

ENVISION: VISUALIZING WATER QUALITY FROM GEOGRAPHICALLY DISTRIBUTED WELLS

Jonathan L. Bradshaw

Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ USA
email: jbradsha@princeton.edu

Patrick L. Gurian

Department of Civil, Architectural, and Environmental Engineering, Drexel University, Philadelphia, PA USA
email: pgurian@drexel.edu

David E. Breen

Department of Computer Science, Drexel University, Philadelphia, PA USA
email: david@cs.drexel.edu

ABSTRACT

Understanding trends in ground water contaminant concentrations is complicated by the fact that ground water quality presents complex three-dimensional trends over multiple spatial scales. To address this issue we have developed a method to jointly display three-dimensional ground water quality and topographic data. This approach allows the user to: 1) identify local areas of elevated contaminant concentrations, 2) identify particular topographic features (river valleys, plateaus, etc) associated with elevated concentrations, 3) assess differences among aquifers and well use types, and 4) identify associations among different constituents. The approach used here links geospatially referenced concentration data with elevation data contained in Digital Terrain Elevation Data (DTED) files within an easy-to-use MATLAB-based visualization system. The method was applied to visualize information on nitrate and arsenic occurrence in a national ground water quality database. The method shows that high arsenic is associated with the transition from plains to piedmont in New Jersey. Nitrate in Iowa is shown to be associated with shallow wells in the southeastern portion of the state. The approach developed here is compatible with any Microsoft Excel Spreadsheet database which follows specific format conventions.

KEY WORDS

water quality visualization, GIS, visualization software.

1 Introduction

Groundwater is an important resource, accounting for 80% of public water supplies and providing 34% of the U.S. population with drinking water [1]. Groundwater is subject to contamination by a variety of organic [2], inorganic [3, 4] and microbial constituents [5]. Understanding trends in the occurrence of these contaminants and their affect on ground water quality is vital to protecting public health. Poor quality source waters need to be identified and potential users provided with appropriate treatment or alternative supplies.

Comprehending trends in contaminant occurrence is

complicated by the fact that ground water quality presents complex three-dimensional patterns over multiple spatial scales. Previous efforts to understand these trends have generally used statistical methods [3, 6]. These methods provide rigor in assessing the significance of trends, but are difficult to implement and interpret. Improved visualization methods are needed both 1) to perform exploratory analyses so that statistical assessments can be focused on areas of interest and 2) to provide complex information in a form readily understandable by the end user or decision-makers who may lack advanced statistical knowledge. While there are many previous examples of studies which have produced two dimensional, static maps of ground water quality [7] and two-dimensional statistical assessments of contaminant occurrence [3, 6], an interactive tool opens up the possibility of three-dimensional visualization and the possibility of simultaneously considering various factors that may influence contaminant concentrations, such as well type (drinking water supply vs. monitoring), aquifer type, and concentrations of other constituents. Such a visualization tool could both permit laypersons to readily understand complex environmental data, and could facilitate the analyses of water quality experts. Water quality specialists themselves can use visualization for exploratory data analysis in order to develop hypotheses which can then be tested mathematically.

While software exists for visualizing water quality data in specific regions, such as the Chesapeake Bay [8, 9] and Northern California [10], and commercial GIS software (e.g. ArcGIS, GMS and RockWorks) is available with spatial analysis tools, to the best of the authors' knowledge, there is no general method of jointly displaying three-dimensional water quality and topographic data at a state or national scale. The goal of our research is to provide support for the visualization of multiple well types and contaminants either individually or simultaneously, and at multiple physical scales. These capabilities allow the user to: 1) identify local areas of elevated contaminant concentrations, 2) identify particular topographic features (river valleys, plateaus, etc) associated with elevated concentrations, 3) assess differences among aquifers and well use types,

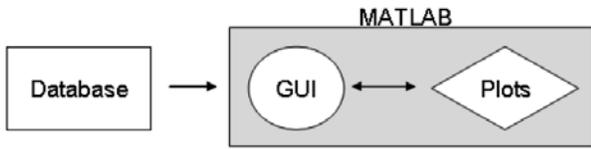


Figure 1. Workflow diagram. MATLAB reads in the database, and the user interacts with the GUI to generate plots until the desired figure is produced.

and 4) identify associations among different constituents. These capabilities have been created in a software system – ENVision – implemented in the MATLAB programming environment and utilize National Geospatial-Intelligence Agency Digital Terrain Elevation Data (DTED) data. MATLAB was chosen as our visualization framework because of the many high-level graphics, visualization, GUI and mapping capabilities that it offers. Since MATLAB is extensively used within the engineering and scientific communities, the system’s designers concluded that building ENVision within MATLAB would facilitate the system’s deployment, use and accessibility.

