DeAnn Presley
Geography 795: Computer Mapping and GeoVisualization
Spring 2006
Dr. Shawn Hutchinson

Literature Review on Geovisualization
Introduction

Much was learned about the meaning and usage of the word “geovisualization” throughout the process of writing this literature review. An often cited definition is from a book chapter by MacEachren (1992) and thus seems like it is the accepted definition for the term.

“Geographic visualization will be defined here as the use of concrete visual representations—whether on paper or through computer displays or other media—to make spatial contexts and problems visible so as to engage the most powerful human information-processing abilities, those associated with vision.”

Geographic visualization has also been defined as “…the ability of maps, graphics, and images to make visible spatial relationships.”(Crampton, 2001). To do this, you find spatial patterns in data sets. It also allows for the three dimensional viewing of data. This can mean that the data itself is three dimensional, as in terrain or construction models, or that it is represented in three dimensions for ease in comprehension.

One other requirement for geovisualization is that it should be interactive (Crampton, 2002). If using paper or physical models, any user can interact with the model and use them to recognize known phenomena. In contrast, computer displays are more easily used for constructing new knowledge from data. Computer displays are also interactive in that the user can often easily make changes to the display.

Sections of the Literature Review

In this paper, the following outline will be used to organize the information learned.

1. Background Information: History, communication, and interactivity
2. Using geovisualization for communication: Physical models, flood risk
3. Using geovisualization for constructing new knowledge from data
   a. Applications
   b. Combined with other technologies

Background Information: History, communication, and interactivity


Baron Frederick W. von Egloffstein produced a map of the Grand Canyon in 1857-1858. This was the first map produced using shading for terrain. At the time, this was revolutionary. This paper discusses how new ideas for visualization and producing knowledge come about. The author calls this visual epistemology, which is an “…explicitly visual way of knowing…”(p.28). The author examines this concept by
considering the Egloffstein terrain map example. The author believed that this map reflects science and culture of the times. In the later 1800’s there was a great expansion of mapping taking place west of the Mississippi River. At this time, the science and culture reflected European Enlightenment and the influence of Alexander von Humboldt for a few reasons: Humboldt believed in basic research (field mapping) of natural resources, and in precise measurements. At this time the federal U.S. government was spending between 1/4 to 1/3 of the annual budget on exploration west of the Mississippi River between 1843-1856. The reports were published in “plain” language, while other reports were more “picturesque,” to enhance the look of nature.

Despite the fact that von Egloffstein’s map was revolutionary in approach, his methods and accuracy have been questioned by many. For one thing, maps made by him and by others do not compare well. Also, he made panoramic maps on flat paper, depicting 180 degree views. Thus, he was using visualization techniques, rather than strictly making maps. The idea of shaded relief maps is attributed to him, with the idea that such rugged landscapes inspired it. He then pioneered a method of creating shaded photographs from plaster terrain models, and published on the subject. It is important to note that he also created flat maps. By doing so, he made his shaded maps and panoramic drawings more credible—he knew what he was doing, but felt the flat maps didn’t allow for visualization.


The map communication model is the traditional theory of cartography, where the cartographer communicates information to the user via the map. In contrast, in geovisualization the “…mapping should proceed through multiple, competing visualizations which are not created by a cartographer and transmitted to the user but made on the spot by the user acting as his or her own cartographer.”(p. 236). In this method, presentation is less important than exploration of data. There may not be any one “best” map—it is up to the user to decide on the display. The user can be the cartographer. In a sense, “maps” are communication tools, and geovisualization uses spatial data for scientific purposes.

