

**EMISSION INVENTORY EVALUATION
AND RECONCILIATION IN THE
HOUSTON-GALVESTON AREA**

**FINAL REPORT
STI-903640-2490-FR2**

By:

**Paul T. Roberts
Steven G. Brown
Stephen B. Reid
Martin P. Buhr
Tami H. Funk
Patricia S. Stiefer**

**Sonoma Technology, Inc.
1360 Redwood Way, Suite C
Petaluma, CA 94954-1169**

**Philip K. Hopke
Eugene Kim
Department of Chemical Engineering
Clarkson University
8 Clarkson Avenue
Potsdam, NY 13699-5208**

**Prepared for:
Houston Advanced Research Center
Texas Environmental Research Consortium
4800 Research Forest Drive
The Woodlands, TX 77381**

June 30, 2004

This page is intentionally blank.

ACKNOWLEDGMENTS

This research was sponsored by the Houston Advanced Research Center (HARC). We thank the current HARC contract manager, Dr. Jay Olaguer, and Ms. Anne Brun, who was the contract manager for the early part of the project. We acknowledge the hard work and assistance given by numerous persons responsible for both surface and aloft air quality data collection, validation, and processing and for emissions inventory compilation. We also thank the following Sonoma Technology, Inc. staff for their contributions: Hilary Hafner for guidance and insight into the technical approach and interpretation of results; Erin Shields, Liz Simon, Mark Lilly, and Michael McCarthy for their invaluable role in processing and displaying air quality and emissions data and analysis results; and Lisa DiStefano and Sandy Smethurst for editing and producing this report.

EXECUTIVE SUMMARY

This study was sponsored by the Houston Advanced Research Center (HARC) in support of the Texas Commission on Environmental Quality (TCEQ) to use innovative methods for evaluating and reconciling the emission inventory for the Houston-Galveston area (HGA). The objective of this project was to evaluate the emissions inventory for large point sources in the HGA for 2000 by comparing emissions estimates to ambient data collected during 2000 and 2001, and to make recommendations on possible improvements to the point source inventory. This project focused on sources of non-methane organic compounds (NMOC, often called volatile organic compounds, VOC) that are most likely to contribute to ozone formation, and on oxides of nitrogen (NO_x) in the HGA. In particular, we focused on the 12 highly-reactive VOCs or groups of compounds that TCEQ has targeted because of their significant contributions to ozone formation. These compounds and groups include the following: acetaldehyde; formaldehyde; ethylene; propylene; 1,3-butadiene; all butenes (butylenes); isoprene; all pentenes; toluene; all xylenes; all ethyltoluenes; and all trimethylbenzenes. However, due to limitations in the measurement systems, we did not address acetaldehyde and formaldehyde, and ethyltoluenes and trimethylbenzenes were often lumped together.

Emission inventories are used to support air quality modeling and develop regulatory control strategies. Because of the complexities involved in developing emission inventories and the implications of errors in the inventory on air quality model performance and control strategy assessment, it is important to evaluate the accuracy and representativeness of any inventory that is intended for use in air quality modeling.

There are several techniques used to evaluate emissions data, including bottom-up evaluations that begin with emissions activity data and estimate the emissions accordingly; and top-down evaluations that compare emission estimates to ambient air quality data or use ambient data to estimate emissions profiles. This project focused on top-down evaluation methods using existing surface and aloft (via aircraft) ambient data. The result is an evaluation independent of the inventory itself, as well as an evaluation independent of the modeling that will later use the inventory to predict ozone concentrations and sensitivities.

The technical approach we used involved a multi-task effort which utilized the extensive collection of surface and aircraft data available for 2000 and 2001 in the HGA. We used three unique but synergistic approaches to perform a top-down emission inventory analysis for emissions estimates in the HSC area. The three analysis tasks included 1) performing source apportionment analyses using positive matrix factorization (PMF) and comparing source contribution profiles with emissions profiles, 2) performing emissions and air quality data reconciliation using surface data; and 3) performing emissions and air quality data reconciliation using aircraft data. The results of each of the three tasks were combined and synthesized to develop recommendations for improving emissions estimates.

