SAND99-2676
Unlimited Release
Printed November 1999

Automation of GIS-Based Population Data-Collection for Transportation Risk Analysis

G. S. Mills and K. S. Neuhauser
Transportation Safety and Security Analysis Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0718

Abstract

Estimation of the potential radiological risks associated with highway transport of radioactive materials (RAM) requires input data describing population densities adjacent to all portions of the route to be traveled. Previously, aggregated risks for entire multi-state routes were adequately estimated from population data with low geographic resolution. Current demands for geographically-specific risk estimates require similar increases in resolution of population density adjacent to route segments. With the advent of commercial geographic information systems (GISs) and databases describing highways, U.S. Census Blocks, and other information that is geographically distributed, it became feasible to determine and tabulate population characteristics along transportation routes with 1-kilometer resolution. This report describes an automated method of collecting population data adjacent to route segments (for calculation of incident-free doses) based on a commercial GIS. It also describes a statistical method of resolving remaining resolution issues, and an adaptation of the automation method to collection of data on population under a hypothetical plume of contamination resulting from a potential transportation accident.

Introduction

Estimation of the potential radiological risks associated with highway transport of radioactive materials (RAM) by use of the RADTRAN computer code [1] requires input data describing population densities adjacent to all portions of the route to be traveled. Typically, population data have been obtained from the HIGHWAY routing code [2], which provides distance-weighted-average population densities in three categories (Rural, Suburban and Urban) along a route or route segment of interest. Population densities in HIGHWAY were derived from U.S. Census Tract data to describe the population within ½ mile (0.8 km) of the route centerline for incident-free risk analyses. These population densities provide adequate accuracy and detail for analysis of an entire route of typical length (hundreds of kilometers or more). However, "stakeholders" often focus on particular points of interest or short route segments (e.g., through cities, towns or other population concentrations) of critical concern to them. In addition, Executive Order 12898 now imposes a requirement on risk assessments that the "environmental justice or equity" of a proposed action be assessed. The underlying structure of the population database employed in the HIGHWAY code does not permit resolution of small-scale population-density variations. Also, the fundamental purpose of HIGHWAY (routing) requires that route-segment definition be based on highway intersections, not population variations.

A further concern regarding the population model for accident-risk analysis is that the analysis of risks associated with potential transportation accidents which might involve RAM packaging and result in the release of a plume of contamination, as it is implemented in RADTRAN, calls for entry of population densities under the plume "footprint." The same population data (i.e., the population densities in the immediate vicinity of a route) have been used in most analyses. Since a plume may travel as far as 50 miles (80 km) from the point of an accident, the current practice, while conservative, does not directly model the downwind population and has been criticized, therefore, as inadequate.

With the advent of commercial geographic information systems (GISs) and databases describing highways, U.S. Census Blocks (identified as "*block*(*s*)" in this report), and other information that is geographically distributed, it became feasible to determine and tabulate population characteristics along transportation routes with 1-kilometer resolution and to tabulate any population-related variable included in the *block* data. A preliminary study of population densities along the proposed WIPP routes in New Mexico revealed that the desired population densities could in fact be tabulated along any major highway and most minor highways on a kilometer-by-kilometer basis. However, the process was labor intensive (though much less so than creating the HIGHWAY databases). Also, in rural areas and most suburban areas, many *blocks* extend over distances which were large in comparison to the 1-km scale of interest in collecting population density data.

This report describes an ancillary code [3] for use with a GIS to automatically tabulate kilometer-by-kilometer population densities along any selected route segment. Second, it describes a statistical method of estimating population densities near the route segment from the average densities in *blocks* (or groups of *blocks*) extending significantly more than 1 km from the route segment. Third, it presents a method of describing population density under a hypothetical plume as a function of downwind radioactive-aerosol concentration that is based on an adaptation of the same ancillary code.