If a map is artwork, there is considerable amounts of influence placed upon the finished product by the cartographer. The map can be interpreted differently by different people. The map also may have an “agenda”, that is, what is the social construction behind making the map? Is it to convey power, as in a map of a territory held by a country? Is it to exaggerate a “problem” that needs to be solved by “map patrons?” (those paying for the map to be drawn). Maps, therefore, have power relations agendas that can not be resolved or removed from the map made by a cartographer.
Interactivity is one of the key aspects of geovisualization, although static maps may also be interactive. The author’s goal in writing this paper is to establish that geovisualization is more interactive than static maps, and more sophisticated. Interactivity requires a computer or technology, including being able to view, navigate, explore, browse, manipulate, etc. One example is being able to zoom in on a map. Another degree of interactivity allows the user to access the database and customize views. The third or deepest degree would allow the user to perform analysis or change the data. The author defines “interactivity in geovisualization simply as a system that changes its visual data display in response to user input” (p. 88). He also describes in detail the four types of interactivity, in order from lowest to highest levels of interactivity. The first type is interaction with the data representation. In this type, the user customizes the view by manipulation of lighting, orientation, viewpoint, zooming, re-scaling, or re-mapping the symbols. The second type is interacting with the temporal dimension. These maps are dynamic, allowing for fly-through or other types of basic navigation, where the user chooses a path through the map. Toggling and sorting also fit into this category. Interaction with the data is another type of interactive method. The user’s goal is to identify unknown patterns in the data. Brushing is done by making a stroke across the map, which selects the datapoints in that area, and creates a scatterplot of that data for the user to examine. Filtering removes inconsequential data, while highlighting accentuates the most relevant data points. (In both cases, relevance is determined by the user, usually using a statistical threshold). The last type of interactivity is contextualizing interaction. This refers to how the map or data are manipulated. One example is multiple views using multiple projections to select the best projection for the data. Combining data layers, window juxtaposition, and linking are additional types. The author mentioned ArcView and MapQuest as two types of interactive mapping environments. Both applications allow for interaction with the map than a traditional static map.

Using geovisualization for communication


Physical models are instantly accessible to the user, and comprehended easily by non-geographers. The author stated that there are certain industries and applications that prefer a physical 3-D model to a virtual 3-D or a 2-D representation on paper. Examples given were for the architecture, military, and fields associated with real estate transactions.

This article summarized the current technologies associated with creating physical terrain models from foam, wood, plastic, etc. The process of creating physical models is no longer time-consuming as it is largely automated. There are three main approaches: Subtractive, additive, and formative. Subtractive processes start with a solid object and cut away the excess foam, wood, or whatever the medium. Additive processes are more
expensive and can be approached in different ways, but they allow for overhangs, such as cliffs or roof overhangs, and for voids such as caves or interior rooms inside buildings. The approach is generally to spray a liquid plastic or resin onto a surface, and the machine that creates the model is basically a very sophisticated printer. The author didn’t provide much detail on formative methods, other than to say that it is accomplished by shaping materials using the “…application of opposing pressures” (p. 66). The author states that additive and subtractive methods are the most commonly used methods in the production of physical terrain models.

The remainder of the article discusses the cartographic decisions that must be made in the design of physical models, such as the size, resolution, vertical exaggeration, terrain characteristics, and material. The construction aspect also has issues, such as the artifacts created by the equipment, as well as the finishing (painting vs. engraving).


Virtual Reality Modeling Language (VRML) was developed for 3-D internet visualization. With the proper computer programming skills, the contents of a database could be linked and used with VRML to allow for internet geovisualization. In this paper, the authors demonstrate how the HyperText Preprocessor (PHP) open source scripting language could be used to create VRML webpages from text databases. They state that there is not much information available on how to do this, so their objective in writing this paper is to inform the reader on how to link a database to a VRML. Their application is the improved, more effective communication of flood-risk potential to the general public, many of whom are not skilled map users but might be more likely to understand 3-D visuals. The authors discuss the history of the development of VRML, which began in 1989. They also review the merits of a database, which include its simple and accessible nature, and the ease of updating and manipulation of the contents. By connecting to a VRML, up-to-date 3-D visualizations are quite possible.

