FINAL REPORT

Source Attribution and Emission Adjustment Study

Task 1: Back-Trajectory Climatology and Identification of Key Source Regions

(Contract No. 582-4-65590, Work for Others Agreement No.46652)

YuLong Xie
Applied Geology & Geochemistry Group

Carl Berkowitz
Atmospheric Chemistry and Meteorology Group

Battelle Memorial Institute, Pacific Northwest Division
Richland, WA 99352

January 10, 2005

CONTENTS

EXECUTIVE SUMMARY v

INTRODUCTION 1

DESCRIPTIONS OF STATISTICS AND DATA 1

Data Available 1
Basic Statistics 2
Diurnal Patterns 4
Scatter Plot and Correlation 5
Time Series and Outlier Detection 5
Backward Wind Trajectories of Seasonal Maxima 6
Categorized Species Percent Contribution 6

WIND DIRECTION AND WIND TRAJECTROY ANALYSES 7

Wind Direction Analysis with Conditional Probability Function (CPF) 7
Olefins: An Overview of Both Day and Nighttime CPF Plots 10
Alkanes: An Overview of Both Day and Nighttime CPF Plots 11
Clustering Analysis on the CPF Plots 14
Potential Source Contribution Function (PSCF) Analysis 17

POSITIVE MATRIX FACTORIZATION (PMF) ANALYSIS 21

Multivariate Receptor Modeling with Positive Matrix Factorization 23
Some Technical Details on the Applications of the PMF 26
Results of PMF Analysis and Discussion 26

REFERENCES 38

TABLES 43

FIGURES 63

APPENDIX A – OVERVIEW STATISTICS, BY SITE AND SPECIES

APPENDIX B – EXTREME VALUES, BY SITE AND SPECIES

APPENDIX C – CPF PLOTS FOR SELECTED SPECIES MEASURED IN THE VICINITY OF THE SHIP CHANNEL

APPENDIX D – CLUSTERING OF CPF PLOTS FOR THE SHIP CHANNEL SITES

EXECUTIVE SUMMARY

- Mean individual and total VOC concentrations at *nighttime* were found to be higher than concentrations measured during the *daytime*. This likely reflects the fact that the VOC emissions are fresher and that there are smaller change in the mixing ratios at *night* than during the *day* due to the absence of photochemical reactions at *night*.

- Mean concentration of the VOCs at the five auto-GC sites near the Ship Channel were found to be generally higher than those measured at the other sites. This is consistent with the proximity of these sites to the Ship Channel.

- The most abundant VOCs are *ethane, propane, butanes, pentanes* followed by *ethene, propylene, benzene, toluene, n-hexane* and *cycloHexane*. The great abundance of the alkanes may be attributed to both high emission rates and their accumulation as a result of having longer photochemical lifetime associated with relatively low reactivities. The abundance of the highly reactive low lifetime alkenes and aromatics is consistent with the high emission rates from local sources. Very high concentrations of *benzene, toluene* and *styrene* were observed at Lynchburg Ferry site relative to the other auto-GC locations, and the maximum concentration for all species, at any of the sites was 7897.5 ppbC for *styrene* (c45220) at Lynchburg Ferry.

- Discrepancies were found between the reported quantities of *TNMHC* / *TNMTC* and values of *TNMHC* / *TNMTC* derived by summing the measured concentrations of the individual VOC species.

- All individual species except *isoprene* showed similar diurnal patterns. The average concentrations have minima around noon and gradually increase to the maxima in the early morning hours (5-8am), then decrease to their noon time minima. This pattern can be explained by the accumulation of fresh emission and the absence of photochemical reactions at night, both leading to enhanced concentrations after sunset and before sunrise. During the day there is a depletion of the VOCs through photochemical destruction and a dilution effect from the increase in the mixing height. The reported total VOC concentrations had patterns similar to the majority of the VOC species.

- *Isoprene* showed a much different diurnal pattern relative to the other species. The concentrations of *isoprene* were lowest during night and started to increase from midnight, reaching maximum in the morning. The broad peak remained for most of the day and then decreased to a minimum around mid-night. This diurnal behavior is consistent with a biogenic source produced at a rate governed by, among other things, photosynthetic activity. The daytime maxima are not obvious at *Lake Jackson* and *Danciger* sites.

- Many isomers were found to have high correlation coefficients suggesting common sources for them.

- At a given site, extreme values in concentration often occur for multiple species at the same time while for a given species there are few patterns suggesting simultaneous or temporally linked patterns across multiple sites. This suggest that extreme events are likely *local* events that do not affect multiple sites simultaneously.

- We found no apparent pattern in the wind trajectories for the extreme events. These events were found to be associated with both backward wind trajectories characterized by straight air flow from distant and a more sinuous air flow that allowed for recirculation in the proximity of the sites. The independence of the extremes in the trajectories patterns suggests that the sources are very *local*.

The following conclusions come from our analysis of wind direction used in the conditional probability functions (CPF) and from the wind trajectory analysis used with the potential source contribution function (PSCF):

1. Wind roses could not explain all of the CPF patterns (which show directions associated with elevated concentrations of specific VOCs). This suggests that local meteorology cannot explain many of the features identified in the CPF analysis.

2. Most lobes in the CPF plot pass through known point source clusters in the inventory but it is hard to pinpoint which specific point source is a main contributor because there are usually many point sources along the axis of the major CPF lobes.

3. Differences between the source direction suggested by the CPF lobes and the existing inventories were observed. For example the *Wallisville* site frequently had elevated values coming from the northeast although no point sources are shown here, and the point sources north of Channel View appear to play a relatively small role in contributing to elevated values observed at this site.

    a. The visual inspection and the clustering analysis of the CPF patterns of the species showed that

        i. the CPF patterns of alkanes are usually quite similar. Chemical related species (*c2/t2-butene* and *c2/t2-pentene*) frequently display similarities in the CPF patterns. Other *butenes* (*1,3butadiene*, *1-butene*) or *pentenes* (1-*pentene*) show similar CPF patterns to these pairs at some sites;
        ii. the alkane isomer groups such as (a) *2,3,4-trimethylPentane* and *2,2,4-trimethylPentane,* (b) *2-methylPentane* and *3-methylPentane,* (c) *2,3-dimethylPentane* and *2,4-dimethylPentane*, (d) *2,2-dimethylButane* and *2,3-dimethylButane,* (e) *2-methyl-Heptane* and *3-methyl-Heptane,* (f) *2-methyl-Hexane* and *3-methyl-Hexane*, often show quite similar CPF patterns;

iii. to some extent, *ethane* and *propane* share similar CPF patterns and in some cases the similarities can be extended to the inclusion of *n-butane* and *i-butane;*
iv. the long-C alkanes of *n-octoane, n-nonane, n-decane* and *n-undecane* are close in the clustering tree suggesting the similarity in the CPF patterns;
v. Often the aromatic compounds form a big group in the clustering analysis of the CPF patterns suggesting the general similarities in their CPF patterns. *Styrene* and *ethylbenzene*, and *m/p-xylene* and *o-xylene* tend to be the two pairs having the most similar CPF patterns among the aromatics compounds. *Toluene* and *benzene* are often display different CPF patterns. The similarities of the species in the CPF patterns may suggest the common sources or transport pathways.

b. The PSCF maps based on data from an individual site usually only reveal the trajectories of the VOC species towards the site. The PSCF maps based on data from multiple sites was able to isolate possible source areas. By inspecting the PSCF maps based on data from both individual sites and multiple sites together it is possible to discern the potential source regions.

Based on the source identification and source appointment with the PMF analysis, the following conclusions can be drawn:

- Common emission categories, or 'factors', were resolved for the five sites close to Ship Channel.

- Very consistently one petrochemical emission factor dominated by *ethene* and *propylene* was found for all five sites. *1-butene*, *1,3-butadiene* and sometime other C4 olefins were included in this factor. It seems that the large refinery, petrochemical and industrial facilities along Ship Channel are responsible for the emissions in this category. The contribution of this category to the total mean VOC concentration measured in the sites ranged from 8.5% at *Haden Road* site to about 20% at *Channel View*, *Clinton* and *Lynchburg Ferry* sites to 31.52% at *WallisVille* site. Overall, this category of emission constitutes the second largest contribution to the ambient VOC at these sites.

- Another very consistent emission category was dominated by *ethane, propane, butanes* and *pentanes* and was identified as coming from sources associated with natural gas and evaporation. The mass contribution of this category ranged from about 20% at *Clinton*, *Haden Road* and *Lynchburg Ferry* to 36% at *WallisVille* to 50% at Channel View site, demonstrating being the largest contributor to the VOC measured in these sites.

- The other major category of sources was characterized by the abundance of mid-weight alkanes and aromatics associated with emissions from the production of solvents and paint. The predominant direction for these emissions was from the

south. The mass contributions are generally less than 10% except at *Haden Road* (23.29%).

- Vehicle exhaust was another common factor identified in the Ship Channel with the exception of *Lynchburg Ferry*. At the *Lynchburg Ferry* site, most of *acetylene* showed up in the accumulation / natural gas and evaporative source category, characterized with the low C alkanes. The percent contribution to the total mean VOC measured ranges from 1.22% at *Haden Road* site to a little bit less than or about 10% at other sites.

- Another common factor found was characterized by *n-butane, n-pentane* and *iso-pentane*. This factor was enriched with C4/C5 especially C5 *pentenes*. Since *pentane*s are markers of gasoline vapors this factor is thought to be associated with fuel evaporation. The contributions of this category to the total mean VOC measured at the sites range from 2% at *WallisVille* site to 19% at *Haden Road* site.

- Some sources were quite unique to specific auto-GC sites. For example the factor enriched with *styrene* and *ethylbenzene* at *Channel View* site (Factor 6) is unique with no similar factor found at the other sites. The *1,3-butadiene* enriched factor (Factor 5), the factor with abundant of *pentanes* and branched *pentanes* (Factor 6) and a factor consisting of only enriched aromatic compounds (Factor 8) were unique at the *Haden Road* site. The biogenic category characterized with enriched *isoprene* was unique at *WallisVille*. At another sites *isoprene* was mixed into the other source categories and with no biogenic source uniquely identified. Although the mass contribution of the biogenic source identified at *WallisVille* site is about 4% and no separate biogenic sources were identified for the other four sites, the contributions of the biogenic sources should not be zero (for other sites) and should be greater than 4% (for *WallisVille* site) considering the biogenic diurnal behavior of *isoprene*.

INTRODUCTION

Ozone pollution is a major problem during the summer months in many U.S. urban areas [TNRCC, 2001]. Ozone is not emitted directly into the air, but instead is formed in sunlight through a series of complex atmospheric photochemical reactions primarily involving nitrogen oxides and volatile organic compound (VOC) emissions. NOx (= NO + NO_2) is produced almost entirely as a by-product of high-temperature combustion and can be continuously measured with good accuracy and precision at source and receptor sites [Atkinson, 2000; Watson et al., 2001]. VOCs include hundreds of organic chemicals originating from many different anthropogenic and natural sources including vapors associated with gasoline and solvents, and emission from trees and plants. These hundreds of VOC species play a major role in the formation of ozone.

As a result, accurate inventories of VOC emissions are crucial to air quality modeling and the development of regulatory control strategies. Data from specific sources or continuous emissions measurements are desirable but not commonly available. Even if they are available they represent a snapshot in time and may not reflect the variability of actual emission over time.

The US Environmental Protection Agency (EPA) initiated the Photochemical Assessment Monitoring Stations (PAMS) networks in major non-attainment urban areas to monitor speciated hydrocarbons, oxides of nitrogen, ozone and surface meteorological data. These observations are being collected to better understand the trends of both precursors and ozone. As a part of the PAMS monitoring efforts the Texas Commission on Environmental Quality (TCEQ) collected hourly speciated VOC data during 2003 using automated gas chromatographs ("auto-GCs") at the nine sites shown in **Figure 1**. The goal of our work reported here was to statistically analyze observations from the network, estimate emission profiles and identify air mass trajectories of the VOCs to the auto-GC sites. A state of art multivariate receptor modeling technique including positive matrix factorization (PMF) [Paatero & Tapper, 1993; Hopke, 2003] was used to apportion the measured VOCs into separate source factor. The number of sources, composition profiles and contribution profiles of the sources were estimated and the results compared with emission profiles based on TCEQ inventories. Meteorological data (wind speed, wind direction) was used in the statistical analyses of so-called conditional probability function (CPF) and potential source contribution function (PSCF) to aid the source apportionment.

DESCRIPTIONS OF STATISTICS AND DATA

Data Available

The auto-GC data measurements came from 9 sites in the Houston/Galveston Area (**Figure 1**) between June 1st and October 31st of 2003. The sampling frequency for each site was 1 hour. Other details of the 9 sites are listed in **Table 1**. All observations were provided by TCEQ, with 68 data fields associated with the VOC measurements that

included methane and a set of 64 individual non-methane hydrocarbons. Also included were two summary measurements: *TNMHC* (total non-methane hydrocarbon) and *TNMTC* (total non-methane PAMS targeted compounds). The non PAMS species were consistently missing; only the 55 PAMS species had reported concentrations. The names and codes of the measured 55 PAMS targeted hydrocarbons are listed in **Table 2**. Local observations of wind speed and wind direction were also available at the auto-GC sites. These meteorological observations were collected every 5 minutes and used for the Conditional Probability Function (CPF) analysis of wind direction and to calculate a set of back wind trajectories for the Potential Source Contribution Function (PSCF) analysis.

Basic Statistics

Much of our analysis partitions the observations into those taken during the day (0700 to 2000 local time) and night (2100 to 0600 local time). We have done this to somewhat remove the role of photochemistry. A comprehensive summary of all the statistics of all species at all sites is presented in **Appendix A (*Appendix_A_basicStat.pdf*)**, and an abbreviated version presented in **Table 3**, the later showing only the maximum, median and mean values of the VOC concentrations. Shown in the very bottom two rows of **Table 3** are the calculated *TNMTC* and *TNMHC* from the individual VOC species. Individual VOC species have been highlighted with colors based on their mean concentrations. Species with mean concentration higher than 10 ppbC are colored pink. Species with mean concentrations between 5 and 10 ppbC are colored brown. Species with mean concentrations between 1 and 5 are colored as green and species with mean concentrations lower than 1 are not colored. We have also listed those species having mixing ratios greater than 10 ppbC in **Table 4**, and present a tabulation of the average number of species within each concentration category in **Table 5**.

The following conclusions follow from the information in **Tables 3-5** (see also the tables in ***Appendix_A_basicStatistics.pdf***):

1. Mean individual and total VOC concentrations at *nighttime* are higher than those measured during the *daytime*. For the five sites near Ship Channels, the mean total VOC mixing ratios range from 107 to 270 ppb during the daytime and 226 to 513 ppb at night. Nighttime VOC probably better reflect the actual VOC emission due the absence of photochemistry at night. The lower concentration at daytime is likely due to such photochemical reactions.

2. The mean concentration of the VOCs at the five auto-GC sites near the Ship Channel are generally higher than those measured at the other sites, but the measured VOC concentration are Mustang and Texas City are comparable to those in *Clinton* and *WallisVille* sites. The higher concentration for sites near the Ship Channels is not surprising given that most of the large emission sources are along the Ship Channel.

3. The most abundant VOCs are *ethane, propane, butanes, pentanes* followed by *ethene, propylene, benzene, toluene, n-hexane* and *cycloHexane*. The great abundance of the alkanes may attribute to both high emission and accumulation resulting from their longer lifetime and lower reactivities. The abundance of the highly reactive low lifetime alkenes and aromatics is consistent with the high emission rates from local sources. Very high concentrations of *benzene, toluene* and *styrene* were observed at Lynchburg Ferry site relative to the other auto-GC locations, and the maximum concentration for all species, at any of the sites was 7897.5 ppbC for *styrene* (c45220) at Lynchburg Ferry. The large difference between the mean and median concentrations of *styrene* at Lynchburg Ferry site suggests great skewness in the distribution of *styrene* concentration, although remove the single maximum value still results in extremely high *styrene* concentrations relative to what was observed at the other sites. For example the nighttime mean *styrene* concentration at Lynchburg Ferry was 5.51 ppbC even after removing the local maximum concentrations that was greater than 1000 ppb.

4. Discrepancies were found between the calculated *TNMHC* / *TNMTC* as the sum of individual species and the measured *TNMHC* / *TNMTC*. The sum of the PAMS targeted hydrocarbons should be very close to the measured *TNMTC* and the sum of all hydrocarbons should be very close to the measured *TNMHC* or "*sum_pol*." This was not found to be the case, as seen in the bottom of **Table 3**. Rather, the percentage of the calculated total VOCs to the measured total VOCs (*TNMHC* and *TNMTC*) differ greatly. **Figure 2** shows the scatter plots of the percentage of the sum of all the hydrocarbons to the measured *TNMHC* (blue) and its surrogate *sum_pol* (red) (No plots were drawn for the Lake Jackson and Mustang sites because no *TNMHC* and *sum_pol* were reported). These discrepancies can be seen clearly in **Figure 2** which shows that the percentages of the sum over all individual VOCs to the measured *sum_pol* were systematically greater than 100%, but the sum over all individual VOCs relative to the measured *TNMHC* itself were below 100%, ranging from average 70% at *Clinton* and *Texas City* to average of 96% at the *WillisVille* site. Such different behaviors of the *TNMHC* and its surrogate *sum_pol* suggest problems with these surrogate species. From **Figure 3** it can be seen that the percentages of the calculated *TNMHC* based on summing from the individual to the measured *TNMHC* or its surrogate *sum_pol* vary depending on both the actual concentration values and the site locations. In general such percentages deviated from 100% more for samples with low total VOC concentrations.

5. We found that most of the minimum concentrations were zero which suggests that concentrations below the minimal detectable limit (MDL) values were assigned this value. We also found no species at any site with an average concentration exceeding 100 ppbC although the *maximum* concentrations of many species exceeded 100, and even 1000 ppbC. Although not shown here, the associated concentration histogram of these species were found to have a strong positively skewed nature.

Diurnal Patterns

The diurnal patterns of the measured concentrations were first inspected through both box plots. Sample box plots (of *ethane* and *isoprene* at *Haden Road)* are shown in **Figure 3**. Although different diurnal patterns can be seen for these two species, the large range of values tends to mask significance diurnal patterns in the box plots. We found that a better alternative to highlight diurnal patterns was to plot the median and mean concentrations as a function of time. To further emphasize these daily changes we have removed information on the spread of the concentration. The diurnal patterns plots of all species at all sites are attached as a supplemental file, (***Supplement_A_DiurnalPatterns.pdf***).

We illustrate the type of information contained in this supplemental file by looking more closely at the characteristic examples in **Figure 4*a* and 4*b*,** which show the diurnal patterns of *n-butane* and *isoprene* for all the auto-GC sites (red lines show the hourly mean values; green lines show the hourly reaction; solid lines apply to weekday values; and dashed lines apply to weekend values). Almost all species except *isoprene* share more or less similar diurnal patterns to that of *n-butane* shown in **Figure 4*a***. The average concentrations of those species have minima around noon and gradually increase to their maxima in the early morning hours (5-8am), then decreasing to their noon time minima. This pattern can be explained by the accumulation of fresh emission during the nighttime in the absence of photochemical reaction, leading to enhanced concentrations after sunset and before sunrise, and to the daytime depletion of the VOCs through photochemical destruction and their dilution associated with the increase in the boundary layer height during the day.

Isoprene shows an almost opposite diurnal pattern (**Figure 4*b***) to the other species. The concentrations of *isoprene* are lowest during nighttime and starts increasing from mid-night to reach maximum in the morning. The broad peak is remained for most of the day and then decreases to minimum around mid-night. This diurnal behavior of *isoprene* is consistent with a biogenic source produced at a rate governed by, among other things, photosynthetic activity. While the increase of *isoprene* at *Haden Road* to a maximum at noon may be explained by natural emission, the sharp increase of *isoprene* during the early morning (around 6am), when there is less sunlight for photosynthetic activity, at *Lake Jackson* and *Danciger* is not so easily explained as a natural emission and probably reflects the presence of an anthropogenic source as well.

It was interesting to find that the times at which *isoprene* reached its maximum value varied from site to site. The *Lake Jackson* site had an earlier maximum and *Danciger* had a later maximum than the other sites. The *Lynchburg Ferry* site and, to a lesser extent, the *Mustang* site, had no obvious noontime maximum nor did they exhibit any large variation in the concentration during the day, again suggesting anthropogenic source components for *isoprene* at these sites.

The diurnal pattern of the total VOC is very similar to that of *n-butane* shown in **Figure 4*a***. As mentioned above that most of the individual VOC species, including the

most abundant species, share more or less similar diurnal patterns with *n-butane*, it is not beyond expectation that the diurnal pattern of the total VOC is similar to what observed for *n-butane.*

Scatter Plot and Correlation

Scatter plots between species provides a quick and simple way to visualize relations between species. The *"TNMHC"* (*total non-methane hydrocarbon measurement*) was found to be a more robust quantity than the measurement for any individual species. By plotting all the individual species against the total non-methane hydrocarbons in the form of a matrix of scatter plots [Lewis et al, 1998], we have a way to quickly assess data quality. One matrix of scatter plots was created for each of the sites. **Figure 5** presents the matrix of scatter plots at *Haden Road* site as an example of how these plots can be used. This sample matrix of scatter plots shows that few species were well correlated with the TNMHC and great variations exist, suggesting that either the *TNMHC* is not dominated by a single source or that most species have a large number of sources. Outliers can also be readily seen in plots having visually obvious correlations.

The scatter plots between for all species are given in the supplemental files ***Supplement_B_Scatplt_*[*siteName*]*.pdf.*** An example of the information contained in these files is given in **Figure 6**, showing the scatter plot for *acetylene* and *ethylene* at the *Haden Road* site. A bifurcation can be seen on the scatter plot between *acetylene* and *ethylene* with one set of points following the horizontal axis and another set of points following the vertical axis. The ratio of *ethylene/acetylene* from the vertical axis is about 3, suggesting gasoline powered vehicle exhaust [Fujita and Campbell, 2003]. This ratio along the horizontal axis suggests a source with very high *ethylene/acetylene* ratio which clearly excludes vehicle exhaust. A very high degree of correlation was found between isomers. Correlation coefficients between all species were calculated and those pairs with high correlation coefficients and other pairs that the two compounds are supposed to be emitted from a common source were checked. **Table 6** lists the species pairs with apparent correlation coefficient equal to and larger than 0.85 at each of the nine sites.

Time Series and Outlier Detection

Time series of concentration data for all species at all the auto-GC sites were examined to locate extreme, zero and missing values that could adversely affect the robustness of the subsequent statistical analyses. Only those spikes with concentrations significantly greater than the surrounding data were identified as "*outliers*." No statistical criterion was used because of the dependence of observations at time *t* with observations at time *t-dt*. Therefore the identification of extreme values was based on visual inspection of the time series. The time series of all species at all sites are provided as appendixes (the file name convention is, i.e. ***Supplement_C_timeSeries_*** **[*siteName*]*.pdf***, where 'siteName is the name of one of the auto-GC sites). **Figure 7** shows an example of an extreme value identified in the time series of *acetylene* for the *Clinton* site. In our subsequent PMF analysis, the error estimates for the extremes were doubled to alleviate their influences on the factorization.

Examination of the identified extremes revealed that (1) at a given site, extremes often occur for multiple species at the same time while (2) for a given species there are few patterns suggesting simultaneous or temporally linked patterns across the 9 sites. Extreme events are more likely *local* events that do not affect multiple sites simultaneously. These flagged extremes values may be useful in future studies using back-trajectories to trace the origin of these localized extreme events. The identified extremes for all sites were documented in the tables in the appendix file ***Appendix_B_Extremes.pdf.***

Backward Wind Trajectories of Seasonal Maxima

The wind trajectories of the seasonal extremes of all species at five sites close to Ship Channels are presented in the supplemented file of ***Supplement_D_Trajectories_SeasonalMaxima.pdf.*** As an example the 6 hour wind trajectories of seasonal extremes of *iso-butane* from the five auto-GC sites close to Ship Channel are presented in **Figure 8**. There is no apparent pattern in the wind trajectories for these seasonal extremes. Extreme values of *iso-butane* could be associated from a back trajectory characterized by straight air flow from distant or a more sinuous air flow allowing for recirculation in the proximity of the sites. The independence of the extremes in the trajectories patterns suggests that the sources are very *local* with the result that they are transported to the sites through a variety of paths.

Categorized Species Percent Contribution

The evaluation of emission inventories involves the comparison of the measured ambient VOC concentration with the assumed emissions. Such a comparison is difficult because no inventory data was available during the sampling period of the auto-GC measurements. In a recent VOC emission study (Roberts, et al., 2004), inventory composition by species categories were prepared for several auto-GC sites near the Ship Channel including the *Clinton* and *Haden Road* sites, using an inventory based on the year 2000 data. When lumping species into categories and in the calculation of the percent composition, *ethane* and *propane* were excluded. The inventory composition was compared with measured ambient VOC concentration during the summer of 2000. Recognizing these limitations, we have used the 2000 inventory to analyze results from the 2003 auto-GC measurements.

The mapping of lumped species (from the inventory) to specific VOCs is given **Table 7**. The pie charts on the left side of **Figure 9** show the percent contribution of observations, lumped according to **Table 7**, at each site while the panels on the right side show the corresponding emission values from the *Clinton* and *Haden Road* sites, reproduced from Figure 2-15*e* and Figure 2-21*e* of Roberts [Roberts, et al., 2004]. As described in Roberts [Roberts, et al., 2004], the emission inventories were based all VOC emission in a 2 km by 2 km area surrounding the sites, with *ethane* and *propane* excluded in the composition

calculation since these species can be formed by other photochemical processes and, because their long lifetimes of days, may not be well characterized well an emission inventory.

For the *Haden Road* site, the composition of the ambient data (*left side of figure*) compared to the emission inventory species composition (*rightside of figure*) shows that *acetylene* (0.5% in ambient data versus 4% in inventory), *benzene* (2% in ambient data and 9% in inventory), *ethene* (5% in ambient data and 8% in inventory) are significantly overestimated while *butanes* (16% and 7% in ambient and 12% and 4% in inventory for *n-butane* & *i-butane* respectively) are underestimated.

For the *Clinton* site, the composition of the ambient data (*left*) compared to the emission inventory species composition (*right*) shows that good agreement on most species with the exception of *benzene* (2% in ambient data and 5% in inventory) which is overestimated and *ethene* (13% in ambient data and 7% in inventory) which is underestimated.

The pie charts in **Figure 9** also show significant differences in the composition among the sites. For example, the percentage of *ethene* ranges from 5% at *Haden Road* site to over 20% at *Channel View* and *WallisVille* sites and even 30% at *Danciger*.

WIND DIRECTION AND WIND TRAJECTROY ANALYSES

Wind Direction Analysis with Conditional Probability Function (CPF)

To identify likely locations of local point sources a conditional probability function (CPF) [Ashbaugh et al. 1985, Kim et al., 2003*a,* 2003*b;* Begum et al., 2004] was calculated using source contribution estimates from the PMF analysis together with surface wind directions measured at the sites. Mathematically, the CPF is defined as

$$\mathrm{CPF} = \frac{m_\theta}{n_\theta}$$

where m_θ is the number of samples in the wind sector θ with mixing ratios greater than some 'high' concentration and n_θ is the total number of samples in the same wind sector. We have defined 'high' as measurements greater than the 75th percentile of all the observations from a given station. A plot in polar coordinates, with the radial distance defined by the magnitude of CPF and the angle defined by the associated wind direction, visually illustrates the fraction of samples coming from a given direction that have 'high' values, and point to regions associated with these high values.

Only direction is indicated from the CPF analysis with the implication that any source along a line indicated by a CPF plot could, in principle, be a contributor to the selected species. The implicit assumption is that the air flow carrying the substance is more or less direct, which is likely to be a good approximation for local sources.

The results reported here apply only to nocturnal observations so as to remove changes in mixing ratios of reactive VOCs associated with photochemical oxidation. All values, including observations at wind flow patterns, are also based only on nighttime conditions.

Each CPF plot is associated with a specific compound and location. An example of a CPF plot is shown below in **Figure 10**. The data for this figure came from observations of *ethane*, made at night from the *Channel View* between June and October of 2003. This figure shows that samples measured when the wind came from approximately 70-80°, 150-180° and 180-210° (on the meteorological compass shown in the figure) had relatively higher probabilities to be greater than or equal to the 75th percentile value of ethane (a value that is not indicated in the figure). An arrow drawn from the center of these figures to the main lobes would point to the directions associated with high concentrations of the measured compound.

There are several technical caveats to be considered in using these figures to identify the direction from which the respective VOC originated.

1. Winds with speeds of less than 1 ms^{-1} were excluded from this set of calculations because of the uncertainty in defining directions for low wind speeds.

2. In order to assure robustness of the results, wind directions were binned into 36 wind sectors of 10 degrees per sector.

3. The CPF values were weighted by the number of occurrences of the wind from each sector. To illustrate the significance of this weighting, we note that if the wind never came from a given direction, there would be no VOC observations from that direction. Conversely, a high frequency of winds from a given sector would enhance the confidence of the CPF associated with that sector. We have taken this uncertainty into account by weighting the results for each sector by the number of observations in that sector. To illustrate this concept, we note that the wind rose associated with observations of *ethane* (**Figure 11**) shows a high occurrence of winds between 60 and 80° and fewer winds between 80 and 140°. Weighting factors to take this into account have been included in the CPF plots to follow.

The original analysis using the CPF approach were based only on nighttime observations. Using the CPF plots to identify the direction on which the species comes from implicitly assumes that the air carrying the emissions have a straight travel path. However a very large degree of variability was found in the paths of trajectories traveled arriving at the receptor sites. **Figure 12** shows 6-hour wind trajectories associated with the seasonal *iso-butane* maxima measured during summer night at the five auto-GC sites close to Ship Channel. Within the same six hours, the trajectory carrying the maximum *iso-butane* to the *Lynchburg Ferry* site was relatively straight, coming from a distant area identified by the trajectory. But the maximum *iso-butane* concentration measured at the *Haden Road* site during the summer nights arrived by a trajectory that circulated around this site. The latter situation would thus be associated with a high degree of uncertainty in defining the

direction of incoming air, with the associated direction in a CPF plot being be much less indicative of the source region than would be the case for relatively straight flow.

To limit the uncertainty introduced into the CPF calculations by using trajectories with very sinuous natures we have made use of only a subset of the auto-GC values (day and night). This subset of auto-GC hydrocarbon values consists only of the observations made during one hour periods for which the flow had a relatively constant direction and steady speed. The selection of time intervals during which the wind came from a relatively constant direction was determined by evaluating the ratio, S, of two distances, D, traveled by an air parcel over a 6 hour period,

$$S = D_{straight}/D_{actual}$$

where $D_{straight}$ is the straight-line distance and D_{actual} is the distance along the trajectory path, calculated as the sum of the distances for the individual segments (calculated over 5 minute intervals). All distance calculations were based on the formula for great circle distance computation based on spherical trigonometry (http://www.ga.gov.au/nmd/geodesy/datums/distance.jsp):

$$D = 1.852 \times \times 60 \times arcos(sin(Lat1) \times sin(Lat2) + cos(Lat1) \times cos(Lat2) \times cos(\Delta Longitude)$$

where *Lat1* and *Lat2* are the latitudes of the first point and the second point respectively, and $\Delta Longitude$ are the difference of the longitude of the second point to the longitude of the first point.

Small values of S reflect trajectories with curved approaching patterns, in contrast to straight trajectories having S = 1. We have selected a cut-off of S = 0.95 to define relatively direct trajectories which, we believe, are better indicative of source regions than are trajectories for which S is less than 0.95. About half of the samples were retained in the computation for the threshold used. The values of S of the trajectories showing in **Figure 12** were given by the numbers besides the site location dots in the figure.

The resulting CPF plots for the summer and fall, for all species, at the 5 sites along the Ship Channel, are contained in the attached file, ***Supplement_E_CPFplots.pdf***. **Table 8** lists the categories of species of the point-source inventory and how they mapped to the observed VOCs measured at the auto-GC sites. Examples of the type of information contained in this supplemental file are shown in **Figure 13,** the wind roses and **Figure 14**, taken from the CPF figures for *ethane* in this supplemental file. We will refer to **Figures 13 and 14** to illustrate patterns contained in the individual CPF plots within the supplemental file. The wind rose is presented to aide in understanding the CPF plots.

As shown in **Figure 13**, the wind is more southerly in the summer than during the fall and the daytime flow more southeasterly than at nighttime for both summer and fall. The domain shown in this, and all the CPF figures in the supplemental file, is about 18.9 km along north-south direction and 38.6 km along east-west direction. This southerly flow is reflected in many, but not all, the characteristic CPF lobes presented in **Figure 14**. A

comparison of the wind roses with the CPF plots will show frequent flow from various directions but no associated CPF lobe, implying an absence of sources from the direction indicated by the wind rose.

Detailed analyses on a selected *olefins* and *alkanes* at *Haden Road* are given in the following sections. The selected *olefins* include *ethene, cis-2/trans-2-butene, cis-2/trans-2-pentene, styrene* and an initial list of alkanes include *n-undecane, n-decance, methylCycloHexane, methylCycloPentane, n-nonane, n-octane.* Several additional *alkanes* were added based on discussions with TCEQ including *C4 butanes, C5 pentanes,* and three *C6 alkanes, n-hexane, 3-methylPentane,* and *2-methylPentane*. The CPF plots of the selected VOC species are included in the Appendix C (*Appendix_C_CPFselected.pdf*).

Olefins: An Overview of Both Day and Nighttime CPF Plots

Anticipating differences in mixing ratio as a function of season (summer, defined here as July and August; and fall, defined as September-October) and nocturnal and daytime circulation patterns, we have partitioned our CPF plots into four categories. Each page of the CPF figures in the **Appendix C** present the summer figures on the left and fall figures on the right. The top row contains the nighttime results and the bottom row the daytime results.

Figure 13 shows the frequency of the wind direction in both summer (July-August) and fall (September-October) during the night (13A & 13C) and day (13B & 13D), respectively. The summer nighttime flow is characterized by southwesterly flow, in the contrast to the daytime pattern which shows more southeasterly flow. The fall nighttime flow is characterized by northeasterly wind patterns, with the daytime flow also having a southeasterly contribution. We would note that there are very few, if any, periods when the flow was from the north during the summer months (day or night). A northerly component is present in the fall (for both day and night).

The daytime-nighttime differences in *ethene* were found to be relatively small (contrast **Figures *Appendix_C_1A*** with ***Appendix_C_1B*,** and **Figures *Appendix_C_1C*** with ***Appendix_C_1D***) relative to the differences the two seasons (see **Figures *Appendix_C_1A*** with ***Appendix_C_1C***, and **Figures *Appendix_C_1B*** with ***Appendix_C_1D***). The *Haden Road* site had elevated *ethene* levels in the summer, both day and night, when local winds were from the south and southeast. In the fall, the elevated levels of *ethene* are associated solely with southeasterly flow but not southerly flow. Similar patterns are found at the other sites shown in these figures.

There are pairs of olefins whose member shows very similar CFP patterns. Examples of species having similar CPF patterns include *c2 butene* and *t2 butene* (see **Figures *Appendix_C_2A*** and ***Appendix_C_3A***, showing similarities in these compounds at night, during the summer) and *c2 pentene* and *t2 pentene* (see **Figures *Appendix_C_4A*** and ***Appendix_C_5A***, also for nighttime, in the summer). The similar CPF patterns suggest that these pairs might have a common source. While these chemically related species

(*c2/t2-butene* and *c2/t2-pentene*) frequently display similarities, there are enough differences to suspect that although there are many common sources, contributing sources are not identical.

The wind patterns illustrated in **Figure 13** can explain many, but not all, of the seasonal differences shown in the seasonal plots of **Figures *Appendix_C_1A*** through ***Appendix_C_5A***. To illustrate this point, we compare the CPF figures for *c2 butene*, shown in **Figures *Appendix_C_2A*** with those for *t2 butene* in ***Appendix_C_3A*** (summer vs. fall, both at night), and **Figures *Appendix_C_2B*** with ***Appendix_C_3B*** (summer vs. fall, both during the day), focusing on the *Haden Road* site. In **Figure *Appendix_C_2A***, the summer plot, the predominant lobe in the *Haden Road* CPF plot is pointing to the southwest. In contrast, **Figure *Appendix_C_2C***, in the fall plot, has three lobes, only two of which (from the northeast and southeast) can be explained by the wind rose in **Figure 13C.** Local meteorology cannot explain all of the features identified in this analysis.

The CPF pattern for nighttime observations of *styrene* (both summer and fall; **Figures *Appendix_C_6A*** and ***Appendix_C_6C***) appear to be inconsistent relative to what would be expected from the associated wind roses. To illustrate this point by way of example, note that the *Channel View* site shown in **Figure *Appendix_C_6A*** has a lobe pointing to the northeast, in contrast to the wind rose for *Channel View* shown in **Figure 13A**, which has no corresponding lobe. This also appears to be true for the daytime observations of *styrene* presented in **Figures *Appendix_C_6B*** and ***Appendix_C_6C*;** there is little, if any relation, between the local wind rose and the observations. Very localized sources may explain this, and also explain the similarities apparent in the fall observations of styrene observed at night (**Figure *Appendix_C_6C***) and during the day (**Figure *Appendix_C_6D***).

Alkanes: An Overview of Both Day and Nighttime CPF Plots

As alkene emissions are reduced, the question naturally arises, are there other less reactive species that may play an increasingly dominant role in oxidation processes over the Houston/Galveston area? One broad category of species that generally play a role of lesser importance, at least relative to the alkenes, are the alkanes. We were first focusing on the six most highly reactive alkanes (with respect to OH): *n-undecane* (k_{OH} = 13.2 s^{-1}), *n-decane* (k_{OH} = 11.6 s^{-1}), *methylcyclohexane* (k_{OH} = 10.4 s^{-1}), *n-nonane* (k_{OH} = 10.2 s^{-1}), *methylcycloPentane* (k_{OH} = 8.81s^{-1}) and *n-octane* (k_{OH} = 8.68 s^{-1}).

The resulting CPF figures are shown in **Figures *Appendix_C_7*** through ***Appendix_C_12***. Many patterns similar to those already noted for alkenes are apparent in these plots. A small, but distinct, directional difference can be observed in the nighttime summer CPF lobes at *Haden Road* site for *n-undecane* (**Figures *Appendix_C_7A***), *n-decane* (**Figure *Appendix_C_8A***), *n-nonane* (**Figure *Appendix_C_11A***) and *n-octane* (**Figure *Appendix_C_12A***), relative to the CPF lobes for *methylCyclopentane* (**Figure *Appendix_C_10A***) and *methylecyclohexane* (**Figure *Appendix_C_9A***). The former are oriented along an axis somewhat less than the 220° axis. The later are oriented

approximately 10° less (closer to the 200° axis). This difference is possibly indicative of difference sources for these two groups of compounds. Similar features can be seen in the nighttime fall CPF plots. One feature of interest is a northeast-pointing CPF lobe in the nighttime summer at the *Haden Road* site for *n-undecane* (**Figure *Appendix_C_7C***), *n-decane* (**Figure *Appendix_C_8C***) and *n-nonane* (**Figure *Appendix_C_11C***) which is absent for *n-octane* (**Figure *Appendix_C_12C***). As noted previously in our overview of alkene CPFs, the wind patterns illustrated in **Figure 13** can explain many, but not all, of the seasonal differences shown in the Fall/Summer alkane plots of **Figures *Appendix_C_7*** through ***Appendix_C_12***.

C4 *butanes,* C5 *pentanes,* and three C6 alkanes, *n-hexane, 3-methylpentane,* and *2-methylpentane,* are alkanes identified by TCEQ as possibly making significant contribution to the total reactivity. In the CPF plots discussed in the following section, we have added arrows originating at the receptor location to emphasize the direction of a key lobe, or, when there is a wide lobe, then added two arrows. No arrows were added when the CPF lobes did not point to an obvious cluster of point-sources, as indicated in the emission map.

*N-Butane (**Figure Appendix_C_13**)*

1. In the summer at the *Clinton, Haden Road* and *Lynchburg Ferry* sites, most of the CPF lobes for both the nighttime and daytime CPF plots (**Figures *Appendix_C_13A & 13B***) are directed at the point source clusters to the south, with the *Channel View* site oriented to point sources to the south. At *WallisVille* site, during daytime (**Figure 3B**), there is no source shown on the inventory map corresponding to the northeasterly lobe shown in this panel.
2. The directions associated with high values at the *Clinton, Haden Road* and *Lynchburg Ferry* sites during the fall are associated with well defined clusters of point sources. However, there are no point sources associated with the northeasterly lobe at the *WallisVille* site in fall (**Figure Appendix_C_13C & 13D**) and no source associated with the lobe pointing to the southeast from *Channel View* (**Figure Appendix_C_13D**). While no CPF lobes point to the cluster of the sources located to the south and southwest for *WallisVille* (**Figures *Appendix_C_13C & 13D***) this may be a simple result of having infrequent wind patterns to the sites during the fall (**Figure 13**).
3. In the fall the relative large cluster of point sources at the northeast corner of the map are along the directions of the lobes of the *fall-nighttime* CPF plot at *Channel View* and *Lynchburg Ferry* sites (**Figure *Appendix_C_13C***), but it is strange that a signal from these point sources was not identified in the CPF plot for *WallisVille,* which is much closer to these emissions.

*iso-Butane (**Figure Appendix_C_14**)*

1. **Figure Appendix_C_14** shows the seasonal and diurnal CPF plots of iso-butane on our background map with point-sources. The inventory map used here is the same as that used for n-butane in **Figure Appendix_C_13** because the inventory is based on the sum of these two species. The CPF plots of iso-butane in **Figure Appendix_C_14** share many common features with those of n-butane in

Figure Appendix_C_13. Many of the observations previously made for n-butane applies equally well to iso-butane.

*Pentanes (**Figures Appendix_C_15** and **Appendix_C_16**)*

1. The inventory map for (**Figures *Appendix_C_15*** & ***Appendix_C_16***) is quite similar to that for the butanes (**Figures *Appendix_C_13*** & ***Appendix_C_14***) with a few notable exceptions.
   a. There is a large source of butane south of *Lynchburg Ferry* (e.g., **Figure *Appendix_C_13A***) that has no corresponding point source of *n-pentane* (e.g., **Figure *Appendix_C_15A***).
   b. A large point source of n-butane is located northeast of *WallisVille* (e.g., **Figure *Appendix_C_13A***) with no corresponding point source shown for *pentane* (e.g., **Figure *Appendix_C_15A***).
2. The *WallisVille* site shows a strong directional dependence for *n-pentane* during the summer, with a high occurrence of elevated values found under southerly flow from the cluster of known pentane sources due south-south-east of this site (**Figure *Appendix_C_15A***). This dependence is not seen during August-September time period (**Figure *Appendix_C_15C***) when winds are more northerly.
3. Although there are large point sources of pentane to the north of *Channel View*, the CPF plots do not implicate them in high values at *Channel View* (e.g., **Figures *Appendix_C_15A*** and ***Appendix_C_16A***). Rather, elevated pentanes at this site come from the south during both the summer and fall seasons. This cluster of point sources north of *Channel View* does not appear to play a major role in producing any of the CPF lobes at other sites, especially in the summer. This lack of influence on the CPF analysis is similar to a lack of influence noted for butanes, as seen in **Figures *Appendix_C_13*** and ***Appendix_C_14*.**
4. Similar to the butane CPF results shown in **Figures *Appendix_C_13*** & ***Appendix_C_14***, we found that the directional dependence of the CPF plots for the *fall*, for *WallisVille*, also point to the northeast (**Figure *Appendix_C_15C, 15D, 16C* and *16D***), but there is no point source of pentane shown in the inventory.
5. The CPF plots of *iso-pentane* (**Figure *Appendix_C_16***) are generally similar to those of n-pentane (**Figure *Appendix_C_15***). The observation previously noted for *n-pentane* apply equally well to *iso-pentane*.

*Hexanes (**Figures Appendix_C_17, Appendix_C_18** and **Appendix_C_19**)*

1. Figures ***Appendix_C_17***, ***Appendix_C_18*** and ***Appendix_C_19*** show the CPF plots of *n-hexane, 2-methylPentane* and *3-methylPentane*. The corresponding inventory used as background is based on the sum of five C6 alkanes that include these three species.
2. The CPF plots for *n-hexane* (**Figure *Appendix_C_17***) show a close relationship to the point sources shown in this inventory.
3. The fall CPF plots, for both the day and nighttime values, for *WallisVille* (**Figures *Appendix_C_17C*** and ***17D***) indicate source of *n-hexane* to the northeast that does not correspond to a close point source on the inventory.

