METHODS TO DETERMINE THE SUSCEPTIBILITY OF SOURCE WATER TO COMMUNITY AND NONCOMMUNITY WATER SUPPLIES IN NEW JERSEY TO CONTAMINATION

Introduction

The 1996 Amendments to the Federal Safe Drinking Water Act require all states to establish a Source Water Assessment Program (SWAP). New Jersey Department of Environmental Protection (NJDEP) elected to evaluate the susceptibility of public water systems to contamination by inorganic constituents, nutrients, volatile organic and synthetic organic compounds, pesticides, disinfection byproduct precursors, pathogens, and radionuclides. Susceptibility to contamination in ground and surface water is a function of many factors, including constituent presence or use in or near the water source, natural occurrence of constituents in geologic material, changes in ambient conditions related to human activities, and location of the source within the flow system. The New Jersey SWAP includes four steps: (1) delineate the source water assessment area of each ground- and surface-water source of public drinking water, (2) inventory the potential contaminant sources within the source water assessment area, (3) determine the public water system's susceptibility to contamination, and (4) incorporate public participation and education (http://www.state.nj.us/dep/swap).

Susceptibility assessment models were developed by U.S. Geological Survey (USGS) to predict the susceptibility of the source water to community and noncommunity water-supply wells and surface-water intakes in New Jersey to contaminants. Susceptibility is defined by variables that describe hydrogeologic sensitivity and potential contaminant-use intensity within the area that contributes water to a ground-water well or surface-water intake. Hydrogeologic conditions, land use and land cover, human activities, and population density are quite variable in New Jersey; thus, some water supplies are more susceptible to contamination than others. Susceptibility assessment models were developed to rate each source of public drinking water as having low, medium, or high susceptibility for groups of constituents (table 1). This report describes methods used to develop these susceptibility assessment models. The database and susceptibility assessment models developed by the USGS will be used by NJDEP as a guide to determine the monitoring needs for public drinking water sources to ensure protection of the public health.

Table 1. Susceptibility assessment contaminant categories		
Contaminant categories	**Ground water**	**Surface water**
Disinfection byproducts precursors	Yes	Yes
Inorganic constituents	Yes	Yes
Nutrients	Yes	Yes
Pathogens	Yes	No[1]
Pesticides	Yes	Yes
Radionuclides	Yes	No[2]
Synthetic organic compounds	No[3]	No[3]
Volatile organic compounds	Yes	Yes

[1] All surface-water sources are considered to be susceptible to contamination by pathogens.
[2] All surface-water sources are not considered to be susceptible to contamination by radionuclides.
[3] Data were insufficient to develop susceptibility rating models.

Contaminant Categories

NJDEP lists eight categories of constituents as contaminants of concern for the SWAP (http://www.state.nj.us/dep/watersupply/swap1.pdf, pg. 29). Constituents within these categories could be from point or nonpoint sources; they pose a health risk because of their possible toxic or carcinogenic effects.

Pathogens

Pathogens include bacteria, protozoa, and viruses. Pathogens from animal and human waste present a risk to human health. Point sources of pathogens include combined sewer overflows, individual septic systems, effluent from sewage-treatment plants, and landfill leachate. Nonpoint sources include runoff from livestock facilities and discharge from multiple septic systems.

Nutrients

The nutrient contaminant category focused on nitrate. Nitrate can occur naturally in the environment or can originate from human sources. Nitrate can adversely affect environmental quality, human health, and the efficiency of drinking-water-treatment plants. An important point source of nitrate is effluent from sewage treatment plants. Nonpoint sources of nitrate include discharge from multiple septic systems, facilities where animal waste is stored, and runoff from agricultural and residential land where fertilizers are applied.

Volatile organic compounds (VOCs)

Volatile organic compounds are the most common organic contaminants in ground water in New Jersey (http://www.state.nj.us/dep/watersupply/swap1.pdf). VOCs can cause a variety of harmful effects to human health. VOCs include chemicals that are used as solvents, degreasers, refrigerants, and gasoline components. VOCs are present in household products, such as air fresheners and cleaning products. They are used extensively for industrial purposes, such as in the manufacturing of paints, plastics, and toiletries. Examples of VOCs are methyl tertiary-butyl ether (MTBE), trichloroethylene (TCE), benzene, vinyl chloride, toluene, and chloroform.

