A Hydrogeologic Database for Ground-Water Modeling

by Charles J. Newell[a], Loren P. Hopkins[b], and Philip B. Bedient[c]

Abstract

A new hydrogeologic database, the HGDB, was developed from a national survey of National Water Well Association (NWWA) members. The database contains general hydrogeologic information from 400 field site investigations across the country and detailed statistical summaries of five ground-water parameters: hydraulic conductivity, seepage velocity, hydraulic gradient, saturated thickness, and depth to top of aquifer. The HGDB was developed to verify and expand statistical distributions used in a Monte Carlo ground-water model developed by EPA for land disposal regulation (*Federal Register,* 1986, 1988).

The database structure is a unique application of the aquifer classification method used in the NWWA's DRASTIC system (Aller et al., 1987). Respondents were asked to classify their aquifers as one of 111 different DRASTIC hydrogeologic settings, and 12 groupings of settings were analyzed to produce statistical distributions of hydrogeologic data based on site geology and geomorphology. Three examples of the hydrogeologic groupings are coastal beaches; alluvial basins, valleys and fans; and outwash settings.

The HGDB can be used for several purposes. First, the HGDB results indicate that the EPA's distributions of seepage velocity and hydraulic conductivity used in the land disposal model are sound. These are the most important hydrogeologic parameters in the model. The HGDB goes a step further, and provides a set of statistical distributions that can be used to make the land disposal regulations more site-specific than the national approach now being used. Finally, the HGDB data can be used for general site characterization and for educational purposes. The database is available as a detailed written report and spreadsheet file from the American Petroleum Institute, and is contained in a graphical computerized decision support system for ground-water modeling called OASIS. The HGDB serves as a framework for organizing hydrogeologic information from different site investigations and can be expanded easily beyond the 400 sites now in the database.

Introduction

A statistical ground-water model, the Environmental Protection Agency's Composite Landfill Model (EPACML), is being proposed by EPA as a screening tool for the land disposal of hazardous wastes (*Federal Register,* 1986, 1988). EPACML uses a Monte Carlo approach, where statistical distributions of aquifer parameters are used to represent the range of aquifer conditions across the country. The proposed EPA regulations make Monte Carlo ground-water modeling and the associated statistics of hydrogeologic parameters important components of the regulatory process. A new hydrogeologic database, the HGDB, was developed to provide an independent source of data for Monte Carlo modeling and for testing the accuracy of the statistical distributions used in the model.

The HGDB takes advantage of the large number of hydrogeologic investigations that have been conducted at waste sites but have not been reported in the technical literature. An extensive technical survey of ground-water professionals was conducted with support from the American Petroleum Institute and assistance from the National Water Well Association (NWWA). The database was structured by adapting hydrogeologic settings from the DRASTIC system, a graphical aquifer vulnerability index developed by the NWWA (Aller et al., 1987). A user selects the hydrogeologic setting that best matches a site or area of interest to access the data of interest.

The HGDB project has three main objectives:

1. Develop a hydrogeologic database based on actual data collected from field investigations.
2. Apply "hydrogeologic settings" from the EPA's DRASTIC system (Aller et al., 1987) to structure the database.
3. Produce a new hydrogeologic database that can be used for:

---

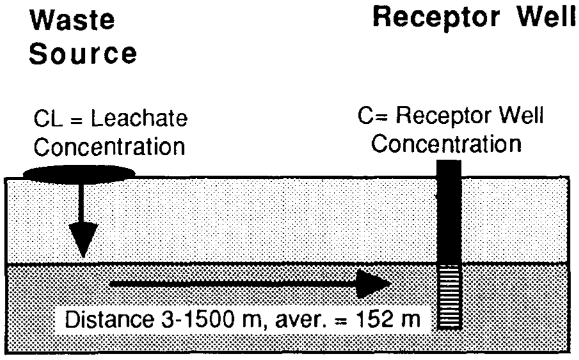
[a] Groundwater Services, Inc., 2503 South Blvd., Houston, Texas 77098.

[b] OHM Corporation, 1 Cielo Center, Suite 240, Austin, Texas 78746.

[c] National Center for Ground Water Research, Department of Environmental Science and Engineering, Rice University, Houston, Texas 77251.

Received April 1989, revised October 1989 and May 1990, accepted May 1990.

Discussion open until March 1, 1991.



General Ground Water Screening Procedure:

1. EPACML Model used to calculate concentration at receptor well (C)
2. Waste is Acceptable for land disposal if concentration at the Receptor Well is below reference dose

Fig. 1. Approach used in proposed land disposal regulations.

- Checking the national distributions of aquifer parameters used in EPACML model;
- More site-specific statistical modeling of waste sites than the existing EPACML approach; and
- General site characterization by hydrogeologists.

Land Disposal Regulation Using the EPACML Model

The 1984 Resource Conservation and Recovery Act (RCRA) directed the Environmental Protection Agency to regulate the land disposal of hazardous wastes by considering the uncertainties involved with land disposal and the need to protect human health. The EPA's proposed approach uses the EPACML model to back-calculate the maximum allowable source concentration of a waste based on a health-based standard, the reference dose (Figure 1). The EPACML approach is designed to be used as a general screening tool for regulating land disposal of wastes, and a petitioning process to account for site-specific cases is to be established sometime in the future.

Modeling Assumptions

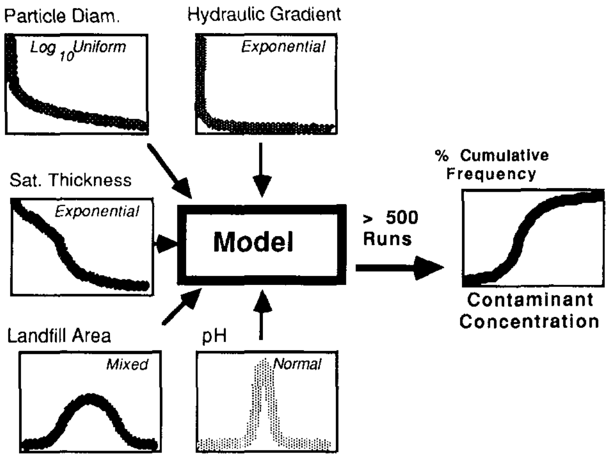
The original version of the EPACML model consisted of a three-dimensional semianalytical model of the advection-dispersion process in the saturated zone (U.S. EPA, 1986). The approach was based on a semianalytic advection dispersion model with a Gaussian distribution of the source concentration underneath a source (such as a landfill) perpendicular to ground-water flow (Huyakorn et al., 1987). In other words, if ground water is flowing in the X direction, then the source concentration in the aquifer is highest underneath the middle of the site, and tails off to zero concentration at the edges of the site in the Y direction. Some of the other key assumptions made in the model are that all engineered facilities will fail (so no consideration of landfill liners was included), aquifers are homogeneous with uniform flow, and that steady-state conditions should be assumed. Many of the modeling assumptions were driven by the conservative philosophy of the RCRA legislation, leading to more restrictive land disposal regulation from this modeling approach.

The original version of the model was modified to improve the screening methodology, and the revised model, EPACML, was introduced in late 1988 (*Federal Register,* 1988; U.S. EPA, 1988a). EPACML incorporated a one-dimensional unsaturated zone model (which has little effect under steady-state conditions for most cases) and a variable location of the receptor well (the measuring point) that was assigned randomly in three dimensions within the general confines of the contaminant plume. A new method for calculating dispersivity was also incorporated into the model (Gelhar, U.S. EPA, 1988a).

Monte Carlo Ground-Water Modeling

The EPACML model uses a Monte Carlo type simulation to account for the differences in aquifer parameters across the country (U.S. EPA, 1988a). To perform a Monte Carlo analysis, an input value for a parameter (such as hydraulic gradient) is selected at random based on the expected frequency described in one of several statistical distributions. The selection process is repeated for each of the input parameters. The model is run to yield a concentration in the aquifer at the point of interest. This process is repeated enough times (several hundred times, for example) until a collection of output concentrations form a statistical frequency distribution. Instead of obtaining a single output value from the model, the user obtains a frequency distribution that can be used to quantify the probability that 85% of the time the output concentration is less than X mg/l. Figure 2 illustrates the Monte Carlo approach to ground-water modeling, where input data distributions are used to generate a frequency curve of concentration at a distribution of various receptor well locations in the aquifer.

