



Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Combining particle-tracking and geochemical data to assess public supply well vulnerability to arsenic and uranium [☆]

Stephen R. Hinkle ^{a,*}, Leon J. Kauffman ^b, Mary Ann Thomas ^c, Craig J. Brown ^d, Kathy A. McCarthy ^a, Sandra M. Eberts ^c, Michael R. Rosen ^e, Brian G. Katz ^f

^a US Geological Survey, 2130 SW 5th Avenue, Portland, OR 97201, USA

^b US Geological Survey, 810 Bear Tavern Road, West Trenton, NJ 08628, USA

^c US Geological Survey, 6480 Doubletree Avenue, Columbus, OH 43229, USA

^d US Geological Survey, 101 Pitkin Street, East Hartford, CT 06108, USA

^e US Geological Survey, 2730 N Deer Run Road, Carson City, NV 89701, USA

^f US Geological Survey, 2010 Levy Avenue, Tallahassee, FL 32310, USA

ARTICLE INFO

Article history:

Received 6 November 2008

Received in revised form 6 May 2009

Accepted 7 July 2009

This manuscript was handled by L. Charlet, Editor-in-Chief, with the assistance of P. Bhattacharya, Associate Editor

Keywords:

Vulnerability

Particle-tracking

Arsenic

Uranium

Classification tree analysis

Principal components analysis

SUMMARY

Flow-model particle-tracking results and geochemical data from seven study areas across the United States were analyzed using three statistical methods to test the hypothesis that these variables can successfully be used to assess public supply well vulnerability to arsenic and uranium. Principal components analysis indicated that arsenic and uranium concentrations were associated with particle-tracking variables that simulate time of travel and water fluxes through aquifer systems and also through specific redox and pH zones within aquifers. Time-of-travel variables are important because many geochemical reactions are kinetically limited, and geochemical zonation can account for different modes of mobilization and fate. Spearman correlation analysis established statistical significance for correlations of arsenic and uranium concentrations with variables derived using the particle-tracking routines. Correlations between uranium concentrations and particle-tracking variables were generally strongest for variables computed for distinct redox zones. Classification tree analysis on arsenic concentrations yielded a quantitative categorical model using time-of-travel variables and solid-phase-arsenic concentrations. The classification tree model accuracy on the learning data subset was 70%, and on the testing data subset, 79%, demonstrating one application in which particle-tracking variables can be used predictively in a quantitative screening-level assessment of public supply well vulnerability. Ground-water management actions that are based on avoidance of young ground water, reflecting the premise that young ground water is more vulnerable to anthropogenic contaminants than is old ground water, may inadvertently lead to increased vulnerability to natural contaminants due to the tendency for concentrations of many natural contaminants to increase with increasing ground-water residence time.

Published by Elsevier B.V.

Introduction

Ground-water vulnerability assessments are frequently used to direct resource monitoring, prioritize aquifer protection efforts, and guide placement of public supply wells. Vulnerability assessments range in sophistication from simple, qualitative, subjective evaluations to complex, quantitative, scientifically defensible

investigations, and a key challenge often is to find an optimal balance among assessment complexity, uncertainty and cost (Focazio et al., 2002). Vulnerability assessments frequently focus on characterization of anthropogenic contaminant loads from near land surface to the water table (Witkowski, 2007), and for simplicity often ignore transport beyond the water table. This focus on anthropogenic contamination was explicitly incorporated in the widely used vulnerability tool "DRASTIC" (Aller et al., 1987), which continues to see widespread worldwide use (Focazio et al., 2002). Focus on anthropogenic contaminants occurs in spite of the fact that many natural contaminants pose human health threats and are widely regulated. Natural contaminants help comprise the contaminant group (inorganic chemicals) that had the highest percent of maximum contaminant level (MCL) violations in United States (US) public water systems that used ground water during the period

[☆] This article is a US Government work and is in the public domain in the USA. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

* Corresponding author. Tel.: +1 503 251 3237; fax: +1 503 251 3470.

E-mail addresses: srhinkle@usgs.gov (S.R. Hinkle), lkauff@usgs.gov (L.J. Kauffman), mathomas@usgs.gov (M.A. Thomas), cjbrown@usgs.gov (C.J. Brown), mccarthy@usgs.gov (K.A. McCarthy), smeberts@usgs.gov (S.M. Eberts), mrosen@usgs.gov (M.R. Rosen), bkatz@usgs.gov (B.G. Katz).

1993–1998 (US Environmental Protection Agency, 1999). Focazio et al. (2002) define a “contaminant” to be a “chemical or physical property of the ground-water resource in question that is not desirable from a health or other perspective such as interference with water-treatment practices.” A natural contaminant is one that has primarily natural (often geologic) sources.

Public supply wells may deserve particular attention in vulnerability assessments because of the heavy reliance on ground water as a source of public supply. For example, in the year 2000, about 31% of the total US population was served by public supply of ground-water origin (Hutson et al., 2004). In spite of the importance of ground water as a source of public supply, though, characterization of the vulnerability of public supply wells is challenging due to difficulties in elucidating sources to wells, a problem compounded by complexities introduced by the mixing of different flow components that results from the typically long screened intervals of public supply wells.

A key challenge to assessing vulnerability to natural contaminants arises from difficulties inherent in accounting for mobilization of natural contaminants. Mobilization of natural contaminants typically is constrained by the kinetically limited nature of water–rock interactions (Hem, 1985). In theory, the kinetic aspects of water–rock interactions can be accounted for with the application of flow modeling and particle-tracking to represent ground-water residence times, and the aforementioned complexities of ground water advection through heterogeneous flow systems to public supply wells also can be simulated with flow modeling and particle-tracking techniques.

The purpose of this work was to test the hypothesis that ground-water flow model analysis and geochemical data that describe aquifer systems at large spatial scales (tens to thousands of square kilometers) can successfully be combined to assess public supply well vulnerability to two natural contaminants, arsenic and uranium. Arsenic and uranium are ubiquitous natural ground-water contaminants of global concern (Cothorn and Lappenbusch, 1983; Wanty and Nordstrom, 1993; Welch et al., 2000; Nordstrom, 2002). Results are presented from seven study areas in six US Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Program study units (Gilliom et al., 1995) in four principal aquifers (Miller, 2000). Three statistical methods are used to analyze the following variables: a suite of time-of-travel and water-flux variables (including variables based on geochemical zonation), aqueous geochemical variables, and solid-phase trace-element variables. Emphasis is on the utility of particle-tracking variables because such flow-model-based variables can be used to account for contaminant mobilization, transport, and fate in heterogeneous aquifers, and because they have the potential to be used in a predictive manner. Predictor variables may be useful in relatively unstudied portions of aquifers where they can be generated prior to resource development and monitoring. The three statistical methods employed are principal components analysis (PCA) (Preisendorfer and Mobley, 1988), Spearman correlation analysis (Helsel and Hirsch, 1992), and classification tree analysis (Breiman et al., 1984). PCA provides a geochemical framework for understanding natural contaminants occurrence patterns in these systems and identifies associations between measured solutes and particle-tracking variables that represent important mobilization and transport processes. Spearman correlation analysis establishes the statistical significance of correlations between particle-tracking variables and arsenic and uranium concentrations. Classification tree analysis demonstrates an application of the use of particle-tracking and solid-phase variables in a screening-level, predictive vulnerability assessment. Thus, the combination of these three statistical approaches allows identification of particle-tracking variables that can be useful in vulnerability analysis, builds upon this by establishing levels of statistical significance for correlations be-

tween particle-tracking variables and natural contaminants, and then allows the demonstration of an application using some of these variables in a vulnerability assessment tool for public supply wells. Although analysis is restricted to arsenic and uranium, the approach of using particle-tracking and other predictor variables could be modified for other contaminants.

The approaches used have several limitations. Flow models and particle-tracking are inherently uncertain, as are estimates of source-term strength for natural contaminants. Also, the large spatial scale of the observations and the reliance on statistical inference do not lend themselves to well developed process-based understanding. Therefore, approaches presented here may most appropriately be used as a method for prioritizing the locations and types of more refined, site-specific vulnerability assessments. However, the combination of particle-tracking analysis and geochemical data to assess public supply well vulnerability to natural contaminants has seen little application to date, and the results presented in this paper demonstrate approaches that could easily be applied in other aquifers.

Study design

Study areas

Data gathering and particle-tracking analysis were done for seven study areas representing a variety of hydrologic, geologic, physiographic, and climatic conditions (Fig. 1). Study areas ranged in size from 60 km² (Basin and Range 1) to 2700 km² (Central Valley). Each study area was nested within a NAWQA study unit. Study areas are synonymous with model areas. Data from study areas were combined for analysis to demonstrate widespread applicability of results. Results presented in this paper are part of a broader, multi-scale assessment of public supply well vulnerability to anthropogenic and natural contaminants (Eberts et al., 2005).

