EVALUATION OF REGIONAL GROUND-WATER QUALITY IN RELATION TO LAND USE: U.S. GEOLOGICAL SURVEY TOXIC WASTE—GROUND-WATER CONTAMINATION PROGRAM

By Dennis R. Helsel and Stephen E. Ragone

U.S. DEPARTMENT OF THE INTERIOR
March 3, 1849

U.S. GEOLOGICAL SURVEY
Water Resources Investigations Report 84-4217
1984

UNITED STATES DEPARTMENT OF THE INTERIOR
WILLIAM P. CLARK, Secretary

GEOLOGICAL SURVEY
Dallas L. Peck, Director

For additional information
write to:

Chief Hydrologist
U.S. Geological Survey, WRD
410 National Center
Reston, Virginia 22092

Copies of the report can be
purchased from:

Open-File Services Section
Western Distribution Branch
U.S. Geological Survey
Box 25425, Federal Center
Denver, Colorado 80225

CONTENTS

Page

ILLUSTRATIONS

ILLUSTRATIONS--continued

TABLE

CONVERSION FACTORS

For the convenience of readers who prefer to use the International System of units (SI), conversion factors for terms used in this report are listed below:

Multiply inch-pound unit	By	To obtain SI unit
foot (ft)	0.3048	meter (m)
gallon (gal)	0.003785	cubic meter (m^3)

EVALUATION OF REGIONAL GROUND-WATER QUALITY IN RELATION TO LAND USE

by

Dennis R. Helsel and Stephen E. Ragone

ABSTRACT

The objective of the ground-water quality appraisals being conducted by the Toxic Waste--Ground-Water Contamination Program of the U.S. Geological Survey is to assess the current quality of the Nation's ground-water reserves and the nature and extent of the ground-water contamination problem. This paper describes a five-step procedure to study the impact of anthropogenic factors on regional ground-water chemistry, with particular emphasis on organics and trace metals, and to determine if this information has application to other parts of the ground-water reserves of the Nation.

The five steps are:

(1) Reconnaissance and the development of a conceptual model--All relevant hydrogeologic, water quality, and cultural information is assembled so that a conceptual model can be developed which proposes the predominant factors affecting ground-water quality.

(2) Experimental design--A plan for the collection of additional hydrogeologic, water quality, and cultural information is developed so that the hydrology and ground-water chemistry can be accurately characterized and that the chances for detecting the factors influencing ground-water quality can be maximized.

(3) Data collection.

(4) Data analysis--The information is evaluated with regard to the conceptual model.

(5) Verification--Information collected for other areas of similar land use in the same hydrogeologic and climatic setting is used to determine if the inferences drawn about the factors affecting ground-water quality in the primary study area are applicable to other parts of the ground-water system.

INTRODUCTION

Problem

Ground-water quality investigations often fail to consider both the complex nature of the subsurface and the variability of contaminant inputs over space and time. Many investigators collect ground-water samples from municipal water supply wells having multiple well screens or single, large screens open to many tens of feet of aquifer. Although such information is useful for defining the chemistry of water currently being supplied, it provides little knowledge about either the spatial variability of the chemistry of the ground-water reserve, or the effect over time of natural or anthropogenic stresses on the reserve. In the ground-water regime, small-scale variability in grain size, mineralogy, grain surface coatings or the activity of microbiological agents may affect ground-water chemistry over relatively short distances. Larger scale variations in the thickness, lithology and areal extent of aquifers tends to influence regional ground-water flow patterns. Geochemical and hydrogeologic factors such as these affect the fate of contaminants in the subsurface and may influence our ability to obtain samples which accurately represent the ground-water reserve.

The contaminant source also will influence the nature and extent of subsurface contamination. Contaminants may enter the subsurface from identifiable "point" sources such as landfills and industrial lagoons, or from less intense and widely dispersed sources. The latter may include "nonpoint" inputs from agriculture and atmospheric deposition, as well as closely spaced multiple-point sources such as septic systems in densely populated suburban areas. Because of the slow rate of ground-water movement--which seldom exceeds 10 feet per day and is often measured in hundredths of a foot per day--point sources often generate relatively localized pockets of contamination which expand only slowly with time. The heterogeneity in subsurface-contaminant concentrations due to these differences in source and in subsurface attributes must be investigated systematically within individual areas, so that information from these areas can be extrapolated to wider ground-water regions of the Nation.

Objective

A program of ground-water quality appraisals is presently being conducted under the Toxic Waste--Ground-Water Contamination Program of the U.S. Geological Survey; the overall objective is to assess the current quality of the Nation's ground-water reserves and the nature and extent of the ground-water contamination problem. Toward this objective, 14 study areas were selected in Fiscal Year 1984 (FY84) covering a wide variety of environments in terms of geohydrology, climate, and human activities. These studies will provide information on ground-water chemistry with emphasis on organic substances and trace metals, and will attempt to explain the water quality in terms of local hydrology and human activities.

The purpose of this report is to outline certain methods for investigating the effect of land use on ground-water chemistry, and to demonstrate how the results from individual studies may be transferred to other areas as part of a national appraisal.

Approach

National Appraisal

The study areas are listed by location in table 1. Study areas range in size from a few tens of square miles to a few thousand square miles; whatever the size, however, each area is characterized by internally uniform climate and includes more than one type of land-use subarea. The transitions from one land use to another are well defined so that correlations between ground-water quality and land use can be studied.

Table 1. Areas selected for preliminary ground-water quality studies.

Gulf Coastal Plain, Louisiana and Mississippi
Long Island, New York
Philadelphia area, Pennsylvania
Potomac-Raritan-Magothy outcrop area, New Jersey
Houston area, Texas
Floridian Aquifer, Florida
Edwards Aquifer, Texas
Combined Regolith, North Carolina
San Joaquin Valley, California
Arkansas River Valley, Colorado
Albuquerque-Belen Basin, New Mexico
High Plains Aquifer, Kansas
High Plains Aquifer, Nebraska
Connecticut River Valley, Connecticut

As mentioned earlier, a number of difficulties face any attempt to characterize regional ground-water quality with respect to land use. Variation in geologic and hydrologic properties causes variability in the direction and velocity of ground-water movement which, in turn, affects the distribution and mixing of contaminants. Variation in geochemical properties of the ground-water system leads to variation in the reactions which affect contaminant concentrations. Moreover, contamination sources may be too diverse, or ground-water contamination may be too localized, to have a predictable or discernible impact on ground-water quality on a regional scale.

This report discusses methods of evaluating the spatial and temporal variability of ground-water contamination within each land use and between land uses within a given ground-water system. Each study will investigate whether confounding factors, such as those listed above, mask the impact of land use on ground-water quality. In those studies where strong correlations between land use and ground-water quality exist, the information collected for the individual study areas will be compared with information from other areas of similar land use and similar hydrogeologic and climatic setting. If similar patterns of ground-water contamination are observed, the extrapolation of information from the study area to similar systems should be possible. On the other hand, large variability in the type or concentration of contaminants within land-use areas will indicate that factors other than land use control ground-water chemistry.

Individual Study Areas

Each of the 14 individual studies in the USGS program will be regional in scope and statistical in approach. The studies are of regional scope in order to take into account the varied and numerous sources of contaminants which may infiltrate ground water. Miller (1980) reports that more than 17 million waste-disposal facilities emplace more than 1,700 billion gallons of contaminated liquid into the ground each year. It is estimated that 75,000 to 100,000 underground tanks of gasoline and related hydrocarbons currently are leaking, and that up to 350,000 will be leaking within 5 years (Shaner, 1982). Agricultural sources may have regional effects on ground waters. For example, Cohen and others (1984) found 12 pesticides of concern in ground waters of 18 States due to widespread agricultural uses. The regional approach described in this report will take into account these varied sources by aggregating point sources, multiple-point sources, and nonpoint sources into land-use categories. Areas containing a significant number of industrial pits, ponds, and lagoons, for instance, will be designated as industrial land use; suburban areas containing multiple-point sources of contamination, such as septic systems, also will constitute a designated land use. In all cases, the designated land use will be based upon predominant, definable sources of contamination.

The 14 studies will be statistical in approach in that formal statistical procedures will test hypotheses which attempt to explain the observed variation of chemical data. In addition to land use, the influence of other factors might be investigated. This approach is midway (number 3) in a continuum of descriptive to deterministic approaches as outlined by R.M. Hirsch (U.S. Geological Survey written communication, 1984):

(1) Comparison (over space and time) of summary statistics (for example, means, variances, and correlations) and exploratory data analysis.

(2) Naive hypothesis testing (for example, tests for association between variables believed to have some causal relationship, or a test for time trends using standard statistical approaches).

(3) Sophisticated hypothesis testing which uses the principles of hydrology and various statistical procedures to eliminate one or more sources of variation in the data (for example, multiple regression, analysis of covariance, residuals analysis).

(4) Parameter estimation and evaluation of models which incorporate explicit mathematical representations of the fundamental physical, chemical, and biological processes, and a stochastic error model.

(5) Calibration and verification of deterministic models based solely on mathematical representations of the physical, chemical, and biological processes.

The approach used here differs from the descriptive approaches, (1) and (2), in that a conceptual model of suspected influences is generated at the outset, on the basis of knowledge of the hydrogeologic system and contaminant sources. On the other hand, the deterministic approaches, (4) and (5), would require more details about subsurface processes and amounts and sources of contaminants than is possible under the time and financial constraints of a regional ground-water quality study.

DESIGN OF REGIONAL STUDIES

Each of the 14 regional studies should be designed using five components:

(1) Reconnaissance and development of a conceptual model. Assemble all relevant information and develop a conceptual model concerning the factors affecting the distribution of solutes in ground water.

(2) Experimental design. Plan for the collection of additional hydrologic, geologic, and water-quality information to accurately characterize the hydrology and water chemistry of the ground-water system and to maximize chances of determining the factors influencing ground-water quality.

(3) Data collection.

(4) Data analysis. Evaluate the information with regard to the conceptual model.

(5) Verification. Use the information collected for other areas of similar land use and similar geohydrologic and climatic setting to determine if inferences about factors affecting ground-water quality can be extrapolated beyond the study area itself.