The EPA plans to use a point on the concentration



Other Parameters: Temperature, Leakage Rate, Dispersivity, Leachate Penetration Depth, Fraction Organic Carbon

Hydraulic Conductivity, Porosity, and Seepage Velocity Derived from Particle Diameter and other parameters

Fig. 2. Monte Carlo modeling approach used in the EPACML model.

frequency curve (probably the 85th percentile) to compare against reference doses for different chemicals, assuming a source concentration of 1.0 mg/l in EPACML. By using the frequency curve and back-calculating from the reference dose, the maximum allowable source concentration of a chemical can be determined. Future land disposal operations will be required to perform a leach test to determine if wastes will exceed the maximum allowable source concentrations.

Input Distributions to EPACML

To perform Monte Carlo modeling, the following information is required for each parameter: the type of statistical distribution, the appropriate summary statistics such as the mean and standard deviation, and the ranges of possible data. The EPACML model uses the data distributions listed in Figure 2 and functional relationships to derive parameters defined as dependent by the EPA. For example, the EPACML developers used particle grain size as an independent "seed" distribution for the Kozeny-Carmen equation (Bear, 1979) to calculate hydraulic conductivity because of the dependency between grain size, porosity, and hydraulic conductivity (U.S. EPA, 1985, 1988a).

As one might expect, the statistical distributions used in the EPACML have high standard deviations and wide ranges. The range of ground-water velocities, for instance, is between 0.01 and 9250 m/year. Although the range appears to be extremely large, it meets the general criteria of the EPA: all possible aquifer types in the country are considered, and the analysis is conservative. The wide range of ground-water velocities and the other input parameters combine to create a wide distribution of output concentrations.

The HGDB project was conceived as an independent check of the distributions and the functional relationships used in the EPACML model, and to provide a mechanism for incorporating site-specific factors into the Monte Carlo modeling process. Original hydrogeologic data were collected and analyzed to determine the type of distribution and calculate the summary statistics that are required for Monte Carlo procedures.

The Hydrogeologic Database: Data Collection

The data collection for the project was based on the assumption that a large amount of data gathered from hydrogeologic field investigations performed across the country are not available in the technical literature. The information is comprised of aquifer tests, soil samples, geologic logs, and water-quality sampling data obtained from intensive field programs at waste sites and other ground-water projects. Many hydrogeologic studies are never described in the technical literature, and only a portion of the data are entered into regulatory agency databases and other types of databases.

The American Petroleum Institute, the National Water Well Association (NWWA), and Rice University designed a technical survey for ground-water professionals who might have hydrogeologic data from field investigations. The target group was the 8700 members of the Association of Ground Water Scientists and Engineers (AGWSE), the technical division of the National Water Well Association. Table 1 lists the information requested in the questionnaire sent to a targeted group of AGWSE members. From this group, 400 useable questionnaires were returned to the project. The distribution of the 400 sites can be seen in Figure 3 which suggests that there were no major problems with the distribution of sites nationally. To maintain the confidentiality of the respondents, the site locations on the map are approximate.

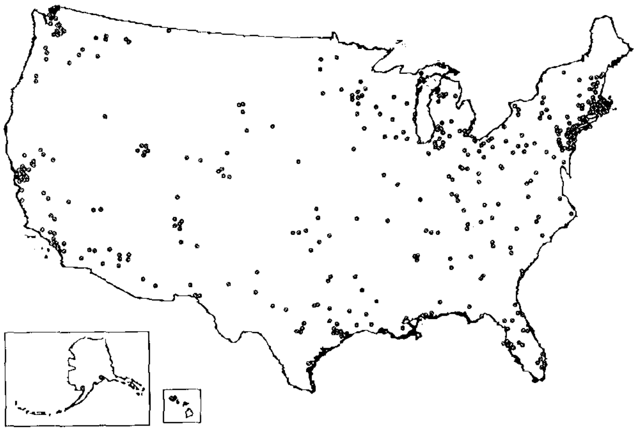


Fig. 3. Approximate locations of sites described by questionnaires.

Organization of the Database

One unique aspect of the HGDB is that the database is structured using hydrogeologic settings from the National Water Well Association's DRASTIC system. DRASTIC is a standardized system for evaluating ground-water pollution potential using hydrogeologic settings (Aller et al., 1987). DRASTIC consists of two main components: a series of 111 mappable units called hydrogeologic settings that are representative of different aquifer types in the United States, and a system for the relative ranking of key hydrogeologic parameters. Both components are designed to be used together to provide a DRASTIC Index, a number indicative of the ground-water pollution vulnerability of a particular aquifer. While the 111 hydrogeologic settings from DRASTIC were used for the HGDB, the ranking system was not applied.

Description of DRASTIC Hydrogeologic Settings

The hydrogeologic settings in the DRASTIC system were developed by combining information from two sources. Heath (1984) divided the United States into separate ground-water regions based on the general occurrence and availability of ground water. Figure 4 shows 13 of these general regions used for the DRASTIC system. The regional map is the top level of classification, and leads to a series of hydrogeologic settings that have been developed for each region. Each region has from 4 to 17 different settings that describe the different types of aquifers found in the region. These settings were developed by a committee of hydrogeologic experts who designed the DRASTIC system

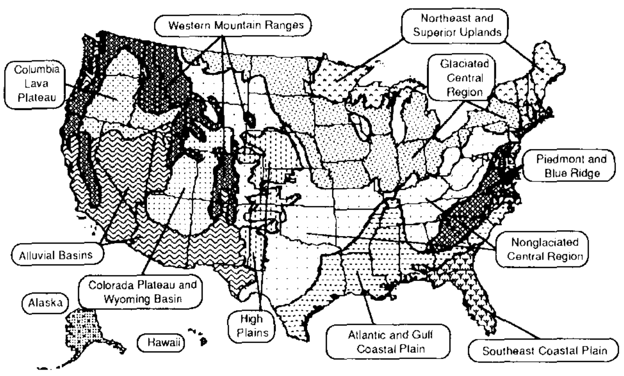
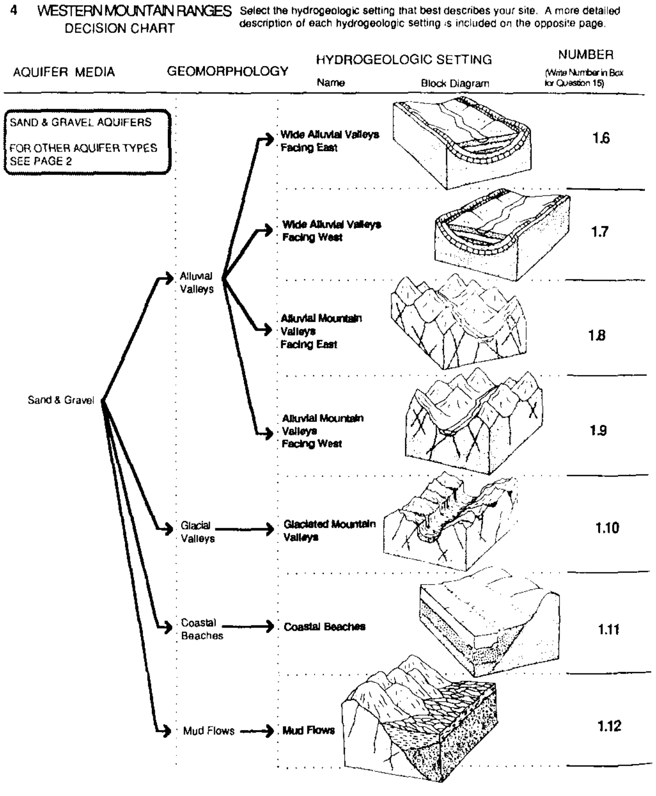


Fig. 4. Ground-water regions of the United States (from Heath, 1984; Aller et al., 1987).

using a consensus building decision-making process (Aller et al., 1987).

