Integration of Particulate Air Modeling with a GIS: An Exposure Assessment of Emissions from Two Phosphate-Processing Plants

Gregory V Ulirsch,* Debra L Gable, Virginia Lee
Agency for Toxic Substances and Disease Registry, US Public Health Service, Atlanta, GA

Abstract

Particulate air emissions from two phosphate-processing plants, which constitute the Eastern Michaud Flats Superfund site in Pocatello, Idaho, will be modeled as part of an ongoing public health assessment being conducted by the federal Agency for Toxic Substances and Disease Registry (ATSDR). The results of the air dispersion modeling will be integrated into a geographic information system (GIS) as a separate coverage. Based on current health-based guidelines for particulate air exposures, the results of the air dispersion modeling will be overlaid on base coverages provided by the TIGER/Line files, integrated with applicable demographic information from the US Census Bureau summary tapes. Demographic information on the population from the census block groups that are located completely within areas exposed at levels of health concern, will be abstracted from this overlay analysis. Other techniques (e.g., simple population density spreading or the kernel density method) will be employed to determine the demographic information from census block groups that are partially exposed at levels of health concern. Maps will be produced that display the areal concentration isopleths of the modeled particulate emissions and show which census block groups are affected. Demographic information from the created attribute tables will be queried and summarized to determine the total population exposed at levels of health concern, age structure, socioeconomic status, and other parameters. The results of the exposure assessment will be used as a basis for a separate study, an ecologic health study of mortality in the population exposed at levels of health concern. This study will be conducted at the University of North Carolina at Chapel Hill.

Keywords: exposure assessment, epidemiologic health studies, particulate matter, air modeling

Study Background

In 1996, the Shoshone-Bannock Tribes in Fort Hall, Idaho, contacted the federal Agency for Toxic Substances and Disease Registry (ATSDR) requesting an evaluation of current and historical exposures to particulate matter (PM) and other air emissions from two nearby phosphate-processing plants (one owned by the FMC Corporation [FMC] and one owned by the JR Simplot Corporation [Simplot]). Together, these corporations constitute the US Environmental Protection Agency's (EPA's) Eastern Michaud Flats Contamination (EMF) National Priorities List (NPL) site. Tribal and non-tribal community members have consistently expressed concern regarding the occurrence of

* Gregory Ulirsch, Centers for Disease Control and Prevention, 1600 Clifton Rd. (E-32), Atlanta, GA 30333 USA; (p) 404-639-0624; (f) 404-639-0653; E-mail: gru1@cdc.gov

asthma and upper respiratory infections that, they believe, are related to exposure to air pollutants emanating from the EMF site. In 1995, ATSDR completed a health study of persons living on the Fort Hall Indian Reservation to investigate concerns related to respiratory and renal disorders being treated by the Indian Health Service clinic. This study concluded that the prevalence of pneumonia and chronic bronchitis was statistically significantly elevated among Fort Hall participants as compared to participants at another Native American reservation. Testing of pulmonary function in the Fort Hall sample showed decreased air flow, but none of these differences were statistically significant. Biological monitoring for cadmium, chromium, and fluoride values in urine samples from both reservations were within normally defined values, and no differences between the two reservations were found (1). ATSDR is also currently investigating the potential for human exposures (past, present, future) to groundwater, surface water and sediment, surface soil, and biota in relation to the EMF site.

Goals and Purpose of Study

The major limitation of the previous ATSDR health study of the residents of Fort Hall was the uncertainty in assigning exposure levels to contaminants emanating from the two phosphate-processing plants (1). In addition, the study recognized that most of the highest exposures to air contaminants may have occurred in the past and that the study methodology could not identify historically exposed persons (1). The current ATSDR exposure assessment will attempt to determine a population that is currently and historically exposed to air emissions (particularly PM) from the two phosphate-processing plants and other potential sources. Using the results of ATSDR's exposure assessment, the University of North Carolina at Chapel Hill (UNC) School of Public Health will conduct an ecologic health study of respiratory and cardiopulmonary mortality in areas where persons have been exposed to PM at the level of health concern.

Study Area and Site Background

The EMF NPL site is made up of the FMC Elemental Phosphorous Plant and the Simplot Don Plant. The nearest major population areas, Pocatello and Chubbuck, Idaho, are located east-southeast and east-northeast, respectively, of the FMC/Simplot plants (Figure 1). The facilities are about 2.5 miles from Pocatello. The FMC plant is located on Fort Hall Reservation land (in the southern part of the reservation) and the Simplot plant is on state land. The Town of Fort Hall is located about 8 miles north-northeast of the facilities.