Pesticides

Pesticides are chemical substances and biological agents used to control weeds, insects, fungi, rodents, bacteria, and other pests. Exposure to some pesticides can cause harmful effects to humans. Common sources of pesticides include land applications in both agricultural and nonagricultural settings (nonpoint source) and manufacturing/distribution centers of pesticides (point source).

Synthetic organic compounds (SOCs)

Synthetic organic compounds, other than VOCs and organic pesticides, are grouped in a separate SOC category. Common sources include chemical manufacturing plants, pharmaceutical plants, sewage-treatment plants, and discharges from contaminated sites. Examples of SOCs are polychlorinated biphenyls (PCBs), anthracene, and phthalates.

Inorganic Constituents

Inorganic constituents in the environment can be naturally occurring or the result of human activities. Sources include discharges from manufacturing plants, release from contaminated sites, past land uses, and geologic material. Examples of inorganic constituents are arsenic, cadmium, copper, lead, mercury, and asbestos.

Radionuclides

Radionuclides are unstable radioactive isotopes such as radium, radon, and uranium. They are a category of contaminant that occurs naturally or can be related to human activities. Sources of radionuclides include the decay of naturally occurring minerals, leaching of subsurface material (for example rocks and sedimentary materials) into ground water, and improper disposal of radioactive waste.

Disinfection byproduct precursors (DBPs)

Disinfection byproducts are formed when the disinfectants used to kill pathogens during the water-treatment process react with organic and inorganic compounds present in the water. Natural organic matter is an important component of the organic material that reacts to form DBPs (Stevens and others, 1976). Some compounds that occur as DBPs in water supplies can occur naturally or as a result of human activities, for example, where septic-system effluent containing both chlorine and organic compounds drains to surface-water bodies or percolates to ground water. The major chemical species formed by chlorination are trihalomethanes, haloacetic acids, haloacetonitriles, and chloral hydrate. The concentration of disinfection byproducts produced is dependent on the amount of organic matter, the type of disinfectant, the concentration of disinfectant, time of contact, pH, and water temperature. Chlorine is the most common disinfectant used in New Jersey (http://www.state.nj.us/dep/watersupply/swap1.pdf).

Susceptibility Model Development

The development of the susceptibility assessment models involved several steps: (1) development of source water assessment areas to community and noncommunity water supplies; (2) building of geographic information system (GIS) and water-quality data sets; (3) exploratory data analysis using univariate and multivariate statistical techniques, and graphical procedures; (4) development of a numerical coding scheme for each variable used in the models; (5) assessment of relations of the constituents to model variables; and (6) use of an independent data set, if available, to verify the models. Multiple lines of evidence were used to select the final variables used in the models.

Development of Source Water Assessment Areas

The New Jersey Geological Survey (NJGS) estimated areas contributing water to more than 2,400 community water-supply wells in New Jersey and New York (Figure of well delineation http://www.state.nj.us/dep/swap/resources.htm) by using the Combined Model/Calculated Fixed Radius Method. These methods use well depth, water-table gradient, water-use data, well characteristics, and aquifer properties to determine the size and shape of the contributing area. The source water assessment area for a well open to an unconfined aquifer was divided into three tiers based on the time of travel from the outside edge to the wellhead: tier 1 (2-year time of travel), tier 2 (5-year time of travel), and tier 3 (12-year time of travel) (http://www.state.nj.us/dep/njgs/whpaguide.pdf). An unconfined aquifer is a permeable water-bearing unit where the water table forms its upper boundary at the interface between unsaturated and saturated zones. The source water assessment area for a well open to a confined aquifer was defined as the area within a 50-foot radius of the well (http://www.state.nj.us/dep/njgs/whpaguide.pdf). Confined aquifers are permeable water-bearing units between hydrogeologic units with low permeability, known as confining units.

