Environ Geol (2007) 52:761–777
DOI 10.1007/s00254-006-0514-8

ORIGINAL ARTICLE

Groundwater monitoring for cement kiln dust disposal units in karst aquifers

Wanfang Zhou · Barry F. Beck · Jie Wang · Arthur J. Pettit

Received: 8 August 2006 / Accepted: 11 September 2006 / Published online: 5 October 2006

Abstract Well-developed karst aquifers tend to be heterogeneous and consist of variable porosities. Groundwater monitoring and the associated data interpretations in such aquifers are often more complicated than porous medium aquifers. Collection of representative data in karst aquifers often requires monitoring at appropriately located wells and/or springs that are proven to connect to the groundwater system. Water samples are to be collected under different flow conditions, including base flow, high-flow, and low-flow. The sampling frequencies may vary from several months for base flows to minutes in response to recharge events. The groundwater monitoring program presented in this paper is for a cement kiln dust mono-fill site in a karst area of southern Indiana. Following dye tracing and extensive geophysical investigations, one spring was selected as a monitoring location. A second spring should be used as a monitoring location when the last cell of the mono-fill begins receiving the wastes. The paper discusses results from the first spring, at which nine background sampling events were completed to evaluate the natural variations of the water quality. Based on the background data, a statistical evaluation plan was developed for 12 water-quality parameters to determine the integrity of the landfill. The statistical power of the statistical analyses was evaluated by Monte Carlo simulations.

W. Zhou (✉) · B. F. Beck · J. Wang · A. J. Pettit
P.E. LaMoreaux and Associates, Inc.,
106 Administration Road, Ste. 4, Oak Ridge,
TN 37830, USA
e-mail: wzhou@pela-tenn.com

Keywords Karst · Landfill · Groundwater monitoring · Sampling and analysis plan · Statistical evaluation plan

Introduction

Well-developed karst aquifers tend to be heterogeneous and consist of variable porosities. Groundwater monitoring in such aquifers is challenging. Permitting of waste disposal units in karst areas requires extensive hydrogeological and geophysical investigations, sinkhole risk assessment, and an effective groundwater monitoring program. Many cases of groundwater contamination have been identified at landfill locations because of poor knowledge of karst conditions beneath the disposal sites, inadequate groundwater monitoring programs, and poor design of landfills (lack of proper liners, for example). Siting of a landfill in karst requires developing and demonstrating a comprehensive understanding of the groundwater flow system at the site. Permitting of the landfill unit is conditional on such a demonstration. In addition to an adequate characterization of the karst geology and hydrology, the investigation must produce recommendations on a water quality monitoring strategy to detect leakage from the landfill during normal and elevated (storm-recharged) water table conditions. A groundwater monitoring program must be in place at any landfill facility. The program should include sampling and analysis procedures to consistently monitor groundwater quality at the sites and statistical evaluation plans to make intelligent decisions on the integrity of the landfills. The monitoring locations, either monitoring wells or springs or both, are to be chosen to produce data representative of the aquifer.

The landfill site discussed in this paper accepts cement kiln dust (CKD) from a cement plant of southern Indiana, USA. According to US Environmental Protection Agency (USEPA) (1998), over 70% of cement plants are in karst areas. USEPA (1998) has identified at least 13 cases of groundwater contamination because of leakage from CKD landfills. CKD samples collected from the cement plant in 2001 indicate that the waste contains very high concentration of potassium (Table 1). The waste samples also contain nitrogen and heavy metals such as lead, zinc, copper, nickel, and selenium. Table 2 shows the analytical results of the leachate derived from the CKD. The leachate has pH greater than 12, and contains high concentrations of dissolved solids and sulfate. Heavy metals such as lead, arsenic, selenium, and cadmium are present in the leachate as well.

Geologically, the site is within a small and isolated karst drainage basin. Clayey overburden, up to 12 m thick overlies the St. Louis and Salem limestone bedrock. As shown in Fig. 1, caves, sinkholes, springs, and seeps are present on-site and in its vicinity. A substantial portion of aquifer recharge occurs through the sinkholes. Groundwater discharge occurs at the springs and seeps. Although certain springs flow continuously throughout the years, others at upper elevations only flow periodically when the uppermost portion of the aquifer becomes filled. A creek is present at the bottom of a small valley within the drainage basin and flows to a regional river ~5 km north of the site.

Table 1 Analytical results of CKD based on samples collected in 2001 at the study site

Parameters	Weight concentration (mg/kg)
% total solids	99.9
Total arsenic	3.8
Total cadmium	4.7
Total copper	11.2
Total lead	52.7
Total mercury	<0.25
Total nickel	8.8
Total selenium	7
Total zinc	24.7
Total molybdenum	<4
PCB	<0.033
Total nitrogen	87.4
Ammonia nitrogen	10.9
Extractable nitrate nitrogen	0.27
Total kjeldahl nitrogen	87.4
Total phosphorus	90
Total potassium	13,000

Detailed geophysical surveys and dye tracing tests determined two major springs, designated as QRS and QNS in Fig. 1, as the suitable monitoring points (Zhou et al. 1999, 2002). Flows and six field water-quality parameters were monitored at 15-minute intervals at the two springs, while groundwater levels were recorded at the same frequency at two monitoring wells installed upgradient of the landfill. The six field water-quality parameters are temperature, pH, specific conductance (SC), turbidity, oxidation reduction potential, and dissolved oxygen. Typical flows for the two springs under low-flow conditions are on the order of 19–38 liters per minute (lpm), and range as high as ~227–341 lpm under high-flow conditions, with occasional higher peaks. Depending on flow conditions, response time of the monitored springs to significant precipitation events ranges from almost immediate to ~2 h. The dye tracing tests further indicate that QRS drains the water from the whole landfill site, while QNS drains the water from a portion of the landfill (cell 3) (Zhou et al. 2002). While background water-quality data were collected at both springs, a phased monitoring approach is used for the compliance monitoring. In the phased approach, QRS is monitored for the whole life of the landfill, and QNS is used only after cell 3 of the landfill begins to receive CKD.

Table 2 Analytical results of leachate from CKD based on samples collected in 2000 at the study site

Leachate parameters	Average concentration (mg/l)
Chlorides	180
Total cyanide	0.003
Fluoride	2.67
Phenols	0.07
Dissolved solids	3,967
Sulfate	1,267
Sulfide	1.0
Barium	0.25
Boron	0.05
Copper	0.01
Iron	2.50
Manganese	0.01
Nickel	0.01
Silver	0.01
Sodium	30.00
Zinc	0.02
Arsenic	0.09
Barium	0.68
Cadmium	0.20
Chromium	0.22
Mercury	0.003
Lead	0.400
Silver	0.05
Selenium	0.18
Ph	12

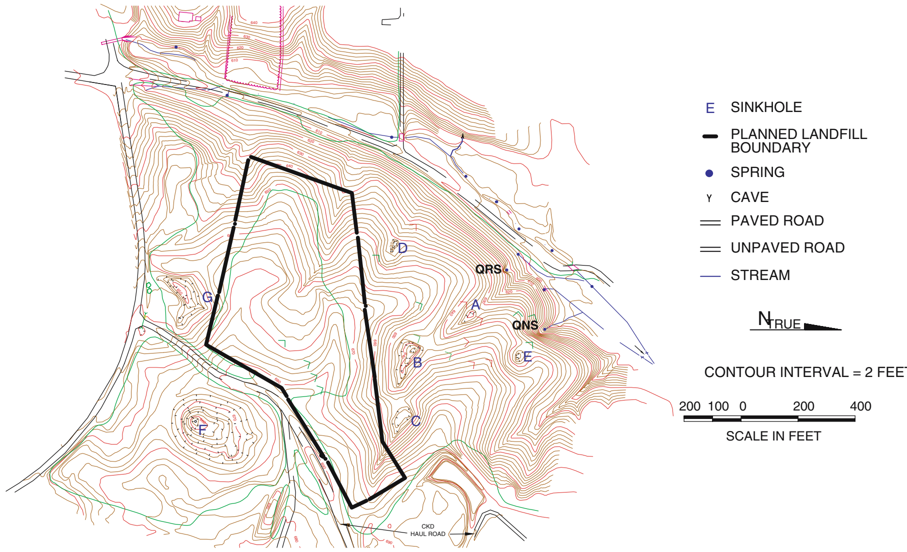

Fig. 1 Site map and groundwater monitoring locations

Background sampling

Nine background sampling events were conducted at the two springs to characterize the background data population. Six of these background sampling events were high-frequency, storm-pulse type events in response to a variety of aquifer, flow, and precipitation conditions, while three events were 24-h base-flow events with no precipitation. Following each background event, samples were judgmentally selected for laboratory analysis based on the spring discharge and field water-quality data collected during each event. Such a sampling approach allows characterization of any changes in natural groundwater chemistry under different flow conditions. Table 3 summarizes the background sampling efforts at one spring (QRS in Fig. 1) and the natural variations of the water-quality parameters. When data are presented in this paper, only data at QRS are presented for clarity.

