

Application of water quality control charts to spring monitoring in karst terranes

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

Received: 20 December 2006 / Accepted: 20 March 2007 / Published online: 12 April 2007
© Springer-Verlag 2007

Abstract Presence of springs in karst terranes provides a unique opportunity to study the rather complex, multi-porosity, and multi-permeability system. When springs are used to evaluate the integrity of storage facilities for hazardous materials or waste disposal facilities constructed in karst areas, the spatial heterogeneity of karst aquifers makes intra-spring comparisons preferred statistical tests. One of the commonly used statistical tests is water quality control procedure such as Shewhart-CUSUM control charts. Appropriate application of the water quality control procedure to intra-spring monitoring depends on whether the assumptions can be justified about the aquifer that drains to the spring and the dataset collected at the spring. Violation of the assumptions would render the statistical tests invalid, which may result in a failure of the groundwater monitoring program. In intra-spring monitoring, it is the temporal variations of water quality at a karst spring need to be addressed, whereas the water quality at the spring is closely associated with the characteristics of the aquifer. The example datasets presented in the paper indicate that both false negative and false positive detections can occur if the temporal variation is not well characterized.

Keywords Karst spring · Groundwater monitoring · Sampling and analysis plan · Statistical tests

Introduction

Most naturally discharging groundwater from karst aquifers occurs from springs. They are natural exits for groundwater to the surface. Based on characteristics of outflows, controlling factors on formation of springs, and responses to recharge events, karst springs have been classified into many types (LaMoreaux and Tanner 2001; Bonacci 1987). Although each spring is unique from others, one common characteristic for karst springs is their dynamic nature. The synergistic relationship between circulation of water and dissolution of soluble rocks tends to lead to such geologic changes as lowering of water levels, enlargement of fractures/conduits, and formation of new springs (LeGrand and LaMoreaux 1975). A perennial spring once draining a whole karst aquifer may become an overflow spring that stops flow during low-flow conditions. The differential solution process, the greater the water flow the greater the dissolution, has the potential to create a hierarchy of porosities and permeabilities in karst aquifers. As karst development becomes mature, a majority of water tends to flow through a few conduits and discharge at a few large springs. Karst springs are products of and closely related to the inexorable dissolution processes of the soluble rocks. Karst springs may discharge the full flow of a drainage basin or share the discharge with other springs as part of a distributary network.

The presence of springs in a karst terrane provides a unique opportunity to understand the rather complex, multi-porosity, and multi-permeability system. Traditionally, monitoring of karst springs has been used to characterize the karst aquifer for purposes of evaluating water resource and preventing or remediating contamination of the karst aquifers. Data recorded at karst springs offers considerable potential insight into the nature and operation

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

of a karst drainage system. The characteristics of the outflow hydrograph recorded at a spring, including the shape, pattern, and rate of recession, provide invaluable information on (a) the storage and structural characteristics of the aquifer system sustaining the spring; (b) the components of autogenic and allogenic inputs at the spring; (c) the percentage of conduit flow versus diffuse flow or fracture flow; and (d) the characteristics of the spring drainage basin.

Water-quality monitoring at a spring is another important aspect of spring monitoring. The chemographs recorded at a spring provide additional information on the aquifer system, as well as the characteristics of solute transport of various water-quality parameters. The water-quality data can be used to determine (a) the residence time of groundwater, (b) potential anthropogenic impact on the spring, (c) remediation plans if the aquifer sustaining the spring is contaminated, and (d) whether groundwater remediation has met pre-established cleanup standards at a contamination site.

A second purpose of monitoring of karst springs is to evaluate the integrity of storage facilities for hazardous materials or waste disposal facilities constructed in karst areas to minimize possible impacts on receiving waters. Construction of these facilities requires not only extensive hydrogeological and geophysical investigations, ecological evaluation, sinkhole risk assessment, but also an effective groundwater monitoring program. Such groundwater monitoring programs are normally driven by regulatory requirements as a means of detecting statistically significant changes in water quality resulting from releases to aquifers by operation of these facilities (EPA 1991). Monitoring of karst springs can be and should be a component of a groundwater monitoring program. Based on several decades of experience working in karst areas, Quinlan (1990) concluded that “the only relevant locations to monitor groundwater quality in a karst terrane are springs, cave streams, and wells that have been shown by tracing tests to include drainage from the facility to be monitored—rather than at wells to which traces have not been run but which were selected because of their convenient downgradient locations.” In some circumstances karst aquifers may present monitoring and remediation environments that are technically superior to most granular aquifers because all discharges are through springs (Ewers et al. 1998). American Society for Testing and Materials (ASTM) recommends karst springs as alternative monitoring points when monitoring wells are unable to provide groundwater samples representative of the aquifer (ASTM 1995).

Spring monitoring is meaningful when the data collected at the spring can be used for its intended purposes. Programs for spring monitoring for landfill evaluation should

include sampling and analysis procedures to consistently monitor groundwater quality at selected springs and statistical evaluation plans to make intelligent decisions on the integrity of the landfill. Hydrologic connections between a landfill and a spring do not automatically guarantee that the spring is monitorable. Other waste disposal facilities in the drainage basin of the spring should be carefully inventoried, and the background water-quality data at the spring should be evaluated. Whether robust statistical tests can be developed for the spring is another important factor to be considered.

