Title:

Exploratory Data Analysis of Benzene and 1,3-Butadiene Measurements for Air Toxics Risk Assessment in Houston

CRC Project No. A-48

Prepared By:

Rick Reiss and John Griffin
Sciences International, Inc.
Alexandria, Virginia

Prepared For:

Coordinating Research Council
3650 Mansell Road, Suite 140
Alpharetta, Georgia 30022

Date:

August 1, 2005

SCIENCES INTERNATIONAL, INC.

LIST OF ACRONYMS

AIRS	Aerometric Information Retrieval System
ANOVA	Analysis of Variance
API	American Petroleum Institute
CV	Coefficient of Variance
DQO	Data Quality Objective
EPA	Environmental Protection Agency
LADCO	Lake Michigan Air Directors Consortium
LCS	Laboratory Control Standard
NATA	National Air Toxics Assessment
NATTS	National Air Toxics Trends Stations
TCEQ	Texas Council on Environmental Quality
UTM	Universal Transverse Mercator

EXECUTIVE SUMMARY

This report presents an analysis of air toxics data in the Houston, Texas, area. Specifically, the report focuses on benzene and 1,3-butadiene, which are key drivers in air toxics risk assessment. The report includes two major parts: (1) an analysis of the sources of variation in the air toxics data, and (2) a comparison of monitoring data with modeling estimates.

All data for benzene and 1,3-butadiene in the Houston area were obtained from the Texas Commission on Environmental Quality (TCEQ) for the period of 1996-2003. The data included 15 stations, including one collocated station at Port Arthur. Most of the stations are east of the Houston city center in the Houston Ship Channel area, and there are several stations to the west of Houston. In Section 2, some summary statistics were calculated and the trends in concentration data over time were analyzed. We reached the following conclusions from this analysis:

- For both benzene and 1,3-butadiene, the concentrations were highest at the stations east of Houston in the Ship Channel as would be expected given the higher amount of industrial emissions.
- There is a general declining trend that is evident in the benzene and 1,3-butadiene measurement data over time. Across all of the monitors, the trend for benzene was -8.1% per year, and the trend for 1,3-butadiene was -14.5% per year. Because of several issues discussed in the report, this analysis cannot be considered definitive.

In Section 3, the variance in the data was analyzed in three separate analyses: (1) variation as the result of the analytical measurement of the data was analyzed by comparing the results of duplicate spiked samples from the TCEQ laboratory, (2) variation as the result of the sampling (which includes the analytical variation) was analyzed by comparing collocated measurements at Port Arthur, and (3) environmental variability, including spatial and temporal variability, was analyzed statistically. The following conclusions were made from these analyses:

- The variance caused by the analytical method was generally low, with the coefficients of variance (CV) due to analytical variability at 1.9% for benzene, and 2.6% for 1,3-butadiene.

- The CV due to sampling variability was 13% for benzene and 29% for 1,3-butadiene, and the agreements between measured and monitored values at a single collocated site were fairly good.
- The estimated CVs for spatial and temporal variability for benzene were 29% and 24%, respectively. These values were lower than found by Battelle in its 10 city study, but the temporal variability is likely to be a low-end estimate as discussed in the report.
- The estimated CVs for spatial and temporal variability for 1,3-butadiene were 98-118% and 185-217%, respectively, depending on how samples below the LOD were handled. These values are very large compared to values in the Battelle report.

In Section 4, the monitoring data were compared with modeling estimates from American Petroleum Institute's (API's) ongoing modeling project in the Houston Ship Channel (conducted by ENSR) and from EPA's National Air Toxics Assessment (NATA) modeling analysis of the continental United States. There are two monitoring stations within the API domain. The NATA domain covers the whole Houston area. The following conclusions were reached from these analyses:

- For benzene, the API model estimates agreed fairly well. Compared to the API modeling estimates, the 1996 average concentration at Haden Road was 20% lower and the average concentration at Deer Park was 36% lower.
- For 1,3-butadiene, the API model estimate agreed fairly well with the average 1996 concentration at Haden Road (+11% for the model estimate). However, the model estimate at Deer Park was significantly higher than the average 1996 concentration.
- When compared to the NATA model estimates, which are available for all of the monitoring sites, the average percent difference for benzene (including the average of all years of data) was -7.4% (monitor concentrations higher than model estimates); however, there were larger differences for some sites, although some large differences canceled out other large differences in the opposite direction. The average absolute difference between the model estimates and measurements was 25%. The model estimates did explain some of the spatial variability in the observed concentrations.
- For 1,3-butadiene, the NATA model estimates were about 3-fold lower than the monitor concentrations, indicating a significant underestimate with the model.

- The declining trends in concentration were not related to the contribution of major or on-road mobile sources at the sites. Therefore, the trends cannot be related to reductions in major or on-road emissions from this analysis.

1.0 INTRODUCTION

This report presents an analysis of air toxics data in the Houston, Texas, area. Specifically, the report focuses on benzene and 1,3-butadiene, which are key drivers in air toxics risk assessment. The first part of the report presents an analysis of the sources of variance in measured concentrations of benzene and 1,3-butadiene. This analysis relies on the publicly available monitoring data from the Texas Commission on Environmental Quality (TCEQ). The data that were obtained from the TCEQ are described in more detail in the next section.

The second part of this report provides a comparison of the benzene and 1,3-butadiene monitoring data with modeling estimates that are being developed in a another project sponsored by the American Petroleum Institute (API) and conducted by the consulting firm ENSR. This modeling analysis focuses on the Houston Ship Channel. Additionally, modeling estimates from the U.S. EPA's National Air Toxics Assessment (NATA), which include the entire continental United States, were compared to the monitoring data.

The remainder of the report is organized as follows:

- Section II – Description of data sources
- Section III – Analysis of sources of variance in the monitoring data
- Section IV – Comparison of modeling and monitoring data

2.0 DESCRIPTION OF DATA SOURCES

2.1 Available Monitoring Stations

All benzene and 1,3-butadiene data were requested from TCEQ from 1996 to the present. We received the following data:

- Concentration data from 15 stations in the Houston area
- Collocated measurement data from a site in Port Arthur
- Recovery data for duplicate spiked samples from the laboratory

Table 2-1 summarizes the list of 15 stations with available data, including the Aerometric Information Retrieval System (AIRS) number, station name, latitude/longitude coordinates, elevation, and years with available data. Most stations had data available from 1996 to 2003. The Allendale site was deactivated in 1998. The Galena Park station began operations in 1997, the Harris County, Bayland Park, and Baytown stations began operations in 1997, and the Shore Acres station began operations in 1998.

The locations of the stations are displayed on Figure 2-1. Most of the stations are to the east of Houston, and many are in the Ship Channel area. The only stations that are not to the east of the Houston city center are Bayland Park (southwest), Aldine (north), and Northwest Harris County (northwest).

2.2 Basic Statistics

Tables 2-2 and 2-3 provide summary statistics for benzene and 1,3-butadiene at each of the stations. For benzene, there were a total of 4,597 measurements, with only 2 samples below the limit of detection (LOD) of 0.005 ppb. The highest average concentrations were measured at Galena Park (1.5 ppb), Lynchburg Ferry (1.4 ppb, though only 9 samples), San Jacinto Monument (1.3 ppb), and Shore Acres (1.1 ppb). These sites are in the Ship Channel area, although Shore Acres is southeast of the Ship Channel. The three sites with the lowest concentrations were all west of Houston, as expected, including Aldine (0.66 ppb), Bayland Park (0.47 ppb), and Northwest Harris County (0.43 ppb).

For 1,3-butadiene, there were 4,596 measurements. Across the sites, an average of 43% of the samples were below the LOD. The LOD was 0.005 ppb. The highest average concentrations were measured at Milby Park (3.0 ppb) and Allendale (1.2 ppb), which are both close to one another, east of Houston, and south of the Ship Channel. The three sites with the lowest concentrations were all west of Houston, as expected, including Aldine (0.17 ppb), Bayland Park (0.10 ppb), and Northwest Harris County (0.06 ppb).

The Milby Park results were highly skewed with some high concentrations significantly impacting the average. There was not a discernible temporal pattern to the peak concentrations. For example, on 6/5/2000, the concentration was 0.04 ppb, while six days later on 6/11/2000, the concentration was 30.9 ppb. There were other similar examples. One of the largest 1,3-butadiene sources, Texas Petrochemical, is within about one kilometer of the Milby Park station, and the Allendale station, which had the second higher average 1,3-butadiene concentration. The Texas Petrochemical plant is southeast of the two monitors. One possibility is that the peak concentrations at these monitors correlate with wind directions that move the emissions from the plant to the monitors. This could be investigated with an analysis of meteorological data.

Table 2-1. Air Toxics Monitoring Sites in the Houston Area

AIRS	Name	Latitude	Longitude	Elevation (meters)	Data Availability
482010024	Aldine	29.901	-95.326	59	1996-2003
482010064	Allendale	29.700	-95.267	n/a	1996-1999
482010055	Bayland Park	29.696	-95.499	59	1998-2003
482010058	Baytown	29.771	-95.031	20	1998-2003
482010026	Channelview	29.803	-95.126	20	1996-2003
482011035	Clinton	29.734	-95.258	20	1996-2003
482011039	Deer Park	29.670	-95.129	20	1996-2003
482010057	Galena Park	29.734	-95.238	20	1997-2003
482010803	HRM-3 Haden Road	29.765	-95.181	20	1996-2003
482011015	Lynchburg Ferry	29.764	-95.078	20	2003
482010069	Milby Park	29.706	-95.261	20	1999-2003
482010029	Northwest Harris Co	30.039	-95.674	180	1998-2003
482450011	Port Arthur West	29.894	-93.988	0	1996-2003
482011041	San Jacinto Monument	29.752	-95.083	20	1996-2003
482010061	Shore Acres	29.615	-95.018	0	1998-2003