Application of Hydrogeologic Settings to the HGDB

The respondents in the survey were asked to classify each of their sites according to the DRASTIC setting that best matched the site using a series of decision charts. After selecting the appropriate ground-water region (Figure 4), the user was asked to use a decision chart (Figure 5) to determine the best setting, and report the information in the questionnaire. Because an accurate classification of each questionnaire was crucial, the reported setting was checked against the rest of the information in the questionnaire, such as the typical well log, saturated thickness, depth to water, and hydraulic conductivity.


4 WESTERN MOUNTAIN RANGES DECISION CHART

Select the hydrogeologic setting that best describes your site. A more detailed description of each hydrogeologic setting is included on the opposite page.

Fig. 5. Sample page from hydrogeologic setting decision chart.

There were not enough data to develop statistical distributions for each of the 111 DRASTIC hydrogeologic settings, and a secondary grouping of the data into 12 hydrogeologic environments was performed (Appendix 1). The environments are based on the natural repetition of settings within the different geographic region in the DRASTIC system (Figure 6). For example, coastal beach settings from each geographic region were grouped together in a hydrogeologic environment called coastal beaches. In some cases, grouping the settings into hydrogeologic environments still did not provide enough observation points to be statistically analyzed (minimum of 20 data points). For those cases, the next larger group, sand and gravel, bedded sedimentary rocks, solution limestone, or metamorphic igneous rocks, was used as that setting's distribution. Once the hydrogeologic environment approach was established, box plots and statistical summary tables of each parameter and environment were developed (see Results section).

Quantitative Analysis Methods

Five of the questions asked the respondent to provide quantitative responses to aquifer conditions, and these data were analyzed in detail: K = Hydraulic Conductivity (cm/sec); V = Seepage Velocity (ft/day); G = Hydraulic Gradient (ft/ft); ST = Saturated Thickness of Aquifer (ft); and DW = Depth to Top of Aquifer (ft).

The entire collection of responses was used for each parameter except in the case of hydraulic conductivity. The hydraulic conductivity dataset was modified to exclude all of the reported values that were derived solely from engineering judgment and literature values only; data obtained from pump tests, slug tests, laboratory permeability tests, or from a combination of these tests were used.

Seepage velocities, on the other hand, were obtained from the entire dataset. The seepage velocity is the velocity of a conservative tracer with no retardation in the ground water, and is not a Darcy velocity or a pressure impulse velocity. Velocity calculations are often more subjective and based on a variety of factors (such as the length of the plume). Thus, the hydraulic conductivity and seepage veloc-

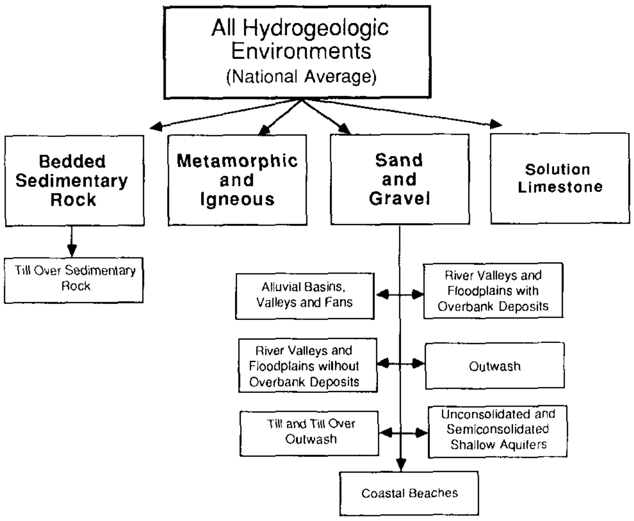


Fig. 6. Hydrogeologic environments used in the HGDB.

ity datasets represent different assumptions: hydraulic conductivity data are measured (pump tests, slug tests and/or lab tests) while the seepage velocity data include some responses (approximately 25% of the total) that were based to some degree on engineering judgment or literature values. Hydraulic conductivity and seepage velocity are reported in units of cm/sec and ft/day, respectively, because these were the most common units reported by the respondents in the questionnaire.

Hydraulic gradient is the slope of the water table or potentiometric surface of an aquifer, and saturated thickness is the thickness of the aquifer. The depth to top of aquifer represents the depth to water in an unconfined aquifer in most cases, and the depth from the surface to the top of the saturated zone for a confined aquifer.

Table 1. Information Requested in Questionnaire for Each Site

General
- Location
- Professional Background of Respondent
- Years Experience with Groundwater Studies

Hydrogeologic Characteristics of Aquifer
- Hydraulic Conductivity:
  - *Maximum at Site*
  - Average
  - *Minimum at Site*
- Accuracy of Hydraulic Cond. Measurments
- Average Ground Water Velocity
- Method Used to Calculate Ground Water Vel.

Aquifer Characteristics
- Mean Grain Size of Aquifer Media
- Bulk Density
- Porosity
- Fraction Organic Carbon
- Depth from Surface to Top of Aquifer
- Saturated Thickness
- Hydraulic Gradient
- pH

Geologic Characteristics
- Soil Type
- Vadose Zone Material
- Aquifer Media

Hydrogeologic Setting

Source Characteristics
- Type of Source
- Surface Area of Source
- Type of Plume(s): Sinking, Floating, etc.
- Length and Width of *Major Dissolved Plume*
- Maximum Concentration of Dissolved Plume
- *Leachate Penetration* Depth Below Source

Untransformed vs. Transformed Data: Analysis Methods

Monte Carlo modeling requires that the underlying distribution of the data be determined. To check the underlying distribution of the raw data and the normality of various data, a graphical analysis with transformation probability plots from the Macintosh SYSTAT package (Wilkinson, 1987) was conducted. The term transformed data refers to the original dataset after a transformation, such as a logarithmic transformation, while untransformed data is the original unaltered dataset. Approximately 70 probability plots were constructed for the national average for each of the five parameters and about half of the environments. Two types of distributions were evaluated: normal and lognormal distributions.

Each of the five parameters appeared to have a lognormal distribution, based on the analysis of the probability plots. The lognormal character of the data suggests that median values are much better indicators of the central tendency of the untransformed data than the mean, and that summary statistics of the logarithmically transformed data are needed for inclusion into a Monte Carlo model such as EPACML.

Statistical Significance Between Means of Different Environments

Statistical significance tests were performed to determine if population means represent the same distribution. If the population means come from distributions which do not overlap at a prescribed level, then the populations are statistically independent. If the distributions of two populations overlap at the ends but only by a small amount (for this study a significance level of .05 was used), then the populations are said to be significantly different from each other. The statistical significance test of the hydrogeologic classification system is a strictly quantitative test.

In order to determine statistically if any of the 12 hydrogeologic environments were significantly different from each other, the means of each environment for the five parameters were compared using analysis of variance, or ANOVA. ANOVA is based on the assumption that the population variances of the groups are the same. Bartlett's test for homogeneity of variance was performed for the five parameters to determine if the variances were equal which would then indicate that an ANOVA procedure could be used. The results indicated that log hydraulic conductivity and log gradient could be analyzed using a one-way ANOVA without violating the variance assumption. (Although a multivariate procedure would have been more appropriate, a univariate test was used because of the exploratory nature of this study.) The other parameters have a range of variance between hydrogeologic environments that is too large for a simple ANOVA test.

Hydrogeologic Database: Results

The questionnaire contained 30 different questions regarding site location, hydrologic characteristics, aquifer characteristics, geologic characteristics, source characteristics, and the professional background of the respondent (Table 1). Most of the questions were descriptive in nature and related to the location of the sites, the type of source, and the type of contaminant plume. Five quantitative parameters (hydraulic conductivity, seepage velocity, hydraulic gradient, saturated thickness, and depth to top of aquifer) were analyzed to compare against the distributions used in the EPACML model, and to develop summary statistics for the hydrogeologic environments.