The data required to implement this three-task approach included: an hourly, gridded, and speciated hydrocarbon and NO_x emission inventory, speciated surface hydrocarbon and NO_x data, speciated aloft hydrocarbon and NO_x data, and wind speed and wind direction data. The

analysis focused on the abundance of ambient NMOC and NO_x data that have been newly collected in the HSC area during recent monitoring and analysis efforts.

The attached report includes separate sections on the PMF source apportionment analysis, the emission inventory reconciliation using surface ambient data, and the emission inventory reconciliation using aloft aircraft measurements, plus an integrated summary of the findings from each of the three tasks, summary recommendations for improvements to the emission inventory, and references. This Executive Summary provides an overview of the project results, of the integrated summary of the findings, and recommendations.

ES.1 RESULTS FOR SOURCE APPORTIONMENT

Hourly speciated volatile VOC data were collected at several sites in 2001 using automatic gas chromatographs (auto-GCs). These are collected to assess the characteristics (e.g., composition, ozone formation potential) of VOCs, investigate spatial and temporal variability in VOCs, and assess the capability of models to simulate the conditions that lead to episodes of high ozone concentrations. The auto-GCs record hourly concentrations of nearly 60 hydrocarbons. Other air quality measurements (such as ozone and NO_x) and meteorological data are collocated at these sites. Exploratory source apportionment of hourly VOC data from three sites near the Houston Ship Channel (HSC), Clinton Drive, Deer Park, and Haden Rd., was performed to isolate emission factors and determine if the emission inventory is representative of ambient data. Factor analysis and the Positive Matrix Factorization (PMF) receptor model were used to identify likely sources. Factor analysis tools such as PMF provide factors that can be related to emission source types and that estimate the quantitative contribution of each factor in every sample. Thus, the variation of source strength by time of day, day of week, and wind direction can be explored. These source apportionment results can reveal types of emission sources, their relative strengths, their temporal behavior, and the direction from a site at which they are located. The results from factor analysis and PMF include the following:

- *PMF was successfully applied to hourly speciated hydrocarbon data from three sites in the Houston area using nighttime data from summer 2001.* Seven factors were isolated at Deer Park and Haden Rd., and eleven factors were identified at Clinton Drive. Mass was well-reconstructed, suggesting that the identified factors accurately represented the ambient data at the three sites. **Table 2-14** summarizes the types of factors identified at each site and the percentage of the total VOC mass they accounted for. The high number of factors found at Clinton Drive is likely due to the better spatial segregation of sources around Clinton Drive and the proximity of the site to major sources. More factors may be resolvable at Haden Rd., as it is also near major sources with a good spatial segregation.
- *Similar factors were found at each site, showing that some sources impact the entire area near the HSC.* Wind direction analysis at each site showed these factors were all strongest from the same area. The following factors were observed at all three sites:
 - accumulation and natural gas usage characterized by ethane and propane;
 - industrial activity using butenes and pentenes;

- evaporative emissions containing butanes and pentanes;
 - an isoprene factor predominantly from biogenic sources but also from industrial activity;
 - petrochemical production characterized by ethene and propene; and
 - mobile sources identified by acetylene, benzene, toluene, and xylenes.
- *Butanes and pentanes are nearly ubiquitous in emissions from sources in the HSC according to the emission inventory, and these sources are spatially distributed, making it difficult for PMF to apportion these hydrocarbons.* These hydrocarbons account for more than one-third of the total VOC mass and were partitioned among factors differently at each site. The large amount of mass associated with these compounds caused similar factors at different sites to account for different amounts of mass.
 - *Mobile sources are spatially distributed and are, therefore, difficult for PMF to completely separate out from point sources in the direction of major freeways.* The “mobile” factors at each site accounted for 6% or less of the mass although, when combined with fractions of other factors (evaporative, aromatic mixes), the total influence from mobile sources is about 10% of the total mass at the three sites.
 - *Industrial sources appear to dominate the VOC composition at all three sites, though not in the same proportion as suggested by the emission inventory.* Both mobile and industrial factors were identified, and industrial factors generally made up 90% of the median VOC mass at each site. This is consistent with the emission inventory mass composition at Deer Park, but the emission inventory split at Haden Rd. was 85/15 industrial/mobile, suggesting that mobile sources may be overestimated by a factor of 1.5, and the emission inventory split at Clinton Drive was 75/25, suggesting that the industrial portion of the emission inventory is underestimated or that the mobile sources are overestimated by a factor of 2.
 - *The industrial factors had most of the reactive olefins and aromatics.* These compounds are the most reactive compounds of the VOCs analyzed and are the most important for ozone formation.
 - *Propene appears to be significantly underestimated at Deer Park.* The mean ambient composition without ethane and propane showed propene at 16% (weight percent) while the emission inventory showed propene only at 5%. Propene was predominantly in the petrochemical factor identified by PMF, suggesting that these specific source types are underestimated. Toluene (6% ambient versus 4% emission inventory weight percent) was also underestimated.
 - *Propene, toluene, and C3/C4/C5 alkylbenzenes appear to be underestimated by a factor of 2 in the emission inventory at Haden Rd.* These compounds were mostly in industrial factors, suggesting that these source types are underestimated in the emission inventory.