Datasets

Three complementary datasets were assembled for this work. They consisted of aqueous geochemical data, flow-model-derived variables (henceforth, “particle-tracking dataset”), and solid-phase geochemical data.

Aqueous geochemical dataset

The aqueous geochemical dataset (Table S1 in supporting information) included sites with analyses of the natural contaminants arsenic and uranium; the redox indicator species dissolved oxygen (O₂), nitrate (NO₃), manganese (Mn), iron (Fe), and sulfate (SO₄); pH; alkalinity (alk); and filtered (dissolved) organic carbon (DOC). Of the 368 sites for which arsenic and/or uranium data were available, 312 contained arsenic data and 289 contained uranium data; data for other analytes were present to varying degrees. This dataset was composed of two groups of data. One group (273 wells) was existing data collected primarily from public supply wells but also including domestic and ambient monitoring wells. Samples were restricted to the time period 1997–2001, and were assembled from USGS and other public databases. Samples that had been treated (e.g., chlorinated) or that came from blended sources (multiple public supply wells) were excluded. The second group (95 wells) consisted of newly collected data from additional public supply wells sampled during 2001–2003 in the seven study areas as part of the NAWQA Program.

Particle-tracking dataset

The particle-tracking dataset (145 sites; Table S2 in supporting information), a subset of the wells represented in the aqueous geochemical dataset, included analysis using three-dimensional

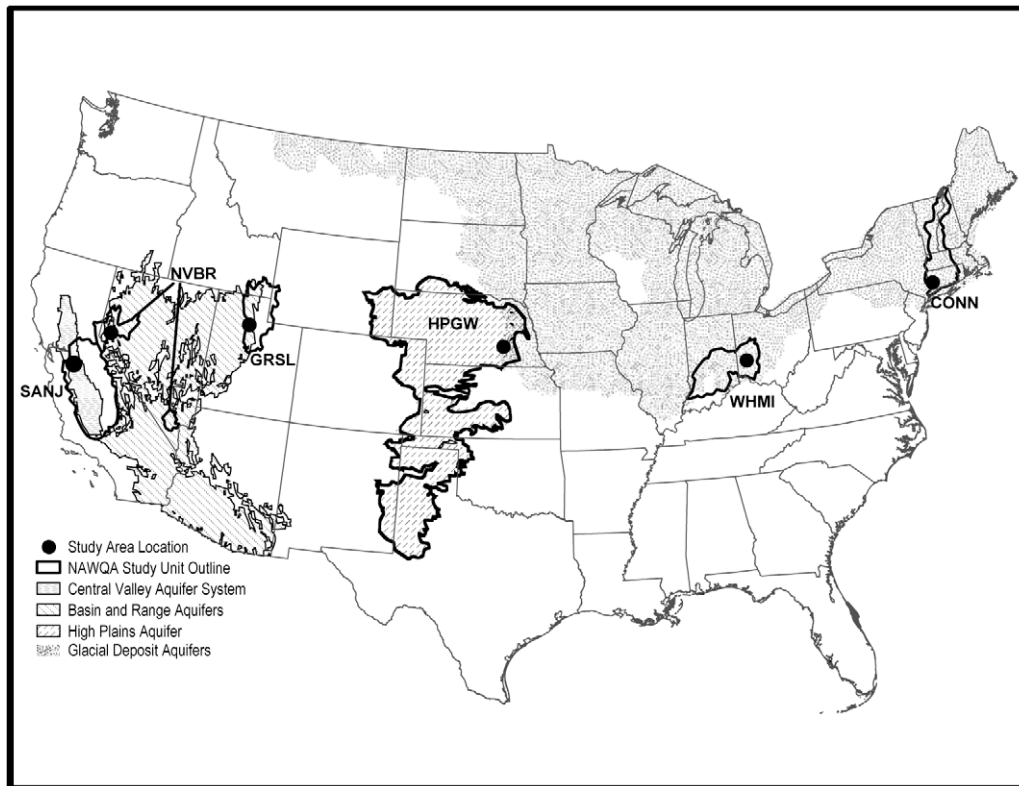


Fig. 1. Location of study areas, NAWQA study units, and principal aquifers. NAWQA study units are SANJ (San Joaquin), NVBR (Nevada Basin and Range), GRSL (Great Salt Lake), HPGW (High Plains Ground Water), WHMI (White-Miami), and CONN (Connecticut); study areas are: Central Valley (SANJ), Basin and Range 1 and Basin and Range 2 (two study areas in northern NVBR), Basin and Range 3 (GRSL), High Plains (HPGW), Glacial 1 (WHMI), and Glacial 2 (CONN); study units and study areas are described in Paschke (2007).

numerical ground-water flow modeling and particle-tracking to derive a variety of time-of-travel and other flow-model-based variables. These variables are different statistical measures of (1) overall simulated time of travel and water flux, and (2) simulated time of travel and water flux through distinct geochemical zones (described below). The wells consisted solely of public supply wells. Of the 145 wells, 138 contained arsenic data and 128 contained uranium data; particle-tracking variables were available for all wells in this dataset.

Steady-state ground-water flow was simulated with the modular finite-difference ground-water flow simulation code MODFLOW-2000 (Harbaugh et al., 2000), and calibrated, using pumping and water-level data, following guidelines of Hill (1998). Hydraulic conditions for the period 1997–2001 were used to facilitate comparisons of modeling results among study areas. The particle-tracking program MODPATH (Pollock, 1994) was used to compute advective time of travel to individual public supply wells. MODPATH uses the flow regime calculated by MODFLOW to simulate the ground-water flow velocity distribution throughout an aquifer. This velocity distribution is then used to determine flow paths of water particles in the aquifer. Time of travel along flow paths is computed by MODPATH using the magnitude of fluxes between model cells, porosity of the aquifer represented by these cells, and the dimensions of the cells. Potential problems associated with weak sinks (model cells with discharge rates insufficient to capture all particles entering the cell) were addressed by using the grid-refinement program of Spitz (2001) for model cells in which public supply wells were located. Well selection criteria for particle-tracking included approximately equal proportions of wells representing each quartile of pumping rates within each study area. Paschke (2007) provides detailed discussion of the conceptualization, simulation and calibration of individual ground-

water flow models; description of the particle-tracking simulation (including implementation of the grid-refinement program) for each study area; and a data dictionary describing the particle-tracking variables.

Uncertainties in flow modeling and particle-tracking arise from simplification and bias in the representation of aquifer properties and stresses. Comparison between particle-tracking-derived measures of time of travel and tritium concentrations in ground water can provide evidence for the reliability of particle-tracking results. In public supply wells, where water typically is composed of mixtures of different residence times, tritium concentrations should generally increase as the proportion of post-bomb ground water increases. Thus, although tritium concentrations have not increased monotonically in recent decades, tritium concentrations—representing contributions of post-bomb water in ground-water mixtures in public supply wells—would be expected to have some correlation with proportions of young particles from particle-tracking simulations if the flow models adequately represent the flow systems being modeled. Tritium data exist for 23 of the public supply wells for which particle-tracking simulations were done (samples collected from six study units, during 1987–2006; Table S3 in supporting information). An inverse linear correlation between tritium concentration and log of median time of travel was observed, with $r^2 = 0.25$ and $p = 0.015$, providing support for the particle-tracking analysis. This p -value, although above the 0.01 significance level used in other correlation analysis in this paper, may reflect the small sample size of this analysis, and is of greater statistical significance than the 0.05 significance level typically used in hydrologic studies (Helsel and Hirsch, 1992, p. 106). The r^2 value likely reflects the complexities of ground-water mixtures in long-screened public supply wells and the nonlinear changes in tritium concentrations over time.

To help account for geochemical controls on arsenic and uranium mobilization, transport, and fate, flow model domains were discretized to represent dominant redox and pH zones, and particle-tracking variables representing simulated time of travel and water flux through these geochemical zones were calculated for each public supply well contributing zone (Paschke, 2007). Thus, non-geochemically based particle-tracking variables capture the general effects of reaction kinetics, whereas geochemically based particle-tracking variables account for residence time in and fluxes through the zones in which different mobilization mechanisms may be operating. As such, geochemically based particle-tracking variables have the potential to contribute geochemical process-based understanding to the characterization of public supply well vulnerability to natural contaminants, and also to contribute to the broad identification of the dominant geochemical reactions responsible for mobilization of natural contaminants within a given dataset. Redox conditions were inferred on the basis of terminal electron accepting processes and were simplified to (1) O₂- and NO₃-reducing, and (2) Fe- and Mn-reducing. (O₂-reducing water is oxic water in which O₂ is available for reduction; it is not anoxic water.) Zones for pH were pH < 8 and pH > 8.

