⊕EPA

Project Summary

Sampling Frequency for Ground-Water Quality Monitoring

Michael J. Barcelona, H. Allen Wehrmann, Michael R. Schock, Mark E. Sievers, and Joseph R. Karny

This project was initiated to collect a benchmark water-quality dataset and evaluate methods to optimize sampling frequency as a network design variable. Ground water was collected biweekly for 18 months from twelve wells at two sites in a shallow sand and gravel aquifer in Illinois. Sampling and analyses were conducted for twenty-six water quality and geochemical constituents with careful quality control measures to allow statistical analysis of variability in ground-water quality data. The results demonstrate that natural variability over time can exceed the variability introduced into the data from sampling and analysis procedures.

Natural temporal variability and the highly autocorrelated nature of ground-water quality data seriously complicate the selection of optimal sampling frequency and the identification of seasonal trends in ground-water quality variables. Quarterly sampling frequency is a good initial starting point for ground-water quality monitoring network design, although bimonthly frequency may be preferred for reactive chemical constituents. Analysis of data collected during this project suggests that the collection of a long-term (i.e., more than two years) dataset is necessary to determine optimal sampling frequency and to identify seasonal trends in ground-water monitoring results.

This Project Summary was developed by EPA's Environmental Monitoring Systems Laboratory, Las Vegas, NV, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

There are two principal sources of variability in ground-water quality data, "natural" variability and variability resultant from the network design and operation The components of "natural" variability arise from temporal or spatial variability related to hydrologic processes such as pumpage, recharge or discharge, as well as influences of these processes on the release and distribution of chemical constituents from a variety of chemical sources. The sources may be natural mineral assemblages, precipitation and percolation through the unsaturated zone, in addition to numerous point and non-point sources of chemical contaminants. In general, "natural" sources of variability cannot be controlled although they may be quantified through effective monitoring network design.

Water-quality data variability may also arise from the sampling and analytical components of monitoring network design. Sampling variability includes variations due to the selection of the locations and construction of sampling points in space, sampling frequency, well purging, and the execution of the sampling protocol. The sampling protocol consists of the procedures used to collect, handle, preserve, and transport water samples to the analytical laboratory. Elements of the sampling protocol have been evaluated for their relative

contributions to variability or errors in water-quality data in previous research.

Analytical variability in water-quality data arises principally from the errors involved in analytical methods and the subsequent data processing steps. These errors can be controlled once suitable water-quality indicators or chemical constituents have been selected and a thorough data quality assurance/quality control program has been designed and executed.

This study was planned to control the sources of variability in water quality data which result from network design components such as sampling location, frequency, sampling methods and analytical procedures. The sampling frequency was held constant at a biweekly interval between sample collection dates. The benchmark dataset that resulted from this experimental design could then be analyzed to determine the optimal sampling frequency for selected water-quality variables at both uncontaminated and contaminated study sites.

The full report describes the level of QA/QC effort which is necessary to achieve control over sources of error and data variability due to sampling and analytical operations. Discussions of temporal variability in groundwater level and water quality results are included in the report to place the dataset in perspective. The results and conclusions of the work are supported by extensive references, where the literature permits. The report should be useful to the planning and execution of regulatory and research activities which demand the cost-effective collection of high quality ground-water quality data.

Variability in Ground-Water Quality

The impact of the sources of variability mentioned above will be influenced by the hydrology of the ground-water system. It is important to understand that although aquifer hydraulic properties may not vary significantly at a single measurement point over time, spatial variability may be substantial. Aquifer hydraulics may be expected to influence chemical constituent distributions in space and time.

Temporal and spatial variations in ground-water elevation may affect ground-water flow rate and the direction of movement. Such changes may influence the quality of the ground water in the vicinity of a sampled well by directing water from a different upgradient area or changing the velocity with which dis-

solved constituents move along a flow path. Examples abound in the literature detailing ground-water response (i.e., elevation change) to a wide variety of influences. In addition to seasonal fluctuations produced in response to short-term (i.e., months to one year) events, ground-water levels also reflect changes in long-term (i.e., years to decades) conditions.