These sampling approaches are non-conventional. To perform high-frequency, storm-induced sampling, accurate selection of the appropriate sampling events based on weather forecasts becomes important. This can be a major difficulty for any karst sampling program. Under some circumstances, individual storm-induced sampling events that have been started may need to be discontinued because forecasted precipitation does not materialize or a storm event may not produce a significant hydrographic or chemical response, even though significant precipitation occurred. Discontinuing a sampling event can cause significant unwanted expenses for clients, especially when significant mobilization times exist. A well-defined plan should be in place for monitoring the weather conditions. When a weather condition is deemed to be favorable for sampling, the flow conditions at the springs and in the aquifer should be immediately evaluated from the field data (groundwater level, discharge, and water-quality parameters) to predict the possible responses of the springs to the precipitation event.

Owing to the complicated nature of the sampling, sampling could occur at any time of the day or night. To facilitate smoother and more efficient handling of sampling related activities, communication between all parties involved is critical to ensuring that the weather is appropriately monitored at all times, and that sampling is initiated at the right times, regardless of the time that this occurs. A groundwater sampling contact list is an effective way to foster such communications (Lounsbury et al. 2003).

Characteristics of water-quality data at karst springs

Applicability of any quantitative analysis depends on the characteristics of the data being collected. Results from statistical procedures are meaningful only when the data conform to the assumptions of the statistical methods. The unique features of the data collected at the karst springs affect the suitability of many classical statistical techniques. Similar to many environmental monitoring data as described by Tasker and Granato (2000), pertinent characteristics of the water quality data at the karst springs include the following:

- *Discrete sampling*: A one-time grab sampling is of little use for estimating the water quality, flow rate and mass loadings because one instantaneous value

Table 3 Variation of water-quality parameters during background monitoring (unit in mg/l)

Event	Duration (days)	Number of samples	Discharge (lpm)	Alkalinity, bicarbonate ($CaCO_3$)	Alkalinity, carbonate ($CaCO_3$)	Alkalinity, total ($CaCO_3$)	Chloride	Fluoride	Solids, dissolved	Solids, total	Sulfate	Arsenic, ICP	Barium, ICP	Cadmium, ICP	Calcium, ICP
1	7.00	21	62–309	210–440	<1	210–440	17–58	0.29–0.38	280–1,000	390–1,008	7.4–180	<0.005	0.062–0.095	<0.005	100–150
2	0.96	4	88–95	180–210	<1	180–210	38–41	0.30–0.32	570–610	590–660	200–220	<0.005–<0.01	0.067–0.070	<0.005	100–110
3	6.79	23	67–234	130–250	<1	130–250	19–60	0.26–0.3	410–740	470–760	<50–250	<0.005–0.022	<0.02–0.13	<0.005	96–140
4	3.86	20	105–430	230–270	<1	230–270	8.4–22	0.23–0.27	380–490	420–510	84–210	<0.005	0.0534–0.0720	<0.001	98–110
5	4.03	20	45–199	190–260	<1	210–260	11–54	0.23–0.29	250–680	410–720	55–170	<0.005	0.0515–0.0818	<0.001	90.6–140
6	5.75	19	6–33	240–330	<1	240–330	43–96	0.25–0.32	570–1,000	650–1,100	150–320	<0.005	0.0597–0.101	<0.001–0.003	114–180
7	0.96	7	13	210–290	<1	210–290	56–59	0.3–0.34	640–670	710–750	170–200	<0.005	0.0634–0.0672	<0.001	103–150
8	0.92	6	42	190–220	<1	190–220	38–40	0.32–0.37	560–590	580–610	180.00	<0.005	0.0658–0.0689	<0.03	111–121
9	3.06	10	57–83	210–240	<1	210–240	32–39	0.33–0.34	530–580	580–620	170–190	<0.005	0.0673–0.0788	<0.001	108–116

Event	Duration (days)	Number of samples	Discharge (lpm)	Chromium, ICP	Copper, ICP	Iron, ICP	Lead, ICP	Magnesium, ICP	Manganese, ICP	Mercury, CVAA	Potassium, ICP	Selenium, ICP	Silver, ICP	Sodium, ICP	Zinc, ICP
1	7.00	21	62–309	<0.01	<0.01	<0.01–0.87	<0.005	12–18	0.021–0.21	<0.0005	49–130	<0.005–0.006	<0.005	9.7–30	<0.02–0.026
2	0.96	4	88–95	<0.01	<0.05	<0.1	<0.005	12.00	<0.01–0.11	<0.0005	80–85	<0.005	<0.005	16–17	<0.02
3	6.79	23	67–234	<0.01	<0.005–0.008	<0.1–0.47	<0.005–<20	11–18	0.018–0.12	<0.005	34–90	0.017–0.08	<0.005–<0.02	7.4–20	<0.02
4	3.86	20	105–430	<0.002	<0.005	<0.1–0.632	<0.001–0.0018	9.9–13.6	0.014–0.174	<0.0002	24.8–45.7	<0.005	<0.0005	5.2–10.2	<0.05
5	4.03	20	45–199	<0.002	<0.005–0.0064	<0.2–1.88	<0.001–0.0026	11.9–19.5	0.015–0.453	<0.0002	21.2–60	<0.025	<0.0005	5.2–17.6	<0.05
6	5.75	19	6–33	<0.002–0.005	<0.005–0.0564	<0.1–0.86	<0.001–0.0184	17.4–28.9	0.036–0.146	<0.0002	48.7–101	<0.025	<0.0005	13.9–29.5	<0.05–0.187
7	0.96	7	13	<0.002	<0.005	<0.1–0.1	<0.001	17.7–22.3	0.027–0.05	<0.0002	53–62.7	<0.01	<0.0005	16–20.1	<0.05
8	0.92	6	42	<0.002	<0.005	<0.1–0.12	<0.001	13.3–14.9	0.015–0.02	<0.0002	78.9–89	<0.005	<0.0005	15.8–17.6	<0.05
9	3.06	10	57–83	<0.005	<0.005–0.0134	<0.1–0.27	<0.001–0.0015	12.9–13.9	0.051–0.076	<0.0002	54.9–65.3	<0.005	<0.0005	12.7–14.5	<0.05

is not representative of the average conditions at the spring; even less, the variability of the concentration and mass changes cannot be described. Site-specific sampling and analysis plans may be needed to characterize the water quality of karst aquifers.

