sample	1	ETAC, Ethane, and Acetylene missing or at LOD
El Rio	9/9/95	5, 12, 16:00	S	Sample	3	High concentration n-Decane
El Rio	9/12/95	5:00	S	Sample	1	High isoprene
El Rio	9/12/95	12:00	S	Sample	1	High concentration n-Decane
El Rio	9/12/95	16:00	S	Sample	1	Missing Isopentane
El Rio	9/19/95	12:00	S	Sample	1	ETAC, Ethane, and Acetylene missing or at LOD
El Rio	9/21/95	12:00	S	Sample	1	High concentration t2bte
El Rio	9/24/95	15:00	S	Sample	1	Only ethylene reported
El Rio	9/27/95	2:00	S	Sample	1	Missing Propane
El Rio	9/30/95	15:00	S	Sample	1	Only ethylene reported
El Rio	10/15/95	12:00	S	Sample	1	Abundant species missing or at LOD
El Rio	10/18/95	15:00	S	Sample	1	Only ethylene reported
El Rio	10/24/95	15:00	S	Sample	1	Only ethylene reported
El Rio	10/27/95	12:00	S	Sample	1	Abundant species missing or at LOD
El Rio	6/5/96	5:00	S	Sample	1	High isoprene
El Rio	6/8/96	5:00	S	Sample	1	High isoprene
El Rio	6/8/96	12:00	S	Sample	1	Missing Isopentane
El Rio	6/11/96	5:00	S	Sample	1	High isoprene
El Rio	6/14/96	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	6/17/96	5:00	S	Sample	1	High isoprene
El Rio	6/17/96	12:00	S	Sample	1	UNID / TNMOC >70%
El Rio	6/29/96	12:00	S	Sample	1	High concentration pdeben
El Rio	7/17/96	5:00	S	Sample	1	High isoprene
El Rio	7/20/96	2:00, 5:00	S	Sample	2	High isoprene
El Rio	7/23/96	5:00	S	Sample	1	High concentration n-Hexane
El Rio	7/29/96	5:00	S	Sample	1	High isoprene
El Rio	8/4/96	16:00	S	Sample	1	UNID / TNMOC >70%
El Rio	8/8/96	5:00	S	Sample	1	High isoprene
El Rio	8/13/96	5:00	S	Sample	1	High isoprene
El Rio	8/13/96	16:00	S	Sample	1	Missing Propane
El Rio	8/19/96	2:00	S	Sample	1	High isoprene

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
El Rio	8/22/96	2:00	S	Sample	1	High isoprene
El Rio	8/22/96	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	8/25/96	12, 16:00	S	Sample	2	UNID / TNMOC >70%
El Rio	8/28/96	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	8/28/96	12:00	S	Sample	1	Benzene > Toluene
El Rio	8/31/96	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
El Rio	9/6/96	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
El Rio	9/12/96	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
El Rio	9/30/96	16:00	S	Sample	1	Missing Toluene
El Rio	10/30/96	12:00	S	Sample	1	Benzene > Toluene
El Rio	6/3/97	12:00	S	Sample	1	Benzene > Toluene
El Rio	7/21/97	12:00	S	Sample	1	Benzene > Toluene
El Rio	7/30/97	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
El Rio	8/2/97	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
El Rio	8/6/97	2:00	S	Sample	1	Abundant species missing or at LOD
El Rio	8/7/97	16:00	S	Sample	1	Missing Toluene
El Rio	8/7/97	2:00	S	Sample	1	UNID / TNMOC >70%
El Rio	8/8/97	12, 16:00	S	Sample	2	UNID / TNMOC >70%
El Rio	8/14/97	2, 16:00	S	Sample	2	Benzene > Toluene
El Rio	8/20/97	2:00	S	Sample	1	Benzene > Toluene
El Rio	9/5/97	2:00	S	Sample	1	Benzene > Toluene
El Rio	9/7/97	12:00	S	Sample	1	Benzene > Toluene
El Rio	9/10/97	12:00	S	Sample	1	Benzene > Toluene
El Rio	9/19/97	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	9/28/97	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	9/29/97	2, 5:00	S	Sample	2	Odd fingerprint, high concentrations
El Rio	10/3/97	5:00	S	Sample	1	Odd fingerprint, high concentrations
El Rio	10/3/97	12:00	S	Sample	1	Benzene > Toluene
El Rio	10/25/97	5:00	S	Sample	1	Benzene > Toluene
					111	

