Spatial Trends in Groundwater Arsenic Concentrations

Jonathan L. Bradshaw, Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, jbradsha@princeton.edu

Patrick L. Gurian, Department of Civil, Architectural, and Environmental Engineering, Drexel University, Philadelphia, PA 19104, pgurian@drexel.edu

Arun Kumar, Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824, akumar@msu.edu

David E. Breen Department of Computer Science, Drexel University, Philadelphia, PA 19104, david@cs.drexel.edu

ABSTRACT

Arsenic presents complex spatial occurrence trends that can be difficult to identify and understand. This project sought to understand geographic trends in arsenic occurrence using a visualization technique. The approach taken was to link geospatially referenced arsenic concentration information from a water quality database with elevation data contained in Digital Terrain Elevation Data (DTED) files. DTED files are available for all land masses across the world for public download. This allows for the development of three-dimensional plots of arsenic concentration and topography. The plots developed in this manner show that high arsenic is associated with the transition from plains to piedmont on the western side of the Delaware River Valley in New Jersey. In Oklahoma high arsenic is found along the North Canadian River Valley. In New Mexico high concentrations are generally high in the Rio Grande Valley but with an area of low concentration in the southern portion of this valley. In California, arsenic concentrations are high in the middle of the Central Valley but moderate somewhat toward the edges. These results are consistent with mobilization of arsenic by reductive processes in the organic-rich sediments of river valleys, but further statistical analysis is required to confirm the significance of this association. The visualization software used here is broadly applicable and a user guide for this software is available on request.

Keywords: Arsenic, Groundwater, Software, Visualization, Water quality

INTRODUCTION

Abundant archival data on arsenic in groundwater are available, but these data are often presented in forms that are not conducive to straight-forward interpretation, such as tables. Previous efforts to understand arsenic spatial occurrence trends have generally used statistical methods [1,2] which provide rigor in assessing the significance of trends, but are difficult to implement and interpret. Visualization can present data in a manner that facilitates the interpretation of complex information for experts and laypersons alike [3-6]; both can use it to visually identify correlations in the data, such as water contaminant hotspots and water quality parameters. This can be especially beneficial to experts, who can develop hypotheses to explain any visual correlations and then test them statistically. Furthermore, visualization can help identify areas where data are sparse and more data collection is needed. This project sought to develop a simple, user-friendly, and interactive visualization technique that visually communicates complex information from otherwise incomprehensible databases. The visualization approach was used to better understand which topographic features are associated with elevated arsenic concentrations.

METHODS

This study used the United States Geological Survey (USGS) arsenic point database, which has water quality data for the United States and its territories. Each row in the database represents a different well. Organized by column, the database includes USGS station identification number, Federal Information Processing Standard (FIPS) state code, FIPS county code, latitude, longitude, depth, water usage code (e.g. private, industrial, municipal, etc.), site use code, aquifer code, and 31 different constituent values (e.g. arsenic, uranium, chloride concentrations, and pH).

MATLAB software was developed which allows the user to specify the subset columns within a Graphical User Interface (GUI). Any number of constituents may be listed in the database. Constituent information can only be numbers, and blank cells are acceptable. The software converts latitude and longitude from Degrees-Minutes-Seconds—written as (d)ddmmss—to decimal degrees, though with only small programming modifications any kind of spatial coordinate system is acceptable.

Digital Terrain Elevation Data (DTED) files provide land elevation data useful for both creating a surface plot and estimating well elevations, facilitating the user's recognition of a geographic area and also permitting the user to view a well's depth when adjusted to local elevation. DTED files are available for all land masses across the world. The National Geospatial-Intelligence Agency (NGIA) provides DTED level 0 (resolution of 30-arcseconds) for public download.

