DYNAMIC REPRESENTATIONS AND ANALYSES OF PROBABILISTIC BOUNDARIES BASED ON DENSITY ESTIMATION FROM POINT-BASED FREQUENCY DATA

Jay LEE
Department of Geography
Kent State University
Kent, Ohio 44242-0001, USA
Tel (216) 672-3222

ABSTRACT

Depending on the nature of geographic phenomena, cartographers need to apply proper techniques to portray the spatial patterns and processes implied in the geographic information being processed and analyzed. In the past, cartographers have been constrained to a limited set of tools for handling various types of geographic information. This restraint, however, is less and less so with the progress of technologies in computer and in geographic information systems (GIS).

In this paper, we discuss an interactive approach to dynamically analyzing and mapping point-based frequency data *through* the technique of density estimation. The point frequency data are processed to form a density surface which in turn is analyzed to construct probabilistic boundaries among the original point data. In this paper, we apply this approach to analyzing trade areas of shopping centers so that we can obtain trade areas of each shopping center and the probabilistic boundaries between them. This paper contributes to cartography by offering a new dimension for dynamic cartographic analysis of geographic data and by offering a new focus on the nature of probabilistic boundaries.

INTRODUCTION

With the increasing availability of digital data and GIS technology, cartographers are now able to process and analyze geographic information in ways that were not possible in the past. The example discussed in this paper is a method that transforms point-based frequency data into density surfaces to provide the necessary information for detecting and describing spatial patterns among the studied point data. The method is based on estimating probability functions of point-based frequency data. Our implementation of this method is able to delineate regions among point-based data with probabilistic boundaries.

Cartographers have long been converting geographic information from one class to another. The most obvious example is interpolating point-based elevation data into a topographic surface. Other examples should be easy to find without much difficulty. In this paper, we focus on delineating trade areas among shopping centers by converting locations of potential customer locations into density surfaces of shopping potential. Because the way customers choosing shopping center is more probabilistic than fuzzy in nature, we refer to the resulting boundaries as probabilistic boundaries.

DENSITY ESTIMATION

To allow generalization of results to other applications, we define potential customer locations as **demand points** and locations of shopping centers as **service centers**. Our approach of converting point data into density surface is *through* the method of density estimation. Previous studies using density estimation to process geographic data can be found in Gatrell (3), Brunsdon (2), and Bithell (1). For detailed technical treatments of the method of density estimation, see Silverman (4) or Scott (5).

Probability density function (`pdf`) in inferential statistics is an efficient tool to describe a numerical distribution. Expanding to three dimensional space, it is directly applicable to processing point-based frequency geographic data. Given a point location with frequency measure of a geographic phenomenon, `pdf` gives how likely the surrounding areas are to have the phenomena happened, or the frequency of the phenomena. The density estimation process can be explained best by starting with one dimensional kernel method. Following the notations in Silverman (1986), a kernel estimator may be:

$$f(x) = \frac{1}{nh}\sum_{i} K\left(\frac{X - X_i}{h}\right)$$

where X is the variable of interest and X_i is an observed frequency of X for location i. The number of points is n and the estimating parameter h is a value that controls the amount of smoothing of pdf. Figure 1 gives an example of kernel estimation. Notice that the pdf is smoothed to different degrees by different kernel sizes. Expanding to three dimensional space, Figure 2 gives a graphic description of a kernel estimator.