Reconnaissance and Development of the Conceptual Model

Reconnaissance involves the review of all pertinent information concerning the hydrology, geology, water quality, and land- and water-use practices in the study area. Using this information, a conceputal model should be developed which relates ground-water quality to controlling factors.

Review of Existing Information

Hydrologic, geologic, water-quality, and land-use information should be collected during reconnaissance from all available sources. To the extent possible, land use information should include historical as well as current data. Hydrologic, geologic and water-quality information should be taken from published reports, from the records of Federal, State and local agencies, and from any private records or consulting reports which are available.

This information should be assembled to create a description of the ground-water flow system. This analysis is essential in order to define the part of the ground-water system to be studied and in order to design the ground-water quality sampling network. The scale of the effort to describe the flow system will vary according to what has been done in the past. In some cases, very little may be required; in others, ground-water models may have to be constructed or updated and simulations carried out to define ground-water flow patterns. At a minimum, ground-water flow patterns must be known in sufficient detail to determine residence times and flow paths of ground water in the different parts of the ground-water system. Generally, waters less than 100 years old may have been affected by anthropogenic contamination and should be the primary interest of the regional appraisals. In some cases, however, the zone of contamination may have to be restricted to ground water of significantly younger age because of changing land use practices which affect the type of contaminant infiltrating ground water. In other cases, ground water which is older than 100 years may appear to be contaminated, particularly where pumpage has caused migration of contaminants to parts of the ground-water system which would otherwise be uncontaminated. Leaky well casings or wells with large well screens may also result in the migration of contaminants into zones of older ground water by "short-circuiting" flow paths.

If possible, zones of uncontaminated ground water should be evaluated in each study area to serve as a control so that the relative impact of cultural practice on ground-water quality can be measured. Data for nitrate, chloride, and sulfate generally are available and may provide a basis for determining whether ground water has been affected by human activities, even where these constituents themselves are not primary targets of investigation.

However, these constituents cannot be used as surrogates for organic or trace metal contaminants, as their transport and fate in the subsurface is generally controlled by different chemical, physical and biological reactions.

The approaches proposed in this paper are based upon the premise that the contamination of shallow ground water by human activities can be considered a function of land use. Specifically, it is assumed that land uses determine the types of chemicals that are applied to the surface and thus the types of contaminants that reach the ground water. The categories of land use considered in an individual investigation should be segregated as specifically as possible with respect to sources of contamination. For example, agricultural areas might be segregated by crop type or irrigation practice, while suburban areas might be separated into sewered and unsewered districts. Care in selecting appropriate land uses will help to minimize variation in the data and improve chances for determining which factors are affecting ground-water quality. Using the knowledge gained during reconnaissance about the hydrogeology, water quality and cultural practices in the study area, the investigator must develop a conceptual model concerning the factor(s) presumed to be affecting ground-water quality. This conceptual model then guides the design of data-collection activities so that the effects on water quality of each factor may be determined. The model may be altered during subsequent stages of the study as new information becomes available.

Example of Reconnaissance Interpretations

Geological Survey experience in Long Island, New York (Figure 1), provides an example of how existing information, assembled during the reconnaissance phase, can be used to develop conceptual models and guide later phases of

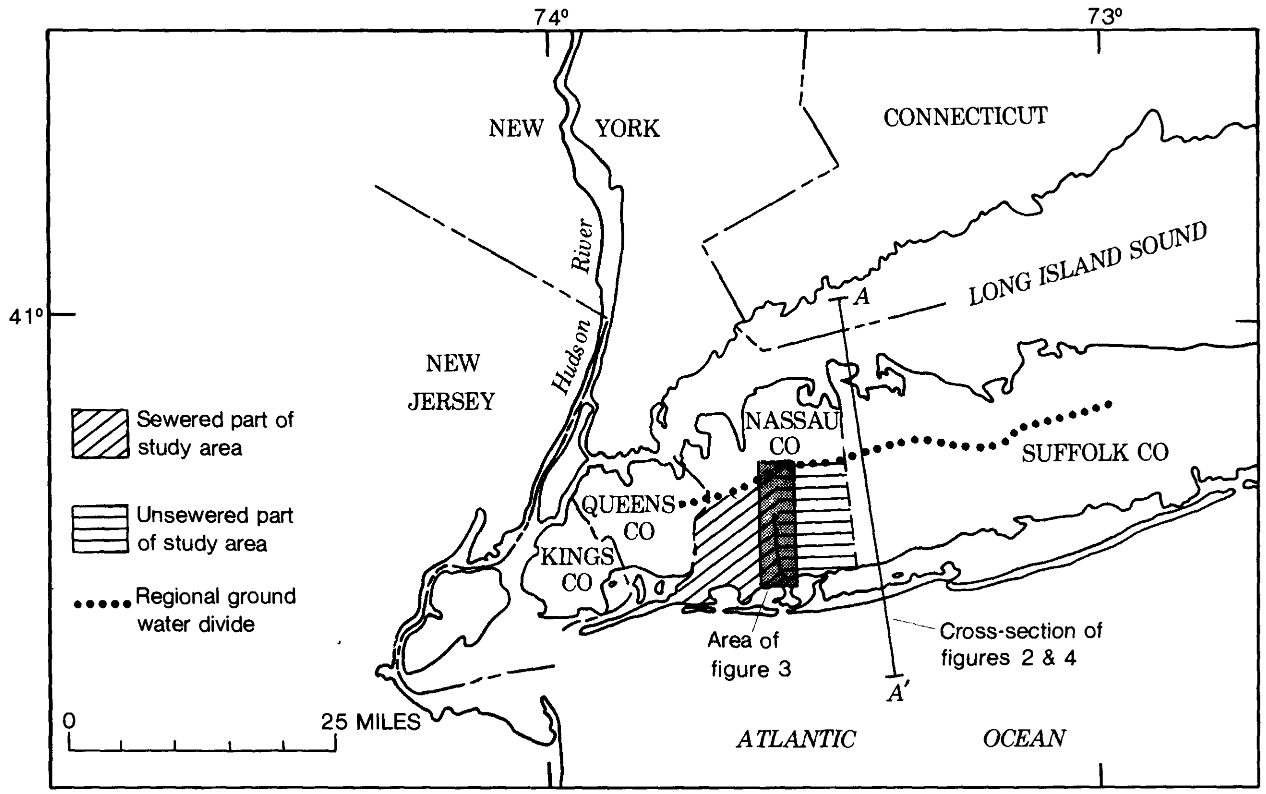


Figure 1. Index map of Long Island, N.Y., showing the regional ground water divide; the sewered and unsewered parts of the study area; the area of figure 3; and the location of cross section A-A' of figures 2 and 4.(after Franke and Cohen, 1972).

the investigation. The example presented here relates primarily to the relation of chlorinated hydrocarbons in ground water to unsewered high density suburban land use, and is not intended as a complete description of a ground-water contamination evaluation in Long Island.

The ground-water flow regime on Long Island is well defined and may be divided into a shallow subsystem which occurs largely within the upper Glacial Aquifer, and in which the discharge is primarily to streams, and a deep subsystem which extends downward into the Magothy and Lloyd Aquifers, and discharges into the bays, the ocean and Long Island Sound (figure 2). Flow paths in the shallow subsystem may range from a few meters to several kilometers in length (Franke and McClymonds 1972). Using a steady-state electric analog model, Franke and Cohen (1972) calculated the times required for water entering the shallow ground-water subsystem in the vicinity of East Meadow Brook to discharge into the stream. Their results are shown as contours in figure 3. For dissolved substances which are assumed to move at the same rate as the water, residence times in the shallow subsystem range from less than 5 years to approximately 30 years. Recharge to the deep subsystem occurs mostly in the zone of the regional ground-water divide (figure 2), where the flow is downward through the upper Glacial Aquifer into the Magothy Aquifer, then outward toward the periphery of the island.

The time required for water to travel from the water table to various locations in the deep subsystem also was estimated by Franke and Cohen (1972) using the analog model. The results are shown as contours in figure 4. The times of travel range from less than 25 years for points near the ground-water divide, to more than 800 years for points in the Magothy Aquifer beneath the coastline.

Residential areas in Nassau County, Long Island can be classified as either sewered or unsewered; sewered areas can be further subdivided according to the time since sewering was completed. In southwestern Nassau County sewering was completed more than 20 years ago. Sewering is presently being completed in southeastern Nassau County. Prior to sewering, residential waste disposal in almost all areas of the county was handled through individual cesspool systems.

The chemistry of ground water in Nassau County has been studied in the past with respect to nitrates, and the results suggest strongly that cultural factors have in fact affected ground-water quality (Ragone and others, 1981). Nitrate concentrations are highest in the shallow glacial aquifer, and in the zone of recharge of the Magothy Aquifer (figures 5a and 5b). Concentrations decrease downgradient, eventually reaching levels which presumably represent uncontaminated conditions. However, it is by no means clear that the nitrates in the shallow ground water are solely related to residential waste disposal. Prior to suburban residential development on Long Island, agriculture was the dominant land use. Thus there was probably an input of nitrate from agricultural use, followed by an input from unsewered residential land use, in most areas of the county.

