

Use of GIS in Population-Based Human Health Risk Assessment

Jamie Cajka, Joan McLean, Pam Birak, and Zach Pekar - RTI International

Abstract

Human health risk estimates often assume that hypothetical human receptors are present and exposed across an entire study area. Using actual population data in a risk assessment enhances the usefulness of the risk estimates by indicating how many people are located in an area where particular exposure concentrations occur. In addition, researchers can augment population data with land use data to draw conclusions about exposure pathways of specific demographic groups, such as rural residents, farmers, and urban gardeners. This type of refined risk analysis allows risk estimation both for specific population subgroups and for the number of individuals in different age cohorts at specific risk levels.

Introduction

Geographic information systems (GIS) have long been used in modeling and assessing human health risk. Individual constituent concentrations, as well as theoretical overall risk values, are often modeled over hypothetical geographic areas. These analyses identify the maximum exposed individual (MEI), which is a hypothetical individual represented by the location and the amount of maximum risk. These analyses fall short, however, because they are based on the assumption that there is population at the point of maximum risk and because they fail to connect levels of risk with the actual human population on the ground. Risk assessors, however, can use a combination of U.S. Census data, land cover data, and GIS to link modeled risk values with specific human receptors in order to enhance the quality of risk estimates. The following sections describe how one sample analysis conducted at an industrial site combined such data and a GIS to generate more refined risk results.

Data

The study used three basic types of data inputs: cancer risk data, land cover data, and U.S. Census data.

Cancer Risk Data

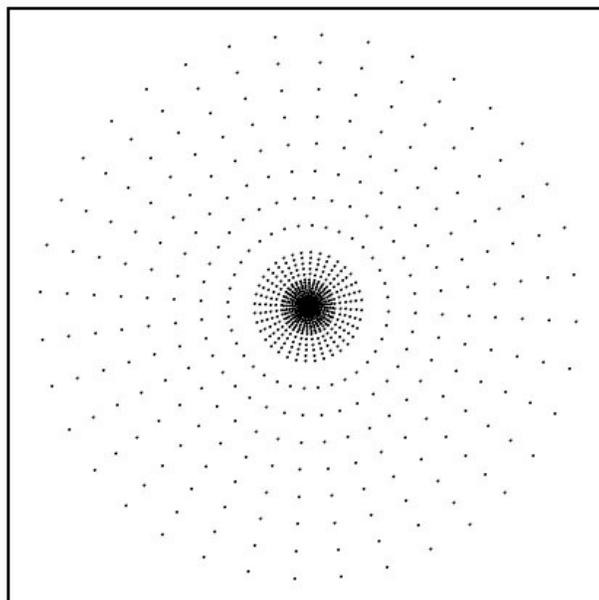
This study modeled 17 separate chemical constituents and their associated cancer risk at a number of industrial facilities. The deposition and concentration amounts were calculated using the U.S. Environmental Protection Agency's (EPA's) Industrial Source Complex Short-Term Model, Version 3 (ISCST3), at 1,152 discrete points extending to a radius of 50 km. The modeled amounts were output in ASCII file format with an x,y location field. Table 1 shows the format of the deposition file for the first few records of one of the constituents.

Table 1 – Sample of ASCII Input file

X	Y	AVERAGE CONCENTRATION	DRY DEPOSITION	WET DEPOSITION
17.36482	98.48078	0.10353	0.03762	0.01427
34.72964	196.9616	0.11381	0.03953	0.01283
52.09445	295.4424	0.12139	0.04016	0.01067
69.45927	393.9231	0.13165	0.04243	0.01013
86.82409	492.4039	0.14413	0.04584	0.0104
104.1889	590.8847	0.16415	0.05091	0.01028
138.9185	787.8461	0.16559	0.04994	0.00876
173.6482	984.8077	0.16297	0.04691	0.00718
208.3778	1181.769	0.14322	0.03971	0.00617
243.1075	1378.731	0.13402	0.03468	0.00548
277.8371	1575.692	0.12797	0.031	0.00496
312.5667	1772.654	0.12482	0.02853	0.00457
347.2964	1969.615	0.13201	0.02775	0.00423
434.1204	2462.019	0.13901	0.02405	0.00349
520.9445	2954.423	0.1195	0.019	0.00297