Geographic Information System

The particular GIS employed in this study is ArcView™ which embodies a graphic user interface and a subset of the functionalities of ARC/INFO™; both are products of Environmental Systems Research Institute, Inc. (ESRI). The capabilities inherent in the ArcView system include **display** of multiple maps of features and data that are geographically distributed (e.g., highway maps, *block* boundaries, household locations, county boundaries, etc.), **selection** of one set of features which is in a specified geographic relationship to another set of features (e.g., *blocks* within 0.8 km of a selected highway, or *blocks* intersected by a graphic figure), and **tabulation** of the characteristics of identified features (e.g., population count, population density, household count and area within identified groups of *blocks*). Functionality may be expanded through incorporation of ancillary codes (scripts) supplied by users or by ESRI.

Correction for Large Census Blocks

As mentioned earlier, a preliminary application of the GIS-based method of population-data acquisition to the proposed WIPP route in New Mexico (primarily I25 from Colorado south) revealed many instances of *blocks* or groups of *blocks* intersected by a 1.6-km^2-rectangle (1.0 along the route and 1.6 km wide, centered on the route) with areas much larger than 1.6 km^2. A single *block* of large area was not likely to be populated uniformly, and the geographic distribution of population within a *block* was not specified in *block* data. Thus, precise determination of the population within 0.8 km of the route centerline was not possible. Simple approximations, such as (1) the average population density over a *block* or group of *blocks*, or (2) the assumption that all of the population in selected *block*(*s*) lies within the 1.6-km bandwidth were expected to yield either underestimates or gross overestimates (by factors of 10 to 100 according to data described below), respectively. Underestimates are unacceptable because they do not yield conservative risk estimates and large over-estimates are unacceptable because the majorities (~90%) of most routes are Rural or Suburban in character.

There are several means of accurately determining population distribution within an area of interest: direct surveys, aerial photographic surveys and special U.S. Census databases. All three expensive and the U.S. Census databases are not always publicly available. The most economical approach for the present need was to acquire ArcView-compatible databases of the coordinates (locations) of households. An initial database of household locations for the 911 emergency system in McKinley County, NM, was the first such database used in this study. It was possible to demonstrate with this database that a distribution of population-density ratios, 1.6PD/AvgPD, could be developed. The "1.6PD" population density was computed by using the GIS to identify the number of households within the 1.6 km^2 area of a selected rectangle and multiplying this number by the number of persons per household, as determined from the *block* data. Dividing this number by 1.6 km^2 yielded an estimate, with acceptable accuracy, of the population density within 0.8 km of the route for the selected kilometer. The average population-density value (AvgPD) was derived from the total number of persons and total area of the *block*(*s*) intersected by the rectangle. Because the number of 1.6PD/AvgPD values obtained from this limited data set was small and related to a particularly rural area, more extensive and varied data sets were sought.

A suitably large database of residence locations was obtained from the Houston-Galveston Area Council (HGAC) in Texas, which provided properly configured coordinates of customers within the Houston Lighting & Power Service Area (covering parts or all of 12 counties). Because these customers were approximately 80% residential and located along Interstate, US, and State highways, their locations provided the data needed for definition of a credible 1.6PD/AvgPD distribution function.

Fifteen highway segments of varying length (11 to 59 km) were analyzed to provide 498 values of 1.6PD/AvgPD. Figure 1 shows 15 cumulative distributions of these ratios, derived from histograms calculated for each of the 15 data sets, plus a lognormal distribution function which was fitted visually to the 15 plots. These 15 data sets are distinguished by road type and county; each may traverse Rural, Suburban or both categories of population density.