- *Response to recharge*: Water quality changes with the increase of discharge at the spring. The lag time may vary, depending on the characteristics of the karst aquifer. Knowledge of within-event constituent concentration fluctuations provides limited information to decision makers regarding control measures that might be required. Sampling under various flow conditions and extensive data analyses are essential to provide interpretations for intelligent decisions.
- *Non-repeatability*: Measurements at the karst spring are observational, not experimental because of the continuous movement of groundwater, antecedent condition, rainfall intensity, and its constant reaction with the surrounding environment. Exactly recreating the many natural and anthropogenic influences that affect each measurement is impossible. This is especially true for the karst aquifer under study. The water flow and water quality have a dynamic response to the rainfall events. Ancillary and metadata (information about a given data set, including explanatory information and data-quality information) pertinent to the statistical characteristics of the sampled population are investigated using the data from nine sampling events. For the reason that statistical regularity cannot be demonstrated through controlled experiments at the site, ancillary data are important to quantify possibly confounding variables that may preclude meaningful interpretation of data. Any modifications on the land use upstream of the landfill site should be documented.
- *A lower limit of zero*: Negative values are not possible for the water quality data at the springs. For example, negative values of precipitation, flow or concentration do not have meaning.
- *Censored data*: Limits in methods for samples collection and analysis cause data to be reported as either above or (more typically) below one or more reporting limits, which produces a censored population of data. The effect of censored data can be especially problematic for interpretation of water-quality data. Laboratory detection limits change with time and can be dramatically different from laboratory to laboratory and may even be different from method to method within a laboratory. Detection-limit artifacts affect statistical properties of individual data sets. When a data set contains values reported as less than one or more detection limits an overestimation of central-tendency measures and an underestimation of dispersion measures will be caused by truncation of the lower tail of the true population. Because the relative uncertainty in the accuracy and precision of individual values tend to increase as reported concentrations approach the detection limit, the percent error expected for measurements near detection limits is much higher than for values well within the measurement range of the method of analysis.
- *Meaningful outliers*: Valid measurements that are considerably higher or lower than most of the measured population are common among data sets at karst springs. Outliers can be extremely problematic in data interpretation. Outliers can arise from a variety of sources including but not limited to transcription errors, inconsistent sampling procedures, instrument failure, calibration or measurement errors, underestimation of spatial or temporal variability, and other factors. The presence of meaningful high-end outliers (actual but extreme values) contributes to the positive skew and is a factor producing non-standard distributions. High-end outliers represent times when, for example, regulatory criteria may be exceeded and the health of the local ecosystems may be affected. Those outliers produce a host of potential problems for interpretation of data sets (Barnett and Lewis 1998).

If an outlier is discovered to have a strong influence on the slope of a regression line (the slope or the correlation coefficient changes significantly when the point is omitted), then it must be determined whether the outlier represents extreme values for a single process or if a secondary process is characterized by the outlier. Measurement and documentation of explanatory variables such as precipitation and flow; real-time measures of water-quality characteristics such as SC, pH, temperature, and turbidity, use of ratios between constituents of interest, and results from a comprehensive quality assurance/quality control (QA/QC) program can be used to identify and explain outliers in terms of the potential effect of real physicochemical processes as opposed to the effect of sampling artifacts. Because of the complexity in sampling and analysis at the spring, a strict QA/QC measure is taken to ensure the meaningful outliers not to be excluded in the statistical analyses.

Elimination of outliers is considered a dangerous and unwarranted practice for the interpretation of water-quality data at the site, unless one has substantial objective evidence demonstrating that the outliers are not representative of the population under study. If outliers are not handled in an appropriate manner, unwanted, and potentially unnoticed bias in statistical calculations can occur, which could result in false-positive and/or false-negative detections.

- *Positive skewness*: Data sets that are not symmetrical around mean or median values are typical for spring water quality data sets because the combined effects of a lower bound of zero, censoring, and meaningful outliers tend to produce data sets in which the right tail of the distribution is extended and the left tail truncated.
- *Autocorrelation and independence*: An event is said to be independent of another event when the occurrence of one does not affect the occurrence of another. Spring peak flow-rates separated by a long period of time may be independent, but two peaks close to one another may not be independent. This is true when the recession limb of the first hydrograph at a spring becomes part of the rising limb of the next hydrograph. Natural and anthropogenic effects tend to cause conditions in which consecutive measurements are correlated. The natural and anthropogenic processes controlling groundwater quality and the methods for sampling, processing, and analysis often cause problems with autocorrelation. Autocorrelation is also referred to as serial correlation or correlation—the dependence of residuals in a time sequence because data reflect the effects of preceding conditions. Time-series effect may also occur between subsequent samples within individual sampling events. When discussing stormwater runoff in Texas, Barrett et al. (1993) indicate that the duration, the volume of runoff, and the intensity of the runoff are significant causal variables in a regression model. Autocorrelation can be important because it affects the optimization of regression coefficients, affects estimates of population variance, invalidates results of hypothesis tests, and produces confidence and prediction intervals that are too narrow for the real population being sampled.
- *Interdependence*: Changes in one characteristic of interest such as rainfall intensity, antecedent flow conditions, or temperature cause changes in other characteristics such as measured flows and concentrations. The interdependence makes it very challenging to characterize the natural variations of water quality at karst springs.
- *Temporal variation*: Temporal variation of water quality is a characteristic of many karst springs. Measured water-quality characteristics vary at time scales of hours, days, seasons, years, and even decades because of both natural and anthropogenic influence. Temporal variation may also increase variability in data and affect the comparability of data between sites.

Flow-weighted concentrations

As shown in Table 3, the number of samples collected in each sampling event varies from 4 to 23. The samples in each sampling event are not statistically independent. A representative parameter that can characterize the water quality over each sampling event is necessary to statistically analyze the data. Because values of measured constituents vary with spring discharge and may be impacted by the antecedent soil moisture conditions, flow conditions, and other factors, a technique is needed to bring all the measurements from each event onto the same 'base level' such that they could be compared.

The approach to deal with these issues is to calculate flow-weighted concentrations (FWCs) for each individual sampling event. Although FWCs do not appear to have been used in regulatory-driven groundwater monitoring programs for statistical evaluation of water-quality data at karst springs, they have been accepted and widely used in storm water runoff evaluation.

For this groundwater monitoring program, sequential discrete water samples were collected across the hydrograph curve for each storm-pulse sampling event. For base-flow sampling (during periods of no precipitation), samples were collected at constant intervals over a period of ~24 h. In each sampling event, one discharge hydrograph (discharge versus time) was obtained, together with one chemograph (concentration versus time) for each water-quality parameter. From the discharge hydrograph and chemograph, the load graph (loading rate versus time) was calculated. The loading rate is defined as the product of concentration and discharge at the time any individual sample is collected. The FWC values were then determined by dividing the cumulative mass of a constituent (loadgraph) by the total volume of water (area under the hydrograph):

$$\mathrm{FWC} = M/V. \tag{1}$$

In Eq. 1, M is the cumulative mass of a water-quality parameter, which was calculated by:

$$\begin{aligned} M &= \sum_{i=1}^{n-1} (L_i + L_{i+1})(t_{i+1} - t_i)/2 \\ &= \sum_{i=1}^{i=n-1} (Q_i C_i + Q_{i+1} C_{i+1})(t_{i+1} - t_i)/2. \end{aligned} \tag{2}$$

V is the total volume of water discharged at the spring, which was calculated from the discharge hydrograph by:

$$V = \sum_{j=1}^{m-1} (Q_j + Q_{j+1})(t_{j+1} - t_j)/2, \quad (3)$$

where, L is the loading rate; Q is flow rate; C is concentration of a water-quality parameter; t is time; index i the ith water sample; index j is jth flow measurement; n is number of samples collected; and m is the number of flow measurements taken.

The FWCs calculated from the above equations represent the average concentrations in the receiving waters when the waters are completely mixed, according to the basic principle of mass conservation. Using either the arithmetic or geometric mean as the average concentration in the receiving waters cannot guarantee the mass balance. The FWC equals the arithmetic mean when the spring flow rate remains unchanged throughout the entire sampling event, as may be observed during base-flow conditions. The geometric mean is always smaller than the arithmetic mean unless all numbers in a dataset are identical (Parkhurst 1998). The FWC values are often between the arithmetic mean and the geometric mean for the background dataset in which both flow and concentration vary with time.

The calculated FWC values for the 11 parameters for the nine background sampling events are provided in Table 4. Table 5 presents the calculated FWCs for SC for 62 selected rain events during the background monitoring period. Statistical evaluation plans were developed for the 12 water-quality parameters listed in Tables 4 and 5.

In calculation of FWCs, several issues need special attentions because of the unique characteristics at karst springs. These issues include factors affecting water quality within each sampling event, outliers, censored data, and independence between sampling events.