2 Data

ENVision was designed to be general enough for use with practically any database of geographically distributed values, but it was specifically developed for use with 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).

ENVision is currently compatible with any Microsoft Excel Spreadsheet database (*.xls) as long as it contains an identification number in the first column and three different subset columns. The subset columns specify the state, the region within the state and either the water usage, site use or aquifer type to be visualized. The user specifies the subset columns within ENVision’s 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. ENVision 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 permit-

ting 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 arc-seconds) for public download.

3 System Overview

As seen in the workflow diagram in Figure 1, the database is loaded directly into ENVision from a Microsoft Excel spreadsheet file. Each column that the user selects determines the plotting criteria and is stored as an individual vector variable. The constituents’ values are stored in one matrix: each column contains values for one constituent, and each row contains values for an individual well. Whenever the user makes a selection decision (i.e. chooses a subset of the database codes and constituents), a new binary selection vector with a length equal to the total number of wells is created to indicate which wells or constituents satisfy the selection criteria. After all selection criteria have been made, the logical AND of all these selection vectors reveals which wells and constituents satisfy all selection criteria. Once data selection has been completed, the user is able to choose two types of visualizations. The first type is a “scanner plot” which presents the data in a simplified or aggregate form. The second type displays all the selected data in a detailed 3-D plot utilizing an expressive icon.

Several different object classes are defined and implemented to facilitate data storage and access for selected wells. 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 reference vector provides the elevation grid’s geographical coordinates and is required for stitching multiple DTED files together. 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 grid points. The accuracy of the well elevation approximation is thus limited by the resolution of the DTED files and the precision of the well’s coordinates in the database.

The three-dimensional terrain is plotted using a MATLAB mapping toolbox function that takes as input the elevation grid and reference vector produced from reading the DTED file. Elevation is relayed not only spatially by the three-dimensional terrain, but also by color, as the colorbar at the right-side of the plots shows how terrain color relates to elevation. The detailed plots represent wells as a combination of surface objects and icons described in the System Details section.

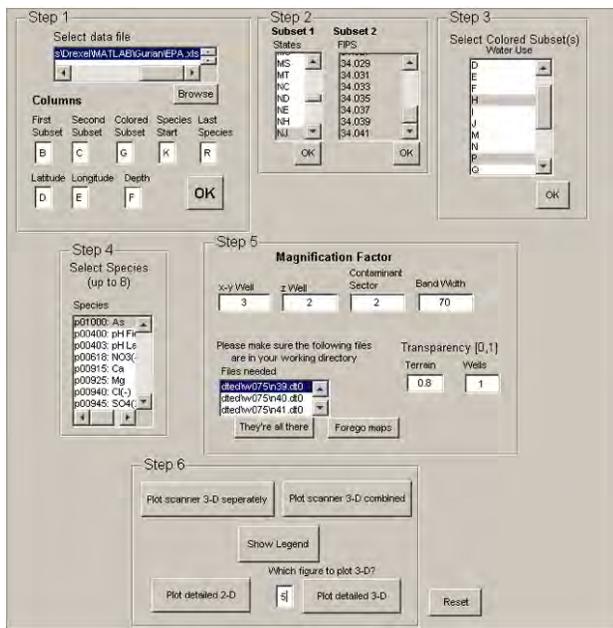


Figure 2. ENVision GUI facilitates the creation and customization of well visualization plots.

4 System Details

4.1 GUI

Using MATLAB's Graphical User Interface (GUI) layout editor, a GUI, seen in Figure 2, was implemented to facilitate visualization creation. Combined with MATLAB's built-in figure interaction tools, the GUI allows the user to move through all the computational steps of visualization – from selecting the input database to figure creation and customization – without using the MATLAB command line and without typing or even seeing any programming code. As all data organization and calculation occur behind-the-scenes, it is possible for someone with very basic computer skills and very little knowledge of either the input database or MATLAB to use ENVision.

4.2 Icon Design

Detailed 3-D plots which visually communicate all of the information selected are available in ENVision. In order to easily convey the most information possible, it is desirable to represent a well with a simple, self-contained icon. Furthermore, since viewing angle is a user preference, the selected information should be visible from all possible viewing angles. To that end, a well icon, presented in Figure 3, was designed. The icon consists of two main components, a central cylindrical shell and two capping circles, both of which consist of additional visual features related to constituent values. To permit the visibility of well information from all angles, the inner and outer circles are also reproduced on the bottom of the well.