Temporal and spatial variability may also result from sample collection and measurement errors inherent to network design and operation. This variability, or "noise," in the data embodies the stochastic distribution of possible values for particular chemical constituents and the effects of both determinate (i.e., systematic) and indeterminate (i.e., random) error. Determinate error can be measured as inaccuracy or bias if the "true value" is known. Indeterminate error can be estimated as imprecision or irreproducibility if a sufficient number of replicate determinations can be made to faithfully estimate the mean or the "true" value.

Statistical measures of short-term temporal variability include seasonal effects (e.g., consequences of recharge or temperature effects) which can be assigned to the seasons of the year, periodic effects (e.g., consequences of anthropogenic contaminant sources or pumping effects) and serial correlation or dependence effects which tend to make data points following maxima or minima in temporal data series higher or lower, respectively, than one would attribute to random processes alone. Trends in data, on the other hand, are long-term variations compared to those which may occur within a hydrologic year.

Procedure

Field Sites

Two sites were chosen to enable the isolation of the effects of network design variables fro those due to natural or contaminant-related sources. The sites were located over an alluvial sand and gravel water table aquifer of moderate to high yield. One site was in a pristine environment far removed from any sources of contamination in the Sand Ridge State Forest near Havana, Illinois. The other site was in an industrial environment under the influence of a leaking anaerobic waste impoundment near Beardstown, Illinois.

Sand Ridge State Forest is an Illinois Department of Conservation (IDOC) facility located 5 miles (8 km) southeast of

the Illinois River in the north-c Havana Lowland. The Illinois State Survey's experimental field site is lc in the middle of the State Forest Havana, Illinois.

Three distinct horizons compris unconsolidated deposits at Sand I at the surface is 30 feet (9 m) of sand (the Parkland sand); from 3D f m) to a depth of 55 feet (17 m) Manito Terrace of the Wisconsinai wash, consisting of a sometimes sometimes coarse sand to me gravel; and from 55 feet (17 m) do bedrock below 110 feet (34 m), anc sibly as deep as 150 feet (46 m), medium sand to fine gravel o Sankoty sand (Kansan outwash).

Depth to the water table is greatei 30 feet (9 m) below the ground su Ground-water movement is general ward the Illinois River. The hydrauli¢ dient measured at the site in 198: approximately 0.0016. Aquifer tests ducted on the water supply wells nearby state fish hatchery indicate the hydraulic conductivity of the san¢ gravel at approximately 100 feet (: depth (in the Sankoty sand) is about gpd/ft^2 (0.094 cm/sec). Tracer e) ments conducted in 1983 indicatec lover hydraulic conductivities (fron to 1900 gpd/ft^2, 0.01 to 0.09 cm/sec, be exhibited by the finer-grained, lower materials. Hydraulic conduc values of 350 to 900 gpd/ft2 (0.02 tc cm/sec) were obtained by emp methods of analysis based on the size distributions of shallow aq samples. The porosity of the satu terrace materials was found to be 25

The "contaminated" field sit located in the vicinity of several I waste impoundments serving a slaughtering facility approximately 1 (1.6 km) southeast of Beardstown, Ill The field site lies two miles southe¢ the river, and it is only about 5 feet m) higher than the floodplain. Farn and wooded areas surround the fa The unconsolidated deposits lying ¢ the bedrock consist of the clayey s of the Beardstown Terrace on Wisconsinan outwash plain. The be¢ surface is of Mississipian age and li about 100 feet (30 m) below the gr surface.

Owing to land surface eleva changes, depth to water varies from 15 feet (1.5 to 4.5 m) below the gr surface. Similar to the Sand Ridge regional ground-water flow is towar¢ Illinois River (hydraulic gradient, 0. Due to the presence of silt and clay aquifer is less permeable than it is a

Sand Ridge site. One falling head permeability test produced a hydraulic conductivity value of only 130 gpd/ft^2 (6 x 10^{-3} cm/sec).

Monitoring Wells

Bore holes for construction of all monitoring wells were drilled with a 4.25-inch (11 cm) inside diameter (I.D.) hollow-stem auger. All auger flights, solid samplers, well casing materials, and well protectors were steam cleaned before use or placement in the bore hole.