Notes (765 samples):
1995- n-Decane was commonly higher than expected. Samples were flagged when n-Decane was reported as the most abundant species.
The 2:00 and 5:00 samples were unusually high in the -C4 species. Only samples that were obviously unrealistic were flagged as suspect
1995- Acetylene appears unusually high in this year.
In 1995, the carbonyl data appear to be reported at 0600 PST while the VOC data are reported at 0500 PST.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
El Cajon	7/11/95	12:00	S	sample	1	Toluene concentration less than Benzene.
El Cajon	8/31/95	2, 5:00	S	sample	2	Uncommonly high concentrations
El Cajon	9/12/95	2, 5:00	S	sample	2	Uncommonly high concentrations
El Cajon	10/3/95	2, 5:00	S	sample	2	Uncommonly high concentrations
El Cajon	10/24/95	2, 5:00	S	sample	2	Uncommonly high concentrations
El Cajon	7/17/96	5:00	S	lspbz	1	High concentration
El Cajon	7/15/97	5:00	S	sample	1	Uncommonly high concentrations
El Cajon	7/3/97	2, 5:00	S	sample	2	Uncommonly high concentrations
El Cajon	10/3/97-10/4/97	all	S	sample	8	Acetylene vs. benzene relationship was different compared to all other samples

Notes to data analysts: (461 samples)

1. Samples taken at 2:00 and 5:00 consistently show uncommonly high concentrations, with the 5:00 sample being higher in concentration. Samples with values that seemed realistic were not flagged.
2. Samples taken within July through August of 1997 show a different relationship between species than at other time periods.

Site	Date	Time	QC	Species or Sample	Flagged Samples	Comments
Overland	7/8/95	12, and 16:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	7/11/95	5:00	S	sample	1	High n-Decane
Overland	7/11/95	12:00	S	sample	1	o-Xylene > m/p-Xylene
	7/11/95	16:00	S	sample	1	High UNID / TNMOC
Overland	7/14/95	12:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	7/17/95	12, and 16:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	7/20/95	02, and 16:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	7/23/95	12, and 16:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	7/26/95	2, 12, 16:00	S	sample	3	o-Xylene > m/p-Xylene
Overland	7/29/95	02, 12:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	8/1/95	12:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	8/10/95	2:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	8/13/95	12, 16:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	8/16/95	2:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	8/16/95	5:00	S	sample	1	Possible misid, 2mpna with 2m1pe
Overland	8/19/95	12, 16:00	S	sample	2	o-Xylene > m/p-Xylene
Overland	8/22/95	2:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	10/3/95	5:00	S	sample	1	Uncommonly High Concentrations
Overland	10/9/95	2:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	10/12/95	2:00	S	sample	1	o-Xylene > m/p-Xylene
Overland	10/15/95	5:00	S	sample	1	mdeben high, Ethane at LOD or missing
Overland	10/15/95	16:00	S	m/p-Xylene	1	High Concentration
Overland	10/24/95	5:00	S	sample	1	Uncommonly High Concentrations
Overland	10/27/95	5:00	S	sample	1	Uncommonly High Concentrations
Overland	7/5/96	2, 12:00	S	sample	2	Ethane at LOD or missing
Overland	7/5/96	16:00	S	sample	1	High UNID / TNMOC
Overland	7/17/96	2:00	S	sample	1	Odd fingerprint, many species at LOD
Overland	7/17/96	5:00	S	sample	1	Acetylene at LOD or missing
Overland	8/25/96	5:00	S	sample	1	Odd fingerprint, many species at LOD
Overland	7/9/97	2:00	S	sample	1	Propane at LOD or missing
Overland	7/9/97	16:00	S	Isopentane	1	High Concentration
Overland	7/15/97	5:00	S	Isopentane	1	High Concentration
Overland	7/24/97	2:00	S	sample	1	Benzene > Toluene
Overland	7/27/97	2, 12, 16:00	S	sample	3	Benzene > Toluene
Overland	8/2/97	16:00	S	sample	1	Benzene > Toluene
Overland	8/17/97	16:00	S	sample	1	High UNID / TNMOC
Overland	10/3/97-10/4/97	all	S	sample	8	Acetylene vs. benzene relationship was different compared to all other samples

Notes to data analysts: (289 samples)

1. Samples at 5:00 consistently show uncommonly high concentrations.
2. Samples taken within July through August of 1997 show a different relationship between species than at other time periods.

APPENDIX B

HYDROCARBON ABBREVIATIONS USED IN THIS REPORT

Table B-1. AIRS code, abbreviation, hydrocarbon name, and species group (O=olefin, P=paraffin, A=aromatic).