Statistical analysis is one tool that geologists and engineers use to determine water quality changes at a facility site. Statistical tests are not intended to be used in isolation of other types of meaningful evidence about a site impact. It is important to use statistical analysis in the context of the site’s hydrogeology to obtain meaningful results. However, the approach that is used to evaluate the landfill integrity determines how a groundwater monitoring program is implemented. A recent study on over 20 landfills indicates that the statistical choices significantly impact the facilities’ groundwater monitoring efforts (Horsey et al. 2001). Different statistical tests make different assumptions about the site hydrogeology and the properties of the data population. Misunderstanding these implicit assumptions can lead to a failure of the entire groundwater monitoring program.

Invalidity of inter-location comparisons in karst terranes

Landfill integrity is often evaluated by data comparisons, either inter-location comparison between up-gradient and down-gradient monitoring locations, or intra-location comparison within each of the down-gradient monitoring locations (Gibbons 1994). In inter-location comparisons, the water quality data collected at the up-gradient wells represents those not impacted by the landfill facility, while the down-gradient wells work on the assumptions that they intercept any contaminant release from the facility. Such a well-to-well comparison assumes that the hydrogeologic conditions in the aquifer being monitored are contiguous and uniform, and potential contaminants disperse down-gradient as a plume.

However, these assumptions are not appropriate in karst terranes, where significant spatial variation in water quality often occurs. Karst aquifers consisting of voids of various sizes are often 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 that 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. The data collected at different locations is not readily comparable if the monitoring wells intercept different types of porosities in the aquifer.

Groundwater in karst flows through discrete paths. Unless one of the down-gradient wells intercepts a conduit that receives water from the landfill site, the wells may not effectively detect any leakage. When springs are used as down-gradient monitoring locations, the scaling effect makes the inter-location comparison even less viable. Data collected at an up-gradient monitoring well may represent the background conditions immediately around the well, whereas data collected at a spring may be representative of a more regional drainage network. Many studies have indicated that permeability in karst aquifer increases with the size of measurement volume, and that the largest values result from analysis of data collected at springs (Sauter 1992). As a result, the dynamic responses (magnitude and lag time) to recharge events can be very different between the up-gradient well and the down-gradient spring. Under such conditions, it is difficult, if possible at all that the background-to-compliance comparisons discern between differences caused by spatial variation and differences caused by a facility impact. Inability of separating the natural variations from water-quality data may lead to erroneous conclusions about facility impacts if an inter-location comparison is used.

Secondly, inter-location comparisons require pooling of a large sample size of data over time to meet the minimum compliance sample requirements (Gibbons 1994). Collecting multiple samples within a single reporting period compromises the statistical requirement that the samples be physically independent. A lack of sample independence leads to reduced variability and ultimately to increased false positives. The site-wide false positive rates increase with the number of statistical tests being performed. If an inter-location statistical test is performed at two locations on ten constituents at a 5% false positive rate per constituent, the site-wide false positive rate can be calculated by $1 - (1 - 5\%)^{20}$, i.e., approximately 64%. Consequently, many facilities have more than a 50% chance of one or more false positives in each reporting period. The site-wide false positive rate is critical because the finding of a

statistically significant difference can potentially move a facility into retesting and/or assessment monitoring.

Intra-spring comparison

An intra-location analysis is fundamentally different from the inter-location analysis. While the inter-location analysis compares compliance locations against a background composed of up-gradient well data, the intra-location analysis compares each compliance point against a background composed of its own historical data. When an intra-location analysis, however, is used in a detection monitoring program, the implicit assumption is that the historical data that is used as background has not been impacted by the facility or any other facilities. The problem becomes complicated if the groundwater monitoring begins after waste had been placed at the facility or the spring to be monitored has been contaminated. The historical data needs to be proved to be “clean” before an intra-location analysis can be used.

The intra-location analysis eliminates the problems associated with the heterogeneity of a karst system. It identifies changes over time at a compliance point instead of changes between locations. The ideal situation to implement an intra-location analysis is at a new facility prior to the placement of any waste. Samples taken at compliance points prior to waste placement can be used to develop the intra-location limits.

When springs are used as the compliance points for intra-location analysis, it is aptly referred to as intra-spring comparison. The watersheds that drain to the springs should be thoroughly investigated to ensure that the springs could be monitored for the intended purposes. Monitoring springs need to be proved to drain the groundwater from the disposal facilities through tracer or other equivalent tests. Some large springs such as deep siphonal or vauculian springs may not be suitable for such a purpose because of their large drainage basins, in which many factors may affect the water quality at the springs. The water quality change caused by disposal facilities may not be significant enough to be detected. Springs that have relatively small and isolated drainage basins are more monitorable.

In intra-spring comparisons, it is the temporal variations rather than spatial variations in the monitoring data that challenge the development of robust statistical tests. Measured water-quality characteristics at springs may vary at time scales of hours, days, seasons, years, and even decades because of both natural and anthropogenic influence. Results from an intra-spring statistical test are meaningful only when the characteristics of the data population conform to the assumptions of the statistical methods.