Solid-phase geochemical dataset

Solid-phase geochemical data (Table 1) were gathered for use in classification tree analysis to represent differences in potential arsenic and uranium source strength among study areas. Three national scale datasets were evaluated. One dataset consisted of NAWQA streambed sediment samples (Stephens, 2003). Streambed sediment integrates near-surface geologic materials upstream from sampling points and provides a means of representing regional differences in geologic materials. NAWQA streambed sediment samples represent primarily NAWQA occurrence and distribution (ambient monitoring) samples (Gilliom et al., 1995). Streambed sediment samples at each site were composited from approximately 5–10 depositional zones, wet sieved on-site to finer than 63 μm (Shelton and Capel, 1994), and analyzed for elemental composition (Sanzolone and Ryder, 1989; Arbogast, 1990; Briggs and Meier, 1999).

The second dataset was a preliminary version of the USGS National Geochemical Survey (NGS) (Grossman, 2004). A similar preliminary dataset (arsenic only) was published by Grosz et al. (2004). The NGS dataset consists of over 20,000 streambed sediment and soil samples. At the time of this analysis, the NGS covered about 75% of the US (arsenic) and 65% of the US (uranium).

The third dataset was a coverage of arsenic concentrations in surficial soils, identical to a map published by Gustavsson et al. (2001), and created for this project using the original raw data and the interpolation methods of Gustavsson et al. (2001). The small number of samples (1323 for the conterminous US) and point-specific nature of soil samples (compared with the integrated nature of streambed sediment samples) are limitations.

Comparison of the three datasets facilitates evaluation of the value of these solid-phase data. Absolute differences in concentrations among the three datasets may reflect, in part, differences in sample medium and in processing and analysis. For example, NAWQA streambed sediment samples were finer than 63 μm, whereas NGS samples were sieved to finer than 150 μm, and soils from the NGS and Gustavsson et al. (2001) datasets may represent a more weathered geologic material than streambed sediment. Relative differences in concentrations among study units and study areas within individual datasets should be more useful in determining patterns of trace element occurrence in solid phases. For arsenic, all three datasets generally show similar patterns of enrichment or depletion in solid phases (Table 1), suggesting that the solid-phase data have utility in representing relative differences in geologic material source strength for arsenic among study

units and study areas. The more limited data for uranium make comparisons for uranium difficult.

For classification tree analysis, median values from NAWQA study unit streambed sediment were used. NGS data were not used because they were not available for all study areas. Soils data of Gustavsson et al. (2001) were not used because they lacked uranium analyses, in addition to having other limitations discussed above.

Approach and methods of analysis

The aqueous geochemical dataset, representing the greatest number of wells, characterizes the status and geochemistry of arsenic and uranium occurrence in the study areas. The particle-tracking dataset can be used to elucidate flow system controls on arsenic and uranium occurrence and, thus, to infer processes of arsenic and uranium mobilization and transport to public supply wells. Particle-tracking variables also can be used in a predictive manner. The solid-phase geochemical variables can be used as categorical predictor variables in classification tree analysis to represent potential source strength. That is, for a given degree of mobilization strength (represented, for example, by exposure of ground water to aquifer materials over a given length of time), higher concentrations of natural contaminants associated with solid phases may lead to the development of higher concentrations of those natural contaminants in the aqueous phase.

Principal components analysis

PCA is used to objectively reduce the number of variables in a dataset to a few principal components (PCs) that capture much of the total dataset variability, and identify structures of or relations among variables, based upon the relative weights (or loadings) of variables contributing to each of the dominant PCs. Principal components are linear combinations of the original variables and are ordered so that PC₁ explains the largest amount of variability in the data, PC₂, the second largest amount, and so on. PC loadings quantitatively express the degree of association between individual variables and the relative importance in describing the variability.

For PCA on the aqueous geochemical dataset, a subset of the entire aqueous geochemical dataset was used. This subset consisted of samples with complete suites of analytes deemed most pertinent for the analysis: arsenic, uranium, O₂, NO₃, Mn, Fe, SO₄, pH, Alk, and DOC. These variables represent the two natural contaminants of interest, redox indicator species (O₂, NO₃, Mn, Fe, SO₄), constituents that may form important complexes with uranium (SO₄, Alk, DOC), and pH (important for arsenic mobilization in some environments). Samples from 125 sites contained all analytes.

A PCA focusing on arsenic and particle-tracking variables was based on nine variables: arsenic, O₂, NO₃, Mn, Fe, median simulated time of travel (TT median), median simulated time of travel in the Fe- and Mn-reducing zone (TT Fe median), median simulated time of travel in the zone of elevated pH (pH > 8) (TT pH8 median), and percent of simulated time of travel in the Fe- and Mn-reducing zone (%TT Fe). These particle-tracking variables were chosen because they represent different aspects of mobilization processes: a general time-of-travel variable (TT median), two redox-zonation based variables (TT Fe median, %TT Fe), and a pH-zonation based variable (TT pH8 median). Samples from 87 sites contained these analytes and particle-tracking variables. This analysis allowed elucidation of associations between arsenic and several important redox indicator species and particle-tracking variables. Analysis of measured redox indicator species and of redox-based particle-tracking variables also provided an opportunity to qualitatively evaluate the effectiveness of flow model redox zonation.

Table 1
Arsenic and uranium concentrations in sediment and soil. (“–”, data not available or insufficient to report.).

Study unit and study area	Region ^a	Arsenic concentration (mg/kg, or parts per million)						Uranium concentration (mg/kg, or parts per million)			
		NAWQA streambed sediment ^b		USGS National Geochemical Survey ^c		Gustavsson et al. (2001) Soils ^d		NAWQA streambed sediment ^b		USGS National Geochemical Survey ^c	
		Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean
CONN Glacial 2	Study unit	5.8 (n = 43)	7.3 (n = 43)	2.4	3.3	5.1	5.2	5.0 (n = 43)	5.4 (n = 43)	4.2	4.6
	Study area	–	–	–	3.2	–	5.3	–	–	–	5.8
GRSL Basin and Range 3	Study unit	14 (n = 12)	80 (n = 12)	–	–	7.9	9.9	3.0 (n = 12)	4.8 (n = 12)	–	–
	Study area	–	–	–	–	–	11	–	–	–	–
HPGW High Plains ^e	Study unit	6.3 (n = 62)	6.7 (n = 62)	–	–	6.0	6.2	1.3 (n = 62)	1.4 (n = 62)	–	–
	Study area	–	–	–	11	–	7.1	–	–	–	3.2
NVBR Basin and Range 1 ^f	Study unit	21 (n = 11)	31 (n = 11)	9.5	13	10	11	4.4 (n = 11)	8.0 (n = 11)	2.3	2.6
	Study area	–	–	–	4.5	–	9.0	–	–	–	1.7
NVBR Basin and Range 2 ^f	Study unit	21 (n = 11)	31 (n = 11)	9.5	13	10	11	4.4 (n = 11)	8.0 (n = 11)	2.3	2.6
	Study area	–	–	–	5.5	–	9.4	–	–	–	3.0
SANJ Central Valley	Study unit	9.6 (n = 17)	12 (n = 17)	4.9	7.2	5.1	5.4	5.7 (n = 17)	8.0 (n = 17)	–	–
	Study area	–	–	–	3.5	–	3.2	–	–	–	–
WHMI Glacial 1	Study unit	8.3 (n = 28)	8.5 (n = 28)	–	–	9.7	10	3.3 (n = 28)	3.2 (n = 28)	–	–
	Study area	–	–	–	–	–	12	–	–	–	–

^a “Study unit” refers to a NAWQA hydrologic component, whereas “Study area” refers specifically to a model area component used in this paper. Study areas are nested in study units.

^b Streambed sediment data are reported for study units but not for study areas because data were too sparse for the latter.

^c National Geochemical Survey data are reported for study units and study areas in which >90% of the study unit or study area is represented in National Geochemical Survey grid coverages.

^d Soils data (arsenic only; soils were not analyzed for uranium) are reported for study units (median and mean) and study areas (mean only; insufficient grid density for meaningful median); these data were derived from a grid coverage of soils data.

^e The High Plains NAWQA study unit encompasses parts of eight mid-western states, so the more appropriately scaled Lower Kansas River Basin NAWQA pilot study unit, which is located in Nebraska and Kansas and encompasses all of the High Plains study area, was used instead of the entire High Plains study unit. The Lower Kansas River Basin study was conducted prior to full implementation of the NAWQA Program. Uranium analysis was done with a single-acid digestion (boiling with nitric acid) and thus the uranium streambed sediment samples likely have a low bias. Other uranium analyses, and all arsenic analyses, represent totals in streambed sediment.