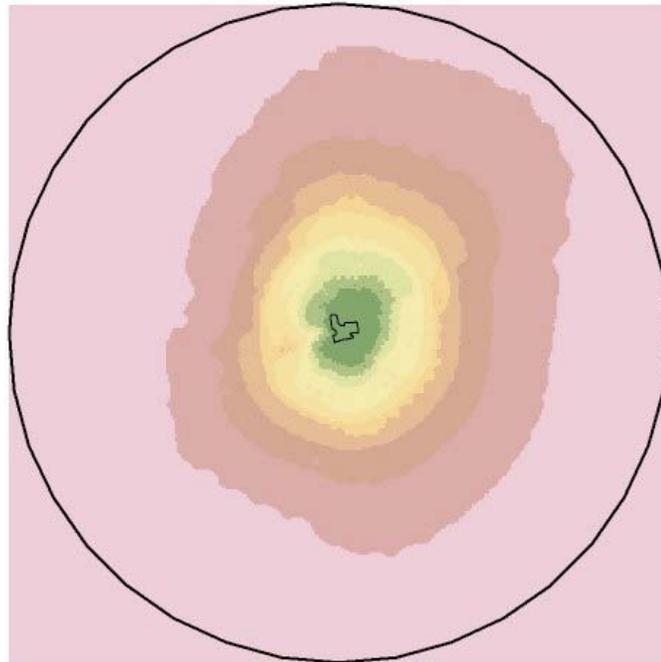
Because each point had an x,y location relative to the origin, a point coverage could be created showing the spatial distribution of cancer risk, as shown in Figure 1.

Figure 1 – Distribution of Modeled Points out to 50 km



By centering the origin of the radial point coverage over the real-world x,y coordinates of the industrial site, the distribution of cancer risk could then be georeferenced. The points were converted to a TIN and then into a GRID with either the constituent's deposition or concentration in the value field. Figure 2 illustrates one of the constituent grids.

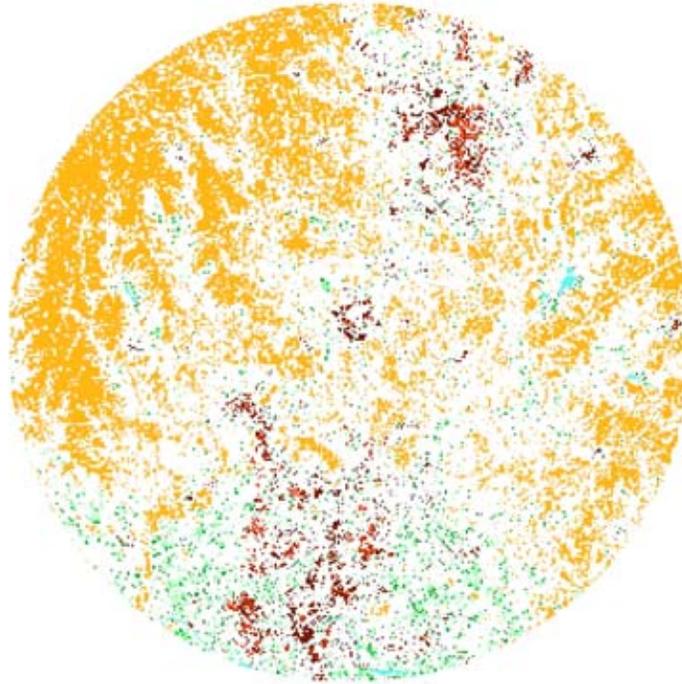
Figure 2 – Rasterized Radial Point Coverage with 50 km Radius



Land Cover Data

The land cover data were taken from the National Land Cover Dataset (NLCD), which stores land cover data in raster format. These data were extracted from imagery obtained between 1988 and 1992. Land cover data were downloaded from the NLCD Web site (<http://landcover.usgs.gov/natl/landcover.htm>) and converted into 30 m resolution grids in the Albers meters NAD83 projection. They were available state by state, so states were appended and clipped to a 50 km radius as needed. An example is shown in Figure 3.

