Environ. Sci. Technol. **1999**, *33,* 4176−4187

Volatile Organic Compounds in Untreated Ambient Groundwater of the United States, 1985−1995

PAUL J. SQUILLACE,*
MICHAEL J. MORAN, WAYNE W. LAPHAM,
CURTIS V. PRICE,
RICK M. CLAWGES, AND
JOHN S. ZOGORSKI

U.S. Geological Survey, 1608 Mt. View Road, Rapid City, South Dakota 57702

As part of the National Water-Quality Assessment Program of the U.S. Geological Survey, an assessment of 60 volatile organic compounds (VOCs) in untreated, ambient groundwater of the conterminous United States was conducted based on samples collected from 2948 wells between 1985 and 1995. The samples represent urban and rural areas and drinking-water and nondrinking-water wells. A reporting level of 0.2 μg/L was used with the exception of 1,2-dibromo-3-chloropropane, which had a reporting level of 1.0 μg/L. Because ambient groundwater was targeted, areas of known point-source contamination were excluded from this assessment. VOC concentrations generally were low; 56% of the concentrations were less than 1 μg/L. In urban areas, 47% of the sampled wells had at least one VOC, and 29% had two or more VOCs; furthermore, U.S. Environmental Protection Agency drinking-water criteria were exceeded in 6.4% of all sampled wells and in 2.5% of the sampled drinking-water wells. In rural areas, 14% of the sampled wells had at least one VOC; furthermore, drinking-water criteria were exceeded in 1.5% of all sampled wells and in 1.3% of the sampled drinking-water wells. Solvent compounds and the fuel oxygenate methyl *tert*-butyl ether were among the most frequently detected VOCs in urban and rural areas. It was determined that the probability of finding VOCs in untreated groundwater can be estimated on the basis of a logistic regression model by using population density as an explanatory variable. Although there are limitations to this national scale model, it fit the data from 2354 wells used for model development and adequately estimated the VOC presence in samples from 589 wells used for model validation. Model estimates indicate that 7% (6−9% on the basis of one standard error) of the ambient groundwater resources of the United States probably contain at least one VOC at a reporting level of 0.2 μg/L. Groundwater is used in these areas by 42 million people (35−50 million based on one standard error); however, human exposure to VOCs from this ambient groundwater is uncertain because the quality of the finished drinking water is generally unknown.

Introduction

Considerable quantities of volatile organic compounds (VOCs) are produced in the United States, and their use is ubiquitous. The production of synthetic organic chemicals (many of which are VOCs) has increased by more than an order of magnitude between 1945 and 1985 (*1*). VOCs are contained in many manufactured products, including paints, adhesives, gasoline, and plastics. The U.S. Environmental Protection Agency (U.S. EPA) Toxics Release Inventory (TRI) (*2*) provides information on the release of toxic chemicals from manufacturing facilities in the United States. These data show that 10 of the 20 chemicals with the largest releases to the environment are VOCs and that the total release for these 10 compounds was almost 500 million kg during 1996 (*2*). There are other anthropogenic sources of VOCs that are not included in the TRI. For example, the inventory does not include emissions and evaporation from mobile sources, such as automobiles, or commercial activities not involved in manufacturing, such as refueling stations and dry-cleaning operations. Leaking underground storage tanks are an important source of groundwater contamination that is not necessarily included in the TRI; only leaking tanks associated with large manufacturing facilities are reported.

VOCs can be important environmental contaminants because many are mobile, persistent, and toxic. The environment consists of a complex system of interacting media (for example, atmosphere, soil, surface water, and groundwater), and VOCs do not necessarily remain in the medium where they originate. In certain media, many VOCs can have a very short half-life of a few hours due to degradation, whereas in other media they can be very persistent and show little degradation over a period of years. In most urban areas, VOCs can contribute substantially to the total cancer risk associated with toxic air pollutants (*3*). The U.S. EPA has established maximum contaminant levels in drinking water for 27 VOCs because of human health concerns (*4*).

Between 1975 and 1981, the U.S. EPA conducted several national surveys to determine the occurrence of VOCs in finished drinking water obtained from groundwater and surface water sources (*5, 6*). The number of groundwater systems sampled ranged from 16 to almost 1000 for the different surveys. The minimum reporting levels for many of the VOCs were greater than 0.2 μg/L. These surveys provide a good indication of the quality of the water that people were consuming. However, they did not give a good indication of the quality of the groundwater resource because VOCs can be gained or lost during the water treatment and distribution process. For example, trihalomethanes can form during chlorination at the treatment plant or in the distribution system, and some VOC concentrations can decrease during water treatment processes such as aeration. Furthermore, finished drinking water can be blended from many wells, and the quality of the finished water may vary depending on which wells are being pumped. There have been no prior national surveys on the occurrence of VOCs in untreated ambient groundwater.

The purpose of this paper is to provide an assessment of the occurrence, distribution, and status of VOCs in untreated ambient groundwater of the conterminous United States on the basis of samples collected between 1985 and 1995. This assessment was done as part of the National Water-Quality Assessment Program (NAWQA) of the U.S. Geological Survey (USGS). The results of this assessment may differ from future assessments as more data are collected from different aquifers; therefore, this paper should be viewed as a progress report on the assessment of VOCs in groundwater. Ambient groundwater is defined here as groundwater in areas where there are no known point sources of contamination prior to sampling. For example, as of March 1997, there were about

* Corresponding author phone: (605)355-4560, ext. 239; fax: (605)-355-4523; e-mail: pjsquill@usgs.gov.

10.1021/es990234m
Published on Web 10/29/1999

330 000 confirmed leaks from underground storage tanks in the United States (*7*). The contaminated areas associated with these sites are excluded from this assessment. All VOC detection frequencies are relative to a reporting level of 0.2 μg/L, with the exception of 1,2-dibromo-3-chloropropane, which had a reporting level of 1.0 μg/L. A 0.2 μg/L reporting level was chosen because it is relatively common and provides the greatest information of low-level concentrations of VOCs in groundwater. Using a common reporting level facilitates comparisons between compounds but required the elimination of some analyses that used higher reporting levels. Occurrence and distribution pertains to the incidence and location of VOCs in groundwater and status pertains to VOC concentrations relative to drinking-water criteria (regulatory or advisory levels) and taste/odor thresholds. This paper also provides an estimate of the area affected by VOCs in ambient groundwater and the number of people using groundwater in areas where VOCs may be in the untreated ambient groundwater resource.

Study Methods

Source, Requirements, and Summary of VOC Data. Data used in this assessment were compiled by the USGS from studies conducted by the USGS and by local, State, and other Federal agencies. A detailed study plan for the compilation of these data has been published (*8, 9*), and the data have been summarized (*9, 10*). When data are aggregated from various sources, inconsistencies among the sources are a concern. Therefore, the study plan was written and implemented to minimize differences in network design, data collection, and laboratory analysis methods used to obtain the aggregated data.

All the wells in the compiled data set were part of well networks that were designed to define ambient water quality in an aquifer or part of an aquifer. If the original network contained a large number of wells, a subset was randomly selected to prevent certain areas of the Nation from being overrepresented, which would bias this national assessment toward water quality in those areas. The well networks included in this assessment generally monitored aquifers used for drinking water. However, some networks monitored aquifers that were considered potential sources of drinking water or aquifers that were hydraulically connected to deeper groundwater used as a source of drinking water. Although data from known point-source contamination sites were excluded from the data set, not all VOC concentrations documented in the data are from nonpoint sources. For instance, large concentrations of VOCs documented in this assessment probably originated from point sources that were not previously identified. Furthermore, low concentrations could have originated from point sources where VOCs migrated beyond the limits of the site investigation. Examples of nonpoint sources of VOCs include stormwater runoff, precipitation scavenging and gaseous diffusion of VOCs from the air, and application of fumigants.

Samples from all networks were collected such that the analyses were representative of the water quality of the aquifer; samples were collected before contact with pressure tanks and water treatment and were analyzed by laboratories certified by the U.S. EPA. The method of analytical analysis was not always reported; however, for at least 50% of the data, the method of analysis was similar to U.S. EPA Method 524.2 (*11*). At least 19 VOCs were analyzed for each sample. Typically, samples with fewer analytes (<30) included the more frequently detected compounds; consequently, these samples did not have fewer VOC detections when compared to samples with more analytes. There were 59 VOCs analyzed at the 0.2 μg/L reporting level. One additional VOC, 1,2-dibromo-3-chloropropane, had a reporting level of 1.0 μg/L because 97.5% of the 1492 analysis were at 1.0 μg/L. A higher reporting level was used for this compound even though the maximum contaminant level for 1,2-dibromo-3-chloropropane is 0.2 μg/L. There were only five detections of 1,2-dibromo-3-chloropropane; therefore, this compound had little effect on the overall VOC detection frequency in this assessment. Furthermore, four of the five samples from wells with 1,2-dibromo-3-chloropropane detections had other detectable VOCs at a reporting level of 0.2 μg/L.