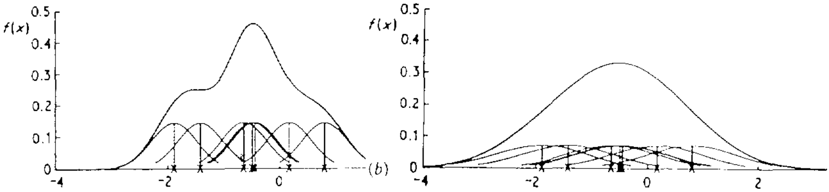


Figure 1: Probability Density Functions and Kernel Sizes

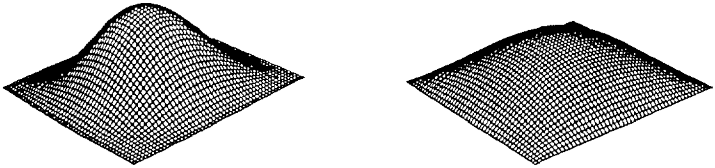


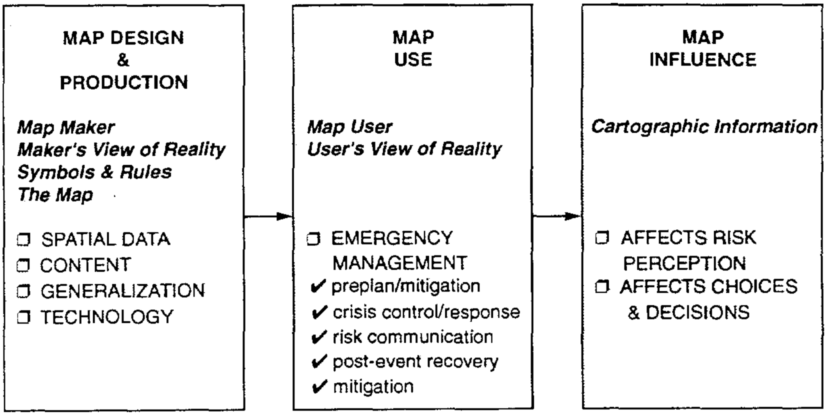
Figure 2: Kernel in Three Dimension

IMPLEMENTATION

When applying density estimation method, one of the most difficult task is to select a proper kernel in terms of its size and shape because the kernel defines how dense the estimated frequencies at each location will be and how smooth the resulting `pdf` will be. In our implementation, we use a normal distribution cone (as in Figure 2) as our kernel shape. This cone has the highest frequency at the center and lower frequencies when moving away from the center, implying the distance decay effect that is widely assumed in many geographic studies. For defining

1.2 Map Communication in the Hazard Management Context

Maps store information, but in terms of map communication and hazard management they do much more. Throughout the hazard management process, three ongoing cartographic processes solve the predicament of securing reliable models of hazardous situations in order to form a spatial basis for informed decision making. Processes of 1) *map design and production*, 2) *map use* and 3) *map influence* transpire. Figure 2 depicts the components of this linear model and illustrates how the concept of cartographic communication, in an expanded fashion, lies within these mapping processes. The seven theoretical elements of Koláčný's map communication model [3] are integrated into Figure 2 in its upper half while the lower portions show items which apply this flow diagram to one specific activity in hazard management, emergency mapping.



- Nancy L. Winter 1995

Figure 2: Three Cartographic Processes in Societal Hazard Management

2 HAZARD CONTROL

Private individuals are the major managers of the overwhelming majority of hazards in everyday life, as opposed to the collective, societal handling of hazards. Among the myriad hazards handled by individuals, few compel people to use maps. Two exceptions are the extremes of hazardous weather and the frustrations of commuter traffic that spawn the ubiquitous daily newspaper and television weather maps and the traffic maps on television. Weather maps are crude risk maps about the probability of natural atmospheric hazards occurring. M. Monmonier [4] claims, "maps also help viewers understand the processes and uncertainty of assessing risk associated with meteorological hazards." Traffic maps are hazard locator maps that help the viewer cope with the effects of the technological hazards of vehicle density on roads of sometimes questionable design.

2.1 Hazard Managers and Monitors

Modern hazard management ensues when society deliberately acts to gain information about hazards, judges what to do about them, and supports programs to control or mitigate their consequences. Participants in this process operate as *managers* or *monitors*. With private **individuals** as the principal managers, four others who function in this role are **technology sponsors** from business and industry, government **officials and policy makers**, **regulators**, and **assessors** such as technical experts and consultants from academia. Hazard monitors include **adversarial groups**, **media** and **regulators**. The availability of appropriate maps is vital to inform these seven major groups.

2.2 Intelligence and Control Functions Require Map Percipients

The first essential activity of hazard managers and monitors is to acquire *intelligence*, to assemble information to aid in ascertaining what problems exist, judging which ones to work on, and evaluating how effective hazard management programs become. Spatial information gleaned from maps is a key ingredient in these intelligence efforts. Hazard managers and monitors must become involved with maps in all the forms described by Robinson and Petchenik [5], not only as map readers looking for a single, specific item or as map users with a specific purpose, but especially as map percipients able to add to their spatial knowledge for geographical understanding of a given hazard situation or situations. Not only the informational details, but the heightened awareness of positional relationships that cartographic aids can grant these decision makers provides them with indispensable insight into the "big picture." *Control* is the second basic activity of hazard managers and monitors, and it calls for creation of methods to prevent, reduce or redistribute technological hazards and/or to mitigate the consequences of natural and technological hazards. Maps aid in the distribution of information, goods and people needed to bring about these measures.