NJDEP estimated 60 areas contributing water to surface-water sources used for drinking water in New Jersey (Figure of intake delineation http://www.state.nj.us/dep/swap/resources.htm); 49 are associated with surface-water intakes and 11 are associated with sources using ground water under the direct influence of surface water (GWUDI). For most surface-water sources, the source water assessment area includes the entire drainage area that contributes to the water that flows past the intake or source (NJDEP Source Water Assessment Reports, 2004). These source water assessment areas include the headwaters and tributaries of a stream and are based on the USGS 14-digit hydrologic unit code (HUC 14) (Ellis and Price, 1995). For intakes or sources with extremely large contributing areas, the source water assessment area is based on the time of travel to the intake or source (NJDEP Source Water Assessment Reports, 2004).

The USGS estimated areas contributing water to 388 surface-water-quality sites in New Jersey for model development and verification. Drainage areas contributing water to a surface-water-quality site were delineated using a GIS macro language program that determines basin area from a digital elevation model (DEM) based on a 1:24,000 scale and 30-meter resolution to contour intervals (L.J. Kauffman, U.S. Geological Survey, written commun., 2002). The flow direction for cells within the DEM grid were determined by the adjacent cell with the greatest slope. A filled DEM was used to indicate the flow direction of each grid cell on the DEM. A coverage, which included locations of sites for which contributing areas were needed, was then used to mark the downstream boundary of each contributing area. The watershed command then created the contributing area for each point in the coverage using the flow directions of each cell.

Development of Data Sets

Data sets were developed for the GIS and water-quality data to assess the variables used to develop the susceptibility models. A relational database was used to store and manipulate water-quality, hydrogeologic-sensitivity, and intensity variables.

GIS

A GIS was used to quantify hydrogeologic-sensitivity and potential contaminant-use intensity variables that could affect ground-water quality within areas contributing water to ground- and surface-water sources, and surface-water-quality sites. The data for the sensitivity and intensity variables are stored as statewide GIS layers where each layer represents a feature from which site-specific variables were calculated. A program was used to cut out a piece of the statewide layer using the boundary of a source water assessment area as a "cookie cutter." A new layer was created that contained a geographic subset of the features from the larger layer. For example, from the statewide land-use layer, a new layer was created that contained only the land use within the source water assessment area for that site. The areas in the new layer were summed by type of land use and results stored in a relational database. Values for each layer were compiled for three ground-water tiers for ground-water sites, and for the entire source water assessment area for both ground- and surface-water-quality sites. Sensitivity variables used in the statistical analysis for ground-water susceptibility models include soil properties, aquifer properties, well-construction characteristics, physiographic province, predominant watershed, and hydrologic unit. Intensity variables for ground-water susceptibility models include land use from coverages based in the early 1970's, 1986, and 1995-97; lengths of roads, railways, and streams; the number of potential contaminant sources; septic-tank, population, and contaminant-site densities; and minimum distances of the well or surface-water source to the various land uses and to potential contaminant sources (Table of variables http://www.state.nj.us/dep/swap/resources.htm). Sensitivity and intensity variables for surface-water-quality sites were generated for the same variables as were developed for ground-water sites, except that 1970's and 1986 land use variables were not developed.

Water-Quality Data

All ground- and surface-water-quality data collected and analyzed by the USGS from June 1980 to October 2002 were retrieved from the National Water Information System (NWIS) (http://water.uggs.gov/pubs/FS/FS-027-98) and imported into a relational database. Data for 3,752 ground-water and 801 surface-water-quality sites were retrieved. Analyses that were determined by older, less accurate, less precise methods, and those with high reporting levels were not used. Sites were eliminated if results of analyses did not include any of the constituents of concern. Sites in northern New Jersey with contributing areas that are predominantly in New York State were eliminated because sensitivity and intensity variables were unavailable. This process yielded a spatially distributed subset of sites with water-quality data for both ground and surface water.

