June 13, 2003

FINAL TECHNICAL REPORT

on

ESTIMATE BACKGROUND CONCENTRATIONS FOR THE NATIONAL-SCALE AIR TOXICS ASSESSMENT

Emissions, Monitoring, and Analysis Division
Office of Air Quality Planning and Standards
U.S. ENVIRONMENTAL PROTECTION AGENCY
Research Triangle Park, North Carolina 27711

DISCLAIMER

This report is a work prepared for the United States Government under contract. In no event shall either the United States Government nor the contractor have any responsibility or liability for any consequences of any use, misuse, inability to use, or reliance on the information contained herein, nor does either warrant or otherwise represent in any way the accuracy, adequacy, efficacy, or applicability of the contents hereof.

TABLE OF CONTENTS

List of Tables

List of Figures

ESTIMATE BACKGROUND CONCENTRATIONS FOR THE NATIONAL-SCALE AIR TOXICS ASSESSMENT

1.0 INTRODUCTION

In the 1990 Clean Air Act Amendments (CAAA), Congress established a list of 188 toxic chemicals designated as hazardous air pollutants (HAPs), also known as air toxics. These HAPs have been associated with a wide variety of adverse health effects and adverse environmental effects. The sources of the air toxics include major point sources, area sources, mobile sources, and transport from other areas. The resulting population exposures are typically a mixture from a multitude of sources emitting many different HAPs. As a part of the implementation of the CAAA, EPA's Office of Air Quality Planning and Standards (OAQPS) currently manages the multi-component Air Toxics Program (ATP). One of the four key activities within the ATP is the National-Scale Air Toxics Assessment (NATA). NATA seeks to quantify the impacts of air toxics emissions on public health and the environment, identify the areas of concern, and eventually track the progress of air toxics reduction strategies. NATA includes several nationwide activities: expanding ambient air toxics monitoring, improving and updating emissions inventories, multi-scale air quality modeling, exposure modeling, and risk assessment.

More specifically, NATA includes a nationwide study of the potential inhalation exposures and health risks associated with 32 HAPs and diesel particulate matter (i.e., diesel PM). As the second of four major steps in the assessment, ambient concentrations are estimated using the Assessment System for Population Exposure Nationwide (ASPEN) air dispersion model. In order to estimate ***total*** ambient concentrations, however, a value for background must be estimated and added to ASPEN's modeled concentrations. As defined, background accounts for natural sources, nearby sources (farther than 50 km), and unidentified sources.

In the recently completed NATA, corresponding to calendar year 1996 emissions and meteorological data, estimates for background were determined from open literature. There are two limitations to this approach. First, estimates based on literature (or data) corresponding to points in time that were far removed from the 1996 temporal scope of the recent NATA study may not have accurately represented backgrounds that had evolved over time. Second, the literature was used to identify a constant, nationwide background estimate for each HAP under study. It is currently believed that such an approach may not be realistic for those HAPs that are expected to exhibit spatially heterogeneous backgrounds across the U.S.

In future assessments, new and improved values for background must be determined. The purpose of this project is to develop a method to improve the estimation of background concentrations for future rounds of NATA. The method, which is to be based on monitoring data, should satisfy the following two key objectives:

1. Use data more consistent with the temporal scope of a given NATA round (e.g., 1999); and

2. Allow for the possibility of spatially heterogeneous backgrounds, as appropriate.

This report summarizes Battelle's progress (approach and results) to date toward developing such a method. Section 2 summarizes the data used for the effort, and describes some of the restrictions applied in order to satisfy objective (1). Section 3 summarizes the approach, which seeks to satisfy objective (2), provides illustrative examples for clarity, and discusses the technical issues that have been encountered and how they have been addressed. Section 4 summarizes the suite of results obtained to date. Section 5 highlights remaining issues and discusses the expected path forward. Appendix A provides the extensive, detailed pollutant by countywide background results from Stage 1 of this project's two stage approach. Appendix B presents the conversion factors applied to convert all monitoring data from their original units to units of $\mu g/m^3$. Appendix C summarizes the comments and responses associated with an internal review of this project by senior Battelle statisticians, as requested by EPA/OAQPS. Finally, Appendix D lists the comments received from EPA/OAQPS on the August 15, 2002, draft report, and summarizes the approach to addressing each comment.

2.0 DATA

The scope of data analyses for this project is limited to air toxics monitoring data and other explanatory information that might be necessary for use in developing statistical models. The following subsections provide further detail on the data included or considered for the project.

2.1 Monitoring Data

The suite of air toxics measurements being considered for this project is restricted to the list of 33 urban HAPs, i.e., 32 HAPs and diesel PM. (Refer to Table 1 of "Air Toxics Monitoring Concept Paper," for the complete list; it is available at http://www.epa.gov/ttn/amtic/airtxfil.html). The focus of this project on these HAPs is consistent with the most recent round of NATA. Subsection 2.1.1 provides an overview of the monitoring data pursued for this project, and the logic for including or excluding various data sources. Subsection 2.1.2 summarizes the final combined database used for analysis.

2.1.1 Overview

In order to satisfy objective (1) above, recent air toxics monitoring data were sought. It is important to recognize that the specification of "recent" is subjective. Ideally, to estimate background concentrations for the 1999 round of NATA, for example, one would restrict available monitoring data to samples generated during calendar year 1999. However, there exists a trade-off between the level of restriction (i.e., quality) applied when accepting or rejecting data for analysis versus the breadth and representativeness of results that might be obtained. To balance these two competing needs, the decision was made to expand the range of acceptable monitoring data to samples generated between 1995 and 2002, inclusive. By

widening the window of data acceptance to 1995 through 2002, results may be obtained for a greater number of observations, monitoring locations, U.S. counties, etc. The implicit assumption associated with this decision is that background concentration levels throughout the country were, at least approximately, stable and consistent with 1999 levels during the years just prior to and subsequent to 1999. Refer to subsection 3.1.6 of this report for a case study example that investigates this assumption further.

Given the window of data acceptance just defined, three major sources of air toxics data were sought:

- *Air Toxics 10-City Pilot Study Data (pilot data)*. For background and further details, see "Final Report — Pilot City Air Toxics Measurements Summary" or "Quarterly Air Toxics Monitoring Newsletter, Jan. 2001." Both are available at http://www.epa.gov/ttn/amtic/airtxfil.html .

- *EPA's Air Toxics Data Archive (archive data)*. For background and further details, see Rosenbaum, A.S., Stiefer, P.S., and Iwamiya, R.K. (1999), "Air Toxics Data Archive and AIRS Combined Data Set: Database Descriptions," Technical Report to U.S. EPA, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina, September.

- *PAMS Air Toxics Data (PAMS data)*. For background and further details, see http://www.epa.gov/oar/oaqps/pams/ .

The inclusion of pilot data in this project was considered important because of their temporal relevance (collected in 2001 and 2002), expected high level of quality, and close association with future air toxics data to be generated as part of the developing national air toxics monitoring network. The archive data, upon restriction to 1995 through 2002, quite simply provide the most comprehensive set of available, relevant data for this project and, hence, are critical to its success. Finally, PAMS data provide yet another important source of air toxics monitoring data.

In terms of obtaining the above-described data for this project, first note that pilot data have been obtained, assembled, managed, and quality assured/quality controlled (QA/QC'd) by Battelle under contract with the Lake Michigan Air Directors Consortium (LADCO). As of the writing of this report, however, note that the quality and completeness of the pilot data are still improving as additional data are received and reviewed. Next, observe that the most recent, and possibly last, version of archive data (compiled January 2002) was provided to Battelle for this project by Systems Applications International (SAI), a wholly-owned subsidiary of ICF Consulting. SAI is the EPA contractor historically responsible for generating the database of archive data. Finally, after some initial limited efforts to obtain a comprehensive set of PAMS data for this project, discussions with SAI revealed that all such data, so long as they are classified as air toxics, are included by default within the archive as part of the algorithm to generate that database. Thus, inclusion of archive data for this project, by default, leads to the inclusion of air toxics PAMS data as well.

Observe that these three data sets are known to have different data collection procedures, analytical instruments, and data quality associated with them. The archive itself includes an assortment of air toxics data from numerous monitoring programs and past research studies. To the degree that data quality varies and biases or other differences exist between the various methodologies included in the combined database, uncertainties may be introduced into the final results of the project. However, any attempt to significantly restrict the combined data to a smaller core of more consistent, and possibly higher quality, measurements could seriously limit the extent or expansiveness of the project's Stage 1 results (see Section 3.1 below). The ultimate consequence of such a decision would likely be the inability to adequately address or complete Stage 2 (see Section 3.2 below).

In particular, different data sets may exhibit varying degrees of uncertainty due to various laboratory sensitivities and specificities. Different technologies for measuring the same pollutant, for example, may lead to different method detection limits (MDLs). Furthermore, independent of the analytical technology, different laboratories may employ different protocols for calculating uncertainty in the form of an MDL. Likewise, due to a range of semantics associated with laboratory uncertainties and measurement reporting thresholds, the specific definition or interpretation of an MDL may vary by laboratory (or data set). All of these issues represent some form of uncertainty introduced into the combined data set. For further discussion and an attempt to clarify some of these issues, refer to "Final Report — Pilot City Air Toxics Measurements Summary" (http://www.epa.gov/ttn/amtic/airtxfil.html).

In addition to the above data sets, several other sources of air toxics data were investigated for inclusion in the project. A summary of these data sources, and some of their potential limitations, is provided as follows:

- The California Air Resources Board (CARB) provided air toxics data corresponding to the city of San Diego. To date, these data have not been incorporated for analyses. Note, however, that even without these data, a significant amount of Stage 1 background results have been obtained for the areas of San Diego and Southern California (see Section 4.1 and Appendix A).

- Sonoma Technologies, Inc., (STI) provided data from Arizona. These data will not be included because it turns out they do not satisfy the 1995 through 2001 temporal requirement discussed above.

- Some preliminary discussion took place regarding the inclusion of speciated metals data from the $PM_{2.5}$ Chemical Speciation network. These data could provide useful information to better understand urban background levels in the case of metals (fine). However, they would not provide direct background information for metals (coarse). To date, these data have not been actively pursued because a significant amount of metals (fine) data is already included in the archive and pilot databases (see discussion on metals data in subsection 4.1.2 below).

- Similar to the above comments on $PM_{2.5}$ Chemical Speciation network data, Interagency Monitoring of Protected Visual Environments (IMPROVE) data might eventually be

considered to better understand rural background levels in the case of metals (fine). Limited primarily to $PM_{2.5}$ measurements, however, these data would not provide direct background information for metals (coarse).

- As with IMPROVE data, Clean Air Status and Trends Network (CASTNet) visibility data might be pursued to further expand on the results for metals (fine). Subject to similar limitations as IMPROVE, CASTNet data have the further limitation of not possessing MDL information.

- Using the secondary analysis approach described below in subsection 3.1.7, National Dioxin Air Monitoring Network (NDAMN) data could be incorporated to provide results for 2,3,7,8-tetrachlorodibenzo-p-dioxin. To date, no background results have been obtained for this urban priority HAP under this project. The NDAMN network consists of approximately 30 well-dispersed nationwide monitoring sites operating since 1999.

- The Mercury Deposition Network (MDN) data could be incorporated to provide results for mercury. The MDN data does not distinguish between fine and coarse mercury and, as with CASTNet, does not contain MDL information. In addition, the MDN collects data on mercury deposition (e.g., by rain) rather than mercury concentrations in the air. The MDN network consists of approximately 70 monitoring sites located primarily in the eastern Unites States.

- Inquiries were made within Battelle to identify air toxics monitoring data from past Battelle studies. The most significant source of data identified was from a 1993 air toxics field study conducted in Columbus, Ohio, and Atlanta, Georgia. Some of these data are already included within the air toxics data archive. However, these data fall outside the temporal requirement discussed above and, therefore, are not included in the results of Section 4.1 and Appendix A.

To date, the majority of the above data sources have not been actively pursued or incorporated into the database for analyses in this project. The primary reason for not pursuing these data is current project time and resource constraints. Their inclusion for obtaining updated or final results should be re-considered at a later date.

2.1.2 Summary

Tables 2.1a and 2.1b together summarize the combined database of archive, PAMS, and pilot monitoring data as described above in subsection 2.1.1. These data are used to conduct the Stage 1 analyses described below in Section 3.1 (see Section 4.1 for results). Table 2.1a corresponds to the subset of data for which the primary quantitative statistical methodology of Stage 1 could be applied (see subsections 3.1.1 through 3.1.6). Table 2.1b corresponds to the subset of data for which a less desirable and less quantitative statistical approach was required in Stage 1 due to small sample sizes or too much data reported below the MDL (see subsection 3.1.7). These tables are presented separately for perspective on the relative amount of data analyzed by either statistical approach.

Tables 2.1a and 2.1b provide summary information for each pollutant considered. The "Number of Observations," "Number of Sites," and "Number of Counties" columns provide an indication of the amount of available data for pursuing the estimation of a given pollutant's background. With respect to each pollutant's available data, the "Date Range of Data" column provides a rough sense of its temporal relevance to 1999. The remaining columns summarize the distribution of concentrations observed in the combined database. Under the general heading of "Method Detection Limit (MDL)," the minimum (min) and maximum (max) columns indicate the range of MDLs associated with a given HAP. The "% of Data Below" column indicates the percentage of a pollutant's data that fall below its respective MDL. Under the general heading of "Above-MDL Distribution of Concentrations ($\mu g/m^3$)", the remaining columns of the tables summarize each pollutant's observed distribution of concentrations for those data that lie above their respective MDLs. When interpreting the Stage 1 results presented in Table 4.1 of subsection 4.1.1, the summary provided in these tables may be used for additional perspective.

Table 2.1a Summary of combined archive, PAMS, and pilot monitoring data for conducting primary Stage 1 analyses. [1]

Class	Pollutant	Number of Observations	Number of Sites	Number of Counties	Date Range of Data	Method Detection Limit (MDL)			Above-MDL Concentration Distribution (µg/m³)				
						Min	Max	% of Data Below	Mean	Standard Deviation	5th Percentile	Median	95th Percentile
Carbonyl	ACETALDEHYDE	21721	204	94	01/02/95 - 06/25/02	0.0005	0.5405	2.27	2.3890	2.5015	0.5044	1.6214	6.8461
	ACROLEIN	1071	13	4	04/03/95 - 06/25/02	0.0030	0.1398	5.70	0.2698	0.3289	0.0080	0.1799	0.8709
	FORMALDEHYDE	25259	234	106	01/02/95 - 06/25/02	0.0026	0.4813	1.20	4.1006	5.4634	0.6215	2.8073	10.1940
VOC	1,1,2,2-TETRACHLOROETHANE	663	3	1	01/10/95 - 12/31/98	0.0090	0.0090	0.00	0.0591	0.1066	0.0309	0.0309	0.2059
	1,2-DIBROMOETHANE	784	4	2	01/03/95 - 12/31/98	0.0100	0.2535	11.22	1.2553	4.7749	0.0384	0.0384	9.7578
	1,2-DICHLOROETHANE	2839	16	9	01/03/95 - 03/26/01	0.0100	0.1619	13.00	0.5802	2.6976	0.0202	0.0809	1.9427
	1,2-DICHLOROPROPANE	680	5	4	01/03/95 - 12/26/99	0.0200	0.3697	29.71	1.1383	4.5584	0.0462	0.0462	4.0204
	1,3-BUTADIENE	19702	115	52	01/03/95 - 06/25/02	0.0177	0.3000	33.04	0.8424	3.1305	0.0664	0.3318	2.1459
	ACRYLONITRILE	1267	6	4	01/03/95 - 12/27/01	0.1302	0.2386	81.93	1.1271	0.8596	0.3688	0.8200	2.9700
	BENZENE	61069	410	154	01/01/95 - 06/25/02	0.0053	0.6400	2.64	1.8314	2.5877	0.2810	1.2118	5.2130
	CARBON TETRACHLORIDE	10930	99	46	01/03/95 - 06/25/02	0.0100	0.6291	2.65	0.7768	1.2490	0.1887	0.6417	1.1324
	CHLOROFORM	9321	71	38	01/03/95 - 05/26/02	0.0070	0.1000	13.76	0.2297	0.4140	0.0488	0.1465	0.5859
	CIS 1,3-DICHLOROPROPENE	NO AVAILABLE DATA											
	TRANS 1,3-DICHLOROPROPENE												
	ETHYLENE OXIDE	268	1	1	01/01/01 - 06/25/02	0.1189	0.1351	36.19	0.2576	0.1289	0.1369	0.2125	0.5205
	METHYLENE CHLORIDE	23162	171	85	01/03/95 - 06/25/02	0.0100	0.5000	19.92	1.8175	10.5251	0.1389	0.6947	4.5156
	TETRACHLOROETHYLENE	12400	100	43	01/03/95 - 06/25/02	0.0100	0.4748	14.40	0.7887	1.3421	0.0678	0.4069	2.7129
	TRICHLOROETHYLENE TCE	11115	78	36	01/03/95 - 06/25/02	0.0100	0.5374	24.38	0.5664	2.8505	0.0537	0.2149	1.6014
	VINYL CHLORIDE	1211	4	2	01/03/95 - 12/20/99	0.0200	0.0511	18.25	0.1279	0.1561	0.0256	0.0665	0.4499
Metal (fine)	ARSENIC	45	1	1	01/01/01 - 02/25/02	0.0000	0.0000	2.22	0.0018	0.0011	0.0002	0.0016	0.0033
	BERYLLIUM	NO AVAILABLE DATA											
	CADMIUM	53	2	1	01/07/01 - 02/25/02	0.0000	0.0000	0.00	0.0013	0.0009	0.0001	0.0011	0.0030
	CHROMIUM	575	3	2	01/03/95 - 02/19/02	0.0005	0.0010	50.78	0.0045	0.0050	0.0008	0.0030	0.0150
	LEAD	2422	14	9	01/03/95 - 02/25/02	0.0004	0.0020	31.63	0.0081	0.0090	0.0030	0.0050	0.0230
	MANGANESE	4303	21	15	01/03/95 - 02/25/02	0.0003	0.0012	35.74	0.0055	0.0076	0.0020	0.0030	0.0190
	MERCURY	NO AVAILABLE DATA											
	NICKEL	612	4	3	01/03/95 - 02/25/02	0.0002	0.0010	38.73	0.0040	0.0040	0.0007	0.0030	0.0120
Metal (coarse)	ARSENIC	NO AVAILABLE DATA											
	BERYLLIUM	NO AVAILABLE DATA											
	CADMIUM												
	CHROMIUM	3986	16	11	01/03/95 - 12/29/00	0.0010	0.0010	49.10	0.0035	0.0031	0.0020	0.0020	0.0090
	LEAD	3446	11	8	01/03/95 - 12/29/00	0.0020	0.0020	59.90	0.0044	0.0024	0.0030	0.0040	0.0090
	MANGANESE	3661	16	10	01/03/95 - 12/29/00	0.0010	0.0010	5.65	0.0135	0.0112	0.0020	0.0110	0.0360
	MERCURY	NO AVAILABLE DATA											
	NICKEL	2605	8	7	01/03/95 - 12/29/00	0.0010	0.0010	61.57	0.0031	0.0021	0.0020	0.0020	0.0070
SVOC [1]	HEXACHLOROBENZENE	NO AVAILABLE DATA											

1 No data were available for all other SVOC pollutants.

Table 2.1b Summary of combined archive, PAMS, and pilot monitoring data for conducting secondary Stage 1 analyses. [1]

Class	Pollutant	Number of Observations	Number of Sites	Number of Counties	Date Range of Data	Method Detection Limit (MDL)			Above-MDL Concentration Distribution (µg/m³)				
						Min	Max	% of Data Below	Mean	Standard Deviation	5th Percentile	Median	95th Percentile
Carbonyl	ACETALDEHYDE	1082	26	22	08/25/95 - 06/25/02	0.0007	1.2984	7.6710	1.5644	1.3612	0.3009	1.1476	3.9635
	ACROLEIN	2562	32	15	01/03/95 - 05/26/02	0.0070	1.3757	87.8610	0.3300	0.2910	0.0080	0.3161	0.7197
	FORMALDEHYDE	479	10	7	08/25/95 - 06/25/02	0.0034	1.2989	32.9854	2.7200	1.8860	0.2579	2.3694	6.0182
VOC	1,1,2,2-TETRACHLOROETHANE	9282	111	58	01/03/95 - 05/26/02	0.0200	13.7244	86.8132	0.3566	0.2402	0.1442	0.3432	0.5492
	1,2-DIBROMOETHANE	14889	177	96	01/03/95 - 05/26/02	0.0200	15.3606	75.8278	0.1656	0.1968	0.0768	0.0768	0.3842
	1,2-DICHLOROETHANE	28919	243	111	01/03/95 - 05/26/02	0.0100	8.0916	77.0324	0.1788	0.4925	0.0202	0.0809	0.4047
	1,2-DICHLOROPROPANE	25705	232	106	01/03/95 - 05/26/02	0.0100	9.2388	88.4925	0.1516	0.1377	0.0231	0.1386	0.2311
	1,3-BUTADIENE	10758	145	72	01/03/95 - 05/26/02	0.0100	10.0000	68.5722	0.4238	1.0971	0.0111	0.2654	1.1061
	ACRYLONITRILE	3218	48	32	01/03/95 - 05/26/02	0.0390	2.1693	85.7365	0.6432	0.6031	0.0824	0.4555	1.9965
	BENZENE	4360	69	44	01/03/95 - 04/26/02	0.0100	6.3868	33.0275	2.7325	15.9370	0.1917	0.9580	5.4308
	CARBON TETRACHLORIDE	21268	175	74	01/03/95 - 06/25/02	0.0200	12.5773	65.9112	0.6924	0.3069	0.3146	0.6917	1.0695
	CHLOROFORM	28793	252	113	01/03/95 - 06/25/02	0.0100	9.7612	78.9810	0.2607	1.0279	0.0488	0.1025	0.5859
	CIS 1,3-DICHLOROPROPENE	1619	28	13	01/01/01 - 05/26/02	0.0544	9.0736	99.8147	0.2420	0.1048	0.1815	0.1815	0.3629
	TRANS 1,3-DICHLOROPROPENE	1602	28	13	01/01/01 - 05/26/02	0.0590	9.0736	99.9376	3.1758	-	3.1758	3.1758	3.1758
	ETHYLENE OXIDE	343	5	2	01/01/01 - 06/25/02	0.1189	0.1351	35.5685	0.2658	0.1693	0.1315	0.2089	0.6051
	METHYLENE CHLORIDE	17241	171	87	01/03/95 - 06/25/02	0.0100	6.9444	65.1296	2.2236	14.3395	0.1355	0.7989	6.0069
	TETRACHLOROETHYLENE	24377	226	110	01/03/95 - 06/25/02	0.0100	13.5593	81.6343	1.0574	1.4143	0.0882	0.6782	3.1864
	TRICHLOROETHYLENE TCE	21332	195	91	01/03/95 - 06/13/02	0.0100	10.7433	82.6739	0.2993	0.3550	0.1021	0.2149	0.8061
	VINYL CHLORIDE	32065	278	122	01/03/95 - 05/26/02	0.0100	5.1104	81.6560	0.2442	0.7501	0.0128	0.1022	0.3834
Metal (fine)	ARSENIC	6522	49	19	01/03/95 - 02/25/02	0.0000	4.0000	96.0902	0.0029	0.0010	0.0007	0.0030	0.0040
	BERYLLIUM	NO AVAILABLE DATA											
	CADMIUM												
	CHROMIUM	36706	124	84	01/03/95 - 02/25/02	0.0005	2.0000	99.0737	0.0026	0.0013	0.0020	0.0020	0.0050
	LEAD	34867	113	80	01/03/95 - 12/29/00	0.0020	3.0000	94.0832	0.0057	0.0038	0.0030	0.0050	0.0120
	MANGANESE	32977	106	78	01/03/95 - 12/29/00	0.0010	2.0000	96.2550	0.0032	0.0032	0.0020	0.0030	0.0060
	MERCURY	5984	26	15	01/03/95 - 12/29/00	0.0020	0.0020	99.8663	0.0030	-	0.0030	0.0030	0.0030
	NICKEL	36675	123	84	01/03/95 - 12/29/00	0.0010	2.0000	98.7566	0.0030	0.0013	0.0020	0.0030	0.0055
Metal (coarse)	ARSENIC	5984	26	15	01/09/95 - 12/29/00	0.0020	0.0020	99.8830	0.0037	0.0015	0.0030	0.0030	0.0070
	BERYLLIUM	NO AVAILABLE DATA											
	CADMIUM												
	CHROMIUM	1998	10	6	01/03/95 - 12/29/00	0.0010	0.0010	51.3514	0.0030	0.0012	0.0020	0.0030	0.0050
	LEAD	2538	15	10	01/03/95 - 12/29/00	0.0020	0.0020	65.8786	0.0052	0.0030	0.0030	0.0040	0.0120
	MANGANESE	2323	10	7	01/03/95 - 12/29/00	0.0010	0.0010	3.7882	0.0138	0.0103	0.0030	0.0110	0.0340
	MERCURY	NO AVAILABLE DATA											
	NICKEL	3379	18	12	01/03/95 - 12/29/00	0.0010	0.0010	80.7635	0.0023	0.0006	0.0020	0.0020	0.0030
SVOC [1]	HEXACHLOROBENZENE	255	7	2	04/19/01 - 12/27/01	0.0476	0.0951	100.0000	-	-	-	-	-

1 No data were available for all other SVOC pollutants.

2.2 Other Data

As part of the project's effort to predict a background estimate for each county in the United States, many different potential explanatory variables were examined from several data sets that were provided by EPA's OAQPS. Each of these variables was plotted against the background estimates for each county that had a stage one background estimate. A simple regression was also run on each of the variables. The plots and the results of the regressions were used to determine variables that may be useful in predicting spatially varying background estimates nationwide. Table 2.2 is a list and brief description of all of the variables that were considered. The variables that were initially chosen for modeling background are shown in italics. The variable pop_density, which is shown in bold, was ultimately chosen to replace the initial four predictor variables as the sole independent variable in the regression as described in Section 3.2.

Table 2.2 Variables considered for analysis (county-level aggregation).

Variable	Description
total	Total emissions (tons per year)
lat_ctr	County centroid latitude (degrees N)
lon_ctr	County centroid longitude (degrees W)
county_area	County area (km^2)
population	County population
pop_density	**Population density (people km^2)**
housing	County housing units
forests	County forest area (km^2)
forest_frac	Fraction of county covered by forests
ag_land	County agricultural land area (km^2)
ag_frac	*Fraction of county covered by agricultural land*
urban	County urban land area (km^2)
urban_frac	*Fraction of county covered by urban land*
rural	County rural land area (km^2)
rural_frac	Fraction of county covered by rural land
land	County land area (km^2)
land_frac	*Fraction of county covered by land*
water	County water area (km^2)
water_frac	Fraction of county covered by water
commercial	County commercial area (km^2)
Industrial	County industrial area (km^2)
institutional	County institutional area (km^2)
comm_frac	*Calculated variable: commercial/(commercial + industrial + institutional)*
com_ind_inst	Sum of commercial, industrial, and institutional area (km^2)
avg_max	Average maximum temperature (K)
avg_min	Average minimum temperature (K)
mean_temp	Average mean temperature (K)
precip	annual rainfall (cm)

3.0 APPROACH

The overall goal for this project is to estimate annual average background concentrations across the nation for the 33 urban HAPs. As defined in the introduction, background accounts for natural sources, nearby sources (farther than 50 km), and unidentified sources. Nearby sources include transport from distant cities, trans-border transport for locations near Canada and Mexico borders, and transport from around the globe. Unidentified sources include sources not in the emissions inventory used by the model.

Before trying to indicate how background concentrations might vary spatially (or temporally) with respect to various explanatory factors, it is important to first consider how monitoring data from a given site or area can provide evidence of local background concentration levels. Once this question has been answered, a statistical algorithm can be developed to estimate background levels at different localities. This represents the first stage of the approach, which can be considered descriptive in nature and is discussed in further detail within Section 3.1. With Stage 1 results in hand, it is then possible to investigate how estimated background levels might vary from one local area to the next and, in turn, extrapolate the results to obtain nationwide estimates. This represents the second stage of the approach, which can be considered explanatory in nature and is discussed in further detail within Section 3.2.

3.1 Stage 1: Descriptive

Since the goal is to estimate annual (or typical) background concentrations, then by definition the approach should not seek to identify what occurs during exceptional events. Rather, the approach needs to identify the typical (or area-wide, long-term) background for a site or area.

With the above perspective in mind, the decision was made to avoid approaches that rely on severely restricting the monitoring data to a small subset of observations corresponding to certain events (e.g., days with persistent winds, sharp frontal passages, or the right wind trajectory). The idea behind such approaches is that background levels are more clearly discernible under certain preconceived notions about ideal conditions. The appeal of these approaches is the atmospheric science, or explanatory nature, underlying their logic. See, for example, "Estimation of Hazardous Air Pollutant Background Concentrations Using Ambient Data: A Pilot Study," EPA Contract No. 68-D-98-052, Work Assignment No. 1-16, August 1999. Their disadvantages include (1) the potential analysis burden associated with collecting meteorological or meta-information to identify extreme events and (2) a reduction in the extensiveness of results due to the failure of many sites or areas to exhibit the required extreme event data. In addition, evidence of background concentration levels is likely contained within all measurements, so any approach that eliminates much of the monitoring data leads to a deliberate reduction of available information about background.

Another somewhat less desirable class of approaches might be described as empirical methods, which amount to calculating some percentile of a given set of monitoring data and treating the resulting threshold as an estimate of background. For example, use the 5-10 percent lowest measurements from a given data set, calculate the mean and standard deviation of those

values, then treat the results as a background and uncertainty estimate, respectively. The appeal of such approaches is their simplicity of application. (This advantage should not be discounted given the multitude of HAPs and volumes of monitoring data to be analyzed in projects such as this.) One disadvantage of such approaches is the fact that, *a priori*, they require an arbitrarily defined empirical metric to represent background. Stakeholders must agree on such a metric, yet there will likely be no theoretical or intuitive justification for any given specification (e.g., What might best represent background, the 1st, 5th, or 10th percentile of a year's worth of monitoring data?). Although somewhat less appealing, note that this type of approach is pursued as a secondary (or back-up) analysis in this project to address the many cases of (1) too few observations or (2) too much below-MDL data, for which the primary analysis approach (see discussion below) could not be applied. See subsection 3.1.7 for further discussion of this secondary analysis approach.

Considering the pros and cons of the above approaches, the decision was made to pursue (as a primary analysis approach) a method that: (1) uses all the information about background that exists within a full set of monitoring data (above or below an MDL); (2) is simple enough to be applied on a routine basis in practice; and, most importantly, and (3) is intuitively consistent with the ASPEN model's conceptualization and treatment of background levels. Specifically, the conceptual view of the proposed approach is that local sources add to a constant background. How much the sources add will vary considerably from day-to-day and site-to-site in response to meteorological conditions, total source output, and other local factors. For urban areas, for example, it is assumed that, except for very rare instances, all data values contain a constant background contribution and a varying, non-negative source-oriented contribution. Similar behavior may be assumed for areas and sites that are more rural or background-like in nature, but to a lesser degree (i.e., source-oriented contributions will likely occur less often and typically be less noticeable). Notice that this conceptual view of monitoring data is consistent with the ASPEN model approach of using source emissions and meteorology to estimate source-oriented concentrations, which are then thought of and modeled as additions to some persistent background level.

The first challenge in applying the above conceptual view to monitoring data is to identify a suitable parametric statistical distribution that is both appropriate for the monitoring data to which it is to be applied and capable of outputting the desired parameter of interest (i.e., background). Non-parametric statistical approaches were not considered, due in part to the intent to treat background as a "parameter" to be estimated in the approach. Subsection 3.1.1 describes the distribution proposed for meeting these needs. The next challenge lies in actually fitting such a distribution to a given set of data in order to obtain the desired parameter estimates. Subsection 3.1.2 describes the approach taken for this project. Next, subsection 3.1.3 discusses a modification to the approach to address below-MDL data issues. Then, subsection 3.1.4 highlights several other issues that impact full-scale implementation of the proposed approach and, therefore, require treatment. Subsection 3.1.5 provides a detailed example for benzene in Portland, Oregon. Finally, in part to test the legitimacy of using data sampled before or after calendar year 1999, subsection 3.1.6 examines year-to-year consistency in background concentration estimates for a benzene case study example.

As mentioned above, subsection 3.1.7 discusses in detail a secondary analysis approach that was necessarily pursued to address those air toxics monitoring data cases for which the primary analysis could not be applied. Such cases typically consist of too few total observations or too many below-MDL observations.

3.1.1 Statistical Distribution Assumptions

It is desired to identify a statistical distribution to represent monitoring data for this project. The resulting distribution must adequately approximate the data's behavior. It must also provide an estimate for background. To these ends, what might be described as a shifted gamma distribution, also known as a probability density function (pdf), was pursued. A detailed description of the gamma pdf and its properties may be found in Cassella, G., and Berger, R. L., (1990), "Statistical Inference," Duxbury Press, Belmont, California. (See Figure 3.3a for an example.)

In general, an ordinary gamma distribution has support on the real interval $[0, \infty)$. That is, it applies to variables with a non-negative range of values. The gamma distribution is defined by two parameters, a shape parameter typically represented by the symbol α and a scale parameter typically represented by the parameter β. The specific values of the parameters α and β impact the appearance of the gamma pdf, and variations of these parameters provide for a very flexible family of data-modeling distributions. The shifted gamma distribution introduces a third parameter, call it μ, which changes the support of the distribution from that of the ordinary gamma, i.e., from $[0, \infty)$ to $[\mu, \infty)$. The basic shape of the gamma pdf is left unchanged by such a shift. In the current context, μ is assumed to be non-negative; however, this constraint is not required in general.

Although not substantiated explicitly in this document, it is widely accepted that environmental data in general, and air quality monitoring data in particular, can typically be well approximated by a log-normal statistical distribution. Furthermore, the behavior modeled by a log-normal pdf, including non-negative support, asymmetry, and a heavy right-tailed distribution (i.e., occasional extreme large values), may be modeled similarly via a gamma distribution with an appropriate specification of α and β. For these reasons, the gamma pdf was also identified as a reasonable statistical distribution choice for this stage's approach. Note that asymmetry in environmental data is by no means guaranteed, and the gamma family of pdfs is flexible enough to address both asymmetric and symmetric data. Both models were considered explicitly, but early modeling efforts revealed application difficulties associated with fitting the parameters of the log-normal distribution in conjunction with addressing the other data analysis issues discussed below. Further, the gamma distribution consistently fit the initial data used to develop the Stage 1 model (see the example below).

While an ordinary gamma (or log-normal) distribution might be used to statistically model monitoring data in general, the additional complexity of the shifted gamma distribution was chosen for this particular application. The advantage of this choice is that the application of a shifted distribution is consistent with the conceptual approach of ASPEN discussed above. More specifically, statistically modeling monitoring data as a shifted gamma distribution yields an explicit and direct estimate of background concentration levels from ASPEN's viewpoint —

namely μ. So, in summary, the shifted gamma distribution appears most appropriate for serving the dual purposes of adequately approximating monitoring data behavior and providing a direct estimate for background.

To illustrate the above discussion and support its logic, consider an example of benzene concentrations observed at five monitoring sites in Portland, Oregon, from July 1999 through July 2000. Figure 3.1 summarizes the data via overlaid, site-specific time series plots. While there are high and low concentration periods, there is no obvious seasonal trend in the Portland data. Notice that there is a distinct lack of data below 0.3 μg/m^3, about half of the data are between 0.3 and 2 μg/m^3, and the remainder of the data are spread out progressively thinner from 2 to 9 μg/m^3. Upon first glance, this behavior would appear consistent with a conceptual viewpoint and statistical approach of a constant background and a distribution of source-oriented additions to that background.

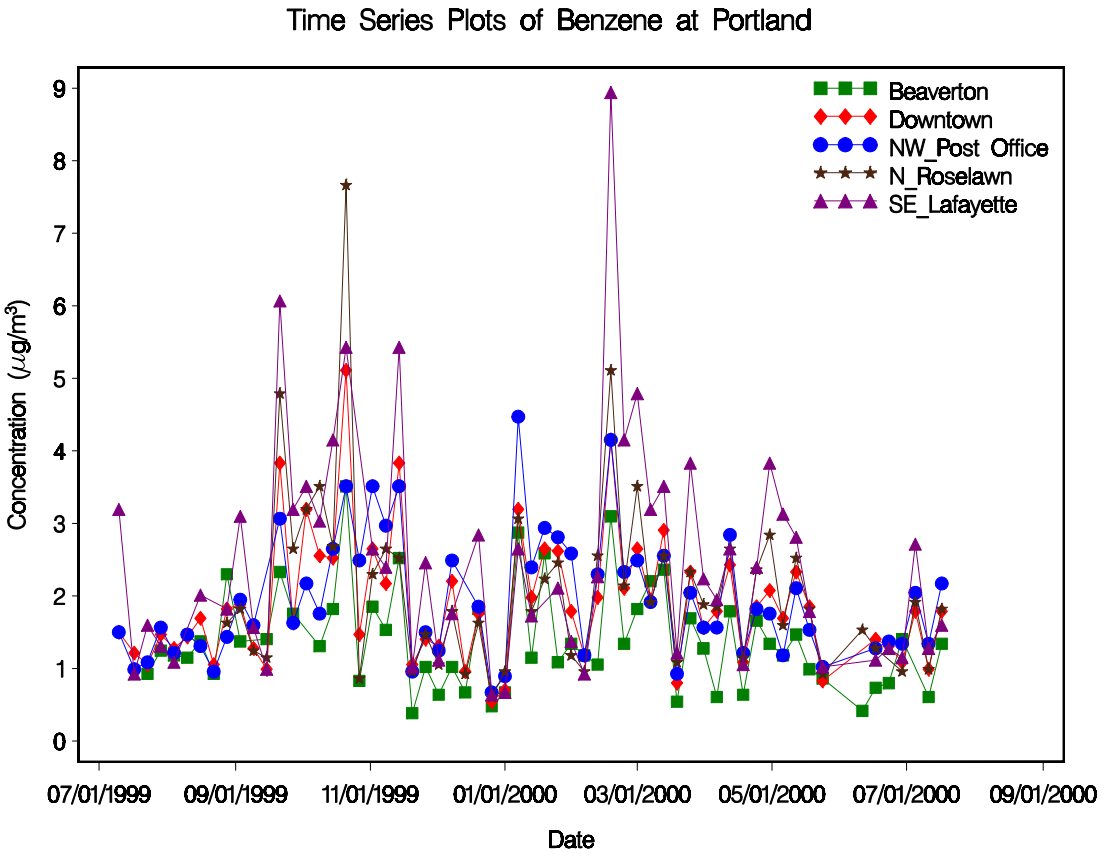


Figure 3.1 Benzene monitoring data (μg/m^3) from five Portland, Oregon, monitoring stations operating from July 1999 through July 2000.

Figure 3.2 presents a quantile-quantile plot (Q-Q plot) of a gamma distribution actually fit to the benzene data at the Downtown site. In general, Q-Q plots demonstrate the fit, or lack thereof, of a proposed statistical distribution to the empirical behavior of a given set of data.

Straight lines in such plots are indicative of a good fit. Other diagnostics include histograms or some type of statistical goodness-of-fit test (to date, not done for this example). Figure 3.2 supports the assertion that the shifted gamma pdf is, in fact, a reasonable choice for the statistical distribution to be used for the approach to this project. It also shows that the data (all the data) contain a positive shift, or background. All of the data for this site are shifted up approximately 0.75 μg/m^3 from a line through the origin. This positive shift is the background in the proposed model. (See Table 3.1 also.)

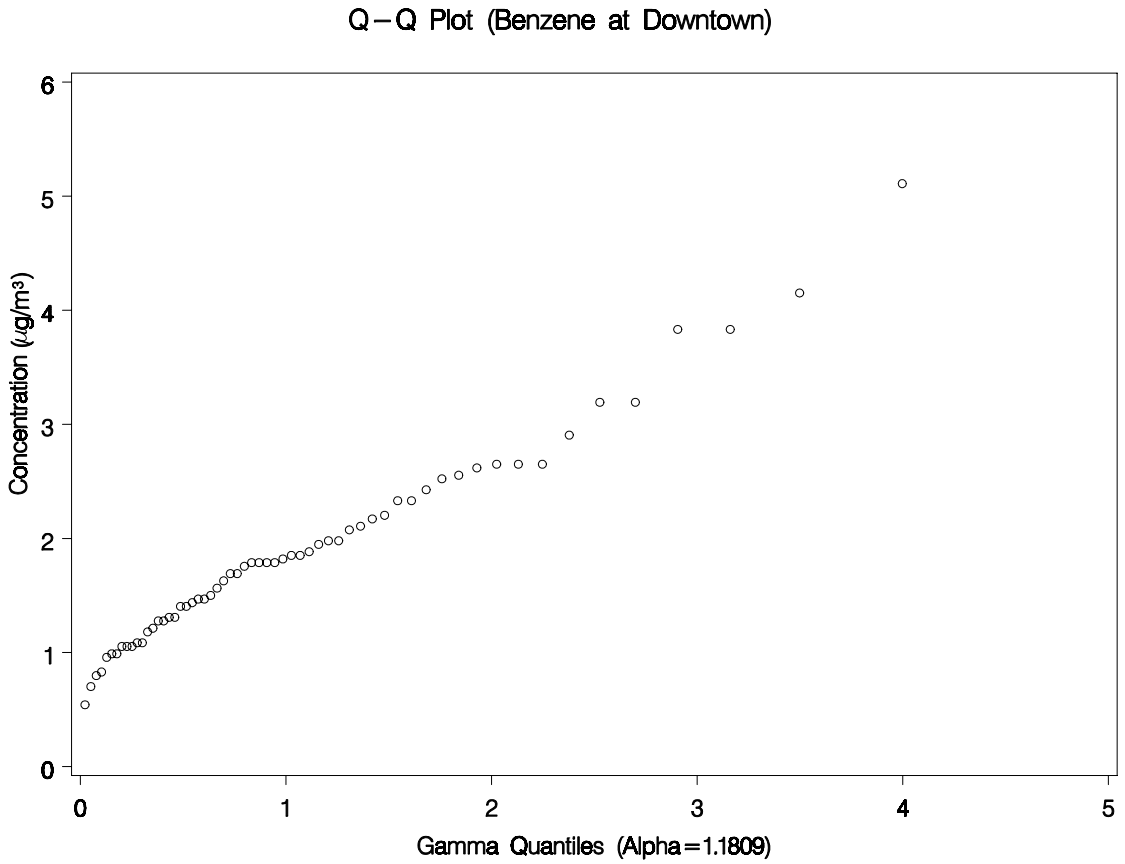


Figure 3.2 Quantile-quantile plot demonstrating the fit of a gamma distribution to benzene data generated from the Downtown site in Portland, Oregon (July 1999 – July 2000).