3 HAZARD MANAGEMENT MAPPING MODEL

A model of hazard assessment by R. W. Kates [6, p. 252] furnishes the basis for a Hazard Management Mapping Model. At its hazard management core are processes of 1) *hazard identification*, 2) *risk estimation*, and 3) *resource allocation and evaluation*. Hazard identification and risk estimation describe the methods society employs to become alerted to modern hazards. Resource allocation and evaluation requires society to resolve what amount of its resources will be spent on given hazards and to substantiate the success of efforts to manage them. The rudiments of each of these hazard management activities call for specific types of maps.

3.1 Hazard Maps

Hazard maps locate threats and are indispensable to the four approaches to *hazard identification*: **research, screening, monitoring** and **diagnosis**. The spatial power of hazard maps was illustrated by the now famous 1855 medical map of cholera hazard in London, revealed when Dr. John Snow mapped the location of cholera deaths and public water pump sites and induced the threat posed by the Broad Street pump. By putting data in a spatial context, patterns can be identified which are not clear otherwise.

Research by modern hazard managers often finds answers from mapping and analysis of spatial distributions. Recent research on the impacts of comets and asteroids on Earth, collectively called bolides, has caused conjecture in the media and has raised public concern about a "new" natural hazard, but one ironically older than all the rest. By mapping the craters at bolide impact sites on Earth, an intriguing revelation supports the controversial Celestial Reference Frame Hypothesis [7]. Its premise is that for at least the last half billion years bolides from space have not crashed into Earth randomly, but along certain predictable circular paths. This theory posits that extradense lumps of mass in Earth's interior pull incoming bolides toward these particular orbits meaning Earth itself guides the landings of bolides threatening it. Figure 3 shows the paths revealed by the patterns of bolide craters on Earth. On the left map, the right halo connects craters 50 to 100 million years old. (On a globe, this line would form a circle.) Circled stars on the left halo mark "cratering nodes", hit often by bolides during of loss of property must be conveyed to the president. *Damage assessment maps*, designed and produced under great time pressure, result from this legal requirement. The availability of local tax assessment maps are a crucial ingredient in composing damage assessment maps. Combining data from local, state and federal efforts to gather information by photo interpretation from overflights or from ground surveys conducted by emergency workers is vital to establish the full extent of damage. Later in the disaster's story these maps are cross-checked when insurance claims come in, and thus damage assessment maps gain even greater economic importance.

Post-event human needs and wants create the necessity for the third major type of emergency mapping, *response maps*. Mapping to meet immediate human needs is facilitated by updating some planning maps such as adding Red Cross shelters to a shelter map and medical aid stations to a medical facilities map. New maps must be plotted for a multitude of kinds of response information, such as food kitchens, water and toilet facilities, tent cities, trash sites, burn sites, FEMA Disaster Application Centers (DACs), and eventually, unemployment centers. The availability of base maps of the local area speeds the creation of these response maps. For certain disasters when all visible landmarks and road signs are destroyed, such as happened after Hurricane Andrew in Florida, simple road maps become invaluable response maps. The critical importance of these response maps of seemingly simple content was highlighted in the recent Kobe earthquake in the Hanshin economic region of Japan. Large amounts of emergency aid supplies arrived in Kobe, but could not be distributed appropriately because no response map existed to show the locations of Red Cross shelters.[9]

In general, emergency maps have at least six potential functions. They serve:

- as the quickest means for locating at a glance all factors in a given geographic area (without having to read large volumes of information),
- to coordinate emergency groups through use of a generally accepted, concrete model of the area impacted by a hazardous event correlated with the key elements of emergency planning,
- to synthesize data from all levels of government,
- to aid in control of the physical agent, if possible,
- to aid in the flow of resources and services, and
- to provide information for possible action by the public.