The study area is located in Myrtleford, which is a regularly-flooded town of 3500 people in Australia and for which there is a variety of flood and water information available, including stream gauging station data. The PHP language uses the stream gauge water height to create a visual representation of flood potential on the landscape in a VRML scene. This is a dynamic process and the scene is updated whenever the database is updated. In summary, this technique will provide flood risk information to the general public using geovisualization.
Using geovisualization for constructing new knowledge from data

Applications


This goes back to the discrete versus continuous argument for the mapping of natural resources, specifically, soil types. In discrete mapping, each polygon belongs to only one class. In continuous mapping or modeling, an area on the landscape (pixel, for lack of a better word) can belong to multiple classes of membership. These classes might correspond to landscape position. In order to produce a final map of soil type that can be useful, it is necessary to ultimately de-fuzzify the data and classify the map based on the highest membership class. The resulting output is usually a raster with a color ramp. The authors of this paper propose using color mixing in the final map output instead. They used fuzzy k-means classification. The color mixing was done in each pixel, mixing red, green and blue colors to correspond to the class membership values. There is more to this method than simply mixing red, green and blue—it also has to do with the strength of the red, green, and blue colors for the strength of the class membership values.

Confusion index (CI) is the “…the difference, or sometimes the ratio, between the first and second highest membership class per pixel”(p.191). In this article, the authors say that CI is not a useful measure of how similar or different two things are—they can be very similar and yet have a high CI, and vice versa (p. 191). They propose calculating the difference from a class center instead, as it seems to give a better comparison for similarity between two classes. For example, an area that is in a transition zone between two soils on a landscape will likely have a high confusion index.

The study dataset was for Baranja Hill in Michigan. Six attribute maps were used in the landform classification: relative elevation to groundwater, slope gradient, wetness index, profile and tangent curvature, and viewshed. Fuzzy memberships were calculated, and maps were created in both the more traditional defuzzification method, and in color mixing. Defuzzified maps don’t really show the user much information about confusion. That is, each pixel ultimately is defuzzified down to being a member of only one set. In CM, the mixed colors themselves show transitional behavior between two classes.

Combining with other technologies (KDD and geocomputation)


As the title suggests, the authors propose combining technologies to construct knowledge from very large databases. Three contributing technologies are described as follows. Knowledge discovery in databases (KDD) is used for pattern recognition and
construction of new knowledge from large data sets. Geocomputation has most commonly been used to answer known questions (rather than recognize patterns and thus find new knowledge) and do so on a more spatial basis than KDD. The advantage to using geocomputation is “…a wealth of experience in machine-learning tools within a geographic setting” (p. 31). Exploratory visual analysis is used to recognize patterns and provide structure to those patterns, which is accomplished via the human-computer interface provided by the visual aspect. “In this case, the goals of portraying the data in visual form is to stimulate pattern recognition and hypothesis generation rather than simply presenting a result or outcome” (p. 31).

By using interactive visualization with these other techniques, the user could manipulate the display parameters in order to recognize and explain patterns in data sets, and at the same time in an integrated environment, use KDD and geocomputation to find new patterns or structure in the data set.


This paper discusses two methods of knowledge construction: Knowledge Discovery in Databases (KDD) and geovisualization. The author advocates using these two methods together to not only recognize patterns, but to understand the processes and structure that create patterns. KDD is a data mining method, while geovisualization is the use of visual representations. Both methods seek to construct knowledge from data as they are used for the “…identification and interpretation of spatio-temporal patterns in very large environmental data sets” (p. 470). Abductive reasoning has to do with producing new hypotheses. In a data set, this method would be used to group data into clusters or sets using a method, such as unsupervised classification. No pre-existing user-knowledge is required or utilized. The author supports using a paired KDD/geovisualization approach to simultaneously evaluate and visually represent data.

Inductive reasoning seeks to reduce uncertainty and is used to construct rules based on patterns, probabilities of class membership, and hierarchies. Examples of these methods are machine learning, rule induction, neural networks, genetic algorithms, case-based learning, and analytical learning (p. 476). In this mode of reasoning, the user is involved in the decision making process, but mostly in the application and development of rules. An example of this method is the use of decision trees. These are applied to data sets in the discovery of new patterns, which can be used to explain or provide structure to observations.