3.1.2 Applied Estimation Approach

In practice, the approach taken for this project is to fit the shifted gamma distribution model using the method of maximum likelihood estimation. A detailed description of maximum likelihood estimation may be found in Cassella, G., and Berger, R. L. (1990), "Statistical Inference," Duxbury Press, Belmont, California. The method is based on the probabilistic structure of the model. There are three parameters that are fit in the process: the background (μ), the shape parameter (α), and the scale parameter (β). These parameters define the shifted gamma distribution (or pdf) discussed in subsection 3.1.1. The parameters may be fit (estimated)

through an iterative procedure using various softwares that optimize an object function. The object function, called a likelihood, is a mathematical description of the probabilistic structure of the data. For this project, the NLMIXED procedure in the SAS® software system was employed.

The likelihood has the same formula as the data's assumed probability density (i.e., shifted gamma distribution), but with a different interpretation. As a pdf, the parameters that are to be estimated are treated as fixed constants and the data treated as random variables. In that setting, the formula describes the probability of observing data in any given range. As a likelihood, the data are treated as fixed constants (i.e., the actual data observed via monitoring are fixed and known once observed) and the parameters are treated as variables. The maximum likelihood estimates are the ones that maximize the likelihood, and essentially represent the parameter values that would assign the highest probability to the observed monitoring outcome.

For illustration, consider a simple example. Suppose a weighted coin is flipped ten times. The probability that X out of the ten flips is heads is a function of X and the long-term probability that the coin comes up heads (say p). When considering the probabilities associated with the different potential values of X, the parameter p is treated as a constant. In maximum likelihood estimation, once data are observed, the X becomes a known value and p is treated as unknown. To estimate p when in fact 8 out of 10 heads are observed, conceptually the probability of observing 8 heads out of 10 flips is calculated for many different candidate values of p. The value of p that is most consistent with the observed data result of 8 is the maximum likelihood estimate. For this example, 80 percent is the maximum likelihood estimate for the long-term probability, p, of obtaining a result of heads from the given coin. That is, the evidence from the data are suggestive of a coin that yields heads 80 percent of the time, although for this example a mere ten flips is obviously not strong evidence.

3.1.3 Handling Data Below the Minimum Detection Level (MDL)

Recall the discussion of the shifted gamma distribution in subsection 3.1.1. The form of such a model and likelihood/distribution may be written as follows:

$$L(\mu, \alpha, \beta) = \text{“shifted gamma”} \text{ (i.e., } y_i = \mu + \varepsilon_i \text{)} , \qquad \text{(Model 1)}$$

where y_i represents the i^{th} individual HAP concentration; μ represents the true unknown background concentration level; and ε_i represents short-term, source-oriented “shocks” that produce positive deviations from the long-term background. (See Figure 3.3a.) The ε_i's are assumed to behave according to an ordinary gamma distribution with parameters α and β, implying the y_i's have a shifted gamma distribution. (See subsection 3.1.1 above for further details.)

More generally, the effects of measurement detection limits must also be considered. Model (1) above ignores MDL issues, hence, it holds only part of the time. Furthermore, MDL values are difficult to quantify. The usual laboratory method involves relating the MDL to a measurement error. For modeling purposes here, such a point of view is adopted. Specifically, it was decided that any data within a threshold of background plus two times the reported MDL (μ+2*MDL) should be treated as random noise, or at least too imprecise to use individually for

estimating the parameters of the assumed shifted gamma distribution. Note that this viewpoint is consistent with what is commonly referred to as a lowest calibration level (LCL), often defined by laboratories as 3*MDL. Many laboratories do not report below-LCL data due to apparent high uncertainties. In effect, all such data are censored and treated simply as an indicator of "below a threshold." Although non-numerical in nature, such censored data may still be used in the statistical modeling process by making the proportion of the data below the threshold consistent with the data above the threshold.

In general, careful consideration of how to treat below-MDL data (or, more generally, censored data) allows for the entire data set to be used in the statistical analysis regardless of MDL related issues. However, the data must be treated in a dichotomous fashion due to the MDL issues (or, more generally, laboratory uncertainty issues) and the associated censoring threshold discussed above. Specifically, Model (1) is first modified as follows:

$$L(\mu, \alpha, \beta) = \begin{cases} \text{"shifted gamma" if } y_i \geq \mu + 2 \cdot MDL \\ \text{"constant" otherwise} \end{cases} \qquad \text{(Model 2)}$$

where y_i and μ are defined to be the same as in Model (1). (See Figure 3.3b.) In application, the NLMIXED procedure in the SAS® software system iteratively searches until converging to a numerically stable solution for the parameters α, β, and μ.

Note that during the initial model development phase, it was recognized that the numerical model fitting procedure of SAS® did not always converge. The decision to censor data up to a threshold of (μ+2*MDL) served to stabilize this procedure, yielding consistent results for virtually every data set considered. The problem came from needing to know if the censoring threshold was larger or smaller than the shift (background) before fitting the models, in particular, when the difference between them was small. By using (μ+2*MDL), the censoring threshold is always greater than the shift (background). Sensitivity analyses were conducted, which suggested that background conclusions were qualitatively unaffected under a range of censoring thresholds, including (μ+1*MDL) and (μ+1.5*MDL). Also note that when, in reality, μ is large relative to the data's MDL, then this approach approximately simplifies to that of Model (1). In other words, if a given data set does not include data near or below its reported MDL, then the statistical modeling approach for addressing below-MDL data reduces to the simpler approach for addressing data without MDL issues.

While the above model can be fit numerically, the dichotomous treatment of the data results in a discontinuity in the likelihood. This discontinuity at (μ+2*MDL) can cause a numerical instability in the estimate of the standard error of the background parameter. Hence, a slight modification to the likelihood was introduced. The modification introduced was to connect the two portions of the likelihood with a positive, finitely-sloped line over a short range; namely, a range of one MDL. (See Figure 3.3c.) This ensures that the object function discussed in subsection 3.1.2 is continuous. Since the function being optimized is now different than for Model (2), the estimates are different from Model (2) as well. However, in most of the test cases considered, the estimates obtained without the continuity correction [i.e., Model (2)] were within a 95 percent confidence interval of the estimates obtained with the continuity correction.

Moreover, the uncertainty of each background estimate is now estimable for almost all cases. In summary, the likelihood used to model the data is modified slightly from that of Model (2) to take the following form:

$$L(\mu, \alpha, \beta) = \begin{cases} \text{"shifted gamma" if } y_i \geq \mu + 2 \cdot MDL \\ \text{"linear" if } \mu + MDL \leq y_i < \mu + 2 \cdot MDL \\ \text{"constant" otherwise} \end{cases} \qquad \text{(Model 3)}$$

Note that there are still some cases where the background estimate's associated standard error estimate is not calculable or is flagged by SAS® as highly uncertain. In these cases, the decision was made to report the standard error for the background estimate as the maximum of: (1) the uncertain standard error given by SAS®, (2) one-half of the background estimate, and (3) one-half of the MDL. This approach provides a conservative uncertainty estimate in such cases. That is, uncertainty will tend to be over-estimated.

Consider Figures 3.3a, 3.3b, and 3.3c to illustrate the approach and the differences among Models (1), (2), and (3). Figure 3.3a displays a shifted gamma pdf for uncensored data, or data having no MDL issue, with parameters α=1.5, β=0.75, and μ=0.4. The appearance of the distribution in Figure 3.3a corresponds to Model (1). In comparison, Figure 3.3b displays a shifted gamma pdf as modified for censored data, or data having an MDL issue, with the same α, β, and μ parameters as above. The appearance of the pdf in Figure 3.3b corresponds to Model (2). Finally, Figure 3.3c displays a continuous, shifted gamma pdf as modified for censored data, again with the same α, β, and μ parameters as above. The appearance of the pdf in Figure 3.3c corresponds to Model (3). (Note that the values of the α, β, and μ parameters in these examples were chosen as reasonable values for benzene based on the data used during Stage 1 model development.)

Of note, the lower portion of the solid curve in Figures 3.3b and 3.3c represents the modeling of all the censored data, jointly, as a random probability of concentrations falling below the (μ+2*MDL) threshold. In modeling the data's entire distribution, the total area under this portion of the curve in these two cases is fixed so that the total area under the entire pdf curve is equal to one, as required for a pdf. However, as discussed above, observe that the associated likelihood has a discontinuity at the censoring threshold of (μ+2*MDL) in the case of Model (2) as depicted by Figure 3.3b. As stated above, in application, this modeling approach typically lacked the ability to estimate the uncertainty of background estimates, hence, the decision was made to apply Model (3) as depicted by Figure 3.3c.

Also of note, recognize that the distributions (pdfs) displayed in Figures 3.3a, 3.3b, and 3.3c are, conceptually at least, no more complicated than the shape and associated interpretation of the familiar bell-shaped curve of the normal distribution (pdf). Any perceived complexities arise from (1) the asymmetry of the curves, (2) the positive shift of the curves to values above zero, and (3) the need to work with censored data. Respectively, these complexities are considered necessary in the Stage 1 approach to (1) reasonably model the true

nature of the monitoring data, (2) explicitly estimate background, and (3) properly handle below-MDL data and associated issues.

In closing this subsection's discussion, it is noted that the primary statistical methodology discussed thus far, while attempting to address below-MDL data, will likely fail to provide a solution when too much below-MDL data are present. Regardless of MDL issues, it will also suffer from too few total observations in a given data set. Specifically, the decision was made to apply this primary approach only in those cases for which at least 20 above-MDL observations are present for a given pollutant at a given monitoring site. This decision was made based on preliminary model development using benzene and carbon tetrachloride data. Subsection 3.1.7 below discusses the secondary data analysis approach necessary to address monitoring data cases not handled by the preferred primary approach, mainly cases of too few total observations or too many below-MDL observations.

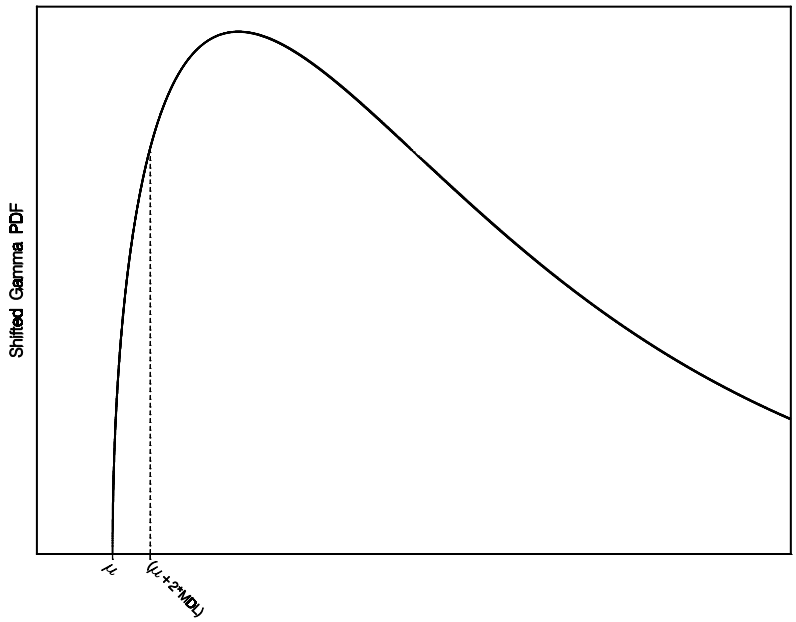


Figure 3.3a Example of a shifted gamma distribution (pdf) with parameters α=1.5, β=0.75, and μ=0.4: i.e., Model (1).

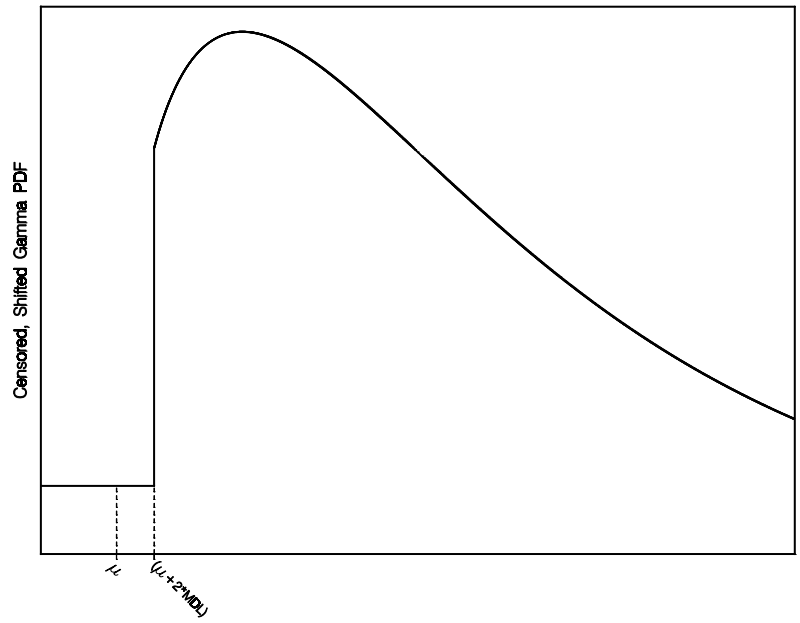


Figure 3.3b Example of a censored, shifted gamma distribution (pdf) with parameters α=1.5, β=0.75, μ=0.4, MDL=0.1, and a censoring threshold of μ+2*MDL: i.e., Model (2).

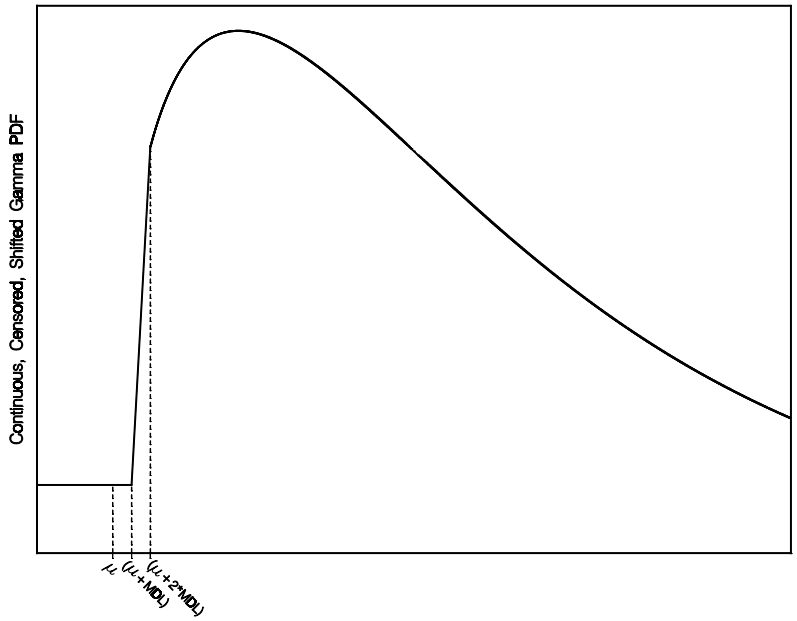


Figure 3.3c Example of a continuous, censored, shifted gamma distribution (pdf) with parameters α=1.5, β=0.75, μ=0.4, MDL=0.1, and a censoring threshold of μ+2*MDL: i.e., Model (3).

3.1.4 Other Issues

The purpose of this subsection of the report is to describe two other important decisions that were made during the development of the Stage 1 approach. Both decisions were arrived at from the perspective of routine, large-scale implementation. That is, the reality of this project is that the approach is to be applied to approximately one-half million monitoring observations, spanning hundreds of monitoring sites and associated counties, and covering up to 33 distinct HAPs. Given budget and time constraints, this level of analysis burden requires a practical approach or algorithm that, in some sense, can be routinely applied over and again without repetitive, painstaking oversight on the part of the responsible data analyst. With that perspective in mind, the following two decisions were made.

First, it was decided that the approach summarized in subsections 3.1.1 through 3.1.3 would always be applied on a site-specific basis. Meanwhile, to be consistent with ASPEN model predictions and to facilitate the approach of Stage 2 (see discussion in Section 3.2 below), background estimates in the current stage are sought on a countywide scale. As a result, final Stage 1 background estimates are obtained from an average of site-specific estimates within a given county in those cases where multiple monitors provide data for that county. Based on initial efforts to develop Stage 1, the advantage of this approach is that reliable background estimates can be obtained routinely in almost every case without concern for convergence or other issues associated with using the NLMIXED procedure in the SAS® software system. Note that other approaches for addressing below-MDL data, such as imputation via simulation, were considered. Such approaches might lead to application methods that do not suffer from convergence issues, and could be considered further in the future.

A more appealing approach from a statistical perspective would be to jointly model all the sites within a given county; however, such an approach is more complex as it requires additional model parameters to account for differences between the data distributions of multiple sites (i.e., site effects). Initial modeling development efforts revealed that this more appealing but more complex approach could lead to convergence difficulties for a larger number of cases during application. Therefore, the decision was made to pursue the simpler approach of obtaining site-specific background estimates and then averaging across sites within counties. In general, we expect the final background estimation results to be similar for the two approaches.

The second decision made with large-scale implementation in mind was to eliminate the highest measurement or top one percent of the data (whichever is greater) from each site's data set prior to analysis. The reason for this decision was to provide an improved fit of the shifted gamma distribution (see subsection 3.1.1) to some of the data sets being analyzed. Recall the quantile-quantile plot of Figure 3.2. As discussed in detail in "Selecting Inorganic Constituents as Chemicals of Potential Concern at Risk Assessments at Hazardous Waste Sites and Permitted Facilities," deviations from a straight line in such plots are suggestive of multiple populations contributing to the overall distribution of a given data set (California EPA, 1997). For example, under the right environmental or atmospheric conditions (e.g., wind direction, day of week, etc.), a given monitoring site may experience a wholly different range or distribution of concentrations from what might be considered more typical for that site. It is these more typical concentrations

(i.e., the lower portion or lowest distinct line of a quantile-quantile plot like Figure 3.2) that are most relevant to the goals of this project.

A potentially more appealing approach would be to fit a shifted gamma distribution to each data set, study the resulting quantile-quantile plot, determine whether a differential population effect is in fact present, and, if so, determine the optimal cut-off point for excluding extreme data at the high end of the distribution. Obviously such an approach would be extremely time consuming and, therefore, impractical for applying across hundreds of monitoring sites and up to 33 distinct HAPs. Moreover, initial efforts to develop Stage 1 using benzene and carbon tetrachloride data, including the investigation of a number of quantile-quantile plots, suggested that such differential population effects were typically confined to a single outlier or very few observations at the high end of a given data set's distribution. These few observations simply did not match well with the distributional behavior of the remainder of the data or the shifted gamma distribution used to model that behavior. Hence, the decision was made to eliminate the highest concentration or top one percent of the data as part of the approach's automated procedures. In general, as the goal of this project and primary purpose of the chosen statistical models is to obtain a sound characterization of the lower portion of each data set's distribution (i.e., background), this decision regarding the highest few observations is expected to have very little impact on any of the final background estimation results.

3.1.5 Example: Estimating Background for Benzene in Portland, Oregon

The approach summarized in subsections 3.1.1 through 3.1.4 was applied to benzene monitoring data in Portland, Oregon. Table 3.1 provides a numerical summary of the data (mean, standard deviation, minimum, and maximum concentration by site), and the results (estimate and standard error) of estimating background in this case. Figure 3.4 provides similar information visually (see Figure 3.1 also). The box plots in Figure 3.4 indicate each site's minimum, 25th percentile, median, 75th percentile, and maximum concentration. The horizontal reference line in the figure identifies the resulting countywide background estimate for Portland ($0.6282\ \mu g/m^3$). As stated in subsection 3.1.4, the countywide background estimate is simply an average of the site-specific background estimates within a given county. To obtain an approximate 95 percent confidence interval for any of the background estimates provided in Table 3.1, add and subtract from the estimate a multiple of 1.96 (or 2) times its associated standard error. (Note that such a confidence interval assumes a normal distribution, which may not be true.)

For this example, notice that the Beaverton site-specific background estimate is lower than the background estimates of the other sites. The site-specific background estimates for the remaining four sites, with the possible exception of the Downtown site, are quite consistent. This is true despite the differences between the lower portions of their respective data distributions, as evidenced by the minimum values of Table 3.1 and the box plots of Figure 3.4. The similarity of background estimates among these four sites lends some credence to the approach presented in subsections 3.1.1 through 3.1.4. Furthermore, the difference in the Beaverton site's background result is to be expected based on Battelle's prior understanding of this site. Specifically, Beaverton is in a different, and more rural, county with a series of hills

separating it from the primary downtown area of Portland and the rest of the monitoring sites. Inclusion of this site in this example serves to pull down the Portland-wide background estimate to a value of 0.6282 μg/m^3. However, based on the approach summarized above in subsection 3.1.4, the Beaverton result is ultimately attributed to a different county, namely Washington county. The results from the remaining sites are averaged and attributed to their associated county, namely Multnomah. Refer to Table A.10 of Appendix A for the project's complete set of countywide benzene background results in Oregon counties (recognizing that data from other sites or years may be incorporated as well.)

Also of note, observe the relatively large standard error for the Beaverton background estimate in Table 3.1. This site's model, as applied in SAS®, did not produce a reliable standard error estimate by default. As discussed in subsection 3.1.3, an alternative standard error was therefore reported in this case, which led to the inflated (conservative) uncertainty result for Beaverton's background in Table 3.1 (and, implicitly, in Table A.10 of Appendix A).

Table 3.1 Numerical summary of July 1999 through July 2000 Portland, Oregon, benzene monitoring data (μg/m^3) and background modeling results [1]

Site	Sample Size	Mean	Standard Deviation	Max	Min	Background Estimate	Standard Error
Beaverton	56	1.3840	0.6948	3.5127	0.3832	0.4067	0.2034
Downtown	60	1.8953	0.8908	5.1094	0.5429	0.5491	0.0593
NW_Post Office	59	1.9204	0.8738	4.4707	0.1000	0.7359	0.0563
N_Roselawn	52	2.0972	1.2321	7.6641	0.6067	0.7127	0.0644
SE_Lafayette	55	2.4844	1.5601	8.9415	0.6387	0.7364	0.0784
All Sites	282	1.9511	1.1342	8.9415	0.1000	0.6282	0.0483

[1] Sample sizes and summary statistics correspond to the complete Portland benzene data set. As described above in subsections 3.1.1 through 3.1.4, a small number of these observations may not be included in the calculation of the background estimate and associated standard error.

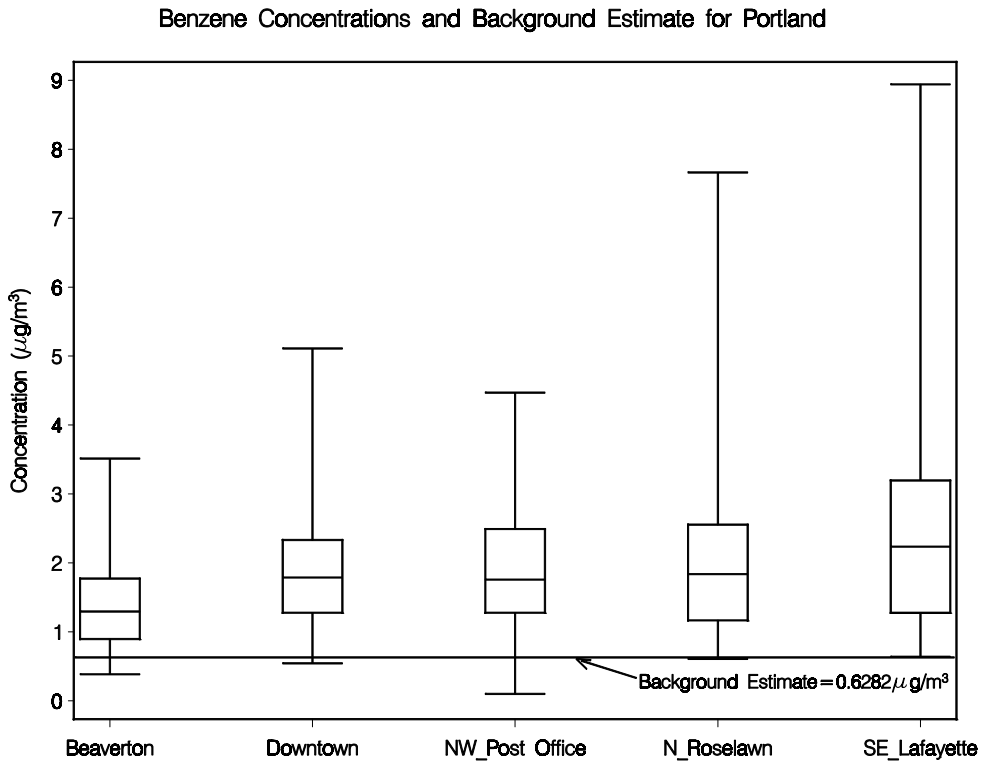


Figure 3.4 Graphical summary of July 1999 through July 2000 Portland, Oregon, benzene monitoring data (μg/m³) and background modeling results.

3.1.6 Year-to-Year Consistency of Background Estimates for Benzene

The statistical approach in Stage 1 assumes that, for a given monitoring program and site, the background does not change substantially from year-to-year within the 1995 through 2002 time frame of the data. If the data were only collected for a 1-year span, as is the case for the Portland example, this is a non-issue (within-year variation is a different issue). However, for some of the data there are sites with observations across multiple years. In particular, the database contains 77 sites in 45 counties with sufficient benzene observations in each year from 1997 through 2000 to obtain separate background estimates for each site and year. This subsection considers applying the Stage 1 approach to these data on a year-specific basis in an effort to test the assumption of relatively constant backgrounds across the time frame in question (at least in the case of benzene within these 45 counties).

Table 3.2 shows summaries of both the concentration data from the 77 case study sites and the distribution of the county-level benzene background estimates. Figure 3.5 shows box plots of the yearly concentration data from those sites and box plots of the background estimates by year. There is a slight downward trend in both the mean and median concentration levels across the sites, which argues against an assumption of constant background levels from 1995 through 2002. However, the downward trends apparent in Figure 3.5 are statistically

insignificant in the cases of both the mean concentration level and mean background level across sites. Furthermore, the magnitude of any apparent background trend, whether statistically insignificant or otherwise, is relatively small according to the results of Table 3.2. In summary, these case study results, while important, do not cause enough concern to merit restricting the project's data to calendar year 1999 only. Such a decision would cause far more serious concerns for the project; namely, a significant reduction in the expansiveness of results obtained in Stage 1. This issue may warrant further investigation.

The authors also note that one of the issues raised during this project has been the feasibility of conducting source apportionment, or another relatively straightforward statistical methodology, in an effort to better understand the sources contributing to background concentrations. Ultimately, such findings may help determine the degree to which background concentration levels are impacted by emissions reductions. The approach to this case study example may begin to provide some evidence toward this end. For example, **assuming** that the slight downward trend in mean benzene levels within these 45 case study counties can indeed be attributed to emissions reductions, Table 3.2 and Figure 3.5 suggest that such reductions may have a residual effect on background levels as well. This assertion should be viewed with extreme caution, however, since the results of Table 3.2 and Figure 3.5 are still in draft form. Also, based solely on this type of data analysis, any apparent trends cannot necessarily be attributed to emissions reductions. This issue needs to be considered further. In particular, it might prove useful to apply more sophisticated source apportionment techniques to obtain more definitive answers. To date, the feasibility of such a source apportionment application has not been considered in detail.

In closing this subsection, it is noted that during discussions with the Work Assignment Manager (WAM) and another EPA/OAQPS Technical Advisor, it was suggested that the sort of exercise conducted and presented in this subsection might be worthwhile pursuing for HAPs other than benzene. In particular, a longer time series for metals was mentioned. Along with the issues highlighted above for benzene, such an exercise might reveal other information, such as trends in metals MDLs over time. Due primarily to time and resource concerns, such activities have not been pursued to date. They may be explored in the future, and are worthy of further consideration.

Table 3.2 Numerical summary of 1997 through 2000 benzene monitoring data ($\mu g/m^3$) and background modeling results (from the same 77 sites in 45 counties)

Year	Number of Observations	Concentration Distribution					Background	
		Mean	Standard Deviation	Min	Median	Max	Mean	Standard Deviation
1997	4,008	2.095	2.645	0.014	1.330	41.594	0.539	0.412
1998	4,267	1.772	1.866	0.009	1.211	19.327	0.515	0.301
1999	4,426	1.767	1.917	0.003	1.150	26.675	0.453	0.298
2000	4,540	1.637	1.954	0.003	1.054	21.819	0.445	0.313

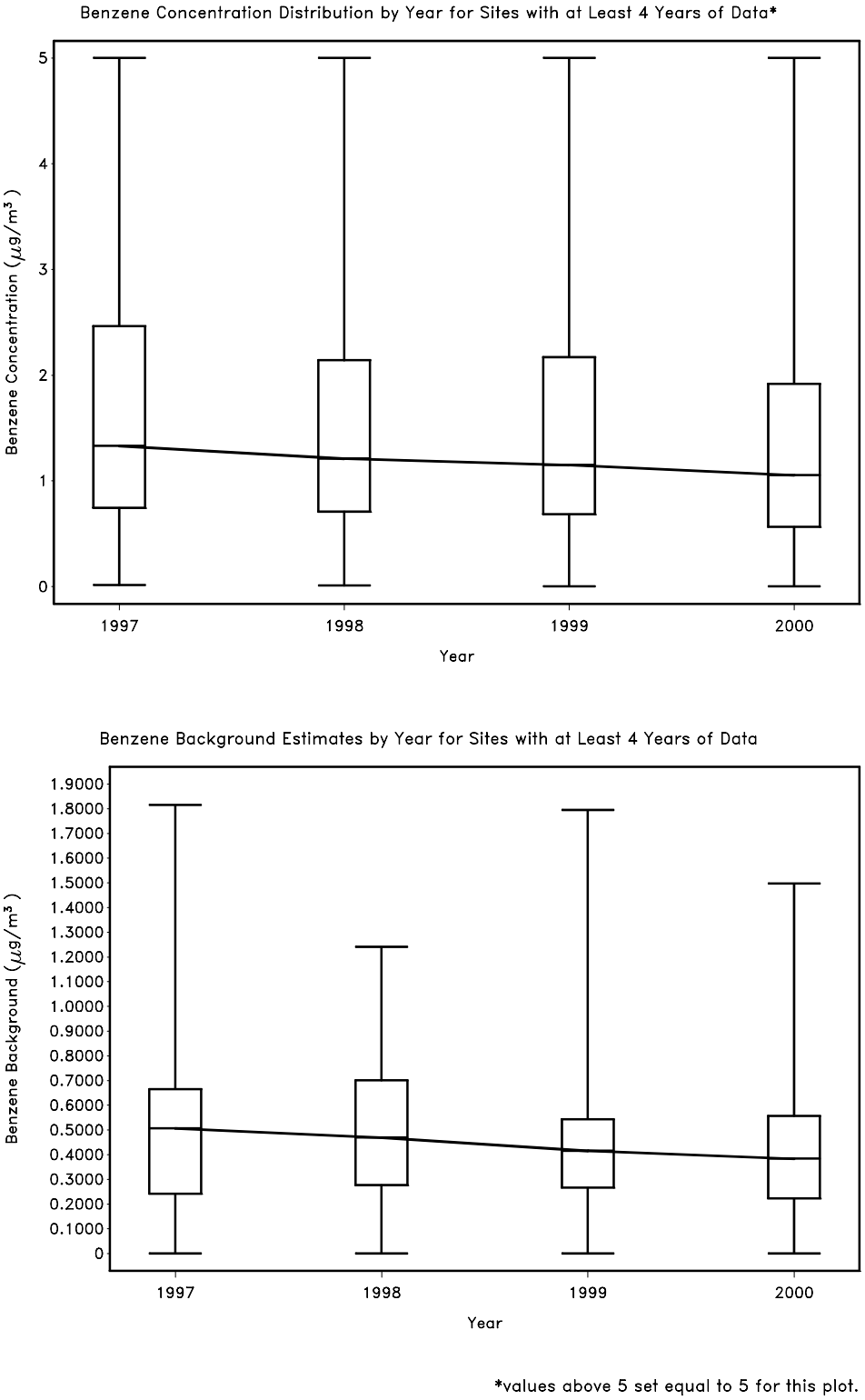


Figure 3.5 Distribution by year of benzene concentrations (top) and distribution by year of the countywide benzene background estimates (bottom).

3.1.7 Secondary Analysis Approach

Model (1), as discussed above, will provide a reasonable background estimate when all monitoring data are above their respective MDLs and enough total observations are available in a given data set for conducting the statistical analysis. Models (2) and (3) extend this capability to address cases when some of the monitoring data are below their respective MDLs. While these models extend the methodology to provide more expansive Stage 1 results, they do not handle all cases. In particular, as discussed throughout this document, the primary statistical analysis approach presented and highlighted in subsections 3.1.1 through 3.1.6 will often fail when presented with too few total observations or too many below-MDL observations in a given data set. Either scenario effectively amounts to too few above-MDL observations. This primary methodology may also fail in a few cases when the size of the data set would appear to be adequate and the data are mostly above their respective MDLs, yet the given application still does not adequately converge to a solution. A secondary analysis approach is required to provide background estimates in all of these cases.

A specific secondary approach is motivated and proposed in this subsection. This approach was applied and the results are included with the results in Section 4. The approach is referred to as secondary because it has not received the same level of review as the primary methodology discussed and highlighted above (see the discussion at the beginning of Section 3.1), and by necessity cannot have the same level of precision. For the combined data used in Stage 1 of this project, Table 2.1a summarizes the subset of data for which the primary statistical approach could be applied. Table 2.1b summarizes the remaining data for which the secondary, or back-up, approach was required, as discussed below.

Motivation. Laboratory analytical sensitivity, which in turn leads to a lack of reported monitoring data below associated MDL thresholds, is generally viewed as a nuisance during data analysis applications. The need for Models (2) and (3) as discussed above represents just one example of this issue. However, in the current context, this apparent nuisance can be capitalized on to provide potentially valuable information about background. To see this, consider a simple example. Suppose a rural background site monitors formaldehyde once every three days for a year, yielding approximately 120 observations. Suppose further that every single observation at this site is reported as below its respective MDL; and, for simplicity, the associated MDLs are all reported as the same value of 0.2 $\mu g/m^3$. So what is known about background in a case like this? The information provided by these data is evidence that background for this pollutant in this particular area of the country must lie somewhere between 0 and 0.2 $\mu g/m^3$. While this may not be ideal quantitative information, it nonetheless gives a very informative bound on background. While the primary statistical methodology of this project cannot be applied in a case like this, a reasonable approach can still be applied to provide some background estimate.

At the opposite extreme of the above example, consider a case where all the observations in a given data set are above their respective MDLs, but the data set consists of only a small total number of observations, say 10, for example. The primary statistical methodology of this project will likely fail to provide a background solution in this case as well. Again, however, potentially useful information is available for estimating background in some manner in a case like this.

Likewise, there will be cases somewhere in between the two extremes of (1) strictly below-MDL data or (2) too few strictly above-MDL observations. The secondary analysis approach discussed below is an attempt to address all such cases in a reasonable and consistent manner. It also recognizes that, intuitively, a reasonable background solution should look somewhat different depending on which of the above two extremes (or somewhere in between) occurs.

So, considering the above examples, how might one use a relatively limited amount of monitoring information to still provide a reasonable estimate of background? Also, how might one characterize the uncertainty of such an estimate? At the same time, to remain consistent with the goals of this project, the approach to addressing these two questions should be relatively straightforward for the sake of large-scale application.

Details of Approach. The approach is to first recognize two extremes in the amount of source activity, and then look for a reasonable means for continuously choosing a background value between those extremes. For both cases, assume that MDL/2 substitution has been done for any data below the MDL prior to investigation.

The first extreme is when a site is not affected by sources and is only measuring background. In this case, the median, or 50th percentile, of the data is considered to provide a reasonable upper bound estimate for the annual mean background concentration. By using the median, outliers will have very little influence on the estimate.

The second extreme is when a site is frequently affected by sources and is rarely measuring just the background. If the site has measured only background on a given day, then that observation would be among the lowest measurements available for the given site. As above with the median, a robust estimator is desired in this case. The minimum is not a robust estimate for the annual mean background. Beyond the statistical properties of the minimum, it may not yield the correct value because conditions that resulted in extremely low source measurements may also have resulted in extremely low background. Further, if the background has some seasonality, then that could also result in the minimum being below the target annual-based background value. For these reasons, the 5th percentile of the data was chosen in this case as a better, hopefully more robust, estimate of the mean annual background.

What about estimating background for cases in between the above two extremes? We start by further distinguishing the above two extremes by the amount of variability in the data, as derived from sources (or lack thereof). In the first case (few source impacts), the data's standard deviation may be assumed to be small, say on the order of two times the MDL, because it is derived mostly from laboratory imprecision and not source-influenced environmental variability. In the second case (many source impacts), the standard deviation is expected to be relatively large, say on the order of the mean of the data, due mostly to the presence of source-influenced variability. So, in between these two extremes, we consider a background estimate that again is some percentile of the data, where the chosen percentile is linearly interpolated to fall somewhere between the 5th and 50th percentile of the data (i.e., somewhere between the estimates chosen for the two extreme cases discussed above). Furthermore, based on the variability differences between the two extreme cases, it seems reasonable to choose an interpolation scheme that is a function of the data's variability. In other words, the interpolated

background estimate between the two extremes of the 5th and 50th percentile of the data should fall closer to the extreme whose presumed variability more closely matches the variability of the data at hand.

This still leaves an ambiguous case. Suppose the mean is less than two times the MDL. In this case, the basis for declaring the standard deviation small or large is backwards. However, the only way for the data to yield this case is for almost all of the data to be less than two times the MDL. Since data within two times the MDL are often considered to be not significantly different from zero, it was decided that the median (as opposed to a lower data percentile) would be the most conservative choice in this case.

Hence, the algorithm for estimating the annual mean background, when Model (3) could not be used, is the following. First, any data below the MDL are replaced by MDL/2. Next, the mean and standard deviation of the data are computed. Then:

Case 1. If the mean is less than two times the maximum MDL, then the background estimate is the median of the data (after MDL/2 substitution).

Case 2. If two times the maximum MDL is less than the mean, then the background estimate is a percentile of the data (after MDL/2 substitution). The specific percentile is a function of the standard deviation, denoted by STD, and is determined as follows:

$$percentile(STD) = \begin{cases} 50th & if \quad STD < 2 \cdot MDL \\ ceil\left(50 - \dfrac{45 \cdot (STD - 2 \cdot MDL)}{(mean - 2 \cdot MDL)}\right) & if \quad 2 \cdot MDL < STD < mean \\ 5th & if \quad mean < STD \end{cases},$$

where the ceil() function rounds the result up to the nearest whole number.

In all of these cases, the standard error of the resulting background estimate was set to one-half of that background estimate. Hence, the usual confidence interval becomes 0 to 2 times the estimate. So, for example, if all of the data are below the MDL then the estimate becomes MDL/2 and the confidence interval ranges from 0 to MDL. Note that this estimate of uncertainty could not be determined using a statistical argument due to the limitations of the data sets being addressed in this secondary analysis.

Attributes and Limitations of Estimator. This secondary methodology was pursued based on discussions with the Work Assignment Manager and a Technical Advisor, and the expressed desire to obtain background estimates in all cases (i.e., whenever monitoring data are not available, including cases when the more rigorous approach may not apply). Further refinement of this proposed approach should be based on additional review and be cognizant of some initial points regarding the properties (or limitations) of the approach described above:

- The estimator yields an estimate in all cases, but that estimate is thought of as an upper bound to the background (conservative).

- The estimate is consistent with intuition as an upper bound for background when most of the data are below the MDL.

- If separate programs both yield essentially below-MDL data, then the one with the smaller MDL will be the one with the sharper background estimate (i.e., the one with the smallest confidence interval).

- Consider a pollutant where background is spatially uniform and all available monitoring data across the country are below their respective MDLs. Suppose further that different MDLs are associated with different counties. This will show up as variability in the background across the country. After accounting for the uncertainty of each estimate, however, the final conclusion would likely be that background is not observed as statistically significantly varying over space.

3.2 Stage 2: Explanatory

The second stage of the approach involves investigating how the background varies spatially with respect to various possible explanatory variables (the "why" question). These potential variables were discussed in Section 2.2. In this stage, the site-specific and county-specific background estimates developed in Stage 1, together with the explanatory variables discussed in Section 2.2, were employed to see how the county estimates change with respect to the explanatory factors. Various multiple regression models of the background as a function of the explanatory factors were explored, thereby enabling the extrapolation of background estimates nationally according to the relationships implied by such models.

In general, these models took the following form:

$$B_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \qquad \text{(Model 4)}$$

where B_i represents the i^{th} background concentration estimate from Stage 1; the β_j's represent the effect on background due to the j^{th} explanatory variable X_{ji} associated with the i^{th} background estimate ($j = 1, \ldots, p$); and ε_i represents residual error in the model due, in part, to any behaviors in B_i that are left unexplained by the model. Once Model (4) is fit, values of $[X_1, \ldots, X_p]$ for areas of the country without monitoring data can be used together with the model to estimate any area's background.

3.2.1 Outlier Cleanup and Responses to Stage 1 Comments

Peer review of the approach and results from Stage 1 suggested that the spatial gradients for some of the background estimates may be too steep. Specifically, the observed background gradients appear to track closer than expected with ambient (or average) conditions in some cases. One interpretation of this phenomenon is that some of the sites used in estimating background may have been overly influenced by local source emissions. The statistical methodology used to estimate background in Stage 1 may not be able to overcome such outlier cases.