Figure 3 – NLCD Data Clipped to 50 km



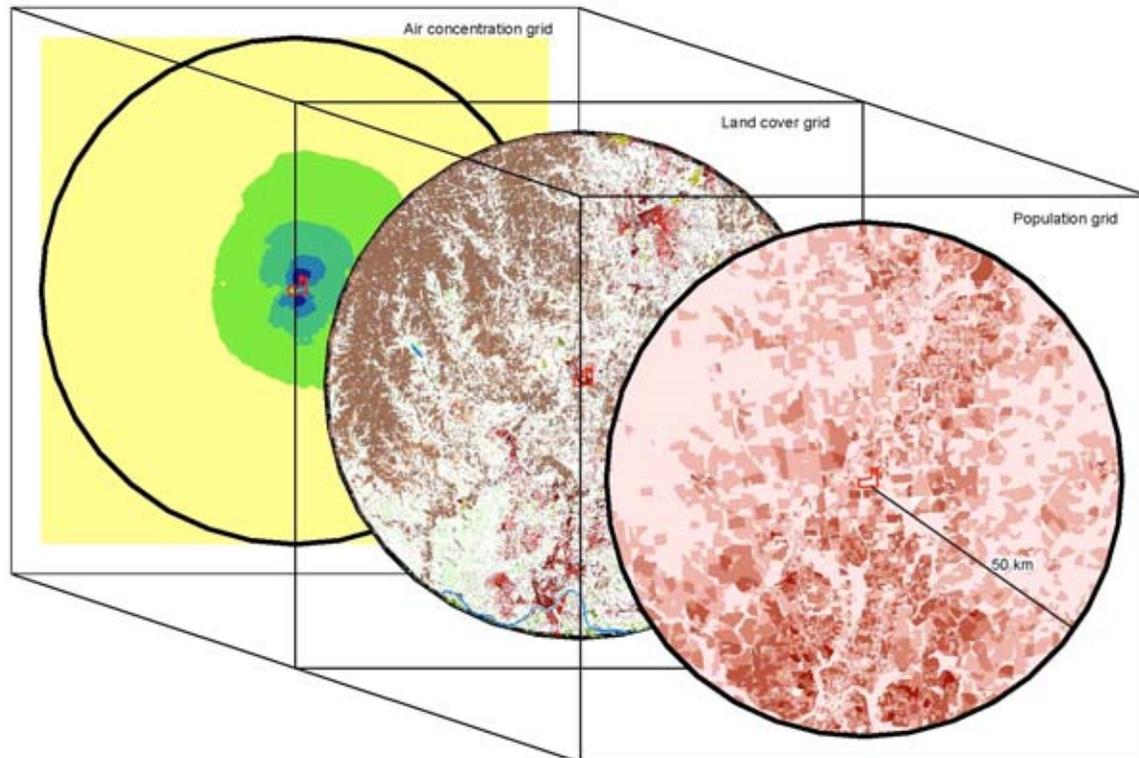
NLCD data come pre-classified into 21 land cover categories, which are listed in Appendix A.

Population Data

The Census data were obtained from the 1990 U.S. Census, which stores Census geography in polygons with links to associated population tables. The 1990 Census data were used because these were the most recent data available at the time of the study. Block group data were downloaded as shapefiles, converted to coverages, projected to Albers NAD83 meters, and then rasterized to GRID format using a 30 m grid cell size. The associated Summary Tape File 3 (STF3) data files were also downloaded, converted to INFO files, and joined to the grids.

Because the NLCD data were already in raster format and the analysis involved large amounts of data, all data were converted to raster format and the analysis was performed using ArcInfo's GRID. These three layers could then be overlaid to produce specific risk/population/land cover combinations, as shown in Figure 4.