A minimum well separation of 1000 m was imposed in the final data set containing 2948 wells to limit autocorrelation among samples. Other spacing criteria were tested and are discussed later. However, using a 1000-m spacing criterion, 50% of the data were sampled for NAWQA; the remaining wells were sampled by local, State, and other Federal agencies or programs.

For this VOC assessment, urban and rural areas were distinguished using only population density; those areas with population density $\geq$386 people/km^2 [1000 people/mi^2, (*12*)] were considered urban, and those areas with <386 people/km^2 were considered rural. This distinction allowed urban areas to be delineated outside of political boundaries. The number of housing units in an area with a population density of 386 people/km^2 varies; however, 0.3 unit/ha (0.7 unit/acre) is typical for Virginia (*12*). Areas designated as "high" population density started at 2000 people/km^2 (*12*). There were five wells used in the assessment for which population density was unknown. Because these five wells were not associated with any major urban area, they were assumed to be in rural areas. For the data summaries presented herein, there were 2542 wells in rural areas and 406 wells in urban areas. Among the wells in urban areas, 356 wells were in or near 38 metropolitan areas, and the remaining 50 wells were in small urban areas or located outside metropolitan areas.

Figure 1 shows the location of the major productive aquifers of the United States (modified from ref *13*) and the location of the wells used in this assessment. For many of the wells, it is not known if the screened intervals were within these productive aquifers. Nevertheless, even if the wells were not screened within the productive aquifer, there may be hydraulic connection to the productive aquifers from where the wells were screened. Almost all the aquifers represented by the data are used at least in part for drinking-water supply. Table A1 (in the Supporting Information) summarizes the location of wells from which data were compiled in relation to the productive aquifers and to urban and rural areas.

The depth and type of the well and the type of aquifer (confined/unconfined) were not documented for most of the data. In fact, for 60% of the wells, depth and well type is not known; for 92% of the wells, type of aquifer is not known. Among the wells with depth information, the median well depth was 40 m, and 75% of the well depths were less than 90 m. Many types of wells were sampled; for example, 80 urban and 718 rural drinking-water wells, 192 public and 606 domestic wells, and 166 designated monitoring wells.

Development and Validation of Logistic Regression Model. Logistic regression was used to develop a national scale model to predict the probability of detecting at least one VOC in groundwater at a reporting level of 0.2 μg/L. Detections of the compound 1,2-dibromo-3-chloropropane, which had a reporting level of 1.0 μg/L, were not included in the data used for model development or validation. Logistic regression was chosen because (i) it permits data with a low frequency of detection to be used and (ii) relations between groundwater quality and explanatory variables can be investigated. Explanatory variables tested for inclusion in this model had to be available on a national scale and would provide relevant information that may help predict the presence or absence of VOCs in ambient groundwater.

Factors that affect the transport and fate of VOCs in the subsurface are well documented; these include source

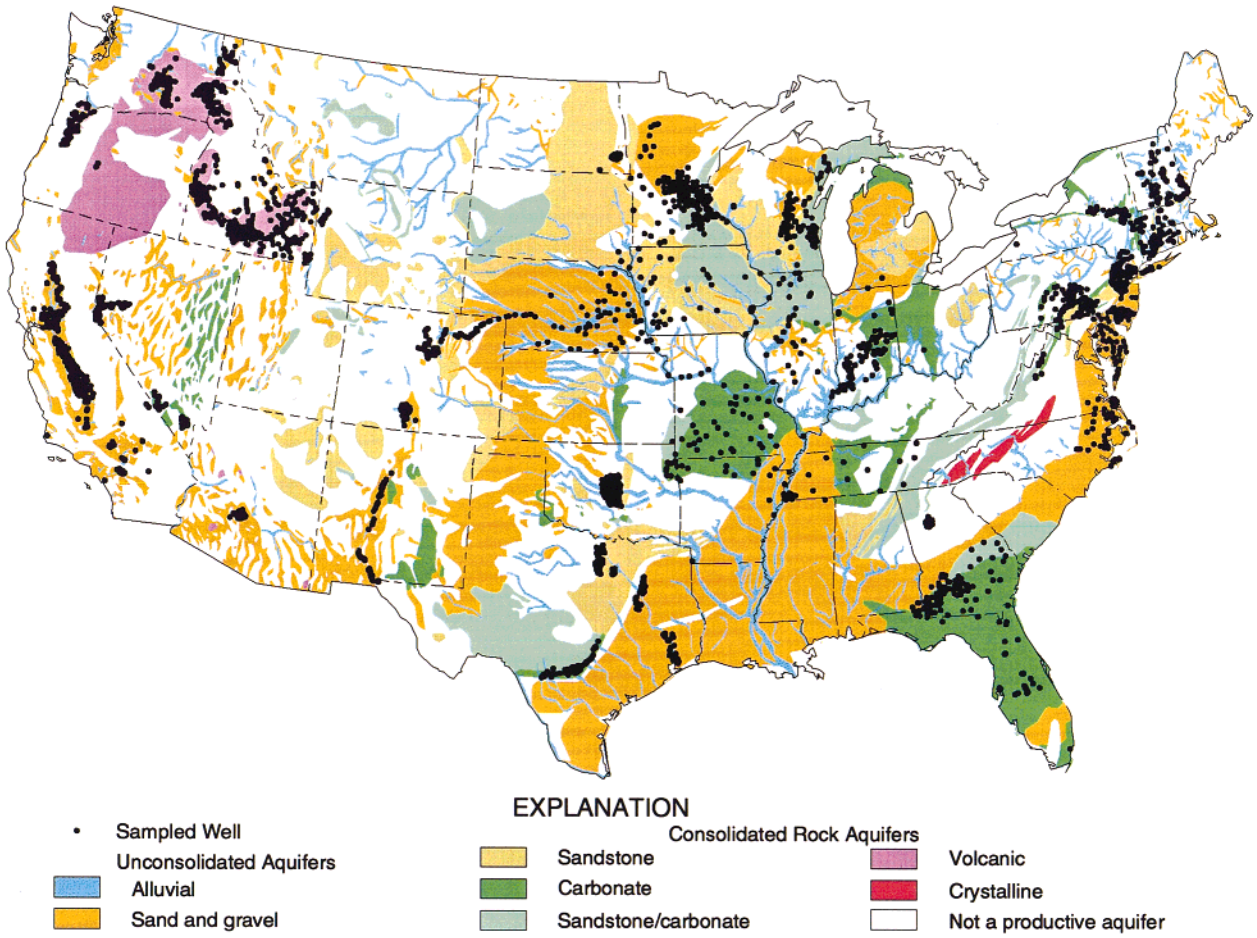


FIGURE 1. Location of 2948 sampled wells used in the final data set for this assessment in relation to productive aquifers in the conterminous United States, [modified from the U.S. Geological Survey (*13*)].

strength, permeability, porosity, recharge, dispersion, biodegradation, and others. These variables could be used in a model that explains the presence or absence of VOCs in groundwater; however, accurate estimates of these factors were not available on a national scale. Nevertheless, various multivariate models were evaluated using explanatory variables such as soil permeability, aquifer hydraulic conductivity, and estimated groundwater recharge. Some of these variables were statistically significant ($p < 0.05$); however, following the model-building strategies outlined by Hosmer and Lemeshow (*14*), it was determined that a univariate model using population density was the best predictive logistic regression model providing estimates of VOC detections on a national scale. If accurate estimates were available for some of the transport variables, they probably would have improved the logistic regression model.

Population density can be used to estimate the probability of detecting VOCs in ambient groundwater. The usefulness of population density as an explanatory variable has been documented by others (*15, 16*). Population data for 1990 are available from the U.S. Bureau of the Census (*17*). Population density was calculated from the census block data by converting the data to 1-km^2 grids and then interpolating the population density for each well. Of the 2948 wells in the data set, five wells did not have population-density information, and 97 had a population density of zero; these 97 wells are located in large unpopulated areas. Those wells where population density was undefined could not be used in the model. The lowest calculated population greater than zero was 0.001 people/km^2. Population densities of zero were replaced by 0.001 in the logistic regression model, as the natural log of zero is undefined.

Wells were divided into subsets—2354 wells for model calibration and 589 wells for model validation. The data were divided by assigning a random number between 0 and 1 to each well and then selecting wells with values greater than 0.8 for the validation data. The validation data were not used during model calibration and had no effect on the model results; these data were used only to verify the model. Model validation is especially important when a model will be used to estimate future outcomes because the model always performs optimally using the calibration data (*14*).