n/a = not available

Figure 2-1. Locations of Air Monitoring Stations in the Houston Area

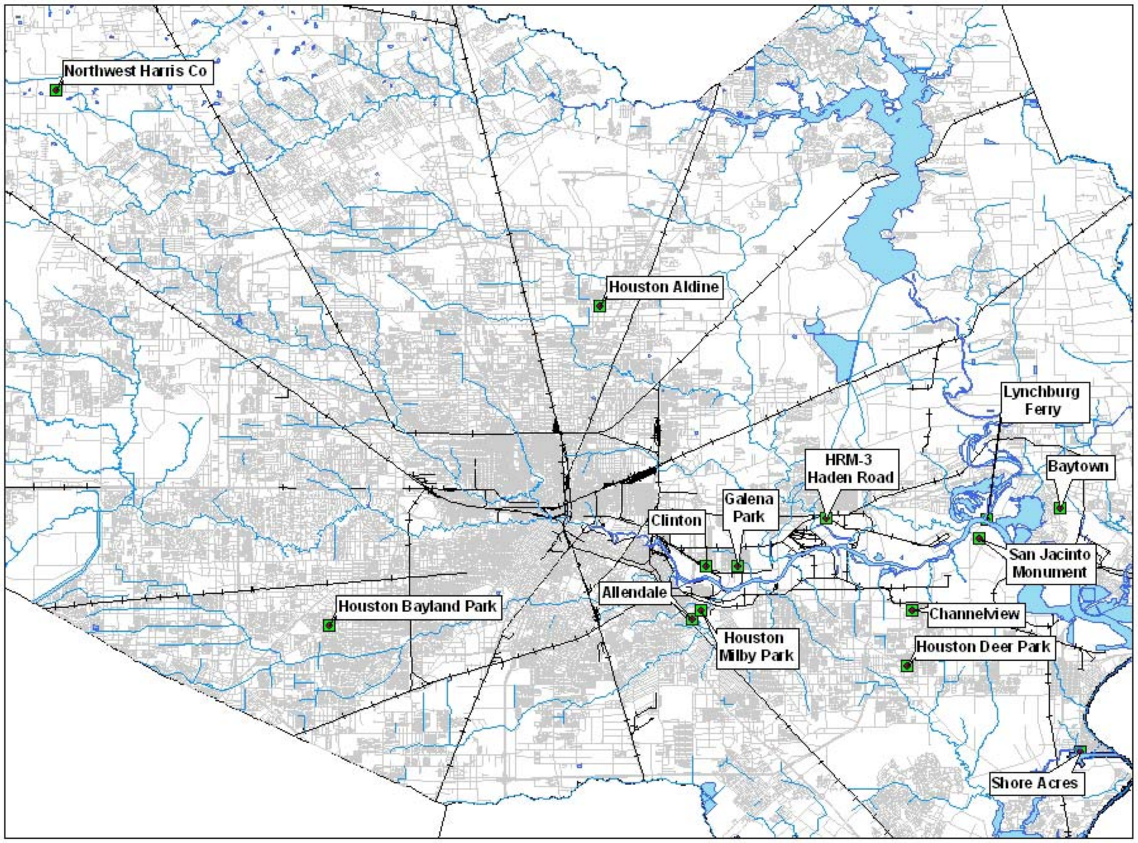


Table 2-2. Summary Statistics for Benzene

Station	Number of Samples	Percent Non-Detect	Concentration (ppb)		
			Average[a]	Min[b]	Max
Aldine	431	0%	0.66	0.20	2.8
Allendale	165	0%	0.89	0.11	8.4
Bayland Park	277	0%	0.47	0.04	2.2
Baytown	269	0%	0.59	0.12	3.6
Channelview	397	0.3%	0.95	0.025	7.7
Clinton	398	0%	0.91	0.01	11.5
Deer Park	363	0%	0.71	0.10	4.3
Galena Park	305	0%	1.5	0.18	10.3
HRM-3 Haden Road	412	0%	0.93	0.17	7.9
Lynchburg Ferry	9	0%	1.4	0.26	5.2
Milby Park	192	0%	0.70	0.15	12.5
Northwest Harris Co	276	0%	0.43	0.11	1.6
Port Arthur - I	378	0%	0.69	0.12	5.1
Port Arthur - II	333	0%	0.71	0.10	3.7
San Jacinto Monument	110	0%	1.3	0.13	10.4
Shore Acres	282	0.4%	1.1	0.025	41.0

[a] Used half the LOD for samples below the LOD.

[b] Not including samples below the LOD.

Table 2-3. Summary Statistics for 1,3-Butadiene

Station	Number of Samples	Percent Non-Detect	Concentration (ppb)		
			Average[a]	Min[b]	Max
Aldine	431	44%	0.17	0.01	3.2
Allendale	165	28%	1.2	0.01	32.8
Bayland Park	277	62%	0.10	0.01	1.8
Baytown	269	43%	0.25	0.02	5.0
Channelview	397	17%	0.47	0.01	6.2
Clinton	398	23%	0.55	0.01	8.4
Deer Park	363	47%	0.20	0.02	2.9
Galena Park	305	18%	0.45	0.01	4.3
HRM-3 Haden Road	412	23%	0.41	0.01	4.1
Lynchburg Ferry	9	63%	0.16	0.02	0.69
Milby Park	192	18%	3.0	0.025	31.6
Northwest Harris Co	276	84%	0.06	0.01	2.8
Port Arthur - I	378	49%	0.17	0.01	3.3
Port Arthur - II	332	54%	0.16	0.01	3.1
San Jacinto Monument	110	46%	0.43	0.01	9.6
Shore Acres	282	66%	0.14	0.01	6.8

[a] Used half the LOD for samples below the LOD.

[b] Not including samples below the LOD.

2.3 Trends over Time

There are sufficient data to estimate linear concentration trends over time. Tables 2.4 and 2.5 show the annual average concentrations for benzene and 1,3-butadiene and the annual percent trend in the concentrations. More sophisticated models could be developed that consider non-linear trends and the affect of meteorology, but this simple model provides useful information given the small dataset and goals of the analysis.

For benzene, there was a decreasing trend at 11 of the stations (there was only one year of data at Lynchberg Ferry so no trend could be calculated). There was no trend at Galena Park and an increasing trend of 5.0% per year at Shore Acres. The decreasing trends ranged from -2.2% per year (Northwest Harris Co.) to -41.1% (San Jacinto Monument). The only trends that were statistically significant were Aldine (-3.5%), Clinton (-7.7%), Haden Road (-8.0%), and San Jacinto Monument (-41.1%), although some other stations had marginally significant trends (i.e., p-values between 0.05 and 0.10). Nonetheless, given the predominance of stations that had decreasing trends, it is reasonable to suggest that there is an actual declining trend in benzene concentrations. It is possible that the trend is caused by changes in the measurement technology; however, this could not be investigated given the available data. A more definitive analysis would include the impact of meteorology and missing data on the concentrations.

For 1,3-butadiene, there was a decreasing trend at all of the stations. The trends ranged from -0.2% per year (Allendale) to -59.0% (Northwest Harris Co.). However, the only trends that were statistically significant were Clinton (-18.1%) and Baytown (-16.5%), although some others were marginally significant. Nonetheless, given that all of the stations had decreasing trends, it is reasonable to suggest that there is an actual declining trend in 1,3-butadiene concentrations. It is possible that the trend is caused by changes in the measurement technology; however, this could not be investigated given the available data. A more definitive analysis would include the impact of meteorology and missing data on the concentrations.

Table 2.4. Annual Average Concentrations and Annual Percent Trend for Benzene

Year	1996	1997	1998	1999	2000	2001	2002	2003	Annual Percent Trend[a]
Aldine	0.69	0.73	0.66	0.77	0.61	0.69	0.56	0.54	-3.5%*
Allendale	0.73	1.27	0.77	0.61					-9.7%
Bayland Park			0.51	0.62	0.44	0.48	0.42	0.39	-6.8%
Baytown			0.62	0.65	0.63	0.52	0.54	0.55	-3.8%
Channelview	1.13	1.18	1.01	0.99	0.82	1.25	0.64	0.72	-6.1%
Clinton	1.12	1.07	1.07	0.99	0.82	0.86	0.62	0.71	-7.7%*
Deer Park	0.77	0.79	0.78	0.82	0.53	0.67	0.69	0.61	-3.7%
Galena Park		1.56	1.51	1.50	1.24	1.71	1.39	1.57	0.0%
HRM-3 Haden Road	1.18	1.29	0.89	0.94	0.84	0.91	0.61	0.77	-8.0%*
Lynchburg Ferry								1.43	n/a
Milby Park				1.27	0.46	0.75	0.64	0.52	-17.3%
Northwest Harris Co			0.40	0.50	0.41	0.48	0.35	0.41	-2.2%
San Jacinto Monument					1.59	1.48	0.84	0.27	-41.1%*
Shore Acres			0.63	1.23	0.86	0.95	1.44	0.82	+5.0%

* Statistically significant, $p<0.05$

[a] Based on a linear trend line. Predicted(year 1)*(1-Trend Per Year)^7= Predicted(year 8), which simplifies to Trend Per Year = 1 – (exp(ln(Pred(8)/Pred(1))/7)

Table 2.5. Annual Average Concentrations and Annual Percent Trend for 1,3-Butadiene

Year	1996	1997	1998	1999	2000	2001	2002	2003	Annual Percent Trend[a]
Aldine	0.10	0.21	0.31	0.21	0.12	0.10	0.08	0.10	-10.5%
Allendale	0.85	1.02	2.02	0.51					-0.2%
Bayland Park			0.15	0.11	0.07	0.10	0.04	0.08	-16.1%
Baytown			0.38	0.31	0.22	0.18	0.18	0.18	-16.5%*
Channelview	0.33	0.70	0.73	0.36	0.47	0.27	0.42	0.44	-4.8%
Clinton	0.69	0.80	1.14	0.43	0.41	0.24	0.28	0.33	-18.1%*
Deer Park	0.05	0.33	0.28	0.22	0.12	0.12	0.11	0.20	-4.1%
Galena Park		0.47	0.78	0.44	0.48	0.40	0.32	0.29	-11.6%
HRM-3 Haden Road	0.39	0.66	0.70	0.46	0.26	0.23	0.28	0.34	-11.0%
Lynchburg Ferry								0.14	n/a
Milby Park				3.47	4.41	2.37	2.12	2.86	-11.1%
Northwest Harris Co			0.12	0.04	0.03	0.03	0.00	0.03	-59.0%
San Jacinto Monument					0.63	0.34	0.24	0.36	-21.5%
Shore Acres			0.10	0.16	0.09	0.18	0.10	0.09	-3.3%

* Statistically significant, $p<0.05$

[a] Based on a linear trend line. Predicted(year 1)*(1-Trend Per Year)^7= Predicted(year 8), which simplifies to Trend Per Year = 1 – (exp(ln(Pred(8)/Pred(1))/7)

3.0 ANALYSIS OF SOURCES OF VARIANCE IN THE MONITORING DATA

This section of the report provides an analysis of the sources of variance in the monitoring data, including:

- Analytical measurement variation, including the uncertainty in the laboratory measurements.
- Sampling variation, which includes the variation associated with sample collection and analysis.
- Environmental variation, including spatial and temporal variation.