Descriptive Questions

A total of 400 useable questionnaires were returned from locations in 48 of the 50 states and the distribution of responses was widespread geographically, with only Arkansas and North Dakota not represented. California was the most frequent location of response, with 41 questionnaires returned; 28 responses came from Texas, 25 from Florida, 22 from Michigan, and 20 from New York. In addition, at least 8 responses were received from Indiana, Illinois, Massachusetts, Minnesota, New Mexico, New Jersey, Ohio, Pennsylvania, Washington, and Wisconsin (Figure 3).

Measurements of hydraulic conductivity were reported by method and by estimated accuracy in 379 questionnaires. About half of the respondents reported using pump tests for the site investigation (Table 2) and about half reported the estimated accuracy of their hydraulic conductivity data at

Table 2. Methods Used by Respondents to Estimate Hydraulic Conductivity

Measurement method	*Number of responses*	*Percent of 379 responses*
Pump Test and/or other methods	184	48.5
Slug Test and/or other methods	196	51.7
Lab Analysis and/or other methods	63	16.6
Grain Size and/or other methods	72	19.9
Literature/Engineering Judgment Only	114	30.1

Table 3. Estimated Accuracy of Hydraulic Conductivity Data as Reported by Respondents

Reported accuracy	*Number responding*	*Percent of responses*
±20%	79	22.7%
±50%	116	33.3%
±100%	56	16.1%
± X 10	88	25.3%
± X 100	9	2.6%

plus or minus 50% or less (Table 3). Darcy's Law was used as a method of determining seepage velocity in nearly 75% of the responses, while measurement of the plume's movement vs. time was mentioned as the method in the remaining responses. The seepage velocities in the HGDB are based more on professional judgment of the respondents than on direct field measurement.

The respondents represented a mix of disciplines and experience levels: 76% were geologists, hydrologists, or hydrogeologists; 16% were engineers; and 7% were from other disciplines. Nearly half had 5 years or less of experience performing hydrogeologic studies, while 26% reported over 10 years of experience. Most of the respondents appeared to work for consulting firms or industry. Approximately 7% appeared to be associated with universities or other research institutions.

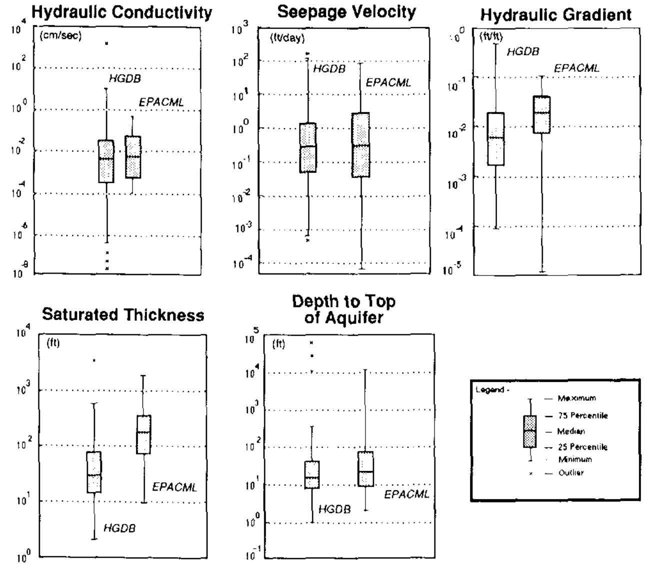


Fig. 7. Comparison of HGDB and EPACML distributions of aquifer parameters.

Table 4. Median EPACML Values Compared to Median HGDB Values for Four Parameters

	EPACML	*HGDB*
Median Seepage Velocity (ft/day)	0.315*	0.274
Median Hydraulic Conductivity (cm/sec)	0.006*	0.005
Median Hydraulic Gradient (ft/ft)	0.0028	0.007
Median Saturated Thickness (ft)	181.2	31.4
Median Depth to Top of Aquifer (ft)	20.0	15.0

*Derived from functional relationships; not derived from actual data.

HGDB Results Compared to EPACML

In order to compare the data distributions used in EPACML to the HGDB national distributions, the EPACML distributions for hydraulic conductivity, seepage velocity, hydraulic gradient, saturated thickness, and depth to top of aquifer were expressed as box plots (Figure 7) and as median values (Table 4). The EPACML hydraulic conductivity distribution and seepage velocity distribution was derived using the functional relationships in the EPACML background documentation (U.S. EPA, 1985, 1988a).

The box plot shows the median (line in the middle of the box plot), the 25th and 75th percentile of the data (top and bottom of box), the maximum and minimum data that are not considered outliers (shown as the whiskers extending out of the box), and outliers (x). The data are plotted on a logarithmic scale, and a box with the median in the middle of the box indicates that the data are described very well by a lognormal distribution. A box with the median line towards one end of the box indicates that the data are not described as well by a lognormal distribution.

Box plots were used to show the differences and the similarities between HGDB and EPACML distributions (Figure 7). The hydraulic conductivity box plots (Figure 7) are similar, with more extreme values shown by the HGDB data. The HGDB data range over nine orders of magnitude, while the EPACML ranges over four orders of magnitude. The two hydraulic conductivity boxes, the middle 50% of the values, are very similar, however, suggesting that the EPACML method to generate hydraulic conductivity values is successful in matching national average values.

The seepage velocity results show that the HGDB and EPACML distributions are also similar, although the EPACML distribution has a larger range of values. The similarity of the box plots suggests that the EPACML seepage velocity distribution is a reasonably accurate representation of the national distribution of seepage velocity, based on the independent check of the data provided by the HGDB.

The hydraulic gradient box plots (Figure 7) show significant differences between the two datasets. The median EPACML value is close to the 75th percentile value of the

Table 5. Statistical Significance Between Hydrogeologic Environments: Results of the Analysis of Variance (ANOVA) Procedure for Means

Hydrogeologic Environment (See Key)	1	2	3	4	5	6	7	8	9	10	11
1											
2											
3											
4											
5		g									
6	k										
7	k	g									
8					g	k g	g				
9	k g	k g						k g			
10	g	g						g			
11		g						g			

Hydrogeologic Environment (See Key)

Parameter Symbols

K = log k hydraulic conductivity
g = log gradient

Environment Abbreviations

1 Bedded Sedimentary Rock
2 Metamorphic and Igneous
3 Sand and Gravel
4 Alluvial Basins, Valleys, and Fans
5 River Valleys and Floodplains with Overbank Deposits
6 River Valleys and Floodplains without Overbank Deposits
7 Coastal Beaches
8 Till and Till Over Outwash
9 Outwash
10 Unconsolidated and Semiconsolidated Shallow Aquifers
11 Solution Limestone

A letter in a box indicates a significant difference between the means of the two environments at the .05 level

HGDB data. The EPACML has a tighter distribution of the middle 50% of the values (shown by the box), but shows a large tail of low gradient values going as low as 10^{-5} ft/ft. The EPACML seepage velocity is similar to the HGDB seepage velocity despite the differences in hydraulic gradient. Seepage velocity is a function of both gradient and hydraulic conductivity; gradient has a smaller effect because of the tighter distribution of gradient data compared to hydraulic conductivity (Tables 6 and 7).

The saturated thickness box plots (Figure 7) also show significant differences, with the HGDB showing thinner aquifers than the saturated thicknesses used in EPACML. The two boxes do not overlap, and the 75th percentile of the HGDB data is close to the 25th percentile of the EPACML data. The reason for the difference is not known. The overall range in the HGDB saturated thickness data is less than three orders of magnitude, about the same for the EPACML data. Preliminary modeling work indicates that the model results do not change significantly if the HGDB saturated thickness data were used; this may be due to the fact that the receptor well location used in this study was at the top of the aquifer. Work is now being performed to further examine the sensitivity of the EPACML to the HGDB saturated thickness values.