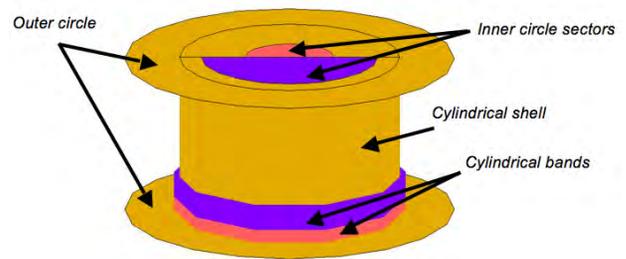


Figure 3. Well icon. This well contains two different constituents represented by the colors red and blue on the cylindrical bands and inner circle sectors. The color of the cylindrical shell and outer circle indicate the type of the well. The sector radius and band width are proportional to the constituent's concentration.

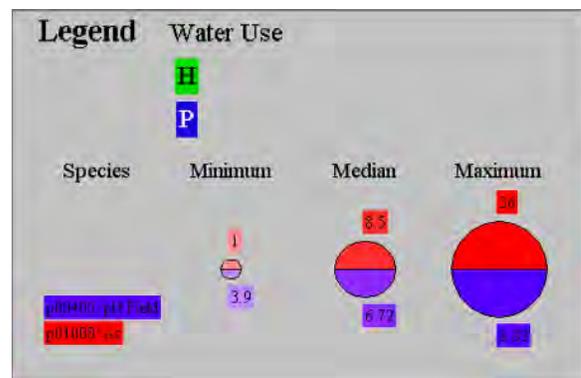


Figure 4. Well legend. This well contains two constituents “pH field” and “As” represented by the colors blue and red, respectively on the cylindrical bands and inner circle sectors. The color of the cylindrical shell and outer circle indicate the type of the well. In this case, a green well indicates a well of water usage ‘H,’ and a blue well represents a well of water usage ‘P.’ The minimum, median, and maximum values for “As” are 1, 8.5, and 36, respectively and for pH are 3.9, 6.72, and 8.83, respectively.

The color of the cylindrical shell and outer circle are identical and correspond to a specific code within one of the types selected by the user, such as water usage code, site usage code, or aquifer code. The inner circle sectors and cylindrical bands both relay the same information about that well's constituent values. Each different color sector or band represents a different constituent. Colors are described using the Hue-Saturation-Value (HSV) color representation system. The hues signifying specific codes and constituent values are spread evenly across the Hue dimension. Values and saturations of the code colors are constant across all wells. The saturations of the sector and band colors indicate how the value of that constituent ranks in a list of unique constituent values compiled from the other wells selected by the user: higher saturations correspond

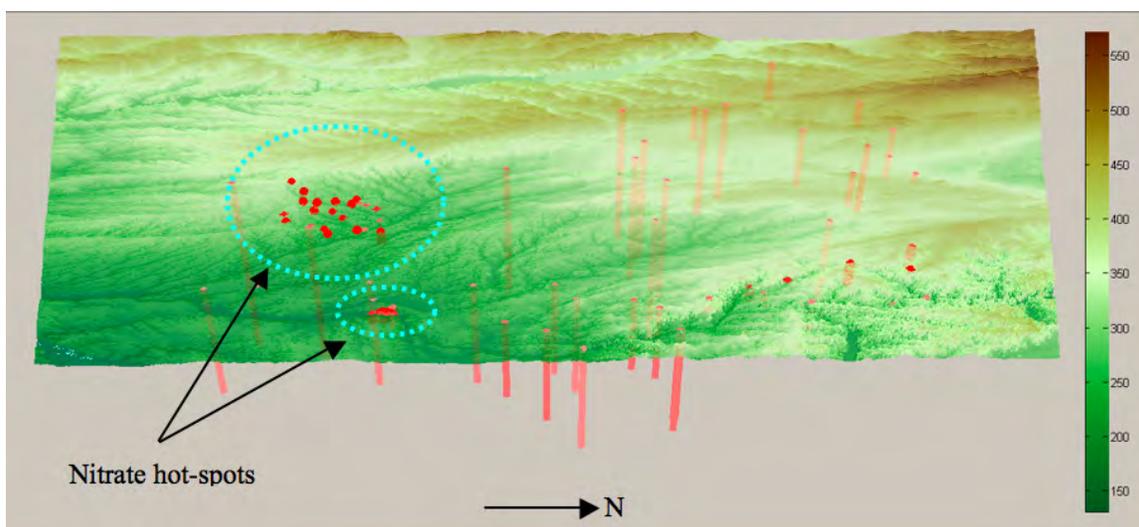


Figure 5. Scanner plot of nitrate concentrations in Iowa highlighting spatial hot spots (dashed cyan circles).