3.1 Analytical Measurement Variation

TCEQ provided results of duplicate laboratory samples of spiked samples. In other words, a sample was spiked with a known amount of benzene or 1,3-butadiene and run through the Laboratory Control Standard (LCS) analysis twice to estimate the analytical precision of the method. There were 255 replicate measurements for both benzene and 1,3-butadiene. As described in the Battelle report (2003), the coefficient of variance (CV) can be estimated from duplicate measurements by computing the replicate differences after the data have been log-transformed (natural log). These differences are multiplied by $1/\sqrt{2}$ so that resulting values follow a normal distribution with a mean of zero and a variance equal to the analytical uncertainty of the log concentrations. The %CV can be estimated with the following formula[1]:

$$\%CV = 100\sqrt{e^{s^2} - 1} \tag{3-1}$$

where s^2 is the sample variance of the replicate differences. The confidence bounds can be estimated based on a chi-squared distribution with *N-1* degrees of freedom, where *N* is the number of replicates.

Figures 3-1 and 3-2 display the results of the replicate analyses. The replicate analyses were run with a 2 ppb control standard, and the results are reported as a percentage recovery. The results show that the analytical precision is fairly good.

[1] The coefficient of variance is defined as the standard deviation divided by the mean, and is a measure of the variation in the data relative to the mean value.

The average percent difference between replicates for the benzene data was 0.36%, and the average percent difference between replicates for the 1,3-butadiene data was 0.59%.

Table 3-1 summarizes the results of the %CV calculations. According to the National Air Toxics Trends Stations' (NATTS) Data Quality Objectives (DQO), the goal is for a laboratory to have a CV below 10%. The CVs for both benzene and 1,3-butadiene are well below the 10% goal, reflecting the relatively close agreement between the replicate measurements. In the Battelle analysis, most of the laboratories met the 10% goal for benzene, but none met the goal for 1,3-butadiene. This shows that the Texas laboratory performed better for 1,3-butadiene than other laboratories across the country.

It is important to note that this analysis is not necessarily representative of the analytical variability for samples at environmental concentrations, which are, on average, somewhat lower than the control standard samples at 2 ppb. In particular, the variability is likely to be greater for samples near the LOD, which was true for many of the 1,3-butadiene measurements.

Table 3-1. Percent Coefficient of Variances for Analytical Precision

Compound	%CV	Lower 95th %CV	Upper 95th %CV
Benzene	1.9	1.8	2.1
1,3-Butadiene	2.6	2.4	2.9

Figure 3-1. Comparison of Replicate Analytical Measurements for Benzene

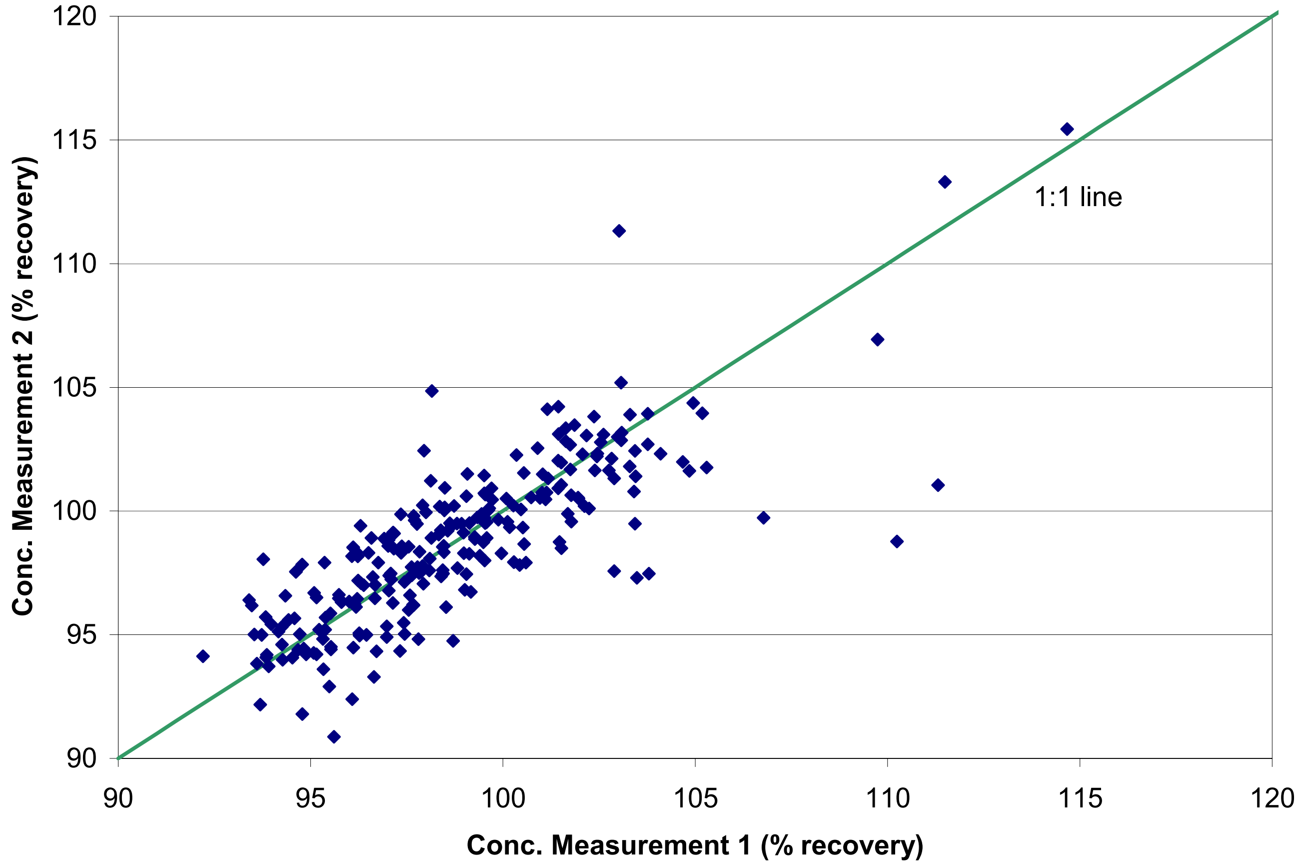
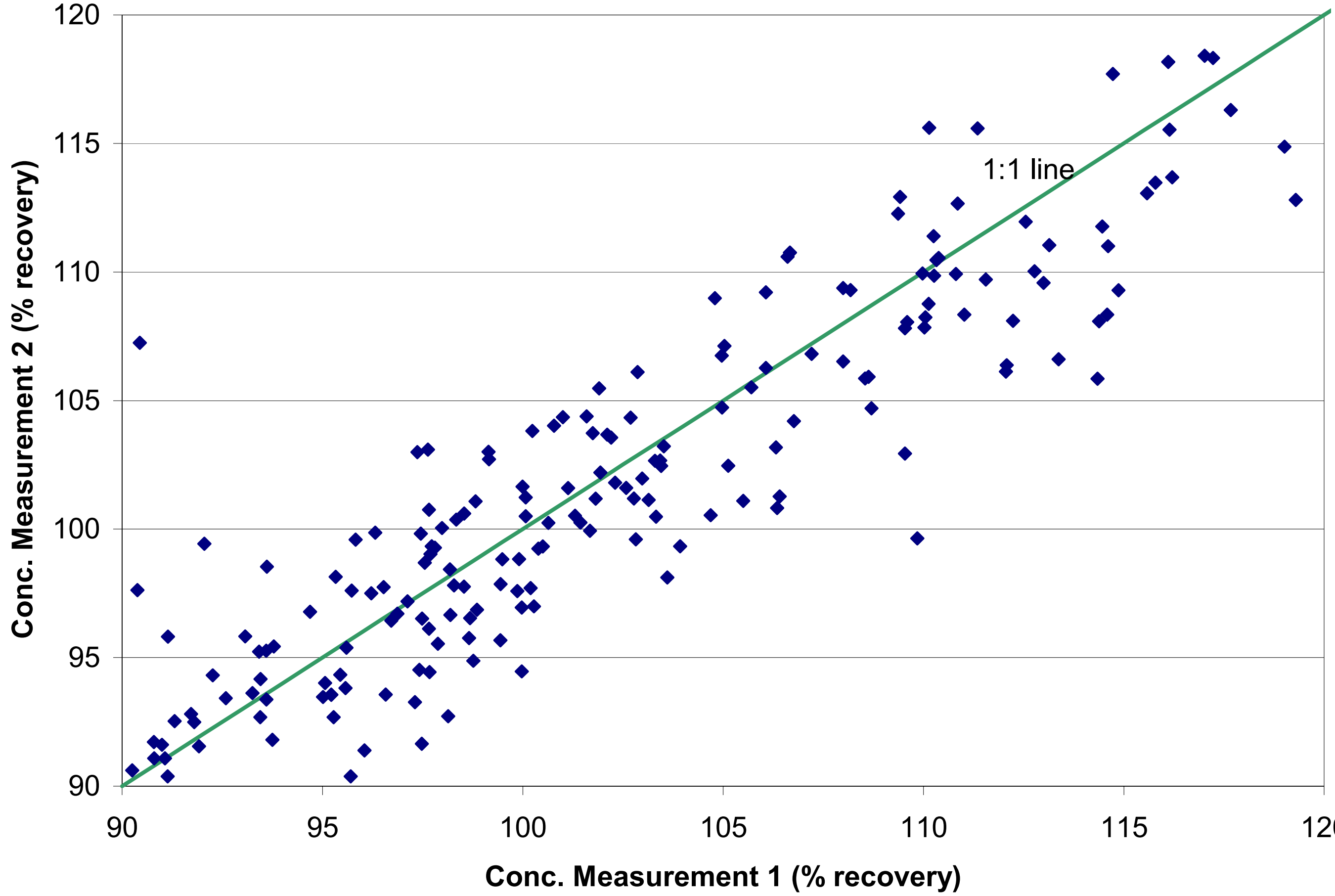


Figure 3-2. Comparison of Replicate Analytical Measurements for 1,3-Butadiene



3.2 Sampling Variation

The sampling variation can be estimated from collocated measurements. For Houston, there is a collocated station at Port Arthur, and the station has both historical 1,3-butadiene and benzene collocated measurements. At the first station at this site, there were 378 measurements of both 1,3-butadiene and benzene. At the second station, there were 332 1,3-butadiene measurements and 333 benzene measurements. When the values were paired, there were 327 paired measurements for 1,3-butadiene and 328 paired measurements for benzene. However, many of the 1,3-butadiene measurements were below the detection limit. To have a valid comparison, it is necessary for both collocated measurements to be above the detection limit. Using this criteria, there were 131 paired measurements for 1,3-butadiene.