The construction details of the sampling wells at both sites are identical in all ways other than the length of casing and casing materials in two wells at Beardstown. One well at Beardstown vas constructed of stainless steel (SS) and one other of polyvinylchloride (PVC). All of the other sampling wells at both sites were constructed with polytetrafluoroethylene (PTFE-Teflon®, DuPont). All wells have 2-inch (5 cm) I.D. flush-threaded casing. Screens were 5 feet (1.5 m) long with 0.01-inch (0.02 cm) slot openings. The four wells at Sand Ridge were completed at depths of 35, 50, 65, and 105 feet (10.6, 15.4, 20, and 32 m), respectively.

The eight wells at Beardstown were completed at several depths at locations upgradient and downgradient from the impoundment.

Results and Discussion

Five preliminary sampling runs were completed between November 1985 and March 1986. Then thirty-nine biweekly sampling trips were conducted during the period of March 10, 1986 through August 25, 1987. These field activities involved purging and sampling the monitoring wells 526 times and measuring more than 2,000 ground-water levels. Only two wells were missed out of the 528 sampling opportunities. Water samples were collected for more than 26 analytical determinations each, including major cations, anions, TOC, TOX, pH, alkalinity, and other species.

During the course of the study, more than 55,000 analytical determinations vere made on blanks, standards and samples. The final dataset was 96% complete, that is, 96% of the maximum possible number of samples and subsequent analytical determinations were successfully completed. Outliers were screened successively at ±3 and ±2 standard deviations from the mean levels. In most cases, this screening revealed apparent errors in calculations, calibration, or data entry which were corrected

prior to data analysis. For all wells and constituents, the maximum number of samples which were identified as possible outliers and for which no documented error was identified was four percent of the total. No adjustment was made to apparent outliers for which no documented error could be identified. QA/QC analyses demonstrated that the analytical methods were within control limits and that good analytical performance was maintained throughout the project period.

Estimation of Sources of Variation

Generally, the natural variations in water quality time series are of interest. For instance, the difference between the time series of a given contaminant at a downgradient and an upgradient well may give an indication of whether contaminant release has occurred. However, the difference series is inevitably corrupted by errors in the field data collection and laboratory analysis procedures, both of which introduce what may be considered "noise" into the time series. Each of these noise processes has a variance, and the total variance is the sum of the three variance terms. this model assumes that the three sources of variation are statistically independent. This is a reasonable assumption because the sources are physically independent and the individual variances were calculated from the analytical results from replicate control samples, lab and field spiked samples.

The results are summarized in Table 1 for three groups of wells (i.e., Sand Ridge wells 1 to 4, Beardstown upgradient wells 5 and 6, and the Beardstown downgradient wells). for almost all of the groups, and for almost all of the chemical constituents, a high fraction of the total variation was natural. In most cases the combined lab and field variances were below ten percent of the total variance. This is consistent with the QA/QC data analyses, which showed that the data collection errors were generally quite small. The entries in the table have been separated into water quality parameters and chemical parameters of geochemical interest. The results confirm that if careful sampling and analytical protocols are used, the analytical and sampling errors can be held to less than about 20%. Therefore, the natural variability in the major ion chemistry of the system can be identified. For TOC and TOX it is clear that "natural" sources of variability are greater than the combined lab and field

variance. However, the level of overall variability in TOX results was quite large in comparison to the mean values for each well. The significance of these determinations at the microgram per liter concentration level is doubtful.

The implication of the results of this study is that network design optimization efforts should focus primarily on the natural or contamination source-related variability. The use of field and laboratory replication for purposes other than QA/QC will be difficult to justify as long as the sampling and analytical protocols are in control. This conclusion must be qualified, however. The chemical constituents present at appreciable concentrations (i.e., mg·L-1) at either site were the major cations and anions and general water quality indicators. The analytical and sampling variances for trace organic contaminants would be expected to be higher, and their analytical recoveries are frequently found to be a function of concentration. For such contaminants, the field and laboratory variations may not be independent, which would violate a basic assumption in this model.