4. Similar to what were observed for butane and pentane categories, the cluster of point sources right to the north of *Channel View* site appear to make no contribution to elevated values of *n-hexane* (**Figure *Appendix_C_17***).
5. Measurements of the two substituted *hexanes, 2-methylPentane* and *3methylPentane*, were not available at *Channel View* (**Figures *Appendix_C_18*** and ***Appendix_C_19***)
6. The CPF figure for *2-methylpentane* and *3-methylPentane,* measured at *WallisVille* in fall (**Figures Appendix_C_18C** and **D**, **Figure Appendix_C_19C** and **D**) indicates a source of these compounds that does show up in the inventory.

Some general observations of the CPF plots:

1. Most lobes in the CPF plot pass through known point source clusters in the inventory. It is hard to pinpoint which specific point source is a main contributor because so many are located along the axis of the major CPF lobes.

2. The *WallisVille* site frequently shows elevated values coming from the northeast although no point source is indicated in the inventory.

3. The point sources north of Channel View appear to play a relatively small role in contributing to elevated values observed at this site.

Clustering Analysis on the CPF Plots

In the sections above we have presented the CPF analysis for individual species and noted that many of the CPF plots had common patterns that could be explained by common sources. However, with over 50 VOCs, and the day/night, summer/fall CPF plots for each VOC, a full comparison of species have a common source area would be difficult, if not impossible to do, by visual inspection alone. To carry out this aspect of the project, we have completed a cluster analysis study of the CPF data, which we present in this section.

Recall that CPF plots show the fraction of auto-GC samples for a VOC species coming from a given direction having 'high' values, defined here by the 75th percentile of each species of VOC. The resulting polar coordinate plot has lobes pointing to regions associated with these high values. Clustering Analysis is a multivariate statistical tool for detecting natural groupings in the data [Hartigan, 1975]. We have selected to use a 'hierarchical clustering' analysis for the work to be shown here. The Pearson correlation coefficient between two CPF plots is taken as a measure of the similarity, or *distance*, measure since it relates CPF patterns rather than the magnitudes of the CPF lobes. The Ward linkage was used to amalgamate the clusters to form the hierarchical cluster tree, using the 'distances' between pairs of objects (species) in different clusters and making adjustments based on covariance to decide how far apart the clusters are. Since the previous CPF analyses on the individual VOC species were conducted separately on the *nighttime* and *daytime* VOCs measurements at each season (*i.e.,* pre-Summer, or June, Summer (July-August)and Fall (September-October) at each of the five sites (*i.e.,*

Channel View, *Clinton, Haden Road, Lynchburg Ferry* and *WallisVille*), a dendrogram (*i.e.,* a cluster tree) was thus obtained for each of the data subset. As an example **Figure 15** shows the dendrograms obtained from the clustering analysis on the *daytime* and *nighttime* VOC CPF plots during pre summer season at *Channel View* site. All the dendrograms from the clustering analysis were presented as figures in **Appendix D (*Appendix_D_CPF_Clustering.pdf*)**. **Figures *Appendix_D_1*** to ***Appendix_D_5*** show the cluster trees of the VOC species based on the CPF plots at the five sites near Ship Channel at each season. The labeling convention for the figures is as follows:

- Figures labeled 'A', 'B' and 'C' refer to observations made during the month of a) June (referred to as 'pre-Summer'), b) July-August (referred to as 'summer') and c) September-October (referred to as 'fall').

- The left side of each figure refers to observations made during the daytime and the right side of each figure referees to observations made at night.

We again clarify the interpretation of these figures by way of an example. **Figure 15** (taken from **Figure *Appendix_D_1***) shows the clustering for the *Channel View* site, with daytime values in the left panel and nighttime values in the right panel. In a hierarchical clustering analysis, VOC species showing most similar CPF patterns join as small groups first. Then small groups of VOCs and/or individual VOCs gradually form bigger groups based on the similarity among the CPF patterns. As the tolerance of the similarity in the patterns relax (larger distance), more small groups of VOCs and/or individual VOC can be placed into a big group. Eventually all VOCs used for the clustering would be put in a single group as the *distance* in the cluster tree (dendrogram) move to the farthest right. By breaking up the cluster tree with different *distance* tolerance, different number of groups of VOCs will be obtained.

Focusing on the daytime values, we note five light gray boxes encompassing a list of VOCs measured at this site. These are the five clusters identified as having similar CPF plots, and hence, having a common directions associated with elevated values. Within each cluster we have further identified some major or more reactive species that have CPF plots that are nearly identical; these are shown by the colored ovals and circles encompassing the species with this property. Arrows have been added to the ovals and circles to link selected species having common CPF plots of interest (*either* major species *or* more reactive species, *e.g.,* see the bottom left side of **Figure *Appendix_D_2C***, where *ethene* and *propylene* had similar patterns but were separated by *cycloPentane*. In this case *cycloPentane* also shares quite similar CPF pattern with *ethane* and *propylene*, but we are not intending to address *cycloPentane* due to the lack of a clear relationship between it and *ethene* or *propylene*). The far right side of each panel shows a single horizontal line indicating that all of these species had a first order commonality in source region. This line splits into two to indicate the next two clustering patterns, and so on, until the specified five clusters is identified on the far left, and subgroups within these five are identified by the ovals or circles.

By visually inspecting all the 28 cluster trees, the following observations on the species grouping can be made (Note: no data from *Clinton* site was available for June):

1. *Alkenes:* CPF patterns of alkanes are usually quite similar. Tight grouping of some species were marked by red circles for easy viewing.

    a. *C2butene* and *t2butene* usually come together. In some cases (*i.e.,* seasonal / diurnal condition) such close linkage can be extended to other *butanes* such as *1,3-butadiene* and *1-butene*. The exceptions are the June nighttime at *Lynchburg Ferry* (**Figure *Appendix_D_4A***), June daytime (**Figure *Appendix_D_5A***), summer daytime (**Figure *Appendix_D_5B***) and both daytime and nighttime in fall (**Figure *Appendix_D_5C***) at *WallisVille* site.
    b. Similarly *C2pentene* and *t2Pentene* is another pair which usually shows very similar CPF patterns. The other *pentenes*, *i.e., 1-pentene* are also often grouped with them. The CPF plots of the June (daytime and nighttime) at *Lynchburg Ferry* site seem like an exception where *c2pentene* and *t2pentene* are separated apart, but they are actually still in the same bigger group in the daytime (**Figure *Appendix_D_4A*** *left*).
    c. The CPF patterns of *butenes* and *pentenes* are different; their clusters in the tree are apart away. In some cases, such as June nighttime (**Figure 15A** or **Figure *Appendix_D_1A***), summer nighttime (**Figure 15B** or **Figure *Appendix_D_1B***) at *Channel View*, summer at *Clinton* site (**Figure *Appendix_D_2B***), fall daytime at *Clinton* site (**Figure *Appendix_D_2C***), fall nighttime at *Lynchburg Ferry* (**Figure *Appendix_D_3C***), the CPF patterns are as shown by their lumping into a single bigger group.
    d. In a less extent *propylene* and *ethene* consists of another pair which shares quite similar CPF patterns in most of the cases. These two species join to form a small cluster or join with others in a bigger group. Sometimes *1,3butadiene*, *1-butene* or *1-pentene* is hanging with them.

2. *Alkanes:* a number of small alkanes groups can be found in the clustering trees. Tight grouping of some species were marked by blue circles for easier viewing.

    a. *Ethane* and *n-propane* form a small group in the clustering tree in many cases (**Figures *Appendix_D_1*** to ***Appendix_D_5***). In some cases the group can be extended to the inclusion of *n-butane* and *i-butane* but in some other cases they separate far away in the clustering tree due to quite distinct CPF patterns (***Figure Appendix_D_2B right, Figure Appendix_D_2C right, Figure Appendix_D_3A left/right, Figure Appendix_D_3C left, Figure Appendix_D_4A left, Figure Appendix_D_4C left, and Figure Appendix_D_3A right***).
    b. Groups of isomers such as (1) *2,3,4-trimethylPentane* and *2,2,4-trimethylPentane,* (2) *2-methylPentane* and *3-methylPentane,* (3) *2,3-dimethylPentane* and *2,4-dimethylPentane*, (4) *2,2-dimethylButane* and *2,3-dimethylButane,* (5) *2-methyl-Heptane* and *3-methyl-Heptane,* (6) *2-methyl-Hexane* and *3-methyl-Hexane*. The species in these isomer pairs either form the

smallest two-component groups or are in the same bigger groups in the clustering trees indicating the similarity in the CPF patterns. These isomer groups could form a common bigger group in some situations at some sites, or locate in separate bigger groups in other situations, suggesting the source and transport complex of these species in difference seasonal / diurnal condition at different sites.

c. We found the grouping of the long-C alkanes of *n-octoane, n-nonane, n-decane* and *n-undecane* not to be as consistent as the other groups. In many cases either two or three of these four long-C alkanes are close in the clustering tree suggesting a common source or pathway of the long-C alkanes. However it is also noted that different groups form in different situations at difference sites.

3. *Aromatics:* Tight grouping of some species were marked by green circles for easier viewing.

   a. In many cases *styrene* tends to be in a separate small group with *ethylBenzene* (**Figures *Appendix_D_1A*, *1B, 1C*, *2B*, *3A*, *3C*, *4A, 4B, 4C***).
   b. *m/p-xylene* and *o-xylene* tend to have the most similar CPF patterns with merge with other aromatics to form a big group of aromatics.
   c. The locations of two important aromatic compounds, *benzene* and *toluene* are noted. The grouping of these two compounds varies. Mostly they are separated in different groups except at the Lynchburg Ferry site (**Figure *Appendix_D_4***). It seems more often that *toluene* shares common wind feature with other aromatic compounds for example at Channel View (**Figure *Appendix_D_1***) and *WallisVille* site (**Figure *Appendix_D_5***). At the Channel View site, *benzene* is more likely to be associated with *acetylene* but *toluene* shared feature with other aromatic compounds.

The grouping of species using the clustering analysis is useful in identifying possible sources. For example, the close similarity on the CPF of the *alkene* groups suggests a common source (*e.g.,* petroleum refinery) and the corresponding CPF plots indicate the direction from which these compounds came. The aromatic compound groups suggest solvent industry sources and the corresponding CPF may reveal the direction of major contributions to the elevated concentration. Still, the caveats previously mentioned regarding the interpretation of CPF plots are relevant. Different emission source categories (*e.g.,* petroleum refinery and solvent industry) may be geographically close to each other and may share similar transport pathways to the receptor sites. Indeed if one moves up to the higher level of grouping in the cluster tree, those smaller *alkene, alkane* and/or *aromatic compounds* groups mingle together to form large groups.

Potential Source Contribution Function (PSCF) Analysis

The CPF analysis is a useful tool to identify the general compass directions associated with elevated values of VOCs measured at each of the 5 auto-GC sites along the Ship Channel. In the following section we extend this concept using not only the local wind directions at each site, but the full set of points derived from back-trajectories for periods

of time associated with elevated VOC levels. The tool to link the trajectory information to the elevated values is called a 'potential source contribution function, or "PSCF."

The PSCF method originated from the residence time probability analysis (RTA) of (Ashbaugh et al., 1985). This is a method that combines chemical data with air parcel trajectories to identify source locations and preferred trajectory paths associated with elevated levels of atmospheric particulate trace elements and aerosols. The PSCF analysis associates each backward trajectory with an observation at the receptor site. The study domain is divided into a grid of many cells. By overlaying all the concentration-wind trajectories over these cells and noting which trajectories are associated with elevated values at the auto-GC at the time of the arrival of the trajectory, one obtains a map showing the frequency with which cells are associated with these elevated VOC levels as measured by the auto-GC sites. The PSCF value is the probability that an air parcel with a concentration higher than a some 'high' threshold value will arrive at the sampling site after residing passing over each cell on the defined grid. The PSCF value is calculated as the ratio between the numbers of trajectory arrivals to the auto-GC site that are associated with concentration greater than the specified threshold to the total number of trajectories arriving at the auto-GC site. The grid cells with high PSCF values are an indication of areas of high potential contributions to the concentrations at the sampling site. A PSCF value of unity at a given grid is interpreted to mean that the given grid contributed to the VOC measurement every time a high VOC value was recorded. A PSCF value of zero means that the given grid never contributed to a high value, at least in terms of a trajectory of air passing over the specified grid and later arriving at the measurement site.

The methods have been widely used in the identification of source locations and preferred transport pathways of atmospheric trace elements and particular species, e.g., sulphate, nitrate, ozone, black carbon, and mercury [Cheng et al., 1993*a* & *b*; Lin et al, 2001; Liu, et al, 2003; Poirot and Wishinski, 1986; Polissar et al, 1999; Polissar et al, 2001*a*; Stohl and Kromp-Kolb, 1994; Zeng and Hopke, 1989]. The method has also been applied to locate the sources or sources categories identified by multivariate receptor models, e.g., principal component analysis, positive matrix factorization [Xie et al, 1999b; Poirot et al, 2001; Polissar et al; 2001*b*]. We have recently used the PSCF methods on the size-segregated aerosol composition measured at Williams Tower during the Houston/southeastern field campaign [Disselkamp**,** et al, 2003].

Several steps are involved in calculating an array of PSCF values relating the passage of air over multiple points within a domain to elevated VOC values at a receptor site. First, back trajectories are calculated for each sample measured at the auto-GC sites, and the latitude/longitude time series of each such trajectories is converted to an (i,j) value that maps each point of the trajectory to a grid that overlays the study domain. With this information in hand, we can tally the total number of trajectories that pass over each grid (which can range from 0 to N, where N is the number of auto-GC samples). Second, we calculate the corresponding number of trajectories to pass over each grid that is associated with an elevated concentration measured at the auto-GC site. This ratio of the trajectories associated with elevated concentrations observed at the receptor site to the

total number of trajectories, is the conditional probability of air parcels carrying *elevated* concentration passing through a given grid cell. We defined *elevated* concentrations as the concentrations exceeding the 75% percentile of the concentrations measured during the sampling period.

It is tempting to interpret grids with high PSCF values as being the sole source for elevated concentrations at the auto-GC sites. However, a grid that is simply along a common pathway of those trajectories carrying elevated concentration to the auto-GC sites will also have a high PSCF value That is, if a trajectory starts at Point A at the edge of the domain and then passes over a large emission source at Point B, and finally arrives at an auto-GC site at Point C, then all points along path A-B-C will be equally implicated in contributing to the high VOC measurement at Point C for this one trajectory.

PSCF analysis has frequently been used by investigators to identify sources associated with long range transport. It has been less frequently applied to identify local sources [Hsu et al., 2003] because sources near the receptor sites can produce high concentrations regardless of the direction from which the air mass is coming resulting in a bias to local grids relative to more distant grids. Recognizing this potential difficulty, we used a very fine mesh grid for the study area with high temporal resolution wind field to define the trajectories. A grid was defined between 28.5° and 30.5° Latitude and -95.8° to -94.0° Longitude with a resolution of 0.02 in both latitude and longitude. This results in mesh with 101 by 91 cells. As a result of the observations previously provided by TCEQ, we were able to calculate trajectories 5 minute time steps. Therefore there were 73 end points for a given 6-hour back trajectory.

Two other steps were taken to avoid bias with local sources. An empirical weighting function was used to remove possible artifacts associated with the high PSCF values in grid cells where the total number of trajectories passing through is relatively small. After trying various weighting options we settled on a scheme in which the PSCF values for cells with fewer end points than a pre-defined threshold were down-weighted by the ratio of the actual number end points to the threshold value. The threshold value was defined as 3 times of the average number of end points in those grid cells passed by at least one trajectory. Finally, our calculations are only for those times when the wind was relatively steady as defined by the sinuosity factor, discussed earlier in this report, which represents the ratio of the straight-line transport distance to the actual travel distance of a simulated air parcel.

In addition to carrying out a PSCF calculation for each individual auto-GC site, an analysis was carried out based on the concentrations and wind trajectories at all five sites near Ship Channel. It is this analysis that potentially shows 'triangulation' information to identify specific locations that contribute to elevated VOC levels. We use the word 'potentially' because this information is available only if trajectories associated with high concentrations at different auto-GC stations intersect one another, thus identifying a specific grid/location. It is much more difficult to identifying specific points in the absence of such crossing trajectory patterns.

The PSCF analysis was applied to the VOC concentrations at the five Ship Channel sites. As was done in our previous analysis using CPF plots, this analysis has been partitioned into individual VOC species on a seasonal / diurnal basis. The complete set of PSCF results for all VOC species at both *nighttime* and *daytime* during each seasons have been included into six supplemented files (file name convention: ***Supplement_F_PSCF_[season]_[day or night].pdf***. The PSCF plots were compared with the CPF plots presented in another part of this report. In general the PSCF plots presented in this section of our final report agree with the corresponding CPF plots presented earlier.

Figure 16 shows examples of PSCF plots. These examples are for the summer observations made at night for *propylene* at each the five Ship Channel sites. The full range of PSCF values is shown in **Figure 16**. An abbreviated form is given in **Figure 17**, where we have color coded the results into three categories: The PSCF values smaller than 0.25 are lumped into a class with a constant value of 0 ('Low'), the PSCF values between 0.25 and 0.50 are lumped into a class with a constant value of 0.5 ('Medium') and those PSCF values larger than and equal to 0.5 are re-assigned with 1.0 ('High').

We illustrate the information contained in PSCF plots with a brief discussion of the results for *propylene* shown **Figures 16 and 17**. A surprising amount of variation is seen in the trajectories associated with high concentrations for each of the five Ship Channel auto-GC sites. The *Channel View* site has two well defined approach paths illustrated in the top row, 2nd panel of **Figures 16 and 17**, one from the northeast and a second from southeast. *Clinton* had a similar path from the northeast, but also recorded high values with more southerly flow (more south than the *Channel View* site). High values at *Haden Road* were associated with the paths shown from the southeast and south, whereas the Lynchburg Ferry site showed elevated values only for southerly flow. *WallisVille* showed a unique set of paths, with elevated values coming largely from the southwest.

Combining the results from all five of the Ship Channel sites into a single plot (labeled "*all-site* PSCF" and shown in the top left panel of these figures) we found a small area (pointed to by the pink arrow in this panel) that appears to be associated with high *propylene* levels at all of the sites. This grid is highlighted in all the other panels except that of the *Clinton* site, and could contain the source of *propylene* affecting the five Ship Channel sites. The comparable CPF plots for propylene at night, during the summer, are shown in **Figure 18**, and show the consistency noted between these two analyses.

As noted above, while PSCF plots for individual stations provide information on trajectory paths associated with elevated VOC levels at the auto-GC sites, they cannot identify individual grids associated with these concentrations. This information can only be provided by the 'all site PSCF' shown in the top left panel of each figure. The all-site PSCF plot reveals candidate sources common to all auto-GC sites if the major paths for the individual sites overlap and pass through common areas. Our point here is that it is necessary to check both the all-site PSCF plot and the individual site PSCF plot to evaluate possible source origins.

POSITIVE MATRIX FACTORIZATION (PMF) ANALYSIS

Multivariate Receptor Modeling with Positive Matrix Factorization

Multivariate Receptor modeling is a widely recognized technique developed that uses ambient air quality data measured at receptor sites to estimate emission profiles from a number of sources [Hopke 1991]. There are two principal approaches for receptor modeling. The Chemical Mass Balance (CMB) approach is used to estimate the contribution of each source to the measure pollution through regression assuming the number and the source composition profiles are known [Cooper, et al., 1984; Watson et al., 1984]. However in many cases source information is not available. Factor analysis based approaches are needed to extract the source information from the measured ambient data.

The fundamental assumptions for receptor modeling is that emission mass is conserved and that the composition of source emissions is unchanged with no transformation occurring during the transport from the source locations (where the source compositions are defined) to the receptor sites. The statistical model requires a series of samples with varying species concentrations measured at the receptor sites. A typical data set can be viewed as an n by m data matrix in which m chemical species are measured at n samples. Under the assumption of unchanging emission profiles from source to receptor locations, the measured concentration of each chemical species in each sample is the sum of the contributions of that species from all p possible sources. Mathematically the model can be written as:

$$X_{n\times m} = G_{n\times p} \bullet F_{p\times m}$$

where F is a p by m matrix describing source composition profiles for all p sources and the source composition profile for each source is actually the relative concentration of the m chemical species co-emitted. G is a matrix describing the mass contribution of all p sources to each individual samples; for each sources it is actually the contribution of the emission to the n samples in the time series. Generally the number of emission sources p is smaller than the number of samples n and the number of chemical species m. The source composition profiles are usually different among sources and are characterized by some unique tracers or key species. The goal of the multivariate receptor modeling is to identify the number of the sources p, the chemical composition profile of the sources, F and the amount of mass contributed by each source to each individual sample, G.

Principal component analysis (PCA) is widely used as a tool for multivariate receptor modeling. By applying the PCA to the measured ambient data matrix $X_{n\times m}$ it is possible to decompose the n by m data matrix $X_{n\times m}$ into a small set (say p) of factors:

$$X_{n\times m} = \hat{G}_{n\times p} \bullet \hat{F}_{p\times m} + E_{n\times m}\,.$$

It is usually assumed that the factor loading $\hat{F}$ and score matrices $\hat{G}$ are the estimates of the actual source composition profiles F and source mass contributions G, respectively. However this is rarely the case. The factor loading and score matrices from the PCA are neither non-negative guaranteed nor unique. Moreover the implicit scaling issue in the PCA leads a distortion in the data analysis. The lack of capacity to handle unreliable data like below detection limits values and missing values also make it inefficient in the utilization of data for the receptor modeling because the occurrence of missing values and values below the detection limits values are common in environmental measurements. For these reasons we have selected to use another, more robust tool in our analysis of the PAMS data.

Positive Matrix Factorization (PMF) is a novel and more powerful tool for receptor modeling [Paatero & Tapper, 1994; Paatero, 1997]. The advantages of the PMF relative to the more widely known PCA technique arises from its use of individual data point weighting as an optimal scaling method. The PMF gives the highest weights to the data in which there is the most confidence. Unreliable data, such as observations below detection limit or missing values, can be included in the analysis by giving them low weights to decrease their influence in the modeling. The common occurrence of extreme values in the heavy positive skewed distribution can also be handled by down-weighting those extreme points to reduce undue influence in the model. In addition, physically meaningful constraints such as non-negativities are an integral part of the method. The details of the algorithm implementation can be found in [Hopke, 2000; Paatero, 2000].

Receptor modeling with the PMF has the same goal as PCA which is to decompose the measured ambient data matrix $X_{n\times m}$ to identify i) the number of the sources p, and to estimate ii) the chemical composition profile of each source, F and iii) the amount of mass contributed by each source to each individual sample, G. PMF requires the measured ambient data matrix $X_{n\times m}$ and an additional uncertainty matrix $S_{n\times m}$ which quantifies of the uncertainty of each data point. The uncertainty matrix is used as a weight in the modeling process and the PMF solve the problem by minimizing the objective function:

$$Q(E) = \left\| \frac{X - GF}{S} \right\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{e_{ij}}{s_{ij}} \right]^2$$

with the constraints that G and F are non-negative.

The utilization of error estimates of data values as weights for optimal data scaling is the strengths of the PMF method and thus the results of the PMF rely on both the data and the uncertainty matrix. However estimating uncertainties for each data point is not trivial. Although analytical uncertainties sometimes may be available for some environmental measurements, those uncertainties only represent part of the unknown total uncertainty. Therefore, in the application of the PMF on the receptor modeling, the uncertainties of the data values are usually estimated empirically and a suitable error estimates is achieved through trial and error.

The results from a PMF analysis produce two matrices (similar to what is produced by PCA). Matrix *F* describes the source composition profiles and matrix *G* describes the mass contribution of each source to each sample. It is the responsibility of the data analyst to infer what source each PMF factor represents, that is, whether a given source is associated with power plant emissions or automotive exhaust. It is very important that the data analyst have information on the emissions and sources in the area, unique tracers or key species associated with each possible source and the relative compositions of the key species in the sources and the emission patterns. With this information, the data analyst looks for tracers, key species and composition relationship of species in each factor in the *F* matrix and inspects the temporal patterns of the corresponding factor in the *G* matrix to see if it matches any specific source in the area. The interpretation of the PMF results on the VOC data to infer the possible sources can be difficult because most VOCs come from multiple sources and the sources generally do not have unique tracers. When the mass contribution matrix *G* is in correct magnitude, the sum of the mass contribution from all the *p* sources to a sample will match the total mass measured on that sample. Since the *G* matrix from the PMF is in arbitrary magnitude, a multiple regression is usually carried out on the *G* matrix against the measured total mass (ppbC) to get the proper scaling factors for each factor. After normalized by the scaling factor, the *G* matrix will have the correct magnitude which in turn will let us evaluate the importance of each source to the total mass by estimating the average (or median) percent contribution of each source to the total mass.

The final source identification and appointment results from the PMF will thus include a) the number of PMF factors indicting the number of possible sources, b) an assessment of what source each factor represents, through the source composition profile represented in *F*, c) the temporal (diurnal) pattern of each source represented in *G*, and d) the average percentage of total mass each source emitted.

Some Technical Details on the Applications of the PMF

Uncertainty Estimates

To apply the PMF analysis on the data, an uncertainty matrix which has the same size as the data matrix needs to be created for the weighting purpose. However estimating uncertainties for each of the data values is not trivial. Although analytical uncertainties sometimes are available for environmental measurements, those uncertainties only represent part of the unknown total uncertainty [Polissar et al, 2001]. Therefore, in the application of the PMF on the receptor modeling, the uncertainties of the data values are usually estimated empirically and various estimate methods are proposed [Polissar et al, 1998; Huang et al., 1999; Qin et al, 2002; Kim et al., 2003; Hopke, 2000; Hopke et al, 2003; Paatero & Hopke, 2003; Xie et al., 1999a]. Most of the uncertainty estimate strategies and their successful applications of the PMF are used for source identification and apportionment of particulate matter. Recently, increasing application of the PMF on the VOCs data have been reported [Roberts, et al., 2004; Zhao et al., 2004; Miller et al., 2002] and error estimates similar to particulate matter have been used. Based on

previous works [Polissar et al, 2003; Lee et al., 2002; Poirot et al., 2001; Polissar et al, 1998; Polissar et al., 2001; Ramadan, et al., 2000; Song et al., 2001; Roberts et al., 2004], similar weighting scheme was used in this study. In general lower uncertainties were assigned to the data values above the detection limit. For values below detection limits, higher uncertainties were assigned. For missing values, the mean values were used as the data values and even higher uncertainties were assigned to them.

Number of Factors

There is no straightforward and universal way to determine the correct number of factors in the PMF analysis. The estimated objective function Q-value is usually checked against the theoretical Q values as a function of the number of factors. Since it is rare to have the well determined error estimates, the estimated Q values themselves generally do not provide much useful information in the determination of the number of the factors, but the change of the Q values usually do. Usually the changes of the Q values on the number of the factors are used to determine the appropriate number of the factors [Zhao et al., 2004; Polissar et al., 2001]. It has been suggested that the number of factors should be determined from multiple aspects of the PMF analysis [Hopke, 2000], including the facts that additional factor will not result in further significant improvement on the Q values, the desirable residual should have symmetric distributions and have the scaled residuals within ±3 standard deviations; no multiple solutions will be obtained for repetitions of the same set of parameters [Hopke, 2000]. The knowledge of possible sources is crucial for obtaining interpretative results from the PMF. In this study, the determination of the number of factors was determined by following this guidance.

Rotation Ambiguity

It is well-known that there is actually infinite possible mathematical valid solutions when using PCA-based approaches in receptor modeling [Henry, 1987], but there is only one *truth* in reality. Although the non-negative constraints embedded in the PMF methods can alle viate this problem, the non-negativity constraints in the PMF alone would normally not result in unique solution. The PMF software package provides several options to rotate the PMF results in order to get more meaningful solutions, which includes setting so-called Fpeak parameters and specifying elements in the so-called key G/F matrices [Paatero, 2000; Paatero et al., 2002]. In a VOC apportionment task like the one in this study, there are more species involved in the data analysis than those usually involved in the particulate matter source appointment. There is generally lack of information to set the elements in the so-called key F and Q matrices for specific species or samples, which leaves the use of the so-called Fpeak parameter as the only option. In this study, the Fpeak parameter was changed between -0.6 to 0.6, and the optimal Fpeak parameter was chosen based on the change on the objective functions and the interpretability of the solutions. It was found that the results with Fpeak = 0 were more interpretable in this study. Therefore the results presented in the Results Section were PMF solution corresponding to such a parameter value.

Data, Species and Samples Included from the PMF Analysis

One assumption in receptor modeling is that the mixing ratios of the sources do not change when air mass moves from the source to the receptor. To meet this assumption we have limited the PMF analysis to only the *nighttime* VOC observations.

The numbers of the missing values and the values below detection limit of the species differ from site to site. The inclusion of species with too many missing values / below detection limit values may jeopardize the PMF analysis and they are usually removed from the analysis [Paatero and Hopke, 2003]. For other species their removal from the analysis may result in the loss of too much information so they are included in the PMF factorization but assigned a relatively large uncertainty to limit their influences [Paatero and Hopke, 2003]. In **Table 9**, the total available number of *nighttime* samples of each species at each of the five Ship Channel sites as well as the number of missing and zero values are listed. If the percentage of the missing and zero values of a species at a site exceeds 63% (coded as 2 and marked by pink background), the species was excluded from the analysis. If the percentage of the missing and zero values of a species at a site between 50% and 63%, the species is considered as having a weak signal and was downweighted in the PMF analysis by doubling the error estimates (coded as 1 and marked by light blue background).

It was mentioned earlier in this report that discrepancies were observed between the measured total VOCs and the total VOCs obtained by summing the individual VOCs, although there was a great variation in such discrepancies among samples. In order to justify the use of measured total VOCs for estimating the scaling factors on the PMF factors, the difference between the measured and calculated total VOCs was defined as a hypothetical species called 'unidentified' that was included in the PMF analysis. However samples with too big deviations on the calculated total VOCs from the measured total VOCs were excluded from the PMF analysis. The final number of species and the number of *nighttime* samples used for the PMF analysis are summarized in **Table 10**.

Scaling of the Composition Profiles

It was mentioned that PMF does not give the absolute source composition profile and the results are actually uncertain relative to a multiplicative scaling coefficients given by

$$X_{ij} = \sum_{k=1}^{p} G_{ik} F_{kj} = \sum_{k=1}^{p} \left(\frac{G_{ik}}{s_k} \right) \left(s_k F_{kj} \right) = \sum_{k=1}^{p} G'_{ik} F'_{kj}, \quad i = 1,\ldots,n;\ j = 1,\ldots,m; \quad \begin{array}{l} n : \text{number of samples} \\ m : \text{number of elements} \\ p : \text{number of identified sources} \end{array}$$

However when the total VOC concentrations are available, the absolute source composition profiles could be derived through the regression of the PMF G factors on the measured total VOC concentrations. By regressing the PMF G factors onto the measured total VOC concentrations, the multiplicative scaling coefficients would be determined. In this study the measured TNMHC was used as the total VOC concentration in such a

regression step. It should be pointed out that negative regression coefficients make no sense in the receptor modeling which simply suggest the numbers of PMF factors are not reasonable. Therefore the regression of the PMF factors on the total measure VOC concentration put another important constraint on the determination of the number of factors.

Results of PMF Analysis and Discussion

PMF Analysis Results at Channel View Site

It was found that 7 PMF factors provided the most consistent and interpretable results for the VOC observations at the Channel View site. The results of the PMF factorization are presented in **Figure 19**. **Figure 19*a*** show the source composition profiles (mixing ratio) and **Figure 19*b*** show the percentage of VOC species variation expressed by each of the 7 PMF factor at Channel View site. In **Figure 19*a* and 19*b*** the factors were listed in the order from bottom to top, *i.e.,* factor 1 is in the very bottom. **Figure 19*c*** presents the contribution of the factors to the total *mean* (**Figure 19*c*** *left*) and *median* (**Figure 19*c*** *right*) VOC mass in pie charts. The scatter plot in **Figure 19*d*** shows the reconstructed total VOC mass by the PMF model versus the measured total mass. Very good total mass reproduction is obtained with this set of factors. The multiple determination R^2 is 0.976 indicating about 98% of data variation is expressed by the PMF model. Checking the scaled residuals also found that over 92% data points have scaled residual between ±2.25. The major wind directions carrying the elevated concentrations of the factors (75% percentile) to Channel View site are shown in **Figure 19*e***. The major species, the designation of the possible source of a PMF factor is based on both the source composition profile and the percentage variation of the species expressed by each PMF factor and is listed in **Table 11**.

Factor 1 consists primarily of *butanes, pentanes*, *C6-C8 paraffins* and aromatics such as *benzene, xylenes* and *toluene* (bottom of **Figure 19*a***). The substantial data variation of mid-weight C6-C8 paraffins, naphthene and aromatics such as *benzene, xylenes* and *toluene* are expressed by this factor (**Figure 19*b***). It is likely to be a mix of solvent chemical plant (C6-C8 paraffins and naphthene), petrochemical industry (aromatics) and fuel evaporation (*butanes* and *pentanes*). It contributes to 2.65% of the mean total VOC mass. This factor was predominately from the south-west and south-east confirming the influence of the Ship Channel region (**Figure 19*e***).

Factor 2 consists primarily of *ethene* and *propylene* as well as other light olefins like *1,3-butadiene, 1-butene, benzene* and *cis-2/trans-2-butenes* and *cis-2/trans-2-pentene.* Elevated alkenes levels especially *ethene* and *propylene* have been a persistent feature of the Houston area for an extended period of time. Tailpipe emissions usually have characteristic ratios of *ethene / acetylene* ranging from 1 to 3 and *propylene / acetylene* from 0.5 to 1.5 determined from airborne VOC measurements in urban cores of Nashville, TN, and Atlanta, GA in 1999, and in Houston and Dallas, TX in 2000 [Ryerson et al., 2003]. Numerous previous studies already found that elevated mixing ratios of *propylene* and *ethane* particular in Ship Channel area for over 20 years [Ryerson

et al., 2003]. The high ratios of *propylene / acetylene* and *ethene / acetylene* (7.1 and 5 for mean nighttime concentration respectively in this data set) are consistent with previously findings in resolved plumes from geographically isolated petrochemical sources [Ryerson et al., 2003]. The observed elevated alkene mixing ratio enhancements suggest that the on-road tailpipe emission were insignificant contributors to the them and therefore this factor is identified as a petrochemical emission source which has 17.21% contribution to the mean total VO mass. It should be pointed out that despite of the hundreds of VOC compounds known to be emitted from petrochemical facilities, *propylene* and *ethane* are the two alkenes make the largest contributions to the prompt O_3 production not only due to their own reactivities and rapidly decay in daytime atmospheric mixing ratios but also due to the photolysis reactivities of their photoxidation products [Ryerson et al, 2003]. **Figure 19*e*** shows that the major directions contributing to the elevated concentration of this factor is southeast.

Factor 3 consists primarily of heavy alkanes like *n-octane, n-nonane, n-decane, 2-methylheptane/methylheptane* and *n-hexane* as well as *xylene, toluene* and other substituted aromatics (**Figure 19*a* and *b***). Highly volatile alkanes such as *n-nonane, n-decane, n-undecane* with *trimethylbenzene, propylbenzene, ethylbenzene* and *xylenes* are caused by combustion process [Edwards et al., 2001] and C10-C11 heavy alkanes, *n-decane* and *n-undecane,* usually are considered as diesel markers [Watson et al., 2001; Brown and Halfne, 2003; Roberts et al, 2004]. *Benzene, toluene, ethylbenzene* and *xylenes* (BTEX) are usually considered from mobile sources and the ratios of BTEX were calculated for the comparison of mobile sources in different regions [Chiang et al., 1996; Ho and Lee, 2002]. The species *n-undecance* was excluded from the analysis at this site for overwhelming missing values. However we should note that *n-decane* is also abundant in solvents from paints and industrial uses, *n-nonane* shows enrichment in printing ink solvent from offset and rotogravure and *toluene* is used for thinning rotogravure inks [Watson et al., 2001]. The enrichment of aromatics may suggest industrial sources [Brown and Hafner, 2003; Roberts et al, 2004] such as chemical and refinery industries [Watson et al., 2001]. So this factor is identified as an industrial source with probable mixing of mobile sources. This factor constitutes 6.81% of the total VOC mass. **Figure 19*e*** shows that Factor 3 share similar contributing directions to the elevated concentration as Factor 1.

Factor 4 appears to be associated with vehicle exhaust emission due to the abundance of *acetylene* and *xylenes, toluene, substituted benzene* [Watson et al., 2001; Roberts et al., 2004], which contributes to 7.46% of total mean VOC mass. Quite little *benzene* was associated with this vehicle exhaust factor. In a recent EXPOLIS-Helsinki study [Edwards et al., 2001] it was found that the lack of *benzene* in the traffic emission sources was attributed to reductions in the *benzene* content of gasoline in Finland. Although earlier work also indicated that *ethene* and *acetylene* were evidence of flare emissions [Brown and Main, 2002] this factor was identified as a vehicle exhaust source. The directions shown in the CPF plots in **Figure 19*e*** suggest the influence of the highways on the west and south of the site.

Factor 5 is dominated by *ethane, propane, butanes* and *pentanes*. Light paraffins like *ethane* and *propane* are likely from aged background air since they tend to accumulate in urban atmosphere due to their low reactivities, but they are also prominent in emissions from natural gas usage [Brown and Hafner, 2003; and Roberts et al., 2004]. The enrichment of *butanes* and *pentanes* are likely from evaporative sources like liquid petroleum gasoline (LPG) or gasoline vapor leaks and evaporated gasoline usually do not consist of the combustion compounds as well as heavier hydrocarbons that volatilize more slowly and [Watson et al., 2001; Jorquera and Rappengluck, 2004]. Therefore this factor may represent a possible natural gas emission and fuel evaporation with possible aged air mass accumulation. This factor contributes about half (50.43%) of the mean total VOC mass at Channel View site. **Figure 19*e*** shows that the broad area to the south of the site contributing most to the elevated concentration of this factor.

The majority of *styrene* and substantial levels of *ethylbenzene* and some *benzene* are contained in Factor 6. The largest portion of *benzene* produced is used in the production of *ethylbenzene/styrene* and *ethylbenzene* is used almost exclusively to produce *styrene*. Most facilities produce both *ethylbenzene* and *styrene* and a patented method co-produces *ethylebenzene* and *styrene* [EPA, 1998]. Some *ethylbenzene* is used as a solvent (often replacing *xylenes*) and in the production of some dyes and *styrene* is abundant in solvents from paints, printing and industrial uses [Watson et al., 2001] and along with other *styrene* was systematically affected by petrochemical industry emissions [Jobson et al., 2004]. Factor 6 is identified as a petrochemical emissions and its mass contribution is 6.16%. The CPF plot in **Figure 19*e*** suggests that the major contributing sources to the elevated concentration of this factor are on the northeast of the site which needs to be confirmed.

Factor 7 consists primarily of *n-butane, n-pentane* and *iso-pentane* which are the major component of gas vapor. However this factor also expressed the majority of *pentenes* (*1-pentene, t2-/c2-pentene*), substantial *butenes* (*t2-/c2-butene, 1,3-butadiene*) as well as some C4/C5 paraffins. *T*2-/c2-*pentenes* were ascribed as contribution from traffic emission [Jorquera and Rappengluck, 2004]. In a recent study in comparing the ratios of *t2-pentene* versus *c2-pentene* and *t2-butene* versus *c2-butene* from measured ambient data to the ratios from tunnel test on the data collected from a site in Houston area (La Porte) it was suggested that the major source of these species was vehicle exhaust [Jobson et al., 2004]. However there exist many point sources of *butenes* and *pentenes* in the Ship Channel area and these species were ascribed to industrial sources [Roberts, et al., 2004] due to the lack of contribution from most aromatics normally appeared in the vehicle exhaust emission such as *benzene, toluene, ethylbenzene, xylenes, substituted benzene.* Thus this factor is identified as a mixture of fuel evaporative and industrial sources. The mass contribution of this factor is 9.0%. The CPF in **Figure 19*e*** indicates that southwest is the major contributing direction to the elevated concentration of this factor.

PMF Analysis Results at Clinton Site

It was found that 7 PMF factors provide most consistent and interpretable results for the VOC measured at *Clinton* site. The results of the PMF factorization are presented in **Figure 20**. **Figure 20*a*** show the source composition profiles (mixing ratio) and **Figure 20*b*** show the percentage of VOC species variation expressed by each of the 7 PMF factor at *Clinton* site. **Figure 20*c*** present the contribution of the factors to the total *mean* (**Figure 20*c*** *left*) and *median* (**Figure 20*c*** *right*) VOC mass in pie charts. The scatter plot in **Figure 20*d*** shows the reconstructed total VOC mass by the PMF model versus the measured total mass. Again good total mass reproduction is obtained with this set of factors. The multiple determination R^2 is 0.883 indicating about 88% of data variation is expressed by the PMF model. Checking the scaled residuals also found that over 98.7% data points have scaled residual between ±2.25. The major wind directions carrying the elevated concentrations of the factors (75% percentile) to *Clinton* site are shown in **Figure 20*e*.** The major species, the designation of the possible source of a PMF factor is based on both the source composition profile and the percentage variation of the species expressed by each PMF factor and is listed in **Table 12**.

Factor 1 consists primarily of *ethene* and *propylene* as well as *1-butene, 1,3-butadiene, benzene* together with some *isoprene.* This factor is similar to Factor 2 found in Channel View site which represents a petrochemical emission. The ratios of the these two light olefins to *acetylene* are 3.44 and 2.55 respectively based on the nighttime observations which are much higher than ratios expected from mobile emissions [Ryerson et al., 2003]. The CPF plot in **Figure 20*e*** shows that the contributing directions to the elevated concentration are from northeast, southeast and south. A similar factor found in this site from previous analyses but with smaller percent contribution to the total VOC concentration [Brown and Hafner, 2003; Roberts et al., 2004].

Factor 2 is similar to Factor 5 found in Channel View site and consists of *ethane, propane, butanes, pentanes* and alkanes like *cyclohexane, cyclopentane, methylcyclopentane, methylcyclohexane* etc. As discussed above this factor may represent a possible natural gas emission and fuel evaporation with possible aged air mass accumulation. This factor is mainly contributed from north-east direction and accounts for 17.62% of the men total VOC mass.

Factor 3 is largely *ethane, propane, butanes* and *pentane.* The existence of *ethane* and *propane* suggest the natural gas emission and/or accumulation of aged air mass. The enrichment of *butanes* and *pentanes* indicate the evaporation sources. However, the C4 olefins like *t2-/c2-butene, 1-butene,* and fair amount of *pentenes* (*1-peneten, t2-/c2-pentenes*) which indicates petrochemical / industrial emissions, are also enriched in this factor. At Channel View site, most of these C4 olefins are located in the same factor as *ethene* and *propylene,* a petrochemical source. Therefore this factor is identified as a petrochemical / industrial source mixed with accumulation / natural gas usage and evaporative emissions. The contribution of this factor to mean total VOC mass is 11.80% and southeast is the major direction contributing to the elevated concentrations.

Factor 4 is dominated by *butanes, pentanes,* and C5-C8 mid-weight alkanes, such as *n-hexane, n-heptane, cyclo-pentane, cyclo-hexane, 2methylhexane, 3methylHexane, methylcyclopentane, 2,2-dimethylButane* and *benzene.* It likely represents a mixture of solvent emission and fuel evaporation from both industrial facilities and oil / gas leaks and spills [Brown and Halfne, 2003; Roberts et al., 2004], similar to factor 1 at Channel View site. The mass contribution to the total VOC concentration of this factor is 16.23% and the major directions contributing to the elevated concentrations are northeast and south.

Factor 6 is enriched in C8 alkanes like *2methylheptane, 3methylheptane,* and other C6 to C10 alkanes *n-hexane, n-heptane, n-octane, n-nonane, n-decane* and is more likely from solvent and paint chemical emission [Roberts et al., 2004; Brown and Hafne 2003]. To some extent this factor is similar to Factor 3 at *Channel View*. The mass contribution of this factor to the total VOC concentration is 9.38% and the major contribution to the elevated concentrations are from southeast (**Figure 20*e***).

Factor 6 is characterized by *acetylene* and *benzene, xylene, toluene* and substituted *benzene*s which are characteristics of a vehicle exhaust emission [Watson et al., 2001]. This factor is similar to Factor 4 found at Channel View site. The mass contribution of this factor is 9.77% and major contributing directions are from northeast and southwest in a less extent.

Factor 7 is similar to the Channel View Factor 7 consisting mainly of *n-butane, n-pentane* and *iso-pentane* but is abundant with C5 olefins (*1-pentene, t2-/c2-pentene*) and paraffines (*n-/iso-pentanes*), branched pentanes (*2,2,4-trimethylpentane, 2,3,4-trimethylpentane*) and C4 olefins and paraffines (*t2-/c2-butene, n-butene, 1,3-butadiene*) which are identified as industrial sources [Roberts et al., 2004]. This factor again is identified as a mixture of fuel evaporative and industrial source. The percent contribution to the total mean VOC concentration of this factor is 14.29% and the major contributing directions are from south and northeast.