Figures 3-3 and 3-4 display the collocated results for benzene and 1,3-butadiene at Port Arthur. The comparison for benzene was fairly good with an r^2=0.87. The 1,3-butadiene comparison was also good (r^2=0.91).

Table 3-2. Percent Coefficient of Variances for Sampling Precision

Compound	%CV	Lower 95th %CV	Upper 95th %CV
Benzene	13.4	12.4	14.5
1,3-Butadiene	29.1	25.9	33.4

Figure 3-3. Collocated Measurements of Benzene at Port Arthur

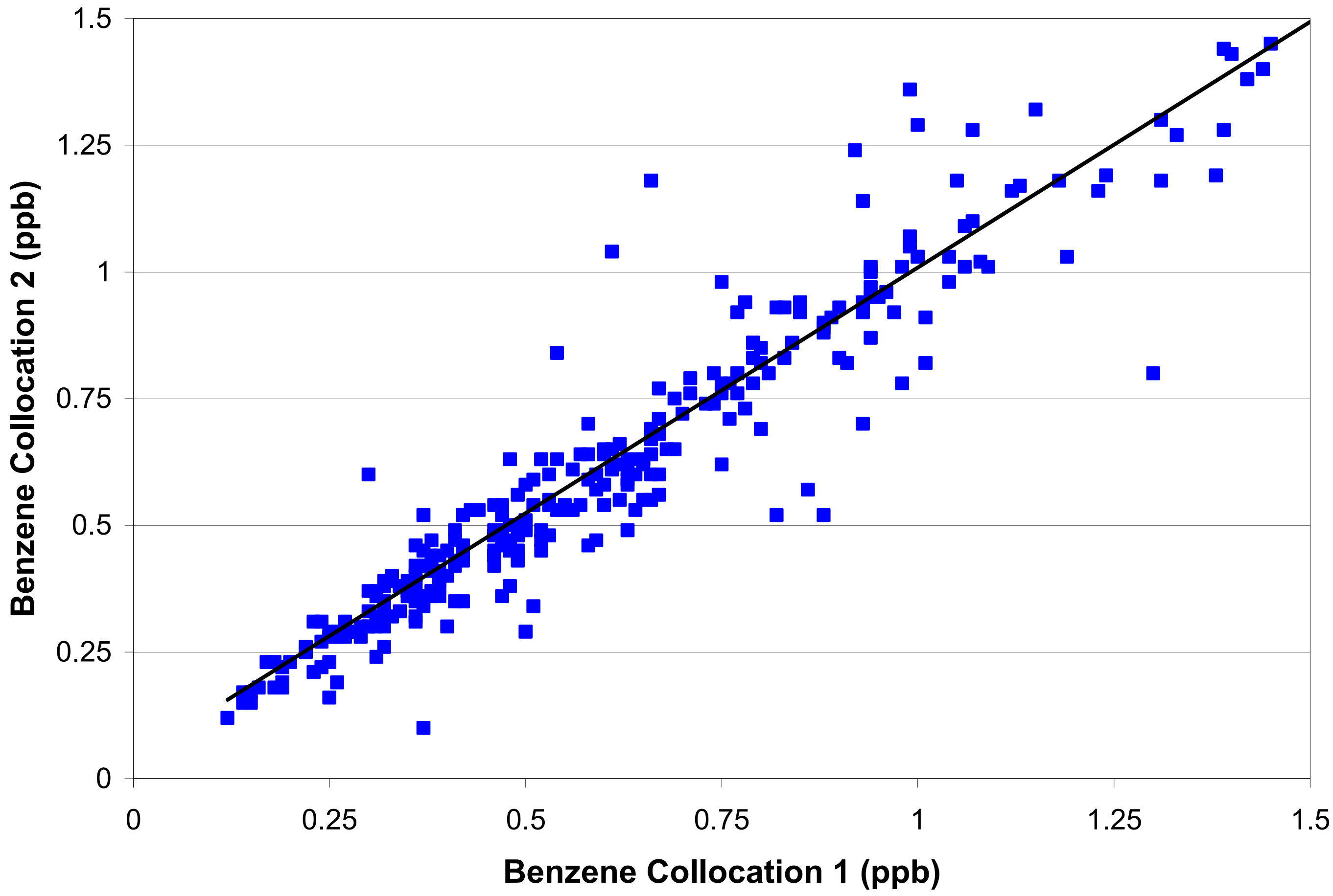
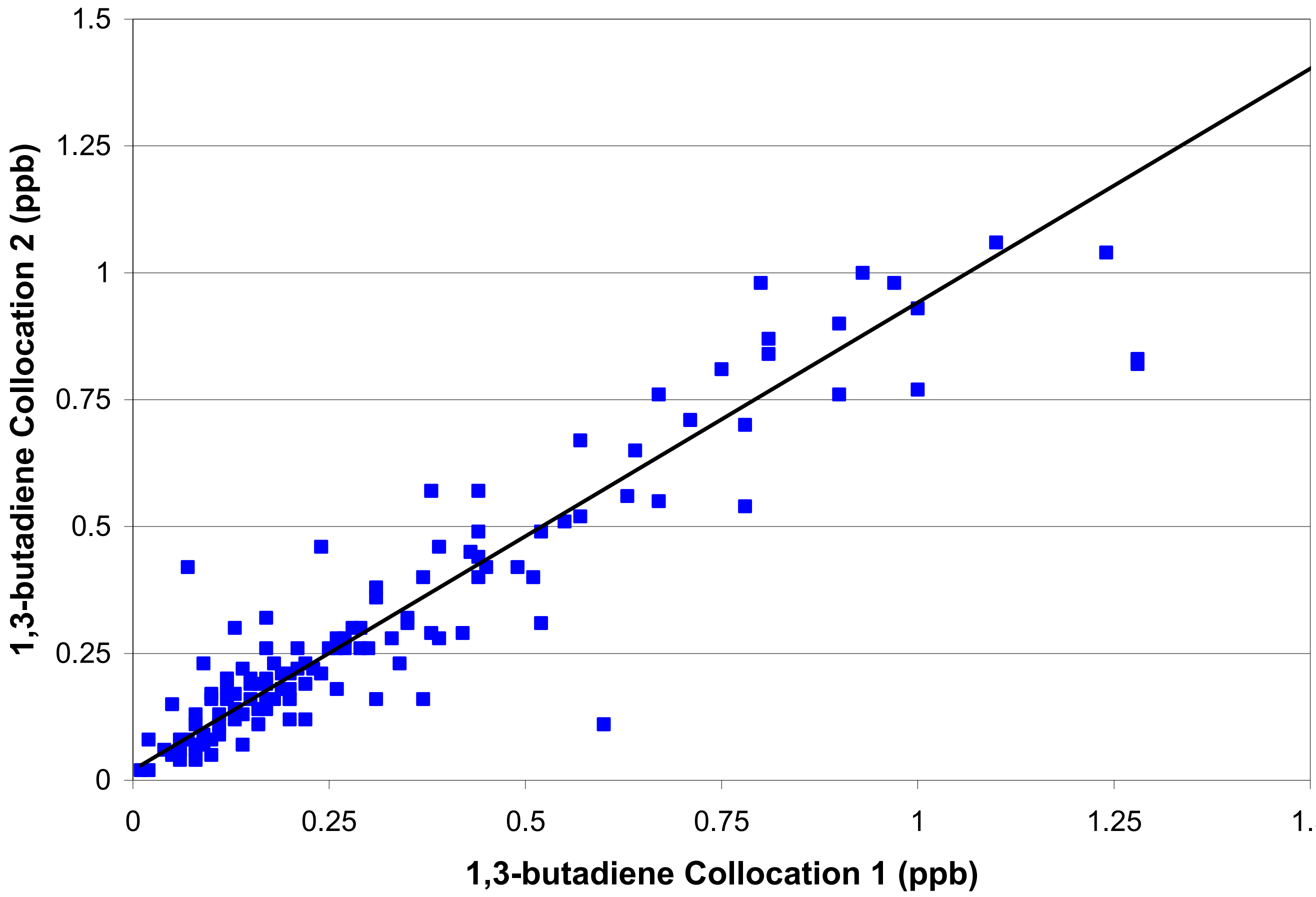


Figure 3-4. Collocated Measurements of 1,3-Butadiene at Port Arthur



3.3 Environmental Variation

The environmental variation includes the spatial variability (i.e., between stations) and temporal variability (i.e., across time). Figures 3.5 though 3.10 are box-whisker plots showing the variability by station (Figures 3.5 and 3.6), by season[2] (Figures 3.7 and 3.8), and by month (Figures 3.9 and 3.10) for benzene and 1,3-butadiene.

The seasonal variability appears modest for benzene. There are slightly elevated concentrations in the colder months (winter and fall), but the difference is not substantial. There is some significant variability between stations. For 1,3-butadiene, the variability looks similar. There is significant variability between stations, but the seasonal and monthly variability is less pronounced. The greater number of outliers above the mean compared to below the mean shows the significant number of data points below the detection limit and the skewness of the data for 1,3-butadiene. These plots use half the detection limit for the non-detect values.