to higher constituent concentrations. A sector's radius and band width are proportional to the quintile of their respective constituent's concentrations within that same list. To make an ENVision plot easier to read, a legend specific to the plot created can be produced by pushing the "Show Legend" button in Step 6 on the GUI. The legend displays which constituents are associated with each color, and also shows the relative size and color of the plotted region's minimum, median, and maximum constituent concentrations, the values of which are also displayed in the legend. Figure 4 is an example of a legend produced for plots very similar to the New Jersey plots described in the Results section.

Sector size and the number of sectors change to accommodate the number of constituents selected. Wells that have no value for a selected constituent have an empty space where that constituent's sector would be. Similarly, these wells have no cylindrical band representing the particular constituent. The height of the well icon is proportional to the well's depth. The well is placed in the visualization according to its latitude and longitude (provided in the database) and elevation (approximated from DTED data).

4.3 Scanner and Combination Plots

Since the well icons are surface objects with a defined size in coordinate space, well size and visual comprehension limitations generally prohibit using the icons in a large-scale scenario. To see and understand these icons on a large scale, they must be quite large themselves, but making them too large can create the problem of well overlap in geographic areas with high well density. Making the well icon size a user-defined parameter in the GUI was included to address this problem. But as seen in Figure 7, there are still cases when wells are too small to be clearly

distinguishable, yet there are other wells at this same scale that overlap. Another feature added to address the well-icon sizing issue was the "scanner plot," which utilizes a simpler well representation comprised of marker and line objects. The size of each of the objects is set to a fixed number of pixels rather than a size in coordinate space, so wells become smaller relative to the coordinate plane when the user zooms in. 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, scanner plots visualize a single constituent with a color and size based on the same ranks used in the detailed 3-D plots.

A combination plot is visually similar to an individual constituent scanner plot except that the well size and color is based on the rank of its combination value. A well's combination value is the sum of its selected constituent's ranks. The combination plot thus highlights where all values for a set of selected species are high. If a well does not have values for all of the selected species, it is not plotted in the combination plot.

5 Results

The ENVision system was utilized by Drexel University faculty and students to explore and evaluate a number of water quality datasets; thus demonstrating its usefulness and effectiveness.

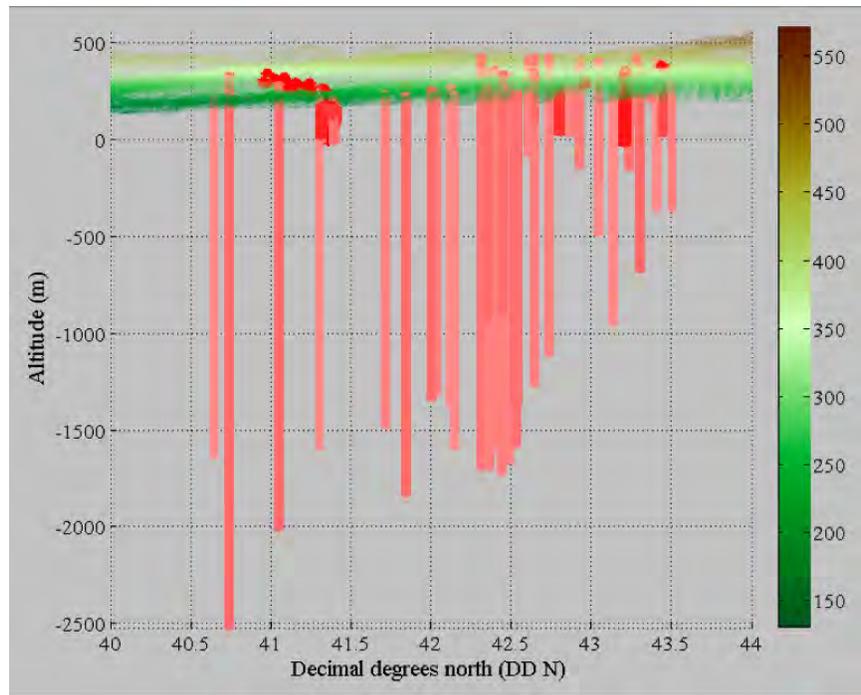


Figure 6. Zoomed-in, rotated view of Iowa nitrate hot spots displays a relationship between well depth and nitrate concentration. This plot shows that shallow wells tend to have the highest concentrations of nitrates.

