PUBLIC HEALTH GIS NEWS AND INFORMATION

February 1997 (No. 14)

Dedicated to CDC/ATSDR scientific excellence and advancement in disease control and prevention using GIS

I. Public Health GIS (and related) Events

☞The 1997 National Conference on Health Education and Health Promotion, Atlanta, May 28 - 30, 1997; The conference will be held in conjunction with the SOPHE Mid-Year Scientific Meeting. The conference is sponsored by CDC, the Association of State and Territorial Directors of Health Promotion and Public Health Education, the Society for Public Health Education and the National Heart, Lung and Blood Institute. Please contact John Korn by email directly or by phone at 770-488-5427.

☛ Fifth International Symposium on Spatial Databases (SSD), Berlin, Germany, July 15-18, 1997; Major topics of interest include, but are not limited to: Interoperability and open GIS, Benchmarks and evaluation, Constraint databases for GIS, Spatio-temporal databases, Spatial query processing and optimization, Spatial modeling, Spatial data and the information highways, Spatial data mining, and Applications (e.g., transport networks, urban planning, environmental studies). For additional SSD conference information, see http://www.inf.fu-berlin. de/~ssd97.

☞ 1997 Joint Meeting of the Public Health Conference on Records and Statistics, and the Data Users Conference, Washington, D.C., July 28-31; A session entitled "Geographic Information Systems: An Exploratory Tool for Disease Surveillance and Analysis" will be conducted; contact Barbara Hetzler, NCHS, (301) 436-7122, ext.148, for information.

II. News from GIS USERS

(Please communicate directly with colleagues on any issues)

A. General News

1. **GIS at Prince George's Community College**: A special welcome (from the GIS Users Group) is extended to six faculty members of Prince George's Community College (PGCC), Largo, MD. Pat Cunniff, Director, Science and Technology Resource Center (STRC), has an NSF grant to promote GIS through the initiative Community Colleges for Innovative Technology Transfer, Inc. (CCITT), a coalition of 12 community colleges (including PGCC as the lead), nine local NASA Centers and ESRI, Inc. Each college is aligned with its local NASA Center. The goals are to host faculty development workshops in image processing, remote sensing and GIS, develop an Earth Systems Science course and laboratory, and support development of a Center for Image Processing at Foothill College, CA. For more information, see the CCITT Home Page, http://earth.fhda.edu, or contact Pat by phone at (301) 322-0445 or through email at pc2@pgstumail.pg.cc.md.us. Information about STRC programming, new courses and faculty development activities that are underway is available at http://www.strc_pgcc.pg.cc.md.us.

2. **GIS and Epidemiology in Cuba**: The Institute of Tropical Medicine "Pedro Kouri" (IPK) from Havana, Cuba, is organizing the V Latin American Congress of Tropical Medicine, to take place in Havana, Cuba, March 3-7, 1997, in the headquarters of the IPK. The Organizing Committee is offering a Symposium and a pre-Congress course on the Application of Geographical Information Systems in Epidemiology and Health, one of the themes of the Congress. Contact: Prof. Gustavo Kouri, President of The Organizing Committee V CLAMT, Director Institute of Tropical Medicine Pedro Kouri, Tel.: 53 (7) 215957, 53 (7) 336051, Fax: 53 (7) 336051, or through email at <ciipk@infomed.sld.cu>.

3. **URISA Conference**: 1997 URISA Conference Call for Presentations, "Integrating Information Technology: Acting Locally, Connecting Globally," Metro Toronto Convention Centre, Toronto, Canada, July 20 - 24, 1997; For a complete Call for Presentations application: see the URISA Home Page at www.urisa.org, call (202) 289-1685 or email <training@urisa.org>.

4. **Airborne GPS**: The First-Ever National Symposium on Airborne GPS convenes in Reston, VA, February 6-8, 1997. Sponsors include the US Geological Survey (National Mapping Division) and the Management Association for Private Photogrammetric Surveyors (MAPPS). For further information, contact John Palatiello, MAPPS, at phone (703) 391-2739 or email <johnmapps@aol.com>.

5. From **Laura Montgomery**, NCHS: Laura reports that the paper session "The Use of GIS (Geographic Information Systems) as a Tool for Diverse Analytical Functions in Public Health", 1996 APHA annual meeting, drew an overflow audience. [Editor: The session organizer was Anita Burney, HRSA. Selected abstracts from this session are reproduced in this edition of the newsletter; see section V. Public Health GIS Literature]

6. From **Roger Friedman**, NIP: FYI, our group is working to increase immunization rates in WIC [Women, Infants and Children] program participants. We are using MapInfo to display the status of WIC immunization interventions and immunization rates among immunization grantees. We start with a Microsoft Access database which contains baseline data, grant information, and grantee report data. A query pulls all the data together for mapping; it is exported in Excel format and imported into MapInfo. Thirteen thematic maps are created, three of which are overlays. The layout windows are then exported from MapInfo as Windows Metafiles and imported into Lotus Freelance; then they are ungrouped and touched up. [Editor: readers interested in viewing these map slides can contact Roger by email at <rdf4@nip1.em.cdc.gov> or phone (404) 639-8789]

B. Technical News

7. **Some Map Comparisons**: A report of 49 maps showing the geographic display of census data in transportation planning and policy decisions ("Census Mapbook for Transportation Planning" or CMPT, by GIS/Trans., Ltd. and sponsored by the Federal Highway Administration, December, 1994) additionally offers GIS Users an opportunity to examine products from a variety of commercial mapping products in one setting. These GIS maps include displays using Intergraph (Unix Workstation, Intergraph Workstation), ARC/INFO (Unix Workstation, Sun Workstation, IBM Workstation, HP Workstation, Sun File Server with Tektronix terminal, and Pentium Desktop), MapInfo (Desktop), pcARC/INFO (Desktop), Atlas GIS (Desktop) and TransCAD (Desktop).

The 1990 census data products used mostly by transportation analysts include: (1) summary tape files, (2) microdata, and (3) geographic files. The maps in this report were submitted mostly by Metropolitan Planning Organizations (MPOs) and State Departments of Transportation (DOTs). Geographic units range from counties, to tracts and traffic analysis zones (TAZs), to 4 km grids for air quality analysis, to network links. GIS will be standard software for all transportation planning agencies by year 2000. For more information, call (202) 366-0160 or for copies contact the FHWA R&T Report Center at (703) 285-2144.

8. **Review of Census Statistical Areas**: The Census Bureau is in the process of reviewing the statistical areas known as "census regions" and "census divisions." These provide a two-tiered classification of nine census divisions nested in four census regions. This configuration of nine census divisions has been used in reporting statistics since 1910 (with Alaska and Hawaii incorporated into the Pacific Division in 1960). The four census regions were configured in 1942, with only a change in the name of the Midwest, from North Central, in 1984. Any comments you may have regarding the appropriateness of this configuration, or the respective names, should be brought to the attention of Don Dahmann, Population

Division, U.S. Bureau of the Census, Washington, D.C. 20233 or phone (301) 457-2413.

9. From **Steve Campbell**, Baltimore City Health Department: I did manage to get that census tract map and the zip code map for Baltimore City created, and, as a result of my work with CDC, they're taking a careful look at EpiMap to make sure it works with all digitizers properly. (There may, I repeat MAY, have been some glitches in version 2 that didn't exist in version 1, and that will NOT exist in the next version.) One thing, by the way, that may not be obvious to users of EpiMap (it hasn't been obvious to me, even though Tom Arner at CDC has pointed it out to me) is that EpiMap doesn't respond extraordinarily well to EpiInfo datasets unless you have a (text) variable named REGIONNAME in the dataset that defines the regions of the map you want to use. If you have that, though, EpiMap works like a charm.