The depth to top of aquifer box plots (Figure 7) are very similar for the HGDB data and the EPACML data. Both datasets show extreme high values over 10,000 ft, representing sites with very thick unsaturated zones. Depth to top of aquifer is of little significance to the existing form of the EPACML model, because of the steady-state assumption and the one-dimensional unsaturated zone model used in EPACML. If there is no biodegradation or other degradation assumed in the unsaturated zone model, then the depth to water has no effect on the outcome of the modeling simulations.

Statistical Differences Between Hydrogeologic Environments

An analysis of variance (ANOVA) procedure was used to examine the statistical significance of the different environments. Two parameters, log hydraulic conductivity and log hydraulic gradient, were found to be suitable for the ANOVA analysis. Table 5 shows the environments that have significant differences in the means of the two parameters. Bedded Sedimentary Rocks (1), Metamorphic/Igneous Rocks (2), and Till-Till Over Outwash (8) showed the greatest differences between the other environments. The HGDB was not intended to be structured around statistically significant groups, but as a semisubjective framework based on the DRASTIC system. The ANOVA procedure is a stringent statistical test, and the ANOVA results show where the classification system used in the HGDB is strongest.

Statistical Summary of the Hydrogeologic Environments

The statistical distributions describing the national average and each of the different hydrogeologic environments are shown as box plots (Figures 8 and 9) and in summary tables (Tables 6 and 7). The summary tables show the median of the untransformed data and the mean and the standard deviation of the transformed data for the national average and each hydrogeologic environment.

The comparison of the data exhibits several trends that conform to the hydrogeology of the different environments:

Hydraulic Conductivity: The Outwash environment generally showed the highest median hydraulic conductivity value and had the tightest distribution, while Till, Till/Sedimentary Rock, Metamorphic/Igneous Rock, and Unconsolidated/Semiconsolidated environments had low median values. River With Overbank Deposits aquifers had lower median hydraulic conductivity compared to the River Without Overbank Deposits aquifers.

Seepage Velocity: Till, Metamorphic/Igneous Rock, Coastal Beaches, and Unconsolidated and Semiconsolidated aquifers had the lowest median seepage velocities. Outwash had the highest median velocities. Till had the widest range of values.

Hydraulic Gradient: Outwash aquifers had lower median hydraulic gradients while Till aquifers and Metamorphic/Igneous Rock aquifers had the higher median gradients, corresponding to the inverse of the hydraulic conductivity data.

Saturated Thickness: Outwash and Limestone aquifers had the highest median saturated thicknesses. Alluvial aquifers and Coastal Beaches had the widest range of values.

Depth to Top of Aquifer: Alluvial Basins had the highest median depth to top of aquifer data, while Coastal Beaches generally had lower values.

One application of the different hydrogeologic environments is to provide a more site-specific alternative to the existing national approach used in the EPACML regulatory process. Work is now ongoing to incorporate the HGDB distributions into the EPACML model, and preliminary results indicate that more stringent protection would be mandated for slower hydrogeologic environments (such as till) compared to fast environments (such as outwash). This result is counterintuitive, as most experienced hydrogeologists would rank an outwash aquifer as more vulnerable to serious ground-water problems than a till aquifer. The cause is the steady-state assumption in the EPACML model, which does not allow any consideration of travel time or the length of the contaminant plume. The effects of incorporating different environments in the existing regulatory framework are now being analyzed and will be reported in a future paper.

Using the HGDB

The database is structured so it can be applied in several different ways. For problems involving national distributions of aquifer parameters, the national average data are reported. For problems involving different types of aquifers, the hydrogeologic environments can be used to provide data related to the hydrogeology of aquifers. The hydrogeologic environments shown in Figure 6 can be used directly, or the DRASTIC system of hydrogeologic settings can be used (Aller et al., 1987). DRASTIC is an established aquifer classification system, and one can use the DRASTIC system of ground-water regions, block diagrams, typical characteristics, and narratives to determine the hydrogeologic setting that best describes a particular area or site. The hydrogeologic environment can be determined from the hydrogeologic setting using Appendix 1 of this paper.

Table 6. Summary Statistics by Hydrogeologic Environment for Hydraulic Conductivity and Seepage Velocity

HYDRAULIC GRADIENT (ft/ft)

Hydrogeologic Environment	Number of Cases	Grad. Median	Log Grad. Mean	Log Grad. St. Dev.
National Average	346	0.006	-2.22	0.8
Metamorphic / Igneous	23	0.019	-1.77	0.7
Bedded Sedimentary Rocks	52	0.009	-1.96	0.6
Till Over Sedimentary Rocks	17	0.010	-2.11	0.6
Sand & Gravel	223	0.005		
River Valleys With Overbank	25	0.004	-2.52	0.5
River Valleys Without Overbank	30	0.005	-2.25	0.6
Alluvial Basins, Valleys & Fans	38	0.005	-2.29	1.0
Outwash	26	0.002	-2.62	0.5
Till & Till Over Outwash	25	0.010	-1.70	0.7
Unconsol. & Semiconsolidated	25	0.005	-2.42	1.0
Coastal Beaches	25	0.004	-2.37	0.8
Solution Limestone	17	0.006	-2.35	0.8

SATURATED THICKNESS (ft)

Hydrogeologic Environment	Number of Cases	B Median	Log B Mean	Log B St. Dev.
National Average	350	30.0	1.58	0.5
Metamorphic / Igneous	22	30.0	1.66	0.5
Bedded Sedimentary Rocks	52	35.0	1.72	0.6
Till Over Sedimentary Rocks	12	13.5	1.27	0.4
Sand & Gravel	228	30.0	1.51	0.5
River Valleys With Overbank	26	25.5	1.41	0.3
River Valleys Without Overbank	30	37.5	1.65	0.4
Alluvial Basins, Valleys & Fans	44	25.0	1.48	0.6
Outwash	26	61.0	1.76	0.4
Till & Till Over Outwash	24	25.0	1.42	0.4
Unconsol. & Semiconsolidated	27	23.8	1.37	0.4
Coastal Beaches	26	35.0	1.63	0.6
Solution Limestone	16	59.1	1.87	0.5

DEPTH TO TOP OF AQUIFER (ft)

Hydrogeologic Environment	Number of Cases	D. Aq. Median	Log D. Aq. Mean	Log D. Aq. St. Dev.
National Average	348	15.0	1.28	0.3
Metamorphic / Igneous	27	17.5	1.25	0.6
Bedded Sedimentary Rocks	60	20.0	1.43	0.5
Till Over Sedimentary Rocks	16	17.5	1.23	0.4
Sand & Gravel	245	12.0	1.22	0.5
River Valleys With Overbank	34	16.5	1.13	0.4
River Valleys Without Overbank	31	15.0	1.22	0.5
Alluvial Basins, Valleys & Fans	51	25.0	1.47	0.5
Outwash	27	15.5	1.29	0.4
Till & Till Over Outwash	25	10.0	1.07	0.4
Unconsol. & Semiconsolidated	27	12.5	1.22	0.4
Coastal Beaches	25	6.0	0.97	0.8
Solution Limestone	20	25.0	1.54	0.6

Table 7. Summary Statistics by Hydrogeologic Environment for Hydraulic Gradient, Saturated Thickness, and Depth to Top of Aquifer

HYDRAULIC CONDUCTIVITY (cm/s)