5.1 Iowa

Figure 5 presents a scanner plot of nitrate concentrations in Iowa and highlights spatial hot spots (dashed cyan circles). Each line represents a well. The length of the line indicates the well's depth. The color and thickness of the well correspond to the well's relative concentration of a constituent – in this example, nitrate. Darker shades and thicker wells have higher concentrations than lighter shades and thinner wells. The colorbar on the right shows how color relates to terrain elevation (m). The figure shows that high nitrate concentrations are mostly clustered in two areas at lower elevations in the southeastern portion of the state. As spatial relationships are difficult to decipher from the database, visualization highlights relationships that may otherwise go unnoticed.

Figure 6 presents a zoomed-in and rotated view of the Iowa nitrate hot spots and shows that nitrate concentrations in Iowa are only high in wells shallower than approximately 50 m below sea level. While this relationship might be seen in a scatterplot of well-depth vs. nitrate concentration, visualization not only demonstrates the correlation between depth and concentration, but also shows the spatial clustering of the high nitrate wells.

5.2 New Jersey

Figure 7 shows a scanner plot created with ENVision. This plot presents arsenic concentration for all New Jersey wells of water use code 'H' (domestic wells) and 'P' (public

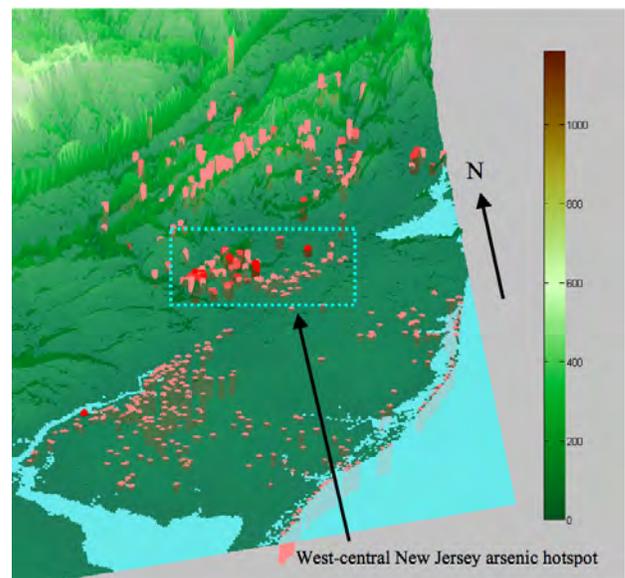
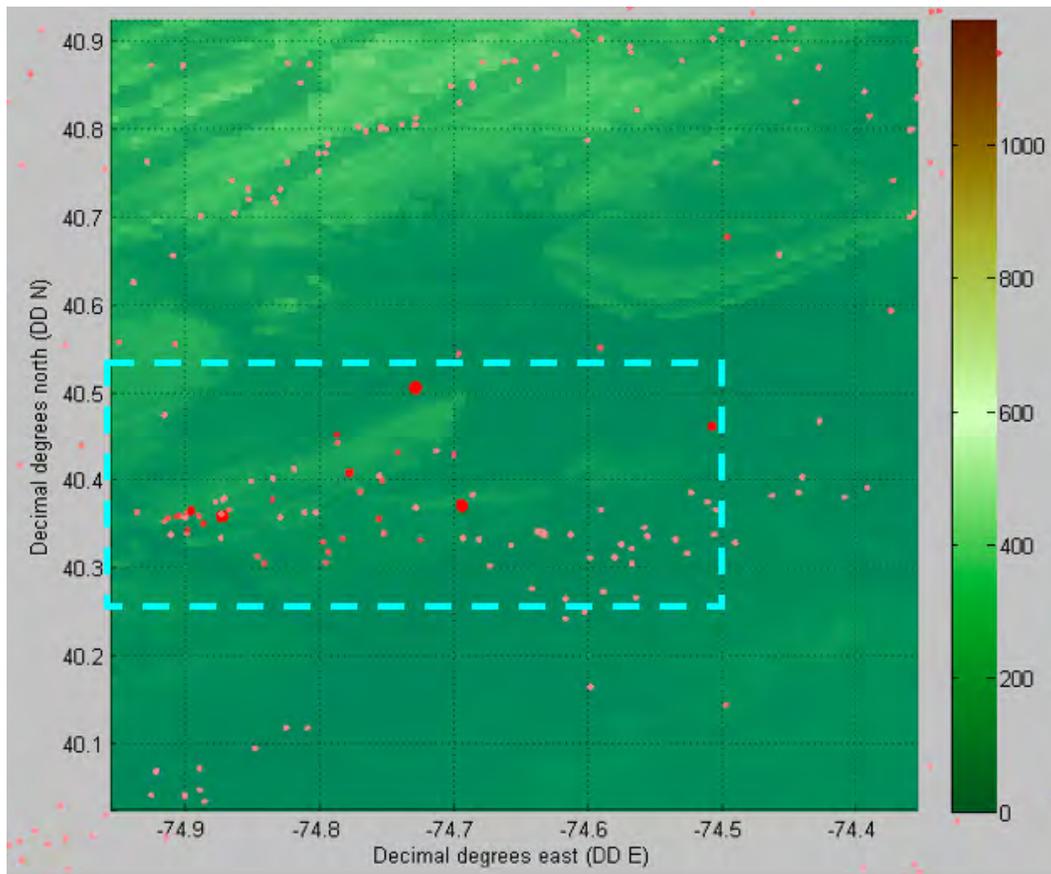
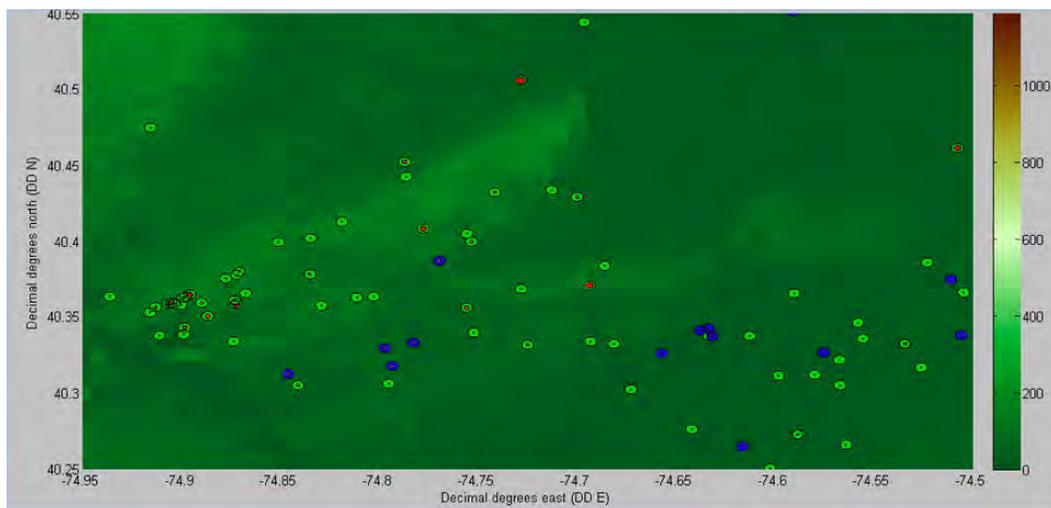