10. From **Tom Richards**, PHPPO: I would like to purchase a computer for a GIS project to map the boundaries of US local health departments. I would appreciate any comments that you might have as to the computer specifications you would ideally purchase if you were starting this project, and especially recommendations for hard drive size; pentium MHz; RAM memory; video card and video memory; CDC-ROM drive; color printer; and/or any other thing that you think would be important to include. [Editor: see Section III below for project description and contact information]

11. **New GIS-related collaboration between Census Bureau and GDT**: A new cooperative agreement between the Census Bureau and Geographic Data Technology, Inc (GDT), Lebanon, NH, is designed to create a comprehensive and consolidated public/private partnership approach to update and share information on new streets, street names and address ranges. This will be especially important to the maintenance of the TIGER database for Census 2000 and other key spatial data sets. This partnership is the first of its kind for the Census Bureau. Editor

12. **Latest GDT Developments**: (a) Dynamap/2000^R v.7.0, GDT's premier street network database includes revamped and more complete (and accurate) landmark layers over previous versions. It includes eight landmark layers: airports, railroads, parks, recreational areas, transportational terminals, institutions, major retail centers and large area landmarks. In addition, there is a new major water layer from a high-altitude perspective; (b) The release of Dynamap^R Carrier Routes v.2 identifies the boundaries of more than 300,000 U.S. postal carrier routes, including both small city routes as well as rural routes, ZIP Code boundaries and landmark features, and; (c) A postal databases ZIP+6 Address Coding Guide. A ZIP+6 combines a ZIP Code, ZIP+4, and a Delivery Point Bar Code (the last two digits of an address). Companies can conduct rooftop geocoding without cumbersome address information and users can quickly interpolate coordinate information associated with a ZIP+6 and associate census variables with individual accuracy. [see *GDT News*, Vol.13, No.1 or checkout their Net site at http://www.geographic.com.]

13. **US Census CD**: The entire US Census, including data on population, housing, occupation, education, income, race, language, ethnic background, etc., is available for use with a powerful Windows interface on one CD-ROM. Called CensusCD, it is the only single CD product on the market with the complete 1990 US Census e.g., 3500+ variables for 16 levels of US geography, from block group level and higher. The Windows interface allows for customized user selections where one can create radii around addresses or other georeferenced locations, search for demographic variables, input addresses for neighborhood characteristics, and many other functions. It is export friendly to SAS, a variety of spreadsheets and GIS mapping software packages. [see http://www. censuscd.com or call GeoLytics at 1-800-577-6717]

C. Internet News

(Selected items picked up from the Internet by GIS Users)

14. **S+SpatialStats**: S+SpatialStats is a new add-on

module for S-PLUS, the most powerful data analysis software available today. S-PLUS is based on the object-oriented S Language, developed by AT&T Bell Laboratories specifically for analyzing scientific and technical data. S-PLUS is the statistics tool used by ESRI Inc, the developers of ARC/INFO GIS. With nearly 2,000 built-in functions and the flexibility of a programming language, S-PLUS offers the user an unrivaled, modern, comprehensive environment for data analysis. A categorical list of the datasets and functions included in S+SpatialStats is shown as an ATTACHMENT (see pp. 16-22) to this edition of *Public Health GIS News and Information*. [Source: <owner-ai-geostats@gis.psu.edu>]

15. **New ARC/INFO ArcView LINK to S-PLUS Statistical Software**: We would like to announce S+GeoTrans, a new SPLUS library to import and export Arc/Info/ArcView SHAPE-Files. The main goal for S+GeoTrans was ease of use: the data analyst does not need to know anything about the underlying GIS system and the GIS user does not need to know anything about SPLUS; no learning curve for both SPLUS user and GIS user; support for ALL platforms where SPLUS is available platform independence: SPLUS and the GIS system may be on different systems; and easy access to existing point data for S+SpatialStats users.

To achieve these goals we have chosen a file-based interface. ESRI established the ArcView shape-file format with ArcView 2.x and ArcInfo Version 7.04. A rising number of other GIS systems support shape-files. A shape-file contains three files, two for the geometric data and one dBaseIII file for the attribute data. S+GeoTrans builds a set of SPLUS-functions for reading and writing shape-files and dBaseIII files. Functions:

tget.shape (filename) : read point and attribute data into a data.frame
tget.shape.points (filename) : read only point data into a data.frame
tget.shape.data (filename) : read only attribute data into a data.frame
tget.shape.info (filename) : get info about the contents of a shape-file
tput.shape (data, filename) : write point and attribute data into a new shape-file
tput.shape.points (data, filename) : write only point data into a new shape-file
tput.shape.data (data, filename) : write only attribute data into a new shape-file
tget.dbf (filename) : read a dBaseIII file into a data.frame
tget.dbf.info (filename) : get info about the table definition of the dbaseIII file
tput.dbf (data, filename) : write a data.frame into a new dBaseIII file

Beta versions of S+GeoTrans for SunOS and Solaris are available now as are the ports for SGI, AIX HP. The rest depends on user request; please contact gis@statsci.co.uk for further information. StatSci Europe, Osney House, Mill Street, OXFORD OX2 0JX, U.K., Tel 01865 200952, FAX 01865 200953. [Source: <owner-ai-geostats@gis.psu.edu>]

16. **Basic Geostatistics Textbook**: An informal survey of basic recommended geostatistics textbooks on this Net site resulted in an overwhelming vote for E.H. Isaaks and R.M. Srivastava. *An Introduction to Applied Geostatistics*, Oxford University Press, 1989. [Source: <owner-ai-geostats@gis.psu.edu>]

17. **Classroom Materials Discussion**: Q- I am currently teaching a topics course in spatial statistics at the advanced undergraduate/beginning graduate level at George Mason University. This is the second time that I have taught the course and we are using the book by Bailey and Gatrell as the primary reference along with the S+ spatial package as a compute engine. In addition the students can buy Cressie's book as an optional reference. The focus of the course is primarily data analysis in flavor and hence I am always looking for interesting data sets. The course provides an introductory treatment to point data, lattice data, geostatistical data, and spatial interaction data. I would appreciate pointers from the group on good data sets that are net accessible. One of my areas of interest is in epidemiological data so pointers in that

area will be particularly appreciated.

R (response)- An exercise that I find useful is to take an exhaustive data set, such as a satellite image of NDVI, and pull out a series of random subsets. give these to the class to calculate variograms, fit models, and krige -- but don't tell them what the data are. Then gather the results and compare them against the "truth". Useful lessons are to be gained about sampling, scale, and challenges of nonstationarity. [Source: <owner-ai-geostats@gis.psu.edu>]

18. **Digital Atlas of California**: The Internet Library of the Department of Geography at California State University Northridge, operated by the Department of Geography and the Center for Geographic Research, is developing a new resource for public access termed the Digital Atlas of California. Under the direction of Professor William Bowen, more than 350 maps (current) are online including maps of California, metro LA, San Francisco and Bay area, metro San Diego and metro Sacramento, with archived census data. Most distributions show census tracts by population and race, citizenship, income, poverty and adult educational attainment. Incidentally, it was Bill Bowen who initiated (first lecturer) the NCHS Cartography and GIS Guest Lecture Series on Sept. 23, 1988, Hyattsville, MD, with his presentation "Mapping Applications and Spatial Analysis." [Source for viewing: http//130.166.124.2/ Capage1.html]

III. GIS Outreach

(Editor: All solutions are welcome and will appear in the next edition; please note that the use of trade names and commercial sources that may appear in *Public Health GIS News and Information* is for identification only and does not imply endorsement by CDC or ATSDR)

☎ From **Sonja Hrabowy**, Kalamazoo County Human Services Department: I'm wondering what is the easiest way to learn how to use GIS. I'm aware of the concept, etc. I have experience as an epidemiologist in using a variety of statistical packages. We have purchased MapInfo in my health department and I need some guidance in how to teach myself this product. OR, are there any classes available in SW Michigan, OR are there resources available to me in this area which I am not utilizing to my full capacity? I would appreciate any comments in this matter. I am very excited to become a full-fledged GIS user!! [Editor: Sonja can be reached by email at sonjah@hline.mdph.state.mi.us]

☎ From **Lois Dean**, HUD: HUD is developing guidelines for community groups that want to use GIS to support their missions. These guidelines will accompany the new version of the Consolidated Plan GIS software and data discs HUD expects to have available in the near future. In the Guide we would like to cite "Lists" that nonprofit organizations and citizens groups may wish to join to help them, especially when starting out with GIS. Also, we would like to find and join such lists for information and feedback for the Guide, in order to tailor it to user's needs. The Net has information on GIS, but not much for GIS at the urban scale, or GIS for public participation. Does anyone have information about a list that focuses on GIS for public participation and any type of community issues? Lois_Dean@hud.gov.