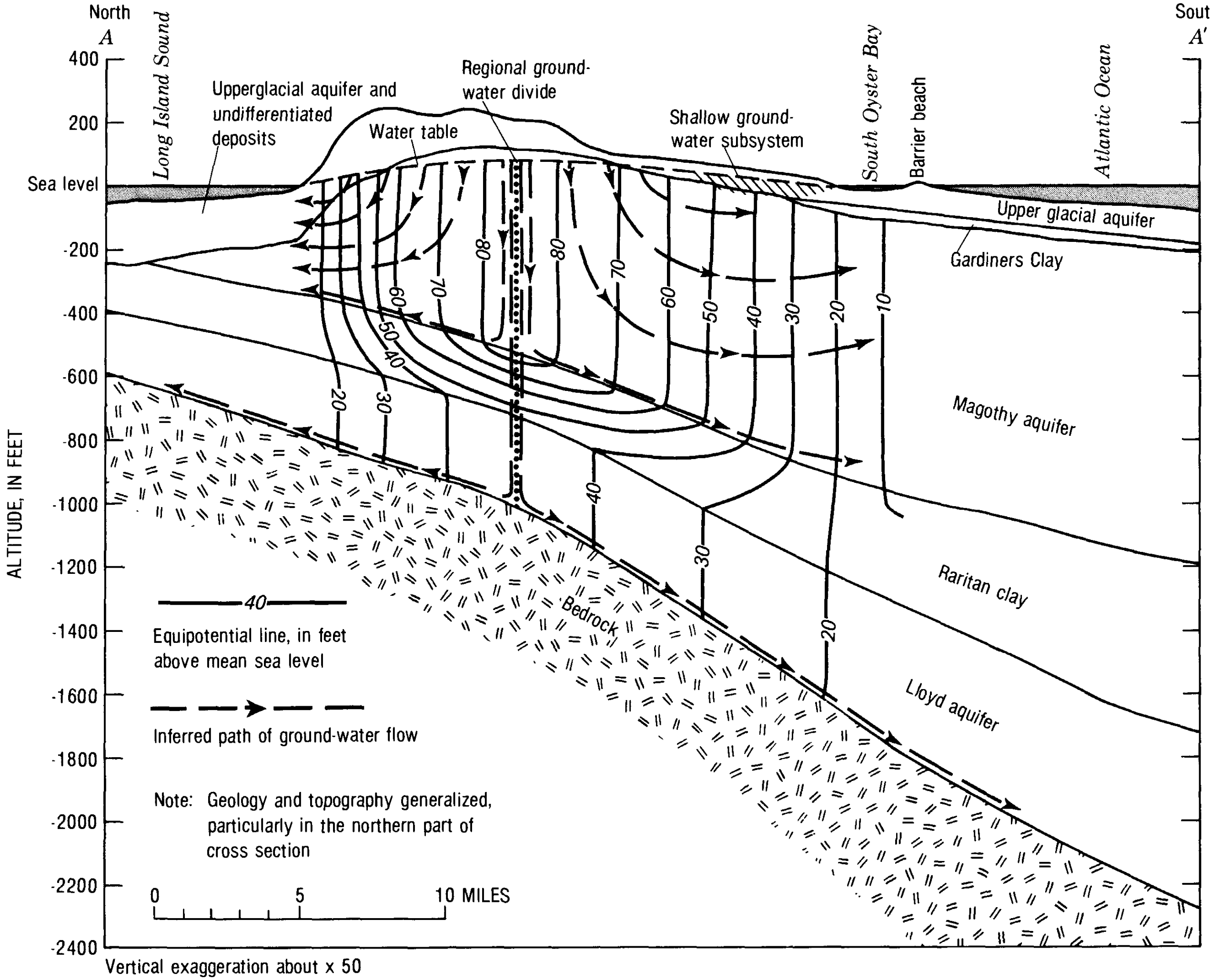


Figure 2. Hydrogeologic section showing generalized equipotential and flow lines in the ground-water system of Long Island, N. Y., under natural conditions (from Franke and Cohen, 1972).

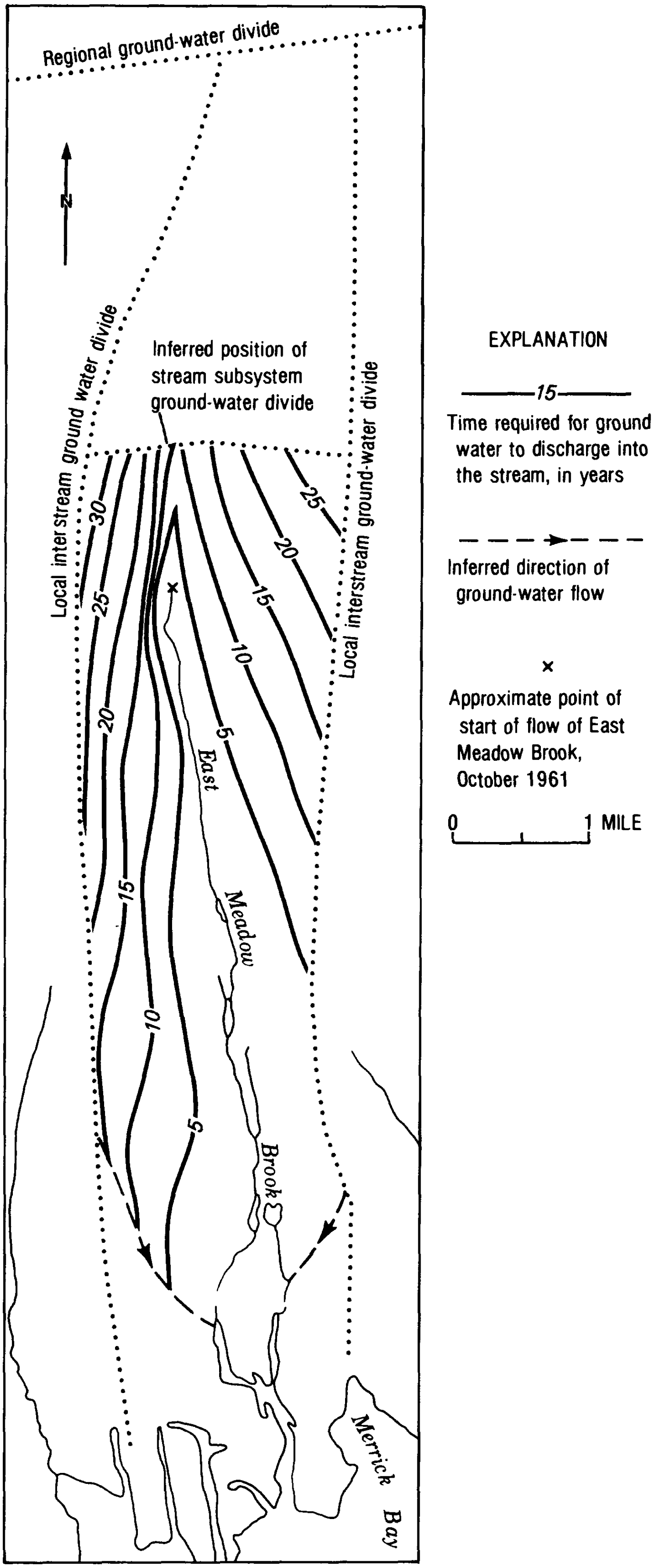


Figure 3. Map showing approximate time required for ground water in shallow-aquifer subsystem to discharge into East Meadow Brook (from Franke and Cohen, 1972).

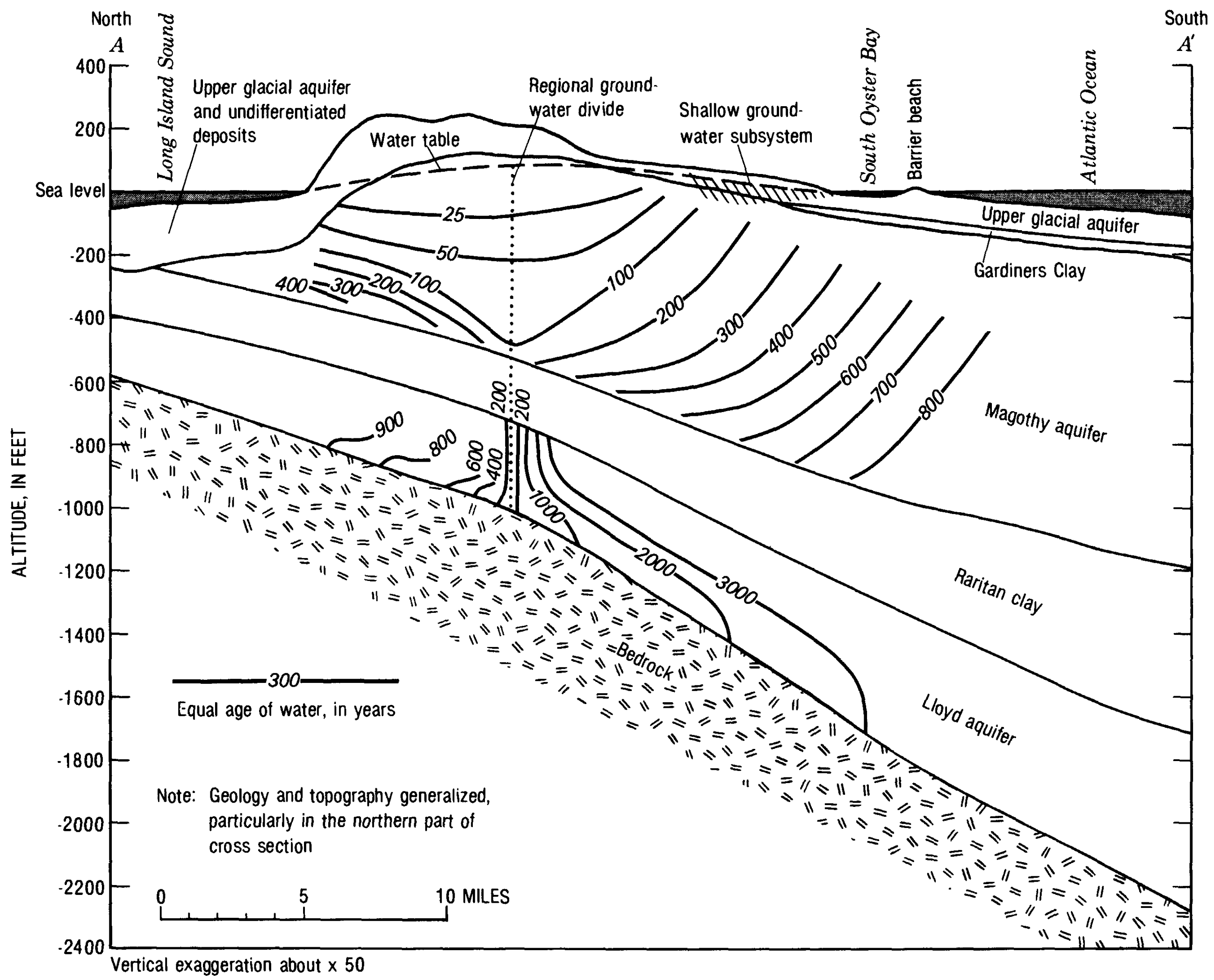


Figure 4. Hydrogeologic section showing time required for water to move from water table to points within the deep-aquifer subsystem of Long Island, N.Y., under natural flow conditions. (from Franke and Cohen, 1972).