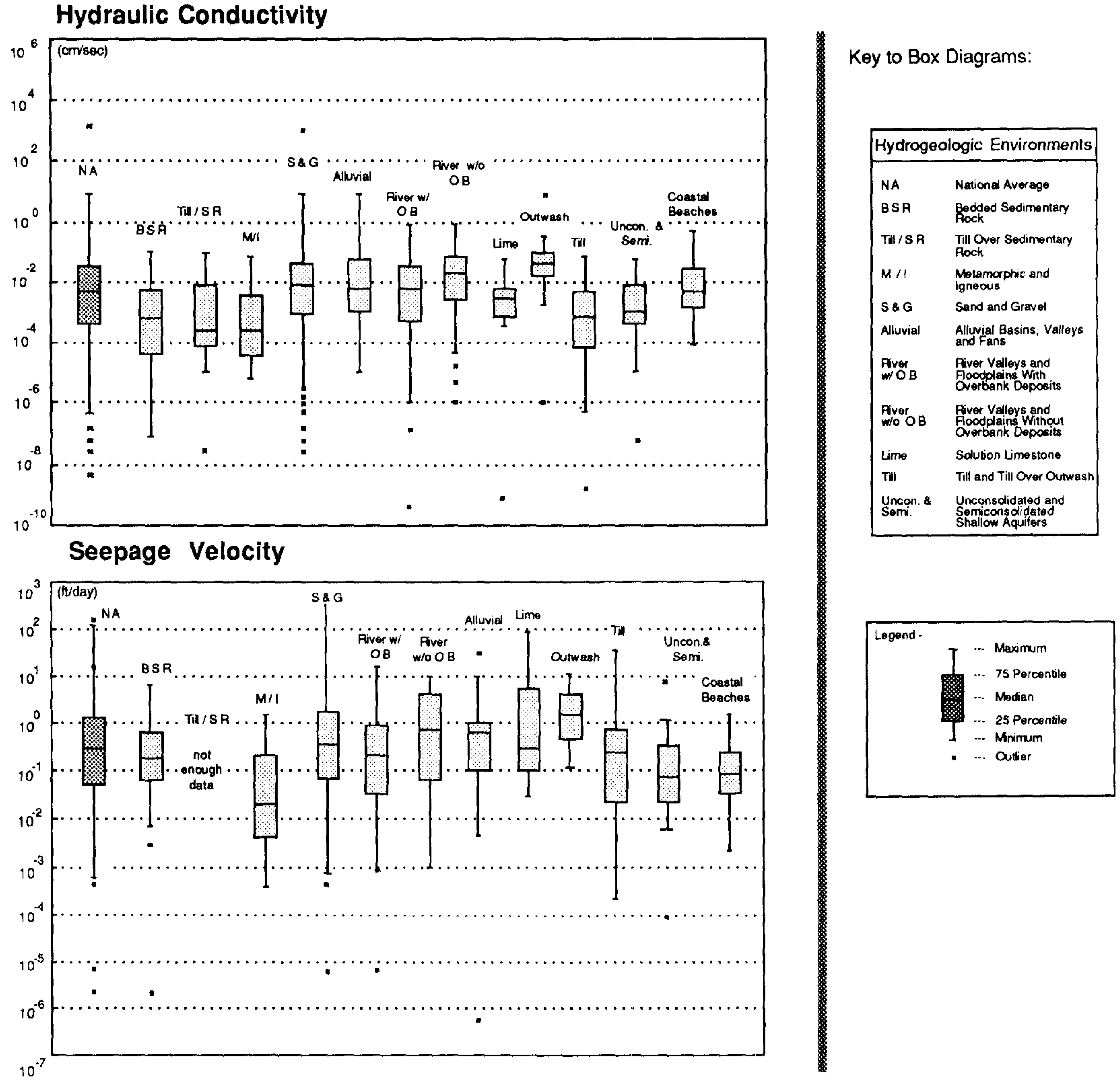
Hydrogeologic Environment	Number of Cases	K Median	Log K Mean	Log K St. Dev.
National Average	287	0.0050	-2.63	1.6
Metamorphic / Igneous	19	0.0003	-3.42	1.3
Bedded Sedimentary Rocks	51	0.0003	-3.40	1.6
Till Over Sedimentary Rocks	13	0.0005	-3.42	1.8
Sand & Gravel	191	0.0080	-2.41	1.6
River Valleys With Overbank	20	0.0060	-2.66	2.2
River Valleys Without Overbank	26	0.0200	-2.16	1.7
Alluvial Basins, Valleys & Fans	36	0.0070	-2.19	1.4
Outwash	22	0.0470	-1.51	1.2
Till & Till Over Outwash	20	0.0009	-3.44	1.5
Unconsol. & Semiconsolidated	18	0.0010	-2.87	1.4
Coastal Beaches	18	0.0065	-2.05	1.0
Solution Limestone	11	0.0040	-2.81	1.2

SEEPAGE VELOCITY (ft/day)

Hydrogeologic Environment	Number of Cases	V Median	Log V Mean	Log V St. Dev.
National Average	290	0.24	-0.59	1.2
Metamorphic / Igneous	21	0.14	-0.69	1.2
Bedded Sedimentary Rocks	38	0.11	-0.92	1.5
Till Over Sedimentary Rocks	11	0.11	-1.37	2.6
Sand & Gravel	193	0.31	-0.62	1.3
River Valleys With Overbank	25	0.35	-0.79	1.5
River Valleys Without Overbank	30	1.20	-0.34	1.1
Alluvial Basins, Valleys & Fans	34	0.63	-0.41	1.1
Outwash	20	1.40	0.31	0.6
Till & Till Over Outwash	24	0.80	-0.88	1.3
Unconsol. & Semiconsolidated	21	0.07	-1.06	1.7
Coastal Beaches	16	0.09	-0.90	0.9
Solution Limestone	12	0.27	0.21	1.6

Once the hydrogeologic setting or environment is determined, one can use the HGDB data in two ways: general site characterization, and for Monte Carlo modeling. The box plots and the median data reported in the summary statistics are useful for seeing the general characteristics of sites in different environments. The data are also useful from an educational standpoint, as they show the large range in reported values and indicate the uncertainty involved in aquifer characterization. The mean and the standard deviation of the logarithmically transformed data are designed to be used for statistical problems, such as Monte Carlo Modeling.

The HGDB only contains general information about the hydrogeologic characteristics of aquifers in different hydrogeologic settings, and cannot provide a comprehen-



Hydrogeologic Environments	
NA	National Average
BSR	Bedded Sedimentary Rock
Till / S R	Till Over Sedimentary Rock
M / I	Metamorphic and Igneous
S & G	Sand and Gravel
Alluvial	Alluvial Basins, Valleys and Fans
River w/ O B	River Valleys and Floodplains With Overbank Deposits
River w/o O B	River Valleys and Floodplains Without Overbank Deposits
Lime	Solution Limestone
Till	Till and Till Over Outwash
Uncon. & Semi.	Unconsolidated and Semiconsolidated Shallow Aquifers

Fig. 8. Box plots by hydrogeologic environment for hydraulic conductivity and seepage velocity.

sive picture of individual site characteristics. The HGDB should not be used as a sole source of data for detailed design calculations for an individual site; actual field measurements are required. Also, caution should be used in applying mean or median values for some problems. For example, flow and contaminant transport at a site with a low average hydraulic conductivity can be dominated by small areas of high hydraulic conductivity (such as fractures or sand channels). The mean or median hydraulic conductivity values from the HGDB would be inappropriate in this case, and the maximum hydraulic conductivity in the HGDB database would be a better estimator of aquifer flow properties. In general, the HGDB is better suited for regulation, planning, and for education than for detailed design.

The database is now available on different media. This paper provides a brief description of the data collection, structure, and the results of the HGDB. A more detailed written report is available from the American Petroleum Institute, 1220 L Street Northwest, Washington, DC 20005 (Newell, Hopkins, and Bedient, 1988). The entire database itself is also available from the API in the form of an MS-DOS text file on a floppy disk which can be used in databases or spreadsheets. The HGDB will also be available in OASIS, a ground-water modeling decision support system being developed for EPA on Macintosh microcomputers (Newell, Haasbeek, and Bedient, 1990).