As such, Stage 2 begins by attempting to address the above described issue. Given the limited resources for this effort, an in depth investigation of each site's (or county's) background estimate for each pollutant can not be conducted. Instead, a simpler and far less resource intensive procedure is required. To that end, Figures 3.6 and 3.7 below provide an example, using benzene, of the results of the chosen approach for Task 1. Specifically, the approach was to eliminate from the analysis, as an outlier, any site whose background estimate from Stage 1 appears extreme in the upper tail of the overall distribution of Stage 1 background estimates. Given the nature of the problem, it seems appropriate to focus on outliers in the upper tail of the distribution. Such results are perhaps due to extreme local emission sources affecting the ability to obtain a relatively pure background estimate based on the sites(s) providing data in such cases. This approach is intended to address the primary technical issue identified from the peer review of the Stage 1 report.

In general, the chosen approach was to eliminate outliers visually, and then to recalculate a county background estimate with the "clean" data set. For benzene, there were initially 177 counties that had a Stage 1 background estimate. Two of those counties (from Puerto Rico) were removed because they had no corresponding explanatory variables. In addition, five sites from five different counties were removed due to high end extreme data. Three of those sites were the only site for their respective counties; therefore, three additional counties were removed for a total of 172 counties. Data on the two remaining affected counties are presented in Table 3.3. The shaded rows in Table 3.3 are the three counties that were removed from the analysis. The comparison of the benzene background estimates in Figures 3.6 and 3.7 illustrates the choice to eliminate the five sites with the highest background estimates. Figure 3.6 displays the distribution of benzene background estimates by site before the outliers were removed, while Figure 3.7 shows the distribution of benzene background estimates by county after the removal of outliers at a site level and after the recalculation of affected county estimates.

Table 3.3 Recalculation of countywide benzene background estimates for affected counties.

State	County	Stage 1 Number of Sites	Stage 1 Background Estimate	Revised Stage 1 Number of Sites	Revised Stage 1 Background Estimate
California	Los Angeles	27	0.8089	26	0.7049
Colorado	Denver	1	3.0349	0	NA
Idaho	Nez Perce	13	0.7916	12	0.6396
Kansas	Sedgwick	1	4.7919	0	NA
Ohio	Scioto	1	4.7200	0	NA

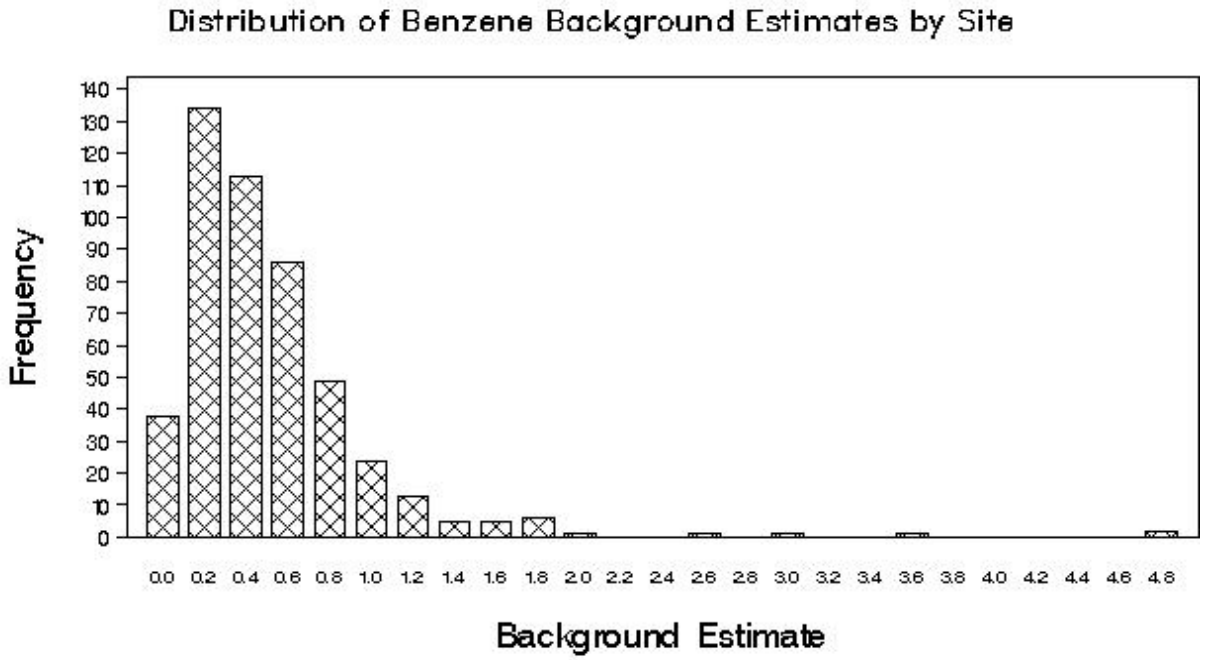


Figure 3.6 All 479 sites with (Stage 1) benzene background estimates.

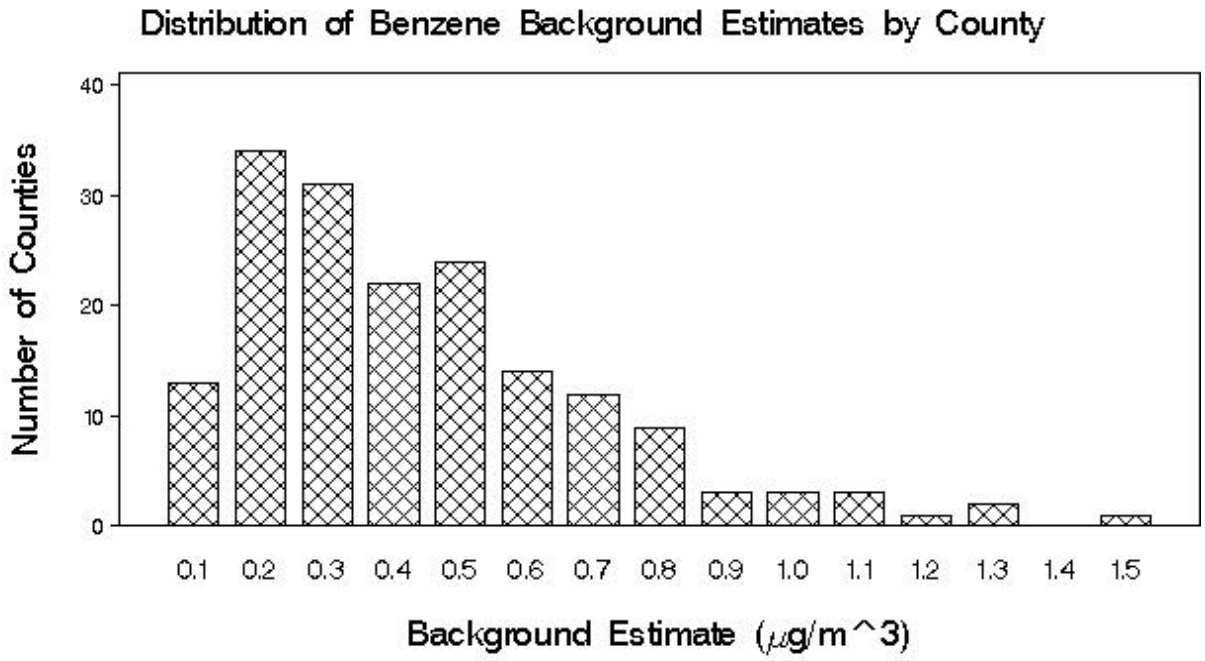


Figure 3.7 172 counties with benzene background estimated from Stage 1 (high outliers have been identified visually and removed).

As described above, peer review of the results from Stage 1 suggested that the observed background gradients may track closer than expected with ambient (or average) conditions in some cases. In other words, it is suspected that some monitoring locations were overly influenced by local source emissions. As a brief investigation of this observation, the correlation between background estimates and emissions was compared to the correlation between ambient concentrations and emissions. This investigation was limited solely to benzene, as the scope and budget of the work assignment did not allow for an in depth investigation of each pollutant.

The approach for the investigation was to run a regression using PROC GLM in SAS®, where total emissions served as the predictor (independent) variable, and background estimates or the average concentration for each county served as the response (dependent) variable. The results of these regressions revealed that, as expected, average benzene concentrations were somewhat correlated with benzene emissions ($r^2 = 0.10$, $p < 0.0001$) while benzene background estimates showed a much weaker correlation, which was not statistically significant ($r^2 = 0.013$, $p = 0.145$). These relationships are represented graphically in Figure 3.8, which plots benzene emissions against benzene background estimates (top graph), and benzene emissions against average benzene concentrations (bottom graph).

In summary, it is felt that this result is very important evidence in favor of the efficacy of the Stage 1 statistical approach, coupled with the outlier removal described in this section of the report. Specifically, as expected, ambient average concentrations correlate with nearby emissions; whereas, as desired, estimated ambient background concentrations do not. That is, the statistical methodology used to estimate background in Stage 1 would appear to have removed from the data the influence of nearby sources, as desired.

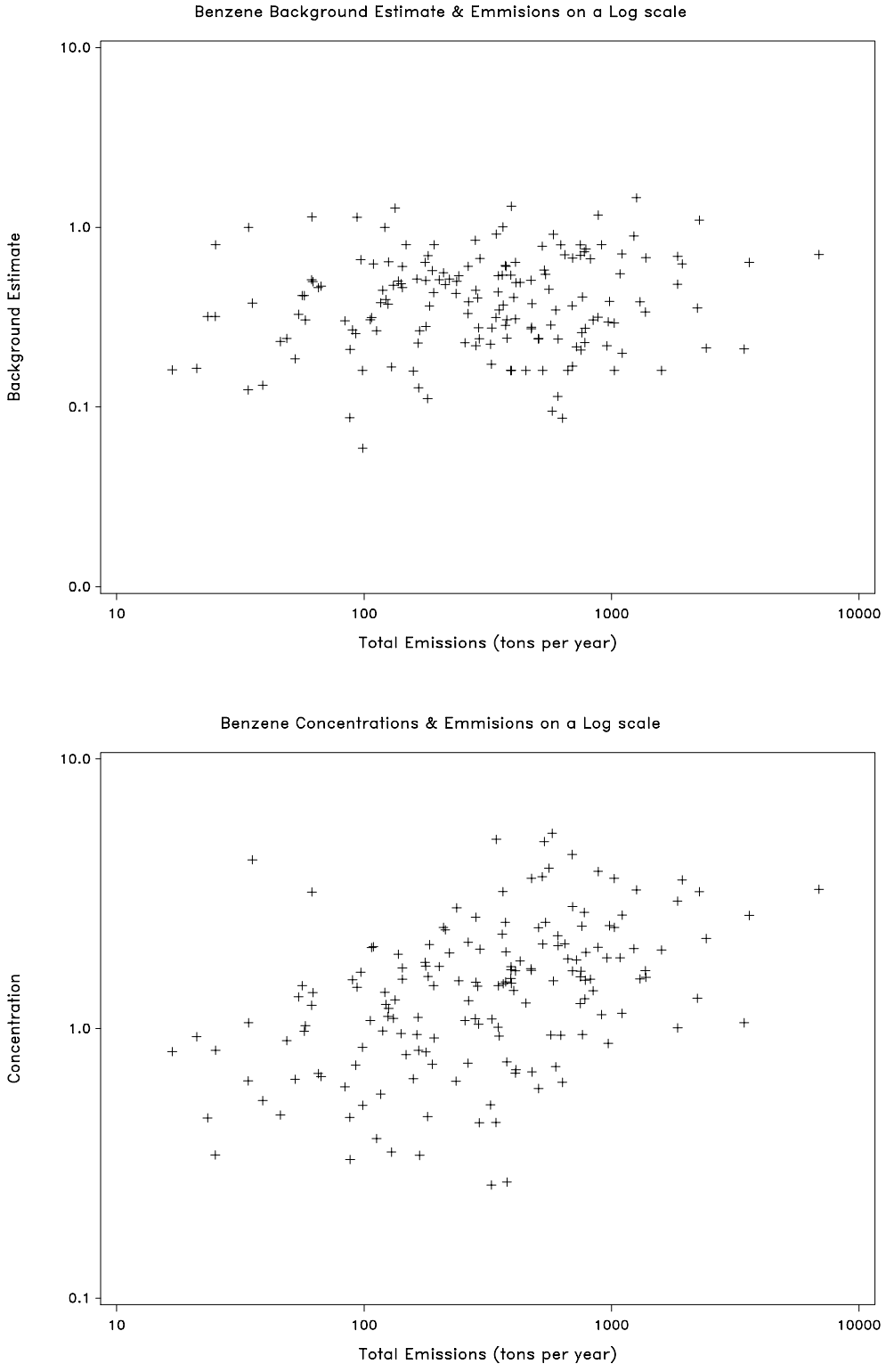


Figure 3.8 Relationship between benzene emissions and background (top), concentrations (bottom).

3.2.2 Initial Development of Explanatory Variables

As discussed in Section 2.2, many different potential explanatory variables were investigated as possible predictors of countywide background estimates. Based on preliminary analyses, it was determined that the variables ag_frac, urban_frac, land_frac, and comm_frac would be good candidates to be predictors of background. Refer to Section 2.2 for a more complete description of these variables.

Once the above variables were chosen as potential predictors, a correlation analysis was conducted to ensure that the variables were not too greatly correlated (i.e., collinearity). "Proc Corr" in SAS® was used to check for correlations between significant variables. Table 3.4 below shows the results, and indicates that it was not necessary to remove additional variables due to collinearity. Note also that although pop_density was initially not directly included in the analysis, it has a reasonably strong positive correlation with urban_frac, which suggests that pop_density is somewhat influential on the spatial variation observed in nationwide background concentrations, as one might expect.

Table 3.4 Correlation between potential predictor variables.

	Land_Frac	Ag_Frac	Urban_Frac	Comm_Frac	Pop_density
Land_Frac	1.000	0.241	-0.104	-0.007	-0.207
Ag_Frac	0.241	1.000	-0.150	-0.077	0.013
Urban_Frac	-0.104	-0.150	1.000	0.112	0.515
Comm_Frac	-0.007	-0.077	0.112	1.000	0.027
Pop_density	-0.207	0.013	0.515	0.027	1.000

Having determined that the correlations between the variables were not problematic, a primary background analysis was conducted. For this analysis, the decision was made to convert the continuous explanatory variables into categorical variables in order to remove uncertainty that arises at the extremes of the continuous variables (i.e., unusual background predictions due to extrapolation outside the observed range of the predictor variable values used to develop the regression model). Therefore, each of the four variables was divided into two categories. The categorical variables were named AgCnty, UrbanCnty, LandCnty, and CommCnty. Counties with percentages greater than or equal to 50 percent represent one category, while counties with percentages less than 50 percent represent another category. Figures 3.9 through 3.12 are plots comparing the original continuous explanatory variable (top graph), to the derived categorical variable (bottom graph).

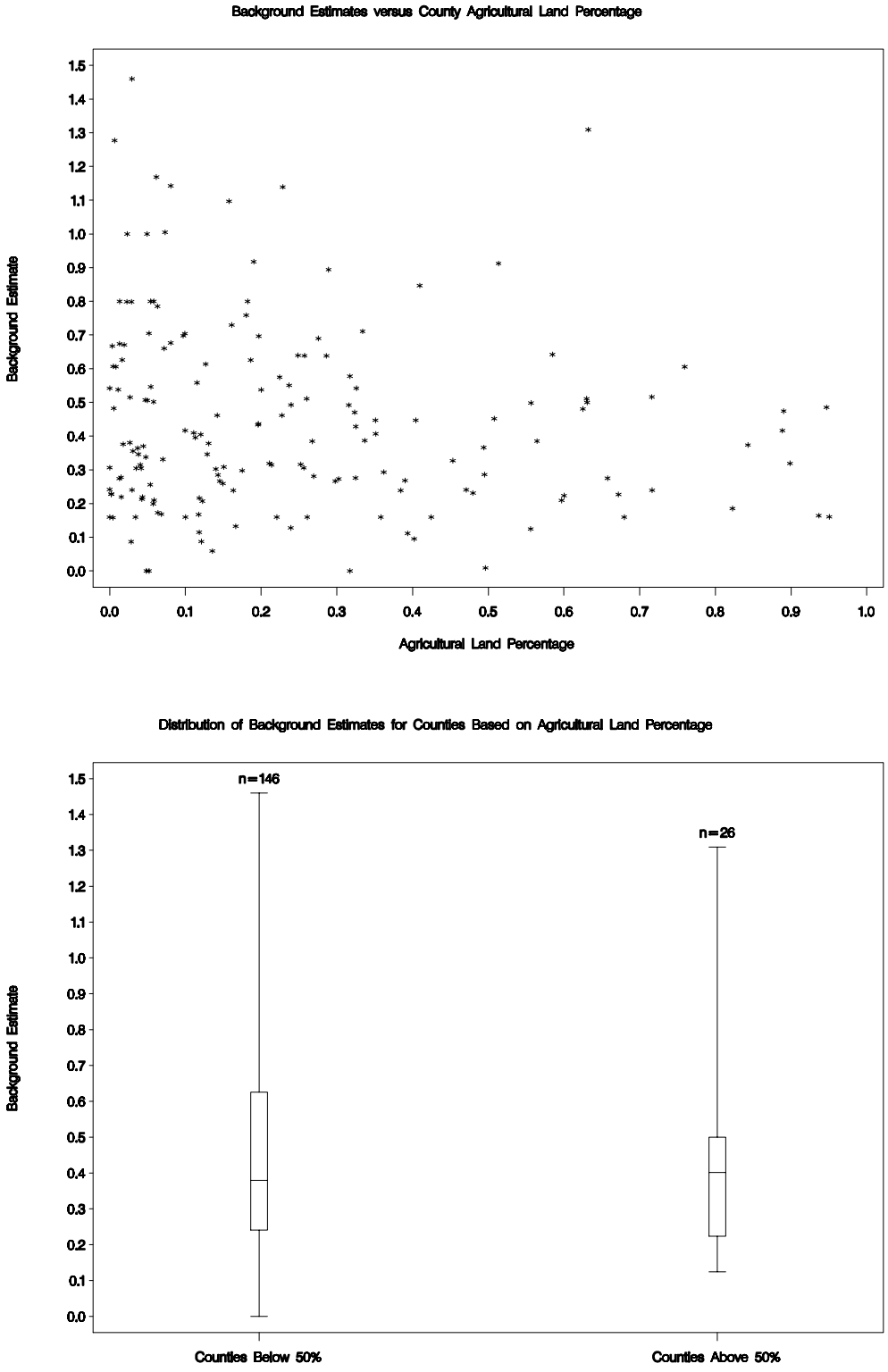


Figure 3.9 Comparison of continuous and categorical explanatory variable (ag).

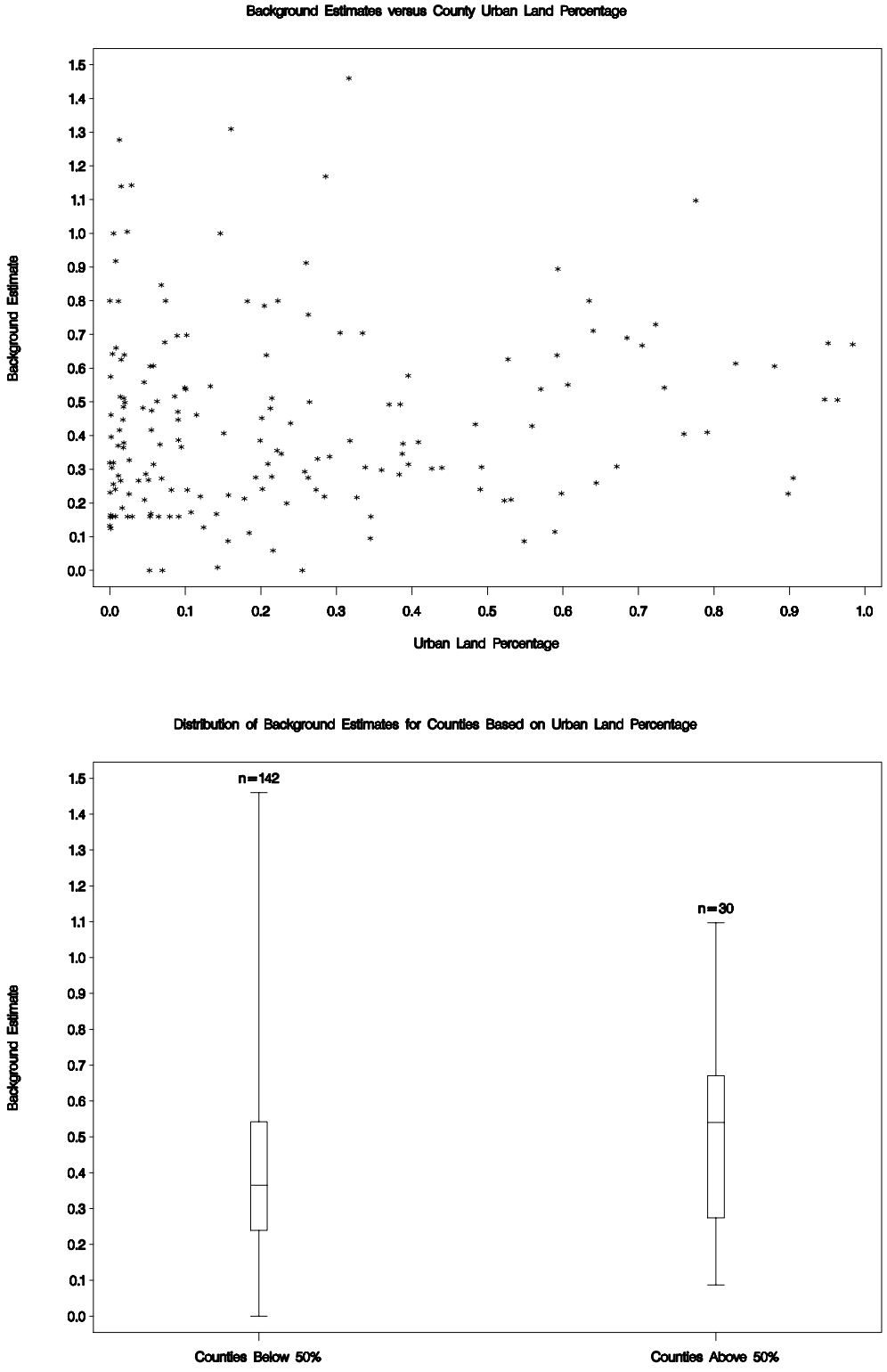


Figure 3.10 Comparison of continuous and categorical explanatory variable (urban).

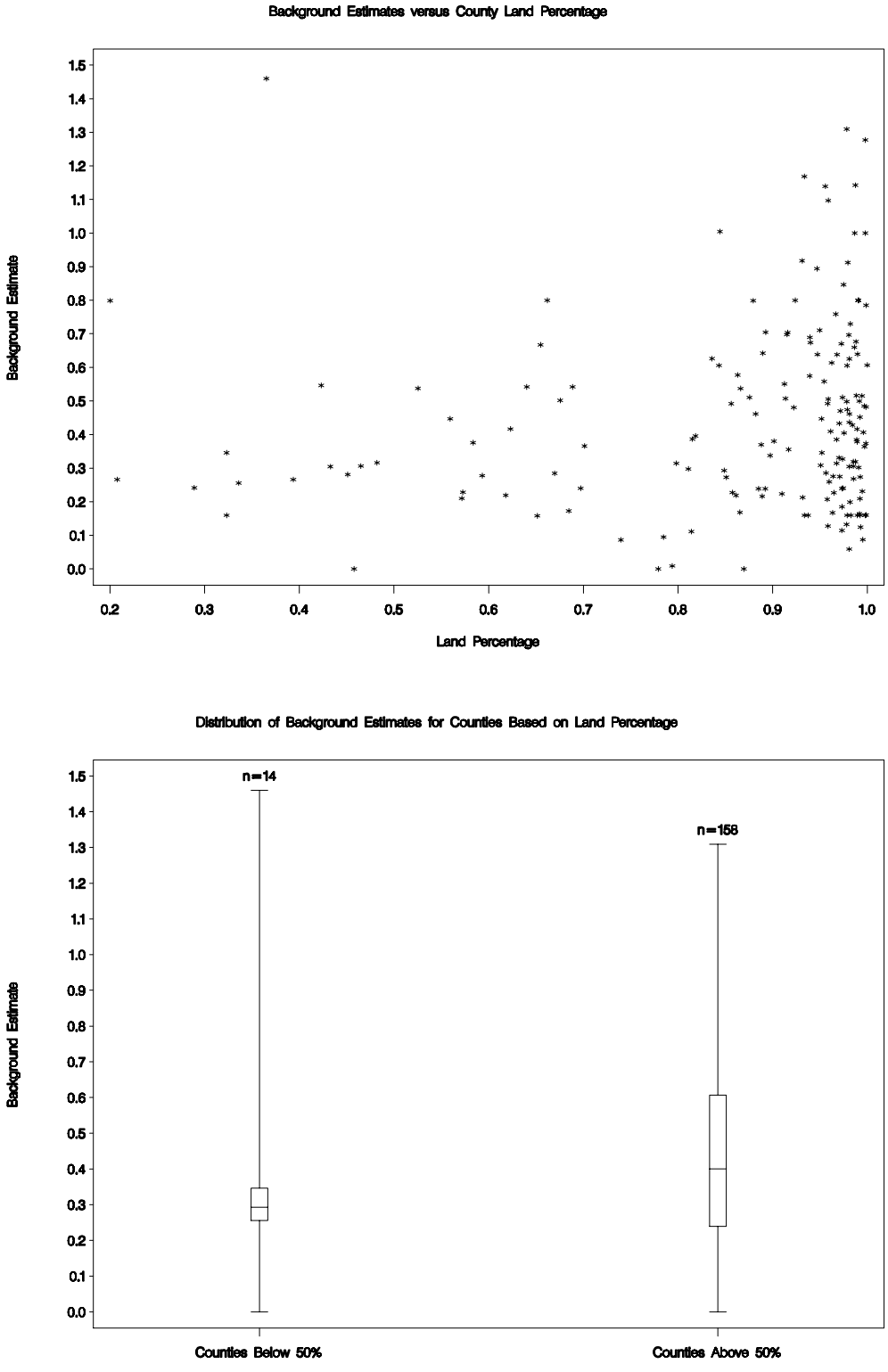


Figure 3.11 Comparison of continuous and categorical explanatory variable (land).

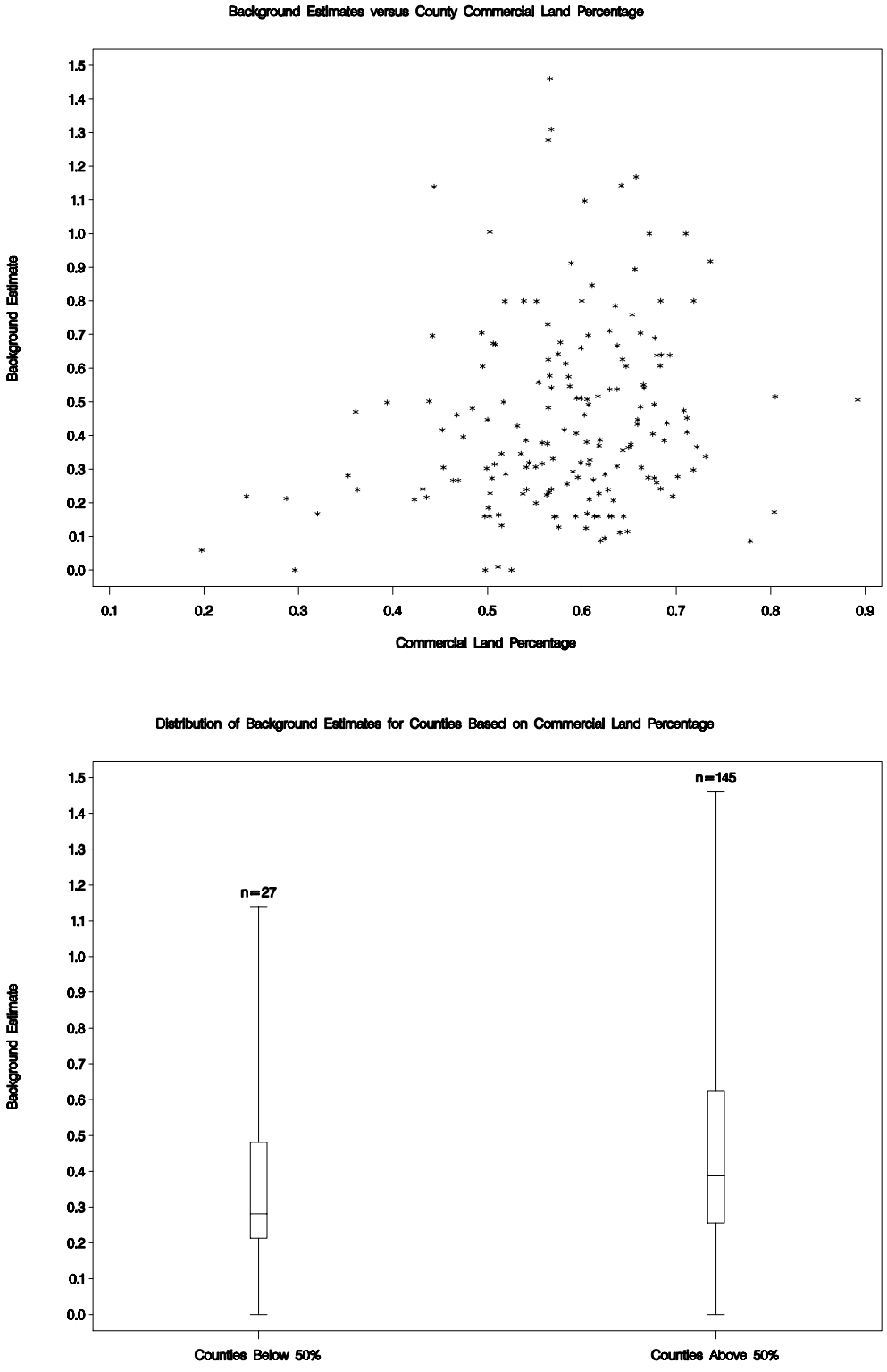


Figure 3.12 Comparison of continuous and categorical explanatory variable (comm).

3.2.3 Regression on Candidate Explanatory Variables

The four categorical variables created in Section 3.2.2 were used in a regression as inputs to Model 4. The results of the regression are an intercept value, which serves as a starting point, along with a positive or negative adjustment which is applied to the intercept depending on the values of each county's categorical variables. Therefore, since the value of each categorical variable for every county in the United States is known, it is possible to compute a background estimate for every county in the country. After following through on this approach and reviewing the results, it was decided to drop the four candidate explanatory variables in favor of relying on (the log of) pop_density as the sole predictor.

3.2.4 Population Density

Subsections 3.2.2 and 3.2.3 describe how the initial four candidate explanatory variables were used as potential predictors of background. As indicated in subsection 3.2.3, these four variables were dropped in favor of pop_density (i.e., population density). Preliminary analysis revealed that MDL limitations could have a significant effect on the results, therefore, prior to commencing the analysis using pop_density, the MDL limitations were dealt with for all pollutants, as described in Section 3.2.5. Instead of converting pop_density into a categorical variable, as with the initial four candidate explanatory variables, it was decided to take the natural log (ln) of pop_density (logPopDens) and leave it as a continuous variable. Next, a regression was run using logPopDens such that the specific form of Model 4 became:

$$\text{BackgroundEstimate} = \beta_0 + \beta_1 * \text{logPopDens} + \varepsilon,$$

where β_0 is the y-intercept, β_1 is the slope of the regression line, and ε is the residual error. As an example of how the model translates into a background estimate, consider Harris County Texas and benzene. Harris County has a population density (pop_density) of 1,086,710.29 people per km^2. Therefore, logPopDens = 13.899. The intercept (β_0) for benzene is 0.124, and the slope (β_1) is 0.024. Therefore, the background estimate for benzene in Harris County Texas = $0.124 + (0.024 * 13.899) \approx 0.46\ \mu g/m^3$.

The slope and intercept values for 13 pollutants are given in Table 4.2. By using the values from Table 4.2 along with known population density data, a background estimate can be calculated for each of the 13 pollutants for every county in the United States. In addition, the data for some pollutants were insufficient for spatial analysis through regression. In these cases, only the intercept value and the number of counties are provided in Table 4.2. This intercept was calculated in a secondary Stage 2 analysis, as summarized below, and represents the constant nationwide background estimate for the entire U.S. for the given pollutant.

3.2.5 Secondary Stage 2 Analysis

In some cases, there were not enough data for a given pollutant to calculate background based on explanatory variables. In each of these instances, a secondary analysis approach was employed to arrive at a Stage 2 background estimate. There were two issues that prevented these pollutants from being analyzed as described in Section 3.2.3 and Section 3.2.4.

The first issue is that there was not an equal enough distribution of counties among categorical variable values. For example, of the 84 counties for which there are background data for Chromium (fine), only two counties fall into the category of greater than or equal to 50 percent ag_frac. This makes it too uncertain to predict background for every county in the country based on, in this case, the explanatory variable's distribution.

The second issue is that most of the data for some pollutants were recorded as below the MDL. As part of the Stage 1 secondary analysis, an MDL/2 substitution was employed (see Section 3.1.7), and a resulting percentile of the data (often the median) was used as the background estimate. For pollutants with a large amount of below MDL data, this resulted in most counties having the same background estimate. Figure 3.13 illustrates the low variability of background estimates for Chromium (fine), a representative example of this issue. The low variability of background estimates again makes it unreasonable to predict background nationwide based on, in this case, the response variable's distribution.

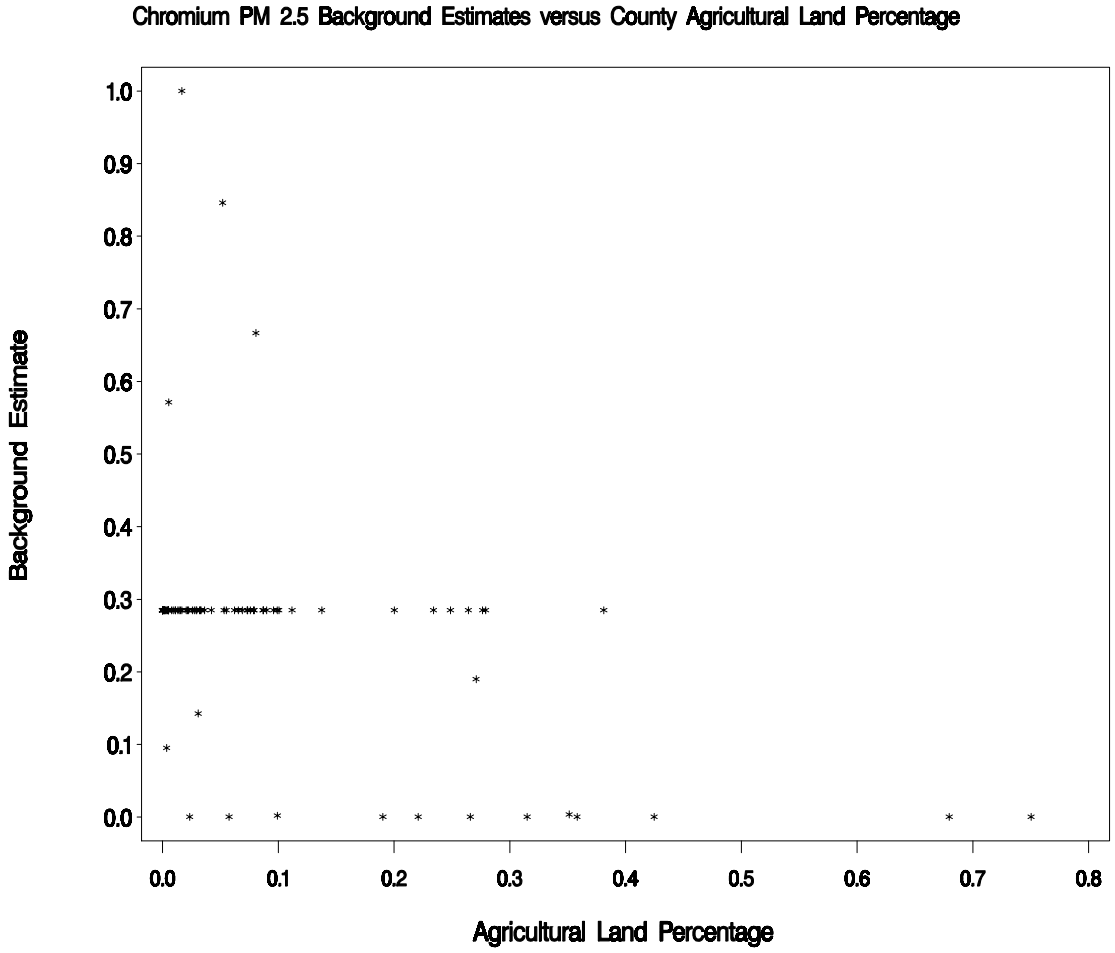


Figure 3.13 Distribution of background estimates by county agricultural land percentage.

A secondary analysis approach was employed to deal with the pollutants that were affected by either or both of the above described issues. The approach first dealt with any MDL issues for the pollutants in question. The first step was to calculate the percentage of the data at each site that were less than or equal to the MDL. The minimum and maximum MDL at each site were identified as well. If the percentage of the data that were below the MDL at any given site was greater than or equal to 50 percent, the background estimate for that site was set to zero. In addition, if the background estimate equaled one-half the minimum or the maximum MDL for the site, the estimate was also set to zero. The reason for setting the MDL to zero in these cases is that data below the MDL are interpreted as not reliably detected. This implies that there is not enough of a given pollutant present within a sample to be able to quantify a non-zero concentration. This interpretation is consistent with the derivation and statistical interpretation of MDLs, i.e., that measurements less than the MDL do not provide sufficient evidence of non-zero concentration. Thus, a decision was made to set the estimate to zero. This decision

also serves to protect against the case of estimating higher (biased) nationwide background concentrations simply because of MDL/2 substitution and imprecise monitoring technologies.

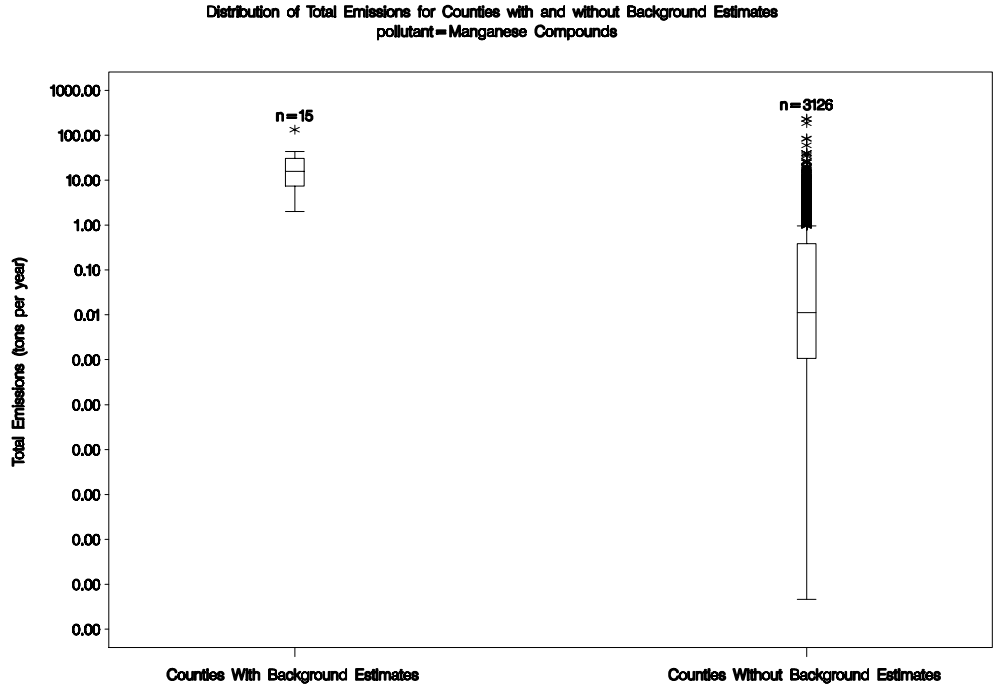
The next step was to recalculate the countywide background estimates from these revised site estimates. Once this was done, the best way to deal with the lack of spatial distribution in the data, given the limited resources of the project, was simply to take the median of all available recalculated Stage 1 background estimates as the constant nationwide estimate. For most of the pollutants addressed in this manner, due to the MDL issue, this resulted in a nationwide background estimate of zero.

Three of the pollutants involved in the secondary Stage 2 analysis, however, had non-zero background estimates (i.e., medians). Two of these three pollutants, ethylene oxide (median = 0.077) and manganese (coarse) (median = 0.014), were investigated one step further to compare emissions in the counties providing data with emissions in all remaining counties. The reason for the continued investigation with these pollutants is that both ethylene oxide (2 counties) and manganese (coarse) (15 counties) had a limited number of counties providing data. Furthermore, it was assumed that the reason for monitoring in these counties was that significant sources of emissions were present in each of the counties.

Figure 3.14 plots total emissions for ethylene oxide (top graph) and for manganese (bottom graph). Note that the emissions data for manganese are not broken down into PM 2.5 and coarse, but are for total manganese. As shown in Figure 3.14, the counties providing data for both ethylene oxide and manganese are in the upper tail of the distribution of total emissions. The mean total emissions for counties with background estimates is 3.09 tons per year and 25.3 tons per year for ethylene oxide and manganese, respectively, while the mean total emissions for counties without background estimates is 0.42 tons per year and 1.06 tons per year for ethylene oxide and manganese, respectively. Therefore, it may not be reasonable to predict non-zero nationwide background estimates for these two pollutants based solely on a very limited number of county estimates that appear to continue to be influenced by high emissions (despite the methods used for Stage 1 and the further adjustments described in Section 3.2.1). As a result, the background estimate (intercept in Table 4.2) for both ethylene oxide and manganese (coarse) was set to zero.

The third pollutant that was estimated to have a non-zero background was carbon tetrachloride. Carbon tetrachloride was initially analyzed using the primary stage 2 analysis (i.e., regression with population density). The results of that regression, however, were counter-intuitive in that the slope was negative (i.e., background estimates decreases as population density increases). In addition to being counter-intuitive, the negative slope (-0.012) was not statistically significant (p=0.142). Therefore, the decision was made to employ the secondary Stage 2 analysis for carbon tetrachloride as well. The result was a uniform nationwide carbon tetrachloride background estimate of 0.027. Unlike ethylene oxide (2 counties) or manganese (coarse) (15 counties), carbon tetrachloride had 105 counties contributing to this nationwide estimate. See Table 4.2 for the results of the secondary Stage 2 analysis, along with the results of the analysis described in Section 3.2.4.