☎ From **Tom Richards**, PHPPO: The following provides an update on activities to develop a national map of the boundaries of US local health departments (LHDs). In the October 1996 issue of *Public Health GIS News and Information*, I asked: "Does anyone have experience with mapping local health department jurisdictions?" Based on the many excellent responses received and my knowledge todate, my conclusion is that: 1) maps of LHD boundaries currently exist ; 2) some state and LHDs use GIS software; 3) however, a national boundary file for local public health jurisdictions currently does not appear to exist in GIS format. Therefore, I should proceed to develop a project in this area. The draft project proposal that I am currently circulating for review and comment follows. Any comments or suggestions about the project proposal would be appreciated. Of course, if anyone has already developed and maintains a national boundary file for local health department jurisdictions in GIS format, please let me know.

Project Title:
Monitoring Local Public Health Performance with GIS. A GIS is a powerful tool that combines geography (georeferenced location), computers and spatial statistical analysis. In a GIS, maps and tabular data bases are stored with linked georeferenced identifiers (e.g., latitude and longitude) so that a computer can manipulate, display, and analyze the information they contain. A GIS makes it easy to comprehend large volumes of data and spatially explore patterns, relationships, trends and outliers that would otherwise go unnoticed. Mapping the boundaries of US local health departments (LHDs) is a key step in the development of a surveillance system for monitoring Healthy People 2000 Objective 8.14 (e.g., "increase to at least 90 percent the proportion of people who are served by a local health department that is effectively carrying out the core functions of public health"), and in establishing a national sampling frame for LHD surveys. Although maps of LHD boundaries currently exist, they are not in digital format that allows easy linkage with other georeferenced data sources. Developing a LHD boundary GIS computer software file would facilitate this type of linkage. Examples of how a national GIS of LHDs might be used include the:

o creation and availability of a national sampling frame for LHD surveys;
o identification and characterization of nonparticipants in LHD surveys;
o updating and maintaining (with relative ease) LHD boundary changes that occur over time e.g., from mergers or consolidations;
o timely linkage of information about activities within local public health jurisdictions (from LHD surveys) with demographic, economic, and health indicator data from other sources including the Bureau of Census, US Environmental Protection Agency, and National Center for Health Statistics. In turn, this will facilitate testing geographic-related hypotheses about the effectiveness of local public health performance of the core functions of public health, conducting spatial statistical analyses, and preparing reports about Healthy People 2000 Objective 8.14, and;
o establishment of a current and systematic basis for identifying where public health interventions need to be targeted, and local public health agencies responsible for the interventions.

Methods:
Using a desktop GIS software package, the boundaries of US local public health jurisdictions will be mapped as completely as possible. This LHD GIS boundary file will be prepared in formats to meet requirements for easy access by other commonly used softwares e.g., MapInfo, Maptitude, ArcView GIS, and EpiMap.

References:
1) Croner, C.M. (ed.). *Public Health GIS News and Information*, Office of Research Methodology, National Center for Health Statistics, CDC (email <cm2@nch09a.em.cdc.gov>).
2) Scanlan, B. Geographic Information Systems: Potential applications for Boards of Health (A tool to prevent health decay). *Massachusetts Association of Health Boards Quarterly*. Fall 1995; 13 (4): 5-7.
3) Rushton, G. Improving public health through geographical information systems. An instructional guide to major concepts and their implementation. CD-ROM, Version 1.0, May 1996. University of Iowa, Iowa City, Iowa 52242 (phone: 319-335-0162; email: <gerard-rushton@uiowa.edu>).
4) Ricketts, T.R., Savitz, L.A., Gesler, W.M., Osborne, D.N. (eds.). *Geographic methods for health services research: a focus on the rural-urban continuum*. Lanham, MD: University Press of America. 1994.
5) Lavin, M.R. *Understanding the census: A guide for marketers, planners, grant writers, and other data users*. Kenmore, NY: Epoch Books. 1996.
6) Pickles, J. *Ground truth: The social implications of geographic information systems*. New York, New York: Guilford Press. 1995.

[Editor: Tom can be reached at (770) 488-2544 or email <tbr1@cdc.gov>]

IV. Special Reports

(Submissions are open to all)

❖ **CD-ROM Technology and Latest Developments-** Excerpts from “The Next Bang: SIGCAT ‘97", Jerry McFaul, *BLER* newsletter, Vol. 3, Issue 1, January 1977: There are exciting developments underway in CD technology that are fundamentally changing the ways in which we manage digital information in our organizations. There are many reasons for the excitement these developments have generated, not the least of which is plummeting costs. CD-ROMs are now so inexpensive to produce that complete multimedia encyclopedias that cost hundreds of dollars only a few years ago are now giveaway items at trade shows! 4X CD-ROM drives are now as cheap as floppy disc drives. A newly announced midrange CD jukebox holds 200 discs, includes two 8X drives, and costs under $5,500. CD-Recorders are available for less than $399. And the price of CD-R media, now hovering around $5 per disc, will drop even lower as new manufacturing plants come online during 1997. All of these developments are fueling a wide variety of CD applications that are now becoming so cost effective that they cannot be ignored.

The World Wide Web has opened up a whole new dimension in digital communications. But as exciting as the possibilities are of linking to information sources anywhere in the world, there are definite limitations imposed by the current telecommunications infrastructure. Moving the large files typically associated with high resolution graphics and video clips is still very tedious, if not downright impractical for the vast majority of Internet users. But CD technology has come to the rescue through innovative companies that have created “connected” CD products in conjunction with the growth of the Web. The user of a connected CD can explore all sorts of spectacular multimedia content and then simply click on a button to connect to the Web and get late-breaking information. We’ll soon take virtual tours of everything from vacation resorts to fashion boutiques to college campuses and then simply connect to complete the transaction. These capabilities will radically redefine the classic business models as the exciting world of CD technology continues to push the limits of our collective imaginations.

But wait! There’s more. An even newer development is emerging in which the connected CD is taken one step further and becomes a Web site on a disc. The capability now exists to readily capture hugh chunks of Web content, including all of the hyperlinked pages, and place them on a CD-R or CD-ROM along with a powerful search engine. A user can then peruse an entire Web site offline, with instantaneous response times and the added luxury (some say necessity) of having everything organized and word searchable. In addition, such discs can still be linked with the Internet at the click of a mouse.

Major advances in CD drives, changers, towers, and automated access mechanisms (commonly referred to as “jukeboxes”) have expanded the scope of enterprise information management applications to the level of gigabyte or even terabyte file sizes. Networked access to these vast CD databases via LANS and intranets as well as the World Wide Web is now of prime importance to most chief information officers and information system managers. CD library systems incorporating massive CD drive arrays accessed through networks are permeating information-intensive organizations such as libraries and law firms.