An intra-spring statistical test is to determine whether statistically significant increases over background levels have occurred at monitoring springs. The statistical test results alone cannot be used to conclude that a waste disposal facility has or has not impacted groundwater (ASTM 1998). But statistically significant increases indicate that the new measurements at the spring are inconsistent with chance expectations based on the background data. If a significant increase occurs, additional investigations or regulatory actions are often required (EPA 1991). Realistic and workable statistical tests need to be carefully designed to minimize both false positive and false-negative rates.

Combined Shewhart-CUSUM control charts

Statistical tests are performed following each sampling event. Over the operating life of a landfill facility, a sequence of decisions, rather than just one decision, is made. The one-decision case is a test of a hypothesis situation in which the significance level is considered. Where decisions are made sequentially over time, one deals with 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; Montgomery 1997).

Combined Shewhart-CUSUM control charts are studied by Starks (1988) for their applications to landfill monitoring. They are the statistical procedures directly recommended by US EPA (1989, 1992) and ASTM (1998) for intra-location monitoring. As implied by its name, a combined Shewhart-CUSUM control chart is a mixture of two statistical schemes: Shewhart control scheme and CUSUM control scheme. The Shewhart control scheme is better than the CUSUM scheme in quickly detecting a large shift in the mean; whereas, the CUSUM scheme is usually faster in detecting a small change in the mean that persists (Lucas 1982). Because landfills constructed in karst terranes are potentially threatened by sinkhole collapses and by less severe types of leakage, both large and small changes may occur in the water-quality data. The sensitivity to both gradual and rapid releases of contaminants makes the combined scheme the choice of statistical tests for intra-spring monitoring in karst terranes.

In terms of individual applications of the decision rule in each sampling period, quality control scheme considers the distributions of run-lengths rather than the probabilities of Type I and II errors (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 landfill. An out-of-control run length is the number of sampling periods from the time that a pollutant plume originating from the landfill discharges at the spring until a decision is made that additional regulatory action is required. Naturally, one wants to use a quality-control scheme that has, on average, long in-control run lengths and short out-of-control run lengths.

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

- \bar{X} : Estimated mean of a water-quality parameter from background samples.
- S : Estimated standard deviation of a water-quality parameter from background samples.
- h : The value against which the cumulative sum will be compared.
- k : A parameter related to the displacement that should be quickly detected.
- SCL: The upper Shewhart limit, which is the number of standard deviation units for an immediate release.

For a new measurement x_i of a water-quality parameter i , the standardized difference z_i is calculated by

$$z_i = \frac{x_i - \bar{X}}{S} \quad (1)$$

And, the cumulative sum Y_i is calculated by

$$Y_i = \max[0, (z_i - k) + Y_{i-1}] \quad (2)$$

In practice, $Y_0 = 0$ (Gibbons 1999), which ensures that only cumulative increases over the background are considered. If a process is in control, the quantity z_i in Eq. (1) is approximately distributed as a $N(0,1)$ random variable and bounces around 0. The quantity $z_i - k$ in Eq. (2) bounces around $-k$. As a result, the upper cumulative sum Y_i will tend to bounce around 0 (Millard and Neerchal 2000).

The procedures can be illustrated by plotting the values of Y_i and z_i 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 $z_i \geq \text{SCL}$.

Assumptions in combined Shewhart-CUSUM control charts

The combined Shewhart-CUSUM control chart procedure assumes that the data are independent and normally distributed with a fixed mean and constant variance.

Interdependence tests

The most important assumption is independence. Violation of this assumption would render selected statistical tests invalid, unless appropriately compensated for. In porous medium aquifers where groundwater moves slowly, collecting independent samples may be challenging. However, in karst aquifer, groundwater moves relatively quickly, the challenge of collecting independent samples becomes less for springs, especially those that drain small areas.

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 are not. 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. 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.

Independence tests can be accomplished by calculating the residence time of water discharging at a spring. Residence time is effectively the average amount of time a particular substance travels within the groundwater system. The average residence time is related to flow conditions and various transport processes. It is important to select the most representative “tracer” for its calculation. Specific conductance (SC) measured at karst springs has long been considered as a representative and sensitive parameter to characterize karst aquifers (Quinlan et al. 1991). When SC at karst springs is one order of magnitude higher than SC in the precipitation, SC is a reasonable tracer for residence time calculation. Factors affecting the spring response in SC to recharge events include total precipitation, precipitation intensity, and antecedent aquifer conditions. Figure 1 shows typical variations of SC over six hydrographs at a karst spring. Because discharge at karst springs changes in response to rain events, the residence time is not a single value but a distribution. Calculation of residence time based on SC measurements needs to determine the beginning and ending times of a storm event. Defining of

the beginning and ending times can be sometimes difficult because of the base flow at springs and the complex responses to storm events. The spring at which the data in Fig. 1 were collected drains a relatively isolated area; thus the residence time is short, with the average varying from 0.2 to 9 days. The residence time distribution thus calculated provides general guidelines for selection of independent sampling events. For springs that drain a large area, the residence time estimation can be more complicated.