NOTE: Total emissions of zero have been replaced by 0.1E−9 in order to plot on a log scale.



NOTE: Total emissions of zero have been replaced by 0.1E−9 in order to plot on a log scale.

Figure 3.14 Total emissions for counties with and without background estimates.

4.0 RESULTS

This section presents the suite of results obtained to date for this project. Section 4.1 provides the Stage 1 results. The associated data analysis approach of Stage 1 is described in further detail above in Section 3.1. Section 4.2 provides the Stage 2 results, which amount to the final results sought for satisfying the goals of the project. The associated data analysis approach of Stage 2 is described in further detail above in Section 3.2.

4. 1 Stage 1 Results

This section presents the results of the Stage 1 approach, as applied to a combined database of archive, PAMS, and pilot monitoring data collected during 1995 through 2002. These descriptive results, namely countywide annual average background estimates for many of the 33 urban priority HAPs, are provided as inputs for Stage 2 of the project. Subsection 4.1.1 below summarizes the actual Stage 1 results to date. This is followed by subsection 4.1.2, which provides further discussion addressing a number of relevant issues.

4.1.1 Summary of Stage 1 Results

Recognize that a large volume of results is generated on application of the Stage 1 approach to the database, summarized in Tables 2.1a and 2.1b. As such, it is useful to condense all the individual outputs down to a more convenient reporting summary. The following describes the summary of Stage 1 results for this project. Appendix A provides the more detailed results, including uncertainty characterization, for every combination of pollutant by county that was modeled.

Table 4.1 provides an overall results summary. Each row of the table presents Stage 1 results for an individual HAP, with the first column of the table indicating compound class. In an effort to remain consistent with the approach of the ASPEN model, metals results are provided for both the fine and coarse size fractions (see discussion on metals data in subsection 4.1.2 below). Following compound class and pollutant, the next two columns of Table 4.1 offer information on the breadth of results (i.e., number of sites and counties). The next column gives the nationwide background estimate used in the 1996 NATA, which may be of interest for comparison. Next, the various columns provided under the heading of "Variability of Background Estimates Across Counties" summarize the distribution of background estimates obtained across all the counties with sufficient data. This information indicates the degree of spatial, or county-to-county, variability in background for each pollutant. However, one should caveat these results based on the uncertainty information provided in the more detailed tables of Appendix A.

From Table 4.1, observe that at least one background result is obtained for all three carbonyls, all sixteen volatile organic compounds (VOCs), seven of eight metals (fine), five of eight metals (coarse), and one of seven semi-volatile organic compounds (SVOCs). In some cases, the results span fewer than 20 counties. It is expected that at least 20 county estimates, and likely more, will be required in order to pursue Stage 2 for a given HAP (see discussion on insufficient data in subsection 4.1.2 below).

To better understand Table 4.1's information, consider a discussion of the acetaldehyde results. A total of 230 monitoring sites provide acetaldehyde background estimates. These site-specific results are combined to form 107 distinct countywide estimates. The 107 countywide background estimates for acetaldehyde have a mean and standard deviation of 0.5551 $\mu g/m^3$ and 0.4744 $\mu g/m^3$, respectively. Individual countywide estimates range from a minimum of 0.0000 $\mu g/m^3$ to a maximum of 3.1267 $\mu g/m^3$, with a median estimate of 0.4832 $\mu g/m^3$ across counties with sufficient data.

Next, observe that there may be a number of ways to evaluate the reasonableness of the results in Table 4.1. For example, the table's 1996 NATA column provides one benchmark for comparison. Focusing on benzene, the updated 1999-based mean and median background estimates roughly match the 1996 estimate. This might be viewed as an affirmation of the benzene results considering that model-to-monitor evaluations of 1996 ASPEN predictions were generally favorable for benzene relative to the other modeled HAPs. Furthermore, the 1996 NATA background estimate is slightly higher than the updated 1999-based median estimate, which agrees with the slight downward trend observed in the case study example presented in subsection 3.1.6 (the 1999-based mean is nearly the same as the 1996 NATA background).

Another approach to evaluating the validity of Table 4.1's results is to compare results across pollutants. For example, more than one study has suggested that, in the absence of short-term source-oriented influences, typical ambient acetaldehyde to formaldehyde ratios are approximately 0.7 (e.g., "Air Toxics Monitoring Data: Analyses and Network Design Recommendations," Battelle's revised draft report under contract to LADCO, October 2001). Observe that the acetaldehyde to formaldehyde ratio for the mean and median background estimates of Table 4.1 is 0.713 and 0.721, respectively. Obviously, these ratios compare quite favorably to a benchmark of 0.7, and note that similar ratios between these two pollutants are observed for the various percentiles of Table 4.1, as well. Keeping in mind that each pollutant's background estimates are obtained separately from the others, this comparison provides some support for both the acetaldehyde and formaldehyde background results.

Other evaluations based on different types of comparisons are certainly a possibility. Likewise, many other characterizations and interpretations can be offered based upon further study of the Table 4.1 results. (See discussion on further interpretation in subsection 4.1.2 below.) Ideally, these results will be input into the Stage 2 analyses for extrapolation to obtain nationwide background estimates. This may not be possible for those pollutants with no results or those with results spanning very few counties. (See discussions on worst-case scenario and insufficient data in subsection 4.1.2 below.) Appendix A provides more detailed results for every pollutant by county combination modeled.

Table 4.1 Summary of Stage 1 results for background (based on archive, PAMS, and pilot monitoring data). [1]

Class	Pollutant	Number of Sites	Number of Counties	1996 NATA	Variability of Background Estimates Across Counties						
					Mean	Standard Deviation	Min	25th Percentile	Median	75th Percentile	Max
Carbonyl	ACETALDEHYDE	230	107	0	0.5551	0.4744	0.0000	0.2897	0.4832	0.7008	3.1267
	ACROLEIN	45	18	0	0.1423	0.1231	0.0000	0.0550	0.1146	0.1524	0.4620
	FORMALDEHYDE	244	109	0.25	0.7782	0.5894	0.0000	0.3665	0.6706	1.0130	3.0853
VOC	1,1,2,2-TETRACHLOROETHANE	114	59	0	0.3626	0.6110	0.0137	0.0686	0.0686	0.3431	3.4500
	1,2-DIBROMOETHANE	181	97	0.0077	0.3802	0.6462	0.0230	0.0384	0.1921	0.3072	3.8500
	1,2-DICHLOROETHANE	259	115	0.061	0.1775	0.2844	0.0000	0.0405	0.0809	0.1821	2.0000
	1,2-DICHLOROPROPANE	237	106	0	0.2110	0.3127	0.0127	0.0462	0.1216	0.2311	2.3000
	1,3-BUTADIENE	260	104	0	0.1322	0.2379	0.0000	0.0221	0.1007	0.1524	2.2122
	ACRYLONITRILE	54	34	0	0.1719	0.1960	0.0000	0.0651	0.0651	0.2278	0.8135
	BENZENE	479	177	0.48	0.5067	0.5665	0.0000	0.2396	0.3854	0.6138	4.7919
	CARBON TETRACHLORIDE	274	107	0.88	0.5027	0.2833	0.0916	0.3146	0.4698	0.6228	1.5500
	CHLOROFORM	323	134	0.083	0.2034	0.2707	0.0000	0.0586	0.1220	0.2441	1.5624
	CIS 1,3-DICHLOROPROPENE	28	13	0	0.3628	0.3621	0.0295	0.1134	0.2268	0.6805	1.2703
	TRANS 1,3-DICHLOROPROPENE	28	13	0	0.3931	0.3466	0.0363	0.2495	0.2495	0.6805	1.2794
	ETHYLENE OXIDE	6	2	0	0.1394	0.0214	0.1243	0.1243	0.1394	0.1545	0.1545
	METHYLENE CHLORIDE	342	145	0.15	0.3812	0.3821	0.0000	0.1216	0.2049	0.5346	1.7368
	TETRACHLOROETHYLENE	326	135	0.14	0.3772	0.4654	0.0000	0.1665	0.2374	0.3391	3.4000
	TRICHLOROETHYLENE TCE	273	112	0.081	0.2769	0.3896	0.0000	0.0984	0.1612	0.2687	2.7000
	VINYL CHLORIDE	282	124	0	0.1772	0.1914	0.0000	0.0256	0.1144	0.2556	1.3000
Metal (fine)	ARSENIC	50	19	0	0.2943	0.5906	0.0004	0.0010	0.0010	0.0296	1.7500
	BERYLLIUM	NO AVAILABLE DATA									
	CADMIUM	2	1	0	0.0015	.	0.0015	0.0015	0.0015	0.0015	0.0015
	CHROMIUM	127	84	0	0.2625	0.1570	0.0005	0.2850	0.2850	0.2850	1.0000
	LEAD	127	84	0	0.0608	0.1642	0.0000	0.0300	0.0300	0.0300	1.0000
	MANGANESE	127	84	0	0.2428	0.1546	0.0000	0.2600	0.2600	0.2600	1.0000
	MERCURY	26	15	0.0015	0.0010	0.0000	0.0010	0.0010	0.0010	0.0010	0.0010
	NICKEL	127	84	0	0.0471	0.1223	0.0003	0.0250	0.0250	0.0250	0.7500
Metal (coarse)	ARSENIC	26	15	0	0.0010	0.0000	0.0010	0.0010	0.0010	0.0010	0.0010
	BERYLLIUM	NO AVAILABLE DATA									
	CADMIUM										
	CHROMIUM	26	15	0	0.0007	0.0014	0.0000	0.0000	0.0000	0.0005	0.0052
	LEAD	26	15	0	0.0012	0.0015	0.0000	0.0000	0.0010	0.0010	0.0045
	MANGANESE	26	15	0	0.0153	0.0098	0.0030	0.0063	0.0141	0.0210	0.0341
	MERCURY	NO AVAILABLE DATA									
	NICKEL	26	15	0	0.0005	0.0007	0.0000	0.0003	0.0005	0.0005	0.0028
SVOC [2]	HEXACHLOROBENZENE	7	2	0.000093	0.0382	0.0036	0.0357	0.0357	0.0382	0.0408	0.0408

1 Refer to Appendix A for a detailed summary of background estimates by pollutant and county, including uncertainty characterization.
2 No data were available for all other SVOC pollutants.

Although Table 4.1 provides results across a number of sites and counties, which give some sense of the expansiveness of the Stage 1 results, it does not provide an indication of the spatial representativeness of these results. To this end, Figure 4.1 completes the results summary begun by Table 4.1, providing a Geographical Information Systems (GIS) summary of results. Figure 4.1 gives the results specific to benzene. Other corresponding figures of results for the remaining HAPs that were studied are not provided in this draft of the report, but could be provided in future versions. Each shaded polygon within Figure 4.1 represents a distinct county from within the 48 conterminous United States for which a benzene background estimate is obtained. The legend of the figure indicates benzene's specific range of estimated background concentrations, in $\mu g/m^3$, using percentile-based cut-offs (i.e., minimum to 25th, 25th to 50th, 50th to 75th, 75th to 1 $\mu g/m^3$, and 1 $\mu g/m^3$ to maximum). Note that the legend's specific breakpoint at 1 $\mu g/m^3$, between the 75th percentile and maximum benzene background estimate, was added to the figure to provide additional resolution for the purposes of interpretation. A comparison of unshaded areas versus shaded counties provides an indication of the spatial representativeness of results.

The reason for using percentile-based cut-offs in the figure is to provide the greatest amount of visual content in the resulting countywide shadings. Another approach would be to determine the color shading legend according to equally-spaced concentration ranges with respect to the overall range of results. This may result in a different interpretation of results by the reader.

As expected, Figure 4.1 indicates that California, Texas, and parts of the North-East and East Coast are relatively well represented with results. Also as expected, several states are not represented at all, with Regions 7, 8, and 10 being particularly sparse on results. Maybe not as expected, Minnesota's counties are rather well represented. To date, an explanation for this unexpected result has not been obtained by Battelle. In general, no obvious trends are discernible according to the shading scheme provided by the figure's legend.

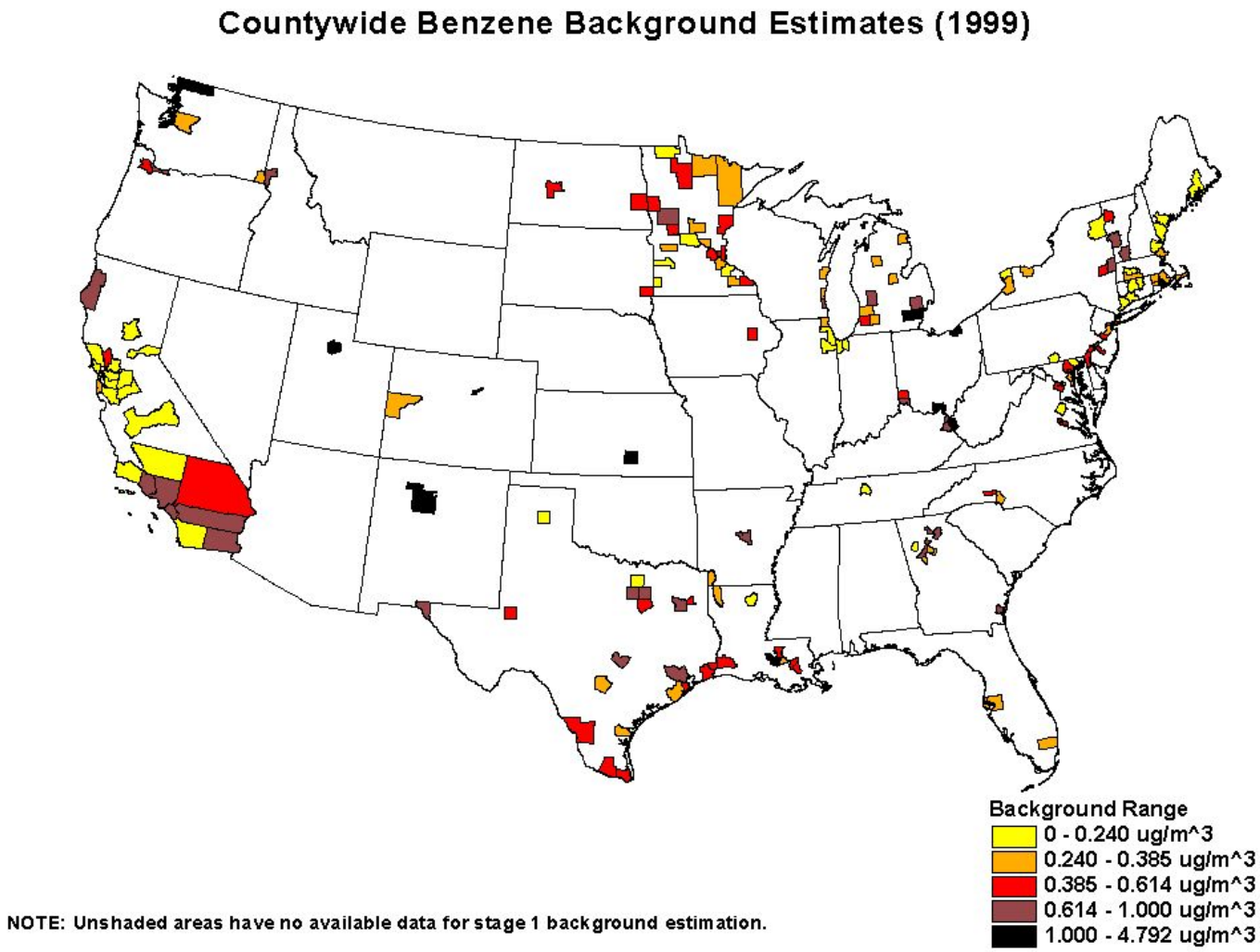


Figure 4.1 GIS summary of Stage 1 results for benzene background.

4.1.2 Further Discussion of Stage 1 Results

A number of topics or issues have come to light during the application of the Stage 1 approach to the combined archive, PAMS, and pilot monitoring database. In no particular order, the following discusses these issues in further detail.

Metals Data. As stated above, in an effort to remain consistent with the approach of the ASPEN model, metals results were pursued, and are provided, for both the fine and coarse size fractions. (Note that fine is typically taken to represent $PM_{2.5}$ and coarse to represent PM_{10} minus $PM_{2.5}$.) However, due to the limited amount of available fine and coarse data, it may be more preferable to conduct analyses on total suspended particulate (TSP) metals. The archive and pilot databases both provide a richer set of TSP data. Although, it is not clear at this time to what degree this alternative would provide for a more expansive set of metals results. Nonetheless, it may be worthy of pursuit at some point in the future.

Further Interpretation. Table 4.1, Figure 4.1, and Appendix A provide only a general summary of the Stage 1 results. Certainly more time and effort could be put into a more careful, detailed exploration and interpretation of these results. For example, county-specific GIS plots might be explored to identify within-county spatial trends or the extent to which monitors providing data for a given county are spatially or otherwise representative of that county. Likewise, various types of outlier analyses might be pursued in an effort to improve the quality of the Stage 1 results. However, it is expected that many such issues will be encountered or revealed as part of the approach to and application of Stage 2. Furthermore, any effort to pursue a more detailed understanding of the Stage 1 results would obviously require resources. The scope and budget of this project could only support the development of an approach and the provision of general results summaries such as those in subsection 4.1.1 and Appendix A.

Worst-Case Scenario. Consider an outcome under which no reliable results are obtained for Stage 2. This is a possibility given that the broadest results currently in Table 4.1 correspond to benzene, and these results span only 177 counties. Meanwhile, the approach in Stage 2 seeks to extrapolate such Stage 1 results to the approximately 3,000 counties spanning the 48 conterminous United States. Under this worst-case scenario, the Stage 1 results by themselves can provide an alternative, albeit less appealing, option for obtaining nationwide background estimates.

Specifically, some measure from the set of countywide estimates for each pollutant (e.g., mean, median, etc.) can be treated as a nationwide constant background estimate. Like the approach in the 1996 NATA, the obvious disadvantage of this approach is that it does not directly address spatial variability in background concentrations. However, spatial variability can still be addressed indirectly by using a measure of the variation between different background estimates (e.g., between-county standard deviation) to explicitly quantify the uncertainty associated with using a constant background for ASPEN. Furthermore, unlike the approach in the 1996 NATA, this approach relies exclusively on temporally relevant monitoring data. For these reasons, even under this worst-case scenario, an improvement over the 1996 NATA approach to background estimation can still be made based on Stage 1 results alone.

***Insufficient Data*.** Ideally, available air toxics monitoring data would have provided some background information for every HAP on the urban list of 33. As Table 4.1 indicates, this is not the case, as a number of HAPs have no Stage 1 results due to insufficient data. Obviously, Stage 2 cannot be applied, at least not directly, in these cases. Nonetheless, background estimates are still needed for these HAPs. A number of possibilities should be explored. One possibility is to conduct an updated literature review from that of the 1996 NATA. This could yield some form of background estimate for the HAPs in question. It might also provide other information that serves to confirm or refute the results of Table 4.1. Another possibility, likely related to any literature review findings, would be to use known or typical ambient ratios of HAPs to infer a missing pollutant's background from that of a non-missing pollutant. The above suggestions are by no means exhaustive. This discussion merely serves to provide an initial stimulus toward further investigation into alternative methods of estimating background for those pollutants with insufficient monitoring data. The current scope and budget of this project cannot support any detailed investigation toward this end.

4.2 Stage 2 Results

This section presents the updated results provided by the Stage 2 analysis. These results build on those that were presented in Section 4.1.1. Using the results from Section 4.1.1 and the explanatory variables (i.e., population density) from Section 2.2, background estimates were obtained for every county in the United States for all but eight of the 33 urban HAPs. For those eight, mostly SVOCs, no data were available in Stage 1 of this project.

4.2.1 Summary of Stage 2 Results

The results of the Stage 2 analysis are presented in Table 4.2. Each row in the table presents Stage 2 results for an individual HAP, with the first column of the table indicating compound class. The second column indicates the HAP. The third column is the number of counties included in the analysis for the respective pollutant. The fourth column is the R^2 from the regression for each of the pollutants that were analyzed as described in Section 3.2.4. The fifth column is the intercept (along with the associated p-value, if applicable) or, in the case of those pollutants analyzed in the secondary Stage 2 analysis, the median of the recalculated countywide background estimates (refer to Section 3.2.5). Note that there is no p-value for each of these pollutants, nor is there a slope or its associated p-value. In these cases, the background estimate is uniform nationwide. Note that many of the pollutants from the primary Stage 2 analysis have negative intercept values. The sixth column is the slope of the regression line and the associated p-value.

Table 4.2 can be used to calculate background estimates for any county in the United States. The following three examples demonstrate using Table 4.2 to calculate a background estimate. Note that due to the number of significant digits used in calculations, the examples given here are approximate.

- Consider Acetaldehyde, which was analyzed using the primary Stage 2 approach as described in Section 3.2.4. The intercept for Acetaldehyde is 0.460 $\mu g/m^3$ and the slope is 0.005 $\mu g/m^3$. The background estimate for a given county for Acetaldehyde is calculated by multiplying the slope (0.005) times the natural log of pop_density,

and adding the intercept (0.460). For example, Butler County, PA, has a population density of 21,987.87 people per km^2. Therefore, logPopDens = ln(21,987.87) = 10.00. Thus, the background estimate for Acetaldehyde in Butler County, PA, is (0.005 * 10.00) + 0.460 = 0.51 $\mu g/m^3$.

- Next, consider Carbon Tetrachloride, which was analyzed using the secondary Stage 2 approach as described in Section 3.2.5. In this case, the intercept given in Table 4.2 is the median of the revised Stage 1 background estimates. The result is a background estimate of 0.270 $\mu g/m^3$ for Carbon Tetrachloride that applies uniformly to every county in the United States.

- Finally, consider Cadmium (coarse). No Stage 1 data were available for Cadmium; however, a background estimate was inferred from similarly behaved pollutants. For example, all of the metals for which data were available resulted in a background estimate of zero. Therefore, it seems reasonable to infer that Cadmium (coarse) also has a background estimate of zero that applies uniformly to every county in the United States.

Refer to Section 3.2.4 for an additional example of calculating background based on Table 4.2.

Note that no data were available for 6 of the 33 HAPs, and are, therefore, not presented in the table. We believe that the results presented in Table 4.2 are an accurate representation of background estimates based on the data available; however, one should caveat these results based on a relatively small amount of source data, much of which have large uncertainties. For example, in the case of cadmium (fine), only a single county contributed to the nationwide background estimate. In fact, even in the case of benzene, which had more contributing counties than any other pollutant, there were only 172 out of the 3,141 counties in the United States (less than 5.5 percent) that contributed data.

Figure 4.2 is a GIS map of the Stage 2 results for benzene. Note that every county in the 48 conterminous United States is represented in the figure. In addition, note that the results of the regression on (the log of) population density as shown in Table 4.2 result in some identifiable patterns on the map. For example, it is possible to pick out a number of cities (e.g., Chicago, Indianapolis, Atlanta, etc.) due to their higher background estimates (a darker shade on the map), due to the high population density in those areas.

4.2.2 Further Discussion of Stage 2 Results

The results of the Stage 2 analysis should be viewed as an estimate of background. The results represent a likely background level for any given county, given that county's population density. Obviously, many other factors could potentially influence background levels; therefore, it is possible that a given county's background estimate does not reflect the actual background level in that county.

Despite that caveat, the Stage 2 results are a step forward from the 1996 NATA for two reasons. First, only temporally relevant data contributed to Stage 2 results. Therefore, Stage 2

results can be viewed as a more accurate estimate of current background levels than the 1996 NATA. Second, spatially varying background estimates were obtained for 13 pollutants. These estimates vary spatially depending on population density, and allow for a potentially more precise estimate than is possible when restricted to a uniform nationwide estimate.

Table 4.2 Summary of Stage 2 results for background

Class	Pollutant	Number of Counties	R^2 from Regression	Intercept (p-value)	Slope (p-value)
Carbonyl	Acetaldehyde	103	0.002	0.460 (<.001)	0.005 (0.620)
	Acrolein[1]	18		0.000	
	Formaldehyde	107	0.174	-0.053 (0.762)	0.074 (<.001)
VOC	1,1,2,2-Tetrachloroethane	56	0.038	-0.040 (0.525)	0.009 (0.152)
	1,2-Dibromoethane	94	0.049	-0.030 (0.203)	0.005 (0.033)
	1,2-Dichloroethane	112	0.038	-0.020 (0.454)	0.005 (0.038)
	1,2-Dichloropropane	103	0.017	-0.012 (0.612)	0.003 (0.185)
	1,3-Butadiene	101	0.162	-0.127 (0.002)	0.015 (<.001)
	Acrylonitrile[1]	32		0.000	
	Benzene	172	0.062	0.124 (0.125)	0.024 (0.001)
	Carbon Tetrachloride	105		0.270	
	Chloroform	132	0.002	0.027 (0.261)	0.001 (0.570)
	Cis 1,3-Dichloropropene[1]	11		0.000	
	Trans 1,3-Dichloropropene[1]	11		0.000	
	Ethylene Oxide[1]	2		0.000	
	Methylene Chloride	143	0.101	-0.187 (0.102)	0.042 (<.001)
	Tetrachloroethylene	132	0.096	-0.123 (0.040)	0.020 (<.001)
	Trichloroethylene TCE	109	0.007	0.031 (0.442)	0.003 (0.004)
	Vinyl Chloride	121	0.080	-0.068 (0.056)	0.011 (0.002)
Metal (fine)	Arsenic[1]	19		0.000	
	Beryllium[2]			0.000	
	Cadmium[1]	1		0.000	
	Chromium[1]	84		0.000	
	Lead[1]	84		0.000	
	Manganese[1]	84		0.000	
	Mercury[1]	15		0.000	
	Nickel[1]	84		0.000	
Metal (coarse)	Arsenic[1]	15		0.000	
	Beryllium[2]			0.000	
	Cadmium[2]			0.000	
	Chromium[1]	15		0.000	
	Lead[1]	15		0.000	
	Manganese[1]	15		0.000	
	Mercury[2]			0.000	
	Nickel[1]	15		0.000	
SVOC[3]	Hexachlorobenzene[1]	2		0.000	

1 Result obtained from secondary analysis. Due in large part to MDL issues, the data for these pollutants were insufficient for spatial analysis through regression.
2 No data were available on these pollutants, result implied from pollutants of similar behavior.
3 No data were available for 2,3,7,8-tetrachlorodibenzo-p-dioxin, coke oven emissions, hydrazine, polycyclic organic matter, polychlorinated biphenyls, nor quinoline. Background was not estimated for these pollutants.

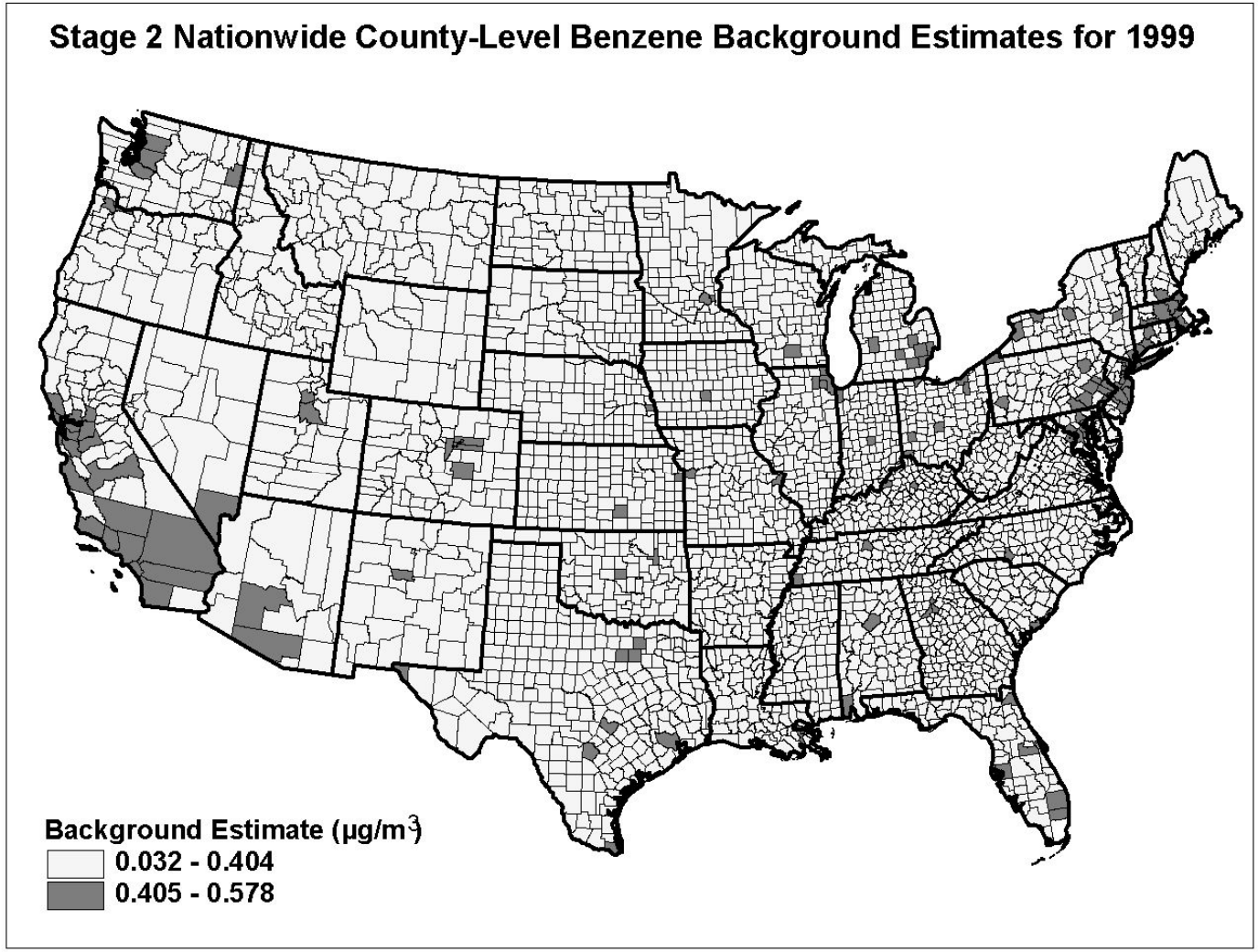


Figure 4.2 GIS summary of Stage 2 results for benzene background.

5.0 REMAINING ISSUES AND PATH FORWARD

There are several issues that remain to be resolved. All of these issues stem from insufficient data. The first of these issues is that eight HAPs have no background estimate. Secondly, 18 HAPs have a uniform background estimate (i.e., no spatial variability). Finally, all of the background estimates in Stage 2 of the analysis are based on a limited dataset.

In order to improve upon the existing results, the first step on the path forward should be to build a more comprehensive, more diverse dataset. This would result in a larger number of counties contributing to the result, and would allow for each of the 33 HAPs to be analyzed with a single method. Ideally, the data would come from a single lab with a constant MDL, however, that may not be a practical option.

APPENDIX A:

DETAILED SUMMARIES OF STAGE 1 RESULTS FOR BACKGROUND

(Background Estimates Given in $\mu g/m^3$)

Appendix A: Detailed Summaries of Stage 1 Results for Background (background estimates given in µg/m^3)

Table A.1 Summary of Countywide Acetaldehyde Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	44	1	0.2702	0.0004	Minnesota	Itasca	21	1	0.2496	0.1147
Arkansas	Pulaski	41	1	0.4350	0.0936	Minnesota	Koochiching	140	2	0.4369	0.1597
California	Alameda	147	1	0.2392	0.0590	Minnesota	Lake	40	1	0.6351	0.3175
California	Butte	156	1	0.2042	0.0731	Minnesota	McLeod	45	1	0.8504	0.4252
California	Contra Costa	233	3	0.2285	0.0488	Minnesota	Morrison	55	1	0.5531	0.2765
California	Fresno	148	1	0.1747	0.0661	Minnesota	Nicollet	23	1	0.5513	0.2756
California	Humboldt	19	1	0.9008	0.4504	Minnesota	Olmsted	59	1	0.2937	0.1468
California	Imperial	166	1	0.4892	0.0642	Minnesota	Otter Tail	97	2	0.5855	0.2140
California	Kern	333	2	0.4351	0.0465	Minnesota	Pine	44	1	0.6477	0.3238
California	Los Angeles	1626	21	0.9973	0.0945	Minnesota	Pipestone	53	1	0.7819	0.0006
California	Orange	48	2	0.2513	0.1281	Minnesota	Ramsey	509	3	0.3563	0.0250
California	Placer	153	1	0.4486	0.0482	Minnesota	Roseau	58	1	0.5459	0.2729
California	Riverside	650	8	0.7826	0.1080	Minnesota	Sherburne	60	1	0.2468	0.1234
California	San Bernardino	158	4	0.7654	0.1442	Minnesota	St. Louis	505	4	0.3558	0.1268
California	San Diego	365	3	0.3998	0.0381	Minnesota	Stearns	55	1	0.4326	0.0696
California	San Francisco	134	1	0.0662	0.0568	Minnesota	Swift	56	1	0.6270	0.3135
California	San Joaquin	312	2	0.2518	0.0373	Minnesota	Washington	676	3	0.2244	0.0388
California	Santa Barbara	169	2	0.4777	0.1343	Minnesota	Winona	58	1	0.6161	0.3081
California	Santa Clara	157	1	0.2451	0.0574	Minnesota	Wright	44	1	0.7729	0.3864
California	Stanislaus	123	2	0.7008	0.0783	Minnesota	Yellow Medicine	49	1	0.1903	0.0585
California	Ventura	144	1	0.3923	0.0670	New Jersey	Camden	286	3	0.7496	0.0683
Colorado	Denver	25	1	2.9555	0.1402	New Jersey	Union	22	1	0.5690	0.1707
Colorado	Mesa	170	2	1.0314	0.0678	New Mexico	Sandoval	52	2	0.1621	0.0699
Connecticut	Hartford	46	2	0.9787	0.3725	New York	Bronx	409	2	0.5301	0.0873
Connecticut	New Haven	20	1	0.7502	0.3751	New York	Queens	248	2	1.4348	0.5669
Florida	Hillsborough	243	3	0.2957	0.1479	New York	Richmond	504	7	0.4780	0.1246
Florida	Pinellas	258	3	0.0000	0.0045	North Dakota	Mercer	59	1	0.3783	0.0957
Georgia	DeKalb	178	2	0.5500	0.2753	Oregon	Multnomah	234	4	0.8888	0.1932
Georgia	Fayette	200	2	0.6037	0.2469	Oregon	Washington	58	1	0.2590	0.0816
Illinois	Cook	196	2	0.4054	0.1396	Pennsylvania	Philadelphia	308	1	0.5901	0.0390
Indiana	Lake	259	4	0.3944	0.0620	Puerto Rico	Barceloneta	25	1	0.0000	0.0079
Iowa	Linn	45	1	0.4558	0.2279	Puerto Rico	San Juan	26	1	0.0000	0.0079
Louisiana	East Baton Rouge Parish	784	3	0.5893	0.2594	Rhode Island	Kent	88	1	2.0170	1.0085
Louisiana	St. Charles Parish	109	1	0.3063	0.1531	Rhode Island	Providence	1272	7	0.7811	0.1693
Louisiana	St. John the Baptist Parish	125	1	0.6522	0.0770	South Carolina	Hampton	493	2	0.0000	0.0424
Maine	Cumberland	19	1	0.5405	0.2702	South Carolina	Lexington	259	1	0.0000	0.0600
Maryland	Baltimore	283	1	0.0000	0.0901	South Carolina	Richland	517	4	0.0000	0.0300
Maryland	Baltimore city	232	2	0.4504	0.1067	Texas	Cameron	56	1	0.0000	0.0901
Massachusetts	Essex	221	2	1.0050	0.3991	Texas	Dallas	63	1	0.4832	0.1995
Massachusetts	Hampden	226	2	1.2195	0.5501	Texas	El Paso	368	2	0.7541	0.1871
Michigan	Allegan	51	1	0.8717	0.0562	Texas	Galveston	58	1	1.1952	0.0572
Michigan	Kalamazoo	336	4	0.7296	0.1659	Texas	Harris	445	4	0.4506	0.0851
Michigan	Kent	196	2	0.4329	0.0697	Texas	Jefferson	62	1	0.0000	0.0901
Michigan	Missaukee	100	2	0.4651	0.2030	Texas	Tarrant	27	1	1.1218	0.0706
Michigan	Oakland	49	1	0.5655	0.0741	Utah	Salt Lake	58	1	0.5945	0.1199
Michigan	Washtenaw	33	1	0.5655	0.2827	Vermont	Chittenden	265	2	0.5670	0.1155
Michigan	Wayne	996	8	0.6235	0.1005	Vermont	Rutland	133	1	0.5810	0.0735
Minnesota	Beltrami	54	1	0.7314	0.3657	Vermont	Windham	113	1	0.7026	0.0560
Minnesota	Clay	57	1	0.3208	0.0834	Virginia	Fairfax	119	1	0.5225	0.2612
Minnesota	Dakota	1196	7	0.2168	0.0174	Washington	Clark	19	1	0.0901	0.0450
Minnesota	Douglas	73	2	0.4414	0.1288	Washington	King	367	6	0.5681	0.0940
Minnesota	Freeborn	59	1	0.2897	0.0667	Washington	Whatcom	64	2	3.1267	1.8127
Minnesota	Goodhue	52	1	0.6216	0.3108	Wisconsin	Milwaukee	242	1	0.2898	0.0481
Minnesota	Hennepin	620	5	0.3454	0.0490						

Table A.2 Summary of Countywide Acrolein Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Maine	Cumberland	19	1	0.1146	0.0573	New York	Queens	176	2	0.4012	0.1743
Maryland	Baltimore	233	1	0.1146	0.0573	New York	Richmond	504	7	0.0010	0.0013
Maryland	Baltimore city	149	1	0.1146	0.0573	Oregon	Multnomah	182	4	0.0550	0.0149
Michigan	Kalamazoo	336	4	0.0459	0.0152	Oregon	Washington	46	1	0.0400	0.0200
Michigan	Kent	163	2	0.1398	0.0502	Pennsylvania	Philadelphia	230	1	0.1146	0.0573
Michigan	Missaukee	71	2	0.1284	0.0456	Rhode Island	Kent	80	1	0.1581	0.0791
Michigan	Oakland	2	1	0.1524	0.0762	Rhode Island	Providence	521	5	0.4620	0.1072
Michigan	Wayne	352	7	0.1341	0.0255	Virginia	Fairfax	119	1	0.1146	0.0573
New York	Bronx	421	3	0.2702	0.1162	Washington	Whatcom	29	1	0.0000	0.0015

Table A.3 Summary of Countywide Formaldehyde Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	44	1	0.0000	0.0614	Minnesota	Hennepin	678	6	0.5424	0.0523
Arkansas	Pulaski	41	1	0.7015	0.1206	Minnesota	Itasca	32	1	0.7962	0.0804
California	Alameda	147	1	0.5517	0.0453	Minnesota	Kandiyohi	32	1	0.3697	0.0005
California	Butte	156	1	0.6890	0.0659	Minnesota	Koochiching	140	2	0.2205	0.0444
California	Contra Costa	233	3	0.8691	0.2057	Minnesota	Lake	53	1	0.2082	0.0770
California	Fresno	148	1	0.7876	0.0825	Minnesota	McLeod	60	1	0.4491	0.0632
California	Humboldt	19	1	1.2282	0.6141	Minnesota	Morrison	55	1	0.2923	0.0624
California	Imperial	166	1	0.3681	0.2287	Minnesota	Nicollet	39	1	0.0381	0.0614
California	Kern	333	2	0.7871	0.0494	Minnesota	Olmsted	59	1	0.1572	0.0974
California	Los Angeles	1749	21	1.5907	0.2337	Minnesota	Otter Tail	112	2	0.4379	0.0410
California	Orange	49	2	0.7597	0.3807	Minnesota	Pine	44	1	0.3375	0.0943
California	Placer	153	1	0.9406	0.0373	Minnesota	Pipestone	53	1	0.4154	0.0785
California	Riverside	689	8	0.3481	0.0394	Minnesota	Ramsey	553	3	0.6888	0.0790
California	San Bernardino	159	4	1.4352	0.3884	Minnesota	Roseau	58	1	0.1911	0.0577
California	San Diego	365	3	0.7907	0.0339	Minnesota	Sherburne	60	1	0.1639	0.0824
California	San Francisco	134	1	0.6646	0.0404	Minnesota	St. Louis	585	5	0.3405	0.0323
California	San Joaquin	312	2	0.6036	0.0371	Minnesota	Stearns	55	1	0.3869	0.0003
California	Santa Barbara	169	2	1.1491	0.0298	Minnesota	Swift	56	1	0.8536	0.4268
California	Santa Clara	157	1	0.9107	0.0446	Minnesota	Washington	752	3	0.2784	0.0292
California	Stanislaus	123	2	1.0565	0.3095	Minnesota	Winona	58	1	0.6374	0.0769
California	Ventura	144	1	0.4091	0.0720	Minnesota	Wright	59	1	0.0000	0.0614
Colorado	Denver	25	1	2.3952	0.6590	Minnesota	Yellow Medicine	49	1	0.2432	0.1175
Colorado	Mesa	168	2	2.4128	0.9464	New Jersey	Camden	366	3	1.3596	0.3875
Connecticut	Fairfield	45	1	0.6698	0.1171	New Jersey	Union	22	1	0.9826	0.4913
Connecticut	Hartford	99	2	1.7829	0.1540	New Mexico	Sandoval	52	2	0.0715	0.0273
District of Columbia	District of Columbia	265	1	3.0853	0.0514	New York	Bronx	512	2	1.0130	0.0675
Florida	Hillsborough	243	3	1.3787	0.3629	New York	Queens	249	2	0.8270	0.1050
Florida	Pinellas	256	3	1.4645	0.3251	New York	Richmond	504	7	0.3588	0.0970
Georgia	DeKalb	258	3	0.4683	0.1191	North Dakota	Mercer	59	1	0.6552	0.0525
Georgia	Fayette	202	2	0.6706	0.2572	Oregon	Multnomah	234	4	1.2555	0.1441
Illinois	Cook	419	3	0.9901	0.2129	Oregon	Washington	58	1	0.0000	0.0333
Indiana	Lake	258	4	0.8566	0.0442	Pennsylvania	Philadelphia	579	2	1.7744	0.0460
Iowa	Linn	45	1	0.1344	0.2761	Puerto Rico	Barceloneta	25	1	1.9726	0.9863
Louisiana	East Baton Rouge Parish	974	3	0.5891	0.0477	Puerto Rico	San Juan	26	1	1.9235	0.9617
Louisiana	St. Charles Parish	109	1	1.0440	0.5220	Rhode Island	Providence	945	7	1.6001	0.1320
Louisiana	St. John the Baptist Parish	125	1	1.2475	0.0462	South Carolina	Hampton	626	2	0.4704	0.0263
Maine	Cumberland	19	1	0.6141	0.3070	South Carolina	Lexington	337	1	0.7600	0.3800
Maryland	Baltimore	461	2	0.3665	0.0521	South Carolina	Richland	571	4	0.2050	0.0985
Maryland	Baltimore city	501	4	1.0010	0.0696	Texas	Cameron	56	1	0.5685	0.0676
Massachusetts	Essex	392	2	0.5525	0.0632	Texas	Dallas	63	1	0.4421	0.2211
Massachusetts	Hampden	327	2	0.8203	0.1391	Texas	El Paso	534	3	0.4994	0.1737
Michigan	Allegan	50	1	1.1133	0.1199	Texas	Galveston	58	1	1.2122	0.1172
Michigan	Kalamazoo	336	4	0.9430	0.2255	Texas	Harris	597	5	1.4188	0.2179
Michigan	Kent	196	2	1.1629	0.1272	Texas	Jefferson	62	1	2.5726	0.1575
Michigan	Missaukee	101	2	0.3924	0.2877	Texas	Tarrant	27	1	1.5230	0.1299
Michigan	Oakland	50	1	0.8053	0.1089	Utah	Salt Lake	58	1	0.9183	0.1170
Michigan	Washtenaw	33	1	0.8228	0.2185	Vermont	Chittenden	265	2	1.4870	0.3082
Michigan	Wayne	1044	9	1.1265	0.1371	Vermont	Rutland	133	1	0.9932	0.0672
Minnesota	Beltrami	54	1	0.4385	0.0655	Vermont	Windham	113	1	0.8106	0.4053
Minnesota	Clay	57	1	0.3403	0.0878	Virginia	Fairfax	119	1	0.5650	0.0715
Minnesota	Crow Wing	36	1	0.4826	0.0732	Washington	Clark	19	1	0.0614	0.0307
Minnesota	Dakota	1314	8	0.2329	0.0199	Washington	King	364	6	0.3582	0.0266
Minnesota	Douglas	90	2	0.1019	0.0674	Washington	Whatcom	64	2	0.0000	0.0307
Minnesota	Freeborn	59	1	0.0000	0.0614	Wisconsin	Milwaukee	246	1	0.9664	0.0220
Minnesota	Goodhue	52	1	0.0000	0.0614						