Factors affecting water quality within individual sampling events

The geochemistry of karst springs is notorious for its high degree of variability, as shown in Table 3. Care must be taken to develop a statistical approach that accommodates the expected natural water-quality variations without leading to false conclusions that groundwater has potentially been impacted by the mono-fill. Water quality at the springs of concern was linearly correlated to the following factors to evaluate the relative importance of these factors and their effect on natural water quality at the springs:

- spring flow;
- rainfall intensity and duration;
- aquifer water levels;
- antecedent conditions;
- soil-moisture conditions;
- seasonal weather patterns;
- water temperature;
- turbidity;
- barometric pressure;
- evaporation;
- air temperature;
- wind-speed.

The results of these analyses indicate that discharge is the major factor that affects water quality during precipitation events. Temperature, turbidity, and barometric pressure were also found to be factors affecting water quality, but to a lesser degree.

Independence by residence time

Most statistical methodologies assume that sampling events are unique, randomly distributed, and completely independent. Violation of this assumption would render selected statistical tests invalid, unless appropriately compensated for. Independence tests completed for this study involved the calculation of residence times of groundwater within the karst aquifer. Residence time is effectively the average amount of time a particular substance travels within the groundwater system.

Because the average residence time is related to flow conditions and various transport processes, the SC measured at the springs was selected as the representative parameter for residence time calculations. It is normally accepted that SC is a representative and sensitive parameter to characterize karst aquifers (Quinlan et al. 1991). At spring QRS, the background conductivity is as high as 900 μS/cm, while the local precipitation has *SC*-values between 20 and 30 μS/cm, which makes SC a reasonable tracer for residence time calculation. Linear correlation analyses between conductivity values and other water-quality parameters within the karst aquifer further confirmed the applicability of conductivity as a representative parameter for residence time calculations.

When discharge at the springs changes in response to rain events, the residence time is not a single value but a distribution. The mean of the residence time distribution can be calculated by (Field 2002):

Table 4 Calculated flow-weighted concentrations for 11 selected parameters for nine sampling events

Event	Alkalinity, bicarbonate ($CaCO_3$) (mg/l)	Alkalinity, carbonate ($CaCO_3$) (mg/l)	Alkalinity, total ($CaCO_3$) (mg/l)	Magnesium, ICP (mg/l)	Potassium, ICP (mg/l)	Selenium, ICP (mg/l)	Sodium, ICP (mg/l)	Sulfate (mg/l)	Chloride (mg/l)	Solids, dissolved (mg/l)	Calcium, ICP (mg/l)
1	257.79	0.5554	257.79	20.87	71.75	0.0025	22.57	131.07	53.58	772.06	143.18
2	195.13	0.9721	195.13	18.52	81.68	0.0049	22.16	211.74	65.66	726.50	120.79
3	211.84	0.4297	211.84	20.17	53.50	0.0544	17.91	149.07	55.52	698.33	129.51
4	255.74	0.5026	255.74	23.00	33.97	0.0025	17.59	93.04	56.09	675.72	134.10
5	242.07	0.4039	242.07	20.35	33.54	0.0101	15.10	107.42	46.14	619.92	132.36
6	273.90	0.3237	273.90	22.69	65.84	0.0081	20.39	205.54	64.51	747.77	140.52
7	254.78	1.0000	254.78	22.16	59.53	0.0100	19.90	188.91	62.94	683.71	129.21
8	203.18	1.0000	203.18	21.86	84.95	0.0050	24.14	180.00	68.93	748.61	138.21
9	232.05	0.8328	232.05	22.52	60.62	0.0041	21.85	177.78	71.04	760.61	137.95

Table 5 Calculated flow-weighted concentrations for SC for 62 rain events

Event	FWC of SC (μS/cm)	Event	FWC of SC (μS/cm)	Event	FWC of SC (μS/cm)
1	431.61	22	265.57	43	943.80
2	622.01	23	381.99	44	687.15
3	658.35	24	586.65	45	747.90
4	676.92	25	431.25	46	640.98
5	619.29	26	377.12	47	557.81
6	702.52	27	254.08	48	441.12
7	676.66	28	271.77	49	429.95
8	648.86	29	344.68	50	821.47
9	443.58	30	346.43	51	659.06
10	550.20	31	344.17	52	385.94
11	602.05	32	307.83	53	407.96
12	461.06	33	405.66	54	422.50
13	389.74	34	217.23	55	461.27
14	421.77	35	202.88	56	509.65
15	394.34	36	278.47	57	415.65
16	541.30	37	743.30	58	530.21
17	662.92	38	724.77	59	667.69
18	606.88	39	1,164.63	60	438.83
19	452.44	40	1,131.74	61	392.77
20	274.93	41	799.04	62	372.11
21	371.16	42	769.34		

$$\bar{t} = \frac{\sum_{i=1}^{N} t_i Q_i (SC)_i \Delta t_i}{\sum_{i=1}^{N} Q_i (SC)_i \Delta t_i}, \tag{4}$$

where, Q_i and $(SC)_i$ are the discharge and conductivity value at time t_i, respectively; and Δt_i is the time interval between two subsequent readings.

The standard deviation (σ) of the residence time is calculated by:

$$\sigma = \left[\frac{\sum_{i=1}^{N} (\bar{t} - t_i)^2 Q_i (SC)_i \Delta t_i}{\sum_{i=1}^{N} Q_i (SC)_i \Delta t_i} \right]^{1/2}. \tag{5}$$

The above equations were used to calculate the residence time distributions of 62 precipitation events for QRS (Fig. 2). The mean residence time varies from 0.2 to 8.7 days, with an average of 2.7 days. For majority of the rain events, the groundwater residence time is between 2 and 3 days. A standard deviation is also associated with each residence time distribution, and their relationship is shown in Fig. 3. In general, the standard deviation increases with the increase in the mean residence time.

The results from residence time analysis provide the time frames when the majority of the water from pre-

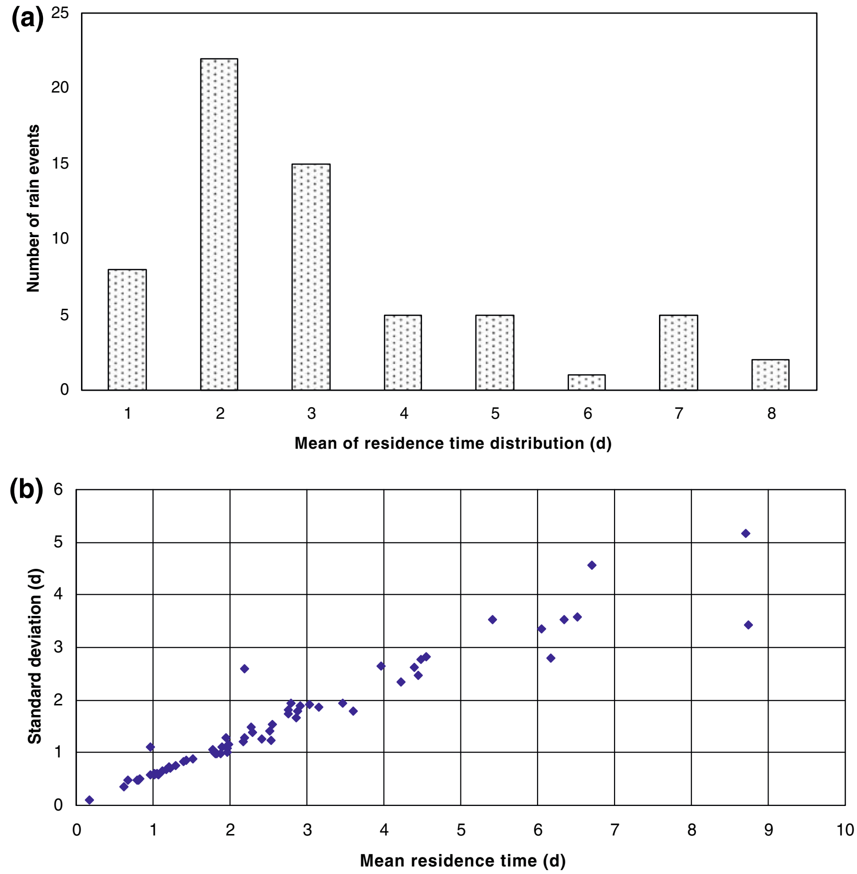

Fig. 2 Residence time calculations for 62 rain events: **a** histogram of mean residence times; **b** mean residence time versus standard deviation

vious precipitation events has completely discharged at the spring. To maintain independence between sampling events, major precipitations should not occur within several days prior to a sampling event; and subsequent sampling events will not be initiated until spring discharge levels and field water-quality parameters have returned to at least 90% of the pre-storm levels. For most storm-pulse events, a period of 5–7 days without precipitation will be sufficient to allow conditions to stabilize.