3.5 Mapping as Risk Communication

The three kinds of maps in the Hazard Management Mapping Model - hazard, risk, and emergency - each has significant and unique power to communicate risk. However, as discussed in the risk map portion of this paper, a complex set of factors influences whether maps convey unintended messages about risk that do not fit with the facts of a case. Investigation of what role maps play in shaping risk perceptions in the general public and especially how maps are employed by the media are neglected topics in cartographic research. Any study of map influence, the third cartographic process shown in Figure 2, especially as applied to the general populace, presents a daunting challenge in creating appropriate cartographic research design.

4 TECHNOLOGY IN THE FUTURE OF HAZARD MANAGEMENT MAPPING

Advances in modern information technology offer a vision of improvements in all phases of hazard management mapping. Microcomputers, digital scanners, satellites, fiber-optics, lasers, geographic information systems (GIS), global positioning systems (GPS) and other rapidly developing communication and cartographic technologies have opened up a heady new horizon. The possibilities of accessing hazard data almost instantaneously over Internet during extreme events, of mapping in real-time or near real-time, and of monitoring natural and technological conditions and effects in awesome detail are seen within reach. Before the promise is realized, however, a eady existing technology systems need to be more firmly established in support of present day carto phic work and integrated more effectively and completely for achievement of hazard manageme .sks such as monitoring, prediction, warning, evacuation and emergency response. U.S. governm agencies still struggle to keep pace and to more fully upgrade and consolidate their hazard ma ient mapping programs. Examples of integrative hazard management and cartographic goals still achieved are illustrated by case studies of failures in hazard management. One such volcanic ter struck Colombia on November 13, 1985 before effective risk communication in the form of w ig messages to the public about the impending eruption of Nevado del Ruiz Volcano and its dea nudflows following was accomplished. Recommendations to evacuate were announced on Septe r 13th, by October 7th a hazard map was issued, but a volcanology team's later report questi ing the reliability of the monitoring of changes in the volcano was contradictory. When the mountain erupted at 9:30 at night, enough of its ice cap melted to send killer mudflows down the Rio Lagunillas toward the town of Armero. Within 15 minutes of the eruption, efforts to warn the town of the coming mudflows via civil defense radio failed, 22,000 people died and half of the 10,000 injuries which resulted were serious [10]. Lifesaving hazard information had been mapped and made public, but the system of communication between scientists and local civil authorities was incomplete and inadequate. Also, the system of radio technology failed. Evaluating these sorts of hazard management failures should help set up conditions for more effective implementation of communication systems and newer technologies.

4.1 Effects of GIS Applications

The effects of applications of geographic information systems (GIS) reverberate through the three cartographic processes embedded in hazard management (Figure 2). In terms of map design and production, grander and more varied amounts of data can be incorporated in speedier fashion than through traditional methods. This widens the capabilities of hazard managers to view problems in a more holistic way. The depth of analysis and synthesis GIS affords the decision maker bodes well for production of more valuable hazard and risk maps. Especially enticing for emergency managers are the advantages GIS use offers for on-line base maps which can speedily be converted into crisis maps, damage assessment maps and response maps in disaster response.

Map use for monitoring and predicting socioenvironmental conditions is strengthened and increased through applications of GIS to problem solving. Paired with other technologies, such as GPS or remote sensing from satellites, GIS brings cartographers and hazard managers much closer to real-time mapping of hazardous conditions. An example, is the way Walsh *et al.* [11] incorporated into a GIS a comparison of data sets employing spatial statistics to turn remotely sensed images into overlays to be included in the analysis. This GIS was used to monitor and predict the paths of snow-avalanches in Glacier National Park, Montana. A risk map depicting high, medium and low avalanche probabilities was the end product. However, the elusive goal of integrating remotely sensed data into a GIS for real-time mapping is still to be realized.