The FMC plant covers an estimated 1,189 acres and adjoins the western boundary of the Simplot plant (2). Elemental phosphorus production at the facility has changed little since the plant operations began in 1949. Phosphorus-bearing shale is shipped to FMC via the Union Pacific Railroad during the summer months. Ore cannot be shipped in the winter because it would freeze in the rail cars; therefore, the ore is stockpiled at the facility during the winter months. Ore from the stockpiles is processed in four electric arc furnaces. The furnaces' reaction yields gaseous elemental phosphorus and byproducts. Some of the byproducts contain radioactive components. The elemental phosphorus is subsequently condensed to a liquid state and eventually shipped off site. About 1.5 million tons of ore are processed at the plant each year. The disposal of

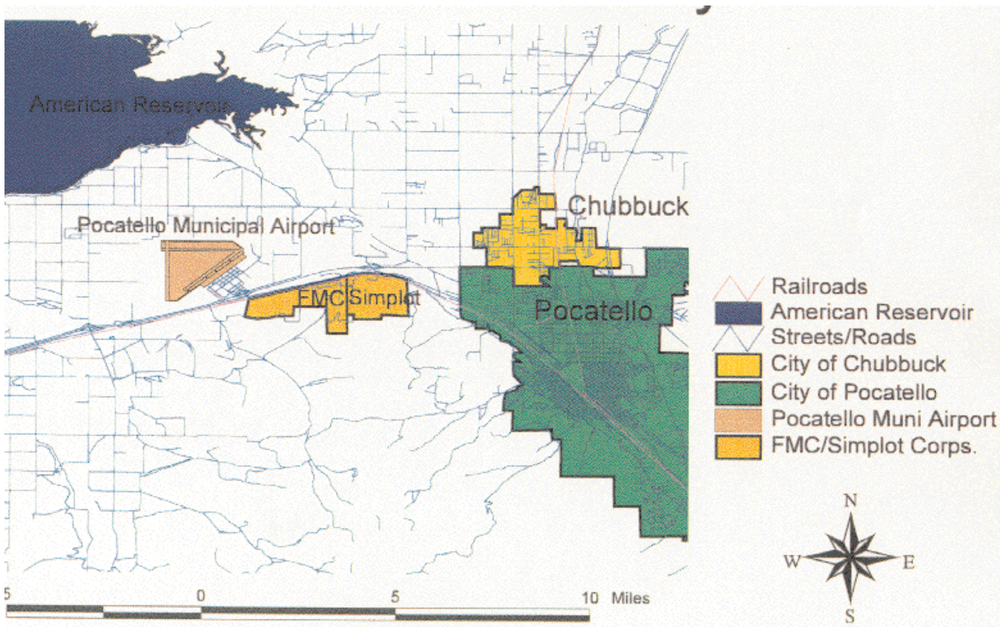

Figure 1 ATSDR/UNC air study area, Pocatello, Idaho.

byproduct waste material at and around the facility has produced slag piles that cover large areas of land (2).

The Simplot plant covers about 745 acres and adjoins the eastern property boundary of the FMC facility (2). The plant began production of single superphosphate fertilizer in 1944. In 1954, the facility began producing phosphoric acid. The phosphoric acid is now produced using an aqueous process. Formerly, phosphate ore was transported from the mines to the facility via rail. As of September 1991, the Simplot plant receives phosphate ore through a slurry pipeline. The phosphate ore slurry is processed at the Simplot plant in phosphoric acid reactors and then further processed into a variety of solid and liquid fertilizers. The plant produces 12 principal products, including five grades of solid fertilizers and four grades of liquid fertilizers (2).

Epidemiologic Studies of Particulate Matter Exposures

"Particulate matter" is the term used for a mixture of solid particles and liquid droplets found in the air. Coarse particles (larger than 2.5 micrometers [μm] in diameter) come from a variety of sources, including windblown dust and grinding operations. Fine particles (smaller than 2.5 μm) result from fuel combustion (from motor vehicles, power generation, and industrial facilities), residential fireplaces, and wood stoves. Before 1987, EPA's standards regulated larger particles (also know as total suspended particles [TSP]). By 1987, research had shown that the particles of greatest health concern were those 10 μm in diameter or smaller, which can penetrate into sensitive regions of the respiratory tract. At that time EPA and the states took action to monitor and regulate PM that was 10 μm and smaller (PM_{10}). In the years since the previous standard was enacted, hundreds of studies have been published on the health effects of PM. These studies suggest that adverse health effects in children and other sensitive populations have been associated with exposure to PM concentrations well below that allowed by the 1987 PM_{10} standard (3).

Moreover, these studies have indicated that the fine particles ($PM_{2.5}$), which penetrate more deeply into the lungs, are more likely than coarse particles to contribute to adverse health effects. Some of the health effects associated with $PM_{2.5}$ exposures are (3):

- Premature death
- Respiratory-related hospital admissions and emergency room visits
- Aggravated asthma
- Acute respiratory symptoms, including aggravated coughing and difficult or painful breathing
- Chronic bronchitis
- Decreased lung function that can be experienced as shortness of breath
- Work and school absences

These studies indicate that the elderly, individuals with pre-existing heart or lung disease, children, and asthmatics are at the most risk for adverse health effects from exposure to $PM_{2.5}$. For these reasons, on July 17, 1997, EPA revised its PM standards to include a primary (health-based) annual average $PM_{2.5}$ standard of 15 micrograms per cubic meter ($\mu g/m^3$) and a 24-hour $PM_{2.5}$ standard of 65 $\mu g/m^3$ (3).

Study Area Topography and Meteorology

The local terrain in the Pocatello area is classified as meteorologically "complex." East of Pocatello, the Pocatello Mountain Range rises from about 4,400 feet to about 6,500 feet above mean sea level. Southeast of the FMC and Simplot facilities is the city of Pocatello, which lies in the funnel-shaped Portneuf Valley. The valley virtually closes at the southern boundary of the city of Pocatello. The northern end of the Bannock Range is immediately south of the FMC and Simplot facilities. This range tapers down to a north-pointing wedge shape just east of the Simplot facility and forms one side of the Simplot gypsum stacks. The ridge just southeast of Simplot rises from Simplot's base elevation of 4,449 feet to about 5,700 feet. The terrain south of the facilities (between them and the Bannock Range) gives way to the Michaud Flats of the Snake River drainage to the north, and to the Arbon Valley to the west. From the southwest, clockwise to north-northeast of the facilities, the terrain is generally flat for several miles (4).