Two diagnostic tests were used to assess the results of the logistic model. First, the −2 log likelihood statistic can be used in univariate models to assess if the regression slope coefficient is significantly different than zero; a slope near zero indicates that the explanatory variable has little or no estimation value. The −2 log likelihood statistic has a χ^2 distribution with an associated probability of significance (p value) that is based on the degrees of freedom in the model (a p value less than 0.05 for this statistic was considered significant).

Second, the Hosmer−Lemeshow goodness-of-fit statistic (*14*) was used to assess how the model fit the data. This statistic was computed by ranking the estimated probabilities of VOC occurrence for each well and then dividing the data into 10 groups, each group having about the same number of wells. Therefore, population densities represented within each group would depend on the data being tested. The model probability estimates were then compared to the VOC

detection frequency, as represented by the data, within these 10 groups. The null hypothesis for the Hosmer–Lemeshow test is that the model fits the data. Therefore, a high p value would indicate a better model fit. Calculation procedures for the Hosmer–Lemeshow test were identical for the calibration data and the validation data, except that the values for the model coefficients were regarded as fixed constants during the validation step rather than estimated values during the calibration step. The Hosmer–Lemeshow test provided a conservative assessment of how the model fit the calibration and validation data.

Results

Occurrence and Distribution of VOCs Based on Data Summaries. At a reporting level of 0.2 μg/L with the exception of 1,2-dibromo-3-chloropropane, which had a reporting level of 1.0 μg/L, VOCs were detected in 47% of the sampled wells in urban areas and 14% of the sampled wells in rural areas. In urban areas, 36 VOCs had a detection frequency greater than 1%; in rural areas, only 6 VOCs had detection frequencies exceeding 1% (Figure 2 and Table A2 in the Supporting Information). In urban areas, 18 VOCs were not detected; in rural areas, 16 VOCs were not detected. Of 60 VOCs, only 14 VOCs were not detected in either urban or rural areas.

Many wells had samples with multiple VOC detections (Figure 3); 29% of the wells in urban areas and 6% in rural areas had two or more VOCs. The most frequently detected VOCs in urban areas were MTBE, tetrachloroethene, trichloroethene, and trichloromethane. These four VOCs commonly co-occur; among wells in urban areas with samples having at least one of these VOCs, 44% had at least two of the four. They co-occur because their occurrence is widespread (had frequencies of detection >10%) not because they are necessarily used together. MTBE is used in gasoline, and the remaining three VOCs are classified as solvents. However, trichloromethane can form during chlorination; lawn irrigation, or leaking water mains and sewers then can introduce this VOC to urban groundwater. Tetrachloroethene can degrade to trichloroethene, and this may partially explain why these two VOCs co-occur.

Well type (domestic water supply, public water supply, monitoring well, etc.) was not defined for 60% of the wells used in this resource assessment; therefore, no definitive analysis was performed to determine the affect of well type on water quality. There were 606 wells designated as domestic drinking-water supply, 192 were designated public water supply, 150 were monitoring, 265 had other designations (such as irrigation), and the remaining 1735 were undesignated (could be any type of well). Data for this assessment were compiled to define the water quality of the resource and not to distinguish the affect of well type on water quality. However, drinking-water quality often is put in a separate class than all other type of well data; consequently, a comparison of the drinking-water data was made with all other types of wells (monitoring and other types of non-drinking water wells and undesignated wells). To make this comparison, wells in similar population density areas were compared to minimize the effect from land use. In urban areas, there were limited public and domestic wells (80 total); however, 41% had a detectable concentration of a VOC. In rural areas, VOCs were detected in 11% of 718 public and domestic water supplies. These frequencies of detection are similar to those calculated for the rest of the wells—49% for wells in urban areas, 15% for wells in rural areas. The Pearson χ^2 test of independence for two-way tables, at the 95% confidence interval, showed that well designation significantly affected VOC detection frequency in rural areas but not in urban areas.

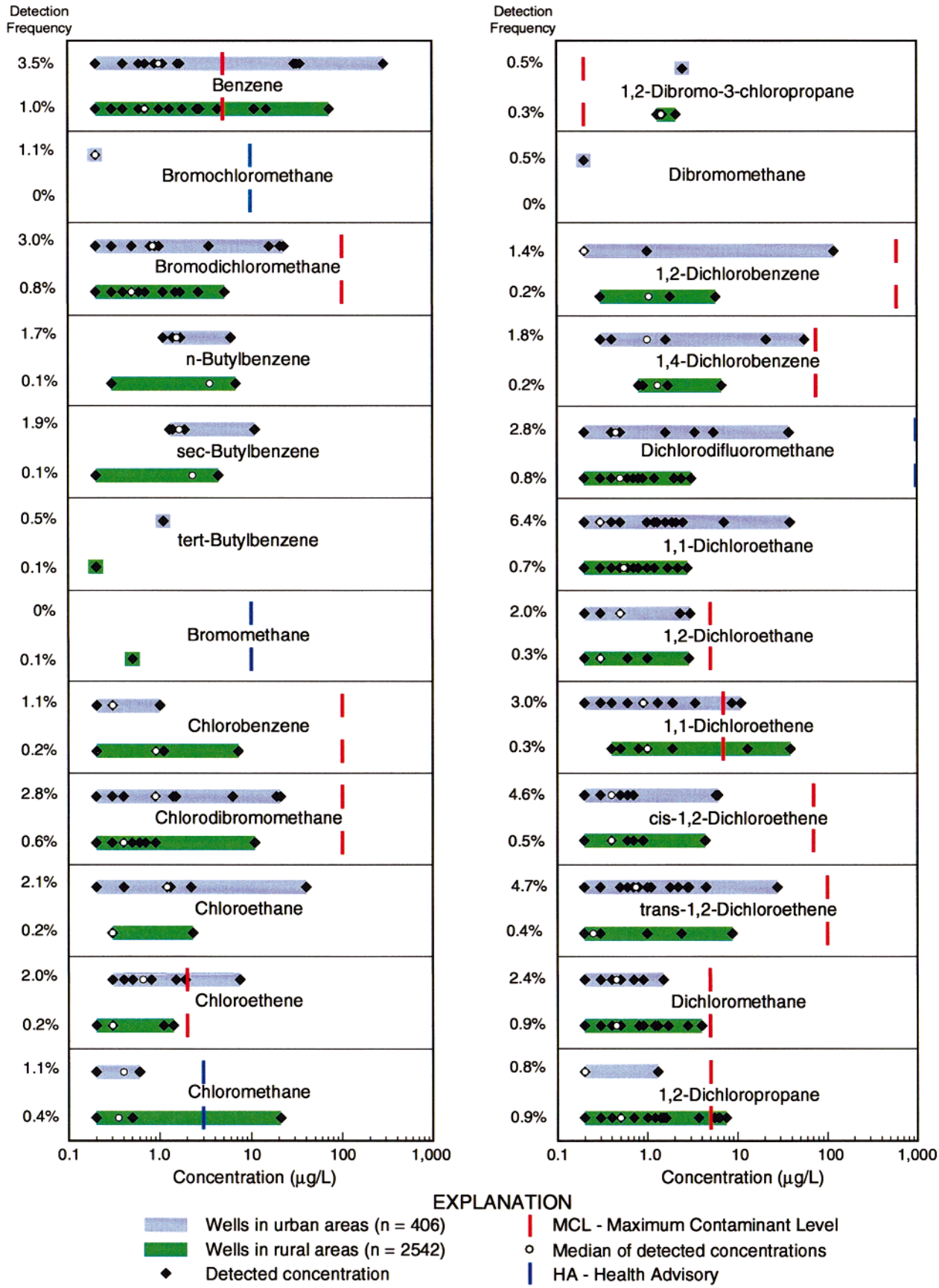
Although there are large differences in the overall detection frequencies of VOCs between urban and rural areas, some surprising similarities were noted. Many of the VOCs detected in urban areas also were detected in rural areas; there were only two VOCs unique to urban areas (bromochloromethane and dibromomethane), and four VOCs were unique to rural areas (bromomethane, 1,2-dibromoethane, 1,3-dichloropropane, and styrene). None of these six compounds was detected frequently; in fact, the detection frequencies were about 1% or less. Bromochloromethane is used in the synthesis of other organic compounds, and dibromomethane is used as a solvent (Table A2 in the Supporting Information). Bromomethane and 1,2-dibromoethane are soil fumigants; 1,3-dichloropropane and styrene are used in the synthesis of other organic compounds (Table A2 in the Supporting Information).