For community water-supply wells, three data sets consisting of wells sampled for each constituent were used in the modeling process to test relations between constituents and sensitivity and intensity variables: (1) all wells in the NWIS database, (2) all community water-supply (CWS) wells, and (3) a subset consisting of 240 CWS wells open to unconfined aquifers. The most recent concentration measured in water from each well was used in each data set because the most recent sample represented current water-quality conditions and probably was analyzed using a method with the lowest minimum reporting level (MRL) and with better precision. If there were sufficient data to run meaningful statistical tests, the 240-well subset was used to develop the model, and the CWS subset was used to verify the model. If not, the data set with all CWS wells was used for model development. Typically, statistical analyses were not run on the data set with all wells from the NWIS database. Many of the samples are from problem-oriented studies, and the results do not necessarily represent typical ground-water conditions for community water-supply wells. This data set was used to determine spatial distribution of constituents within New Jersey. Source water assessment areas were not generated for all wells, and consequently, values for sensitivity and intensity variables were not determined.

A statewide network of 388 USGS surface-water-quality sites was selected for use in the development the models for the surface-water intakes. Many of these sites are part of the USGS New Jersey District systematic data-collection program. Some are sites from the National Water Quality Assessment program (NAWQA), and others are from regional and local investigations. The maximum concentration was selected to represent surface-water-quality sites. Subsets of the set of 388 surface-water-quality sites that had water-quality data for a particular constituent were used in the modeling process to test relations between that constituent and sensitivity and intensity variables.

The pH and dissolved oxygen concentration of water samples were used as variables in statistical tests and in the application of models to CWS wells and surface-water sources. The results of the most recent analyses stored in the NWIS database were used to represent the pH and dissolved oxygen concentration of the wells. If data were unavailable in the NWIS database for pH, the most recent value from the NJDEP water-quality database was used. These analyses are unlike analyses in the NWIS database in that they are often obtained from samples that were collected from facilities that receive water from more than one well, and the water may have been treated. This most recent value was used for all wells that contribute to that facility. NJDEP does not require water suppliers to provide dissolved oxygen concentration data; consequently, only data from the NWIS database were used. If results of analyses were unavailable in either database, no value of pH or dissolved oxygen concentration was used.

Data Analysis

Statistical tests and graphical procedures were used to evaluate the relation between constituents and sensitivity and intensity variables to determine those variables that best describe the concentrations of constituents in source waters (table 2). These data-analysis tools were used as multiple lines of evidence when evaluating which variables to include in each susceptibility assessment model.

Table 2. Statistical methods used for susceptibility model development		
Statistical Procedure	Symbol	Application of procedure in model development
Kruskal-Wallis	K	Initial exploration of univariate relations
Spearman's rho	P	Univariate non-parametric testing of significance of explanatory variables used for screening insignificant variables from further consideration
Logistic regression	-	Multivariate regression for determining relative significance of, and interaction among, explanatory variables in models that predict the probability of exceeding specific concentration values
Principal component analysis	-	Multivariate graphical method used for observing trends in direction and magnitude

Univariate statistical tests were run on all variables and used for exploratory data analysis. Univariate tests included the Kruskal-Wallis test and Spearman's rho rank correlation. The Kruskal-Wallis test was used to compare the distributions of variables among two or more groups and to determine whether distributions differed among groups. The size of the Kruskal-Wallis test statistic and corresponding p-value were used as a measure of the strength of differences between the groups; the larger the test statistic and the smaller the p-value relative to the other values within the data set, the more significant the test result (http://water.usgs.gov/pubs/twri/twri4a3). The magnitude of the test statistic depends on the size of the data set; the larger the data set, the larger the test statistic relative to the test statistic from a smaller data set. The test statistic is not influenced by outlying values because the Kruskal-Wallis test is nonparametric. Spearman's rho correlation coefficients were determined to evaluate the magnitude of positive or negative relations between the explanatory and response variables. The Spearman's rho coefficient minimizes the effects of outlying values (http://water.usgs.gov/pubs/twri/twri4a3). Spearman's rho, the nonparametric equivalent of a correlation coefficient, was used because environmental variables rarely are normally distributed (Helsel and Hirsch, 2002). Correlation coefficients were calculated for paired combinations of constituent concentration values in relation to sensitivity, intensity, and many water-quality variables.