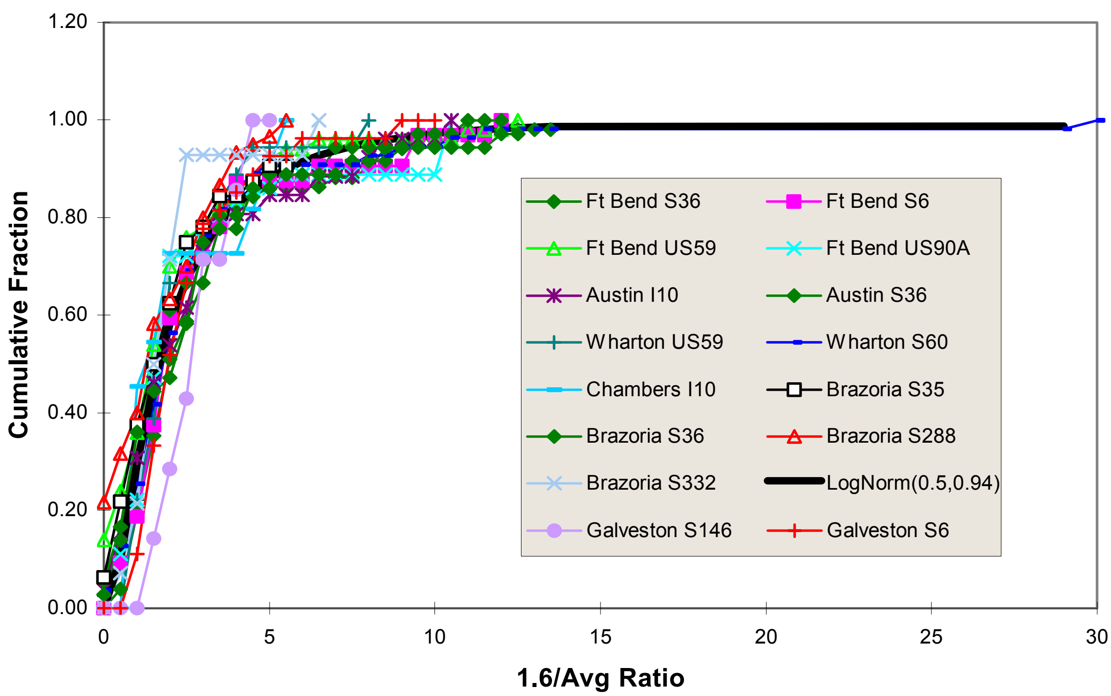

Figure 1 – Distributions of 1.6PD/AvgPD Ratios for all HGAC Data

In order to discover possible correlations among the route types, histograms of data from multiple counties for a particular highway type were calculated; the resultant cumulative distributions are presented in Figure 2. Figure 2 also includes a distribution (labeled Grand Comb.) which is derived from a histogram of the aggregated ratios from all highways and all counties. In Figure 3, the Grand Comb. distribution is compared with the distributions from McKinley County, NM and with a lognormal distribution fitted (least squares) to the Grand Comb.