As can be seen in Figure 4, the types of VOCs detected were similar in urban and rural areas. To create this figure, each VOC was classified into one of four categories on the basis of its use (Table A2 in the Suporting Information): (i) solvents, organic synthesis (used in production of other organic compounds), and refrigerants; (ii) gasoline aromatic hydrocarbons; (3) oxygenates (the only VOC in this category was methyl *tert*-butyl ether, MTBE); and (4) fumigants. When a particular VOC did not clearly belong in a single category, a dominant category was inferred. VOCs used as solvents and in organic synthesis were at times difficult to categorize; therefore, these VOCs were combined with refrigerants into a single category in Figure 4. The most frequently detected VOCs were those used as solvents for organic synthesis and as refrigerants. Fumigants were a small fraction of VOCs detected in urban and rural areas. MTBE and gasoline aromatic hydrocarbons combined contributed about equally to the overall detection of VOCs in urban and rural areas.

Overall, the concentrations of VOCs in untreated ambient groundwater from urban areas were similar to those detected in rural areas. The Wilcoxon signed ranks test can be used to determine whether the median difference between paired observations equals zero, and the null hypothesis is that this median is equal to zero. For this comparison, the concentrations of a particular VOC measured in urban areas were paired with the concentrations measured in rural areas. A paired comparison of the maximum concentrations detected for each VOC showed that there was no significant difference in urban and rural areas ($p > 0.05$). A paired comparison of the median concentrations for each VOC showed that there was a significant difference ($p < 0.05$) with urban areas having larger concentrations. Although the median concentrations were statistically different, the magnitude of this difference was small; the median difference was only 0.28 μg/L. In addition, the results of a paired t-test showed that there was no difference ($p > 0.05$) in the median concentrations between urban and rural areas. For both the Wilcoxon test and the t-test, the median concentrations were based on measured concentrations, and the "true" median concentrations would fall below 0.2 μg/L because none of the VOCs had detection frequencies exceeding 50%.

MTBE and some of the VOCs used as solvents were among the most frequently detected VOCs in urban and rural areas. MTBE is added to gasoline as a fuel oxygenate or to increase the octane. In urban areas, there were six VOCs with a detection frequency 5% or greater (Table A2 in the Supporting Information). Among these six VOCs, five were solvent alkanes or alkenes with multiple chlorines. In rural areas, there were six VOCs with a frequency of detection of 1% or greater. Among these six VOCs, four were solvent alkanes or alkenes with multiple chlorines. The presence of chlorine in these solvents tends to make them less water-soluble, more volatile, and usually less reactive and more environmentally stable than nonchlorinated compounds (*24*).

Occurrence and Distribution of VOCs Based on Univariate Logistic Model. A national scale, univariate, logistic


Detection Frequency
3.5%
1.0%
Benzene
1.1%
0%
Bromochloromethane
3.0%
0.8%
Bromodichloromethane
1.7%
0.1%
n-Butylbenzene
1.9%
0.1%
sec-Butylbenzene
0.5%
0.1%
tert-Butylbenzene
0%
0.1%
Bromomethane
1.1%
0.2%
Chlorobenzene
2.8%
0.6%
Chlorodibromomethane
2.1%
0.2%
Chloroethane
2.0%
0.2%
Chloroethene
1.1%
0.4%
Chloromethane
0.1
1.0
10
100
1,000
Concentration (µg/L)
Detection Frequency
0.5%
0.3%
1,2-Dibromo-3-chloropropane
0.5%
0%
Dibromomethane
1.4%
0.2%
1,2-Dichlorobenzene
1.8%
0.2%
1,4-Dichlorobenzene
2.8%
0.8%
Dichlorodifluoromethane
6.4%
0.7%
1,1-Dichloroethane
2.0%
0.3%
1,2-Dichloroethane
3.0%
0.3%
1,1-Dichloroethene
4.6%
0.5%
cis-1,2-Dichloroethene
4.7%
0.4%
trans-1,2-Dichloroethene
2.4%
0.9%
Dichloromethane
0.8%
0.9%
1,2-Dichloropropane
0.1
1.0
10
100
1,000
Concentration (µg/L)
EXPLANATION
Wells in urban areas (n = 406)
Wells in rural areas (n = 2542)
Detected concentration
MCL - Maximum Contaminant Level
Median of detected concentrations
HA - Health Advisory

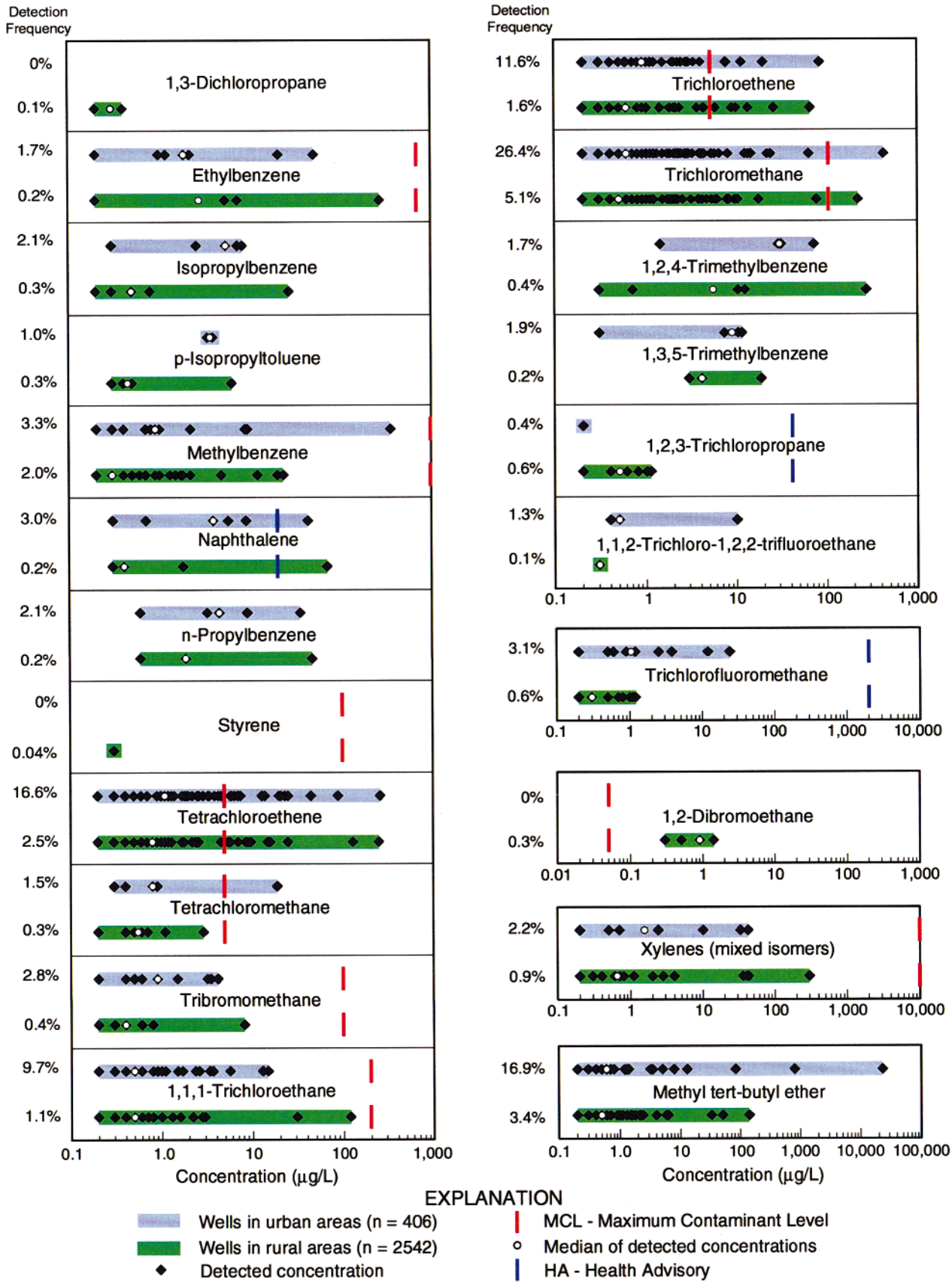


FIGURE 2. Detection frequency and concentrations of selected VOCs in sampled wells in urban and rural areas, 1985–1995.