GIS applications prompt a fundamental change in the customary view of cartographic communication. The interactive exploratory aspects of GIS software shatters the traditional concept of a cartographer designing a map for a specific interpretation or purpose and usually with a certain map user in mind. The GIS user typically aspires to solve a many-faceted problem. A rich assortment of data layers and varying analytic approaches allows the GIS user to make and examine a number of maps derived through a GIS program. The outcome is that in the process of exploring data to answer problematic questions map makers become map users, or more accurately, map makers become map percipients, capable of deriving rich geographical understanding from the GIS maps they make. A new dynamism and flexibility is gained in this mapping and modelling. Gone are the traditional cartographers' relationships to map users. There may be negative outcomes to this. This sort of map use may

3.10 Feature Object Information

How the features/objects in the standard are defined and coded is examined in this section. The first item specifies how the features/objects are defined in terms of an internal dictionary, reference to external dictionary, user specified, or by some other means, and whether the definitions can be included in the transfer itself. The second item specifies whether the feature/object definitions are structured in a hierarchical, nonhierarchical, or by some other means. A following query asks how the feature/object are defined in terms of the real world, map scale, or some other way. The respondent is then asked to list the classes of feature/objects that have been defined, and the number of individual feature/object definitions provided for the user. Then the attention focuses on how the feature/objects are encoded in terms of numeric codes, alphabetic codes, alphanumeric codes, full names, or by some other means. All of the standards have some means of providing quality information about the feature/objects being transferred as part of the transfer process.

3.11 Attribute Information

This section focuses on just how the attributes are defined in the standard. The first item examines just how the attributes are defined: internal dictionary, reference to external dictionary, user specified, or by some other means, and whether the definitions can be included in the transfer itself. Attention then turns to how the attributes are structured in terms of hierarchical, nonhierarchical or on some other basis, and whether they are defined in terms of the real world, map scale, or some other means. Then the questions are asked concerning on just how the attribute name is encoded, and just how the attribute value is encoded by numeric, alphabetic, alphanumeric, full name, or some other means. The attribute types available are requested in terms of text, numeric, pictorial, user definitions, or by some other means. Attention then turns to whether it is possible to attach multiple values to a single attribute, and whether the range of possible values associated with the attributes are defined within the standard or somewhere else. The final query deals with the attribute information can be transferred at differing aggregation levels such as specified geographic areas, themes, individual elements, attributes or relationships. This information goes hand in hand with the feature/object information coded per the specifications in the previous section.

3.12 Relationship Information

This section focuses on how the relationships between the feature/objects are defined and transferred. The first item focuses on how the relationships are defined in terms of internal dictionary, external dictionary, user specified, or by some other means, and whether the definitions can be definitions in the transfer file. The attention then turns to how the relationships are defined in terms of the real world, map scale or some other way, and how many feature/objects can participate in a single relationship instance in terms of only two, or as many feature/objects are required. Whether the relationship types are explicitly defined as also of interest. Finally the assessor is asked to provide information as to how the relationship information is named and encoded.

3.13 Metadata Information

Metadata information is little understood by the average spatial data processing person, yet in the future this sort of information will be used very widely because it provides information about the data in the transfer file. The first query concerns just how the metadata is defined. Then several questions concerning how the metadata is encoded in terms whether it is limited to the quality information, whether the metadata can be transferred at differing aggregation levels, whether the standard supports pure metadata only transfers, and whether the structure of the metadata information is specified in terms of structured text, unstructured, text, code values, reference to external standards, or by some other means.

4 Summary and Conclusion

This paper has explored the scientific and technical characteristics associated with spatial database transfer standards. Initially a brief summary is provided of the basic concepts involved such as real/virtual maps, deep/surface structure, Nyerges data levels, and the concept of syntax/semantics. One must understand these concepts in order to fully understand and fully appreciate the scope and nature of these transfer standards. The organization of the work by the Commission has resulted in a set of rather specific goals being developed, and a brief discussion of the challenges facing the Commission. The result is a very extensive and detailed specification of 13 fundamental classes of characteristics with about 85 subcharacteristics and more than 220 detailed characteristics. The 13 fundamental classes of information are then summarized and discussed. These items are provided in full detail in a Commission Technical Report (Moellering and Wortman, 1994).

These characteristics and their detailed aspects provide an internationally developed and agreed means to assess any spatial database transfer standard on an even handed and uniform manner using material that has been defined on a consistent scientific and technical basis. As such these characteristics can be used to understand any particular standard in more detail and develop a better idea how that standard actually works. These characteristics are best used along with the standard document itself. These characteristics can also be used by an interested individual to compare and contrast two or more standards of interest and better understand the similarities and differences between them.