

GEOSPATIAL ANALYSIS APPROACHES FOR DRINKING WATER SOURCE PROTECTION AREAS

Jamie Cajka, RTI International

William Cooter, RTI International

Jay Rineer, RTI International

James Sinnott, RTI International

Roger Anzzolin, U.S. Environmental Protection Agency, Office of Water, Office of Ground Water and Drinking Water

Abstract

A variety of geographic information systems (GIS) layers for drinking water source areas are now available to assist in identifying potential contaminant risks for surface and groundwater sources used by public drinking water systems. Because these areas are typically defined for the specific intakes or wellheads for each system, the complete set of source area polygons for entire states, or the entire country, will involve complex patterns, with polygons being overlapped or embedded within other polygons. This paper presents examples on approaches for using these polygons to process information using other GIS data layers or raster-based information, such as the National Land Cover Dataset (NLCD). ESRI tools, such as Spatial Analyst, Model Builder, and Zonal Statistics commands, were applied using special Python-based ArcGIS scripts. The outputs from these geospatial analyses can be incorporated into ArcIMS-based tools, such as the Office of Ground Water and Drinking Water's (OGWDW's) Drinking Water Mapping Application (DWMA).

Introduction

The U.S. Environmental Protection Agency's (EPA's) Office of Ground Water and Drinking Water (OGWDW) Drinking Water Mapping Application (DWMA) is a Web-based geospatial application that enables queries of the Safe Drinking Water Information System (SDWIS/FED), as well as the identification of potential contaminant risks to surface water and groundwater used for public drinking water supplies. The DWMA provides a secure application that EPA staff can easily use to obtain reports and maps that help manage programs under the Safe Drinking Water Act (SDWA). The core data sources, Web interface, functionality, and system architecture of the DWMA are described in an earlier paper entitled *EPA Drinking Water Mapping Application: Client Focus and Architecture Design*, which was presented at the 23rd Annual ESRI International User Conference and is available online at <http://gis.esri.com/library/userconf/proc03/abstracts/a0149.pdf>.

This paper will provide a case study of the creation of the impervious surface data layer being used by the DWMA. The process description will highlight the use of current ArcGIS tools and the flexibility of this technique in creating future data layers.

2.0 Creation of Source Protection Area Polygons

SWAs are the land areas that contribute water and pollutants to the water supply for surface water intakes or groundwater wells that supply drinking water (paraphrased from <http://www.epa.gov/safewater/protect/assessment.html>). EPA is in the process of collecting state-supplied SWAs for public water supplies, but these SWAs were only available for a small number of states during the time this work is conducted. To complete this task, it was necessary to build and use analytical Source Protection Areas (SPAs). When state-supplied SWAs become widely available, they will probably be used in conjunction with the analytically defined SPAs because both of these types of polygons may be of value for different types of analyses.

Because there are more than 8,000 surface water intakes associated with public water systems in the United States, it was necessary to automate the process of building analytical surface SPAs. The first step in building the SPAs was to build upstream drainage networks for each drinking water intake based on medium resolution National Hydrography Dataset (NHD) (<http://nhd.usgs.gov>). To build the upstream drainage network, an upstream distance had to be chosen. A value of fifteen miles upstream was selected as a robust approximation of the distance water would travel in a single day. This distance is used by some states as part of their analysis methodologies for source water assessments (Tennessee Department of Environment and Conservation, 2003). Similar robust estimates can be found in older EPA guidance for watershed modeling (U.S. EPA, 1982), as well as in more recent landscape ecology studies (Heggem, 2004). A recursive upstream search algorithm was encoded in a procedure to collect the upstream drains (i.e., a small stream segment) within 15 miles upstream. This procedure was exercised for each of the 8,000+ surface water intakes. Figure 1 shows the derived upstream network (shown in red) for a single drinking water intake (shown as a blue triangle).