^f The Nevada Basin and Range NAWQA study unit contains separate northern (Truckee/Carson Basins) and southern (Las Vegas Valley) regions, so the more appropriately located northern component was used to represent the Basin and Range 1 and 2 study units.

Finally, a PCA focusing on uranium and particle-tracking variables was based on nine variables, including a larger suite of particle-tracking variables: uranium, five general particle-tracking variables (TT median; minimum simulated time of travel: TT Min; maximum simulated time of travel: TT Max; percent of simulated water flux with time of travel <10 years: %TT 10–; and percent of simulated water flux with time of travel \geq 200 years: %TT 200+), two redox-zonation based particle-tracking variables (median simulated time of travel in the O₂- and NO₃-reducing zone: TT O₂ median; percent of simulated water flux from the O₂- and NO₃-reducing zone: %Flux O₂), and one pH-zonation based particle-tracking variable (TT pH8 median). Samples from 128 sites were represented. This analysis highlighted associations between uranium and a more extensive suite of particle-tracking variables.

S-PLUS software was used for PCA (Insightful Corporation, 2001; Venables and Ripley, 2002). Data variables were transformed to utilize correlations instead of covariances, circumventing potential problems associated with use of disparate units.

Spearman correlation analysis

Spearman correlation analysis is a nonparametric correlation analysis based upon the ranks of the data—that is, it measures the strength of monotonic (including nonlinear) correlations, and as such, it is well suited for non-normally distributed data. Although simple parametric or nonparametric correlations commonly have insufficient explanatory power for use in vulnerability analysis by themselves, identification of statistically significant correlations can provide insight for other applications in which the same or similar variables are used. SAS software (SAS Institute, 1989) was used for Spearman correlation analysis.

Classification tree analysis

Classification tree analysis is a nonparametric method in which a dataset is partitioned recursively into increasingly homogeneous subsets. Partitioning is optimized by maximizing the decrease in “deviance” (Venables and Ripley, 2002) or “impurity” (Breiman et al., 1984) in the split groups; that is, choosing the predictor variable and splitting value that yield the most homogeneous subsets. The recursive splitting yields models that schematically resemble trees, similar to dichotomous classification keys commonly used in botanical identification and medical decision schemes. These recursive models can uncover relations that are logical, but may be difficult to identify with linear statistical models. This can be advantageous where interactions among predictor variables are neither additive nor multiplicative. Both continuous and categorical data can be used in classification tree analysis. Classification tree analysis has been used to evaluate factors controlling pesticide concentration variability in streams (Qian and Anderson, 1999) and to infer nitrate sources in ground water (Spruill et al., 2002).

S-PLUS software was used for classification tree analysis (Insightful Corporation, 2001; Venables and Ripley, 2002). Model building was done using predictive variables: particle-tracking variables (including geochemistry-based ones) from the Spearman correlation analysis, plus study-unit-level arsenic and uranium concentrations in streambed sediment. The use of study-unit-level streambed sediment data results in a categorical variable in which all sites in a given study unit are assigned the same streambed sediment value. A variety of response variable classes (ranges of arsenic and uranium concentrations), based on trial-and-error, was used during model development. Random numbers were assigned to samples for the purposes of dividing the overall dataset into learning and testing datasets. The learning dataset was used to develop classification tree models, whereas the testing dataset was used to evaluate the predictive capability of the models.

Results and discussion

Arsenic and uranium concentrations

In the aqueous geochemical dataset, arsenic concentrations ranged from <2 to 139 $\mu\text{g/L}$. Samples from 12% of these sites exceeded the US Environmental Protection Agency (USEPA) MCL of 10 $\mu\text{g/L}$, with exceedances in six of the seven study areas. Uranium concentrations ranged from <0.02 to 1240 $\mu\text{g/L}$. Samples from 11% of these sites exceeded the USEPA MCL of 30 $\mu\text{g/L}$, with exceedances in three of the seven study areas.

In the particle-tracking dataset, arsenic concentrations ranged from <2 to 30 $\mu\text{g/L}$. Samples from 12% of these sites exceeded the USEPA MCL for arsenic, with exceedances in five of the seven study areas. Uranium concentrations ranged from 0.02 to 67 $\mu\text{g/L}$. Samples from 8% of these sites exceeded the USEPA MCL for uranium, with exceedances in three of the seven study areas.

These results are similar to results of other studies at local, regional, and national scales that demonstrate widespread occurrence of natural contaminants in ground and drinking water (e.g., Ryker, 2003; Focazio et al., 2006). These data demonstrate that natural contaminants deserve attention in vulnerability analysis.

Principal components analysis

PCA was used to characterize geochemical occurrence patterns in the aqueous geochemical dataset, and then used to demonstrate that the combination of particle-tracking analysis and geochemical data could be useful in understanding public supply well vulnerability to arsenic and uranium. Geochemical controls over arsenic and uranium mobilization and fate in ground water have been well studied, and evaluation of occurrence patterns in the aqueous geochemical dataset was done to place this dataset in the context of these known processes. However, the degree to which arsenic and uranium occurrence in public supply wells might be related to and characterized by a combination of flow-model and geochemical variables has received little attention, and it is elucidation of associations between geochemical and particle-tracking variables and the usefulness of these relations in vulnerability assessments that is of primary interest in this paper.

The first three PCs from a PCA on the aqueous geochemical dataset are shown in Fig. 2a. These three PCs account for 51% of the total variance. Component 1 is dominated by Mn (positive), arsenic (positive), and O₂ (negative). The opposing loadings of Mn and O₂ reflect the thermodynamically different redox environments that these redox indicator species represent; loadings for Fe (positive) and NO₃ (negative), although smaller than the loadings of Mn and O₂, may reflect the common association of Mn and Fe in reducing environments, and the common association of O₂ and NO₃ in oxidizing environments (McMahon and Chapelle, 2008). The association of arsenic with Mn in component 1 is consistent with the widespread importance of reducing conditions for arsenic mobilization (Smedley and Kinniburgh, 2002). Arsenic mobilization under Fe- and Mn-reducing conditions has been attributed to reductive dissolution of iron and manganese oxyhydroxides, a process by which arsenic that is sorbed to or co-precipitated with iron and manganese oxyhydroxides is released (Smedley and Kinniburgh, 2002; Plant et al., 2005), and to reduction of sorbed arsenate to less strongly sorbed arsenite with subsequent release (Zobrist et al., 2000). However, the geochemistry of arsenic is complex. Desorption of arsenic, an oxyanion, increases as pH increases from circumneutral pH to elevated pH (>8) (Dzombak and Morel, 1990). Arsenic mobilization through pH-induced desorption can occur under oxidizing or reducing conditions; however, pH-induced arsenic desorption is stronger for the oxidized form, arsenate, than for the reduced form, arsenite (Dzombak