- *The light olefins (C2-C5) at Clinton Drive appear to be underestimated to the south of Clinton Drive.* Analysis suggests that sources of these reactive compounds to the East and South should be of similar magnitude (as they are also a similar distance), but the emission inventory shows significant sources only to the East.
- *Extremely high concentrations of industrial factors can occur at any hour of the night, any day of week, and any month during the summer.* Mobile factors were fairly constant and below 100 ppbC, while industrial factors all showed periods of extremely high concentrations (e.g., in excess of 1000 ppbC).
- *Industrial factors were consistent with the locations of similar industrial emissions according to the emission inventory.* Analysis using wind direction data helped identify the likely source areas by finding the directions where each factor had the highest concentration. Sources of C-2 to C-5 olefins, C-2 to C-8 paraffins, and alkyl benzenes were all consistent between PMF results and the emission inventory, although the magnitude was not examined.
- *Hourly speciated VOC data provide a rich and useful database for receptor-based source apportionment.* Hourly, weekday-weekend and wind direction patterns in factor strength were vital in aiding the identification of factors; these patterns can only be fully inspected with hourly data.

ES.2 RESULTS FOR EMISSIONS RECONCILIATION WITH SURFACE DATA

The purpose of performing emission inventory reconciliation is to identify sectors of the emission inventory that are may not accurately represent real-world emissions levels, spatial patterns, and/or temporal patterns. When these sectors of the inventory are identified, corrective action can be taken to improve the inventory prior to air quality modeling. Top-down comparative analyses of emissions data with surface air quality data are routinely termed “emissions reconciliation,” for the purpose of improving the inventory. An emissions reconciliation is a selective, quantitative comparison of inventory- and ambient-derived molar pollutant ratios (e.g, NMOC/NO_x) and NMOC speciation profiles. Findings often point toward weaknesses or omissions in the emission inventory, which can be iteratively remedied until the inventory data and ambient data reconcile with one another (Haste et al., 1998; Korc et al., 1995; Fujita et al., 1992).

Emissions reconciliation requires careful selection of data sets to minimize the effects of two critical confounding factors: (1) transport of aged pollutants from distant sources and (2) transformation of fresh pollutants due to photochemical reactions. When significant, these factors invalidate any direct comparison of emissions data with ambient data. However, it is possible to minimize these factors by applying emissions reconciliation techniques only to geographic areas with high emission densities and time periods when chemical reaction rates are low. Thus, one of the most basic principles of emissions reconciliation is to carefully select and analyze emissions and ambient data sets that represent morning periods in urban or industrial areas. Under these conditions, emission rates are high, mixing depths are low, and long-range transport and chemical reactions are minimal.

A second important principle of the analysis is that NO_x is the best species to select as the denominator of inventory- and ambient-derived pollutant ratios (e.g., NMOC/NO_x). This is because NO_x is a product of combustion only, and therefore is more likely to be accurately inventoried than hydrocarbons or NMOC species, which are typically emitted by a variety of combustion and non-combustion sources. Furthermore, ambient NO_x is relatively straightforward to characterize in fresh air parcels than are mixtures of NMOC.