One methodology for apportioning the sources of variability is to use a random effects analysis of variance (ANOVA) model. For example, for Lake Michigan Air Directors Consortium (LADCO), Battelle analyzed air toxics data from the 10 pilot cites in EPA's Air Toxics Program, using the following model (Battelle, 2003):

$$\ln(y_{ijkl}) = \mu + \alpha_i + \beta_{j(i)} + \tau_{k(ij)} + \varepsilon_{l(ijk)} \qquad (3\text{-}2)$$

where $\ln(y_{ijkl})$ is the natural logarithm of the air toxic concentration for the l^{th} observation on the k^{th} day at the j^{th} monitoring location in the i^{th} city, μ is the overall mean of the logarithm of concentration, α_i is the random deviation attributable to the effect of city *i*, $\beta_{j(i)}$ is the random deviation from a random-city specific mean that is attributable to the effect of monitoring location *j* with city *i*, $\tau_{k(ij)}$ is the random deviation from a random (city x site)-specific mean that is attributable to the effect of the k^{th} sample collection day in the i^{th} city and j^{th} monitoring location, and $\varepsilon_{l(ijk)}$ is the residual error term of the model. The l^{th} observation corresponded to either collocated monitors, duplicate samples, or replicate analyses. The logarithm of the concentration was used to better approximate normality.

[2] Winter is defined as Dec-Feb, Spring is defined as Mar-May, Summer is defined as Jun-Aug, and Fall is defined as Sep-Nov.

The LADCO project included data from different cities, while this project only includes data from Houston. Therefore, the analogous model parameterization is as follows:

$$\ln(y_{jkl}) = \mu + \beta_j + \tau_{k(j)} + \varepsilon_{l(jk)} \tag{3-3}$$

where $\ln(y_{jkl})$ is the natural logarithm of the air toxic concentration for the l^{th} observation on the k^{th} day at the j^{th} monitoring location, μ is the overall mean of the logarithm of concentration, β_j is the random deviation that is attributable to the effect of monitoring location j, $\tau_{k(j)}$ is the random deviation from a random station-specific mean that is attributable to the effect of the k^{th} sample collection day at the j^{th} monitoring location, and $\varepsilon_{l(jk)}$ is the residual error term of the model.

For 1,3-butadiene, there were a significant number of measurements below the LOD. Therefore, the model was run two ways for 1,3-butadiene: (1) with replacement of the LOD values with the one-half the LOD, and (2) with deletion of the LOD values.

This model was fit using the S statistical programming language. The estimated variances for spatial and temporal variability, and the residuals (i.e., unexplained variability) are summarized in Table 3.3. There were large CVs for the residuals (67% and 108% for benzene and 1,3-butadiene, respectively). The likely reason is that there was only a small amount of replication in the dataset (i.e., the Port Arthur station). This means that the $\tau_{k(j)}$ term in the model (representing temporal variability) is confound by the residual term, which has the same factors, when there is no replication. Therefore, the temporal variability terms must be considered low-end estimates, as some of the temporal variability cannot be separated from the residuals (Pinheiro and Bates, 2000).

Table 3.3. Estimated CVs for Spatial and Temporal Variability

Compound	Coefficients of Variance (95% CI)		
	Spatial Variability	Temporal Variability	Residuals
Benzene	29% (17-29%)	24% (19-27%)	67% (64-70%)
1,3-Butadiene (with replacement)	118% (80-191%)	185% (169-207%)	108% (98-121%)
1,3-Butadiene (deletion of samples below LOD)	98% (53-224%)	217% (204-235%)	30% (24-35%)

For benzene, the spatial and temporal variability were less than estimated in the Battelle analysis. The spatial variability in the Battelle analysis was 52% compared to 29% for Houston[3]. The temporal variability was 60% compared to 24% for Houston (but the 24% estimate is likely a low-end value). It is hard to speculate on a reason for these differences, but one cause may be that so many of the monitors in Houston are in the Ship Channel area causing much less variability in the results.

For 1,3-butadiene, the CVs were very large, much larger than the values in the Battelle study for 1,3-butadiene or any other air toxic. The residual CV was significantly reduced for the model with replacement. The likely reason is that the removal of the values below the LOD resulted in a distribution that more closely approximately normality. The high temporal variability despite the modest seasonal effect, suggests that particular meteorological conditions that may occur anytime of the year result in high concentrations.

The lower spatial and temporal variability for benzene suggests that benzene concentrations are dominated by a continuous local background from the many benzene sources in the area. Benzene emitted from the vehicles in the area is likely re-circulating for several days which dampens the variability caused by fresh emissions. By contrast, the spatial and temporal variability for 1,3-butadiene is larger. One reason for the higher variability is the right skewness of the 1,3-

[3] Battelle did its analysis three ways. We compared our results with their "replacement" method, which replaced sample values below the LOD with ½ the LOD.

butadiene data, which results in infrequent peaks in the concentration time-series. The possible reasons for these peaks are meteorological conditions that either concentrate the emissions or direct them at a particular monitor on a particular day, or upset conditions at particular facilities. The hypothesis regarding the meteorological variability could be investigated by examining the correlation between meteorological parameters and concentrations.

Figure 3.5. Variability by Monitoring Station for Benzene Concentrations

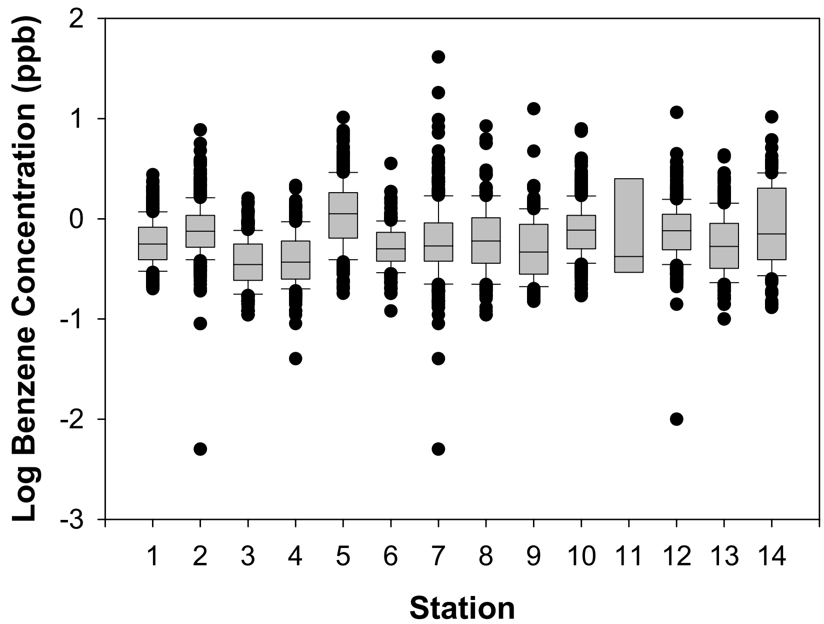
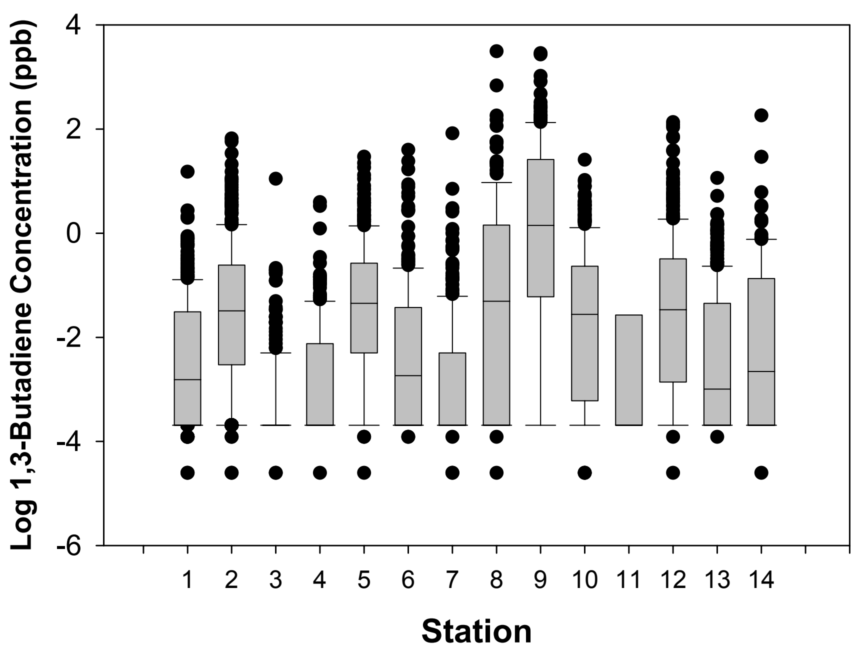


Figure 3.6. Variability by Monitoring Station for 1,3-Butadiene Concentrations



Station IDs: (1) Aldine, (2) Channelview, (3) Northwest Harris Co., (4) Bayland Park, (5) Galena Park, (6) Baytown, (7) Shore Acres, (8) Allendale, (9) Milby Park, (10) Haden Road, (11) Lynchburg Ferry, (12) Clinton, (13) Deer Park, and (14) San Jacinto Monument.