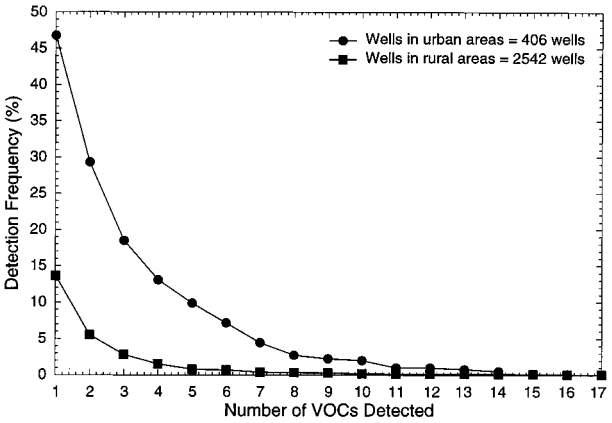


FIGURE 3. Detection frequency of one or more VOCs in wells located in urban and rural areas.

regression model was developed to estimate the presence or absence of VOCs in untreated groundwater based on population density. The concentrations of VOCs were not considered in this model. The probability of detecting VOCs in untreated groundwater (P) from this model can be calculated from population density (x) according to the following equation:

$$P = \frac{e^{[-3.1+0.40 \ln (x)]}}{1 + e^{[-3.1+0.40 \ln(x)]}} \quad (1)$$

where the exponent (e) is the logistic regression equation fitted by the final calibrated model. Probabilities were calculated using a 1-km^2-grid overlay of population density in the United States. Equation 1 is based on a reporting level of 0.2 μg/L; if a different reporting level were used, the estimated probabilities would differ. Figure 5 shows how increasing the hypothetical reporting level would affect the detection frequency of VOCs. From this figure, it can be calculated that 56% of the detected VOCs were at concentrations less than 1.0 μg/L; this was calculated by subtracting the frequency of detection at 0.2 μg/L minus the frequency of detection at 1.0 μg/L divided by the frequency of detection at 0.2 μg/L.

Figure A1 in the Supporting Information shows that, throughout the range of population density, the final model fit the (a) calibration data and (b) validation data. Probabilities begin to increase substantially for population densities greater than 20 people/km^2. Even when only those wells with population densities greater than 20 people/km^2 were evaluated (1436 wells) separately, aquifer hydraulic conductivity and groundwater recharge were not statistically significant explanatory variables. Soil permeability is significant explanatory variable ($p < 0.001$) among these wells, but the slope coefficient for this parameter is 0.09, which is very close to 0. As more data become available, other explanatory variables will be investigated. Nevertheless, population density among the 1436 wells was still a very good explanatory variable, and the model fit the data (Hosmer–Lemeshow test $p = 0.78$). Additional model analysis also showed that a 1000-m minimum well-spacing criterion provided the optimum well spacing while preserving the maximum number of wells in the data set; increasing the well-spacing criteria to 1500 m resulted in the deletion of some wells from the data set (Table 1). The parameter estimate of the ln(population density), standard error, and −2 log likelihood did not change drastically by using the three different well-spacing criteria (Table 1); furthermore, the best Hosmer–Lemeshow statistic ($p = 0.47$) was calculated by using 1000-m well spacing.

At a national scale, the model fit the calibration and validation data; however, model results may not always accurately reflect actual VOC presence or absence at a local scale. Table 1 shows that the coefficients for population density were significantly different than zero ($p < 0.001$), and the p value for the Hosmer–Lemeshow test indicates that the model fits both the calibration data ($p = 0.31$) and the validation data ($p = 0.43$). However, the model did not distinguish between aquifers but rather assumed a single

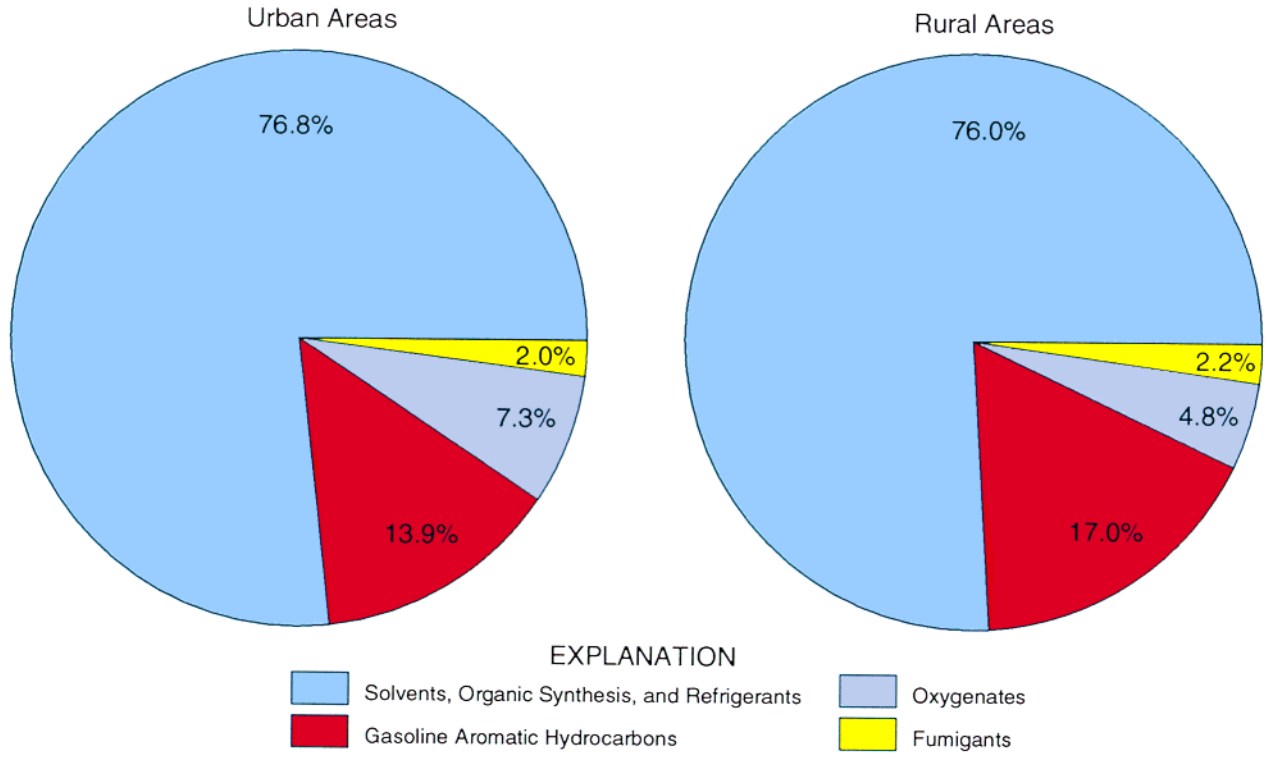


FIGURE 4. Percent detection of selected classes of VOCs in urban and rural areas.

TABLE 1. Diagnostics for Selected Logistic Regression Models

purpose of model	well-spacing criteria (m)	no. of wells	slope of population density (ln) estimate	standard error	−2 log likelihood value	*p*-value	intercept estimate	standard error	goodness-of-fit Hosmer–Lemeshow (*14*) *p*-value
(1) final calibrated model (excludes data for model validation)	1000	2354	0.40	0.027	262	<0.001	−3.05	0.132	0.31
(2) model validation	1000	589							0.43
(3) test well-spacing criteria	500	3145	0.40	0.023	361	<0.001	−3.00	0.113	0.18
(4) test well-spacing criteria	1000	2942	0.41	0.025	340	<0.001	−3.05	0.117	0.47
(5) test well-spacing criteria	1500	2768	0.43	0.026	322	<0.001	−3.10	0.122	0.24

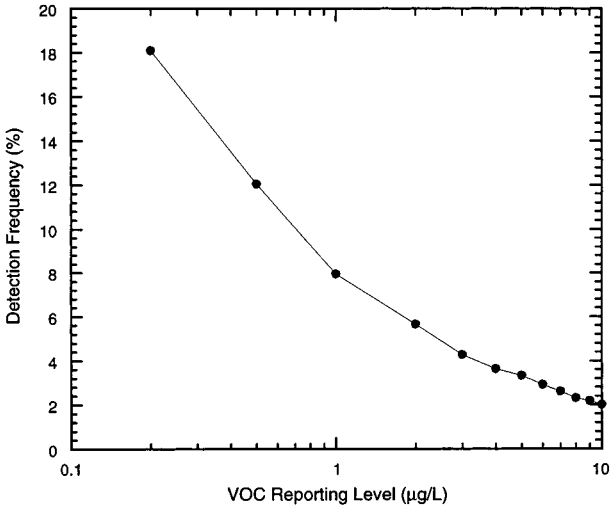


FIGURE 5. Comparison of VOCs detection frequency with different reporting levels for final data used in this assessment.

contiguous aquifer for the nation. Consequently, a sample from a well was assumed to be representative of all groundwater at a location even where multiple aquifers are present. Given these assumptions, the limitation and uses of this national scale model are discussed in a later section of this paper.