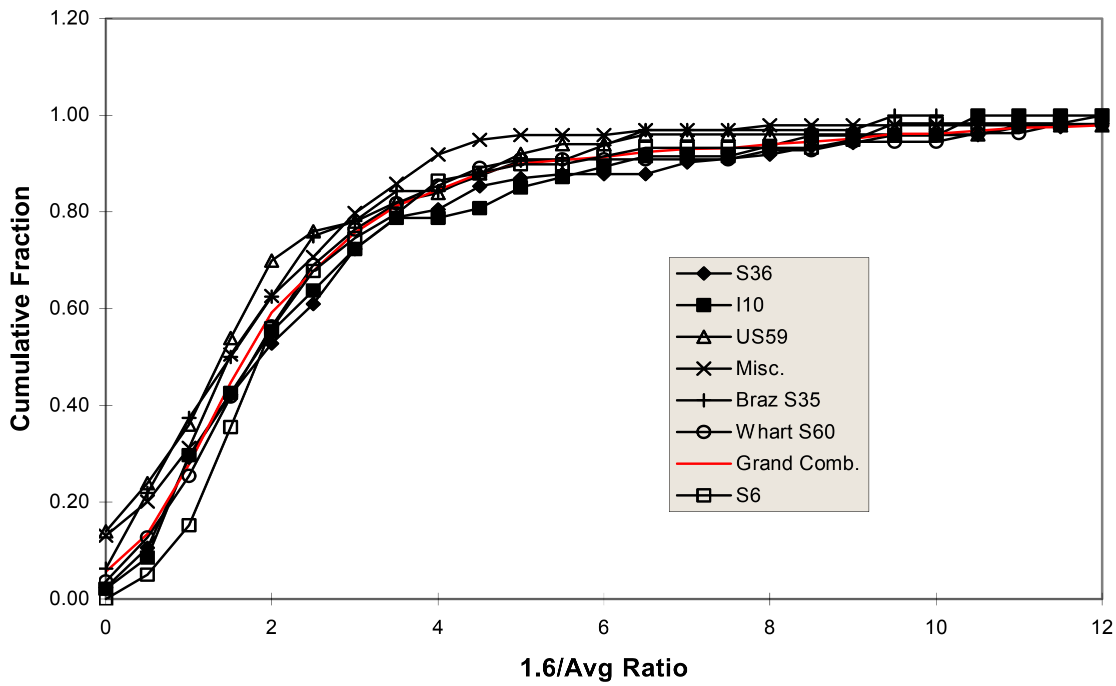

Figure 2 – Comparison of Distributions by Highway Type

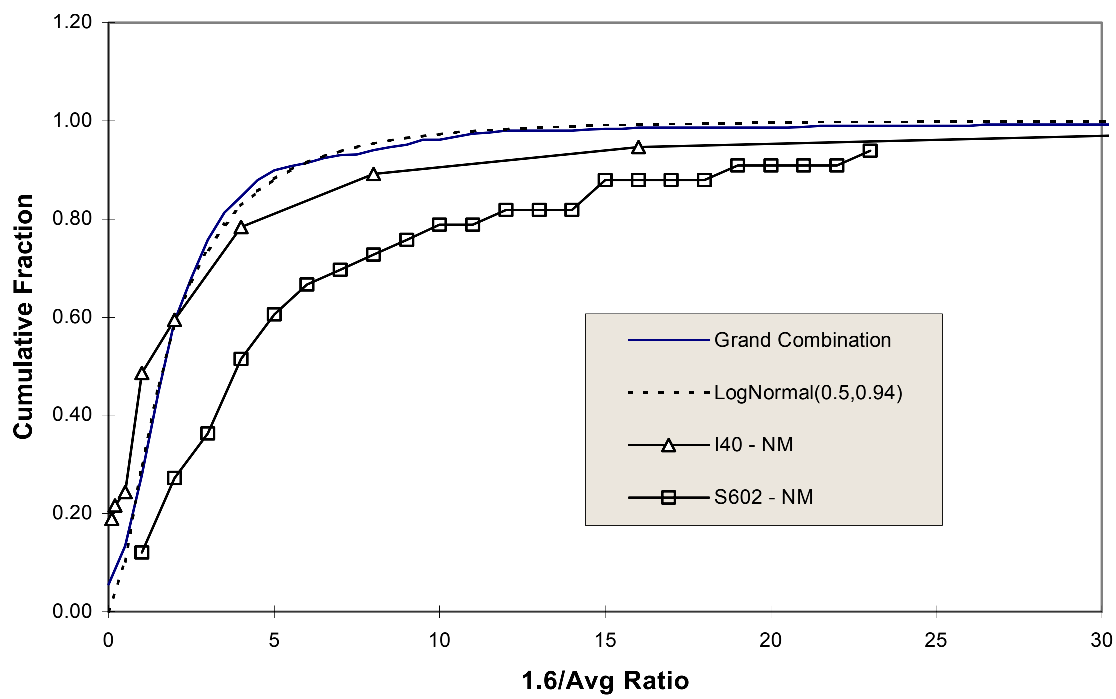

Figure 3 – Grand Comparison of Distributions

Automated Data Collection

Population Data for Incident-Free Analysis

In the implementation of an automated means of collecting and tabulating population data for incident-free risk analysis along a selected route, the basic method used in manual collection is maintained: for each kilometer of the route, identify all *blocks* with any portion of their areas lying within 0.8 km of the route centerline and compute an average population density (the quotient of the total population and the area of the *blocks*). Identification of the *blocks* is accomplished by a slightly different means:

1) The route (or any portion) is tagged using a standard tool in the GIS
2) A 1.6-km-wide* "worm" overlying the selected route is generated
3) This "worm" is divided into 1-km* segments
4) *Blocks* intersected by each segment are identified and tabulated
5) Selected categories (total population, total area and others of interest) of *block* data are summarized by a modified GIS tool for each km of the route
6) A database file is exported for further processing in a spreadsheet

*Other bandwidth or length values can be selected by the user.

For a route of a few hundred kilometers these 6 steps require ~1 hr compared to 1/2 day or more for the manual method. Automation enables routine GIS-based acquisition of route population data with greatly reduced potential for human error.

Population Data for Accident-Risk Analysis

The population model employed in accident-risk analysis with RADTRAN (releases 4 and earlier) was based on the incident-free data for each segment of a route and an assumption that the population density for a particular kilometer of the route covers the entire area under the dispersion plume for all accidents that might occur in that 1-km segment. This model acknowledged the impossibility of knowing where an accident might occur or what the wind velocity (direction and speed) might be at an arbitrary location and time. This same approach may be used with RADTRAN 5, but separate population densities for sections of the plume footprint also may be specified. In order to quantify the accuracy of the original model and to enable analysis of selected potential accident sites of interest with the new model in RADTRAN 5, a modification of the incident-free, automated GIS tool was developed.

Typically, a dispersion plume is defined by a set of contours of constant time-integrated concentration called isopleths; a set of 18 precalculated isopleths, described by their areas and associated time-integrated concentrations, has been available in RADTRAN [1] for average weather and ground-level release conditions. A GIS script was written which draws these isopleths aligned at a point and compass angle of choice. The *blocks* intersected by each of these isopleths are selected and the desired data are extracted by the appropriate scripts developed previously; population densities are tabulated for individual isopleths aligned on a point rather than for individual kilometers along a route.

Sample Applications

Population Data for Incident-Free Analysis

An Interstate highway route from St. Louis to Kansas City, Missouri, along I70 (and loops in the cities) was selected for demonstration purposes since it includes a representative mix of Rural, Suburban and Urban segments, and has been analyzed previously with a manual prototype. The route starting-point was selected as the IL/MO border on I270; it proceeds west along I70 and ends at the MO/KS border on I435 (Figure 4). A portion of the route (suburbs of Kansas City) and the overlay of a 1.6 km wide band ("worm") with nominal 1.0-km segmentation are shown in Figure 5. Note that the segmentation is not precisely 1.0 km over sections of highway that include intersections with other Interstates or that have radii-of-curvature on the order of 1 km. Such instances represent a negligible fraction of a typical route on Interstate highways. The Rural, Suburban and Urban population densities derived from the 6-step automated procedure, uncorrected and corrected for large *blocks*, are summarized in Table 1.