Normality tests

The assumption of normality is another concern in combined Shewhart-CUSUM control charts. If the measured data does not follow normal distribution, natural log or square root transformation or other methods should be conducted prior to developing the control charts. 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. Several methods for testing normality of environmental data including empirical cumulative distribution function plots, probability plots, and Shapiro–Wilk goodness-of-fit tests are described by Gibbons (1994) and EPA (1992). No single method, however, is suitable for generic use because of the complexity in data patterns.

Very often data collected at karst springs are not symmetrical around a mean because combined effects of a lower bound of zero, censoring, and meaningful outliers. The distribution of data is characterized by a right tail extended and a left tail truncated. 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 (Tasker and Granato 2000).

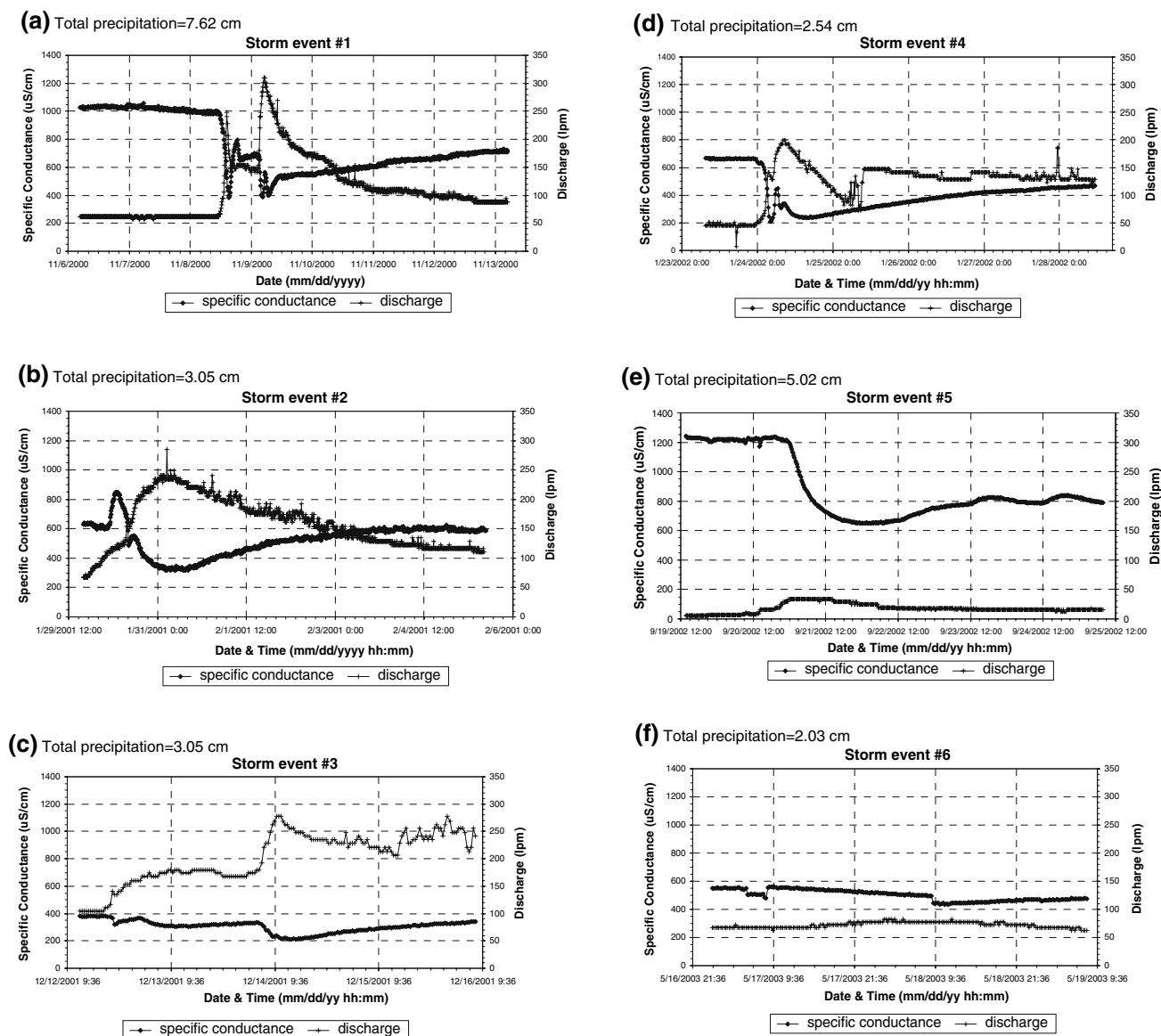


Fig. 1 Variations of specific conductance measured at a karst spring in response to various storm events

Gibbons and Coleman (2001) indicate 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. A commonly used method, for handling low levels of censored data is to use a simple substitution method where half of the detection limit (C_{DL}) is used in place of the censored value (EPA 1989). 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}) be used as the censoring mechanism instead of the DL because values above the C_{DL} and below the C_{QL} are detected but not quantifiable, and use of the C_{DL} produces data with a wide variety of uncertainty and also violates the

assumption of homoscedasticity (Gibbons and Coleman 2001).