The explosion in jukebox technology is forcing the migration of major databases in the gigabyte and terabyte range from mainframe-based disc farms to networked Cds for reasons of lower costs and higher accessibility. Placing a 600-disc jukebox on a network dramatically expands the number of users who can directly access data. This next generation of CD technology is about to emerge from the world’s major technology companies and will increase the storage capacity of a single CD-ROM disc to over 18 gigabytes. When the new DVD (Digital Versatile Disc) capacity is eventually coupled with some of the larger jukebox systems now coming to market, systems capacities of over 10 terabytes in a single box start to become feasible.

Another development changing the ground rules in information management is the explosive growth in CD-R publishing systems. Many companies now have systems capable of automatically producing hundreds of professional-looking discs overnight. An entire cottage industry of digital publishing is about to be unleashed, driven by the combination of truly affordable CD-R publishing systems and a potential worldwide customer base with CD capable computer systems that should number over half a billion by the year 2000. [**Editor**: Jerry McFaul is a renown expert in this area and the creator of SIGCAT (The Special Interest Group on CD Applications & Technology). For more information on this newsletter or upcoming SIGCAT meetings (e.g., SIGCAT ‘97, May 18-22, 1997, Reston, VA), see http://www.sigcat.org.]

❖ **Spherekit** (an integrated toolkit for the spatial interpolation and comparison of spatial interpolation algorithms): Spherekit was developed at the NCGIA (National Center for Geographic Information and Analysis) and at the University of California, Santa Barbara (UCSB). The initial version has been released into the public domain in October 1996. It is UNIX based and includes a complete graphical user interface (GUI). It uses Generic Mapping Tools (GMT) for display of interpolated fields. The package features several unique capabilities:

Large-scale interpolation- Spherekit permits interpolation over continental or global scales because its computations are based upon spherical distances and orientations. Conventional interpolations are based upon planar projections of the earth, which produce distortions of some kind over large distances. In Spherekit, projections are applied only for display purposes, after the interpolation has been carried out in spherical geometry. The user can select from a wide range of interpolation algorithms and can experiment with any associated parameter settings.

Smart interpolation- Spherekit permits the user to incorporate knowledge or information about the processes that produced the spatial variations. This strategy, also known as "smart interpolation," is implemented through the interpolation of user-defined, derived variables. A simple equation editor is available to produce combinations of observation variables, and several nonlinear transforms are built-in. A digital elevation model (DEM) is included so that elevation can be treated consistently with other variables.

Error analysis- Error analysis is an integrated component of Spherekit. The performance of an interpolation method and its associated set of parameters is evaluated using cross-validation. The error at each observation point is defined as the difference between its actual value and its estimated value using the remaining n-1 points. The resulting error field can be displayed either at the data points or interpolated to a regular grid, to reduce any spatial biases. Error difference fields, comparing a pair of methods, can be easily displayed.

The available Interpolation methods are:
Inverse distance weighting
Triangulation
Kriging
Multiquadric
Thin plate spline

Spherekit is available for download by anonymous ftp from the NCGIA. Check the homepage at http://www.ncgia.ucsb.edu/pubs/spherekit/main.html

❖ **"Geographic Information Systems: Application in Cancer Research"**, November 25, 1996 Workshop, Chantilly, VA, sponsored by the Division of Cancer Epidemiology and Genetics, Extramural Epidemiology and Genetics Program, National Cancer Institute (NCI). Kumi Iwamoto, Program Director, Epidemiology, served as workshop organizer. Kumi is a recent GIS User and can be reached at (301) 496-9600 or email iwamotok@epndce.nci.nih.gov. [**Editor**: The following were my pre-meeting thoughts to address three questions posed to participants of the workshop. I welcome any comments or observations you might wish to share]

<u>What are the chief advantages/strengths of GIS in cancer epidemiology?</u>: The technology (and science) of GIS is especially designed to support and enable

exploratory spatial data analysis and hypothesis generation, at virtually any geographic scale on earth. Through time, GIS should evolve into an important tool in the study of environmental cancer epidemiology. First, digital environmental databases can be used more effectively to discern the relative risks of exposure when tied to case-control cancer epidemiology study designs. Ecologic error in distinguishing real environmental risks from exposure databases can be minimized when incorporated into case-control exposure and outcome measures of individuals. Second, GIS is uniquely qualified to help assess spatial structure and anisotropic departures from stationarity in spatial surfaces. This is important to the surveillance and detection of disease occurrence in both time and space. Cancers with short-term latencies e.g., childhood leukemias, present GIS with opportunities to associate environmental exposures with cancer etiology. Even more extended cancer latencies e.g., lung, have been successfully related through GIS to patterns of surface geology and corresponding radon exposures, as well as to occupational and lifestyle exposures. Third, hypotheses of cancer and community relationships can be developed through databases on population-based socioeconomic status indicators. Discoveries of community barriers to cancer screening and prevention programs e.g., including ethnicity and racial origin, income, health insurance coverage, educational attainment and accessibility to care, are integral to our development of effective intervention strategies. Lastly, and even at the smallest national scales of county, county aggregations, states and regions (trend surfaces), cancer outcomes can be successfully modeled with suspected associative covariates as in the case of malignant melanoma and solar irradiation, lifestyle and occupation, and lead to important etiologic discoveries in followup case-control studies.

<u>What are the primary weaknesses of GIS in cancer epidemiology?</u>: The primary weaknesses associated with GIS in cancer epidemiology appear to be found not in the technology itself but in the availability and completeness of quality documented georeferenced events and measurements, over time. Given the recent migration of GIS to public health, and now to cancer epidemiology, it is understandable that many environmental exposure databases were not planned with GIS capabilities in mind. To conduct GIS analyses of the possible relationships between cancers with long-term latencies e.g., breast, and environmental exposures, the technology requires good historical measurements of environmental and personal exposures. GIS in cancer epidemiology is going to cause us to question the uses and limitations of these databases. For example, exploring areal associations between high black male prostate cancer death rates and possible heavy metal exposures in ground water is prohibited due to the limited sample and coverage of constituents measured in the national water dataset. In time, this may be remedied as the effort to build a national water quality database evolves. In terms of personal exposures, many existing population-based databases have limitations in the documentation of potential confounding factors including prior changes in residential mobility, occupation, lifestyle and diet, and the periodicity and intensity of potential carcinogenic exposures. For long-term cancer latencies, we cannot build robust spatial paradigms of exposure around "address at date of diagnosis" information collected in cancer registries. One encouraging note is that although long-term cancer latencies create uncertainties in modeling past environmental exposures, there are ongoing developments to improve dose reconstruction and dispersion modeling with GIS as in the case of radiologic exposures at Department of Energy installations.

<u>Is GIS worth future research support by NCI?</u>: I believe NCI is positioned to promote the role of GIS and cancer epidemiology research. It will require a broad interdisciplinary approach that can help to define what essential data elements and measurements are needed to make better GIS use of current and future environmental and personal exposure databases. It will require a standardized approach that takes into account cumulative and peak lifetime exposures. Institutional partnerships need to be included that

consider similar experiences in cancer surveillance and case-control programs, such as those in progress at Health Canada, and new activities such as the genetics initiative at NCHS. We need to cultivate our collective wisdom and consider a wide range of potential participants to help delineate and measure the environmental, occupational, genetic and behavioral risks, and plan for the best GIS uses of data collection mechanisms. As we begin to comprehend new uses of geocoding and record linkage, we may anticipate new discoveries in spatial relationships, structure and form pertaining to cancer epidemiology. Given the emerging analytical suite of compatible GIS spatial statistical software, historical and prospective study designs related to geographic location will become more model comprehensive (log linear, hierarchical, contextual, etc.) in terms of space-time events and the associated multiplicity of interactions. I envision a challenging future for GIS in cancer epidemiology research and a timely role for NCI support.