Long-term meteorological data in the study area are primarily obtained from the National Weather Service (NWS) station at the Pocatello Airport, located west-northwest of the FMC and Simplot facilities (Figure 1). The wind rose (Figure 2) shows a marked preference for west to south winds with a prevailing wind direction from the southwest. A secondary preference for wind direction is indicated from the northeast. The lowest frequency of wind direction is from the east to south-southeast. Wind data from Simplot's meteorological station, located north of the Simplot facility, indicate a prevailing wind direction from the southwest to west-southwest, with a strong second predominant wind direction from the southeast to east-southeast. This secondary flow is clearly out of the Portneuf Valley and is a nighttime drainage wind flow (4).

Emissions from the phosphate plants and area topography both contribute to local

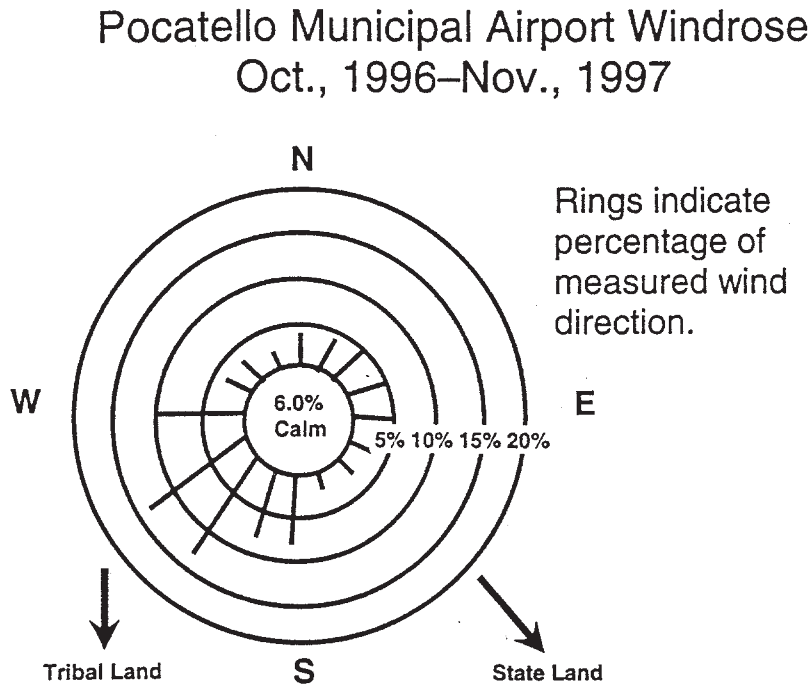

Figure 2 Pocatello Municipal Airport wind rose, October 1996 through November 1997. Source: Idaho Department of Environmental Quality.

air pollution. Particulates, phosphate pentoxide, metals, and radionuclides escape into the atmosphere from the stacks during production of phosphorus products. Fugitive dust from the waste ponds, ore stacks, and waste piles on the site is also a concern. The effect of these industrial emissions on air quality is compounded by the complex local topography and climatic conditions. Winds from the southwest sector carry pollutants from these plants toward population areas in northern Pocatello, Chubbuck, and the Fort Hall Reservation. In the study area, temperature inversions, caused by high-pressure subsidence and radiative cooling, can occur year-round. During these inversions, emissions from the industrial plants might become trapped and form a dense brown cloud about 1,000 feet above the ground, extending 4 to 5 miles in length and 32 miles in width, or the emissions might stay at ground level (1).

Area Monitoring Network and Historical PM Levels

The Idaho Department of Environmental Quality (IDEQ) began monitoring the air in the study area for TSP in 1975. Basing its decision on TSP data collected from 1975 to 1977, EPA designated the area as in nonattainment status (5). The original air monitoring network consisted of three stations: the sewage treatment plant (STP), the Chubbuck School (CS), and Idaho State University (ISU) (Figure 3). Monitoring data for the maximum 24-hour and average annual TSP concentrations are available from these three monitoring stations for 1977 through 1987. EPA's primary health-based average annual TSP standard at that time was 75 $\mu g/m^3$. During this 11-year period, this