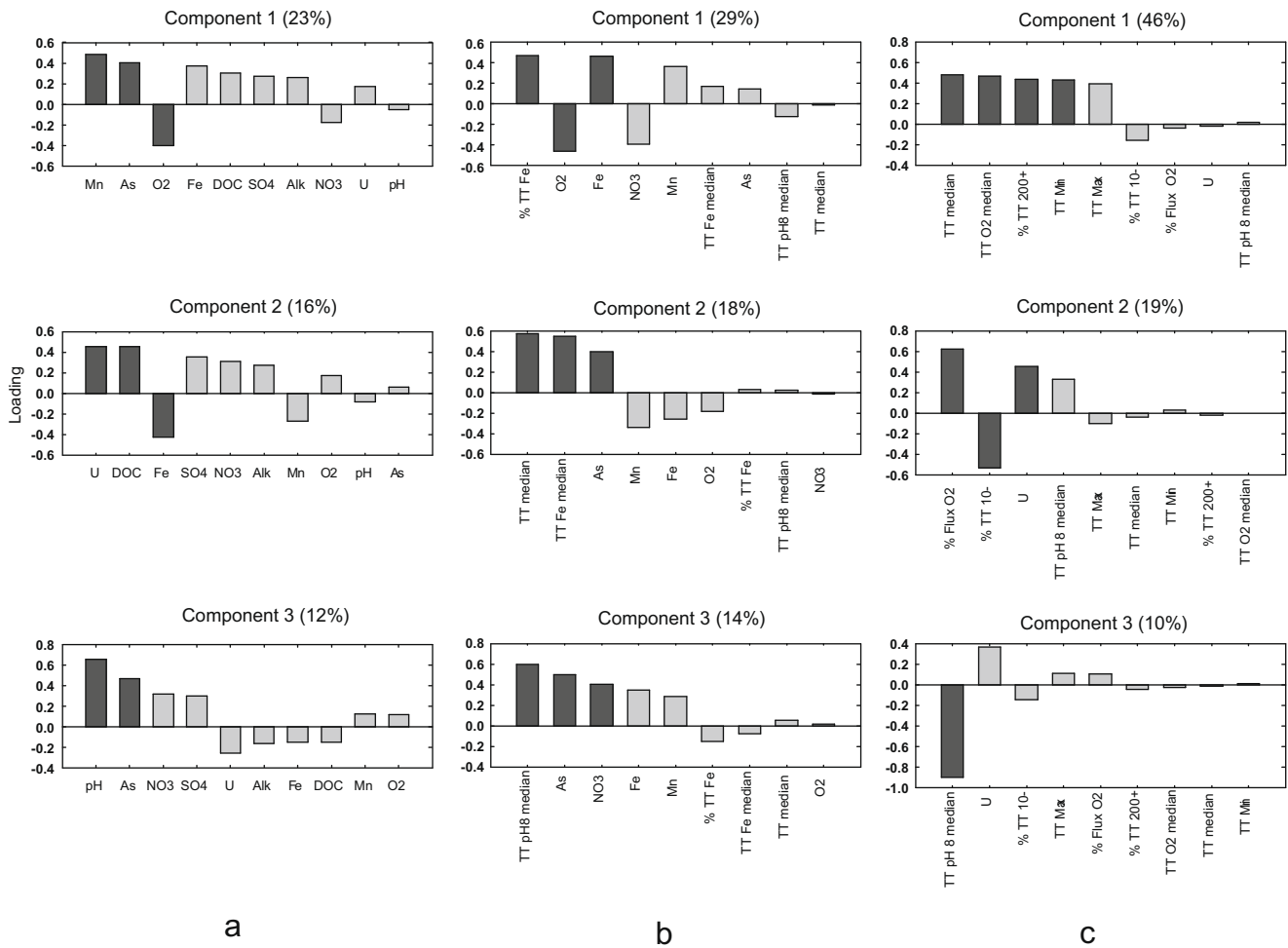


Fig. 2. Loadings for each of the first three components from principal components analysis (darker shade indicates loading >0.4), and percent of variance explained by each component: (A) arsenic, uranium, and aqueous geochemical variables (left); (B) arsenic, four redox variables, and four particle-tracking variables (center); (C) uranium and eight particle-tracking variables (right). Abbreviations: Mn, manganese; As, arsenic; O₂, dissolved oxygen; Fe, iron; DOC, filtered (dissolved) organic carbon; SO₄, sulfate; Alk, alkalinity; NO₃, nitrate; U, uranium; %TT Fe, percent of simulated time of travel in the Fe- and Mn-reducing zone; TT Fe median, median simulated time of travel in the Fe- and Mn-reducing zone; TT pH8 median, median simulated time of travel in the zone of elevated pH (pH > 8); TT median, median simulated time of travel; TT O₂ median, median simulated time of travel in the O₂- and NO₃-reducing zone; %TT 200+, percent of simulated water flux with time of travel \geq 200 years; TT Min, minimum simulated time of travel; TT Max, maximum simulated time of travel; %TT 10-, percent of simulated water flux with time of travel <10 years; %Flux O₂, percent of simulated water flux from the O₂- and NO₃-reducing zone.

and Morel, 1990). Oxidation of arsenic-bearing sulfide minerals and competitive sorption are other arsenic mobilization processes that occur in oxic environments (Welch et al., 2000; Smedley and Kinniburgh, 2002). Thus, the association between arsenic and Mn (and, to a lesser extent, Fe) in this analysis likely reflects the overriding importance of reductive desorption and dissolution reactions for arsenic mobilization in the overall dataset, but oversimplifies the true complexity of arsenic mobilization processes in these different environments.

Component 2 of the first PCA is dominated by uranium (positive), DOC (positive), and Fe (negative). The association of uranium with DOC, and to a lesser extent with SO₄ and Alk, is consistent with importance of complexation reactions for uranium mobilization. Uranium solubility is strongly increased by complexation with DOC, SO₄, and carbonate (Langmuir, 1978; Garrels and Christ, 1990; Davis et al., 2004). Although SO₄ is a redox indicator species, sulfate-reducing conditions reflect strongly reducing conditions that may be difficult to identify in the absence of hydrogen sulfide (H₂S) concentration data, and the association between uranium and SO₄ in this analysis may reflect a stronger role of SO₄ in uranium complexation than in characterization of redox environment. The opposite loading of uranium and Fe is consistent with redox

controls on uranium mobilization: the oxidized form of uranium, uranium(VI), is mobile under O₂- and NO₃-reducing conditions, whereas under Fe- and Mn-reducing conditions, uranium tends to be reduced to low-solubility uranium(IV) (Langmuir, 1978; Garrels and Christ, 1990; Lovley et al., 1991).

Component 3 of the first PCA is dominated by pH and arsenic (both positive). This association may reflect the importance of pH-driven desorption for arsenic mobilization.

The results of the first PCA demonstrate associations between geochemical conditions and the occurrence of arsenic and uranium in these samples from seven study areas across the US. These findings reinforce the understanding of dominant geochemical relations that control the occurrence of these natural contaminants, understanding that could be combined with flow-model-derived variables to better understand public supply well vulnerability to arsenic and uranium.

The second PCA demonstrates associations among arsenic, redox indicator species, and selected particle-tracking variables. The first three PCs, shown in Fig. 2b, account for 60% of the total variance. Component 1 is dominated by the percent of simulated time of travel in the Fe- and Mn-reducing zone (positive), O₂ (negative), and Fe (positive). The association of simulated time of travel

in reducing zones with the measured Fe, and in inverse association with measured O₂, indicates that the redox zones represented in the flow models capture the general redox conditions in these systems.

Component 2 of the second PCA is dominated by median simulated time of travel (positive), median simulated time of travel in the Fe- and Mn-reducing zone (positive), arsenic (positive). These associations suggest that arsenic concentrations generally increase with increasing residence time in aquifers and with increasing residence time in reducing portions of aquifers. The association between arsenic and median simulated time of travel reflects the role of kinetically limited reactions in the mobilization of arsenic, whereas the association between arsenic and median simulated time of travel in the Fe- and Mn-reducing zone indicate that the kinetically limited reactions that appear to control arsenic mobilization may be dominated by reductive desorption and dissolution reactions rather than by oxidizing processes.

Component 3 of the second PCA is dominated by median simulated time of travel in the zone of elevated pH (pH > 8) (positive), arsenic (positive), and NO₃ (positive). These associations suggest, again, that elevated pH may be important in the mobilization of arsenic, and the similarity between the associations in component 3 of the second PCA (Fig. 2b) with the associations in component 3 of the first PCA (Fig. 2a) suggests that pH zonation in the flow model domains captures the dominant pH patterns as reflected in measured pH in these systems. The positive loading for NO₃ could indicate a tendency for elevated pH conditions to be better represented in oxidizing environments in this dataset, but similar and only slightly lower loadings for Fe and Mn indicate that associations between zones of elevated pH and redox conditions are not well defined in the overall dataset.

The results of the second PCA demonstrate associations between arsenic concentrations and overall time of travel, time of travel through reducing zones, and time of travel through zones of elevated pH. These findings indicate that particle-tracking variables capture important hydrodynamic controls over public supply well vulnerability to arsenic.

The third PCA demonstrates associations between uranium and selected particle-tracking variables. The first three PCs, shown in Fig. 2c, account for 75% of the total variance. Component 1 is dominated by median simulated time of travel, median simulated time of travel in the O₂- and NO₃-reducing zone, percent of simulated water flux with time of travel ≥ 200 years, and minimum simulated time of travel (all positive). The association among these particle-tracking variables is not unexpected, given the cross-correlation of time-of-travel variables in particle-tracking analysis.

Component 2 of the third PCA is dominated by percent of simulated water flux from the O₂- and NO₃-reducing zone (positive), percent of simulated water flux with time of travel < 10 years (negative), and uranium (positive). The association between uranium and water flux from oxidizing zones is consistent with the known importance of redox controls on uranium mobilization, as discussed above. The inverse association between uranium and percent of simulated water flux with time of travel < 10 years may indicate a tendency for young ground water to contain relatively low concentrations of uranium and old ground water to contain relatively high concentrations of uranium, a kinetic effect analogous to the apparent importance of kinetic effects for arsenic mobilization (Fig. 2b). The general propensity for ground water to evolve towards increasingly reduced conditions as residence time increases presents a condition in which, on the one hand, older water has more time for kinetically limited reactions such as mineral dissolution to proceed, potentially increasing uranium concentrations, and on the other hand, older water has more time to progress to increasingly reduced conditions, potentially decreasing uranium concentrations. The two dominant particle-tracking vari-

ables in the second component capture aspects of these competing effects by accounting for contributions of both oxidizing ground water and young ground water.