AIRS code	Abbreviation	Hydrocarbon	Species Group
43295	3ethex	3-ethylhexane	P
43955	25mhex	2,5-dimethylhexane	P
43293	hex24m	2,4-dimethylhexane	P
43294	hex23m	2,3-dimethylhexane	P
43222	propa22m	2,2-dimethylpropane	P
43347	2m&c	2mhexa&cyhexa	P
43270	ibute	Isobutene	O
43239	c3m2pne	Cis-3-Methyl-2-pentene	O
43145	1octe	1-Octene	O
43240	mcpne	Methylcyclopentene	O
43264	hpte	Heptene	O
43269	cyhxe	Cyclohexene	P
43277	225tmhxa	2,2,5-Trimethylhexane	P
43278	c2octe	Cis-2-Octene	O
43279	1none	1-Nonene	O
43395	4mhpte	4-Methylheptane	P
45216	sbutb	Sec-Butylbenzene	A
43144	prpyn	Propyne	O
43210	1butyn	1-Butyne	O
43000	pamshc	Sum PAMS Target Species	

Table B-1. AIRS code, abbreviation, hydrocarbon name, and species group (O=olefin, P=paraffin, A=aromatic).

AIRS code	Abbreviation	Hydrocarbon	Species Group
43206	acety	Acetylene	O
43203	ethyl	Ethylene	O
43202	ethan	Ethane	P
43205	prpyl	Propylene	O
43204	propa	Propane	P
43214	isbta	Isobutane	P
43280	1bute	1-Butene	O
43212	nbuta	n-Butane	P
43216	t2bte	trans-2-Butene	O
43217	c2bte	cis-2-Butene	O
43282	3mlbe	3-Methyl-1-Butene	O
43221	ispna	Isopentane	P
43224	1pnte	1-Pentene	O
43220	npnta	n-Pentane	P
43243	ispre	Isoprene	O
43226	t2pne	trans-2-Pentene	O
43227	c2pne	cis-2-Pentene	O
43228	2m2be	2-Methyl-2-Butene	O
43244	22dmb	2,2-Dimethylbutane	P
43283	cypne	Cyclopentene	O
43234	4mlpe	4-Methyl-1-Pentene	O
43242	cypna	Cyclopentane	P
43284	23dmb	2,3-Dimethylbutane	P
43285	2mpna	2-Methylpentane	P
43230	3mpna	3-Methylpentane	P
43246	2m1pe	2-Methyl-1-Pentene	O
43231	nhexa	n-Hexane	P
43289	t2hex	trans-2-Hexene	O
43290	c2hex	cis-2-Hexene	O
43262	mcpna	Methylcyclopentane	P
43247	24dmp	2,4-Dimethylpentane	P
45201	benz	Benzene	A
43248	cyhxa	Cyclohexane	P
43263	2mhxa	2-Methylhexane	P
43291	23dmp	2,3-Dimethylpentane	P
43249	3mhxa	3-Methylhexane	P
43250	224tmp	2,2,4-Trimethylpentane	P

Table B-1. AIRS code, abbreviation, hydrocarbon name, and species group (O=olefin, P=paraffin, A=aromatic).

AIRS code	Abbreviation	Hydrocarbon	Species Group
43232	nhept	n-Heptane	P
43261	mcyhx	Methylcyclohexane	P
43252	234tmp	2,3,4-Trimethylpentane	P
45202	tolu	Toluene	A
43960	2mhep	2-Methylheptane	P
43253	3mhep	3-Methylheptane	P
43233	noct	n-Octane	P
45203	ebenz	Ethylbenzene	A
45109	m/pxy	m/p-Xylene	A
45206	pxyl	p-Xylene	A
45220	styr	Styrene	A
45204	oxyl	o-Xylene	A
43235	nnon	n-Nonane	P
45210	ispbz	Isopropylbenzene	A
45209	npbz	n-Propylbenzene	A
43256	apine	alpha-Pinene	O
45207	135tmb	1,3,5-Trimethylbenzene	A
45208	124tmb	1,2,4-Trimethylbenzene	A
43257	bpine	beta-Pinene	O
45211	oetol	o-Ethyltoluene	A
45212	metol	m-Ethyltoluene	A
45213	petol	p-Ethyltoluene	A
45218	mdeben	m-diethylbenzene	A
45219	pdeben	p-diethylbenzene	A
45225	123tmb	1,2,3-trimethylbenzene	A
43238	ndec	n-Decane	P
43954	nundc	n-Undecane	P
43102	tnmoc	Total Non-Methane Organic Compounds	
43391	43mpne	3/4 Methylpentene	
43599	ETAC	Ethane & Acetylene	
43392	223tmb	2,2,3 Trimethylbutane	
43393	c13dmp	Cis-1,3-Dimethylcyclopentane	
43394	t13dmp	Trans-1, 3-Dimethylcyclopentane	
43218	13buta	1,3-butadiene	O
43225	2m1bte	2-methyl-1-butene	O
43245	1hexe	1-hexene	O
43275	p2e244	224-trimethyl-2-pentene	O

APPENDIX C

SUMMARY OF COEFFICIENT OF VARIATION SCREENING OF 1997 DATA BY SITE

To investigate the variation in the composition from day to day, we tabulated the hydrocarbons that had COV values (based on the mean and standard deviations in the 1997 weight fractions) greater than 50 percent at each site. An analysis of the COV helps us to visualize the variability in the hydrocarbon weight fraction.