Censored data

Censored data are present in the background data sets for three of the 11 parameters subject to statistical analysis. Sulfate, selenium, and carbonate alkalinity contain 1, 80, and 100% censoring of data, respectively. Although censored data are quite common in environmental monitoring data, there is still considerable disagreement in the literature regarding appropriate methods for the handling of censored data in statistics.

Review of Gibbons and Coleman (2001) indicates that the specific amount (or percentage) of censoring for any given parameter is a key factor in determining how the censored data for that parameter should be handled. More specifically, parameters with up to 15–20% censoring need to be handled differently from those parameters with 50% or more censoring.

A commonly used method, for handling low levels of censored data is to use a simple substitution method where ½ of the detection limit (C_{DL}) is used in place of the censored value. This substitution method is generally accepted and used by many regulatory agencies including the USEPA and other state regulators. This procedure generally yields reasonable results in most cases when censoring is <20%. However, Gibbons and Coleman (2001) propose that the quantitation limit (C_{QL}) should be used as the censoring mechanism instead of the detection limit. The reasons are:

- values above C_{DL} and below C_{QL} are detected but not quantifiable;
- use of C_{DL} produces data with a wide variety of uncertainty and also violates the assumption of homoscedasticity (Gibbons and Coleman 2001).

The concentrations of carbonate alkalinity are less than 1.0 mg/l, its C_{QL}, for all background samples. The concentrations of selenium are above its C_{QL} in each sample from event 3 (Table 3), whereas, they are below C_{QL} in the other eight background sampling events. For data that contain 100% censoring, the simple substitution method is not applicable.

Correlation analysis of data for each storm-pulse sampling event indicates that discharge is the main factor that affects concentrations of the water-quality parameters at QRS. Although such correlation could not be established between carbonate alkalinity and selenium because of censored data, it is reasonable to assume that their concentrations are also related to the discharge at the spring. Because both carbonate alkalinity and selenium can be considered to be trace elements when they are below their C_{QL}, it can further be assumed that the total mass of these elements remains constant in the karst aquifer at any given time. It is also assumed that the censored elements are completely mixed within the aquifer. Under these assumptions specifically for carbonate alkalinity and selenium, the loading rate remains constant, therefore:

$$Q_i C_i = Q_j C_j, \tag{6}$$

where, Q is the discharge at the spring, C is concentration for either carbonate alkalinity or selenium, i and j are the time indexes when water samples were collected and $i \neq j$. It should be noted that C is a value below C_{QL}.

At the site, it is reasonable to assume that this maximum concentration corresponds to the smallest discharge at the spring within a sampling event. When a compound is present at unquantifiable values for all the samples within a sampling event, C_{QL} can be considered as the maximum concentration, as used in many environmental applications (Gibbons 1994; Helsel and Hirsch 2002). The concentration at any sampling point within each sampling event was then calculated by:

$$C_i = \frac{Q_{\min} C_{\mathrm{QL}}}{Q_i}, \tag{7}$$

where $Q_{\min}$ is the minimum flow rate at the spring among the samples within each individual sampling event.

The obtained value for each censored sample was then used to calculate the FWCs for carbonate alkalinity and selenium.

Sensitivity analyses

The prerequisite for the FWCs to be regarded as variables in statistical analyses is that they are independent and random. Necessary adjustments should be made if the FWCs are affected by natural factors. Based on the site characteristics and all available data, the following factors were evaluated to determine sensitivity of the FWCs by linear correlation analyses:

- Sampling event duration.
- Average spring water temperature.
- Residence time.
- Flow conditions: the flow conditions are defined by:
  a. Discharge at the beginning of the sampling event;
  b. Peak discharge over the sampling event;
  c. Discharge at the end of the sampling event;
  d. Total volume of water discharged during the sampling event;
  e. Time elapsed to reach peak flow;
  f. Time elapsed for increased discharge following onset of precipitation.
- Precipitation conditions at site and adjacent areas: for each weather station, the following parameters are considered the rainfall condition:
  a. Total rainfall;
  b. Peak rainfall intensity;
  c. Rainfall intensity.
- Antecedent moisture conditions: the antecedent moisture conditions are characterized by the following parameters:
  a. Dry days preceding the sampling event;
  b. Rainfall total 10 days preceding the sampling event;
  c. Rainfall total 20 days preceding the sampling event;
  d. Rainfall total 30 days preceding the sampling event.
- Aquifer storage conditions: the average water levels in monitoring wells M1 and M4 are used to represent the average storage condition of the aquifer.
- Sampling methods: starting time, ending time, and sampling frequencies.
- Methods to select water samples for laboratory analysis: constant interval selection versus judgmental selection.

The correlation analyses indicate that FWCs of the 12 water-quality parameters in Tables 4 and 5 are not sensitive to the factors listed above except the following six parameters:

- SC,
- magnesium,
- sodium,
- chloride,

- total dissolved solids (TDS), and
- calcium.

The FWCs of the six parameters listed are linearly related to the water levels in monitoring well M1. Their values tend to decrease with the increase of the water level in the aquifer. Table 6 provides equations to accommodate the impact of the water level in M1 on the FWCs.

Karst aquifers like the one under study are often conceived and modeled as storage reservoirs penetrated by trunk conduits (Smart 1999). In response to recharge events, conduits permit exceptionally rapid transfer of water and chemical constituents, while the storage reservoirs (fractures and matrix blocks), which contain the majority of water in the karst aquifer, slowly adjust to autogenic recharge from sinkholes and backflooding from the primary conduits. During recession periods, the head loss in conduits is often much lower than in the surrounding fractures or matrix blocks and the water stored in the matrix gradually drains into the conduits to sustain the spring flow. If time permits, equilibrium between the water in the conduits and that in the matrix occurs, and a unified water level is obtainable. The constant exchanging process also leads toward equilibrium between the chemical constituents within the conduits and matrix.

Because the majority of water is stored in the non-conduit portion of the karst aquifer, the water level in monitoring well M1 is an indication of the storage conditions. The unconfined nature of the aquifer makes the relationship between the volume of water stored in the aquifer and the water level more obvious. Like surface reservoirs, the volume of water in the aquifer is directly proportional to its water level. Because of a dilution effect in the aquifer, the concentrations of the water-quality parameters tend to be larger when the total volume of water stored in the aquifer is smaller (or under low-flow conditions). This relationship between chemical concentration and water level in M1 explains why the average water level in the aquifer is related to the FWCs for the six mentioned parameters. This dilution effect is also a reflection of the seasonal effects. Simple binary division into wet and dry seasons or high- and low-flow conditions may over-simplify the site conditions.