Deductive reasoning uses what we already know to locate patterns of interest in a data set. It is probably the most applied of the three modes of reasoning. It is often used to determine implications of existing knowledge and/or improve decision making. Space-time cube representations are straight-forward, in that the clusters often match what is already known about the data. However, the author discusses using integrated KDD and geovisualization approaches so that we don’t make mistakes in the interpretation of the
data. The author discusses one example in which the normalized difference vegetation index (NDVI) values for a given area in multiple years and a variety of land use/land cover types were plotted in a space-time cube. Normally, you’d expect the data to cluster into land use/land cover types, but there was confounding effect from the different years, i.e., the between-years variability was high with respect to the within-year variability between land use/land cover. In this example, much was already known about NDVI and vegetation, but the decision making process was improved through visualization.


Modeling has been used to evaluate the effects of silvicultural practices on sustainable forest management. In this study, modeling, GIS, and data visualization are combined. The authors want to use the combined technologies in exploratory spatial data analysis (ESDA) as a decision making tool in forestry, both spatially and temporally. The study site was located in Central European Russia. A forest ecosystem model called EFIMOD 2 was used to approximate various forest ecology parameters. CommonGIS was a program used for visualization and analysis of data. Simulations were run using known parameters, such as soil type and forest composition, and under four management practices. The simulation was run for 200 years. The authors used experts in the decision making process. The experts performed exploratory analysis of the simulation results, and then selected criteria and indicators for comparing the management practices. Their results were as predicted: Soil and forest parameters (for example, species composition) were most desirable under the least intensive silvicultural practices, and less desirable under the more intensive practices, one of which was clear-cutting. Soil carbon sequestration was highest under the protective management regime (such as in a national forest). Soil carbon levels were also higher under a selective cutting management regime, although this practice doesn’t really sequester much carbon as burning of the residue is allowed in this practice. The authors concluded that integrated modeling and geovisualization proved very useful in their first experiment, and that they will continue to refine the methods in future experiments.


In this paper, the author demonstrates how parallel coordinate plots (PCP) can be linked to maps and used in decision making processes. A PCP is useful in the display of results that pertain to a variety of characteristics. In a PCP, “…observations are represented…as a series of unbroken line segments, passing through parallel axes, each of which represents a different variable” (p. 606). A PCP allows for the representation of multivariate data in relation to all of the variables explored.
Two case studies are presented in this paper. In both, PCP was used with geovisualization. In the first study, PCP was used with KDD in climate modeling. In the second case study, PCP was used in the study of health statistics (cancer) data. In both cases, the volume of multivariate data was tremendous, and georeferenced, hence the use of PCP and geovisualization. By “brushing” the PCP graph, which means to click and drag a box over a given section of interest on the plot, the user can explore the data set itself and examine statistics on that selected region on the PCP plot. That is, rather than wait for the statistical output and final result, the user can interactively examine the data for things such as outliers, or for dependency between variables that probably were not anticipated by the user. Again, the idea is that the user is looking for new knowledge, rather than looking for the information that might be already known. For example, in the second study it was observed that areas with very low uncertainty on the PCP plot were for large cities and that lung cancer occurred at a consistently high rate in these areas. Low uncertainty means a high probability. However, for the same variables, prostate cancer was not consistently high. This could lead the user to more closely examine the environmental factors in the data set for lung cancer, such as those variables affecting air quality in cities.

Summary

Geovisualization’s beginnings are traced to the creation of a terrain relief map by Baron von Egloffstein in the 1800’s. He was perhaps the first person to realize the advantages of using communication techniques to enhance the appearance of a flat, paper map. Physical models are often useful to portray information about known phenomena, such as a terrain model or a construction model. Geovisualization is different from a flat, paper map in that it allows the user to interact with the data to either comprehend known phenomena, or to extract new knowledge from data sets. Geovisualization has been paired with many other computing techniques to extract and apply new knowledge. It has been paired with modeling, geocomputation, knowledge discovery in databases, and parallel coordinate plots.

Possible Future Research

The expanded use of 3-D internet geovisualization to communicate information to the public might be one of the most important future research areas for geovisualization. Basic and Nuwantwee (2004) used database contents and VRML to create a flood-risk website for public use. Similar climatic information could be used in the Great Plains to assess and minimize various risks associated with production agriculture.

Also, the expanded use of paired geovisualization and KDD in natural resources mapping/management has a lot of untapped potential for use in multidisciplinary scientific studies, such as the use of multiple data sources in global climate modeling.
References