V. Public Health GIS Literature

(This section may include literature citations, abstracts, syntheses, etc., and submissions are open to all)

Selected Abstracts: 124th Annual Meeting of the American Public Health Association, New York, November, 1996

❒ GIS Applications in Environmental Health Research: This presentation demonstrates the application of geographic information systems (GIS) to environmental health research. Spatial databases that are available from government agencies and vendors, including data available over the Internet, and potentially relevant to environmental health research are described. The spatial and cartographic display functions of GIS are reviewed. Database characteristics and GIS functions are illustrated in the context of environmental health research problems conducted to evaluate problems of childhood lead poisoning, electromagnetic fields associated with power transmission lines, and geographic patterns of health outcomes in relation to waste sites where contaminants were measured at levels of health concern. The presentation emphasizes potential sources of error in GIS applications and how these can be addressed in addition to the value of analyzing public health data in its spatial context. Presenter: Ellen Cromley, Associate Professor, Department of Geography, U. of Connecticut, at email <ecromley@uconnvm.uconn.edu>.

❒ Low Birthweight Risk and Services in New York City: A GIS Analysis: Low birthweight is a significant public health problem in many cities of the United States. To better understand and address this problem in New York City, we developed a geographic information system (GIS) that links spatial data on low birthweight with information on women's social and economic responsibilities and the locations of prenatal care services. The GIS is used to model and display spatial variation in low birthweight risk and evaluate the geographical accessibility of prenatal services to women in high-risk areas. Using suitability analysis, we create a multivariate index of low birthweight risk by health area. The index incorporates health and socio-economic indicators, as well as data on prenatal care utilization. Maps of risk scores show the locations of high-risk areas. We use the GIS to assess the accessibility of pre-natal services to those high-risk areas and to identify areas most in need of services. Presenter: Sarah McLafferty (with Barbara Tempalski and Zhiqiang Guan), Associate Professor, Department of Geology and Geography, Hunter College, New York at email <slm@everest.hunter.cuny.edu>.

❑ Developing On-Line Map Library Services for the Public Health Professional: Experiences from the University of Connecticut: Approximately 1,400 libraries receive information from the U.S. government through the Depository Library Program. Depository libraries agree to hold the material and make it accessible and available to the public. Increasingly, this information is provided in electronic format--CD-ROMS, diskettes and more recently through the Internet. Since about 1989, TIGER (Topographically Integrated Encoded and Referencing system) files, developed by the U.S. Bureau of the

Census, have been distributed to depository libraries. As a result of this added resource, many libraries are now able to offer GIS (Geographic Information System) consultation and technical assistance to their clients. The Map and Geographic Information Center (MAGIC) at the University of Connecticut is the primary map and geographic reference center for the State of Connecticut. As such it serves the public health community for demographic data and street files for geocoding and address matching for epidemiological analysis. This presentation will also touch upon a project initiated by the Association of Research Libraries. That project is known as the ARL/GIS project. Two of the key goals of this effort are to develop GIS expertise in the library community and to encourage connections and communications between GIS users and government agencies. Presenter: Patrick McGlamery, Map Librarian, University of Connecticut at email <pmc@lib.uconn.edu>.

Selected abstracts of papers and posters presented at The International Congress on Hazardous Waste: Impact on Human and Ecological Health, Atlanta, June, 1995

❑ GIS Studies of Cancer Around National Priority List Sites, Tim E. Aldrich, Ph.D., M.P.H., Karen Graham, M.S., Kryn Krautheim, David Popson, and Carol Hanchette, M.S., State Center for Health and Environmental Statistics, Raleigh, North Carolina: States are increasingly operating population-based disease registries. With rapid case ascertainment, available population data, and the technology of Geographic Information Systems (GIS), these disease registries provide an immense opportunity for environmental epidemiology studies of health effects of hazardous exposures. Attempts to perform multi-site studies of National Priority List (NPL) sites have often relied on county-level mortality data.

The present study is an attempt to improve this method by using 1) cancer incidence data, 2) geographic levels below the county scale, and 3) cancers for strategic sites so that greater validity of assessment is achieved. This assessment uses incidence data for pediatric cancers because they are relatively rare and are often regarded as appropriate sentinels for hazardous exposures. The discussion will cover the conceptual progression for inferences beginning at relatively superficial geographic scales and progressing to true case mapping for actual proximity studies and exposure characterization. These exploratory methods represent a form of "shoe leather" epidemiology for the computer age. Many of the best elements of deductive reasoning and evaluating progressive spatial evidence for etiologic associations will be discussed. These techniques form the kernel of public health surveillance methods found in population-based disease registries.

❑ Race, Ethnicity, and Poverty Status of Populations Living Near Cement Plants and Commercial Incinerators, Lisa M. Harris, Sara Rasmussen, Paul Balserak, Gary Ballard, Environmental Protection Agency, Washington, D.C. Ried Harvey, ICF Incorporated, Michael Paquette, Tetra Tech, Inc., Fairfax, Virginia, Judy English, Environmental Consulting Services, Burke, Virginia: The Environmental Protection Agency (EPA) completed analyses that identified demographic characteristics of populations near cement plants and commercial hazardous waste incinerators and compared the populations of county and state. The agency focused on the spatial relationship between cement plants and incinerators and minority and low-income populations.

The studies address the proximity of populations to sites but do not describe the actual health status of these populations or how their health might be affected by proximity to these facilities. EPA used digital geographic and demographic data to develop population estimates around the cement plants and incinerators. EPA processed the census geographic and demographic data along with the incinerator and cement plant location using Geographic Information System (GIS) software. The agency generated concentric circles or buffers for six radii (0.5, 1, 2, 3, 4, 5 miles) from the site location and computed population estimations for the six buffer areas. EPA then compared the demographics of the buffers to the demographics of the county and state that the facility was in. EPA would like to present the GIS approach and solicit comments and opinions from participants regarding the methodology and its advantages and disadvantages.

❑ Small-Area Analysis Techniques for Environmental Justice in Minority and Disadvantaged Communities, Emmanuel A. Taylor, M.Sc., Dr.P.H., Karen E. Harris, M.P.H., Rueben C. Warren, D.D.S., Dr.P.H., Centers for Disease Control and Prevention, E.B. Attah, Ph.D., Clark Atlanta University, Atlanta, Georgia: Minority populations suffer disproportionately from preventable morbidity and mortality, regardless of income, education, or geographic locale. However, the health impact of the environment on minority populations has not been adequately characterized. This presentation will explore alternative demographic and epidemiologic models for small-area analysis of the distribution of health and exposure phenomena in defined populations. In 1994, President William J. Clinton signed Executive Order No. 12898, mandating that federal agencies collect, maintain, and analyze information to assess and compare environmental and human health risks borne by populations identified by race, national origin, or income. However, current statistical models that document disproportionate impacts of environmental stressors on these populations are limited. Geographic-information-system (GIS) mapping for quantitative analyses provides only anecdotal information.

The Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry are collaborating with other federal and non-federal agencies and organizations to identify and reduce the disparities of environmental hazards experienced by disadvantaged communities and persons of color in the Mississippi Delta Region. A critical first step is to complete a needs assessment. This needs assessment will be developed around four areas, including an environmental profile of key environmental hazards and a demographic profile of communities near the sources of environmental hazards. To maintain scientific rigor in the environmental justice movement, analytic models are needed for intra-racial and site- referenced indices of disproportionate impact of environmental hazards. The use of GIS mapping coupled with analytic capabilities of the PC software, "Cluster," for public health assessment of disproportionate burden among disadvantaged populations in a defined area will be discussed. The proposed methodologies will yield local-level information needed to plan and evaluate environmental health intervention programs in communities throughout the United States.