A number of “value added attributes” are currently being computed for each drain within the NHD. One of these attributes is the average stream flow, slopes, and velocities. When these data are available, they could be used to improve the rough approximation of “15 miles upstream equals 1 day of travel.” The distance upstream from each drinking water intake could be tailored to the individual velocities associated with each drain upstream of the intake location. A national navigation database-level service is also in progress and is dependent on the completion of the value-added attributes. Once this navigation service is available, the DWMA plans to make use of its improved algorithm via this shared service environment.

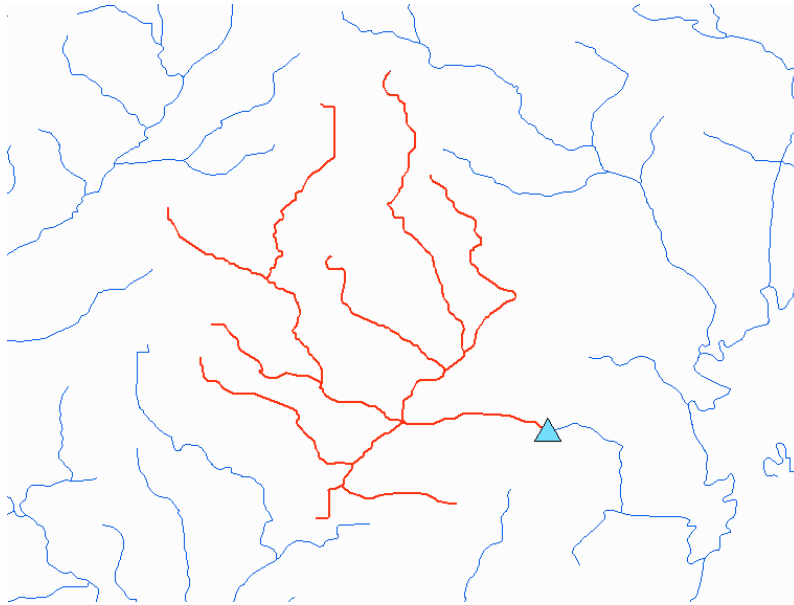


Figure 1. Example Drinking Water Intake Upstream Network

The next step for building surface SPAs was to create the watershed for the upstream drainage network. There is also work in progress to define a national set of “drain catchments” based on methods initially implemented by the U.S. Geological Survey (USGS) for several New England states; this method is often referred to as the New England method (Moore et al., 2004). Once these data are available, drain catchments for the upstream network of drains for a particular drinking water intake could be merged to create the watershed for this intake. Because these drain catchments were not available at the time this task was being performed, analytical watersheds were built using a “convex hull” approach. A convex hull is defined by the set of points that is the smallest convex set that includes all of the drains that comprise the upstream network. Figure 2 shows an analytical surface SWA polygon (shown in green) defined by the convex hull that surrounds the derived upstream network (shown in red).

Although this technique worked well for the majority of surface drinking water intakes, it did not work well for two categories of intakes. First, the concept of a simple upstream drainage network for intakes situated offshore (e.g., within a Great Lake) does not exist. For these intakes, simple buffers with a radius of 1 mile were created. These simple buffers could be relevant to protection from contaminant spills. Second, if a drinking water intake draws water from a perfectly straight drain with no upstream drains, a convex hull could not be created (i.e., it would have no area). For these intakes, a buffer was created using the single, perfectly straight drain as the diameter of the buffer. The introduction of the drain catchments described earlier will greatly improve the quality of the SPAs for these types of intakes.