In recent data analyses on the VOC data collected during the TexAQS 2000 field campaign up to 11 and 15 factors have been resolved from the PMF analyses [Brown and Hafner, 2003; Roberts et al, 2004]. In those analyses three [Roberts et al., 2004] and eleven factors [Brown and Hafner, 2003] with less than 5% of the mass contributions were resolved. More factors were also tried in this analysis, and up to 9 factors were also resolved. However, it was found that the mass contributions of some contributing factors are quite small (around or below 1%) and the variation was big in multiple random runs when the number of the factors increased. On the other hand, the 7 factor solutions are quite consistent and seem robust.

PMF Analysis Results at Haden Road Site

It was found that 8 PMF factors provide most consistent and interpretable results for the VOC measured at *Haden Road* site. The results of the PMF factorization are presented in **Figure 21**. **Figure 21*a*** show the source composition profiles (mixing ratio) and

Figure 21*b* show the percentage of VOC species variation expressed by each of the 7 PMF factor at *Haden Road* site. **Figure 21*c*** present the contribution of the factors to the total *mean* (**Figure 21*c*** *left*) and *median* (**Figure 21*c*** *right*) VOC mass in pie charts. The scatter plot in **Figure 21*d*** shows the reconstructed total VOC mass by the PMF model versus the measured total mass. Very good total mass reproduction is obtained with this set of factors. The multiple determination R^2 is 0.881 indicating about 88% of data variation is expressed by the PMF model. Checking the scaled residuals also found that over 85% data points have scaled residual between ±2.25. The major wind directions carrying the elevated concentrations of the factors (75% percentile) to *Haden Road* site are shown in **Figure 21*e***. The major species, the designation of the possible source of a PMF factor is based on both the source composition profile and the percentage variation of the species expressed by each PMF factor and is listed in **Table 13**.

Factor 1 is enriched with C6-C11 alkanes specifically with high contributions from *n-hexane, n-heptane, n-octane, n-nonane, n-decane, n-undecane, 2methylhexane, 3methylhexane, 2methylHeptane, 3methylHeptane,* and naphthene *e.g., cyclo-pentane, cyclohexane, methylcyclopentane* and *methylcyclohexane.* This factor seems like the combination of Factors 4 and 5 found at the *Clinton* site and Factor 3 at Channel View site. These compounds are associated with solvent / paint emission and have a mass contribution factor is 23.29%. Mostly the elevated concentration of this factor is from south of the site (**Figure 21*e***).

Factor 2: *ethane, propane, butanes* and *pentane*, similar to Factor 5 at Channel View and Factor 2 at the *Clinton* site, ascribed as the a possible natural gas emission and fuel evaporation with possible aged air mass accumulation. It contributes 21.04% of the total mean VOC mass and major contributing directions to the elevated concentrations are from northeast and southeast.

Factor 3: enriched in light olefins, *i.e., ethene* and *propylene* typically from petrochemical emissions. This factor is similar to Factor 2 at Channel View and Factor 1 at *Clinton*. The mass contribution of this factor is 8.52% and the major contributing directions include northeast, southeast and south.

Factor 4 is a vehicle exhaust emission characterized by *acetylene* and aromatics which has a mass contribution of 1.22%. The major coming direction is northeast. A similar factor was identified at *Channel View* and *Clinton* sites as Factor 4 and Factor 6, respectively.

Factor 5 indicates another petrochemical emission source characterized by C4 olefins, mainly *1,3-butadiene* with some *1-butene, cis-2/trans-2-butene*. No such a factor was identified at Channel View and *Clinton* site. At the Channel View site, most of *1,3-butadiene* was lumped with *ethene* and *propylene* from the petrochemical sources and at *Clinton*, the majority *1,3-butadiene* was co-emitted in the vehicle exhaust factor. Although *1,3-butadiene* is a component of exhaust emission from both gasoline- and diesel-powered vehicles, it is not a constituent of the fuel but is produced by the incomplete combustion of olefins. It is also an important chemical in certain industrial

processes, particularly the manufacture of polymers including polybutadiene, styrene-butadiene rubbers and latexes, and nitrile-butadiene rubbers. It is unclear that this *1,3-butadiene* factor is from vehicle exhaust emission or on-site industrial usage. Considering the Ship Channel area is the largest manufacturers of *1,3-butadiene* in the US (http://www.nsc.org/library/chemical/13-Butad.htm) , this factor is most likely a industrial sources. It contributes to 11.19% of mean total VOC mass and the major contributing directions include northeast, southeast and southwest.

Factor 6 consists primarily of branched pentanes together with *n-butane* and *n-pentane*, which is likely fuel evaporative emission. The mass contribution of this factor is 4.56% and southwest is the major direction contributing to the elevated concentrations. No similar factor was resolved from the data at *Channel View* and *Clinton* sites.

Factor 7 consisted mainly of *n-butane, n-pentane* and *iso-pentane* and is enriched with C4/C5 olefins, *e.g., 1-pentenes, cis-2/trans-2-pentene, cis-2/trans-2-butene* and is similar to Factor 7 found in both *Channel View* and *Clinton*. It was identified as a mixture of fuel evaporative and industrial source and has 19.48% of mass contribution to the total mean VOC mass. The contributing direction of this factor is almost the same as the one for Factor 6.

Factor 8 was enriched with substituted aromatics like *1,2,4-timethylbenzene, 1,3,5-trimethylbenzne, p,o,m-ethyltoluene, n-/iso-propylbenzene* and is probably an industrial source as well as of vehicle emission influences. The emission from welding and printing industries involves the use of such heavy aromatics [Roberts et al., 2004]. This factor has a 10.69% contribution to the total mean VOC mass with major broad coming direction from south.

PMF Analysis Results at Lynchburg Ferry Site

Half of the samples collected at *Lynchburg Ferry* site did not have the corresponding total VOC concentrations. Although the PMF factorization does not need the total VOC concentration, this information is required to obtain the scaling factors by regressing the PMF G factors to the measured total VOC concentrations. Therefore, only those samples with available total VOC concentration were used in the PMF analyses. This resulted in a 6 factor PMF solution. **Figure 22*a*** show the source composition profiles (mixing ratio) and **Figure 22*b*** show the percentage of VOC species variation expressed by each of the 7 PMF factor at *Lynchburg Ferry* site. **Figure 22*c*** present the contribution of the factors to the total *mean* (**Figure 22*c*** *left*) and *median* (**Figure 22*c*** *right*) VOC mass in pie charts. The scatter plot in **Figure 22*d*** shows the reconstructed total VOC mass by the PMF model versus the measured total mass. Very good total mass reproduction is obtained with this set of factors. The multiple determination R^2 is 0.771 indicating about 77% of data variation is expressed by the PMF model. Checking the scaled residuals also found that over 92.4% data points have scaled residual between ±2.25. The major wind directions carrying the elevated concentrations of the factors (75% percentile) to *Lynchburg Ferry* site are shown in **Figure 22*e***. The major species, the designation of the possible source of a PMF factor is based on both the source composition profile and

the percentage variation of the species expressed by each PMF factor and is listed in **Table 14**.

Factor 1 was enriched with most of the aromatic compounds as well as long alkanes like *n-octane, n-nonane, n-decane* and *n-undecane.* The heavy alkanes and aromatics could be ascribed to solvent industries. This factor has some features of Factors 1 and 8 found at *Haden Road* site. The contribution of this factor to the total mean VOC is 8.66% and the major contributing direction is southwest.

Factor 2 was characterized by light olefins, *ethene, propylene, 1-butene, n-hexane* which is the same factor found in Channel View (Factor 2), *Clinton* (Factor 1) and *Haden Road* (Factor 3). It is from petrochemical production with a mass contribution of 21.76% to the total mean VOC. The major contributing direction is south.

Factor 3 was abundant in *butanes, pentanes,* other C6/C7 paraffins and naphthenes, *toluene* and *benzene*. This factor consists of likely fuel evaporative and solvent mission. The mass contribution of this factor is 25.22%. The major contributing direction is from southwest. It is noticed that most *styrene* variation has been expressed by this factor. A 7 factor PMF solution on the complete data set without deleting the samples without available total VOC concentration seemed splitting this factor into two factors, one of the split factor was dominated by *styrene* which accounted for about 17% of the total VOC concentration. Such a difference comes from the data difference used for the PMF factorization. By retaining all the available data, the extreme high values in the *styrene* concentrations were maintained. It should pointed out that one *styrene* observation at *Lynchburg Ferry* site represents the highest measured VOC concentration for all VOC species at all sites. Beside that extreme values there are several other high values in *styrene* measurements. Although the extreme values have been given much higher error estimate to minimize their influence in the PMF factorization in order to avoid distorted PMF factors, they affect the percent contribution calculation of the factors. The results shown in **Figure 22** were based on the tailored data set by deleting all the samples without available total VOCs and the extreme high values of *styrene* were deleted thereafter. This may explain why no separate *styrene* factor was resolved.

Similar to Factor 5 at Channel View site, Factor 2 at *Clinton* site and Factor 2 at *Haden Road* site, Factor 4 at the *Lynchburg Ferry* site, shown in **Figure 22*a* and 8*b***, is abundant with *ethane, propane, butanes* and *pentanes* which represents the accumulation of aged air mass or natural gas usage (*ethane* and *propane*) and fuel evaporation / refinery (*butanes* and *pentanes*). The majority variation of *acetylene* is expressed by this factor, suggesting the possible mixture with the vehicle exhaust emission. Increasing the number of factors did not resolve a separate vehicle exhaust factor featured by *acetylene* as the three sites discussed already. The mass contribution to the total mean VOC mass is 22.96% the major contributing direction is southwest.

Factor 5 mainly consists of *propane, n-butane* and *pentanes* but is enriched with C5 paraffins and olefins (*t2/c2-pentenes, 1-pentene, c2-/t2-butenes, butanes, pentanes*) as well as C6-C9 mid-weight alkanes. It is similar to Factor 7 at *Channel View*, *Clinton* and

Haden Road sites. It is likely a fuel evaporative emission but complicated by solvent chemical and petrochemical sources. The mass contribution of this factor is 6.96% and the major contributing direction to the elevated concentration is southwest.

Factor 6 was characterized by C4 olefins (*cis-2/trans-2-butenes, 1,3-butadienen, 1-butene*) and aromatics (*benzene, toluene* and *xylenes*). Substantial *isoprene* was also expressed by this factor. The light olefins are likely from petrochemical sources. *Isoprene* is likely to be either anthropogenic or biogenic sources, and aromatics compounds *benzene, toluene* and *xylenes* are probably from industrial / petrochemical sources. The percent contribution to the mean total VOC concentration is 14.44% and most elevated concentrations of this factor are from southwest direction.

PMF Analysis Results at WallisVille

At *WallisVille* site, the 7 factor solutions from PMF seem the most reliable and easily explained. **Figure 23*a*** show the source composition profiles (mixing ratio) and **Figure 23*b*** show the percentage of VOC species variation expressed by each of the 7 PMF factor at *WallisVille* site. **Figure 23*c*** present the contribution of the factors to the total *mean* (**Figure 23*c*** *left*) and *median* (**Figure 23*c*** *right*) VOC mass in pie charts. The scatter plot in **Figure 23*d*** shows the reconstructed total VOC mass by the PMF model versus the measured total mass. Very good total mass reproduction is obtained with this set of factors. The multiple determination R^2 is 0.944 indicating about 94% of data variation is expressed by the PMF model. Checking the scaled residuals also found that over 92.5% data points have scaled residual between ±2.25. The major wind directions carrying the elevated concentrations of the factors (75% percentile) to *WallisVille* site are shown in **Figure 23*e***. The major species, the designation of the possible source of a PMF factor is based on both the source composition profile and the percentage variation of the species expressed by each PMF factor and is listed in **Table 15**.

Factor 1 consists of abundant *ethane, propane, butanes, pentanes* and other C6-C8 alkanes which we think represent mixed sources such as accumulation of aged air or natural gas usage (*ethane* and *propane*), fuel evaporative / refinery (*butanes* and *pentanes*) and solvent emissions (C6-C8 alkanes). The contribution of this factor to the total mean VOC is 35.70% and the major coming direction for elevated concentration is northwest. This factor is similar to previous found Factor 5 at *Channel View*, Factor 2 at *Clinton*, Factor 2 at *Haden Road* and Factor 4 at *Lynchburg Ferry*. It is common for these factors to have enriched with *ethane, propane, butanes* and *pentanes*, but the other co-existing species differ from site to site.

Factor 2 was enriched with C6 to C9 alkanes such as *n-hexne, n-heptane, n-octane, n-nonane, 2methylheptane, 3methylheptane, toluene* and *xylenes,* and is identified as solvent / paint chemical source. This factor is similar to Factor 3 at Channel View, Factor 5 at *Clinton*, Factor 1 at *Haden Road* and Factor 1 at *Lynchburg Ferry* site. The mass contribution of this factor is 6.20% and the major contributing direction to the elevated concentration is southwest.

Factor 3 was fond to be abundant in *isoprene*, a representative species of biogenic source. The daytime maximum diurnal patterns of *isoprene* shown in **Figure 4*b*** indicates that biogenic *isoprene* exists in almost all the five sites under investigation (less evident for *Lynchburg Ferry* site). Since the PMF analyses were applied *only* on the nighttime data to which the biogenic emissions were not significant, no separate *isoprene* factor was resolved for the four sites discussed above. The furthest distance of the *WallisVille* site away from the industrial and chemical facilities along Ship Channel and the closer to the northern hardwood forest than the other sites may be the reason to have separate *isoprene* factor resolved from the data. The biogenic emission has been the subject of research for many studies [Wiedinmyer et al., 2001; Wert et al., 2003; Ryerson et al., 2003; Vizuete et al., 2002] and the *isoprene* from forested region northeast of Houston could be an important sources of VOC in the urban area [Kleinman et al., 2002; Vizuete et al., 2002]. The major contributing direction from northeast may suggest that the biogenic sources of this *isoprene* factor. The mass contribution of this biogenic source to the total mean VOC concentration is 3.87%. It was claimed that the emissions from biogenic sources have been most likely underestimated in ozone modeling study and suggested that a significant role of biogenic hydrocarbon emissions in many urban and suburban locations in ozone formation [Choi and Ehrman, 2004; Chameides et al., 1988, 1992; Cardeline and Chameides, 2000]. Indeed in these five sites under investigation, there is a biogenic source contribution which is not resolved from the PMF analyses (except *WallisVille* site) because of the use of nighttime observation alone.

Factor 4 mainly consists of *ethane, propane* which could be ascribed to the possible natural gas emission and aged air mass accumulation, but this factor is enriched with *cis-2/trans-2-butene* which suggests an industrial source. There is also substantial of *acetylene* existing in this factor. Considering the abundant *acetylene* in the vehicle exhaust emission identified in the next factor (Factor 5), it is suggested that *acetylene* in this factor may be associated with industrial source. Therefore this factor is identified as an industrial source probably mixing with accumulation and natural gas emission. The mass contribution of this factor is 10.09% and the major contributing direction is northeast.

The coexisting of *acetylene* and aromatic compounds like *benzene, toluene* and *xylenes in* Factor 5 suggests a vehicle exhaust emission source which is similar to Factor 4 at *Channel View*, Factor 6 at *Clinton* and Factor 4 at *Haden Road*. The substantial contribution of *butane* and *pentanes* indicates the factor is mixed with fuel evaporative emissions. The mass contribution of this factor is 10.04% and the major contributing direction is from southwest.

The major constituents such as *n-butane, n-pentane* and *iso-pentane* and the enrichment of species like *t2-/c2-pentene* make this factor similar to Factor 7 at Channel View, *Clinton* and *Haden Road* sites, which represents a mixture of fuel evaporative and industrial source. However the co-existence of heavy alkanes (*n-decane* and *n-undecane*) and aromatics suggests the contribution of other chemical industrial source (or a diesel exhaust). The mass contribution of this factor is 2.18% and the contributing direction is southwest and southeast.

The enrichment of *ethene, propylene, 1-butene* and *1,3-butadiene* in Factor 7 clearly demonstrates a petrochemical sources. The mass contribution of this factor is 31.92% and the major contributing direction to the elevated concentration is southwest.

Discussion

Table 16 summarizes the PMF results of all five sites close to Ship Channel which are presented in **Tables 11 to 15** separately. It was found from **Figures 19 to 23** and **Tables 11 to 15** that some common factors were resolved for the five sites under investigated. One petrochemical emission dominated by *ethene* and *propylene* were found for all five sites. *1-butene*, *1,3-butadiene* and sometime other C4 olefins were also grouped into this factor. The major contributing wind directions revealed by the CPF plots were summarized in **Figure 24*a*** which indicates that the vast refinery, petrochemical and industrial facilities along Ship Channel are responsible for the emission of this category. The contribution of this category to the total mean VOC concentration measured in the sites ranged from 8.5% at *Haden Road* site to about 20% at *Channel View*, *Clinton* and *Lynchburg Ferry* sites to 31.52% at *WallisVille* site. Overall, this category of emission constitutes the second largest contribution to the ambient VOC at these sites.

Another common source is dominated by *ethane, propane, butanes* and *pentanes* and was identified as sources from natural gas and evaporation. The major contributing directions were revealed by the CPF plots shown in **Figure 24*b***. It is should be noted that the major contributing direction at *Clinton*, *Haden Road* and *WallisVille* sites is the northeast, away from the Ship Channel. The mass contribution of this category ranged from about 20% at *Clinton*, *Haden Road* and *Lynchburg Ferry* to 36% at *WallisVille* to 50% at Channel View site, demonstrating being the largest contributor to the VOC measured in these sites.

The other common category of sources is characterized by the abundance of mid-weight alkanes and aromatics and is identified as from solvent, paint industry. The contributing directions of this category are disclosed by the CPF plots presented in **Figure 24*c***. The major contributing directions of this category of emission at all sites are from south. The mass contributions are generally less than 10% except at *Haden Road* (23.29%). As shown in **Figure 21*b*** (bottom), more mid-weight alkanes are enriched in this factor. As discussed above this factor at *Haden Road* seems like a combination of Factors 4 and 5 at *Clinton* site, which may explain the bigger percent contribution at *Haden Road*.

Vehicle exhaust is another common factor identified in these sites except at *Lynchburg Ferry*. At *Lynchburg Ferry* site, most of *acetylene* shows up in the accumulation / natural gas and evaporative source category characterized with the low C alkanes. **Figure 24*d*** shows the major contributing direction of vehicle exhaust and most of the lobes in the CPF plots are pointing to the direction of nearby highways. The percent contribution to the total mean VOC measured ranges from 1.22% at *Haden Road* site to a little bit less than or about 10% at other sites.

Another common factor found is characterized by *n-butane, n-pentane* and *iso-pentane* and the factor is enriched with C4/C5 especially C5 *pentenes*. Since *pentane*s are markers of gasoline vapors this factor is identified as fuel evaporation. In previous research a similar factor enriched with C4/C5 olefins and paraffins was identified and indicated an industrial source [Roberts et al., 2003]. Such factors resolved from data at Channel View, *Clinton* and *Haden Road* sites are quite similar to one another (**Figure 19 to Figure 21**). At *Lynchurg Ferry* site, the existence of branched pentanes is enriched in this factor and at *WallisVille* site this factor is complicated by the enrichment of aromatic compounds. The CPF plots of **Figure 24*e*** shows the major directions contributing to the elevated concentration of this factor. Except these is a lobe in the CPF plot at *Clinton* site points to northeast direction, all other lobes at all sites are pointing to the south part of the sites. The contributions of this category to the total mean VOC measured at the sites range from 2% at *WallisVille* site to 19% at *Haden Road* site.

One factor enriched with *n-hexane*, *pentanes* was identified at *Channel View* (Factor 1) as another solvent/paint industrial source. A similar factor was found at *Clinton* (Factor 4) but without enriched aromatics. Factor 3 at *Lynchburg Ferry* shares some similarity to this factor but the greater enrichment of *butanes* and *pentanes* may suggest the mixing with fuel evaporative source of this factor at *Lynchburg Ferry*.

A less well identified source category is represented by Factor 3 at *Clinton*, Factor 6 at *Lynchburg Ferry* and Factor 4 at *WallisVille* site. A common feature of these factors is the enrichment of *t2-/c2-butene* and sometime *1-butene*. The major compositions of the factors differ. At the *Clinton* site, the factor was enriched with *ethane, propane, butanes* and *pentanes*, but at *Lynchburg Ferry*, it was found to be enriched with *benzene* and *toluene* additionally, which may suggest that the industrial emission (*t2-/c2-butene*) are mixed with fuel evaporative emission or petrochemical sources. For *WallisVille*, the major component of *ethane* and *propane* may suggest the industrial emission (*t2-/c2-butene*) mixed with aged air or natural gas emission.

The factor enriched with *styrene* and *ethylbenzene* at *Channel View* site (Factor 6) is unique and no similar factor was found at other sites. The *1,3-butadiene* enriched factor (Factor 5), the factor with abundant of *pentanes* and branched *pentanes* (Factor 6) and a factor consisting of only enriched aromatic compounds (Factor 8) at *Haden Road* site are also unique in the PMF analyses of the sites. The biogenic category characterized with enriched *isoprene* at *WallisVille* shows another unique source category resolved from the nighttime VOC data. At other sites *isoprene* was mixed into other source categories and no separate biogenic source was identified. Although the mass contribution of the biogenic source identified at *WallisVille* site is about 4% and no separate biogenic sources were identified for the other four sites, the contributions of the biogenic sources should not be zero (for other sites) and should be greater than 4% (for *WallisVille* site) considering the biogenic *isoprene* behavior revealed by the diurnal patterns in **Figure 4*b***.

REFERENCES

Ashbaugh, L.L., Malm, W.C., Sadeh, W.Z., *A residence time probability analysis of sulfur concentrations at ground canyon national park,* **Atmospheric Environment,** 19(8), 1263-1270 (1985).

Atkinson, R. 2000, *Atmospheric chemistry of VOCs and NOx*, **Atmospheric Environment**, 34, 2063-2102.

Disselkamp, R.S., Xie, Y.L., Shutthanandan, V., Barrie, L.A., Berkowitz, C.M., Cliff, S.S., Cahill, T.A., *Analysis of Size-Segregated Aerosol Compositional Data Collected as part of the Texas 2000 Air Quality*, **Atmospheric Environment**, Submitted, October 2003

Begum, Bilkis, Kim, E., Biswas, S.K., and Hopke, P.K., *Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh*, **Atmospheric Environment,** 38, 3025-3038 (2004).

Brown, S. G. and Hafner, H. R., *Exploratory Source Apportionment of Houston's Clinton Drive Auto-GC 1998-2001 Data,* STI 900700-2317-FR, March 2003

Brown, S. G. and Main, H.H., *Acquisition, review, and analysis of auto-GC VOC data in Houston area, 1998-2001,* **STI 900670-2224-FR**, July 2002

Cardeline, C.A. and Chameides, W.L. 2000, *The application of data from photochemical assessment monitoring stations to the observations-based model*, **Atmospheric Environment** 34, 2325-2332

Chameides, W.L.,, Fehsenfekd, D., Rodgers, M.O., Cardelino, C., Martinez, J., Parrish, D., Lonneman, W., Lawson, D.R., Ramussen, R.A., Zimmerman, P., Greenberg, J., Middleton, P., Wang, T., 1992, *Ozone precursor relationships in the ambient atmosphere,* **J Geophys. Res** 97, 6037-6055,

Chameides, W.L., Lindsay, R.W., Richardson, J., Kiang, C.S., 1988, *The role of biogenic hydrocarbon in urban photochemical smog: Atlanta as a case study*. **Science** 241, 1473-1475

Cheng, M.D., Hopke, P.K., Zeng, Y., 1993*a*, *A receptor-oriented methodology for determining source regions of particulate sulfate at Dorset, Ontario,* **Journal of Geophysical Research-Atmosphere**, 98, 16839-16849.

Cheng, M.D., Hopke, P.K., L. Barrie, A. Ripple, M. Olson, S. Landsberger**,** 1993*b*, *Qualitative determination of source regions of aerosol in Canadian high Arctic,* **Environmental Science and Technology,** 27, 2063.

Chiang, P.C., Ching, Y.C., Ching, E.E., Chang, S.C., *Characterization of hazardous air pollution emitted from motor vehicles,* **Toxicol. Environ. Chem.,** 1996, 56, 85-104,

Choi, Yu-Jin and Ehrman, Sheryl H., *Investigation of source of volatile organic carbon in the Baltimore area using highly time-resolved measurements*, **Atmospheric Environment** 38, 775-791(2004)

Cooper, J.A., Watson, J.G., Huntzicker, J.J., **Atmospheric Environment**, 1984, 8, 1347-1355.

Edwards, Rufus D. ; Jurvelin, J. ; Koistinen, K. ; Saarela, K. ; Jantunen, M., *VOC source identification from personal and residential indoor, outdoor and workplace microenvironment samples in EXPOLIS-Helsinki, Finland*, **Atmospheric Environment** Volume: 35, Issue: 28, October, 2001, pp. 4829-4841

Environment Protection Agency (EPA), *Locating and Estimating Air Emissions from Sources of Benzene,* **EPA-454/R-98-011,** June 1998.

Fujita, E. M. and D. E. Campbell, *Validation and application protocol for surface apportionment of photochemical assessment monitoring stations (PAMS) ambient volatile organic compound (VOC) data,* **Report to EPA #GR826237-01-0,** August 2003

Hartigan, J.A., 1975, Clustering Algorithms, New York, John Wiley & Sons, Inc.

Henry, RC. *Current factor analysis receptor models are ill-posed.* **Atmospheric Environment** 21, 1815-1820. 1987.

Ho, K.F. ; Lee, S.C. ; Chiu, Gloria M.Y., *Characterization of selected volatile organic compounds, polycyclic aromatic hydrocarbons and carbonyl compounds at a roadside monitoring station*, **Atmospheric Environment** Volume: 36, Issue: 1, January, 2002, pp. 57-65

Ho, K.F. ; Lee, S.C., *Identification of atmospheric volatile organic compounds (VOCs), polycyclic aromatic hydrocarbons (PAHs) and carbonyl compounds in Hong Kon*g, **The Science of the Total Environment** Volume: 289, Issue: 1-3, April 22, 2002, pp. 145-158

Hopke, P.K., **Receptor Modeling in Environment Chemistry,** Wiley, New York, 1985.

Hopke, P.K., 1991, **Receptor Modeling for Air Quality Management,** Elsevier, Amsterdam, The Netherlands.

Hopke, P.K., 2003, *Recent development in receptor modeling,* **Journal of Chemometrics, 17,**12.

Hopke, P.K., 2000, *A Guide to Positive Matrix Factorization***,** EPA WORKSHOP PROCEEDINGS *Materials from the Workshop on UNMIX and PMF as Applied to PM2.5*, February 14-16, 2000

Huang,SL, K A Rahn, R Arimoto, 1999**,** *Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode Island:* **Atmospheric Environment**, v. 33, p. 2169-2185.

Hsu, Ying-Kuang, T. M. Holsen and P. K. Hopke, 2003, *Comparison of hybrid receptor models to locate PCB sources in Chicago,* **Atmospheric Environment**, 37(4) 545-562.

http://www.ga.gov.au/nmd/geodesy/datums/distance.jsp

Jobson, B. T., C. M. Berkowitz, W. C. Kuster, P. D. Goldan, E. J. Williams, F. C. Fesenfeld, E. C. Apel, T. Karl, W. A. Lonneman, and D. Riemer (2004), *Hydrocarbon source signatures in Houston, Texas: Influence of the petrochemical industry*, **J. Geophys. Res**., *109*, D24305

Jorquera, Héctor ; Rappenglück, Bernhard, *Receptor modeling of ambient VOC at Santiago, Chile*, **Atmospheric Environment** Volume: 38, Issue: 25, August, 2004, pp. 4243-4263

Kim, E., Hopke, P.K., Edgerton, E., *Source identification of Atlanta aerosol by positive matrix factorization,* **Journal of the Air and Waste Management Association,** 53, 731-739 (2003*a*).

Kim, E., Larson, T.V., Hopke, P.K., Slaughter, C., Sheppard, L.E., Claiborne, C., *Source identification of PM2,5 in an arid northwest US city by positive matrix factorization,* **Atmospheric Research,** 66, 291-305, 2003*b*.

Kleinman, L.I., Daum, P.H., Imre, D., Lee, Y.N., Nunnemacker, L.J., Springston, S.R., Weinstein-Lloyd, J., and Rudolph, J., *Ozone production rate and hydrocarbon reactivity in 5 urban areas: A cause of high ozone concentration in Houston,* **Geophys. Res. Lett.,** 29(10), 1467, doi:10.1029/2001GL014569, 2002.

Lwis,C.W., R.C. Henry and Jack H. Shreffler, *An exploratory look at Hydrocarbon data from the photochemical assessment monitoring stations network;* **Journal of the Air & Waste Management Association***,* 48, 71-76 (1998).

Lin, Che-Jen; Cheng, Meng-Dawn; Schroeder, William H., 2001, *Transport patterns and potential sources of total gaseous mercury measured in Canadian high Arctic in 1995***; Atmospheric Environment** Volume: 35(6), 1141-1154.

Liu, W., Hopke, P.K., Han, Y.J., Yi, S.M., Holsen, T.M., Cybart, S., Kozlowski, K. and Milligan, M., 2003, *Application of receptor modeling to atmospheric constituents at Potsdam and Stockton, NY,* **Atmospheric Environment**, 37(36), 4997-5007.

Liu, W., P.K.Hopke, Y-J Han, S-M Yi, T.M. Holsen, S. Cybart, K.Kozlowski, M.Milligan, *Application of receptor modeling to atmospheric constituents at Potsdam and Stockton, NY,* **Atmospheric Environment** 37, 4997-5007, 2003.

Paatero,P, U Tapper. *Analysis of different modes of factor analysis at least squares fit problems.* **Chemometrics and Intelligent Laboratory Systems***,*18, 183-194. 1993.

Paatero,P, U Tapper. *Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values.* **Environmetrics***,* 5, 111-126. 1994.

Paatero, P., P.K.Hopke, 2003, *Discarding or downweighting high-noise variables in factor analytic models,* **Analytica Chimica Acta***,* 490, 277-289.

Paatero,P, *Least squares formulation of robust non-negative factor analysis***, Chemometrics and Intelligent Laboratory Systems,**37, 23-35, 1997.

Paatero, P., 2000, *User's Guide for Positive Matrix Factorization programs PMF2 and PMF3***.**

Paatero,P, P.K.Hopke, X.H.Song, Z. Ramadan, 2002, *Understanding and controlling rotations in factor analytic models***, Chemometrics and Intelligent Laboratory Systems***,*60, 253-264.

Roberts, Paul T., Brown, Steven G., Reid, Stephen B., Buhr, Martin P., Funk, Tami H., Stiefer, Patricia S., *Emission inventory evaluation and reconciliation in the Houston-Galveston area,* **STI-903640-2490-FR,** March 19, 2004.

Ryerson, T. B.; Trainer, M.; Angevine, W. M.; Brock, C. A.; Dissly, R. W.; Fehsenfeld, F. C.; Frost, G. J.; Goldan, P. D.; Holloway, J. S.; Hübler, G.; Jakoubek, R. O.; Kuster, W. C.; Neuman, J. A.; Nicks, D. K.; Parrish, D. D.; Roberts, J. M.; Sueper, D. T.; Atlas, E. L.; Donnelly, S. G.; Flocke, F.; Fried, A.; Potter, W. T.; Schauffler, S.; Stroud, V.; Weinheimer, A. J.; Wert, B. P.; Wiedinmyer, C.; Alvarez, R. J.; Banta, R. M.; Darby, L. S.; Senff, C. J., *Effect of petrochemical industrial emissions of reactive alkenes and NO_x on tropospheric ozone formation in Houston*, *Texas*, **J. Geophys. Res.** *Vol. 108 No. D8,* 10.1029/2002JD003070, 25 April 2003

Poirot, R.L., Wishinski, P.R., 1986, *Visibility, sulfate and air mass history associated with the summertime aerosol in northern Vermont,* **Atmospheric Environment**, 20, 1457.

Poirot, R.L., Wishinski, P.R., Hopke, P.K., Polissar, A.V., 2001, *Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont*, **Environmental Science and Technology***,* 35, 4622-4636.

Polissar, A.V., Hopke, P.K., Harris, J.M., 2001*a, Source regions for atmospheric aerosol measured at Barrow, Alaska,* **Environmental Science and Technology**, 35, 4214-4226

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001*b*, *Atmospheric aerosol over Vermont: chemical composition and sources,* **Environmental Science and Technology,** 35, 4604-4621.

Polissar,AV, P K Hopke, P Paatero, 1998, *Atmospheric aerosol over Alaska - 2. Elemental composition and sources:* **Journal of Geophysical Research-Atmospheres,** v. 103, p. 19045-19057.

Polissar,AV, P K Hopke, P Paatero, Y J Kaufmann, D K Hall, B A Bodhaine, E G Dutton, J M Harris, 1999, *The aerosol at Barrow, Alaska: long-term trends and source locations:* **Atmospheric Environment,** v. 33, p. 2441-2458.

Polissar, A.V., P.K.Hopke, R.L. Poirot, *Atmosperic Aerosol over Vermont: Chemical Composition and Sources,* **Environ. Sci. Technol.**, 35, 4604-4621, 2001.

Qin, Y., K. Oduyemi, L.Y.Chan, 2002, *Comparative testing of PMF and CFA models,* **Chemometrics and Intelligent Laboratory Systems,** 61, 75-87..

Ramadan,Z, X H Song, P K Hopke, 2000, *Identification of sources of Phoenix aerosol byositive matrix factorization:* **Journal of the Air & Waste Management Association,** v. 50, p. 1308-1320.

Song, X.H., A.V.Polissar, P.K.Hopke, *Sources of fine particle composition in the northeastern US,* **Atmospheric Environment,** 35, 5277-5286, 2001.

Stohl, A., Kromp-Kolb, H., 1994, *Origin of ozone in Vienna and surroundings, Austria,* **Atmospheric Environment**, 28, 1255-1266.

Stohl, A**.,** 1996, *Trajectory statistics-a new method to establish source-receptor relationship of air pollutants and its application to the transport of particulate sulfate in Europe,* **Atmospheric Environment,** 30, 579.

Texas Natural Resource Conservation Commission (TNRCC), **Ozone Air Pollution Facts**, GI-273, Feb., 2001.

Vizuete, W., Junquer, V., McDonald-Buller, E., McGaughey, G., Yarwood, G., and Allen, D.T., *Effects of Temperature and Land Use on Predictions of Biogenic Emissions in Eastern Texas,* **Atmospheric Environment,** 36, 3321-3337, 2002

Watson, John G.; Chow, Judith C.; Fujita, Eric M. *Review of volatile organic compound source apportionment by chemical mass balanc*e, **Atmospheric Environment** Volume: 35, Issue: 9, March, 2001, pp. 1567-1584.

Watson, J.G.; Cooper, J.A.; and Huntzicker, J.J. (1984). *The effective variance weighting for least squares calculations applied to the mass balance receptor model.* **Atmos. Environ**., 18(7):1347-1355.

Wiedinmyer, C., Friedfeld, S., Baugh, W., Greenberg, J., Guenther, Alex, Fraser, Matthew and Allen, David, *Measurement and analysis of atmospheric concentrations of isoprene and its reaction products in central Texas,* **Atmospheric Environment**, 35, 1001-1013, 2001

Wert, B.P., Traine, M. , Fried, A., Ryerson, T.B., Henry, B. , Potter, W. Angevine, W.M., Atlas, E., Donnelly, S.G., Fehsenfeld, F.C., Frost, G.J., Golden, P.D., Hansel, A., Holloway, J.S., Hubler, G., Kuster, W., Nicks Jr., D.K., Neuman, J.A., Parrish, D.D. , Schauffler, S. , Stutz, J. , Sueper, D.T., Widinmyer, C. and Wisthaler, A. , *Signatures of terminal alkene oxidation in airbone formaldehyde measurements during TexAQS 2000*, **Journal of Geophysical Research**, 108, D3, 4104-, 2003.

Xie, Y.L, P K Hopke, P Paatero, L A Barrie, S M Li, 1999a**,** *Identification of source nature and seasonal variations of arctic aerosol by positive matrix factorization:* **Journal of the Atmospheric Sciences,** v. 56, p. 249-260.

Xie, Y.L, P K Hopke, P Paatero, L A Barrie, S M Li, 1999b**,** *Locations and preferred pathways of possible sources of Arctic aerosol*: **Atmospheric Environment**, v. 33, p. 2229-2239.

Zhao, W., Hopke, P.K., and Karl, T., *Source identification of volatile organic compounds in Houston,* **Texas, Environ. Sci Technol.,** 38, 1338-1347(2004).

Zeng, Y. and P.K.Hopke, 1989, *A study of the sources of acid precipitation in Ontario, Canada,* **Atmospheric Environment**, 23, 1499-1509

Table 1. Information of the 9 auto-GC sites in the Houston-Galveston area

Site Abbr.	Site Name	Site Code	Latitude	Longitude
Danciger	Danciger C618	480390618	29.148889	-95.765
Mustang	Mustang Bayou C619	480390619	29.313611	-95.201389
LakeJackson	Lake Jackson C1016	480391016	29.043611	-95.472778
TexasCity	Texas City 34Th St. C620	481670056	29.402222	-94.946389
ChannelView	ChannelView C15/A115	482010026	29.8025	-95.125556
Wallisville	Wallisville Road C617	482010617	29.821389	-94.99
HadenRoad	HRM-3 Haden Road C603/A114	482010803	29.765278	-95.181111
LynFerry	Lynchburg Ferry C1015/A165	482011015	29.764444	-95.078056
Clinton	Clinton C403/C304/A113	482011035	29.733889	-95.2575

Table 2. Name and codes of the measured hydrocarbons in the auto-GC sites at Houston-Galveston area

field index[a]	agc code	abbre	name	MW[b]	K_{OH}[c]	MIR[d]	Formula	Lifetime (hr)
6	43202	ethane	ETHANE	30.07	0.27	0.08	C2H6	207.3
7	43203	ethene	ETHYLENE	28.05	8.52	2.16	C2H4	6.52
8	43204	n_prop	PROPANE	44.1	1.15	0.15	C3H8	48.31
9	43205	prope	PROPYLENE	42.08	26.3	2.75	C3H6	2.11
10	43206	acetyl	ACETYLENE	26.04	0.9	0.14	C2H2	61.73
11	43212	n_buta	N-BUTANE	58.12	2.54	0.31	C4H10	21.87
12	43214	i_buta	ISO-BUTANE	58.12	2.34	0.37	C4H10	23.74
13	43216	t2bute	TRANS-2-BUTENE	56.11	64	2.92	C4H8	0.87
14	43217	c2bute	CIS-2-BUTENE	56.11	56.4	2.92	C4H8	0.99
15	43218	bute13	1,3-BUTADIENE	54.09		13.5	C4H6	
16	43220	n_pent	N-PENTANE	72.15	3.94	0.31	C5H12	14.1
17	43221	ipenta	ISO-PENTANE	72.15	3.9	0.41	C5H12	14.25
18	43224	pente1	1-PENTENE	70.13	31.4	1.81	C5H10	1.77
19	43226	t2pene	TRANS-2-PENTENE	70.13	67	2.57	C5H10	0.83
20	43227	c2pene	CIS-2-PENTENE	70.13	65	2.57	C5H10	0.85
21	43230	pena3m	3-METHYLPENTANE	86.17	5.7	0.45	C6H14	9.75
22	43231	n_hex	N-HEXANE	86.17	5.61	0.98	C6H14	9.9
23	43232	n_hept	N-HEPTANE	100.2	7.15	0.24	C7H16	7.77
24	43233	n_oct	N-OCTANE	114.22	8.68	0.18	C8H18	6.4
25	43235	n_non	N-NONANE	128.26	10.2	0.16	C9H20	5.45
26	43238	n_dec	N-DECANE	142.29	11.6	0.17	C10H22	4.79
27	43242	cpenta	CYCLOPENTANE	70.13	5.16	0.7	C5H10	10.77
28	43243	i_pren	ISOPRENE	68.11	101	2.58	C5H8	0.55
29	43244	bu22dm	2,2-DIMETHYLBUTANE	86.17	2.32	0.25	C6H14	23.95
30	43247	pen24m	2,4-DIMETHYLPENTANE	100.2	5.1	0.45	C7H16	10.89
31	43248	cyhexa	CYCLOHEXANE	84.16	7.49	0.37	C6H12	7.42
32	43249	hexa3m	3-METHYLHEXANE	100.2	7.16	0.42	C7H16	7.8
33	43250	pa224m	2,2,4-TRIMETHYLPENTA	114.23	3.68	0.28	C8H18	15.1
34	43252	pa234m	2,3,4-TRIMETHYLPENTA	114.23	7	0.48	C8H18	7.94
35	43253	hep3me	3-METHYLHEPTANE	114.23	8.56	0.29	C8H18	6.49
36	43261	mecyhx	METHYLCYCLOHEXANE	98.19	10.4	0.53	C7H14	5.34
37	43262	mcypna	METHYLCYCLOPENTANE	84.16	8.81	0.82	C6H12	6.31
38	43263	hexa2m	2-METHYLHEXANE	98.19	6.79	0.32	C7H16	8.18
39	43280	lbut1e	1-BUTENE	56.11	31.4	2.6	C4H8	1.77
40	43284	bu23dm	2,3-DIMETHYLBUTANE	86.17	6.2	0.25	C6H14	8.96
41	43285	pena2m	2-METHYLPENTANE	86.17	5.6	0.45	C6H14	9.92
42	43291	pen23m	2,3-DIMETHYLPENTANE	100.2	4.87	0.39	C7H16	11.41
43	43954	n_unde	N-UNDECANE	156.3	13.2	0.12	C11H24	4.2
44	43960	hep2me	2-METHYLHEPTANE	114.2	8.18	0.29	C8H18	6.8
45	45109	mp_xyl	M&P-XYLENE	106.16	18.95	2.05	C8H10	4.71
46	45201	benze	BENZENE	78.11	1.23	0.11	C6H6	45.17
47	45202	tolue	TOLUENE	92.1	5.96	0.74	C7H8	9.32
48	45203	etbz	ETHYLBENZENE	106.16	7.1	0.75	C8H10	7.82
49	45204	o_xyl	O-XYLENE	106.17	13.7	1.8	C8H10	4.06
50	45207	bz135m	1,3,5-TRI-M-BENZENE	120.2	57.5	2.81	C9H12	0.97
51	45208	bz124m	1,2,4-TRI-M-BENZENE	120.2	32.5	2.45	C9H12	1.71
52	45209	n_prbz	N-PROPYLBENZENE	120.2	6	0.58	C9H12	9.26
53	45210	iprbz	ISO-PROPYLBENZENE	120.2	6.5	0.6	C9H12	8.55
54	45211	o_etol	O-ETHYLTOLUENE	120.2	12.3	1.48	C9H12	4.52
55	45212	m_etol	M-ETHYLTOLUENE	120.2	19.2	1.48	C9H12	2.89
56	45213	p_etol	P-ETHYLTOLUENE	120.2	12.1	1.48	C9H12	4.59
57	45219	detbz2	P-DIETHYLBENZENE	134.22	14.2	1.33	C10H14	3.9
58	45220	styr	STYRENE	104.14	58	0.6	C8H8	0.96
59	45225	bz123m	1,2,3-TRI-M-BENZENE	120.2	32.7	2.6	C9H12	1.7
66	45218	detbz1	M-DIETHYLBENZENE	134.22	14.2	1.33	C10H14	3.9
67	sum_pol	sum_pol	surrogate of TNMHC					
73	43000	TNMTC	Total Non-Methane Targeted hHydroCarbons					
74	43102	TNMHC	Total Non-Methane HydroCarbons					

a:) index of data field in the data file;
b) Molecular weight;
c) Rate constants *k* at 298 K for the reaction of OH radicals with VOCs
d) MIR maximum incremental reactivity.