Table C-1. Summary of hydrocarbons in 1997 that had a coefficient of variation greater than 50 percent (for Ventura County, greater than 100 percent).

Hydrocarbon	Total	Elk Grove	Emma Wood	Arvin	Del Paso Manor	Bakers-field	Clovis	Fresno	San Diego	Los Angeles	El Rio	Simi Valley	Folsom	Parlier
Site type		1	1	"1/3	2	2	2	Urban	Urban	Urban	2	3	3	3
Acetylene	3		x			x			x					
Ethene	1		x											
Ethane	2					x	x							
Propene	5	x	x	x							x	x		
Propane	4	x			x	x							x	
1-Butene	1		x											
i-Butene	2								x	x				
i-Butane	7	x		x		x	x	x					x	x
n-Butane	3					x	x		x					
i-Pentane	1			x										
t-2-butene	2		x									x		
c-2-Butene	3		x								x	x		
1-Pentene	7	x	x			x	x				x	x		x
n-Pentane	2	x											x	
Isoprene	9		x		x	x	x	x		x	x	x	x	
t-2-Pentene	3		x					x				x		
c-2-Pentene	3		x								x	x		
2,2,-Dimethylbutane	1											x		
2-Methyl-2-butene	1						x							
Cyclopentene	3					x							x	x
Cyclopentane	4								x		x	x		x
2-Methyl-1-pentene	3		x								x	x		
n-Hexane	2	x							x					
2,4-Dimethylpentane	1	x												
Benzene	2							x	x					
2,3-Dimethylpentane	1	x												
2,2,4-Trimethylpentane	1	x												
2-Methylhexane	1			x										
n-Heptane	1			x										

Table C-1. Summary of hydrocarbons in 1997 that had a coefficient of variation greater than 50 percent (for Ventura County, greater than 100 percent).

Hydrocarbon	Total	Elk Grove	Emma Wood	Arvin	Del Paso Manor	Bakers-field	Clovis	Fresno	San Diego	Los Angeles	El Rio	Simi Valley	Folsom	Parlier
Site type		1	1	"1/3	2	2	2	Urban	Urban	Urban	2	3	3	3
3-Methylheptane	1											x		
2-Methylheptane	1											x		
2,3,4-Trimethylpentane	3	x	x											x
Toluene	3	x					x	x						
n-Octane	5					x		x	x			x		x
Ethylbenzene	1	x												
m-&-p-Xylenes	2	x					x							
o-Xylene	3	x											x	x
Styrene	5		x			x		x	x			x		
n-Nonane	6						x	x	x		x	x	x	
i-Propylbenzene	3		x								x	x		
n-Propylbenzene	1											x		
135TMB	3	x	x									x		
124TMB	5	x		x			x	x					x	
123TMB	5			x		x		x				x		x
o-Ethyltoluene	3		x		x						x			
m-Ethyltoluene	2										x		x	
p-Ethyltoluene	2		x								x			
m-Diethylbenzene	4		x					x			x	x		
p-Diethylbenzene	5		x			x				x	x	x		
n-Decane	8			x	x	x	x	x	x			x		x
Undecane	10	x		x	x	x	x	x		x		x	x	x
Total		17	19	9	5	14	12	13	10	4	14	23	10	10

APPENDIX D

LINEAR REGRESSION RESULTS FOR CLOVIS 1995-1997 CONTINUOUS NMHC AND PAMS NMHC CONCENTRATIONS

D.1 ALL DATA

Linear regression of 3-hr averaged continuous NMHC data and PAMS NMHC data. All sampling periods are included in this analysis.

247 case(s) deleted due to missing data.

Dep Var: TNMOC N: 336 Multiple R: 0.793 Squared multiple R: 0.629

Adjusted squared multiple R: 0.628 Standard error of estimate: 77.642

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	31.024	6.947	0.000	.	4.466	0.000
AVGOFNMHCPP	0.477	0.020	0.793	1.000	23.801	0.000