Temporal Variations in Ground-Water Quality

There are numerous examples of both short- and long-term variability in ground-water quality in the literature. Significant short-term temporal concentration variability has been observed in low-yield wells (i.e., monitoring and observation wells) largely resulting from purging effects. Similar variations from one to ten times the initial or background concentrations have been noted in samples from high-volume production wells due to pumping rate, initial pumping after periods of inactivity, and cone of depression development.

In general, the major ionic chemical constituents determined in this study showed differences between their overall maximum and minimum values from the mean for each well on the order of one or two times the mean value. One or two times the mean value places the variability noted in this study in the same range as long-term, seasonal variability. The magnitude of overall long-term variations observed in this study and the literature is often much lower than those noted for short-term variations due to pumping and local recharge effects. The magnitude of short-term concentration variations noted in the literature strongly suggests that the analysis of ambient resource, water quality datasets must be undertaken with careful attention to the

Table 1. Percentage of Variance Attributable to Laboratory Error, Field Error, and Natural Variability by Chemical and Site

Type of Parameter	Sand Ridge			Beardstown (Upgradient)			Beardstown (Downgradient)		
	Lab	Field	Nat	Lab	Field	Nat	Lab	Field	Nat
Water Quality									
NO_3^-	0.0	0.0	100.0	0.1	NA*	99.9	0.2	NA	99.8
$SO_4^=$	0.0	0.0	100.0	0.2	NA	99.8	1.4	0.0	98.6
$SiO_2=$	0.0	NA	100.0	0.0	20.0	80.0	0.0	6.8	93.2
$o\text{-}PO_4=$	1.2	1.2	97.6	0.0	0.0	100.0	0.0	0.0	100.0
$T\text{-}PO_4$	0.0	NA	100.0	2.8	NA	97.2	0.9	NA	99.1
Cl^-	7.2	NA	92.8	0.0	3.3	96.7	0.0	17.2	82.8
Ca	0.0	45.7	54.3	0.0	2.3	97.7	0.0	3.6	96.4
Mg	0.0	20.0	80.0	0.0	2.2	97.8	0.0	2.8	97.2
Na	0.0	NA	100.0	0.0	0.3	99.7	0.0	7.1	92.9
K	0.0	NA	100.0	33.9	NA	66.1	87.1	NA	12.9
Geochemical									
$NH_3\text{-}$	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0
NO_2	NA	NA	NA	0.1	NA	99.9	0.3	NA	99.7
$S^=$	NA	NA	NA	NA	NA	NA	NA	NA	NA
Fe^{2+}	NA	NA	NA	0.0	0.1	99.9	0.0	5.9	94.1
Fe_T	0.0	NA	100.0	0.0	0.0	100.0	0.0	NA	100.0
Mn_T	0.0	NA	100.0	0.0	40.1	59.9	0.0	73.6	26.4
Contaminant Indicator	Lab + Field			Lab + Field			Lab + Field		
TOC^{**}	15.4		84.6	29.9		70.1	40.5		59.5
TOX^{**}	0.0		100.0	12.5		87.5	24.6		75.4

*NA indicates that the number of observations on which the estimated variance was based was less than 5, or the estimated variance was negative.

**True field spiked standards not available for these constituents, demanding combined estimates of laboratory and field variability.

pumping procedures used in purging and sample collection. This observation is particularly critical in relatively sparse datasets where annual "mean" concentrations may be determined from programs with low sampling frequency (i.e., annually, biannually, etc.). Similar cautions in interpretations of long-term datasets apply in the analysis of trends at varying or unequal sampling frequencies due to the relatively short duration of the records in comparison to the length of apparent annual to multi-year variations.

Sampling Frequency

The primary purpose of the project was to investigate the optimal sampling frequency for ground-water quality monitoring. Strictly speaking, there is no minimum sampling frequency. However, there is a relationship between the information content of the data and the sampling frequency. The term "information" is sometimes used loosely, but in a statistical context, it can be given a more precise definition, depending on the use of the data. The most common definition

of information (e.g., in the Fisher sense) is in terms of the variance of the mean, $Var(x) = \sigma^2/n$, where x is the sample mean, n is the sample size, and σ^2 is the variance of the data. The reciprocal of the variance of the mean is a measure of the information content of the data. If the σ^2 is large, or the sample size small, the information content is low. While this definition of information applies to estimation of the mean, the power of trend detection (in space or time) is related to the variance of the mean as well.