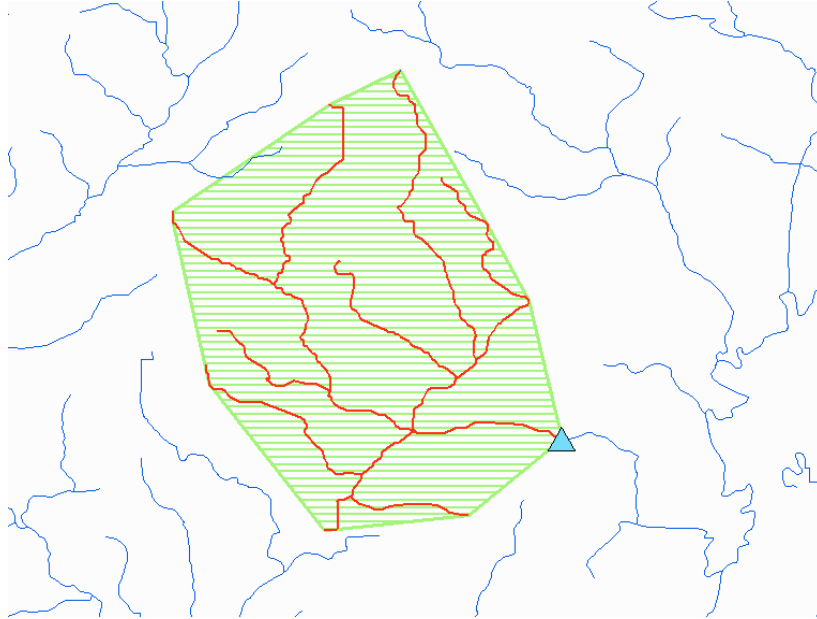


Figure 2. Drinking Water Intake Analytical Watershed

In addition to the 8,000+ surface water intakes associated with U.S. public water systems, there are almost 200,000 wells. Analytical wellhead SWAs were created for each of these wells by generating simple buffers with a radius of 1 mile around each wellhead.

3.0 National Land Cover Dataset Selection

The National Land Cover Dataset (NLCD) is a nationwide raster land cover data set distributed by USGS and derived from aerial photography taken between 1988 and 1992 (<http://landcover.usgs.gov/natl/landcover.asp>). This data set is available on a state-by-state basis at a spatial resolution of 30 meters (Figure 3).

The NLCD classifies land cover into 21 broad categories (<http://landcover.usgs.gov/classes.asp>). The NLCD was chosen as the initial data set for the impervious surface layer creation because it is a nationwide, consistent data set with land cover categories that could be mapped to a percent-impervious value. An updated NLCD data set is in progress, but only a few isolated areas of the United States are currently available. When this data set is complete, there will be nationwide land cover, impervious surface, and forest canopy layers (<http://gisdata.usgs.net/website/MRLC/viewer.php>).