For this study, we have selected data sets that meet the conditions described above and computed molar-based pollutant ratios (e.g., NMOC/NO_x) and speciation profiles from ambient and emission inventory data for direct comparison. The results from emissions reconciliation with surface data include the following:

- *Overall, the ratio comparisons for the median ambient and emission inventory ratios including elevated sources indicate that the ambient ratios are consistently higher than emission inventory ratios at all three sites. At Clinton, ambient ratios are about a factor of 2 to 5 higher than emission inventory ratios; at Deer Park, ambient ratios are about a factor of 4 to 12 higher than emission inventory ratios; at Haden, ambient ratios are about a factor 5 higher than emission inventory ratios. These findings indicate that overall, the emission inventory is substantially under-representing the relative amounts of TNMOC and NO_x that are observed in the ambient air during the 0500 to 0900 CST time period.*
- *In general, the TNMOC/NO_x ratios by site show the greatest agreement between the ambient data and emission inventory at the Clinton Drive site. For the entire 0500 to 0900 CST period, the median ambient ratio is about two times higher than the emission inventory ratio (with elevated sources included) for Clinton Drive in 2000 and about three times higher in 2001. However, for Deer Park and Haden Rd., the median ambient ratio is about five times higher than the emission inventory ratio. On an hour-to-hour basis, there tends to be better agreement between the ambient and emission inventory ratios after 0700 CST, when a significant increase in non-road, mobile TNMOC emissions is observed in the emission inventory data.*
- *These findings indicate that the TNMOC/NO_x ratios for Deer Park and Haden Rd. show better agreement when the ambient data is influenced more by mobile sources than industrial sources, as is the case at the Clinton Drive site. This conclusion is strengthened when TNMOC/NO_x ratios by wind quadrant are examined for each site.*
- *Analysis of the total relative compositions of the three species groups by site indicate that there is generally good agreement between the ambient data and emission inventory, with paraffins consistently making up about 60% of total TNMOC.*
- *For comparisons at all sites, the paraffin/ NO_x ratios show the largest discrepancies—the inventory ratios are about 3 to 8 times lower than observed ratios; however, the olefin/ NO_x and aromatic/ NO_x ratios are also substantially higher in the ambient median data compared to the emission inventory data including elevated sources (about 2 to 5 times higher). Moreover, the relative amounts of olefins and aromatics in the inventory have a greater impact on the ozone formation potential represented by the inventory. Because ambient levels of these species groups are much higher than reported in the*

inventory, it is likely that the inventory is under-representing the reactive potential of emissions in the Houston/Galveston area, particularly when the spatial distribution of these data is considered.

- *At the Deer Park site, when winds are from the northeast, the ambient-derived data show a spike in the fraction of propylene, whereas the emission inventory-derived data show an elevated fraction of benzene. The emission inventory-derived compositions of benzene are also significantly higher than the ambient-derived compositions at the Haden Rd. site when the winds are from the east. Examination of pre-gridded point source records indicates that the sources of these emissions at both sites are chemical manufacturing operations at Shell Oil Company and Rohm & Haas facilities northeast of Deer Park. It appears that the chemical speciation profiles used to speciate the point source inventory over-represent the relative amount of benzene (by about a factor of 2 to 5) and under-represent the relative amount of propylene emitted by these sources (by about a factor of 2).*
- *For individual aromatic species, benzene is consistently over-represented in the emissions inventory; toluene is under-represented in the emissions inventory at Clinton and to the southwest of Deer Park and the northwest at Haden Road, but fairly consistent in other directions at Deer Park; and the C-3, C-4, C-5 alkylbenzene group is under-represented to the northwest and southeast at Clinton Drive and to the southeast at Deer Park, and significantly under-represented to the northwest at Haden Road.*
- *The ambient-derived compositions of ethane and propane are consistently higher than the emission inventory-derived compositions, and the difference is especially pronounced at the Deer Park site. However, because of the low reactivity of these species, it is likely that the monitors detected aged ethane and propane emissions.*
- *The emission inventory-derived compositions of C-5 paraffins tend to be higher than the ambient-derived compositions. At the Clinton Drive site, the difference is less pronounced with winds from the west and than when winds were from the east, suggesting that point source speciation profiles may be overestimating paraffins.*
- *When winds are from the southeast, the ambient-derived data at the Clinton Drive site show a spike in the fraction of n-butane for 2000. This wind quadrant contains large petroleum and chemical manufacturing facilities, and it is likely that the chemical speciation profiles used to speciate the point source emission inventory are under-representing the relative mass of n-butane emitted by these industrial sources.*