Table 6 Adjustments for water-quality parameters affected by aquifer storage conditions

Parameter	FWC adjustment equation
Calcium	FWC* = FWC + 6.88 (H-190)
Dissolved solids	FWC* = FWC + 52.2 (H-190)
Chloride	FWC* = FWC + 9.3 (H-190)
Sodium	FWC* = FWC + 2.22 (H-190)
Magnesium	FWC* = FWC + 2.35 (H-190)
Specific conductance (SC)	FWC* = FWC + 111 (H-190)

FWC flow-weighted concentration calculated from discrete water samples, *FWC** flow-weighted concentration adjusted by the average water level in piezometer M1, *190* the lowest historical value in M1, *H* average water level in piezometer M1 (m)

The linear correlation between the average water level in the aquifer and the FWCs is more significant in 6 of 12 water-quality parameters. It is obvious that other geochemical processes are involved in the aquifer, especially for alkalinity (carbonate, bicarbonate, and total), selenium, potassium, and sulfate. Karstification in a carbonate aquifer is the process of mineral dissolution. The dissolution process is, however, complex and is affected by many factors (Dreybrodt 1988). Gypsum layers are known to be present at the site, which further complicates the geochemical interactions of various constituents. As in most limestone aquifers (Drever 1997), water at the site contains total alkalinity that approximately equals the bicarbonate alkalinity. This is because carbonate alkalinity is only present in trace amounts below a pH of 8.3, and as a result, carbonate alkalinity is below the quantitation limit for all of the samples at the site. Because of the high percentage of censoring in the carbonate alkalinity data, the linear correlation analysis is inconclusive. The same applies to selenium, which has ~80% censored data.

Temporal variation

Evaluating the nature of possible effects in environmental data caused by temporal variation can be useful in data interpretation and in understanding site conditions. For data obtained during the nine background sampling events, little can be done to evaluate the specific effects of temporal variation because the sampling events were random. To properly evaluate temporal variation, data need to be obtained at regular frequencies such as quarterly, monthly, weekly, or hourly. Although temporal variation could not be evaluated directly, the calculation and use of FWCs may have compensated for any temporal variation in the analytical data. As indicated in the sensitivity analyses, six water-quality parameters are affected by the aquifer storage conditions, which already reflect the seasonal impacts to a certain extent.

Statistical techniques

Statistical methods are used to determine whether statistically significant increases in chemical parameters are occurring for any given sampling event, with the overall goal of balancing false-negative and false-positive detections.

Probability distribution

Statistical methods should be developed with careful consideration of the distributional properties of the data. This is because many of these methods assume that measurements are normally distributed, or that they can be mathematically transformed into normal distributions. In this study, FWCs are the variables used in statistical analyses. For SC, TDS, chloride, sodium, magnesium, and calcium, the adjusted FWCs are the statistical variables.

Several methods for testing non-normality of environmental data have been described (USEPA 1992; Gibbons 1994). No single method, however, is suitable for generic use because of the complexity in data patterns. The primary reason to test FWCs for normal distribution is to determine whether parametric test procedures can be employed. The null hypothesis for all tests of normality is that the data are normally distributed (Helsel and Hirsch 2002). Rejection of this hypothesis indicates that it is doubtful that the data are normally distributed. However, failure to reject the hypothesis does not prove that the data are normally distributed, especially for small sample sizes. It simply says that normality cannot be rejected with the evidence at hand.

Four methods were used to search for non-normality including Histograms, Empirical Cumulative Distribution Function Plots, Probability Plots, and Shapiro–Wilk Goodness-of-fit tests. The first three methods provided a visual assessment of the data distribution, while the more rigorous Shapiro–Wilk-tests computed a test statistic to determine whether the normality assumption is unacceptable to a significant level. Histograms were not used as a reference tool for judging the normality of the data. Additionally, the limitations of the Shapiro–Wilk-test procedure have been acknowledged as these tests that do not really provide a measure of whether the fit is good enough to allow accurate use of normal-based statistics (D. F. Parkhurst, personal communication 2004). A summary of the non-normality test results is presented in Table 7.

Statistical quality-control

It is very important to recognize that on the basis of a statistical analysis alone, it can never be concluded that a waste disposal facility has or has not impacted groundwater (ASTM 1998). A statistically significant increase over background levels indicates that the new measurement from a particular monitoring location is inconsistent with chance expectations based on the background data available. Although carefully designed statistical tests and procedural consistency will help in minimizing false positives and false-negatives, knowledge in how to apply these specific techniques to a karst groundwater setting is also critical.

Like groundwater monitoring programs at other waste disposal facilities, the operating permit for the mono-fill requires statistical methods as the basis for investigating the mono-fill integrity. Statistical tests are to be performed on each of the 12 water-quality parameters in Tables 4 and 5. Statistical decision procedures for these tests were based on the properties of data collected during the background monitoring collection period. Following future statistical tests made after each compliance-monitoring event, appropriate decisions will be made to determine whether additional investigations or regulatory actions are required. Therefore, it is essential that a realistic and workable model for the measurement data be formulated and used both in the construction and evaluation of decision procedures. Obviously, no statistical measurement model is perfect for a karst groundwater monitoring program. However, the model for a decision procedure should be reasonable and as simple as possible.

During the compliance-monitoring phase, statistical tests will be performed following each sampling event. Over the operating life of the mono-fill facility, a sequence of decisions, rather than just one decision, will be made. The one-decision case is a test of a hypothesis situation in which it makes sense to consider the significance level and power of the statistical test. Where decisions are made sequentially over time, one is dealing with quality-control schemes and should be more interested in distributions of run lengths (Starks 1988).

An in-control run length is the number of sampling periods from start-up until a decision is made, on the basis of water sample measurements, that additional regulatory action is required when, in fact, there is no leakage from the mono-fill. An out-of-control run length is the number of sampling periods from the time that a pollutant plume originating from the mono-fill discharges at QRS until a decision is made that additional regulatory action is required.

In comparing decision procedures for the mono-fill, consideration should be given to the distributions of their run-lengths rather than to the false-positive and false-negative probabilities on individual applications of the decision rule in each sampling period (Starks 1988). Naturally, one wants to use a quality-control

Table 7 Results of non-normality tests

Specific conductance (SC)	855.6	121.0	Normal
Alkalinity bicarbonate ($CaCO_3$)	236.3	27.5	Normal
Alkalinity carbonate ($CaCO_3$)	–0.49	0.44	Log-normal
Alkalinity total ($CaCO_3$)	236.3	27.5	Normal
Magnesium, ICP	21.3	1.5	Normal
Potassium, ICP	60.6	18.3	Normal
Selenium, ICP	–5.0	0.9	Log-normal
Sodium, ICP	20.2	2.9	Normal
Sulfate	160.5	42.6	Normal
Chloride	60.5	8.1	Normal
Solids, dissolved	714.8	49.4	Normal
Calcium, ICP	134.0	6.9	Normal

scheme that has, on average, long in-control run lengths and short out-of-control run lengths. From this perspective, it is apparent that the decision procedures pertaining to the groundwater monitoring program at for this site are in truth quality-control schemes.

Quality-control schemes have proved successful in many industrial applications (Lucas 1982, 1985; Lucas and Crosier 1982). Shewhart control charts, Cumulative Sum (CUSUM) control charts, and combined Shewhart-CUSUM control charts are typical graphical and statistical methods of assessing the performance of a system over time and have been widely used to maintain process control (Ryan 2002).

Combined Shewhart-CUSUM control charts were used for this study. The USEPA (1989) and ASTM (1998) recommend using combined Shewhart-CUSUM control charts for groundwater monitoring as an alternative to prediction or tolerance limits for intra-locational comparison of monitoring constituent concentrations. It takes advantages of good properties of both Shewhart and CUSUM control charts. The Shewhart control scheme is better than the CUSUM scheme in quickly detecting a large (>3σ) shift in the mean μ; whereas, the CUSUM scheme is usually faster in detecting a small change in μ that persists (Lucas 1982). Because the integrity of the mono-fill liner is potentially threatened by sudden or catastrophic failure, in addition to more gradual and less severe types of failure, it is essential to account for both large and small changes in the water-quality data. Combined Shewhart-CUSUM control charts are sensitive to gradual and rapid releases. Using both charts can reduce the false-negative rate and increase the overall correct-positive rate.

Design of combined Shewhart-CUSUM control charts

Design of a combined control chart requires determination of the following five parameters:

- $\overline{X}$ Estimated mean of the FWC or adjusted FWC from the background samples.
- S Estimated standard deviation of the FWC or adjusted FWC from the background samples.
- H The value against which the CUSUM will be compared.
- k A parameter related to the displacement that should be quickly detected.
- SCL The upper Shewhart limits, which is the number of standard deviation units for an immediate release. In this application, the upper control limits are of interest.