Component 3 of the third PCA is dominated by median simulated time of travel in the zone of elevated pH (pH > 8). Consisting of one dominant variable and explaining only a small amount (10%) of the variance of the data, this component does not provide useful information but is retained for consistency of presentation.

The most notable results of the third PCA are associations between uranium concentrations and both general and redox-based particle-tracking variables. These results again indicate that particle-tracking variables can be useful in understanding public supply well vulnerability, in this case for vulnerability to uranium.

Spearman correlation analysis

Spearman correlation analysis was used to evaluate correlations of arsenic and uranium with selected explanatory particle-tracking variables that represent processes responsible for arsenic and uranium mobilization and transport to public supply wells (Table 2). Although individual correlations were only moderately strong (absolute value of Spearman's rho ≤ 0.49), all correlations listed in Table 2 exhibited statistical significance ($p < 0.01$, and commonly < 0.0001).

Simulated time of travel was positively correlated with arsenic, and to a lesser extent with uranium concentrations (Table 2). Also, the percent of water flux to wells that was simulated to be young was inversely correlated with arsenic and uranium concentrations. These results are consistent with the hypothesis that particle-tracking variables can be used to characterize natural contaminant concentrations in public supply wells.

Redox particle-tracking variables were evaluated with the expectation that geochemically based particle-tracking variables might be useful for characterizing arsenic and uranium vulnerability by capturing both components of residence time to account for the kinetics of reactions and components of geochemical conditions to account for specific mobilization mechanisms. Analysis of arsenic may have been complicated by the potential for mobilization under both oxidizing and reducing conditions. Arsenic was positively correlated with several variables representing simulated time of travel through O₂- and NO₃-reducing zones, but not with variables representing simulated time of travel through Fe- and Mn-reducing zones (Table 2). However, these correlations with redox-based particle-tracking variables were weaker than were the correlations with general particle-tracking variables (variables without geochemical zonation), and likely reflect correlation between general time of travel and time of travel through O₂- and NO₃-reducing zones. For example, median simulated time of travel was correlated with median simulated time of travel in the O₂- and NO₃-reducing zone (Spearman's rho 0.83, $p < 0.0001$), but the correlation with median simulated time of travel in the Fe- and Mn-reducing zone was not statistically significant. Thus, the weaker correlations with variables representing simulated time of travel through O₂- and NO₃-reducing zones may reflect the importance of time of travel in general, rather than time of travel through oxidizing zones.

For uranium, Spearman correlation analysis of redox particle-tracking variables supports their use in characterizing uranium mobilization and fate (Table 2). Uranium was positively correlated with variables representing simulated time of travel through O₂- and NO₃-reducing zones, and inversely correlated with variables representing simulated time of travel through Fe- and Mn-reducing zones. Fluxes of water from O₂- and NO₃-reducing zones also were positively correlated with uranium concentrations. Statistically significant correlations between uranium and particle-tracking variables that account for redox conditions were more numerous and

Table 2
Results of Spearman correlation analysis with particle-tracking variables associated with public supply wells and their contributing zones. Only significant ($p < 0.01$) results are shown, such that a blank indicates $p \geq 0.01$ (rho, Spearman's correlation coefficient; TT, simulated time of travel; O₂NO₃, oxygen- and nitrate-reducing; FeMn, iron- and manganese-reducing; variable names in parentheses are abbreviated names used in Fig. 2).

Variable	Arsenic (138 wells)		Uranium (128 wells)	
	rho	p	rho	p
<i>General particle-tracking variables</i>				
Mean simulated TT	0.37	<0.0001		
Minimum simulated TT (TT min)	0.38	<0.0001	0.30	0.0005
Simulated TT, 10th percentile	0.39	<0.0001	0.25	0.0044
Simulated TT, 25th percentile	0.39	<0.0001	0.24	0.0058
Simulated TT, 50th percentile (TT median)	0.38	<0.0001		
Simulated TT, 75th percentile	0.40	<0.0001		
Simulated TT, 90th percentile	0.36	<0.0001		
Maximum simulated TT (TT max)				
Percent of simulated flux with TT < 10 years (%TT 10–)	–0.23	0.0062	–0.36	<0.0001
Percent of simulated flux with TT < 20 years	–0.33	<0.0001		
Percent of simulated flux with TT < 30 years	–0.37	<0.0001		
Percent of simulated flux with TT < 40 years	–0.41	<0.0001		
Percent of simulated flux with TT < 50 years	–0.41	<0.0001		
Percent of simulated flux with TT < 60 years	–0.40	<0.0001		
Percent of simulated flux with TT < 100 years	–0.34	<0.0001		
Percent of simulated flux with TT < 200 years	–0.28	0.0010		
Percent of simulated flux with TT \geq 200 years (%TT 200+)	0.28	0.0010		
<i>Redox particle-tracking variables</i>				
Percent of simulated flux from O ₂ NO ₃ zone (%Flux O ₂)			0.48	<0.0001
Percent of simulated flux from FeMn zone			–0.48	<0.0001
Simulated percent of zone of contribution = O ₂ NO ₃			0.49	<0.0001
Simulated percent of zone of contribution = FeMn			–0.49	<0.0001
Mean simulated TT through O ₂ NO ₃ zone	0.29	0.0005	0.28	0.0014
10th Percentile of simulated TT through O ₂ NO ₃ zone	0.29	0.0006	0.31	0.0003
25th Percentile of simulated TT through O ₂ NO ₃ zone	0.27	0.0015	0.30	0.0005
50th Percentile of simulated TT through O ₂ NO ₃ zone (TT O ₂ median)	0.28	0.0010	0.30	0.0006
75th Percentile of simulated TT through O ₂ NO ₃ zone	0.32	0.0001	0.28	0.0013
90th Percentile of simulated TT through O ₂ NO ₃ zone	0.30	0.0004	0.28	0.0015
Mean simulated TT through FeMn zone			–0.36	<0.0001
10th Percentile of simulated TT through FeMn zone			–0.43	<0.0001
25th Percentile of simulated TT through FeMn zone			–0.40	<0.0001
50th Percentile of simulated TT through FeMn zone (TT Fe median)			–0.48	<0.0001
75th Percentile of simulated TT through FeMn zone			–0.38	<0.0001
90th Percentile of simulated TT through FeMn zone			–0.37	<0.0001
Maximum simulated TT for particles from O ₂ NO ₃ zone			0.28	0.0013
Maximum simulated TT for particles from FeMn zone			–0.37	<0.0001
Percent of simulated TT in O ₂ NO ₃ zone			0.45	<0.0001
Percent of simulated TT in FeMn zone (%TT Fe)			–0.45	<0.0001
<i>pH particle-tracking variables</i>				
Mean simulated TT through pH > 8 zone				
10th Percentile of simulated TT through pH > 8 zone				
50th Percentile of simulated TT through pH > 8 zone (TT pH8 median)				
90th Percentile of simulated TT through pH > 8 zone				

generally stronger than those with particle-tracking variables that did not account for redox conditions. Thus, redox-based particle-tracking variables that capture the uranium mobilization and attenuation processes of mobilization under oxic conditions and attenuation under reducing conditions explain important aspects of public supply well vulnerability to natural contaminants.

Spearman correlation analysis primarily focused on particle-tracking-derived variables because they can be used in predictive applications. However, Spearman correlation analysis using aqueous geochemical data can sometimes elucidate correlations between natural contaminants and redox conditions. With the aqueous geochemical dataset, uranium showed statistically significant ($p < 0.01$) correlations with NO₃ (positive; Spearman's rho +0.41) and Fe (inverse, but weak; Spearman's rho –0.17). This association of uranium with oxidizing conditions is consistent with the results of the PCA. Arsenic was not significantly correlated with NO₃, Mn, or Fe, and was significantly but only weakly correlated with O₂ (inverse; Spearman's rho = –0.20). These results indicate that at this scale of investigation, it is difficult to identify competing arsenic mobilization processes in a combination of diverse aquifers with the application of simple correlation analysis.

Classification tree analysis

Classification trees were constructed using predictor (particle-tracking and solid-phase) variables to determine if these variables could be used to predict general patterns of arsenic or uranium occurrence in water from public supply wells (response variables). In this paper, we present one classification tree and test it against an independent dataset. More complex classification tree models could be developed and applied at other scales, given sufficient data and understanding; such applications were beyond the scope of this paper.