Water quality at karst springs tends to be related to discharge. When a compound is present at unquantifiable values for all the samples within a sampling event, the C_{QL} is considered as the maximum concentration when the discharge is the smallest. The concentrations of the samples collected at other times can be calculated by multiplying C_{QL} by the ratio of the minimum discharge over the discharge at the sampling time. The calculated concentration is a value below the C_{QL} .

Outliers produce a host of potential problems as well for interpretation of data sets at karst springs. Outliers can arise from a variety of sources including transcription

errors, inconsistent sampling procedures, instrument failure, calibration or measurement errors, and underestimation of spatial or temporal variability. 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.

If an outlier is discovered in a dataset, 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 the outlier 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 springs, strict QA/QC measures are required to ensure the meaningful outliers not to be excluded in statistical analyses. Elimination of outliers is a dangerous and unwarranted practice for the interpretation of water-quality data at karst springs, 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 interpretations can occur, which could result in false-positive and/or false-negative detections.

Parameters in combined Shewhart-CUSUM control charts

Parameters h , k , and SCL

EPA (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 (EPA 1992). For easy application, ASTM (1998) suggested the use of $h = SCL = 4.5$, which is slightly more robust in detecting leakage and thus reducing the rate of false negatives. When the number of background sampling events is more than 12, Starks (1988) and EPA (1992) suggest using $k = 0.75$ and $h = SCL = 4$.

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 US EPA’s own review of the literature