Table A.4 Summary of Countywide 1,1,2,2-Tetrachloroethane Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Riverside	1	1	0.3431	0.1716	Minnesota	Douglas	57	1	0.0687	0.0343
Colorado	Mesa	163	2	0.2059	0.0728	Minnesota	Goodhue	53	1	0.0687	0.0343
Delaware	New Castle	55	1	0.0687	0.0343	Minnesota	Hennepin	427	4	0.0687	0.0172
District of Columbia	District of Columbia	169	1	0.0687	0.0343	Minnesota	Koochiching	139	2	0.0687	0.0243
Florida	Broward	52	2	0.2746	0.0971	Minnesota	Morrison	55	1	0.0687	0.0343
Florida	Hillsborough	37	3	0.0915	0.0268	Minnesota	Olmsted	55	1	0.0687	0.0343
Florida	Pinellas	514	7	0.2591	0.0552	Minnesota	Otter Tail	47	1	0.0687	0.0343
Georgia	Dawson	25	1	1.7000	0.8500	Minnesota	Pine	50	1	0.0687	0.0343
Georgia	Fayette	207	2	0.2059	0.0728	Minnesota	Pipestone	46	1	0.0687	0.0343
Georgia	Fulton	22	1	1.7000	0.8500	Minnesota	Ramsey	378	3	0.0687	0.0198
Georgia	Glynn	20	1	1.7000	0.8500	Minnesota	Roseau	46	1	0.0687	0.0343
Georgia	Hall	21	1	1.7000	0.8500	Minnesota	Sherburne	57	1	0.0687	0.0343
Indiana	Lake	277	2	0.1716	0.0607	Minnesota	St. Louis	377	3	0.0687	0.0198
Indiana	Porter	22	1	0.1716	0.0858	Minnesota	Stearns	51	1	0.0687	0.0343
Iowa	Linn	36	1	0.2059	0.1030	Minnesota	Swift	53	1	0.0687	0.0343
Kansas	Sedgwick	54	1	0.6865	0.3432	Minnesota	Washington	304	2	0.0687	0.0243
Maine	Oxford	25	1	3.4500	1.7250	Minnesota	Winona	50	1	0.0687	0.0343
Maryland	Anne Arundel	406	2	0.0687	0.0243	Minnesota	Yellow Medicine	42	1	0.0687	0.0343
Maryland	Baltimore	347	1	0.0687	0.0343	Ohio	Hamilton	25	1	0.3432	0.1716
Maryland	Baltimore city	1398	6	0.0687	0.0140	Oregon	Multnomah	277	5	0.3432	0.0768
Maryland	Harford	109	1	0.0687	0.0343	Pennsylvania	Delaware	663	3	0.0219	0.0063
Michigan	Alpena	130	5	0.3383	0.0757	Pennsylvania	Philadelphia	269	1	0.0687	0.0343
Michigan	Kent	163	3	0.8006	0.2493	Puerto Rico	Barceloneta	37	1	0.2059	0.1029
Michigan	Missaukee	87	2	0.6863	0.2713	Puerto Rico	San Juan	26	1	0.2059	0.1029
Michigan	Oakland	43	1	0.2059	0.1029	Rhode Island	Kent	51	1	0.0137	0.0069
Michigan	Washtenaw	54	2	0.6863	0.2713	Rhode Island	Providence	352	5	0.0192	0.0049
Michigan	Wayne	877	7	1.4313	0.5027	Texas	Nueces	55	1	0.6865	0.3432
Minnesota	Beltrami	54	1	0.0687	0.0343	Virginia	Fairfax	117	1	0.0687	0.0343
Minnesota	Clay	51	1	0.0687	0.0343	Washington	Whatcom	30	1	0.5492	0.2746
Minnesota	Dakota	337	5	0.0687	0.0154						

Table A.5 Summary of Countywide 1,2-Dibromoethane Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Pulaski	19	1	0.1921	0.0960	Minnesota	Koochiching	139	2	0.0384	0.0136
California	Alameda	463	4	0.0768	0.0192	Minnesota	Morrison	55	1	0.0384	0.0192
California	Contra Costa	454	6	0.0768	0.0157	Minnesota	Olmsted	55	1	0.0384	0.0192
California	Marin	223	2	0.0768	0.0272	Minnesota	Otter Tail	47	1	0.0384	0.0192
California	Napa	120	1	0.0768	0.0384	Minnesota	Pine	50	1	0.0384	0.0192
California	San Francisco	121	1	0.0768	0.0384	Minnesota	Pipestone	46	1	0.0384	0.0192
California	San Mateo	120	1	0.0768	0.0384	Minnesota	Ramsey	378	3	0.0384	0.0111
California	Santa Clara	239	2	0.0768	0.0272	Minnesota	Roseau	46	1	0.0384	0.0192
California	Solano	120	1	0.0768	0.0384	Minnesota	Sherburne	57	1	0.0384	0.0192
California	Sonoma	119	1	0.0768	0.0384	Minnesota	St. Louis	377	3	0.0384	0.0111
Colorado	Denver	19	1	0.1921	0.0960	Minnesota	Stearns	51	1	0.0384	0.0192
Colorado	Mesa	163	2	0.3072	0.1086	Minnesota	Swift	53	1	0.0384	0.0192
Delaware	New Castle	55	1	0.0384	0.0192	Minnesota	Washington	304	2	0.0384	0.0136
District of Columbia	District of Columbia	169	1	0.0384	0.0192	Minnesota	Winona	50	1	0.0384	0.0192
Florida	Hillsborough	28	3	0.0896	0.0264	Minnesota	Yellow Medicine	42	1	0.0384	0.0192
Florida	Pinellas	491	7	0.1123	0.0212	New Jersey	Camden	60	1	0.1921	0.0960
Georgia	Dawson	25	1	1.9000	0.9500	New Jersey	Union	22	1	0.1921	0.0960
Georgia	Fayette	207	2	0.3072	0.1086	North Dakota	Cass	60	1	0.1921	0.0960
Georgia	Fulton	22	1	1.9000	0.9500	North Dakota	Mercer	56	1	0.1921	0.0960
Georgia	Glynn	20	1	1.9000	0.9500	Ohio	Hamilton	25	1	0.3842	0.1921
Georgia	Hall	21	1	1.9000	0.9500	Oregon	Multnomah	309	6	0.3521	0.0734
Indiana	Lake	277	2	0.0317	0.0653	Pennsylvania	Delaware	643	3	0.0284	0.0082
Indiana	Porter	22	1	0.0634	0.0317	Pennsylvania	Philadelphia	269	1	0.0384	0.0192
Iowa	Linn	36	1	0.3073	0.1537	Puerto Rico	Barceloneta	37	1	0.3072	0.1536
Kansas	Sedgwick	31	1	1.2500	0.6250	Puerto Rico	San Juan	26	1	0.3072	0.1536
Louisiana	Ascension Parish	125	1	0.1921	0.0960	Rhode Island	Kent	59	1	0.0230	0.0115
Louisiana	Bossier Parish	263	1	0.1921	0.0960	Rhode Island	Providence	360	5	0.0269	0.0063
Louisiana	Calcasieu Parish	258	1	0.1921	0.0960	South Dakota	Minnehaha	41	1	0.1921	0.0960
Louisiana	East Baton Rouge Parish	186	2	0.3842	0.1358	Texas	Bexar	69	1	0.1921	0.0960
Louisiana	Ouachita Parish	104	1	0.1921	0.0960	Texas	Brazoria	129	1	0.1921	0.0960
Louisiana	St. Charles Parish	24	1	0.1921	0.0960	Texas	Cameron	127	1	0.1921	0.0960
Louisiana	St. John the Baptist Parish	27	1	0.1921	0.0960	Texas	Carson	41	2	0.0500	0.0177
Maine	Oxford	25	1	3.8500	1.9250	Texas	Dallas	83	1	1.8824	0.9412
Maryland	Anne Arundel	406	2	0.0384	0.0136	Texas	Ector	110	1	1.8824	0.9412
Maryland	Baltimore	347	1	0.0384	0.0192	Texas	El Paso	231	4	0.6147	0.2389
Maryland	Baltimore city	1398	6	0.0384	0.0078	Texas	Ellis	111	1	1.8824	0.9412
Maryland	Harford	109	1	0.0384	0.0192	Texas	Galveston	124	2	0.1921	0.0679
Michigan	Alpena	130	5	0.3847	0.0860	Texas	Harris	570	6	0.4738	0.1609
Michigan	Kent	163	3	0.8961	0.2790	Texas	Jefferson	426	4	0.1921	0.0480
Michigan	Missaukee	87	2	0.7681	0.3036	Texas	Nueces	169	2	0.1921	0.0679
Michigan	Oakland	43	1	0.3072	0.1536	Texas	Orange	78	1	0.1921	0.0960
Michigan	Washtenaw	54	2	0.7681	0.3036	Texas	Smith	75	1	1.8824	0.9412
Michigan	Wayne	877	7	1.6239	0.5630	Texas	Tarrant	96	3	1.3190	0.4448
Minnesota	Beltrami	54	1	0.0384	0.0192	Utah	Salt Lake	57	1	0.1921	0.0960
Minnesota	Clay	51	1	0.0384	0.0192	Vermont	Chittenden	75	2	0.1921	0.0679
Minnesota	Dakota	340	6	0.0384	0.0078	Vermont	Rutland	38	1	0.1921	0.0960
Minnesota	Douglas	57	1	0.0384	0.0192	Vermont	Windham	38	1	0.1921	0.0960
Minnesota	Goodhue	53	1	0.0384	0.0192	Virginia	Fairfax	117	1	0.0384	0.0192
Minnesota	Hennepin	427	4	0.0384	0.0096						

Table A.6 Summary of Countywide 1,2-Dichloroethane Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.0810	0.0405	Minnesota	Douglas	57	1	0.0202	0.0101
Arkansas	Pulaski	51	1	0.0810	0.0405	Minnesota	Goodhue	53	1	0.0688	0.0344
California	Alameda	463	4	0.2024	0.0506	Minnesota	Hennepin	427	4	0.0304	0.0088
California	Contra Costa	454	6	0.2024	0.0413	Minnesota	Koochiching	139	2	0.0435	0.0174
California	Marin	223	2	0.2024	0.0715	Minnesota	Morrison	55	1	0.0607	0.0304
California	Napa	120	1	0.2024	0.1012	Minnesota	Olmsted	55	1	0.0486	0.0243
California	San Francisco	121	1	0.2024	0.1012	Minnesota	Otter Tail	47	1	0.0526	0.0263
California	San Mateo	120	1	0.2024	0.1012	Minnesota	Pine	50	1	0.0607	0.0304
California	Santa Clara	239	2	0.2024	0.0715	Minnesota	Pipestone	46	1	0.0202	0.0101
California	Solano	120	1	0.2024	0.1012	Minnesota	Ramsey	378	3	0.0202	0.0058
California	Sonoma	119	1	0.2024	0.1012	Minnesota	Roseau	46	1	0.0465	0.0233
Colorado	Denver	19	1	0.0810	0.0405	Minnesota	Sherburne	57	1	0.0202	0.0101
Colorado	Mesa	163	2	0.1214	0.0429	Minnesota	St. Louis	377	3	0.0310	0.0100
Delaware	New Castle	55	1	0.0202	0.0101	Minnesota	Stearns	51	1	0.0202	0.0101
District of Columbia	District of Columbia	169	1	0.0202	0.0101	Minnesota	Swift	53	1	0.0202	0.0101
Florida	Broward	52	2	0.0405	0.0143	Minnesota	Washington	304	2	0.0202	0.0072
Florida	Hillsborough	90	3	0.1618	0.0467	Minnesota	Winona	50	1	0.0202	0.0101
Florida	Pinellas	588	7	0.2429	0.0330	Minnesota	Yellow Medicine	42	1	0.0526	0.0263
Georgia	Dawson	25	1	1.0000	0.5000	New Jersey	Camden	164	1	0.0810	0.0405
Georgia	Fayette	207	2	0.1214	0.0429	New Jersey	Union	22	1	0.0810	0.0405
Georgia	Fulton	22	1	1.0000	0.5000	New York	Erie	151	1	0.0810	0.0405
Georgia	Glynn	20	1	1.0000	0.5000	New York	Essex	146	2	0.0810	0.0286
Georgia	Hall	21	1	1.0000	0.5000	New York	Kings	136	1	0.0810	0.0405
Indiana	Lake	277	2	0.0167	0.0344	New York	Monroe	166	2	0.0810	0.0286
Indiana	Porter	22	1	0.0334	0.0167	New York	Niagara	139	1	0.0810	0.0405
Kansas	Sedgwick	54	1	0.4047	0.2024	New York	Rensselaer	124	1	0.0810	0.0405
Louisiana	Ascension Parish	125	1	0.0000	0.0809	New York	Richmond	2399	19	0.0725	0.0085
Louisiana	Bossier Parish	263	1	0.0810	0.0405	North Dakota	Cass	60	1	0.0810	0.0405
Louisiana	Calcasieu Parish	258	1	0.0000	0.0809	North Dakota	Mercer	56	1	0.0810	0.0405
Louisiana	East Baton Rouge Parish	283	3	0.0540	0.0330	Ohio	Hamilton	25	1	0.2024	0.1012
Louisiana	Ouachita Parish	104	1	0.0810	0.0405	Oregon	Multnomah	309	6	0.1821	0.0383
Louisiana	St. Charles Parish	112	1	0.0810	0.0405	Pennsylvania	Delaware	1095	3	0.0205	0.0059
Louisiana	St. John the Baptist Parish	112	1	0.0810	0.0405	Pennsylvania	Philadelphia	269	1	0.0678	0.0123
Maine	Oxford	25	1	2.0000	1.0000	Puerto Rico	Barceloneta	37	1	0.1214	0.0607
Maryland	Anne Arundel	406	2	0.0304	0.0113	Puerto Rico	San Juan	26	1	0.1214	0.0607
Maryland	Baltimore	347	1	0.0405	0.0202	Rhode Island	Kent	60	1	0.0202	0.0101
Maryland	Baltimore city	1398	6	0.0337	0.0072	Rhode Island	Providence	369	5	0.0202	0.0045
Maryland	Harford	109	1	0.0202	0.0101	South Dakota	Minnehaha	41	1	0.0810	0.0405
Michigan	Alpena	130	5	0.3809	0.0912	Tennessee	Davidson	30	1	0.0810	0.0405
Michigan	Kalamazoo	670	4	0.0810	0.0202	Texas	Bexar	261	1	0.1012	0.0506
Michigan	Kent	163	3	0.4720	0.1470	Texas	Brazoria	321	1	0.1012	0.0506
Michigan	Midland	1504	4	0.5000	0.1250	Texas	Cameron	359	2	0.0911	0.0324
Michigan	Missaukee	87	2	0.4046	0.1599	Texas	Carson	834	8	0.0252	0.0052
Michigan	Oakland	43	1	0.1214	0.0607	Texas	Dallas	466	3	0.4182	0.1770
Michigan	Van Buren	51	1	0.0810	0.0405	Texas	Denton	28	1	0.1012	0.0506
Michigan	Washtenaw	54	2	0.4046	0.1599	Texas	Ector	250	1	0.1012	0.0506
Michigan	Wayne	877	7	0.8439	0.2964	Texas	El Paso	1377	7	0.0983	0.0186
Minnesota	Beltrami	54	1	0.0202	0.0101	Texas	Ellis	330	2	0.5767	0.2643
Minnesota	Clay	51	1	0.0202	0.0101	Texas	Galveston	670	4	0.0961	0.0241
Minnesota	Dakota	340	6	0.0202	0.0041	Texas	Gregg	161	1	0.1012	0.0506

Table A.6 Summary of Countywide 1,2-Dichloroethane Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Texas	Harris	2466	13	0.1743	0.0427	Texas	Webb	244	1	0.1012	0.0506
Texas	Hidalgo	422	2	0.1012	0.0358	Utah	Salt Lake	57	1	0.0810	0.0405
Texas	Jefferson	1542	7	0.0983	0.0186	Vermont	Chittenden	260	2	0.0810	0.0286
Texas	Nueces	955	4	0.1012	0.0253	Vermont	Rutland	126	1	0.0810	0.0405
Texas	Orange	276	1	0.1012	0.0506	Vermont	Windham	105	1	0.0810	0.0405
Texas	Smith	89	1	1.0523	0.5262	Virginia	Fairfax	117	1	0.0202	0.0101
Texas	Tarrant	242	4	0.5717	0.1867	Washington	Whatcom	60	2	0.3025	0.1326
Texas	Travis	208	3	0.1012	0.0292						

Table A.7 Summary of Countywide 1,2-Dichloropropane Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.2311	0.1155	Minnesota	Swift	53	1	0.0462	0.0231
Arkansas	Pulaski	51	1	0.2311	0.1155	Minnesota	Washington	304	2	0.0462	0.0163
Colorado	Denver	19	1	0.2311	0.1155	Minnesota	Winona	50	1	0.0462	0.0231
Colorado	Mesa	163	2	0.1617	0.0572	Minnesota	Yellow Medicine	42	1	0.0462	0.0231
Delaware	New Castle	55	1	0.0462	0.0231	New Jersey	Camden	164	1	0.2311	0.1155
District of Columbia	District of Columbia	169	1	0.0462	0.0231	New Jersey	Union	22	1	0.2311	0.1155
Florida	Broward	52	2	0.0693	0.0245	New Mexico	Sandoval	40	2	0.2079	0.0735
Florida	Hillsborough	36	3	0.1155	0.0333	New York	Erie	155	1	0.1848	0.0924
Florida	Pinellas	503	7	0.1277	0.0287	New York	Essex	143	2	0.2080	0.0740
Georgia	Dawson	25	1	1.1500	0.5750	New York	Kings	142	1	0.1848	0.0924
Georgia	Fayette	207	2	0.1617	0.0572	New York	Monroe	147	2	0.1155	0.1090
Georgia	Fulton	22	1	1.1500	0.5750	New York	Niagara	90	1	0.2311	0.1155
Georgia	Glynn	20	1	1.1500	0.5750	New York	Rensselaer	137	1	0.1848	0.0924
Georgia	Hall	21	1	1.1500	0.5750	New York	Richmond	1273	15	0.2126	0.0276
Indiana	Lake	277	2	0.0127	0.0387	North Dakota	Cass	60	1	0.2311	0.1155
Indiana	Porter	22	1	0.0254	0.0127	North Dakota	Mercer	56	1	0.2311	0.1155
Iowa	Linn	36	1	0.1617	0.0809	Ohio	Hamilton	25	1	0.2311	0.1155
Kansas	Sedgwick	54	1	0.4621	0.2311	Oregon	Multnomah	531	10	0.1786	0.0300
Louisiana	Ascension Parish	125	1	0.2311	0.1155	Oregon	Washington	55	1	0.1000	0.0500
Louisiana	Bossier Parish	263	1	0.2311	0.1155	Pennsylvania	Delaware	805	3	0.0396	0.0117
Louisiana	Calcasieu Parish	258	1	0.2311	0.1155	Pennsylvania	Philadelphia	269	1	0.0462	0.0231
Louisiana	East Baton Rouge Parish	283	3	0.2311	0.0667	Puerto Rico	Barceloneta	37	1	0.1617	0.0808
Louisiana	Ouachita Parish	104	1	0.2311	0.1155	Puerto Rico	San Juan	26	1	0.1617	0.0808
Louisiana	St. Charles Parish	112	1	0.2311	0.1155	Rhode Island	Kent	59	1	0.0231	0.0115
Louisiana	St. John the Baptist Parish	112	1	0.2311	0.1155	Rhode Island	Providence	360	5	0.0254	0.0058
Maine	Oxford	24	1	2.3000	1.1500	South Dakota	Minnehaha	41	1	0.2311	0.1155
Maryland	Anne Arundel	406	2	0.0462	0.0163	Tennessee	Davidson	30	1	0.2311	0.1155
Maryland	Baltimore	347	1	0.0462	0.0231	Texas	Bexar	261	1	0.0693	0.0347
Maryland	Baltimore city	1398	6	0.0462	0.0094	Texas	Brazoria	321	1	0.0693	0.0347
Maryland	Harford	109	1	0.0462	0.0231	Texas	Cameron	359	2	0.1502	0.0603
Michigan	Alpena	130	5	0.2304	0.0515	Texas	Carson	834	8	0.0298	0.0057
Michigan	Kent	163	3	0.5390	0.1678	Texas	Dallas	466	3	0.2388	0.0977
Michigan	Missaukee	87	2	0.4620	0.1826	Texas	Denton	28	1	0.0693	0.0347
Michigan	Oakland	43	1	0.1617	0.0808	Texas	Ector	250	1	0.0693	0.0347
Michigan	Washtenaw	54	2	0.4620	0.1826	Texas	El Paso	1424	9	0.0873	0.0168
Michigan	Wayne	877	7	0.9701	0.3385	Texas	Ellis	330	2	0.3235	0.1454
Minnesota	Beltrami	54	1	0.0462	0.0231	Texas	Galveston	670	4	0.1098	0.0325
Minnesota	Clay	51	1	0.0462	0.0231	Texas	Gregg	161	1	0.0693	0.0347
Minnesota	Dakota	337	5	0.0462	0.0103	Texas	Harris	2514	14	0.1056	0.0225
Minnesota	Douglas	57	1	0.0462	0.0231	Texas	Hidalgo	422	2	0.0693	0.0245
Minnesota	Goodhue	53	1	0.0462	0.0231	Texas	Jefferson	1542	7	0.0924	0.0205
Minnesota	Hennepin	427	4	0.0462	0.0116	Texas	Nueces	955	4	0.0693	0.0173
Minnesota	Koochiching	139	2	0.0462	0.0163	Texas	Orange	276	1	0.0693	0.0347
Minnesota	Morrison	55	1	0.0462	0.0231	Texas	Smith	89	1	0.5776	0.2888
Minnesota	Olmsted	55	1	0.0462	0.0231	Texas	Tarrant	242	4	0.3639	0.1065
Minnesota	Otter Tail	47	1	0.0462	0.0231	Texas	Travis	208	3	0.0693	0.0200
Minnesota	Pine	50	1	0.0462	0.0231	Texas	Webb	244	1	0.0693	0.0347
Minnesota	Pipestone	46	1	0.0462	0.0231	Utah	Salt Lake	57	1	0.2311	0.1155
Minnesota	Ramsey	378	3	0.0462	0.0133	Vermont	Chittenden	260	2	0.2311	0.0817
Minnesota	Roseau	46	1	0.0462	0.0231	Vermont	Rutland	126	1	0.2311	0.1155
Minnesota	Sherburne	57	1	0.0462	0.0231	Vermont	Windham	105	1	0.2311	0.1155
Minnesota	St. Louis	377	3	0.0462	0.0133	Virginia	Fairfax	117	1	0.0462	0.0231
Minnesota	Stearns	51	1	0.0462	0.0231	Washington	Whatcom	60	2	0.1606	0.0620

Table A.8 Summary of Countywide 1,3-Butadiene Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.1106	0.0553	Minnesota	Ramsey	83	3	0.0221	0.0064
Arkansas	Pulaski	51	1	0.1106	0.0553	Minnesota	St. Louis	59	2	0.0221	0.0078
California	Alameda	540	4	0.2015	0.0749	Minnesota	Stearns	23	1	0.0221	0.0111
California	Butte	179	1	0.0000	0.1106	Minnesota	Swift	22	1	0.0221	0.0111
California	Contra Costa	789	8	0.1545	0.0511	Minnesota	Washington	77	2	0.0221	0.0078
California	Fresno	182	1	0.0000	0.1106	Minnesota	Winona	24	1	0.0221	0.0111
California	Humboldt	23	1	0.1106	0.0553	New Jersey	Camden	164	1	0.1106	0.0553
California	Imperial	155	1	0.0221	0.1106	New Jersey	Union	22	1	0.1106	0.0553
California	Kern	361	2	0.0000	0.0782	New Mexico	Sandoval	40	2	0.0885	0.0313
California	Los Angeles	812	16	0.2613	0.0407	North Dakota	Cass	120	2	0.1106	0.0391
California	Marin	223	2	0.1818	0.1118	North Dakota	Mercer	56	1	0.1106	0.0553
California	Napa	120	1	0.0318	0.1500	Ohio	Hamilton	145	4	0.1989	0.0514
California	Orange	42	2	0.2697	0.1137	Oregon	Multnomah	309	10	0.1064	0.0168
California	Placer	179	1	0.0000	0.1106	Oregon	Washington	31	1	0.1000	0.0500
California	Riverside	281	5	0.1017	0.0392	Pennsylvania	Delaware	1231	3	0.0985	0.0367
California	San Bernardino	191	4	0.3836	0.1032	Pennsylvania	Philadelphia	268	1	0.3097	0.1549
California	San Diego	386	3	0.0148	0.0551	Puerto Rico	Barceloneta	75	1	0.1050	0.0525
California	San Francisco	174	1	0.0000	0.1106	Puerto Rico	San Juan	55	1	0.2875	0.1437
California	San Joaquin	183	1	0.0000	0.1106	Rhode Island	Kent	324	3	0.0735	0.0235
California	San Mateo	120	1	0.0318	0.1500	Rhode Island	Providence	903	7	0.1425	0.0317
California	Santa Barbara	152	1	0.1106	0.1106	South Dakota	Minnehaha	82	2	0.1106	0.0391
California	Santa Clara	299	2	0.0270	0.0932	Tennessee	Davidson	30	1	0.1106	0.0553
California	Solano	120	1	0.0318	0.1500	Texas	Bexar	261	1	0.0000	0.0332
California	Sonoma	119	1	0.0318	0.1500	Texas	Brazoria	321	1	0.0000	0.0332
California	Stanislaus	156	1	0.0000	0.1106	Texas	Cameron	359	2	0.0553	0.0322
California	Ventura	172	1	0.0362	0.0246	Texas	Carson	834	8	0.0208	0.0047
Colorado	Denver	19	1	0.4203	0.2102	Texas	Dallas	466	3	0.2323	0.1172
Colorado	Mesa	163	2	0.1161	0.0433	Texas	Denton	28	1	0.0332	0.0166
Delaware	New Castle	54	1	0.2655	0.1327	Texas	Ector	250	1	0.0000	0.0332
District of Columbia	District of Columbia	169	1	0.1770	0.0885	Texas	El Paso	1424	9	0.0270	0.0155
Florida	Broward	39	1	0.1240	0.0502	Texas	Ellis	330	2	0.3650	0.1744
Florida	Hillsborough	158	3	0.0885	0.0255	Texas	Galveston	670	4	0.0277	0.0199
Florida	Pinellas	224	3	0.1014	0.0297	Texas	Gregg	161	1	0.0332	0.0166
Georgia	Fayette	410	2	0.1050	0.0371	Texas	Harris	3099	15	0.0557	0.0247
Iowa	Linn	73	2	0.0913	0.0326	Texas	Hidalgo	422	2	0.0166	0.0186
Kansas	Sedgwick	52	1	2.2122	1.1061	Texas	Jefferson	1542	7	0.0000	0.0196
Louisiana	East Baton Rouge Parish	97	1	0.1106	0.0553	Texas	Nueces	955	4	0.0166	0.0131
Louisiana	St. Charles Parish	112	1	0.1106	0.0553	Texas	Orange	276	1	0.0000	0.0332
Louisiana	St. John the Baptist Parish	112	1	0.1106	0.0553	Texas	Smith	89	1	0.6968	0.3484
Maryland	Anne Arundel	406	2	0.0636	0.0161	Texas	Tarrant	242	4	0.3761	0.1242
Maryland	Baltimore	347	1	0.1770	0.0885	Texas	Travis	208	3	0.0000	0.0192
Maryland	Baltimore city	1395	6	0.1502	0.0249	Texas	Webb	244	1	0.0000	0.0332
Maryland	Harford	109	1	0.0221	0.0111	Utah	Salt Lake	114	2	0.2848	0.1007
Michigan	Kent	37	2	0.3317	0.1173	Vermont	Chittenden	347	4	0.1894	0.0513
Michigan	Midland	1486	4	0.2200	0.0550	Vermont	Rutland	169	2	0.1106	0.0391
Michigan	Missaukee	30	1	0.3317	0.1659	Vermont	Windham	146	2	0.1106	0.0391
Michigan	Oakland	43	1	0.1548	0.0774	Virginia	Fairfax	117	1	0.1204	0.0087
Michigan	Washtenaw	33	1	0.3317	0.1659	Washington	King	460	6	0.0894	0.0193
Michigan	Wayne	560	5	0.2808	0.0668	Washington	Whatcom	60	2	0.1377	0.0496
Minnesota	Beltrami	21	1	0.0221	0.0111	Wisconsin	Manitowoc	32	1	0.0277	0.0138
Minnesota	Dakota	155	5	0.0221	0.0049	Wisconsin	Milwaukee	211	1	0.0000	0.0277
Minnesota	Hennepin	72	2	0.0221	0.0078	Wisconsin	Ozaukee	54	1	0.0277	0.0138

Table A.9 Summary of Countywide Acrylonitrile Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.0651	0.0326	Michigan	Wayne	557	5	0.6334	0.1659
Arkansas	Pulaski	51	1	0.0651	0.0326	New Jersey	Camden	91	1	0.0651	0.0326
California	Riverside	13	2	0.2072	0.0811	New Jersey	Union	22	1	0.0651	0.0326
Colorado	Denver	19	1	0.0651	0.0326	North Dakota	Cass	60	1	0.0651	0.0326
Colorado	Mesa	163	2	0.2278	0.0805	North Dakota	Mercer	56	1	0.0651	0.0326
Florida	Hillsborough	160	3	0.2784	0.0987	Oregon	Multnomah	32	1	0.0651	0.0326
Florida	Pinellas	226	3	0.3471	0.1023	Puerto Rico	Barceloneta	37	1	0.2278	0.1139
Georgia	Fayette	207	2	0.2278	0.0805	Puerto Rico	San Juan	26	1	0.2278	0.1139
Iowa	Linn	36	1	0.2279	0.1139	Rhode Island	Kent	59	1	0.0195	0.0098
Louisiana	East Baton Rouge Parish	47	1	0.0651	0.0326	Rhode Island	Providence	374	5	0.0282	0.0064
Louisiana	St. Charles Parish	57	1	0.0651	0.0326	South Dakota	Minnehaha	41	1	0.0000	0.0651
Louisiana	St. John the Baptist Parish	56	1	0.0651	0.0326	Texas	El Paso	65	1	0.0651	0.0326
Michigan	Kent	37	2	0.8135	0.3032	Texas	Tarrant	27	1	0.0651	0.0326
Michigan	Midland	1488	4	0.0250	0.0451	Utah	Salt Lake	57	1	0.0000	0.0651
Michigan	Missaukee	30	1	0.5423	0.2712	Vermont	Chittenden	137	2	0.0651	0.0230
Michigan	Oakland	43	1	0.2278	0.1139	Vermont	Rutland	68	1	0.0651	0.0326
Michigan	Washtenaw	33	1	0.5423	0.2712	Vermont	Windham	69	1	0.0651	0.0326

Table A.10 Summary of Countywide Benzene Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.3274	0.0633	Indiana	Porter	22	1	0.0088	0.0044
Arkansas	Pulaski	51	1	0.6389	0.3195	Iowa	Linn	73	2	0.5162	0.1349
California	Alameda	541	4	0.2191	0.0653	Kansas	Sedgwick	55	1	4.7919	2.3960
California	Butte	179	1	0.1597	0.1597	Kentucky	Boyd	26	1	1.0000	0.5000
California	Contra Costa	769	7	0.2165	0.0544	Kentucky	Lawrence	23	1	1.0000	0.5000
California	Fresno	184	1	0.1597	0.1749	Louisiana	Ascension Parish	125	1	0.2758	0.0311
California	Humboldt	23	1	0.7987	0.3993	Louisiana	Bossier Parish	263	1	0.3146	0.0384
California	Imperial	192	2	0.9175	0.3294	Louisiana	Calcasieu Parish	258	1	0.4472	0.2236
California	Kern	361	2	0.1597	0.1129	Louisiana	East Baton Rouge Parish	1089	6	0.4927	0.1142
California	Los Angeles	2505	27	0.8089	0.0927	Louisiana	Iberville Parish	350	1	1.1394	0.5697
California	Marin	104	1	0.2195	0.1097	Louisiana	Ouachita Parish	104	1	0.1278	0.0639
California	Napa	120	1	0.5584	0.2792	Louisiana	St. Charles Parish	112	1	0.5018	0.0439
California	Orange	46	2	0.6264	0.1896	Louisiana	St. John the Baptist Parish	112	1	0.4169	0.0353
California	Placer	179	1	0.1597	0.1597	Maine	Cumberland	830	2	0.1730	0.0666
California	Riverside	844	7	0.6766	0.1364	Maine	Hancock	577	3	0.1582	0.0576
California	San Bernardino	657	6	0.4822	0.0345	Maine	York	393	1	0.0000	0.0027
California	San Diego	385	3	0.2130	0.1627	Maryland	Anne Arundel	749	4	0.2848	0.0101
California	San Francisco	174	1	0.1597	0.1597	Maryland	Baltimore	1015	3	0.5775	0.1363
California	San Joaquin	184	1	0.1597	0.1597	Maryland	Baltimore city	1731	8	0.6058	0.0466
California	San Mateo	119	1	0.2778	0.2500	Maryland	Harford	494	3	0.1116	0.0407
California	Santa Barbara	203	2	0.1687	0.0810	Massachusetts	Barnstable	90	2	0.2415	0.0882
California	Santa Clara	299	2	0.1993	0.0997	Massachusetts	Bristol	161	2	0.3145	0.0971
California	Solano	120	1	0.2389	0.1195	Massachusetts	Essex	1078	5	0.3761	0.0629
California	Sonoma	119	1	0.2389	0.1195	Massachusetts	Hampden	730	4	0.3311	0.0714
California	Stanislaus	155	1	0.1597	0.1597	Massachusetts	Hampshire	389	2	0.1674	0.0598
California	Ventura	224	2	0.6980	0.3192	Massachusetts	Suffolk	37	1	0.3064	0.0268
Colorado	Denver	19	1	3.0349	1.5174	Michigan	Allegan	198	1	0.2810	0.0147
Colorado	Mesa	163	2	0.3646	0.3367	Michigan	Alpena	130	5	0.2559	0.0599
Connecticut	Fairfield	413	2	0.0867	0.0454	Michigan	Kalamazoo	675	4	0.3854	0.0586
Connecticut	Hartford	434	2	0.1146	0.0588	Michigan	Kent	162	3	0.9123	0.2976
Connecticut	New Haven	153	1	0.2405	0.0037	Michigan	Midland	1458	4	0.2682	0.1175
Connecticut	Tolland	475	2	0.0590	0.0133	Michigan	Missaukee	68	2	0.3194	0.1262
Delaware	New Castle	453	2	0.4921	0.0153	Michigan	Oakland	43	1	0.8941	0.4471
District of Columbia	District of Columbia	637	2	0.2277	0.0277	Michigan	Van Buren	52	1	0.4472	0.2236
Florida	Broward	462	10	0.3378	0.0572	Michigan	Washtenaw	44	2	1.3095	0.4631
Florida	Hillsborough	169	3	0.2980	0.1059	Michigan	Wayne	846	7	1.0971	0.2349
Florida	Pinellas	677	7	0.3048	0.0731	Minnesota	Beltrami	54	1	0.3961	0.1981
Georgia	Dawson	25	1	0.8000	0.4000	Minnesota	Clay	51	1	0.4165	0.0202
Georgia	DeKalb	858	4	0.2741	0.0628	Minnesota	Dakota	340	6	0.2753	0.0275
Georgia	Fayette	410	2	0.3060	0.0726	Minnesota	Douglas	57	1	0.5111	0.0215
Georgia	Fulton	22	1	0.8000	0.4000	Minnesota	Goodhue	53	1	0.1853	0.0926
Georgia	Glynn	20	1	0.8000	0.4000	Minnesota	Hennepin	427	4	0.5510	0.0736
Georgia	Hall	20	1	0.8000	0.4000	Minnesota	Koochiching	139	2	0.3048	0.0162
Georgia	Paulding	240	2	0.0873	0.0116	Minnesota	Morrison	55	1	0.2407	0.0265
Georgia	Rockdale	451	2	0.3024	0.0074	Minnesota	Olmsted	55	1	0.3738	0.1869
Idaho	Nez Perce	749	13	0.7916	0.1387	Minnesota	Otter Tail	47	1	0.6422	0.0178
Illinois	Cook	1052	4	0.2103	0.0129	Minnesota	Pine	50	1	0.4616	0.2308
Illinois	Lake	412	2	0.3461	0.1730	Minnesota	Pipestone	46	1	0.1607	0.0275
Illinois	Will	314	2	0.2396	0.1198	Minnesota	Ramsey	378	3	0.5074	0.1323
Indiana	Lake	946	3	0.0949	0.0507	Minnesota	Roseau	46	1	0.1246	0.0623

Table A.10 Summary of Countywide Benzene Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Minnesota	Sherburne	57	1	0.2860	0.0340	Rhode Island	Providence	1343	7	0.3462	0.0486
Minnesota	St. Louis	375	3	0.3702	0.0557	South Dakota	Minnehaha	82	2	0.4744	0.1046
Minnesota	Stearns	51	1	0.2268	0.1134	Tennessee	Davidson	30	1	0.2073	0.1413
Minnesota	Swift	53	1	0.3195	0.1597	Texas	Bexar	261	1	0.3850	0.0242
Minnesota	Washington	304	2	0.4806	0.0145	Texas	Brazoria	321	1	0.2726	0.0188
Minnesota	Winona	50	1	0.4984	0.0002	Texas	Cameron	359	2	0.5420	0.2674
Minnesota	Yellow Medicine	42	1	0.1641	0.0311	Texas	Carson	834	8	0.2314	0.0416
New Hampshire	Rockingham	254	1	0.0000	0.0027	Texas	Dallas	466	3	0.6899	0.1412
New Jersey	Camden	502	2	0.4048	0.0881	Texas	Denton	28	1	0.2236	0.1118
New Jersey	Mercer	523	1	0.4283	0.0009	Texas	Ector	250	1	0.6070	0.3035
New Jersey	Middlesex	420	1	0.3088	0.1544	Texas	El Paso	1459	10	0.7851	0.0783
New Jersey	Union	22	1	0.6709	0.3354	Texas	Ellis	330	2	0.6055	0.2119
New Mexico	Sandoval	38	2	1.2774	0.4655	Texas	Galveston	670	4	0.5465	0.0286
New York	Albany	72	1	0.4366	0.0242	Texas	Gregg	161	1	0.5111	0.2556
New York	Bronx	669	2	0.5420	0.0159	Texas	Harris	3076	15	0.6383	0.0669
New York	Erie	141	1	0.2933	0.4216	Texas	Hidalgo	422	2	0.4518	0.1283
New York	Essex	137	2	0.1597	0.1786	Texas	Jefferson	1928	9	0.3868	0.0504
New York	Kings	124	1	0.6674	0.3337	Texas	Nueces	955	4	0.3662	0.0608
New York	Monroe	149	2	0.3162	0.1952	Texas	Orange	276	1	0.4617	0.0172
New York	Niagara	130	1	0.0000	0.3195	Texas	Smith	89	1	0.8466	0.4233
New York	Queens	389	2	0.2284	0.0238	Texas	Tarrant	242	4	0.7110	0.1512
New York	Rensselaer	117	1	0.6964	0.3482	Texas	Travis	208	3	0.7589	0.1954
New York	Richmond	2473	19	0.5377	0.0728	Texas	Webb	244	1	0.5153	0.0189
North Carolina	Lincoln	269	1	0.4709	0.2354	Utah	Salt Lake	114	2	1.1687	0.4132
North Carolina	Mecklenburg	1771	7	0.2594	0.0293	Vermont	Chittenden	347	4	0.5375	0.0981
North Dakota	Cass	120	2	0.4852	0.1760	Vermont	Rutland	169	2	0.6254	0.0288
North Dakota	Mercer	56	1	0.5750	0.2875	Vermont	Windham	146	2	0.6602	0.0227
Ohio	Butler	57	1	0.5000	0.2500	Virginia	Caroline	472	2	0.1324	0.0082
Ohio	Cuyahoga	158	2	1.4600	0.5243	Virginia	Fairfax	441	3	0.4097	0.0084
Ohio	Hamilton	358	4	0.7297	0.2025	Virginia	Henrico	44	1	0.4334	0.2167
Ohio	Scioto	27	1	4.7200	2.3600	Virginia	Richmond city	113	1	0.5058	0.2529
Oregon	Multnomah	531	10	0.7043	0.0268	Washington	Asotin	48	1	0.3783	0.1178
Oregon	Washington	55	1	0.4067	0.2034	Washington	King	472	6	0.3557	0.0418
Pennsylvania	Adams	222	1	0.2092	0.0015	Washington	Whatcom	60	2	1.0049	0.2770
Pennsylvania	Delaware	1387	3	0.6138	0.0849	West Virginia	Wayne	120	5	1.1430	0.2635
Pennsylvania	Philadelphia	1061	6	0.6741	0.1189	Wisconsin	Manitowoc	32	1	0.2662	0.1331
Puerto Rico	Barceloneta	75	1	0.1729	0.0731	Wisconsin	Milwaukee	221	1	0.7987	0.3245
Puerto Rico	San Juan	55	1	0.7087	0.0908	Wisconsin	Ozaukee	54	1	0.2662	0.1331
Rhode Island	Kent	536	3	0.3806	0.1171						