Figure 4 – Overlay of Data Input Grids



Procedures

The main goal of this study was to assign risk to actual human populations. Calculating maximum, minimum, range and mean risk values over a Census block group is fine, but this method assumes that the population is spread evenly throughout the block group. If land cover type is known (and land use can therefore be deduced), then population subgroups can be placed within a block group to further refine the risk estimates.

There were five specific population subgroups of interest in this study: rural resident, urban resident, farmer, adult receptor, and MEI. These subgroups have different levels of exposure and are exposed through different numbers and types of pathways; for instance, urban residents may be exposed only through ambient air concentrations, whereas farmers may be exposed through ambient air concentrations, consumption of contaminated produce from a home garden, ingestion of contaminated fish from a farm pond, as well as other routes. Placing the population subgroups allows risk assessors to conduct a multi-pathway, multimedia risk assessment that accounts for the varying exposure levels within a population.

This study included a three-tiered analysis: tier 1 consisted of a detailed deterministic risk assessment to generate risk scenarios for a *range* of input parameters, while tier 2

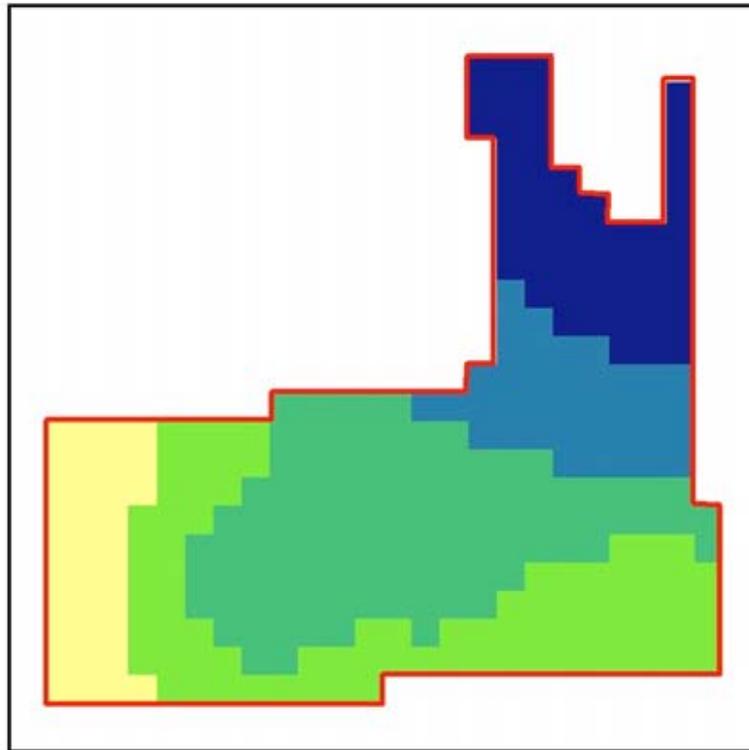
and 3 consisted of a probabilistic risk analysis that generated a risk scenario for *specific* input parameter values. The first three subgroups were required to conduct the tier 1 analysis, and the last two were required to conduct the tier 2 and 3 analyses. In order to place these receptors, the NLCD data had to be reclassified into specific land cover types. Table 2 lists the subgroups, their Census fields, and the required land cover codes that were needed to create the population subgroups.

Table 2 – Population Subgroup Parameters and Definitions

Receptor Population	U.S. Census Data Attributes	Land Cover Category	NLCD Land Use Data
Rural resident (Tier 1)	P0060004 "persons in rural area not on farm"	Rural	41 (deciduous forest); 42 (evergreen forest); 43 (mixed forest); 51 (shrubland); and 71 (grasslands/herbaceous).
Urban resident (Tier 1)	P0060001 "persons inside urbanized areas"; P0060002 "persons outside urbanized areas"	Residential	21 (low-intensity residential) and 22 (high-intensity residential).
Farmer (Tier 1)	P0060003 "persons in rural area on farm"	Cropland	Beef cattle/dairy cattle grazing: 71 (grasslands/herbaceous); 81 (pasture/hay) Vegetable/fruit (crop) cultivation: 61 (orchards/vineyards/other); 82 (row crops); 83 (small grains); 84 (fallow).
Adult receptor (Tier 2)	POP18_65	Habitable	41 (deciduous forest); 42 (evergreen forest); 43 (mixed forest); 51 (shrubland); 71 (grasslands/herbaceous); 21 (low-intensity residential) and 22 (high-intensity residential).
MEI analysis (Tier 3)	NA (population count data not used in this assessment)	Potential Future Habitable	"Potential future habitable" land use areas: commercial/industrial (23).