Summary and Conclusions

1. A new hydrogeologic database (the HGDB) was developed from 400 technical questionnaires supplied by members of the National Water Well Association. The data were grouped into 12 hydrogeologic environments for analysis. A method based on the EPA's DRASTIC hydrogeologic settings was developed to allow users to determine the

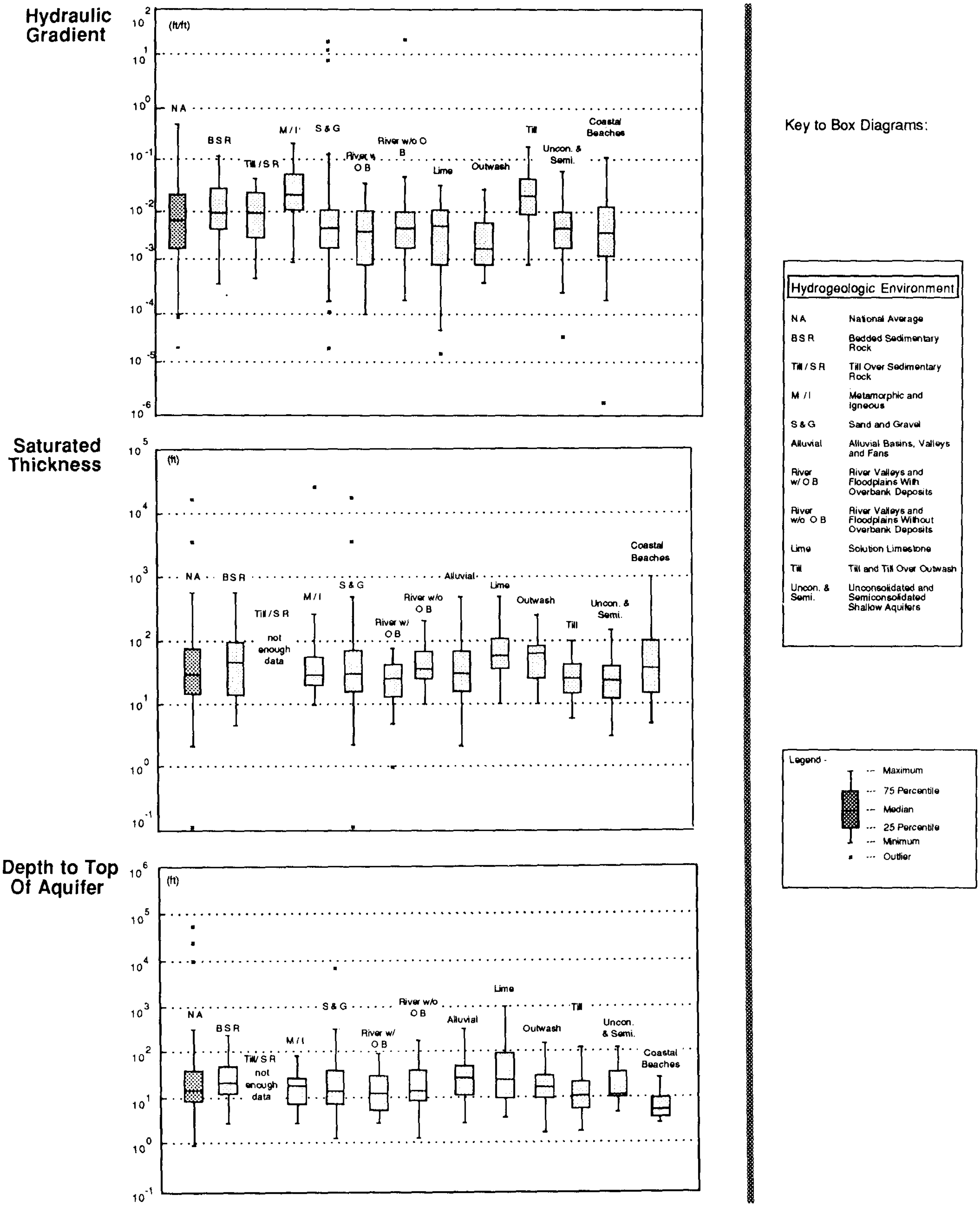


Fig. 9. Box plots by hydrogeologic environment for hydraulic gradient, saturated thickness, and depth to top of aquifer.

hydrogeologic environment of a site of interest. Five parameters were evaluated in detail: hydraulic conductivity, seepage velocity, hydraulic gradient, saturated thickness, and depth to the top of the aquifer.

2. The HGDB provides a unique framework for organizing hydrogeologic data from site investigations. The framework is based on the NWWA's DRASTIC system, an aquifer classification method based on hydrogeologic settings. The existing HGDB now has data from 400 sites, and data from additional site investigations can be added easily

Appendix 1. Hydrogeologic Environments of DRASTIC Hydrogeologic Settings

HYDROGEOLOGIC SETTING (from DRASTIC)	# of Qst.	HYDROGEOLOGIC ENVIRONMENT (from HGDB)
Western Mountain Ranges		
Mountain Slopes Facing East	1	M/I
Mountain Slopes Facing West	0	M/I
Mountain Flanks Facing East	1	BSR
Mountain Flanks Facing West	1	BSR
Swamp/ Marsh	0	BSR
Wide Alluvial Valleys Facing East	3	Alluvial
Wide Alluvial Valleys Facing West	2	Alluvial
Alluvial Mountain Valleys Facing East	2	Alluvial
Alluvial Mountain Valleys Facing West	2	Alluvial
Glaciated Mountain Valleys	0	S&G
Coastal Beaches	2	Coastal Beaches
Mud Flows	0	S&G
Alluvial Basins		
Mountain Slopes	4	M/I
Swamp/ Marsh	0	BSR
Alternating Sedimentary Rocks	3	BSR
River Alluvium with Overbank Deposits	14	River w/OB
River Alluvium without Overbank Deposits	8	River w/o OB
Coastal Lowlands	5	S&G
Alluvial Mountain Valleys	0	Alluvial
Mud Flows	0	S&G
Alluvial Fans	11	Alluvial
Alluvial Basins with Internal Drainage	3	Alluvial
Playa Lakes	2	Alluvial
Continental Deposits	15	Alluvial
Columbia Lava Plateau		
Mountain Slopes	0	M/I
Swamp/ Marsh	0	M/I
Lava Flows: Hydraulically Connected	1	M/I
Lava Flows: Not Hydraulically Connected	0	BSR
Alluvial Fans	1	Alluvial
Alluvial Mountain Valleys	0	Alluvial
River Alluvium	3	River w/o OB
Colorado Plateau and Wyoming Basin		
Resistant Ridges	2	BSR
Consolidated Sedimentary Rocks	3	BSR
Alluvium and Dune Sand	1	S&G
River Alluvium	2	River w/o OB
Swamp/ Marsh	0	S&G
High Plains		
Swamp/ Marsh	0	S&G
River Alluvium with Overbank Deposits	2	River w/OB
River Alluvium without Overbank Deposits	1	River w/o OB
Braided River Deposits	0	River w/o OB
Sand Dunes	0	S&G
Alluvium	0	Alluvial
Playa Lakes	1	Alluvial
Ogalalla	2	Alluvial
Alternating Sedimentary Rocks	0	BSR
Non-Glaciated Central Region		
Metam. & Igneous Domes and Fault Blocks	0	M/I
Triassic Basins	8	BSR
Mountain Slopes	3	BSR
Mountain Flanks	5	BSR
Alt. Beds of SS, LS,or SH Under Thin Soil	9	BSR
Alt. Beds of SS, LS,or SH Under Regolith	3	BSR
Swamp/ Marsh	0	BSR
Alluvial Mountain Valleys	1	Alluvial
Braided River Deposits	1	River w/o OB
River Alluvium With Overbank Deposits	8	River w/OB
River Alluvium Without Overbank Deposits	4	River w/o OB
Unconsol. & Semi-Consolidated Aquifers	9	Alluvial
Solution Limestone	10	Lime
Glaciated Central Region		
Till Over Solution Limestone	6	Lime
Outwash Over Solution Limestone	2	Lime
Till Over Bedded Sedim. Rock	3	Till/SR
Thin Till Over Bedded Sedim. Rock	4	Till/SR
Outwash Over Bedded Sedim. Rock	4	BSR
Till Over Sandstone	4	Till/SR
Till Over Shale	6	Till/SR
Glacial Lake Deposits	2	S&G
Outwash	27	Outwash
Till Over Outwash	3	Till
Moraine	6	S&G
Buried Valley	5	S&G
River Alluvium With Overbank Deposits	7	River w/OB
River Alluvium Without Overbank Deposits	8	River w/o OB
Beaches, Beach Ridges, and Sand Dunes	2	Coastal Beaches
Swamp/ Marsh	3	S&G
Till (HGDB Setting)	12	Till
Northeast and Superior Uplands		
Outwash	13	Outwash
Alluvial Mountain Valleys	1	Alluvial
River Alluvium With Overbank Deposits	3	River w/OB
River Alluvium Without Overbank Deposits	1	River w/o OB
Till (HGDB Setting)	8	Till
Atlantic & Gulf Coast		
Confined Regional Aquifers	3	S&G
Unconsolidated & Semiconsolidated Surficial Aq.	27	Uncon. & Semi.
River Alluvium With Overbank Deposits	3	River w/OB
River Alluvium Without Overbank Deposits	2	River w/o OB
Swamp	3	S&G
Southeast Coastal Plain		
Solution Limestone and	5	Lime
Unconsolidated & Semiconsolidated Surficial Aq.	1	S&G
Swamp	4	Coastal Beaches
Beaches & Bars	17	Coastal Beaches
Hawaii		
Volcanic Uplands	1	M/I
Mountain Slopes	0	M/I
Alluvial Mountain Valleys	0	Alluvial
Coastal Beaches	1	Coastal Beaches
Alaska		
Bedrock of the Mountains	0	BSR
Coastal Lowland Deposits	1	S&G
Alluvium	0	S&G
Glacial and Glacio-lacustrine Deposits of the Interior Valleys	2	S&G