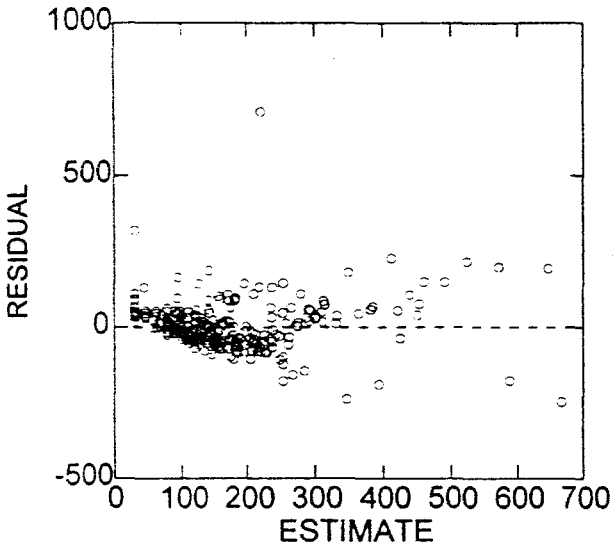
Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	3414824.752	1	3414824.752	566.468	0.000
Residual	2013442.570	334	6028.271		

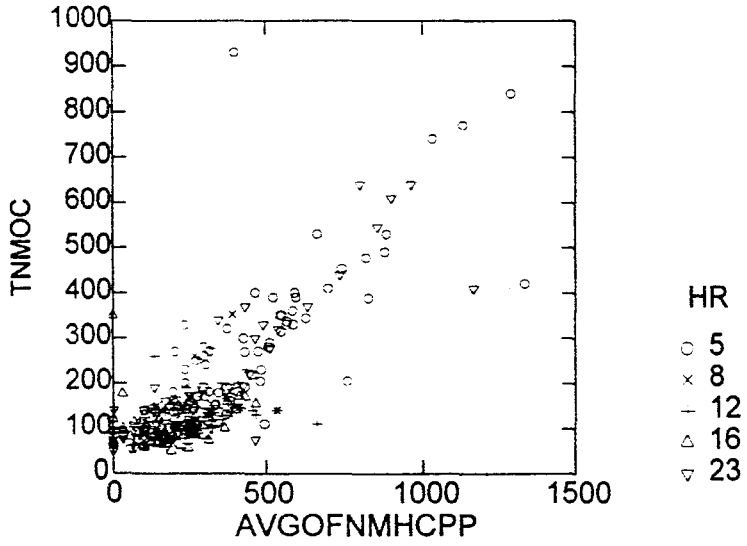
*** WARNING ***
Case 211 is an outlier (Studentized Residual = 4.228)
Case 234 is an outlier (Studentized Residual = 10.534)
Case 293 has large leverage (Leverage = 0.056)
Case 294 has large leverage (Leverage = 0.078)
Case 314 has large leverage (Leverage = 0.052)
Case 457 has large leverage (Leverage = 0.072)

Durbin-Watson D Statistic 1.383
First Order Autocorrelation 0.308

Plot of Residuals against Predicted Values



Scatter plot of 3-hr averaged continuous NMHC data and PAMS NMHC data. All sampling periods are included in this analysis.



D.2 0500-0700 PST DATA

Linear regression of 3-hr averaged continuous NMHC data and PAMS NMHC data. Only data from the 0500-0700 sampling period are included in this analysis.

83 case(s) deleted due to missing data.

Dep Var: TNMOC N: 84 Multiple R: 0.783 Squared multiple R: 0.613

Adjusted squared multiple R: 0.609 Standard error of estimate: 108.043

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	51.245	21.663	0.000	.	2.365	0.020
AVGOFNMHCPP	0.503	0.044	0.783	1.000	11.403	0.000

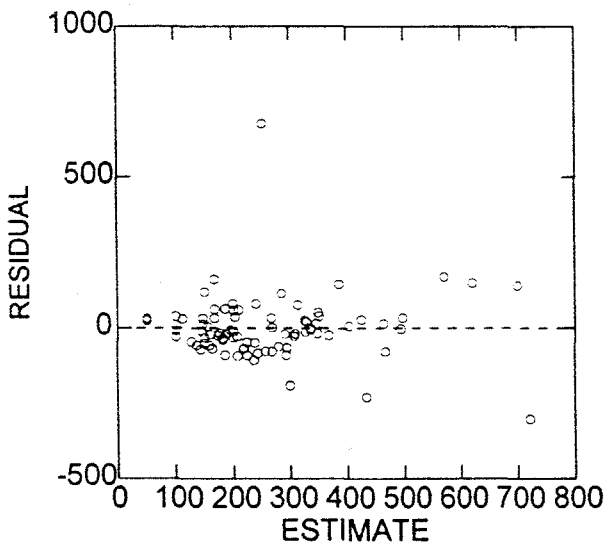
Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	1517730.495	1	1517730.495	130.017	0.000
Residual	957210.493	82	11673.299		