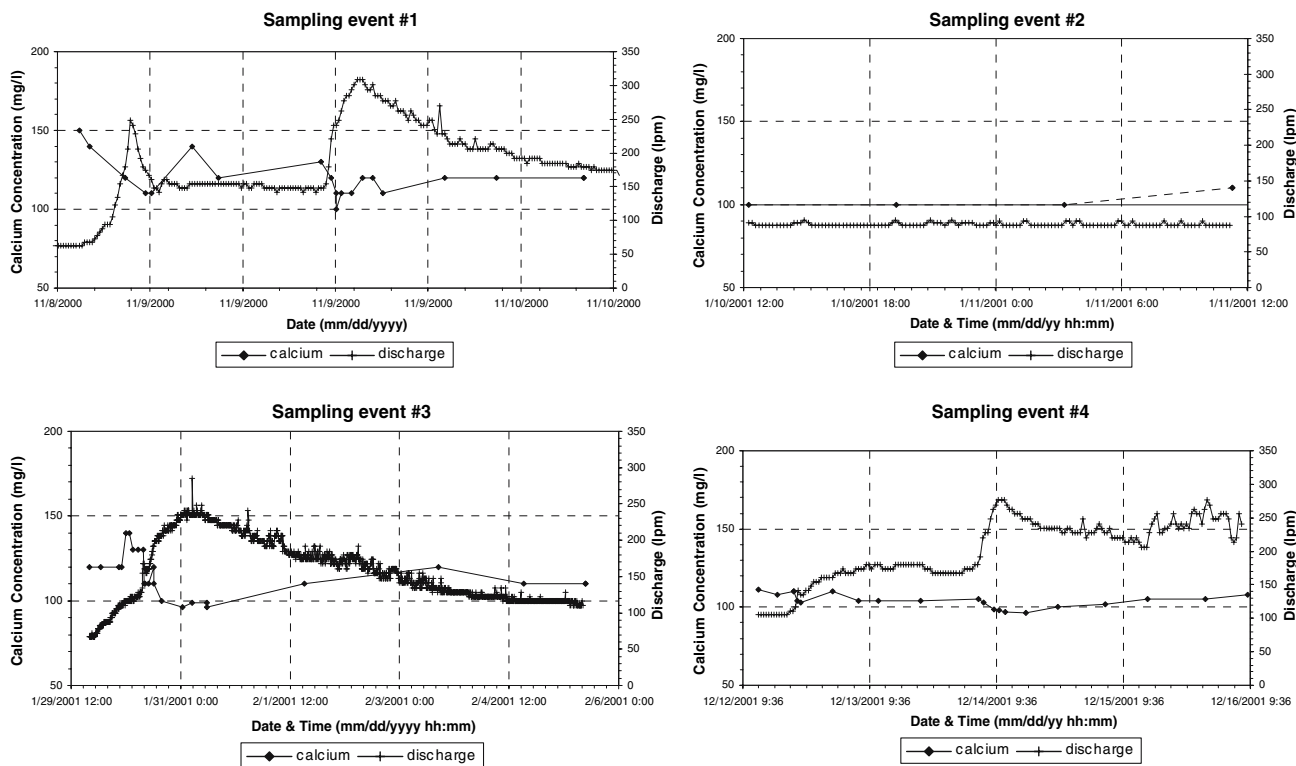


Fig. 2 Calcium chemographs in nine sampling events at a karst spring

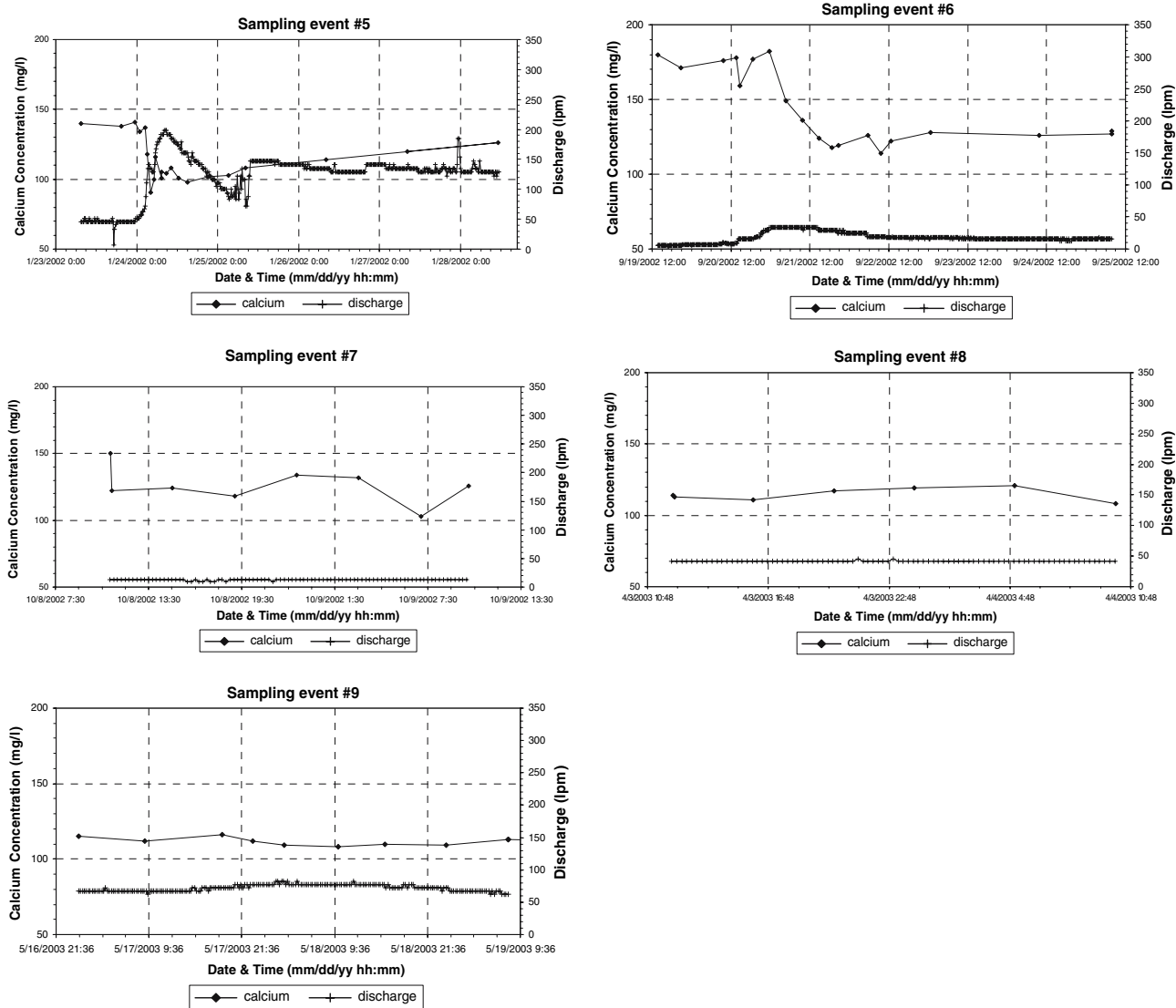


Fig. 2 continued

and simulation (Lucas 1982; Starks 1988). Since 1.96 standard deviation units correspond to 95% confidence on a normal distribution, there is approximately 95% confidence for this method as well for each comparison.

The recommendations for h , k , and SCL are based on analysis of a single statistical comparison. In practice, more than one statistical comparison is made. 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. Verification resampling is challenging at karst springs because of the dynamic nature of groundwater flow in karst aquifers and its constant reaction with the surrounding environment. Exactly recreating the many natural and anthropogenic

influences that affect each measurement is impossible for any given storm pulse sampling event.

Pooling is a method of updating background data to increase the overall size of the background data set (Gibbons 1994). 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 should be performed only after data and the monitoring process are shown to be in control.

Estimated mean (\bar{X}) and standard deviation (S)

The seemingly straight forward estimations for mean and standard deviation become complicated for measurements

at karst springs. If sequential water samples are collected over a complete hydrograph, as recommended by ASTM (1995), a chemograph is obtained for each water-quality parameter. Measurements of the samples within each individual sampling event cannot be directly used to establish the combined Shewhart-CUSUM control charts because these samples are not independent. Figure 2 shows the chemograph of calcium in nine sampling events at a karst spring, together with the discharge hydrographs. Clearly, a one-time grab sample is of little use for representing the average conditions at the spring. Each water sample represents the water quality at one particular time, and flow weighted concentration (FWC) may be used to represent the average concentrations in the receiving waters within each sampling event. Compared with the arithmetic or geometric mean, the FWC is a better parameter for the average concentration because it guarantees mass balance. 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 dataset in which both flow and concentration vary with time.