Key to Hydrogeologic Environments:

BSR: Bedded Sedimentary Rocks
Till/SR: Till Over Sedimentary Rocks
M/I: Metamorphic/Igenous Rocks
S&G: Sand & Gravel
Alluvial: Alluvial Basins, Valleys and Fans
River w/ OB: River Valleys and Floodplains with Overbank Deposits
River w/o OB: River Valleys and Floodplains without Overbank Deposits
Lime: Solution Limestone
Till: Till and Till Over Outwash

to the HGDB by determining the DRASTIC hydrogeologic setting for the site. It is envisioned that, with additional sites, the grouping of data into the 12 hydrogeologic environments can be expanded by developing statistical summaries of some of the 111 different DRASTIC settings.

3. The HGDB was used to provide an independent check against the input distributions contained in the EPACML model, an analytical model for proposed land disposal regulations. EPACML is a Monte Carlo groundwater model that uses statistical distributions of input data, many of which are generated from physical relationships. The seepage velocity distribution derived in EPACML was very similar to the observed distribution from the HGDB, indicating that the EPACML model accurately accounts for the distribution of seepage velocities found at sites across the country. Because the HGDB is a completely independent source from the database used in EPACML, it suggests that the calculation used to generate seepage velocities in EPACML is reasonably accurate. The HGDB and EPACML hydraulic conductivity distributions were also similar, as were the depth to the top of aquifer distributions, while greater differences were observed with hydraulic gradient and saturated thickness (Figure 7). Preliminary work on the sensitivity of the model to the new HGDB distributions in the model indicate that the results of the EPACML modeling would not change significantly if the HGDB distributions are used. A paper is now being prepared discussing the sensitivity of the model to each of these parameters.

4. The HGDB data can be used to determine the effect of hydrogeologic environment on the proposed land disposal regulations. Preliminary results indicate that more stringent protection would be mandated for slower hydrogeologic environments (such as till) compared to fast environments (such as outwash) if hydrogeologic environments were incorporated into existing national regulatory process. This result is counterintuitive as fast environments are generally considered more vulnerable to contamination problems. Additional work is now being conducted to further examine the implications of using hydrogeologic environments as part of the regulations, and will be reported in a future paper.

Acknowledgments

The authors would like to thank the American Petroleum Institute for the funding and technical support for this research. We would also like to thank Jay Lehr, Linda Aller, and Rebecca Petty of the National Water Well Association, Paul Lewandowski of API, the members of the API Land and Groundwater Modeling Task Force for their contributions to the development of the hydrogeologic database, and Jill Oglesby, Julie Smythe, Jessica A. Gould, and J. Peter Nevin for their assistance in developing the database.

References

Aller, L. A., T. Bennett, J. H. Lehr, and G. Hackett. 1987. DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. Washington, DC. EPA/600-2-85/018.

Bear, Jacob. 1979. Hydraulics of Groundwater. McGraw-Hill.

Davis, S. N. 1969. Porosity and permeability of natural materials. In: Flow Through Porous Media. Ed., Roger J. M. DeWiest. Academic Press.

Federal Register. 1986. 51 Federal Register 1602, January 14, 1986.

Federal Register. 1988. 53 Federal Register 28892, August 1, 1988.

Freeze, R. A. and J. A. Cherry. 1979. Groundwater. Prentice-Hall, Inc., New Jersey.

Heath, Ralph C. 1984. Ground water regions of the United States. Washington, DC. U.S. Geological Survey Water Supply Paper 2242.

Huyakorn, P. S., M. J. Ungs, L. A. Mulkey, and E. A. Sudicky. 1987. A three-dimensional analytical method for predicting leachate migration. Ground Water. v. 25, no. 5, p. 588.

Newell, C. J., J. H. Haasbeek, and P. B. Bedient. 1990. OASIS: A graphical, Hypertext decision support system for groundwater modeling. Ground Water. v. 28, no. 2, pp. 224-234.

Newell, C. J., L. P. Hopkins, and P. B. Bedient. 1988. Hydrogeologic Database for Ground Water Modeling. American Petroleum Institute, Washington, DC.

Meinzer, O. E. 1923. Outline of ground water hydrology. U.S. Geological Survey Water Supply Paper 494.

Rosner, Bernard. 1986. Fundamentals of Biostatistics. Harvard Univ. Duxbury Press, PWS Publishers, Boston, MA. ISBN 0-87150-981-4.

U.S. Environmental Protection Agency. 1985. Development of Land Disposal Banning Decisions Under Uncertainty. Environmental Research Laboratory, U.S. EPA, Athens, GA.

U.S. Environmental Protection Agency. 1986. Background Document: Ground Water Screening Procedure. Office of Solid Waste, U.S. EPA, Washington, DC.

U.S. Environmental Protection Agency. 1988a. Background Document on the Subsurface Fate and Transport Model. Office of Solid Waste, U.S. EPA, Washington, DC.

U.S. Environmental Protection Agency. 1988b. User's Manual for EPA's Composite Landfill Model (EPACML). Office of Solid Waste, U.S. EPA, Washington, DC.

Wilkinson, Leland. 1987. SYSTAT: The System for Statistics. SYSTAT, Inc., Evanston, IL.

* * * * *

Charles J. Newell received a Ph.D. in Environmental Engineering from Rice University in 1989, and designed the HGDB hydrogeologic database and the OASIS Ground Water Modeling Decision Support System as part of his dissertation research. Newell graduated with a B.S. in Chemical Engineering in 1978 and an M.S. in Environmental Engineering in 1980 from Rice University, and has six years experience as an environmental consultant. His research and professional interests include ground-water modeling, design of in-situ biorestoration systems, and field studies of hydrocarbon residual dissolution. He is now an Environmental Engineer and Vice President of Groundwater Services, Inc., of Houston, Texas.

Loren P. Hopkins graduated from Rice University in 1989 with a Master of Science in Environmental Science and from the University of Texas in 1986 with a B.S. in Geophysics. Her research interests are in the area of hydrogeologic databases, hydrogeologic classification, and stochastic ground-water modeling. She is now employed as a Project Scientist with OHM Corporation of Austin, Texas.

Philip B. Bedient is a Professor of Environmental Engineering at Rice University and holds the Shell Distinguished Chair of Environmental Science. He graduated with a B.S. in Physics in 1969 and an M.S. and Ph.D. in Environmental Engineering from the University of Florida, Gainesville, in 1975. Bedient directed the development of the BIOPLUME II biodegradation model and the OASIS Groundwater Modeling Decision Support System, and has directed contaminant transport monitoring and modeling work at 12 sites across the country.