Figure 7. Scanner plot of arsenic in New Jersey. Cyan dashed box (added by authors) highlights a hot spot near the transition to piedmont in west-central New Jersey.



(a)



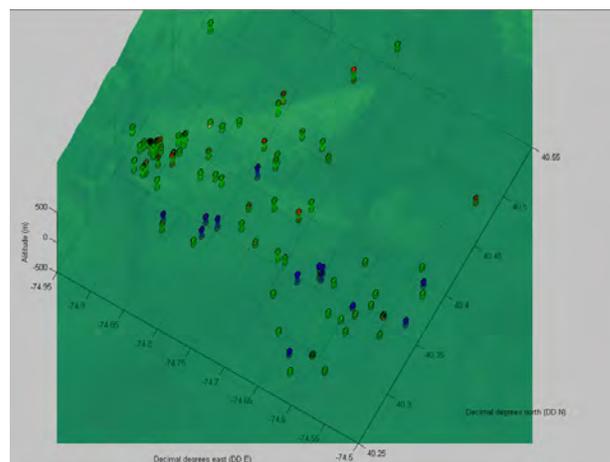
(b)

Figure 8. (a) Plan view close-up scanner plot of arsenic hot spot (dashed cyan rectangle). (b) Plan view close-up detailed plot of arsenic hot spot. High arsenic, signified by larger, brighter red circles, appears in green wells, but no blue wells.