Using FWC as a statistical variable, one needs to demonstrate that FWC is random and independent of natural factors. Natural factors that systematically affect FWC should be appropriately compensated. Factors that potentially affect FWC at karst springs include

- Rainfall intensity and duration
- Aquifer structure and water levels
- Antecedent aquifer conditions
- Soil-moisture conditions
- Seasonal weather patterns
- Water temperature
- Turbidity
- Evaporation.

Interpretation of data at karst springs requires ancillary information pertinent to local weather conditions and the aquifer under study. When a spring is used as a compliance monitoring point for a landfill, data collection is not limited to the spring only. Precipitation data and data pertinent to the characteristics of the karst aquifer should also be collected. Inadequate data collection may lead to false conclusions about the landfill sites. \bar{X} and S should be estimated after the water-quality parameter has been adjusted for the effects of any natural factors. Figure 3 shows the relationship between FWC of calcium at a karst spring (Fig. 2) and average water level in a monitoring well 500 m upstream of the spring. The FWC decreases with the

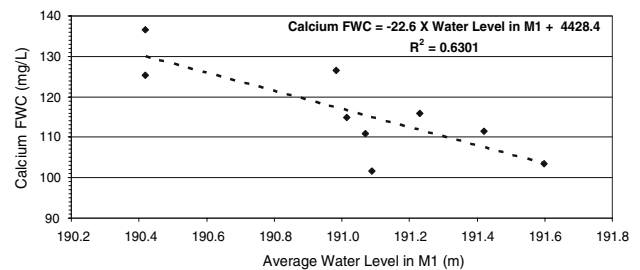


Fig. 3 Relationship between calcium concentration at a karst spring and water level in a monitoring well 500 m up-gradient of the spring

increase of the water level, which indicates a seasonal effect. The following example illustrates the potential consequence if the seasonal effect is not compensated. More detailed studies of the statistical power of combined Shewhart-CUSUM control charts are performed by Starks (1988), Gibbons (1999) and Zhou et al. (2006).

Assume that calcium has a background mean of $\bar{X} = 134$ mg/l and standard deviation of $S = 6.9$ mg/l (data in Fig. 2), $SCL = 4.5$, $k = 1$, and $h = 5$. Table 1 lists the hypothetical FWCs of calcium in the next eight sampling events in response to a continual decrease of water level in M1 and the combined Shewhart-CUSUM control procedures.

The process is out of control in terms of the CUSUM control procedure in the fifth sampling event. Based on the Shewhart control procedure, the process is out of control in the seventh sampling event. Therefore, the in-control length of the combined Shewhart-CUSUM procedure is only five. Such an out-of-control detection is false because it is not caused by the landfill operation. It results from the effect of the water level on the water quality at the spring. Therefore, it is very important to understand the geochemical processes in the aquifer so that the developed control charts are effective in detection of landfill leakage.

If the water level continuously increases for the next eight sampling events, relationship between water level and

Table 1 Example dataset for eight future sampling events in response to decrease of water level

Sampling event	Water level (m)	FWC (mg/l)	Standardized z_i	$z_i - k$	Y_i
1	189.94	135.78	0.26	-0.74	0.00
2	189.63	142.67	1.26	0.26	0.26
3	189.33	149.56	2.25	1.25	1.51
4	189.02	156.45	3.25	2.25	3.76
5	188.72	163.34	4.25	3.25	7.02
6	188.41	170.23	5.25	4.25	11.27
7	188.11	177.12	6.25	5.25	16.52
8	187.80	184.01	7.25	6.25	22.76

Note: Values in bold and italics are out of control

Table 2 Example dataset for eight future sampling events in response to increase of water level

Sampling event	Water level (m)	FWC (mg/l)	Leakage 3XS (mg/l)	Composite concentration (mg/l)	Standardized z_i	$z_i - k$	Y_i
1	189.94	135.78		135.78	0.26	-0.74	0.00
2	190.24	128.89	20.70	149.59	2.26	1.26	1.26
3	190.55	122.00	20.70	142.70	1.26	0.26	1.52
4	190.85	115.11	20.70	135.81	0.26	-0.74	0.78
5	191.16	108.22	20.70	128.92	-0.74	-1.74	0.00
6	191.46	101.33	20.70	122.03	-1.74	-2.74	0.00
7	191.77	94.44	20.70	115.14	-2.73	-3.73	0.00
8	192.07	87.55	20.70	108.25	-3.73	-4.73	0.00

calcium concentration can lead to false negative detection. Table 2 lists the hypothetical FWCs of calcium in response to an increase trend of the water level. Starting at the second sampling event, a leakage with 3 standard deviation units over the background is added to the concentrations. However, the combined Shewhart-CUSUM control procedure indicates that the process stays in control (Table 2).

Clearly, the combined control charts cannot detect such a leakage because of the effect of water level. The out-of-control length is very long. In fact, the CUSUM control procedure shows out-of-control only when the leakage is 4 standard deviation units, whereas the Shewhart control procedure shows out-of-control when the leakage is 5 standard deviation units. The false negative detection for leakage less than 4 standard deviation units is caused by the effect of water level on the water quality at the spring. Therefore, inadequate characterization of the temporal variations of water quality and their influencing factors at karst springs can lead to failure of groundwater monitoring programs at landfill sites. Springs are viable monitoring locations in karst terranes at landfill sites, however, development of robust statistical evaluation plan requires a comprehensive understanding of the karst system and knowledge of the assumptions involved in the statistical tests.