Solid-phase arsenic and uranium concentrations among study areas were included as variables in classification tree model building to represent source strength. Site-specific solid-phase data were not available for this study, and the use of study-unit scale streambed sediment variables rather than site-specific solid-phase data is a substantial limitation. Mobilization processes often overwhelm effects of solid-phase arsenic concentrations (e.g., van Geen et al., 2003). However, several studies at local to regional scales have demonstrated relations between solid-phase and aqueous arsenic concentrations (Robertson, 1989; Peters et al., 1999; Schrei-

ber et al., 2000; Ayotte et al., 2003; Gotkowitz et al., 2004; Polizzotto et al., 2008). Ayotte et al. (2006) found that arsenic in streambed sediment was a significant and useful predictive variable for a ground-water arsenic logistic-regression model. Streambed arsenic concentrations were correlated with arsenic concentrations in both bedrock and in ground water in a multi-state region in New England (Robinson and Ayotte, 2006). Relations between aqueous and solid-phase uranium concentrations appear to have received less attention than has been the case for arsenic.

Classification trees that were reasonably accurate (those that correctly specified aqueous concentration ranges in 70% or more of the cases) were elusive. Classification trees generally require larger datasets than many other types of statistical analysis (Balk and Elder, 2000); the particle-tracking variables, although numerous, were largely redundant; and the solid-phase data were non-site-specific. Nevertheless, one satisfactory classification tree model was derived for arsenic (Fig. 3). The model is simple, having three splits and four nodes. The variables are: minimum simulated time of travel, streambed sediment arsenic concentration, and percent of simulated water flux with time of travel ≥ 200 years. The model assigns ground water from a public supply well in this dataset with minimum simulated time of travel <5.1 years to the “As ≤ 2 $\mu\text{g/L}$ ” category. Conceptually, this split is consistent with the results of the PCA and Spearman’s correlation analysis, where arsenic concentrations were associated with time of travel.

The model next splits on streambed sediment arsenic concentration. Within this category (minimum simulated time of travel >5.1 years), water from public supply wells in regions with streambed sediment arsenic concentration <12 mg/kg are considered most likely to be in the intermediate “2 $<$ As ≤ 7 $\mu\text{g/L}$ ” category. Ground water in regions with streambed sediment arsenic concentration >12 mg/kg is split on a third variable, percent of simulated water flux with time of travel ≥ 200 years. Ground water in public supply wells with a greater proportion of long simulated time of travel is assigned to the “As > 7 $\mu\text{g/L}$ ” category; otherwise, the assignment is to “As ≤ 2 $\mu\text{g/L}$ ” category. The splits on streambed sediment arsenic concentration and percent of simulated water flux with time of travel ≥ 200 years are consistent with the expectation that older water in environments with greater arsenic concentrations in solid phases should yield greater arsenic concentrations in the aqueous phase.

We were not successful in generating robust uranium classification trees. In addition to the challenges discussed above, a classification tree for uranium might be limited by the particular importance exerted by complexing agents on uranium solubility, hinted at by PCA results. Representing the effects of aqueous complexes in a predictive model could require a priori knowledge of

the distribution of complexing species, requiring a separate predictive model. Particle-tracking and solid-phase variables have the potential to contribute to vulnerability assessments, as demonstrated by the arsenic classification tree. However, successful application of these variables in vulnerability assessments will depend, among other factors, on the particular contaminants of concern, the scale of investigation, and on the other types of process-based variables that might be available.

Conclusions

Evaluation of ground-water vulnerability to natural contaminants remains a challenging endeavor. Combining flow-model analysis with geochemical data to evaluate ground-water vulnerability, in spite of its promise, has seen little application to date. Here, particle-tracking analysis and geochemical data were used to assess public supply well vulnerability to arsenic and uranium at large spatial scales in a multi-study-area dataset. Three statistical methods were used to characterize patterns of arsenic and uranium occurrence; identify relevant particle-tracking and geochemical variables that represent the dominant processes controlling their mobilization, transport and fate; and demonstrate the use of these variables in vulnerability assessment.

- PCA results were consistent with uranium mobilization under oxidizing conditions and attenuation under reducing conditions. Uranium was associated with particle-tracking variables that accounted for residence time and that accounted for redox conditions encountered by water particles. Analysis of arsenic with PCA suggested that mobilization is stronger under reducing conditions than oxidizing conditions in the overall dataset. Arsenic was associated with particle-tracking variables that accounted for overall residence time as well as residence time in reducing zones and in zones of elevated pH.
- Spearman correlation analysis demonstrated statistical significance in the correlation between uranium concentrations and particle-tracking variables, with strongest correlations for particle-tracking variables that accounted for aquifer redox conditions. Arsenic concentrations were statistically correlated primarily with general (independent of redox) particle-tracking variables.
- A classification tree model demonstrated an application of some of these variables in a screening-level predictive model for public supply well vulnerability to arsenic.

Vulnerability assessments often focus on anthropogenic contaminants, and often implicitly or explicitly attribute a lower vulnerability to relatively old water. A potentially underappreciated consequence of this focus could be greater vulnerability of ground water to natural contaminants. This could arise from the tendency for concentrations of many natural contaminants to increase with increasing ground-water residence time, a pattern that was observed in this multi-study-area dataset and shown to be useful in vulnerability analysis. Comprehensive vulnerability assessments might benefit from expanded use of estimates of ground-water age and other flow-model-based variables in combination with an awareness of the possible balances between anthropogenic and natural contaminant vulnerability.

Acknowledgements

This work was funded by the USGS NAWQA Program. Sarah J. Ryker and Alan H. Welch contributed to the development of the project design. Jeffrey N. Grossman, Andrew E. Grosz, David B. Smith, Steven M. Smith and V. Cory Stephens provided solid-phase data and guidance.

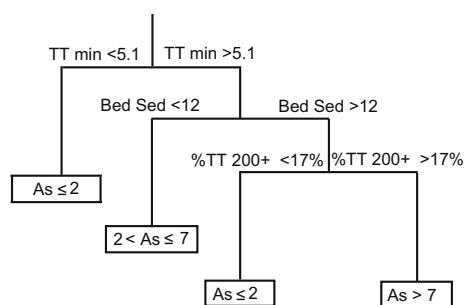


Fig. 3. Classification tree model for arsenic in the combined study areas. TT min, minimum simulated time of travel, in years. Bed Sed, streambed sediment arsenic concentration, in milligrams per kilogram. %TT 200+, percent of simulated water flux with time of travel ≥ 200 years. As, predicted aqueous arsenic concentration in ground water, in $\mu\text{g/L}$. Arsenic classification tree model accuracy on learning data subset, 70% ($n = 109$); on testing data subset, 79% ($n = 29$).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhydrol.2009.07.020.