Seemingly, the information content of the data could be increased arbitrarily, since it depends linearly on the sample size. In practice, though, ground-water quality data are correlated in time (autocorrelated), and the autocorrelation increases with the sampling frequency. When the data are autocorrelated, the variance of the mean can be reexpressed as $Var(x) = \sigma^2/n_{ef}$, is an effective independent sample size, which depends on the autocorrelation. The value of n_{ef} is always less than n, the actual sample size, if the autocorrelation is positive, as it usually is in practice. If the model that

describes the autocorrelation is the l one Markov process, n_{ef} approaches upper limit as the sampling frequer increases, regardless of how large becomes. The lag-one process has be found to provide a reasonable descript of many water quality time series. It often difficult to extend the analysis water quality data beyond lag-one l cause the autocorrelation function l comes excessively noisy.

The ratio n_{ef}/n can be considered to a measure of the loss of information c to autocorrelation in the data. Althou n_{ef} always increases with n for posit autocorrelation, n_{ef} may increase qu slowly if the autocorrelation is high. I this reason, one of the analyses c ducted was to estimate a model of serial dependence (i.e., autocorrelati in the observed chemical series.

To illustrate the effect of the au correlation on sampling frequency, solved for the sampling interval, in wee that would result in ratios $n_{ef}/n = 0.5$, C and 0.9. Alternatively, these can be int preted as relative losses of informat due to autocorrelation in the data of

20, and 10 percent. The results are given in Table 2. At Sand Ridge, the implied loss of information was about 50 percent for many variables at a weekly sampling frequency, 20 percent for many variables at sampling intervals in the range of 4-8 weeks, and 10 percent for the majority of variables at a sampling interval of 8 weeks or more. At the Beardstown wells, the loss of information at high sampling frequencies was much greater. At the upgradient wells, which had the highest autocorrelation, the inferred loss of information of 50 percent occurred for several variables at a sampling interval of over 26 weeks. Information loss of between 20 and 10 percent was inferred for some variables at sampling intervals exceeding one year. This effect was particularly evident for Na^+, Cl^- and well-head temperature (TEMPW) which showed an increasing trend over the study period.

The results of the study indicate that, for the major chemical constituents (i.e., water quality or contaminant indicator), quarterly sampling represents a good starting point for a preliminary network design. Some estimated ranges of sampling frequency to maintain information losses below ten percent are shown in Table 3. This frequency, of course, must be evaluated with respect to the purpose and time-frame over which the network will be conducted. Under the conditions of this study, sampling four to six times per year would provide an estimated information loss below 20% and minimize redundancy. The results for reactive, geochemical constituents suggest that bimonthly sampling frequency would be a good starting point if chemical reactivity and transformation are of concern.

Caution must be exercised in interpretation of these results due to the effects of seasonality and long-term trends. However, it should be clear that there is considerable redundancy in the data at the two-week sampling interval, and that, at similar sites and for most of the variables studied, operational sampling programs would be inefficient at sampling intervals more frequent than bimonthly.

It is important to emphasize that the information from sampling depends on the effective independent sample size, not just the ratio n_{ef}/n. Therefore, if the autocorrelation is large so that a relatively low sampling frequency is necessary to avoid sampling redundancy, the total length of the sampling period must be increased to achieve sufficient information return. These results cannot simply be interpreted to mean, for instance, that quarterly sampling is adequate, unless that interpretation is couched in terms of the time horizon of the sampling program.