Each NLCD grid was reclassified using the reclass function in ArcInfo GRID to produce the desired output grid; one grid was produced for each land cover type of interest. Each land cover grid was then overlaid with the Census block group grid to produce a grid of block groups (whole and partial) for each land cover type. An example of this can be seen in Figure 5 where the red line represents the area of a given block group that contains residential land cover, The shading within the boundary represents different levels of a chemical constituents deposition or concentration.

Figure 5 – Residential Land Cover/Census Block Group Combination



These unique land cover/block group combinations were used to define likely areas of receptor population activity within each block group. For example, NLCD land use data identifying high- and low-density residential activity were used to determine those portions of a given block group that were likely to contain residential receptors.

The hybrid NLCD/U.S. Census block group coverages were then combined with the continuous air concentration/deposition coverages to generate air concentration/deposition exposure levels for each receptor population. Specifically, those raster (cell-level) air concentration/deposition values that intersect the portion of a given block group associated with a particular receptor population were used to generate a series of statistics of characterizing air concentration/deposition exposure for that receptor (e.g., mean, median, variance, minimum, maximum from the zonalstats

GRID function). This procedure was then repeated for all blocks/block groups within a given study area to generate a set of air concentration/deposition statistics for each receptor population within specific risk characterization (i.e., air concentration/deposition statistics were generated for each source associated with the facility).

In addition to receptor-population-specific block group level air concentration/deposition statistics, block group level population counts were also needed to generate population risk estimates. The Tier 1 deterministic population risk assessment required block group level population counts for farmers and rural/urban residents, including four age cohorts for each receptor. The Tier 2 probabilistic site-level population risk analysis required a combination of block group level population count data, but only for a more generalized adult receptor. The block group level population counts for modeled receptor population were generated by (1) identifying those block groups that fell within the study areas of interest; (2) area-apportioning population counts for those block groups that intersect the study area boundary between that portion of the block group that falls inside versus outside of the study area; and (3) generating the final block group level population count of interest within the study area by aggregating age cohorts as appropriate.

Conclusions

In this study, the use of GIS in human health risk modeling allowed risk assessors to calculate more refined risk estimates both for the general population and for specific population subgroups within the 50 km radius of the industrial site. This was a greatly improved result over simply identifying the theoretical maximum exposed individual (MEI). The two data sets used to refine the risk (population and land cover) are both readily available at minimal cost. This type of risk analysis is straightforward and easily applied to other types of human health analyses that involve placing human receptors on the ground.

Acknowledgements

The development of these methods was partially funded under EPA contract 68-D-01-073

Contact Information

Appendix A

NLCD Land Cover Classification System Key - Rev. July 20, 1999

Water

- 11 Open Water
- 12 Perennial Ice/Snow

Developed

- 21 Low Intensity Residential
- 22 High Intensity Residential
- 23 Commercial/Industrial/Transportation

Barren

- 31 Bare Rock/Sand/Clay
- 32 Quarries/Strip Mines/Gravel Pits
- 33 Transitional

Forested Upland

- 41 Deciduous Forest
- 42 Evergreen Forest
- 43 Mixed Forest

Shrubland

- 51 Shrubland

Non-natural Woody

- 61 Orchards/Vineyards/Other

Herbaceous Upland

- 71 Grasslands/Herbaceous

Herbaceous Planted/Cultivated

- 81 Pasture/Hay
- 82 Row Crops
- 83 Small Grains
- 84 Fallow
- 85 Urban/Recreational Grasses

Wetlands

- 91 Woody Wetlands
- 92 Emergent Herbaceous Wetlands

NLCD Land Cover Classification System Land Cover Class Definitions

Water - All areas of open water or permanent ice/snow cover.