Figure 3.7. Seasonal Variability for Benzene Concentrations

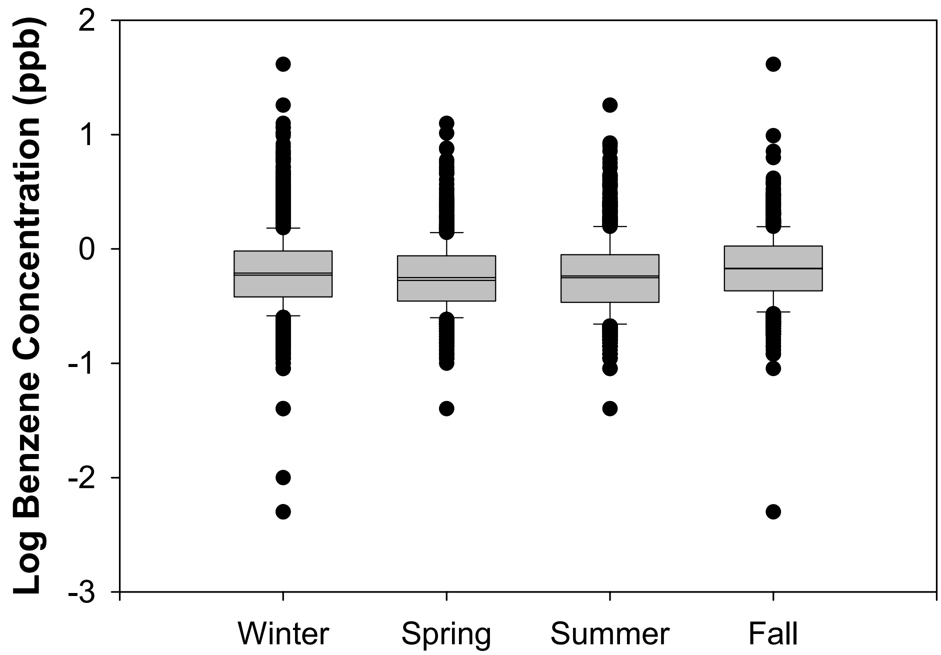


Figure 3.8. Seasonal Variability for 1,3-Butadiene Concentrations

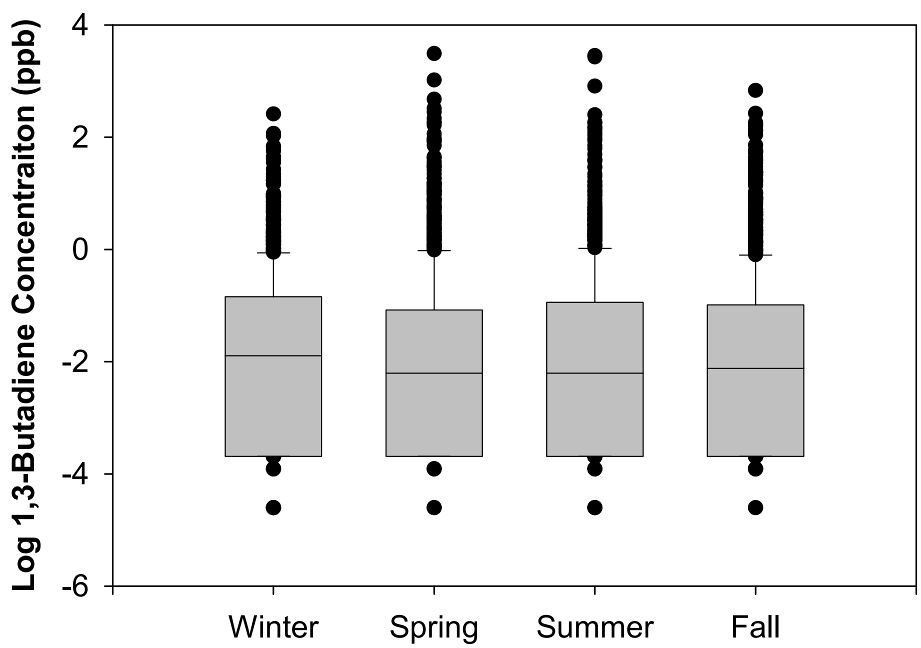


Figure 3.9. Monthly Variability for Benzene Concentrations

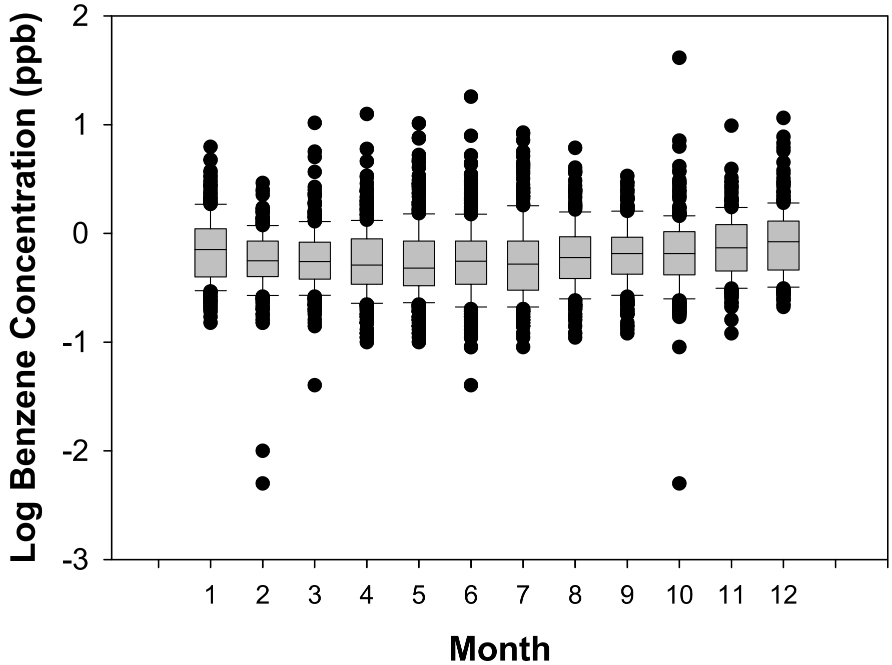
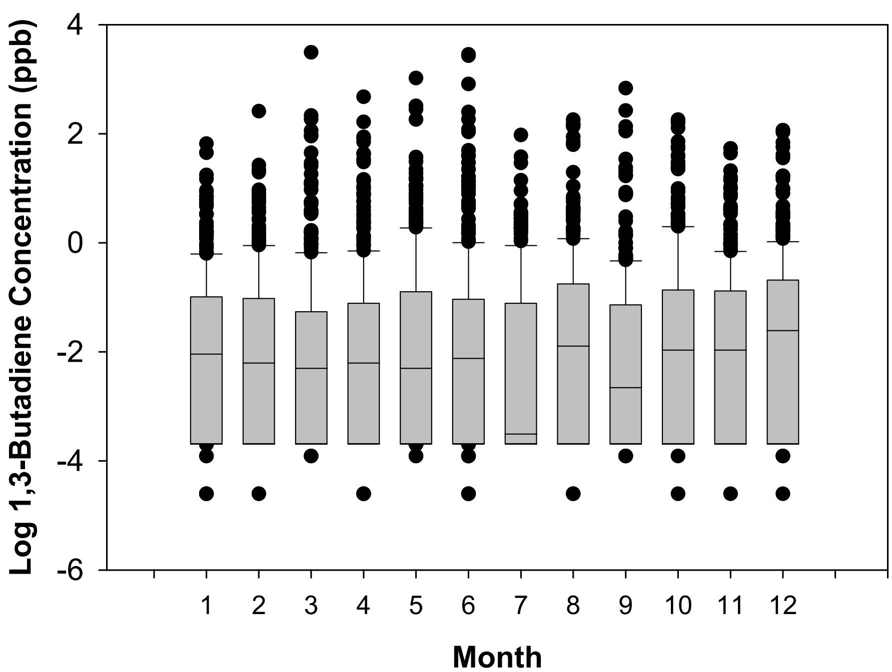


Figure 3.10. Monthly Variability for 1,3-Butadiene Concentrations



4.0 COMPARISON OF MODELING AND MONITORING RESULTS

4.1 Comparison of Modeled and Monitored Results for the Houston API Study

API provided preliminary results of their modeling analysis in the Houston Ship Channel area (Dick Karp, personal communication). This analysis is being conducted by ENSR, and has focused on a small domain in the Ship Channel with a southwest corner at Universal Transverse Mercator (UTM) zone 15, easting of 284km, northing of 3282 km. The modeling is based on 1996 emissions and meteorology data, and the estimates are annual averages. Both the Haden Road and Deer Park monitors are located within the API domain. Tables 4-1 and 4-2 summarize the estimated modeled concentrations at the modeling grids nearest to the Haden Road and Deer Park monitors for API's baseline modeling run. For comparison, the table also includes the average 1996 monitor concentrations, and the average concentrations for all available years at the monitors (1996-2003). The modeling results are specific to 1996, but the use of annual averages for a wider period for comparison purposes is justified given the small trend in concentrations over time.

There are several uncertainties in making these comparisons. First, the monitored concentrations only include measurements every 6 days (approximately), while the model estimates include every day of the year. Also, there may be significant small-scale gradients in concentrations that influence monitor concentrations and that cannot be accounted for in the modeling.

For benzene, the API modeling results agree fairly well with the monitored concentrations, although the model estimates were higher than the monitored concentrations. At Haden Road, the 1996 monitored concentration was 20% below the model estimate, and the annual average concentration was 34% lower than the model estimate. At Deer Park, the 1996 monitored concentration was 36% below the API model estimate, and the annual average concentration was 41% below the monitored concentration.

For 1,3-butadiene, the API model estimates for Haden Road agreed well with the monitored concentrations. The 1996 average monitor concentration was 11% higher than the model estimate, and the annual average monitor concentration was 17% higher than the monitor concentration. For Deer Park, the API model estimates were significantly higher than the monitored concentrations. The 1996 monitored concentration was 85% lower than the API model estimate, and the annual average monitored concentration was 45% lower.

Table 4.1. Summary of API Model Estimates and Monitored Concentrations for Benzene

Monitoring Station	Benzene Concentrations (ppb)		
	API Model Estimate	1996 Average Monitor Concentration[a]	Annual Average Monitor Concentration[a]
Haden Road	1.48	1.18 (-20%)	0.97 (-34%)
Deer Park	1.20	0.77 (-36%)	0.71 (-40%)

[a] Value in parentheses is the percent difference from the API model estimate

Table 4.2. Summary of API Model Estimates and Monitored Concentrations for 1,3-Butadiene

Monitoring Station	1,3-Butadiene Concentrations (ppb)		
	API Model Estimate	1996 Average Monitor Concentration	Annual Average Monitor Concentration
Haden Road	0.35	0.39 (+11%)	0.41 (+17%)
Deer Park	0.33	0.050 (-85%)	0.18 (-45%)

[a] Value in parentheses is the percent difference from the API model estimate

4.2 Comparison of Modeled and Monitored Results for the NATA

In addition to the API refined modeling of a portion of the Ship Channel, the U.S. EPA conducted an air toxics modeling analysis of the entire continental United States for the NATA program. The NATA modeling is done at a census tract level and includes both benzene and 1,3-butadiene. While not as refined as the API modeling, the NATA estimates provide a means to compare modeled versus monitored concentrations for all of the monitoring stations available in Houston. The NATA modeling was done for 1996 (an update for 1999 will be released relatively soon, according to EPA). Therefore, the model estimates were compared to the monitoring results for 1996 only, and for the annual average of all available years since 1996. The model estimate for the census tract that includes the monitor was used. Table 4.3 summarizes the comparison for benzene, and Table 4.4 summarizes the comparison for 1,3-butadiene.