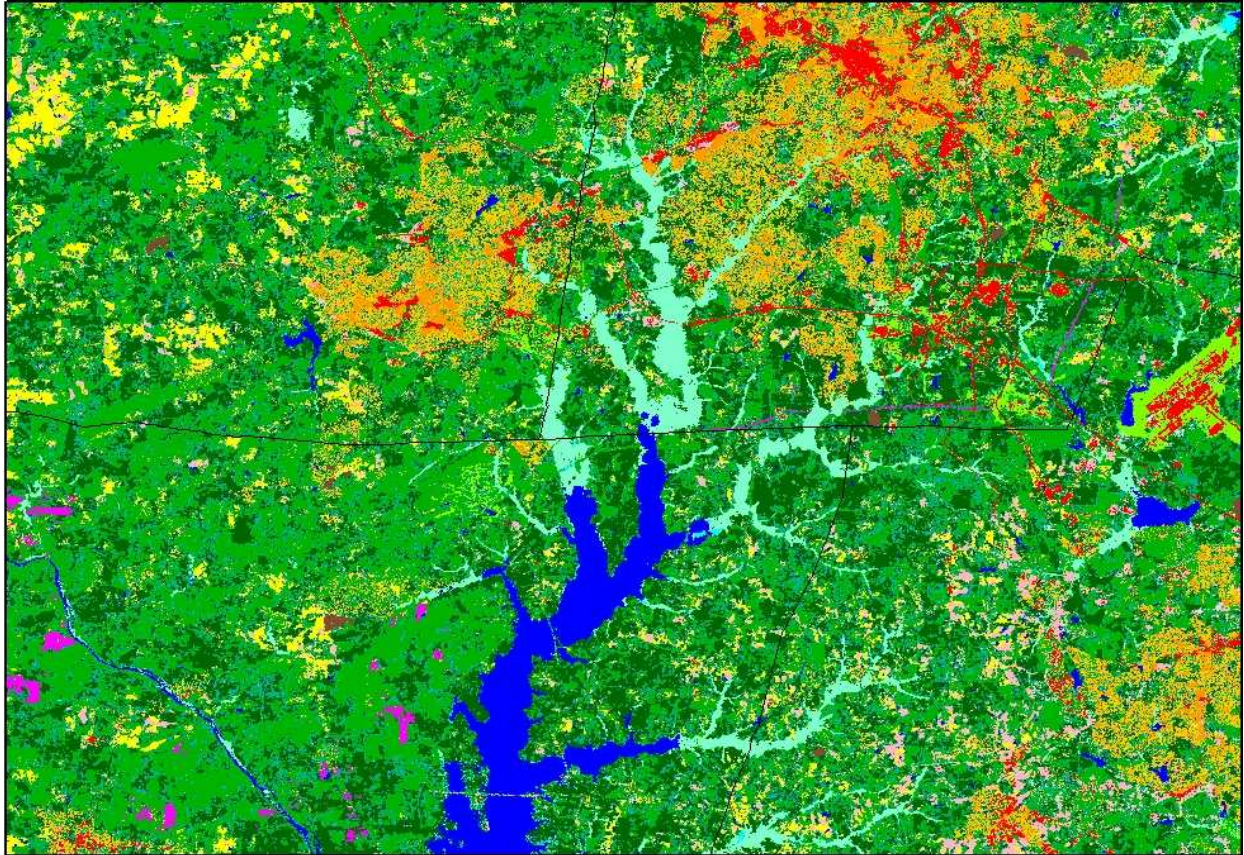


Figure 3. Example of NLCD Imagery

4.0 Spatial Analyst and Model Builder

The goal of this procedure was to generate a cell count, by land cover category, for each of the approximately 200,000 SPA polygons. There is no ArcGIS command that could create summary statistics for each SWA polygon for all cell values within the NLCD for any given state. The closest command is the zonal statistics tool within the Spatial Analyst toolbox found in ArcGIS. This generated a number of statistical measures, including the minimum, maximum, range, and count for the cells of a raster data set whose centers fell within the boundaries of each SPA polygon within an input polygon data layer. It would have been possible to cursor through each SPA polygon, clip the NLCD, and then perform a frequency, but the size of the polygon feature layer (more than 200,000 separate features) precluded this action. The most efficient data manipulation path was to extract each NLCD land cover class from the original state data and to perform the zonal statistics on each of the extracted land cover grids. The count of each land cover class by SPA could then be summarized as a post process in Microsoft (MS) Access.

Model Builder is a graphic user interface (GUI) that allows the user to drag and drop tools from Arc Toolbox and data from Arc Catalog. The user can then connect the data to the tools, specifying input

parameters, intermediate data sets, and final data sets. The model built for this application was relatively simple and straightforward and is illustrated in Figure 4.