11. Open Water - All areas of open water; typically 25 percent or greater cover of water (per pixel).

12. Perennial Ice/Snow - All areas characterized by year-long cover of ice and/or snow.

Developed - Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc).

21. Low Intensity Residential - Includes areas with a mixture of constructed materials and vegetation. Constructed materials account for 30-80 percent of the cover. Vegetation may account for 20 to 70 percent of the cover. These areas most commonly include single-family housing units. Population densities will be lower than in high intensity residential areas.

22. High Intensity Residential - Includes highly developed areas where people reside in high numbers. Examples include apartment complexes and row houses. Vegetation accounts for less than 20 percent of the cover. Constructed materials account for 80 to 100 percent of the cover.

23. Commercial/Industrial/Transportation - Includes infrastructure (e.g. roads, railroads, etc.) and all highly developed areas not classified as High Intensity Residential.

Barren - Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen cover may be extensive.

31. Bare Rock/Sand/Clay - Perennially barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, beaches, and other accumulations of earthen material.

32. Quarries/Strip Mines/Gravel Pits - Areas of extractive mining activities with significant surface expression.

33. Transitional - Areas of sparse vegetative cover (less than 25 percent of cover) that are dynamically changing from one land cover to another, often because of land use activities. Examples include forest clearcuts, a transition phase between forest and agricultural land, the temporary clearing of vegetation, and changes due to natural causes (e.g. fire, flood, etc.).

Forested Upland - Areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25-100 percent of the cover.

41. Deciduous Forest - Areas dominated by trees where 75 percent or more of the tree species shed foliage simultaneously in response to seasonal change.

42. Evergreen Forest - Areas dominated by trees where 75 percent or more of the tree species maintain their leaves all year. Canopy is never without green foliage.

43. Mixed Forest - Areas dominated by trees where neither deciduous nor evergreen species represent more than 75 percent of the cover present.

Shrubland - Areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching or interlocking. Both evergreen and deciduous species

of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.

51. Shrubland - Areas dominated by shrubs; shrub canopy accounts for 25-100 percent of the cover. Shrub cover is generally greater than 25 percent when tree cover is less than 25 percent. Shrub cover may be less than 25 percent in cases when the cover of other life forms (e.g. herbaceous or tree) is less than 25 percent and shrubs cover exceeds the cover of the other life forms.

Non-natural Woody - Areas dominated by non-natural woody vegetation; non-natural woody vegetative canopy accounts for 25-100 percent of the cover. The non-natural woody classification is subject to the availability of sufficient ancillary data to differentiate non-natural woody vegetation from natural woody vegetation.

61. Orchards/Vineyards/Other - Orchards, vineyards, and other areas planted or maintained for the production of fruits, nuts, berries, or ornamentals.

Herbaceous Upland - Upland areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75-100 percent of the cover.

71. Grasslands/Herbaceous - Areas dominated by upland grasses and forbs. In rare cases, herbaceous cover is less than 25 percent, but exceeds the combined cover of the woody species present. These areas are not subject to intensive management, but they are often utilized for grazing.

Planted/Cultivated - Areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75-100 percent of the cover.

81. Pasture/Hay - Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.

82. Row Crops - Areas used for the production of crops, such as corn, soybeans, vegetables, tobacco, and cotton.

83. Small Grains - Areas used for the production of graminoid crops such as wheat, barley, oats, and rice.

84. Fallow - Areas used for the production of crops that are temporarily barren or with sparse vegetative cover as a result of being tilled in a management practice that incorporates prescribed alternation between cropping and tillage.

85. Urban/Recreational Grasses - Vegetation (primarily grasses) planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, golf courses, airport grasses, and industrial site grasses.

Wetlands - Areas where the soil or substrate is periodically saturated with or covered with water as defined by Cowardin et al.

91. Woody Wetlands - Areas where forest or shrubland vegetation accounts

for 25-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

92. Emergent Herbaceous Wetlands - Areas where perennial herbaceous vegetation accounts for 75-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.