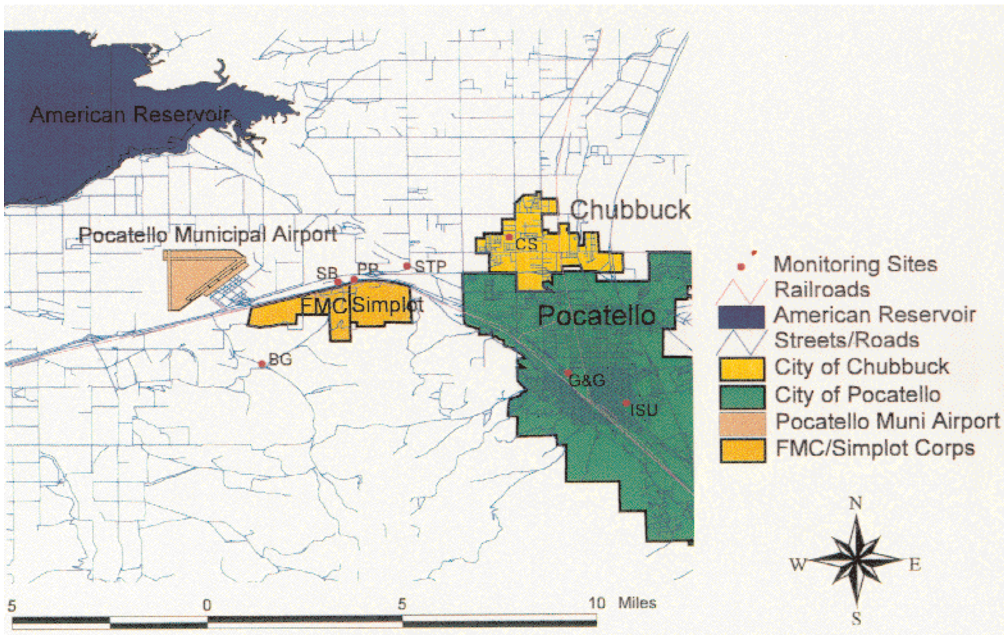

Figure 3 Particulate monitoring sites in study area.

standard was exceeded 10 times at the STP site, 4 times at the CS site, and 0 times at the ISU site. Average annual and 24-hour maximum PM_{10} monitoring has been ongoing at the STP site since 1986, at the CS and ISU sites since 1988, and at a new site (Garrett and Gould [G&G]) since 1990 (Figure 3). EPA's health-based annual average PM_{10} standard (established in 1987) is 50 µg/m^3. During the time period of PM_{10} monitoring, the standard has been violated 4 out of the 12 times at which samples were taken at STP, 0 out of 9 times at ISU, 0 out of 10 times at CS, and 0 out of 7 times at G&G. Since 1996, the Shoshone-Bannock Tribes and EPA have operated three air monitoring sites: two located just north of the FMC facility (Primary Particulate [PP] and Shoshone-Bannock [SB]) and one background site (BG) located about 4 kilometers west-southwest of the PP and SB monitoring sites (Figure 3) (6). Monitoring at these sites is primarily for PM_{10}; however, every third day, samples from a dichotomous sampler, located at the PP site, provide data for both the PM_{10} and the $PM_{2.5}$ fraction (6). Although average annual monitoring data are not currently available for the PP and SB sites, EPA's 24-hour PM_{10} standard of 150 µg/m^3 (established in 1987) has been exceeded at least 44 times between October 1996 and September 1997.

Included in EPA's 1997 revision of the PM standards were regulations that called for implementation of a monitoring network for $PM_{2.5}$. The IDEQ is currently in the process of implementing this monitoring network within the study area. Except for the more recent $PM_{2.5}$ data available from the PP monitoring site and several other seasonal studies that included monitoring for $PM_{2.5}$, historical $PM_{2.5}$ data are very sparse. PM_{10} monitoring has also not produced a complete database dating back to 1977. To obtain a better understanding of the historical $PM_{2.5}$ and PM_{10} levels in the study area, several site-specific ratios were used to estimate these levels where monitoring data were not available. Long-term data indicate an average PM_{10}/TSP ratio of about 0.5 (7). Three different studies have calculated ratios of $PM_{2.5}$ to PM_{10} of 0.5, 0.66, and 0.76 (8). Although the 1978–1979 study ($PM_{2.5}/PM_{10}$=0.76) was considered the best study with respect to the number of samples taken (9), for the sake of this evaluation, the middle value of 0.66 was chosen. Based on this ratio, actual and estimated historical values for PM_{10} and $PM_{2.5}$ were calculated and plotted (Figures 4 and 5). From these data, it can readily be observed that PM levels at these monitoring sites have dramatically declined since 1992. It is not known whether this decline is due to Simplot's switch to a wet process or due to other measures to reduce other sources of PM in the study area, or due to both. From a public health standpoint, these data are also very illustrative. The data indicate that, since 1977, the only monitoring station at which the PM_{10} standard has been exceeded has been the STP site, which is located in a relatively sparsely populated area. Because the CS, ISU, and G&G sites are all located in more densely populated areas, one could conclude that the ambient air levels of PM, based on PM_{10} levels alone, do not indicate a large public health impact. As previously indicated, however, studies have shown that $PM_{2.5}$ is the more important PM fraction from a public health standpoint. As shown in Figure 5, since 1977, the health-based $PM_{2.5}$ standard of 15 $\mu g/m^3$ may have been exceeded frequently at all three monitoring stations located in populated areas. These data provide evidence for public health concern that persons in more densely populated areas have been exposed to $PM_{2.5}$ at levels that may result in adverse health effects.