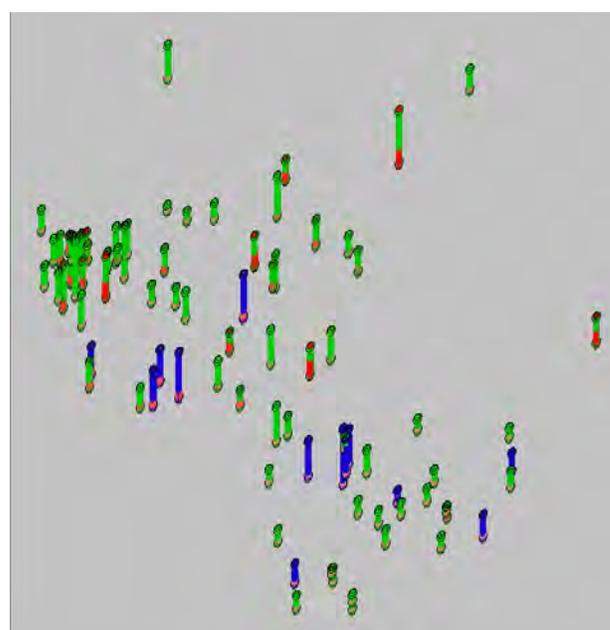
wells). The scanner plot reveals a west-central New Jersey arsenic hot spot. Identifying such a spatial trend from the arsenic point database is difficult, but visualization makes this spatial trend obvious. Since this visualization also visually communicates elevation data, it is possible to see that with the exception of one well in southwest New Jersey, arsenic concentrations in the coastal plane are very low. Figure 7 shows that higher arsenic concentrations are not found until the coastal plain begins to transition into the piedmont on the east 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 [11]. As seen in Figures 7, 8 and 9 the high arsenic wells exist in a line roughly perpendicular to the elevation gradient, suggesting that they may tap the same geological stratum.

Figures 8, 9 and 10 show different views of the arsenic hot spot, lying roughly within 40.25 to 40.55 DD N and -74.5 to -74.95 DD E. Each of these views help to identify that within the west-central New Jersey hot spot, high arsenic concentrations are found almost exclusively in private, but not public, wells. As in the legend shown in Figure 4, green wells represent wells of water use code 'H,' meaning that they are domestic wells. Blue wells are of water use code 'P' (public wells), and these wells only contain low to moderate levels of arsenic. This trend is again one that would be difficult to extract from the spreadsheet, but the ENVision plots makes the trend easy to recognize.

Because it is not always easy to identify a trend from one view, we include Figures 8 through 10 to show the New Jersey arsenic hot spot from several viewpoints, in part as a demonstration of ENVision's capabilities, but also as a walkthrough of a sample analysis using ENVision. The standard view is the two-dimensional plan view (Figure 8). It is possible, though difficult, to see that the trend of large, bright red, circles appears in green wells and not in blue wells. To see this trend more clearly, Figure 9 shows a rotation of Figure 8, taking advantage of ENVision's 3-dimensional rotation capability, which is facilitated by ENVision's GUI. Figure 9(a) is in some ways an improvement on Figure 8(b) because the user can read the well's information from the top, as in Figure 8, and now also on the side. The advantage of this information redundancy is clearer in Figure 9(b), which is the same view as 9(a), but the terrain and elevation data have been removed in order to more clearly view the wells. Making plots without the terrain is an option in ENVision, simply by pressing the "Forego Maps" button on the GUI before plotting. In Figure 9(b), the user can begin to clearly see that the public wells only show low to moderate concentrations of arsenic; all the high arsenic readings are found in private wells. Using ENVision to rotate this plot even more, to the horizontal view in Figure 10, the user is given another visual confirmation that high arsenic concentrations are found in private wells. This horizontal view also would give the reader the advantage of being able to recognize a trend related to well



(a)



(b)

Figure 9. Detailed 3-D plot close-up of New Jersey arsenic hot spot. (a) Plotted with terrain. (b) The same plot, but without the terrain and elevation data in order to see the wells more clearly. Both (a) and (b) show that high arsenic concentrations are mostly found in green (private) wells.

elevation or well depth, if any existed.