Table 3*a*. Maximum, median and mean values of the measured VOC on each auto-GC site at daytime

			ChannelView (day)			Wallisville (day)			Clinton (day)			LynFerry (day)			HadenRoad (day)			TexasCity (day)			Danciger (day)			Mustang (day)			LakeJackson (day)		
field	species	code	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men
6	43202	ethane	493.18	16.61	23.33	137.29	8.6	12.84	189.69	14.46	18.59	154.98	11.32	15.86	129.71	11.355	14.87	853.02	9.335	14.60	48.65	6.155	7.56	752.4	10.04	16.54	190.08	4.33	6.50
7	43203	ethene	147.76	3.91	7.49	170.5	1.86	5.28	76.38	5.555	8.41	536.08	3.17	9.07	646	3.92	8.24	127.57	1.64	3.61	27.89	0.98	1.98	185.56	1.08	2.61	396.25	1.06	3.29
8	43204	n_prop	720.84	16.7	28.65	180.53	7.985	13.82	223.42	12.68	18.32	543.67	12.45	21.95	232.5	11.3	17.33	905.56	7.695	13.57	57.24	4.79	6.80	648.86	8.77	18.32	165.26	3.6	5.94
9	43205	prope	722.48	4.08	12.30	361.19	1.16	7.04	184.58	3.01	5.74	716.27	3.075	18.78	1271.01	3.14	8.51	125.84	1.43	2.81	42.38	0.48	1.37	96.99	0.705	2.48	136.4	0.56	1.61
10	43206	acetyl	26.64	1.03	1.59	10.2	0.51	0.71	156.72	1.77	2.67	43.15	0.375	0.78	20.78	0.68	1.06	215.4	0.5	1.35	10.21	0.27	0.35	4.11	0.26	0.41	7.67	0.31	0.42
11	43212	n_buta	1559.89	10.52	20.04	124.82	4.32	7.99	355.25	9.14	17.47	1155.8	6.74	15.33	4280.02	7.755	21.17	417.74	3.82	6.77	48.61	2.32	3.51	341.24	4.38	9.76	253.74	1.58	3.17
12	43214	i_buta	412.39	7.98	15.71	269.76	3.68	7.91	176.71	5.68	10.45	372.84	6.18	13.38	172.42	6.085	11.38	470.59	3.4	7.03	45.64	1.9	3.12	271.71	4.3	9.54	84.78	1.125	2.51
13	43216	t2bute	11.13	0.45	0.71	5.84	1.61	1.56	46	0.39	0.83	25.89	0.34	0.72	29.14	0.61	0.92	6.33	0.83	0.86	4.91	0.41	0.43	49.41	0.5	0.59	1.48	0.42	0.42
14	43217	c2bute	12.2	0.28	0.55	4.15	0.47	0.49	31.96	0.31	0.67	79.48	0.18	0.55	22.83	0.35	0.65	10.84	0.26	0.30	3.74	0.12	0.16	8.44	0.09	0.15	1.07	0.13	0.15
15	43218	bute13	144.16	0.485	1.97	18.59	0.22	0.49	39.96	0.49	1.20	174.14	0.95	1.81	123.77	0.45	1.60	27.08	0.11	0.25	4.64	0.08	0.17	51.92	0.06	0.35	16.51	0.12	0.20
16	43220	n_pent	106.33	4.25	7.78	55.87	2.12	3.66	156.79	5.05	10.05	652.7	3.62	8.72	200.16	4.44	8.38				16.96	1.25	1.75	69.02	0.9	2.17	22.12	1.03	1.64
17	43221	ipenta	798.38	9.64	15.81	418.09	4.21	7.51	374.95	10.73	21.36	814.48	7.93	17.35	658.61	9.85	18.29				57	2.82	3.79	140.6	1.23	4.50	24.63	2.1	3.05
18	43224	pente1	6.89	0.18	0.29	2.08	0.08	0.13	33.02	0.35	0.76	17.04	0.17	0.44	19.6	0.16	0.41	3.78	0.07	0.16	7.18	0.06	0.10	14.23	0.04	0.09	0.72	0.06	0.08
19	43226	t2pene	4.83	0.15	0.26	2.29	0.09	0.14	28	0.48	0.94	36.32	0.18	0.58	96.96	0.2	0.77	6.59	0.09	0.21	4.54	0.06	0.09	2.1	0.03	0.10	1.18	0.07	0.10
20	43227	c2pene	3.68	0.11	0.17	1.02	0.06	0.08	14.95	0.24	0.46	11.01	0.13	0.32	57.5	0.11	0.41	3.32	0.05	0.12	2.33	0.06	0.09	6.99	0.03	0.08	0.51	0.06	0.08
21	43230	pena3m				17.99	0.64	1.12				120.8	0.99	2.37	108.63	1.25	2.42	62.63	0.58	1.13	17	0.19	0.36	34.15	0.45	1.03	23.1	0.24	0.45
22	43231	n_hex	59.32	2.205	4.18	72.23	1.04	2.22	190.36	3.03	5.85	575.13	3.04	6.50	93.93	3.3	5.17	104.79	1.19	2.31	17.86	0.48	0.90	82.22	1.005	2.35	40.27	0.47	0.92
23	43232	n_hept	44.26	0.71	1.21	9.2	0.35	0.57	127.22	1.36	2.16	345.53	0.76	1.85	25.76	0.98	1.50	51.07	0.63	1.12	9.19	0.19	0.30	86.73	0.2	0.61	14.77	0.22	0.37
24	43233	n_oct	18.43	0.44	0.70	14.23	0.2	0.36	28.21	0.64	1.03	145.79	0.46	1.09	48.56	0.62	1.22	34.93	0.25	0.51	4.18	0.09	0.16	52.57	0.18	0.49	17.17	0.115	0.19
25	43235	n_non	6.02	0.35	0.52	15.56	0.12	0.21	12.58	0.44	0.67	52.1	0.26	0.71	24.01	0.29	0.46	18.29	0.16	0.27	1.78	0.06	0.09	51.62	0.11	0.27	5.91	0.08	0.11
26	43238	n_dec	11.27	0.49	0.71	15.53	0.21	0.28	7.95	0.51	0.71	114.32	0.4	0.99	21.95	0.63	0.75	9.1	0.15	0.22	1	0.07	0.09	49.87	0.3	0.39	1.38	0.08	0.10
27	43242	cpenta	12.54	0.62	1.01	5.86	0.26	0.41	206.93	0.57	1.16	45.17	0.455	1.00	884.02	0.51	1.41	13.58	0.35	0.48	3.03	0.16	0.21	28.35	0.16	0.38	5.1	0.14	0.20
28	43243	i_pren	36.24	2.28	2.93	14.61	1.03	1.44	8.07	1.28	1.48	55.57	0.84	1.38	23.28	3.005	3.98	4.62	0.55	0.71	52.8	3.58	5.04	6.51	0.6	0.89	26.56	1.77	2.49
29	43244	bu22dm	8.15	0.18	0.28	3.06	0.11	0.18	13.21	0.29	0.64	16.46	0.17	0.38	57.31	0.19	0.46	24.84	0.14	0.29	7.63	0.08	0.12	31.53	0.12	0.27	0.83	0.08	0.11
30	43247	pen24m	30.29	0.56	1.56	5.29	0.23	0.30	18.13	2.33	4.02	65.37	0.72	1.23	16.01	0.31	0.54	6.77	0.24	0.39	3.94	0.18	0.23	6.54	0.435	0.63	1.09	0.18	0.21
31	43248	cyhexa	324.2	1.06	2.53	32.33	0.42	1.07	47.01	1	1.93	256.17	1.08	3.41	124.06	1.15	2.44	60.47	0.62	1.30	4.55	0.3	0.49	50.88	0.59	1.45	6.5	0.275	0.55
32	43249	hexa3m	11.22	0.69	1.03	5.87	0.37	0.55	102.71	1.27	2.09	443.91	0.7	1.77	44.55	0.81	1.30	29.98	0.41	0.69	11.75	0.17	0.26	16.31	0.2	0.40	5	0.24	0.35
33	43250	pa224m	81.67	1.22	1.98	11.75	0.59	0.84	68.57	1.91	2.97	50.19	1.08	2.19	28.47	1.13	1.89	10.1	0.49	0.72	16.56	0.26	0.39	13.51	0.32	0.58	3.52	0.33	0.48
34	43252	pa234m	5.39	0.27	0.39	4.73	0.17	0.25	16.7	0.52	0.84	15.71	0.35	0.68	8.12	0.31	0.54	3.1	0.14	0.21	6.85	0.08	0.13	1.59	0.1	0.16	2.37	0.13	0.18
35	43253	hep3me	2.43	0.24	0.30	5.45	0.11	0.18	44.22	0.49	0.72	220.19	0.31	0.77	12.59	0.33	0.49	16.29	0.18	0.30	4.93	0.07	0.10	13.91	0.11	0.24	5.33	0.08	0.13
36	43261	mecyhx	18.83	0.85	1.39	11.85	0.4	0.72	96.03	1.26	1.98	74.59	0.65	1.44	21.82	0.97	1.48	60.86	0.62	1.26	10.57	0.25	0.44	98.01	0.53	1.54	8.25	0.18	0.39
37	43262	mcypna	24.89	0.785	1.72	13.08	0.51	0.93	113.22	1.35	3.11	132.35	1.24	2.44	82.1	1.37	2.17	56.26	0.5	0.97	7.52	0.24	0.42	35.29	0.49	1.11	8.39	0.28	0.49
38	43263	hexa2m	7.22	0.39	0.62	5.36	0.32	0.47	66.24	0.735	1.47	341.28	0.63	1.55	49.11	0.69	1.10	26.88	0.26	0.55	9.71	0.14	0.21	13.81	0.26	0.55	5	0.2	0.32
39	43280	lbut1e	82.86	0.57	1.68	118.49	0.24	1.06	27.65	0.68	1.36	129.51	0.47	1.67	239.44	0.62	3.04	35.11	0.15	0.41	4.29	0.15	0.26	50.79	0.14	0.60	4.22	0.16	0.23
40	43284	bu23dm				4.44	0.18	0.31				20.53	0.24	0.61	29.79	0.3	0.64	21.62	0.17	0.31	6.37	0.09	0.15	5.36	0.15	0.29	2.03	0.09	0.15
41	43285	pena2m				22.81	0.74	1.28				92.01	0.98	2.22	126.02	1.29	2.61	74.56	0.71	1.27	23.44	0.24	0.44	28.38	0.49	1.04	15.57	0.24	0.42
42	43291	pen23m	41.04	0.51	0.78	2.32	0.26	0.34	40.61	1.265	1.92	139.47	0.65	1.26	16.66	0.35	0.58	15	0.32	0.51	5.21	0.14	0.19	8.28	0.22	0.40	2.05	0.21	0.25
43	43954	n_unde	3.94	0.26	0.43	12.74	0.1	0.16				55.03	0.3	0.73	15.03	0.27	0.38	7.49	0.16	0.22	0.82	0.04	0.06	1.39	0.16	0.24	1.76	0.06	0.09
44	43960	hep2me	4.54	0.17	0.26	4.71	0.11	0.17	34.92	0.5	0.79	148.72	0.25	0.63	10.21	0.25	0.42	15.64	0.13	0.26	3.62	0.07	0.09	15.73	0.11	0.24	4.9	0.08	0.11
45	45109	mp_xyl	84.59	1.35	2.17	11.84	0.55	0.89	57.09	2.325	3.33	100.19	1.43	3.09	149.67	2.06	3.45	68.42	0.8	1.48	36.72	0.15	0.28	38.31	0.24	0.64	31.09	0.23	0.40
46	45201	benze	247.68	2.67	5.21	22.92	1.1	1.80	127.39	2.04	3.48	1055.68	2.97	10.07	85.51	2.68	4.28	407.99	1.43	5.13	26.92	0.67	0.97	35.33	1.16	1.92	25.89	0.59	0.90
47	45202	tolue	159.38	3.835	6.83	78.41	1.53	2.51	67.4	4.67	6.73	182.34	3.4	6.67	347.46	4.45	6.73	227.36	1.77	3.53	49.6	0.55	0.86	50.55	0.85	1.70	29.23	0.6	1.02
48	45203	etbz	51.51	0.58	0.94	5.74	0.24	0.36	7.38	0.79	1.04	424.76	0.62	2.00	40.35	0.64	1.01	17.28	0.32	0.53	11.77	0.09	0.14	4.19	0.15	0.23	2.64	0.14	0.22
49	45204	o_xyl	25.58	0.52	0.81	4.42	0.24	0.36	12.81	0.9	1.27	382.79	0.655	2.25	55.27	0.67	1.07	35.23	0.3	0.55	15.18	0.07	0.13	20.07	0.13	0.29	1.68	0.1	0.16
50	45207	bz135m	17.4	0.26	0.45	5.1	0.09	0.14	4.69	0.3	0.46	26.18	0.2	0.43	6.18	0.19	0.31	6.75	0.11	0.17	5.41	0.05	0.08	13.99	0.06	0.14	0.72	0.06	0.08
51	45208	bz124m	41.07	0.84	1.21	10.64	0.24	0.35	11.14	1.165	1.62	30.64	0.575	1.03	17.22	0.63	1.02	6.45	0.29	0.38	16	0.1	0.16	117.51	0.08	0.42	32.82	0.11	0.19
52	45209	n_prbz	7.79	0.18	0.25	2.99	0.08	0.12	1.96	0.23	0.32	7.53	0.18	0.30	2.79	0.13	0.19	1.27	0.09	0.11	3.23	0.04	0.06	43.76	0.05	0.14	34.94	0.06	0.09
53	45210	iprbz	15.43	0.17	0.39	3.07	0.14	0.23	5.51	0.17	0.26	387.39	0.25	1.61	34.84	0.08	0.39	1.59	0.05	0.07	0.34	0.04	0.06	2.92	0.05	0.10	0.97	0.05	0.07
54	45211	o_etol	1.05	0.14	0.21	3.66	0.08	0.12				12.52	0.21	0.37	6.64	0.16	0.24	1.7	0.06	0.10	4.13	0.05	0.07	3.6	0.04	0.08	18.65	0.05	0.08
55	45212	m_etol				7.11	0.18	0.30				20.02	0.37	0.64	5.89	0.42	0.63	4.73	0.2	0.28	11.02	0.25	0.37	13	0.11	0.27	1.54	0.13	0.18
56	45213	p_etol				4.24	0.18	0.25				22.23	0.3	0.51	3.86	0.23	0.34	3.35	0.12	0.17	5.83	0.11	0.37	8.79	0.12	0.23	0.54	0.13	0.15
57	45219	detbz2				3.29	0.13	0.18				12.08	0.27	0.42	2.86	0.22	0.28	4.82	0.14	0.17	0.35	0.04	0.05				2.22	0.07	0.11
58	45220	styr	20.16	0.3	0.65	3.88	0.14	0.21	14.87	0.32	0.66	1266.45	0.39	6.52	91.12	0.31	1.42	4.28	0.19	0.28	0.96	0.07	0.10	6.1	0.1	0.16	2.3	0.09	0.12
59	45225	bz123m	10.95	0.52	0.75	5.79	0.17	0.26	13	0.605	0.99	16.64	0.36	0.61	6.16	0.42	0.60	3.48	0.22	0.31	1.31	0.435	0.47	12.49	0.15	0.32	1.69	0.14	0.20
66	45218	detbz1				4.25	0.075	0.12				1.77	0.23	0.31	2.33	0.19	0.23												
67	sumpol	sum_pol	1611.59	112.26	161.63				1443.18	116.13	176.83																		
73	43000	TNMTC	3023.46	127.98	258.34	942.53	59.11	96.02				1445.92	134.9	205.14	5445.69	118.195	172.59	4488.82	83.905	115.27	496.11	33.4	44.53						
74	43102	TNMHC	3105.77	134.75	270.88	970.12	69.095	107.84	2233.35	123.43	185.23	1561.23	147.05	223.23	5540.87	136.5	194.99	4606.75	91.235	124.67	582.4	37.49	49.53						
77	c43102	cTNMHC	2956.54	111.8	172.31	942.51	53.675	86.97	1766.43	100.57	157.59	3468.38	107.665	179.80	5432.09	112.57	165.94	3965.09	49.36	79.26	485.13	31.85	42.92	2226.14	41.93	79.71	670.68	23.37	35.54
78	c43000	cTNMTC	2814.89	111.44	171.34	940.81	53.21	86.48	1766.43	100.57	156.98	3467.06	106.66	178.01	5414.76	111.91	164.35	3964.93	49.25	79.06	482.96	31.77	42.78	2224.98	41.685	79.49	670.62	23.32	35.42

Table 3*b*. Maximum, median and mean values of the measured VOC on each auto-GC site at nighttime

			ChannelView (night)			Wallisville (night)			Clinton (night)			LynFerry (night)			HadenRoad (night)			TexasCity (night)			Danciger (night)			Mustang (night)			LakeJackson (night)		
field	species	code	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men	max	med	men
6	43202	ethane	528.82	35.045	50.37	241.54	18.71	24.84	311.59	22.88	29.27	221.66	24.36	32.38	176.63	18.54	23.68	482.23	14.32	28.07	117.04	10.23	12.26	707.23	30.68	59.19	348.69	6.335	11.19
7	43203	ethene	128.84	7.645	12.95	109.56	4.49	10.15	92	6.86	9.14	454.8	7.18	18.05	201.9	5.545	10.93	63.56	2.36	5.14	43.69	1.38	2.39	205.71	2.26	7.02	372.9	2.005	6.01
8	43204	n_prop	789.24	40.3	68.98	180.82	20.03	29.68	236.61	18.9	25.40	377.94	26.71	45.14	267.92	18.77	26.43	1213.71	12.68	28.96	160.7	12.4	17.54	1027.65	33.42	76.01	258.23	6.985	11.84
9	43205	prope	522.69	8.68	18.47	528.58	6.09	26.38	84.89	4.41	6.78	1281.88	12.02	46.53	307.08	4.82	9.88	137.04	2.68	4.86	113.16	1.1	2.87	110.04	1.83	6.85	77.25	1.095	2.66
10	43206	acetyl	80.64	1.28	2.59	8.65	0.62	0.91	33.04	1.78	2.66	50.48	0.685	1.48	10.29	0.85	1.25	160.61	0.53	3.78	2.46	0.33	0.41	6.49	0.34	0.58	2.68	0.43	0.57
11	43212	n_buta	656.93	23.81	41.15	329.83	9.835	16.39	529.38	11.96	23.55	812.84	14.04	32.78	2263.75	13.95	37.12	613.24	6.94	14.26	66.09	4.38	5.94	522.23	16.85	38.29	352.78	3.14	6.92
12	43214	i_buta	370.4	19.71	32.35	289.45	8.835	15.82	540.41	8.22	15.26	616.17	12.38	28.56	523.59	10.41	16.71	650.57	6.95	14.96	103.47	3.62	5.29	534.95	18.33	40.22	114.47	2.24	4.92
13	43216	t2bute	26.36	0.69	1.25	8.23	1.57	1.58	796.43	0.55	2.72	117.87	0.44	1.77	43.6	0.72	1.59	3.59	0.92	0.97	10.63	0.46	0.57	25.7	0.54	0.79	1.51	0.42	0.43
14	43217	c2bute	19.78	0.53	1.04	3.62	0.47	0.55	555.43	0.44	2.05	74.06	0.29	1.40	39.56	0.59	1.37	15.94	0.3	0.38	9.17	0.14	0.26	12.39	0.14	0.32	1.18	0.14	0.17
15	43218	bute13	122.12	1.325	4.38	31.79	0.36	1.01	57.05	0.615	2.08	87.98	1.41	3.94	97.96	0.7	2.53	68.73	0.16	0.72	27.52	0.14	0.29	152.48	0.12	1.83	3.45	0.13	0.26
16	43220	n_pent	127.48	9.67	16.03	49.25	4.695	7.13	312.13	6.06	13.31	1164.65	6.945	18.97	318	7.09	16.64				18.89	2.31	2.80	99.84	4.095	8.37	158.15	1.62	3.88
17	43221	ipenta	510.84	19.37	30.25	161.8	8.115	12.40	771.13	12.82	26.74	564.33	11.895	26.32	1337.55	14.62	37.74				59.46	4.47	5.46	212.39	6.635	19.64	132.38	2.87	5.63
18	43224	pente1	8.67	0.34	0.61	2.1	0.1	0.20	44.96	0.44	1.14	17.6	0.29	0.77	33.92	0.3	1.15	4.07	0.11	0.22	3.49	0.09	0.13	10.55	0.06	0.27	0.86	0.09	0.11
19	43226	t2pene	16.27	0.32	0.69	3.94	0.12	0.22	53.44	0.66	1.48	289.79	0.27	1.15	108.29	0.49	2.32	6.27	0.15	0.30	0.88	0.08	0.11	3.9	0.05	0.27	1.37	0.08	0.13
20	43227	c2pene	9.4	0.2	0.43	1.72	0.07	0.13	27.31	0.32	0.72	11.04	0.19	0.46	46.45	0.25	1.12	3.35	0.07	0.16	0.52	0.05	0.07	2.86	0.04	0.17	0.62	0.06	0.09
21	43230	pena3m				13.75	1.31	1.91				249.87	1.89	4.62	62.66	2.315	5.02	107.83	1.06	2.24	11.18	0.44	0.60	66.4	1.825	3.50	22.27	0.54	0.87
22	43231	n_hex	57.95	5.4	8.14	94.32	2.61	4.69	203.63	3.34	7.96	302.62	6.45	12.83	79.43	5.6	8.56	174.58	2.2	4.51	22.53	1.115	1.59	166.85	4	7.56	47.48	0.97	1.87
23	43232	n_hept	15.13	1.47	2.15	10.01	0.62	0.90	41.3	1.555	2.76	73.45	1.28	2.99	24.95	1.75	2.79	95.21	1.13	2.19	4.12	0.36	0.44	104.74	0.17	2.73	12.56	0.39	0.66
24	43233	n_oct	7.48	0.81	1.09	7.69	0.33	0.52	24.6	0.72	1.29	36.79	0.78	1.74	29.24	1.05	2.11	57.29	0.47	0.99	2.87	0.16	0.22	50.83	0.64	1.51	3.24	0.17	0.30
25	43235	n_non	11.58	0.59	0.79	2.59	0.13	0.20	16.47	0.59	0.88	25.98	0.33	0.86	6.8	0.48	0.76	22.9	0.235	0.41	1.45	0.08	0.11	48.87	0.26	0.80	0.96	0.1	0.15
26	43238	n_dec	13.77	0.82	1.09	2.97	0.25	0.32	10.02	0.75	0.96	33.17	0.58	1.08	9.23	0.8	1.02	6.91	0.2	0.29	0.34	0.09	0.11	45.71	0.38	0.74	1.05	0.09	0.13
27	43242	cpenta	18.16	1.305	1.98	4.7	0.46	0.67	478.23	0.655	1.79	1320.42	0.77	3.63	103.66	0.77	1.68	22.29	0.47	0.73	1.44	0.27	0.31	12.66	0.67	1.24	6.14	0.21	0.34
28	43243	i_pren	29.14	0.84	1.36	40.69	0.29	0.82	17.2	0.45	0.87	13.6	0.63	1.34	30.88	0.52	1.00	7.26	0.13	0.26	6.49	0.26	0.56	7.87	0.135	0.44	6.59	0.16	0.37
29	43244	bu22dm	26.9	0.36	0.60	5.34	0.18	0.28	15.68	0.35	0.83	29.37	0.27	0.68	71.15	0.36	1.11	44.1	0.26	0.59	3.36	0.11	0.15	34.77	0.41	0.85	2.18	0.14	0.20
30	43247	pen24m	2.05	0.55	0.76	2.8	0.26	0.37	15.03	2.22	3.70	22.56	0.84	1.43	15.44	0.57	1.14	5.03	0.44	0.58	0.96	0.21	0.26	5.37	0.72	0.90	1.5	0.26	0.31
31	43248	cyhexa	287.82	3.06	6.23	155.21	1.23	3.08	55.97	1.215	2.59	361.85	2.945	10.18	133.77	2.04	4.41	102.12	1.32	2.63	4.7	0.59	0.80	76.91	2.63	5.26	10.57	0.48	1.04
32	43249	hexa3m	18.27	1.24	1.91	5.41	0.6	0.84	55.17	1.41	2.89	94.96	1.06	2.37	26.64	1.44	2.40	54.71	0.76	1.37	2.47	0.33	0.41	21.62	0.15	1.21	5.98	0.41	0.59
33	43250	pa224m	47.07	2.08	3.30	15.89	0.77	1.13	70.34	2.005	3.60	86.29	1.23	3.08	52.41	2.09	4.04	20.5	0.68	1.10	6.1	0.42	0.59	32.4	0.93	1.62	5.57	0.6	0.73
34	43252	pa234m	9.47	0.43	0.67	4.8	0.2	0.31	23.11	0.56	1.06	42.62	0.43	1.05	13.03	0.59	1.19	7.74	0.21	0.35	1.96	0.12	0.18	6.96	0.21	0.33	1.62	0.21	0.25
35	43253	hep3me	3.5	0.39	0.50	2.63	0.14	0.21	13.08	0.55	0.91	44.52	0.37	0.90	8.2	0.61	0.91	26.95	0.31	0.57	1.46	0.09	0.12	15.56	0.32	0.69	1.68	0.11	0.20
36	43261	mecyhx	30.16	2.17	3.10	25.67	1.01	1.58	24.63	1.96	2.80	86.24	1.56	3.02	24.74	2.14	2.84	106.22	1.17	2.94	5.66	0.67	0.84	118.09	2.59	6.90	13.24	0.375	0.89
37	43262	mcypna	31.9	2.16	3.83	14.14	1.26	1.86	119.88	1.75	4.50	332.81	2.65	4.85	49.8	2.47	4.03	109.12	0.955	2.14	7.1	0.555	0.73	47.42	2.01	3.74	9.74	0.53	0.96
38	43263	hexa2m	12.11	0.74	1.22	5.05	0.53	0.71	38.46	0.93	2.11	76.39	0.89	2.06	29.36	1.22	2.07	48.87	0.59	1.13	2.33	0.27	0.32	23.38	1.08	1.75	5.78	0.38	0.55
39	43280	lbut1e	100.87	1.55	3.97	45.43	0.65	1.90	59.23	0.82	1.80	247.87	1.395	4.67	388.82	1.08	4.35	141.01	0.22	1.06	9.66	0.23	0.44	80.09	0.19	1.86	17.86	0.16	0.37
40	43284	bu23dm				5.26	0.33	0.51				45.84	0.385	1.07	28.32	0.57	1.49	37.13	0.31	0.64	2.94	0.15	0.20	16.01	0.46	0.94	2.17	0.18	0.25
41	43285	pena2m				19.69	1.465	2.23				189.52	1.755	4.46	81.52	2.455	5.58	123.2	1.28	2.54	11.31	0.56	0.71	34.38	1.71	4.10	17.01	0.48	0.80
42	43291	pen23m	14.55	0.83	1.30	2.48	0.33	0.45	26.99	1.565	2.59	34.93	0.77	1.44	12.62	0.66	1.16	25.86	0.555	0.89	1.37	0.19	0.24	9.14	0.27	0.70	2.42	0.29	0.37
43	43954	n_unde	6.16	0.47	0.65	2.69	0.12	0.18				42.23	0.35	0.76	8.8	0.44	0.61	3.95	0.2	0.26	0.23	0.07	0.07	1.07	0.19	0.24	1.65	0.07	0.09
44	43960	hep2me	4.76	0.24	0.42	2.89	0.13	0.21	19.01	0.55	0.97	30.5	0.32	0.85	8.16	0.47	0.77	25.66	0.21	0.48	1.43	0.08	0.10	17.41	0.32	0.72	1.73	0.11	0.17
45	45109	mp_xyl	58.35	2.34	3.24	22.99	0.94	1.45	24.4	3.33	4.57	114.33	2.62	5.02	71.11	3.3	5.04	48.44	1.2	2.24	3.3	0.46	0.57	33.42	0.93	1.83	6.18	0.49	0.74
46	45201	benze	425.73	5.52	8.26	30.43	1.71	2.61	165.14	2.23	4.54	2828.57	5.38	29.57	152.21	3.485	5.38	462.96	2.055	7.95	17.54	1.19	1.53	60.92	3.33	5.47	18.43	0.97	1.55
47	45202	tolue	113.31	5.95	8.60	49.95	2.24	3.50	60.03	5.25	8.25	370.77	7.5	13.50	98.91	7.5	11.03	165.49	2.39	5.31	7.19	1.25	1.50	66.06	2.77	4.91	18.43	1.12	1.83
48	45203	etbz	23.97	0.92	1.50	5.97	0.29	0.46	6.93	1.07	1.41	452.97	0.98	3.08	17.54	1.07	1.56	14.49	0.41	0.74	1.05	0.18	0.21	5.27	0.21	0.52	1.87	0.22	0.29
49	45204	o_xyl	14.22	0.93	1.27	9.34	0.34	0.51	9.17	1.245	1.67	716.63	1.09	3.82	15.58	1.14	1.69	29.05	0.41	0.82	1.1	0.18	0.22	20.17	0.4	0.84	2.49	0.21	0.30
50	45207	bz135m	10.04	0.58	0.86	2.1	0.13	0.20	4.36	0.49	0.68	21.55	0.34	0.64	3.75	0.46	0.64	4.74	0.17	0.27	1.05	0.09	0.12	13.79	0.14	0.32	0.89	0.09	0.12
51	45208	bz124m	27.62	1.34	1.94	5.55	0.39	0.57	9.11	1.99	2.50	43.96	1.02	1.90	24.64	1.375	2.02	6.69	0.46	0.63	2.39	0.26	0.33	18.42	0.285	0.59	1.63	0.25	0.33
52	45209	n_prbz	3.37	0.29	0.37	1.96	0.12	0.16	2.04	0.35	0.45	25.63	0.26	0.48	2.34	0.25	0.35	1.99	0.13	0.18	0.56	0.07	0.09	5.41	0.12	0.20	1.02	0.1	0.11
53	45210	iprbz	19.99	0.3	0.66	2.8	0.13	0.24	2.9	0.2	0.29	568.02	0.3	2.22	26.47	0.1	0.51	1.26	0.05	0.07	1	0.04	0.06	2.07	0.07	0.16	0.79	0.06	0.08
54	45211	o_etol	1.34	0.24	0.34	1.88	0.1	0.15				17.1	0.29	0.55	2.95	0.31	0.43	2.39	0.09	0.14	0.59	0.07	0.08	3.79	0.09	0.16	0.41	0.07	0.09
55	45212	m_etol				5.09	0.445	0.56				64.21	0.65	1.20	7.04	1.04	1.34	7.69	0.33	0.49	5.92	0.58	0.74	11.07	0.48	0.71	1.44	0.28	0.34
56	45213	p_etol				2.66	0.3	0.40				37.24	0.47	0.82	3.27	0.44	0.59	5.94	0.18	0.31	0.24	0.11	0.10	8.75	0.29	0.52	0.51	0.165	0.18
57	45219	detbz2				2.1	0.18	0.23				5.84	0.38	0.60	3.07	0.36	0.47	1.52	0.19	0.24	0.2	0.06	0.07				0.7	0.12	0.14
58	45220	styr	37.04	0.785	1.47	33.54	0.2	0.39	15.25	0.42	1.01	7897.54	0.755	20.16	90.87	0.61	1.90	7.56	0.21	0.28	1.23	0.09	0.13	8.3	0.15	0.34	2.24	0.1	0.14
59	45225	bz123m	8.13	1.27	1.67	6.56	0.44	0.62	11.07	1.125	1.75	22.06	0.62	0.98	9.74	0.93	1.31	4.45	0.33	0.51	2.66	0.68	0.77	12.98	0.53	0.81	1.38	0.22	0.28
66	45218	detbz1				1.08	0.11	0.19				3.9	0.45	0.57	1.65	0.26	0.34												
67	sumpol	sum_pol	1946.11	245.56	319.64				2386.93	176.68	260.23																		
73	43000	TNMTC	3380.69	249.73	490.58	1003.94	134.555	189.28				3409.12	341.41	456.62	3526.5	201.535	283.04	5187.28	107.39	188.02	560.86	57.25	71.97						
74	43102	TNMHC	3504.55	260.9	513.31	1058.11	146.345	204.16	2506.46	138.03	226.63	3576.75	371.16	492.37	4377.47	230.59	320.41	5402.08	118.45	201.56	575.04	63.7	78.63						
77	c43102	cTNMHC	3268.81	243.51	338.04	1003.96	123.44	176.06	2415.6	130.58	216.28	8795.74	253.58	383.24	3525.72	195.265	272.62	4362.19	82.99	150.93	558.97	55.84	70.42	3030.36	162.95	298.96	1191.21	38.845	63.84
78	c43000	cTNMTC	3254.2	242.53	335.90	988.24	122.97	175.06	2388.46	130.53	215.24	8792.98	249.51	379.35	3519.1	192.955	270.11	4361.85	82.63	150.30	531.45	55.68	70.14	3029.81	162.39	297.55	1191.14	38.72	63.65

Table 4. Species with relative great abundance at each auto-GC site

Site	species with avg conc. >= 10 ppbC	species with avg conc. between 5 and 10 ppbC
ChannelView	43202 (ethane); 43203 (ethene); 43204 (n_propane); 43205 (propene); 43212 (n-butane); 43214 (i_butane); 43220 (n-pentane); 43221 (i-pentane)	43231 (n-hexane); 43248 (cycloheane); 45201 (benzene); 45202 (toluene)
Clinton	43202 (ethane); 43204 (n_propane); 43212 (n-butane); 43214 (i_butane); 43220 (n-pentane); 43221 (i-pentane)	43203 (ethene); 43205 (propene); 43231 (n-hexane); 45202 (toluene)
Danciger	43202 (ethane); 43204 (n_propane);	43212 (n-butane); 43214 (i_butane); 43221 (i-pentane)
Wallisville	43202 (ethane); 43203 (ethene); 43204 (n_propane); 43205 (propene); 43212 (n-butane); 43214 (i_butane); 43221 (i-pentane)	43220 (n-pentane); 43231 (n-hexane);
HadenRoad	43202 (ethane); 43203 (ethene); 43204 (n_propane); 43205 (propene); 43212 (n-butane); 43214 (i_butane); 43220 (n-pentane); 43221 (i-pentane); 45202 (toluene)	43230 (3-methylpentane); 43231 (n-hexane); 43285 (2-methylpentane); 45201 (benzene)
LakeJackson	43202 (ethane); 43204 (n_propane);	43212 (n-butane); 43214 (i_butane); 43221 (i-pentane)
Mustang	43202 (ethane); 43204 (n_propane); 43212 (n-butane); 43214 (i_butane); 43221 (i-pentane)	43203 (ethene); 43205 (propene); 43220 (n-pentane); 43231 (n-hexane); 43248 (cyclohexane); 43261 (methylcyclohexane); 45201 (benzene); 45202 (toluene)
TexasCity	43202 (ethane); 43204 (n_propane); 43212 (n-butane); 43214 (i_butane);	43203 (ethene); 43205 (propene); 43206 (aceylene); 45201 (benzene); 45202 (toluene)
LynFerry	43202 (ethane); 43203 (ethene); 43204 (n_propane); 43205 (propene); 43212 (n-butane); 43214 (i_butane); 43220 (n-pentane); 43221 (i-pentane); 43231 (n-hexane); 45201 (benzene); 45202 (toluene)	43230 (3-methylpentane); 43242 (cyclopentane); 43248 (cyclohexane); 43262 (methylcyclopentane); 43280 (l-butene); 43285 (2-methylpentane); 45109 (m&p xylene);

Table 5. Average number of species in concentration category

site code	site name	missing	C=0	0<C<=0.01	0.01<C<=0.1	0.1<C<=1	1<C<=5	5<C<=10	C>10	sum
482010026	ChannelView	21	1	0	2	20	12	3	5	64
482011035	Clinton	25	2	0	1	15	13	3	4	64
480390618	Danciger	16	7	0	11	21	6	2	1	64
480390619	MustangBayou	15	11	1	8	17	7	2	3	64
480391016	LakeJackson	13	13	0	10	21	5	1	1	64
481670056	TexasCity	13	4	0	8	26	9	2	2	64
482010617	Wallisville	10	6	0	7	26	10	2	3	64
482010803	HadenRoad	11	1	0	3	26	15	4	5	64
482011015	LynFerry	14	5	0	3	22	12	3	5	64

Table 6*a*. Species pairs with apparent high correlation coefficient at *Channel View* site

ETHANE	PROPANE	0.97
PROPANE	METHYLCYCLOPENTANE	0.85
TRANS-2-BUTENE	CIS-2-BUTENE	0.94
N-PENTANE	METHYLCYCLOHEXANE	0.88
	METHYLCYCLOPENTANE	0.95
1-PENTENE	TRANS-2-PENTENE	0.88
	CIS-2-PENTENE	0.9
TRANS-2-PENTENE	CIS-2-PENTENE	0.98
N-HEXANE	METHYLCYCLOPENTANE	0.88
N-HEPTANE	METHYLCYCLOPENTANE	0.88
N-NONANE	N-DECANE	0.9
	N-UNDECANE	0.87
N-DECANE	N-UNDECANE	0.92
3-METHYLHEXANE	3-METHYLHEPTANE	0.9
	METHYLCYCLOPENTANE	0.87
	2-METHYLHEXANE	0.97
3-METHYLHEPTANE	METHYLCYCLOPENTANE	0.85
	2-METHYLHEXANE	0.87
	O-ETHYLTOLUENE	0.86
METHYLCYCLOHEXANE	METHYLCYCLOPENTANE	0.94
M&P-XYLENE	O-XYLENE	0.93
O-XYLENE	135-TRI-M-BENZENE	0.86
	124-TRI-M-BENZENE	0.85
	N-PROPYLBENZENE	0.85
	O-ETHYLTOLUENE	0.89
135-TRI-M-BENZENE	124-TRI-M-BENZENE	0.96
	N-PROPYLBENZENE	0.94
	O-ETHYLTOLUENE	0.91
124-TRI-M-BENZENE	N-PROPYLBENZENE	0.94
	O-ETHYLTOLUENE	0.95
N-PROPYLBENZENE	O-ETHYLTOLUENE	0.93

Table 6*b*. Species pairs with apparent high correlation coefficient at *Clinton* site

ETHANE	PROPANE	0.88
TRANS-2-BUTENE	CIS-2-BUTENE	1
N-PENTANE	ISO-PENTANE	0.92
1-PENTENE	TRANS-2-PENTENE	0.9
	CIS-2-PENTENE	0.89
TRANS-2-PENTENE	CIS-2-PENTENE	1
N-HEXANE	22-DIMETHYLBUTANE	0.86
	CYCLOHEXANE	0.92
	3-METHYLHEXANE	0.89
	METHYLCYCLOPENTANE	0.98
	2-METHYLHEXANE	0.91
	23-DIMETHYLPENTANE	0.92
N-HEPTANE	3-METHYLHEXANE	0.95
	3-METHYLHEPTANE	0.93
	METHYLCYCLOHEXANE	0.95
	2-METHYLHEXANE	0.94
	23-DIMETHYLPENTANE	0.91
N-OCTANE	N-NONANE	0.87
CYCLOHEXANE	3-METHYLHEXANE	0.85
	METHYLCYCLOPENTANE	0.91
	2-METHYLHEXANE	0.9
	23-DIMETHYLPENTANE	0.9
3-METHYLHEXANE	METHYLCYCLOHEXANE	0.87
	METHYLCYCLOPENTANE	0.9
	2-METHYLHEXANE	0.99
	23-DIMETHYLPENTANE	0.97
224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.95
3-METHYLHEPTANE	METHYLCYCLOHEXANE	0.91
	2-METHYLHEPTANE	0.91
METHYLCYCLOHEXANE	2-METHYLHEXANE	0.87
	23-DIMETHYLPENTANE	0.85
	2-METHYLHEPTANE	0.86
METHYLCYCLOPENTANE	2-METHYLHEXANE	0.92
	23-DIMETHYLPENTANE	0.92
2-METHYLHEXANE	23-DIMETHYLPENTANE	0.98
M&P-XYLENE	ETHYLBENZENE	0.86
	O-XYLENE	0.92
ETHYLBENZENE	O-XYLENE	0.89
135-TRI-M-BENZENE	124-TRI-M-BENZENE	0.9
	N-PROPYLBENZENE	0.9
124-TRI-M-BENZENE	N-PROPYLBENZENE	0.94

Table 6*c*. Species pairs with apparent high correlation coefficient at *Danciger* site

ACETYLENE	TOLUENE	0.86
	ETHYLBENZENE	0.85
N-BUTANE	ISO-BUTANE	0.88
	N-PENTANE	0.9
TRANS-2-BUTENE	CIS-2-BUTENE	0.99
N-PENTANE	ISO-PENTANE	0.86
	CYCLOPENTANE	0.88
ISO-PENTANE	CYCLOPENTANE	0.86
TRANS-2-PENTENE	CIS-2-PENTENE	0.97
3-METHYLPENTANE	N-HEXANE	0.87
	22-DIMETHYLBUTANE	0.87
	23-DIMETHYLBUTANE	0.96
	2-METHYLPENTANE	0.98
	BENZENE	0.87
N-HEXANE	METHYLCYCLOPENTANE	0.89
N-HEPTANE	3-METHYLHEXANE	0.88
	3-METHYLHEPTANE	0.87
	2-METHYLHEXANE	0.89
	23-DIMETHYLPENTANE	0.86
N-DECANE	N-UNDECANE	0.93
CYCLOPENTANE	METHYLCYCLOPENTANE	0.85
22-DIMETHYLBUTANE	23-DIMETHYLBUTANE	0.87
	2-METHYLPENTANE	0.9
CYCLOHEXANE	METHYLCYCLOHEXANE	0.86
	METHYLCYCLOPENTANE	0.86
3-METHYLHEXANE	3-METHYLHEPTANE	0.89
	2-METHYLHEXANE	0.99
	23-DIMETHYLPENTANE	0.95
	M&P-XYLENE	0.87
	TOLUENE	0.94
	ETHYLBENZENE	0.88
	O-XYLENE	0.87
224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.97
	23-DIMETHYLPENTANE	0.91
	TOLUENE	0.85
234-TRIMETHYLPENTA	23-DIMETHYLPENTANE	0.85
N-PROPYLBENZENE	O-ETHYLTOLUENE	0.88
124-TRI-M-BENZENE	N-PROPYLBENZENE	0.95
	O-ETHYLTOLUENE	0.93
	P-ETHYLTOLUENE	0.87

3-METHYLHEPTANE	2-METHYLHEXANE	0.9
	2-METHYLHEPTANE	0.91
	M&P-XYLENE	0.86
	TOLUENE	0.85
	ETHYLBENZENE	0.86
	O-XYLENE	0.86
2-METHYLHEXANE	23-DIMETHYLPENTANE	0.96
	M&P-XYLENE	0.88
	TOLUENE	0.95
	ETHYLBENZENE	0.88
	O-XYLENE	0.88
23-DIMETHYLBUTANE	2-METHYLPENTANE	0.99
2-METHYLPENTANE	BENZENE	0.86
23-DIMETHYLPENTANE	TOLUENE	0.91
M&P-XYLENE	TOLUENE	0.92
	ETHYLBENZENE	0.98
	O-XYLENE	0.99
	135-TRI-M-BENZENE	0.85
	124-TRI-M-BENZENE	0.96
	N-PROPYLBENZENE	0.89
	O-ETHYLTOLUENE	0.9
	P-ETHYLTOLUENE	0.89
TOLUENE	ETHYLBENZENE	0.93
	O-XYLENE	0.92
	124-TRI-M-BENZENE	0.88
ETHYLBENZENE	O-XYLENE	0.98
	124-TRI-M-BENZENE	0.94
	N-PROPYLBENZENE	0.87
	O-ETHYLTOLUENE	0.88
	P-ETHYLTOLUENE	0.87
O-XYLENE	135-TRI-M-BENZENE	0.85
	124-TRI-M-BENZENE	0.96
	N-PROPYLBENZENE	0.88
	O-ETHYLTOLUENE	0.91
	P-ETHYLTOLUENE	0.91
135-TRI-M-BENZENE	124-TRI-M-BENZENE	0.9
	N-PROPYLBENZENE	0.87
	O-ETHYLTOLUENE	0.86
O-ETHYLTOLUENE	P-ETHYLTOLUENE	0.86
M-ETHYLTOLUENE	123-TRI-M-BENZENE	0.93

Table 6*d*. Species pairs with apparent high correlation coefficient at *Haden Road* site

TRANS-2-BUTENE	CIS-2-BUTENE	0.97
	1-PENTENE	0.85
CIS-2-BUTENE	1-PENTENE	0.9
N-PENTANE	ISO-PENTANE	0.9
ISO-PENTANE	1-PENTENE	0.87
	TRANS-2-PENTENE	0.85
	24-DIMETHYLPENTANE	0.85
TRANS-2-PENTENE	CIS-2-PENTENE	0.99
3-METHYLPENTANE	N-HEPTANE	0.85
	24-DIMETHYLPENTANE	0.91
	3-METHYLHEXANE	0.91
	3-METHYLHEPTANE	0.86
	METHYLCYCLOPENTANE	0.93
	2-METHYLHEXANE	0.93
	23-DIMETHYLBUTANE	0.91
	2-METHYLPENTANE	0.97
	23-DIMETHYLPENTANE	0.94
	2-METHYLHEPTANE	0.85
N-HEPTANE	3-METHYLHEXANE	0.9
	3-METHYLHEPTANE	0.88
	METHYLCYCLOHEXANE	0.92
	METHYLCYCLOPENTANE	0.88
	2-METHYLHEXANE	0.87
	23-DIMETHYLPENTANE	0.92
	2-METHYLHEPTANE	0.9
N-NONANE	N-DECANE	0.93
	N-UNDECANE	0.89
N-DECANE	N-UNDECANE	0.93
24-DIMETHYLPENTANE	3-METHYLHEXANE	0.88
	224-TRIMETHYLPENTA	0.93
	234-TRIMETHYLPENTA	0.9
	METHYLCYCLOPENTANE	0.88
	2-METHYLHEXANE	0.89
	23-DIMETHYLBUTANE	0.87
	2-METHYLPENTANE	0.86
	23-DIMETHYLPENTANE	0.96
3-METHYLHEXANE	3-METHYLHEPTANE	0.9
	METHYLCYCLOHEXANE	0.86
	METHYLCYCLOPENTANE	0.94
	2-METHYLHEXANE	0.98
	2-METHYLPENTANE	0.86
	23-DIMETHYLPENTANE	0.96
	2-METHYLHEPTANE	0.9