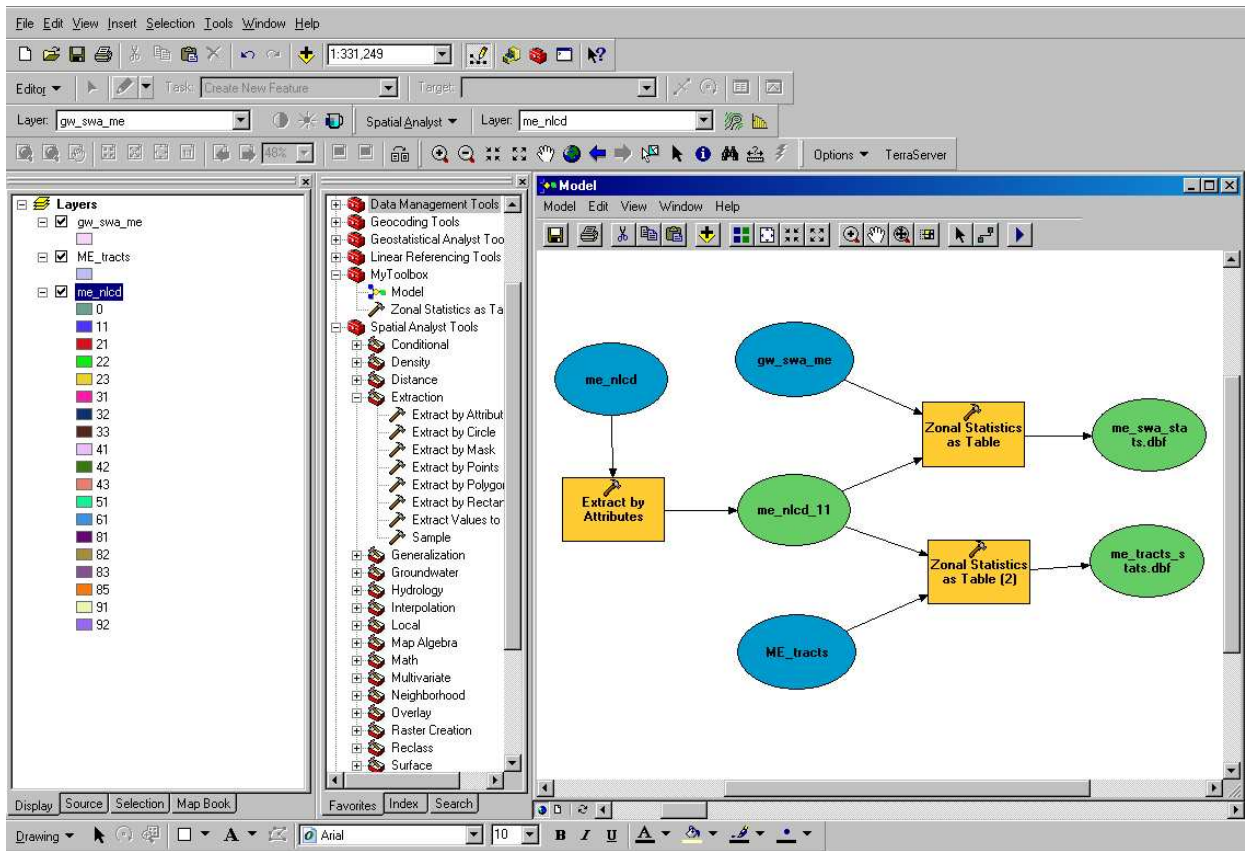


Figure 4. Model Builder Interface and Model

5.0 Export to Python

ESRI’s Model Builder allows models to be exported to Jscript, VBScript, or Python. Python was chosen for its ease of use and the promise of samples from the ESRI technical support Web site. Python is an interpreted, interactive, object-oriented open-source programming language. Its implementation is copyrighted, but it is freely usable and distributable, even for commercial use (www.python.org).