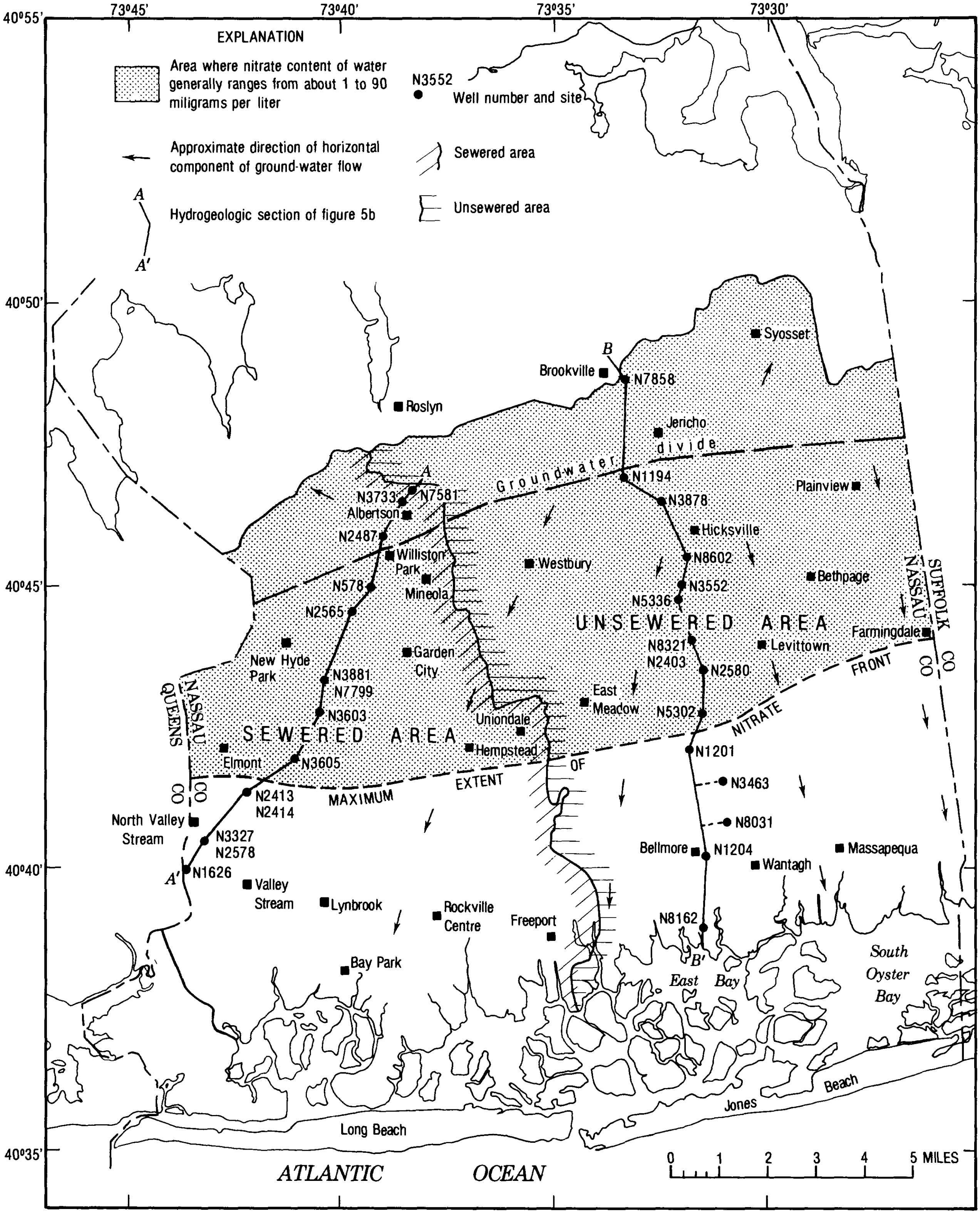


Figure 5a. Approximate areal extent of main body of nitrate enriched water and maximum southern position of nitrate front in Magothy aquifer, Nassau County, Long Island, N.Y., 1969-70. Nitrate content of Magothy aquifer south of nitrate front is generally less than 1 mg/l. (after Perlmutter and Koch, 1972).

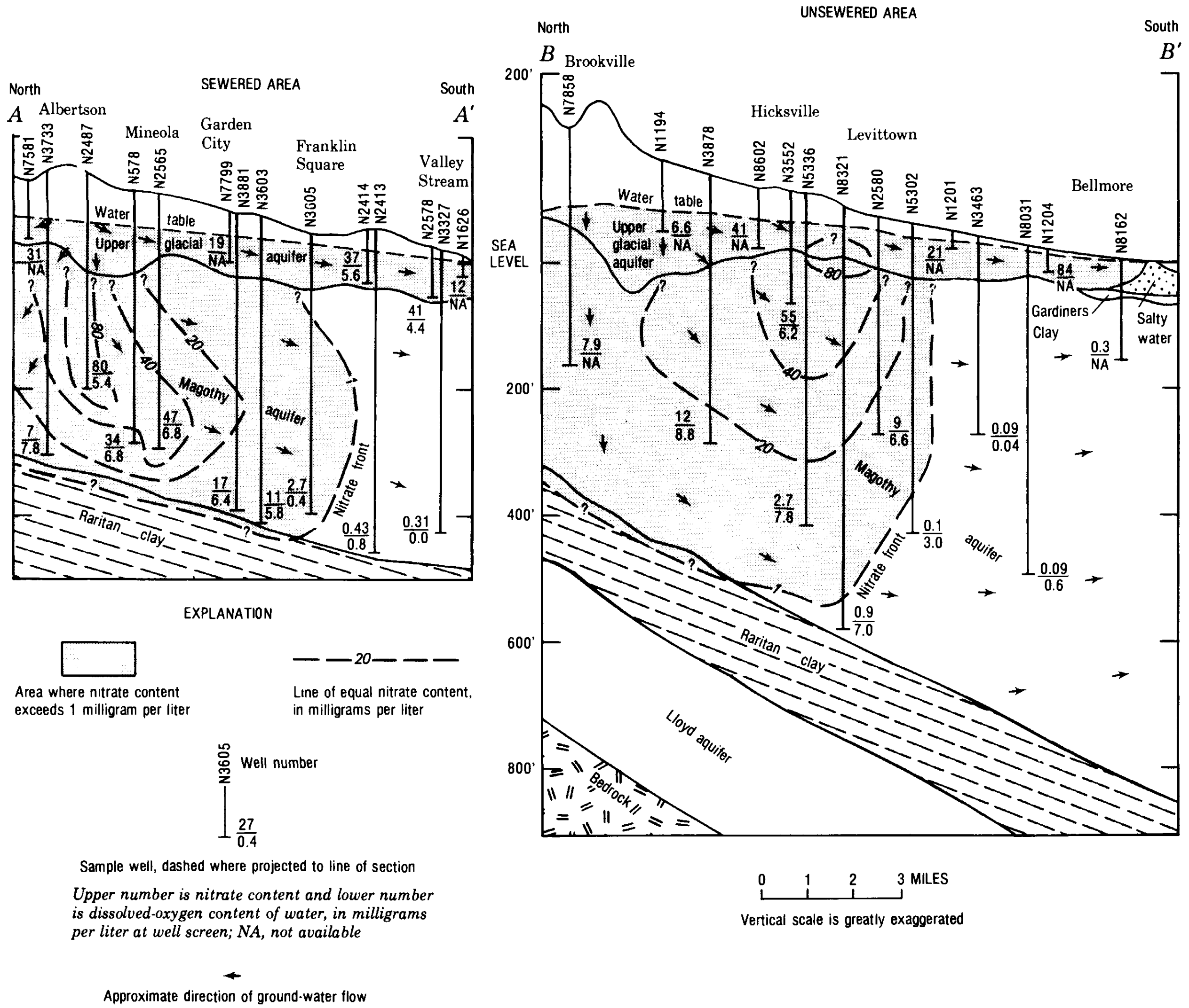


Figure 5b. Vertical distribution of nitrate in water in upper glacial and Magothy aquifers, southern Nassau County, Long Island, N.Y. (from Perlmutter and Koch, 1972).

A relation between land use and ground-water contamination by chlorinated organic compounds, however, may be less ambiguous. During the past 30 years chlorinated organic solvents such as trichloroethylene (TCE) and methylene chloride have been used extensively in the cleaning of residential cesspools. According to the Council of Environmental Quality (1981) "An estimated 40,000 gallons of TCE were used by homeowners on Long Island in 1979 alone."

These patterns of chemical use, coupled with the characteristics of the ground-water flow regime outlined above, provide a number of guidelines for an investigation of ground-water contamination in Nassau County. Such an investigation should include nitrates, but new data collection should be highly focused to answer specific questions. A broader survey of chlorinated organic solvents would clearly also be included, as there is no existing large data base for organics such as for nitrates. A great deal of sampling and chemical analysis would be involved. These compounds are expected to reside in ground waters less than 30 years of age. This would include virtually all ground water in the Upper Glacical Aquifer, as well as a limited zone of shallow ground water in the Magothy Aquifer, close to the regional ground-water divide on figure 4. Finally, it would be reasonable to expect the distribution of these compounds to be different in sewered and unsewered areas.

Collection of New Information

In the balance of this paper it is assumed that detailed land-use information and a detailed hydrogeologic evaluation, similar to those in the Nassau County example, are available for the area of investigation. If this is not the case, these topics must be addressed through the collection of new information during the reconnaissance phase of the study. In particular, enough geohydrologic data must be collected and analyzed to define flow patterns, residence times and target sampling populations, as in the Long Island example.