224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.96
	23-DIMETHYLPENTANE	0.89
234-TRIMETHYLPENTA	23-DIMETHYLPENTANE	0.87
3-METHYLHEPTANE	METHYLCYCLOHEXANE	0.88
	METHYLCYCLOPENTANE	0.87
	2-METHYLHEXANE	0.91
	23-DIMETHYLPENTANE	0.91
	2-METHYLHEPTANE	0.97
METHYLCYCLOHEXANE	METHYLCYCLOPENTANE	0.9
	2-METHYLHEXANE	0.86
	23-DIMETHYLPENTANE	0.88
	2-METHYLHEPTANE	0.89
METHYLCYCLOPENTANE	2-METHYLHEXANE	0.96
	2-METHYLPENTANE	0.87
	23-DIMETHYLPENTANE	0.94
	2-METHYLHEPTANE	0.87
2-METHYLHEXANE	2-METHYLPENTANE	0.88
	23-DIMETHYLPENTANE	0.96
	2-METHYLHEPTANE	0.9
23-DIMETHYLBUTANE	2-METHYLPENTANE	0.96
	23-DIMETHYLPENTANE	0.86
2-METHYLPENTANE	23-DIMETHYLPENTANE	0.89
23-DIMETHYLPENTANE	2-METHYLHEPTANE	0.9
M&P-XYLENE	O-XYLENE	0.9
ETHYLBENZENE	O-XYLENE	0.92
135-TRI-M-BENZENE	N-PROPYLBENZENE	0.92
	O-ETHYLTOLUENE	0.87
	M-ETHYLTOLUENE	0.92
	P-ETHYLTOLUENE	0.94
	P-DIETHYLBENZENE	0.91
N-PROPYLBENZENE	O-ETHYLTOLUENE	0.86
	M-ETHYLTOLUENE	0.89
	P-ETHYLTOLUENE	0.93
	P-DIETHYLBENZENE	0.88
M-ETHYLTOLUENE	P-ETHYLTOLUENE	0.88
	P-DIETHYLBENZENE	0.88
	123-TRI-M-BENZENE	0.88
P-ETHYLTOLUENE	P-DIETHYLBENZENE	0.89

Table 6*e.* Species pairs with apparent high correlation coefficient at *Lake Jackson* site

PROPANE	N-BUTANE	0.88
	ISO-BUTANE	0.95
	N-PENTANE	0.91
	ISO-PENTANE	0.94
	N-HEXANE	0.88
	CYCLOPENTANE	0.93
	METHYLCYCLOPENTANE	0.87
	2-METHYLPENTANE	0.86
N-BUTANE	ISO-BUTANE	0.85
	N-PENTANE	0.99
	ISO-PENTANE	0.94
	N-HEXANE	0.91
	CYCLOPENTANE	0.9
ISO-BUTANE	N-PENTANE	0.89
	ISO-PENTANE	0.96
	N-HEXANE	0.9
	CYCLOPENTANE	0.94
	3-METHYLHEXANE	0.85
	METHYLCYCLOPENTANE	0.89
	2-METHYLHEXANE	0.86
	2-METHYLPENTANE	0.87
TRANS-2-BUTENE	CIS-2-BUTENE	0.9
CIS-2-BUTENE	TRANS-2-PENTENE	0.86
	CIS-2-PENTENE	0.85
N-PENTANE	ISO-PENTANE	0.94
	N-HEXANE	0.93
	CYCLOPENTANE	0.92
ISO-PENTANE	N-HEXANE	0.92
	CYCLOPENTANE	0.98
	3-METHYLHEXANE	0.85
	METHYLCYCLOPENTANE	0.87
	2-METHYLHEXANE	0.86
	2-METHYLPENTANE	0.85
TRANS-2-PENTENE	CIS-2-PENTENE	0.95
3-METHYLPENTANE	N-HEXANE	0.93
	CYCLOPENTANE	0.85
N-HEXANE	CYCLOPENTANE	0.93
	METHYLCYCLOPENTANE	0.85
	2-METHYLPENTANE	0.87
N-HEPTANE	CYCLOPENTANE	0.87
	3-METHYLHEXANE	0.85
	2-METHYLHEXANE	0.86

N-OCTANE	N-NONANE	0.96
	3-METHYLHEPTANE	0.93
	2-METHYLHEPTANE	0.94
N-NONANE	3-METHYLHEPTANE	0.91
	2-METHYLHEPTANE	0.92
CYCLOPENTANE	3-METHYLHEXANE	0.89
	METHYLCYCLOPENTANE	0.91
	2-METHYLHEXANE	0.9
	2-METHYLPENTANE	0.92
	23-DIMETHYLPENTANE	0.87
22-DIMETHYLBUTANE	METHYLCYCLOHEXANE	0.9
	METHYLCYCLOPENTANE	0.89
	2-METHYLHEXANE	0.85
24-DIMETHYLPENTANE	23-DIMETHYLPENTANE	0.86
CYCLOHEXANE	METHYLCYCLOHEXANE	0.9
	METHYLCYCLOPENTANE	0.89
3-METHYLHEXANE	METHYLCYCLOPENTANE	0.93
	2-METHYLHEXANE	0.99
	23-DIMETHYLBUTANE	0.87
	2-METHYLPENTANE	0.89
	23-DIMETHYLPENTANE	0.96
224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.9
3-METHYLHEPTANE	2-METHYLHEPTANE	0.97
METHYLCYCLOHEXANE	METHYLCYCLOPENTANE	0.92
	2-METHYLHEXANE	0.86
METHYLCYCLOPENTANE	2-METHYLHEXANE	0.94
	23-DIMETHYLBUTANE	0.91
	2-METHYLPENTANE	0.9
	23-DIMETHYLPENTANE	0.93
2-METHYLHEXANE	23-DIMETHYLBUTANE	0.87
	2-METHYLPENTANE	0.89
	23-DIMETHYLPENTANE	0.97
23-DIMETHYLBUTANE	2-METHYLPENTANE	0.89
	23-DIMETHYLPENTANE	0.87
2-METHYLPENTANE	23-DIMETHYLPENTANE	0.87
M&P-XYLENE	TOLUENE	0.85
	124-TRI-M-BENZENE	0.88
BENZENE	TOLUENE	0.89
TOLUENE	O-XYLENE	0.89
O-XYLENE	135-TRI-M-BENZENE	0.86
124-TRI-M-BENZENE	N-PROPYLBENZENE	0.96

Table 6*f*. Species pairs with apparent high correlation coefficient at *Lynchburg Ferry* site

TRANS-2-PENTENE	CIS-2-PENTENE	0.97
3-METHYLPENTANE	2-METHYLHEXANE	0.86
	23-DIMETHYLBUTANE	0.9
	2-METHYLPENTANE	0.94
	23-DIMETHYLPENTANE	0.85
N-HEPTANE	N-OCTANE	0.93
	24-DIMETHYLPENTANE	0.92
	3-METHYLHEXANE	0.97
	3-METHYLHEPTANE	0.95
	2-METHYLHEXANE	0.97
	23-DIMETHYLPENTANE	0.97
	2-METHYLHEPTANE	0.97
N-OCTANE	3-METHYLHEXANE	0.86
	3-METHYLHEPTANE	0.89
	2-METHYLHEXANE	0.86
	23-DIMETHYLPENTANE	0.86
	2-METHYLHEPTANE	0.92
N-DECANE	N-UNDECANE	0.85
24-DIMETHYLPENTANE	3-METHYLHEXANE	0.94
	3-METHYLHEPTANE	0.89
	2-METHYLHEXANE	0.94
	23-DIMETHYLPENTANE	0.97
	2-METHYLHEPTANE	0.9
3-METHYLHEXANE	3-METHYLHEPTANE	0.98
	2-METHYLHEXANE	1
	23-DIMETHYLPENTANE	0.99
	2-METHYLHEPTANE	0.97
224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.91
3-METHYLHEPTANE	2-METHYLHEXANE	0.97
	23-DIMETHYLPENTANE	0.96
	2-METHYLHEPTANE	0.99
2-METHYLHEXANE	23-DIMETHYLPENTANE	0.99
	2-METHYLHEPTANE	0.96
23-DIMETHYLBUTANE	2-METHYLPENTANE	0.94
23-DIMETHYLPENTANE	2-METHYLHEPTANE	0.96
135-TRI-M-BENZENE	124-TRI-M-BENZENE	0.9
	P-ETHYLTOLUENE	0.92
	P-DIETHYLBENZENE	0.86
124-TRI-M-BENZENE	P-ETHYLTOLUENE	0.86
N-PROPYLBENZENE	M-ETHYLTOLUENE	0.95
	P-ETHYLTOLUENE	0.93
M-ETHYLTOLUENE	P-ETHYLTOLUENE	0.87

Table 6*g*. Species pairs with apparent high correlation coefficient at *Mustang* site

ETHANE	PROPANE	0.92
	N-BUTANE	0.88
	N-PENTANE	0.91
	ISO-PENTANE	0.87
	METHYLCYCLOPENTANE	0.85
ETHYLENE	N-HEPTANE	0.85
	STYRENE	0.85
PROPANE	N-BUTANE	0.98
	ISO-BUTANE	0.91
	N-PENTANE	0.95
	ISO-PENTANE	0.96
N-BUTANE	ISO-BUTANE	0.9
	N-PENTANE	0.95
	ISO-PENTANE	0.97
	3-METHYLPENTANE	0.86
ISO-BUTANE	N-PENTANE	0.93
	ISO-PENTANE	0.88
N-PENTANE	ISO-PENTANE	0.95
	3-METHYLPENTANE	0.91
	N-HEXANE	0.88
	CYCLOPENTANE	0.9
	22-DIMETHYLBUTANE	0.9
	2-METHYLPENTANE	0.9
	BENZENE	0.87
ISO-PENTANE	3-METHYLPENTANE	0.9
	N-HEXANE	0.88
	CYCLOPENTANE	0.93
	22-DIMETHYLBUTANE	0.94
	2-METHYLPENTANE	0.94
3-METHYLPENTANE	CYCLOHEXANE	0.9
	METHYLCYCLOPENTANE	0.9
	2-METHYLHEXANE	0.85
	BENZENE	0.86
N-HEXANE	METHYLCYCLOPENTANE	0.87
N-HEPTANE	3-METHYLHEXANE	0.9
	STYRENE	0.85
N-OCTANE	N-NONANE	0.92
	N-DECANE	0.89
	3-METHYLHEPTANE	0.95
	2-METHYLHEPTANE	0.93
	O-XYLENE	0.91
	135-TRI-M-BENZENE	0.92
	P-ETHYLTOLUENE	0.87
N-NONANE	N-DECANE	0.99
	3-METHYLHEPTANE	0.86
	2-METHYLHEPTANE	0.86
	O-XYLENE	0.9
	135-TRI-M-BENZENE	0.98
	O-ETHYLTOLUENE	0.86
	P-ETHYLTOLUENE	0.94
N-DECANE	O-XYLENE	0.88
	135-TRI-M-BENZENE	0.98
	O-ETHYLTOLUENE	0.86
	P-ETHYLTOLUENE	0.94
CYCLOHEXANE	METHYLCYCLOPENTANE	0.97
	2-METHYLHEXANE	0.92
	BENZENE	0.86
3-METHYLHEXANE	23-DIMETHYLPENTANE	0.86
	ETHYLBENZENE	0.89
3-METHYLHEPTANE	METHYLCYCLOHEXANE	0.85
	2-METHYLHEXANE	0.91
	2-METHYLHEPTANE	0.99
	O-XYLENE	0.91
	135-TRI-M-BENZENE	0.86
METHYLCYCLOHEXANE	2-METHYLHEXANE	0.87
	2-METHYLHEPTANE	0.88
METHYLCYCLOPENTANE	2-METHYLHEXANE	0.92
	BENZENE	0.86
2-METHYLHEXANE	2-METHYLPENTANE	0.85
	2-METHYLHEPTANE	0.9
	TOLUENE	0.87
2-METHYLHEPTANE	O-XYLENE	0.91
	135-TRI-M-BENZENE	0.87
BENZENE	TOLUENE	0.86
TOLUENE	O-XYLENE	0.87
O-XYLENE	135-TRI-M-BENZENE	0.95
	O-ETHYLTOLUENE	0.93
	P-ETHYLTOLUENE	0.9
135-TRI-M-BENZENE	O-ETHYLTOLUENE	0.93
	P-ETHYLTOLUENE	0.97
O-ETHYLTOLUENE	P-ETHYLTOLUENE	0.92
M-ETHYLTOLUENE	123-TRI-M-BENZENE	0.91

Table 6*h*. Species pairs with apparent high correlation coefficient at *Texas City* site

ETHANE	PROPANE	0.95
	N-BUTANE	0.87
	ISO-BUTANE	0.91
	22-DIMETHYLBUTANE	0.85
PROPANE	N-BUTANE	0.96
	ISO-BUTANE	0.96
	3-METHYLPENTANE	0.86
	CYCLOPENTANE	0.89
	22-DIMETHYLBUTANE	0.9
	CYCLOHEXANE	0.85
	METHYLCYCLOPENTANE	0.85
	23-DIMETHYLBUTANE	0.9
	2-METHYLPENTANE	0.9
N-BUTANE	ISO-BUTANE	0.95
	3-METHYLPENTANE	0.86
	CYCLOPENTANE	0.89
	22-DIMETHYLBUTANE	0.89
	23-DIMETHYLBUTANE	0.9
	2-METHYLPENTANE	0.9
ISO-BUTANE	3-METHYLPENTANE	0.86
	CYCLOPENTANE	0.89
	22-DIMETHYLBUTANE	0.9
	METHYLCYCLOPENTANE	0.85
	23-DIMETHYLBUTANE	0.91
	2-METHYLPENTANE	0.89
CIS-2-BUTENE	22-DIMETHYLBUTANE	0.86
	23-DIMETHYLBUTANE	0.85
1-PENTENE	TRANS-2-PENTENE	0.9
	CIS-2-PENTENE	0.91
TRANS-2-PENTENE	CIS-2-PENTENE	0.99
3-METHYLPENTANE	N-HEXANE	0.96
	N-HEPTANE	0.92
	CYCLOPENTANE	0.96
	22-DIMETHYLBUTANE	0.93
	CYCLOHEXANE	0.92
	3-METHYLHEXANE	0.96
	METHYLCYCLOHEXANE	0.88
	METHYLCYCLOPENTANE	0.98
	2-METHYLHEXANE	0.96
	23-DIMETHYLBUTANE	0.97
	2-METHYLPENTANE	0.99
	23-DIMETHYLPENTANE	0.97
N-HEXANE	N-HEPTANE	0.9
	CYCLOPENTANE	0.94
	22-DIMETHYLBUTANE	0.86
	CYCLOHEXANE	0.91
	3-METHYLHEXANE	0.94
	METHYLCYCLOHEXANE	0.85
	METHYLCYCLOPENTANE	0.96
	2-METHYLHEXANE	0.93
	23-DIMETHYLBUTANE	0.92
	2-METHYLPENTANE	0.95
	23-DIMETHYLPENTANE	0.94
N-HEPTANE	N-OCTANE	0.92
	CYCLOPENTANE	0.87
	CYCLOHEXANE	0.92
	3-METHYLHEXANE	0.98
	3-METHYLHEPTANE	0.95
	METHYLCYCLOHEXANE	0.95
	METHYLCYCLOPENTANE	0.94
	2-METHYLHEXANE	0.97
	23-DIMETHYLBUTANE	0.87
	2-METHYLPENTANE	0.9
	23-DIMETHYLPENTANE	0.97
	2-METHYLHEPTANE	0.94
N-OCTANE	N-NONANE	0.92
	3-METHYLHEXANE	0.89
	3-METHYLHEPTANE	0.97
	METHYLCYCLOHEXANE	0.92
	2-METHYLHEXANE	0.88
	23-DIMETHYLPENTANE	0.87
	2-METHYLHEPTANE	0.96
N-NONANE	N-DECANE	0.93
	3-METHYLHEPTANE	0.91
	2-METHYLHEPTANE	0.89
N-DECANE	N-UNDECANE	0.87
	135-TRI-M-BENZENE	0.86
CYCLOPENTANE	22-DIMETHYLBUTANE	0.92
	CYCLOHEXANE	0.91
	3-METHYLHEXANE	0.92
	METHYLCYCLOHEXANE	0.87
	METHYLCYCLOPENTANE	0.95
	2-METHYLHEXANE	0.93
	23-DIMETHYLBUTANE	0.95
	2-METHYLPENTANE	0.97
	23-DIMETHYLPENTANE	0.93
22-DIMETHYLBUTANE	CYCLOHEXANE	0.88
	3-METHYLHEXANE	0.87
	METHYLCYCLOPENTANE	0.92
	2-METHYLHEXANE	0.9
	23-DIMETHYLBUTANE	0.98
	2-METHYLPENTANE	0.95
	23-DIMETHYLPENTANE	0.88
CYCLOHEXANE	3-METHYLHEXANE	0.94
	3-METHYLHEPTANE	0.86
	METHYLCYCLOHEXANE	0.95
	METHYLCYCLOPENTANE	0.94
	2-METHYLHEXANE	0.95
	23-DIMETHYLBUTANE	0.91
	2-METHYLPENTANE	0.91
	23-DIMETHYLPENTANE	0.94
	2-METHYLHEPTANE	0.86
3-METHYLHEXANE	3-METHYLHEPTANE	0.92
	METHYLCYCLOHEXANE	0.94
	METHYLCYCLOPENTANE	0.97
	2-METHYLHEXANE	0.99
	23-DIMETHYLBUTANE	0.92
	2-METHYLPENTANE	0.94
	23-DIMETHYLPENTANE	0.99
	2-METHYLHEPTANE	0.91
224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.9
3-METHYLHEPTANE	METHYLCYCLOHEXANE	0.94
	2-METHYLHEXANE	0.92
	23-DIMETHYLPENTANE	0.9
	2-METHYLHEPTANE	0.98
METHYLCYCLOHEXANE	METHYLCYCLOPENTANE	0.91
	2-METHYLHEXANE	0.96
	23-DIMETHYLBUTANE	0.85
	2-METHYLPENTANE	0.86
	23-DIMETHYLPENTANE	0.95
	2-METHYLHEPTANE	0.93
METHYLCYCLOPENTANE	2-METHYLHEXANE	0.97
	23-DIMETHYLBUTANE	0.96
	2-METHYLPENTANE	0.98
	23-DIMETHYLPENTANE	0.97
2-METHYLHEXANE	23-DIMETHYLBUTANE	0.94
	2-METHYLPENTANE	0.95
	23-DIMETHYLPENTANE	0.99
	2-METHYLHEPTANE	0.91
23-DIMETHYLBUTANE	2-METHYLPENTANE	0.99
	23-DIMETHYLPENTANE	0.93
2-METHYLPENTANE	23-DIMETHYLPENTANE	0.95
23-DIMETHYLPENTANE	2-METHYLHEPTANE	0.9
M&P-XYLENE	ETHYLBENZENE	0.88
	O-XYLENE	0.92
	135-TRI-M-BENZENE	0.89
BENZENE	TOLUENE	0.97
	O-XYLENE	0.88
TOLUENE	ETHYLBENZENE	0.87
	O-XYLENE	0.94
	135-TRI-M-BENZENE	0.9
	123-TRI-M-BENZENE	0.86
ETHYLBENZENE	O-XYLENE	0.91
	135-TRI-M-BENZENE	0.87
O-XYLENE	135-TRI-M-BENZENE	0.92
	123-TRI-M-BENZENE	0.85
135-TRI-M-BENZENE	124-TRI-M-BENZENE	0.9
	O-ETHYLTOLUENE	0.88
124-TRI-M-BENZENE	O-ETHYLTOLUENE	0.9
	123-TRI-M-BENZENE	0.85
M-ETHYLTOLUENE	123-TRI-M-BENZENE	0.86

Table 6*i*. Species pairs with apparent high correlation coefficient at *WallisVille* site

ETHANE	PROPANE	0.92
N-PENTANE	3-METHYLPENTANE	0.94
	CYCLOPENTANE	0.97
	22-DIMETHYLBUTANE	0.88
	METHYLCYCLOPENTANE	0.88
	2-METHYLHEXANE	0.85
	23-DIMETHYLBUTANE	0.9
	2-METHYLPENTANE	0.94
	23-DIMETHYLPENTANE	0.85
TRANS-2-PENTENE	CIS-2-PENTENE	0.98
3-METHYLPENTANE	N-HEPTANE	0.88
	CYCLOPENTANE	0.94
	22-DIMETHYLBUTANE	0.88
	3-METHYLHEXANE	0.91
	METHYLCYCLOPENTANE	0.9
	2-METHYLHEXANE	0.92
	23-DIMETHYLBUTANE	0.95
	2-METHYLPENTANE	0.98
	23-DIMETHYLPENTANE	0.9
N-HEPTANE	CYCLOPENTANE	0.86
	3-METHYLHEXANE	0.95
	2-METHYLHEXANE	0.95
	23-DIMETHYLBUTANE	0.86
	2-METHYLPENTANE	0.9
	23-DIMETHYLPENTANE	0.9
	2-METHYLHEPTANE	0.85
N-OCTANE	3-METHYLHEPTANE	0.87
	2-METHYLHEPTANE	0.88
N-NONANE	N-DECANE	0.89
	N-UNDECANE	0.91
N-DECANE	N-UNDECANE	0.92
CYCLOPENTANE	22-DIMETHYLBUTANE	0.88
	3-METHYLHEXANE	0.86
	METHYLCYCLOPENTANE	0.87
	2-METHYLHEXANE	0.88
	23-DIMETHYLBUTANE	0.92
	2-METHYLPENTANE	0.95
	23-DIMETHYLPENTANE	0.86
22-DIMETHYLBUTANE	23-DIMETHYLBUTANE	0.91
	2-METHYLPENTANE	0.91
24-DIMETHYLPENTANE	23-DIMETHYLBUTANE	0.85
	23-DIMETHYLPENTANE	0.85
3-METHYLHEXANE	2-METHYLHEXANE	0.98
	23-DIMETHYLBUTANE	0.9
	2-METHYLPENTANE	0.92
	23-DIMETHYLPENTANE	0.96
224-TRIMETHYLPENTA	234-TRIMETHYLPENTA	0.92
	23-DIMETHYLPENTANE	0.86
3-METHYLHEPTANE	2-METHYLHEPTANE	0.95
METHYLCYCLOPENTANE	2-METHYLHEXANE	0.86
	23-DIMETHYLBUTANE	0.85
	2-METHYLPENTANE	0.88
2-METHYLHEXANE	23-DIMETHYLBUTANE	0.91
	2-METHYLPENTANE	0.93
	23-DIMETHYLPENTANE	0.97
23-DIMETHYLBUTANE	2-METHYLPENTANE	0.97
	23-DIMETHYLPENTANE	0.91
2-METHYLPENTANE	23-DIMETHYLPENTANE	0.91
M&P-XYLENE	ETHYLBENZENE	0.91
	O-XYLENE	0.97
ETHYLBENZENE	O-XYLENE	0.91
135-TRI-M-BENZENE	124-TRI-M-BENZENE	0.96
	N-PROPYLBENZENE	0.89
	O-ETHYLTOLUENE	0.96
	M-ETHYLTOLUENE	0.85
	P-DIETHYLBENZENE	0.86
124-TRI-M-BENZENE	N-PROPYLBENZENE	0.9
	O-ETHYLTOLUENE	0.96
	M-ETHYLTOLUENE	0.87
	P-DIETHYLBENZENE	0.87
N-PROPYLBENZENE	O-ETHYLTOLUENE	0.92
	M-ETHYLTOLUENE	0.85
	P-DIETHYLBENZENE	0.86
O-ETHYLTOLUENE	M-ETHYLTOLUENE	0.86
	P-DIETHYLBENZENE	0.86

Table 7. Categories of species for comparing with inventories

Index	Category	Abbr.	VOCs
1	1,3-butadiene	BUTA13	1,3-butadiene
2	aceytlene	aceyt	aceytlene
3	benzene	benzene	benzene
4	C10 Paraffins	C10P	n-decane
5	C11 Paraffins	C11P	n-undecane
6	C3/C4/C5 alkylBenzene	alkylBZ	1,2,4-triMethylBenzene, 1,2,4-triMethylBenzene, 1,2,3-triMethylBenzene, n-propylBenzene, iso-propylBenzene, m-dimethylBenzene, p-dimethylBenzene, o,p,m-ethylToluene
7	C4 Olefins	C4O	1-butene, cis-2-butene, trans-2-butene
8	C5 Olefins	C5O	1-pentene, cis-2-pentene, trans-2-pentene
9	C5 Paraffins	C5P	n-pentane, iso-pentane, cycloPentane
10	C6 Paraffins	C6P	n-hexane, cycloHexane, methylCycloPentane,2-methylPentane, 3-methylPentane, 2,2-dimethylButane, 23-dimethylButane
11	C7 Paraffins	C7P	n-heptane, 2-methylHexane, 3-methylHexane, methylCycloHexane, 2,3-dimethylPentane, 2,4-dimethylPentane
12	C8 Paraffins	C8P	n-octane, 2-methylHeptane, 3-methylHeptane, 2,2,4-trimethylPentane, 2,3,4-triMethylPentane
13	C9 Paraffins	C9P	n-nonane
14	ethylBenzene	EBENZ	ethylBenzene
15	ethylene	ethylene	ethylene
16	iso-butane	i-BUTA	iso-butane
17	isoprene	ispre	isoprene
18	n-butane	n-BUTA	n-butane
19	propylene	prope	propylene
20	toluene	toluene	toluene
21	xylenes	xylenes	m,p,o-xylene

Table 8. Categories of the Inventories

Index	**Category**	**VOCs**
1	Propylene	propylene
2	Ethene	Ethene
3	C2C3	Ethane, propane, acetylene
4	Butadiene	Butadiene
5	Butenes	1-butene, *cis*-butene, *trans*-butene
6	**Butanes**	***n*-butane, *iso*-butane**
7	**Pentanes**	***n*-pentane, *iso*-pentane**
8	Pentenes	1-pentene, *cis*-pentene, *trans*-pentene, 2-methyl-2-bentene, 3-methyl-1-butene
9	Isoprene	Isoprene
10	C5Cyclos	cycloPentane, cycloPentene
11	**Hexanes**	***n*-hexane,** 2,2-dimethylButane, 2,3-dimethylButane, **2-methylPentane, 3-methylPentane**
12	Hexenes	1-hexene, *cis*-2hexene, *trans*-2hexene, 2-methyl-1-Pentene, 4-methyl-1-Pentene
13	C6Arom	Benzene
14	C6Cyclos	cycloHexane, methylCycloPentane
15	C7C10Arom	Toluene, *p*-ethylToluene, *o*-ethylToluene, *m*-ethylToluene, 1,3,5-trimethylBenzene, 1,2,4-trimethylBenzene, 1,2,3-trimethylBenzene, ethylBenzene, *m*-diethylBenzene, *p*-diethylBenzene, *n*-propylBenzene, *iso*-propylBenzene
16	C7C11Other	*n*-heptane, *n*-octane, *n*-nonane, *n*-decane, *in*-undecane, 2-methylHexane, 3-methylHexane, 2-methylHeptane, 3-methylHeptane, methylCycloHexane, 2,4-dimethylPentane, 2,3-dimethylPentane, 2,2,4-trimethylPentane, 2,3,4-trimethylPentane
17	Styrene	Styrene
18	Xylenes	*m&p*-xylene, *o*-xylene

Table 9. Species excluded from (code 2) and downweighted (code 1) in the PMF analysis

	VOC	VOC	VOC	Clinton (Night)				ChannelView (night)				LynFerry (night)				Wallisville (night)				HadenRoad (night)			
field	code	name	abbrv.	no	no999	no0	PMF#	no	no999	no0	PMF#	no	no999	no0	PMF#	no	no999	no0	PMF#	no	no999	no0	PMF#
6	43202	ETHANE	ethane	891	0	0	0	1034	33	0	0	1063	5	1	0	1090	8	0	0	1134	12	0	0
7	43203	ETHYLENE	ethene	463	428	0	1	1034	33	0	0	1053	5	11	0	1088	8	2	0	1134	12	0	0
8	43204	PROPANE	n_prop	867	24	0	0	1034	33	0	0	1063	5	1	0	1090	8	0	0	1135	11	0	0
9	43205	PROPYLENE	prope	891	0	0	0	1034	33	0	0	1063	5	1	0	1089	8	1	0	1135	11	0	0
10	43206	ACETYLENE	acetyl	889	2	0	0	1033	33	1	0	692	366	11	1	1090	8	0	0	900	246	0	1
11	43212	N-BUTANE	n_buta	891	0	0	0	1034	33	0	0	1063	5	1	0	1090	8	0	0	1135	11	0	0
12	43214	ISO-BUTANE	i_buta	891	0	0	0	1034	33	0	0	1063	5	1	0	1090	8	0	0	1135	11	0	0
13	43216	TRANS-2-BUTENE	t2bute	891	0	0	0	1034	33	0	0	1061	7	1	0	1034	64	0	0	1135	11	0	0
14	43217	CIS-2-BUTENE	c2bute	857	34	0	0	1033	33	1	0	1059	5	5	0	1087	8	3	0	1135	11	0	0
15	43218	1,3-BUTADIENE	bute13	446	445	0	1	522	544	1	1	1053	14	2	0	1089	8	1	0	1135	11	0	0
16	43220	N-PENTANE	n_pent	891	0	0	0	872	195	0	1	820	248	1	1	706	392	0	1	913	233	0	1
17	43221	ISO-PENTANE	ipenta	891	0	0	0	1033	33	1	0	624	445	0	1	706	392	0	1	913	233	0	1
18	43224	1-PENTENE	pente1	874	16	1	0	855	204	8	1	876	147	46	1	1035	8	55	0	1130	11	5	0
19	43226	TRANS-2-PENTENE	t2pene	891	0	0	0	1013	33	21	0	872	110	87	1	983	8	107	0	1115	11	20	0
20	43227	CIS-2-PENTENE	c2pene	882	8	1	0	1001	33	33	0	863	17	189	1	834	8	256	1	1104	11	31	0
21	43230	3-METHYLPENTANE	pena3m	0	891	0	2	0	1067	0	2	787	272	10	1	1089	8	1	0	1124	15	7	0
22	43231	N-HEXANE	n_hex	890	1	0	0	1067	0	0	0	1005	47	17	0	1097	1	0	0	1146	0	0	0
23	43232	N-HEPTANE	n_hept	884	1	6	0	876	190	1	1	1025	17	27	0	1090	0	8	0	1146	0	0	0
24	43233	N-OCTANE	n_oct	890	1	0	0	1067	0	0	0	1017	17	35	0	1082	0	16	0	1145	0	1	0
25	43235	N-NONANE	n_non	494	397	0	1	1067	0	0	0	993	22	54	0	1005	0	93	0	1146	0	0	0
26	43238	N-DECANE	n_dec	494	397	0	1	1067	0	0	0	992	72	5	0	1091	0	7	0	1146	0	0	0
27	43242	CYCLOPENTANE	cpenta	890	0	1	0	1034	33	0	0	1062	5	2	0	1090	8	0	0	1129	17	0	0
28	43243	ISOPRENE	i_pren	451	436	4	1	1034	33	0	0	785	272	12	1	1064	9	25	0	1118	19	9	0
29	43244	2,2-DIMETHYLBUTANE	bu22dm	891	0	0	0	875	190	2	1	913	114	42	0	1073	8	17	0	1128	17	1	0
30	43247	2,4-DIMETHYLPENTANE	pen24m	35	1	855	2	7	651	409	2	349	17	703	2	600	0	498	1	1067	0	79	0
31	43248	CYCLOHEXANE	cyhexa	886	1	4	0	1061	0	6	0	978	17	74	0	1079	1	18	0	1145	0	1	0
32	43249	3-METHYLHEXANE	hexa3m	879	1	11	0	1020	41	6	0	972	17	80	0	1089	0	9	0	1146	0	0	0
33	43250	2,2,4-TRIMETHYLPENTA	pa224m	888	1	2	0	1017	41	9	0	742	225	102	1	1084	0	14	0	1139	7	0	0
34	43252	2,3,4-TRIMETHYLPENTA	pa234m	867	1	23	0	1054	0	13	0	859	17	193	1	924	0	174	0	1143	0	3	0
35	43253	3-METHYLHEPTANE	hep3me	881	1	9	0	1065	0	2	0	949	17	103	0	1007	0	91	0	1145	0	1	0
36	43261	METHYLCYCLOHEXANE	mecyhx	890	1	0	0	1067	0	0	0	1037	17	15	0	1094	1	3	0	1146	0	0	0
37	43262	METHYLCYCLOPENTANE	mcypna	886	1	4	0	408	651	8	1	1003	17	49	0	1088	0	10	0	1146	0	0	0
38	43263	2-METHYLHEXANE	hexa2m	628	228	35	1	1020	41	6	0	925	17	127	0	1079	1	18	0	1146	0	0	0
39	43280	1-BUTENE	lbut1e	891	0	0	0	1033	33	1	0	1056	7	6	0	1089	8	1	0	1135	11	0	0
40	43284	2,3-DIMETHYLBUTANE	bu23dm	0	891	0	2	0	1067	0	2	704	358	7	1	1088	8	2	0	1126	15	5	0
41	43285	2-METHYLPENTANE	pena2m	0	891	0	2	0	1067	0	2	794	272	3	1	1090	8	0	0	1130	15	1	0
42	43291	2,3-DIMETHYLPENTANE	pen23m	356	228	307	2	762	143	162	1	584	17	468	1	884	0	214	1	1125	0	21	0
43	43954	N-UNDECANE	n_unde	0	891	0	2	221	845	1	2	939	119	11	0	1025	0	73	0	1146	0	0	0
44	43960	2-METHYLHEPTANE	hep2me	882	1	8	0	1065	0	2	0	954	17	98	0	1033	1	64	0	1146	0	0	0
45	45109	M&P-XYLENE	mp_xyl	494	397	0	1	1067	0	0	0	1047	17	5	0	1097	1	0	0	1146	0	0	0
46	45201	BENZENE	benze	890	1	0	0	1067	0	0	0	1041	17	11	0	1096	2	0	0	1146	0	0	0
47	45202	TOLUENE	tolue	890	1	0	0	1067	0	0	0	1011	57	1	0	1097	1	0	0	1146	0	0	0
48	45203	ETHYLBENZENE	etbz	494	397	0	1	1066	0	1	0	1018	33	18	0	1094	0	4	0	1146	0	0	0
49	45204	O-XYLENE	o_xyl	494	397	0	1	1067	0	0	0	1019	22	28	0	1089	0	9	0	1146	0	0	0
50	45207	1,3,5-TRI-M-BENZENE	bz135m	488	397	6	1	1065	0	2	0	979	17	73	0	947	0	151	1	1141	3	2	0
51	45208	1,2,4-TRI-M-BENZENE	bz124m	463	428	0	1	1067	0	0	0	932	129	8	0	1085	1	12	0	1146	0	0	0
52	45209	N-PROPYLBENZENE	n_prbz	476	397	18	1	1064	0	3	0	867	56	146	1	919	1	178	1	1135	0	11	0
53	45210	ISO-PROPYLBENZENE	iprbz	469	397	25	1	1043	0	24	0	761	30	278	1	703	0	395	1	1103	0	43	0
54	45211	O-ETHYLTOLUENE	o_etol	0	891	0	2	220	845	2	2	884	17	168	1	831	0	267	1	1135	1	10	0
55	45212	M-ETHYLTOLUENE	m_etol	0	891	0	2	0	1067	0	2	1019	17	33	0	862	72	164	1	1117	28	1	0
56	45213	P-ETHYLTOLUENE	p_etol	0	891	0	2	0	1067	0	2	651	17	401	1	371	72	655	2	1035	33	78	0
57	45219	P-DIETHYLBENZENE	detbz2	0	891	0	2	0	1067	0	2	742	285	42	1	825	1	272	1	1125	0	21	0
58	45220	STYRENE	styr	396	397	98	1	1066	0	1	0	940	57	72	0	1019	0	79	0	1141	0	5	0
59	45225	1,2,3-TRI-M-BENZENE	bz123m	494	397	0	1	1067	0	0	0	961	72	36	0	1071	0	27	0	1146	0	0	0
66	45218	M-DIETHYLBENZENE	detbz1	0	891	0	2	0	1067	0	2	197	855	17	2	574	384	140	1	450	696	0	1

Table 10. Numbers of species and *nighttime* samples used in the PMF analyses

Site Name	No of VOCs	no. of Samples
ChannelView	46	1063
Clinton	45	880
HadenRoad	56	1139
Lynchburg Ferry	54	1023
WallisVille	55	1097

Table 11. PMF Results at ChannelView

Factor Index	Major Species	Mean Conc (ppbC)	Mean (%)	Median Conc. (ppbC)	Median (%)	Wind Direction	Possible Sources
1	n-hexane, C6-C8 paraffins, Pentanes, butanes, aromatics	9.05	**2.65**	5.31	2.85	SW, SE	Solvent industry, fuel evaporation, petrochemical
2	Ethene, propylene, 1-butene, 1,3-butadiene, benzene and C4/C5 olefins	58.87	**17.21**	34.17	18.35	SE	Petrochemical industry
3	n-hexane, n-octane, n-Nonane, n-decane, 2methylheptane, 3methylheptane, aromatics	23.30	**6.81**	17.48	9.39	SW, SE	Solvent, Painting Industry / Mobile
4	Acetylene, Aromatics	25.52	**7.46**	13.83	7.43	SW, SE, NE	Vehicle exhaust
5	Ethane, propane, butanes, pentanes, C5/C6 alkanes	172.51	**50.43**	89.25	47.93	S, NE	Accumulations/Natural Gas; Evaporative
6	Styrene, ethylbenzene	21.06	**6.16**	11.52	6.18	NE	Petrochemical plant
7	C4/C5 Olefins; n-butane; pentanes	31.80	**9.30**	14.67	7.88	SW	Fuel evaporative

Table 12. PMF Results at Clinton

Factor Index	Major Species	Mean Conc (ppbC)	Mean (%)	Median Conc. (ppbC)	Median (%)	Wind Direction	Possible Sources
1	Ethene, propylene, 1-butene, 1,3-butadiene, benzene and C4/C5 olefins	47.02	**20.92**	29.51	24.23	NE, S, SE	Petrochemical industry
2	Ethane, propane, butanes, pentanes,	39.60	**17.62**	30.15	24.75	NE	Accumulations/Natural Gas; Evaporative
3	Ethane, propane, butanes, pentanes, c2-/t2-butene, 1-butene	26.53	**11.80**	9.36	7.68	SE	petrochemical / accumulation / Natural Gas / Evaporation
4	n-hexane, Butanes, pentanes, C5-C8 alkanes	36.48	**16.23**	13.31	10.93	NE, S	Solvent
5	C6-C10 alkanes	21.07	**9.38**	11.84	9.72	SE	Solvent, Painting Industry
6	Acetylene, Aromatics	21.95	**9.77**	13.48	11.07	NE	Vehicle exhaust
7	C4/C5 Olefins; n-butane; pentanes	32.11	**14.29**	14.15	11.62	NE, S	Fuel evaporation

Table 13. PMF Results at HadenRoad

Factor Index	Major Species	Mean Conc (ppbC)	Mean (%)	Median Conc. (ppbC)	Median (%)	Wind Direction	Possible Sources
1	C6-C11 alkanes	73.67	**23.29**	44.38	26.32	SW	Solvent, Paint industry
2	Ethane, propane, butanes, pentanes,	66.55	**21.04**	47.56	28.20	NE, SE	Accumulations/Natural Gas; Evaporative
3	Ethene, propylene	26.96	**8.52**	16.29	9.66	NE, SE, S	Petrochemical
4	Acetylene, aromatics	3.85	**1.22**	2.40	1.42	NE	Vehicle Emission
5	1,3-butadiene	35.38	**11.19**	14.51	8.60	NE, SE, SW	Industrial
6	Pentane and branched pentanes	14.41	**4.56**	6.84	4.05	SW	evaporation
7	C4/C5 Olefins; n-butane; pentanes	61.62	**19.48**	12.72	7.55	SW	Fuel evaporation
8	aromatics	33.81	**10.69**	23.93	14.19	S	industry

Table 14. PMF Results at Lynchburg Ferry

Factor Index	Major Species	Mean Conc (ppbC)	Mean (%)	Median Conc. (ppbC)	Median (%)	Wind Direction	Possible Sources
1	Long alkanes and aromatics	38.21	**8.66**	21.36	9.58	SW	Solvent industry
2	Ethene, propylene	95.98	**21.76**	50.71	22.74	S	petrochemical
3	Butane, pentane, C6/C7 paraffins/naphthene	111.28	**25.22**	72.13	32.34	SW	solvent
4	Ethane, propane, butanes, pentanes,	101.31	**22.96**	53.18	23.85	SW	Accumulations/Natural Gas; Evaporative
5	Butane, Pentanes, C4/C5 olefins, C6-C9 alkanes	30.72	**6.96**	9.02	4.05	SW	Evaporative, Solvent chemical/petrochemical
6	C4 olefins, isoprene, benzene, toluene.	63.69	**14.44**	16.61	7.45	SW	petrochemical

Table 15. PMF Results at WallisVille

Factor Index	Major Species	Mean Conc (ppbC)	Mean (%)	Median Conc. (ppbC)	Median (%)	Wind Direction	Possible Sources
1	Ethane, propane, butanes, pentanes	71.96	**35.70**	43.22	38.75	NE	Accumulations/Natural Gas; Evaporative
2	C6-C9 alkanes, aromatics	12.50	**6.20**	8.35	7.49	SW	Solvent / Painting
3	isoprene	7.80	**3.87**	3.70	3.31	NE, SW	biogenic
4	C4 olefins / acetylene / ethane, propane	20.33	**10.09**	20.35	18.24	NE	Industrial / accumulation / natural gas
5	Acetylene, butanes, pentanes, aromatics	20.25	**10.04**	9.80	8.79	SW	Vehicle emission / evaporation
6	Butane, pentane, C4/C5 olefins, aromatics	4.39	**2.18**	3.29	2.95	SW, SE	Fuel evaporative
7	Ethene, propylene, 1-butene	64.35	**31.92**	22.84	20.48	SW	petrochemical

Table 16. Tentative source identification and appointment for the five sites near Ship Channel

		Channelview			Clinton			HadenRoad			Lynchburg Ferry			Wallisville		
Possible Source	**Feature Species**	Factor	Mean(%)	Note	Factor	Mean(%)	Note	Factor	Mean(%)	Note	Factor	Mean(%)	Note	Factor	Mean(%)	Note
Petrochemical	Ethene, propylene, 1-butene, 1,3-butadiene, benzene and C4/C5 olefins	2	**17.21**		1	**20.92**		3	**8.52**		2	**21.76**		7	**31.92**	
Accumulations/Natural Gas; Evaporative	Ethane, propane, butanes, pentanes, C5/C6 alkanes	5	**50.43**		2	**17.62**		2	**21.04**		4	**22.96**		1	**35.70**	
Solvent, Painting Industry	n-hexane, n-octane, n-Nonane, n-decane, 2methylheptane, 3methylheptane, aromatics	3	**6.81**		5	**9.38**		1	**23.29**		1	**8.66**		2	**6.20**	
Vehicle exhaust	Acetylene, Aromatics	4	**7.46**		6	**9.77**		4	**1.22**					5	**10.04**	
Fuel evaporative	n-butane; pentanes ; C4/C5 Olefins;	7	**9.30**		7	**14.29**		7	**19.48**		5	**6.96**		6	**2.18**	
Solvent industry	n-hexane, Pentanes, butanes,	1	**2.65**		4	**16.23**	No benzene/toluene				3	**25.22**	Too much butanes/pentane			
petrochemical / accumulation / Natural Gas / Evaporation	Ethane, propane, butanes, pentanes, c2-/t2-butene, 1-butene				3	**11.80**					6	**14.44**		4	**10.09**	
Petrochemical plant	Styrene, ethylbenzene	6	**6.16**													
Industrial	1,3-butadiene							5	**11.19**							
evaporation	Pentane and branched pentanes							6	**4.56**							
industry	aromatics							8	**10.69**							
biogenic	isoprene													3	**3.87**	

Figure 1. Site Locations

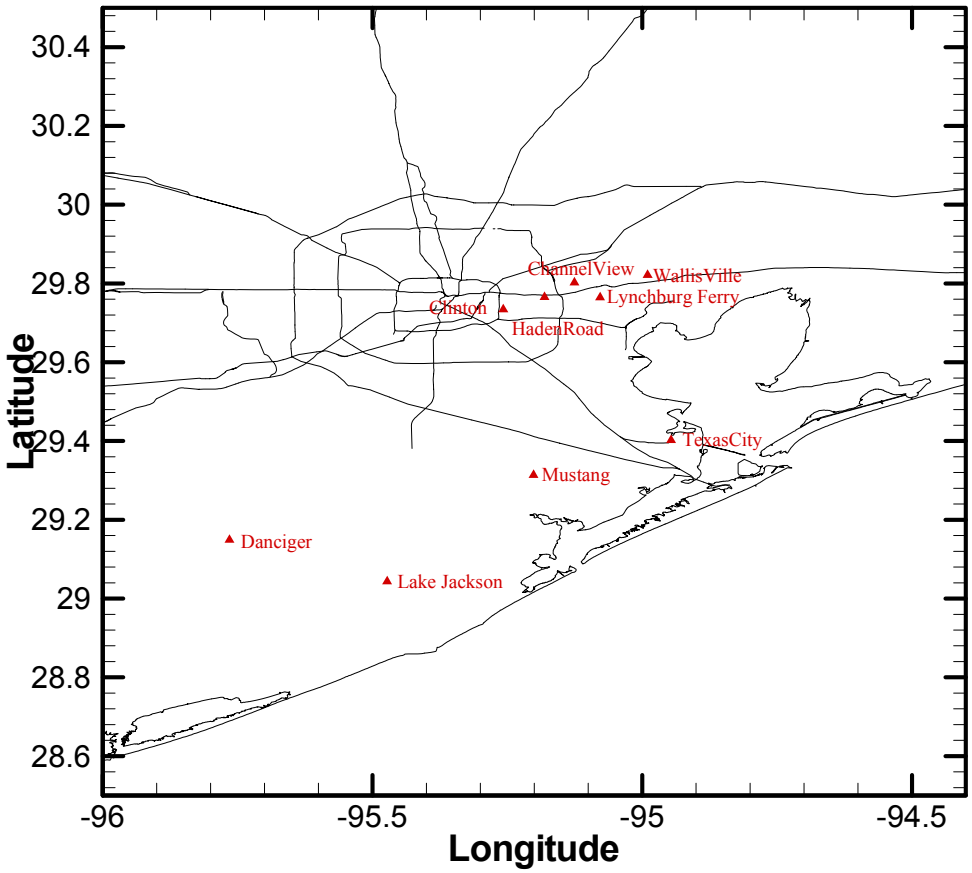


Figure 2. Scatter plots of the percentages of the sum of all the hydrocarbons to the measured *TNMHC* (blue) and its surrogate *sum_pol* (red)

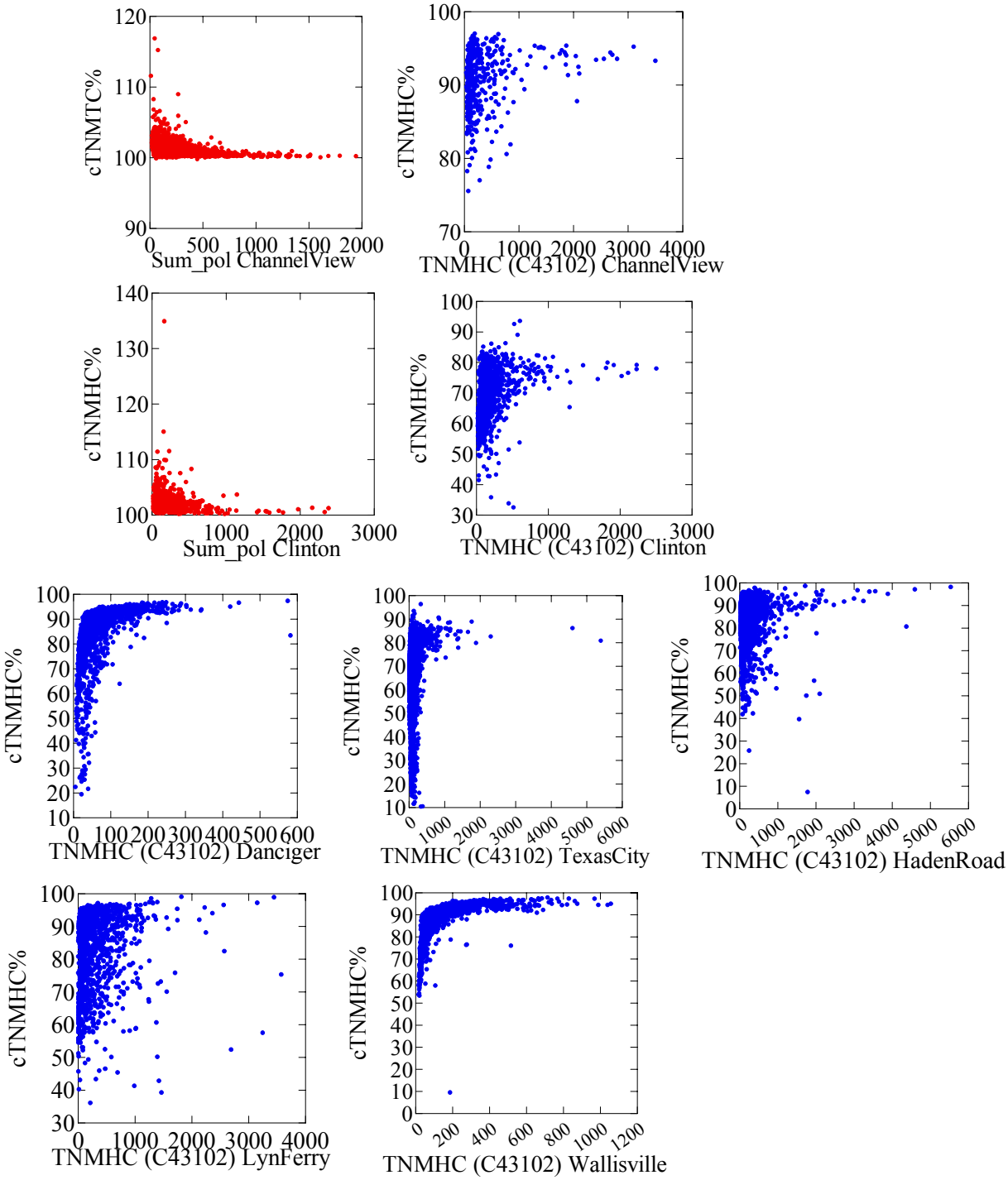


Figure 3. Diurnal patterns of *ethane* and *isoprene* at Haden Road site shown as box plots. The out boundary of the notched box shows the first and third quartile of the data (lower and upper edges, respectively), the point between the notched box denotes the median; The vertical lines extended from the out boundary of the notched box stand for median values plus or minus 1.5 times of the box length (inter quartile range); outliers are denoted by star symbols.