Table A.11 Summary of Countywide Carbon Tetrachloride Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Alameda	272	2	0.4304	0.1576	Minnesota	Otter Tail	47	1	0.6228	0.3114
California	Butte	140	1	0.3146	0.1573	Minnesota	Pine	50	1	0.7097	0.0249
California	Contra Costa	352	4	0.4876	0.1295	Minnesota	Pipestone	46	1	0.7131	0.0227
California	Fresno	138	1	0.3146	0.1573	Minnesota	Ramsey	378	3	0.3964	0.1048
California	Humboldt	23	1	0.3146	0.1573	Minnesota	Roseau	46	1	0.6039	0.3020
California	Imperial	130	1	0.3146	0.1573	Minnesota	Sherburne	57	1	0.4403	0.0214
California	Kern	303	2	0.4718	0.1758	Minnesota	St. Louis	375	3	0.4047	0.0989
California	Los Angeles	603	10	0.4815	0.0762	Minnesota	Stearns	51	1	0.2794	0.0212
California	Orange	24	1	0.5273	0.0160	Minnesota	Swift	53	1	0.2957	0.0266
California	Placer	135	1	0.3146	0.1573	Minnesota	Washington	304	2	0.2994	0.0774
California	Riverside	423	5	0.5032	0.1177	Minnesota	Winona	50	1	0.3146	0.1573
California	San Bernardino	147	3	0.4900	0.1464	Minnesota	Yellow Medicine	42	1	0.6543	0.3271
California	San Diego	329	3	0.3146	0.0908	New Mexico	Sandoval	29	2	0.2515	0.0889
California	San Francisco	146	1	0.6291	0.3146	New York	Erie	143	1	0.1887	0.0944
California	San Joaquin	154	1	0.3146	0.1573	New York	Essex	143	2	0.3583	0.1401
California	Santa Barbara	112	1	0.4718	0.2359	New York	Kings	134	1	0.1887	0.0944
California	Santa Clara	152	1	0.3146	0.1573	New York	Monroe	165	2	0.3685	0.1450
California	Stanislaus	119	1	0.3146	0.1573	New York	Niagara	134	1	0.1887	0.0944
California	Ventura	135	1	0.3146	0.1573	New York	Rensselaer	120	1	0.1887	0.0944
Colorado	Mesa	163	2	0.5345	0.1893	New York	Richmond	2468	19	0.1729	0.0201
Delaware	New Castle	55	1	0.4926	0.0160	Ohio	Cuyahoga	57	2	0.3848	0.1662
District of Columbia	District of Columbia	169	1	0.5033	0.2516	Ohio	Hamilton	25	1	0.3146	0.1573
Florida	Broward	382	9	0.3090	0.0244	Oregon	Multnomah	499	9	0.7626	0.1463
Florida	Hillsborough	169	3	0.6289	0.1815	Oregon	Washington	55	1	0.5000	0.2500
Florida	Pinellas	677	7	0.7774	0.0837	Pennsylvania	Delaware	1387	3	0.5867	0.1005
Georgia	Dawson	25	1	1.5500	0.7750	Pennsylvania	Philadelphia	269	1	0.4880	0.0080
Georgia	Fayette	207	2	0.5660	0.2001	Puerto Rico	Barceloneta	37	1	0.5660	0.2830
Georgia	Fulton	22	1	1.5500	0.7750	Puerto Rico	San Juan	26	1	0.5660	0.2830
Georgia	Glynn	20	1	1.5500	0.7750	Rhode Island	Kent	333	3	0.4784	0.1318
Georgia	Hall	21	1	1.5500	0.7750	Rhode Island	Providence	902	7	0.4454	0.0865
Indiana	Lake	277	2	0.2202	0.1217	Texas	Bexar	261	1	0.3146	0.1573
Indiana	Porter	22	1	0.1038	0.0519	Texas	Brazoria	311	1	0.6291	0.3146
Iowa	Linn	36	1	0.0916	0.0458	Texas	Cameron	312	1	0.6291	0.3146
Kansas	Sedgwick	54	1	0.6291	0.3146	Texas	Carson	665	7	0.4665	0.0799
Maryland	Anne Arundel	406	2	0.5033	0.1779	Texas	Dallas	425	3	0.6815	0.2473
Maryland	Baltimore	347	1	0.4974	0.0063	Texas	Denton	28	1	0.3146	0.1573
Maryland	Baltimore city	1398	6	0.4986	0.0839	Texas	Ector	210	1	0.3146	0.1573
Maryland	Harford	109	1	0.5023	0.0092	Texas	El Paso	1296	8	0.3146	0.0556
Michigan	Alpena	130	5	0.5038	0.1178	Texas	Ellis	330	2	0.8650	0.3625
Michigan	Kent	161	3	0.7337	0.2284	Texas	Galveston	612	3	0.4194	0.1284
Michigan	Midland	1486	4	0.6300	0.1575	Texas	Gregg	161	1	0.3146	0.1573
Michigan	Missaukee	63	2	0.6289	0.2486	Texas	Harris	2471	14	0.3932	0.0704
Michigan	Oakland	43	1	0.6289	0.3144	Texas	Hidalgo	422	2	0.3146	0.1112
Michigan	Washtenaw	28	2	0.6289	0.2486	Texas	Jefferson	1379	6	0.3670	0.0786
Michigan	Wayne	727	7	0.5930	0.1185	Texas	Nueces	955	4	0.3146	0.0786
Minnesota	Beltrami	54	1	0.3208	0.1604	Texas	Orange	275	1	0.3146	0.1573
Minnesota	Clay	51	1	0.2957	0.1478	Texas	Smith	89	1	1.4155	0.7077
Minnesota	Dakota	340	6	0.3104	0.0447	Texas	Tarrant	215	3	1.0485	0.3377
Minnesota	Douglas	57	1	0.7109	0.3554	Texas	Travis	208	3	0.3146	0.0908
Minnesota	Goodhue	53	1	0.7361	0.0186	Texas	Webb	244	1	0.3146	0.1573
Minnesota	Hennepin	427	4	0.4698	0.0868	Virginia	Fairfax	117	1	0.4680	0.0165
Minnesota	Koochiching	139	2	0.5316	0.1977	Washington	King	492	6	0.4520	0.0858
Minnesota	Morrison	55	1	0.7172	0.3586	Washington	Whatcom	30	1	0.5033	0.2516
Minnesota	Olmsted	55	1	0.4404	0.2202						

Table A.12 Summary of Countywide Chloroform Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.0244	0.0122	Maryland	Baltimore	347	1	0.2441	0.1221
Arkansas	Pulaski	51	1	0.0244	0.0122	Maryland	Baltimore city	1398	6	0.1980	0.0369
California	Alameda	527	4	0.0827	0.0311	Maryland	Harford	109	1	0.0977	0.0488
California	Butte	151	1	0.2441	0.1221	Michigan	Alpena	130	5	0.3887	0.0908
California	Contra Costa	714	8	0.1390	0.0308	Michigan	Kalamazoo	666	4	0.0122	0.0096
California	Fresno	153	1	0.2441	0.1221	Michigan	Kent	163	3	0.5694	0.1773
California	Humboldt	23	1	0.2441	0.1221	Michigan	Midland	1468	4	0.4850	0.1213
California	Imperial	148	1	0.2441	0.1221	Michigan	Missaukee	70	2	0.4881	0.1929
California	Kern	339	2	0.2441	0.0863	Michigan	Oakland	43	1	0.1220	0.0610
California	Los Angeles	577	7	0.2175	0.0421	Michigan	Van Buren	50	1	0.0244	0.0122
California	Marin	223	2	0.0488	0.0173	Michigan	Washtenaw	44	2	0.4881	0.1929
California	Napa	120	1	0.0777	0.0388	Michigan	Wayne	875	7	1.0110	0.3574
California	Placer	152	1	0.2441	0.1221	Minnesota	Beltrami	54	1	0.0244	0.0122
California	Riverside	393	4	0.2441	0.0610	Minnesota	Clay	51	1	0.0244	0.0122
California	San Bernardino	156	3	0.2116	0.0625	Minnesota	Dakota	329	6	0.0374	0.0085
California	San Diego	358	3	0.2441	0.0705	Minnesota	Douglas	57	1	0.1709	0.0854
California	San Francisco	159	1	0.2441	0.1221	Minnesota	Goodhue	53	1	0.1025	0.0513
California	San Joaquin	166	1	0.2441	0.1221	Minnesota	Hennepin	423	4	0.0989	0.0257
California	San Mateo	120	1	0.0288	0.0144	Minnesota	Koochiching	139	2	0.1172	0.0423
California	Santa Barbara	126	1	0.2441	0.1221	Minnesota	Morrison	55	1	0.1025	0.0513
California	Santa Clara	284	2	0.1365	0.0615	Minnesota	Olmsted	55	1	0.0830	0.0415
California	Solano	120	1	0.0488	0.0244	Minnesota	Otter Tail	47	1	0.0830	0.0415
California	Sonoma	119	1	0.0288	0.0144	Minnesota	Pine	50	1	0.0952	0.0476
California	Stanislaus	132	1	0.2441	0.1221	Minnesota	Pipestone	46	1	0.0895	0.0122
California	Ventura	148	1	0.2441	0.1221	Minnesota	Ramsey	373	3	0.0874	0.0189
Colorado	Denver	19	1	0.0244	0.0122	Minnesota	Roseau	46	1	0.0903	0.0452
Colorado	Mesa	163	2	0.1220	0.0431	Minnesota	Sherburne	57	1	0.0684	0.0342
Delaware	New Castle	55	1	0.1465	0.0732	Minnesota	St. Louis	377	3	0.0667	0.0211
District of Columbia	District of Columbia	169	1	0.0488	0.0244	Minnesota	Stearns	51	1	0.0586	0.0293
Florida	Broward	52	2	0.0288	0.0102	Minnesota	Swift	53	1	0.0244	0.0122
Florida	Hillsborough	169	3	0.1464	0.0438	Minnesota	Washington	302	2	0.0732	0.0268
Florida	Pinellas	677	7	0.1341	0.0268	Minnesota	Winona	50	1	0.0391	0.0195
Georgia	Dawson	25	1	1.2000	0.6000	Minnesota	Yellow Medicine	42	1	0.0854	0.0427
Georgia	Fayette	207	2	0.1220	0.0431	New Jersey	Camden	164	1	0.0000	0.0244
Georgia	Fulton	22	1	1.2000	0.6000	New Jersey	Union	22	1	0.0244	0.0122
Georgia	Glynn	20	1	1.2000	0.6000	New Mexico	Sandoval	40	2	0.2928	0.1035
Georgia	Hall	21	1	1.2000	0.6000	New York	Erie	150	1	0.0977	0.0488
Idaho	Nez Perce	749	13	0.1506	0.0238	New York	Essex	147	2	0.0977	0.0345
Indiana	Lake	277	2	0.0806	0.0285	New York	Kings	136	1	0.0977	0.0488
Indiana	Porter	22	1	0.0806	0.0403	New York	Monroe	166	2	0.0977	0.0345
Iowa	Linn	36	1	0.1221	0.0610	New York	Niagara	138	1	0.0977	0.0488
Kansas	Sedgwick	53	1	0.4883	0.2441	New York	Rensselaer	125	1	0.0977	0.0488
Louisiana	Ascension Parish	125	1	0.0000	0.0244	New York	Richmond	2497	19	0.1028	0.0121
Louisiana	Bossier Parish	263	1	0.0000	0.0244	North Dakota	Cass	60	1	0.0244	0.0122
Louisiana	Calcasieu Parish	208	1	0.1465	0.0732	North Dakota	Mercer	56	1	0.0244	0.0122
Louisiana	East Baton Rouge Parish	283	3	0.0651	0.0244	Ohio	Hamilton	36	1	0.1441	0.0721
Louisiana	Ouachita Parish	104	1	0.0000	0.0244	Oregon	Multnomah	531	10	0.1645	0.0291
Louisiana	St. Charles Parish	112	1	0.0244	0.0122	Oregon	Washington	55	1	0.1000	0.0500
Louisiana	St. John the Baptist Parish	112	1	0.0244	0.0122	Pennsylvania	Delaware	1369	3	0.1308	0.0402
Maryland	Anne Arundel	406	2	0.1748	0.0125	Pennsylvania	Philadelphia	269	1	0.1953	0.0977

Table A.12 Summary of Countywide Chloroform Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Puerto Rico	Barceloneta	37	1	0.1220	0.0610	Texas	Harris	2514	14	0.3383	0.0640
Puerto Rico	San Juan	26	1	0.1220	0.0610	Texas	Hidalgo	422	2	0.2441	0.0863
Rhode Island	Kent	333	3	0.1050	0.0311	Texas	Jefferson	1543	7	0.2093	0.0429
Rhode Island	Providence	902	7	0.1012	0.0183	Texas	Nueces	955	4	0.2441	0.0610
South Dakota	Minnehaha	41	1	0.0244	0.0122	Texas	Orange	276	1	0.2441	0.1221
Tennessee	Davidson	30	1	0.0244	0.0122	Texas	Smith	89	1	1.5624	0.7812
Texas	Bexar	261	1	0.2441	0.1221	Texas	Tarrant	242	4	0.8483	0.2779
Texas	Brazoria	321	1	0.2441	0.1221	Texas	Travis	208	3	0.2441	0.0705
Texas	Cameron	359	2	0.1343	0.0613	Texas	Webb	244	1	0.2441	0.1221
Texas	Carson	834	8	0.0350	0.0067	Utah	Salt Lake	57	1	0.0244	0.0122
Texas	Dallas	466	3	0.6836	0.2667	Vermont	Chittenden	260	2	0.0122	0.0136
Texas	Denton	28	1	0.2441	0.1221	Vermont	Rutland	126	1	0.0000	0.0244
Texas	Ector	250	1	0.2441	0.1221	Vermont	Windham	105	1	0.0000	0.0244
Texas	El Paso	1424	9	0.2197	0.0384	Virginia	Fairfax	117	1	0.1465	0.0732
Texas	Ellis	330	2	0.9033	0.3953	Washington	Asotin	48	1	0.1158	0.0579
Texas	Galveston	670	4	0.1831	0.0532	Washington	King	492	6	0.1375	0.0274
Texas	Gregg	161	1	0.2441	0.1221	Washington	Whatcom	60	2	0.0425	0.0245

Table A.13 Summary of Countywide CIS 1,3-Dichloropropene Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Colorado	Mesa	163	2	0.2268	0.0802	Michigan	Washtenaw	33	1	0.6805	0.3403
Florida	Hillsborough	26	3	0.0983	0.0290	Michigan	Wayne	564	5	1.2703	0.4649
Florida	Pinellas	33	3	0.1134	0.0327	Puerto Rico	Barceloneta	37	1	0.2268	0.1134
Georgia	Fayette	207	2	0.2268	0.0802	Puerto Rico	San Juan	26	1	0.2268	0.1134
Michigan	Kent	38	2	0.6805	0.2406	Rhode Island	Kent	59	1	0.0295	0.0147
Michigan	Missaukee	30	1	0.6805	0.3403	Rhode Island	Providence	360	5	0.0295	0.0066
Michigan	Oakland	43	1	0.2268	0.1134						

Table A.14 Summary of Countywide Trans 1,3-Dichloropropene Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Colorado	Mesa	163	2	0.2495	0.0882	Michigan	Washtenaw	33	1	0.6805	0.3403
Florida	Hillsborough	18	3	0.2193	0.0645	Michigan	Wayne	564	5	1.2794	0.4651
Florida	Pinellas	24	3	0.2495	0.0720	Puerto Rico	Barceloneta	37	1	0.2495	0.1248
Georgia	Fayette	207	2	0.2495	0.0882	Puerto Rico	San Juan	26	1	0.2495	0.1248
Michigan	Kent	38	2	0.6805	0.2406	Rhode Island	Kent	59	1	0.0363	0.0181
Michigan	Missaukee	30	1	0.6805	0.3403	Rhode Island	Providence	360	5	0.0363	0.0081
Michigan	Oakland	43	1	0.2495	0.1248						

Table A.15 Summary of Countywide Ethylene Oxide Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Rhode Island	Kent	87	1	0.1243	0.0621	Rhode Island	Providence	524	5	0.1545	0.0377

Table A.16 Summary of Countywide Methylene Chloride Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.4863	0.2431	Louisiana	St. John the Baptist Parish	112	1	0.1910	0.0955
Arkansas	Pulaski	51	1	0.1910	0.0955	Maryland	Anne Arundel	406	2	0.2002	0.0091
California	Alameda	547	4	0.7868	0.2108	Maryland	Baltimore	347	1	0.3632	0.0074
California	Butte	176	1	1.0421	0.5210	Maryland	Baltimore city	1398	6	0.3017	0.0056
California	Contra Costa	717	8	1.0421	0.1905	Maryland	Harford	109	1	0.3474	0.1737
California	Fresno	170	1	1.0421	0.5210	Michigan	Alpena	130	5	0.2765	0.0646
California	Humboldt	23	1	1.0421	0.5210	Michigan	Kalamazoo	673	4	0.0912	0.0753
California	Imperial	167	1	1.0421	0.5210	Michigan	Kent	162	3	1.2951	0.4950
California	Kern	375	2	1.0421	0.3684	Michigan	Midland	1494	4	0.3450	0.0863
California	Los Angeles	717	10	0.8031	0.1369	Michigan	Missaukee	86	2	0.3473	0.1373
California	Marin	222	2	0.6184	0.2505	Michigan	Oakland	41	1	0.2357	0.0704
California	Napa	120	1	0.8684	0.4342	Michigan	Van Buren	52	1	0.1910	0.0955
California	Orange	22	1	0.4947	0.2474	Michigan	Washtenaw	50	2	0.3473	0.1373
California	Placer	167	1	1.0421	0.5210	Michigan	Wayne	831	7	0.9400	0.2949
California	Riverside	474	5	0.6602	0.1609	Minnesota	Beltrami	54	1	0.0000	0.0174
California	San Bernardino	165	3	0.8596	0.2591	Minnesota	Clay	51	1	0.0000	0.0174
California	San Diego	399	3	1.0421	0.3008	Minnesota	Crow Wing	32	1	0.0174	0.0087
California	San Francisco	178	1	1.0421	0.5210	Minnesota	Dakota	784	8	0.0204	0.0106
California	San Joaquin	190	1	1.0421	0.5210	Minnesota	Douglas	88	2	0.0834	0.0376
California	San Mateo	119	1	0.0472	0.1253	Minnesota	Freeborn	53	1	0.0000	0.0174
California	Santa Barbara	130	1	1.0421	0.5210	Minnesota	Goodhue	53	1	0.3091	0.1546
California	Santa Clara	300	2	0.6002	0.2622	Minnesota	Hennepin	659	6	0.0932	0.0265
California	Solano	120	1	0.2186	0.0891	Minnesota	Itasca	22	1	0.0174	0.0087
California	Sonoma	119	1	0.8684	0.4342	Minnesota	Kandiyohi	35	1	0.0174	0.0087
California	Stanislaus	139	1	1.7368	0.8684	Minnesota	Koochiching	139	2	0.1799	0.0408
California	Ventura	160	1	1.0421	0.5210	Minnesota	Lake	45	1	0.0000	0.0174
Colorado	Denver	19	1	0.1910	0.0955	Minnesota	McLeod	56	1	0.2418	0.0217
Colorado	Mesa	163	2	0.1389	0.0582	Minnesota	Morrison	55	1	0.1216	0.0680
Delaware	New Castle	55	1	0.2358	0.0181	Minnesota	Nicollet	21	1	0.0174	0.0087
District of Columbia	District of Columbia	169	1	0.2696	0.0076	Minnesota	Olmsted	55	1	0.2049	0.1025
Florida	Broward	193	6	0.2532	0.0689	Minnesota	Otter Tail	100	2	0.1042	0.0528
Florida	Hillsborough	169	3	0.4282	0.1240	Minnesota	Pine	50	1	0.1980	0.0990
Florida	Pinellas	677	7	0.2632	0.0561	Minnesota	Pipestone	46	1	0.1667	0.0834
Georgia	Dawson	25	1	0.8500	0.4250	Minnesota	Ramsey	513	3	0.1401	0.0554
Georgia	Fayette	207	2	0.0521	0.0582	Minnesota	Roseau	46	1	0.1598	0.0799
Georgia	Fulton	22	1	1.2950	0.6475	Minnesota	Sherburne	57	1	0.1077	0.0538
Georgia	Glynn	20	1	0.8500	0.4250	Minnesota	St. Louis	547	5	0.1390	0.0279
Georgia	Hall	20	1	0.8500	0.4250	Minnesota	Stearns	51	1	0.0000	0.0174
Indiana	Lake	277	2	0.0000	0.0405	Minnesota	Swift	53	1	0.0000	0.0174
Indiana	Porter	22	1	0.0573	0.0287	Minnesota	Washington	522	3	0.0000	0.0100
Iowa	Linn	36	1	0.1042	0.0521	Minnesota	Winona	50	1	0.0000	0.0174
Kansas	Sedgwick	40	1	1.2500	0.6250	Minnesota	Wright	59	1	0.0000	0.0174
Kentucky	Boyd	26	1	1.0000	0.5000	Minnesota	Yellow Medicine	42	1	0.1112	0.0556
Kentucky	Lawrence	23	1	1.0000	0.5000	New Jersey	Camden	164	1	0.0194	0.0464
Louisiana	Ascension Parish	85	1	0.2779	0.1910	New Jersey	Union	22	1	0.7468	0.3734
Louisiana	Bossier Parish	279	1	0.0000	0.1910	New Mexico	Sandoval	40	2	1.2153	0.4340
Louisiana	Calcasieu Parish	272	1	0.0000	0.1910	New York	Erie	150	1	0.3474	0.1737
Louisiana	East Baton Rouge Parish	283	3	0.1274	0.0780	New York	Essex	145	2	0.3474	0.1228
Louisiana	Ouachita Parish	112	1	0.0000	0.1910	New York	Kings	135	1	0.3474	0.1737
Louisiana	St. Charles Parish	112	1	0.1910	0.0955	New York	Monroe	159	2	0.3474	0.1228

Table A.16 Summary of Countywide Methylene Chloride Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
New York	Niagara	137	1	0.3474	0.1737	Texas	Denton	43	1	0.1737	0.0868
New York	Rensselaer	122	1	0.3474	0.1737	Texas	Ector	250	1	0.1737	0.0868
New York	Richmond	2502	19	0.3052	0.0369	Texas	El Paso	1493	9	0.1281	0.0587
North Dakota	Cass	60	1	0.1910	0.0955	Texas	Ellis	344	2	0.4342	0.1790
North Dakota	Mercer	56	1	0.0000	0.1910	Texas	Galveston	710	4	0.1520	0.0583
Ohio	Butler	30	1	0.2217	0.1179	Texas	Gregg	175	1	0.1737	0.0868
Ohio	Cuyahoga	106	2	0.1661	0.0261	Texas	Harris	2690	14	0.1056	0.0469
Ohio	Hamilton	341	4	0.5346	0.1306	Texas	Hidalgo	449	2	0.1737	0.0614
Oregon	Multnomah	529	10	0.1268	0.0290	Texas	Jefferson	1631	7	0.1762	0.0333
Oregon	Washington	31	1	0.1000	0.0500	Texas	Nueces	1069	4	0.1737	0.0434
Pennsylvania	Delaware	1380	3	0.2511	0.0355	Texas	Orange	291	1	0.1737	0.0868
Pennsylvania	Philadelphia	269	1	0.3885	0.0127	Texas	Smith	89	1	0.6947	0.3474
Puerto Rico	Barceloneta	37	1	1.0417	0.2757	Texas	Tarrant	242	4	0.4342	0.1317
Puerto Rico	San Juan	26	1	0.4514	0.2257	Texas	Travis	222	3	0.1737	0.0501
Rhode Island	Kent	333	3	0.0972	0.0577	Texas	Webb	258	1	0.1737	0.0868
Rhode Island	Providence	902	7	0.1948	0.0390	Utah	Salt Lake	57	1	0.4168	0.2084
South Dakota	Minnehaha	41	1	0.1910	0.0955	Vermont	Chittenden	260	2	0.1910	0.0675
Tennessee	Davidson	30	1	0.1910	0.0955	Vermont	Rutland	126	1	0.1910	0.0955
Texas	Bexar	273	1	0.1737	0.0868	Vermont	Windham	105	1	0.0000	0.1910
Texas	Brazoria	336	1	0.0115	0.0436	Virginia	Fairfax	117	1	0.3821	0.1910
Texas	Cameron	374	2	0.1824	0.0645	Washington	Whatcom	60	2	1.0955	2.0621
Texas	Carson	834	8	0.0157	0.0069	West Virginia	Wayne	120	5	1.0000	0.2236
Texas	Dallas	495	3	0.2316	0.1418						

Table A.17 Summary of Countywide Tetrachloroethylene Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.2374	0.1187	Maryland	Baltimore	347	1	0.4069	0.2035
Arkansas	Pulaski	51	1	0.2374	0.1187	Maryland	Baltimore city	1398	6	0.3460	0.0676
California	Alameda	169	1	0.3391	0.1696	Maryland	Harford	109	1	0.2035	0.1017
California	Butte	155	1	0.3391	0.1696	Michigan	Alpena	130	5	0.5406	0.1264
California	Contra Costa	391	4	0.3391	0.0848	Michigan	Kalamazoo	660	4	0.2374	0.0593
California	Fresno	152	1	0.3391	0.1696	Michigan	Kent	163	3	0.7910	0.2463
California	Humboldt	23	1	0.3391	0.1696	Michigan	Midland	1464	4	0.6750	0.1688
California	Imperial	148	1	0.3391	0.1696	Michigan	Missaukee	87	2	0.6780	0.2680
California	Kern	335	2	0.3391	0.1199	Michigan	Oakland	43	1	0.2034	0.1017
California	Los Angeles	744	15	0.7912	0.1221	Michigan	Van Buren	49	1	0.2374	0.1187
California	Marin	104	1	0.2513	0.1256	Michigan	Washtenaw	49	2	0.6780	0.2680
California	Napa	120	1	0.1156	0.0578	Michigan	Wayne	838	7	1.5303	0.5017
California	Orange	40	2	0.7021	0.3393	Minnesota	Beltrami	54	1	0.0339	0.0170
California	Placer	150	1	0.3391	0.1696	Minnesota	Clay	51	1	0.0339	0.0170
California	Riverside	411	6	0.6443	0.1545	Minnesota	Dakota	339	6	0.0577	0.0158
California	San Bernardino	185	4	0.7419	0.2226	Minnesota	Douglas	57	1	0.3730	0.0003
California	San Diego	362	3	0.3391	0.0979	Minnesota	Goodhue	53	1	0.2102	0.0267
California	San Francisco	159	1	0.3391	0.1696	Minnesota	Hennepin	427	4	0.1306	0.0271
California	San Joaquin	169	1	0.3391	0.1696	Minnesota	Koochiching	139	2	0.1255	0.0587
California	Santa Barbara	129	1	0.3391	0.1696	Minnesota	Morrison	55	1	0.2035	0.1017
California	Santa Clara	165	1	0.3391	0.1696	Minnesota	Olmsted	55	1	0.1560	0.0780
California	Stanislaus	134	1	0.3391	0.1696	Minnesota	Otter Tail	47	1	0.0248	0.0422
California	Ventura	148	1	0.3391	0.1696	Minnesota	Pine	50	1	0.1153	0.0576
Colorado	Denver	19	1	0.2374	0.1187	Minnesota	Pipestone	46	1	0.1560	0.0780
Colorado	Mesa	163	2	0.2034	0.0719	Minnesota	Ramsey	378	3	0.0113	0.0170
Delaware	New Castle	55	1	0.1368	0.0443	Minnesota	Roseau	46	1	0.1492	0.0746
District of Columbia	District of Columbia	169	1	0.3717	0.0213	Minnesota	Sherburne	57	1	0.1696	0.0848
Florida	Broward	134	6	0.8734	0.2873	Minnesota	St. Louis	377	3	0.1572	0.0168
Florida	Hillsborough	160	3	0.2034	0.0587	Minnesota	Stearns	51	1	0.0339	0.0170
Florida	Pinellas	673	7	0.1665	0.0367	Minnesota	Swift	53	1	0.0339	0.0170
Georgia	Dawson	25	1	1.7000	0.8500	Minnesota	Washington	301	2	0.0000	0.0240
Georgia	Fayette	207	2	0.2034	0.0719	Minnesota	Winona	50	1	0.0339	0.0170
Georgia	Fulton	22	1	1.7000	0.8500	Minnesota	Yellow Medicine	42	1	0.1831	0.0916
Georgia	Glynn	20	1	1.7000	0.8500	New Jersey	Camden	164	1	0.2374	0.1187
Georgia	Hall	21	1	1.7000	0.8500	New Jersey	Union	22	1	0.2374	0.1187
Indiana	Lake	277	2	0.0280	0.0577	New Mexico	Sandoval	40	2	0.3051	0.1079
Indiana	Porter	22	1	0.0560	0.0280	New York	Erie	193	1	0.1356	0.0678
Iowa	Linn	36	1	0.2035	0.1017	New York	Essex	177	2	0.1356	0.0480
Kansas	Sedgwick	53	1	0.7460	0.3730	New York	Kings	180	1	0.9909	1.1454
Kentucky	Boyd	26	1	1.0000	0.5000	New York	Monroe	206	2	0.0000	0.0959
Kentucky	Lawrence	23	1	1.0000	0.5000	New York	Niagara	132	1	0.0000	0.1356
Louisiana	Ascension Parish	125	1	0.2374	0.1187	New York	Rensselaer	169	1	0.0000	0.1356
Louisiana	Bossier Parish	263	1	0.2374	0.1187	New York	Richmond	2406	19	0.1844	0.0736
Louisiana	Calcasieu Parish	258	1	0.0000	0.2374	North Dakota	Cass	60	1	0.2374	0.1187
Louisiana	East Baton Rouge Parish	283	3	0.2374	0.0685	North Dakota	Mercer	56	1	0.2374	0.1187
Louisiana	Ouachita Parish	104	1	0.2374	0.1187	Ohio	Cuyahoga	51	2	0.4741	0.1676
Louisiana	St. Charles Parish	112	1	0.2374	0.1187	Ohio	Hamilton	81	3	0.3652	0.1244
Louisiana	St. John the Baptist Parish	112	1	0.2374	0.1187	Oregon	Multnomah	529	10	0.2033	0.0379
Maine	Oxford	22	1	3.4000	1.7000	Oregon	Washington	55	1	0.1000	0.0500
Maryland	Anne Arundel	406	2	0.2299	0.1020	Pennsylvania	Delaware	1348	3	0.2381	0.0755

Table A.17 Summary of Countywide Tetrachloroethylene Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Pennsylvania	Philadelphia	269	1	0.3665	0.0205	Texas	Harris	2514	14	0.3052	0.0770
Puerto Rico	Barceloneta	37	1	0.2034	0.1017	Texas	Hidalgo	422	2	0.1696	0.0599
Puerto Rico	San Juan	26	1	0.2034	0.1017	Texas	Jefferson	1541	7	0.1550	0.0401
Rhode Island	Kent	331	3	0.1379	0.0419	Texas	Nueces	955	4	0.1696	0.0424
Rhode Island	Providence	902	7	0.2782	0.0517	Texas	Orange	276	1	0.1696	0.0848
South Dakota	Minnehaha	41	1	0.2374	0.1187	Texas	Smith	89	1	2.0686	1.0343
Tennessee	Davidson	30	1	0.2374	0.1187	Texas	Tarrant	242	4	1.1360	0.3675
Texas	Bexar	261	1	0.1696	0.0848	Texas	Travis	208	3	0.1696	0.0489
Texas	Brazoria	321	1	0.1696	0.0848	Texas	Webb	244	1	0.1696	0.0848
Texas	Cameron	359	2	0.2035	0.0729	Utah	Salt Lake	57	1	0.2374	0.1187
Texas	Carson	834	8	0.0334	0.0065	Vermont	Chittenden	260	2	0.2374	0.0839
Texas	Dallas	466	3	0.7460	0.3505	Vermont	Rutland	126	1	0.2374	0.1187
Texas	Denton	28	1	0.1696	0.0848	Vermont	Windham	105	1	0.2374	0.1187
Texas	Ector	250	1	0.1696	0.0848	Virginia	Fairfax	117	1	0.0886	0.0256
Texas	El Paso	1424	9	0.1771	0.0297	Washington	King	492	6	0.1415	0.0328
Texas	Ellis	330	2	1.1191	0.5189	Washington	Whatcom	60	2	0.1534	0.0618
Texas	Galveston	670	4	0.1865	0.0472	West Virginia	Wayne	120	5	1.0000	0.2236
Texas	Gregg	161	1	0.1696	0.0848						

Table A.18 Summary of Countywide Trichloroethylene TCE Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Alameda	530	4	0.2284	0.0574	Minnesota	Beltrami	54	1	0.0537	0.0269
California	Butte	157	1	0.2687	0.1343	Minnesota	Clay	51	1	0.0537	0.0269
California	Contra Costa	713	8	0.2318	0.0417	Minnesota	Dakota	334	5	0.0430	0.0152
California	Fresno	152	1	0.2687	0.1343	Minnesota	Douglas	57	1	0.0860	0.0632
California	Humboldt	23	1	0.2687	0.1343	Minnesota	Goodhue	53	1	0.3869	0.1935
California	Imperial	150	1	0.2687	0.1343	Minnesota	Hennepin	427	4	0.2432	0.0741
California	Kern	368	3	0.1970	0.0640	Minnesota	Koochiching	139	2	0.0699	0.0300
California	Los Angeles	583	7	0.2671	0.0811	Minnesota	Morrison	55	1	0.2483	0.0777
California	Marin	223	2	0.2149	0.0760	Minnesota	Olmsted	55	1	0.1935	0.0967
California	Napa	112	1	0.2149	0.1075	Minnesota	Otter Tail	47	1	0.1559	0.0708
California	Orange	20	1	0.5374	0.2687	Minnesota	Pine	50	1	0.1827	0.0914
California	Placer	150	1	0.2687	0.1343	Minnesota	Pipestone	46	1	0.2203	0.1031
California	Riverside	277	4	0.2686	0.0672	Minnesota	Ramsey	377	3	0.0358	0.0310
California	San Bernardino	138	2	0.2687	0.0950	Minnesota	Roseau	46	1	0.0547	0.1370
California	San Diego	363	3	0.2687	0.0776	Minnesota	Sherburne	57	1	0.2987	0.0487
California	San Francisco	162	1	0.2687	0.1343	Minnesota	St. Louis	377	3	0.0967	0.0127
California	San Joaquin	171	1	0.2687	0.1343	Minnesota	Stearns	51	1	0.0537	0.0269
California	San Mateo	120	1	0.1349	0.0675	Minnesota	Swift	53	1	0.0537	0.0269
California	Santa Barbara	130	1	0.2687	0.1343	Minnesota	Washington	304	2	0.0000	0.0380
California	Santa Clara	278	2	0.2418	0.0860	Minnesota	Winona	50	1	0.0537	0.0269
California	Solano	120	1	0.2149	0.1075	Minnesota	Yellow Medicine	42	1	0.0537	0.0269
California	Sonoma	119	1	0.1349	0.0675	New Mexico	Sandoval	40	2	0.2417	0.0855
California	Stanislaus	136	1	0.2687	0.1343	New York	Erie	202	1	0.0537	0.0269
California	Ventura	148	1	0.2687	0.1343	New York	Essex	190	2	0.0537	0.0190
Colorado	Mesa	163	2	0.1880	0.0665	New York	Kings	182	1	0.1075	0.0537
Delaware	New Castle	55	1	0.1612	0.0806	New York	Monroe	221	2	0.1069	0.0598
District of Columbia	District of Columbia	169	2	0.0537	0.0190	New York	Niagara	136	1	0.0000	0.0537
Florida	Broward	52	2	0.1075	0.0380	New York	Rensselaer	176	1	0.0537	0.0269
Florida	Hillsborough	47	3	0.1611	0.0465	New York	Richmond	2440	19	0.0534	0.0084
Florida	Pinellas	571	7	0.1450	0.0551	Ohio	Cuyahoga	30	1	0.0687	0.0500
Georgia	Dawson	25	1	1.3500	0.6750	Ohio	Hamilton	25	1	0.2687	0.1343
Georgia	Fayette	207	2	0.1880	0.0665	Oregon	Multnomah	499	9	0.1937	0.0352
Georgia	Fulton	22	1	1.3500	0.6750	Oregon	Washington	55	1	0.1000	0.0500
Georgia	Glynn	20	1	1.3500	0.6750	Pennsylvania	Delaware	1376	3	0.1355	0.0397
Georgia	Hall	21	1	1.3500	0.6750	Pennsylvania	Philadelphia	269	1	0.1612	0.0806
Indiana	Lake	277	2	0.0000	0.0150	Puerto Rico	Barceloneta	37	1	0.1880	0.0940
Indiana	Porter	22	1	0.0134	0.0067	Puerto Rico	San Juan	26	1	0.1880	0.0940
Iowa	Linn	36	1	0.1881	0.0940	Rhode Island	Kent	308	3	0.1146	0.0382
Kansas	Sedgwick	53	1	0.5374	0.2687	Rhode Island	Providence	887	7	0.1800	0.0390
Maine	Oxford	22	1	2.7000	1.3500	Texas	Bexar	261	1	0.1343	0.0672
Maryland	Anne Arundel	406	2	0.0585	0.0145	Texas	Brazoria	321	1	0.0000	0.1343
Maryland	Baltimore	347	1	0.0765	0.0088	Texas	Cameron	312	1	0.1343	0.0672
Maryland	Baltimore city	1398	6	0.1149	0.0220	Texas	Carson	834	8	0.0365	0.0080
Maryland	Harford	109	1	0.0537	0.0269	Texas	Dallas	466	3	0.5822	0.2733
Michigan	Alpena	130	5	0.2695	0.0603	Texas	Denton	28	1	0.1343	0.0672
Michigan	Kent	163	3	0.5372	0.2100	Texas	Ector	250	1	0.1343	0.0672
Michigan	Missaukee	87	2	0.5372	0.2123	Texas	El Paso	1296	8	0.1343	0.0237
Michigan	Oakland	43	1	0.1880	0.0940	Texas	Ellis	330	2	0.8732	0.4044
Michigan	Washtenaw	50	2	0.5372	0.2123	Texas	Galveston	613	3	0.1343	0.0388
Michigan	Wayne	869	7	1.2163	0.3976	Texas	Gregg	161	1	0.1343	0.0672

Table A.18 Summary of Countywide Trichloroethylene TCE Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Texas	Harris	2514	14	0.2399	0.0601	Texas	Tarrant	215	3	1.1195	0.3806
Texas	Hidalgo	422	2	0.1343	0.0475	Texas	Travis	208	3	0.1343	0.0388
Texas	Jefferson	1482	6	0.1343	0.0274	Texas	Webb	244	1	0.1343	0.0672
Texas	Nueces	955	4	0.1343	0.0336	Virginia	Fairfax	117	1	0.0537	0.0269
Texas	Orange	276	1	0.1343	0.0672	Washington	King	492	6	0.1603	0.0319
Texas	Smith	89	1	1.6121	0.8061	Washington	Whatcom	30	1	0.1343	0.0672