For benzene, the 1996 model to monitor comparisons were fairly good for Aldine (-3.4% difference between modeling estimate and monitor concentration), Allendale (+5.3%), and Deer Park (-13%). The comparisons for Clinton (+23% difference) and Haden Road (-30%) were adequate. The monitored concentration at Channelview was about two-fold higher than the modeling estimate. All of the monitors to model comparisons are within an order of magnitude. For the annual average of the monitoring data, the comparison was within 10% for Aldine, Deer Park, Galena Park, Haden Road, Milby Park, and Northwest Harris County. The difference was more significant for several of the monitors including Baytown (+52%), Channelview (-35%), Clinton (+38%), Lynchburg Ferry (+64%), San Jacinto Monument (-64%), and Shore Acres (-49%). The average percent difference was -7.4%, indicating a small underestimate for the modeling, compared to the modeling data. However, this average was affected by several high differences than canceled out one another. The average absolute percent difference was 25%.

Figure 4-1 provides a scatter plot of the pairs of NATA modeling estimates and annual average monitoring concentrations for each of the monitoring stations. The plots show that there is some correlation among the sites (r^2=0.25). Given the narrow range of concentrations and low levels, it is encouraging that the model is able to account for about a quarter of the spatial variability.

For 1,3-butadiene, the comparisons between the modeling and monitoring data were significantly worse than for benzene. For the 1996 monitoring data, the modeled concentration was higher for 4 of the 5 comparisons. The modeled and monitored concentration agreed fairly well at Deer Park (+6.7% difference). For the annual average monitoring data, the monitoring concentrations were lower than the model estimates for all of the monitors, and the estimates were more than two-fold lower for 12 of the 14 monitors. Clearly, the agreement between the model and monitor concentrations is poor for 1,3-butadiene.

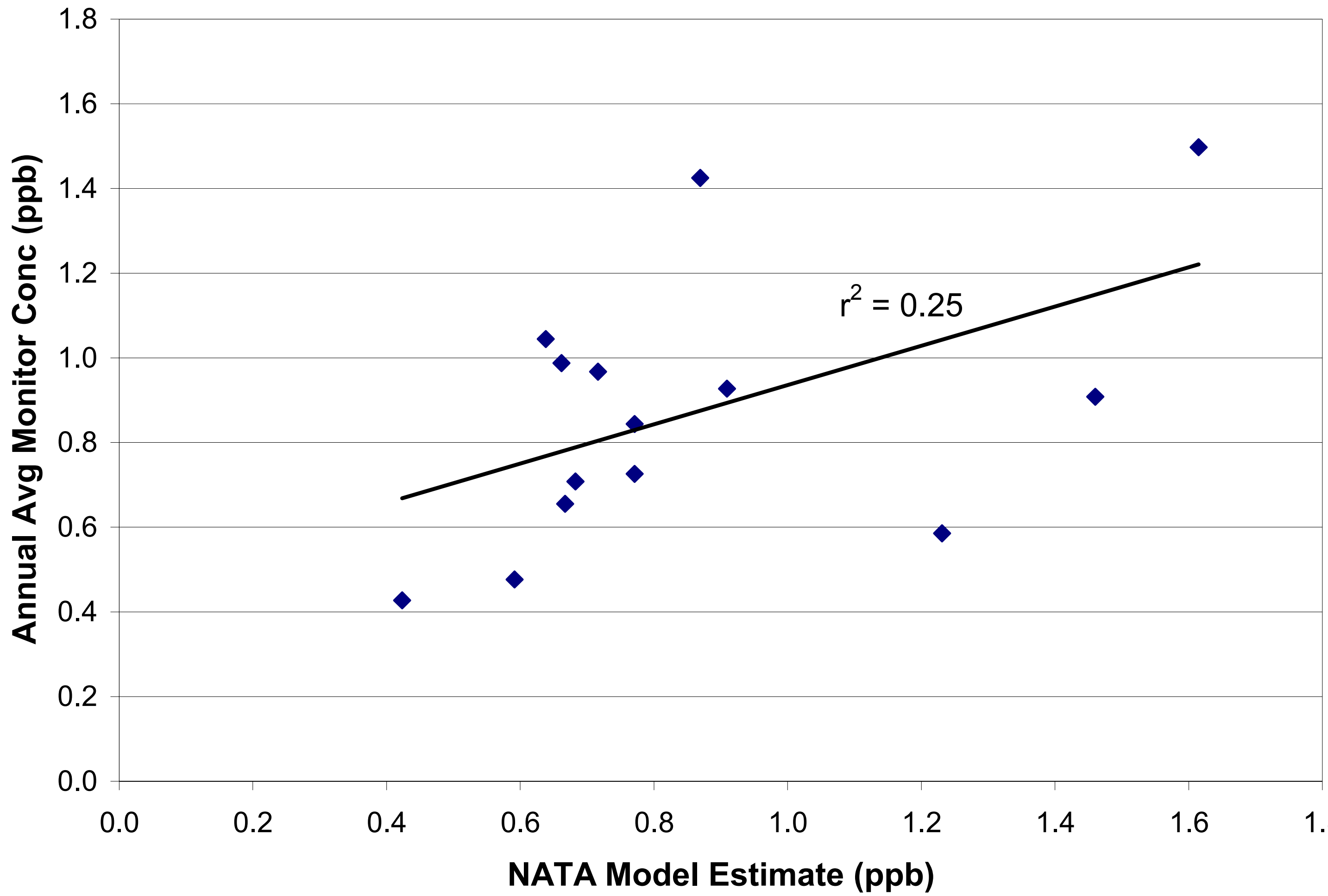
Table 4.3. Summary of NATA Model Estimates and Monitored Concentrations for Benzene

Monitoring Station	NATA Modeling Estimate	1996 Data		Annual Average	
		Average Monitoring Concentration (ppb)	Percent Difference from Modeling Estimate	Average Monitoring Concentration (ppb)	Percent Difference from Modeling Estimate
Aldine	0.67	0.69	-3.4%	0.66	+1.8%
Allendale	0.77	0.73	+5.3%	0.84	-9.4%
Bayland Park	0.59			0.48	+19%
Baytown	1.23			0.59	+52%
Channelview	0.72	1.13	-58%	0.97	-35%
Clinton	1.46	1.12	+23%	0.91	+38%
Deer Park	0.68	0.77	-13%	0.71	-3.7%
Galena Park	1.61			1.50	+7.3%
HRM-3 Haden Road	0.91	1.18	-30%	0.93	-2.0%
Lynchburg Ferry	0.87			1.43	-64%
Milby Park	0.77			0.73	+5.8%
Northwest Harris Co	0.42			0.43	-1.0%
San Jacinto Monument	0.64			1.04	-64%
Shore Acres	0.66			0.99	-49%

Table 4.4. Summary of NATA Model Estimates and Monitored Concentrations for 1,3-Butadiene

Monitoring Station	NATA Modeling Estimate	1996 Data		Annual Average	
		Average Monitoring Concentration (ppb)	Percent Difference from Modeling Estimate	Average Monitoring Concentration (ppb)	Percent Difference from Modeling Estimate
Aldine	0.044	0.10	-126%	0.15	-248%
Allendale	0.35			1.1	-214%
Bayland Park	0.041			0.092	-124%
Baytown	0.069			0.24	-253%
Channelview	0.053	0.33	-527%	0.47	-785%
Clinton	0.44	0.69	-58%	0.54	-24%
Deer Park	0.054	0.050	+6.7%	0.18	-236%
Galena Park	0.11			0.45	-333%
HRM-3 Haden Road	0.059	0.39	-559%	0.41	-600%
Lynchburg Ferry	0.084			0.14	-70%
Milby Park	0.35			3.1	-770%
Northwest Harris Co	0.020			0.043	-112%
San Jacinto Monument	0.059			0.39	-572%
Shore Acres	0.050			0.12	-139%

Figure 4-1. Comparison of NATA Model Estimates and Annual Average Monitoring Concentrations for Benzene



4.3 Comparison of NATA Model Estimates and Monitoring Data Relative to Source Contribution

The NATA model estimates include the contribution of the modeled concentration from the primary source categories, including major, area, on-road mobile, off-road mobile, and background. In the NATA, the background contribution is understood to mean the amount of the chemical recirculating in the atmosphere from the previous day, as opposed to only a natural background. Tables 4.5 and 4.6 summarize the source contributions of benzene and 1,3-butadiene for each of the monitoring sites.

For benzene, the sites with the largest contribution from major sources are Baytown (61%), Clinton (52%), Galena Park (64%), Haden Road (51%), and Lynchberg Ferry (50%). The sites with the largest contribution from on-road mobile sources are Aldine (45%), Bayland Park (53%), and Northwest Harris County (56%). These are the three sites that are generally not in the easterly direction of Houston. It makes sense that the sites that are nearest to the industrial sources in Houston or in the Ship Channel have the highest contributions from major sources. For 1,3-butadiene, the sites with the largest contribution from major sources are Allendale (89%), Clinton (87%), and Milby Park (89%). These sites are generally the closest to Houston, in the easterly direction. The sites with the largest contribution from on-road mobile sources are Bayland Park (71%), Northwest Harris County (80%), and Shore Acres (61%). Bayland Park and Northwest Harris County are west of Houston and the Ship Channel. Shore Acres is fairly south of Houston, in the southeast direction.

One interesting analysis is to compare the annual trend results in the previous section with the contribution from the different source types to see if these trends can be attributed to a particular source category. Figures 4.2 through 4.5 provide plots of the annual average trend for benzene and 1,3-butadiene versus the percent contribution from major and on-road mobile sources. There is no evident pattern in any of the graphs. Therefore, the observed downtrend trends in benzene and 1,3-butadiene concentrations cannot be attributed to a given source type from this analysis.