If adequate land-use and geohydrologic information already exist, the acquisition of new information during the reconnaissance phase will involve primarily the collection and chemical analysis of ground-water samples. The distribution of solutes in the ground-water system is three dimensional; to the extent that samples are taken from wells with unknown screen depths, multiple screens, or long open intervals, the depth of origin of the samples will be unknown, or the samples may represent an integration of the water quality at different depths. In either case, the three dimensional characterization of the water-quality distribution will be lost. For this reason, it is recommended that sampling always be done from wells equipped with a single short screen, in a known depth interval. In the balance of this report, it is assumed that sampling can be done in this matter.

Unless extensive analyses of organic contaminants and trace metals are available for the study area, valuable information can be gained by limited chemical analysis of samples collected during reconnaissance. Gas Chromatograph--Flame Ionization Detector (GC-FID) scans or other semi-quantitative methods could be used to indicate the presence of semi-volatile and extractable organic compounds for a relatively low cost. Methods must be chosen with sufficiently low detection limits for the compounds of interest. Such scans of samples collected throughout the study area may provide valuable information to improve the conceptual model prior to a full-scale sampling effort. On Long Island, for example, scans should show peaks at the proper retention time of TCE in samples under unsewered areas if the conceptual model is correct. Scans may also indicate that previously overlooked land uses are contributing contaminants.

One of the major uses of limited reconnaissance sampling is to estimate the variability in concentrations of important contaminants at and between sites (wells). Gas Chromatograph-Mass Spectrograph (GC-MS) analyses provide reliable identification and concentration data for organic compounds. Atomic absorption provides reliable concentration data for most trace metals. Using a small number of representative organic compounds (one soluble, one strongly sorbed and one weakly sorbed on aquifer materials) plus one or two trace metals of interest, the variability of concentrations within subgroups of the study area can be quantified. This will aid in determining the optimal number of samples for the large-scale data collection program, as outlined in the next section.

Experimental Design

In designing an experiment to test the conceptual model, hydrologic, chemical, and statistical expertise is required. The study's experimental design should:

(1) Accurately characterize the hydrology and chemistry of the study area through representative sampling.

(2) Avoid mistaking the effects of one factor for a second through random sampling.

(3) Maximize detection of a factor's influence for a given cost through optimal sample allocation.

Representative sampling occurs when the data collected (using waters from a small number of wells) have characteristics identical to the entire population under study (all water in a specific part of the aquifer). This may be thwarted in several ways. Suppose the ground-water quality in a several-county area is to be described, and 10 percent of all wells in the area are sampled. If 75 percent of the area's wells are located around landfills, the resulting data will be biased toward water quality around those landfills. The wells selected will not adequately represent general water quality of the overall aquifer. A design which segregated wells near landfills into a separate category would provide better descriptions of overall water quality. As a second example, suppose the organics of interest adsorb to plastic well casings. To representatively sample aquifer conditions, wells having these casings should be avoided. Representative sampling results solely from good professional judgment based on hydrologic and chemical knowledge.

Once all wells which will truly represent aquifer conditions are designated, some method is needed to select those which will actually be sampled. Within each subgroup of the study area, random sampling occurs when each well has an equal chance of being sampled. If wells are not selected randomly, hidden factors not previously accounted for may produce effects mistakenly attributed to one of the factors being investigated. For example, suppose two areas of an aquifer were compared using data from seven wells in each area. Assume further that, because they were not randomly selected, wells in one area were screened exclusively in coarse gravels, while wells in the second area were screened in a variety of aquifer materials. Concentration differences associated with this difference in aquifer lithology might then be mistakenly attributed to the primary effect (land use) being tested. Had formal procedures for random selection of wells been used, the "aquifer lithology effect" would have been added to the background variability within each group and could be ignored as a particular influence. The major obstacle to random selection is convenience--if it is more convenient to select sampling points in a particular fashion, the process is probably not random and should be avoided.

Finally, optimization methods can be used to allocate numbers of samples so that detection of true differences can be maximized for a given cost. This may be accomplished using analysis of variance (ANOVA). ANOVA compares the data variance within each subgroup (a land-use type, for example) with the variance of all data combined. If the total variance is greater than the variance within single subgroups, at least one subgroup is discernible from the others (fig. 6a). However, if the total and within variances are similar, no differences between subgroups can be detected (fig. 6b).

To illustrate the allocation procedure, suppose we were interested in comparing water quality from wells representing two land-use types. Twenty-five wells are examined in each area, and two samples are collected from each well. The important questions are:

(1) Do the concentrations differ between areas?

(2) How much do concentrations differ between wells in one area? Will this obscure the true differences between areas?

(3) How much do concentrations vary between samples from one well?

If the variance is large between wells within one area or between samples from one well, any differences between land-use types may be obscured. If this were the case, more samples collected at the appropriate level (within one area or within one well) would reduce the variance and allow any differences between land use to be seen.

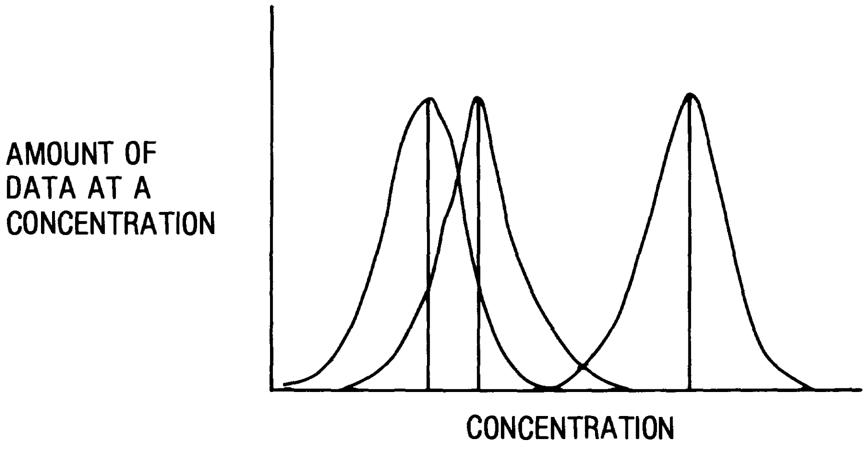
Sokal and Rohlf (1969) present equations to determine optimal sample allocations. Applied to the above example, they would be

$$n_c = \max\left(\sqrt{\frac{C_b \; S_c^2}{C_c \; S_b^2}} \; , 1\right) \qquad (1)$$

$$C = n_a \left(n_b \, C_b + n_c \, n_b \, C_c\right) \qquad (2)$$

where C is the overall project cost;

a. When total variance for all data combined > within group variance



b. When total variance for all data combined ≅ within group variance

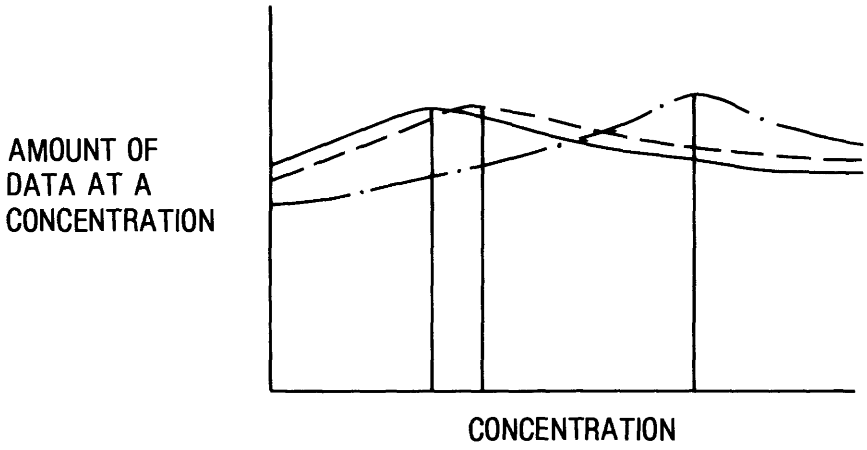


Figure 6. Graphs illustrating discernment of means as a function of variance.

n_a is the number of areas to be compared (fixed by the experimenter);
n_b is the number of wells per area;
n_c is the number of samples per well;
C_b is the cost of sampling an additional well (the field visit cost not including costs dependent on the number of samples taken);
C_c is the cost of an additional sample at a well (shipment and laboratory costs);

S_b^2 is the variance estimate for concentrations between wells, or:

$$\frac{n_c \sum^{a} \sum^{b} (\overline{Y}_c^{\,2} - \overline{Y}_{bc}^{\,2})}{n_a \, (n_b - 1)}$$

S_c^2 is the variance estimate for concentrations of samples from the same well, or:

$$\frac{\sum^{a} \sum^{b} \sum^{c} (Y^2 - \overline{Y}_c^{\,2})}{n_a \, n_b \, (n_c - 1)}$$

Y is the concentration of a sample;

$$\overline{Y}_{bc} = \frac{\sum^{b} \sum^{c} Y}{n_b \, n_c} \text{ , and}$$

$$\overline{Y}_c = \frac{\sum^{c} Y}{n_c}$$

Note that when the average variance increases within a well (S_c^2 increases), the number of samples required (n_c) also increases.

Sample sizes n_b and n_c are to be optimized for a given overall cost C. Required estimates of the variance components S_b^2 and S_c^2 should usually be obtained from GC-MS analyses during reconnaissance. Using the results of approximately four GC-MS analyses per subgroup, the variation between wells (S_b^2) within a subgroup can be estimated. Variation within one well (S_c^2) should be estimated by taking several samples from one or two wells per subgroup during reconnaissance. Only one sample should be taken per visit; replicates taken at the same time will not provide a proper estimate of within-well variation. Variance estimates for both sorbed and soluble constituents should be determined using the same quantitative analytical methods employed in the later detailed data collection program. Surrogate inorganic species such as nitrate should not be used to estimate S_b^2 and S_c^2 unless the source areas for both the surrogate and the organic species of interest are identical, and their transport rates and mechanisms are known to be similar.