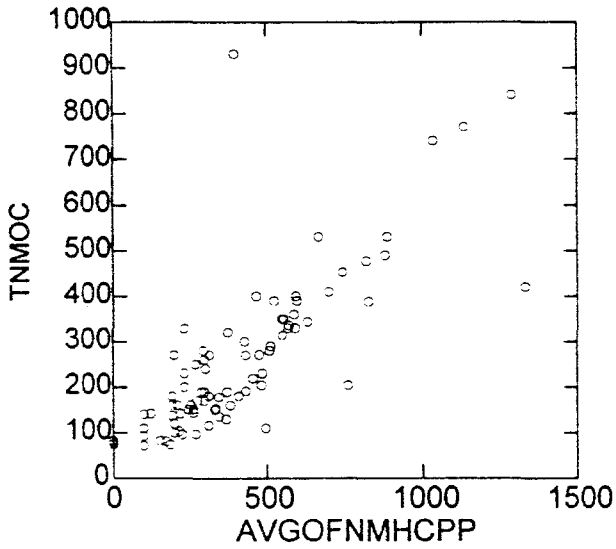
*** WARNING ***
Case 234 is an outlier (Studentized Residual = 8.741)
Case 294 has large leverage (Leverage = 0.153)

Durbin-Watson D Statistic 1.610
First Order Autocorrelation 0.194

Plot of Residuals against Predicted Values



Scatter plot



D.3 1200-1400 PST DATA

Linear regression of 3-hr averaged continuous NMHC data and PAMS NMHC data. Only data from the 1200-1400 sampling period are included in this analysis.

81 case(s) deleted due to missing data.

Dep Var: TNMOC N: 82 Multiple R: 0.465 Squared multiple R: 0.217

Adjusted squared multiple R: 0.207 Standard error of estimate: 37.479

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	82.450	8.841	0.000	.	9.326	0.000
AVGOFNMHCPP	0.159	0.034	0.465	1.000	4.702	0.000

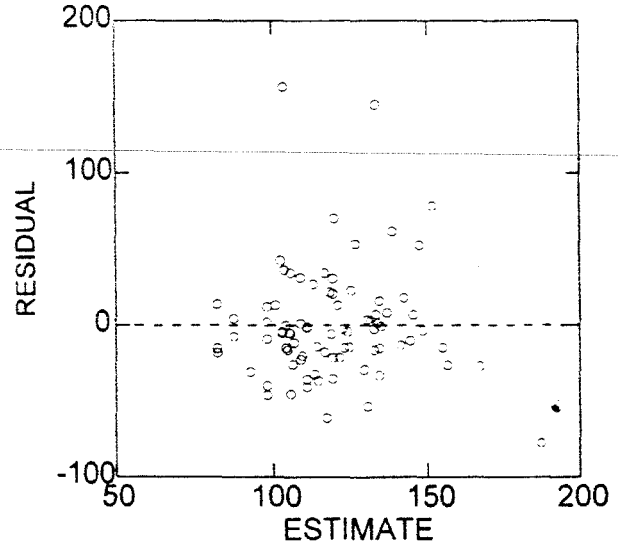
Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	31058.189	1	31058.189	22.111	0.000
Residual	112374.067	80	1404.676		

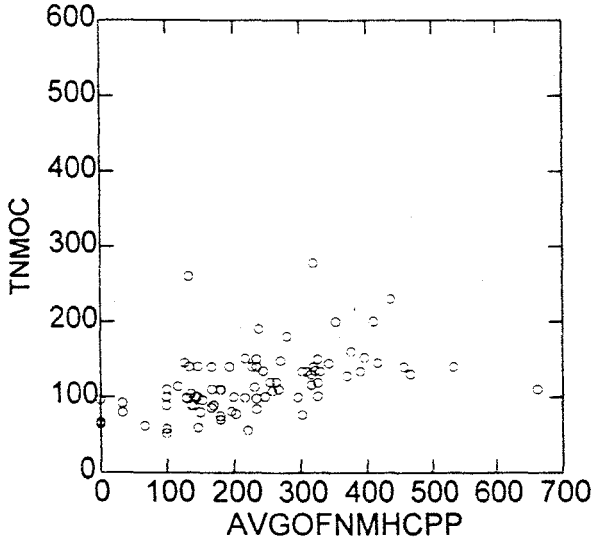
*** WARNING ***
Case 226 is an outlier (Studentized Residual = 4.749)
Case 416 has large leverage (Leverage = 0.163)
Case 558 is an outlier (Studentized Residual = 4.305)

Durbin-Watson D Statistic 2.061
First Order Autocorrelation -0.032

Plot of Residuals against Predicted Values



Scatter plot



D.4 1600-1800 PST DATA

Linear regression of 3-hr averaged continuous NMHC data and PAMS NMHC data. Only data from the 1600-1800 sampling period are included in this analysis.