The contaminant database is loaded directly into the MATLAB programming environment from a Microsoft Excel spreadsheet file (Figure 1). Each column that the user selects determines the plotting criteria and is stored as an individual vector variable. The constituents' values are all stored in one matrix: each column contains values for one constituent, and each row contains values for an individual well. All constituent rank calculations (which determine color) are made after well and constituent selection. A function from MATLAB's mapping toolbox determines which DTED files are required to approximate the selected wells' elevations, and the appropriate DTED file is assigned to each well. Another mapping toolbox function translates the DTED *.dt0 files into an elevation grid and reference vector. The elevation of each well is approximated by locating the well's position within the elevation grid and calculating the mean elevation of the nine closest points. The three-dimensional terrain is plotted using a mapping toolbox function that takes as input the elevation grid and reference vector produced from reading the DTED file.

For detailed visualization of smaller geographic areas (roughly on the county level or smaller) wells are indicated by icons that indicate well depth (by height), well type (by color of the cylindrical shell and outer circle) and constituent concentrations (by the darkness of colors of different sectors within the cylindrical shell of the icon). Wells are represented on the plot by using vertical (i.e. parallel to the altitude axis) lines with dot markers on each end. The line's length is equal to the well's depth. The location of a well is determined by its latitude, longitude, and approximated elevation. Instead of incorporating and presenting all of a well's constituent and type information at once, less detailed plots visualize a single constituent with a color and size used in the detailed plots.

RESULTS

The visualization system was utilized to explore and evaluate arsenic occurrence in New Jersey, New Mexico, Oklahoma, and California to demonstrate the usefulness and effectiveness of this visualization approach. These states were identified by previous studies as areas with elevated arsenic levels [7,8] and during exploratory analysis were found to show spatial trends of interest.

Figure 2 shows a horizontal and vertical plot of the distributions of arsenic in New Jersey. The area displayed extends from 38 to 42 decimal degrees N and -76 to -73 decimal degrees E. Wells are shown in circles and changes in arsenic concentrations can be seen by variation in darkness of color of different circle (i.e., a change of color from pink to red indicates an increase in arsenic concentration, as indicted by the legend of the figure). Arsenic concentrations in the coastal plain are mostly low. Higher arsenic concentrations are not found frequently until the coastal plain begins to transition into the piedmont on the west side of the Delaware River Valley. High arsenic concentrations have been associated with the sediments in major river valleys which are rich in organic material and hence create the reducing conditions capable of mobilizing arsenic [9]. The high arsenic wells are relatively shallow wells roughly in a line perpendicular to the elevation gradient (Figure 2), suggesting that they may tap the same geological stratum.

The spatial distribution of arsenic is also shown for Oklahoma (Figure 3). The plot reveals serious data gaps in the state as few observations are available for the north-central and southwest portions of the state. The large number of wells in the center of the plot represents the greater Oklahoma City area. The eastern side of this area has mostly low arsenic groundwater while the western side has noticeably more high arsenic wells. A string of high arsenic wells extend northwestward from the Oklahoma City area. This area of high arsenic corresponds to the transition from the lower elevations at the center of the state to the higher elevations at the western side of the state. These wells fall along the North Canadian River Valley.

New Mexico (Figure 4) also shows high arsenic concentrations in this case in the northern portion of the Rio Grande River Valley, in the Albuquerque area. Farther south along the same river valley, in the Truth or Consequences area, a region of lower arsenic concentration is found. Clearly even within the same river valley arsenic concentrations vary with localized geological conditions.

In California (Figure 5) elevated concentrations are found in the Central Valley region. However, the concentrations are somewhat lower towards the sides of the valley than in the center, another indication that even within the same valley, local geology influence arsenic occurrence trends. The plot also reveals a data gap as there are few observations in the northwestern portion of the valley.

CONCLUSIONS

Visualization of the arsenic concentration data revealed a number of localized hotspots. The visualization also identifies regions where data gaps exist. This information may be useful to alert public health agencies to potential water quality problems or where additional sampling is needed. A common theme of elevated arsenic being associated with portions of particular river valleys was noted. This is consistent with mobilization of arsenic under the reducing conditions found in organic-rich sediments of river valleys. However, not all river valleys have elevated arsenic and not all arsenic hot spots are associated with major river valleys. Further statistical analysis is needed to assess whether there is a significant association between particular topographic features and groundwater arsenic concentrations.