Design of Air Study

ATSDR is currently designing an exposure assessment methodology that will not only address community concerns regarding past and present exposures to PM and other

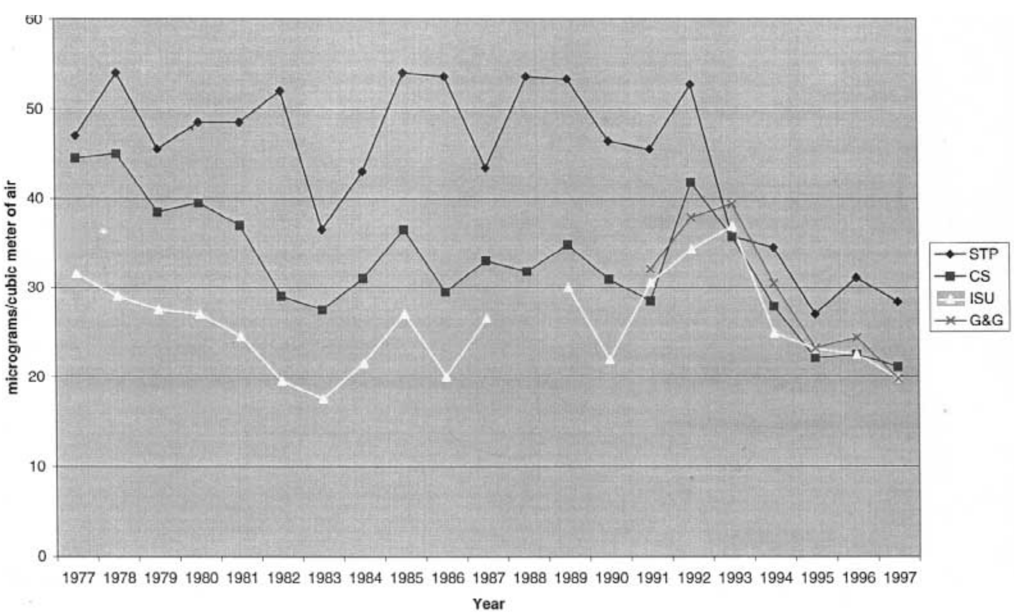

Figure 4 Estimated and actual annual average PM_{10} concentrations in study area.

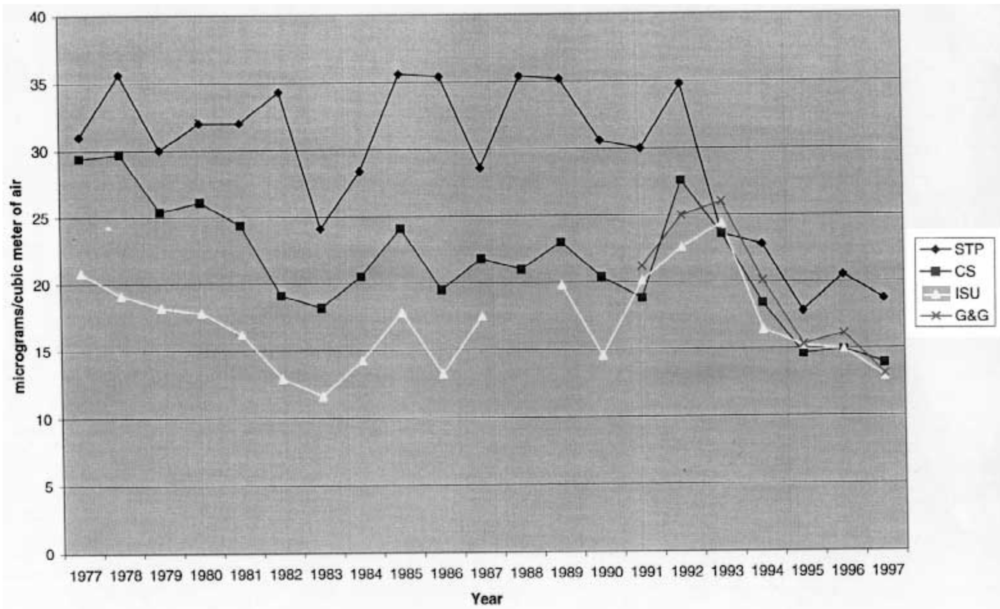

Figure 5 Estimated annual average $PM_{2.5}$ concentrations in study area.

contaminants emanating from FMC/Simplot facilitates, but will also be scientifically defensible. The second phase of the planned study, an ecologic health study of respiratory and cardiopulmonary mortality, to be conducted by the UNC's School of Public Health, is also under design. The basic designs of the exposure assessment and ecologic health study have already been conceptualized. Some of the methodological issues that have been and will be evaluated to determine the final design approach are discussed in the next section.

A geographic information system (GIS) will be a main feature of the design of the studies because (1) a GIS environment is an excellent platform for bringing together disparate databases, (2) a GIS can be used to manipulate data to uncover the underlying spatial associations between various data layers (making it a "value-added" product), and (3) the "value-added" data produced within a GIS environment can be linked with statistical packages outside a GIS.

The first step in the basic design of the exposure assessment will be to use an air dispersion model, like EPA's Industrial Source Complex Model, to determine concentration contours within which persons have been exposed to levels of $PM_{2.5}$ above the health-based standard of 15 $\mu g/m^3$. The results of the air dispersion model will be imported into a GIS and will then be edited to produce concentration polygons. Within the GIS, an analysis will be performed to overlay these $PM_{2.5}$ concentration polygons onto the TIGER/Line files integrated with the 1990 or 1980 US Census Bureau summary tape data. This overlay analysis will "clip" out the demographic information of persons who have been exposed to $PM_{2.5}$ above 15 $\mu g/m^3$, as predicted by air dispersion modeling. Demographic information about the total population exposed, total susceptible populations exposed (e.g., persons over 65 years old and children under 17 years old), and socioeconomic status of persons exposed will be obtained. This analysis will be performed in five-year periods beginning with 1977 and ending with 1996.