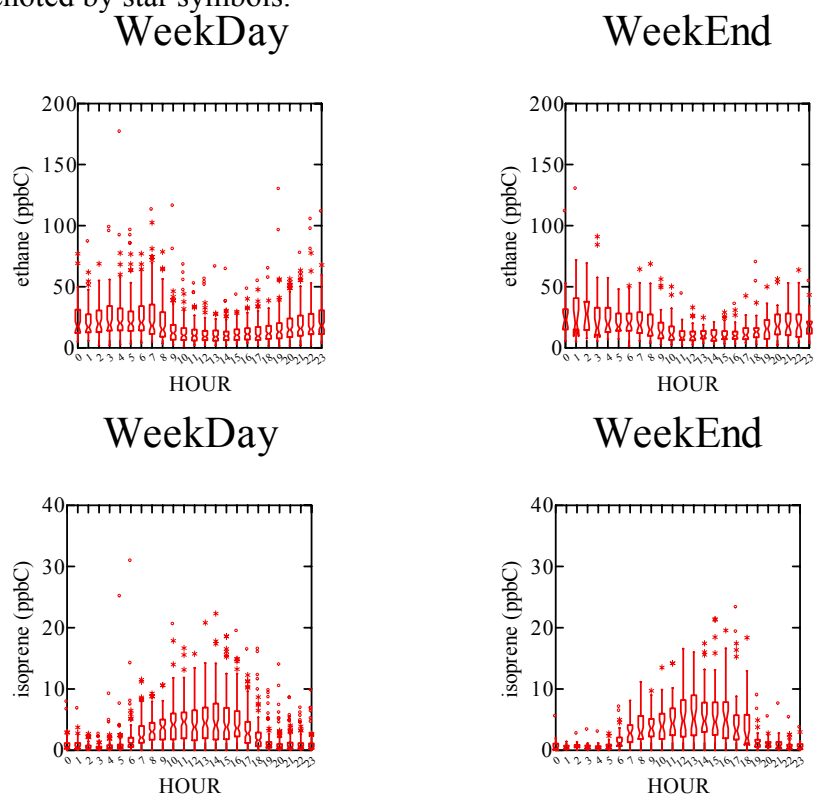


Figure 4*a*. Diurnal pattern of *n-Butane* at all nine sites (solid red: weekday mean, solid green: weekday median, dash read: weekend mean & dash green: weekend median)

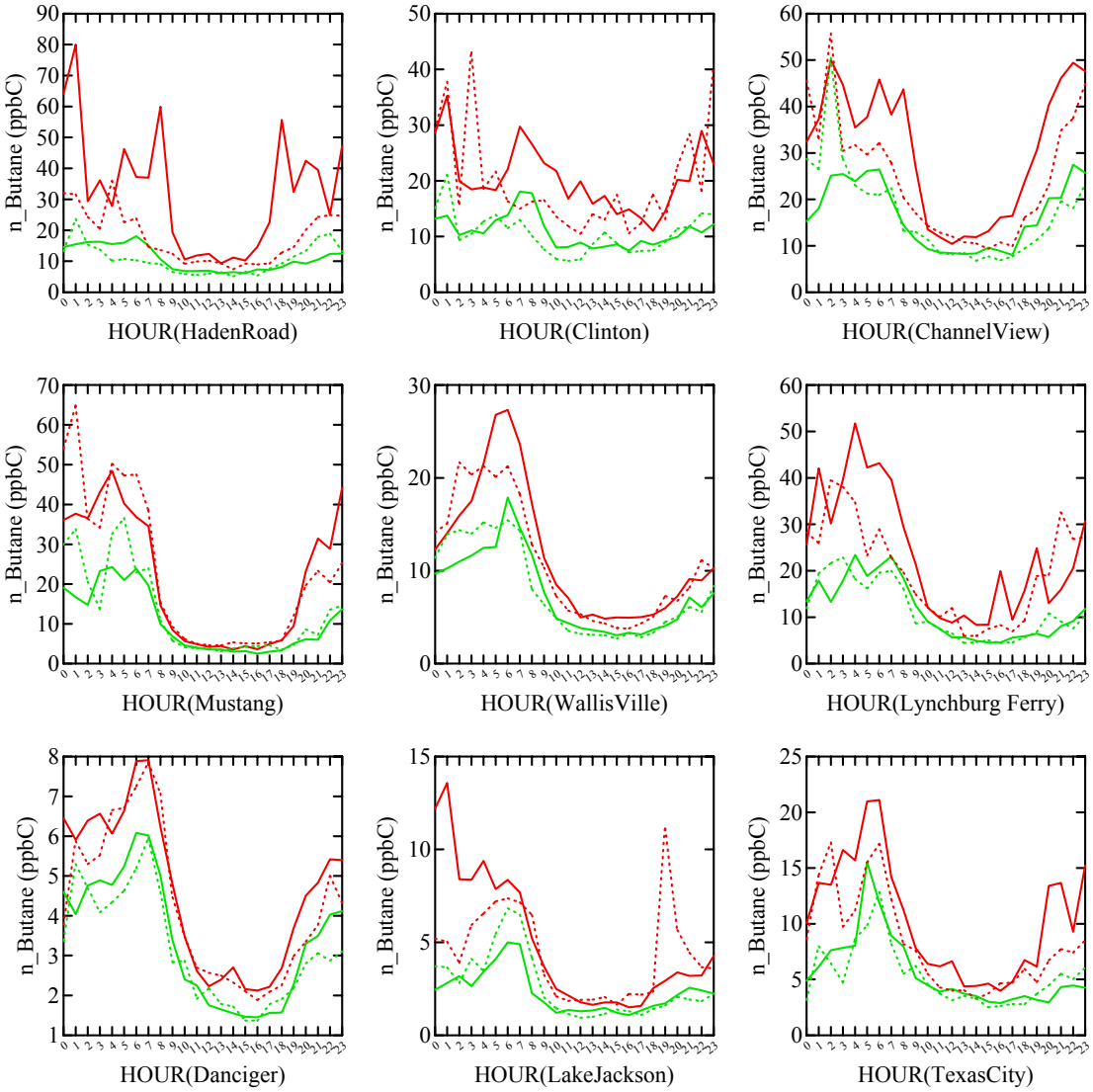


Figure 4*b*. Diurnal pattern of *isoprene* at all nine sites (solid red: weekday mean, solid green: weekday median, dash red: weekend mean & dash green: weekend median)

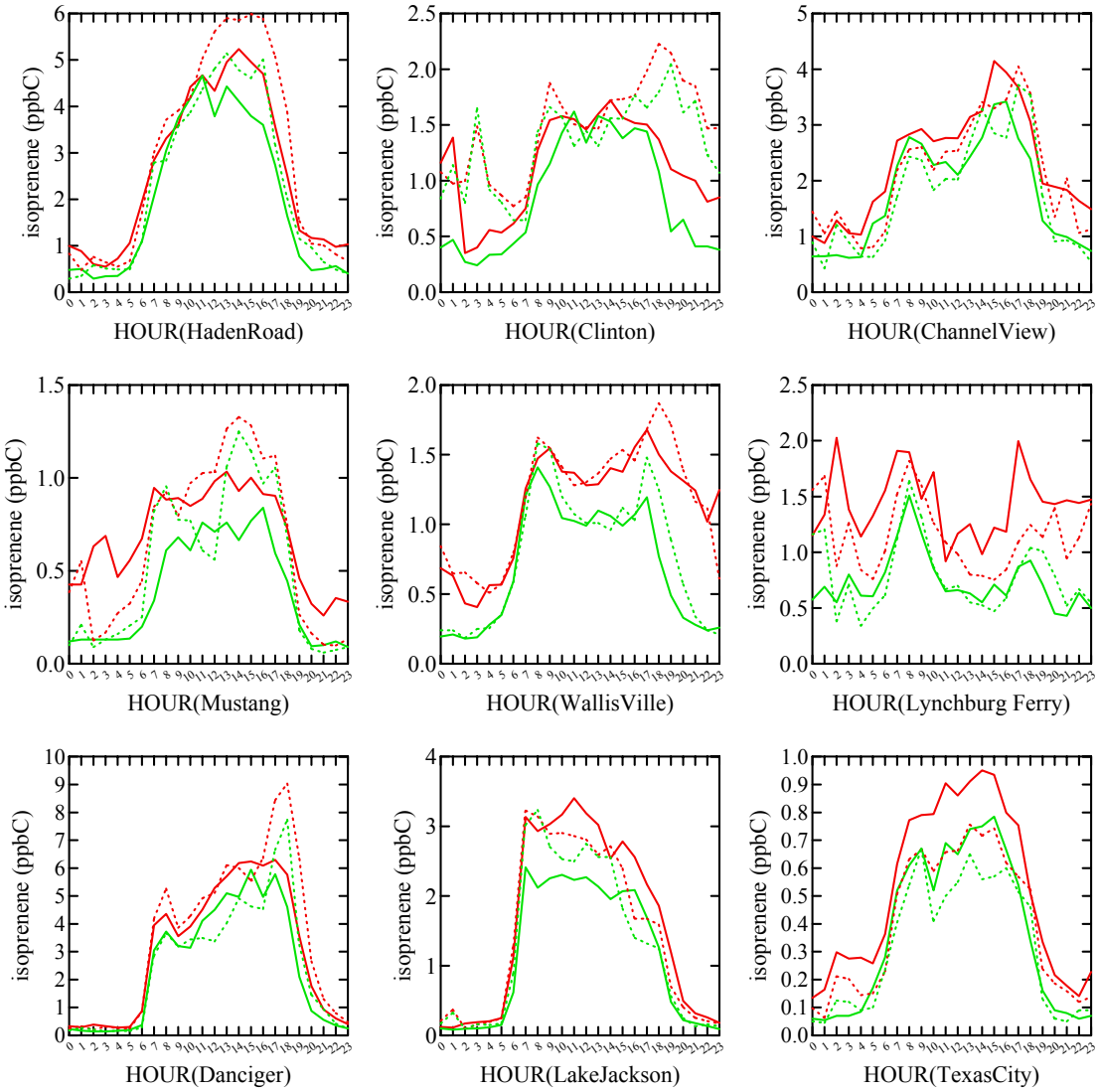


Figure 5. Scatter plot of species vs TNMHC at Haden Road site

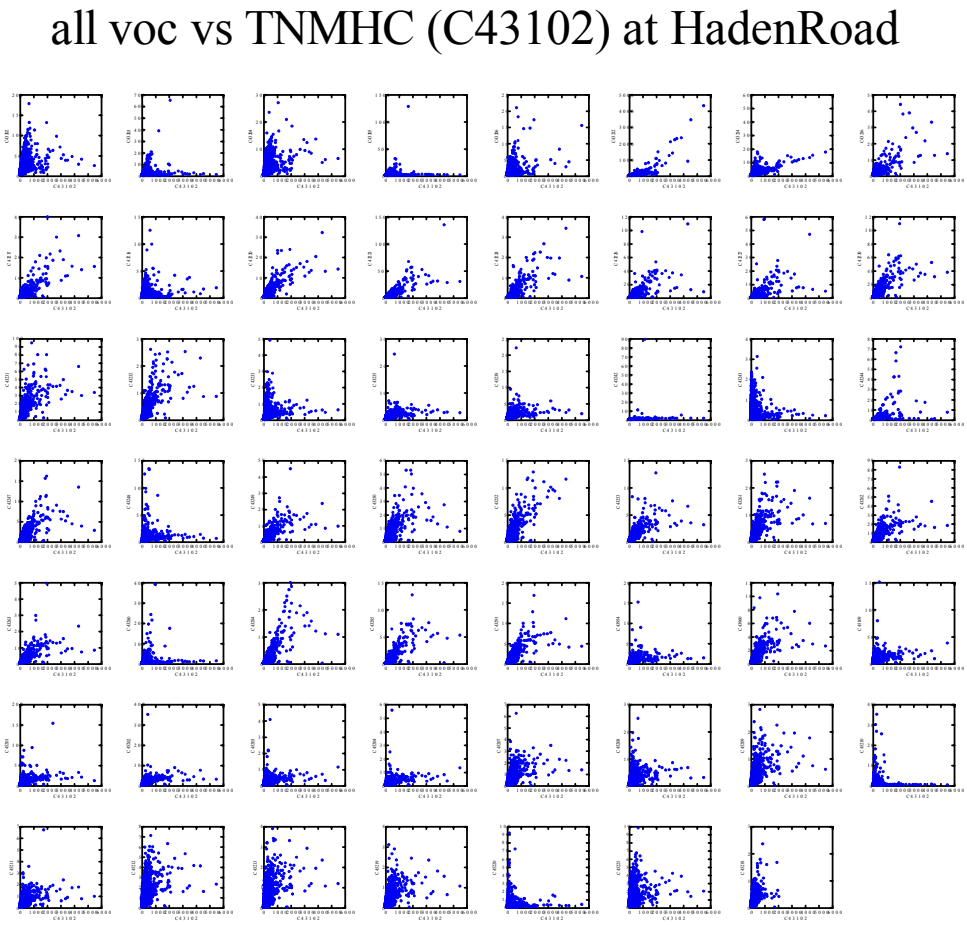


Figure 6. Scatter plot of *acetylene* against *ethane* at Haden Road site

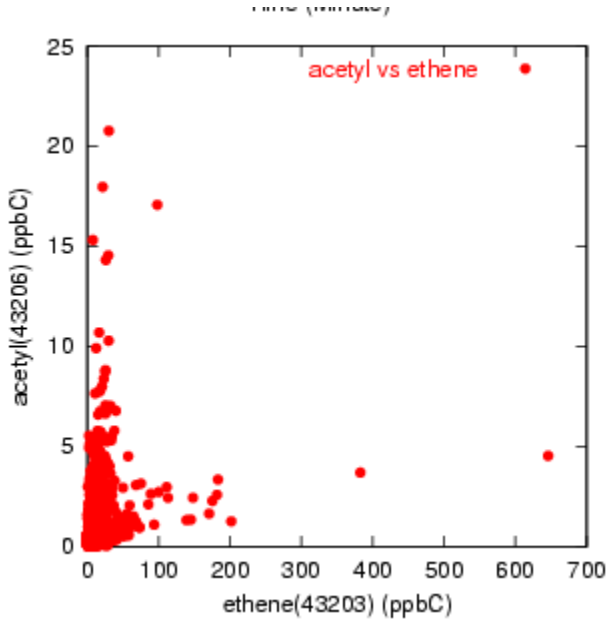


Figure 7. An example of the extreme value: in the time series of *acetylene* at Clinton site

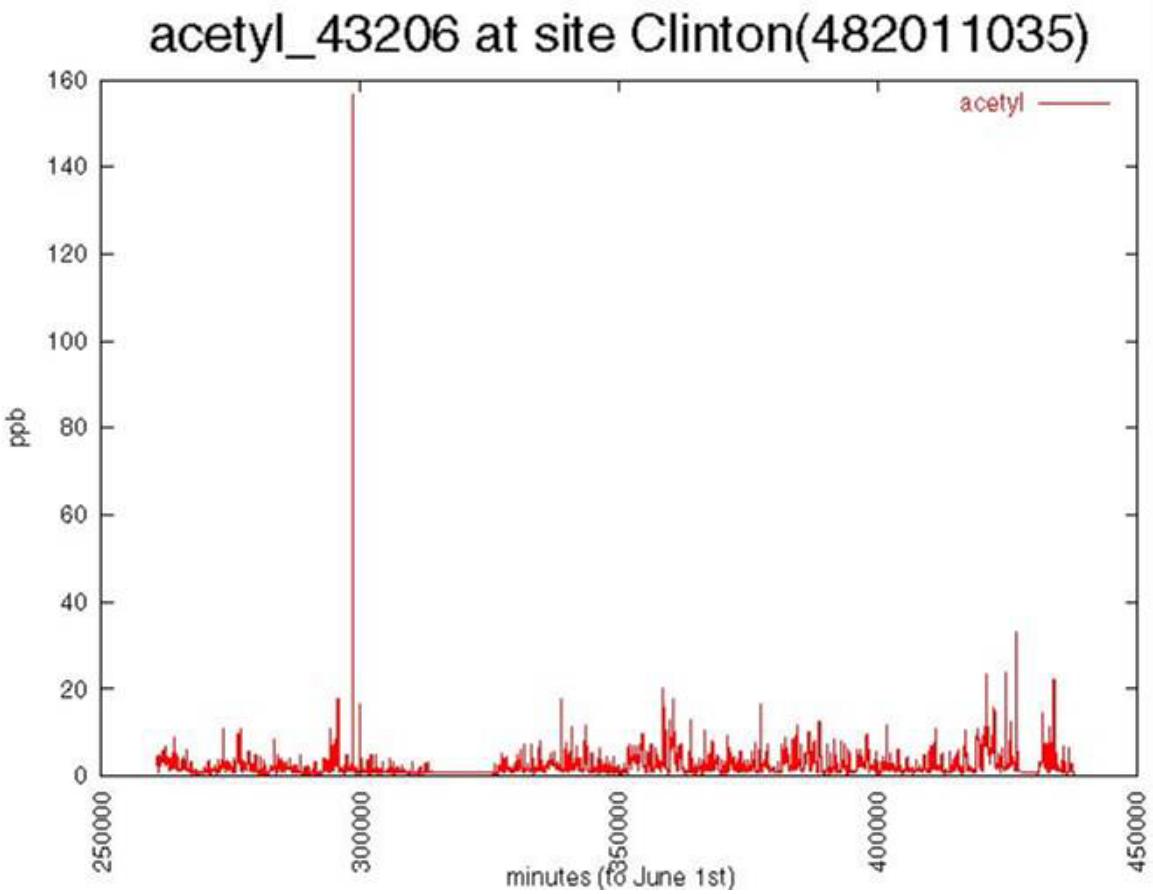


Figure 8: wind trajectories of seasonal extremes of *iso-butane* at five sites near ship channel. The date and time of the trajectories differ from site to site, as indicated by the last 8 digits (*mmddyyhh*) of each line in the text box.

iso-butane

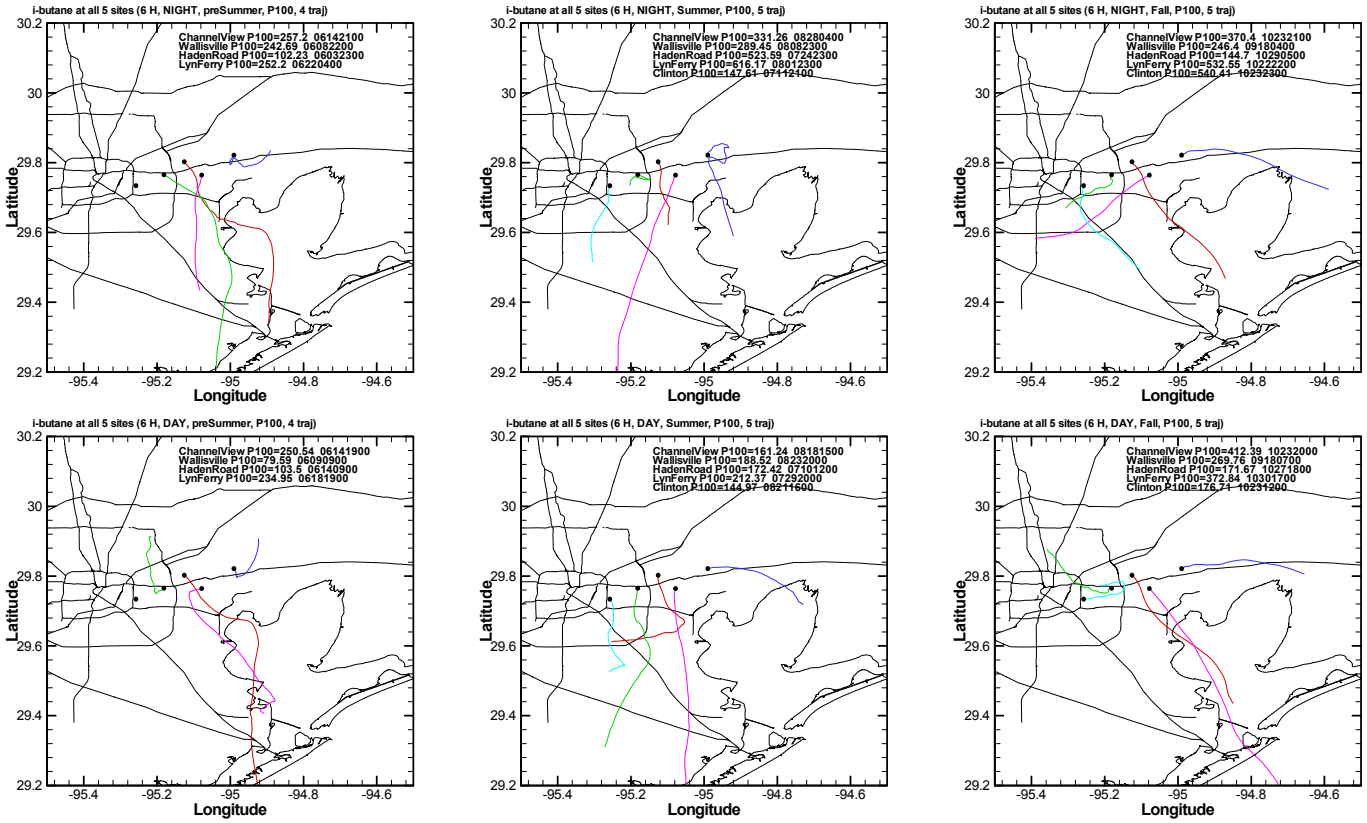
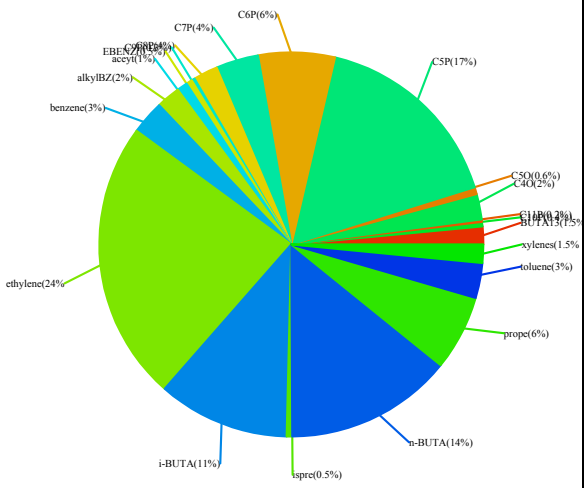
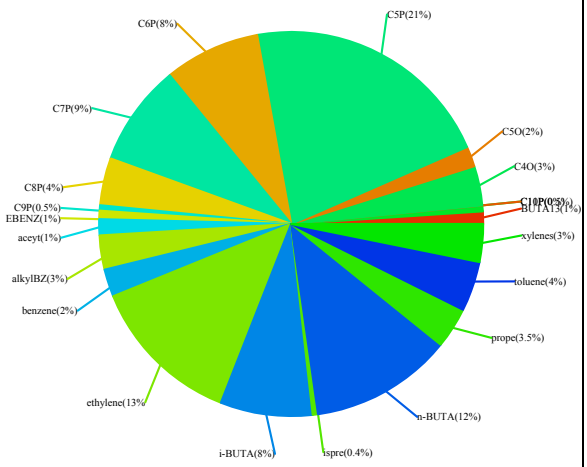
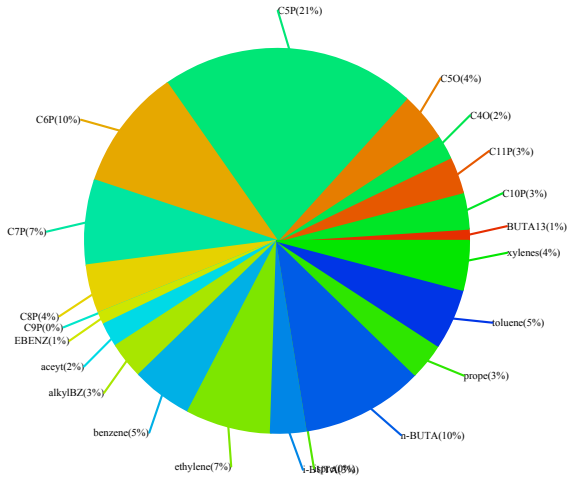
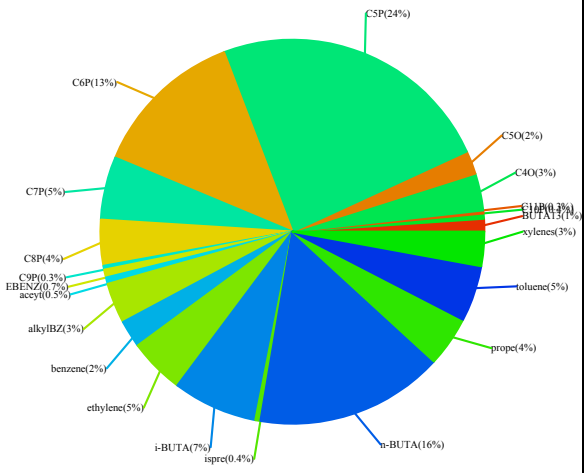
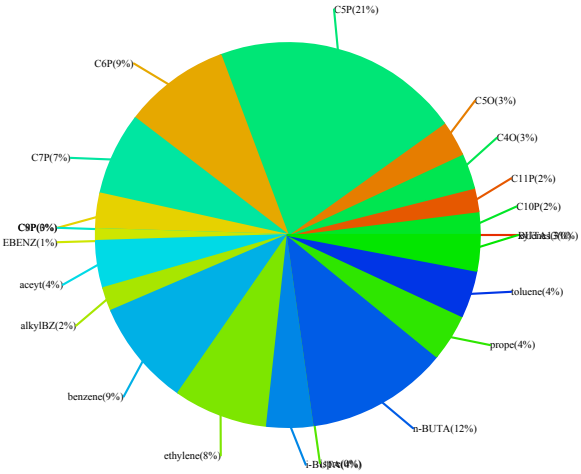
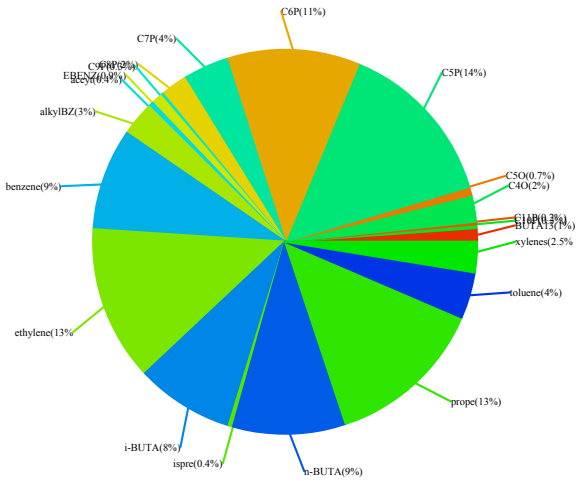
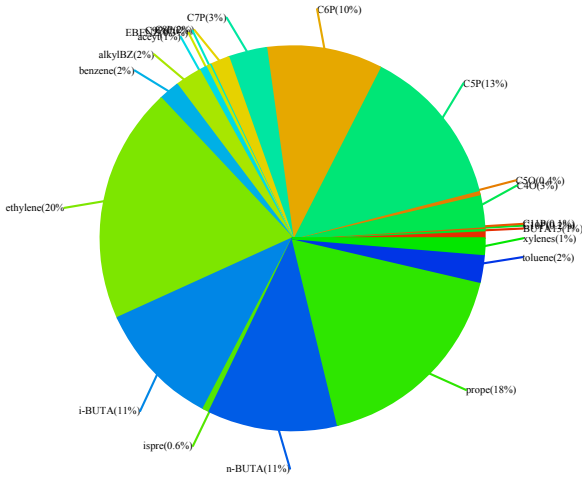
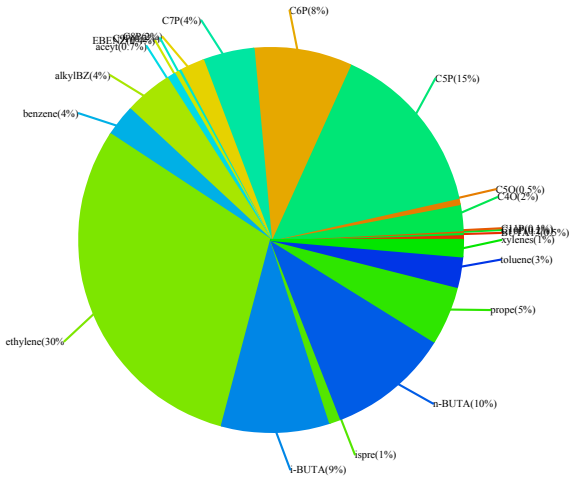


Figure 9. Average Categorized Species Contributions

ChannelView		
Clinton		
HadenRoad		

LynFerry		
WallisVille		
Danciger		

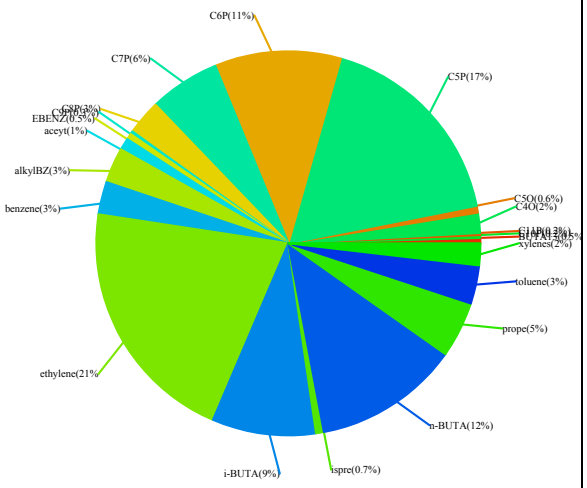
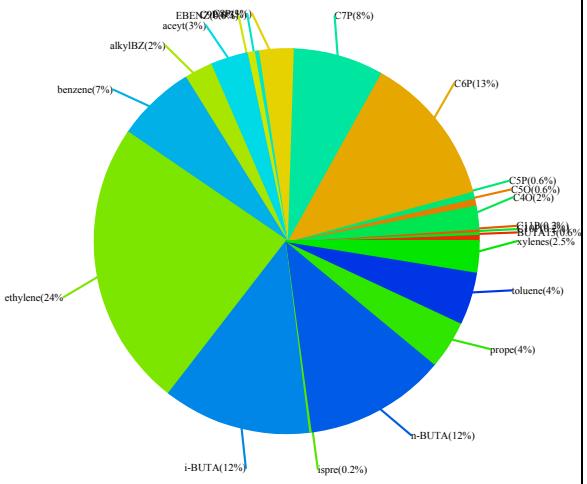
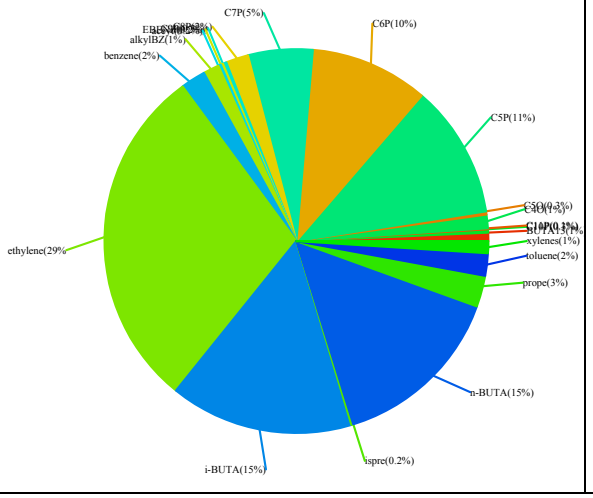
LakeJackson		
TexasCity		
Mustang		

Figure 10. Sample CPF plot.

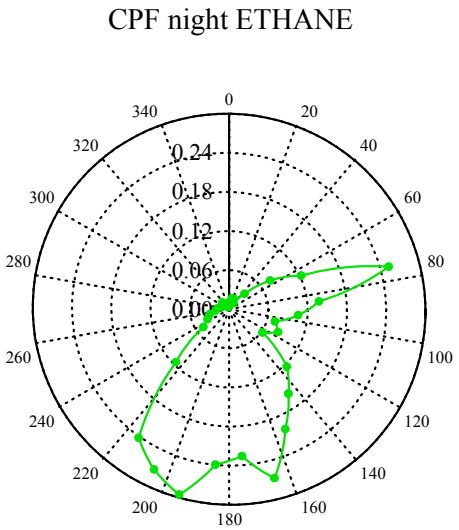


Figure 11. Wind rose associated with observations of ethane. Note the lack of symmetry in directions from which the wind came.

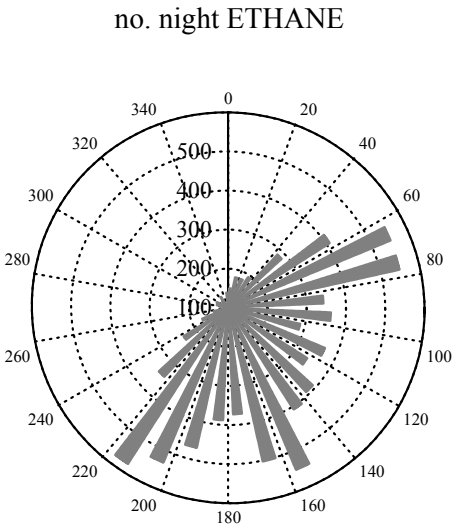


Figure 12. Trajectories corresponding to the seasonal maxima of *iso*-butane. Note: date and time of seasonal maximum are different at different sites as shown in the end of inset text box. The number besides each site is the corresponding sinuosity value of the trajectory.