Although this visualization approach was specifically developed for use with the USGS arsenic point database, it can support a wide variety of databases, consisting of different water quality characteristics and geological information. Combining these sources of information with only minimal adjustments to the software or database, the software could be used to visualize relationships of different parameters or contaminant concentrations. As this system is made on a MATLAB platform, it could be further appended with statistical package to perform statistical analysis to summarize contaminant concentrations and identify significant trends. In addition, it can also visualize other environmental factors such as soil quality or air quality or even social science values. Just as it possible to visualize arsenic values in domestic and public wells of New Jersey, Oklahoma, New Mexico, and California, it is also possible to visualize other spatially-distributed information, such as water quality parameters or median neighborhood income for other places, such as India, as long as the appropriate data exists. Such visualization may help to identify emerging environmental problems and reveal environmental justice issues. A user guide to the visualization software is available on request from the authors.

ACKNOWLEDGMENTS

This research was funded by the Engineering Cities REU Program through NSF Grant No. EEC 0552792 [Engineering Cities]. All Digital Terrain Elevation Data (DTED) level 0 files are courtesy of the National Geospatial-Intelligence Agency and are available for free download at http://geoengine.nga.mil/muse-cgi-bin/rast_roam.cgi

REFERENCES

[1] Lockwood, J.R. et al., 2001. Characterization of Arsenic Occurrence in U.S. Drinking Water Treatment Facility Source Waters. *Journal of the American Statistical Association, 96*(456):1184-1193.

[2] Lockwood J.R. et al., 2004. Analysis of Contaminant Co-occurrence in Community Water Systems. *Journal of the American Statistical Association, 99*(465):45-56.

[3] Focazio M.J. et al., 2000. *A Retrospective Analysis on the Occurrence of Arsenic in Ground-water Resources of the United States and Limitations in Drinking-water-Supply Characterizations*. U.S. Geological Survey.

[4] Forgang, A B. et al., 1996. Visualization of Water Quality Data for the Chesapeake Bay. In *Proceedings of the 7th conference on Visualization '96*, 417-ff. San Francisco, California, United States: IEEE Computer Society Press. http://portal.acm.org/citation.cfm?id=245747.

[5] Stein, R. J. et al., 2000. Scientific Visualization of Water Quality in the Chesapeake Bay. In *Proceedings of the 11th IEEE Visualization 2000 Conference (VIS 2000)*. IEEE Computer Society. http://portal.acm.org/citation.cfm?id=832272.833941&coll=&dl=.

[6] Sreevalsan-Nair et al., 2007. An Interactive Visual Exploration Tool for Northern California's Water Monitoring Network. *Visualization and Data Analysis.*

[7] Welch, A.H. et al., 1999. Arsenic in Ground Water Supplies of the United States, *in*: Arsenic Exposure and Health Effects, W.R. Chappell, C.O. Abernathy and R.L. Calderon, Eds., Elsevier Science, New York, pp. 9-17.

[8] Kumar, A. et al., 2009. Arsenic Exposure in US Public and Domestic Drinking Water Supplies: A Comprehensive Risk assessment. *Journal of Exposure Science and Environmental Epidemiology.*,(in press) doi: 10.1038/jes.2009.24