Suppose reconnaissance information shows $S_b^2 = 100$ and $S_c^2 = 25$. Suppose we also know the field cost of sampling a well, C_b, is $300 and the cost of taking an additional sample at a well, C_c (essentially the laboratory cost) is $150. The optimum design with a project cost of $10,000 for distinguishing between three areas would be:

$$n_c = \max \left(\frac{300 \times 25}{150 \times 100} , 1 \right)$$

$$= \max \left(0.5, 1 \right)$$

$$= 1$$

From equation 2:

$$\$10,000 = 3\ [n_b\ (300) + (1)n_b\ (150)]$$

or $n_b = 7.4$. These results indicate that only one sample should be taken per well, and that seven to eight wells should be sampled per area. Because $n_c<1$, any additional funds should be spent on sampling more wells, rather than on more samples per well (although periodic quality assurance replicates are still necessary). Note that if sampling costs at a well and laboratory costs were the only costs considered (travel and salary costs c_b are equal to 0), n_c would always be 1. On the other hand, for non-zero c_b there is some variance ratio $S_c{}^2/S_b{}^2$ for which multiple samples are warranted.

Data Collection

Description of sampling and analytical methods is beyond the scope of this report; the reader is referred to recent reports by Scalf and others (1981), Dunlap and others (1981), and Claassen (1982) for discussions of the well emplacement and construction techniques and sampling methodologies. Analytical methods and quality assurance are described by Barnett and Mallory (1971), Skougstad and others (1979), Friedman and Erdmann (1982), and Wershaw and others (1983).

The investigator must exercise professional judgment when using published procedures to collect data in order to ensure that the procedures are applicable to field conditions under study. Often, sampling methods for one type of chemical compound cannot be used for others; volatile organic compounds should not be sampled with a centrifugal pump, for instance, although this sampling method would be acceptable for certain inorganic chemicals. Accuracy of analytical methods may suffer as a result of matrix effects or the presence of a second phase (particulate matter or gases). The investigation should include in the experimental design a rigorous test of methodology that would include blanks, replicate samples, and spiked samples.

Data Analysis

Three categories of statistical inference procedures will be discussed: (1) parametric analysis of variance (ANOVA), (2) nonparametric procedures for comparing several groups, and (3) contingency-table analysis for comparing frequencies. Situations in which each might be used will be illustrated.

Two terms require definition at this point, the significance level (α) and the power ($1-\beta$) of a statistical test procedure. The significance level or error risk is the probability of falsely detecting differences due to some factor when there actually are none. The power is the probability of detecting differences when they do exist.

Data Analysis Using ANOVA

After collecting data from the sampling network, ANOVA can again be used to interpret whether hypothesized influences on water chemistry are significant. Suppose TCE concentrations are to be compared in four subgroups of an aquifer, each subgroup receiving recharge from a different land use. Are TCE concentrations in these four subgroups the same, or different? This can be answered by a "one-factor" ANOVA; the influence of only one factor (subgroup) is investigated. A null hypothesis is proposed that the mean concentration, μ, in each subgroup is identical

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

This is tested against the alternative

H_1: at least one μ_i is unequal.

The ANOVA procedure is fully explained by Sokal and Rohlf (1969) and by Kleinbaum and Kupper (1978). Briefly, the estimated total variance and the estimated within-groups variance (within land-use region) are calculated:

$$\text{Total} = \frac{\sum\sum^{an}(y - \overline{\overline{y}})^2}{a\,(n - 1)}$$

$$\text{Variance Within Groups} = \frac{\sum^{a}\sum^{n}(y - \overline{y})^2}{a\,(n - 1)}$$

where

a is the number of land-use regions (groups),
n is the number of data in each group,
$\underline{y}$ is the concentrations of a sample,
y is the average concentration of a group, and
y is the average concentration over all groups.

An F statistic is calculated as the ratio:

$$\frac{\text{(total variance)-(within-group variance)}}{\text{(within-group variance)}}$$

For large F values, the variance within groups is small in comparison to the total variance (as in fig. 6a), and differences in group mean concentrations can be detected. For example, if the total variance = 100 and the within group variance = 20, then F = 4.0. The degrees of freedom, the denominators of the above variance equations, are used to compare this calculated F value to that in an F table. For a significance level of 0.05, and a = 4 while n = 8,

$$\begin{aligned} F_{\alpha}, df_{num}, df_{denom} &= F_{.05}, 3, 28 \\ &= 2.29 \end{aligned}$$

When the calculated F value exceeds the tabled value, as it does above, the (total-within) variance is large enough to detect a group effect at that significant level. H_0 is therefore rejected, and at least one group mean concentration differs from the others. Multiple comparison procedures next determine which group means are different. Stoline (1981) has reviewed the various types of multiple comparisons, and argues for using Tukey's honest significant difference test. These tests use the within-group variance to calculate the minimum difference in mean concentration between groups required to consider the groups significantly different.

The one-factor ANOVA may be expanded in two ways. Two or more influencing factors may be simultaneously investigated with what is known as a factorial ANOVA. A two-factor ANOVA is diagrammed in figure 7. The two suspected influencing factors are A and B. Each combination of A and B defines a subgroup (or treatment, in ANOVA terminology). One example would be if contaminant concentrations in three aquifers (factor B) are to be compared, and wells of two casing types (factor A) are sampled. The hypothesis to be tested is that all subgroup means are identical. If they are not, this difference may be due to different aquifers, different casing type, or to both factors. The factorial ANOVA design is appropriate when one factor is not a subset of the other, or when each B (or A) level is identical across the other factor. In the above example, each aquifer (b_1, b_2, b_3) is the same for the two casing types.

The second type of expanded ANOVA is the nested design. Nested ANOVA's utilize subsets of random repetitions to (1) filter out or (2) measure the background variability ("noise") that may be obscuring any effects due to the factor of interest. For example, again suppose areas of different land use are to be compared in terms of shallow, aquifer water quality. Randomly selected subgroups (wells) within each area are sampled, which allow the variation between wells ("noise") to be evaluated separately from the land-use effect between areas. If several samples are taken from each well, there are "sub-subgroups" (samples) within each subgroup (wells). More of the background variation, the differences between samples from the same well, is now accounted for; the land-use effect (if present) will stand out more clearly.

		Factor A	
		a_1	a_2
Factor B	b_1	x_{11}	x_{12}
	b_2	x_{21}	x_{22}
	b_3	x_{31}	x_{32}

Figure 7. Diagram of a two-factor analysis of variance

During reconnaissance, the nested ANOVA aids in experimental design. Estimates of the variance for each level of subgroup (wells within land-use areas, samples within wells) allow estimates to be made of the optimal allocation of samples.

Figure 8 diagrams a nested ANOVA. The main factor of interest (A) is represented by two groups, a_1 and a_2 (urban and rural land use, for example). Within each group are n randomly chosen subgroups (b's), such as wells within each area. For each of the n subgroups are two replicate measurements (r_1 and r_2). The nested design is evident because the b's are random--b_1 of a_1 has no relationship to b_1 of a_2.

The F tests associated with nested ANOVA differ slightly from those of the factorial design (Sokal and Rohlf, 1969; Kleinbaum and Kupper, 1978). Following the ANOVA significance tests for the main effect, multiple comparison procedures, again, are used to determine which groups are, in fact, different from others.

ANOVA is a flexible and powerful procedure for simultaneously evaluating the significance of multiple suspected influences. It is the basic tool for evaluating discontinuous-group influences such as land use. Hazardous-waste investigations should rely on ANOVA rather than on qualitative, informal interpretation schemes.

Nonparametric Comparisons Among Several Groups

ANOVA is a classical parametric procedure. Nonparametric (or "distribution free") procedures do not assume a specific distribution of data in order to make inferences about the data. When data actually are normally distributed, nonparametric tests will be only 85 to 95 percent as powerful as parametric procedures. However, when data are not normal, the nonparametric tests will be more powerful, perhaps many hundreds of times more powerful. Nonparametric tests often use the ranks of the data (lowest = rank 1, up to highest = rank n) rather than the data themselves. If the data within a subgroup (a_i b_j combination) are normally distributed, then ANOVA is the most powerful procedure for detecting differences between subgroups. When the data do not closely fit a normal curve, however, ANOVA may not be able to detect even large differences between subgroups; it has low power. Nonparametric procedures are much more powerful in this situation.

In general, chemical-quality data rarely are close to a normal distribution. They cannot have negative values, and usually are at low concentrations with less-frequent higher values. This produces a distribution shaped not like the normal (fig. 9a), but more like the lognormal (fig. 9b). By taking logarithms of the data, a shape more like 9a might be derived; parametric procedures such as ANOVA can then be used on the transformed data. Tests for the normality of data distributions are discussed in Bhattacharyya and Johnson (1977) and many other statistics texts, and should be performed prior to an ANOVA.

A problem arises when the data do not approximate either a normal or lognormal pattern. This is often true for trace metals or organics, which may have a large percentage of essentially zero values. A second problem