Table 7 shows the calculated mean $(\overline{X})$ and standard deviation (S) of the FWC or adjusted FWC for each water-quality parameter from the background sampling events. $\overline{X}$ and S are computed from nine sampling events, except SC that was calculated from 62 sampling events. They are used as reasonable estimates of their population values μ and σ (Helsel and Hirsch 2002). The means and standard deviations for carbonate alkalinity and selenium were computed as logarithms of the FWCs, as indicated in Table 5.

Let the FWC or adjusted FWC for any one water-quality parameter during the new sampling event i be x_i. Then the standardized difference z_i is calculated by:

$$z_i = \frac{x_i - \overline{X}}{S}. \tag{8}$$

And, the CUSUM Y_i is calculated by:

$$Y_i = \max[0, (z_i - k) + Y_{i-1}]. \tag{9}$$

In practice, $Y_0 = 0$ (Gibbons 1999), which ensures that only cumulative increases over the background are considered. If a process is in control, such that all future observations come from a normal distribution with a fixed mean and standard deviation, the quantity z_i in Eq. 8 is approximately distributed as a $N(0,1)$ random variable and bounces around 0. The quantity z_i– in Eq. 9 bounces around –k. As a result, the ith upper CUSUM Y will tend to bounce around 0 (Millard and Neerchal 2000).

The procedures can be illustrated graphically. The values of Y_i and z_i are plotted against t_i. An out-of-control situation is declared on sampling event i if for the first time, the cumulative increase of one water-quality parameter over its background $Y_i \geq h$ or $z_i \geq$ SCL.

The USEPA (1989) recommends using SCL = 4.5, k = 1, and h = 5, based on the recommendations of Lucas (1982) and Starks (1988). These values are suggested because they allow a displacement of two standard deviations to be detected quickly (USEPA

1992). For easy application, ASTM (1998) suggests the use of h = SCL = 4.5, which is slightly more robust in detecting leakage and thus reducing the rate of false negatives.

Unlike prediction limits, which provide a fixed confidence level (e.g., 95%) for a given number of future comparisons, control charts do not adjust for the number of future comparisons. The selection of $h = 5$, SCL = 4.5, and $k = 1$ is based on the USEPA's own review of the literature and simulation (Lucas 1982; Starks 1988). Since 1.96 standard deviation units correspond to 95% confidence on a normal distribution, there is ~95% confidence for this method as well for each comparison.

When the number of background sampling events is over 12, Starks (1988) and USEPA (1992) suggest using $k = 0.75$ and h = SCL = 4. Since 62 FWCs were calculated for SC in the background, $k = 0.75$ and h = SCL = 4 will be used for SC control charts. For the other 11 parameters, there are only nine background sampling events and the values recommended in ASTM (1998) are used.

It should be noted that these recommendations for values h, k, and SCL were based on analysis of a single statistical comparison. They do not account for adjusting the control chart parameters to account for the number of monitoring points and the number of constituents monitored during each sampling event (Gibbons 1994; Davis 1998). In practice, the statistical comparisons are often more than one. For this project, 12 statistical comparisons need to be performed at the end of each sampling event. ASTM (1998) suggests several possible modifications to control charts by allowing re-sampling and updating background data to attempt to control the overall false-positive rate and keep the statistical power high on each monitoring occasion.

Trigger for regulatory actions

Based on the relative importance of the 12 water-quality parameters of concern, regulatory actions are to be initiated under either of the two following scenarios:

- Selenium is out of control on the combined control chart.
- Three or more of the other 11 constituents of concern are out of control on the combined control charts.

Pooling

Pooling is a method of incorporating newly obtained data into the 'background' data set as a means of increasing the overall size of the background data set. Through pooling, uncertainty in the sample-based mean and standard deviation decrease, as does the size of the prediction limit, thereby minimizing both false-positives and false-negatives (ASTM 1998). Pooling is recommended by USEPA (1992) and Gibbons (1994). Pooling is recommended every 2 years once an appropriate amount of compliance monitoring data becomes available and it will continue for the life of the monitoring program. Pooling should only be performed if data and the monitoring process are shown to be in control.

Verification resampling

To help control false-positive and false-negative rates, the immediate re-collection of groundwater samples to refute or confirm unexpected measurement results, also known as verification resampling, has been recommended by many including the USEPA. However, in karst groundwater settings, verification resampling is generally not possible because groundwater concentrations are constantly changing in response to aquifer recharge and many other factors, and no single concentration can ever be duplicated for any given storm pulse sampling event. Although verification resampling cannot be completed for this study, minimizing false-positives and false-negatives was accomplished by:

- Stringent QA/QC in sampling and analysis,
- Appropriate/consistent outlier handling and treatment,
- Intra-spring statistical comparison,
- Use of FWCs or adjusted FWCs,
- Combined Shewhart-CUSUM control charts, and
- Pooling techniques.

Statistical power

To investigate statistical power, the combined Shewhart-CUSUM control scheme has been evaluated using Monte Carlo simulations. For these simulations, up to 20,000 adjusted FWC values of each of the 12 parameters of concern were generated based on properties as characterized during the background sampling phase. The simulation process was repeated 100 times, and information regarding the empirical distribution of the run lengths was obtained.

The simulation results indicate that the average run length of the developed control charts depends on the standard deviation unit, which reflects the amount of displacement in the concentration means over the background. As expected, the average run length decreases rapidly with the increase in the displacement of the concentration means. When the displacement of the mean concentrations is five standard deviation units, the control scheme can immediately detect the change at the next sampling event. When the displacement is relatively minor, for example 0.01 and 0.1, the average run lengths suggest that the control scheme has little chance to detect the change over the life of the groundwater monitoring project. Figure 3 shows the average run lengths versus the standard deviation units.

The USEPA (1992) has established a general guideline for using control charts in statistical approaches, in which an increase of two standard deviation units should be detected easily. The Monte Carlo simulations indicate that the average run length for the increase of two standard deviation units is 5.4 with the range from 1 to 15, which is very similar to values obtained by Starks (1988) when he suggested the values for *h*, SCL, and *k*. The simulation results obtained for this project are also comparable to those obtained by Lucas (1982) who calculated the average run lengths with the Markov chain approach (Brooke and Evans 1972). Figure 4 shows the distribution of the run lengths when the increase of the mean concentrations is two standard deviation units. Based on the 100 realizations in the Monte Carlo simulation, the probability of detecting an increase of two standard deviation units at QRS at the next sampling event is 2%. Assuming six compliance monitoring sampling are completed per year, the probabilities of detecting the same increase are 76 and 99% following the first and second years of compliance monitoring, respectively.

With an in-control state (standard deviation unit = 0), the average run length is ~4,248. If six sampling events are conducted annually, this average run length is

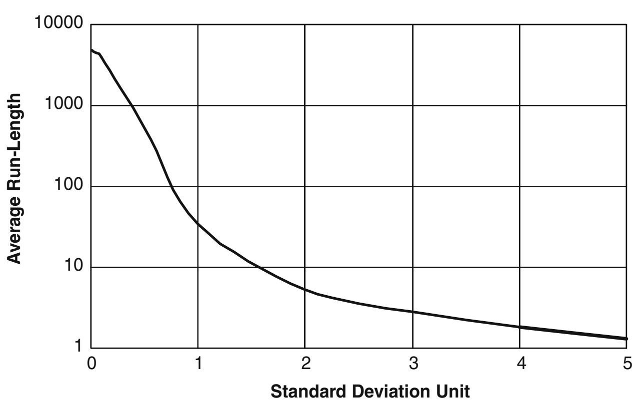

Fig. 3 Average run-length versus standard deviation unit based Monte Carlo simulations

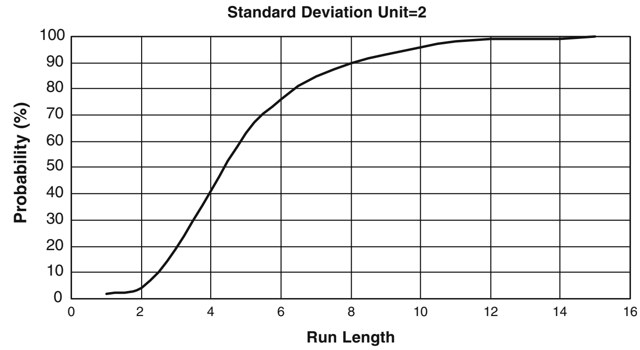

Fig. 4 Run length distribution when increase of mean concentrations is two standard deviation units

equivalent to over 700 years of sampling. On average, the chance to have a false-negative detection at QRS is negligible. Because the minimum in-control run length is 129 and two simulations (of 100) produced run lengths less than 360, there is a slight chance (~2%) that a false-positive detection may occur over 30 years of groundwater monitoring.