Multivariate statistical tests were used to determine those variables that collectively are the best predictors of contamination of water from community water-supply wells and surface-water intakes. Multivariate statistical tests were conducted on selected statistically significant and conceptual variables to narrow the list of variables to be used in the susceptibility assessment model and to determine those variables that collectively are best predictors of potential contamination of water from water-supply wells (http://water.usgs.gov/pubs/twri/twri4a3). Multivariate tests included logistic regression and principal components analysis. Variables used in the susceptibility assessment model were selected on the basis of results of summary statistics, univariate and multivariate statistical tests, and graphical procedures.

Scatter plots were used to look at relations between constituent concentrations and each sensitivity and intensity variable (figs. 1A and 1B). Boxplots were used to compare differences between two or more groups of data (fig. 1C). Minimum, median, maximum, and range of outlier concentrations were compared to determine whether differences between groups could be quantified. A boxplot is constructed by drawing a box between the lower and upper quartiles (25th and 75th percentiles) of a range of data. This range of data, known as the interquartile range (IQR), includes half of the observations. A line is drawn across the box to represent the median, which is the middle observation in the range of data. A straight line, called a whisker, is drawn from the box to the largest value within the upper quartile plus 1.5 times the IQR, and another whisker is drawn from the box to the smallest value within the lower quartile minus 1.5 times the IQR. Any extreme values lying outside of the whiskers are called outliers (Ott, 1993). Scatter plots and boxplots also were used to determine scores to apply when rating the susceptibility of wells and intakes to constituents. Maps were generated to assess the spatial distribution of water-quality data in relation to hydrogeologic sensitivity and potential contaminant-use intensity variables.