Mortality data for respiratory and cardiopulmonary deaths (*International Classification of Diseases, Ninth Revision* [ICD-9] [10], codes 400–440 and 485–496) along with lung cancer (ICD-9 code 162), will be obtained from the Idaho State Department of Health for the counties that encompass the study area (Power and Bannock Counties) for the years 1977 through 1996. The mortality data will be grouped by the same five-year periods used in the exposure assessment. The addresses for these mortalities will be matched and geocoded to street addresses from the TIGER/Line files. A point-in-polygon analysis will then be performed to determine if the geocoded addresses for the cardiopulmonary and lung cancer deaths are located within or outside each of the polygons defining a geographic area where persons have been exposed at levels of health concern. Data on the total number of mortality cases and the total exposed population for each of the five-year periods will be used to calculate the crude mortality rates. These rates and the aggregated age, sex, and socioeconomic status data for the exposed areas will be used to control for ecologic confounding. The rates will be compared with the rates for the state of Idaho. The analytical technique to be used in the health study to control for ecologic confounding is still being evaluated (see the discussion below).

ATSDR Exposure Assessment Methodological Issues

The first basic issue in designing any exposure assessment that uses GIS is the choice of the analytic method to use to define the areal extent of exposure and the advantages and limitations of available methods. As indicated above, ATSDR has chosen air dispersion modeling to define the areal extent of exposure to $PM_{2.5}$.

There are other impediments and potential data fallacies that need to be addressed before designing any study using GIS or interpreting any result of a GIS analysis (11). Three of the major impediments to the use of GIS within ATSDR's exposure assessment are the areal interpolation problem, the fallacy of the homogeneous polygon, and the fallacy related to fuzzy boundaries (11). The areal interpolation problem arises when data obtained from one reporting unit must be combined with data from a different reporting unit; questions arise about the correct way to interpolate the overlays of these units as well as about what assumptions are to be made. The fallacy of the homogeneous polygon arises when it is assumed that a polygon delineates an area as homogeneous when indeed the phenomenon being represented by the polygon is not evenly distributed across the area. The other fallacy, related to fuzzy boundaries, occurs when it is assumed that the boundary between two polygons is discrete when it actually represents some sort of gradient (11). These methodological issues, as they relate to ATSDR's exposure assessment, are discussed below.

Exposure and risk assessment in relation to environment and health is essentially an attempt to estimate the level of exposure to specified pollutants, either for individuals or particular population groups. Direct measurements of exposure are rare, and usually only exist for a relatively small sample of people. Instead, levels of exposure commonly have to be estimated by indirect means. Two main approaches available are spatial interpolation from measurements of ambient pollution levels and modeling based on data on emission levels and sources (12). Air dispersion models are usually based on Gaussian plume dispersion equations and take into account emission source and meteorological and terrain effects. The use of dispersion modeling to assess exposure has a number of advantages. These include the fact that dispersion modeling does not rely on measured concentrations, meaning that it can be extended to areas and pollutants for which no monitored data are available. Dispersion modeling also allows for knowledge about dispersion processes and can take into account the pollution surface and local factors that influence these processes (e.g., terrain, weather) (12).

Dispersion models suffer from three major limitations. The first is the demand for data inputs to the models, which require various large databases that can be of questionable quality. The second limitation is that these models only work for areas relatively close to the emission source, where Gaussian dispersion processes can be assumed. The final limitation is that these models, for the most part, are designed to work under simple conditions; that is, they operate best with limited sources of emissions and under relatively simple terrain and weather condition (12,13). Very few attempts at linking air dispersion models within a GIS are available in the literature (12). However, there are a few examples in the published (14,15) and unpublished (16) literature that have incorporated the results of air or (more often) groundwater modeling with GIS in order to perform a risk or exposure assessment. Many of these examples are from applications of GIS to help local, state, and federal agencies evaluate risk from or exposure to contaminants (17).

As indicated above, another method available for evaluating exposure is spatial interpolation—a way to estimate pollutant levels at unsampled sites. In this type of analysis, GIS can provide a range of interpolation techniques that can be used to produce a concentration surface. However, spatial interpolation is not without its limitations and considerations. The first consideration is the method of interpolation used (e.g., inverse distance weighting, kriging, spline). The performance of different interpolation methods depends upon a number of factors, including the nature of the underlying spatial variation in the phenomenon under consideration and the sample density and distribution (12).

In exposure and risk assessment, it is vitally important to assess the link between the environment and health, especially if the results of the analysis will be used for further epidemiologic studies. Misclassification of exposure and misclassification of disease play a major role in the degree of confidence one has in the results of environmental epidemiologic investigations (18,19). Moreover, in contrast to the relatively well-defined exposure characterization variables in an occupational setting, some of the variables in environmental settings are not well defined or are not defined at all. Understanding exposures in the residential setting is even more complicated. These exposures are strongly influenced by seasonal or even daily lifestyle preferences, travel and excursion habits, and indoor/outdoor concentration differences, as well as complexities involved in the estimation of exposures in general (13). The most common approach to determining the populations at risk (exposed) has been to assume that people living near the point source are at greater risk than those who live farther away. Various summary articles (18,19) contain many examples that illustrate the use of GIS to construct a buffer around a point, line, or area source. Buffers are, however, inevitably crude indicators of risk or exposure, and without an understanding of the dispersion processes and pathways involved, it is easy to use buffers of an inappropriate size and/or shape. GIS clearly allows for modeling (either within the GIS or imported from an outside model) of more complex search areas, taking into account dispersion patterns and other effects, where data permit (12).