Estimates of Area Affected and Number of People Using Groundwater in These Areas. Defining the extent of VOC occurrence is basic to any assessment, but its presence in groundwater is most consequential in aquifers used for drinking water. However, the number of people using groundwater where VOCs may be present in the ambient groundwater resource may not necessarily reflect actual exposure to VOCs in the finished drinking water for at least three reasons. First, VOCs in finished drinking water may be affected by the well depth, well type, land-use practices around the well, pumping rates, etc. Second, the reporting levels of VOCs in compliance data of finished drinking water may be greater than 0.2 μg/L. Third, VOCs such as the trihalomethanes can be introduced into the finished drinking water as a disinfection byproduct. Currently, the USGS is working with other agencies to measure, summarize, and define VOCs in the source water and finished drinking-water supplies of the United States.

Estimates of the area affected and number of people using groundwater in these areas were calculated using two methods and scales of analysis; therefore, a comparison of the results provides corroboration. The first method of analysis used information compiled from 1-km^2 grids, including (i) population density, (ii) probabilities of detecting at least one VOC obtained through the logistic model, and (iii) groundwater-use information. This method provided the most detailed estimates that probably are more accurate than those derived from the second method. The second method of analysis used information compiled at a national scale, including (i) the average frequency of detection of VOCs in urban and rural areas, (ii) the size of urban and rural areas, and (iii) the percentage of people dependent on groundwater in urban and rural areas.

By use of the first method of analysis (information compiled from 1-km^2 grids), it was estimated that 7% of the area of the United States has ambient groundwater that contains at least one VOC at a concentration of $\geq$0.2 μg/L. By use of one standard error associated with logistic regression, the interval for the percentage of area ranges from about 6 to 9%. The area potentially with VOCs in untreated ambient groundwater is the product of the area of a grid cell (1 km^2) and the probability of detecting a VOC. The sum of these areas divided by the total area in the Nation yields an estimate of the percentage of the area with VOCs in untreated ambient groundwater.

On the basis of this first method of analysis, it was estimated that 42 million people may use groundwater in areas where the untreated ambient groundwater resource contains at least one VOC at a concentration $\geq$0.2 μg/L. This number is calculated as the product of the probability of detecting a VOC and the number of people using groundwater in an area; the number of people using groundwater from public or domestic water supplies is available at a county scale for 1995 (*25*), and these data were reduced to a 1-km^2 grid. By using one standard error associated with the logistic regression, the total number of people potentially using groundwater in these areas ranges from 35 to 50 million, with about 75% of the people living in urban areas; an estimated 32 million people in urban areas and 10 million people in rural areas use groundwater in areas where the ambient groundwater resource contains at least one VOC. These estimates are based on the assumption that the used groundwater resource lies beneath the area where people live; that is, groundwater is not piped in from some distance. Thus, these national scale estimates have definite limitations at the local scale.

The second method of analysis (information compiled at the national scale) indicated that 15% of the area of the contiguous United States has VOCs in the untreated groundwater resource, with 35 million people living in urban areas and 9 million in rural areas. Thus, the estimated area that has VOCs in untreated groundwater as determined by the second method of analysis was about twice as large as the method based on information compiled at a 1-km^2 grid; however, the number of people using groundwater in these areas was similar using either method. The second method of analysis was based on the following assumptions, information, and calculations. It was assumed that that the 406 wells in urban areas and 2542 wells in rural areas had sufficient spatial distribution to provide an assessment of VOC presence in these two areas on a national scale. The area affected was calculated by taking the product of the area and the frequency of detection of VOCs for urban and rural areas. VOCs were

TABLE 2. Frequency of Sampled Wells That Exceeded Drinking-Water Criteria and Taste or Odor Thresholds for VOCs

well location	well type	no. of wells	% of sampled wells that exceeded at least one drinking-water criterion[a]	% of sampled wells that exceeded taste or odor[b] or drinking-water criterion
urban areas	all wells	406	6.4	7.6
	drinking water wells (public and domestic)	80	2.5	2.5
rural areas	all wells	2542	1.5	1.9
	drinking water wells (public and domestic)	718	1.3	1.4
all areas	domestic wells	606	1.5	1.5
	public supply wells	192	2.1	3.1

[a] Drinking-water criteria are listed in Table A2 in the Supporting Information. [b] Only considered if the taste or odor threshold was less than the health criteria or if there were no health criteria. References and concentrations for taste or odor thresholds are listed in Table A2 in the Supporting Information.

detected in samples from 47% of the wells in urban areas and 14% of the wells in rural areas. On the basis of 1990 population density, there were 121 000 km^2 of urban land ($\geq$386 people/km^2) and 7 731 000 km^2 of rural land (<386 people/km^2) in the United States. The number of people potentially using groundwater where VOCs may be present in the groundwater resource was calculated by taking the product of the frequency of detection of VOCs, population, and the percentage of population dependent on groundwater. In 1990, the urban population was 187 million, and the rural population was 62 million (*26*). The urban population was 40% dependent on groundwater, whereas the rural population was nearly 100% dependent on groundwater (*27*).

Status of VOCs in Groundwater. Concentrations of VOCs among all sampled wells were compared with (i) drinking-water criteria established by the U.S. EPA (Figure 2) or (ii) the lowest published taste or odor thresholds if the thresholds were lower than the drinking-water criteria or if there were no drinking-water criteria (Table A2 in the Supporting Information). This comparison provides a measure of the overall quality of the untreated groundwater resource. There were three different drinking-water criteria considered (Table A2 in the Supporting Information), but only one was selected for each VOC. Preference was given for drinking-water criteria in the following order: (i) the maximum contaminant level (MCL), (ii) health advisory (HA) level, and (iii) the risk-specific dose (RSD5). RSD5 is associated with a risk of one additional person in 100 000 contracting cancer over a 70-year life span (*4*). In using this selection scheme, none of the VOC concentrations were compared to a RSD5 (Figure 2) because VOCs with a RSD5 also had either a MCL or a HA. VOC concentrations also were compared to taste or odor thresholds because groundwater can also be unusable because of aesthetic qualities; therefore, taste or odor thresholds were considered only if these thresholds were lower than the drinking-water criteria or if there were no drinking-water criteria for the VOC. Actual values for taste or odor thresholds will vary depending on the individuals tested, design of the test, and matrix water. There are odor thresholds lower than 20 μg/L for MTBE in the literature (*28*), but 20 μg/L (*19*) was used because the U.S. EPA has established a 20−40 μg/L taste or odor advisory for MTBE (*29*).

The concentrations of VOCs in sampled drinking-water wells also were compared with drinking-water criteria and taste or odor thresholds. Almost all samples in the data set have relevance to drinking water because this assessment was designed to investigate the groundwater resource that is used as a source of drinking water, has the potential to be used as a source of drinking water, or is hydraulically connected to a used aquifer. Nevertheless, the quality of water being consumed is of special interest to some; therefore, drinking-water wells were considered separately.

Among all wells, untreated groundwater in urban areas was four times more likely to exceed a drinking-water criterion than untreated groundwater in rural areas (Table 2). Taste or odor thresholds did not substantially increase the percentage of wells that would be unsuitable for consumption beyond those already not meeting health criterion. An exception may be observed among all wells in urban areas and public supply wells where an additional 1% of the wells exceeded taste or odor thresholds but met all drinking-water criteria (Table 2). Drinking-water criteria were exceeded in samples from 6.4% of all sampled wells in urban areas and 2.5% of the wells in urban areas used as a source of drinking water (Table 2). This compares with 1.5% of the wells in rural areas and 1.3% of wells in rural areas used as a source of drinking water. VOCs that were detected in more than 10% of the wells sampled in urban areas were also those that frequently exceeded drinking-water criteria.

Limitations and Use of Assessment and Model

Studies designed to investigate the presence of VOCs in groundwater associated with a specific land use in a particular area of the United States may have different results than this resource assessment. Data for this assessment were compiled from numerous studies with various objectives and scopes; therefore, the wells were located in a variety of urban and rural land-use areas. Studies with a more narrow scope may have different results because of the proximity to potential sources of contamination or differences in the hydrogeologic conditions.

This statistical summary of the monitoring data is a valid assessment of VOCs in untreated ambient groundwater in the conterminous United States only to the degree that the data represent an unbiased random sampling of groundwater in the conterminous United States. An assessment of the presence of VOCs requires data from representative, or preferably, all parts of the United States. Although there are 2948 wells in the data set, these wells were not spatially distributed throughout the entire United States; furthermore, the wells were not equally distributed even within areas where data exist (Figure 1).