Conclusions

Inter-location comparisons are not valid in karst spring monitoring because of spatial variability of karst aquifers. Intra-spring monitoring is preferred. In intra-spring comparisons, the statistical tests must address the temporal variations of the water-quality at the springs. When storm-driven sampling take place over hydrographs, FWC can be used to represent the average concentrations of each sampling event, and combined Shewhart-CUSUM control procedures can be developed to evaluate whether the processes are in control. The most important issue that affects

the power of the combined control procedures is whether the natural factors that may affect the FWC of each chemical constituent has been thoroughly studied and properly compensated. Springs are viable monitoring locations in karst terranes at landfill sites, however, successful interpretation of data collected at the springs requires development of a robust statistical evaluation plan, which relies on a comprehensive understanding of the karst system and knowledge of the assumptions involved in the statistical tests.

References

- ASTM (1995) ASTM D5717-95, Standard guide for design of ground-water monitoring systems in karst and fractured-rock aquifers: 1995 annual book of standards, Sect. 4. West Conshohocken, Pennsylvania, pp 440–456
- 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, Pennsylvania, pp 1–14
- Bonacci O (1987) Karst hydrology with special reference to the dinaric karst, Springer series in physical environmental. Springer, Herdelberg, p 184
- EPA (1989) Statistical analysis of groundwater monitoring data at RCRA facilities, Interim Final Guidance. EPA/530-SW-89-026, Office of Solid Waste, US Environmental Protection Agency, Washington, DC
- EPA (1991) Proposed modifications to Title 40 CFR Part 264—standards for owners and operators of hazardous waste treatment, storage, and disposal facilities. US Environmental Protection Agency, pp 40–44 and 51–52
- EPA (1992) Statistical analysis of groundwater monitoring data at RCRA facilities: addendum to interim final guidance. In: Statistical training course for groundwater monitoring data analysis, EPA530-R-93-003
- Ewers R, Idstein P, Mingus E (1998) Carbonate (karst) aquifers may present monitoring and remediation environments which are superior to most granular aquifers, presentation at “Friends of Karst Meeting” during UNESCO’s international geological correlation program (IGCP), Project 379 “karst processes and the global carbon cycle,” Bowling Green, Kentucky, September 23–25, 1998
- Gibbons RD (1994) Statistical methods for groundwater monitoring. Wiley, 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, 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, Chapt. 3*, US Geological Survey
- Horseý HR, Carosone-Link P, Sullivan WR, Loftis J (2001) The effectiveness of intrawell ground water monitoring statistics at older subtitle d facilities. The eGovernment Company, Indianapolis
- LaMoreaux PE, Tanner JT (2001) Springs and bottled waters of the world: ancient history, source, occurrence, quality and use. Springer, Heidelberg, p 315
- LeGrand HE, LaMoreaux PE (1975) Chapt. I: hydrogeology and hydrology of karst, In: Burger A, Dubertret L (eds) *Hydrogeology of karstic terrains, Series B(3)*. International Association of Hydrogeologists, Paris, pp 9–19
- Lucas JM, Crosier RB (1982) Robust CUSUM: A robustness study for CUSUM quality control schemes. *Commun Stat 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. *Commun Stat Theor Method* 14(11):2689–2704
- Millard SP, Neerchal NK (2000) *Environmental Statistics*. CRC Press, Boca Raton
- Montgomery DC (1997) *Introduction to statistical quality control*, 3rd edn. John Wiley & Sons, New York
- Parkhurst DF (1998) Arithmetic versus geometric means for environmental data. *Environ Sci Technol/News* 92A–98A
- Quinlan JF (1990) Special problems of ground-water monitoring in karst terranes. In: Nielsen DM, Johnson AL (eds) *Ground water and vadose zone monitoring*. American society for testing and materials STP 1053, Philadelphia, Pennsylvania, 275–304
- 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. Proceedings of the third conference on hydrogeology, ecology, monitoring, and management of groundwater in karst terranes, December 4–6, Nashville, Tennessee
- Ryan TP (2002) *Statistical methods for quality improvement*, 2nd edn. Wiley, New York, p 446
- Sauter M (1992) Assessment of hydraulic conductivity in a karst aquifer at local and regional scale, Proceedings of the Third Conference on hydrogeology, ecology, monitoring, and management of groundwater in karst terranes. National Ground Water Association, Dublin, pp 39–57
- Smart CC (1999) Subsidiary conduit system: a hiatus in aquifer monitoring, modeling. In: Palmer AN, Palmer MV, Sasowsky ID (eds) *Karst modeling*. Karst Water Institute, Special Publication 5, 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. US Geological Survey, Open-File Report 00-491
- Zhou W, Beck BF, Wang J, Pettit AJ (2006) Groundwater monitoring for cement kiln dust disposal units in karst aquifers. *Environ Geol*. doi:10.1007/s00254-006-0514-8