40 case(s) deleted due to missing data.

Dep Var: TNMOC N: 85 Multiple R: 0.169 Squared multiple R: 0.029

Adjusted squared multiple R: 0.017 Standard error of estimate: 39.055

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	91.535	8.395	0.000	.	10.904	0.000
AVGOFNMHCPP	0.063	0.041	0.169	1.000	1.561	0.122

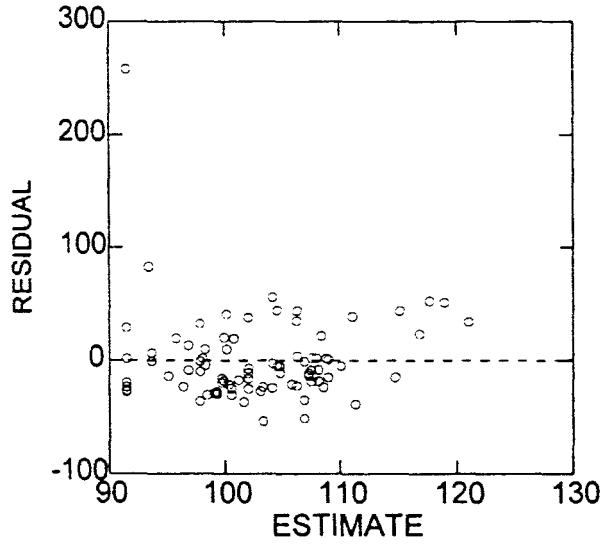
Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	3715.806	1	3715.806	2.436	0.122
Residual	126601.205	83	1525.316		

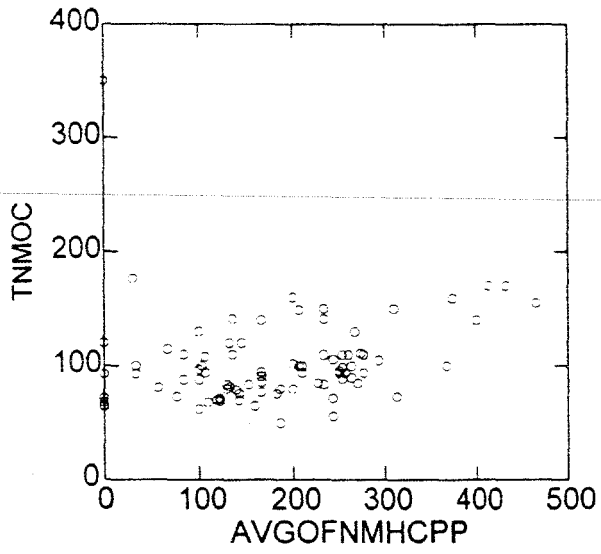
*** WARNING ***
Case 211 is an outlier (Studentized Residual = 10.077)

Durbin-Watson D Statistic 2.067
First Order Autocorrelation -0.045

Plot of Residuals against Predicted Values



Scatter plot



D.5 2300-0200 PST DATA

Linear regression of 3-hr averaged continuous NMHC data and PAMS NMHC data. Only data from the 2300-0200 sampling period are included in this analysis.

40 case(s) deleted due to missing data.

Dep Var: TNMOC N: 73 Multiple R: 0.881 Squared multiple R: 0.775

Adjusted squared multiple R: 0.772 Standard error of estimate: 66.375

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	25.336	11.924	0.000	.	2.125	0.037
AVGOFNMHCPP	0.516	0.033	0.881	1.000	15.658	0.000

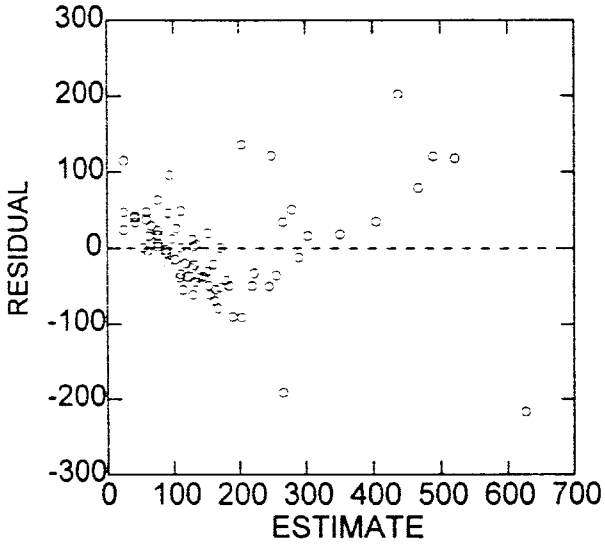
Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	1080220.523	1	1080220.523	245.187	0.000
Residual	312804.463	71	4405.697		