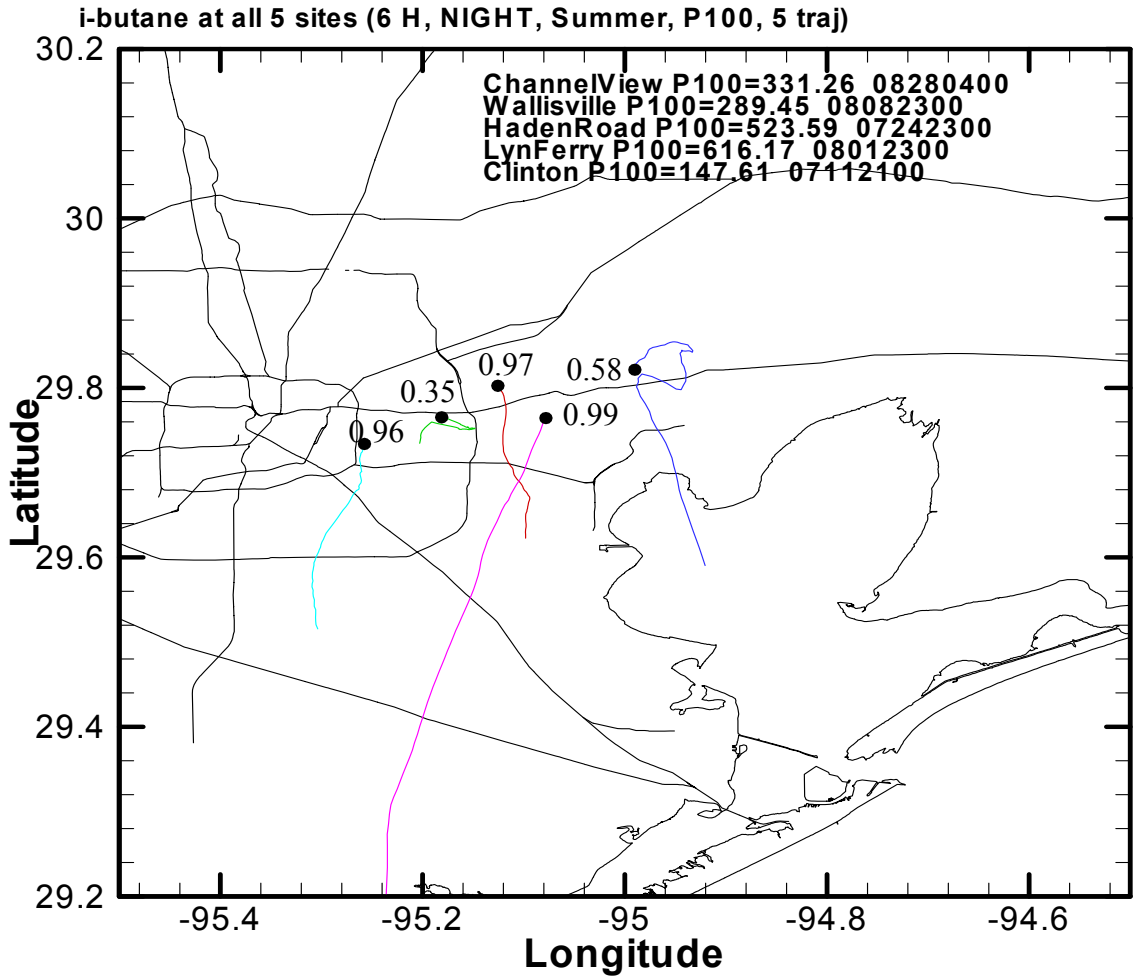


Figure 13. Wind roses of *nighttime* and *daytime* at *summer* and *fall*

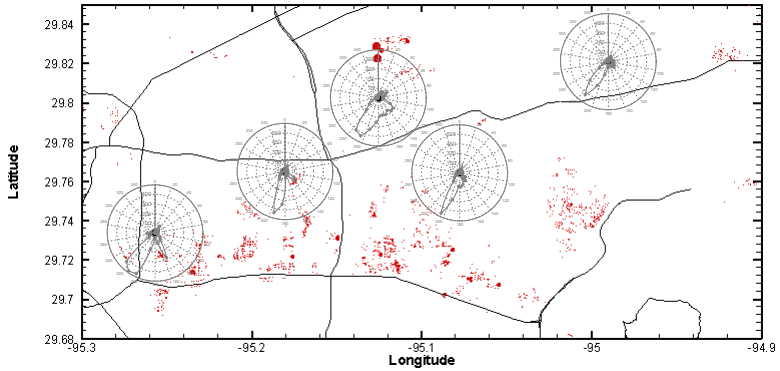


Figure 13A. Wind rose of **Nighttime** at **Summer**

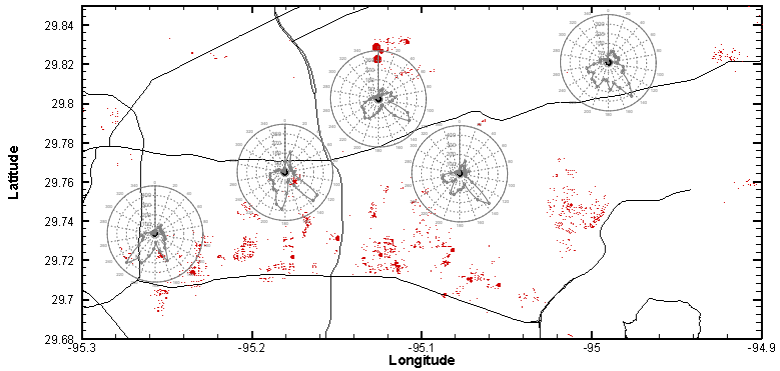


Figure 13B. Wind rose of **Daytime** at **Summer**

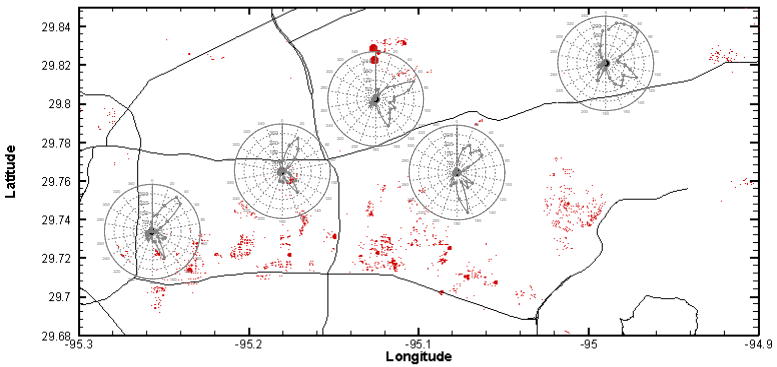


Figure 13C. Wind rose of **Nighttime** at **Fall**

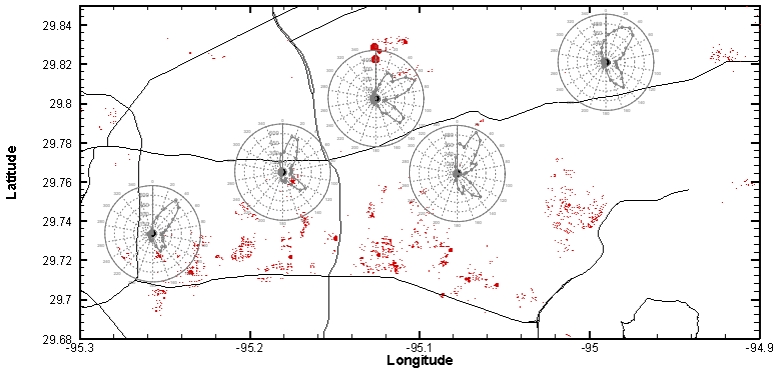


Figure 13D. Wind rose of **Daytime** at **Fall**

Figure 14. CPF plots of *ethene* of *nighttime* and *daytime* at *summer* and *fall*

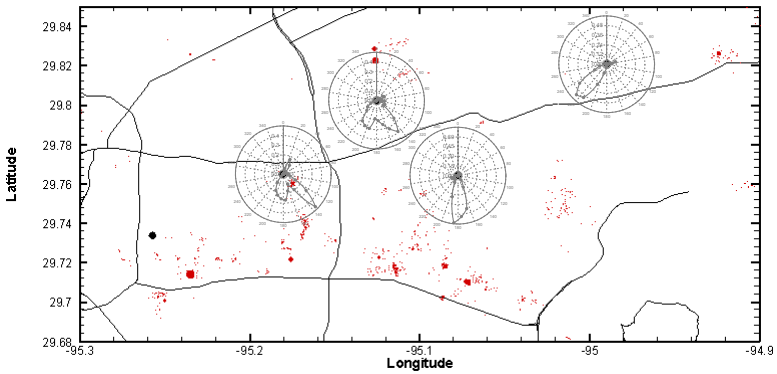


Figure 14A. CPF plot of **Nighttime** *ethene* at **Summer** (Figure 5 in last report)

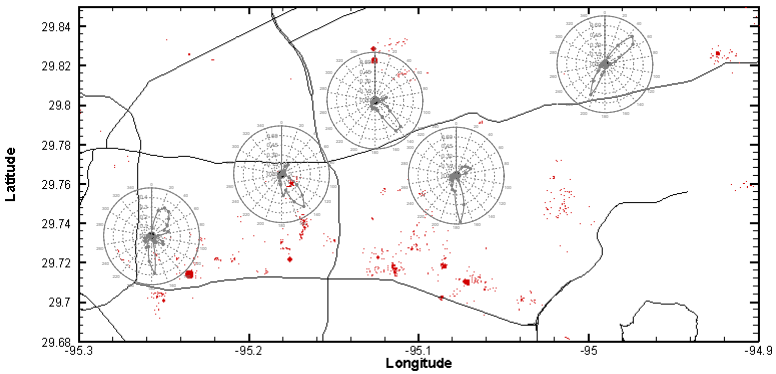


Figure 14C. CPF plot of **Nighttime** *ethene* at Fall (Figure 6 in last report)

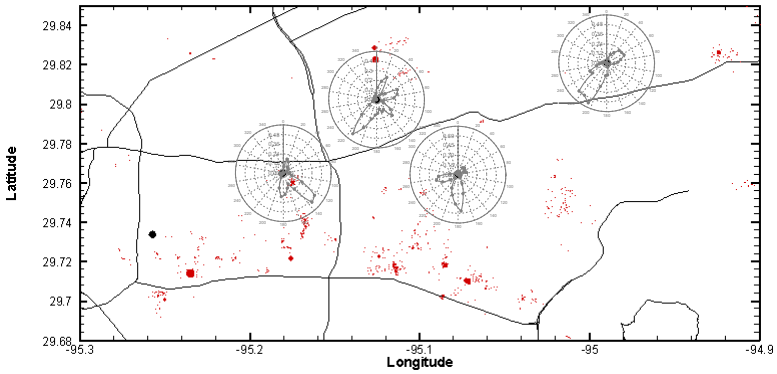


Figure 14B. CPF plot of **Daytime** *ethene* at **Summer**

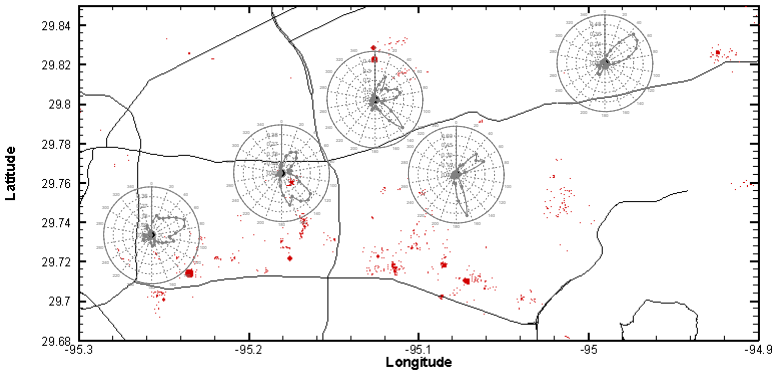


Figure 14D. CPF plot of **Daytime** *ethene* at

Figure 15A. Cluster tree (dendrogram) from the hierarchical clustering analysis on the CPF plots of all VOC species measured at *daytime* (*left*) and *nightime*) during pre-summer (*i.e.,* June) at ChannelView site.

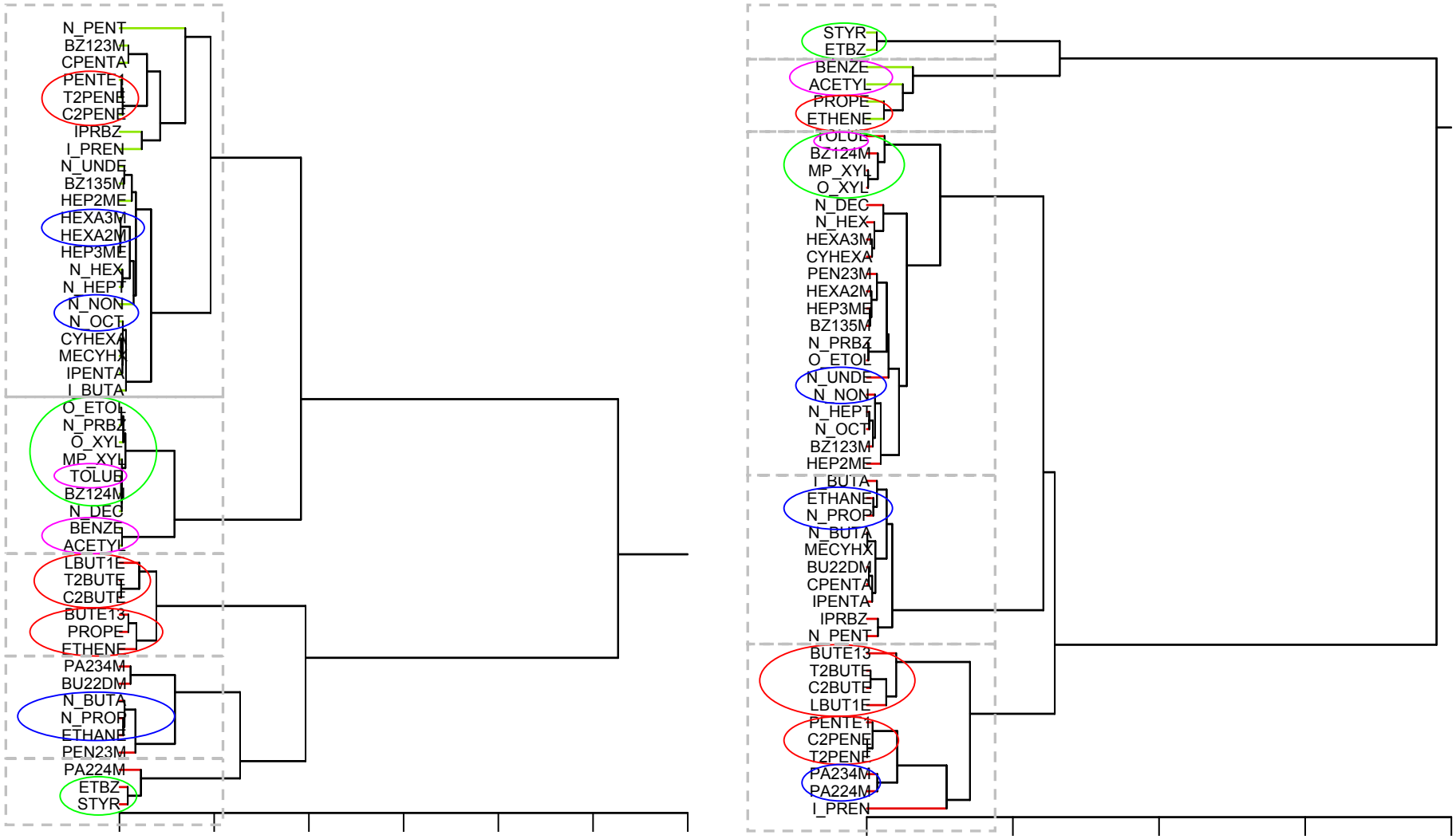


Figure 15B. Cluster tree (dendrogram) from the hierarchical clustering analysis on the CPF plots of all VOC species measured at *daytime* (*left*) and *nightime*) during summer (*i.e.,* July-August) at ChannelView site.

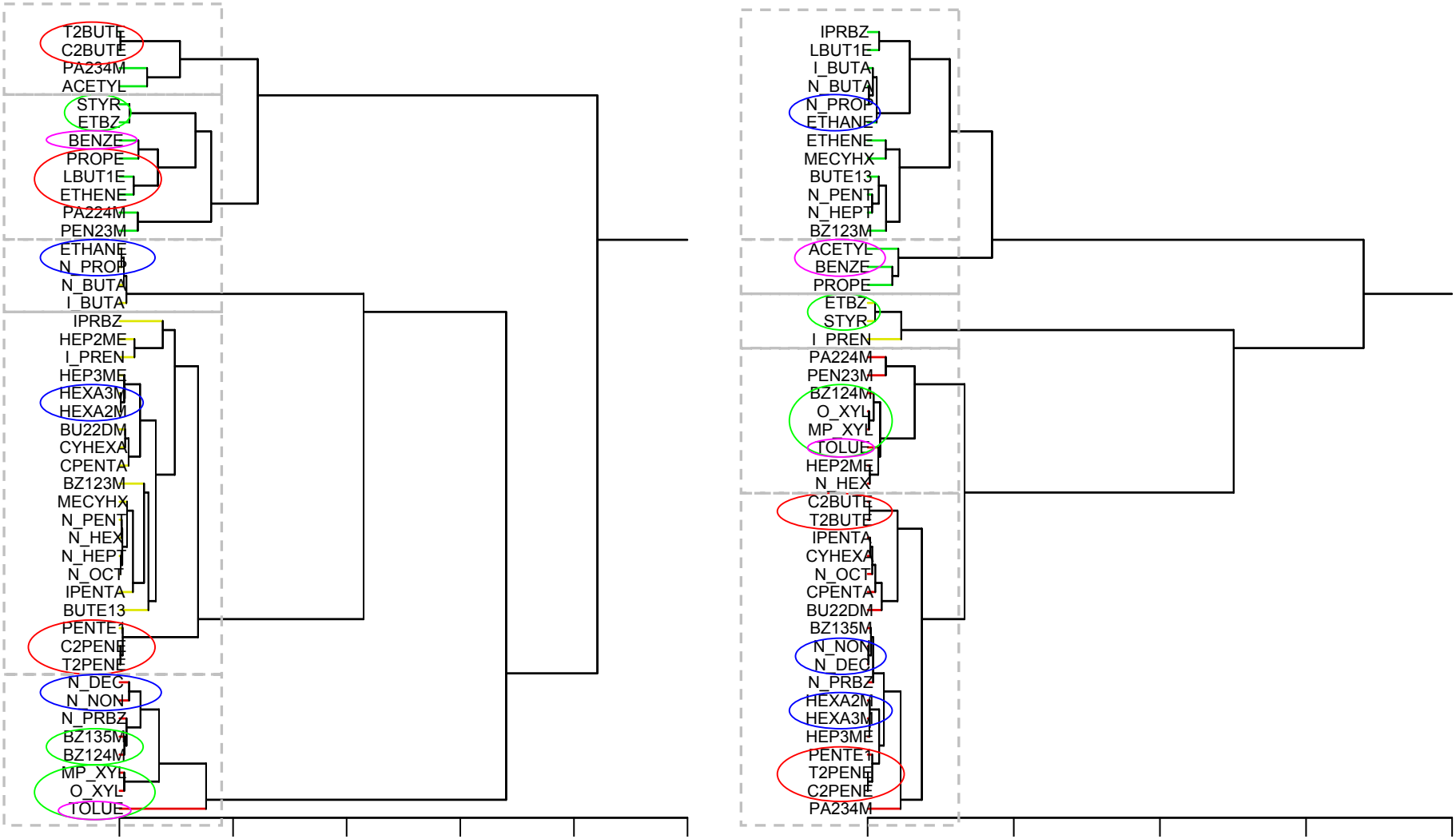


Figure 15C. Cluster tree (dendrogram) from the hierarchical clustering analysis on the CPF plots of all VOC species measured at *daytime* (*left*) and *nightime*) during fall (*i.e.,* September-October) at ChannelView site

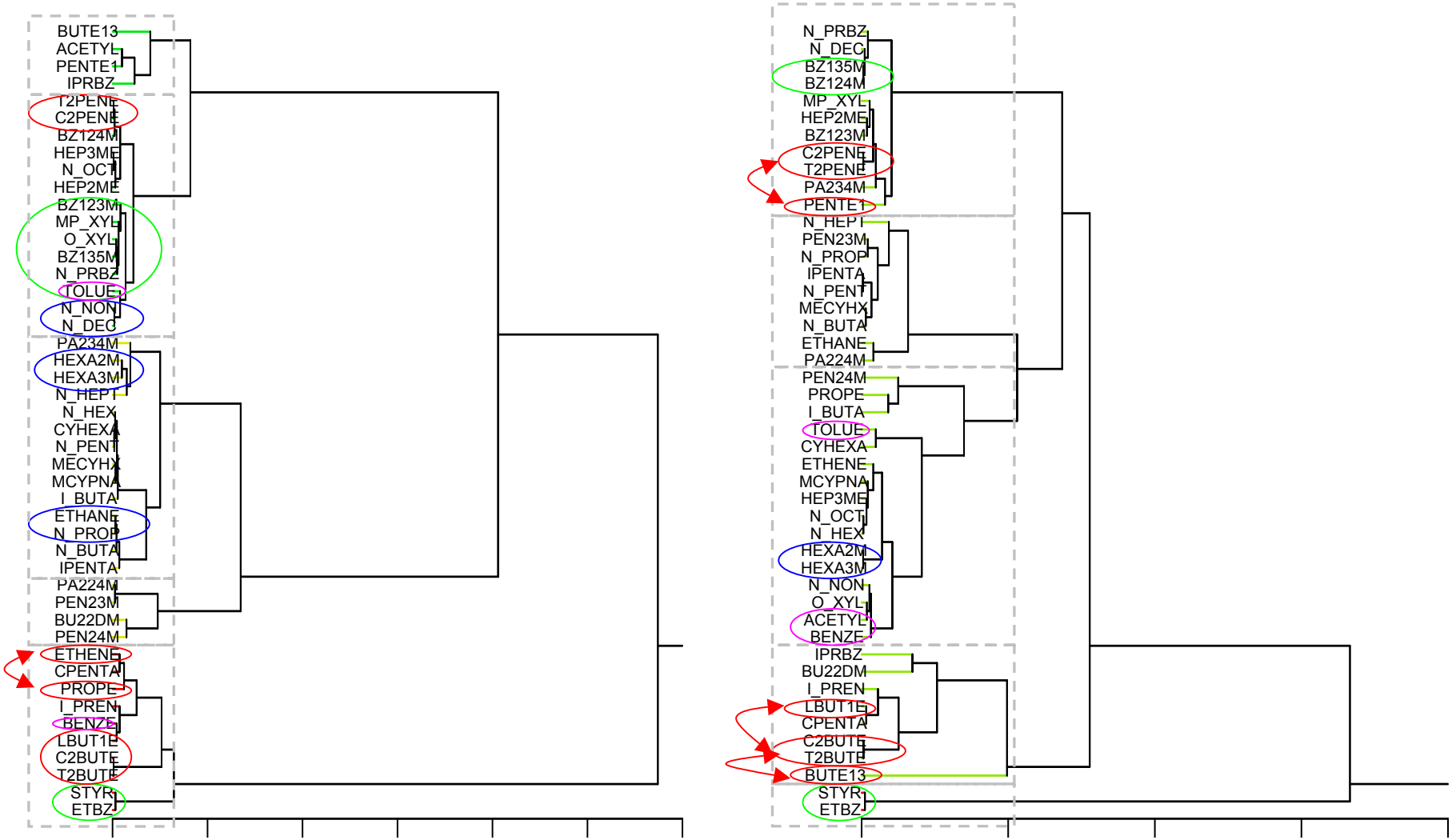


Figure 16. PSCF plots of summer nighttime propylene

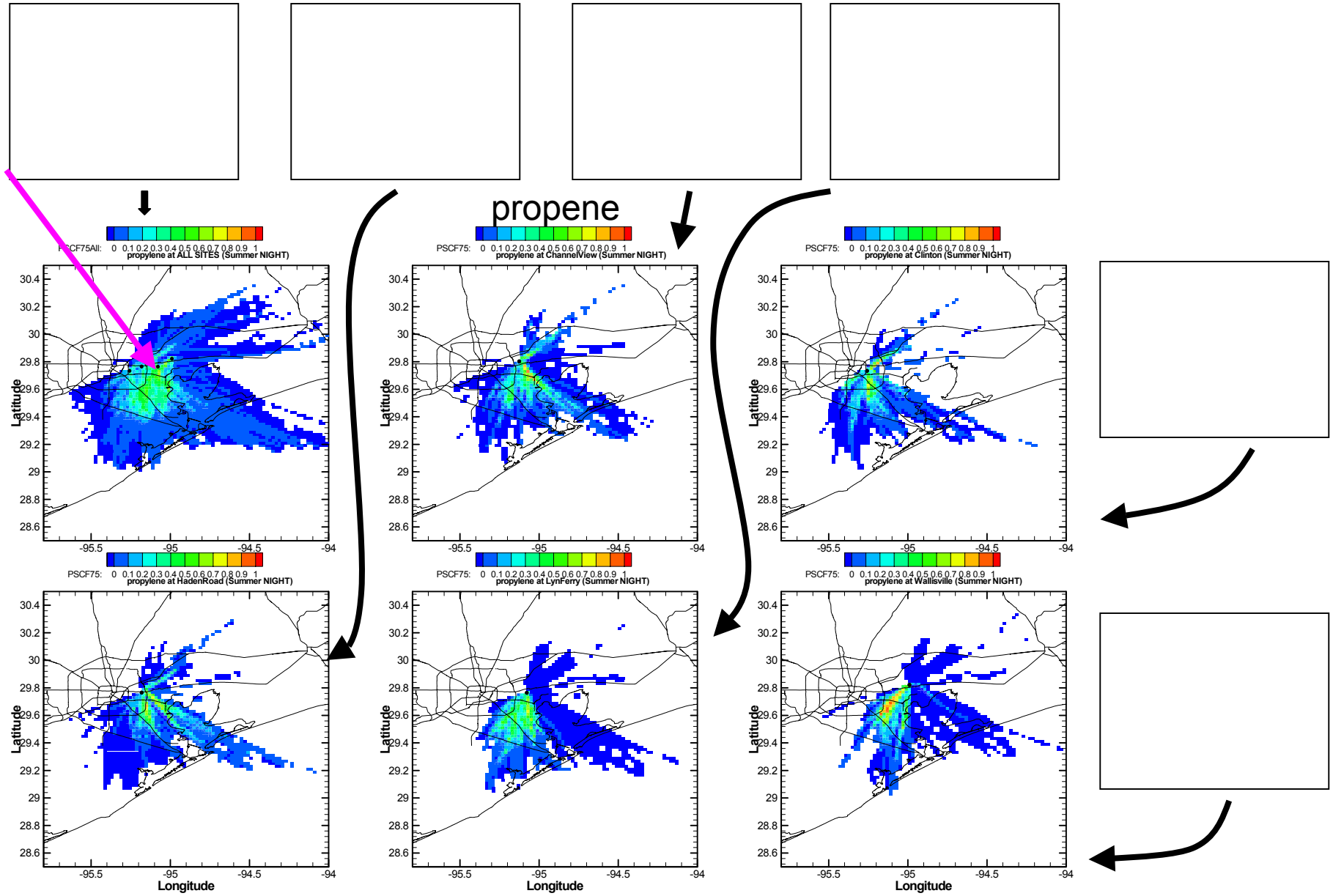


Figure 17. Re-categorized PSCF plots of summer nighttime propylene (**blue**: PSCF values 0-0.25; **green**: PSCF value 0.25-0.50; **red**: PSCF values 0.5 – 1.0)

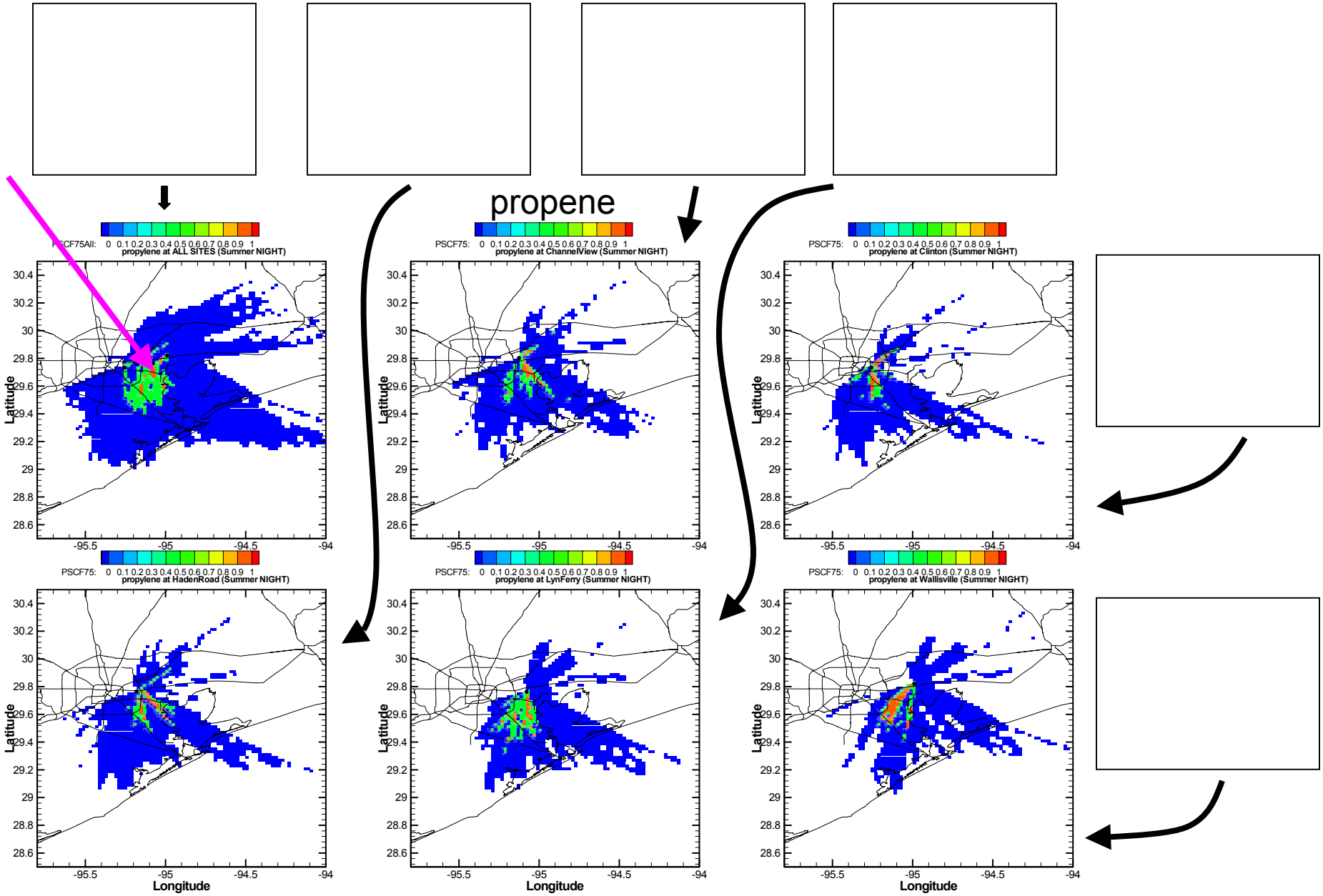


Figure 18. CPF plot of summer nighttime propylene

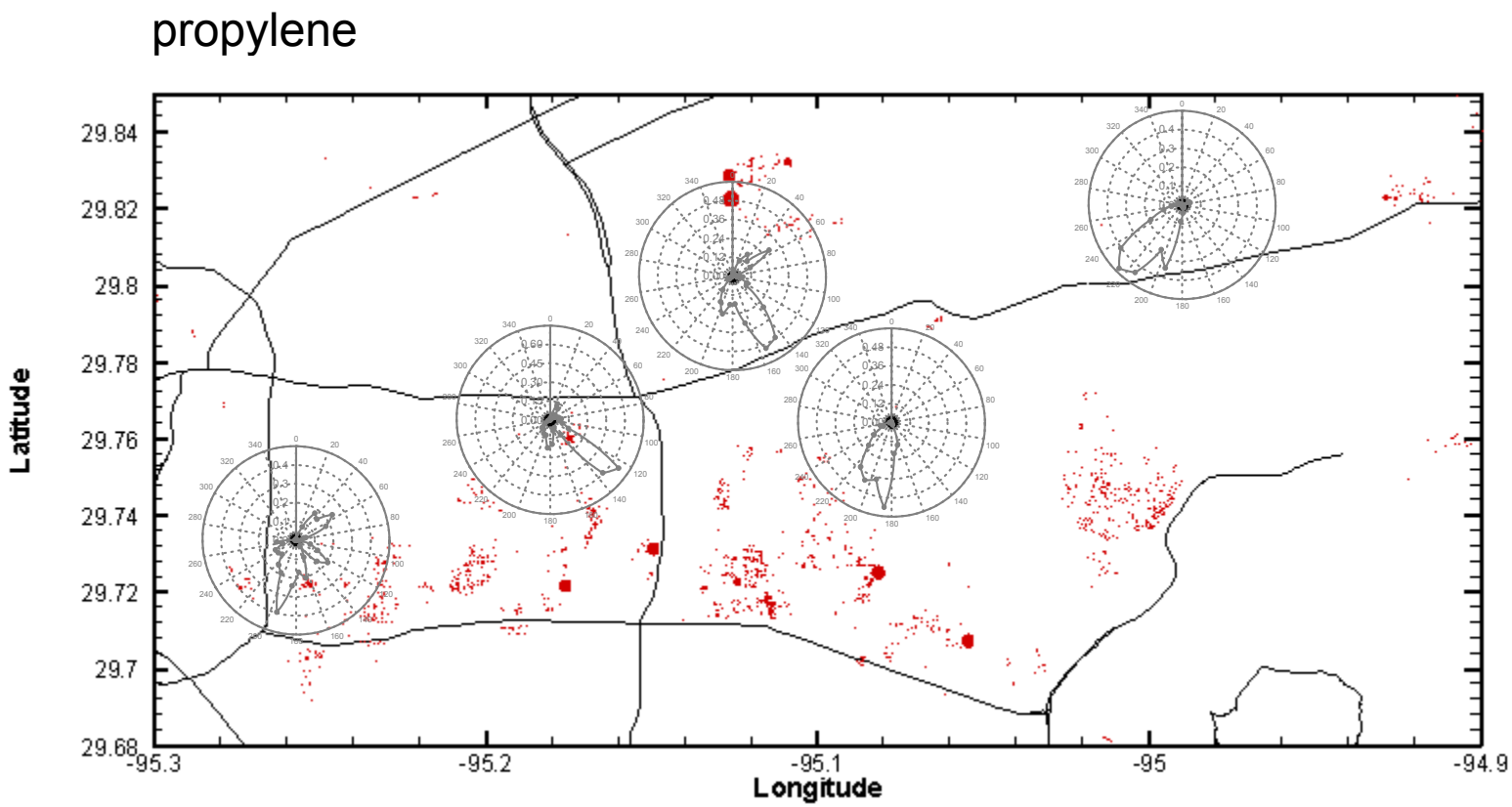


Figure 19*a* (*left*). Source Composition Profiles of the 7 factor PMF solution at ChannelView site (factor 1 to 7 are from bottom to top)
Figure 19*b* (*right*). Percentage of VOC species variation expressed by each of the 7 PMF factor at ChannelView site (factor 1 to 7 are from bottom to top)

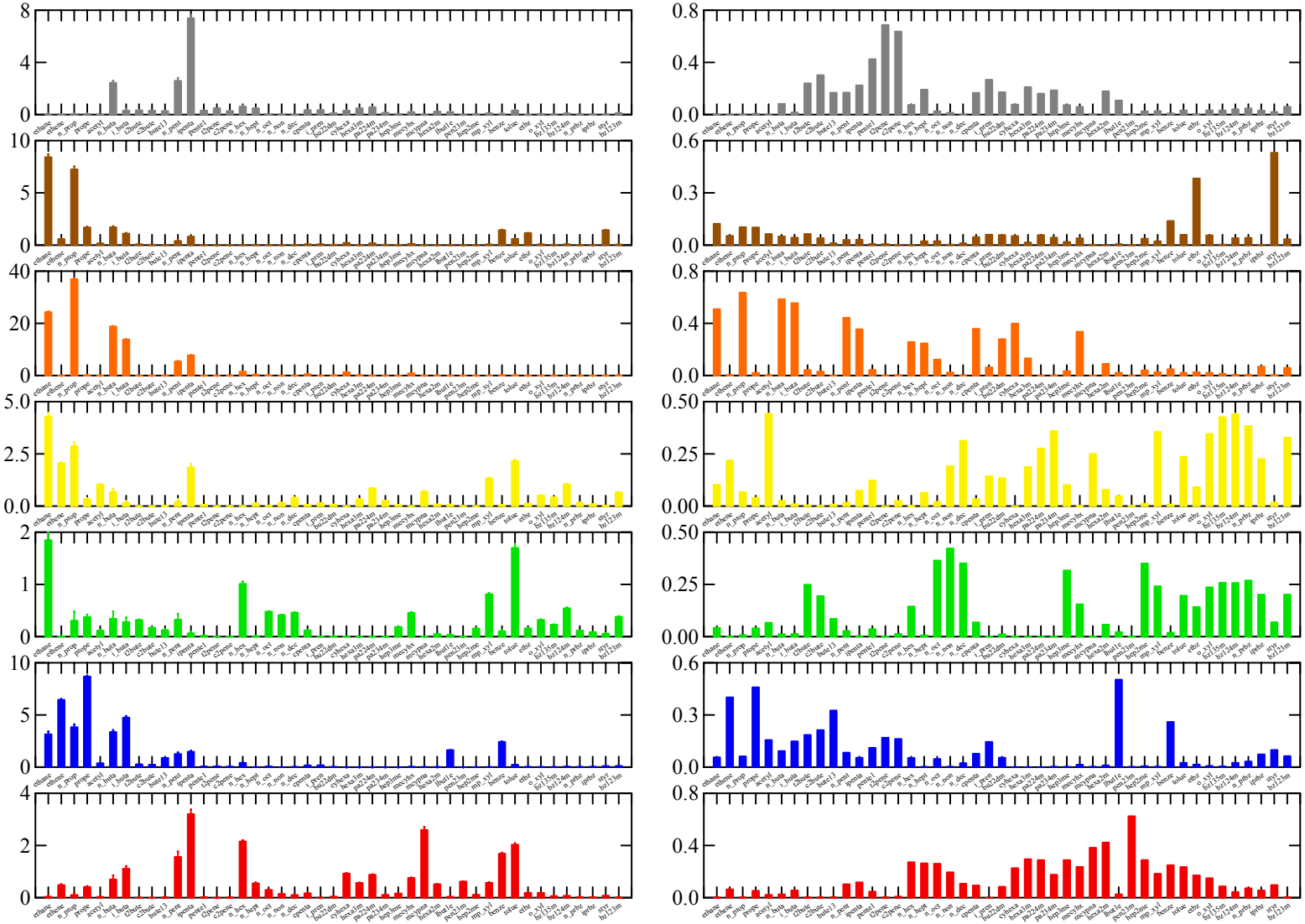


Figure 19*c*. Mean (*left*) and median (*right*) VOC composition of each factor at ChannelView site

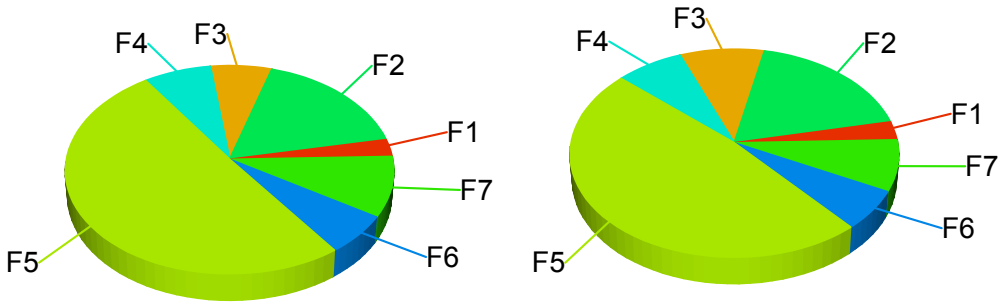


Figure 19*d*. Scatter plot of reproduced total VOC concentration versus the measured total VOC concentration at ChannelView site

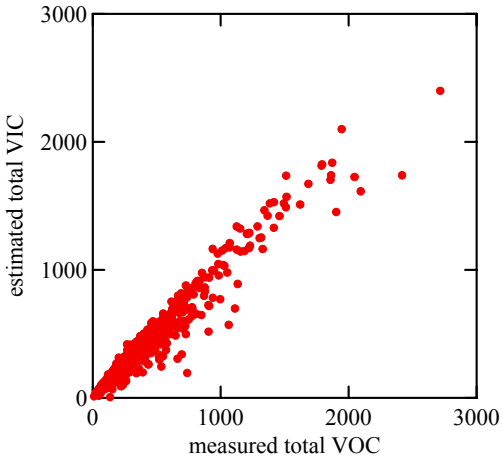


Figure 19*e*. CPF plots of each factor at ChannelView site

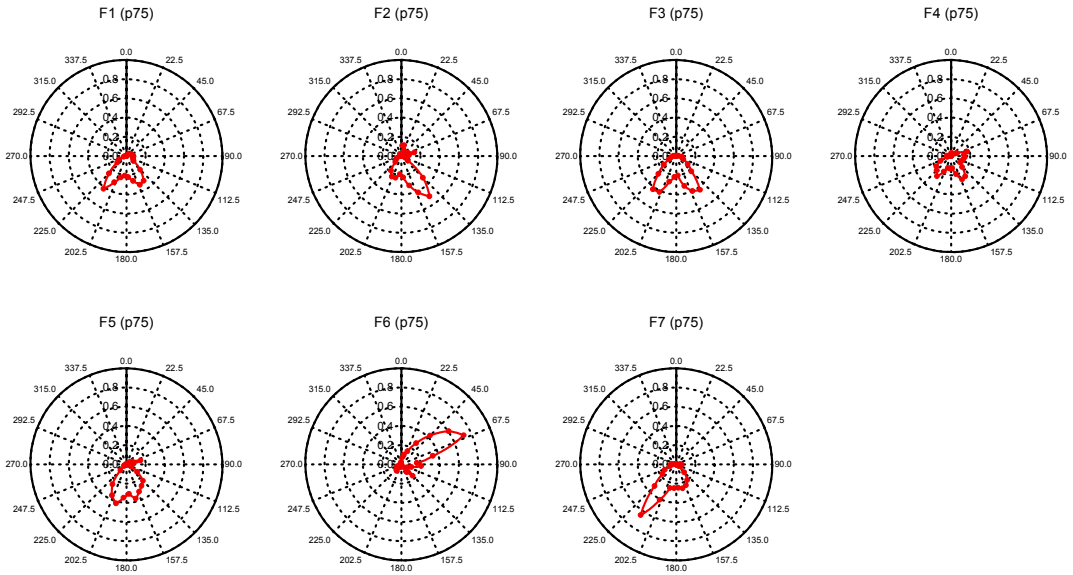


Figure 20*a* (*left*). Source Composition Profiles of the 7 factor PMF solution at Clinton site
Figure 20*b* (*right*). Percentage of VOC species variation expressed by each of the 7 PMF factor at Clinton site

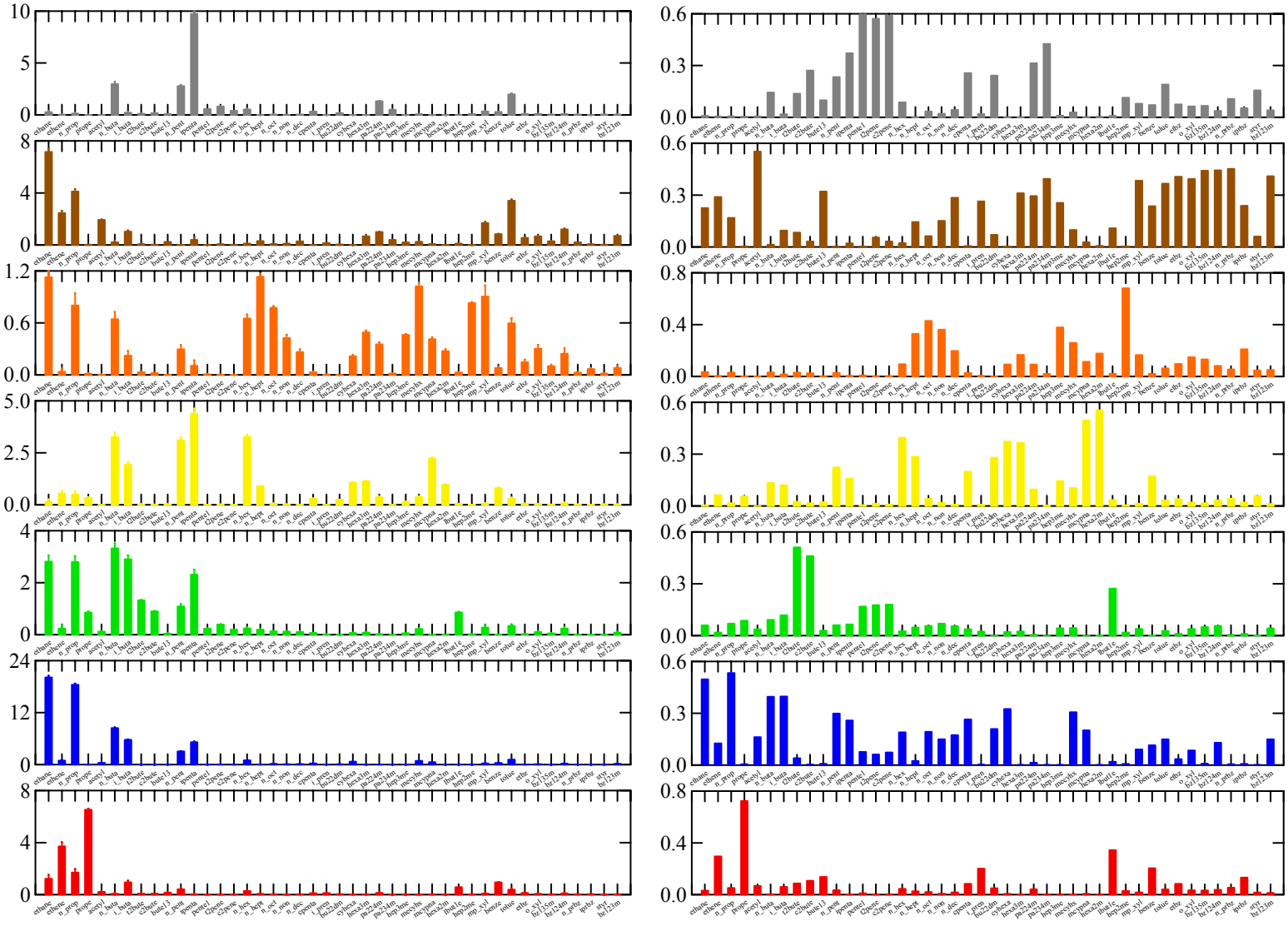


Figure 20*c*. Mean (*left*) and median (*right*) VOC composition of each factor at Clinton site

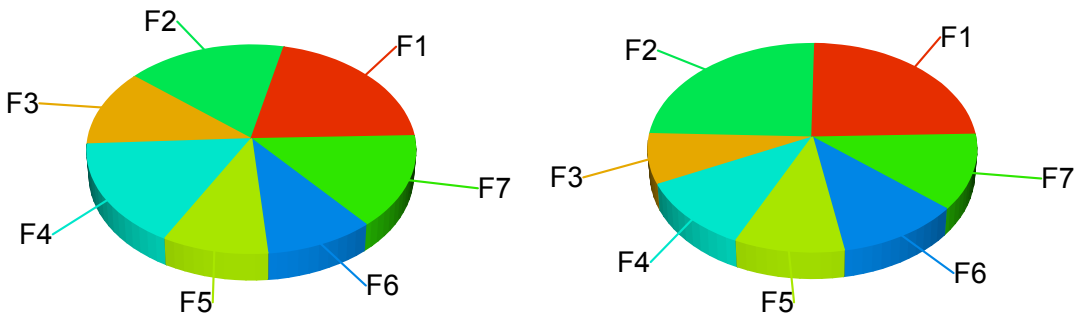


Figure 20*d*. Scatter plot of reproduced total VOC concentration versus the measured total VOC concentration at Clinton site