The above discussion clearly provides a strong argument for the use of air dispersion modeling as the preferred approach for the ATSDR exposure assessment. The use of modeling to define an exposed population can help overcome some of the impediments and data fallacies mentioned above. For example, if a radial buffer were used in the analysis, the defined zone would represent an area assumed to have homogenous exposure; however, in reality, a buffer-defined area could represent very low to very high exposures. Moreover, air dispersion modeling versus the use of buffers allows a researcher to define polygons that represent various exposure gradients, thus alleviating the problem of fuzzy boundaries. However, as previously indicated, Gaussian plume models must be used with care. The validity of using air dispersion modeling for this exposure assessment could be questioned because the FMC and Simplot sites contain well over 100 different point, line, and area PM sources, the topographic and meteorological conditions of the study area are not simple, and most of the exposed population is not near the primary sources of PM emissions. Many of these drawbacks can be overcome by incorporating corrective equations that calibrate the model with available monitoring data. For this reason, historical exposures before 1977, when air monitoring began in the study area, will not modeled.

The problems of areal interpolation and the fallacy of the homogeneous polygon must also be considered carefully when evaluating the method used to determine the demographics of the exposed population defined by the air dispersion model. The polygons that define the various exposure levels predicted by the air dispersion model will not correspond to the US Census Bureau's reporting units (e.g., block groups). Furthermore, the populations within the census units are not evenly distributed. Therefore, an overlay analysis method that does not provide some estimate of the population densities within each census unit will likely produce much exposure misclassification. ATSDR uses an area proportion program (a script written in Avenue, the programming language of ArcView GIS [ESRI, Redlands, CA]) that is easy to use and is good for many applications; however, it assumes that a population within a given census reporting unit is evenly distributed. This method may provide reasonable estimates of an exposed population in a completely urban setting; however, for this study, it is likely that the exposed population resides in urban, suburban, and rural areas. Because the results of the exposure assessment will be used as the basis for an epidemiologic study of the population exposed at levels of health concern, it is of vital importance that an accurate estimate of the "truly" exposed population be obtained. As previously mentioned, misclassification of exposure or disease in environmental epidemiologic investigations is a primary source of error. For these reasons, other methods are being evaluated that provide better estimates of population densities within the census reporting units. The two methods currently being evaluated are the kernel density method (20) and the census control population method (21). Both of these methods use techniques that "disaggregate" the census reporting units, helping to alleviate the areal interpolation problem and helping to avoid the fallacy of the homogeneous polygon.

UNC Ecologic Health Study Methodological Issues

Ecologic studies have been featured prominently in environmental epidemiology because exposures are often already measured at the group level or because the limited resources of some studies prohibit collection of individual-level data (22). Because of the various methods that can be used within and outside a GIS environment to define an exposed population, the choice of an ecologic study design is common when using GIS in the analysis. Researchers using such a study design with GIS must address unique methodological issues beyond those they might encounter using other epidemiologic study designs (e.g., cohort or case-control studies). First, researchers must be careful in interpreting and conveying the results of an ecologic health study. Although ecologic studies can provide valid information on associations of exposure and outcome as related to a defined exposed group, they do not provide reliable information on individual-level risk. That is, ecologic bias (fallacy) can arise when inferences are drawn about associations at the individual level based on analyses conducted at the group level (22). Moreover, in addition to the usual sources of bias that threaten individual-level analyses, using ecologic analyses to estimate biological effects has an underlying problem: reflecting the heterogeneity of exposure level and covariate levels within groups. This heterogeneity is not fully captured with ecologic data because of missing information on the joint distribution of exposure, disease, and covariates (22). Ecologic fallacy can be easily avoided by not making any assertions regarding individual risks from the results of an ecologic study. The UNC study follows this admonition in that it attempts to determine the association between PM exposures to the community (a geographically defined exposed group) and cardiopulmonary mortality rates as compared with statewide rates for Idaho (another geographically defined group). The only data that will be collected on an individual level will be the mortality data.

The quality of an exposure assessment will determine the validity of an environmental epidemiology study. Furthermore, errors in measurement of exposure can introduce both bias and imprecision into the estimates of health effects (22). The methodological issues related to the analysis techniques used in the ATSDR exposure assessment have been discussed above. Disease misclassifications can also be a major source of bias within any environmental epidemiologic study. The use of GIS in an analysis does add methodological considerations related to disease misclassification beyond those that can be found in other epidemiology studies (e.g., increased disease diagnosis and systematic over-reporting of a disease). For example, within the current air study design, inaccurate address-matching or low address-matching rates, using a GIS or other software package, can provide incorrect or missing classification of a disease case during the point-in-polygon analysis. The conditions for confounding differ for individual-level versus group-level or ecologic analyses, and some types of confounding cannot be controlled in ecologic analyses (22). Even when all variables are accurately measured for all groups, adjustment for extraneous risk factors may not reduce the ecologic bias produced by these risk factors. In fact, it is possible for such ecologic adjustment to increase bias (22). There are two methods currently being evaluated to control for ecologic confounders.