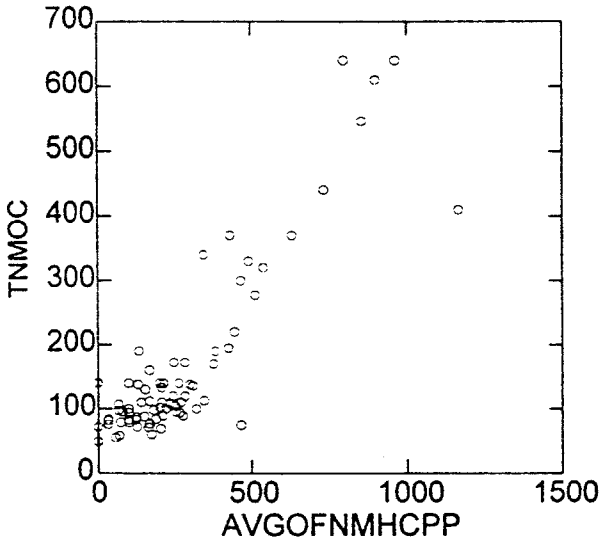
*** WARNING ***
Case 281 is an outlier (Studentized Residual = 3.406)
Case 293 has large leverage (Leverage = 0.210)
Case 293 is an outlier (Studentized Residual = -4.058)

Durbin-Watson D Statistic 1.838
First Order Autocorrelation 0.081

Plot of Residuals against Predicted Values



Scatter plot



APPENDIX E

AMBIENT DATA FILES FOR CMB8.0

Table of species names and abbreviations used in the ambient data CMB8.0 input files.

Table E-1. Naming convention used in CMB input files.

Species Abbreviation	Species Uncertainty	Species Full Name
TNMOC	TNMOCU	total non-methane hydorcarbons
ETHANE	ETHANEU	ethane
ETHENE	ETHENEU	Ethylene
ACETYL	ACETYLU	acetylene
N PROP	N PROPU	n-propane
PROPE	PROPEU	propylene
I BUTA	I BUTAU	isobutane
N BUTA	N BUTAU	n-butane
T2BUTE	T2BUTEU	trans-2-butane
C2BUTE	C2BUTEU	cis-2-butane
B1E3ME	B1E3MEU	3-methyl-1-butene
IPENTA	IPENTAU	isopentane
PENTE1	PENTE1U	1-pentene
N PENT	N PENTU	n-pentane
I PREN	I PRENU	isoprene
T2PENE	T2PENEU	trans-2-pentene
C2PENE	C2PENEU	cis-2-pentene
B2E2M	B2E2MU	2-methyl-2-butene
BU22DM	BU22DMU	2,2-dimethylbutane
CPENTE	CPENTEU	cyclopentene
P1E4ME	P1E4MEU	4-methyl-1-pentene
CPENTA	CPENTAU	cyclopentane
BU23DM	BU23DMU	2,3-dimethylbutane
PENA2M	PENA2MU	2-methylpentane
PENA3M	PENA3MU	3-methylpentane
P1E2ME	P1E2MEU	2-methyl-1-pentene
N HEX	N HEXU	n-hexane
T2HEXE	T2HEXEU	trans-2-hexane
C2HEXE	C2HEXEU	cis-2-hexane
MCYPNA	MCYPNAU	methylcyclopentane
PEN24M	PEN24MU	2,4-dimethylpentane
BENZE	BENZEU	benzene
CYHEXA	CYHEXAU	cyclohexane
HEXA2M	HEXA2MU	2-methylhexane
PEN23M	PEN23MU	2,3-dimethylpentane
HEXA3M	HEXA3MU	3-methylhexane
PA224M	PA224MU	2,2,4-trimethylpentane
N HEPT	N HEPTU	n-heptane
MECYHX	MECYHXU	methylcyclohexane
PA234M	PA234MU	2,2,3-trimethylpentane
TOLUE	TOLUEU	toluene
HEP2ME	HEP2MEU	2-methylheptane
HEP3ME	HEP3MEU	3-methylheptane
N OCT	N OCTU	n-octane

Table E-1. Naming convention used in CMB input files.

Species Abbreviation	Species Uncertainty	Species Full Name
ETBZ	ETBZU	ethylbenzene
MP XYL	MP XYLU	m/p-Xylene
STYR	STYRU	Styrene
O XYL	O XYLU	o-Xylene
N NON	N NONU	n-nonane
IPRBZ	IPRBZU	isopropylbenzene
N PRBZ	N PRBZU	n-propylbenzene
BZ135M	BZ135MU	1,3,5-trimethylbenzene
BZ124M	BZ124MU	1,2,4-trimethylbenzene
UNID	UNIDU	unidentified