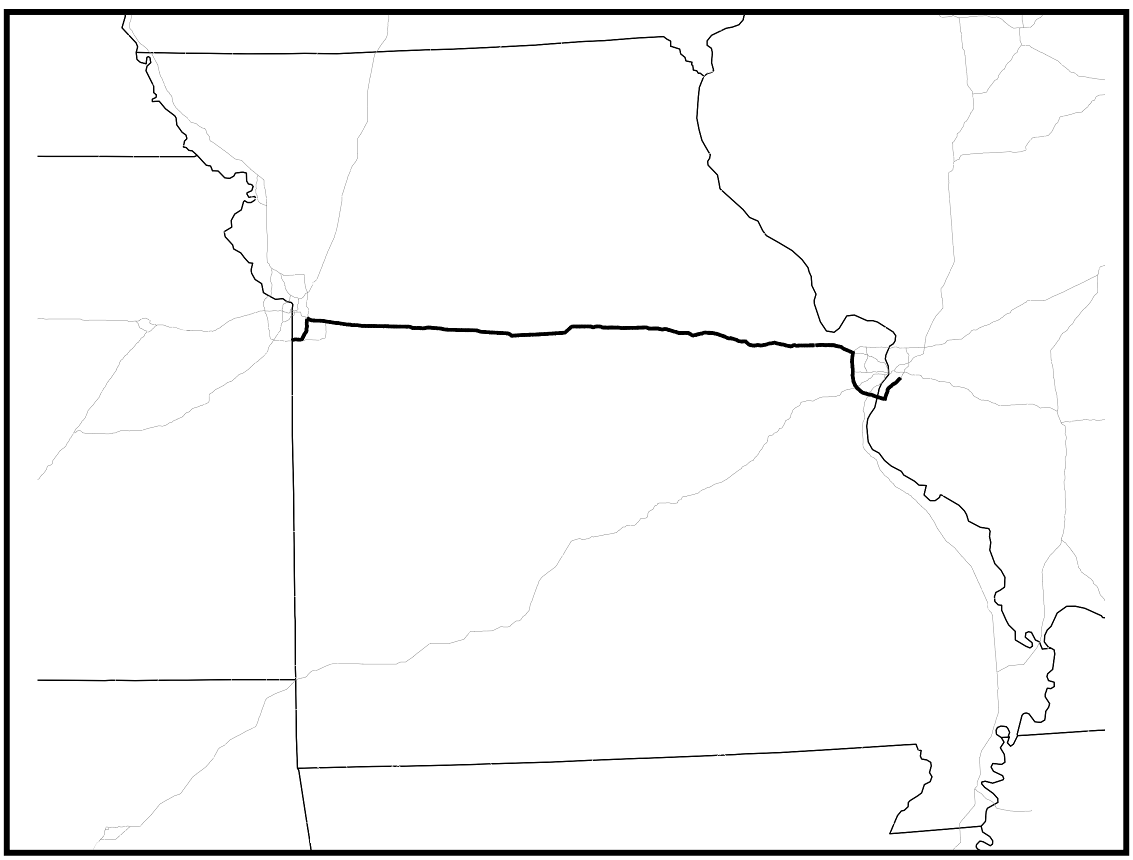

Figure 4 – Sample Route Across Missouri (from Kansas City to St. Louis)