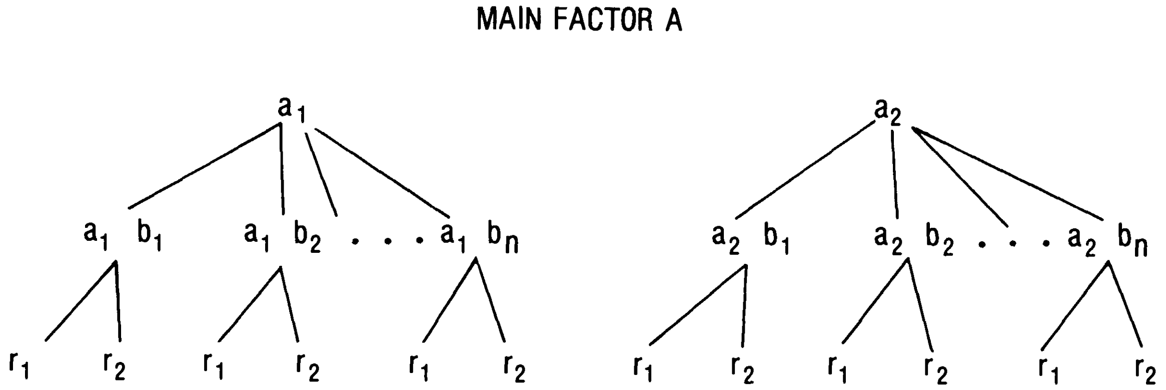


Figure 8. Diagram of a nested analysis of variance

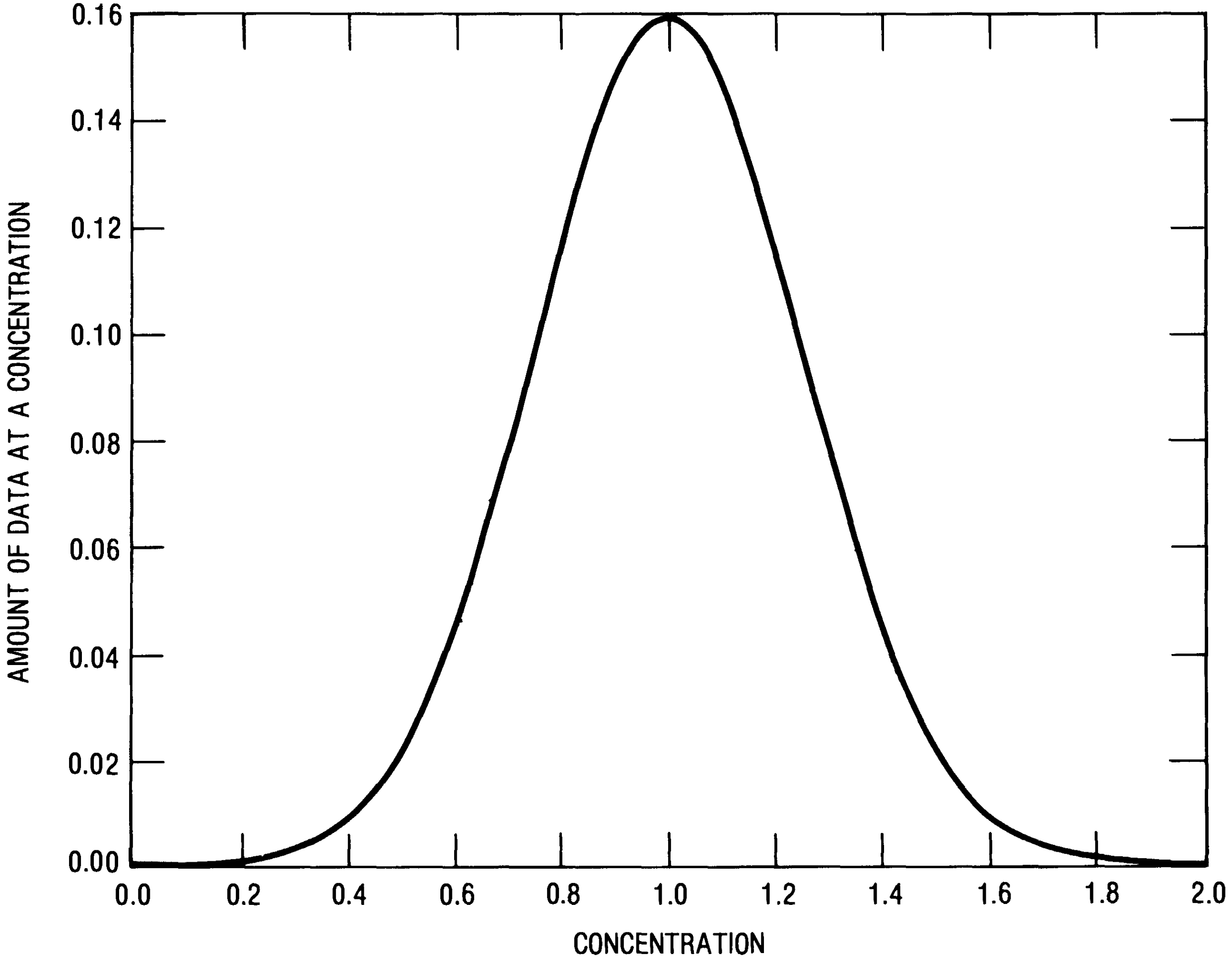


Figure 9a. Graph showing normal distribution

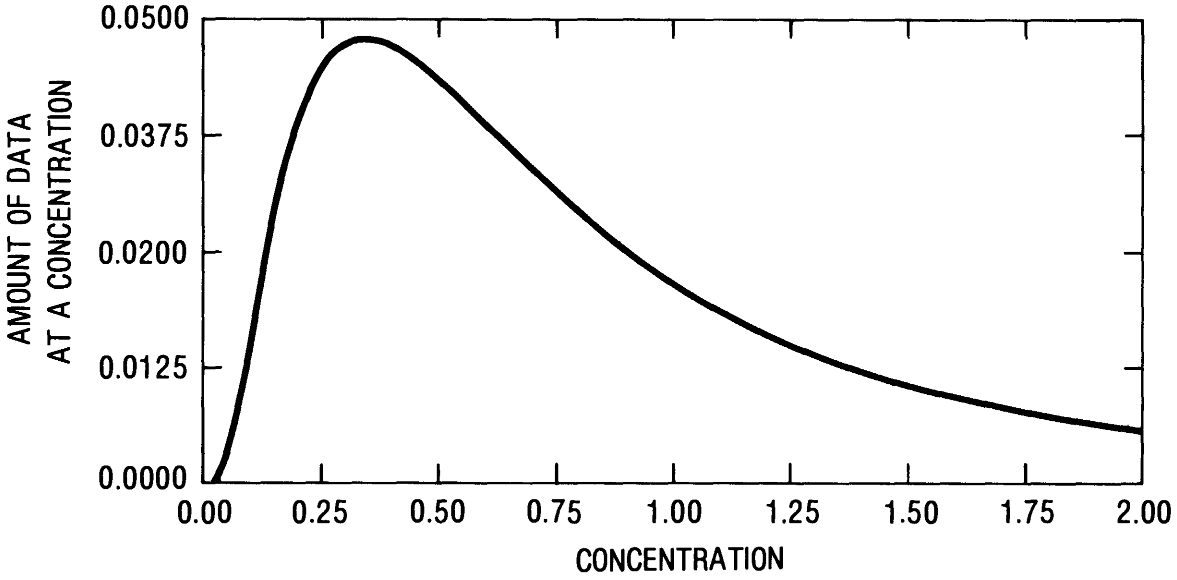


Figure 9b. Graph showing lognormal distribution

occurs when data from some of the subgroups appear normal, but others do not. No transformation (such as logarithms) will produce normal distributions for all subgroups in this case. Nonparametric procedures are again useful in these situations.

The nonparametric Kruskal-Wallis (KW) test is analagous to the one-factor ANOVA (Conover, 1980). KW tests the hypothesis that all subgroups produce similar data against the alternative that some subgroups tend to produce different values. Consider the following example in which the lead concentrations associated with three land use groups are compared.

	Group 1					Group 2					Group 3				
Concentration	10	12	8	7	16	15	18	20	23	17	9	4	12	10	14
Rank	5.5	7.5	3	2	11	10	13	14	15	12	4	1	7.5	5.5	9
Summed Rank	29					64					27				

The concentrations are ranked together from lowest to highest, and tied observations are given the average of the two possible ranks. The sum of all ranks for each group is then calculated, and from these is computed the Kruskal-Wallis T statistic (Conover, 1980), a measure of the heterogeneity among groups.

$$T = \frac{12}{N\,(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3\,(N+1)$$

where R_i is the sum of ranks for group i

k is the number of groups compared, and

and N is the total number of samples, $\Sigma\, n_i$.

For our example,

$$T = \frac{12}{15(16)} \left[\frac{29^2}{5} + \frac{64^2}{5} + \frac{27^2}{5} \right] - 3\,(16)$$

$$= 8.66$$

A slight correction (ignored here) is necessary when tied observations occur. This T statistic is then compared to standard tables; if it is higher in absolute value than the tabled statistic, the null hypothesis that all groups have similar values is rejected. A multiple comparison procedure is also given by Conover (1980) to determine which groups differ. Nonparametric versions of correlation, regression, t-tests, and trend tests are also frequently useful.

Nonparametric analogues to factorial and nested ANOVA can be obtained by ranking all data from lowest to highest, performing ANOVA on the ranks rather than on the data themselves (Conover and Iman, 1981). An application of this procedure to water-quality data was done by Helsel (1983). Multiple comparisons also may be performed on the ranks.

Contingency-Table Analysis of Frequency Data

Trace organics or metals analyses often result in values discerned only as "below detection limit." Can these values provide information on suspected influences of interest? In most cases, they can.

For data sets with few "not detecteds" (ND), perhaps less than 20 percent, Kruskal-Wallis or similar rank-transform procedures again can be used. All NDs are assigned the same lowest rank and the procedure is performed as usual. The ability to analyze data sets containing NDs (without substituting arbitrarily guessed values) is another strength of nonparametric procedures.

Larger percentages of NDs can be analyzed using the frequency of exceeding the detection limit rather than on the concentration itself. Such methods are called categorical or nominal methods. For example, concentrations could be categorized into ND, low, or high; this would, of course, require the choice of a suitable low/high boundary. Contingency tables would then be used to compare the subgroups' percentage of observations falling into each category. Conover (1980), Bhattacharyya and Johnson (1977), and other basic texts describe contingency-table procedures. Should the frequency of detection be small, only two categories (ND and detected) probably would be used.

As a simple example, suppose wells tapping two aquifers are compared. The tested hypothesis is that the probability of exceeding the detection limit in aquifer A (P_A) equals or exceeds that for aquifer B (P_B). The alternative is that P_B is greater (due to some contamination). This test is "one-sided"; only when B exceeds A in detections is contamination proven. The number of detections and NDs are recorded in a 2 x 2 contingency table:

	ND	Detected	Total
Aquifer A	$x_{a1} = 35$	$x_{a2} = 6$	$n_a = 41$
B	$x_{b1} = 21$	$x_{b2} = 25$	$n_b = 46$
Total	$C_1 = 56$	$C_2 = 31$	$N = 87$

where $n_a = x_{a1} + x_{a2}$

$n_b = x_{b1} + x_{b2}$

$c_1 = x_{a1} + x_{b1}$

$c_2 = x_{a2} + x_{b2}$

and $N = n_a + n_b = c_1 + c_2$.