[9] 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

	A	B	C	D	E	F	G	H	I	AG	AH	AI	AJ
1	staid	States	FIPS Cour	Latitude	Longitude	Well Deptl	Water Use	Site Use	Aquifer Cod	p22703	p70301	p71888	p81366
2	15S	2S	3S	9N	9N	9N	1S	1S	8S	9N	9N	9N	9N
3	295737090051201	LA	22.071	295737	900512	811	A	W	112GZNO		534		
4	300054090012601	LA	22.071	300054	900126	658	A	W	112GZNO		574		
5	353026097274801	OK	40.109	353026	972748	164	A	W	318GRBR	3.05	711		
6	393931074482101	NJ	34.001	393932	744820	165	A	W	121CKKD	0.01			1.2
20254	484304099511301	ND	38.079	484304	995113	35	n/a	O	112SLVL		743		
20255	484416099523901	ND	38.079	484416	995239	25	n/a	T	112SLVL		392		
20256	484632099513001	ND	38.079	484632	995130	54	n/a	A	n/a		503		
20257	484731099504104	ND	38.079	484731	995041	38	n/a	O	112SLVL		493		
20258	485316122322601	WA	53.073	485316	1223226	20	n/a	W	112SUMS		155		
20259	485848122272701	WA	53.073	485848	1222727		n/a	W	112SUMS		90		

EPA

Ready NUM

Figure 1. United States Geological Survey (USGS) arsenic point database, consists of almost 800,000 data points.