❑ Birth Defects, Fetal Loss, and Developmental Disorders in an 11-Year Study of an Herbicide Manufacturing Facility with Dioxin Contamination, Marge A. Brewster, Ph.D., Russell Kirby, Ph.D., Arkansas Reproductive Health Monitoring System (ARHMS) and University of Arkansas for Medical Sciences, and Carole Canino, ARHMS, Arkansas Children s Hospital, Little Rock, Arkansas: Adverse reproductive outcomes (major birth defects, fetal loss, and developmental disorders) have been studied for temporal and geographic relationships to a Superfund facility that previously manufactured Agent Orange (2,3,5-T) and other herbicides and had documented on- and off-site dioxin contamination. Population-based case ascertainment was conducted actively by an established monitoring system, the Arkansas Reproductive Health Monitoring System, in cooperation with ATSDR and the Arkansas Department of Health.

The study area was Pulaski County, Arkansas, and the birth years covered were 1980-1990. Cases were sought from records of all known institutions providing diagnosis or care for this population, including hospitals, diagnostic centers, community programs, and public schools. Cases of birth defects and fetal loss were ascertained for the entire 11-year period and developmental disorders for the last 6 years of this period. Cases were linked to birth and death records for this period. Births and cases were mapped via Geographic Information Systems. The overall and race-specific rates of these outcome categories will be presented for the whole county, the facility community, and for potentially exposed areas. The temporal relationships of rates to the clean-up activities and the search for clusters of adverse events will be described.

❒ A Community-Based Study of Adverse Pregnancy Outcomes Near a Large Hazardous Waste Landfill in California, Martin Kharrazi, Ph.D., Julie Von Behren, M.P.H., Margot Smith, Dr.P.H., Tim Lomas, Michael Armstrong, Rachel Broadwin, M.P.H, Elinor Blake, M.S.W, Bob McLaughlin, Gordon Worstell, Lynn Goldman, M.D., M.P.H., California Department of Health Services, Emeryville, California: Various adverse pregnancy outcomes were investigated in residents living within 3 miles of a large hazardous waste landfill in Southern California to determine whether a dose-response relationship existed to fugitive air emissions from the site. Outcome data derived from 9 years of live birth, fetal death, and infant death records (N=25,216). Exposure measures were based on distance from the site, waste manifests, odor complaints, vinyl chloride monitoring, topography, and meteorologic data.

A geographic information system was used to assign exposure levels to births according to the mother's street address and month of conception. Adverse pregnancy outcomes included fetal, early neonatal, and one year mortality; developmental indices based on birthweight and gestational age; and rates of fertility and multiple births. Numerous potentially confounding factors from vital records and from census block groups were used in conjunction with multiple linear and logistic regression modeling to control for the effect of differing sociodemographic factors across the study area. Rates of adverse outcomes did not increase with decreasing distance from the landfill. However, two findings were noteworthy. During the 4 years with the highest air emissions, the rate of early neonatal death was elevated in an area with a medium level of reported odors (adjusted odds ratio = 3.8, $p<0.05$, total N=175), and mean birthweight and mean gestational age in term--but not low-birthweight--infants were lower in a high odor area (adjusted mean difference = -59 grams and -1.8 days, respectively, both p values <0.05, total N=255) than in a large comparison area without odors (total N=9240). Overall, although no large-scale, dose-response effects were observed around the landfill, small effects could be detected in the residential area with the heaviest odors.

❑ Use of Tumor Registry Data and Geographic Information Systems in a Cancer Cluster Investigation Near a Superfund Site, Diane D. Aye, M.P.H., Gary V. Archambault, Connecticut Department of Health, Deborah M. Dumin, Connecticut Department of Environmental Protection, Hartford, Connecticut, Morris L. Maslia, P.E., Agency for Toxic Substances and Disease Registry, Atlanta, Georgia: Connecticut Tumor Registry data, census data, sampling data, and Geographic Information Systems (GIS) are being used to examine cancer incidence within a Connecticut town from 1968 to 1991. This study was conducted in response to citizen concerns that exposure to environmental contaminants from a local Superfund site was the cause of suspected elevated cancer occurrence. Age-specific rates of cancer incidence and standard incidence ratios are being calculated in relation to estimates of exposure to contaminants released into the public drinking water system and into the air. GIS was used to calculate population estimates, locate the addresses of persons with cancer, and derive the spatial distribution of concentrations of water-system contaminants. These concentrations were obtained from a water quality computation program. Maps have been produced of the estimated distribution of trichloroethylene in the public water system and the air. The population-based Connecticut Tumor Registry provides complete data on cases, however, methodologic difficulties exist in determining reliable population estimates and estimates of environmental contaminant exposure. Use of GIS is anticipated to allow more precise analysis of the relationship between exposure to contaminants and development of disease.

❒ Spatial Analysis of Animal Location Data With Other Data Sources: Investigating the Link Between Exposure and Effect Using a Geographic Information System, William D. Henriques, M.S.P.H., Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, The Institute of Wildlife and Environmental Toxicology (IWET), Clemson University, Pendleton, South Carolina, Dale J. Hoff, MS, Kenneth R. Dixon, Ph.D., and Michael J. Hooper, Ph.D., IWET: In determining the probable cause of health effects from exposure to environmental contaminants, many factors strengthen the link between exposure and effect. One of the most useful new methods for assessing exposure and potential effects is the use of a Geographic Information System (GIS) and data-layering techniques based on a geographic coordinate system.

Spatial data from different sources must each be analyzed for quality, accuracy of locations through ground-truthing techniques, completeness of coverage for the area of interest, and assessment of other environmental factors that may play integral roles in exposure or effects. Predator-prey relationships and the relation of the sentinel species to its habitat are also important considerations that require unique analysis techniques, such as home range analysis. Use of tissue concentrations of potentially adverse toxicants adds additional strength to an investigation, and biomarkers of exposure and effect in the sentinel species have a predictive ability to show subtle effects in a population before histologic effects can be observed. Examples of data-overlay techniques and methods of analysis used at the Rocky Mountain Arsenal National Wildlife Refuge will be presented and discussed. Methods will be presented to determine statistical significance of multivariate data. Supported by NIEHS ES04696.

❒ Using GIS to Estimate Potentially Exposed Populations, Danika Holm, Orkand Corporation, John Mann, P.G., John Crellin, Ph.D., Don Gibeaut, P.E., Agency for Toxic Substances and Disease Registry, Atlanta, Georgia: This poster will illustrate how a geographic information system (GIS) was used to estimate the number of children (72 months of age and younger) living within 100, 300, and 500 meters of lead-contaminated stream deposits at a Superfund site in Utah. The advantages and limitations of GIS methods are also illustrated.

❒ Geographic Information Systems: Applications for ATSDR Public Health Assessments, Jennifer C. Kertanis, M.P.H., Diane D. Aye, M.P.H., Connecticut Department of Health, Hartford, Connecticut, Greg Ulirsch, M.S., Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, Marsha D. Henley, and Mark A. Tuttle, M.S., Oak Ridge National Laboratory, Oak Ridge, Tennessee: Applications of Geographic Information Systems (GIS) have helped in the integration, analysis, and visualization of three primary sources of data used in ATSDR public health assessments of hazardous waste sites: environmental data, community health concerns, and health outcome data. As part of a demonstration project, GIS applications have been used in the public health assessment process in Stratford, Connecticut. Waste material including lead, asbestos, polychlorinated biphenyls, and dioxin were used as fill material throughout the town of Stratford. Surface soil contamination has been identified at more than 15 commercial and public sites and at many residential properties.