From Conover (1980, page 145) the test statistic is

$$T = \frac{\sqrt{N}\,(x_{a1}\,x_{b2} - x_{a2}\,x_{b1})}{\sqrt{n_a\,n_b\,c_1\,c_2}} = \frac{\sqrt{87}\,(35(25)-6(21))}{\sqrt{(41)(46)(56)(31)}} = 3.86$$

We would reject the hypothesis that $P_B \leq P_A$ in favor of $P_B > P_A$ if T is large, or

$$T > z_{1-\alpha},$$

the (1-α) quantile of the standard normal distribution. For $\alpha = .05$, $z_{.95} = 1.64$. We would therefore reject $P_B \leq P_A$, and conclude that aquifer B is contaminated in comparison to aquifer A.

Contingency tables for more than two subgroups and more than two data categories are solved similarly. Where few NDs are present, contingency table methods will be less powerful than KW or other rank-transform methods. Contingency tables also will not give estimates of "typical" concentrations, but only of the observed frequency of falling into some category or range of concentrations. Experimental design procedures for contingency-table analysis are available, but some parameter estimates are required. Again, a pilot study may provide these. An estimate of one of the frequencies is necessary. Often the percentage of detections for a "clean" (control) area is available (P_A). Also required is the minimum difference in detection frequency for the other area that should be discernible. Finally, the significance level (α) and power (1-β) of the test procedure need to be specified.

For example, suppose that previous reconnaissance sampling in a clean area has indicated that a given compound is detected 10 percent of the time ($P_A = 0.10$). In the study area, we wish to declare detection frequencies of 35 percent or more to be significantly different ($P_B = 0.35$). We set $\alpha = 0.05$ whereas power = 0.80.

Natrella (1963) presents the following procedure for calculating the sample size n, which is the size required for each of the groups:

$$d = (\arcsin\sqrt{P_B} - \arcsin\sqrt{P_A})\ 0.03491$$

$$\text{and } n = \frac{2\,(z_{1-\alpha} + z_{1-\beta})^2}{d^2}$$

where the z's are, again, from the standard normal table. For the above example

$$d = (\arcsin\sqrt{.35} - \arcsin\sqrt{.10})\ 0.03491$$

$$= (36.27 - 18.43)\ (.03491)$$

$$= 0.623$$

$$n = \frac{2\,(z_{.95} + z_{.80})^2}{d^2}$$

$$= \frac{2\,(1.645 + 0.84)^2}{(0.623)^2}$$

$$= 31.86$$

Thus, 32 samples should be taken from each area.

For more than two subgroups, the procedure is more arduous. One method of approximation is to compare all subgroups with the same control group having P_1 detections.

Verification

Verification provides a test of the transfer value of the water-quality information collected in the 14 study areas to other areas of similar land use in the same hydrologic and climatic setting. Each of the 14 studies should reserve funds sufficient to conduct additional, limited sampling in secondary study areas that are completely separate from the primary study areas. These secondary study areas should be selected during reconnaissance, and should be similar to the primary study areas in land use, hydrogeology, and climate.

Sampling in the secondary areas can be concurrent with that of the primary study. Ten percent or less of all samples would be collected from the secondary areas. Following the conclusions concerning land use and other factors' influences in the primary study area, these influences should be verified using the secondary data.

Successful verification of organic and trace-metal concentrations would suggest that the factors affecting ground-water chemistry in the primary study area are similar to those of the secondary study area. Therefore, extrapolation of the findings to larger geohydrologic and climatic regions would be more feasible. Unsuccessful verification would indicate that different factors are controlling ground-water chemistry in the primary and secondary study areas. Additional study would be required before extrapolation of results beyond the primary study area could be accomplished.

APPLICATION TO A NATIONAL APPRAISAL

The 14 individual studies will provide an understanding of the variation in concentrations of trace metals and organic compounds in selected parts of the Nation's ground-water reserves, and of the effects of land use within a variety of hydrogeologic and climatic environments. This information also will allow comparisons between studies for similar land uses. Thus, any differences in patterns of contamination between hydrogeologic regions or differences in climate within those regions may be discussed. The verification procedure will demonstrate the transferability of inferences made about the effect of land use on ground-water quality within regions. In those studies where transferability cannot be demonstrated, additional studies would be necessary to further understand which factors are affecting ground-water quality. In those studies where causality and transferability is demonstrated, projections of ground-water quality to wider parts of the Nation may be possible. To complete a truly national appraisal of toxics in the Nation's ground waters, an expansion of the 14 studies currently begun to other parts of the Nation will be required.

SUMMARY

The purpose of this report is to provide a method to study parts of the regional ground-water systems of the Nation in order to provide information on ground-water chemistry--with emphasis on organic substances and trace metals--and to explain water quality in terms of local hydrology and human activities. The individual studies are regional in scope and statistical in approach in order to accommodate the large number and variety of contaminant sources and the complex and heterogeneous nature of ground-water systems. The report describes how results from studies of individual areas can be used in a national appraisal of ground-water quality.

The procedure includes five major steps: reconnaissance and development of the conceptual model, experimental design, data collection, data analysis, and verification. Reconnaissance and development of the conceptual model involve assembling all relevant chemical, earth-science and cultural information so that a conceptual model can be proposed concerning the predominant factors affecting ground-water quality. Experimental design provides a sampling plan to representatively measure aquifer water quality in order to test the conceptual model. Data collection should follow previously published guidelines for toxic wastes. Analysis of variance (ANOVA) methods are used to test the hypotheses concerning the influences on ground-water chemistry. Examples are given that demonstrate the use of ANOVA for parametric and nonparametric comparisons of data. Verification provides a test of the transfer value of the information from land-use subareas of an individual study area to similar land-use areas in independent but similar hydrologic and climatic settings. Successful verification will allow the investigator the opportunity to extrapolate information from the intensively studied areas to other areas.

The 14 individual studies will provide an understanding of the variation in the concentration of trace metals and organic compounds in selected parts of the Nation's ground-water reserves, and of the effects of land use on ground-water quality. In those studies where causality and transferability can be demonstrated, projections of ground-water quality for wider parts of the Nation should be possible.

REFERENCES

Barnett, P. R., and Mallory, E. C., 1971, Determination of minor elements in water by emission spectroscopy: U.S. Geological Survey Techniques of Water-Resources Investigations, Chap. 5, Book A-2, 31 p.

Bhattacharyya, G. K., and Johnson, R. A., 1977, Statistical concepts and methods: New York, John Wiley, 639 p.

Claassen, H. C., 1982, Guidelines and techniques for obtaining water samples that accurately represent the water chemistry of an aquifer, U.S. Geological Survey Open-File Report 82-1024, 49 p.

Cohen, S. Z., Creeger, S. M., Carsel, R. F., and Enfield, C. G., 1984, Potential for pesticide contamination of ground water resulting from agricultural uses: Treatment and disposal of pesticide wastes, R. F. Krueger and J. N. Seiver, eds., American Chemical Society Symposium Series 259, p. 297-326.

Conover, W. J., 1980, Practical nonparametric statistics, 2nd edition: New York, John Wiley, 493 p.

Conover, W. J., and Iman, R. L., 1981, Rank transformations as a bridge between parametric and nonparametric statistics: American Statistician, v. 35, p. 124-129.

Council on Environmental Quality, 1981, Contamination of ground water by toxic organic chemicals: Washington, U.S. Government Printing Office, 84 p.

Dunlap, W. J., Cosby, R. L., and Freiberger, John, 1981, Manual of ground-water sampling procedures: National Water Well Association, U.S. Environmental Protection Agency series, 93 p.

Franke, O. L., and Cohen, Philip, 1972, Regional rates of ground-water movement on Long Island, New York: U.S. Geological Survey Professional Paper 800-C, p. C271-C277.

Franke, O. L., and McClymonds, N. E., 1972, Summary of the hydrologic situation on Long Island, New York, as a guide to water-management alternatives: U.S. Geological Survey Professional Paper 627-F, 57 p.

Friedman, L. C., and Erdmann, D. E., 1982, Quality-assurance practices for the chemical and biological analyses of water and fluvial sediments: U.S. Geological Survey Techniques of Water-Resources Investigations, Chap. 5, Book A-6, 181 p.

Helsel, D. R., 1983, Mine drainage and rock type influences on Eastern Ohio streamwater quality: Water Resources Bull., v. 19, p. 881-887.

Kleinbaum, D. G., and Kupper, L. L., 1978, Applied regression analysis and other multivariate methods: North Scituate, Mass., Duxbury Press, 556 p.

Miller, D. W., 1980, Waste disposal effects on ground water: Berkeley, Premier Press, 512 p.

Natrella, M. G., 1963, Experimental Statistics: U.S. Dept. of Commerce, National Bureau of Standards Handbook 91, 450 p.

Ragone, S. E., Katz, B. G., Kimmel, G. E., and Lidner, J. B., 1981, Nitrogen in ground water from sewered and unsewered areas, Nassau County, Long Island, New York: U.S. Geological Survey Water-Resources Investigations 80-21, 64 p.

Scalf, M. R., McNabb, J. F., Dunlap, W. J., Cosby, R. L., and Freiberger, John, 1981, Manual of ground-water sampling procedures: National Water Well Association, U.S. Environmental Protection Agency series, 93 p.

Shaner, J. R., 1982, How non-detection can cost you--The tank leak mess: National Petroleum News, v. 74, p. 36-40.

Skougstad, M. W., Fishman, M.J., Friedman, LL.C., Erdmann, D. E., and Duncan, S.S., eds, 1979, Methods for determination of inorganic substances in water and fluvial sediments: U.S. Geological Survey Techniques of Water-Resources Investigations, Chap. 5, Book A-1, 626 p.

Sokal, R. R., and Rohlf, F. J., 1969, Biometrry: San Francisco, W. H. Freeman, 776 p.

Stoline, M. R., 1981, The status of multiple comparisons--simultaneous inference of all pairwise comparisons in one-way ANOVA designs: American Statistician, v. 35, p. 134-141.

Wershaw, R. L., Fishman, M. J., Grabbe, R. R., and Lowe, L. E. eds., 1983, Methods for the determination of organic substances in water and fluvial sediments: U.S. Geological Survey Open-File Report 82-1004, 173 p.