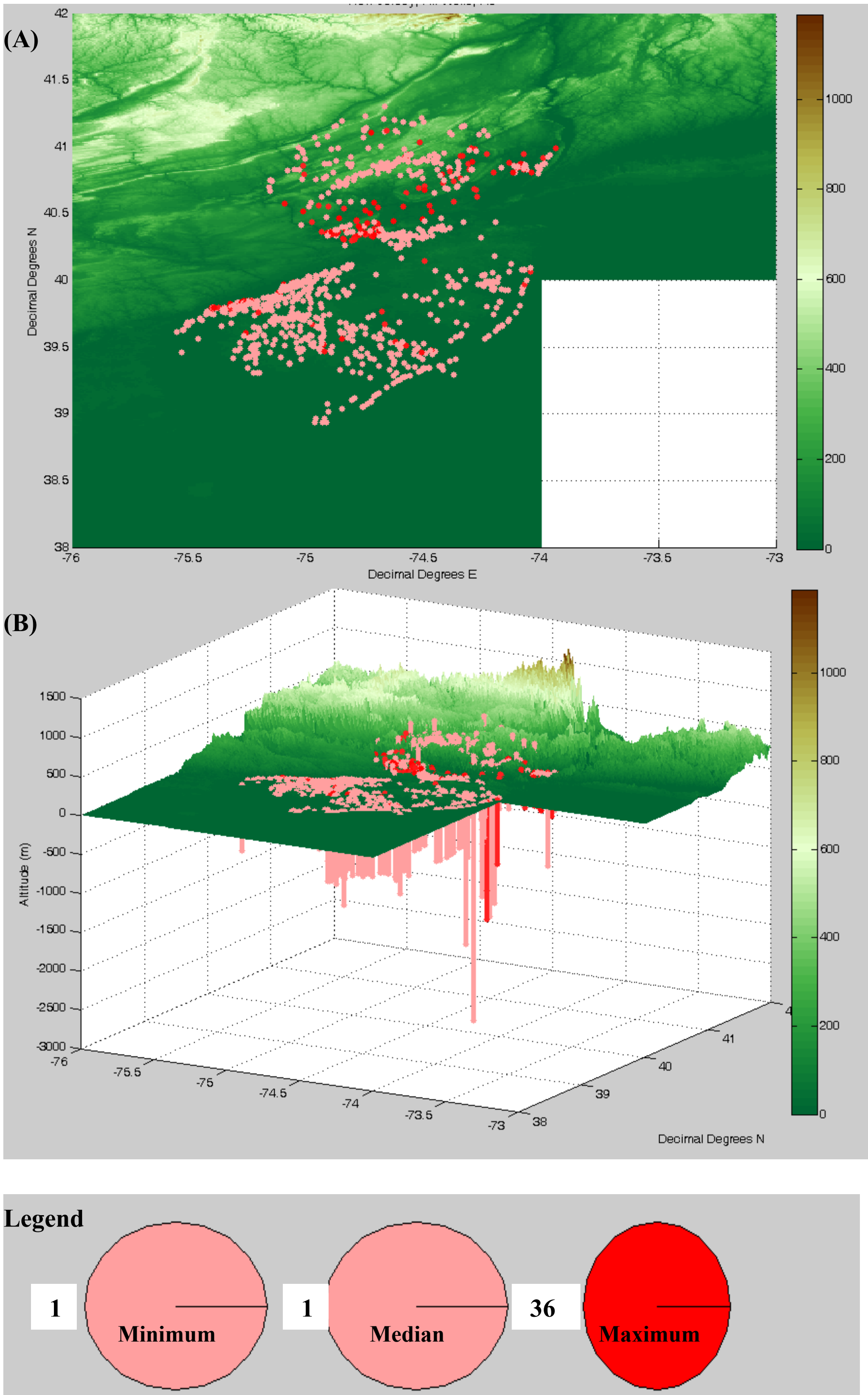

Figure 2: Visualization of arsenic distribution (in µg/L) in New Jersey for the focus area (38 to 42 decimal degrees N and -76 to -73 decimal degrees E) : (a) Horizontal spatial distribution of arsenic, (b) Vertical distribution of arsenic. An increase in arsenic concentration is indicated by an increase in darkness of color of circles.