Table 2. Sampling Intervals (in Weeks) for Given Ratio of Effective to Independent Sample Size, Based on the Estimated Lag One Markov Model

| | n_{ef}/n | | |
	0.5	0.8	0.9
Sand Ridge			
NO_{2+}· N	2	4	5
Fe	1	1	2
pH	4	7	9
S⁻	2	3	4
NH_3	2	4	5
SiO_2	3	5	6
Mn_T	4	7	9
Probe O_2	3	5	7
t-PO_4	1	2	3
O-PO_4	1	2	3
Eh	3	6	8
NO_3NO_2-N	8	16	21
TOC	3	6	8
SO_4	5	9	12
Fe_T	2	3	4
K	2	4	5
Ca	3	6	8
Mg	4	7	9
Cl^-	2	3	4
Na	3	6	8
Alk	7	14	19
Ion Balance	7	14	19
Temp Cell	4	8	10
VOC	4	8	10
Cond	10	10	27
TOX	10	20	27
Temp Well	6	11	15
NVOC	6	11	15
Beardstown Upgradient			
NO_{2+}· N	3	6	7
Fe	15	29	39
pH	3	6	8
S⁻	3	5	6
NH_3	11	22	30
SiO_2	8	16	22
Mn_T	3	6	8
Probe O_2	6	11	15
t-PO_4	2	3	4
O-PO_4	2	3	4
Eh	5	9	12
NO_3NO_2-N	3	5	6
TOC	5	9	12
SO_4	4	7	10
Fe_T	21	42	56
K	19	38	51
Ca	26	53	71
Mg	23	47	62
Cl^-	53	107	144
Na	42	85	114
Alk	6	12	16
Ion Balance	6	12	16
Temp Cell	26	53	71
VOC	26	53	71
Cond	35	71	95
TOX	35	71	95
Temp Well	71	143	192
NVOC	71	143	192

Table 2. (continued)

	n_{ef}/n		
	0.5	0.8	0.9
Beardstown			
Downgradient			
NO_{2+} N	3	5	6
Fe	4	8	11
pH	2	3	4
S⁻	6	11	15
NH_3	2	4	5
SiO_2	2	4	5
Mn_T	2	3	4
Probe O_2	3	6	8
t-PO_4■	15	29	39
O-PO_4■	23	47	62
Eh	5	9	12
NO_3NO_2-N	3	6	7
TOC	5	9	12
SO_4■	4	7	9
Fe_T	6	11	15
K	7	13	18
Ca	6	11	15
Mg	5	11	14
Cl⁻	8	16	21
Na	5	11	14
Alk	8	16	22
Ion Balance	8	16	22
Temp Cell	10	19	25
VOC	10	19	25
Cond	8	16	21
TOX	8	16	21
Temp Well	9	18	24
NVOC	9	118	24

Table 3. Estimated Ranges of Sampling Frequency (in Months) to Maintain Information Loss at < 10% for Selected types of Chemical Parameters

Type of Parameter	Pristine Background Conditions	Contaminated	
		Upgradiant	Downgradient
Water Qulaity			
Trace constituents	2 to 7	1 to 2	2 to 10
(<1.0 mg·L-1)			
Major constituents	2 to 7	2 to 38	2 to 10
Geochemical			
Trace constituents	1 to 2	2	1 to 5
(<1.0 mg·L-1)			
Major constituents	1 to 2	7 to 14	1 to 5
Contaminant Indicator			
TOC	2	3	3
TOX	6 to 7	24	7
Conductivity	6 to 7	24	7
pH	2	2	1

Conclusions

Sampling and analytical errors car controlled to within about ±20% of annual mean inorganic chemical cons ent concentration in ground water if protocols are properly designed and (cuted. The use of previously publis guides for ground-water monitoring provide reproducible, accurate results such studies.

The results of the study concent mainly on inorganic chemical constitu(in ground water. The statisti characteristics of the time-series data reactive chemical constituents (e Fe(II), sulfide, H_2O_2, O_2 and NO_2⁻) close that temporal variability is o lower than the magnitude of concentra changes observed during purging stagnant water prior to sampling. 1 means that improper well purging result in gross errors and the introduc of artifacts into ground-water quality d sets.

6

Michael J. Barcelona, H. Allen Wehrmann, Michael R. Schock, Mark E. Sievers, and Joseph R. Karny are with the Illinois Department of Energy and Natural Resources, Champaign, IL 61820-7495.

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

EPA/600/S4-89/032