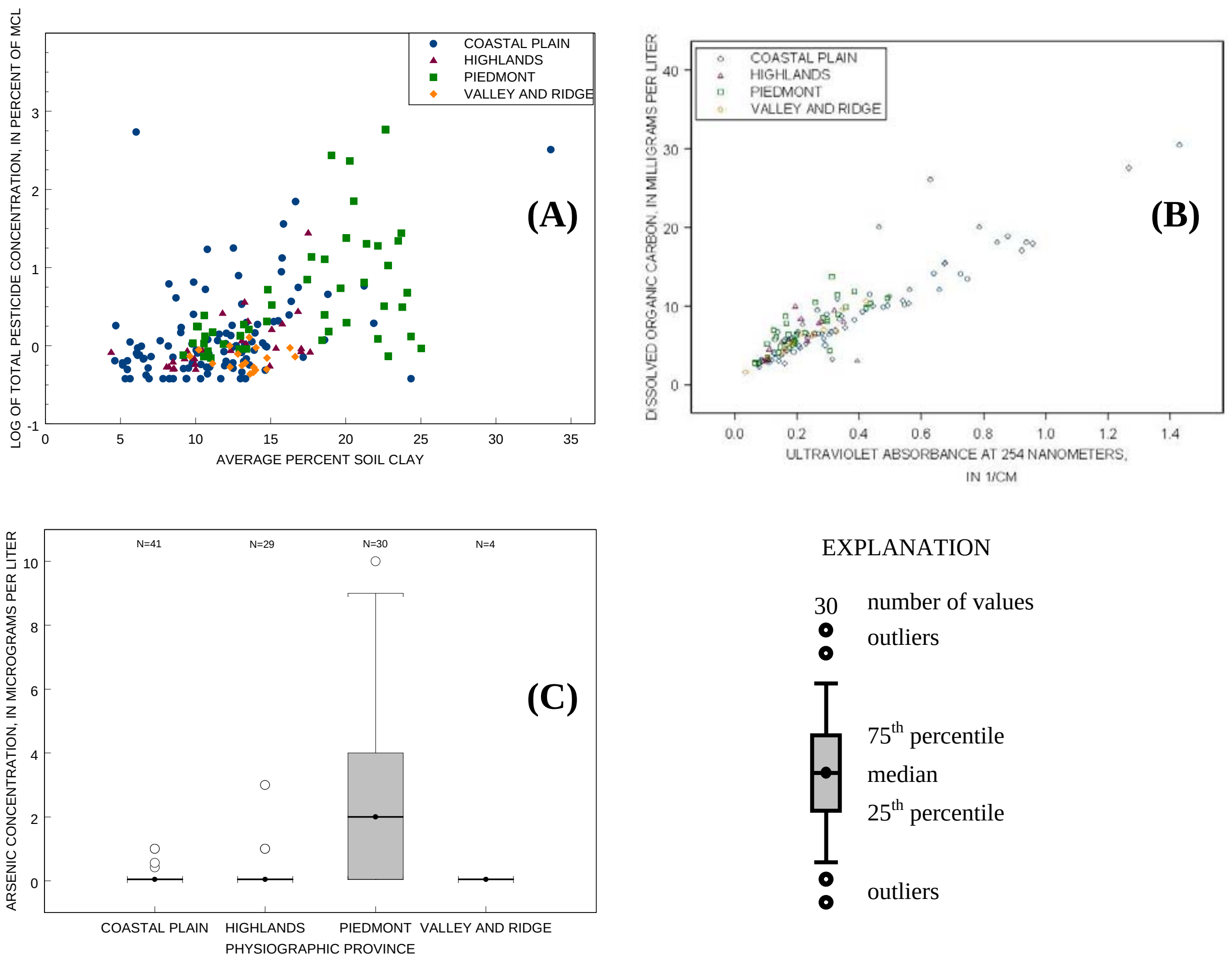

Figure 1. Example of boxplots and scatter plots used to select susceptibility model variables.

Results of univariate statistical tests (Spearman's rho and Kruskal-Wallis) and graphs (scatter plots and boxplots) were used to identify potential predictors of contamination at selected concentration levels relative to the MCL. In some cases, variables thought to be a good predictor of contamination did not produce a significant univariate statistical relation. In this report, conceptual variables are variables with possible graphical relations for which results of univariate statistical tests were not significant but that have been shown in a previous scientific investigation to be related to the concentrations of a constituent. Conceptual variables also are variables for which results of univariate statistical tests were or were not significant but that improve the model and may represent a surrogate for other unidentified variables associated with the concentration of a constituent, although no evidence was found in previous investigations of a relation. Conceptual variables that did not produce significant univariate statistical relations may, however, produce a significant relation when used with other variables in multivariate statistical tests. Only variables that were used in the model will be discussed.

Some variables that proved to be statistically significant were not used in the model. Some possible reasons for exclusion were (1) the variable was not a known source of the constituent modeled, (2) the use of the variable in the model was not supported by scientific investigations, (3) the variable did not show a graphical relation to the constituent, or (4) the variable was found to have a relation to the constituent similar to the relation with another variable.

Rating Scheme

A numerical scoring method was developed for each model that assigned points to each variable used in the model. Relations observed in scatter plots or boxplots were used as the starting points for devising the numerical code. In some cases, the scoring interval was based on a weighting scheme relative to the strength of the statistical relation. The maximum number of points was given to variables that appeared to work best statistically (both univariate and multivariate tests) and that graphically approached a linear relation. If, for example, when the average percent soil clay was statistically related (a positive Spearman's correlation, Kruskal-Wallis score of 40.25, and p-value of 0.000) to the concentration of pesticides and the percent soil clay within the contributing area for a surface-water site was large, a score of 5 was assigned. When the percent soil clay was small for a site, a score of 0 was assigned. Fewer points were given to variables that were less significant statistically, that had lower correlation coefficients, that appeared graphically to be grouped, or that did not show changes over the entire range of values. Relations observed in the graphs presented in this report were used as the starting point for devising the numerical code.