It is apparent that using lower control limits for SC has improved the ability of the statistical tests to detect potential leakage from the mono-fill, especially when the leakage is less than 0.1 standard deviation units. When the standard deviation units are equal to or greater than 0.5, the lower control limits for *SC* still change the run length distribution slightly. However, the average run lengths of an out-of-control state remain essentially the same. This is partially caused by the fact that the control chart for Selenium plays the dominant role when the release from the mono-fill is greater than or equal to 0.5 standard deviation units. Selection of lower control limits for *SC* can improve the statistical power of the tests.

Conclusions

Groundwater monitoring in karst is challenging. First, one needs to demonstrate that the groundwater system at the study site can be monitored. Both springs and wells are potential monitoring locations if:

- They are effectively connected to the groundwater system and,
- They are not impacted by any other disposal facilities.

Second, due to dynamic responses to recharge events, particularly at karst springs, multiple-parameter, long-term, and high-frequency monitoring may be required. Sampling and analysis plans should be designed to reflect the unique characteristics of the monitoring locations. Characterization of the natural variations in water quality may require sampling efforts under different flow conditions.

Third, evaluation of the potential impact of waste disposal units on the groundwater system requires an effective statistical evaluation plan. Due to heterogeneity of karst aquifers, inter-locational comparison is generally preferred to inter-locational comparison. In the study presented here, various specific quality-control procedures are developed for the springs using FWC based on either 9 or 62 background sampling events. It is the authors' intention that these approaches may shed light on groundwater monitoring programs in other karst areas, although they may not be directly applicable.

Acknowledgments The authors wish to thank the following people who participated in the project and contributed to its success at various stages: Brad J. Stephenson of Shaw Groups, Rich A. Lounsbury of AMEC, David Parkhurst of Indiana University, Joseph Woods of Indiana Department of Environmental Management, Angela L. Adams of MACTEC, Gheorghe Ponta and Desiree Sturdivant of P. E. LaMoreaux and Associates, Inc.

References

ASTM (1998) ASTM D 6312-98, Standard guide for developing appropriate statistical approaches for ground-water detection monitoring programs. Annual book of standards, West Conshohocken, PA, pp 1–14

Barnett V, Lewis T (1998) Outliers in statistical data. Wiley series in probability and mathematical statistics, 3rd edn. Wiley, West Sussex, England, p 584

Barrett ME, Zuber RD, Collins ER, Malina JF Jr, Charbeneau RJ (1993) A review and evaluation of literature pertaining to the quantity and control of highway runoff and construction. Technical Report CRWR 239, 1st edn. Center for Research in Water Resources, University of Texas, Austin, TX

Brooke D, Evans DA (1972) An approach to the probability distribution of CUSUM run length. Biometrika 59:539–549

Davis CB (1998) Power comparisons for control chart and prediction limits procedures. EnviroStat Technical Report 98-1. Environmetrics & Statistics Limited, Henderson, NV

Drever JI (1997) The Geochemistry of natural waters, 3rd edn. Prentice-Hall, Upper Saddle River, NJ, p 436

Dreybrodt W (1988) Processes in Karst Systems—Physics, Chemistry, and Geology. Springer Series in Physical Environment. Springer-Verlag, Germany

Field MS (2002) The QTRACER2 program for tracer-breakthrough curve analysis for tracer tests in karstic aquifers and other hydrologic systems. EPA/600/R-02/001

Gibbons RD (1994) Statistical methods for groundwater monitoring. Wiley Inc., New York

Gibbons RD (1999) Use of combined Shewhart-CUSUM control charts for ground water monitoring application. Ground Water 37(5):682–691

Gibbons RD, Coleman DE (2001) Statistical methods for detection and quantification of environmental contamination. Wiley Inc., New York

Helsel DR, Hirsch RM (2002) Statistical methods in water resources. Techniques of Water-resources investigations of the United States geological survey, book 4, hydrologic analysis and interpretation, Chap. 3. US Geological Survey, Reston, VA

Lounsbury RA, Pettit AJ, Zhou W, Beck BF (2003) Effective practices for initiating high-frequency, storm-induced sampling of two karst springs. In: Beck BF (ed) Sinkholes and engineering/environmental impacts of Karst. Geotechnical special publication #122, American society of civil engineers, Reston, VA, pp 328–338

Lucas JM, Crosier RB (1982) Robust CUSUM: a robustness study for CUSUM quality control schemes. Communications in statistics. Theor Method 11(23):2669–2687

Lucas JM (1982) Combined Shewhart-CUSUM quality control scheme. J Qual Technol 14:51–59

Lucas JM (1985) Cumulative Sum (CUSUM) control schemes. Communications in statistics. Theor Method 14(11):2689–2704

Millard SP, Neerchal NK (2000) Environmental statistics. CRC Press, Boca Raton, FL

Parkhurst DF (1998) Arithmetic versus geometric means for environmental data. Environ Sci Technol/News, Feb. 1, 92A–98A

Quinlan JF, Smart PL, Schindel GM, Alexander EC Jr, Edwards AJ, Smith AR (1991) Recommended administrative/regulatory definition of karst aquifer, principles for classification of vulnerability of karst aquifers, and determination of optimum sampling frequency at springs. In: Proceedings of the 3rd conference on hydrogeology, ecology, monitoring, and management of ground water in karst terranes. Nashville, TN, 4–6 December

Ryan TP (2002) Statistical methods for quality improvement, 2nd edn. Wiley, New York, p 446

Smart CC (1999) Subsidiary conduit system: a hiatus in aquifer monitoring and modeling. In: Palmer AN, Palmer MV, Sasowsky ID (eds) Karst modeling. Special publication 5. Karst Water Institute, Charles Town, WV, pp 146–157

Starks TH (1988) Evaluation of control chart methodologies for RCRA waste sites. Draft Report by Environmental Research Center, University of Nevada, Las Vegas, for Exposure Assessment Research Division. Environmental Monitoring Systems Laboratory—Las Vegas, Nevada. EPA Technical Report CR814342-01-3

Tasker GD, Granato GE (2000) Statistical approaches to interpretation of local, regional, and national highway-runoff and urban-stormwater data. Open-File Report 00–491. US Geological Survey, VA

USEPA (1998) Draft technical background document on groundwater controls at CKD Landfills. Office of solid wastes, US Environmental Protection Agency, Washington, DC

USEPA (1989) Statistical analysis of ground-water monitoring data at RCRA Facilities. Interim Final Guidance. EPA/530-SW-89-026, Office of solid waste, US Environmental Protection Agency, Washington, DC

USEPA (1992) Statistical analysis of ground-water monitoring data at RCRA Facilities: addendum to interim final guidance, in statistical training course for ground-water monitoring data analysis. EPA530-R-93-003, Washington, DC

Zhou W, Beck BF, Stephenson JB (1999) Application of electrical resistivity tomography and natural-potential technology to delineate potential sinkhole collapse areas in a covered karst terrane. In: Beck BF, Pettit AJ, Herring JG (eds) Hydrogeology and engineering geology of sinkholes and karst. A. A. Balkema, Rotterdam, pp 187–193

Zhou W, Beck BF, Pettit AJ, Stephenson JB (2002) A groundwater tracing investigation as an aid of locating groundwater monitoring stations on the Mitchell Plain of Southern Indiana. Environ Geol 41:842–851