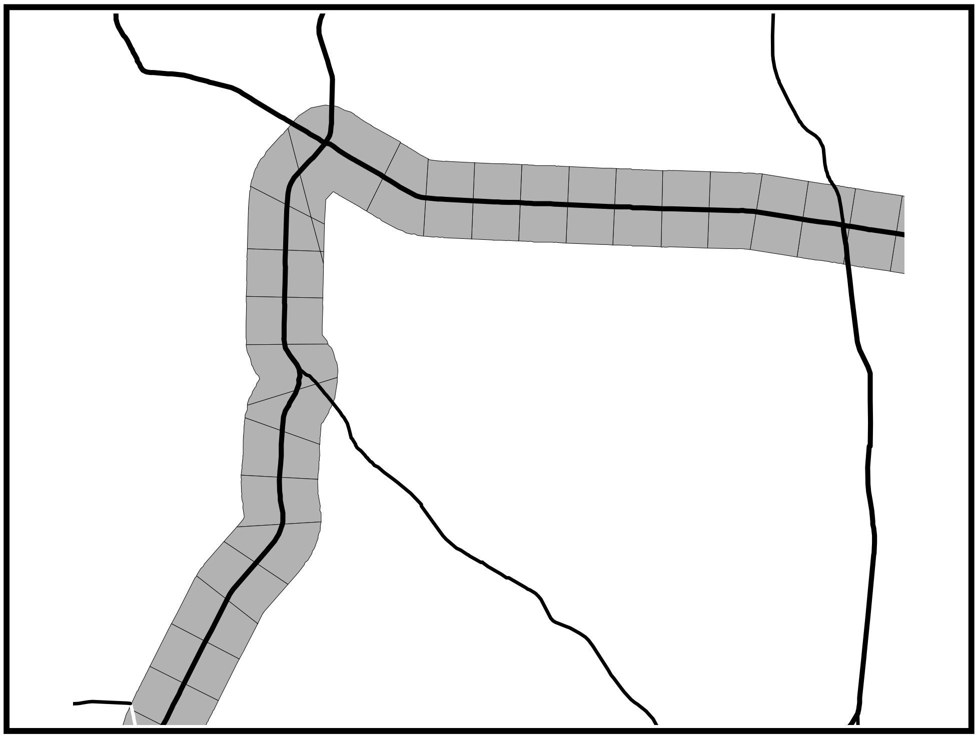

Figure 5 – Route with 1.6 km Band and 1.0 km Segment Overlay

Table 1 – Summary of Incident-Free Population Data from Automated Method

	Uncorrected Data		Corrected Data	
	Persons/km^2	kilometers	Persons/km^2	kilometers
Rural	12.6	270	15.9	252
Suburban	486	150	515	137
Urban	1798	1	3453	32

As a check on the accuracy of route segmentation by the automated method, the total length of 421 km may be compared with 419 km obtained with the manual prototype employed previously and 417 km obtained from the HIGHWAY database. The total potentially-exposed populations also were compared: 122,339 persons (manual) and 124,960 persons (automatic). These values are sums of the three uncorrected population densities multiplied by their respective lengths and 1.6 km.

Population Data for Accident-Risk Analysis

A point on I270 in the suburban area west of St. Louis, MO, and a random angle (105 deg.) were selected for demonstration of the accident-risk population tool. Figure 6 shows a portion of the 18 isopleths, highways in the St. Louis area (darkest lines are Interstates) and the *block* outlines in the affected counties in Missouri and Illinois. Table 2 presents data tabulated by the automated GIS tool for the largest 13 isopleths. Areas of the smallest 5 isopleths are less than

Figure 6 – Sample Overlay of Isopleths on St. Louis, MO and Western IL

1.6 km^2; their data are set equal to the values for the sixth ("0") isopleth. As may be discerned from the sequences of values, the data in Table 2 are totals for each entire isopleth; population density for each isopleth, exclusive of the smaller isopleths within it, is calculated according to the following equation:

$$PopDen_n = \frac{Pop_n - Pop_{n-1}}{Area_n - Area_{n-1}}$$

where *n* is the isopleth number. The data in Table 2 are easily processed according to this equation by means of a spreadsheet, and the results are the values needed for input to RADTRAN.

Table 2 – Population Data Automatically Tabulated for Isopleths in Figure 6

Isopleth	Number of *Blocks*	Total Population	Total Area (km^2)
0	7	1121	3.86
1	7	1121	3.86
2	9	1252	4.07
3	11	1340	4.80
4	38	4825	8.71
5	51	6141	10.10
6	133	21021	22.04
7	265	30396	35.74
8	466	45890	87.41
9	862	70619	257.27
10	1612	103685	662.52
11	2041	119777	1023.83
12	2566	131733	1602.24

Conclusions

Examination of the cumulative distributions in Figures 1 – 3 indicate no clearly discernable correlations of the 1.6PD/AvgPD ratios with county (i.e. population density range) or highway type except S602 in McKinley County, NM, which is a state highway. This might suggest that a general model for estimation of population densities along routes in regions consisting of large *blocks* should not be based on a single distribution function. However, use of a minor (state) road in such a sparsely populated region for RAM transport would be unusual. Interstate highways are used in most long-distance transportation by truck mode (their use is mandated for highway route controlled quantities, e.g., spent nuclear fuel). Lesser highways are employed, if at all, as access routes between the nearest Interstate highway and shipment origin and destination points. Therefore, the lognormal distribution shown in Figure 3 has been incorporated into the GIS-based population model for estimation of population densities within 0.8 km of RAM-shipment truck routes in areas for which *blocks* do not provide adequate resolution. This distribution may be sampled through use of the Latin Hypercube Sampling [4] code to provide correction-ratio values to be multiplied by average population densities obtained from selected *block*(*s*) with a total area greater than 3.2 km^2. The latter area was chosen as the threshold for requiring correction because *blocks* generally extend beyond the approximately 1.0 x 1.6 km "worm" segment, even in areas in which *blocks* are small compared to the segment.