Table 4.5. Contribution of Different Source Types to Benzene Concentrations Estimated in the 1996 NATA at the Houston Monitor Locations

Station Name	Census Tract ID	Contribution of Different Sources				
		Major (%)	Area and Other (%)	Onroad Mobile (%)	Nonroad Mobile (%)	Background (%)
Aldine	022302	19%	7%	45%	10%	18%
Allendale	032003	30%	7%	37%	11%	16%
Bayland Park	042503	9%	5%	53%	13%	21%
Baytown	026500	61%	7%	17%	5%	10%
Channelview	036001	36%	8%	30%	8%	17%
Clinton	021002	52%	6%	27%	6%	8%
Deer Park	036004	28%	8%	36%	10%	18%
Galena Park	021100	64%	6%	17%	5%	8%
HRM-3 Haden Road	023300	51%	9%	20%	7%	14%
Lynchburg Ferry	026300	50%	8%	22%	6%	14%
Milby Park	032003	30%	7%	37%	11%	16%
Northwest Harris Co	055200	1%	3%	56%	12%	28%
San Jacinto Monument	036200	23%	10%	20%	32%	15%
Shore Acres	036602	24%	9%	39%	9%	19%

Table 4.6. Contribution of Different Source Types to 1,3-Butadiene Concentrations Estimated in the 1996 NATA at the Houston Monitor Locations

Station Name	Census Tract ID	Contribution of Different Sources				
		Major (%)	Area and Other (%)	Onroad Mobile (%)	Nonroad Mobile (%)	Background (%)
Aldine	022302	32%	4%	51%	13%	0%
Allendale	032003	89%	1%	7%	3%	0%
Bayland Park	042503	10%	2%	71%	16%	0%
Baytown	026500	52%	14%	28%	6%	0%
Channelview	036001	57%	4%	31%	8%	0%
Clinton	021002	87%	2%	9%	2%	0%
Deer Park	036004	42%	3%	43%	11%	0%
Galena Park	021100	64%	9%	22%	6%	0%
HRM-3 Haden Road	023300	67%	6%	20%	6%	0%
Lynchburg Ferry	026300	71%	7%	17%	4%	0%
Milby Park	032003	89%	1%	7%	3%	0%
Northwest Harris Co	055200	3%	3%	80%	14%	0%
San Jacinto Monument	036200	42%	4%	15%	39%	0%
Shore Acres	036602	19%	9%	61%	11%	0%

Figure 4-2. Annual Trend Percent versus the Percent of Major Source Contribution to Monitoring Site's Concentrations in NATA for Benzene

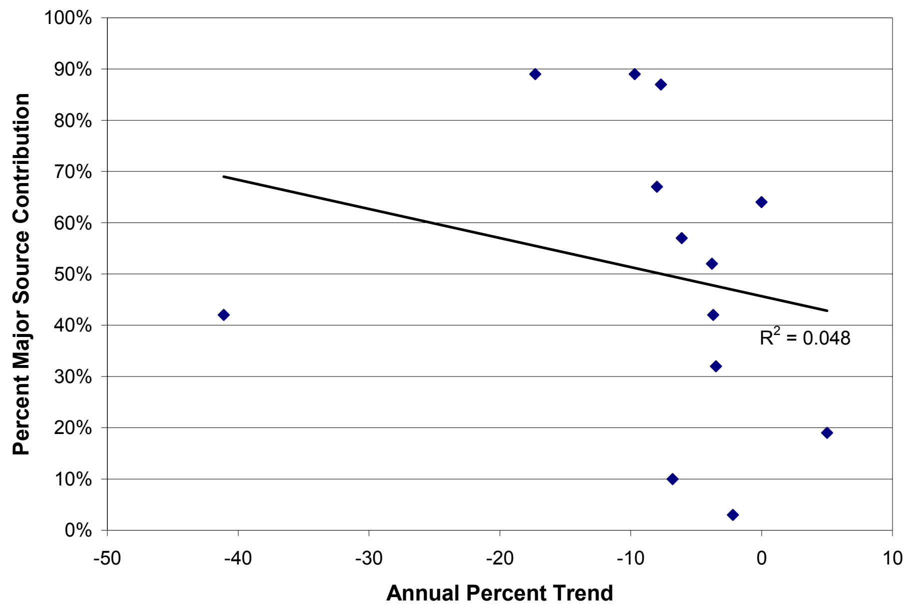


Figure 4-3. Annual Trend Percent versus the Percent of On-Road Mobile Source Contribution to Monitoring Site's Concentrations in NATA for Benzene

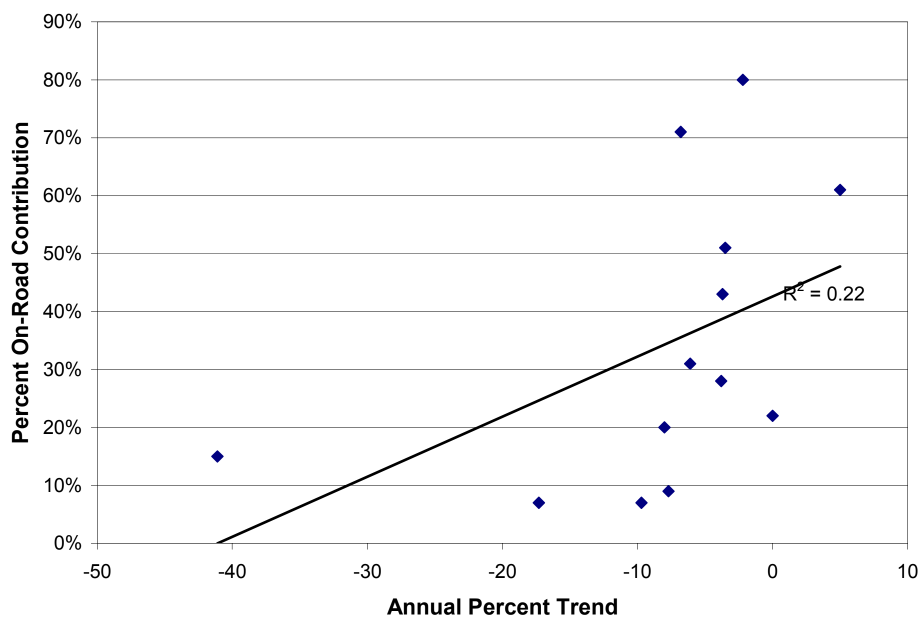


Figure 4-4. Annual Trend Percent versus the Percent of Major Source Contribution to Monitoring Site's Concentrations in NATA for 1,3-Butadiene

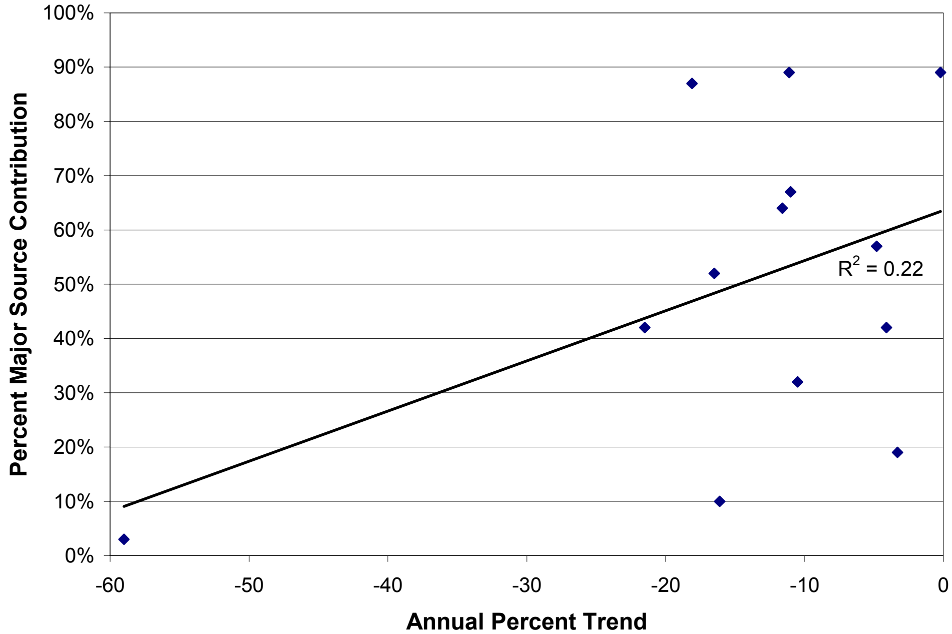
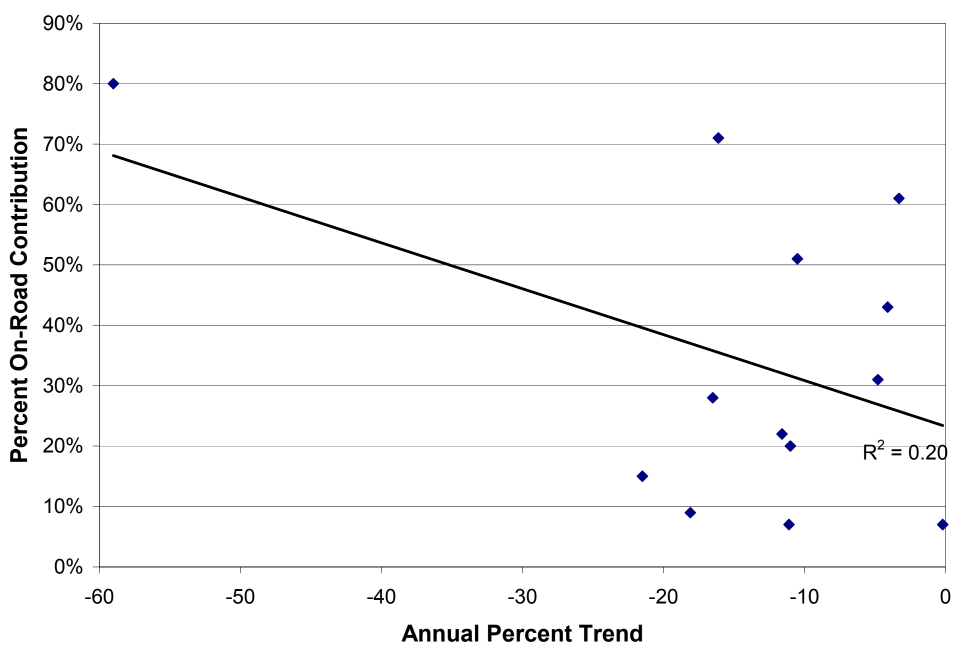


Figure 4-5. Annual Trend Percent versus the Percent of On-Road Mobile Source Contribution to Monitoring Site's Concentrations in NATA for 1,3-Butadiene



REFERENCES

Battelle Corporation (2003) “Air toxics monitoring data: analyses and network design recommendations,” Report prepared for LADCO.

Karp, D. (2004) API/ENSR modeling analysis provided as personal communication.

Pinheiro, J.C and Bates, D.M. (2000) Mixed-effects models in S and S-Plus, Springer.