ES.3 RESULTS FOR EMISSIONS RECONCILIATION WITH ALOFT DATA

The continuous reactive nitrogen data and VOC canister samples collected by Baylor University during 2001 and 2002 demonstrate that airborne sampling is an effective means to evaluate emissions from individual point sources and facilities. The analysis of that data presented here demonstrates the utility of the data for evaluation of VOC/NO_x emission ratios

from a number of facilities. The principal findings and recommendations based on the work presented in this section are:

- Analysis of the data collected by Baylor University during the 2001 and 2002 sampling seasons resulted in that isolation and evaluation of 9 major point source facilities in the Houston region using 55 VOC canister samples.
- Comparison of the observed Σ VOC/NO_y ratios to the expected Σ VOC/NO_x ratios indicated that the emission ratios from the facilities examined except for Oil Tanking Houston may be underestimated by factors ranging from 2 to 42.
- Comparison of the observed Σ Alkenes/NO_y ratios to the expected Σ Alkenes/NO_x ratios indicated that the emission ratios from all of the facilities examined may be underestimated by factors ranging from 2 to 64.
- Comparison of the observed Σ Aromatics/NO_y ratios to the expected Σ Aromatics/NO_x ratios indicated that the emission ratios from the facilities examined except for Oil Tanking Houston and Shell may be underestimated by factors ranging from 1.4 to 20. The observed and expected Σ Aromatics/NO_y for Oil Tanking Houston and Shell were in good agreement.
- Of the source facilities examined, only Oil Tanking Houston, for all comparisons except Σ Alkenes/NO_x, and Shell Oil, for Σ Aromatics/NO_x, had expected emission ratios comparable with the observed values.
- The method of evaluating the observed VOC/NO_y ratios as a function of the NO_x/NO_y “photochemical clock” did not provide compelling reason to use the extrapolated value as an indication of the emission ratio at the source. However, examination of the data presented in this way for several of the sources does suggest that if one samples too close to a facility, the resulting measured ratio may be biased high because inadequate mixing of the VOC and NO_x sources has taken place. At the same time, the data examined in this way for the Chevron-Phillips facility for the alkenes and aromatics do show a trend to higher values at shorter processing times. This observation can be generalized to note that for the alkene and aromatic ratios, the measured values reported may be lower limits to the zero-processing-time emission ratio.
- Uncertainty with respect to the upwind concentration of VOCs was a limiting factor in this analysis. At one extreme, this uncertainty sets the observed VOC/NO_y ratios as an upper limit for the target facility.

ES.4 CONSENSUS RESULTS

The following results are supported by results from at least two or more detailed investigations and typically by results from more than one of the three analysis approaches that were used.

- Total VOC, as identified using VOC/NO_x ratios for multiple sources, was typically under-represented in the emissions inventory by a factor of 2 to 10. For single sources, total VOC, as identified using VOC/NO_x ratios, was typically under-represented in the emissions inventory by a factor of 2 to 42.
- Most of the under-representation seems to be in the industrial facilities category.
- Olefins, and specifically propylene, are under-represented in the emissions inventory by a factor of 2 to 5. For single sources, total alkenes, as identified using VOC/NO_x ratios, was typically under-represented in the emissions inventory by a factor of 2 to 64.
- Total aromatics were also under-represented in the emissions inventory by a factor of 2 to 5. For single sources, total aromatics were typically under-represented in the emissions inventory by a factor of 1.4 to 20. For individual aromatic species, benzene is consistently over-represented in the emissions inventory; toluene is under-represented in the emissions inventory at Clinton and to the southwest of Deer Park and the northwest at Haden Road, but fairly consistent in other directions at Deer Park; and the C-3, C-4, C-5 alkylbenzene group is under-represented to the northwest and southeast at Clinton Drive and to the southeast at Deer Park, and significantly under-represented to the northwest at Haden Road.
- Total paraffins were also under-represented in the emissions inventory by factor of 3 to 8. For individual paraffin species, ethane and propane are consistently under-represented in the emission inventory, especially at the Deer Park site, while the C-5 paraffins tended to be over-represented in the emission inventory. However, because of the low reactivity of these species, these discrepancies in the emissions inventories are less important than ones with more-reactive species.