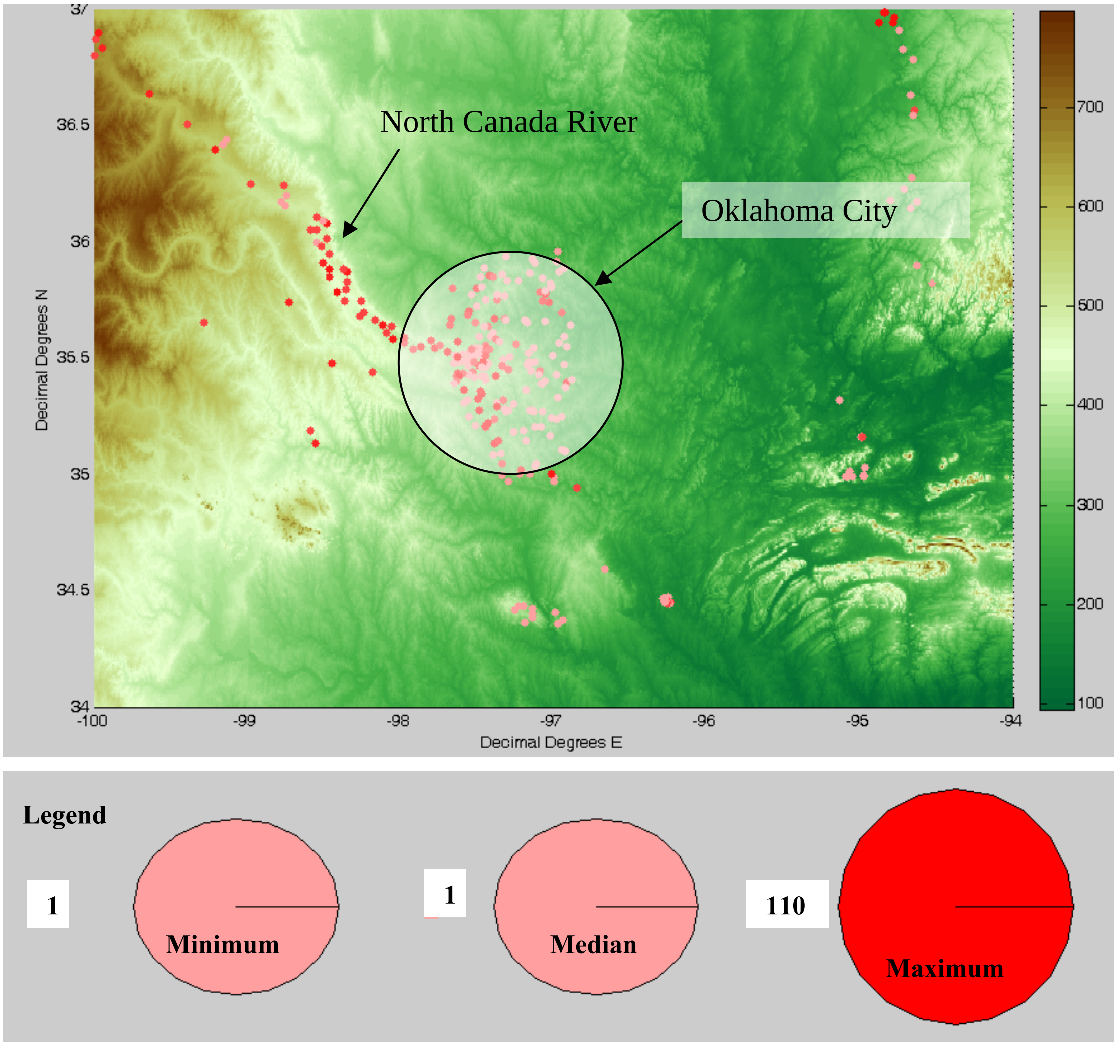

Figure 3: Visualization of spatial distribution of arsenic (in µg/L) in Oklahoma for the focus area (34 to 37 decimal degrees N and -100 to -94 decimal degrees E). An increase in arsenic concentration is indicated by an increase in darkness of color of circles.