References

- Aller, L., Bennett, T., Lehr, J.H., Petty, R.J., Hackett, G., 1987. DRASTIC: a standardized system for evaluating ground water pollution potential using hydrogeologic settings. US Environmental Protection Agency Report EPA 600-2-87-035, 622 pp.
- Arbogast, B.F., 1990. Quality assurance manual for the Branch of Geochemistry. US Geological Survey Open-File Report 90-668, 184 pp.
- Ayotte, J.D., Montgomery, D.L., Flanagan, S.M., Robinson, K.W., 2003. Arsenic in groundwater in eastern New England: occurrence, controls, and human health implications. *Environmental Science and Technology* 37, 2075–2083.
- Ayotte, J.D., Nolan, B.T., Nuckols, J.R., Cantor, K.P., Robinson Jr., G.R., Baris, D., Hayes, L., Karagas, M., Bress, W., Silverman, D.T., Lubin, J.H., 2006. Modeling the probability of arsenic in groundwater in New England as a tool for exposure assessment. *Environmental Science and Technology* 40, 3578–3585.
- Balk, B., Elder, K., 2000. Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed. *Water Resources Research* 36, 13–26.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. CRC Press, Boca Raton, Louisiana, 358 pp.
- Briggs, P.H., Meier, A.L., 1999. The determination of forty-two elements in geological materials by inductively coupled plasma-mass spectrometry. US Geological Survey Open-File Report 99-166, 15 pp.
- Cothern, C.R., Lappenbusch, W.L., 1983. Occurrence of uranium in drinking water in the US. *Health Physics* 45, 89–99.
- Davis, J.A., Meece, D.E., Kohler, M., Curtis, G.P., 2004. Approaches to surface complexation modeling of uranium(VI) adsorption on aquifer sediments. *Geochimica et Cosmochimica Acta* 68, 3621–3641.
- Dzombak, D.A., Morel, F.M.M., 1990. *Surface Complexation Modeling—Hydrous Ferric Oxide*. Wiley, New York, 393 pp.
- Eberts, S.M., Erwin, M.L., Hamilton, P.A., 2005. Assessing the vulnerability of public-supply wells to contamination from urban, agricultural, and natural sources. US Geological Survey Fact Sheet 2005-3022, 4 pp.
- Focazio, M.J., Reilly, T.E., Rupert, M.G., Helsel, D.R., 2002. Assessing ground-water vulnerability to contamination: providing scientifically defensible information for decision makers. US Geological Survey Circular 1224, 33 pp.
- Focazio, M.J., Tipton, D., Shapiro, S.D., Geiger, L.H., 2006. The chemical quality of self-supplied domestic well water in the United States. *Ground Water Monitoring & Remediation* 26, 92–104.
- Garrels, R.M., Christ, C.L., 1990. *Solutions, Minerals, and Equilibria*, second ed. Jones and Bartlett, Boston, 450 pp.
- Gilliom, R.J., Alley, W.M., Gurtz, M.E., 1995. Design of the National Water-Quality Assessment Program: occurrence and distribution of water-quality conditions. US Geological Survey Circular 1112, 33 pp.
- Gotkowitz, M.B., Schreiber, M.E., Simo, J.A., 2004. Effects of water use on arsenic release to well water in a confined aquifer. *Ground Water* 42, 568–575.
- Grossman, J.N., 2004. US Geological Survey, Written Communication.
- Grosz, A.E., Grossman, J.N., Garrett, R., Friske, P., Smith, D.B., Darnley, A.G., Vowinkel, E., 2004. A preliminary geochemical map for arsenic in surficial materials of Canada and the United States. *Applied Geochemistry* 19, 257–260.
- Gustavsson, N., Bølviken, B., Smith, D.B., Severson, R.C., 2001. Geochemical landscapes of the conterminous United States—new map presentations for 22 elements. US Geological Survey Professional Paper 1648, 38 pp.
- Harbaugh, A.W., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW-2000, The US Geological Survey modular ground-water model—user guide to modularization concepts and the ground-water flow process. US Geological Survey Open-File Report 2000-92, 121 pp.
- Helsel, D.R., Hirsch, R.M., 1992. *Statistical Methods in Water Resources*. Elsevier, Amsterdam, 522 pp.
- Hem, J.D., 1985. Study and interpretation of the chemical characteristics of natural water, third ed. US Geological Survey Water-Supply Paper 2254, 263 pp.
- Hill, M.C., 1998. Methods and guidelines for effective model calibration. US Geological Survey Water-Resources Investigations Report 98-4005, 90 pp.
- Hutson, S.S., Barber, N.L., Kenny, J.F., Linsey, K.S., Lumia, D.S., Maupin, M.A., 2004. Estimated use of water in the United States in 2000. US Geological Survey Circular 1268, 46 pp.
- Insightful Corporation, 2001. S-PLUS 6 for Windows User's Guide. Insightful Corporation, Seattle, 688 pp.
- Langmuir, D., 1978. Uranium solution-mineral equilibria at low temperatures with applications to sedimentary ore deposits. *Geochimica et Cosmochimica Acta* 42, 547–569.
- Lovley, D.R., Phillips, E.J.P., Gorbey, Y.A., Landa, E.R., 1991. Microbial reduction of uranium. *Nature* 350, 413–416.
- McMahon, P.B., Chapelle, F.H., 2008. Redox processes and water quality of selected principal aquifer systems. *Ground Water* 46, 259–271.
- Miller, J.A. (Ed.), 2000. *Ground water atlas of the United States*. US Geological Survey, Reston, Virginia.
- Nordstrom, D.K., 2002. Worldwide occurrences of arsenic in ground water. *Science* 296, 2143–2145.
- Paschke, S.S. (Ed.), 2007. Hydrogeologic settings and ground-water flow simulations for regional studies of the transport of anthropogenic and natural contaminants to public-supply wells – studies begun in 2001. US Geological Survey Professional Paper 1737-A, 288 pp.
- Peters, S.C., Blum, J.D., Klaue, B., Karagas, M.R., 1999. Arsenic occurrence in New Hampshire drinking water. *Environmental Science and Technology* 33, 1328–1333.
- Plant, J.A., Kinniburgh, D.G., Smedley, P.L., Fordyce, F.M., Klinck, B.A., 2005. Arsenic and selenium. In: Lollar, B.S. (Ed.), *Environmental Geochemistry*. Elsevier, Amsterdam, pp. 17–66.
- Polizzotto, M.L., Kocar, B.D., Benner, S.G., Sampson, M., Fendorf, S., 2008. Near-surface wetland sediments as a source of arsenic release to ground water in Asia. *Nature* 454, 505–508.
- Pollock, D.W., 1994. User's guide for MODPATH/MODPATH-PLOT, Version 3; a particle tracking post-processing package for MODFLOW, the US Geological Survey finite-difference ground-water flow model. US Geological Survey Open-File Report 94-464, 248 pp.
- Preisendorfer, R.W., Mobley, C.D., 1988. *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, Amsterdam, 425 pp.
- Qian, S.S., Anderson, C.W., 1999. Exploring factors controlling the variability of pesticide concentrations in the Willamette River Basin using tree-based models. *Environmental Science and Technology* 33, 3332–3340.
- Robertson, F.N., 1989. Arsenic in ground-water under oxidizing conditions, south-west United States. *Environmental Geochemistry and Health* 11, 171–185.
- Robinson Jr., G.R., Ayotte, J.D., 2006. The influence of geology and land use on arsenic in stream sediments and ground waters in New England, USA. *Applied Geochemistry* 21, 1482–1497.
- Ryker, S.J., 2003. Arsenic in ground water used for drinking water in the United States. In: Welch, A.H., Stollenwerk, K.G. (Eds.), *Arsenic in Ground Water*. Kluwer, Boston, pp. 165–178.
- Sanzolone, R.F., Ryder, J.L., 1989. Quality assessment program and results for the NAWQA surface water pilot studies. US Geological Survey Open-File Report 89-658, 22 pp.
- SAS Institute, 1989. SAS/STAT®, User's Guide, Version 6, fourth ed., vol. 2. SAS Institute, Cary, North Carolina.
- Schreiber, M.E., Simo, J.A., Freiberg, P.G., 2000. Stratigraphic and geochemical controls on naturally occurring arsenic in groundwater, eastern Wisconsin, USA. *Hydrogeology Journal* 8, 161–176.
- Shelton, L.R., Capel, P.D., 1994. Guideline for collecting and processing samples of stream bed sediment for analysis of trace elements and organic contaminants for the National Water-Quality Assessment Program. US Geological Survey Open-File Report 94-458, 20 pp.
- Smedley, P.L., Kinniburgh, D.G., 2002. A review of the source, behavior and distribution of arsenic in natural waters. *Applied Geochemistry* 17, 517–568.
- Spitz, F.J., 2001. Method and computer programs to improve pathline resolution near weak sinks representing wells in MODFLOW and MODPATH ground-water-flow simulations. US Geological Survey Open-File Report 2000-392, 41 pp.
- Spruill, T.B., Showers, W.J., Howe, S.S., 2002. Application of classification-tree methods to identify nitrate sources in ground water. *Journal of Environmental Quality* 31, 1538–1549.
- Stephens, V.C., 2003. US Geological Survey, Written Communication.
- US Environmental Protection Agency, 1999. A review of contaminant occurrence in public water systems. US Environmental Protection Agency Report EPA 816-R-99-006, 78 pp.
- van Geen, A., Zheng, Y., Versteeg, R., Stute, M., Horneman, A., Dhar, R., Steckler, M., Gelman, A., Small, C., Ahsan, H., Graziano, J.H., Hussain, I., Ahmed, K.M., 2003. Spatial variability of arsenic in 6000 tube wells in a 25 km² area of Bangladesh. *Water Resources Research* 39, HWC 3-1–HWC 3-16.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, fourth ed. Springer, New York, 495 pp.
- Wanty, R.B., Nordstrom, D.K., 1993. Natural radionuclides. In: Alley, W.M. (Ed.), *Regional Ground-Water Quality*. Van Nostrand Reinhold, New York, pp. 423–441.
- Welch, A.H., Westjohn, D.B., Helsel, D.R., Wanty, R.B., 2000. Arsenic in ground water of the United States: Occurrence and geochemistry. *Ground Water* 38, 589–604.
- Witkowski, A.J., 2007. Groundwater vulnerability assessment and mapping important current issues in groundwater protection. *Environmental Geology* 53, 477–478.
- Zobrist, J., Dowdle, P.R., Davis, J.A., Oremland, R.S., 2000. Mobilization of arsenite by dissimilatory reduction of adsorbed arsenate. *Environmental Science and Technology* 34, 4747–4753.