A public health assessment is defined as the evaluation of data and information on the release of hazardous substances into the environment to assess any current or future impact on public health and identify actions needed to evaluate and mitigate or prevent human health effects. A relational evaluation of the data is made easier through the application of GIS and provides for a more comprehensive database from which conclusions and recommendations can be made. Maps have been produced using GIS that (1) Incorporate environmental sampling data to depict the location of waste in the community, (2) Use 1990 census data to display demographic data in relation to the waste, and (3) Help evaluate the effectiveness of a voluntary lead screening program in the community.

Selected Publications

Abstract: Thioulouse, J., Chessel, D. and Champely, S. (1995). Multivariate analysis of spatial patterns: a unified approach to local and global structures, *Enviromental and Ecological Statistics*, 2, 1-14. We propose a new approach to the multivariate analysis of data sets with known sampling site spatial positions. A between-sites neighbouring relationship must be derived from site positions and this relationship is introduced into the multivariate analyses through neighbouring weights (number of neighbours at each site) and through the matrix of the neighbouring graph. Eigenvector analysis methods (e.g., principal component analysis, correspondance analysis) can then be used to detect total, local and global structures. The introduction of the ***D***-centring (centring with respect to the neighbouring weights) allows us to write a total variance decomposition into local and global components, and to propose a unified view of several methods. After a brief review of the matrix approach to this problem, we present the results obtained on both simulated and real data sets, showing how spatial structure can be detected and analysed. Freely available computer programs to perform computations and graphical displays are proposed. Keywords: Correspondance analysis; Geary's index; global structure; local structure; Moran's index; neighbouring relationship; principal component analysis; spatial correlation analysis; spatial ordination.

New Journal: *Journal of Environment, Disease and Health Care Planning*, Akhtar, R. and Izhar, N. (eds.). The journal is dedicated to advancing knowledge on spatial perspectives of disease and health care planning, particularly in the Third World. The contents encompass an environmental and behavioral framework of disease theory as well as empirical discussions on health care planning. Biannual, first issue released May, 1996. For information, contact Prof. Rais Akhtar, Chair, Department of Geography, University of Kashmir, Srinagar-190006 (J&K), India.

New Journal: *Earth Interactions*. Focusing on research results in the interdisciplinary Earth system sciences, this electronic (structured hypertext) journal contains scientific articles dealing with the interactions between the lithosphere, hydrosphere, atmosphere, and biosphere in the context of global issues and global change. It is published jointly by the American Geophysical Union, and the Association of American Geographers with support from NASA's Mission to Planet Earth, and input from the Ecological Society of America and The Oceanography Society. [see http://EarthInteractions.org or SoftQuad's home page at http://www.sq.com/products/panorama/panor-fe.htm]

New Book: Arlinghaus, SL, Griffith, DA, Arlinghaus, WC, Drake, WD and Nystuen, JD (eds.), 1996. *Practical Handbook of Spatial Statistics*. CRC Press, New York: Chapter 1- Introduction: The Need for Spatial Statistics; Chapter 2- Visualization of Spatial Dependence: An Elementary View of Spatial Autocorrelation; Chapter 3- Spatial Sampling; Chapter 4- Some Guidelines for Specifying the Geographic Weights Matrix Contained in Spatial Statistical Models; Chapter 5- Aggregation Effects in Geo-referenced Data; Chapter 6- Implementing Spatial Statistics on Parallel Computers; Chapter 7- Spatial Statistics and GIS Applied to Internal Migration in Rwanda, Central Africa; Chapter 8- Spatial Statistical Modeling of Regional Fertility Rates: A Case Study of He-Nan Province, China; Chapter 9- Spatial Statistical/ Econometric Versions of Simple Urban Population Density Models; and, Chapter 10- Spatial Statistics for Analysis of Variance of Agronomic Field Trials. Excerpts from the Introduction:
"...Generally speaking, spatial statistics is concerned with the statistical analysis of geo-referenced data;
...*Spatial statistics differs from classical statistics in that the observations analyzed are not independent; this single assumption violation is the crux of the difference*; and,
...observations are correlated strictly due to their relative locational positions (referred to as spatial autocorrelation), resulting in spill-over of information from one location to another (locational information). This spill-over causes redundant information to be present in data values. The redundancy increases as the degree of locational dependence increases. This duplication of information produces complications in the statistical analysis of geo-referenced data that remain dormant in the statistical analysis of traditional data composed of independent observations. That is, invoking an assumption of independent observations suppresses potential data complexities. The net result is that classical statistics applied to georeferenced data fail to capture locational information, raising questions

of estimator sufficiency, bias, efficiency, and consistency. These four cardinal statistical properties might also be coupled with two others: robustness and minimum variance. Geo-referenced data are highly complex, with spatial dependence introducing further complications."

Net Site of Interest for this Edition: http://atsdr1.atsdr.cdc.gov:8080/cx.html

Final Thought(s)

Superfund site Munisport landfill in North Miama, Florida has exposure pathways that include air, contaminated soil, contaminated surface water, and ingestion of fish and oysters from Biscayne Bay. Among contaminants of concern are arsenic, ammonia, benzene, barium, cadmium, carbon disulfide, chromium, chlordane, dieldrin, lead, manganese, PCBs, strontium, and styrene. The City of N. Miami intends to build a rock concert amphitheater on the dump for 25,000 attendees. [see Net Site of Interest for this Edition]

Did you know the major predictors of blood lead levels for children living in Mexico City were the lead content of the glazed ceramic used to prepare children's food, exposure to airborne lead from vehicular emissions, and the lead content of dirt from children's hand? [see Net Site of Interest for this Edition]

And a reminder: Stay current with the Federal Geographic Data Committee's (FGDC) efforts to develop consistency and standards for digital geospatial data in building the National Spatial Data Infrastructure (NSDI). NSDI is a key initiative to "reinventing government" as identified by Vice President Al Gore. FGDC standards will be mandatory for federal agencies. [see http://www.fgdc.gov/swg/swg.html]

Charles M. Croner, Ph.D., Editor, ***PUBLIC HEALTH GIS NEWS AND INFORMATION***, Office of Research and Methodology, National Center for Health Statistics <cmc2@nch09a.em.cdc.gov>

Brave the winter of '97 ... and stay in GIS touch

ATTACHMENT: Complete List of S+SpatialStats Functions

The listing below only includes functions from S+SpatialStats, a new add-on module for S-PLUS. S-PLUS contains a vast array of additional functions which may also be used to analyze spatial data. S-Plus is the statistics tool used by ESRI Inc, the developers of ARC/INFO GIS. Documentation for the following datasets and functions is available via the on-line help files in S+SpatialStats.

Add to Existing Plot	
hex.legend	Add a Legend Hexagonal Lattice Plot
hexagons	Add Hexagonal Cells to Plot of "hexbin" Object
identify.hexbin	Identify Points On a Hexagonal Binned Plot
identify.vgram.cloud	Identify Points On a Variogram Cloud Plot
Computations Related to Plotting	
boxplot.vgram.cloud	Boxplots of a Variogram Cloud Object
Datasets	
aquifer	Wolfcamp Aquifer Data
bramble Bramble	Cane Data
coal.ash	Coal Ash Data
iron.ore	Iron Ore Data
lansing	Lansing Woods Tree Data
quakes.bay	Bay Area Earthquakes
quakes.wash	Washington State Earthquakes In 1980
scallops	Scallop Abundance Data
sids	Sudden Infant Death Syndrome Data 1974-1978
sids.neighbor	Neighbors for Sudden Infant Death Syndrome Data
sids2	Sudden Infant Death Syndrome Data 1979-1984
wheat	Wheat Grain and Straw Yield
Geostatistical Data Analysis	
anisotropy.plot	Explore Corrections for Geometric Anisotropy
boxplot.vgram.cloud	Boxplots of a Variogram Cloud Object