Table A.19 Summary of Countywide Vinyl Chloride Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Arkansas	Miller	41	1	0.2556	0.1278	Minnesota	Douglas	89	2	0.0256	0.0090
Arkansas	Pulaski	51	1	0.2556	0.1278	Minnesota	Freeborn	53	1	0.0256	0.0128
California	Alameda	463	4	0.3834	0.0959	Minnesota	Goodhue	53	1	0.0256	0.0128
California	Contra Costa	454	6	0.3834	0.0783	Minnesota	Hennepin	661	6	0.0256	0.0052
California	Marin	223	2	0.3834	0.1356	Minnesota	Itasca	23	1	0.0256	0.0128
California	Napa	120	1	0.3834	0.1917	Minnesota	Kandiyohi	36	1	0.0256	0.0128
California	San Francisco	121	1	0.3834	0.1917	Minnesota	Koochiching	139	2	0.0256	0.0090
California	San Mateo	120	1	0.3834	0.1917	Minnesota	Lake	45	1	0.0256	0.0128
California	Santa Clara	239	2	0.3834	0.1356	Minnesota	McLeod	56	1	0.0256	0.0128
California	Solano	120	1	0.3834	0.1917	Minnesota	Morrison	55	1	0.0256	0.0128
California	Sonoma	119	1	0.3834	0.1917	Minnesota	Nicollet	23	1	0.0256	0.0128
Colorado	Denver	19	1	0.2556	0.1278	Minnesota	Olmsted	55	1	0.0256	0.0128
Colorado	Mesa	163	2	0.0767	0.0271	Minnesota	Otter Tail	100	2	0.0256	0.0090
Delaware	New Castle	55	1	0.0256	0.0128	Minnesota	Pine	50	1	0.0256	0.0128
District of Columbia	District of Columbia	169	1	0.0256	0.0128	Minnesota	Pipestone	46	1	0.0256	0.0128
Florida	Broward	52	2	0.0383	0.0136	Minnesota	Ramsey	514	3	0.0256	0.0074
Florida	Hillsborough	85	3	0.1022	0.0295	Minnesota	Roseau	46	1	0.0256	0.0128
Florida	Pinellas	516	7	0.0679	0.0135	Minnesota	Sherburne	57	1	0.0256	0.0128
Georgia	Dawson	25	1	0.6500	0.3250	Minnesota	St. Louis	549	5	0.0256	0.0057
Georgia	Fayette	207	2	0.0767	0.0271	Minnesota	Stearns	51	1	0.0256	0.0128
Georgia	Fulton	22	1	0.6500	0.3250	Minnesota	Swift	53	1	0.0256	0.0128
Georgia	Glynn	20	1	0.6500	0.3250	Minnesota	Washington	523	3	0.0256	0.0074
Georgia	Hall	21	1	0.6500	0.3250	Minnesota	Winona	50	1	0.0256	0.0128
Indiana	Lake	277	2	0.0316	0.0118	Minnesota	Wright	59	1	0.0256	0.0128
Indiana	Porter	22	1	0.0211	0.0105	Minnesota	Yellow Medicine	42	1	0.0256	0.0128
Iowa	Linn	36	1	0.0767	0.0383	New Jersey	Camden	164	1	0.2556	0.1278
Kansas	Sedgwick	54	1	0.4473	0.2237	New Jersey	Union	22	1	0.2556	0.1278
Louisiana	Ascension Parish	141	1	0.2556	0.1278	New Mexico	Sandoval	40	2	0.3833	0.1355
Louisiana	Bossier Parish	279	1	0.2556	0.1278	New York	Erie	155	1	0.1022	0.0511
Louisiana	Calcasieu Parish	272	1	0.2556	0.1278	New York	Essex	143	2	0.1022	0.0361
Louisiana	East Baton Rouge Parish	283	3	0.2556	0.0738	New York	Kings	141	1	0.1022	0.0511
Louisiana	Ouachita Parish	112	1	0.2556	0.1278	New York	Monroe	147	2	0.1022	0.0361
Louisiana	St. Charles Parish	112	1	0.2556	0.1278	New York	Niagara	90	1	0.1022	0.0511
Louisiana	St. John the Baptist Parish	112	1	0.2556	0.1278	New York	Rensselaer	136	1	0.1022	0.0511
Maine	Oxford	25	1	1.3000	0.6500	New York	Richmond	2095	19	0.1170	0.0135
Maryland	Anne Arundel	402	2	0.0256	0.0090	North Dakota	Cass	60	1	0.2556	0.1278
Maryland	Baltimore	347	1	0.0256	0.0128	North Dakota	Mercer	56	1	0.2556	0.1278
Maryland	Baltimore city	1393	6	0.0256	0.0052	Ohio	Hamilton	25	1	0.1278	0.0639
Maryland	Harford	109	1	0.0256	0.0128	Oregon	Multnomah	531	10	0.1295	0.0216
Michigan	Alpena	131	5	0.3511	0.0885	Oregon	Washington	55	1	0.1000	0.0500
Michigan	Kent	163	3	0.2981	0.0928	Pennsylvania	Delaware	945	3	0.0056	0.0058
Michigan	Midland	1549	4	0.5000	0.1250	Pennsylvania	Philadelphia	266	1	0.0000	0.0256
Michigan	Missaukee	87	2	0.2555	0.1010	Puerto Rico	Barceloneta	37	1	0.0767	0.0383
Michigan	Oakland	43	1	0.0767	0.0383	Puerto Rico	San Juan	26	1	0.0767	0.0383
Michigan	Washtenaw	54	2	0.2555	0.1010	Rhode Island	Kent	59	1	0.0141	0.0070
Michigan	Wayne	877	7	0.5330	0.1872	Rhode Island	Providence	360	5	0.0141	0.0031
Minnesota	Beltrami	54	1	0.0256	0.0128	South Dakota	Minnehaha	41	1	0.2556	0.1278
Minnesota	Clay	51	1	0.0256	0.0128	Tennessee	Davidson	30	1	0.2556	0.1278
Minnesota	Crow Wing	33	1	0.0256	0.0128	Texas	Bexar	273	1	0.1278	0.0639
Minnesota	Dakota	789	8	0.0256	0.0045	Texas	Brazoria	336	1	0.1278	0.0639

Table A.19 Summary of Countywide Vinyl Chloride Background Estimates (continued)

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Texas	Cameron	374	2	0.1917	0.0714	Texas	Nueces	1069	4	0.1118	0.0288
Texas	Carson	834	8	0.0221	0.0048	Texas	Orange	291	1	0.1278	0.0639
Texas	Dallas	495	3	0.2769	0.1005	Texas	Smith	89	1	0.5751	0.2876
Texas	Denton	43	1	0.1278	0.0639	Texas	Tarrant	242	4	0.3675	0.1069
Texas	Ector	250	1	0.0639	0.0320	Texas	Travis	222	3	0.0852	0.0261
Texas	El Paso	1493	9	0.1207	0.0222	Texas	Webb	258	1	0.1278	0.0639
Texas	Ellis	344	2	0.3515	0.1473	Utah	Salt Lake	57	1	0.2556	0.1278
Texas	Galveston	710	4	0.1598	0.0423	Vermont	Chittenden	260	2	0.2556	0.0904
Texas	Gregg	175	1	0.1278	0.0639	Vermont	Rutland	126	1	0.2556	0.1278
Texas	Harris	2690	14	0.1552	0.0260	Vermont	Windham	105	1	0.2556	0.1278
Texas	Hidalgo	449	2	0.1278	0.0452	Virginia	Fairfax	117	1	0.0256	0.0128
Texas	Jefferson	1630	7	0.1461	0.0289	Washington	Whatcom	60	2	0.1981	0.0729

Table A.20 Summary of Countywide Arsenic (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0010	0.0004	California	Sacramento	385	1	0.0010	0.0005
California	Imperial	378	3	0.0010	0.0003	California	San Bernardino	260	3	1.1670	0.4167
California	Inyo	432	2	0.0010	0.0004	California	San Joaquin	343	1	0.0010	0.0005
California	Kern	1176	3	0.0010	0.0003	California	Santa Clara	280	1	0.0010	0.0005
California	Kings	402	2	0.0010	0.0004	California	Stanislaus	531	4	0.0010	0.0003
California	Los Angeles	763	13	1.4617	0.2226	California	Tulare	307	1	0.0010	0.0005
California	Madera	104	1	0.0010	0.0005	Oregon	Multnomah	89	4	0.0042	0.0012
California	Orange	43	2	1.7500	0.6250	Oregon	Washington	28	1	0.0296	0.0148
California	Plumas	117	1	0.0010	0.0005	Washington	King	93	2	0.0004	0.0002
California	Riverside	143	3	1.1670	0.4167						

Table A.21 Summary of Countywide Cadmium (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Washington	King	53	2	0.0015	0.0006						

Table A.22 Summary of Countywide Chromium (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Alabama	Lawrence	369	1	0.2850	0.1425	Idaho	Custer	465	1	0.2850	0.1425
Alaska	Yukon-Koyukuk Census Area	466	1	0.2850	0.1425	Idaho	Lemhi	387	1	0.2850	0.1425
Arizona	Apache	428	1	0.2850	0.1425	Kentucky	Edmonson	469	1	0.2850	0.1425
Arizona	Cochise	462	1	0.2850	0.1425	Maine	Hancock	470	1	0.2850	0.1425
Arizona	Coconino	856	3	0.2850	0.0823	Maine	Washington	449	1	0.2850	0.1425
Arizona	Gila	461	1	0.2850	0.1425	Minnesota	Lake	345	1	0.2850	0.1425
Arkansas	Newton	472	1	0.2850	0.1425	Montana	Flathead	478	1	0.2850	0.1425
California	Del Norte	443	1	0.2850	0.1425	Montana	Ravalli	471	1	0.2850	0.1425
California	El Dorado	404	2	0.2850	0.1008	Nevada	Elko	409	1	0.2850	0.1425
California	Fresno	693	2	0.0005	0.0002	Nevada	White Pine	475	1	0.2850	0.1425
California	Imperial	378	3	0.0005	0.0001	New Hampshire	Coos	171	1	0.2850	0.1425
California	Inyo	896	3	0.0953	0.0475	New Jersey	Atlantic	445	1	0.2850	0.1425
California	Kern	1176	3	0.0005	0.0001	New Mexico	Catron	25	1	0.2850	0.1425
California	Kings	402	2	0.0005	0.0002	New Mexico	Dona Ana	174	1	0.2850	0.1425
California	Los Angeles	763	13	0.8462	0.1276	New Mexico	Los Alamos	457	1	0.2850	0.1425
California	Madera	104	1	0.0005	0.0003	North Carolina	Haywood	405	1	0.2850	0.1425
California	Marin	388	1	0.2850	0.1425	North Carolina	Swain	477	1	0.2850	0.1425
California	Mariposa	462	1	0.2850	0.1425	Oregon	Klamath	455	1	0.2850	0.1425
California	Orange	43	2	1.0000	0.3536	Oregon	Linn	467	1	0.2850	0.1425
California	Plumas	117	1	0.0005	0.0003	Oregon	Multnomah	91	4	0.0021	0.0007
California	Riverside	143	3	0.6668	0.2357	Oregon	Washington	28	1	0.0037	0.0019
California	Sacramento	385	1	0.0005	0.0003	South Carolina	Charleston	433	1	0.2850	0.1425
California	San Benito	438	1	0.2850	0.1425	South Dakota	Jackson	477	1	0.2850	0.1425
California	San Bernardino	645	4	0.5714	0.1803	Texas	Brewster	421	1	0.2850	0.1425
California	San Joaquin	343	1	0.0005	0.0003	Texas	Culberson	460	1	0.2850	0.1425
California	Santa Clara	280	1	0.0005	0.0003	Utah	Garfield	466	1	0.2850	0.1425
California	Shasta	438	1	0.2850	0.1425	Utah	San Juan	451	1	0.2850	0.1425
California	Stanislaus	531	4	0.0005	0.0001	Utah	Utah	416	1	0.2850	0.1425
California	Tulare	862	3	0.1902	0.0672	Vermont	Bennington	418	1	0.2850	0.1425
Colorado	Alamosa	479	1	0.2850	0.1425	Vermont	Chittenden	42	1	0.2850	0.1425
Colorado	La Plata	451	1	0.2850	0.1425	Virgin Islands	St. John	305	1	0.2850	0.1425
Colorado	Larimer	896	2	0.2850	0.1008	Virginia	Page	438	1	0.2850	0.1425
Colorado	Montezuma	442	1	0.2850	0.1425	Virginia	Rockbridge	465	1	0.2850	0.1425
Colorado	Pitkin	459	1	0.2850	0.1425	Washington	King	890	4	0.1427	0.0504
Colorado	Routt	446	1	0.2850	0.1425	Washington	Klickitat	445	1	0.2850	0.1425
District of Columbia	District of Columbia	467	1	0.2850	0.1425	Washington	Pierce	471	1	0.2850	0.1425
Florida	Citrus	475	1	0.2850	0.1425	Washington	Skamania	175	1	0.2850	0.1425
Florida	Dade	405	1	0.2850	0.1425	Washington	Whatcom	39	1	0.2850	0.1425
Georgia	Charlton	460	1	0.2850	0.1425	West Virginia	Tucker	467	1	0.2850	0.1425
Hawaii	Hawaii	859	2	0.2850	0.1008	Wyoming	Albany	464	1	0.2850	0.1425
Hawaii	Maui	439	1	0.2850	0.1425	Wyoming	Sublette	463	1	0.2850	0.1425
Idaho	Butte	919	2	0.2850	0.1008	Wyoming	Teton	417	1	0.2850	0.1425

Table A.23 Summary of Countywide Lead (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Alabama	Lawrence	369	1	0.0300	0.0150	Idaho	Custer	465	1	0.0300	0.0150
Alaska	Yukon-Koyukuk Census Area	467	1	0.0300	0.0150	Idaho	Lemhi	388	1	0.0300	0.0150
Arizona	Apache	428	1	0.0300	0.0150	Kentucky	Edmonson	469	1	0.0300	0.0150
Arizona	Cochise	462	1	0.0300	0.0150	Maine	Hancock	470	1	0.0300	0.0150
Arizona	Coconino	856	3	0.0300	0.0087	Maine	Washington	449	1	0.0300	0.0150
Arizona	Gila	461	1	0.0300	0.0150	Minnesota	Lake	345	1	0.0300	0.0150
Arkansas	Newton	472	1	0.0300	0.0150	Montana	Flathead	478	1	0.0300	0.0150
California	Del Norte	443	1	0.0300	0.0150	Montana	Ravalli	471	1	0.0300	0.0150
California	El Dorado	405	2	0.0300	0.0106	Nevada	Elko	409	1	0.0300	0.0150
California	Fresno	693	2	0.0040	0.0014	Nevada	White Pine	475	1	0.0300	0.0150
California	Imperial	378	3	0.0160	0.0009	New Hampshire	Coos	171	1	0.0300	0.0150
California	Inyo	896	3	0.0107	0.0050	New Jersey	Atlantic	446	1	0.0300	0.0150
California	Kern	1176	3	0.0023	0.0007	New Mexico	Catron	25	1	0.0300	0.0150
California	Kings	402	2	0.0026	0.0009	New Mexico	Dona Ana	174	1	0.0300	0.0150
California	Los Angeles	763	13	0.8854	0.1477	New Mexico	Los Alamos	457	1	0.0300	0.0150
California	Madera	104	1	0.0036	0.0003	North Carolina	Haywood	405	1	0.0300	0.0150
California	Marin	388	1	0.0300	0.0150	North Carolina	Swain	477	1	0.0300	0.0150
California	Mariposa	462	1	0.0300	0.0150	Oregon	Klamath	455	1	0.0300	0.0150
California	Orange	43	2	1.0000	0.3953	Oregon	Linn	467	1	0.0300	0.0150
California	Plumas	117	1	0.0010	0.0005	Oregon	Multnomah	91	4	0.0048	0.0012
California	Riverside	143	3	0.6692	0.2635	Oregon	Washington	28	1	0.0030	0.0015
California	Sacramento	385	1	0.0030	0.0015	South Carolina	Charleston	433	1	0.0300	0.0150
California	San Benito	438	1	0.0300	0.0150	South Dakota	Jackson	477	1	0.0300	0.0150
California	San Bernardino	645	4	0.5078	0.1977	Texas	Brewster	421	1	0.0300	0.0150
California	San Joaquin	343	1	0.0025	0.0002	Texas	Culberson	461	1	0.0300	0.0150
California	Santa Clara	280	1	0.0000	0.0010	Utah	Garfield	466	1	0.0300	0.0150
California	Shasta	438	1	0.0300	0.0150	Utah	San Juan	451	1	0.0300	0.0150
California	Stanislaus	531	4	0.0028	0.0005	Utah	Utah	416	1	0.0300	0.0150
California	Tulare	862	3	0.0210	0.0071	Vermont	Bennington	418	1	0.0300	0.0150
Colorado	Alamosa	479	1	0.0300	0.0150	Vermont	Chittenden	42	1	0.0300	0.0150
Colorado	La Plata	451	1	0.0300	0.0150	Virgin Islands	St. John	305	1	0.0300	0.0150
Colorado	Larimer	896	2	0.0300	0.0106	Virginia	Page	438	1	0.0300	0.0150
Colorado	Montezuma	442	1	0.0300	0.0150	Virginia	Rockbridge	466	1	0.0300	0.0150
Colorado	Pitkin	459	1	0.0300	0.0150	Washington	King	891	4	0.0191	0.0056
Colorado	Routt	446	1	0.0300	0.0150	Washington	Klickitat	445	1	0.0300	0.0150
District of Columbia	District of Columbia	467	1	0.0300	0.0150	Washington	Pierce	471	1	0.0300	0.0150
Florida	Citrus	475	1	0.0300	0.0150	Washington	Skamania	175	1	0.0300	0.0150
Florida	Dade	405	1	0.0300	0.0150	Washington	Whatcom	39	1	0.0300	0.0150
Georgia	Charlton	460	1	0.0300	0.0150	West Virginia	Tucker	467	1	0.0300	0.0150
Hawaii	Hawaii	860	2	0.0300	0.0106	Wyoming	Albany	464	1	0.0300	0.0150
Hawaii	Maui	439	1	0.0300	0.0150	Wyoming	Sublette	463	1	0.0300	0.0150
Idaho	Butte	919	2	0.0300	0.0106	Wyoming	Teton	417	1	0.0300	0.0150

Table A.24 Summary of Countywide Manganese (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Alabama	Lawrence	369	1	0.2600	0.1300	Idaho	Custer	465	1	0.2600	0.1300
Alaska	Yukon-Koyukuk Census Area	466	1	0.2600	0.1300	Idaho	Lemhi	387	1	0.2600	0.1300
Arizona	Apache	428	1	0.2600	0.1300	Kentucky	Edmonson	469	1	0.2600	0.1300
Arizona	Cochise	462	1	0.2600	0.1300	Maine	Hancock	470	1	0.2600	0.1300
Arizona	Coconino	856	3	0.2600	0.0751	Maine	Washington	449	1	0.2600	0.1300
Arizona	Gila	461	1	0.2600	0.1300	Minnesota	Lake	345	1	0.2600	0.1300
Arkansas	Newton	472	1	0.2600	0.1300	Montana	Flathead	478	1	0.2600	0.1300
California	Del Norte	443	1	0.2600	0.1300	Montana	Ravalli	471	1	0.2600	0.1300
California	El Dorado	404	2	0.2600	0.0919	Nevada	Elko	409	1	0.2600	0.1300
California	Fresno	693	2	0.0000	0.0004	Nevada	White Pine	475	1	0.2600	0.1300
California	Imperial	378	3	0.0071	0.0004	New Hampshire	Coos	171	1	0.2600	0.1300
California	Inyo	896	3	0.0868	0.0433	New Jersey	Atlantic	445	1	0.2600	0.1300
California	Kern	1176	3	0.0013	0.0005	New Mexico	Catron	25	1	0.2600	0.1300
California	Kings	402	2	0.0000	0.0004	New Mexico	Dona Ana	174	1	0.2600	0.1300
California	Los Angeles	763	13	0.8487	0.1276	New Mexico	Los Alamos	457	1	0.2600	0.1300
California	Madera	104	1	0.0000	0.0005	North Carolina	Haywood	405	1	0.2600	0.1300
California	Marin	388	1	0.2600	0.1300	North Carolina	Swain	477	1	0.2600	0.1300
California	Mariposa	462	1	0.2600	0.1300	Oregon	Klamath	455	1	0.2600	0.1300
California	Orange	43	2	1.0000	0.3536	Oregon	Linn	467	1	0.2600	0.1300
California	Plumas	117	1	0.0005	0.0003	Oregon	Multnomah	90	4	0.0023	0.0008
California	Riverside	143	3	0.6681	0.2357	Oregon	Washington	28	1	0.0016	0.0008
California	Sacramento	385	1	0.0000	0.0005	South Carolina	Charleston	433	1	0.2600	0.1300
California	San Benito	438	1	0.2600	0.1300	South Dakota	Jackson	477	1	0.2600	0.1300
California	San Bernardino	645	4	0.5655	0.1797	Texas	Brewster	421	1	0.2600	0.1300
California	San Joaquin	343	1	0.0044	0.0001	Texas	Culberson	460	1	0.2600	0.1300
California	Santa Clara	280	1	0.0000	0.0005	Utah	Garfield	466	1	0.2600	0.1300
California	Shasta	438	1	0.2600	0.1300	Utah	San Juan	451	1	0.2600	0.1300
California	Stanislaus	531	4	0.0014	0.0004	Utah	Utah	416	1	0.2600	0.1300
California	Tulare	862	3	0.1733	0.0613	Vermont	Bennington	418	1	0.2600	0.1300
Colorado	Alamosa	479	1	0.2600	0.1300	Vermont	Chittenden	42	1	0.2600	0.1300
Colorado	La Plata	451	1	0.2600	0.1300	Virgin Islands	St. John	305	1	0.2600	0.1300
Colorado	Larimer	896	2	0.2600	0.0919	Virginia	Page	438	1	0.2600	0.1300
Colorado	Montezuma	442	1	0.2600	0.1300	Virginia	Rockbridge	465	1	0.2600	0.1300
Colorado	Pitkin	459	1	0.2600	0.1300	Washington	King	890	4	0.1331	0.0460
Colorado	Routt	446	1	0.2600	0.1300	Washington	Klickitat	445	1	0.2600	0.1300
District of Columbia	District of Columbia	467	1	0.2600	0.1300	Washington	Pierce	471	1	0.2600	0.1300
Florida	Citrus	475	1	0.2600	0.1300	Washington	Skamania	175	1	0.2600	0.1300
Florida	Dade	405	1	0.2600	0.1300	Washington	Whatcom	39	1	0.2600	0.1300
Georgia	Charlton	460	1	0.2600	0.1300	West Virginia	Tucker	467	1	0.2600	0.1300
Hawaii	Hawaii	859	2	0.2600	0.0919	Wyoming	Albany	464	1	0.2600	0.1300
Hawaii	Maui	439	1	0.2600	0.1300	Wyoming	Sublette	463	1	0.2600	0.1300
Idaho	Butte	919	2	0.2600	0.0919	Wyoming	Teton	417	1	0.2600	0.1300

Table A.25 Summary of Countywide Mercury (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0010	0.0004	California	Riverside	104	1	0.0010	0.0005
California	Imperial	378	3	0.0010	0.0003	California	Sacramento	385	1	0.0010	0.0005
California	Inyo	432	2	0.0010	0.0004	California	San Bernardino	221	1	0.0010	0.0005
California	Kern	1176	3	0.0010	0.0003	California	San Joaquin	343	1	0.0010	0.0005
California	Kings	402	2	0.0010	0.0004	California	Santa Clara	280	1	0.0010	0.0005
California	Los Angeles	511	2	0.0010	0.0004	California	Stanislaus	531	4	0.0010	0.0003
California	Madera	104	1	0.0010	0.0005	California	Tulare	307	1	0.0010	0.0005
California	Plumas	117	1	0.0010	0.0005						

Table A.26 Summary of Countywide Nickel (Fine) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Alabama	Lawrence	369	1	0.0250	0.0125	Idaho	Custer	465	1	0.0250	0.0125
Alaska	Yukon-Koyukuk Census Area	467	1	0.0250	0.0125	Idaho	Lemhi	388	1	0.0250	0.0125
Arizona	Apache	428	1	0.0250	0.0125	Kentucky	Edmonson	469	1	0.0250	0.0125
Arizona	Cochise	462	1	0.0250	0.0125	Maine	Hancock	470	1	0.0250	0.0125
Arizona	Coconino	856	3	0.0250	0.0072	Maine	Washington	449	1	0.0250	0.0125
Arizona	Gila	461	1	0.0250	0.0125	Minnesota	Lake	345	1	0.0250	0.0125
Arkansas	Newton	472	1	0.0250	0.0125	Montana	Flathead	478	1	0.0250	0.0125
California	Del Norte	443	1	0.0250	0.0125	Montana	Ravalli	471	1	0.0250	0.0125
California	El Dorado	405	2	0.0250	0.0088	Nevada	Elko	409	1	0.0250	0.0125
California	Fresno	693	2	0.0005	0.0002	Nevada	White Pine	475	1	0.0250	0.0125
California	Imperial	378	3	0.0003	0.0002	New Hampshire	Coos	171	1	0.0250	0.0125
California	Inyo	896	3	0.0087	0.0042	New Jersey	Atlantic	446	1	0.0250	0.0125
California	Kern	1176	3	0.0005	0.0001	New Mexico	Catron	25	1	0.0250	0.0125
California	Kings	402	2	0.0005	0.0002	New Mexico	Dona Ana	174	1	0.0250	0.0125
California	Los Angeles	763	13	0.6541	0.1036	New Mexico	Los Alamos	457	1	0.0250	0.0125
California	Madera	104	1	0.0005	0.0003	North Carolina	Haywood	405	1	0.0250	0.0125
California	Marin	388	1	0.0250	0.0125	North Carolina	Swain	477	1	0.0250	0.0125
California	Mariposa	462	1	0.0250	0.0125	Oregon	Klamath	455	1	0.0250	0.0125
California	Orange	43	2	0.7500	0.2795	Oregon	Linn	467	1	0.0250	0.0125
California	Plumas	117	1	0.0005	0.0003	Oregon	Multnomah	91	4	0.0021	0.0005
California	Riverside	143	3	0.5002	0.1863	Oregon	Washington	28	1	0.0016	0.0008
California	Sacramento	385	1	0.0005	0.0003	South Carolina	Charleston	433	1	0.0250	0.0125
California	San Benito	438	1	0.0250	0.0125	South Dakota	Jackson	477	1	0.0250	0.0125
California	San Bernardino	645	4	0.3814	0.1398	Texas	Brewster	421	1	0.0250	0.0125
California	San Joaquin	343	1	0.0005	0.0003	Texas	Culberson	461	1	0.0250	0.0125
California	Santa Clara	280	1	0.0005	0.0003	Utah	Garfield	466	1	0.0250	0.0125
California	Shasta	438	1	0.0250	0.0125	Utah	San Juan	451	1	0.0250	0.0125
California	Stanislaus	531	4	0.0005	0.0001	Utah	Utah	416	1	0.0250	0.0125
California	Tulare	862	3	0.0168	0.0059	Vermont	Bennington	418	1	0.0250	0.0125
Colorado	Alamosa	479	1	0.0250	0.0125	Vermont	Chittenden	42	1	0.0250	0.0125
Colorado	La Plata	451	1	0.0250	0.0125	Virgin Islands	St. John	305	1	0.0250	0.0125
Colorado	Larimer	896	2	0.0250	0.0088	Virginia	Page	438	1	0.0250	0.0125
Colorado	Montezuma	442	1	0.0250	0.0125	Virginia	Rockbridge	466	1	0.0250	0.0125
Colorado	Pitkin	459	1	0.0250	0.0125	Washington	King	889	4	0.0137	0.0044
Colorado	Routt	446	1	0.0250	0.0125	Washington	Klickitat	445	1	0.0250	0.0125
District of Columbia	District of Columbia	467	1	0.0250	0.0125	Washington	Pierce	471	1	0.0250	0.0125
Florida	Citrus	475	1	0.0250	0.0125	Washington	Skamania	175	1	0.0250	0.0125
Florida	Dade	405	1	0.0250	0.0125	Washington	Whatcom	39	1	0.0250	0.0125
Georgia	Charlton	460	1	0.0250	0.0125	West Virginia	Tucker	467	1	0.0250	0.0125
Hawaii	Hawaii	860	2	0.0250	0.0088	Wyoming	Albany	464	1	0.0250	0.0125
Hawaii	Maui	439	1	0.0250	0.0125	Wyoming	Sublette	463	1	0.0250	0.0125
Idaho	Butte	919	2	0.0250	0.0088	Wyoming	Teton	417	1	0.0250	0.0125

Table A.27 Summary of Countywide Arsenic (Coarse) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0010	0.0004	California	Riverside	104	1	0.0010	0.0005
California	Imperial	378	3	0.0010	0.0003	California	Sacramento	385	1	0.0010	0.0005
California	Inyo	432	2	0.0010	0.0004	California	San Bernardino	221	1	0.0010	0.0005
California	Kern	1176	3	0.0010	0.0003	California	San Joaquin	343	1	0.0010	0.0005
California	Kings	402	2	0.0010	0.0004	California	Santa Clara	280	1	0.0010	0.0005
California	Los Angeles	511	2	0.0010	0.0004	California	Stanislaus	531	4	0.0010	0.0003
California	Madera	104	1	0.0010	0.0005	California	Tulare	307	1	0.0010	0.0005
California	Plumas	117	1	0.0010	0.0005						

Table A.28 Summary of Countywide Chromium (Coarse) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0000	0.0004	California	Riverside	104	1	0.0000	0.0005
California	Imperial	378	3	0.0018	0.0006	California	Sacramento	385	1	0.0000	0.0005
California	Inyo	432	2	0.0005	0.0002	California	San Bernardino	221	1	0.0000	0.0005
California	Kern	1176	3	0.0013	0.0005	California	San Joaquin	343	1	0.0000	0.0005
California	Kings	402	2	0.0000	0.0004	California	Santa Clara	280	1	0.0000	0.0005
California	Los Angeles	511	2	0.0052	0.0003	California	Stanislaus	531	4	0.0001	0.0002
California	Madera	104	1	0.0005	0.0003	California	Tulare	307	1	0.0000	0.0005
California	Plumas	117	1	0.0005	0.0003						

Table A.29 Summary of Countywide Lead (Coarse) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0000	0.0007	California	Riverside	104	1	0.0030	0.0015
California	Imperial	378	3	0.0044	0.0003	California	Sacramento	385	1	0.0000	0.0010
California	Inyo	432	2	0.0010	0.0004	California	San Bernardino	221	1	0.0010	0.0005
California	Kern	1176	3	0.0003	0.0005	California	San Joaquin	343	1	0.0000	0.0010
California	Kings	402	2	0.0010	0.0004	California	Santa Clara	280	1	0.0000	0.0010
California	Los Angeles	511	2	0.0045	0.0016	California	Stanislaus	531	4	0.0008	0.0003
California	Madera	104	1	0.0010	0.0005	California	Tulare	307	1	0.0000	0.0010
California	Plumas	117	1	0.0010	0.0005						

Table A.30 Summary of Countywide Manganese (Coarse) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0210	0.0074	California	Riverside	104	1	0.0310	0.0155
California	Imperial	378	3	0.0100	0.0030	California	Sacramento	385	1	0.0137	0.0003
California	Inyo	432	2	0.0063	0.0008	California	San Bernardino	221	1	0.0080	0.0040
California	Kern	1176	3	0.0176	0.0012	California	San Joaquin	343	1	0.0141	0.0071
California	Kings	402	2	0.0290	0.0075	California	Santa Clara	280	1	0.0030	0.0015
California	Los Angeles	511	2	0.0060	0.0022	California	Stanislaus	531	4	0.0144	0.0015
California	Madera	104	1	0.0155	0.0007	California	Tulare	307	1	0.0341	0.0006
California	Plumas	117	1	0.0050	0.0025						

Table A.31 Summary of Countywide Nickel (Coarse) Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
California	Fresno	693	2	0.0003	0.0003	California	Riverside	104	1	0.0005	0.0003
California	Imperial	378	3	0.0003	0.0002	California	Sacramento	385	1	0.0000	0.0005
California	Inyo	432	2	0.0005	0.0002	California	San Bernardino	221	1	0.0005	0.0003
California	Kern	1176	3	0.0003	0.0002	California	San Joaquin	343	1	0.0000	0.0005
California	Kings	402	2	0.0003	0.0003	California	Santa Clara	280	1	0.0005	0.0003
California	Los Angeles	511	2	0.0028	0.0003	California	Stanislaus	531	4	0.0005	0.0001
California	Madera	104	1	0.0005	0.0003	California	Tulare	307	1	0.0005	0.0003
California	Plumas	117	1	0.0005	0.0003						

Table A.32 Summary of Countywide Hexachlorobenzene Background Estimates

State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate	State	County	Sample Size	No. Sites	Background Estimate	Uncertainty of Background Estimate
Michigan	Oakland	39	1	0.0357	0.0178	Michigan	Wayne	216	6	0.0408	0.0084

APPENDIX B:

CONVERSION FACTORS

Appendix B: Conversion Factors

Table B.1 consists of three original units and the general equation used to convert original concentration values into $\mu g/m^3$. In Table B.1 MWGT is the molecular weight of the pollutant, and NUMC is the number of carbon atoms for the pollutant. Table B.2 consists of 19 pollutants and the equations used to convert original concentration values into $\mu g/m^3$. In both tables, X is the original concentration value given in the unit as indicated by the Original Units column of the table. All conversions assume standard temperature and pressure (STP), i.e., a temperature of 0 C (32 F) and 1 atmosphere of pressure (14.7 psi). For example, suppose a measurement of Acetaldehyde is recorded as 1.64 PPBV, then the converted concentration value would be 1.64 * (44.05/24.46) = 2.95 $\mu g/m^3$.

Table B.1. General Form of Conversions

Original Units	Conversion Equation
PPBV	X * (MWGT/24.46)
PPBC	(X/NUMC) * (MWGT/24.46)
NG/M^3	X/1000

Table B.2. Conversions by Pollutant

Class	Pollutant	Original Units	Conversion to $\mu g/m^3$
Carbonyl	ACETALDEHYDE	PPBV	X*(44.05/24.46)
	ACROLEIN	PPBV	X*(56.06/24.46)
	FORMALDEHYDE	PPBV	X*(30.03/24.46)
Metal	ALL (when applicable)	NG/M^3	X/1000
VOC	1,1,2,2-TETRACHLOROETHANE	PPBV	X*(167.85/24.46)
	1,2-DIBROMOETHANE	PPBV	X*(187.86/24.46)
	1,2-DICHLOROETHANE	PPBV	X*(98.96/24.46)
	1,2-DICHLOROPROPANE	PPBV	X*(112.99/24.46)
	1,3-BUTADIENE	PPBV	X*(54.09/24.46)
		PPBC	(X/4)*(54.09/24.46)
	ACRYLONITRILE	PPBV	X*(53.06/24.46)
	BENZENE	PPBV	X*(78.11/24.46)
		PPBC	(X/6)*(78.11/24.46)
	CARBON TETRACHLORIDE	PPBV	X*(153.82/24.46)
	CHLOROFORM	PPBV	X*(119.38/24.46)
	CIS 1,3-DICHLOROPROPENE	PPBV	X*(110.97/24.46)
	ETHYLENE OXIDE	PPBV	X*(44.053/24.46)
	METHYLENE CHLORIDE	PPBV	X*(84.93/24.46)
	TETRACHLOROETHYLENE	PPBV	X*(165.83/24.46)
	TRICHLOROETHYLENE	PPBV	X*(131.39/24.46)
	VINYL CHLORIDE	PPBV	X*(62.5/24.46)

APPENDIX C:

RESPONSES TO
BATTELLE SENIOR STATISTICAL REVIEW

Appendix C: Responses to Battelle Senior Statistical Review

EPA/OAQPS requested that Battelle conduct an internal statistical review of the approach and methods used to obtain background concentration estimates for this project. Two senior Battelle statisticians were identified and agreed to participate in this effort: Dr. Bruce Buxton and Dr. Paul Feder. This appendix provides their comments and the associated responses of the principal researchers for this project, Dr. Steve Bortnick and Dr. Basil Coutant. (Note that page number and paragraph references correspond to the August 15, 2002, draft report.)

C.1 Comments From and Responses to Dr. Bruce Buxton

Dr. Bruce Buxton is a Senior Program Manager in Battelle's Statistics and Data Analysis Systems (SDAS) department. He obtained his Ph.D. in Geostatistics from Stanford University in 1986. Dr. Buxton has over 20 years of experience working on many different applications in the fields of statistics, primarily focused on environmental problems. Dr. Buxton's comments, and responses to these comments, are as follows:

1. Overall, the proposed model and report seem fine, although I do have several questions and comments.

Response: No response.

2. Page 7, Para. 3 – Your first sentence jumps into the assumption of a parametric statistical distribution as if it is mandatory. You might want to consider non-parametric approaches, or at least acknowledge that they exist.

Response: Sentence added.

3. Page 8, Para. 3 – I do not think that asymmetry is absolutely required for all background distributions. You might want to point out that the gamma distribution you are using has the flexibility to model both heavily asymmetric or reasonably symmetric distributions.

Response: Sentence added.

4. Page 10, Figure 3.2 and related text – How about showing a histogram of your data to go along with the Q-Q plot? Is a visual inspection of this figure all you recommend to determine that your model selection is adequate? Should you consider some goodness-of-fit or similar statistical test?

Response: While this could be explored further, other tests of fit adequacy were not explored for this draft of the report, mostly due to resource and time constraints. A sentence mentioning these possibilities has been added.

5. Page 11, Para. 4 – It strikes me that your definition of a "threshold" at mu+2*MDL appears similar to the definition of a Limit of Quantification (although I am not an expert in LOQs). If there is a natural analogy here, that might be a better way to present it rather than an arbitrary threshold.

Response: Good point. Assuming our semantics are in line, LOQs are also referred to as the lowest calibration level (LCL), typically defined as 3*MDL. Many laboratories choose not to report below-LCL data due to apparent high uncertainties. Discussion has been added to the report.

6. Page 12, Para. 2 – You talk about censoring data below the threshold, but the threshold involves the parameter mu which is unknown. It is not clear to me what is happening here. Maybe this is more a part of the model, rather than a step in the data analysis. I suggest you clarify the discussion here.

Response: It is a data analysis decision to help the model (software) fit the data in real-world application; i.e., the model does not have to guess where the background lies relative to the censoring threshold, because of the data analysis decision that is made up front (the answer is always the same – the background lies to the left of the censoring threshold). No changes have been made to the text to date. It is explained further in the subsequent discussion and example Figures 3.3a through 3.3c.

7. Page 13, Para. 2 – Three times you state that the figures are "consistent with the model", as if that is noteworthy. Aren't the figures just plots of the models, so that they have to be consistent by definition? Maybe I am missing something here.

Response: Correct. "consistent with" has been changed to "corresponds to".

8. Page 15, Para. 1 – I do not agree that your analysis needs to be automated and routine just because your data set is large. I think the more important issue is how frequently the analysis needs to be performed or updated, for example, hourly, daily, or weekly.

Response: The real point in this case is that the analysis has to be routine relative to the budget of the work assignment. The work assignment budget could not afford for us to manually fit "the best" statistical distribution to every site for every pollutant. The limitations of the budget required a more generic approach that could be run over and over on a large number of sites and pollutants. This was a project management decision. The routine requirement will come into play only to the degree that EPA/OAQPS needs to repeat this effort for the 2002 cycle of NATA. No modifications were made to the report in response to this comment.

9. Page 15, Para. 2 – **Major Issue** – Is it the data below the MDL that cause the convergence issues? You discuss convergence problems in several places in the report, so often, in fact, that it appears convergence is the single greatest factor determining your technical modeling approach. Maybe you should consider ways of replacing the non-detect data with continuous data (e.g., simulation) so that algorithms without convergence issues can be used.

Response: This was considered in the early phases of the project. It is certainly a valid approach. By the time the models and applied approach were substantially developed and the project was well underway, we felt it would have been too late to change the approach due to convergence issues. Convergence and other issues are mentioned quite often throughout the report, but only due to our desire to explicitly describe as many details of the process as possible. The desire on our part, based on feedback from the Work Assignment Manager, was to clearly elucidate the theory and applied process so that others would have ample opportunity to review and possibly critique the methods. A discussion has been added to subsection 3.1.4 mentioning your suggestion.

10. Page 16, Para. 2 – The 1 percent data trimming seems arbitrary at best. Honestly, it appears that you have a lack of fit of your model to your data. And, in response, instead of searching for a better model, you decided to alter your data set to eliminate the problem data and, hence, obtain a better model fit. Also, if you attribute the problematic distribution to a mixture of data due to changing meteorological or other conditions, why would this only affect the upper tail of the distribution?

Response: To answer the last question first, the hypothesis for this assumption is explained in the California EPA report cited in the text of this section. Otherwise, as we state throughout the report, this was a pragmatic decision that was made in full awareness of the objection you raise from the perspective of scientific rigor. Bottom line: it was a quick and easy solution that seemed to fix a potential problem, versus a possibly more complex (but possibly more appealing) solution of considering different distributions, developing site-specific distributions, etc. Furthermore, as we state in the report, we are interested in estimating the lower tail of the distribution, so the actual effect of this decision should be minimal at most (note the comment to this effect from Dr. Feder below). No changes were made to the report in response to this comment.

11. Page 16, Para. 3, Last sentence – I would expect trimming the highest data from your data set to have a potentially dramatic impact on the parameter estimates.

Response: We disagree, and the initial sensitivity analyses we conducted to check this assumption all seemed to confirm our conclusion. Again, see comment from Dr. Feder below to this effect.

12. Page 17, Para. 1 – Did you check on the appropriateness of applying a standard normal confidence interval to the estimation of mu? Do you have a reference for this?

Response: No. A parenthetical sentence has been added as a caveat.

13. Page 17, Table 3.1 – I suggest that you add a column for N, the number of data in each data set.

Response: Done.

14. Page 18, Figure 3.4 – Your box plots look reasonably symmetric to me. I am not sure if you should point this out to the reader, but the report does state that one reason for choosing the gamma distribution was the expected asymmetric data distributions.

Response: Our response to Comment 3 above addresses this issue.

15. Page 19, Para. 2 – I suggest you delete this paragraph; it is distracting at best. I do not believe the apparent downward trend in benzene levels over time says anything meaningful about source apportionment, that is, what the sources are or how much each source contributes to pollution levels.

Response: This paragraph was put in based on direct conference call discussions with EPA/OAQPS and the Work Assignment Manager. It is left in for now, but could be deleted in the future upon EPA review and comment.

16. Page 21, Para. 2 – You should probably support your opinion about an R-squared of 0.3 with some references.

Response: No reference has been added to date. This could be done in the future. For now, the sentence has been modified to recognize the lack of reference support.

17. Page 22, Para. 4 – When you state that some of your results "may not be credible", it makes this reader nervous about all the other results. Also, later in this paragraph you present "crude" results. I think crude results should probably not be included.

Response: The sentence with "may not be credible" was removed, and the following discussion with "crude" was removed.

18. Page 22, Para. 5 – I suggest you delete "for better or worse".

Response: Done.

19. Page 23, Para. 2 – You should provide references for the "number of studies" you cite in the second sentence.

Response: Modified text accordingly.

20. Page 25, Para. 1 – I believe you mean Geographical (not Graphical) Information Systems.

Response: You are correct. Correction made.

21. GIS maps – There is so much white space that these figures seem pretty uninformative, at least from a national perspective.

Response: All GIS maps have been dropped from the report except for one example in Chapter 4 for benzene, which provides the most comprehensive set of results (i.e., the least white space).