Conclusions

It is apparent that GIS is an excellent platform for bringing together disparate databases and, through spatial analysis, assessing the exposure and demographics of an area.

It must not be assumed, however, that the results of a sophisticated analysis or well-laid-out map actually are valid representations of the world they are trying to model.

Methods of estimating areas of health concern in exposure assessments (e.g., Gaussian air dispersion models), either outside or within a GIS environment, must be used with an understanding of their limitations and the site-specific factors that may affect the validity of their results.

The use of air dispersion modeling in a study of air exposures can help to alleviate some of the data issues related to the fallacy of the homogeneous polygon and the problem of fuzzy boundaries.

Methods of estimating the demographics captured by modeling techniques that define an area of health concern must be used with care so that the problem of areal interpolation and the fallacy of the homogeneous polygon can be alleviated. Analytical techniques like the area proportion method are excellent for some applications; however, if exposure assessment information is to be used in an epidemiologic study of an exposed population, the distribution of population densities within demographic geographic units must be evaluated before one can feel confident that exposure misclassification has been reduced to acceptable levels.

An ecologic study design based on a GIS analysis carries with it unique methodological issues beyond those that may be encountered in other epidemiologic designs. Ecologic fallacy, disease and exposure misclassification, and control for confounding must be carefully considered when designing an ecologic study and in interpreting its results.

The design of ATSDR's exposure assessment and UNC's ecologic health study will be refined to estimate the association of PM exposures with cardiopulmonary mortality with as much validity and precision as an ecologic approach allows.

References

1. Agency for Toxic Substances and Disease Registry. 1995. *Fort Hall air emissions study—Fort Hall Indian Reservation, Fort Hall, Idaho.* Atlanta: US Department of Health and Human Services. November.
2. Bechtel Environmental, Inc. 1996. *Remedial investigation/feasibility study report for the EMF site.* San Francisco: Bechtel Environmental, Inc. August.
3. US Environmental Protection Agency, Office of Air and Radiation and Office of Air Quality Planning and Standards. Web site.
4. Bechtel Environmental, Inc. 1993. *Remedial investigation/feasibility study: 1993 air data interpretation report for the EMF site.* San Francisco: Bechtel Environmental, Inc. January.
5. Idaho Department of Environmental Quality. 1988. *Idaho air quality annual report—1987*. Boise: Idaho Department of Environmental Quality. July.
6. Air Resources Specialists, Inc. 1997. *Draft monitoring and quality assurance plan for the Shoshone-Bannock/EPA particulate monitoring program, Pocatello, Idaho.* Fort Collins, CO: Air Resources Specialists, Inc. January.
7. Personal communication. 1998. Steven Body, US Environmental Protection Agency, Region 10, Seattle. July 8.
8. IDEQ data report. No date.

9. Personal communication. 1998. Tom Edwards, Idaho Department of Environmental Quality, Pocatello, ID. July 8.
10. World Health Organization (WHO). 1977–78. *Manual of the international statistical classification of diseases, injuries, and causes of death: Based on the recommendations of the Ninth Revision Conference, 1975, and adopted by the Twenty-Ninth World Health Assembly*. Geneva: WHO.
11. National Environmental Health Association. Web site.
12. Briggs DJ, Elliott P. 1995. The use of geographical information systems in studies on environment and health. *World Health Statistics Quarterly* 48(2):85–94.
13. Esmen NA, Marsh GM. 1996. Applications and limitations of air dispersion modeling in environmental epidemiology. *Journal of Exposure Analysis and Environmental Epidemiology* 104(4):414–20.
14. Maslia ML, Aral MM, Williams RC, Susten AS, Heitgard JL. 1994. Exposure assessment of population using environmental modeling, demographic analysis, and GIS. *Water Resources Bulletin* 30(6):1025–41.
15. Von Braun M. 1993. The use of GIS in assessing exposure and remedial alternatives at Superfund sites. In: *Environmental modeling with GIS.* Ed. M Goodchild, B Parks, L Steyaert. New York: Oxford Press. 339–47.
16. Koontz MD, Zarus GM, Stunder MJ, Nagda NL. 1991. Air toxics risk assessment. Unpublished paper. Germantown, MD: GEOMET Technologies, Inc.
17. Croner CM, Sperling J, Broome FR. 1996. Geographic information systems (GIS): New perspectives in understanding human health and environmental relationships. *Statistics in Medicine* 15:1961–77.
18. Vine MF, Degnan D, Hanchette C. 1997. Geographical information systems: Their use in environmental epidemiology. *Environmental Health Perspectives* 105(6):598–605.
19. Nuckols JR, Berry JK, Stallones L. 1994. Defining populations potentially exposed to chemical waste mixtures using computer-aided mapping and analysis. In: *Toxicology of chemical mixtures: Case studies, mechanisms, and novel approaches.* Ed. RSH Yang. Ann Arbor, MI: Academic Press. 473–503.
20. Page PH. 1996. *A variable kernel density estimation procedure for generating raster GIS population data layers from US Census data*. Master's degree thesis. Chapel Hill, NC: University of North Carolina at Chapel Hill, Department of Geography.
21. Knapp MB, Archambault GV. 1998. Census control population methodology. Unpublished paper. Connecticut Department of Public Health.
22. Rothman KJ, Greenland S. 1998. *Modern epidemiology.* 2d Ed. Philadelphia: Lippincott-Raven Press.