correlogram	Empirical Correlogram and Covariogram
covariogram	Empirical Correlogram and Covariogram
exp.cov Theoretical	Distance Based Covariance Functions
exp.vgram	Theoretical Variogram Functions
gauss.cov	Theoretical Distance Based Covariance Functions
gauss.vgram	Theoretical Variogram Functions
identify.vgram.cloud	Identify Points On a Variogram Cloud Plot
krige	Ordinary and Universal Kriging
linear.vgram	Theoretical Variogram Functions
loc	Correct Spatial Locations for Geometric Anisotropy
model.variogram	Display a Variogram Object and Theoretical Model
plot.correlogram	Plot a Variogram, Covariogram or Correlogram
plot.covariogram	Plot a Variogram, Covariogram or Correlogram
plot.variogram	Plot a Variogram, Covariogram or Correlogram
power.vgram	Theoretical Variogram Functions
predict.krige	Ordinary and Universal Kriging Prediction
rfsim	Simulate a Random Field Process
spher.cov	Theoretical Distance Based Covariance Functions
spher.vgram	Theoretical Variogram Functions
twoway.formula	Fit of a Two-Way Table (Formula Method)
variogram	Empirical Variogram
variogram.cloud	Calculate Variogram Cloud
Hexagonal Binning	
cell2xy	Compute x,`y' Coordinates From Hexagon Cell Ids
erode.hexbin	Erode a Hexagonally Binned Image
hex.legend	Add a Legend Hexagonal Lattice Plot

hexagons	Add Hexagonal Cells to Plot of "hexbin" Object
hexbin	Bivariate Binning into Hexagonal Cells
identify.hexbin	Identify Points On a Hexagonal Binned Plot
plot.hexbin	Plot A Hexagonal Lattice
rayplot	Adds Rays with Optional Confidence Arcs (Sectors)
smooth.hexbin	Hexagonal Bin Smoothing
summary.hexbin	Summary Method for a Hexagonally Binned Object
xy2cell	Compute Hexagon Cell Id's From x and `y'
High-Level Plots	
anisotropy.plot	Explore Corrections For Geometric Anisotropy
boxplot.vgram.cloud	Boxplots of a Variogram Cloud Object
plot.correlogram	Plot a Variogram, Covariogram or Correlogram
plot.covariogram	Plot a Variogram, Covariogram or Correlogram
plot.hexbin	Plot A Hexagonal Lattice
plot.spp	Plots an Object of class spp
plot.variogram	Plot a Variogram, Covariogram or Correlogram
plot.vgram.cloud	Plot a Variogram Cloud
rayplot	Adds Rays with Optional Confidence Arcs (Sectors)
scaled.plot	Equal Scales Plot
Input/Output--Files	
read.geoeas	Read A GEO-EAS Data File
read.neighbor	Read ASCII Files Containing Spatial Contiguity
Informat	
ion	
write.geoeas	Write A GEO-EAS Data File
Interacting with Plots	

identify.hexbin	Identify Points On a Hexagonal Binned Plot
identify.vgram.cloud	Identify Points On a Variogram Cloud Plot
Lattice Data Analysis	
CAR	Conditional Spatial Autoregression Object
MA	Moving Average Spatial Regression Object
SAR	Simultaneous Spatial Autoregression Object
check.islands	Detect Isolated Spatial Regions
cov.family.object	Spatial Family Object
find.neighbor	Find the Nearest Neighbors of a Point
lrt.slm	Likelihood Ratio Test for Spatial Linear Models
neighbor.grid	Creates a "spatial.neighbor" Object for a Grid
print.slm	Use print() on a `slm' Object
print.spatial.cor	Use print() on a `spatial.cor' Object
print.spatial.neighbor	Print a spatial.neighbor Object
print.summary.slm	Use print() on a `summary.slm' Object
quad.tree	Order a Multicolumn Real Matrix into a Quad Tree
read.neighbor	Read ASCII Files Containing Spatial Contiguity
Informat	
ion	
slm	Fit a Spatial Linear Regression Model
slm.fit.spatial	Fit a Spatial Linear Model (Generalized Least Squares)
slm.nlminb	Fit a Profile Likelihood in a Spatial Regression
Model	
spatial.cg.solve	Solve S b = x
spatial.condense	Remove Redundancy in "spatial.neighbor" Objects
spatial.cor	Measures of Spatial Correlation

spatial.determinant	Compute Sparse Matrix Determinant
spatial.multiply	Compute Sparse Matrix Vector Product A x
spatial.neighbor	Create a "spatial.neighbor" Object
spatial.neighbor.object	Class "spatial.neighbor"
spatial.solve	Solve S b = x
spatial.subset	Subset an Object of Class "spatial.neighbor"
spatial.sum	Sum of Two Objects of Class "spatial.neighbor"
spatial.weights	Compute a Spatial Weight Matrix
summary.slm	Summary Method for Spatial Linear Models
Multivariate Techniques	
find.neighbor	Find the Nearest Neighbors of a Point
quad.tree	Order a Multicolumn Real Matrix into a Quad Tree
twoway	Fit of a Two-Way Table
twoway.default	Fit of a Two-Way Table
Point Pattern Analysis	
Fhat	EDF of Origin-to-Point Nearest Neighbor Distances
Ghat	EDF of Point-to-Point Nearest Neighbor Distances
Kenv	Compute Simulations of Khat
Khat	Ripley's K Function for a Spatial Point Pattern
Object	
Lenv	Compute Simulations of Khat
Lhat	Ripley's K Function for a Spatial Point Pattern
Object	
as.spp	Spatial Point Pattern Objects
bbox	Bounding Box for a Spatial Point Pattern
intensity	Estimate the Intensity of a Spatial Point Pattern

is.spp	Spatial Point Pattern Objects
kern2d	Kernel Smoothing of a Two-Dimensional Density
make.pattern	Generate a Spatial Point Process
plot.spp	Plots an Object of class spp
points.in.poly	Find Points Inside a Given Polygon
poly.area	Computes the Area of a Polygon
poly.grid	Generate a Grid Inside a Given Polygonal Boundary
scaled.plot	Equal Scales Plot
spp	Spatial Point Pattern Objects
summary.spp	Summary Method for a Spatial Point Pattern Object
triangulate	Delaunay's Triangulation
Printing	
print.slm	Use print() on a `slm' Object
print.spatial.cor	Use print() on a `spatial.cor' Object
print.spatial.neighbor	Print a spatial.neighbor Object
print.summary.slm	Use print() on a `summary.slm' Object
Regression	
slm	Fit a Spatial Linear Regression Model
summary.slm	Summary Method for Spatial Linear Models
Robust/Resistant Techniques	
twoway	Fit of a Two-Way Table
twoway.default	Fit of a Two-Way Table
twoway.formula	Fit of a Two-Way Table (Formula Method)
Spatial Regression	
CAR	Conditional Spatial Autoregression Object
MA	Moving Average Spatial Regression Object
SAR	Simultaneous Spatial Autoregression Object

cov.family.object	Spatial Family Object
lrt.slm	Likelihood Ratio Test for Spatial Linear Models
print.slm	Use print() on a `slm' Object
slm	Fit a Spatial Linear Regression Model
slm.fit	Fitting Function for Spatial Linear Models
slm.fit.spatial	Fit a Spatial Linear Model (Generalized Least Squares)
slm.nlminb	Fit a Profile Likelihood in a Spatial Regression
Model	
summary.slm	Summary Method for Spatial Linear Models
Statistical Inference	
verbatimcmd	
lrt	Computes a Likelihood Ratio Test
lrt.slm	Likelihood Ratio Test for Spatial Linear Models

Source: StatSci Europe (GIS & Training Departments), Osney House, Mill Street, Oxford, OX2 OJX, UK or email gis@statsci.co.uk.