The results in Table 2 demonstrate that for a single location and angle, the population densities can differ significantly from the value obtained within 0.8 km of the route, i.e., the "0" isopleth values. At a minimum, multiple angles at a selected site must be measured and averaged to obtain a representative evaluation of a single, potential accident site. Note that for transportation-accident risk analysis, in contrast to fixed sites, wind-direction frequency data are not available for the great majority of all potential accident sites on a route. A statistical comparison of a potential RAM transport route has been performed [5] in which total accident-dose risks were computed with RADTRAN and the GIS-based population densities within 0.8 km of the route. The results were compared to the average of results based on data obtained with isopleth-specific population data. The latter were computed for the midpoints of 29 separate Rural, Suburban or Urban segments and four angles (randomly selected in each compass quadrant). These two average accident-risk doses for an approximately 400 km route were in substantial agreement. The difference between them was much less than the standard deviation of the GIS-based risk estimates.

Automation of GIS-based population data-collection has made practical new application of this sophisticated, commercial computer tool to transportation risk analysis. In addition to providing geographically specific Rural, Suburban and Urban population densities for incident-free dose estimates, geographically detailed examination of populations affected by potential accidents (at selected locations or in statistical studies) may also be conducted for RAM transportation. In both cases, public concerns regarding selected population centers can be addressed with speed and economy while the results represent a significant improvement in accuracy in calculating aggregated doses over entire transport routes.

The method developed for approximating population densities near routes in areas of relatively large U.S. Census Blocks leads to an expected shift to higher population densities (especially in the Urban category) and to somewhat greater fractions of the route being placed in the Urban category. This has the effect of making the resultant estimate of total incident-free dose more conservative but not by the factors of 10 to 100 associated with simpler methods. As more states and communities continue to develop databases specifying residence locations, the distribution(s) of 1.6PD/AvgPD can be refined.

References

1. Neuhauser, K. S., and Kanipe, F. L. "RADTRAN 4 - Vol. 3: User Guide," SAND89-2370, Sandia National Laboratories, Albuquerque, NM (1992).

2. Oak Ridge National Laboratory (ORNL), "HIGHWAY 3.1 - An Enhanced Highway Routing Model: Program Description, Methodology, and Revised User's Manual," ORNL/TM-12090, ORNL, Oak Ridge, TN (1992)
3. Earth Touch Solutions, LLC, Albuquerque, NM, 87109-5058

4. Iman, R. L. and M. J. Shortencarier, "A FORTRAN Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models," NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, NM, March 1984.

5. Mills, G. S., Neuhauser, K. S., “Statistical Evaluation of Population Data for Calculation of RAM Transport Accident Risks”, Risk Analysis: An International Journal, to be published.

Distribution

1	MS0701 L.E. Shephard, 6100
1	MS0718 C.D. Massey, 6141
10	MS0718 K.S. Neuhauser, 6141
10	MS0718 G.S. Mills, 6141
1	MS0718 F.L. Kanipe, 6141
1	MS0718 R.F. Weiner, 6141
1	MS1138 J.H. Ganter, 6535
1	MS0720 D.R. Bleakly, 6804
1	MS9018 Central Technical Files, 8940-2
2	MS0899 Technical Library, 4916
1	MS0612 Review & Approval Desk, 4912 For DOE/OSTI
1	U.S. Dept. of Energy National Transportation Program Office Attn.: S.C. Hamp P.O. Box 5400 Albuquerque, NM 87185