Figure 6 shows the probability of detecting a VOC in groundwater based on the logistic regression model; however, probabilities are only shown where more than 5 people/km^2 use groundwater as a water supply. Areas with 5 or less people/km^2 using groundwater generally are rural areas, but there are urban areas where groundwater is not used as a water supply. The areas where probabilities are shown are believed to be most relevant because they encompass (i) 82% of all people in the United States, (ii) 95% of all people dependent on groundwater, and (iii) 99% of people that use groundwater in areas that the ambient groundwater resource probably contains a VOC. This national assessment may not always accurately portray the presence of VOCs on the local scale. Nevertheless, it does provide a national assessment of where VOCs are likely to be found and where they may be problematic based on water-use information. For example,

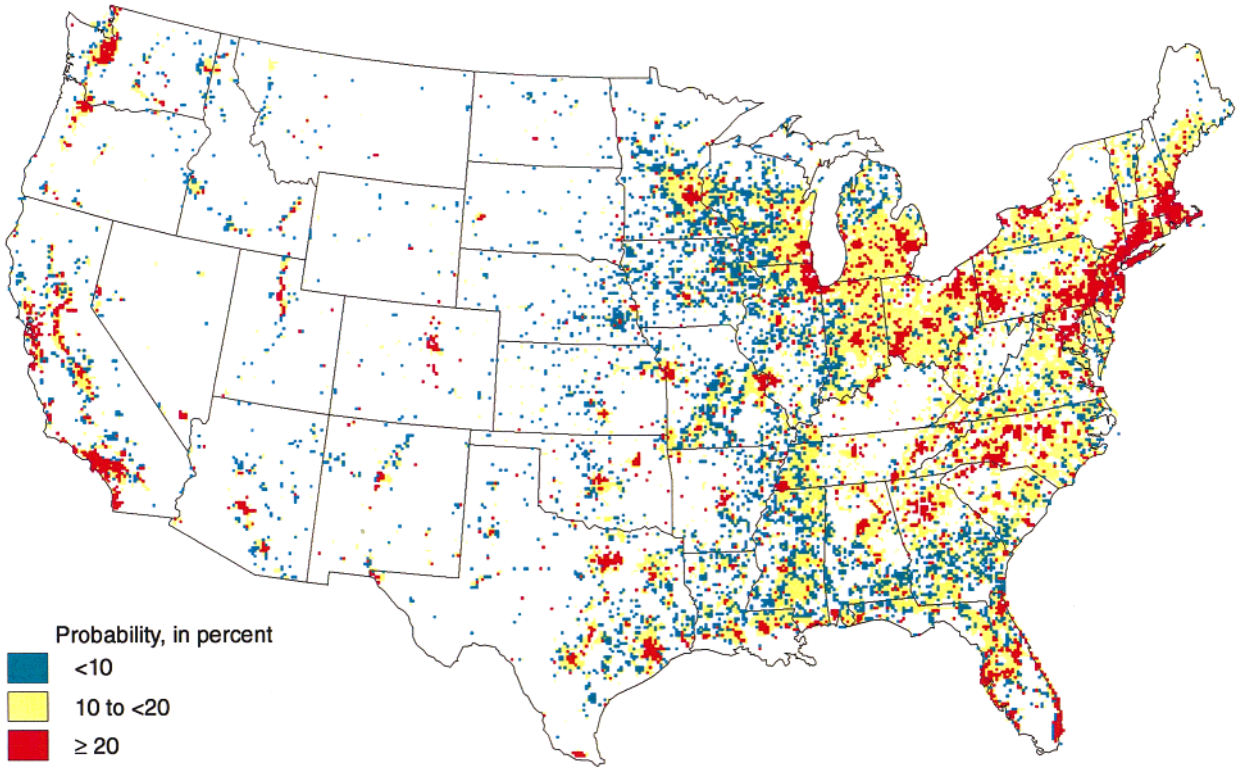


FIGURE 6. Probability of detecting a VOC in ambient untreated groundwater in areas where more than 5 people/km² use groundwater. Probabilities were calculated using eq 1.

people in the eastern half of the United States are more likely to use groundwater in areas where the ambient groundwater resource contains a VOC than those in the western half of the United States. People living in the populated areas of the Northeast and West Coast also are more likely to use groundwater where the ambient groundwater resource contains a VOC.

Although this assessment contained a limited number of known drinking-water wells, it has considerable relevance to drinking water. To establish the relevance for urban areas, data from 406 wells in urban areas were compared to the U.S. EPA finished drinking-water assessment of 186 randomly selected water supplies serving more than 10 000 people (*5*). Although the population densities surrounding the wells sampled by the U.S. EPA are unknown, they probably are urban because the wells served more than 10 000 people. The well types for the 406 wells in urban areas were 15% public, 5% domestic, 7% monitoring, 5% other, and 68% unknown. Trihalomethanes were excluded from this comparison because they can be formed during chlorination. Overall, 22 VOCs were detected in at least one assessment. The detection frequencies of these VOCs are compared in Figure 7 by using the higher reporting level in the U.S. EPA data (*5*). An analysis of the paired detection frequencies for the 22 VOCs using the Wilcoxon signed-ranks test indicated that detection frequencies were not significantly higher in one data set versus the other. However, as indicated in Figure 7, the chemicals most frequently detected in the urban wells evaluated in the study (those indicated by name in Figure 7) do tend to be detected at slightly lower rates in the U.S. EPA finished drinking-water data (except for trichloroethene). The similarity of the U.S. EPA finished drinking-water survey and the data collected from urban wells in this assessment indicate that data in this resource assessment have drinking-water relevance even if the wells were not all designated as drinking-water wells.

This assessment provides information on detection frequencies of specific VOCs and the ranges of concentrations

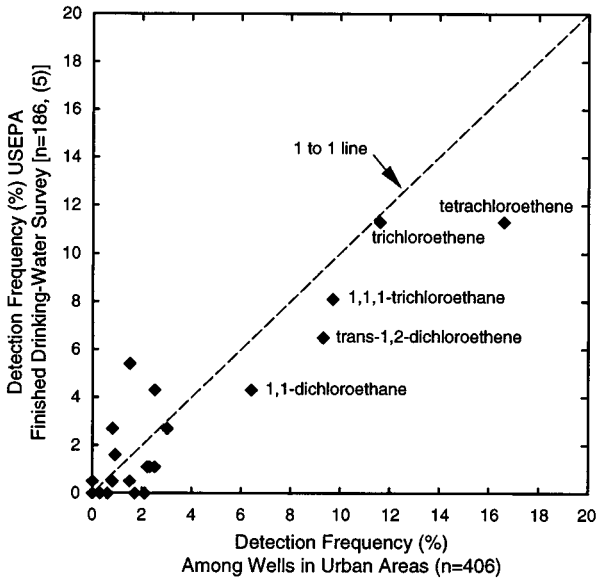


FIGURE 7. Comparison of the detection frequencies of VOCs common to U.S. EPA's finished drinking-water survey of large community water supplies using groundwater (*5*) and samples from wells in urban areas used in this assessment.

detected, and both of these factors need to be considered when evaluating potential effects of a VOC on the groundwater resource. Concentrations that exceed drinking-water criteria are of immediate concern. Large concentrations associated with point-source contamination were infrequent in this assessment, but this is not surprising because known point-source contamination sites were specifically excluded. Concentrations of VOCs less than drinking-water criteria are also important. These data might be critical, for example, in the event that drinking-water criteria would be lowered. This possibility is particularly important for VOCs that have criteria

based on limited information. Furthermore, low concentrations of a VOC can mean two things about the source of contamination: (i) the nonpoint source of contamination produces only a small concentration, or (ii) the sampling point is some distance from point sources and source concentrations are much larger than what was detected.

Implications of Findings for Protecting and Managing the Groundwater Resource

Most of the detected VOC concentrations were less than current drinking-water criteria; however, this should not be seen as a definitive appraisal of the health risks for several reasons. Drinking-water criteria are based on current toxicity information, and as new information becomes available, these criteria may be revised downward. Drinking-water criteria also are based on the effects of individual compounds. The criteria do not, for example, consider the co-occurrence of VOCs or the co-occurrence of VOCs with other potential contaminants, such as pesticides or nitrate; however, co-occurring VOCs are common. This assessment indicates that samples from 29% of the wells in urban areas had two or more VOCs, and in rural areas, 6% of the wells had two or more VOCs. Finally, there were 14 VOCs that were detected that presently lack drinking-water criteria.

Because health criteria may change, prevention of VOC contamination that is based on a solid understanding of the sources and transport of these compounds may be preferable to remediation or treatment, which can be difficult and expensive. The frequent detection of VOCs in some aquifers is a clear indication that these aquifers are susceptible to anthropogenic sources of contamination. If an aquifer is currently used or is a likely future source of drinking water, then initiation of activities to proactively protect the water quality of the aquifer would seem prudent. This protection is also desirable when one considers the uncertainty of appropriate drinking-water criteria and potential health effects of co-occurring contaminants. Monitoring groundwater quality on a regular basis at low reporting levels can provide an indication of potential contaminant concentrations before it becomes a serious problem.