Once the basic zonal statistics tool was created and modified in Model Builder, it was exported to a Python script. Python scripts contain a .py extension and, although these scripts can be edited in any text editor, it is recommended that PythonWin be installed on the user’s PC. PythonWin is an Interactive Development Environment (IDE) and GUI framework that contains a script editor with a debugger and limited intellisense-like tools. The initial model worked for a single land cover code for a single state; therefore, all that needed was to create two looping mechanisms: one for each state, and one for each landcover code within the NLCD. A sample of the Python code used is presented in Figure 5.

```

fullcode_dwma5.py
- for state in ['AZ', 'AR', 'CA', 'CO', 'CT', 'DE', 'DC', 'FL', 'GA', 'ID', 'IL', 'IN', 'IA', 'KS', 'KY', 'LA', 'ME', 'MD', 'M
-   for code in [11, 12, 21, 22, 23, 31, 32, 33, 41, 42, 43, 51, 61, 71, 81, 82, 83, 84, 85, 91, 92]:

    # Local variables...
    st_tract_sum = "//pubnts35/GIS35/DWMA/tracts/" + st + "_tract_" + str(code) + ".dbf"
    st_tract = "//Gisnts01/projects/dwma_08673.004/data/" + st + "_tract.shp"
    st_swa_sum = "//pubnts35/GIS35/DWMA/swa/" + st + "_swa_" + str(code) + ".dbf"
    st_code = "//pubnts35/GIS35/NLCD/" + st + str(code)
    Delete_succeeded = "true"
    st_nlcd = "//pubnts35/GIS35/NLCD/" + st + "_nlcd"
    tracts = "//Gis450/Data1/base/us_covers/tract_us_stname.shp"
    st_swa = "//Gisnts01/projects/dwma_08673.004/data/gw_swa_" + st + ".shp"

    # Process: Extract by Attributes...
    gp.ExtractByAttributes_sa(st_nlcd, "Value = " + str(code), st_code)

    # Process: Select...
    gp.Select_analysis(tracts, st_tract, "STATE = '" + st + "'")
    tempEnvironment0 = gp.extent
    gp.extent = st_nlcd

    # Process: Zonal Statistics as Table...
    gp.ZonalStatisticsAsTable_sa(st_tract, "TRACT_FULL", st_code, st_tract_sum, "DATA")

    # Process: Zonal Statistics as Table (2)...
    gp.ZonalStatisticsAsTable_sa(st_swa, "ENTITY_ID", st_code, st_swa_sum, "DATA")
    gp.extent = tempEnvironment0

    # Process: Delete...
    gp.Delete_management(st_code, st_tract_sum, )
    # Process: Delete...
    gp.Delete_management(st_tract, st_tract_sum, )

```

Figure 5. Sample of Python Code

The Python code could have potentially produced a total of 1122 data files (51 states, including DC X 22 land cover categories). In reality, the number of files was less because not all land cover categories existed in all states.

6.0 Post Processing Using Access

The output files were dbase IV files by land cover code and state, containing the SPA entity_id (i.e., the unique identifier for each SPA), the count, and a variety of other statistical measures. In order to create an impervious surface data layer, the total number of cells (the count) for each land cover type needed to be known for each SPA entity_id. Because SPAs could cross state lines, MS Access queries were created and run against all state files to sum the count of each land cover code by entity_id to produce the final output dbase file. Since the spatial resolution of the NLCD is 30 meters, each cell was 900 square meters in area; therefore, if a given entity_id had a count of 10 for a given land cover code, that land cover code comprised 9000 square meters within the entity_id.

7.0 Creation of Impervious Surface Data Layer

For each SPA polygon, the original raster information was converted into a more compact, lumped-parameter format to facilitate determining the distribution patterns for the 21 potential different

land use cover classes defined in the NLCD. Several types of indicator scores for entire SPA polygons can be developed from lumped-parameter data. Possibilities include simple percentages of certain land cover types (e.g., the percent in urbanized land cover) or indicators where weighting factors are applied to the land cover information. Weighted indicators can be developed to provide estimates of annual loadings for such nutrients as total phosphorous or total nitrogen. Another common weighted indicator is an estimate of the percent of impervious land cover over an entire SPA. Default impervious weighting factors from EPA's Analytical Tools Interface for Landscape Assessments (ATtILA) tool (information available at the following Internet address: <http://www.pagis.org/CuttingATtILA.htm>) were applied to develop robust percent-impervious cover indicators for all the SPA polygons. Percent-impervious cover indicators for small landscape units or watershed areas help identify situations where the introduction of streets and buildings and the replacement of natural forest or grasslands vegetation will seriously alter natural hydrological patterns of groundwater infiltration or surface runoff following wet weather events (CWP, 2003; MWCOG, 1992). Impervious indicator values greater than approximately 10 percent are generally interpreted to suggest noticeable alterations to natural hydrologic cycles, whereas indicator values greater than 20 percent usually suggest high degrees of paved or similar impervious surfaces in a landscape unit.