ES.5 CONSENSUS RECOMMENDATIONS

This work has shown the utility of using source apportionment tools such as PMF to identify sources of VOCs and quantify their impacts at a monitoring site. This work provides a basis for subsequent, similar analyses. We recommend the following analyses be considered:

- Inspect PMF time series results to see if reported industrial upsets can be identified in the data.
- Inspect events of extremely high concentrations observed in the ambient data to see if they match reported upsets.
- Triangulate similar factors found at the three sites to see if a more specific source area can be identified using wind speed, direction, and relative strength of individual factors.
- Run PMF on data limited to the morning hours used in the emission inventory analysis (0500-0900 CST) to better examine the temporal variations of the ambient data, and to

investigate how accurate the temporal profiles of specific emission inventory types (such as off-road mobile) are.

- Run PMF on data limited to individual weeks of interest to further isolate minor sources that can have infrequent but significant emissions; this approach has been pioneered by Dr. Matt Fraser (Rice University) using Clinton Drive data (Buzcu and Fraser, 2003).
- Scale PMF results by their ozone formation potential to examine the reactivity of individual factors and whether these factors are high on the mornings of ozone episodes.
- Further analyze identified factors using ratios with TNMOC and NO_x to compare to the emission inventory.
- Collect sufficient carbonyl data at one of the auto-GC sites and run PMF including these species. The carbonyl compounds are important to ozone formation, are air toxics, and can comprise a significant part of the TNMOC.

This work has identified specific improvements in the emissions inventory that could be made, such as the following:

- Non-road emissions appear to be high in general, particularly TNMOC emissions occurring from 0700 to 0900 CST around the Clinton Drive site, where non-road TNMOC emissions are approximately the same as point source TNMOC emissions. Lawn and garden equipment activity is the main contributor to non-road TNMOC emissions for Harris County as a whole. Emissions of NO_x from activities at Ellington Field (south of the Deer Park site) appear to be overestimated and should be re-evaluated.
- Since TNMOC/NO_x ratios show the greatest discrepancy between the ambient- and emission inventory-derived values at sites (and wind quadrants) dominated by industrial sources, it appears that point source TNMOC emissions are underestimated, NO_x emissions are overestimated, or both. However, given the prevalence of refineries and chemical manufacturing facilities in the region and the difficulties inherent in estimating NMOC emissions, it seems more likely that NMOC emissions are underestimated. The point source NMOC inventory should be evaluated and corrected as necessary.
- Overall, the speciation profiles used to prepare the emission inventory track relatively well with the weight percentages of organic species (and species groups) detected by the ambient monitoring system. However, propylene, n-butane, and ethane were under-represented and benzene and C-5 paraffins were over-represented in the emission inventory.
- The point source emission inventory should be assessed to determine which sources contribute the most mass for the compounds listed above. The chemical speciation profiles used to speciate these sources should be assessed for the compounds listed above. Also, speciation profiles used to speciate emissions from refineries and chemical manufacturing should be assessed.

Additional recommendations regarding future emissions inventory reconciliation include the following:

- This same emission inventory reconciliation analysis should be performed for more sites within the Houston/Galveston area.
- The recommended changes to the emission inventory should be implemented and the emission inventory reconciliation should be performed again based on the updated inventory.

Similar ambient-to-emission inventory comparisons using aloft data should be performed with the data collected in 2003. In addition, future data collection campaigns using aircraft should strive to include an upwind canister for each set of downwind samples collected. The resulting data would significantly reduce the uncertainty that exists in the current data with respect to the upwind concentration of the VOC species.