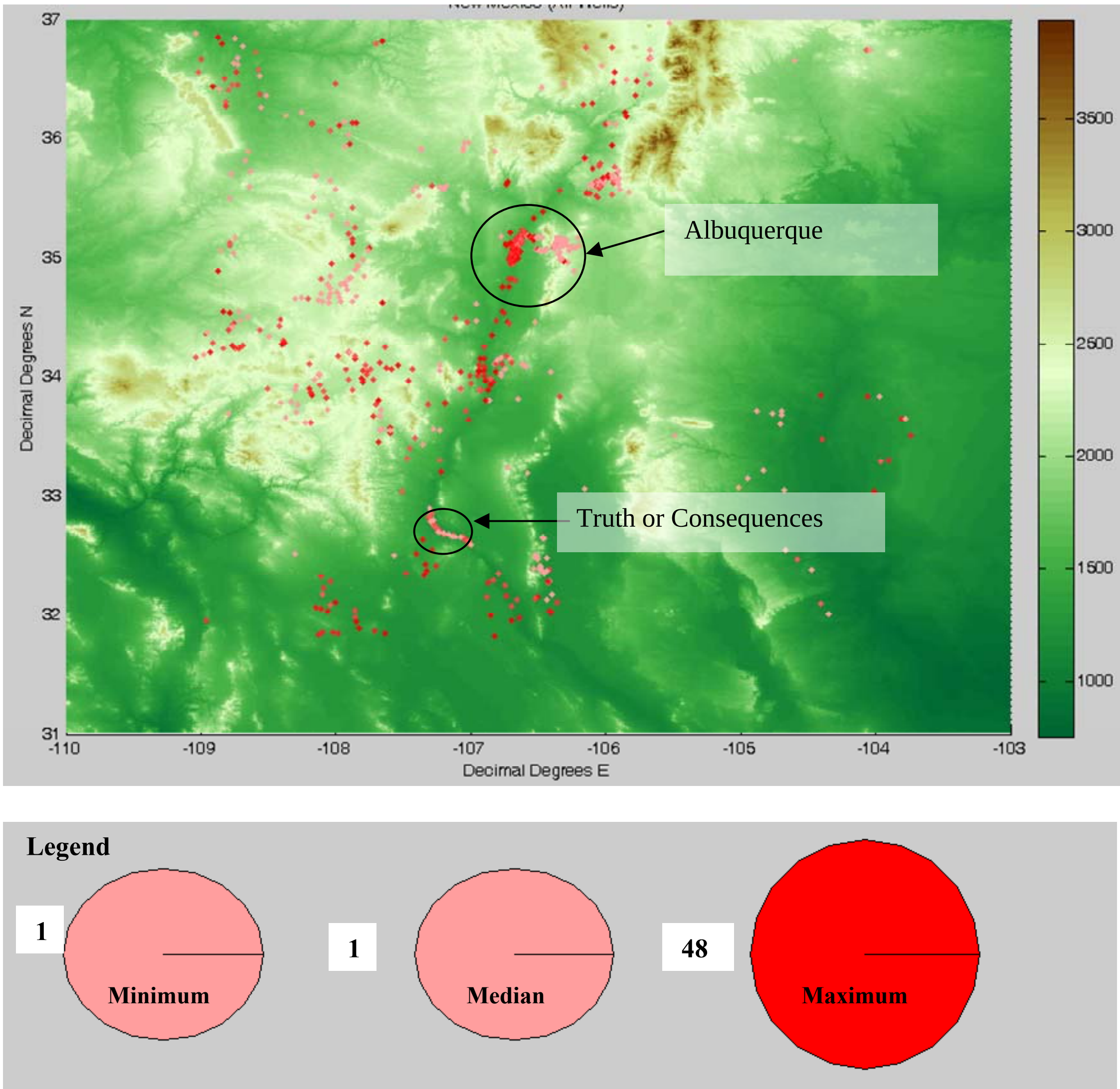

Figure 4: Visualization of spatial distribution of arsenic (in µg/L) in New Mexico for the focus area (31 to 37 decimal degrees N and -110 to -103 decimal degrees E). An increase in arsenic concentration is indicated by an increase in darkness of color of circles.

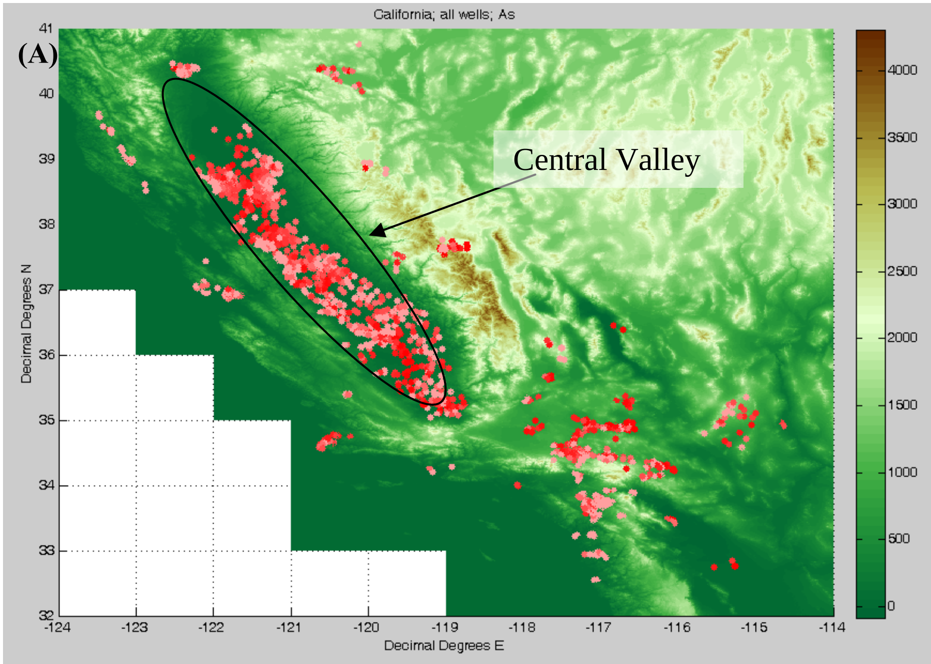

Figure 5: Visualization of spatial distribution of arsenic (in µg/L) in California (30 to 41 decimal degrees N and -124 to -114 decimal degrees E). An increase in arsenic concentration is indicated by an increase in darkness of color of dot.