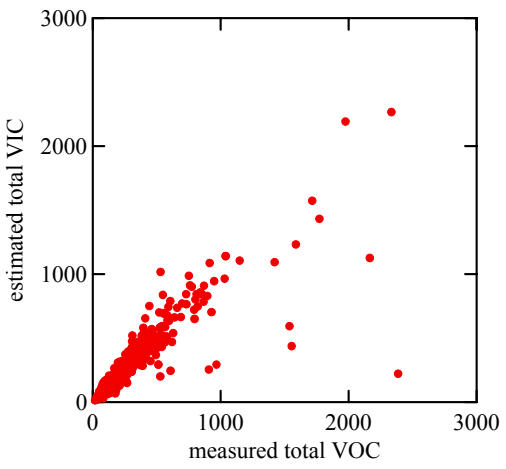


Figure 20*e*. CPF plots of each factor at Clinton site

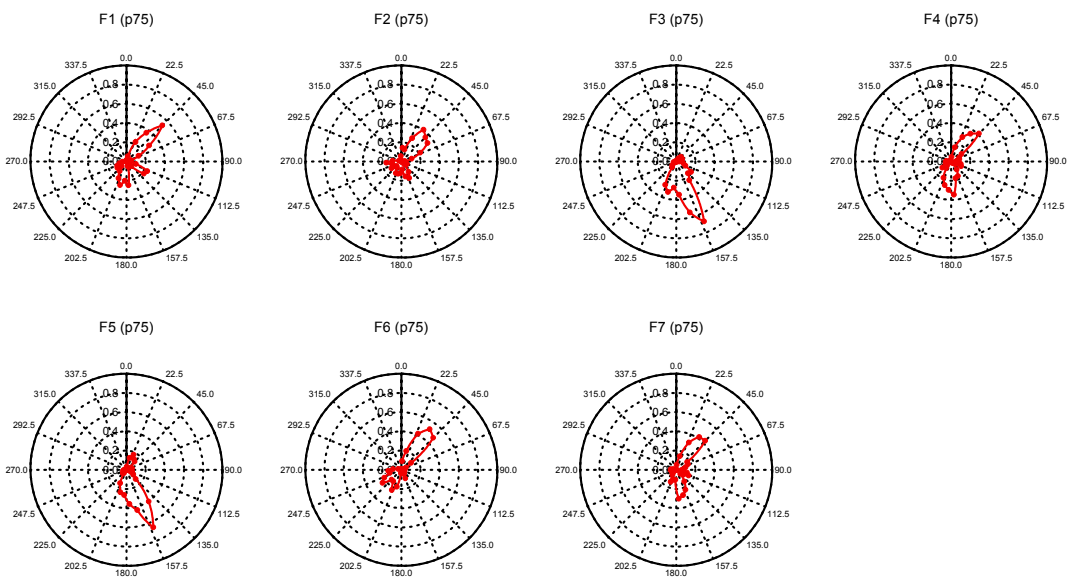


Figure 21*a* (*left*). Source Composition Profiles of the 7 factor PMF solution at HadenRoad site
Figure 21*b* (*right*). Percentage of VOC species variation expressed by each of the 7 PMF factor at HadenRoad site

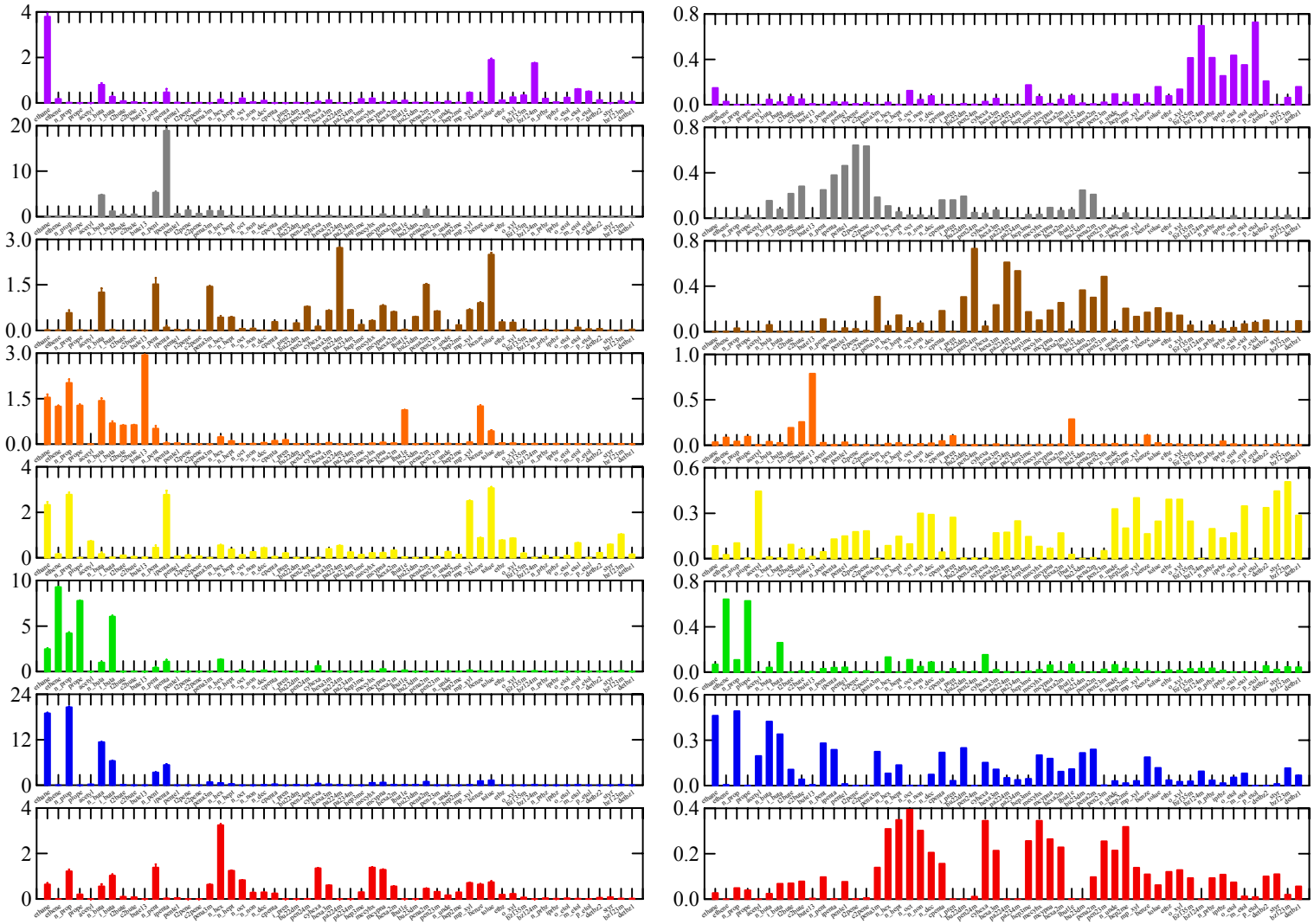


Figure 21*c*. Mean (*left*) and median (*right*) VOC composition of each factor at HadenRoad site

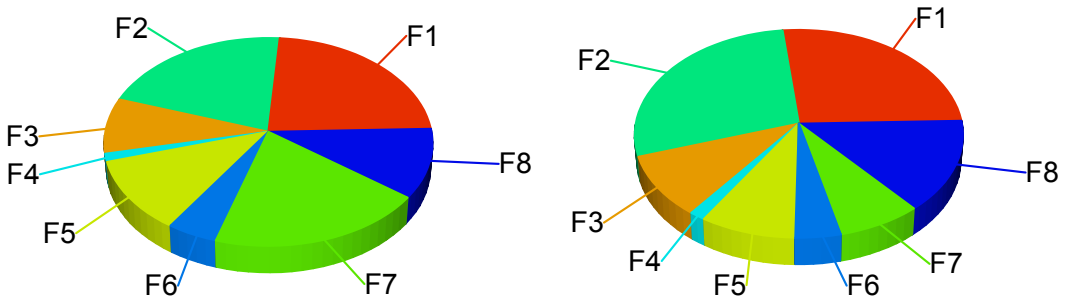


Figure 21*d*. Scatter plot of reproduced total VOC concentration versus the measured total VOC concentration at HadenRoad site

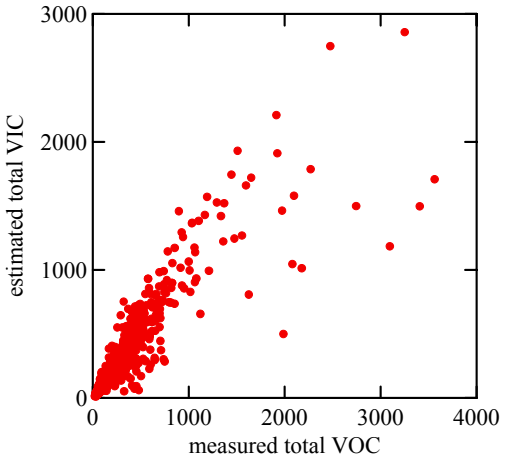


Figure 21*e*. CPF plots of each factor at HadenRoad site

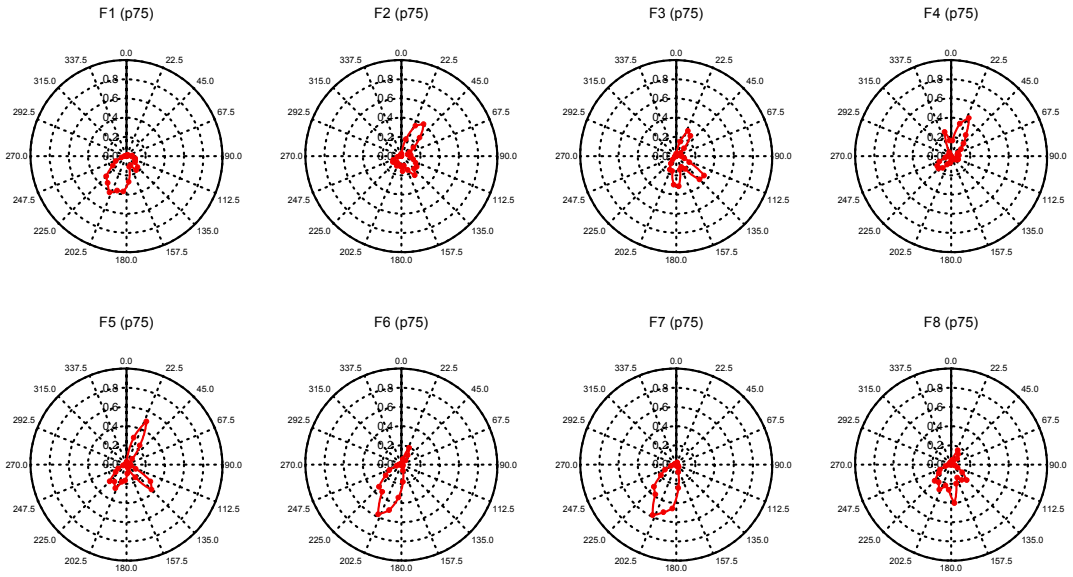


Figure 22*a* (*left*). Source Composition Profiles of the 7 factor PMF solution at Lynchburg Ferry site
Figure 22*b* (*right*). Percentage of VOC species variation expressed by each of the 7 PMF factor at Lynchburg Ferry site

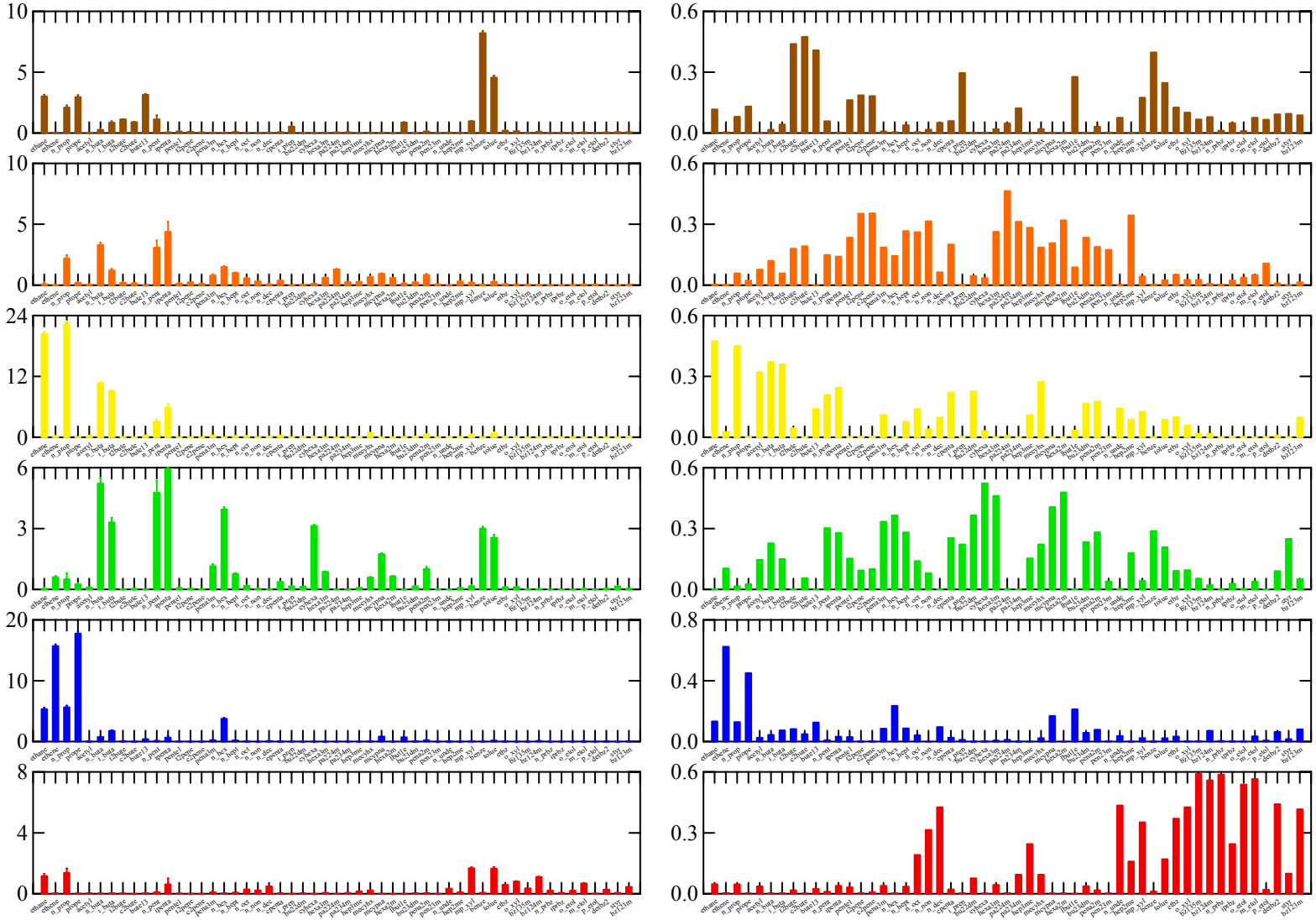


Figure 22*c*. Mean (*left*) and median (*right*) VOC composition of each factor at Lynchburg Ferry site

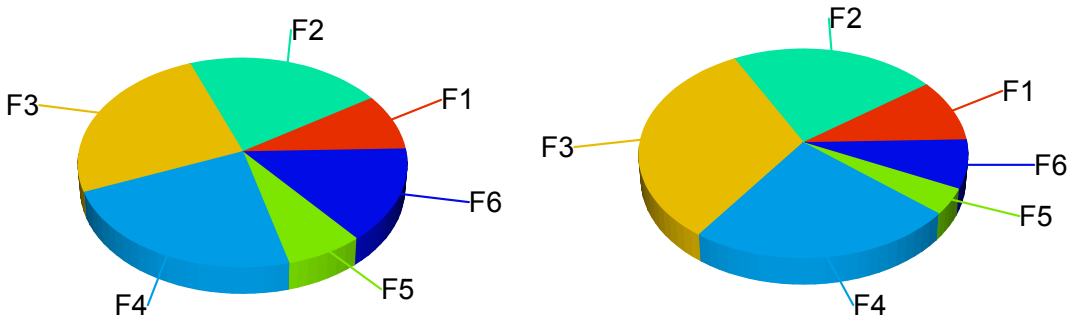


Figure 22*d*. Scatter plot of reproduced total VOC concentration versus the measured total VOC concentration at Lynchburg Ferry site

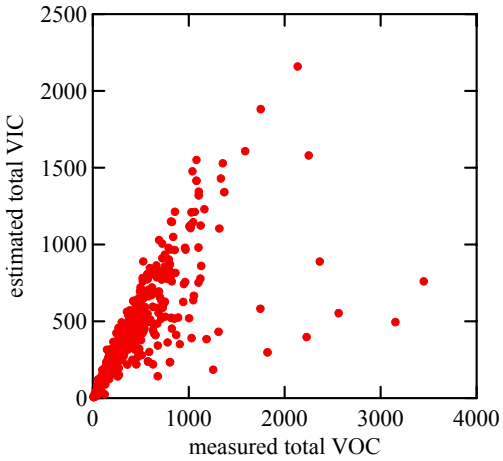


Figure 22*e*. CPF plots of each factor at Lynchburg Ferry site

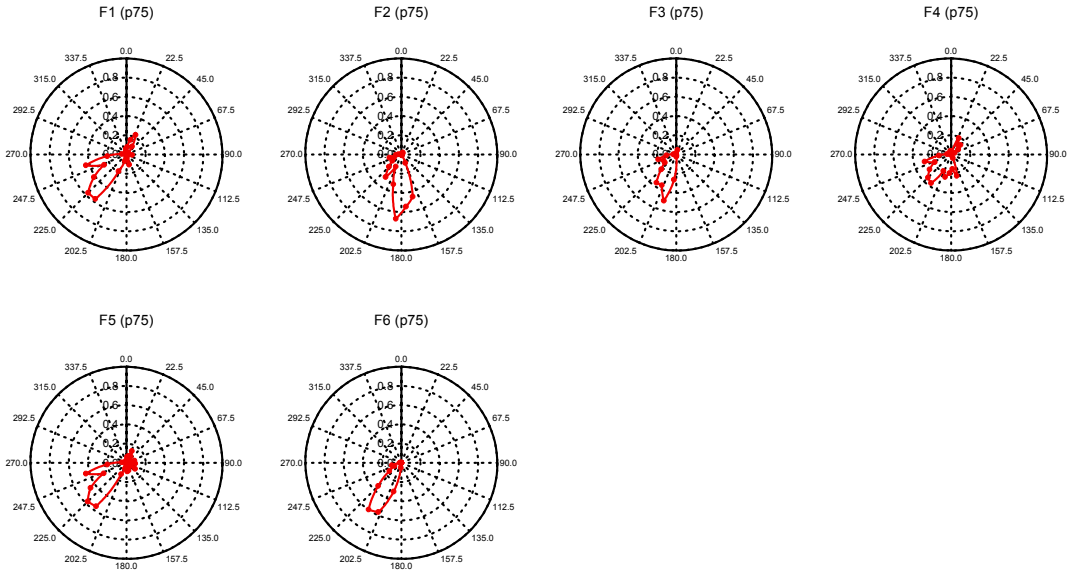


Figure 23*a* (*left*). Source Composition Profiles of the 7 factor PMF solution at WallisVille site
Figure 23*b* (*right*). Percentage of VOC species variation expressed by each of the 7 PMF factor at WallisVille site

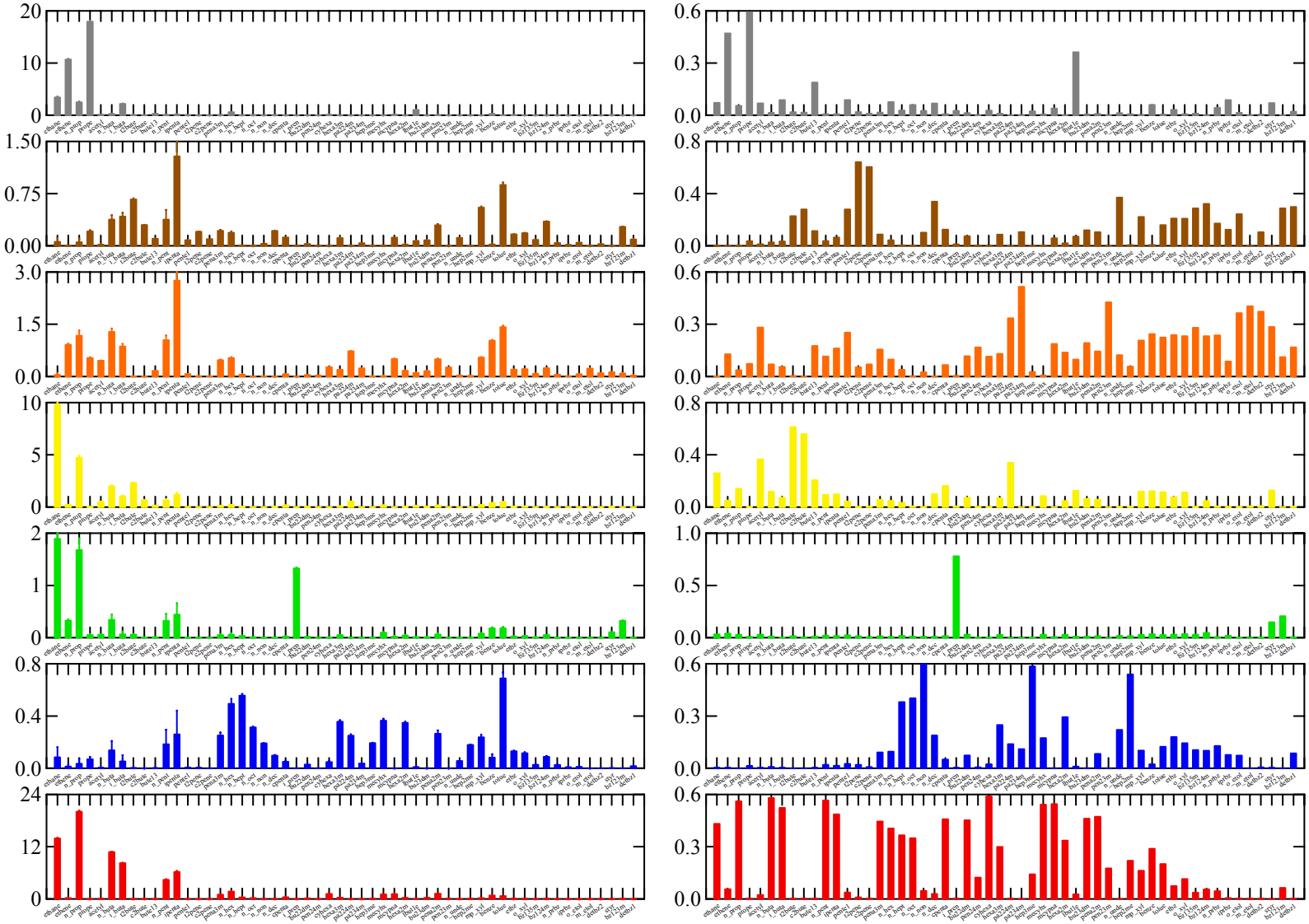


Figure 23*c*. Mean (*left*) and median (*right*) VOC composition of each factor at WallisVille site

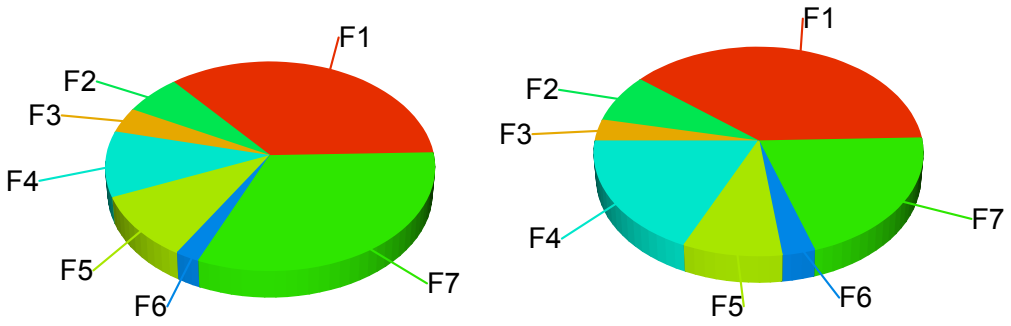


Figure 23*d*. Scatter plot of reproduced total VOC concentration versus the measured total VOC concentration at Wallisville site

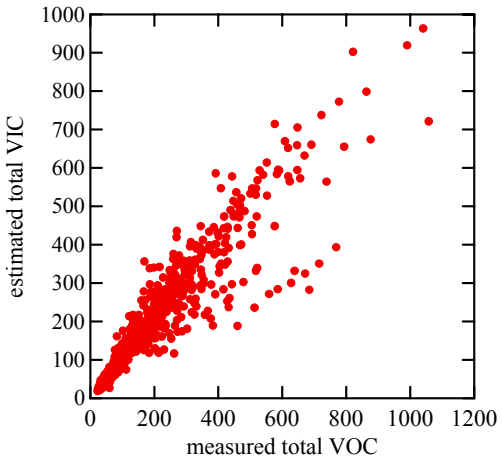


Figure 23*e*. CPF plots of each factor at WallisVille site

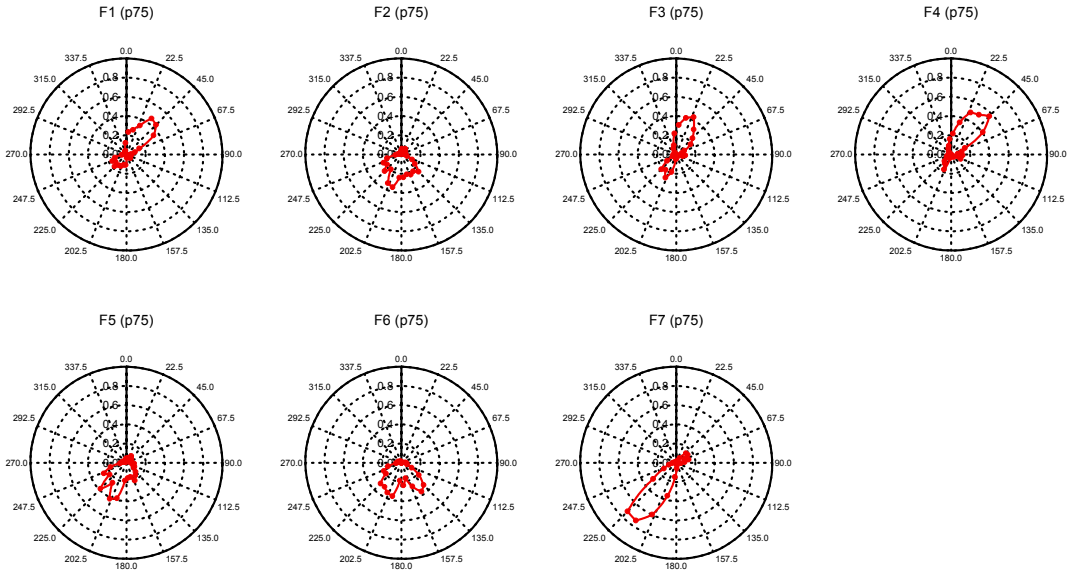


Figure 24*a*. CPF plots of petrochemical sources

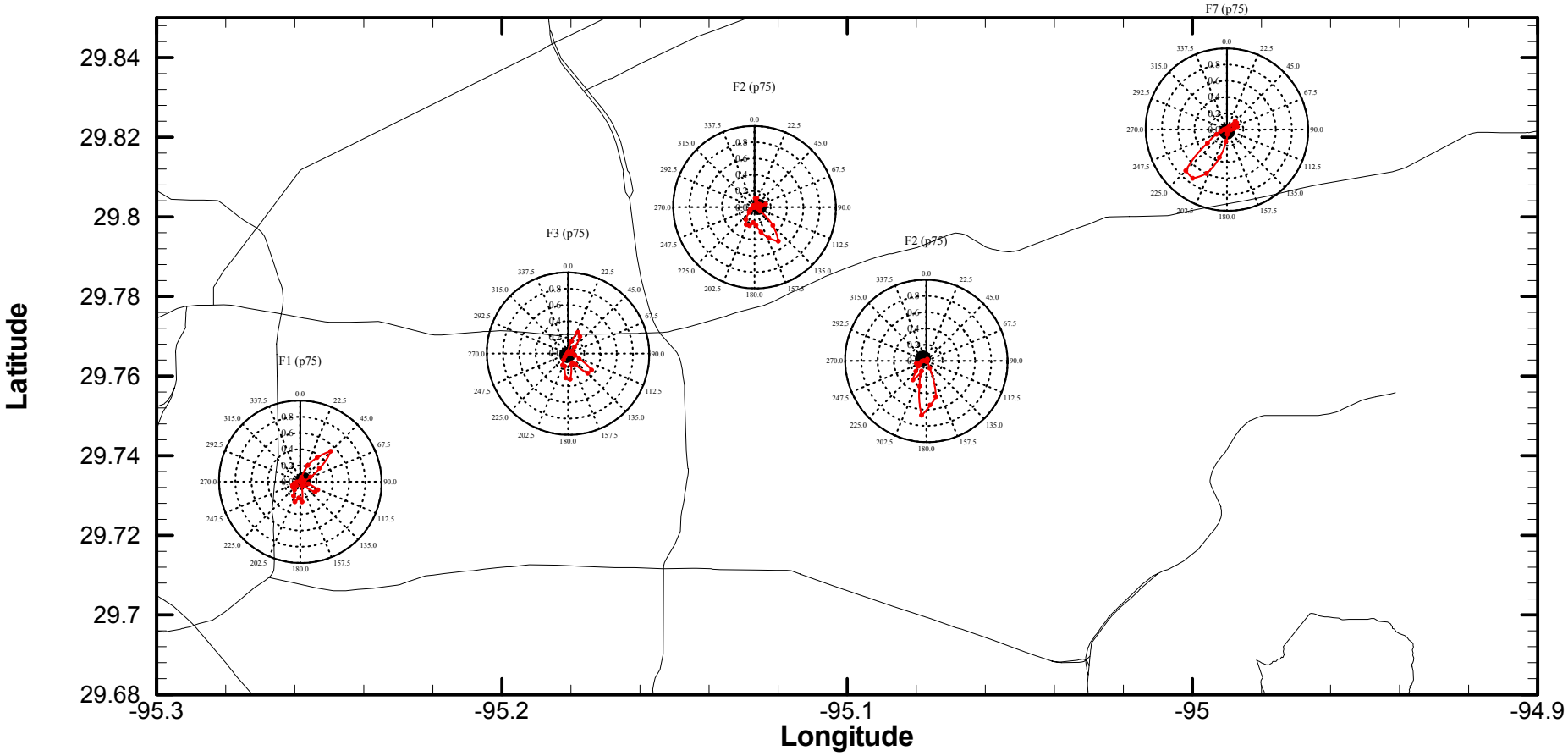


Figure 24*b*. CPF plots of natural gas, evaporative sources / accumulation

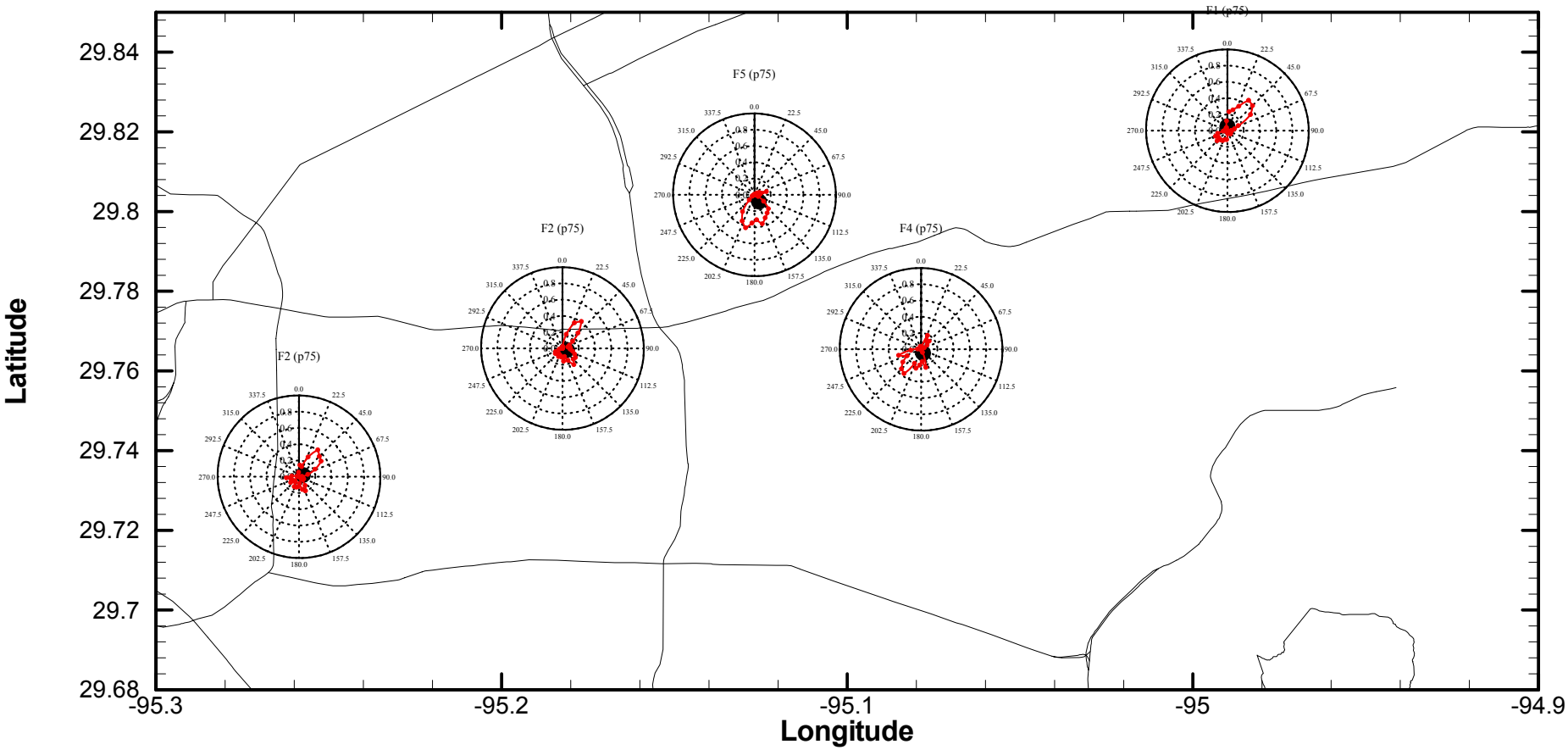


Figure 24*c*. CPF plots of solvent industrial sources

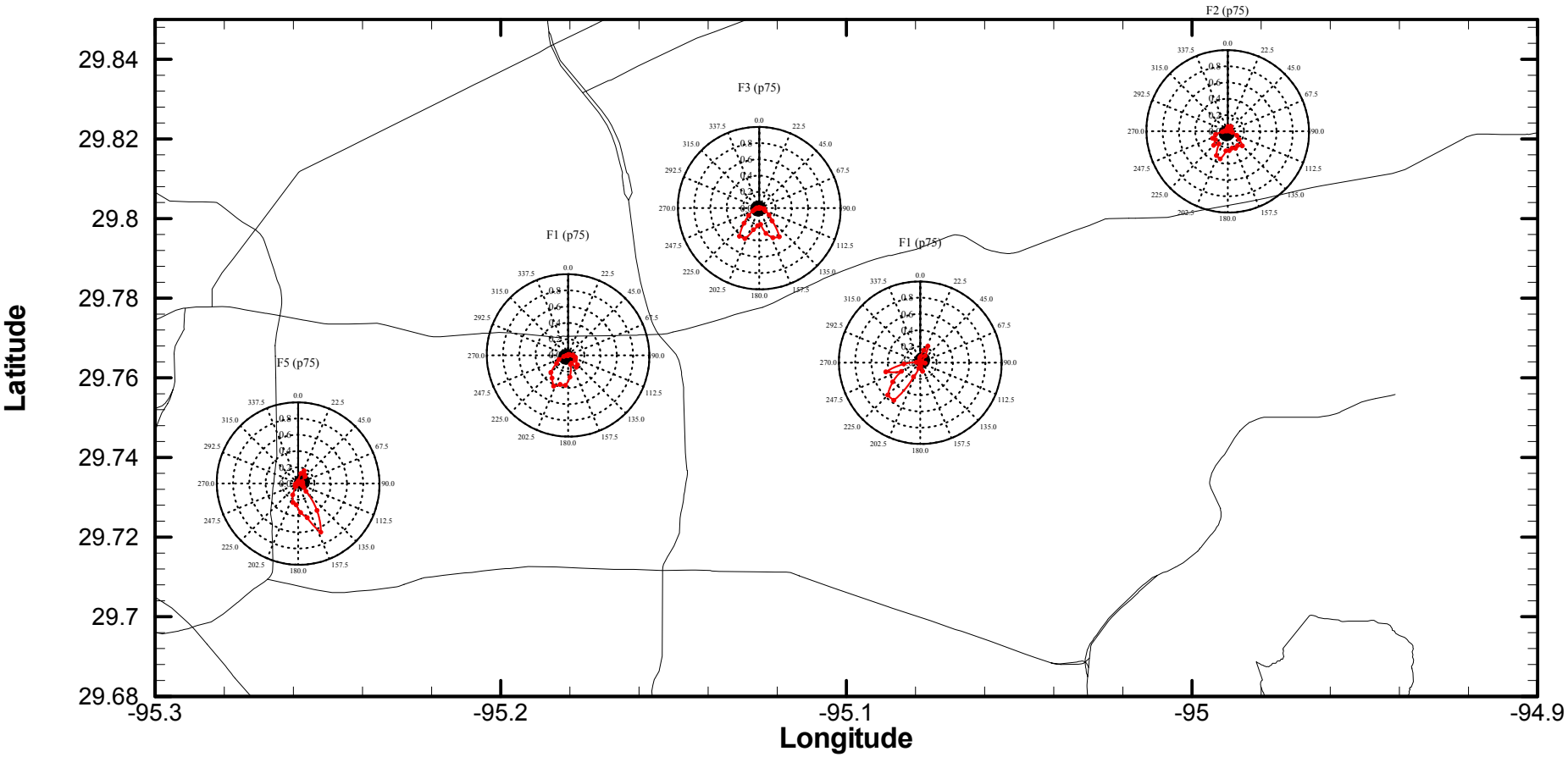


Figure 24*d.* CPF plot of vehicle exhaust emission

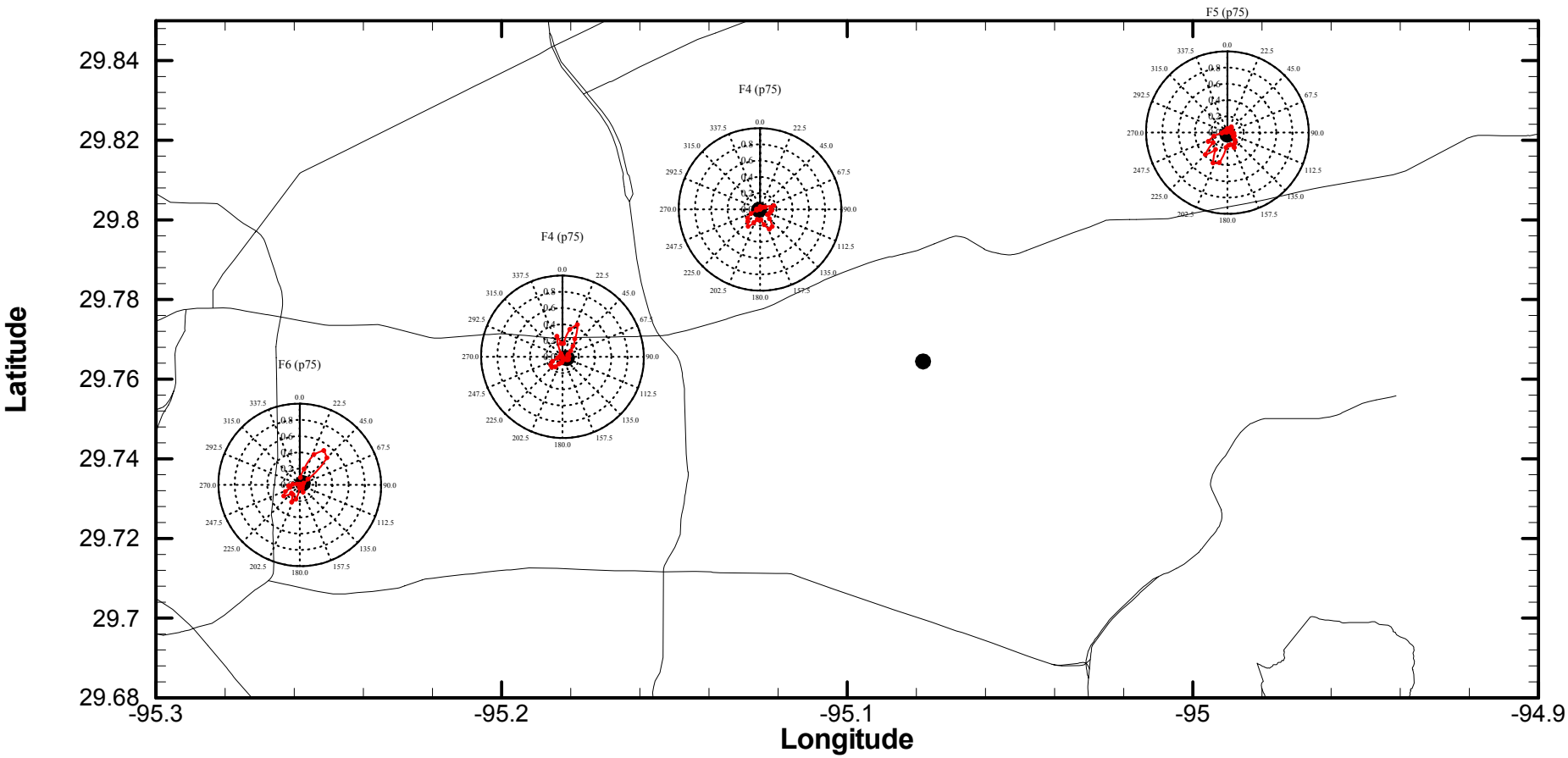
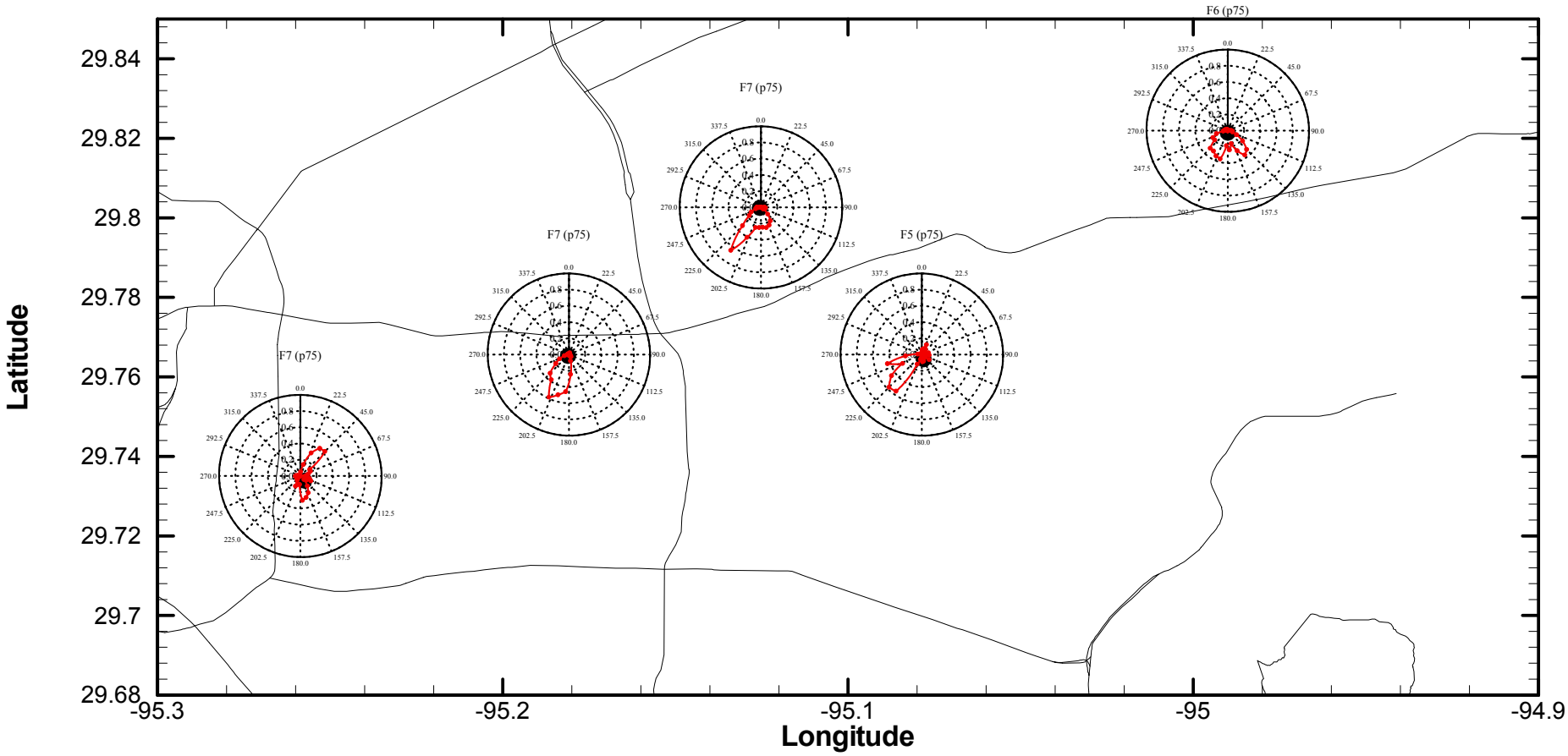



Figure 24*e.* CPF plot of fuel evaporation