Model Application

All community water-supply wells and surface-water intakes were categorized as having low, medium, or high susceptibility based on scores of the susceptibility assessment model for each contaminant category. The low, medium, and high categories were grouped according to guidelines established by NJDEP. In general, the low category includes wells and intakes for which constituent values are not predicted to exceed one-tenth of the maximum contaminant level. The medium category includes wells and intakes for which constituent values are not predicted to exceed one-half of the maximum contaminant level. The high category includes wells and intakes for which constituent values may exceed one-half of the maximum contaminant level. If multiple constituent susceptibility assessment models were developed within a contaminant category, the susceptibility rating for that contaminant category is based on individual susceptibility assessment models. The contaminant category rating for a well or intake is the highest susceptibility rating of the individual constituent susceptibility ratings within that category.

Model Verification

Susceptibility assessment models were verified when an independent data set was available. Ground water VOC and ground water and surface water DBP precursor models were verified using boxplots to compare concentrations in water from the site with the susceptibility assessment of the site (figure #). In general, the medians of the boxplots should increase as the susceptibility assessment increases. Differences between data used to develop the model and data used to verify the model may explain differences in predicted susceptibility and actual concentrations.

Discussion

Several limitations to the susceptibility assessment models should be noted. These models should be used only as screening tools to identify potential contamination problems. The concentrations used for a well in the analysis were those measured in the most recently analyzed sample and do not take into account fluctuations in concentrations that may occur.

Some of the components of the analysis were subjective, especially the coding scheme used for the susceptibility assessment model. Problems could exist in the interpretation of data at a local scale and projecting to statewide scales. Using different scales for various GIS layers could bias statistical results, and land-use changes could cause spurious relations. The method used to determine source water assessment areas and tiers representing times of travel of water to the well is inexact and produces only estimates of the actual contributing area and the length of time the water is in transit before it reaches the well.

Results of statistical tests performed on groups of VOCs, relative to the threshold of concern of the NJDEP, might differ if performed on individual VOCs. The susceptibility rating represents a combination of both sensitivity and intensity, and in some cases may be inconsistent with the results of water-quality analyses. For example, a well may be considered highly susceptible to contamination by VOCs and have no detections in samples from that well if VOCs are not used within the contributing area, or if contamination has not yet reached the well.

The database, GIS coverages, statistical analysis, and susceptibility assessment models can be used by scientists and water managers to help determine effects of hydrogeology and land use on the quality of water of public supplies. The relations between water quality and susceptibility variables shown in figures, graphs, and tables can be used in determining and evaluating monitoring requirements for water purveyors to ensure public health.

Summary

The susceptibility of more than 2,300 CWS wells, and 60 surface-water intakes and sources using ground water under the direct influence of surface water, was determined. Sensitivity and intensity variables were related to well, surface-water-quality site, and intake source water assessment areas by using a GIS. Statistical and graphical relations between sensitivity and intensity variables and water-quality concentrations were determined to select variables to use to assess susceptibility. Constituents were grouped into eight categories for susceptibility assessments.

Susceptibility ratings from the application of susceptibility assessment models for wells and intakes are available (NJDEP web site?). SOC models for ground and surface water were not developed because of insufficient data. All surface-water intakes are considered by NJDEP to be susceptible to contamination by pathogens and none are considered to be susceptible to contamination by radionuclides.

References Cited

Ellis, W.H., Jr., and Price, C.V., 1995, Development of a 14-digit hydrologic coding scheme and boundary data set for New Jersey: U.S. Geological Survey Water-Resources Investigations Report 95-4134, 1 sheet.

Helsel, D.R,. and Hirsch, R.M., 2002, Statistical methods in water resources, in U.S. Geological Survey Techniques of Water Resources Investigations, book 4, Hydrologic analysis and interpretation, chapter A3, 510 p.

N.J. Department of Environmental Protection Source Water Assessment Reports, 2004, Surface Water Delineation Methodology.

Ott, R.L., 1993, An introduction to statistical methods and data analysis: Belmont, CA, Duxbury Press, 1170 p.

Stevens, A.A., Slocum, C.J., Seeger, D.P., and Roebeck, G.G., 1976, Chlorination of organics in drinking water: Journal of the American Water Works Association, v. 68, p. 615.