8.0 Integration of Impervious Surface Data Layer in DWMA

Impervious surface percentages were calculated for each SPA polygon. These values, in conjunction with other calculated and dynamically generated attribute values, are used in the DWMA to provide drinking water intake query functionality at the national, state, and local levels. For a discussion of the DWMA query architecture, refer to the aforementioned paper, *EPA Drinking Water Application: Client Focus and Architecture Design*. The percent-impervious query filter is grouped together with other query options, including the following:

- SDWIS system type (Community, Non- Community)
- SDWIS source type (Intakes, Wells, Reservoirs, Streams)
- SDWIS-reported drinking water violation
- Potential point sources (PCS, CSO, RCRA, CERCLA, UIC)
- Potential non-point sources (Nitrogen, Atrazine, Manure)
- Population in SWA (Census derived)
- Median household income in county served (Census derived)
- Intersections with
 - Karst
 - U.S. Department of Defense (DOD) lands
 - Sole Source Aquifers.

These query options and the DWMA framework provide a powerful means to investigate issues within SPAs. Figure 6 shows the results of a sample query displaying the states where at least one SWA has a percent-impervious cover greater than 20 percent and a population of greater than 100,000 living in or near the SPA. The DWMA provides tools to drill down and investigate the individual drinking water sources and corresponding SWAs.