C.2 Comments From and Responses to Dr. Paul Feder

Dr. Paul Feder is a Research Leader in Battelle's Statistics and Data Analysis Systems (SDAS) department. He obtained his Ph.D. in Statistics from Stanford University in 1968. He has over 30 years of experience on a wide variety of research projects for government agencies and private industry. Dr. Feder has led a large variety of research programs involving experimental design and data analysis in the areas of in vitro, mammalian, and aquatic toxicology; preclinical investigations of developmental drug efficacy and safety; multi-center pharmaceutical clinical trials; pharmacokinetic modeling; ecological monitoring and experimentation; nuclear safety; product performance field testing; among many other applications. Dr. Feder's comments, and responses to these comments, are as follows:

1. Treatment of MDL values. Should do censored data analysis.

Response: We essentially are using censored analysis techniques except that we had to modify the traditional approach to take into consideration that we do not have sharp censoring (i.e., censoring at a clearly distinguishable threshold).

2. Sensitivity of distributional assumptions. E.g., log normal or Weibull versus gamma.

- Average value of background
- Standard error of background
- Should do sensitivity analysis

Response: A portion of this was done. The gamma distribution was compared with the log normal distribution. The differences in the background estimates were insignificant, where both models converged. However, the numerical routines failed to converge more frequently when the log normal distribution was used.

3. Objectives of background estimates - use of standard error of background estimates

- I have questions about validity
- Background estimate $\hat{\mu}$ does not satisfy standard conditions for maximum likelihood estimates.
- Usual theory may not hold. Asymptotic normality, standard error.
- Jackknife or bootstrap standard errors better.

Response: It is true that the standard conditions do not hold (for Model 1) and, hence, the usual theory may not hold. However, the standard error estimate still holds as a lower bound (via the Cramer-Rao theorem). Regardless, theory only states that the estimates are approximately correct, the quality of those approximations is governed more by limitations of the numerical methods than the theory in this case. The suggestion to use jackknife or bootstrap methods is not practical given the frequency with which the mathematical algorithms fail to converge.

4. Autocorrelation not accounted for.

Response: In many cases, the autocorrelation from day-to-day is known to be small (and was being investigated as part of WA 5-12 under this contract). Much of the data in the archive are based on one-in-six-day sampling and, hence, autocorrelation is negligible. In any case, autocorrelation generally will not affect the background estimate, but can change the standard error estimates (as stated above, not expected to be a significant issue in this case due to the typical sampling frequencies encountered).

5. Page 5: I question how you report "above MDL" distribution

Response: The report is clear that a certain percentage of each pollutant's data fall below the MDL, and a certain percentage above. For each of these distinct groups of data, a summary of its distribution is provided.

6. Page 8: Shifted gamma versus alternative distribution. Basis of choice? Sensitivity analysis.

Response: See response to Comment 2 above.

7. Page 9: Did you check for autocorrelation at your sites? Impact on analysis results

- Main impact may be on standard errors.
- All analyses assume i.i.d. observations. Implication of assumptions.

Response: See response to Comment 4 above.

8. Page 11, 12: I disagree with the way you are treating MDL values. From page 5, Table 2.1, for some chemicals the proportion of MDLs is appreciable (10-55%).

- MDLs have nothing to do with background. They reflect analytical limitations and may be less than or greater than background.
- Treat as left censored data at the MDL, irrespective of background level.
- Treat as one process with measurable values.

Response: It is true that the MDLs have nothing to do with background. However, the mathematical form of the likelihood does depend on which is larger. The treatment used yields a consistent mathematical form, and a consistent censoring regardless of which is larger. Hence, no assumption needs to be made before modeling, and this allows for a mathematical treatment as one process with measured values.

9. Page 15: Weighted or unweighted average across sites within County?

- Weight by population or inversely to (standard errors)^2 ?

Response: An unweighted average is used. This example is generated from a set of consistently collected data of approximately the same amount from nearby sites. Note that the background estimates are similar, but the standard errors vary by more than a factor of three. Statistical theory states that weighting with the (inverse of the square of the) standard errors yields estimates with smaller standard errors. However, in this case, it was felt that the numerical precision of the standard error estimates made their use in the manner suggested inappropriate.

10. Page 16: Eliminate upper portion of model. I agree; you can even eliminate more. I would treat it as right censored data. Your model is then more robust in lower portion.

Response: This opinion is not shared with Dr. Buxton. Initial investigations showed that essentially all of the data could be used, and so this issue does not turn out to be significant.

11. Page 17, Table 3.1: NW Post office

- Background estimate is 0.7359.
- Figure 3.4: minimum is much lower than 0.5. How can that be?! It violates your model.

Response: It does seem to violate the initial model, but in fact it works by design. The day-to-day background is almost certainly not constant. The daily values are not estimable from the data. However, we are estimating the annual average of that background. True daily values could be less than this value. Further, the model allows for (in fact, assumes) non-trivial measurement error near or below the annual average background. Hence, measured values could be substantially less than the estimate.

C.3 Other Remarks

EPA/OAQPS requested that Dr. Buxton and Dr. Feder also comment on the secondary data analysis methodology described in subsection 3.1.7 of the report. This methodology was not developed in time for the August 15 draft report, and has only since been added to address all the cases for which the primary methodology, as discussed in subsections 3.1.1 through 3.1.6, does not give a solution. Due to the timing of incorporating the secondary methodology into the current draft of the report, a detailed review could not be provided by the reviewers. However, their general sentiments on the issue of addressing cases where most/all data are below the MDL or too few total observations are available for rigorous analysis may be summarized as follows:

- Dr. Buxton felt that any attempt to analyze data where most/all measurements were below the MDL is effectively "non-statistical". Essentially, there's no variation in the data in order to conduct a statistical analysis. [The same is essentially true if too few total observations are available in a data set, regardless of MDL issues.] Therefore, no one approach (e.g., set background equal to MDL/2) can easily be singled out and defended as clearly better or worse than any other in such cases. In essence, there is little to no information to confirm or refute any associated estimate, so long as the estimate passes some sort of basic reasonableness criterion. For example, if 100 percent of the data are below some constant MDL, then any estimate of background that fails to lie within the interval (0,MDL) would appear to be unreasonable.

- Dr. Feder felt that the primary methodology used for the project, as discussed in subsections 3.1.1 through 3.1.6, might be extended to address these cases. For example, assume some sort of shifted gamma distribution still holds for describing below-MDL data. However, the parameters of this distribution are such that the data's observed distribution essentially falls within the range of (0,MDL). Due to the MDL limitation, that distribution cannot be distinguished in the recorded data. The main reason for attempting this sort of approach would be to maintain the motivation and rationale associated with the primary approach of the report. The main problem with this approach is that it would require further assumptions in order to infer the "unobserved" gamma distribution assumed to be producing the below-MDL data (see Dr. Buxton's comments above). Furthermore, it would not directly address the other scenarios to be handled via the secondary methodology of the report, namely cases with above-MDL data but too few total observations.

With the above comments in mind, an approach was pursued that handles both cases of (1) too much below-MDL data and (2) too few total observations in a given data set. It also handles cases that fall somewhere in between these two extremes. In general, the secondary data analysis methodology described in subsection 3.1.7 was pursued for the following reasons:

1. It can address every single case not addressed by the primary methodology discussed in subsections 3.1.1 through 3.1.6, so that any available monitoring data

will somehow be used to estimate background. This dramatically extends the breadth of results from those provided in the August 15 report.

2. It provides a reasonable result (see Dr. Buxton's comment above) when all data are below the MDL.

3. It provides a reasonable result when all data are above the MDL but too few total observations are available for pursuing the primary methodology discussed in subsections 3.1.1 through 3.1.6.

4. It seamlessly addresses all cases in between Items 2 and 3 above.

5. It provides conservative estimates of uncertainty for the resulting background estimates.

Refer to subsection 3.1.7 for further details on the secondary statistical analysis methodology pursued for this project. Refer to Table 2.1b for a summary of the amount of data to which this secondary methodology was applied (compare to Table 2.1a for the primary methodology). Refer to Table 4.1 and Appendix A for associated results.

APPENDIX D:

RESPONSES TO EPA REVIEW OF THE AUGUST 15, 2002, DRAFT

Appendix D: Responses to EPA Review of the August 15, 2002, Draft

E-mail received from Joe Touma August 16, 2002, 11:11 am

The shading scheme you are using in Figures A. to A.9 does not show up well in print. Can you please re-do with different scheme (cross hatched, dots, etc.).

Response: For the most part, these figures have been dropped from the report in favor of more detailed tabular summaries (see Appendix A). This change was made in response to the next e-mail comment below and other discussions with EPA/OAQPS. However, Figure 4.1 for benzene is still provided in the report as an example. This figure is now made in color to address the above comment. The use of "cross hatched, dots, etc." was explored, but due to the small sizes of many counties in the figure, the decision was made that color was a better choice for addressing the primary issue in the above comment.

E-mail received from Joe Touma August 16, 2002, 3:17 pm

Another suggestion on data presentation ... use tabular presentation instead of the more expensive geographical presentation.

Response: Done. See Appendix A.

E-mail received from Joe Touma August 21, 2002, 3:31 pm

1. Explore adding years 1995 and 1996 to the database.

Response: Done.

2. Find why some data are not being retrieved. For example, there are 1997 monitoring data for Atlanta on AQS (formally AIRS).

Response: Battelle conducted an experiment to explore the air toxics data archive for two commonly monitored HAPs, benzene and formaldehyde, with respect to both Atlanta (Fulton county) and St. Louis. Limiting the search to as far back as 1995, only one site (23 observations) in the archive provided benzene data for Fulton county, no data were present for benzene in St. Louis, and no data were present for formaldehyde in either Fulton county or St. Louis. Going back through the entire archive, a total of 17 sites (347 observations) provided Fulton County benzene data. Fulton county formaldehyde data also exist within the archive, but all were collected prior to 1993. Likewise, St. Louis benzene and formaldehyde data also exist within the archive (2 sites), but all measurements were made prior to 1991. For this example, this summarizes the data in the latest version of the archive we worked with for this project.

Note that the post-1995 Fulton county benzene data mentioned above are included in the results of the report (see, specifically, Table A.10 of Appendix A).

3. Provide tabular values for the counties instead of the graphs that are hard to see. (Please provide us this table for Harris county, Texas, and any surrounding counties as soon as practical). Regarding graphs used in your report, please explain rational for the legend scale.

Response: [Provide tabular values for the counties instead of the graphs that are hard to see.] Appendix A has been modified accordingly. [Please provide us this table for Harris county, Texas, and any surrounding counties as soon as practical.] See Appendix A. [Regarding graphs used in your report, please explain rational for the legend scale.] A rationale discussion has been added to subsection of 4.1.1 of the report.

4. Provide conversion factors used in going from AQS to Table 4.1. [If possible also provide a Table (e.g., 4.1a) that gives units in ppbc for use by monitoring staff.]

Response: Appendix B has been added to the report to describe conversion factors. Due to time and resource constraints, a Table 4.1a has not been added. Any individual interested in making such a conversion could do so via comparing Table 4.1 with the details provided in Appendix B.

5. Explore how easy it is to get databases that include other pollutants, e.g., mercury, dioxin, etc.

Response: The issue of "how easy it is" is not discussed explicitly in the report. However, a more detailed discussion of many such databases has been added to subsection 2.1.1 of the report. For most of the databases described there, Battelle has reasonably straightforward, if not direct, access to the data. The issue of "how easy it is" is more a matter of potential data management and QA/QC complications versus any potential difficulties associated with obtaining the data.

6. Explore other ideas for estimating background mentioned in your report.

Response: Done. See discussion at the beginning of Section 3.1 and detailed discussion in subsection 3.1.7. Also, refer to results presented in Table 4.1 (compare to Table 4.1 results provided in August 15 report). Also, compare Tables 2.1a and 2.1b.

E-mail received from Joe Touma August 22, 2002, 9:48 am

I asked one of my colleagues whether there are any air toxics data for Atlanta on AQS. Here is what he came up with. Please include in your database if not already included.

Response: These data were explored. Ultimately they were not included, mainly because no MDL information was provided along with the data. Also, the data were provided as hourly measurements (i.e., 1-hour integrated samples), as opposed to the daily measurements

(i.e., 24-hour integrated samples or appropriately averaged 1-hour or 3-hour samples) used for the data analyses of this project. To analyze the Atlanta AQS data in a manner consistent with all the other data analyses conducted for this project would require appropriate pre-processing or averaging. This presents an additional data management and QA/QC effort before the use of the data. Finally, the data were limited to a very small number of HAPs (namely, benzene) for which the existing combined database is already relatively rich in data. For all these reasons, and considering project budget and time constraints, these data were not added.

E-mail received from Joe Touma August 27, 2002, 9:55 am

1. The reason I sent you the Atlanta data was to have you examine your extraction protocol to ensure that all relevant data have been extracted. I had wondered, and others will too, as to why some large data sources, that they are familiar with or have heard of, such as Atlanta, St. Louis, etc., are not included. If your response is they were not included because they did not include an MDL, then that should be clearly stated in the report.

Response: Done. See response to Item 2 of August 21 e-mail and response to August 22 e-mail above.

2. Also, are you sure that the reported MDL values are equipment-specific or generic? If generic, then it seems that you can apply the same MDL values that you use in other sites. Perhaps you can include a section on this topic, that you may have written elsewhere, that explains this issue.

Response: Yes, MDL values are equipment-specific. They are also specific to a number of other issues for that matter. A discussion and reference addressing some of the MDL issues have been added to subsection 2.1.1.

3. As to whether you should contact Susan [Zimmer-Dauphinee], that is up to you. I do not know how much work is involved in doing this nor whether these data will add any additional information (i.e., costs versus benefits).

Response: Due primarily to time and resource constraints, this has not been done to date. Susan is the Program Manager of the Ambient Monitoring Program located in Atlanta, Georgia. Note that results for benzene in Atlanta (Fulton county) are now included in the report. Also, refer to the comments and responses related to Atlanta data that have been provided above.

E-mail received from Joe Touma August 30, 2002, 3:40 pm

1. Table 2.1. Are you saying that Acetaldehyde has an MDL range from 0.0067 to 0.5405? If so, this looks like a wide range and is there a significance to it?

Response: Yes, and now the table has been updated as Table 2.1a with a range from 0.0005 to 0.5405. It is likely a function of changing MDLs over time, different measurement technologies, varying laboratory protocols for calculating MDLs, and alternative definitions or interpretations (i.e., semantics) made by different technicians when reporting MDLs. A discussion and reference has been added to Section 2.

2. Section 3.2.3, just for curiosity, how many pollutants passed Models 1, 2, and 3?

Response: The model development was done using benzene and carbon tetrachloride data. Models 1 and 2 were not applied to other pollutants. Further, since Model 3 reduces to Model 1 when the data are sufficiently far above the MDL, there is no need to run a separate model in that case.

3. Table 3.1, is the value for "All Sites, Background Estimate" (0.6282) a weighted value? I could not duplicate the standard error of 0.0483.

Response: The standard error is computed assuming that the errors for each of the estimates are mutually independent. Hence,

$$0.0483 = (0.2034^2 + 0.0593^2 + 0.0563^2 + 0.0644^2 + 0.0784^2)^(0.5)/5.$$

4. Table 3.2. This is a very important topic. Should this be done in specific counties? It seems when you are averaging so much data, the number is not meaningful. Also, should you also present by pollutant?

Response: [Should this be done in specific counties?] Table 3.2 summarizes the results of modeling each site and year separately. Then, for each year, the results were aggregated up to the county level. Individual county estimates varied (up and down) from year-to-year. The "trend" shown is less than the typical standard error bounds for an individual county and year estimate. Hence, it can only be seen at the aggregated level. [It seems when you are averaging so much data the number is not meaningful.] In the sense that it does not reflect the fact that individual estimates do go up and down, this is true. Averaging smoothes out the noise, which is what we were looking for in this exercise and this section of the report. [Also, should you also present by pollutant?] The results shown were a case study; benzene was chosen for the richness of the benzene data. While we could try other pollutants, the number of sites that would yield data across multiple years would be a limiting factor. A comment to this effect has been added to the report.

5. Section 4.4.1, 4th paragraph, 7th line. Can you explain your note as to not being credible and needing further investigation.

Response: This was referred to in conference call on 9/5/02. Because the model is fit using numerical approximation algorithms, we initially flagged some results and have since removed some results from consideration. These were places where the model converged to what we considered to be suspect values. They were suspect in that the model converged to what we considered to be an extreme. It is possible that the algorithm designed to find the maximum likelihood, instead found a set of parameters such that small changes to those values would result in a decrease in the likelihood. This will cause the algorithm to stop, even though the set of parameters found may not, in fact, maximize the likelihood. The automated routines have built-in checks to guard against this, but it is still a possibility. Also, note that standard errors are based on the assumption that maximization is correct. Since we have removed these estimates from consideration, the notes in the text have been removed also (and replaced with an appropriate text). Similarly, the associated footnote to Table 4.1 has been removed.

Conference call with Joe Touma and Ted Palma on September 5, 2002, 10:15 am

1. Provide background results in those cases for which the approach detailed in the August 15 report did not give a solution.

Response: Done. See discussion at the beginning of Section 3.1 and detailed discussion in subsection 3.1.7. Also, refer to results presented in Table 4.1 (compare to Table 4.1 results provided in August 15 report). Also, compare Tables 2.1a and 2.1b.

2. Mention in the report the minimum number of observations in the data set considered for conducting the primary analyses.

Response: Done. A discussion has been added to the end of subsection 3.1.3.

2. Update Table 4.1.

Response: Done.

4. Note in the text that the Beaverton site is in a different county than the downtown Portland, Oregon, sites.

Response: Done. The discussion in subsection 3.1.5 has been modified accordingly.

5. The Senior Battelle Statistical Review should include a review of any new methodology used to address data not handled by the primary method presented in the August 15 report.

Response: Done. Refer to Section C.3 of Appendix C.

6. Similar to the benzene case study presented in subsection 3.1.6, consider other pollutants (in particular a longer time series for metals).

Response: Not done, primarily due to time and resource constraints. A discussion about this possibility has been added to the text of subsection 3.1.6.

Other verbal comments received

1. Why are results for St. Louis not showing up?

Response: Refer to the investigation discussed in the response to Item 2 in the August 21 e-mail above.

2. Why are so many results showing up for Minnesota?

Response: In a conversation with Joe Touma and Tesh Rao, Tesh seemed to confirm that his experience with the archive often yielded an unusually large number of results for Minnesota as well (more than one might expect relative to other areas of the country). To date, no further investigation has been pursued to address this issue. A future discussion with ICF/SAI, the contractor who built the archive, may be warranted. Further discussion about this issue has been added in the results subsection 4.1.1.

3. Look into other toxics data that might be available through Battelle.

Response: This was explored, but did not result in any new data being added to the combined data set used for this project. The effort and results are discussed in more detail in subsection 2.1.1 of the report.

APPENDIX E:

EXTERNAL PEER REVIEW OF THE SEPTEMBER 27, 2002, DRAFT

APPENDIX E: EXTERNAL PEER REVIEW OF THE SEPTEMBER 27, 2002, DRAFT

An independent, external peer review was conducted on the September 27, 2002, revised draft version of this technical report (Contract 68-D-98-030, Work Assignment 5-09). Copies of the report were sent to three experts at different organizations (names withheld), with the following "charge to reviewers."

Charge 1: "Basic Methodology"

Subcharge 1.1	What is the validity of the technical approach described in Chapter 3.1 and results presented in Chapter 4.0? What are the limitations?
Subcharge 1.2	This method results in high background estimates where ever the monitoring values are high. Is background being overestimated in high pollution areas? Is this due to the influence of nearby sources and can this influence be statistically addressed?
Subcharge 1.3	There is high spatial variability in estimated background concentrations. Is this an artifact of the approach or is this realistic? Is this also due to the influence of nearby sources?

Charge 2: "Extension of Methodology"

Subcharge 2.1	What is the validity of technical approach described in Chapter 3.1 for extrapolating these background (point) estimates to other unmonitored area?
Subcharge 2.2	Can you suggest alternative methods for extrapolating estimated background concentrations to other areas based on readily available data such as land use, population density, emissions density, meteorology, terrain, etc?

Because of budget constraints, a line-by-line response to these peer reviews was not within the scope of the current work assignment. The bulk of the peer review comments concerned the decision regarding data that were below the method detection limit, or that were missing.

One editorial comment made by Reviewer #2, regarding an editing error in §3.1, 4th paragraph, was corrected in the current report. Reviewer #2 also noted the apparent discrepancy

in defining "nearby" sources as greater than 50km, as stated in §1.0, 2nd paragraph and again in §3.0, 1st paragraph. Battelle notes that this definition was provided to Battelle by EPA.

For future reference, the external peer review comments are presented below, essentially as received from the reviewers. Personal identifying information and affiliations have been removed.

Comments of Reviewer #1

Summary Comments on Estimating Background Concentrations: The purpose of this study is to develop a method for estimating the background concentration at an air monitoring site. For these purposes the background concentration is defined as the concentration contribution from emission sources **not** captured by the ASPEN model: natural sources, anthropogenic sources more than 50 km away (non-local anthropogenic sources), and unidentified sources. Neglecting the issue of unidentified sources, the background concentration is generally a measure of what we would expect the concentration to be in the absence of any local anthropogenic sources.

The general approach taken in this study is use all the measured values to estimate the background concentration, essentially defined as the minimum possible concentration at the receptor. (More precisely, if the MDL is zero or negligible, then Model 1 applies and the statistical shifted gamma model defines the true background as the minimum possible concentration. Otherwise, Model 3 applies, for which the true background parameter will be close to the minimum possible concentration). The estimated background is, thus, approximately equal to the minimum measured concentration. So a fundamental question about the approach is whether the minimum measured concentration at the location is likely to reflect only natural sources and non-local anthropogenic sources. In order for this to be the case, at least some of the measurements would have to be made at a time when there were no significant contributions from anthropogenic sources within 50 km (i.e., local anthropogenic sources). It is unlikely that any urban monitors would meet this criterion for many important air toxics. For example, virtually all urban locations are subject to significant contributions from mobile emission sources. Therefore, the estimated "background" concentrations of mobile source related pollutants estimated with this method are likely to include a local anthropogenic component. Supporting evidence for this view is the correlation between estimated background concentrations and total measured concentrations.

An alternative conceptual approach to estimating background concentrations, at least for the non-local source component (both anthropogenic and natural), would be to use measured values at locations greater than 50 km upwind of the target location. Of course, implementation would be complicated variability in wind direction and the sparseness of appropriate monitoring data. Still, it would be useful to assess how many metropolitan areas would have sufficient data for making such an estimate. For example, an area with a small number of predominant wind directions (e.g., one or two) and corresponding upwind monitors could be used to test this method. If it proved fruitful, this type of monitoring could be incorporated into future monitoring network design.

Charge 1: "Basic methodology"

Subcharge 1.1. What is the validity of the technical approach described in Chapter 3.1 and results presented in Chapter 4.0? What are the limitations?

- It is very much better to use the shifted gamma with zero probability below μ and censored at the MDL (Model 1 with censoring) than it is to use Model 3, which allows for concentrations below the "background," is censored at μ + MDL, linear up to μ + 2*MDL, and then shifted gamma. The Model 3 gives a definition of background (μ) that has no simple physical interpretation. For example, measured concentrations near zero are possible for a positive background. The Model 3 allows the distribution fitted to the measured physical concentrations to depend in an unusual way upon the MDL of the measuring instrument (it is plausible that a different MDL might change the shape of the distribution of the measurements but the form in Figure 3.3c is not very plausible).
- The likelihood is either incorrectly or not clearly stated for Models 2 and 3: for values below the censoring threshold C, the likelihood term should be F(C) rather than "constant," where F is the cumulative distribution function. Perhaps this is what was meant by the unclear last full paragraph on page 16.
- Some of the convergence problems reported for Model 1 are likely due to the fact that a shifted distribution with an unknown endpoint often leads to a non-regular maximum likelihood problem with likelihoods that approach infinity as the endpoint parameter approaches the smallest observed value. See Cheng and Traylor, Journal of the Royal Statistical Society series B, 57, 1, 1995, pp 3-44 for some ideas of how to deal with the convergence problem. For example, the maximum product of spaces method or grouping of the observations will lead to fewer convergence problems than maximum likelihood.
- A better modeling approach might be to assume that the basic distribution is the sum of a shifted gamma and an error term that could be assumed to be normal with mean zero and an unknown variance. The fitted distribution will be the basic distribution censored at the MDL. This allows for the possibility of observing measured values below the background level μ because of random measurement variation.
- The report does not explain whether the concentration data for each site used for these analyses are daily averages, averages over shorter monitoring periods (e.g., hourly, 3-hourly), or both. If the data have multiple monitoring periods, then the true distribution will be a mixture of several distributions and the assumed model will not be correct. If only daily averages were used, how were they computed and how was the daily MDL computed if there were several measurements on the same day?
- What was the treatment of concentration values with a missing or zero MDL?
- The discussion on page 15 treats the reported MDL's as a measurement error arguing that μ + 2*MDL is like 3*MDL which is often the value of the lowest calibration level (LCL). The report argues that values below the LCL are often not reported due to high uncertainties. In some cases, the so-called MDL value reported in the database could in fact be the LCL, implying that the censoring threshold used for the analysis, μ + 2*MDL, is of the order of 7*MDL and is far too high. In the old AIRS manual the "MDL" to be reported is defined as "Minimum Detectable Value" which could have several possible interpretations including both of Battelle's MDL and LCL.

- A bootstrap (or similar) estimate of the standard error is preferred in cases where SAS flags the standard error value as highly uncertain.
- The arguments for the secondary analysis approach using a background that is between the 5th and 50th percentile depending upon the MDL are not convincing and the method seems arbitrary. The argument for using the 50th percentile assumes the (measured) background varies daily, although the primary Model 3 assumes a fixed background throughout the year. Consistent use of the minimum or a low percentile (e.g., 5th) in these cases of non-convergence is preferable.

Subcharge 1.2. This method results in high background estimates wherever the monitoring values are high. Is background being overestimated in high pollution areas? Is this due to the influence of nearby sources and can this influence be statistically addressed?

- The statistical approach estimates the background for each site by fitting a model to the measured values at that site. The background for the county is estimated by averaging the estimated background across the sites in the county. The method will overestimate background at sites or counties where there is a strong nearby point or area source with consistent values throughout the year. In this case the estimated background will include the source contribution and will be overestimated. It would probably be better to use the county minimum estimated background (admittedly the minimum is more uncertain than the average) or an average or minimum background across a wider area. Another statistical approach would be to fit the statistical model simultaneously to multiple sites in an area, assuming that all the sites have the same background value but the other model parameters may vary between sites. Yet another approach would be to exclude sites known to be near large sources from the analysis.

Subcharge 1.3. There is high spatial variability in estimated background concentrations. Is this an artifact of the approach or is this realistic? Is this also due to the influence of nearby sources?

- The high spatial variability of the estimated background concentrations is an artifact of the approach since each site is independently modeled and may be influenced by a constantly emitting nearby source. See under Subcharge 1.2 for some suggestions to reduce this problem.

Charge 2: "Extension of Methodology"

Subcharge 2.1. What is the validity of the technical approach described in Chapter 3.1 for extrapolating these background (point) estimates to other unmonitored area?
Subcharge 2.2. Can you suggest alternative methods for extrapolating estimated background concentrations to other areas based on readily available data such as land use, population density, emissions density, meteorology, terrain, etc?

- The "Stage 2" multiple regression approach described in Section 3.2 is a reasonable initial general approach for estimating background at non-monitored sites. The selection of which explanatory variables to use, and which functional transformations (e.g., include squared or cubed latitude) or interaction terms should be included, is the hardest part of

any regression-based method. Two important variables for this approach are the latitude and longitude, and it would be a good idea to also use their squares and product as explanatory variables to fit a quadratic rather than linear surface. Land use is a useful surrogate for natural sources. Suitable summary statistics of the emission density at least 50 km to, say, about 75 km away could address impacts from non-local sources. A suitable function of the average wind speed (e.g., reciprocal) could summarize important meteorological impacts.

Comments of Reviewer #2

Overall, this report needs a great deal of clarification about both the methodology and the analysis performed. Assumptions need to be justified and there needs to be discussion about how realistic the assumptions are and how the conclusions are affected when the assumptions are not met. This report requires substantial revision before it can be considered for publication as an EPA report.

A number of points I make here were raised in the review of this report by Battelle statisticians. Most of the comments of the Battelle statisticians were not addressed at all or not addressed satisfactorily by the authors. These comments need to be properly addressed.

The definition of background concentrations provided to the authors makes estimation of background almost intractable, since the background concentrations depend on the modeling emissions inventory. Thus, if the same scenario is modeled (e.g., using ASPEN) twice, the second time with an improved emissions inventory, the true background concentrations will be different. Clearly it is problematic to estimate background using only monitored data, since one important factor (the inventory) is ignored.

****

Working definition of background concentrations

For the purposes of Battelle's analysis, background concentrations are to be estimated so that, when added to ASPEN model predictions, the resulting concentrations approximate the ambient concentrations. From page 1:

> *In order to estimate **total** ambient concentrations, however, a value for background must be estimated and added to ASPEN's modeled concentrations. As defined, background accounts for natural sources, nearby sources (farther than 50 km), and unidentified sources.*

****

Conceptually, this method for estimating background concentrations should have a hard time distinguishing the impact of sources with small temporal variation within 50 km of the monitor from sources further away than 50 km. The monitor will see variation from both sources due to varying wind direction, but there will not be any additional variability in the

closer source due to emissions fluctuations. As many sources of toxics may have fairly constant daytime emissions, this may be a serious shortcoming of the method, and lead to overestimation of background. If so, this would be manifest in a positive relationship between the ambient concentrations and the estimated background concentrations.

There should be a discussion of the difference between uninventoried sources (part of background) and underestimated sources (not part of background), in relation to estimating background concentrations.

Statistical goodness-of-fit tests quantitatively describing the adequacy of the proposed model should be presented.

The failure of the method to converge in some cases is mentioned in several places in the report, even when the amount of data should be adequate. Why is this happening? This is important to know as it might indicate how the methodology could be improved. Failure to converge might indicate that the form of the model is inappropriate. (Of course, it is difficult to judge this without goodness-of-fit tests)

This methodology might be improved on by treating background concentrations probabilistically by modeling ambient concentrations as a mixture of two distributions, one for background concentrations and one for source-oriented concentrations. It is worth exploring.

SPECIFIC COMMENTS

§3.0 The definition of "nearby sources" includes transport from distant cities and transport from around the globe. This terminology is not apt.

§3.1, 2nd paragraph.

I disagree with the statement "*The idea behind such approaches is that background levels are more clearly discernible under certain preconceived notions about ideal conditions*".

The statement "*Their disadvantages include … (2) a reduction in the extensiveness of results due to the failure of many sites or areas to exhibit the required extreme event data*" is unclear. What is meant by this?

The following statement is not true. With some techniques, eliminating some data can reduce noise obscuring the background signal, resulting in a better estimate of background, which should not be characterized as a disadvantage of the method.

> *In addition, evidence of background concentration levels is likely contained within all measurements, so any approach that eliminates much of the monitoring data leads to a deliberate reduction of available information about background.*

§3.1, 3rd paragraph.

Empirical, or nonparametric, approaches are described as "less desirable" and are simplistically dismissed:

> *Another somewhat less desirable class of approaches might be described as empirical methods, which amount to calculating some percentile of a given set of monitoring data and treating the resulting threshold as an estimate of background. For example, use the 5-10 percent lowest measurements…*

First, nonparametric approaches are not necessarily less desirable. Second, they don't necessarily "amount to calculating some percentile." Some nonparametric approaches are quite sophisticated. It may be that a nonparametric approach would yield better results than the approach taken in the report.

I question the use of a nonparametric approach in cases of "too few observations." Generally, nonparametric approaches require more data than parametric approaches.

§3.1, 4th paragraph.

> *the decision was made to pursue (as a primary analysis approach) a method that: (1) uses all the information about background that exists within a full set of monitoring data (above or below an MDL); (2) is simple enough to be applied on a routine basis in practice; and, most importantly, (3) is consistent with the ASPEN model's conceptualization and treatment of background levels.*

This is good. A method that satisfies these three points is appropriate.

§3.1, 4th paragraph.

> *For urban areas it is assumed that, except for very rare instances, all data values contain a constant background contribution and a varying, non-negative source-oriented contribution.*

This assumption is clearly not true (nearby and uninventoried sources will have time-varying impacts). The question that needs to be addressed is: "How will the background estimates be impacted when the variance of the background concentrations is X percent of the variance of the known-source-oriented concentrations?" where X is in a realistic range.

§3.1.1 I agree that the gamma distribution is appropriate for modeling concentrations. However, justification for using it is not there. Essentially, the justification in the report is that the gamma distribution has certain features in common with the lognormal distribution and that the lognormal distribution is known to adequately approximate concentration distributions, but the lognormal distribution can't be used because of "application difficulties."

One way to address this would be to remove the sentence "*So, in summary, the shifted gamma distribution appears most appropriate for serving the dual purposes of adequately approximating monitoring data behavior and providing a direct estimate for background.*"

§3.1.1, last paragraph.

The authors claim that "Figure 3.2 supports the assertion that the shifted gamma pdf is, in fact, a reasonable choice for the statistical distribution to be used for the approach to this project." However, this figure indicates a poor fit in the most important region, the lowest concentration quantiles. (It would be helpful to have a 1-1 line on this plot.)

It is imperative to present statistical goodness-of-fit tests quantitatively describing the adequacy of the proposed model.

§3.1.3, 2nd paragraph.

> *Specifically, it was decided that any data within a threshold of background plus two times the reported MDL (+2*MDL) should be treated as random noise, or at least too imprecise to use individually for estimating the parameters of the assumed shifted gamma distribution.*

Why is the background level μ part of the threshold? For some pollutants μ can be above the MDL. If μ is well above the MDL, then measurements well above the MDL are considered too imprecise? Suppose (as could happen in a rural area) μ ≈ the median concentration and all measurements are above the detection limit? Then more than half of the data are censored and none should be. Perhaps I'm missing something here.

§3.1.3, 3rd paragraph.

In Model 2, what is the "constant?" Is it a fourth parameter to be estimated? Is it μ?

More information should be provided on how SAS was used to fit this model. Since the parameter μ to be estimated is used to define the model and, therefore, the model changes for different estimates of μ, the procedure is presumably not straightforward.

§3.1.3, 4th paragraph.

Model 2 uses all the data without censoring (as stated in paragraph 3). However, in this paragraph, the authors state that data below the censoring threshold are censored: "*The decision to censor data up to a threshold of (+2*MDL) served to stabilize this procedure*…" This discussion of the model needs to be clarified.

§3.1.3, 6th paragraph.

The rationale for reporting "*the standard error for the background estimate as the maximum of: (1) the uncertain standard error given by SAS , (2) one-half of the background estimate, and (3) one-half of the MDL*" should be described for (2) and (3).

The statements "*This approach provides a conservative uncertainty estimate in such cases. That is, uncertainty will tend to be over-estimated*," if true, need to be justified.

§3.1.7, page 25.

The failure of the method to converge in some cases is mentioned in several places in the report, even when the amount of data should be adequate:

> *This primary methodology may also fail in a few cases when the size of the data set would appear to be adequate and the data are mostly above their respective MDLs, yet the given application still does not adequately converge to a solution.*

Why is this happening?

§3.1.7, page 25+ describes the Secondary Analysis Approach. There are several assumptions made in this section for which there is no justification, and some of them appear to be incorrect. This section should be revised to include such justification.

For example, the report states:

> *The first extreme is when a site is not affected by sources and is only measuring background. In this case, the median, or 50th percentile, of the data is considered to provide a reasonable upper bound estimate for the annual mean background concentration.*

If all the measurements reflect background, then clearly the median of the data does not provide a reasonable upper bound estimate for the annual mean. An upper bound could be constructed from confidence intervals for the mean.

§3.2, Stage 2: Spatial extrapolation of background estimates

Why restrict the model to be linear?

It will be important to have estimates of uncertainty for any estimates of background concentrations in areas without any monitoring data.

At the end of this section the authors give the opinion that the proposed model will provide only a poor fit (an R^2 of at best 0.3). If so, why waste time pursuing this model?

§4.1.1 Summary of Stage I results

There are some results that beg for explanation. For example (page A-3, top right) Kandiyohi County, MN has a background estimate of 0.37 with an uncertainty of 0.0005, based on a sample size of only 32 measurements at one site. This uncertainty is clearly much too small.

Then there is Hennepin County, MN which has 678 measurements at six sites with a background estimate of 0.54 and an uncertainty of 0.05. This county has 20 times as much data as Kandiyohi, yet the uncertainty is 100 times greater. ?
An uncertainty of 0.05 also seems small.

RESPONSE TO CHARGE QUESTIONS

The rationale for these responses is provided in the comments above.

Charge 1: " Basic Methodology"

Subcharge 1.1 What is the validity of the technical approach described in Chapter 3.1 and results presented in Chapter 4.0? What are the limitations?

The report's description of the technical approach is sufficiently incomplete that it is difficult to assess its validity and limitations. There appear to be a number of unjustified assumptions. The results indicate that the method does not work well.

Subcharge 1.2 This method results in high background estimates where ever the monitoring values are high. Is background being overestimated in high pollution areas? Is this due to the influence of nearby sources and can this influence be statistically addressed?

Yes, with this method the background will tend to be overestimated in areas with high ambient concentrations due to local sources. Due to the nature of the definition of background, this cannot be addressed statistically without using some type of air quality model.

Subcharge 1.3 There is high spatial variability in estimated background concentrations. Is this an artifact of the approach or is this realistic? Is this also due to the influence of nearby sources?

I don't think that this can be determined without further analysis. I would expect that there would be significant spatial variability in background concentrations, due to the influence of "nearby" sources and uninventoried local sources. (Note that "nearby" sources includes distant sources.) However, the high variability in the results could certainly be an artifact of the approach.

Charge 2: "Extension of Methodology"

Subcharge 2.1 What is the validity of technical approach described in Chapter 3.1 for extrapolating these background (point) estimates to other unmonitored area?

There is some discussion about such an approach in §3.2, where the approach is to develop a multiple regression model of the background based on various possible explanatory variables. The authors doubt that this approach will be satisfactory and so do I.

Subcharge 2.2 Can you suggest alternative methods for extrapolating estimated background concentrations to other areas based on readily available data such as land use, population density, emissions density, meteorology, terrain, etc?

One way to extrapolate estimated background concentrations to other areas is to use the mean or median of the estimates based on available measurements. This would not capture the geographic variability of background concentrations and will be biased (since the vast majority of areas have no data). An attempt would have to be made to estimate upper confidence bounds for this. Zero can be used as a lower bound. In my opinion, the only way to improve on this is to bring in additional data such as land use, population density, emissions density, meteorology, and terrain. The best way to use these kinds of data to estimate concentrations (background or ambient) is through multi-scale air quality dispersion modeling combined with a statistical approach.

Comments of Reviewer #3

My overall impression is that this is a good piece of work on an important problem. The approach used to estimate the background concentration of air toxics is sound. The conceptual model used is that there is a regional "background" concentration caused by non-local and/or un-inventoried sources that can be estimated at the left-hand side of the distribution of observed concentrations. For this approach to work, the component of the observed concentration that is not regional background must follow a known probability density function (distribution). Since ambient observations often follow a skewed distribution (log-normal or gamma), the portion of the observation that is not fit to the distribution (at the left hand tail) represents the background.

The authors have selected the shifted gamma distribution for this analysis. The advantage of the shifted gamma distribution is that the shift-parameter (μ) represents an estimate of the background. This approach also allows for censored data, that is data where some of the observations are below the analytical limit of detection. Following are specific concerns and suggestions.

I. The conceptual model specifies that "background" is a constant. However, if background also includes un-inventoried sources, it may not be constant and may be better represented as a distribution. However, I do not believe that most air toxic data sets

contain sufficient precision at low concentrations to support a more complex conceptual model and the constant assumption is a reasonable starting point.

II. I believe that the shifted gamma is a reasonable choice. However, I am not sure why the data were censored at the background + 2 MDL. Why is this extra censoring necessary? It would be more straightforward to include all measures above the MDL in the analysis.

III. The model selected specifies that the distribution of concentrations is approximately shifted gamma if greater than μ + MDL and "constant" otherwise. This is shown in figure 3.3c. However, the figures shows that the concentration is not constant below the μ + MDL point, but rather the concentration distribution is uniform below the μ + MDL point. I don't see what this adds. If the background concentration is "constant", than the shifted gamma should be anchored at the background and the model in figure 3.3a should be used. The degree of censoring should not affect the shape of the distribution.

IV. The key question is does the shifted gamma distribution fit observed concentrations? If it does, great. If not, why not? It is stated on page 15 that the numerical model did not always converge and that increased censoring was needed. However, I am worried that lack of convergence may have been simply due the lack of fit between the observed data and gamma distribution (e.g., the actual impact of inventoried sources may have been bimodal or some other distribution). Increased censoring could have had the effect of simplifying the actual distribution allowing convergence for a higher percentage of data sets. If this is true, then the estimates of background will be high. (Note that "high" estimates of background were occasionally found.)

V. If you go back to the simple shifted gamma (Figure 3.3a), what percentage of cases fail to converge? For the data sets that converge, are the estimates of background lower and/or more consistent that estimates from the model in Figure 3.3c? It may be better to live with a higher number of cases where the model does not converge if it produces more consistent results. This should be a testable assumption. Also, an analysis of why the model fails to converge may provide additional insight. Finally, what is the effect of the level of censoring on the estimated background? If the model is working properly, the level of censoring should not affect the estimated background. This should also be testable.

VI. The decision to apply the model to each monitor and then produce a regional average by averaging between monitors is good.

VII. The decision to censor the highest measurements in interesting. It has been my experience that fitting observed data to distributions is always problematic at the tails so censoring the highest points has appeal. Did you test the effect of clipping different amounts of data off of the top of the distribution? Why 1 %? Why not 5 or 10%?

VIII. The things in figure 3.4 are not box plots. A standard box plot gives much more information than included in figure 3.4.

IX. Include the shifted gamma parameters in table 3.1. It would also be interesting to see the parameters (or a summary of the parameters) in Appendix A.

X. The results in Table 3.2 are very interesting. Benzene concentrations are supposed to be decreasing due primarily to changes in motor vehicle fuels. What this table is showing is that the background estimate is not completely separate from motor vehicle emissions.

XI. I understand the need for a secondary analysis approach. Overall, how do the background estimates developed by the shifted gamma approach compare to the estimates from the secondary analysis?

XII. The approach proposed for the stage 2 explanatory analysis is reasonable. However, finding a model with an R^2 of 0.3 may be a bit optimistic

XIII. Figure 4.1 needs to be printed in color to be readable. However, even in color, I am not sure what it will add.

Overall, I think this is a very interesting project.