6 Conclusions

ENVision provides more flexibility and is capable of displaying more information than previous water quality visualization software. Instead of only visualizing specific constituents in a specific region, this program is general enough to visualize any variety of constituents in any land-mass. Even though the software is very general, through the GUI, the user can create visualizations specific to

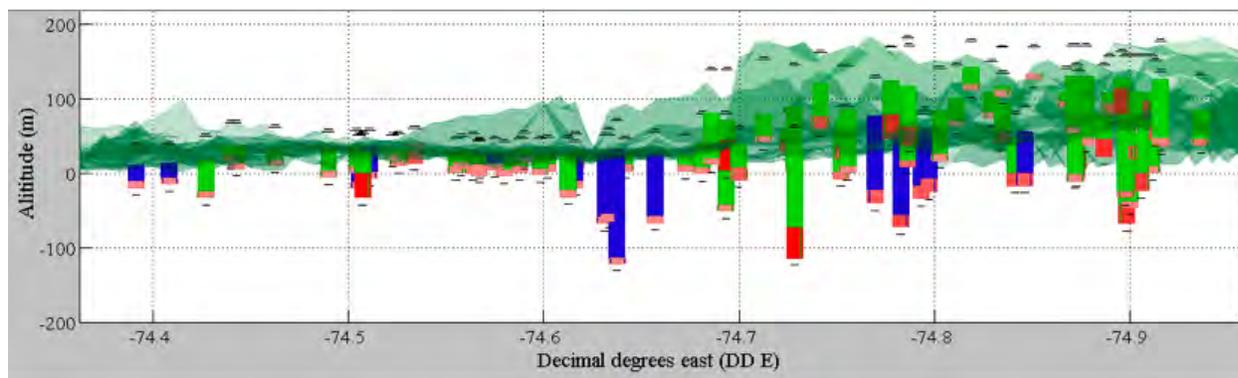


Figure 10. Rotated and focused, detailed 3-D plot of the arsenic hot spot identified in Figure 7 in west-central New Jersey. Detailed 3-D plots show each well's location, depth, water usage type, and constituent concentration, which in this plot is arsenic only. Cylindrical bands' color and height display relative arsenic concentrations. This specific plot shows that high arsenic is found mostly in private wells (green wells) rather than public wells (blue wells).

his/her needs.

Although this visualization tool was specifically developed for use with the USGS arsenic point database, it can support a wide variety of databases. With only minimal adjustments to the software or database, ENVision can visualize other environmental factors such as soil quality or air quality or even social science values. Just as it possible to visualize arsenic and pH values in New Jersey's domestic and public wells, it is also possible to visualize water quality and median neighborhood income in India, as long as the appropriate data exists. Such visualization may help to reveal environmental justice issues.

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.

References

- [1] Environmental Protection Agency. Ground water & drinking water: Frequently asked questions, September 2007. <http://www.epa.gov/safewater/faq/faq.html>.
- [2] P.J. Squillace, M.J. Moran, W.W. Lapham, C.V. Price, R.M. Clawges, and J.S. Zogorski. Volatile organic compounds in untreated ambient groundwater of the United States, 1985-1995. *Environmental Science and Technology*, 33(23):4176-4187, 1999.
- [3] J.R. Lockwood, M.J. Schervish, P.L. Gurian, and M.J. Small. Analysis of contaminant co-occurrence in community water systems. *Journal of the American Statistical Association*, 99(465):45-56, 2004.
- [4] M.J. Focazio, D. Tipton, S.D. Shapiro, and L.H. Geiger. The chemical quality of self-supplied domes-

tic well water in the United States. *Ground Water Monitoring & Remediation*, 26(3):92-104, 2006.

- [5] G. Serres et al. Molecular confirmation of Hepatitis A virus from well water: Epidemiology and public health implications. *The Journal of Infectious Diseases*, 179(1):37-43, 1999.
- [6] J.R. Lockwood, M.J. Schervish, P.L. Gurian, and M.J. Small. Characterization of arsenic occurrence in U.S. drinking water treatment facility source waters. *Journal of the American Statistical Association*, 96(456):1184-1193, 2001.
- [7] M.J. Focazio, A.H. Welch, S.A. Watkins, D.R. Helsel, and M.A. Horn. *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, 2000.
- [8] A.B. Forgang, B. Hamann, and C.F. Cerco. Visualization of water quality data for the Chesapeake Bay. In *Proc. IEEE Visualization*, pages 417-420, 1996.
- [9] R.J. Stein, A.M. Shih, M.P. Baker, C.F. Cerco, and M.R. Noel. Scientific visualization of water quality in the Chesapeake Bay. In *Proc. IEEE Visualization*, pages 509-512, 2000.
- [10] J. Sreevalsan-Nair, E. Van Nieuwenhuysse, I. Hotz, L. Linsen, and B. Hamann. An interactive visual exploration tool for northern California's water monitoring network. In *Proc. SPIE Conference on Visualization and Data Analysis*, pages 1-12, 2007.
- [11] P.L. Smedley and D.G. Kinniburgh. A review of the source, behavior and distribution of arsenic in natural waters. *Applied Geochemistry*, 17:517-568, 2002.