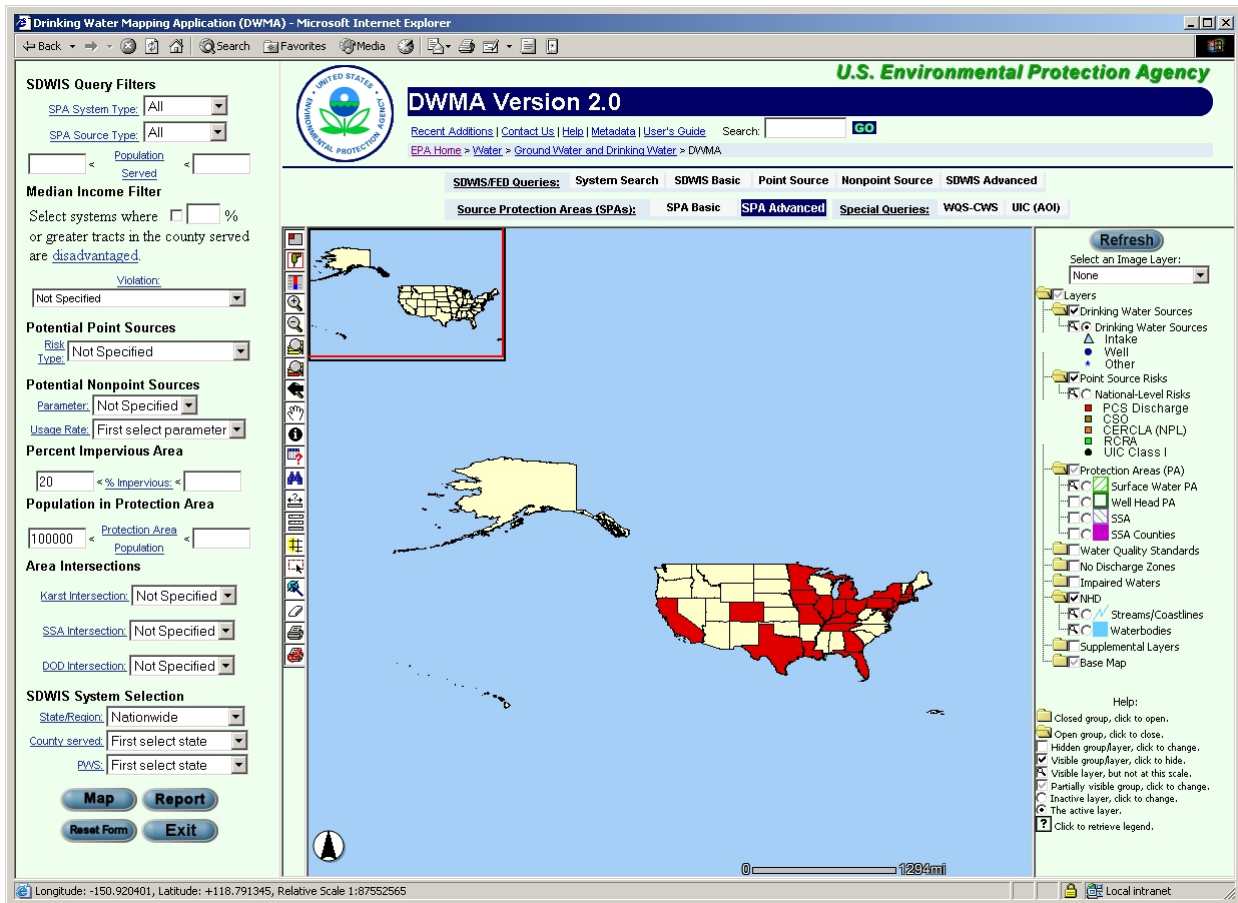


Figure 6. Results of DWMA Query Using the Impervious Surface Layer

9.0 Conclusions

Ongoing refinements to the DWMA have focused on using sets of analytical SPAs as a convenient way to process a broad range of vector and raster-based data to identify potential contaminant risks in the vicinities of surface intakes or wellheads used by public drinking water systems. Particularly in the case of surface intakes, the application of a convex hull technique to create polygons associated with networks of NHD drains provides a robust upstream analysis framework to simplify the identification of potential risks and the presentation of the information using ESRI ArcIMS-based GIS displays.

References

- CWP (Center for Watershed Protection). 2003. *Impacts of Impervious Cover on Aquatic Systems*. CWP, Ellicott City, MD.
- Heggen, D. 2004. Development of landscape indicators for potential nutrient impairment of streams in EPA Region 8. *EMAP Symposium 2004*. May 3-6, Newport, RI. Available at <http://www.epa.gov/emap/html/pubs/docs/groupdocs/symposia/symp2004/index.html>.
- MWCOG (Metropolitan Washington Council of Governments). 1992. *Watershed Restoration Sourcebook*. Department of Environmental Programs, MWCOG, Washington, DC.
- Moore, R.B., Johnston, C. M., Robinson, K.W. and Deacon, J.R. 2004. Estimation of Total Nitrogen and Phosphorus in New England Streams Using Spatially Referenced Regression Models. *U.S. Geological Survey Scientific Investigations Report 2004-5012* Available at <http://water.usgs.gov/pubs/sir/2004/5012/>.
- Sinnott, J., Anzzolin, R.A., Cooter, W., and Lowry, M. 2004. *EPA Drinking Water Mapping Application: Client Focus and Architecture Design*. 23rd Annual ESRI International User Conference. Available at <http://gis.esri.com/library/userconf/proc03/abstracts/a0149.pdf>.
- Tennessee Department of Environment and Conservation. 2003. *Source Water Assessment Report*. Knoxville, TN: Tennessee Department of Environment and Conservation, Division of Water Supply (available at the following Internet address: <http://www.state.tn.us/environment/dws/dwassess.php>).
- U. S. EPA 1982. *Water Quality Assessment: A Screening Procedure for Toxic and Conventional Pollutants - Part 1*. EPA-600/6-82-004a. Athens, GA: Office of Research and Development, Environmental Research Laboratory.

Author Information

Jamie Cajka
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709
jcajka@rti.org

William Cooter, Ph.D.
RTI International
3040 Cornwallis Road
RTP, NC 27709
sid@rti.org

James (Jay) Rineer
RTI International
3040 Cornwallis Road
RTP, NC 27709
jrin@rti.org

James Sinnott
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709
jsinnott@rti.org

Roger Anzzolin
US EPA Office of Water, Office of Ground Water and Drinking Water
Ariel Rios Building
1200 Pennsylvania Avenue NW (4606M)
Washington, DC 20460
anzzolin.roger@epa.gov