In urban areas where groundwater is used as a water supply, there is a large probability of detecting VOCs in the ambient untreated groundwater. Only 40% of the urban population are dependent on groundwater as a water supply; however, most people using groundwater in areas of the United States where the ambient groundwater contains a VOC live in large urban areas. In fact, 70% of these people live in metropolitan areas with populations greater than 250 000. Protection of the groundwater resource underlying large urban areas that use groundwater as a water supply may be difficult but could substantially reduce the number of people using an ambient groundwater resource with a VOC. Research is needed on the sources, transport, and fate of VOCs in large metropolitan areas. Once the sources have been identified, they can be dealt with accordingly. However, the hydrology in many urban areas is poorly understood. For example, very little is known about groundwater recharge through the extensively disturbed unsaturated zone in urban areas. The effect of roads, sewers, managed lawns, and commercial areas on the groundwater is also poorly documented. The need for research will continue to increase in the future as the urban population continues to grow. Finally, the similarities between the types and concentrations of VOCs detected in urban and rural areas would seem to indicate that the results of urban investigations would have some transfer value to rural areas.

Large-scale groundwater monitoring networks, which are designed to assess the presence of VOCs, need to consider stratifying sample locations that are based on population density. Simple random selection is the basis for most sampling schemes in use today in which wells are selected at random within a study area. When simple random selection is used, no prior knowledge of the groundwater system is used and every area has an equal and known chance of being selected. However, stratified random selection makes use of prior information by dividing the target study area into smaller areas that are likely to be more homogeneous than the area as a whole (*30*). Population density has been shown to be a significant explanatory variable for the presence of VOCs. Therefore, this variable in conjunction with other variables could be used on a national, regional, or state level to delineate areas where VOC occurrence is likely to be similar. Other stratification layers could include variables, such as type of aquifer, depth to water, well depth, depth to well screen, and recharge rate. Once these layers are identified, wells are located randomly within these smaller areas. Stratified random selection can produce large gains in the precision of the sample estimates over simple random selection (*31*).

Supporting Information Available

One table summarizes the location of sampled wells in relation to productive aquifers and land-use setting; another table lists by land-use setting the primary use of each VOC, detection frequency and concentrations, drinking-water criteria, and taste or odor thresholds; one figure shows a comparison of model probability estimates of detecting VOCs with calibration and validation data (6 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Ashford, N. A.; Miller, C. S. *Chemical Exposures—Low Levels and High Stakes*; Van Nostrand Reinhold: New York, 1991.

(2) U.S. Environmental Protection Agency. *1996 Toxics Release Inventory Public Data Release*; Office of Pollution Prevention and Toxics: Washington, DC, 1998; EPA 745-R-98-005.

(3) Cohen, Y. In *Volatile Organic Compounds in the Environment*; Wang, W., Schnoor, J. L., Doi, J., Eds.; ASTM STP 1261; American Society for Testing and Materials: West Conshohocken, PA, 1996; pp 7−32.

(4) U.S. Environmental Protection Agency. *Drinking Water Regulations and Health Advisories*; Office of Water: Washington, DC, 1996.

(5) Westrick, J. J.; Mello, W.; Thomas, R. F. *J. Am. Water Works Assoc.* **1984**, *76*, 52−59.

(6) Westrick, J. J. In *Significance and Treatment of Volatile Organic Compounds in Water Supplies*; Ram, N. M., Christman, R. F., Cantor, K. P., Eds.; Lewis Publishers: Chelsea, MI, 1990; pp 103−125.

(7) U.S. Environmental Protection Agency. *Straight Talk On Tanks—Leak Detection Methods for Petroleum Underground Storage Tanks and Piping*; Office of Underground Storage Tanks: Washington, DC, 1997; EPA 510-B-97-007.

(8) Lapham, W. W.; Tadayon, S. *Plan For Assessment of the Occurrence, Status, and Distribution of Volatile Organic Compounds in Aquifers of the United States*; USGS Open-File Report 96-199; U.S. Geological Survey: Reston, VA, 1996.

(9) Lapham, W. W.; Neitzert, K. M.; Moran, M. J.; Zogorski, J. S. *Ground Water Monit. Rem.* **1997**, *Fall*, 147−157.

(10) Lapham, W. W.; Moran, M. J.; Zogorski, J. S. In *National Monitoring Conference, July 7−9, 1998, Reno, NV [Proceedings]*; National Water-Quality Monitoring Council: 1998; pp III371−III381.

(11) Rose, D. L.; Schroeder, M. P. *Method of Analysis by the U.S. Geological Survey National Water Quality Laboratory—Determination of Volatile Organic Compounds in Water by Purge and Trap Capillary Gas Chromatography/Mass Spectrometry*; USGS Open-File Report 94-708; U.S. Geological Survey: Reston, VA, 1995.

(12) Hitt, K. J. *Refining 1970s Land-Use Data with 1990 Population Data to Indicate New Residential Development*; Water-Resources Investigative Report 94-4250; U.S. Geological Survey: Reston, VA, 1994.

(13) U.S. Geological Survey. *National Atlas of the United States of America*; U.S. Geological Survey: Reston, VA, 1970.
(14) Hosmer, D. W.; Lemeshow, S. *Applied Logistic Regression*; John Wiley and Sons: New York, 1989.
(15) Eckhardt, D. A.; Stackelberg, P. E. *Ground Water* **1995**, *33*, 1019−1033.
(16) Nolan, B. T.; Ruddy, B. C.; Hitt, K. J.; Helsel, D. R. *Environ. Sci. Technol.* **1997**, *31*, 2229−2236.
(17) Bureau of the Census. *Statistical Abstracts of the United States—1994*; U.S. Government Printing Office: Washington, DC, 1995; CD-SA-94.
(18) Amoore, J. E.; Hautala E. *J. Appl. Toxicol.* **1983**, *3*, 272−290.
(19) Young, W. F.; Horth, H.; Crane, R.; Ogden, T.; Arnott, M. *Water Res.* **1996**, *30*, 331−340.
(20) Verschueren, K. *Handbook of Environmental Data on Organic Chemicals*; Van Nostrand Reinhold: New York, 1983.
(21) Prager, J. C. *Environmental Contaminant Reference Databook Volume 1*; Van Nostrand Reinhold: New York, 1997.
(22) Verschueren, K. *Handbook of Environmental Data on Organic Chemicals*; Van Nostrand Reinhold: New York, 1997.
(23) American Water Works Association Research Foundation & Lyonnaise des Eaux. *Identification and Treatment of Tastes and Odors in Drinking Water*; Mallevialle, J., Suffet, I. H., Eds.; American Water Works Association: Denver, CO, 1987.
(24) Capel, P. D. In *Regional Ground Water Quality*; Alley, W. M., Ed.; Van Nostrand Reinhold: New York, 1993; pp 155−179.
(25) Solley, W. B.; Pierce, R. R.; Perlman, H. A. *Estimated Use of Water in the United States in 1995*; U.S. Geological Survey Circular 1200; U.S. Geological Survey: Denver, CO, 1998.
(26) U.S. Bureau of the Census. *Statistical Abstracts of the United States—1992*, 112th ed.; U.S. Government Printing Office: Washington, DC, 1992.
(27) Carr, J. E.; Chase, E. B.; Paulson, R. W.; Moody, D. W. *National Water Summary 1987—Hydrologic Events and Water Supply and Use*; USGS Water-Supply Paper 2350; U.S. Geological Survey: Reston, VA, 1990.
(28) Shen, Y. F.; Yoo L. J.; Fitzsimmons, S. R.; Yamamoto, M. K. In *American Chemical Society Division of Environmental Chemistry Preprints of Papers*, 213th Meeting, San Francisco, CA; American Chemical Society: Washington, DC, 1997; Vol. 37, pp 407−409.
(29) U.S. Environmental Protection Agency. *Drinking water advisory—Consumer acceptability advice and health effects analysis on methyl tertiary-butyl ether (MTBE)*; EPA-822-F-97-009; Office of Water: Washington, DC, 1997.
(30) Alley, W. M. In *Regional Ground Water Quality*; Alley, W. M., Ed.; Van Nostrand Reinhold: New York, 1993; pp 63−85.
(31) Stuart, A. *Basic Ideas of Scientific Sampling*, 2nd ed.; Griffin's Statistical Monographs and Courses 4; Hafner Press: New York, 1976.

Received for review March 1, 1999. Revised manuscript received September 9, 1999. Accepted September 13, 1999.

ES990234M